Differential regulation of alternative promoters emerges from unified kinetics of enhancer-promoter interaction

Jingyao Wang[1,2], Shihe Zhang[2,3,*], Hongfang Lu[1,2], and Heng Xu[2,3,*]

[1]School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

[2]Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai 200240, China

[3]School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China

*Correspondence: hsx_sjtu@sjtu.edu.cn (S.Z.), Heng_Xu@sjtu.edu.cn (H.X.)

Abstract

Many eukaryotic genes contain alternative promoters with distinct expression patterns. How these promoters are differentially regulated remains elusive. Here, we apply single-molecule imaging to quantify the transcriptional regulation of two alternative promoters (P1 and P2) of the Bicoid (Bcd) target gene *hunchback* in syncytial blastoderm *Drosophila* embryos. Contrary to the previous notion that Bcd only activates P2, we find that Bcd activates both promoters via the same two enhancers. P1 activation is less frequent and requires binding of more Bcd molecules than P2 activation. Using a theoretical model to relate promoter activity to enhancer states, we show that the two promoters follow common transcription kinetics driven by sequential Bcd binding at the two enhancers. Bcd binding at either enhancer primarily activates P2, while P1 activation relies more on Bcd binding at both enhancers. These results provide a quantitative framework for understanding the dynamics of complex eukaryotic gene regulation.

Introduction

Promoters initiate gene transcription by interacting with specific *cis*-regulatory sequences (enhancers)[1-3]. In eukaryotic genomes, many genes contain alternative promoters[4,5], which can produce functionally distinct transcript isoforms under the regulation of multiple enhancers[6,7]. During development, the differential expression of these isoforms is critical to cell fate decisions[8,9]. Misregulation of these isoforms can lead to diseases, including cancer[6,10-12]. Thus, eukaryotic gene regulation needs to be understood at the level of individual alternative promoters and enhancers.

Enhancer activation of a single promoter involves a series of molecular events, including transcription factor (TF) binding, local chromatin opening, and physical proximity between the two elements[1,2,13]. Typically, TF binding determines the spatial expression pattern[1,14], whereas other events set the expression amplitude[15-18]. When multiple promoters and enhancers are present, interactions between these elements often result in complex and varied promoter behaviors[19]. For example, multiple enhancers may contact each other and synergistically drive a promoter[20,21]. Alternatively, different enhancers may be mutually exclusive and compete for promoter activation[22,23]. In these cases, enhancers combine their regulatory effects differently, ranging from superadditive to subadditive or even repressive summation[12]. Similarly, a single enhancer may activate multiple promoters simultaneously[24,25] or one at a time[26]. So far, although specific mechanisms were proposed for some of these phenomena, it is unclear whether universal mechanisms exist for the complex interactions between multiple enhancers and promoters.

An ideal model to investigate alternative promoter regulation is the *Drosophila* gap gene *hunchback* (*hb*), which contains two alternative promoters (P1 and P2, Fig. 1a)[27-30]. In early embryogenesis (syncytial blastoderm stage, nuclear division cycles (nc) 10–13), *hb* is expressed in a bursty manner[31,32] throughout the anterior half of the embryo in response to the concentration gradient of the maternal TF Bicoid (Bcd)[29,33,34]. This expression is believed to be purely from the P2 promoter mediated by two enhancers: a proximal enhancer located next to P2[29,34] and a distal "shadow" enhancer located 4.5 kb upstream[35]. Cooperative Bcd binding at either enhancer can activate P2[35]. Competitive action of these partially redundant enhancers helps suppress expression noise and ensure a robust expression pattern[23,35].

Unlike P2, the P1 promoter is believed to be inactive during early development[36,37], due to its lack of Zelda binding sites and TATA box necessary for local chromatin opening[30,38]. Instead, the promoter is activated in late nc14 by a Bcd-independent stripe enhancer[28,36]. However, previous measurements of endogenous P1 activity relied on traditional *in situ* hybridization methods, which have limited sensitivity to detect weak mRNA signals. In fact, *hb*-reporter experiments showed that P1 could respond to Bcd when placed adjacent to the proximal or distal enhancers[30]. Without precise quantification of endogenous P1 activity, the understanding of early *hb* regulation is incomplete. It is unclear how P1 interacts with different *hb* enhancers and whether the mechanisms of P1 and P2 regulation are intrinsically related.

Here, we used single-molecule fluorescence *in situ* hybridization (smFISH)[31,32,39,40] to quantify the expression of individual P1 and P2 promoters for each endogenous *hb* gene locus in nc11–

13 embryos. Contrary to the previous notion, we found that P1 contributes a modest yet non-negligible fraction of bursty *hb* transcription in early development. Using different TF dosages and enhancer deletion, we showed that Bcd activates both promoters via the proximal and distal enhancers. Compared with P2, P1 activation requires cooperative binding of more Bcd molecules and a synergistic (as opposed to competitive) action of the two enhancers. Analyzing the statistics of nascent mRNA signals from individual promoter loci revealed that both promoters follow a unified scheme of three-state transcription kinetics. Cooperative Bcd binding at either enhancer can drive a promoter to a weak active state, while additional Bcd binding at the second enhancer can turn the promoter to full-power transcription. The two promoters differ in their responses to different Bcd binding configurations. P2 transcription is primarily driven by Bcd binding at a single enhancer, while P1 transcription relies more on Bcd binding at both enhancers. In concert, these results provide a simple and quantitative mechanism for the differential regulation of alternative promoters. Our quantitative approach may be generalized as a framework for deciphering complex eukaryotic gene regulation involving multiple promoters and enhancers.

Results

hb P1 and P2 promoters are both active in early embryogenesis.

Previous studies reported that each *hb* promoter produces a unique transcript isoform (P1: *hb*-RB, P2: *hb*-RA)[29,41]. To quantify the endogenous P1 and P2 activities in early embryos, we applied smFISH with four sets of oligonucleotide probes designed for different regions of *hb* mRNAs. Specifically, two P1-specific probe sets targeted the 5' untranslated region (UTR) and intron region of *hb*-RB. A P2-specific probe set targeted the 3'UTR of *hb*-RA. Finally, a probe set targeted the coding sequence (CDS) region shared by *hb*-RA and *hb*-RB (Fig. 1a, Supplementary Table 1).

Confocal imaging and automated image analysis identified actively transcribing *hb* loci as bright FISH spots in wild-type (WT) syncytial blastoderm embryos (Fig. 1b). We quantified the instantaneous transcriptional activity of every locus in units of individual cytoplasmic mRNAs[32] (Fig. 1b; See Methods). For the CDS signal, we observed two bright FISH spots in most anterior nuclei of nc11–13 embryos (Fig. 1c, Supplementary Fig. 1a), consistent with previous literature[31]. It indicates that at least one of the two *hb* promoters is active in the anterior side of the embryo.

Similar to the CDS signal, promoter-specific FISH signals were also concentrated in the anterior half of nc11–13 embryos (Fig. 1c). Specifically, ~50% of the anterior loci (within the range of 0.20–0.40 embryo length (EL)) showed bright P2-3'UTR spots (Fig. 1d), consistent with previous reports of active P2 expression in early development[27,30]. Surprisingly, ~18%–34% of the anterior loci in nc11–13 embryos also contained bright P1-5'UTR and intron spots (Fig. 1d). The percentage of P1-active anterior loci increased with the nuclear cycle (Fig. 1d), suggesting that P1 becomes increasingly active during development. In total, P1- and P2-specific probe signals were exhibited in ~64%–76% of the anterior loci, lower than that of the CDS signal. This

percentage difference may be because the P2-specific probes target the very end of *hb*-RA, which is missing in many incomplete nascent transcripts. Alternatively, some P2 transcripts may terminate at a proximal site without passing the target region of our probes[42].

To analyze the instantaneous transcription of P1 and P2, we plotted, for each probe set, the number of nascent mRNAs against the nuclear position for all loci in the embryo (Fig. 1b). During nc11–13, the average expression profile for each probe set exhibited a reverse-sigmoidal shape within the range 0.25–0.75 EL (Fig. 1e). By fitting the profile to a logistic function, we estimated the maximal transcription level r_{max} and the boundary position x_0 of the anterior expression domain (See Methods). As expected, r_{max} of P1-specific signals was much smaller than that of the CDS signal (Fig. 1f), indicating that P1 contributes less than P2 to early *hb* expression. r_{max} of the P2-specific signal also remained low (Fig. 1f), as the probe set targets the 3'UTR of *hb*-RA. In contrast, active P1 and P2 loci exhibited much more nascent mRNAs (Supplementary Fig. 1b), suggesting that both promoters perform intensive expression once active. The boundary position x_0 of the P1-specific signals was significantly lower than that of the P2-specific and CDS signals (~0.42 EL v.s. ~0.46EL, $p<0.05$, Student's t-test, Fig. 1g), indicating that P1 is activated at a more anterior position than P2.

In addition to the mean expression level, the activities of individual promoter loci exhibited substantial variability (Fig. 1b). The Fano factors of anterior P1- and P2-specific signals were much larger than one (Fig. 1h; See Methods), indicating bursty transcription from both promoters[43,44]. Further analysis showed that such burstiness resulted from the intrinsic

stochasticity of *hb* transcription, with the two promoters behaving independently (Supplementary Fig. 1c, d; See Methods).

P1 contributes a modest yet non-negligible fraction of nascent *hb* transcription.

To quantify the contributions of P1 and P2 to nascent *hb* transcription, we note that the *hb* CDS signal reflects either P1 or P2 transcription. At steady state, the nascent CDS signal from a specific promoter should be, on average, proportional to the signal of promoter-specific probes (See Methods). Thus, the observed nascent CDS signal (r_{CDS}) should be a linear combination of the nascent P1-5'UTR and P2-3'UTR signals (r_{P1} and r_{P2}):

$$r_{CDS} = a_1 r_{P1} + a_2 r_{P2} \quad (1)$$

where a_1 and a_2 are ratio parameters that depend on the mRNA elongation and termination dynamics (See Methods)[45]. The two terms on the right-hand side of the equation distinguish the contribution of each promoter to the CDS signal.

To examine Equation (1), we compared P1-5'UTR, P2-3'UTR, and CDS signals in the same embryo (Fig. 2a). The three signals from P1- and P2-active gene loci satisfied a linear relationship (Fig. 2b) with a_1 = 0.53 ± 0.07 and a_2 = 2.74 ± 0.16 (Fig. 2c). In a simple transcription model with deterministic elongation and termination processes[45,46], the above a_1 and a_2 values indicate a post-elongation residence time (T_R) in terminating a nascent mRNA (Fig. 2d, Supplementary Fig.

2a; See Methods). The estimated T_R for P1 and P2 was 142 s and 46 s, respectively, consistent with previous estimations of transcription termination[45,47,48]. These results validate Equation (1) and enable the decomposition of the CDS expression profile into different promoter activities (Fig. 2e, Supplementary Fig. 2b–d; See Methods). In the anterior expression domain, the contribution of P1 increased with the nuclear cycle to ~13% (Fig. 2f), which was modest yet non-negligible.

In addition to the mean expression level, we investigated the contributions of P1 and P2 to the variability of nascent *hb* transcription quantified by the intrinsic noise of the single-locus data[49]:

$$\eta^2 = \frac{\langle (m_1 - m_2)^2 \rangle}{2\langle m_1 \rangle \langle m_2 \rangle} \tag{2}$$

where m_1 and m_2 are the numbers of nascent mRNAs at two homologous loci in the same nucleus, respectively (See Methods). During nc11–13, the intrinsic noise of P1 was 2–5 times higher than that of P2 in the anterior expression domain (Fig. 2g). This was mainly due to the low expression level of P1. Since the CDS signal is the sum of P1 and P2 activities, its intrinsic noise is significantly lower than that of P2 by ~5 folds ($p<0.01$, Student's t-test, Fig. 2g). Considering that P1 and P2 activities were independent (Supplementary Fig. 1d), their contributions to CDS noise could be summed as:

$$\eta^2_{\text{CDS}} = \eta^2_{\text{P1-CDS}} f^2_{\text{P1-CDS}} + \eta^2_{\text{P2-CDS}} f^2_{\text{P2-CDS}} \tag{3}$$

where $\eta^2_{\text{P1-CDS}}$ and $\eta^2_{\text{P2-CDS}}$ denote CDS noises originating from P1 and P2 activities, respectively,

while $f_{\text{P1-CDS}}$ and $f_{\text{P2-CDS}}$ indicate the relative contributions of each promoter to the mean CDS signal (See Methods). Equation (3) revealed that P1 contributed to up to ~20% of the anterior CDS noise during nc11–13 (Fig. 2h), which exceeded its contribution to the mean CDS signal.

Activating P1 requires cooperative binding of more Bcd molecules than activating P2.

In early embryogenesis, the exponential gradient of Bcd is the primary driver of anterior *hb* transcription[29,34]. To examine whether Bcd directly regulates both promoters, we used a transgenic fly line with 1× functional *bcd* gene[50]. With reduced Bcd dosage, the anterior expression domains of P1 and P2 both retreated towards the anterior pole (Fig. 3a). Specifically, P1 and P2 expression boundaries in 1× *bcd* embryos significantly shifted to the anterior side by ~0.07 EL ($p<0.01$, Student's t-test, Fig. 3b). Thus, Bcd activates both promoters in nc11–13 embryos.

To quantify the regulation of each promoter by Bcd, we combined promoter-specific smFISH of *hb* mRNA with immunofluorescence of Bcd (Fig. 1b). We estimated the absolute Bcd concentration in each nucleus using a previously developed image analysis method[32] (See Methods). For each promoter, we plotted the average transcriptional response of a gene locus (0.25–0.75 EL) to nuclear Bcd concentration in the embryo (Fig. 3c), known as the gene regulation function (GRF)[51]. Previous studies reported that *hb* GRF fitted well to a Hill function with a Hill coefficient $h \approx 5$–6[32,52]. A common explanation was that Bcd activates *hb* by cooperatively binding multiple sites in the regulatory sequence[33,53,54]. Here, we found that the Hill coefficient of P1-

specific GRFs (P1-5'UTR: h = 7.1 ± 0.4, P1-intron: h = 7.2 ± 0.3, mean ± s.e.m.) was significantly higher than that of P2-specific (h = 4.6 ± 0.2) and CDS (h = 5.3 ± 0.3) GRFs during nc11–13 ($p<0.05$, Fig. 3d). These results indicate that P1 activation corresponds to higher-order cooperative binding than P2 activation. Moreover, P1-specific signals exhibited a higher concentration threshold C_0 for activation than P2-specific and CDS signals, with their ratios being consistently larger than one (Fig. 3e). This agrees with our observation of P1 and P2 expression boundaries (Fig. 1g).

To directly quantify Bcd binding corresponding to P1 and P2 activation, we measured local enrichment of the Bcd signal in the vicinity of P1- and P2-active *hb* loci[32] (Fig. 3f; See Methods). In the anterior expression domain of nc11–12 embryos, we estimated an average binding of ~6.2 Bcd molecules at P1-active loci, exceeding that of ~4.7 Bcd molecules at P2-active loci (Fig. 3f). This result confirmed that P1 activation requires cooperative binding of more Bcd molecules than P2 activation. In nc13, along with the increase in P1 activity (Fig. 1d, f), the number of bound Bcd molecules at P1-active loci dropped to ~5.4. Examining how Bcd binding at active promoter loci varied with nuclear Bcd concentration (or nuclear position) revealed that both P1- and P2-specific binding curves plateaued at ~4–5 Bcd molecules, while the P1-specific binding curve exhibited an additional plateau with ~8–10 Bcd molecules in nc11–12 (Fig. 3g, Supplementary Fig. 3). As suggested previously, these plateaus may correspond to distinct Bcd binding states at *hb* enhancers[32]. The additional Bcd binding plateau for P1 implies that P1 activation may involve more Bcd-binding steps than P2 activation. This additional plateau decreased to ~6 Bcd

molecules in nc13, implying a change in Bcd binding dynamics.

Two Bcd-dependent enhancers synergistically drive P1 activation.

Two Bcd-dependent enhancers are involved in early *hb* regulation. To distinguish their roles in P1 and P2 activation, we used transgenic fly lines derived from a bacterial artificial chromosome (BAC) containing the *hb* gene and its regulatory sequence[35] (Fig. 4a; See Methods). The transgenes retained the *hb* promoters, introns, and UTRs, whereas the *hb* CDS was replaced with a *yellow* reporter gene. The distal and proximal enhancers of the two transgenes were substituted, respectively, while the regulatory sequence of the third transgene was kept intact as a control.

To measure the promoter activities of the transgene, we labeled transgenic embryos with three sets of smFISH probes targeting P1 intron, *hb* CDS, and *yellow*, respectively (Fig. 4b). The P1 signal corresponding to the endogenous *hb* gene was identified and excluded based on its colocalization with the *hb* CDS signal. We compared the anterior P1 and *yellow* (mainly from P2) signals in the control and enhancer-removed transgenes in nc11–13 embryos (Fig. 4c, Supplementary Fig. 4a, b). The control transgene was active (with a nascent *yellow* signal) in >70% of the anterior loci (Fig. 4d), similar to the endogenous *hb* gene. Removal of either enhancer significantly lowered the percentage of the *yellow*-positive anterior loci by >16% ($p<0.05$, Student's t-test) in nc11–12. This finding is consistent with previous reports that both enhancers

are required for authentic *hb* expression[23,30]. In contrast, removing the proximal but not the distal enhancer lowered the percentage of P1-active anterior loci by >26% in nc11–13 embryos (Fig. 4c, d), suggesting that early P1 activation relies more on the proximal enhancer.

Comparing the *yellow* expression profiles between the control and enhancer-removed transgenes (Fig. 4e, Supplementary Fig. 4c, d) revealed a decrease in the maximum anterior expression level r_{max} by ~34.5% in response to proximal or distal enhancer removal (Fig. 4f). This agrees with a previous report that P2 subadditively integrates regulatory inputs from different enhancers via enhancer competition[23]. In contrast, removing the proximal enhancer decreased the maximum expression level r_{max} of P1 by >50% in nc11–12, while removing the distal enhancer barely changed r_{max} of P1 (Fig. 4f). These results confirmed that the early P1 expression amplitude is primarily determined by the proximal enhancer.

Besides affecting the expression amplitude, both enhancers are critical for promoter expression boundaries. Specifically, removing the proximal enhancer shifted the *yellow* expression boundary towards the anterior pole, while removing the distal enhancer caused a posterior shift of the *yellow* profile (Fig. 4g). These results agree with the enhancer competition model for P2 activation[23], in which the expression boundary resulting from two enhancers lies between that from individual ones. In contrast, the deletion of either enhancer caused an anterior shift of the P1 expression boundary (Fig. 4h), which is inconsistent with the enhancer competition model. It shows that the existence of a second enhancer helps activate P1 at lower Bcd concentration. Thus, the two enhancers may interact synergistically to drive P1 activation.

Promoter-specific transcription kinetics reveal a unified scheme of enhancer-promoter interaction

Nascent mRNA copy number statistics reflect the microscopic mechanisms of gene regulation[55]. Previous studies reported a super-Poissonian distribution of the nascent mRNA copy number on individual *hb* gene loci[31,32,45]. Such distribution can be explained by a minimal model of two-state transcription kinetics[43,46,55-57] with Bcd modulating the frequency of stochastic transition from an inactive to an active transcription state[32].

To uncover the kinetics of P1 and P2 transcription, we measured, for each embryo, the distributions of P1 and P2 nascent mRNA signals in different Bcd concentration ranges (Fig. 5a). Contrary to the prediction from the two-state model[46], each promoter exhibited a wide distribution with more than one population of active loci characterized by different expression levels (Fig. 5a). A natural explanation of this phenomenon is that some of the observed promoter loci may indeed be a pair of closely located sister loci that are indistinguishable under the microscope[31,45]. However, the nascent mRNA signal from optically resolved sister loci pairs also exhibited two active populations, with >38% of P1 and >17% of P2 active sister loci corresponding to the minor population (Fig. 5b; See Supplementary Note). Thus, P1 and P2 transcription must be modeled as a three-state process.

In this model, a promoter randomly switches between an inactive state (state 0) and two

active states (states 1 and 2) with Poissonian rates k_{ij} (i, j = 0, 1, 2), whereas new transcripts are only initiated in active states with rates $k_{\mathrm{INI},i}$. Following transcription initiation, each nascent mRNA elongates with constant speed V_{EL}, and resides on the gene for an extra termination period T_{R} before being released (Fig. 5a). With a detailed balance between states, we solved the steady-state distribution of nascent mRNAs per promoter locus and compared it with experimental data to infer the kinetic parameters (Fig. 5a; See Supplementary Note). We found that P1 and P2 both followed a sequential activation scheme from state 0 to state 2 (Supplementary Fig. 5a; Supplementary Note). Bcd mainly regulated the activation rates, k_{01} and k_{12}, while other kinetic rates remained constant (Supplementary Fig. 5b).

For both promoters, the Bcd dependence of k_{01} and k_{12} satisfied Hill functions (Fig. 5c), with the corresponding Hill coefficient h and concentration threshold C_0 identical between the promoters (Fig. 5d, e). Specifically, the Hill coefficient of k_{01} was ~4.5 (P1: h = 5.2 ± 0.5, P2: h = 4.0 ± 0.5, Fig. 5d), close to that of P2-GRF. In contrast, the Hill coefficient of k_{12} was ~10 (P1: h = 10.0 ± 1.0, P2: h = 10.1 ± 1.0, Fig. 5d). h of k_{01} and k_{12} each matched a plateau in the Bcd binding curves of *hb* promoters (Fig. 3g), suggesting that each promoter activation step may involve a cooperative Bcd binding event. The concentration threshold of k_{01} was ~24 nM (P1: C_0 = 23.5 ± 2.5, P2: C_0 = 23.9 ± 2.6, Fig. 5e), which was close to that of P2-GRF. In contrast, the concentration threshold of k_{12} was only ~17 nM (P1: C_0 = 17.1 ± 2.6, P2: C_0 = 17.2 ± 2.4, Fig. 5e), suggesting that the second Bcd-binding event is easier to happen than the first one.

In addition to the Bcd dependence, the ratio between the activation and inactivation rates

determines the probability of each transcription state. In the anterior expression domain, k_{01}/k_{10} for P1 was smaller than that for P2 by >5 folds (Supplementary Fig. 5c), indicating that P2 is more likely to be activated than P1. In contrast, k_{12}/k_{21} for P1 exceeded that for P2 by >2 folds (Supplementary Fig. 5d), indicating that active P1 is more inclined to reach state 2 than active P2. For each active state (i = 1, 2), the transcription initiation rate $k_{\mathrm{INI},i}$ for P1 were lower than that for P2 (Fig. 5f). However, since $k_{\mathrm{INI},2}$ of P1 was close to $k_{\mathrm{INI},1}$ of P2, mRNA production from active P1 loci was comparable to that of active P2 loci (Supplementary Fig. 1b). Specifically, State 2 contributed much more to P1 transcription than to P2 transcription (57.7% vs. 25.7%, Fig. 5g). This explains the difference (in h and C_0) between the P1- and P2-GRFs. I.e., the differential regulation of P1 and P2 results from their preference for different active states.

To relate transcription kinetics with enhancer activities, we applied theoretical analysis to enhancer deletion experiments. For early P1 transcription (nc11-12), the removal of either enhancer yielded a uniform decrease of k_{01} by ~30% in the anterior expression domain (Fig. 5h, Supplementary Fig. 5e). Deleting the proximal enhancer also caused a modest anterior shift of the k_{01} boundary (Fig. 5h). These results suggest that either enhancer can drive promoter activation to state 1. In contrast, deleting either enhancer caused a dramatic anterior shift of the k_{12} boundary (Fig. 5h, Supplementary Fig. 5f). This indicates that state 2 is primarily driven by the synergistic action of both enhancers. When the Bcd concentration is sufficiently high, state 2 may also be driven by a single enhancer, possibly due to extra Bcd binding at weak binding sites. These mechanisms explain the difference between P1 and P2 in the expression boundary shift

upon enhancer removal (Fig. 4g, h).

Altogether, we propose that the two *hb* promoters follow a common kinetic scheme of Bcd-dependent activation (Fig. 5i; See Supplementary Note). Cooperative Bcd binding at either enhancer can drive a promoter to a basal active state, while full-power transcription requires additional Bcd binding at the second enhancer. The two promoters differ in their responses to different Bcd binding configurations. P2 is primarily activated by Bcd binding at a single enhancer, which results in competitive action of the two enhancers. In contrast, P1 activation relies more on Bcd binding at both enhancers, which leads to synergistic enhancer action. The predicted joint distribution of P1- and P2-specific nascent mRNAs agreed well with experimental results (Fig. 5j; See Supplementary Note).

Discussion

Traditional FISH and lacZ reporter experiments reported that Bcd-dependent *hb* transcription in early embryogenesis only involves the P2 promoter[33,34,36,37], while P1 is Bcd-insensitive and remains silent until late nc14[27,30,36]. Here, using smFISH method with single-molecule sensitivity, we showed that Bcd activated both promoters in the anterior domain of nc11–13 embryos. P1 contributes up to ~13% of nascent *hb* signal and up to ~20% of nascent *hb* noise. These fractions

are modest compared to those of P2, yet significant enough to affect the shape and fluctuation of the *hb* expression pattern. Thus, the Bcd-dependent P1 expression is not negligible.

According to previous studies, the difference in P1 and P2 expression levels is due to specific motif codes in promoter sequences. P2 contains Zelda binding sites and a strong TATA box to facilitate chromatin opening and promoter activation[16,30,58]. In contrast, the lack of Zelda binding sites and TATA box in the P1 promoter significantly impedes local chromatin opening and promoter activation[17,38,59]. However, our results suggest that the lack of these motif codes may be insufficient to completely block early P1 activity. In fact, the local chromatin state of a gene has been reported to be highly dynamic and can randomly switch between “open” and “close” to allow transient gene expression[13,60]. This effect has been proposed as a mechanism for transcriptional bursting[60-62], consistent with our observation of bursty P1 transcription. Moreover, our observation that the percentage of active P1 loci increased with the nuclear cycle suggests a gradual increase in chromatin opening frequency during development. The stronger P1 activity in late nc14 may be a continuation of this trend.

In addition to the expression amplitude, the Bcd dependence of P1 and P2 showed quantitative differences. P1 activation requires a higher Bcd concentration and binding of more Bcd molecules than P2 activation. One possible mechanism for this difference is that the two promoters are driven by different enhancers. However, we showed that early P1 and P2 activation relied on the same pair of enhancers[35]. Another possibility is that the two promoters distinguish their behaviors by interacting with each other (e.g., mutual repression[63]). However, we found little

correlation between P1 and P2 activities. This is consistent with a previous promoter-deletion experiment[30] and indicates no interaction between the two promoters. Thus, the difference in Bcd dependence originates from complex enhancer-promoter interactions.

One form of complex enhancer-promoter interaction is the competitive action of the proximal and shadow enhancers for *hb* P2 activation[23]. Here, each enhancer corresponds to a unique Bcd-dependent regulatory effect, and the two enhancers compete for promoter activation[12]. The resulting P2 expression is a sub-additive integration of individual enhancers' regulatory effect. Such a mechanism can lead to a different Bcd-dependent relationship if the enhancers' relative weights in the integration change. However, the P1 expression boundary always shifts anteriorly in response to the deletion of either enhancer, suggesting a synergistic, as opposed to competitive, action of the two enhancers for P1 activation. This result also revealed that the Bcd dependence of P1 and P2 differed even in the case of a single enhancer, indicating that the TF dependence of an enhancer's regulatory effect may be promoter-specific.

To understand the mechanism of promoter-specific interactions with enhancers, we analyzed the transcription kinetics of individual promoters. Unlike the previous model of two-state *hb* transcription[31,32,45], we found that both P1 and P2 activities need to be described by three-state kinetics with sequential activation steps. Multi-step transcriptional activation was previously proposed to model enhancer-, promoter-, or chromatin-related intermediate states during activation[13,60,62]. However, these intermediate states are rarely experimentally detectable. In the case of *hb*, previous measurements mixed signals from P1 and P2, whose difference easily

overwhelmed the subtle signatures of different transcription states. Here, the identification of multiple active states was possible owing to promoter-specific demixing of the *hb* nascent mRNA signal. Thus, it is important to distinguish alternative promoters in the study of complex gene regulation.

Our model revealed that Bcd regulates both activation steps (k_{01} and k_{12}) through cooperative binding. This suggests that the two active states of a promoter may correspond to different Bcd binding states at *hb*. Previous literature has shown evidence of multiple Bcd binding states at *hb*[32]. However, only one binding state (with ~4–5 Bcd molecules) was reported to coincide with gene activation[32]. The biological functions of the other binding states were unclear. Here, we showed that the binding of ~4–5 Bcd molecules happened at a single enhancer (either proximal or distal). Enhancers at this binding state can competitively drive promoter activation, consistent with a previous study of *hb* P2 activation[23]. In contrast, the binding of ~8–10 Bcd molecules involved both enhancers, leading to synergistic activation of a promoter. Such a mechanism was proposed for some eukaryotic genes[23,64] but has never been reported for *hb*.

P1 and P2 differ in their preference for different Bcd binding states. Specifically, the P1 response to the lower Bcd binding state is much less than that of P2. A possible reason is that P1 lacks specific motif codes for chromatin opening[30]. Thus, its activation may require TF binding at enhancers to help open the local chromatin configuration[61,62]. Binding of more Bcd molecules at both enhancers may be more effective for chromatin opening. However, since P1 locates between the two enhancers, binding of Bcd (and other regulatory factors) at the proximal enhancer may

physically block P1 transcription. It is unclear how P1 coordinates its activation and transcription. One possibility is that TF binding at enhancers is only transiently needed to form the preinitiation complex. Future experiments using high-resolution live imaging techniques would be likely to solve this puzzle[65,66].

Altogether, these results showed that a single kinetic scheme could create apparently different types of enhancer-promoter interactions. Such a unified scheme may be shared by alternative promoter regulation in many eukaryotic systems and may be crucial for phenotypic complexity in higher eukaryotes [7]. Moreover, our combined experimental and theoretical approach directly relates TF binding at individual enhancers with the stochastic activity of each promoter. A generalization of this approach (e.g., by including more regulatory factors) will help understand the regulatory mechanisms for cell fate decisions and enable the precise design of synthetic gene circuits[67].

Methods

Fly strains

Oregon-R (OreR) strain was used as the wild type. 1×*bcd* strain (*+/CyO-bcd+*; *E1s*) was developed previously[50] and was obtained as a gift from Dr. Jun Ma (Zhejiang University). *hb*-BAC reporter constructs were previously developed[35]. The distal-enhancer-removed BAC construct

and its control were obtained as gifts from Dr. Michael Perry (University of California San Diego) and Dr. Alistair Boettiger (Stanford University). The proximal-enhancer-removed construct and its control were rebuilt as previously described[35]. Transgenes in these constructs were constructed from BAC CH322-55J23, which contains a 20-kbp *Drosophila* genomic sequence encompassing the *hb* gene and its proximal and distal enhancers. A *yellow-kanamycin* fusion was used to replace the *hb* CDS in all BACs, while the proximal or distal enhancers in the BACs were substituted with *ampicillin*. All BACs were integrated on chromosome 2 of *Drosophila*. The distal-enhancer-removed transgene and its control were integrated into landing site VK37 (Bloomington Stock Center number 24872). The proximal-enhancer-removed transgene and its control were integrated into landing site attP40 (Bloomington Stock Center number 25709).

smFISH probe design

Sets of DNA oligonucleotides complementary to the target transcripts (nine probes for *hb* P1 5'UTR, 32 probes for *hb* P1 intron, 48 probes for *hb* CDS, eight probes for *hb* P2 3'UTR, and 43 probes for *yellow*) were designed (Supplementary Table 1) and synthesized (Biosearch Technologies). Each probe was ordered with a 3' amine group (mdC(TEG-Amino)) and was conjugated to various fluorophores, as described previously[32]. *hb* P1-5'UTR and intron probes were conjugated with tetramethylrhodamine (TAMRA; Thermo Fisher Scientific, C6123). *hb* P2-3'UTR and *yellow* probes were conjugated with Alexa Fluor™ 647 (Invitrogen, A20106). *hb* CDS probes were conjugated with either Alexa Fluor™ 647 (Invitrogen, A20106) or Alexa Fluor™ 488 (Invitrogen, A20100).

Embryo staining

Embryo collection, fixation, and labeling were performed according to a previously published protocol[32]. Briefly, 2-hour-old embryos were collected at 25°C, fixed with 4% paraformaldehyde solution, and stored in 100% methanol at −20°C. For smFISH, fixed embryos were rehydrated (4 × 10 min) in PBTx (1× PBS, 0.1% (v/v) Triton X-100) at room temperature, washed (2 × 10 min) in hybridization wash buffer at 30°C, and incubated with the probe-containing hybridization buffer (2× SSC, 20% (w/v) formamide, 0.1% (v/v) Triton X-100) at 30°C overnight. After hybridization, embryos were washed in hybridization wash buffer at 30°C (2 × 10 min) and in 2× SSC at room temperature (2 × 10 min). For immunofluorescence (IF), embryos were washed (4 × 10 min) in PBTx and blocked in PBT-B (1× PBS, 20% (v/v) western blocking reagent (Roche, 11921673001), 2 mM ribonucleoside vanadyl complex (NEB, S1402S), 0.1% (v/v) Triton X-100) at room temperature for 1 h. The preabsorbed rabbit anti-Bcd primary antibody (Santa Cruz Biotechnology, SC-66818) was incubated with embryos at 4°C for 20 h. Following primary antibody staining, embryos were further washed (4 × 10 min) in PBTx, blocked in PBT-B at room temperature for 1 h, and incubated with goat anti-rabbit IgG secondary antibodies conjugated with Alexa Fluor™ 488 (Invitrogen, A11008) at room temperature for 1 h. For DNA counterstaining, embryos were washed (4 × 10 min) in PBTx and stained with Hoechst 33342 at room temperature for 10 min. Following additional washes (4 × 10 min) in PBTx, embryos were mounted in Aqua-Poly/Mount (Polysciences, 18606). Imaging was performed after the samples were completely solidified.

Imaging

Most embryos were imaged using a Zeiss LSM 880 laser scanning confocal microscope equipped with a GaAsP detector and a 63× oil-immersion objective (1.4 NA). 16-bit image stacks were acquired with a pixel size of 71 × 71 nm^2 and a z-step size of 0.32 μm. A small number of embryos (n = 5) were imaged using a Leica TCS SP8 confocal microscope equipped with a GaAsP detector and a 63× oil-immersion objective (1.4 NA). 12-bit image stacks were acquired with a pixel size of 81 × 81 nm^2 and a z-step size of 0.3 μm. Nc11–13 embryos at the mitotic interphase were selected based on the number and shape of the nuclei (Hoechst signal). Approximately 10 μm of the cortex layer of each embryo was imaged.

Preprocessing and nuclear segmentation

Image processing and data analysis followed a previously developed pipeline[32] with updated algorithms to improve accuracy and efficiency. Briefly, raw images were divided by a normalized flat-field image to correct for monochromatic aberrations. Three-dimensional (3D) segmentation of nuclei from the Hoechst image stack was done using a combination of local threshold (to optimize a circularity parameter) and watershed (to separate the merged nuclei). The nuclear cleavage cycle of the embryo was determined by the number of recognized nuclei. Embryo boundary was identified from the averaged Hoechst image by thresholding image pixels outside the nuclear area. This boundary was then used to determine the AP position of each nucleus.

mRNA quantification

For mRNA quantification, spot candidates in smFISH images were identified as 3D local maxima in the image stack. Since the splicing process occurs inside the nucleus, intron spot candidates were only identified in the nuclear region. The local intensity profile of each candidate was fitted to a two-dimensional (2D) Gaussian function to extract the peak height (I_{peak}) and radius (σ_0). The spot intensity was calculated as $I = 2\pi I_{peak}\sigma_0^2$. By comparing the joint distribution of peak height and radius between the anterior and posterior spots, a 2D threshold was determined to distinguish real mRNA spots from background noise. The typical intensity, I_0, of a single mRNA molecule was extracted by fitting the primary peak of the spot intensity distribution to a multi-Gaussian function. A threshold of $3I_0$ was defined to identify sites of active transcription from mRNA spots inside each nucleus. The equivalent number of nascent transcripts at each transcription site was estimated by dividing the intensity of the transcription site by I_0.

To identify signals corresponding to the same *hb* locus in two different smFISH channels, we calculated the mutual distance between every possible pair of active transcription sites detected in different smFISH channels. The distribution of all the mutual distances in an embryo exhibited two distinct populations. Using a threshold distance of 0.55 μm, the colocalized pairs of active transcription sites were identified. Each pair corresponded to a single *hb* locus. In contrast, each of the unpair active transcription sites belonged to a different *hb* locus, whose activity in the other smFISH channel was zero.

In some anterior nuclei, three or four bright FISH spots appeared because of the replication of the *hb* gene[31] (Supplementary Fig. 1a). To identify nascent mRNA signals corresponding to

sister loci pairs, we calculated the mutual distance between every possible pair of active transcription sites in each nucleus. The distribution of all the mutual distances in the embryo exhibited two distinct populations. Using a threshold distance of 0.71 μm, sister loci pairs were distinguished from unpaired homologous loci.

Protein quantification

For protein quantification, the average immunofluorescence (IF) intensity of each nucleus was calculated from the central *z*-slice of the nucleus. IF spots in the cytoplasm were identified and quantified to determine the typical intensity, I_1, of a single protein molecule, following the same procedure used for the smFISH signal. The absolute protein concentration of a nucleus was estimated by dividing the average IF intensity of the nucleus by $(2\pi)^{1/2}\sigma_z I_1$, where σ_z is the half-width of the single-protein intensity profile in the *z* dimension.

In the preceding steps, nuclear segmentation, embryo boundary identification, and detection of active transcription sites could be further refined and corrected manually using custom MATLAB graphical user interfaces.

Measuring the spatial profile of promoter activity

We analyzed the expression profile of a promoter using embryos in the mid-to-late mitotic interphase to ensure steady-state promoter activity. To identify silent loci (no nascent mRNA) in

the embryo, we assumed that each nucleus had two (for nuclei containing ≤2 active transcription sites) or four (for nuclei containing >2 active transcription sites) *hb* loci. The number of silent loci was estimated accordingly. For each embryo, we plotted the number of nascent mRNAs (r) against nuclear position (x) for all loci and binned individual data points by x. Within the range of 0.25–0.75 EL, we used a least-square algorithm to fit the binned data to a logistic function:

$$r = r_{\max} \frac{e^{-(x-x_0)/d}}{e^{-(x-x_0)/d} + 1} + r_0 \tag{4}$$

where x_0 is the boundary position of the expression domain, d is the half-width of the transition region, $r_{\max}$ is the maximal transcription level induced in the anterior region, and r_0 denotes the basal activity in the posterior part.

Measuring the fluctuation of promoter activity

At the single-molecule level, gene expression constantly varies over time and between different cells. Based on the correlation between the homologous loci in the same cell, the fluctuation or noise of gene expression may be divided into two parts: the intrinsic noise due to the inherent stochasticity of biochemical reactions and the extrinsic noise caused by cell-to-cell variability of the microenvironment[68].

To characterize the expression variability of different *hb* promoters, we computed, for each promoter, the Fano factor (F) of nascent mRNA copy number (r) per locus in the anterior expression domain (0.2–0.4 EL)[32]:

$$F = \frac{\sigma^2}{\langle r \rangle} \tag{5}$$

where $\langle r \rangle$ and σ are the mean and standard deviation of the single-locus data, respectively. The Fano factors of P1- and P2-specific signals were much larger than one (Fig. 1h), indicating bursty transcription from both promoters[43,44].

We quantified the intrinsic noise of P1 and P2 expression in the anterior expression domain (0.2–0.4 EL) of an embryo using the following formula[49]:

$$\eta^2 = \frac{\langle (m_1 - m_2)^2 \rangle}{2 \langle m_1 \rangle \langle m_2 \rangle} \tag{6}$$

where m_1 and m_2 are the nascent mRNA signals at two homologous loci in the same nucleus measured using a given probe set, respectively.

Measuring the correlation between promoter loci

To distinguish the intrinsic and extrinsic noise of different promoter activities, we computed the correlation coefficient (ρ) of the nascent mRNA copy number between the two homologous copies of a given promoter within the same nucleus. Specifically, we divided the single-locus data of the promoter activity (equivalent number of nascent mRNA) in an embryo into two groups, r_1 and r_2. Each group corresponded to one of the two homologous loci in the nucleus. In a given region of the embryo, we applied the following formula:

$$\rho = \frac{\left\langle \left(r_1 - \langle r_1 \rangle\right) \cdot \left(r_2 - \langle r_2 \rangle\right) \right\rangle}{\sigma_1 \sigma_2} \tag{7}$$

where $\langle r_1 \rangle$, $\langle r_2 \rangle$, σ_1, and σ_2 are the mean and standard deviation of each group, respectively. In the anterior expression domain (0.2–0.4 EL), P1, P2, and CDS signals showed little correlation (<0.15, Supplementary Fig. 1c), agreeing with previous reports of loci independence[31,32]. Thus, intrinsic noise dominates P1 and P2 expression. To further evaluate the interaction between the two promoters, we computed the correlation coefficient between P1 and P2 signals from the same (intra-allele) or different (inter-allele) *hb* loci in the nucleus using the same formula, where r_1 and r_2 denote the activities of individual P1 and P2 copies, respectively. Both quantities were at low levels (<0.06, Supplementary Fig. 1d), indicating that the two promoters do not interact during expression.

Estimating promoter contributions to *hb* activity

To estimate the contribution of P1 and P2 activities to nascent *hb* transcription, we co-labeled fly embryos with P1-5'UTR, P2-3'UTR, and CDS probes (Fig. 2a). For P1- and P2-active *hb* loci in the position range of 0.2–0.7 EL of an embryo, we plotted the three FISH signals (in units of the number of nascent mRNAs) against each other (Fig. 2b). We applied linear regression to fit the single-locus data to Equation (1). With the inferred a_1 and a_2 values, we computed the P1 contribution to the average nascent CDS signal as

$$f_{\text{P1-CDS}} = \frac{a_1 r_{\text{P1}}}{r_{\text{CDS}}} \tag{8}$$

where r_{CDS} and r_{P1} are the average nascent CDS and P1-5'UTR signals per *hb* locus, respectively. The rest of the nascent CDS signal should come from P2. However, directly computing the P2 contribution using a similar formula revealed an extra component in the nascent CDS signal that was not labeled by promoter-specific probes (Supplementary Fig. 2b). In the anterior expression domain, this component corresponded to >20% of the nascent CDS signal (Supplementary Fig. 2c). This phenomenon agrees with our observation that some CDS-positive *hb* loci lack P1-5'UTR and P2-3'UTR signals. To determine the origin of this extra component, we fitted its expression profile to a logistic function. The estimated expression boundary position (x_0 = 0.47 EL ± 0.02 EL) matched the P2-3'UTR signal (Supplementary Fig. 2d), indicating that the component is from P2 transcription. The P2-specific probes failed to label this component because the probes target the 3'UTR of *hb*-RA. It could be that mRNA termination complexes on the 3' UTR block FISH probes from binding[47]. Moreover, a previous study reported alternative termination of *hb* transcription[42], which may account for unlabeled P2 transcripts. By summing the contributions from labeled and unlabeled P2 transcripts, we decomposed the CDS profile into different promoter activities (Supplementary Fig. 2e, f).

To compute promoter contributions to the intrinsic noise of nascent *hb* transcription, we estimated $\eta^2_{\text{P1-CDS}}$ in Equation (3) from the intrinsic noise of the nascent P1-5'UTR signal (see *Mathematical modeling of transcriptional kinetics* of Supplementary Note). The contribution of P1

to the intrinsic noise of nascent CDS signal was calculated as:

$$w_{\text{P1-CDS}} = \frac{\eta^2_{\text{P1-CDS}} f^2_{\text{P1-CDS}}}{\eta^2_{\text{CDS}}} \tag{9}$$

The rest of the intrinsic noise was contributed by P2.

Measuring the gene regulation function (GRF)

To analyze the regulation of a promoter by Bcd, we plotted, for each embryo, the nascent mRNA signal of each gene locus versus the nuclear Bcd concentration within the position range of 0.25–0.75 EL. To extract the GRF, we binned individual data points by Bcd concentration and fitted them to a Hill function using a least-squares algorithm:

$$y = a\frac{[\text{Bcd}]^h}{[\text{Bcd}]^h + C_0^h} + d \tag{10}$$

where h is the Hill coefficient, C_0 is the concentration threshold for promoter activation, a indicates the maximal level of Bcd-dependent activity, and d denotes the basal activity.

Measuring Bcd binding

To quantify Bcd binding at a specific *hb* promoter, we used active P1 and P2 transcription sites to locate individual (active) promoter loci. Near each locus, a “locus-integration region” (*xy* distance ≤ 3 pixel and *z* distance = 0 from the locus) and an “out-of-locus region” (*xy* distance ≤ 6 pixels

and ≥ 3 pixels, z distance = 0) were defined, which covered the nuclear volumes of V_I and V_o, respectively. The nuclear IF signal (in units of the number of Bcd molecules) within these two regions was integrated and denoted as I_I and I_o, respectively. The enriched Bcd signal was defined as the difference between I_I and I_o in consideration of the volume difference between the two regions, i.e., $I_{\text{enrich}} = I_I - I_o \cdot V_I / V_o$. Data from multiple embryos in the sample nuclear cycle were pooled to increase the sample size. Data in the nuclear position range of 0.2–0.35EL were averaged to estimate the mean Bcd binding level of the anterior expression domain. To plot the Bcd binding curve, we binned the single-locus data by nuclear position and related mean Bcd enrichment with mean nuclear position or Bcd concentration.

Mathematical modeling of transcriptional kinetics

Stochastic modeling and inference of transcriptional kinetics are described in detail in the Supplementary Note.

Data and code availability

The raw image data reported in this paper are available at http://gofile.me/4yuzx/wKna2V9pK. Custom scripts for data analysis and mathematical modeling were written in MATLAB 2018a (MathWorks) and are available at https://github.com/Dr-xu-lab/Quantify-the-transcriptional-regulation.

References

1. Levine M. Transcriptional enhancers in animal development and evolution. *Curr Biol* **20**, 754-763 (2010).

2. Vernimmen D, Bickmore WA. The Hierarchy of Transcriptional Activation: From Enhancer to Promoter. *Trends Genet* **31**, 696-708 (2015).

3. Haberle V, Stark A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol* **19**, 621-637 (2018).

4. Rach EA, Yuan HY, Majoros WH, Tomancak P, Ohler U. Motif composition, conservation and condition-specificity of single and alternative transcription start sites in the Drosophila genome. *Genome Biol* **10**, R73 (2009).

5. Carninci P*, et al.* Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* **38**, 626-635 (2006).

6. Davuluri RV, Suzuki Y, Sugano S, Plass C, Huang TH. The functional consequences of alternative promoter use in mammalian genomes. *Trends Genet* **24**, 167-177 (2008).

7. Landry JR, Mager DL, Wilhelm BT. Complex controls: the role of alternative promoters in mammalian genomes. *Trends Genet* **19**, 640-648 (2003).

8. Pozner A*, et al.* Developmentally regulated promoter-switch transcriptionally controls Runx1 function during embryonic hematopoiesis. *BMC Dev Biol* **7**, 84 (2007).

9. Lu D, Sin HS, Lu C, Fuller MT. Developmental regulation of cell type-specific transcription by novel promoter-proximal sequence elements. *Genes Dev* **34**, 663-677 (2020).

10. Sendoel A*, et al.* Translation from unconventional 5' start sites drives tumour initiation. *Nature* **541**,

494-499 (2017).

11. de Klerk E, t Hoen PA. Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. *Trends Genet* **31**, 128-139 (2015).

12. Kvon EZ, Waymack R, Gad M, Wunderlich Z. Enhancer redundancy in development and disease. *Nat Rev Genet* **22**, 324-336 (2021).

13. Voss TC, Hager GL. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat Rev Genet* **15**, 69-81 (2014).

14. Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**, 613-626 (2012).

15. Hnisz D, Day DS, Young RA. Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. *Cell* **167**, 1188-1200 (2016).

16. Foo SM*, et al.* Zelda potentiates morphogen activity by increasing chromatin accessibility. *Curr Biol* **24**, 1341-1346 (2014).

17. Sun Y*, et al.* Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation. *Genome Res* **25**, 1703-1714 (2015).

18. Li XY, Harrison MM, Villalta JE, Kaplan T, Eisen MB. Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition. *Elife* **3**, e03737 (2014).

19. Furlong EEM, Levine M. Developmental enhancers and chromosome topology. *Science* **361**, 1341-1345 (2018).

20. Ghavi-Helm Y*, et al.* Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96-100 (2014).

21. Oudelaar AM*, et al.* Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. *Nat Genet* **50**, 1744-1751 (2018).

22. Scholes C, Biette KM, Harden TT, DePace AH. Signal Integration by Shadow Enhancers and Enhancer Duplications Varies across the Drosophila Embryo. *Cell Rep* **26**, 2407-2418 (2019).

23. Bothma JP, Garcia HG, Ng S, Perry MW, Gregor T, Levine M. Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. *Elife* **4**, e07956 (2015).

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

25. Lim B, Heist T, Levine M, Fukaya T. Visualization of Transvection in Living Drosophila Embryos. *Mol Cell* **70**, 287-296 (2018).

26. Su W, Jackson S, Tjian R, Echols H. DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1. *Genes Dev* **5**, 820-826 (1991).

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

28. Perry MW, Bothma JP, Luu RD, Levine M. Precision of hunchback expression in the Drosophila embryo. *Curr Biol* **22**, 2247-2252 (2012).

29. Driever W, Nusslein-Volhard C. The bicoid protein is a positive regulator of hunchback transcription in the early Drosophila embryo. *Nature* **337**, 138-143 (1989).

30. Ling J, Umezawa KY, Scott T, Small S. Bicoid-Dependent Activation of the Target Gene hunchback Requires a Two-Motif Sequence Code in a Specific Basal Promoter. *Mol Cell* **75**, 1178-1187 (2019).

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

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

33. Driever W, Thoma G, Nusslein-Volhard C. Determination of spatial domains of zygotic gene expression in the Drosophila embryo by the affinity of binding sites for the bicoid morphogen. *Nature* **340**, 363-367 (1989).

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

35. Perry MW, Boettiger AN, Levine M. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. *Proc Natl Acad Sci U S A* **108**, 13570-13575 (2011).

36. Margolis JS, Borowsky ML, Steingrimsson E, Shim CW, Lengyel JA, Posakony JW. Posterior stripe expression of hunchback is driven from two promoters by a common enhancer element. *Development* **121**, 3067-3077 (1995).

37. Wu X, Vasisht V, Kosman D, Reinitz J, Small S. Thoracic patterning by the Drosophila gap gene hunchback. *Dev Biol* **237**, 79-92 (2001).

38. Blythe SA, Wieschaus EF. Establishment and maintenance of heritable chromatin structure during early Drosophila embryogenesis. *Elife* **5**, e20148 (2016).

39. Femino AM, Fay FS, Fogarty K, Singer RH. Visualization of single RNA transcripts in situ. *Science* **280**, 585-590 (1998).

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

41. Tautz D*, et al.* Finger protein of novel structure encoded by hunchback, a second member of the gap

class of Drosophila segmentation genes. *Nature* **327**, 383-389 (1987).

42. Bender M, Horikami S, Cribbs D, Kaufman TC. Identification and expression of the gap segmentation gene hunchback in Drosophila melanogaster. *Dev Genet* **9**, 715-732 (1988).

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

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

45. Zoller B, Little SC, Gregor T. Diverse Spatial Expression Patterns Emerge from Unified Kinetics of Transcriptional Bursting. *Cell* **175**, 835-847 (2018).

46. Xu H, Skinner SO, Sokac AM, Golding I. Stochastic Kinetics of Nascent RNA. *Phys Rev Lett* **117**, 128101 (2016).

47. Bentley DL. Coupling mRNA processing with transcription in time and space. *Nat Rev Genet* **15**, 163-175 (2014).

48. Lenstra TL, Rodriguez J, Chen H, Larson DR. Transcription Dynamics in Living Cells. *Annu Rev Biophys* **45**, 25-47 (2016).

49. Waymack R, Fletcher A, Enciso G, Wunderlich Z. Shadow enhancers can suppress input transcription factor noise through distinct regulatory logic. *Elife* **9**, e59351 (2020).

50. Liu J, Ma J. Dampened regulates the activating potency of Bicoid and the embryonic patterning outcome in Drosophila. *Nat Commun* **4**, 2968 (2013).

51. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB. Gene regulation at the single-cell level. *Science* **307**, 1962-1965 (2005).

52. Gregor T, Tank DW, Wieschaus EF, Bialek W. Probing the limits to positional information. *Cell* **130**, 153-164 (2007).

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

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

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

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

57. Senecal A*, et al.* Transcription factors modulate c-Fos transcriptional bursts. *Cell Rep* **8**, 75-83 (2014).

58. Schulz KN*, et al.* Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo. *Genome Res* **25**, 1715-1726 (2015).

59. Nien CY*, et al.* Temporal coordination of gene networks by Zelda in the early Drosophila embryo. *PLoS Genet* **7**, e1002339 (2011).

60. Bintu L*, et al.* Dynamics of epigenetic regulation at the single-cell level. *Science* **351**, 720-724 (2016).

61. Lammers NC, Galstyan V, Reimer A, Medin SA, Wiggins CH, Garcia HG. Multimodal transcriptional control of pattern formation in embryonic development. *Proc Natl Acad Sci U S A* **117**, 836-847 (2020).

62. Eck E, Liu J, Kazemzadeh-Atoufi M, Ghoreishi S, Blythe SA, Garcia HG. Quantitative dissection of transcription in development yields evidence for transcription-factor-driven chromatin accessibility. *Elife* **9**, e56429 (2020).

63. Li X, *et al.* Heritable, Allele-Specific Chromosomal Looping between Tandem Promoters Specifies Promoter Usage of SHC1. *Mol Cell Biol* **38**, e0065817 (2018).

64. Choi J, *et al.* Evidence for additive and synergistic action of mammalian enhancers during cell fate determination. *Elife* **10**, e65381 (2021).

65. Li JR, *et al.* Single-Molecule Nanoscopy Elucidates RNA Polymerase II Transcription at Single Genes in Live Cells. *Cell* **178**, 491-506 (2019).

66. Li J, *et al.* Single-gene imaging links genome topology, promoter-enhancer communication and transcription control. *Nat Struct Mol Biol* **27**, 1032-1040 (2020).

67. Doshi J, Willis K, Madurga A, Stelzer C, Benenson Y. Multiple Alternative Promoters and Alternative Splicing Enable Universal Transcription-Based Logic Computation in Mammalian Cells. *Cell Rep* **33**, 108437 (2020).

68. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. *Science* **297**, 1183-1186 (2002).

Acknowledgments

We thank Jun Ma, Michael Perry, and Alistair Boettiger for the generous gift of fly lines. We thank Ido Golding, Anna Sokac, and Jun Ma for insightful discussion and valuable comments on the manuscript. This work was supported by the National Key R&D Program of China (grant no.

2018YFC0310803), the National Science Foundation of China (grant no. 11774225, 41921006), the National Science Foundation of Shanghai (grant no. 18ZR1419800), and the Burroughs Wellcome Fund Career Award at the Scientific Interface (grant no. 1013907). We gratefully acknowledge the imaging and computing resources provided by the Instrumental Analysis Center and the Student Innovation Center at Shanghai Jiao Tong University.

Author contributions

Conceptualization, J.W. and H.X.; Methodology, J.W., H.L., S.Z., and H.X.; Software, S.Z. and H.X.; Formal Analysis, J.W. and S.Z.; Investigation, J.W., S.Z., and H.X.; Writing – Original Draft, J.W. and H.X.; Funding Acquisition, H.X.; Resources, H.X.; Supervision, H.X.

Competing interests

The authors declare no competing interests.

Figure legends

Fig. 1 Absolute quantification of *hb* transcription reveals P1 and P2 activities in early embryogenesis.

a Schematic of the endogenous *hb* locus with two Bcd-dependent enhancers (green: distal enhancer, brown: proximal enhancer) and two promoters (P1 and P2). Each promoter drives the expression of a specific transcript isoform. Four smFISH probe sets were used to label different regions of *hb* mRNAs. **b** Confocal images of two wild-type *Drosophila* embryos, each labeled for two regions of *hb* mRNAs, Bcd protein, and DNA at nc12. Scale bars, 50 μm. Insets, magnified views of anterior nuclei. Scale bars, 5 μm. The number of nascent mRNAs at individual *hb* loci was plotted against the anterior-posterior (AP) position for different probe signals. **c** Percentage of active *hb* loci as a function of the AP position for different probe signals during nc11–13. Marked region, 0.2–0.4 EL. Shadings indicate s.e.m. **d** Average percentage of active *hb* loci in the position range of 0.2–0.4 EL for different probe signals during nc11–13. Error bars represent s.e.m. **e** The number of nascent mRNAs at individual *hb* loci as a function of the AP position for different probe signals during nc11–13. Marked region, 0.25–0.75 EL. Shadings indicate s.e.m. **f, g** The maximal transcription level (**f**) and the boundary position (**g**) of the anterior expression domain for different probe signals during nc11–13. Error bars represent s.e.m. **h** The Fano factor for the number of nascent mRNAs at individual *hb* gene loci in the position range of 0.2–0.4 EL for different probe signals during nc11–13. **c-h** Data averaged from ≥5 embryos for each nuclear cycle. P-values were from Student's t-test: *, $p<0.05$; **, $p<0.01$.

Fig. 2 P1 contributes a modest yet non-negligible fraction of early *hb* transcription.

a Confocal image of a wild-type *Drosophila* embryo colabeled for P1-5'UTR, P2-3'UTR, CDS of *hb* mRNA, and DNA at nc13. Scale bar, 50 μm. Inset, magnified view of anterior nuclei. Scale bar, 5 μm. **b** The numbers of nascent mRNAs for P1-5'UTR, P2-3'UTR, and CDS signals at individual active *hb* gene loci were plotted against each other and fitted to a linear function. Data from a single nc12 embryo. **c** Parameters of the linear fit between P1-5'UTR, P2-3'UTR, and CDS signals during nc11–13. Data averaged from two embryos for nuclear cycle 11 and seven embryos each for nuclear cycles 12 and 13. Error bars represent s.e.m. **d** Schematic of P1 and P2 transcription. Nascent mRNAs were elongated on the gene with a constant speed and stayed on the transcription site for an extra period before being released. The P1-specific intron was spliced during or after transcription. **e** The contributions of P1 and P2 transcription to the nascent CDS signal at individual *hb* loci as a function of the AP position during nc11–13. Marked region, 0.2–0.4 EL. **f** Average P1 and P2 contributions to the nascent CDS signal at individual *hb* loci in the position range of 0.2–0.4 EL during nc11–13. **g** The intrinsic noise for P1-5'UTR, P2-3'UTR, and CDS signals at individual *hb* gene loci in the position range of 0.2–0.4 EL during nc11–13. Error bars represent s.e.m. P-values were from Student's t-test: *, $p<0.05$; **, $p<0.01$. **h** The contributions of P1 and P2 to the intrinsic noise of the nascent CDS signal at individual *hb* gene loci in the position range of 0.2–0.4 EL during nc11–13. **e-h** Data averaged from ≥5 embryos for each nuclear cycle.

Fig. 3 Bcd activates P1 and P2 through different regulatory relations.

a The average number of nascent mRNAs at individual *hb* loci for P1-5'UTR and P2-3'UTR signals as a function of the AP position in a 1× *bcd* (solid line) embryo was compared with that in the wild-type (dashed line, data averaged from five embryos). Shadings indicate s.e.m. **b** The boundary position of the anterior expression domain for P1-5'UTR and P2-3'UTR signals in wild-type and 1× *bcd* embryos during nc11–13. Data averaged from ≥5 wild-type embryos and four 1× *bcd* embryos for each nuclear cycle. Error bars represent s.e.m. P-values were from Student's t-test: *, $p<0.05$; **, $p<0.01$. **c** The number of nascent mRNAs at individual *hb* loci for P1-5'UTR (upper) and P2-3'UTR (lower) signals was plotted against nuclear Bcd concentration (single wild-type embryo, >1200 nuclei, 0.25–0.75 EL). The single-locus data were binned along the Bcd axis (mean ± s.e.m.) and fitted to Hill functions. **d** Hill coefficient of the gene regulation function for different probe signals during nc11–13. Error bars represent s.e.m. P-values were from Student's t-test: *, $p<0.05$; **, $p<0.01$. **e** The ratios of the concentration threshold between P1-intron and CDS signals, and between P1-5'UTR and P2-3'UTR signals, during nc11–13. Error bars represent s.e.m. **f** The average number of Bcd molecules bound at P1- and P2-active *hb* loci in the anterior expression domain (0.2–0.35 EL) during nc11–13. Error bars represent s.e.m. Inset: an anterior nucleus labeled for P1-5'UTR and P2-3'UTR of *hb* mRNAs and Bcd protein. The enriched Bcd signal in the vicinity of the promoter (yellow and purple circles) was measured. Scale bar, 2 μm. **g** Bcd binding at P1- and P2-active *hb* loci as a function of nuclear Bcd concentration at nc12.

The binned data were fitted to multi-Hill functions. Dashed lines highlight discrete binding plateaus for each promoter. Error bars represent s.e.m. **d-g** Data averaged from ≥5 embryos for each nuclear cycle.

Fig. 4 Two enhancers combine differently to drive P1 and P2 activation.

a Schematic of *hb* reporter constructs with enhancer replacements. A construct without enhancer deletion was used as a control. Two smFISH probe sets were used to label different regions of reporter mRNAs: blue, P1-intron probes; red, *yellow* probes. **b** Confocal image of a distal-enhancer-removed embryo labeled for P1-intron, *yellow*, and *hb* CDS at nc12. Scale bar, 50 μm. Inset, magnified view of a single anterior nucleus. Scale bar, 2 μm. **c** Percentage of active reporter gene loci as a function of the AP position for P1-intron and *yellow* signals in nc12 embryos of different constructs. Shadings indicate s.e.m. **d** The average percentage of active reporter gene loci in the position range of 0.2–0.4 EL for P1-intron and *yellow* signals in different constructs during nc11–13. Error bars represent s.e.m. **e** The number of nascent mRNAs at individual reporter gene loci as a function of the AP position for P1-intron and *yellow* signals in nc12 embryos of different constructs. Shadings indicate s.e.m. **f** The maximal transcription level of the anterior expression domain for P1-intron and *yellow* signals in different constructs during nc11–13. Error bars represent s.e.m. **g, h** The boundary shift of the anterior expression for *yellow* (**g**) and P1-intron (**h**) signals upon removing one enhancer. Error bars represent s.e.m. Right: schematic of boundary shift. **c-h** Data compared between the distal- or proximal-enhancer-removed constructs

and their controls. Data averaged from ≥4 embryos for each reporter construct and each nuclear cycle. P-values were from Student's t-test: *, $p<0.05$; **, $p<0.01$.

Fig. 5 Three-state promoter kinetics reveal a unified scheme of P1 and P2 regulation.

a Histograms of nascent mRNA at individual *hb* gene loci for P1-5'UTR (upper) and P2-3'UTR (lower) signals in different position ranges (single embryo, the center of each position range is indicated above the histogram). Each histogram was fitted to a three-state transcription model (right). **b** Histograms of *hb* nascent mRNA at active sister loci for P1-5'UTR (upper) and P2-3'UTR (lower) signals in the position range of 0.2–0.4 EL. Each histogram was fitted to two Poisson distributions. Data pooled from seven embryos at nc13. Insets, images of a single anterior nucleus with active sister loci pairs. **c** Promoter activation rates for P1 and P2 estimated from five embryos at nc12 were plotted against nuclear Bcd concentration and fitted to Hill functions. **d, e** The Hill coefficients (**d**) and concentration thresholds (**e**) of promoter activation rates for P1 and P2 during nc11–13. Error bars represent s.e.m. **f** The transcription initiation rates of the two active states for P1 and P2 during nc11–13. Error bars represent s.e.m. **g** The contributions of states 1 and 2 to P1 and P2 transcription in the position range of 0.2–0.4 EL during nc11–13. Error bars represent s.e.m. **h** P1 activation rates as functions of the AP position for different constructs at nc12 (≥4 embryos for each reporter construct). Data were normalized and compared between the distal- or proximal-enhancer-removed constructs and the control. **i** Schematic of P1 and P2 regulation by cooperative Bcd binding at the proximal and distal enhancers. **j** The joint distribution of nascent

P1-5'UTR and P2-3'UTR signals at individual *hb* gene loci in the position range of 0.2–0.4 EL compared with model prediction. Data pooled from 17 embryos during nc11–13. White circles, individual loci. Color code, probability estimated from a unified model of P1 and P2 transcription. **d-g** Data averaged from ≥5 embryos for each nuclear cycle.

Figure 1

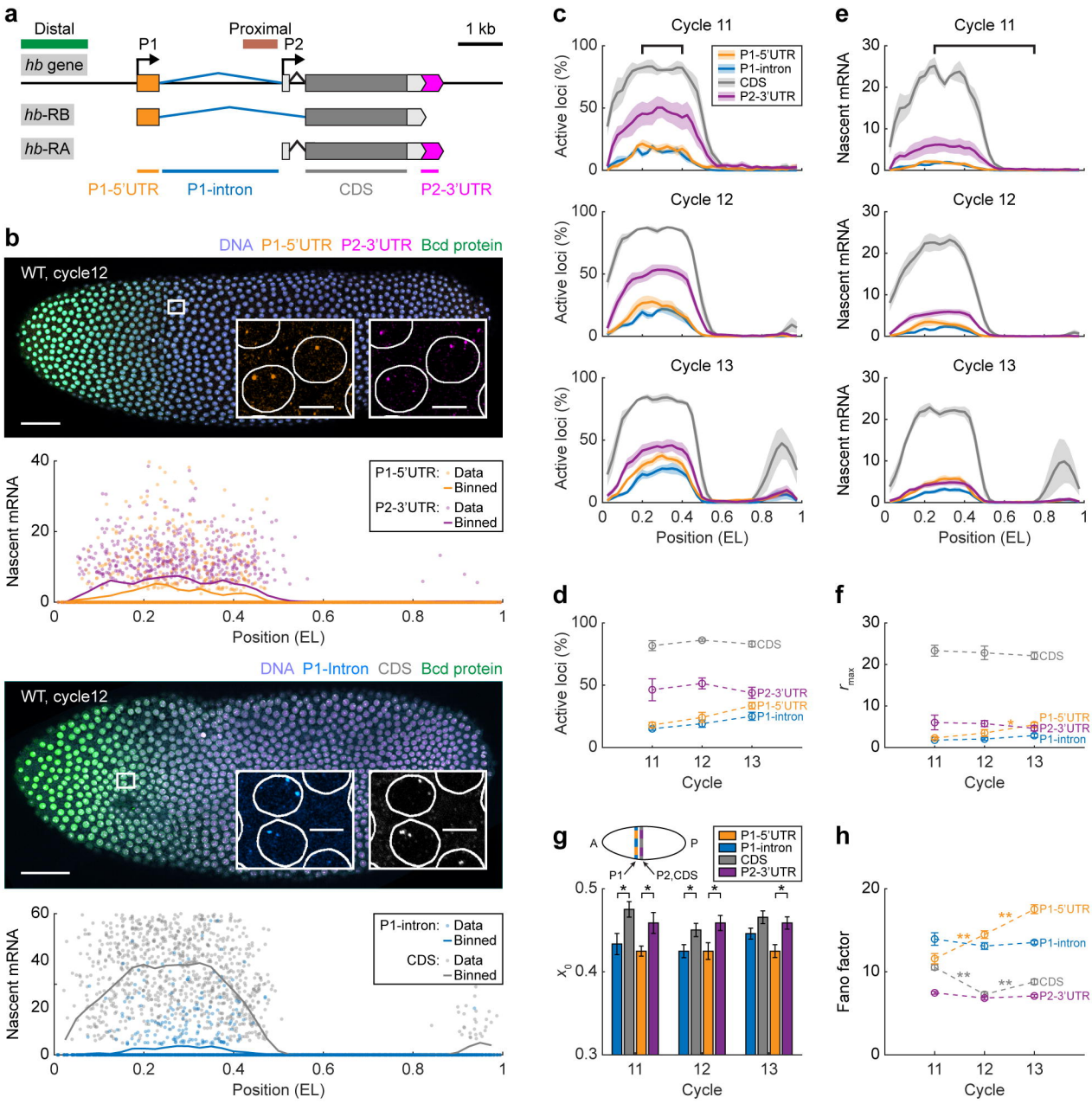

a
Distal
Proximal
1 kb
P1
P2
hb gene
hb-RB
hb-RA
P1-5'UTR
P1-intron
CDS
P2-3'UTR
b
DNA P1-5'UTR P2-3'UTR Bcd protein
WT, cycle12
Nascent mRNA
Position (EL)
P1-5'UTR: Data Binned
P2-3'UTR: Data Binned
DNA P1-Intron CDS Bcd protein
WT, cycle12
P1-intron: Data Binned
CDS: Data Binned
c
Cycle 11
Cycle 12
Cycle 13
Active loci (%)
Position (EL)
P1-5'UTR
P1-intron
CDS
P2-3'UTR
e
Nascent mRNA
d
Active loci (%)
Cycle
f
r_{max}
g
A
P
P1
P2,CDS
x_0
h
Fano factor

Figure 2

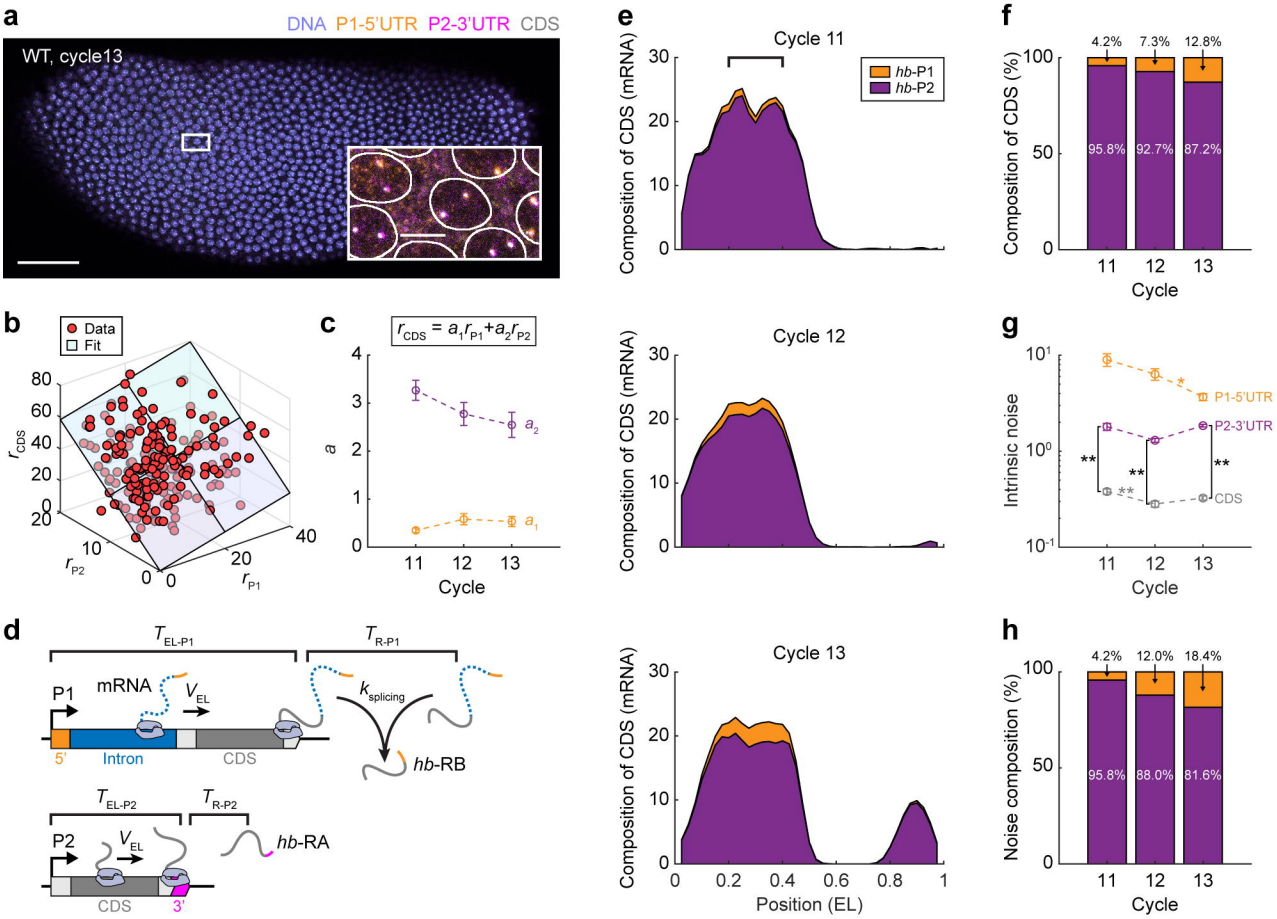

a
DNA P1-5'UTR P2-3'UTR CDS
WT, cycle13
b
Data
Fit
r_{CDS}
r_{P2}
r_{P1}
c
$r_{CDS} = a_1 r_{P1} + a_2 r_{P2}$
a
a_2
a_1
Cycle
d
T_{EL-P1}
T_{R-P1}
P1
mRNA
V_{EL}
$k_{splicing}$
5'
Intron
CDS
hb-RB
T_{EL-P2}
T_{R-P2}
P2
hb-RA
3'
e
Cycle 11
Cycle 12
Cycle 13
Composition of CDS (mRNA)
hb-P1
hb-P2
Position (EL)
f
4.2%
7.3%
12.8%
95.8%
92.7%
87.2%
Composition of CDS (%)
Cycle
g
Intrinsic noise
P1-5'UTR
P2-3'UTR
CDS
Cycle
h
4.2%
12.0%
18.4%
95.8%
88.0%
81.6%
Noise composition (%)
Cycle

Figure 3

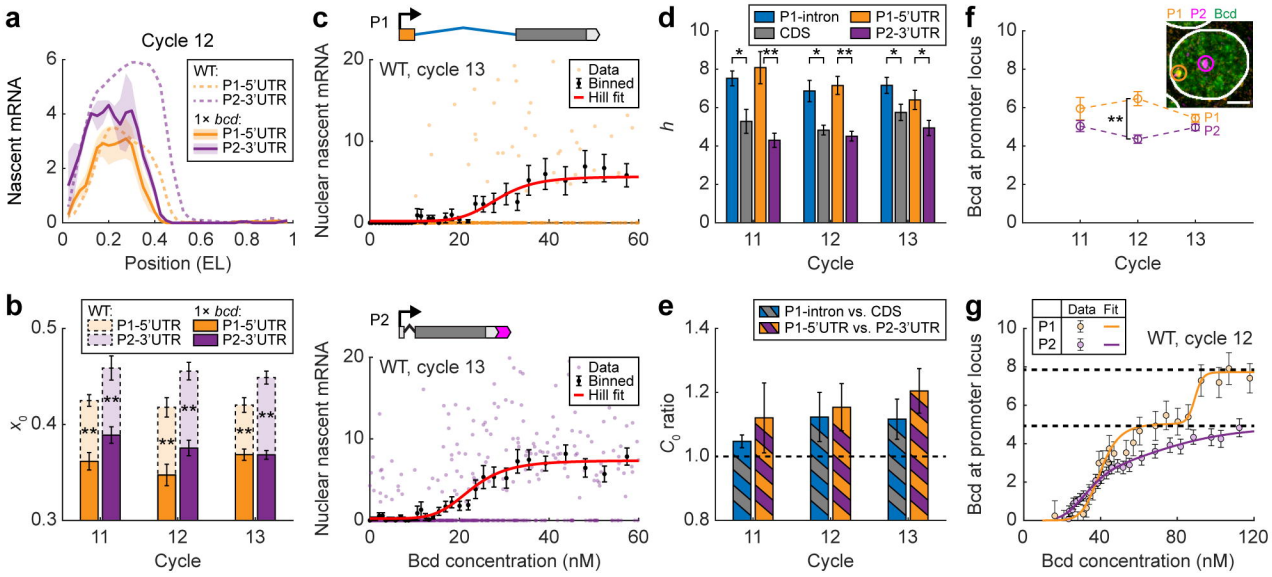
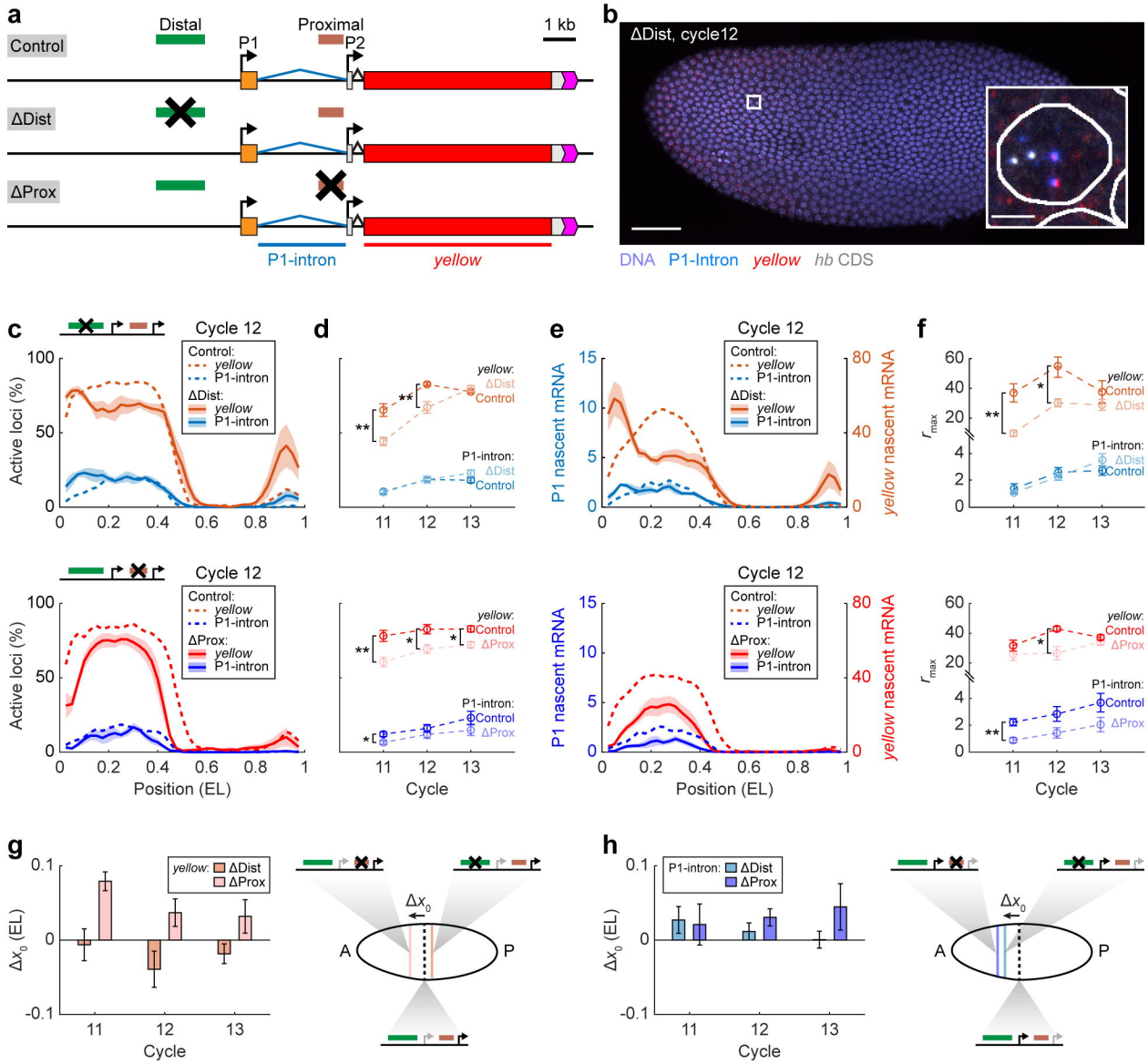
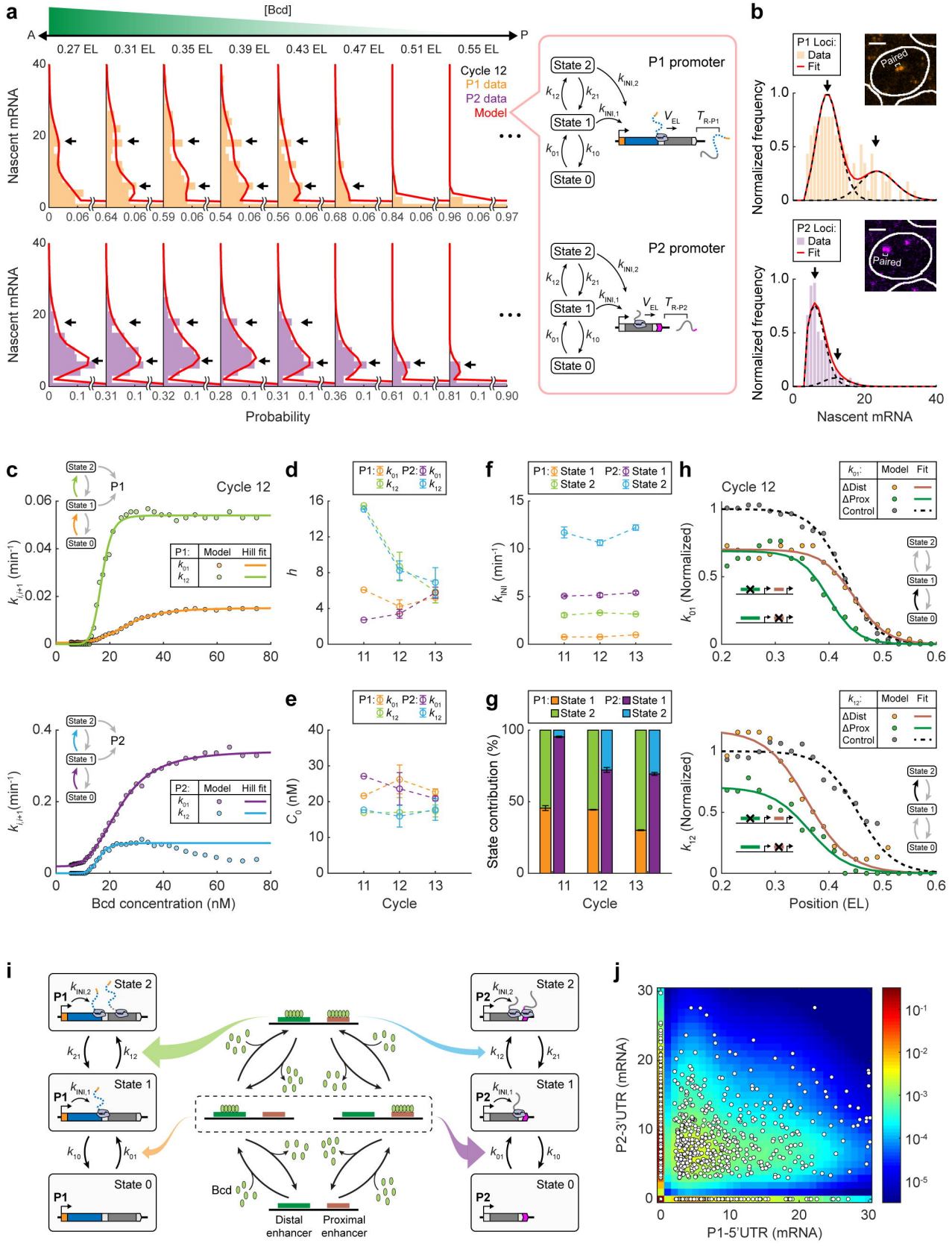

a
Cycle 12
Nascent mRNA
Position (EL)
WT:
P1-5'UTR
P2-3'UTR
1× bcd:
P1-5'UTR
P2-3'UTR
b
WT:
P1-5'UTR
P2-3'UTR
1× bcd:
P1-5'UTR
P2-3'UTR
Cycle
c
P1
WT, cycle 13
Data
Binned
Hill fit
Nuclear nascent mRNA
P2
WT, cycle 13
Bcd concentration (nM)
d
P1-intron
CDS
P1-5'UTR
P2-3'UTR
h
Cycle
e
P1-intron vs. CDS
P1-5'UTR vs. P2-3'UTR
Cycle
f
P1 P2 Bcd
P1
P2
Bcd at promoter locus
Cycle
g
Data Fit
P1
P2
WT, cycle 12
Bcd at promoter locus
Bcd concentration (nM)

Figure 4



a
Distal
Proximal
1 kb
Control
P1
P2
ΔDist
ΔProx
P1-intron
yellow
b
ΔDist, cycle12
DNA
P1-Intron
yellow
hb CDS
c
Cycle 12
Control:
yellow
P1-intron
ΔDist:
yellow
P1-intron
Active loci (%)
Position (EL)
d
yellow:
ΔDist
Control
P1-intron:
ΔDist
Control
Cycle
e
Cycle 12
P1 nascent mRNA
yellow nascent mRNA
Position (EL)
f
yellow:
Control
ΔDist
P1-intron:
ΔDist
Control
Cycle
Cycle 12
ΔProx:
yellow
P1-intron
Position (EL)
Cycle
g
yellow: ΔDist
ΔProx
Δx0 (EL)
Cycle
A
P
h
P1-intron: ΔDist
ΔProx
Δx0 (EL)
Cycle
A
P

Figure 5


a
[Bcd]
A
P
0.27 EL
0.31 EL
0.35 EL
0.39 EL
0.43 EL
0.47 EL
0.51 EL
0.55 EL
Cycle 12
P1 data
P2 data
Model
Nascent mRNA
Probability
P1 promoter
P2 promoter
State 2
State 1
State 0
b
P1 Loci:
Data
Fit
Paired
P2 Loci:
Normalized frequency
Nascent mRNA
c
Cycle 12
P1
P2
Model
Hill fit
Bcd concentration (nM)
d
e
Cycle
f
g
State contribution (%)
h
ΔDist
ΔProx
Control
Position (EL)
i
Bcd
Distal enhancer
Proximal enhancer
j
P2-3'UTR (mRNA)
P1-5'UTR (mRNA)