Journal of Molecular Evolution (2024) 92:834–841
https://doi.org/10.1007/s00239-024-10211-x

ORIGINAL ARTICLE

Stress-Induced Constraint on Expression Noise of Essential Genes in *E. coli*

Perry A. LaBoone[1] · **Raquel Assis**[1,2]

Received: 18 April 2024 / Accepted: 19 September 2024 / Published online: 11 October 2024
© The Author(s) 2024

Abstract

Gene expression is an inherently noisy process that is constrained by natural selection. Yet the condition dependence of constraint on expression noise remains unclear. Here, we address this problem by studying constraint on expression noise of *E. coli* genes in eight diverse growth conditions. In particular, we use variation in expression noise as an analog for constraint, examining its relationships to expression level and to the number of regulatory inputs from transcription factors across and within conditions. We show that variation in expression noise is negatively associated with expression level, implicating constraint to minimize expression noise of highly expressed genes. However, this relationship is condition dependent, with the strongest constraint observed when *E. coli* are grown in the presence of glycerol or ciprofloxacin, which result in carbon or antibiotic stress, respectively. In contrast, we do not observe evidence of constraint on expression noise of highly regulated genes, suggesting that highly expressed and highly regulated genes represent distinct classes of genes. Indeed, we find that essential genes are often highly expressed but not highly regulated, with elevated expression noise in glycerol and ciprofloxacin conditions. Thus, our findings support the hypothesis that selective constraint on expression noise is condition dependent in *E. coli*, illustrating how it may play a critical role in ensuring expression stability of essential genes in unstable environments.

Keywords Gene expression · Expression noise · Essential gene · Evolution · Natural selection · *E. coli*

Introduction

Gene expression is the result of a series of interactions among regulatory molecules, including transcription factors (TFs). Because TFs have limited intracellular availabilities, they are subject to the stochasticity of diffusion and binding (van Zon et al. 2006). Consequently, gene expression is a noisy process driven by a combination of extrinsic and intrinsic factors. Extrinsic noise affects all genes in the same way and depends on the characteristics of a cell, such as its size, position in the cell cycle, and concentrations of various TFs (Elowitz et al. 2002; Barroso et al. 2018; Thomas 2019). Intrinsic noise varies across genes and depends on the characteristics of a gene, such as its genomic position, regulatory sequences, and stability of transcribed mRNAs. (Elowitz et al. 2002; Barroso et al. 2018; Thomas 2019). Both types of noise play key roles in the overall cell-to-cell variation in gene expression observed in isogenic populations living in homogeneous environments (Elowitz et al. 2002; Hodgins-Davis et al. 2015).

Gene expression noise can be beneficial in some scenarios, such as when the environment is in flux (Thattai and van Oudenaarden 2001; Acar et al. 2008; Beaumont et al. 2009; Liu et al. 2015; Bódi et al. 2017; Duveau et al. 2018; Payne and Wagner 2019; Schmutzer and Wagner 2020; Urchueguía et al. 2021). Indeed, studies in the unicellular organisms *E. coli* and *S. cerevisiae* show that noisy gene expression in genetically identical cells produces heterogeneous phenotypes conferring selective advantages in stressful or fluctuating environments (Thattai and van Oudenaarden 2001; Acar et al. 2008; Liu et al. 2015; Wolf et al. 2015; Duveau et al. 2018; Schmutzer

Handling Editor: **KerryGeiler-Samerotte**

✉ Raquel Assis
rassis@fau.edu

[1] Department of Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL 33431, USA

[2] Institute for Human Health and Disease Intervention, Florida Atlantic University, Boca Raton, FL 33431, USA

and Wagner 2020; Urchueguía et al. 2021). Further demonstrating this effect, *S. cerevisiae* populations with high phenotypic heterogeneity evolve greater antifungal resistance and are more robust to extinction than those with low phenotypic heterogeneity (Bódi et al. 2017; Payne and Wagner 2019). However, expression noise is typically deleterious (Barkai and Leibler 2000; Fraser et al. 2004; Wang and Zhang 2011; Schmiedel et al. 2019), and stabilizing selection maintains expression levels in most scenarios (Gilad et al. 2006). The fitness effects of expression noise vary across genes, with gene-specific sensitivity to noise hypothesized to depend on an evolutionary trade-off between energy efficiency and noise reduction (Thattai and van Oudenaarden 2001; Hausser et al. 2019). Regardless, multiple theoretical and empirical studies have uncovered support for widespread negative selection to minimize gene expression noise (Fraser et al. 2004; Lehner 2008; Wang and Zhang 2011), suggesting that it is an important trait that should be considered in evolutionary studies.

As for other biologically important traits, the fitness effects of expression noise are expected to vary across environments (Raser and O'Shea 2005; Wang and Zhang 2011; Duveau et al. 2018; Schmiedel et al. 2019). Yet because most genome-wide studies have only assayed expression noise in a single condition, noise response to different environments is not clearly understood. A recent study addressed this question by quantifying and comparing expression noise in 1,103 *E. coli* genes across eight growth conditions (Urchueguía et al. 2021). Their findings demonstrate that gene expression noise indeed varies across conditions and, moreover, that the condition dependence of genome-wide noise levels is primarily determined by the structure of the gene regulatory network (Urchueguía et al. 2021). However, the condition-dependent role of selection on gene expression noise remains unclear.

To tackle this problem, we used the dataset of Urchueguía et al. (2021) to examine and compare the impacts of expression level, number of regulatory inputs from TFs, and essentiality on variation in expression noise across and within eight growth conditions in *E. coli*. Though several studies have interrogated relationships between each of these traits and mean expression noise (Fraser et al. 2004; Sánchez and Kondev 2008; Wang and Zhang 2011; Silander et al. 2012; Sharon et al. 2014; Wolf et al. 2015; Urchueguía et al. 2021), knowledge of their contributions to variation in expression noise across conditions can shed light on the role of selection in moderating expression noise. In particular, pinpointing differences among these conditions can elucidate how selection may act to minimize expression noise variation in diverse environments, shaping our understanding of the plasticity of this critical biological trait.

Results and Discussion

As a first step toward understanding condition-dependent selection on expression noise, we examined how expression noise changes as a function of expression level across and within the eight growth conditions in *E. coli* (Fig. 1; see *Methods*). In particular, because our expression noise estimates account for the natural dependency between expression level and noise due to measurement error (see *Methods*) (Urchueguía et al. 2021), this analysis enabled us to assay the relationship between expression level and the remaining "biological" noise. Consistent with a prior genome-wide analysis in *E. coli* (Silander et al. 2012), we observed a weak positive nonlinear correlation between expression level and noise across conditions (Fig. 1A; see *Methods*). However, a White test (White 1980) also uncovered strong evidence of non-constant variance, or heteroskedasticity (James et al. 2021), in expression noise as a function of expression level (Fig. 1A; see *Methods*). In particular, despite its small increase in magnitude, expression noise becomes much less variable as expression level increases. Heteroskedasticity is an important statistical property, as it indicates that the variance of one variable is dependent on the value of another, violating a common assumption of regression analysis (James et al. 2021). In this case, heteroskedasticity signals a negative relationship between expression level and variation in expression noise, suggesting that noise is more tightly controlled in highly expressed genes. As highly expressed genes tend to evolve slowly at the sequence (Pál et al. 2001; Krylov et al. 2003; Subramanian and Kumar 2004; Pál et al. 2006; Drummond and Wilke 2008; Marek and Tomala 2018; Shibai et al. 2022) and expression (Lemos et al. 2005; Liao and Zhang 2006; Gu et al. 2019) levels, it is not surprising that their expression noise is also constrained. Stronger constraint on such genes may be due in part to their ubiquity and therefore importance in many cellular processes (Krylov et al. 2003), particularly as expression noise is often deleterious (Barkai and Leibler 2000; Fraser et al. 2004; Wang and Zhang 2011; Schmiedel et al. 2019).

To glean insight into which conditions drive the observed relationships of expression level with expression noise and with expression noise variation, we explored associations between these variables in individual conditions (Fig. 1B). Our analysis revealed that the relationships of expression level with expression noise and with expression noise variation are both condition dependent (Fig. 1B). Specifically, expression level and noise are weakly positively correlated in five of the eight conditions: stationary phase 30 h, glucose, glycerol, lactose, and ciprofloxacin. Comparisons of the magnitudes of correlation coefficients suggest that glycerol and ciprofloxacin

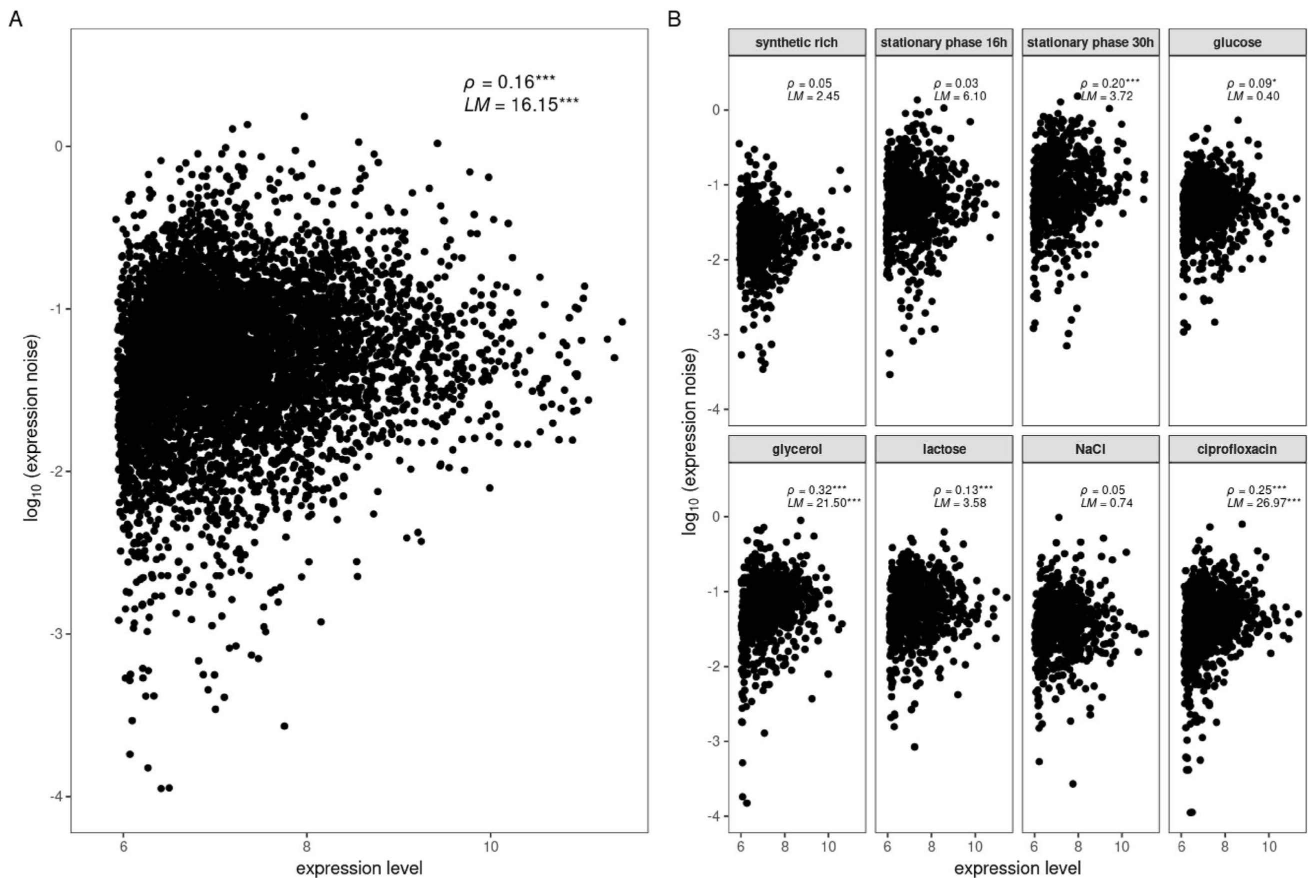

Fig. 1 Relationships of expression level with expression noise across (**A**) and within (**B**) eight conditions in *E. coli*. Spearman correlation coefficients (ρ) and White test statistics (*LM*) are shown in the upper right corner of each plot. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, *NS* not significant (see *Methods*)

are the two major contributors to the correlation observed across conditions (Fig. 1A), as they are not significantly different from each other but significantly larger than those of all other conditions aside from stationary phase 30 h (Table S1; see *Methods*). Intriguingly, heteroskedasticity is only evident for these two conditions (Fig. 1B), which have White test statistics that are not significantly different from each other but significantly larger than those of all but one condition for glycerol and all other conditions for ciprofloxacin (Table S2; see *Methods*). Thus, glycerol and ciprofloxacin appear to be the primary or sole drivers of the heteroskedasticity observed when considering all conditions together (Fig. 1A) and therefore, the stronger constraint on expression noise of highly expressed genes. Indeed, both substances perturb the cell—glycerol is an inefficient nutrient source that triggers a carbon stress response (Martínez-Gómez et al. 2012), and ciprofloxacin is an antibiotic that causes irreversible cellular damage and death (Smirnova et al. 2017; Adamus-Białek et al. 2019). Hence, perhaps growth of *E. coli* in these conditions initiates stress-related pathways that naturally increase fluctuations in expression noise, which must then be tightly controlled to ensure stable expression of highly expressed and presumably biologically important genes.

Because TFs are thought to be key drivers of expression noise (van Zon et al. 2006) and contribute to condition-dependent expression noise in *E. coli* (Urchueguía et al. 2021), we hypothesized that TFs may also play a role in the observed condition-dependent constraint on expression noise. To evaluate this hypothesis, we next examined relationships of the number of regulatory inputs from TFs with expression level and with expression noise across and within the eight growth conditions in *E. coli* (Fig. 2; see *Methods*). Across conditions, we observed weak-positive nonlinear correlations for the number of regulatory inputs with expression level (Fig. 2A) and with expression noise (Fig. 2B), suggesting that both expression level and noise only increase slightly as the number of regulatory inputs increases. There is also support for heteroskedasticity in the relationship of the number of regulatory inputs with expression level (Fig. 2A), but not with expression noise (Fig. 2B). These relationships are consistent with constraint

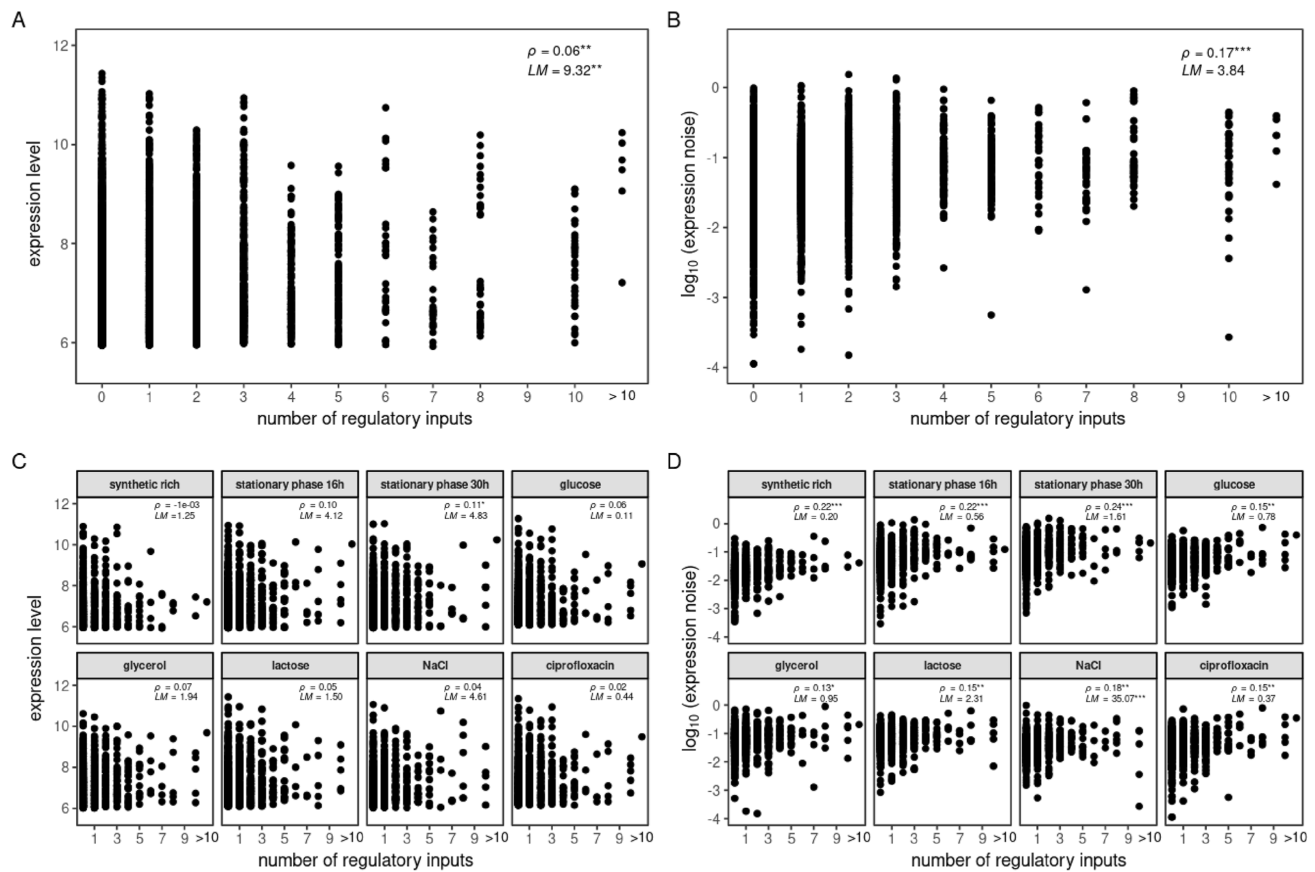

Fig. 2 Relationships of the number of regulatory inputs from TFs with expression level (left) and expression noise (right) across (**A**, **B**) and within (**C**, **D**) eight conditions in *E. coli*. Spearman correlation coefficients (ρ) and White test statistics (*LM*) are shown in the upper right corner of each plot. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, *NS* not significant (see *Methods*)

on expression level but not on expression noise of highly regulated genes. The observed constraint on expression noise of highly expressed, but not highly regulated, genes suggests that genes that are highly expressed tend not to be highly regulated. Hence, perhaps highly expressed genes and highly regulated genes compose two distinct classes of genes whose noise is constrained in different ways.

We next examined the condition dependence for observed relationships of the number of regulatory inputs with expression level and with expression noise (Fig. 2A and B). We found that conditions manifesting changes in expression level and noise as a function of the number of regulatory inputs (Fig. 2C and D) differ from those identified for the relationship between expression level and noise (Fig. 1B). In particular, there is a weak positive correlation between the number of regulatory inputs and expression level in only one condition (stationary phase 30 h; Fig. 2C), whereas there are weak-positive correlations between the number of regulatory inputs and expression noise in all conditions (Fig. 2D). Additionally, there are no significant differences between correlation coefficients in any pair of conditions for either expression level (Table S3) or noise (Table S4), suggesting that there is likely no condition dependence for the correlation between the number of regulatory inputs and either trait. When considering heteroskedasticity, we obtained no statistical support for an association between the number of regulatory inputs and variation in expression level in any of the conditions (Fig. 2C). Although the NaCl condition demonstrates heteroskedasticity in the relationship between the number of regulatory inputs and expression noise (Fig. 2D), there are again no significant differences between White test statistics in any pair of conditions (Table S5). Thus, unlike highly expressed genes, highly regulated genes do not appear to experience condition-dependent noise constraint, supporting the hypothesis that they represent a different class of genes with unique evolutionary constraints.

Last, inspired by studies demonstrating that essential genes often display high expression levels and low noise (Fraser et al. 2004; Bhardwaj and Lu 2005; Wang and Zhang 2011; Silander et al. 2012; Wang et al. 2015; Wu et al. 2017), we considered associations of gene essentiality with expression level and with noise across and within the eight conditions in *E. coli* (Fig. 3). Across conditions, our findings mirror prior studies (Fraser et al. 2004; Bhardwaj and

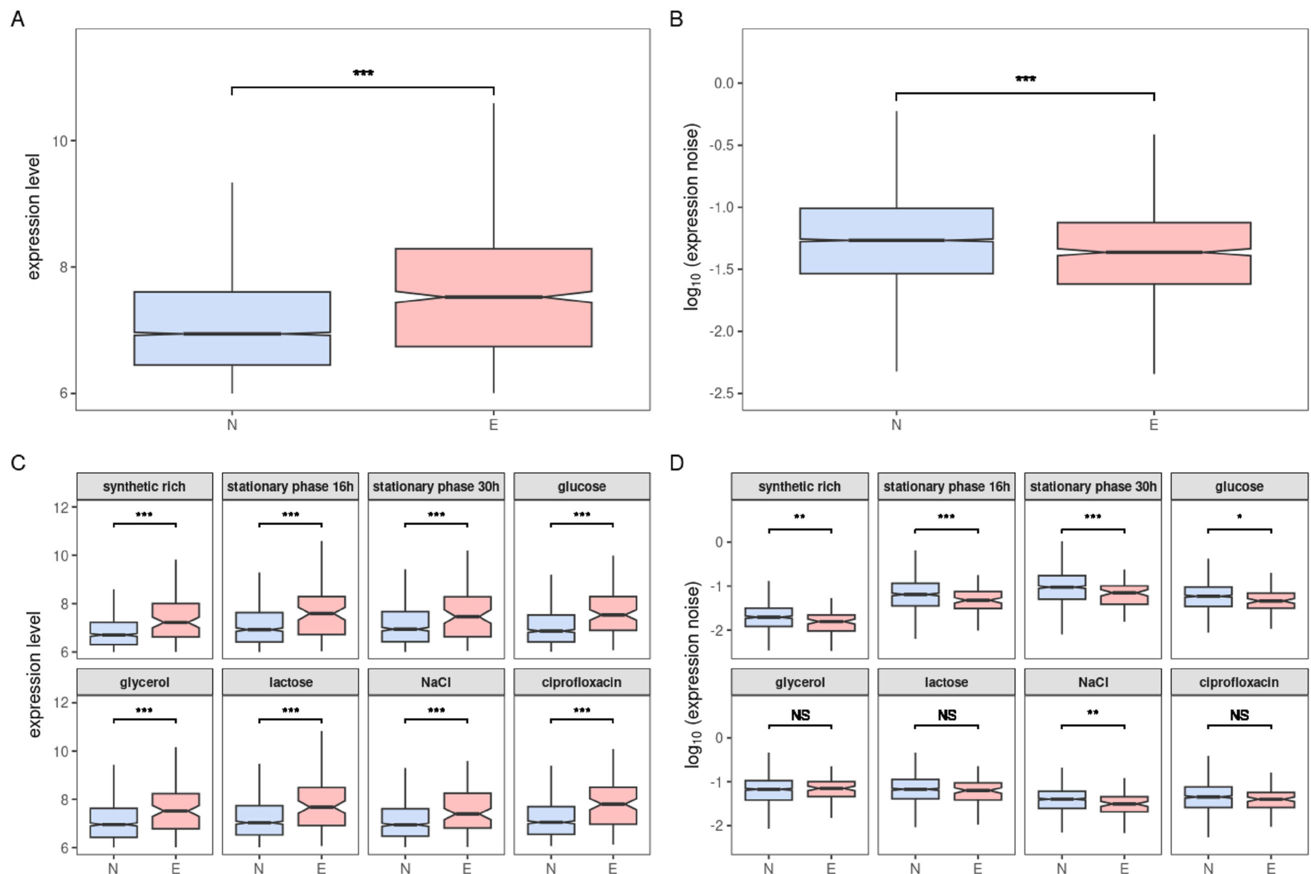

Fig. 3 Distributions of expression level (left) and expression noise (right) for nonessential ("N") and essential ("E") genes across (**A**, **B**) and within (**C**, **D**) eight conditions in *E. coli*. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, *NS* not significant (see *Methods*)

Lu 2005; Wang and Zhang 2011; Silander et al. 2012; Wang et al. 2015), showing that essential genes tend to be more highly expressed and less noisy than nonessential genes in *E. coli* (Fig. 3A and B). Consistent with a genome-wide study in yeast (Yu et al. 2004), essential *E. coli* genes also tend to have fewer regulatory inputs from TFs than nonessential genes (Fig. S1), indicating that they are not highly regulated despite being highly expressed and therefore providing further support for the hypothesis that highly expressed and highly regulated genes form distinct classes. For expression level, the association between essentiality and expression level appears to be condition independent, as essential genes are more highly expressed in all eight conditions (Fig. 3C). In contrast, the association between essentiality and expression noise is condition dependent, with no difference in expression noise between essential and nonessential genes in glycerol, lactose, and ciprofloxacin conditions (Fig. 3D). This result is interesting, as despite the relationship between essentiality and expression level across and within conditions (Fig. 3A and C), significant associations between essentiality and expression noise (Fig. 3D) exist only for conditions with no evidence of heteroskedasticity between expression level and noise (Fig. 1B). That is, expression noise of essential genes is only lower in conditions where it is not constrained as a function of expression level. This result may point to an important role of gene essentiality in expression noise constraint, such that noise is only constrained in conditions where essential genes tend to be noisier. Intriguingly, the pertinent conditions in our study are glycerol and ciprofloxacin, which also demonstrate the strongest associations between expression level and noise. Taken together, our results suggest that these conditions generate extreme perturbations of the cellular environment that increase expression noise of essential genes, perhaps necessitating selection to constrain noise and minimize its deleterious effects on critical biological processes.

Methods

Data Acquisition and Processing

We utilized the Urchueguía et al. (2021) estimations of expression level and noise for 1103 genes in *E. coli* grown

in eight conditions: synthetic-rich media with 0.2% glucose, M9 minimal media with 0.2% glucose for 16 h, M9 minimal media with 0.2% glucose for 30 h, M9 minimal media with 0.2% glucose, M9 minimal media with 0.2% glycerol, M9 minimal media with 0.2% lactose, M9 minimal media with 0.4M NaCl, and M9 minimal media with 1.5 ng/ml ciprofloxacin. For simplification, we refer to these conditions in the manuscript as synthetic rich, stationary phase 16 h, stationary phase 30 h, glucose, glycerol, lactose, NaCl, and ciprofloxacin, respectively.

Urchueguía et al. (2021) used flow cytometry to measure the mean and variance of expression levels (log-fluorescence). To estimate expression noise, they decomposed each raw measurement of expression variance into two terms: a "Poissonian" term that decreases with expression level and appears to be primarily driven by measurement noise of the flow cytometer (Galbusera et al. 2020) and a "noise floor" term that is independent of expression level. They then extracted the noise floor component of expression variance and computed expression noise as the difference between the variance and the noise floor. As a result, these expression noise estimates do not depend on expression levels and are also likely to represent true "biological" noise because they do not incorporate variation due to measurement error. Additionally, similar to other methods (Barroso et al. 2018; Laloum and Robinson-Rechavi 2021), these estimates take into account mean expression level and neutralize the noise floor as a function of mean expression, which is important when comparing expression noise across conditions.

We also obtained numbers of regulatory inputs from TFs for each of the 1103 *E. coli* genes from Urchueguía et al. (2021). They extracted all gene-TF regulation annotations from the RegulonDB database for *E. coli* (Santos-Zavaleta et al. 2019) and then counted the number of unique TFs known to regulate each gene. These values ranged from 0 (no known regulatory inputs) to 14 for the genes in our study. Of the 1103 genes in our study, 644 do not have any known regulatory inputs and were therefore excluded from analyses utilizing the number of regulatory inputs. Additionally, because only one gene (*gadX*) has more than ten known regulatory inputs, we assigned it to the group "> 10" for easier visualization in Fig. 2. Note that the actual number of regulatory inputs (14) was used for statistical analyses.

Lists of essential genes were sourced from Dasmeh et al. (2017) and Goodall et al. (2018). Those from Goodall et al. (2018) were determined through transposon-directed insertion site sequencing (TraDIS), which combines transposon mutagenesis with short-fragment DNA sequencing of transposon junctions (Gawronski et al. 2009; Goodman et al. 2009; Langridge et al. 2009; van Opijnen et al. 2009). Because genes can be conditionally essential, they defined a gene as essential only if the transposon insertion data showed that at least a portion of the protein-coding sequence of the gene is required for growth in all four tested conditions (Goodall et al. 2018). To decrease false positives, they also implemented a statistical model to correct for both gene length and genome length when predicting essentiality (Goodall et al. 2018). To maximize our sample size, we considered a gene essential if it was defined as essential by at least one of the two studies (Dasmeh et al. 2017; Goodall et al. 2018). However, it is important to note that essentiality status may be dynamic, and it is therefore possible for a gene to be essential in some conditions and nonessential in others.

Statistical Analyses

All statistical analyses were performed in R (R Core Team 2022) with the Posit Cloud IDE (RStudio Team 2024). Before performing any analyses discussed in the manuscript, we evaluated linearity in the relationships of expression level with expression noise (Fig. 1) and the number of regulatory inputs with expression level and expression noise (Fig. 2). Specifically, we first used the lm() function in the stats package (R Core Team 2022) to fit linear regression models to the data, and then we applied the shapiro.test() function in the stats package (R Core Team 2022) to evaluate normality of the residuals (errors) of the fitted models with Shapiro–Wilk tests (Shapiro and Wilk 1965). Because the null hypothesis of normality was rejected in nearly all cases (Figs. S2 and S3), we conservatively chose to employ statistical tests that did not assume linearity for all analyses in our study.

We used the cor.test() function in the stats package (R Core Team 2022) to estimate (nonlinear) Spearman correlation coefficients (ρ) (Spearman 1907) and evaluate their statistical significance for the relationships of expression noise with expression level (Fig. 1) and the number of regulatory inputs with expression level and expression noise (Fig. 2). Because the commonly used Breusch–Pagan test (Breusch and Pagan 1979) can only detect linear forms of heteroskedasticity, we instead used its nonlinear equivalent, the White test (White 1980), to evaluate heteroskedasticity in these relationships. In particular, we performed bootstrapped White tests with the white_test() function in the whitestrap package (Jeong and Lee 1999). The test statistic for the White test is the Lagrange multiplier (*LM*), which follows a chi-squared distribution. Two-tailed Mann–Whitney *U* tests (Mann and Whitney 1947), implemented with the wilcox.test() function In the stats package (R Core Team 2022), were used to evaluate differences between expression level and expression noise distributions of nonessential and essential genes (Fig. 3), as well as differences between the number of regulatory inputs from TFs of nonessential and essential genes (Fig. S1).

Two-tailed permutation tests were used to evaluate differences between statistically significant Spearman correlation and White tests depicted in Figs. 1B, 2C, and 2D and

those for all other conditions. Specifically, we used 10,000 permutations to compare each pair of conditions, setting the test statistic as the mean difference between either computed values of ρ (for Spearman correlation tests) or LM (for White tests). Each permutation P value was Bonferroni-corrected for the seven comparisons performed with the p.adjust() function in the stats package (R Core Team 2022).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00239-024-10211-x.

Acknowledgements This work was supported by the National Institutes of Health Grant R35GM142438 and the National Science Foundation Grant DBI-2130666.

Data Availability All R code and processed datasets are available at: https://github.com/PerryFAU/Expresson_Noise.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

References

Acar M, Mettetal JT, van Oudenaarden A (2008) Stochastic switching as a survival strategy in fluctuating environments. Nat Genet 40:471–475

Adamus-Białek W, Wawszczak M, Arabski M, Majchrzak M, Gulba M, Jarych D, Parniewski P, Głuszek S (2019) Ciprofloxacin, amoxicillin, and aminoglycosides stimulate genetic and phenotypic changes in uropathogenic *Escherichia coli* strains. Virulence 10:260–276

Barkai N, Leibler S (2000) Circadian clocks limited by noise. Nature 403:267–268

Barroso GV, Puzovic N, Dutheil JY (2018) The evolution of gene-specific transcriptional noise is driven by selection at the pathway level. Genetics 208:173–189

Beaumont HJE, Gallie J, Kost C, Ferguson GC, Rainey PB (2009) Experimental evolution of bet hedging. Nature 462:90–93

Bhardwaj N, Hui L (2005) Correlation between gene expression profiles and protein–protein interactions within and across genomes. Bioinformatics 21:2730–2738

Breusch TS, Pagan AR (1979) A simple test for heteroscedasticity and random coefficient variation. Econometrica 47:1287–1294

Bódi Z, Farkas Z, Nevozhay D, Kalapis D, Lázár V, Csörgő B, Nyerges Á, Szamecz B, Fekete G, Papp B, Araújo H, Oliveira J, Moura G, Santos M, Székely T, Balázsi G (2017) Phenotypic heterogeneity promotes adaptive evolution. PLoS Biol 15:e2000644

Dasmeh P, Girard É, Serohijos AWR (2017) Highly expressed genes evolve under strong epistasis from a proteome-wide scan in *E. coli*. Sci Rep 7:15844

Drummond DA, Wilke CO (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell 134:341–352

Duveau F, Hodgins-Davis A, Metzger BPH, Yang Bing, Tryban S, Walker EA, Lybrook T, Wittkopp PJ (2018) Fitness effects of altering gene expression noise in Saccharomyces cerevisiae. eLife 7:e37272

Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. Science (New York) 297:1183–1186

Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB (2004) Noise minimization in eukaryotic gene expression. PLoS Biol 2:e137

Galbusera L, Bellement-Theroue G, Urchueguia A, Julou T, van Nimwegen E (2020) Using fluorescence flow cytometry data for single-cell gene expression analysis in bacteria. PLoS ONE 15:e0240233

Gawronski JD, Wong SMS, Giannoukos G, Ward DV, Akerley B (2009) Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. Proc Nat Acad SciD 106:16422–16427

Gilad Y, Oshlack A, Rifkin S (2006) Natural selection on gene expression. Trends Genet. 22:456–461

Goodall ECA, Robinson A, Johnston IG, Jabbari S, Turner KA, Cunningham AF, Lund PA, Cole JA, Henderson IR (2018) The essential genome of *Escherichia coli* K-12. mBio 9:89. https://doi.org/10.1128/mbio.02096-17

Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone Catherine A, Knight Rob, Gordon Jeffrey I (2009) Identifying genetic determinants needed to establish a human gut symbiont in its habitat. Cell Host & Microbe 6:279–289

Xun G, Ruan H, Yang J (2019) Estimating the strength of expression conservation from high throughput RNA-seq data. Bioinformatics 35:5030–5038

Hausser J, Mayo A, Keren L, Alon U (2019) Central dogma rates and the trade-off between precision and economy in gene expression. Nat Commun 10:68

Hodgins-Davis A, Rice DP, Townsend JP (2015) Gene expression evolves under a house-of-cards model of stabilizing selection. Mol Biol Evolut 7:32

James G, Witten D, Hastie T, Tibshirani R (2021) An introduction to statistical learning: with applications in R. Springer, New York

Jeong J, Lee K (1999) Bootstrapped white's test for heteroskedasticity in regression models. Econ Lett 63:261–267

Krylov DM, Wolf YI, Rogozin IB, Koonin EV (2003) Gene Loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. Genome Res 13:2229–2235

Laloum D, Robinson-Rechavi M (2021) Two levels of selection of rhythmicity in gene expression: energy saving for rhythmic proteins and noise optimization for rhythmic transcripts. bioRxiv 04:20

Langridge GC, Phan M-D, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, Wain J, Parkhill J, Turner AK (2009) Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. Genome Res 19:2308–2316

Lehner B (2008) Selection to minimise noise in living systems and its implications for the evolution of gene expression. Mol Syst Biol 4:170

Lemos B, Bettencourt BR, Meiklejohn CD, Hartl DL (2005) Evolution of proteins and gene expression levels are coupled in drosophila and are independently associated with MRNA abundance, protein length, and number of protein–protein interactions. Mol Biol Evol 22:1345–1354

Liao B-Y, Zhang J (2006) Low rates of expression profile divergence in highly expressed genes and tissue-specific genes during mammalian evolution. Mol Biol Evol 23(6):1119–1128

Liu J, Martin-Yken H, Bigey F, Dequin S, François J-M, Capp J-P (2015) Natural yeast promoter variants reveal epistasis in the generation of transcriptional-mediated noise and its potential benefit in stressful conditions. Genome Biol Evolut 7:969–984

Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat 18:50–60

Marek A, Tomala K (2018) The contribution of purifying selection, linkage, and mutation bias to the negative correlation between gene expression and polymorphism density in yeast populations. Genome Biol Evol 10:2986–2996

Martínez-Gómez K, Flores N, Castañeda HM, Martínez-Batallar G, Hernández-Chávez G, Ramírez OT, Gosset G, Encarnación S, Bolivar F (2012) New insights into *Escherichia coli* metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol. Microb Cell Fact 11:46

Payne JL, Wagner A (2019) The causes of evolvability and their evolution. Nat Rev Genets 20:24–38

Pál C, Papp B, Hurst LD (2001) Highly expressed genes in yeast evolve slowly. Genetics 158:927–931

Pál C, Papp B, Lercher MJ (2006) An integrated view of protein evolution. Nat Rev Genet 7:337–348

R Core Team (2022) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Raser JM, O'Shea EK (2005) Noise in gene expression: origins, consequences, and control. Science (New York, N.Y.) 309:2010–2013

RStudio Team (2024) RStudio: Integrated Development Environment for R

Santos-Zavaleta A, Salgado H, Gama-Castro S, Sánchez-Pérez M, Gómez-Romero L, Ledezma-Tejeida D, García-Sotelo JS, César MC, Carlos-Francisco GJ, Collado-Vides J (2019) RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. Nucleic Acids Res 47:D212–D220

Schmiedel JM, Carey LB, Lehner B (2019) Empirical mean-noise fitness landscapes reveal the fitness impact of gene expression noise. Nat Commun 10:3180

Schmutzer M, Wagner A (2020) Gene expression noise can promote the fixation of beneficial mutations in fluctuating environments. PLOS Comput Biol 16:e1007727

Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). Biometrika 52:591–611

Sharon E, van Dijk D, Kalma Y, Keren L, Manor O, Yakhini Z, Segal E (2014) Probing the effect of promoters on noise in gene expression using thousands of designed sequences. Genome Res 113:168773

Shibai A, Kotani H, Sakata N, Furusawa C, Tsuru S (2022) Purifying selection enduringly acts on the sequence evolution of highly expressed proteins in *Escherichia coli*. G3 12:235

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

Smirnova GV, Tyulenev AlV, Muzyka NG, Peters MA, Oktyabrsky ON (2017) Ciprofloxacin provokes SOS-dependent changes in respiration and membrane potential and causes alterations in the redox status of *Escherichia coli*. Res Microbiol 168:64–73

Spearman C (1907) Demonstration of formulæ for true measurement of correlation. Am J Psychol 18:161–169

Subramanian S, Kumar S (2004) Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. Genetics 168:373–381

Sánchez Á, Kondev J (2008) Transcriptional control of noise in gene expression. Proc Nat Acad Sci 105:5081–5086

Thattai M, van Oudenaarden A (2001) Intrinsic noise in gene regulatory networks. Proc Nat Acad Sci 98:8614–8619

Thomas P (2019) Intrinsic and extrinsic noise of gene expression in lineage trees. Sci Rep 9:474

Urchueguía A, Galbusera L, Chauvin D, Bellement G, Julou T, van Nimwegen E (2021) Genome-wide gene expression noise in *Escherichia coli* is condition-dependent and determined by propagation of noise through the regulatory network. PLoS Biol 19:e3001491

van Opijnen T, Bodi KL, Camilli A (2009) Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat Methods 6:767–772

van Zon J, Morelli M, Tănase-Nicola S, Wolde PR (2006) Diffusion of transcription factors can drastically enhance the noise in gene expression. Biophys J 91:4350–4367

Wang T, Birsoy K, Hughes NW, Krupczak KM, Post Y, Wei Jenny J, Lander Eric S, Sabatini David M (2015) Identification and characterization of essential genes in the human genome. Science (New York, N.Y.) 350:1096–1101

Wang Z, Zhang J (2011) Impact of gene expression noise on organismal fitness and the efficacy of natural selection. Proc Nat Acad Sci 108:E67–E76

White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48:817–838

Wolf L, Silander OK, van Nimwegen E (2015) Expression noise facilitates the evolution of gene regulation. eLife 4:e05856

Shaohuan W, Li K, Li Y, Zhao T, Li T, Yang Yu-Fei, Qian Wenfeng (2017) Independent regulation of gene expression level and noise by histone modifications. PLoS Comput Biol 13:e1005585

Yu H, Greenbaum D, Lu H-X, Zhu X, Gerstein M (2004) Genomic analysis of essentiality within protein networks. Trends Genet 20:227–231