bioRxiv

ARTICLE PREPRINT

Promoters adopt distinct dynamic manifestations depending on transcription factor context

Anders S. Hansen[1,2*] and Christoph Zechner[3,4*]

[1]Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA; [2]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA; [3]Max Planck Institute of Molecular Cell Biology & Genetics, Germany [4]Center for Systems Biology Dresden, Germany; *corresponding authors, alphabetical order
Contact: anders.sejr.hansen@berkeley.edu and zechner@mpi-cbg.de

Cells respond to external signals and stresses by activating transcription factors (TF), which induce gene expression changes. Previous work suggests that signal-specific gene expression changes are partly achieved because *different* gene promoters exhibit varying induction dynamics in response to the *same* TF input signal. Here, using high-throughput quantitative single-cell measurements and a novel statistical method, we systematically analyzed transcription in individual cells to a large number of dynamic TF inputs. In particular, we quantified the scaling behavior among different transcriptional features extracted from the measured trajectories such as the gene activation delay or duration of promoter activity. Surprisingly, we found that even the *same* gene promoter can exhibit qualitatively distinct induction and scaling behaviors when exposed to different dynamic TF contexts. That is, promoters can adopt context-dependent "manifestations". Our analysis suggests that the full complexity of signal processing by genetic circuits may be significantly underestimated when studied in specific contexts only.

INTRODUCTION

Exquisite regulation of gene expression underlies essentially all biological processes, including the remarkable ability of a single cell to develop into a fully formed organism. Transcription factors (TFs) control gene expression by binding to the promoters of genes and recruiting chromatin remodelers and the general transcriptional machinery. Recruitment of RNA Polymerase II enables the initiation of transcription, which produces mRNAs that are exported to the cytoplasm, where they are finally translated into proteins by the ribosome. Gene expression is primarily regulated at the level of promoter switching dynamics and initiation of transcription in a highly stochastic manner[1]. For practical reasons, however, gene expression is typically analyzed at the level of mRNAs (e.g. FISH) or proteins (e.g. immunofluorescence or GFP-reporters) using bulk or single-cell approaches. Although powerful, these data provide only partial and indirect information about the underlying promoter states and transcription initiation dynamics. Moreover, although gene regulation is complex in both time (e.g. time-varying signals) and space (e.g. signaling gradients), experimental measurements tend to be limited to simple perturbations such as ON/OFF or dose-dependent responses under steady-state conditions.

Ideally, gene regulation should be studied at the level of promoter switching dynamics and transcription initiation events, using experimental approaches that capture gene expression in a sufficiently large number of cells in response to a broad range of dynamic inputs. Several elegant studies have addressed some, but not all, of these challenges[1–5]. Here, through an integrated experimental and computational approach, we make a first attempt to realize this goal. We focus on a simple feedback-free system, where a single inducible TF activates a target gene. Surprisingly, our approach reveals that even the simplest gene networks can display complex and counter-intuitive behaviors, which cannot be explained by simple kinetic models. In particular, we show that genes exhibit "context-dependent manifestations", such that the same gene can switch between qualitatively different kinetic behaviors depending on which dynamic input it is exposed to.

RESULTS

Inferring promoter induction dynamics from quantitative single-cell trajectories

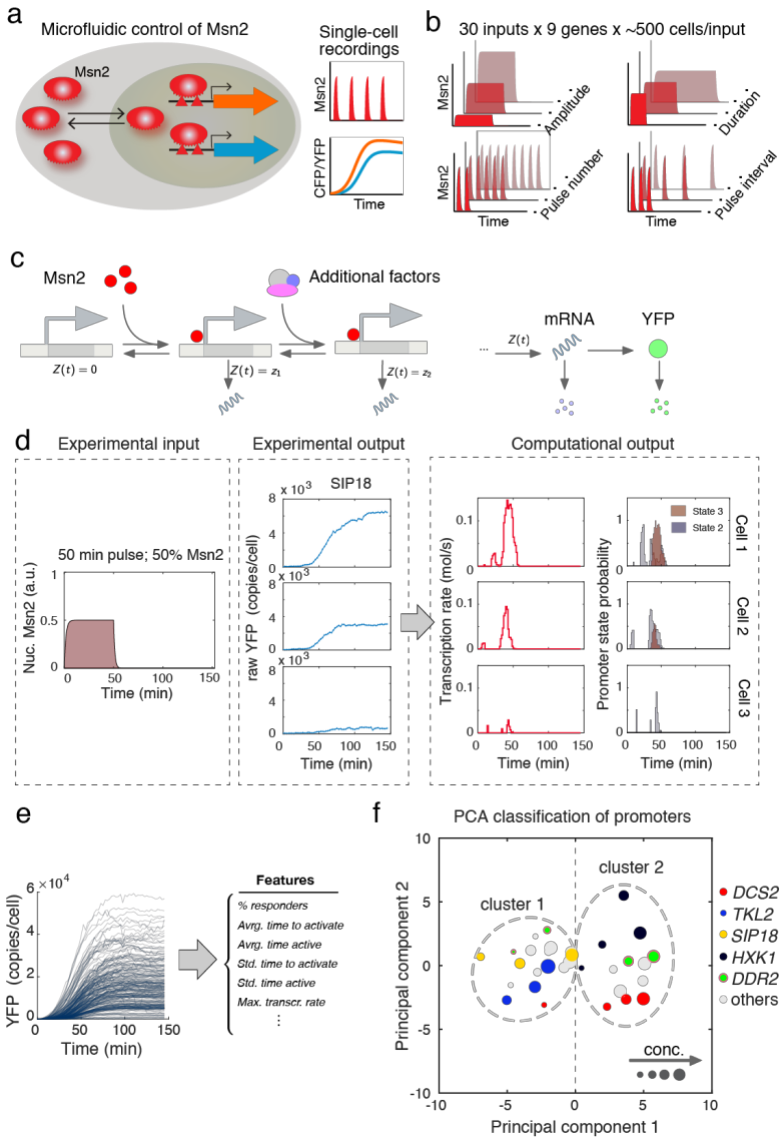


Figure 1. Overview of Msn2 system and inference approach. (**a**) Overview of microfluidic control of Msn2 activity and read-out of gene expression. (**b**) Overview of range of Msn2 input dynamics. **(c)** Stochastic model of gene expression. The promoter (left) can switch from its inactive state to its active state in an Msn2-dependent manner. Once active, mRNA can be transcribed at a certain rate z_1. Transcription can be further tuned by recruitment of additional factors, which is captured by a third state with distinct transcription rate z_2. Messenger RNA and protein dynamics are described as a two-stage birth-and-death process (right). A detailed description of the model can be found in Supplementary text S.1.1). **(d)** Statistical reconstruction of promoter switching and transcription dynamics. Gene expression output trajectories were quantified for diverse Msn2 inputs in a large number of cells. From the experimental inputs and outputs, transcription rates and promoter state occupancies were reconstructed using Bayesian inference (Box 1 and Supplementary Text S.1). **(e)** Determination of promoter features from single-cell data. Several features characterizing the promoter and transcription dynamics were calculated from the single-cell reconstructions for each promoter and experimental condition. **(f)** Clustering of promoters revealed by PCA (Supplementary text S.2.3).

To study how genes respond to complex and dynamic TF inputs, we focus on a large data-set that we previously generated (**Supplementary Fig. 1**)[4,6]. Here, addition of a small molecule causes the budding yeast TF, Msn2, to rapidly translocate to the nucleus and activate gene expression (**Fig. 1a**). Using microfluidics, rapid addition or removal of 1NM-PP1 allowed us to control both pulse length, pulse interval and pulse amplitude of the TF (fraction of Msn2 that is activated) and simultaneously measure the single-cell response of natural and mutant Msn2 target genes using fluorescent reporters[4,6–8] (**Fig. 1a**). We note that Msn2 naturally exhibits complex signal-dependent activation dynamics[3], that the system is not subject to feedback[3,4], that we replaced the target gene ORF with YFP and measured the endogenous gene response[4,6] and that the target genes are strictly Msn2-dependent[4]. Our extensive dataset contains 30 distinct dynamical Msn2 inputs for 9 genes (210 conditions) and ~500 cells per condition, numbering more than 100,000 single-cell trajectories in total (**Fig. 1b**).

Gene promoters can generally exist in a number of transcriptionally active and inactive states[9,10]. To understand the occupancy, dynamics and switching between these states, we used mathematical modeling in combination with our experimental data (**Fig. 1c**). Here we focus on a three-state promoter architecture (**Supplementary text S.1**), which was the simplest model that could recapitulate the single-cell responses of all promoters at individual Msn2 concentrations (**Supplementary text S.2.1**). This model accounts for Msn2-dependent activation of the promoter after which mRNA can be transcribed at a certain rate. Transcription can be further tuned (for instance by recruitment of additional factors), which is captured by a third state with distinct transcription rate (**Fig. 1c**).

This model in combination with a robust statistical approach allows us to determine the promoter switching- and transcription dynamics in individual cells across different promoters and conditions (see **Box 1**). However, existing inference approaches for trajectory data are limited in throughput and thus cannot handle extensive datasets like the one considered here[11–13]. We have therefore developed a novel recursive Bayesian inference method, which achieves accurate reconstructions while maintaining scalability (**Supplementary Text S.1**). The method processes single-cell trajectories and extracts from them time-varying transcription rates and promoter state occupancies (**Fig. 1d**). From the reconstructions, in turn, we computed a number of transcriptional features that summarize the single-cell expression dynamics of each promoter and condition (**Fig. 1e** and Supplementary text S.1.7). This combined experimental and computational approach allowed us to quantitatively compare different promoters under a wide range of Msn2 contexts.

To gain an overview of this high-dimensional dataset, we analyzed the gene expression responses to single pulses of nuclear Msn2 of different amplitudes (25%, 50%, 75% or 100%) for each promoter. Using Principal Component Analysis (PCA) to reduce the dimensionality (**Supplementary text S.2.3**), we uncovered the known categories of the promoters[4] for most conditions (**Fig. 1f**): slow activation, high amplitude threshold promoters (*SIP18, TKL2*) clustered together and fast activation, low amplitude threshold promoters (*HXK1*, *DCS2*) also clustered together. Surprisingly, however, *DDR2* (green; **Fig. 1f**) clustered with the slow, high threshold promoters at low Msn2 amplitudes (25%, 50%), but with the fast, low threshold promoters at high Msn2 amplitudes (75%, 100%). To explain this phenomenon, we introduce the concept of "context-dependent manifestations". Operationally, we define a context-dependent manifestation of a promoter as a situation where the same promoter exhibits qualitatively distinct kinetic behaviors under different input contexts. Here for instance, *DDR2* behaves like one promoter class at low Msn2 amplitudes, but a distinct class at high Msn2 amplitudes.

Promoters exhibit context-dependent scaling behaviors

Having established that a single gene promoter can exhibit distinct context-dependent manifestations, we next considered specific examples of this. We analyzed the scaling behavior among different transcriptional features under all input contexts. While certain promoters behaved consistently under all contexts, others exhibited complex and context-dependent scaling behaviors. In **Fig. 2a,** for instance, we plotted the time it took to activate the promoter and the time the promoter was active (**Fig. 2b**) against the maximal transcription rate for single pulse inputs for *DDR2*. Surprisingly, at low amplitude Msn2 input, the time it takes to activate *DDR2* for the first time increases with pulse length (**Fig. 2a**) without strongly changing the active duration (**Fig. 2b**) while the maximal transcription rate remained constant (**Fig. 2a**). In contrast, at high Msn2 amplitude, the time to activate is fixed at around 5 min, but now the maximal transcription rate increases with pulse duration. Thus, we observe qualitatively distinct scaling behaviors depending on Msn2 context (here, amplitude) and refer to this phenomenon as a context-dependent manifestation.

Having observed qualitatively distinct scaling behaviors for a natural promoter (*DDR2*), we next tested whether context-dependent behaviors are tunable by analyzing wild-type *SIP18* and two mutants[6] of *SIP18* (**Fig. 2c**).

BOX 1. Bayesian reconstruction of promoter switching dynamics from time-lapse data.

We developed an efficient Bayesian method to quantify promoter switching and transcription dynamics from time-lapse fluorescent reporter measurements (**Fig 1, Supplementary text S.1.1**). The dynamic state of the gene expression system at time t is denoted by $s(t) = (z(t), m(t), n(t))$, with $z(t)$ as the instantaneous transcription rate and $m(t)$ and $n(t)$ as the mRNA and YFP reporter copy numbers, respectively. The dynamics of $s(t)$ are described by a continuous-time Markov chain (CTMC) as described in **Supplementary Text S.1**. Let us denote by $s_{0:K} = \{s(t) \mid 0 \le t \le t_K\}$ a complete trajectory of $s(t)$ on the time interval $t \in [0, t_K]$. We assume that we can collect a sequence of K partial and noisy measurements $y_1, \dots, y_K$ at times $t_1 < t_2 < \cdots < t_K$ along the trajectory. The statistical relationship between the measurements and the underlying state of the system is captured by a measurement density $p(y_k \mid s_k)$ with $s_k = s(t_k)$ for all $k = 1, \dots, K$. In the scenario considered here, the measurements $y_1, \dots, y_K$ represent noisy readouts of the reporter copy number extracted from time-lapse fluorescence movies (Materials and Methods). In order to infer $s_{0:K}$ from a measured trajectory $y_1, \dots, y_K$, we employ Bayes' rule, which is stated as

$$p(s_{0:K} \mid y_1, \dots, y_K) \propto p(y_1, \dots, y_K \mid s_{0:K}) p(s_{0:K}) = \prod_{k=1}^{K} p(y_k \mid s_k)\, p(s_{0:K}),$$

with $p(s_{0:K})$ as the probability distribution over trajectories $s_{0:K}$. The latter is governed by the CTMC model of gene expression, which summarizes the statistical knowledge about the system that is available *prior* to performing the experiment. The corresponding posterior distribution on the left-hand side captures the knowledge about $s_{0:K}$ that we gain once we take into account the experimental measurements. However, analytical expressions to analyze the posterior distribution are generally lacking and one is typically left with numerical approaches. Sequential Monte Carlo (SMC) methods[22] have been successfully applied to address this problem in the context of time-lapse reporter measurements[11–13]. The core idea of these approaches is to generate a sufficiently large number of random sample paths $s_{0:K}^{(i)}$ consistent with the posterior distribution from eq. (3). This is performed sequentially over individual measurement time points, which allows splitting the overall sampling problem into a sequence of smaller ones that can be solved more efficiently (**Supplementary Text S.1.4**). The resulting SMC methods, however, are still computationally very expensive since the generation of an individual sample path $s_{0:K}^{(i)}$ can span thousands or even millions of chemical events when considered on realistic experimental time scales. In the Msn2 induction system, for instance, trajectories involve a large number of transcription and translation events, which would render conventional SMC approaches inefficient. To address this problem, we developed a hybrid SMC algorithm, in which only the promoter switching events have to simulated explicitly, while the transcription and translation dynamics are eliminated from the simulation. More precisely, the scheme targets the marginal posterior distribution

$$p(z_{0:k} \mid y_1, \dots, y_k) \propto p(y_1 \mid z_{0:1}) \prod_{k=2}^{K} p(y_k \mid y_1, \dots, y_{k-1}, z_{0:k})\; p(z_{0:K}),$$

where the mRNA and reporter dynamics $m(t)$ and $n(t)$ have been integrated out. We derived expressions for the marginal likelihood functions $p(y_k \mid y_1, \dots, y_{k-1}, z_{0:k})$ using an analytical approximation based on conditional moment equations (**Supplementary Text S.1.5**). Using this hybrid SMC algorithm, the sampling space could be significantly reduced, which makes the method scale to large datasets as the one considered in this study. Prior to applying the scheme to the dataset, the model parameters were inferred using a subset of the data (**Supplementary Text S.2.1** and **Supplementary Table 1**). A complete description of the method and additional details can be found in **Supplementary Text S.1** and **S.2**.

Specifically, we analyzed the mean number of state transitions between active and inactive promoter states for a single Msn2 pulse of variable duration (10, 20, 30, 40 or 50 min) as a function of the amplitude (25%, 50%, 75% or 100% Msn2 activation) and calculated the mean number of state transitions normalized by pulse duration. Within the context of a traditional "random-telegraph" model, high transcription is normally associated with a large number of state transitions[1,4]. Surprisingly however, *SIP18* exhibited an unusual "U-shaped" scaling where low and high Msn2 amplitudes were associated with fast switching kinetics, whereas intermediate Msn2 amplitudes showed ~2-3 fold slower switching kinetics (**Fig. 2d**). A simple model where the switching rates increase with Msn2 amplitude cannot explain this behavior. To further investigate this scaling behavior, we compared the behavior of the wild-type *SIP18* promoter to two mutants[6], A4 and D6 (**Fig. 2a**). While D6 (**Fig. 2f**) displayed the expected scaling behavior of monotonically increasing switching kinetics with increasing Msn2 amplitude, A4 (**Fig. 2e**) also showed unusual scaling behavior of "increasing→decreasing→increasing" switching kinetics with Msn2 concentration. In summary, these results demonstrate that a single promoter can exhibit context-dependent scaling behavior that cannot be explained by one simple model, and additionally, that only a few promoter mutations are sufficient to completely alter this behavior.

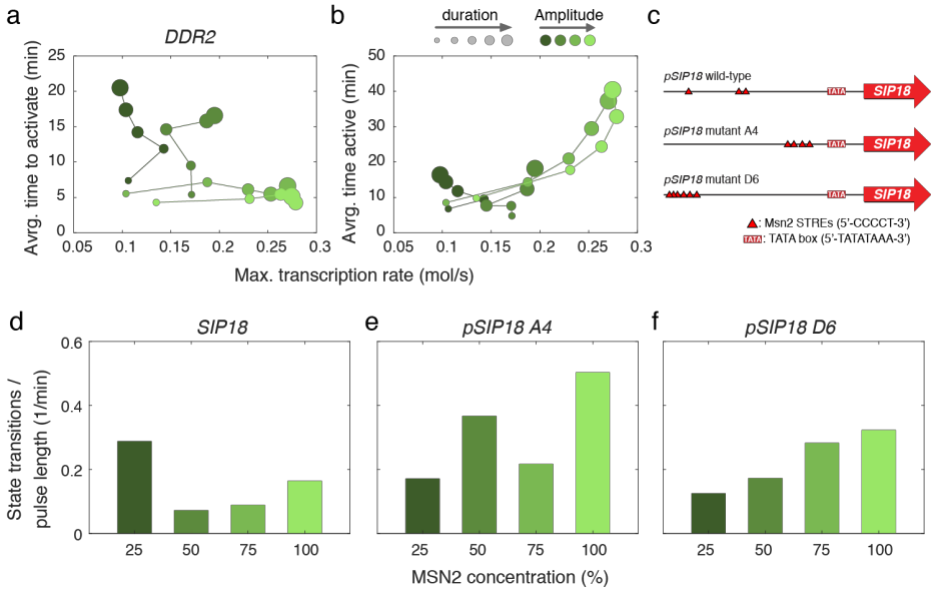


Figure 2: Context-dependent scaling behaviors. (**a**) Scaling behavior for *DDR2*. Scaling of time to activate (top) and total time active (bottom) for *DDR2* with the maximal transcription rate. Time to activate is defined as the time from when Msn2 enters the nucleus until the promoter converts into a transcriptionally active state. Similarly, time active is defined as the total duration the promoter is in a transcriptionally active state. Cells that never switched into a transcriptionally active state (non-responders) were excluded from the analysis. The maximal transcription rate is defined as the maximum of the average transcription rate calculated across the whole population. (**b**) Overview of wild-type and mutant *SIP18* promoters[6]. (**c-f**) Plots of the total number of state transitions normalized by pulse duration for 100% Msn2 input for wild-type (c), mutant A4 (d) and mutant D6 (f).

Context-dependent promoter manifestations control gene expression noise

Having analyzed population-averaged properties, we next asked whether manifestations exist at the single-cell level. We began by analyzing the correlation between transcriptional output (the total number of mRNA produced) and the time the promoter was active (**Fig. 3a**) for *TKL2* and *DCS2*. For *DCS2*, transcriptional output at the single-cell level is a nearly deterministic function of the time the promoter spends in an active state. To validate this, we performed a regression analysis and found that a simple linear model where transcriptional output is proportional to time spent in the active promoter states (with coefficient β) can explain nearly all the variation in transcriptional output (R^2~1; **Fig. 3a**). Thus, for a given Msn2 amplitude, the rate of *DCS2* transcription is fixed and the single-cell expression level is determined almost exclusively by the time the promoter is active. However, the rate of transcription is set by the Msn2 amplitude (i.e. β increases with Msn2 amplitude). Thus, *DCS2* is remarkably simple and regulation by time active and transcription rate can be decoupled.

In contrast, *TKL2* resembles *DCS2* at low and high Msn2 amplitudes, but at 50% Msn2 activation (**Fig. 3a**, blue), *TKL2* becomes highly stochastic and time active becomes a poor predictor of transcriptional output (R^2<<1; **Fig. 3a**). Thus, surprisingly, time active is an almost deterministic predictor of *TKL2* transcriptional output at low and high Msn2 amplitudes (R^2~1), but a fairly poor predictor at intermediate Msn2 amplitude (R^2<<1). This demonstrates that *TKL2* exhibits an unusual "inverse-U" "deterministic→stochastic→deterministic" scaling behavior and further highlights how promoters exhibit context-dependent single-cell manifestations, in this case dependent on the Msn2 amplitude.

The analysis above considered the single-cell responses to a single 50 min Msn2 pulse at different amplitudes. To generalize our analysis, we next analyzed how transcriptional noise (quantified using CV: std/mean) scales with mean transcriptional output under all conditions (**Fig. 3b**). As expected from previous studies[14,15], transcriptional noise uniformly decreases as transcriptional output increases for some genes such as *DCS2*. In contrast, *SIP18* exhibits the same "inverse-U" scaling as *TKL2*: low noise during low transcription, high noise during intermediate levels of transcription and again low noise during high levels of transcription (**Fig. 3c**). To further investigate the "inverse-U" scaling, we compared the behavior of the wild-type *SIP18* promoter to the two mutants A4 and D6[6] (**Fig. 2c**). Mutant A4 behaves like *DCS2,* demonstrating that only a few

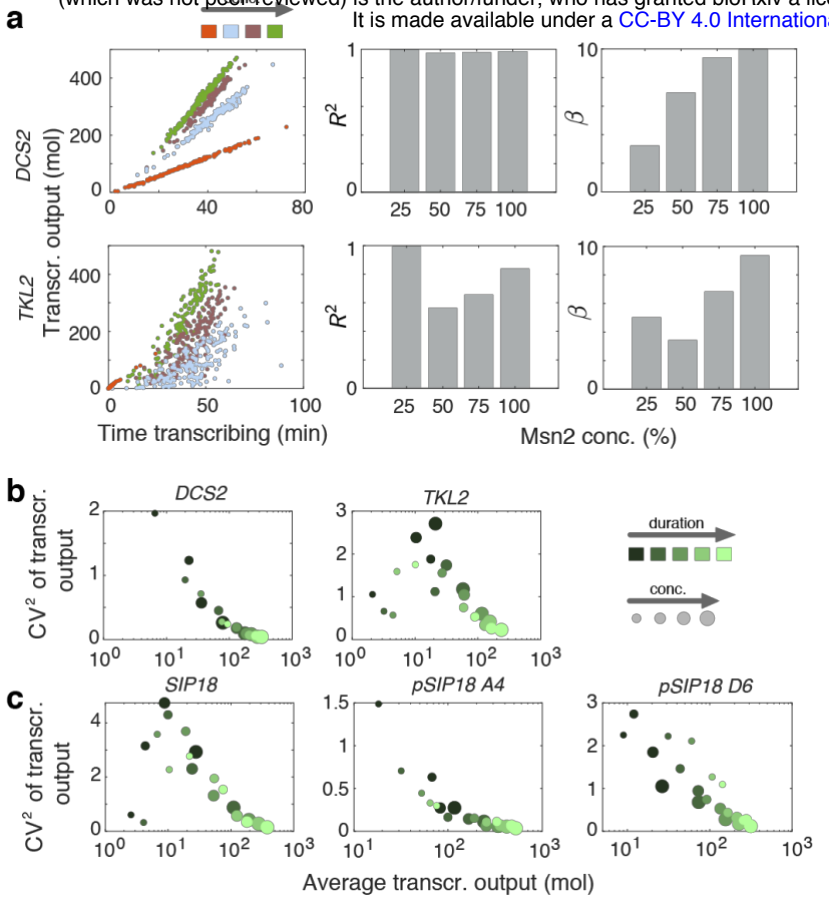


Figure 3: Single-cell manifestations control gene expression noise. **(a)** Scaling of transcriptional output with time transcribing, defined as the time the promoter spends in any of the two transcriptionally permissive states (see **Fig. 1c**). Transcriptional output is defined as the expected number of transcripts produced along the duration of the experiment. The left panel plots transcr. output against time transcribing for individual cells for a 50min pulse with 25%, 50%, 75% and 100% Msn2 input. Linear regression analysis was performed to determine the R^2 and slope β between transcr. output and time transcribing as shown in the center and right panels. **(b)** Scaling of noise with average transcriptional output for different pulse lengths and Msn2 induction levels. Noise is defined as the squared coefficient variation of the transcriptional output calculated across individual cells. **(c)** Like (**b**), but for wild-type and mutant *SIP18* promoters[6].

promoter mutations are sufficient to switch scaling and manifestation behavior.

Memory dependent promoter manifestations revealed by pulsatile Msn2 activation

We next analyzed how promoters respond to pulsatile Msn2 activation. Cells were exposed to four 5-min Msn2 pulses separated by 5, 7.5, 10, 15 or 20 min intervals. Some promoters behaved relatively simply, e.g. *DCS2* (**Fig. 4a**). Essentially all cells activate the *DCS2* promoter during the first pulse (**Fig. 4b**), and the promoter displays limited "positive memory" between pulses. By "positive memory", we refer to the fact that successive pulses of Msn2 activation increase the susceptibility of the promoter to become activated and induce higher gene expression. This has also been termed the "head-start" effect[3]. In contrast, the *SIP18* mutant D6 promoter[6] exhibited very curious behavior: at 5-min intervals (**Fig. 4b**, top row), there was significant positive memory and most cells only activated expression in the 2nd, 3rd or 4th pulse (**Fig. 4b**, top row). In contrast, with 20 min intervals, we observed "negative memory": there was much lower expression during pulse 2-4, than during pulse 1. In other words, exposure to 1 pulse of Msn2 inhibited transcription during subsequent pulses. Furthermore, comparing the different pulse intervals we observed a transition from positive memory at 5 and 7.5 min intervals to negative memory at 15 and 20 min intervals (**Fig. 4b**). While positive memory has previously been reported[3,4], a context-dependent switch from positive to negative memory has not. We note that no simple model can explain a sharp transition from positive to negative promoter memory and that this type of behavior only becomes visible once the response to diverse dynamic inputs are analyzed. Although the underlying molecular mechanism is unknown, we show in **Supplementary Fig. 2** a hypothetical toy model that could explain this switch from positive to negative memory. In conclusion, these data provide another example of how the same promoter can exhibit very different quantitative and qualitative behaviors depending on the context – in this case, depending on the interval between Msn2 pulses.

DISCUSSION

Here we quantitatively analyze the input-output relationship in a simple feedback-free system. We show that even under these relatively simple conditions, the same promoter can exhibit context-dependent scaling and induction behaviors. To describe this observation, we introduce the concept of context-dependent "manifestations". The underlying number of molecular states of a promoter is potentially enormous: when we measure a dose-response, we likely observe only certain rate-limiting regimes or manifestations of the system (**Fig. 4c**). What we show here is that the particular observed rate-limiting manifestation is highly context-dependent and very distinct quantitative behaviors can be observed under different contexts – even in systems that are seemingly simple. It appears likely that many other genes and pathways exhibit similar context-dependencies, when analyzed at high resolution and under a wide range of experimental conditions.

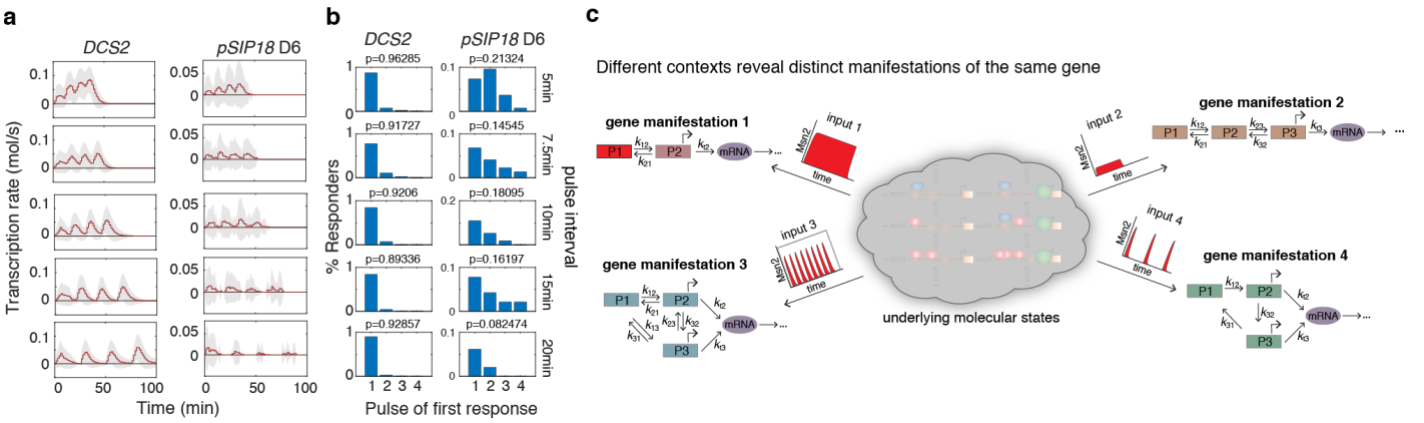


Figure 4. Promoter memory and model. (a-b) Interval-dependent regulation of promoter memory. Single-cell reconstructions are shown for *DCS2* and *SIP18* mutant D6 for the pulse-sequence experiments in which cells where treated with four consecutive Msn2 pulses (75% induction level) with 5-minute duration. The interval between the pulses were chosen to be 5, 7.5, 10, 15 and 20min. (**a)** Reconstructed transcription rates across all cells. Solid lines correspond to population means and shaded areas mark one standard deviation above and below the mean. **(b)** Distribution of responding cells across individual pulses. The histograms count for each pulse *i* the percentage of cells that had their first activation event during pulse *i*. The total percentage of responding cells *p* is provided for each histogram. **(c)** Model of context-dependent promoter manifestations.

Our results have two important implications. First, system identification efforts based on limited sets of experimental conditions within complex systems are unlikely to be successful in the sense of capturing the full range of behaviors of the underlying molecular pathways. In extreme cases, we may arrive at very different and possibly contradictory conclusions about a pathway's inner working depending on which experimental context we choose. The only solution to this problem is to resort to experimental and computational approaches that capture a pathway's response to a sufficiently broad range of dynamic experimental contexts. Much more work on simple systems will be necessary to truly understand the relevant complexity of signal processing in cells.

Second, a major conundrum in quantitative biology has been how to reconcile the remarkable spatiotemporal precision of biological systems with the high degree of gene expression noise observed at the single-cell level[16,17]. For example, when information transduction capacities have been measured for simple pathways, such systems appear to be barely capable of reliable distinguishing ON from OFF (~1 bit)[18–21]. Since these studies were done under strict experimental conditions, it is likely that they captured only one out of multiple manifestations. Our results suggest that if all physiologically relevant manifestations could be captured, the estimated information transduction capacity of biochemical pathways could be substantially greater than previously estimated. This may, in part, explain the remarkable signal processing capacity of biological systems.

MATERIALS AND METHODS; SUPPLEMENTARY FIGURES

Overview of experiments and source data

The data in concentration units of arbitrary fluorescence were previously described[4,6]. Here, we performed absolute abundance quantification as previously described[23] to convert the data to absolute numbers of YFP and CFP proteins per cell. All the source data supporting this manuscript are freely available together with a detailed ReadMe file at https://zenodo.org/record/2755026 .

Microfluidics and time-lapse microscope

Since the unnormalized data was previously acquired, here we only briefly describe the experimental methods. Microfluidic devices were constructed as previously described[4]. We furthermore refer the reader to a detailed protocol describing how to construct microfluidic devices and computer code for controlling the solenoid valves[8]. Briefly, for microscopy experiments, diploid yeast cells were grown overnight at 30°C with shaking at 180 RPM to an $OD_{600\ nm}$ of ca. 0.1 in low fluorescence medium without leucine and tryptophan, quickly collected by suction filtration and loaded into the five channels of a microfluidic device pretreated with concanavalin A (4 mg/mL). The setup was mounted on an inverted fluorescence microscope kept at 30°C. The microscope automatically maintains focus and acquires phase-contrast, YFP, CFP, RFP and iRFP images from each of five microfluidic channels for 64 frames with a 2.5 min time resolution corresponding to imaging from -5 min to 152.5 min. Solenoid valves control delivery of 1-NM-PP1 to each microfluidic channel. For full details on the range of input conditions, please see source data at https://zenodo.org/record/2755026 .

Image analysis and YFP quantification and normalization

Time-lapse movies were analyzed using custom-written software (MATLAB) that automatically segments yeast cells

based on phase-contrast images and tracks cells between frames. The image analysis software and a protocol describing how to use it is available elsewhere[8]. The arbitrary fluorescence units were converted to absolute abundances by comparing fluorescence to strains with known absolute abundances and by segmenting the cell to calculate the total number of YFP molecules per cell per timepoint as described previously[23].

Quantification of nuclear Msn2 dynamics

Msn2 was visualized as an Msn2-mCherry fusion protein. This allows accurate quantification of the nuclear concentration of Msn2 over time (Msn2 only activates gene expression when nuclear) as previously described[3,4,6]. From the resulting time courses, we extracted continuous functions $u(t)$, which served as inputs to our stochastic promoter model (**Supplementary Text S.1**). Since we found nuclear Msn2 concentration to vary very little between cells (Supplementary Figure 1), we considered $u(t)$ to be deterministic. We performed this as described previously[4] and elaborated on here. We model nuclear Msn2 import with first-order kinetics:

$$u(t)=A(1-e^{-c_1 t}).$$

That is, if Msn2 is cytoplasmic at time t=0, the nuclear level of Msn2 at a later time, t, is given by the above expression where A is the maximal level of nuclear Msn2 for the given concentration of 1-NM-PP1. We chose the 1-NM-PP1 concentrations as 100 nM, 275 nM, 690 nM and 3000 nM such that they would correspond to approximately 25%, 50%, 75% and 100% of maximal nuclear Msn2. c_1 is a fit parameter describing the rate of nuclear import, which we found to vary slightly depending on the 1-NM-PP1 concentration.

Similarly, we model export of Msn2 from the nucleus as a first-order process:

$$u(t_2) = u(t_1)e^{-c_2(t_2-t_1)}.$$

Here, $u(t_1)$ is the nuclear level of Msn2 when the microfluidic device was switched to medium without 1-NM-PP1. Correspondingly, $u(t_2)$ is the nuclear level of Msn2 at some later time $t_2>t_1$. This is to account for the fact that, depending on the pulse duration, Msn2 may not have reached its maximal nuclear level, A.

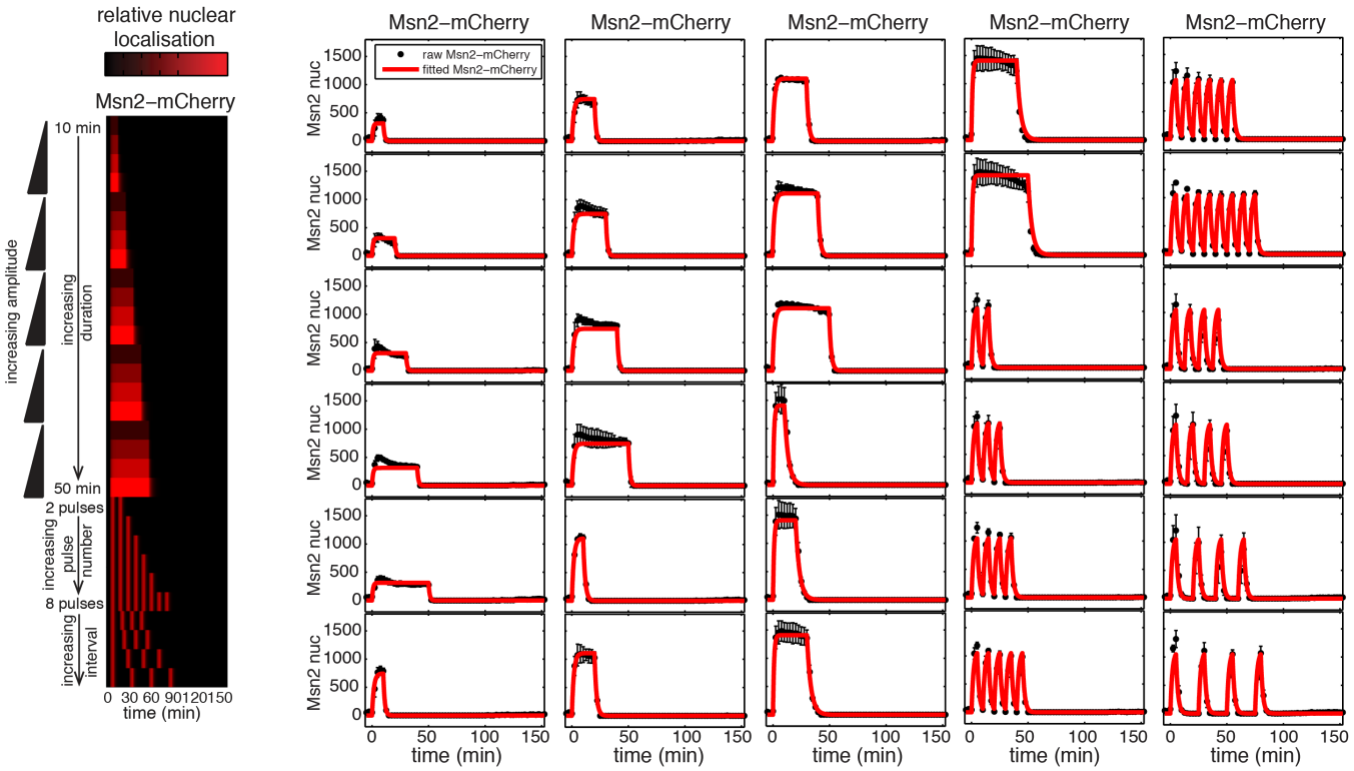
The parameters A, c_1 and c_2 were determined through fitting. Specifically, we took the full 30 different pulses and inferred the best-fit values for A, c_1 and c_2 using least squares fitting.

Acknowledgements

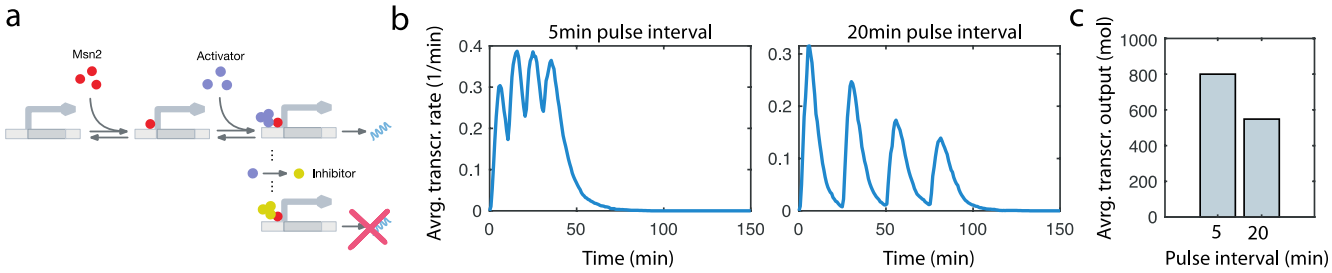
ASH acknowledges support from Howard Hughes Medical Institute (to Erin K. O'Shea), Siebel Stem Cell Foundation (post-doctoral fellowship) and the National Institutes of Health (NIGMS K99GM130896) during parts of this work. CZ acknowledges support from the Max Planck Society, the MPI-CBG as well as the Center for Advancing Electronics Dresden (cfaed). We thank Nan Hao, Nadine Vastenhouw, Stephan Grill, Andre Nadler, Carl Modes, Alf Honigmann, Pavel Tomancak and Tony Hyman for insightful comments on the manuscript.

REFERENCES

1. Coulon, A., Chow, C. C., Singer, R. H. & Larson, D. R. Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat. Rev. Genet.* **14,** 572 (2013).
2. Suter, D. M. *et al.* Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics. *Science (80-.).* **332,** 472–474 (2011).
3. Hao, N. & O'Shea, E. K. Signal-dependent dynamics of transcription factor translocation controls gene expression. *Nat. Struct. &Amp; Mol. Biol.* **19,** 31 (2011).
4. Hansen, A. S. & O'Shea, E. K. Promoter decoding of transcription factor dynamics involves a trade-off between noise and control of gene expression. *Mol. Syst. Biol.* **9,** 704 (2013).
5. Zoller, B., Nicolas, D., Molina, N. & Naef, F. Structure of silent transcription intervals and noise characteristics of mammalian genes. *Mol. Syst. Biol.* (2015). doi:10.15252/msb.20156257
6. Hansen, A. S. & O'Shea, E. K. Cis Determinants of Promoter Threshold and Activation Timescale. *Cell Rep.* (2015). doi:10.1016/j.celrep.2015.07.035
7. Hao, N. & O'Shea, E. K. Signal-dependent dynamics of transcription factor translocation controls gene expression. *Nat. Struct. &Amp; Mol. Biol.* **19,** 31 (2011).
8. Hansen, A. S., Hao, N. & OShea, E. K. High-throughput microfluidics to control and measure signaling dynamics in single yeast cells. *Nat. Protoc.* (2015). doi:10.1038/nprot.2015.079
9. Coulon, A., Chow, C. C., Singer, R. H. & Larson, D. R. Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat. Rev. Genet.* **14,** 572 (2013).
10. Neuert, G. *et al.* Systematic Identification of Signal-Activated Stochastic Gene Regulation. *Science (80-.).* **339,** 584–587 (2013).
11. Zechner, C., Unger, M., Pelet, S., Peter, M. & Koeppl, H. Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nat. Methods* **11,** 197–202 (2014).
12. Feigelman, J. *et al.* Analysis of Cell Lineage Trees by Exact Bayesian Inference Identifies Negative Autoregulation of Nanog in Mouse Embryonic Stem Cells. *Cell Syst.* **3,** 480–490.e13 (2016).
13. Kuzmanovska, I., Milias-Argeitis, A., Mikelson, J., Zechner, C. & Khammash, M. Parameter inference for stochastic single-cell dynamics from lineage tree data. *BMC Syst. Biol.* **11,** (2017).
14. Bar-Even, A. *et al.* Noise in protein expression scales with natural protein abundance. *Nat. Genet.* **38,** 636 (2006).
15. Newman, J. R. S. *et al.* Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441,** 840 (2006).
16. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic Gene Expression in a Single Cell. *Science (80-.).* **297,** 1183 LP-1186 (2002).
17. Cai, L., Friedman, N. & Xie, X. S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440,** 358–362 (2006).
18. Cheong, R., Rhee, a., Wang, C. J., Nemenman, I. & Levchenko, a. Information Transduction Capacity of Noisy Biochemical Signaling Networks. *Science (80-.).* **334,** 354–358 (2011).
19. Selimkhanov, J. *et al.* Accurate information transmission through dynamic biochemical signaling networks. *Science* **346,** 1370–3 (2014).
20. Uda, S. *et al.* Robustness and compensation of information transmission of signaling pathways. *Science* **341,** 558–61 (2013).
21. Voliotis, M., Perrett, R. M., McWilliams, C., McArdle, C. a & Bowsher, C. G. Information transfer by leaky, heterogeneous, protein kinase signaling systems. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E326-33 (2014).
22. Doucet, A., Godsill, S. & Andrieu, C. On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. Comput.* **10,** 197–208 (2000).
23. Huang, L. *et al.* Reconstructing dynamic molecular states from single-cell time series. *J. R. Soc. Interface* (2016). doi:10.1098/rsif.2016.0533



Supplementary Figure 1. Overview of Msn2 input experiments. Left: heatmap overview of the 30 different Msn2-mCherry input. Right: Raw experimentally measured Msn2-mCherry input (black) and standard deviation (black error bars) for each of the 30 Msn2 inputs. The fitted Msn2-mCherry input is overlaid in red. This figure has been partially reproduced with permission from *Molecular Systems Biology*[4].



Supplementary Figure 2. Toy model with interval-dependent promoter memory. (**a**) Model scheme. Once Msn2 binds to the promoter, activator molecules can be recruited, which causes the promoter to switch into a transcriptionally active state with a rate proportional to the number of activators present. Once the promoter switches back into the Msn2-unbound state, the activator can be converted into an inhibitor, which causes the promoter to switch into a transcriptionally inactive state with a rate proportional to the number of inhibitors present. (**b**) Average transcription rate for 5 min and 20 min pulse intervals as a function of time obtained by forward simulation of the model. Blue lines indicate averages computed from stochastic simulations (n=2000). (**c**) Corresponding average transcriptional output for 5min and 20min pulse intervals. A detailed reaction scheme and parameters used for simulation can be found in Supplementary text S.3. We emphasize that this Toy model only serves to illustrate one possible scenario, which could result in a pulse-interval dependent switch from positive to negative memory, as observed in **Fig. 4a-b**. We do not currently understand the mechanism underlying the observation in **Fig. 4a-b**.

Supplementary Text

Promoters adopt distinct dynamic manifestations depending on transcription factor context

Anders S. Hansen and Christoph Zechner

S.1 Statistical inference of transcription dynamics from time-course reporter measurements

S.1.1 Stochastic model of Msn2-dependent gene expression

We describe Msn2-dependent gene expression using a canonical three-state model as shown in Figure S.1. The promoter is described as a continuous-time Markov chain, which switches stochastically between three states of different transcriptional activity (Figure S.1a). Correspondingly, the rate of transcription at time t is governed by a stochastic process $Z(t) \in \{z_0, z_1, z_2\}$, whose value changes discontinuously whenever the promoter transitions from one state into another. In the absence of nuclear Msn2, the promoter is in its transcriptionally inactive state ($z_0 = 0$), where no transcripts are produced. Upon recruitment of Msn2 to the promoter, it can switch into a transcriptionally permissive state in which transcription takes place with propensity z_1. To account for Msn2-dependent promoter activation, we consider the switching rate from z_0 to z_1 to depend on the nuclear Msn2 concentration. For simplicity, we consider a linear dependency, i.e., $q_{01}(t) = \gamma u(t)$, with $u(t)$ as the Msn2 abundance at time t. The corresponding reverse rate q_{10} is considered to be constant. We assume that transcription can be further enhanced by recruitment of additional factors such as chromatin remodeling complexes and general transcriptional factors. This is captured in our model by introducing a third state with transcription rate z_2 and corresponding transition rates q_{12} and q_{21}. With this, we can describe the time-dependent probability distribution over the transcription rate $P_Z(t) = (P(Z(t) = z_0 \mid \theta), P(Z(t) = z_1 \mid \theta), P(Z(t) = z_2 \mid \theta))^T$ in terms of a forward equation

$$\frac{\mathrm{d}}{\mathrm{d}t} P_Z(t) = QP(t) = \begin{pmatrix} -q_{01}(t) & q_{10} & 0 \\ q_{01}(t) & -q_{10} - q_{12} & q_{21} \\ 0 & q_{12} & -q_{21} \end{pmatrix} P_Z(t), \tag{1}$$

with $P_Z(0) = p_{z,0}$ as some initial distribution over $Z(t)$ and $\theta = \{\gamma, q_{10}, q_{12}, q_{21}\}$ as a set of parameters. In the following, we denote by $\mathbf{z}_t = \{z(s) \mid 0 \leq s \leq T\}$ a complete realization of $Z(t)$ on a fixed time interval $[0, t]$. Furthermore, we introduce the conditional path distribution $p(\mathbf{z}_t \mid \theta)$ which measures the likelihood of observing a particular trajectory $\mathbf{z}_t$ for a given parameter set θ. Note that it is straightforward to draw random sample paths $\mathbf{z}_t$ from this distribution using Gillespie's stochastic simulation algorithm [2] or its variants.

Transcription and translation are modeled as a two-stage reaction network as shown in Figure S.1b. We denote by $M(t)$ and $N(t)$ the copy numbers of mRNA and protein at time t, respectively. The parameters c_1 and c_2 are the mRNA and protein degradation rates and A is the protein translation rate. To account for cell-to-cell variability in protein translation, we consider the latter to be randomly distributed across isogenic cells, i.e., $A \sim p(a \mid \beta)$, with $p(a \mid \beta)$ as an arbitrary probability density function (pdf) with positive support and β as a set of hyperparameters characterizing this distribution [3, 4]. Here we consider as hyperparameters the average and coefficient of variation (CV) of A such that $\beta = \{\langle A \rangle, CV[A]\}$. Consequently, β captures the magnitude and variability associated with protein translation. In the following, we denote by $\omega = \{c_1, c_2, \beta\}$ the set of parameters corresponding to transcription and translation.

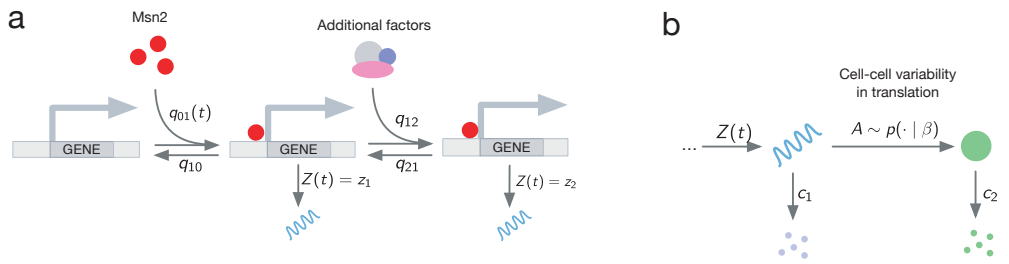


Figure S.1: Stochastic model of Msn2-inducible reporter expression. (a) Promoter model. (b) Model of transcription and translation.

For a given set of parameters θ and ω and a concrete realization of the translation rate A, the overall dynamics of the joint system state $(Z(t), M(t), N(t))$ can be described by a Markov chain. However, due to the random variability over A, each cell is associated with a differently parameterized Markov chain. This results in a so-called mixed-effect Markov model, whose computational analysis turns out to be challenging [3]. One way to address this issue is to augment the state space by the random variable A and to formulate a master equation on this extended space. For $S(t) = (Z(t), M(t), N(t), A)$, such master equation reads

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}P(z_0,m,n,a,t) &= z_0 P(z_0,m-1,n,a,t) + c_1(m+1)P(z_0,m+1,n,a,t) \\
&\quad + amP(z_0,m,n-1,a,t) + c_2(n+1)P(z_0,m,n+1,a,t) \\
&\quad - \left[z_0 + c_1 m + am + c_2 n\right] P(z_0,m,n,a,t) \\
&\quad + q_{10}P(z_1,m,n,a,t) - q_{01}(t)P(z_0,m,n,a,t) \\
\frac{\mathrm{d}}{\mathrm{d}t}P(z_1,m,n,a,t) &= z_1 P(z_1,m-1,n,a,t) + c_1(m+1)P(z_1,m+1,n,a,t) \\
&\quad + amP(z_1,m,n-1,a,t) + c_2(n+1)P(z_1,m,n+1,a,t) \\
&\quad - \left[z_1 + c_1 m + am + c_2 n\right] P(z_1,m,n,a,t) \\
&\quad + q_{01}(t)P(z_0,m,n,a,t) - q_{10}(t)P(z_1,m,n,a,t) \\
&\quad + q_{21}P(z_2,m,n,a,t) - q_{12}(t)P(z_1,m,n,a,t) \\
\frac{\mathrm{d}}{\mathrm{d}t}P(z_2,m,n,a,t) &= z_2 P(z_2,m-1,n,a,t) + c_1(m+1)P(z_2,m+1,n,a,t) \\
&\quad + amP(z_2,m,n-1,a,t) + c_2(n+1)P(z_2,m,n+1,a,t) \\
&\quad - \left[z_2 + c_1 m + am + c_2 n\right] P(z_2,m,n,a,t) \\
&\quad + q_{12}(t)P(z_1,m,n,a,t) - q_{21}(t)P(z_2,m,n,a,t)
\end{aligned}
\tag{2}
$$

with $P(z_i, m, n, a) := P(Z(t) = z_i, M(t) = m, N(t) = n, A \in [a + \mathrm{d}a) \mid \theta, \omega)$. Differential equations for arbitrary moments $\mathbb{E}\left[f(Z(t), M(t), N(t), A)\right]$ with f as a polynomial can be computed by multiplying (3) with f and summing or integrating over all possible values of m, n, z_i and a, respectively [4]. Since the network is linear in its propensities, the resulting moment equations turn out to be closed and can thus be computed without approximation. Moreover, we will denote by $\mathbf{s}_t = \{s(u) \mid 0 \leq u \leq t\}$ a complete sample path of the full system state between time zero and t.

S.1.2 Conditional dynamics of transcription and translation

One major difficulty in inferring gene networks like the one in Figure S.1 is that they involve both very lowly and highly abundant components. This is why moment-based descriptions of the full system state $S(t)$ are of limited use for the time-series inference problem considered here as will be discussed

later in Section S.1.3. On the other hand, approaches purely based on stochastic simulation become computationally expensive, since transcription and translation often involve thousands or even millions of events over the duration of a time-course experiment. In such cases, hybrid approaches can be beneficial, where only the lowly abundant components are described stochastically, whereas the remaining components are handled using moment equations. In the scenario considered here, for instance, the time evolution of the transcription rate $\mathbf{z}_t$ can be efficiently simulated using stochastic simulation since the number of times the promoter switches between states is comparably small. For a given $\mathbf{z}_t$, one could then calculate a corresponding set of conditional moments characterizing the dynamics of mRNA and protein. More technically, this can be understood by the fact that the path distribution over the total system state factorizes into $p(\mathbf{s}_t \mid \omega, \theta) = p(\mathbf{x}_t \mid \mathbf{z}_t, \omega) p(\mathbf{z}_t \mid \theta)$. Correspondingly, we can describe the dynamics over $X(t) = (M(t), N(t), A)$ as a conditional Markov process $X(t) \mid \mathbf{z}_t$, whose state probability distribution $P(m, n, a, t) := P(M(t) = m, N(t) = n, A \in [a + \mathrm{d}a) \mid \mathbf{z}_t)$ satisfies

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} P(m,n,a,t) = {} & z(t) P(m-1,n,a,t) + c_1 (m+1) P(m+1,n,a,t) \\ & + am P(m,n-1,a,t) + c_2 (n+1) P(m,n+1,a,t) \\ & - \left[z(t) + c_1 m + am + c_2 n \right] P(m,n,a,t),\end{aligned} \tag{3}$$

whereas we assume for the initial condition $P(m, n, a, t = 0) = P(M(0) = m, N(0) = n \mid Z(0) = z_0) p(a \mid \beta)$. In order the derive conditional moments $\mathbb{E}\left[X(t) \mid \mathbf{z}_t\right]$, we multiply (3) with polynomials in x and sum and integrate over all m, n and a, respectively. Here, we consider moments of mRNA and protein up to order two, which can be fully described by the system of differential equations

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left[M(t) \mid \mathbf{z}_t\right] = z(t) - \mathbb{E}\left[M(t) \mid \mathbf{z}_t\right] c_1 \\
&\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left[N(t) \mid \mathbf{z}_t\right] = \mathbb{E}\left[M(t) A \mid \mathbf{z}_t\right] - \mathbb{E}\left[N(t) \mid \mathbf{z}_t\right] c_2 \\
&\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left[M(t)^2 \mid \mathbf{z}_t\right] = z(t) + 2\mathbb{E}\left[M(t) \mid \mathbf{z}_t\right] z(t) + \mathbb{E}\left[M(t) \mid \mathbf{z}_t\right] c_1 - 2\mathbb{E}\left[M(t)^2 \mid \mathbf{z}_t\right] c_1 \\
&\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left[M(t) N(t) \mid \mathbf{z}_t\right] = \mathbb{E}\left[N(t) \mid \mathbf{z}_t\right] z(t) - \mathbb{E}\left[M(t) N(t) \mid \mathbf{z}_t\right] c_1 - \mathbb{E}\left[M(t) N(t) \mid \mathbf{z}_t\right] c_2 + \mathbb{E}\left[M(t)^2 A \mid \mathbf{z}_t\right] \\
&\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left[M(t) A \mid \mathbf{z}_t\right] = \mathbb{E}\left[A \mid \mathbf{z}_t\right] z(t) - \mathbb{E}\left[M(t) A \mid \mathbf{z}_t\right] c_1 \\
&\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left[N(t)^2 \mid \mathbf{z}_t\right] = \mathbb{E}\left[N(t) \mid \mathbf{z}_t\right] c_2 + \mathbb{E}\left[M(t) A \mid \mathbf{z}_t\right] - 2\mathbb{E}\left[N(t)^2 \mid \mathbf{z}_t\right] c_2 + 2\mathbb{E}\left[M(t) N(t) A \mid \mathbf{z}_t\right] \\
&\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left[N(t) A \mid \mathbf{z}_t\right] = \mathbb{E}\left[M(t) A^2 \mid \mathbf{z}_t\right] - \mathbb{E}\left[N(t) A \mid \mathbf{z}_t\right] c_2 \\
&\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left[M(t)^2 A \mid \mathbf{z}_t\right] = \mathbb{E}\left[A \mid \mathbf{z}_t\right] z(t) + 2\mathbb{E}\left[M(t) A \mid \mathbf{z}_t\right] z(t) + \mathbb{E}\left[M(t) A \mid \mathbf{z}_t\right] c_1 - 2\mathbb{E}\left[M(t)^2 A \mid \mathbf{z}_t\right] c_1 \\
&\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left[M(t) N(t) A \mid \mathbf{z}_t\right] = \mathbb{E}\left[N(t) A \mid \mathbf{z}_t\right] z(t) - \mathbb{E}\left[M(t) N(t) A \mid \mathbf{z}_t\right] c_1 - \mathbb{E}\left[M(t) N(t) A \mid \mathbf{z}_t\right] c_2 \\
&\qquad + \mathbb{E}\left[M(t)^2 A^2 \mid \mathbf{z}_t\right] \\
&\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left[M(t) A^2 \mid \mathbf{z}_t\right] = \mathbb{E}\left[A^2 \mid \mathbf{z}_t\right] z(t) - \mathbb{E}\left[M(t) A^2 \mid \mathbf{z}_t\right] c_1 \\
&\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left[M(t)^2 A^2 \mid \mathbf{z}_t\right] = \mathbb{E}\left[A^2 \mid \mathbf{z}_t\right] z(t) + 2\mathbb{E}\left[M(t) A^2 \mid \mathbf{z}_t\right] z(t) + \mathbb{E}\left[M(t) A^2 \mid \mathbf{z}_t\right] c_1 \\
&\qquad - 2\mathbb{E}\left[M(t)^2 A^2 \mid \mathbf{z}_t\right] c_1 .
\end{aligned} \tag{4}$$

Note that (4) involves all first and second order moments, but also a few additional moments of order three and four, which are needed in order to obtain a closed set of differential equations.

S.1.3 Statistical model of time-series reporter measurements

As detailed in Materials and Methods, we performed quantitative single-cell time-lapse measurements of reporter expression for different Msn2-inducible promoters and Msn2 activation profiles. We denote by $t_1, \ldots, t_K$ the time points at which measurements were taken. Correspondingly, we define by $\mathbf{s}_{l:k}$ a complete sample path of the gene expression system between times t_l and t_k. If $l = 0$, we refer the state at time $t = 0$, which does not necessarily coincide with the first measurement time point t_1. The measurements – denoted by Y_k for $k = 1, \ldots, K$ – provide noisy information about the system state $S(t_k)$ according to a measurement density

$$Y_k \mid (S(t_k) = s_k) \sim p(\cdot \mid s_k).$$

We consider the measurement noise to be independent among time-points such that

$$p(y_1, \ldots, y_K \mid s_1, \ldots, s_K) = \prod_{k=1}^{K} p(y_k \mid s_k). \tag{5}$$

In our particular case, the measurements correspond to the reporter abundance $N(t)$ corrupted by measurement noise such that

$$p(y_k \mid s_k) = p(y_k \mid x_k) = p(y_k \mid n_k).$$

For a given set of parameters $\{\theta, \omega\}$, the relation between a complete sample path $\mathbf{s}_{0:K}$ and the observed measurements is captured by a joint distribution

$$p\left(y_1, \ldots, y_K, \mathbf{s}_{0:K} \mid \omega, \theta\right) = P(s_0) p(\mathbf{s}_{0:K} \mid \omega, \theta) \prod_{k=1}^{K} p(y_k \mid s_k), \tag{6}$$

with $P(s_0) := P(S(0) = s_0)$ as the initial distribution over the system state and $p(\mathbf{s}_{0:K} \mid \omega, \theta)$ as the distribution over complete sample paths $\mathbf{s}_{0:K}$. Correspondingly, the posterior distribution over $\mathbf{s}_{0:K}$ is proportional to (6), i.e.,

$$p\left(\mathbf{s}_{0:K} \mid y_1, \ldots, y_K, \omega, \theta\right) \propto P(s_0) p(\mathbf{s}_{0:K} \mid \omega, \theta) \prod_{k=1}^{K} p(y_k \mid s_k). \tag{7}$$

S.1.4 Recursive Bayesian Estimation

The posterior distribution (7) is generally intractable but several approximate techniques have been proposed previously [3]. Most of them rely on Bayesian filtering methods, which construct an approximation of (7) recursively over measurement time points. In those approaches, one exploits the fact that the posterior distribution at any measurement time t_k can be written recursively as

$$p\left(\mathbf{s}_{0:k} \mid y_1, \ldots, y_k, \omega, \theta\right) \propto p(y_k \mid s_k) p(\mathbf{s}_{k-1:k} \mid s_{k-1}, \omega, \theta) p\left(\mathbf{s}_{0:k-1} \mid y_1, \ldots, y_{k-1}, \omega, \theta\right), \tag{8}$$

with $p\left(\mathbf{s}_{0:k-1} \mid y_1, \ldots, y_{k-1}, \omega, \theta\right)$ as the posterior distribution at time t_{k-1}. In order to solve the Bayesian recursion between consecutive time steps, one can either employ analytical approximations, or Monte Carlo methods. In a recent study, for instance, we have proposed a Gaussian approximation of the Bayesian filtering problem, which relies on the time-evolution of the first and second order moments of the gene network dynamics [5]. While computationally efficient, the underlying Gaussian approximation is not suitable for lowly abundant, switch-like components, such as the transcription rate $Z(t)$ in our promoter model. Alternative approaches are mostly based on sequential Monte Carlo techniques [6, 3], which approximate (8) using a sufficiently large number of Monte Carlo samples drawn by SSA. The main advantage of these techniques is that they are exact up to sampling variance but on their downside, suffer from limited scalability. In particular, forward-simulation via SSA can become prohibitively slow, especially when RNAs and proteins are highly abundant. They are therefore not able to tackle large datasets like the one considered here. In the following, we will present a novel hybrid inference algorithm, which bypasses expensive SSA simulations of highly abundant species, making it sufficiently scalable to deal with datasets that span tens- or even hundreds of thousands of single-cell trajectories.

S.1.5 Hybrid sequential Monte Carlo

One strategy to improve the scalability of sequential Monte Carlo techniques is to analytically eliminate variables that are not of direct interest to a particular inference problem [7, 3]. In our case, for instance, we are specifically interested in the promoter switching dynamics and the corresponding transcription rate $Z(t)$. From this perspective, it would therefore suffice to calculate the marginal posterior distribution

$$p(\mathbf{z}_{0:K} \mid y_1, \ldots, y_K, \theta, \omega) \propto p(y_1, \ldots, y_K, \mathbf{z}_{0:K} \mid \theta, \omega) = \mathbb{E}\left[p(y_1, \ldots, y_K, \mathbf{x}_{0:K}, \mathbf{z}_{0:K} \mid \theta, \omega)\right], \tag{9}$$

in which the dynamics of $X(t)$ have been "integrated out". In order to perform this integration, we first realize that the joint distribution can be rewritten as

$$\begin{aligned} p\left(y_1, \ldots, y_K, \mathbf{s}_{0:K} \mid \omega, \theta\right) &= p\left(y_1, \ldots, y_K, \mathbf{x}_{0:K}, \mathbf{z}_{0:K} \mid \omega, \theta\right) \\ &= P(s_0) p(\mathbf{s}_{0:K} \mid \omega, \theta) \prod_{k=1}^{K} p(y_k \mid s_k) \\ &= P(x_0, z_0) p(\mathbf{x}_{0:K} \mid \mathbf{z}_{0:K}, \omega) p(\mathbf{z}_{0:K} \mid \theta) \prod_{k=1}^{K} p(y_k \mid x_k) \\ &= P(x_0, z_0) p(\mathbf{z}_{0:K} \mid \theta) \prod_{k=1}^{K} p(y_k \mid x_k) p(\mathbf{x}_{k-1:k} \mid x_{k-1}, \mathbf{z}_{k-1:k}, \omega), \end{aligned} \tag{10}$$

where we have made use of the fact that $p(\mathbf{s}_{0:K} \mid \omega, \theta) = p(\mathbf{x}_{0:K} \mid \mathbf{z}_{0:K}, \omega) p(\mathbf{z}_{0:K} \mid \theta)$ as shown in Section S.1.2. Next, we integrate (10) over all subpaths $\{\mathbf{x}_{k-1:k} \setminus x_k\}$ such that only the values of $X(t)$ at the time points $t_0, \ldots, t_K$ remain in the model. This integration is straightforward and can be carried out by replacing the path distribution $p(\mathbf{x}_{k-1:k} \mid x_{k-1}, \mathbf{z}_{k-1:k}, \omega)$ by the state transition kernel $P(x_k \mid x_{k-1}, \mathbf{z}_{k-1:k}, \omega)$, i.e.,

$$p\left(y_1, \ldots, y_K, x_0, \ldots, x_K, \mathbf{z}_{0:K} \mid \omega, \theta\right) = P(x_0, z_0) p(\mathbf{z}_{0:K} \mid \theta) \prod_{k=1}^{K} p(y_k \mid x_k) P(x_k \mid x_{k-1}, \mathbf{z}_{k-1:k}, \omega). \tag{11}$$

The marginalization over the remaining variables $x_0, \ldots, x_K$ then reduces to a summation

$$p(y_1, \ldots, y_K, \mathbf{z}_{0:K} \mid \theta, \omega) = \sum_{x_0} \cdots \sum_{x_K} p\left(y_1, \ldots, y_K, x_0, \ldots, x_K, \mathbf{z}_{0:K} \mid \omega, \theta\right). \tag{12}$$

Most conveniently, this summation can be solved iteratively, by first summing over x_0, subsequently over x_1 and so forth. The first summation yields

$$\begin{aligned} p\left(y_1, \ldots, y_K, x_1, \ldots, x_K, \mathbf{z}_{0:K} \mid \omega, \theta\right) &= \sum_{x_0} P(x_0 \mid z_0) p(y_1 \mid x_1) P(x_1 \mid x_0, \mathbf{z}_{0:1}, \omega) p(\mathbf{z}_{0:K} \mid \theta) \\ &\quad \times \prod_{k=2}^{K} p(y_k \mid x_k) P(x_k \mid x_{k-1}, \mathbf{z}_{k-1:k}, \omega) \\ &= p(y_1 \mid x_1) P(x_1 \mid \mathbf{z}_{0:1}, \omega) p(\mathbf{z}_{0:K} \mid \theta) \times \prod_{k=2}^{K} p(y_k \mid x_k) P(x_k \mid x_{k-1}, \mathbf{z}_{k-1:k}, \omega) \\ &= p(\mathbf{z}_{0:K} \mid \theta) P(x_1 \mid y_1, \mathbf{z}_{0:1}, \omega) p(y_1 \mid \mathbf{z}_{0:1}, \omega) \times \prod_{k=2}^{K} p(y_k \mid x_k) P(x_k \mid x_{k-1}, \mathbf{z}_{k-1:k}, \omega), \end{aligned} \tag{13}$$

whereas the last step follows from the fact that $p(y_1 \mid x_1)P(x_1 \mid \mathbf{z}_{0:1}, \omega) = P(x_1 \mid y_1, \mathbf{z}_{0:1}, \omega)p(y_1 \mid \mathbf{z}_{0:1}, \omega)$ via Bayes' rule. Repeating the same procedure for x_1 yields

$$
\begin{aligned}
&p\left(y_1, \ldots, y_K, x_2, \ldots, x_K, \mathbf{z}_{0:K} \mid \omega, \theta\right) \\
&\quad = \sum_{x_1} p(\mathbf{z}_{0:K} \mid \theta)P(x_1 \mid y_1, \mathbf{z}_{0:1}, \omega)p(y_1 \mid \mathbf{z}_{0:1}, \omega) \times \prod_{k=2}^{K} p(y_k \mid x_k)P(x_k \mid x_{k-1}, \mathbf{z}_{k-1:k}, \omega) \\
&\quad = p(\mathbf{z}_{0:K} \mid \theta)p(y_2 \mid x_2) \sum_{x_1} P(x_2 \mid x_1, \mathbf{z}_{1:2}, \omega)P(x_1 \mid y_1, \mathbf{z}_{0:1}, \omega)p(y_1 \mid \mathbf{z}_{0:1}, \omega) \\
&\qquad \times \prod_{k=3}^{K} p(y_k \mid x_k)P(x_k \mid x_{k-1}, \mathbf{z}_{k-1:k}, \omega) \\
&\quad = p(\mathbf{z}_{0:K} \mid \theta)p(y_2 \mid x_2)P(x_2 \mid y_1, \mathbf{z}_{0:2}, \omega)p(y_1 \mid \mathbf{z}_{0:1}, \omega) \times \prod_{k=3}^{K} p(y_k \mid x_k)P(x_k \mid x_{k-1}, \mathbf{z}_{k-1:k}, \omega) \\
&\quad = p(\mathbf{z}_{0:K} \mid \theta)P(x_2 \mid y_2, y_1, \mathbf{z}_{0:2}, \omega)p(y_2 \mid y_1, \mathbf{z}_{0:2}, \omega)p(y_1 \mid \mathbf{z}_{0:1}, \omega) \times \prod_{k=3}^{K} p(y_k \mid x_k)P(x_k \mid x_{k-1}, \mathbf{z}_{k-1:k}, \omega).
\end{aligned}
\tag{14}
$$

Continuing the above procedure for $x_2, \ldots, x_K$ finally leads to

$$
p\left(y_1, \ldots, y_K, \mathbf{z}_{0:K} \mid \omega, \theta\right) = p(\mathbf{z}_{0:K} \mid \theta)p(y_1 \mid \mathbf{z}_{0:1}, \omega) \prod_{k=2}^{K} p(y_k \mid y_{k-1}, \ldots, y_1, \mathbf{z}_{0:k}, \omega). \tag{15}
$$

Therefore, the marginal posterior distribution over the transcription dynamics $\mathbf{z}_{0:K}$ is proportional to (15), which can again be expressed recursively as

$$
\begin{aligned}
p\left(\mathbf{z}_{0:K} \mid y_1, \ldots, y_K, \omega, \theta\right) &\propto p\left(y_1, \ldots, y_K, \mathbf{z}_{0:K} \mid \omega, \theta\right) \\
&\propto p(y_K \mid y_{K-1}, \ldots, y_1, \mathbf{z}_{0:K}, \omega)p(\mathbf{z}_{K-1:K} \mid z_{K-1}, \theta)p\left(\mathbf{z}_{0:K-1} \mid y_1, \ldots, y_{K-1}, \omega, \theta\right).
\end{aligned}
\tag{16}
$$

Importantly, using eq. (16) we can perform a sequential Monte Carlo algorithm on a significantly reduced sampling space, where only the transcription dynamics $\mathbf{z}_{0:K}$ have to be simulated explicitly. However, in order to perform this algorithm, we need to be able to calculate the marginal likelihood terms $p(y_k \mid y_{k-1}, \ldots, y_1, \mathbf{z}_{0:k}, \omega)$, which are given by

$$
\begin{aligned}
p(y_k \mid y_{k-1}, \ldots, y_1, \mathbf{z}_{0:k}, \omega) &= \sum_{x_k} p(y_k \mid x_k)P(x_k \mid y_{k-1}, \ldots, y_1, \mathbf{z}_{0:k}, \omega) \\
&= \sum_{x_k} p(y_k \mid x_k) \sum_{x_{k-1}} P(x_k \mid x_{k-1}, \mathbf{z}_{k-1:k}, \omega)P(x_{k-1} \mid y_{k-1}, \ldots, y_1, \mathbf{z}_{0:k-1}, \omega).
\end{aligned}
\tag{17}
$$

The two sums in (17) are generally intractable, but analytical solutions exist if the measurement likelihood function $p(y_k \mid x_k)$ and the state transition kernel $P(x_k \mid x_{k-1}, \mathbf{z}_{k-1:k}, \omega)$ belong to certain classes of distributions. This is the case, for instance, if both are Gaussian. However, this is likely not a good assumption in the scenario considered here, since both the measurement- and state distributions are generally positive and asymmetric. As it turns out, however, eq. (17) has an analytical solution also if both $p(y_k \mid x_k)$ and $P(x_k \mid x_{k-1}, \mathbf{z}_{k-1:k}, \omega)$ are log-normally distributed, which is in good agreement with previous studies[8, 3, 9] and also the experimental data considered here. We therefore assume

$$
\begin{aligned}
Y_k \mid (N(t) = n) &\sim \mathcal{LN}\left(\log(n), \eta^2\right) \\
X(t) \mid \mathbf{z}_t &\sim \mathcal{LN}\left(\mu(t), \Sigma(t)\right),
\end{aligned}
\tag{18}
$$

where η^2 corresponds to the strength of the measurement noise and $\mu(t) \in \mathbb{R}^3$ and $\Sigma(t) \in \mathbb{R}^{3\times 3}$ characterize the distribution over $X(t) = (M(t), N(t), A)$ conditionally on a particular realization of $\mathbf{z}_{0:K}$. More precisely, $\mu(t)$ and $\Sigma(t)$ are the mean and covariance of $\log X(t)$ and we therefore refer to them as logarithmic moments in the following.

Now, assuming that the posterior distribution over $X(t)$ is log-normally distributed at time t_{k-1},

$$p(x_{k-1} \mid y_{k-1}, \dots, y_1, \mathbf{z}_{0:k-1}, \omega) \approx \mathcal{LN}\left(x_{k-1} \mid \mu(t_{k-1}), \Sigma(t_{k-1})\right), \tag{19}$$

it will – based on our assumption – remain log-normal upon applying the state transition kernel, i.e.,

$$\begin{aligned} p(x_k \mid y_{k-1}, \dots, y_1, \mathbf{z}_{0:k}, \omega) &\approx \int p(x_k \mid x_{k-1}, \mathbf{z}_{k-1:k}, \omega) \mathcal{LN}\left(x_{k-1} \mid \mu(t_{k-1}), \Sigma(t_{k-1})\right) \mathrm{d}x_{k-1} \\ &= \mathcal{LN}\left(x_k \mid \mu(t_k), \Sigma(t_k)\right), \end{aligned} \tag{20}$$

whereas the sum has now been replaced by an integral. In order to calculate the logarithmic moments $\mu(t_k)$ and $\Sigma(t_k)$ for a given $\mu(t_{k-1})$ and $\Sigma(t_{k-1})$, one first has to calculate the first and second order moments from the log-normal distribution, propagate those forward in time until t_k using the conditional moment dynamics described in Section S.1.2 and subsequently convert them back into the logarithmic domain to obtain $\mu(t_k)$ and $\Sigma(t_k)$. The relationship between logarithmic and standard moments is given by

$$\begin{aligned} \mathbb{E}\left[X_i(t)\right] &= e^{\mu_i(t) + \frac{1}{2}\Sigma_{ii}(t)} \\ \mathbb{E}\left[X_i(t) X_j(t)\right] &= e^{\mu_i(t) + \mu_j(t) + \frac{1}{2}\left(\Sigma_{ii}(t) + 2\Sigma_{ij}(t) + 2\Sigma_{ji}(t) + \Sigma_{jj}(t)\right)}. \end{aligned} \tag{21}$$

In order to determine the posterior distribution at the next measurement time t_k, we multiply (20) with the log-normal measurement density such that

$$\begin{aligned} p(x_k \mid y_k, \dots, y_1, \mathbf{z}_{0:k}, \omega) &\propto p(y_k \mid x_k) p(x_k \mid y_{k-1}, \dots, y_1, \mathbf{z}_{0:k}, \omega) \\ &= \mathcal{LN}\left(y_k \mid \log(n_k), \eta^2\right) \times \mathcal{LN}\left(x_k \mid \mu(t_k), \Sigma(t_k)\right), \end{aligned} \tag{22}$$

with $n_k = N(t_k)$ as the protein abundance at time t_k. It is straightforward to show that the product of two log-normal distributions in (22) is again a log-normal distribution such that

$$p(x_k \mid y_k, \dots, y_1, \mathbf{z}_{0:k}, \omega) = \mathcal{LN}\left(x_k \mid \mu^+(t_k), \Sigma^+(t_k)\right), \tag{23}$$

with

$$\Sigma^+(t_k) = \left[\frac{1}{\eta^2} w w^T + \Sigma(t_k)^{-1}\right]^{-1} \tag{24}$$

$$\mu^+(t_k) = \Sigma^+(t_k)\left[\frac{1}{\eta^2}\log(y_k) w + \Sigma(t_k)^{-1}\mu(t_k)\right], \tag{25}$$

with $w = (0, 1, 0)^T$ as a vector that reflects the fact that from $X(t) = (M(t), N(t), A)$, the second component (i.e., the protein abundance) is measured experimentally.

For the likelihood term $p(y_k \mid y_{k-1}, \dots, y_1, \mathbf{z}_{0:k}, \omega)$ we obtain

$$\begin{aligned} p(y_k \mid y_{k-1}, \dots, y_1, \mathbf{z}_{0:k}, \omega) &= \int p(y_k \mid x_k) p(x_k \mid y_{k-1}, \dots, y_1, \mathbf{z}_{0:k}, \omega) \mathrm{d}x_k \\ &= \int \mathcal{LN}\left(y_k \mid \log(n_k), \eta^2\right) p(n_k \mid y_{k-1}, \dots, y_1, \mathbf{z}_{0:k}, \omega) \mathrm{d}n_k \\ &= \int \mathcal{LN}\left(y_k \mid \log(n_k), \eta^2\right) \mathcal{LN}\left(n_k \mid \mu_2(t_k), \Sigma_{22}(t_k)\right) \mathrm{d}n_k, \end{aligned} \tag{26}$$

whereas the last line follows from the fact that each dimension i of a multivariate log-normal distribution with logarithmic moments μ and Σ is marginally log-normal with parameters μ_i and Σ_{ii}. This integral can be solved in closed form such that we obtain for the logarithm of the marginal likelihood function

$$\begin{aligned}\log p(y_k \mid y_{k-1}, \ldots, y_1, \mathbf{z}_{0:k}, \omega) = \\ -\frac{1}{2}\left[\frac{(\log y_k - \mu_2(t_k))^2}{\eta^2 + \Sigma_{22}(t_k)} - \log\left(\frac{1}{\eta^2} + \Sigma_{22}(t_k)^{-1}\right) - \log \Sigma_{22}(t_k) - \log \eta^2\right] + const.\end{aligned} \tag{27}$$

Together, eqs. (4), (24), (25) and (27) define a recursive Bayesian filter, which allows us to eliminate the components $X(t)$ from the inference problem. As mentioned above, the remaining component $Z(t)$ can then be inferred efficiently using a conventional sequential importance sampler. To this end, we define a set of J particles, each of them consisting of a path $\mathbf{z}^{(i)}$, a set of logarithmic moments $\mu^{(i)}$ and $\Sigma^{(i)}$ as well as a particle probability $p^{(i)}$. This set of particles serves as a finite sample approximation of the posterior distribution at each iteration k. At the kth time step, J new particles are drawn randomly according to the particle probabilities $p^{(i)}$. For each particle i, the path $\mathbf{z}^{(i)}$ is first extended to the next measurement t_{k+1} using SSA. The new probability of this particle is then determined by first propagating the corresponding logarithmic moments until t_{k+1} using eq. (4) and then evaluating eq. (27). The particle probabilities are then normalized across the J particles such that they sum up to one. Subsequently, $\mu^{(i)}$ and $\Sigma^{(i)}$ are updated using (24) and (25) and the algorithm proceeds with the next iteration. At the final time t_K, the paths $\mathbf{z}^{(i)}$ associated with the particles represent samples from the desired marginal posterior distribution, which can be used for further analysis.

S.1.6 Quantitative characterization of promoter dynamics

The inference algorithm described above allows as to compute an arbitrary number of samples $\mathbf{z}_{0:K}^{(i)}$ from the desired posterior distribution. In order to compare the dynamics of the different promoters under various experimental conditions, we extracted a number of features from these samples that characterize the transcriptional response for each individual cell. More technically, these features can be defined as functionals that map a random path $\mathbf{z}_{0:K}^{(i)}$ to a real or discrete number. This functional can then be averaged with respect to the posterior distribution associated with a particular cell, i.e.,

$$\mathbb{E}\left[f(\mathbf{z}_{0:K}) \mid y_1, \ldots, y_K\right] \approx \frac{1}{J}\sum_{j=1}^{J} f(\mathbf{z}_{0:K}^{(j)}), \tag{28}$$

with $y_1, \ldots, y_K$ as the measurements of this cell and $\mathbf{z}_{0:K}^{(j)}$ as samples from the posterior distribution obtained from the inference method. The following list summarizes the different features that were used in this study.

- **Percentage of responders.** A cell is considered a responder if it managed to switch into a state of significant transcriptional activity at least once. To this end, we defined a functional $r(\mathbf{z}_{0:K}) \in \{0, 1\}$, which is one only if a promoter state was reached which had a transcription rate of at least 20% of the maximum transcription rate taken over all Msn2 induction levels. Depending on the promoter and condition, this could encompass one, two or none of the promoter states. We then estimated the response probability $p_a = \mathbb{E}\left[r(\mathbf{z}_{0:K}) \mid y_1, \ldots, y_K\right]$ for each cell by averaging over all the individual samples paths $\mathbf{z}_{0:K}$ obtained from the sequential Monte Carlo algorithm. A cell was then classified as a responder if $p_a > 0.99$. Subsequently, we calculated the percentage of responders for each promoter and condition.
- **Time to activate.** For all responding cells, we calculated the posterior expectation of the time it took until the cell switched into a transcriptionally significant state, i.e., $\mathbb{E}\left[\tau_S(\mathbf{z}_{0:K}) \mid y_1, \ldots, y_K\right]$ with $\tau_S(\mathbf{z}_{0:K}) \in \mathbb{R}^+$ as a functional that measures the time until the first transition into a responsive state happened. We further calculated the mean and variance of the time until activation over all cells in an experiment.

- **Total time active.** Analogously to the time to activate, we quantified the total time the promoter was active, i.e., $\mathbb{E}\left[\tau_A(\mathbf{z}_{0:K}) \mid y_1, \ldots, y_K\right]$ with $\tau_S(\mathbf{z}_{0:K}) \in \mathbb{R}^+$ as a functional that extracts the total time the promoter spent in any of the active states.
- **Time spent in state i.** We calculated the total time the promoter spent in any of the three states, i.e., $\mathbb{E}\left[\tau_i(\mathbf{z}_{0:K}) \mid y_1, \ldots, y_K\right]$ with $\tau_i(\mathbf{z}_{0:K}) \in \mathbb{R}^+$.
- **Number of transitions from state i to j.** For each cell we calculated how often the promoter switched between states 1 and 2, and 2 and 3, respectively. More technically, we estimated the expectation $\mathbb{E}\left[n_{ij}(\mathbf{z}_{0:K}) \mid y_1, \ldots, y_K\right]$ with $n_{ij}(\mathbf{z}_{0:K}) \in \mathbb{N}$ as a functional that counts the number of transitions from state i to j. Additionally, we counted the total number of transitions that took place over the duration of the time course experiment.
- **Maximum transcription.** We calculated the maximum transcription rate that the promoter achieved during a time course experiment. In particular, we computed the expected transcription rate for each cell $\lambda(t) = \mathbb{E}\left[Z(t) \mid y_1, \ldots, y_K\right]$ and subsequently the corresponding population average $\langle\lambda(t)\rangle$. We then determined the maximum of this average, i.e, $\lambda_{max} = \max_t\langle\lambda(t)\rangle$.
- **Time to maximum transcription.** Next to the maximum transcription, we also determined the time when this maximum was achieved, i.e., $\tau_{max} = \arg\max_t\langle\lambda(t)\rangle$.

S.1.7 Evaluation of the inference method using synthetic data

In order to study the accuracy of the proposed inference method, we tested it using artificially generated data. In particular, we considered two differently parameterized versions of the stochastic model in Figure S.1. The first one resembled a fast promoter like *DCS2* or *HXK1* whereas the second one had slow and switch-like promoter activation kinetics like *SIP18* or *TKL1*. In particular, the parameters of the system were chosen to be $\gamma = 0.05$, $q_{10} = 0.055$, $q_{12} = 0.001\kappa$, $q_{21} = 0.004\kappa$, $z_1 = 0.0035$, $z_2 = 0.728$, $c_1 = 0.0013$, $c_2 = 1.67e-5$, $\langle A\rangle = 0.1$, $CV[A] = 0.02$, whereas $\kappa = \{1, 10\}$ for the slow and fast promoter model, respectively. All rate parameters are given in units s^{-1}.

For each promoter, we generated 30 single cell trajectories between time zero and $t_K = 150\,\text{min}$ using SSA and sampled the protein abundance at 55 equidistant time points $t_1, \ldots, t_K$. For the Msn2 activation function $u(t)$, we used the experimentally determined profile for a single pulse experiment (75% Msn2 induction level, 40min duration). The measurements were then simulated from a log-normal measurement density $\mathcal{LN}\left(y_k \mid \log(n_k), \eta^2\right)$, with n_k as the protein copy number at time t_k and η as the logarithmic standard deviation of this density. For this study, we set $\eta = 0.05$.

We applied the hybrid sequential Monte Carlo algorithm to reconstruct the promoter dynamics and compared it to the true realization. In particular, we analyzed three of the path functionals described in Section S.1.6: total time active, time to activate and transcriptional output. We estimated posterior expectations of these functionals using $J = 400$ Monte Carlo samples and analyzed how they compared to the true values extracted from the exact sample paths $\mathbf{z}_{0:K}$. We first assumed perfect knowledge of all process parameters. The top panels in Figure S.2a and b show the inferred values plotted against the ground truth. For all three features we found a strict linear relationship with a slope close to one. Moreover, the corresponding high R^2 values show that the inference method operates at a very high accuracy. We furthermore analyzed the robustness of the method with respect to parameter mismatch. To this end, we randomly perturbed all of the parameters using a log-normal distribution $\mathcal{LN}\left(\log a, 0.1^2\right)$ with a as the underlying true value. Note that this was performed for each of the considered trajectories separately. In case of poor robustness, we would thus expect a significantly reduced correlation. However, we found for all three features that both the R^2 and slope changed only marginally indicating a high robustness of the method. This is an important feature in practical scenarios where knowledge about process parameters is generally imperfect.

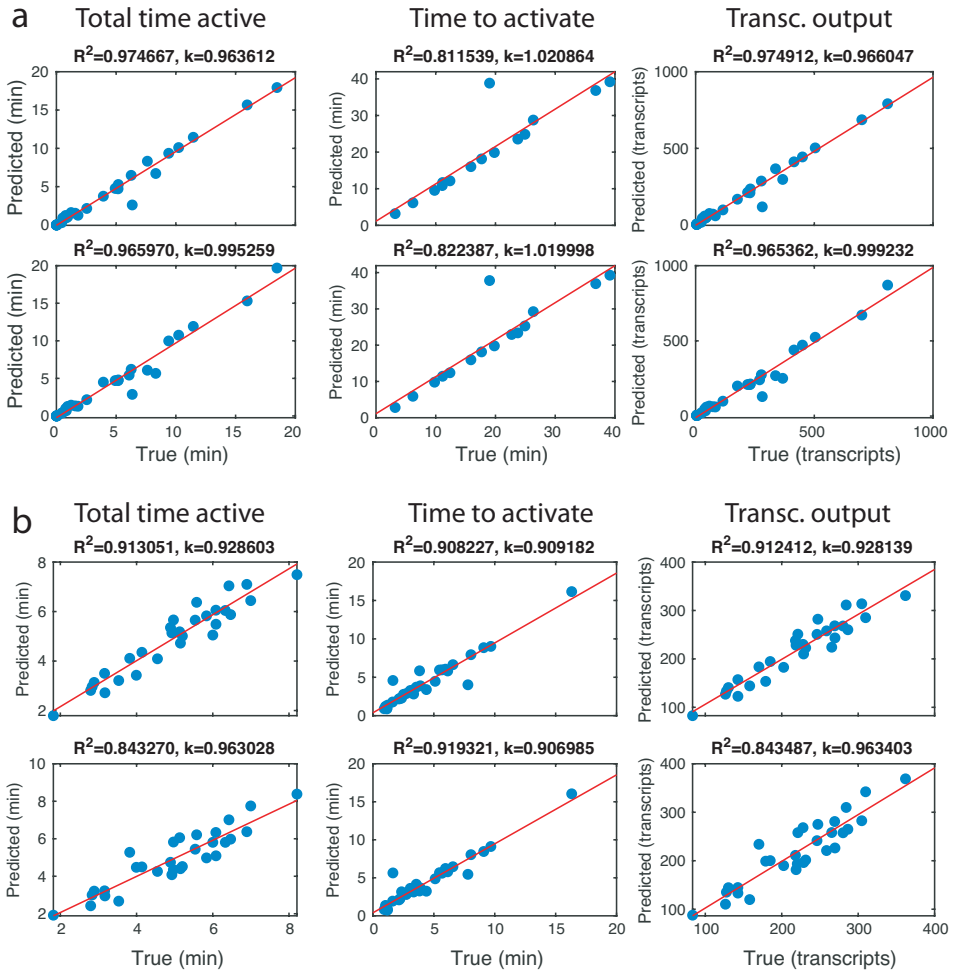


Figure S.2: Evaluation of the inference method using synthetic data. The inference method was performed using artificially generated time-course data as described in the text. (a) Inference results for the slow promoter model. (b) Inference results for the fast promoter model. Respective top panels show the results assuming perfect knowledge of the model parameters. Bottom panels show the corresponding results under parameter mismatch. The R^2 and slope k between the true and predicted features were determined using linear regression (red lines).

S.2 Statistical analysis of Msn2-dependent gene expression

In the following we provide details on the statistical analysis of Msn2-dependent gene expression as shown in the main text. In this case, the function $u(t)$ corresponds to the nuclear Msn2 concentration that was measured experimentally for each condition (Materials and Methods and Supplementary Figure 1). In combination with the measured YFP time series, this allowed us to infer and quantitate the input-output relationship of different promoters under different experimental conditions using the recursive inference method described in Section S.1.4. However, before this method could be applied, the stochastic model from Figure S.1 had to be parameterized. To for this purpose, we used a portion of the experimental single-cell trajectories to infer the kinetic parameters of the model (Section S.2.1). Subsequently, we reconstructed the transcription dynamics of each promoter and condition as described

in Section S.2.2.

S.2.1 Statistical inference of kinetic parameters

In order to parameterize the stochastic gene expression model for different promoters and experimental conditions, we used an established moment-based inference method [4]. This method uses a Markov chain Monte Carlo sampler to match the first and second order moments of the stochastic gene network to the experimentally determined ones. For detailed information on this approach, the reader shall refer to [4]. The kinetics of the same gene expression system have been previously quantified using a simpler, deterministic model [10]. We incorporated this additional information in the form of suitable prior distributions over some of the kinetic parameters. In particular, we considered Gamma prior distributions $p(c_1) = \Gamma(20, 20/1.3e-3s^{-1})$ and $p(\langle A \rangle) = \Gamma(20, 20/0.05s^{-1})$ for the mRNA degradation and average protein translation rates, respectively. Additionally, the protein degradation rate was fixed to $c_2 = 1.67e-5s^{-1}$. For the switching parameters q_{ij}, we used a prior distribution $p(q_{ij}) = \Gamma(1, 1/30s^{-1})$. For each promoter, we first estimated the total set of parameters ω and θ using the single pulse experiments with maximum level and duration (100% Msn2, 50min). In particular, we used a Metropolis-Hastings sampler with lognormal proposal distributions to generate $2e4$ samples, whereas the first $3e3$ samples were discarded. From the resulting samples we extracted maximum a posterior (MAP) estimates of the model parameters. Since the promoter switching dynamics is generally concentration-dependent, we re-estimated θ for the 25%, 50% and 75% Msn2 pulse experiments (also 50min pulse duration). However, we expected the transcriptional and translational dynamics to remain the same among different experiments of the same promoter and we thus kept ω fixed across different Msn2 induction levels. Maximum a posterior (MAP) estimates of the parameters for the different promoters and conditions are summarized in Supplementary Table 1. The resulting model fits are shown Figure S.3.

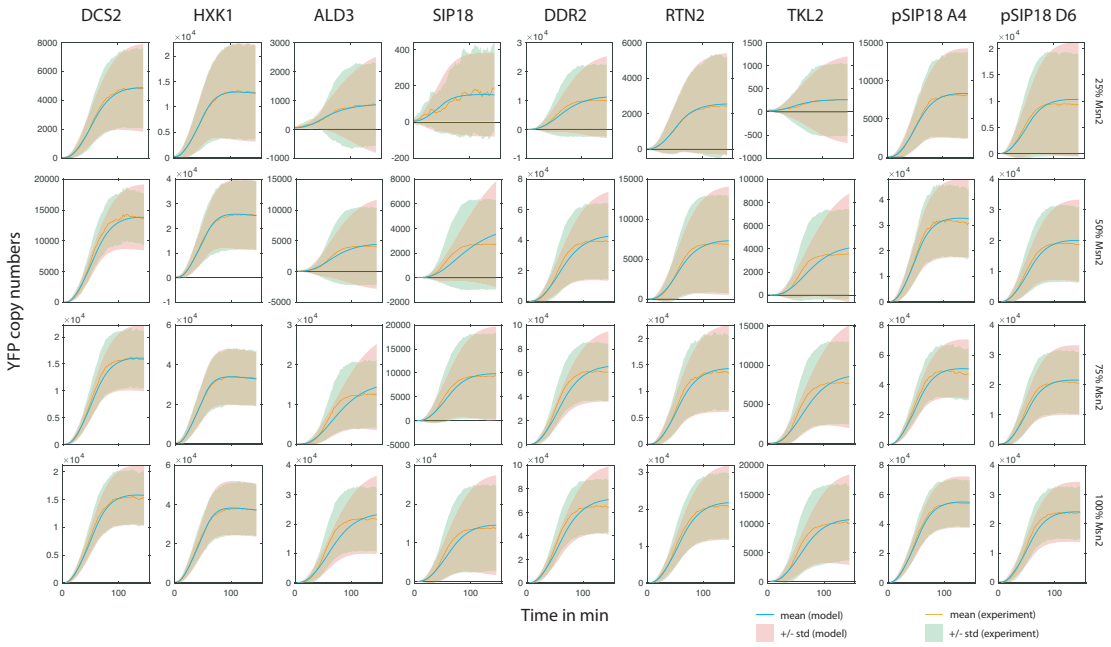


Figure S.3: Comparison of the calibrated model with the experimental data.

S.2.2 Statistical inference of transcription dynamics

Based on the calibrated models, we inferred the transcription and promoter switching dynamics using the recursive Bayesian inference scheme from Section S.1.4. Based on our previous study [3] which uses a similar data processing and calibration pipeline, we set the measurement noise parameter to $\eta = 0.15$ corresponding to an expected relative variation of around 15 percent. For each condition and promoter, we processed each individual cell using $J = 400$ particles. From the resulting samples, we calculated Monte Carlo estimates of the promoter features summarized in Section S.1.6.

S.2.3 Clustering of promoters

To get an overview of how promoters behave under different contexts, we performed a principal component analysis (PCA) of the previously inferred promoter features. For this purpose, we considered all single pulse experiments (10-50min duration, 25-100% Msn2 induction). For each condition, we calculated the percentage of responders, the average time to activate, the average time active and the maximal transcription rate. For each promoter and Msn2 concentration, we concatenated the respective features for all pulse lengths, giving rise to a 20-dimensional feature vector. Each individual feature (e.g., average time active) was normalized across all five pulse lengths and promoters to preserve the relative scaling of this feature with pulse length. In total this leads to 36 20-dimensional data points (4 concentrations for 9 promoters), which we analyzed using PCA. Figure 1f in the main text shows the first two leading principal components plotted against each other.

S.3 Toy model of interval-dependent promoter memory

For the simulations shown in Supplementary Fig. 2, we considered a simple promoter model described by a reaction network

$$\begin{aligned} P_0 &\underset{c_2}{\overset{c_1 u(t)}{\rightleftharpoons}} P_1 \\ P_1 &\underset{c_4}{\overset{c_3 A(t)}{\rightleftharpoons}} P_2 \\ P_0 &\underset{c_6}{\overset{c_5 I(t)}{\rightleftharpoons}} P_3 \\ P_1 &\xrightarrow{c_7} P_1 + A \\ P_0 + A &\xrightarrow{c_8} P_0 + I, \end{aligned} \tag{29}$$

with $u(t)$ as the experimentally measured nuclear Msn2 abundance (Materials and Methods). Transcription takes place with rate z when the promoter is in state P_2. The parameters used for simulation were chosen to be $c_1 = 0.02$, $c_2 = 0.06$, $c_3 = 0.003$, $c_4 = 0.02$, $c_5 = 0.0006$, $c_6 = 0.001$, $c_7 = 0.9$, $c_8 = 7e-6$ and $z = 0.01$ in units s^{-1}.

References

[1] H. Koeppl, C. Zechner, A. Ganguly, S. Pelet, and M. Peter. Accounting for extrinsic variability in the estimation of stochastic rate constants. *Int J Robust Nonlin*, 22(10):1103–1119, 2012.

[2] Daniel T. Gillespie. Stochastic simulation of chemical kinetics. *Annual Review of Physical Chemistry*, 58(1):35–55, 2007.

[3] C. Zechner, M. Unger, S. Pelet, Peter M., and H. Koeppl. Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nat Methods*, 11(2):197–202, 2014.

[4] C. Zechner, J. Ruess, P. Krenn, S. Pelet, M. Peter, J. Lygeros, and H. Koeppl. Moment-based inference predicts bimodality in transient gene expression. *Proc Natl Acad Sci USA*, 109(21):8340–8345, 2012.

[5] Lirong Huang, Loic Pauleve, Christoph Zechner, Michael Unger, Anders S Hansen, and Heinz Koeppl. Reconstructing dynamic molecular states from single-cell time series. *Journal of The Royal Society Interface*, 13(122):20160533, 2016.

[6] Andrew Golightly and Darren J Wilkinson. Bayesian sequential inference for stochastic kinetic biochemical network models. *Journal of Computational Biology*, 13(3):838–851, 2006.

[7] Arnaud Doucet, Nando De Freitas, Kevin Murphy, and Stuart Russell. Rao-blackwellised particle filtering for dynamic bayesian networks. In *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence*, pages 176–183. Morgan Kaufmann Publishers Inc., 2000.

[8] Y. Taniguchi, P. J. Choi, G. 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(5991):533–538, 2010.

[9] Chikara Furusawa, Takao Suzuki, Akiko Kashiwagi, T. Yomo, and K. Kaneko. Ubiquity of log-normal distributions in intra-cellular reaction dynamics. *Biophysics*, 1:25–31, 2005.

[10] Anders S Hansen and Erin K O'shea. Promoter decoding of transcription factor dynamics involves a trade-off between noise and control of gene expression. *Molecular systems biology*, 9(1):704, 2013.