Live dynamical tracking of slncRNA speckles in single *E. coli* cells reveals bursts of fluorescence degradation

Naor Granik[1], Noa Katz[2], Yoav Shechtman[1], and Roee Amit[2,3*]

Affiliation:

[1] Department of Biomedical Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel.
[2] Department of Biotechnology and Food Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel.
[3] The Russell Berrie Nanotechnology Institute, Technion - Israel Institute of Technology, Haifa 32000, Israel.
***** To whom correspondence should be addressed: roeeamit@technion.ac.il.

Abstract

We study synthetic fluorescent speckles in bacteria by encoding two types of synthetic long-non-coding RNA genes (slncRNA) that incorporate RNA-binding phage-coat-protein (RBP) binding sites downstream from a pT7 promoter in BL21 *E. coli* cells. For both slncRNAs studied, fluorescent speckles containing ~15-30 RBP-bound slncRNA molecules form in cell poles. Measurement of speckle fluorescence reveals both positive and negative bursts spaced by exponentially distributed periods of non-classified activity lasting 10 minutes on average. We identify positive events with transcriptional bursts, and term the negative events fluorescence degradation bursts. Our data indicates that negative bursts correspond to shedding of multiple slncRNAs back to cytoplasm. Finally, we composed a kinetic model that reproduced the negative bursts and showed that a multi-burst mechanism can reduce gene expression intrinsic noise. These results may have implications to transcriptional dynamics, synthetic speckle formation, and to the understanding of the role of intracellular phase-separated compartments.

Introduction

Transcription is a complex process that depends on successive stochastic interactions between many molecular species (transcription factors, promoters, polymerases, etc.). This randomness leads to variability in gene expression levels, even in a genetically identical population of cells (Blake et al., 2006; Elowitz, 2002; McAdams and Arkin, 1997). In recent years, single molecule studies of transcription in differing cell types (varying from bacteria to mammalian cells), have unexpectedly revealed dynamics characterized by bursts of transcriptional events that are separated by periods of quiescence in which transcription is barely observed (Chong et al., 2014; Golding et al., 2005; Tutucci et al., 2018a). These observations have been predominantly obtained using two methods offering single molecule resolution. The first is single-molecule fluorescent in-situ hybridization (SM-FISH), and the second, phage-coat-protein labeling of cassettes containing multiple binding sites in living cells. SM-FISH facilitates quantitative analysis of the number of transcripts at a given time point in a population of single fixed cells by labeling mRNA molecules with fluorescently tagged DNA probes complementary to the transcript sequence. Although highly quantitative, this approach does not allow for the direct exploration of the temporal dynamics of transcription, and instead these are inferred from population statistics (Femino et al., 1998; Raj et al., 2008).

In order to directly study the dynamics of transcription, Singer and colleagues (Bertrand et al., 1998) introduced a second method, whereby a set of stem loops is encoded either into the 5' or 3' end of a transcript. The loops encode a binding site for the coat-protein of either MS2 or PP7 bacteriophages. In addition, the cells separately express a chimera of the particular coat-protein fused to a fluorescent protein. When co-expressed, the coat-protein-bound cassettes yield bright puncta or speckles which can be tracked in living cells. Thus, in theory, evaluation of the spot intensity allows one to interrogate the dynamics of processes at the single-molecule-in-single-cell level (Bertrand et al., 1998; Tutucci et al., 2018a). Despite extensive efforts to optimize this technology yielding a commonly used cassette developed by Tutucci and Singer (Tutucci et al., 2018b) consisting of 12 or 24 binding sites, this approach has still not reached the single-molecule-in-single-cell threshold necessary for a direct evaluation of transcriptional dynamics. This is a result of several critical drawbacks. First, it is thought that the synthetic binding sites disrupt natural degradation, effectively artificially extending transcript lifetimes (Jones and Elf, 2018). Second, sequences of multiple binding sites suffer from severe occupancy issues (Fusco et al., 2003), making it impossible to accurately correlate fluorescence to transcript number. Third, puncta are often composed of multiple RNA molecules, making it difficult to disentangle signals from single molecules (Fusco et al., 2003). Finally, due to the repeating binding sites, cassettes are prone both to mutation or general instability (Wells et al., 2007). For example, in the context of bacteria, Golding et al (Golding et al., 2005) engineered an MS2-coat protein binding site cassette containing 96 hairpin repeats. These were inserted downstream of the Plac/ara promoter, providing the first live evidence for transcriptional

bursts in bacteria. However, individual transcriptional events were not resolved, and this relatively large cassette was not used in follow-up studies. Instead, later studies opted to use SM-FISH, RT-PCR, and RNA-seq to provide further proof for the pervasiveness of bursty transcription throughout the microbial genome (So et al., 2011; Taniguchi et al., 2010).

In recent years, naturally occurring puncta-like complexes of lncRNAs that are bound by various RNA binding proteins (RBPs) have been discovered in many Eukaryotic cell types. The most well-studied examples of these natural puncta-like complexes are paraspeckles (Clemson et al., 2009; Fox et al., 2018; Staněk and Fox, 2017) and nuclear-speckles (Spector and Lamond, 2011) which are composed predominantly of the lncRNAs and RNA binding proteins. These particles are an example of a wider phenomenon, which has received increased attention by the research community over the past decade, of liquid-liquid phase-separated micro- and nano-compartments within cells (Li and Fu, 2019). Given the similarity in optical microscopy observations between paraspeckles (for instance) and the RNA-phage coat protein puncta, it is possible to view the latter retrospectively as studies which showed that synthetic liquid-liquid phase-separated compartments (or speckles) can be formed *in vivo* (in any cell type) with phage coat proteins and synthetic lncRNA molecules that encode cassettes of their binding sites.

In this work, we were motivated by this revisionist perspective of the phage coat protein experiments to revisit transcriptional bursting in bacteria by tracking the expression of a pT7 promoter using both the 24x state-of-the-art slncRNA, and a new and significantly shorter slncRNA encoding four PP7 and five Qβ phage-coat-protein binding sites in interlaced fashion (Qβ-5x-PP7-4x). In the new slncRNA molecule, the respective coat proteins do not recognize the other binding site family (Katz et al., 2018), and thus upon expression only half of the hairpins are bound, while the unbound half increases the stability of the entire structure. For both slncRNAs, we showed that individual speckles within single cells can be observed and dynamically tracked. Upon analyzing the resulting intensity vs. time signals, we detected not only the characteristic transcriptional bursts, but also for the first time, bursts of signal degradation suggesting that speckles not only can accrue new slncRNAs upon transcription but also shed them in a burst-like fashion. Further analysis revealed important differences between the PP7-24x and the Qβ-5x-PP7-4x speckles, implying that macroscopic characteristics of these compartments are dependent on the RNA cassette design. Finally, we studied a three-state random telegraph kinetic model to show that degradation bursts can reduce mRNA noise, thus providing a potential biological function for speckle-like compartments in the regulation of gene expression.

Results

Qβ-5x-PP7-4x slncRNA cassette leads to positive and negative fluorescent bursts in synthetic speckles

In order to study synthetic speckle formation with slncRNA-RBP complexes, we first designed a short mRNA binding-site cassette, consisting of four native PP7, and five native Qβ binding sites in an interlaced manner (Figure 1a). The cassette was cloned downstream to a pT7 promoter on a BAC (Addgene plasmid # 13422), and transformed, together with a plasmid encoding for Qβ-mCherry from a pRhlR inducible promoter, to BL21 *E. coli* cells. Single cells expressing the Qβ-5x-PP7-4x together with Qβ-mCherry (data gathered from these cells is denoted as Qβ-5x data) were imaged every 10 seconds for 60 minutes under constant conditions (200 msec integration time, 37° c), and subsequently the intensity of bright speckles (Figure 1b) resulting from the bound cassette in each cell were analyzed for every timepoint (see Methods and Figure S1 for image analysis pipeline) resulting in a trace of intensity vs. time. During processing each trace was smoothed by a 10-point moving average and subsequentially normalized by subtracting the background of the cellular environment surrounding the speckle (which was smoothed in similar manner). In Figure 1c we plot several processed traces obtained from multiple speckles tracked in different fields of view on separate days. The signals are either decreasing or increasing in overall intensity, and occasionally signals that initially show an increase and subsequent decrease are observed

To determine whether the "sharp" increases or decreases in intensity correspond to a distinct signal, and are not part of the underlying noise, we employed a scheme, whereby statistically significant changes in intensity are identified as burst events (see Figure S2 and Methods for analysis details and definitions). In brief, we assume the total fluorescence is comprised of three distinct signal processes: total transcript fluorescence, background fluorescence and noise. We further assume that background fluorescence is slowly changing, as compared with total transcript fluorescence which depends on the dynamic and frequent processes of transcription and degradation. Finally, we consider noise to be a symmetric, memory-less process. We define a "burst" as a change or shift in the level of signal intensity leading to either a higher or lower new sustainable signal intensity level. To identify such shifts in the baseline intensity we search for continuous segments of gradual increase or decrease whose probability of occurrence, under our assumption of random symmetric noise, is 1 in 1000. From this probability we set a threshold for the minimum length of a gradual shift, where events lasting longer than this threshold are classified as burst events. Segments within the signal that are not classified as either a negative or positive burst event are considered unclassified. Unclassified segments are typically signal elements whose noise profile does not allow us to make a classification into one or the other event-type. Such segments can consist of multiple event types, for example: bursts that do not pass the false positive threshold that we set, or events where no transcriptional or degradation processes are recorded. In Figure 1c-right we mark the classifications on three separate speckle traces with positive "burst", negative "burst", and non-classified events in green, red, and blue respectively. We confine our segment analysis between the first and last significant segments identified in a given signal, since we cannot correctly classify signal sections that extend beyond the

observed trace. These unprocessed segments (before the first significant event, and after the last) are marked in a dashed black line.

Next, using this classification criteria for bursts, we annotated the three features of our signal (increasing bursts, decreasing bursts, and non-classified events) for ~1500 speckle traces. In Figure 1d-e we plot the amplitude (ΔI) distribution and rate of intensity change ($\Delta I/ \Delta t$) for all three event types. The plots show distributions with three separated populations of non-classified, increasing, and decreasing bursts, with the number of positive and negative burst events being approximately equal (<15% difference regardless of moving average window and the statistical threshold set for identification- see Fig. S2), Finally, to check that our analysis is independent of our choice of image background and segmentation results, we repeated the analysis on all experimental traces using the same definition for a "burst", but with an alternative definition for cell background noise (see Figure S3 and SI for details). The results of this alternative analysis confirm that the distributions shown in Fig. 1d-e are largely independent of the cell background definition, indicating that our definition of bursts is robust to noise that might arise from the image analysis process.

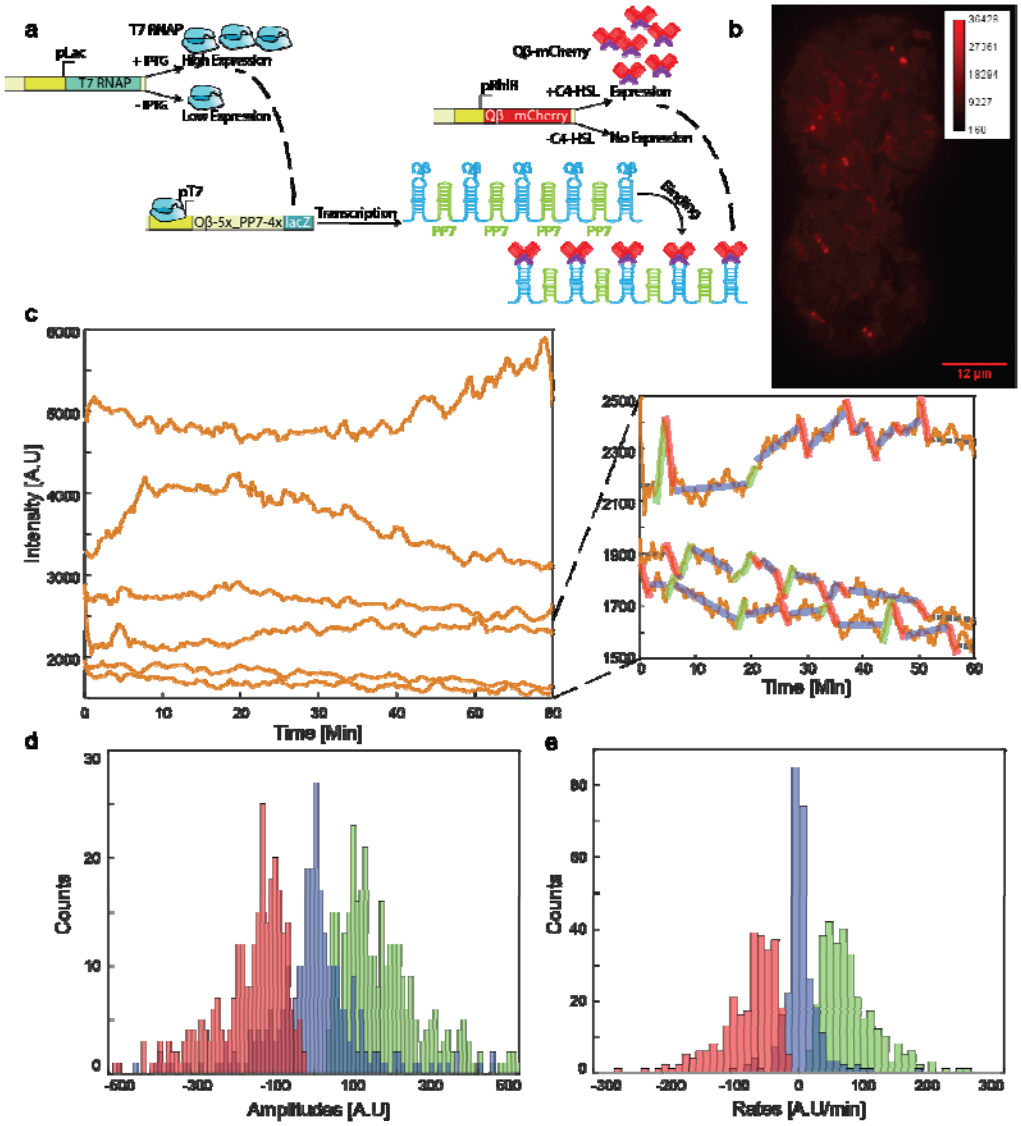

Figure 1: Biological setup, sample signals and amplitude distributions. **a.** Biological scheme of the experimental system comprised of the slncRNA with binding site configuration Qβ-5x-PP7-4x and the RBP-FP fusion protein Qβ-mCherry. **b.** A region of interest (ROI) from a field of view of an experiment showing dark background, cells (dim red) and bright speckles (bright red), resulting in a high dynamic range as shown by the color bar presenting intensity values in A.U. **c.** Sample intensity signals taken from different speckles from different experiments on separate days showing a range of behaviors. Zoom-in shows the three lower signals with overlaid segments presenting three signal states: strong increase (green), strong decrease (red), non-classified (blue). The black dashed lines mark data that was not analyzed (segments extend beyond signal range). **d.** Empirical amplitude distributions gathered from ~1000 signals. Green, blue and red correspond to increasing, decreasing and non-classified signal segments. **e.** Rate distributions gathered from ~1000 signals.

Negative and positive burst distributions indicate 1-3 molecules per burst

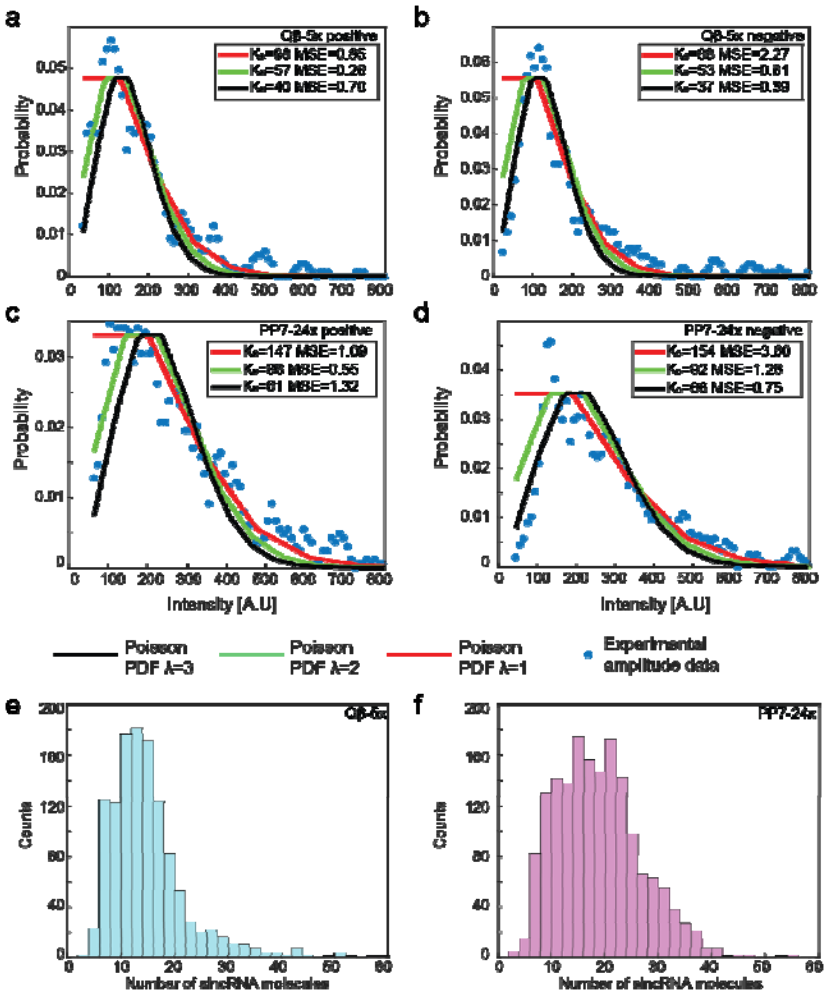

Figure 2: Poisson distribution fitting of empirical amplitude data. **a-d.** Experimental data is presented by a scatter plot, overlaid by theoretical Poisson probability distribution functions (PDFs) with parameter values 1-3, presented in red, green, and black lines, respectively. Theoretical fits normalized to correct x-axis by the factor K_0 **a.** Qβ-5x positive amplitudes. **b.** Qβ-5x negative amplitudes. **c.** PP7-24x positive amplitudes. **d.** PP7-24x negative amplitudes. **e-f.** Average number of binding sites cassettes per signal, evaluated by dividing the average signal intensity by the value of K_0 calculated from the Poisson PDF with λ=1 fit. **e.** Qβ-5x data. **f.** PP7-24x data.

To further explore the burst distributions, we first take into consideration the assumed nature of burst events. To a first approximation, molecular bursting events are thought to vary over a small range of integer values (e.g. 1-10 as for transcriptional bursting (Golding et al., 2005)), and should thus exhibit a Poisson-like distribution. Consequently, the accumulation of the fluorescent amplitudes of a large number of burst events should behave according to a Poisson distribution as well. Since each burst amplitude should be directly proportional to the number of events which comprise it, the mean number of burst events and fluorescence signature per molecule can then be extracted from such distributions.

We plot separately the distributions for the positive and negative burst amplitudes as blue dots in Figure 2a-b. Each plot is overlaid with three Poisson distribution fits with parameter λ = 1 (red), 2

(green), and 3 (black) respectively, corresponding to a mean of 1-3 transcripts per burst. Given the fact that we cannot directly infer the fluorescence intensity associated with a single RNA cassette, we fitted the distributions with a modified Poisson function of the form:

$$p(I) = \frac{\lambda^{\frac{I}{k_0}} e^{-\lambda}}{\left(\frac{I}{k_0}\right)!}, \qquad (1)$$

where I is the experimental fluorescence amplitude, λ is the Poisson parameter (rate), and K_0 is a fitting parameter whose value corresponds to the amplitude associated with a single RBP-bound slncRNA molecule within the burst. For each rate we chose the fit to K_0 that minimizes the deviation from the experimental data. The fits show that while the λ=3 distribution provides the best fit to the data (corresponding to a mean of three transcriptional slncRNA's per burst), the λ=1 distribution provides the best fit to the tail of the distribution, but fails at lower amplitude values which may be due to our analysis threshold that treats many of these small amplitude events as unclassified. Since higher rate distributions provide a progressively worse fit (as Poisson distribution resembles a Gaussian curve in higher rates), we conclude that both the transcriptional and degradation distributions provide a reasonable match to Poisson distributions with λ=1 to 3. This match suggests that the range of fluorescence amplitude spanning ~40-95 (A.U) most likelihood corresponds to the amplitude generated by the addition or subtraction of a single Qβ-5x-PP7-4x slncRNA into the speckle under our experimental conditions. Finally, given the match with the lower rate Poisson distributions, it seems likely that the number of cassettes involved in both the positive and negative bursts varies between 1 and 3 molecules per burst.

To provide additional confirmation that we are able to detect a signal from a single slncRNA molecule, we repeated the experiment with a strain expressing PP7-mCherry with a 24x PP7 binding site-cassette (sequence obtained from Addgene plasmid #31864 (Larson et al., 2011)). We plot the relevant distributions in Figure 2c-d. For this cassette (denoted PP7-24x), the plots show similar distributions as to the one observed for Qβ-5x. The distributions appear to be well described by Poisson distributions with λ=1-3 as well, but with an increase of 62-64% in the fitted K_0 per λ. This result is consistent with past observations, which have shown that these cassettes are only occupied by 8-14 proteins, resulting in a reduced intensity relative to the intensity expected from the number of designed binding sites (Fusco et al., 2003). The consistency of the Poisson fit results for both cassettes indicate that in the PP7-24x case, we are also observing 1-3 mRNAs per burst, and that the nature of the reporting cassette does not have a large effect on the outcome of the experiment.

Given the values extracted for the fluorescence intensity that is associated with a single reporter cassette, we computed a lower estimate for the average number of RBP-bound slncRNAs that make-up a single speckle. To do so we took the average value of the Poisson 1- K_0 value (93 and 151 A.U. for the Qβ-5x and PP7-24x respectively) and computed the average number of cassettes per trace

by dividing the average trace intensity with the appropriate K_0. We then plotted the resultant distribution of average number of "cassettes" for both cassette types in Fig. 2e-f. The results show that a single speckle can be estimated to be made up of at least 10-30 slncRNA molecules on average.

Speckles in steady state exhibits a three-state random telegraph signal

To check that our analysis is consistent with an underlying random burst signal, we simulated three types of base signals with an added white Gaussian noise of magnitude 30 [A.U] peak-to-peak amplitude, matching the value calculated from the experimental traces. For each simulation type, 1000 signals of 360 time-points were simulated and analyzed using the same data analysis process used for the experimental signals.

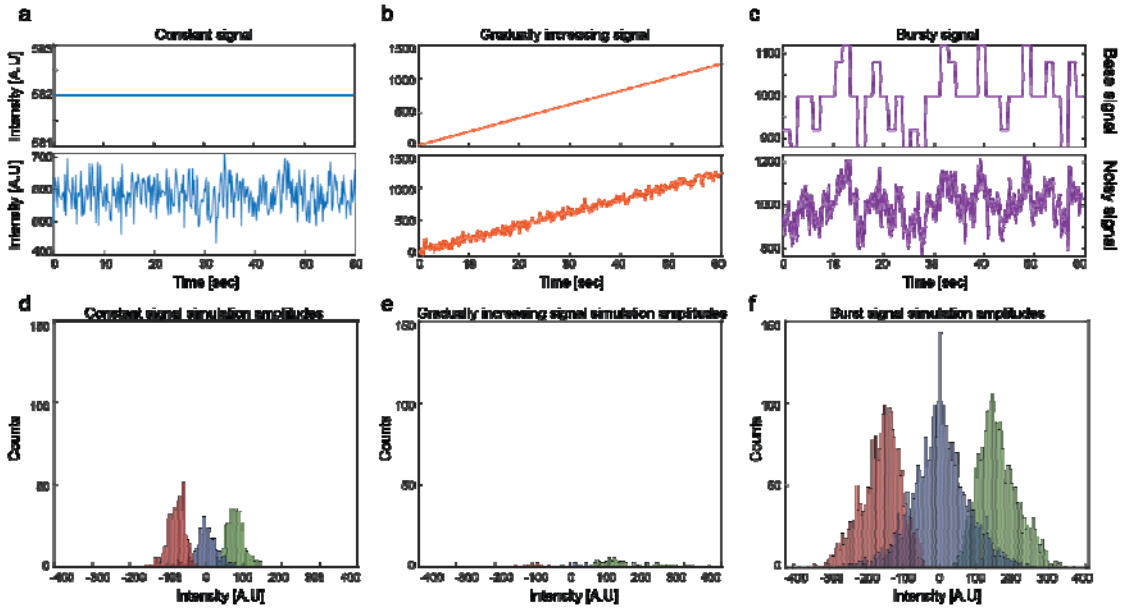

Figure 3: Numerical simulations of potential signals. **a-c.** Simulated signal without noise (top) and with Gaussian noise (bottom) for a constant (**a**), gradually increasing (**b**), and intermittent/bursty (**c**) signals. **d-f.** Amplitude distributions computed with our signal analysis algorithm for the constant (d), gradually increasing (e), and intermittent/bursty (f) signal simulations. In all three panels red, blue, and green bars correspond to strongly decreasing, non-classified, and strongly increasing events, respectively. Given the close match of the intermittent signal simulation to the experimental data, this result indicates that in our experimental data, no more than ~10% of the called "events" are false positives.

We simulated a flat, constant signal with noise (Figure 3a - top), a gradually ascending signal with noise (Figure 3b - top), and a three-state random telegraph signal with noise (Figure 3c - top). We then applied our burst-detection algorithm described above and found that for the flat signal (Figure 3d) positive and negative bursts (green and red respectively) and non-classified events are detected. However, a closer examination of the results reveals that the burst amplitude width is smaller by a factor of ~3 as compared with the experimental data bursts, and the total number of events observed is significantly smaller than the experimental data (i.e. 371 positive, 439 negative,

and 274 non-classified segments found), indicating less than one event per signal, as expected from our base assumption that a rare noise event occurs once in a thousand time points. For the gradually increasing signal with additional noise, (Figure 3e) a negligible number of burst-like events was detected by our algorithm, with a pronounced bias towards positive events. The scarcity of events can be explained by the positive bias in the signal which results in a steep increase in the statistical threshold for event identification (See SI). Similar simulations with a decreasing signal show a mirror image of amplitude distribution (data not shown).

Finally, a signal designed to mimic our interpretation of the experimental data containing randomly distributed instantaneous bursts, both increasing and decreasing with multiple possible amplitudes was analyzed (Figure 3f). Our simulated signals resulted in a symmetric amplitude distribution, comprising of non-Gaussian or skewed amplitude distributions. Additionally, the range of amplitudes observed is 2x larger as compared with the case for the constant signal, with the non-classified amplitudes presenting a wider distribution. A total of 2297 positive, 2221 negative and 2981 non-classified segments were found, which is approximately an order of magnitude larger than the number of events observed for the constant signal, and is similar to the density of events observed experimentally.

Duration of events further supports the three-state random telegraph model

We next compared frequency and duration of burst and non-classified states (Figure 4). Burst states (Figure 4a-b) for both the Qβ-5x-PP7-4x and PP7-24x cassettes appear to last approximately 2.5 minutes, irrespective of cassette size or burst-type, indicating a possible temporal resolution issue arising from the statistical analysis process wherein any event shorter than the temporal threshold (see SI) will be missed. In contrast to these narrow duration distributions, non-classified state durations (Figure 4c) are exponentially distributed, with an average decay rate of about 10 minutes for both Qβ-5x-PP7-4x and PP7-24x. The non-classified amplitude distribution (Figure 4d) for the PP7-24x cassette shows a slight preference for slow signal degradation trends (sample skewness of -0.55), that may be consistent with increased mRNA stability that has been previously attributed to these cassettes (Fusco et al., 2003) or to the underlying structural characteristic of the speckle which differ from the Qβ-5x-PP7-4x example. Specifically, the binding sites cassette either slows down, or entirely halts the signal degradation process, resulting in negative amplitudes that do not meet our statistical criteria, but instead appears in the non-classified distribution. Finally, we checked if the bursts occurred at random or whether there was some bias in the order of the bursts. To do so we examined whether after a non-classified period that lasted more than 2.5 minutes there was a bias for one type of burst or the other. The results are depicted in Figure S4 and show that no such bias seems to exist, i.e. either a positive or negative burst seems to occur after non-classified events with equal probability. This observation is consistent with the fact that we measured fluorescence from bright speckles, which

appear after accumulation of multiple binding-sites cassettes, meaning the transcript levels in the cell are at a steady state.

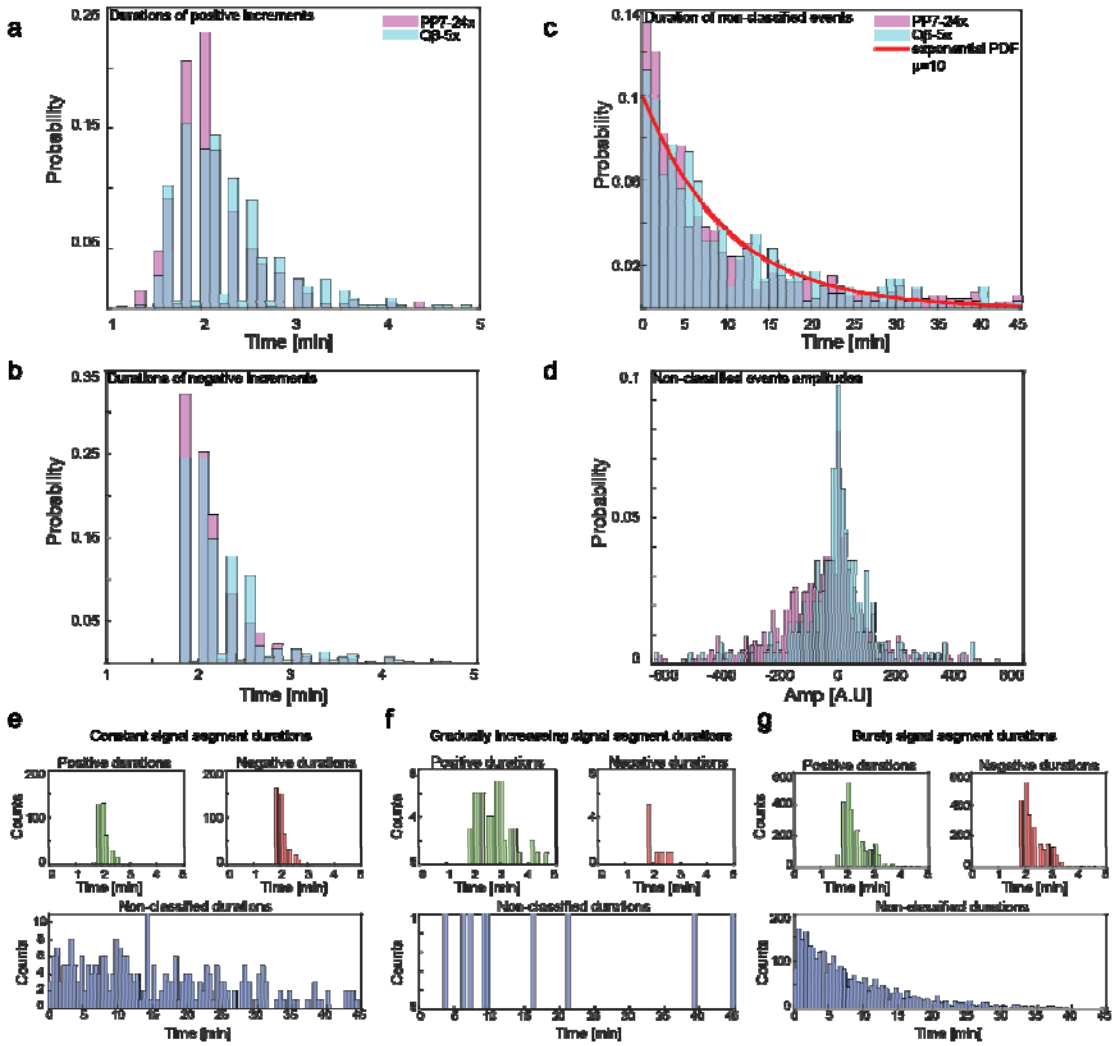

Figure 4: Temporal statistics and non-classified amplitudes for experimental and simulated signals. **a.** Duration of positive segments (transcription bursts) showing a Poisson-like distribution shifted by 1 minutes, in both cases. **b.** Duration of negative segments (degradation bursts), showing an exponential-like distribution shifted by ~2 minutes for both cases. **c.** Duration of non-classified segments, showing an exponential distribution decay time $\tau \approx 10$ (min) for both cases. **d.** Amplitude distribution of non-classified segments. The PP7-24x data is skewed toward negative values (sample skewness of -0.55 for PP7-24x, and 0.83 for Qβ-5x data). For all cases, Qβ-5x-PP7-4x, and PP7-24x cassettes are presented in cyan and magenta, respectively. **e-g.** Temporal statistics for the simulated signals for the (e) constant, (f) gradually increasing, (g) and bursty signals respectively.

In order to provide context to the information generated by the duration data, we studied the duration of events in our simulated signals. Interestingly, both positive and negative durations for the burst events in each signal bear a striking resemblance to the experimental data (Figure 4g - top). This result is consistent with an interpretation that burst duration measurements are limited by the resolution of the experiment (1 frame every 10 seconds) and choice of smoothing algorithm (ten

experimental frames or simulated time points – See SI). Together, these constraints result in a lower bound of 150 seconds, or 15 frames on the temporal resolution, in which a burst can be detected. Any process occurring faster would be obscured by the smoothing algorithm and missed. By contrast, important information can be deduced from the simulated duration of the non-classified events (Figure 4e-g – bottom). Here, the three different signals generate visually distinguishable duration signals, which correspond to a distinct fingerprint for each signal type in this case. While both the increasing and constant signal generate a gradually declining and spread-out non-classified duration distributions, the random telegraph signal generated an exponentially distributed duration distribution. This is consistent with the experimental observations and provides further evidence that the underlying signal in our experimental data is a multi-state random telegraph noise.

A multi-state random telegraph noise model describes steady state speckle intensity

The existence of negative bursts that appear to be independent of labelling cassette implies that the well-studied two-state model for transcription dynamics (Jones and Elf, 2018; Paulsson, 2005; Sanchez et al., 2011), which incorporates one state with a strong transcriptional rate and another state with a weak rate, must be modified, since the kinetics of this model cannot generate two isolated sets of bursts. We therefore add another degradation rate, extending the two-state to a four state-model, as follows: strong transcription/strong degradation, strong transcription/weak degradation, weak transcription/strong degradation, weak transcription/weak degradation (see SI for a detailed description of the model and Tables S1-S3). In practical terms, the first and second states result in transcriptional bursts and cannot be discerned neither in simulation nor experiments when we assume a stronger transcription rate than degradation rate for both states. The third state results in negative bursts when the degradation rate is larger than the weak transcriptional rate, while the fourth state results in a quiescent state of apparent inactivity when both transcription and degradation rates are sufficiently small. Therefore, in order to account for the negative-burst finding, we are left with three discernible states, forming a model similar to the three-state random telegraph noise (RTN) model (Gong et al., 2018) (Figures 5a, S5).

To test this new model, we implemented both it, and the two-state model, using kinetic Monte Carlo simulations (Battaile, 2008). We simulated time-lapse sequences, modelling the number of transcripts as a function of time. Both simulations were run for 3600 steps (corresponding to 1 hour, simulated at 1 sec intervals), with sampling every 10 steps to emulate the sampling conditions of the experimental data. The transition parameters between states for the two-state model, and for the transcriptional burst part of the three-state model, were taken from literature (Sanchez et al., 2011). The transition parameters into and out of the degradation burst state were set such that the distribution of number of mRNA molecules would be similar between the two models, meaning that there should be no apparent change in the number of transcripts (Figure 5b – for a list of the parameters – see Table S1).

To confirm that a three-state RTN model is better suited for describing our experimental data, we used two different methods to compare its performance to that of the two-state model. As a first test, we analyzed the simulated sequences in the same manner used for the experimental data, to generate the amplitude histograms in Figure 5c-d (3 state and 2-state respectively). The most notable difference between the two models lies in the distribution of negative amplitudes. While the two-state model produces a Gaussian distribution centered at -1, representing constant degradation, the three-state simulation generates a more flattened, spread-out distribution, similar to the one appearing in the experimental data in Figure 1d.

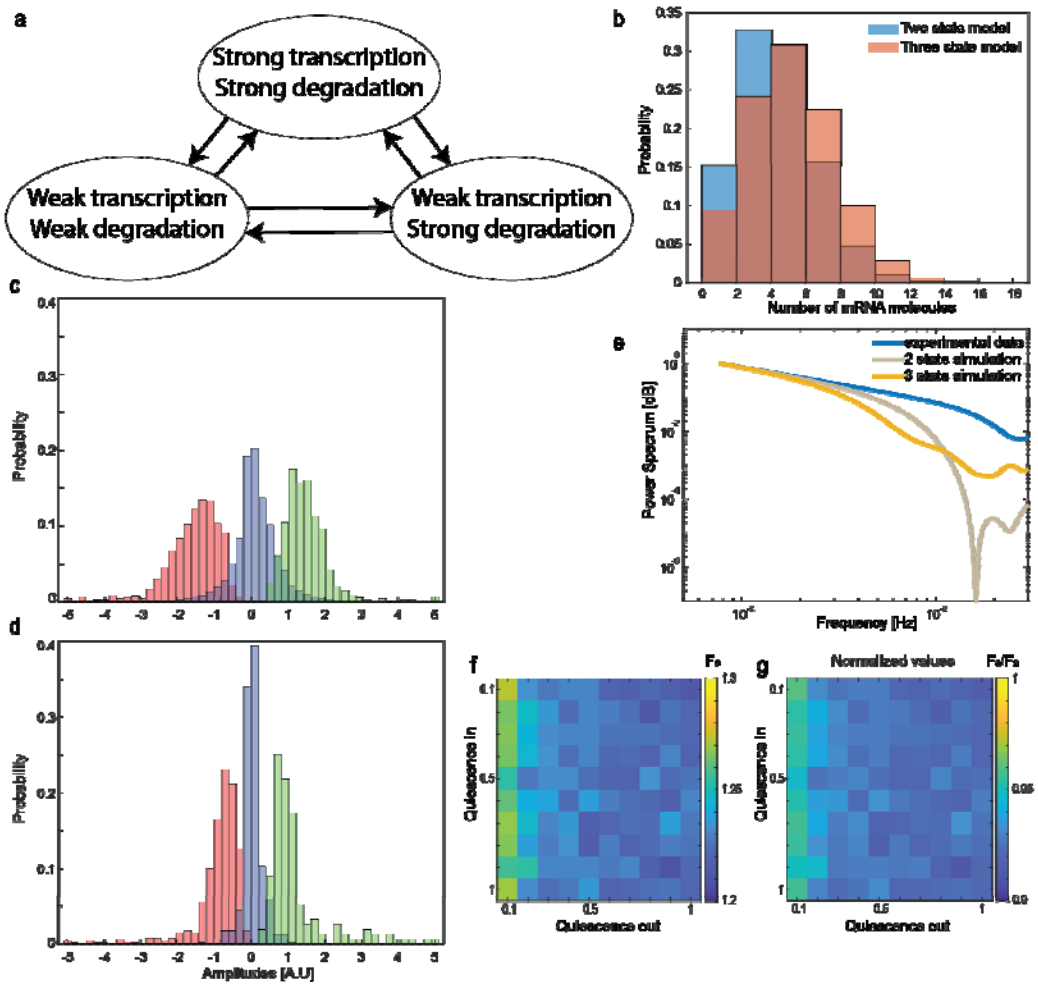

Figure 5: Three state model analysis. **a.** Model schematic showing transition possibilities between the three discernible states. **b.** Number of mRNA molecules arising from simulations of three-state (orange) and two-state (blue) models. **c-d.** Intensity amplitude analysis for the three-state model (**c**) and two-state model (**d**). **e.** Power spectrum density of experimental Qβ-5x, three-state model and two-state model. **f-g.** Fano factor analysis of the three-state model as a function of transition probabilities into, and out of, the quiescent state. Each pixel represents 500 simulations. (**f**) – Original values. (**g**) – Values normalized by the two-state model Fano factor.

The second test is rooted in telegraph processes theory, which provided the basis for the historic two-state model (Larson et al., 2009; Lionnet and Singer, 2012). A common analysis of signals generated from such processes is based on the power spectral density (PSD), which is proportional to the square of the Fourier transform of the signal (Miller and Childers, 2012). To generate a suitable signal from our data, we concatenated the slopes of all identified segments from all signals (under an assumption of memoryless noise), generating a signal resembling a telegraph process (Figure S6). We calculated the PSD from both simulations and experimental data as shown in Figure 5e. From a heuristic viewpoint, the experimental PSD bears a closer similarity to that of the three-state model in the low frequency range, as compared with the PSD of the two-state model, providing further evidence that the three-state model provides a better description for the experimental data.

Finally, we characterized the three-state-model in terms of the Fano factor, which is defined as the variance divided by the mean of a stochastic distribution, a value typically used to quantify deviation from Poisson statistics. In the context of mRNA, the Fano factor is used to characterize the relation between mRNA noise and the noise of a Poisson distribution with the same mean (Sanchez et al., 2011). A large Fano factor translates directly to a larger variance, which indicates greater fluctuations in mRNA copy number. The Fano factor for the given model was computed over a range of transition probabilities to and from the quiescent state (weak transcription/weak degradation – see Table S3 for parameter values), iterating from 0.1 to 1 [s^{-1}]. This range was selected to maintain the mRNA distribution roughly the same as the distribution predicted by the two-state model. For each pair of transition values, the three-state model simulation was repeated 500 times, each iteration yielding an mRNA copy number distribution. The Fano factor was calculated separately for each of the 500 distribution. Figures 5f presents the average Fano factor as a function of transition probabilities, while figure 5g presented the same values normalized by the average Fano factor calculated from 500 iterations of the two-state model simulation. As can be seen, the average values (i.e. not normalized) are consistently greater than one, while still being smaller than the two-state Fano factor. The three-state model thus leads to a narrower mRNA copy number distribution in comparison to the two-state model (i.e. smaller fluctuations in mRNA copy number), while still being super-Poissonian, in full agreement with previous experimental data (Sanchez and Golding, 2013).

Discussion

In the present study, we used the pT7 promoter in bacteria to express synthetic lncRNA (slncRNA) molecules that incorporated several RBP binding sites, which together with their cognate RBP formed fluorescent speckles in the cell poles. Our speckles were composed of two different types of slncRNA molecules or binding-site cassettes: a new Qβ-5x-PP7-4x and the common-place PP7-24x cassette. Our findings reveal that at steady-state, speckles are likely composed of 15-30 RBP bound

slncRNA molecules. Changes in speckle fluorescence is characterized by exponentially distributed random bursts of either positive or negative amplitudes, spaced by periods where no classifiable change in speckle fluorescence can be made. To our knowledge there was one previous study (Golding et al., 2005) in bacteria that tracked synthetic fluorescent speckles in single cells. In that study, the RNA component was an mRNA gene that was labelled at its 5'UTR by a ~5 kbp cassette of 96 binding sites for the phage coat protein for MS2 (MCP), which in turn self-assembled into speckles that only exhibited bursts of increasing fluorescence intensity that were attributed to bursts of transcriptional activity. In this study, we identified both positive and negative bursts in approximately equal proportions in synthetic speckles composed of two significantly shorter RNA molecules. As in the Golding *et al.* (Golding et al., 2005) paper, we also interpret positive bursts as a near simultaneous addition of multiple fluorescent slncRNA molecules to the speckle resulting from bursts of transcriptional activity. By contrast, a negative burst can be interpreted as a simultaneous removal of slncRNA molecules from the speckle. It is therefore reasonable to conclude that negative bursts may correspond to bursts of degradation.

Unlike transcriptional bursts, which have been attributed to kinetic and structural nucleoid-related processes, there is no particular reason why degradation of speckle signal in bacteria should also present itself in bursts. There seem to be two possible kinetic scenarios; either slncRNA molecules are actively degraded in an intermittent fashion directly within a given speckle, or the speckle itself sheds intermittently a number of slncRNA molecules which are then further degraded in the cytoplasm. The implications of the first scenario are that RnaseE and potentially other major degradation enzymes may also be expressed in bursts themselves leading to an intermittent set of degradation events. The second scenario implies that speckles are nano-particles which are comprised of entangled protein-RNA complexes that are effectively phase-separated from the rest of the cytoplasm and thus do not allow access to other molecular species such as the ribosome or degradation enzymes. When considering the observed increased negative skewedness observed in the non-classified set of events for the 24x lncRNA as compared with its 5x shorter counterpart (Fig. 4d), the latter scenario offers a more compelling explanation. Namely, synthetic speckles that are composed of longer slncRNA molecules are likely to be more entangled leading to occasionally a slower or more gradual release of the molecules from the biomolecular complex as compared with their shorter counter parts. An increased entanglement due to length is also consistent with the wider distribution for the estimated "number of slncRNAs" within speckles observed for the 24x as compared with the shorter example (Fig. 2e-f), and the lack of negative bursts observed for the 96x speckle by Golding *et al.* (Golding et al., 2005).

Observing both positive and negative fluorescent bursts, and taking into account the presence of a phase separated compartment which stores RNA and proteins led us to a new modelling scheme for the regulation of gene expression. While a two-state RTN kinetic model suffices for describing the mRNA distribution resulting from bursty gene expression process, in order to take the effects of

speckles into account a multi-state RTN model is needed. Here, the presence of the speckles was accounted for by adapting the two-state model to one which included an additional state where degradation is a rare event. This, in turn, led to the desired three-state RTN prediction, and to a reduced variability in the number of RNAs over a cell population. The latter prediction, in particular, provides an evolutionary impetus for having bursts of degradation as a mechanism for mitigating the cell-to-cell variability of the number of RNA molecules. Consequently, this prediction may be relevant to not only this synthetic context, but to natural systems as well. This, therefore, become particularly pertinent in lieu of many recent observations in Eukaryotes, which suggests that intra-cellular liquid-liquid phase-separation into molecularly isolated compartments (Hyman et al., 2014) (e.g. paraspeckles (Clemson et al., 2009; Fox et al., 2018; Staněk and Fox, 2017), nuclear speckles (Spector and Lamond, 2011), etc.) may be a common and generic cellular phenomenon that affects the regulation of gene expression. Thus, bursting may not just be a process that characterizes transcription but may be characteristic of other molecular processes related to the regulation of gene expression as well.

Finally, we believe that the ability to reach the negative burst finding is due in large part to our new binding-site cassette design. Our cassette is ~450 bps in length, which is about 5 times shorter than the 24x, and an order of magnitude shorter than the cassette used by Golding *et al.* (Golding et al., 2005). Comparison of our cassette to the results generated by the PP7-24x indicates that our cassette is likely to be fully occupied by RBPs, as compared with <50% occupancy for the PP7-24x, as has been previously reported (Fusco et al., 2003). Therefore, continued exploration of additional designs of such synthetic long non-coding RNA molecules have the potential to provide important biophysical insight into both the assembly and characteristics of natural membrane-free intracellular compartments in all cell-types. Given the increased importance that these compartments are now thought to have in many biological processes, constructing and studying such objects synthetically has the potential to provide important biophysical insight for this new class of intracellular compartments.

References

Battaile CC. 2008. The Kinetic Monte Carlo method: Foundation, implementation, and application. *Computer Methods in Applied Mechanics and Engineering*, Recent Advances in Computational Study of Nanostructures **197**:3386–3398. doi:10.1016/j.cma.2008.03.010

Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM. 1998. Localization of ASH1 mRNA particles in living yeast. *Molecular cell* **2**:437–445.

Blake WJ, Balázsi G, Kohanski MA, Isaacs FJ, Murphy KF, Kuang Y, Cantor CR, Walt DR, Collins JJ. 2006. Phenotypic Consequences of Promoter-Mediated Transcriptional Noise. *Molecular Cell* **24**:853–865. doi:10.1016/j.molcel.2006.11.003

Chong S, Chen C, Ge H, Xie XS. 2014. Mechanism of Transcriptional Bursting in Bacteria. *Cell* **158**:314–326. doi:10.1016/j.cell.2014.05.038

Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, Lawrence JB. 2009. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell* **33**:717–726. doi:10.1016/j.molcel.2009.01.026

Elowitz MB. 2002. Stochastic Gene Expression in a Single Cell. *Science* **297**:1183–1186. doi:10.1126/science.1070919

Femino AM, Fay FS, Fogarty K, Singer RH. 1998. Visualization of Single RNA Transcripts in Situ. *Science* **280**:585–590. doi:10.1126/science.280.5363.585

Fox AH, Nakagawa S, Hirose T, Bond CS. 2018. Paraspeckles: Where Long Noncoding RNA Meets Phase Separation. *Trends in Biochemical Sciences* **43**:124–135. doi:10.1016/j.tibs.2017.12.001

Fusco D, Accornero N, Lavoie B, Shenoy SM, Blanchard J-M, Singer RH, Bertrand E. 2003. Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. *Current Biology* **13**:161–167.

Golding I, Paulsson J, Zawilski SM, Cox EC. 2005. Real-Time Kinetics of Gene Activity in Individual Bacteria. *Cell* **123**:1025–1036. doi:10.1016/j.cell.2005.09.031

Gong T, Luo Q, Xu X, Yu J, Dong D, Lv H, Yuan P, Chen C, Yin J, Tai L, Zhu X, Liu Q, Long S, Liu M. 2018. Classification of Three-Level Random Telegraph Noise and Its Application in Accurate Extraction of Trap Profiles in Oxide-Based Resistive Switching Memory. *IEEE Electron Device Letters* **39**:1302–1305. doi:10.1109/LED.2018.2858245

Hyman AA, Weber CA, Jülicher F. 2014. Liquid-Liquid Phase Separation in Biology. *Annual Review of Cell and Developmental Biology* **30**:39–58. doi:10.1146/annurev-cellbio-100913-013325

Jones D, Elf J. 2018. Bursting onto the scene? Exploring stochastic mRNA production in bacteria. *Current Opinion in Microbiology* **45**:124–130. doi:10.1016/j.mib.2018.04.001

Katz N, Cohen R, Solomon O, Kaufmann B, Atar O, Yakhini Z, Goldberg S, Amit R. 2018. An in Vivo Binding Assay for RNA-Binding Proteins Based on Repression of a Reporter Gene. *ACS Synth Biol* **7**:2765–2774. doi:10.1021/acssynbio.8b00378

Larson DR, Singer RH, Zenklusen D. 2009. A Single Molecule View of Gene Expression. *Trends Cell Biol* **19**:630–637. doi:10.1016/j.tcb.2009.08.008

Larson DR, Zenklusen D, Wu B, Chao JA, Singer RH. 2011. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science* **332**:475–478. doi:10.1126/science.1202142

Li X, Fu X-D. 2019. Chromatin-associated RNAs as facilitators of functional genomic interactions. *Nature Reviews Genetics* 1. doi:10.1038/s41576-019-0135-1

Lionnet T, Singer RH. 2012. Transcription goes digital. *EMBO Rep* **13**:313–321. doi:10.1038/embor.2012.31

McAdams HH, Arkin A. 1997. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences* **94**:814–819.

Miller S, Childers D. 2012. Probability and Random Processes With Applications to Signal Processing and Communications. San Diego, CA, USA: Elsevier Science & Technology Books.

Paulsson J. 2005. Models of stochastic gene expression. *Physics of Life Reviews* **2**:157–175. doi:10.1016/j.plrev.2005.03.003

Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. 2008. Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods* **5**:877–879. doi:10.1038/nmeth.1253

Sanchez A, Garcia HG, Jones D, Phillips R, Kondev J. 2011. Effect of Promoter Architecture on the Cell-to-Cell Variability in Gene Expression. *PLoS Computational Biology* **7**:e1001100. doi:10.1371/journal.pcbi.1001100

Sanchez A, Golding I. 2013. Genetic determinants and cellular constraints in noisy gene expression. *Science* **342**:1188–1193. doi:10.1126/science.1242975

So L, Ghosh A, Zong C, Sepúlveda LA, Segev R, Golding I. 2011. General properties of transcriptional time series in Escherichia coli. *Nature Genetics* **43**:554–560. doi:10.1038/ng.821

Spector DL, Lamond AI. 2011. Nuclear speckles. *Cold Spring Harb Perspect Biol* **3**. doi:10.1101/cshperspect.a000646

Staněk D, Fox AH. 2017. Nuclear bodies: news insights into structure and function. *Curr Opin Cell Biol* **46**:94–101. doi:10.1016/j.ceb.2017.05.001

Taniguchi Y, Choi PJ, Li G-W, Chen H, Babu M, Hearn J, Emili A, Xie XS. 2010. Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science* **329**:533–538. doi:10.1126/science.1188308

Tutucci E, Livingston NM, Singer RH, Wu B. 2018a. Imaging mRNA In Vivo, from Birth to Death. *Annual Review of Biophysics* **47**:85–106. doi:10.1146/annurev-biophys-070317-033037

Tutucci E, Vera M, Biswas J, Garcia J, Parker R, Singer RH. 2018b. An improved MS2 system for accurate reporting of the mRNA life cycle. *Nature Methods* **15**:81–89. doi:10.1038/nmeth.4502

Wells AL, Condeelis JS, Singer RH, Zenklusen D. 2007. Imaging Real-Time Gene Expression in Living Systems with Single-Transcript Resolution: Construct Design and Imaging System Setup. *Cold Spring Harbor Protocols* **2007**:pdb.top28. doi:10.1101/pdb.top28

Acknowledgements

This project received funding from the I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation (Grant No. 152/11); and Marie Curie Reintegration Grant No. PCIG11-GA-2012-321675, and by the European Union's Horizon 2020 Research And Innovation Programme under grant agreement no. 664918 - MRG-Grammar.

Author contributions

N.G. designed, carried out the experiments, and performed the analysis of the data. N.K. designed the RNA binding sites cassette. R.A. and Y.S. supervised the study. N.G. and R.A. wrote the manuscript.

Competing interest

The authors declare that there is no conflict of interest regarding the publication of this article

Materials & Correspondence

Correspondence and requests for materials should be addressed to roeeamit@technion.ac.il.

Supplementary Information

Materials and Methods

Plasmids and BAC preparation

Bacterial strains

- E. coli BL21-DE3 cells which encode the gene for T7 RNAP downstream from an inducible pLac/Ara promoter.
- E. coli TOP10 (Invitrogen, Life Technologies, Cergy-Pontoise) was used for cloning experiments.

Addgene plasmids

- pCR4-24XPP7SL was a gift from Robert Singer (Addgene plasmid # 31864; http://n2t.net/addgene:31864; RRID: Addgene_31864)
- pBAC-lacZ was a gift from Keith Joung (Addgene plasmid # 13422; http://n2t.net/addgene:13422; RRID: Addgene_13422)

Construction of the pBAC-5xQβ-4xPP7 binding sites array

The T7 promoter followed by the binding sites sequence: cctaggcgattatgacgttattctactttgattgtgatgcatgtctaagacagcatcgcctgctggtcgtgactaaggagtttatatggaaacccttacga gacaatgctaccttaccggtcgggcccacttgtttttacccatgatgcatgtctaagacagcatcgcctgctggtcgtgactaaggagtttatatggaa acccttagaaacagccgtcgccttgaagccgagaacaatgcatgtctaagacagcatatggattgcctgtctgttaaggagtttatatggaaacccctta catcaggcttcgcagtatgcaacgcttgcgatgcatgtctaagacagcatttcaccgctttcctaagtaaggagtttatatggaaacccttagtactaac tcgcagatgcatgtctaagacagcatcagaaacgtcacgtcctggc. (Qβ and PP7 binding sites marked in red and green respectively),

was ordered from GenScript, Inc. (Piscataway, NJ), as part of a Puc plasmid, flanked by EcoRI and HindIII restriction sites.

pBAC-LacZ backbone plasmid was obtained from Addgene plasmid #13422. Both insert and vector were digested using the above restriction sites and ligated to form pBAC-5xQβ-4PP7-lacZ.

Sample preparation

BL21 cells expressing the BAC-Qβ-5x_PP7-4x and the Qβ-mCherry expression plasmid were grown O/N in Luria Broth (LB), in 37□ with appropriate antibiotics (CM, AMP). O/N culture was diluted 1:100 into 3μl solution of BioAssay (BA)-LB (95%-5% v:v) with appropriate antibiotics, and induced with 1μl IPTG (final concentration 1mM) and 1.5μl *N*-butanoyl-l-homoserine lactone (C4-HSL) (final concentration 60μM) to induce expression of T7 RNA polymerase and the RBP-FP respectively. Culture was shaken for 3 hours in 37□ before being applied to a gel slide (3μl Dulbecco's Phosphate-Buffered Saline (Biological Industries) x1, mixed with 0.045g SeaPlaque low melting Agarose (Lonza, Switzerland), heated for 20 seconds and allowed to cool for 25 minutes). 1.5 μl cell culture was deposited on a gel slide and allowed to settle for an additional 30 minutes before imaging.

Image Analysis

A single experiment was carried out by tracking a field of view for 60 minutes on Nikon Eclipse Ti-E epifluorescent microscope (Nikon, Japan) using the Andor iXon Ultra EMCCD camera at 6 frames-per-minute with a 200 msec exposure time per frame to avoid photo-bleaching and sufficient recovery of fluorescence signal. Excitation was performed at 585 [nm] wavelength by a CooLED (Andover, UK) PE excitation system. Subsequently, the brightest spots (top 10%) in the field of view were tracked using the algorithm developed by Sbalazarini and Koumoutsakos (Sbalzarini and Koumoutsakos, 2005). A typical field of view usually contained dozens of cells, a portion of which were not fluorescent while others presented distinct bright speckles, localized at the cell poles as expected from literature (Fei and Sharma, 2018).

The tracking data, (x,y,t coordinates of the bright spots centroids), together with the raw microscopy images were fed to a custom built Matlab (The Mathworks, Natick, MA) script designed to normalize the relevant spot data. Normalization was carried out as follows: for each bright spot, a 20-pixel wide sub-frame was extracted from the field of view, with the spot at its center. Each pixel in the sub-frame was classified to one of three categories according to its intensity value. The brightest pixels were classified as 'spot region' and would usually appear in a cluster, corresponding to the spot itself. The dimmest pixels were classified as 'dark background', corresponding to an empty region in the field of view. Lastly, values in between were classified as 'cell background'. Classification was done automatically using Otsu's method (Otsu, 1979), (Figure S1). From each sub-frame, two values were extracted, the mean of the 'spot region' pixels and the mean of the 'cell background' pixels, corresponding to spot intensity value and cell intensity value. This was repeated for each spot from each frame in the data, resulting in sequences of intensity vs. time. Sequences were filtered for high frequency noise by a moving average filter with a window of 10 time points. Normalization was done by subtracting the cell intensity values from the spot intensity values.

Identifying bursts events

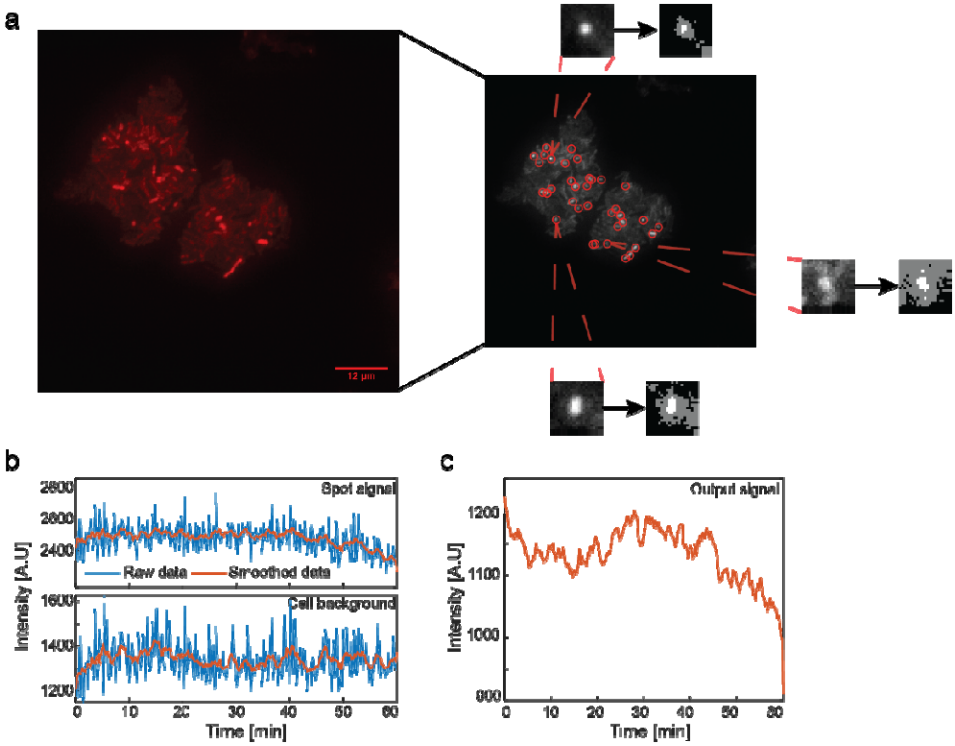

Figure S1: Image analysis and signal acquisition. **a.** Leftmost image – Typical field of view showing bright fluorescent spots, along with bright cells. Dim red background is cells which are not fluorescing. Middle image – Capturing top 10% of bright spots (marked in red circles). Presented are three example sub-frames, each showing a spot and its immediate surroundings. For each sub-frame, each pixel is classified to one of three intensity levels – bright spot, cell background and dark background corresponding to white, grey and black colors in the segmented images. **b.** Sample spot signal (top), and its corresponding cell background signal. Blue line is raw data, orange line is

smoothed data after a 10-point moving average. **c.** Output signal resulting from the subtraction of the background signal from the spot signal.

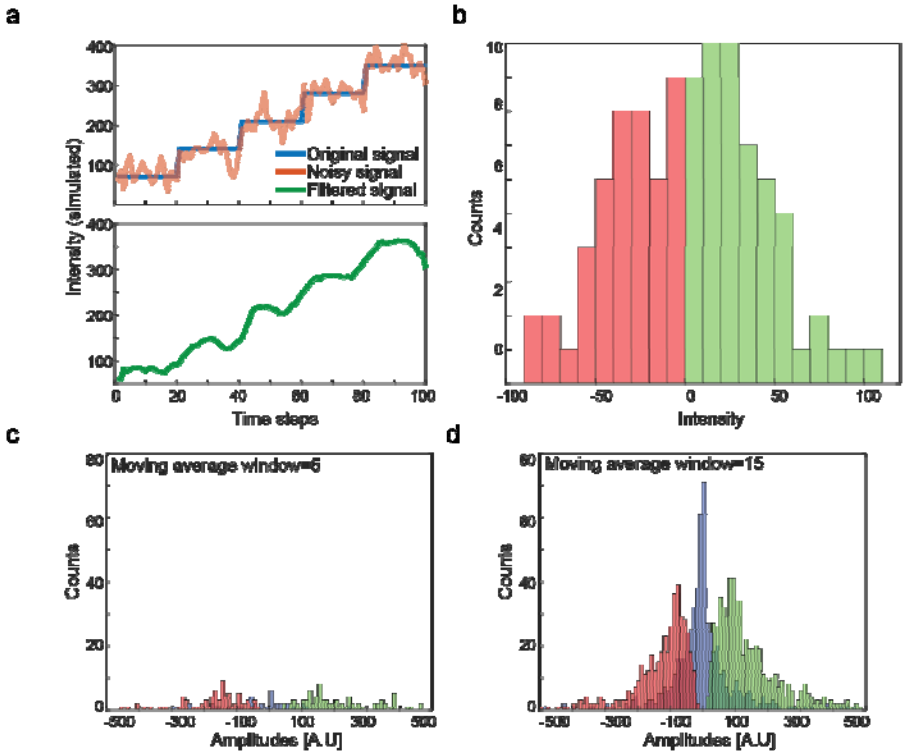

Figure S2: Identification of burst events. **a.** Effects of noise, and noise-filtering, on a bursty signal. Underlying simulated signal is comprised of instantaneous increases in intensity (top plot – blue line), however, this feature disappears with the addition of Gaussian noise with a standard deviation of 30 [A.U] (top plot - orange line). The noisy signal was filtered with a 10-point moving average filter and appears continuous (bottom plot). **b.** Distribution of intensity difference between successive time points of the simulated signal appearing in (**a**), showing a slight bias towards the positive side (54% of the derivative is positive). This distribution is used to calculate the threshold for a significant event. **c-d.** Amplitude distribution of the Qβ-5x experimental data analyzed with different moving average windows. **c.** 5 time-points moving average. **d.** 15 time-points moving average.

We define a *burst* as a sudden change or shift in the level of the speckle's fluorescence intensity leading to either a sustainable higher or lower new signal level (Fig. S2A-top). To identify such shifts in the base-line fluorescence intensity, we use a moving-average window of ten points to smooth the data. The effect of such an operation is to bias the fluctuations of the smoothed noisy signal in the immediate vicinity of the bursts towards either a gradual increase or decrease in the signal (Figure S2a-bottom). Random fluctuations, which do not settle on a new baseline level are not expected to generate a gradual and continuous increase over multiple time-points in a smoothed signal. As a result, we search for contiguous segments of gradual increase or decrease, and record only those whose probability for occurrence are 1 in 1000 or less given a Null hypothesis of randomly

fluctuating noise. As an example, we consider a constant base-line intensity amplitude with white Gaussian noise. For any particular time-point the probability that the next point will exhibit either a stronger or weaker signal is 50% respectively. Since the noise is independent and identically distributed (IID), the probability for an increase in the signal lasting 10 consecutive time points is $\frac{1}{2^{10}} = \frac{1}{1024}$. Given that our traces typically take place over ~60 minutes and include 360 frames in total, we expect <1 such random 1 in 1000 events per trace. This probability remains roughly the same even after the moving average smoothing.

For the general case, the underlying empirical signal is not constant, and may be either be trending up or down. Therefore, it is necessary to normalize the probability per signal and determine a threshold – *m*, such that the probability for a consecutive increase of *m* time points is $\frac{1}{1024}$ given the underlying signal trend. For every trace, we first compute the intensity difference distribution (Fig. S2b). The probability that the signal increases at a random time point is calculated by summing the number of points in which the signal derivative is positive and dividing by the total length of the signal.

$$p = \frac{length\left(\frac{dS}{dt} > 0\right)}{length(S)} \tag{1}$$

This in turn allows us to compute *m* as follows:

$$p^m = \frac{1}{2^{10}} \Rightarrow m log_2(p) = -10 \Rightarrow m = -\frac{10}{\log_2(p)} \tag{2}$$

The threshold is calculated for each signal separately and is usually in the range of 7-13 time points. An analogous threshold is calculated for decrements in the signal and is usually in the range $[m-1, m+1]$. We mark each trace with the number of events that exceed this threshold and define those as bursts.

While the choice of smoothing window is somewhat arbitrary, it was chosen to be sufficiently large to allow for both an identification of a gradual increase or decrease due to the burst and a stable base-line shift, without compromising our ability to properly characterize the signal on a longer time-scale. To check, that our choice of smoothing parameter does not affect the interpretation of the data, we applied both shorter and longer moving-average windows showing that the over-all nature of the results remains unchanged (Fig. S2c-d). The main difference between the averaging windows lies in the number of significant events identified. The 5-point window results in a total of 91 positive, and 86 negative events found, while the 15-point window results in 577 positive and 424 negative events.

Measuring the effect of background choice on observed signal

The foundation of the analysis method is the tracking and gathering of intensity data from fluorescent spots in microscopy time-lapse movies. This is done by an image processing scheme designed to be as simple as possible, to mitigate undesirable noise effects that might originate from such operations. However, arbitrary choices were made in the process that might themselves bring about unwanted noise into the statistics. Namely, the selection of the sub-frame size used to calculate the background intensity for normalization. A large sub-frame would undoubtedly include other cells, with possibly different spots of themselves, inserting a bias to the background intensity signal. On the other hand, a small sub-frame might not have a sufficient spot-to-background area ratio, resulting in an underestimated background signal.

To test whether this has a meaningful effect on the data, we repeated the entire data analysis process with a sub-frame length of 10 pixels, i.e. a 10-pixel-wide sub frame was extracted from the FOV, with the spot at its center, in contrast to the 20-pixel wide described in the methods section. Smaller lengths would cause a loss of spot intensity data and larger lengths would bring about biases originating from other cells and therefore were not considered.

Figure S3 shows main comparison points that were considered for this test. Panel A shows a sample comparison between two sub-frames with 20-pixel width (top) and 10-pixel width (bottom). Panel b shows a sample intensity signal (after moving average and background normalization). This plot directly shows the effects of a small sub-pixel, namely the difference in values between the two signals, resulting from the number of pixels being classified as ‘spot’ being smaller in the 10-pixel wide sub-frame. Despite this difference, both signals show similar behavior (i.e. similar ‘hills’ and ‘valleys’).

Panel c shows a sample comparison between positive amplitude statistics for the PP7-24x experiments calculated with the two sub-frame sizes. The comparison does reveal a slight difference in amplitude values however these are not significant from a statistical viewpoint.

In a similar manner, panel d shows no discernible difference in the durations of quiescent periods between the two background analysis variations. Lastly, panels e-f show fittings of the amplitude data gathered using 10-pixel wide sub-frames, to theoretical Poisson probability density functions with rates 1,2,3, similarly to the process described in eq. 1 of the main text. The optimal K_0 values in this case differ by no more than 10% in all cases, compared to their 20-pixel-wide counterparts presented in Figure 2.

Overall, this control comes to show that the statistical results presented in the text, and the conclusions drawn from them do not suffer any change upon altering this step in the data processing. That being said, this control does show that absolute quantitative results, i.e. intensity value per binding site, will be difficult to correctly calculate without further work and calibrations.

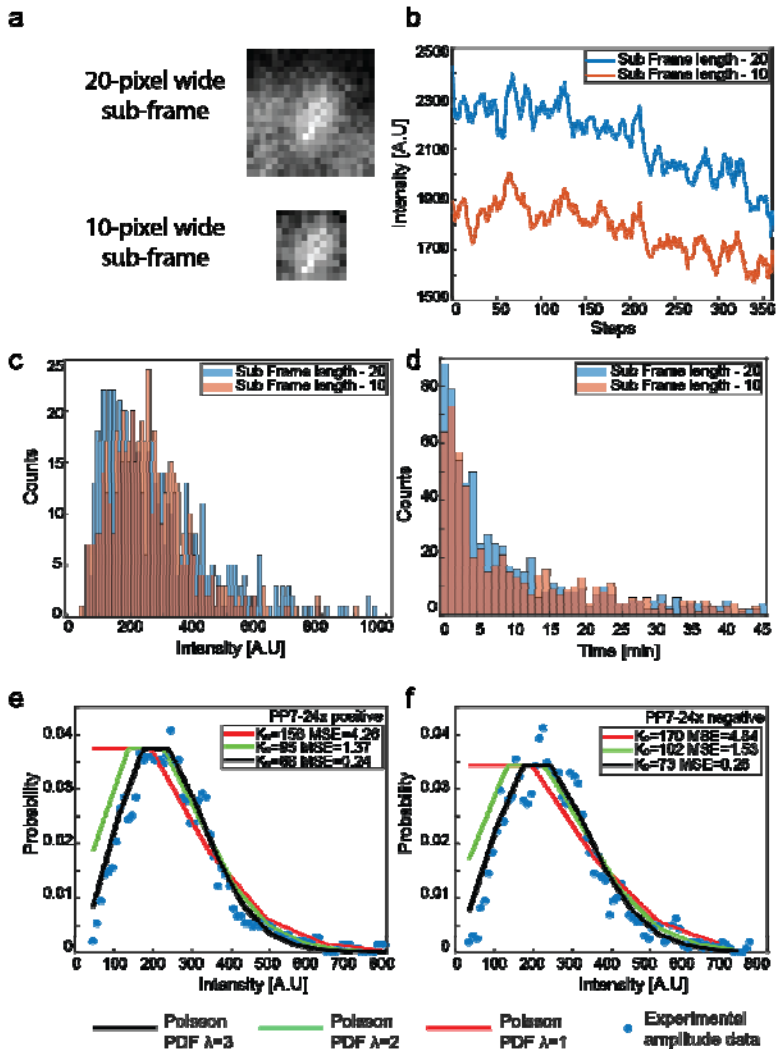

Figure S3: Background selection controls. **a.** Sample sub-frames extracted from an experiment showing 10- and 20-pixel wide sub frames. **b.** Intensity plots of a single spot calculated with a 10- and 20-pixel wide sub frames. **c.** Increment amplitude distribution of the PP7-24x calculated with a 10- and 20-pixel wide sub frames. **d.** Quiescent segments durations of the PP7-24x. **e-f.** Poisson fits for positive amplitudes **(e),** and negative amplitudes **(f)**, following the same fitting procedure as described in the main text.

Three state model sequential order

We tested whether burst events had a "memory" bias, thus generating a preference for a particular burst-type following another particular event. We check both next events (Fig S4a-c), and events following a quiescent event (Fig. S4d-f). In the case of negative and positive bursts (Fig. S4a,c), a clear preference for a quiescent event is observed in the data, while no memory is observed for burst events following a quiescent event (Fig. S4b,e-f).

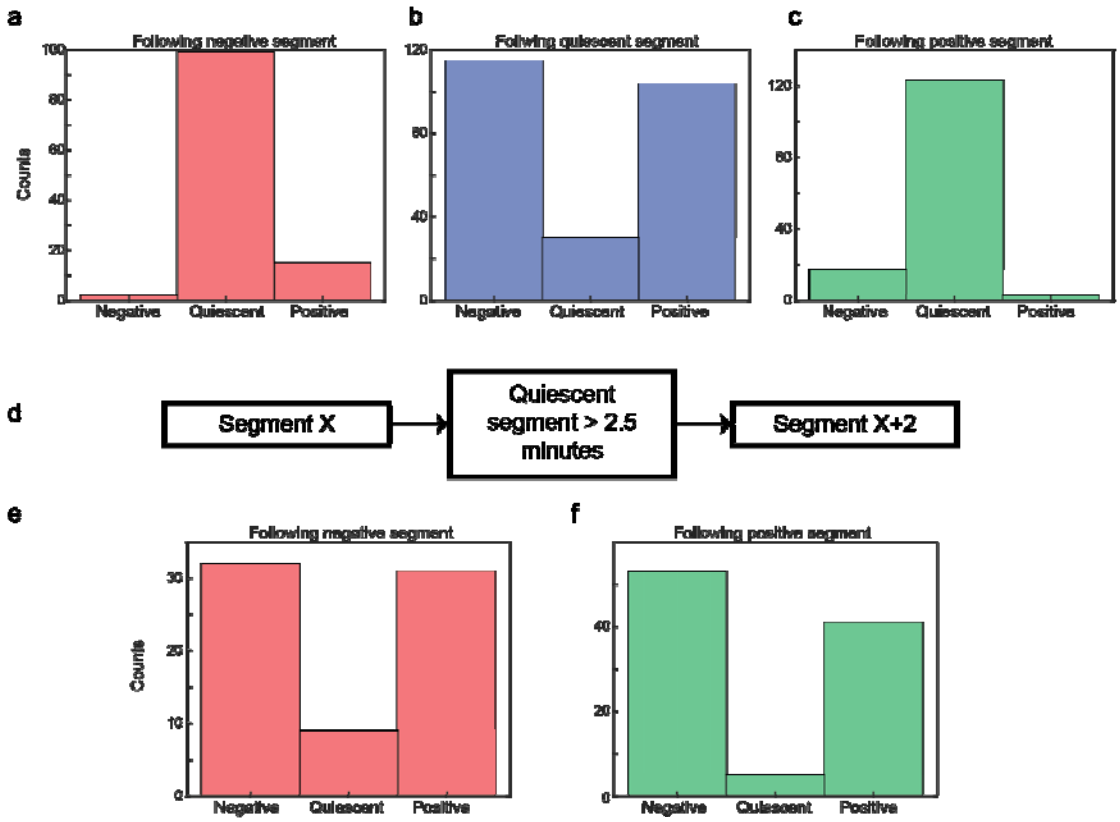

Figure S4: Distribution of burst sequence. **a-c.** Distribution of successive bursts according to current burst type: negative (**a**), quiescent (**b**), positive (**c**). **d.** Schematic of the data presented in (**e,f**). Distribution of bursts following a quiescent event longer than 2.5 minutes, which follows a particular burst-type. **e.** Following a negative segment. **f.** Following a positive segment.

Mathematical derivation of the four-state degradation burst model

The mathematical derivations presented here are based on the work done by Sanchez et al(Sanchez et al., 2011). The general form of the model contains two stochastic variables: the number of mRNA molecules *n*, and the state of the system *S*, with the system being the two major biochemical processes controlling speckle fluorescence, transcription and degradation.

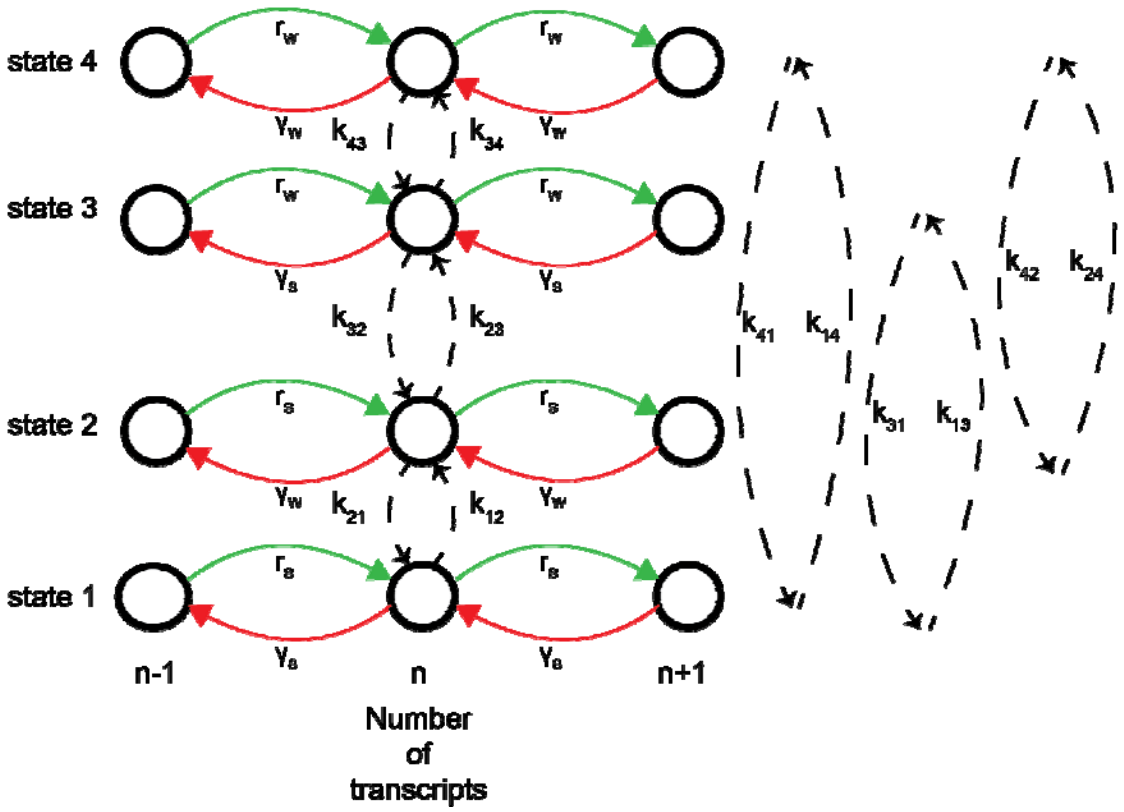

Figure S5: Scheme of the general four-state model. Black circles represent the possible states of the system for any *n>1* number of RNA molecules. Red and green arrows indicate degradation and transcription processes, respectively, governing the number of molecules. Dashed arrows stand for transitions between system states.

The general model (Figure S5), is comprised of four possible states for the system, strong transcription and strong degradation; strong transcription and weak degradation; weak transcription and strong degradation; weak transcription and weak degradation. The rate of transcription is determined by the state of the promoter (open/closed), and the rate of degradation is similarly determined by the accessibility of mRNA molecules to Rnase binding or the availability of the degradation enzymes.

The temporal dynamics of this stochastic system are given by the master equation, which can be derived by listing all possible reactions leading to a change in *S* or in *n*.

$$
\begin{aligned}
\frac{d}{dt}P_{1,n} &= r_S P_{1,n-1} + \gamma_S(n+1)P_{1,n+1} + k_{21}P_{2,n} + k_{31}P_{3,n} + k_{41}P_{4,n} - P_{1,n}[\gamma_S n + r_S + k_{12} + k_{13} + k_{14}] \\
\frac{d}{dt}P_{2,n} &= r_S P_{2,n-1} + \gamma_W(n+1)P_{2,n+1} + k_{12}P_{1,n} + k_{32}P_{3,n} + k_{42}P_{4,n} - P_{2,n}[\gamma_W n + r_S + k_{21} + k_{23} + k_{24}] \\
\frac{d}{dt}P_{3,n} &= r_W P_{3,n-1} + \gamma_S(n+1)P_{3,n+1} + k_{13}P_{1,n} + k_{23}P_{2,n} + k_{43}P_{4,n} - P_{3,n}[\gamma_S n + r_W + k_{31} + k_{32} + k_{34}] \\
\frac{d}{dt}P_{4,n} &= r_W P_{4,n-1} + \gamma_W(n+1)P_{4,n+1} + k_{14}P_{1,n} + k_{24}P_{2,n} + k_{34}P_{3,n} - P_{4,n}[\gamma_W n + r_W + k_{41} + k_{42} + k_{43}]
\end{aligned}
\tag{1}
$$

Here the subscripts ‘S’ and ‘W’, stand for strong and weak respectively, indicating the strength of the biochemical process.

Using the following definitions:

$$
\vec{P}(n) = \begin{pmatrix} P_{1,n} \\ P_{2,n} \\ P_{3,n} \\ P_{4,n} \end{pmatrix}
$$

$$
\hat{K} = \begin{pmatrix} -[k_{12} + k_{13} + k_{14}] & k_{21} & k_{31} & k_{41} \\ k_{12} & -[k_{21} + k_{23} + k_{24}] & k_{32} & k_{42} \\ k_{13} & k_{23} & -[k_{31} + k_{32} + k_{34}] & k_{43} \\ k_{14} & k_{24} & k_{34} & -[k_{41} + k_{42} + k_{43}] \end{pmatrix} \tag{2}
$$

$$
\hat{R} = \begin{pmatrix} r_S & 0 & 0 & 0 \\ 0 & r_S & 0 & 0 \\ 0 & 0 & r_W & 0 \\ 0 & 0 & 0 & r_W \end{pmatrix} ; \ \hat{\Gamma} = \begin{pmatrix} \gamma_S & 0 & 0 & 0 \\ 0 & \gamma_W & 0 & 0 \\ 0 & 0 & \gamma_S & 0 \\ 0 & 0 & 0 & \gamma_W \end{pmatrix}
$$

We can write the above equations in matrix form.

$$
\frac{d}{dt}\vec{P}(n) = \left[\hat{K} - \hat{R} - n\hat{\Gamma}\right]\vec{P}(n) + \hat{R}\vec{P}(n-1) + (n+1)\hat{\Gamma}\vec{P}(n+1) \tag{3}
$$

The matrix $\hat{K}$ describes the transition rates between system states, the matrices $\hat{R}$ and $\hat{\Gamma}$ describe the rates of transcription initiation rates, and degradation rates, respectively.

At this point, we can derive equations from which the first and second moments of the mRNA distribution in steady state can be computed. The derivation is similar to the one presented by Sanchez et al (Sanchez et al., 2011) and therefore will not be repeated here. Since it is convenient to write the resulting equations in terms of partial moments of the mRNA distribution, they will be defined here as:

4

$$\vec{n}_0 = \sum_{n=0}^{\infty} n^0\, \vec{P}(n) = \begin{pmatrix} \sum_{n=0}^{\infty} n^0 P_{1,n} \\ \sum_{n=0}^{\infty} n^0 P_{2,n} \\ \sum_{n=0}^{\infty} n^0 P_{3,n} \\ \sum_{n=0}^{\infty} n^0 P_{4,n} \end{pmatrix} = \begin{pmatrix} P_1 \\ P_2 \\ P_3 \\ P_4 \end{pmatrix} \;\; ; \;\; \vec{n}_1 = \sum_{n=0}^{\infty} n\, \vec{P}(n) = \begin{pmatrix} \sum_{n=0}^{\infty} nP_{1,n} \\ \sum_{n=0}^{\infty} nP_{2,n} \\ \sum_{n=0}^{\infty} nP_{3,n} \\ \sum_{n=0}^{\infty} nP_{4,n} \end{pmatrix};$$

$$\vec{n}_2 = \sum_{n=0}^{\infty} n^2\, \vec{P}(n) = \begin{pmatrix} \sum_{n=0}^{\infty} n^2 P_{1,n} \\ \sum_{n=0}^{\infty} n^2 P_{2,n} \\ \sum_{n=0}^{\infty} n^2 P_{3,n} \\ \sum_{n=0}^{\infty} n^2 P_{4,n} \end{pmatrix}$$

$\vec{n}_0$ can be found from the equation:

$$\hat{K}\vec{n}_0 = 0 \qquad (5)$$

Together with the normalization condition:

$$P_1 + P_2 + P_3 + P_4 = 1 \qquad (6)$$

$\vec{n}_1$ from the equation:

$$\left(\hat{K} - \hat{\Gamma}\right)\vec{n}_1 + \hat{R}\vec{n}_0 = 0 \qquad (7)$$

$\vec{n}_2$ from the equation:

$$\hat{K}\vec{n}_2 + \hat{R}(2\vec{n}_1 + \vec{n}_0) + \hat{\Gamma}(\vec{n}_1 - 2\vec{n}_2) = 0 \qquad (8)$$

Using these equations, the Fano factor can be extracted and computed numerically as shown in Figure 5. Biological assumptions can be made to simplify the model in order to simulate it using Monte Carlo methods. To this end, we give the different states biological meaning.

Model State	Biological state
1	Strong transcription; strong degradation
2	Strong transcription; weak degradation
3	Weak transcription; strong degradation
4	Weak transcription; weak degradation

Table S1: Biological meaning of model states

First, we assume that a simultaneous transition of both transcription and degradation is unlikely, i.e. $k_{14}, k_{41}, k_{23}, k_{32} = 0$. Second, transition rates in a specific system do not change between the different states, meaning that, switching of a promoter from ‘open’ to ‘closed’ will have the same rate regardless of degradation and vice versa. In mathematical terms:

$k_{12} = k_{34} = k_{off}^{d}$	Strong degradation → Weak degradation
$k_{21} = k_{43} = k_{on}^{d}$	Weak degradation → Strong degradation
$k_{13} = k_{24} = k_{off}^{t}$	Strong transcription → Weak transcription
$k_{31} = k_{42} = k_{on}^{t}$	Weak transcription → Strong transcription

Table S2: Parameter designations for biological model

Finally, for simplicity we set $\gamma_W = 0$, as we believe this value to be insignificant compared to γ_L

The matrices then take the following form:

$$\hat{K} = \begin{pmatrix} -\left[k_{off}^{d} + k_{off}^{t}\right] & k_{on}^{d} & k_{on}^{t} & 0 \\ k_{off}^{d} & -\left[k_{on}^{d} + k_{off}^{t}\right] & 0 & k_{on}^{t} \\ k_{off}^{t} & 0 & -\left[k_{on}^{t} + k_{off}^{d}\right] & k_{on}^{d} \\ 0 & k_{off}^{t} & k_{off}^{d} & -\left[k_{on}^{t} + k_{on}^{d}\right] \end{pmatrix} \qquad 9$$

$$\hat{R} = \begin{pmatrix} r_S & 0 & 0 & 0 \\ 0 & r_S & 0 & 0 \\ 0 & 0 & r_W & 0 \\ 0 & 0 & 0 & r_W \end{pmatrix} ; \ \hat{\Gamma} = \begin{pmatrix} \gamma_S & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma_S & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

We note that we cannot differentiate between all four possible states experimentally, as two of these result in similar mRNA behavior. Depending on parameter choices, the indistinguishable pair could be states 1 and 2 (assuming the effect of transcription is stronger regardless of degradation state), or states 1 and 4 (assuming the effect of transcription to be similar to that of degradation).

Monte Carlo simulation details

Monte Carlo simulations of the two-state and four-state models were generated with the following parameters:

Kinetic transition	Mathematical Notation	Biological meaning	Value [s^{-1}]
Closed promoter -> Open promoter	k_{on}^{t}	RNAP association	0.0027
Open promoter -> Closed promoter	k_{off}^{t}	RNAP disassociation	0.0023
Open promoter -> mRNA	r_S	Strong transcription	0.04
Close promoter -> mRNA	r_W	Weak transcription	0.003
mRNA -> Quiescent state	k_{off}^{d}	Halted degradation	0.2
Quiescent state -> mRNA	k_{on}^{d}	Resumed degradation	0.4
mRNA degradation	γ_S	Degradation rate	0.011

Table S3: Parameters used for the kinetic modelling

The above association/disassociation rates were gathered from the work by Sanchez et al (Sanchez et al., 2011) , and are valid for the pLac promoter in E.coli. These rates do not represent empirical biological kinetic data, as no *in-vivo* association/disassociation rates could be found for the T7 RNAP in literature. Instead, these should be regarded as qualitative only, representing the order of magnitude to the association of a protein (e.g. RNAP, transcription factor) to a promoter, in a crowded cellular environment.

The T7 RNAP transcription rate, originates from empirical in-vitro transcription rates of ~40 [bp/sec] (Nichola, 2010) and therefor might be overestimating the actual rate in the cell. The rate of 0.04 [mRNA/sec] was calculated under the assumption of 1000 [bp/mRNA]. The Quiescent state transition rates were determined to be the rates in which the distribution of mRNA molecules is similar to that of the two-state model. It should be noted that these are not unique in any way, different rates in the same order of magnitude would still hold the condition of similar mRNA distribution.

Analysis of experimental signal as a telegraph process

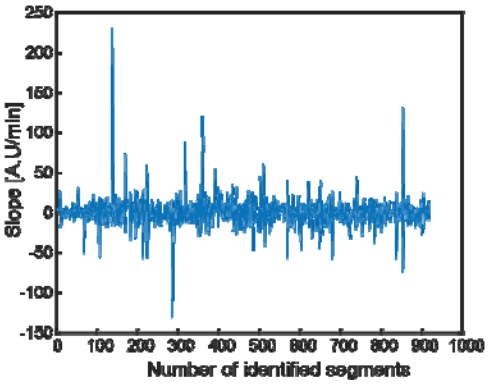

Figure S6: Experimental slope-signal representing a three-state telegraph noise.

One of the tests used to determine the validity of the three state model is the comparison of the power spectral density (PSD) of the experimental data to that of the simulated data generated by Monte Carlo simulations. In order to compute the PSD for the experimental, and simulated data, we computed the slope of the identified segments from all intensity sequences. The slopes were concatenated to a single sequence, with the assumption of memory-lessness, to generate a telegraph-model-like signal. Figure S6 shows the signal generated from experimental data of the qβ_5x-PP7_4x slncRNA. In terms of a three-state telegraph model, the positive slopes correspond to state '1', negative to state '-1', and anything in between to '0' state.

Supplementary References

Fei J, Sharma CM. 2018. RNA localization in bacteria. *Microbiol Spectr* **6**. doi:10.1128/microbiolspec.RWR-0024-2018

Nichola D. 2010. Determining the Rate of Transcription of T7 RNA Polymerase Using Single Molecule Fluorescence Imaging. *Theses, Dissertations and Capstones*.

Otsu N. 1979. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics* **9**:62–66. doi:10.1109/TSMC.1979.4310076

Sanchez A, Garcia HG, Jones D, Phillips R, Kondev J. 2011. Effect of Promoter Architecture on the Cell-to-Cell Variability in Gene Expression. *PLoS Computational Biology* **7**:e1001100. doi:10.1371/journal.pcbi.1001100

Sbalzarini IF, Koumoutsakos P. 2005. Feature point tracking and trajectory analysis for video imaging in cell biology. *J Struct Biol* **151**:182–195. doi:10.1016/j.jsb.2005.06.002