nature structural & molecular biology

Supplementary information

https://doi.org/10.1038/s41594-025-01615-4

A conserved coupling of transcriptional ON and OFF periods underlies bursting dynamics

In the format provided by the authors and unedited

Contents

1 Supplementary Figures **3**

2 Supplementary Videos **12**

3 Fly strains and genetics **13**

3.1 Plasmid construction . . . 13

3.2 Transgenic fly generation . . . 13

3.3 Genetic crosses for live imaging . . . 14

4 Live imaging **15**

4.1 Sample preparation . . . 15

4.2 Imaging settings . . . 15

5 Image processing **15**

5.1 Nuclei tracking . . . 15

5.2 Transcription spot tracking and signal integration . . . 16

5.2.1 Detecting spot candidates and transcription trajectories . . . 16

5.2.2 Signal integration and transcription time-series . . . 16

5.2.3 AP position assignment . . . 16

6 Data processing, calibration, and measurement errors **17**

6.1 Spatiotemporal alignment of embryos . . . 17

6.1.1 Developmental time normalization and temporal alignment . . . 17

6.1.2 Embryo spatial alignment . . . 17

6.2 Embryo-to-embryo variability and pooling . . . 18

6.2.1 Embryo-to-embryo variability assessment . . . 18

6.2.2 Pooling nuclei across multiple embryos . . . 19

6.3 Calibration of absolute units in live imaging data . . . 19

6.3.1 Predicted live activity from smFISH . . . 19

6.3.2 Computing effective gene lengths . . . 20

6.3.3 Estimating conversion factor for profile matching . . . 20

6.3.4 Comparison of higher cumulants . . . 21

6.4 Measurement error . . . 22

6.4.1 Dual-color measurement error assessment . . . 22

6.4.2 Signal correlation-based imaging error assessment . . . 23

7 Data analysis and modeling **24**

7.1 Deconvolution of the single-allele transcription rate . . . 24

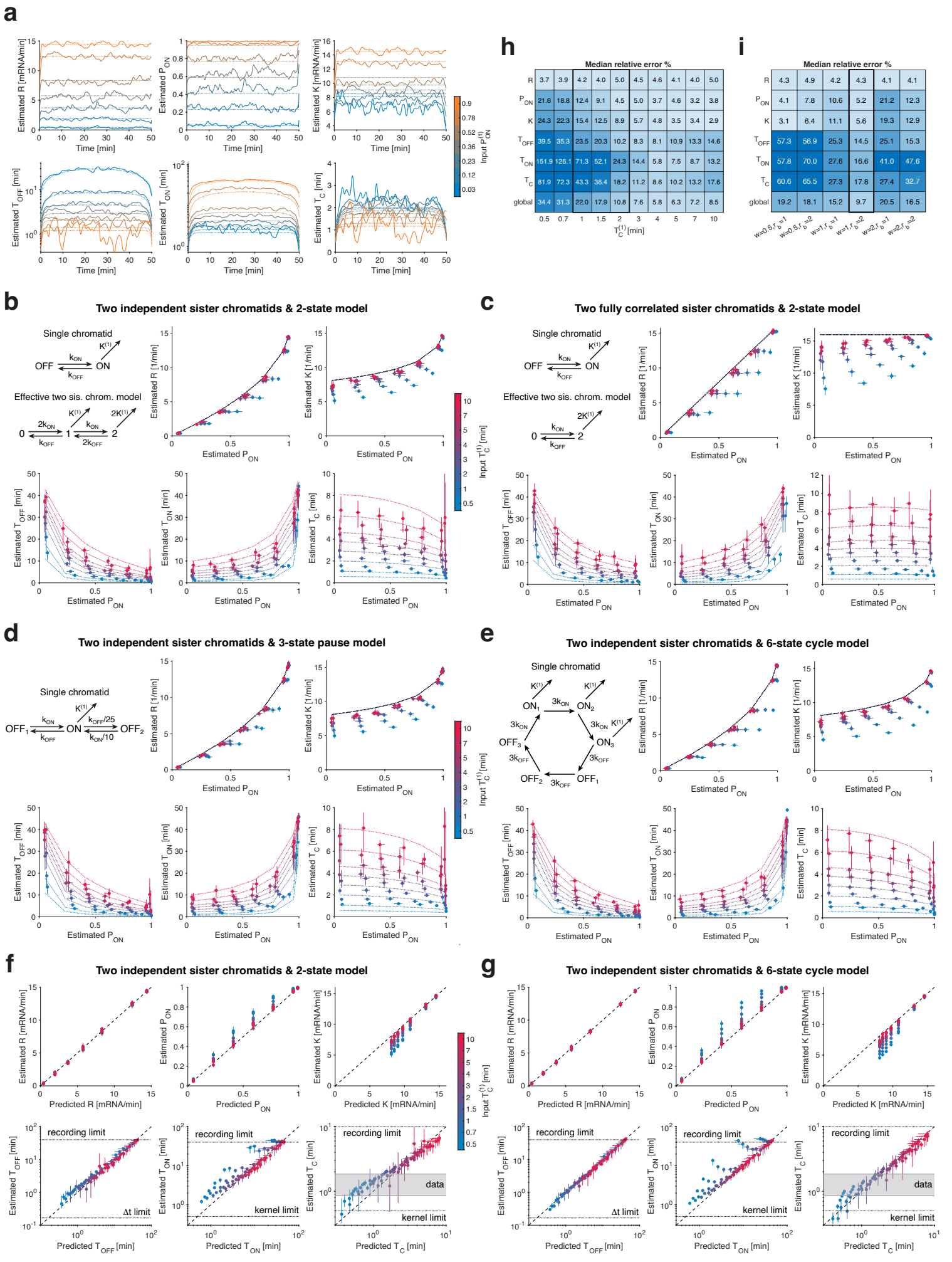
7.1.1 Modeling initiation events . . . 24

7.1.2 Bayesian deconvolution of transcription initiation events . . . 24

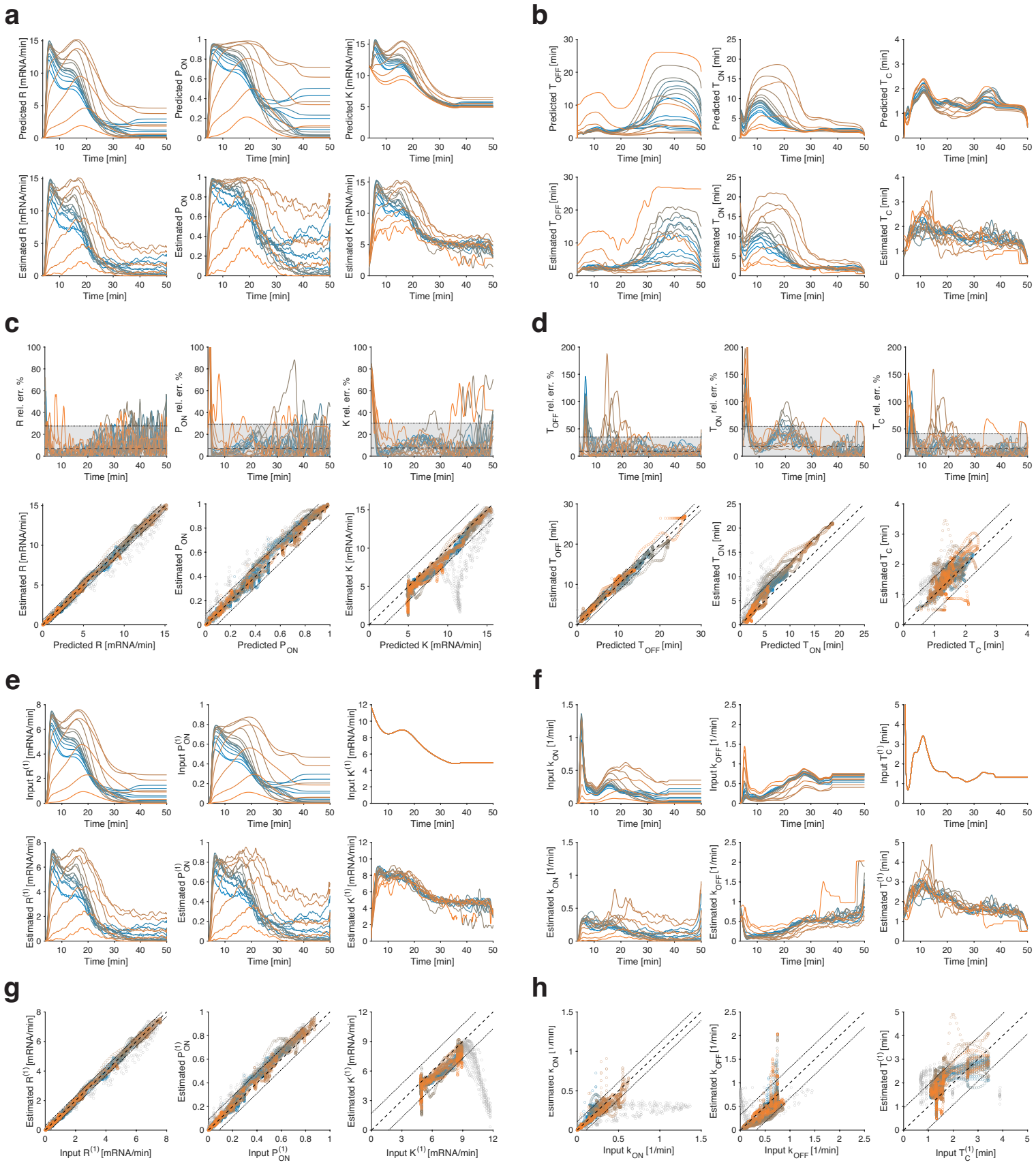
7.1.3 Prior on initiation configurations . . . 25
7.1.4 MCMC sampling of initiation configurations . . . 26
7.1.5 Estimating single-allele transcription rate . . . 27
7.1.6 Validation of deconvolution and estimation of elongation rate . . . 27
7.2 Gap gene mean expression pattern . . . 29
7.2.1 Transcription expression pattern . . . 29
7.2.2 Comparison with protein pattern . . . 29
7.3 Transcription fluctuation analysis . . . 30
7.3.1 Distributions of single-allele transcription rates . . . 30
7.3.2 Constitutive transcriptional regime . . . 31
7.3.3 Transcription rate auto-correlation . . . 32
7.3.4 Two-state model auto-correlation . . . 33
7.3.5 Validating auto-correlation on simulated data . . . 36
7.4 Burst calling and transcriptional parameter estimation . . . 37
7.4.1 Calling bursts from deconvolved time series . . . 37
7.4.2 Estimating transcriptional parameters . . . 38
7.4.3 Sister chromatids and single copy parameters . . . 40
7.4.4 Predicting time-dependent bursting parameters . . . 43
7.4.5 Validation using synthetic, simulation-generated data . . . 46
7.4.6 Validation on real data . . . 50
7.5 Exploring the generality of bursting rules . . . 51
7.5.1 *Drosophila* data . . . 51
7.5.2 Yeast data . . . 52
7.5.3 Human data . . . 52
7.5.4 Mouse scRNA-seq data . . . 54
7.6 Bursting parameters and noise . . . 57

8 References **58**

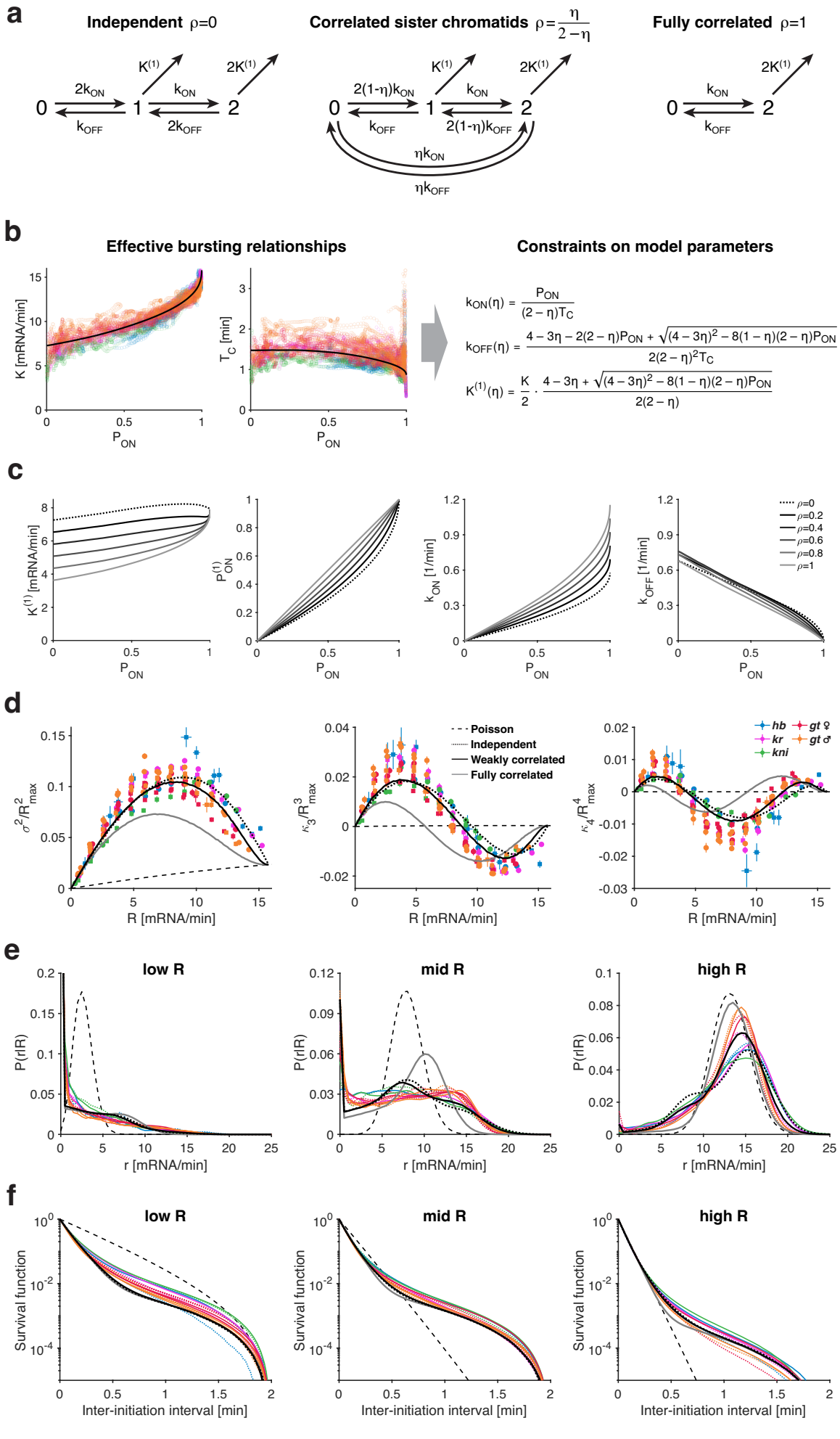
1 Supplementary Figures


a
Estimated R [mRNA/min]
Estimated P_ON
Estimated K [mRNA/min]
Estimated T_OFF [min]
Estimated T_ON [min]
Estimated T_C [min]
Time [min]
Input T_C^(1) [min]
Input P_ON^(1)
h
Median relative error %
R 3.7 3.9 4.2 4.0 5.0 4.5 4.6 4.1 4.0 5.0
P_ON 21.6 18.8 12.4 9.1 4.5 5.0 3.7 4.6 3.2 3.8
K 24.3 22.3 15.4 12.5 8.9 5.7 4.8 3.5 3.4 2.9
T_OFF 39.5 35.3 23.5 20.3 10.2 8.3 8.1 10.9 13.3 14.6
T_ON 151.9 126.1 71.3 52.1 24.3 14.4 5.8 7.5 8.7 13.2
T_C 81.9 72.3 43.3 36.4 18.2 11.2 8.6 10.2 13.2 17.6
global 34.4 31.3 22.0 17.9 10.8 7.6 5.8 6.3 7.2 8.5
0.5 0.7 1 1.5 2 3 4 5 7 10
T_C^(1) [min]
i
Median relative error %
R 4.3 4.9 4.2 4.3 4.1 4.1
P_ON 4.1 7.8 10.6 5.2 21.2 12.3
K 3.1 6.4 11.1 5.6 19.3 12.9
T_OFF 57.3 56.9 25.3 14.5 25.1 15.3
T_ON 57.8 70.0 27.6 16.6 41.0 47.6
T_C 60.6 65.5 27.3 17.8 27.4 32.7
global 19.2 18.1 15.2 9.7 20.5 16.5
w=0.5, r_o=1
w=0.5, r_o=2
w=1, r_o=1
w=1, r_o=2
w=2, r_o=1
w=2, r_o=2
b
Two independent sister chromatids & 2-state model
Single chromatid
OFF
ON
k_ON
k_OFF
K^(1)
Effective two sis. chrom. model
0
1
2
2k_ON
2k_OFF
Estimated P_ON
c
Two fully correlated sister chromatids & 2-state model
d
Two independent sister chromatids & 3-state pause model
OFF_1
OFF_2
k_OFF/25
k_ON/10
e
Two independent sister chromatids & 6-state cycle model
ON_1
ON_2
ON_3
OFF_1
OFF_2
OFF_3
3k_ON
3k_OFF
f
Two independent sister chromatids & 2-state model
Predicted R [mRNA/min]
Predicted P_ON
Predicted K [mRNA/min]
Predicted T_OFF [min]
Predicted T_ON [min]
Predicted T_C [min]
recording limit
Δt limit
kernel limit
data
g
Two independent sister chromatids & 6-state cycle model

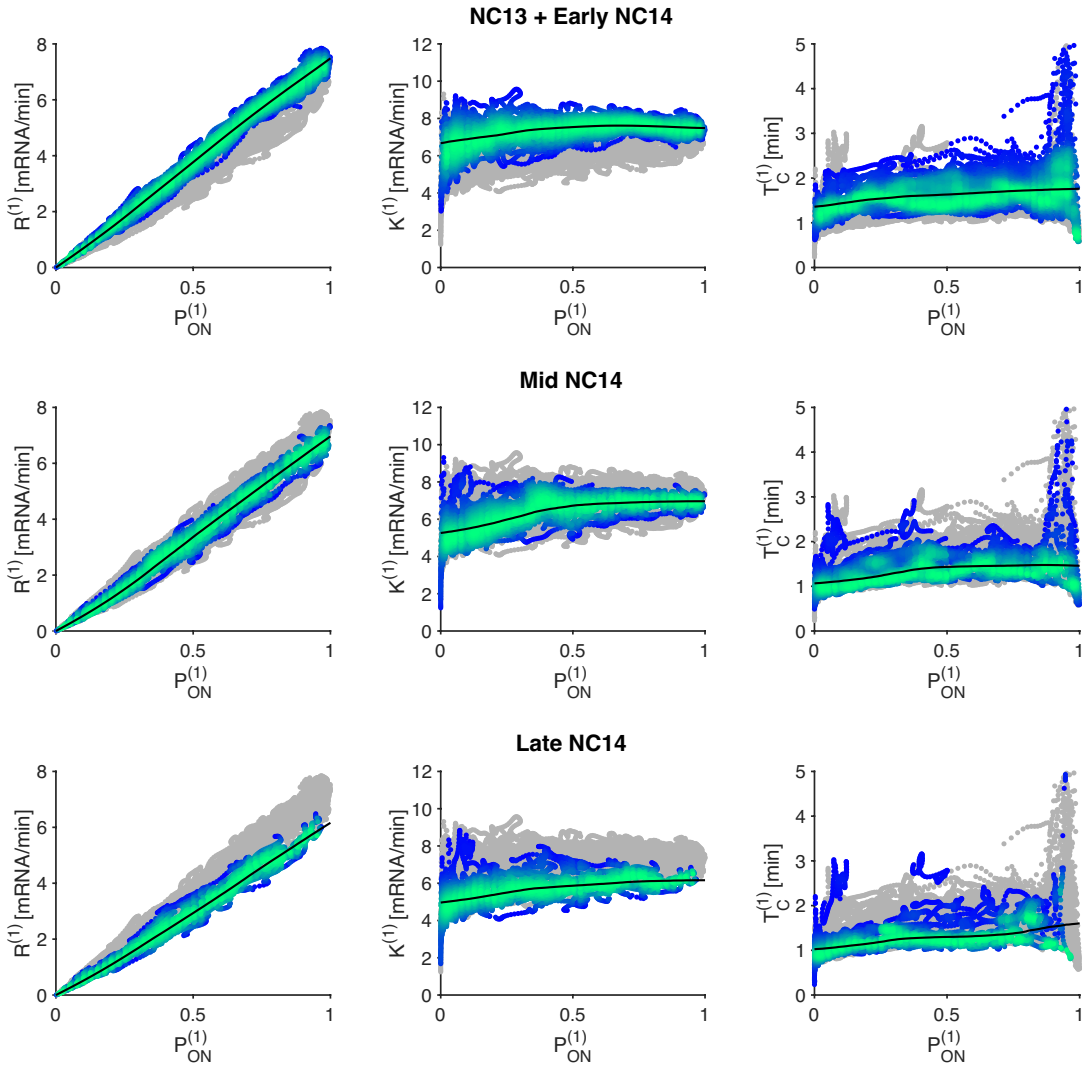
Supplementary Fig. 1: Validation of transcription parameter estimation using stationary simulated data. Burst calling permits effective estimation of transcription parameters. For each set of input parameters ($K^{(1)}$, $T_C^{(1)}$, $P_{\mathrm{ON}}^{(1)}$), we generated simulated data (200 alleles and 50 min long time series at 10 s intervals and with the same imaging noise as measured in real data). Using the 2-state model (Extended Data Fig. 3a) and the Gillespie algorithm, we generated time series for each individual sister chromatid. We summed the initiation events of both chromatids assuming independence. The resulting initiation events were convolved with the elongation kernel to generate synthetic single allele signal data over which measurement noise was added. We performed single allele deconvolution and burst calling as described (Figs. 1d,e and 2a) to estimate the effective mean transcription parameters (R, K, P_{ON}, T_{ON}, T_{OFF} and T_{C}) for each simulated set of alleles. **(a)** Effective parameters R, P_{ON}, K, T_{OFF}, and T_{ON} as a function of time (solid lines). Parameters were estimated using our burst calling routine and stationary (constant in time) input parameters ($T_C^{(1)} = 2$ min, $K^{(1)} = 8\,\mathrm{mRNA/min}$. $P_{\mathrm{ON}}^{(1)}$ is color-coded from blue to orange) and varies from 0.03 to 0.9. For comparison theoretically expected parameters are depicted as dotted lines, respectively. Overall, the estimated effective parameters recovered well the theoretically predicted ones. Despite the stationary nature of the input parameters $T_C^{(1)}$ and $P_{\mathrm{ON}}^{(1)}$, biases for T_{ON}, T_{OFF} and T_{C} (bottom row) are expected due to the finite length of the simulated time series (censoring), especially noticeable near the beginning and the end, where the estimations are "bent" (i.e., the resulting mean time estimate cannot exceed the width of the time window used to perform the estimate). However, our parameter estimation (solid lines) aligns well with the expected biases (dotted lines). **(b)** Estimated parameters for the stationary case (input parameters are constant in time) assuming independent sister chromatids. The parameter estimation was performed on a large simulated data set that includes data in a and b ($T_C^{(1)} = 2$ min, $P_{\mathrm{ON}}^{(1)}$ from 0.03 to 0.9) and data for other values of $T_C^{(1)}$ comprised between $T_C^{(1)} = 0.5$ (blue dots) to $T_C^{(1)} = 10$ (red dots). Each dot results from one combination of input parameters ($K^{(1)}$, $T_C^{(1)}$, $P_{\mathrm{ON}}^{(1)}$) and corresponds to the median effective parameter and the error bars to the 68% confidence interval estimated over 50 min. Dotted lines show the expected effective parameters given the input parameters. While we observe biases when the input $T_C^{(1)}$ approaches 0.5 min, the estimated parameters reproduce the predicted trend (dotted lines) sufficiently well to detect relationships as a function of P_{ON}. See f for the characterization of absolute errors. **(c)** Same as b except the sister chromatids were assumed to be fully correlated. Accordingly, simulated data were generated using an effective single chromatid 2-state model with doubled input parameters (see top left). While biases on estimated P_{ON} and K appear more pronounced, the recovered P_{ON} to K relationships are quantitatively distinct enough from those in C to discriminate one versus two sister chromatids confidently. Overall, the estimated effective T_{OFF}, T_{OFF}, and T_{C} are properly recovered, even when sister chromatids are strongly correlated. This consistency aligns with our burst calling approach, unassuming of specific distributions for ON and OFF intervals and therefore correctly estimating the mean ON and OFF times with little impact from the generative model. **(d)** Same as b except each sister chromatid was modeled by a 3-state pause model with one ON state and two alternative transitions toward an OFF state. This model was previously used to investigate promoter architecture in fly embryos [1]. Consistent with this study, we chose a reaction rate such that the pause state (OFF_2) is less frequent but more persistent than the basal OFF state (OFF_1). Such a model generates a hyperexponential waiting distribution for the effective OFF intervals instead of exponentially distributed as in the 2-state model. As in c, the effect of the generative model on our estimated bursting parameters is minimal (see b as a reference). **(e)** Same as b except each sister chromatid was modeled by a 6-state cycle model with three sequential ON states followed by three sequential OFF states. This model is similar to other cycle models previously used to model mammalian promoters [2]. Here, the model generates gamma waiting-time distributions (with shape parameter $k = 3$) for the effective ON and OFF intervals instead of exponential distributions as in the 2-state model. As in the test cases in d and e, the parameter estimation is not particularly sensitive to the generative model. It stems from the intrinsic property of our burst calling approach, which estimates average ON and OFF intervals without making any assumptions on the underlying model. See g for the characterization of absolute errors. **(f-g)** Global comparison between expected and estimated parameters for the simulated dataset in b (f) and e (g). Estimated parameters (estimated with our deconvolution and burst calling approaches) and predicted (i.e., expected) parameters are in good agreement (i.e., diagonal would be identical) over a large range of $T_C^{(1)}$ and $P_{\mathrm{ON}}^{(1)}$ values. Noticeable biases in P_{ON}, K and T_{ON} arise when $T_C^{(1)}$ approaches 0.5 min (light blue dots). Such biases increase with the likelihood of merging successive ON-periods, i.e. when $T_C^{(1)}$ is small compared to the size of our burst calling window ($T_C^{(1)} < 1\,\mathrm{min}$). However, the relative deviation in T_C never exceeds a factor of two, supporting our ability to detect its constancy in the experimental data sets (of which the 5^{th} to 95^{th} percentiles are highlighted in gray, i.e., $[0.85, 1.85]$). Furthermore, T_C estimated from experimental data is well within the range mapped out by our detection limit. Indeed, the maximal length of our recording, which is around 40 min in NC14, sets the upper bound in T_{OFF} and T_{ON}, which converts into an upper bound of 10 min for T_{C} (using $T_{\mathrm{ON}} = T_{\mathrm{OFF}} = T_{\mathrm{max}} = 40$ min, we have $(T_{\mathrm{max}}/2)^2/(2(T_{\mathrm{max}}/2)) = 10$ min as one needs to observe both an ON and OFF interval to estimate T_{C}). Our sampling time of $\Delta t = 10$ s and the size of the window used in the burst calling routine set the lower bound in T_{OFF} and T_{ON}, respectively. The minimal detectable T_{ON} is 30 s and corresponds to two successive initiation events (when the tip of our ~ 1 min peaked-kernel used for averaging crosses the 2-molecules threshold). Thus, a conservative lower bound on estimated T_{C} is 30 s. Lastly, we stress that the two generative models in f and g display very similar bias-behavior, pointing to the limited impact of the underlying model on the parameter estimates. **(h)** Summary of median relative error for each effective parameter estimated from the data in f as a function of input $T_C^{(1)}$. The parameter estimated from real data (Extended Data Fig. 8c) suggests that $T_C^{(1)}$ lies within $1-2$ min (black border). **(i)** Summary of median relative error for each effective parameter estimated from the whole data set in c as a function of the burst calling parameters. Burst calling depends on two free parameters: the time window w over which the rate is estimated, and the rate threshold r_b applied to call the burst (see Fig. 2a). Our default parameter values are $w = 1u$ and $r_b = 2/u$ (with $u = 5\Delta t = 5/6\,\mathrm{min}$), which should be close to optimal given the estimated correlation time of $\tau_{\mathrm{AC}} \sim 1$ min (Fig. 1g) and our measurement sensitivity of $1-2$ mRNA. When testing the effect of different w and r_b values on the median relative error, our default choice leads to the lowest global relative error.


a
b
c
d
e
f
g
h
Predicted R [mRNA/min]
Predicted P_{ON}
Predicted K [mRNA/min]
Estimated R [mRNA/min]
Estimated P_{ON}
Estimated K [mRNA/min]
Predicted T_{OFF} [min]
Predicted T_{ON} [min]
Predicted T_C [min]
Estimated T_{OFF} [min]
Estimated T_{ON} [min]
Estimated T_C [min]
Time [min]
R rel. err. %
P_{ON} rel. err. %
K rel. err. %
T_{OFF} rel. err. %
T_{ON} rel. err. %
T_C rel. err. %
Predicted R [mRNA/min]
Predicted P_{ON}
Predicted K [mRNA/min]
Predicted T_{OFF} [min]
Predicted T_{ON} [min]
Predicted T_C [min]
Input $R^{(1)}$ [mRNA/min]
Input $P_{ON}^{(1)}$
Input $K^{(1)}$ [mRNA/min]
Input k_{ON} [1/min]
Input k_{OFF} [1/min]
Input $T_C^{(1)}$ [min]
Estimated $R^{(1)}$ [mRNA/min]
Estimated $P_{ON}^{(1)}$
Estimated $K^{(1)}$ [mRNA/min]
Estimated k_{ON} [1/min]
Estimated k_{OFF} [1/min]
Estimated $T_C^{(1)}$ [min]

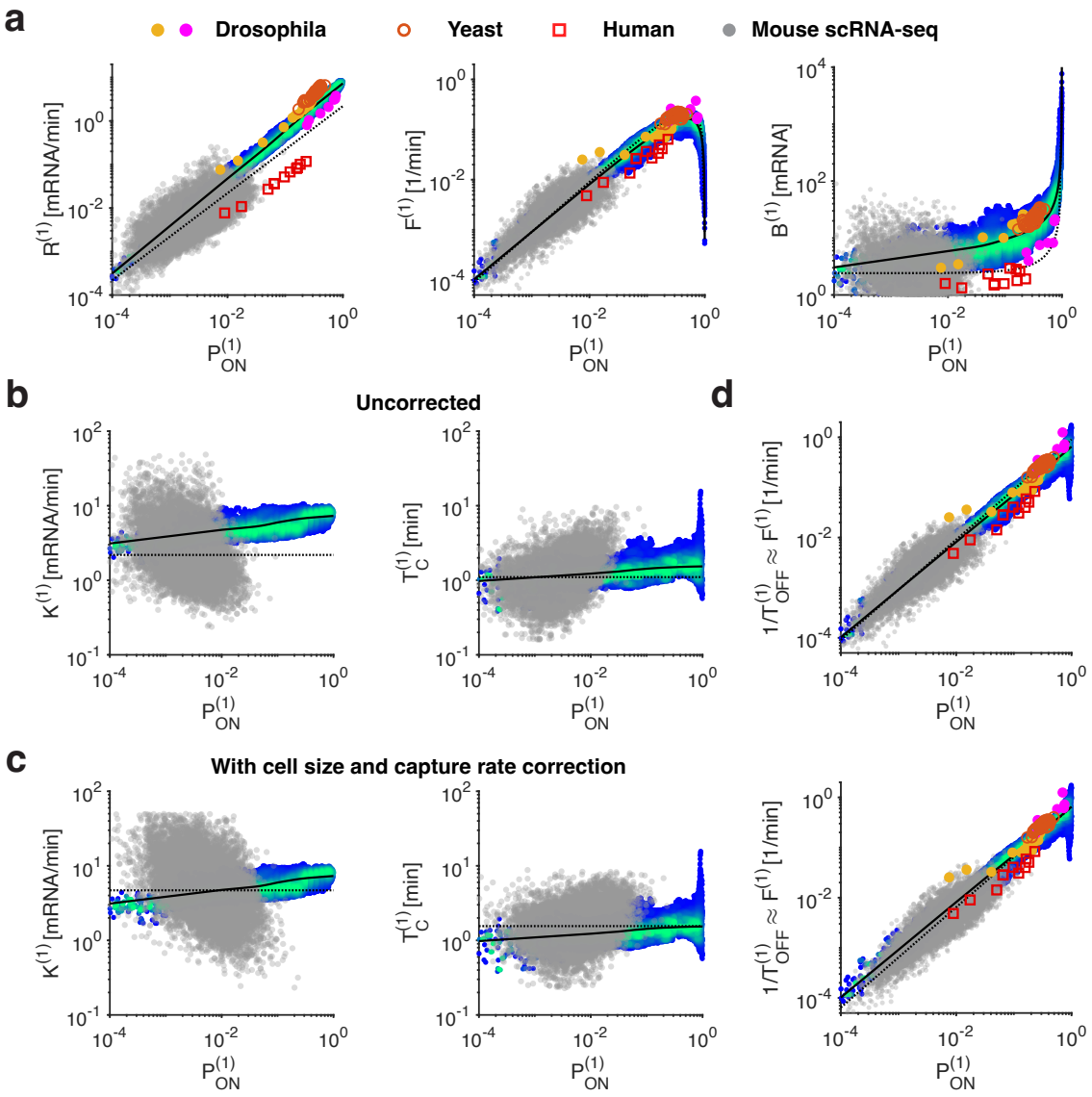
Supplementary Fig. 2: Validation of transcription parameter estimation using *hb*-like non-stationary simulated data. Simulated data is generated as in Supplementary Fig. 1, except that the input parameters are non-stationary (time-dependent). **(a-b)** Comparison between expected (a, b, top row) and estimated effective parameters (a, b, bottom row) for time-dependent input parameters (a, b, top row), mimicking the transcriptional output of *hb* in NC14. Color code (blue to orange) stands for virtual AP position. Estimated effective parameters tightly recapitulate the time-dependence of the expected parameters, allowing us to also capture the temporal regulation of the gap genes. **(c-d)** We show with this realistic test case that our estimated parameters (a, b, bottom row) using burst calling very closely match the expected effective ones (a, b, top row). Indeed, the median (dashed line) and 95 percentiles (dotted line) of the relative errors beyond the transient ($t > 7.5$ min) are all smaller than 18% and 55% respectively (top row). Furthermore, most of the time points (circles, bottom row) lie on the line of slope one (dashed line), which strongly supports our ability to precisely characterize the transcription parameters from real data. The dotted lines indicate the 95 percentiles of the absolute errors. **(e–f)** Comparison between input (e, f, top row) and burst calling estimated single-gene copy parameters (e, f, bottom row) as a function of time. The time-dependent input parameters (e, f, top row) were used to generate the synthetic *hb* data in a and b. Color code (blue to orange) stands for virtual AP position as in a and b. The estimated single-gene copy parameters were computed from the effective ones (a, b, bottom row) assuming the latter originated from two independent sister chromatids. Namely, we get $R^{(1)} = R/2$, $P_{\mathrm{ON}}^{(1)} = 1-(1-P_{\mathrm{ON}})^{1/2}$, $K^{(1)} = K(1+(1-P_{\mathrm{ON}})^{1/2})/2$, which are exact, and assuming steady state $T_{\mathrm{C}}^{(1)} = 2T_{\mathrm{C}}/(1+(1-P_{\mathrm{ON}})^{1/2})$, $k_{\mathrm{ON}} = P_{\mathrm{ON}}^{(1)}/T_{\mathrm{C}}^{(1)}$ and $k_{\mathrm{OFF}} = (1-P_{\mathrm{ON}}^{(1)})/T_{\mathrm{C}}^{(1)}$. **(g-h)** Even though the single gene copy parameters are deeply buried in the data, our simple burst calling procedure still manages to recover them correctly. Some discrepancies are observed for k_{ON}, k_{OFF} and $T_{\mathrm{C}}^{(1)}$, mostly near the beginning ($t < 7.5$ min) when the transient after mitosis violates the near-steady-state assumption. But these are expected as the relationships $k_{\mathrm{ON}} = P_{\mathrm{ON}}^{(1)}/T_{\mathrm{C}}^{(1)}$ and $k_{\mathrm{OFF}} = (1-P_{\mathrm{ON}}^{(1)})/T_{\mathrm{C}}^{(1)}$ are only valid for near-steady state.


a
Independent ρ=0
Correlated sister chromatids ρ=η/(2−η)
Fully correlated ρ=1
b
Effective bursting relationships
Constraints on model parameters
c
d
Poisson
Independent
Weakly correlated
Fully correlated
hb
kr
kni
gt ♀
gt ♂
R [mRNA/min]
e
low R
mid R
high R
r [mRNA/min]
f
Survival function
Inter-initiation interval [min]

Supplementary Fig. 3: Bursting relationships provide strong empirical constraints on transcription models. This figure demonstrates how the derived bursting relationships help to discriminate between transcription models. Also shows how models that specifically account for two indistinguishable sister chromatids provide a better explanation for the data, suggesting that two weakly correlated sister chromatids form transcription loci in the embryo. **(a)** Minimal class of models describing two identical and possibly correlated sister chromatids. The gene copy on each sister chromatid behaves as a 2-state model whose transitions between the active and inactive state can be coupled ($0 < \eta \leq 1$). Assuming two identical gene copies, the model reduces to three effective states corresponding to 0, 1, and 2 active copies. Transition rates between the effective states are a function of the single gene copy switching rates k_{ON} and k_{OFF}. Production of mRNAs in the active states ('1' or '2') is determined by the single gene copy initiation rate $K^{(1)}$. The coupling parameter $\eta \in [0, 1]$ determines the Pearson correlation coefficient $\rho \in [0, 1]$ between the two bursting gene copies, which is given by $\rho = \frac{P(2)-(P(2)+P(1)/2)^2}{(P(2)+P(1)/2)(1-P(2)-P(1)/2)} = \frac{\eta}{2-\eta}$ with $P(1)$ and $P(2)$ the occupancy probability of states '1' and '2'. In the limiting case, $\eta = 0$ the two gene copies are independent and $\rho = 0$, whereas in the case $\eta = 1$ the two gene copies are fully correlated and $\rho = 1$. **(b)** The identified bursting relationships (derived without fitting) impose strong constraints on possible models. Any model of transcription (characterized by a set of active and inactive states) must satisfy the relationships between the effective parameters K, T_{C} and P_{ON} (left, data from NC13 + early NC14, see Extended Data Fig. 8a). Using the class of models in a, we can map the effective parameters onto the single copy parameters by satisfying the following equations: $T_{\mathrm{OFF}} = T_0$, $T_{\mathrm{ON}} = T_{1\cup 2}$ and $KP_{\mathrm{ON}} = K^{(1)}P(1) + 2K^{(1)}P(2)$. Doing so (see Section 7.4.3), we find expressions for k_{ON}, k_{OFF} and $K^{(1)}$, as functions of the sole free parameter η. We have thus reduced a four-parameter model into a single-parameter model, which can easily be tested against data (see d, e, and f). **(c)** Single gene copy parameters ($K^{(1)}$, $P^{(1)}_{\mathrm{ON}} = k_{\mathrm{ON}}/(k_{\mathrm{ON}}+k_{\mathrm{OFF}})$, k_{ON} and k_{OFF}, see b) as a function of P_{ON}, predicted by the different class of models in a given our measured bursting relationships (black lines in b). The color code highlights the effect of the correlation coefficient $\rho \in [0, 1]$ (see legend). The correlation between sister chromatids makes a difference in single gene copy parameters at most of factor two. However, such a difference could have a noticeable signature in the data, which we investigate in d, e, and f. **(d)** Variance (2^{nd} cumulant), 3^{rd} cumulant and 4^{th} cumulant of single allele transcription rate r as a function of mean transcription rate $R = 2K^{(1)}P^{(1)}_{\mathrm{ON}}$ in NC13 (square) and early NC14 ($7.5 \leq t < 20.5$ min; circle). The single allele transcription rates are estimated within 2-min-intervals and the cumulants are normalized by $\max R = R_{\max} = 15.77$ mRNA/min. Data color code as in Extended Data Fig. 3. The dashed line corresponds to the Poisson limit (i.e., a single gene copy constitutively transcribed). In contrast, the dotted and solid curves are predictions made by the models in A that satisfy the derived bursting relationships in b. The dotted curve corresponds to independent sister chromatids ($\rho = 0$), the solid black line to weakly correlated sister chromatids ($\rho = 0.16$) and the solid gray line to fully correlated sister chromatids ($\rho = 1$). We immediately see that not accounting for a pair of sister chromatids, or equivalently only considering highly correlated pair $\rho \sim 1$, provides a poor explanation for the data. Models that include little to no correlations between chromatids match well the data. The best fit to the cumulants is obtained for $\rho \sim 0.16$, which is close to independence. The predicted cumulants are computed by sampling from each model to account for the Pol II footprint estimated at around 60 bp. **(e)** Distribution $P(r|R)$ of single allele transcription rates estimated within 2-min-intervals in both NC13 (color dotted lines) and early NC14 (color solid lines). These distributions are computed over all the nuclei from time points and AP bins whose mean transcription rate R corresponds either to a low $[2.1, 3.2]$, mid $[7.5, 8.5]$ or high transcription level $[12.8, 13.9]$ (as in Extended Data Fig. 3). Black dashed line corresponds to the Poisson distribution for a single constitutive gene. Solid black and gray lines are predicted distributions by the models as in d (weakly correlated and fully correlated sister chromatids, respectively). While the model with fully correlated sister chromatids fails to account for the empirical distributions at mid and high R, a small amount of correlation ρ leads to a good match. **(f)** Survival function of elapsed time between successive PolII initiation events (inter-initiation intervals) estimated within 2-min intervals in both NC13 (color dotted lines) and early NC14 (color solid lines). These survival functions (1-"the cumulative distribution") are computed over the same nuclei, time points, and AP bins as in d. The black dashed line corresponds to the Poisson for a single constitutive gene. As in d and e, solid black and gray lines are model-predicted survival functions. A model describing a pair of weakly correlated bursting sister chromatids generates survival functions that closely mimic the data on a 2-min-interval (over which the system can be considered roughly stationary, i.e. $T_{\mathrm{C}} \sim 1.5$ min)



Supplementary Fig. 4: Common bursting relationships single gene copy assuming weakly correlated sister chromatids. Global scatter of single gene copy transcription parameters as a function of P_{ON} for all gap genes estimated within three time windows, corresponding to NC13 ($6.5 \leq t$ min) plus early NC14 ($7.5 \leq t < 20.5$ min), mid NC14 ($20.5 \leq t < 34.5$ min), and late NC14 ($34.5 \leq t < 48$ min). The single gene copy parameters were computed from the effective ones assuming weakly correlated ($\rho = 0.16$) sister chromatids (see Supplementary Fig. 3 and Section 7.4.3). The difference between independent and weakly correlated chromatids is very subtle (See Extended Data Fig. 8c for comparison with independent sister chromatids). Interestingly, in both cases, the relationship between SGC transcription rate $R^{(1)}$ and SGC ON-probability $P^{(1)}_{\mathrm{ON}}$ is almost linear, confirming that the SGC initiation rate $K^{(1)}$ does not depend strongly on $P^{(1)}_{\mathrm{ON}}$. Thus, the apparent dependence of K on P_{ON} is only effective and results from measuring two sister chromatids (two gene copies) together, instead of an isolated single gene copy.



Supplementary Fig. 5: Comparison between bursting parameters from mouse scRNA-seq analysis and live imaging in other organisms. (a) Transcription rate R, burst frequency F and burst size B for a single gene copy versus $P_{\rm ON}$. Transcription parameters estimated from allelic single-cell RNA-seq in mouse cells (gray circles [3]) are mostly consistent with our relationships (black lines). Black dotted lines correspond to expected relationships in scRNA-seq data assuming constant K and $T_{\rm C}$, see b. The scRNA-seq parameters were estimated as in the original study by fitting (maximum-likelihood) a Beta-Poisson distribution to the single-cell mRNA counts. Contrary to the original study, we regularized the inference through a weak prior on T_C around 1 min. Remarkably, this is sufficient to alleviate the parameter identifiability issue related to long mRNA lifetimes. Importantly, as in the original study, the estimated parameters displayed here are not corrected for low mRNA capture rate and cell-to-cell variability in total counts due to cell size variation (e.g. related to cell-cycle). Once corrected for these factors (methods), we recover very similar trends (see Fig. 7b), albeit with a larger mean initiation rate (~ 2.2 times larger, see B and C for comparison). **(b-c)** Initiation rate K and switching correlation time $T_{\rm C}$ for a single gene copy versus $P_{\rm ON}$, estimated from allelic single-cell RNA-seq. Dotted lines stand for the mean scRNA-seq parameters. Parameters estimated without (b) and with (c) correction for cell-to-cell variation in total mRNA counts and low capture rate (assumed to be 20%). Even with the corrections above, estimated parameters in absolute units remain very noisy due to technical noise in both scRNA-seq and mRNA lifetime measurements. **(d)** Inverse mean OFF-time $1/T_{\rm OFF}$ versus $P_{\rm ON}$. $1/T_{\rm OFF}$ is commonly used as an approximation for burst frequency F in the regime of short ON-time $T_{\rm ON} \ll T_{\rm OFF}$, or equivalently low $P_{\rm ON}$. Indeed, comparison with the exact burst frequency $F = 1/(T_{\rm OFF} + T_{\rm ON})$ in Fig. **??**B demonstrates that $F \approx 1/T_{\rm OFF}$ is an excellent approximation, notably for all mouse and human, as these are lowly expressed, i.e., $P_{\rm ON} \lesssim 0.5$.

2 Supplementary Videos

Video V1-V4. Representative videos of 4 gap gene transcription rate measurements: for *hb*, *Kr*, *kni*, and *gt* (female), respectively. Anterior is on the left. We measured the dorsal side of the embryos. Red channel shows nuclei marked by Histone-RFP label; green channel shows the MCP-GFP signal, in particular highlighting one site of nascent transcription in each nucleus when it binds to MS2 stem-loops. Each video was mean projected in the axial (z) direction. Projected image size is about 210x120 μm^2. Timestamp is in units of minutes. The video display frame rate is 15 Hz, which is equivalent to 2.5 min of imaging time per display second. Video contrast was adjusted for better visualization.

Video V5. The temporal progression of transcriptional activity along developmental time. X-axis: normalized embryo length (0-1). (Top) The mean activity directly from the MS2 signal. (Bottom) Calculated mean transcription rate (Section 7.2) that is no longer gene length dependent.

Video V6. The temporal progression of relative protein abundance along developmental time. (Top) Protein accumulation predicted from measured mean transcription rate based on a simple model. (Bottom) Previously measured protein patterns from carefully staged gap gene antibody staining [4]. Major differences observed for *hb* are likely related to maternal mRNAs, whose contribution is not observed in our live measurements reflecting only zygotic gene expression.

Video V7. The mean transcription rate for wildtype *hb*-MS2 (dark blue) and its distal enhancer deletion mutant (light blue).

Video V8. The mean transcription rate for wildtype *kni*-MS2 (green) and its distal enhancer deletion mutant (dark yellow-green).

Video V9. The mean transcription rate for wildtype *kni*-MS2 (green) and its expression in the background of *hb* null mutant (light yellow-green).

3 Fly strains and genetics

3.1 Plasmid construction

MS2 and PP7 stem-loops cassettes were produced by a series of cloning steps, duplicating the annealed oligos below. The final cassette consists of 24 stem-loops (12 repetitions of the initial annealed oligos)

- MS1 oligo 1: CTAGTTACGGTACTTATTGCCAAGAAAGCACGAGCATCAGCCGTGCCTCC AGGTCGAATCTTCAAACGACGACGATCACGCGTCGCTCCAGTATTCCAGGGTTCATCC
- MS2 oligo 2: CTAGGGATGAACCCTGGAATACTGGAGCGACGCGTGATCGTCGTCGTTTG AAGATTCGACCTGGAGGCACGGCTGATGCTCGTGCTTTCTTGGCAATAAGTACCGTAA
- PP7 oligo 1: CTAGTTACGGTACTTATTGCCAAGAAAGCACGAGACGATATGGCGTCCGT GCCTCCAGGTCGAATCTTCAAACGACGAGAGGATATGGCCTCCGTCGCTCCAGTATTC CAGGGTTCATCC
- PP7 oligo 2: CTAGGGATGAACCCTGGAATACTGGAGCGACGGAGGCCATATCCTCTCGT CGTTTGAAGATTCGACCTGGAGGCACGGACGCCATATCGTCTCGTGCTTTCTTGGCAA TAAGTACCGTAA

All 2attP-dsRed plasmids were made by cloning homology arms into a previously used 2attp-dsRed plasmid [5]. All 2attB-insert plasmid were made by cloning the inserts into a previously used 2attB-insert plasmid [6]. Plasmid maps and cloning details are available upon request.

3.2 Transgenic fly generation

For the endogenous tagging of the gap genes (*hb*, *kni*, *Kr*, and *gt*) a two-step transgenic strategy was used. First, a CRISPR-mediated replacement of each locus was performed. Specifically, the upstream regulatory regions (including annotated enhancers) and coding regions were replaced by a 2attp-dsRed cassette. This CRISPR step was performed with the following guides:

Gene	Guide1	Guide2	Stem loops insertion coordinates (in dm6)
kni	CTTGAAGCTCAT TAATTCCACGG	GGGAGGGCTTGA TTCGGGAAAGG	Intronic (3L:20694142)
hb	ATGAACACTCAT ACATATCCTGG	GTCACGGCTAAG ACGCCTTAAGG	Intronic (3R:8694188), 3utr (3R:8691669*)
gt	TCTTACGTGTAA GAATTCATGGG	CGGCCGGCGAGG AAGTGAACGGG	Intronic (X:2428789)
Kr	GTAAATCCCAGA TGTATAATTGG	AAGACTTGAACC AAATACACAGG	Intronic (2R:25227036), 3utr (2R:25228873)

* An additional ~ 1.5 kb fragment of the gene yellow was placed downstream of the stem-loop cassette to insure sufficient length for signal detection.

Homology arms were amplified from genomic DNA of the *nos-Cas9/CyO* injection line (BDSC #78781). For *kni* and *gt*, loss of gap gene proteins was verified by antibody staining as previously described [4]. For all four genes segmentation defects previously ascribed to the loss of protein were

observed. PCR verification was performed (i.e., from the dsRed to the flanking genomic regions). These lines are referred to as a gap gene null line (e.g., *hb*-null).

In a second step, the deleted region for each gap gene was PCR amplified from the *nos-Cas9/CyO* line and cloned into a 2attB plasmid. MS2 stem-loops (see description above) were cloned into the gene (see insertion position in the above table and Extended Data Fig. 1A). This 2attB-insert was subsequently delivered into the 2attp site of the corresponding gap gene line from step one by co-injection with phiC31 integrase (RMCE injection with $\sim 0.25\mu g/\mu l$ [DNA] and hsp-PhiC31 DNA $\sim 0.1\mu g/\mu l$). Flies were screened for loss of dsRed and PCR verified for the presence of the insert in the correct orientation, with primers from inside the insert to the flanking genomic regions.

In addition to MS2 lines for each gap gene, two lines with dual (orthogonal) stem-loop systems (MS2 and PP7) on the same gene were produced. They were primarily used for elongation measurements (see Section 7.1.5 and Extended Data Fig. 2). We generated 1) a *hb* line with intronic insertion of MS2 and a 3'utr insertion of PP7, 2) a *Kr* line with intronic insertion of PP7 and 3'utr insertion of MS2 (insertion positions are as in the single stem-loop lines, see table above). An additional control line with a cassette of alternating MS2 and PP7 stem-loops [6] in the intronic *Kr* gene was used to assess imaging error (Extended Data Fig. 1e,f).

A *hb*-MS2 fly line was generated without the distal *hb* enhancer. To this end, the 2attB-insert included the deleted region from the *hb* locus, the MS2 (intronic) insertion, and a replacement of the distal enhancer (3R:8698553-8700369, dm6) by a fragment of the same length with the *lacZ* gene. The modified 2attB-insert was delivered into the *hb*-2attp site as described above. Similarly a *kni*-MS2 line lacking the distal enhancer was made by replacing the enhancer sequence (3L:20696340-20697752, dm6) with a lacZ fragment. Plasmid maps and cloning details are available upon request.

3.3 Genetic crosses for live imaging

A fly line with the fluorophore (*yw; His2Av-mRFP; nanos>MCP-eGFP*) [7] was crossed with wild-type (Ore-R) flies to reduce the background fluorescence. Female offspring (*yw/+; His2Av-mRFP/+; nanos>MCP-eGFP/+*) was subsequently crossed with our CRISPR-MS2 transgenic fly lines.

Male *gt*-MS2 crossing scheme: Since *gt* is on the X-chromosome, the above crossing scheme only works for female *gt-MS2* embryos. For male embryos, we followed a different crossing strategy, where Female *gt-MS2* flies are crossed with a 3-color fluorophore line (*yw; His2Av-mRFP; nanos>MCP-eGFP,nanos>PCP-mCherry*). The female offspring of that cross (*gt-MS2/+; His2Av-mRFP/+; nanos>MCP-eGFP,nanos>PCP-mCherry/+*) was then crossed with (*gt-MS2PP7-interlaced*) males. To ensure we measure male embryos, imaging was performed only with embryos expressing *gt-MS2* signal but not PP7-PCP-mCherry signal.

***kni-MS2* in *hb-null* background crossing scheme.** Since *kni* and *hb* are both on Chromosome-III, we adopted a two-generation crossing scheme for this line. The first generation had two sets of crosses. A first fly line (*hb-null/Tm3sb*) was crossed with *hb-3'UTR-MS2*. A second fly line (*hb-null,kni-MS2/Tm3sb*) was crossed with a dual fluorophore line (*yw; His2Av-mRFP; nanos>MCP-eGFP*). Male offspring from the first cross (*hb-null/hb-3'UTR-MS2*, selected against Tm3sb) was then crossed with female offspring from the second (*His2Av-mRFP/+; nanos>MCP-eGFP/hb-null,kni-MS2*, selected against Tm3sb). Imaging was performed with embryos that had *kni-MS2* expression but without the *hb-3UTR-MS2* signal (spatially distinguishable within embryos). This ensured the measured embryos had a geno-

type of *hb-null,kni-MS2/hb-null* on Chromosome-III.

4 Live imaging

4.1 Sample preparation

Sample preparation for live imaging is adapted from previous work [7, 8]. We prepare an air-permeable membrane (roughly 2 cm by 2 cm) on a sample mounting slide. Heptane glue is evenly distributed on the membrane. Since we record embryos starting from late nuclear cycle 12, the flies are caged on agar plates for 2 to 2.5 hours. Embryos laid on these plates in that time window are transferred to a piece of double-sided tape by a dissection needle with which the embryos are also hand-dechorionated and placed on the glued membrane. After mounting, we immerse embryos in Halocarbon 27 oil (Sigma) and compress the embryos with a cover glass (Corning #1 1/2, 18x18 mm). Excess oil is removed by a tissue if needed. All embryos are mounted dorsal up (dorsal side facing objective lens).

4.2 Imaging settings

Live embryo imaging is performed from late NC12 to the end of NC14 (onset of gastrulation) using a custom-built inverted two-photon laser scanning microscope (similar in design to previous studies [9, 10]). The laser sources for two-photon excitation are a Chameleon Ultra (wavelength at 920 nm, for channel 1) and a HighQ-2 from Spectra Physics (wavelength at 1045 nm, for channel 2). Excitation and emission photons are focused and collected by a 40x oil-immersion objective lens (1.3NA, Nikon Plan Fluor). The average laser power measured at the objective back aperture are 20 mW and 4 mW, for Ultra and HighQ-2 lasers, respectively. In dual color Imaging Error Experiments (Section 4.4) 20 mW was used for both channels. Laser scanning and image acquisition are controlled by ScanImage 5.6-1 (Vidrio Technologies, LLC). Fluorescence signal from both channels is simultaneously detected by separated GaAsP-PMTs (both Hamamatsu H10770P-40). The pixel size is 220 nm, with an image size of 960x540 pixels. Each image stack contains 12 frames, separated by 1 μm in the axial z-direction. Pixel dwell time is 1.4 μs. The overall temporal resolution for one image stack is around 10 s.

5 Image processing

5.1 Nuclei tracking

Nuclei tracking was performed on a specifically dedicated image acquisition channel with red fluorophores, i.e. red-fluorescently labeled histone proteins (His2Av-mRFP) that are provided maternally in all imaged embryos. Typical imaging windows span from late NC12 to late NC14 during embryonic development, when the embryos undergo two rounds of nuclear division. As during this period, the size of the nuclei changes continuously (from $\sim 12\,\mu$m to $\sim 7\,\mu$m in diameter), we use adaptive parameters for nuclei filtering and identification. First, as a pre-filter, a simple 2D Gaussian filter was applied to the Z-projections of the image stack of the red nuclei channel. This procedure roughly captures the average nuclear size of the population at each time point. Second, this time-dependent mean nuclear size is used to construct dynamic parameters for a 3D difference of Gaussian (DOG) filter to smooth the nuclei images, whose spherical shape is enhanced with a disk filter. For each 3D embryo image stack (per time point), filtered images are normalized to the maximum pixel intensity and subsequently binarized and segmented using

a watershed algorithm. From each of these 3D segmented image stacks we constructed a 2D Voronoi diagram. For consecutive time points, the identity for all nuclei is propagated based on the shortest pair-wise distance of centroids and the largest overlapping fraction of Voronoi surfaces.

5.2 Transcription spot tracking and signal integration

To detect low-intensity signal-to-noise ratio (SNR) spots, our detection algorithm is tuned to high sensitivity at the cost of increasing the false positive rate when no obvious bright object throughout the detection region is present. Thus, we could find multiple spot candidates per nucleus at a given time point. However, in all single-allele-labeling experiments, we expect at most one spot per nucleus at all times. With this knowledge, a dedicated custom detection and tracking (modified Viterbi) algorithm identifies and tracks the real low SNR transcription spots based on the spot candidates.

5.2.1 Detecting spot candidates and transcription trajectories

Identification of transcription spots is simplified by dividing the original images into 3D image stacks containing individual nuclei. The transcription channel for each such stack, tracked across time, is filtered for salt-and-pepper noise with a 3D median filter and subsequently, a 3D difference of Gaussian (DOG) filter is applied to detect round, concave objects. These objects are initial candidates for identifying transcription active sites. If consecutive time points present multiple spot candidates, we determined the most likely association using the pair-wise distance between potential pairs and a diffusion-based potential. The latter is constructed iteratively from the random walk of all transcription spots across all time points. This potential is penalizing for less likely displacements. This procedure identifies the most likely trajectory for each nucleus.

5.2.2 Signal integration and transcription time-series

We calculated the mean pixel intensity of a 3D ellipsoid volume centered at the centroid coordinates of each transcription spot as identified by the spot-tracking procedure. The ellipsoid volume has diameters of (X, Y, Z)=(7, 7, 5) pixels, determined empirically based on the objective point spread function and the image pixel size (see image settings). Similarly, the mean local fluorescence background was estimated by averaging the intensities of all pixels in a larger 3D ellipsoid (i.e. (X, Y, Z)=(19, 19, 7) pixels, also centered at the spot centroid), except for the pixels of the inner spot ellipsoid. The fluorescence intensity of transcription spots is then quantified by subtracting this mean local background value from the mean spot intensity. This procedure is performed to avoid intensity contributions of freely floating MCP-GFP molecules within the nuclei from the transcription spot intensities. Our detection and tracking pipeline results in background-subtracted intensity time series, each corresponding to the real-time transcriptional activity of a single labeled allele within a nucleus.

5.2.3 AP position assignment

The anterior-posterior axis (AP) was determined based on a mid-sagittal image of the full embryo in the histone-RFP channel. The tips of the long axis of the full embryo set the positions of both the anterior and posterior poles. Both poles are routinely checked manually. The XY coordinates of the nuclei identified in the surface images are registered onto the mid-sagittal image to determine their AP coordinates in that reference frame, relative to the poles.

6 Data processing, calibration, and measurement errors

6.1 Spatiotemporal alignment of embryos

All data sets (consisting of multiple embryos) are realigned in time (using mitotic events as a set point) and space to minimize inter-embryo variabilities. These stem from differences in the speed of developmental progression (mainly due to temperature fluctuations at the sample) and from misalignment in space resulting from the AP-axis determination.

6.1.1 Developmental time normalization and temporal alignment

To determine the speed of developmental progression, we compute for each nucleus in each embryo the time between mitosis 12 and 13, i.e., the duration of nuclear cycle 13 (NC13) t_{13}. t_{13} is thus the time for individual nuclei between two mitotic events (that way accounting for the well-documented mitotic waves that propagate along the AP-axis [11]. It is identified when our nucleus-segmentation algorithm detects two separate entities (calling error $\sim 10\,\mathrm{s}$). The embryo-intrinsic developmental time is determined by estimating the average duration of NC13 over all nuclei of a given embryo, $\langle t_{13}\rangle = T_{13}$. The global mean measured over $n = 123$ embryos is $\langle T_{13}\rangle = 18.4 \pm 1.2\,\mathrm{min}$, corresponding to 6.5% variability at $24 \pm 1°\mathrm{C}$.

While most of our transcription measurements are performed during NC14, the mean duration of NC14 T_{14} is difficult to measure as the end of the cycle occurs beyond gastrulation and the completion of the cycle is asynchronous among cells. Even the onset of gastrulation (cells no longer form a monolayer) is difficult to time precisely due to the way our imaging is performed: specimen labeling is optimized to measure transcription, which does not allow for optimal determination of the onset of gastrulation. However strong correlations of the durations of earlier cycles (T_{11}, T_{12} and T_{13}, Pearson correlation $\rho_{11-12} = 0.75$ and $\rho_{12-13} = 0.81$) suggest that T_{13} serves as a good proxy for the speed of developmental progression. Thus each embryo is aligned in time to match this unified speed. Temporal alignment is achieved in each embryo by normalizing the sampling time Δt for both NC13 and NC14 by $T_{13}/\langle T_{13}\rangle$, leading to $\Delta t' = \Delta t \langle T_{13}\rangle / T_{13}$. Consequently, all embryos are aligned with respect to the onset and the end of NC13 (and thus the onset of NC14), making further translational temporal shifts between embryos unnecessary. The resulting temporal alignment typically provides a reduction in embryo-to-embryo variability (defined as the variance across mean embryo transcriptional activity) by a factor of two at 15 min into NC13 (near the end of NC13).

6.1.2 Embryo spatial alignment

All nuclei are projected on each embryo's internally determined AP-axis. However, axis determination is error-prone due to image analysis constraints, small changes in the azimuthal angle of the embryo with respect to the optical axis, and small tissue deformation due to the compression from the coverslip. We thus adjusted the spatial alignment for each embryo by correcting for possible errors stemming from these technical constraints. Indeed, the embryo elliptic mask fit (see Section 5.2.3) together with small deformation of the embryo surface and natural variability typically leads to a positional error of $\sim 1\%$ egg length (corresponding to approximately one cell diameter, which is statistically significant to our subsequent analyses, see also [12]). Spatial alignment is achieved directly on the image stack that measures gene activity. The mean and variance of gene activity are measured across nuclei for each embryo as a function of time and space within AP bins (bin sizes of 2.5% and 1.5% egg length in NC13 and NC14, respectively). For each embryo i and each nuclear cycle, we thus obtain an activity surface

(kymograph) for the mean $\mu_i(x,t)$ and variance $\sigma_i^2(x,t)$ as a function of AP position x and time t with respect to mitosis. We then introduce new coordinates for each embryo that account for spatial shift Δx, spatial dilatation α, and temporal dilatation β, i.e., $x' = \Delta x + \alpha(x - \langle x \rangle) + \langle x \rangle$ and $t' = \beta t$, such that $\tilde{\mu}_i(x',t') = \mu_i(x,t)$ and $\tilde{\sigma}_i^2(x',t') = \sigma_i^2(x,t)$. Next, we define least squares sums based on the shifted and dilated surfaces $\tilde{\mu}_i(x,t)$ and $\tilde{\sigma}_i^2(x,t)$:

$$\chi_1(x,t) = \sum_{i=1}^{N_e} (\tilde{\mu}_i(x,t) - \langle \tilde{\mu}_i(x,t) \rangle)^2,$$

$$\chi_2(x,t) = \sum_{i=1}^{N_e} (\tilde{\sigma}_i^2(x,t) - \langle \tilde{\sigma}_i^2(x,t) \rangle)^2,$$

$$\chi(\{\Delta x_i, \alpha_i, \beta_i\}) = \int\int (\chi_1(x,t) + \sqrt{\chi_2(x,t)}) dx dt,$$

where N_e is the total number of embryos for a given gene.

We then minimized $\chi(\{\Delta x_i, \alpha_i, \beta_i\})$ to learn Δx_i, α_i and β_i for each embryo, under the following constraints $\sum_i \Delta x_i = 0$, $\sum_i \log \alpha_i = 0$ and $\sum_i \log \beta_i = 0$. These constraints ensure that on average (over embryos), position and time remain the same, i.e., $\langle x_i \rangle = 0$, $\langle \alpha_i \rangle = 1$ and $\langle \beta_i \rangle = 1$. When learning the new alignment in NC13, no time dilation was needed, and $\beta_i = 0$, $\forall i$. The β_i were only learned for NC14 since the absence of well-defined events for the exact onset of gastrulation made the temporal alignment harder in that case. In the end, the standard deviation (std) for $\{\Delta x_i\}$ is 1.2% (NC13) and 0.8% (NC14) egg length, the std for α_i is 8% (NC13) and 6% (NC14), and the std for β_i is 5% (NC14).

6.2 Embryo-to-embryo variability and pooling

To increase our allele sample size in each condition, we use a pooling strategy, effectively pooling alleles from multiple spatially and temporally aligned embryos (see Section 6.1). This is facilitated by the high reproducibility of the *Drosophila* embryos, which intrinsically exhibit low embryo-to-embryo variability.

6.2.1 Embryo-to-embryo variability assessment

We characterize embryo-to-embryo variability on the measured transcriptional activity across embryos. As we have done for aligning embryos, we compute the mean and variance of gene activity across nuclei for each embryo as a function of time and space within AP bins (bin sizes of 2.5% and 1.5% egg length in NC13 and NC14, respectively). We thus obtain for each embryo i and each nuclear cycle the mean $\mu_i(x,t)$ and variance $\sigma_i^2(x,t)$ as a function of AP position x and time t with respect to mitosis. We define embryo-to-embryo variability as the variance of the mean σ_{emb}^2 across N_e embryos, i.e.,

$$\sigma_{\text{emb}}^2(x,t) = \sum_{i=1}^{N_e} w_i(x,t) \left(\mu_i(x,t) - \mu(x,t)\right)^2, \tag{1}$$

where the weight are given by $w_i(x,t) = 1/N_i(x,t)$ with $N_i(x,t)$ the number of alleles at time t in the AP-bin centered at position x for embryo i, and the global mean $\mu(x,t) = \sum_{i=1}^{N_e} w_i(x,t)\mu_i(x,t)$. Applying the law of total variance, the total variance in the measured transcriptional activity σ^2 is given by

$$\sigma^2(x,t) = \sigma_{\text{emb}}^2(x,t) + \sum_{i=1}^{N_e} w_i(x,t)\sigma_i^2(x,t), \tag{2}$$

where the last term corresponds to intra-embryo variability, i.e. variability across nuclei (plus a small contribution coming from measurement error, see Section 6.4). To determine the contribution of the embryo-to-embryo variability on the measured total variance, we compute the ratio $\sigma^2_{\text{emb}}(x,t)/\sigma^2(x,t) \leq 1$ for all gap genes in NC13 and NC14. As it has been previously measured in fixed embryos [12], the embryo-to-embryo variability remains small compared to the total variance throughout developmental time, overall averaging to less than 10% of the total variance (see Extended Data Fig. 1g for specific time points).

6.2.2 Pooling nuclei across multiple embryos

Having shown that embryo-to-embryo variability represents only a small fraction ($\sim 10\%$) of the total variability, we can treat (to a good approximation) individual embryos as independent samples drawn from the same underlying distribution. Thus, to characterize the fluctuations in transcriptional activity for a given gene, we pool together all the nuclei (at the same time and position along the AP axis) from all measured embryos (aligned in space and time beforehand, see Section 6.1) generating a large sample of nuclei (~ 200 nuclei). Specifically, once the embryos are spatiotemporally aligned, we define common spatial bins along the AP axis (bin sizes of 2.5% and 1.5% egg length in NC13 and NC14, respectively) allowing us to pool nuclei at same position and at each time point. Pooling provides a large gain in sample size (by a factor N_e =10–20 embryos), facilitating the estimation of the mean profiles and higher moments of the alleles' transcriptional activity (Section 6.3), as well as the estimation of the bursting parameters (Section 7). Moreover, embryo pooling is justified as 1) embryo-to-embryo variability is almost negligible, and 2) we focus explicitly on variability across nuclei, which mostly reflects transcription dynamics and the underlying bursts (provided the measurement noise is small). Hence we assume that for a given AP bin nuclei sample the same distribution function in each embryo.

6.3 Calibration of absolute units in live imaging data

We calibrate our live measurements to absolute units, which are defined as cytoplasmic units (C.U.) and correspond to equivalent single mRNA molecule counts at the site of transcription. The calibration is achieved by matching mean activity profiles in NC13 from live and previously calibrated fixed smFISH measurements [12]. Here we explain this procedure in detail.

6.3.1 Predicted live activity from smFISH

We use calibrated smFISH measurements of nascent transcriptional activity (in C.U.) for all trunk gap genes in late NC13 ($10-15$ min into NC13) to compute the predicted live activity. From these calibrated fixed measurements, we compute mean activity profiles for each gene by pooling the nuclei from multiple embryos ($n = 10-20$) within AP bins of 2.5% egg length (~ 2 nuclei wide region along the AP axis) (see Zoller et al. [12], Fig. 1C). The resulting smFISH mean activity profiles $\bar{A}_F(x)$ predict the expected levels of live mean activity $\bar{A}_L(x)$ (in C.U.) as a function of position x along the AP axis. We compute the expected $\bar{A}_L$ by adjusting $\bar{A}_F$ for differences in gene copy number ($n_F = 2$ alleles imaged in fixed vs $n_L = 1$ allele in live), gene length (endogenous length in fixed versus an additional ~ 1.5 kb length due to the insertion of the MS2 cassette in live measurements), and the specificity of targeted locations by FISH probes and stem-loops. We thus get

$$\bar{A}_L = \frac{n_L \tilde{L}_L}{n_F \tilde{L}_F} \bar{A}_F,$$

Gene	Annotated length [bp]	Physical length L_g [bp]	Effective length $\tilde{L}$ [bp]
hb-intron	4464	4464	3851
hb-3utr	3516	3516	2903
Kr-intron	3765	4565	3952
Kr-3utr	1915	2715	2102
kni	3049	3849	3236
gt	2940	3740	3127

Table 1: Gene length of the measured gap genes. The annotated gene length was obtained from UCSC Genome Browser (BDGP Release 6 + ISO1 MT/dm6) Assembly and it includes the length of the inserted MS2 cassette. The physical gene length is obtained by further adding 800bp to all genes except *hb*, and it accounts for the retention of elongated nascent transcripts at the sites. This extra length was estimated from dual color smFISH measurements [12]. The effective gene length is calculated from the physical gene length according to Equ. 4.

where $\tilde{L}_F$ and $\tilde{L}_L$ are effective gene lengths that account for the effect of probe locations on the signal.

6.3.2 Computing effective gene lengths

We compute the effective gene length $\tilde{L}$ for all constructs. Knowing the exact location of the binding sequences for the fluorescent smFISH probes (fixed) and of the MS2 sequences (live) along the gene, we calculate the relative contribution of a single nascent transcript (compared to a fully tagged transcript corresponding to one C.U.) to the signal $s(l)$ as a function of its length l:

$$s(l) = \frac{1}{m}\sum_{i=1}^{m} H(l - l_i), \tag{3}$$

where H is the (Heaviside) unit step function, l_i the end position of the i^{th} probe binding sequences and m the total number of probes. Integrating $s(l)$ over the gene length L_g (from TSS to polyA site) leads to the effective length:

$$\tilde{L} = \int_0^{L_g} s(l)dl = \frac{1}{m}\sum_{i=1}^{m}(L_g - l_i) = L_g - \langle l_i \rangle. \tag{4}$$

The resulting physical gene length L_g and the effective gene length $\tilde{L}$ are given in Table 1. The physical length L_g is slightly longer than the annotated gene length, as it accounts for a small retention time ($\sim 30\,\text{s}$) of nascent transcripts at the transcription site, due to termination and possibly Pol II running further than the annotated 3'UTR. We have previously estimated the equivalent quantity of this retention time as an extra transcribed length in dual color smFISH measurements [12].

6.3.3 Estimating conversion factor for profile matching

We compute the non-calibrated live mean activity profiles $\bar{S}(x)$ for all constructs by averaging the single-allele activity time series over all nuclei in a given 2.5% AP bin over a $10-15\,\text{min}$ time window in NC13, effectively pooling together nuclei from $n = 10 - 20$ embryos. The conversion factor ν (allowing us to

express S in absolute units) is estimated by defining a pseudo-likelihood function for a given gene g:

$$\mathcal{L}_g(\bar{S}|\nu,\Delta x)=\prod_{i=1}^{N_x}\mathcal{N}(\nu\bar{S}(x_i)|\mu=\bar{A}_L(x_i-\Delta x),\sigma^2=\sigma_A^2(x_i-\Delta x)+\nu^2\sigma_S^2(x_i)),$$

where $\mathcal{N}$ is the normal probability density function (i.e. a Gaussian), N_x the number of positions along the profiles, and Δx a possible small shift in space registration (typically $< 1\%$ egg length) between fixed and live experiments. σ_A and σ_S are standard errors of the mean based on the smFISH prediction and live averaging, respectively. Typically, ν accounts for the different imaging parameters (such as quantum yield and PMT amplification, etc.) between the fixed (confocal microscopy) and live (two-photon microscopy) experiments. The inference of the factor ν was performed globally over all genes at once, i.e., by maximizing

$$\mathcal{L}(\{\bar{S}_g\}|\nu,\{\Delta x_g\})=\prod_g\mathcal{L}_g(\bar{S}_g|\nu,\Delta x_g),$$

leading to a single ν that was used to define our global calibration unit $1/\nu$. We also inferred ν per individual gene to assess our calibration error (Extended Data Fig. 1b).

6.3.4 Comparison of higher cumulants

To further validate the quantitative nature and the proper calibration of our measurements, we compute the mean-to-higher cumulant relationships and compare them to previously extracted relationships where we used smFISH [12]. The first four cumulants of the measured activity A are defined as

$$\begin{aligned}
\kappa_1 &\equiv \mu = \langle A\rangle\\
\kappa_2 &\equiv \sigma^2 = \langle A^2\rangle - \langle A\rangle^2\\
\kappa_3 &= \langle A^3\rangle - 3\langle A^2\rangle\langle A\rangle + 2\langle A\rangle^3\\
\kappa_4 &= \langle A^4\rangle - 4\langle A^3\rangle\langle A\rangle - 3\langle A^2\rangle^2 + 12\langle A^2\rangle\langle A\rangle^2 - 6\langle A\rangle^4.
\end{aligned} \tag{5}$$

The cumulants are extensive, meaning that the contributions from independent random variables are additive, which is convenient as the measured transcriptional activities result from the sum of n_g independent gene copies. Assuming replicated sister chromatids, $n_g = 2$ for our live data (only one allele per nucleus has MS2 stem-loops) and $n_g = 4$ for smFISH data (both alleles in each nucleus are tagged with probes). Using the definition above, we compute the higher cumulants of the calibrated activity from both, smFISH and live measurements. The same pooling strategy as for the means is applied (see previous paragraphs), using 2.5% AP bins over a $10-15$ min time window in NC13.

We have previously demonstrated that the higher cumulants can be normalized, assuming Poisson background, such that we account for differences in gene length, probe locations, and gene copy number [12]. We compute the normalized cumulants κ_k' for a single gene copy as follows:

$$\kappa_k' = \frac{1}{n_g C_k}\left(\frac{L'}{L_g}\right)^k\kappa_k, \tag{6}$$

where n_g is the gene copy number, L_g the original gene length, L' the normalized gene length. The k-th order coefficients C_k are defined as

$$C_k = \frac{1}{L_g}\int_0^{L_g} s(l)^k dl = \sum_{i=1}^{m}\left(\frac{i}{m}\right)^k\frac{l_{i+1}-l_i}{L_g},$$

where $s(l)$ is the relative contribution of a single nascent transcript (Eq. 3). Note that for $k = 1$, we have $C_1 L_g = \tilde{L}$, which is the effective gene length.

We compare the normalized cumulants computed from live and smFISH data using a normalized gene length of $L' = 3.3$ kb (which corresponds to the average gap gene length in the absence of stem-loops, also used in Zoller et al. [12]). Once normalized, the first cumulant κ'_1 corresponds to the mean number of nascent transcripts $\langle g \rangle$ on a single gene copy of length L'. We plot the normalized higher cumulants κ'_2, κ'_3, and κ'_4 as a function of that $\langle g \rangle$ (Extended Data Fig. 1c). We further normalized all the cumulants by g_0^k. The value g_0 corresponds to the largest intercept between a k-th order polynomial fit to mean-cumulant relationships (continuous line for live and dotted line for smFISH data) and the straight line of the Poisson background (dashed line) set by $\kappa'_k = \langle g \rangle$. The number g_0 provides the maximal possible value of $\langle g \rangle$ based on the assumption that transcriptional variability κ'_2 cannot be lower than Poisson. Thus, g_0 can be interpreted as the mean number of nascent transcripts (on a 3.3 kb long gene) at maximal activity. We get $g_0 = 13.6$ from live and $g_0 = 15.2$ with smFISH measurements, a difference of 12% based on higher cumulants, not too far from the 5% error obtained from calibrating the means (Extended Data Fig. 1b). Overall, the higher cumulants versus mean relationships obtained from live (Extended Data Fig. 1c left column) and from smFISH (Extended Data Fig. 1c right column) measurements match almost perfectly (black solid versus dotted line), strongly confirming the quantitative nature and the proper calibration of both our live and fixed assay.

6.4 Measurement error

We evaluate imaging error through two distinct approaches: a direct measurement conducted through a specially designed experiment, and an auto-correlation analysis of the recorded transcription time series. The utilization of these two independent methods not only provides a robust internal control but also validates the reliability and effectiveness of our chosen approach. The convergence of results obtained from both approaches strengthens the credibility of our findings.

6.4.1 Dual-color measurement error assessment

To experimentally assess imaging or measurement error, we performed the same experiment twice and simultaneously with two different colors. A discrepancy in both experiments from a perfect match can be assigned to measurement error. We inserted a cassette of interlaced (alternating) MS2 and PP7 stem-loops in the first intron of *Kr*. The MS2 stem-loops were tagged with MCP-GFP (green) and the PP7 stem-loops with PCP-mCherry (red). By plotting the green channel against the red channel (Extended Data Fig. 1e, left), we characterize the spread of the data along the expected line of slope one (when correctly calibrated and without noise both channels are expected to perfectly correlate). We build a simple effective model to describe measurement noise in each channel:

$$P(A|G) = \mathcal{N}(A|G, \sigma^2(G)),$$

where, for a given channel, the measured transcriptional activity A (in C.U.) is normally distributed with mean G and variance $\sigma^2(G)$. G is the corresponding activity (in C.U.) in the absence of noise, and the variance σ^2 explicitly depends on G to account for the heteroscedasticity (i.e., heterogeneity of variance) in the data as the observed spread in the data increases with transcriptional activity.

To estimate the variance σ^2, we fit a straight line $y = \alpha x + \gamma$ assuming error on both $x \equiv A_{\text{green}}$ and

$y \equiv A_{\text{red}}$. We expand the variance as a function of the scalar projection along the line v:

$$\sigma^2(v) = \sigma_b^2 + \beta_1 v + \beta_2 v^2,$$
$$v = \frac{x + \alpha(y - \gamma)}{\sqrt{1 + \alpha^2}}.$$

Assuming the same error for *green* (x) and *red* (y), we maximize the following likelihood to estimate the set of parameters $\theta = \{\alpha, \gamma, \sigma_b, \beta_1, \beta_2\}$:

$$\mathcal{L}(\{x_i, y_i\}|\theta) = \prod_{i=1}^{N_d} \frac{1}{\sqrt{2\pi\sigma^2(v_i)}} \exp\left(-\frac{(y_i - \alpha x_i - \gamma)^2}{2(1+\alpha^2)\sigma^2(v_i)}\right),$$

where N_d is the total number of pairs of red-green data points.

Since only the green channel is calibrated in absolute units, we normalize the red channel such that the slope $\alpha = 1$. Using the Akaike information criterion [13], we find that the best model was parameterized by only (σ_b, β_1) with $\alpha = 1$, $\gamma = 0$ and $\beta_2 = 0$. The best fitting parameters are $\sigma_b = 1.22$ and $\beta_1 = 0.11$. The imaging error derived from the noise measurement model follows:

$$\sigma(G) = \sqrt{\sigma_b^2 + \tilde{\beta}_1 G}, \tag{7}$$

with $\tilde{\beta}_1 = \sqrt{2}\beta_1 = 0.16$. The resulting imaging noise is shown in Extended Data Fig. 1e,f. The functional form of the imaging error is exactly what one would expect from laser scanning microscopy, that is a background term σ_b^2 and a Poisson shot noise term $\tilde{\beta}_1 G$, the latter being proportional to the activity G.

6.4.2 Signal correlation-based imaging error assessment

Due to the elongation of individual nascent transcripts adding a persistent signal to the measured activity, we expect a strong temporal correlation in the measured single allele activity $a(t)$. Indeed, based on gene length, i.e. Table 1), and elongation rate, i.e. $K_{\text{elo}} = 1.8$kb/min, (see Section 7.1.6 for validation of deconvolution and estimation of elongation rate), imaged transcripts should persist at the transcription site for at least $\tau = 2\,\text{min}$ on average. Thus, at a short timescale near our sampling time $\Delta t = 10\,\text{s}$ ($\Delta t \ll \tau$), most biological variability should be correlated in time. Consequently, the remaining uncorrelated variability σ_u^2 is likely related to the imaging error as the latter is also expected to be temporally uncorrelated.

We estimate the uncorrelated variability σ_u^2 from the transcriptional activity time series $A(t)$ within each AP bin for each gene individually. To this end, we compute the time-dependent mean activity $\mu(t) = \langle A(t)\rangle$, variance $\sigma^2(t) = \langle (A(t) - \mu(t))^2\rangle$ and covariance $\text{Cov}(t, t+\Delta t) = \langle (A(t) - \mu(t))(A(t+\Delta t) - \mu(t+\Delta t))\rangle$, where Δt is the sampling time (and the averaging is performed over all alleles in each AP bin pooled from multiple embryos). For a Δt sufficiently small compared to the correlation time τ imposed by elongation, $\sigma_u^2(t) = \sigma^2(t) - \text{Cov}(t, t + \Delta t)$ should be a good estimate of the uncorrelated variability in the data.

When we plot $\sigma_u^2(t)$ as a function $\mu(t)$ for all genes and AP bins across NC13 and NC14 (Extended Data Fig. 1e right), we observe a strong linear relationship between $\mu(t)$ and $\sigma_u^2(t)$. That is, $\sigma_u^2(t) = \sigma_b^2 + \beta\mu(t)$ with $\sigma_b = 0.75$ and $\beta = 0.19$, which is reminiscent of the functional form that we find using the interlaced approach above (Eq. 7). Indeed, we find again that the variance is given by a background σ_b^2 and a Poisson shot noise term $\beta\mu$ that is proportional to the mean activity. The interlaced- and the correlation-based quantifications of the imaging error are numerically very close over the entire range of measured activities with a sensitivity close to a single mRNA (Extended Data Fig. 1f).

7 Data analysis and modeling

7.1 Deconvolution of the single-allele transcription rate

7.1.1 Modeling initiation events

We designed a simple model to reconstruct the individual initiation events of productive transcription by Pol II from the calibrated single allele transcriptional activity $A(t)$ (in C.U.). Our model for measurement noise $\sigma(G)$ (Eq. 7) links the activity $A(t)$ to the signal in the absence of measurement noise $G(t)$ by:

$$A(t) = G(t) + \sigma(G(t))\eta(t), \tag{8}$$

where $\eta(t)$ is an uncorrelated Gaussian white noise with zero mean and unit standard deviation. The signal $G(t)$ is expressed as a convolution between a kernel $\kappa(t)$ (modeling the fluorescent signal resulting from the Pol II elongation process through the stem-loop cassette) and a function $I(t)$ that models Pol II initiation events:

$$G(t) = \kappa(t) * I(t). \tag{9}$$

The initiation function I(t) can be expressed as

$$I(t) = \sum_{i=1} \delta(t - t_i), \tag{10}$$

where δ are unit pulses (δ-Dirac) representing to individual initiation events at time t_i. The kernel function $\kappa(t)$ is built according to the signal contribution of individual transcripts (see Eq. 3):

$$\kappa(t) = \frac{1 - H\left(t - \frac{L_g}{K_{\text{elo}}}\right)}{m} \sum_{i=1}^{m} H\left(t - \frac{l_i}{K_{\text{elo}}}\right), \tag{11}$$

where H is the unit (Heaviside) step function, L_g the physical length of the gene, l_i the end position of the i^{th} MS2 stem-loop sequence and m the total number of loops (in our construct $m = 24$). The key assumptions behind this kernel are i) constant elongation rate K_{elo}, ii) deterministic elongation (limited fluctuations, no pausing in the gene body), iii) fast termination, and iv) no co-transcriptional splicing (with short gene lengths, facilitating fast transcription). Importantly, all the parameters in Eq. 11 are determined by the DNA sequence, except the elongation rate K_{elo} that needs to be inferred from measurements.

7.1.2 Bayesian deconvolution of transcription initiation events

We performed Bayesian deconvolution to reconstruct the possible configuration of initiation events I from individual single allele activity time series $A = (A_1, A_2, \ldots, A_{N_t})$, where $A_i \equiv A(t_i)$, $t_{i+1} - t_i = \Delta t \approx 10$s $\forall i$ and N_t the number of time points. We first wrote the likelihood of the measured activity time series A given the true signal time series G (discretely sampled at the same time points as A):

$$P(A|G) = \prod_{i=1}^{N_t} \mathcal{N}(A_i|G_i, \sigma^2(G_i)). \tag{12}$$

Individual time points, A_i, are assumed to be normally distributed with mean G_i and variance $\sigma^2(G_i)$ given by our measurement noise model (see Eq. 7).

According to our model above (Eq. 9), the signal G results from the convolution of $\kappa * I$. To compute this convolution, we first discretize the time variable t in the kernel $\kappa(t)$ and the initiation function $I(t)$ (see Eq. 10 and 11) using a smaller sampling time $\Delta t' = f/(2K_{\text{elo}}) = 1\,\text{s}$, where $f = 60\,\text{bp}$ is a conservative estimate on the Pol II footprint on the DNA, $K_{\text{elo}} = 1.8\,\text{kb/min}$ is the measured elongation rate (see Section 7.1.6) and the factor 2 accounts for having two unresolved active sister chromatids. $\Delta t' = 1\,\text{s}$ is the minimal physically possible time interval between two consecutive Pol II loading events (oversampling beyond $\Delta t'$ is superfluous). When discretized, the initiation configuration I becomes a binary vector $I = (I_1, I_2, \ldots, I_{N_I})$ with $I_i \equiv I(t'_i) \in \{0, 1\}$, $t'_{i+1} - t'_i = \Delta t'\ \forall i$ and $N_I = \lceil N_t \Delta t / \Delta t' \rceil - 1$. Thus G can be computed as a discrete convolution:

$$G_i = \sum_{j=1}^{\min(k,i)} I_{i-j+1} \kappa_j,$$

where $k = \lceil L_g/(K_{\text{elo}} \Delta t') \rceil + 1$ is the number of time points needed to discretize the kernel κ and L_g the physical length of the gene (see Table 1). The resulting time series G is then downsampled to evaluate $P(A|G)$ (Eq. 12) at the same time points as the measured activity time series A. We thus have effectively designed a simple procedure to evaluate $P(A|I)$, i.e., the likelihood of A given an initiation configuration I:

$$P(A|I) = \sum_G P(A|G) P(G|I), \tag{13}$$

where $P(G|I) = 1$ if $G = \kappa * I$ and $P(G|I) = 0$ otherwise.

Applying Bayes' theorem gives us the posterior distribution $P(I|A)$, i.e., the distribution of possible configurations of initiation events I given a measured single allele activity time series A:

$$P(I|A) = \frac{P(A|I)P(I)}{\sum_I P(A|I)P(I)}, \tag{14}$$

where $P(I)$ is the prior distribution on the configuration of initiation events (see Section 7.1.3). The posterior $P(I|A)$ is our Bayesian representation of the deconvolved initiation configurations I and naturally provides the uncertainty on these configurations as a probability distribution. Importantly, $P(I|A)$ accounts for measurement noise and incorporates our assumptions about the elongation process underlying the fluorescent signal. Overall, the Bayesian approach to our deconvolution problem has a few advantages over maximum likelihood estimation (MLE). First, in the presence of measurement noise, there could be a risk with MLE that the unique resulting configuration I (the one that maximizes $P(I|A)$) overfits the noise. In contrast, Bayesian estimation readily provides all the possible configurations compatible with measurement noise. Second, when the transcriptional activity is high and many Pol II/transcripts contribute to the measured activity A, the underlying initiation configurations I may be degenerate (meaning that multiple I once convolved with the kernel κ would lead to the same or very close A. In that context, choosing a single configuration seems arbitrary, as MLE would do, whereas Bayesian estimation naturally captures the symmetries underlying the initiation configurations.

7.1.3 Prior on initiation configurations

A prior $P(I)$ over the initiation configurations needs to be defined in order to calculate the posterior distribution $P(I|A)$ in Eq. 14. As $I_i \in \{0, 1\}$, expressing the prior $P(I)$ as a product of multiple independent Bernoulli distributions represents a simple and natural solution:

$$P(I) = \prod_{i=1}^{N_I} p_i^{I_i} (1 - p_i)^{1 - I_i}, \tag{15}$$

with the parameters $p_i \in [0, 1]$ for each time point. For a given gene and AP bin, we set these parameters to $p_i = R(t_i)\Delta t'$, where $R(t)$ is the time-dependent mean transcription rate in the AP bin. A good proxy for $R(t)$ is given by $R(t) \approx \mu(t + \tau_{\text{elo}}/2)/\tau_{\text{elo}}$, where $\mu(t)$ is the mean activity in the AP bin and τ_{elo} the effective elongation time of the gene given by $\tilde{L}/K_{\text{elo}}$ (see Eq. 23 and Table 1). Thus, our prior mainly encapsulates our knowledge about the mean transcription rate and not much about the underlying single allele initiation configurations. Interestingly, a constant prior $P(I) = \text{cst.}$ implies that $p_i = 0.5\ \forall i$, which is equivalent to a constant mean transcription rate set at $R_{\max}/2$ with $R_{\max} = 1/\Delta t' = 60\,\text{mRNA/min}$ the maximal physical rate given the Pol II footprint and the measured elongation rate. However, in practice, the prior only has a small impact on inferred initiation configurations as the data is strong, i.e., given our imaging sensitivity and a rather short elongation time the likelihood $P(A|I)$ strongly constrains the possible configurations I allowed by the prior.

7.1.4 MCMC sampling of initiation configurations

We designed a Monte-Carlo Markov Chain (MCMC) algorithm [14], to directly sample from the posterior distribution $P(I|A)$ in Eq. 14. This method relies on building a Markov Chain $P(I'|I)$, whose stationary distribution $\pi(I)$ is the target distribution, i.e., $\pi(I) \equiv P(I|A)$. Such a Markov chain is obtained by imposing detailed balance $P(I'|I)\pi(I) = P(I|I')\pi(I')$, which ensures that the chain has the correct stationary distribution. Once the chain has asymptotically reached after a few iterations the stationary distribution, samples drawn from $P(I'|I)$ at each subsequent iteration are equivalent to samples directly drawn from $P(I'|A)$ (where I' is the new sample given the previous one I).

In practice, the sampling procedure at each iteration is performed in two steps. First, a new sample I' is drawn from a proposal distribution $Q(I'|I)$. Second, the proposed sample I' is accepted ($I' \to I$) or rejected ($I \to I$), with acceptance probability $\alpha(I', I)$:

$$\alpha(I, I') = \min\left(1, \frac{P(A|I')P(I')Q(I, I')}{P(A|I)P(I)Q(I', I)}\right), \tag{16}$$

where $P(A|I)$ is the likelihood (Eq. 13) and $P(I)$ is the prior (Eq. 15). The proposal distribution $Q(I'|I)$ must be carefully chosen to ensure a good acceptance probability (a good acceptance is typically between 10–70%) [15, 16]. Indeed, if the acceptance is too high (I' is too close to I) or too low (I' is too far from I), the chain will generate highly correlated samples (poor mixing), and the sampling will be inefficient (slow convergence). In principle, a good proposal is a distribution that closely mimics the target distribution. Since the target distribution is a priori unknown, we use an adaptative method to learn $Q(I'|I)$ on the fly such that it approaches the target distribution [15, 16]. In that case, the proposal distribution can typically be chosen as independent from the previous sample $Q(I'|I) = Q(I')$, and our prior (Eq. 15) represents a decent candidate proposal distribution:

$$Q(I) = \prod_{i=1}^{N_I} p_i^{I_i}(1 - p_i)^{1-I_i} \tag{17}$$

Here, the goal is to learn the parameters $q_i \in [0, 1]$ while sampling the configurations I. As initial values for the q_i, we chose $q_i = \tilde{A}(t_i')\Delta t'$, where $\tilde{A}(t) = A(t + \tau_{\text{elo}}/2)/\tau_{\text{elo}}$ with $A(t)$ the activity time series and τ_{elo} the elongation time. After each iteration of the MCMC sampler (after the acceptance or rejection of the proposed I'), we then update the q_i according to the following adaptative scheme:

$$q_i' = q_i(1 - l) + lI_i,$$

where q_i' is the new q_i, I the current configuration, and the learning rate $l = 1/\sqrt{3j}$ with j the iteration number. The function l, which vanishes when $j \to \infty$, is chosen such that the Markov chain converges despite adaptation. Importantly, we enforce $q_i > 0.05\ \forall i$ at each iteration to ensure that all time points of configuration I are sampled.

In practice, even with adaptation, attempting to update the whole configuration I at once according to the proposal (Eq. 17) can lead to sub-optimal acceptance. Indeed, the I_i's are often quite correlated, whereas our proposal assumes independent I_i. A simple workaround is to update the I_i's by block and increase the number of iterations: only the I_i from a subset $i \in \{b, \ldots, b+w-1\} \subset \{1, \ldots, N_I\}$ are drawn from the proposal (Eq. 17) at each iteration (while the remaining I_i are left unchanged), with b a random integer and w the size of the block. The size w is then tuned to achieve the desired acceptance ratio. We found that block size w covering approximately 6 min led to a good compromise between a decent acceptance ratio and sampling speed. To sample I, we performed $N_s N_I/w$ iterations with $N_s = 3500$ and we discarded the first $N_b N_I/w$ iteration, where $N_b = 500$.

7.1.5 Estimating single-allele transcription rate

Once the posterior distribution $P(I|A)$ of possible initiation configuration has been sampled by MCMC, we compute the single-cell transcription rate r. For each configuration I (Eq. 10 and 13), we can compute a single $r(t)$ as follows:

$$r(t) = \frac{1}{v\Delta t} \int_{t-v\Delta t}^{t} I(t')dt' = \frac{1}{v\Delta t} g(t), \tag{18}$$

where $g(t)$ is the number of initiation events during the time interval $t \in [t - v\Delta t, t]$ given by

$$g(t) = \sum_{i \in \{i : t-v\Delta t \leq t_i' \leq t\}} I_i.$$

In the equations above, $\Delta t \approx 10\,\mathrm{s}$ is our sampling time, and v is a positive integer defining the time window over which r is estimated. Except when explicitly specified otherwise, we estimated r using $v = 1$ which we defined as the instantaneous single allele transcription rate.

Using our MCMC samples, $P(r|A)$ is easily estimated by

$$P(r|A) = \sum_{I} P(r|I)P(I|A), \tag{19}$$

where $P(r|I) = 1$ if Eq. 18 is satisfied, and $P(r|I) = 0$ otherwise. The distribution $P(r|A)$ enables the computation of the instantaneous mean and standard deviation of $r(t)$ for individual single-allele activity time series A (Fig. 1e). Of note, $P(r|A)$ is actually a discrete probability distribution since we have a finite amount of MCMC samples and r is discrete (as is g).

7.1.6 Validation of deconvolution and estimation of elongation rate

To validate the hypothesis behind our kernel-based deconvolution (see Section 7.1.1), we tested our approach on dual color data, where the MS2 and PP7 stem-loop cassettes were inserted in the first intron and 3'UTR of *hb*, respectively. Similarly, our approach was tested on a data set for an additional gene, *Kr*, where the stem-loop insertions were permuted (PP7 in intron and MS2 in 3'UTR). In both cases MS2 stem-loops were tagged with MCP-GFP (green) and PP7 stem-loop via PCP-mCherry (red).

After calibrating the green and red channels in absolute units (see Section 6.3), we expressed the likelihood of obtaining the green A_{g} and the red activity A_{r} together, which is simply $P(A_{\mathrm{g}}|G_{\mathrm{g}})P(A_{\mathrm{r}}|G_{\mathrm{r}})$, where G_{g} and G_{r} are the true signals in each channel. Both $P(A_{\mathrm{g}}|G_{\mathrm{g}})$ and $P(A_{\mathrm{r}}|G_{\mathrm{r}})$ are given by Eq. 12, with $\sigma^2(G)$ that is specific to each channel (see Section 6.4). In principle, the true signals G_{g} and G_{r} should be highly correlated since they should result from the same initiation configuration I. $G_{\mathrm{g}} = \kappa_{\mathrm{g}} * I$ and $G_{\mathrm{r}} = \kappa_{\mathrm{r}} * I$ contain two kernels κ_g and κ_r that are determined from Eq. 11. Whereas these kernels mainly differ in the location of the stem-loop insertions, they share one free parameter, i.e., the elongation rate K_{elo} that is a priori unknown. Importantly, the overall delay between G_{g} and G_{r} should be given by the relative position $|l_{\mathrm{g}} - l_{\mathrm{r}}|$ of the MS2 and PP7 stem-loop cassettes divided by K_{elo}. It implies that K_{elo} can be extracted from these dual-color data. With these ingredients, the likelihood function for the pair A_{g} and A_{r} is as follows:

$$P(A_{\mathrm{g}}, A_{\mathrm{r}}|I, K_{\mathrm{elo}}) = \sum_{G_{\mathrm{g}}, G_{\mathrm{r}}} P(A_{\mathrm{g}}|G_{\mathrm{g}})P(A_{\mathrm{r}}|G_{\mathrm{r}})P(G_{\mathrm{g}}, G_{\mathrm{r}}|I, K_{\mathrm{elo}}), \tag{20}$$

where $P(G_{\mathrm{g}}, G_{\mathrm{r}}|I, K_{\mathrm{elo}})$ is trivially determined from $G_{\mathrm{g}} = \kappa_{\mathrm{g}}(K_{\mathrm{elo}}) * I$ and $G_{\mathrm{r}} = \kappa_{\mathrm{r}}(K_{\mathrm{elo}}) * I$ (see Eq. 13).

Next, from our dual-color data, we inferred I and K_{elo} for each nucleus using the likelihood in Eq. 20. Specifically, we sampled from the posterior distribution $P(I, K_{\mathrm{elo}}|A_{\mathrm{g}}, A_{\mathrm{r}})$ according to our MCMC sampler (see Section 7.1.4). We set the prior of K_{elo} to log-uniform $P(K_{\mathrm{elo}}) \sim 1/K_{\mathrm{elo}}$. For both *hb* and *Kr* data, we assessed the quality of the deconvolution by computing the normalized residuals for each channel $\mathrm{c} \in \{\mathrm{g}, \mathrm{r}\}$ and each nucleus:

$$z_{\mathrm{c}} = \frac{A_{\mathrm{c}} - \bar{G}_{\mathrm{c}}}{\sigma_{\mathrm{c}}(\bar{G}_{\mathrm{c}})},$$

where $\sigma_{\mathrm{c}}(G)$ is the measurement error of channel c, and $\bar{G}_{\mathrm{c}}$ the mean reconstructed signal computed as the marginal mean of

$$P(G_{\mathrm{g}}, G_{\mathrm{r}}|A_{\mathrm{g}}, A_{\mathrm{r}}) = \sum_{I} \int P(G_{\mathrm{g}}, G_{\mathrm{r}}|I, K_{\mathrm{elo}})P(I, K_{\mathrm{elo}}|A_{\mathrm{g}}, A_{\mathrm{r}})dK_{\mathrm{elo}}.$$

Since the residuals z_{c} are given by a time series, we thus reported the mean $\mu(z_{\mathrm{c}})$ and the standard deviation of $\sigma(z_{\mathrm{c}})$ in the two channels for each nucleus (Extended Data Fig. 2c). We then compared the $\mu(z_{\mathrm{c}})$ and $\sigma(z_{\mathrm{c}})$ to a perfect model for which the z_{c} have the same length than our data and each $z_{\mathrm{c}}(t)$ is independent and normally distributed with $\mu = 0$ and $\sigma = 1$. We computed the 95% confidence ellipse of $\mu(z_{\mathrm{c}})$ and $\sigma(z_{\mathrm{c}})$ for both our data (solid black line in Extended Data Fig. 2c) and the perfect model (dotted black line in Extended Data Fig. 2c). Our reconstructions are close to the perfect model, meaning that the two channels can be reconstructed from the same initiation configurations within confidence intervals, overall justifying our kernel assumptions. Indeed, the good agreement between the predicted and the measured dual color signal implies that the elongation process must be close to deterministic (with approximately constant elongation rate) and co-transcriptional splicing must be limited for these short genes. Otherwise, we would have observed a striking mismatch between the model prediction and the measurements in the 5'-to-3' amplitude ratio and the 5'-to-3' delay.

Lastly, we estimate the elongation rate K_{elo} for each individual nucleus numerically as the mean of the marginal posterior distribution $P(K_{\mathrm{elo}}|A_{\mathrm{g}}, A_{\mathrm{r}})$. Within each embryo and for both *hb* and *Kr*, the average K_{elo} across nuclei at different AP bins remains stable along the AP axis (Extended Data Fig. 2d). We estimated the global mean K_{elo} by averaging all nuclei for each embryo, the resulting averages are very similar across genes and embryos (Extended Data Fig. 2e). The overall mean elongation rate is $\bar{K}_{\mathrm{elo}} = 1.8 \pm 0.1\,\mathrm{kb/min}$.

7.2 Gap gene mean expression pattern

7.2.1 Transcription expression pattern

From our deconvolution-based estimates of the single-allele transcription rate $r(t)$, we reconstruct the mean expression pattern as a further test for the validity of our approach. For each gene, we align the imaged embryos in space and time, allowing us to pool nuclei from different embryos (see Section 6.2). Thus, for a given gene and nuclear cycle, we obtain a set $\mathcal{A}(x,t)$ of activity time series $A \in \mathcal{A}(x,t)$ coming from all the nuclei in the same AP bin at time t (of width 2.5% and 1.5% egg length in NC13 and NC14, respectively) centered at position x. The pooling procedure results in the distribution of single-allele transcription rate $P(r|x,t)$ defined as the following mixture:

$$P(r|x,t) = \sum_{A \in \mathcal{A}(x,t)} P(r(t)|A)P(A), \tag{21}$$

where $P(r|A)$ is the distribution of deconvolved r given a single-allele time series A (defined in Eq. 19) and $P(A) = 1/N_A(x,t)$ with $N_A(x,t) = \#\mathcal{A}(x,t) \sim 200$ the total number of nuclei in the AP bin at position x and time t. From $P(r|x,t)$, we compute the time-dependent mean transcription rate $R(x,t)$ at position x:

$$R(x,t) = \langle r(x,t) \rangle = \sum_r r P(r|x,t). \tag{22}$$

Using $R(x,t)$, we reconstruct the mean profiles for all genes as a function of time t, which are displayed for three different time points in Extended Data Fig. 3a (color dots).

To validate our estimation of the mean transcription rate $R(x,t)$, we compare it to an alternative approximation that is independent of the deconvolution procedure:

$$R(x,t) \approx \frac{\mu(x, t + \tau_{\text{elo}}/2)}{\tau_{\text{elo}}}, \tag{23}$$

where $\mu(x,t)$ is the time-dependent mean activity in the AP bin centered at position x and τ_{elo} the effective elongation time of the gene given by $\tilde{L}/K_{\text{elo}}$ (see Table 1). We see in Extended Data Fig. 3a that both approaches (color and black dashed line for deconvolution approach and approximation via Eq. 23, respectively) strongly agree and can thus be used interchangeably to estimate the mean transcription rate $R(x,t)$.

7.2.2 Comparison with protein pattern

We further tested whether our mean gap gene patterns measured directly from the transcriptional output, i.e., MS2-signal (Extended Data Fig. 4a, left column) are consistent with previously measured protein patterns (Extended Data Fig. 4a, right column) [4]. To this end, we extrapolate protein accumulation from the characterized transcriptional mean output using a minimal model (Extended Data Fig. 4a, middle column).

We model the accumulation of protein as a convolution between the mean transcription rate $R(x,t)$ and a kernel $\kappa(x,t)$ that accounts for diffusion, degradation, and a delay in protein production:

$$\kappa(x,t) = \begin{cases} \exp\left(-\dfrac{t-\delta}{\tau_m}\right) \exp\left(-\dfrac{x^2}{4D(t-\delta)}\right) & \text{if} \quad t - \delta \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

Thus $\kappa(x,t)$ above is parameterized by three free parameters: a diffusion constant D, a mean lifetime τ_m, and a delay δ. We compute the predicted protein pattern $\tilde{P}(x,t)$ as bidimensional convolution (along x and t, hence $**$ symbol) $\tilde{P}(x,t) = R(x,t)**\kappa(x,t)$, where $R(x,t)$ is estimated using Eq. 23. During NC14, we multiply $R(x,t)$ by a factor of two to account for the doubling of nuclei during mitosis 13. In addition, the resulting $\tilde{P}(x,t)$ is normalized such that its maximum value is one to facilitate direct comparison with previously similarly normalized protein data [4].

To estimate the three kernel parameters $\{D, \tau_m, \delta\}$ (Eq. 24), we minimize the squared difference between the predicted protein pattern $\tilde{P}(x,t)$ and the measured protein pattern $P(x,t)$. We performed the optimization for all gap genes individually and together (Extended Data Fig. 4b). The resulting parameters are in line with previous estimates and the modeled accumulation is consistent with previously measured protein patterns (Extended Data Fig. 4a). **Given our minimal model, the agreement between the modeled and measured protein patterns suggests that protein levels are mainly determined by transcription.** Indeed, low absolute errors (Extended Data Fig. 4c,d) and high structural correlations (Extended Data Fig. 4e,f) between the modeled and measured patterns support this hypothesis.

7.3 Transcription fluctuation analysis

7.3.1 Distributions of single-allele transcription rates

To characterize the transcriptional fluctuations around the mean transcription rate $R(x,t)$, we aimed to reconstruct the conditional distribution $P(r|R)$ of single-allele transcription rate r for each gene. In principle, $P(r|R)$ should exhibit key signatures of the transcriptional regime at play. Importantly, to magnify these signatures, here we specifically estimate the single-allele transcription rate r over a $\tau = 1$–min time window ($v = 6$, in Eq. 18). Indeed, with limited time-averaging, i.e., r estimated over $\Delta t = 10$s, the distribution $P(r|R)$ should approach the Poisson distribution at all R.

For each gene and nuclear cycle, our deconvolution method enables estimating the distribution of single-allele transcription rate $P(r|x,t)$ by pooling nuclei at position x and time point t (see Eq. 21). The mean transcription rate is then given by $R(x,t) = \langle r(x,t)\rangle = \sum_r rP(r|x,t)$ (Eq. 22).

We reconstruct the conditional distribution $P(r|R)$ by pooling together all nuclei from the distributions $P(r|x,t)$ at all position x and time t that satisfy a specific range of means $|R(x,t) - R_0| \leq \epsilon$ (where ϵ is a positive parameter). By defining the set $\mathcal{V}(R_0) = \{(x,t) : |R(x,t) - R_0| \leq \epsilon\}$, we express the conditional as a mixture of the single-allele transcription rate distributions $P(r|x,t)$:

$$P(r|R_0) = \sum_{(x,t)\in\mathcal{V}(R_0)} P(r|x,t)P(x,t) \tag{25}$$
$$P(x,t) = \frac{N_A(x,t)}{\sum_{(x,t)\in\mathcal{V}(R_0)} N_A(x,t)},$$

where $N_A(x,t)$ is the number of nuclei in the bin at position x and time t. Importantly, a reconstruction of $P(r|R_0)$ through pooling is only valid under the condition that the distributions $P(r|x,t)$ for which $(x,t) \in \mathcal{V}(R_0)$ are mostly similar.

To verify the validity of our reconstruction of $P(r|R)$ for each gene, we binned the range of observed R values in 15 equal bins from 0 to 16 mRNA/min. Each respective R_0 is thus defined as the centers of these R-bins, which sets $\epsilon = 0.53$ (the half width of the R bins). We then computed the overall $P(r|R)$ by applying Eq. 25 for each R_0. Given $P(r|R_0)$, we determined the 95% confidence envelope on its cumulative

distribution. We assessed what fraction of the individual $P(r|x,t)$ (with $(x,t) \in \mathcal{V}(R_0)$) are within the global 95% envelope. It turns out that when pooling alleles irrespective of developmental time, we observe a significant fraction of (x,t)-bins (position and time point) in NC14 for which $P(r|x,t)$ significantly deviates from the global distribution (Extended Data Fig. 3b, bottom right). When performing the same operation on three separate periods – NC13 ($t \geq 6.5$ min after mitosis) plus early NC14 ($7.5 \leq t < 20.5$ min), mid-NC14 ($20.5 \leq t < 34.5$ min) and late NC14 ($34.5 \leq t < 48$ min) – the agreement between the individual $P(r|x,t)$ and the overall distribution $P(r|R)$ improves significantly (Extended Data Fig. 3b). This suggests that transcription might be undergoing slow kinetic changes over developmental time, warranting a partitioning of NC14 into broad periods that are roughly stationary.

Having checked the pooling validity underlying the reconstruction of $P(r|R)$, we then analyze the single-allele transcription rates underlying the period encompassing NC13 and early NC14 ($7.5 \leq t < 20.5$ min). Within this period and for each of the 15 R_0 values defined as the center of the R-bin, we reconstruct the conditional $P(r|R)$ distribution according to Eq. 25. We show the resulting distributions for each gene in Extended Data Fig. 3c. Strikingly, the conditional $P(r|R)$ at a given $R = R_0$ value looks very similar across genes and periods (NC13 vs early NC14). Moreover, these distributions differ from the typical distribution one would expect in a constitutive (i.e., non-bursting) regime.

7.3.2 Constitutive transcriptional regime

In a constitutive regime, initiation events are assumed to happen as a simple Poisson process at a constant rate of R. Thus over some time τ, we would expect to observe g initiation events with probability given by the Poisson distribution:

$$P_P(g|\lambda = R\tau) = \frac{\lambda^g \exp(-\lambda)}{g!}.$$

Given our definition of the single-allele transcription rate (see Eq. 18), we have $r = g/\tau$ implying that in a constitutive regime $P(r|R) = P_P(g = r\tau|\lambda = R\tau)$. Such a description of the constitutive regime implicitly assumes that Pol II has no footprint, meaning that initiation could happen in very quick succession. However, this is unrealistic as the Pol II footprint $f = 60$bp (conservative estimate) should set a refractory period $\Delta t' = f/(2K_{\text{elo}}) = 1$s below which successive initiations cannot occur (see Section 7.1.2). As a simple correction to account for such a refractory period, we discretize the period $\tau > \Delta t'$ using the interval $\Delta t'$ as we did for our deconvolution procedure. In that case, the probability to observe g initiation events over some time τ would be given by the Binomial distribution:

$$P_B(g|n_t = \lceil \tau/\Delta t' \rceil, p = R\Delta t') = \binom{n_t}{g} p^g (1-p)^{n_t - g}. \tag{26}$$

Thus, considering the Pol II footprint, the distribution of single-allele transcription rate in a constitutive regime should be well approximated by $P(r|R) = P_B(g = r\tau|n_t = \lceil \tau/\Delta t' \rceil, p = R\Delta t')$. Of note, when $\tau \gg \Delta t'$ ($\Delta t' \to 0$), we recover the Poisson distribution and $P(r|R) = P_P(g = r\tau|\lambda = R\tau)$. This means that the binomial correction to the Poisson regime only matters when $\Delta t' \lesssim \tau$ and it becomes negligible when $\Delta t' \ll \tau$.

To highlight the difference between the empirically determined conditional distributions $P(r|R)$ and the constitutive "Poisson" regime, we plotted $P_B(g = r\tau|n_t = \lceil \tau/\Delta t' \rceil, p = R\Delta t')$ in Extended Data Fig. 3c (dashed line), where we used $\tau = 1$ min. We also computed the 2$^{\text{nd}}$ to 4$^{\text{th}}$ cumulants from the empirically determined conditional distributions $P(r|R)$ and the constitutive "Poisson" regime as a

function of R. The cumulants are defined as follows:

$$\begin{aligned}\sigma^2(R) &= \langle r^2\rangle - \langle r\rangle^2\\ \kappa_3(R) &= \langle r^3\rangle - 3\langle r^2\rangle\langle r\rangle + 2\langle r\rangle^3\\ \kappa_4(R) &= \langle r^4\rangle - 4\langle r^3\rangle\langle r\rangle - 3\langle r^2\rangle^2 + 12\langle r^2\rangle\langle r\rangle^2 - 6\langle r\rangle^4\end{aligned}$$

where $\langle r^k\rangle = \sum_r^k P(r|R)$. Assuming a constitutive regime (Eq. 26), we get the following cumulants:

$$\begin{aligned}\sigma^2(R) &= n_t p(1-p)/\tau^2 = R(1-R\Delta t')/\tau\\ \kappa_3(R) &= n_t p(1-p)(1-2p)/\tau^3 = R(1-R\Delta t')(1-2R\Delta t')/\tau^2\\ \kappa_4(R) &= n_t p(1-p)(1-6p(1-p))/\tau^4 = R(1-R\Delta t')(1-6R\Delta t'(1-R\Delta t'))/\tau^3\end{aligned}$$

We plot the cumulants from the empirically determined conditional distributions $P(r|R)$ (color) and the constitutive "Poisson" regime (black dashed line) as a function of R in Extended Data Fig. 3d. Overall, the empirical cumulants deviate significantly from the "Poisson" regime, except at the extreme ends of the R spectrum where we observe a convergence towards the constitutive regime. This can be interpreted as evidence for a bursting regime spanning all the way from $P_{\mathrm{ON}} = 0$ to $P_{\mathrm{ON}} = 1$ [12]. Errors on R and the cumulants were computed from bootstrapping.

7.3.3 Transcription rate auto-correlation

To characterize the transcriptional fluctuations over time, we compute the auto-correlation of the instantaneous single-allele transcription rate $r(t)$ ($v = 1$, in Eq. 18). For a given gene, nuclear cycle, and for each AP-bin (centered at position x), pooling gives a set $\mathcal{A}(x)$ of activity time series $A \in \mathcal{A}(x)$ for all nuclei in that particular AP bin. Our Bayesian deconvolution approach enables sampling from the posterior distribution $P(r|A)$. It generates MCMC samples of the instantaneous single-allele transcription rate given an activity time series A (see Eq. 19). We denote such a sample $r_s(t)|A$, where s labels a specific MCMC sample drawn from $P(r|A)$. Thus, for a given AP-bin (centered at position x), we compute the normalized auto-correlation function $\rho_s(\tau)$ as follows:

$$\rho_s(\tau) = \frac{\langle (r_s(t)|A - \mu_s(t))(r_s(t+\tau)|A - \mu_s(t+\tau)\rangle}{\sigma_s(t)\sigma_s(t+\tau)}, \tag{27}$$

where τ is the lag (which is a multiple of Δt) and the averaging is performed over all the nuclei in the AP-bin, i.e., $\langle F(r_s(t)|A)\rangle = \frac{1}{N_A(x)}\sum_{A\in\mathcal{A}(x)} F(r_s(t)|A)$ with $N_A(x) = \#\mathcal{A}(x,t) \sim 200$ nuclei. Hence, $\mu_s(t) = \langle r_s(t)|A\rangle$ is the time-dependent mean and $\sigma_s(t) = \sqrt{\langle (r_s(t)|A - \mu(t))^2\rangle}$ the time-dependent standard deviation in the AP-bin. The final estimate of the normalized auto-correlation $\rho(\tau)$ and the corresponding error $\sigma_\rho(\tau)$ are obtained by averaging over MCMC samples:

$$\begin{aligned}\rho(\tau) &= \frac{1}{N_s}\sum_{s=1}^{N_s}\rho_s(\tau)\\ \sigma_\rho^2(\tau) &= \frac{1}{N_s}\sum_{s=1}^{N_s}\left(\rho_s(\tau) - \rho(\tau)\right)^2,\end{aligned} \tag{28}$$

where $N_s = 10^3$ is the total number of samples used per allele.

We show auto-correlation functions $\rho(\tau)$ for *hb* in early NC14 ($7.5 \le t < 20.5$ min after mitosis) in Fig. 1f (color code stands for the AP-bin) and for other genes in Extended Data Fig. 5c. An auto-correlation is characterized by two key quantities: the magnitude of correlated fluctuations $\Sigma_{\mathrm{AC}} := \rho(\Delta t)$

(where $\Delta t = 10\,\mathrm{s}$ is the sampling time) and the correlation time τ_{AC} that captures the temporal scale of the fluctuations. We estimated Σ_{AC} and τ_{AC} for all gap genes at all AP positions in NC13 and early NC14 (Extended Data Fig. 5d and Fig. 1g). The correlation time τ_{AC} was estimated by fitting $\rho(\tau)$ with the following exponential function for $\Delta t \leq \tau < 10\,\mathrm{min}$:

$$\tilde{\rho}(\tau) = \Sigma_{\mathrm{AC}}(1-\beta)\exp\left(-\frac{\tau-\Delta t}{\tau_{\mathrm{AC}}}\right)+\beta, \tag{29}$$

where Σ_{AC} is estimated from $\rho(\Delta t)$ and $\{\tau_{\mathrm{AC}}, \beta\}$ are obtained by maxmimum likelihood. The parameter β accommodates for a possible noise floor in $\rho(\tau)$, which typically results from residual embryo-to-embryo variability; on average $\beta = 0.04$. Errors on Σ_{AC} and τ_{AC} are computed from bootstrapping the fit to $\rho_s(\tau)$ (Eq. 27) over MCMC samples.

7.3.4 Two-state model auto-correlation

To help us interpret the auto-correlation function, we investigated the prediction of the 2-state model of transcriptional bursting [12, 17]. In the 2-state model, the gene toggles stochastically between an ON ($n = 1$) and OFF ($n = 0$) states with rates k_{ON} and k_{OFF}, respectively. In the ON state, the gene promoter can initiate transcription with initiation rate k leading to the synthesis of g transcripts. Within that model, we further define the switching correlation time $T = 1/(k_{\mathrm{ON}} + k_{\mathrm{OFF}})$ and steady-state mean occupancy $\eta = k_{\mathrm{ON}}/(k_{\mathrm{ON}} + k_{\mathrm{OFF}})$. In what follows, we focus on the stationary case (k_{ON}, k_{OFF}, and k don't change over time), which is easier to interpret. From the master equation [18, 19], we derive the equations that describe the time evolution of the moments:

$$\begin{cases} \dfrac{d}{dt}\langle n(t)\rangle = \dfrac{1}{T}(\eta - \langle n(t)\rangle) \\ \dfrac{d}{dt}\langle g(t)\rangle = k\langle n(t)\rangle \\ \dfrac{d}{dt}\langle g(1,t)\rangle = k\langle n(t)\rangle + \dfrac{1}{T}(\eta\langle g(t)\rangle - \langle g(1,t)\rangle) \\ \dfrac{d}{dt}\langle g^2(t)\rangle = k(\langle n(t)\rangle + 2\langle g(1,t)\rangle) \end{cases} \tag{30}$$

$\langle n(t)\rangle$ is the mean occupancy of the ON-state (or equivalently the ON-probability P_{ON}), $\langle g(t)\rangle$ is the mean number of initiation events (or transcripts), $\langle g(1,t)\rangle = \langle g(t)\rangle - \langle g(0,t)\rangle$ is the mean number of initiation events conditioned on the gene being ON, and $\langle g^2(t)\rangle$ is the second moment of g. Solving the equations above (Eq. 30) with initial conditions $\langle n(t=0)\rangle = n_0$, $\langle g(t=0)\rangle = \langle g(1,t=0)\rangle = \langle g^2(t=0)\rangle = 0$, we determine the mean $\mu(t) \equiv \langle g(t)\rangle$ and the variance $\sigma^2(t) = \langle g^2(t)\rangle - \langle g(t)\rangle$ of the number of transcription initiation events g over the period $[0,t]$. In the special case $t = \tau_{\mathrm{elo}}$ (the elongation time), $\mu(\tau_{\mathrm{elo}})$ and $\sigma^2(\tau_{\mathrm{elo}})$ correspond to the mean and the variance of the number of nascent transcripts on the gene (i.e., the transcriptional activity), under the assumption that elongation is a deterministic process at constant rate K_{elo}.

Assuming the gene is initially at steady state ($n_0 = \eta$), we solve Eq. 30 and get:

$$\begin{aligned} \langle g(t)\rangle &= k\eta t \\ \langle g(1,t)\rangle &= k\eta^2 t + k\eta(1-\eta)T(1-\exp(-t/T)) \\ \langle g(0,t)\rangle &= k\eta(1-\eta)t - k\eta(1-\eta)T(1-\exp(-t/T)) \\ \langle g^2(t)\rangle &= k\eta t + k^2\eta^2 t^2 + 2k^2\eta(1-\eta)(T^2\exp(-t/T) + tT - T^2). \end{aligned} \tag{31}$$

To compute the auto-correlation, we further need the transient of the first moments obtained by solving Eq. 30 with generic initial condition $\langle n(t=0)\rangle = n_0 \in \{0,1\}$:

$$\begin{aligned}\langle n(t|n_0)\rangle &= \eta(1-\exp(-t/T)) + n_0\exp(-t/T)\\ \langle g(t|n_0)\rangle &= k\eta t + k(n_0-\eta)T(1-\exp(-t/T)).\end{aligned} \tag{32}$$

The transcriptional activity at time t is given by $g(t;t-\tau_{\text{elo}})$, which is the total number of nascent transcripts (on the gene) initiated over the period $[t,t-\tau_{\text{elo}}]$. In the stationary case, the transcriptional activity is simply given by $\langle g(t;t-\tau_{\text{elo}})\rangle = \langle g(\tau_{\text{elo}})\rangle$. Using this definition, we then define the normalized auto-correlation of the transcriptional activity as

$$\begin{aligned}\rho_A(\tau) &= \frac{\langle (g(t;t-\tau_{\text{elo}})-\mu(\tau_{\text{elo}}))(g(t+\tau;t+\tau-\tau_{\text{elo}})-\mu(\tau_{\text{elo}}))\rangle}{\sigma^2(\tau_{\text{elo}})}\\ &= \frac{\langle g(t+\tau;t+\tau-\tau_{\text{elo}})g(t;t-\tau_{\text{elo}})\rangle - \mu^2(\tau_{\text{elo}})}{\sigma^2(\tau_{\text{elo}})}.\end{aligned} \tag{33}$$

And similarly for the normalized auto-correlation of the transcription rate $r(t) = g(t;t-\Delta t)/\Delta t$ (where $\Delta t = 10$s is the sampling time), we have:

$$\begin{aligned}\rho_R(\tau) &= \frac{\langle (g(t;t-\Delta t)-\mu(\Delta t))(g(t+\tau;t+\tau-\Delta t)-\mu(\Delta t))\rangle/(\Delta t)^2}{\sigma^2(\Delta t)/(\Delta t)^2}\\ &= \frac{\langle g(t+\tau;t+\tau-\Delta t)g(t;t-\Delta t)\rangle - \mu^2(\Delta t)}{\sigma^2(\Delta t)}.\end{aligned} \tag{34}$$

As it turns out, these auto-correlation functions are almost identical and only differ in the integration period $\tilde{\tau}$, which is set to $\tilde{\tau} = \tau_{\text{elo}}$ to compute $\rho_A(\tau)$ and to $\tilde{\tau} = \Delta t$ to compute $\rho_R(\tau)$. In addition, the mean $\mu(t) \equiv \langle g(t)\rangle$ and the variance $\sigma^2(t) = \langle g^2(t)\rangle - \langle g(t)\rangle$ are fully determined by the solutions in Eq. 31. However, we still need to calculate the term $\langle g(t+\tau;t+\tau-\tilde{\tau})g(t;t-\tilde{\tau})\rangle$. This term is easily calculated by splitting the calculation into two separate cases:

1. When $g(t+\tau;t+\tau-\tilde{\tau})$ and $g(t;t-\tilde{\tau})$ share the same underlying trajectory, i.e., when $0 \le \tau < \tilde{\tau}$.
2. When $g(t+\tau;t+\tau-\tilde{\tau})$ and $g(t;t-\tilde{\tau})$ are disjoint, i.e., when $\tau \ge \tilde{\tau}$.

We can then take advantage of the fact that $g(t_1;t_3) = g(t_1;t_2) + g(t_2;t_3)$, provided we preserve the underlying connection of the gene state n at t_2.

Case 1: $0 \le \tau < \tilde{\tau}$. Introducing $t_1 := t+\tau$, $t_2 := t$, $t_3 := t-\tilde{\tau}+\tau$ and $t_4 := t-\tilde{\tau}$ such that $t_1 \ge t_2 \ge t_3 \ge t_4$, we get:

$$\begin{aligned}\langle g(t+\tau;t-\tilde{\tau}+\tau)g(t;t-\tilde{\tau})\rangle &= \langle (g(t_1;t_2)+g(t_2;t_3))(g(t_2;t_3)+g(t_3;t_4))\rangle\\ &= \langle g(t_1-t_2)\rangle_{n|n}\langle g(t_2-t_3)\rangle + \langle g(t_1-t_2)\rangle_{n|n'}\langle g(t_3-t_4)\rangle\\ &\quad + \langle g(t_2-t_3)\rangle_{n|n}\langle g(t_3-t_4)\rangle + \langle g^2(t_2-t_3)\rangle,\end{aligned} \tag{35}$$

where the symbol $_{n|n}$ means that the products are "connected" through the temporal correlation in n.

These products are computed as follows:

$$\langle g(t_1-t_2)\rangle_{n|n}\langle g(t_2-t_3)\rangle = \sum_{n=0}^{1}\langle g(\tau|n)\rangle\langle g(n,\tilde{\tau}-\tau)\rangle$$

$$\langle g(t_1-t_2)\rangle_{n|n'}\langle g(t_3-t_4)\rangle = \sum_{n,n'=0}^{1}\langle g(\tau|n)\rangle P(n|n';\tilde{\tau}-\tau)\langle g(n',\tau)\rangle$$

$$\langle g(t_2-t_3)\rangle_{n|n}\langle g(t_3-t_4)\rangle = \sum_{n=0}^{1}\langle g(\tilde{\tau}-\tau|n)\rangle\langle g(n,\tau)\rangle,$$

where both $\langle g(t|n)\rangle$ and $\langle g(n,t)\rangle$ have been calculated above (Eq. 31 and 32), and $P(n|n';t)$ is the propagator (i.e., the transition probabilities) of the telegraph process:

$$P(n|n';t) = (n\eta + (1-n)(1-\eta))(1-\exp(-t/T)) + \delta_{nn'}\exp(-t/T), \tag{36}$$

where $\delta_{nn'}$ is the Kronecker delta. We thus have determined all the terms in Eq. 35 and we can finally obtain:

$$\begin{aligned}\langle g(t+\tau;t-\tilde{\tau}+\tau)g(t;t-\tilde{\tau})\rangle = {} & k\eta(\tilde{\tau}-\tau) + k^2\eta^2\tilde{\tau}^2 + 2k^2\eta(1-\eta)T(\tilde{\tau}-\tau) \\ & + k^2\eta(1-\eta)T^2\left(\exp(-(\tilde{\tau}+\tau)/T) + \exp(-(\tilde{\tau}-\tau)/T) - 2\exp(-\tau/T)\right).\end{aligned} \tag{37}$$

When $\tau = 0$, the expression above reduces to $\langle g^2(\tilde{\tau})\rangle$, consistent with the fact that all the terms in Eq. 35 vanish except $\langle g^2(t_2-t_3)\rangle = \langle g^2(\tilde{\tau})\rangle$. Since, $\langle g^2(\tilde{\tau})\rangle - \langle g(\tilde{\tau})\rangle^2 = \sigma^2(\tilde{\tau})$, we thus properly recover $\rho(0) = 1$ as expected.

Case 2: $\tau \geq \tilde{\tau}$. Introducing $t_1 := t+\tau$, $t_2 := t+\tau-\tilde{\tau}$, $t_3 := t$ and $t_4 := t-\tilde{\tau}$ such that $t_1 \geq t_2 \geq t_3 \geq t_4$, we get:

$$\begin{aligned}\langle g(t+\tau;t-\tilde{\tau}+\tau)g(t;t-\tilde{\tau})\rangle &= \langle (g(t_1;t_2)g(t_3;t_4)\rangle \\ &= \langle g(t_1-t_2)\rangle_{n|n'}\langle g(t_3-t_4)\rangle,\end{aligned}$$

where the connected product is given by

$$\langle g(t_1-t_2)\rangle_{n|n'}\langle g(t_3-t_4)\rangle = \sum_{n,n'=0}^{1}\langle g(\tilde{\tau}|n)\rangle P(n|n';\tau-\tilde{\tau})\langle g(n',\tilde{\tau})\rangle.$$

Again, using the expression for $\langle g(t|n)\rangle$ and $\langle g(n,t)\rangle$ (Eq. 31 and 32), and for $P(n|n';t)$ (Eq. 36), we finally get:

$$\langle g(t+\tau;t-\tilde{\tau}+\tau)g(t;t-\tilde{\tau})\rangle = k^2\eta^2\tilde{\tau}^2 + k^2\eta(1-\eta)T^2(1-\exp(-\tilde{\tau}/T))^2\exp(-\tau/T) \tag{38}$$

Conclusion. We have calculated the normalized auto-correlation functions $\rho_A(\tau)$ (Eq. 33, $\tilde{\tau} = \tau_{\text{elo}}$) and $\rho_R(\tau)$ (Eq. 34, $\tilde{\tau} = \Delta t$) using the two-state model. Indeed, using Eq. 31, 37 and 38, we found

$$\rho(\tau) = \begin{cases} \dfrac{\tilde{\tau}-\tau+k(1-\eta)\left(2(\tilde{\tau}-\tau)T + T^2\left(\exp(-\frac{\tilde{\tau}+\tau}{T}) + \exp(-\frac{\tilde{\tau}-\tau}{T}) - 2\exp(-\frac{\tau}{T})\right)\right)}{\tilde{\tau}+2k(1-\eta)\left(T^2\exp(-\frac{\tilde{\tau}}{T}) + \tilde{\tau}T - T^2\right)} & \text{if} \quad \tau < \tilde{\tau} \\ \dfrac{k(1-\eta)T^2\left(1-\exp(-\frac{\tilde{\tau}}{T})\right)^2\exp(-\frac{\tau}{T})}{\tilde{\tau}+2k(1-\eta)\left(T^2\exp(-\frac{\tilde{\tau}}{T}) + \tilde{\tau}T - T^2\right)} & \text{if} \quad \tau \geq \tilde{\tau} \end{cases} \tag{39}$$

Using the expression above, we plot $\rho_A(\tau)$ and $\rho_R(\tau)$ for different values of η in Extended Data Fig. 5a (left and right respectively). For $\rho_A(\tau)$, we clearly see two regimes: one that is dominated by the elongation process $\tau < \tau_{\text{elo}}$ and one that is only dictated by the switching process $\tau \geq \tau_{\text{elo}}$. As $\rho_R(\tau)$ is unencumbered by the elongation process, only the exponential decay due to promoter switching remains. Importantly, these auto-correlation functions are identical for two independent gene copies (2 sister chromatids), since the correlated contributions and the variances add (in the independent case).

From Eq. 39, the auto-correlation of the single-allele transcription rate $\rho_R(\tau)$ (with $\tilde{\tau} = \Delta t$) can be further simplified. Indeed, since for real data τ becomes a multiple of the sampling time Δt, we simply get

$$\rho_R(\tau) = \begin{cases} 1 & \text{if} \quad \tau = 0 \\ \Sigma_{\text{AC}} \exp\left(-\dfrac{\tau - \Delta t}{T}\right) & \text{if} \quad \tau \geq \Delta t \end{cases} \tag{40}$$

Here the amplitude of the correlated variability Σ_{AC} is given by

$$\Sigma_{\text{AC}} = \frac{k\Delta t(1-\eta)\phi_2(\Delta t/T)}{1 + k\Delta t(1-\eta)\phi_1(\Delta t/T)}, \tag{41}$$

with filtering functions $\phi_1(x) \in [0,1]$ and $\phi_2(x) \in [0,1]$ given by

$$\phi_1(x) = 2\frac{\exp(-x) + x - 1}{x^2}$$
$$\phi_2(x) = \frac{(1 - \exp(-x))^2}{x^2}.$$

From Eq. 40, we clearly see that in the 2-state model, the time scale of the exponential decay is solely dictated by the switching correlation time T. This suggests that the measured correlation time τ_{AC} results from bursts and is in fact related to T (Fig. 1f and 1g). Importantly, Σ_{AC} displays two opposite behaviors depending on whether the mean transcription rate $R = k\eta$ is varied through k or η (Eq. 41 and Extended Data Fig. 5b). Indeed, an increase in k leads to an increase in Σ_{AC}, while an increase in η leads to a decrease Σ_{AC}. Interestingly, the scaling of Σ_{AC} with R observed in data is very much in line with the latter scenario (Extended Data Fig. 5d), i.e. changes in R driven by the ON-probability η. Lastly, we also see that Σ_{AC} vanishes when $\Delta t \to 0$ whereas the exponential decay is lost when $\Delta t \gtrsim T$. This implies there is a range of suitable Δt that provide sufficient integration $\Delta t \gg 0$ but still remain sufficiently small $\Delta t < T$ to observe the correlation in the signal. Given, the range of measured $\Sigma_{\text{AC}} \sim 0.2$–$0.6$ and $\tau_{\text{AC}} \sim 1$–2min (see Fig. 1g and Extended Data Fig. 5d), it appears that our sampling time $\Delta t \approx 10\,\text{s}$ is adequate.

7.3.5 Validating auto-correlation on simulated data

To assess our ability to estimate the auto-correlation (AC) function properly, we performed single allele deconvolution on simulated data mimicking experimental conditions. Specifically, we compared the estimated AC parameters (Σ_{AC} and τ_{AC}) obtained through deconvolution to the known input parameters used to generate the simulated data. For each combination of input parameters, we generated 200 synthetic alleles and 50 min long activity time series using the Gillespie algorithm [20], including measurement noise consistent with real data and assuming two independent gene copies (i.e., sister chromatids) described by the 2-state model (See Section 7.4.5 for further details). Overall, we probed 56 combinations of (single gene copy) input parameters with $k \equiv K^{(1)} = 8$ mRNA/min (consistent with real data), $\eta \equiv P_{\text{ON}}^{(1)} \in \{0.03, 0.12, 0.23, 0.36, 0.52, 0.78, 0.90\}$ and $T \equiv T_{\text{C}}^{(1)} \in \{0.5, 1, 2, 3, 4, 5, 7, 10\}$ min.

For each combination of input parameters, we calculated the input Σ_{AC} according to Eq. 41, while the input τ_{AC} is simply given by the input T. Using the simulated data, we estimated Σ_{AC} and τ_{AC} from the instantaneous single-allele transcription rate, as described in Section 7.3.3. In Extended Data Fig. 5e, we show that for both Σ_{AC} and τ_{AC}, the input and the estimated AC parameters strongly correlate over a large range of $P_{\mathrm{ON}}^{(1)}$ and $T_{\mathrm{C}}^{(1)}$ values, demonstrating our ability to faithfully determine Σ_{AC} and τ_{AC} from deconvolved instantaneous single-allele transcription rates. Thus, the striking collapse of Σ_{AC} and τ_{AC} observed in real data (Fig. 1g and Extended Data Fig. 5d), across genes and activity, should not be markedly affected by our approach and appears to be a solid result.

Lastly, to assess the potential impact of errors in the elongation rate on the deconvolution process and the auto-correlation estimation, we compared the estimated AC parameters using the correct deconvolution kernel (used to generate the simulated data and computed here with $K_{\mathrm{elo}} = 2\,\mathrm{mRNA/min}$) and incorrect kernels (computed with either with $K_{\mathrm{elo}} = 1.5$ or $2.5\,\mathrm{mRNA/min}$, i.e., a $\pm 25\%$ error). We focused on the analysis of the $T_{\mathrm{C}}^{(1)} = 2\,\mathrm{min}$ simulated data set (that is closest to real data) and show the estimated τ_{AC} and Σ_{AC} using the three kernels for all values of input $P_{\mathrm{ON}}^{(1)}$ (in Extended Data Fig. 5f). The analysis determines that while the effect of a 25% deviation on K_{elo} has only a very limited impact on Σ_{AC} (resulting error on Σ_{AC} $\sim$4%), the estimation of τ_{AC} is almost entirely unaffected by a 25% overestimation of K_{elo} (error on τ_{AC} $\sim$7%). However, an underestimation of K_{elo} leads to a significant underestimation of τ_{AC}, albeit with errors on τ_{AC} not exceeding 25% on average. Overall, this demonstrates the robustness of our approach to changes in the elongation rate.

7.4 Burst calling and transcriptional parameter estimation

7.4.1 Calling bursts from deconvolved time series

To call bursts from individual single-allele transcriptional activity time series A, we opt for a simple approach with minimal mechanistic assumptions. We simply aim to cluster the deconvolved configurations of initiation events I underlying A over time (see Eq. 14 and Fig. 2a). We first compute the single-allele transcription rate $r(t)$ from I (see Eq. 18). We then perform a (centered) temporal moving average of $r(t)$ using the following kernel:

$$\kappa(t) = \frac{1}{Z(w)} \exp\left(-\left(\frac{2t}{w}\right)^4\right), \tag{42}$$

where w is the width of the moving window and $Z(w) = \sum_{i=1}^{N_t} \kappa(t_i)$ is the normalization. We compute the moving average through discrete convolution resulting in the following rate $r_w(t) = r(t) * \kappa(t)$. Subsequently, a threshold $r_b = g_b/w$ is applied to the resulting rate estimate r_w, where g_b corresponds to the desired minimum number of initiations events for a burst. This threshold allows us to determine the cluster of initiation events (bursts) defining the allele's ON-state $n(t) \in \{0, 1\}$:

$$n(t) = \begin{cases} 1 & \text{if} \quad r_w(t) \geq r_b \\ 0 & \text{if} \quad r_w(t) < r_b \end{cases} \tag{43}$$

It follows that the clustering approach described above and the ensuing definition of the ON state $n(t)$ depend on two free parameters: the width of the averaging window w and the minimal number of initiation events g_b. As a rule of thumb, w needs to be large enough to average the Poisson initiation fluctuations and prevent the calling of spurious bursts, but small enough to preserve the temporal correlations attributed to ON-OFF switching in the data. In practice, we set $w = 5\Delta t \approx 50\,\mathrm{s}$, which is slightly below the measured correlation time $\tau_{\mathrm{AC}} \sim$ 1–2 min (Fig. 1g). Regarding the threshold, we set $g_b = 2$ corresponding to a

detection of at least 2 initiation events within the moving window, which is motivated by our detection sensitivity of around 1–2 mRNAs (see Extended Data Fig. 1f. and Section 6.4). Notably, the shape of the kernel (Eq. 42) and the chosen parameter values ($w = 5\Delta t$ and $g_b = 2$) set a lower bound on the minimal ON interval duration that can be reliably detected, which is $3\Delta t \approx 30\,\mathrm{s}$. This value corresponds to the duration of the ON interval calculated from $n(t)$ with input $I(t) = g_b\delta(t)$. To verify the impact of our choice of clustering parameters on the estimation of bursting parameters go to Section 7.4.5 Validation using synthetic data.

Importantly, our estimation of the ON state n through clustering can be applied to each initiation configuration I sampled from the posterior distribution $P(I|A)$ (Eq. 14). It implies that we can easily reconstruct $P(n|A)$ for each allele, as we did for the instantaneous distribution of the single-allele transcription rate $P(r|A)$ (see Section 7.1.5). Using our MCMC samples we get

$$P(n|A) = \sum_I P(n|I)P(I|A), \tag{44}$$

where $P(n|I) = 1$ when Eq. 43 is satisfied and $P(n|I) = 0$ otherwise. The distribution $P(n|A)$ allows us to assess the uncertainty on burst calling for each time series of transcriptional activity A.

7.4.2 Estimating transcriptional parameters

To characterize the bursting dynamics of each gap gene in space and time, we define a set of bursting parameters and estimate them empirically from the single-allele transcription rates $r(t)$ and the ON-state $n(t)$ (see Fig. 2b,c). For a given single-allele transcriptional activity time series A, we sampled individual initiation configurations I_s from the posterior distribution $P(I|A)$ using MCMC (see Section 7.1). The index $s \in \{1, 2, \ldots, N_s\}$ denotes one specific MCMC sample drawn from the posterior, where $N_s = 10^3$ is the total number of samples used per allele. For each I_s, we further computed $r_s(t)|A$ (Eq. 18) and $n_s(t)|A$ (Eq. 43), where the index s and the notation $|A$ specify that these come from the sample I_s given the time series A. We then defined the mean transcription rate R and the ON-probability P_{ON} over all nuclei in a specific spatiotemporal (x,t)-bin, that is over all the $A \in \mathcal{A}(x,t)$. For a single MCMC sample s per A we get

$$R_s(x,t) = \frac{1}{N_A(x,t)} \sum_{A\in\mathcal{A}(x,t)} r_s(t)|A \tag{45}$$

$$P_{\mathrm{ON},s}(x,t) = \frac{1}{N_A(x,t)} \sum_{A\in\mathcal{A}(x,t)} n_s(t)|A, \tag{46}$$

where $N_A(x,t) = \#\mathcal{A}(x,t)$. We then compute a single point estimate per bursting parameter Θ for each position x and time t by averaging over our MCMC samples as follows

$$\Theta(x,t) := \langle \Theta_s(x,t) \rangle = \frac{1}{N_s} \sum_{s=1}^{N_s} \Theta_s(x,t) \tag{47}$$

$$\sigma_\Theta(x,t) = \sqrt{\langle (\Theta_s(x,t) - \Theta(x,t))^2 \rangle}, \tag{48}$$

where σ_Θ is our estimation of the error on Θ. Thus, the mean transcription rate and the ON-probability are given by $R(x,t) = \langle R_s(x,t) \rangle$ and $P_{\mathrm{ON}}(x,t) = \langle P_{\mathrm{ON},s}(x,t) \rangle$ (see Fig. 2d,f). Importantly, this construction of $R(x,t)$ is equivalent to our previous calculation using the mixture of posterior distributions (see Eq. 22).

The mean initiation rate K is defined over all nuclei in a spatiotemporal (x,t)-bin by averaging the single-allele transcription rate $r(t)$ conditioned on the allele being ON ($n(t)=1$). It is expressed as

$$\begin{aligned} K_s(x,t) &= \frac{1}{Z_K} \sum_{A \in \mathcal{A}(x,t)} r_s(t)|A \cdot n_s(t)|A \\ Z_K &= \sum_{A \in \mathcal{A}(x,t)} n_s(t)|A, \end{aligned} \tag{49}$$

where Z_K counts the number of ON alleles at time t. Averaging over MCMC samples (Eq. 47) leads to the the mean initiation rate $K(x,t) = \langle K_s(x,t) \rangle$ (see Fig. 2e). By construction $R = K \cdot P_{\mathrm{ON}}$, which we verify empirically (see Extended Data Fig. 6a). As additional control, we define a "leaking" rate K_L by averaging the single-allele transcription rate $r(t)$ conditioned on the allele being OFF ($n(t)=0$):

$$\begin{aligned} K_{L,s}(x,t) &= \frac{1}{Z_L} \sum_{A \in \mathcal{A}(x,t)} r_s(t)|A \cdot (1 - n_s(t)|A) \\ Z_L &= \sum_{A \in \mathcal{A}(x,t)} 1 - n_s(t)|A. \end{aligned} \tag{50}$$

We verified that $K_L(x,t) = \langle K_{L,s}(x,t) \rangle$ is indeed negligible compared to $K(x,t)$, supporting our ability to identify well-demarcated bursts (Extended Data Fig. 6b,c).

To calculate the mean durations of the ON and OFF periods (T_{ON} and T_{OFF}) for each spatiotemporal (x,t)-bin, we first determine the boundaries of the ON ($n(t)=1$) and OFF ($n(t)=0$) periods for individual alleles. Using $n(t)$, we built a function $b(t)$ such that $b(t)=1$ whenever time point t corresponds to an ON-OFF boundary:

$$b(t) = \begin{cases} 1 & \text{if} \quad |n(t+\Delta t) - n(t)| > 0 \\ 0 & \text{if} \quad |n(t+\Delta t) - n(t)| = 0 \end{cases} \tag{51}$$

where Δt is the sampling time in our measurements. $b(t)$ defines ON-OFF boundaries that are consistent with our definition of the instantaneous single-allele transcription rate $r(t)$ (see Eq. 18), i.e., we associate the state $n(t)$ and the rate $r(t)$ to the same time interval $(t-\Delta t, t]$. For convenience, we also enforce $b(t_1)=1$ and $b(t_{N_t})=1$ for the first and last time point t_1 and t_{N_t} of the time series.

From $b(t)$ we construct the set of boundary time points, i.e, $\mathcal{T} = \{t : b(t)=1\} = \{t_{b_1}, \ldots, t_{b_k}\}$ such that $b_j < b_{j+1}\ \forall j \in \{1,2,\ldots,k-1\}$ where $k \geq 2$ (in each case $t_{b_1} = t_1$ and $t_{b_k} = t_{N_t}$). To measure the durations of individual ON and OFF periods, we introduce a function $\tau(t)$ such that

$$\tau(t) = \begin{cases} t_{b_2} - t_{b_1} & \text{if} \quad t = t_1 \\ t_{b_{j+1}} - t_{b_j} & \text{else if} \quad t \in (t_{b_j}, t_{b_{j+1}}] \end{cases} \tag{52}$$

We further introduce a function $\pi(t) = \tau(t)/\Delta t$ that counts the number of Δt intervals underlying $\tau(t)$. We calculate T_{ON} from the activity time series A as

$$\begin{aligned} T_{\mathrm{ON},s}(x,t) &= \frac{1}{Z_{\mathrm{ON}}} \sum_{A \in \mathcal{A}(x,t)} \tau_s(t)|A \cdot \frac{n_s(t)|A}{\pi_s(t)|A} \\ Z_{\mathrm{ON}} &= \sum_{A \in \mathcal{A}(x,t)} \frac{n_s(t)|A}{\pi_s(t)|A}, \end{aligned} \tag{53}$$

where Z_{ON} counts the number of ON alleles at time t weighted by the inverse of the length of the underlying ON periods (s denotes an individual MCMC sample as before). The rationale behind the

$\frac{n_s(t)|A}{\pi_s(t)|A}/Z_{\mathrm{ON}}$ weights is as follows: as we aim to estimate mean durations across nuclei at time t (Extended Data Fig. 6k), we need to account for the length of each period $\tau(t)$ and correct for their overall contribution to $\pi(t)$ time points (e.g. longer periods contribute to more time points, thus we need weights that balance over-counting).

Finally, by averaging over MCMC samples (Eq. 47), we obtain the mean ON duration $T_{\mathrm{ON}}(x,t) = \langle T_{\mathrm{ON},s}(x,t)\rangle$ (see Fig. 3a). Likewise, for T_{OFF}, we have

$$\begin{aligned} T_{\mathrm{OFF},s}(x,t) &= \frac{1}{Z_{\mathrm{OFF}}} \sum_{A\in\mathcal{A}(x,t)} \tau_s(t)|A \cdot \frac{1-n_s(t)|A}{\pi_s(t)|A} \\ Z_{\mathrm{OFF}} &= \sum_{A\in\mathcal{A}(x,t)} \frac{1-n_s(t)|A}{\pi_s(t)|A}. \end{aligned} \tag{54}$$

Again, averaging over MCMC samples (Eq. 47) leads to the mean OFF duration $T_{\mathrm{OFF}}(x,t) = \langle T_{\mathrm{OFF},s}(x,t)\rangle$ (see Fig. 3a).

When estimating the initiation rate K (Eq. 49), and the mean durations of the ON and OFF periods T_{ON} (Eq. 53) and T_{OFF} (Eq. 54), we sometimes encounter missing values, e.g. absence of ON or OFF periods over all nuclei at a given time point t. This typically occurs when $P_{\mathrm{ON}}(t) \to 0$ or $P_{\mathrm{ON}}(t) \to 1$. Therefore, we filter each bursting parameter estimate $\Theta_s(x,t)$ using a Gaussian kernel over a short timescale of 1 min. This "smoothing" interpolates the eventual missing values and prevents gaps in our time-dependent parameter estimation.

Using K, T_{ON} and T_{OFF}, we can further define the following bursting parameters:

$$T_C := \frac{T_{\mathrm{ON}}T_{\mathrm{OFF}}}{T_{\mathrm{ON}}+T_{\mathrm{OFF}}} \tag{55}$$

$$B := K \cdot T_{\mathrm{ON}} \tag{56}$$

$$F := \frac{1}{T_{\mathrm{ON}}+T_{\mathrm{OFF}}}, \tag{57}$$

where T_C is the switching correlation time, B the burst size and F the burst frequency. These parameters were directly computed from our estimates of K, T_{ON}, and T_{OFF} as defined above. Importantly, we tested extensively our ability to estimate static and time-dependent bursting parameters correctly using simulated data (see Section 7.4.5; Validation using synthetic data).

7.4.3 Sister chromatids and single copy parameters

In the blastoderm phase of the *Drosophila* embryo, DNA replication occurs during the first few minutes of each nuclear cycle (2–3 min after mitosis). Therefore, as our microscopy does not allow for separating individual sister chromatids, imaged spots decorated with fluorescently labeled nascent transcripts typically correspond to two optically unresolved sites of transcription. Since these sites result from replication of the same allele, the underlying gene copies must be *identical*. However, the activity of the two copies could be correlated. As recent observations favor independence or weak correlation [21, 12, 22], we first transform our effective bursting parameters into single-gene-copy (SGC) parameters assuming *independent* gene copies. In a subsequent step, we extend our parameter transformation to an arbitrary degree of correlation between gene copies, allowing us to test the independence assumption and to confront it with our data.

Two independent gene copies. In that case, the mean SGC transcription rate $R^{(1)}$ is simply given by $R^{(1)} = R/2$, which is half the effective one. In addition, the SGC ON-probability $P_{\rm ON}^{(1)}$ must satisfy

$$\underbrace{(1-P_{\rm ON}^{(1)})^2}_{P(0 \text{ copy active})} + \underbrace{2P_{\rm ON}^{(1)}(1-P_{\rm ON}^{(1)}) + {P_{\rm ON}^{(1)}}^2}_{P(1\geq \text{ copy active})} = 1,$$

where the probability of observing at least one active gene copy corresponds to the effective $P_{\rm ON}$. Thus, $P_{\rm ON}^{(1)}(2-P_{\rm ON}^{(1)}) = 1-(1-P_{\rm ON}^{(1)})^2 = P_{\rm ON}$ and we obtain the SGC ON-probability given by

$$P_{\rm ON}^{(1)} = 1-(1-P_{\rm ON})^{1/2}. \tag{58}$$

Regarding the SGC mean initiation rate $K^{(1)}$, it must satisfy $K^{(1)} = R^{(1)}/P_{\rm ON}^{(1)} = KP_{\rm ON}/(2P_{\rm ON}^{(1)})$. Thus, using Eq. 58 we get

$$K^{(1)} = \frac{K}{2}\left(1+(1-P_{\rm ON})^{1/2}\right). \tag{59}$$

For a single gene copy, the switching correlation time $T_{\rm C}^{(1)}$ is exactly given by $T_{\rm C}^{(1)} = 1/(k_{\rm ON}+k_{\rm OFF})$, where $k_{\rm ON}$ and $k_{\rm OFF}$ are the ON and OFF rates of the 2-state model (see Section 7.3.4). For a system made of two identical and independent gene copies we can write the following kinetic equation:

$$\lambda_0 \underset{k_{\rm OFF}}{\overset{2k_{\rm ON}}{\rightleftharpoons}} \lambda_1 \underset{2k_{\rm OFF}}{\overset{k_{\rm ON}}{\rightleftharpoons}} \lambda_2, \tag{60}$$

where λ_0, λ_1, and λ_2 are states corresponding to zero, one, and two active gene copies, respectively. Again, the effective OFF-state is equivalent to λ_0, while the effective ON-state is either λ_1 or λ_2. Further assuming steady-state, from Eq. 60 we calculate the mean residence time in the OFF and ON states (using phase-type distributions), i.e., $T_{\rm OFF}$ and $T_{\rm ON}$ as a function of $k_{\rm ON}$ and $k_{\rm OFF}$:

$$\begin{aligned} T_{\rm OFF} &= \frac{1}{2k_{\rm ON}} \\ T_{\rm ON} &= \frac{k_{\rm ON}+2k_{\rm OFF}}{2k_{\rm OFF}^2}. \end{aligned} \tag{61}$$

One can easily express the SGC switching rates as a function of $T_{\rm OFF}$ and $T_{\rm ON}$:

$$\begin{aligned} k_{\rm ON} &= \frac{1}{2T_{\rm OFF}} \\ k_{\rm OFF} &= \frac{T_{\rm OFF}+(T_{\rm OFF}^2+T_{\rm OFF}T_{\rm ON})^{1/2}}{2T_{\rm OFF}T_{\rm ON}}. \end{aligned} \tag{62}$$

From our definition of the effective switching correlation time $T_{\rm C}$ (Eq. 55), and the steady state relationships $k_{\rm ON} = P_{\rm ON}^{(1)}/T_{\rm C}^{(1)}$ and $k_{\rm OFF} = (1-P_{\rm ON}^{(1)})/T_{\rm C}^{(1)}$, we get

$$T_{\rm C} := \frac{T_{\rm ON}T_{\rm OFF}}{T_{\rm ON}+T_{\rm OFF}} = \frac{k_{\rm ON}+2k_{\rm OFF}}{2(k_{\rm ON}+k_{\rm OFF})^2} = \frac{1}{2}T_{\rm C}^{(1)}(2-P_{\rm ON}^{(1)}).$$

Using Eq. 58, we express the SGC switching correlation time as a function of $T_{\rm C}$ and $P_{\rm ON}$:

$$T_{\rm C}^{(1)} = \frac{2T_{\rm C}}{1+(1-P_{\rm ON})^{1/2}} \tag{63}$$

Together Eq. 58, 59, and 63 provide the key relationships to transform our effective bursting parameters to SGC parameters (see Supplementary Fig. 2e-h, Extended Data Figs. 6e-j and 8c). Note that the relationships between T_{OFF}, T_{ON}, and $T_{\mathrm{C}}^{(1)}$ (Eq. 61 and 63) are in principle valid only near steady-state. Nevertheless, these relationships provide good approximations in the non-stationary case given that the temporal changes of the underlying kinetics over developmental time are slow (see Section 7.4.5; Validation using synthetic data).

Two correlated gene copies. To address the SGC parameter transformation in the two correlated gene copies case, we introduce a new reaction scheme (Supplementary Fig. 3a), allowing continuous deformation through a control parameter $\eta \in [0, 1]$ the "independent gene copies" model into the "fully correlated gene copies" model. From the master equation of the system, we solve the steady state occupancy $P(\lambda_2)$ and $P(\lambda_1)$ with $P(\lambda_0) = 1-P(\lambda_1)-P(\lambda_2)$, where λ_0, λ_1, and λ_2 are states corresponding to zero, one, and two active gene copies, respectively. As function of the SGC switching rates k_{ON} and k_{OFF}, we find:

$$\begin{aligned} P(\lambda_2) &= \frac{\eta k_{\mathrm{OFF}} k_{\mathrm{ON}} + (2-\eta) k_{\mathrm{ON}}^2}{(2-\eta)(k_{\mathrm{OFF}} + k_{\mathrm{ON}})^2} \\ P(\lambda_1) &= \frac{4(1-\eta) k_{\mathrm{OFF}} k_{\mathrm{ON}}}{(2-\eta)(k_{\mathrm{OFF}} + k_{\mathrm{ON}})^2}. \end{aligned} \tag{64}$$

The effective ON-probability is then given by $P_{\mathrm{ON}} = P(\lambda_2) + P(\lambda_1)$. Using, the SGC steady state relationships $k_{\mathrm{ON}} = P_{\mathrm{ON}}^{(1)}/T_{\mathrm{C}}^{(1)}$ and $k_{\mathrm{OFF}} = (1 - P_{\mathrm{ON}}^{(1)})/T_{\mathrm{C}}^{(1)}$, we can express P_{ON} as function of the SGC ON-probability $P_{\mathrm{ON}}^{(1)}$:

$$P_{\mathrm{ON}} = \frac{\eta P_{\mathrm{ON}}^{(1)} + 2(1-\eta) P_{\mathrm{ON}}^{(1)} (2 - P_{\mathrm{ON}}^{(1)})}{2-\eta}. \tag{65}$$

Solving the equation above, we can express $P_{\mathrm{ON}}^{(1)}$ as function of P_{ON}, we get

$$P_{\mathrm{ON}}^{(1)} = \frac{2(2-\eta) P_{\mathrm{ON}}}{4(1-\eta) + \eta + \sqrt{(4(1-\eta)+\eta)^2 - 8(1-\eta)(2-\eta) P_{\mathrm{ON}}}}. \tag{66}$$

When $\eta = 0$ (uncorrelated), we correctly recover $P_{\mathrm{ON}}^{(1)} = 1-(1-P_{\mathrm{ON}})^{1/2}$ (see Equ. 58), while when $\eta = 1$ (fully correlated) we get $P_{\mathrm{ON}}^{(1)} = P_{\mathrm{ON}}$ as expected.

Next, we can calculate the (Pearson) correlation coefficient ρ between the two gene copies, we get

$$\rho = \frac{\langle n_1 n_2 \rangle - \langle n_1 \rangle \langle n_2 \rangle}{\sigma_n^2} = \frac{P(\lambda_2) - (P(\lambda_2) + \frac{1}{2} P(\lambda_1))^2}{(P(\lambda_2) + \frac{1}{2} P(\lambda_1))(1 - (P(\lambda_2) + \frac{1}{2} P(\lambda_1)))}, \tag{67}$$

where $n_1, n_2 \in [0, 1]$ are the individual transcriptional state of each gene copy (OFF and ON, respectively). Above, we use the following identities: $\langle n_1 n_2 \rangle = P(n_1 = 1, n_2 = 1) = P(2)$, $\langle n \rangle \equiv \langle n_1 \rangle = \langle n_2 \rangle = P(\lambda_2) + \frac{1}{2} P(\lambda_1)$, and lastly the binomial variance $\sigma_n^2 = \langle n \rangle (1 - \langle n \rangle)$. Using Eq. 64 and the SGC steady state relationships, the expression for ρ simplifies, and we get

$$\rho = \frac{\eta}{2-\eta}. \tag{68}$$

Thus, ρ only depends on η! We correctly have $\rho = 0$ when $\eta = 0$ and $\rho = 1$ when $\eta = 1$. For an intermediate degree of correlation ($0 < \rho < 1$), it is straightforward to solve the equation above, and we get $\eta = 2\rho(1+\rho)^{-1}$.

Next, we calculate the SGC initiation rate $K^{(1)}$. Since $2R^{(1)} = R = KP_{\mathrm{ON}}$, it must satisfy $K^{(1)} = R^{(1)}/P_{\mathrm{ON}}^{(1)} = KP_{\mathrm{ON}}/(2P_{\mathrm{ON}}^{(1)})$. Thus, using Eq. 66, we get

$$K^{(1)} = \frac{K}{2} \cdot \frac{4(1-\eta)+\eta+\sqrt{(4(1-\eta)+\eta)^2 - 8(1-\eta)(2-\eta)P_{\mathrm{ON}}}}{2(2-\eta)}. \tag{69}$$

Here again, we recover the correct limits: when $\eta = 0$ (uncorrelated) we get $K^{(1)} = \frac{K}{2}\left(1+(1-P_{\mathrm{ON}})^{1/2}\right)$ (see Equ. 59), while when $\eta = 1$ (fully correlated) we get $K^{(1)} = K/2$ as expected.

As in the independent case, and assuming at steady state, we calculate the mean residence time in the OFF and ON states (using phase-type distributions), i.e., T_{OFF} and T_{ON} as a function of k_{ON} and k_{OFF}:

$$\begin{aligned} T_{\mathrm{OFF}} &= \frac{1}{(2-\eta)k_{\mathrm{ON}}} \\ T_{\mathrm{ON}} &= \frac{(4(1-\eta)+\eta)k_{\mathrm{OFF}} + (2-\eta)k_{\mathrm{ON}}}{(2-\eta)^2 k_{\mathrm{OFF}}^2 + \eta(2-\eta)k_{\mathrm{OFF}}k_{\mathrm{ON}}}. \end{aligned} \tag{70}$$

Solving for k_{ON} and k_{OFF} and substituting $T_{\mathrm{OFF}} = T_{\mathrm{C}}/P_{\mathrm{ON}}$ and $T_{\mathrm{ON}} = T_{\mathrm{C}}/(1-P_{\mathrm{ON}})$, we find

$$\begin{aligned} k_{\mathrm{ON}} &= \frac{P_{\mathrm{ON}}}{(2-\eta)T_{\mathrm{C}}} \\ k_{\mathrm{OFF}} &= \frac{4(1-\eta)+\eta-2(2-\eta)P_{\mathrm{ON}}+\sqrt{(4(1-\eta)+\eta)^2 - 8(1-\eta)(2-\eta)P_{\mathrm{ON}}}}{2(2-\eta)^2 T_{\mathrm{C}}}. \end{aligned} \tag{71}$$

Using these equations, we can further calculate the SGC switching correlation time $T_{\mathrm{C}}^{(1)} = (k_{\mathrm{ON}}+k_{\mathrm{OFF}})^{-1}$:

$$T_{\mathrm{C}}^{(1)} = \frac{2(2-\eta)^2 T_{\mathrm{C}}}{4(1-\eta)+\eta+\sqrt{(4(1-\eta)+\eta)^2 - 8(1-\eta)(2-\eta)P_{\mathrm{ON}}}} \tag{72}$$

We see that the limits are correct: when $\eta = 0$ (uncorrelated) we get $T_{\mathrm{C}}^{(1)} = 2T_{\mathrm{C}}\left(1+(1-P_{\mathrm{ON}})^{1/2}\right)^{-1}$ (see Equ. 63), while when $\eta = 1$ (fully correlated) we get $T_{\mathrm{C}}^{(1)} = 2T_{\mathrm{C}}$ as expected.

Together Eq. 66, 69 and 72 provide general relationships to transform our effective bursting parameters to SGC parameters when sister chromatids are correlated ($\rho > 0$). Importantly, by comparing the parameters in the independent and fully correlated case, we see that the effect of correlation on the SGC parameters is very subtle. Indeed, correlation introduces a correction of at most a factor two to the SGC parameters ($K^{(1)}$, $T_{\mathrm{C}}^{(1)}$, and even less for $P_{\mathrm{ON}}^{(1)}$). However, these differences could still have a noticeable effect on the predicted single-cell transcription rate distribution and higher moments measured in the population. We have thus used the derived generalized relationships above to assess the correlation between sister chromatids, we find that sister chromatids are close to independent (see Section 7.4.6).

7.4.4 Predicting time-dependent bursting parameters

To test our deconvolution and burst calling approaches on simulated data (modeled via the master equation), we need to predict the time dependence of the effective bursting parameters (ground truth) given the underlying input kinetic parameters. Indeed, most bursting parameters (i.e., $P_{\mathrm{ON}}(t)$, $T_{\mathrm{ON}}(t)$, $T_{\mathrm{OFF}}(t)$ and $T_{\mathrm{C}}(t)$; but not $K(t)$) result from non-trivial time integration of the underlying gene switching rates $k_{\mathrm{OFF}}(t)$ and $k_{\mathrm{ON}}(t)$. In addition, the time scales $T_{\mathrm{ON}}(t)$, $T_{\mathrm{OFF}}(t)$ and $T_{\mathrm{C}}(t)$ when estimated from data are typically subjected to biases due to discreteness (e.g., missing events) and the finite nature (e.g.,

censoring) of the measured time intervals. To account for these effects, we compute the expected $P_{\mathrm{ON}}(t)$, $T_{\mathrm{ON}}(t)$, $T_{\mathrm{OFF}}(t)$ and $T_{\mathrm{C}}(t)$ given a 2-state model of transcription (for either a single allele or for each of the two independent alleles of the replicated sister chromatids) and the kinetic rates $k_{\mathrm{OFF}}(t)$ and $k_{\mathrm{ON}}(t)$ [12, 17].

Starting with the time evolution of the gene state n (with $n \in \{0, 1\}$ for a single gene copy and $n \in \{0, 1, 2\}$ in the two gene copies system), the non-stationary master equation of the system [18, 19] is given by

$$\frac{d}{dt}\mathbf{P}(t) = \hat{M}(t)\mathbf{P}(t), \tag{73}$$

where $\mathbf{P}(t)$ is the time-dependent vector whose components are the probabilities to find the system in any of the states n at time t. Thus, $\sum_n P_n(t) = 1$ with the sum taken over all components of $\mathbf{P}(t)$. $\hat{M}(t)$ is the time-dependent state transition matrix containing the propensity functions of the different reactions. For a single gene copy, $\hat{M}(t)$ is given by

$$\hat{M}_1(t) = \begin{pmatrix} -k_{\mathrm{ON}}(t) & k_{\mathrm{OFF}}(t) \\ k_{\mathrm{ON}}(t) & -k_{\mathrm{OFF}}(t) \end{pmatrix}. \tag{74}$$

For two independent and identical gene copies, $\hat{M}(t)$ is given by

$$\hat{M}_2(t) = \begin{pmatrix} -2k_{\mathrm{ON}}(t) & k_{\mathrm{OFF}}(t) & 0 \\ 2k_{\mathrm{ON}}(t) & -(k_{\mathrm{OFF}}(t) + k_{\mathrm{ON}}(t)) & 2k_{\mathrm{OFF}}(t) \\ 0 & k_{\mathrm{ON}}(t) & -2k_{\mathrm{OFF}}(t) \end{pmatrix}. \tag{75}$$

The solution of the non-stationary master equation for the system (Eq. 73) is given by:

$$\mathbf{P}(t) = \hat{U}(t, t_1)\mathbf{P}(t_1), \tag{76}$$

where $\mathbf{P}(t_1)$ is the initial condition vector, and the propagator $\hat{U}(t, t_1) \equiv \mathrm{OE}[\hat{M}](t, t_1)$ is the time-ordered exponential of $\hat{M}(t)$. Intuitively, $\hat{U}(t, t_1)$ is a matrix describing the time evolution of the system whose entries correspond to the transition probabilities, i.e., $\hat{U}_{nn'}(t, t_1) = P(n; t|n'; t_1)$, such that $P_n(t) = \sum_{n'} \hat{U}_{nn'}(t, t_1)P_{n'}(t_1)$. There are multiple ways to define the time-ordered exponential $\mathrm{OE}[\hat{M}](t, t_1)$; here we define it such that our computation becomes an infinite product of matrix exponential:

$$\mathrm{OE}[\hat{M}](t, t_1) = \lim_{N\to\infty} \left(\exp\left(\int_{t_{N-1}}^{t_N} \hat{M}(t')dt' \right) \ldots \exp\left(\int_{t_2}^{t_3} \hat{M}(t')dt' \right) \exp\left(\int_{t_1}^{t_2} \hat{M}(t')dt' \right) \right), \tag{77}$$

where $t_i = t_1 + (i - 1)\Delta t$ for $i \in \{1, \ldots, N\}$ and $\Delta t = (t - t_1)/(N - 1)$. Provided the temporal changes in $k_{\mathrm{OFF}}(t)$ and $k_{\mathrm{ON}}(t)$ are small over individual Δt, the product above provides a good approximation to the time-ordered exponential when N is finite. In practice, we set Δt to our sampling time ($\sim 10\,\mathrm{s}$) and we approximated the integral using the trapezoidal rule such that $\int_{t_{i-1}}^{t_i} \hat{M}(t)dt \approx \frac{1}{2}(\hat{M}(t_i) + \hat{M}(t_{i-1}))\Delta t$ leading to

$$\mathrm{OE}[\hat{M}](t_i, t_1) \approx \exp\left(\frac{1}{2}(\hat{M}(t_i) + \hat{M}(t_{i-1}))\Delta t \right) \ldots \exp\left(\frac{1}{2}(\hat{M}(t_2) + \hat{M}(t_1))\Delta t \right). \tag{78}$$

This approximation is quite convenient for discretized time intervals as is often the case for real data. Moreover, the matrix exponential is straightforward to compute and rather well-behaved (i.e., it preserves the normalization of the columns to one) [23].

We compute $P_{\mathrm{ON}}(t)$ using the transition probabilities determined by the propagator (Eq. 76 and 78), namely $P_{\mathrm{ON}}(t) = \sum_{n>0} P_n(t)$ with $P_n(t) = \sum_{n'} \hat{U}_{nn'}(t, t_1)P_{n'}(t_1)$. Incidentally, for a single gene copy in

the stationary case ($\hat{M}$ given by Eq. 74 with k_{ON} and k_{OFF} constant in time), the computation of $P_{\text{ON}}(t)$ above becomes equivalent to solving the moment equation for $\langle n(t) \rangle$, introduced previously (see Eq. 30).

To determine the expected $T_{\text{ON}}(t)$ and $T_{\text{OFF}}(t)$, we compute the discrete residence time distribution of the ON and OFF states in the non-stationary case. We define the sets of residence states $\mathcal{V}_r$ and absorbing states $\mathcal{V}_a$ such that $n \in \mathcal{V}_r \cup \mathcal{V}_a$ with $\mathcal{V}_r \cap \mathcal{V}_a = \emptyset$. For the ON-time we have $\mathcal{V}_r = \{1\}$ and $\mathcal{V}_a = \{0\}$ (1 gene copy) or $\mathcal{V}_r = \{1, 2\}$ and $\mathcal{V}_a = \{0\}$ (2 gene copies). For the OFF-time we have $\mathcal{V}_r = \{0\}$ and $\mathcal{V}_a = \{1\}$ (1 gene copy) or $\mathcal{V}_r = \{0\}$ and $\mathcal{V}_a = \{1, 2\}$ (2 gene copies). We define the probability $P_r(\tau = t_j - t_i | t_i)$ that the system has just settled at time t_i in any of the residence states ($n(t_i) \in \mathcal{V}_r$) and is found for the first time after a duration τ in any absorbing state ($n(t_j) \in \mathcal{V}_a$). Given that time is discrete and finite, with $t_{i+1} - t_i = \Delta t \ \forall i \in \{1, \dots, N_t - 1\}$ and N_t the total number of time points, the residence time (or first-passage time) distribution $P_r(\tau = t_j - t_i | t_i)$ can be written as

$$P_r(\tau = t_j - t_i | t_i) \equiv W_{ji}, \tag{79}$$

where W is a triangular matrix (with $W_{ji} = 0$ if $j < i$) whose columns correspond to truncated discrete phase-type distributions (truncated due to a finite number of time points N_t). The coefficients W_{ji} are defined as follows

$$W_{ji} = \begin{cases} \displaystyle\sum_{n \in \mathcal{V}_a} Q_n(t_j, t_i) & \text{if} \quad i \leq j < N_t \\ \\ \displaystyle 1 - \sum_{k=i}^{N_t - 1} W_{ki} & \text{if} \quad j = N_t \end{cases} \tag{80}$$

where the value for $j = N_t$ ensures that the W_{ji} form normalized distributions along j, i.e. $\sum_{j=i}^{N_t} W_{ji} = 1$. In practice, these distributions were computed by defining $Q_n(t_j, t_i)$ recursively for $1 \leq i < j \leq N_t$ using the propagator (Eq. 76 and 78):

$$Q_n(t_j, t_i) = \sum_{n' \in \mathcal{V}_r} \hat{U}_{nn'}(t_j, t_{j-1}) Q_{n'}(t_{j-1}, t_i) \tag{81}$$

The initial condition of our recursion $Q_n(t_i, t_i)$ (when $i = j$) corresponds to the probability that the system has just settled in any of the residence states at time t_i. For $1 < i \leq N_t$, this probability is given by

$$Q_n(t_i, t_i) = \begin{cases} \displaystyle\frac{1}{Z_i} \sum_{n' \in \mathcal{V}_a} \hat{U}_{nn'}(t_i, t_{i-1}) P_{n'}(t_{i-1}) & \text{if} \quad n \in \mathcal{V}_r \\ \\ 0 & \text{otherwise} \end{cases} \tag{82}$$

where $Z_i = \sum_{n \in \mathcal{V}_r} \sum_{n' \in \mathcal{V}_a} \hat{U}_{nn'}(t_i, t_{i-1}) P_{n'}(t_{i-1})$ with $P_n(t)$ the occupancy of the states given by Eq. 76. For $i = 1$, $Q_n(t_1, t_1)$ is actually ill-defined and we chose

$$Q_n(t_1, t_1) = \begin{cases} \displaystyle\frac{1}{Z_1} P_n(t_1) & \text{if} \quad n \in \mathcal{V}_r \\ \\ 0 & \text{otherwise} \end{cases} \tag{83}$$

where $Z_1 = \sum_{n \in \mathcal{V}_r} P_n(t_1)$. Consistent with our definition of $P_r(\tau | t_i)$, the probability that the system remains in a given state for a null duration is zero. Indeed, by construction $P_r(\tau = 0 | t_i) \equiv W_{ii} = 0$ since $W_{ii} = \sum_{n \in \mathcal{V}_a} Q_n(t_t, t_i)$ where $Q_n(t_t, t_i) = 0 \ \forall n \in \mathcal{V}_a$.

To compute the expected $T_{\text{ON}}(t)$ and $T_{\text{OFF}}(t)$ according to our definition (see Eq. 53 and 54), we renormalize the computed residence time distributions. Specifically, we want to compute the average

of $\tau = t_j - t_i$ for all possible t_i given t_j, meaning we must properly normalize the W_{ij} along i (they only form distributions along j). To do so, we first need to weight each t_i according to the probability that the system settled in any residence state at that particular time, which is given by the previously computed Z_i (see Eq. 82 and 83). In addition, we need to take into account that any τ contributes to $j - i$ time points along j. Thus, with that in mind, we defined the following renormalized coefficient $\tilde{W}_{ji}$ for $1 \leq i < j \leq N_t$:

$$\tilde{W}_{ji} = \frac{W_{ji}}{j-i}\frac{Z_i}{\tilde{Z}_j}, \tag{84}$$

with $\tilde{Z}_j = \sum_{i<j} \frac{W_{ji}}{j-i}\frac{Z_i}{\tilde{Z}_j}$, such that the $\tilde{W}_{ji}$ properly form distributions along i. Note that in practice, the Z_i's in the equation above cancel out with those in Eq. 82 and 83. The expected $T_{\mathrm{ON}}(t)$ and $T_{\mathrm{OFF}}(t)$ are then computed using the $\tilde{W}_{ji}$ as

$$T(t_j) = \sum_{i<j} \tilde{W}_{ji}(t_j - t_i), \tag{85}$$

where the appropriate residence state ($n \in \mathcal{V}_r$) and absorbing state ($n \in \mathcal{V}_a$) have to be used to determine the $\tilde{W}_{ji}$. Thus, Eq. 85 allows us to computationally predict the expected $T_{\mathrm{ON}}(t)$ and $T_{\mathrm{OFF}}(t)$, while accounting for discrete and finite number of time points time.

7.4.5 Validation using synthetic, simulation-generated data

To validate our Bayesian deconvolution and burst calling approaches, we tested our method on synthetic data mimicking experimental conditions. We initially used a 2-state model to generate single-gene copy initiation events I given a set of input parameters $\Theta = \{k, k_{\mathrm{ON}}, k_{\mathrm{OFF}}\}$, where k is the initiation rate and k_{ON} and k_{OFF} are the gene switching rates. The resulting initiation events are convolved with the elongation kernel κ of the gene *hb* (see Eq. 11) to generate transcriptional activity time series G, i.e., $G = \kappa * I$. Assuming two independent sister chromatids (as suggested by data, see Section 7.4.6), we sum over two independently generated transcriptional activity time series G per nucleus. Finally, we add the characterized measurement noise on the resulting activity according to Eq. 7 and 8, thus generating realistic single allele activity time series A per simulated nucleus.

Stationary bursting parameters. We first investigate the performance of our method in recovering stationary bursting parameters (i.e., not varying in time). We generated an initial simulated data set using the following input parameter ranges: $k \equiv K^{(1)} = 8\,\mathrm{mRNA/min}$ (consistent with real data), $k_{\mathrm{ON}} = \eta/T$ and $k_{\mathrm{OFF}} = (1-\eta)/T$ with $\eta \equiv P_{\mathrm{ON}}^{(1)} \in \{0.03, 0.12, 0.23, 0.36, 0.52, 0.78, 0.90\}$ and $T \equiv T_{\mathrm{C}}^{(1)} \in \{0.5, 1, 2, 3, 4, 5, 7, 10\}$ min. For each of the 56 combinations of input parameters $\{K^{(1)}, T_C^{(1)}, P_{\mathrm{ON}}^{(1)}\}$, we generate 200 alleles and 50 min long activity time series A at 10 s intervals (as in real data). For each allele, the underlying initiation events of the 2-state model are sampled using the Gillespie algorithm [20].

We next perform single-allele deconvolution and burst calling (see Sections 7.1 and 7.4) to estimate the effective mean transcription parameters (R, K, P_{ON}, T_{ON}, T_{OFF} and T_{C}) for each simulated set of alleles. As an example, in Supplementary Fig. 1a we show both the predicted parameters based on input (see 7.4.4) and the resulting estimated parameters (see 7.4.2) for $T = 2\,\mathrm{min}$ as a function of time and P_{ON} (color code). Overall, the estimated parameters align well with the predicted ones, albeit with fluctuations around the mean, which is expected with a sample of finite size (200 alleles).

We observe that the parameters T_{ON} and T_{OFF} are affected by different biases (Supplementary Fig. 1a): 1) bias due to censoring that leads to underestimation and the bending observed near the beginning

and the end of the time interval, 2) bias due to missing short ON/OFF periods compared to the sampling time leading to overestimation for small values of T_{ON} and T_{OFF}, 3) bias due to the finite length of the recording leading to underestimation for values of T_{ON} and T_{OFF} that are close or larger than the recording duration. Nevertheless, the estimates for T_{ON} and T_{OFF} remain good for the range of probed values. Interestingly, biases on T_{C} are rather limited, whereas fluctuations seem larger than observed for T_{ON} and T_{OFF}.

To highlight our ability to reconstruct bursting relationships, we test whether the estimated bursting parameters follow the predicted ones as P_{ON} varies. We compute the median (over time) and 68% confidence intervals of the estimated bursting parameters for each data set (i.e., a combination of input parameters), and plot these as a function of the estimated P_{ON} in Supplementary Fig. 1b. We further compute the predicted bursting parameters shown as dotted lines for the different values of input $T_C^{(1)}$ (see Section 7.4.4). While we observe biases when the input $T_C^{(1)}$ approaches 0.5 min, the estimated parameters reproduce the predicted trend sufficiently well to recover the relationships as a function of P_{ON}, over a large range of input $P_{\mathrm{ON}}^{(1)} \in [0.02, 0.90]$ and input $T_{\mathrm{C}}^{(1)} \in [0.5, 10]\,\mathrm{min}$.

Testing different models. We further explore how different generative models would impact the estimated bursting relationships. Since our burst calling approach does not assume any specific distributions of ON and OFF intervals, the estimated bursting parameters should not depend on the underlying generative model employed. To validate this property, we generated an alternative simulated data set using the same input parameters as above, but instead assuming fully correlated sister chromatids. This is achieved by simulating a single transcriptional activity time series G per nucleus, using a 2-state model whose initiation rate is doubled $k = 2K^{(1)}$. We see in Supplementary Fig. 1c that overall the estimated bursting parameters follow the expected relationships. Although we observe a larger bias in the estimation of K under the correlated case compared to the independent case (Supplementary Fig. 1b), the errors in T_{OFF}, T_{ON} and T_{C} remain comparable between the two scenarios.

We generated a third simulated data set using a 3-state pause model for each chromatid (Supplementary Fig. 1d), previously employed to explain transcription data in the early fly embryo [1]. Consistent with that study, we set reaction rates such that the pause state (Pause $\equiv \mathrm{OFF}_2$) is less frequent $k_{\mathrm{pause}} = k_{\mathrm{ON}}/25$ but more persistent $k_{\mathrm{unpause}} = k_{\mathrm{OFF}}/10$ than the basal OFF state (OFF_1). This model produces a hyperexponential waiting time distribution for effective OFF intervals, contrasting with the exponential distribution of the 2-state model. Despite these differences, Supplementary Fig. 1d shows that the estimated bursting relationships are properly recovered and are nearly indistinguishable from those of the 2-state model (see Supplementary Fig. 1b).

Lastly, we generated a fourth simulated data set using a 6-state cycle model for each chromatid (Supplementary Fig. 1e). This model produces gamma-distributed ON and OFF intervals, with shape parameter $\alpha = 3$ and rate parameter $\beta = 3k_{\mathrm{OFF}}$ and $\beta = 3k_{\mathrm{ON}}$. These distributions differ markedly from the exponential ON and OFF interval distributions of the 2-state model. Nevertheless, as shown in Supplementary Fig. 1e, the estimated bursting relationships are accurately recovered and are nearly indistinguishable from the 2-state model results (see Supplementary Fig. 1b). Together these findings confirm that our burst-calling approach remains robust across diverse generative models.

Error characterization. To fully characterize the errors in our estimates, we return to the realistic scenario of two independent sister chromatids. We compare the estimated effective bursting parameters to the predicted ones for datasets generated by the 2-state (Supplementary Fig. 1f) and 6-state cycle models (Supplementary Fig. 1g), which produce markedly different ON and OFF period distributions (exponential vs. gamma). For each dataset, we compute the median (over time) and 68% confidence intervals of the estimated bursting parameters (as in Supplementary Fig. 1b,e). Across a broad range of

input parameters—$P_{\mathrm{ON}}^{(1)} \in [0.02, 0.90]$ and $T_{\mathrm{C}}^{(1)} \in [0.5, 10]$ min— we observe excellent agreement between the median estimated and predicted bursting parameters (see Supplementary Fig. 1f,g).

Under both models, we observe increasing biases in P_{ON}, K and T_{ON} as $T_C^{(1)}$ approaches 0.5 min (light blue dots in Supplementary Fig. 1f,g). When $T_C^{(1)}$ becomes smaller than the default burst-calling window ($T_C^{(1)} < 1$ min), short OFF intervals between successive ON intervals are more likely to be missed. This leads to the merging of detected ON intervals, artificially inflating the estimated T_{ON}, which in turn increases P_{ON} and decreases K (since R is accurately estimated in all conditions). Nevertheless, these systematic biases do not prevent accurate estimation of T_{C} within the detection limit set by our experiment and burst-calling method.

Our detection limits for estimating T_{OFF} and T_{ON} are determined by the duration of our recordings ($\sim$ 40 min in NC14) and the sampling time ($\Delta t = 10$ s). The upper limit for T_{C} is approximately 10 min, which corresponds to $T_{\mathrm{ON}} = T_{0\mathrm{FF}} = 40/2$ min. An additional lower limit on T_{ON} is set by the burst-calling window size. Specifically, the minimal detectable T_{ON} is 30 s, which reflects the minimal time required for detecting two successive initiation events (when the tip of our ~ 1 min peaked-kernel used for averaging crosses the 2-molecule threshold). As a result, a conservative lower bound for T_{C} is 30 s. As shown in Supplementary Fig. 1f,g, we can accurately estimate T_{C} within the detection range of $0.5 - 10$ min, with relative deviations remaining below a factor of two.

Lastly, we evaluated the impact of the burst detection parameters w and r_b (see Eqn. 43) on the accuracy of the estimated bursting parameters. Using our default burst-calling parameter settings, i.e. $w = 1u$ and $r_b = 2/u$, where $u = 5\Delta t = 5/6$ min, we quantified the median relative error between the estimated and predicted parameters for the 2-state dataset. This was assessed both for individual parameters and globally, as a function of the input $T_{\mathrm{C}}^{(1)}$ (see Supplementary Fig. 1h). Despite using a fixed moving window of size $w = 1u$, our approach reliably recovered the bursting parameters (global median relative error $< 35\%$) across a broad range of correlation times ($T_{\mathrm{C}} \sim 0.5$–8 min), which includes the values estimated from experimental data ($T_{\mathrm{C}} \sim 1$–2 min). Thus, our choice of burst-calling parameters did not restrict our ability to detect either smaller or larger switching correlation times than those observed in experimental data (see Figs. 4, 5 and Supplementary Fig. 1f,g). Finally, we explored how varying w and r_b affect the global median relative error (see Supplementary Fig. 1i). This analysis confirmed that our default burst-calling parameters yield the lowest global relative error across the range of tested parameter values.

Time-dependent bursting parameters. We assess the performance of our method for estimating time-dependent bursting parameters in a realistic setting by generating a data set that mimics the spatiotemporal transcriptional output of *hb* in NC14 within $x/L \sim$0.20–0.46. The corresponding time-dependent input bursting parameters for 18 virtual AP bins (color-coded) are displayed in Supplementary Fig. 2e,f top. For each virtual position, we generate 200 alleles and 50 min long activity time series A at 10 s intervals (as in real data). Instead of sampling the initiation times series I given the 2-state model using a time-dependent Gillespie algorithm, we use the propagator (Eq. 76) to sample both the allele state n and the resulting initiation events g at fixed and short time interval such that $I_j = g(t_j)$.

To compute the full propagator of the system, we expand the previously defined state transition matrix (Eq. 75) to also include the initiation process. The new state transition matrix $\hat{M}'(t)$ can be written using tensor products:

$$\hat{M}'(t) = \hat{I}_g \otimes \hat{M}_2(t) + \hat{K}_g(t) \otimes \hat{R}_2, \tag{86}$$

where $\hat{I}_g$ is the identity matrix of size N_g, $\hat{M}_2(t)$ is the transition matrix of the allele state for two independent sister chromatids (see Eq. 75), $\hat{K}_g(t)$ is a square matrix of size N_g that describes the

initiation of transcripts and is given by

$$\hat{K}_g(t) = \begin{pmatrix} -k(t) & 0 & \dots & \dots & 0 \\ k(t) & -k(t) & \ddots & \ddots & \vdots \\ 0 & k(t) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -k(t) & 0 \\ 0 & \dots & 0 & k(t) & -k(t) \end{pmatrix},$$

where $k \equiv K^{(1)}$ corresponds to the single gene copy (SGC) initiation rate, and lastly the matrix $\hat{R}_2$ indicates (based on the allele state) how many gene copies participate in the initiation process:

$$\hat{R}_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Given the expanded state transition matrix (Eq. 86), we compute the complete time-dependent propagator of the system using the trapezoidal rule:

$$\hat{U}(t_{j+1}, t_j) \approx \exp\left(\frac{1}{2}\left(\hat{M}'(t_{j+1}) + \hat{M}'(t_j)\right)\Delta t'\right). \tag{87}$$

For accurate sampling, we set $t_{j+1} - t_j = \Delta t'$ to the minimal discretized time interval $\Delta t' = f/(2K_{\text{elo}}) = 1\text{s}$ given by the Pol II footprint of $f = 60\,\text{bp}$ and the measured elongation rate $K_{\text{elo}} = 1.8\,\text{kb/min}$. In addition, we set $N_g = 6$, which is a safe cutoff given the chosen input parameters and the small $\Delta t'$.

Once the propagator has been computed (Eq. 87), sampling the allele state and the initiation events becomes straightforward. Given a specific allele state $n(t_j) \in \{0, 1, 2\}$ at time t_j, we define the expanded state $\lambda = n(t_j)N_g + 1$ and the corresponding zero vector $\mathbf{\Lambda}(t_j)$ of size $3N_g$, whose λ-th entry has value 1. The vector $\mathbf{\Lambda}(t_j)$ and the propagator enables computing the probability $P_{\lambda'}(t_{j+1})$ to find the system in any state $\lambda' = n(t_{j+1})N_g + g(t_{j+1}) + 1 \in \{1, 2, \dots, 3N_g\}$ at time $t_{j+1} = t_j + \Delta t'$, where $g(t_{j+1}) \in \{0, 1, \dots, N_g - 1\}$ is the number of initiation events during the time interval $\Delta t'$. We compute the vector $\mathbf{P}(t_{j+1})$ whose entries are the probability $P_{\lambda'}(t_{j+1})$ as

$$\mathbf{P}(t_{j+1}) = \hat{U}(t_{j+1}, t_j)\mathbf{\Lambda}(t_j). \tag{88}$$

Using $\mathbf{P}(t_{j+1})$, we sample the state λ' at time t_{j+1} by drawing a single random number $\lambda' \in \{1, 2, \dots, 3N_g\}$ according to $P_{\lambda'}(t_{j+1})$. The corresponding new allele state is then given by $n(t_{j+1}) = \lfloor(\lambda' - 1)/N_g\rfloor$ and the number of initiation events during $\Delta t'$ by $g(t_{j+1}) = (\lambda' - 1) \mod N_g$. Setting the new state vector $\mathbf{\Lambda}(t_{j+1})$ as the zero vector whose λ''-th entry is set to 1, with $\lambda'' = n(t_{j+1})N_g + 1$, defines an iterative procedure to generate times series according to the 2-state model with time-dependent input parameters. Specifically, this approach enables sampling single initiation time series I such that $I_j = g(t_j)$, i.e., the number of initiation events during the time interval $(t_{j_1}, t_j]$. Given the short interval $\Delta t' = 1\text{s}$ and the probed value of the initiation rate $k \leq 8$ mRNA/min, the sampled number of initiation events $g(t_j)$ is usually comprised between 0 and 1, and rarely exceeds 2.

Capitalizing on the sampling procedure described above, we generate 200 initiation time series I and the corresponding activity time series A at 10 s intervals for each virtual position. We first compared the estimated time-dependent bursting parameters with the predicted ones (Supplementary Fig. 2a-d). The temporal changes in effective bursting parameters are recovered well (Supplementary Fig. 2a,b). Indeed, both the predicted and estimated parameters are highly correlated (Supplementary Fig. 2c,d).

We further estimate the single gene copy (SGC) input parameters from the estimated effective ones using the relationships derived in Section 7.4.3. Even though the SGC input parameters are deeply buried within the data, our deconvolution and burst calling approaches lead to a well-behaved time-dependent estimation of the latter (Supplementary Fig. 2e,h). Albeit with some distortions, we recover the input parameters adequately, confirming our ability to estimate bursting parameters in a real-case scenario.

7.4.6 Validation on real data

To further validate our burst-calling approach, we used the gap gene bursting parameter estimates (Fig. 4) to predict key statistical properties of the data. Specifically, we tested whether the estimated parameters accurately predict the distributions of single-allele transcription rates and inter-initiation intervals. Achieving this required postulating an explicit transcriptional model. Given that the measured nascent transcription spots represent two unresolved gene copies (see Section 7.4.3), we modeled their signal as two coupled 2-state systems, linked by a coupling parameter $\eta \in [0, 1]$. The resulting model is shown in Supplementary Fig. 3a, and the corresponding SGC (single gene copy) parameter transformations are detailed in the above-mentioned section.

To facilitate the interpretation of the distributions of single-allele transcription rate and inter-initiation intervals, we focus on data from NC13 and early NC14 ($7.5 \leq t < 20.5\,\mathrm{min}$), where effective bursting relationships are rather stable in time (see Extended Data Fig. 8a)). For that time interval, we reconstruct the distribution of single-allele transcription rates $P(r|R)$ using 2-min-intervals (see Section 7.3.1). From the distributions, we estimate the variance (2^{nd} cumulant), 3^{rd} cumulant, and 4^{th} cumulant of single allele transcription rate r as a function of mean transcription rate R (see Supplementary Fig. 3d). The cumulants are normalized by the maximal mean transcription rate $\max R = R_{\max} = 15.77\,\mathrm{mRNA/min}$.

The estimated bursting relationships set strong constraints on our model parameters (see Supplementary Fig. 3b). Indeed, together with Equ. 69 and 71, they fully determine the SGC parameters $K^{(1)}$, k_{ON}, and k_{OFF} as a function of the coupling parameter η (see Supplementary Fig. 3c). Given these SGC parameters and our model, we then aimed to compute the predicted distributions of single-allele transcription rate $P(r|R)$, where $R = 2K^{(1)}P^{(1)}_{\mathrm{ON}}$ and $P^{(1)}_{\mathrm{ON}} = k_{\mathrm{ON}}/(k_{\mathrm{ON}}+k_{\mathrm{OFF}})$. Given the short 2-minute interval over which these distributions are computed, it is crucial to account for the Pol II footprint (see Section 7.3.2). Thus, to compute these distributions, we opt for a sampling approach similar to the one used to generate synthetic data (see Section 7.4.5).

Assuming a steady state on the 2-min interval, we first compute the propagator of the gene states $\hat{U} = \exp(\hat{M}(\eta)\Delta t')$. The state rate matrix $\hat{M}(\eta)$ for two arbitrarily correlated gene copies can be written as

$$\hat{M}(\eta) = \begin{pmatrix} -(2-\eta)k_{\mathrm{ON}}(\eta) & k_{\mathrm{OFF}}(\eta) & \eta k_{\mathrm{OFF}}(\eta) \\ 2(1-\eta)k_{\mathrm{ON}}(\eta) & -(k_{\mathrm{OFF}}(\eta)+k_{\mathrm{ON}}(\eta)) & 2(1-\eta)k_{\mathrm{OFF}}(\eta) \\ \eta k_{\mathrm{ON}}(\eta) & k_{\mathrm{ON}}(\eta) & -(2-\eta)k_{\mathrm{OFF}}(\eta) \end{pmatrix}, \tag{89}$$

where $k_{\mathrm{ON}}(\eta)$ and $k_{\mathrm{OFF}}(\eta)$ are given by Equ. 71. We set $\Delta t'$ to the minimal discretized time interval $\Delta t' = f/(2K_{\mathrm{elo}}) = 1\,\mathrm{s}$ given by the Pol II footprint of $f = 60\,\mathrm{bp}$ and the measured elongation rate $K_{\mathrm{elo}} = 1.8\,\mathrm{kb/min}$. At each time point $t_{j+1} = t_j + \Delta t'$, we compute the new gene state probability vector $\mathbf{P}(t_{j+1})$ according to $\mathbf{P}(t_{j+1}) = \hat{U}\mathbf{\Lambda}(t_j)$, where $\mathbf{\Lambda}(t_j)$ is the state indicator vector at time t_j. We then sample the new gene state $n(t_{j+1}) \in \{0, 1, 2\}$ according to $\mathbf{P}(t_{j+1})$, thus generating the new $\mathbf{\Lambda}(t_{j+1})$. Using this iterative sampling strategy, we generate 2-min gene state sequences $n = n(t_0), n(t_1), ...$ For each sequence n, we sample a single configuration of initiation events I, such that $I(t_j) \equiv I_j \in \{0, 1\}$ is drawn from

$$P(I_j|p_j = n(t_j)K^{(1)}(\eta)\Delta t') = p_j^{I_j}(1-p_j)^{1-I_j}, \tag{90}$$

where $K^{(1)}(\eta)$ is given by Equ. 69. For each I, we then calculate the single allele transcription rate r according to Equ. 18 with $v = 12$.

Thus, by sampling more than 10^5 n and I samples per condition (defined by the choice of input SGC parameters), we can easily reconstruct the distribution of single allele transcription rate $P(r|R = 2K^{(1)}P_{\mathrm{ON}}^{(1)}$ shown in Supplementary Fig. 3e. Similarly, using the sampled initiation configuration I, we can easily compute the distributions of inter-initiation events (i.e., the time interval between events) shown in Supplementary Fig. 3f as survival functions.

By varying the coupling parameter $\eta \in [0, 1]$, we immediately see that the data favors weak coupling between sister chromatids (see Supplementary Fig. 3d-f). Indeed, the best fit to the cumulants leads to $\eta = 0.27$, with a small correlation coefficient $\rho = \eta/(2 - \eta) = 0.16$. Fully correlated sister chromatids ($\rho = 1$) provide a poor match to the data, whereas sister chromatids that are independent ($\rho = 0$) to weakly correlated ($\rho = 0.16$) lead to an excellent match. Such a good match, both at the level of distributions of single-allele transcription rate and inter-initiation events, not only suggest that sister chromatids are indeed close to independent but also further validates our estimated bursting parameters that successfully recapitulate key statistical properties of the data.

7.5 Exploring the generality of bursting rules

We test the validity of our derived bursting relationships beyond the *Drosophila* gap genes. To this end, we extract bursting parameters from four other studies performed in different laboratories on different genes and systems.

7.5.1 *Drosophila* data

While our study mainly focuses on the *Drosophila* gap genes, two other studies performed in early *Drosophila* embryos (i.e., one on dorsoventral genes [24] and another one on synthetic core promoter modifications [1]) generated a similar set of bursting parameters that allow for a test of our established rules. For both studies, we assume that the estimated parameters are effective ones (for two indistinguishable sister chromatids) and not single gene copy parameters.

From the first study (Figure 5D in [24]), we recover the ON-probability (occupancy) P_{ON}, the initiation rate (loading rate) K, and the burst frequency F for *ush*-wt and st2-*dpp* for the ten positions along the dorsoventral axis considered in that study. We calculate the mean ON-time $T_{\mathrm{ON}} = P_{\mathrm{ON}}/F$, the mean OFF-time $T_{\mathrm{OFF}} = (1 - P_{\mathrm{ON}})/F$, and the switching correlation time and plot the resulting parameters in yellow in Fig. 6g.

From the second study [1], we directly use the estimated parameters (T_{OFF}, T_{ON}, K and P_{ON}) for the 2-state model provided in Supplementary Table 2. These correspond to the following seven constructs: snaE>*sna*>24xMS2-y, snaE>*snaTATAlight*>24xMS2-y, snaE>*snaTATAmut*>24xMS2-y, snaE>*snaTATAlight+INR*>24xMS2-y, snaE>*kr-INR1*>24xMS2-y, snaE>*kr-INR2*>24xMS2-y and snaE>*Ilp4-INR*>24xMS2-y. The resulting seven data points are plotted in pink in Fig. 6g. We did not use the estimated parameters for the 3-state model due to some arbitrariness in the potential definitions of an equivalent 2-state T_{OFF} in that particular model.

7.5.2 Yeast data

We analyzed estimated bursting parameters from a recent study on yeast genes [25]. From their Extended Data Fig. 6, we extracted the mean ON-time (burst duration) T_{ON} and mean OFF-time (time between bursts) T_{OFF} for *GAL10* under 42 different growth conditions. The ON-probability was computed assuming steady state as $P_{\mathrm{ON}} = T_{\mathrm{ON}}/(T_{\mathrm{ON}} + T_{\mathrm{OFF}})$.

To infer the absolute initiation rate K, we calibrated the unscaled burst intensity $\tilde{K}$, necessitating both an estimated mRNA residence time on the gene and absolute units. To this end, we first estimated the residence time of nascent mRNA on *GAL10* from the estimated genomic gene length (2200 bp + 862 bp for the PP7 cassette) and the estimated elongation rate of 3.9 kb/min [26]. As there might be some retention of nascent transcripts at the transcription site, we round up the gene length to 3.9kb, leading to a residence time for nascent mRNAs of approximately 1 min.

To obtain absolute units, we used smFISH analysis results (in triplicate) of the -RSC (DMSO) experiment [25], allowing us to calculate the mean number of nascent transcripts on the gene. With a conservative definition of transcription sites (spots having at least 5 nascent transcripts), we obtain on average 13.2 ± 0.2 nascent transcripts per transcription site. Given that the mRNA residence time is short relative to T_{ON}, i.e., 1 min compared to $T_{\mathrm{ON}} \sim 1.7$ min, we assume that nascent transcripts accumulate primarily during the ON period. This yields a lower-bound estimate of the absolute initiation rate as $K = 13.2\tilde{K}/\langle\tilde{K}\rangle$, where the average is performed over the replicates.

We report the resulting parameters across all 42 conditions as orange circles in Fig. 7a. While these yeast parameters are extracted from a single gene copy system, the difference between effective and single gene copy is not very pronounced below $P_{\mathrm{ON}} < 5$ (see Section 7.4.3).

7.5.3 Human data

We re-analyzed single-cell transcriptional time traces of eleven endogenous human genes obtained through live imaging of a MS2 reporter [27, 28]. Two key reasons motivated such an analysis: 1) the definition of the ON period in the original studies is different from ours, conflating both the mRNA dwell-time (resulting from elongation, cleavage and termination) with the promoter ON-time; 2) the fitted generalized transcription model relies on different assumptions (3 promoter states, maximum number of nascent mRNA between 2 and 5, non-trivial dwell-time distribution due to early intron ejection and initiation prevented by early PolII engagement), which make the mapping of their parameters onto ours difficult. Thus, we chose to apply our deconvolution method to the original data and reconstruct the initiation events, allowing us to estimate bursting parameters whose definitions are identical to ours.

The single-cell time traces from the original studies were obtained using a 24xMS2 stem-loops cassette inserted either in the 3'UTR (*TFF1*, [27]) or within introns (10 other genes, [28]). The fluorescent signal resulted from MCP-GFP imaged using a confocal microscope (Zeiss LSM780), with a temporal resolution of $\Delta t = 100$ s. The original dataset comprised 35 to 123 cells per gene, whose time traces last on average 11.5 hours. As these human genes are lowly expressed, estimating the fluorescent background in the data was straightforward. Indeed, only a few bursts are noticeable in each single-cell time trace and the majority of the time points correspond to background fluctuations. We thus computed the histogram of intensities for each cell, resulting in a mixture of a Gaussian (background fluctuation) and a heavy-tailed distribution (nascent mRNA counts plus noise). We kernel smoothed the resulting distributions to robustly estimate the mode and the 2.5^{th} percentile. For each cell, we defined the mean of the background μ_b as the mode of the mixture and the standard deviation of the background σ_b as the half difference between the mode and 2.5^{th} percentile. We checked the consistency of our assessment of the background

distribution across all cells for a given gene, noticing only a very limited number of outlier time traces corresponding to highly active alleles. For these traces, we replaced μ_b and σ_b by med(μ_b) and med(σ_b), the median over all cells of the corresponding gene. Lastly, we subtracted the cell-specific background μ_b to the intensities of all traces, such that in the absence of transcription the signal is centered around the true zero.

Next, we calibrated the fluorescent signal in absolute units, i.e., nascent mRNA counts. We used the gene *TFF1* as nascent mRNA counts were characterized by smFISH (see Fig. S2C-D in [27]). Importantly, these smFISH distributions do not include non-transcribing alleles (as with dual color smFISH, the detection of transcription sites necessitates intron signal). Thus, to match the histogram of live fluorescent intensities with the smFISH counts for calibration, one needs to account for the unknown fraction of non-transcribing alleles. We fitted the live histogram of intensities (maximum likelihood) with a mixture distribution composed of Gaussian background noise $\mathcal{N}(s|\mu, \sigma_b)$ and smFISH mRNA count distribution $P(g)$, which is given by

$$P(s|\rho, \alpha) = \frac{1}{Z}\left(\rho\mathcal{N}(s|0, \sigma_b) + (1-\rho)\sum_{g=1}^{\infty}\mathcal{N}(s|\alpha g, \sigma_b)P(g)\right),$$

where s is the live intensity, $\rho \in [0, 1]$ the fraction of time nascent mRNA are present at the locus, and $\alpha \in [0, +\infty)$ the calibration factor (i.e., the live intensity of a single nascent mRNA). Z is the normalization constant such that $\int P(s)ds = 1$ and $P(g)$ is normalized such that $\sum_{g=1}^{\infty} P(g) = 1$ (thus excluding $P(g = 0)$ corresponding to the unknown fraction of non-transcribing alleles). By maximizing the likelihood above for the *TFF1* data, we obtained $\rho = 0.81$ and $\alpha = 38$. We managed to calibrate the other genes in absence of available smFISH data by extrapolation. We first noticed that med(σ_b) (median over all cells for a given gene) scales linearly with med(μ_b) across genes and over a 2-fold range ($R^2 = 0.96$), which points mainly to a gain difference between datasets. As the different genes were imaged using the same MCP-GFP in the same condition, the background fluctuations should in principle be the same. We thus rescaled the background-subtracted signal of the other 10 genes using med(σ_b)/(αmed(σ_b^{TFF1})), giving us absolute units for all the dataset.

To deconvolve the single-cell time traces, we first need to characterize the gene-specific mRNA dwell-time and assess measurement noise. We computed the auto-correlation (AC) function for each gene using all the time traces. From the AC (at lag $\tau = 0$), we estimated the fraction of uncorrelated variability $\sigma^2_{\text{img}}/\sigma^2$, where σ^2 is the total variance (Table 2, see Extended Data Fig. 1e,f for comparison with our data). Measurements noise represents 34% of total variability on average (5% in ours). Given these numbers, we extrapolated a simple measurement noise model in absolute units, i.e., $\sigma(g) = \sqrt{\sigma_b^2 + \beta g}$, with $\sigma_b = 0.5$ and $\beta = 0.38$. This noise model includes Poisson shot noise as in ours (see Equ. 7) and complies with single mRNA sensitivity, as claimed in the original study.

Regarding the correlated component of the AC (for lag $\tau > 0$), it is well approximated by a single exponential that decays with the lag τ over a timescale T_{d}. We thus fitted these AC functions with an exponential estimating T_{d} for each gene (Table 2). Notably, these time scales are much larger than what we observed in flies (6-fold larger on average). For the gap genes, we know that the mRNA dwell-time is rather small (∼2 min), due to the rapid elongation of short mRNAs (see Section 7.1.6 and Table 1) and their small retention time at the locus [12]. On the other hand, it has been demonstrated in the original studies [27, 28] that the mRNA-dwell time of these human genes is rather long (typically around 10 min, even for short mRNAs), due to a combination of transcription elongation, splicing, and termination. Following this observation, we assumed that the measured T_{d} for each gene mainly correspond to mRNA dwell-times, thus setting an upper bound on possible switching correlation time (see Sec. 7.3.3 and Equ.

Gene	$\sigma^2_{\text{img}}/\sigma^2$	T_{d} [min]
TFF1	0.21 ± 0.02	10.5 ± 0.8
CANX	0.53 ± 0.04	7.4 ± 1.5
DNAJC5	0.29 ± 0.02	10.4 ± 0.9
ERRFI1	0.49 ± 0.06	15.7 ± 2.6
KPNB1	0.26 ± 0.03	11.8 ± 1.4
MYH9	0.24 ± 0.03	6.2 ± 0.8
RAB7A	0.25 ± 0.02	12.4 ± 1.3
RHOA	0.32 ± 0.03	9.0 ± 1.2
RPAP3	0.62 ± 0.04	7.9 ± 1.7
SEC16A	0.39 ± 0.04	15.5 ± 2.1
SLC2A1	0.15 ± 0.03	14.9 ± 2.0

Table 2: Estimated fractional imaging noise and dwell-time for the 11 human genes from [27, 28].

39).

Having calibrated the data and characterized both the measurement noise and the mRNA dwell-time, we proceeded to re-analyze all the transcriptional time traces according to our deconvolution approach (see Section 7.1). We assumed a squared elongation kernel whose duration was set by T_{d} (Table 2), thus neglecting the discrete effect of the individual MS2 loops. Indeed, the whole MS2 sequence length is ~ 1.4 kb long and thus transcribed in 0.7 min ($K_{\text{elo}} = 2$ kb/min in humans [28], which is negligible compared to the total dwell-time. To sample the initiation events, we used $\Delta t' = 10$ s (instead of $\Delta t' = 1$ for our data), which was small enough given the temporal resolution ($\Delta t = 100$ s) and the maximal amount of nascent mRNAs (~ 10). Once the initiation events have been reconstructed, we performed our burst calling routine (see Section 7.4). For the burst calling, we set the averaging window to $w = 3\Delta t = 300$ s and $g_b = 1$. The window w is roughly half the average dwell time (~ 10 min), which is the limit of what we can reliably detect based on simulated data (Supplementary Fig. S1). The threshold g_b corresponds to one mRNA, consistent with their claimed single mRNA sensitivity [27, 28]. We then estimated the bursting parameters as described in Section 7.4.2. For individual time points, the bursting parameter relative errors are on average 6-fold larger than for our data (27% vs 4% relative error on average). However, since these human genes are at steady state (their bursting parameters are stable in time), we only report the time-averaged parameters. These parameters are given in Table 3 and 4, and shown in Fig. 7.

7.5.4 Mouse scRNA-seq data

We reanalyzed single-cell allelic RNA-seq data from mouse fibroblasts and embryonic stem cells [3]. We followed a very similar approach as in the original study to estimate bursting parameters from UMI (unique molecular identifier) counts. Namely, we fitted the steady state distribution of UMI counts for each gene through maximum likelihood, using a Beta-Poisson distribution (the steady state distribution of the 2-state model of transcription). We also used mRNA lifetime estimates [29] to get physical time units. Unlike the original study, we explicitly imposed a prior on the switching correlation time T_{C}, effectively regularizing the parameter inference. Indeed, for the vast majority of genes, the empirical steady-state distribution is not sufficient to determine all the three independent parameters of the 2-state model (K, P_{ON}, T_{C}). Typically, using mature mRNA counts, one cannot resolve a T_{C} value that is

Gene	R [1/min]	K [1/min]	P_{ON}	T_{OFF} [min]	T_{ON} [min]	T_{C} [min]
TFF1	0.027 ± 0.001	0.54 ± 0.02	0.050 ± 0.016	71.3 ± 11.7	4.4 ± 0.7	4.1 ± 0.6
CANX	0.037 ± 0.017	0.55 ± 0.03	0.065 ± 0.035	37.7 ± 11.2	2.7 ± 0.6	2.5 ± 0.5
DNAJC5	0.084 ± 0.016	0.52 ± 0.01	0.163 ± 0.029	17.0 ± 1.9	3.5 ± 0.4	2.9 ± 0.3
ERRFI1	0.008 ± 0.004	0.57 ± 0.03	0.009 ± 0.007	213.6 ± 44.7	2.8 ± 1.2	2.8 ± 1.2
KPNB1	0.101 ± 0.041	0.55 ± 0.04	0.182 ± 0.067	19.9 ± 4.4	4.7 ± 0.8	3.8 ± 0.5
MYH9	0.081 ± 0.033	0.51 ± 0.02	0.157 ± 0.056	24.9 ± 4.6	5.5 ± 1.2	4.5 ± 0.8
RAB7A	0.052 ± 0.017	0.53 ± 0.02	0.097 ± 0.032	25.3 ± 5.7	3.0 ± 0.5	2.6 ± 0.4
RHOA	0.117 ± 0.032	0.51 ± 0.01	0.231 ± 0.060	12.1 ± 2.4	3.8 ± 0.6	2.8 ± 0.3
RPAP3	0.011 ± 0.005	0.57 ± 0.03	0.018 ± 0.012	117.2 ± 31.3	2.4 ± 0.6	2.3 ± 0.5
SEC16A	0.036 ± 0.015	0.55 ± 0.02	0.065 ± 0.026	37.1 ± 9.7	2.9 ± 0.5	2.6 ± 0.4
SLC2A1	0.067 ± 0.029	0.53 ± 0.03	0.122 ± 0.049	32.9 ± 8.0	5.5 ± 1.6	4.5 ± 1.2

Table 3: Estimated bursting parameters using our deconvolution approach for the 11 human genes from [27, 28].

Gene	F [1/min]	B
TFF1	0.014 ± 0.002	2.33 ± 0.39
CANX	0.026 ± 0.007	1.50 ± 0.28
DNAJC5	0.049 ± 0.005	1.79 ± 0.19
ERRFI1	0.005 ± 0.001	1.59 ± 0.58
KPNB1	0.042 ± 0.007	2.62 ± 0.53
MYH9	0.034 ± 0.005	2.83 ± 0.66
RAB7A	0.037 ± 0.007	1.58 ± 0.25
RHOA	0.064 ± 0.009	1.91 ± 0.27
RPAP3	0.009 ± 0.002	1.33 ± 0.26
SEC16A	0.026 ± 0.007	1.58 ± 0.30
SLC2A1	0.027 ± 0.005	2.94 ± 0.87

Table 4: Estimated burst frequency and size using our deconvolution approach for the 11 human genes from [27, 28].

smaller than the mRNA lifetime T_{M}, which is of the order of a few hours ($\mathrm{med}(T_{\mathrm{M}}) = 5.3$ h). To alleviate this identifiability issue, we imposed a prior on T_{C} such that the effective log-likelihood $\log \tilde{\mathcal{L}}$ is given by

$$\log \tilde{\mathcal{L}}(D|K, P_{\mathrm{ON}}, T_{\mathrm{C}}) = \log \mathcal{L}(D|K, P_{\mathrm{ON}}, T_{\mathrm{C}}) - \frac{1}{2}\left(\frac{\log_{10} T_{\mathrm{C}} - \mu}{\sigma}\right)^2, \tag{91}$$

where $\log \mathcal{L}(D|\Theta)$ is the Beta-Poisson distribution (see [3]), and the prior parameters are set to $\mu = 0$ and $\sigma = 0.5$. The prior corresponds to a Gaussian in log-space centered around $T_{\mathrm{C}} = 1$ min, whose 95% coverage spans two orders of magnitude ($T_{\mathrm{C}} \in [0.1, 10]$ min). Given the range of estimated T_{C} values (Fig. 7a), this prior is not strongly informative and simply enforces a plausible range compatible with observations.

For each gene, we estimated the bursting parameters K, P_{ON} and T_{C} by maximizing the effective likelihood (Equ. 91) on UMI counts. We used the following range for the optimization: $P_{\mathrm{ON}} \in [10^{-5}, 1]$, $T_{\mathrm{C}} \in [10^{-2}, 10^{3}]$ min and $K \in [10^{-2.5}, 10^{2.5}]$ min^{-1}. To assess inference errors σ_Θ, we bootstrapped the estimated parameters by resampling the UMI distributions. We only reported genes whose parameters satisfy $2\sigma_\Theta < 0.5$ in log-space and $B = KT_{\mathrm{C}}/(1 - P_{\mathrm{ON}}) > 1$. This effectively removes 32% of the genes, which is comparable to the effect of the thresholding applied in the original study. We show the mean transcription rate $R = KP_{\mathrm{ON}}$, the burst frequency $F = P_{\mathrm{ON}}(1-P_{\mathrm{ON}})/T_{\mathrm{C}}$, the burst size B, the initiation rate K and the switching correlation time T_{C} in Supplementary Fig. 5a,b. We then assessed the impact of our prior on the maximum log-likelihood values. The median penalty imposed by our prior on the log-likelihood is 0.06, and for 80% of the genes, this penalty is less than one. It turns out that the resulting best-fit distributions are barely different from the ones obtained without priors, suggesting that indeed the mRNA distributions contain very little information about T_{C} when $T_{\mathrm{C}} \ll T_{\mathrm{M}}$ (the mRNA lifetime).

As in the original study [3], the analysis above was performed without correcting the UMI counts for low mRNA capture rate (claimed to be ∼10–30% on average in recent scRNA-seq, likely closer to $5-6$%, see [30]). This capture typically varies across cells, and possibly significantly across genes. Furthermore, extrinsic noise, notably related to the cell-cycle, generates significant variability in total count numbers across cells. Following the approach proposed by recent scRNA-seq analysis [30, 31, 32], we then aimed to further correct for the low capture rate and the extrinsic variability. We modified the likelihood above (Equ. 91) to include a cell-dependent correction β_j on the initiation rate K:

$$\log \mathcal{L}(D|K, P_{\mathrm{ON}}, T_{\mathrm{C}}) = \sum_{j=1}^{N_{\mathrm{cell}}} \log \mathcal{L}(x_j|\beta_j K, P_{\mathrm{ON}}, T_{\mathrm{C}}), \tag{92}$$

where x_j is the UMI count in cell j for a given gene. The correction factor is defined as follows:

$$\beta_j = \frac{T_j}{\bar{T}}\bar{\beta}, \tag{93}$$

where T_j is the total UMI counts (over all genes) in cell j, $\bar{T}$ is the average total counts across cells and $\bar{\beta}$ is the average mRNA capture rate that we fixed to 20%. The main rationale behind the β_j correction factor is that the low capture rate and cell size mainly affect the initiation rate [30, 31], introducing biases that could further propagate to other parameters. Combining Equ. 91 and 92, we re-estimated the bursting parameters K, P_{ON} and T_{C} for each gene. We show the resulting mean transcription rate $R = KP_{\mathrm{ON}}$, the burst frequency $F = P_{\mathrm{ON}}(1-P_{\mathrm{ON}})/T_{\mathrm{C}}$ and the burst size B in Fig. 7b, and the corrected initiation rate K and the switching correlation time T_{C} in Supplementary Fig. 5c. Beyond a rescaling of K by a factor ~ 2.2 (Supplementary Fig. 5b,c), the parameter relationships remain mostly unaffected by the correction (Supplementary Fig. 5a and Fig. 7B).

7.6 Bursting parameters and noise

To provide a simple explanation as to why low switching correlation T_{C} could be advantageous, we simply highlight here its link to expression noise. The variance in accumulated mRNA molecules given by a 2-state model is as follows:

$$\sigma_{\mathrm{M}}^2 = KT_{\mathrm{M}}P_{\mathrm{ON}} + (KT_{\mathrm{M}})^2 P_{\mathrm{ON}}(1 - P_{\mathrm{ON}})\phi(T_{\mathrm{M}}/T_{\mathrm{C}}), \tag{94}$$

where T_{M} is the mRNA life-time and $\phi(x) = 1/(1+x)$ a filtering function. The noise is then given by $\eta^2 = \sigma_{\mathrm{M}}^2/\mu_{\mathrm{M}}$, where $\mu_{\mathrm{M}} = KT_{\mathrm{M}}P_{\mathrm{ON}}$. The mRNA noise can only be minimized by making the binomial term (second term in Equ. 94) vanish. Given that K is observed to be constant, there are two ways to do that: 1) set P_{ON} to zero, which is counterproductive as regulation can no longer be performed, or 2) make T_{C} as small as possible compared to T_{M} such that $\phi \to 0$. The second option doesn't interfere with the expression range and ensures minimal noise, namely Poisson noise $\eta = 1/\mu_{\mathrm{M}}$. Thus, having a small T_{C} compared to the typical mRNA lifetime (tens of minutes for the gap genes, a few hours for typical mammalian genes), appears as a reasonable strategy to reduce expression noise over a large range of expression levels. The argument above remains the same at the protein levels, namely $T_{\mathrm{C}} \ll T_{\mathrm{P}}$, where T_{P} is the protein lifetime.

8 References

References

[1] Virginia L. Pimmett, Matthieu Dejean, Carola Fernandez, Antonio Trullo, Edouard Bertrand, Ovidiu Radulescu, and Mounia Lagha. Quantitative imaging of transcription in living Drosophila embryos reveals the impact of core promoter motifs on promoter state dynamics. *Nature Communications*, 12(1):4504, 2021.

[2] Benjamin Zoller, Damien Nicolas, Nacho Molina, and Felix Naef. Structure of silent transcription intervals and noise characteristics of mammalian genes. *Molecular Systems Biology*, 11(7):823, 07 2015.

[3] Anton J. M. Larsson, Per Johnsson, Michael Hagemann-Jensen, Leonard Hartmanis, Omid R. Faridani, Björn Reinius, Åsa Segerstolpe, Chloe M. Rivera, Bing Ren, and Rickard Sandberg. Genomic encoding of transcriptional burst kinetics. *Nature*, 565(7738):251–254, 12 2018.

[4] Julien O Dubuis, Reba Samanta, and Thomas Gregor. Accurate measurements of dynamics and reproducibility in small genetic networks. *Molecular Systems Biology*, 9(1):639–639, 2013.

[5] William A. Rogers, Yogesh Goyal, Kei Yamaya, Stanislav Y. Shvartsman, and Michael S. Levine. Uncoupling neurogenic gene networks in the Drosophila embryo. *Genes & Development*, 31(7):634–638, 4 2017.

[6] Hongtao Chen, Michal Levo, Lev Barinov, Miki Fujioka, James B. Jaynes, and Thomas Gregor. Dynamic interplay between enhancer–promoter topology and gene activity. *Nature Genetics*, 50(9):1296–1303, 2018.

[7] Hernan G Garcia, Mikhail Tikhonov, Albert Lin, and Thomas Gregor. Quantitative imaging of transcription in living drosophila embryos links polymerase activity to patterning. *Current biology*, 23:2140–5, 11 2013.

[8] Jacques P. Bothma, Matthew R. Norstad, Simon Alamos, and Hernan G. Garcia. LlamaTags: A Versatile Tool to Image Transcription Factor Dynamics in Live Embryos. *Cell*, 173(7):1810–1822.e16, 06 2018.

[9] Thomas Gregor, Eric F. Wieschaus, Alistair P. McGregor, William Bialek, and David W. Tank. Stability and Nuclear Dynamics of the Bicoid Morphogen Gradient. *Cell*, 130(1):141–152, 07 2007.

[10] Feng Liu, Alexander H. Morrison, and Thomas Gregor. Dynamic interpretation of maternal inputs by the Drosophila segmentation gene network. *Proceedings of the National Academy of Sciences*, 110(17):6724–6729, 04 2013.

[11] Luke Hayden, Woonyung Hur, Massimo Vergassola, and Stefano Di Talia. Manipulating the nature of embryonic mitotic waves. *Current Biology*, 32(22):4989–4996.e3, 2022.

[12] Benjamin Zoller, Shawn C. Little, and Thomas Gregor. Diverse Spatial Expression Patterns Emerge from Unified Kinetics of Transcriptional Bursting. *Cell*, 175(3):835–847.e25, 10 2018.

[13] J. deLeeuw. Introduction to Akaike (1973) Information Theory and an Extension of the Maximum Likelihood Principle. In *Breakthroughs in Statistics, Foundations and Basic Theory*, Springer Series in Statistics, pages 599–609. Springer, 10 1992.

[14] Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I Jordan. An introduction to MCMC for machine learning. *Machine learning*, 50(1):5 – 43, 2003.

[15] Christophe Andrieu and Johannes Thoms. A tutorial on adaptive MCMC. *Statistics and Computing*, 18(4):343–373, 12 2008.

[16] Jeffrey S. Rosenthal. Optimal Proposal Distributions and Adaptive MCMC. *Handbook of Markov Chain Monte Carlo*, 4, 2010.

[17] J. Peccoud and B. Ycart. Markovian modeling of gene-product synthesis. *Theoretical Population Biology*, 48:222–234, 10 1995.

[18] Ioannis Lestas, Johan Paulsson, Nicholas E. Ross, and Glenn Vinnicombe. Noise in Gene Regulatory Networks. *IEEE Transactions on Automatic Control*, 53(Special Issue):189–200, 2008.

[19] Aleksandra M. Walczak, Andrew Mugler, and Chris H. Wiggins. Computational Modeling of Signaling Networks. *Methods in Molecular Biology*, 880:273–322, 2012.

[20] Daniel T. Gillespie. Stochastic Simulation of Chemical Kinetics. *Annual Review of Physical Chemistry*, 58(1):35–55, 2007.

[21] Shawn C. Little, Mikhail Tikhonov, and Thomas Gregor. Precise Developmental Gene Expression Arises from Globally Stochastic Transcriptional Activity. *Cell*, 154(4):789–800, 08 2013.

[22] Nicholas C. Lammers, Vahe Galstyan, Armando Reimer, Sean A. Medin, Chris H. Wiggins, and Hernan G. Garcia. Multimodal transcriptional control of pattern formation in embryonic development. *Proceedings of the National Academy of Sciences*, 117(2):836–847, 2020.

[23] Robert B Sidje. Expokit: a software package for computing matrix exponentials. *ACM Transactions on Mathematical Software (TOMS)*, 24(1):130 – 156, 1998.

[24] Caroline Hoppe, Jonathan R. Bowles, Thomas G. Minchington, Catherine Sutcliffe, Priyanka Upadhyai, Magnus Rattray, and Hilary L. Ashe. Modulation of the Promoter Activation Rate Dictates the Transcriptional Response to Graded BMP Signaling Levels in the Drosophila Embryo. *Developmental Cell*, 54(6):727–741.e7, 2020.

[25] Ineke Brouwer, Emma Kerklingh, Fred van Leeuwen, and Tineke L. Lenstra. Dynamic epistasis analysis reveals how chromatin remodeling regulates transcriptional bursting. *Nature Structural & Molecular Biology*, 30(5):692–702, 2023.

[26] Benjamin T Donovan, Anh Huynh, David A Ball, Heta P Patel, Michael G Poirier, Daniel R Larson, Matthew L Ferguson, and Tineke L Lenstra. Live-cell imaging reveals the interplay between transcription factors, nucleosomes, and bursting. *The EMBO Journal*, 38, 6 2019.

[27] Joseph Rodriguez, Gang Ren, Christopher R Day, Keji Zhao, Carson C Chow, and Daniel R Larson. Intrinsic dynamics of a human gene reveal the basis of expression heterogeneity. *Cell*, 176:213–226.e18, 1 2019.

[28] Yihan Wan, Dimitrios G Anastasakis, Joseph Rodriguez, Murali Palangat, Prabhakar Gudla, George Zaki, Mayank Tandon, Gianluca Pegoraro, Carson C Chow, Markus Hafner, and Daniel R Larson. Dynamic imaging of nascent rna reveals general principles of transcription dynamics and stochastic splice site selection. *Cell*, 184:2878–2895.e20, 5 2021.

[29] Veronika A Herzog, Brian Reichholf, Tobias Neumann, Philipp Rescheneder, Pooja Bhat, Thomas R Burkard, Wiebke Wlotzka, Arndt von Haeseler, Johannes Zuber, and Stefan L Ameres. Thiol-linked alkylation of RNA to assess expression dynamics. *Nature Methods*, 14(12):1198–1204, 2017.

[30] Pawel Trzaskoma, SeolKyoung Jung, Aleksandra Pekowska, Christopher H. Bohrer, Xiang Wang, Faiza Naz, Stefania Dell'Orso, Wendy D. Dubois, Ana Olivera, Supriya V. Vartak, Yongbing Zhao, Subhashree Nayak, Andrew Overmiller, Maria I. Morasso, Vittorio Sartorelli, Daniel R. Larson, Carson C. Chow, Rafael Casellas, and John J. O'Shea. 3D chromatin architecture, BRD4, and Mediator have distinct roles in regulating genome-wide transcriptional bursting and gene network. *Science Advances*, 10(32):eadl4893, 2024.

[31] Wenhao Tang, Andreas Christ Sølvsten Jørgensen, Samuel Marguerat, Philipp Thomas, and Vahid Shahrezaei. Modelling capture efficiency of single-cell RNA-sequencing data improves inference of transcriptome-wide burst kinetics. *Bioinformatics*, 39(7):btad395, 2023.

[32] Ramon Grima and Pierre-Marie Esmenjaud. Quantifying and correcting bias in transcriptional parameter inference from single-cell data. *Biophysical Journal*, 123(1):4–30, 2024.