# Optimal scaling between mRNA and protein maximizes cell fitness

Yichen Yan[1] and Jie Lin[1,2,3]

[1] *Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China*
[2] *Center for Quantitative Biology, Peking University, Beijing, China*
[3] *School of Physics, Peking University, Beijing, China*

(Dated: March 13, 2025)

Gene expression regulation is crucial for cell fitness, balancing the costs of mRNA and protein synthesis with the benefits of their production. Here, we identify the optimal scaling between mRNA and protein levels that maximizes cell fitness. For highly expressed genes for which protein noise is negligible, the mRNA copy number scales with the square root of the protein copy number ($m \sim p^{0.5}$), while for lowly expressed genes for which protein noise is significant, a linear relationship ($m \sim p$) emerges. Simulations of *in silico* cells using biological parameters predict that *Saccharomyces cerevisiae* is close to the square-root regime while *Escherichia coli* is close to the linear regime, in perfect agreement with experimental measurements. We show that the universal predominance of transcription regulation in determining protein abundance is a natural result of optimal fitness, in agreement with the mRNA-protein scaling across different organisms. We also establish a constraint on the ratio between the translation and transcription rate, rationalizing the absence of highly expressed mRNAs with low translational activity. Our findings uncover a universal principle of gene expression regulation.

## INTRODUCTION

A cell's capacity to thrive in a given environment hinges upon its ability to synthesize mRNAs and proteins at the appropriate levels, neither lacking essential components nor wasting resources on superfluous products [1–5]. This intricate balancing necessitates a regulatory system, the principles governing which remain incompletely understood. In this work, we seek to answer a fundamental question: what are the optimal expression levels of mRNA and protein that maximize cell fitness? On the one hand, the protein level should be above some minimum value to perform its function appropriately [6], and the mRNA level also cannot be too low; otherwise, the resulting drastic protein noise can impair cell growth [7–11]. On the other hand, the gene expression process also imposes a cost on cell growth [3, 12–18].

To answer the question, we examine the trade-off between gene expression cost and the benefit of mRNA and protein production. We present a minimal model incorporating the benefits of protein synthesis, inspired by the massive measurements of the effects of gene expression levels on fitness [19] and the costs associated with transcription and translation [3]. Our analysis reveals a previously unrecognized relationship between the mRNA copy number ($m$) and the protein copy number ($p$) that maximizes cell fitness. For highly expressed genes, maximizing fitness predicts that the mRNA copy number is proportional to the square root of protein copy number ($m \sim p^{0.5}$); in contrast, for lowly expressed genes, optimal growth requires that the mRNA copy number is proportional to the protein copy number ($m \sim p$). The transition between the two different scaling behaviors happens at a critical gene expression level at which the fitness cost due to protein deficiency equals the one due to mRNA deficiency.

These different scaling behaviors between $m$ and $p$ are validated by the excellent agreement between the predictions from *in silico* cells and experimental data of diverse organisms [9, 20–23]. Intriguingly, we find that *Saccharomyces cerevisiae* and *Escherichia coli* appear near the two limiting scenarios of our model predictions. In *S. cerevisiae*, the expression levels of most genes are above the critical levels, and our model predicts a square root scaling between $m$ and $p$, in agreement with experimental data. In *E. coli*, we find that the expression levels of most genes are below the critical levels, which means a linear scaling between $m$ and $p$, also in agreement with experimental data. We unify the different scaling behaviors between $m$ and $p$ observed across different organisms and demonstrate a continuous spectrum of gene expression strategies across diverse organisms. For *E. coli*, the dominant knob of gene expression is at the transcription level, and the mRNA copy number mainly sets the protein copy number. For *S. cerevisiae*, post-transcriptional regulation is significant and approximately contributes equally to the protein level compared with transcription regulation. Other organisms, including human, mouse, and fission yeast, are between the two limiting scenarios.

Our model reveals that the universal predominance of mRNA level in regulating protein abundance [21, 24–27] is a natural consequence of optimal growth, in agreement with the experimental data of all organisms we study (even for *S. cerevisiae*, the contribution of transcription regulation is slightly larger than that of post-transcription regulation). Our minimal model also demonstrates a lower bound on the ratio between the translation rate ($\beta_p$) and the transcription rate ($\beta_m$), rationalizing the absence of highly expressed mRNA with low translational activity. The principle of gene expres-

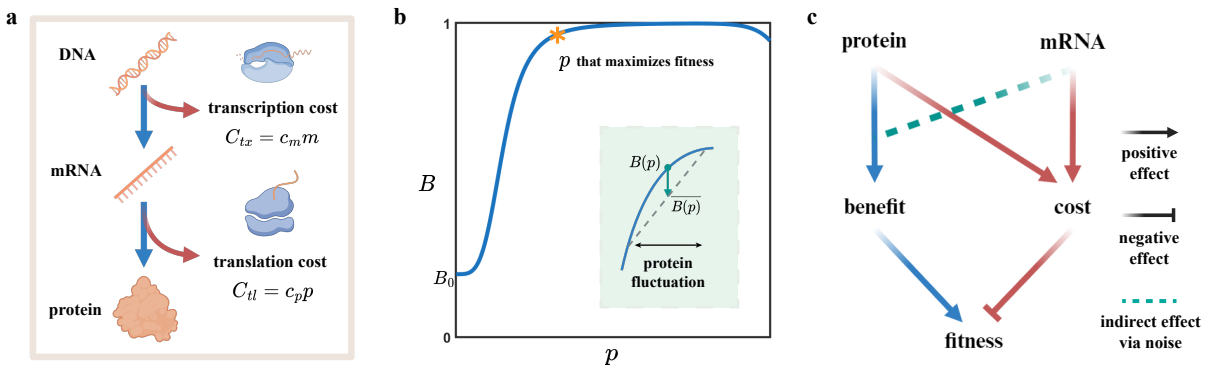


FIG. 1. **The cost and benefit of gene expression determine the fitness.** (a) The cost of gene expression. The transcription and translation processes impose a fitness cost to cell growth, proportional to mRNA and protein copy numbers. (b) A schematic benefit function increases from $B_0$ and approaches 1. The yellow asterisk marks the optimal $p$ value that maximizes the total fitness. The reduction of the benefit function at large $p$ is irrelevant to the optimization. The inset shows a reduction in the average benefit due to protein noise. (c) The causal maps of how protein and mRNA influence cell fitness.

sion regulation generated by the benefit-cost trade-off we unveil here may also guide the design of synthetic gene circuits with optimized performance.

## RESULTS

### Cost and benefit of gene expression

We consider a gene whose expression benefits cell growth but also imposes a cost. Its benefit comes from the encoded protein involved in various biological processes [6, 19, 28, 29]. Its cost comes from the gene-expression load imposed on the cell, which originates from the competition for finite resources, including ribosome, RNA polymerase, ATP, etc [1, 12, 30–36]. Experiments showed that the fitness cost imposed by gene expression depends on the produced mRNA and protein [1, 3, 37–39]. In this work, we introduce the cost $C$ as the relative growth rate reduction due to gene expression and decompose it into a transcription cost ($C_{tx}$) and a translation cost ($C_{tl}$), each proportional to the mRNA copy number ($m$) and protein copy number ($p$):

$$C(m,p) = C_{tx} + C_{tl} = c_m m + c_p p, \tag{1}$$

where $c_m$ is the cost per mRNA copy, and $c_p$ is the cost per protein copy (Fig. 1a). Together with this work, we also uncover the mechanistic meanings of the coefficients $c_m$ and $c_p$, and they are proportional to the fraction of resources consumed in the gene expression process of the particular gene [40]; we do not go into the details of the mechanistic origin of $c_m$ and $c_p$ since it is not essential to our conclusions in this work.

We define the benefit function $B(p)$ as the relative increase in growth rate due to the produced protein, which approaches a constant value that we normalize as 1. $B(p)$ may decrease for a very large $p$; nevertheless, the decreasing regime is irrelevant to us since the optimal $p$ maximizing fitness must be away from the decreasing regime given the cost of protein production (Fig. 1b). This scenario appears to agree with multiple *S. cerevisiae* genes if not all [19, 41, 42] (Fig. S1). To simplify without losing the essential ingredient, we set

$$B(p) = B_0 + (1 - B_0)\frac{p^s}{p^s + p_0^s}, \tag{2}$$

where $B_0$ is the basis fitness without protein ($0 < B_0 < 1$), $p_0$ is the characteristic protein number needed for cell growth, and $s$ is the sensitivity factor quantifying how sensitive cell growth is to the protein copy number.

One may expect that given a fixed protein copy number, the smaller the mRNA copy number, the higher fitness since it minimizes the transcription cost [Eq. (1)]. However, a small mRNA copy number also leads to a high protein noise: the protein number fluctuates significantly around its mean value, reducing the benefit function (Fig. 1b). Thus, the complete benefit function is a function of both $p$ and $m$, where $m$ enters the benefit function through a term proportional to the local curvature of the benefit function $B''(p)$,

$$\begin{aligned} B(m,p) = \overline{B(p)} &= B(p) + \frac{1}{2}B''(p)p^2\mathrm{CV}^2 \\ &= 1 - D_p - D_m. \end{aligned} \tag{3}$$

Here, CV is the coefficient of variation for the fluctuating protein number, which depends on the mRNA copy number and can be well approximated as $\mathrm{CV}^2 = \frac{\tau_m}{\tau_p m}$ with $\tau_m$ and $\tau_p$ the lifetimes of mRNA and protein, respectively [7, 43–46]. A Taylor expansion is used at the average protein number in deriving Eq. (3). We note that $B(p,m)$ is below 1 because of two mechanisms: one is due to protein inadequacy, which we quantify by a deficiency function $D_p = 1 - B(p)$; the other is due to mRNA inadequacy

that generates high protein noise, which we quantify by another deficiency function $D_m = -\frac{1}{2}B''(p)p^2\mathrm{CV}^2$. Given the cost and benefit, we define the fitness function as

$$F(m,p) = B(m,p)[1 - C(m,p)]. \tag{4}$$

The mRNA and protein copy number affect the fitness through coupled pathways (Fig. 1c).

## Emerging scaling between mRNA and protein

Numerous experiments suggested that gene expression levels are under evolutionary selection [2, 19, 41, 42, 47–53]. In the following, we find the optimal values of $m$ and $p$ that maximize fitness, which can be approximated as

$$F(m,p) = 1 - D_p - D_m - c_m m - c_p p \tag{5}$$

Here, we neglect higher-order corrections because $c_m$ and $c_p$ are typically small (e.g., $c_m \approx 10^{-6}$, $c_p \approx 10^{-8}$ for *S. cerevisiae* [3, 5, 40, 54]). Because $c_p$ is small, the benefit of protein production overwhelms its cost, which means that cells can produce adequate proteins until $B(p) \to 1$. Therefore, the optimal $p$ should be overabundant, i.e., $p \gg p_0$, which means that $D_p$ is approximately proportional to $p^{-s}$ and the ratio between $D_p$ and $D_m$ is simplified as

$$\frac{D_m}{D_p} = \frac{s(s+1)}{2}\mathrm{CV}^2. \tag{6}$$

The $D_m$ and $D_p$ ratio allows us to define a critical CV of protein number: $\mathrm{CV}_c^2 = 2/s(s+1)$. For a given protein, if its CV $<$ CV$_c$, the total deficiency is dominated by protein inadequacy; if CV $>$ CV$_c$, the total deficiency is dominated by noise due to mRNA inadequacy.

Maximization of fitness against $m$ and $p$ leads to the following intriguing results:

$$C_{tx} = D_m, \quad C_{tl} = s(D_p + D_m). \tag{7}$$

At the optimal fitness condition, the transcription cost equals the deficiency due to mRNA inadequacy; the translation cost equals the total deficiency multiplied by the sensitivity factor $s$. Combining Eqs. [6-7], we obtain a simple and insightful relationship between the optimal $m$ and the optimal $p$:

$$\frac{m}{m_c} + \left(\frac{m}{m_c}\right)^2 = 2\frac{p}{p_c}. \tag{8}$$

Here, $m_c = \frac{\tau_m}{\tau_p}\frac{s(s+1)}{2}$ and $p_c = \frac{c_m\tau_m}{c_p\tau_p}s^2(s+1)$. They are the critical mRNA number and protein number, respectively, corresponding to the critical gene expression level at which CV = CV$_c$ and $D_p = D_m$.

Eq. (8) shows that for a highly expressed gene, i.e., $p \gg p_c$, a square-root relationship emerges between $m$ and $p$:

$$m \sim p^{0.5}. \tag{9}$$

Biologically, for a highly expressed gene, protein noise is mild; therefore, CV $\ll$ CV$_c$ and $D_m \ll D_p$ according to Eq. (6). The deficiency due to mRNA inadequacy is much smaller than the deficiency due to protein inadequacy, which means the cell does not need to produce too many copies of mRNA to reduce protein noise. For a specific organism, if $p \gg p_c$ is satisfied for most genes (equivalently, $m \gg m_c$ and CV $\ll$ CV$_c$), a square-root scaling emerges between the mRNA copy number and the protein copy number across the genome (Fig. 2a). In this scenario, since $\mathrm{CV}^2 \sim 1/m$, we also predict that $\mathrm{CV}^2 \sim p^{-0.5}$ (Fig. 2b). Later, we show this is approximately the case for *S. cerevisiae*.

On the contrary, for a lowly expressed gene, i.e., $p \ll p_c$, the left side of Eq. [8] is dominated by the linear term such that

$$m \sim p. \tag{10}$$

In this case, protein noise is significant and $D_m \gg D_p$, which means that the cell needs to produce enough copies of mRNA to reduce the protein noise. For a specific organism, if $p \ll p_c$ is satisfied for most genes (equivalently, $m \ll m_c$ and CV $\gg$ CV$_c$), a linear scaling emerges between the mRNA copy number and the protein copy number across the genome (Fig. 2a) along with the inverse scaling between $\mathrm{CV}^2$ and $p$ (Fig. 2b). Later, we show this is approximately the case for *E. coli*. In the intermediate cases where the ratio between $p$ and $p_c$ can be

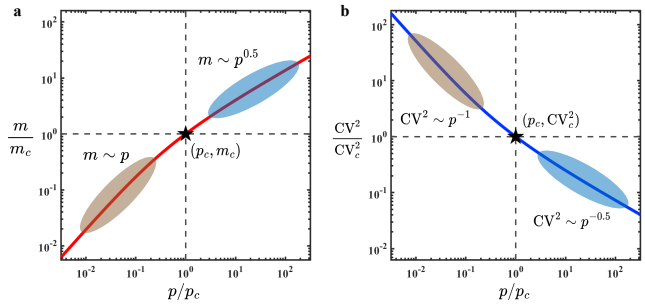


FIG. 2. **The scaling relationship between the mRNA copy number, the protein copy number, and the protein noise under the condition of maximum fitness.** (a) For an organism, if the condition $p/p_c > 1$ (or $m/m_c > 1$) is satisfied for most genes, a square-root scaling appears between $m$ and $p$ across the genome: $m \sim p^{0.5}$ (the blue shade). In contrast, if $p/p_c < 1$ (or $m/m_c < 1$) is satisfied for most genes, a linear scaling appears between $m$ and $p$ across the genome: $m \sim p$ (the brown shade). The scaling relationship can be generalized for the intermediate cases as $m \sim p^\gamma$ with $0.5 < \gamma < 1$. (b) The corresponding relationship between the protein noise (quantified as $\mathrm{CV}^2$) and the protein copy number: $\mathrm{CV}^2 \sim p^{-\gamma}$.

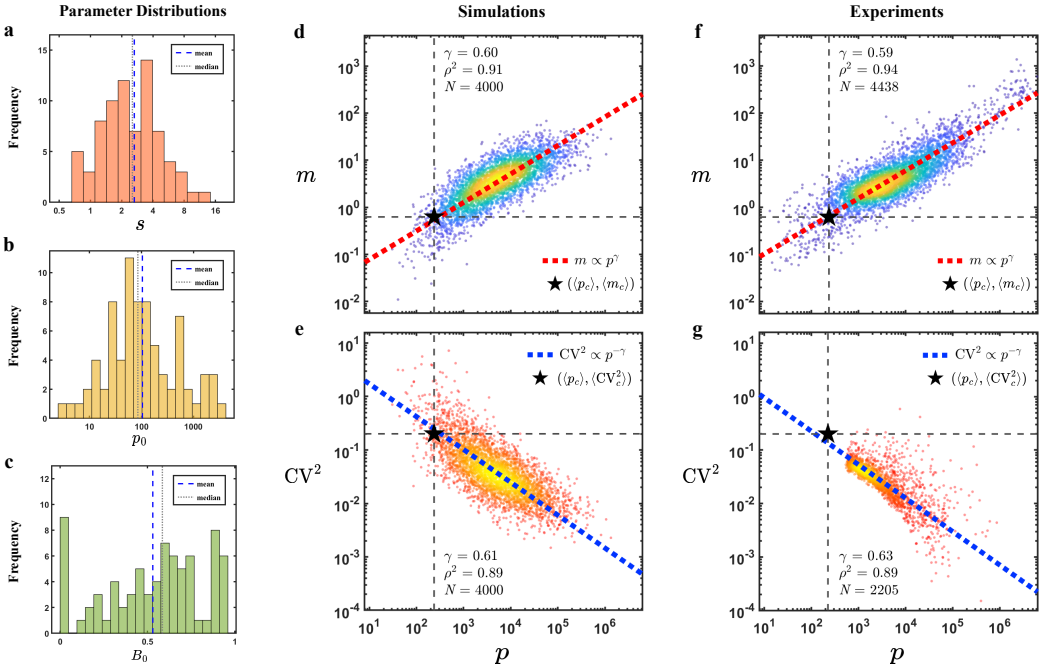


FIG. 3. **Comparison between the model predictions and experimental data for** *S. cerevisiae***.** (a-c) The parameters of the benefit functions ($s$, $p_0$, and $B_0$) extracted from experiments measuring the effects of protein level on fitness [19]. Here, we present the results for cells grown in glucose; similar results for cells grown in galactose are shown in Fig. S2. (d-e) The scaling of $m$-$p$ and CV$^2$-$p$ for the *in silico S. cerevisiae* cell that optimize cell fitness. Each data point represents a gene ($N = 4000$). The stars mark the locations of the genome-wide average $\langle m_c \rangle$, $\langle p_c \rangle$, and $\langle \mathrm{CV}_c^2 \rangle$. Since the protein numbers for most genes are above $\langle p_c \rangle$, we predict a scaling exponent $\gamma$ close to 0.5, in agreement with measurements. (f-g) The experimental data of $m$ vs. $p$ (from Ref. [20]) and CV$^2$ vs. $p$ (from Ref. [9]) across the genome for *S. cerevisiae*. All the experimental data used in this figure are from conditions of rapid growth. The $\gamma$ exponent is calculated by the total-least-squares (TLS) method (Appendix B), which is also used in other relevant figures unless otherwise mentioned. $\rho^2$ is the fraction of total variance explained by the power-law scaling $m \sim p^\gamma$. In (d-g), the color of the dots represents the local density, with yellow as the high density and blue or red as the low density.

either above 1 or below 1 across the genome, we expect a scaling relationship $m \sim p^\gamma$ (also, $\mathrm{CV}^2 \sim p^{-\gamma}$) and the $\gamma$ exponent must be between 0.5 and 1, which we confirm later using data from human, mouse, and fission yeast.

### Simulations vs. experiments for *S. cerevisiae*

In the following, we numerically simulate an *in silico* cell using parameters of *S. cerevisiae* [3], for which the benefit functions have been massively measured for about 80 genes [19]. We analyze the data from which we infer the distributions of the parameters $s$, $p_0$, and $B_0$ based on Eq. (2) (Fig. 3a-c; Fig. S1). For each gene of the *in silico* cell, we randomly sample the parameters from distributions that make good approximations of the experimental data with the average values of the *in silico* cell the same as those of experimental data (Table I). For each gene, we compute the optimal mRNA and protein copy numbers that maximize cell fitness, and also the CV$^2$ of the protein numbers (Fig. 3d-e). To verify the assumption we make in the analytical analysis, we confirm that the protein copy numbers are indeed in large surplus, $p > p_0$, particularly significant for lowly expressed genes (Fig. S3a), in agreement with previous findings that proteins are typically in excess of what is required for function and that large-magnitude genetic perturbation is needed to generate strong phenotypes [5, 19].

One should note that $m_c$, $p_c$, and $\mathrm{CV}_c$ are different for each gene; nevertheless, we can still compute their average values using the genome-wide averaged values of parameters (e.g., $\langle s \rangle$, see Table I), and find that $\langle m_c \rangle \approx 0.6$, $\langle p_c \rangle \approx 240$, and $\langle \mathrm{CV}_c^2 \rangle \approx 0.2$. Intriguingly, $\langle m_c \rangle$ and $\langle p_c \rangle$ are below the corresponding values for most genes for the *in silico* cell, while $\langle \mathrm{CV}_c \rangle$ is over the CV of most genes (Fig. 3d-e). These observations predict that the scaling exponent between $m$ and $p$ should be close to 0.5, and the scaling exponent between CV$^2$ and $p$ should be close to $-0.5$. We fit the *in silico* data of $m$ vs. $p$ and CV$^2$ vs. $p$ and indeed find the exponents $\gamma \approx 0.59$ for both scalings (Fig. 3d-e).

Surprisingly, the experimental data are very close to the simulation results in terms of both the scaling exponents and the absolute values (Fig. 3f-g), which strongly

TABLE I. **Parameters for the simulations of** *S. cerevisiae* **cell.** $\langle X \rangle$ represent the mean of the distribution, extracted from experiments [3, 19] (Appendix A). All of the parameters are sampled from a lognormal distribution ($log_{10} \frac{X}{\langle X \rangle}$ follows a normal distribution with mean zero and standard deviation $\sigma$) except $B_0$, which is sampled from a uniform distribution between 0 and 1. The details of the estimation of the parameter values are in Appendix A.

| Parameter | $c_p$ | $c_m$ | $s$ | $p_0$ | $B_0$ | $\tau_m$ (min) | $\tau_p$ (min) |
|---|---|---|---|---|---|---|---|
| $\langle X \rangle$ | $1.2 \times 10^{-8}$ | $8.4 \times 10^{-7}$ | 2.7 | 100 | 0.5 | 15 | 120 |
| $\sigma$ | 0.3 | 0.3 | 0.25 | 0.6 | —— | 0.3 | 0.15 |

supports the validity of our model: balancing the cost and benefit of gene expression is the driving force of gene expression regulation. We remark that for *S. cerevisiae*, post-transcriptional regulation is significant because $p/m \sim m^{0.69}$, in particular for highly expressed genes: the higher protein copy numbers, the more copy numbers of protein produced per mRNA copy.

## Simulations vs. experiments for *E. coli*

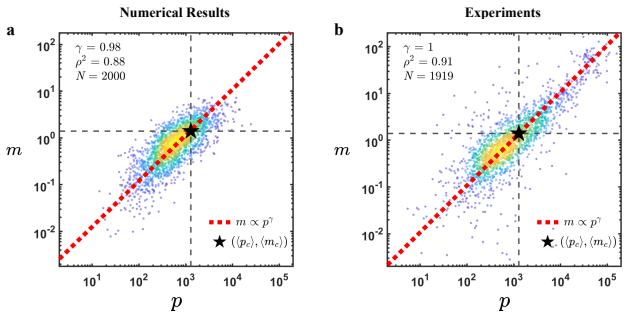


FIG. 4. **The scaling of $m$-$p$ for the *in silico E. coli* cell from optimizing the fitness function (a) and the experimental data (b).** The stars mark the locations of $\langle m_c \rangle$ and $\langle p_c \rangle$. Since the protein numbers for most genes are below $\langle p_c \rangle$, we predict a scaling exponent $\gamma$ close to 1, in agreement with the experimental data from Ref. [21]. The color of the dots (yellow to blue) represents the local density (high to low).

We perform a similar analysis for *E. coli*. As far as we realize, no systematic measurements of the benefit functions have been done for *E. coli*. We expect $p_0$ of *E. coli* should be smaller than *S. cerevisiae* due to its smaller size. Meanwhile, we argue that the overabundance of protein copy numbers should be less significant in *E. coli* than *S. cerevisiae* since all proteins are constrained within a smaller volume. This corresponds to a higher sensitivity factor where the benefit function saturates quickly, and overabundance is futile or even detrimental [2, 55].

Given these reasonable assumptions, we set the relevant parameters of the benefit functions and other parameters for *E. coli* (Table S1 and Supplementary Information). We obtain the genome-wide averaged values of $m_c$, $p_c$, and $\mathrm{CV}_c^2$: $\langle m_c \rangle \approx 1.4$, $\langle p_c \rangle \approx 1.3 \times 10^3$, and $\langle \mathrm{CV}_c^2 \rangle \approx 0.05$. These values are above most genes in this case. Indeed, the scaling exponent $\gamma$ is close to 1 for the *in silico E. coli* cell (Fig. 4a). Intriguingly, the scaling exponent measured from the genome-wide experimental data [21] is also close to 1 (Fig. 4b). The agreement between simulations and experiments strongly suggests that most genes in the experimental data are below the critical gene expression level; therefore, a linear scaling between $m$ and $p$ emerges. We remark that for *E. coli*, post-transcriptional regulation is insignificant because $p/m \sim 1$, that is, the protein expression level is mainly set by the mRNA expression level. Again, we observe an overabundance of protein copy numbers ($p > p_0$), especially notable for low-demand genes (Fig. S3b), which ensures robust growth against genetic perturbations [56].

## Predominance of transcription regulation

We further analyze the genome-wide data of $m$ vs. $p$ for multiple organisms, including *Schizosaccharomyces pombe* (fission yeast) [22], *Mus musculus* (mouse) [23, 54], and *Homo sapiens* (human) [23, 54]. In agreement with our model predictions, the scaling exponents between $m$ and $p$ are always between 0.5 and 1 (Fig. 5a-c). We introduce a transcription-predominant factor, defined as $\theta = 2\gamma - 1$. $\theta = 1$ means that all the regulation of protein levels is implemented at the transcription level, and $\theta = 0$ means that transcription and translation equally contribute to the protein level.

We remark that the $\gamma$ exponent is presumably condition- and strain-dependent since the gene expression levels and other parameters (e.g., the sensitivity factor $s$) can differ. We analyze six other datasets, including three for *E. coli* [21, 28, 57], and one for *S. cerevisiae* [58], *M. musculus* [59] and *H. sapiens* [60]. Intriguingly, although the $\gamma$ exponents can be different for the same organism across different conditions (Fig. S4), the transcription-predominant factor is relatively conservative for a given organism (Fig. 5d). While *E. coli* and the two yeasts (*S. cerevisiae* and *S. pombe*) are close to

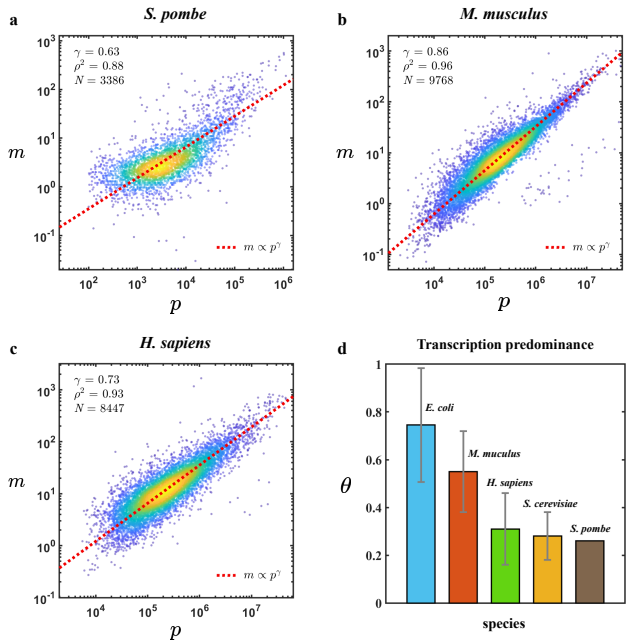


FIG. 5. **The $m$-$p$ scaling across multiple organisms.** (a-c) The experimental data of $m$ vs. $p$ for multiple organisms. *S. pombe* data from [22]; *M. musculus* (NIH3T3 cell) and *H. sapiens* (Hela cell) data from [23]. The color of the dots (yellow to blue) represents the local density (high to low). (d) The transcription-predominant factor. $\theta = 1$ means that all the protein regulation is implemented at the transcription level, and $\theta = 0$ means that transcription and translation equally contribute to the protein level. The error bar represents the standard error among different datasets, if available.

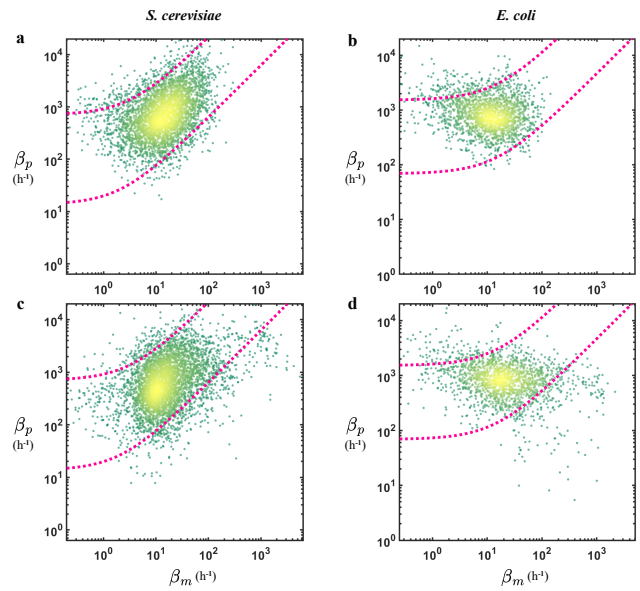


FIG. 6. **The occupied area in the phase space of translation rate and transcription rate is constrained.** (a-b) The simulation data are from the same simulations as Fig. 3d-e and Fig. 4a. The two boundaries, which are Eq. (11) with $k$ and $\beta_{p,\min}$ taken as the 5th and 95th percentiles from the sampled data, well constrain the occupied area of the phase space of $\beta_p$ and $\beta_m$. (c-d) The corresponding experimental data with the two boundaries the same as (a) and (b). Data in (c) is from the same dataset in Fig. 3f, and data in (d) is from the same dataset in Fig. 4b. The color of the dots (yellow to green) represents the local density (high to low).

the opposite limiting scenarios, transcription regulation is predominant (i.e., $\theta > 0$) in determining protein levels for all the organisms we study. We remark that the predominance of transcription regulation is a natural result of fitness optimization because $\gamma$ has a lower bound 0.5 according to our model; meanwhile, no biochemical constraints require that $\gamma$ has to be larger than 0.5.

### Constraints of transcription and translation rates

We introduce the transcription rate $\beta_m$, the copy number of mRNA produced per unit time for a given gene, and the translation rate $\beta_p$, the copy number of protein produced per mRNA per unit time. Using the steady-state conditions, $m = \beta_m \tau_m$, $p = m\beta_p\tau_p$ together with Eq. [8], we obtain the expression of the ratio of $\beta_p$ to $\beta_m$ that maximizes fitness:

$$\frac{\beta_p}{\beta_m} = k + \frac{\beta_{p,\min}}{\beta_m} \tag{11}$$

where $k = \frac{2c_m}{(s+1)c_p}$ and $\beta_{p,\min} = \frac{c_m s}{c_p \tau_p}$. Eq. (11) has two intriguing predictions. First, it suggests that in the small $\beta_m$ limit, a lower bound exists for $\beta_p$, which is $\beta_{p,\min}$. Second, in the large $\beta_m$ limit, $\beta_p/\beta_m = k$.

Imagine a group of genes, and they share the same $k$, $\beta_{p,\min}$ in Eq. [11]. In a log-log plot of $\beta_p$ vs. $\beta_m$, all these genes will be on the same curve of Eq. (11). In real scenarios, heterogeneity among genes must exist. Therefore, for the *in silico* cells, we plot two boundary lines of Eq. (11) using the 5th and 95th percentiles of $k$ and $\beta_{p,min}$ from the sampled data, respectively. For both *S. cerevisiae* and *E. coli*, the two predicted boundaries well describe the occupied region in the phase space $\beta_m$ and $\beta_p$ (Fig. 6a-b), with a small fraction of genes crossing the boundaries as expected. For the experimental data, the same boundaries also well constrain the occupied space of $\beta_p$ and $\beta_m$ (Fig. 6c-d), in agreement with simulations. Interestingly, our simulations automatically present an upper bound for the translation rate (Fig. 6a-b), which coincides with the maximal possible translation rate, $\beta_{p,\max} \sim 10^{3.6}$ h$^{-1}$, in the limit ribosomes follow each other closely along the mRNA [33, 54].

## DISCUSSION

In this work, we unveil a universal relationship between the mRNA and protein copy numbers, which maximizes cell fitness. Surprisingly, the scaling exponent $\gamma$ between

$m$ and $p$ must be between 0.5 and 1; *S. cerevisiae* and *E. coli* appear near the opposite limiting scenarios. The lower bound 0.5 of the $\gamma$ exponent suggests the predominance of transcription regulation in setting the protein level, which is simply a consequence of optimal growth; no mechanistic constraint prevents $\gamma$ from being below 0.5. Our prediction is in agreement with the numerous experiments showing that the mRNA levels make either primary [21, 24–26, 61–66] or roughly equal [20, 67–71] contribution to the protein abundance compared to other regulations like translation rate and protein degradation.

Optimizing fitness under a cost-benefit trade-off has been applied to determine the expression levels of particular genes both theoretically and experimentally [2, 28, 72, 73]. However, the consequences of the cost-benefit trade-off on the genome-wide mRNA-protein correlations are still elusive. Hausser et al. found the upper bound of transcription rate for optimal growth by considering the trade-off between the transcription cost and the benefit of noise reduction by transcription [54]. In their model, the only degree of freedom is the transcription rate, while in our case, we let both the mRNA and protein levels change, which allows us to identify the mRNA-protein scaling and the constraints in the space of transcription and translation rates Eq. (11).

Lo et al. built a mathematical model to explore the trade-off between metabolic load and growth robustness [5], where they defined fitness as the population growth rate and assumed that cells may stop growing if the protein copy number is below a threshold value. The authors found that optimal protein expression levels for many genes are vastly overabundant and predicted a linear relationship between translation rate and mRNA copy number, which aligns with the $m \sim p^{0.5}$ scaling of our model. However, their model cannot reconcile the discrepancy in the scaling behaviors of *E. coli* and *S. cerevisiae*.

Our model maximizes cell growth to find the optimal expression levels of mRNA and protein, which may not be always valid [74]. For example, in the stationary phase of microorganisms, cell fitness involves the survival ability in a starving environment instead of growth [75]. Another case is multicellular organisms, as the principle of optimal growth may not apply to some well-developed tissues like brain [76, 77] or to cells experiencing complex biological processes like differentiation [78]; nevertheless, we still observe robust a power-law scaling between the mRNA and protein copy numbers for mouse and human, although the conclusions may depend on cell types [79, 80]. In the future, it will be exciting to explore the mechanisms that determine the $\gamma$ exponent and understand the origin of the discrepancy between different organisms.

## APPENDIX A: EXTRACTING PARAMETERS FROM EXPERIMENTS

Kafri et al. [3] measured the cost of unnecessary protein production by integrating different mCherry constructs into the genome, from which we calculated the transcription cost per mCherry mRNA to be $4.2 \times 10^{-7}$ and the translation cost per mCherry protein to be $5.9 \times 10^{-9}$ [40]. Considering the mCherry protein is about half the size of an average protein in *S. cerevisiae*, we have higher costs for an average gene: $c_m = 8.4 \times 10^{-7}$ and $c_p = 1.2 \times 10^{-8}$. We set $\sigma = 0.3$ (Table I) for $c_m$ and $c_p$ so that 90% of variables are within one order of magnitude. The heterogeneity of $c_m$ and $c_p$ across the genome can originate from differences in gene length, elongation speed, amino acid compositions, etc [37].

Keren et al. [19] measured the effects of protein level on the fitness for 81 genes, which we fit using the expression of the benefit function, $B_0 + (1 - B_0)\frac{p^s}{p^s + p_0^s}$ (Fig. S1). One should note the decrease of the fitness due to gene expression cost is very weak in the regime we fit the data, that is, $c_p p \ll 1$ and $c_m m \ll 1$; therefore, the benefit function can be well approximated by the measured fitness. We exclude genes with a decreasing fitness before the wild-type expression, leaving 77 genes remaining for the glucose condition and 70 genes remaining for the galactose condition. If the fitness decreases after the maximum, we set the decreasing part after the maximum to be 1 for the fitting since we only utilize the properties of the increasing interval where the optimal protein number that maximizes fitness is located. We take the conversion factor between the protein number and fluorescence level as 60, which is calibrated using the wild-type expression levels [19].

The average mRNA degradation rate is set to be 15 min for *S. cerevisiae*, which is among several experimental measurements (12min [81], 14 min [82], 16 min [83]). We set $\sigma$=0.3 for $\tau_m$ so that 90% of them are within one order of magnitude. The protein lifetime is calculated by $\tau_p = \frac{1}{ln2}/(\frac{1}{1.6\text{h}} + \frac{1}{10\text{h}}) = 2$ h where 1.6 h [3] is the cell doubling time (dilution effect) and 10 h is the average protein degradation half-life [84]. We notice that the protein half-life is mainly set by the cell doubling time [84–86], such that using a longer reported protein degradation half-life [87] has a mild effect on $\tau_p$. Therefore, the standard deviation of $\tau_p$ should be small, and we set $\sigma$=0.15 for $\tau_p$.

## APPENDIX B: FITTING OF THE $\gamma$ EXPONENT

The ordinary-least-squares (OLS) linear regression considers only the noise in the vertical-axis-variable and assumes the horizontal-axis-variable is perfectly measured, which is not true in actual measurements [88]. Therefore, the OLS regression fails to accurately capture

the $\gamma$ exponent; particularly, the $m$-vs-$p$ slope is not the inverse of the $p$-vs-$m$ slope [68]. In contrast, the total-least-squares (TLS) method accounts for errors or uncertainties in both dimensions and minimizes the sum of the squared perpendicular distances from the data points to the fitted line [89]. Consequently, TLS more faithfully and robustly reflects the relationship between mRNA and protein levels with a stronger explanatory power [90, 91].

In this article, different statistical methods are applied to multiple datasets (Table S2 and Fig. S5). We show that the OLS regression leads to a biased lower slope, an effect called regression dilution bias [92, 93]. Such deviation is especially significant when regressing $p$ on $m$ due to large noise in mRNA levels, which might lead to obscureness regarding the $m$-$p$ correlation. Therefore, the $\gamma$ exponent of all relevant figures in this article is calculated by the TLS method.


[1] A. Wagner, Energy constraints on the evolution of gene expression, Mol. Biol. Evol. **22**, 1365 (2005).

[2] E. Dekel and U. Alon, Optimality and evolutionary tuning of the expression level of a protein, Nature **436**, 588 (2005).

[3] M. Kafri, E. Metzl-Raz, G. Jona, and N. Barkai, The cost of protein production, Cell reports **14**, 22 (2016).

[4] J. M. Peters, A. Colavin, H. Shi, T. L. Czarny, M. H. Larson, S. Wong, J. S. Hawkins, C. H. Lu, B.-M. Koo, E. Marta, A. L. Shiver, E. H. Whitehead, J. S. Weissman, E. D. Brown, L. S. Qi, K. C. Huang, and C. A. Gross, A comprehensive, crispr-based functional analysis of essential genes in bacteria, Cell **165**, 1493 (2016).

[5] T. W. Lo, H. J. Choi, D. Huang, and P. A. Wiggins, Noise robustness and metabolic load determine the principles of central dogma regulation, Science Advances **10**, eado3095 (2024).

[6] N. M. Belliveau, G. Chure, C. L. Hueschen, H. G. Garcia, J. Kondev, D. S. Fisher, J. A. Theriot, and R. Phillips, Fundamental limits on the rate of bacterial growth and their influence on proteomic composition., Cell systems (2021).

[7] J. Paulsson, Summing up the noise in gene networks, Nature **427**, 415 (2004).

[8] J. Raser and E. O'Shea, Noise in gene expression: origins, consequences, and control, Science **309**, 2010 (2005).

[9] J. R. S. Newman, S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. Derisi, and J. S. Weissman, Single-cell proteomic analysis of s. cerevisiae reveals the architecture of biological noise, Nature **441**, 840 (2006).

[10] A. Sanchez, S. Choubey, and J. Kondev, Regulation of noise in gene expression, Annual Review of Biophysics **42**, 469 (2013).

[11] J. Lin and A. Amir, The effects of stochasticity at the single-cell level and cell size control on the population growth, Cell Systems **5**, 358 (2017).

[12] W. E. Bentley, N. Mirjalili, D. C. Andersen, R. H. Davis, and D. S. Kompala, Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant bacteria, Biotechnology and bioengineering **35**, 668 (1990).

[13] H. Dong, L. Nilsson, and C. G. Kurland, Gratuitous overexpression of genes in escherichia coli leads to growth inhibition and ribosome destruction, Journal of bacteriology **177**, 1497 (1995).

[14] C. Rang, J. E. Galen, J. B. Kaper, and L. Chao, Fitness cost of the green fluorescent protein in gastrointestinal bacteria, Canadian journal of microbiology **49**, 531 (2003).

[15] M. Scott, C. Gunderson, E. Mateescu, Z. Zhang, and T. Hwa, Interdependence of cell growth and gene expression: origins and consequences, Science **330**, 1099 (2010).

[16] M. Lynch and G. K. Marinov, The bioenergetic costs of a gene, Proceedings of the National Academy of Sciences **112**, 15690 (2015).

[17] A. San Millan, M. Toll-Riera, Q. Qi, and R. C. MacLean, Interactions between horizontally acquired genes create a fitness cost in pseudomonas aeruginosa, Nature communications **6**, 6845 (2015).

[18] Y. Eguchi, K. Makanae, T. Hasunuma, Y. Ishibashi, K. Kito, and H. Moriya, Estimating the protein burden limit of yeast cells by measuring the expression limits of glycolytic proteins, Elife **7**, e34595 (2018).

[19] L. Keren, J. Hausser, M. Lotan-Pompan, I. V. Slutskin, H. Alisar, S. Kaminski, A. Weinberger, U. Alon, R. Milo, and E. Segal, Massively parallel interrogation of the effects of gene expression levels on fitness, Cell **166**, 1282 (2016).

[20] D. E. Weinberg, P. Shah, S. W. Eichhorn, J. A. Hussmann, J. B. Plotkin, and D. P. Bartel, Improved ribosome-footprint and mrna measurements provide insights into dynamics and regulation of yeast translation, Cell reports **14**, 1787 (2016).

[21] R. Balakrishnan, M. Mori, I. Segota, Z. Zhang, R. Aebersold, C. Ludwig, and T. Hwa, Principles of gene regulation quantitatively connect dna to rna and proteins in bacteria, Science **378**, eabk2066 (2022).

[22] S. Marguerat, A. Schmidt, S. Codlin, W. Chen, R. Aebersold, and J. Bähler, Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells, Cell **151**, 671 (2012).

[23] S. W. Eichhorn, H. Guo, S. E. McGeary, R. A. Rodriguez-Mias, C. Shin, D. Baek, S.-h. Hsu, K. Ghoshal, J. Villén, and D. P. Bartel, mrna destabilization is the dominant effect of mammalian micrornas by the time substantial repression ensues, Molecular cell **56**, 104 (2014).

[24] J. J. Li, P. J. Bickel, and M. D. Biggin, System wide analyses have underestimated protein abundances and the importance of transcription in mammals, PeerJ **2**, e270 (2014).

[25] J. J. Li and M. D. Biggin, Statistics requantitates the central dogma, Science **347**, 1066 (2015).

[26] Y. Liu, A. Beyer, and R. Aebersold, On the dependency of cellular protein levels on mrna abundance, Cell **165**, 535 (2016).

[27] C. Buccitelli and M. Selbach, mrnas, proteins and the emerging principles of gene expression control, Nature Reviews Genetics **21**, 630 (2020).

[28] G.-W. Li, D. Burkhardt, C. Gross, and J. S. Weissman, Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources, Cell **157**, 624 (2014).

[29] M. N. Price, K. M. Wetmore, A. M. Deutschbauer, and A. P. Arkin, A comparison of the costs and benefits of

bacterial gene expression, PLOS ONE **11**, 1 (2016).
[30] P. Shah, Y. Ding, M. Niemczyk, G. Kudla, and J. B. Plotkin, Rate-limiting steps in yeast protein translation, Cell **153**, 1589 (2013).
[31] F. Ceroni, R. Algar, G.-B. Stan, and T. Ellis, Quantifying cellular capacity identifies gene expression designs with reduced burden, Nature methods **12**, 415 (2015).
[32] G. Wu, Q. Yan, J. A. Jones, Y. J. Tang, S. S. Fong, and M. A. Koffas, Metabolic burden: cornerstones in synthetic biology and metabolic engineering applications, Trends in biotechnology **34**, 652 (2016).
[33] J. Lin and A. Amir, Homeostasis of protein and mrna concentrations in growing cells, Nature Communications **9**, 4496 (2018).
[34] Y. Han and F. Zhang, Control strategies to manage trade-offs during microbial production, Current opinion in biotechnology **66**, 158 (2020).
[35] M. P. Swaffer, G. K. Marinov, H. Zheng, L. Fuentes Valenzuela, C. Y. Tsui, A. W. Jones, J. Greenwood, A. Kundaje, W. J. Greenleaf, R. Reyes-Lamothe, and J. M. Skotheim, Rna polymerase ii dynamics and mrna stability feedback scale mrna amounts with cell size, Cell **186**, 5254 (2023).
[36] L. Calabrese, L. Ciandrini, and M. C. Lagomarsino, How total mrna influences cell growth, Proceedings of the National Academy of Sciences **121**, e2400679121 (2024).
[37] I. Frumkin, D. Schirman, A. Rotman, F. Li, L. Zahavi, E. Mordret, O. Asraf, S. Wu, S. F. Levy, and Y. Pilpel, Gene architectures that minimize cost of gene expression, Molecular cell **65**, 142 (2017).
[38] K. Tomala and R. Korona, Evaluating the fitness cost of protein expression in saccharomyces cerevisiae, Genome biology and evolution **5**, 2051 (2013).
[39] D. M. Stoebel, A. M. Dean, and D. E. Dykhuizen, The cost of expression of escherichia coli lac operon proteins is in the process, not in the products, Genetics **178**, 1653 (2008).
[40] Y. Yan and J. Lin, Disentangling the cost of gene expression, To be published (2025).
[41] J. S. Rest, C. M. Morales, J. B. Waldron, D. A. Opulente, J. Fisher, S. Moon, K. Bullaughey, L. B. Carey, and D. Dedousis, Nonlinear fitness consequences of variation in expression level of a eukaryotic gene, Molecular biology and evolution **30**, 448 (2012).
[42] F. Duveau, P. V. Zande, B. P. Metzger, C. J. Diaz, E. A. Walker, S. Tryban, M. A. Siddiq, B. Yang, and P. J. Wittkopp, Mutational sources of trans-regulatory variation affecting gene expression in saccharomyces cerevisiae, Elife **10**, e67806 (2021).
[43] J. Paulsson, Models of stochastic gene expression, Physics of Life Reviews **2**, 157 (2005).
[44] A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, E. O'Shea, Y. Pilpel, and N. Barkai, Noise in protein expression scales with natural protein abundance, Nature genetics **38**, 636 (2006).
[45] V. Shahrezaei and P. S. Swain, Analytical distributions for stochastic gene expression, Proceedings of the National Academy of Sciences **105**, 17256 (2008).
[46] M. Mundt, A. Anders, S. M. Murray, and V. Sourjik, A system for gene expression noise control in yeast, ACS synthetic biology **7**, 2618 (2018).
[47] F. Duveau, W. Toubiana, and P. J. Wittkopp, Fitness effects of cis-regulatory variants in the saccharomyces cerevisiae tdh3 promoter, Molecular biology and evolution **34**, 2908 (2017).
[48] F. Duveau, A. Hodgins-Davis, B. P. Metzger, B. Yang, S. Tryban, E. A. Walker, T. Lybrook, and P. J. Wittkopp, Fitness effects of altering gene expression noise in saccharomyces cerevisiae, Elife **7**, e37272 (2018).
[49] G. I. Lang, A. W. Murray, and D. Botstein, The cost of gene expression underlies a fitness trade-off in yeast, Proceedings of the National Academy of Sciences **106**, 5755 (2009).
[50] F. J. Poelwijk, P. D. Heyning, M. G. de Vos, D. J. Kiviet, and S. J. Tans, Optimality and evolution of transcriptionally regulated gene expression, BMC systems biology **5**, 1 (2011).
[51] A. Y. Weiße, D. A. Oyarzún, V. Danos, and P. S. Swain, Mechanistic links between cellular trade-offs, gene expression, and growth, Proceedings of the National Academy of Sciences **112**, E1038 (2015).
[52] P. Vande Zande, M. S. Hill, and P. J. Wittkopp, Pleiotropic effects of trans-regulatory mutations on fitness and gene expression, Science **377**, 105 (2022).
[53] A. Gyorgy, Competition and evolutionary selection among core regulatory motifs in gene expression control, Nature communications **14**, 8266 (2023).
[54] J. Hausser, A. Mayo, L. Keren, and U. Alon, Central dogma rates and the trade-off between precision and economy in gene expression, Nature communications **10**, 68 (2019).
[55] S. Klumpp, M. Scott, S. Pedersen, and T. Hwa, Molecular crowding limits translation and cell growth, Proceedings of the National Academy of Sciences **110**, 16754 (2013).
[56] S. Donati, M. Kuntz, V. Pahl, N. Farke, D. Beuter, T. Glatter, J. V. Gomes-Filho, L. Randau, C.-Y. Wang, and H. Link, Multi-omics analysis of crispri-knockdowns identifies mechanisms that buffer decreases of enzymes in e. coli metabolism, Cell Systems **12**, 56 (2021).
[57] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, Quantifying e. coli proteome and transcriptome with single-molecule sensitivity in single cells, Science **329**, 533 (2010).
[58] F. W. Albert, D. Muzzey, J. S. Weissman, and L. Kruglyak, Genetic influences on translation in yeast, PLoS genetics **10**, e1004692 (2014).
[59] B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach, Global quantification of mammalian gene expression control, Nature **473**, 337 (2011).
[60] N. Nagaraj, J. R. Wisniewski, T. Geiger, J. Cox, M. Kircher, J. Kelso, S. Pääbo, and M. Mann, Deep proteome and transcriptome mapping of a human cancer cell line, Molecular systems biology **7**, 548 (2011).
[61] P. Lu, C. Vogel, R. Wang, X. Yao, and E. M. Marcotte, Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation, Nature biotechnology **25**, 117 (2007).
[62] E. Lundberg, L. Fagerberg, D. Klevebring, I. Matic, T. Geiger, J. Cox, C. Älgenäs, J. Lundeberg, M. Mann, and M. Uhlen, Defining the transcriptome and proteome in three functionally different human cell lines, Molecular systems biology **6**, 450 (2010).
[63] M. Jovanovic, M. S. Rooney, P. Mertins, D. Przybylski, N. Chevrier, R. Satija, E. H. Rodriguez, A. P. Fields, S. Schwartz, R. Raychowdhury, *et al.*, Dynamic profiling of the protein life cycle in response to pathogens, Science

**347**, 1259038 (2015).
[64] A. M. Franks, G. Csárdi, D. A. Drummond, and E. M. Airoldi, Estimating a structured covariance matrix from multilab measurements in high-throughput biology, Journal of the American Statistical Association **110**, 27 (2015).
[65] A. Koussounadis, S. P. Langdon, I. H. Um, D. J. Harrison, and V. A. Smith, Relationship between differentially expressed mrna and mrna-protein correlations in a xenograft model system, Scientific reports **5**, 10775 (2015).
[66] D. Choe, J. H. Lee, M. Yoo, S. Hwang, B. H. Sung, S. Cho, B. Palsson, S. C. Kim, and B.-K. Cho, Adaptive laboratory evolution of a genome-reduced escherichia coli, Nature communications **10**, 935 (2019).
[67] C. Vogel, R. de Sousa Abreu, D. Ko, S.-Y. Le, B. A. Shapiro, S. C. Burns, D. Sandhu, D. R. Boutz, E. M. Marcotte, and L. O. Penalva, Sequence signatures and mrna concentration can explain two-thirds of protein abundance variation in a human cell line, Molecular systems biology **6**, 400 (2010).
[68] G. Csárdi, A. Franks, D. S. Choi, E. M. Airoldi, and D. A. Drummond, Accounting for experimental noise reveals that mrna levels, amplified by post-transcriptional processes, largely determine steady-state protein levels in yeast, PLoS genetics **11**, e1005206 (2015).
[69] F. Edfors, F. Danielsson, B. M. Hallström, L. Käll, E. Lundberg, F. Pontén, B. Forsström, and M. Uhlén, Gene-specific correlation of rna and protein levels in human cells and tissues, Molecular systems biology **12**, 883 (2016).
[70] A. Franks, E. Airoldi, and N. Slavov, Post-transcriptional regulation across human tissues, PLoS computational biology **13**, e1005535 (2017).
[71] B. Ho, A. Baryshnikova, and G. W. Brown, Unification of protein abundance datasets yields a quantitative saccharomyces cerevisiae proteome, Cell systems **6**, 192 (2018).
[72] T. Kalisky, E. Dekel, and U. Alon, Cost-benefit theory and optimal design of gene regulation functions, Phys. Biol. **4**, 229 (2007).
[73] M. Scott, S. Klumpp, E. M. Mateescu, and T. Hwa, Emergence of robust growth laws from optimal regulation of ribosome synthesis, Molecular systems biology **10** (2014).
[74] B. D. Towbin, Y. Korem, A. Bren, S. Doron, R. Sorek, and U. Alon, Optimality and sub-optimality in a bacterial growth law, Nature communications **8**, 14123 (2017).
[75] P. K. Herman, Stationary phase in yeast, Current Opinion in Microbiology **5**, 602 (2002).
[76] A. L. Bauernfeind and C. C. Babbitt, The predictive nature of transcript expression levels on protein expression in adult human brain, BMC genomics **18**, 1 (2017).
[77] C. P. Moritz, T. Mühlhaus, S. Tenzer, T. Schulenborg, and E. Friauf, Poor transcript-protein correlation in the brain: negatively correlating gene products reveal neuronal polarity as a potential cause, Journal of neurochemistry **149**, 582 (2019).
[78] A. R. Kristensen, J. Gsponer, and L. J. Foster, Protein synthesis rate is the predominant regulator of protein expression during differentiation, Molecular systems biology **9**, 689 (2013).
[79] D. Popovic, B. Koch, M. Kueblbeck, J. Ellenberg, and L. Pelkmans, Multivariate control of transcript to protein variability in single mammalian cells, Cell systems **7**, 398 (2018).
[80] V. M. Peterson, K. X. Zhang, N. Kumar, J. Wong, L. Li, D. C. Wilson, R. Moore, T. K. McClanahan, S. Sadekova, and J. A. Klappenbach, Multiplexed quantification of proteins and transcripts in single cells, Nature biotechnology **35**, 936 (2017).
[81] P. Eser, C. Demel, K. C. Maier, B. Schwalb, N. Pirkl, D. E. Martin, P. Cramer, and A. Tresch, Periodic mrna synthesis and degradation co-operate during cell cycle gene expression, Molecular systems biology **10**, 717 (2014).
[82] B. Neymotin, R. Athanasiadou, and D. Gresham, Determination of in vivo rna kinetics using rate-seq, Rna **20**, 1645 (2014).
[83] C. Miller, B. Schwalb, K. Maier, D. Schulz, S. Dümcke, B. Zacher, A. Mayer, J. Sydow, L. Marcinowski, L. Dölken, *et al.*, Dynamic transcriptome analysis measures rates of mrna synthesis and decay in yeast, Molecular systems biology **7**, 458 (2011).
[84] R. Christiano, N. Nagaraj, F. Fröhlich, and T. C. Walther, Global proteome turnover analyses of the yeasts s. cerevisiae and s. pombe, Cell reports **9**, 1959 (2014).
[85] M. Martin-Perez and J. Villén, Determinants and regulation of protein turnover in yeast, Cell systems **5**, 283 (2017).
[86] R. Christiano, H. Arlt, S. Kabatnik, N. Mejhert, Z. W. Lai, R. V. Farese, and T. C. Walther, A systematic protein turnover map for decoding protein degradation, Cell reports **33** (2020).
[87] J. M. Pratt, J. Petty, I. Riba-Garcia, D. H. Robertson, S. J. Gaskell, S. G. Oliver, and R. J. Beynon, Dynamics of protein turnover, a missing dimension in proteomics, Molecular & Cellular Proteomics **1**, 579 (2002).
[88] R. C. Jansen, J.-P. Nap, and L. Mlynárová, Errors in genomics and proteomics, Nature biotechnology **20**, 19 (2002).
[89] G. H. Golub and C. F. Van Loan, An analysis of the total least squares problem, SIAM journal on numerical analysis **17**, 883 (1980).
[90] I. Markovsky and S. Van Huffel, Overview of total least-squares methods, Signal processing **87**, 2283 (2007).
[91] S. Van Huffel and P. Lemmerling, *Total least squares and errors-in-variables modeling: analysis, algorithms and applications* (Springer Science & Business Media, 2013).
[92] N. Draper, *Applied regression analysis* (McGraw-Hill. Inc, 1998).
[93] S. Weisberg, Applied linear regression (2005).

## ACKNOWLEDGEMENTS

The research was funded by the National Key Research and Development Program of China (2024YFA0919600) and supported by Peking-Tsinghua Center for Life Sciences grants.