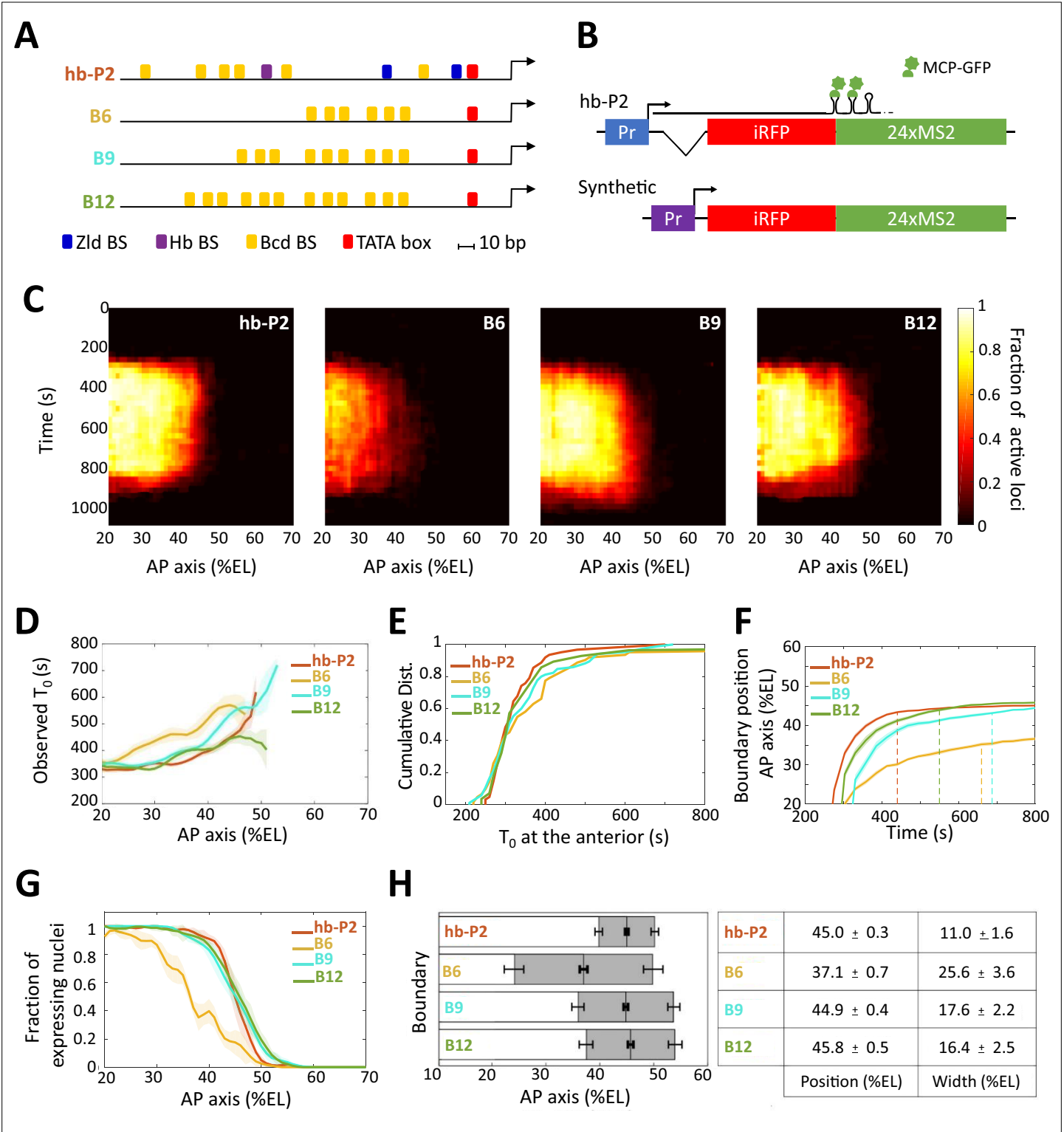


---

## Figures and figure supplements

Synthetic reconstruction of the *hunchback* promoter specifies the role of Bicoid, Zelda and Hunchback in the dynamics of its transcription

**Gonçalo Fernandes et al**

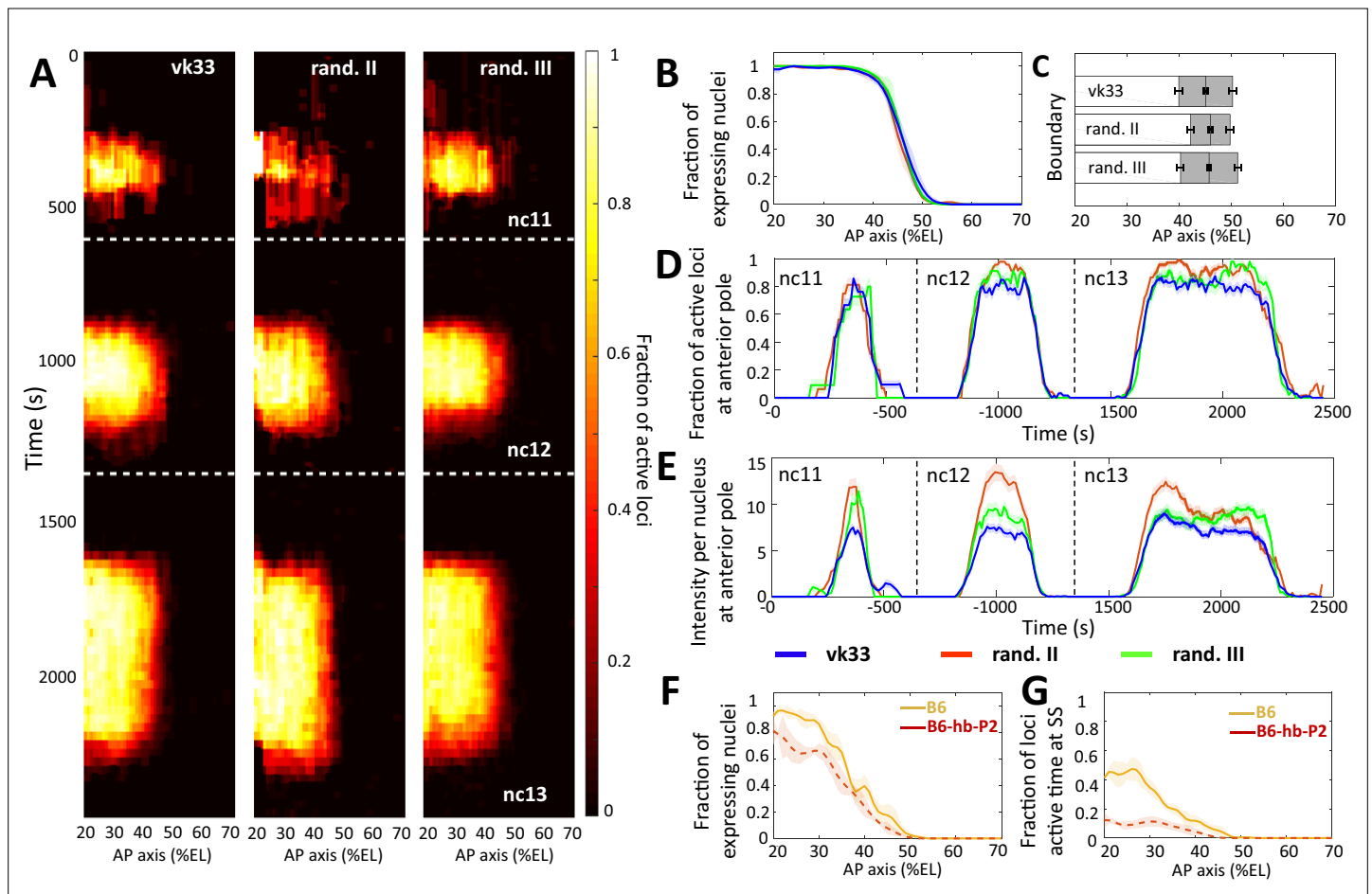


**Figure 1.** Transcription dynamics of the hb-P2, B6, B9 and B12 reporters. **(A)** Arrangement of the binding sites for Bcd (yellow), Hb (purple), and Zld (blue) upstream of the TATA box (red) and the TSS (broken arrow) of each reporter. **(B)** The MS2 reporters express the iRFP coding sequence followed by the sequence of the 24 MS2 stem loops. In the hb-P2 reporter, the hb-P2 promoter, 5'UTR sequence of the endogenous *hb* and its intron are placed just upstream of the iRFP sequence. In the synthetic reporters, the minimal promoter of the *hsp70* gene was used. Of note, replacing the minimal promoter of *hsp70* in B6 by the *hb* minimal promoter leads to a reporter with lower activity (**Figure 1—figure supplement 1**, (F–G)). **(C)** Kymographs of mean fraction of active loci (colormap on the right) as a function of time (Y axis in s) and nuclei position along the AP axis (X axis in %EL) at nc13. **(D)** Along the

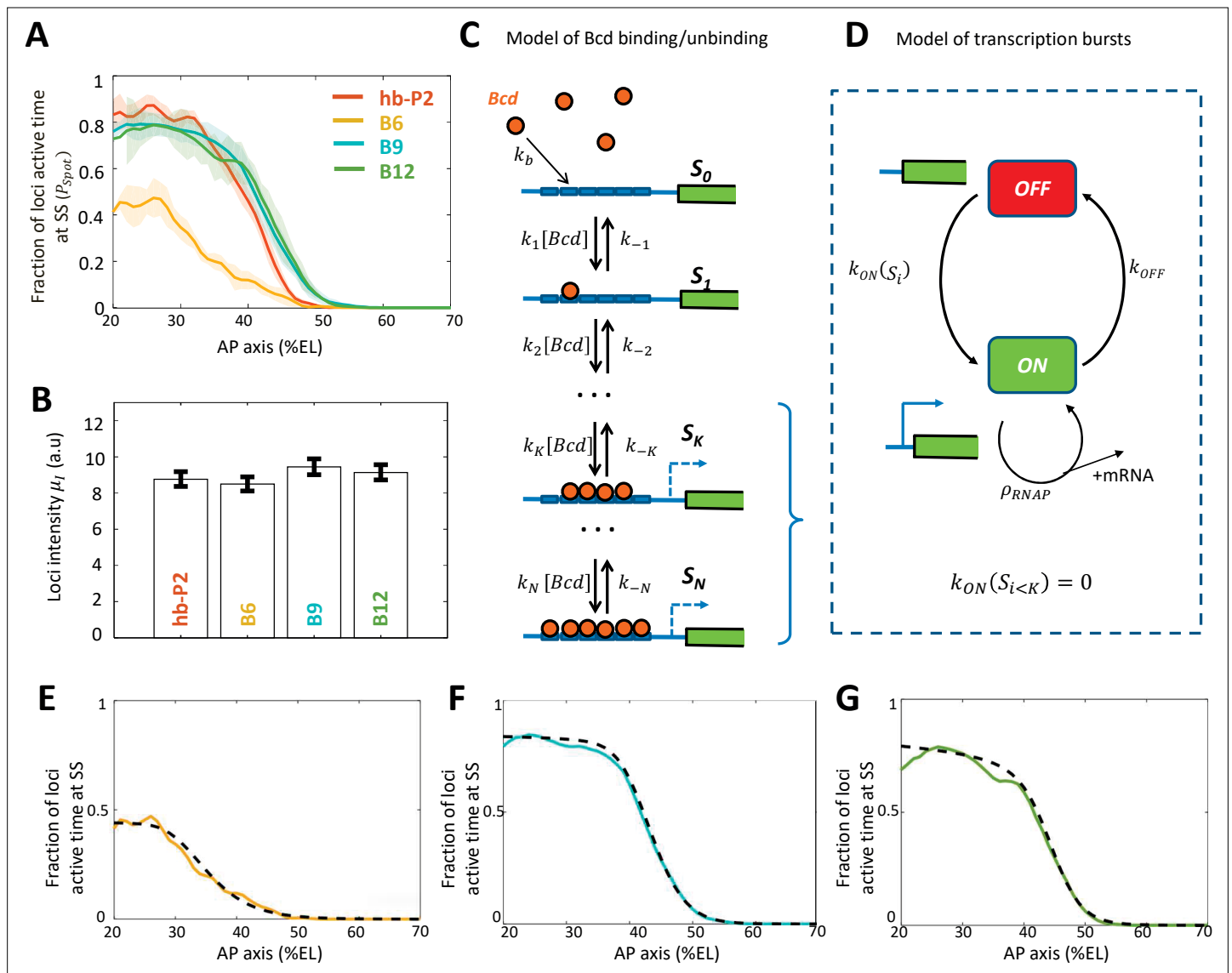
Figure 1 continued on next page

## Figure 1 continued

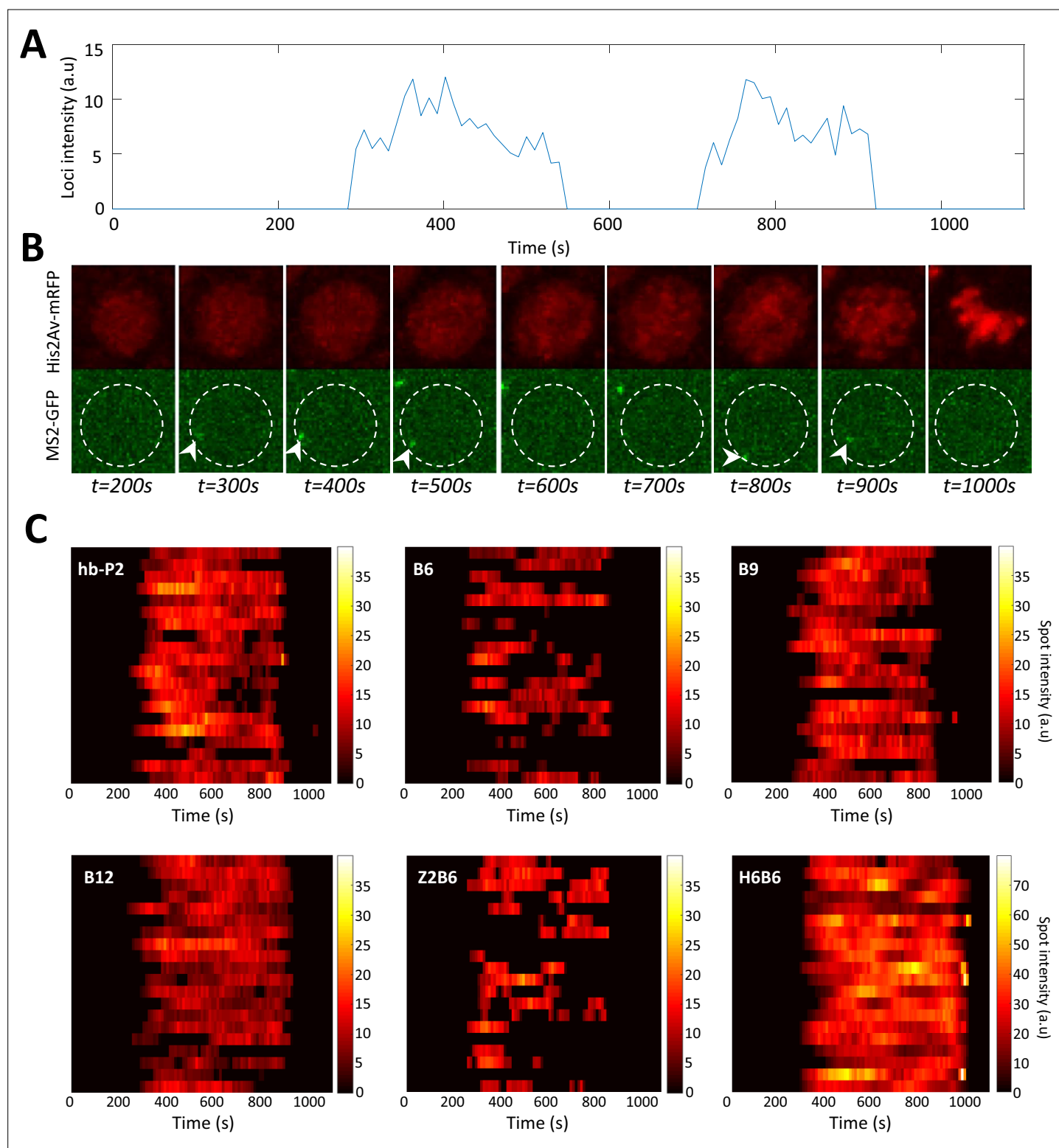
AP axis (%EL), mean time of first spot appearance  $T_0$  (s) with shaded standard error of the mean and calculated only for loci with observed expression. (E) Cumulative distribution function of  $T_0$  (s) in the anterior ( $20\% \pm 2.5\%$  %EL). (F) Boundary position (%EL) of fraction of nuclei with MS2 signal along AP axis, with shaded 95% confidence interval, as a function of time. The dash vertical lines represent the time to reach the final decision boundary position ( $\pm 2\%$  %EL). (G) Fraction of nuclei with any MS2 signal, averaged over  $n$  embryos, with shaded standard error of the mean, along the AP axis (%EL), at nc13. (H) Boundary position and width were extracted by fitting the patterns fraction of expressing nuclei, (G) with a sigmoid function. Bar plots with 95% confidence interval for boundary position and width as the gray region placed symmetrically around the boundary position. Average values and confidence intervals are indicated in the adjacent table. (D–H) reporter data are distinguished by color: hb-P2 (orange,  $n = 5$  embryos), B6 (yellow,  $n = 5$  embryos), B9 (cyan,  $n = 6$  embryos), and B12 (green,  $n = 4$  embryos).



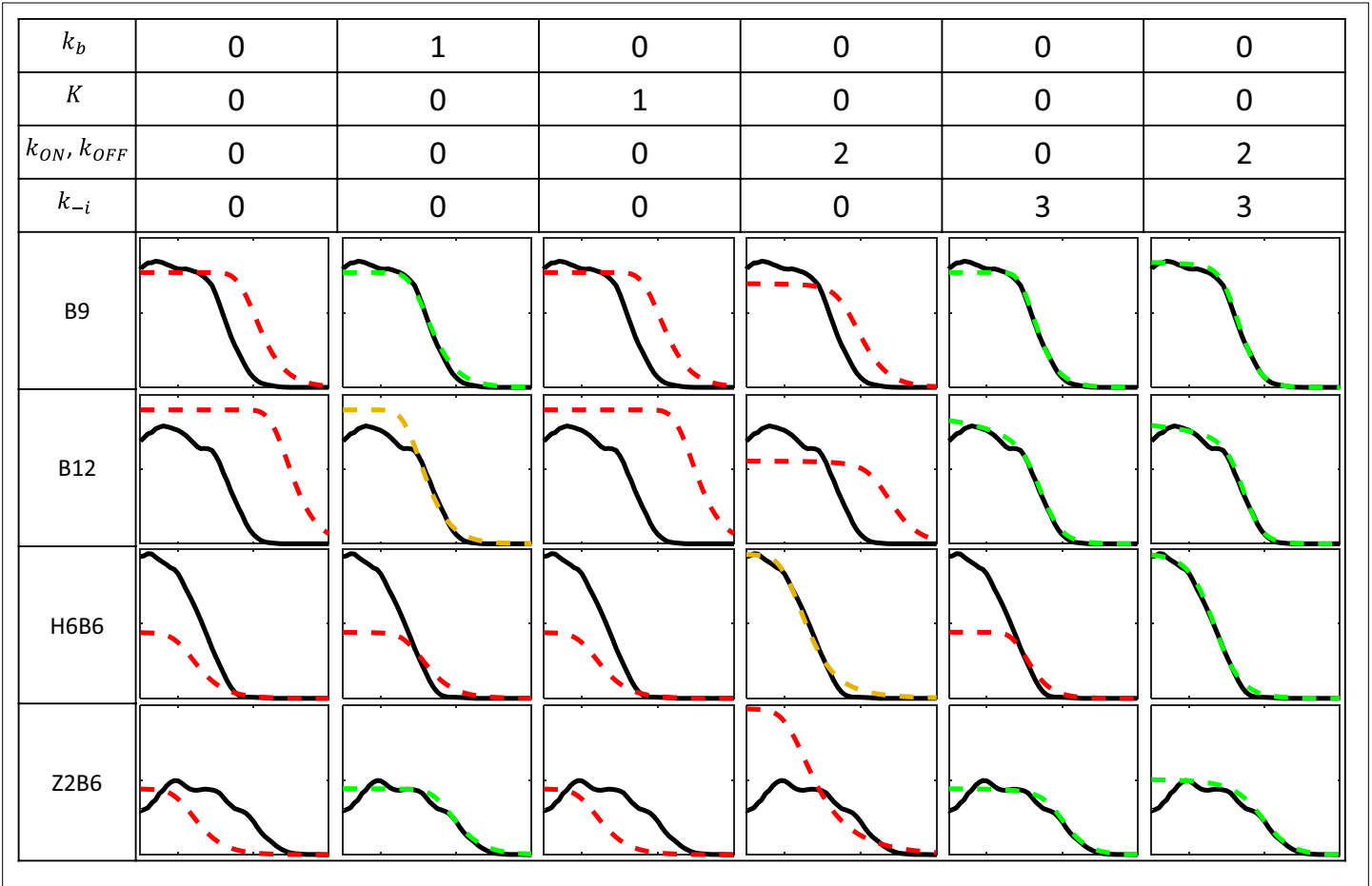
**Figure 1—figure supplement 1. Comparison of transcription dynamics with hb-P2 reporters.** inserted at the vk33 site on the 3rd chromosome and randomly in the 2nd chromosome and 3rd chromosome (Lucas *et al.*, 2018). **(A)** Kymographs representing the fraction of nuclei with active MS2 loci (represented as by the colormap) as a function of time and nuclei position along the AP axis. The white dashed horizontal lines represent mitoses between nuclear cycles. **(B)** Fraction of nuclei with any MS2 expression in nc13, averaged over multiple embryos, with shaded error of the mean as a function of nuclei position along AP axis (%EL). **(C)** Bar plots with 95% confidence interval for the expression boundary position for MS2 reporters, based on the fraction of expressing loci (panel B). For each reporter, also shown is the boundary width as the gray region placed symmetrically around the boundary position. **(D)** Time evolution of the fraction of active MS2 loci near the anterior pole (from 15% to 25%EL) in nc11, nc12, and nc13. **(E)** Time evolution of the MS2 locus intensity per nucleus at the anterior pole (from 15% to 25%EL) in nc11, nc12, and nc13. In panels B, E, and D, data is shown for the hb-P2 reporter inserted at the vk33 site (blue,  $n = 5$  embryos), randomly in the 2nd chromosome (red,  $n = 3$ ) and in the 3rd chromosome (green,  $n = 6$ ). **(F)** Fraction of nuclei with MS2 expression, averaged over multiple embryos, with shaded standard error of the mean, during nc13 along the AP axis (%EL). **(G)** Fraction of MS2 loci active time at steady state, averaged over multiple embryos, with shaded standard error of the mean, during nc13 along the AP axis (%EL). In panel F and G, data shown for B6 reporters with TATA box of HSBG promoter (solid orange) and of hb-P2 promoter (dashed red).



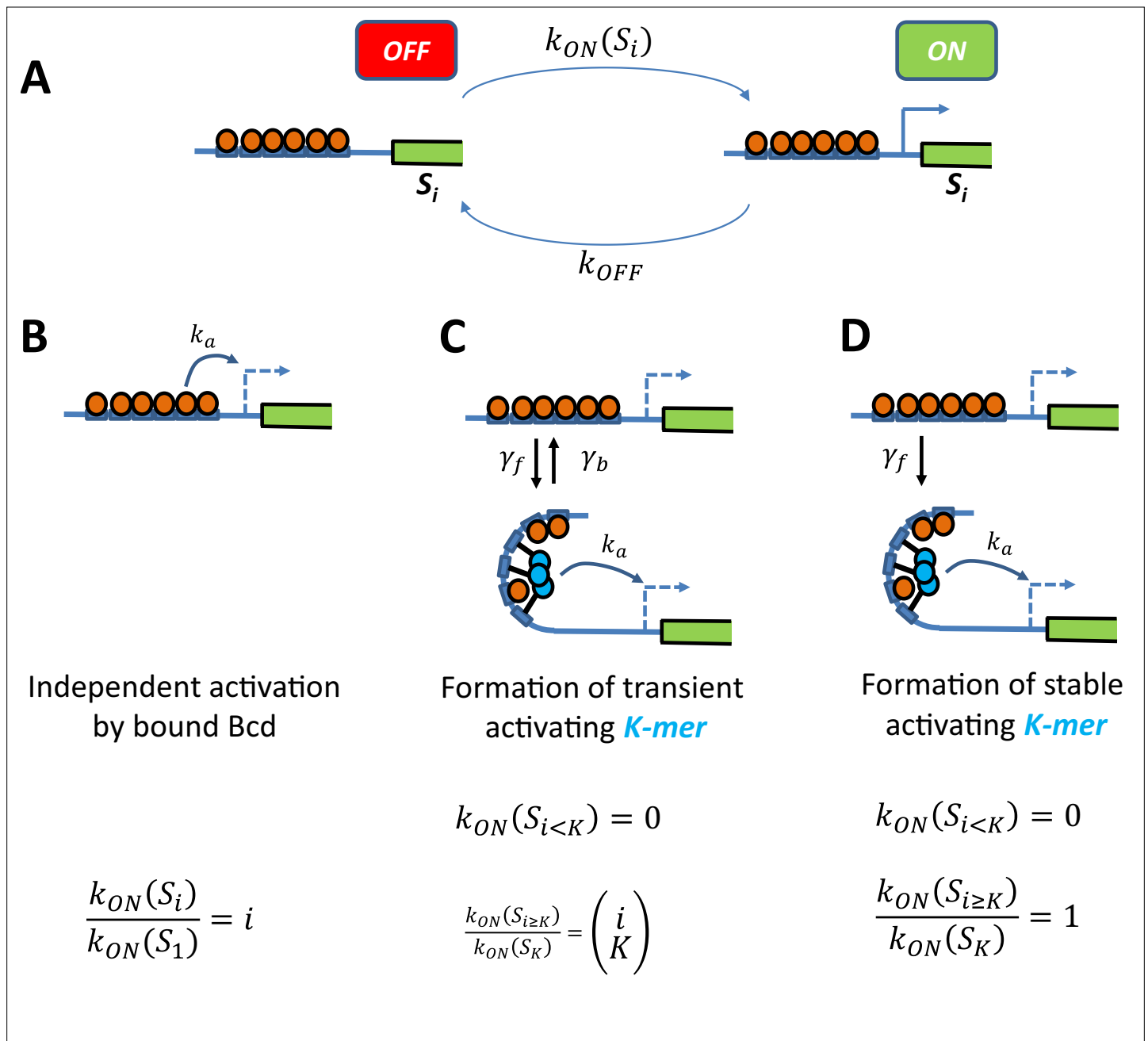
**Figure 2.** Modeling transcription dynamics at steady state. **(A)** Fraction of loci active time ( $P_{spot}$ ) at steady state (time window of 600–800 s into nc13), averaged over  $n$  embryos, as a function of nuclei position along AP axis (%EL). **(B)** Mean fluorescent intensity ( $\mu_I$ ) with standard error of active MS2 loci detected in the anterior region ( $\sim 20\% \pm 2.5\%$  EL) at steady state. In **(A–B)** reporter data are distinguished by color: hb-P2 (orange,  $n = 5$  embryos), B6 (yellow,  $n = 5$  embryos), B9 (cyan,  $n = 6$  embryos), and B12 (green,  $n = 4$  embryos). **(C)** Model of Bicoid binding and unbinding to an array of  $N$  identical binding sites: nuclear Bcd molecules can bind independently to individual binding sites at rate  $k_b$ . The binding array state is denoted by  $S_i$  where  $i$  is the number of bound sites. The forward rate constants  $k_i$  are the binding rates of Bcd to the free remaining sites of  $S_{i-1}$ :  $k_i = (N - i + 1) k_b$ . The backward rate constants  $k_{-i}$  are the unbinding rates of bound Bcd from  $S_i$ . **(D)** Transcription dynamics is modeled as a bursty two-state ON/OFF model with the switching rate constants  $k_{ON}(S_i)$  and  $k_{OFF}$ . The switching rate  $k_{ON}(S_i)$  depends on  $i$  the number of bound Bcd molecules. Transcription is not activated with fewer than  $K$  bound Bcd ( $k_{ON}(S_{i < K}) = 0$ ). Only during the ON state can RNAPs arrive and initiate transcription at rate constant  $\rho_{RNAP}$ . **(E–G)** Fraction of active loci at steady state obtained experimentally for B6 (E, solid yellow), B9 (F, solid cyan), B12 (G, solid green) compared to the fraction of active loci at steady state from the best fitting models (dashed black) for corresponding BS numbers  $N=6$  for B6 (E),  $N=9$  for B9 (F) or  $N=12$  for B12 (G). In these models, the free parameters are the unbinding rate constant ( $k_{-i}$ ), the promoter switching rates with  $K$  bound Bcd molecules ( $k_{ON}(S_K)$  and  $k_{OFF}$ ).  $K$  is set to 3. The switching ON rates at higher bound states are set  $k_{ON}(S_{i > K}) = k_{ON}(S_K) \left(\frac{i}{K}\right)$ , given the synergistic activation of transcription by bound Bcd (see **Supplementary file 5**). The binding rate constant  $k_b$  is determined by assuming that Bcd binding is diffusion limited (Appendix 2).



**Figure 2—figure supplement 1.** Example of timely transcription dynamics (bursting) monitored with MS2-MCP system. **(A)** Example of a fluorescent intensity trace as a function of time for the B6 reporter in a nucleus in the anterior region (~20 %EL). **(B)** Image of the nucleus which exhibited the trace observed in panel A with hisAV-mRFP (red, top) and the MS2-GFP (green, bottom). Active transcription can be detected as a bright spot in the green channel (white arrowheads). The white dashed circle corresponds to the estimated nuclear envelope. **(C)** Kymographs of MS2-GFP intensity traces at the hb-P2 and the synthetic B6, B9, B12, H6B6, and Z2B6 reporters in the anterior region (~20%  $\pm$  2.5 %EL). For each construct, 20 traces are shown as horizontal bars with the intensity described by the color heatmap (right). Of note, the heatmap scale is shared by hb-P2, B6, B9, B12, and Z2B6 while it is twice higher for H6B6.

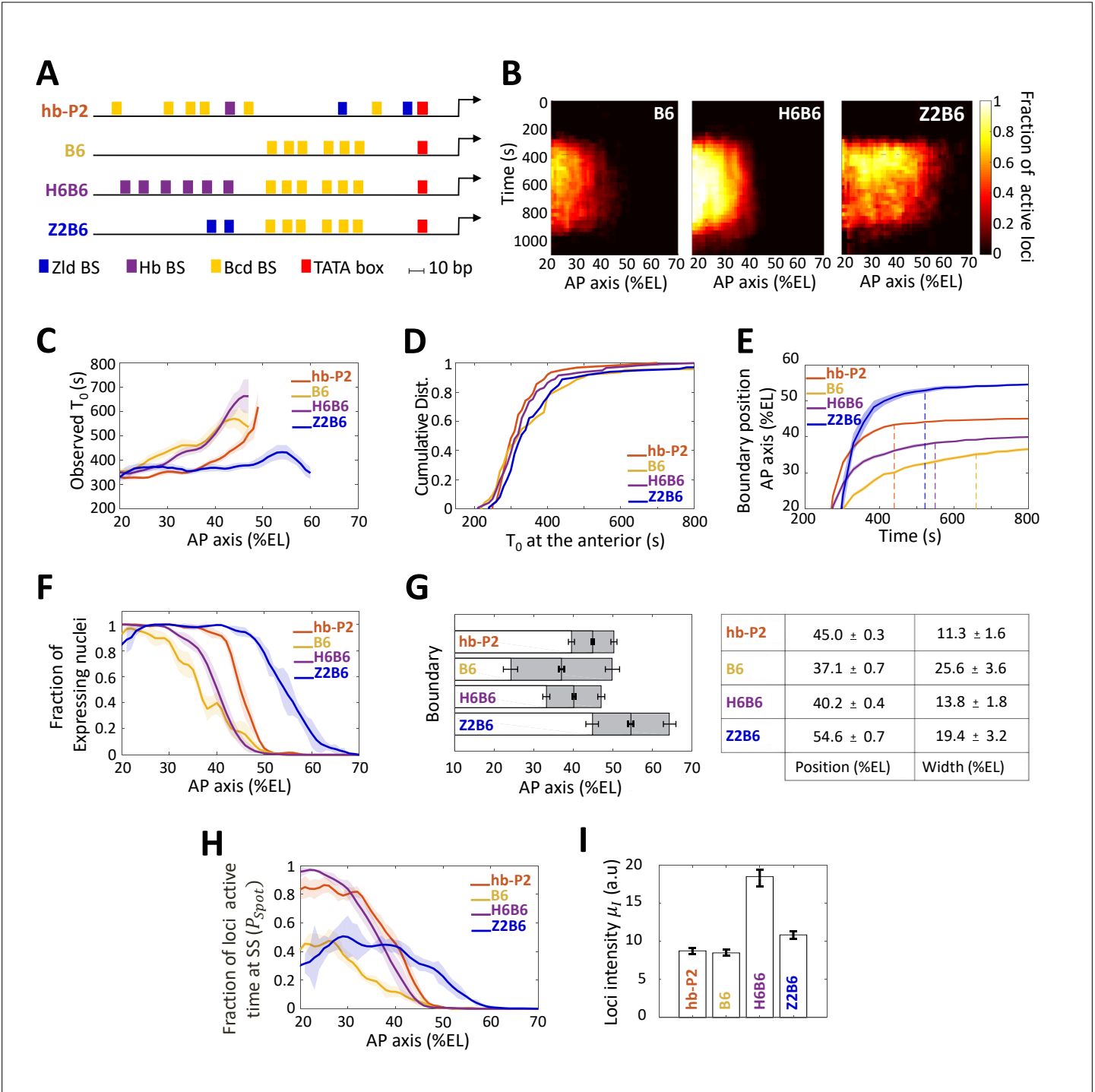


**Figure 2—figure supplement 2.** Fitting models to experimental data for the B9, B12, Z2B6, and H6B6 with the least degree of freedom when compared to B6. Each column corresponds to a fitting configuration with the number indicating the degree of freedom from the fitted B6 model for each group of kinetic parameters. A value of 0 indicates that the value of the kinetic parameter is imposed from the fitted model of B6 pattern (**Figure 2E**). Parameters include the Bcd binding rate constant  $k_b$ , the minimal number of bound Bcd for transcription activation  $K$ , the promoter switching rate when bound by  $K$  Bcd molecules  $k_{ON}(S_K)$  and  $k_{OFF}$ , and the unbinding rate constants of Bcd from the BS array  $k_{-i}$  (see Appendix 2). The switching ON rates at higher bound states are set to  $k_{ON}(S_{i>K}) = k_{ON}(S_K) \binom{i}{K}$ , given the synergistic activation of transcription by bound Bcd (see Appendix 3). For each fitting configuration and reporter are shown the patterns from the data (solid black) and from the fitted model (dashed color). On the y axis: fraction of loci active time at steady state (from 0 to 1). On the x axis: position along the AP axis (from -30 to 20%EL). The quality of the model fit to the data (visually determined) is indicated by the color of the dashed curves as follow: good (green), moderate (orange) and red (bad).



**Figure 2—figure supplement 3.** Schemes of transcription activation by bound Bcd molecules. **(A)** Context: When the binding sites are occupied by  $i$  Bcd molecules, the promoter can switch between two transcriptional states ON and OFF. The activation rate  $k_{ON}(S_i)$ , not the deactivation rate, depends on the binding state  $S_i$  and how it interacts with the promoter (see schemes in panel B, C and D). **(B)** Independent activation: Bound Bcd molecules (orange balls) can independently activate transcription at rate  $k_a$ . Activation can be enabled with even a single bound Bcd:  $k_{ON}(S_1) = k_a$ . The activation rate with binding state  $S_i$  is proportional to the number of bound Bcd:  $k_{ON}(S_i)/k_{ON}(S_1) = i$ . **(C-D)** Formation of a stable activating “*K-mer*”: Bound Bcd molecules (orange balls) can randomly form protein complexes (cyan balls) containing  $K$  molecules (hence called “*K-mers*”), which can activate transcription at rate  $k_a$ . There is no activation of transcription with less than  $K$  bound Bcd ( $k_{ON}(S_{i < K}) = 0$ ). In panel C, The “*K-mer*” is constantly formed and degraded at rates  $\gamma_f$  and  $\gamma_b$  respectively. Assuming that these rates are fast enough (so that the formation and degradation of “*K-mers*” are always in equilibrium) and that  $\gamma_b \ll \gamma_f$ , the fold change in the activation rate is given by the number of ways to choose a subset of  $K$  molecules from  $i$  bound Bcd:  $\frac{k_{ON}(S_i)}{k_{ON}(S_K)} = \binom{i}{K} = \frac{i!}{K!(i-K)!}$ . In panel D, the protein complex is stable once formed ( $\gamma_b = 0$ ). Assuming a negligible formation time when enough Bcd is bound ( $1/\gamma_f \approx 0$ ), the activation rate is always constant  $k_{ON}(S_{i \geq K}) = k_a$ .



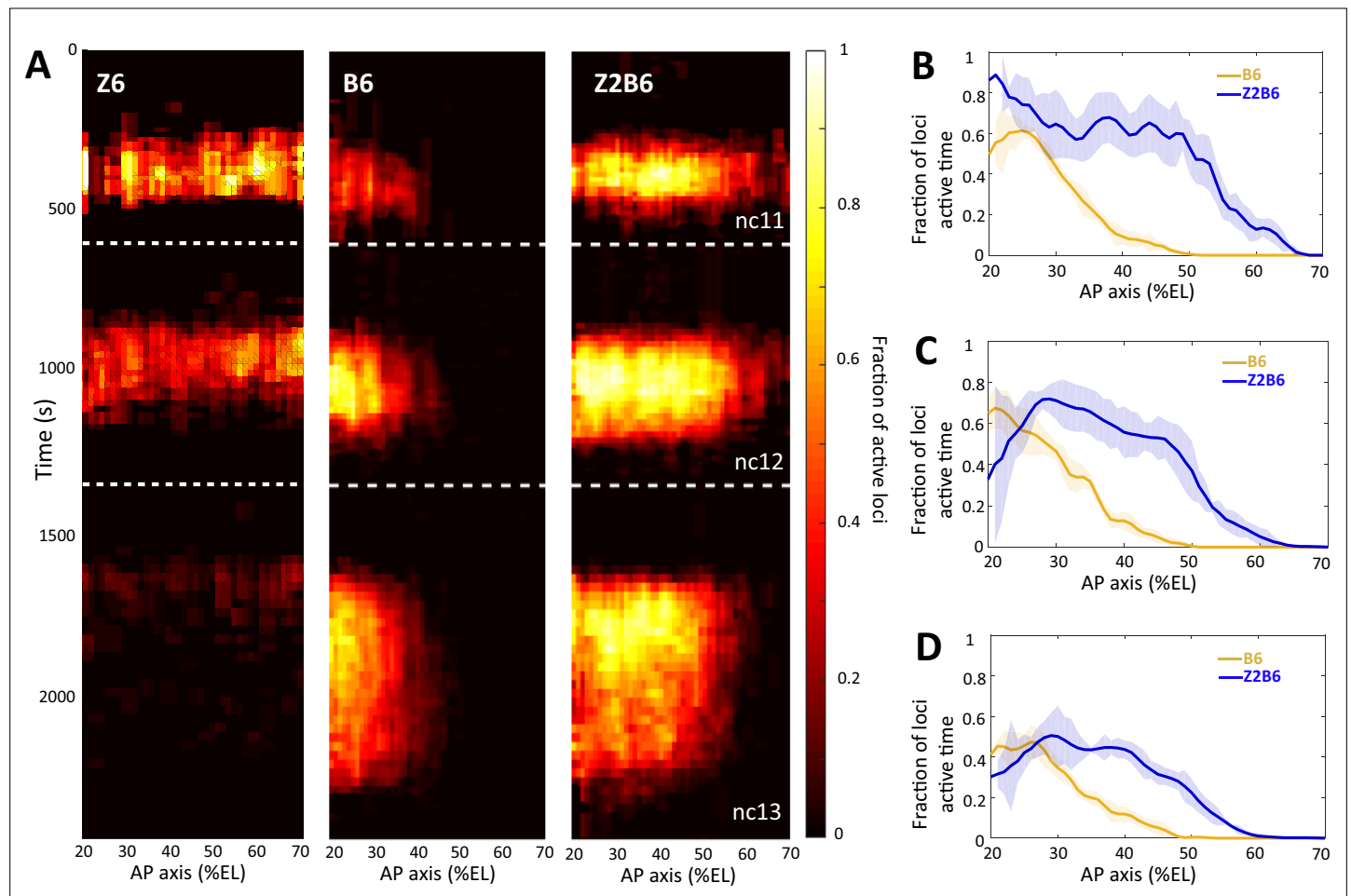


**Figure 3.** Transcription dynamics of the B6, H6B6, and Z2B6 reporters. **(A)** Arrangement of the binding sites for Bcd (yellow), Hb (purple), and Zld (blue) upstream of the TATA box (red) and the TSS (broken arrow) of each reporter. **(B)** Kymographs of mean fraction of active loci (colormap on the right) as a function of time (Y axis in s) and nuclei position along the AP axis (X axis in %EL) at nc13. **(C)** Mean time of first spot appearance  $T_0$  (s) along the AP axis with shaded standard error of the mean and calculated only for loci with observed expression. **(D)** Cumulative distribution function of  $T_0$  (s) at the anterior (20% ± 2.5%EL). **(E)** Boundary position (as %EL) of fraction of nuclei with MS2 signal along AP axis, with shaded 95% confidence interval, as a function of time. The dash vertical curves represent the time to reach the final decision boundary position (± 2 %EL). **(F)** Fraction of nuclei with any MS2 signal along the AP axis (%EL) with shaded standard error of the mean. **(G)** Boundary position and width were extracted by fitting the patterns fraction of expressing nuclei, **(F)** with a sigmoid function. Bar plots with 95% confidence interval for boundary position and width as the grey region placed symmetrically around the boundary position. Average values and confidence intervals are indicated in the adjacent table. **(H)** Fraction of loci active time at steady state ( $P_{Spot}$ ) at steady state (time window of 600–800 s into nc13) as a function of nuclei position along AP axis. **(I)** Mean intensity ( $\mu_I$ ) with standard error of

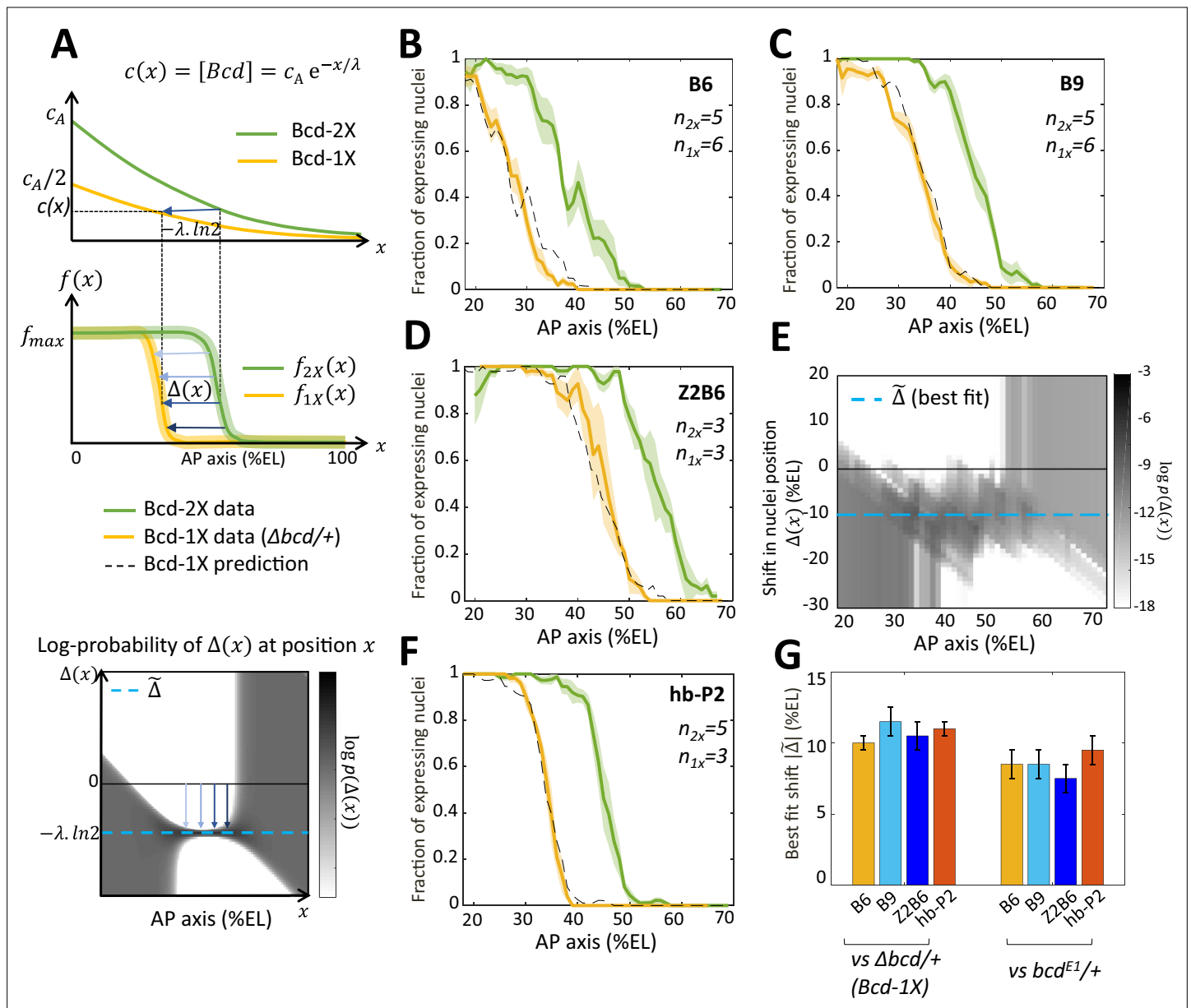
Figure 3 continued on next page

*Figure 3 continued*

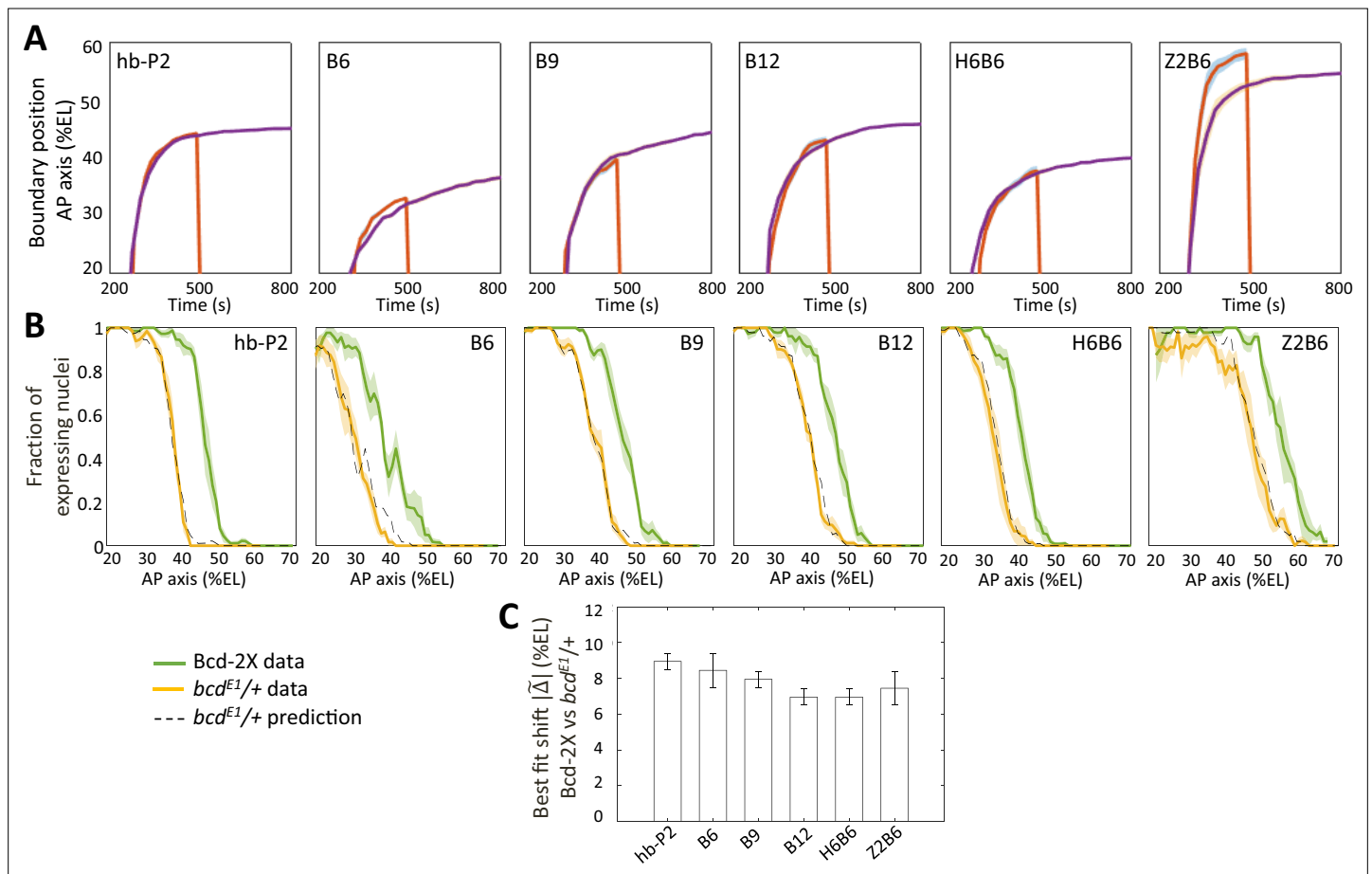
active fluorescent loci detected in the anterior region ( $\sim 20\% \pm 2.5\%$  EL) at steady state. (**C–I**) reporter data are distinguished by color: hb-P2 (orange, n = 5 embryos), B6 (yellow, n = 5 embryos), H6B6 (purple, n = 7 embryos), and Z2B6 (blue, n = 3 embryos).



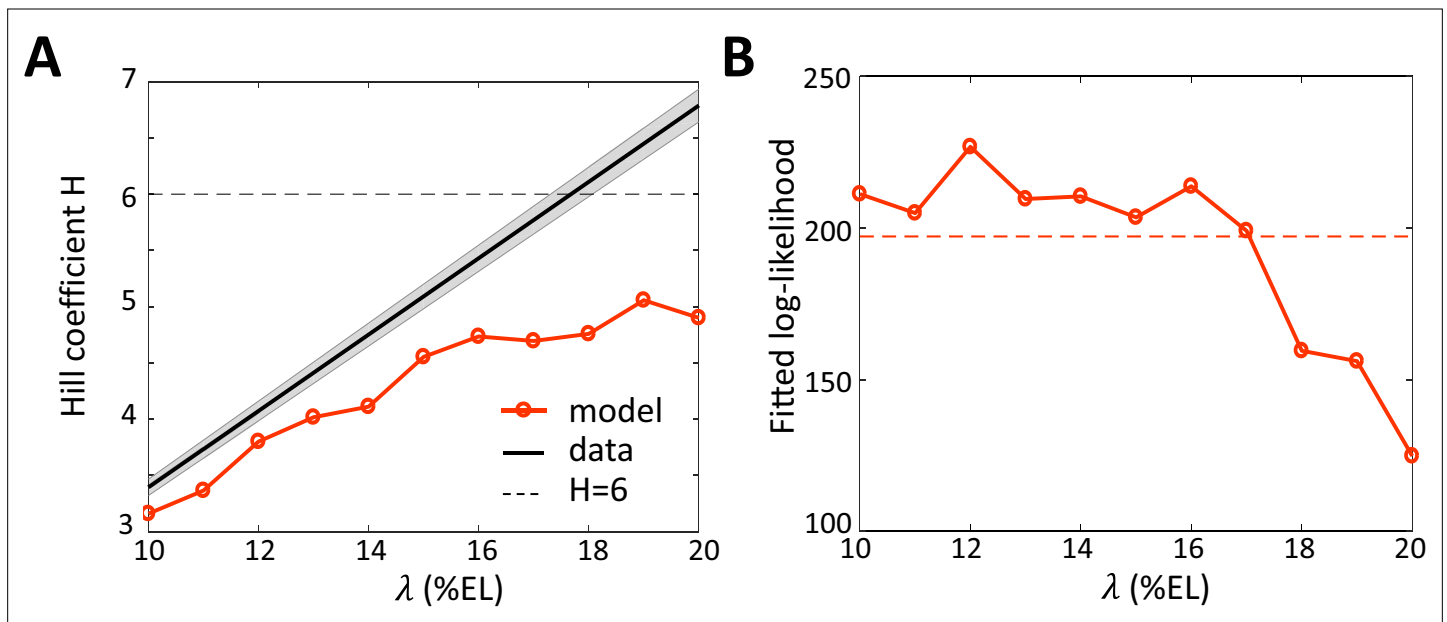
**Figure 3—figure supplement 1.** Transcription dynamics of the Z6 reporter ( $n = 2$  embryos), B6 ( $n = 5$  embryos) and Z2B6 ( $n = 3$  embryos) expression patterns. **A**) Kymographs representing the fraction of nuclei with active MS2 loci (represented by the colormap) as a function of time and nuclei position along the AP axis. The white dashed horizontal lines represent mitoses between nuclear cycles. **(B–D)** Comparison of B6 (orange) and Z2B6 (blue) patterns based on the fraction of loci active time, averaged over multiple embryos, with shaded standard error of the mean, in different time windows: **(B)** 450 s-550s into nc12, **(C)** 450 s-550s into nc13 and **(D)** 700 s-800s into nc13.



**Figure 4.** Bicoid thresholds measurements by the Bcd-only synthetic reporters. **(A)** Modeling the pattern shifts between Bcd-2X and Bcd-1X embryos. Top: The Bcd concentration gradient along the AP axis with its exponential decay length  $\lambda$ . At the anterior pole, Bcd concentration is  $c_A$  in Bcd-2X embryos (solid green line) and  $c_A/2$  in Bcd-1X embryos (solid yellow line). The distance between any two nuclei columns in Bcd-2X and Bcd-1X that have the same Bcd concentration (blue horizontal arrow) is given by  $-\lambda \ln 2$ . Middle: along the AP axis, expression pattern of a Bcd-dependent reporter in Bcd-2X embryos ( $f_{2X}(x)$ , solid green line) and in Bcd-1X embryos ( $f_{1X}(x)$ , solid yellow line).  $\Delta(x)$ : the shift in position (blue horizontal arrows) from a nuclei column in Bcd-2X embryos at position  $x$  to one at Bcd-1X embryos with the same expression level, such that  $f_{2X}(x) = f_{1X}(x - \Delta(x))$ . Bottom: Cartoon of log-probability map of the shift  $\Delta(x)$  based on the expression patterns in Bcd-2X and Bcd-1X (i.e.  $f_{1X}(x)$  and  $f_{2X}(x)$ ). Its value  $\log p(\Delta(x))$  is represented on the grey scale. The blue vertical arrows denoting the shift correspond to the horizontal arrows with similar shade observed in the middle panel. If the Bcd gradient is the only source of positional information for the expression patterns, then the best fit value of  $\Delta(x)$  given the probability map is  $\tilde{\Delta} = -\lambda \ln 2$  (horizontal blue dashed line). **(B–D and F)** Expression patterns of B6 **(B)**, B9 **(C)**, Z2B6 **(D)** and hb-P2 **(F)** reporters in embryos from wild-type (Bcd-2X, solid green lines with shaded errors) and  $\Delta bcd/+$  (Bcd-1X, solid yellow lines with shaded errors) females. In each panel, the numbers of embryos for each construct and condition are also shown. Prediction of Bcd-1X patterns from the Bcd-2X patterns assuming a fitted constant shift (values in panel G) are shown as dashed black lines. **(E)** Log-probability map ( $\log p(\Delta(x))$ ) of the shift  $\Delta(x)$  (in %EL) at a given nuclei position in Bcd-2X embryos ( $x$ , in %EL), extracted from combined B6, B9, and Z2B6 reporters' data. The horizontal cyan dashed line represents the best fit value  $\tilde{\Delta} = 10.5$  %EL from the log-probability map. **(G)** Comparison of the shift, with 95% confidence interval, in nuclei position from wild-type embryos to embryos from  $\Delta bcd/+$  females (left bars) and from wild-type embryos to embryos from  $bcd^{E1}/+$  females (right bars) fitted individually to B6, B9, Z2B6 and hb-P2 reporters' data.



**Figure 4—figure supplement 1.** Comparison of fraction of expressing nuclei between nuclear cycles and between Bcd-2X and *bcd<sup>E1/+</sup>* flies. **(A)** Boundary position of fraction of expressing nuclei along AP axis as a function time in nc12 (red) and nc13 (purple), shown for hb-P2 and synthetic reporters, shown with 95% confidence interval. The hb-P2 reporter reach position of expression boundary very rapidly therefore, the position of the boundary is the same at nc12 and nc13 for the hb-P2 reporter even though nc12 is very short. In contrast, positions of the Bcd-only reporters are different at the end of nc12 and nc13. This also true for H6B6 and Z2B6. **(B)** Expression patterns of hb-P2 and synthetic reporters in embryos from wild-type (Bcd-2X, solid green lines with shaded errors) and *bcd<sup>E1/+</sup>* (solid yellow lines with shaded errors) females. Projection of *bcd<sup>E1/+</sup>* pattern (black dashed) from the Bcd-2X pattern assuming a fitted constant shift  $\tilde{\Delta} = 9.0\% \pm 0.5\%$ EL for hb-P2 ( $n_{2x} = 5$  embryos,  $n_{bcdE1/+} = 4$  embryos),  $\tilde{\Delta} = 8.5\% \pm 1.0\%$ EL for B6 ( $n_{2x} = 5$ ,  $n_{bcdE1/+} = 4$ ),  $\tilde{\Delta} = 8.0\% \pm 0.5\%$ EL for B9 ( $n_{2x} = 6$ ,  $n_{bcdE1/+} = 6$ ),  $\tilde{\Delta} = 7.0\% \pm 0.5\%$ EL for B12 ( $n_{2x} = 4$ ,  $n_{bcdE1/+} = 6$ ),  $\tilde{\Delta} = 7.0\% \pm 0.5\%$ EL for H6B6 ( $n_{2x} = 7$ ,  $n_{bcdE1/+} = 5$ ),  $\tilde{\Delta} = 7.5\% \pm 1.0\%$ EL for Z2B6 ( $n_{2x} = 3$ ,  $n_{bcdE1/+} = 5$ ). **(C)** Comparison of the best fitted shift constant from Bcd-2X to *bcd<sup>E1/+</sup>* flies for hb-P2 and synthetic reporters.



**Figure 5.** Fitting the data with models assuming different values for the Bcd gradient decay length  $\lambda$ . **(A)** Hill coefficient  $H$  (see Materials and methods) (solid red) in the steady state window (600 s–800s into nc13 interphase) calculated numerically from the best fitted models. Given that the pattern sharpness  $\eta = H/\lambda$  is measured to be 0.34 from the hb-P2 data, the observed Hill coefficient as a function of  $\lambda$  is given by the black line with shaded error. The physical limit of equilibrium sensing model with 6 BS ( $H = 6$ , black dashed line). **(B)** Log-likelihood of the best fitted models (solid red). The dashed line corresponds to the log-likelihood thresholds for a significantly worse fit (p-value = 0.05).