A general model of the dynamics of genes with dual σ factor preference

Ines S. C. Baptista[1,+], Vinodh Kandavalli[1,2,+], Vatsala Chauhan[1], Mohammed Bahurudeen[1], Bilena Almeida[1], Cristina Palma[1], Suchintak Dash[1], and Andre S. Ribeiro[1,3,*]

[1] Laboratory of Biosystem Dynamics, Faculty of Medicine and Health Technology, Tampere University, Tampere, 33520, Finland.

[2] Department of Cell and Molecular Biology, Uppsala University, Uppsala, 752 37, Sweden.

[3] CA3 CTS/UNINOVA. Faculty of Science and Technology, NOVA University of Lisbon, Quinta da Torre, 2829-516, Caparica, Portugal.

[+] Equal contributions.

* To whom correspondence should be addressed. E-mail: andre.sanchesribeiro@tuni.fi.

Present Address: Andre S. Ribeiro, Arvo Ylpön katu 34, P.O. Box 100, 33014, Tampere University, Finland.

Abstract

Escherichia coli uses the ability of σ factors to recognize specific DNA sequences in order to quickly control large gene cohorts. While most genes respond to only one σ

factor, approximately 5% have dual σ factor preference. The ones in significant numbers are ‘σ^{70+38} genes’, responsive to σ^{70}, which controls housekeeping genes, as well as to σ^{38}, which activates genes during stationary growth and stresses. We show that σ^{70+38} genes are almost as upregulated in stationary growth as genes responsive to σ^{38} alone. Also, their response strengths to σ^{38} are predictable from their promoter sequences. Next, we propose a sequence- and σ^{38} level-dependent, analytical model of σ^{70+38} genes applicable in the exponential, stationary, and in the transition period between the two growth phases. Finally, we propose a general model, applicable to other σ factors as well. This model can guide the design of synthetic circuits with sequence-dependent sensitivity and plasticity to transitions between the exponential and stationary growth phases.

Author Summary

Present challenges in Synthetic Biology include the design of genetic circuits that are robust to growth phase transitions and whose responsiveness is sequence-dependent, and, thus predictable prior to design. We present and validate an empirical-based, sequence-dependent analytical model of *E. coli* genes with dual responsiveness to the regulators σ^{70} and σ^{38}. These genes, supported by our sequence-dependent model, could become building blocks for synthetic genetic circuits functional in both the exponential and the stationary growth phases.

1. Introduction

In *Escherichia coli (E. coli*), genes are expressed by RNA polymerase (RNAP), which has 5 subunits ($\alpha_2\beta\beta'\omega$) and can synthesize RNA by starting and ending at sequence-specific DNA locations, with the former being known as 'promoter' regions [1]. Transcription is constantly regulated, mostly at the promoter regions, which typically harbor transcription factor (TF) binding sites (TFBS) and other regulatory sequence motifs [2–7]. Such regulation is an essential survival skill of adaptability to transitions in both internal as well as external conditions [8,9]

One mechanism of large-scale, relatively synchronous gene expression regulation is executed by σ factors [4,10–13]. *E. coli* has seven different σ factors [14]. During exponential growth in optimal conditions, RNAP mostly transcribes genes with preference for σ^{70}, responsible for basic cell functions [15]. Other σ factors are expressed under specific stresses [16], and activate a specific gene cohort. For example, after growing exponentially at the cost of environment components, *E. coli* usually switches to stationary growth. This is triggered, among other, by RpoS (σ^{38}), whose appearance activates ~10% of the genome [17,18], which leads to key phenotypic modifications [18–24].

During this adaptation, the concentration of RNAP remains relatively constant [25] and, since the number of RNAP core enzymes is limited, σ factors compete for them [10,12,14,20,26,27]. Thus, when σ^{38} numbers increase, not only are the genes responsive to σ^{38} activated, but also other genes, previously active, are indirectly negatively regulated, due to the reduction in RNAPs carrying σ^{70} [10,26,28].

For this regulatory system to be efficient, promoters need to have high specificity to one and only one σ factor. Nevertheless, there is a small fraction of promoters that can recognize more than one σ factor [6,29,30]. The most common (84%) are the

promoters responsive to both σ^{70} and σ^{38} [31]. Other cases are too rare to obtain sufficient statistics on their kinetics (Table S3 in S3 Appendix).

Here we investigate how the sequences of promoters recognized by both σ^{70} and σ^{38} relate to the dynamics of genes that they control (here named 'σ^{70+38} genes'). From flow cytometry and RNA-seq measurements, along with sequence analysis, we establish a relationship between RNA and protein levels in each growth phase with the promoter sequence (logos of positions -41 to -1 shown in Fig S1 in S2 Appendix). From this, we propose a model of the dynamics σ^{70+38} genes that fits their behavior prior, during, and following the transition to stationary growth. The model is dependent on both the promoter-sequence as well as the levels of σ factors in the cell.

2. Results

2.1. RNA fold changes when shifting to stationary growth

To study how σ^{70+38} genes respond to stationary growth, we first identified when are cells in exponential and stationary growth. We used both wild type (WT, control) cells and a MGmCherry strain [19] carrying fluorescently tagged σ^{38} to measure its levels prior to, during, and after shifting from exponential to stationary growth (Section 4.2). Cells at 150 min after being placed in fresh medium are used as representatives of cells in mid-exponential growth (Fig 1A). Meanwhile, at 500 min and onwards, cells exhibit stationary growth (Fig 1A).

From the measurements, σ^{38} levels are small at 150 min, while from 500 min onwards they are stably high (Fig 1B). Considering that the cell areas are ~15% smaller in the

stationary phase (Figs S3A and S3B in S2 Appendix), at 700 min the absolute concentration of σ^{38} is ~32 times higher than at 150 min (Fig 1C and Fig S3C in S2 Appendix). Meanwhile, RNAP concentration is largely stable (only ~7% higher during stationary growth) (Fig 1C and Fig S3D in S2 Appendix).

Next, we measured changes in RNA levels, by RNA-seq, between moments 150 min and 500 minutes by calculating the $\log_2$ fold changes in RNA levels (LFC_{RNA}) (Sections 4.2 and 4.4). Gaussian fits to the LFC_{RNA} distributions (Fig 1D) of σ^{70+38} genes, σ^{70} genes, and σ^{38} genes, as well as all other genes, have high coefficient of determination (R^2) values of the fits suggesting that they capture well the shapes of the empirical distributions.

In general, σ^{70+38} genes are nearly as upregulated as σ^{38} genes (Fig 1D). On the other hand, σ^{70} genes are weakly downregulated, likely due to indirect negative regulation caused by the increase in σ^{38} levels combined with the limited pool of RNAP core enzymes [10,12,14,20,26,27]. Finally, most other genes (~2737 out of the 4308 genes) are relatively unresponsive.

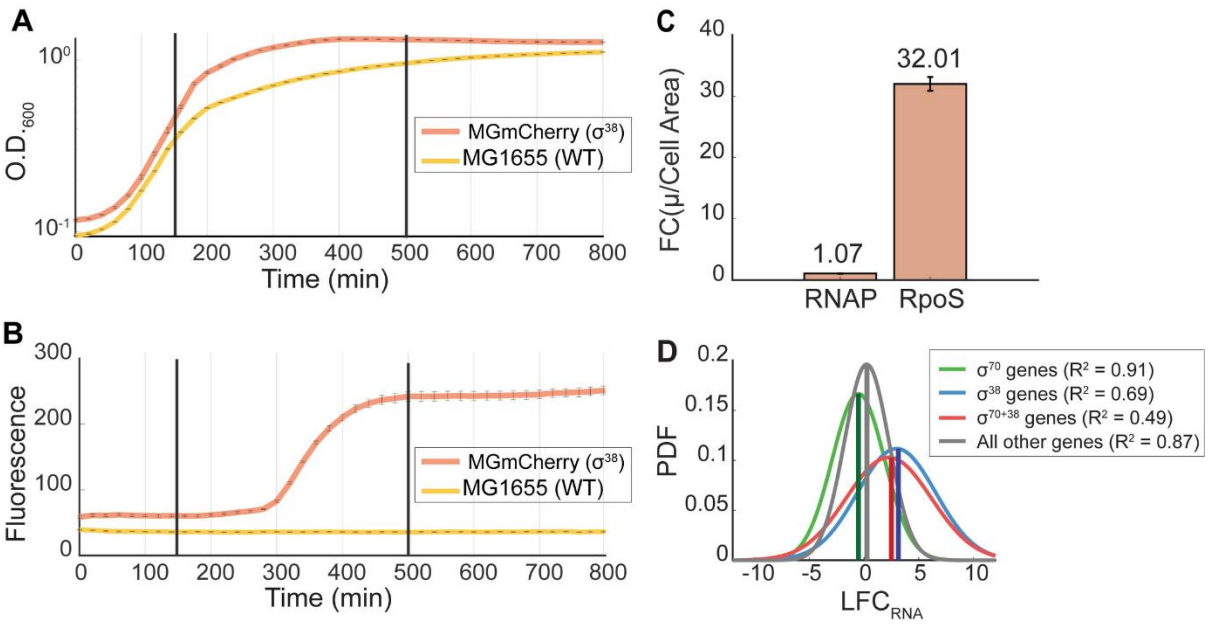

Fig 1. Cell growth, RNAP, RpoS and genome-wide RNA levels when changing from exponential to stationary growth. (A) Optical densities 'O.D.$_{600}$' of wild-type MG1655 (control) and MGmCherry (*rpoS::mcherry*) strains (Section 4.1). **(B)** Mean population fluorescence of WT and MGmCherry strains. Black vertical bars show the timing of measurements in exponential (150) and stationary (500 minutes onwards) growth. At 0 minutes, cells were moved to fresh medium (Section 4.2). **(C)** Fold-change between the exponential and stationary growth phases of mean cellular fluorescence relative to cell area (FC(μ/Cell Area)) due to changes in RNAP-GFP and in RpoS-mCherry, respectively. Error bars are the standard error of the mean (SEM). **(D)** Gaussian fits to the distributions of LFC_{RNA} of gene cohorts. Vertical lines mark the mean**.**

2.2. Propagation of shifts in RNA numbers to protein numbers

We used a YFP strain library to measure single-cell protein levels (Section 4.5) of 9 out of the 64 σ^{70+38} genes (only 15 of the 64 are represented in this library and some exhibited too weak signals to be detected by the flow cytometer). These 9 genes, according to their LFC_{RNA}, cover well the state space of response strengths of σ^{70+38} genes (Section S1.2 in S1 Appendix).

From the single-cell distributions of protein levels in exponential and stationary growth, we extracted means, μ_P, and squared coefficient of variations, CV^2_P. We also calculated the $\log_2$ fold changes in μ_P, LFC_P. Since LFC_P and the corresponding LFC_{RNA} are linearly positively correlated, with a high R^2 (Fig 2A), we conclude that changes in RNA levels of σ^{70+38} genes during the growth phase transition propagate to their protein levels.

We then analyzed the noise in the dynamics of these genes, prior and after the growth phase transition. In Figs 2B1 and 2B2, CV^2_P decreases quickly with μ_P for small μ_P,

but only weakly for high μ_P. This is well described by a function of the form: $CV_P^2 = C \times \mu_P^{-1} + N$ [32] (where *N* is the noise floor, a lower bound on the cell-to-cell variability of protein levels in clonal populations due to extrinsic factors [33]), in agreement with past reports [32,34,35].

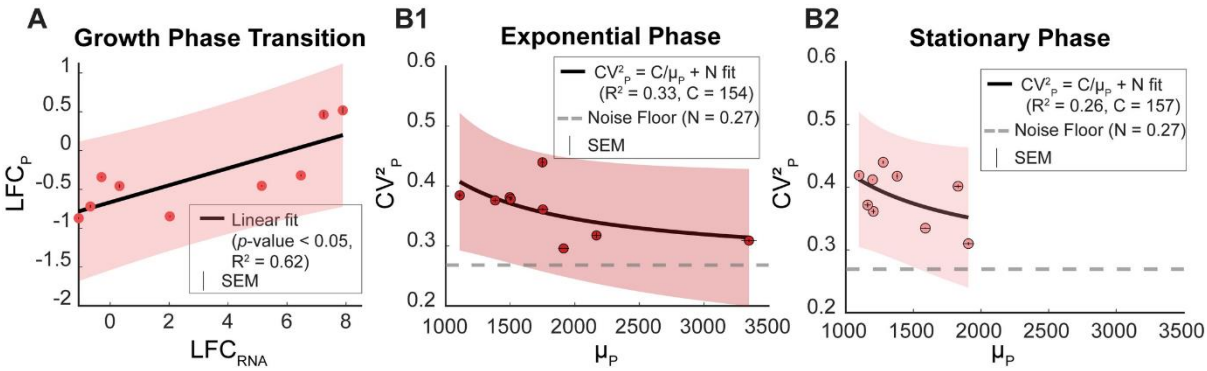

Fig 2. Single-cell protein levels of σ^{70+38} genes in the exponential and stationary growth phases. (A) Log_2 fold-change of protein levels, LFC_P, plotted against the corresponding LFC_{RNA}, during the shifting from exponential to stationary growth. **(B1)** and **(B2)** CV^2_P plotted against μ_P of σ^{70+38} genes in the exponential and stationary growth phases, respectively. Error bars (small) are the standard error of mean. All figures show the best fitting curves and their 95% prediction bounds (shadow areas). Horizontal dashed lines are the noise floors.

2.3. Full sequence-dependent model of the shift in the transcription dynamics of σ^{70+38} genes

Here, we present a sequence-dependent model of transcription of σ^{70+38} genes. Based on past works, we assume that only σ^{70} is present in high numbers (in holoenzyme form) in the exponential phase and that only σ^{38} increases significantly in numbers when shifting to stationary growth [10,14,20,26,36–38].

The model therefore accounts for the competition between σ^{70} and σ^{38} for binding to RNAP core enzymes, since these exist in limited numbers in *E. coli*

[10,12,14,20,26,27] and since σ factor numbers differ between the two growth phases. For this, we set reactions for binding and unbinding of σ^{70} and σ^{38} to free floating RNAP core enzymes, (R1a) and (R1b), where $K^{\sigma 70}$ and $K^{\sigma 38}$ are the ratios between the association and dissociation rate constants:

$$\mathrm{RNAP} + \sigma^{70} \xleftrightarrow{K^{\sigma 70}} \mathrm{RNAP}.\sigma^{70} \qquad \text{(R1a)}$$

$$\mathrm{RNAP} + \sigma^{38} \xleftrightarrow{K^{\sigma 38}} \mathrm{RNAP}.\sigma^{38} \qquad \text{(R1b)}$$

Given (R1a) and (R1b), RNA production dynamics depends on the ratio between σ^{70} and σ^{38} numbers if, and only if, $K^{\sigma 70}$ and $K^{\sigma 38}$ differ. Meanwhile, the limited pool of RNAPs is accounted for by setting:

$$\left[RNAP_{total}\right] = \left[RNAP.\sigma^{70}\right] + \left[RNAP.\sigma^{38}\right]. \qquad (1)$$

Next, to introduce the dual responsiveness to σ^{70} and σ^{38}, we define two competing, sequence-dependent reactions of transcription, (R2a) and (R2b), differing in which holoenzyme binds to the promoter. Further, to also introduce sequence dependence, we define two sequence-based distances: $D_{\sigma 38}$ and $D_{\sigma 70}$. $D_{\sigma 38}$ is the distance, in number of nucleotides, of a promoter sequence differing from the consensus (average) sequence of promoters with σ^{38} dependency. $D_{\sigma 70}$ is the distance in number of nucleotides differing from the consensus sequence of promoters with σ^{70} dependency (Section 4.8.3):

$$\mathrm{Pro} + \mathrm{RNAP}.\sigma^{70} \xrightarrow{k_t^{\sigma 70}(D_{\sigma 70})} \mathrm{Pro} + \mathrm{RNA} + \mathrm{RNAP} + \sigma^{70} \qquad \text{(R2a)}$$

$$\mathrm{Pro} + \mathrm{RNAP}.\sigma^{38} \xrightarrow{k_t^{\sigma 38}(D^{\sigma 38})} \mathrm{Pro} + \mathrm{RNA} + \mathrm{RNAP} + \sigma^{38} \qquad \text{(R2b)}$$

The rate constants of these reactions, which control the affinities to the holoenzymes, are expected to differ between genes, as they depend on the promoter sequences (and potentially which transcription factors are acting on the promoters, here not represented for simplicity). Nevertheless, here, also for simplicity, $k_t^{\sigma 70}$ and $k_t^{\sigma 38}$ are assumed to be a function of only $D_{\sigma 70}$ and $D_{\sigma 38}$ respectively.

Together, the set of reactions (R1a), (R1b), (R2a) and (R2b), model the transcription kinetics of σ^{70+38} genes before, during and, after shifting from exponential to stationary growth. The rates $k_t^{\sigma 70}(D_{\sigma 70})$ and $k_t^{\sigma 38}(D_{\sigma 38})$ are dissected below.

2.4. Reduced model of the shift in transcription kinetics

To reduce the model, we assume that the numbers of RNAP.σ^{70} and RNAP.σ^{38} in the cells are significantly larger than 1, which is expected. Also, we assume that the rate-limiting steps in transcription initiation (during which a promoter is occupied and thus new events cannot start [39,40]) are shorter in time-length than the intervals between transcription events. If so, for any σ^{70+38} gene:

$$\text{Pro} \xrightarrow{\left[RNAP.\sigma^{70}\right]\cdot k_t^{\sigma 70}(D_{\sigma 70})} \text{Pro} + \text{RNA} + \text{RNAP} + \sigma^{70} \qquad \text{(R3a)}$$

$$\text{Pro} \xrightarrow{\left[RNAP.\sigma^{38}\right]\cdot k_t^{\sigma 38}(D_{\sigma 38})} \text{Pro} + \text{RNA} + \text{RNAP} + \sigma^{38} \qquad \text{(R3b)}$$

We then merge (R3a) and (R3b) into a single transcription process:

$$\text{Pro} \xrightarrow{\left[RNAP.\sigma^{70}\right]\cdot k_t^{\sigma 70}(D_{\sigma 70})+\left[RNAP.\sigma^{38}\right]\cdot k_t^{\sigma 38}(D_{\sigma 38})} \text{Pro} + \text{RNA} \qquad \text{(R4)}$$

Finally, to complete the model, we included a reaction for translation of RNAs into proteins (R5), as well as events of RNA (R6) and protein (R7) decay due to degradation and dilution by cell division:

$$\mathrm{RNA} + \mathrm{Rib} \xrightarrow{k_{tr}} \mathrm{RNA} + \mathrm{Rib} + \mathrm{P} \qquad \text{(R5)}$$

$$\mathrm{RNA} \xrightarrow{\gamma_{RNA}} \varnothing \qquad \text{(R6)}$$

$$\mathrm{P} \xrightarrow{\gamma_{P}} \varnothing \qquad \text{(R7)}$$

2.5. Analytical solution of the reduced model

Based on the reduced model, we obtained an analytical solution for the expected fold change in RNA numbers of a gene whose promoter has preference for both σ^{70} as well as σ^{38}. From (R4), (R5), (R6), and (R7) we derived the expected RNA numbers produced by a σ^{70+38} gene:

$$RNA = \frac{\left[RNAP.\sigma^{70}\right] \cdot k_t^{\sigma 70}(D_{\sigma 70}) + \left[RNAP.\sigma^{38}\right] \cdot k_t^{\sigma 38}(D_{\sigma 38})}{\gamma_{RNA}} \qquad (2)$$

For simplicity, the rate constants are not expected to differ between the growth phases (as they depend on biophysical parameters that should not change significantly, such as binding affinities, etc.). Consequently, the fold-change in RNA numbers between the two growth phases should equal:

$$FC_{RNA} = \frac{RNA_{\mathrm{sta}}}{RNA_{\mathrm{exp}}} = \frac{\left[RNAP.\sigma^{70}\right]_{\mathrm{sta}} \cdot k_t^{\sigma 70}(D_{\sigma 70}) + \left[RNAP.\sigma^{38}\right]_{\mathrm{sta}} \cdot k_t^{\sigma 38}(D_{\sigma 38})}{\left[RNAP.\sigma^{70}\right]_{\mathrm{exp}} \cdot k_t^{\sigma 70}(D_{\sigma 70}) + \left[RNAP.\sigma^{38}\right]_{\mathrm{exp}} \cdot k_t^{\sigma 38}(D_{\sigma 38})} \cdot \Gamma \qquad (3)$$

where γ is the ratio between the RNA decay rates in the two growth phases:

$$\Gamma = \frac{\gamma_{RNA_{\mathrm{exp}}}}{\gamma_{RNA_{\mathrm{sta}}}} \qquad (4)$$

Since $[RNAP_{total}]$ is similar in the two growth phases (Fig 1C):

$$FC_{RNA} = \frac{\frac{\left[RNAP.\sigma^{70}\right]_{\text{sta}}}{\left[RNAP_{total}\right]} \cdot k_t^{\sigma 70}(D_{\sigma 70}) + \frac{\left[RNAP.\sigma^{38}\right]_{\text{sta}}}{\left[RNAP_{total}\right]} \cdot k_t^{\sigma 38}(D_{\sigma 38})}{\frac{\left[RNAP.\sigma^{70}\right]_{\text{exp}}}{\left[RNAP_{total}\right]} \cdot k_t^{\sigma 70}(D_{\sigma 70}) + \frac{\left[RNAP.\sigma^{38}\right]_{\text{exp}}}{\left[RNAP_{total}\right]} \cdot k_t^{\sigma 38}(D_{\sigma 38})} \Gamma \quad (5)$$

Also, considering equation (1), then:

$$\frac{\left[RNAP.\sigma^{70}\right]}{\left[RNAP_{total}\right]} + \frac{\left[RNAP.\sigma^{38}\right]}{\left[RNAP_{total}\right]} \approx 1 \quad (6)$$

For simplicity, let ρ_{exp} and ρ_{sta} be:

$$\rho_{\text{exp}} = \frac{\left[RNAP.\sigma^{38}\right]_{\text{exp}}}{\left[RNAP_{total}\right]} = \frac{\left[RNAP_{total}\right] \cdot \left[\sigma^{38}\right]_{\text{exp}} \cdot K^{\sigma 38}}{\left[RNAP_{total}\right] \cdot \left(\left[\sigma^{70}\right]_{\text{exp}} \cdot K^{\sigma 70} + \left[\sigma^{38}\right]_{\text{exp}} \cdot K^{\sigma 38}\right)} \quad (7a)$$

$$\rho_{sta} = \frac{\left[RNAP.\sigma^{38}\right]_{\text{sta}}}{\left[RNAP_{total}\right]} = \frac{\left[RNAP_{total}\right] \cdot \left[\sigma^{38}\right]_{\text{sta}} \cdot K^{\sigma 38}}{\left[RNAP_{total}\right] \cdot \left(\left[\sigma^{70}\right]_{\text{sta}} \cdot K^{\sigma 70} + \left[\sigma^{38}\right]_{\text{sta}} \cdot K^{\sigma 38}\right)} \quad (7b)$$

From (5):

$$FC_{RNA} = \frac{(1-\rho_{\text{sta}}) \cdot k_t^{\sigma 70}(D_{\sigma 70}) + \rho_{\text{sta}} \cdot k_t^{\sigma 38}(D_{\sigma 38})}{(1-\rho_{\text{exp}}) \cdot k_t^{\sigma 70}(D_{\sigma 70}) + \rho_{\text{exp}} \cdot k_t^{\sigma 38}(D_{\sigma 38})} \cdot \Gamma \quad (8)$$

Finally, since, in the exponential growth phase, $\left[RNAP.\sigma^{70}\right] >> \left[RNAP.\sigma^{38}\right]$, then $\rho_{\text{exp}} \approx 0$. Thus, (8) can be simplified:

$$FC_{RNA} = \gamma \cdot (1-\rho_{\text{sta}}) + \rho_{\text{sta}} \cdot \Gamma \cdot \frac{k_t^{\sigma 38}(D_{\sigma 38})}{k_t^{\sigma 70}(D_{\sigma 70})} \quad (9)$$

All parameters in (9) can be measured directly, except $k_t^{\sigma 70}(D_{\sigma 70})$ and $k_t^{\sigma 38}(D_{\sigma 38})$. Meanwhile, in the stationary phase: $\left[\sigma^{38}\right]^{\text{sta}} = 0.3 \cdot \left[\sigma^{70}\right]$ [20,26,37]. Also, $K^{\sigma 70} = 5K^{\sigma 38}$ [41]. Thus:

$$\rho^{sta} = \frac{\left[RNAP.\sigma^{38}\right]_{\text{sta}}}{\left[RNAP_{total}\right]} \approx \frac{0.3 \cdot \left[\sigma^{70}\right]_{\text{sta}} \cdot \frac{K^{\sigma 70}}{5}}{\left[\sigma^{70}\right]_{\text{sta}} \cdot K^{\sigma 70} + 0.3 \cdot \left[\sigma^{70}\right]_{\text{sta}} \cdot \frac{K^{\sigma 70}}{5}} = 0.0566 \qquad (10)$$

2.6. Promoter sequence affects the expression of σ^{70+38} genes

To model the influence of the promoter sequence (i.e., $D_{\sigma 70}$ and $D_{\sigma 38}$) on the response strength of σ^{70+38} genes to σ^{38}, we studied three models. In all, we assumed the basal levels of transcription rates, '$k_{t0}^{\sigma 70}$' and '$k_{t0}^{\sigma 38}$'. The overall transcription rates are tuned by a function *f* (Table 1) of the influence of $D_{\sigma 70}$ and $D_{\sigma 38}$ on the basal transcription rate:

$$k_t^{\sigma 70}(D_{\sigma 70}) = k_{t0}^{\sigma 70} \cdot f\left(D_{\sigma 70}\right) \qquad (11a)$$

$$k_t^{\sigma 38}(D_{\sigma 38}) = k_{t0}^{\sigma 38} \cdot f\left(D_{\sigma 38}\right) \qquad (11b)$$

Consider that, in general, as $D_{\sigma i}$ of a promoter increases, the promoter should become more distinct from the 'average' promoter with preference for σ^i. Consequently, its affinity to σ^i is expected to decrease. If this holds true, then the promoter consensus sequence of genes with preference for a given σ factor should have strong affinity to that σ factor. We hypothesize, for simplicity, it has the strongest affinity. If true, it follows that as $D_{\sigma 38}$ increases, the transcription rate by RNAP.σ^{38} should decrease.

Having this in account, to model how a promoter sequence controls the transcription kinetics of promoters recognized by σ^{70} and by σ^{38}, we considered a linear model

(model I in Table 1), a rational model (model II), where the transcription rate is inversely proportional to $D_{\sigma i}$, and an exponential model (model II), similar to the one in [6,29,30] for σ^{70} genes.

Table 1. Models of how a sequence's p-distance, D_{σ}, affects a promoter's response to changes in the concentration of its preferred σ factor. Here, *m* is an empirical-based constant.

I. Linear model	$f(D_{\sigma i}) = 1 - m_i \cdot D_{\sigma i}$, $m_i > 0$ & $m_i \times D_{\sigma i} < 1$, *i* = 38 or 70
II. Rational model	$f(D_{\sigma i}) = \frac{1}{1 + m_i \cdot D_{\sigma i}}$, $m_i > 0$, *i* = 38 or 70
III. Exponential model	$f(D_{\sigma i}) = e^{-m_i \cdot D_{\sigma i}}$, $m_i > 0$, *i* = 38 or 70

We fitted the models (Fig 3) to the empirical fold changes (FC) of σ^{70+38} genes (FC_{RNA}) with known input TFs (Table S4 and Table S6) (FC rather than LFC were used, because this allowed for simpler equations for the fitting surfaces). From the R^2 values of the best fitting models (Table 2), we find that the model III best describes how $D_{\sigma 70}$ and $D_{\sigma 38}$ affect $k_t^{\sigma 70}$ and $k_t^{\sigma 38}$.

Next, we validated the fitted models by quantifying how well they predict the dynamics (Fig 3) of σ^{70+38} genes without input TFs (Table S4 and Table S6). Given the R^2 values, we conclude that the tuned model III predicts well their dynamics. Further, the fact that the model obtained from genes with TF regulation fits well (with a higher R^2) the genes without TF regulation is suggestive that, in our measurement conditions, such TF regulation is not exerting a significant role.

Table 2. Goodness of fit, measured by the R^2 of the surface fitting of FC_{RNA} as a function of $D_{\sigma 38}$ and $D_{\sigma 70}$. Shown are the models and best fitting parameter values, where $K_0 = \frac{k_{t0}^{\sigma 38}}{k_{t0}^{\sigma 70}}$. The surfaces were fitted to σ^{70+38} genes without input TFs and validated on genes σ^{70+38} genes with input TFs.

I. Linear model	Surface Equation	$FC_{RNA} = \gamma \cdot (1-\rho_{\text{sta}}) + \gamma \cdot \rho_{\text{sta}} \cdot K_0 \cdot \left(\frac{1-m_{38} \times D_{\sigma 38}}{1-m_{70} \times D_{\sigma 70}} \right)$ (12a)		
	Coefficients	m_{38}	-834.80	
		m_{70}	-1.01×10^{4}	
		Γ	85.60	
		K_0	2.39×10^{-4}	
		ρ_{sta}	0.06 (Equation (10))	
	R^2	σ^{70+38} genes with input TFs		<0
		σ^{70+38} genes without input TFs		<0
II. Rational model	Surface Equation	$FC_{RNA} = \gamma \cdot (1-\rho_{\text{sta}}) + \gamma \cdot \rho_{\text{sta}} \cdot K_0 \cdot \left(\frac{1+m_{70} \cdot D_{\sigma 70}}{1+m_{38} \cdot D_{\sigma 38}} \right)$ (12b)		
	Coefficients	m_{38}	-0.38	
		m_{70}	26.51	
		Γ	0.21	
		K_0	241.70	
		ρ_{sta}	0.06 (Equation (10))	
	R^2	σ^{70+38} genes with input TFs		0.04

		σ^{70+38} genes without input TFs	0.00
III. Exponential model	Surface Equation	$FC_{RNA} = \gamma \cdot (1-\rho_{sta}) + \gamma \cdot \rho_{sta} \cdot K_0 \cdot e^{-(m_{38} \cdot D_{\sigma 38} - m_{70} \cdot D_{\sigma 70})}$ (12c)	
	Coefficients	m_{38}	30.78
		m_{70}	56.90
		Γ	28.20
		K_0	3.841×10^{-9}
		ρ_{sta}	0.06 (Equation (10))
	R^2	σ^{70+38} genes with input TFs	0.55
		σ^{70+38} genes without input TFs	0.98

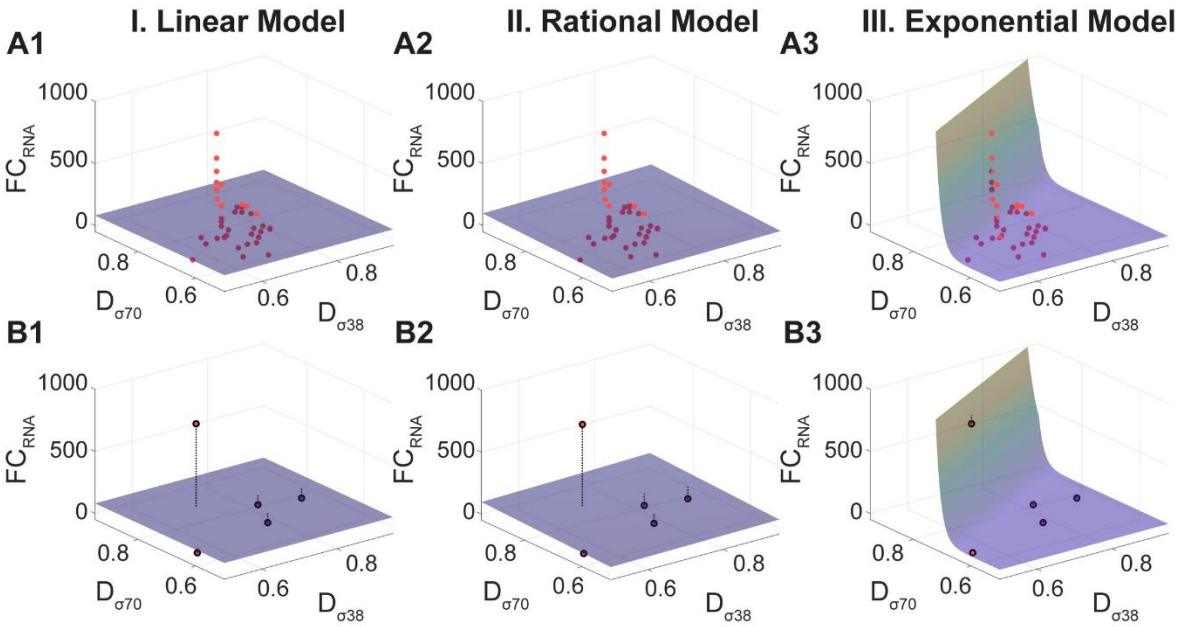

Fig 3. Fold changes in RNA levels of σ^{70+38} genes plotted against their promoter sequence *p*-distances to the consensus sequence of σ^{70} dependent promoters ($D_{\sigma 70}$) and of σ^{38} dependent promoters ($D_{\sigma 38}$). (A1-A3) show the best fitting surfaces to FC(μ_{RNA})

assuming **(A1)** a linear function, **(A2)** a rational function, and **(A3)** an exponentially decreasing function. Only σ^{70+38} genes with input TFs and FDR < 0.05 are included. Light red points are above the surface, while dark ones are below. **(B1-B3)** Same plots but the surfaces are those obtained from (A1-A3) and applied to σ^{70+38} genes without input TFs and FDR < 0.05. The dashed black lines depict the vertical distances between the estimated and measured FC(μ_{RNA}).

Finally, from Fig 2A, the LFC of protein numbers (LFC_P) correlates linearly with LFC_{RNA}. The same is predictable from the model (reactions R5 to R7) (Section S1.3 in S1 Appendix). Here, to obtain the best fitting model for protein log2 fold changes, we fitted a line to the data in Fig 2A, and extracted from it a scaling factor, α, between LFC_{RNA} and LFC_P. From this:

$$LFC_P = \log_2\left[\Gamma\cdot(1-\rho_{\text{sta}})+\Gamma\cdot\rho_{\text{sta}}\cdot K_0\cdot e^{-(m_{38}\cdot D_{\sigma 38}-m_{70}\cdot D_{\sigma 70})}\right]\cdot\alpha+\beta \qquad (13)$$

where β equals -0.67 and is the intercept between the y axis and the best fitting line, while α is the scaling factor between RNA and protein $\log_2$ fold changes and equals 0.1 (Fig 2A).

2.7. σ^{70+38} genes without known input transcription factors

Next, we fitted the exponential model III directly to σ^{70+38} genes without input TFs (not to mistake with when they were used to validate the surface created from σ^{70+38} genes with TF inputs). The best fitting model (R^2 of 1) has similar m_{38} and m_{70} (49.04 and 52.48, respectively). Thus, for this restricted set of genes, since $\Delta = D_{\sigma 38} - D_{\sigma 70}$ (Section 4.8.3), the surface equation (12c) in Table 2 can be approximated to:

$$FC_{RNA} \approx \Gamma\cdot(1-\rho_{\text{sta}})+\Gamma\cdot\rho_{\text{sta}}\cdot K_0\cdot e^{-r\times D} \quad \text{, where } r=\frac{m_{38}+m_{70}}{2}$$

Applying the $\log_2$ to both sides, one has:

$$LFC_{RNA} \approx \log_2\left(\Gamma\cdot(1-\rho_{\text{sta}})+\Gamma\cdot\rho_{\text{sta}}\cdot K_0\cdot e^{-r\times D}\right)=C_1+\log_2\left(\Gamma\cdot\rho_{\text{sta}}\cdot K_0\right)+\log_2\left(e^{-r\cdot D}\right)=C_1+C_2+\log_2\left(e^{-r\cdot D}\right)$$

If (C_1+C_2) is much smaller than $\log_2\left(e^{-r\cdot D}\right)$, we expect that

$$LFC_{RNA} \propto \log_2\left(e^{-r\cdot D}\right) \approx -\frac{r}{\ln(2)}\cdot D \qquad (14)$$

Therefore, and even though $D_{\sigma 38}$ and $D_{\sigma 70}$ are not significantly correlated (Fig 4A), one finds a linear correlation between LFC_{RNA} and Δ (Fig 4B).

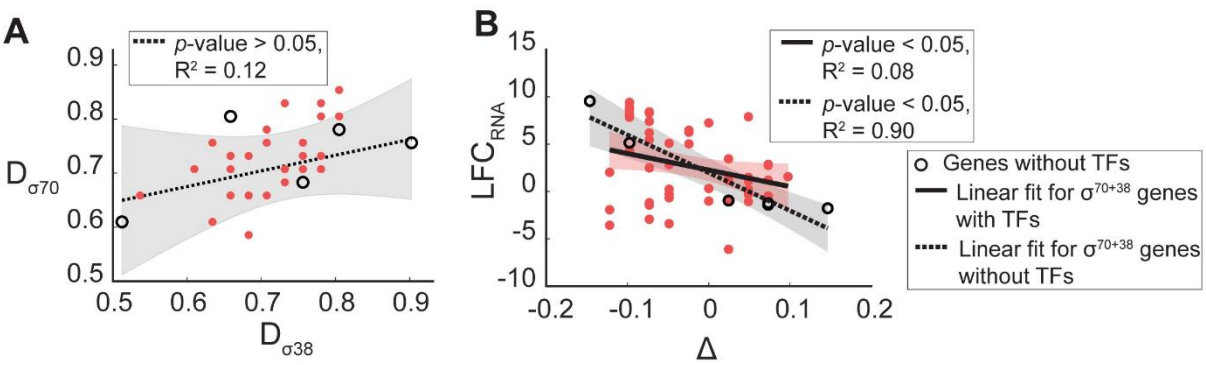

Fig 4. Log_2 fold changes in RNA levels of σ^{70+38} genes and their promoter *p*-distances to the consensus sequence of σ^{70} dependent promoters ($D_{\sigma 70}$) and of σ^{38} dependent promoters ($D_{\sigma 38}$). (A) Scatter plot of $D_{\sigma 38}$ plotted against $D_{\sigma 70}$. **(B)** Scatter plot of $\Delta = D_{\sigma 38} - D_{\sigma 70}$ plotted against LFC_{RNA}. The shadows of the best fitting lines are the 95% prediction bounds.

For comparison, we produced similar figures in search for linear correlations between Δ and LFC_{RNA} in the cohorts of genes whose promoters have preference for σ^{70}, σ^{38}, σ^{24}, σ^{28}, σ^{32}, and σ^{54} (Figs 5A-5F). No significant correlation with a high value of R^2 was found. We further studied the two small cohorts of genes whose promoters have preference for two σ factors other than σ^{70+38}. Here, correlations are found, but they

are not statistically significant, likely due to the small numbers of genes (Figs 5G and 5H).

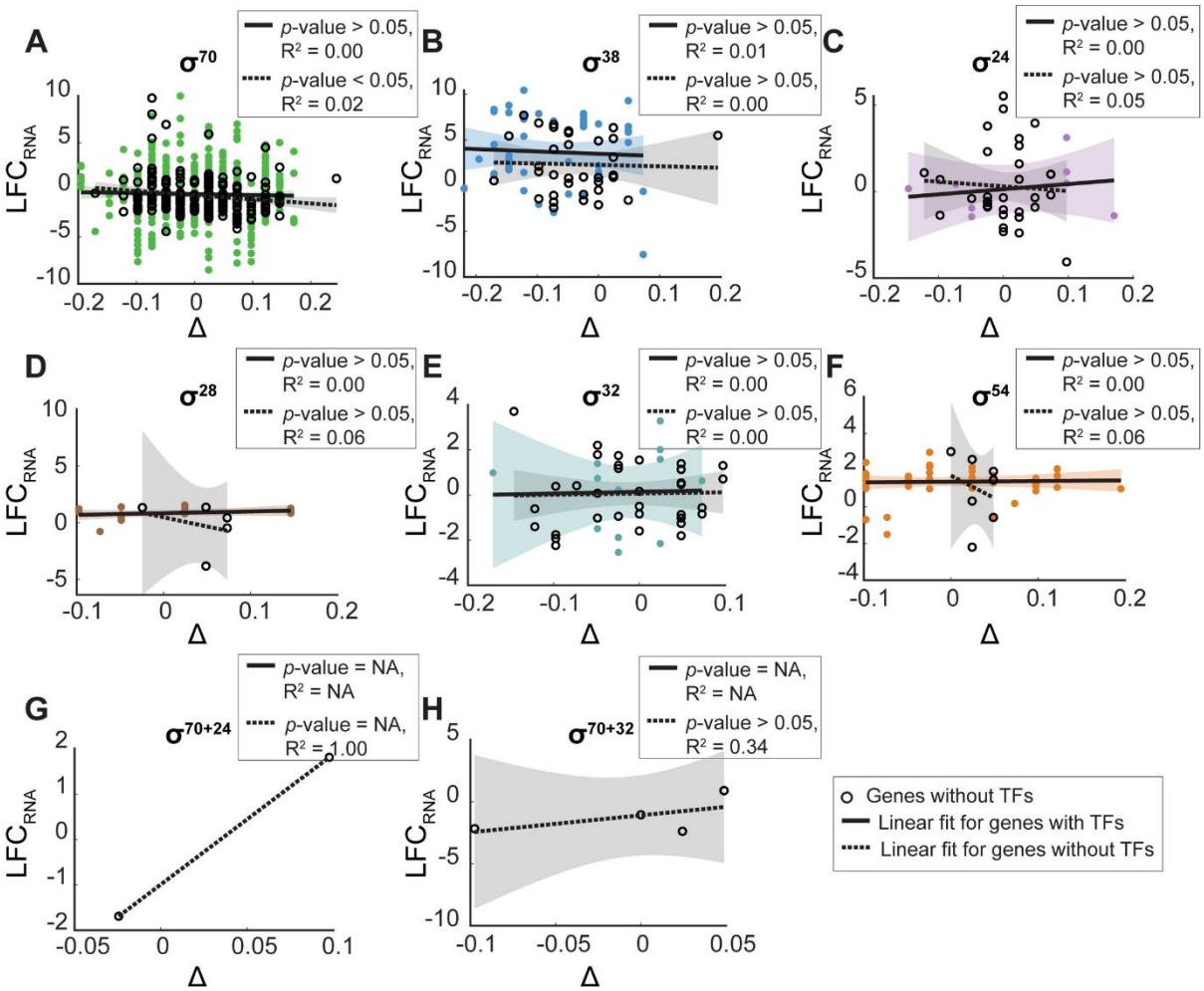

Fig 5. Fold changes in RNA levels and Δ. Genes whose promoters have preference for **(A)** σ^{70} **(B)** σ^{38} **(C)** σ^{24} **(D)** σ^{28} **(E)** σ^{32} **(F)** σ^{54} **(G)** σ^{70+24} and **(H)** σ^{70+32} are plotted against $\Delta = D_{\sigma 38} - D_{\sigma 70}$. The shadows of the best fitting lines are the 95% prediction bounds.

2.8. Expanding the model to the growth phase transition period

We expanded the model to include the transition period between the growth phases. For this, we collected temporal data on the protein numbers of three σ^{70+38} genes, specifically pstS, aidB and asr. These genes are used as representative of σ^{70+38}

genes with, strong, mild, and weak response strengths to the phase transition, respectively (Fig 6A). All data was well fitted by Hill functions (Fig 1B and Fig 6A). Moreover, in all three genes, there is a linear relationship between the fold changes in protein levels and in σ^{38} levels over time (Fig 6B).

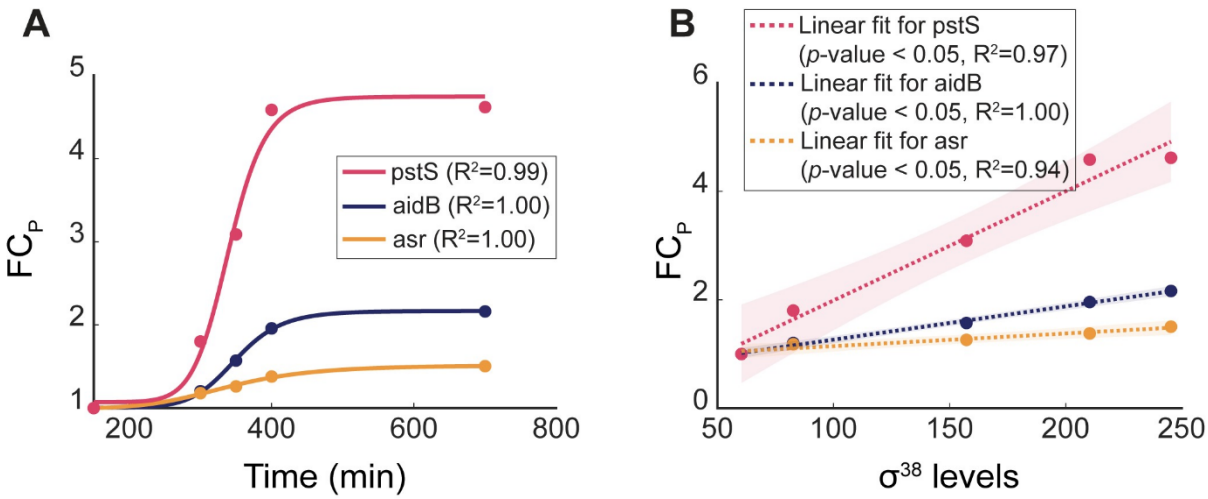

Fig 6. Temporal changes in the fold-change of protein levels as σ^{38} changes. (a) Protein levels of three σ^{70+38} genes prior to, during, and after the transition from exponential to stationary growth phases. The balls are the empirical mean fold changes (FC_P) in protein expression levels relative to the first moment. The lines are the corresponding best fitting Hill functions (parameter values in Table S5 in S3 Appendix). **(b)** Scatter plot of FC_P against the corresponding σ^{38} levels (data from Fig 1B) over time. The shadows are the 95% prediction bounds.

The dependency on σ^{38} levels is accounted for in model (ρ_{sta} is a function of σ^{38} levels in equation (7b), which are time dependent). Consequently, given the goodness of fit of the Hill functions, we propose the following time-dependent model:

$$FC_P(t) = \frac{\left(FC_P^{\max} - 1\right)\cdot t^s}{h^s + t^s} + b\,, \qquad (15a)$$

where:

$$LFC_P^{\max} = \left(\log_2\left[\Gamma\cdot(1-\rho_{sta}^{\max}) + \Gamma\cdot\rho_{sta}^{\max}\cdot K_0\cdot e^{-\left(m_{38}\cdot D_{\sigma 38} - m_{70}\cdot D_{\sigma 70}\right)}\right]\right)\cdot\alpha - \beta \qquad (15b)$$

$$FC_P^{\max} = \frac{\left[\Gamma\cdot(1-\rho_{sta}^{\max}) + \Gamma\cdot\rho_{sta}^{\max}\cdot K_0\cdot e^{-\left(m_{38}\cdot D_{\sigma 38} - m_{70}\cdot D_{\sigma 70}\right)}\right]^{\alpha}}{2^{\beta}} \qquad (15c)$$

Here, $\rho_{sta}^{\max}$ is the final (thus, maximum) concentration of RNAP.σ^{38} relative to the total concentration of bound RNAPs. Meanwhile, $FC_P^{\max}$ is the expected fold change in protein numbers of a σ^{70+38} gene, which is reached after the transition to stationary growth is complete. Finally, *b* (the intercept) is the expression level controlled by the promoter of interest, when in exponential growth phase. Meanwhile, *s* (the slope) is the corresponding response strength, and *h* (the half-activation coefficient) is a measure of response time to changes in σ^{38} levels.

2.9. Model generalization

While here we only found a correlation between the promoter sequence and the gene's response in σ^{70+38} genes, it may be that genes whose promoters have different σ preferences exhibit similar sequence-dependent behaviors during the stresses that they are responsive to. Thus, we generalized the model to be applicable to any stress and responsive gene cohort. As an example, we set a model for genes responsive to all seven σ factors of *E. coli*. For this, first, we expand reactions (R1a) and (R1b), to include all σ factors as follows:

$$\mathrm{RNAP} + \sigma^i \xleftrightarrow{K^{\sigma i}} \mathrm{RNAP}.\sigma^i \text{, where } i = 70, 54, 38, 32, 28, 24 \text{ or } 19 \qquad \text{(R8a-R8g)}$$

Next, we generalize eq. (1) to account for the σ factors in holoenzyme form:

$$\left[RNAP_{total}\right] = \sum_{i}\left[RNAP.\sigma^{i}\right], \text{ where } i = 70, 54, 38, 32, 28, 24 \text{ or } 19 \qquad (16)$$

Given this, we generalize (R4) as follows:

$$\text{Pro} \xrightarrow{\sum_{i}\left[RNAP.\sigma^{i}\right]\times k_{t}^{\sigma i}(D_{\sigma i})} \text{Pro} + \text{RNA}, \text{ where } i = 70, 54, 38, 32, 28, 24 \text{ or } 19 \qquad (R9)$$

This general model can be tuned based on the numbers of each σ factor present in the conditions considered, and the consensus sequences to each σ factor (Table S5). In addition, following the findings in Section 2.8, it should be feasible to introduce factors to account for the timing of the changes in the transition period.

3. Discussion

From information on the promoters' sequence and dynamics, we proposed and validated a sequence-dependent, kinetic model of genes controlled by promoters that are responsive to both σ^{70} and σ^{38}. This model of σ^{70+38} genes, which accurately predicts how their dynamics change for exponential and stationary growth, is an expansion of a past model of promoters with preference for σ^{70} alone [26], but has two competing reactions for transcription by RNAP when bound to σ^{70} and when bound to σ^{38}. In addition, these reactions' rate constants are sequence-dependent, in line with the hypothesis that the consensus sequence of promoters with preference for a given σ factor should provide the strongest affinity to that σ factor. The model also accounts for the period of the growth phase transition of the cell, which is linked to σ^{38} levels (from lowest to highest). Finally, we proposed (yet to validate) a general model, applicable to promoters with any given σ factor(s) preference.

Relevantly, we identified a simple correlation between the promoter sequence of σ^{70+38} genes and their response strength to the growth phase transition, making it predictable. So far, this predictability has not been reported for any cohort of natural promoters of *E. coli*, even in stable growth conditions. Similar relationships have only been reported for synthetic libraries of promoter variants under stable growth [5,6,30,42].

Since it remains challenging to predict if a sequence can act as a promoter and, if so, with which strength and under what regulatory mechanisms [5], this predictability of the dynamics from the sequence for the natural cohort of σ^{70+38} genes is of interest for three main reasons. First, these promoters and their variants (with varying p-distances from the consensus sequences of σ^{70} and σ^{38} genes) could become key components of a toolbox for engineering circuits whose response kinetics is predictable from the promoter sequences. Further, such circuits should be functional in both exponential and stationary growth phases, with tunable responsiveness to the phase transition. Finally, their model could be a starting point for broader models of sequence-predictable adaptability of natural promoters to various stresses.

It may prove difficult to find similar relationships between the sequences of other natural promoters and their responsiveness to other σ factors (or global regulators). E.g., we failed to find a relationship between the sequences of promoters responsive to σ^{38} alone and their response strength to σ^{38} (even for genes without TF inputs (Fig 5b)). Also, it remains to be proved whether promoters responsive to σ^{70}, σ^{54}, σ^{32}, σ^{28}, σ^{24}, or dual combinations, have similar behaviors when those σ factors change.

It also remains to be studied whether the sequence-dynamics relationship in σ^{70+38} genes is causal or coincidental. While the large numbers of promoters studied here support the causality, a definitive proof likely requires producing synthetic promoters similar to the natural ones but differing in p-distances to the consensus sequences of promoters with preference for σ^{38} and for σ^{70}, respectively.

Also, in the future, it should be investigated if the sequence-dependent responsiveness of σ^{70+38} genes to σ^{38} levels contributes to why their promoters (from positions -41 to -1) are highly conserved (Fig S1 in S2 Appendix) and/or why the TFs that they code for commonly serve as input TFs to essential genes [43] (2.5 times more than average, Fisher test *p*-value < 0.05).

Finally, over-representation tests of the ontology [44,45] of σ^{70+38} genes suggest that they are commonly involved in respiration (Fig S4 in S2 Appendix, Table S4 in S3 Appendix). In agreement, this process is highly affected when changing from exponential to stationary growth [23,46], since aerobic respiration is reduced, while fermentation and anaerobic respiration are enhanced [18]. Moreover, we observed that σ^{70+38} genes are amongst the most responsive to the change from growth phase out of the genes associated with these functions (Fig S4 in S2 Appendix). This suggests that they may control the processes that need most alterations during the adaptation. Therefore, externally regulating these genes may thus give significant control over these processes. This is particularly appealing since this control could be exerted by tuning their promoter sequences (i.e., their p-distances) making the effects of their integration on a genetic circuit's response to the growth phase shift largely predictable as well. This strategy may thus contribute to the engineering of synthetic circuits with tailored responses to growth phase transitions.

4. Materials and Methods

4.1. Bacterial strains and media

E. coli strains and plasmids are listed in Tables S1 and S2 in S3 Appendix. In short, we used YFP strains from the genetic stock center (CGSC) of Yale University, U.S.A [32] and, as support, a low-copy plasmid fusion library of fluorescent reporter strains using GFP to track promoter activity [47]. In both, one measures the levels of fluorescent proteins under the control of the promoters of our genes of interest which are a good proxy for the native protein levels [35][39]. For simplicity, we refer to these fluorescence levels as ‘protein levels’.

We used a RL1314 strain (rpoC::GFP) generously provided by Robert Landick, University of Wisconsin-Madison [48]), to measure RNA Polymerase levels. Their *rpoC* gene codes for β’ sub-unit endogenously tagged with GFP. Since *rpoC* codes for the β' subunit, a limiting factor in the assembly of RNAP holoenzyme [49,50], its numbers serve as a good proxy for RNAP numbers. For simplicity, [RNAP] refers to the sum of RNAP core and holoenzymes in a cell. Further, we used a MG*mCherry* (*rpoS*::mCherry) strain to measure RpoS levels (kind gift from James Locke [19]). Their *rpoS* gene codes for σ^{38} endogenously tagged with mCherry. Finally, we used wild-type K12 MG1655 strain for control.

We used M9 medium (1xM9 Salts, 2 mM MgSO4, 0.1 mM CaCl2; 5xM9 Salts 34 g/L Na2HPO4, 15 g/L KH2PO4, 2.5 g/L NaCl, 5 g/L NH4Cl) supplemented with 0.2% Casamino acids and 0.4% glucose, and Luria-Bertani (LB) medium with 10 g peptone,

10 g NaCl, and 5 g yeast extract in 1000 ml distilled water. We used the antibiotics kanamycin and chloramphenicol from Sigma Aldrich, U.S.A.

4.2. Growth rates and growth phases

Growth rates were measured by spectrophotometry (BioTek Synergy HTX Multi-Mode Microplate Reader). From a glycerol stock (-80 °C), cells were streaked on LB agar plates (2%) and incubated at 37 °C, overnight. Next, a single colony was picked, inoculated in LB medium with antibiotics (Section 4.1), and incubated at 30°C overnight with shaking. Overnight cultures were further diluted into fresh medium to an optical density of 600 nm ($O.D._{600}$) of 0.01 and incubated for growth by shaking at 250 rpm at 37°C. OD_{600} was recorded every 20 min, for 800 min. Cells were extracted at 150 min and at 700 min after inoculation into fresh medium to represent cells in exponential and stationary growth phases, respectively (Fig 1A).

4.3. Microscopy and Image analysis

To collect microscopy data, cells were placed between a coverslip and agarose gel pad (2%) and visualized by a confocal laser-scanning system, using a 100× objective. Green fluorescence images were captured using a 488 nm laser and a 514/30 nm emission filter. Phase-contrast images were simultaneously acquired for purposes of segmentation and to assess health, morphology, and physiology.

Images were analyzed using “CellAging” software [51]. After automatically segmenting cells and manually correcting errors, we applied 2D Gaussian filters to remove measurement noise, and extracted each cell’s total fluorescent intensity.

4.4. RNA-seq

4.4.1. Sample preparation

Cells were grown until reaching either the exponential or stationary growth phases, respectively. At each of these moments, they were treated with a double volume of RNA protect bacteria reagent (Qiagen, Germany) for 5 min, at room temperature, to prevent RNA degradation. Cells were then pelleted and frozen immediately at -80 ºC. After unfreezing the cells, total RNA was extracted using RNeasy mini-kit (Qiagen) according to the instructions. RNA was treated twice with DNase (Turbo DNA-free kit, Ambion, Life Technologies, U.S.A.) and quantified using Qubit 2.0 Fluorometer RNA assay (Invitrogen, Carlsbad, CA, USA). Quality of total RNA was determined by gel electrophoresis, using 1% agarose gel stained with SYBR safe (Invitrogen, U.S.A.). RNA was detected using UV with a Chemidoc XRS imager (BioRad, U.S.A.).

4.4.2. Sequencing

Sequencing was performed by GENEWIZ, Inc. (Leipzig, Germany). The RNA integrity number (RIN) of the samples was obtained with Agilent 4200 TapeStation (Agilent Technologies, Palo Alto, CA, USA). Ribosomal RNA depletion was performed using Ribo-Zero Gold Kit (Bacterial probe) (Illumina, San Diego, CA, USA). RNA-seq libraries were constructed using NEBNext Ultra RNA Library Prep Kit (NEB, Ipswich, MA, USA). Sequencing libraries were multiplexed and clustered on 1 lane of a flow-cell. Samples were sequenced using single-index, 2x150bp paired-end (PE) configuration on an Illumina HiSeq instrument. Image analysis and base calling were conducted with HiSeq Control Software (HCS). Raw sequence data (.bcl files) were converted into fastq files and de-multiplexed using Illumina bcl2fastq v.2.20. One mismatch was allowed for index sequence identification.

4.4.3. Data analysis

RNA-seq data analysis pipeline: i) RNA sequencing reads were trimmed to remove adapter sequences and nucleotides with poor quality with Trimmomatic [52] v.0.39. ii) Trimmed reads were then mapped to the reference genome (*E. coli* MG1655, NC_000913.3) using STAR v.2.5.2b aligner [53]. iii) *featureCounts* from Rsubread R package (v.1.34.7) was used to calculate unique gene hit counts [54]. Genes with less than 5 counts in more than 3 samples, and genes whose mean counts are less than 10, were removed. iv) Next, the unique gene hit counts were used for subsequent differential expression analysis, using DESeq2 R package (v.1.24.0) [55] to compare gene expression between samples and calculate p-values and log2 of fold changes (LFC) using Wald tests (function *nbinomWaldTest*). P-values were adjusted for multiple hypotheses testing (Benjamini–Hochberg, BH procedure [56]) and genes with adjusted p-values (False discovery rate (FDR)) smaller than 0.05 were classified as differentially expressed (D.E.).

Due to filtering (iii), only 783 (777 D.E.) out of the 931 σ^{70} genes, 87 (all D.E.) out of the 93 σ^{38} genes, 57 (56 D.E.) out of the 64 σ^{70+38} genes, and 3257 (2737 D.E.) out of 3607 remaining genes were assessed by RNA-seq. Other genes did not produce enough reads.

4.5. Flow cytometry

We used a ACEA NovoCyte Flow Cytometer equipped with yellow (561 nm) and blue lasers (488 nm), controlled by Novo Express (V1.50). Cells were diluted 1:10000 into 1 ml of PBS vortexed for 10 sec. For each gene and condition, we performed 3 biological replicates, acquiring 50,000 events in each. For GFP and YFP, we used the

blue laser for excitation and the FITC-H channel (520/20 nm filter) for emission. For mCherry, we used a yellow laser for excitation and PE-Texas Red (615/20 nm filter) for emission. We collected events at a flow rate of 14 µl/minute, a core diameter of 7.7 µM, and adjusted PMT voltage for each parameter. We have set the FSC-H detection threshold to 5000.

We collected single-cell fluorescence distributions. To remove outliers due to debris, cell doublets, and other undesired events, we performed gating by setting the maximum values of SSC-H to 5×10^4 and of FSC-H to 10^5. These limits allowed observing the whole distributions, without cutting their natural tails. The results did not differ widely from using unsupervised gating as in [57,58], if setting α to 0.95.

Finally, we discarded “far out” events using Tukey’s fences [59]. This had little effect on the results. Furthermore, background fluorescence was not subtracted since doing so did not affect the results (Section S1.2 in S1 Appendix and Fig S2 in S2 Appendix).

4.6. Spectrophotometry

Time-lapse protein fluorescence of MG*mCherry* cells was measured by a BioTek Synergy HTX Multi-Mode Microplate Reader. Overnight cultured cells were diluted 1:1000 times into fresh M9 medium, aliquoted into 24 well dark bottom microplates, and kept at 37°C with shaking of 250 rpm. Fluorescence intensities were recorded for 14 hours, every 20 min, using excitation (590/20 nm) and emission (645/20 nm) filters. Temporal profiles of fluorescence intensity were extracted using Gen5 software, based on 6 biological repeats.

4.7. Statistical analyses of the data

Data points from microscopy, flow cytometry, etc. are based on at least 3 biological replicates each. Meanwhile, to assess if genes measured by flow cytometry were a good representative of the set of all σ^{70+38} genes, we used two-sample Kolmogorov-Smirnov tests (KS tests) to compare their sequences ($D_{\sigma 38}$ and $D_{\sigma 70}$) and dynamics (LFC_{RNA}, μ, and CV^2). Further, gaussian fittings were made by the Distribution Fitter app (MATLAB). To evaluate its fitting, we calculated the coefficient of determination (R^2) between the PDF of the distributions and the PDF of the fitting. Finally, linear correlations were estimated by least-squares fitting (FITLM of MATLAB) and considered significant if the *p*-value of an *F*-test on the fitted regression line was smaller than 0.05. Other curves were fitted using MATLAB's curve fitting toolbox (apart from the fitting of Hill functions, which was done using the Hill function [60]).

As for gene ontology representations, we perform overrepresentation tests using PANTHER Classification [61] for finding significant overrepresentations by Fisher's exact tests. For *p*-values < 0.05, the null hypothesis that there are no associations between the gene cohort and the corresponding GO of the biological process is rejected. This *p*-value is corrected by calculating if the overall False Discovery Rate (FDR) is < 0.05. FDR is calculated by the Benjamini-Hochberg procedure [56].

4.8. Features of the promoter sequences

4.8.1. σ factor preference

Table S3 in S3 Appendix informs on the σ factor preference of the genes. From Regulon DB v10.5 (August 14, 2019), we obtained lists of all transcription units (TUs), promoters (including σ factor preferences), and genes of *E. coli* [62]. Recently, we compared our lists with information on July 1st, 2021 and found no changes that would

affect the conclusions. TUs only differed by ~1%, promoters by ~0.5%, and genes by ~1%.

From the 3548 TUs (gene(s) transcribed from a single promoter), we extracted 2179 with known promoters, containing 2713 genes in total. To minimize interference to the classification of σ factor preferences arising from transcription by multiple promoters, we narrowed our list to 1824 genes transcribed by only one promoter and, of those, to the 1328 genes with known σ factor preference. From those, 1242 have a preference for only one σ factor, including 931 with a preference for σ^{70}, and 93 with a preference for σ^{38}. Conversely, 76 genes have promoters with a preference for two σ factors. Out of these, 64 are transcribed by only one promoter with a preference for σ^{70} and σ^{38}.

4.8.2. Promoter sequence logos

Promoter sequence logos were created using WebLogo [63]. In each position (from -41 to -1), we counted how many times a nucleotide is present in all promoters considered. Then, we stacked all nucleotides (A, C, T, G) on top of each other, and sorted from the least found one in the bottom to the most present one on the top. For each position, we quantified its 'bit', as the difference between the maximum information possible (entropy given the 4 nucleotides) and the information considering the variability of the nucleotides (sum of the entropy for each of the 4 nucleotides) in that position (observed entropy): $\mathrm{Bit} = \log_2 4 - \sum_{n=1}^{4} f_n \times \log_2(f_n)$. From this, the height of each letter is set to be proportional to its frequency of occurrence. Finally, the height of each position was normalized, so as to equal its corresponding amount of information (with the more conserved positions having more bits) [64].

4.8.3. Promoter sequence p-distance

The p-distance D of a promoter [65] is the fraction of its nucleotides between positions -41 to -1 (assuming that transcription start site starts at position +1) that differ from the consensus (most common) nucleotide in that position of a cohort of genes (here, genes with preference for a given σ factor). We extracted the consensus sequences related to each σ from RegulonDB [62]. To measure the D of promoters with preference for both σ^{38} and σ^{70}, we calculated the p-distances $D_{\sigma 38}$ and $D_{\sigma 70}$, and then an overall p-distance, Δ, defined as: $\Delta = D_{\sigma 38} - D_{\sigma 70}$.

5. Supporting Information

S1 Appendix. Extended Methods and Materials. (Word)

S2 Appendix. Supporting Figure. (Word)

S3 Appendix. Supporting Tables. (Word)

S4 Table. List of genes classified as σ^{70}, and σ^{38} or as both σ^{70} and σ^{38} dependent (Excel)

S5 Table. Consensus sequences of promoters with preference for one σ factor. (Excel)

S6 Table. Fold changes in RNA levels of genes with a promoter with preference for σ^{70} and σ^{38}. Measurements by RNA-seq in the exponential and stationary growth phases. (Excel)

S7 Table. Statistics of single-cell distributions of fluorescence of cells measured by flow cytometry. (Excel)

6. References

1. Chen J, Chiu C, Gopalkrishnan S, Chen AY, Olinares PDB, Saecker RM, et al. Stepwise Promoter Melting by Bacterial RNA Polymerase. Mol Cell. 2020;78: 275-288.e6. doi:10.1016/j.molcel.2020.02.017

2. Browning DF, Busby SJW. Local and global regulation of transcription initiation in bacteria. Nat Rev Microbiol. 2016;14: 638–650. doi:10.1038/nrmicro.2016.103

3. Browning DF, Busby SJW. The regulation of bacterial transcription initiation. Nat Rev Microbiol. 2004;2: 57–65. doi:10.1038/nrmicro787

4. Cho BK, Kim D, Knight EM, Zengler K, Palsson BO. Genome-scale reconstruction of the sigma factor network in Escherichia coli: Topology and functional states. BMC Biol. 2014;12: 1–11. doi:10.1186/1741-7007-12-4

5. Urtecho G, Tripp AD, Insigne KD, Kim H, Kosuri S. Systematic Dissection of Sequence Elements Controlling σ70 Promoters Using a Genomically Encoded Multiplexed Reporter Assay in Escherichia coli. Biochemistry. 2019;58: 1539–1551. doi:10.1021/acs.biochem.7b01069

6. Einav T, Phillips R. How the avidity of polymerase binding to the –35/–10 promoter sites affects gene expression. Proceedings of the National Academy of Sciences. 2019;116: 13340 LP – 13345. doi:10.1073/pnas.1905615116

7. Jones DL, Brewster RC, Phillips R. Promoter architecture dictates cell-to-cell variability in gene expression. Science. 2014;346: 1533–1536. doi:10.1126/science.1255301

8. Stoebel DM, Hokamp K, Last MS, Dorman CJ. Compensatory evolution of gene regulation in response to stress by Escherichia coli lacking RpoS. PLoS Genet. 2009;5: e1000671. doi:10.1371/journal.pgen.1000671

9. Kannan G, Wilks JC, Fitzgerald DM, Jones BD, Bondurant SS, Slonczewski JL. Rapid acid treatment of Escherichia coli: transcriptomic response and recovery. BMC Microbiol. 2008;8: 37. doi:10.1186/1471-2180-8-37

10. Farewell A, Kvint K, Nyström T. Negative regulation by RpoS: A case of sigma factor competition. Mol Microbiol. 1998;29: 1039–1051. doi:10.1046/j.1365-2958.1998.00990.x

11. Rouvière PE, De Las Peñas A, Mecsas J, Lu CZ, Rudd KE, Gross CA. rpoE, the gene encoding the second heat-shock sigma factor, sigma E, in Escherichia coli. EMBO J. 1995;14: 1032–1042. doi:10.1002/j.1460-2075.1995.tb07084.x

12. Dong T, Schellhorn HE. Global effect of RpoS on gene expression in pathogenic Escherichia coli O157:H7 strain EDL933. BMC Genomics. 2009;10: 349. doi:10.1186/1471-2164-10-349

13. Rahman M, Hasan MR, Oba T, Shimizu K. Effect of rpoS gene knockout on the metabolism of Escherichia coli during exponential growth phase and early stationary phase based on gene expressions, enzyme activities and intracellular

metabolite concentrations. Biotechnol Bioeng. 2006;94: 585–595. doi:10.1002/bit.20858

14. Maeda H, Fujita N, Ishihama A. Competition among seven Escherichia coli σ subunits: Relative binding affinities to the core RNA polymerase. Nucleic Acids Res. 2000;28: 3497–3503. doi:10.1093/nar/28.18.3497

15. Feklístov A, Sharon BD, Darst SA, Gross CA. Bacterial sigma factors: A historical, structural, and genomic perspective. Annu Rev Microbiol. 2014;68: 357–376. doi:10.1146/annurev-micro-092412-155737

16. Helmann JD, Chamberlin MJ. Structure and function of bacterial sigma factors. Annu Rev Biochem. 1988;57: 839–872. doi:10.1146/annurev.bi.57.070188.004203

17. Hengge R. Proteolysis of σS (RpoS) and the general stress response in Escherichia coli. Research in Microbiology. 2009. pp. 667–676. doi:10.1016/j.resmic.2009.08.014

18. Weber H, Polen T, Heuveling J, Wendisch VF, Hengge R. Genome-wide analysis of the general stress response network in Escherichia coli: sigmaS-dependent genes, promoters, and sigma factor selectivity. J Bacteriol. 2005;187: 1591–1603. doi:10.1128/JB.187.5.1591-1603.2005

19. Patange O, Schwall C, Jones M, Villava C, Griffith DA, Phillips A, et al. Escherichia coli can survive stress by noisy growth modulation. Nat Commun. 2018;9. doi:10.1038/s41467-018-07702-z

20. Jishage M, Iwata A, Ueda S, Ishihama A. Regulation of RNA polymerase sigma subunit synthesis in Escherichia coli: Intracellular levels of four species of sigma subunit under various growth conditions. J Bacteriol. 1996;178: 5447–5451. doi:10.1128/jb.178.18.5447-5451.1996

21. Lange R, Hengge-Aronis R. The cellular concentration of the σ(S) subunit of RNA polymerase in Escherichia coli is controlled at the levels of transcription, translation, and protein stability. Genes and Development. 1994;8: 1600–1612. doi:10.1101/gad.8.13.1600

22. Gruber TM, Gross CA. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. Annu Rev Microbiol. 2003;57: 441–466. doi:10.1146/annurev.micro.57.030502.090913

23. Riedel TE, Berelson WM, Nealson KH, Finkel SE. Oxygen consumption rates of bacteria under nutrient-limited conditions. Appl Environ Microbiol. 2013;79: 4921–4931. doi:10.1128/AEM.00756-13

24. Abantika Ganguly DC. A Comparative Kinetic and Thermodynamic Perspective of the σ-Competition Model in Escherichia coli. Biophys J. 2012;103: 1325. doi:10.1016/j.bpj.2012.08.013

25. Dennis PP, Bremer H. Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. EcoSal Plus. 2008;3. doi:10.1128/ecosal.5.2.3

26. Kandavalli VK, Tran H, Ribeiro AS. Effects of σ factor competition are promoter initiation kinetics dependent. Biochim Biophys Acta. 2016;1859: 1281–1288. doi:10.1016/j.bbagrm.2016.07.011

27. Grigorova IL, Phleger NJ, Mutalik VK, Gross CA. Insights into transcriptional regulation and sigma competition from an equilibrium model of RNA polymerase binding to DNA. Proc Natl Acad Sci U S A. 2006;103: 5332–5337. doi:10.1073/pnas.0600828103

28. Mauri M, Klumpp S. A Model for Sigma Factor Competition in Bacterial Cells. PLoS Comput Biol. 2014;10: e1003845. doi:10.1371/journal.pcbi.1003845

29. Kinney JB, Murugan A, Callan CG, Cox EC. Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. Proc Natl Acad Sci U S A. 2010;107: 9158–9163. doi:10.1073/pnas.1004290107

30. Brewster RC, Jones DL, Phillips R. Tuning Promoter Strength through RNA Polymerase Binding Site Design in Escherichia coli. PLoS Comput Biol. 2012;8: e1002811. doi:10.1371/journal.pcbi.1002811

31. Tanaka K, Takayanagi Y, Fujita N, Ishihama A, Takahashi H. Heterogeneity of the principal sigma factor in Escherichia coli: the rpoS gene product, sigma 38, is a second principal sigma factor of RNA polymerase in stationary-phase Escherichia coli. Proc Natl Acad Sci U S A. 1993;90: 3511–3515. doi:10.1073/pnas.90.8.3511

32. Taniguchi Y, Choi PJ, Li G-W, Chen H, Babu M, Hearn J, et al. Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. Science. 2010;329: 533–538. doi:10.1126/science.1188308

33. Hausser J, Mayo A, Keren L, Alon U. Central dogma rates and the trade-off between precision and economy in gene expression. Nat Commun. 2019;10: 1–15. doi:10.1038/s41467-018-07391-8

34. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y, et al. Noise in protein expression scales with natural protein abundance. Nat Genet. 2006;38: 636–643. doi:10.1038/ng1807

35. Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, et al. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature. 2006;441: 840–846. doi:10.1038/nature04785

36. Ishihama A. Functional modulation of Escherichia coli RNA polymerase. Annu Rev Microbiol. 2000;54: 499–518. doi:10.1146/annurev.micro.54.1.499

37. Jishage M, Ishihama A. Regulation of RNA polymerase sigma subunit synthesis in Escherichia coli: intracellular levels of sigma 70 and sigma 38. J Bacteriol. 1995;177: 6832–6835. doi:10.1128/jb.177.23.6832-6835.1995

38. Piper SE, Mitchell JE, Lee DJ, Busby SJW. A global view of Escherichia coli Rsd protein and its interactions. Mol Biosyst. 2009;5: 1943–1947. doi:10.1039/B904955J

39. McClure WR. Mechanism and Control of Transcription Initiation in Prokaryotes. Annu Rev Biochem. 1985;54: 171–204. doi:10.1146/annurev.bi.54.070185.001131

40. Muthukrishnan AB, Martikainen A, Neeli-Venkata R, Ribeiro AS. In vivo transcription kinetics of a synthetic gene uninvolved in stress-response pathways in stressed Escherichia coli cells. PLoS One. 2014;9. doi:10.1371/journal.pone.0109005

41. Colland F, Fujita N, Ishihama A, Kolb A. The interaction between sigmaS, the stationary phase sigma factor, and the core enzyme of Escherichia coli RNA polymerase. Genes Cells. 2002;7: 233–247. doi:10.1046/j.1365-2443.2002.00517.x

42. Liu M, Tolstorukov M, Zhurkin V, Garges S, Adhya S. A Mutant Spacer Sequence between -35 and -10 Elements Makes the P_{lac} Promoter Hyperactive and cAMP Receptor Protein-Independent. Proc Natl Acad Sci U S A. 2004;101: 6911–6916. Available: http://www.jstor.org/stable/3372027

43. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol. 2006;2: 2006.0008-2006.0008. doi:10.1038/msb4100050

44. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25: 25–29. doi:10.1038/75556

45. Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 2021;49: D325–D334. doi:10.1093/nar/gkaa1113

46. Andersen KB, von Meyenburg K. Are growth rates of Escherichia coli in batch cultures limited by respiration? J Bacteriol. 1980;144: 114–123. doi:10.1128/JB.144.1.114-123.1980

47. Zaslaver A, Bren A, Ronen M, Itzkovitz S, Kikoin I, Shavit S, et al. A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. Nat Methods. 2006;3: 623–628. doi:10.1038/nmeth895

48. Bratton BP, Mooney RA, Weisshaar JC. Spatial distribution and diffusive motion of RNA polymerase in live Escherichia coli. J Bacteriol. 2011;193: 5138–5146. doi:10.1128/JB.00198-11

49. Izard J, Gomez Balderas CDC, Ropers D, Lacour S, Song X, Yang Y, et al. A synthetic growth switch based on controlled expression of RNA polymerase. Mol Syst Biol. 2015;11: 840. doi:10.15252/msb.20156382

50. Chamberlin MJ. The selectivity of transcription. Annu Rev Biochem. 1974;43: 721–775. doi:10.1146/annurev.bi.43.070174.003445

51. Häkkinen A, Muthukrishnan A-B, Mora A, Fonseca JM, Ribeiro AS. CellAging: a tool to study segregation and partitioning in division in cell lineages of Escherichia coli. Bioinformatics. 2013;29: 1708–1709. doi:10.1093/bioinformatics/btt194

52. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30: 2114. doi:10.1093/bioinformatics/btu170

53. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9: 357–359. doi:10.1038/nmeth.1923

54. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res. 2019;47: e47. doi:10.1093/nar/gkz114

55. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15: 550. doi:10.1186/s13059-014-0550-8

56. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Series B Stat Methodol. 1995;57: 289–300. Available: http://www.jstor.org/stable/2346101

57. Startceva S, Kandavalli VK, Visa A, Ribeiro AS. Regulation of asymmetries in the kinetics and protein numbers of bacterial gene expression. Biochimica et Biophysica Acta - Gene Regulatory Mechanisms. 2019;1862: 119–128. doi:10.1016/j.bbagrm.2018.12.005

58. Razo-Mejia M, Barnes SL, Belliveau NM, Chure G, Einav T, Lewis M, et al. Tuning Transcriptional Regulation through Signaling: A Predictive Theory of Allosteric Induction. Cell Systems. 2018;6: 456-469.e10. doi:10.1016/j.cels.2018.02.004

59. Tukey JW. Schematic Summaries. In: Mosteller F, editor. Exploratory Data Analysis. Reading, Massachusetts: Addison-Wesley Publishing Company; 1977. pp. 27–56.

60. Olsen T. Hill function. In: MATLAB Central File Exchange [Internet]. 27 Sep 2018 [cited 27 Oct 2021]. Available: https://www.mathworks.com/matlabcentral/fileexchange/68935-hill-function)

61. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. 2019;47: D419–D426. doi:10.1093/nar/gky1038

62. Santos-Zavaleta A, Salgado H, Gama-Castro S, Sánchez-Pérez M, Gómez-Romero L, Ledezma-Tejeida D, et al. RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in E. coli K-12. Nucleic Acids Res. 2018;47: D212–D220. doi:10.1093/nar/gky1077

63. Crooks G, Hon G, Chandonia J, Brenner S. NCBI GenBank FTP Site\nWebLogo: a sequence logo generator. Genome Res. 2004;14: 1188–1190. doi:10.1101/gr.849004.1

64. Schneider TD, Stephens RM. Sequence logos: A new way to display consensus sequences. Nucleic Acids Res. 1990;18: 6097–6100. doi:10.1093/nar/18.20.6097

65. Nei M, Zhang J. Evolutionary Distance: Estimation. Encyclopedia of Life Sciences. 2006; 1–4. doi:10.1038/npg.els.0005108