

RESEARCH ARTICLE

Real-time single-cell characterization of the eukaryotic transcription cycle reveals correlations between RNA initiation, elongation, and cleavage

Jonathan Liu¹, Donald Hansen², Elizabeth Eck³, Yang Joon Kim³, Meghan Turner³, Simon Alamos⁴, Hernan G. Garcia^{1,3,5,6*}

1 Department of Physics, University of California at Berkeley, Berkeley, California, United States of America, **2** Institute of Pharmacy and Molecular Biotechnology, University of Heidelberg, Heidelberg, Germany, **3** Biophysics Graduate Group, University of California at Berkeley, Berkeley, California, United States of America, **4** Department of Plant and Microbial Biology, University of California at Berkeley, Berkeley, California, United States of America, **5** Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California, United States of America, **6** Institute for Quantitative Biosciences-QB3, University of California at Berkeley, Berkeley, California, United States of America

* hggarcia@berkeley.edu



OPEN ACCESS

Citation: Liu J, Hansen D, Eck E, Kim YJ, Turner M, Alamos S, et al. (2021) Real-time single-cell characterization of the eukaryotic transcription cycle reveals correlations between RNA initiation, elongation, and cleavage. *PLoS Comput Biol* 17(5): e1008999. <https://doi.org/10.1371/journal.pcbi.1008999>

Editor: James R. Faeder, University of Pittsburgh, UNITED STATES

Received: October 28, 2020

Accepted: April 23, 2021

Published: May 18, 2021

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1008999>

Copyright: © 2021 Liu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All software is available on GitHub at <https://github.com/>

Abstract

The eukaryotic transcription cycle consists of three main steps: initiation, elongation, and cleavage of the nascent RNA transcript. Although each of these steps can be regulated as well as coupled with each other, their *in vivo* dissection has remained challenging because available experimental readouts lack sufficient spatiotemporal resolution to separate the contributions from each of these steps. Here, we describe a novel application of Bayesian inference techniques to simultaneously infer the effective parameters of the transcription cycle in real time and at the single-cell level using a two-color MS2/PP7 reporter gene and the developing fruit fly embryo as a case study. Our method enables detailed investigations into cell-to-cell variability in transcription-cycle parameters as well as single-cell correlations between these parameters. These measurements, combined with theoretical modeling, suggest a substantial variability in the elongation rate of individual RNA polymerase molecules. We further illustrate the power of this technique by uncovering a novel mechanistic connection between RNA polymerase density and nascent RNA cleavage efficiency. Thus, our approach makes it possible to shed light on the regulatory mechanisms in play during each step of the transcription cycle in individual, living cells at high spatiotemporal resolution.

Author summary

Live cell imaging using fluorescence microscopy provides an exciting way to visualize the transcription cycle in living organisms with great amounts of precision. However, the output of these technologies is often complex and can be hard to interpret. We have

GarciaLab/TranscriptionCycleInference Data is attached as a supplementary zip file [S1 Data](#).

Funding: This work was supported by the Burroughs Wellcome Fund Career Award at the Scientific Interface (<https://www.bwfund.org/grant-programs/interfaces-science/career-awards-scientific-interface>), the Sloan Research Foundation (<https://sloan.org/>), the Human Frontiers Science Program (<https://www.hfsp.org/>), the Searle Scholars Program (<https://www.searlescholars.net/>), the Shurl and Kay Curci Foundation (<http://thecurcifoundation.org/>), the Hellman Foundation (<https://www.hellmanfoundation.org/>), the NIH Director's New Innovator Award (DP2 OD024541-01, <https://commonfund.nih.gov/newinnovator>), and an NSF CAREER Award (1652236, https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503214) (HGG), an NSF GRFP (DGE 1752814, <https://www.nsfgrfp.org/>) (EE, MT), a UC Berkeley Chancellor's Fellowship (EE, <https://grad.berkeley.edu/admissions/apply/fellowships-entering/>), a KFAS scholarship (YJK, <https://www.kfas.or.kr/ScholarShip/ScholarShip0201.aspx?pCulture=en>), and an 430 DoD NDSEG graduate fellowship (JL, <https://ndseg.sysplus.com/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

developed a computational framework for analyzing the transcription cycle that quantifies rates of RNA initiation, elongation, and cleavage, given input datasets from live cell imaging. Using the developing fruit fly embryo as a case study, we demonstrate that our methodology can quantitatively describe the whole transcription cycle at single-cell resolution. These results allow us to investigate a plethora of avenues, from couplings between different aspects of the transcription cycle at the single-cell level to comparisons with theoretical predictions of distributions of elongation rates across cells. We envision our methodology to provide a unified computational framework for the analysis of transcriptional data obtained from live cell imaging.

Introduction

The eukaryotic transcription cycle consists of three main steps: initiation, elongation, and cleavage of the nascent RNA transcript (Fig 1A; [1]). Crucially, each of these three steps can be controlled to regulate transcriptional activity. For example, binding of transcription factors to enhancers dictates initiation rates [2], modulation of elongation rates helps determine splicing efficiency [3], and regulation of cleavage controls aspects of 3' processing such as alternative polyadenylation [4].

The steps of the transcription cycle can be coupled with each other. For example, elongation rates contribute to determining mRNA cleavage and RNA polymerase (RNAP) termination efficiency [5–8], and functional linkages have been demonstrated between transcription initiation and termination [9, 10]. Nonetheless, initiation, elongation, and transcript cleavage have largely been studied in isolation. In order to dissect the entire transcription cycle, it is necessary to develop a holistic approach that makes it possible to understand how the regulation of each step dictates mRNA production and to unearth potential couplings among these steps.

To date, the processes of the transcription cycle have mostly been studied in detail using *in vitro* approaches [11, 12] or genome-wide measurements that require the fixation of cellular material and lack the spatiotemporal resolution to uncover how the regulation of the transcription cycle unfolds in real time [13–18]. Only recently has it become possible to dissect these processes in living cells and in their full dynamical complexity using tools such as MS2 or PP7 to fluorescently label nascent transcripts at single-cell resolution [19–22]. These technological advances have yielded insights into, for example, intrinsic transcriptional noise in yeast [23], kinetic splicing effects in human cells [24], elongation rates in *Drosophila melanogaster* [25, 26], and transcriptional bursting in mammalian cells [27], *Dictyostelium* [28–30], fruit flies [25, 31–35] and *Caenorhabditis elegans* [36].

Despite the great promise of MS2 and PP7, using these techniques to comprehensively analyze the transcription cycle is hindered by the fact that the signal from these *in vivo* RNA-labeling technologies convolves contributions from all aspects of the cycle. Specifically, the fluorescence signal from nascent RNA transcripts persists throughout the entire cycle of transcript initiation, elongation, and cleavage; further, a single gene can carry many tens of transcripts. Thus, at any given point, an MS2 or PP7 signal reports on the contributions of transcripts in various stages of the transcription cycle [37]. Precisely interpreting an MS2 or PP7 signal therefore demands an integrated approach that accounts for this complexity.

Here, we present a method for analyzing live-imaging data from the MS2 and PP7 techniques in order to dynamically characterize the steps—initiation, elongation, and cleavage—of the full transcription cycle at single-cell resolution. While the transcription cycle is certainly more nuanced and can include additional effects such as sequence-dependent pausing [38], we

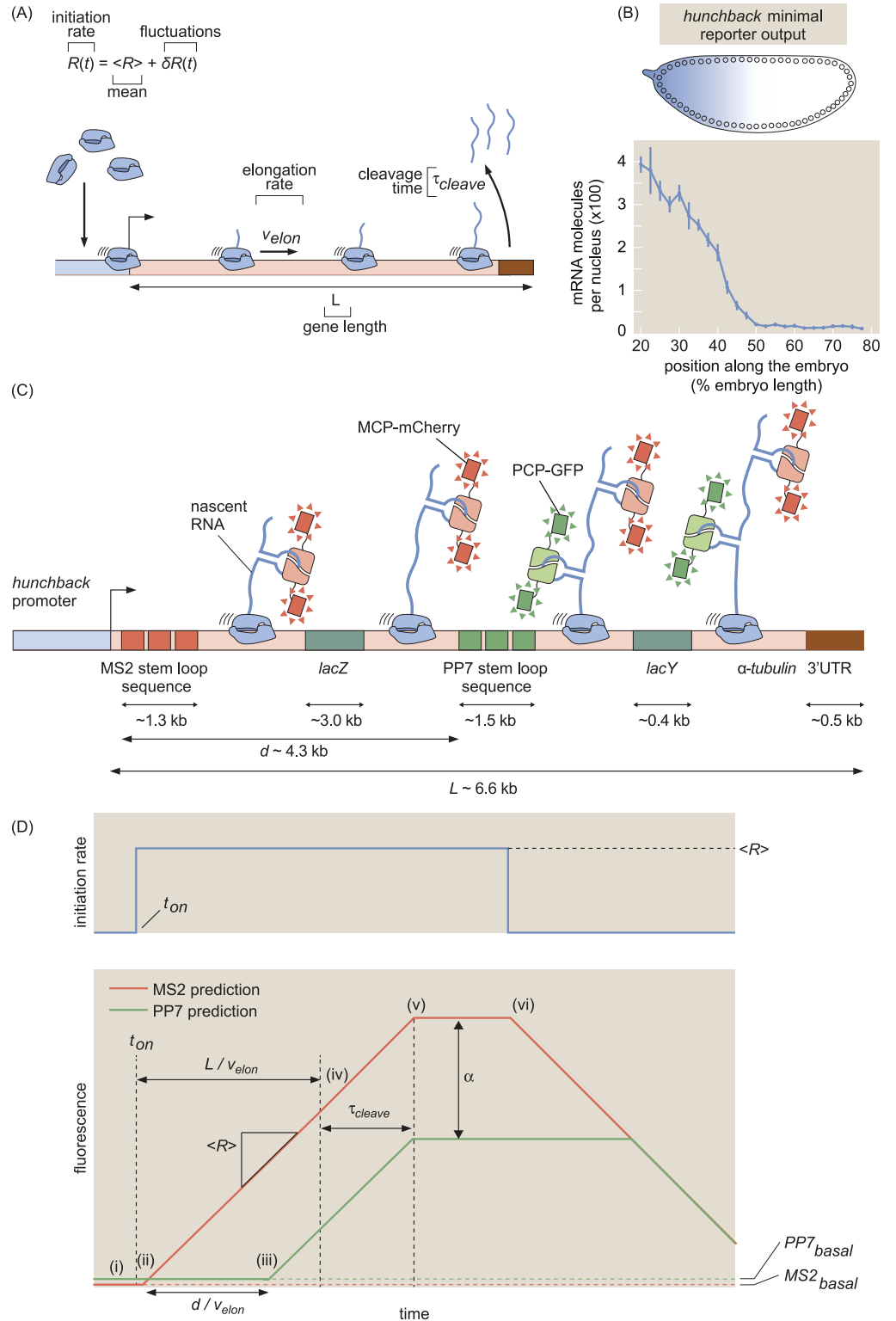


Fig 1. Theoretical model of the transcription cycle and experimental setup. (A) Simple model of the transcription cycle, incorporating nascent RNA initiation, elongation, and cleavage. (B) The reporter construct, which is driven by the *hunchback* P2 minimal enhancer and promoter, is expressed in a step-like fashion along the anterior-posterior axis of the fruit fly embryo. (C) Transcription of the stem loops results in fluorescent puncta with the 5' mCherry signal appearing before the signal from 3' GFP. Only one stem loop per fluorophore is shown for clarity, but the actual construct contains 24 repeats of each stem loop. (D, top) Relationship between fluorescence trace profiles and model parameters for an initiation rate consisting of a pulse of constant magnitude $\langle R \rangle$. (D, bottom, i) At first, the zero

initiation rate results in no fluorescence other than the basal levels $MS2_{basal}$ and $PP7_{basal}$ (red and green dashed lines). (ii) When initiation commences at time t_{on} , RNAP molecules load onto the promoter and elongation of nascent transcripts occurs, resulting in a constant increase in the MS2 signal (red curve). (iii) After time $\frac{d}{v_{elon}}$, the first RNAP molecules reach the PP7 stem loops and the PP7 signal also increases at a constant rate. (iv) After time $\frac{L}{v_{elon}}$, the first RNAP molecules reach the end of the gene, and (v) after the cleavage time τ_{cleave} , these first nascent transcripts are cleaved. The subsequent loss of fluorescence is balanced by the addition of new nascent transcripts, resulting in a plateauing of the signal. (vi) Once the initiation rate shuts off, no new RNAP molecules are added and both fluorescence signals will start to decrease due to cleavage of the nascent transcripts still on the gene. Because elongation continues after initiation has ceased, the 5' MS2 signal begins decreasing before the 3' PP7 signal. The MS2 and PP7 fluorescent signals are rescaled to be in the same arbitrary units with the calibration factor α . (Data in (B) adapted from [25] with the line representing the mean and error bars representing the standard error across 24 embryos).

<https://doi.org/10.1371/journal.pcbi.1008999.g001>

view the quantification of these effective parameters as a key initial step for testing theoretical models. This method combines a dual-color MS2/PP7 fluorescent reporter [23, 24, 26] with Bayesian statistical inference techniques and quantitative modeling. As a proof of principle, we applied this analysis to the transcription cycle of a *hunchback* reporter gene in the developing embryo of the fruit fly *Drosophila melanogaster*. We validate our approach by comparing our inferred average initiation and elongation rates with previously reported results.

Crucially, our analysis also delivered novel single-cell statistics of the whole transcription cycle that were previously unmeasurable using genome-wide approaches, making it possible to generate distributions of parameter values necessary for investigations that go beyond simple population-averaged analyses [39–60]. We show that, by taking advantage of time-resolved data, our inference is able to filter out uncorrelated noise, such as that originating from random measurement error, in these distributions and retain sources of correlated variability (such as biological and systematic noise). By combining these statistics with theoretical models, we revealed substantial variability in RNAP stepping rates between individual molecules, demonstrating the utility of our approach for testing hypotheses of the molecular mechanisms underlying the transcription cycle and its regulation.

This unified analysis enabled us to investigate couplings between the various transcription cycle parameters at the single-cell level, whereby we discovered a surprising correlation of cleavage rates with nascent transcript densities. These discoveries illustrate the potential of our method to sharpen hypotheses of the molecular processes underlying the regulation of the transcription cycle and to provide a framework for testing those hypotheses.

Results

To quantitatively dissect the transcription cycle in its entirety from live imaging data, we developed a simple model (Fig 1A) in which RNAP molecules are loaded at the promoter of a gene of total length L with a time-dependent loading rate $R(t)$. For simplicity, we assume that each individual RNAP molecule behaves identically and independently: there are no interactions between molecules. While this assumption is a crude simplification, it nevertheless allows us to infer effective average transcription cycle parameters.

We parameterize this $R(t)$ as the sum of a constant term $\langle R \rangle$ that represents the mean, or time-averaged, rate of initiation, and a small temporal fluctuation term given by $\delta R(t)$ such that $R(t) = \langle R \rangle + \delta R(t)$. This mean-field parameterization is motivated by the fact that many genes are well approximated by constant rates of initiation [25, 31, 35, 61]. The fluctuation term $\delta R(t)$ allows for slight time-dependent deviations from the mean initiation rate. As a result, this term makes it possible to account for time-dependent behavior that can occur over the course of a cell cycle once the promoter has turned on. After initiation, each RNAP molecule traverses the gene at a constant, uniform elongation rate v_{elon} . Upon reaching the end of

the gene, there follows a deterministic cleavage time, τ_{cleave} , after which the nascent transcript is cleaved.

We do not consider RNAP molecules that do not productively initiate transcription [62] or that are paused at the promoter [16], as they will provide no experimental readout. Based on experimental evidence [25], we assume that these RNAP molecules are processive, such that each molecule successfully completes transcription, with no loss of RNAP molecules before the end of the gene (see Section H in [S1 File](#) for a validation of this hypothesis).

Dual-color reporter for dissecting the transcription cycle

As a case study, we investigated the transcription cycle of early embryos of the fruit fly *D. melanogaster*. Specifically, we focused on the P2 minimal enhancer and promoter of the *hunchback* gene during the 14th nuclear cycle of development; the gene is transcribed in a step-like pattern along the anterior-posterior axis of the embryo with a 26-fold modulation in overall mRNA count between the anterior and posterior ends ([Fig 1B](#) [25, 63–65]). As a result, the fly embryo provides a natural modulation in mRNA production rates, with the position along the anterior-posterior axis serving as a proxy for mRNA output.

To visualize the transcription cycle, we utilized the MS2 and PP7 systems for live imaging of nascent RNA production [25, 31, 33]. Using a two-color reporter construct similar to that reported in [23], [24], and [26], we placed the MS2 and PP7 stem loop sequences in the 5' and 3' ends, respectively, of a transgenic *hunchback* reporter gene ([Fig 1C](#); see [S1 Fig](#) for more construct details). The *lacZ* sequence and a portion of the *lacY* sequence from *Escherichia coli* were placed as a neutral spacer [66] between the MS2 and PP7 stem loops.

As an individual RNAP molecule transcribes through a set of MS2/PP7 stem loops, constitutively expressed MCP-mCherry and PCP-GFP fusion proteins bind their respective stem loops, resulting in sites of nascent transcript formation that appear as fluorescent puncta under a laser-scanning confocal microscope ([Fig 2A](#) and [S1 Video](#)). The fluorescent signals did not exhibit noticeable photobleaching (Section B in [S1 File](#) and [S2 Fig](#)). Since *hunchback* becomes transcriptionally active at the start of the nuclear cycle before slowly decaying into a transcriptionally silent state [25, 67, 68], we restrict our analysis to the initial 18 minute window after mitosis where the promoter remains active.

The intensity of the puncta in each color channel is linearly related to the number of actively transcribing RNAP molecules that have elongated past the location of the associated stem loop sequence [25], albeit with different arbitrary fluorescence units. After reaching the end of the gene, which contains the 3'UTR of the α -tubulin gene [66], the nascent RNA transcript undergoes cleavage. Because the characteristic timescale of mRNA diffusion is about two order of magnitudes faster than the time resolution of our experiment, we approximate the cleavage of a single transcript as resulting in the instantaneous loss of its associated fluorescent signal in both channels (Section C in [S1 File](#); [69]). We included a few additional parameters in our model to make it compatible with this experimental data: a calibration factor α between mCherry and eGFP intensities, a time of transcription onset t_{on} after mitosis at which the promoter switches on, and basal levels of fluorescence in each channel $MS2_{basal}$ and $PP7_{basal}$ (see Section A in [S1 File](#) for more details). The qualitative relationship between the model parameters and the fluorescence data is described in [Fig 1D](#), which considers the case of a pulse of constant initiation rate.

Transcription cycle parameter inference using Markov Chain Monte Carlo

We developed a statistical framework to estimate transcription-cycle parameters ([Fig 1A](#)) from fluorescence signals. Time traces of mCherry and eGFP fluorescence intensity are extracted

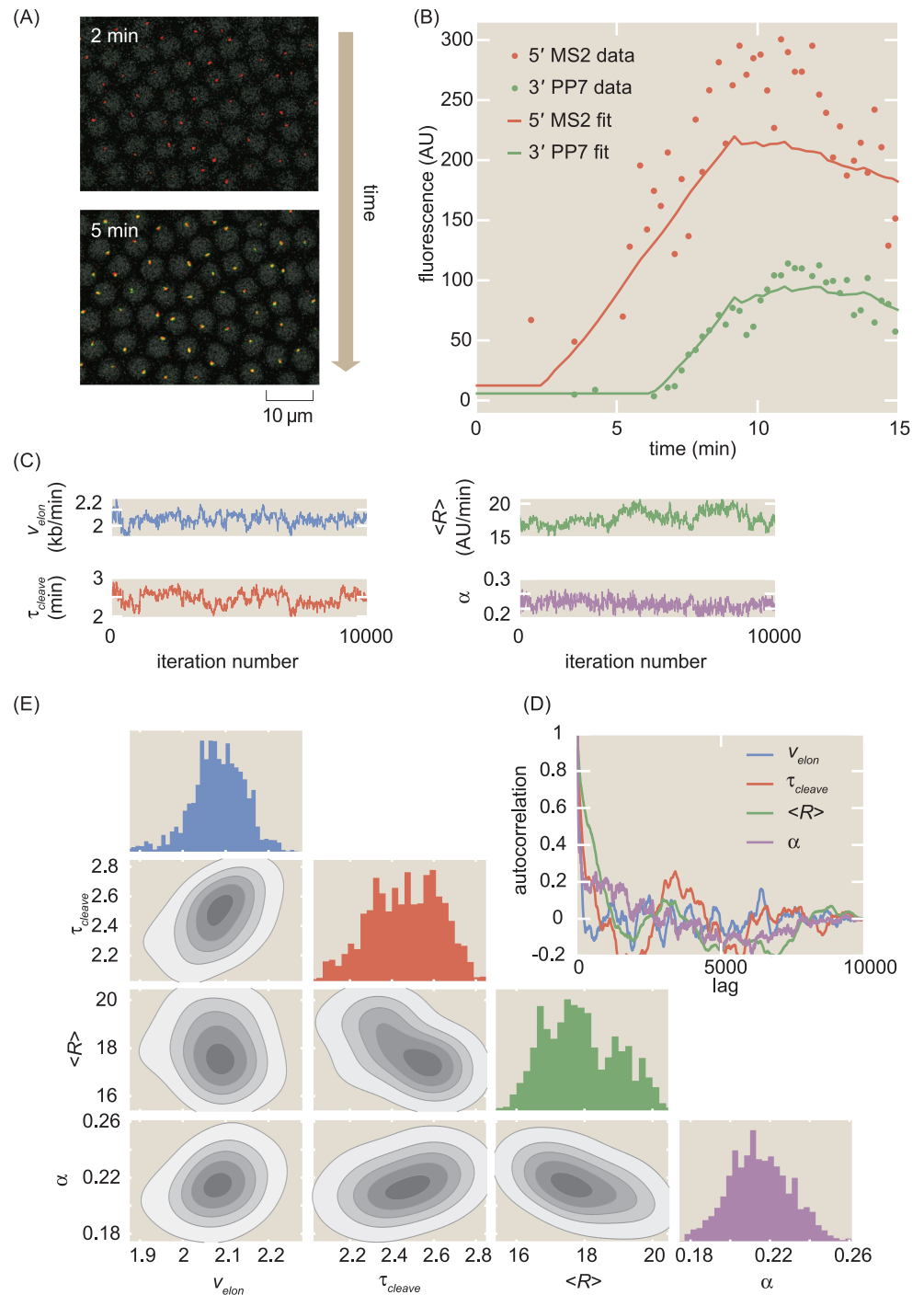


Fig 2. MCMC inference procedure. (A) Snapshots of confocal microscopy data over time, with MS2-mCherry (red) and PP7-eGFP (green) puncta reporting on transcription activity. Gray circles correspond to iRFP-labeled histones whose fluorescence is used as a fiduciary marker for cell nucleus segmentation (see [Methods and materials](#) for details). (B) Sample single-cell MS2 and PP7 fluorescence (points) together with best-fits of the model using MCMC inference (curves). (C) Raw MCMC inference chains for the elongation rate v_{elon} , cleavage time τ_{cleave} , mean initiation rate $\langle R \rangle$, and calibration factor α for the inference results of a sample single cell. (D) Auto-correlation function for the raw chains in (A) as a function of lag (i.e. inference sample number). (E) Corner plot of the raw chains shown in (C).

<https://doi.org/10.1371/journal.pcbi.1008999.g002>

from microscopy data such as shown in Fig 2A and S1 Video to produce a dual-signal readout of nascent RNA transcription at single-cell resolution (Fig 2B, data points; see Methods and materials for details). To extract quantitative insights from the observed fluorescence data, we used the established Bayesian inference technique of Markov Chain Monte Carlo (MCMC) [70] to infer the effective parameter values in our simple model of transcription: the calibration factor between mCherry and eGFP intensities α , the time-dependent transcription initiation rate, separated into the constant term $\langle R \rangle$ and fluctuations $\delta R(t)$, the elongation rate v_{elon} , the cleavage time τ_{cleave} , the time of transcription onset t_{on} , and the basal levels of fluorescence in each channel $MS2_{\text{basal}}$ and $PP7_{\text{basal}}$.

The details of the inference procedure are described in Section D in S1 File. Briefly, the inference was run separately for each single cell, yielding chains of sampled parameter values (Fig 2C). These resulting chains exhibited rapid mixing and rapidly decaying auto-correlation functions (Fig 2D), indicative of reliable fits. Corner plots of the fits indicated reasonable posterior distributions (Fig 2E).

From these single-cell fits, the mean value of each parameter's chain was retained for further analysis. The final dataset was produced by filtering with an automated procedure that relied on overall fit quality (Section F in S1 File and S4 Fig). This curation procedure did not introduce noticeable bias in the results (S4(G)–S4(I) Fig). A small minority of the rejected cells (S4(E) Fig) exhibited highly time-dependent behavior reminiscent of transcriptional bursting [71], which lies outside the scope of our model and is explored more in the Discussion. A sample fit is shown in Fig 2B. To aggregate the results, we constructed a distribution from the inferred parameter from each single-cell. Intra-embryo variability between single cells was greater than inter-embryo variability (Section I in S1 File and S6 Fig). As a result, unless stated otherwise, all statistics reported here were aggregated across 355 single cells combined between 7 embryos, and all shaded errors reflect the standard error of the mean.

MCMC successfully infers calibration between eGFP and mCherry intensities

Due to the fact that the MS2 and PP7 stem loop sequences were associated with mCherry and eGFP fluorescent proteins, respectively, the two experimental fluorescent signals possessed different arbitrary fluorescent units, related by the scaling factor α given by

$$\alpha = \frac{F_{MS2}}{F_{PP7}}, \quad (1)$$

where F_{MS2} and F_{PP7} are the fluorescence values generated by a fully transcribed set of MS2 and PP7 stem loops, respectively. Although α has units of AU_{MS2}/AU_{PP7} , we will express α without units in the interest of clarity of notation.

We inferred single-cell values of α using the inference methodology. As shown in the blue histogram in Fig 3A, our inferred values of α possessed a mean of 0.145 ± 0.004 (SEM) and a standard deviation of 0.068.

As an independent validation, we measured α by using another two-color reporter, consisting of 24 alternating, rather than sequential, MS2 and PP7 loops [72–74] inserted at the 5' end of our reporter construct (Fig 3B). Thus, this reporter had a total of 48 stem loops, with 24 each of MS2 and PP7.

Fig 3C shows a representative trace of a single spot containing our calibration construct (see S2 Video for full movie). For each time point, the mCherry fluorescence in all measured single-cell traces was plotted against the corresponding eGFP fluorescence (Fig 3D, yellow points). The mean α was then calculated by fitting the resulting scatter plot to a line going

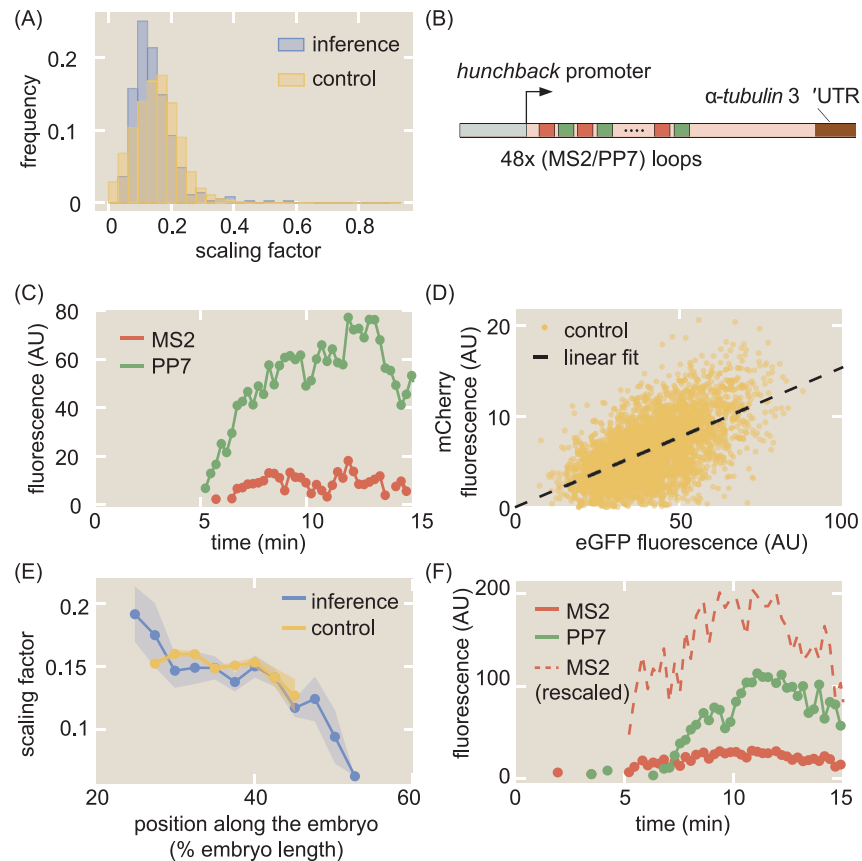


Fig 3. Calibration of MS2 and PP7 fluorescence signals. (A) Histogram of inferred values of α at the single-cell level from inference (blue), along with histogram of α values from the control experiment (yellow). (B) Schematic of construct used to measure the calibration factor α using 24 interlaced MS2/PP7 loops (48 loops in total). (C) Sample single-cell MS2 (red) and PP7 (green) traces from this control experiment. (D) Scatter plot of MS2 and PP7 fluorescence values for each time point (yellow) along with linear best fit (black) resulting in $\alpha = 0.154 \pm 0.001$. (E) Position-dependent mean value of α in both the inference (blue) and the control experiment (yellow). (F) Representative raw and rescaled MS2 and PP7 traces for a sample single cell in the inference data set. (A, D, E, data were collected for 314 cells across 4 embryos for the interlaced reporter, and for 355 cells across 7 embryos for the reporter with MS2 on the 5' and PP7 on the 3' of the gene (Fig 1C); shaded regions in (E) reflect standard error of the mean. Measurement conditions for both experiments are described in Methods and materials).

<https://doi.org/10.1371/journal.pcbi.1008999.g003>

through the origin (Fig 3D, black line). The best-fit slope yielded the experimentally calculated value of $\alpha = 0.154 \pm 0.001$ (SEM). A distribution for α was also constructed by dividing the mCherry fluorescence by the corresponding eGFP fluorescence for each datapoint in Fig 3D, yielding the histogram in Fig 3A (yellow), which possessed a standard deviation of 0.073. Our independent calibration agreed with our inference, thus validating the inferred values of α .

Interestingly, binning the cells by position along the embryo revealed a slight position dependence in the scaling factor. As shown in Fig 3E, both the directly measured and inferred α displayed higher values in the anterior, about 0.15, and lower values in the posterior, about 0.1. The fact that this position dependence is observed in both in the calibration experiments and inference suggests that this spatial modulation in the value of α is not an artifact of the constructs or our analysis, but a real feature of the system. We speculate that this spatial dependence could stem from differential availability of MCP-mCherry and PCP-GFP along the embryo, leading to a modulation in the maximum occupancy of the MS2 stem loops versus the PP7 stem loops [75].

Regardless, our data demonstrate that the inferred and calibrated α can be used interchangeably, obviating the need for the control. Thus, the MS2 signals for each single cell could be rescaled to the same units as the PP7 signal (Fig 3F) within a single experiment, greatly increasing the power of the inference methodology. All plots, unless otherwise stated, reflect these rescaled values using the overall mean value of $\alpha = 0.145$ obtained from the inference.

Inference of single-cell initiation rates recapitulates and improves on previous measurements

After validating the accuracy of our inference method in inferring transcription initiation, elongation, and cleavage dynamics using simulated data (Section G in S1 File and S5 Fig), we inferred these transcriptional parameters for the *hunchback* reporter gene as a function of the position along the anterior-posterior axis of the embryo. The suite of quantitative measurements on the transcription cycle produced by the aggregated inference results is shown in Fig 4A, 4C, 4E and 4F. Full distributions of these parameters can be found in S7 Fig.

Control of initiation rates is one of the predominant, and as a result most well-studied, strategies for gene regulation [2, 13, 76]. Thus, comparing our inferred initiation rates with previously established results comprised a crucial benchmark for our methodology. Our inferred values of the mean initiation rate $\langle R \rangle$ exhibited a step-like pattern along the anterior-posterior axis of the embryo, qualitatively reproducing the known *hunchback* expression profile (Fig 4A, blue). As a point of comparison, we also examined the mean initiation rate measured by [25], which was obtained by manually fitting a trapezoid (Fig 1D) to the average MS2 signal (Fig 4A, black). The quantitative agreement between these two dissimilar analysis methodologies demonstrates that our inference method can reliably extract the average rate of transcription initiation across cells.

Measurements of cell-to-cell variability in transcription initiation rate have uncovered, for example, the existence of transcriptional bursting and mechanisms underlying the establishment of precise developmental boundaries [39, 40, 45, 47, 48, 55, 57]. Yet, to date, these studies have mostly employed techniques such as single-molecule FISH to count the number of nascent transcripts on a gene or the number of cytoplasmic mRNA molecules [39–41, 43, 45, 48–52, 56–58, 77–84]. In principle, these techniques do not report on the variability in transcription initiation alone; they convolve this measurement with variability in other steps of the transcription cycle [76, 81].

Our inference approach isolates the transcription initiation rate from the remaining steps of the transcription cycle at the single-cell level, making it possible to calculate, for example, the coefficient of variation (CV; standard deviation divided by the mean) of the mean rate of initiation. Our results yielded values for the CV along the embryo that were fairly uniform, with a maximum value of around 40% (Fig 4F, blue). This value is roughly comparable to that obtained for *hunchback* using single-molecule FISH [45, 50, 57].

One of the challenges in measuring CV values, however, is that informative biological variability is often convolved with undesired experimental noise, such as experimental measurement noise inherent to fluorescence microscopy. In general, this experimental noise can contain both random, uncorrelated components as well as systematic components, the latter of which combines with actual biological variability to form overall correlated noise. Although we currently cannot entirely separate biological variability from experimental noise with our data and inference method, a strategy for at least separating uncorrelated noise from correlated noise was recently implemented in the context of snapshot-based fluorescent data [57]. By utilizing a dual-color measurement of the same biological signal, one can separate the total

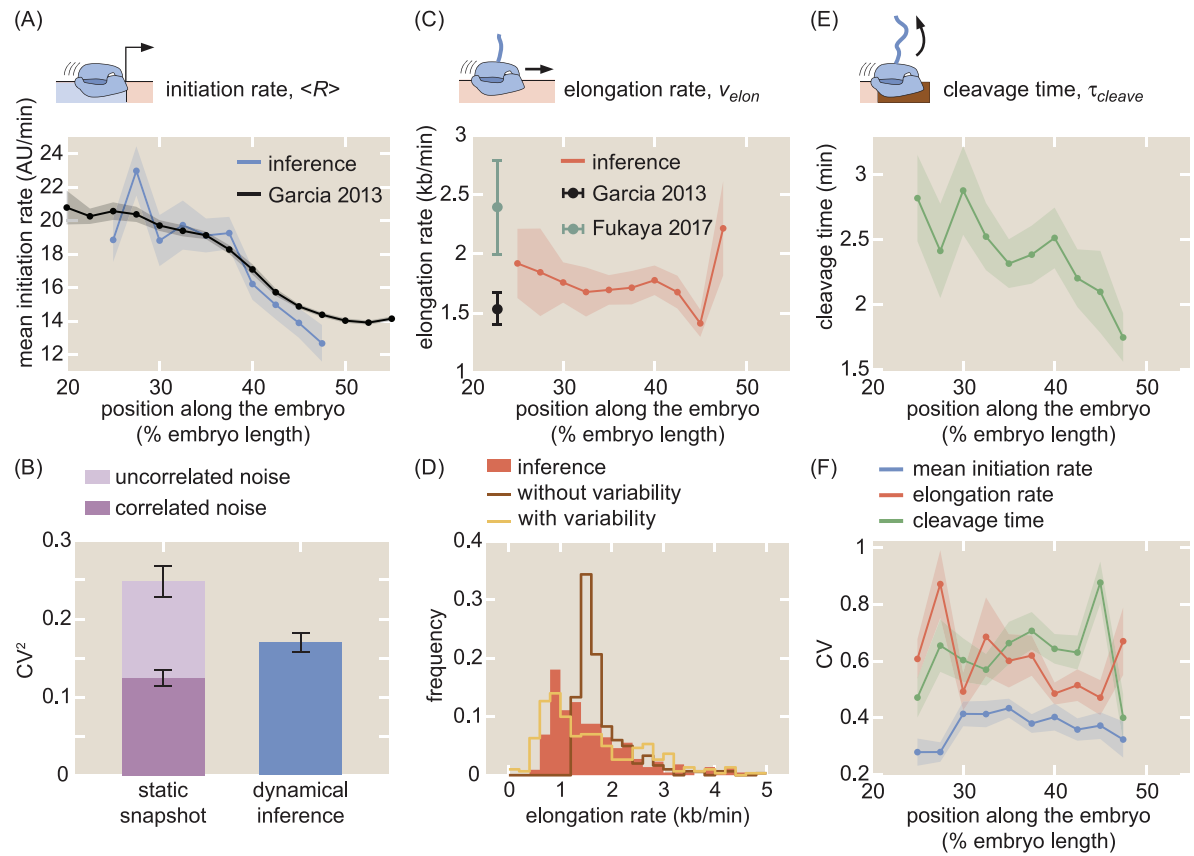


Fig 4. Inferred transcription-cycle parameters. (A) Mean inferred transcription initiation rate as a function of embryo position (blue), along with rescaled previously reported results (black, [25]). (B) Comparison of the squared CV of the mean initiation rate inferred using our approach (blue) or obtained from examining the fluorescence of transcription spots in a single snapshot (light plus dark purple). While snapshots captured a significant amount of uncorrelated noise (light purple), our inference accounts mostly for correlated noise (compare blue and dark purple). See Section K in S1 File and S8 Fig for details. (C) Inferred elongation rate as a function of embryo position (red), along with previously reported results (black, [25]; teal, [26]). (D) Distribution of inferred single-cell elongation rates in the anterior 40% of embryo (red), along with best fit to mean and standard deviation using single-molecule simulations with and without RNAP-to-RNAP variability (gold and brown, respectively, see Section M in S1 File for details). (E) Inferred cleavage time as a function of embryo position. (F) CV of the mean initiation rate (blue), elongation rate (red), and cleavage time (green) as a function of embryo position. (A, C, E, shaded error reflects standard error of the mean across 355 nuclei in 7 embryos, or of previously reported mean results; B, F, shaded error or black error bars represent bootstrapped standard errors of the CV or CV² for 100 bootstrap samples each; C, error bars reflect standard error of the mean for [25] and lower (25%) and upper (75%) quintiles of the full distribution from [26]).

<https://doi.org/10.1371/journal.pcbi.1008999.g004>

variability in a dataset into uncorrelated measurement noise and correlated noise, which includes components such as true biological variability and systematic measurement error.

Building on this strategy, we first took a single snapshot from our live-imaging data and calculated the total squared CV of the fluorescence of spots at a single time point (Fig 4B, dark plus light purple). Compared to the squared CV from the inferred mean initiation rate (Fig 4B, blue), the squared CV from the snapshot was larger by about 0.1, suggesting that the inference method reported on a somewhat lower level of overall variability.

To investigate this disparity in measured variability further, we then rewrote the squared CV from the snapshot approach as the sum of uncorrelated and correlated components

$$CV_{total}^2 = CV_{uncorrelated}^2 + CV_{correlated}^2 \quad (2)$$

The magnitudes of each noise component were estimated by using the data from the interlaced reporter introduced in Fig 3B. To do so, we utilized the fact that, in principle, the mCherry and GFP signals from this experiment reflected the same underlying biological process, and assumed that deviations between the two signals were a result of uncorrelated measurement noise. Thus, we could apply the two-color formalism introduced in [85] to calculate the uncorrelated and correlated noise components from snapshots taken from the interlaced reporter construct (see Section K in S1 File and S8 Fig for more details).

The bar graph shown in Fig 4B shows that, once the uncorrelated noise (light purple) is subtracted from the total noise of our snapshot-based measurement, the remaining correlated variability (dark purple), which includes the biological variability, is slightly lower than the variability of our inference results (blue). Thus, our inference mostly captures correlated variability and filters out the bulk of the uncorrelated noise, similarly to techniques such as single-molecule FISH [57] but with the added advantage of also being able to resolve temporal information. Because such fixed tissue techniques ultimately provide static measurements that con-volve signals from transcription initiation with those of elongation and cleavage, it is important to note that this is a qualitative comparison between the ability of fixed-tissue and live-imaging to separate correlated and uncorrelated variability. Thus, our results further validate our approach and demonstrate its capability to capture measures of cell-to-cell variability in the transcription cycle with high precision.

Elongation rate inference reveals single-molecule variability in RNAP stepping rates

Next, we investigated the ability of our inference approach to report on the elongation rate v_{elon} . Nascent RNA elongation plays a prominent role in gene regulation, for example, in dosage compensation in *Drosophila* embryos [86], alternative splicing in human cells [3, 87], and gene expression in plants [88]. Our method inferred an elongation rate v_{elon} that was relatively constant along the embryo (Fig 4C), lending support to previous reports indicating a lack of regulatory control of the elongation rate in the early fly embryo [26]. We measured a mean elongation rate of 1.72 ± 0.05 kb/min (SEM; $n = 355$), consistent with previous measurements of the fly embryo (Fig 4C, black and teal; [25, 26]), as well as with measurements from other techniques and model organisms, which range from about 1 kb/min to upwards of 4 kb/min [20, 23, 24, 27, 62, 76, 77, 89–92]. In addition, the CV of the elongation rate was roughly uniform across embryo position (Fig 4E, red).

Like cell-to-cell variability in transcription initiation, single-cell distributions of elongation rates can provide crucial insights into, for example, promoter-proximal pausing [54], traffic jams [93, 94], transcriptional bursting [95, 96], and noise propagation [59]. While genome-wide approaches have had huge success in measuring mean properties of elongation [16, 97], they remain unable to resolve single-cell distributions of elongation rates. We examined the statistics of single-cell elongation rates in the anterior 40% of the embryo, where the initiation rate was roughly constant, and inferred a broad distribution of elongation rates with a standard deviation of around 1 kb/min and a long tail extending to values upwards of 4 kb/min (Fig 4D, red). This large spread was consistent with observations of large cell-cell variability in elongation rates [76, 91] using a wide range of techniques, as well as with measurements from similar two-color live imaging experiments ([23, 26]; Section L in S1 File; S9 Fig).

To illustrate the resolving power of examining elongation rate distributions, we performed theoretical investigations of cell-to-cell variability in this transcription cycle parameter. Following [93], we considered a model where RNAP molecules stochastically step along a gene and cannot overlap or pass each other (Section M in S1 File). The model simulated MS2 and

PP7 fluorescences that were then run through the inference procedure, in order to account for the presence of inferential noise (Section G in [S1 File](#)).

First, we considered a scenario where the stepping rate of each RNAP molecule is identical. In this case, the sole driver of cell-to-cell variability is the combination of stochastic stepping behavior with traffic jamming due to steric hindrance of RNAP molecules. As shown in brown in [Fig 4D](#), this model cannot account for the wide distribution of observed single-cell elongation rates.

In contrast, by allowing for substantial variability in the elongation rate of individual RNAP molecules, the model can reproduce the empirical distribution of single-cell elongation rates. As shown in gold in [Fig 4D](#), the model can quantitatively approximate the inferred distribution within error ([S9\(D\) Fig](#)). This single-molecule variability is consistent with *in vitro* observations of substantial molecule-to-molecule variability in RNAP elongation rates [98, 99], thus demonstrating the ability of our approach to engage in the *in vivo* dissection of the transcription cycle at the single-molecule level.

Inference reveals functional dependencies of cleavage times

Finally, we inferred values of the cleavage time τ_{cleave} . Through processes such as alternative polyadenylation [4, 100] and promoter-terminator crosstalk [9, 10], events at the 3' end of a gene exert substantial influence over overall transcription levels [101]. Although many investigations of mRNA cleavage and RNAP termination have been carried out in fixed-tissue samples [102, 103], live-imaging studies with single-cell resolution of this important process remain sparse; some successes have been achieved in yeast and in mammalian cells [76]. We inferred a mean mRNA cleavage time in the range of 1.5–3 min ([Fig 4E](#)), consistent with values obtained from live imaging in yeast [22] and mammalian cells [24, 27, 62, 89]. Interestingly, as shown in [Fig 4E](#), the inferred mRNA cleavage time was dependent on anterior-posterior positioning along the embryo, with high values (~ 3 min) in the anterior end and lower values toward the posterior end (~ 1.5 min). While the reasons for this position dependence are unknown, such dependence could result from the presence of a spatial gradient of a molecular species that regulates cleavage. Importantly, such a modulation could not have been easily revealed using genome-wide approaches that, by necessity, average information across multiple cells.

The CV of the cleavage time slightly increased toward the posterior end of the embryo ([Fig 4F](#), green). Thus, although cleavage remains an understudied process compared to initiation and elongation, both theoretically and experimentally, these results provide the quantitative precision necessary to carry out such mechanistic analyses.

Uncovering single-cell mechanistic correlations between transcription cycle parameters

In addition to revealing trends in average quantities of the transcription cycle along the length of the embryo, the simultaneous nature of the inference afforded us the unprecedented ability to investigate single-cell correlations between transcription-cycle parameters. We used the Spearman rank correlation coefficient (ρ) as a non-parametric measure of inter-parameter correlations. The mean initiation rate and the cleavage time exhibited a negative correlation ($\rho = -0.52$, $p\text{-val} \approx 0$; [Fig 5A](#)). This negative correlation at the single-cell level should be contrasted with the positive relation between these magnitudes at the position-averaged level, where the mean initiation rate and cleavage time both increased in the anterior of the embryo ([Fig 4A and 4E](#)). Thus, our analysis unearthed a quantitative relationship that was obscured by a naive investigation of spatially averaged quantities, an approach often used in fixed [57] and live-

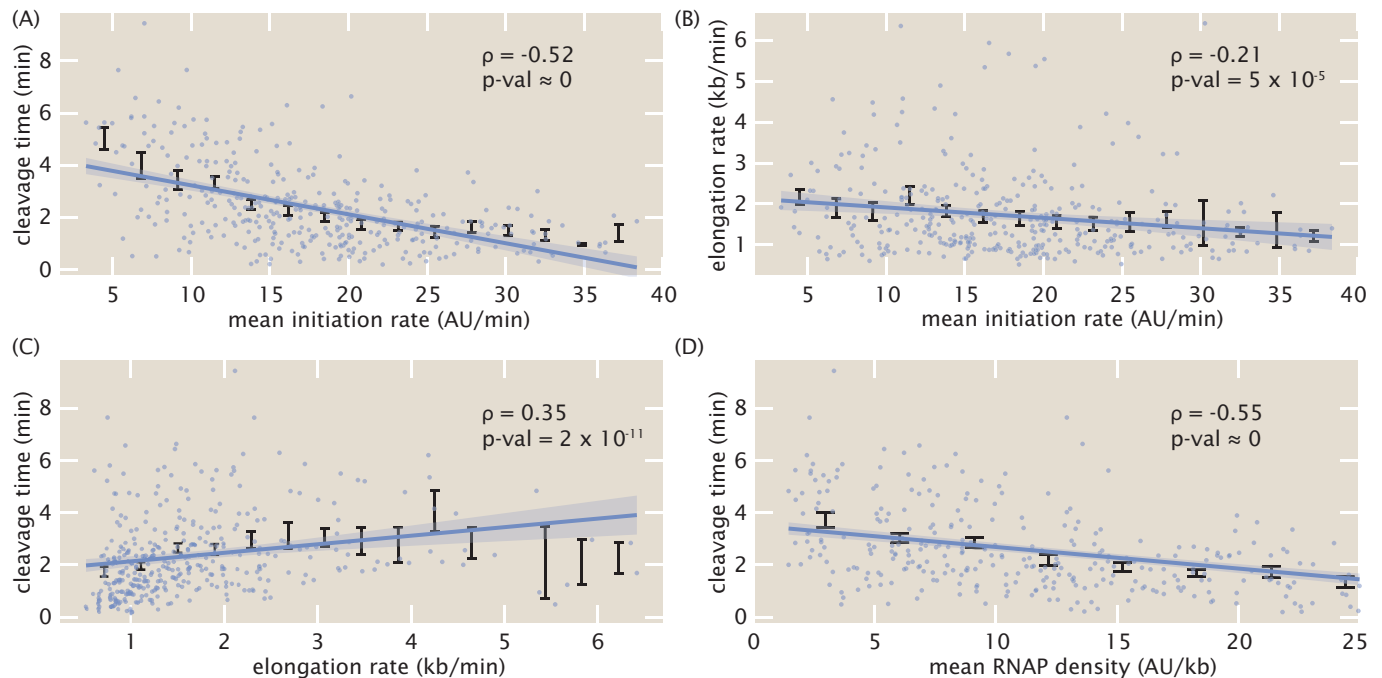


Fig 5. Single-cell correlations between transcription cycle parameters. Spearman rank correlation coefficients and associated p-values between (A) mean initiation rate and cleavage time, (B) mean initiation rate and elongation rate, (C) elongation rate and cleavage time, and (D) mean RNAP density and cleavage time. Blue points indicate single-cell values; black points and error bars indicate mean and SEM, respectively, binned across x-axis values. Lines and shaded regions indicate generalized linear model fit and 95% confidence interval, respectively, and are shown for ease of visualization (see [Methods and materials](#) for details).

<https://doi.org/10.1371/journal.pcbi.1008999.g005>

imaging [35] studies, as well as in genome-wide investigations [104, 105]. We also detected a small negative correlation ($\rho = -0.21$, $p\text{-val} = 5 \times 10^{-5}$) between elongation rates and mean initiation rates (Fig 5B). Finally, we detected a small positive correlation ($\rho = 0.35$, $p\text{-val} = 2 \times 10^{-11}$) between cleavage times and elongation rates (Fig 5C). These results are consistent with prior studies implicating elongation rates in 3' processes such as splicing and alternative polyadenylation: slower elongation rates increased cleavage efficiency [3, 5].

The observed negative correlation between cleavage time and mean initiation rate (Fig 5A), in conjunction with the positive correlation between cleavage time and elongation rate (Fig 5C), suggested a potential underlying biophysical control parameter: the mean nascent transcript density on the reporter gene body ρ , given by

$$\rho = \frac{\langle R \rangle}{v_{\text{elon}}}. \quad (3)$$

Possessing units of (AU/kb), this mean transcript density estimates the average number of nascent RNA transcripts per kilobase of template DNA. Plotting the cleavage time as a function of the mean transcript density yielded a negative correlation ($\rho = -0.55$, $p\text{-val} \approx 0$) that was stronger than any of the other correlations between transcription-cycle parameters at the single-cell level (Fig 5D). Mechanistically, the correlation between cleavage time and mean transcript density suggests that, on average, more closely packed nascent transcripts at the 3' end of a gene cleave faster.

Further investigations using simulations indicated that this relationship did not arise from spurious correlations in the inference procedure itself (Section G in [S1 File](#) and [S5\(E\)–S5\(H\) Fig](#)),

but rather captured real correlations in the data. Furthermore, although the four inter-parameter correlations investigated here only used mean values obtained from the inference methodology, a Monte Carlo simulation involving the full Bayesian posterior distribution confirmed the significance of the results (Section N in [S1 File](#) and [S11 Fig](#)).

Using an absolute calibration for a similar reporter gene [25] led to a rough scaling of 1 AU \approx 1 molecule corresponding to a maximal RNAP density of about 20 RNAP molecules/kb in [Fig 5D](#). With a DNA footprint of 40 bases per molecule [106], this calculation suggests that, in this regime, RNAP molecules are densely distributed, occupying about 80% of the reporter gene. We hypothesize that increased RNAP density could lead to increased pausing as a result of traffic jams [93, 94]. Due to this pausing, transcripts would be more available for cleavage, increasing overall cleavage efficiency. Regardless of the particular molecular mechanisms underlying our observations, we anticipate that this ability to resolve single-cell correlations between transcription parameters, combined with perturbative experiments, will provide ample future opportunities for studying the underlying biophysical mechanisms linking transcription processes.

Discussion

Over the last two decades, the genetically encoded MS2 [19] and PP7 [21] RNA labeling technologies have made it possible to measure nascent and cytoplasmic RNA dynamics *in vivo* in many contexts [20, 22–28, 31–36, 61, 62, 73, 76, 107–110]. However, such promising experimental techniques can only be as powerful as their underlying data-analysis infrastructure. For example, while initial studies using MS2 set the technological foundation for revealing transcriptional bursts in bacteria [20], single-celled eukaryotes [28, 111], and animals [25, 31], only recently did analysis techniques become available to reliably obtain parameters such as transcriptional burst frequency, duration, and amplitude [24, 35, 112–114].

In this work, we established a novel method for inferring quantitative parameters of the entire transcription cycle—initiation, elongation and cleavage—from live imaging data of nascent RNA dynamics. Notably, this method offers high spatiotemporal resolution at the single-cell level, resolving aspects of transcriptional activity within the body of an organism and at sub-minute resolution. Furthermore, while our experimental setup utilized two fluorophores, we found that the calibration between their intensities could be inferred directly from the data ([Fig 3](#)), rendering independent calibration and control experiments unnecessary.

After validating previously discovered spatial modulations in the mean initiation rate, we discovered an unreported modulation of the cleavage time with respect to embryo position that mirrored that of the mean initiation rate ([Fig 4E](#)). Although such a relationship at first would suggest a positive correlation between initiation and cleavage, the presence of significant negative correlation at the single-cell level refutes this idea ([Fig 5A](#)). Instead, we speculate that the spatial modulation of the cleavage time could underlie a coupling with a spatial gradient of some molecular factor that controls this transcription cycle parameter [115], possibly due to effects such as gene looping [116, 117].

These features are unattainable by widespread, but still powerful, genome-wide techniques that examine fixed samples, such as global run-on sequencing (GRO-seq) to measure elongation rates *in vivo* [118, 119]. Additionally, while fixed-tissue technologies such as single-molecule RNA-FISH provide superior spatial and molecular resolution to current live imaging technologies [45, 57], the fixation process necessarily prevents temporal analysis of the same single cell to study these dynamic transcriptional processes. Thus, live imaging approaches offer a complementary approach to widespread RNA-FISH studies of transcriptional dynamics [39–41, 43, 45, 48–52, 56–58, 77–84].

Dissecting the transcription cycle at the single-cell level

From elucidating the nature of mutations [120] and revealing mechanisms of transcription initiation [23, 40, 42, 43, 45–48, 50, 57, 60, 95, 96], transcription elongation [54, 59, 121], and translational control [122], to enabling the calibration of fluorescent proteins in absolute units [123–128], examining single-cell distributions through the lens of theoretical models has made it possible to extract molecular insights about biological function that are inaccessible through the examination of averaged quantities. The single-cell measurements afforded by our approach made it possible to infer full distributions of transcription parameters (Fig 4B, 4D and 4F). This single-cell resolution motivates a dialogue between theory and experiment for studying transcription initiation, elongation, and cleavage at the single-cell level.

We showed how our inferred distributions of initiation rates effectively filter out most uncorrelated measurement noise, which we expect to be dominated by experimental noise, while retaining information on sources of correlated noise, including underlying biological variability (Fig 4B). Additionally, our theoretical model of elongation rate distributions make it possible to test mechanistic models of RNAP transit along the gene. While still preliminary and far from conclusive, our results suggest that cell-to-cell variability in elongation rates arises from single-molecule variability in stepping rates, and that processes such as stochasticity in stepping behavior and traffic jamming due to steric hindrance alone cannot account for the observed elongation rate distributions (Fig 4D). Such statistics could then be harnessed to make predictions for future perturbative experiments that utilize, for example, mutated RNAP molecules with altered elongation rates [129] or reporter genes with differing spacer lengths between MS2 and PP7 stem loops sequences.

Finally, the simultaneous single-cell inference of transcription-cycle parameters granted us the novel capability to investigate couplings between transcription initiation, elongation, and cleavage, paving the way for future studies of mechanistic linkages between these processes. In particular, the observed coupling of the mRNA cleavage time with RNAP density (Fig 5D) suggests future experiments utilizing, for example, orthogonal stem loops on either side of the 3'UTR as potential avenues for investigating mechanisms such as RNAP traffic jams [93, 94], inefficient or rate-limiting nascent RNA cleavage [7, 100], and promoter-terminator looping [130]. Other potential experiments could include perturbative effects, such as introducing inhibitors of transcription initiation, elongation, and/or cleavage and assessing the downstream impact on the inferred transcriptional parameters to see if the perturbed effects are separable or convolved between parameters.

Comparison to existing analysis techniques

Our method provides a much-needed framework for applying statistical inference for the analysis of live imaging data of nascent transcription, complementing existing Bayesian approaches [131, 132] as well as expanding the existing repertoire of model-driven statistical techniques to analyze single-cell protein reporter data [133–136]. In particular, compared to auto-correlation analysis of transcriptional signals [137], another powerful method of analyzing live imaging transcription data, our method is quite complementary.

First, auto-correlation analysis typically requires a time-homogeneous transcript initiation process [137], and benefits immensely from having experimental data acquired over long time windows to enhance the auto-correlation signal (although recent work has improved on the ability to analyze short time windows [112]). In contrast, our model-driven inference approach can account for slight time dependence and can fit short time traces. This is of particular

relevance to the fly embryo, where each cell cycle in early development is incredibly short (here, we only examined 18 minutes of data) and transcription initiation switches from OFF to ON and back to OFF within that timeframe.

Second, auto-correlation analysis depends strongly on signal-noise ratio, namely the ability to resolve single-or-few-transcript fluctuations in the number of actively transcribing polymerases on a gene [22, 24]. Our approach, however, can be applied even if the signal-noise ratio can only resolve differences in transcript number of several transcripts, rather than just one.

Third, our model-driven approach benefits from explicitly parameterizing the various steps of the transcription cycle, allowing for the separation of processes such as elongation and cleavage. In contrast, while the auto-correlation technique has the advantage of not relying on a particular specific model, it does rely on unknown parameters such as the overall transcript dwell time, which is a combination of elongation and cleavage. Thus, it becomes harder to separate contributions from these different processes. Additionally, auto-correlation approaches cannot produce absolute rates of transcriptional processes, such as the quantified rates of mean transcription initiation obtained in this work.

Future improvements

Future improvements to experimental or inferential resolution could sharpen precision of single-cell results, increasing confidence in the distributions obtained through this methodology. For example, technologies such as lattice light-sheet microscopy [138–142] would vastly improve spatiotemporal imaging resolution and reduce uncertainty in measurements. While this increased resolution is unlikely to dramatically change the statistics reported here, it could potentially push the analysis regime to the single-molecule level, necessitating the parallel development of increasingly refined models that can account for stochasticity and fluctuations that are not resolved with bulk measurements. In addition, while our analysis restricted itself to consider only nascent RNA labeling technologies, this methodology could be extended to also examine mature labeled RNA in the nucleus and cytoplasm of an organism, providing a more complete picture of transcription.

One important caveat of our method is the failure to account for genes that undergo transcriptional bursting [71]. Here, the initiation rate fluctuates much more rapidly in time such that our assumption of a constant mean transcription initiation rate breaks down. We chose not to address this regime in this work because only a small minority of cells (4%) studied exhibited bursting behavior. Nevertheless, although our model does not capture bursting behavior (Section F in S1 File; S4(E) and S4(F) Fig), transcriptional bursting remains a prevalent phenomenon in eukaryotic transcription and thus motivates extensions to this work to account for its behavior. For example, one possible implementation to account for transcriptional bursting could first utilize the widespread two-state model used to describe this phenomenon [141] in order to partition a time trace into ON and OFF time windows. Then the MCMC inference method developed in this work could be used to quantify the transcription cycle during the ON and OFF windows with finer precision.

Outlook

To conclude, while we demonstrated this inference approach in the context of the regulation of a *hunchback* reporter in *Drosophila melanogaster*, it can be readily applied to other genes and organisms in which MS2 and PP7 have been already implemented [20, 25, 27, 28, 31, 36, 62, 142], or where non-genetically encoded RNA aptamer technologies such as Spinach [142, 143] are available. Thus, we envision that our analysis strategy will be of broad applicability to

the quantitative and molecular *in vivo* dissection of the transcription cycle and its regulation across many distinct model systems.

Methods and materials

DNA constructs

The fly strain used to express constitutive MCP-mCherry and PCP-eGFP consisted of two transgenic constructs. The first construct, MCP-NoNLS-mCherry, was created by replacing the eGFP in MCP-NoNLS-eGFP [25] with mCherry. The second construct, PCP-NoNLS-eGFP, was created by replacing MCP in the aforementioned MCP-NoNLS-eGFP with PCP, sourced from [22]. Both constructs were driven with the *nanos* promoter to deliver protein maternally into the embryo. The constructs lacked nuclear localization sequences because the presence of these sequences created spurious fluorescence puncta in the nucleus that decreased the overall signal quality. Both constructs were incorporated into fly lines using P-element transgenesis, and a single stable fly line was created by combining all three transgenes.

The reporter construct P2P-MS2-lacZ-PP7 was cloned using services from GenScript. It was incorporated into the fly genome using PhiC31-mediated Recombinase Mediated Cassette Exchange (RMCE) [144], at the 38F1 landing site.

Full details of construct and sequence information can be found in a public [Benchling folder](#).

Fly strains

Transcription of the *hunchback* reporter was measured by imaging embryos resulting from crossing *yw;MCP-NoNLS-mCherry*, *Histone-iRFP;MCP-NoNLS-mCherry*, *PCP-NoNLS-GFP* female virgins with *yw;P2P-MS2-LacZ-PP7* males. The *Histone-iRFP* transgene was provided as a courtesy from Kenneth Irvine and Yuanwang Pan.

Sample preparation and data collection

Sample preparation followed procedures described in [32], [145], and [35]. To summarize, embryos were collected, dechorinated with bleach and mounted between a semipermeable membrane (Lumox film, Starstedt, Germany) and a coverslip while embedded in Halocarbon 27 oil (Sigma). Excess oil was removed with absorbent paper from the sides to flatten the embryos slightly. Data collection was performed using a Leica SP8 scanning confocal microscope (Leica Microsystems, Biberach, Germany). The MCP-mCherry, PCP-eGFP, and Histone-iRFP were excited with laser wavelengths of 488 nm, 587 nm, and 670 nm, respectively, using a White Light Laser. Average laser powers on the specimen (measured at the output of a 10x objective) were 35 μ W and 20 μ W for the eGFP and mCherry excitation lasers, respectively. Three Hybrid Detectors (HyD) were used to acquire the fluorescent signal, with spectral windows of 496–546 nm, 600–660 nm, and 700–800 nm for the eGFP, mCherry, and iRFP signals, respectively. The confocal stack consisted of 15 equidistant slices with an overall z-height of 7 μ m and an inter-slice distance of 0.5 μ m. The images were acquired at a time resolution of 15 s, using an image resolution of 512 x 128 pixels, a pixel size of 202 nm, and a pixel dwell time of 1.2 μ s. The signal from each frame was accumulated over 3 repetitions. Data were taken for 355 cells over a total of 7 embryos, and each embryo was imaged over the first 25 min of nuclear cycle 14.

Image analysis

Images were analyzed using custom-written software following the protocols in [25] and [35]. This software contains MATLAB code automating the analysis of all microscope images obtained in this work, and can be found on a [public GitHub repository](#). Briefly, this procedure involved segmenting individual nuclei using the Histone-iRFP signal as a nuclear mask, segmenting each transcription spot based on its fluorescence, and calculating the intensity of each MCP-mCherry and PCP-eGFP transcription spot inside a nucleus as a function of time. The Trainable Weka Segmentation plugin for FIJI [146], which uses the FastRandomForest algorithm, was used to identify and segment the transcription spots. The final intensity of each spot over time was obtained by integrating pixel intensity values in a small window around the spot and subtracting the background fluorescence measured outside of the active transcriptional locus. When no activity was detected, a value of NaN was assigned.

Data analysis

Inference was done using *MCMCstat*, an adaptive MCMC algorithm [147, 148]. Figures were generated using the open-source *gramm* package for MATLAB, developed by Pierre Morel [149]. Generalized linear regression used in Fig 5 utilized a normally distributed error model and was performed using MATLAB's *glmfit* function. All scripts relating to the MCMC inference method developed in this work are available at the [associated GitHub repository](#).

Supporting information

S1 File. Detailed information supporting main text. Table A: Mean and standard deviation of model parameters used in single-cell simulations. **Table B:** Comparison of Spearman rank correlation coefficients and p-values between experimental and simulated single-cell correlations. **Table C:** Parameters used in single-molecule Monte Carlo simulation of elongation rates.

(PDF)

S1 Fig. Detailed description of reporter construct used in this work.

(EPS)

S2 Fig. Investigation of photobleaching in experimental setup.

(EPS)

S3 Fig. Scaling of fluorescence measurement noise with overall fluorescence intensity.

(EPS)

S4 Fig. Automated curation of data.

(EPS)

S5 Fig. Overview of MCMC inference validation.

(EPS)

S6 Fig. Comparison of intra- and inter-embryo variability for inferred mean initiation rates, elongation rates, and cleavage times, as a function of embryo position.

(EPS)

S7 Fig. Single cell distributions of inferred parameters.

(EPS)

S8 Fig. Comparison of coefficients of variation (CV) between inferred mean initiation rates and instantaneous counts of number of nascent RNA transcripts.

(EPS)

S9 Fig. Comparison of distribution of elongation rates with previous studies.

(EPS)

S10 Fig. Single-molecule simulations of elongation dynamics require molecular variability to describe empirical distributions.

(EPS)

S11 Fig. Monte Carlo simulation of error in single-cell analysis.

(EPS)

S1 Video. Measurement of main reporter construct.

(AVI)

S2 Video. Measurement of interlaced reporter construct.

(AVI)

S1 Data. Dataset containing results from inference procedure and simulations.

(ZIP)

Acknowledgments

We thank Sandeep Choubey, Antoine Coulon, Jane Kondev, Anders Sejr Hansen, Mustafa Mir, Rob Phillips, Manuel Razo-Mejia, and Matthew Ronshaugen for thoughtful comments on the manuscript. We also are grateful to Florian Jug, Nick Lammers, and Armando Reimer for their crucial work in developing the image analysis code used here.

Author Contributions

Conceptualization: Jonathan Liu, Hernan G. Garcia.

Data curation: Jonathan Liu, Donald Hansen.

Formal analysis: Jonathan Liu.

Funding acquisition: Hernan G. Garcia.

Investigation: Jonathan Liu, Donald Hansen.

Methodology: Jonathan Liu, Donald Hansen.

Project administration: Jonathan Liu, Hernan G. Garcia.

Resources: Jonathan Liu, Elizabeth Eck, Yang Joon Kim, Meghan Turner, Simon Alamos.

Software: Jonathan Liu.

Supervision: Jonathan Liu.

Validation: Jonathan Liu.

Visualization: Jonathan Liu.

Writing – original draft: Jonathan Liu.

Writing – review & editing: Jonathan Liu, Hernan G. Garcia.

References

1. Alberts B. *Molecular biology of the cell*. Sixth edition. ed. New York, NY: Garland Science, Taylor and Francis Group; 2015.
2. Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet*. 2012; 13(9):613–26. <https://doi.org/10.1038/nrg3207>
3. De La Mata M, Alonso CR, Kadener S, Fededa JP, Blaustein M, Pelisch F, et al. A slow RNA polymerase II affects alternative splicing in vivo. *Molecular Cell*. 2003; 12(2):525–532. <https://doi.org/10.1016/j.molcel.2003.08.001> PMID: 14536091
4. Tian B, Manley JL. Alternative polyadenylation of mRNA precursors. *Nature Reviews Molecular Cell Biology*. 2016; 18(1):18–30. <https://doi.org/10.1038/nrm.2016.116>
5. Pinto PAB, Henriques T, Freitas MO, Martins T, Domingues RG, Wyrzykowska PS, et al. RNA polymerase II kinetics in polo polyadenylation signal selection. *EMBO Journal*. 2011; 30(12):2431–2444. <https://doi.org/10.1038/emboj.2011.156> PMID: 21602789
6. Hazelbaker DZ, Marquardt S, Wlotzka W, Buratowski S. Kinetic Competition between RNA Polymerase II and Sen1-Dependent Transcription Termination. *Molecular Cell*. 2013. <https://doi.org/10.1016/j.molcel.2012.10.014>
7. Fong N, Brannan K, Erickson B, Kim H, Cortazar MA, Sheridan RM, et al. Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Molecular Cell*. 2015; 60(2):256–267. <https://doi.org/10.1016/j.molcel.2015.09.026> PMID: 26474067
8. Liu X, Freitas J, Zheng D, Oliveira MS, Hoque M, Martins T, et al. Transcription elongation rate has a tissue-specific impact on alternative cleavage and polyadenylation in *Drosophila melanogaster*. *Rna*. 2017; 23(12):1807–1816. <https://doi.org/10.1261/rna.062661.117> PMID: 28851752
9. Moore MJ, Proudfoot NJ. Pre-mRNA Processing Reaches Back to Transcription and Ahead to Translation. *Cell*. 2009; 136(4):688–700. <https://doi.org/10.1016/j.cell.2009.02.001>
10. Mapendano CK, Lykke-Andersen S, Kjems J, Bertrand E, Jensen TH. Crosstalk between mRNA 3' End Processing and Transcription Initiation. *Molecular Cell*. 2010; 40(3):410–422. <https://doi.org/10.1016/j.molcel.2010.10.012>
11. Bai L, Santangelo TJ, Wang MD. Single-molecule analysis of RNA polymerase transcription. *Annual Review of Biophysics and Biomolecular Structure*. 2006; 35:343–360. <https://doi.org/10.1146/annurev.biophys.35.010406.150153>
12. Herbert KM, Greenleaf WJ, Block SM. Single-Molecule Studies of RNA Polymerase: Motoring Along. *Annual Review of Biochemistry*. 2008; 77(1):149–176. <https://doi.org/10.1146/annurev.biochem.77.073106.100741>
13. Roeder RG. The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. *Trends in Biochemical Sciences*. 1991; 16:402–408. [https://doi.org/10.1016/0968-0004\(91\)90164-Q](https://doi.org/10.1016/0968-0004(91)90164-Q)
14. Saunders A, Core LJ, Lis JT. Breaking barriers to transcription elongation. *Nature Reviews Molecular Cell Biology*. 2006; 7(8):557–567. <https://doi.org/10.1038/nrm1981>
15. Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, Grissom SF, et al. RNA polymerase is poised for activation across the genome. *Nature Genetics*. 2007; 39(12):1507–1511. <https://doi.org/10.1038/ng.2007.21> PMID: 17994021
16. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*. 2008; 322(5909):1845–1848. <https://doi.org/10.1126/science.1162228>
17. Fuda NJ, Ardehali MB, Lis JT. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature*. 2009; 461(7261):186–92. <https://doi.org/10.1038/nature08449>
18. Churchman LS, Weissman JS. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*. 2011; 469(7330):368–+. <https://doi.org/10.1038/nature09652> PMID: 21248844
19. Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM. Localization of ASH1 mRNA particles in living yeast. *Mol Cell*. 1998; 2(4):437–45. [https://doi.org/10.1016/S1097-2765\(00\)80143-4](https://doi.org/10.1016/S1097-2765(00)80143-4)
20. Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. *Cell*. 2005; 123(6):1025–36. <https://doi.org/10.1016/j.cell.2005.09.031>
21. Chao JA, Patskovsky Y, Almo SC, Singer RH. Structural basis for the coevolution of a viral RNA-protein complex. *Nat Struct Mol Biol*. 2008; 15(1):103–5. <https://doi.org/10.1038/nsmb1327>
22. Larson DR, Zenklusen D, Wu B, Chao JA, Singer RH. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science*. 2011; 332(6028):475–8. <https://doi.org/10.1126/science.1202142>

23. Hocine S, Raymond P, Zenklusen D, Chao JA, Singer RH. Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. *Nature Methods*. 2013; 10(2):119–121. <https://doi.org/10.1038/nmeth.2305>
24. Coulon A, Ferguson ML, de Turrís V, Palangat M, Chow CC, Larson DR. Kinetic competition during the transcription cycle results in stochastic RNA processing. *Elife*. 2014; 3. <https://doi.org/10.7554/eLife.03939>
25. Garcia HG, Tikhonov M, Lin A, Gregor T. Quantitative imaging of transcription in living *Drosophila* embryos links polymerase activity to patterning. *Curr Biol*. 2013; 23(21):2140–5. <https://doi.org/10.1016/j.cub.2013.08.054>
26. Fukaya T, Lim B, Levine M. Rapid Rates of Pol II Elongation in the *Drosophila* Embryo. *Curr Biol*. 2017; 27(9):1387–1391. <https://doi.org/10.1016/j.cub.2017.03.069>
27. Tantale K, Mueller F, Kozulic-Pirher A, Lesne A, Victor JM, Robert MC, et al. A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. *Nat Commun*. 2016; 7:12248. <https://doi.org/10.1038/ncomms12248> PMID: 27461529
28. Chubb JR, Trcek T, Shenoy SM, Singer RH. Transcriptional pulsing of a developmental gene. *Curr Biol*. 2006; 16(10):1018–25. <https://doi.org/10.1016/j.cub.2006.03.092>
29. Muramoto T, Cannon D, Gierlinski M, Corrigan A, Barton GJ, Chubb JR. Live imaging of nascent RNA dynamics reveals distinct types of transcriptional pulse regulation. *Proc Natl Acad Sci U S A*. 2012. <https://doi.org/10.1073/pnas.1117603109>
30. Corrigan AM, Chubb JR. Regulation of transcriptional bursting by a naturally oscillating signal. *Current Biology*. 2014; 24(2):205–211. <https://doi.org/10.1016/j.cub.2013.12.011>
31. Lucas T, Ferraro T, Roelens B, De Las Heras Chanes J, Walczak AM, Coppey M, et al. Live imaging of bicoid-dependent transcription in *Drosophila* embryos. *Curr Biol*. 2013; 23(21):2135–9. <https://doi.org/10.1016/j.cub.2013.08.053> PMID: 24139736
32. Bothma JP, Garcia HG, Esposito E, Schlissel G, Gregor T, Levine M. Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living *Drosophila* embryos. *Proc Natl Acad Sci U S A*. 2014; 111(29):10598–10603. <https://doi.org/10.1073/pnas.1410022111>
33. Fukaya T, Lim B, Levine M. Enhancer Control of Transcriptional Bursting. *Cell*. 2016; 166(2):358–368. <https://doi.org/10.1016/j.cell.2016.05.025>
34. Faló-Sanjuan J, Lammers NC, Garcia HG, Bray SJ. Enhancer Priming Enables Fast and Sustained Transcriptional Responses to Notch Signaling. *Dev Cell*. 2019; 50(4):411. <https://doi.org/10.1016/j.devcel.2019.07.002>
35. Lammers NC, Galstyan V, Reimer A, Medin SA, Wiggins CH, Garcia HG. Multimodal transcriptional control of pattern formation in embryonic development. *Proc Natl Acad Sci U S A*. 2020; 117(2):836–847. <https://doi.org/10.1073/pnas.1912500117>
36. Lee C, Shin H, Kimble J. Dynamics of Notch-Dependent Transcriptional Bursting in Its Native Context. *Dev Cell*. 2019; 50(4):426–435 e4. <https://doi.org/10.1016/j.devcel.2019.07.001>
37. Ferraro T, Lucas T, Clemot M, De Las Heras Chanes J, Desponds J, Coppey M, et al. New methods to image transcription in living fly embryos: the insights so far, and the prospects. *Wiley Interdiscip Rev Dev Biol*. 2016; 5(3):296–310. <https://doi.org/10.1002/wdev.221> PMID: 26894441
38. Gaertner B, Zeitlinger J. RNA polymerase II pausing during development. *Development*. 2014; p. 1179–1183. <https://doi.org/10.1242/dev.088492>
39. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol*. 2006; 4(10):e309. <https://doi.org/10.1371/journal.pbio.0040309>
40. Zenklusen D, Larson DR, Singer RH. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol*. 2008; 15(12):1263–71. <https://doi.org/10.1038/nsmb.1514>
41. Wyart M, Botstein D, Wingreen NS. Evaluating gene expression dynamics using pairwise RNA fish data. *PLoS Computational Biology*. 2010; 6(11). <https://doi.org/10.1371/journal.pcbi.1000979> PMID: 21079668
42. Sanchez A, Garcia HG, Jones D, Phillips R, Kondev J. Effect of Promoter Architecture on the Cell-to-Cell Variability in Gene Expression. *PLoS Comput Biol*. 2011; 7(3):e1001100. <https://doi.org/10.1371/journal.pcbi.1001100>
43. So LH, Ghosh A, Zong C, Sepulveda LA, Segev R, Golding I. General properties of transcriptional time series in *Escherichia coli*. *Nat Genet*. 2011; 43(6):554–60. <https://doi.org/10.1038/ng.821>
44. Coulon A, Chow CC, Singer RH, Larson DR. Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat Rev Genet*. 2013; 14(8):572–84. <https://doi.org/10.1038/nrg3484>
45. Little SC, Tikhonov M, Gregor T. Precise developmental gene expression arises from globally stochastic transcriptional activity. *Cell*. 2013; 154(4):789–800. <https://doi.org/10.1016/j.cell.2013.07.025>

46. Sanchez A, Choubey S, Kondev J. Regulation of noise in gene expression. *Annu Rev Biophys*. 2013; 42:469–91. <https://doi.org/10.1146/annurev-biophys-083012-130401>
47. Sanchez A, Golding I. Genetic determinants and cellular constraints in noisy gene expression. *Science*. 2013; 342(6163):1188–93. <https://doi.org/10.1126/science.1242975>
48. Jones DL, Brewster RC, Phillips R. Promoter architecture dictates cell-to-cell variability in gene expression. *Science*. 2014; 346(6216):1533–6. <https://doi.org/10.1126/science.1255301>
49. Senecal A, Munsky B, Proux F, Ly N, Braye FE, Zimmer C, et al. Transcription factors modulate c-Fos transcriptional bursts. *Cell Rep*. 2014; 8(1):75–83. <https://doi.org/10.1016/j.celrep.2014.05.053> PMID: 24981864
50. Xu H, Sepulveda LA, Figard L, Sokac AM, Golding I. Combining protein and mRNA quantification to decipher transcriptional regulation. *Nat Methods*. 2015; 12(8):739–42. <https://doi.org/10.1038/nmeth.3446>
51. Albayrak C, Jordi CA, Zechner C, Lin J, Bichsel CA, Khammash M, et al. Digital Quantification of Proteins and mRNA in Single Mammalian Cells. *Molecular Cell*. 2016; 61(6):914–924. <https://doi.org/10.1016/j.molcel.2016.02.030> PMID: 26990994
52. Gomez-Schiavon M, Chen LF, West AE, Buchler NE. BayFish: Bayesian inference of transcription dynamics from population snapshots of single-molecule RNA FISH in single cells. *Genome Biol*. 2017; 18(1):164. <https://doi.org/10.1186/s13059-017-1297-9>
53. Shaffer SM, Dunagin MC, Torborg SR, Torre EA, Emerit B, Krepler C, et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature*. 2017; 546(7658):431–435. <https://doi.org/10.1038/nature22794> PMID: 28607484
54. Serov AS, Levine AJ, Mani M. Abortive Initiation as a Bottleneck for Transcription in the Early *Drosophila* Embryo. *ArXiv e-prints*. 2017; p. 1701.06079.
55. Lucas T, Tran H, Perez Romero CA, Guillou A, Fradin C, Coppéy M, et al. 3 minutes to precisely measure morphogen concentration. *PLoS Genet*. 2018; 14(10):e1007676. <https://doi.org/10.1371/journal.pgen.1007676> PMID: 30365533
56. Munsky B, Li G, Fox ZR, Shepherd DP, Neuert G. Distribution shapes govern the discovery of predictive models for gene regulation. *Proceedings of the National Academy of Sciences of the United States of America*. 2018; 115(29):7533–7538. <https://doi.org/10.1073/pnas.1804060115>
57. Zoller B, Little SC, Gregor T. Diverse Spatial Expression Patterns Emerge from Unified Kinetics of Transcriptional Bursting. *Cell*. 2018; 175(3):835–847 e25. <https://doi.org/10.1016/j.cell.2018.09.056>
58. Miura M, Dey S, Ramanayake S, Singh A, Rueda DS, Bangham CRM. Kinetics of HTLV-1 reactivation from latency quantified by single-molecule RNA FISH and stochastic modeling. *PLoS Pathogens*. 2019; 15(11):1–20. <https://doi.org/10.1371/journal.ppat.1008164>
59. Ali MZ, Choubey S, Das D, Brewster RC. Probing Mechanisms of Transcription Elongation Through Cell-to-Cell Variability of RNA Polymerase. *Biophysical Journal*. 2020; 118(7):1769–1781. <https://doi.org/10.1016/j.bpj.2020.02.002>
60. Filatova T, Popovic N, Grima R. Statistics of nascent and mature RNA fluctuations in a stochastic model of transcriptional initiation, elongation, pausing, and termination. *bioRxiv*. 2020;.
61. Eck E, Liu J, Kazemzadeh-Atoufi M, Ghoreishi S, Blythe SA, Garcia HG. Quantitative dissection of transcription in development yields evidence for transcription factor-driven chromatin accessibility. *Elife*. 2020; 9. <https://doi.org/10.7554/eLife.56429>
62. Darzacq X, Shav-Tal Y, de Turris V, Brody Y, Shenoy SM, Phair RD, et al. In vivo dynamics of RNA polymerase II transcription. *Nat Struct Mol Biol*. 2007; 14(9):796–806. <https://doi.org/10.1038/nsmb1280> PMID: 17676063
63. Driever W, Nusslein-Volhard C. The bicoid protein is a positive regulator of hunchback transcription in the early *Drosophila* embryo. *Nature*. 1989; 337(6203):138–43. <https://doi.org/10.1038/337138a0>
64. Margolis JS, Borowsky ML, Steingrimsson E, Shim CW, Lengyel JA, Posakony JW. Posterior stripe expression of hunchback is driven from two promoters by a common enhancer element. *Development*. 1995; 121(9):3067–77. <https://doi.org/10.1242/dev.121.9.3067>
65. Perry MW, Bothma JP, Luu RD, Levine M. Precision of hunchback expression in the *Drosophila* embryo. *Curr Biol*. 2012; 22(23):2247–52. <https://doi.org/10.1016/j.cub.2012.09.051>
66. Chen H, Xu Z, Mei C, Yu D, Small S. A system of repressor gradients spatially organizes the boundaries of bicoid-dependent target genes. *Cell*. 2012; 149(3):618–29. <https://doi.org/10.1016/j.cell.2014.01.034>
67. Liu F, Morrison AH, Gregor T. Dynamic interpretation of maternal inputs by the *Drosophila* segmentation gene network. *Proc Natl Acad Sci U S A*. 2013; 110(17):6724–9. <https://doi.org/10.1073/pnas.1220912110>

68. Liu J, Ma J. Modulation of temporal dynamics of gene transcription by activator potency in the drosophila embryo. *Development (Cambridge)*. 2015; 142(21):3781–3790. <https://doi.org/10.1242/dev.126946>
69. Gorski SA, Dunder M, Misteli T. The road much traveled: trafficking in the cell nucleus. *Current Opinion in Cell Biology*. 2006; 18(3):284–290. <https://doi.org/10.1016/j.ceb.2006.03.002>
70. Geyer CJ. Practical Markov Chain Monte Carlo. *Statistical Science*. 1992; 7(4):473–483. <https://doi.org/10.1214/ss/1177011147>
71. Rodriguez J, Larson DR. Transcription in Living Cells: Molecular Mechanisms of Bursting. *Annu Rev Biochem*. 2020; 89:189–212. <https://doi.org/10.1146/annurev-biochem-011520-105250>
72. Wu B, Chen J, Singer RH. Background free imaging of single mRNAs in live cells using split fluorescent proteins. *Sci Rep*. 2014; 4:3615.
73. Chen H, Levo M, Barinov L, Fujioka M, Jaynes JB, Gregor T. Dynamic interplay between enhancer-promoter topology and gene activity. *Nat Genet*. 2018; 50(9):1296–1303. <https://doi.org/10.1038/s41588-018-0175-z>
74. Child MB, Bateman JR, Jahangiri A, Reimer A, Lammers NC, Sabouni N, et al. Live imaging and biophysical modeling support a button-based mechanism of somatic homolog pairing in *Drosophila*. *bioRxiv*. 2020; p. 265108.
75. Wu B, Chao JA, Singer RH. Fluorescence Fluctuation Spectroscopy Enables Quantitative Imaging of Single mRNAs in Living Cells. *Biophysical Journal*. 2012; 102:2936. <https://doi.org/10.1016/j.bpj.2012.05.017>
76. Lenstra TL, Rodriguez J, Chen H, Larson DR. Transcription Dynamics in Living Cells. *Annual Review of Biophysics*. 2016; 45(1):25–47. <https://doi.org/10.1146/annurev-biophys-062215-010838>
77. Femino AM, Fay FS, Fogarty K, Singer RH. Visualization of single RNA transcripts in situ. *Science*. 1998; 280(5363):585–90. <https://doi.org/10.1126/science.280.5363.585>
78. Pare A, Lemons D, Kosman D, Beaver W, Freund Y, McGinnis W. Visualization of individual *Scr* mRNAs during *Drosophila* embryogenesis yields evidence for transcriptional bursting. *Curr Biol*. 2009; 19(23):2037–42. <https://doi.org/10.1016/j.cub.2009.10.028>
79. Boettiger AN, Levine M. Rapid transcription fosters coordinate snail expression in the *Drosophila* embryo. *Cell Rep*. 2013; 3(1):8–15. <https://doi.org/10.1016/j.celrep.2012.12.015>
80. Fei J, Singh D, Zhang Q, Park S, Balasubramanian D, Golding I, et al. Determination of in vivo target search kinetics of regulatory noncoding RNA. *Science*. 2015; 347(6228):1371–1374. <https://doi.org/10.1126/science.1258849> PMID: 25792329
81. Padovan-Merhar O, Nair GP, Biaesch AG, Mayer A, Scarfone S, Foley SW, et al. Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. *Mol Cell*. 2015; 58(2):339–52. <https://doi.org/10.1016/j.molcel.2015.03.005> PMID: 25866248
82. Skinner SO, Xu H, Nagarkar-Jaiswal S, Freire PR, Zwaka TP, Golding I. Single-cell analysis of transcription kinetics across the cell cycle. *Elife*. 2016; 5:e12175. <https://doi.org/10.7554/eLife.12175>
83. Bartman CR, Hsu SC, Hsiung CC, Raj A, Blobel GA. Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Mol Cell*. 2016; 62(2):237–247. <https://doi.org/10.1016/j.molcel.2016.03.007>
84. Hendy O, Campbell JJ, Weissman JD, Larson DR, Singer DS. Differential context-specific impact of individual core promoter elements on transcriptional dynamics. *Mol Biol Cell*. 2017; 28(23):3360–3370. <https://doi.org/10.1091/mbc.e17-06-0408>
85. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. *Science*. 2002; 297(5584):1183–6. <https://doi.org/10.1126/science.1070919>
86. Larschan E, Bishop EP, Kharchenko PV, Core LJ, Lis JT, Park PJ, et al. X chromosome dosage compensation via enhanced transcriptional elongation in *Drosophila*. *Nature*. 2011; 471(7336):115–8. <https://doi.org/10.1038/nature09757> PMID: 21368835
87. Batsché E, Yaniv M, Muchardt C. The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nature Structural and Molecular Biology*. 2006; 13(1):22–29. <https://doi.org/10.1038/nsmb1030>
88. Wu B, Eliscovich C, Yoon YJ, Singer RH. Translation dynamics of single mRNAs in live cells and neurons. *Science*. 2016; 352(6292):1430–5. <https://doi.org/10.1126/science.aaf1084>
89. Boireau S, Maiuri P, Basyuk E, de la Mata M, Knezevich A, Pradet-Balade B, et al. The transcriptional cycle of HIV-1 in real-time and live cells. *J Cell Biol*. 2007; 179(2):291–304. <https://doi.org/10.1083/jcb.200706018> PMID: 17954611
90. Ardehali MB, Lis JT. Tracking rates of transcription and splicing in vivo. *Nat Struct Mol Biol*. 2009; 16(11):1123–4. <https://doi.org/10.1038/nsmb1109-1123>

91. Palangat M, Larson DR. Complexity of RNA polymerase II elongation dynamics. *Biochimica et Biophysica Acta—Gene Regulatory Mechanisms*. 2012; 1819(7):667–672. <https://doi.org/10.1016/j.bbagr.2012.02.024>
92. Fuchs G, Voickek Y, Benjamin S, Shlomit G, Amit I, Moshe O. 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biology*. 2014; 15(5):R69. <https://doi.org/10.1186/gb-2014-15-5-r69>
93. Klumpp S, Hwa T. Stochasticity and traffic jams in the transcription of ribosomal RNA: Intriguing role of termination and antitermination. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105(47):18159–18164. <https://doi.org/10.1073/pnas.0806084105>
94. Klumpp S. Pausing and Backtracking in Transcription Under Dense Traffic Conditions. *Journal of Statistical Physics*. 2011; 142(6):1252–1267. <https://doi.org/10.1007/s10955-011-0120-3>
95. Choubey S, Kondev J, Sanchez A. Deciphering Transcriptional Dynamics In Vivo by Counting Nascent RNA Molecules. *PLoS Comput Biol*. 2015; 11(11):e1004345. <https://doi.org/10.1371/journal.pcbi.1004345>
96. Choubey S, Kondev J, Sanchez A. Distribution of Initiation Times Reveals Mechanisms of Transcriptional Regulation in Single Cells. *Biophysical Journal*. 2018; 114(9):2072–2082. <https://doi.org/10.1016/j.bpj.2018.03.031>
97. Carrillo Oesterreich F, Preibisch S, Neugebauer KM. Global analysis of nascent rna reveals transcriptional pausing in terminal exons. *Molecular Cell*. 2010; 40(4):571–581. <https://doi.org/10.1016/j.molcel.2010.11.004>
98. Tolić-Nørrelykke SF, Engh AM, Landick R, Gelles J. Diversity in the Rates of Transcript Elongation by Single RNA Polymerase Molecules. *Journal of Biological Chemistry*. 2004; 279(5):3292–3299. <https://doi.org/10.1074/jbc.M310290200>
99. Larson MH, Landick R, Block SM. Single-Molecule Studies of RNA Polymerase: One Singular Sensation, Every Little Step It Takes. *Molecular Cell*. 2011; 41(3):249–262. <https://doi.org/10.1016/j.molcel.2011.01.008>
100. Jung W, Tschaplinski T, Wang L, Glazebrook J, Greenberg J. Priming in systemic plant immunity. *Science*. 2009; 324(5923):89–91. <https://doi.org/10.1126/science.1170025>
101. Bentley DL. Coupling mRNA processing with transcription in time and space. *Nat Rev Genet*. 2014; 15(3):163–75. <https://doi.org/10.1038/nrg3662>
102. Richard P, Manley JL. Transcription termination by nuclear RNA polymerases. *Genes Dev*. 2009; 23(11):1247–69. <https://doi.org/10.1101/gad.1792809>
103. Kuehner JN, Pearson EL, Moore C. Unravelling the means to an end: RNA polymerase II transcription termination. *Nature Reviews Molecular Cell Biology*. 2011; 12(5):283–294. <https://doi.org/10.1038/nrm3098>
104. Combs PA, Eisen MB. Genome-wide measurement of spatial expression in patterning mutants of *Drosophila melanogaster*. *F1000Res*. 2017; 6:41. <https://doi.org/10.12688/f1000research.9720.1>
105. Haines JE, Eisen MB. Patterns of chromatin accessibility along the anterior-posterior axis in the early *Drosophila* embryo. *PLoS Genet*. 2018; 14(5):e1007367. <https://doi.org/10.1371/journal.pgen.1007367>
106. Selby CP, Drapkin R, Reinberg D, Sancar A. RNA polymerase II stalled at a thymine dimer: Footprint and effect on excision repair. *Nucleic Acids Research*. 1997; 25(4):787–793. <https://doi.org/10.1093/nar/25.4.787>
107. Lenstra TL, Coulon A, Chow CC, Larson DR. Single-Molecule Imaging Reveals a Switch between Spurious and Functional ncRNA Transcription. *Mol Cell*. 2015; 60(4):597–610. <https://doi.org/10.1016/j.molcel.2015.09.028>
108. Dufourt J, Trullo A, Hunter J, Fernandez C, Lazaro J, Dejean M, et al. Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs. *Nat Commun*. 2018; 9(1):5194. <https://doi.org/10.1038/s41467-018-07613-z> PMID: 30518940
109. Fritzschn C, Baumgartner S, Kuban M, Steinshorn D, Reid G, Legewie S. Estrogen-dependent control and cell-to-cell variability of transcriptional bursting. *Mol Syst Biol*. 2018; 14(2):e7678. <https://doi.org/10.15252/msb.20177678>
110. Li J, Dong A, Saydaminova K, Chang H, Wang G, Ochiai H, et al. Single-Molecule Nanoscopy Elucidates RNA Polymerase II Transcription at Single Genes in Live Cells. *Cell*. 2019. <https://doi.org/10.1016/j.cell.2019.05.029> PMID: 31155237
111. Larson DR, Singer RH, Zenklusen D. A single molecule view of gene expression. *Trends Cell Biol*. 2009; 19(11):630–7. <https://doi.org/10.1016/j.tcb.2009.08.008>
112. Desponds J, Tran H, Ferraro T, Lucas T, Perez Romero C, Guillou A, et al. Precision of Readout at the hunchback Gene: Analyzing Short Transcription Time Traces in Living Fly Embryos. *PLoS Comput Biol*. 2016; 12(12):e1005256. <https://doi.org/10.1371/journal.pcbi.1005256> PMID: 27942043

113. Corrigan AM, Tunnaclyffe E, Cannon D, Chubb JR. A continuum model of transcriptional bursting. *Elife*. 2016; 5. <https://doi.org/10.7554/eLife.13051>
114. Bowles JR, Hoppe C, Ashe HL, Rattray M. Scalable inference of transcriptional kinetic parameters from MS2 time series data. *bioRxiv*. 2020; p. 2020.12.04.412049.
115. El Kaderi B, Medler S, Raghunayakula S, Ansari A. Gene looping is conferred by activator-dependent interaction of transcription initiation and termination machineries. *Journal of Biological Chemistry*. 2009; 284(37):25015–25025. <https://doi.org/10.1074/jbc.M109.007948>
116. O'Sullivan JM, Tan-Wong SM, Morillon A, Lee B, Coles J, Mellor J, et al. Gene loops juxtapose promoters and terminators in yeast. *Nature Genetics*. 2004; 36(9):1014–1018. <https://doi.org/10.1038/ng1411> PMID: 15314641
117. Tan-Wong SM, French JD, Proudfoot NJ, Brown MA. Dynamic interactions between the promoter and terminator regions of the mammalian BRCA1 gene. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105(13):5160–5165. <https://doi.org/10.1073/pnas.0801048105>
118. Danko CG, Hah N, Luo X, Martins AL, Core L, Lis JT, et al. Signaling Pathways Differentially Affect RNA Polymerase II Initiation, Pausing, and Elongation Rate in Cells. *Molecular Cell*. 2013; 50(2):212–222. <https://doi.org/10.1016/j.molcel.2013.02.015> PMID: 23523369
119. Jonkers I, Lis JT. Getting up to speed with transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*. 2015; 16(3):167–177. <https://doi.org/10.1038/nrm3953>
120. Luria SE, Delbruck M. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics*. 1943; 28(6):491–511. <https://doi.org/10.1093/genetics/28.6.491>
121. Boettiger AN, Ralph PL, Evans SN. Transcriptional Regulation: Effects of Promoter Proximal Pausing on Speed, Synchrony and Reliability. *PLoS Computational Biology*. 2011; 7(5). <https://doi.org/10.1371/journal.pcbi.1001136> PMID: 21589887
122. Cai L, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. *Nature*. 2006; 440(7082):358–62. <https://doi.org/10.1038/nature04599>
123. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB. Gene regulation at the single-cell level. *Science*. 2005; 307(5717):1962–5. <https://doi.org/10.1126/science.1106914>
124. Rosenfeld N, Perkins TJ, Alon U, Elowitz MB, Swain PS. A fluctuation method to quantify in vivo fluorescence data. *Biophys J*. 2006; 91(2):759–66. <https://doi.org/10.1529/biophysj.105.073098>
125. Teng SW, Wang Y, Tu KC, Long T, Mehta P, Wingreen NS, et al. Measurement of the copy number of the master quorum-sensing regulator of a bacterial cell. *Biophys J*. 2010; 98(9):2024–31. <https://doi.org/10.1016/j.bpj.2010.01.031> PMID: 20441767
126. Brewster RC, Weinert FM, Garcia HG, Song D, Rydenfelt M, Phillips R. The transcription factor titration effect dictates level of gene expression. *Cell*. 2014; 156(6):1312–23. <https://doi.org/10.1016/j.cell.2014.02.022>
127. Kim NH, Lee G, Sherer NA, Martini KM, Goldenfeld N, Kuhlman TE. Real-time transposable element activity in individual live cells. *Proc Natl Acad Sci U S A*. 2016; 113(26):7278–83. <https://doi.org/10.1073/pnas.1601833113>
128. Bakker E, Swain PS. Estimating numbers of intracellular molecules through analysing fluctuations in photobleaching. *Sci Rep*. 2019; 9(1):15238. <https://doi.org/10.1038/s41598-019-50921-7>
129. Chen Y, Chafin D, Price DH, Greenleaf AL. Drosophila RNA polymerase II mutants that affect transcription elongation. *Journal of Biological Chemistry*. 1996; 271(11):5993–5999. <https://doi.org/10.1074/jbc.271.11.5993>
130. Hampsey M, Singh BN, Ansari A, Lainé JP, Krishnamurthy S. Control of eukaryotic gene expression: Gene loops and transcriptional memory. *Advances in Enzyme Regulation*. 2011; 51(1):118–125. <https://doi.org/10.1016/j.advenzreg.2010.10.001>
131. Gupta S, Hainsworth L, Hogg J, Lee R, Faeder J. Evaluation of Parallel Tempering to Accelerate Bayesian Parameter Estimation in Systems Biology. *Proceedings—26th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, PDP 2018*. 2018; p. 690–697.
132. Gupta S, Lee REC, Faeder JR. Parallel Tempering with Lasso for model reduction in systems biology. *PLoS Computational Biology*. 2020; 16(3):1–22. <https://doi.org/10.1371/journal.pcbi.1007669>
133. Heron EA, Finkenstädt B, Rand DA. Bayesian inference for dynamic transcriptional regulation; the Hes1 system as a case study. *Bioinformatics*. 2007; 23(19):2596–2603. <https://doi.org/10.1093/bioinformatics/btm367>
134. Finkenstädt B, Heron EA, Komorowski M, Edwards K, Tang S, Harper CV, et al. Reconstruction of transcriptional dynamics from gene reporter data using differential equations. *Bioinformatics*. 2008; 24(24):2901–2907. <https://doi.org/10.1093/bioinformatics/btn562> PMID: 18974172

135. Suter DM, Molina N, Gattfield D, Schneider K, Schibler U, Naef F. Mammalian genes are transcribed with widely different bursting kinetics. *Science*. 2011; 332(6028):472–4. <https://doi.org/10.1126/science.1198817>
136. Zechner C, Unger M, Pelet S, Peter M, Koepl H. Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nat Methods*. 2014; 11(2):197–202. <https://doi.org/10.1038/nmeth.2794>
137. Coulon A, Larson DR. Fluctuation Analysis: Dissecting Transcriptional Kinetics with Signal Theory. *Methods in Enzymology*. 2016;. <https://doi.org/10.1016/bs.mie.2016.03.017> PMID: 27241754
138. Chen BC, Legant WR, Wang K, Shao L, Milkie DE, Davidson MW, et al. Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution. *Science*. 2014; 346(6208):1257998. <https://doi.org/10.1126/science.1257998> PMID: 25342811
139. Mir M, Reimer A, Haines JE, Li XY, Stadler M, Garcia H, et al. Dense Bicoid hubs accentuate binding along the morphogen gradient. *Genes Dev*. 2017; 31(17):1784–1794. <https://doi.org/10.1101/gad.305078.117> PMID: 28982761
140. Mir M, Stadler MR, Ortiz SA, Hannon CE, Harrison MM, Darzacq X, et al. Dynamic multifactor hubs interact transiently with sites of active transcription in *Drosophila* embryos. *Elife*. 2018; 7:e40497. <https://doi.org/10.7554/eLife.40497> PMID: 30589412
141. Peccoud J, Ycart B. Markovian modeling of gene product synthesis. *Theor Popul Biol*. 1995; 48:222–234. <https://doi.org/10.1006/tpbi.1995.1027>
142. Sato H, Das S, Singer RH, Vera M. Imaging of DNA and RNA in Living Eukaryotic Cells to Reveal Spatiotemporal Dynamics of Gene Expression. *Annual review of biochemistry*. 2020; p. 1–29.
143. Paige JS, Wu KY, Jaffrey SR. RNA Mimics of Green Fluorescent Protein. *Science*. 2011; 333(July). <https://doi.org/10.1126/science.1207339> PMID: 21798953
144. Bateman JR, Lee AM, Wu CT. Site-specific transformation of *Drosophila* via phiC31 integrase-mediated cassette exchange. *Genetics*. 2006; 173(2):769–77. <https://doi.org/10.1534/genetics.106.056945>
145. Garcia HG, Gregor T. In: Gaspar I, editor. Live Imaging of mRNA Synthesis in *Drosophila*. *Methods Mol Biol*. 2018; 1649:349–357. https://doi.org/10.1007/978-1-4939-7213-5_23
146. Arganda-Carreras I, Kaynig V, Rueden C, Eliceiri KW, Schindelin J, Cardona A, et al. Trainable Weka Segmentation: A machine learning tool for microscopy pixel classification. *Bioinformatics*. 2017; 33(15):2424–2426. <https://doi.org/10.1093/bioinformatics/btx180> PMID: 28369169
147. Haario H, Saksman E, Tamminen J. An Adaptive Metropolis Algorithm. *Bernoulli*. 2001; 7(2):223. <https://doi.org/10.2307/3318737>
148. Haario H, Laine M, Mira A, Saksman E. DRAM: Efficient adaptive MCMC. *Statistics and Computing*. 2006; 16(4):339–354. <https://doi.org/10.1007/s11222-006-9438-0>
149. Morel P. Grammar: grammar of graphics plotting in Matlab. *The Journal of Open Source Software*. 2018; 3(23):568. <https://doi.org/10.21105/joss.00568>