Genetic properties underlying transcriptional variability across different perturbations

Saburo Tsuru[1,*] & 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

*Corresponding authors:
Saburo Tsuru
E-mail: saburotsuru@gmail.com
Chikara Furusawa
chikara.furusawa@riken.jp

Abstract

The rate and direction of phenotypic evolution depend on the availability of phenotypic variants induced genetically or environmentally. It is widely accepted that organisms do not display uniform phenotypic variation, with certain variants arising more frequently than others in response to genetic or environmental perturbations. Previous studies have suggested that gene regulatory networks channel both environmental and genetic influences. However, how the gene regulatory networks influence phenotypic variation remains unclear. To address this, we characterized transcriptional variations in *Escherichia coli* under environmental and genetic perturbations. Based on the current understanding of transcriptional regulatory networks, we identified genetic properties that explain gene-to-gene differences in transcriptional variation. Our findings highlight the role of gene regulatory networks in shaping the shared phenotypic variability across different perturbations.

Introduction

The rate and direction of phenotypic evolution depend not only on natural selection but also on the availability of phenotypic variants, whether induced genetically or environmentally, upon which selection may act[1]. Even environmentally induced phenotypic variability or plasticity has the potential to initiate and direct adaptive evolution through processes such as genetic assimilation[2]. It is widely acknowledged that organisms exhibit heterogeneous phenotypic variation, with certain variants emerging more frequently than others in response to genetic or environmental perturbations[3, 4, 5, 6]. This bias in phenotypic variability can facilitate and constrain adaptive evolution across different phenotypic dimensions[7], contingent on the alignment between the bias in phenotypic variability and the direction of selection[8]. Thus, without a thorough understanding of the nature of phenotypic variability, our understanding of why evolution has progressed in the manner that it does or why some phenotypes evolve repeatedly while others are one-offs remains incomplete.

Seemingly different types of phenotypic variability may actually be driven by a common mechanism. Theoretical studies have demonstrated that the bias in phenotypic variability is shared by different perturbations[9, 10, 11, 12, 13, 14]. Notably, several quantitative genetic studies identified similarity in the magnitude of variance, where traits sensitive to environmental perturbations tend to be sensitive to genetic perturbations[15, 16, 17, 18]. Additionally, there is a consistent pattern of directional variation among traits, with organisms frequently exhibiting correlated variation between traits[19]. Environmental influences on traits often act in the same direction as genetic influences, thereby shaping phenotypic variation in a consistent direction[6]. These similarities imply the existence of a common mechanism underlying the bias in phenotypic variability against any kinds of perturbations[12, 13, 14, 20, 21].

What molecular mechanisms bias the phenotypic variability? In biological systems, most traits arise from genetic interactions within networks represented by gene regulatory networks[22]. Genetic and environmental perturbations exert phenotypic effects via identical regulatory networks[23]. Consequently, phenotypic variance is expected to be influenced by the properties of the regulatory networks, regardless of the underlying causes of perturbations[8, 10]. Despite a growing number of studies demonstrating empirical evidence supporting the similarity both in susceptibility and in directionality of variations in gene expression levels between environmental and genetic perturbations[17, 18, 24], it remains largely unknown how the real regulatory networks account for the observed bias in variability among genes[8].

To address this issue, we characterized the variability of transcriptome profiles in *Escherichia coli* in response to different environmental and genetic perturbations. *E. coli* is an ideal model organism for this investigation because of its well-characterized gene regulatory network. Our analysis revealed that genes showing higher transcriptional variability against environmental perturbations tended to have higher sensitivity to genetic perturbations, indicating a shared gene-to-gene bias in variability across different types of perturbations. In addition, the major directions, or the top principal

components, of correlated variation among thousands of genes were shared by different perturbations, demonstrating a common bias in the directionality of transcriptional changes. Most importantly, we found that the transcriptional regulators account for these common biases. These empirical findings support the relevance of the gene regulatory networks in explaining the bias in phenotypic variability shared across different perturbations. Our results suggest that these biases are not random but are functionally significant in coping with fluctuating environments.

Results

Similarity in transcriptional variance between environmental and genetic perturbations

We initially analyzed two public transcriptome datasets to evaluate the transcriptional variability across different perturbations. To this end, we curated transcriptome profiles from the PRECISE database[25, 26], focusing on a single strain of *E. coli*, K12 MG1655, cultured under different environmental conditions (160 profiles, 76 unique environmental conditions) (**Supplementary Table 1, Supplementary Fig. 1a,b**), termed the Env dataset (**Fig. 1a**). Additionally, to evaluate transcriptional variability in response to genetic perturbations via natural evolution, we used transcriptome profiles from 16 natural strains of *E. coli* cultured under an identical environmental condition[27], termed the Evo dataset (**Fig. 1b**). These isolates spanned the most common phylogroups and lifestyles of *E. coli*. Importantly, genetic mutations among these strains might reflect their diverse natural habitats, potentially confounding the estimation of transcriptional variability in response to genetic perturbations independent of environmental perturbations. To mitigate this, we conducted a mutation accumulation (MA) experiment[28] using a mutator strain of *E. coli*, K12 MG1655, lacking the *dnaQ* gene[29, 30] (**Fig. 1c**). Here, mutations primarily accumulate through genetic drift owing to small population bottlenecks, thus minimizing positive selection. We established two founder colonies (founders A and B) and propagated single colonies on lysogeny broth (LB) agar plates (yielding three MA lineages per founder). Consequently, we obtained transcriptome profiles from six MA lineages cultured in LB broth using the reliable DNA microarray technique[31] (**Supplementary Fig. 1c,d**), termed the Mut dataset (**Supplementary Tables 2**,**3**). Importantly, each dataset was generated within single laboratories using a common protocol, minimizing technical variation within each dataset. We used quantile-normalized and $\log_2$-transformed transcriptome data for the analysis.

The means and standard deviations of expression levels were computed for each gene across each dataset (**Fig. 1d** for six representative genes). While standard deviation might directly measure transcriptional variability, it can be influenced by the mean[32], leading to a confounding effect in estimating similarities in transcriptional variability between datasets. Notably, the standard deviations exhibited a statistically significant mean dependency for all datasets (**Fig. 1e**). To alleviate this influence, we calculated the vertical deviation of each standard deviation from a smoothed spline of the running median of the standard deviations (DM), which served as a measure of the magnitude of transcriptional variability. The DM values of each gene in the dataset was designated as DM_{env}, DM_{evo}, DM_{mut} for the Env, Evo, and Mut datasets, respectively (**Fig. 1f**, **Supplementary Table 4**).

Comparing DM values between datasets revealed positive correlations (Spearman's rank correlation coefficient (R) =0.43–0.56, $p<0.05$) (**Fig. 2a**). These results indicate that genes with higher sensitivity to environmental perturbations tend to have higher sensitivity to genetic perturbations in terms of their expression levels. This phenomenon of canalization/decanalization could occur not only at the gene level but also at more complex phenotypes subject to natural selection. Canalization/decanalization in transcriptional changes has been observed in some functional categories

of genes in yeast[18, 24]. For instance, genes involved in stress or amino acid metabolism tend to have higher transcriptional variability [24], while genes related to cell growth, maintenance, and cell cycles do not exhibit higher transcriptional variability against genetic perturbations [18]. To explore whether functional associations in DM values were present in bacteria, we utilized a refined subset of Gene Ontology (GO) terms as functional categories (**Methods**) (**Supplementary Table 5**). Using the mean DM values of genes within each GO term as a measure of transcriptional variability, we observed positive correlations between the datasets even at the functional category level (**Fig. 2b**). To further investigate functional associations in DM values, we performed gene set enrichment analysis (GSEA[33]) with GO terms for both higher and lower DM values for each dataset independently. We found several GO terms significantly enriched in higher or lower DM values (BH-adjusted p-value (q)<0.05, GSEA). For instance, we identified genes related to 13 GOs that were significantly overrepresented at higher DM values compared with a uniform distribution. We identified several common GO terms across all datasets (**Fig. 2c**, **Supplementary Fig. 2**). For higher DM values, more than three GOs were found at every intersection between pairs of datasets. This overlap was significantly more than expected (p<0.0001) based on random drawing from a pool of 81 GOs without replacement. Notably, genes related to flagellum-dependent cell motility and amino acid catabolic processes were consistently enriched with higher DM values in all three datasets (q<0.05, GSEA). Tuning the activity of these functional categories in response to environmental changes is crucial for cells to cope with nutritionally fluctuating environments[34, 35]. In contrast, genes related to fundamental biological processes related to cell-wall construction and maintenance were consistently enriched at lower DM values for all datasets (q<0.05, GSEA, **Fig. 2c right**), indicating that the expression levels of genes within these categories were relatively robust to perturbations. The number of these common GOs (7) in lower DM values was significantly larger (p<0.0001) than expected (0.06) based on random drawing from a pool of GOs without replacement. We obtained relevant results using the functional categories defined by the KEGG pathways (**Supplementary Fig. 3**,**4**, **Supplementary Table 6**). There were positive correlations in DM values at the pathway level between the three datasets (Spearman's R=0.5–0.67, p<0.05). Additionally, genes related to chemotaxes and peptidoglycan biosynthesis were overrepresented at higher and lower DM values, respectively, for all datasets (q<0.05, GSEA). These KEGG pathways were closely related to flagellum-dependent cell motility and cell-wall construction in GO terms, respectively. These results indicate frequent associations between canalization/decanalization in transcription level of genes and their functions, suggesting a certain common mechanism underlying the heterogeneous transcriptional variability associated with functional categories.

Transcriptional regulators influencing transcriptional variability

To explore the underlying mechanism, we examined the influence of transcriptional regulators on transcriptional variability. First, we conducted GSEA for each dataset independently based on known regulatory interactions (**Supplementary Table 7**) to identify transcriptional regulators that affect target genes exhibiting higher and lower DM values, termed TR_{high} and TR_{low}, respectively. We analyzed two

147 golden standard databases, RegulonDB[36] and EcoCyc[37], collecting known regulatory interactions
148 between genes. We focused on the transcription factors and sigma factors that regulate more than four
149 and less than 1001 target genes. For the Env dataset, we identified 30 transcriptional regulators
150 enriched in higher DM_{env}, such as FlhDC and Fur, and termed them TR_{env} ($q<0.05$, GSEA, **Fig. 3a**).
151 This enrichment means that the target genes of these regulators were overrepresented at higher DM_{env}
152 compared with a uniform distribution. In contrast, we identified only four transcriptional regulators
153 enriched in the lower DM_{env} ($q<0.05$, GSEA). Similar analyses were performed for DM_{evo} and DM_{mut},
154 independently (**Supplementary Fig. 5**). Consequently, we identified 29 (TR_{evo}) and three
155 transcriptional regulators enriched in higher and lower DM_{evo}, respectively. Similarly, 28 (TR_{mut}) and
156 two transcriptional regulators were enriched in the higher and lower DM_{mut}, respectively ($q<0.05$,
157 GSEA). Thus, fewer transcriptional regulators were overrepresented at lower DM values, suggesting
158 that regulation by transcriptional regulators tends to increase the transcriptional variability of target
159 genes, with some regulatory motifs, such as negative feedback, potentially repressing fluctuations in
160 gene expression, as a minor exception[38].

161 Next, we explored the similarity between TR_{env}, TR_{evo}, and TR_{mut}, which were identified in
162 each dataset independently. We found that 13 transcriptional regulators, referred to as TR_{com}, were
163 common among the three datasets, corresponding to more than one-third of the transcriptional
164 regulators enriched in each dataset (**Fig. 3b**). The number of TR_{com}s (13) was significantly larger
165 ($p<0.0001$) than the expected number (1.7) for the intersection of the three sets, constructed by random
166 drawing from a pool of 118 TRs without replacement. These results suggest that the target genes
167 regulated by TR_{com} exhibited higher transcriptional variability in response to both environmental and
168 genetic perturbations, indicating that TR_{com}s function as a common mechanism to bias transcriptional
169 variability against various perturbations.

170 **Network signatures of transcriptional regulators influencing transcriptional variability**

171 How do TR_{com}s enhance transcriptional variability of target genes against a variety of perturbations
172 both inside and outside of the cells? We found that TR_{com}s are global regulators with a large number
173 of target genes, including TR_{com}s themselves (**Fig. 3c**). Based on these findings, we speculated that
174 TR_{com}s might regulate each other frequently, which could increase the probability of the propagation
175 of different perturbations to their target genes. To test this hypothesis, we explored the regulatory
176 structure of a network comprising TR_{com}s alone (**Fig. 3d**). We found that most TR_{com}s had both an in-
177 degree and an out-degree (11 out of 13 regulators). FliA had no out-degrees but four in-degrees. Only
178 IscR was isolated with no regulatory interactions within TR_{com}s. This network structure indicates that
179 the network within TR_{com}s is not top-down structured but has many mutual interactions (activations
180 and inhibitions) within TR_{com}s. The edge density within the TR_{com}s (**Fig. 3e**) was remarkably high
181 (0.17), with statistical significance ($p<0.0001$), compared to the edge density of random expectations
182 (0.019). These results indicate that the TR_{com}s form a relatively densely connected network that is
183 proficient in propagating the influence of perturbations within this group.

184 Next, we explored the effects of TR_{com}s on transcriptional variability. First, we compared the
185 DM values of the target genes with TR_{com}-containing regulation with those of the target genes with
186 TR_{com}-less regulation (**Fig. 4a**). The target genes of TR_{com}s tended to have higher DM values in all
187 datasets than other genes with the same number of unique regulators. We found none or very weak
188 positive correlations between the DM values of the genes without regulations by TR_{com}s and the
189 number of unique regulators that the genes received in all datasets (Spearman's R=~0.13). On the other
190 hand, the target genes of TR_{com}s exhibited relatively high positive correlations in all datasets
191 (Spearman's R=0.23–0.35, $p<0.05$). These results indicate that the recruitment of TR_{com}s is more
192 effective in enhancing transcriptional variability. To further investigate this hypothesis, we explored
193 the relationship between DM values and the number of unique TR_{com}s among the genes with a fixed
194 number of regulators. In the case of the genes that received five unique regulators (**Fig. 4b**), we found
195 a positive correlation between DM_{env} and the number of unique TR_{com}s among the five regulators
196 (Spearman's R=0.29, $p<0.05$). We performed similar analyses for all the datasets and found pervasive
197 positive correlations across broad classes of the number of unique regulators (**Fig. 4c**). These results
198 highlight the role of TR_{com}s in enhancing transcriptional variability through dense mutual regulation
199 within TR_{com}s and co-regulation by multiple TR_{com}s, likely because both increase *trans*-target sizes[18]
200 that affect the expression levels of the genes regulated by TR_{com}s in response to both environmental
201 and genetic perturbations.

202 To understand how TR_{com}s access a variety of perturbations, we explored the relationship
203 between TR_{com}s and the known molecular machinery for sensing stimuli both inside and outside the
204 cells. First, we performed bow-tie decomposition of the TRN to identify the subnetworks to which the
205 TR_{com}s belonged (**Fig. 5a–d**, **Supplementary Table 8**). The *E. coli* TRN was formally dissected into
206 seven sectors defined by Yang's bow-tie structure[39] based on the directed regulatory links between
207 genes[40]. The primary sectors are IN, OUT and the greatest Strongly Connected Component (SCC).
208 The SCC sector represents the subnetwork, in which each node (TR) is reachable by any other node
209 following the direction of regulation from the TR to its target. The IN sector comprises TRs that are
210 not included in the SCC and can reach TRs in the SCC. The OUT sector comprises nodes (TRs or non-
211 regulatory genes) not included in SCC but reachable by TRs in SCC. The numbers of nodes in these
212 primary sectors in *E. coli* TRN were seven TRs, 62 TRs and 2715 genes including the TRs for IN,
213 SCC, and OUT, respectively (**Fig. 5e**). As expected from their many target genes, TR_{com}s (12 of 13
214 TR_{com}s) and TR_{high}s (26 of 48 TR_{high}s in Env, Evo, and Mut) belonged to SCC (**Fig. 5d**). This implies
215 that TR_{com}s can receive broad perturbations from approximately 70 genes in the IN and SCC sectors.
216 We noted that TR_{com}s in SCC still showed a higher edge density (0.19) than other TRs in SCC
217 ($p<0.005$), supporting the especially dense connections within the TR_{com}s.

218 Next, we investigated how TR_{com}s sense different intracellular and extracellular stimuli. *E.*
219 *coli* employs at least 145 sensory regulators to transmit various stimuli to the TRN (**Fig. 5b**,
220 **Supplementary Table 8**). These regulators, termed SRs, change their regulatory activity depending

on their interactions with corresponding effectors, as summarized by Martínez-Antonio *et al*[41]. For visualization, we defined the STIMULI block as an aggregated group of the known effectors that interact with SRs, such as nutrients, stressors and metabolites. We found that the IN and SCC sectors contained 39 SRs (27% of the total SRs, **Fig. 5f**), which was reasonable agreement with the random expectation based on the number of TRs in these sectors. Interestingly, more than half of TR_{com}s (8 of 13) were SRs. These results indicate that TRs in the SCC sectors, TR_{com}s included, not only serve as intermediate managers in information processing but also play a role in sensing different kinds of stimuli. These characteristics of TR_{com}s, particularly their presence in SCC with many SRs, may enable TR_{com}s to access a variety of stimuli both inside and outside the cells.

Similarity in direction of transcriptional variance between environmental and genetic perturbations

Next, we explored the major directions of correlated transcriptional variation among thousands of genes. Theoretically, genetically caused phenotypic variance is expected to be the greatest along the major axis or the principal component of environmentally caused phenotypic variance[10, 12, 13]. To identify such principal components (PC) and test this prediction, we performed principal component analysis (PCA) for each dataset (**Fig. 6a**, **Supplementary Figs. 6–8**). The first three principal components (PC1–3) explained 46%, 53%, and 57% of the total variance in the Env, Evo, and Mut datasets, respectively, indicating their significance as the major axes of transcriptional variation for each dataset. Correlation analysis based on the loadings of these major axes revealed similarities in their directions across the datasets (**Fig. 6b**). We used the absolute values of Spearman's R because of the relative nature of the signs of the principal components. Specifically, PC1 in the Env dataset exhibited relatively strong correlations with PC1 and PC2 in the Evo dataset (Spearman's R = 0.45 and 0.41, respectively, $p<0.05$) and with PC1 and PC2 in the Mut dataset (Spearman's R = 0.34 and 0.41, respectively, $p<0.05$). Similarly, PC1 in the Evo dataset showed a relatively strong correlations with PC1–3 in the Env and Mut datasets. These correlations in loadings were much stronger than expected (~0.03 in Spearman's R) based on independent datasets comprising no correlations in expression levels between genes (**Supplementary Fig. 9c**). To validate the high similarity of the PCs between the three datasets more formally, we assessed the extent to which the principal components in the Env dataset explained the transcriptional variance in the other two datasets. Overall, all principal components in the Env dataset explained up to 37% and 39% of the total variance in the Evo and Mut datasets, respectively (**Fig. 6c**). These fractions were much larger than random expectations (~3%, **Supplementary Fig. 9d**). Moreover, the cumulative variance plot displayed a convex upward curve for both the Evo and Mut datasets, as well as Env, which was distinct from a linear-like curve in random expectations (blue line in **Supplementary Fig. 9d**). These findings support the notion that the major directions of transcriptional variation in response to environmental perturbations are considerably similar to those observed in response to genetic perturbations, despite a large number of genes.

A directional bias in transcriptional variations is expected to reflect transcriptional

258 regulations[8]. To validate this prediction, we explored the transcriptional regulators that contribute to
259 the major directions. Utilizing the loadings of PC1–3 in each dataset, we conducted GSEA (**Fig. 6d**,**e**).
260 We found the transcriptional regulations enriched in the positive and negative directions of PC1–3 in
261 the Env dataset (**Fig. 6d**). For instance, activation by FadR and FlhDC or inhibition by DksA and PdhR
262 was enriched in the positive direction of PC1. Conversely, activation by RcdA and IscR or inhibition
263 by CecR and MarA was enriched in the negative direction of PC1. We conducted a similar GSEA for
264 the other two datasets (**Supplementary Fig. 10**), identifying 31 regulations common across all three
265 datasets (**Fig. 6e**), including 11 regulations by TR_{coms}, which is consistent with the notion that TR_{com}s
266 are global regulators collectively controlling numerous targets. The number of common regulations
267 (31) was significantly larger ($p<0.0001$) than the expected number (0.2) for the intersection of the three
268 sets constructed by random drawing from a pool of 267 regulations without replacement. Overall, these
269 findings support the idea that the similarity in the direction of transcriptional changes in response to
270 different types of perturbations arises from shared transcriptional regulation. To further confirm the
271 relevance of these findings, we analyzed transcriptome profiles of *E. coli* mutants harboring rewired
272 transcriptional regulations[42] that perturb native transcriptional regulations between major global
273 regulators (**Supplementary Fig. 11**). Each of the rewired strains has a unique shuffled synthetic operon
274 comprising a plasmid copy of the native promoter region of a global regulator ('A') and a plasmid copy
275 of a gene coding for another global regulator ('B'). Accordingly, promoter-mediated transcriptional
276 regulations controlling expression of 'A' are newly connected to 'B' without mutating native
277 chromosomal regulations. Consequently, these rewired strains have new regulatory "links" between
278 randomly selected pairs of global regulators, perturbing native transcriptional regulations between
279 global regulators. The transcriptome profiles of the rewired strains were originally obtained by
280 Baumstark *et al*[43] under identical environmental conditions. Using this dataset, termed the Rwr dataset,
281 we performed PCA and explored how the directionality in transcriptional variance differed from the
282 common directionality shared by the three main datasets (**Supplementary Fig. 11c,d**). We confirmed
283 that the top two PCs in Rwr showed lower correlations (~0.18 at most) with those in the three main
284 datasets, which was very different from the higher correlations between the three main datasets (**Fig.
285 6b**). Relatively higher correlations were found between the third PC in Rwr and the top two PCs in the
286 three main datasets. Consistently, the PCs defined by the Env dataset explained at most 29% of the
287 variance in the Rwr dataset, which was lower than those of the Mut and Evo datasets (**Supplementary
288 Fig. 11e**). These results indicated that the directionality in transcriptional variation between rewired
289 strains was less similar to the common directionality shared in the other datasets. These results
290 highlighted the importance of regulatory networks between global regulators in major directions in
291 transcriptional variability, supporting the relevance of our findings.

292 Furthermore, functional enrichment in GO terms was observed in the top three PCs across all
293 three datasets (**Fig. 6f**, **Supplementary Figs. 12–14**). The most major direction in Env (PC1) revealed
294 enrichment of translation, cell division, and cell-wall-related biological processes in one direction
295 (positive sign), and responses to starvation and amino acid catabolic/biosynthetic processes in the

opposite direction (negative sign). We termed these opposing functional families growth and starvation families, respectively, because of their close association with cellular growth and starvation states in response to environmental conditions[44]. The correlation between PC1 in Env and PC1 and PC2 in Evo and Mut suggested that approximately half to one-thirds of the functional categories in both families were enriched in these PCs, with opposite directions between the families retained across datasets. Importantly, the mutually opposite directionality observed in these functional families tended to be preserved across different perturbations, suggesting a consistent response pattern at the functional category level (**Fig. 6g**).

Similarity of phenotypic variability across various perturbations

Heterogeneous phenotypic variability is not a random occurrence devoid of biological significance (**Fig. 2c** and **6f,g**), but rather it may result from or be facilitated by the adaptive generalization of past environments[9, 10, 11, 12, 13]. In addition, several studies suggest that evolution tends to produce global canalization that buffers phenotypes against not only genetic and environmental perturbations[12, 13, 21], but also perturbations caused by stochastic molecular noise[45, 46]. These findings imply a similarity in variability in gene expression levels across a broader range of perturbations than previously examined. For instance, genes with higher/lower DM values in Env might also have higher/lower transcriptional variability in response to other environmental perturbations or other nongenetic perturbations (molecular noises) untested in the Env dataset. Similarly, genes with higher/lower DM values in response to random mutations accumulated through genetic drift (Mut) might exhibit higher/lower DM values in response to beneficial mutations accumulated by strong selection in adaptive evolution. To address these possibilities and validate a broader applicability of the findings from the main datasets, we analyzed five additional datasets for *E. coli*, termed Strs_hp, Strs_he, Strs, Prtn, and Noise_ch (**Fig. 7a**, **Supplementary Figs. 15–22**, **Supplementary Note**). These datasets include transcriptome or proteome profiles obtained under relatively narrower or more specific conditions than the three main datasets (Env, Evo, and Mut), but under several perturbations absent in the three main datasets, except for Prtn. The Prtn dataset, which contains several environmental conditions similar to those used in Env, was used to test whether the bias in transcriptional variability was retained at the protein level. The phenotypic variabilities based on these datasets, either at the transcriptional or protein level, were termed DM_{strs_hp}, DM_{strs_he}, DM_{strs}, and DM_{noise_ch} as detailed in **Supplementary Note**. In short, DM_{strs_hp} and DM_{prtn} show variability in their expression levels (mRNA for Strs_hp and protein for Prtn) in response to various environmental perturbations. DM_{strs_he} and DM_{strs} represent transcriptional variability in response to genetic mutations accumulated through adaptive laboratory evolution under various stressors. DM_{noise_ch} represents the protein noise of clonal populations under identical environmental and genetic conditions. We found that these DM values were positively correlated with those of the three main datasets (**Fig. 7b**). Notably, the correlations with DM_{env} show relatively strong in all pairs (Spearman's $R = 0.31–0.68$, $p<0.05$). Next, we conducted PCA to identify the directionality of each dataset. To understand the similarity in the major directions across the datasets, we compared PC1–3 of these additional datasets with those of the three main datasets (**Fig. 7c**). Correlation analysis

using the loadings of PC1–3 revealed that PC1s of the additional datasets tended to correlate with PC1s or PC2s of the three main datasets. In particular, PC1 in the Env dataset showed the strongest correlation with the PC1s of the additional datasets. Thus, transcriptional variability tended to be retained at the protein level, as shown in the Prtn dataset, suggesting the impact of transcriptional regulation on the variability in phenotypes governed by proteins. The similarity in transcriptional variability with Strs_hp supported the impact of the transcriptional variability identified in the three main datasets on broad environmental conditions. Importantly, the similarity with Strs and Strs_he highlighted the impact of transcriptional variability under relaxed selection (Env or Mut) on the evolutionary diversification in transcriptional phenotypes accompanied by adaptive evolution. Interestingly, the transcriptional variability in response to mutations (Mut) was also coupled with stochastic protein noise (Noise_ch), which was previously reported in yeast [18]. Taken together, these results indicated similar trends in gene-to-gene differences in both the magnitude and major directions of expressional variations across various perturbations, suggesting global canalization/decanalization across different perturbations[12, 13, 14, 21, 45].

Discussion

In this study, we characterized the transcriptional variability of *E. coli* against different perturbations and explored the regulatory commonalities between genetic and environmental perturbations. We identified common global transcriptional regulators that underlie the shared variability across different perturbations. These transcriptional regulators interact relatively densely with each other and strongly connect with many sensory regulators, forming a network structure that is favorable to facilitate the frequency of propagation of various stimuli both inside and outside the cells. In addition, the more regulated genes by these influential regulators showed greater changes in expression effectively. PCA revealed that these global transcriptional regulators simultaneously induced changes in the expression levels of multiple genes and contributed significantly to the formation of major directions of genome-wide changes in expression levels. These results provide implications for the molecular mechanisms behind non-uniform phenotypic variability. The observed transcriptional bias was consistent across diverse conditions, including adaptive evolution, and was closely related to crucial functions for coping with fluctuating environments, suggesting its pervasive impact on both genetic and nongenetic adaptation to environments.

A major limitation of our study is that it focused solely on the transcriptome regulatory networks of a single strain of *E. coli* (MG1655 including its closely-related strains). Currently, most of the reliable knowledge about transcriptional regulatory networks in *E. coli* is heavily restricted to this model strain. Ideally, gathering comprehensive information on the networks of different strains of *E. coli* or other bacterial species with different configurations of transcriptional regulatory networks would enable us to explore how transcriptional variability depends on their transcriptional regulatory networks. This future exploration would also require many transcriptomic profiles under different environmental conditions for each strain to capture the transcriptional variability unique to each. This extension would further underscore the impact of transcriptional regulatory networks on the magnitude and directionality of transcriptional variability.

375 # Methods

376 **Bacterial strains and media**

377 We constructed a mutator strain of *E. coli* MG1655 by deleting the *dnaQ* gene. Deletion was performed
378 using conventional λ RED recombination[47]. In brief, a kanamycin resistance gene flanked by FLP
379 recognition sites was amplified from *E. coli* BW25113 *ΔmoaA::kan*, a strain of the Keio collection[48],
380 by PCR with chimeric primers containing homology arms for the *dnaQ* region. We used the same
381 chimeric primers as those used to knock out *dnaQ* in a previous study[48]. The PCR fragment was used
382 for lambda RED-mediated homologous recombination with pKD46. The obtained strain after
383 obviating pKD46 was called as MG1655 *ΔdnaQ* and used for the MA experiment. We used YENB
384 broth[49] supplemented with 60 mM arabinose (FUJIFILM Wako, Osaka, Japan, 012-04581) and 50
385 μg/mL ampicillin (FUJIFILM Wako, Osaka, Japan, 012-23303) to prepare electrocompetent cells for
386 λ RED recombination. LB agar plates (Miller, BD Difco, 244620) supplemented with 25 μg/mL
387 kanamycin (FUJIFILM Wako, Osaka, Japan, 115-00342) were used as the selective media after
388 electroporation. We used LB agar and broth for the MA experiments and transcriptome analysis of the
389 mutants, respectively.

390

391 **MA experiments**

392 The MA experiments were performed as previous studies[30]. In brief, we used two independent colonies
393 of MG1655 *ΔdnaQ* as founder strains, founder A and B, to alleviate a potential bias of the mutations
394 before the MA procedure that might affect the accumulated mutations of the descendants during the
395 MA procedures[50]. We initially established six independent lineages for each founder in preparation for
396 extinction during MA experiments. The MA experiments were conducted by transferring randomly
397 chosen single colonies to fresh LB agar plates daily. The plates were incubated at 32 °C. The MA
398 procedure was repeated for 30 times. The MA lineages after the 10th and final rounds were stored at –
399 80 °C. Three randomly selected three MA lineages from each founder were subjected to transcriptome
400 analyses (**Supplementary Table 2**).

401

402 **Transcriptome profiling using microarray technology**

403 The detailed procedure for transcriptome profiling has been described in previous study[31]. All strains
404 obtained by the MA experiments were grown in 200 μL of LB medium in 96-well microplates at 32 °C.
405 The cell density was monitored using a 1420 ARVO microplate reader (PerkinElmer Inc., Waltham,
406 Massachusetts, USA). The cell broth was sampled at the exponential phase, when OD_{600} reached in
407 the 0.072–0.135 range. The cell broth was immediately added to an equal volume of ice-cold ethanol
408 containing 10% (w/v) phenol. The cell pellets were obtained by centrifugation at 20,000 × g at 4 °C
409 for 5 min. Total RNA was extracted using an RNeasy Mini kit with on-column DNase I digestion for
410 15 min (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Purified RNA samples
411 were analyzed for quality control using an Agilent 2100 Bioanalyzer and an RNA 6000 Nano kit
412 (Agilent Technologies). Microarray experiments were performed using a custom-designed Agilent

413 8 × 60 K array for the *E. coli* W3110 strain, comprising 12 probes for each gene. Purified total RNA
414 samples (100 ng) were labeled with Cyanine3 (Cy3) using a Low Input Quick Amp WT labeling kit
415 (One-color; Agilent Technologies). The Cy3-labeled cRNA samples were checked for amount
416 (>825 ng) and specific activity (>15 pmol/μg) using a NanoDrop ND-2000 (Thermo Scientific).
417 Subsequently, the labeled cRNA (600 ng) was fragmented and hybridized to a microarray for 17 h
418 while rotating at 10 rpm at 65 °C in a hybridization oven (Agilent Technologies). The hybridized
419 microarray was washed and scanned according to the manufacturer's instructions. Microarray images
420 were analyzed using Feature Extraction version 10.7.3.1 (Agilent Technologies). Duplicate biological
421 data of the founder strains under the same culture conditions were obtained to confirm technical noise.
422 The compendium of transcriptome data obtained from the MA experiment is called the Mut dataset.
423
424 **Compilation of the TRN in *E. coli***
425 We compiled the reported TRN using all interactions with experimental evidence from RegulonDB
426 11.0[36] and EcoCyc[37] for both transcription factors and sigma factors (**Supplementary Table 7**). For
427 transcription factors, the mode of regulation (activation or repression) was also included. We set the
428 mode of regulation by the sigma factor to activation. When multiple modes were reported for a given
429 regulatory mechanism between a transcription factor and its target gene, they were designated as
430 unknown.
431
432 **Analysis of network architecture of TRN using bow-tie decomposition**
433 First, we analyzed the architecture of TRN in *E. coli* based on bow-tie decomposition[40]. We followed
434 the formal definition of the bow-tie structures proposed by Yang *et al*[39]. Their bow-tie structure has
435 seven sectors. The three main sectors are IN, OUT, and SCC. The SCC sector consists of the largest
436 subnetwork in which each node (gene or transcriptional regulator) is reachable by any other node,
437 following the direction of regulation (activation and/or inhibition). The IN sector represents a group
438 of nodes that are not included in the SCC but can reach nodes in the SCC. The OUT sector represents
439 a group of nodes not included in the SCC that can be reached by the nodes in the SCC. The remaining
440 nodes, other than those in the three main sectors, can be further categorized into TUBES,
441 INTENDRILS, OUTTENDRILS, and OTHERS, according to reachability.

442 To understand how the internal and external perturbations of the cells propagate TRN, we
443 embedded links between the internal/external effectors and TRs, as follows. In addition to the bow-tie
444 structure, we also defined the STIMULI block. The STIMULI block does not consist of a gene or
445 regulator, but consists of intracellular and extracellular effectors, such as nutrients, stressors,
446 metabolites, and stimulating transcriptional regulators by allosteric interactions and so forth. We used
447 the compendium of known links between effectors and TRs collected by Martínez-Antonio *et al*[41].
448 These target regulators are defined as SR. SRs change their regulatory activity depending on whether
449 they bind to effectors. The networks obtained are listed in **Supplementary Table 8**.

450 **Data analysis of the three main datasets (Env, Evo and Mut)**

451 In addition to the Mut dataset, we analyzed two datasets of transcriptome in *E. coli*. To estimate the
452 variability in the expression levels of *E. coli* MG1655 in response to environmental changes, we used
453 an updated version of PRECISE[25, 26], which contains the RNA-seq compendium of *E. coli* MG1655
454 obtained from a single laboratory using a standardized protocol. In particular, the transcriptome
455 profiles of PRECISE 2.0 were downloaded from iModulonDB[51]. We filtered the transcriptome data of
456 any mutant or evolved strains derived from the MG1655 strain or from other strains to avoid mutational
457 effects on transcriptional variance. Consequently, 160 profiles under 76 unique environmental
458 conditions were used (**Supplementary Table 1**). The environmental conditions were mainly
459 combinations of seven base media (LB, M9, W2, CAMHB, Tris-maleic acid minimal medium, M9
460 without phosphate, RPMI supplemented with 10% LB), 12 carbon sources (acetate, D-ribose, fructose,
461 galactose, glucarate, gluconate, glucose, glycerol, N-acetylglucosamine, pyruvate, sorbitol, xylose),
462 three nitrogen sources (NH_4Cl, cytosine, glutamine), two electron donors (O_2, KNO_3), three trace
463 elements (aebersold trace element mixture, sauer trace element mixture with or without $MgSO_4$), 35
464 supplements (cytidine, thiamine, leucine, tyrosine, HCl, lactic acid, adenosine, 2, 2′-dipyridyl, $FeCl_2$,
465 NaCl, paraquat, glutathione, adenine, dibucaine, methionine, salicylate, acetoacetate/LiCl, H_2O_2, LiCl,
466 ceftriaxone, ciprofloxacin, glutamate, glycine, L-tryptophan, meropenem, threonine, trimethoprim-
467 sulfamethoxazole, KCl, phenylalanine, uracil, L-arginine, pyruvate, $CuSO_4$, ethanol, $ZnCl_2$), two
468 temperatures (37 °C, 42 °C), three pHs (5.5, 7, 8.5) and two culture types(batch, chemostat). The
469 filtered compendium was called the Env dataset. To estimate the variability in expression levels during
470 the evolution of *E. coli* in nature, we analyzed the RNA-seq compendium of natural isolates of *E. coli*
471 during growth in the exponential phase in LB broth, where the RNA-seq profiles were obtained in a
472 single laboratory using a common protocol[27]. We analyzed the RNA-seq profiles of 16 natural isolates
473 that covered all five primary phylogroups of *E. coli*. The compendium of this dataset was called the
474 Evo dataset.

475 We used 2622 genes that were common to the three datasets (Env, Evo, and Mut)
476 (**Supplementary Table 3**). For all three datasets, the $\log_2$-transformed gene expression levels were
477 used in subsequent analyses. Expression levels were subsequently quantile-normalized. The means
478 and standard deviations of the expression levels of each gene were calculated for each dataset. For the
479 Mut dataset, we first calculated the means and variances of expression levels among the MA lineages
480 of the final rounds, which were derived from the same founder. Subsequently, we averaged the mean
481 and variance of the MA lineages derived from different founders. We obtained the means and standard
482 deviations of the expression levels of MA lineages that underwent 30 rounds of the MA procedure. We
483 found that the standard deviations were dependent on the mean expression levels for every dataset,
484 which can confer a confounding effect on the correlation analysis between the datasets. To compensate
485 for this dependency, we calculated the vertical distance of each standard deviation from a smoothed
486 running median of standard deviations, referred to as the DM, following the method outlined in a
487 previous study[52, 53]. For each gene, we calculated the DM for each dataset and named it DM_{env}, DM_{evo}

488 and DM_{mut} for the Env, Evo, and Mut datasets, respectively. We also calculated variability at the
489 functional category level based on the DM values. Functional categories were defined by Gene
490 Ontologies (GO)[54, 55]. We constructed a subset (82) of GO by removing redundant GO terms based on
491 semantic similarity among GO terms using the GOSemSim[56], AnnotationDbi[57], rrvgo[58] R packages
492 with default parameters. Alternatively, we used the KEGG pathway database[59]. GO terms and KEGG
493 pathways that consisted of fewer than five genes or more than 300 genes were omitted. The mean DM
494 values among the genes belonging to a given GO term or KEGG pathway were used as a measure of
495 variability at the functional category level **(Supplementary Tables 5**,**6**).

496 Gene set enrichment analysis (GSEA)[33] was performed using the clusterprofiler R package[60].
497 We used the refined subset of GO terms and KEGG pathways as functional categories. We filtered out
498 the GO terms and KEGG pathways that consisted of fewer than five genes or more than 300 genes.
499 For GSEA of transcriptional regulators, we also filtered out the transcriptional regulators that regulated
500 fewer than five target genes or more than 1000 genes. The q-values were calculated using the
501 Benjamini-Hochberg (BH) procedure[61]. Transcriptional regulators whose target genes were
502 significantly enriched at higher DM in each dataset were called TR_{env}, TR_{evo} and TR_{mut} for the Env,
503 Evo, and Mut datasets, respectively. The 13 transcriptional regulators present in all three groups were
504 called TR_{com}.

505 The edge density of a given gene set belonging to TRN was calculated as the total number of
506 edges (i.e. regulations) within the gene set divided by the total number of possible edges within the
507 gene set. Self-loops (i.e. autoregulations) were ignored. To compare the edge density within TR_{com}s
508 with that within randomly chosen regulators, we generated 10,000 subsets by drawing the same
509 number of TR_{com}s (i.e. 13) transcriptional regulators at random from a pool of transcriptional regulators.
510 The transcriptional regulators that regulated fewer than five target genes were omitted.

511 PCA was conducted using the prcomp function in R. For the Mut dataset, we used all
512 transcriptional profiles, including isolates at rounds 10 and 30, as well as the founders. In PCA, the *k*-
513 th principal component, PC_k, is represented as a linear combination of loadings *a*, and expression levels
514 *x*, as follows:

515 $$PC_{k,D} = \sum_{i}^{N} a_{k,i,D} x_{i,D}$$

516 where *D*, *N* and *i* represent the datasets (ex. Env), total number of genes, and *i*-th gene,
517 respectively. Correlation analysis was performed for the loadings of the top three PCs, that is, PC1–3
518 between the datasets. The correlation was evaluated using absolute values of the Spearman's rank
519 correlation coefficients (R), as the signs of the PCs are arbitrary.

520 **Data analysis of the Rwr dataset**

521 The Rwr dataset was derived from transcriptome profiles reported by Baumstark *et al* [43]. In the Rwr
522 dataset, *E. coli* TOP10 (K12) strains, which harbor synthetic shuffled operons [42] (referred to as rewired

strains), were cultured under identical environmental conditions. Transcriptome profiles were obtained using the DNA microarray technique (**Supplementary Fig. 11a,b**). A total of 2442 genes, shared across the three main datasets, were extracted. The expression profiles were $\log_2$-transformed and quantile-normalized. To facilitate comparison with the Env dataset, which exhibits high reproducibility, transcriptional profiles that showed relatively low quality (less than 0.95 in mean Spearman's R between biological triplicates) were excluded from the Rwr dataset. The transcriptional profiles were subsequently averaged between the triplicates for each rewired strain. We used a computationally generated random dataset, following a multivariate normal distribution, to confirm that this filtration and averaging process yielded high reliability, comparable to the Env dataset. This filtration resulted in transcriptional profiles of 28 rewired strains (**Supplementary Fig. 11c**).

Data analysis of the other omics datasets (Strs_hp, Strs_he, Strs, Prtn and Noise_ch)

The Strs_hp and Strs_he datasets were obtained from transcriptome profiles reported by Horinouchi *et al*[62]. In Strs_hp, *E. coli* MDS42[63], a derivative of MG1655, was cultured under different stress conditions, and transcriptome profiles were obtained using the DNA microarray technique (**Supplementary Fig. 17a**). The transcriptome profiles obtained from the same stressors used in the Env dataset were removed. Transcriptome profiles for eight stressful conditions (cobalt chloride, sodium carbonate, l-malate, methacrylate, crotonate, methylglyoxal, *n*-butanol and cetylpyridinium chloride) and one stress-free condition (modified M9 medium) were analyzed (**Supplementary Table 9**). Thus, the difference in transcriptome profiles within the obtained dataset, called the Strs_hp dataset, represents transcriptional variability in response to environmental perturbations. For Strs_he, transcriptome profiles were obtained for mutants resistant to each of the different stressors cultured under stressor-free conditions using DNA microarrays (**Supplementary Fig. 17b**). These resistant mutants were obtained through adaptive laboratory evolution experiments using MDS42 as the ancestral strain (**Supplementary Fig. 17c**). We removed the transcriptional profiles obtained from the mutants that experienced the same stressors used in the Env dataset. We analyzed the transcriptional profiles of the mutants obtained under the eight stress conditions described above (**Supplementary Table 9**). Thus, the difference in transcriptional profiles within the resultant dataset, called the Strs_he dataset, reflected the transcriptional variability against genetic mutations accumulated through adaptive evolution. Both datasets contained expression levels of 3741 genes (**Supplementary Fig. 17d**).

The Strs dataset was obtained from the transcriptome profiles reported by Maeda and Iwasawa *et al*[31]. In the Strs dataset, the different MDS42 mutants isolated from adaptive laboratory evolution experiments in the presence of different stressors were subjected to DNA microarray analysis (**Supplementary Fig. 20a–c**). The evolutionary experiment was aimed at obtaining resistance mutants against each of the different stressors using MDS42 as an ancestral strain. The transcriptome profiles of the mutant and ancestral strains were obtained under stressor-free conditions, similar to those of Strs_hp. We removed the transcriptome profiles of mutants that evolved under the same stressors used

560 in the Env dataset (example. hydrogen peroxide, meropenem). We analyzed the transcriptional profiles
561 of the mutants obtained under the 47 stress conditions as detailed in **Supplementary Table 10**. Thus,
562 the difference in transcriptome profiles within the obtained dataset, called the Strs dataset, contained
563 transcriptional variability against genetic mutations that accumulated through adaptive evolution. In
564 addition to the transcriptome profiles of the resistant mutants, Maeda and Iwasawa *et al*.[31] identified
565 several genes that were commonly mutated in multiple evolved lines among independently evolved
566 lines under the same stress conditions. Among these commonly mutated genes, we analyzed those
567 encoding transcriptional regulators, called commonly mutated transcriptional regulators, as listed in
568 **Supplementary Table 11**.

569 For the Strs_hp, Strs_he and Strs datasets, the mRNA expression levels obtained using DNA
570 microarrays were $\log_2$-transformed and quantile-normalized, similar to the Mut dataset.

571 The Prtn dataset was obtained from proteomic profiles reported by Schmidt *et al*[64]. Proteomic
572 profiles with the growth rate were determined for *E. coli* BW25113 cultured under 22 different
573 environmental conditions (including different carbon sources). The difference in proteomic profiles
574 within the resultant dataset, called the Prtn dataset, represented the proteomic variability (2002
575 proteins) in response to environmental perturbations (**Supplementary Fig. 15**). The absolute
576 abundance of proteins (in copies/cell) was $\log_2$-transformed.

577 For all four datasets (Strs_hp, Strs_he, Strs, and Prtn), the means and standard deviations of
578 the expression levels of each gene were calculated across either the mutant or environment in each
579 dataset (**Supplementary Figs. 15**,**17**,**20**). All datasets showed that the standard deviations depended
580 on the mean expression levels. To compensate for this dependency, we calculated the DM values for
581 each gene as the vertical distance of each standard deviation from a smoothed running median of the
582 standard deviations for each dataset as detailed above. Consequently, we obtained DM_{strs_hp}, DM_{strs_he},
583 DM_{strs} and DM_{prtn} for Strs_hp, Strs_he, Strs, and Prtn, respectively. PCA was performed as described
584 previously (**Supplementary Figs. 16**,**19**,**21**).

585 The Noise_ch dataset was obtained from the cell-to-cell heterogeneity of protein expression
586 levels reported by Taniguchi *et al*[65] (**Supplementary Fig. 22**). Briefly, a library of yellow fluorescent
587 protein (YFP) fusion strains of *E. coli* K12 was cultured under identical conditions, and each
588 chromosomal copy of the gene was labeled with YFP fusion. YFP fluorescence was measured using
589 microscopy at the single-cell level, which yielded the fluorescence distribution for each gene (1018
590 genes). For each gene, the protein noise was defined as the variance divided by the square of the mean,
591 which was identical to the coefficient of variation (CV^2)[65]. Therefore, the protein noise in the
592 Noise_chr dataset represents isogenic cell-to-cell heterogeneity in protein expression levels. It is well
593 known that protein noise is mainly composed of two sources: intrinsic noise dominating lowly
594 expressed proteins and extrinsic noises dominating highly expressed proteins. Owing to the Poisson-
595 like statistical nature of intrinsic noise, the protein noise depends on the mean expression levels of the

596 lowly expressed genes (**Supplementary Fig. 22c**). To compensate for this dependency, we calculated
597 the DM values for each gene as the vertical distance of each protein noise ($\log_{10}$-transformed CV^2)
598 from a smoothed running median of the protein noise; this conventional compensation is widely used
599 for protein noise[52, 66]. Consequently, we obtained DM_{noise_ch}, as a measure of variability against
600 stochastic perturbations (**Supplementary Fig. 22d**). Using the noise limits defined by Taniguchi *et al*.,
601 we determined the expression level at which the intrinsic noise limit was equal to the extrinsic noise
602 limit. Genes with expression levels below this threshold were defined as lowly expressed, whereas
603 those with expression levels beyond this threshold were defined as highly expressed. We compared
604 DM_{noise_ch} with the DM values of the other datasets for lowly and highly expressed genes separately.

605 **Data visualization**

606 All figures were generated in R[67] and the following packages were used. The phylogenetic tree of the
607 16 natural isolates of *E. coli* was visualized using the ape[68] and ggtree[69] packages. Pairwise
608 phylogenetic distances were based on core genome alignments and were obtained from a previous
609 study[27]. Network visualization was performed using the ggraph[70], igraph[71] and tidygraph[72] packages.
610 The Venn diagram was visualized using the ggVenndiagram[73], gplots[74] and venneuler[75] packages.
611 Other illustrative plots were generated using the ggplot2[76] and ggpubr[77] packages.

Data availability

All microarray data are available in the National Center for Biotechnology Information's Gene Expression Omnibus functional genomics data repository under the accession number GSE260863.

Code availability

Custom scripts used in the manuscript are available at https://github.com/tsuruubi/Variability_E_coli.

Ethics declarations

The authors declare no competing interests.

Acknowledgements

The authors thank H. Koike for performing MA experiments. The authors acknowledge H. Kotani for providing technical assistance with the DNA microarray experiments. This work was supported by, the Japan Society for the Promotion of Science (JSPS) KAKENHI (18H02427, 22H05403, 24K21985 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

1. Lynch M, Walsh B. *Genetics and analysis of quantitative traits*. Sinauer Sunderland, MA (1998).

2. Levis NA, Pfennig DW. Evaluating 'Plasticity-First' Evolution in Nature: Key Criteria and Empirical Approaches. *Trends Ecol Evol* **31**, 563-574 (2016).

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

4. Waddington CH. *The strategy of the genes*. Routledge (1957).

5. Smith JM*, et al.* Developmental constraints and evolution: a perspective from the Mountain Lake conference on development and evolution. *The Quarterly Review of Biology* **60**, 265-287 (1985).

6. Noble DWA, Radersma R, Uller T. 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 (2019).

7. Arnold SJ, Pfrender ME, Jones AG. The adaptive landscape as a conceptual bridge between micro- and macroevolution. *Genetica* **112-113**, 9-32 (2001).

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

9. Parter M, Kashtan N, Alon U. Facilitated variation: how evolution learns from past environments to generalize to new environments. *PLoS Comput Biol* **4**, e1000206 (2008).

10. Draghi JA, Whitlock MC. Phenotypic plasticity facilitates mutational variance, genetic variance, and evolvability along the major axis of environmental variation. *Evolution* **66**, 2891-2902 (2012).

11. Kouvaris K, Clune J, Kounios L, Brede M, Watson RA. How evolution learns to generalise: Using the principles of learning theory to understand the evolution of developmental organisation. *PLoS Comput Biol* **13**, e1005358 (2017).

12. Brun-Usan M, Rago A, Thies C, Uller T, Watson RA. Development and selective grain make

plasticity 'take the lead' in adaptive evolution. *BMC Ecol Evol* **21**, 205 (2021).

13. Milocco L, Uller T. Utilizing developmental dynamics for evolutionary prediction and control. *Proc Natl Acad Sci U S A* **121**, e2320413121 (2024).

14. Brun-Usan M, Zimm R, Uller T. Beyond genotype-phenotype maps: Toward a phenotype-centered perspective on evolution. *Bioessays* **44**, e2100225 (2022).

15. Waddington CH. Genetic assimilation of an acquired character. *Evolution*, 118-126 (1953).

16. Wagner GP, Booth G, Bagheri-Chaichian H. A Population Genetic Theory of Canalization. *Evolution* **51**, 329-347 (1997).

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

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

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

20. Cheverud JM. A Comparison of Genetic and Phenotypic Correlations. *Evolution* **42**, 958-968 (1988).

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

22. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5**, 101-113 (2004).

23. Voordeckers K*, et al.* Identification of a complex genetic network underlying *Saccharomyces cerevisiae* colony morphology. *Mol Microbiol* **86**, 225-239 (2012).

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

25. Sastry AV, *et al.* The *Escherichia coli* transcriptome mostly consists of independently regulated modules. *Nat Commun* **10**, 5536 (2019).

26. Lamoureux CR, *et al.* A multi-scale transcriptional regulatory network knowledge base for *Escherichia coli*. *bioRxiv*, 2021.2004.2008.439047 (2022).

27. Rousset F, *et al.* The impact of genetic diversity on gene essentiality within the *Escherichia coli* species. *Nat Microbiol* **6**, 301-312 (2021).

28. Halligan DL, Keightley PD. Spontaneous mutation accumulation studies in evolutionary genetics. *Annual Review of Ecology, Evolution, and Systematics* **40**, 151-172 (2009).

29. Fijalkowska IJ, Schaaper RM. Mutants in the Exo I motif of Escherichia coli dnaQ: defective proofreading and inviability due to error catastrophe. *Proc Natl Acad Sci U S A* **93**, 2856-2861 (1996).

30. Tsuru S, *et al.* Genomic confirmation of nutrient-dependent mutability of mutators in *Escherichia coli*. *Genes Cells* **20**, 972-981 (2015).

31. Maeda T, *et al.* High-throughput laboratory evolution reveals evolutionary constraints in *Escherichia coli*. *Nat Commun* **11**, 5970 (2020).

32. Singh GP. Coupling between noise and plasticity in *E. coli*. *G3 (Bethesda)* **3**, 2115-2120 (2013).

33. Subramanian A, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).

34. Honda T, Cremer J, Mancini L, Zhang Z, Pilizota T, Hwa T. Coordination of gene expression with cell size enables *Escherichia coli* to efficiently maintain motility across conditions. *Proc Natl Acad Sci U S A* **119**, e2110342119 (2022).

35. Zampieri M, Horl M, Hotz F, Muller NF, Sauer U. Regulatory mechanisms underlying coordination of amino acid and glucose catabolism in *Escherichia coli*. *Nat Commun* **10**, 3354 (2019).

36. Tierrafria VH, *et al.* RegulonDB 11.0: Comprehensive high-throughput datasets on transcriptional regulation in *Escherichia coli* K-12. *Microb Genom* **8**, (2022).

37. Keseler IM, *et al.* The EcoCyc Database in 2021. *Front Microbiol* **12**, 711077 (2021).

38. Becskei A, Serrano L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590-593 (2000).

39. Yang R, Zhuhadar L, Nasraoui O. Bow-tie decomposition in directed graphs. In: *14th International Conference on Information Fusion*). IEEE (2011).

40. Ghosh Roy G, He S, Geard N, Verspoor K. Bow-tie architecture of gene regulatory networks in species of varying complexity. *J R Soc Interface* **18**, 20210069 (2021).

41. Martinez-Antonio A, Janga SC, Salgado H, Collado-Vides J. Internal-sensing machinery directs the activity of the regulatory network in *Escherichia coli*. *Trends Microbiol* **14**, 22-27 (2006).

42. Isalan M, *et al.* Evolvability and hierarchy in rewired bacterial gene networks. *Nature* **452**, 840-845 (2008).

43. Baumstark R, *et al.* The propagation of perturbations in rewired bacterial gene networks. *Nat Commun* **6**, 10105 (2015).

44. Shoval O, *et al.* Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. *Science* **336**, 1157-1160 (2012).

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

46. Furusawa C, Kaneko K. Formation of dominant mode by evolution in biological systems. *Phys Rev E* **97**, 042410 (2018).

47. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640-6645 (2000).

48. Baba T, *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006 0008 (2006).

49. Sharma RC, Schimke RT. Preparation of electrocompetent *E. coli* using salt-free growth medium. *Biotechniques* **20**, 42-44 (1996).

50. Niccum BA, Lee H, MohammedIsmail W, Tang H, Foster PL. The Spectrum of Replication Errors in the Absence of Error Correction Assayed Across the Whole Genome of *Escherichia coli*. *Genetics* **209**, 1043-1054 (2018).

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

52. Newman JR, *et al.* Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441**, 840-846 (2006).

53. Uchida Y, Shigenobu S, Takeda H, Furusawa C, Irie N. Potential contribution of intrinsic developmental stability toward body plan conservation. *BMC Biol* **20**, 82 (2022).

54. Ashburner M, *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).

55. Gene Ontology C, *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224**, (2023).

56. Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* **26**, 976-978 (2010).

57. Pagès H, Carlson M, Falcon S, Li N. AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor. *R package version 1641*, https://bioconductor.org/packages/AnnotationDbi (2023).

58. Sayols S. rrvgo: a Bioconductor package for interpreting lists of Gene Ontology terms. *microPublication Biology*, (2023).

59. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).

60. Wu T, *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141 (2021).

61. Benjamini Y, Hochberg Y. 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 (1995).

62. Horinouchi T, *et al.* Prediction of Cross-resistance and Collateral Sensitivity by Gene Expression profiles and Genomic Mutations. *Sci Rep* **7**, 14009 (2017).

63. Posfai G, *et al.* Emergent properties of reduced-genome *Escherichia coli*. *Science* **312**, 1044-1046 (2006).

64. Schmidt A, *et al.* The quantitative and condition-dependent *Escherichia coli* proteome. *Nat Biotechnol* **34**, 104-110 (2016).

65. Taniguchi Y, *et al.* Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533-538 (2010).

66. Silander OK, *et al.* A genome-wide analysis of promoter-mediated phenotypic noise in *Escherichia coli*. *PLoS Genet* **8**, e1002443 (2012).

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

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

69. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 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 (2017).

70. Pedersen TL. ggraph: An Implementation of Grammar of Graphics for Graphs and Networks. *R package version 221*, https://CRAN.R-project.org/package=ggraph (2024).

71. Csardi G, Nepusz T. The igraph software. *Complex syst* **1695**, 1-9 (2006).

72. Pedersen TL. tidygraph: A Tidy API for Graph Manipulation. *R package version 131*, https://tidygraph.data-imaginist.com (2024).

73. Gao CH, Yu G, Cai P. ggVennDiagram: An Intuitive, Easy-to-Use, and Highly Customizable R Package to Generate Venn Diagram. *Front Genet* **12**, 706907 (2021).

74. Warnes GR, *et al.* gplots: Various R Programming Tools for Plotting Data. *R package version 3131*, https://CRAN.R-project.org/package=gplot (2024).

75. Wilkinson L. Venn and Euler Diagrams. https://www.rforge.net/venneuler/ (2024).

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

77. Kassambara A. ggpubr: 'ggplot2' Based Publication Ready Plots. *R package version 060*, https://rpkgs.datanovia.com/ggpubr/ (2023).

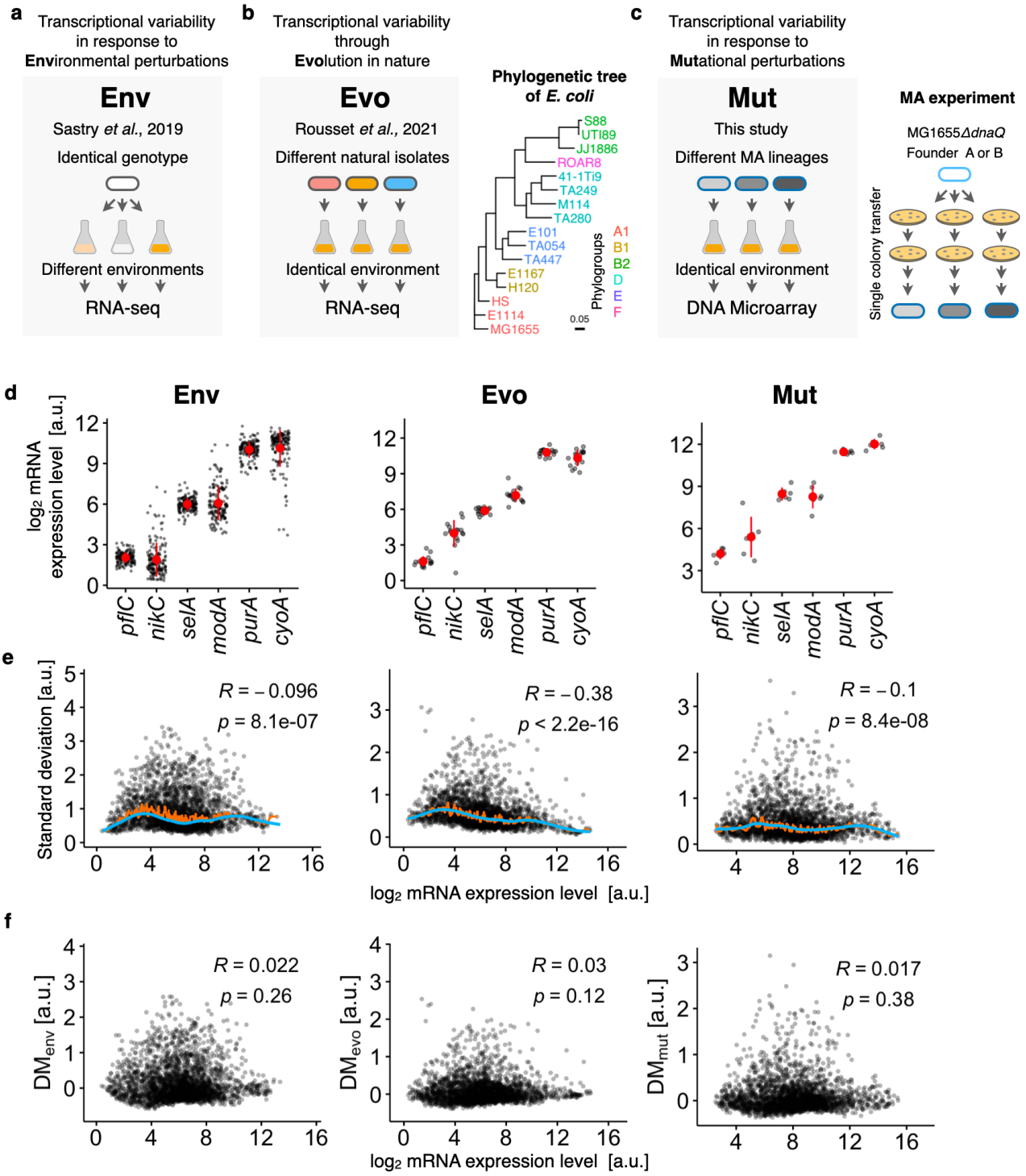
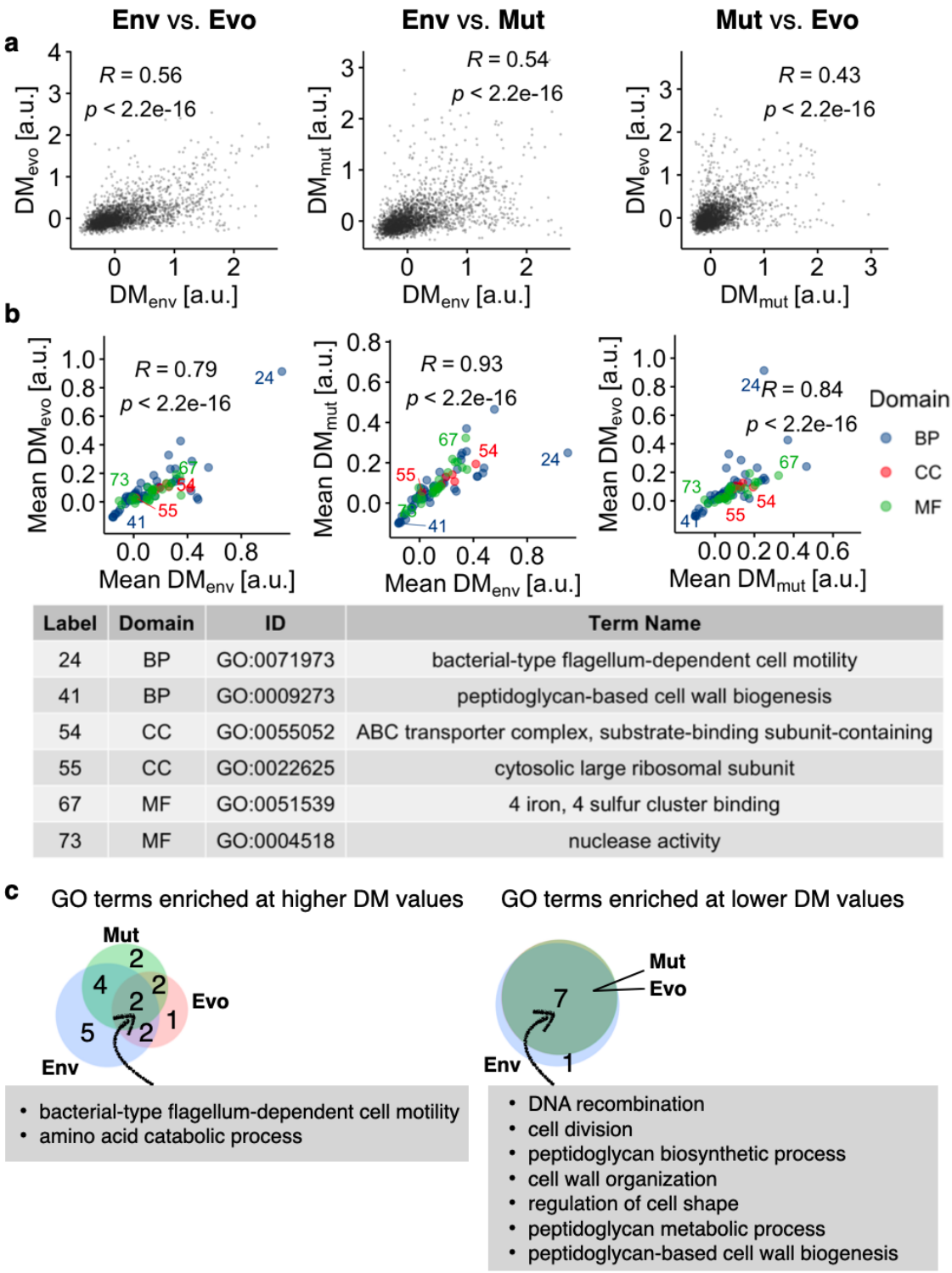


Fig. 1: Transcriptional variability in *E. coli*.

a, A schematic diagram of the Env dataset. The transcriptome profiles of *E. coli* MG1655 were obtained under various growth conditions[25, 26]. **b**, Schematic of the Evo dataset. The transcriptome profiles of 16 natural isolates of *E. coli* were obtained under a common growth condition[27]. The phylogenetic tree was based on the core genome alignment. **c**, Schematics of the MA and Mut dataset. The MA experiment was initiated by two founders of the mutator strain, *E. coli* MG1655*ΔdnaQ*. Single colonies were randomly picked and transferred daily to fresh LB agar plates for 30 rounds. Three MA

lineages from rounds 10 and 30 for each founder were subjected to DNA microarray analysis. Transcriptome profiles were obtained during the exponential growth phase in LB. **d–f**, Transcriptional variability in Env, Evo, and Mut datasets (from left to right). The transcriptome profiles of the MA lineages in the final (30th) round were used for the Mut dataset. **d**, Expression levels of six representative genes. The expression levels were $\log_2$-transformed and quantile-normalized for each dataset. Each grey circle corresponds to a different bacterial culture. Red closed circles and error bars represent means and standard deviations, respectively. **e**, Relationship between means and standard deviations in the expression levels of 2622 genes present in all three datasets. Each grey circle corresponds to a different gene. The variability metric for each gene was defined as the vertical deviation from a smoothed spline (sky blue) calculated from the running median (dark orange) of the standard deviations. The calculated distance to the median is called DM, with subscripts of the corresponding datasets. For the Mut dataset, the standard deviations were obtained by calculating the variances in expression levels separately in each of the two groups distinguished by the founders (founders A and B) and subsequently averaging them. **f**, Relationship between DMs and mean expression levels. The insets represent Spearman's rank correlation coefficients (R) and p-values.



Label	Domain	ID	Term Name
24	BP	GO:0071973	bacterial-type flagellum-dependent cell motility
41	BP	GO:0009273	peptidoglycan-based cell wall biogenesis
54	CC	GO:0055052	ABC transporter complex, substrate-binding subunit-containing
55	CC	GO:0022625	cytosolic large ribosomal subunit
67	MF	GO:0051539	4 iron, 4 sulfur cluster binding
73	MF	GO:0004518	nuclease activity

Fig. 2: Positive correlations of transcriptional variability between datasets.
a, Pairwise correlations are shown for DMs at the gene levels between the Env, Evo, and Mut datasets. **b**, Correlations of DMs at the functional category level between pairs of datasets. The genes were grouped into functional categories based on their GO terms. To reduce the redundancy of GO terms, a custom subset (82) of GO terms was used as detailed in the **Methods** section. Each dot corresponds to a different GO term. The mean DM was calculated by averaging the DM of the genes belonging to

each GO term. The three domains of GO–biological processes (BP), cellular components (CC), and molecular functions (MF)–are labeled in different colors. Spearman's R and p-values are shown in the panels. Six representative GO terms are labeled with different numbers, as detailed at the bottom. **c**, Venn diagram of GO terms enriched in higher (left) and lower (right) DM values in each dataset (GSEA, BH-adjusted p-values (q)<0.05, **Supplementary Fig. 2**). The Evo and Mut sets completely overlap in the right Venn diagram. The numbers indicated in the diagram show the number of GO terms enriched in the corresponding set.

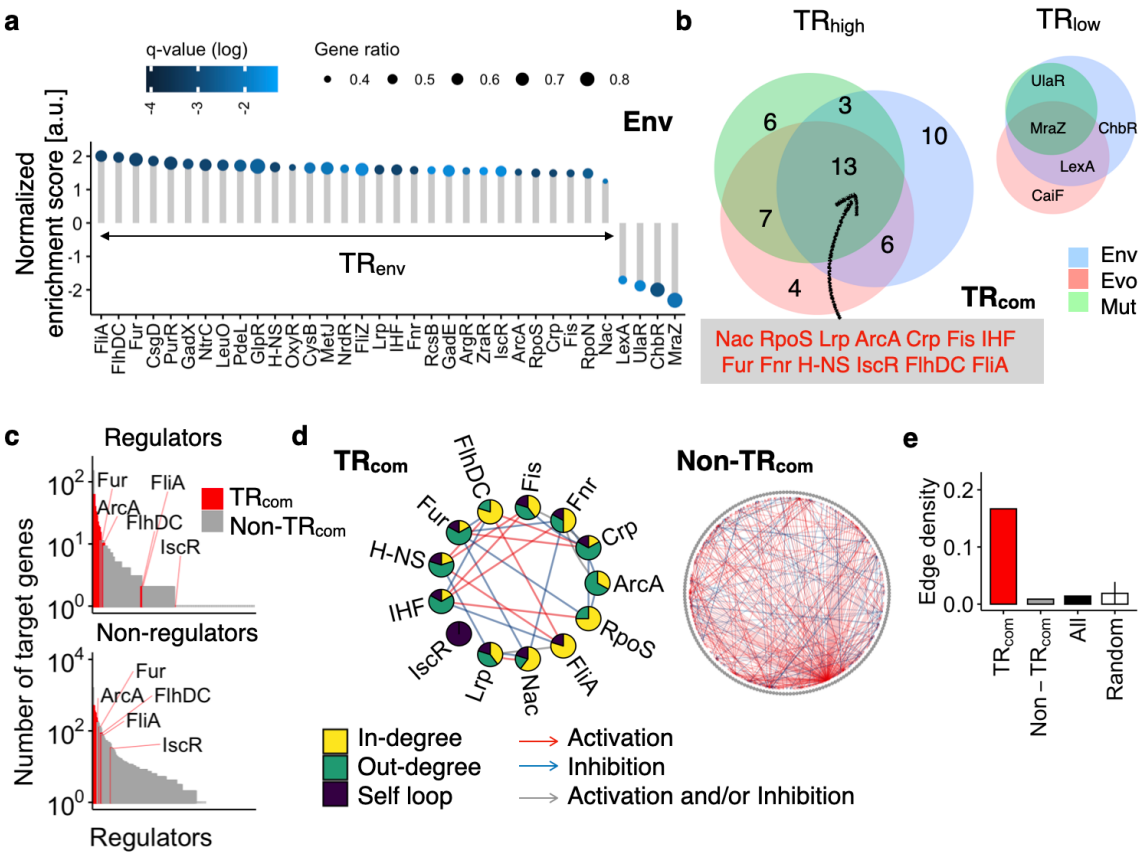


Fig. 3: Transcriptional regulators and transcriptional variability.

a, Transcriptional regulators (TR) enriched in genes with higher and lower DM_{env}. GSEA identified statistically significant TRs ($q<0.05$). Positive and negative enrichment scores indicate enrichment at higher and lower DM_{env}, respectively. The gene ratio was calculated as the number of core target genes contributing to enrichment scores divided by the number of target genes. TRs with positive enrichment scores are labeled as TR_{env}. The q-values (adjusted p-values) were computed using the BH procedure. **b**, Venn diagram of the enriched TRs in each dataset. TR_{evo} and TR_{mut} represent the TRs with positive enrichment scores in the Evo and Mut datasets, respectively. The intersection of the three sets represents TR_{com}. **c**, Rank plot of the number of target genes in the TRs. The TRs were sorted in decreasing order. Target genes were divided into regulators (transcription factors and sigma factors, upper panel) and non-regulators (lower panel). The red and grey bars represent TR_{com}s and the other TRs (Non-TR_{com}s), respectively. The labels are subsets of TR_{com}. **d**, Circular representation of *E. coli* transcriptional regulatory network (TRN) within TR_{com}s (left) and Non-TR_{com}s (right). The red, blue, and grey lines represent activation, inhibition, and activation and/or inhibition, respectively. Pie charts indicate the percentage of each degree within the TR_{com} nodes. The edges corresponding to the self-loops were omitted from the visualization. **e**, Edge densities within regulatory groups. Random groups were generated by drawing 13 TRs at random 10,000 times from a pool of TRs, where TRs that regulate fewer than five target genes were omitted. Error bars represent standard deviation. Edge density was calculated as the total number of edges within the regulators divided by the total number of possible edges within the regulators, excluding the self-loops.

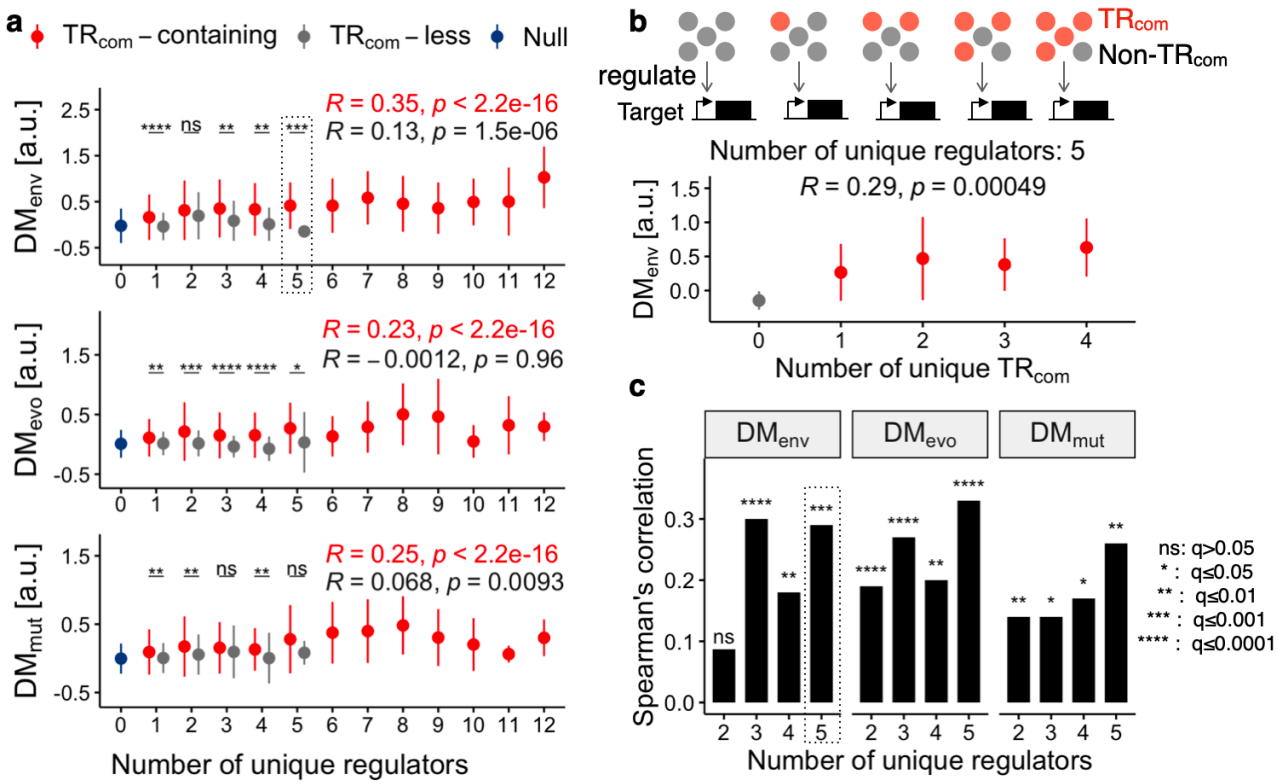


Fig. 4: Relationship between transcriptional variability and the number of transcriptional regulators

a, Relationship between DM values and the number of unique regulators. The top, middle, and bottom panels show the DM_{env}, DM_{evo} and DM_{mut}, respectively. Red, grey, and blue symbols represent target genes receiving TR_{com}-containing regulation, TR_{com}-less regulation, and genes without known regulators, respectively. Means and standard deviations are shown as the points and the error bars, respectively. Data points for TR_{com}-containing and TR_{com}-less regulations were restricted to those comprising more than four target genes. Spearman's R was calculated for data points 0–5 on the x-axis for both TR_{com}-containing and TR_{com}-less regulations. Asterisks represent the statistical significance levels (q-values) of the Wilcoxon test between TR_{com}-containing and TR_{com}-less genes in each class. The target genes within the dotted square are used in panel **b**. **b**,**c**, The correlation between DM values and the number of TR_{com} regulations. **b**, Relationship between DMenv and the number of TR_{com}s for target genes regulated by any of the five regulators. Means and standard deviations are shown as the points and the error bars, respectively. Spearman's R and p-values are shown. **c**, Spearman's R for DM_{env} (left), DM_{evo} (center), and DM_{mut} (right). The enclosed square corresponds to panel **b**.

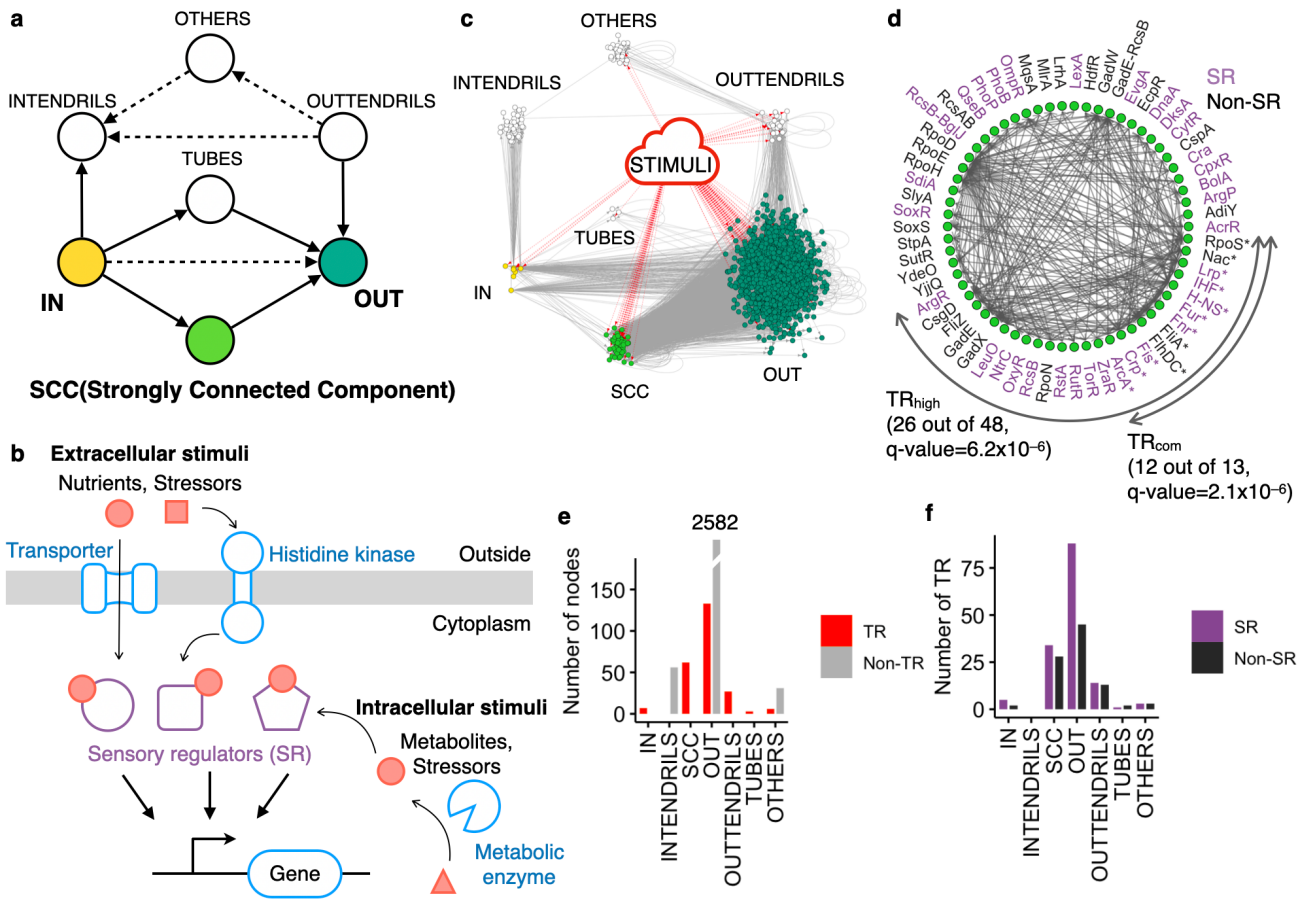
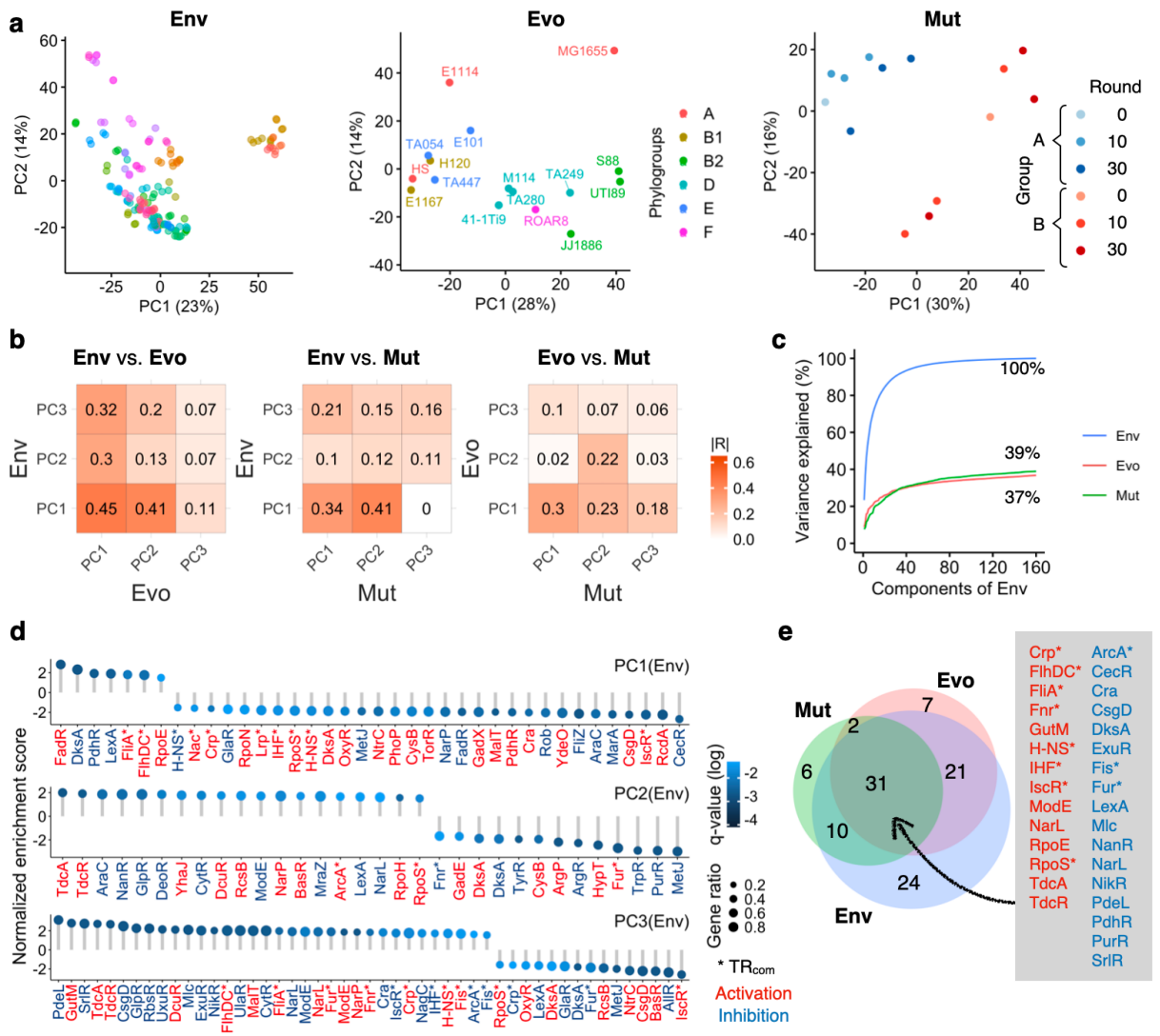


Fig. 5: Bow-tie decomposition of the gene regulatory network in *E. coli*.
a, Yang's bow-tie structure[39] is delineated into seven sectors, with the main sectors highlighted: IN (yellow), OUT (blue green), and the strongly connected component (SCC, light-green). Arrows indicate the direction of regulation, with solid arrows defining each sector and dashed arrows representing possible regulations without violating the sector definitions. **b**, Schematic illustrating how sensory regulators (SR, open purple symbols) influence the expression levels of target genes by interacting with intracellular and extracellular effectors (red closed symbols). **c**, Bow-tie representation of the gene regulatory network in *E. coli* based on panel **a**. Each node represents a TR or target gene. Grey arrows denote regulation (activation and/or inhibition). The STIMULI block represents an aggregated group of known effectors interacting with SRs, with red dotted arrows indicating the interactions between the effectors in STIMULI and SRs. **d**, Regulatory network within the SCC. SR and non-SR groups are colored purple and black, respectively, with asterisks representing TR_{com}. **e**, Distribution of nodes in each sector. TR and non-TR groups are colored red and grey, respectively. IN, SCC, TUBES, and OUTTENDRILS consist solely of TR, whereas INTENDRILS contains only non-TRs by definition. **f**, Number of TR in each sector. SR and non-SR groups are colored purple and black, respectively.



Representative GO terms enriched in the **positive** direction of PC1 in Env (growth family)

Domain	ID	Term Name
MF	GO:0140640	catalytic activity, acting on a nucleic acid
MF	GO:0003735	structural constituent of ribosome
BP	GO:0006412	translation
BP	GO:0140694	non-membrane-bounded organelle assembly
BP	GO:0001510	RNA methylation

Representative GO terms enriched in the **negative** direction of PC1 in Env (starvation family)

Domain	ID	Term Name
CC	GO:0055052	ABC transporter complex, substrate-binding subunit-containing
BP	GO:0008652	amino acid biosynthetic process
BP	GO:0009401	phosphoenolpyruvate-dependent sugar phosphotransferase system
BP	GO:0006006	glucose metabolic process
BP	GO:0042594	response to starvation

Fig. 6: Alignment of genetic and environmental correlations in transcriptional variability. **a**, First and second principal components (PC) of Env, Evo and Mut arranged from left to right. Points are colored by environmental condition in Env. **b**, Correlation analysis for loadings of the first three

963 PCs among the three datasets. The color scale and numbers in the tiles represent the absolute values of
964 Spearman's R. **c**, Cumulative variance explained by the PCs defined by the Env dataset. Overall, the
965 Env dataset explained 37% and 39% of the variance in the Evo and Mut datasets, respectively. **d**, TRs
966 were significantly enriched ($q<0.05$) in greater loadings, eithter positive or negative, of the first three
967 PCs in the Env dataset. GSEA was used to identify regulations, distinguishing between activatory (red)
968 and inhibitory (blue) regulation. Asterisks denote regulations by TR_{com}s. **e**, Venn diagram illustrating
969 the TRs enriched in the first three PCs of each dataset. GSEA was performed for each dataset, as shown
970 in panel **d**. Numbers indicate the number of regulations in the corresponding set. The detailed
971 regulations at the intersection of the three sets are shown on the right. **f**, GO terms enriched with greater
972 loadings in the positive (top) and negative (bottom) directions in PC1 of the Env dataset (five
973 representatives). The former and latter GO terms were categorized into the growth and starvation
974 families. **g**, Number (left) and fraction (right) of GO terms enriched in each sign of each PC in each
975 dataset. GSEA was conducted for each PC in each dataset. The fraction was calculated by dividing the
976 number of GO terms in each family by the total number of GO terms in each PC sign. A full list of the
977 enriched GO terms is presented in **Supplementary Fig. 12–14**.

a

Dataset	Perturbation	mRNA/protein	Variability	Reference
Strs_hp	Environmental	mRNA	DM_{strs_hp}	Horinouchi et al (2017)
Strs_he	Genetic	mRNA	DM_{strs_he}	Horinouchi et al (2017)
Strs	Genetic	mRNA	DM_{strs}	Maeda & Iwasawa et al (2020)
Prtn	Environmental	protein	DM_{prtn}	Schmidt et al (2016)
Noise_ch	Molecular noise	protein	DM_{noise_ch}	Taniguchi et al (2010)

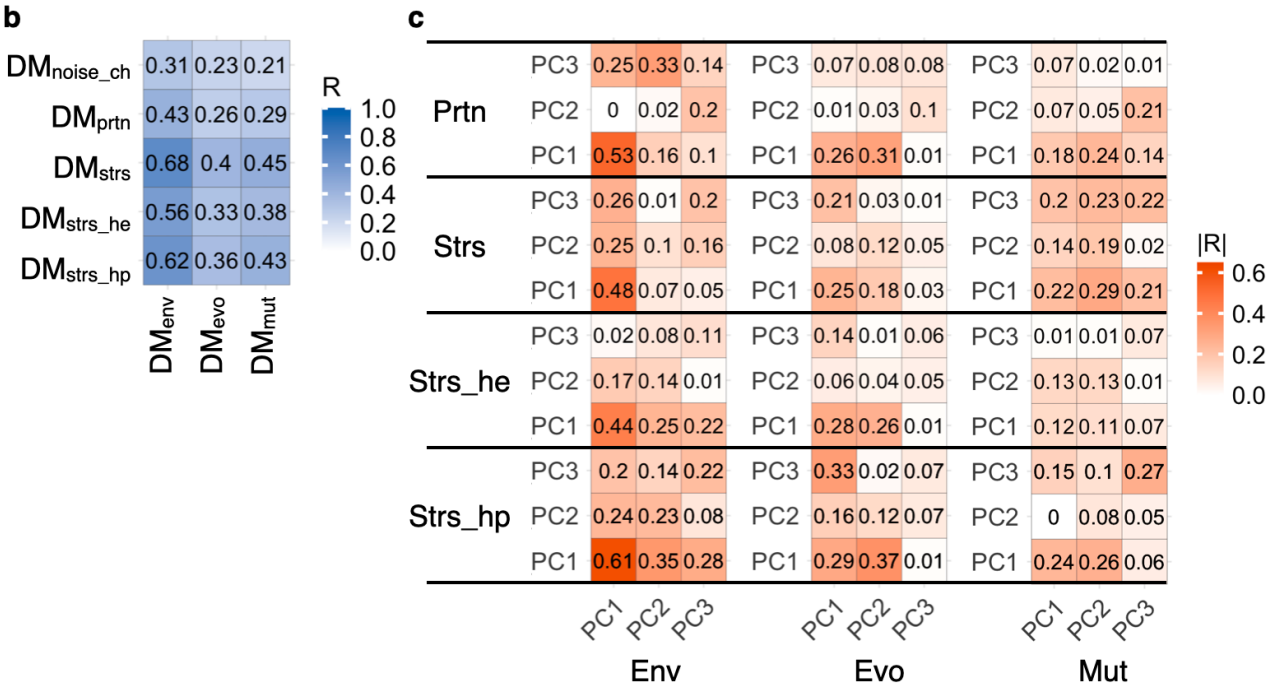


Fig. 7: Similarity of phenotypic variability across different datasets

a, Datasets compared to three main datasets. **b**, Correlation analysis of the variability across different datasets. The color bar and numbers in the tiles represent Spearman's R. Correlation analysis for DM_{noise_ch}, was conducted only for highly expressed proteins with extrinsic noise dominance. The results of the lowly expressed proteins are detailed in **Supplementary Discussion 2**. **c**, Correlation analysis of loadings of the first three PCs between the datasets. The color bar and numbers in the tiles represent absolute values of Spearman's R.