**Supplementary Information**

Formation of synthetic RNP granules using engineered phage-coat-protein -RNA complexes

**Authors:** Naor Granik1, Noa Katz2, Or Willinger2, Sarah Goldberg2, and Roee Amit2,3\*

**Affiliations:**

1 Department of Applied Mathematics, 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.

**Supplementary Results**

****

**Figure S1**: **QQ-plots of modified Poisson fits**. Quantile-quantile (QQ) plots showing agreement between sample data (experimental observations) and the theoretical Poisson distribution for the fits shown in figure 1d.

****

**Figure S2**: **PCP-24x granules amplitude distribution**. **a**, Empirical amplitude distributions gathered from 391 traces *in vivo* from cells expressing the PCP-24x slncRNA together with the tdPCP-mCherry protein. Green – Positive amplitudes (insertion events), red – negative amplitudes (shedding events), blue – unclassified events.

****

**Figure S3**: **Fitting of amplitude data to Poisson distributions.** (a-b) Poisson functions fits for the amplitude distribution of insertion events assuming 1, 2, or 3 mean events (λ values). MSE values represent mean squared error between the empirical distribution and the theoretical modified Poisson functions. **a**, Data collected from 255 PCP-4x/QCP-5x signal traces. **b**, Data collected from 391 PCP-24x signal traces.

**Supplementary Methods**

**Image Analysis**

The brightest spots (top 10%) in the field of view were tracked over time and space via the imageJ MosaicSuite plugin1–3. A typical field of view usually contained dozens of granules (*in-vitro*) or cells containing puncta (*in vivo*) (Figure S3a,b).

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 14-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’ (Figure S3c).We note that for the *in vitro* experiments the ‘dark background’ and ‘cell background’ pixel groups yield similar intensity values. This, however, does not affect the performance of the algorithm for *in vitro* experiments. Classification was done automatically using Otsu’s method4. 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 for the spot itself and for the cell background. (Figure S3d)



**Figure S4**: **Image processing scheme**. **a**, Raw microscopy image showing bacterial cells containing bright spots. **b**, Bright spots are identified and their position over time and space is recorded. **c**, The environment of each spot is classified into 3 regions, based on intensity values. The brightest pixels are classified as ‘spot’ (marked in white), the darkest pixels are classified as ‘dark background’ indicating empty space, and pixels with intermediate values are classified as cell background (marked in gray). **d**, The mean values of the spot pixels and cell background pixels are recorded over time resulting in the spot signal (green) and cell signal (blue). The spot signal is then normalized to remove photobleaching and global background effects (red).

**Signal Analysis**

We assume a noise model comprised of both additive and exponential components, corresponding to fluorescent proteins (bound or unbound) not relating to the spot itself, and photobleaching. This can be described as follows:

 

 

where is the observed spot signal,  is the underlying spot signal which we try to extract,  is the observed cell background signal, is the underlying background signal and  is the photobleaching component.

To find , we assume:

 

This leads to:

 

 

To get , we filter the measured spot signal with a moving average of span 13, in order to remove high frequency noise effects, and smooth out fluctuations (see section – Identifying burst events). To get , we fit the measured cell background signal to a 3rd degree polynomial (fitting to higher degree polynomials did not change the results). This is done to capture the general trend of the signal while completely eliminating fluctuations due to random noise.

**Identifying burst events**

We assume the total fluorescence is comprised of three distinct signal processes: RNP granule fluorescence, background fluorescence and noise. We further assume that background fluorescence is slowly changing, as compared with granule fluorescence which depends on the dynamic and frequent insertion and shedding events occurring in the droplet. Finally, we consider noise to be a symmetric, memory-less process. Based on these assumptions, we define a “signal-burst” event 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 base-line fluorescence intensity, we use a moving-average filter of 13 points (i.e., 2 minutes) 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 S4a). Random single fluctuations, which do not settle on a new baseline level are not expected to generate a gradual and continuous increase or decrease over multiple time-points in a smoothed signal. Following this, we search for contiguous segments of gradual increase or decrease and record only those whose probability for occurrence is 1 in 1000 or less given a Null hypothesis of randomly fluctuating noise.

To translate this probability to a computational threshold, we first compute the intensity difference distribution for every trace separately. This distribution is computed by collecting all the instantaneous differences in signal (ΔS(ti)= S(ti)- S(ti-1)) and binning them (Figure S4b). Given a particular trace the likelihood for observing an instantaneous signal increase event in a time-point (ti) can therefore be computed as follows:

 

where *N(*$ΔS(t\_{i})$*>0)* and *Ntot* correspond to the number of increasing instantaneous events and total number of events in a trace respectively. Likewise, the number of decreasing instantaneous events is defined as:

 

This in turn allows us to compute the number of consecutive instantaneous signal increase events (*m*) to satisfy our 1 in 1000 threshold for a significant signal increase burst event *m* as follows:

 

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 typically in the range $[m-1,m+1]$.

To account for the presence of the occasional strong instantaneous noise fluctuations appearing in experimental signals, we allow isolated reversals in the signal directionality (e.g., an isolated one time point decrease in an otherwise continuous signal increase environment). Furthermore, since the moving average filter itself can induce correlations in the signal, we determined that the minimum allowed threshold is the moving average window span. This means that any calculated threshold lower than the moving average size is increased to this bare minimum.

We mark each trace with the number of events whose duration exceeds the threshold and define those as bursts. 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. For each identified segment we record the amplitude (), and duration (). In Figure S4c we mark the classifications on a sample trace 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 each signal, since we cannot correctly classify signal sections that extend beyond the observed trace.



**Figure S5**: **Identification of burst events**. **a**, Top: simulated step signal (blue) with added white Gaussian noise (orange). Bottom: noisy signal after moving average filter. **b**, Intensity difference distribution for the signal presented in panel A. **c**, Sample experimental signal (orange) overlaid with markers indicating identified segments in green, blue, and red, corresponding to positive bursts, quiescent segments, and negative bursts.

Estimating the signal amount per slncRNA-RBP complex

Given the fact that we cannot directly infer the fluorescence intensity associated with a single RNA-RBP complex, we fitted the distributions with a modified Poisson function of the form:

 

where *I* is the experimental fluorescence amplitude, λ is the Poisson parameter (rate), and 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 tothat minimizes the deviation (MSE) from the experimental data. Fits were validated by observing the resulting QQ-plots.

**Numerical simulations of signal types**

To check that our analysis is consistent with an underlying random burst signal, we simulated three types of base signals with added noise components. For each simulation type, 1000 signals of 360 time-points were simulated and analyzed using the same data analysis process described in the methods section.

We simulated flat constant signals, gradually ascending signals, and signals containing multiple burst events. Two noise components were added to all signals, based on our noise model. White Gaussian noise of magnitude 40 [A.U] peak-to-peak amplitude, matching the value estimated from experimental traces, and an exponential component, simulating photobleaching (Figure S5).

We then applied our burst-detection algorithm described above and found that for the flat signal (Figure S5a) 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 ~5-10 as compared with the experimental data bursts, and the total number of events observed (458 positive, 452 negative, and 298 non-classified segments found) is significantly smaller than the experimental data, indicating roughly 1 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 S5c) a negligible number of negative burst-like events was detected by our algorithm, with a pronounced bias towards positive events (1111 positive, 9 negative and 467 non-classified). 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. 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 S5e). Our simulated signals resulted in a symmetric amplitude distribution, comprising of non-Gaussian or skewed amplitude distributions. Additionally, the range of amplitudes observed is 2-3x larger as compared with the case for the constant signal, with the non-classified amplitudes presenting a wider distribution. A total of 2298 positive, 1831 negative and 2489 non-classified segments were found.

**Estimating statistical significance of burst events in all traces recorded**

To compute whether the number of burst events identified via our algorithm is statistically significant, we simulated a constant base-line intensity amplitude with overlaid white Gaussian noise. For each numerical trace, we simulated 360 times points (corresponding to a ~60-minute experimental trace) and identified the total number of “increasing” and “decreasing” burst events in accordance with the algorithm described in detailed above. Here, we used m = 10 (see eqn. 1.8) consecutive increasing or decreasing instantaneous signal difference events as our threshold. We identified 458 and 298 increasing and decreasing burst events respectively in 1000 simulated traces with constant baseline. By comparison, we found 2298 and 1831 increasing and decreasing burst events respectively in 1000 simulated traces containing bursts, which using Fisher’s test yield a p-value of 4e-309 and 2e-310 for the significance of the increasing and decreasing burst findings.

We repeated this statistical test for experimental data, comparing the PP7-4x data against traces measured from cell containing only tdPP7-mCherry with no expression of our RNA cassettes, using the latter as a baseline akin to the constant signal simulations. We identified 7 increasing and 6 decreasing burst events in 150 traces gathered from the cells lacking RNA binding sites, while for the PP7-4x data we identified 112 increasing and decreasing burst events in 255 experimental traces, which using Fisher’s test yields a p-value of 2e-13.



**Figure S6**: **Signal type simulations**. **a**, Simulated constant signal (blue), with photobleaching (orange), and added noise (cyan). **b**, Amplitude distributions of burst events identified from 1000 constant signals. **c**, Simulated signal with slope (blue), with photobleaching (orange), and added noise (cyan). **d**, Amplitude distributions of burst events identified from 1000 sloped signals. **e**, Simulated signal with burst events (blue), with photobleaching (orange), and added noise (cyan). **f**, Amplitude distributions of burst events identified from 1000 bursty signals.

**Signal analysis parameter selection**

Subframe length

As part of the analysis process of the *in-vivo* microscopy experiments, the immediate surroundings of each discovered bright spot are recorded as a sub-frame containing the spot at its center, from this sub-frame the mean spot intensity and mean background intensity are calculated. The selection of the sub-frame length used to calculate the background intensity is an important parameter in the analysis process that might bring about unwanted noise into the resulting statistics when analyzing *in vivo* images. A large sub-frame might include other cells, with possibly different bright spots of themselves, inserting a bias into both the cell background intensity, and spot intensity signals. On the other hand, a small sub-frame might not have a sufficient spot-to-background area ratio, resulting in an underestimated cell background signal.

To select the appropriate sub-frame length, we analyzed the PCP-24x data with sub-frames of different lengths – 10, 14, 20, and 30 pixels. Figure S6a shows an example of this where the orange, yellow and red squares correspond to sub-frames of 10,14 and 20 pixels in length, and the panel itself constitutes a 30-pixel wide sub-frame. The criteria for this selection process are the mean ratio between cell area to spot area; percentage of frames where this ratio is less than one; and the ratio between the spot mean intensity to the cell mean intensity without any filtering or fitting. These criteria are designed to find the length that does not cause an overestimation of cell background against spot or vice versa (as could be the case where more than one bright spot fall inside the sub-frame). From these tests we learned that lengths of 10 and 14 pixels result in a mean ratio of less than two (i.e., on average the sizes of the bright spot and of its surrounding environment are equal) (Figure S6b). However, a sub-frame length of 10 pixels results in nearly a fifth of frames where the cell background is less than one and thus potentially underestimated (Figure S6c). Finally, the intensity ratios show that the mean ratio does not vary much between the different options, however the spread is more conserved for lengths of 10 and 14 pixels (Figure S6d). Following these tests, we chose a sub-frame length of 14 pixels for our analysis process.

Moving average span

The moving average window span is a critical component in the signal analysis process. It is used both as a noise reduction filter, and as a means to bias sharp signal jumps. The filter span plays another significant role, as it is the minimal allowed length for a burst duration. Choosing a small value might introduce false positives into the statistics, while a large value would cause many actual burst events to be discarded. To find the optimal span length we compare the number of events found in a simulated flat signal, such a signal should not produce any bursts under noise-less conditions. For this we simulated 1000 constant signals, 360 time points each, with an added white Gaussian noise and an exponential component and applied our data analysis procedure. (Figure S7a). An ideal result for this test would be less than one event of each type, i.e., positive, and negative bursts, per signal (Figure S7b).

We further show that using intermediate span length values (9-13 time points), has little effect on the qualitative nature of the results (Figure S7c, d).

Following these tests, we decided on a span of 13 time points. This value results in one event or less of each type per simulated signal, while still allowing us to record the statistical nature of the experimental signals.



**Figure S7: Subframe length selection**. **a**, Example of different sub-frame lengths. Image is a sub-frame with length of 30 pixels. Red, yellow, and orange squares correspond to sub-frames of length 20, 14 and 10 pixels accordingly. **b**, Ratio between cell background area to spot area (both are in number of pixels). **c**, Percentage of cells where the area ratio presented in (b) is less than one, indicating probable underestimation of the cell background. **d**, Ratio between spot mean intensities to cell background mean intensities (i.e., each spot is divided by its corresponding cell background). Horizontal lines represent 25 and 75 percentiles.



**Figure S8: Moving average span length selection**. **a**, Sample simulated signal used for testing. Blue line is the underlying constant signal, orange line represents the same signal with an added photobleaching component. Cyan signal is the orange signal with added white Gaussian noise. **b**, Total number of identified events of any kind per simulated signal. **c**, Positive amplitude histograms of PCP-24x data analyzed using a moving average filter of 9 time points (blue) and 13 time points (orange). **d**, Duration between positive events of PCP-24x data analyzed using a moving average filter of 9 time points and 13 time points.

**References**

(1) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. NIH Image to ImageJ: 25 Years of Image Analysis. *Nature Methods* **2012**, *9* (7), 671–675. https://doi.org/10.1038/nmeth.2089.

(2) Schindelin, J.; Arganda-Carreras, I.; Frise, E.; Kaynig, V.; Longair, M.; Pietzsch, T.; Preibisch, S.; Rueden, C.; Saalfeld, S.; Schmid, B.; Tinevez, J.-Y.; White, D. J.; Hartenstein, V.; Eliceiri, K.; Tomancak, P.; Cardona, A. Fiji: An Open-Source Platform for Biological-Image Analysis. *Nature Methods* **2012**, *9* (7), 676–682. https://doi.org/10.1038/nmeth.2019.

(3) Sbalzarini, I. F.; Koumoutsakos, P. Feature Point Tracking and Trajectory Analysis for Video Imaging in Cell Biology. *J. Struct. Biol.* **2005**, *151* (2), 182–195. https://doi.org/10.1016/j.jsb.2005.06.002.

(4) Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics* **1979**, *9* (1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076.