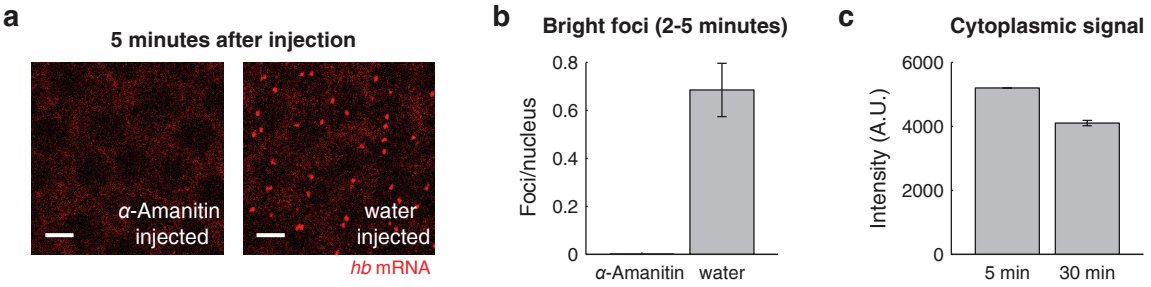
Supplementary Information

Combining Protein and mRNA Quantification to Decipher Transcriptional Regulation

Heng Xu, Leonardo A. Sepulveda, Lauren Figard, Anna Marie Sokac and Ido Golding

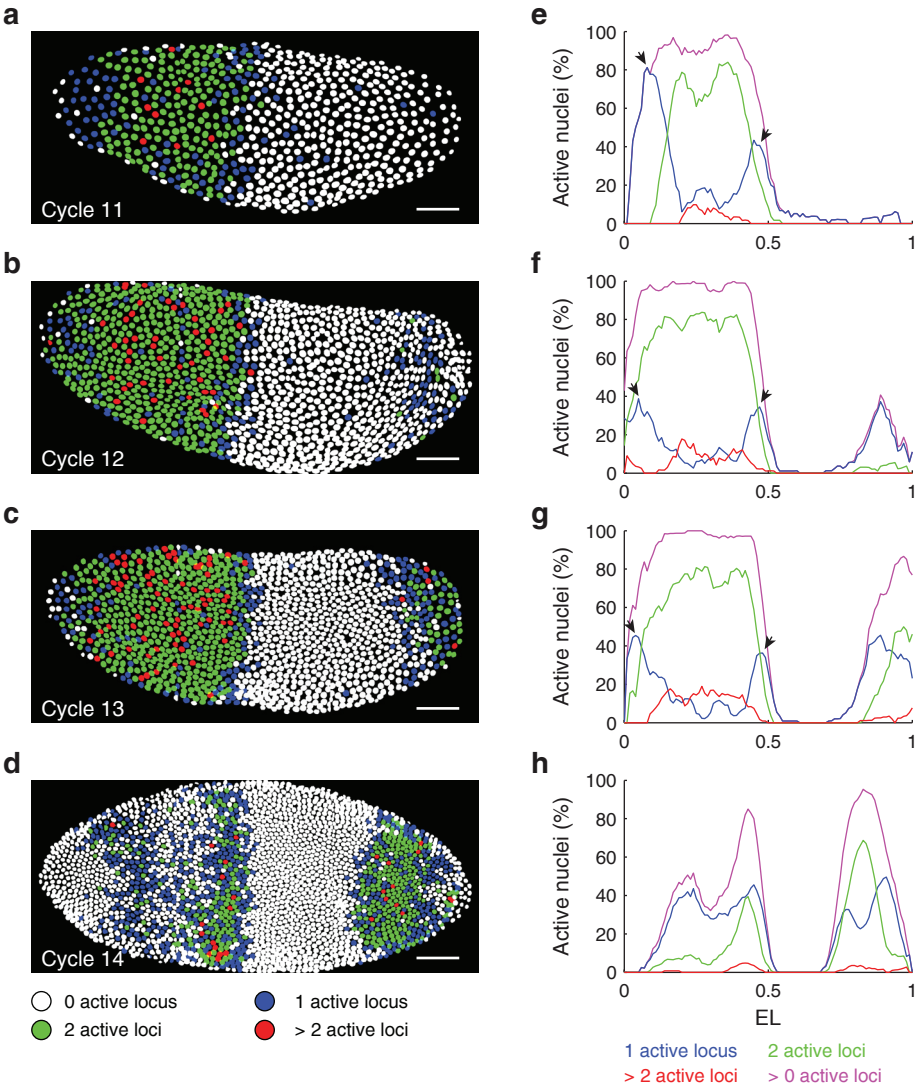
Supplementary Figure 18	Relating *hb* transcription to Bcd and Hb binding
Supplementary Table 1	Sequences of smFISH probes
Supplementary Note	1. Image acquisition and processing 2. mRNA quantification 3. Protein quantification 4. Measuring the gene regulation function (GRF) 5. Measuring Bcd binding 6. Modeling transcription kinetics 7. Deriving the fluctuation method for measuring protein concentration 8. Error estimates Figures SN1-SN10

Supplementary Figure 1. Inhibiting *hb* transcription by injecting embryos with α-amanitin.



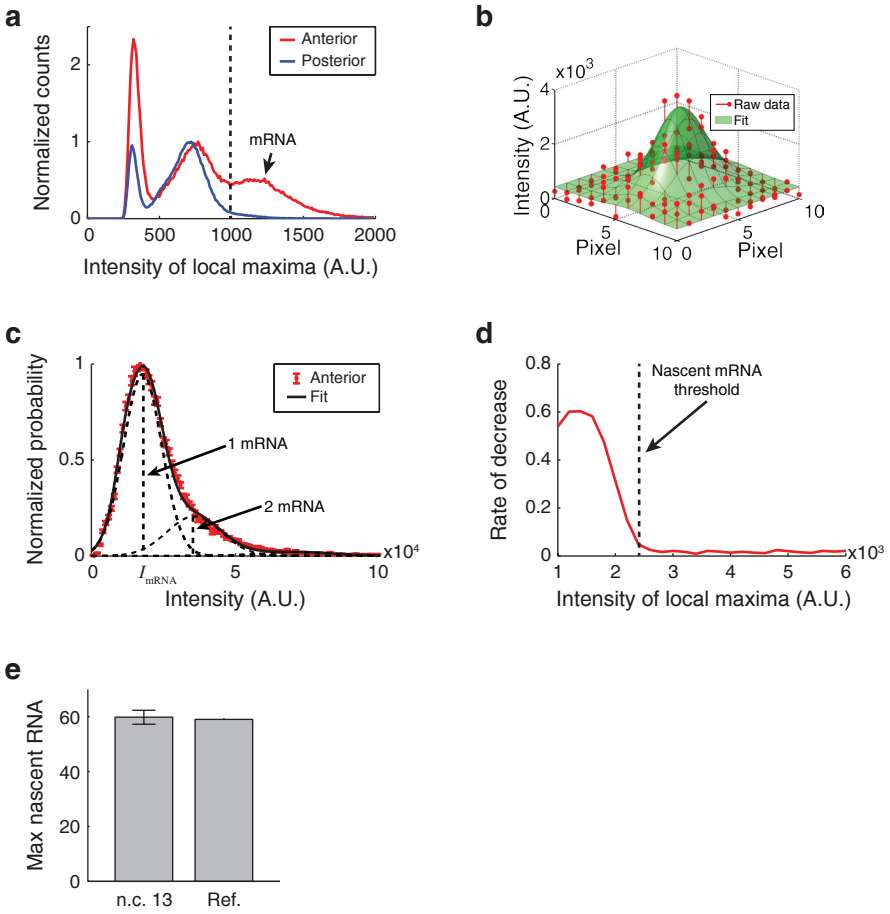
(**a**) *hb* mRNAs (red) after injection of α-amanitin (left) or water (right). Scale bar, 2 μm. (**b**) A comparison of the number of bright foci between α-amanitin-injected embryos (5 embryos, cycles 11-14) and the control (3 embryos, cycles 11-14). After injection and a brief incubation period (~2-5 minutes), foci were counted within the AP position range 0.3-0.7 EL. Error bars indicate S.E.M. Bright foci vanished completely within <5 minutes of inhibiting transcription, consistent with the release of nascent mRNA from the gene upon completion of synthesis (Garcia et al., 2013; Larson et al., 2011). (**c**) The change in cytoplasmic mRNA signal at 5 and 30 minutes after α-amanitin injection. The cytoplasmic signal was measured as the difference in average intensity between the anterior cytoplasmic region (0.3-0.45 EL) and the posterior cytoplasmic region (0.55-0.7 EL) of the embryo (data from 4 embryos, cycles 11-14). Error bars indicate S.E.M. The cytoplasmic signals survived for >30 minutes, consistent with the estimated lifetime of mature hb mRNA (~1 hour, (Little et al., 2013)).

Supplementary Figure 2. The spatial profile of *hb* transcription at different nuclear cycles.



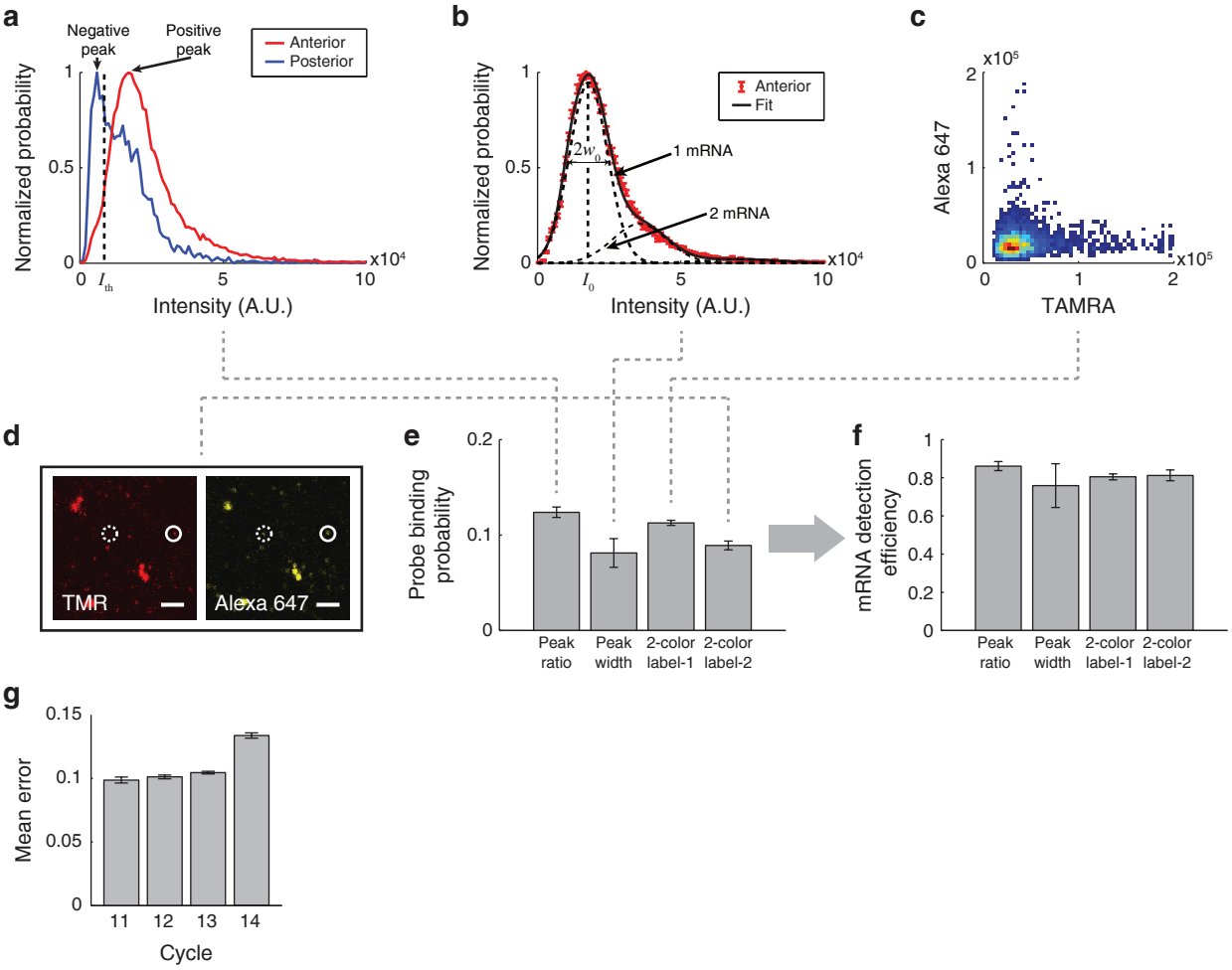
(**a**) – (**d**) The transcriptional activity of individual nuclei in wild-type embryos during cycles 11-14. The color of each nucleus indicates the number of active *hb* loci (see legend). Scale bar, 50 μm. (**e**) – (**h**) Percentage of nuclei containing different numbers of active *hb* loci as a function of the AP position (data from the same embryos shown in Panels **a**-**d**). For cycles 11-13 embryos, more than 90% of the nuclei in the anterior expression band of the embryo (0.15-0.45 EL, (Porcher et al., 2010)) contained at least one active *hb* locus (23 embryos examined), consistent with previous reports using both live and fixed embryos (Garcia et al., 2013; Lucas et al., 2013; Porcher et al., 2010). In particular, the majority of nuclei in this region had both *hb* copies active, while nuclei with a single active locus were enriched at the anterior tip and the posterior border of the expression band (arrows), again consistent with previous observations (Porcher et al., 2010).

Supplementary Figure 3. Quantifying *hb* mRNA.



(**a**) The distribution of pixel intensity of candidate mRNA spots in the anterior (red) and posterior (blue) parts of the embryo (>300,000 spots from a single cycle 12 embryo). A threshold (dashed line) was used to identify spots corresponding to real *hb* mRNAs (arrow), and discard false-positive particles resulting e.g. from nonspecific probe binding. (**b**) Fitting the local intensity profile (red) of a candidate spot using a 2D Gaussian function (green). The fitting result was used to calculate the total fluorescence intensity of the spot. (**c**) The distribution of total fluorescence intensity of smFISH spots at the anterior part of the embryo (red, mean ± S.E.M., >10,000 spots from a single cycle 12 embryo). By fitting the histogram to the sum of Gaussian functions (black), the intensity corresponding to a single mRNA molecule (I_{mRNA}) was identified. This procedure was repeated independently for each imaged embryo, to correct for any differences due to labeling or imaging. (**d**) The rate of decrease in number of candidates for different threshold values (>300,000 candidate spots from a single cycle 12 embryo). The lowest intensity value corresponding to <8% rate-of-decrease (dashed line) was selected as the threshold to identify candidates for active transcription sites. (**e**) Comparing *hb* expression level in the strongest expression region during cycle 13 (0.34 ± 0.02 EL, mean ± S.E.M., data from 7 embryos) with data from (Little et al., 2013). Error bars indicate S.E.M.

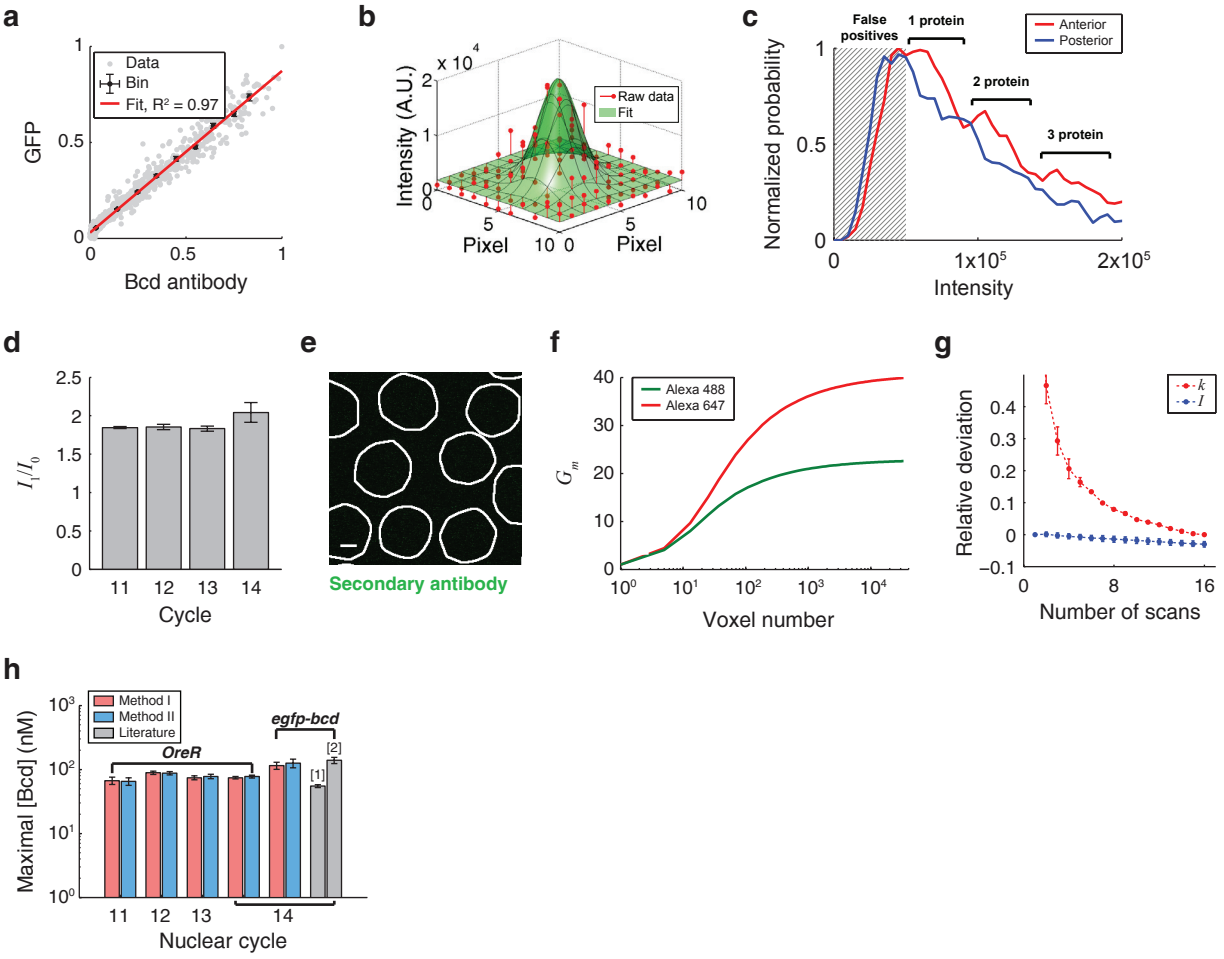
Supplementary Figure 4. Estimating the error in mRNA quantification.



(**a**) The intensity histogram of *hb* smFISH spots in the anterior (red) and posterior (blue) parts of the embryo (>10,000 spots from a single cycle 11 wild-type embryo). A threshold I_{th} (dashed line) was used to discriminate false positive spots from real mRNA spots. The most probable intensity values of the two species (arrows) were used to calculate the probe binding probability. (**b**) The intensity histogram of *hb* smFISH spots in the anterior part of the embryo (red, mean ± S.E.M., >10,000 spots from a single cycle 11 wild-type embryo). By fitting the histogram to the sum of Gaussian functions (black), the typical intensity I_0 and the half-width w_0 corresponding to a single *hb* mRNA were extracted. (**c**) The 2D spot-intensity histogram for the competitive two-color smFISH experiment (>20,000 spots from a single cycle 12 *bcd3-lacZ* embryo). Two identical sets of *lacZ* smFISH probes, conjugated with TAMRA and Alexa 647, respectively, were used. (**d**) Comparing the two channels in the noncompetitive two-color smFISH experiment. An image of a small area of a cycle 11 *bcd3-lacZ* embryo is shown. Scale bar, 2 μm. To label the *lacZ* mRNA, 30 probes corresponding to the 5' half of the *lacZ* gene were conjugated with TAMRA, while another 30 probes corresponding to the 3' half of the *lacZ* gene were conjugated with Alexa 647. smFISH spots that appeared in both channels (solid circle) were distinguished from those that only appeared in one channel (dashed circle). The probability of detecting a spot in both channels was used to calculate the probe binding probability. (**e**) A comparison of the estimated probe binding probability obtained using different methods. For each method, data is from 3-4 embryos. Error

bars indicate S.E.M. The estimated binding probabilities are comparable to values reported in previous studies (Boettiger and Levine, 2013; Little et al., 2013). (**f**) A comparison of the estimated *hb* mRNA detection efficiency obtained using probe binding probabilities extracted from different methods. For each method, data is from 3-4 embryos. Error bars indicate S.E.M. (**g**) A comparison of the average error in nascent mRNA quantification between different cycles (cycle 11, 7 embryos; cycle 12, 9 embryos; cycle 13, 7 embryos; cycle 14, 7 embryos). Error bars indicate S.E.M.

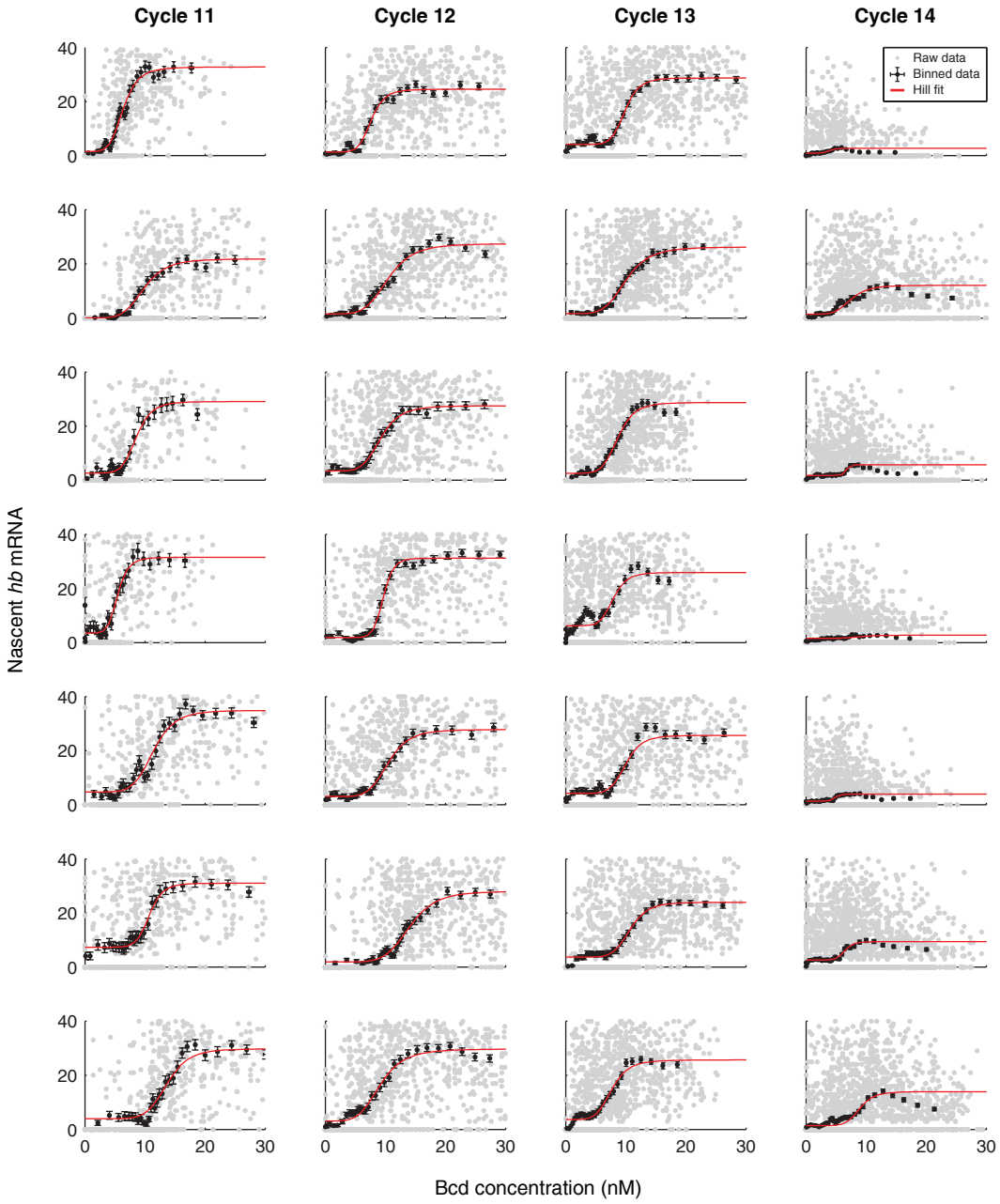
Supplementary Figure 5. Quantifying Bcd concentration.



(**a**) A comparison of nuclear fluorescence signals between Bcd antibody and GFP fluorescence (~2000 nuclei of a single *egfp-bcd* embryo, cycle 14). Data from individual nuclei (gray) was binned along the Bcd axis (black, mean ± S.E.M.) and fitted to a straight line (red, R^2 = 0.91 ± 0.02, mean ± S.E.M. when averaged over 12 embryos). This comparison reveals a linear relation covering the full range of protein concentrations, indicating reliable representation of Bcd protein level by the antibody signal. (**b**) Fitting of the local intensity profile (red) of a candidate spot in the cytoplasm using a 2D Gaussian function (green). The fitting result was used to calculate the total fluorescence intensity of the spot. (**c**) The intensity histogram of cytoplasmic Bcd spots in the anterior (red) and posterior (blue) parts of the embryo (>20,000 spots from a single cycle 12 embryo). The histogram was used to discard false-positive spots (shaded region) and to distinguish spots corresponding to different numbers of Bcd molecules (brackets). (**d**) The ratio between the typical intensity of a single Bcd molecule (I_1) and the typical intensity of a false-positive spot (I_0). Data from different cycles is shown (cycle 11, 7 embryos; cycle 12, 9 embryos; cycle 13, 7 embryos; cycle 14, 7 embryos). Error bars represent S.E.M. (**e**) Confocal image of the anterior region of a cycle 12 embryo. The embryo was only labeled using secondary antibody (green). Nuclear boundaries are marked in white. No fluorescence spots were identified. Scale bar, 2 μm. (**f**) The parameter G_m as a function of the voxel number for different fluorophores. The nucleus was modeled as a 2D square lattice of m voxels. (**g**) The change in mean and

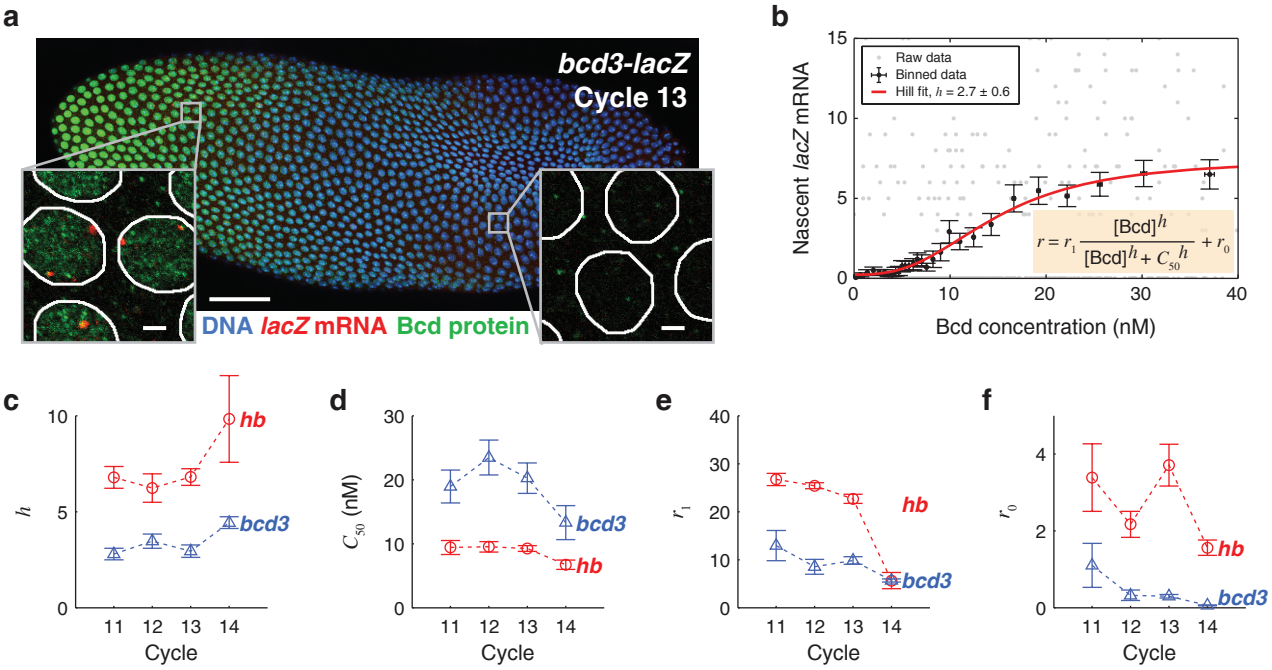
fluctuation of nuclear Bcd signal for different numbers of confocal scans. Red, the relative deviation of k (the slope between intensity variance and mean) from its minimal value (obtained from a 16× scan). Blue, the relative deviation of the mean intensity I from its maximal value (obtained from a 1× scan). Data from the anterior region of a single cycle 11 embryo. Error bars represent S.E.M. (**h**) A comparison of the maximal Bcd concentration between different cycles and different fly strains. The average concentration in the brightest 1% nuclei of individual embryos was calculated using the two methods described in the text. In wild type embryos, Bcd concentration during cycles 11-14 does not change much, and the two methods show good agreement (cycle 11, 7 embryos; cycle 12, 9 embryos; cycle 13, 7 embryos; cycle 14, 7 embryos). *egfp-bcd* transgenic embryos (5 embryos) have higher Bcd concentrations than wild-type during cycle 14. Values from the literature ([1] Gregor et al. 2007; [2] Abu-Arish et al. 2010) are shown in grey. Error bars indicate S.E.M.

Supplementary Figure 6. The regulatory relation between Bcd and *hb* for multiple individual embryos.



The gene regulation function measured in 28 wild-type embryos between cycles 11 and 14 (7 embryos per cycle). The number of nascent mRNAs from individual *hb* loci is plotted versus the nuclear Bcd concentration (AP positions 0.25-0.7 EL). Gray, individual loci. Black, data binned along the Bcd axis (mean ± S.E.M.). Red, fit to a Hill function.

Supplementary Figure 7. The regulatory relation between Bcd and the *bcd3-lacZ* reporter gene.

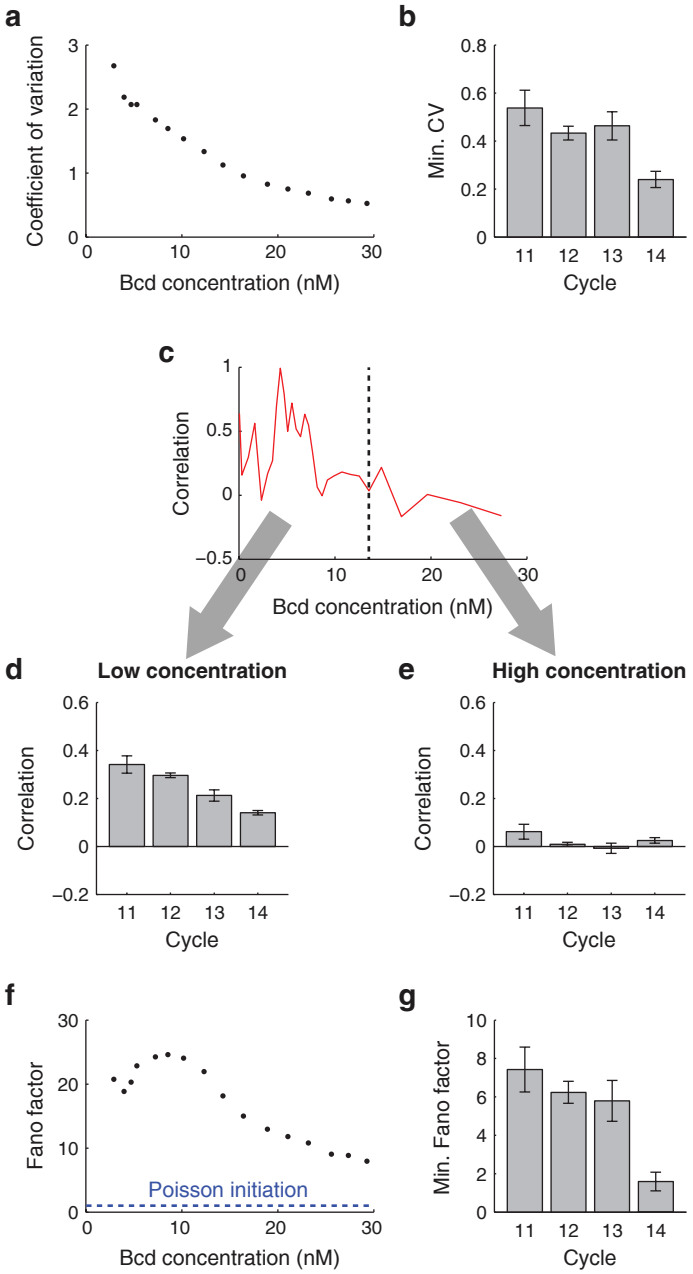


(**a**) Confocal image of a *bcd3-lacZ* embryo during cycle 13. Bcd (green) was labeled using immunofluorescence, *lacZ* mRNA (red) was labeled using smFISH, and DNA (blue) was stained with Hoechst. Scale bar, 50 μm. Insets: Magnified views of a small anterior region (left) and a small posterior region (right). Scale bar, 2 μm. (**b**) The gene regulation function measured in the embryo shown in Panel **a**. The number of nascent mRNAs from individual *lacZ* loci is plotted versus the nuclear Bcd concentration (data from >800 nuclei, AP positions 0.15-0.7 EL). Gray, individual loci. Black, data binned along the Bcd axis (mean ± S.E.M.). Red, fit to a Hill function. (**c**) – (**f**) Parameters of the gene regulation function, extracted from the fit to a Hill function. Red circles, *hb* gene in wild-type embryos (cycle 11, 7 embryos; cycle 12, 9 embryos; cycle 13, 7 embryos; cycle 14, 7 embryos). Blue triangles, *bcd3-lacZ* in transgenic flies (cycle 11, 6 embryos; cycle 12, 7 embryos; cycle 13, 7 embryos; cycle 14, 3 embryos). Error bars represent S.E.M.

The concentration threshold for *hb* activation, C_{50} = 9.4 ± 0.5 nM, is consistent with previous estimates using other methods (Burz et al., 1998; Gregor et al., 2007a; Ma et al., 1996). In contrast, the *bcd3-lacZ* gene showed a ~2 fold higher Bcd activation threshold C_{50} relative to endogenous *hb*, consistent with previous reports regarding the effect of reducing the number of transcription-factor binding sites (Becskei et al., 2005).

Compared to cycles 11-13, cycle 14 embryos displayed a distinctly different regulatory relation between Bcd and *hb*: A higher Hill coefficient, a lower activation threshold, and a lower level of maximum *hb* expression. These differences from earlier cycles indicate changes in the regulation of *hb* expression. They are consistent with previous reports on the involvement of additional transcription factors (Lopes et al., 2012) and other enhancers (Perry et al., 2012).

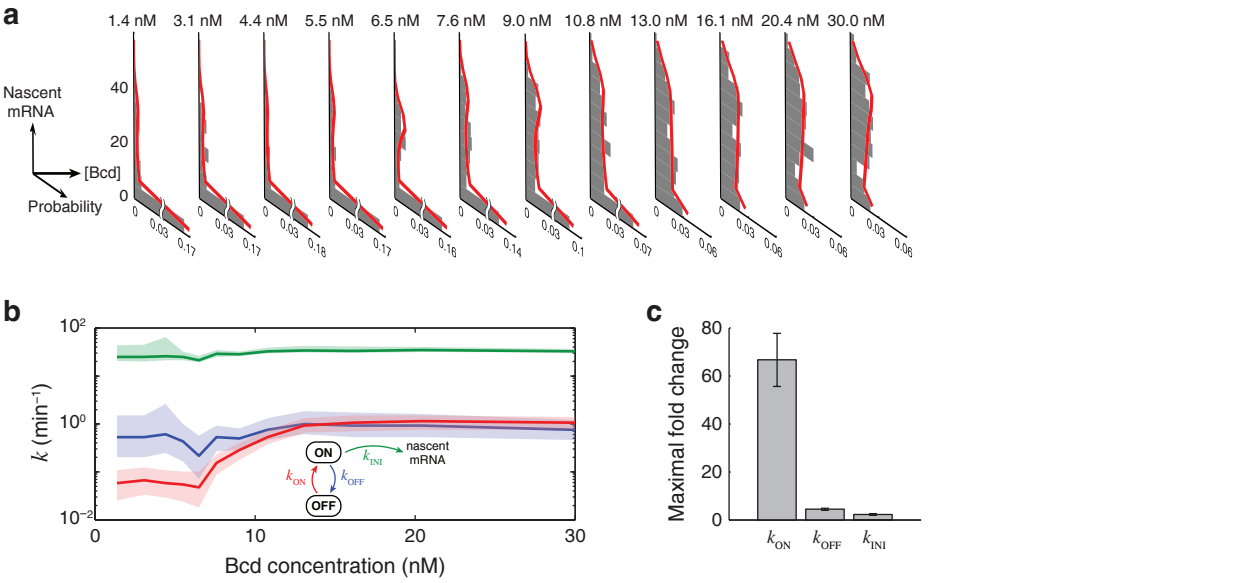
Supplementary Figure 8. Fluctuations in the number of nascent mRNAs at individual *hb* gene loci.



(**a**) The coefficient of variation (defined as the ratio of the standard deviation, σ, to the mean, μ) for the number of nascent mRNAs at individual *hb* gene loci, as a function of nuclear Bcd concentration (data from >600 nuclei, AP positions 0.25-0.7 EL, single wild-type embryo at cycle 12). During cycles 11-13, the coefficient of variation for the number of nascent *hb* mRNAs per locus, at a given Bcd concentration, was in the range ~50-300% (values from 23 individual embryos). (**b**) A comparison of the lower bound on the coefficient of variation between different cycles. Data averaged over multiple wild-type embryos (cycle 11, 7 embryos; cycle 12, 9 embryos; cycle 13, 7 embryos; cycle 14, 7 embryos). Error bars represent S.E.M. The lower bound on *hb* variability was in good agreement with previous studies (He et al., 2011; Little et al., 2013); it is also significantly higher than our estimated error in nascent-mRNA quantification

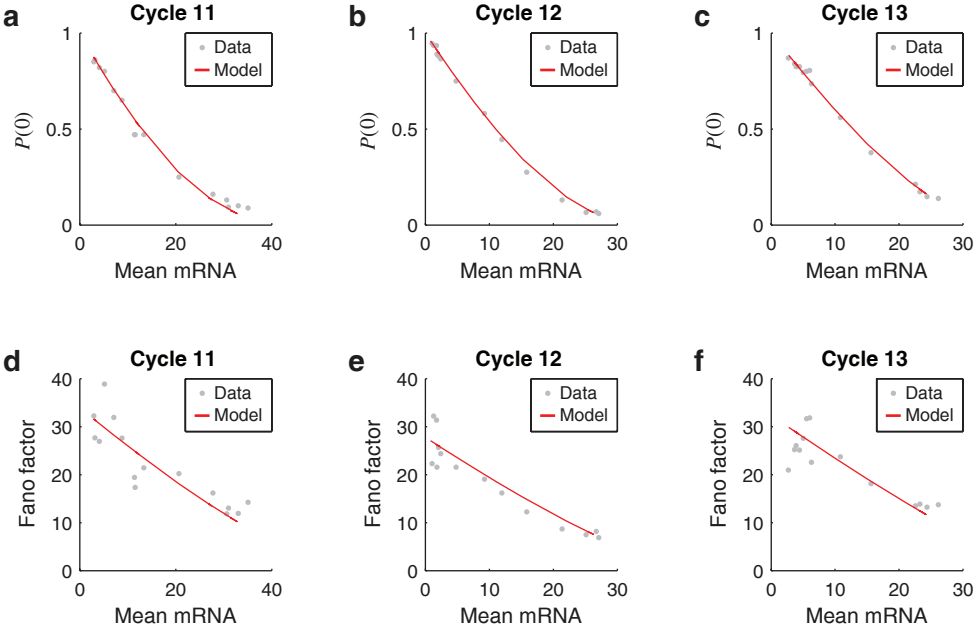
(<10%, see **Supplementary Note**). (**c**) The correlation coefficient of number of nascent mRNAs between different *hb* gene loci in the same nucleus, as a function of nuclear Bcd concentration (data from >600 nuclei, AP positions 0.25-0.7 EL, single wild-type embryo at cycle 12). A threshold Bcd concentration value (dashed line) was set to distinguish regions of high and low *hb* expression. (**d**) A comparison of the correlation coefficient between different cycles in the low Bcd concentration range (data from wild-type embryos; cycle 11, 7 embryos; cycle 12, 9 embryos; cycle 13, 7 embryos; cycle 14, 7 embryos). Error bars represent S.E.M. Positive correlation was detected in the activity of the two *hb* gene copies within a given nucleus. (**e**) A comparison of the correlation coefficient between different cycles in the high Bcd concentration range (data from wild-type embryos; cycle 11, 7 embryos; cycle 12, 9 embryos; cycle 13, 7 embryos; cycle 14, 7 embryos). Error bars represent S.E.M. Unlike panel **d**, no correlation was detected here, consistent with a negligible contribution by extrinsic noise (Little et al., 2013). (**f**) The Fano factor (defined as the ratio of the variance, σ^2, to the mean, μ (Golding et al., 2005; Ozbudak et al., 2002)) for the number of nascent mRNAs at individual *hb* gene loci, as a function of nuclear Bcd concentration (data from >600 nuclei, AP positions 0.25-0.7 EL, a single wild-type embryo at cycle 12). For comparison, the Fano factor is also plotted (blue) for the case of transcription events that are initiated with constant probability over time (a Poisson process), where the copy-number distribution of nascent mRNAs per locus is expected to be ~1 (the deviation from 1 is due to the deterministic effect of elongation, (Choubey et al., 2013; Zenklusen et al., 2008) and see **Supplementary Note**). (**g**) A comparison of the lower bound on the Fano factor between different cycles (data from wild-type embryos; cycle 11, 7 embryos; cycle 12, 9 embryos; cycle 13, 7 embryos; cycle 14, 7 embryos). Error bars represent S.E.M. During cycles 11-13, the Fano factor is consistently higher than 5 and can be as high as ~30 at low Bcd concentrations (value from 23 individual embryos). These values suggested that, rather than occurring regularly in time, *hb* transcription initiation is intermittent (bursty), consisting of periods of intense activity separated by periods of quiescence (Golding et al., 2005; Raj et al., 2006; Sanchez and Golding, 2013). In contrast, embryos in cycle 14 do reach Fano ≈ 1 at high Bcd concentrations (value from 7 individual embryos), consistent with the presence of different regulatory mechanisms (and resulting stochastic kinetics) at this later stage.

Supplementary Figure 9. Fitting the copy-number statistics of nascent *hb* mRNA using a two-state model.



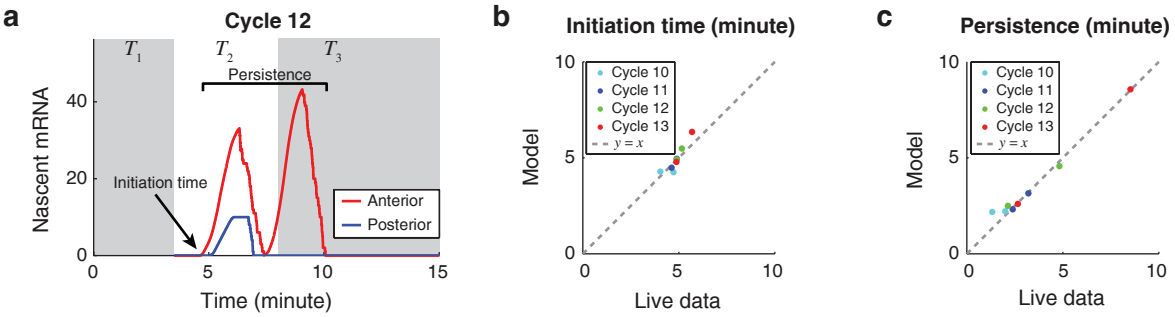
(**a**) Fitting the histograms of number of nascent *hb* mRNAs (gray) using a two-state model (red). The single-locus data from a single embryo was binned into multiple Bcd concentration windows, and the fitting of each window was performed independently of others using the maximum-likelihood method (data from the same embryo used in **Figure 2**, >1300 loci, AP positions 0.25-0.7 EL, cycle 12). A good agreement between model and experiment was achieved over the full range of Bcd concentrations (R^2 = 0.88± 0.01, 23 embryos). (**b**) The extracted parameters of the two-state model as a function of nuclear Bcd concentration (data from the same embryo used in Panel **a**). Red, k_{ON}. Blue, k_{OFF}. Green, k_{INI}. Shaded area, Parameter ranges corresponding to >0.7% of the maximum likelihood. (**c**) A comparison of the maximal fold change in different kinetic parameters when varying the Bcd concentration (data from wild-type embryos; cycle 11, 7 embryos; cycle 12, 9 embryos; cycle 13, 7 embryos). Error bars represent S.E.M. k_{ON} changed much more dramatically than k_{OFF} and k_{INI}.

Supplementary Figure 10. The two-state model reproduces the observed nascent *hb* mRNA statistics.



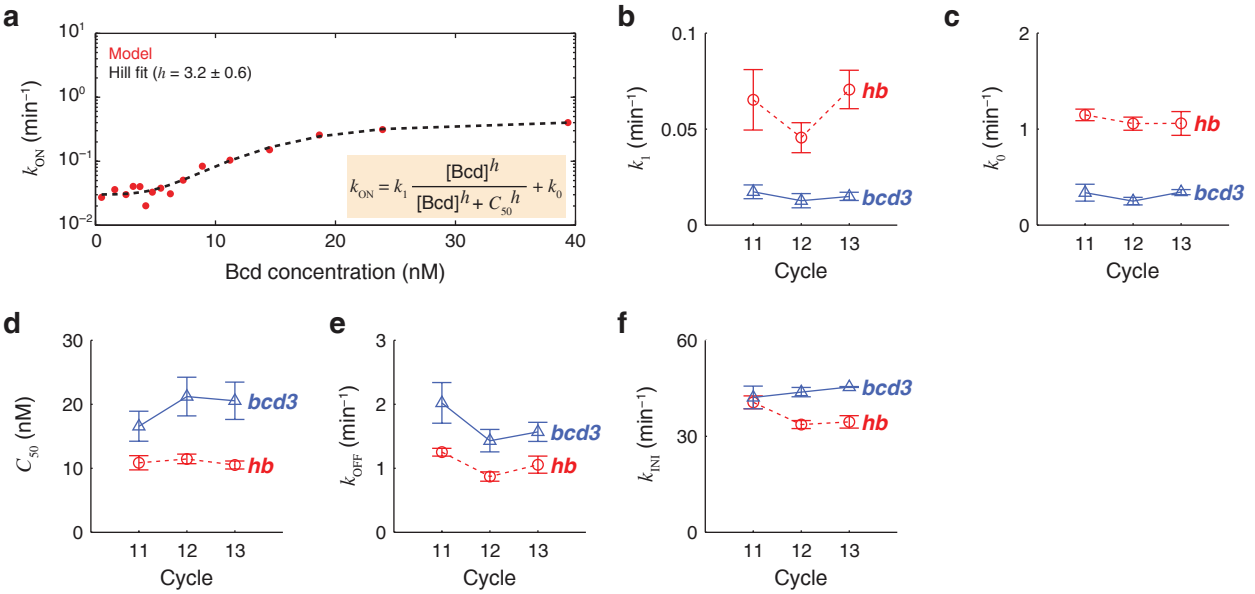
(**a**) – (**c**) The fraction of inactive *hb* loci, $P(0)$, versus the mean number of nascent *hb* mRNAs, for embryos at different cycles. Gray, experimental data. Red, results of the two-state model. (**d**) – (**f**) The Fano factor versus the mean number of nascent *hb* mRNAs, for embryos at different cycles. Gray, experimental data. Red, results of the two-state model.

Supplementary Figure 11. The two-state model reproduces live-embryo data.



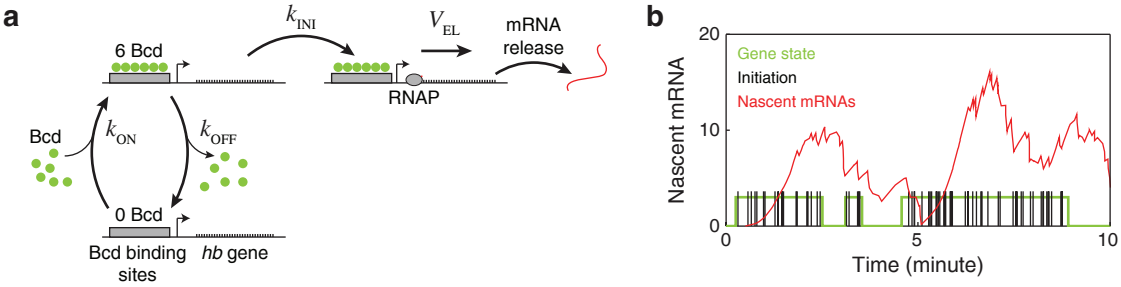
(**a**) Simulated nascent mRNA signal at an anterior (red) and a posterior (blue) gene loci during cycle 12. The entire nuclear cycle was modeled as consisting of three time periods. Transcription can only be initiated during period T_2. The initiation time and the persistence of the transcriptional activity are marked for the anterior locus. (**b**) – (**c**) Comparison of the initiation time (**b**) and the persistence (**c**) between the live-imaging data (Lucas et al., 2013) and our simulation. For each nuclear cycle, the mean values for the anterior side and the posterior side of the embryo are plotted (dots). Dashed line indicates perfect match between experiment and simulation.

Supplementary Figure 12. Kinetic parameters of the two-state model.



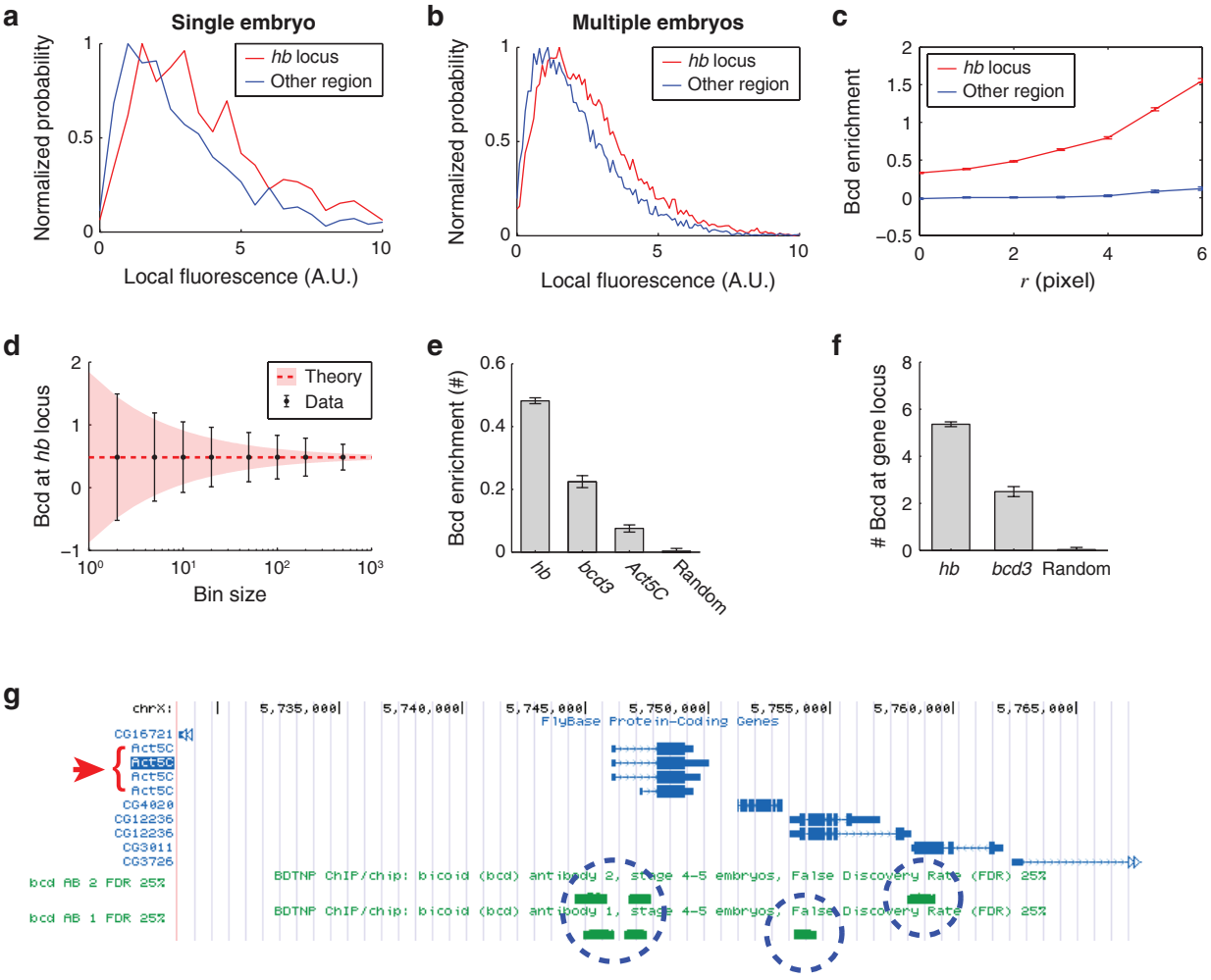
(**a**) The rate of stochastic gene activation, k_{ON}, for the *bcd3-lacZ* reporter gene is plotted against nuclear Bcd concentration, and fitted to a Hill function (dashed line). Fitting was done using data from >1700 loci, AP positions 0.15-0.7 EL, in a single *bcd3-lacZ* embryo at cycle 12. (**b**) – (**f**) Kinetic parameters extracted from fitting the nascent mRNA statistics to the two-state model. Red circles, *hb* gene in wild-type embryos (cycle 11, 7 embryos; cycle 12, 9 embryos; cycle 13, 7 embryos). Blue triangles, *bcd3-lacZ* in transgenic flies (cycle 11, 6 embryos; cycle 12, 7 embryos; cycle 13, 7 embryos). Error bars represent S.E.M.

Supplementary Figure 13. Bcd modulates the probability of *hb* activation through cooperative binding.



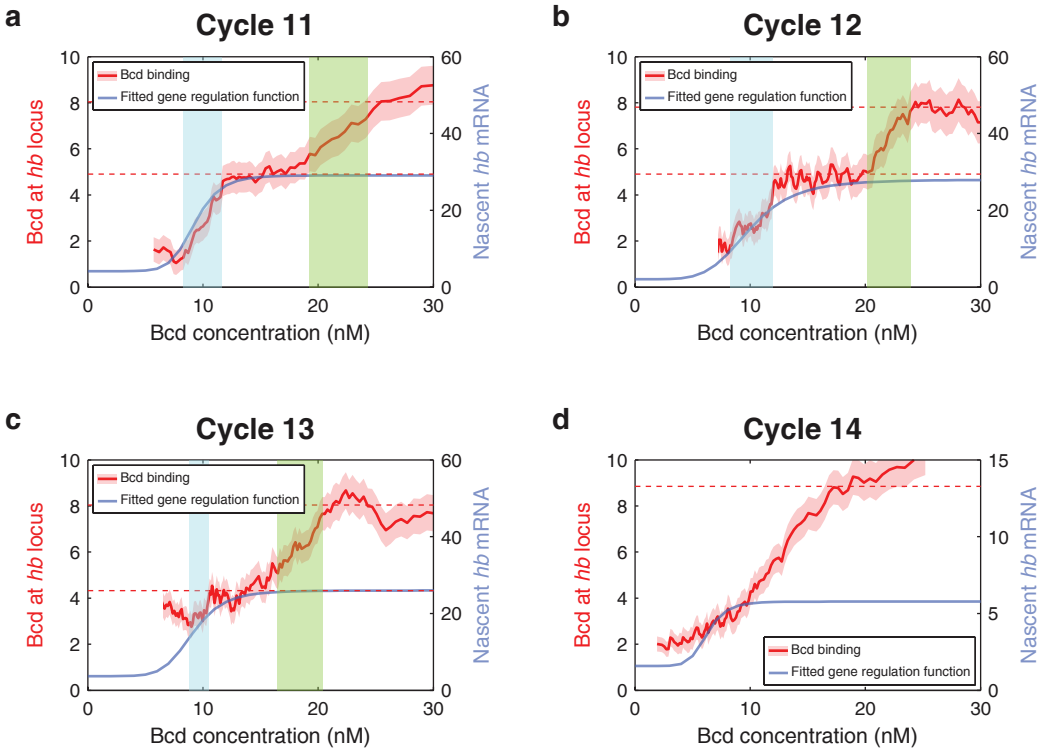
(**a**) A two-state model of Bcd regulation. Bcd molecules stochastically bind and unbind to the *hb* gene. Cooperative binding of 6 Bcd molecules switches the *hb* gene to the active state, whereas the unbinding of Bcd molecules switches the gene to the inactive state. In the active state, *hb* transcription is stochastically initiated at rate k_{INI}. Transcript elongation occurs deterministically with speed V_{EL}. Once elongation is completed, the transcript is released from the gene locus. (**b**) A simulation of *hb* transcription using the above model. Plotted are time traces of *hb* gene state (green), transcription initiation (black), and nascent mRNA (red).

Supplementary Figure 14. Quantifying Bcd binding at the *hb* locus.



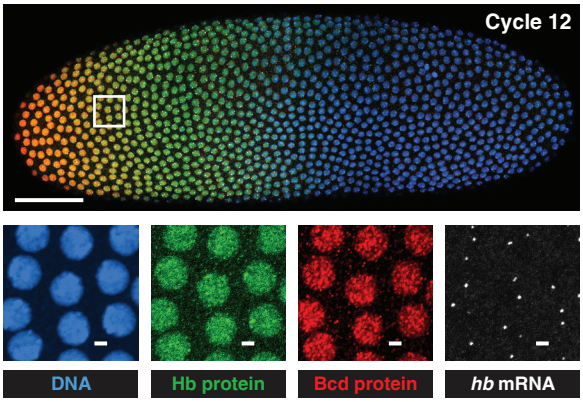
(**a**) – (**b**) Comparing the distribution of local Bcd signal between the *hb* locus and other nuclear regions. Panel **a** is for a single embryo (data from >1200 *hb* loci, wild-type embryo at cycle 12). Panel **b** is for multiple embryos (data from >20,000 *hb* loci, 30 wild-type embryos, cycles 11-14). In both cases, the distribution of the Bcd signal measured at the *hb* locus was distinguishable (and had a higher mean) from that of the signal at other nuclear regions. (**c**) The average Bcd enrichment signal versus the integration radius r. Data from >20,000 *hb* loci, 30 wild-type embryos, cycles 11-14. (**d**) The fluctuations of the binned Bcd enrichment signal as a function of bin size. The mean and standard deviation of the binned data (black, data from >20,000 *hb* loci, 30 wild-type embryos, cycles 11-14) are compared with a model assuming random distribution of Bcd molecules inside the nucleus (red). (**e**) The average Bcd enrichment signal at different gene loci. Values are plotted without rescaling. Each data point was obtained from >5000 nuclei in >7 embryos. Error bars represent S.E.M. (**f**) The average number of bound Bcd molecules at different gene loci. All values were scaled to the value for *bcd3-lacZ*. Each data point was obtained from >5000 nuclei in >7 embryos. Error bars represent S.E.M. (**g**) The ChIP-chip signal in the vicinity of the *Act5C* gene (red arrow). Data is from the Berkeley *Drosophila* Transcription Network Project (Li et al., 2008), and is plotted using the UCSC Genome Browser (www.genome.ucsc.edu). ChIP peaks from two different Bcd antibodies (green) were used to estimate the total signal and the error in the data.

Supplementary Figure 15. Bcd binding and *hb* transcription at different cycles.



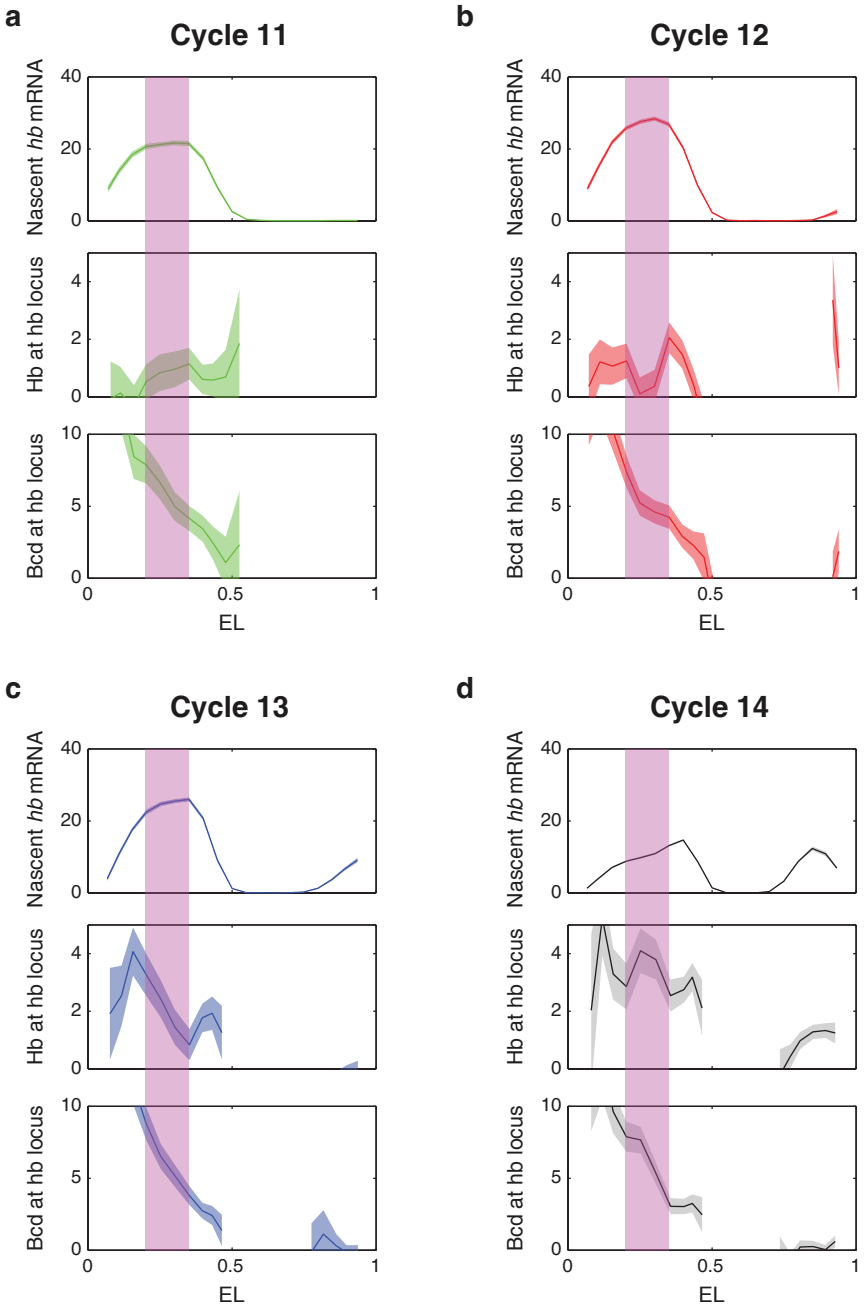
(**a**) – (**d**) Bcd binding at the *hb* locus (solid red line, mean value; red shading, S.E.M.) and *hb* transcription (blue line, Hill fit of the gene regulation function) as a function of nuclear Bcd concentration, for different cycles (cycle 11, 7 embryos; cycle 12, 9 embryos; cycle 13, 7 embryos; cycle 14, 7 embryos, AP range 0.25-0.7 EL). Dashed red lines highlight binding plateaus. During cycles 11-13, the binding curve exhibits two sharp steps (cyan and green shading). The first one (cyan) is accompanied by activation of *hb*, while the second (green) is not.

Supplementary Figure 16. Simultaneous labeling of Bcd protein, Hb protein and *hb* mRNA.



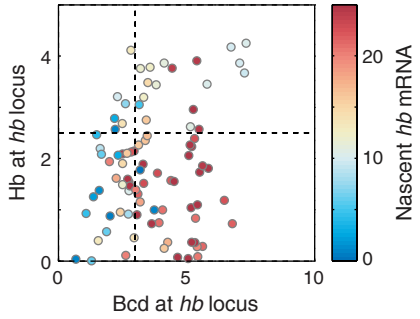
Confocal image of a wild-type embryo during cycle 12. Bcd protein (red) and Hb protein (green) were labeled using immunofluorescence, *hb* mRNA (white) was labeled using smFISH, and DNA (blue) was stained with Hoechst. Scale bar, 50 μm. Bottom: A magnified view of individual channels from a small anterior region (corresponding to the white box in the main panel). Scale bar, 2 μm.

Supplementary Figure 17. Quantifying *hb* transcription, Hb binding and Bcd binding.



(**a**) – (**d**) *hb* transcription, Hb binding, and Bcd binding at different AP positions for different cycles (cycle 11, 6 embryos; cycle 12, 3 embryos; cycle 13, 5 embryos; cycle 14, 4 embryos). Lines, mean values. Shading, S.E.M. Purple shading indicates the range 0.20-0.35 EL, the area immediately adjacent to the future *hb* middle stripe (Perry et al., 2012). Within that AP range, Hb binding increases while *hb* expression decreases between earlier cycles (11-13) and cycle 14.

Supplementary Figure 18. Relating *hb* transcription to Bcd and Hb binding.



hb transcription as a function of both Bcd and Hb binding at the *hb* locus. 18 embryos between cycles 11 and 14 were analyzed. The numbers of nascent *hb* mRNAs and bound Bcd and Hb proteins were measured for each locus in AP positions 0.25-0.7 EL (>7600 loci). The data for each cycle was binned by AP position, and plotted on a 2D color plot (x and y axes, Bcd and Hb binding; color, nascent *hb* mRNAs).

SUPPLEMENTARY TABLE

Supplementary Table 1. Sequences of smFISH probes

Transcript	Probe sequences (5' to 3')
hb	TTGTGCTGCTCGTAGTTGGT
	GAACATGCTGTTGTACCAGG
	GCTCCTGTTTGATATTTGCC
	TATTCCCGTCGAGATGATGA
	AACTGTTCCAGGTGATTGGT
	ATCCATGGGTTGCTGCTGAA
	TTTGATCGTTTTGGCTGGGT
	TTAGCATCGTAATGCTGCAG
	TTGCTGCAGCAACTGTTGCT
	TGGAAATGCTGCTGGTACTG
	ATGGTGATGATGTTGCTGCT
	TTGAATCCACCCATCAGATG
	TAGAAGTGCTGCATGGGATT
	TGTTAGTGCCTGCAACTTCT
	TATTCGACTGACTCGACTTG
	ATGTACTTCATGTCCTCGCT
	ATGTTGGTATCATCGTCCTC
	TGCGAATTGTAGATGGGCAT
	GGTCTTGCACTTGTAGTTCT
	TTGTCTGGTTTCATGTGGGT
	TACTCCAAGTGGTGCTTGAA
	GTTCTTGTGCTTCCGGATAT
	ACACGTGTAGCTGCATTTGT
	GCGAGTTTAGCATGGATTTG
	TACACAGAACTGTGCGACTT
	TAATCACAATCCGCACAACG
	AAGCTGTGGCAATACTTGGT
	ATACTTGCGCAGATGCAGCT
	AAACATCGATGACCAACGAG
	ATTCTTGCTCTTCGGACCAC
	AGCTGCAACATTTGACTTCC
	TGGCTGAGATTGCTGTTGCT
	TTGAACCAGAGGGAATCCTT
	AAGAAGGCCATGTTGCGGTT
	TGGAGATTGAGGTTCCAGTA
	TCGCATTCTTGGCGACAATT
	TGGTTCTGTTGCTGCAGTTG
	TGACTTACGCTCGTACTCAT
	TTCCTTGGGACAGATCCATG
	TTGTTGCTGCTGCTCATCCT
	TCCTCCACCTTGAGATTCAT

SUPPLEMENTARY TABLE

	TGTGCTGGGTACTTTCAGTT TTGCTATTGCTGCTGGCATT TTCCATTGCTGCTGGAATTG AGTACTTGCACTCGTAGATG GCGTCCTTGAAGAAGATATC CATGTGAATGGTGTAGAGCA TTGCACTTGAACACATCGTC
lacZ	GTGAATCCGTAATCATGGTC TCACGACGTTGTAAAACGAC ATTAAGTTGGGTAACGCCAG TATTACGCCAGCTGGCGAAA ATTCAGGCTGCGCAACTGTT AAACCAGGCAAAGCGCCATT AGTATCGGCCTCAGGAAGAT AACCGTGCATCTGCCAGTTT TAGGTCACGTTGGTGTAGAT AATGTGAGCGAGTAACAACC GTAGCCAGCTTTCATCAACA AATAATTCGCGTCTGGCCTT AGATGAAACGCCGAGTTAAC AATTCAGACGGCAAACGACT TTTCTCCGGCGCGTAAAAAT ATCTTCCAGATAACTGCCGT AACGAGACGTCACGGAAAAT GCTGATTTGTGTAGTCGGTT TTAAAGCGAGTGGCAACATG AACTGTTACCCGTAGGTAGT ATAATTTCACCGCCGAAAGG TTTCGACGTTCAGACGTAGT ATAGAGATTCGGGATTTCGG TTCTGCTTCAATCAGCGTGC ACCATTTTCAATCCGCACCT TTAACGCCTCGAATCAGCAA ATGCAGAGGATGATGCTCGT TCTGCTCATCCATGACCTGA TTCATCAGCAGGATATCCTG CACGGCGTTAAAGTTGTTCT TGGTTCGGATAATGCGAACA TTCATCCACCACATACAGGC TGCCGTGGGTTTCAATATTG ATCGGTCAGACGATTCATTG TGATCACACTCGGGTGATTA ATACAGCGCGTCGTGATTAG GATCGACAGATTTGATCCAG

SUPPLEMENTARY TABLE

	AAATAATATCGGTGGCCGTG TTTGATGGACCATTTCGGCA TATTCGCAAAGGATCAGCGG AAGACTGTTACCCATCGCGT TGCCAGTATTTAGCGAAACC AAACGGGGATACTGACGAAA TAATCAGCGACTGATCCACC GGGTTGCCGTTTTCATCATA TCGGCGTATCGCCAAAATCA TTCATACAGAACTGGCGATC TGGTGTTTTGCTTCCGTCAG ACGGAACTGGAAAAACTGCT TATTCGCTGGTCACTTCGAT GTTATCGCTATGACGGAACA TTTACCTTGTGGAGCGACAT GTTCAGGCAGTTCAATCAAC TTGCACTACGCGTACTGTGA AGCGTCACACTGAGGTTTTC ATTTCGCTGGTGGTCAGATG ACCCAGCTCGATGCAAAAAT CGGTTAAATTGCCAACGCTT CTGTGAAAGAAAGCCTGACT GGCGTCAGCAGTTGTTTTTT TACGCCAATGTCGTTATCCA TAAGGTTTTCCCCTGATGCT ATCAATCCGGTAGGTTTTCC GTAATCGCCATTTGACCACT AGTTTTCTTGCGGCCCTAAT ATGTCTGACAATGGCAGATC ATAATTCAATTCGCGCGTCC TGATGTTGAACTGGAAGTCG TCAGTTGCTGTTGACTGTAG ATTCAGCCATGTGCCTTCTT AATCCCCATATGGAAACCGT AGACCAACTGGTAATGGTAG
Act5C	GGAAGGGGAAAGCTACTTAC GTTCGATGGTGTTTTGTTGT TTGCTTTTTCTTGGTGGGAG ACATTTTCCGCGGCTTTTCT CTTTCAGCACTGGCACATAA CTTTAGCTCGTGGCTAGTAT GGGTATTTTTGGAGTTTTGG TTGACAATTTAGGGCCAAAC GATGTCTTTCGACCTATTGG

SUPPLEMENTARY TABLE

	GGTATTTGAAACCCTGGTTT GCAGTTGTAATCCTGATTCT GGGCTATTGAATTGCTTTGG CGAGCTTGGCTTTAAACACT TTTGGGTGAAATCGTGACAC GTGTTGGCGAATTGATTTGC TAGGGCGTGGGAAAAAAAGA CATTTATGGATGATCTGGTG CGTCACTATTTTCAAGAACC TTTCTGCTTCAGCCAAAACG CAACTTTGAAAAACCGGCAA GCTCGAAGGTTTTCTCCATT GAGTGTGTGTATATGAACGA TCTCTCTCGTTATTTTGCGC GCTCTCTCACACACACATAC TTTCGAGGCGAAAAGCAAAC GGCCAAGTTTTGTGACCATA AGCGCTGGTGTGTTGAATTG TTTTCAATCGCTGCCTGCGT GTTGAATGCCAGTTAACTGC TAAACCGACTGAAAGTGGCT GGTTTGAAAGGAATGACTGG ACTACTGTAAACGCAAGTGG GCCACGAAACTTTTCAAAGC CACTAAACGCACTCTAGAAA CTCGAATTGGAACCCGAAAT TTAAAATGGCGGCTGCCTAA ACAGTTGCCTGTGTGTATGC CAAAAGAAAGCCGCAAAGAG AGACGACACTTAACGAATGC ACGAATATGCCGAAAAGGGA CCCCGAAAAATTAAAACCAC AGTTACAAAATAGAAGGCCC CTGATTTTCACTTGCGATCG CACTCAATGCAATGTACGTT CTAACTTGTTTTTCGTGTAC GCTGGTCTACCAAAGTACAA ATTTGCGTGGTTTCCTTGGA GAATGCGGATTTCCCATTAG
nullo	ACTGTTGTTGGACAACTGAT TTTGGAGAACGTTGTTAGCG GGTATAGTCACAAGTTCGTA GCTGCCCATTTTGAGATTTC TTACCTTCTCAGCGGAATGT

SUPPLEMENTARY TABLE

	AGAAGATGCCAGGATTTGCA TTATAAAACCGGGCAGGGGA AGGACGAGTCTCTGGATGTT TTTGTTTTCTGGCCGAGATG ATGCTTGGAGCGCTTGTTTA ATGTTGGATGCGGACTTCTT TCGAAACTGCACTTCCATTG ATGGAGACGCGGCTGCTTTT ACACCTCTTGGGATCCGTGT AACTGGAGTCCATAACCATC ATGTCGTAGAGCGGTGACTC ACCGTAGTAATCCTCCAGAT TAAACCTTCACCAATCGCTC GCTATACAAATGGCTCTATG GGCTTACGAGCTATCACAAT

SUPPLEMENTARY NOTE

1 Image acquisition and processing

1.1 Correction for optical aberrations

Image processing and data analysis were performed using custom MATLAB scripts (**Supplementary Software**). First, confocal data stored in Zeiss *.lsm* files was converted to *.tif* format, and the intensity profile of each channel was corrected for optical aberrations as follows. Monochromatic aberrations (such as spherical aberration) cause uneven fluorescence detection efficiency at different regions of the confocal field of view (Dunn and Wang, 2000; Pawley, 2006). To compensate for this effect, we imaged a sample of uniformly mixed fluorophores, and then used polynomial regression (MATLAB function "polyfitn") to fit the resulting intensity profile (**Fig. SN1a**) to a 2D quadratic function,

$$I(x,y) = Ax^2 + By^2 + Cxy + Dx + Ey + F \quad (1)$$

The relative detection efficiency, r, at a position (x, y) was calculated as

$$r(x,y) = \frac{I(x,y)}{I_{max}} \quad (2)$$

and then used to rescale the intensity values of each pixel in the embryo image stacks.

To correct for the spatial mismatch between different channels, known as chromatic aberration (Dunn and Wang, 2000; Pawley, 2006), we labeled a set of smFISH probes with two different colors and applied them in equal molarity to the same sample. By comparing smFISH spots from both channels (see **Supplementary Note Section 2.1** for details of smFISH spot recognition), we identified inter-channel pairs corresponding to the same spot (**Fig. SN1b**). The inter-channel spatial mismatch of these spots indicates the amount of chromatic aberration in different parts of the field of view. For the 63× oil immersion objective used in this study, the mismatch in the z dimension was negligible (much smaller than 1 voxel, **Fig. SN1c**), while the mismatch in the xy plane was linearly related to the xy coordinates (**Fig. SN1d**), such that,

$$\begin{pmatrix}\Delta x_{21}\\ \Delta y_{21}\end{pmatrix} = \begin{pmatrix}x_2 - x_1\\ y_2 - y_1\end{pmatrix} = \begin{pmatrix}a & b\\ c & d\end{pmatrix}\begin{pmatrix}x_1\\ y_1\end{pmatrix} \quad (3)$$

where (x_1, y_1) and (x_2, y_2) denote the xy coordinates of channels 1 and 2, respectively. In later analysis of smFISH/immunofluorescence (IF) images, every smFISH spot position was shifted according to Equation (3) to match the IF channel.

1.2 Nuclear segmentation and nuclear cycle determination

In each embryo, we reconstructed the 3D shape of individual nuclei using the Hoechst signal (**Fig. SN2a**). For each z slice, the reconstruction started by smoothing the Hoechst image using a Gaussian filter (resulting in image I_G). This was followed by a difference-of-Gaussians filter to enhance contrast (resulting in image I_D). A watershed algorithm was then applied to image I_D to identify nuclei candidates (Boettiger and Levine, 2013). For each candidate, we determined its accurate shape by thresholding the smoothed image I_G. To find the optimal threshold, we tested a series of values (from 0 to 65535 with a step size of 256) by calculating the nuclear circularity $4\pi A/P^2$ (where A is the area and P is the perimeter length, (McHale et al., 2011)). Because syncytial blastoderm embryos typically have circular nuclei, we chose the lowest threshold value that gave circularity >0.7. For a group of nuclei that were too close to each other, and therefore could not be separated by pure thresholding, we applied an additional watershed step to separate the merged nuclei. Results from individual z slices were stacked together to form a 3D mask. Nuclear segments from neighboring z slices were stitched together as one nucleus if the

centroid *xy* position of one was within the area of the other. Further refinements and corrections were done manually through a custom MATLAB GUI.

To establish the anterior-posterior (AP) axis of the embryo, we first thresholded the maximum intensity projection of the Hoechst image to identify the embryo boundary. A pair of boundary points with the longest mutual distance were then defined as the anterior and posterior poles, respectively, based on their Bcd intensities (the anterior pole has higher Bcd intensity than the posterior pole). The AP position of each nucleus was then calculated as its distance (normalized by the length of AP axis) to the anterior pole along the AP axis.

The nuclear cleavage cycle of each embryo was first estimated using the formula:

$$\text{Cycle number} = \text{ceil}(\log_2 N_{\text{nuclei}}) + 2 \quad (4)$$

where N_{nuclei} is the number of nuclei recognized from the image stack. This equation takes into account the fact that our imaging only covers the surface layer of one side (left or right from the midsagittal plane) of the embryo. In some cases, the automatic estimation was corrected (±1) following visual inspection of the embryo. To test the accuracy of our cycle estimation, we measured the nuclear-to-cytoplasmic area ratio of each imaged wild-type embryo, and plotted it against the estimated cycle number. The measured ratios were found to increase monotonically with nuclear cycles (**Fig. SN2b**), as expected (Edgar et al., 1986).

2 mRNA quantification

2.1 Identification and quantification of individual mRNAs

To identify smFISH spots and quantify their intensities, we followed the approach of (Little et al., 2011; Skinner et al., 2013; Zenklusen et al., 2008). For each *z* slice, the smFISH image was first smoothed using a 2D Gaussian filter with a radius of 3 pixels. Candidate mRNA particles were identified as 3D local maxima in the filtered image stacks, whose *xy* positions also appeared in two consecutive *z* slices as 2D local maxima (allowing a ±1 pixel shift in *xy* dimensions).

Intensity values from the candidate positions exhibited a multi-modal distribution (**Supplementary Fig. 3a**). To determine which signals actually correspond to target mRNA, we compared the intensity histogram with one from a negative sample (for *lacZ* mRNA, OreR embryos; for *hb* mRNA, the posterior part of OreR embryos). The right peak of the histogram (marked by a black arrow in **Supplementary Fig. 3a**) was only present in the positive sample, and was therefore identified as target mRNA. In contrast, the left two peaks were present in both positive and negative samples, and were therefore identified as false positives. A threshold ("single mRNA" threshold) was set to filter out these false-positive spots.

For each mRNA candidate, we used a least-square algorithm (MATLAB function "lsqcurvefit") to fit its local intensity profile (7x7 pixels centered on the local peak) to a 2D elliptical Gaussian function (Little et al., 2011; Skinner et al., 2013; Zenklusen et al., 2008) (**Supplementary Fig. 3b**),

$$f(x,y) = I_p e^{-(a(x-x_0)^2+2b(x-x_0)(y-y_0)+c(y-y_0)^2)} + I_b \quad (5)$$

with

$$a = \frac{\cos^2\theta}{2{\sigma_1}^2} + \frac{\sin^2\theta}{2{\sigma_2}^2}$$
$$b = -\frac{\sin 2\theta}{4{\sigma_1}^2} + \frac{\sin 2\theta}{4{\sigma_2}^2} \quad (6)$$
$$c = \frac{\sin^2\theta}{2{\sigma_1}^2} + \frac{\cos^2\theta}{2{\sigma_2}^2}$$

where I_p is the intensity amplitude of the particle, I_b is the background level, (x_0, y_0) are the center coordinates of the particle, (σ_1, σ_2) are the half-widths of the Gaussian in major and minor axes, and θ is the angle between the major axis of the Gaussian and the x axis of the image. For candidates that were close to each other in the xy plane (center-to-center distance <15 pixels in both dimensions), we fitted their superposed intensity profiles to a multiple-Gaussian function in order to decompose signals from individual candidates. Fitting results with σ_1 or σ_2 <0.4 pixel were discarded to further eliminate false positives. The total fluorescence intensity of an mRNA spot was defined as the background-subtracted integral of the Gaussian profile,

$$I = 2\pi I_p \sigma_1 \sigma_2 \quad (7)$$

To extract the typical intensity, I_0, of individual mRNA molecules, we used a least-square algorithm (MATLAB function "fit") to fit the intensity distribution of identified smFISH spots to a sum of Gaussian functions (Skinner et al., 2013; Zenklusen et al., 2008) (**Supplementary Fig. 3c**),

$$p(I) = A_1 e^{-\frac{(I-I_0)^2}{2{w_0}^2}} + A_2 e^{-\frac{(I-2I_0)^2}{4{w_0}^2}} + A_3 e^{-\frac{(I-3I_0)^2}{6{w_0}^2}} \quad (8)$$

where w_0 is the half-width. The second and the third terms in the expression describe the contribution from closely located molecules.

2.2 Identification and quantification of transcription sites

The identification of active transcription sites among the population of mRNA spots was done in two steps: (1) Before fitting the local intensity profile, candidates corresponding to active transcription sites were pre-selected from the mRNA spot candidates. To do this, we tested a series of intensity threshold values, I_i (from 1000 to 8000 intensity units, with an interval of 100 intensity units), by counting the number of detected candidates (N) using a given threshold and calculating the rate of decrease in number of candidates, $r_d(I_i) = N(I_i) / N(I_{i+1}) - 1$, at each threshold value (**Supplementary Fig. 3d**). The intensity value corresponding to the minimum in r_d (or <8%/100 intensity units, whichever occurred first) was used as the "nascent mRNA" threshold for pre-selection. (2) After fitting the local intensity profile of each pre-selected candidate, an additional threshold of $3I_0$ was used to define sites of active transcription (This value is similar to the one in (Little et al., 2013). The possible errors resulting from the specific choice of threshold are discussed in **Supplementary Note Section 8** below). Finally, we estimated the number of nascent transcripts at each transcription site by dividing the intensity of the transcription site by I_0, and rounding the value up or down to the nearest integer. This followed the approach of (Femino et al., 1998; Little et al., 2013; Zenklusen et al., 2008).

3 Protein quantification

SUPPLEMENTARY NOTE

We calculated the average IF intensity of each nucleus from the central z slice of the nucleus, defined as the slice with the strongest Hoechst signal. Before averaging, the nuclear mask of the slice was eroded by 10 pixels to eliminate boundary effects. Background fluorescence was estimated from the average intensity of posterior nuclei (AP positions 0.8-0.9 EL), and subtracted from the IF intensity profile (This follows the approach of (Gregor et al., 2007a)).

For embryos in cycles 11-14, Bcd immunofluorescence exhibited an exponential gradient along the AP axis, and the exponential decay length was in good agreement with previous measurements (Abu-Arish et al., 2010; Gregor et al., 2007b; Houchmandzadeh et al., 2002; Little et al., 2011) (**Fig. SN3**). In the anterior part of the embryo, nuclei exhibited strong fluorescence, while cytoplasmic regions were characterized by low signal with sparsely distributed fluorescent particles (**Fig. 1c**, middle panel), as expected from the nuclear localization of Bcd during interphase (Gregor et al., 2007b).

3.1 Quantifying nuclear Bcd concentration using IF spot intensity

We recognized IF spots in the cytoplasmic region of the embryo, where Bcd concentration is typically much lower than in the nucleus (Driever and Nusslein-Volhard, 1988; Gregor et al., 2007b). Spot identification and quantification followed the same procedure as for smFISH spots. Briefly, spot candidates were recognized as 3D local maxima, whose xy locations also appeared in consecutive z slices as 2D maxima. Each candidate was then fitted to a 2D Gaussian function to extract its intensity (**Supplementary Fig. 5b**).

To identify spots corresponding to individual Bcd molecules, we compared the intensity histograms from the anterior and posterior sides of the embryo (**Supplementary Fig. 5c**). Since most Bcd molecules are in the anterior part (Driever and Nusslein-Volhard, 1988; Gregor et al., 2007b), and are believed to remain as monomers before binding to DNA (Burz et al., 1998; Ma et al., 1996), the extra peaks seen in the anterior distribution were assumed to correspond to individual Bcd molecules, as well as multiple adjacent Bcd molecules. In contrast, the lower-intensity peak seen in the posterior distribution was assumed to correspond to nonspecific bindings of the primary antibody. Following the same procedure as for smFISH spots, we used a least-square algorithm (MATLAB function “fit”) to fit both the anterior and the posterior distributions to a multi-Gaussian function,

$$p(I) = B_0 e^{-\frac{(I-I_0)^2}{2{w_0}^2}} + B_1 e^{-\frac{(I-I_1)^2}{2{w_1}^2}} + B_2 e^{-\frac{(I-2I_1)^2}{4{w_1}^2}} + B_3 e^{-\frac{(I-3I_1)^2}{6{w_1}^2}} \quad (9)$$

where I_0 is the typical intensity of spots resulting from nonspecific antibody binding, and I_1 is the typical intensity of a single antibody-labeled Bcd molecule. In practice, since the nonspecific binding term of the anterior distribution typically has a very small amplitude and overlaps heavily with the Bcd terms (and vice versa for the Bcd terms in the posterior distribution), direct extraction of both I_0 and I_1 from a single distribution (anterior or posterior) was inaccurate. To precisely measure I_0 and I_1, we first fitted the posterior distribution to extract the value of I_0, and then applied this value to the fitting of the anterior distribution, to extract I_1. We note that I_1 was equal to approximately $2I_0$ (**Supplementary Fig. 5d**). This is consistent with the interpretation that I_0 represents nonspecific binding of individual primary antibodies, while the specific labeling of a Bcd molecule involves two primary antibodies on average. Treating the embryo using only secondary antibodies led to no detectable signal and no recognized spots (**Supplementary Fig. 5e**), consistent with the same interpretation.

We next used the value of I_1 to convert the background-subtracted mean nuclear IF intensity, I_{nu} (per μm^3), to the absolute protein concentration, C_{nu}, using the following formulas:

$$C_{\mathrm{nu}} = I_{\mathrm{nu}}/A$$
$$A = (2\pi)^{1/2}\sigma_z I_1 \quad (10)$$

where σ_z is the half-width of the single-protein intensity profile in the z dimension (in the unit of z spacing, see **Supplementary Note Section 7** for mathematical details).

3.2 Quantifying nuclear Bcd concentration using IF intensity fluctuations

As an alternative method, nuclear protein concentration was also estimated based on the spatial fluctuation of the IF signal. Specifically, for each nucleus, we computed the variance of pixel intensity values at the central z slice (using the same set of pixels as for calculating the average IF intensity) and then plotted the variance versus mean for all nuclei in a single embryo (**Fig. 1c**). Next, we fitted this plot to a straight line. According to our derivation (see **Supplementary Note Section 7**), the slope k of the fit obeys

$$k \approx \frac{A}{8(\pi)^{3/2}\sigma_x\sigma_y\sigma_z} \quad (11)$$

where A is the calibration constant between fluorescence and protein concentration, as in Equation (10) above. σ_x, σ_y, σ_z are the half-widths of the single-protein intensity profile in all three dimensions, respectively (in the unit of pixel number or z spacing).

4 Measuring the gene regulation function (GRF)

4.1 Measuring the mean regulatory response

For the analysis of transcriptional regulation, we used cycles 11-14 embryos, whose gene-of-interest has not yet been replicated (prior to G2 phase). The identification of nuclear cycle phase was based on nuclear morphology and the presence of active transcription sites (Boettiger and Levine, 2013). Since our strains are homozygous, we chose embryos where most nuclei (≥ 90%) show 2 or fewer active transcription sites (Boettiger and Levine, 2013).

We next recorded the amount of nascent mRNA at each gene locus in the embryo. To identify silent loci (no nascent mRNA), we applied the following criteria: (1) Each nucleus was assumed to contain two copies of the gene. (2) For nuclei containing fewer than two active loci (recognized from the smFISH channel), corresponding numbers of silent loci were added to satisfy criterion #1. (3) Nuclei containing more than two active loci were discarded.

For a given gene, we examined its transcription in a defined AP position range. For the case of *hb* regulation by Bcd, we used 0.25-0.7 EL. This range covers the AP expression boundary of *hb* in early embryos (Gregor et al., 2007a; Houchmandzadeh et al., 2002; Struhl et al., 1989), while excluding the anterior pole and the posterior expression band of the embryo, both of which are believed to involve other transcription factors (Janody et al., 2000; Margolis et al., 1995; Perry et al., 2012; Ronchi et al., 1993). When examining *bcd3-lacZ*, we extended the range to 0.1-0.7 EL. The purpose was to include nuclei with higher Bcd concentration, since the activation of *bcd3-lacZ* occurs at a higher concentration.

For each embryo and the chosen AP position range, we plotted the number of nascent mRNA, r, against the Bcd concentration in the corresponding nucleus (**Fig. 2a**, **Supplementary Fig. 6**, **7b**). To extract the GRF, we binned individual data points by Bcd concentration and used a least-square algorithm (MATLAB function “nlinfit”) to fit the binned data to a Hill function:

$$r = r_1 \frac{[\mathrm{Bcd}]^h}{[\mathrm{Bcd}]^h + {C_{50}}^h} + r_0 \tag{12}$$

where h is the Hill coefficient, C_{50} is the concentration threshold for *hb* activation, r_1 indicates the maximal level of Bcd-dependent activity, and r_0 denotes the basal activity of *hb* (see **Supplementary Fig. 7c-f**).

To test the robustness of the fitting results, we tested different binning methods (equal population versus equal concentration interval) and different bin sizes. Fitting results were found to be insensitive to the choice of particular binning (**Fig. SN4**).

4.2 Measuring mRNA fluctuations and correlation

For each Bcd concentration bin, we calculated the coefficient of variation (CV), defined as the ratio between the standard deviation (σ) and the mean (μ) of the number of nascent mRNAs:

$$\mathrm{CV} = \frac{\sigma}{\mu} \tag{13}$$

We also calculated the Fano factor (F) for each Bcd concentration bin as the ratio between the variance (σ^2) and the mean (μ) of the nascent mRNA level:

$$F = \frac{\sigma^2}{\mu} \tag{14}$$

For a transcription model with constant probability of initiation, the Fano factor was estimated to be 22/27 (see **Supplementary Note Section 6** for details).

To calculate the Pearson correlation coefficient (ρ) between the numbers of nascent mRNAs at *hb* loci within the same nucleus, for each Bcd concentration bin, we first divided the single-locus data (number of nascent mRNAs) from each embryo into two groups, R_1 and R_2, each containing one of the two loci from each nucleus. Next, we applied the following formula:

$$\rho = \frac{\sum_i (x_i - r_1)(y_i - r_2)}{\sigma_1 \sigma_2} - \frac{\sum_i (x_i - r_1)(y_{i+1} - r_2)}{\sigma_1 \sigma_2} \tag{15}$$

where $x_i \in R_1$ and $y_i \in R_2$ belong to nucleus i (the index was sorted by Bcd concentration). r_1, r_2 and σ_1, σ_2 are the mean and standard deviation of each group, respectively. The second term in Equation (15) is used to correct for the so-called "input noise" (Tkacik et al., 2008), created by the variation of Bcd concentration within the bin. This is achieved by re-assigning the single-locus data, x_i, to pair with y_{i+1} coming from a different nucleus with similar Bcd concentration. The correlation profile of a typical cycle 12 embryo is shown in **Supplementary Figure 8**.

5 Measuring Bcd binding

5.1 The enrichment calculation

The method for quantifying transcription-factor binding at specific gene loci was motivated by the observation in (He et al., 2011), that the IF signal (corresponding to antibody-labeled protein) is enriched at active transcription sites. Here, we used the presence of nascent mRNA (smFISH signal) to identify the spatial position of (active) gene loci. For each such locus, we then defined a "locus-integration region", V_l (in units of voxel number), to cover the nearby nuclear voxels within xy distance $\leqslant r$ and z distance $\leqslant h$, and measured the total IF signal within that volume, I_l. In addition, we chose a random position somewhere else in the nucleus and measured the total IF signal, I_r, in its vicinity in the same manner. To examine protein

enrichment at the gene locus, we compared the distributions of I_l and I_r for multiple nuclei. **Supplementary Figure 14a-b** shows the results for the Bcd signal at *hb* loci, (Panel A, single embryo; Panel B, multiple embryos pooled together). A Kolmogorov-Smirnov test (Massey Jr, 1951) indicated that these two signals follow different distributions ($\alpha \leqslant 0.05$ for all embryos). To increase the accuracy of calculating the enriched IF signal, we replaced I_r with an average over multiple out-of-locus positions. Specifically, we defined the faraway voxels from the same nucleus with xy distance $>2r$ and z distance $\leqslant h$ as the "out-of-locus region", V_o, and measured the total IF signal of that region, I_o. The enriched IF signal was then calculated as

$$\boldsymbol{I_{\mathrm{enrich}} = I_l - I_o \frac{V_l}{V_o}} \tag{16}$$

To convert I_{enrich} to the absolute number of bound molecules, N_{bound}, we used the calibration constant, A (see **Supplementary Note Sections 3**, **7**), as follows:

$$\begin{aligned} N_{\mathrm{locus}} &= \frac{I_l}{A \cdot R_l} \\ N_{\mathrm{background}} &= \frac{I_o}{V_o}\frac{V_l}{A \cdot R_l} \\ N_{\mathrm{bound}} &= \frac{I_{\mathrm{enrich}}}{A \cdot R_l} = N_{\mathrm{locus}} - N_{\mathrm{background}} \end{aligned} \tag{17}$$

Here, N_{locus} and $N_{\mathrm{background}}$ denote the number of molecules in V_l and at the background region respectively, and R_l is the fraction of single-protein fluorescence signal covered by V_l (assuming that the molecule is located at the center of V_l), calculated as

$$R_l = \frac{\iiint_{V_l} d\boldsymbol{r}\, f(\boldsymbol{r})}{\|f\|} \tag{18}$$

where $f(\boldsymbol{r})$ is the intensity profile of individual molecules (see **Supplementary Note Section 7**).

We note that the specific choice of V_l is expected to affect our quantification result. Since the typical size of a smFISH spot corresponding to active gene locus is ~2 times the optical diffraction limit, a small V_l may miss some of the enrichment signal, while a large V_l may collect additional signal from other nearby genes. To determine the optimal size of V_l, we examined N_{bound} for different r's and h's. We found that N_{bound} does not vary much with h, but it increases significantly when r > 3 pixels (**Supplementary Fig. 14c**). We therefore chose r = 2 pixels and h = 0 (single image plane) for the data analysis. Under this setting, V_l is 13 voxels, and R_l = 0.335.

5.2 Estimating the required sample size

The accuracy in estimating N_{bound} is limited by the random spatial distribution of protein molecules in the nucleus. To estimate the noise level of a single measurement, we combined Equations (60) and (17), to obtain

$$B_{\mathrm{locus}} \approx N_{\mathrm{locus}} = \frac{I_l}{V_l}\frac{V_l}{A \cdot R_l} = C\frac{V_l}{R_l} \tag{19}$$

where B_{locus} denotes the background number of Bcd molecules in V_l, and C is the nuclear Bcd concentration. Under our parameter settings, V_l / R_l = 1.44 μm^3. Hence, in the anterior expression region where C > 10 nM (the transition Bcd concentration, see **Supplementary Fig. 7d**), we have $B_{\mathrm{locus}} \gtrsim 10$ molecules. The noise caused by random distribution of Bcd molecules

is calculated as $\sigma_{\mathrm{locus}} \approx \sqrt{B_{\mathrm{locus}}} \gtrsim 3$ molecules. In contrast, $N_{\mathrm{background}}$ is measured from a much larger region ($V_o \approx V_{\mathrm{nucleus}} \gg V_l$), and its noise level is thus negligible. Therefore, the noise level of a single enrichment measurement is

$$\sigma_{\mathrm{bound}} \approx \sigma_{\mathrm{locus}} \gtrsim 3 \text{ molecules} \tag{20}$$

Averaging over N loci will reduce this noise level by a factor $1/\sqrt{N}$. For the case of Bcd binding at *hb*, the estimated average enrichment is ≈ 0.5 (**Supplementary Fig. 14d**), and therefore requires averaging over at least ≈ 40 loci to emerge above the noise (**Supplementary Fig. 14d**).

5.3 Recalibrating enrichment levels using the *bcd3-lacZ* strain

Measuring N_{bound} for Bcd protein at different gene loci (*hb*, *Act5C*, *nullo* and the *bcd3-lacZ* reporter) consistently yielded values smaller than 1 (**Supplementary Fig. 14e**). Specifically, for both *hb* and *bcd3-lacZ*, the measured numbers were ~10 times smaller than the expected values assuming full occupancy of the cognate binding sites at saturating Bcd concentrations. This apparent discrepancy may represent an experimental artifact; for example, a lower efficiency of antibody labeling for DNA-bound proteins. Alternatively, the lower numbers may truly reflect the average number of bound Bcd proteins. This would be consistent with recent reports of less-than-full occupancy by transcription factors (Poorey et al., 2013), possibly due to rapid binding and unbinding (McNally et al., 2000).

To circumvent this issue, we used an alternative method to convert IF enrichment to the number of bound Bcd proteins. We measured the Bcd binding curve for the *bcd3-lacZ* reporter gene, which exhibited a plateau at high Bcd concentration (**Fig. SN5**, see **Supplementary Note Sections 5.6** for details). Postulating that this plateau corresponds to the binding of 3 Bcd molecules, we rescaled the enrichment of Bcd at all other genes accordingly.

5.4 Ruling out cross-talk between imaging channels

To verify that the enrichment signal did not simply represent cross-talk between the smFISH and IF channels, we measured the fluorescence at both channels in samples that were only labeled using smFISH. Plotting the intensity value of each pixel in the two channels revealed that the cross-talk from the smFISH channel to the IF channel was linear in the smFISH intensity, with a slope of ≈ 6×10^{-4} (data from 8 images). For the *hb* locus, the contribution to enrichment from cross-talk would therefore be ~0.001 molecule, much lower than the measured enrichment signal (**Supplementary Fig. 14e**).

5.5 Comparing the enrichment values with ChIP

To validate the quantification of Bcd binding, we compared our results with ChIP-chip data, taken from the online database of the Berkeley *Drosophila* Transcription Network Project http://bdtnp.lbl.gov/Fly-Net/, (Li et al., 2008; MacArthur et al., 2009). For a given gene, we summed all ChIP peak values within ± 100kbp of the gene (the range was set based on (Ghavi-Helm et al., 2014)). Since the database provided signals from two different Bcd antibodies, we estimated the error of ChIP measurements from the difference between these two signals (**Supplementary Fig. 14g**). We then plotted the ChIP data from each gene against the corresponding Bcd enrichment values, and fitted the data to a straight line (**Fig. 3b**).

5.6 The Bcd binding curve

To plot the Bcd binding curve at the *hb* gene, we binned the enrichment data from individual gene loci by Bcd concentration (**Fig. 3c**, **Supplementary Fig. 15**). As a robustness test, we compared different binning methods (equal population versus equal concentration interval) and different bin sizes, and found no significant differences in terms of the shape and amplitude of the curve (**Fig. SN6**).

At the lower end of the enrichment curve, where Bcd concentration is lower than the threshold for *hb* activation, our estimation of the average Bcd enrichment is expected to be higher than the actual value, since we only detect active *hb* loci, which are more likely to have Bcd bound. This explains why the lower bound of the enrichment signal is higher than zero (**Fig. 3c**, **Supplementary Fig. 15**). To limit the level of overestimation, we calculated, for each bin, the fraction of *hb* loci that are active. For cycle 11-13 embryos, we excluded bins that have <25% active *hb* loci.

The Bcd binding curves at different cycles exhibited plateaus, where binding is approximately constant over a range of Bcd concentrations (**Fig. 3c**, **Supplementary Fig. 15**). To estimate the enrichment value of each plateau, we plotted the distribution of enrichment values from all data points on the binding curve (**Fig. SN5**). This distribution exhibited multiple peaks, each of which corresponds to a plateau. We fitted these peaks to Gaussian functions to extract the mean enrichment value and width of each plateau (**Fig. SN5**).

5.7 Relating Hb and Bcd binding with *hb* transcription

To simultaneously measure the binding of Bcd and Hb proteins at *hb* loci, we applied Bcd and Hb antibodies simultaneously, following the smFISH labeling of *hb* mRNAs (**Supplementary Fig. 16**). We then quantified Bcd and Hb enrichment, as well as nascent *hb* mRNA, at each active *hb* locus. For each nuclear cycle (11-14), we combined the single-locus data (all three species) from multiple embryos, and binned the data by the embryo AP position (**Supplementary Fig. 17**).

To describe *hb* transcription as a 2D function of Bcd and Hb binding, we plotted the average amount of nascent *hb* mRNAs at each AP position against the corresponding Bcd and Hb enrichment (x and y axes) for all cycles (**Supplementary Fig. 18**). Based on the Bcd and Hb enrichment values at the transition AP position (~0.4 EL), we set a Bcd enrichment threshold (3 Bcd molecules) and a Hb enrichment threshold (2.5 Hb molecules; the conversion from enrichment to number of bound molecules was done using the same calibration factor calculated for Bcd in **Supplementary Note Section 5.3** above). We used those threshold values to divide the entire data set into 4 groups, corresponding to “high” and “low” binding of both Bcd and Hb, and calculated the average *hb* transcription level for each group (**Fig. 3d**).

6 Modeling transcription kinetics

6.1 A two-state model of initiation, followed by deterministic elongation

We model nascent transcription at the gene locus using a modified version of the two-state model (Neuert et al., 2013; Peccoud and Ycart, 1995; Raj et al., 2006). In this model, the gene can be in one of the two states: (1) an “ON” (active) state, where new transcripts can be initiated, and (2) an “OFF” (inactive) state, where the gene is silent. Transitions between states and the initiation of new transcripts (during the ON state) are all assumed to be Poisson processes (**Fig. 2c**). The possible reactions can be written as

$$\mathrm{OFF} \xrightarrow{k_{\mathrm{ON}}} \mathrm{ON}$$

SUPPLEMENTARY NOTE

$$\mathrm{ON} \xrightarrow{k_{\mathrm{OFF}}} \mathrm{OFF}$$
$$\mathrm{ON} \xrightarrow{k_{\mathrm{INI}}} \mathrm{ON} + \text{nascent mRNA}$$

where, k_{ON}, k_{OFF} and k_{INI} are the rates of gene activation, gene inactivation and transcription initiation, respectively. Next, we assume that each nascent mRNA molecule elongates with a constant speed, V_{EL}; thus it takes $T_{\mathrm{EL}} = 1/V_{\mathrm{EL}}$ to complete a transcript. Upon completion, mRNA is immediately released from the gene. This scenario can be abbreviated as follows:

$$\text{nascent mRNA} \xrightarrow{V_{\mathrm{EL}}} \text{full length mRNA} \xrightarrow{k_{\mathrm{FALL}}=\infty} \text{mature mRNA}$$

6.2 Deriving equations for the distribution of nascent mRNA numbers

We begin by noting that, at a given time point t_{ob} (the observation time), the number of nascent mRNA molecules N_{m} (the total length of nascent transcripts) on a gene is completely determined by transcription events that happened within a time window from t_{ob}–T_{EL} to t_{ob}, and the contribution from each event depends on its initiation time, t_i. Without loss of generality, let us set $t_{\mathrm{ob}} = 0$. For an given series of initiation events, $n(t) = \Sigma_i\, \delta(t_i)$, we then have

$$N_{\mathrm{m}} = \int_{-T_{\mathrm{EL}}}^{0} n(t)g(t)dt \tag{21}$$

Here, $g(t)$ is the contribution function quantifying the signal from a single transcript initiated at time t. A transcript that initiated at $t = -T_{\mathrm{EL}}$ will contribute a full length mRNA at $t_{\mathrm{ob}} = 0$, therefore $g(-T_{\mathrm{EL}}) = 1$. On the other hand, a transcript that initiated at $t = 0$ has had no time to elongate, and hence $g(0) = 0$. For intermediate values of t, g(t) is given by

$$g(t) = -t/T_{\mathrm{EL}} \tag{22}$$

(see **Fig. SN7a**).

When analyzing the smFISH signal in our experiments, we have to modify the expression for $g(t)$ to account for the positions of smFISH probes along the gene. In the simplest case, where smFISH probes are uniformly distributed along the entire gene (as is the case for *lacZ*), Equation (22) still stands. For *hb*, however, the probes cover only the coding region of the gene, leaving the 5'-UTR (including an intron) and the 3'-UTR unlabeled. Considering the lengths of the relevant features (Schroder et al., 1988), we obtain

$$g(t) = \begin{cases} 1, & -1.6T_{\mathrm{C}} \le t < -1.2T_{\mathrm{C}} \\ -0.2 - t/T_{\mathrm{C}}, & -1.2T_{\mathrm{C}} \le t < -0.2T_{\mathrm{C}} \\ 0, & -0.2T_{\mathrm{C}} \le t \le 0 \end{cases} \tag{23}$$

where $T_{\mathrm{C}} = T_{\mathrm{EL}}/1.6$ is the elongation time for the coding region (**Fig. SN7b**).

To describe the distribution of number of nascent mRNAs, we first define the state of a transcription site at time t ($-T_{\mathrm{EL}} \le t \le 0$) using two parameters: i, denoting the gene state (1 = OFF, 2 = ON); and $m = \int_{-T_{\mathrm{EL}}}^{t} n(\tau)g(\tau)d\tau$, which denotes the amount of nascent mRNAs at time 0 created by transcripts initiated between $-T_{\mathrm{EL}}$ and t. Note that m is non-decreasing with time. Specifically, $m = 0$ at $t = -T_{\mathrm{EL}}$ and $m = N_{\mathrm{m}}$ (the number of nascent mRNA molecules) at $t = 0$. Next, we write the Kolmogorov forward equations describing the time evolution of the probability distribution function, $P_i(m,t)$, as follows:

$$\begin{cases} \dfrac{dP_1(m,t)}{dt} = -\left(k_{12} + k_{ini,1}\right)P_1(m,t) + k_{21}P_2(m,t) + k_{\mathrm{INI},1}P_1(m - g(t),t) \\ \dfrac{dP_2(m,t)}{dt} = -\left(k_{21} + k_{ini,2}\right)P_2(m,t) + k_{12}P_1(m,t) + k_{\mathrm{INI},2}P_2(m - g(t),t) \end{cases} \tag{24}$$

where $k_{12} = k_{\mathrm{ON}}$, $k_{21} = k_{\mathrm{OFF}}$, $k_{\mathrm{INI},1} = 0$, and $k_{\mathrm{INI},2} = k_{\mathrm{INI}}$. The distribution of number of nascent mRNAs at the observation time $t = 0$ can be written as

$$P(N) = P_1(m,0) + P_2(m,0) \tag{25}$$

Before we proceed to solve the full model, we note that in the special case where the gene is always in the active state, we can write a simpler equation,

$$\frac{dP(m,t)}{dt} = -k_{\mathrm{INI}}P(m,t) + k_{\mathrm{INI}}P(m - g(t),t) \tag{26}$$

By multiplying both sides of Equation (26) by m and m^2, we can easily obtain equations for the mean and variance of m, such that,

$$\frac{d\langle m\rangle}{dt} = k_{\mathrm{INI}}g(t) \tag{27}$$

$$\frac{d\sigma_m^2}{dt} = k_{\mathrm{INI}}g^2(t) \tag{28}$$

Substituting $g(t)$ from Equation (23) above, we can calculate the Fano factor,

$$F = \left.\frac{\sigma_m^2}{\langle m\rangle}\right|_{t=0} = \frac{\int_{-T_{\mathrm{EL}}}^{0} g^2(\tau)d\tau}{\int_{-T_{\mathrm{EL}}}^{0} g(\tau)d\tau} = \frac{\int_{-1.6}^{-1.2} d\tau + \int_{-1.2}^{-0.2}(\tau + 0.2)^2 d\tau}{\int_{-1.6}^{-1.2} d\tau + \int_{-1.2}^{-0.2}(\tau + 0.2)d\tau} = \frac{22}{27} \tag{29}$$

6.3 Solving the Kolmogorov forward equations using the Finite State Projection (FSP) method

To solve Equation (24), we use the Finite State Projection (FSP) algorithm (Munsky and Khammash, 2008; Neuert et al., 2013). Since m belongs to the uncountable set $[0,\infty)$, we first discretize it to $m = 0, \Delta m, 2\Delta m, \ldots$, with $\Delta m \ll 1$. The probability distribution functions, $P_1(m,t)$ and $P_2(m,t)$, can then be replaced by a vector $\boldsymbol{P}(t)$, such that,

$$\boldsymbol{P}(t) = \begin{bmatrix} \boldsymbol{P}_0 \\ \boldsymbol{P}_{\Delta m} \\ \vdots \end{bmatrix}_t = \begin{bmatrix} \begin{bmatrix} P_{1,0} \\ P_{2,0} \end{bmatrix} \\ \begin{bmatrix} P_{1,\Delta m} \\ P_{2,\Delta m} \end{bmatrix} \\ \vdots \end{bmatrix}_t \tag{30}$$

where the vector $\boldsymbol{P}_m$ groups the probabilities of observing m when the gene is either in the ON or OFF states. Similarly, we discretize the contribution function, $g(t)$, by rounding up its values to the nearest multiplications of Δm.

We next rewrite Equation (24) as

SUPPLEMENTARY NOTE

$$\frac{d}{dt}\boldsymbol{P}(t) = \boldsymbol{Q}(t)\boldsymbol{P}(t) = \begin{bmatrix} \boldsymbol{K}-\boldsymbol{T} & 0 & 0 & 0 \\ 0 & \boldsymbol{K}-\boldsymbol{T} & 0 & \cdots \\ 0 & 0 & \boldsymbol{K}-\boldsymbol{T} & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ \boldsymbol{T} & 0 & 0 & \ddots \\ 0 & \boldsymbol{T} & 0 & \ddots \\ 0 & 0 & \boldsymbol{T} & \ddots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \boldsymbol{P}_0 \\ \boldsymbol{P}_{\Delta m} \\ \boldsymbol{P}_{2\Delta m} \\ \vdots \\ \boldsymbol{P}_{1-g(t)} \\ \boldsymbol{P}_{1-g(t)+\Delta m} \\ \boldsymbol{P}_{1-g(t)+2\Delta m} \\ \vdots \end{bmatrix}_t \tag{31}$$

where $\boldsymbol{K}$ and $\boldsymbol{T}$ are matrices describing gene-state transition and transcription initiation, respectively,

$$\boldsymbol{K} = \begin{bmatrix} -k_{12} & k_{21} \\ k_{12} & -k_{21} \end{bmatrix}; \boldsymbol{T} = \begin{bmatrix} k_{\mathrm{INI},1} & 0 \\ 0 & k_{\mathrm{INI},2} \end{bmatrix} \tag{32}$$

$\boldsymbol{Q}(t)$ is the propagation matrix describing the transition between different i's (gene states) and m's (nascent mRNA numbers) at time t. Note that $\boldsymbol{Q}(t)$ is time dependent. Specifically, at $t = -T_{\mathrm{EL}}$ we have

$$\frac{d}{dt}\boldsymbol{P}(-T_{\mathrm{EL}}) = \boldsymbol{Q}(-T_{\mathrm{EL}})\boldsymbol{P}(-T_{\mathrm{EL}}) = \begin{bmatrix} \boldsymbol{K}-\boldsymbol{T} & 0 & 0 & 0 \\ 0 & \boldsymbol{K}-\boldsymbol{T} & 0 & \cdots \\ 0 & 0 & \boldsymbol{K}-\boldsymbol{T} & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ \boldsymbol{T} & 0 & 0 & \ddots \\ 0 & \boldsymbol{T} & 0 & \ddots \\ 0 & 0 & \boldsymbol{T} & \ddots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \boldsymbol{P}_0 \\ \boldsymbol{P}_{\Delta m} \\ \boldsymbol{P}_{2\Delta m} \\ \vdots \\ \boldsymbol{P}_1 \\ \boldsymbol{P}_{1+\Delta m} \\ \boldsymbol{P}_{1+2\Delta m} \\ \vdots \end{bmatrix}_{t=-T_{\mathrm{EL}}} \tag{33}$$

To calculate the distribution of number of nascent mRNAs, we need to time-propagate the initial state, $\boldsymbol{P}(-T_{\mathrm{EL}})$, all the way to $t = 0$, i.e.

$$\boldsymbol{P}(0) = (\boldsymbol{I} + \boldsymbol{Q}(-\Delta t)\Delta t) \cdots (\boldsymbol{I} + \boldsymbol{Q}(-T_{\mathrm{EL}} + \Delta t)\Delta t)(\boldsymbol{I} + \boldsymbol{Q}(-T_{\mathrm{EL}})\Delta t)\boldsymbol{P}(-T_{\mathrm{EL}}) \tag{34}$$

where $\boldsymbol{I}$ is the unit matrix and Δt is the time step. Since m is always zero at $t = -T_{\mathrm{EL}}$, we can write the initial state, $\boldsymbol{P}(-T_{\mathrm{EL}})$, as

$$\boldsymbol{P}(-T_{\mathrm{EL}}) = \begin{bmatrix} \boldsymbol{P}_0(-T_{\mathrm{EL}}) \\ 0 \\ \vdots \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \frac{k_{21}}{k_{12}+k_{21}} \\ \frac{k_{12}}{k_{12}+k_{21}} \end{bmatrix} \\ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \vdots \end{bmatrix} \tag{35}$$

where the distribution of gene states in $\boldsymbol{P}_0(-T_{\mathrm{EL}})$ satisfies the steady-state condition. Plugging in Equation (24) into Equation (34) will allow us to calculate $\boldsymbol{P}(0)$. Equation (25) can then be used to convert $\boldsymbol{P}(0)$ to the distribution of number of nascent mRNAs.

We note that, in Equation (34), the kinetic rates (k_{ON}, k_{OFF}, and k_{INI}) always appear multiplied by the time step, Δt. We can therefore define a time constant, T_0, and scale all parameters in Equation (34) to dimensionless ones. Specifically, the value of T_{EL} is rescaled to T_{EL}/T_0, the value of Δt is rescaled to $\Delta t/T_0$, and the values of k_{ON}, k_{OFF}, and k_{INI} are rescaled to $k_{\mathrm{ON}} \cdot T_0$, $k_{\mathrm{OFF}} \cdot T_0$, $k_{\mathrm{INI}} \cdot T_0$, respectively. To simply the expression for $g(t)$ (see Equations (22) and (23)), we choose

$T_0 = T_{EL}$ for *lacZ* and $T_0 = T_{EL}/1.6$ for *hb*, where T_{EL} = 2 min for *lacZ* and T_{EL} = 2.5 min for *hb* (values estimated based on the length of the two genes).

Until this step, m has no upper boundary. As a result, $\boldsymbol{P}(t)$ has infinite dimensions. To enable numerical calculation of the relevant equations, we follow the FSP approach (Munsky and Khammash, 2008; Neuert et al., 2013), and truncate the dimensions of vectors and matrices in Equation (34) by setting an upper limit of m. The truncated $\boldsymbol{P}(t)$ and $\boldsymbol{Q}(t)$ must still be large enough to cover the main portion of the distribution. According to our experimental data, the number of nascent mRNAs at individual *hb* loci seldom exceeds 100 (less than 1 per 5000 loci), we therefore set the upper limit to be $m = 100$.

We solve Equation (34) using MATLAB. To compromise accuracy and speed, we set $\Delta m = 0.1$ and $\Delta t = 0.001$. After code optimization, we are able to achieve a speed of <0.05 second per solution. Such a speed is necessary for the large-scale parameter search (see **Online Methods Section 8.4**). To ensure the accuracy of computation, we compare the computed probability distribution with Gillespie simulations (Gillespie, 1977; Zenklusen et al., 2008) at different parameter settings (**Fig. SN8**). The two calculation methods agree well over the full range of parameters tested ($k_{ON} \cdot T_0$ and $k_{OFF} \cdot T_0$: 0.01-10, $k_{INI} \cdot T_0$: 1-100; $R^2 = 0.975 \pm 0.004$ from 64 simulations).

6.4 Maximum Likelihood Estimation (MLE) of transcription parameters

To fit our experimental data to the theoretical model and estimate the kinetic parameters, we apply a Maximum Likelihood Estimation (MLE) method, following the approach of (Neuert et al., 2013). Briefly, for a given parameter set ($\boldsymbol{K} = \{k_{ON}, k_{OFF}, k_{INI}\}$), if the probability of observing N_m nascent mRNAs at a single locus is $P(N|\boldsymbol{K})$, then the likelihood of $\boldsymbol{K}$ for a set of observations $\boldsymbol{N} = \{N_1, N_2, \ldots\}$ is defined as

$$L(\boldsymbol{N}|\boldsymbol{K}) = \prod_j P\left(N_j|\boldsymbol{K}\right) \tag{36}$$

The best fit to our experimental data should satisfy the highest likelihood (or log-likelihood), namely,

$$\boldsymbol{K}_{\text{best}} = \underset{\boldsymbol{K}}{\text{argmax}}\left(L(\boldsymbol{N}|\boldsymbol{K})\right) = \underset{\boldsymbol{K}}{\text{argmax}}\left(\log\left(L(\boldsymbol{N}|\boldsymbol{K})\right)\right) = \underset{\boldsymbol{K}}{\text{argmax}}\left(\sum_j \log\left(P\left(N_j|\boldsymbol{K}\right)\right)\right) \tag{37}$$

For a given embryo, we divide the single-locus data set into multiple subsets according to the nuclear Bcd concentration (the subset is marked as $\boldsymbol{N}([\text{Bcd}])$). To ensure a sufficient number of data points in each subset, we use equal-population binning. We adjust the bin size for each nuclear cycle, to limit the range of Bcd concentration in a single bin (bin size for cycle 11: 100; for cycles 12-13: 200; for cycle 14: 250). In all cases, we end up with ~15 bins per embryo.

To fit the nascent-mRNA distribution in a given bin, we test a broad range of $\{k_{ON}, k_{OFF}, k_{INI}\}$ values (covering what we considered the "physiologically plausible" space): k_{ON} and k_{OFF} from 0 to 10 min^{-1}, and k_{INI} from 0 to 100 min^{-1}. We then look for maxima of the likelihood function (see **Supplementary Fig. 9**).

We followed the free parameter fit with a modified fit, using the assumption that Bcd only regulates k_{ON}. In that case, we fit all bins ($n = 1, \ldots, n_0$) together by constructing a total-likelihood function for the embryo,

SUPPLEMENTARY NOTE

$$L_{all}(\boldsymbol{N}|\boldsymbol{K}) = \prod_{n=1}^{n_0} P(\boldsymbol{N}([\mathrm{Bcd}]_n)|\{k_{\mathrm{ON}}(n), k_{\mathrm{OFF}}, k_{\mathrm{INI}}\}) \quad (38)$$

Here k_{OFF} and k_{INI} are both single valued, whereas k_{ON} is bin-specific. Therefore, an embryo with n_0 bins will have n_0+2 parameters. To perform parameter search in such high dimensions, we use a combination of simplex and simulated annealing methods (MATLAB function: "simulannealbnd"). To increase the accuracy, the search is applied to each embryo 24 times, and the result with the highest L_{all} is chosen. To further verify that the fitting did not get stuck in local solutions, we follow up with a complimentary approach: scanning all three parameters simultaneously within a similar range (k_{ON} and k_{OFF} from 0.01 min^{-1} to 10 min^{-1}, and k_{INI} from 1 min^{-1} to 100 min^{-1}), and calculating the likelihood function for each parameter combination. We found that the scanning and fitting yielded very similar estimates for the kinetic parameters, suggesting that the solution found using the combined simplex and simulated annealing algorithm is global, not local.

For all fitting and scanning approaches, computation was performed on a single node (12 cores, 48 GB memory) of a computer cluster (at the Computational and Integrative Biomedical Research Center of Baylor College of Medicine), using MATLAB Parallel Computing Toolbox. For a typical embryo, it took ~24 hours to fit each bin independently (all k's), and ~48 hours to fit all bins together (k_{ON} only).

6.5 Gillespie simulation and comparison with live data

We used stochastic simulation to generate virtual kinetic data, and compared it with data from the literature, obtained using live imaging of an *hb*-MS2 reporter gene (Lucas et al., 2013). We first created a set of "virtual embryos" (cycles 10-13), each containing 200 *hb*-MS2 gene loci evenly distributed within the AP position range 0.15-0.85 EL (The number of loci, AP position range, and all other parameters were chosen to match the details in (Lucas et al., 2013). To model the regulation by Bcd, we created a virtual Bcd concentration, $C(x) = C_{\max}\, e^{-x/d}$, where x denotes the AP position. $C_{\max}$ = 75 nM and d = 0.225 EL were set according to our findings in the current study (**Fig. SN3**). For each gene locus, the transcriptional activity during a nuclear cycle was divided into three time periods: (1) During period T_1 (lasting 3.5 min for all cycles), the gene was at the OFF state, with no nascent mRNA produced. (2) During period T_2 (lasting 1 min for cycle 10, 2.5 min for cycle 11, 4.5 min for cycle 12, 8 min for cycle 13), transcription initiation followed two-state kinetics and was modulated by Bcd concentration. (3) During period T_3 (the remainder of the nuclear cycle), the gene is OFF again, but any nascent transcripts initiated during T_2 continue to elongate until they are completed.

To simulate the two-state kinetics during T_2, we used the kinetic parameters found for the endogenous *hb* gene, namely, k_{OFF} = 1 min^{-1}, k_{INI} = 32 min^{-1}, $k_{\mathrm{ON}} = k_1\, C^h / (C^h + {C_{50}}^h) + k_0$, k_1 = 1.1 min^{-1}, k_0 = 0.05 min^{-1}, h = 6, C_{50} = 10 nM (**Supplementary Fig. 12**). Based on the length of the *hb*-MS2 reporter gene in (Lucas et al., 2013), we estimated its elongation time to be T_{EL} = 1.70 min. Since only the array of MS2 binding sites is fluorescently detected and quantified, we defined the response function as follows:

$$g(t)_{hb-\mathrm{MS2}} = \begin{cases} -0.95 - 1.95 \cdot t, & -1 \le t < -0.49 \\ 0, & -0.49 \le t \le 0 \end{cases} \quad (39)$$

where the time scale has been normalized by T_{EL}.

Following the procedure in (Lucas et al., 2013), we recorded the following two quantities from each active gene locus: (1) The initiation time, defined as the time when the first transcription

event takes place; and (2) The persistence, defined as the time interval during which nascent mRNA signal is detected (**Supplementary Fig. 11a**). Next, we defined an AP position threshold (cycles 10-12: 0.5 EL; cycle 13: 0.4 EL), and used this threshold to partition all loci in an embryo into "anterior" and "posterior" groups. We then calculated the average values of the initiation time and the persistence for each group, during each cycle. Comparing our simulated data with the values from (Lucas et al., 2013) showed good agreement (**Supplementary Fig. 11b-c**). Consistent with the findings in (Lucas et al., 2013), the simulated initiation time did not show a noticeable trend with either the nuclear cycle or the AP position. In contrast, the persistence increased with nuclear cycle and decreased with AP position.

6.6 Generalization to more than two states

Our theoretical formalism can be extended to describe a model with multiple gene states. The general form of Equation (31) remains the same, but the dimensions of vectors and matrices involved are increased, to describe the additional states. Thus, for a three-state model (**Fig. SN9a**), we have

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{P}_0 \\ \boldsymbol{P}_{\Delta m} \\ \vdots \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} P_{1,0} \\ P_{2,0} \\ P_{3,0} \end{bmatrix} \\ \begin{bmatrix} P_{1,\Delta m} \\ P_{2,\Delta m} \\ P_{3,\Delta m} \end{bmatrix} \\ \vdots \end{bmatrix} \tag{40}$$

$$\boldsymbol{K} = \begin{bmatrix} -k_{12} - k_{13} & k_{21} & k_{31} \\ k_{12} & -k_{21} - k_{23} & k_{32} \\ k_{13} & k_{23} & -k_{31} - k_{32} \end{bmatrix}; \boldsymbol{T} = \begin{bmatrix} k_{\mathrm{INI},1} & 0 & 0 \\ 0 & k_{\mathrm{INI},2} & 0 \\ 0 & 0 & k_{\mathrm{INI},3} \end{bmatrix} \tag{41}$$

$$\boldsymbol{P}(-T_{\mathrm{EL}}) = \begin{bmatrix} \boldsymbol{P}_0(-T_{\mathrm{EL}}) \\ 0 \\ \vdots \end{bmatrix} \tag{42}$$

where $\boldsymbol{P}_0$ obeys the steady-state requirement:

$$\boldsymbol{K}\boldsymbol{P}_0(-T_{\mathrm{EL}}) = 0 \tag{43}$$

A three-state model can be used to reproduce the observed correlation between *hb* alleles within the same nucleus (**Supplementary Fig. 9c**). To show that, we first assume that only one of the three states is transcriptionally active ($k_{\mathrm{INI},1} = k_{\mathrm{ini},2} = 0$, $k_{\mathrm{ini},3} > 0$), and that transitions only occur between neighboring ones. Next, we define both state 1 and state 2 as "nuclear" states, such that transitions between them must happen simultaneously for both alleles. In contrast, state 3 describes the state of individual *hb* alleles (**Fig. SN9b**). If we further assume that k_{23} responds to Bcd concentration in a power-law form ($k_{23} = k_1 (C^h / C_{50})^h + k_0$, k_1 = 1.2 min^{-1}, k_0 = 0.05 min^{-1}, h = 6, C_0 = 10 nM) while all other transition rates are constant (k_{12} = 1.2 min^{-1}, k_{21} = 0.9 min^{-1}, k_{32} = 1.2 min^{-1}, $k_{\mathrm{INI},3}$ = 32 min^{-1}), we are then able to recreate the observed trend of intra-nuclear correlations between nascent *hb* mRNA levels (**Fig. SN9c**).

7 Deriving the fluctuation method for measuring protein concentration

SUPPLEMENTARY NOTE

7.1 Modeling the immunofluorescence signal in the nucleus

To model the immunofluorescence signal, we start with the following assumptions: (1) Bcd molecules are distributed randomly in a 3D volume V (the nucleus), with uniform probability over V. (2) Each Bcd molecule exhibits the same intensity profile in space, which is not affected by the presence of other Bcd molecules. (3) The size of a single Bcd molecule is much smaller than the optical diffraction limit, therefore the intensity profile of a single Bcd molecule can be described by the point spread function (PSF) $f(\boldsymbol{r}-\boldsymbol{r}_0)$, where $\boldsymbol{r} = (x,y,z)$ is the imaging position, and $\boldsymbol{r}_0 = (x_0,y_0,z_0)$ is the position of the Bcd molecule.

Suppose a nucleus contains N Bcd molecules at positions $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N$. The observed fluorescence signal I at a position $\boldsymbol{r}$ is then the sum of individual PSFs, or,

$$I(\boldsymbol{r}|\boldsymbol{r_1},\cdots,\boldsymbol{r_N}) = \sum_{i=1}^{N} f(\boldsymbol{r}-\boldsymbol{r_i}) \tag{44}$$

Therefore, by averaging over all possible positions of all molecules, the expectation value of I at the position $\boldsymbol{r}$ is

$$\overline{I(\boldsymbol{r})} = \sum_{i=1}^{N} \int_V d\boldsymbol{r_i}\, f(\boldsymbol{r}-\boldsymbol{r_i}) p(\boldsymbol{r_i}) = \frac{N}{V}\int_V d\boldsymbol{r_0}\, f(\boldsymbol{r}-\boldsymbol{r_0}) \tag{45}$$

where $p(\boldsymbol{r}_i) = 1/V$ is the probability of having the i'th Bcd molecule at position $\boldsymbol{r}_i$.

7.2 The mean and variance of nuclear immunofluorescence intensity

Suppose that the image of a nucleus contains m voxels, with each voxel corresponding to an imaging position $\boldsymbol{r}_j$ ($j = 1, \ldots, m$). The fluorescence signal from each voxel obeys Equation (45) with $\boldsymbol{r} = \boldsymbol{r}_j$. Therefore, the average fluorescence over all voxels should have the following expectation value:

$$\overline{\mathrm{Mean}(I)} = \langle \overline{I(\boldsymbol{r_j})} \rangle_{j=1,\cdots,m} = \frac{N}{V}\frac{1}{m}\sum_{j=1}^{m}\int_V d\boldsymbol{r_0}\, f(\boldsymbol{r_j}-\boldsymbol{r_0}) \tag{46}$$

where "$\langle\cdot\rangle$" denotes averaging over different imaging positions. Similarly, the expectation value for the variance of fluorescence is

$$\overline{\mathrm{Var}(I)} = \overline{\langle I(\boldsymbol{r_j})^2\rangle - \langle I(\boldsymbol{r_j})\rangle^2} = \overline{\langle I(\boldsymbol{r_j})^2\rangle} - \overline{\langle I(\boldsymbol{r_j})\rangle^2} \tag{47}$$

where,

$$\begin{aligned}\overline{\langle I(\boldsymbol{r_j})^2\rangle} &= \frac{1}{m}\sum_{j=1}^{m}\overline{I(\boldsymbol{r_j})^2} \\ &= \frac{1}{m}\sum_{j=1}^{m}\left[\frac{N}{V}\int_V d\boldsymbol{r_0}\, f(\boldsymbol{r_j}-\boldsymbol{r_0})^2 + \frac{N(N-1)}{V^2}\left(\int_V d\boldsymbol{r_0}\, f(\boldsymbol{r_j}-\boldsymbol{r_0})\right)^2\right]\end{aligned} \tag{48}$$

$$\overline{\langle I(\boldsymbol{r_j})\rangle^2} = \frac{1}{m^2}\overline{\left(\sum_{j=1}^{m}\sum_{i=1}^{N} f(\boldsymbol{r_j}-\boldsymbol{r_i})\right)^2} \tag{49}$$

$$= \frac{1}{m^2}\sum_{j=1}^{m}\sum_{j'=1}^{m}\left[\frac{N}{V}\int_V d\boldsymbol{r_0}\, f(\boldsymbol{r_j}-\boldsymbol{r_0})f(\boldsymbol{r_{j'}}-\boldsymbol{r_0}) + \frac{N(N-1)}{V^2}\int_V d\boldsymbol{r_0}\, f(\boldsymbol{r_j}-\boldsymbol{r_0})\int_V d\boldsymbol{r_{0'}}\, f(\boldsymbol{r_{j'}}-\boldsymbol{r_{0'}})\right]$$

Next, we note that the size of the PSF (~200 nm in x and y directions and ~500 nm in z direction) is much smaller than the whole nucleus (~6 μm in all three dimensions). Therefore, by excluding the nuclear boundary voxels (see **Supplementary Note Section 3** for details), we can replace the integrals in Equations (46), (48), (49) with the following approximations:

$$\int_V d\boldsymbol{r_0}\, f(\boldsymbol{r_j}-\boldsymbol{r_0}) \approx \iiint d\boldsymbol{r}\, f(\boldsymbol{r}) \equiv \|f\| \tag{50}$$

$$\int_V d\boldsymbol{r_0}\, f(\boldsymbol{r_j}-\boldsymbol{r_0})^2 \approx \iiint d\boldsymbol{r}\, f(\boldsymbol{r})^2 \equiv \|f^2\| \tag{51}$$

where "∭" denotes integration over the entire 3D space. Equations (50) and (51) allow us to rewrite Equations (46) and (47) as

$$\overline{\mathrm{Mean}(I)} \approx \frac{N}{V}\|f\| \tag{52}$$

$$\overline{\mathrm{Var}(I)} \approx \frac{N}{V}\|f^2\| - \frac{N}{Vm^2}\sum_{j,j'=1}^{m}\|f_j f_{j'}\| \tag{53}$$

where $f_j = f(\boldsymbol{r}_j - \boldsymbol{r}_0)$.

In Equation (53), since f is a localized function, it is easy to see that the magnitude of $\|f_j f_{j'}\|$ decreases with the distance between two imaging positions $\boldsymbol{r}_j$ and $\boldsymbol{r}_{j'}$. More specifically, a significant contribution is obtained only when this distance is smaller than the PSF size. Therefore, the double sum in Equation (53) should be proportional to m instead of m^2, namely,

$$\sum_{j,j'=1}^{m}\|f_j f_{j'}\| = mG_m\|f^2\| \tag{54}$$

where G_m is a function of the voxel number m, and should saturate at large m. We calculated G_m numerically, by simulating the image of a "virtual nucleus" using the actual form of the PSF for different fluorescent channels (see **Supplementary Note Section 7.3**). The results are shown in **Supplementary Figure 5f**. Since G_m is always smaller than 40, while m > 1000 voxels for a typical nucleus, we can neglect the second term on the right hand side of Equation (53), such that,

$$\overline{\mathrm{Var}(I)} \approx \frac{N}{V}\|f^2\| \tag{55}$$

7.3 Obtaining the single-Bcd intensity from nuclear intensity fluctuations

As seen from Equation (52), the key to converting the average pixel intensity of a nucleus to its Bcd concentration is $\|f\|$, the integrated fluorescence signal of a single Bcd molecule. To obtain this quantity, we measured the ratio between the variance and mean of the fluorescence from individual nuclei, k (equivalently, the slope of our linear fit for the variance vs. mean plot, see

Supplementary Note Section 3.3). According to Equations (52) and (55), k is a constant that depends only on the shape and magnitude of the PSF, i.e.

$$k = \frac{\overline{\mathrm{Var}(I)}}{\overline{\mathrm{Mean}(I)}} \approx \frac{\|f^2\|}{\|f\|} \tag{56}$$

For confocal imaging, we can approximate the PSF using a 3D Gaussian function (Zhang et al., 2007),

$$f = \frac{A}{(2\pi)^{3/2}\sigma_x\sigma_y\sigma_z} e^{-\frac{x^2}{2\sigma_x^2}-\frac{y^2}{2\sigma_y^2}-\frac{z^2}{2\sigma_z^2}} \tag{57}$$

where σ_x, σ_y, σ_z are half-widths of the PSF in the x, y, z dimensions, respectively, and A is the amplitude, i.e.

$$\|f\| = A \tag{58}$$

Therefore, by applying Equation (57) to Equation (56), we get

$$A \approx 8(\pi)^{3/2}\sigma_x\sigma_y\sigma_z k \tag{59}$$

This directly relates the single-Bcd fluorescence to the measurable quantity k. Bcd concentration can now be calculated using Equation (52), i.e.

$$C = \frac{N}{V} = \frac{\overline{\mathrm{Mean}(I)}}{A} \tag{60}$$

Recall that, as an alternative method, we also extracted the single-Bcd fluorescence A by identifying individual antibody-labeled molecules (see **Supplementary Note Section 3.2**). In this case, the typical spot intensity I_1 is essentially the integrated intensity (over xy dimensions) at the central z slice, therefore,

$$I_1 = \int_S f_{z=0} dS = \frac{A}{(2\pi)^{1/2}\sigma_z} \tag{61}$$

The agreement between the two methods in estimating Bcd concentration (**Fig. 1d**) indicates that they both provide similar estimates for the single-molecule intensity *A*.

In the above derivations, the PSF half-widths, σ_x, σ_y, σ_z, were determined experimentally from individual Bcd particles. Specifically, for Alexa 488 channel, σ_{xy} = 1.35 pixels (110 nm) and σ_z = 0.8 voxel (280 nm). For Alexa 647 channel, σ_{xy} = 1.8 pixels (150 nm) and σ_z = 1 voxel (350 nm).

7.4 Estimating the effect of imaging noise

Experimental noise due to the imaging process may cause an overestimation of the variance in fluorescence, leading to an error in our measurement of Bcd concentration. To characterize this effect, we consider the imaging "shot noise", which is due to the stochastic nature of photon emission and counting (Pawley, 2006). The shot noise obeys Poisson statistics,

$$P(q|Q) = \frac{Q^q}{q!} e^{-Q} \tag{62}$$

where Q is the expected number of photons (per pixel and per acquisition), and q is the actual number of photons being observed (for a given pixel and during a given acquisition).

To incorporate the effect of the shot noise, we need to rewrite the expressions for the fluorescence mean and variance (Equations (46) and (47)) to include averaging over different

q's. We assume that the observed fluorescence intensity of a voxel is proportional to the number of detected photons ($I = \beta q$, where β is a constant), and likewise the expected fluorescence intensity of a voxel is proportional to the expected photon number ($\Sigma f_j = \beta Q$). By re-defining "$\|\cdot\|$" as both an integration over space and a summation over all possible q values (q = 0,1,2,…), we can rewrite Equations (46) and (47) as

$$\overline{\mathrm{Mean}(I)} \approx \frac{N}{V}\|f\| = \frac{N\beta}{V}\left\|\frac{f(\boldsymbol{r})}{\beta}\right\| = \frac{N\beta}{V}\iiint d\boldsymbol{r}\sum_{q=0}^{\infty} qP\left(q\middle|\frac{f(\boldsymbol{r})}{\beta}\right) = \frac{N}{V}\iiint d\boldsymbol{r}\, f(\boldsymbol{r}) \tag{63}$$

$$\begin{aligned}\overline{\mathrm{Var}(I)} &\approx \frac{N}{V}\|f^2\| = \frac{N\beta^2}{V}\left\|\frac{f(\boldsymbol{r})^2}{\beta^2}\right\| = \frac{N\beta^2}{V}\iiint d\boldsymbol{r}\sum_{q=0}^{\infty} q^2P\left(q\middle|\frac{f(\boldsymbol{r})}{\beta}\right)\\ &= \frac{N\beta^2}{V}\iiint d\boldsymbol{r}\left(\frac{f(\boldsymbol{r})}{\beta}+\frac{f(\boldsymbol{r})^2}{\beta^2}\right) = \frac{N}{V}\left[\beta\iiint d\boldsymbol{r}\, f(\boldsymbol{r}) + \iiint d\boldsymbol{r}\, f(\boldsymbol{r})^2\right]\end{aligned} \tag{64}$$

In the final expression of Equation (64), the first term indicates the contribution of the shot noise. Therefore, the effect of noise can be diminished by lowering β. This can be achieved by scanning each pixel multiple times and taking the average intensity value (this is done using a standard setting of the Zeiss Zen imaging software). When varying the number of scans, we found that the slope k between intensity variance and mean (see **Supplementary Note Section 7.3**) initially decreases with the scan number and then approaches a constant value (**Supplementary Fig. 5g**). This indicates that the noise level of a single scan is significant (>40%), and averaging over multiple scannings is necessary for accurate quantification of k. In our experiments, we chose to use 8 scannings per pixel, which was enough to lower the shot noise effect to <10%, and only caused a photobleaching of <1.5%, while allowing the imaging of an entire embryo in ~3 hours.

8 Error estimates

8.1 Estimating the efficiency of mRNA labeling and detection

Our accuracy of quantifying mRNA depends critically on the efficiency of labeling the mRNA using smFISH probes. This efficiency can be described through the parameter p_0, defined as the probability of a probe binding site being occupied. We developed four different methods to estimate p_0:

(1) In the histogram of smFISH spot intensities, real mRNAs and false-positive spots appear as separate peaks (**Supplementary Fig. 4a**). Assuming that false-positive spots mainly correspond to the nonspecific binding of individual smFISH probes, we estimated the most probable number of probes binding on an individual mRNA molecule as the ratio between the intensity of the positive and the negative peaks. The probe binding probability is therefore

$$p_0 = \frac{I_{\mathrm{positive}}}{N_{\mathrm{set}}\cdot I_{\mathrm{negative}}} \tag{65}$$

where N_{set} is the total number of probes in the probe set (N_{set} = 48 for *hb*), I_{positive} is the typical intensity of individual mRNA spots, I_{negative} is the intensity of false-positive spots. This calculation yielded p_0 = 0.12 ± 0.01 (**Supplementary Fig. 4e**, data averaged from 3 embryos).

(2) As described in **Supplementary Note Section 2.1**, we can use a Gaussian fit to extract the typical intensity I_0 and the half-width w_0 of the intensity distribution for positive mRNA spots (**Supplementary Fig. 4b**). If we assume that the binding and unbinding of each probe is

independent of all other probes, then the number of bound probes follows simple binomial statistics. If we further assume that the fluorescence from each probe has a fixed value A_0, we can then express I_0 and w_0 as

$$\begin{aligned} I_0 &= N_{\text{set}} \cdot p_0 \cdot A_0 \\ w_0 &= \sqrt{N_{\text{set}} \cdot p_0 \cdot (1-p_0)} \cdot A_0 \end{aligned} \tag{66}$$

Accordingly, the probe binding probability p_0 can be estimated as

$$p_0 = \frac{1}{N_{\text{set}} \cdot (\frac{w_0}{I_0})^2 + 1} \tag{67}$$

This method yielded p_0 = 0.08 ± 0.02 (**Supplementary Fig. 4e**, data averaged from 3 embryos).

(3) Assuming that the probe binding probability for different sequences is similar, we performed a two-color experiment similar to the one described in **Supplementary Note Section 1.1**, where we labeled the same set of *lacZ* smFISH probes with two different colors, and applied them in equal molarity to the same embryo. To compare the signal from the two different channels, we identified mRNA spots that appeared in both channels (for spot matching procedure, see **Supplementary Note Section 1.1**), and plotted the 2D histogram of spot intensities (**Supplementary Fig. 4c**). To estimate p_0 from the histogram, we consider the Pearson correlation coefficient, ρ, between two channels. Because probes with different colors compete to hybridize with the target sequence, we expect negative correlation between channels. To extract the correlation from experimental data, we fitted the histogram to a 2D Gaussian function,

$$p(I,J) = Ae^{-(a(I-I_0)^2 - 2b(I-I_0)(J-J_0) + c(J-J_0)^2)} \tag{68}$$

where I and J are the intensity values from the two channels, respectively. Applying this distribution function, we get

$$\rho = \frac{b}{2\sqrt{ac}} \tag{69}$$

If we now assume that, for each channel, probe binding/unbinding follows binomial statistics with a probe binding probability of $0.5p_0$, then the two-color system follows multinomial statistics. As a result, we can write p_0 as a function of the correlation coefficient, ρ (Evans et al., 2000), such that,

$$p_0 = \frac{-2\rho}{1-\rho} \tag{70}$$

Using this method, we obtained p_0 = 0.11 ± 0.01 (**Supplementary Fig. 4e**, data averaged from 4 embryos).

(4) Following the same assumption as in method #3, we created a set of smFISH probes for the *lacZ* gene, where probes corresponding to the 5' half of the gene were conjugated with one color, while those corresponding to the 3' half were conjugated with another color. We applied this probe set to the embryo and identified the mRNA spots that appear in each channel (**Supplementary Fig. 4d**). Next, we counted for each channel the fraction of spots that also appeared in the other channel. Assuming that probes of different colors bind independently, this conditional probability should be identical to the marginal probability of mRNA detected in a single channel: ($P(1|2) = P(1)$, $P(2|1) = P(2)$). To relate this mRNA detection efficiency to p_0, we again assume binomial statistics of probe binding, to obtain

SUPPLEMENTARY NOTE

$$P(1) = P(2) = \sum_{k=N_{\text{th}}}^{N_{\text{half}}} \binom{N_{\text{half}}}{k} p_0^k (1-p_0)^{N_{\text{half}}-k} \quad (71)$$

where N_{half} denotes the number of probes with one color (half of a set), and N_{th} is the detection threshold for individual mRNA molecules. Based on our threshold setting (**Supplementary Fig. 4a**), N_{th} roughly corresponds to half of the value of the positive smFISH peak, or $N_{\text{all}} \cdot p_0/2$. By applying Equation (71), we obtained p_0 = 0.09 ± 0.01 (**Supplementary Fig. 4e**, data averaged from 4 embryos).

To summarize, all four methods above give similar estimates of the probe binding probability, p_0 ~10% (**Supplementary Fig. 4e**). We next used this value to estimate the mRNA detection efficiency, P_{RNA}. Assuming binomial statistics of probe binding, we obtain

$$P_{\text{RNA}} = \sum_{k=N_{\text{th}}}^{N_{\text{set}}} \binom{N_{\text{set}}}{k} p_0^k (1-p_0)^{N_{\text{set}}-k} \quad (72)$$

where N_{set} is the number of probes in the full set. As mentioned above (method #4), $N_{\text{th}} \approx N_{\text{all}} \cdot p_0/2$. This yields a detection efficiency of ~80% for *hb* mRNA (0.87 ± 0.02 for method 1, 0.76 ± 0.11 for method #2, 0.80 ± 0.02 for method #3, 0.81 ± 0.03 for method #4; **Supplementary Fig. 4f**).

The above estimations are likely to be lower bounds rather than the actual values. This is due to the following reasons: (1) In method #1, we assumed that false-positive spots correspond to individual smFISH probes. However, it is possible that these spots correspond to multiple smFISH probes bound together. In addition, there is no guarantee that individual probes can be detected under the confocal imaging, therefore the weakest detectable spots may again correspond to multiple probes. As a result, we may have overestimated the intensity of a single smFISH probe, and underestimated p_0. (2) In method #2, we assumed that the half-width w_0 of the intensity distribution for positive mRNA spots arises solely from the stochasticity of probe binding. However, additional factors likely contribute to the variability in the fluorescence signal, for example the photon shot noise of individual smFISH probes (discussed in **Supplementary Note Section 7.4**). Thus, we may have overestimated the variance caused by stochastic probe binding, and underestimated p_0. (3) Similarly, the fluctuation in single-probe intensity will also affects method #3, where the correlation coefficient is defined as the ratio between covariance and variance. Since different fluorophores emit independently, the fluctuations in single-probe fluorescence (for each channel) contribute to the variance but not the covariance of the spot intensity distribution. Thus, we may have underestimated the correlation coefficient and p_0. (4) In method #4, we again did not consider the fluctuations in single-probe intensity in Equation (72), which lower our estimation of $P(1|2)$ and p_0.

A previous study (Little et al., 2013) reported that smFISH labeling of *hb* using 114 probes yielded an estimated mRNA detection efficiency of ~94%. Applying Equation (72) above using this number of probes suggests a probe binding probability of $p_0 \approx 0.1$. Thus, we estimate that our probe binding probability is comparable to that of (Little et al., 2013). The smaller number of smFISH probes used in the current work (48 versus 114) reflects our choice to target only the coding region of the *hb* gene and to maintain optimal probe design parameters (Raj et al., 2008).

8.2 Estimating the error in nascent mRNA quantification

SUPPLEMENTARY NOTE

To estimate the error in nascent mRNA quantification due to finite probe binding probability, we assume binomial distribution for all of the probe binding sites on nascent mRNAs at a given gene locus. Thus,

$$\begin{aligned} n_{\mathrm{nas}} &= N_{\mathrm{nas}} \cdot N_{\mathrm{set}} \cdot p_0 \\ \sigma_{\mathrm{nas}} &= \sqrt{N_{\mathrm{nas}} \cdot N_{\mathrm{set}} \cdot p_0 \cdot (1-p_0)} \end{aligned} \tag{73}$$

$$\frac{\sigma_{\mathrm{nas}}}{n_{\mathrm{nas}}} = \sqrt{\frac{(1-p_0)}{N_{\mathrm{nas}} \cdot N_{\mathrm{set}} \cdot p_0}} \tag{74}$$

where n_{nas} and σ_{nas} are the mean and standard deviation of the number of bound probes, respectively, N_{nas} is the number of nascent mRNAs, N_{set} is the total number of probes in the probe set, and p_0 is the probe binding probability. Following our measurement of N_{nas}, we applied Equation (74) to each active *hb* locus within a single embryo, and calculated the average error in quantifying nascent mRNA, $\langle \sigma_{nas}/n_{nas} \rangle$, to be ~10% for cycles 11-13 embryos, and ~13% for cycle 14 embryos (**Supplementary Fig. 4g**).

8.3 Estimating the error in estimating model parameters

When modeling *hb* transcription kinetics, errors in mRNA and protein quantification are expected to propagate into our estimation of kinetic parameters of the two-state model. Specifically, we consider below the effect of three types of errors: (1) The error in quantifying nascent mRNA; (2) The error in identifying active transcription sites; and (3) The error in quantifying nuclear Bcd concentration.

To characterize how the errors above propagate into our estimation of model parameters, we simulated our experimental data as follows. We created a "virtual embryo" containing 700 nuclei, evenly distributed on the AP axis (x = 0.25-0.75 EL). Bcd concentration in the nuclei follows an exponential gradient, $C(x) = C_{\max} e^{-x/d}$ with $C_{\max}$ = 75 nM and d = 0.225 EL (see **Fig. SN3**). To simulate *hb* transcription, we created two *hb* gene loci in each nucleus, and applied a Gillespie algorithm of the two-state model with the following parameters: T_{EL} = 2.5 min, k_{OFF} = 0.9 min^{-1}, k_{INI} = 32 min^{-1}, $k_{\mathrm{ON}} = k_1 C^h / (C^h + C_{50}{}^h) + k_0$, k_1 = 1.2 min^{-1}, k_0 = 0.05 min^{-1}, h = 6, C_0 = 10 nM (see **Supplementary Fig. 12**). After 5 minutes of transcription, we recorded the number of nascent mRNA molecules at each gene locus. We then applied different type of quantification errors to this data set.

As discussed in **Supplementary Note Section 8.2**, the finite binding probability by smFISH probes leads to errors in quantifying nascent mRNA. To test how this error propagates into our model of *hb* transcription kinetics, we simulated the smFISH labeling and quantification processes. First, we multiplied the number of nascent mRNAs at each *hb* locus by 48, to calculate the number of probe binding sites. We then applied a binomial statistics with p_0 = 0.1 to estimate the number of bound probes. Following our method of quantifying the smFISH signal (see **Supplementary Note Section 2.2**), we used the typical signal of a single mRNA (Np_0, N = 48) to calibrate the measured nascent mRNA level, and then filtered the value using a threshold of N_{th} = 3 mRNAs (see **Supplementary Note Section 2.2**) to distinguish active from inactive loci. With this new set of single-locus data, we then performed parameter estimation for the two-state model (see **Supplementary Note Section 6.4**), and compared the resulting parameter values (h, C_{50}, k_1, k_0, k_{OFF}, k_{INI}) with the original setting (see the preceding paragraph). We found that for all parameters (except for k_0, whose absolute value is small), the deviation from the original setting was ≤ 15% (**Fig. SN10a**).

SUPPLEMENTARY NOTE

Another source of error we considered is due to the specific choice of intensity threshold for detecting active transcription sites (N_{th}, see **Supplementary Note Section 2.2**). The value of 3 mRNAs was chosen to avoid false classification of two closely-spaced individual mRNAs as a transcription site. We varied the value of N_{th} between 1 and 4 mRNAs and performed the same fitting procedure as described in the preceding paragraph. We found that among all kinetic parameters (except for k_0, whose absolute value is small), k_1 and k_{OFF} were the main species that vary with N_{th}. For both of them, the deviation from the original setting was less than 20% for $N_{th} \leq 4$ (**Fig. SN11a**).

Finally, we estimated the effect of errors in quantifying nuclear Bcd concentration. These in turn are due to two factors: (1) the inaccuracy in estimating the single-Bcd fluorescence (see **Supplementary Note Section 3**), and (2) fluctuations of antibody labeling, e.g. heterogeneity in the number of antibodies labeling each Bcd molecule. The former factor only leads to a scaling factor in Bcd quantification, while not affecting the distribution of nascent mRNAs. Thus, it only affects h and C_{50}, but not the other k's. Here we mainly consider the latter factor, which results in an error between Bcd concentration and the immunofluorescence signal. In analyzing *hb* statistics, nuclei may be assigned to the wrong Bcd bin, affecting the two-state fitting procedure. To quantify this effect, we added to our simulation a Bcd measurement step, in which we applied a Gaussian error to the measurement. As a simple case, we assumed that the noise amplitude, σ, is proportional to the real Bcd concentration, C, of each nucleus (e.g. the relative noise amplitude, σ/C, is a constant). We next repeated the fitting procedure as above, and binned the data using the "measured" Bcd concentration. Based on our measurement of single-Bcd fluorescence intensity (**Fig. 1c**), the relative noise amplitude for single-Bcd quantification is ~30%, resulting in an even smaller value for the error in estimating Bcd concentration. By trying different values of σ/C, we found that the deviation of the kinetic parameters from the original setting was ≤ 20% for $\sigma/C \leq 20\%$. (**Fig. SN10b**).

SUPPLEMENTARY NOTE

REFERENCES

Abu-Arish, A., Porcher, A., Czerwonka, A., Dostatni, N., and Fradin, C. (2010). High mobility of bicoid captured by fluorescence correlation spectroscopy: implication for the rapid establishment of its gradient. Biophysical journal *99*, L33-35.

Boettiger, A.N., and Levine, M. (2013). Rapid transcription fosters coordinate snail expression in the Drosophila embryo. Cell reports *3*, 8-15.

Burz, D.S., Rivera-Pomar, R., Jackle, H., and Hanes, S.D. (1998). Cooperative DNA-binding by Bicoid provides a mechanism for threshold-dependent gene activation in the Drosophila embryo. The EMBO journal *17*, 5998-6009.

Driever, W., and Nusslein-Volhard, C. (1988). A gradient of bicoid protein in Drosophila embryos. Cell *54*, 83-93.

Dunn, K.W., and Wang, E. (2000). Optical aberrations and objective choice in multicolor confocal microscopy. BioTechniques *28*, 542-544, 546, 548-550.

Edgar, B.A., Kiehle, C.P., and Schubiger, G. (1986). Cell cycle control by the nucleo-cytoplasmic ratio in early Drosophila development. Cell *44*, 365-372.

Evans, M., Hastings, N.A.J., and Peacock, J.B. (2000). Statistical distributions, 3rd edn (New York: Wiley).

Femino, A.M., Fay, F.S., Fogarty, K., and Singer, R.H. (1998). Visualization of single RNA transcripts in situ. Science *280*, 585-590.

Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E.M. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. Nature *advance online publication*.

Gillespie, D.T. (1977). Exact stochastic simulation of coupled chemical reactions. The journal of physical chemistry *81*, 2340-2361.

Gregor, T., Tank, D.W., Wieschaus, E.F., and Bialek, W. (2007a). Probing the limits to positional information. Cell *130*, 153-164.

Gregor, T., Wieschaus, E.F., McGregor, A.P., Bialek, W., and Tank, D.W. (2007b). Stability and nuclear dynamics of the bicoid morphogen gradient. Cell *130*, 141-152.

He, F., Ren, J., Wang, W., and Ma, J. (2011). A multiscale investigation of bicoid-dependent transcriptional events in Drosophila embryos. PloS one *6*, e19122.

Houchmandzadeh, B., Wieschaus, E., and Leibler, S. (2002). Establishment of developmental precision and proportions in the early Drosophila embryo. Nature *415*, 798-802.

Janody, F., Sturny, R., Catala, F., Desplan, C., and Dostatni, N. (2000). Phosphorylation of bicoid on MAP-kinase sites: contribution to its interaction with the torso pathway. Development *127*, 279-289.

Li, X.Y., MacArthur, S., Bourgon, R., Nix, D., Pollard, D.A., Iyer, V.N., Hechmer, A., Simirenko, L., Stapleton, M., Luengo Hendriks, C.L.*, et al.* (2008). Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. PLoS biology *6*, e27.

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

SUPPLEMENTARY NOTE

Little, S.C., Tkacik, G., Kneeland, T.B., Wieschaus, E.F., and Gregor, T. (2011). The formation of the Bicoid morphogen gradient requires protein movement from anteriorly localized mRNA. PLoS biology *9*, e1000596.

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. Current biology : CB *23*, 2135-2139.

Ma, X., Yuan, D., Diepold, K., Scarborough, T., and Ma, J. (1996). The Drosophila morphogenetic protein Bicoid binds DNA cooperatively. Development *122*, 1195-1206.

MacArthur, S., Li, X.Y., Li, J., Brown, J.B., Chu, H.C., Zeng, L., Grondona, B.P., Hechmer, A., Simirenko, L., Keranen, S.V.*, et al.* (2009). Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. Genome biology *10*, R80.

Margolis, J.S., Borowsky, M.L., Steingrimsson, E., Shim, C.W., Lengyel, J.A., and Posakony, J.W. (1995). Posterior stripe expression of hunchback is driven from two promoters by a common enhancer element. Development *121*, 3067-3077.

Massey Jr, F.J. (1951). The Kolmogorov-Smirnov test for goodness of fit. Journal of the American statistical Association *46*, 68-78.

McHale, P., Mizutani, C.M., Kosman, D., MacKay, D.L., Belu, M., Hermann, A., McGinnis, W., Bier, E., and Hwa, T. (2011). Gene length may contribute to graded transcriptional responses in the Drosophila embryo. Developmental biology *360*, 230-240.

McNally, J.G., Muller, W.G., Walker, D., Wolford, R., and Hager, G.L. (2000). The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. Science *287*, 1262-1265.

Munsky, B., and Khammash, M. (2008). The finite state projection approach for the analysis of stochastic noise in gene networks. Automatic Control, IEEE Transactions on *53*, 201-214.

Neuert, G., Munsky, B., Tan, R.Z., Teytelman, L., Khammash, M., and van Oudenaarden, A. (2013). Systematic identification of signal-activated stochastic gene regulation. Science *339*, 584-587.

Pawley, J.B. (2006). Handbook of biological confocal microscopy, 3rd edn (New York, NY: Springer).

Peccoud, J., and Ycart, B. (1995). Markovian modeling of gene-product synthesis. Theoretical population biology *48*, 222-234.

Perry, M.W., Bothma, J.P., Luu, R.D., and Levine, M. (2012). Precision of hunchback expression in the Drosophila embryo. Current biology : CB *22*, 2247-2252.

Poorey, K., Viswanathan, R., Carver, M.N., Karpova, T.S., Cirimotich, S.M., McNally, J.G., Bekiranov, S., and Auble, D.T. (2013). Measuring chromatin interaction dynamics on the second time scale at single-copy genes. Science *342*, 369-372.

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

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nature methods *5*, 877-879.

Ronchi, E., Treisman, J., Dostatni, N., Struhl, G., and Desplan, C. (1993). Down-regulation of the Drosophila morphogen bicoid by the torso receptor-mediated signal transduction cascade. Cell *74*, 347-355.

SUPPLEMENTARY NOTE

Schroder, C., Tautz, D., Seifert, E., and Jackle, H. (1988). Differential regulation of the two transcripts from the Drosophila gap segmentation gene hunchback. The EMBO journal *7*, 2881-2887.

Skinner, S.O., Sepulveda, L.A., Xu, H., and Golding, I. (2013). Measuring mRNA copy number in individual Escherichia coli cells using single-molecule fluorescent in situ hybridization. Nature protocols *8*, 1100-1113.

Struhl, G., Struhl, K., and Macdonald, P.M. (1989). The gradient morphogen bicoid is a concentration-dependent transcriptional activator. Cell *57*, 1259-1273.

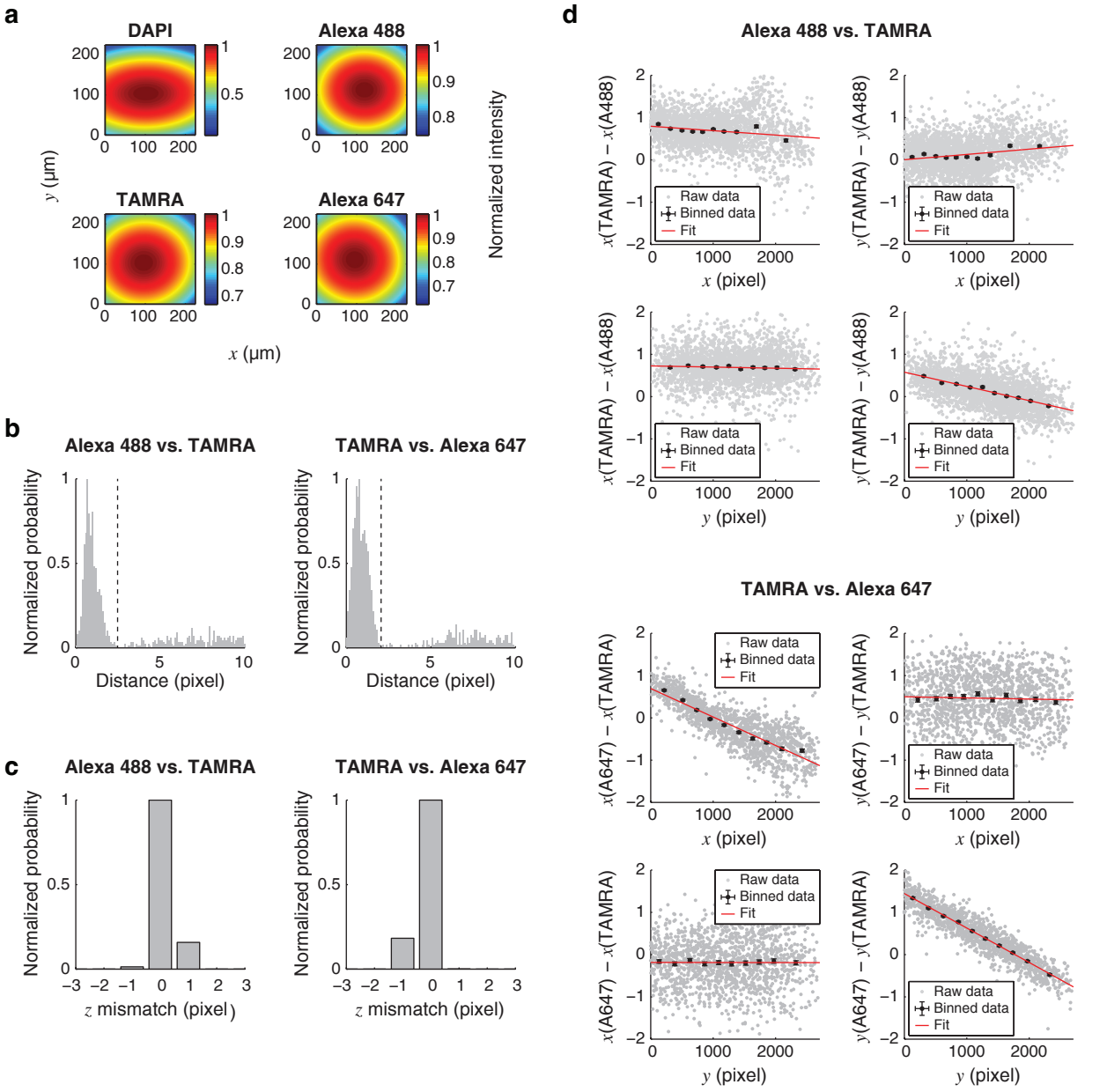
Tkacik, G., Gregor, T., and Bialek, W. (2008). The role of input noise in transcriptional regulation. PloS one *3*, e2774.

Zenklusen, D., Larson, D.R., and Singer, R.H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. Nature structural & molecular biology *15*, 1263-1271.

Zhang, B., Zerubia, J., and Olivo-Marin, J.-C. (2007). Gaussian approximations of fluorescence microscope point-spread function models. Applied Optics *46*, 1819-1829.

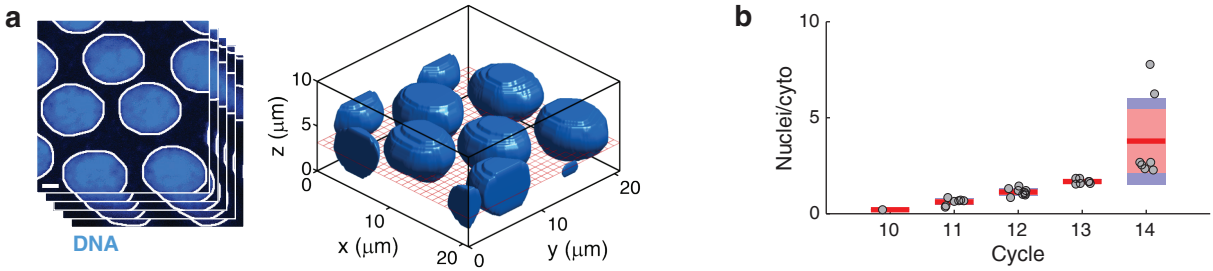
SUPPLEMENTARY NOTE

Figure SN1. Measuring optical aberrations.



(**a**) Intensity profiles for samples of uniformly mixed fluorophores, using a 63× oil immersion objective. Each profile was obtained by averaging measurements from 6 different areas of the sample, with each measurement first normalized by the maximal intensity. (**b**) The distribution of distance between smFISH spots from different channels. Left: Alexa 488 vs. TAMRA, data from >20,000 spots in a single cycle 12 embryo. Right: TAMRA vs. Alexa 647, data from >20,000 spots in a single cycle 12 embryo. Two groups of spot pairs were recognized from the distribution. A distance threshold was used to identify inter-channel spot pairs corresponding to the same mRNA. (**c**) The distribution of spatial mismatch in the z dimension for inter-channel spot pairs corresponding to the same mRNA (the same data set as in Panel **b**). (**d**) The relation between the inter-channel spatial mismatch and spot position in x and y dimensions. Upper four plots: Alexa 488 vs. TAMRA, data from >20,000 spots in a single cycle 12 embryo. Lower four plots: TAMRA vs. Alexa 647, data from >20,000 spots in a single cycle 12 embryo. Data from individual spot pairs (gray) was binned along the x or y axis (black, mean ± S.E.M.) and fitted to a linear function (red).

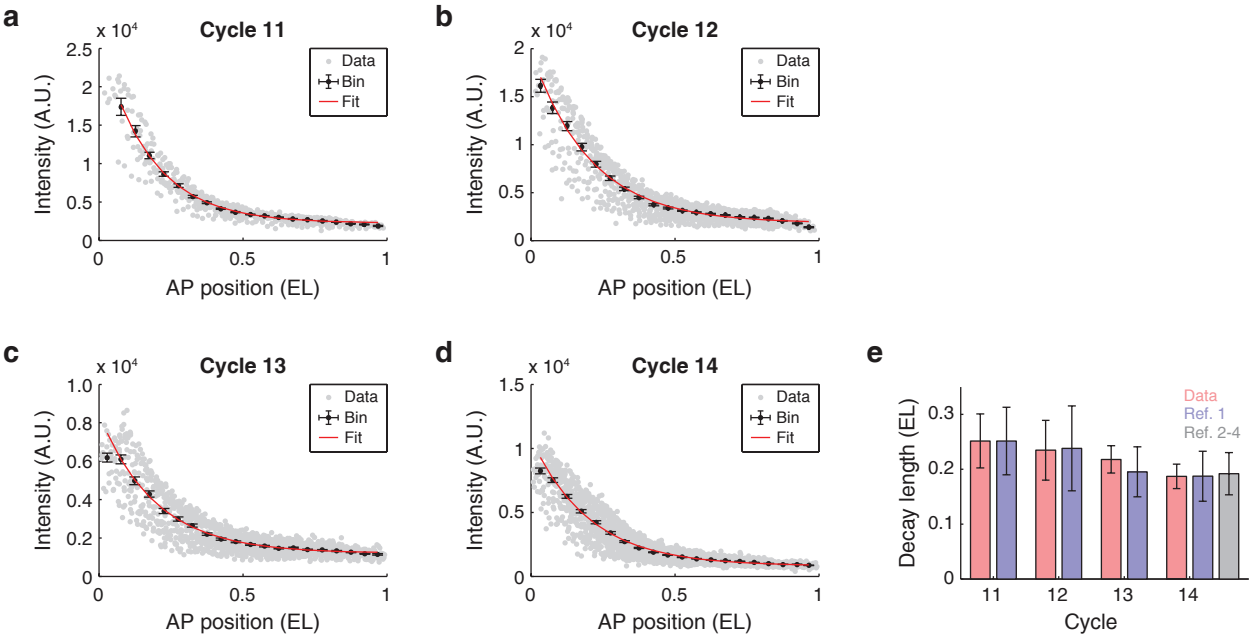
Figure SN2. The nuclear segmentation.



(**a**) Automated segmentation of nuclei. Left: At each image plane, the DNA signal was used to identify nuclear boundaries (white). Scale bar, 2 μm. Right: The nuclear boundaries from different image planes were stitched together to reconstruct each nucleus in 3D. (**b**) Comparison of the nuclear-to-cytoplasmic area ratio between different nuclear cycles. Data from 31 embryos, cycles 10-14. Gray, individual embryos. Red, mean value. Pink area, 95% confidence interval. Blue area, ±1 standard deviation.

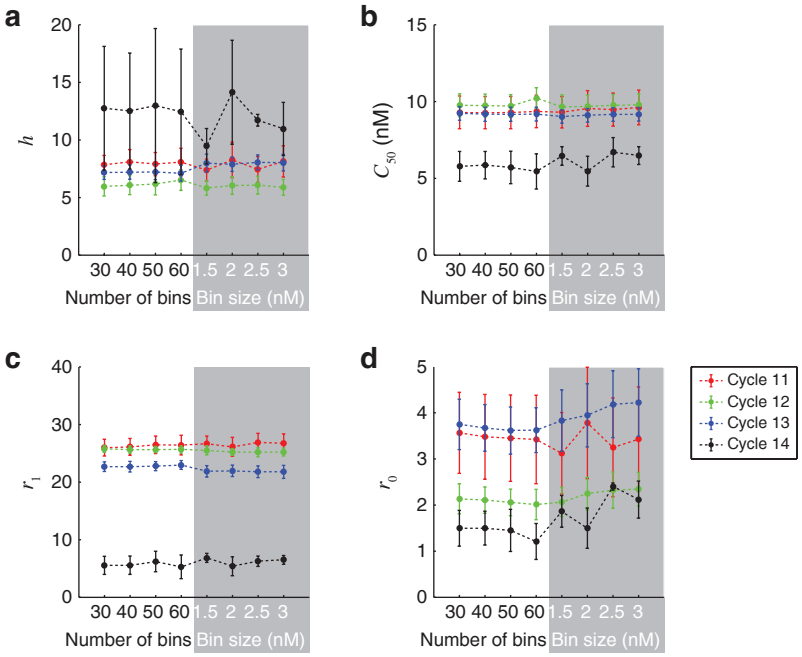
SUPPLEMENTARY NOTE

Figure SN3. The spatial profile of Bcd immunofluorescence.



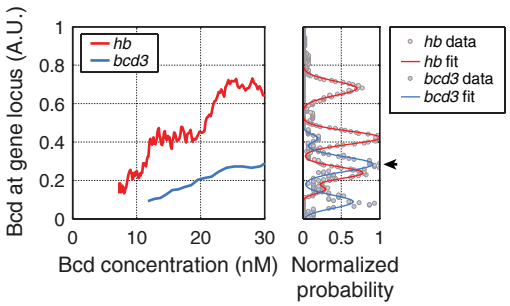
(**a**) – (**d**) The exponential gradient of nuclear Bcd immunofluorescence along the AP axis of individual wild-type embryos during cycles 11-14. Data from individual nuclei (gray) was binned along the AP axis (black, mean ± S.E.M.) and fitted to an exponential function (red). (**e**) A comparison of the exponential decay length between different cycles. Our data (red; cycle 11, 7 embryos; cycle 12, 9 embryos; cycle 13, 7 embryos; cycle 14, 7 embryos) shows good agreement with the literature (Blue: [1] Abu-Arish *et al*. 2007. Gray: [2] Houchmandzadeh *et al*. 2002; [3] Gregor *et al*. 2007; [4] Little *et al*. 2011). Error bars indicate S.E.M.

Figure SN4. The robustness of fitting the gene regulation function.



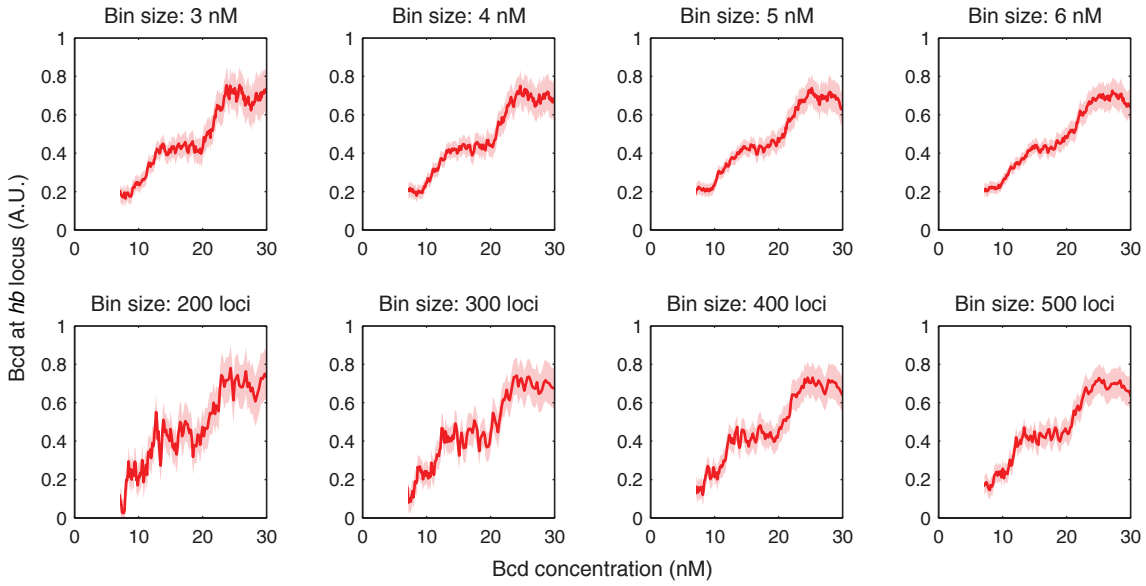
(**a**) – (**d**) Comparing parameters extracted from the gene regulation function using different binning criteria. The x axis of each plot discriminates equal-population binning (left) and equal-concentration-interval binning (right). Data is from the *hb* gene in wild-type embryos (cycle 11, 7 embryos; cycle 12, 9 embryos; cycle 13, 7 embryos; cycle 14, 7 embryos). Error bars represent S.E.M.

Figure SN5. Extracting the plateau levels of Bcd enrichment.



Left: The Bcd binding curves for the *hb* gene (red, data from 9 wild-type embryos, cycle 12) and the *bcd3-lacZ* reporter gene (blue, data from 23 transgenic embryos, cycles 11-14). Right: The distributions of Bcd enrichment values from all data points on each binding curve (*hb*, red dots; *bcd3-lacZ*, blue dots). Each distribution was fitted to a sum of Gaussian functions (lines) to extract the mean and width of each enrichment plateau. The plateau of Bcd enrichment at *bcd3-lacZ*, measured at high Bcd concentration (arrow), was used to rescale Bcd enrichment at other genes.

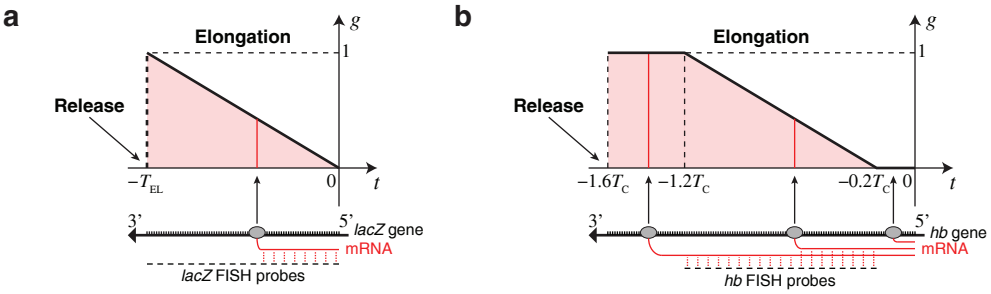
Figure SN6. The robustness of measuring the Bcd binding curve.



Comparing the Bcd binding curve at the *hb* gene using different binning criteria (data from 9 embryos, cycle 12). Top, equal-population binning. Bottom, equal-concentration-interval binning. Lines, mean values. Shading, S.E.M.

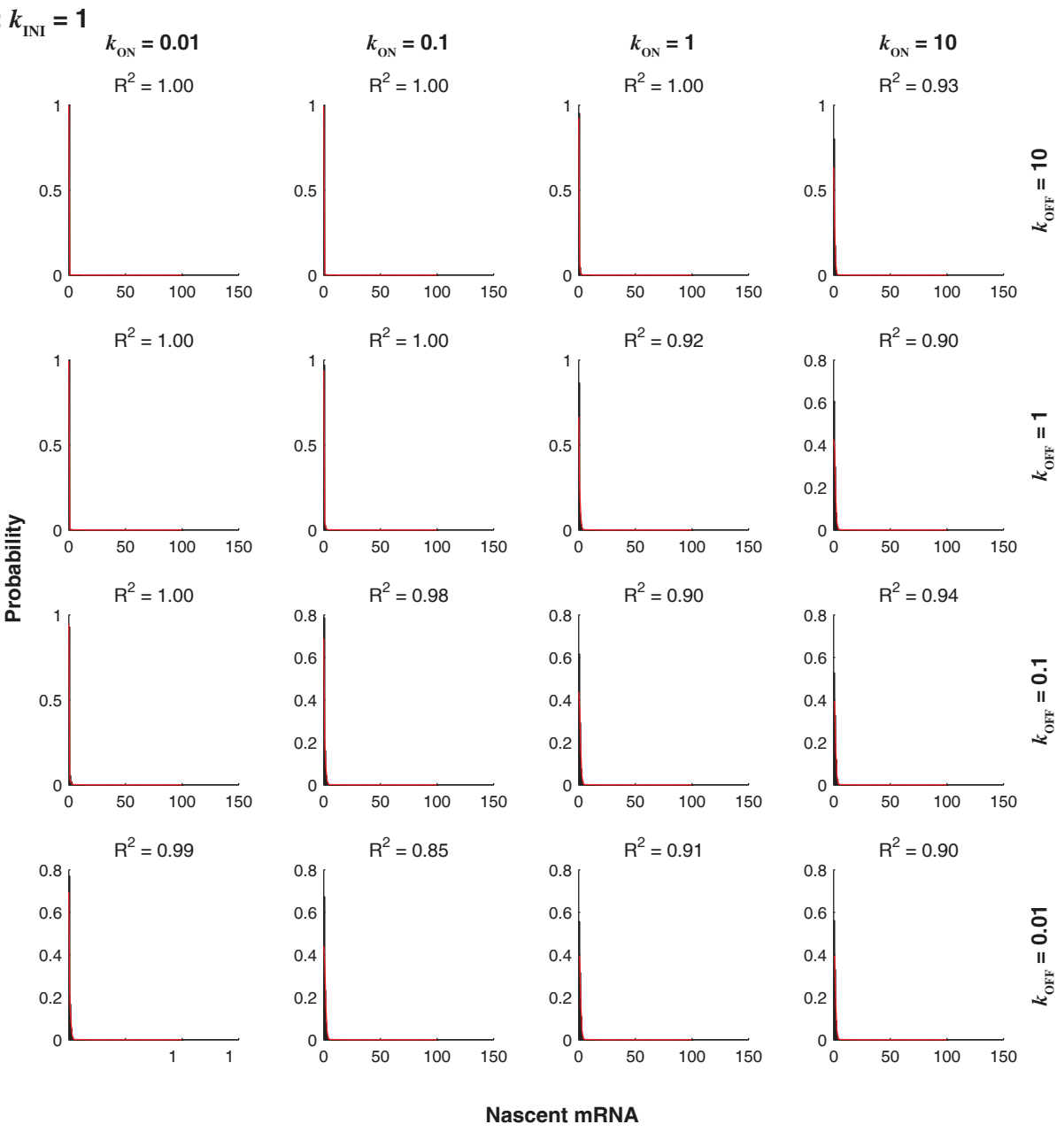
SUPPLEMENTARY NOTE

Figure SN7. The contribution function for modeling the distribution of nascent mRNA numbers.

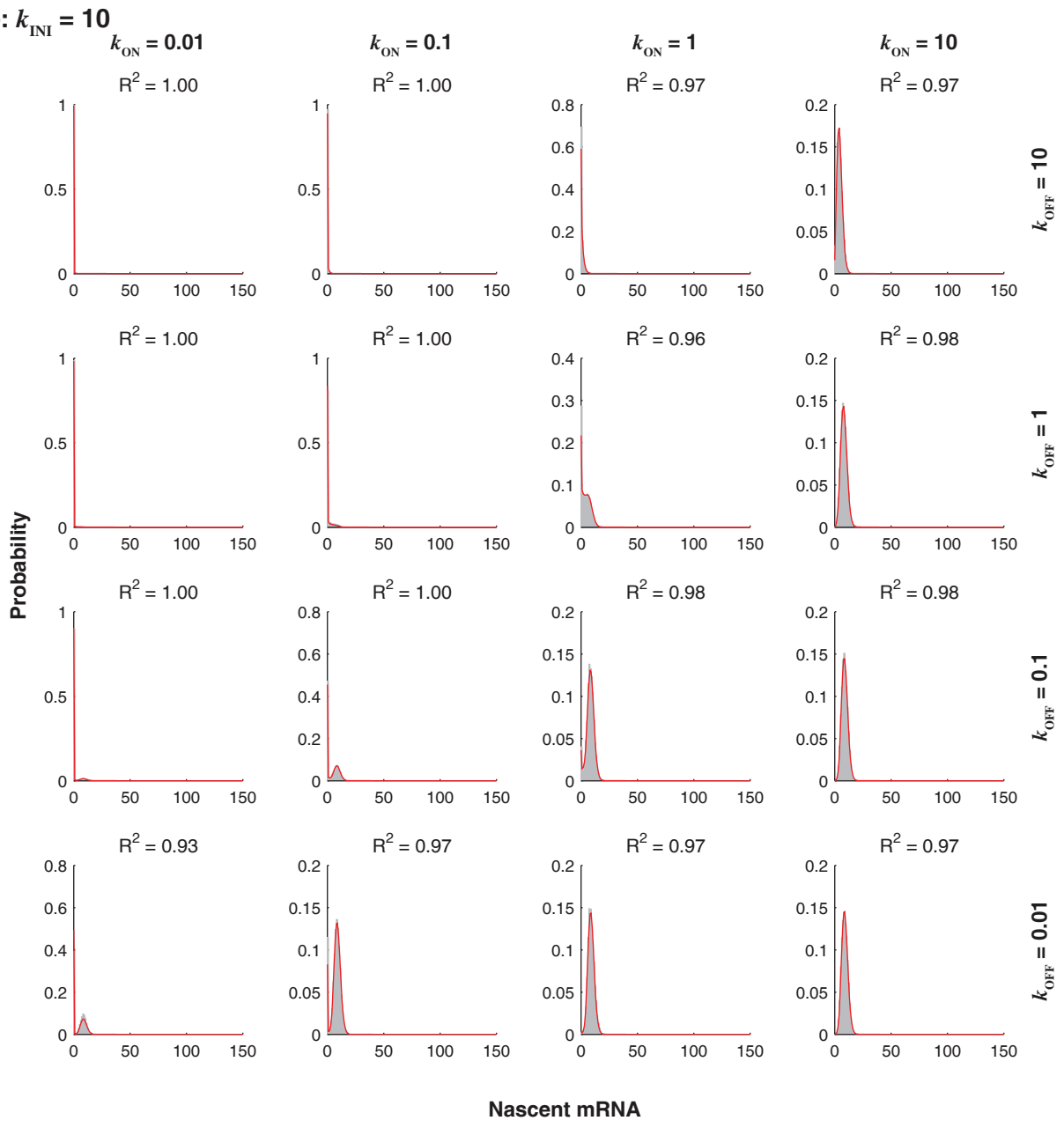


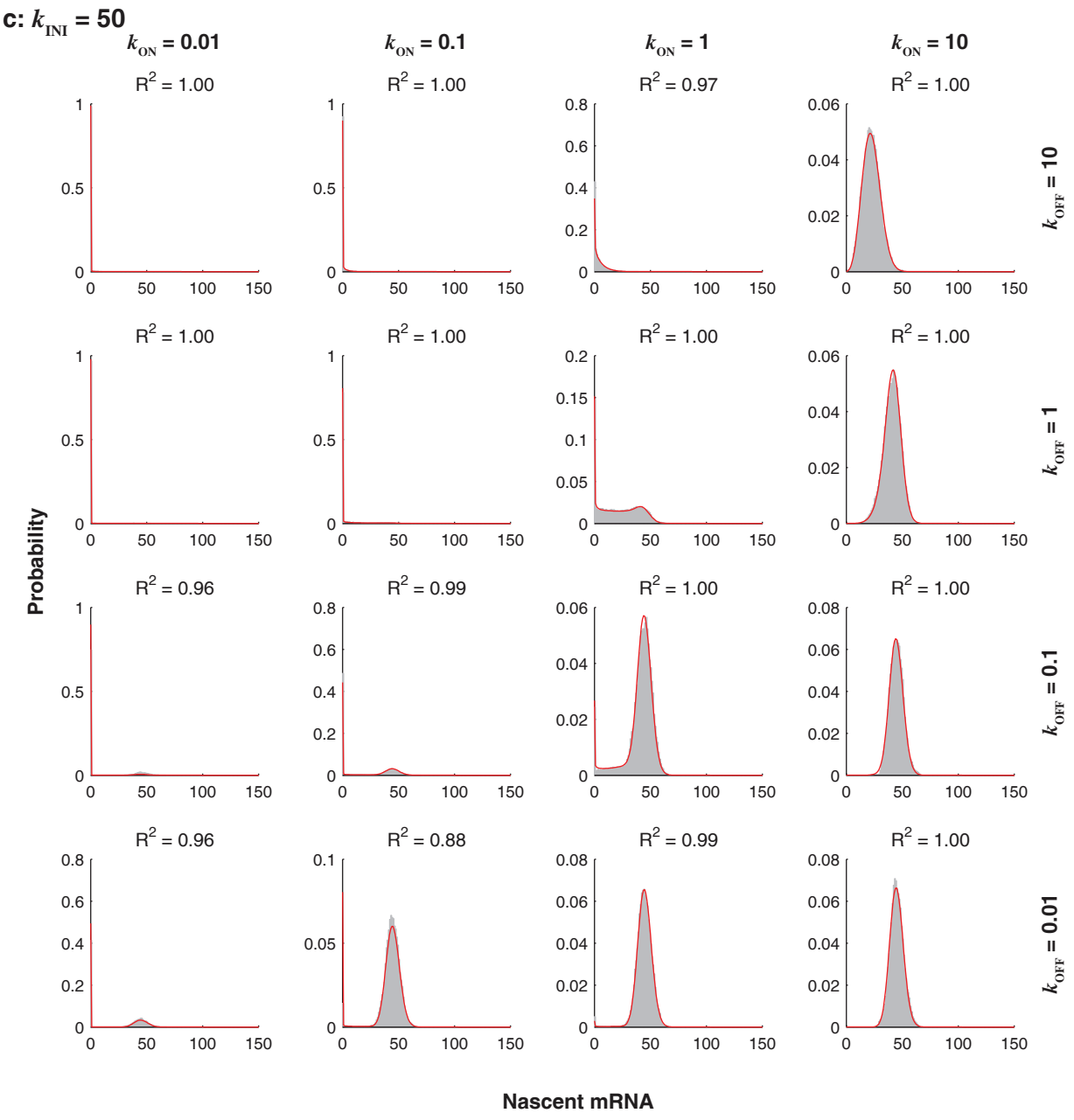
(**a**) The contribution function of the *lacZ* gene. The observed smFISH signal for a nascent transcript is plotted as a function of the initiation time of the transcript. The signal, represented as the number of bound smFISH probes, is proportional to the length of the nascent transcript (bottom schematic). (**b**) The contribution function of the *hb* gene. Because the smFISH probes only cover the coding region of the gene, the observed smFISH signal of a nascent *hb* transcript is a piecewise-linear function of the initiation time, which only varies with the initiation time in a particular time window.

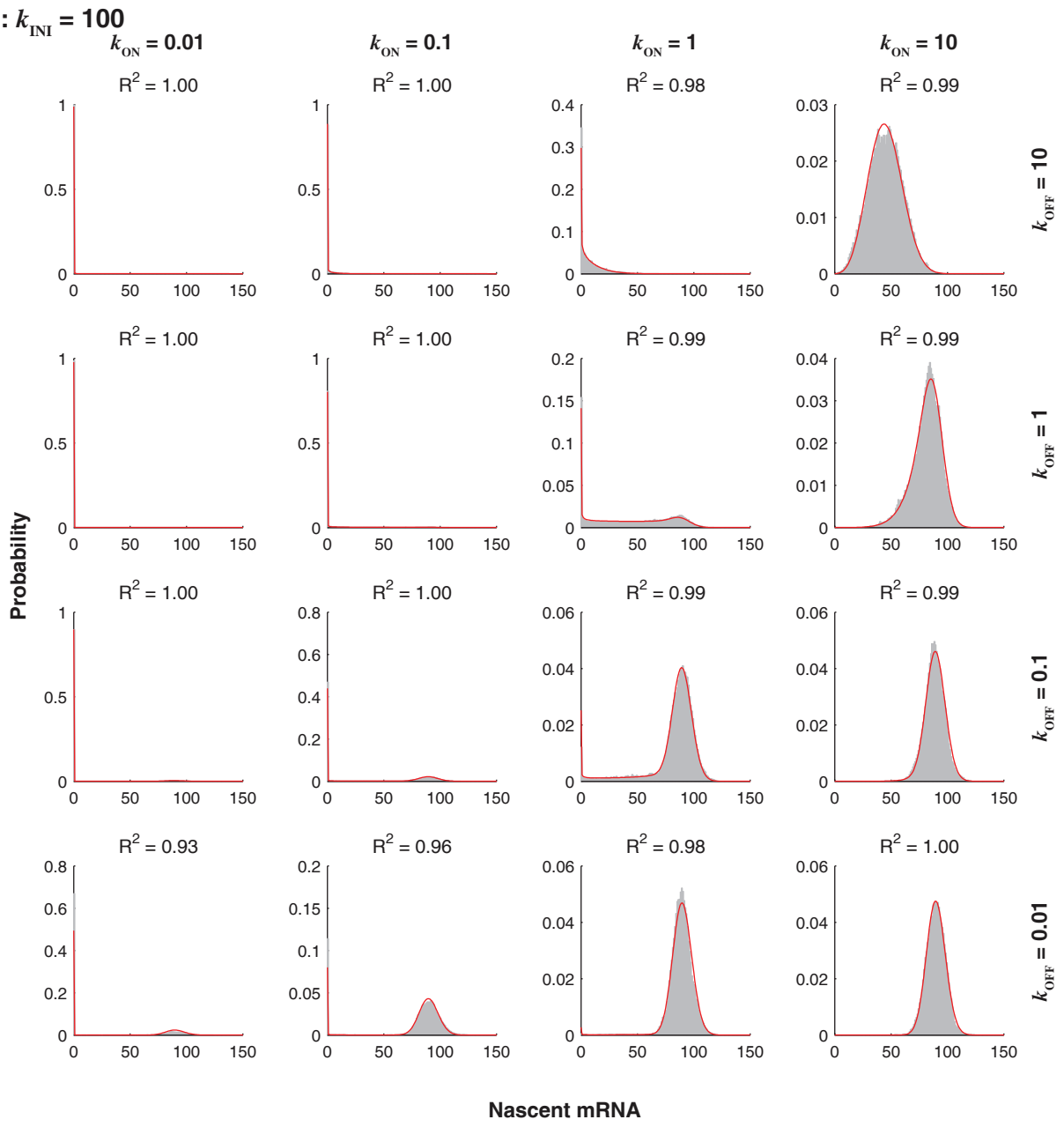
Figure SN8. Validating the accuracy of the finite state projection method.



SUPPLEMENTARY NOTE

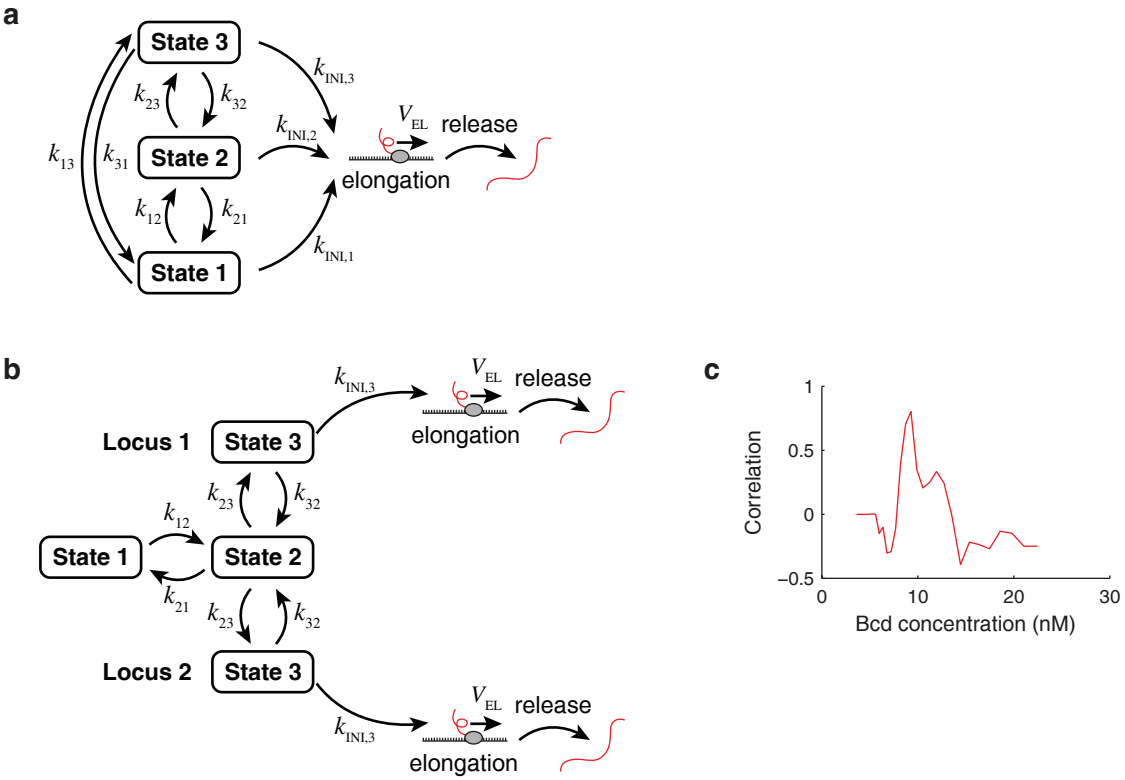



c: k_{INI} = 50
k_{ON} = 0.01
k_{ON} = 0.1
k_{ON} = 1
k_{ON} = 10
k_{OFF} = 10
k_{OFF} = 1
k_{OFF} = 0.1
k_{OFF} = 0.01
R² = 1.00
R² = 1.00
R² = 0.97
R² = 1.00
R² = 1.00
R² = 1.00
R² = 1.00
R² = 1.00
R² = 0.96
R² = 0.99
R² = 1.00
R² = 1.00
R² = 0.96
R² = 0.88
R² = 0.99
R² = 1.00
Probability
Nascent mRNA



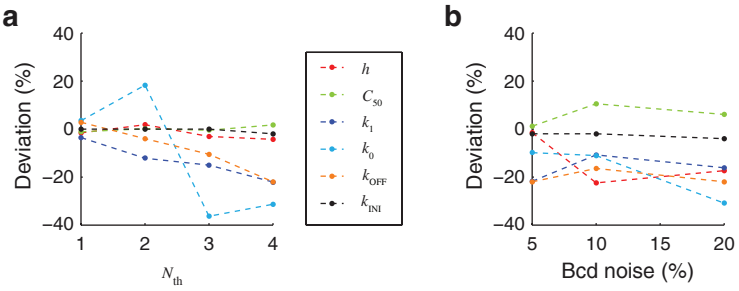
(**a**) – (**d**) A comparison of the distribution of number of nascent mRNAs obtained using the Gillespie simulation (gray) with the results of the finite state projection method (red). The two computational methods agree well over the full range of kinetic parameters tested.

Figure SN9. The three-state model.



(**a**) A general three-state model of transcription. The gene stochastically switches between three different states, with transition rates k_{ij} $(i, j = 1, 2, 3)$. In each state, transcription is stochastically initiated at rate $k_{\mathrm{INI},i}$ (i = 1, 2, 3). Transcript elongation occurs deterministically with speed V_{EL}. Once elongation is completed, the transcript is released from the gene locus. (**b**) Modeling two *hb* alleles in the same nucleus using a correlated three-state model. State 1 and state 2 represent "nuclear" states that are shared by the two alleles. Both states are silent. State 3 represents an active gene state of individual alleles. (**c**) The calculated correlation coefficient between two *hb* loci in the same nucleus, as a function of nuclear Bcd concentration. The model described in Panel **b** was simulated. The simulation agrees qualitatively with the experimental data (**Supplementary Figure 8c**).

Figure SN10. Error estimation for the two-state model parameters.



(**a**) Estimated error (deviation from the pre-set value) in the extracted kinetic parameters as a function of the threshold for detecting active transcription sites, N_{th} (data from a single virtual embryo, 700 nuclei, AP positions 0.25-0.7 EL, no Bcd quantification noise). (**b**) Estimated error (deviation from the pre-set value) in the extracted kinetic parameters as a function of the noise in Bcd quantification (data from a single virtual embryo, 700 nuclei, AP positions 0.25-0.7 EL, N_{th} = 3).