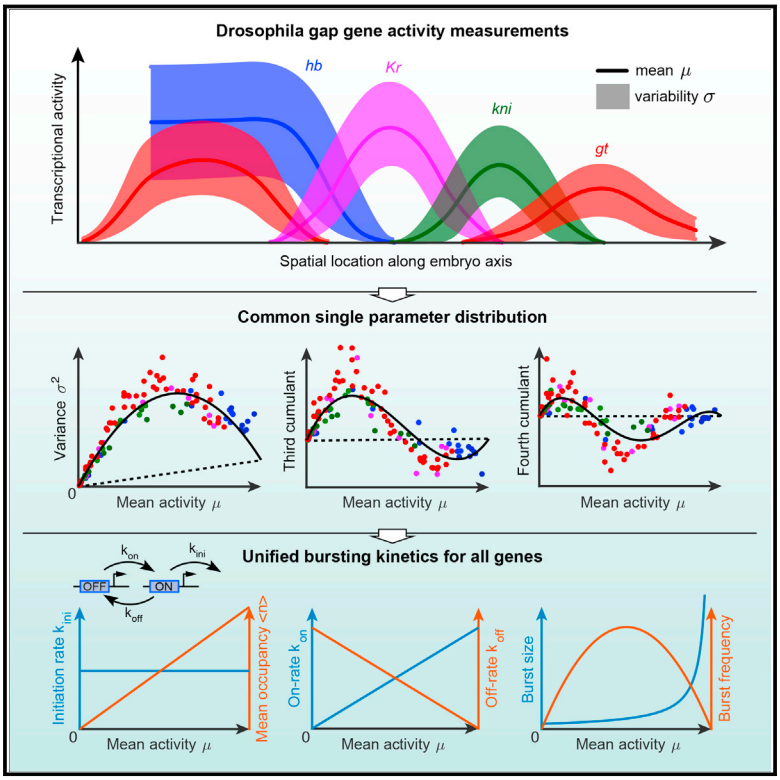
Theory

Cell

Diverse Spatial Expression Patterns Emerge from Unified Kinetics of Transcriptional Bursting

Graphical Abstract



Authors

Benjamin Zoller, Shawn C. Little, Thomas Gregor

Correspondence

tg2@princeton.edu

In Brief

Multiple gene expression boundaries in a developing embryo arise from the same strategy of modulated transcriptional bursting rates.

Highlights

- Measurements of transcriptional bursting for endogenous genes in intact embryos
- Gene expression boundaries arise from a shared regulatory strategy
- Bursting parameters are tightly constrained across genes
- Steady-state synchrony achieved by constant correlation time across genes

Zoller et al., 2018, Cell *175*, 835–847
October 18, 2018
https://doi.org/10.1016/j.cell.2018.09.056

CellPress

Cell

Diverse Spatial Expression Patterns Emerge from Unified Kinetics of Transcriptional Bursting

Benjamin Zoller,[1,5] Shawn C. Little,[2,3,5] and Thomas Gregor[1,4,6,*]
[1]Joseph Henry Laboratories of Physics and the Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA
[2]Department of Molecular Biology and Howard Hughes Medical Institute, Princeton University, Princeton, NJ 08544, USA
[3]Department of Cell and Developmental Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[4]Department of Developmental and Stem Cell Biology, Institut Pasteur, 75015 Paris, France
[5]These authors contributed equally
[6]Lead Contact
*Correspondence: tg2@princeton.edu
https://doi.org/10.1016/j.cell.2018.09.056

SUMMARY

How transcriptional bursting relates to gene regulation is a central question that has persisted for more than a decade. Here, we measure nascent transcriptional activity in early *Drosophila* embryos and characterize the variability in absolute activity levels across expression boundaries. We demonstrate that boundary formation follows a common transcription principle: a single control parameter determines the distribution of transcriptional activity, regardless of gene identity, boundary position, or enhancer-promoter architecture. We infer the underlying bursting kinetics and identify the key regulatory parameter as the fraction of time a gene is in a transcriptionally active state. Unexpectedly, both the rate of polymerase initiation and the switching rates are tightly constrained across all expression levels, predicting synchronous patterning outcomes at all positions in the embryo. These results point to a shared simplicity underlying the apparently complex transcriptional processes of early embryonic patterning and indicate a path to general rules in transcriptional regulation.

INTRODUCTION

A central question in gene regulation concerns how discrete molecular interactions generate a continuum of expression levels observed at the transcriptome level (Lionnet and Singer, 2012; Scholes et al., 2017). A large set of molecular activities is required to elicit RNA transcription, including transcription factor binding, chromatin modifications, and long-range enhancer-promoter interactions (Voss and Hager, 2014). However, in most cases, it is unclear which of these interactions predominantly regulate RNA synthesis rates and variability for a given gene (Coulon et al., 2013). In general, for genes whose transcription rates depend on levels of external inputs, we do not know which regulatory steps are preferably tuned to achieve required mRNA expression levels. Overall, it is unknown whether constraints exist that might select common mechanisms for modulating transcriptional activity across genes, space, and time.

Addressing these questions requires measuring the kinetic rates of transcription in absolute units. Many studies using single-molecule counting approaches have documented the inherently stochastic nature of transcription (Little et al., 2013; Raj et al., 2006; Taniguchi et al., 2010; Zenklusen et al., 2008). In organisms ranging from bacteria to vertebrates, genes exhibit transcription bursts characterized by intermittent intervals of mRNA production followed by protracted quiescent periods (Bothma et al., 2014; Golding et al., 2005; Suter et al., 2011). This inherent stochasticity in gene activation results in higher cell-to-cell variability than expected from constitutive expression (Blake et al., 2003). A simple telegraph or two-state model has been used to explain the measured variability in the context of transcriptional bursting (Peccoud and Ycart, 1995). In this model, a locus switches at random between inactive and active states, with only the latter permitting transcription initiation. Despite its prevalent use, it is not largely known which molecular events determine the kinetic rates of this model (Coulon et al., 2013), nor is it widely understood which of these kinetic rates are modulated by external input signals or to what extent. However, with precise measurements and quantitative modeling, it is possible to gain intuition for the mechanisms of transcriptional bursting based on their signature in the measured variability (Jones et al., 2014; Larson et al., 2013; Molina et al., 2013; Senecal et al., 2014; Zoller et al., 2015).

Drosophila embryos provide an ideal model to investigate transcriptional regulation (Gregor et al., 2014). Early embryos express many genes in graded patterns in response to modulatory inputs (Struhl et al., 1992). Spatial domains, where gene expression levels transition from highly active to nearly silent, are functionally the most critical for the developing embryo, as they determine specification of cell identities (Kornberg and Tabata, 1993). Among the earliest expressed genes in *Drosophila* development are the gap genes, which encode transcription factors responsible for anterior-posterior (AP) patterning (Jaeger, 2011). Each gap gene is expressed in its own unique domain, and the expression boundaries arise at distinct and precise positions (Dubuis et al., 2013). Gene expression levels are spatially graded across several cell diameters, and the intermediate levels

of these gap genes confer patterning information necessary for segmentation (Lawrence, 1992). Therefore, the precise control of expression levels is essential for properly patterned cell fate specification.

The regulation of gap genes appears highly complex. Many activating and repressing factors determine expression boundaries through complex layers of homo- and heterotypic protein interactions at multiple promoters and enhancers (Estrada et al., 2016; Jaeger et al., 2004; Kvon et al., 2014; Perry et al., 2011; Segal et al., 2008). The collective activity of these factors generates expression rates that vary with position in the embryo (Briscoe and Small, 2015; Lawrence, 1992; Manu et al., 2009). Given the diversity of *cis*-regulatory architecture and *trans*-acting factors regulating these genes, an intuitive expectation is that expression rates emerge from carefully tuned transcription factor concentrations and binding affinities. Since various bursting kinetics could achieve such rates, a straightforward prediction is that the underlying bursting kinetics will differ between boundaries. This expectation is consistent with prior studies in cultured cells suggesting that many regulatory strategies exist (Carey et al., 2013; Dey et al., 2015; Molina et al., 2013; Senecal et al., 2014). However, it is unknown how bursting rates are modulated across multiple expression boundaries in intact tissues.

To address these questions, we developed a single-molecule fluorescent *in situ* hybridization (smFISH) method that generates accurate counts of nascent RNA molecules in individual nuclei. We applied this method to assess absolute transcriptional activity of the gap genes in terms of the number and variability of RNA polymerase II (Pol II) molecules at transcribing loci. This approach reveals a common principle that unifies transcriptional activity across expression boundaries. Surprisingly, a single common control parameter globally determines the distribution of transcriptional activity. We use a simple telegraph model to interpret our measurements. We show that the key regulatory parameter is the fraction of time a gene is in a transcriptionally active state, while the Pol II initiation rate is constant. Contrary to the expectation of diverse bursting kinetics, the promoter switching rates are tightly constrained across boundaries. This constraint highlights the conservation of the switching correlation time and predicts synchronous transcriptional outcomes regardless of expression level, gene identity, or position in the embryo. We propose that this synchronicity is important for ensuring precise patterning. Moreover, our results suggest an emergent simplicity in the modulation of bursting that governs the apparently complex process of embryo segmentation. Overall, our quantitative approach provides a framework for uncovering unifying principles of transcriptional regulation that can be applied across genes in any biological context.

RESULTS

Precise Measurements of Transcriptional Activity

During early fly development, gene expression boundaries arise from spatially varying transcription factor concentrations. Early embryos thus provide a natural context in which to ask how input factors shape transcription dynamics. Here, we enhanced a previously developed smFISH method (Little et al., 2013) to yield a 3- to 4-fold increase in sensitivity, enabling precise counting of nascent transcripts and measurement of transcriptional activity across boundaries (STAR Methods). We performed confocal imaging with fluorescent oligonucleotide probes to label single mRNA molecules in fixed embryos followed by analysis to estimate intensities of transcription sites (i.e., spatially co-localized nascent transcripts) and individual cytoplasmic mRNAs. This method measures instantaneous activity per nucleus in terms of intensity units of individual cytoplasmic mRNAs, the "cytoplasmic unit" (C.U.) by normalizing the total intensity of each locus to that of cytoplasmic mRNAs (Figures 1A and 1B).

We measured the transcriptional activity of the four major gap genes *hunchback* (*hb*), *Krüppel* (*Kr*), *knirps* (*kni*), and *giant* (*gt*) along the embryo's AP axis. These genes are expressed early in development in broad spatial domains, permitting measurements of thousands of synchronized nuclei across small numbers of embryos; these factors all favor low measurement error (Figures 1C and 1D; N $\sim$ 15 embryos per combination of gene/genotype). Analysis of expression levels in mid- to late interphase 13 ensures sufficient time to attain steady-state levels of transcribing Pol II (Figures S1A–S1D and STAR Methods), and DNA replication occurs in early interphase (Blumenthal et al., 1974) such that these observations eliminate ambiguity arising from varying numbers of loci. Since loci on recently duplicated chromatids are often closely apposed in space, we measure total transcription per nucleus (Little et al., 2013) then infer properties of individual loci. As a control, we generated data from embryos heterozygous for a *hb* deficiency and observed half the wild-type level of expression per nucleus (Figure 1C). Importantly, we observe a corresponding decrease in variance to half of wild-type (Figure 1D), supporting previous findings that all loci behave independently (Little et al., 2013). These results demonstrate the suitability of using total transcriptional activity per nucleus to infer the behavior of individual loci.

Since biological variance greatly constrains models of regulatory processes, we needed to determine how variability arises from measurement error, embryo-to-embryo differences, and intrinsic fluctuations in individual nuclei. The performance of our measurements was assessed by labeling each mRNA in alternating colors along the length of the strand. This allowed us to perform independent normalization in each channel, thus characterizing sources of measurement error, such as noise stemming from imaging and normalization (Figure 2A). Estimation of the variance of the mean across embryos (Figure 2B) enables further splitting of the variability in terms of embryo alignment along the AP axis and inherent embryo-to-embryo variability (Figures S1E–S1H and STAR Methods). For all genes and at all positions, measurement variability (imaging and spatial alignment) represents less than 7% of the total variance on average (Figure 2C), indicating that biological variability dominates our measurements (Dubuis et al., 2013). Importantly, this variability arises almost entirely from differences between nuclei rather than differences between embryos (Figure 2D); the low embryo-to-embryo variability in the maximally expressed regions (16 $\pm$ 4% coefficient of variation [CV]; Figure 2E) emphasizes that the mean expression levels across embryos are reproducible in absolute units (Figure 1C). Therefore, the measured expression noise mainly stems from zygotic transcription and is intrinsic to the molecular processes of transcription rather

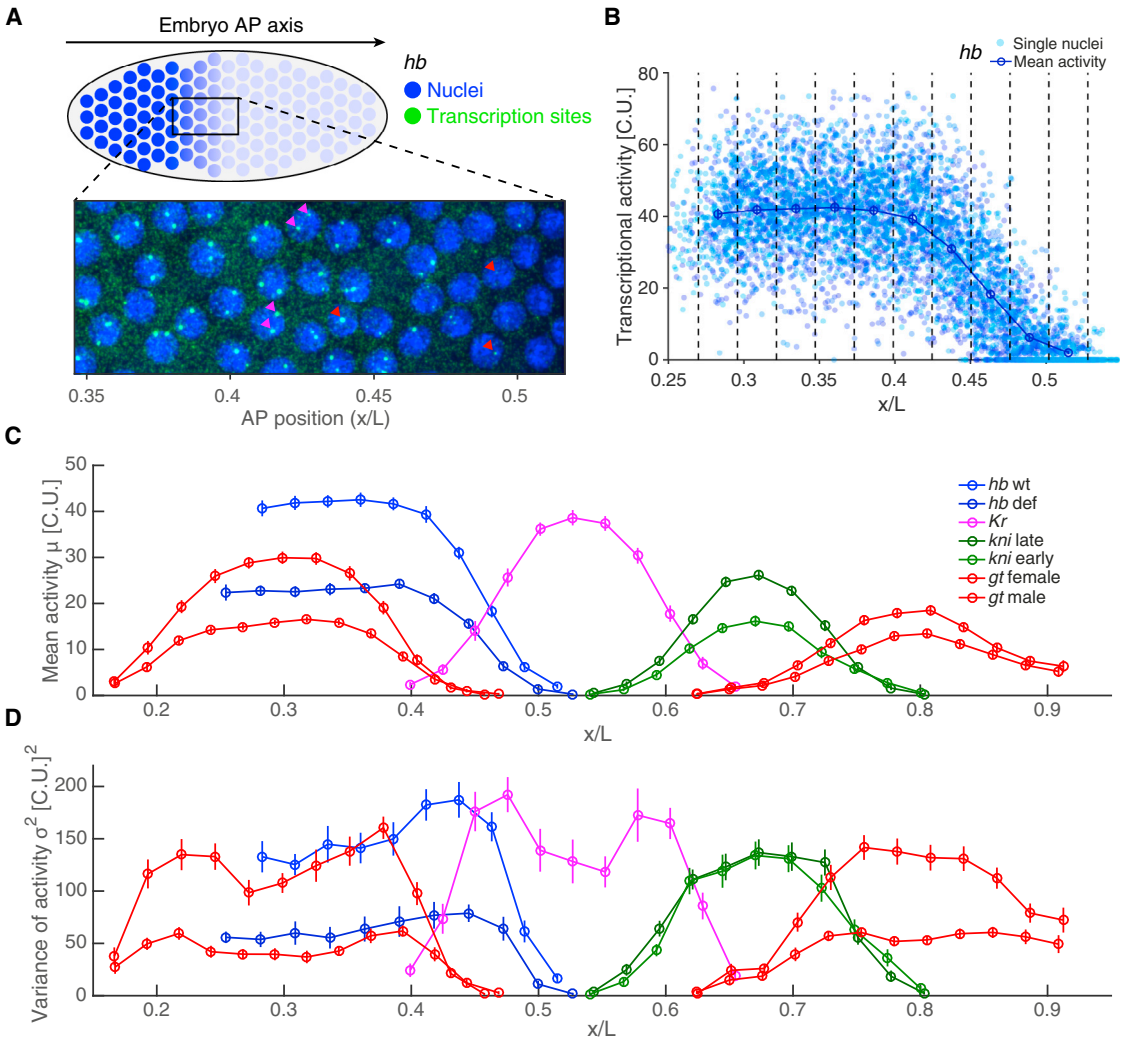


Figure 1. Absolute Quantification of Gap Gene Transcriptional Activity

(A) Activity of individual nuclei (blue) for the gene *hunchback* (*hb*) measured by single-molecule mRNA-FISH (green) in nuclear cycle 13 (nc13) of the blastoderm embryo of length *L*. Red arrowheads: nuclei with two sites of transcription; magenta arrowheads: single site of transcription.

(B) Activity profile of *hb* as a function of AP position in % egg length for 18 embryos. Activity of individual nuclei from the summed intensity of all transcription sites per nucleus, normalized by the average intensity of a single cytoplasmic mRNA (C.U.), is shown. Blue dots: total mean intensity per nucleus; vertical dashed lines: AP bins; circles: mean activity in each bin.

(C) Mean activity in C.U. as a function of AP position during nc13 for *hunchback* in wild-type (labeled *hb* wt in blue, N = 18 embryos), *hunchback* deficiency with half the *hb* dosage (*hb* def, light blue, N = 7), *Krüppel* (*Kr*, magenta, N = 11), *knirps* during early (*kni* early, green, N = 14) and late (*kni* late, light green, N = 16) nc13, *giant* in females with two alleles (*gt* female, red, N = 20) and in males with one (*gt* male, light red, N = 16).

(D) Total variance of transcriptional activity as a function of AP position (color code as in C).

All error bars are 68% confidence intervals. See also Figure S1.

than from extrinsic sources of variability. Low measurement error and the predominance of intrinsic variability facilitates analysis of the noise-mean relationship, permitting inference of bursting kinetics from several hundred nuclei at each position along the AP axis (Figure 1B) as detailed below.

Single-Parameter Distribution of Transcriptional Activity across All Expression Boundaries

The expression patterns of the gap genes are determined by multiple enhancer elements at varying distances from their promoters (Kvon et al., 2014; Perry et al., 2011). Each enhancer contains a variable number of binding sites for multiple patterning input factors with cross-regulatory interactions (Ochoa-Espinosa et al., 2005; Schroeder et al., 2004). These features and evidence from genetic manipulations (Hoch et al., 1990; Jacob et al., 1991; Pankratz et al., 1992) indicate that many molecular processes regulate transcription rates generating observed mRNA levels with their stereotypical modulation as a function of position (Figure 1C). Given the diversity of input factors and molecular control elements, it would appear likely that different

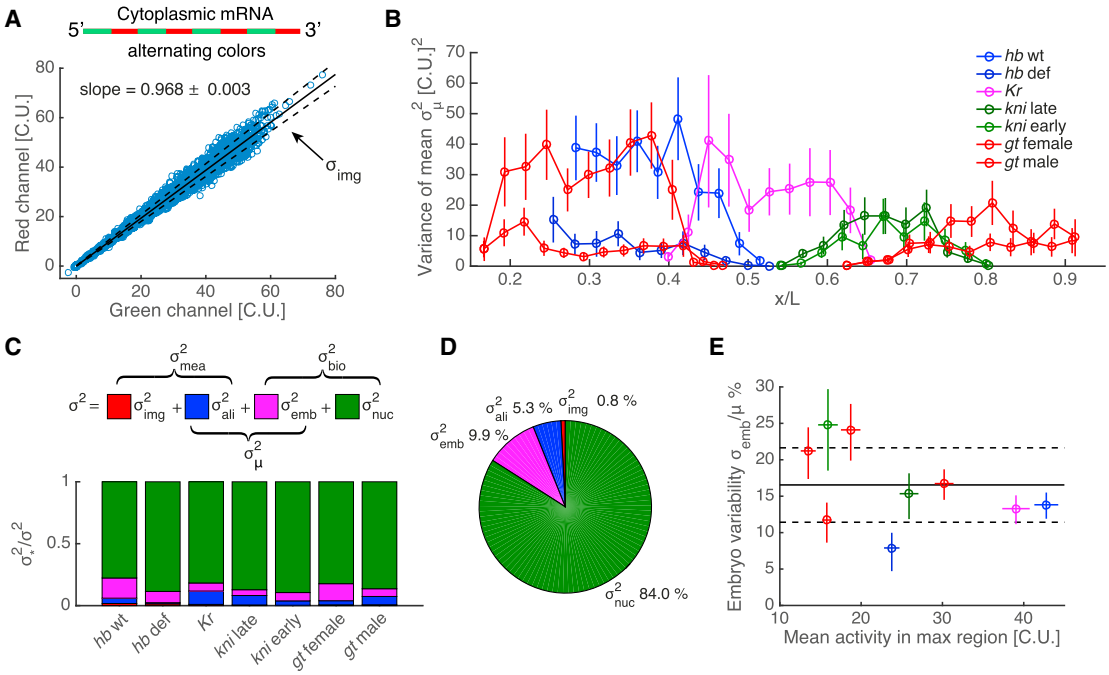


Figure 2. Decomposition of the Total Variance

(A) Imaging noise estimation with dual-color smFISH. mRNA molecules are tagged with an alternating probe configuration. Blue circles: activity of single nuclei in 15 *hb* embryos. In absence of measurement noise or normalization error, both channels should perfectly correlate with slope 1. We characterized the spread along the fitted line (solid line), assuming error in both channels. Dashed lines: 1σ envelope.

(B) Variance of the mean σ_μ^2 across embryos as a function of AP position (color code as in Figure 1C).

(C) Decomposition of the total variance σ^2 into measurement error and biological variability. Estimates of imaging error (red), alignment error (blue), and embryo-to-embryo variability (magenta) are decoupled from the total variance. The remaining variance corresponds to biological variability and is termed intrinsic nucleus-to-nucleus variability in the text (green).

(D) Decomposition of total variance for all the genes pooled together. Nucleus-to-nucleus variability dominates (~84%).

(E) Fractional embryo-to-embryo variability (CV) as a function of mean activity (solid black line: mean ratio; dashed lines: 68% confidence intervals) reaches 16 ± 4% (CV) in the maximally expressed regions that are the most reproducible. This represents absolute reproducibility, as all embryos peak at comparable means.

Error bars are 68% confidence intervals. See also Figure S1.

genes should exhibit vastly different, uniquely defined transcriptional kinetics. To make progress in understanding these complex relationships, we capitalize on the fact that the kinetics of the processes underlying transcription determines not only mean expression levels but also the variability (Figure 1D). Therefore, we can use the noise-mean relationship to characterize the transcription kinetics for individual genes.

To characterize noise-mean relationships in our system, we examined the dependence of variability on mean transcription levels (Figure 3A). In agreement with prior measurements (Little et al., 2013), genes span a similar dynamic range of expression levels across boundaries, from nearly zero to a maximum value of 34 ± 6 C.U. (Figure 1C). Moreover, transcription is inherently variable: at all positions and for all genes, variability exceeds that expected from a simple model of constitutive activity, with noise (measured as CV^2) approximately 10 times larger than Poisson for mean transcriptional activity below 10 C.U. (Figure 3A). However, the noise-mean relationship follows an unexpectedly similar overall trend (Figure 3A and STAR Methods). Unlike many other systems (bacteria, yeast, mammalian cell culture), there is no clearly identifiable noise floor at high expression (Keren et al., 2015; Taniguchi et al., 2010; Zoller et al., 2015). The absence of such an extrinsic noise floor is likely a key feature of early embryo development: nuclei are highly synchronized within the cell cycle and share the same environment of the syncytial blastoderm. Sources of extrinsic noise that affect gene expression in cultured cells are thus minimized. Moreover, the collapse on a unique curve is unexpected and atypical given the different promoter-enhancer architectures (Hornung et al., 2012; Sanchez and Golding, 2013).

This result is even more striking when we convert our units of transcriptional activity from C.U. to the actual number count of Pol II molecules, *g*. Such a conversion is necessary, as the intensity at a given active transcription locus is dependent on the length of the individual gene, the copy number, and the probe arrangement (Figure S2A and STAR Methods). Accounting for these factors, we can describe the shape of the distribution of Pol II counts per locus by calculating the second, third, and fourth cumulants for each gene across each boundary. While again, the expectation is that Pol II counts should differ between different genes, an extra data collapse is observed instead: the second, third, and fourth cumulants for all data points are nearly uniquely determined by a single parameter, the mean activity $\langle g \rangle$ (Figures 3B–3D and S2B–S2D). Thus, transcriptional activity for

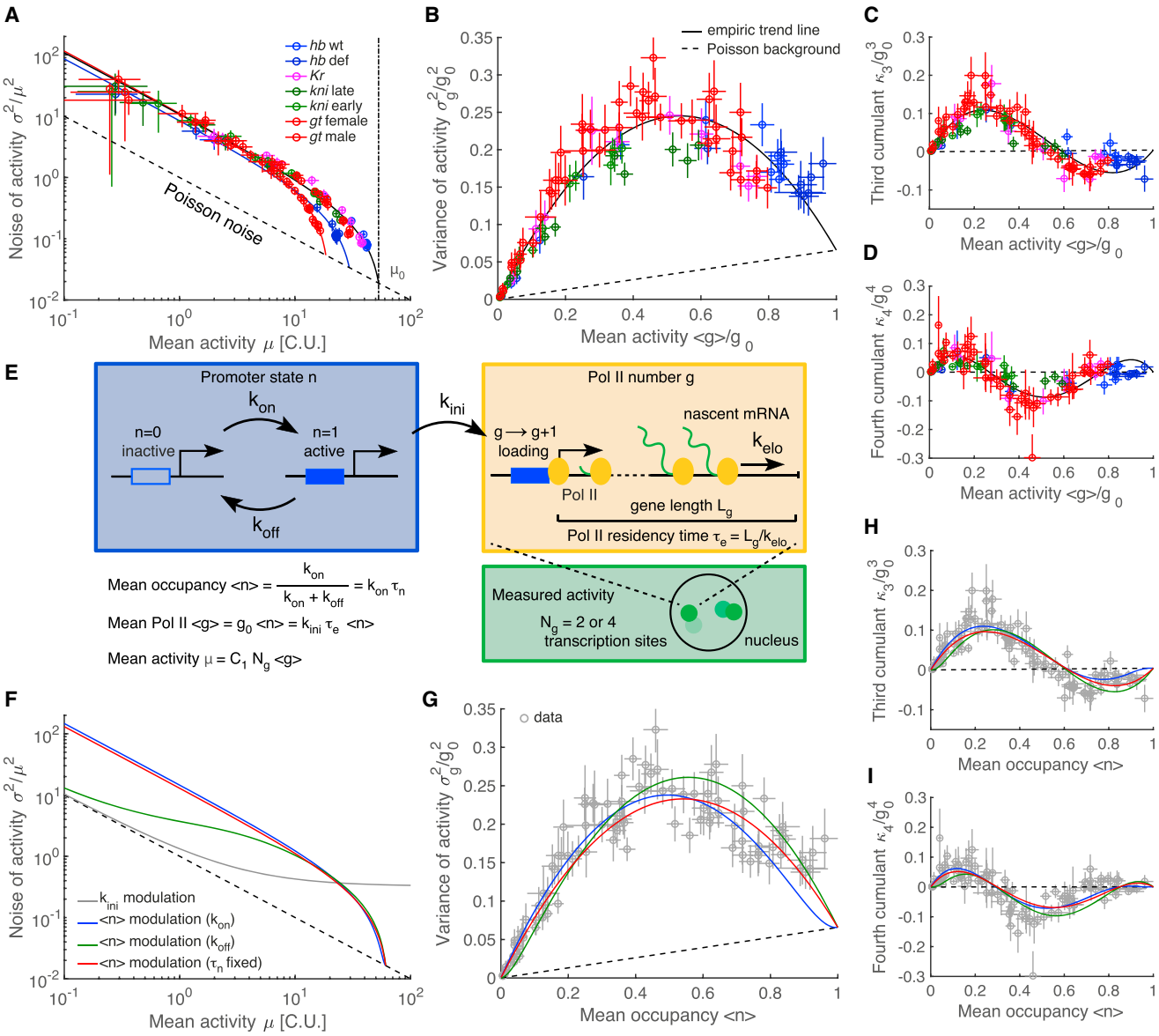


Figure 3. A Two-State Model Recapitulates Data Collapse and Single-Parameter Modulation

(A) Noise-mean relationship (noise = CV^2). Dashed line: Poisson background, the lowest attainable noise; solid lines: fitting for each gene the following functional form of the noise $CV^2 = (1 + a(1 - \mu/\mu_0))/\mu$, where a and μ_0 are fitted parameters. The collapse of the trend to Poisson noise ($1/\mu$) at high expression implies an upper limit of attainable expression levels, μ_0 (vertical dashed line). Color code as in Figure 1C.

(B–D) Normalized second (B), third (C), and fourth (D) cumulant as a function of normalized Pol II counts for a single gene copy. Activities in C.U. were converted into Pol II counts g by using fluorescent probe locations and gene lengths L_g. Assuming independence, the mean and the cumulant were divided by the gene copy number $N_g = 2, 4$. Dashed lines: Poisson background; solid lines: fitting the cumulants with second-, third-, and fourth-order polynomials, respectively, constrained to match the Poisson level at maximum Pol II counts $g_0 = \mu_0\langle L_g\rangle/(C_1 N_g L_g)$, where $\langle L_g\rangle$ is the average gene length and $C_1 \in [0, 1]$ is a conversion factor that depends on the probe locations on transcripts. Color code as in Figure 1C.

(E) Two-state model for measured transcriptional activity. The mean activity in Pol II counts is $\langle g\rangle = k_{ini}\tau_e\langle n\rangle$ with initiation rate k_{ini}, elongation time $\tau_e = L_g/k_{elo}$, and mean promoter activity $\langle n\rangle = k_{on}/(k_{on} + k_{off}) \in [0, 1]$. The maximal Pol II count is given by $g_0 = k_{ini}\tau_e$. The measured mean activity in C.U. is $\mu = C_1 N_g\langle g\rangle$, where N_g is the gene copy number and $C_1 \in [0, 1]$ a conversion factor as in (B–D).

(F–I) Noise-mean relationship (F) and normalized second (G), third (H), and fourth (I) cumulants predicted by the two-state model under different single-parameter mean activity modulation schemes: Pol II initiation rate k_{ini} (gray), off rate k_{off} (green), on rate k_{on} (blue), and promoter occupancy $\langle n\rangle$ at constant switching correlation time $\tau_n = 1/(k_{on} + k_{off})$ (red). Modulation of $\langle n\rangle$ by means of k_{on} or at constant τ_n achieve numerical values that closely match the trends of our data. Error bars are 68% confidence intervals. See also Figures S2 and S3.

Table 1. Terminology and Parameterization of Transcription Rates

Kinetic Rates	Units	Parameterization $\{k_{ini}, k_{on}, k_{off}\}$	Parameterization $\{k_{ini}, \langle n \rangle, \tau_n\}$
Pol II initiation rate k_{ini}	$[min^{-1}]$	k_{ini}	k_{ini}
Promoter switching on rate k_{on}	$[min^{-1}]$	k_{on}	$\frac{\langle n \rangle}{\tau_n}$
Promoter switching off rate k_{off}	$[min^{-1}]$	k_{off}	$\frac{1-\langle n \rangle}{\tau_n}$
Bursting Parameters	**Units**	**Parameterization $\{k_{ini}, k_{on}, k_{off}\}$**	**Parameterization $\{k_{ini}, \langle n \rangle, \tau_n\}$**
Promoter mean occupancy $\langle n \rangle$	#	$\frac{k_{on}}{(k_{on}+k_{off})} = k_{on}\,\tau_n$	$\langle n \rangle$
Switching correlation time τ_n	[min]	$\frac{1}{(k_{on}+k_{off})}$	τ_n
Burst size b	#	$\frac{k_{ini}}{k_{off}}$	$\frac{k_{ini}\tau_n}{1-\langle n \rangle}$
Burst frequency f	$[min^{-1}]$	$\frac{k_{on}\,k_{off}}{(k_{on}+k_{off})} = k_{off}\,\langle n \rangle$	$\frac{\langle n \rangle(1-\langle n \rangle)}{\tau_n}$
Mean transcript synthesis rate	$[min^{-1}]$	$\frac{k_{ini}\,k_{on}}{(k_{on}+k_{off})}$	$k_{ini}\langle n \rangle$

Within the context of the two-state model, the most intuitive parameterization is given by the kinetic rates k_{ini}, k_{on}, and k_{off}. However, fluctuation analysis in transcriptional activity and inference approach both revealed that the three independent and uncorrelated variables, k_{ini}, τ_n, and $\langle n \rangle$, provide a more natural parameterization in which only $\langle n \rangle$ is modulated, while k_{ini} and τ_n are both constant. Bursting parameters are clearly identified in both parameterizations.

all genes and across the entire expression range is characterized by a unique, common single-parameter distribution. This observation is model-free and indicates that a single parameter determines the generation of all gene expression boundaries. The uniqueness of the Pol II count distribution suggests that despite the well-documented diversity of *cis*-regulatory elements and *trans*-acting factors, a common conserved set of processes is regulated to determine transcription kinetics across all boundaries in the early embryo.

Two-State Model Identifies the Unique Control Parameter

The shared Pol II count distribution suggests that a common general model can describe the regulation of all gap genes. The observed intrinsic super-Poissonian variability in our data suggests that these genes operate in a bursting regime. While constitutive genes can be modeled by a single parameter—i.e., the effective initiation rate—multiple independent parameters are required to model transcription kinetics of bursting genes. A popular minimalist model accounting for bursting is the "two-state" or "telegraph" model (Peccoud and Ycart, 1995). It has been widely used to describe the distribution of mature mRNA and protein counts (Bar-Even et al., 2006; Raj et al., 2006; Zenklusen et al., 2008). Such a simple mechanistic model enables estimation of kinetic rates underlying bursting (Figure 3E and Table 1), i.e., the switching rates between promoter states (k_{on} and k_{off}), as well as the effective initiation rate k_{ini} (Larson et al., 2013; Senecal et al., 2014; Suter et al., 2011).

Our measurements of nascent transcriptional activity represent instantaneous counts of the number of Pol II molecules engaged in transcription, providing a more direct measurement of transcriptional activity compared to counts of mature mRNAs or proteins. The two-state model presents a straightforward and parameter-sparse means to describe how discrete randomly occurring events generate a continuum of expression rates. Assuming the Pol II elongation rate k_{elo} is constant and identical for all gap genes (Garcia et al., 2013; O'Brien and Lis, 1993), this model predicts the dependence of variability on mean activity for different scenarios of parameter modulation. Specifically, it predicts which kinetic rates are modulated to form gene expression boundaries.

Given that the first four cumulants of our data are uniquely determined by the mean activity, we sought to explore modulation of the mean arising from varying a single parameter, where such parameters could consist of combinations of the kinetic rates. When we solve the master equation for such a model (STAR Methods), a comparison of predicted noise (Figure 3F) with our data (Figure 3A) eliminates modulation of k_{ini}. Indeed, solely varying k_{ini} leads to saturation of noise at high activity, which is not observed. This is true no matter the values of k_{on} and k_{off}, which only affect the level of the plateau. Instead, our measurements are consistent with modulation of the fractional mean promoter occupancy $\langle n \rangle$, defined as $\langle n \rangle = k_{on}/(k_{on} + k_{off})$. (Here, occupancy refers to the active or "ON" state; thus $\langle n \rangle$ is bound between zero and one.) This value is the fraction of time spent in the active state and is equivalent to the probability of finding a locus in the active state (Lucas et al., 2013; Xu et al., 2015). Varying $\langle n \rangle$ is the only solution leading to a concave function for the variance observed in the data (Figures 3B and 3G and STAR Methods, Equation 8). Modulation of the mean production rate is thus determined by $\langle n \rangle$ rather than the rate at which Pol II molecules enter into productive elongation.

In principle, either or both of the rates k_{on} and k_{off} may be tuned to modulate $\langle n \rangle$. To test which of these scenarios reproduces the noise and the shape of the cumulants (Figures 3A–3D), we first set the value of k_{ini} to match the Poisson background in the data (Figure 3B, dashed line; and STAR Methods). For the special case in which both switching rates are modulated simultaneously, we achieved effective single-parameter modulation by fixing the switching correlation time $\tau_n = 1/(k_{on} + k_{off})$, the characteristic timescale for changes in promoter activity. This quantity reveals how fast the switching occurs, how much time is required for the mean number of Pol II molecules engaged in transcription to reach steady state, and what fraction of the switching noise is filtered by the elongation process (STAR Methods). When τ_n is fixed, both switching rates, k_{on} and k_{off}, are fully determined by $\langle n \rangle$, i.e.,

$$k_{on} = \frac{\langle n \rangle}{\tau_n} \quad \text{and} \quad k_{off} = \frac{(1 - \langle n \rangle)}{\tau_n}$$

In the three scenarios (tuning k_{on}, k_{off}, or $\langle n \rangle$), the single free parameter (either k_{off}, k_{on} or τ_n) was estimated by fitting the set of modeled cumulants to the data, assuming steady-state Pol II levels (Figures 3G–3I and S2E). Modulation of k_{off} alone is ruled out, since this does not capture the noise below 10 C.U. (Figure 3F). However, modulation of k_{on} alone or of $\langle n \rangle$ at fixed τ_n recapitulates the noise and the cumulants (Figures 3F–3I). Thus, in addition to conserved k_{ini}, the model predicts a second conserved quantity across genes and positions—either k_{off} alone or a combination of k_{on} and k_{off}.

Finally, we examined whether the fitted cumulants assuming steady state are compatible with the finite duration of the nuclear cycle (~15 min). The time during which a gene relaxes from an inactive state devoid of elongating Pol II (start of interphase 13) to steady state is determined by the correlation time τ_n (Figure S3A). Since each parameter modulation predicts a different dependency of τ_n on $\langle n \rangle$ (Figure S3B), we tested under each scenario whether the mean and the cumulants at mid-cycle would be attained in time. It follows that modulation of $\langle n \rangle$ through k_{on} alone or at fixed τ_n predicts a time-dependent solution at mid-cycle that is consistent with the steady-state assumption above (Figures S3C–S3G and STAR Methods). Thus, the two-state model explains the data collapse and predicts that tuning only the mean occupancy $\langle n \rangle$ uniquely describes the formation of expression boundaries regardless of their position in the embryo.

Transcriptional Bursting in Absolute Units

Further insight into transcriptional mechanisms requires the absolute scales of kinetic parameters. To go beyond arguments based on cumulants, we adopted an approach that is agnostic to the modulation strategy. To resolve whether k_{off} or τ_n is constant and exclude other non-trivial forms of modulation (i.e., multiple rates changing simultaneously), we inferred all kinetic rates from the full distribution of transcriptional activity for each gene and at each position independently. We performed dual-color smFISH, tagging the 5′ and 3′ regions of the transcripts with differently colored probe sets that provide two complementary readouts of nascent activity (Figure 4A) (Brody et al., 2011; Xu et al., 2016). The measured activities are correlated via a finite Pol II elongation time (Figures S4A–S4C and STAR Methods) and provide two snapshots of the state of the gene. Jointly measuring the 5′ and 3′ activities constrains the possible configurations of nascent transcripts and Pol II configurations at each locus (Figure 4B).

Given a stochastic model of transcription, it is possible to extract the transcriptional parameters underlying the activities of each gene (Figures 4C, 4D, and S5). Using the two-state model, we calculated the likelihood of the joint distribution of 5′ and 3′ activities at each AP position while accounting for measurement noise (Figure 4E and STAR Methods). The rate parameters k_{ini}, k_{on}, and k_{off} for each AP position were inferred from the likelihood of the data according to Bayes's rule. We sampled the joint posterior distribution of the parameters (Hastings, 1970), which provides a probability for each parameter combination given the observed data. All inferred parameters with respective errors were estimated from the sampled joint posterior distribution (Figures 4E and S5C). Validating our approach, inference on synthetic data clearly shows that the parameters are identifiable as long as the Pol II elongation rate is measured independently (Figures S6A–S6F). Moreover, the previously measured Pol II elongation rate $k_{elo} = 1.5$ kb/min (Garcia et al., 2013) provides an absolute timescale, enabling inference of endogenous kinetics from chemically cross-linked, inert embryos.

The inferred kinetic rates revealed nearly identical modulation across all expression boundaries, regardless of gene identity or boundary position (Figure 5). Consistent with predictions based on cumulants (Figure 3), the initiation rate k_{ini} is constant at 7.2 ± 1.0 Pol II initiations per minute and does not change across genes or positions (Figure 5A). Thus, while in the ON state, these genes share the same rate-limiting step(s) in the cascade of molecular interactions leading to productive Pol II elongation as reported for constitutive genes (Choubey et al., 2015). We also observe close agreement between measured and inferred mean activity, as well as good agreement between all other cumulants (Figures S6G–S6J). Our inference confirms that all expression boundaries are generated through modulation of the mean promoter occupancy (Figure 5B). This result supports the view that the processes that determine k_{ini} are disfavored as mechanisms for controlling mRNA synthesis rates. Because these rates are determined by $\langle n \rangle$ for all genes and span a similar dynamic range for all boundaries (Figure S6K), we advocate that promoter occupancy represents the key control parameter describing expression boundary formation.

Current models of boundary formation imply a careful tuning of multiple input factor concentrations and DNA binding affinities (Briscoe and Small, 2015; Jaeger, 2011). The complexity and diversity of these inputs leads to an intuitive expectation that kinetic switching rates will differ between genes. This expectation seems all the more reasonable given that many combinations of k_{on} and k_{off} generate the same $\langle n \rangle$. Surprisingly, both k_{on} and k_{off} are tightly constrained for all genes and across all boundaries when portrayed as a function of mean occupancy $\langle n \rangle$ (Figures 5C and 5D). This suggests that some combination of k_{on} and k_{off} must be conserved. Indeed, as predicted by the cumulant analysis above, our measurements confirm that the conserved combination is in fact the correlation time of the switching

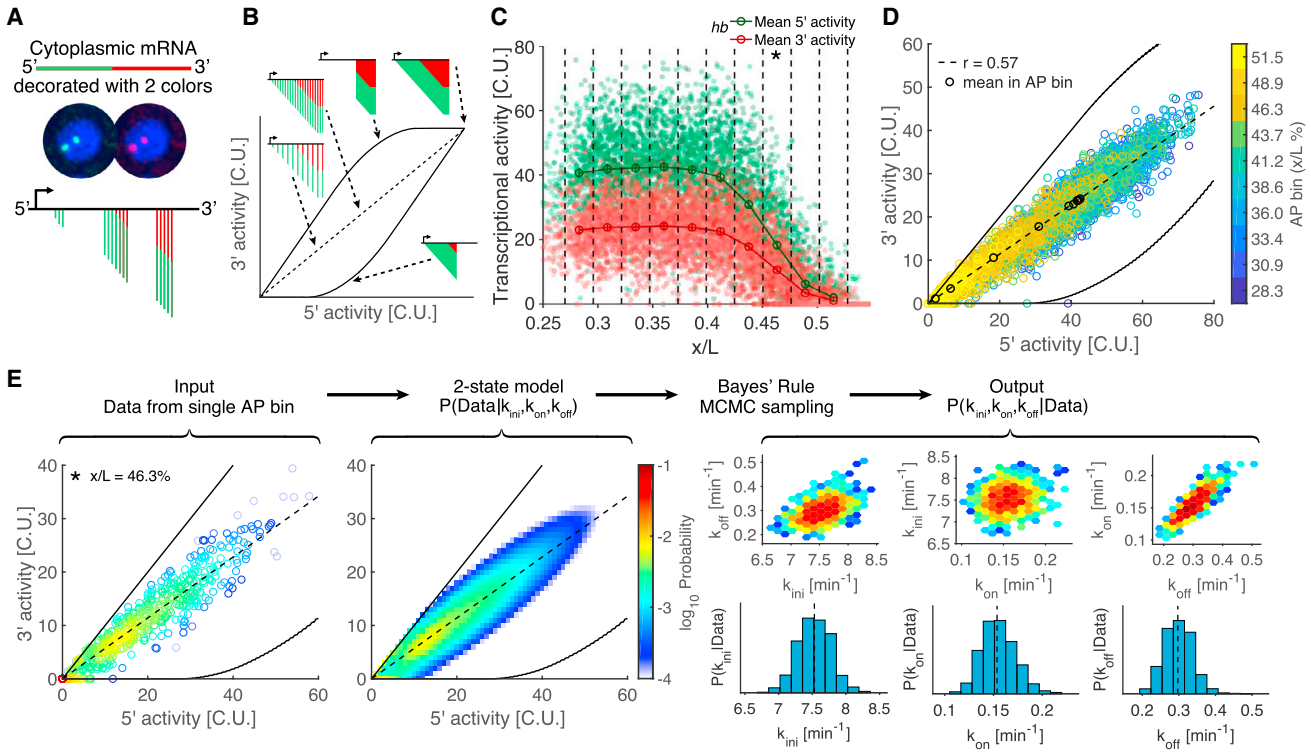


Figure 4. Estimation of Transcription Parameters via Dual-Color smFISH

(A) Schematic of the dual-color single-molecule mRNA-FISH technique. Two independent probe sets hybridized to different fluorophore target the 5′ (green) and 3′ region (red). The combination of readouts constrains the possible configurations of nascent transcript locations and numbers.

(B) Dual-color smFISH measurement space represented as 5′ versus 3′ activity. Solid black line: border of possible measurements given probe set configuration, gene length, and maximal possible Pol II density (here, we assumed a Pol II holoenzyme footprint of 90 bp); dashed black line: expected ratio of 3′ to 5′ activity defining the subset of configurations for which nascent transcripts are equally spaced along the entire gene length but at different densities.

(C) Activity profile for *hb* as a function of AP position for both 5′ and 3′ channels. Dots: total intensity of nascent transcripts in C.U. in a single nucleus. N = 18 embryos aligned and overlaid. Vertical dashed lines: AP bins; circles: mean activity in each bin; error bars: 68% confidence intervals.

(D) Empirical distributions of 5′ versus 3′ activity for *hb*; colored circles: individual nuclei. Color code represents different AP bins. Black circles: mean of each AP bin (see B). The measurements are enclosed by the envelope of maximal Pol II density (black line as in B).

(E) Inference of parameters defined by the two-state model. Parameters are estimated from the empirical distribution individually at each AP bin (the data, C and D). We calculated the likelihood of the data given a set of parameters $P(\text{Data} \mid k_{\text{ini}}, k_{\text{on}}, k_{\text{off}})$. By applying Bayes's rule, we obtained the posterior probability $P(k_{\text{ini}}, k_{\text{on}}, k_{\text{off}} \mid \text{Data})$, the probability of the parameters given the observed data; the posterior probability was sampled by Monte-Carlo Markov chain (MCMC). Final estimates of the parameters are given by the median of the marginal posterior (vertical dashed line in histogram). The color code of the distributions stands for the $\log_{10}$ probability.

See also Figures S4, S5, and S6.

process $\tau_n = 1/(k_{\text{on}} + k_{\text{off}})$, which is roughly constant at all positions over the entire expression range for every gene (Figure 5E).

Our inference thus revealed that the more natural parameterization of this system is expressed in terms of the three independent, uncorrelated variables $\{k_{\text{ini}}, \tau_n, \langle n \rangle\}$ in which only $\langle n \rangle$ is modulated (Table 1). The conservation of correlation time implies that k_{on} and k_{off} must be carefully coordinated such that all boundaries emerge from quantitatively identical modulation of switching rates. In addition, these conclusions are unaffected by changes in elongation rate, which only rescales the kinetic parameters (Figures S6L–S6N and STAR Methods).

Our observation of constant k_{ini} and τ_n has several implications. Much prior work has characterized bursting in terms of burst size $b = k_{\text{ini}}/k_{\text{off}}$ (the average number of transcripts produced per burst) and burst frequency $f = \langle n \rangle \cdot k_{\text{off}}$ (which reduces to k_{on} for short burst durations, i.e., small $\langle n \rangle$) (Dar et al., 2012; Dey et al., 2015). Interestingly, by virtue of the constancy of k_{ini} and τ_n, at high activity ($\langle n \rangle > 0.5$), mainly the burst size changes (Figure 5F), while for $\langle n \rangle < 0.5$, it is the burst frequency that changes (Figure 5G). These results recapitulate recent observations of frequency modulation (Bartman et al., 2016; Fukaya et al., 2016; Larson et al., 2013; Li et al., 2018; Senecal et al., 2014) and might explain previously observed global trends in burst size (Sanchez and Golding, 2013).

Provided all genes become transcriptionally competent at the same time following mitosis (Blythe and Wieschaus, 2015, 2016), the conserved correlation time we measure here implies that all genes reach steady state simultaneously (Figures S3C and S3F). Consistent with prior observations (Dubuis et al., 2013; Garcia

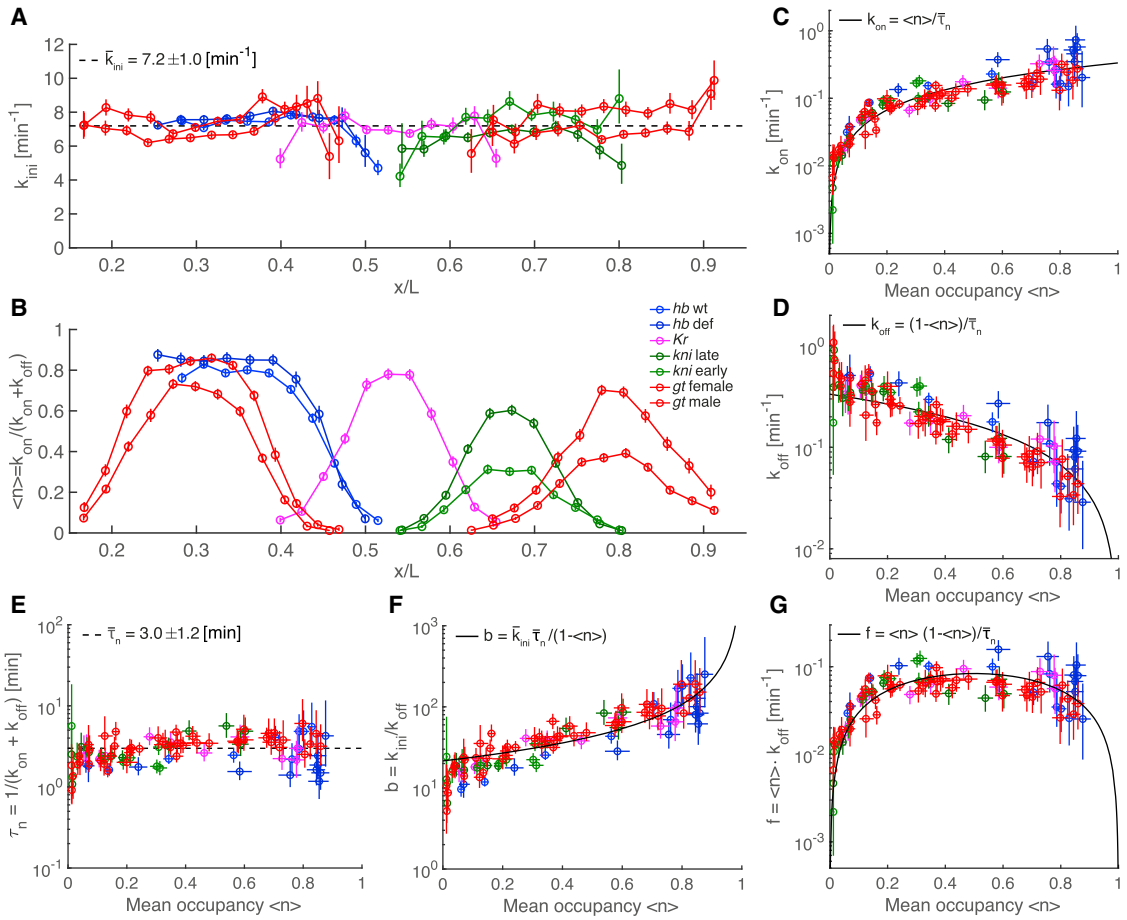


Figure 5. Inferred Transcription Parameters Are Tightly Constrained across Gap Genes

(A and B) Inferred Pol II initiation rate k_{ini} (A) and promoter mean occupancy $\langle n \rangle$ (B) for all genes across AP position.

(C and D) Inferred on-rate k_{on} (C) and off-rate k_{off} (D) as a function of mean occupancy $\langle n \rangle$ for all genes. Solid black lines represent the global trend using the mean value of τ_n (see formula in inset).

(E) Inferred switching correlation time τ_n as a function of mean occupancy $\langle n \rangle$ for all genes with a mean value of $\bar{\tau}_n = 3.0 \pm 1.2$ min (dashed line).

(F and G) Inferred burst size b (F) and burst frequency f (G) as a function of the mean occupancy $\langle n \rangle$. Solid black lines represent the global trend using the mean value of k_{ini} and τ_n (see formula in inset).

Color code as in Figure 1C. Error bars are the 10th to 90th percentiles of the posterior distribution. See also Figure S6.

et al., 2013), synchronicity suggests that the relative mean synthesis rates are maintained (i.e., unmodulated) across the patterning boundaries during early development (Figure S3F). In addition, a short correlation time ($\bar{\tau}_n = 3.0 \pm 1.2$ min, small relative to ~15 min duration of interphase 13) ensures effective temporal averaging of the switching noise by accumulation of stable transcripts, further suggesting that both expression timing and noise minimization jointly constrain switching rates. These dynamic constraints may be essential for precise and reproducible patterning outcomes, affecting the range of permissible values of k_{on} and k_{off}. Together, these results show that for the gap genes, the apparently complex process of regulating expression rates is explained by a conceptually simple, shared modulation strategy of bursting kinetics. Our approach opens a path to uncovering general principles to unify the regulation of transcription across genes.

DISCUSSION

A multitude of processes influence eukaryotic transcription rates. It is not clear which events might be more likely than others to determine the kinetics of bursting—either globally or in a gene specific manner, nor is it known how bursting kinetics compare across endogenous genes over a range of expression levels. Our quantitative bursting measurements reveal that all gap gene expression boundaries arise from the same underlying kinetics regardless of the differences in regulatory elements. Thus, from the complex combination of diverse interactions specific to each gene emerges a simple, common strategy for transcriptional regulation.

Our recognition of shared regulation surfaced only upon development of a highly precise single-molecule method of quantification. Conclusions about bursting depend heavily upon

understanding sources and extent of measurement error and minimizing variability from extrinsic sources. Extrinsic processes, such as cell growth and division, DNA duplication, and mRNA transport and decay, can significantly affect the apparent variability between cells and thus also bursting rates (Battich et al., 2015; Bahar Halpern et al., 2015; Zopf et al., 2013). We minimize these effects by measuring transcription at nascent sites in an endogenous system with synchronized cell divisions. Moreover, explicit quantification of measurement error resulted in a noise model that significantly constrained our inference framework. All these approaches are generally applicable to enable precise quantification in any system.

The fundamental mean-cumulant relationships we uncovered demonstrate that a single-parameter distribution globally determines transcriptional activity (Figures 3B–3D). Employing the telegraph model (Peccoud and Ycart, 1995), we find that the modulation of mean occupancy $\langle n \rangle$ predicts mean mRNA synthesis rates comparable with previous measurements (Figure S6O) (Garcia et al., 2013) and reproduces the distribution of nascent activity (Figures S6G–S6J), whereas k_{ini} and τ_n are conserved. The global behavior we observe is surprising, given that bursting is generally believed to be gene and promoter specific. Multiple factors and processes, including enhancer-promoter interactions, chromatin context, nucleosome occupancy, Pol II pausing, and transcription factor interactions, all impinge on bursting rates (Bartman et al., 2016; Brown and Boeger, 2014; Carey et al., 2013; Dar et al., 2012; Dey et al., 2015; Fukaya et al., 2016; Molina et al., 2013; Senecal et al., 2014; Suter et al., 2011; Weinberger et al., 2012; Zenklusen et al., 2008). It remains to be determined whether the same processes are modulated in the same manner or, conversely, whether different regulatory strategies have converged to generate identical transcriptional activity across genes.

These observations raise the question of whether the common transcriptional bursting kinetics carry a functional advantage (Eldar and Elowitz, 2010). In early embryos, the precise positioning of cell fates requires minimizing variability between nuclei, which is achieved by a combination of long mRNA lifetimes permitting accumulation and spatial averaging through the syncytial cytoplasm (Little et al., 2013). In principle, modulating k_{ini} at a constitutive promoter would generate the theoretical minimal (Poisson) transcriptional noise at all levels (Sanchez et al., 2013). The fact that neither constitutive activity ($\langle n \rangle \leq 0.85$) nor Pol II saturation ($k_{elo}/k_{ini} \sim 215$ bp $\gg$ Pol II footprint) is ever observed suggests that some constraint prohibits this system from maintaining a continuous active state and/or it is not straightforward to alter k_{ini}. Instead, a constant switching correlation time suggests that this value is important in facilitating robust patterning. We propose that both expression timing and noise minimization jointly constrain switching rates.

The mechanistic origins of the conserved parameters are unknown. One possibility is that protein-DNA affinities have been individually selected to confer the switching rates we observe. However, it is unclear how transient transcription factor interactions, usually on the order of seconds, could generate bursts on the order of minutes (Elf et al., 2007; Izeddin et al., 2014; Karpova et al., 2008). Another possibility is that the fast transcription factor binding kinetics are masked by the slower dynamics of common general factors involved in the transcription process. In fact, recent evidence suggests that mediator and TATA-binding protein binding, as well as the core promoter and its shape, play a key role in bursting (Li et al., 2018; Schor et al., 2017; Tantale et al., 2016). Alternatively, processes of potentially even slower dynamics, such as long-range enhancer-promoter interactions, chromatin modification, or Pol II pausing, may determine common bursting kinetics (Chen et al., 2018; Henriques et al., 2018; Nicolas et al., 2018).

The observed constancy of τ_n will guide further modeling and identification of the molecular mechanisms. This constancy is connected to the binomial noise level (STAR Methods, Equation 8). Extensions of the two-state model must provide similar filtering of the binomial noise, which will restrict the possible class of models. For example, we tested two particular extensions of the two-state model. One possibility is a three-state model consisting of a two-step reversible activation (Rieckh and Tkačik, 2014). Alternatively, a model with an additional noise term, such as input noise stemming from input transcription factor diffusion (Kaizu et al., 2014; Tkačik et al., 2008), could explain dual modulation of switching rates observed under the two-state model. However, distinguishing these models will require live imaging.

The common transcriptional parameters of the gap genes highlight a form of complexity reduction: despite the variety of upstream regulatory elements, all expression boundaries result from similar bursting kinetics. Whether this signature results from an underlying molecular simplicity has yet to be determined. Regardless of the mechanistic means by which these similarities are achieved, the convergence suggests the general constraints that limit the range of permitted bursting rates and/or minimize transcription variability. The unexpected conservation of the initiation rate and the correlation time might indicate a path to general rules in transcriptional regulation. It is now possible to inquire about the breadth of these generalities and whether they apply to the same gene expressed in different cell types, to the transcriptome as a whole, or even across organisms. Indeed, it appears plausible that other classes of genes share similarly constrained bursting kinetics (Sanchez and Golding, 2013). The methods we utilize here are applicable in a variety of systems and permit the discovery of the molecular mechanism(s) conferring unified transcription kinetics.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCE TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Fly strains
- METHOD DETAILS
  - DNA oligonucleotides
  - smFISH protocol
  - Imaging
  - Image analysis

- Calibration in absolute units
- Measurement error
- Single parameter distribution of transcriptional activity
- Two-state model of transcriptional activity
- Cumulant analysis
- Inferring transcription kinetics of endogenous genes from dual color smFISH

- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures, three tables, and one data file and can be found with this article online at https://doi.org/10.1016/j.cell.2018.09.056.

ACKNOWLEDGMENTS

We thank C. Bartman, W. Bialek, P. Francois, M. Levo, J. Mozziconacci, F. Naef, A. Raj, T. Sokolowski, G. Tkačik, and E. Wieschaus for insightful discussion and valuable comments on the manuscript. B.Z. was partially supported by the Swiss National Science Foundation early Postdoc Mobility fellowship. This study was funded by grants from the National Institutes of Health (U01 EB021239, U01 DA047730, and R01 GM097275) and the National Science Foundation (PHY-1734030).

AUTHOR CONTRIBUTIONS

Conceptualization, methodology, formal analysis, investigation, writing—original draft, writing—review & editing, funding acquisition, resources, and supervision: B.Z., S.C.L., and T.G.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 5, 2018
Revised: July 2, 2018
Accepted: September 26, 2018
Published: October 18, 2018

REFERENCES

Bahar Halpern, K., Tanami, S., Landen, S., Chapal, M., Szlak, L., Hutzler, A., Nizhberg, A., and Itzkovitz, S. (2015). Bursty gene expression in the intact mammalian liver. Mol. Cell *58*, 147–156.

Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O'Shea, E., Pilpel, Y., and Barkai, N. (2006). Noise in protein expression scales with natural protein abundance. Nat. Genet. *38*, 636–643.

Bartman, C.R., Hsu, S.C., Hsiung, C.C.S., Raj, A., and Blobel, G.A. (2016). Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. Mol. Cell *62*, 237–247.

Battich, N., Stoeger, T., and Pelkmans, L. (2015). Control of Transcript Variability in Single Mammalian Cells. Cell *163*, 1596–1610.

Blake, W.J., KAErn, M., Cantor, C.R., and Collins, J.J. (2003). Noise in eukaryotic gene expression. Nature *422*, 633–637.

Blumenthal, A.B., Kriegstein, H.J., and Hogness, D.S. (1974). The units of DNA replication in Drosophila melanogaster chromosomes. Cold Spring Harb. Symp. Quant. Biol. *38*, 205–223.

Blythe, S.A., and Wieschaus, E.F. (2015). Zygotic genome activation triggers the DNA replication checkpoint at the midblastula transition. Cell *160*, 1169–1181.

Blythe, S.A., and Wieschaus, E.F. (2016). Establishment and maintenance of heritable chromatin structure during early *Drosophila* embryogenesis. eLife *5*, 1752–1765.

Bothma, J.P., Garcia, H.G., Esposito, E., Schlissel, G., Gregor, T., and Levine, M. (2014). Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos. Proc. Natl. Acad. Sci. USA *111*, 10598–10603.

Briscoe, J., and Small, S. (2015). Morphogen rules: design principles of gradient-mediated embryo patterning. Development *142*, 3996–4009.

Brody, Y., Neufeld, N., Bieberstein, N., Causse, S.Z., Böhnlein, E.M., Neugebauer, K.M., Darzacq, X., and Shav-Tal, Y. (2011). The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. PLoS Biol. *9*, e1000573.

Brown, C.R., and Boeger, H. (2014). Nucleosomal promoter variation generates gene expression noise. Proc. Natl. Acad. Sci. USA *111*, 17893–17898.

Carey, L.B., van Dijk, D., Sloot, P.M.A., Kaandorp, J.A., and Segal, E. (2013). Promoter sequence determines the relationship between expression level and noise. PLoS Biol. *11*, e1001528.

Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J.B., and Gregor, T. (2018). Dynamic interplay between enhancer-promoter topology and gene activity. Nat. Genet. *50*, 1296–1303.

Choubey, S., Kondev, J., and Sanchez, A. (2015). Deciphering Transcriptional Dynamics In Vivo by Counting Nascent RNA Molecules. PLoS Comput. Biol. *11*, e1004345.

Coulon, A., Chow, C.C., Singer, R.H., and Larson, D.R. (2013). Eukaryotic transcriptional dynamics: from single molecules to cell populations. Nat. Rev. Genet. *14*, 572–584.

Dar, R.D., Razooky, B.S., Singh, A., Trimeloni, T.V., McCollum, J.M., Cox, C.D., Simpson, M.L., and Weinberger, L.S. (2012). Transcriptional burst frequency and burst size are equally modulated across the human genome. Proc. Natl. Acad. Sci. USA *109*, 17454–17459.

Dubuis, J.O., Samanta, R., and Gregor, T. (2013). Accurate measurements of dynamics and reproducibility in small genetic networks. Mol. Syst. Biol. *9*, 639.

Eldar, A., and Elowitz, M.B. (2010). Functional roles for noise in genetic circuits. Nature *467*, 167–173.

Elf, J., Li, G.-W., and Xie, X.S. (2007). Probing transcription factor dynamics at the single-molecule level in a living cell. Science *316*, 1191–1194.

Estrada, J., Wong, F., DePace, A., and Gunawardena, J. (2016). Information Integration and Energy Expenditure in Gene Regulation. Cell *166*, 234–244.

Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer Control of Transcriptional Bursting. Cell *166*, 358–368.

Fukaya, T., Lim, B., and Levine, M. (2017). Rapid Rates of Pol II Elongation in the Drosophila Embryo. Curr. Biol. *27*, 1387–1391.

Garcia, H.G., Tikhonov, M., Lin, A., and Gregor, T. (2013). Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning. Curr. Biol. *23*, 2140–2145.

Gillespie, D.T. (1977). Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. *81*, 2340–2361.

Golding, I., Paulsson, J., Zawilski, S.M., and Cox, E.C. (2005). Real-time kinetics of gene activity in individual bacteria. Cell *123*, 1025–1036.

Gregor, T., Garcia, H.G., and Little, S.C. (2014). The embryo as a laboratory: quantifying transcription in Drosophila. Trends Genet. *30*, 364–375.

Hastings, W.K. (1970). Monte Carlo Sampling Methods Using Markov chains and Their Applications. Biometrika *57*, 97–109.

Henriques, T., Scruggs, B.S., Inouye, M.O., Muse, G.W., Williams, L.H., Burkholder, A.B., Lavender, C.A., Fargo, D.C., and Adelman, K. (2018). Widespread transcriptional pausing and elongation control at enhancers. Genes Dev. *32*, 26–41.

Hoch, M., Schröder, C., Seifert, E., and Jäckle, H. (1990). cis-acting control elements for Krüppel expression in the Drosophila embryo. EMBO J. *9*, 2587–2595.

Hornung, G., Bar-Ziv, R., Rosin, D., Tokuriki, N., Tawfik, D.S., Oren, M., and Barkai, N. (2012). Noise-mean relationship in mutated promoters. Genome Res. *22*, 2409–2417.

Izeddin, I., Récamier, V., Bosanac, L., Cissé, I.I., Boudarene, L., Dugast-Darzacq, C., Proux, F., Bénichou, O., Voituriez, R., Bensaude, O., et al. (2014). Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. eLife *3*, e02230.

Jacob, Y., Sather, S., Martin, J.R., and Ollo, R. (1991). Analysis of Krüppel control elements reveals that localized expression results from the interaction of multiple subelements. Proc. Natl. Acad. Sci. USA *88*, 5912–5916.

Jaeger, J. (2011). The gap gene network. Cell. Mol. Life Sci. *68*, 243–274.

Jaeger, J., Surkova, S., Blagov, M., Janssens, H., Kosman, D., Kozlov, K.N., Manu, Myasnikova, E., Vanario-Alonso, C.E., Samsonova, M., et al. (2004). Dynamic control of positional information in the early Drosophila embryo. Nature *430*, 368–371.

Jones, D.L., Brewster, R.C., and Phillips, R. (2014). Promoter architecture dictates cell-to-cell variability in gene expression. Science *346*, 1533–1536.

Kaizu, K., de Ronde, W., Paijmans, J., Takahashi, K., Tostevin, F., and ten Wolde, P.R. (2014). The Berg-Purcell limit revisited. Biophys. J. *106*, 976–985.

Karpova, T.S., Kim, M.J., Spriet, C., Nalley, K., Stasevich, T.J., Kherrouche, Z., Heliot, L., and McNally, J.G. (2008). Concurrent fast and slow cycling of a transcriptional activator at an endogenous promoter. Science *319*, 466–469.

Keren, L., van Dijk, D., Weingarten-Gabbay, S., Davidi, D., Jona, G., Weinberger, A., Milo, R., and Segal, E. (2015). Noise in gene expression is coupled to growth rate. Genome Res. *25*, 1893–1902.

Kornberg, T.B., and Tabata, T. (1993). Segmentation of the Drosophila embryo. Curr. Opin. Genet. Dev. *3*, 585–594.

Kvon, E.Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J.O., Pagani, M., Schernhuber, K., Dickson, B.J., and Stark, A. (2014). Genome-scale functional characterization of Drosophila developmental enhancers in vivo. Nature *512*, 91–95.

Larson, D.R., Fritzsch, C., Sun, L., Meng, X., Lawrence, D.S., and Singer, R.H. (2013). Direct observation of frequency modulated transcription in single cells using light activation. eLife *2*, e00750.

Lawrence, P.A. (1992). The making of a fly: The genetics of animal design.

Lestas, I., Paulsson, J., Ross, N.E., and Vinnicombe, G. (2008). Noise in Gene Regulatory Networks. IEEE Trans. Automat. Control *53*, 189–200.

Li, C., Cesbron, F., Oehler, M., Brunner, M., and Höfer, T. (2018). Frequency Modulation of Transcriptional Bursting Enables Sensitive and Rapid Gene Regulation. Cell Syst. *6*, 409–423.

Lionnet, T., and Singer, R.H. (2012). Transcription goes digital. EMBO Rep. *13*, 313–321.

Little, S.C., Tikhonov, M., and Gregor, T. (2013). Precise developmental gene expression arises from globally stochastic transcriptional activity. Cell *154*, 789–800.

Lucas, T., Ferraro, T., Roelens, B., De Las Heras Chanes, J., Walczak, A.M., Coppey, M., and Dostatni, N. (2013). Live imaging of bicoid-dependent transcription in Drosophila embryos. Curr. Biol. *23*, 2135–2139.

Manu, Surkova, S., Spirov, A.V., Gursky, V.V., Janssens, H., Kim, A.R., Radulescu, O., Vanario-Alonso, C.E., Sharp, D.H., Samsonova, M., et al. (2009). Canalization of gene expression in the Drosophila blastoderm by gap gene cross regulation. PLoS Biol. *7*, e1000049.

Molina, N., Suter, D.M., Cannavo, R., Zoller, B., Gotic, I., and Naef, F. (2013). Stimulus-induced modulation of transcriptional bursting in a single mammalian gene. Proc. Natl. Acad. Sci. USA *110*, 20563–20568.

Munsky, B., and Khammash, M. (2006). The finite state projection algorithm for the solution of the chemical master equation. J. Chem. Phys. *124*, 044104.

Nicolas, D., Zoller, B., Suter, D.M., and Naef, F. (2018). Modulation of transcriptional burst frequency by histone acetylation. Proc. Natl. Acad. Sci. USA *115*, 7153–7158.

O'Brien, T., and Lis, J.T. (1993). Rapid changes in Drosophila transcription after an instantaneous heat shock. Mol. Cell. Biol. *13*, 3456–3463.

Ochoa-Espinosa, A., Yucel, G., Kaplan, L., Pare, A., Pura, N., Oberstein, A., Papatsenko, D., and Small, S. (2005). The role of binding site cluster strength in Bicoid-dependent patterning in Drosophila. Proc. Natl. Acad. Sci. USA *102*, 4960–4965.

Pankratz, M.J., Busch, M., Hoch, M., Seifert, E., and Jäckle, H. (1992). Spatial control of the gap gene knirps in the Drosophila embryo by posterior morphogen system. Science *255*, 986–989.

Peccoud, J., and Ycart, B. (1995). Markovian Modeling of Gene-Product Synthesis. Theor. Popul. Biol. *48*, 222–234.

Perry, M.W., Boettiger, A.N., and Levine, M. (2011). Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. Proc. Natl. Acad. Sci. USA *108*, 13570–13575.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. PLoS Biol. *4*, e309.

Rieckh, G., and Tkačik, G. (2014). Noise and information transmission in promoters with multiple internal States. Biophys. J. *106*, 1194–1204.

Sanchez, A., and Golding, I. (2013). Genetic determinants and cellular constraints in noisy gene expression. Science *342*, 1188–1193.

Sánchez, A., and Kondev, J. (2008). Transcriptional control of noise in gene expression. Proc. Natl. Acad. Sci. USA *105*, 5081–5086.

Sanchez, A., Choubey, S., and Kondev, J. (2013). Regulation of noise in gene expression. Annu. Rev. Biophys. *42*, 469–491.

Scholes, C., DePace, A.H., and Sánchez, Á. (2017). Combinatorial Gene Regulation through Kinetic Control of the Transcription Cycle. Cell Syst. *4*, 97–108.

Schor, I.E., Degner, J.F., Harnett, D., Cannavò, E., Casale, F.P., Shim, H., Garfield, D.A., Birney, E., Stephens, M., Stegle, O., and Furlong, E.E. (2017). Promoter shape varies across populations and affects promoter evolution and expression noise. Nat. Genet. *49*, 550–558.

Schroeder, M.D., Pearce, M., Fak, J., Fan, H., Unnerstall, U., Emberly, E., Rajewsky, N., Siggia, E.D., and Gaul, U. (2004). Transcriptional control in the segmentation gene network of Drosophila. PLoS Biol. *2*, E271.

Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U., and Gaul, U. (2008). Predicting expression patterns from regulatory sequence in Drosophila segmentation. Nature *451*, 535–540.

Senecal, A., Munsky, B., Proux, F., Ly, N., Braye, F.E., Zimmer, C., Mueller, F., and Darzacq, X. (2014). Transcription factors modulate c-Fos transcriptional bursts. Cell Rep. *8*, 75–83.

Dey, S.S., Foley, J.E., Limsirichai, P., Schaffer, D.V., and Arkin, A.P. (2015). Orthogonal control of expression mean and variance by epigenetic features at different genomic loci. Mol. Syst. Biol. *11*, 806.

Sidje, R.B. (1998). Expokit: A Software Package for Computing Matrix Exponentials. ACM Trans. Math. Softw. *24*, 130–156.

Struhl, G., Johnston, P., and Lawrence, P.A. (1992). Control of Drosophila body pattern by the hunchback morphogen gradient. Cell *69*, 237–249.

Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics. Science *332*, 472–474.

Taniguchi, Y., Choi, P.J., Li, G.W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science *329*, 533–538.

Tantale, K., Mueller, F., Kozulic-Pirher, A., Lesne, A., Victor, J.-M., Robert, M.-C., Capozi, S., Chouaib, R., Bäcker, V., Mateos-Langerak, J., et al. (2016). A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. Nat. Commun. *7*, 12248.

Tkačik, G., Gregor, T., and Bialek, W. (2008). The role of input noise in transcriptional regulation. PLoS ONE *3*, e2774.

Voss, T.C., and Hager, G.L. (2014). Dynamic regulation of transcriptional states by chromatin and transcription factors. Nat. Rev. Genet. *15*, 69–81.

Weinberger, L., Voichek, Y., Tirosh, I., Hornung, G., Amit, I., and Barkai, N. (2012). Expression noise and acetylation profiles distinguish HDAC functions. Mol. Cell *47*, 193–202.

Xu, H., Sepúlveda, L.A., Figard, L., Sokac, A.M., and Golding, I. (2015). Combining protein and mRNA quantification to decipher transcriptional regulation. Nat. Methods *12*, 739–742.

Xu, H., Skinner, S.O., Sokac, A.M., and Golding, I. (2016). Stochastic Kinetics of Nascent RNA. Phys. Rev. Lett. *117*, 128101–128106.

Zenklusen, D., Larson, D.R., and Singer, R.H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. Nat. Struct. Mol. Biol. *15*, 1263–1271.

Zoller, B., Nicolas, D., Molina, N., and Naef, F. (2015). Structure of silent transcription intervals and noise characteristics of mammalian genes. Mol. Syst. Biol. *11*, 823.

Zopf, C.J., Quinn, K., Zeidman, J., and Maheshri, N. (2013). Cell-cycle dependence of transcription dominates noise in gene expression. PLoS Comput. Biol. *9*, e1003161.

STAR★METHODS

KEY RESOURCE TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Organisms/Strains		
D. melanogaster: Oregon-R, wild-type laboratory stock	N/A	Flybase: FBst1000077
D. melanogaster: chromosomal deletion spanning *hb* w[1118]; Df(3R)BSC477/TM6C, Sb[1] cu[1]	Bloomington *Drosophila* Stock Center	Flybase: FBab0045343 BDSC: 24981
Oligonucleotides		
smFISH probes for *hb*, see Table S1	This paper	N/A
smFISH probes for *Kr*, see Table S1	This paper	N/A
smFISH probes for *kni*, see Table S1	This paper	N/A
smFISH probes for *gt*, see Table S1	This paper	N/A
Software and Algorithms		
FiSH Toolbox	Little et al. (2013)	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Thomas Gregor (tg2@princeton.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Fly strains

Oregon-R (Ore-R) embryos were used as wild-type. Embryos heterozygous for a deficiency spanning *hb* were collected from crosses of heterozygous adults of the strain w^{1118}; Df(3R)BSC477/TM6C. Heterozygotes of the *hb* deficiency, as well as wild-type male and female embryos stained for *gt*, were distinguished from siblings by visual inspection of nascent transcription sites.

METHOD DETAILS

DNA oligonucleotides

Oligonucleotide sequences complementary to the open reading frames of each gene of interest were chosen using the Biosearch Technologies Stellaris RNA FISH probe designer (https://www.biosearchtech.com/support/tools/design-software/stellaris-probe-designer). Amine-modified oligonucleotides were obtained from Biosearch Technologies, chemically coupled to NHS-ester-Atto565 (Sigma-Aldrich; 72464) or -Atto633 (Sigma-Aldrich; 01464) and purified by HPLC. Probes are listed in Table S1.

smFISH protocol

We modified our smFISH protocol (Little et al., 2013) to minimize background and maximize signal. Embryos were crosslinked in 1xPBS containing 16% paraformaldehyde for 2 min before devitellinization. Embryos were washed four times in methanol, 5 min per wash, with gentle rocking at room temperature, followed by an extended 30-60 min wash in methanol. Fixed embryos were then used immediately for smFISH without intervening storage. Embryos washed three times in 1X PBS, 5 min per wash, at room temperature with rocking. Embryos were then washed 3 times in smFISH wash buffer (Little et al., 2013), 10 min per wash, at room temperature. During this time, probes diluted in hybridization buffer (Little et al., 2013) were preheated to 37°C. Hybridization was performed for 1.5 hr at 37C with vigorous mixing every 15 min. During hybridization, smFISH wash buffer was preheated to 37C. Embryos were washed four times with large excess volumes of wash buffer for 3-5 min per wash, rinsed twice briefly in PBS, stained with DAPI, and mounted in VECTASHIELD (Vector Laboratories; H-1000). Imaging was performed within 48 hr to ensure high quality signal.

Imaging

Imaging was performed by laser-scanning confocal microscopy on a Leica SP5 inverted microscope. We used a 63x HCX PL APO CS 1.4 NA oil immersion objective with pixels of 76×76 nm^2 and z spacing of 340 nm. We typically obtained stacks representing 8μm in total axial thickness starting at the embryo surface. The microscope was equipped with "HyD Hybrid Detector" avalanche

photodiodes (APDs) that we utilized in photon counting mode. This is in contrast to our prior approach (Little et al., 2013) in which standard photomultiplier tubes (PMTs) were used to collect two separate smFISH image stacks at two different laser intensities: a low power stack for measuring transcription intensities, and a high power stack to distinguish single mRNAs. The use of low-noise photon-counting APDs in place of standard photomultipliers provided sufficient dynamic range to capture high signal transcription sites and to separate relatively dim cytoplasmic single mRNAs from background fluorescence with a single laser power. This also abrogated the need to calibrate the high- and low-power stacks for comparison. The removal of the calibration step provided an additional reduction in measurement error.

Image analysis

Raw data are processed according to previously developed image analysis pipeline (Little et al., 2013). Briefly, raw images are filtered using a Difference-of-Gaussians (DoG) filter to detect spot objects. A master threshold is applied to separate candidate spots from background. True point-like sources of fluorescence are identified, as they appeared on multiple consecutive z-slices (>3) at the same location. All candidate particles are then labeled as transcription sites, cytoplasmic transcripts or noise based on global thresholds. The threshold separating cytoplasmic transcripts from noise is defined as the bottom of the valley between the two peaks on the particle intensity distribution. The threshold for transcription sites depends both on intensity and position, as transcription sites cluster in z and are enclosed in nuclei (segmented from DAPI staining). Intensity of transcription sites is obtained by integrating the signal over a fixed cylinder volume ($V_s = \pi \times 0.76^2 \times 3.06\ \mu m^3$, determined from the objective's PSF).

Calibration in absolute units

We calibrated the integrated intensity of transcription sites F_s by first characterizing the relationship between the fluorescence signal and the density of cytoplasmic transcripts. We defined summation volumes in the embryo $(V \cong 3.8 \times 3.8 \times 8\ \mu m^3)$ avoiding region of high tissue deformation and excluding transcription site location. For each summation volume we counted the number of detected cytoplasmic transcripts and integrated the fluorescence intensity. At low count density, the fluorescence per summation volume F scales linearly with density D (Little et al., 2013). Fitting a simple linear relationship $F = \alpha D + \beta$, where β corresponds to background, enables estimation of a scaling factor α to calibrate transcription sites in "cytoplasmic units" (C.U.) for each embryo. Namely, the intensity in C.U. is given by $f = (F_s - bV_s)/\alpha$ where b is the background intensity per pixel in each nucleus. The resulting quantification of transcriptional activity for all gap genes is provided in Supplemental Data.

Measurement error

Embryo staging

In order to assess the timing of the different embryos, we first manually ranked the different embryos based on timing estimation from DAPI staining. We estimated the interphase stage relying on morphological features of the nuclei (shape and textures) in the DAPI channel. We then verified whether accumulation of cytoplasmic mRNAs correlates with our manual ranking (Figure S1A). Both approaches lead to similar results and provide a decent proxy for timing. By comparing the average activity of the different embryos in the maximally expressed regions with the cytoplasmic density, we assessed the effect of timing on the mean activity (Figure S1B). We estimated the Pearson correlation coefficient ρ for the different genes and regions (*gt* anterior and posterior regions). Overall, timing explain up to $\rho^2 = 44\%$ of the embryo variability (defined as the variance of the mean activity among embryos σ_μ^2) in the maximally expressed regions (Figure S1C), with the exception of *kni* that is highly correlated $\rho \sim 0.8$. We thus separated the *kni* embryos in two sub-populations, early and late embryos. We performed the splitting by finding the cytoplasmic density threshold that minimizes the sum of within-population variance in mean activity. We then calculated the staging variability $\sigma_{sta} = \rho\sigma_\mu$, defined as the variability in mean activity explained by timing between late and early embryos (Figure S1D). Given the overall small staging variability <14%, the total mean activity is stable enough to warrant the assumption of steady-state.

Imaging noise model

We quantified measurements noise due to imaging and calibration using a two-color smFISH approach, labeling each mRNA in alternating colors along the length of the mRNA. We included 15 *hb* embryos in the analysis, which corresponds to approximately 4,000 nuclei activity measurements. We then normalized the activity (fluorescence signal) of the nuclei in cytoplasmic units independently in each channel. In absence of noise and provided accurate normalization, both channels would perfectly correlate with slope one. By plotting one channel against the other (Figure 2A), we assessed the slope and characterized the spread of the data along the expected line.

We build a simple effective model to describe measurement noise:

$$P\left(S^{(5)}, S^{(3)} | G^{(5)}, G^{(3)}\right) = \mathcal{N}\left(S^{(5)} \mid \mu = G^{(5)}, \sigma_5^2\left(G^{(5)}\right)\right) \cdot \mathcal{N}\left(S^{(3)} \mid \mu = G^{(3)}, \sigma_3^2\left(G^{(3)}\right)\right) \quad \text{(Equation 1)}$$

where S stands for the fluorescent signal in cytoplasmic units and G the total nascent transcripts (in C.U.) in absence of noise. We assumed that the measurement errors were normally distributed and independent in both channels, which was motivated by the absence of correlation in the background. We further assumed that the variance would depend on activities, consistent with the

increasing spread observed in the data. In order to estimate the variance specific to each channel, we fitted a straight line $y = ax + h$ assuming error on both $x \equiv S^{(5)}$ and $y \equiv S^{(3)}$. We expanded the variance as a function of the scalar projection along the line v:

$$\sigma^2(v) = \sigma_b^2 + b_1 v + b_2 v^2 + \cdots$$

$$v = \frac{x + ay}{\sqrt{1 + a^2}}$$

Assuming the same error along x and y, we then maximized the following likelihood to estimate the parameters $\theta = \{a, h, \sigma_b, b_1, b_2, \cdots\}$:

$$P(\{x_i, y_i\} \mid \theta) = \prod_{i=1}^{N_d} \frac{1}{\sqrt{2\pi\sigma^2(v_i)}} \exp\left(-\frac{(y_i - ax_i - h)^2}{2(1 + a^2)\sigma^2(v_i)} \right)$$

Using the Akaike information criterion, we selected the best model which was parameterized by (a, σ_b, b_1, b_2) with $h = 0$. The best fitting parameters were: $a = 0.968$, $\sigma_b = 4.59 \cdot 10^{-2}$, $b_1 = 9.31 \cdot 10^{-3}$ and $b_2 = 9.23 \cdot 10^{-4}$ (Figure 2A). The variances in the noise measurement model (Equation 1) are then given by:

$$\sigma_5^2(G) = \sigma^2\left(v = \sqrt{G^2 + (aG)^2} \right)$$

$$\sigma_3^2(G) = \sigma^2\left(v = \sqrt{G^2 + (G/a)^2} \right)$$

where $\sigma^2(v) = \sigma_b^2 + b_1 v + b_2 v^2$. The resulting imaging noise is shown in Figure S1E. In the maximally expressed regions, we measure transcriptional activity with an error of 5% and relate it to absolute units with an uncertainty below 3.5% (the largest deviation of the slope 0.968 ± 0.003 from 1). This represents an error reduction by 3- to 4-fold compared to our previous measurements (assuming multiplicative errors; 6% versus 20%) (Little et al., 2013).

Splitting of the total variance

The Anterior-Posterior axis (AP) was determined based on a mid-sagittal elliptic mask of the embryo in the DAPI channel (Little et al., 2013). Position is obtained by registration of high- and low-magnification DAPI images of the surface. We then fitted constrained splines to approximate the mean activity as a function of the AP position. We used different features of the mean profiles such as maxima and inflection points to refine the alignment between the different embryos. Overall, this realignment procedure enables us to estimate an alignment error of the order of 2% egg length.

After alignment, we defined spatial bins along the AP-axis with a width of 2.5% of egg length. Such a width was a good compromise to balance the sampling and binning error. We next sought to decompose the measured total variance of the transcriptional activity σ^2 (Figure 1D) into different components related to imaging, alignment, embryo and nuclei variability (Figures 2B–2D). We first estimated the variability of the mean across embryos σ_μ^2 in each bin (Figure 2B); we split the total variance σ^2 in each bin according to the law of total variance:

$$\sigma^2 = \underbrace{\frac{1}{N_e}\sum_{i=1}^{N_e} \sigma_i^2}_{\sigma_i^2} + \underbrace{\frac{1}{N_e}\sum_{i=1}^{N_e} (\mu_i - \mu)^2}_{\sigma_\mu^2}$$

where N_e is the total number of embryos and μ the global mean.

Next we aimed to determine what fraction of σ_μ^2 is explained by residual misalignment. Assuming that all the variability in the mean at boundaries results from spatial misalignment of the different embryos, one can find an upper bound on the residual alignment error σ_x:

$$\sigma_\mu^2 \geq \sigma_{\text{ali}}^2 = \left(\frac{d\mu}{dx} \right)^2 \sigma_x^2$$

where μ is the global mean profile as function of AP position x. For each gene, we estimated the residual alignment error σ_x required to explain as much embryo variability as possible (Figure S1F, diagonal dash line). Overall we found that σ_x is of the order of 1% egg length. The total embryo variability in the maximally expressed regions cannot be explained by misalignment as $\left(\frac{d\mu}{dx} \right) \approx 0$ and leads to a noise floor (Figure S1F, horizontal dash line). This noise floor can be partly explained by variability in the stage (early versus late

interphase) of the different embryos (Figures S1C and S1D). In the following we thus split $\sigma^2_\mu = \sigma^2_{\text{ali}} + \sigma^2_{\text{emb}}$ where σ^2_{emb} is the residual embryo to embryo variability.

Finally, we assessed what fraction of the total variance σ^2 corresponds to combined measurement noise $\sigma^2_{\text{mea}} = \sigma^2_{\text{img}} + \sigma^2_{\text{ali}}$ where σ^2_{img} was estimated in subsection (STAR Methods, *Imaging noise model*). Total measurement noise σ^2_{mea} remains below 20% of the total variance for all genes and all position (Figure S1G), and on average reaches $6.1 \pm 3.5\%$. The remaining variability corresponds to biological variability $\sigma^2_{\text{bio}} = \sigma^2_{\text{nuc}} + \sigma^2_{\text{emb}}$ where σ^2_{nuc} is the nuclei variability and was defined as:

$$\sigma^2_{\text{nuc}} = \sigma^2 - \sigma^2_{\text{img}} - \sigma^2_{\text{ali}} - \sigma^2_{\text{emb}}$$

Overall, the non-nuclear variability $(\sigma^2_{\text{img}} + \sigma^2_{\text{ali}} + \sigma^2_{\text{emb}})$ remains below 33% of the total variance for all genes and all position (Figure S1H), and on average reaches $16.0 \pm 6.4\%$. Thus, the nuclei variability σ^2_{nuc} largely dominates in our data and represents 84% of the total variance on average (Figures 1E and 1F).

Single parameter distribution of transcriptional activity

Noise-mean relationship in the FISH data

In practice, we measure transcriptional activity in cytoplasmic units (intensity in equivalent number of fully elongated transcripts) and not in Pol II counts g directly. The measured mean activity μ in cytoplasmic units is proportional to the mean Pol II counts for a single gene copy $\langle g \rangle$, i.e., $\mu = C_1 N_g \langle g \rangle$ where $C_1 \in [0, 1]$ is a conversion factor accounting for the FISH probe locations on the gene and N_g the number of gene copies (for most gap genes $N_g = 4$, except for *gt* male and *hb* deficient that only have 2 copies). Assuming independence of loci, the measured variance σ^2follows a similar relationship, i.e., $\sigma^2 = C_2 N_g \sigma^2_g$ with $C_2 \in [0, 1]$. The conversion factors C_1 and C_2 are constants that are unique for each gene and are calculated further (STAR Methods, *Conversion factor for Pol II counts*).

As we will see later (Equation 8), one can derive the following functional form for the variance in Pol II counts for a single gene copy:

$$\sigma^2_g = \langle g \rangle + \langle g \rangle (g_0 - \langle g \rangle) \Phi$$

where g_0 is the maximal mean Pol II counts on the gene that is determined by the Pol II initiation rate k_{ini} and elongation time τ_e, and Φ a quantity that is related to the dynamics of the promoter activity and bounded $\Phi \in [0, 1]$. Of note, $\Phi = 0$ for a constitutively expressed gene such that the variance reduces to $\sigma^2_g = \langle g \rangle$ (Poisson variance). In principle, the values of both g_0 and Φ are gene-specific and could have specific dependency on $\langle g \rangle$. The interpretation of the equation above and the quantities g_0 and Φ will be discussed in greater details later on (STAR Methods, *Two-state model of transcriptional activity*). Using the relationships between the cytoplasmic units and Pol II counts for the mean and variance above, we can express the measured noise as:

$$\frac{\sigma^2}{\mu^2} = \frac{C_2}{C_1}\left(\frac{1}{\mu} + \frac{1}{C_1 N_g}\frac{\mu_0 - \mu}{\mu}\Phi\right)$$

where $\mu_0 = C_1 N_g g_0$ is the maximal mean expression level in cytoplasmic units. In practice, $C_2/C_1 \approx 1$ and $C_1 \approx 0.7$ (Table S2, 5′ probe location) such that the Poisson noise background in cytoplasmic units is approximately $1/\mu$. By setting $C_2/C_1 = 1$, we further simplify the equation above and obtain:

$$\frac{\sigma^2}{\mu^2} = \frac{1}{\mu}(1 + a(1 - \mu/\mu_0)) \qquad \text{(Equation 2)}$$

with $a = g_0 \Phi$. By assuming a and μ_0 constant, we found that the above noise-mean relationship (Equation 2) captures the overall trend in the data well (Figure 3A), with $a = 9.93 \pm 0.35$ and $\mu_0 = 53.07 \pm 1.73$ ($R^2 = 0.99$). Although both *gt* male and *hb* deficient follow a similar trend, they deviate from the black line, ($a = 10.66 \pm 0.35$, $\mu_0 = 18.52 \pm 0.28$) and ($a = 7.68 \pm 1.00$, $\mu_0 = 29.57 \pm 1.59$) respectively. Interestingly, despite the fact that g_0 and Φ could *a priori* be gene-specific, a is roughly conserved across genes and differences in μ_0 can be explained by variation in gene copies ($N_g = 2$ copies for *gt* male and *hb* deficient instead of 4) and gene length (*gt* is shorter than *hb*, Table S3). This suggests that some key quantities underlying transcription are conserved among the gap genes and can be highlighted by proper normalization of the measured activity.

Normalized cumulants for a single gene copy

To further investigate the transcriptional commonalities of the gap genes, we calculated the 2$^{\text{nd}}$, 3$^{\text{rd}}$ and 4$^{\text{th}}$ cumulants from the data (Figures 3B–3D). For independent random variables, the cumulants have the property to be extensive, which is convenient as the measured transcriptional activities result from the sum of 2 or 4 independent gene copies. We first converted the k^{th} cumulants $\tilde{\kappa}_k$ computed from the data in cytoplasmic units to Pol II counts (or number of nascent transcripts) for a single gene copy with a normalized gene length:

$$\kappa_k = \frac{1}{C_k N_g}\left(\frac{\langle L_g \rangle}{L_g}\right)^k \tilde{\kappa}_k$$

where κ_k is the k^{th} cumulant in Pol II counts for a single gene copy, L_g the gene length, N_g the gene copy number (4 for most genes, except *gt* male and *hb* deficient that only have 2 copies) and C_k a conversion factor for the k^{th} cumulant to ensure proper normalization of the Poisson background (Equation 3 and Table S2). The annotated gene length L_g varies between 1.8 to 3.6 kb for the gap genes. In the following we used an effective gene length that is slightly larger and takes into account the possible lingering of fully elongated transcripts at the loci (Table S3). This effective gene length can be estimated from the dual color FISH data (STAR Methods, *Dual color smFISH and effective gene length*). For the normalization, we used a normalized gene length of $\langle L_g \rangle = 3.3$ kb.

We then fitted a second order polynomial of the mean activity $\langle g \rangle$ to the variance σ_g^2 (Figures 3B and S2A) in order to estimate the maximal activity g_0, which was defined as the second crossing point between the Poisson background (Figure S2A dash line) and the fitted variance (solid line). We found $g_0 = 15.21 \pm 0.20$ Pol II for a normalized gene length of 3.3 kb. Similarly, we fitted 3rd and 4th order polynomial of the mean activity to the cumulants $\kappa_3$and κ_4 (Figures 3C and 3D), constrained to reach the Poisson limit at g_0. Of note, the cumulants of the Poisson distribution are all equal to the mean. As we observed in Figures 3B–3D, the polynomial fits (solid lines) capture the main trend observed in the data, suggesting a simple relationship between the cumulants and the mean. It follows that the underlying activity distribution is essentially a universal single parameter distribution whose parameter is the mean activity. To test the extent of the universality, we repeated the analysis above of each gap gene individually (Figures S2B–S2D). The individual fits (colored solid lines) remain relatively close to each other. Although the fits for *hb* slightly deviate from the other genes, the global shape of the cumulants is conserved.

Conversion factor for Pol II counts

As mentioned above, the cumulants of the transcriptional activity in cytoplasmic units are related to the cumulants in number of nascent transcripts or Pol II counts on the gene by conversion factors C_k. We calculated these conversion factors to ensure proper normalization of the Poisson background, meaning that the conversion of cumulants in C.U. for a constitutive gene would yield the correct cumulants in Pol II counts. Knowing the exact location of the fluorescent probe binding regions along the gene, one can calculate the contribution of a single nascent transcript to the signal in C.U. as a function its length l:

$$s(l) = \frac{1}{N}\sum_{i=1}^{N} H(l - l_i) = \frac{1}{N} b(l)$$

where H is the unit step function, l_i the end position of the i^{th} probe binding region and N the total number of probes. Here, we made the assumptions that each fluorescent probe contributes equally to the signal and each transcribed probe region bound. The number of probes bound to a transcript of length l is given by $b(l)$ and will be denoted b_i for $l \in (l_i, l_{i+1}]$ with $l_{N+1} = L_g$ the length of a fully elongated transcript. The total fluorescent signal s in cytoplasmic units for g transcripts is given by

$$s = \frac{1}{N}\sum_{i=1}^{N} b_i g_i$$

where $g = \sum_{i=1}^{N} g_i$, with g_i the number of transcripts whose length l belongs to the length interval $(l_i, l_{i+1}]$. Assuming that g_i follows a Poisson distribution with parameter $\lambda_i = k_{\text{ini}}\tau_i$ where $\tau_i = (l_{i+1} - l_i)/k_{\text{elo}}$, the mean fluorescent signal $\langle s \rangle$ is then given by

$$\langle s \rangle = \frac{1}{N}\sum_{i=1}^{N} b_i \langle g_i \rangle = \frac{1}{N}\sum_{i=1}^{N} b_i k_{\text{ini}}\tau_i = \underbrace{\left(\frac{1}{N}\sum_{i=1}^{N} b_i \tilde{\tau}_i\right)}_{C_1} k_{\text{ini}}\tau_e = C_1 \langle g \rangle$$

where $\tilde{\tau}_i = \tau_i/\tau_e = (l_{i+1} - l_i)/L$ and C_1 the conversion factor that relates the mean number of transcripts $\langle g \rangle$ to the mean fluorescent signal $\langle s \rangle$ in cytoplasmic units. This relation remains valid for the two-state model with $\langle g \rangle = k_{\text{ini}}\tau_e \langle n \rangle$ (Equation 7).

As for the mean, one can calculate the conversion factors for the higher moments and cumulants assuming a Poisson background. The second moment is given by

$$\langle s^2 \rangle = \frac{1}{N^2}\left\langle \sum_{ij} b_i b_j g_i g_j \right\rangle = \frac{1}{N^2}\left(\sum_{i \neq j} b_i b_j \langle g_i \rangle \langle g_j \rangle + \sum_i b_i^2 \langle g_i^2 \rangle\right)$$

$$= \frac{1}{N^2}\left(\sum_{i \neq j} b_i b_j k_{\text{ini}}^2 \tau_i \tau_j + \sum_i b_i^2 \left(k_{\text{ini}}^2 \tau_i^2 + k_{\text{ini}}\tau_i\right)\right)$$

$$= \frac{1}{N^2}\left(\sum_{ij} b_i b_j k_{\text{ini}}^2 \tau_i \tau_j + \sum_i b_i^2 k_{\text{ini}}\tau_i\right)$$

where $\langle g_i g_j \rangle = \langle g_i \rangle \langle g_j \rangle$ since initiation events are assumed independent. This only holds for the Poisson background and is no longer exact for the two-state model as the switching process would introduce correlations. Nevertheless, the conversion factors for the

higher moments and cumulants calculated below remain a good approximation under the two-state model, provided most probes are located in the 5′ region. The variance of the signal is finally given by

$$\left\langle (s-\langle s\rangle)^2 \right\rangle = \langle s^2\rangle - \langle s^2\rangle = \frac{1}{N^2}\sum_{i=1}^{N} b_i^2 k_{\text{ini}}\tau_i = \underbrace{\left(\frac{1}{N^2}\sum_{i=1}^{N} b_i^2\tilde{\tau}_i\right)}_{C_2}\langle g\rangle = C_2\langle g\rangle$$

The calculation above can be generalized to the 3rd and 4th cumulants. We found the following correction factor for the Poisson background:

$$C_k = \frac{1}{N^k}\sum_{i=1}^{N} b_i^k\tilde{\tau}_i \text{ for } k = 1,\cdots,4 \qquad \text{(Equation 3)}$$

Calculated values of C_k for each gene and two different configurations of probe locations (5′ or 3′ region) are given in Table S2.

Two-state model of transcriptional activity

Master equation

Transcriptional activity of a single gene copy was modeled as a telegraph process (on-off promoter switching) with transcript initiation occurring as a Poisson process during the 'on' periods (Peccoud and Ycart, 1995). Within the two-state model (Figure 3E), the distribution of nascent transcripts on a gene results from random Pol II initiation in the active state coupled with elongation and termination (Choubey et al., 2015; Senecal et al., 2014; Xu et al., 2016). For simplicity, we combined elongation and termination as an effective process that was modeled as a deterministic progression (constant Pol II elongation rate). In addition, we assumed that all the kinetic rates of the model are constant in time and identical across embryos. The kinetic parameters of the model are the initiation rate k_{ini}, the promoter switching rates k_{on} and k_{off}, and the elongation time $\tau_e = L_g/k_{\text{elo}}$.

The master equation that governs the temporal evolution of nascent transcripts at loci is given by

$$\frac{d}{dt}P_t(g,n) = k_{\text{ini}}\delta_{n1}(P_t(g-1,n) - P_t(g,n)) + k_n P_t(g,n-1) - k_{n+1}P_t(g,n) \qquad \text{(Equation 4)}$$

with g the number of nascent transcripts (or alternatively the number of Pol II) on the gene and n the promoter state. We used the convention that $n=1$ and $n=0$ correspond to the 'on' state and 'off' state respectively, and the following periodic conditions $n=-1\equiv 1$ and $n=2\equiv 0$. Here, δ stands for the Kronecker delta since initiation only occurs in the active state. Of note, we only considered the promoter switching and the initiation of elongation (Eq. 5); we did not explicitly model release of transcripts after termination. The rationale is the following; only the initiation events occurring during the time interval $[t-\tau_e, t]$ contributes to the signal at time t, i.e., the elongation time τ_e determined the 'memory' of the system. This is correct as long as the release events are instantaneous and termination is fast compared to elongation. Thus, the dynamics of nascent transcripts accumulation on the gene for $t \leq \tau_e$ is obtained by solving the master equation with zero initial transcript on the gene $P_{t_0}(g) = \delta_{g0}$ and an arbitrary initial distribution of promoter state.

Summary statistics

We can derive the temporal evolutions of the central moments from the master equation (Equation 4) (Lestas et al., 2008; Sánchez and Kondev, 2008). The means of nascent transcripts g and promoter states n satisfy the following equations:

$$\begin{cases} \dfrac{d}{dt}\langle g(t)\rangle = k_{\text{ini}}\langle n(t)\rangle \\ \dfrac{d}{dt}\langle n(t)\rangle = k_{\text{on}} - (k_{\text{on}}+k_{\text{off}})\langle n(t)\rangle \end{cases} \qquad \text{(Equation 5)}$$

At steady state $\left(\dfrac{d}{dt}\langle n\rangle = 0\right)$, the mean occupancy of the promoter is simply given by $\langle n\rangle = k_{\text{on}}/(k_{\text{on}}+k_{\text{off}})$. Similarly, the covariance satisfies the following set of equations:

$$\begin{cases} \dfrac{d}{dt}\sigma_g^2(t) = 2k_{\text{ini}}\sigma_{gn}(t) + k_{\text{ini}}\langle n(t)\rangle \\ \dfrac{d}{dt}\sigma_{gn}(t) = k_{\text{ini}}\sigma_n^2(t) - (k_{\text{on}}+k_{\text{off}})\sigma_{gn}(t) \\ \dfrac{d}{dt}\sigma_n^2(t) = -2(k_{\text{on}}+k_{\text{off}})\sigma_n^2(t) + k_{\text{on}}(1-\langle n(t)\rangle) + k_{\text{off}}\langle n(t)\rangle \end{cases} \qquad \text{(Equation 6)}$$

Assuming zero initial transcripts and promoter at steady state, one can solve both the mean and variance for g. Thus, the initial conditions are given by $\langle g(t_0)\rangle = 0$, $\langle n(t_0)\rangle = k_{\text{on}}/(k_{\text{on}}+k_{\text{off}})$, $\sigma_g^2(t_0) = 0$, $\sigma_{gn}(t_0) = 0$ and $\sigma_n^2(t_0) = \langle n(t_0)\rangle(1-\langle n(t_0)\rangle)$. Solving these equations (Equations 5 and 6) for the elongation time $t=\tau_e$ leads to:

$$\langle g\rangle = g_0\langle n\rangle \qquad \text{(Equation 7)}$$

$$\sigma_g^2 = g_0\langle n\rangle + g_0^2\langle n\rangle(1-\langle n\rangle)\Phi(\tau_e/\tau_n) \quad \text{(Equation 8)}$$

where $g_0 = k_{\text{ini}}\tau_e$ is the maximal mean nascent transcript number or equivalently the mean number of transcripts in a constitutive regime (gene always 'on') and $\Phi \in [0,1]$ a noise filtering function that takes into account the fluctuation correlation times. Here, the relevant timescales are the elongation time τ_e and the promoter switching correlation time $\tau_n = 1/(k_{\text{on}} + k_{\text{off}})$. The variance σ_g^2 results from the sum of two contributions; the Poisson variance $g_0\langle n\rangle$ stemming from the stochastic initiation of transcript and the propagation of switching noise:

$$\left(\frac{d\langle g\rangle}{d\langle n\rangle}\right)^2 \sigma_n^2 \Phi(\tau_e/\tau_n) = g_0^2 \underbrace{\langle n\rangle(1-\langle n\rangle)}_{\text{binomial variance}} \Phi(\tau_e/\tau_n)$$

For deterministic elongation, we find that the noise filtering function is given by:

$$\Phi(x) = 2\frac{\exp(-x)+x-1}{x^2}$$

In the limit of fast and slow promoter switching respectively, the noise filtering function reduces to

$$\tau_e \gg \tau_n \lim_{x\to\infty} \Phi(x) = 0$$

$$\tau_e \ll \tau_n \lim_{x\to 0} \Phi(x) = 1$$

Thus, the noise is minimal in the fast switching regime $\tau_e \gg \tau_n$ and reaches the Poisson limit $\sigma_g^2 = g_0\langle n\rangle$. While in the slow switching regime $\tau_e \ll \tau_n$, none of the switching noise is filtered and the variance is described by a second order polynomial of the mean occupancy $\langle n\rangle$, i.e., $\sigma_g^2 = g_0\langle n\rangle + g_0^2\langle n\rangle(1-\langle n\rangle)$. Of note, for exponentially distributed life-time of transcripts, such as cytoplasmic mRNA subject to degradation, the results above remain valid except that the noise averaging function becomes $\Phi(x) = 1/(1+x)$ with τ_e the average life-time of the transcripts.

Following a similar approach as in the previous paragraph, higher order moments and cumulants are analytically calculated from the master equations (Equation 4). The cumulants up to order 3 are equal to the central moments while higher order cumulants can be expressed as a combination of central moments. The 4th cumulant is given by $\kappa_4 = \mu_4 - 3\mu_2^2$, where μ_4 is the 4th central moment and μ_2 the variance. Assuming promoter at steady state, we solved the equations for 3rd and 4th moments of g and derive the following analytical expressions for 3rd and 4th cumulants, κ_3 and κ_4:

$$\kappa_3 = g_0\langle n\rangle + 3g_0^2\langle n\rangle(1-\langle n\rangle)\Phi_1(\tau_e/\tau_n) + g_0^3\langle n\rangle(1-\langle n\rangle)(1-2\langle n\rangle)\Phi_2(\tau_e/\tau_n) \quad \text{(Equation 9)}$$

$$\kappa_4 = g_0\langle n\rangle + 7g_0^2\langle n\rangle(1-\langle n\rangle)\Phi_1(\tau_e/\tau_n) + 6g_0^3\langle n\rangle(1-\langle n\rangle)(1-2\langle n\rangle)\Phi_2(\tau_e/\tau_n) + g_0^4\langle n\rangle(1-\langle n\rangle)(\Phi_3(\tau_e/\tau_n) - 6\langle n\rangle(1-\langle n\rangle)\Phi_4(\tau_e/\tau_n)) \quad \text{(Equation 10)}$$

where Φ_1, Φ_2, Φ_3 and Φ_4 are noise filtering functions that vanish in the fast switching regime ($\tau_e \gg \tau_n$) and tend to one in the slow switching regime ($\tau_e \ll \tau_n$):

$$\Phi_1(x) = 2\frac{\exp(-x)+x-1}{x^2}$$

$$\Phi_2(x) = 6\frac{x\exp(-x)+2\exp(-x)+x-2}{x^3}$$

$$\Phi_3(x) = 12\frac{x^2\exp(-x)+4x\exp(-x)+6\exp(-x)+2x-6}{x^4}$$

$$\Phi_4(x) = 2\frac{\exp(-x)^2+4x^2\exp(-x)+20x\exp(-x)+28\exp(-x)+10x-29}{x^4}$$

The above expressions for the cumulants are exact and were tested numerically. The cumulants are polynomials of the mean promoter activity $\langle n \rangle$, which follows from the propagation of the binomial cumulants from the switching process. Since the cumulants are extensive, the cumulants for N_g independent gene copies are obtained by multiplying by N_g the expression for a single gene copy (Equations 8, 9, and 10).

Cumulant analysis

Noise-mean relationship and cumulants predicted by the two-state model

Within the context of the two-state model, we tested whether any transcriptional parameter modulations could explain the global trends in the noise and the cumulants (Figures 3A–3D). Since we showed based on the cumulants that the distribution of activity is a single parameter distribution, we restricted the analysis to single parameter modulations of the mean activity (Figures 3F–3I). It is worth mentioning a few important observations that will simplify this task.

First, we see by close inspection of the steady state cumulants (Equations 8, 9, and 10) that τ_e sets the scale, i.e., all parameters are defined with respect to τ_e. In practice, the cumulants only depends on the three following independent parameters $\tilde{k}_{\text{ini}} = k_{\text{ini}}\tau_e$, $\tilde{k}_{\text{on}} = k_{\text{on}}\tau_e$ and $\tilde{k}_{\text{off}} = k_{\text{off}}\tau_e$. Thus, there is some freedom to set the scale of these rates. Here, we used $\tau_e = 2.2$ min that is approximately the Pol II elongation time for the normalized gene length (3.3 kb and $k_{\text{elo}} = 1.5$ kb/min; (Garcia et al., 2013)) and it will be considered fixed. Second, the magnitude of k_{ini} determines whether the Poisson (first term $\propto k_{\text{ini}}$) or the binomial (second term $\propto k_{\text{ini}}^2$) components dominates in the expression of the variance (Equation 6). We immediately see that increasing the mean Pol II number on the gene $\langle g \rangle$ by only modulating k_{ini} cannot explain the data, since it would lead to a monotonic increase of the variance whereas the observed trend is concave with a global maxima at mid-expression levels. The only way of achieving such a trend is by modulating $\langle n \rangle$ provided the binomial term dominates the Poisson one. This condition implies that k_{ini} has to be sufficiently large for intermediate value of $\langle n \rangle$, i.e., $k_{\text{ini}} \gg 1/(\tau_e(1 - \langle n \rangle)\Phi(\tau_e/\tau_n))$. Alternatively, if k_{ini} is known, this inequality sets some constraints on the possible values of τ_n. Third, it is possible to give an estimate of k_{ini} from the polynomial fit of the measured variance (Figure S2A and Figure 3B). The second intercept of the fitted curve (black line) with the Poisson background (dash line), which should occur at $\langle n \rangle = 1$, allows us to estimate g_0. Assuming k_{ini} is maintained constant as $\langle n \rangle$ is modulated, we have $g_0 = k_{\text{ini}}\tau_e = 15.21$, which gives $k_{\text{ini}} = 6.99$ min^{-1} for $\tau_e = 2.2$ min (see above).

We then investigated three different type of single parameter modulation to vary the mean Pol II number $\langle g \rangle$ consistent with the observation above, namely, modulations of the mean occupancy $\langle n \rangle$ from 0 to 1 by either varying k_{on} alone, k_{off} alone or both k_{on} and k_{off} while keeping the switching correlation time τ_n constant. The latter modulation also corresponds to single parameter modulation since $k_{\text{on}} = \langle n \rangle/\tau_n$ and $k_{\text{off}} = (1 - \langle n \rangle)/\tau_n$ are then fully determined by $\langle n \rangle$. For each of these three types of modulation, one parameter is free (either k_{off}, k_{on} or τ_n) and sets the amplitude of the cumulants (Figure S2E). In order to infer these free parameters, we fitted (maximum likelihood) the measured cumulants with the modeled ones (Equations 8, 9, and 10) predicted by each modulation strategy (Figures 3G–3I). We found:

1) k_{on} modulation: $k_{\text{off}} = 0.142$ min^{-1} and $k_{\text{on}} = k_{\text{off}}\langle n \rangle/(1 - \langle n \rangle)$
2) k_{off} modulation: $k_{\text{on}} = 0.075$ min^{-1} and $k_{\text{off}} = k_{\text{on}}(1 - \langle n \rangle)/\langle n \rangle$
3) $\langle n \rangle$ modulation at fixed τ_n: $\tau_n = 2.9993$ min with $k_{\text{on}} = \langle n \rangle/\tau_n$ and $k_{\text{off}} = (1 - \langle n \rangle)/\tau_n$

We then calculated the noise-mean relationship (Equation 2). We also show an example of a k_{ini} modulation alone (Figure 3F, gray line); no matter the value of $\langle n \rangle$ and τ_n this modulation cannot reproduce the trend in the data as explained above. The modulation of k_{off} alone (green line) fails to capture the noise at low expression (Figure 3F). On the other hand both the modulation of k_{on} alone (blue line) and n at constant τ_n (red line) provides good qualitative agreement with the data (Figures 3F–3I). As mentioned above, it is important to keep in mind that the units of k_{ini}, k_{on}, k_{off} and τ_n estimated here depends on the value of the elongation rate. Here, we used a conservative estimate of $k_{\text{elo}} = 1.5$ kb/min (Garcia et al., 2013), which is possibly too small for the gap genes (Fukaya et al., 2017). A different elongation rate would simply imply a rescaling of the rates and the correlation time without affecting the fitting results (STAR methods, *Effect of elongation rate on inference*). Namely, the mean occupancy $\langle n \rangle$ would remain unchanged while the rates would be rescaled by a factor $k^*_{\text{elo}}/k_{\text{elo}}$ and the correlation time by $k_{\text{elo}}/k^*_{\text{elo}}$, where k^*_{elo} corresponds to the new elongation rate.

Time-dependent cumulant analysis

Next, we investigated whether the single parameter modulation fitted above assuming steady state are consistent with the finite duration of the nuclear cycle (approximately 15 min in nc13). Namely, assuming all the data were taken at mid cycle, we asked under each modulation scenario whether steady state could be reached in a timely manner (mid cycle), as supported by our staging analysis (Figures S1A–S1D) and other studies (Garcia et al., 2013). The relaxation time to steady state is determined by the switching correlation time τ_n. By solving the equation for the temporal evolution of the mean Pol II number $\langle g(t) \rangle$ (Equation 5) with initial condition $\langle g(t = 0) \rangle = 0$ (no Pol II on the gene) and $\langle n(t = 0) \rangle = 0$ (gene initially 'off'), one finds:

$$\langle g(t) \rangle = \begin{cases} g_0\langle n \rangle \left(\frac{t}{\tau_e} + \frac{\tau_n}{\tau_e}(\exp(-t/\tau_n) - 1) \right) t \leq \tau_e \\ g_0\langle n \rangle \left(1 + \frac{\tau_n}{\tau_e}\exp(-t/\tau_n)(1 - \exp(\tau_e/\tau_n)) \right) t > \tau_e \end{cases}$$

As mentioned above, the relaxation of the mean $\langle g(t)\rangle$ to its steady state value $\langle g\rangle = g_0\langle n\rangle$ is determined by the correlation time τ_n through the exponential factor $\exp(-t/\tau_n)$. As τ_n increases, the relaxation gets slower and slower (Figure S3A). It follows that the finite duration of nc13 should set some upper bound on the possible value of τ_n. According to Figure S3A, τ_n should not exceed 3 min for $\langle g(t)\rangle$ to reach approximately 90% of the maximum activity g_0 ($\langle g\rangle = g_0$ for $\langle n\rangle = 1$) at mid cycle as observed in the data (Figure 3B).

Each of the three single parameter modulations fitted above predicts different dependency of τ_n on the mean occupancy $\langle n\rangle$ (Figure S3B). Importantly, these values of τ_n were obtained for $k_{elo} = 1.5$ kb/min (Garcia et al., 2013). A larger elongation rate would lead to smaller correlation times (Fukaya et al., 2017) (STAR methods, *Effect of elongation rate on inference*). The main benefit of using a potentially smaller elongation rate, it provides a stronger guarantee that the time-dependent solution reaches steady state in time (as the relaxation is slower). For each modulation (Figure S3C), we estimated what fraction of the steady state value $\langle g(t)\rangle/\langle g\rangle$ is attained as a function of $\langle n\rangle$ at mid cycle ($t = 7.5$ min). It turns out that the k_{off} modulation clearly fails to reach steady state in time for higher occupancy, whereas both modulation of k_{on} and n at fixed τ_n cover the measured range of activity at mid cycle (0 to 90% of g_0). Each modulation predicts different boundary formation dynamics (Figures S3D–S3F). For k_{on}, the highly expressed regions (large $\langle n\rangle$) relax much faster than the lowly expressed ones (small $\langle n\rangle$), whereas for k_{off} it is the opposite. Interestingly, at fixed τ_n, each position relaxes in synchrony and the activity ratio between them is conserved. The latter modulation appears more consistent with previous experimental observations (Dubuis et al., 2013; Garcia et al., 2013).

Next, we investigated the shape of the higher order time-dependent cumulants. Although the higher order time-dependent cumulants can be calculated from the moment equations, their analytical expressions are cumbersome. Alternatively, one can calculate the time-dependent cumulants directly from the time-dependent distribution of Pol II $P_t(g)$, which is easily computed numerically. With the same initial condition as the mean above, the time-dependent distribution of Pol II $P_t(g)$ is given by:

$$P_t(g) = \begin{cases} \sum_n P_t(g, n \mid g' = 0, n' = 0) & t \le \tau_e \\ \sum_{n,n'} P_{\tau_e}(g, n \mid g' = 0, n') P_{t-\tau_e}(n' \mid n'' = 0) & t > \tau_e \end{cases}$$

where $P_t(g, n \mid g', n')$ is the propagator of the telegraph model (STAR Methods, *Distribution of nascent transcripts*, Equation 12) and $P_t(n \mid n')$ the propagator of the switching process alone:

$$P_t(n, n') = (\delta_{n1}\langle n\rangle + \delta_{n0}(1 - \langle n\rangle))(1 - \exp(-t/\tau_n)) + \delta_{nn'}\exp(-t/\tau_n)$$

We then computed the 2nd, 3rd and 4th time-dependent cumulants from $P_t(g)$ for each fitted modulation (Figure S3G). Provided the elapsed time is sufficiently large compared to the correlation time and the elongation time, the time-dependent cumulants closely follow the steady state solution. Thus, both the modulation of k_{on} alone and n at fixed τ_n fitted assuming steady state predicts time-dependent mean versus cumulant curves at mid cycle ($t = 7.5$ min) that are consistent with the data. In addition, under these conditions, the time-dependent mean activity closely reflect the time-dependent mean occupancy $\langle n(t)\rangle$:

$$\frac{\langle g(t)\rangle}{g_0} = \underbrace{\frac{1 + \frac{\tau_n}{\tau_e}\exp(-t/\tau_n)(1 - \exp(\tau_e/\tau_n))}{1 - \exp(-t/\tau_n)}}_{\approx 1} \langle n(t)\rangle \; t > \tau_e$$

Together it implies that even away from steady state, provided the elapsed time is sufficiently large ($t \gg \tau_e, \tau_n$), the inference based on steady state solutions should yield good estimates of the parameters. Indeed, for fixed τ_e, the relationships between the mean and the cumulants at steady state are uniquely determined by k_{ini}, $\langle n\rangle$ and τ_n. As long as time dependent-cumulants run along the steady state curves (Figure S3G), the estimation of k_{ini} and τ_n will be correct while the estimation of the mean occupancy will in fact corresponds to the instantaneous mean occupancy $\langle n(t)\rangle$ as $\langle g(t)\rangle/g_0 \approx \langle n(t)\rangle$.

Inferring transcription kinetics of endogenous genes from dual color smFISH

Dual color smFISH and effective gene length

We performed dual-color smFISH tagging the 5′ and 3′ regions of the transcripts with different probe sets (Figure 4A and Table S1). After normalization in cytoplasmic units, both channels offer a consistent readout of the mean and the variability (Figures S4A and S4B). For each gene, given the 5′ and 3′ FISH probe configurations and assuming constant elongation rate, we calculated the expected ratio of 3′ over 5′ signal $r = C_1^{(3)}/C_1^{(5)}$ according to Equation 3 using the annotated gene length (Figures S4C and S4D and Table S3). The predicted ratios are consistent with the measured ones, albeit with small deviations likely stemming from termination (Figure S4E). This suggests that nascent transcripts might be retained at transcription sites for a short duration. We then calculated for each gene, the effective length that would be consistent with the measured ratio (Figure S4F and Table S3). Assuming an elongation rate $k_{elo} = 1.5$ kb/min (Garcia et al., 2013), we estimated the lag consistent with the length difference between the effective and an-

notated length (Figure S4F inset). Nascent transcripts remain at the loci for at most 35 s, which remains small compared to the typical elongation time for the gap genes $\tau_e \sim 2$ min. In this study, we used the effective elongation time for each gene that includes the short lingering time, which was calculated from the effective gene length.

The two channels enable estimation of the total nascent transcripts (5′ channel) and the fractional occupancy of transcripts along the 5′ and 3′ portions of the gene at each locus (Figures 4B and 4C). Because the 5′ and 3′ activities are temporally correlated through the elongation process additional information about transcription can be extracted that is not available with a single channel/color (Figures 4B and 4D). Combining measurements from multiple embryos (Figures 4C and 4D), we select nuclei at similar positions (bins of 2.5% egg length) to generate the joint distribution of 5′ and 3′ activity across AP position bins (Figure 4D).

Distribution of nascent transcripts

Modeling the joint distribution of 5′ and 3′ activity based on the two-state model requires first to calculate two key distributions, namely the steady-state distribution of nascent transcripts (or Pol II number) on the gene and the propagator that describes the temporal evolution of an arbitrary distribution of nascent transcripts. Both distributions can be derived from the master equation (Equation 4). Although the master equation can be solved using generating functions (Xu et al., 2016), we followed another route that can be easily extended to multi-state system and remains computationally tractable. The master equation can be written in terms of an operator $\widehat{A}$ containing the propensity functions of the different reactions:

$$\frac{d}{dt}P_t(g,n) = \widehat{A}P_t(g,n)$$

After appropriate truncation on the transcript number (setting an upper bound for the maximum number of nascent transcripts) (Munsky and Khammash, 2006), the $\widehat{A}$ operator can be written in terms of a sum of tensor products of different matrices:

$$\widehat{A} = I_G \otimes N_2 + K_G \otimes R_2 \qquad \text{(Equation 11)}$$

with I_G standing for the identity matrix of size $G+1$ where G is the maximum number of transcripts after truncation. The matrix N_2 encodes the rates of the possible transitions for the two-state promoter and R_2 indicates in which promoter state initiation occurs:

$$N_2 = \begin{bmatrix} -k_{\text{on}} & k_{\text{off}} \\ k_{\text{on}} & -k_{\text{off}} \end{bmatrix} R_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

while K_G describes the initiation of transcripts:

$$K_G = \begin{bmatrix} -k_{\text{ini}} & 0 & 0 & 0 \\ k_{\text{ini}} & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & k_{\text{ini}} & -k_{\text{ini}} \end{bmatrix}$$

The propagator of the resulting finite system can be expressed as a matrix exponential of the $\widehat{A}$ operator:

$$P_t(g,n|g',n',\theta) = \exp\left(\widehat{A}t\right) \qquad \text{(Equation 12)}$$

where θ stands for the set of kinetic parameters $(k_{\text{ini}}, k_{\text{on}}, k_{\text{off}})$. Although the propagator explicitly depends on the kinetic parameters, we chose to omit θ in the following for readability. The propagator dictates how an initial joint distribution of transcript and promoter state $P(g', n')$ evolves after time t in $P(g,n)$:

$$P(g,n) = \sum_{g',n'} P_t(g,n|g',n')P(g',n')$$

The distribution of nascent transcripts $P(g)$ for a gene of length L_g is typically calculated using the propagator above with $t = \tau_e \equiv L_g/k_{\text{elo}}$ the elongation time and the initial conditions. Since τ_e sets the 'memory' of the system, $P(g)$ can be calculated with initially zero nascent transcript on the gene and is then given by:

$$P(g) = \sum_{n,g',n'} P_{\tau_e}(g,n|g',n')\delta_{g'0}P(n') \qquad \text{(Equation 13)}$$

where $P(n)$ specifies the initial distribution of promoter state. The distribution $P(g)$ can be computed efficiently by directly estimating the action of the initial vector on the matrix exponential (Sidje, 1998). Assuming the promoter at steady state, $P(n)$ is then given by:

$$P(n) = \begin{cases} \langle n \rangle \text{ for } n = 1 \\ 1 - \langle n \rangle \text{ for } n = 0 \end{cases}$$

with the mean occupancy $\langle n \rangle = k_{\text{on}}/(k_{\text{on}} + k_{\text{off}})$.

Provided each gene copy is independent and undistinguishable, the combination of two and four gene copies can be represented by a three- and five-state promoter model. The corresponding N and R matrices are given by:

$$N_3 = \begin{bmatrix} -2k_{on} & k_{off} & 0 \\ 2k_{on} & -(k_{off}+k_{on}) & 2k_{off} \\ 0 & k_{on} & -2k_{off} \end{bmatrix} R_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$$N_5 = \begin{bmatrix} -4k_{on} & k_{off} & 0 & 0 & 0 \\ 4k_{on} & -(k_{off}+3k_{on}) & 2k_{off} & 0 & 0 \\ 0 & 3k_{on} & -2(k_{off}+k_{on}) & 3k_{off} & 0 \\ 0 & 0 & 2k_{on} & -(3k_{off}+k_{on}) & 4k_{off} \\ 0 & 0 & 0 & k_{on} & -4k_{off} \end{bmatrix}$$

$$R_5 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

The distribution of nascent transcripts is calculated according to Equation 13, with the propagator $P_t(g,n|g',n')$ computed from the updated $\widehat{A}$ operator (Equations 11 and 12). The steady-state distribution of the N_g-gene copy system is given by:

$$P(n) = \binom{N_g}{n} \langle n \rangle^n (1 - \langle n \rangle)^{N_g - n} \text{ with } n \in \{0, 1, 2, \cdots, N_g\} \quad \text{(Equation 14)}$$

where $n = k_{on}/(k_{on}+k_{off})$ is the steady state mean occupancy of a single promoter.

Joint distribution of 5′ and 3′ activity

Here, we lay out the approach used to calculate the joint distribution of 5′ and 3′ activity for an arbitrary configuration of 5′ and 3′ FISH probes. Analytic solutions for steady-state distributions with idealistic single color probe configuration exist (Xu et al., 2016), but solutions for arbitrary probe configurations and multi-color FISH are cumbersome. Here, the computational approach is general enough and can be applied to a large class of transcription model, at or out of steady-state (transient relaxation), provided the elongation process is assumed deterministic.

The measured 5′ and 3′ transcriptional activities result from partially elongated nascent transcripts. Each fluorescent probe is assumed to be instantaneously bound and to contribute equally to the total fluorescence. Thus, the fluorescent signal of each nascent transcript is proportional to the number of probe binding regions that have been transcribed. In order to calculate the joint distribution, one needs to proceed backward in time. Starting from the 3′ end up to the 5′ end of the gene, we accumulate the contribution of nascent transcripts to the signal that could have been initiated in the interval separating two successive probe regions. Since we assumed elongation to occur at constant speed, the distance between two successive probe regions can be converted into a time. Doing so for each interval leads to the following temporal hierarchy (Figure S4G). We used the following naming conventions for the durations $t_i^{(C)}$: the superscript $(C) \in \{(3),(5)\}$ stands for the probe channel, either (3) for the 3′ probes (red channel) or (5) for the 5′ probes (green channel), whereas the subscript i denotes the interval separating probe i from probe $i-1$ where increments are performed along the 3′ end to 5′ end direction.

For instance, if the 5′ and 3′ signal is measured at time $t = \tau_e$, only transcripts initiated during the time interval $[0, t_1^{(3)}]$ fully contribute (1 C.U.) to the 3′ (red) signal, since only those get fully bound by 3′ FISH probes. On the other hand, transcripts initiated during $[t_1^{(3)}, t_1^{(3)} + t_2^{(3)}]$ will contribute less to the signal since the last probe region has not yet been transcribed at the time of the measurement $t = \tau_e$. Thus, the individual contribution of these transcripts to the total 3′ signal is $(k-1)/k$ C.U., where k is the total number of probes for the 3′ channel. As we will see below, the probability to initiate g nascent transcripts during any duration $t_i^{(C)}$ is given by the propagator $P_{t_i^{(C)}}(g,n|0,n')$ (Equation 12), where n and n' are the promoter states before and after $t_i^{(C)}$.

For any model of promoter activity that only consider the stochastic initiation of transcripts (as a Poisson process) and deterministic elongation with instantaneous release, the propagator will satisfy the following equality:

$$P_t(g,n|g',n') = P_t(g-g',n|0,n')$$

Thus, one only needs to calculate $P_t(g,n|0,n') \equiv P_t(g,n|n')$, which can be computed much faster than the matrix exponential (Equation 12) (Sidje, 1998). It then follows that the Chapman-Kolmogorov equation for the time propagation reduces to a discrete convolution:

$$P_{t_2+t_1}(g_2,n_2|n_0) = \sum_{n_1} \sum_{g_1=0}^{g_2} P_{t_2}(g_2-g_1,n_2|n_1) P_{t_1}(g_1,n_1|n_0)$$

This property is used extensively in the following calculation of the joint distribution.

The computation of the joint distribution is performed according to a dynamic programming approach that can in principle be applied to an arbitrary number of color probes. We first calculate recursively the 3′ contribution (red probes) to the signal $P^{(3)}(\tilde{G}_k, G_k, n_k)$, where $\tilde{G}_k$ stands for the total signal in probe space, G_k the total number of nascent transcripts, n_k the promoter state and k the total number of probes covering the 3′ region. We then calculate the 5′ contribution in a similar fashion, $P^{(5)}(\tilde{G}_k, n_0)$. Lastly, we combine both components to generate the final joint distribution $P(\tilde{G}^{(5)}, \tilde{G}^{(3)})$ in probe space.

Step 1: calculate the 3′ contribution. The initial distribution is given by:

$$P^{(3)}\left(\tilde{G}_2, G_2, n_2\right) = P^{(3)}\left(\overbrace{(k-1)g_2 + kg_1}^{\&doublehyphen;38pt\tilde{G}_2}, \overbrace{g_2+g_1}^{G_2}, n_2\right) = \sum_{n_0,n_1} P_{t_2^{(3)}}(g_2, n_2|n_1) P_{t_1^{(3)}}(g_1, n_1|n_0) P(n_1) P(n_0)$$

where $P(n_0)$ and $P(n_1)$ are the initial distributions of promoter state at time $t_0 = 0$ and $t_1^{(3)}$ respectively. Assuming promoters at steady state, both distributions are then given by Equation 14 for a multi-gene system. We then perform the following recursion scheme for $i = \{3, \cdots, k\}$:

$$P^{(3)}\left(\tilde{G}_i, G_i, n_i\right) = \sum_{n_{i-1}} \sum_{g_i=0}^{g_{\max}} P_{t_i^{(3)}}(g_i, n_i|n_{i-1}) P^{(3)}\left(\underbrace{\tilde{G}_i - (k-i+1)g_i}_{\tilde{G}_{i-1}}, \underbrace{G_i - g_i}_{G_i}, n_{i-1}\right)$$

where $g_{\max} = \min\left(\lfloor \tilde{G}_i/(k-i+1) \rfloor, G_i\right)$.

Step 2: calculate the 5′ contribution. The initial distribution is given by:

$$P^{(5)}\left(\tilde{G}_1, n_1|n_0\right) \equiv P^{(5)}(kg_1, n_1|n_0) = P_{t_1^{(5)}}(g_1, n_1 \mid n_0)$$

We then perform the following recursion scheme for $i = \{2, \cdots, k\}$:

$$P^{(5)}\left(\tilde{G}_i, n_i|n_0\right) = \sum_{n_{i-1}} \sum_{g_i=0}^{g_{\max}} P_{t_i^{(5)}}(g_i, n_i \mid n_{i-1}) P^{(5)}\left(\underbrace{\tilde{G}_i - (k-i+1)g_i}_{\tilde{G}_{i-1}}, n_{i-1}|n_0\right)$$

where $g_{\max} = \lfloor \tilde{G}_i/(k-i+1) \rfloor$. Lastly, we sum out n_k:

$$P^{(5)}\left(\tilde{G}_k|n_0\right) = \sum_{n_k} P^{(5)}\left(\tilde{G}_k, n_k|n_0\right)$$

Step 3: combine 3′ and 5′ contributions. The final joint distribution of 5′ and 3′ activity in probe space is then given by:

$$P\left(\tilde{G}^{(5)}, \tilde{G}^{(3)}\right) = \sum_{n} \sum_{G=0}^{G_{\max}} P^{(5)}\left(\tilde{G}^{(5)} - kG|n\right) P^{(3)}\left(\tilde{G}^{(3)}, G, n\right)$$

where $G_{\max} = \left\lfloor \tilde{G}^{(5)}/k \right\rfloor$. $P^{(3)}$ and $P^{(5)}$ are the joint distributions computed at step 1 and 2. Since the actual signal resolution is of the order of 1 cytoplasmic unit (a fully tagged transcript with k fluorescent probes), the joint distribution can be coarse-grained by aggregating the states $\tilde{G}$ by a block of size k corresponding to a single cytoplasmic unit. The coarse-grained distribution will be denoted $P(G^{(5)}, G^{(3)})$ in the following. In addition, it is possible to compute $P(G^{(5)}, G^{(3)})$ faster and with good accuracy using a reduced effective number of probes k, provided the original probe configuration is well approximated. Lastly, we remind the readers that $P(G^{(5)}, G^{(3)})$implicitly depends on the kinetic parameters $(k_{\text{ini}}, k_{\text{on}}, k_{\text{off}})$ through the two-state model propagator, the elongation rate and the position of the probes through the temporal hierarchy (Figure S4G).

Likelihood and inference

We modeled the joint distribution of 5′ and 3′ activity based on the two-state model and the exact probe location assuming steady state and constant Pol II elongation rate (Figure 4E; STAR Methods, *Joint distribution of 5′ and 3′ activity*). The resulting modeled activity distribution, together with the measurement noise model (Figure 2A; STAR Methods, *Imaging noise model*), enable calculating the likelihood of the 5′ and 3′ activities in C.U. (i.e., Data) given a set of kinetic parameters $(k_{\text{ini}}, k_{\text{on}}, k_{\text{off}})$. Specifically, the likelihood of the data Data $= \{S^{(5)}, S^{(3)}\}$ given the parameters $\theta = (k_{\text{ini}}, k_{\text{on}}, k_{\text{off}})$ is expressed in terms of the measurement noise model $P(S^{(5)}, S^{(3)}|G^{(5)}, G^{(3)})$ (Equation 1) and the joint distribution $P(G^{(5)}, G^{(3)} \mid \theta)$:

$$P(\text{Data}|\theta) = \prod_{i=1}^{N_D} \sum_{G^{(5)}, G^{(3)}} P\left(S_i^{(5)}, S_i^{(3)}|G^{(5)}, G^{(3)}\right) P\left(G^{(5)}, G^{(3)} \mid \theta\right)$$

where N_D is the total amount of data, i.e., the total number of measured nuclei per AP-bin for a given gene.

The general idea underlying "classical" inference is to maximize the probability of the data under some model, namely to find the parameters $(k_{ini}, k_{on}, k_{off})$ that maximize the likelihood of the data $P(\text{Data}|k_{ini}, k_{on}, k_{off})$. In this manuscript we adopted a Bayesian approach, estimating the probability of the kinetic rate parameters of the two-state model given the observed data (i.e., the joint posterior distribution) $P(k_{ini}, k_{on}, k_{off}|\text{Data})$ using Bayes' rule:

$$P(k_{ini}, k_{on}, k_{off}|\text{Data}) = \frac{P(\text{Data}|k_{ini}, k_{on}, k_{off})P(k_{ini}, k_{on}, k_{off})}{P(\text{Data}) \equiv \int P(\text{Data}|k_{ini}, k_{on}, k_{off})P(k_{ini}, k_{on}, k_{off})dk_{ini}dk_{on}dk_{off}}$$

where $P(k_{ini}, k_{on}, k_{off})$ is the prior that encodes for prior knowledge about the parameter values. We used a non-informative and independent prior for each kinetic parameter, which was chosen as log-uniform $P(k_{ini}, k_{on}, k_{off}) = 1/(k_{ini} \cdot k_{on} \cdot k_{off})$. Note that in absence of a prior $P(k_{ini}, k_{on}, k_{off})$, the most likely parameters are the ones that maximize $P(\text{Data}|k_{ini}, k_{on}, k_{off})$. In that case, the Bayesian approach is essentially equivalent to "classical" maximum likelihood. The main advantage of the Bayesian approach over maximum likelihood is that it provides a natural way to estimate the uncertainty on the parameters through the joint posterior and allows us to determine whether the parameters are identifiable. Indeed, as the uncertainty grows, the posterior distribution becomes wider/flatter, which directly reflects on the range of the parameter confidence intervals.

Importantly, we set the elongation rate k_{elo} to the experimentally measured value of 1.5 kb/min (Garcia et al., 2013). At steady state, a known value of k_{elo} is required to set the temporal scale of the other transcriptional parameters, which can be seen by inspecting the expressions of the various cumulants of the nascent transcript distribution (Equations 8, 9, and 10). Since all cumulants can be parameterized by the three independent parameters $g_0 = k_{ini}/k_{elo}$, $n = k_{on}/(k_{on} + k_{off})$ and the ratio $\tau_e/\tau_n = (k_{on} + k_{off})/k_{elo}$, it follows that the model is not identifiable when the temporal scale is not set.

We then sampled the joint posterior distribution $P(k_{ini}, k_{on}, k_{off}|\text{Data})$ using a Markov chain Monte Carlo (MCMC) algorithm (Hastings, 1970), for each gene and at each AP position individually. The sampled joint posterior distribution enables estimation of the marginal posterior distribution for each kinetic rate and any combination of these rates, such as $\langle n \rangle$ and τ_n. All the parameters of the model and the error bars were estimated from the marginal posterior distribution, as the median and the percentiles respectively (Figure 4E). The best-fitting distributions predicted by the model match the data closely (Figure S5B), and outliers are mainly explained by measurement and binning noise. Importantly, our inference approach does not require any *a priori* assumptions about the underlying parameter modulation, nor does it assume any continuity between datasets. In principle, the inferred parameters could be different for each gene and be modulated in any arbitrary way.

Parameter identifiability and performance

As mentioned above, the two-state model is fully identifiable (structural identifiability) as long as k_{elo} is fixed. Indeed, in that case the steady state and time-dependent solution depend on three independent parameters, such as $(k_{ini}, k_{on}, k_{off})$ or $(k_{ini}, \langle n \rangle, \tau_n)$. In principle, provided one has enough data and measurement noise is small, each parameter can be resolved individually. On the other hand, it is true that some regimes might require a very large/infinite amount of data to infer the different parameters without ambiguity (practical identifiability). For instance, in the case of instantaneous bursts, namely when k_{off} and k_{ini} become large (i.e., approach infinity, but with finite ratio), only the burst size $b = k_{ini}/k_{off}$ and the burst frequency $f = k_{on}$ are well defined. Thus it is not possible to infer the exact values of k_{ini} and k_{off} individually. Such a scenario can be clearly diagnosed based on the marginal posterior distributions $P(k_{ini}|\text{Data})$ and $P(k_{off}|\text{Data})$ (from which the median and the error bars of the parameters are estimated). Indeed, since we used non-informative priors, the variance of these marginal posterior distributions would become extremely large and thus less informative. More intuitively, k_{ini} and k_{off} would no longer be sharply peaked around a mean value, but would take all possible values (consistent with the prior) that satisfy $b = k_{ini}/k_{off} \pm$ some error on b. This would consequently lead to to very large error bars on k_{ini} and k_{off}. Thus, the error bars extracted from the marginal posterior distribution are indicative for whether or not we can estimate these parameters.

To validate our inference framework, we tested the inference on simulated data using a broad range of parameter values and in presence of measurement noise. Using the Gillespie algorithm (Gillespie, 1977), we generated simulated nuclei activity data based on 4 independent gene copies modeled by the telegraph model. We used the probe configuration and gene length of *hb* and assumed a typical elongation rate of 1.5 kb/min (Garcia et al., 2013). Measurement noise was included in the simulated data according to the characterization performed previously on real data (*Imaging noise model*). We investigated different parameter regimes and modulation schemes of the mean activity $\langle g \rangle$, to test whether the input parameters used to generate the data could be inferred properly (Figures S6A–S6E). Namely, we tested:

1) Modulation of the initiation rate k_{ini} alone with $\tau_n = 2$ min and $\langle n \rangle = 0.35$ (cyan dash line).
2) Modulation of the on-rate k_{on} alone with $k_{ini} = 7$ min^{-1} and $k_{off} = 0.25$ min^{-1} (green dash line).
3) Modulation of the off-rate k_{off} alone with $k_{ini} = 7$ min^{-1} and $k_{on} = 0.25$ min^{-1} (blue dash line).
4) Modulation of the mean occupancy $\langle n \rangle$ alone with $k_{ini} = 7$ min^{-1} and $\tau_n = 2$ min (red dash line).

For each scenario, we generated 8 batches of data covering the range of normalized activity $\langle g \rangle / g_0$. Each batch was made of 10 independently sampled datasets of 500 nuclei activity measurements. We performed the inference on each dataset individually and reported the mixture of posterior distribution over the 10 datasets to take into account the finite size variability in the generated data. We conclude that the inference framework performs well, since all the inferred quantities cover the true values within error bars.

In addition, we estimated globally for all synthetic data the fractional inference error $|\theta_{inf} - \theta_{true}|/\theta_{true}$ from the MCMC sampled parameters θ_{inf}. For all inferred parameters, the median of the error never exceeds 20% (S6F). Overall, the inference allows us to distinguish the different tested modulation strategies without ambiguities. In addition, the sampled joint posterior distributions $P(k_{ini}, k_{on}, k_{off}|\text{Data})$ are clearly peaked in the parameter space (Figure S5C), indicating that practical identifiability is not an issue with real data.

Effect of elongation rate on inference

As discussed above, the elongation rate k_{elo} sets the temporal scale of the transcriptional parameters, thus a different elongation rate would lead to different values of the parameters. In the manuscript, we used a value of $k_{elo} = 1.5$ kb/min which we previously measured (Garcia et al., 2013). A recent study suggests that this value might be overall larger in the blastoderm embryo, of the order of 2.5 kb/min (Fukaya et al., 2017). We thus sought to determine to which extent this new value would affect our results.

In principle, a different value of k_{elo} rescales the transcriptional parameters in a very predictable way. No matter the elongation rate, the three quantities k_{ini}, $\langle n \rangle$ and τ_n should be perfectly identifiable. It follows that the new parameters (denoted by the * superscript) have to satisfy the following equations:

$$k_{ini}\frac{k^*_{elo}}{k_{elo}} = k^*_{ini}$$

$$\langle n \rangle = \langle n \rangle *$$

$$\tau_n \frac{k_{elo}}{k^*_{elo}} = \tau^*_n$$

Inferring the transcriptional parameters from the data with $k_{elo} = 2.5$ kb/min instead of $k_{elo} = 1.5$ kb/min (as in the main text) confirms the rescaling above (Figures S6L–S6N). As predicted, k_{ini} and k_{ini} are rescaled by a factor $2.5/1.5 = 1.67$ and $1.5/2.5 = 0.6$ respectively, whereas $\langle n \rangle$ is conserved.

QUANTIFICATION AND STATISTICAL ANALYSIS

We imaged *hunchback* wild-type (labeled *hb* wt) in N = 18 embryos; a *hunchback* deficiency fly line with half the *hb* dosage (*hb* def) N = 7; *Krüppel* (*Kr*) N = 11; *knirps* during early (*kni* early) N = 14 and late nc13 (*kni* late) N = 16; *giant* females with two alleles (*gt* female) N = 20 and *giant* males with one allele (*gt* male) N = 16. On average the number of quantified nuclei per AP bin (2.5% egg length) is n = 499 (*hb* wt), n = 157 (*hb* def), n = 270 (*Kr*), n = 354 (*kni* early), n = 302 (*kni* late), n = 397 (*gt* female anterior region), n = 387 (*gt* female posterior region), n = 310 (*gt* male anterior region) and n = 277 (*gt* male posterior region). The confidence intervals for all point estimators of the data (mean, variance, noise, third cumulant and fourth cumulant; Figures 1, 2, and 3) were built by bootstrapping the empirical distribution of activity in each individual embryo. We used the 68% confidence intervals for the point estimators. All the error bars for the inferred parameters (Figure 5) correspond to the 10^{th} to 90^{th} percentiles of the marginal posterior distributions.

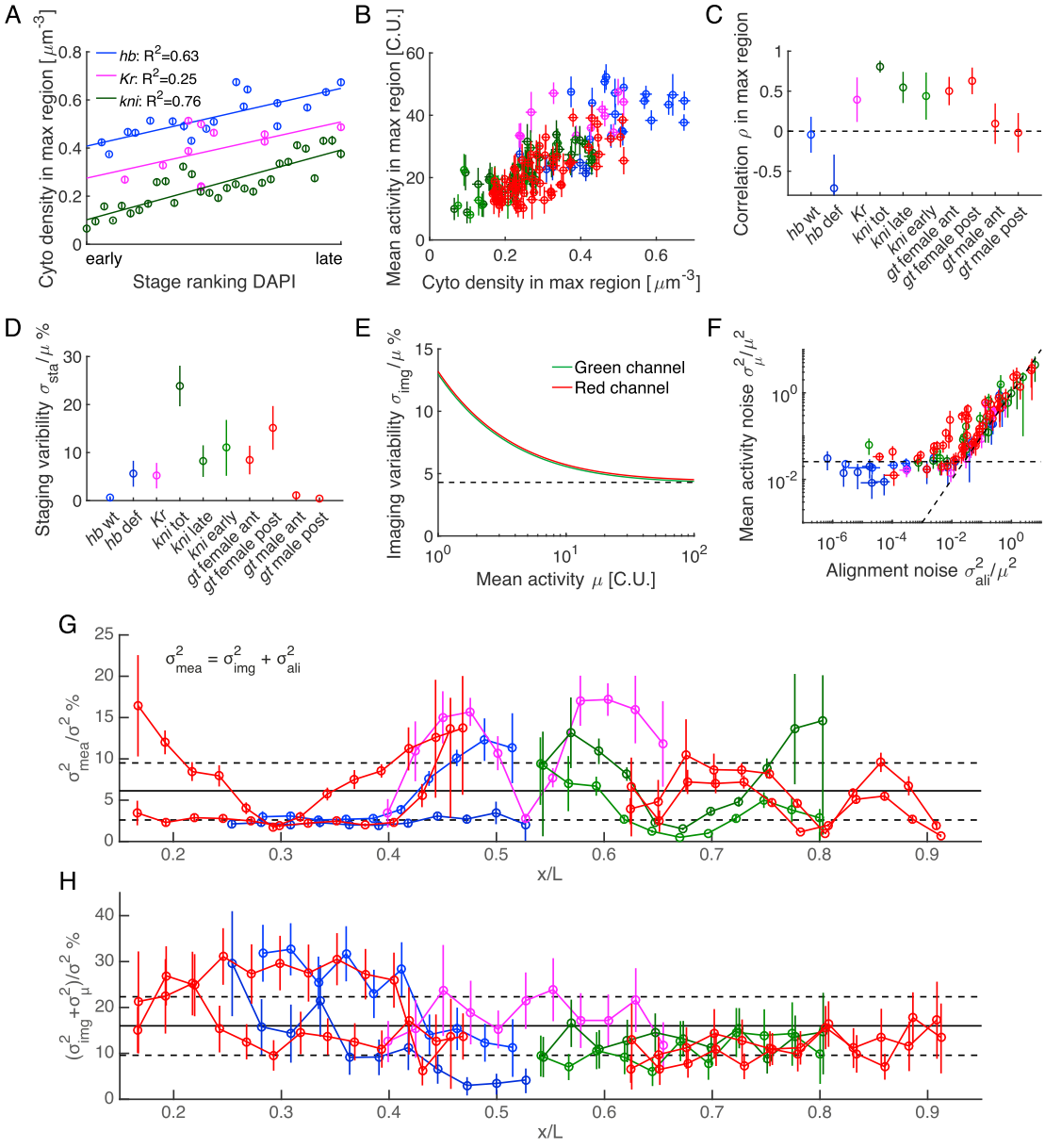


Figure S1. Temporal Staging, Measurement Error, and Embryo-to-Embryo Variability, Related to Figures 1 and 2

(A) Cytoplasmic mRNA density as a function of developmental stage during the 13th interphase as estimated from DAPI staining by eye-inspection. Each data point corresponds to a single embryo; cytoplasmic density was measured for each gene in the maximally expressed spatial region along the AP axis. Good correlation between manual ranking and the cytoplasmic mRNA accumulation correlation justifies the latter as a convenient proxy for time, and thus the developmental age of the embryos within nuclear cycle 13.

(B) Mean activity in the maximally expressed regions as a function of the cytoplasmic mRNA density. Each data point corresponds to a single embryo. Color code as in Figure 1C.

(C) Pearson correlation coefficient ρ between the mean activity and the cytoplasmic mRNA density calculated over the population of embryos in (B). Values indicate that up to 44% (ρ^2) of the variance in mean activity across embryos can be explained by staging uncertainty. The large correlation for *kni* (dark green) led to splitting the population of *kni* stained embryos into early and late stages to minimize the staging uncertainty in each subpopulation. We performed the splitting by finding the cytoplasmic density threshold that minimizes the sum of within-population variance in mean activity.

(D) Staging variability σ_{sta} in percent of the total mean activity μ for each gene in the respective maximally expressed regions. The staging variability corresponds to the variability in mean activity among embryos, which is explained by staging uncertainty between early and late embryo as estimated from cytoplasmic mRNA density. The staging variability σ_{sta} is defined as $\sigma_{sta} = \rho\sigma_\mu$, where σ_μ is the standard deviation of the mean activity across embryos. Note that the splitting of *kni* stained embryos into early and late stages was justified as the staging variability is significantly reduced. The overall small staging variability, which never exceeds 14%, indicates that the mean activity is sufficiently stable in time to warrant a steady state assumption.

(legend continued on next page)

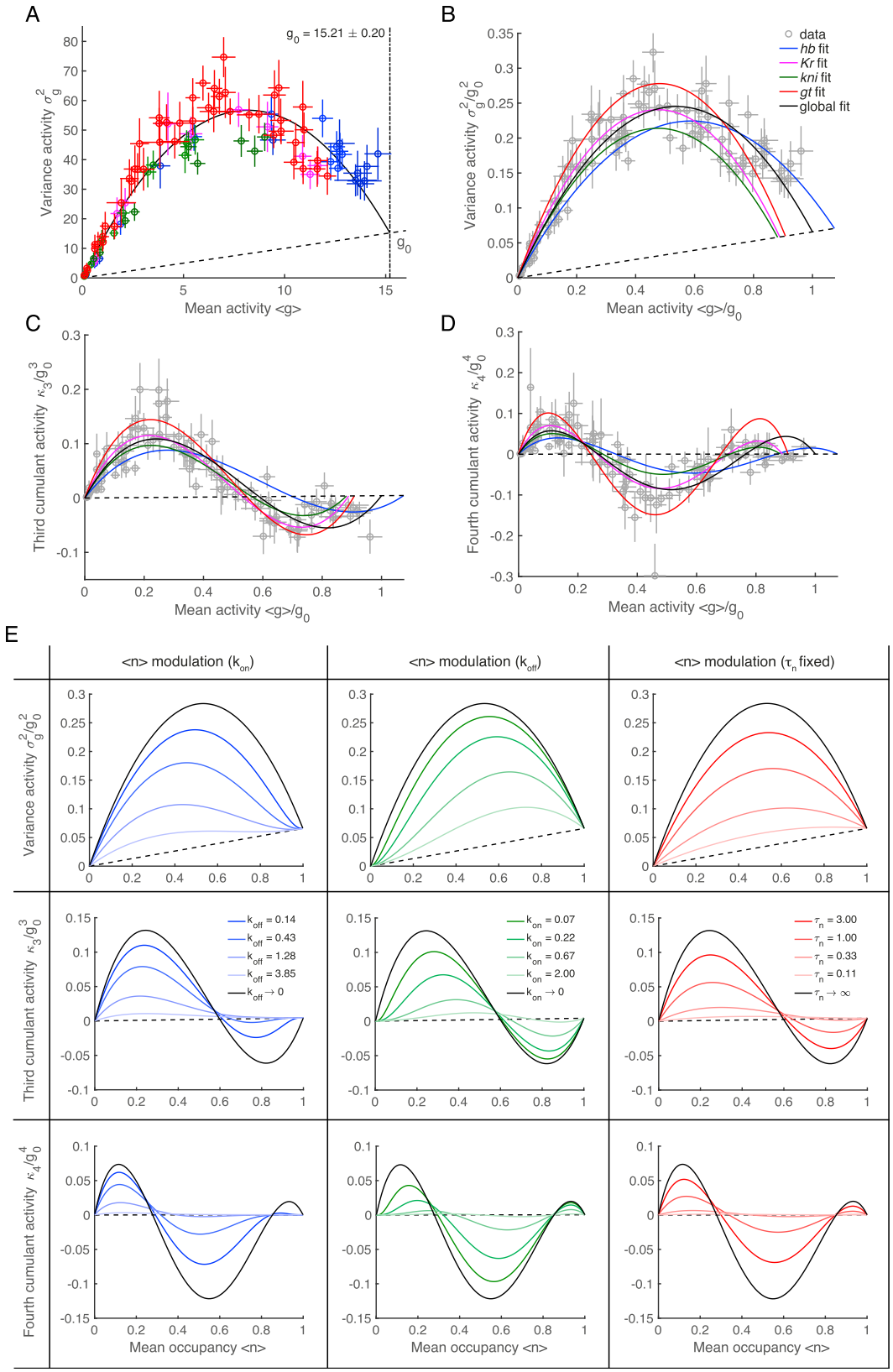
(E) Modeled imaging noise (CV) as a function of the mean activity for both channels. The imaging noise model was built from dual-color smFISH data using an alternating probe configuration (see Figure 2A). Imaging error σ_{img} was determined from the spread along the regression line between both channels (STAR Methods). Errors were assumed normally distributed, independent, and of equal magnitude in both channels. Thus, the modeled imaging error σ_{img} is characterized as the orthogonal spread along the fitted regression line, which was parameterized as $\sigma_{img}(v) = \sqrt{\sigma_b^2 + b_1 v + b_2 v^2}$, where (σ_b^2, b_1, b_2) are fit parameters, and v is the scalar projection of each data point onto the regression line. After fitting, the modeled imaging noise (CV) is given by $\sigma_{img}(v)/\mu$ with $v = \sqrt{\mu^2 + (a\mu)^2}$ for the green channel (green line) and $v = \sqrt{\mu^2 + (\mu/a)^2}$ for the red channel (red line), where a is the slope of the fitted line and μ is the mean activity.

(F) Variability of the mean across embryos (CV^2) as a function of alignment noise. Each data point corresponds to a single AP bin (2.5% egg length). The diagonal dashed line (slope = 1) highlights the correlation between the two quantities at the boundaries while the horizontal dash line corresponds to the embryo variability in the maximally expressed regions for each gene (Figure 2E). The correlation indicates that most variability across embryos in the transition regions can be explained by alignment noise, whereas the remaining variability in the maximally expressed regions reflects staging variability (C and D) and other extrinsic noise sources.

(G) Fraction of the total variance σ^2 corresponding to the measurement variance as a function of the AP position. Measurement variability σ^2_{mea} is defined as the combination of imaging σ^2_{img} and alignment variability σ^2_{ali} . The solid and dashed vertical lines are the overall mean fraction across genes and the 68% confidence interval, respectively.

(H) Fraction of the total variance σ^2 corresponding to the non-nuclear variance as a function of the AP position. The non-nuclear variance is the sum of the imaging σ^2_{img}, the alignment σ^2_{ali} and embryo variability σ^2_{emb}. The remaining variance $\sigma^2_{nuc} = \sigma^2 - \sigma^2_{img} - \sigma^2_{ali} - \sigma^2_{emb}$ is defined as the *nuclear variance* and is deemed intrinsic to transcription. Overall, the nuclear variance largely predominates as it represents 84% of the total variance, on average. The solid and dashed vertical lines are the overall mean fraction across genes and the 68% confidence interval. Color code as in Figure 1C.

All error bars are the 68% confidence intervals.



(legend on next page)

Figure S2. Mean-Cumulant Activity Relationships for a Single Gene Copy, Related to Figure 3

(A–D) The mean and the cumulants were corrected for different gene length, probe configuration and copy number. Each data point corresponds to a single AP bin and the error bars are the 68% confidence intervals. The dashed line stands for the Poisson background. Color code as in Figure 1C.

(A) Estimation of the maximal activity g_0 by fitting a 2^{nd} order polynomial of the mean activity to the variance. The maximal activity g_0 is determined as the second intercept of the fit with the Poisson background (vertical dashed line). In Figures 3B–3D and S2B-D the mean and the cumulants are normalized by the respective powers of g_0. Notably, $\langle g \rangle / g_0 = \langle n \rangle$ for constant k_{ini}.

(B–D) Normalized cumulants as a function of normalized mean activity. The solid lines are 2^{nd} (B), 3^{rd}(C) and 4^{th} (D) order polynomial fits, respectively. Fits were performed for each gene independently (colored lines); black line corresponds to the global fit of all genes (Figures 3B–3D). Individual fits are qualitatively similar, suggesting global trends in the data.

(E) Steady state two-state model cumulants as a function of the mean occupancy $\langle g \rangle / g_0 = \langle n \rangle$ for different scenarios of single parameter modulation (modulation of $\langle n \rangle$ through either k_{on} or k_{off} alone, or modulation of $\langle n \rangle$ at fixed correlation time τ_n by changing both k_{on} and k_{off}; Figures 3G–3I). For each considered modulation, only a single parameter is free since the value of g_0 (determined from A) has been fixed and the initiation rate k_{ini} is assumed constant. Varying the free parameters (graded colored lines) mainly affects the amplitude of the cumulants. The solid black lines stand for the common maximal amplitude limit attained when the correlation time goes to infinity.

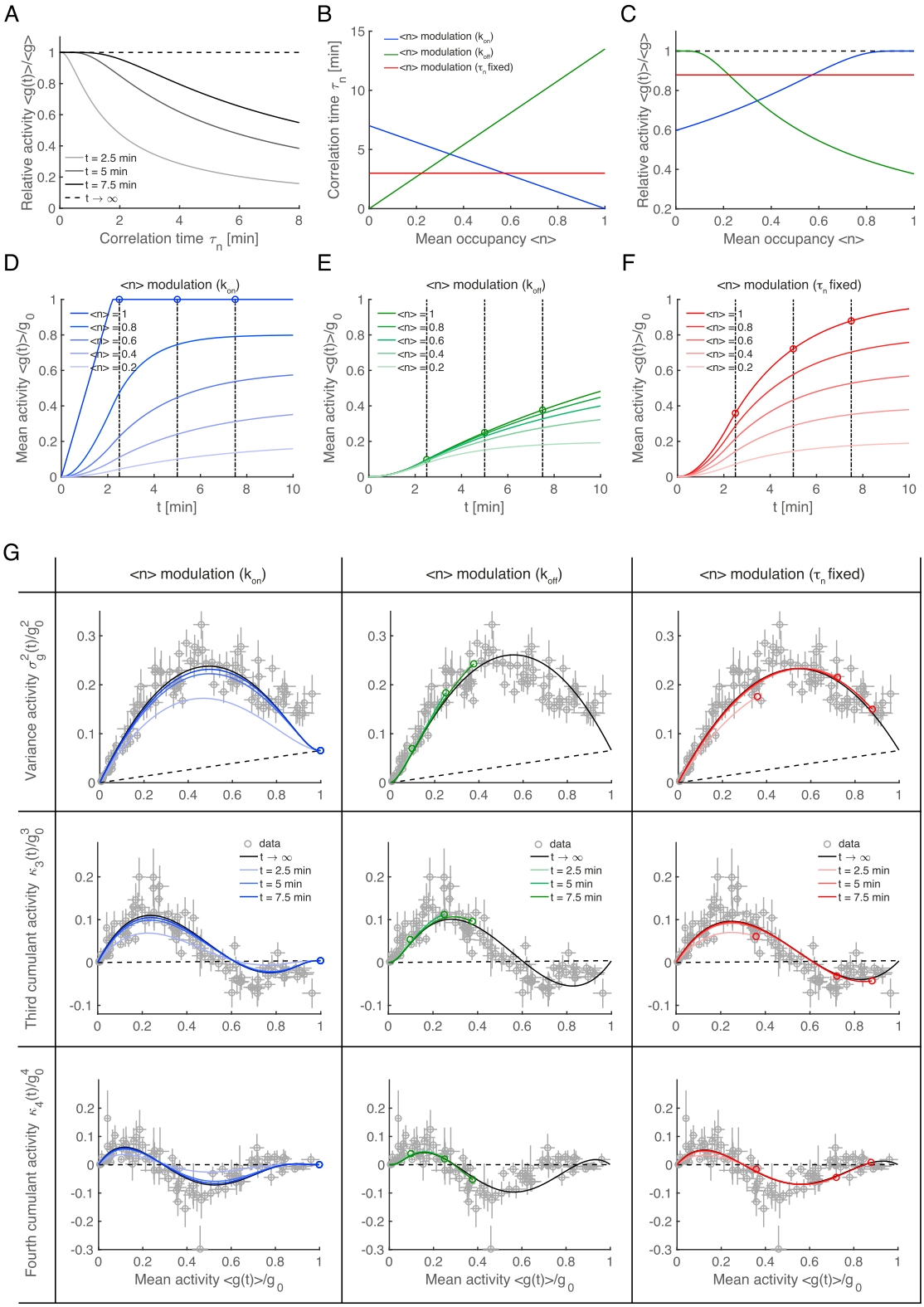


Figure S3. Time-Dependent Cumulant Analysis, Related to Figure 3

All time dependent-solutions of the two-state model were calculated with initial conditions $g(t=0)=0$ (no Pol II on the gene) and $n(t=0)=0$ (gene initially in the 'off' state).

(legend continued on next page)

(A) Time-dependent mean activity $\langle g(t)\rangle$ normalized by its steady state value $g(t\rightarrow\infty)\equiv g$ at three different times ($t = 2.5$ min, $t = 5$ min and $t = 7.5$ min) as a function of the switching correlation time τ_n. At steady state, the ratio is thus equal to one (horizontal dashed line). The correlation time is the only parameter that affects the relaxation to steady state. As τ_n increases, the relaxation becomes slower. For $t = 7.5$ min, a correlation time no larger than 3 min is required to reach approximately 90% of the maximal activity as observed in the data (Figure 3B).

(B) Correlation time τ_n as a function of the mean occupancy $\langle n\rangle$ for each best-fit single parameter modulation (from Figures 3G–3I). Modulation of k_{off} alone predicts a correlation time that is too large ($\tau_n \gg 3$ min) at high $\langle n\rangle$ to reach the maximal activity of the data at mid cycle (7.5 min).

(C) Time-dependent relative activity as a function of the mean occupancy $\langle n\rangle$ for each best-fit single parameter modulation (as in Figures 3G–3I). Same color code as in (B). The relative activity was calculated as the mean activity $\langle g(t)\rangle$ at $t = 7.5$ min normalized by its steady state value $\langle g\rangle$. Modulation of k_{off} alone clearly fails to reach steady state in time at high $\langle n\rangle$, as it only reaches 40% of the maximal activity. On the other hand, both modulation of k_{on} alone and of $\langle n\rangle$ at fixed τ_n reach a sufficiently large maximal activity to explain the data (100% and 88%, respectively).

(D–F) Normalized time-dependent mean activity $\langle g(t)\rangle/g_0$ as a function of time for each best-fit single parameter modulation (as in Figures 3G–3I). The circles correspond to the maximal attainable activity ($\langle n\rangle = 1$) after $t = 2.5, 5$ and 7.5 min (vertical dashed lines). Each modulation predicts different dynamics for boundary formation; for k_{on} modulation high $\langle n\rangle$ regions relax faster than low $\langle n\rangle$ regions (D), while it is the opposite for k_{off} (E). For fixed τ_n, all regions relax in synchrony independently of $\langle n\rangle$ (F). In the latter case, during interphase 13 the ratio of any two curves is constant in time, and thus these ratios are conserved across the patterning boundaries, which are uniquely determined by $\langle n\rangle$.

(G) Normalized time-dependent cumulants as a function of the normalized time-dependent mean activity for each best-fit single parameter modulation. The solid black lines correspond to the steady state best fits in Figures 3G–3I. The data in gray are identical to Figures 3B–3D and the error bars are given by the 68% confidence intervals. For sufficiently large t (i.e., $t > \{\tau_n, \tau_e\}$), the time-dependent mean and cumulant relationships closely follow the steady state ones. In addition, at fixed elongation time τ_e, the set of steady state cumulants are uniquely determined by k_{ini}, $\langle n\rangle$ and τ_n. Together, these two observations imply that even when far from steady state, fitting the steady state cumulants would still provide good estimates of the parameters, except that the estimated $\langle n\rangle$ would instead corresponds to the instantaneous mean occupancy $\langle n(t)\rangle \approx \langle g(t)\rangle/g_0$ (STAR Methods).

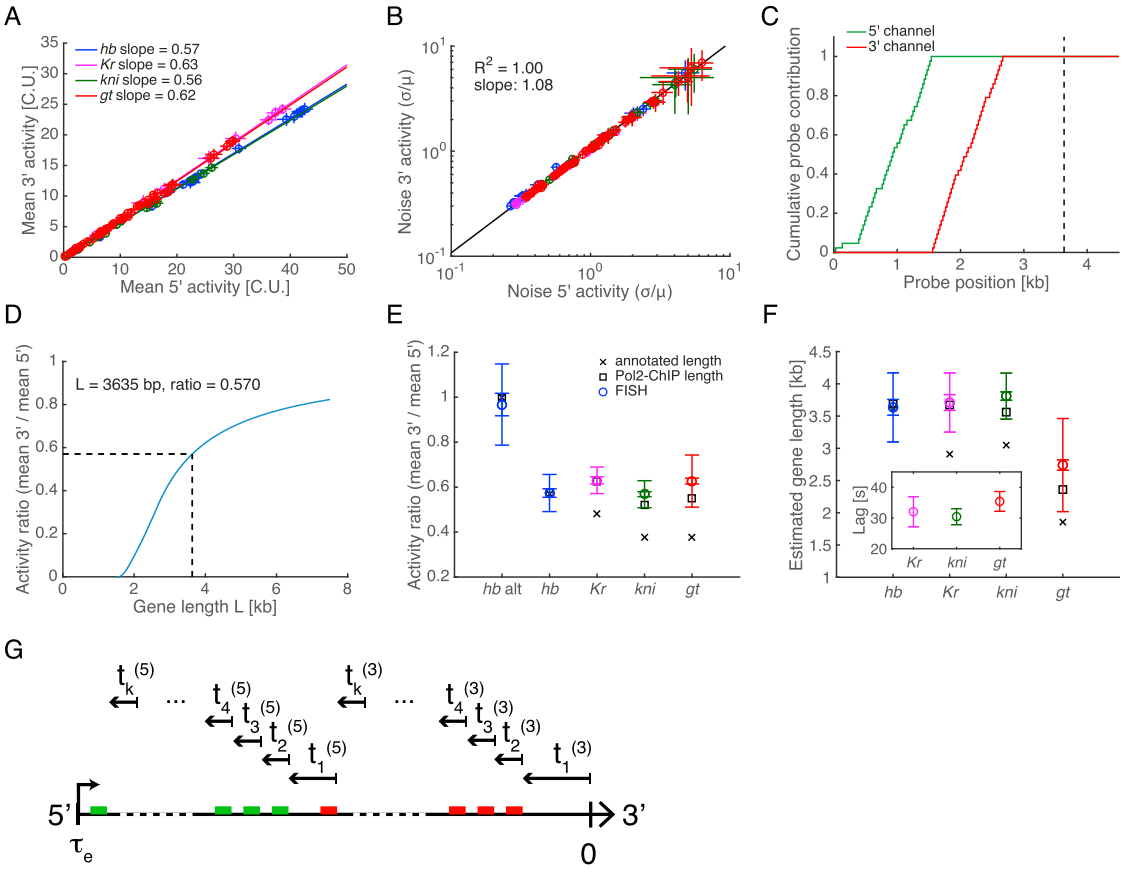


Figure S4. Link between Signal Properties from Dual-Color smFISH and Probe Configuration for Each Gene, Related to Figure 4

(A) Mean 3′ versus 5′ activity for all gap genes. Each data point corresponds to the mean activity over all embryos in a single AP bin. The slopes for the different genes depend on the exact probe configuration. Error bars are the 68% confidence intervals.

(B) 3′ versus 5′ noise (CV). The excellent correlation and the slope close to one suggest that the switching correlation time τ_n is on the order of the elongation time τ_e. Indeed, if $\tau_n < \tau_e$, one would have expected more buffering of the switching noise on the 5′ end compared to the 3′ end, whereas if $\tau_n \gtrsim \tau_e$ the magnitude of the noise should be similar on both ends. Error bars as in (A).

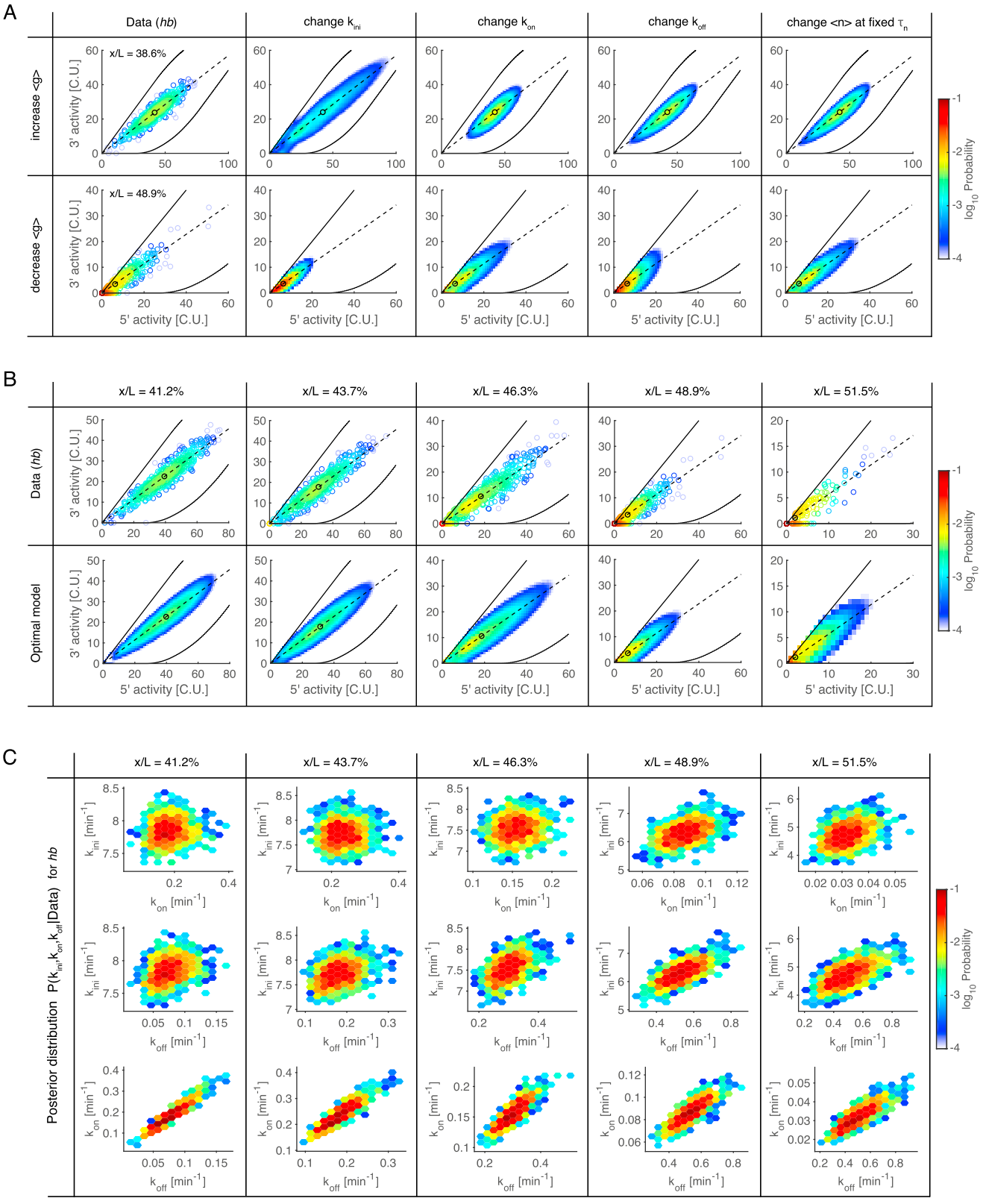
(C) Cumulative *hb* probe contribution to the fluorescence signal as a function of transcript length. The vertical dashed line corresponds to the length of a cytoplasmic mRNA for *hb* (3635 bp). Transcripts whose length is larger than 2667 bp would contribute as 1 cytoplasmic unit in both channels.

(D) Activity ratio (mean 3′ signal over mean 5′ signal) as a function of gene length for *hb* (blue line). Assuming elongation to occur at constant speed and instantaneous release of transcripts, the ratio is fully determined by the probes' location and the gene length (transcribed region). The activity ratio results from the ratio of the integrals of the cumulative probe contribution in (C).

(E) Activity ratio for each gene. The circles stand for the measured ratio with error bars (both standard errors and standard deviations are shown) obtained from the propagation of the normalization errors in both channels for all embryos. The crosses correspond to the predicted ratio based on the annotated gene length. The squares are derived from Pol2 occupancy data (Pol2-ChIP; Blythe and Wieschaus, 2015). For *Kr*, *kni* and *gt*, Pol2 signal is found a few hundreds bp away from the annotated length suggesting extra processing related to termination. Similarly, the larger measured ratios (compared to the predicted ones based on annotated gene length (crosses)) likely reflect retention of nascent transcripts at the loci due to termination.

(F) Effective gene length for each gene as determined from the activity ratio. Symbols and error bars as in (E). Assuming an elongation speed of 1.5 kb/min, the difference between the effective and annotated gene length can be translated in time (inset). The lag or extra residence time of transcripts at the loci is at most 35 seconds.

(G) Temporal hierarchy used to calculate the 5′ and 3′ joint distribution of transcriptional activities. The measured signal result from partially tagged nascent transcripts and is proportional to the number of probe binding regions that have been transcribed. In order to calculate the joint distribution, we accumulate the distinct contribution of nascent transcripts, between each probe region, from the 3′ end up to the 5′ end of the gene. At constant elongation rate, the distance separating each successive probe region is converted into a time $t_i^{(C)}$, where the superscript $(C) \in \{(3),(5)\}$ stands for the probe channel and the subscript i denotes the interval separating probe i from probe $i-1$ (from the 3′ end to 5′ end direction). The joint distribution of activity is obtained by subsequent convolution of the distribution of Pol II initiated during each time $t_i^{(C)}$. Each of these convolutions are properly weighted to take into account the proper contribution of each probe region to the activity.



(legend on next page)

Figure S5. Parameter Inference from Dual-Color smFISH Activity Distribution Using the Two-State Model, Related to Figure 4

(A and B) The data correspond to the measured distribution of 3′ versus 5′ activity across AP position for *hb*. Data distributions were constructed based on the 2.5%-AP-bins defined in Figure 4C. Dashed black line represents the expected ratio of 3′ versus 5′ activity ($r = 0.57$ for *hb*); black circle corresponds to the mean of the distribution and lies on the dashed line.

(A) Qualitative change of the distribution predicted by the 2-state model as the parameters vary for AP-bin at $x/L = 38.6\%$ (top row) and at $x/L = 48.9\%$ (bottom row). Changes in the transcriptional parameters k_{ini}, k_{on}, k_{off}, and in $\langle n \rangle$ at fixed τ_n set at the same mean activity $\langle g \rangle$ as in the data leads to qualitatively different distributions. Thus, all information regarding the kinetic parameters is contained in the distribution of 3′ versus 5′ activity, which enables inference of these rates.

(B) Side by side representation of the empirical (data, top row) and modeled (bottom row) distributions with best-fit parameters for different AP bins. The empirical distributions are used as input in our inference framework enabling precise inference of the underlying transcriptional kinetics at each AP position. Of note, the displayed modeled distributions are devoid of measurement noise and represent the theoretical output of the two-state model given the probe-set configuration and the effective elongation time. Thus, the likelihood of the data is essentially the convolution of the activity distribution calculated from the two-state model with the noise measurement distribution. Overall, the best-fit distributions reproduced the data well.

(C) Joint posterior distribution of the parameters given the data in (B) for each AP position. These distributions are generated as the output of our inference framework, namely we sampled the posterior distributions calculated from the likelihood according to Bayes' rule using a Markov chain Monte Carlo (MCMC) algorithm. As the joint posterior distributions are highly peaked in the parameter space, it indicates that the parameters of the model are identifiable for all AP positions. The optimal kinetic rates k_{ini}, k_{on} and k_{off}, which were used to generate the modeled distribution in (B), are estimated from these joint posteriors as the median of the marginal posterior distributions.

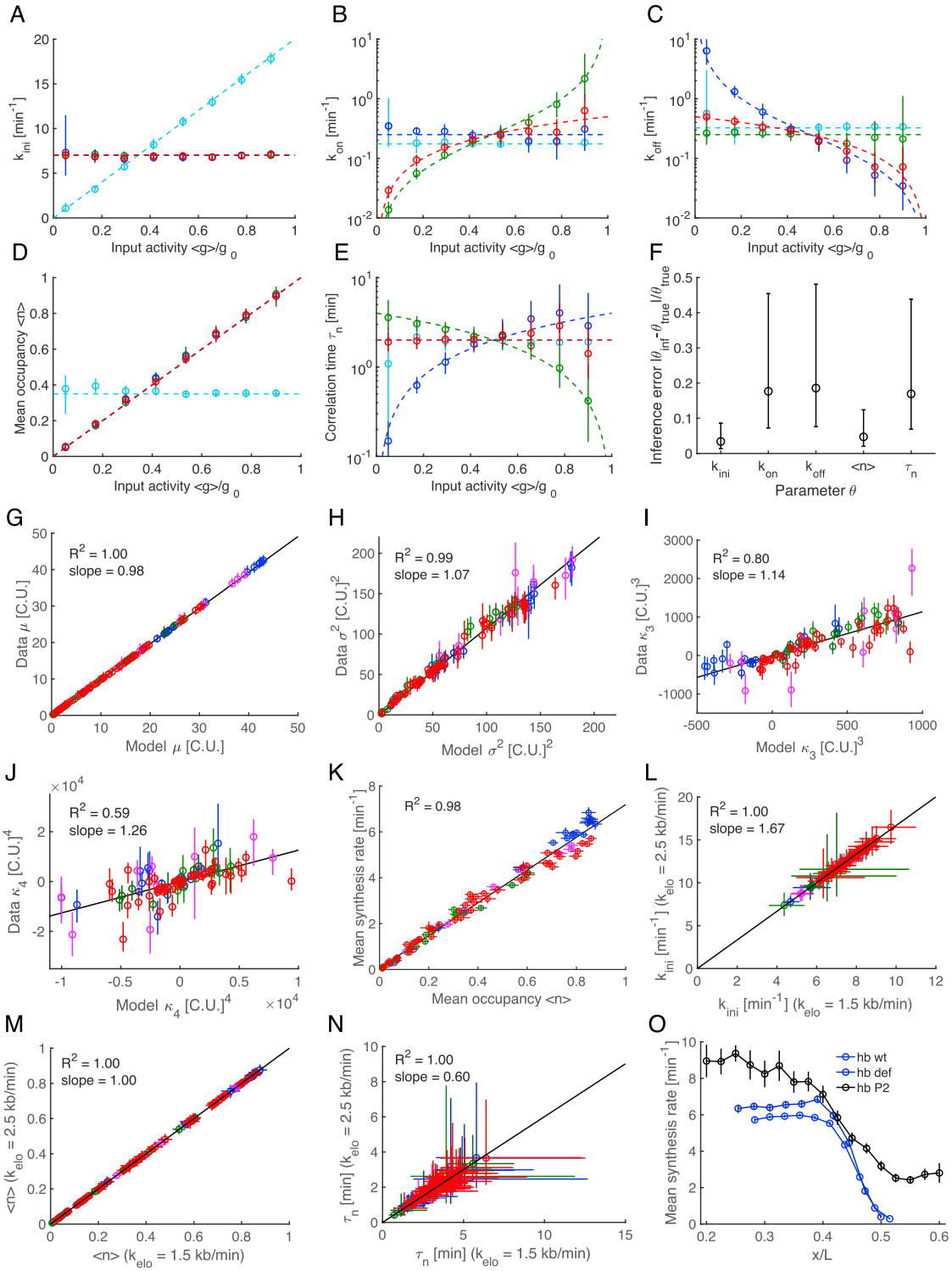


Figure S6. Validation of the Inference Framework for Dual-Color smFISH and Synthesis Rates, Related to Figures 4 and 5

(A–F) We simulated synthetic 3′ and 5′ nuclear activity data based on four gene copies (two alleles with two sister chromatids each) modeled by a two-state model with measurement noise, using the probe configuration for *hb*. To test the performance of our inference, we generated four different datasets by modulating the mean input activity $\langle g \rangle$ in the data through: 1) initiation rate k_{ini} alone (cyan), 2) on-rate k_{on} alone (green), 3) off-rate k_{off} alone (blue) and the mean occupancy $\langle n \rangle$ at constant switching correlation time τ_n (red). The constant g_0 corresponds to the maximal activity for each dataset, defined as $g_0 = \max(k_{ini})\tau_e$, where the maximum is taken over the dataset when k_{ini} varies (cyan) and τ_e is the elongation time. Importantly, the inference of the kinetic parameters was performed for

(legend continued on next page)

each sub-dataset independently (individual circles; 500 nuclei), without assuming any continuity in the dataset. To take into account finite size sampling variation in the data, we inferred parameters on 10 replicates for each synthetic dataset. Thus, the estimated posterior distributions are aggregated over all replicates.
(A–E) Inferred kinetic rates k_{ini} (A), k_{on} (B), and k_{off} (C), mean occupancy $\langle n \rangle = k_{on}/(k_{on}+k_{off})$ (D) and switching correlation time $\tau_n = 1/(k_{on}+k_{off})$ (E) as a function of the mean input activity $\langle g \rangle / g_0$. All quantities are estimated from the sampled joint posterior distribution of the kinetic rates. Colored circles stand for the inferred parameters as a function of input activity, i.e., the median of the marginal posterior distribution. Error bars correspond to the 10^{th} and 90^{th} percentiles of the posterior distribution. The colored dashed lines represent the input (true) parameters used to simulate the data.
(F) Global relative inference error $|\theta_{inf} - \theta_{true}|/\theta_{true}$ calculated for each parameter θ. These errors are estimated over all synthetic datasets and replicates and correspond to the median with error bars given by the 68% confidence intervals. Notably, k_{ini} and $\langle n \rangle$ are easier to infer than the switching rates k_{on} and k_{off} or the correlation time τ_n, which have more subtle effect on the shape of the activity distribution. Still, the inference is able to distinguish between small differences in parameter modulation. Overall, the errors remain small, as the medians of the inference errors never exceed 20% of the true values.
(G–J) Four first cumulants of data (unnormalized, in cytoplasmic units) as a function of the ones predicted by the two state-model with best fitting parameters for multiple gene copies ($N_g = 2, 4$). Each data point corresponds to a single AP-bin. Error bars are the 68% confidence intervals. Overall, the slopes close to one and the large R^2 indicate that the model captures the first four cumulants of data well. Color code as in Figure 1C.
(K) Inferred mean synthesis rate $k_{ini}\langle n \rangle$ as a function of the mean occupancy $\langle n \rangle$ for all genes. Modulation of transcript mean synthesis rate across boundaries is fully determined by the mean occupancy. Color code as in Figure 1C. All error bars correspond to the 10^{th} and 90^{th} percentiles of the posterior distribution.
(L–N) Comparison of the inferred transcriptional parameters k_{ini}, $\langle n \rangle$ and τ_n assuming two different elongation speeds k_{elo} (1.5 kb/min versus 2.5 kb/min). Both k_{ini} and τ_n are rescaled while $\langle n \rangle$ remains the same. Thus, our results are unaffected by the exact value of k_{elo}; it only leads to a rescaling of the inferred parameters that have time units. Color code and error bars as in (K).
(O) Comparison of the estimated mean synthesis rate for a single gene copy of endogenous *hb* (wt and deficient) and the synthetic *hb* P2 reporter live-imaged by Garcia et al. (2013) during interphase 13. The reporter corresponds to a minimal version of the *hb* gene that is driven by the P2 promoter and the P2 (proximal) enhancer alone. The mean synthesis rate of the P2 reporter was obtained by multiplying the estimated effective initiation rate and the fraction of active nuclei divided by two (two sister chromatids per locus), as reported in Garcia et al. (2013). Excluding the posterior region ($x/L > 0.45$), where the reporter shows ectopic expression, the estimated mean synthesis rates only differ by approximately 30 to 50%. This difference, in the case of the reporter, likely stems from both larger live-imaging measurement and calibration errors, and potentially reflects different expression rates between the endogenous gene and the synthetic reporter. Nevertheless, the reported synthesis rates estimated through different models and techniques are consistent. Error bars as in (K) except for the *hb* P2 reporter which are standard errors over multiple embryos.