Global and gene-specific translational regulation in *Escherichia coli* across different conditions

Di Zhang[1,2], Sophia Hsin-Jung Li[3,4,5], Christopher G. King[3], Ned S. Wingreen[3, 6*], Zemer Gitai[3*], Zhiyuan Li[1, 2*]

[1] Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, 100871, China

[2] Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, 100871, China

[3] Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA

[4] Institute of Bioengineering, School of Life Sciences, Swiss Federal Institute of Technology Lausanne, CH-1015 Lausanne, Switzerland

[5] Global Health Institute, School of Life Sciences, Swiss Federal Institute of Technology Lausanne, CH-1015 Lausanne, Switzerland

[6] Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA.

* To whom correspondence should be addressed. Tel: +86 10 62750190 ; Email: zhiyuanli@pku.edu.cn

Correspondence may also be addressed to Ned S. Wingreen. Tel: +1 6092588476; Email: wingreen@princeton.edu

Correspondence may also be addressed to Zemer Gitai. Tel: +1- 6092585419; Email: zgitai@princeton.edu

The authors wish it to be known that, the first two authors should be regarded as joint First Authors

Abstract

How well mRNA transcript levels represent protein abundances has been a controversial issue. Particularly across different environments, correlations between mRNA and protein exhibit remarkable variability from gene to gene. Translational regulation is likely to be one of the key factors contributing to mismatches between mRNA level and protein abundance in bacteria. Here, we quantified genome-wide transcriptome and relative translation efficiency (RTE) under 12 different conditions in *Escherichia coli*. By quantifying the mRNA-RTE correlation both across genes and across conditions, we uncovered a diversity of gene-specific translational regulations, cooperating with transcriptional regulations, in response to carbon (C), nitrogen (N), and phosphate (P) limitations. Intriguingly, we found that many genes regulating translation are themselves subject to translational regulation, suggesting possible feedbacks. Furthermore, a random forest model suggests that codon usage partially predicts a gene's cross-condition variability in translation efficiency; such cross-condition variability tends to be an inherent quality of a gene, independent of the specific nutrient limitations. These findings broaden the understanding of translational regulation under different environments, and provide novel strategies for the control of translation in synthetic biology.

Introduction

The central dogma connects DNA, RNA, and protein through transcription and translation. While transcriptional regulation has been extensively studied in the past century[1], how well transcript level represents protein abundance remains controversial[2, 3]. Despite the overall positive correlation between mRNA and protein abundance when different genes are compared in bacteria, genes with similar mRNA abundance may show large differences in protein abundance[4]. Given that a non-negligible portion of the bacterial genome exists as polycistrons, the difference in protein abundance can be traced to widespread differences in translational capacity between genes, e. g. arising from mRNA secondary structure, codon usage bias, ribosome binding sites, riboswitches, and leader peptides[5-9]. Not only do these factors account for differential protein synthesis under steady-state conditions, but some of them also respond to external stressors, e. g. the hairpin structure in the 5'UTR of *pfrA* in Listeria opens at high temperatures to improve translation[10].

Cells regulate their protein expression profiles in response to environmental challenges. While this regulation is conventionally thought to occur primarily at transcription[11], several recent studies based on the translatome have revealed gene-specific translational regulation when bacteria are exposed to heat stress, oxidative stress, or amino-acid starvation[12-15]. These studies specifically focused on translational regulation in specific genes contributing to stress responses, such as the heat shock protein (HSP) gene family[16]. However, the universality of translational regulation as a response to general environmental challenges remains largely unexplored. For instance, nutrient limitations are fundamental challenges for cells, and *E. coli* cells are known to cope with different nutrient limitations by changing ribosomal synthesis and usage strategies [17, 18]. But we do not yet know if cells also regulate the translation of specific genes in response to nutrient limitations. Moreover, as transcription and translation are coupled in bacteria[19], it is worth quantifying the extent to which transcription and translation are regulated in concert to cope with environmental stresses.

To understand the general principles of gene expression, theoretical models have been developed to provide an integrative picture of translational regulation [20-22]. However, the factors that contribute to gene-specific translational regulation upon environmental changes are still poorly understood[23]. In this regard, some researchers have suggested that protein synthesis rates are tightly linked to tRNA composition and modification, which is known to vary across conditions [24-26]. However, others suggested that translation efficiency is highly correlated with ORF mRNA structure

rather than other mRNA features such as tRNA adaptation index (tAI)[27]. Therefore, the underlying mechanisms of translational regulation in response to environmental changes are worth exploring.

Here we extended our previous research[18] and systematically quantified the total number, bound fraction, and elongation rates of ribosomes under carbon (C), nitrogen (N), and phosphorus (P) limitations at different growth rates, and several other growth conditions. Then we extended our perspective on translational regulation from the global scale to the level of individual genes. We aimed to examine whether there is gene-specific translational regulation in *E. coli* responding to different nutrient limitations, and then explore possible mechanisms. Combining global ribosome profiling with RNA-seq, we quantified the correlation between mRNA level and translation efficiency both across genes and across conditions, as well as the variability of translation efficiencies across conditions. We uncovered a diverse range of gene-specific translational regulations concerted with transcriptional regulations in response to environmental deficiencies. Intriguingly, several translational regulation genes are themselves subject to translational regulation, suggesting possible feedbacks. Further analysis suggested that codon usage may play an important role in gene-specific translational regulation. Using a random forest model, we quantified the contribution of codon usage towards condition-dependent translational regulation. This analysis revealed that the cross-condition variability tends to be an inherent feature of individual genes, independent of particular conditions. These findings expand our understanding of translational regulation in response to environmental changes, and suggest novel strategies for effective translation in future synthetic biology.

Results

Cells adapt to different nutrient conditions through global translational regulations

To explore the translational regulation in *E. coli* under different environments, we utilized 12 different growth conditions (Table 1). In an effort to focus on distinct yet stable steady-state growth conditions, we grew *E. coli* in chemostats with dilution rates of 0.1 and 0.6 h^{-1} under limitations for carbon (C), nitrogen (N), and phosphate (P). We also grew two mutant strains in chemostats, Δ*rplA* and Δ*leuB*, with the same dilution rates of 0.1 and 0.6 h^{-1}. *rplA* encodes a component of the 50S ribosome subunit [28, 29] and *leuB* is involved in leucine biosynthesis[30]. These two mutant strains thus enabled us to probe how single-gene mutations that disrupt distinct aspects of the translation process affect the overall pattern of translation. In addition, wild type *E. coli* was also grown in batch culture using both glucose minimal media and defined rich MOPS media, with measured growth rates of 0.9 and 1.8 h^{-1}, respectively.

As observed in a previous study[18], P-limited cells consistently exhibited lower RNA-to-protein ratios than C-limited or N-limited cells. Δ*rplA* cells exhibited a higher RNA-to-protein (R/P) ratio than other conditions at the growth rate of 0.6 h^{-1}, consistent with the significantly reduced activity of ribosomes for Δ*rplA* cells[31]. All other conditions are located on a single line of R/P ratio versus growth rate (Fig. 1A). The free ribosome pools decreased as growth rate increased across all the nutrient-limited conditions (Fig. 1B). Our previous study suggested that *E. coli* differentially tune multiple ribosomal features, including ribosome total number, elongation rate, and active fraction, to achieve the same growth rate of 0.1 h^{-1} under different nutrient limitations [18]. The current results confirmed that this pattern extends to a higher growth rate of 0.6 h^{-1} (Fig. 1C-E). Meanwhile, under batch conditions, all three of these ribosomal features reached very high levels (Fig. 1F). The distribution of ribosome density along mRNAs also revealed differences between conditions: For C- and N-limited conditions, there was a higher ribosome occupancy near the start codon, particularly at the lower growth rate (Fig. 1G and the inset). Across all conditions, after the first few codons, the ribosome density exhibited no significant decrease along mRNAs (Fig. 1G). In summary, cells adapt to different nutrient conditions by differentially tuning multiple ribosome-related features, which act globally on the translation efficiencies of all genes. In addition to such global translational regulation, we wondered whether there could be gene-specific translational regulation in response to different environment conditions.

Quantifying transcriptome and translatome in *E. coli* under multiple nutrient limitations

To explore individual gene expression regulation in *E. coli* under various nutrient conditions, we quantified the genome-wide transcriptome and translatome by performing RNA-seq and global ribosome profiling for all the conditions above[32]. After filtering out ribosomal RNA (rRNA) and transfer RNA (tRNA) species, a total of 4321 genes were used as the reference for mapping.

We then combined ribosome profiling and RNA-seq data to quantify the relative translation efficiency (RTE, defined as the ratio of the transcript abundance in the ribosome profiling to relative mRNA level) of each gene under different conditions (see Methods for details). It is worth noting that the RTE represents the relative occupancy of ribosomal resources devoted to translation, rather than the absolute protein production rate per mRNA molecule. To avoid the high noise caused by low mRNA levels, we filtered genes with a cut-off of $\log_{10}$(mRNA RPKM) > 1.5. After filtering, a total of 2914 genes were retained for further analysis. Scatter plots and correlation analysis of per gene mRNA and ribosome level showed high data reproducibility across different replicates (Sup Fig. 1-2). Since RTE is the ratio of footprint densities to RNA-seq read densities, it could be sensitive to the changes of mRNA levels. To test whether RTE truly reflected differences in translation between genes, we analyzed the expression pattern of genes from the *dusB-fis* operon and the F_0F_1 ATP complex, which are two typical cases that controlled at translation level with similar mRNA abundances. *dusB* and *fis* are coregulated as part of the same operon, and we observed that their mRNA levels were comparable. However, because of the highly different mRNA structure[27], their RTEs showed significant disparities (Sup Fig. 3A), consistent with the results of previous study[27]. Eight subunits of the F_0F_1 ATP complex were from a single polycistronic transcript, and the mRNA levels of these genes were similar (Sup Fig. 3B-C). However, their RTEs varied substantially and were proportional to their stoichiometry in the F_0F_1 ATP complex (Sup Fig. 3D), consistent with a previous study by Li *et al.*[17] on the quantification of absolute translation efficiency. Thus, these results confirmed the reliability of RTE in quantifying translational differences among genes.

mRNA-RTE correlation analysis suggests the preponderance of both gene-specific and condition-specific translational regulation

To quantify the correlation between mRNA level and RTE, we analyzed two types of correlation: across genes and across conditions (Fig. 2A). For cross-gene correlations, we confirmed that, on average, mean mRNA levels positively correlated with RTEs (Fig. 2B), with a coefficient of determination (R^2) of 0.3. This means that in an average sense, if one gene has a higher mRNA level than another, it is also likely to have a larger RTE.

Next, we wondered whether the mRNA levels and RTEs of individual genes change in concert across different conditions. To answer this question, we examined the cross-condition correlation between mRNA level and RTE for each gene. In contrast to the positive cross-gene correlation, we found a broad distribution of the 2914 Spearman's rank correlation coefficients between a gene's mRNA levels and its RTEs across the 12 conditions (Fig. 2C, blue). The distribution ranged from -1 to +1, asymmetrically biased toward negative values. The median of this distribution was -0.23, and 24.5% of the genes exhibited a smaller than -0.5 correlation between their mRNA levels and RTEs across conditions; only 4.7% of the genes exhibited a larger than 0.5 correlation. To test the significance of this asymmetric and mostly negative distribution, we randomly scrambled the RTEs among conditions for each gene and recalculated the 2914 correlation coefficients to obtain a null distribution (Fig. 2C, grey). As confirmed by theoretical analysis, this null distribution was symmetric with zero mean (see supplement for details), and visibly distinct from the actual distribution. The sizable, statistically significant difference between the actual distribution and the null distribution implied the widespread existence of gene-specific translational regulation, where the RTE of an individual gene changed in response to different environmental conditions (Sup Fig. 4).

To further explore the possible roles of gene-specific translational regulation, we examined two genes with highly negative and highly positive mRNA-RTE correlations. For the gene *fieF*, which mediates metal-ion transport in response to iron poisoning[33], the correlation coefficient was -0.91 (Fig, 2D). Iron homeostasis is essential for cell survival[33]. For the gene *ycaO*, which is involved in the β-methylthiolation of the ribosome complex S12[34], the correlation coefficient was 0.73 (Fig. 2E). Ribosome abundance has been known to change with growth conditions or cellular status[35]. These two examples raise one more general question: Do genes with negative correlation and those with positive correlations perform different classes of biological functions?

Correlations between mRNA level and RTE link to gene function

To test the hypothesis that genes with distinctive mRNA-RTE correlations fall into different functional categories, we performed gene ontology (GO) enrichment analysis for the top 300 genes with the strongest negative mRNA-RTE correlations across conditions, as well as the top 300 genes with the strongest positive correlations. We found that the top 300 most negatively correlated genes were mainly involved in biological processes that may not be affected by our nutrient limitations (Fig. 3A). These genes, such as *nagC*, *ascG*, and *kdgR*, are essential for the homestasis of metabolism. Other genes in this group, such as *rpoE*, *lacI*, and *frmR*, respond to inputs such as heat shock, lactose, or formaldehyde which were not assayed in our experimental condition. It is conceivable that the negative correlation between mRNA level and RTE for genes in this group reduces the dependence of protein abundance on conditions.

By contrast, the top 300 genes with the strongest positive correlation were mainly involved in nutrient utilization, stimulus response, and translational regulation itself (Fig. 3B). These genes are the key cellular factors that respond to the imposed nutrient limitations. Interestingly, the observation of strong positive correlation for genes involved in translational regulation hinted at possible direct feedback, i.e. that genes regulating translation are themselves subject to translational regulation. Take two typical genes as examples: *rplA*, encoding a component of the 50S ribosome subunit, functions in translational regulation [28, 29]. The mRNA level of r*plA* was significantly upregulated at a growth rate of 0.6 h^{-1} comparing to 0.1 h^{-1}. In concert, the RTE of *rplA* was also significantly upregulated at the faster growth rate (Fig. 3C). This phenomenon was robust under all three nutrient limitations, C, N, and P. Similarly, *rmf*, a translation inhibitor, is also subject to translational regulation. RMF is a ribosome modulation factor that reversibly converts active 70S ribosomes to a dimeric form, which is associated with a decrease in overall translational activity during the transition from exponential growth to stationary phase[36]. Our data show that for *rmf*, both mRNA level as well as RTE were significantly down-regulated at faster growth rates, regardless of which nutrient was limiting (Fig. 3D).

Gene-specific translational regulation in response to nutrient limitations

Despite the mostly negative distribution of mRNA-RTE correlations for all genes, our former analysis suggests concerted regulation of both mRNA level and RTE for genes responsive to environmental changes. To systematically examine such concerted regulation, we analyzed the relative changes at

both mRNA and RTE levels between pairs of nutrient limitations under the same growth rate of 0.1 h^{-1}. We first compared the expression between C-limited and N-limited cells. We used a cutoff of $\log_2$(C-/N-limited mRNA fold change) > 4 and *p*-value < 0.05 to select a group of differentially expressed genes at the mRNA level. For genes with significantly upregulated mRNA levels under C limitation, almost all of their RTE fold changes were also greater than 1, thus exhibiting concerted regulation of transcription and translation (Fig. 4A, red dots). In the same way, genes with significantly upregulated mRNA levels under N limitation also showed upregulated RTEs (Fig. 4A, blue dots). Similar phenomena can be observed when comparing N-limited and P-limited cells: for genes with upregulated mRNA levels under P limitation, their RTEs were also significantly upregulated (Fig. 4B, green dots), and similarly for genes upregulated under N limitation (Fig. 4B, blue dots – same genes as in Fig. 4A). Next, we compared the mRNA level and RTE across C, N, and P limitations in parallel for the three gene groups selected above. The results further confirmed that mRNA and RTE change in concert for genes that are specifically expressed under specific nutrient limitations (Fig. 4C-E).

To verify whether these three groups of genes are actually involved in utilization of specific nutrients, we performed GO analysis for each group. The resulting functional enrichment confirmed our hypothesis. Genes with concerted upregulation of both mRNA level and RTE under C limitation were mainly involved in the transport of carbon-containing compounds or cell locomotion (Fig. 4F). For example, the gene *yjcH* (marked in red in the upper right region of Fig. 4A) encodes a protein involved in acetate catabolism and transport[37], while the genes *mglA-C* are involved in galactose transport, which also responds to C limitation[38]. Other genes in this group, *flgC-F*, are involved in flagellar assembly. Genes with concerted upregulation of both mRNA level and RTE under N limitation were mainly involved in nitrogen utilization (Fig. 4G). The genes *rutA-G* (marked in blue in the lower-left region of Fig. 4A) in the *rut* pathway are typical examples: they contribute to derivation of nitrogen from pyrimidines [39, 40]. Similarly, genes with concerted upregulation of both mRNA level and RTE under P limitation were mainly involved in phosphorus metabolism (Fig. 4H). These genes mainly consist of the *phn* gene cluster (marked in green in the lower-left region of Fig. 4B), which is induced under phosphate limitation and plays an important role in deriving phosphate from phosphonate degradation [41, 42]. The same analyses were performed for the growth rate of 0.6 h^{-1}, with consistent results, except for no significant observed difference in RTE between C and N limitations for the genes involved in nitrogen utilization (Sup. Fig. 5).

To test for gene-specific translational regulation between different growth rates, we performed similar analyses at the growth rates of 0.1 h^{-1} and 0.6 h^{-1} under the same nutrient limitation. Intriguingly, there were no evident differences in translational regulation between the two different growth rates (Sup. Fig. 6A-C). Even the highlighted subsets of genes identified in Fig. 4, involved in utilization of specific nutrients, showed no significant difference of RTE between different growth rates under the same types of nutrient limitation (Sup. Fig. 6D-F). In summary, these results strongly suggest gene-specific translational regulation in response to different nutrient limitations but not different growth rates.

Translational regulation patterns associate with codon usage

Our findings revealed translational regulation of genes in response to different nutrient limitations but not in response to different growth rates, implying the existence of two classes of genes: those that change RTEs across conditions, and those with stable RTEs. We therefore investigated the variability of RTE across conditions (Fig. 5A). To obtain a global view, we first calculated the mean and variance of RTE for each gene across the 12 different conditions (Fig. 5B). The results show an overall positive correlation between the mean and variance of RTE. However, genes with similar mean RTE still exhibit remarkable differences in their RTE variance.

To zoom in to a function-related view, we compared the translational regulation patterns of the 82 pathways in *E. coli*[43], and they seem to be distinguished in the mean-variance biplot of RTE (Sup, Fig. 7). For example, four pathways with different biological functions occupied two distinguishable regions in the biplot (Fig. 5B, colored dots). Compared with the overall transcription and translation pattern of background genes (Fig. 5C), the TCA cycle and the pyruvate metabolism pathways shared similar translation patterns, with a small RTE variance (Fig. 5D). These two pathways are both involved in basic metabolic processes[44, 45]. Intriguingly, we found that they shared highly similar codon usage, with a Spearman's rank correlation coefficient between their codon frequencies of 0.93 (Fig. 5G). By contrast, the flagellar assembly pathway and the bacterial chemotaxis pathway both exhibit large RTE variance (Fig. 5E). In addition, their mean RTEs were significantly positively correlated across the 12 conditions (Fig. 5F). These two pathways are both involved in cell motion[46, 47]. They also shared similar codon usage and the Spearman's rank correlation coefficient between

their codon frequencies was 0.52 (Fig. 5H). Altogether, we found that pathways with similar translational regulation patterns tend to share similar codon usage.

Codon usage partially predicts the cross-condition variability of RTE

The observation above concerning specific pathways inspired us to quantify how much codon usage contributes to this cross-condition RTE variance. Overall, the mean codon frequencies for the top 200 genes with the largest RTE variance and those for the bottom 200 genes with the smallest RTE variance exhibited negative correlation (Fig. 6A, Spearman's rank correlation coefficient -0.55). Globally, certain codons appeared with changing frequencies for genes with different RTE variabilities. We singled out four codons with discrepant frequencies between the high-RTE-variability genes and the low-RTE-variability genes: The frequencies of AAA and GAT in the 2914 genes showed an overall increasing trend with increasing RTE variance (Fig. 6B, upper). On the other hand, the frequencies of CGT and CTG showed an overall decreasing trend with increasing RTE variance (Fig. 6B, lower). None of them is rare codon.

Beside codon usage, RTE variance positively correlated with the mean value of RTE. In our data, there was also a weak correlation between RTE variance and mRNA level. Therefore, we needed to carefully separate the influences of the absolute value of RTE and mRNA level to examine whether codon usage directly contributes to the cross-condition RTE variability. We utilized a random forest model to quantify the contribution of different features to the prediction of RTE variance. The flowchart of the algorithm is shown in Fig. 6C. First of all, according to the median of RTE variance, we divided the 2914 genes into two clusters, which represent large and small RTE variance respectively. Then 80% of the genes were randomly sampled as the training set, leaving 20% as the test set. For the training set, Breiman's random forest algorithm was used to train a random forest model until the error converged. Different combinations of the features were separately used for training. By comparing the results from different feature combinations used for classification, we were able to quantify how much each single feature contributes to RTE variance. The receiver operating characteristic (ROC) curves suggested that the absolute value of RTE contributes most of the classification accuracy (Fig. 6D, yellow line). The addition of the feature mRNA level only improved the classification accuracy slightly (Fig. 6D, purple line). Nevertheless, the addition of the feature codon frequency improved the classification accuracy by approximately 10% (Fig. 6D, red line, and Table 2), suggesting a

297 nonnegligible and independent contribution from codon frequency to the cross-condition RTE
298 variability.
299 An advantage of random forest models is that the contribution of each feature to the classification
300 result can be quantified. The rank of codons contributing to classification from our random forest
301 model (Sup Fig. 8A-B) is consistent with the anti-correlated codons in Fig. 6A.
302 Furthermore, we examined whether other features contribute to RTE variance, such as the
303 distribution of the third base for codons, gene length, and translation pause motifs consisting of
304 adjacent double or triple codons[48-50]. The results showed that these features have little effect on
305 classification accuracy (Sup Fig. 8C, Sup Table 1). In addition, we used two other evaluation indices
306 to test whether the conclusion was robust with respect to different definitions of RTE variability: the
307 Fano factor and the coefficient of variation (CV, see Methods). In both cases, the addition of the
308 feature codon frequency markedly improved the classification accuracy (Sup Fig. 9), consistent with
309 our results using the index of RTE variance. In summary, codon usage contributed to the cross-
310 condition RTE variability of genes and the result was robust according to our tests.
311
312 **Codon-related RTE variability is an inherent feature of genes**
313 An intuitive hypothesis is that codon-related RTE variability could be due to the adaptation of tRNA
314 pools to the environment. Indeed, codon usage has been suggested as a mechanism of translational
315 regulation under oxidative stress or heat shock, as codon usage can be coupled to environment-
316 dependent factors such as the tRNA pool composition[25, 51]. An analogous extrapolation to explain
317 our observed codon-related RTE variability would be as follows: different nutrient conditions lead to
318 distinct compositions of the tRNA pool, so that genes with codon frequencies matching a particular
319 tRNA pool would have increased translation efficiency in the corresponding nutrient condition, thus
320 producing high cross-condition variability. This hypothesis predicts that codon-related RTE variability
321 would be condition-dependent. That is, there would be different sets of codons for high RTE genes for
322 each nutrient condition, and the identification of "high variability genes" would depend on which
323 conditions are being compared.
324 However, this hypothesis was found to test negative in our dataset. We compared the C-, N-, and
325 P-limited conditions in pairs. When any pair of conditions A and B were compared, genes with a
326 significantly higher RTE in A and those with significantly higher RTE in B actually share similar codon

frequencies: there are no "condition-specific" codons that distinguish high-RTE genes in A from those in B (Fig. 6E). By contrast, negative correlations of codon frequencies were observed between highly variable RTE genes and stable genes, between any pairs of conditions (Fig. 6E). These observations indicate that genes can be divided into two classes according to their RTE variability, which have to do with their codon usage, but are independent of nutrient conditions.

To further confirm that the codon-related RTE variability does not rely on specific conditions, we randomly selected sets of conditions from the 12 conditions to calculate RTE variance. Then the top 200 and bottom 200 genes of RTE variance were used to calculate the correlation coefficient of codon frequency. We found a clear downward trend of the correlation coefficient with increasing number of conditions, asymptoting to a strongly negative correlation of $r \sim -0.55$ when more than 8 conditions were picked. This indicates that codon-related RTE variability is an inherent feature of genes that applies across multiple conditions.

Discussion

How well transcript level represents protein abundance remains a controversial issue[2, 3]. Translational regulation is one of the key factors affecting the correlation between transcript level and protein abundance in bacteria[2]. In this work, we systematically examined the ribosomal behaviors in response to various nutrient conditions. Then combining ribosome profiling and RNA-seq in *E. coli*, we quantified genome-wide RTE under 12 conditions and observed a diverse range of gene-specific translational regulations in response to nutrient conditions. Furthermore, using a random forest model, we discovered that codon usage partially predicts the cross-condition RTE variability, such that a particular subset of codons, especially AAA (Lysine) and GAT (Aspartate), favors variability across all the nutrient conditions. By contrast, CGT (Arginine) and CTG (Leucine) disfavor RTE cross-condition variability (Sup Fig. 8). These findings broaden the understanding of translational regulation under environmental changes. What is more, our quantification of the contribution of codon usage to translational regulation can assist in the design of effective translation strategies in synthetic biology, as well as guide theoretical efforts to predict gene expression in response to environmental changes.

One important note is that the notion of RTE used in this work is slightly different from the TE in previous studies[17, 52]. RTE represents the relative ribosomal resources allocated by per unit length of mRNA molecules. It does not stand for the absolute translation efficiency (TE), which also includes global translation-related factors such as the total number, the working fraction, and the elongation speed of ribosomes. These global factors affect the TE of all genes as a whole[18], while RTE involves translational differences between individual genes. Therefore, by quantifying RTE, we capture the ribosomal resources devoted to translation at the single-gene level, and thus can compare translational regulation among different conditions, excluding the effect of global translation-related factors. In fact, according to comparison with previous studies on the translation efficiency of operons[17, 27], RTE reliably reflects translation differences between genes (Sup Fig. 3).

Protein biosynthesis consumes a large amount of building blocks and energy in fast growing bacteria[53]. To ensure efficient allocation of translation resources and so maximize cell growth, the protein synthesis rate is precisely controlled in proportion to the stoichiometry of complexes or hierarchical functions[17]. We found that the overall mRNA-RTE correlation across genes is not affected by mutations in single genes such as Δ*rplA* and Δ*leuB* which are involved in translation processes (Sup. Fig. 11). Previous studies reported gene-specific translational regulation in bacteria

under various stimuli[12, 13], which enables a faster response to environmental stresses than through transcriptional regulation[2, 54]. In our findings, both negative and positive mRNA-RTE correlations are likely biologically meaningful. For genes not sensitive to environmental changes, the mRNA level and RTE may be negatively correlated to stabilize protein production rate. For genes responding to specific nutrient limitations, the RTE may positively correlate with its mRNA level to amplify the change of protein synthesis rate, thus leading to a stronger correlation between mRNA level and protein abundance[55]. Gene-specific translational regulation is observed under C-, N-, and P-limitations. Therefore, the concerted regulation of transcription and translation may be a general strategy for cells to amplify their adaptation to environmental changes. In addition, the variance of RTE across conditions displays a large range, indicating that different genes are subject to varying degrees of translational regulation.

Also, according to our data, we suspected that translational regulation not only acts on genes responding to specific stressful conditions, but also acts on genes regulating translation itself, forming possible feedbacks[56]. Studies have revealed certain ribosomal proteins as feedback regulators, such as L1, S4, and S7 [57, 58]. Previously, this kind of feedback regulation was believed to be associated with growth-rate-dependent ribosome synthesis[28]. In our findings, the RTEs of several proteins involved in translational regulation correlate strongly with their mRNA levels, indicating concerted translational regulation. Feedback regulation on translation allows for better regulation in the overall translation activity of cells, providing one additional possible strategy for bacteria to rapidly and effectively respond to environment changes.

Our analysis suggested that codon usage not only contributes to condition-independent translation efficiency, but also partially predicts the variability of RTE across conditions. For condition-independent translation efficiency, multiple factors encoded in mRNA sequences affect the initiation, elongation, and termination of translation[5-7]. In particular, genome-scale studies have revealed significant association between codon usage and translation efficiency[59]. Codon usage *per se* mainly contributes during the elongation process[60], as it couples translation rates to the composition of the tRNA pool. However, our analysis indicated that under environmental stresses, the codon-related RTE variability across conditions was an inherent feature of genes, independent of specific conditions. Therefore, such RTE variability cannot be simply attributed to coupling between codon usage and the tRNA pool under any specific nutrient condition. This finding is consistent with the

speculation of a previous study that the change of tRNA composition leads to different translation efficiencies between stress-response and non-stress-response genes[25].

One limitation of our study is the lack of a detailed mechanism for how codons contribute to gene-specific translational regulation. As the translation process from mRNA to protein involves many factors, the differences in codon frequencies among mRNAs cannot be directly mapped to differences in translation efficiencies. In fact, it has been reported that there are complex interactions among multiple factors affecting translation, making it difficult to characterize the relation between codon frequency and translation efficiency[23]. For example, trade-offs between tRNA-mediated codon selection and mRNA structure entangle their separate roles[61]. Therefore, it remains an intriguing puzzle how codon frequency, a condition-invariant innate property of a gene, influences a gene's ability to respond to different conditions. We believe that in future research, a combination of technical approaches such as tRNA sequencing, mRNA structure probing, and translation-site-specific ribosome profiling will help uncover more mechanistic features of translational regulation[62, 63].

Methods

Cell strains and growth conditions

Escherichia coli strain NCM3722 was grown in batch or continuous cultures. Dilution rates of 0.1 h-1 and 0.6 h-1 were used to define slow and fast growth rates in chemostats. We utilized a 300mL volume chemostat (Sixfors, HT) with oxygen and pH probes to monitor the culture. The aeration rate was set at 4.5 l/h and pH was kept at 7.2 +/- 0.1. For minimal glucose media, 40 mM MOPS media (M2120, Teknova) was utilized with glucose (0.4%, Sigma G8270), ammonia (9.5 mM NH4Cl, Sigma A9434) and phosphate (1.32 mM K2HPO4, Sigma P3786) added separately. For defined rich media, the minimal media is supplemented with 10x ACGU (M2103, Teknova) and 5X Supplement EZ (M2104, Teknova). For carbon- and nitrogen-limited media, glucose and ammonia concentrations were reduced by 5-fold (0.08% and 1.9mM respectively). Phosphorus-limited medium contains 0.132 mM K_2HPO_4. Δ*leuB* and Δ*rplA* mutants were produced by P1 transduction from the KEIO collection[64] into Escherichia coli strain NCM3722.

Total RNA measurement

The method for RNA measurement was adapted from You et al.[65]. The culture was 1.5 mL and centrifuged at 13,000g for 1 min to form pellets. The pellet was frozen on dry ice and the supernatant was used to measure absorbance for cell loss at 600 nm. Then the pellet was washed twice with 0.6 M $HClO_4$, digested with 0.3 M KOH at 37 ℃ for 1h, and precipitated with 3 M $HClO_4$ to collect the supernatant. Then the pellets were extracted again with 0.5 M $HClO_4$. The supernatant was mixed and the absorbance was measured at 260 nm using Tecan Infinite 200 Pro (Tecan Trading AG, Switzerland). Finally, the total RNA concentration was the multiplication product of the absorbance value of A260 and the extinction coefficient (31 μg RNA mL^{-1}).

Total protein measurement

The protein measurement method was adapted from You et al.[65]. The culture was 1.5 mL and centrifuged at 13,000g for 1 min to form pellets. The cells were washed with 1mL MOPS buffer once, suspended in 200 μL water again, and then placed on dry ice. All the supernatant was collected and cell loss was measured with A600nm. Then the samples were thawed to measure protein content. The samples were added with 100 μL 3M NaOH and heated at 98 ℃ for 5 min. The samples were

cooled to 20 ℃ for 5min. After that, 300 μL 0.1% $CuSO_4$ was added in the samples for biuret assay. The samples were incubated at room temperature for 5 min and centrifuged at 13,000g for 1 min. The supernatant was then collected and the absorbance of 200 μL sample volume was measured at 555 nm using software Gen5 in a Microplate reader (Synergy HT, BioTek). The total protein concentration in the cell was inferred using a known concentration of appropriately diluted albumin (23209, Thermo).

Quantification of the total number and fraction of ribosomes

The calculation of the total number of ribosomes was adapted from Li et al.[18]. The total number of ribosomes was calculated as

$$R_{\mathrm{t}} = V_{\mathrm{c}} \cdot C_{\mathrm{p}} \cdot RPR \cdot \frac{f_r}{m_{\mathrm{r}}},$$

where the V_c is cell volume (m^3)[66], C_p is concentration of proteins (g/m^3)[67], *RPR* represents RNA-to-protein ratio, m_r is the mass of the rRNA component of a ribosome (g)[68], and f_r is the fractional mass of rRNA among total RNA. The quantification of f_r was adapted from Li et al.[18].

Polysome profiling was performed to quantify the ribosome fraction. The experimental methods were adapted from Li et al.[18]. The polysome profiling data was processed using customized MATLAB codes. The baseline absorbance was estimated using the average of the last 50 readings where RNA was not detected, and this background was subtracted. By fitting the exponential decay function to the first peak of the non-ribosomal signal source, free nucleotides and tRNA backgrounds were removed. Then each ribosome peak was selected and quantified by the area under the curve.

In order to quantify different kinds of ribosomes in the 70S peak, 170 mM KCl was used instead of 100 mM NH_4Cl. Cytolysis products were loaded into 10-30% linear gradient and centrifuged at 35,000 r.p.m. in a SW41Ti barrels for 5h at 4 ℃. Then the MATLAB file-exchange scheme, Peakfit (2.0) esd used to fit the three overlapped peaks (50S subunit, 70S without mRNA, and 70S with mRNA binding) into three Gaussian distributions.

***lacZ* induction and translational elongation rate measurement**

The measurement of ribosome elongation rate was adapted from Zhu et al.[69]. Isopropyl-β-D-thiogalactoside (IPTG) (I2481C-25, Gold Biotechnology) with concentration of 5mM was added to the culture. Every 15 seconds, 1 mL of culture medium was taken and placed in a tube containing 10 μL

of 100mm chloramphenicol, immediately frozen in liquid nitrogen and stored at -20℃, followed by subsequent measurements. After thawing, 400 μL of the sample was added to 100 μL of 5xZ buffer solution (0.3M $Na_2HPO_4.7H_2O$, 0.2M $NaH_2PO_4.H_2O$, 50mM KCl, 5mM $MgSO_4$, 20 mM β-mercaptoethanol) and incubated at 37 ℃ for 10 minutes. 100 μL 4 mg mL^{-1} 4-methylumbelliferyl-β-D-galactopyranoside (MUG, 337210010, ACROS Organics) in DMSO was added to each sample every 10 s for precise control of the reaction time. The samples were incubated in Eppendorf Thermomixer R at 37 ℃ at a mixing rate of 1400 r.p.m. for 30 min to 2 h, according to the enzyme expression levels. Then we added 300 μL 1 M Na_2CO_3 to stop the reaction. The tube was spun down at 16,000g for 3 min to precipitate cell debris. Finally, the fluorescence of 200 μL supernatant was measured with a microplate reader (365 nm excitation and 450 nm emission filter). We integrated the signals and performed a linear fit to infer the ribosome elongation rate. According to the previous study[69], the elongation time was corrected by subtracting 10 s from the measured delay time.

RNA extraction and ribosome profiling

The method of RNA extraction and ribosome profiling is described in Li et al.[18]. The cell collection step was the same as for polysome profiling in Li et al.[18] except that 1mM chloramphenicol was utilized in the sucrose solution. The footprinting and library preparation steps were adapted from Li et al.[17] After quantification of RNA concentration with NanoDrop, samples with 500μg RNA were digested with 750U MNase (10107921001, Roche) for 1 hour at 25˚C before being quenched with 6mM EGTA. The lysates were then layered onto a 10%-55% sucrose gradient and centrifuged. The monosome fraction was collected and snap frozen in liquid nitrogen. There were no observed polysome peaks, which indicated thorough digestion. The RNA was separated using hot phenol and size selected on 15% TBE-Urea PAGE gels run for 1 hour at 210V. Gels were stained with SYBR Gold and visualized using Dark Reader (Clare Chemical Research). Finally, RNA fragments with size between 25-40 nt were extracted using isopropanol precipitation.

Library preparation and sequencing

RNA fragments from footprints were dephosphorylated at the 3' end by PNK (M0201, NEB). The repaired fragments were linked to the Universal miRNA Cloning Linker (S1315S, NEB), reverse transcribed (18080044, Thermo), and circularized (CL4111K, Epicentre). rRNA was subtracted from

the circularized samples before PCR amplification (M0531L, NEB) and size selection. High quality PCR samples were checked by Bioanalyzer highly sensitive DNA chip. Deep sequencing was performed by Illumina HiSeq 2500 on Rapid flowcells with settings of single end and 75 nt-long read length.

Mapping and sequencing data analysis

Data processing including barcode splitting, linker trimming, and mapping were performed using Galaxy[70]. The processed reads were mapped to Escherichia coli genome escherichia_coli_k12_nc_000913_3 from the NCBI database with the BWA short read mapping algorithm[71]. Only the reads between 20-45 nt that aligned to the coding region were extracted for further analysis.

To infer the ribosome A-site position, python package Plastid[72] was used to align the 3' end of reads to the stop and start codons[73], which are known to have higher ribosome densities. We found that the offsets were 12 nt for stop codon and 15 nt for start codon. Therefore, we utilized 11nt for A site position and 14nt for P site. Further analysis was done using MATLAB and R codes.

Analysis of deep-sequencing data

The counts from ribosome profiling and RNA-seq were used to calculate relative translational efficiency (RTE) for each transcript:

$$[\mathrm{RTE}] = \frac{[footprinting\ counts]/[gene\ length]}{[relative\ mRNA\ level]},$$

where the footprinting counts were normalized by the total counts in one experiment, reflecting the percentage of ribosomes occupied by a gene. The ratio of footprinting counts to gene length reflects the relative ribosome density: the percentage of ribosomes occupied by per unit length of a gene. The relative mRNA levels were also normalized to the total counts and gene length as reads per kilobase million (RPKM). In addition, genes with $\log_{10}$(RNA-seq RPKM) > 1.5 were selected for subsequent analysis (selected genes n = 2914).

Mean levels were taken as the average of the 12 conditions for analyzing the correlation between the mRNA level and RTE across genes. The Spearman's rank correlation coefficient was used for correlations both across genes and across conditions. In order to test the significance of the

distribution of correlation coefficients between mRNA and RTE across conditions, the RTE values for each gene were randomly scrambled among the 12 conditions. The resulting randomly ordered RTEs were used to recalculate the distribution of correlation coefficients, which was considered as the null distribution. Then we used the package *ttest2* in MATLAB to test whether the two distributions are significantly different, and calculated the *p*-value.

When comparing two different nutritional restriction conditions, the RNA-seq RPKM were averaged for three biological replicates. Then we screened for differential gene groups with $\log_2$(mRNA fold change) > 4 or < -4 and *p*-value < 0.05. To test the significance of RTE fold changes for the genes with differentially expressed mRNA, we first calculate the RTE fold change distribution for this group of genes. Then the distribution of the RTE fold changes for the whole set of 2914 genes was considered as the null distribution. A *p*-value was calculated using student's *t*-test for the two distributions. All the above processes were performed with Matlab2020a.

GO analysis and KEGG pathway analysis

Functional enrichment analysis was carried out using function *enrichGO* in R package *clusterProfiler*[74]. In addition, genome wide annotation *org.EcK12.eg.db* for *E. coli* strain K12 was used. The enrichment results were filtered with an adjusted *p*-value < 0.05. Furthermore, function *dropGO* was used to refine gene ontology level. Besides, KEGG pathway enrichment analysis was carried out using function *enrichKEGG* in R package *clusterProfiler*[74]. Genes contained in the 82 pathways of *E. coli* strain K-12 MG1655 were obtained from https://www.genome.jp/kegg/pathway.html.

Codon usage analysis

The codon frequency of a gene was defined as the ratio of the number of a certain codon to the total number of codons. The frequencies of 64 codons constituted the codon frequency vector of a gene. Then we calculated the background codon frequencies from the complete set of analyzed genes. To characterize the bias for a gene towards certain codons, the background codon frequencies were subtracted from the codon frequency vector.

Before comparing the codon usage between different pathways, the overlapped genes were removed. Then we calculated the average codon frequencies for all genes in a pathway. As shown in Fig 5G, the rarity of codons was ranked according to their background frequencies.

Evaluation indices for RTE variability

We used three different evaluation indices: the variance, the Fano factor, and the coefficient of variation (CV). The variance is defined as

$$\mathrm{var}(\mathrm{RTE}) = \frac{\sum\left(\mathrm{TE} - \overline{\mathrm{RTE}}\right)^2}{n-1},$$

where $\overline{\mathrm{RTE}}$ is the sample mean of RTE, and the n is the sample size of RTE. The Fano factor is defined as

$$\mathrm{Fano}(\mathrm{RTE}) = \frac{\sigma_{\mathrm{RTE}}^2}{\mu_{\mathrm{RTE}}},$$

where σ_{RTE}^2 is the variance of RTE, and the μ_{RTE} is the sample mean of RTE. The CV is defined as

$$\mathrm{CV}(\mathrm{RTE}) = \frac{\sigma_{\mathrm{RTE}}}{\mu_{\mathrm{RTE}}},$$

Where σ_{RTE} is the standard deviation of RTE, and the μ_{RTE} is the sample mean of RTE.

Random forest algorithm

We used the package *TreeBagger* in MATLAB to build the binary classification model. The number of trees was set to 200 and the minimum number of observations per tree leaf was set to 5. The number of variables to select at random for each decision split was set to the square root of the total variable number. In our model, the total variable number is 64, corresponding to the 64 codons. Finally, Breiman's random forest algorithm was invoked to perform the training[75].

As stated in the main text, features such as frequencies of the 64 codons, mRNA level, RTE absolute value, the distribution of the third base of codons, and gene length were selected and combined to determine their contribution to classification results. In addition, the frequencies of typical translation pause motifs were also used as classification features.

1000 random samplings of the dataset were performed to exclude the contingency of results. We used true positive rate to evaluate the sensitivity, defined as

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

where TP and FN refer to the number of true positives and false negatives, respectively. The specificity is defined as

$$\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}},$$

The area under curve (AUC) was calculated as the area under ROC curve. To calculate the sensitivity and specificity, a classification threshold is needed. The score for each gene from the model is in the range of [0, 1]. If the score is above the threshold, it is considered a positive sample, otherwise it is considered a negative sample. The results shown in Table 2 used 0.5 as the classification threshold.

Availability of data and materials

The RNA-seq and ribosome profiling data is available in the GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE182100).

Accession numbers

The RNA-seq and ribosome profiling data is available in the GEO database with accession number GSE182100. Data will be available after the article is published.

Supplementary information

Supplementary information is available at bioRxiv online.

Acknowledgements

This work was supported by grants from Peking-Tsinghua Center for Life Sciences. This work was supported in part by the National Science Foundation, through the Center for the Physics of Biological Function (PHY-1734030).

Funding

This work was supported by the National Natural Science Foundation of China (grant number: 32071255); U.S. Department of Health and Human Services, National Institutes of Health (grant number: R01 GM082938, DP1AI124669); and National Science Foundation (PHY-1607612) and through the Center for the Physics of Biological Function (PHY-1734030). Funding for open access charge: National Natural Science Foundation of China.

Competing interests

The authors declare no competing interests.

References

1. Ptashne M: **Regulation of transcription: from lambda to eukaryotes.** *Trends Biochem Sci* 2005, **30:**275-279.
2. Liu Y, Beyer A, Aebersold R: **On the Dependency of Cellular Protein Levels on mRNA Abundance.** *Cell* 2016, **165:**535-550.
3. Nie L, Wu G, Zhang W: **Correlation of mRNA expression and protein abundance affected by multiple sequence features related to translational efficiency in Desulfovibrio vulgaris: a quantitative analysis.** *Genetics* 2006, **174:**2229-2243.
4. Lu P, Vogel C, Wang R, Yao X, Marcotte EM: **Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation.** *Nat Biotechnol* 2007, **25:**117-124.
5. Mauger DM, Cabral BJ, Presnyak V, Su SV, Reid DW, Goodman B, Link K, Khatwani N, Reynders J, Moore MJ, McFadyen IJ: **mRNA structure regulates protein expression through changes in functional half-life.** *Proc Natl Acad Sci U S A* 2019, **116:**24075-24083.
6. Boel G, Letso R, Neely H, Price WN, Wong KH, Su M, Luff J, Valecha M, Everett JK, Acton TB, et al: **Codon influence on protein expression in E. coli correlates with mRNA levels.** *Nature* 2016, **529:**358-363.
7. Fredrick K, Ibba M: **How the sequence of a gene can tune its translation.** *Cell* 2010, **141:**227-229.
8. Serganov A, Nudler E: **A decade of riboswitches.** *Cell* 2013, **152:**17-24.
9. Arenz S, Meydan S, Starosta AL, Berninghausen O, Beckmann R, Vázquez-Laslop N, Wilson DN: **Drug sensing by the ribosome induces translational arrest via active site perturbation.** *Molecular cell* 2014, **56:**446-452.
10. Johansson J, Mandin P, Renzoni A, Chiaruttini C, Springer M, Cossart P: **An RNA thermosensor controls expression of virulence genes in Listeria monocytogenes.** *Cell* 2002, **110:**551-561.
11. Chen WJ, Zhu T: **Networks of transcription factors with roles in environmental stress response.** *Trends Plant Sci* 2004, **9:**591-596.
12. Bucca G, Pothi R, Hesketh A, Moller-Levet C, Hodgson DA, Laing EE, Stewart GR, Smith CP: **Translational control plays an important role in the adaptive heat-shock response of Streptomyces coelicolor.** *Nucleic Acids Res* 2018, **46:**5692-5703.
13. Zhang Y, Xiao Z, Zou Q, Fang J, Wang Q, Yang X, Gao N: **Ribosome Profiling Reveals Genome-wide Cellular Translational Regulation upon Heat Stress in Escherichia coli.** *Genomics Proteomics Bioinformatics* 2017, **15:**324-330.
14. Chionh YH, McBee M, Babu IR, Hia F, Lin W, Zhao W, Cao J, Dziergowska A, Malkiewicz A, Begley TJ, et al: **tRNA-mediated codon-biased translation in mycobacterial hypoxic persistence.** *Nature Communications* 2016, **7:**13302.
15. Samatova E, Daberger J, Liutkute M, Rodnina MV: **Translational Control by Ribosome Pausing in Bacteria: How a Non-uniform Pace of Translation Affects Protein Production and Folding.** *Frontiers in microbiology* 2021, **11:**619430-619430.
16. Garbuz DG: **Regulation of heat shock gene expression in response to stress.** *Molecular Biology* 2017, **51:**352-367.
17. Li GW, Burkhardt D, Gross C, Weissman JS: **Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources.** *Cell* 2014, **157:**624-635.

18. Li SH, Li Z, Park JO, King CG, Rabinowitz JD, Wingreen NS, Gitai Z: **Escherichia coli translation strategies differ across carbon, nitrogen and phosphorus limitation conditions.** *Nat Microbiol* 2018, **3:**939-947.
19. Webster MW, Takacs M, Zhu C, Vidmar V, Eduljee A, Abdelkareem M, Weixlbaumer A: **Structural basis of transcription-translation coupling and collision in bacteria.** *Science* 2020, **369:**1355-1359.
20. Dykeman EC: **A stochastic model for simulating ribosome kinetics in vivo.** *PLoS Comput Biol* 2020, **16:**e1007618.
21. Levin D, Tuller T: **Whole cell biophysical modeling of codon-tRNA competition reveals novel insights related to translation dynamics.** *PLoS Comput Biol* 2020, **16:**e1008038.
22. Shaham G, Tuller T: **Genome scale analysis of Escherichia coli with a comprehensive prokaryotic sequence-based biophysical model of translation initiation and elongation.** *DNA Res* 2018, **25:**195-205.
23. Rodnina MV: **The ribosome in action: Tuning of translational efficiency and protein folding.** *Protein Sci* 2016, **25:**1390-1406.
24. Chan C, Pham P, Dedon PC, Begley TJ: **Lifestyle modifications: coordinating the tRNA epitranscriptome with codon bias to adapt translation during stress responses.** *Genome Biol* 2018, **19:**228.
25. Torrent M, Chalancon G, de Groot NS, Wuster A, Madan Babu M: **Cells alter their tRNA abundance to selectively regulate protein synthesis during stress conditions.** *Science Signaling* 2018, **11:**eaat6409.
26. Zhong J, Xiao C, Gu W, Du G, Sun X, He QY, Zhang G: **Transfer RNAs Mediate the Rapid Adaptation of Escherichia coli to Oxidative Stress.** *PLoS Genet* 2015, **11:**e1005302.
27. Burkhardt DH, Rouskin S, Zhang Y, Li G-W, Weissman JS, Gross CA: **Operon mRNAs are organized into ORF-centric structures that predict translation efficiency.** *eLife* 2017, **6:**e22037.
28. Cole JR, Nomura M: **Translational regulation is responsible for growth-rate-dependent and stringent control of the synthesis of ribosomal proteins L11 and L1 in Escherichia coli.** *Proceedings of the National Academy of Sciences* 1986, **83:**4129.
29. Sander G: **Ribosomal protein L1 from Escherichia coli. Its role in the binding of tRNA to the ribosome and in elongation factor g-dependent gtp hydrolysis.** *Journal of Biological Chemistry* 1983, **258:**10098-10103.
30. Wallon G, Yamamoto K, Kirino H, Yamagishi A, Lovett ST, Petsko GA, Oshima T: **Purification, catalytic properties and thermostability of 3-isopropylmalate dehydrogenase from Escherichia coli.** *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* 1997, **1337:**105-112.
31. Subramanian AR, Dabbs ER: **Functional Studies on Ribosomes Lacking Protein L1 from Mutant Escherichia coli.** *European Journal of Biochemistry* 1980, **112:**425-430.
32. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS: **Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling.** *Science* 2009, **324:**218.
33. Grass G, Otto M, Fricke B, Haney CJ, Rensing C, Nies DH, Munkelt D: **FieF (YiiP) from Escherichia coli mediates decreased cellular accumulation of iron and relieves iron stress.** *Archives of Microbiology* 2005, **183:**9-18.

34. Strader MB, Costantino N, Elkins CA, Chen CY, Patel I, Makusky AJ, Choy JS, Court DL, Markey SP, Kowalak JA: **A proteomic and transcriptomic approach reveals new insight into beta-methylthiolation of Escherichia coli ribosomal protein S12.** *Molecular & cellular proteomics : MCP* 2011, **10:**M110.005199-M005110.005199.
35. Petibon C, Malik Ghulam M, Catala M, Abou Elela S: **Regulation of ribosomal protein genes: An ordered anarchy.** *WIREs RNA* 2021, **12:**e1632.
36. Wada A, Igarashi K, Yoshimura S, Aimoto S, Ishihama A: **Ribosome Modulation Factor: Stationary Growth Phase-Specific Inhibitor of Ribosome Functions from Escherichia coli.** *Biochemical and Biophysical Research Communications* 1995, **214:**410-417.
37. Gimenez R, Nuñez MF, Badia J, Aguilar J, Baldoma L: **The Gene yjcG, Cotranscribed with the Gene acs, Encodes an Acetate Permease in <em>Escherichia coli</em&gt.** *Journal of Bacteriology* 2003, **185:**6448.
38. Harayama S, Bollinger J, Iino T, Hazelbauer GL: **Characterization of the mgl operon of Escherichia coli by transposon mutagenesis and molecular cloning.** *Journal of bacteriology* 1983, **153:**408-415.
39. Loh KD, Gyaneshwar P, Markenscoff Papadimitriou E, Fong R, Kim K-S, Parales R, Zhou Z, Inwood W, Kustu S: **A previously undescribed pathway for pyrimidine catabolism.** *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103:**5114.
40. Parales RE, Ingraham JL: **The Surprising Rut Pathway: an Unexpected Way To Derive Nitrogen from Pyrimidines.** *Journal of Bacteriology* 2010, **192:**4086.
41. Metcalf WW, Wanner BL: **Involvement of the Escherichia coli phn (psiD) gene cluster in assimilation of phosphorus in the form of phosphonates, phosphite, Pi esters, and Pi.** *Journal of Bacteriology* 1991, **173:**587.
42. Stasi R, Neves HI, Spira B: **Phosphate uptake by the phosphonate transport system PhnCDE.** *BMC Microbiology* 2019, **19:**79.
43. Kanehisa M, Goto S: **KEGG: Kyoto Encyclopedia of Genes and Genomes.** *Nucleic Acids Research* 2000, **28:**27-30.
44. Swim HE, Krampitz LO: **Acetic acid oxidation by Escherichia coli; evidence for the occurrence of a tricarboxylic acid cycle.** *Journal of bacteriology* 1954, **67:**419-425.
45. de Kok A, Hengeveld AF, Martin A, Westphal AH: **The pyruvate dehydrogenase multi-enzyme complex from Gram-negative bacteria.** *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* 1998, **1385:**353-366.
46. Li H, Sourjik V: **Assembly and stability of flagellar motor in Escherichia coli.** *Molecular Microbiology* 2011, **80:**886-899.
47. Wadhams GH, Armitage JP: **Making sense of it all: bacterial chemotaxis.** *Nature Reviews Molecular Cell Biology* 2004, **5:**1024-1037.
48. Peil L, Starosta AL, Lassak J, Atkinson GC, Virumae K, Spitzer M, Tenson T, Jung K, Remme J, Wilson DN: **Distinct XPPX sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor EF-P.** *Proc Natl Acad Sci U S A* 2013, **110:**15265-15270.
49. Samatova E, Daberger J, Liutkute M, Rodnina MV: **Translational Control by Ribosome Pausing in Bacteria: How a Non-uniform Pace of Translation Affects Protein Production and Folding.** *Front Microbiol* 2020, **11:**619430.

50. Tesina P, Lessen LN, Buschauer R, Cheng J, Wu CC, Berninghausen O, Buskirk AR, Becker T, Beckmann R, Green R: **Molecular mechanism of translational stalling by inhibitory codon combinations and poly(A) tracts.** *EMBO J* 2020, **39:**e103365.
51. Frumkin I, Lajoie MJ, Gregg CJ, Hornung G, Church GM, Pilpel Y: **Codon usage of highly expressed genes affects proteome-wide translation efficiency.** *Proc Natl Acad Sci U S A* 2018, **115:**E4940-E4949.
52. Li GW: **How do bacteria tune translation efficiency?** *Curr Opin Microbiol* 2015, **24:**66-71.
53. Russell JB, Cook GM: **Energetics of bacterial growth: balance of anabolic and catabolic reactions.** *Microbiol Rev* 1995, **59:**48-62.
54. Tollerson R, 2nd, Ibba M: **Translational regulation of environmental adaptation in bacteria.** *The Journal of biological chemistry* 2020, **295:**10434-10445.
55. Lahtvee PJ, Sanchez BJ, Smialowska A, Kasvandik S, Elsemman IE, Gatto F, Nielsen J: **Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast.** *Cell Syst* 2017, **4:**495-504 e495.
56. Nomura M, Dean D, Yates JL: **Feedback regulation of ribosomal protein synthesis in Escherichia coli.** *Trends in Biochemical Sciences* 1982, **7:**92-95.
57. Nomura M, Yates JL, Dean D, Post LE: **Feedback regulation of ribosomal protein gene expression in Escherichia coli: structural homology of ribosomal RNA and ribosomal protein MRNA.** *Proceedings of the National Academy of Sciences* 1980, **77:**7084.
58. Yates JL, Nomura M: **Feedback regulation of ribosomal protein synthesis in E. coli: Localization of the mRNA target sites for repressor action of ribosomal protein L1.** *Cell* 1981, **24:**243-249.
59. Tuller T, Waldman YY, Kupiec M, Ruppin E: **Translation efficiency is determined by both codon bias and folding energy.** *Proc Natl Acad Sci U S A* 2010, **107:**3645-3650.
60. Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, Pan T, Dahan O, Furman I, Pilpel Y: **An evolutionarily conserved mechanism for controlling the efficiency of protein translation.** *Cell* 2010, **141:**344-354.
61. Gorochowski TE, Ignatova Z, Bovenberg RA, Roubos JA: **Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate.** *Nucleic Acids Res* 2015, **43:**3022-3032.
62. Pang Yan Ling J, Abo R, Levine SS, Dedon PC: **Diverse cell stresses induce unique patterns of tRNA up- and down-regulation: tRNA-seq for quantifying changes in tRNA copy number.** *Nucleic Acids Research* 2014, **42:**e170-e170.
63. Wang Y, Zhang H, Lu J: **Recent advances in ribosome profiling for deciphering translational regulation.** *Methods* 2020, **176:**46-54.
64. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H: **Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection.** *Molecular systems biology* 2006, **2:**2006.0008-2006.0008.
65. You C, Okano H, Hui S, Zhang Z, Kim M, Gunderson CW, Wang Y-P, Lenz P, Yan D, Hwa T: **Coordination of bacterial proteome with metabolism by cyclic AMP signalling.** *Nature* 2013, **500:**301-306.
66. Bremer H, Dennis PP: **Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates.** *EcoSal Plus* 2008, **3**.

67. Albe KR, Butler MH, Wright BE: **Cellular concentrations of enzymes and their substrates.** *J Theor Biol* 1990, **143:**163-195.
68. Sundararaj S, Guo A, Habibi-Nazhad B, Rouani M, Stothard P, Ellison M, Wishart DS: **The CyberCell Database (CCDB): a comprehensive, self-updating, relational database to coordinate and facilitate in silico modeling of Escherichia coli.** *Nucleic Acids Res* 2004, **32:**D293-295.
69. Zhu M, Dai X, Wang Y-P: **Real time determination of bacterial in vivo ribosome translation elongation speed based on LacZα complementation system.** *Nucleic acids research* 2016, **44:**e155-e155.
70. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C, et al: **The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update.** *Nucleic Acids Research* 2016, **44:**W3-W10.
71. Li H, Durbin R: **Fast and accurate long-read alignment with Burrows–Wheeler transform.** *Bioinformatics* 2010, **26:**589-595.
72. Dunn JG, Weissman JS: **Plastid: nucleotide-resolution analysis of next-generation sequencing and genomics data.** *BMC Genomics* 2016, **17:**958.
73. Woolstenhulme CJ, Guydosh NR, Green R, Buskirk AR: **High-precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP.** *Cell reports* 2015, **11:**13-21.
74. Yu G, Wang L-G, Han Y, He Q-Y: **clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters.** *OMICS: A Journal of Integrative Biology* 2012, **16:**284-287.
75. Breiman L: **Random Forests.** *Machine Learning* 2001, **45:**5-32.

Tables

Condition ID	1	2	3	4	5	6	7	8	9	10	11	12
Nutrient limitation	C-limited		N-limited		P-limited		glucose minimal	defined rich MOPS	N-limited for ΔrplA		Leu-limited for ΔleuB	
Growth rate (h^{-1})	0.1	0.6	0.1	0.6	0.1	0.6	0.9	1.8	0.1	0.6	0.1	0.6
Culture environment	chemostat						batch culture		chemostat			

Table 1. List of the 12 different conditions for *Escherichia coli* in our measurements. *E. coli* was grown under glucose (C, carbon), ammonia (N, nitrogen) and phosphate (P, phosphorus) limited conditions in chemostats at two different dilution rates of 0.1 and 0.6 h^{-1} (equal to growth rates). ΔrplA and ΔleuB mutant strains were grown under ammonia and leucine limitations, respectively. Three biological replicates were performed for all the 12 conditions.

Features	Sensitivity	Specificity	Accuracy	AUC
mRNA	53.4±3.19%	53.47±3.08%	53.41±1.81%	0.54±0.02
Codon	64.27±3.08%	71.96±2.9%	68.07±1.76%	0.75±0.02
RTE	72.58±2.68%	72.5±2.61%	72.52±1.61%	0.8±0.02
Codon + mRNA	66.44±3.07%	72.53±2.91%	69.45±1.87%	0.77±0.02
Codon + RTE	81.95±2.33%	81.53±2.31%	81.72±1.52%	0.89±0.01
Codon + mRNA + RTE	**84.36±2.19%**	**82.99±2.24%**	**83.66±1.5%**	**0.91±0.01**

Table 2. Classification results of the random forest model. The table shows the average results with S.D. from a thousand random samples.

Figures

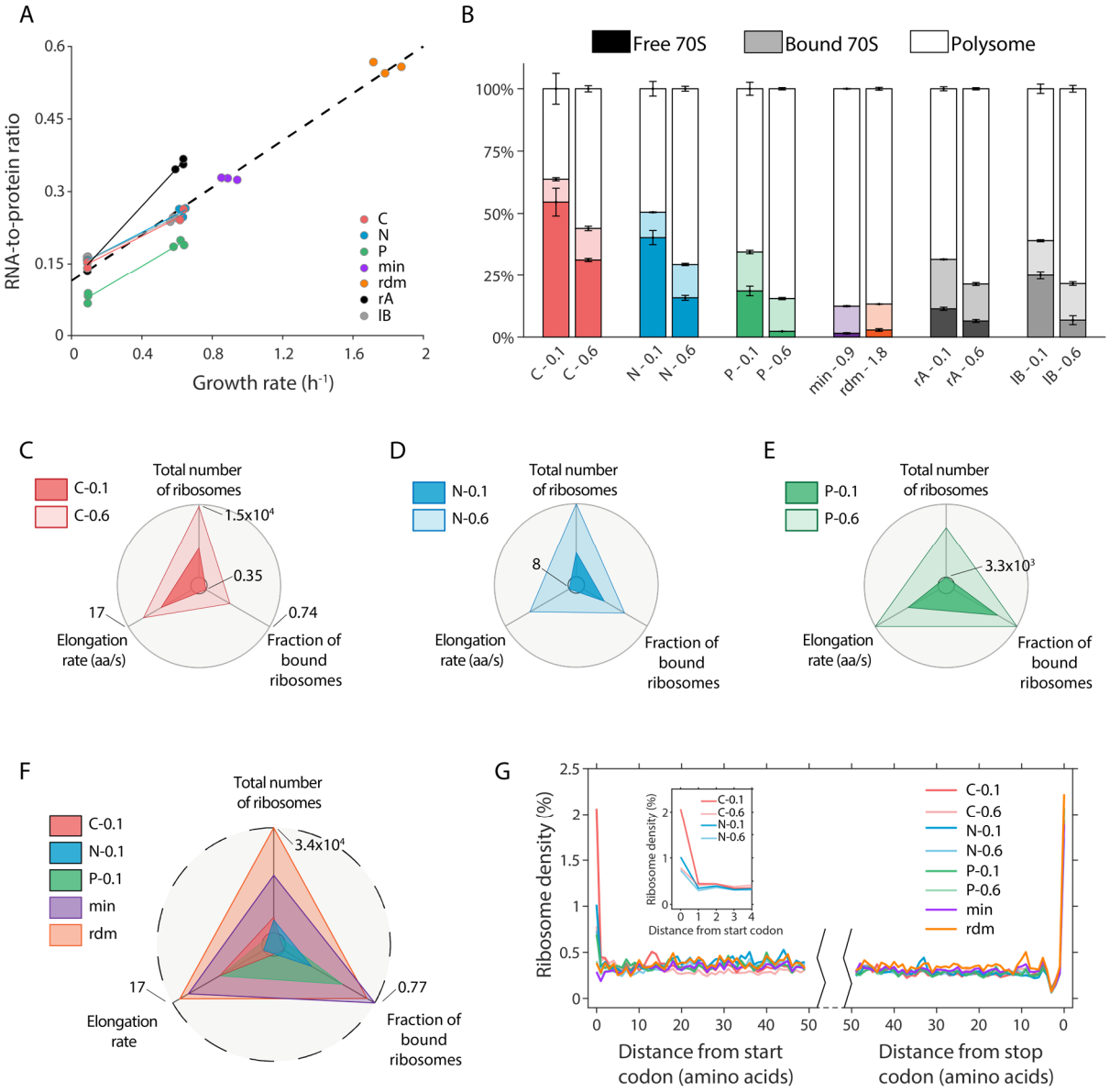

Figure 1. Cells adapt to nutrient conditions through different ribosomal strategies.

(A) RNA-to-protein ratios for 12 conditions at different growth rates. Each data point represents one experimental measurement.

(B) Fractions of assembled (70S) ribosomes under 12 conditions. The assembled ribosomes include free 70S monosomes, mRNA-bound 70S monosomes, and mRNA-bound 70S polysomes (multiple ribosomes on one mRNA). Free 70S, bound 70S, and polysome are represented in white, light, and dark colors, respectively. The bar heights represent mean values with error bars indicating s.e.m. from three biological replicates.

(C-E) Cells differentially regulate three ribosomal features in response to C-, N-, and P-limitations at the growth rates of 0.1 and 0.6 h^{-1}. The three features include total number of ribosomes per average cell (see Methods), elongation rate, and fraction of bound ribosomes. These features are scaled linearly between the inner circle and the outer circle, which represent the minimum and maximum among all conditions, respectively. The scales of the three indicated axes are the same for panels C-E.

(F) Same as (C-E), but showing the differences between chemostat cultures and two batch conditions. The value of the outermost circle is larger than in C-E, especially the total number of ribosomes.

(G) Averaged A-site ribosome density within the first and last 50 codons of the transcripts, from ribosome profiling analysis. Each curve shows the mean value from three replicates at each condition. Inset: ribosome density at the beginning of the transcripts.

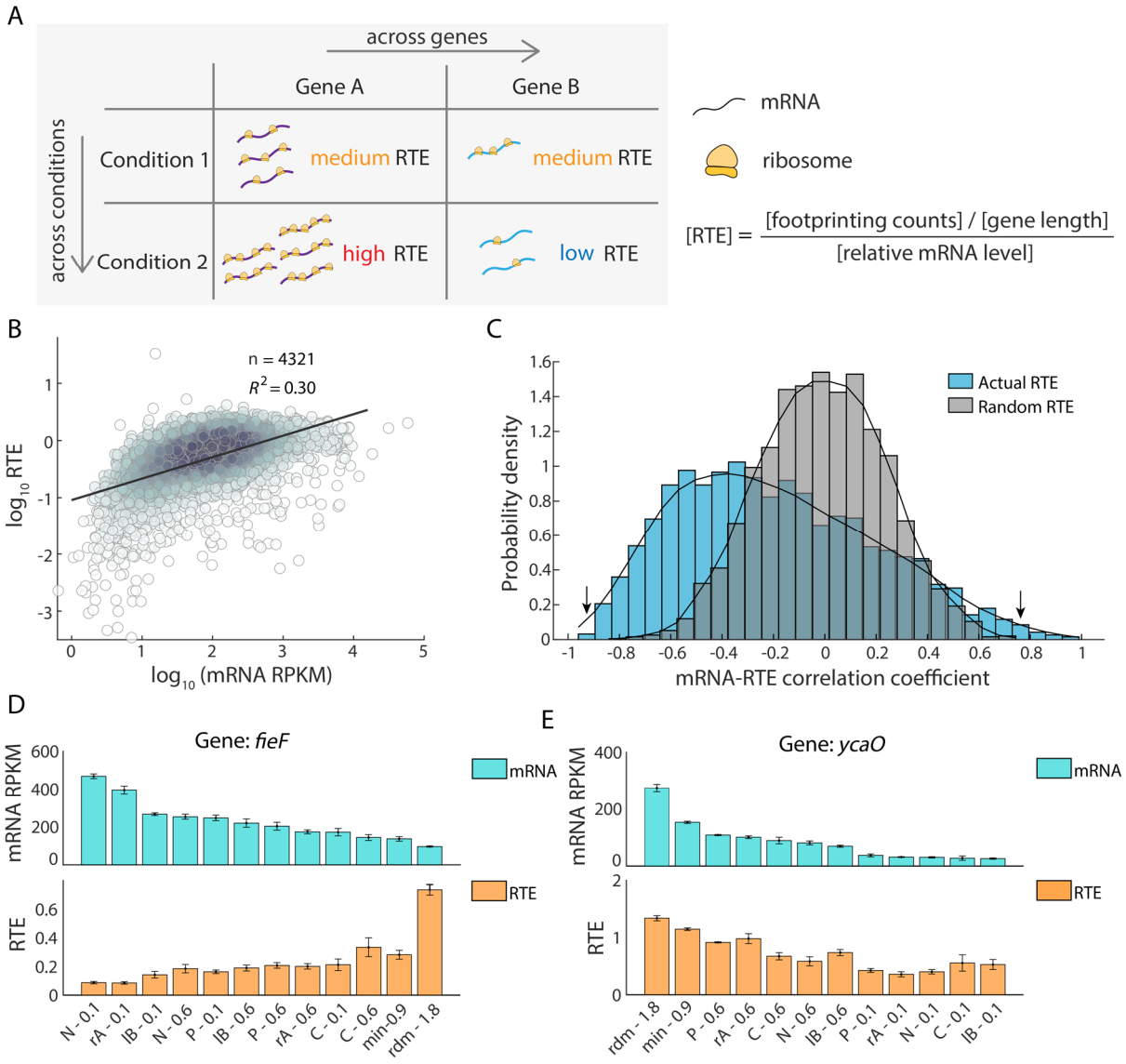

Figure 2. Global view of mRNA-Relative Translation Efficiency (RTE) correlations across genes and across conditions.

(A) Two types of correlations between mRNA level and RTE: across genes and across conditions.

(B) Correlation between mean mRNA level and mean RTE, across different genes. Mean levels were taken as the average of all 12 conditions (Table 1). Each dot represents one gene, and color depth depicts the density of points.

(C) Distribution of Spearman's rank correlation coefficients between mRNA level and RTE across the 12 different conditions. Each gene provides one such correlation coefficient, and distributions are shown for 2914 genes (blue bars – original data; gray bars – scrambling the RTEs among conditions for each gene).

(D-E) Example of two genes with negative correlation (D, gene *fieF*, left arrow in B) and positive correlation (E, gene *ycaO*, right arrow in B) between their mRNA levels and RTEs. Error bars represent s.e.m. from three biological replicates.

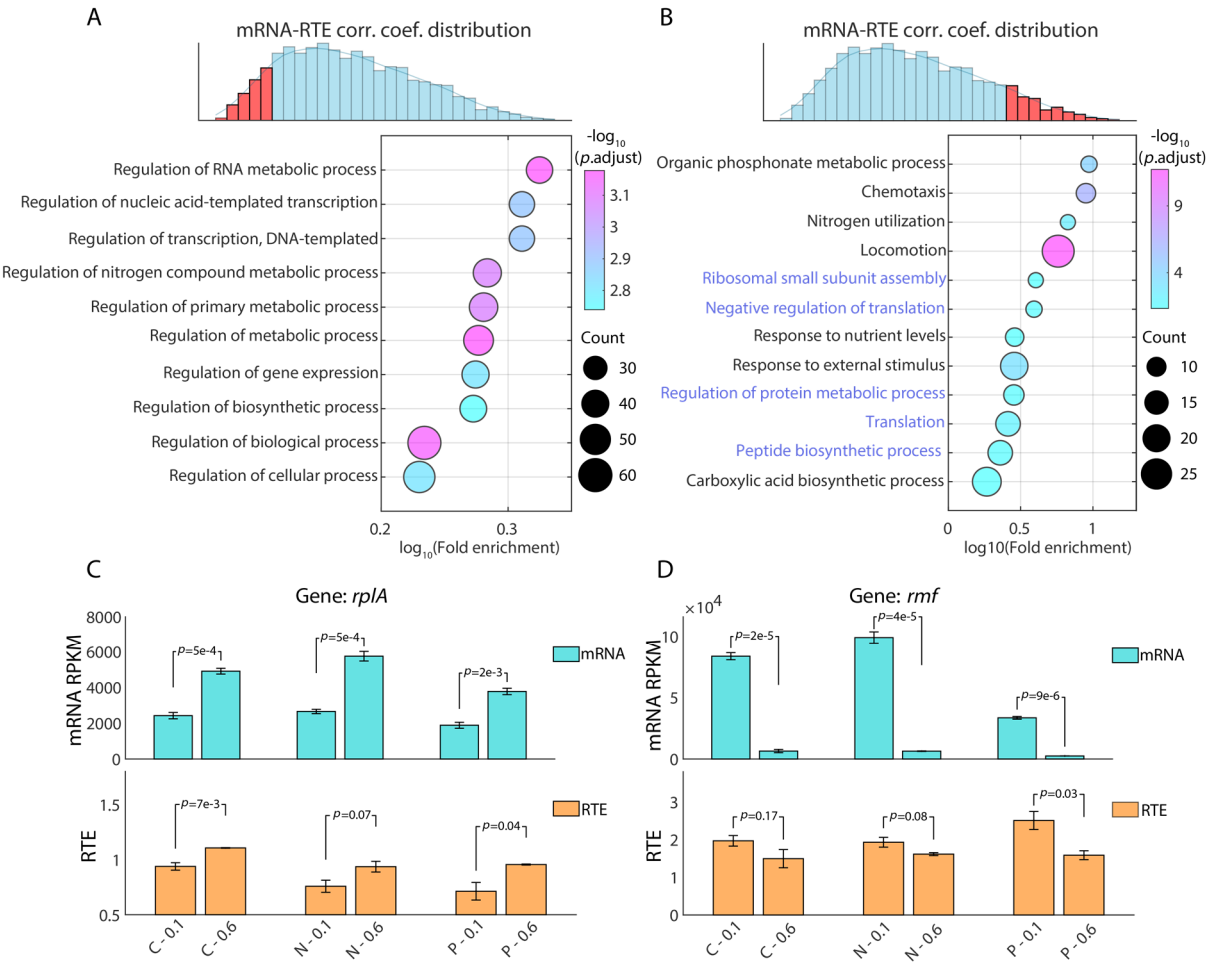

Figure 3. The correlation between mRNA level and RTE is related to specific gene functions.

(A-B) Gene ontology (GO) enrichment for the top 300 genes with the most negative correlation (A) and the most positive correlation (B) between mRNA levels and RTEs. The color of the dots represents the -$\log_{10}$ adjusted *p*-value, and the dot size represents the number of genes appearing in each biological process.

(C-D) Genes regulating translation are themselves subject to translational regulation. Examples of positive correlation between mRNA level and RTE across different growth rates for one gene that promotes translation (C, gene *rplA*) and one that inhibits translation (D, gene *rmf*). Student's *t*-test was used to calculate the *p*-value. Reads Per Kilobase Million (RPKM) is used for mRNA level.

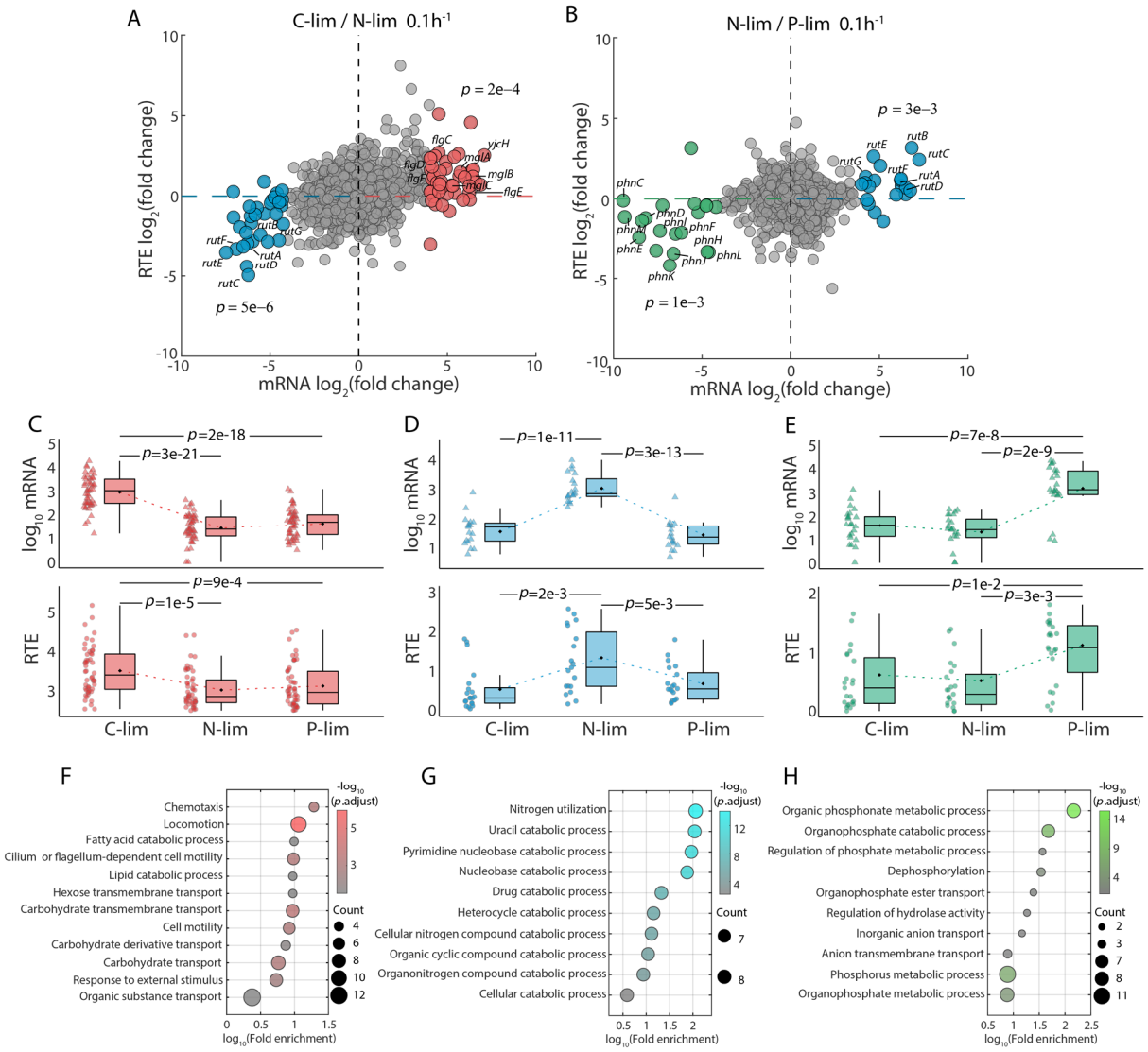

Figure 4. Transcription and translation couple together to respond to nutrient limitations.

(A) Comparison of transcription changes (log_2 mRNA fold change, *x*-axis) and translation changes (log_2 RTE fold change, *y*-axis) between carbon limitation and nitrogen limitation at the growth rate of 0.1 h^{-1}. The averages of three biological replicates are shown. Red dots represent genes with log_2 mRNA fold change (C-limited / N-limited) > 4. Blue dots represent genes with log_2 mRNA fold change (C-limited / N-limited) < -4. *p*-value was used to test the significance of the RTE fold change between the highlighted genes and the background genes.

(B) Same as (A), but showing the change between nitrogen limitation and phosphate limitation. Green dots represent genes with log_2 mRNA fold change (N-limited / P-limited) < -4.

(C-E) mRNA level (upper panel) and RTE level (lower panel) of the three groups of highlighted genes in (A) and (B). The three highlighted groups of genes are upregulated under carbon (C), nitrogen (D), and phosphate (E) limitations.

(F-H) Gene ontology (GO) enrichment analysis for the three highlighted gene groups in (A) and (B). The color of the dots represents the -$\log_{10}$ adjusted *p*-value, and the size represents the number of genes.

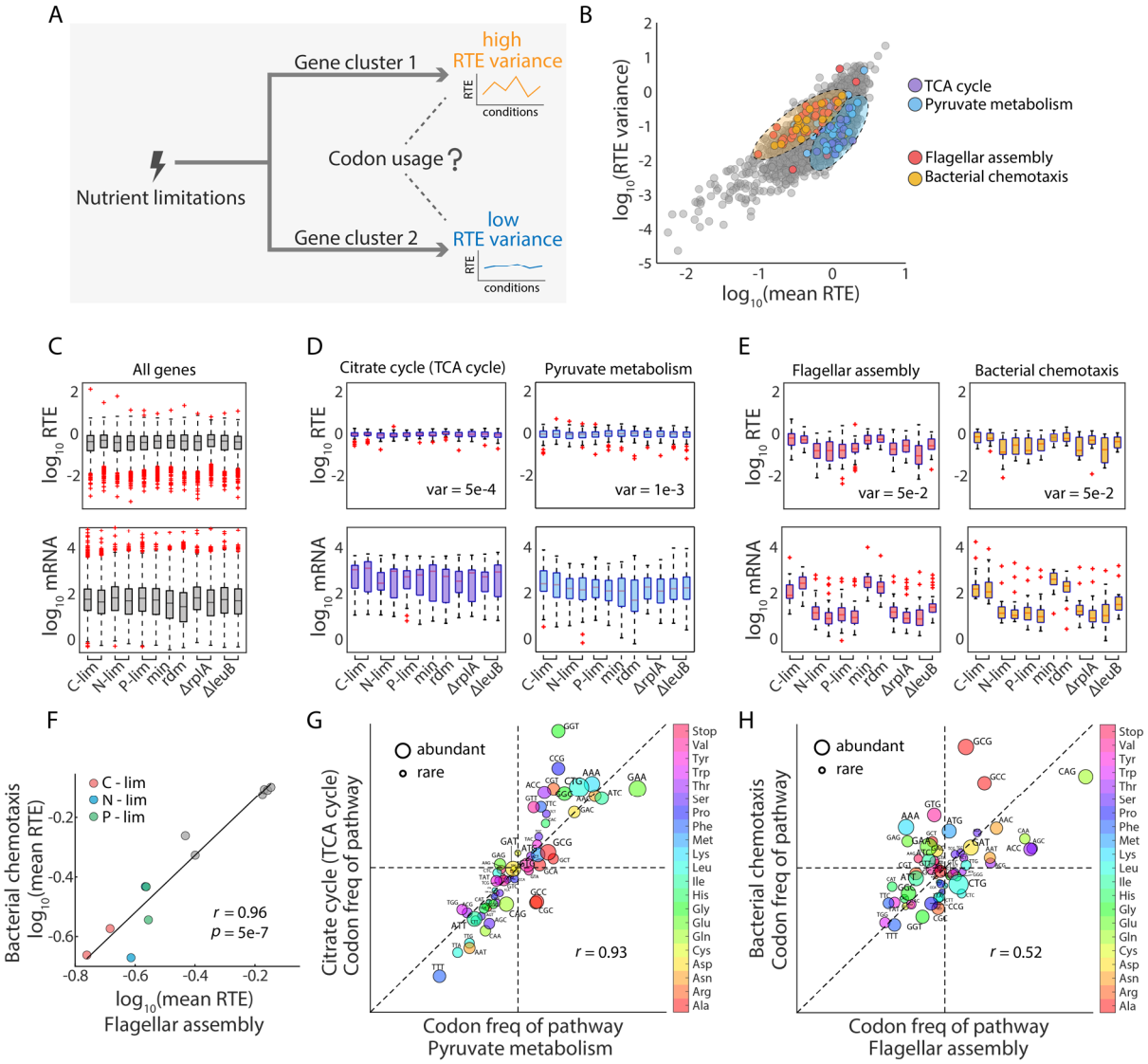

Figure 5. Pathways with similar gene expression patterns share similar codon usage bias.

(A) Illustration of the question we investigated: when genes are classified by their cross-condition variance of RTE, what is the cause of this high-or-low variance classification?

(B) Relationship between mean and variance of RTE across 12 conditions. Each dot represents one gene. Colored dots are genes involved in four selected pathways with different patterns of translational regulation.

(C) Distribution of RTEs (upper) and mRNA levels (lower) of all genes, under the 12 different conditions.

(D) Distribution of RTEs (upper) and mRNA levels (lower) for genes involved in TCA cycle (left panel) and those involved in pyruvate metabolism (right panel).

(E) Distribution of RTEs (upper) and mRNA levels (lower) for genes involved in flagellar assembly (left panel) and those involved in chemotaxis (right panel).

(F) Correlation of mean RTE across 12 conditions between genes involved in flagellar assembly and bacterial chemotaxis.

(G) Correlation of codon frequencies between genes in the two pathways described in (B). After removing the overlapped genes, there are 38 and 13 genes involved in each pathway, respectively. The 64 codons are dotted with sizes representing the rarity of codons in *E. coli*.

(H) Correlation of codon frequencies between genes in the two pathways described in (C). After removing the overlapped genes, there are 38 and 15 genes involved in them, respectively.

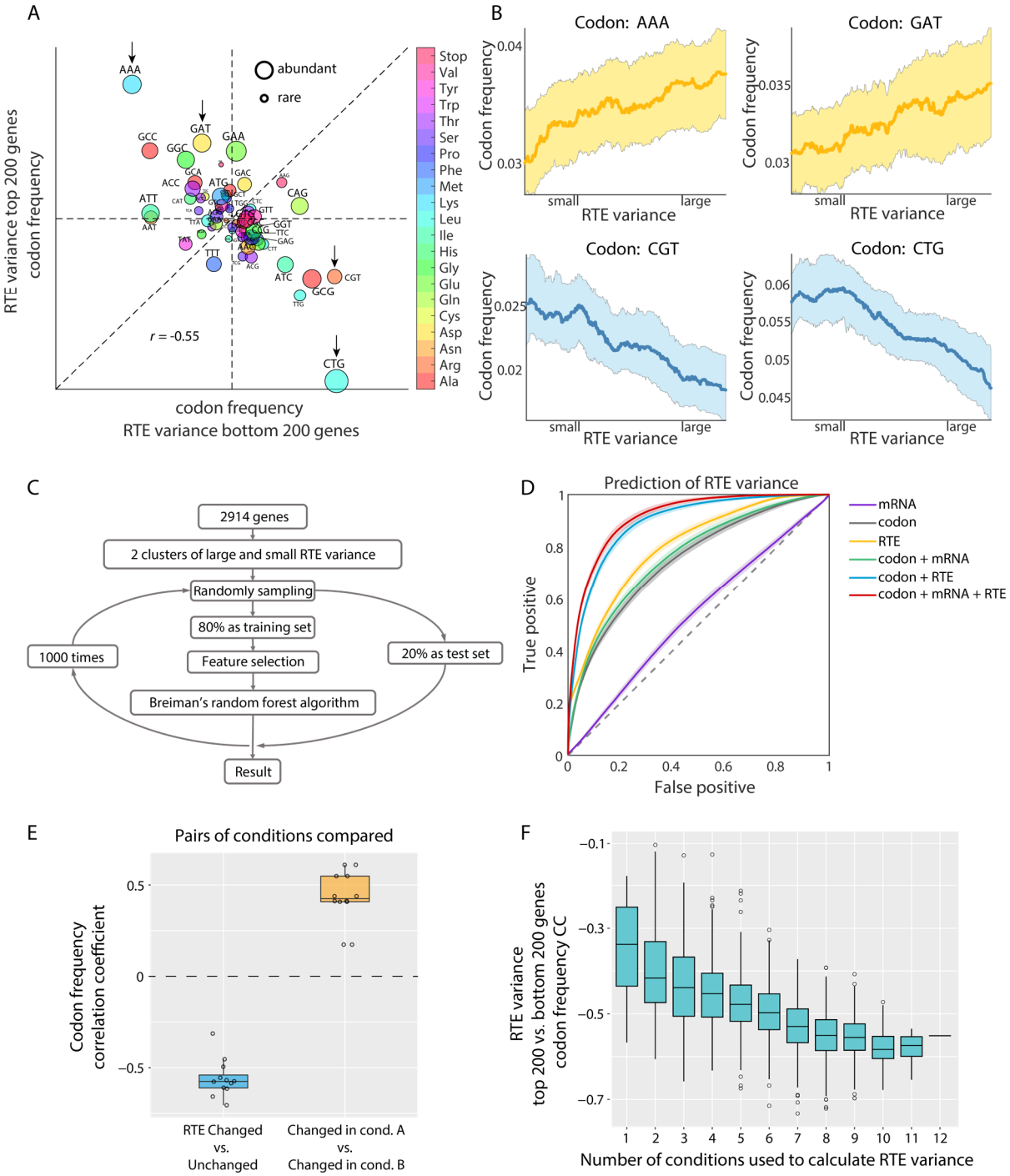

Figure 6. Codon usage contributes to RTE variability across nutrient conditions.

(A) Negative correlation between the codon frequencies for the top 200 and bottom 200 genes in their RTE cross-condition variances. Four anti-correlated codons are indicated by arrows.

(B) Relationship between the gene-by-gene RTE variance (*x*-axis) and codon frequencies (*y*-axis), for the four codons highlighted in (A). The average over 2914 genes is shown. The shadings represent the fluctuation of codon usage frequencies and the highlighted lines show smoothed mean results.

(C) Flowchart for predicting the classification of RTE cross-condition variance using a random forest model.

(D) The ROC curves of the classification accuracy using different combinations of features. An average of results for 1000 trainings was used. The shaded areas represent the S.D. Codon, mRNA, and RTE stand for the codon frequency, the mRNA level, and the RTE absolute value, respectively.

(E) Codon frequency correlations (evaluated as in A) between different gene clusters when pairs of conditions were compared (Sup. Fig. 10). The yellow box shows correlations between two clusters of RTE up-regulated genes in each of the paired conditions. The blue box shows correlations between RTE up-regulated and unchanged genes in each of the paired conditions.

(F) The correlation coefficient exactly as obtained in (A), but for different numbers of conditions used to calculate the RTE variance. Each box was derived from all possibilities of taking n from the 12 conditions.