

# SUPPLEMENT

## **FISH-quant: automatic counting of transcripts in 3D FISH images**

Florian Mueller<sup>1,2</sup>, Adrien Senecal<sup>2</sup>, Katjana Tantale<sup>3</sup>, Hervé Marie-Nelly<sup>1</sup>, Nathalie Ly<sup>2</sup>, Olivier Collin<sup>2</sup>, Eugenia Basyuk<sup>3</sup>, Edouard Bertrand<sup>3</sup>, Xavier Darzacq<sup>2</sup>, Christophe Zimmer<sup>1</sup>

<sup>1</sup> Institut Pasteur, Unité Imagerie et Modélisation, Centre National de la Recherche Scientifique, Unité de Recherche Associée 2582, 25-28 rue du Docteur Roux, 75015 Paris, France

<sup>2</sup> Institut de Biologie de l'Ecole Normale Supérieure, Functional Imaging of Transcription, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 8197, 46 rue d'Ulm, 75005 Paris, France

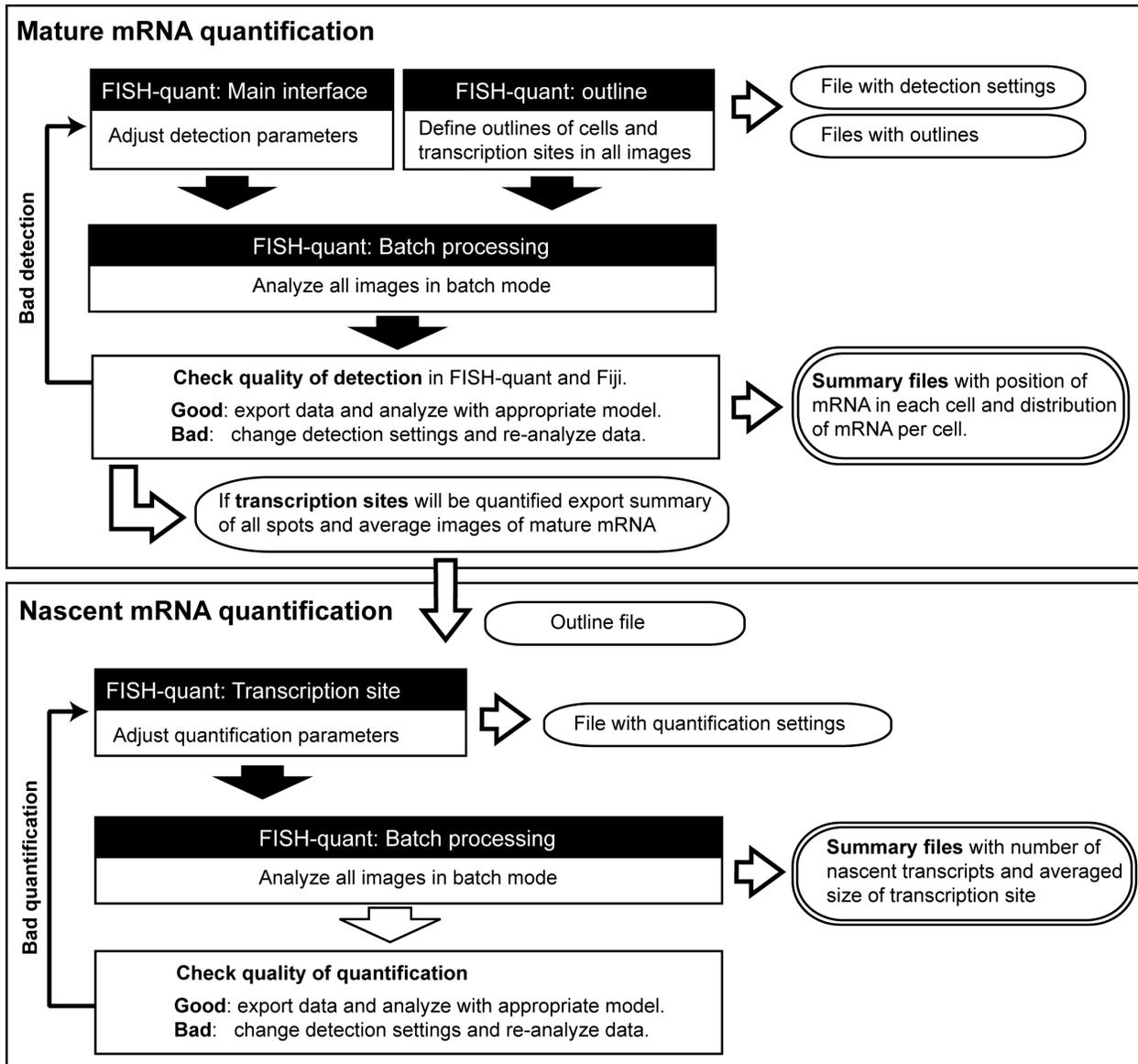
<sup>3</sup> Institut de Génétique Moléculaire de Montpellier, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 5535, 34293 Montpellier Cedex 5, France

*Correspondence to:* Edouard Bertrand (edouard.bertrand@igmm.cnrs.fr), Xavier Darzacq (darzacq@ens.fr), Christophe Zimmer (czimmer@pasteur.fr)

<b>1. Workflow of FISH-quant</b> .....	<b>2</b>
<b>2. Algorithm for mature RNA detection and counting</b> .....	<b>3</b>
<b>3. Validation of mature RNA detection in simulations and experiments</b> .....	<b>6</b>
<b>4. Algorithm for transcription site quantification and detection</b> .....	<b>9</b>
<b>5. Validation of transcription site quantification on simulated images</b> .....	<b>16</b>
<b>6. Validation of transcription site quantification with experimental data</b> .....	<b>22</b>
<b>Supplementary Methods</b> .....	<b>25</b>
<b>References</b> .....	<b>28</b>

# 1. Workflow of FISH-quant

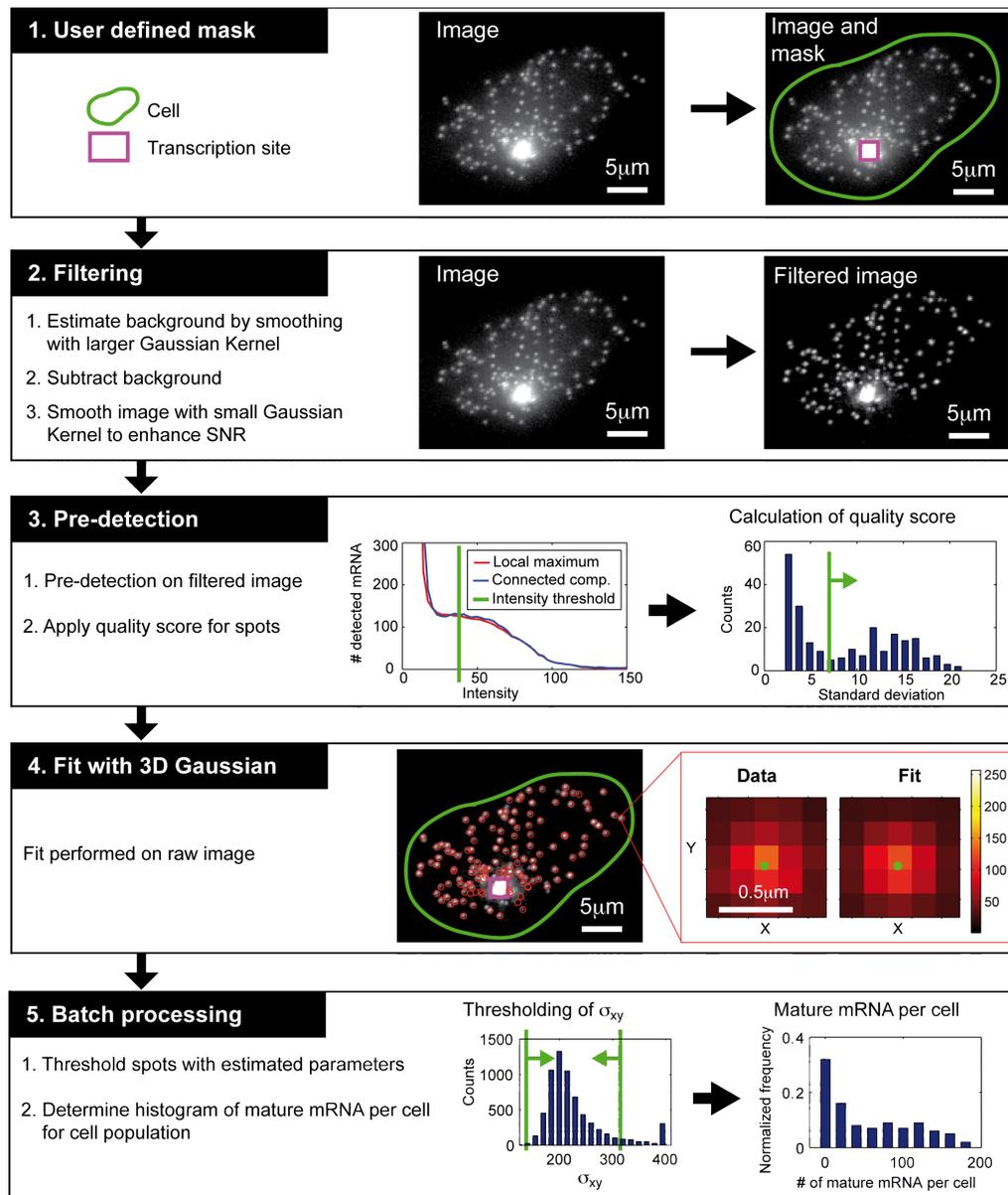
The general workflow for the counting of mature and nascent mRNA in FISH-quant is summarized in the following two schematics. The entire functionality can be controlled via graphical user-interfaces (GUI's). These interfaces are indicated by boxes with white headings on black background. A detailed description of how to use FISH-quant can be found in the documentation distributed together with source-code (<http://code.google.com/p/fish-quant/>). In this document also screen shots of the different GUI's can be found together with already processed example data. A more detailed description of the respective algorithms will be presented in the two following sections.



**Figure S1.** Schematic of workflow for counting of mature and nascent mRNA in FISH-quant. Full arrows indicate processing steps, empty arrows point to generated data. Boxes with white headings on black background indicate GUI's. Output files are indicated with boxes with rounded edges. Final output files that can be analyzed with mathematical models are indicated with a doubled frame.

## 2. Algorithm for mature RNA detection and counting

The detection and counting of mature mRNA is based on established methods for single molecule detection in 3D<sup>1</sup>. In short, pre-detection of spots is performed on a filtered image followed by a fit with a 3D Gaussian function. Remaining spots are then counted in each cell. Figure S2 summarizes the different steps involved. Each step will be explained in more detail below.



**Figure S2. Algorithm for mature mRNA detection.** For illustration purposes images are shown as maximum intensity projections (MIP) in XY but analysis is performed in 3D. **(1)** User defines masks for the outline of cells, transcription sites, and optionally nuclei. **(2)** Images are filtered for improved pre-detection. **(3)** Pre-detection by local maximum detection or connected components. User has to define a minimum intensity threshold. Plot on the left shows number of detected spots as a function of this detection threshold. Similar results are obtained for either method (compare blue and red curve). A characteristic plateau is observed for the correct thresholds. Chosen threshold (green vertical line) yields a slight over-detection but gives a safety margin for batch detection. For each spot candidate a quality score is calculated (here the standard deviation of the intensity in the neighborhood of the

spot). User sets a minimum score (green line), (4) Each candidate spot (red circles) is fit with a 3D Gaussian. Plots on the right show the MIP of the spot (left) and the best fit (right). The green circle indicates the identified center of the spot with sub-pixel localization accuracy. (5) Analysis can be performed in batch mode and results of many cells are pooled together. Spots can then be thresholded based on the different fitting parameters, e.g.  $\sigma_{xy}$ , as shown in the left plot. Final result is the number of mature mRNA per cell (right plot).

## 2.1. Define mask with outline of cell and transcription site

The user draws a mask to outline the individual cells and the transcription sites (and optionally nuclei). The subsequent analysis is only performed within the cells; transcription sites are excluded from the analysis of mature mRNA (but will be processed separately, see Supplementary Note 4). FISH-quant also provides different methods to automatically detect transcription sites (See Supplementary Note 4.4).

## 2.2. Filtering of image for better pre-detection

We implemented a two-step filtering process to remove inhomogeneous background and increase the SNR<sup>2</sup>. This is achieved by a 2-step convolution of the image with a Gaussian Kernel using the function `gaussSmooth`<sup>3</sup>. First, the raw image  $I_{raw}$  is convolved with a large Gaussian Kernel to blur it and obtain a good approximation of the background. By default the standard deviation of this Kernel is set to 5 times the standard deviation of a Gaussian that best matches the theoretical PSF for the optical setup used<sup>4</sup>. Then this image is subtracted from the raw image. The resulting image is then filtered with a small Gaussian Kernel to enhance the SNR. By default the standard deviation of this Kernel is set as the standard deviation of the Gaussian best describing the theoretical PSF<sup>4</sup>. The filtered images  $I_{filt}$  is therefore obtained by

$$I_{filt} = G_{\sigma_{th}} * (I_{raw} - G_{5\sigma_{th}} * I_{raw}), \quad [1]$$

where  $*$  indicates convolution with the indicated Gaussian Kernel  $G_{\sigma_{th}}$  (the integrated intensity of this Kernel is 1).

## 2.3. Pre-detection of spots

Next, candidate spots are identified that will be subsequently fit with a 3D Gaussian function in the next step. We implemented two different methods to identify these candidates. Both methods are applied to the filtered image  $I_{filt}$  obtained in step B.

- a. **3D local maximum detection**<sup>1</sup>. Identifies the local maxima with values greater than or equal to all voxels in the surrounding area with the function `nonMaxSupr`<sup>3</sup>. FISH-quant sets the radius of this area by default to twice the standard deviation of the Gaussian best describing the theoretical PSF<sup>4</sup>.
- b. **Connected components**<sup>5</sup>. Spot candidates are identified as connected components in 3D after thresholding the image with the Matlab function `bwconncomp`.

In either method the minimum intensity of a spot candidate must first be specified. When plotting the number of detected spots as a function of this intensity threshold a characteristic plateau can be found for a range of intensity values that yield an optimal detection (also described by Raj et al.<sup>6</sup>). We found identical curves for both pre-detection methods (Fig. S2). We manually place the intensity threshold towards the left part of the plateau, which leads to a slight over-detection, but the subsequent steps will remove false-positive detections.

Given the signal-to-noise ratio of typical FISH experiments (Supplementary Note 3.1), the determined spot candidates will encompass only a few false positives. Nevertheless we implemented an **additional quality check** to discriminate true spots from background noise. For this purpose we consider the

intensity distribution surrounding the spot candidates<sup>1</sup>. In FISH-quant either the 3D curvature based on the Hessian matrix<sup>1</sup> or the standard deviation of the surrounding voxels can be estimated and serve as quality scores. We found that the standard deviation works more robustly for lower quality image. For both methods larger values of the quality score are obtained for good spots, and lower value for background and so a second threshold can be set to separate noise from actual spots. The remaining spots will then be fit with a 3D Gaussian function.

**Note:** Signal from individual, non-specifically bound probes is detected for some FISH experiments, especially if only a limited number of probes can be used to target the mRNA. Here the quality score alone might not be sufficient to differentiate background noise from real spots. A careful combination of intensity and quality score thresholding has to be applied. The estimated amplitude from the fit with the 3D Gaussian can be also used as an additional thresholding parameter (see below).

## 2.4. Spot fitting with 3D Gaussian

The remaining spot candidates are then fit with the following function, a 3D Gaussian integrated over the voxel. The fitting is performed in the raw image since filtering affects the localization accuracy and the intensity estimates<sup>1</sup>.

$$I_{ijk} = B + A \frac{1}{x_{i,u}-x_{i,l}} \frac{1}{y_{j,u}-y_{j,l}} \frac{1}{z_{k,u}-z_{k,l}} \int_{x_{i,l}}^{x_{i,u}} \int_{y_{j,l}}^{y_{j,u}} \int_{z_{k,l}}^{z_{k,u}} G(x, y, z) dx dy dz, \quad [2]$$

where  $I_{ijk}$  is the modeled intensity of voxel  $i$ ,  $x_{i,l}$ ,  $y_{j,l}$  and  $z_{k,l}$  denote the lower border of the voxel,  $x_{i,u}$ ,  $y_{j,u}$ , and  $z_{k,u}$  denote the upper border of the voxel,  $G(x,y,z)$  is the Gaussian function give by Eq. [3],  $B$  is background of the image, and  $A$  is the amplitude of the Gaussian.

$$G(x, y, z) = e^{-\frac{(x-\mu_x)^2+(y-\mu_y)^2}{2\sigma_{xy}^2}} e^{-\frac{(z-\mu_z)^2}{2\sigma_z^2}}, \quad [3]$$

where  $\sigma_{xy}$  and  $\sigma_z$  are the width of the Gaussian in xy and z.  $\mu_x$ ,  $\mu_y$ , and  $\mu_z$  are the coordinates of its center in x, y, and z. The solution of this integral is provided in Matlab with the function `erf` and can be used after a simple renormalization. Images of individual spots  $I_{spot}$  are then fit with the Matlab function `lsqcurvefit` to minimize the squared sum of residuals  $R$

$$R = R = \sum_i \sum_j \sum_k (I_{spot,ijk} - I_{ijk})^2 \quad [4]$$

thus yielding estimates of  $\sigma_{xy}$ ,  $\sigma_z$ ,  $\mu_x$ ,  $\mu_y$ ,  $\mu_z$ ,  $A$ , and  $B$ .

In a last step, spots can be selected by thresholding  $\sigma_{xy}$ ,  $\sigma_z$ , as well as  $A$ . False positives, resulting from noise, usually have large  $\sigma_{xy}$  and  $\sigma_z$  compared to real spots and can therefore be easily removed. The remaining spots are then counted in each cell, providing the estimated number of mature mRNA.

## 2.5. Batch mode

We found that the various detection parameters can be defined robustly for images taken on the same day under identical imaging conditions. Therefore such a set of images can be analyzed with the same parameters. Accordingly FISH-quant offers a batch-processing module. The user can first define all outlines for all cells in the images and then process them fully automatically. The final thresholding based on the fitting parameters can then be adjusted based on the results of all fitted spots in all images.

### 3. Validation of mature RNA detection in simulations and experiments

In this section we report quantitative validations of mature mRNA detection in simulations and in experimental data.

#### 3.1. Localization accuracy of mature mRNA on simulated images

We validated the mRNA localization accuracy of FISH-quant in simulated 2D images. We compared the FISH-quant estimates to two other localization methods: a Maximum-Likelihood Estimation (MLE) and a recently developed algorithm based on radial symmetry<sup>7</sup>. Both of these methods reach accuracies that are near theoretical limits and were implemented in Matlab, facilitating their implementation and comparison with FISH-quant.

##### Signal-to-noise (SNR) in FISH images

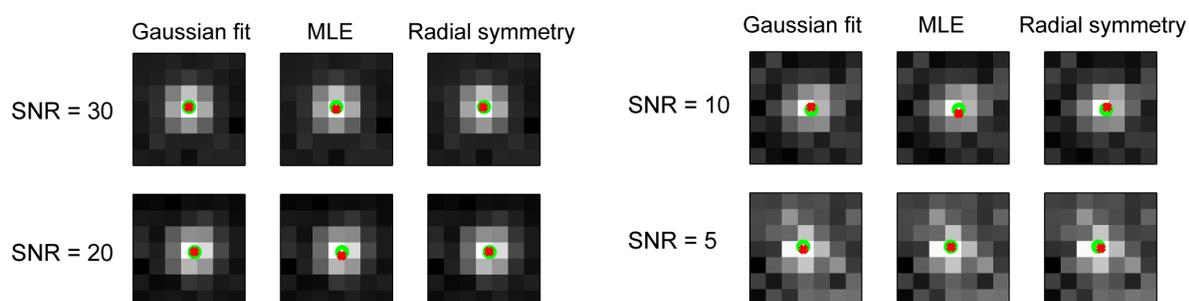
It is well known that the localization accuracy of these methods depends strongly on the signal-to-noise ratio (SNR) of the images<sup>8</sup>. We therefore first quantified the signal-to-noise-ratio (SNR) in typical FISH images. We defined SNR as the ratio of estimated amplitude of the Gaussian  $A$  over the standard deviation of the background  $\sigma$ :  $SNR = A/\sigma$ . We obtained  $A$  from the actual fits with the Gaussian function, and  $\sigma$  by computing the mean and standard deviation of regions in cells containing only background (see equation [2] on page 5). The obtained SNR (Table S1) are high because in FISH individual mRNA molecules are labeled with several tens of fluorophores. Higher SNR was obtained for RPB1 probes labeled with Cy3 compared to Alexa 488, as expected given the difference of autofluorescence in the two colors and brightness of the dyes.

FISH experiment	SNR
RPB1: labeled with Cy3	33
RPB1: labeled with Alexa 488	7
Hygro-MS2x96-bGH	27

**Table S1.** SNR of FISH images for different experiments presented in this study.

##### Validation of localization accuracy in noisy images

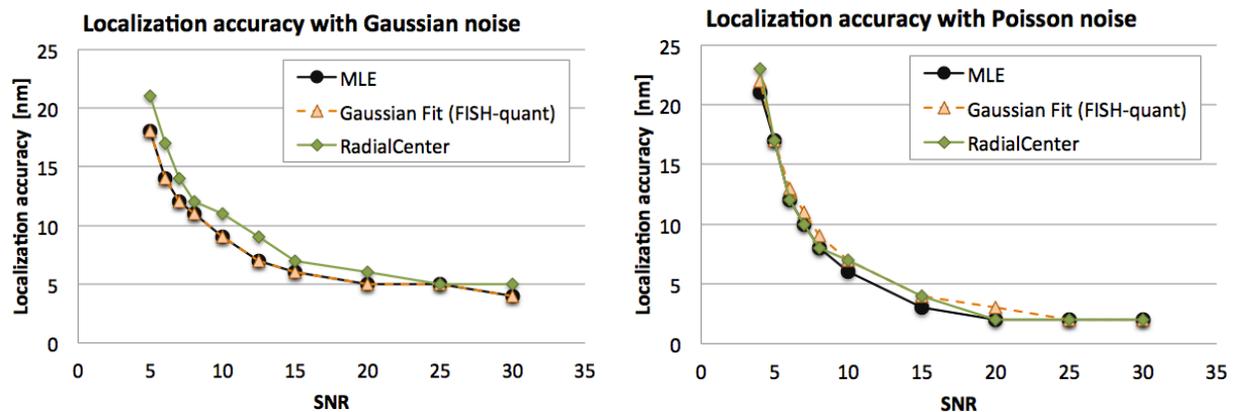
To estimate the pointing accuracy of the three methods, we simulated pixelated images of a diffraction-limited point source with different, known sub-pixel localizations and different noise levels (Fig. S3). For each SNR, we simulated 800 images and fit them with the three methods.



**Figure S3.** Fit of simulated pixelated images in 2D. PSF was obtained from PSF ImageJ plugin PSF-Generator<sup>9</sup> on a fine pixel grid of 5nm (Emission wavelength = 568 nm, numerical aperture = 1.25, refractive index = 1.46). PSF was then placed at random sub-pixel locations and an image on larger pixel grid (100nm) was generated. We included noisy background by additive Gaussian noise and varied its standard deviation to obtain different SNR levels. Open green circles indicate the true locations of the PSF center, red spots indicate the position estimated by each localization method. Fig. S10 summarizes localization accuracy for different SNR.

We then computed for each SNR the median of the individual absolute localization errors  $e = \sqrt{(x_c - x_m)^2 + (y_c - y_m)^2}$ , where  $x_c$  and  $y_c$  are the known center coordinates of the point source, and  $x_m$  and  $y_m$  are the measured coordinates (Fig. S4, left). All three methods achieved similar detection accuracy for the simulated range of SNR.

We also analyzed simulated images with Poisson noise<sup>7</sup> for the same range of SNR and found again that the methods yielded comparable results (Fig. S4, right).



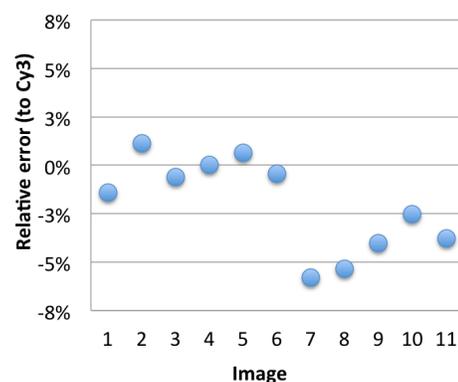
**Figure S4.** Localization accuracy of Gaussian fit compared to MLE and radial center method. All methods provide accurate estimates for simulated images. **(Left plot)** Images with simulated additive Gaussian noise with different SNR. **(Right plot)** Images with simulated Poisson noise<sup>7</sup> with different SNR (SNR is defined relative to peak signal intensity).

In summary, these simulations indicate that the localization accuracy of the Gaussian fit used in FISH-quant is high for realistic SNR levels of FISH images, and is very similar to accuracies achieved by state-of-the-art localization methods. This stems from the high SNR of FISH images where each mRNA molecule is labeled by tens of fluorophores.

### 3.2. Experimental validation with dual-color FISH

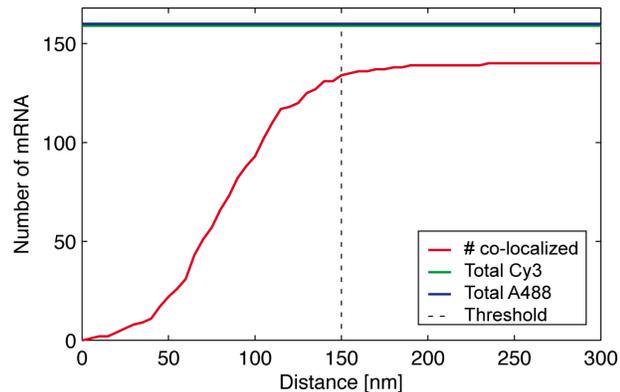
We further tested the reliability of the mature mRNA detection experimentally by labeling the same target mRNA simultaneously in two colors. We defined a total of 30 probes for RPB1 and labeled them interleaved with Alexa 488 and Cy3 (see Supplementary Methods for details). We collected 11 images in both colors and used FISH-quant for the mature mRNA detection (Fig. 1b). We obtained excellent agreement for the number of mature mRNA and the estimates were within +/- 5% (Fig. S5 and Fig. 1c).

**Figure S5.** Relative error between estimated amount of mature mRNA molecules in dual color FISH against Alexa 488 and Cy3. Error is shown with respect to Cy3.



Next we investigated if the detected spots **co-localize**, i.e. if two spots detected in Alexa 488 and Cy3 correspond to the same mRNA molecule. For this purpose we first needed a quantitative definition of co-localization. To do this we calculated the 3D distance between each spots detected in Cy3 to all spots detected in Alexa 488 and vice-versa. We then determined for each spot the distance to the closest detected spot in the other color. Then we subtracted the average shift as a first order correction for chromatic aberration effects. We then counted for one color the number of spots that have at least one detected spot in the other color within a given distance (Fig. S6). These numbers increase with distance and reach a plateau after 125 nm indicating that this is the maximum distance between two co-localized spots corresponding to the same mRNA molecule. We then used a value of 150 nm to define co-localization of spots detected in two colors. We estimated the percentage of all spots that have a co-localized spot in the other color and determined the amount of co-localization for all images (Fig. 1d). We found that in each image 85%-90% of all detected spots co-localize with a spot in the other color.

**Figure S6.** Number of detected spots having a neighboring detected spot in the other color within a given distance. Green line shows the total number of spots detected in Cy3, blue in Alexa 488. The red line shows the number of co-localized spots in Alexa 488 with respect to Cy3. The dashed black line shows the distance chosen to define co-localization. The co-localization for Alexa 488 is around 86% and for Cy3 85%.



In summary, the dual-color FISH experiment demonstrates the high reliability of mature mRNA detection in FISH-quant.

## 4. Algorithm for transcription site quantification and detection

The following sections describe in detail how FISH-quant quantifies the amount of nascent mRNA— or more precisely **the equivalent amount of full-length transcripts**. The quantification of the FISH signal at the transcription site yields the corresponding number of full length transcripts that would give rise to this signal. However, the signal could also stem from a larger number of partially transcribed transcripts. It is, however, not possible to differentiate between those two scenarios since the resulting signal will be the same. However, appropriate experimental design can minimize this problem and even be used to infer important properties of transcription<sup>10–12</sup>. When placing FISH probes towards the 3' end of the transcript only almost completed transcripts are visible. Alternatively, probes can be placed towards the 5' region to detect also incomplete nascent transcripts. Comparing the results of these two placement strategies for the same gene can be used to study polymerase clustering and transcriptional bursting<sup>10</sup>, or to estimate the relative time taken by elongation versus 3'-end processing and release<sup>12</sup>. Further, by designing probe sets in different colors against different parts of the transcript the position of polymerase on the gene can be investigated<sup>11</sup>.

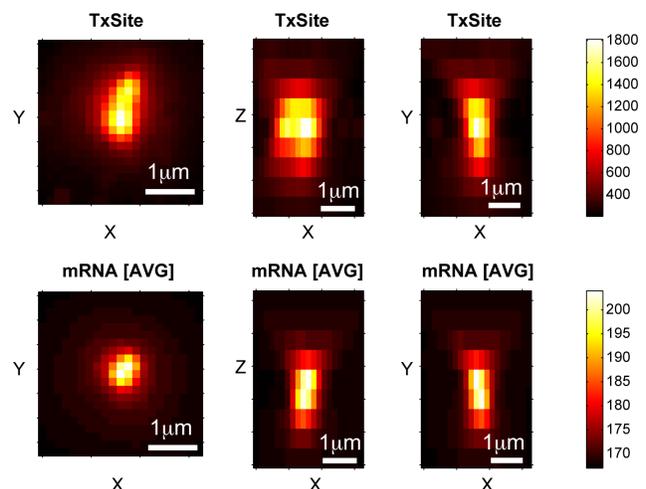
Lastly, we note that FISH-quant is not limited to transcription sites, but can be **readily applied to other structures** with a dense accumulation of mRNA, e.g. P-bodies or stress granules<sup>13</sup>.

First a brief motivation is given, followed by a detailed explanation of the implementation. The last section describes two different approaches implemented in FISH-quant to automatically detect transcription sites.

### 4.1. Motivation: spatially extended transcription sites

Transcription sites can have complex topologies such as elongated structures or V-shapes as observed for viruses, genes transcribing large repeated non-coding RNAs, and gene arrays<sup>12,14–20</sup>. We routinely observed transcription site that are larger than the diffraction limit (Fig. S7 and S19). Figure S7 shows the image of a typical transcription site and to the averaged image of individual mRNA molecules. This comparison shows that the transcription site is substantially larger than individual mRNA molecule (and the PSF).

**Figure S7.** Comparison of a typical transcription site (upper row) to the averaged image of 7500 individual mature mRNA molecules (lower row) for *Hygro-MS2x96-bGH*. Images are maximum intensity projections along the major axis as indicated in the axis label. The transcription site is larger than the individual mRNA molecule. Both images show defocusing pattern.



To our knowledge no method is available to accurately quantify the number of nascent transcripts for such large and spatially extended transcription sites in 3D. In previous studies, the maximum intensity of the transcription site was divided by the averaged maximum intensity of the brightest voxel of individual mature mRNA in the cell<sup>12</sup>. Alternatively, the number of nascent mRNA can be inferred by calculating

the ratio of the estimated amplitudes of the transcription site to that of the individual mRNA molecules. Either method neglects the spatial extent of transcriptions sites and therefore implicitly assumes that all transcripts are within a sub-resolution region. The example in Figure S7 illustrates, however, that's transcription sites can be larger and extended in 3D. We therefore implemented two new methods for the quantification of such sites as detailed below.

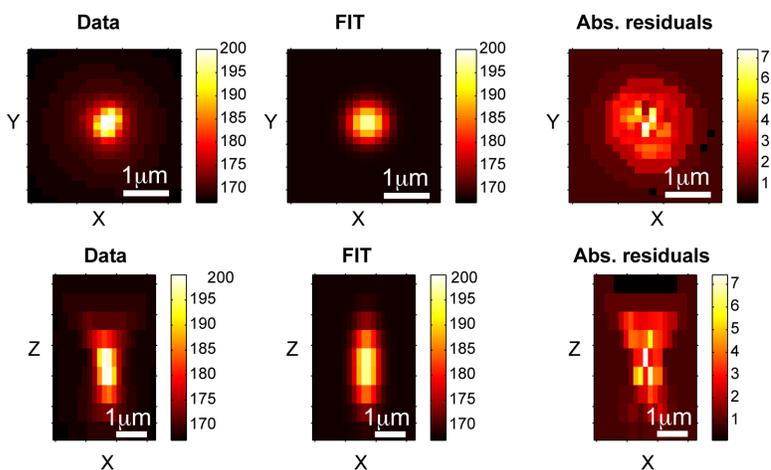
## 4.2. Transcription site quantification method 1: integrated intensity in 3D

We consider the spatial extent by comparing the total integrated intensity of the transcription site to the total integrated intensity of the individual mRNA molecules. FISH-quant directly considers the 3D image (quantifications based on integrated intensity have previously been applied to 2D maximum intensity projections<sup>10</sup>). We first average the images of the individual mRNA molecules detected as detailed in Supplementary Note 2. We then fit this image with Eq. [2] and calculate the integrated intensity under the fitted curve. Each transcription site is then fit with Eq. [2] and the integrated intensity under the Gaussian is calculated. The number of nascent transcripts is then inferred by dividing the integrated intensity of the transcription site by the integrated intensity of the individual mRNA molecules.

## 4.3. Transcription site quantification method 2: superposition of PSFs

In the second approach we considered the spatial extent by using the average image of the individual mRNA molecules to construct an image that best describes the recorded image of the transcription site. This approach is inspired by Gaussian mixture models (GMM) where a weighted sum of Gaussian functions is used to describe complex signals resulting from a superposition of overlapping Gaussians<sup>1</sup>. There are, however, two limitations to GMM that impede its application to FISH. First, a 3D Gaussian function can be satisfyingly used to fit and localize diffraction limited spot<sup>1</sup> but it fails to describe the observed complex diffraction patterns and other aberration effects due to misalignments of the microscope (Fig. S8). We therefore use directly the averaged image of all detected mature mRNA as described above rather than a Gaussian function to describe the signal of individual mRNA molecules. Second, in GMM the weight (amplitude) of the individual Gaussian functions are not restricted<sup>1</sup>. This can lead to an overestimation of the number of mRNA at the transcription site when Gaussian functions with increasingly small amplitudes are used to further improve the fit. In FISH-quant, we therefore restrict the range of the allowed amplitude to the range measured on the individual mRNA molecules.

**Figure S8.** Fit of averaged image of mRNA with 3D Gaussian. First row shows maximum intensity projections in XY, second row in XZ. First column shows the image, second column the best fit, and third column the absolute residuals. The fit describes the signal well but small systematic deviations can be seen for the diffraction patterns. This poses no problem for localization but it can result in an overestimation of the number of Gaussians used in the GMM to describe a bright transcription site. For such bright sites the diffraction patterns can become prominent and these additional Gaussians would be necessary to model them.



We first compared the averaged image of individual mRNA molecules to images of 100 nm fluorescent beads (TetraSpeck, Invitrogen). We analyzed the images of the beads with the same workflow as the FISH data. While the beads were brighter than the individual mRNA molecules their estimated size

was similar (Table S2). This argues that for our experimental system individual mRNA molecules are diffraction limited in size and we can safely average them without losing spatial information.

Parameter	Beads (N=1,500)	Hygro-MS2x96-bGH (N=13,000)
$\sigma_{xy}$	151 +/- 5 nm	175 +/- 35 nm
$\sigma_z$	561 +/- 53 nm	577 +/- 128 nm
<b>A</b>	428 +/- 101	84 +/- 24
<b>B</b>	205 +/- 7	177 +/- 15

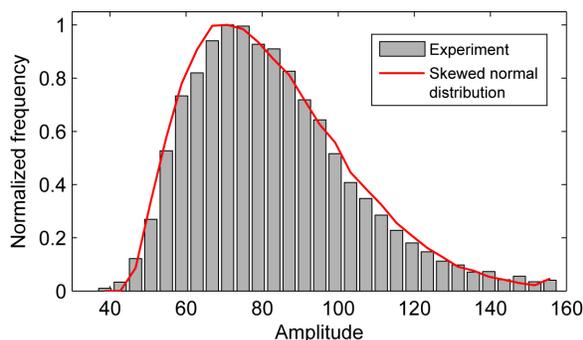
**Table S2.** Comparison of fitting results for beads and mature mRNA in FISH-quant. Listed values are mean +/- standard deviation. Numbers in parenthesis indicate how many spots were considered.

Next we analyzed the **distribution of the estimated amplitudes** of the fit with the 3D Gaussian for the individual mRNA. We found that this distribution is well described by a skewed normal distribution (Fig. S9). We used the Matlab command `normfit` to determine the mean value  $\mu$  and standard deviation  $\sigma$ . Skewness  $s$ , and kurtosis  $k$  were determined with the Matlab commands `skewness` and `kurtosis` and are defined as follows:

$$s = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^3}{\left( \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \right)^{3/2}}, \quad k = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^4}{\left( \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \right)^2}. \quad [5]$$

This distribution is caused by a number of different factors including detection noise, variable labeling efficiency of the FISH probes, variable number of hybridized probes per mRNA, and stochasticity of fluorescence.

**Figure S9.** Fit of distribution of estimated amplitudes with skewed normal distribution yields  $\mu = 83$ ,  $\sigma = 21$ ,  $s = 0.8$ , and  $k = 3.5$ . Red curve shows normalized histogram of 10,000 random numbers simulated with the Matlab function `pearsrnd` with the specified values.



### 4.3.1. Algorithm for superposition of PSFs

The quantification method is summarized below and in Fig. S10. In short, the algorithm attempts to find the most probable 3D positions of mRNA's that give rise to the recorded image of the transcription site. This is achieved by an iterative process where individual mRNAs are placed in a model image until the best description of the actual image is obtained. Each step will be explained next in more detail.

#### a) Analyze transcription site and background of cell

The algorithm starts with a homogenous background image in which the individual mRNAs are placed. Different cells have different background values so we implemented an automated method to determine the best background. First, the background of the cell is analyzed by extracting the voxel intensities within the same z-planes as the transcription site. Then their mean  $\mu_{cell}$  and standard deviation  $\sigma_{cell}$  are calculated and used to determine a possible range of background values  $B$  that will be tested (By default 10 values in the range  $[\mu_{cell} - \sigma_{cell}, \mu_{cell} + \sigma_{cell}]$ ). The image is further cropped around the transcription site to restrict the area of analysis ( $I_{TS}$ ).

## b) Properties of mature mRNA molecules

As an input for the algorithm the averaged image of the individual mature mRNA molecules (Fig. S7) and the distribution of the estimated amplitudes (Fig. S9) are imported (Supplementary Note 2).

## c) Determine background

The algorithm (see below) is then applied 50 times for each background value  $B$  in the range  $[\mu_{cell} - \sigma_{cell}, \mu_{cell} + \sigma_{cell}]$ . These repetitions are needed because the algorithm uses random numbers.

## d) Detailed analysis

Then the background value with the lowest residuals is chosen for a subsequent analysis and the algorithm is performed 100 times. Additionally, the averaged size of the transcription site is determined by calculating the average distance of all individual placed mRNA molecules to their center of mass.

### Algorithm

1. Generate homogenous background image  $B$  with  $B \in [\mu_{cell} - \sigma_{cell}, \mu_{cell} + \sigma_{cell}]$
2. Calculate sum of absolute residuals  $R_0$  between  $B$  and the cropped image of the transcription site  $I_{TS}$ :

$$R_0 = \sum_{x,y,z} |B - I_{TS}|. \quad [6]$$

3. Iteratively add one image of the mature mRNAs  $I_M$  to the background  $B$  to obtain the model image  $I_{G,N}$  (see below):

$$I_{G,N} = B + \sum_{i=1}^N I_M(x_{c,i}, y_{c,i}, z_{c,i}, A_i), \quad [7]$$

where  $I_{G,N}$  is the image obtained after placing  $N$  mRNA images. Each placed individual mRNA  $I_M$  has a different center specified by  $x_{c,i}, y_{c,i}, z_{c,i}$  and amplitude  $A_i$ .

### Iterative placement of mRNA

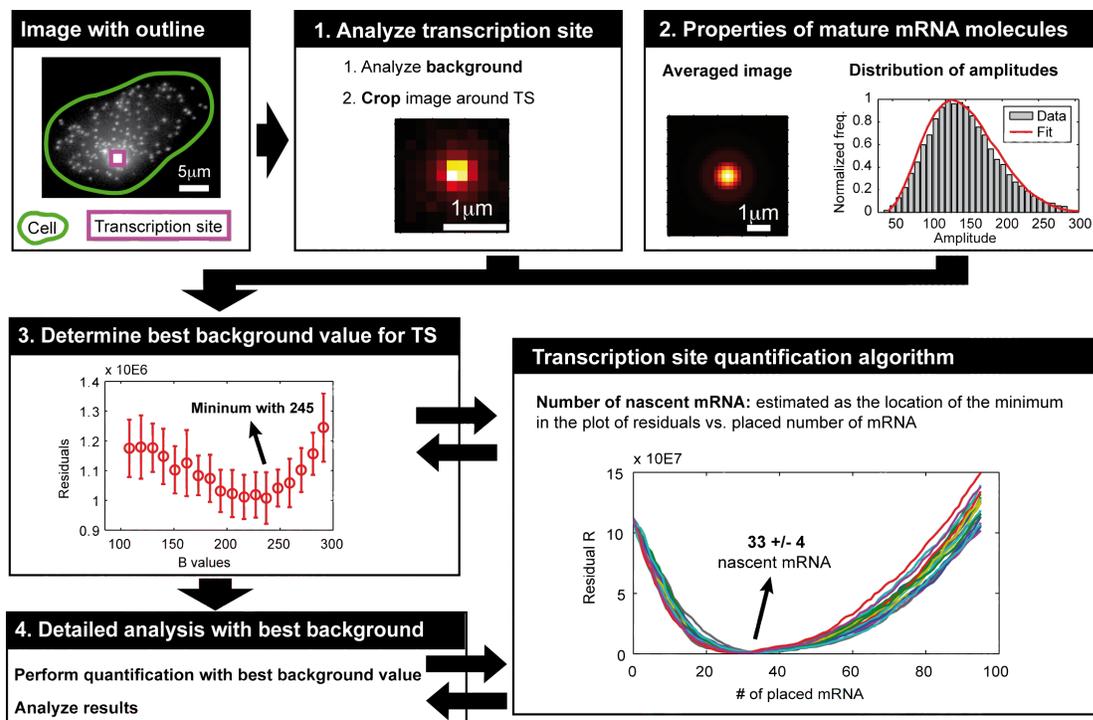
- 3.1. Subtract the model image from the preceding iteration ( $I_{G,N-1}$ ) from the image of the transcription site  $I_{TS}$ . Note that  $I_{G,0} = B$ .

$$I_D = I_{G,N-1} - I_{TS} \quad [8]$$

- 3.2. Find voxel with maximum intensity in image  $I_D$  from Eq. [8].
- 3.3. Pick an amplitude  $A$  by random sampling of the skewed Gaussian distribution of the estimated amplitudes (Fig. S9).
- 3.4. Renormalize average image of mRNA to match amplitude from step 3.3.
- 3.5. Add this image to  $I_{G,N-1}$  at location from step 3.2 to obtain the new model image  $I_{G,N}$  as described in Eq. [7].
- 3.6. Calculate sum of absolute residuals  $R_N$  between model image and image of transcription site:
$$R_N = \sum_{x,y,z} |I_{G,N} - I_{TS}|. \quad [9]$$
- 3.7. Back to 3.1. until residuals  $R_N$  are larger than residuals  $R_0$  estimated in step 2.

### Analysis of results

4. The residuals  $R_N$  as a function of  $N$  follow a characteristic U-form (Fig. S10). For each run the number of mRNA's with the minimum residuals is determined and serves as an estimate of the number of nascent transcripts. Runs are repeated several times and the averaged number of nascent mRNA and the standard deviation is calculated.



**Figure S10.** Schematic of PSF superposition approach to quantify the amount of nascent transcripts.

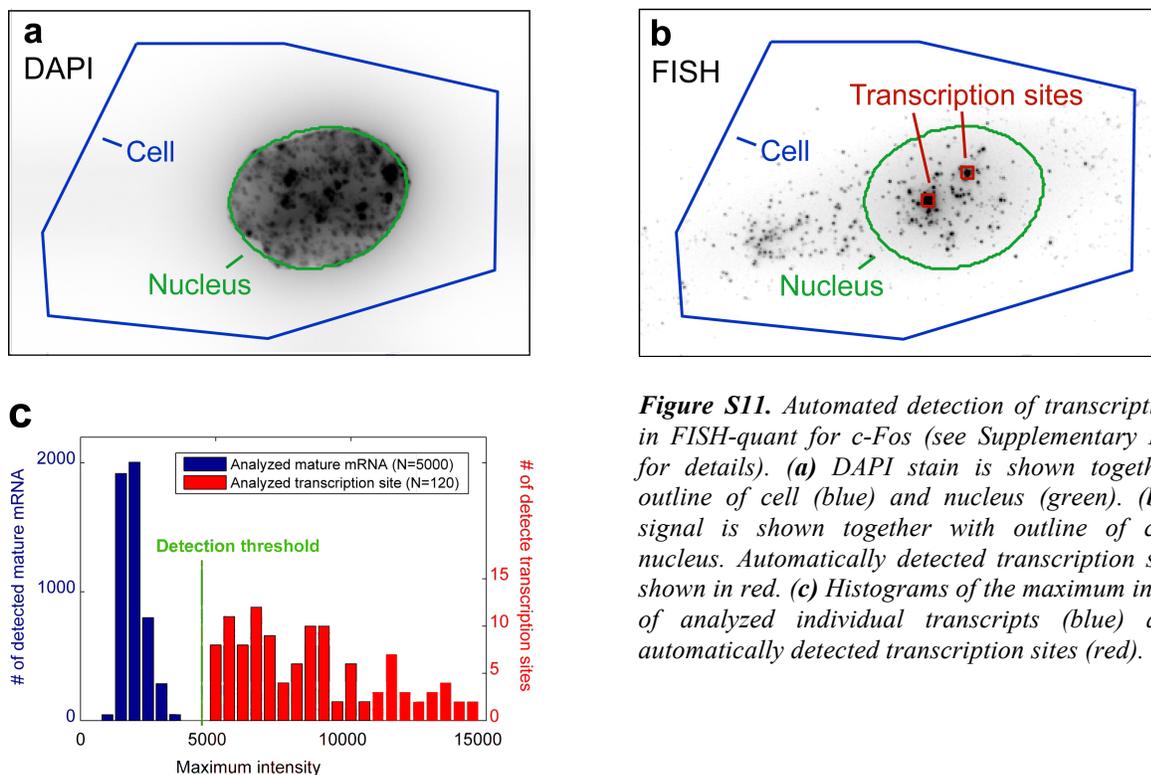
## 4.4. Automated detection of transcription sites

FISH-quant provides two different methods to automatically detect transcription sites. First, transcription sites can be identified as spots in the nucleus with higher intensities than the mature counterparts<sup>10</sup> (Supplementary Note 4.4.1.). Second, images of an independent label of the transcription site, like LacI, can be used to locate transcription sites in the FISH image (Supplementary Note 4.4.2.).

### 4.4.1. Automated detection based on intensity alone

Transcription sites are identified based on a user defined intensity threshold that separates them from mature mRNA. To reduce the number of false-positives the detection can be further restricted to the nucleus of each cell. This can be done either by loading a DAPI image, or an image of any other nuclear stain, and defining an additional intensity threshold for the DAPI signal or by defining the outline of the nucleus (Fig. S11a).

We applied this method on RNA-FISH images against the c-Fos gene in human fibroblasts 20 min after serum induction (Fig S11b). This gene is less expressed as the other genes used in this study ( $\beta$ -actin Hygro-MS2x96-bGH reporter), thus making transcription site identification more challenging. After a first round of FISH-quant analysis we found that the intensity of individual transcripts did not exceed 4000 units (Fig. S11c). We therefore set the intensity threshold for transcription site detection to 5000. We then restricted the automated detection to the outlined nuclei. Detected transcription sites were substantially brighter than mature transcripts (Fig. S11c). We visually verified more than 100 cells and found excellent agreement between the automatically detected and manually outlined sites.



**Figure S11.** Automated detection of transcription sites in FISH-quant for *c-Fos* (see Supplementary Methods for details). (a) DAPI stain is shown together with outline of cell (blue) and nucleus (green). (b) FISH signal is shown together with outline of cell and nucleus. Automatically detected transcription sites are shown in red. (c) Histograms of the maximum intensities of analyzed individual transcripts (blue) and the automatically detected transcription sites (red).

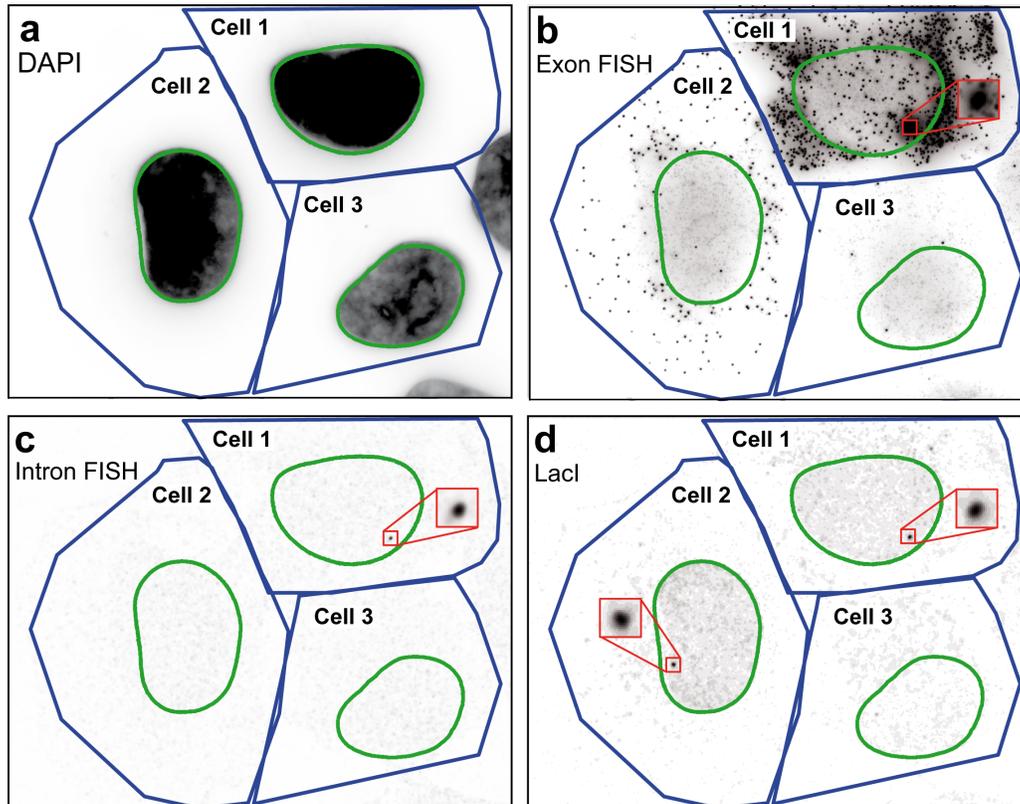
#### 4.4.2. Automated detection based on second marker

The detection method described in Supplementary Note 4.4.1. only works for sites that are sufficiently bright compared to mature mRNA. This poses, however, a problem at only weakly transcribing genes where in the lowest limit only one transcript is attached to the transcription site. The image of such a transcription site will result in the same diffraction limited spot as the image of its mature counterpart. It is therefore impossible to distinguish the two based on their intensity alone. Experimental approaches have been developed to circumvent this limitation by independently labeling the transcription site with a second marker. The most frequently used method is the LacI tagging approach<sup>21</sup>. Here the lac repressor (LacI) is fluorescently tagged and binds to arrays of lac operator sequences inserted close to the transcription sites on the chromosome. Alternatively, DNA FISH can be performed against the target gene to obtain independent labeling<sup>11</sup>. Lastly, mRNA FISH can be performed with special probes designed against the intron of the studied genes<sup>11</sup>. Most transcripts are spiced co-transcriptionally<sup>22,23</sup>, so transcripts will only be visible at the transcription sites, while mature mRNA will generally not be detected. Ultimately, each of these methods produces a second image stack where the transcription sites are marked independently. In FISH-quant, these additional images can be used to automatically detect transcription sites also in the absence of FISH signal (Fig. S12).

We demonstrate the different approaches with an artificial reporter ( $\beta$ -globin-Luc-CFP-24MS2, Supplementary Methods). We constructed this reporter such that FISH can be performed against exons and introns. The reporter has also binding sites for LacI, so an independent visualization of the gene locus is possible with LacI-YFP (which was transfected as a plasmid). We acquired 4 image stacks for each field of view: DAPI, FISH against exons, FISH against introns, and LacI-YFP. This allows a direct comparison of these techniques. The exon FISH image shows a large number of mature mRNA, but also allows to detect a transcription site in cell 1 (Fig. S12b). In the intron FISH image, mature mRNA molecules are not visible, as expected, but the active transcription site can be clearly detected (Fig. S12c).

Finally, the image of LacI-YFP also shows a silent transcription site in cell 2 that was not visible in the other two images (Fig. S12d). No transcription site could be detected in cell 3, because this cell did not express LacI-YFP.

In summary, FISH-quant provides different options for the automated detection of transcription sites. For strongly transcribing genes a detection based on the intensity of the transcription site alone can be sufficient. For weakly transcribing genes an independent label of the site can be used to reliably detect its location.



**Figure S12.** Automated detection of transcription sites for  $\beta$ -globin-Luc-CFP-24MS2 in FISH-quant. (a) DAPI stain was used to outline the nuclei in the cells. (b) Detection based on intensity of FISH (against exon) signal. Only cell 1 contains a detected transcription site. (c) Detection with FISH against introns yields the same site as in b, since only transcriptionally active sites can be detected. (d) Detection with LacI (transiently expressed) yields two transcription sites: the same site in cell 1 detected in b and c, and a site detected in cell 2 containing only one transcript; in cell 3, the transfected LacI-YFP was not expressed and therefore did not allow to visualize the transcription site (this could be avoided by stably expressing LacI).

## 5. Validation of transcription site quantification on simulated images

We first evaluated the transcription site quantification methods on simulated data. In the following section we will refer to the different quantification methods with the following abbreviations:

- FISH-quant method based on integrated intensity: FQ-IntInt
- FISH-quant method based on superposing PSFs: FQ-PSFsup
- Method based on estimated amplitude: AMP
- Method based on maximum intensity: MaxInt

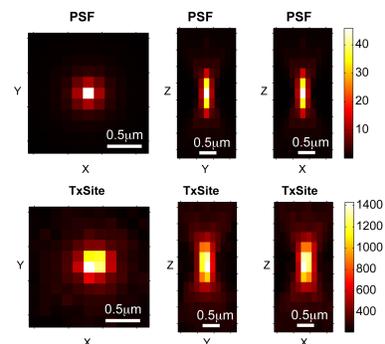
Further, we utilize the term FISH-quant methods to refer to FQ-IntInt and FQ-PSFsup, and simpler methods to refer to AMP and MaxInt, which both ignore the three-dimensional extent of the transcription sites.

### 5.1. Generation of artificial images of transcription sites

We showed that individual mature mRNA molecules are diffraction limited, i.e. their image can be described by the point-spread function (PSF) of the microscopes (Table S2). For the following simulations we therefore used a realistic 3D PSF obtained by the ImageJ plugin PSF-Generator<sup>9</sup> (Fig. S13, lower row).

We generated an image of a transcription site by superposing a pre-defined number of individual PSF's in a certain area. The resulting image of the transcription site depends a number of parameters: number of nascent mRNAs, amplitude of each placed mRNA, size of the transcription site, and noisy background (Fig. S13 shows an example for such a simulated site).

**Figure S13.** Images of theoretical PSF and simulated transcription site. Images are shown as maximum intensity projections along the major axis. (**Upper row**) Theoretical PSF was generated with ImageJ PSF-Generator. Emission wavelength = 568 nm, numerical aperture = 1.25, refractive index = 1.46. (**Lower row**) Simulated transcription sites with 50 nascent transcripts and a radius of. No noise was added.



We then simulated transcription sites with different spatial extent and varying amounts of nascent mRNA. We repeated the simulations for each condition 5 times and averaged the obtained estimates for each of the different quantification methods.

### 5.2. Transcription sites without spatial extent and no noise

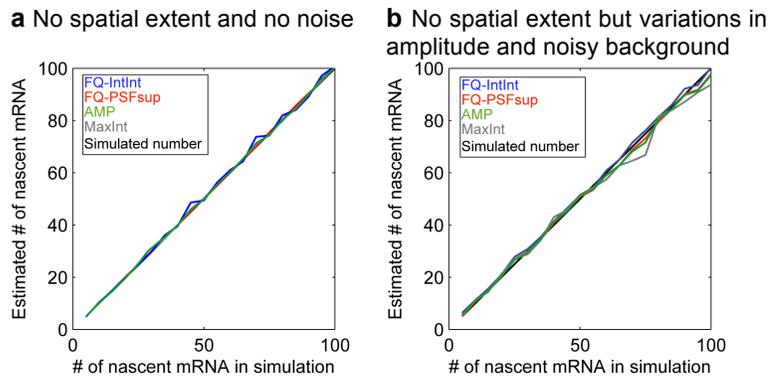
First, we simulated transcription sites without any spatial extent and in absence of noise. All mRNA were placed at the same location with the same amplitude. No noisy background was added. Under these idealized conditions all quantification methods worked well for the entire tested range of nascent mRNA abundance (5-100) (Fig. S14a).

### 5.3. Transcription site with noise

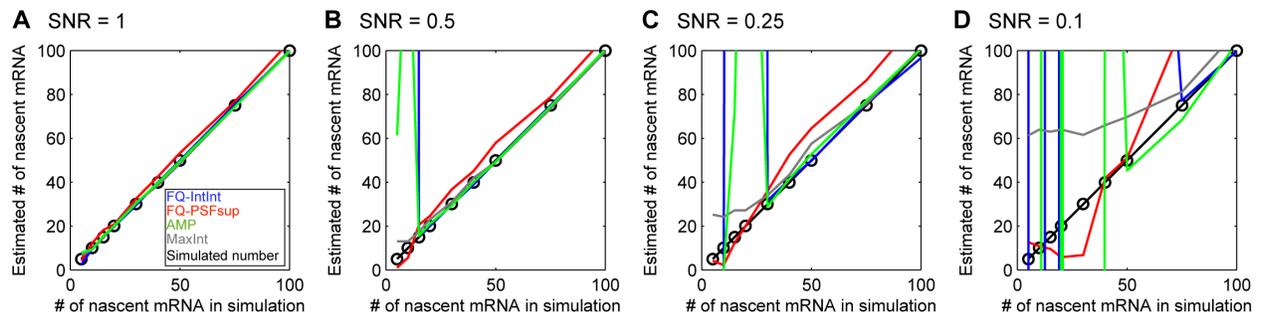
For the next simulations, we placed the mRNA again at the exact same location but now considered the experimentally observed variability of their brightness and the effect of noise. We generated the

amplitudes of the placed mRNA molecules randomly following a Gaussian distribution with a mean of 100 and standard deviation of 30, roughly similar to the empirically measured distribution of Figure S7. Furthermore, we simulated a noisy background by adding Gaussian noise of mean 500 and variable standard deviation to achieve a range of signal-to-noise ratios (Supplementary Note 3.1.). For the lowest experimentally observed SNR of 5, all methods again yielded accurate results (Fig. S14b).

**Figure S14.** Nascent mRNA counting in simulated images of sub-diffraction transcription sites. Plots show estimated number of nascent transcripts as a function of the simulated number of transcripts. Each data-point is the average of 5 individual simulations. FQ-PSFsup: red, FQ-IntInt: blue, MaxInt: green, AMP: gray. **(a)** Without fluctuations of the intensity of the placed mRNA's and no noisy background. **(b)** With variable amplitudes and noisy background (SNR = 5). In either scenario all methods yield accurate estimates.



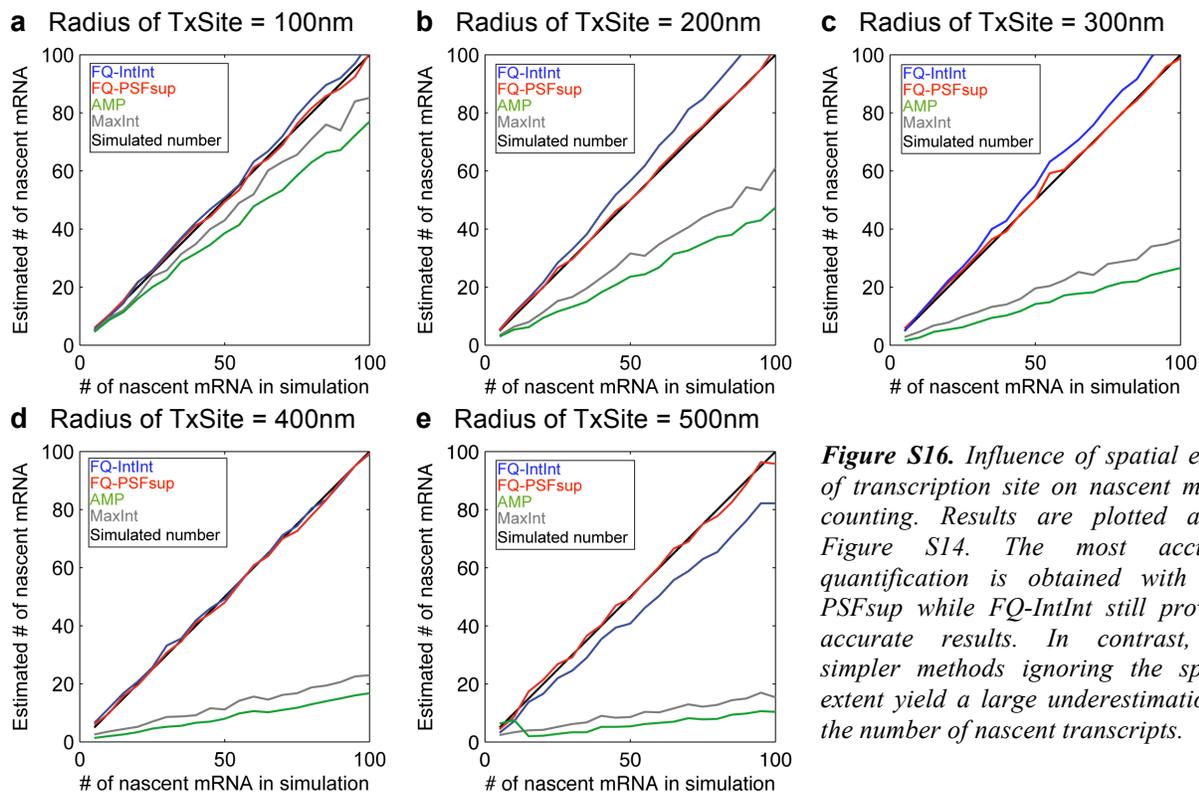
We then investigated lower SNR than experimentally observed (Fig. S15). We still obtained accurate quantification for SNR as low as 1, but the methods started to fail for SNR below 1. As expected, the quantifications fail first for sites with fewer transcripts (Fig S15), whereas sites more enriched in transcripts still have sufficiently high signal to be quantified. However, such low SNR will not typically occur in typical FISH images (Supplementary Note 3.1.), otherwise individual mRNA could no longer be detected.



**Figure S15.** Accuracy of transcription site quantification for very low SNR for individual mRNA molecules. At these noise levels individual mRNA molecules cannot be detected, so noise-free images of the individual mRNA molecules were used for the quantification.

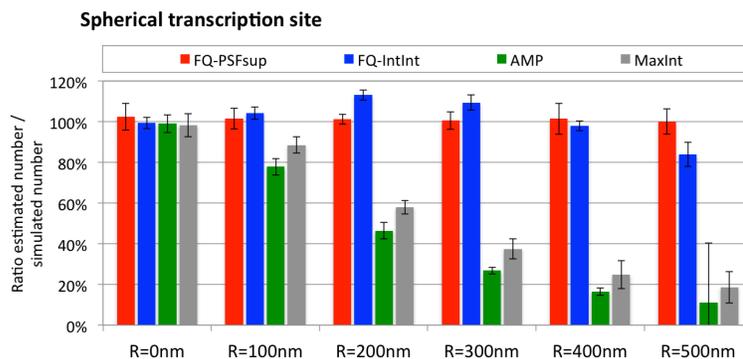
#### 5.4. Spherical transcription sites

Next we investigated the impact of spatially extended transcription sites on the quantification results. We therefore simulated spherical transcription sites with increasing radius in which the mRNAs were randomly placed. We considered variations in the amplitude and added a noisy background with SNR=5 as described above. We obtained excellent agreement with the FQ-PSFsup for all radii and the estimates stayed within 3% of the true number (Compare red to black lines in Fig. S16 and Fig. S17). FQ-IntInt yielded good agreement as well and the estimates were within 15% error (Compare blue to black lines in Fig. S16, Fig. S17). The method based on comparisons of amplitude or maximum intensity, however, significantly underestimated the number of nascent transcripts by up to 80% (compare green and gray lines to black lines in Fig. S16 and Fig. S17).



**Figure S16.** Influence of spatial extent of transcription site on nascent mRNA counting. Results are plotted as in Figure S14. The most accurate quantification is obtained with *FQ-PSFsup* while *FQ-IntInt* still provides accurate results. In contrast, the simpler methods ignoring the spatial extent yield a large underestimation of the number of nascent transcripts.

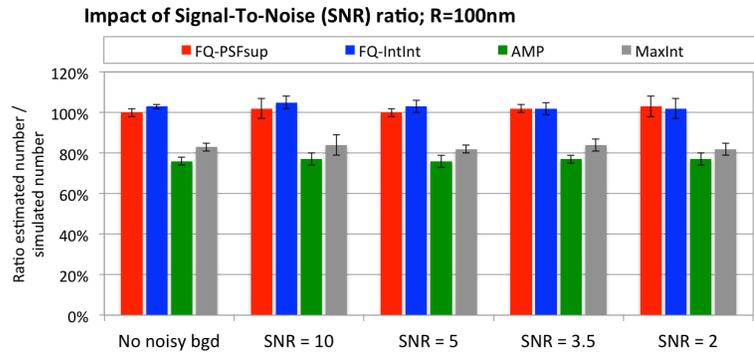
**Figure S17.** Accuracy of transcription site quantification for spherical sites. For each individual simulation the ratio of the estimated number of nascent transcripts and the actually simulated number of nascent transcripts was calculated. Then median value and standard deviation of these ratios are shown as bar plots.



### 5.5. Spherical transcription sites and noisy background

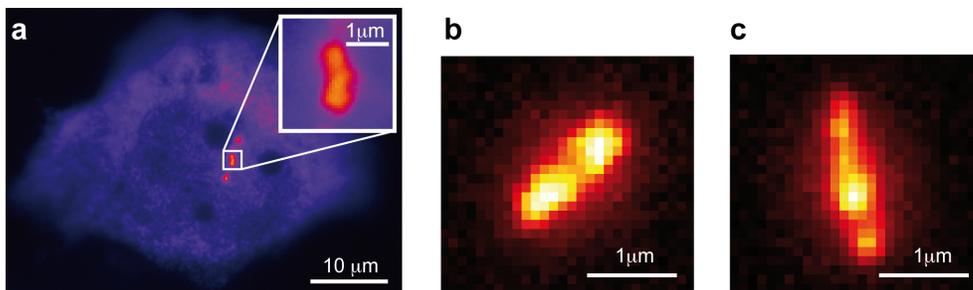
We then simulated the impact of different SNR on spherical transcription sites with a radius of 100nm (Fig. S16a and S17). As before, we found that varying SNR did not affect the quantification results. However, the simpler methods always underestimate the number of transcripts, while the FISH-quant methods provide accurate estimates (Compare Fig. S17 and Fig. S18). Similar results were found for larger transcription sites (data not shown).

**Figure S18.** Influence of SNR on accuracy of transcription site quantification for spherical site of radius 100nm. Results of quantification are presented as in Fig. S17.



## 5.6. Ellipsoidal transcription sites

In the above simulations we assumed spherical transcription sites. As described above, many biological samples will, however, not show such perfect symmetry and more complex topologies such as elongated structures or V-shaped transcriptions sites can be observed, e.g. for viruses, genes transcribing large repeated non-coding RNAs, and gene arrays<sup>12,14-19</sup> (Fig. S19a). We therefore simulated ellipsoidal transcription sites to investigate the effect of less symmetrical sites. Ellipsoids were simulated with different ratios of the three semi-axes and rotated randomly in 3D to consider different spatial orientations (Fig 19b, c).



**Figure S19.** Elongated transcriptions sites are simulated as ellipsoids. **(a)** U2OS cells expressing an HIV-1 reporter gene expressed from a gene array (Exo1 cells<sup>12</sup>). Cells were hybridized in situ with a probe against the reporter RNA, with single molecule sensitivity. **(b, c)** Simulated ellipsoidal transcription site. Semi-axes are 900 nm, 300 nm, and 300 nm **(b)** and 1000 nm, 250 nm and 250 nm **(c)**. Ellipsoid are rotated counter-clockwise by 45°**(b)** and 100°**(c)**. Rotation is 2D for illustration purposes only, in simulations 3D rotations were applied.

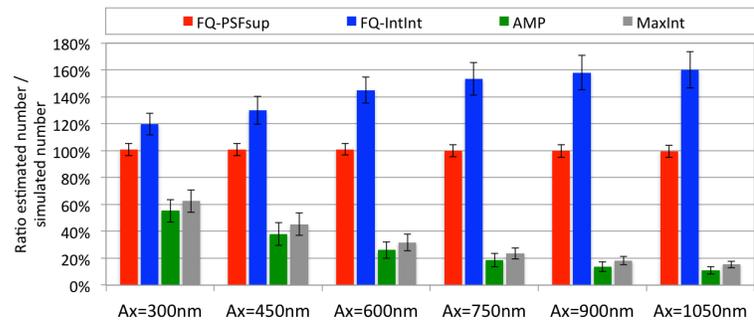
We simulated different mRNA abundances (5-100) and repeated each simulation 50 times. Validation results are shown in Fig. S20. The quantification with FQ-PSFsup stayed within 4% of the true number, while the quantification with FQ-IntInt led to an over-estimation of up to 60%. The simpler methods underestimated mRNA counts by up to 85%.

**Figure S20.** Summary of quantification for ellipsoidal transcription sites. Sites were simulated with different ratios and lengths of the semi-axes. In addition, sites were rotated randomly in 3D to consider different spatial orientations. Results of quantification are presented as in Fig. S17.

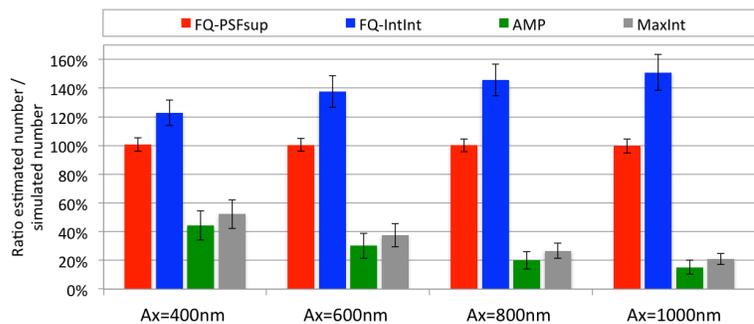
**(Upper plot)** Ellipsoidal transcription site with a ratio of the semi-axes  $AX:AY:AZ = 3:1:1$ . Length of the longest axis is indicated below each group of bars.

**(Lower plot)** Ellipsoidal transcription site with a ratio of the semi-axes  $AX:AY:AZ = 4:1:1$ .

**Ellipsoidal transcription sites:  $AX:AY:AZ = 3:1:1$**

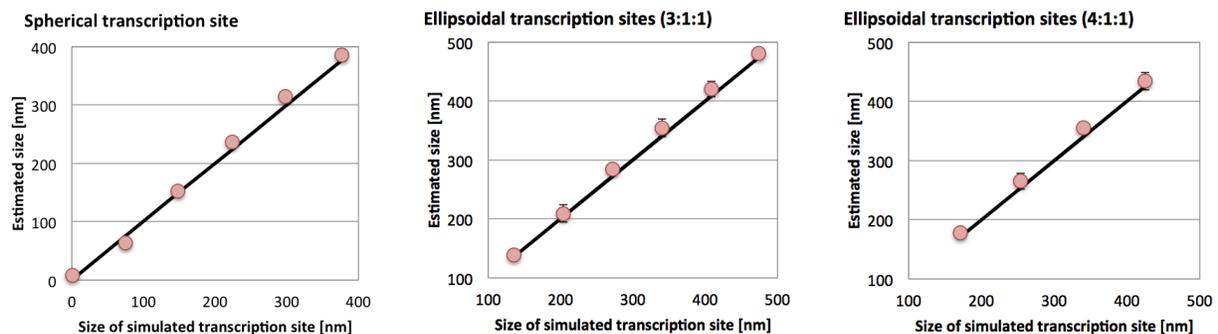


**Ellipsoidal transcription sites:  $AX:AY:AZ = 4:1:1$**



## 5.7. Analysis of the spatial extent of the transcription site

The PSF superposition approach does not only yield the number of nascent transcripts but also information about their spatial positioning. While the precise locations of individual mRNA molecules cannot be determined, we can still calculate ensemble quantities such as the averaged distance from their center of mass to measure the spatial extent of the transcription site. Figure S21 summarizes the results of the size measurement for the transcription sites simulated in the preceding sections. For each simulated size we averaged the estimated size by FQ-PSFsup (for sites with more than 20 transcripts). The estimated size was in good agreement to the simulated size for all considered transcription site geometries.



**Figure S21.** Size of the transcription sites from Fig S17, S20. Size is measured as the averaged distance of each individual mRNA to the center of mass of the transcription site. Plots show the estimated size ( $y$ ) vs. the actual size of the simulated size ( $x$ ). The estimated size is in good agreement with the true size.

## 5.8. Summary of validation with simulations

In this section we investigated the impact of noise and spatial extent on the different transcription site quantification methods. Our results indicate that the quantification accuracy is not strongly affected by the typical noise observed in FISH, but instead depends on a proper consideration of the spatial extent.

We found that all methods are robust to experimental noise. This robustness can be explained by two factors. First, the signal-to-noise ratio (SNR) of individual mRNA molecules is typically very high in FISH images (Supplementary Note 3.1.), and transcription sites will have an even higher SNR. Second, by considering the averaged image of all detected mRNA molecules (frequently several thousands) we minimize the impact of noise in the analysis of individual mRNAs.

We found, however, that considering the spatial extent is important for a reliable quantification. We found that **FQ-PSsup** performed reliably for all simulated sites, independently of their spatial extent and simulated geometry, whereas the simpler methods ignoring this extent can grossly underestimate the number of transcripts. **FQ-IntInt** yielded accurate estimates for symmetrical, spherical sites but overestimated the number of transcripts for elongated, ellipsoidal sites.

The choice of the quantification methods therefore depends on the typical shape of the observed transcription site. For rather compact, symmetrical sites, both FISH-quant methods yield accurate results. We would therefore recommend using both methods and verify if the results obtained are comparable as an internal quality-check for the quantification. FQ-IntInt has the advantage of being computationally faster than FQ-PSFsup. So if computational time becomes an issue FQ-IntInt can be used alone. For spatially elongated transcription sites, e.g. as can be found for viruses, genes transcribing large repeated non-coding RNAs, and gene arrays, we recommend using FQ-PSsup since only this method accurately quantifies the number of nascent transcripts for these more complex structures. The methods based on a comparison of amplitude or maximum intensity underestimated the number of nascent transcripts for larger, spatially extended site. However, these methods still provided accurate results for diffraction limited sites.

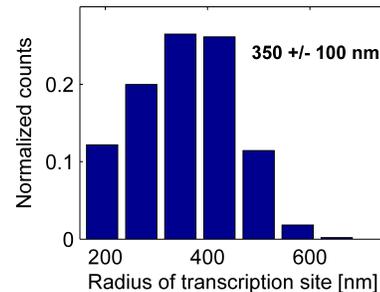
**We quantified the spatial extent of the transcription site** by calculating the average distance of each transcript to the center of the site. The distance estimated by FQ-PSFsup was in good agreement with the actual size of the transcription site. These estimates could therefore be used to quantify the spatial extent of the sites and relate this to biological properties such as the decondensation state of a transcriptionally active locus.

## 6. Validation of transcription site quantification with experimental data

### 6.1. Hygro-MS2x96-bGH: transcription site quantification

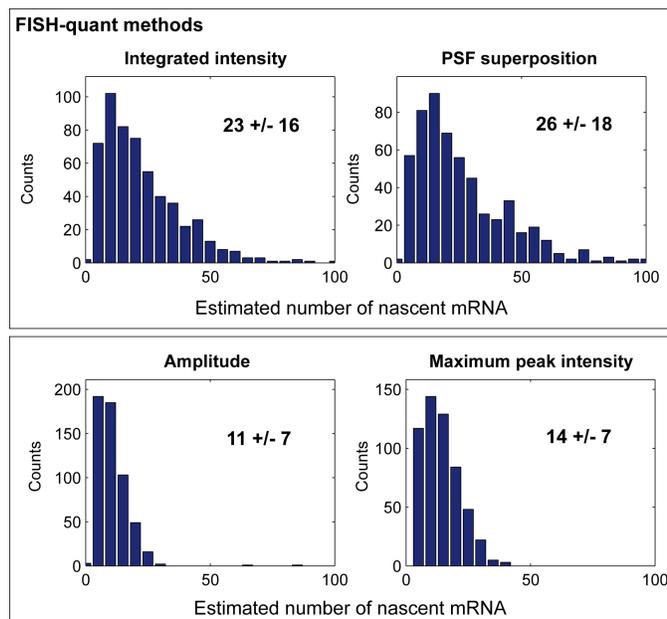
For experimental validation, we analyzed transcription sites in Hygro-MS2x96-bGH cells. The estimated size with FQ-PSFsup reveals that the majority of the sites are larger than the diffraction limit (Fig. S22).

**Figure S22.** Estimated radius of the transcription sites for Hygro-MS2x96-bGH. In addition the mean value +/- standard deviation is reported.



In our simulations we found that for sites that are larger than the diffraction limit the FISH-quant methods estimated larger numbers of nascent transcripts than the simpler methods (Supplementary Note 5). The quantification results of the amount of the nascent mRNA in Hygro-MS2x96-bGH cells with the four methods revealed identical trends (Fig. S23). The FISH-quant methods estimated twice as many nascent transcripts than the simpler methods.

**Figure S23.** Transcription site quantification with the different quantification methods. A total of 552 transcription sites were analyzed. Each plot shows the histogram of the amount of nascent mRNA per site as estimated with the method indicated in the title. In addition the mean value +/- standard deviation is reported.

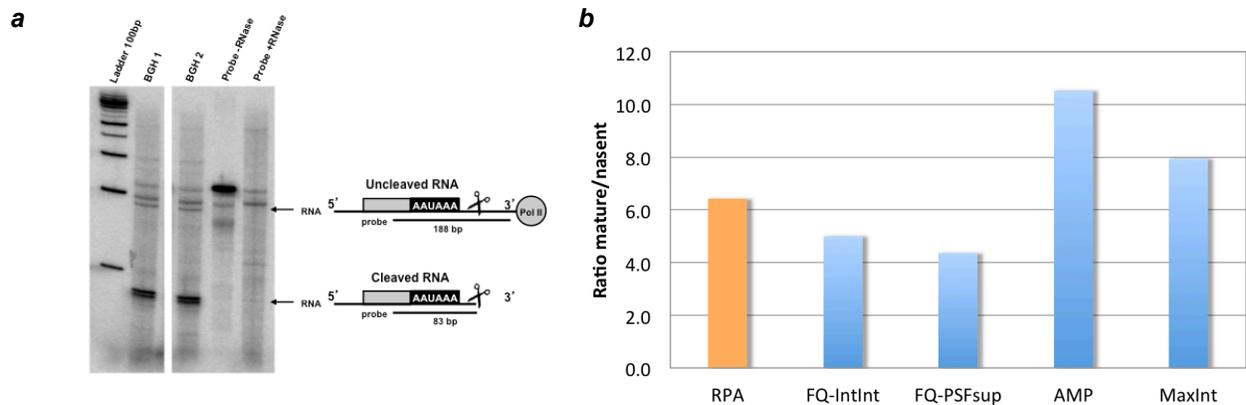


### 6.2. Hygro-MS2x96-bGH: FISH-quant vs. RNase protection assay

We attempted to further validate the transcription site quantification by comparing the ratio of mature vs. nascent mRNA estimated by FISH to the values estimated by RNase protection assay (RPA, Supplementary Methods). Using a probe that spans the 3'-end cleavage and polyadenylation site, RPA allows to detect 3'-end cleaved and uncleaved mRNA. The ratio of cleaved vs. uncleaved mRNA can then be used as an approximation of the ratio of mature vs. nascent mRNA and thus compared to the ratio estimated by FISH-quant. We used samples from the same day and experiments were performed in

triplicates. Nevertheless, the accuracy of RPA is limited. RPA measurements are based on the quantification of bands from a gel (Fig. S24a). Because nascent mRNA is not very abundant, it appears as a dim band and therefore its quantification is prone to uncertainties. Despite these limitations, RPA can be used to estimate the order of magnitude of nascent vs. mature mRNA.

The cleaved/uncleaved ratio estimated by RPA fell between the values obtained by the FISH-quant methods and the simpler methods (Fig. S24b). Because of its limited accuracy, RPA cannot be used to favor one method over the other. Nevertheless, it confirms the general validity of using imaging based methods to measure mRNA content.

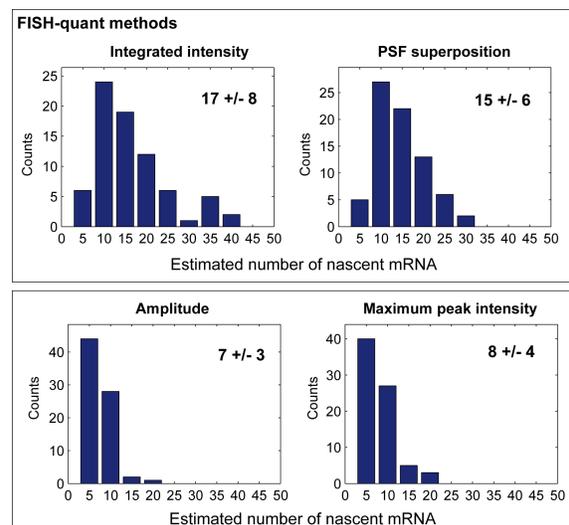


**Figure S24.** Cleaved (mature) vs Uncleaved (nascent) mRNA ratio determination by Ribonuclease Protection Assay (RPA) and FISH methods. **(a)** The protected radiolabeled probe hybridized to complementary RNA is separated on polyacrylamide gel. Schematization of Cleaved and Uncleaved mRNAs with the position and the length of the hybridized probe indicated on the right. **(b)** Ratio of mature vs. nascent Hydro-MS2x96-bGH mRNA as estimated by RPA and FISH methods.

### 6.3. Validation of transcription site quantification on $\beta$ -actin

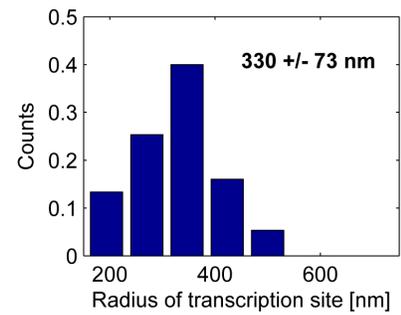
We then analyzed the transcription activity of  $\beta$ -actin. This gene has been studied by mRNA FISH in the pioneering study of Femino et al<sup>11</sup> and showed strong activation after serum induction. We repeated the experiment in U2OS cells and obtained similar results. We see practically no active transcription sites before induction and a strong activity 20 min after serum induction. As before the FISH-quant methods estimated larger numbers of nascent transcripts than the simpler methods (Fig. S25).

**Figure S25.** Transcription site quantification for  $\beta$ -actin 20 min after serum induction with the different quantification methods. Each plot shows the histogram of the amount of nascent mRNA per site as estimated with the method indicated in the title. In addition the mean value +/- standard deviation is reported.



The estimated size of the transcription is larger than the diffraction limit (Fig. S26) providing evidence why the FISH-quant methods estimated larger number of nascent transcripts.

**Figure S26.** Estimated radiuses of the transcription sites for  $\beta$ -actin 20 min after serum induction. In addition the mean value +/- standard deviation is reported.



## Supplementary Methods

### Plasmids and cell lines

#### Hygro-MS2x96-bGH reporter

Frt-hygro fragment from pCDNA5/FRT was inserted in BamHI site of BAC2+bs86 (XXV)<sup>24</sup>. Bovine growth hormone (bGH) polyadenylation signal was inserted in MluI site of BAC2+bs96-frthygro to generate the Hygro-MS2x96-bGH reporter. The reporter was stably integrated in frt site of flp-in-293 cell line (Life Technologies) as recommended by manufacturer. The resulting cell lines flp-in-293-Hygro-MS2x96-bGH were cultured at 37°C in DMEM with 10% FBS and 50 µg/ml of hygromycin.

#### β-globin-Luc-CFP-24MS2

The plasmids pFRT/LacZeo, pOG44 and pCDNA5/FRT were supplied by Invitrogen, pTet-On by Clontech, pBslacO containing 40 LacI binding sites is a gift from M. Ackermann, Institute of Virology, Zurich, Switzerland<sup>25</sup> and pSV2-EYFP/ lac repressor is a gift from DL. Spector, Cold Spring Harbor Laboratory, New York, USA<sup>26</sup>. A pTet-globin-Luc-CFP-24MS2 was generated by inserting in the pTet-globin-CFP-18MS2-2 construct<sup>27</sup> the PCR amplified Luc gene at the BstXI site and 24MS2 repeats from pSL-24X<sup>28</sup> replacing the 18MS2. The FRT-TOLCM vector was generated by inserting pTet-globin-Luc-CFP-24MS2 in the pFRT/LacZeo backbone with restriction enzymes ApaI and NruI. The cell line U2OS was cultured at 37°C in low glucose DMEM with 10% FBS (optionally supplemented with 150µg/ml Zeocin or 100 mg/ml Hygromycin). Following calcium phosphate transfection of a mixture of 1:10 of pFRT/LacZeo and 9:10 of pBslacO, U2OS cell clones having integrated pFRT/LacZeo were selected on Zeocin. The clones containing a single tandem array of lac operator sites next to a single FRT site were selected. One of these clones (A33-8) was then co-transfected with pOG44 and FRT-TOLCM by FuGENE (Roche). One clone (A33-8-T1) that had integrated the single copy of FRT-TOLCM plasmid at the FRT site via Flp recombinase mediated DNA recombination was selected with Hygromycin.

#### c-Fos: Normal Human Dermal Fibroblast Cell

FISH against c-FOS was performed in Normal Human Dermal Fibroblast Cell (NHDF). Cells were isolated from the dermis of adult skin (Promocell, C-12302). Primary fibroblast cultures were maintained at 37°C in DMEM-F12 (Invitrogen), supplemented with 10% FBS. Primary cell cultures were transferred into collagen (Gibco) coated culture plates and maintained in DMEM-F12, supplemented with 10% FBS. The culture medium was removed 1 day after passage to start serum starvation, the cells were washed with PBS and fresh medium containing no FBS was added for 24h. The culture medium was removed and fresh medium containing 10% FBS was added for 20min before fixation.

#### RBP1 and β-actin: U2OS

FISH against RBP1 and β-actin was performed in U2OS cells. Cells were cultured at 37°C in DMEM with 10% FBS.

### In situ hybridization and imaging

In situ hybridization was performed as described previously<sup>28</sup>. The formamide concentration for FISH in hybridization and washing mixture was 30% against MS2, and 40% against RPB1, β-actin, β-globin-Luc-CFP-24MS2, and c-Fos. 10ng (MS2, RPB1, β-globin-Luc-CFP-24MS2, and c-Fos) or 5ng (β-actin) of probes were used per 50 microliters of the hybridization mixture.

For β-globin-Luc-CFP-24MS2 cells were co-transfected with pTet-On and pSV2-EYFP/ lac repressor by FuGENE (Roche) 18h before fixation.

**Sequence of MS2 probe** (X stands for amino-allyl T). This probe binds 48 times to the Hygro-MS2x96-bGH reporter. It was labeled with Cy3 (GE Healthcare).

AXACATGGGTGATCCTCATGTTACCCAXGCTCTAGCACACATGGGTGATCCTCATGTXA

**Sequence of probes against RBP1** (X stands for amino-allyl T). Even-numbered probes were labeled with Alexa 488 (Life technology) and odd-numbered probes with Cy3 (GE Healthcare) following the protocols of suppliers. 0.5ng of each probe was used for hybridization.

Probe 01	CCCXCAGTCGTCXCTGGGTATXTGATGCCACCCXCCGTCACAGACATXCG
Probe 02	TXTCTTTGGTCAGAXCCTCGTCACCCXCAGGTTGTGCCACACCGAACTXG
Probe 03	GGAXCTTCTTCTCCXGAGAGTCCTCAXTAACGTGCTXCCATTCCGCAXAC
Probe 04	AAACACXCCTCATCXGAGATGCGTXXGAAGATCTCAXGCACTCGCTCXG
Probe 05	TXGATCTTCACGAXGTCAGCCAGTXXGTGAGTCAGGXCATCCTGGTXACG
Probe 06	XGGAGGAGCTTCACAXCCTCTGCAAXGACATGGGCCGCXGCGCCGTTXCG
Probe 07	CGAXCACCATTGTCXCGGATGATGTACXTGGCGCCTGGGXA CTGGCTGTXC
Probe 08	ACAXGTGCCGTTCCACCXTATAGCCGGXCTGCAGGXGAAGGTCACXGGG
Probe 09	XCGGAGTTGXCACACTAAGATXCAAGCGAAAGGXAGACCATGGGAGAAXGC
Probe 10	CXCTGCTCGCGXCTCCAGAGACXGTGGCAGGXGCAAGTTCATCTCAXCC
Probe 11	XGGGACCTXCCATCCCACGXCAGACAGGAACAXCAGGAGGTTCAXCAC
Probe 12	XCGGGATGGGTACXGTGGGTACGGAXACAATTGATGXGACCAGGTATGAXG
Probe 13	XCGATGAGGAGCCAGXTGTTAATGACAGXCTGAATGTXXGGAGTAGAAGAGXA
Probe 14	XGCTTGGCCTTCTXAATAGTGTTCXGAATGCTCTGGXAAGTCTTAGAAXC
Probe 15	XGGAACCTTAGCXCCGGACACGACCAXAGACTGAAGXTATTGTATXCAG
Probe 16	XCACATCAAAGXCAGGCATTTCAAGTAGACATXCACCCATTCTTGAXCC
Probe 17	GXGAGCTTCCGGXCAGTCATGTGCXTCCGATCCAGCXCCACCCGCAACAGXA
Probe 18	XCTTCTCCTCXTGCATCTTGTXTCATCGCTGTXCATGATGCGAAXACG
Probe 19	XCAGCATGXTGGACTCGAXGCAGCGCAGGAAGACAXCATCATCCATCTXG
Probe 20	CCAXCCTCCGTGAXGATGATCTTCTTCTXGTTGTCTGTCTXGTTGGCAAGXG
Probe 21	XTTCTCACXCAGCACCCGCAXCAAGCTCACGCCXCCGTCTCCAGGAXCCAC
Probe 22	AAGXGTCGGTAATXGACATAGGAGCCAXCAAAGGAGAXGACGTGGTACAGXA
Probe 23	XAAGGAACACTXCATGAGTGGTCCXGTGTCTGGCGGXTGACTCCGTGXCG
Probe 24	ACGXTGGCGAGTAGCXGGGAGACAXGGCACCACCTGGXGAAGGGATGXAG
Probe 25	GGAGAGGXCGGTGAGTAGCXGGGTGACGTXXGGCGAATAGCXGGGTGATGXG
Probe 26	AAXTGGGACTGGTXXGAGAATAGTXCGGGCTGGXGGGTGAGTAACTXGGG
Probe 27	AXAGGTGGGACXGGTAGGCGAGXACTTGGGAGAGGXGGGTGAATATTXGG
Probe 28	TCXCCTCGTCACTGXCATCCGGGCTGAXAGCCGGGCTXGTGAGACTGXAG
Probe 29	TCXGCATCAGAAACGGGAXCCAGAAGTXCACCCGGGAGCXCTGCCACAAGGXT
Probe 30	XCTTTGTTCTXCCCAGGATCAGCXGTAACCACXCACAGCAGGAACXACCC

**Sequence of probes against  $\beta$ -actin** (X stands for amino-allyl T). Probes were labeled with Cy3 (GE Healthcare).

Probe 01	AXTGTAGAAGGXGTGGTGCCAGAXTTTCTCCATGXCGTCCCAGTTGGXGA
Probe 02	GCCXGGATAGCAACGXACATGGCTGGGGXGTTGAAGGXCTCAAACAXGAT
Probe 03	GAAGXCCAGGGCGACGXAGCACAGCTXCTCCTTAATGXCACGCACGATXT
Probe 04	AXGTCCACGTCACACXTCATGATGGAGXTGAAGGTAGXTTCGTGGAXGCC
Probe 05	XAACGCAACTAAGTCAXAGTCCGCCXAGAAGCATTXGCGGTGGACGAXGGA

**Sequence of  $\beta$ -globin-Luc-CFP-24MS2 probes** (X stands for amino-allyl T). Exonic probes were labeled with Cy3 (GE Healthcare) and intronic probes with Cy5 (GE Healthcare) following the protocols of suppliers. 0.5ng of each probe was used for hybridization.

β-globine Exon 1	AGGAGXCAGGTGCACCAAXGGTGTCTGTTXGAGGTTGCTAGXGAACACAGTA
β-globine Exon 2	GCCCAXAACAGCAXCAGGAGTGGACAGAXCCCCAAAGGACXCAAAGAACC
β-globine Exon 3-CFP	XGAACAGCTCCXCGCCCTTGXCACCATGAATTCXTTGCCAAAGTGAXGG
Luciferase 1	GCGGXTCCATCCTCXAGAGGATAGAAAXGGCGCCGGGCCXTTCTTTATGXT
Luciferase 2	TGTXCCAGGAACCAGGGCGXATCTCTTCAXAGCCTTATGCAGXTGCTCTXA
Luciferase 3	XCCAACCGAACGGACAXTTTCAAGTAXTCCGCGTACGXGATGTTTACCXCG
Luciferase 4	XAACCGGGAGGXAGATGAGATGXGACGAACGTGTACAXCGACTGAAAXCCC
Luciferase 5	XAAAATAGGAXCTCTGGCAXGCGAGAATCXGACGCAGGCAGTTCTAXGCGG
MS2 NBX (12 repeat)	CXAGGCAATXAGGTACCTXAGGATCTAAXGAACCCGGGAATACXGCAGAC
β-globine Exon1-Intron 1	GXCTTGTAACTXGATACCAACCXGCCAGGGCCXCACCACCAACTTCATA
β-globine Intron 1	XCAGTGCCTAXCAGAAACCCAAGAGXCTTCTCTGTCCACATGCCAGXA
β-globine Intron 2	XAGCAAAGGGCCXAGCTTGGACXCAGAATAAXCCAGCCTTAXCCCAACCA

**Sequence of probes against c-Fos** (X stands for amino-allyl T). Probes were labeled with Cy3 (GE Healthcare).

c-Fos 188E	CXCGTAGTCTGCGTXGAAGCCCCGAGAACAXCATCGTGGCGGXTAGGCAAAXA
c-Fos 288	XGACAGGCGAGCCCAXGCTGGAGAAGGAGXCTGCGGGTGAGTGGXAGTAAGXA
c-Fos 1123	XCCGGACTGGXCGAGATGGCAGXGACCGTGGGAAXGAAGTTGGCACXGGAG
c-Fos 1806E	XTGCGGCATTXGGCTGCAGCCAXCTTATTCCTTXXCCCTTCGGATTXCCT
c-Fos 2007E	XGGCAATCTCGGXCTGCAAAGCAGACXTCTCATCTTCXAGTTGGTCTGXC
c-Fos 2083	AGGXCATCAGGGATCTXGACAGCAGGXCGGTGAGCXGCCAGGATGAACTA
c-Fos 2270E	GAAGXCATCAAAGGGCXCGTCTTCAGCXCCATGCTGCXGATGCTCTXGA
c-Fos 2382	XAGCCACTGXGCAGAGGCTCCCAGXCTGCTGCAXAGAAGGACCCAGAXAGG
c-Fos 2485	XGAAGACGAAGGAAGCAGXGTAAGCAGXGACAGCTGGGAGXACAGGTGACXT
c-Fos 2676	AXGTGTTTCTCCXCTGTAAAXGCACCAGCXCGGGCAGTGGCAGTGGCAGG
c-Fos 2727	TXCACGCACAGAXAAGGTCCXCCTAGGTTCXACAGGAACCCXCTAGGGAA
c-Fos 2781	CXTGAGTCCACACAXGGATGCTTXXCAAGTCTCXGAGGCCACAGCCXGGT
c-Fos 2832	XGGAACAATACACACXCCATGCGTTTXXGCTACATCXCCGGAAGAGGXAAAG
c-Fos 2883	CCAGGCCXGGCTCAACAXGCTACTAACXACCAGCTCTCXGAAGTGTACXG

The modified oligonucleotide probes for MS2, RPB1, and c-Fos were synthesized by J-M. Escudier (Plateforme de synthèse d'Oligonucléotides modifiés de l'Interface Chimie Biologie de l'ITAV, Toulouse, France). The modified oligonucleotide probes for β-actin and β-globin-Luc-CFP-24MS2 were synthesized by Eurogentec (Seraing, Belgium).

**Imaging of MS2.** 3D image stacks of cells after in situ hybridization were captured on a 100x NA 1.4 wide-field microscope (DMRA; Leica) equipped with a camera (CoolSNAP HQ; Roper Scientific) and controlled by MetaMorph software (Universal Imaging Corp.). Pixel-size of 160 nm. Z-stacks of 61 images with a 300-nm Z-step were used.

**Imaging of RPB1, β-actin, and β-globin-Luc-CFP-24MS2.** 3D image stacks of cells after in situ hybridization were captured on a 100x NA 1.4 wide-field microscope (ECLIPSE Ti; Nikon) equipped with a camera (CoolSNAP HQ; Roper Scientific) and controlled by MetaMorph software (Universal Imaging Corp.). Pixel-size of 160 nm. Z-stacks of 51 images with a 200-nm Z-step were used.

## RNase protection assay

188 nucleotides fragment of bGH poly-adenylation signal spanning RNA cleavage site was amplified by PCR and cloned in pCRII-TOPO vector (Invitrogen). 32P-UTP labeled antisense RNA probe was synthesized with T7 RNA polymerase using Riboprobe in vitro transcription kit (Promega). RNase protection assay was performed using Ambion kit RPAIII, according to manufacturer protocol. The protected fragments were run on 6% denaturing acrylamide gel, which was dried and exposed in Phosphorimager. 188 nucleotides band corresponded to noncleaved mRNA and 83 nucleotides band corresponded to cleaved mRNA. The intensities of the bands were quantified by ImageJ.

## References

1. Thomann, D., Rines, D. R., Sorger, P. K. & Danuser, G. Automatic fluorescent tag detection in 3D with super-resolution: application to the analysis of chromosome movement. *J Microsc* **208**, 49–64 (2002).
2. Henriques, R. *et al.* QuickPALM: 3D real-time photoactivation nanoscopy image processing in ImageJ. *Nat. Methods* **7**, 339–340 (2010).
3. Piotr's Matlab Toolbox. at <<http://vision.ucsd.edu/~pdollar/toolbox/doc/>>
4. Zhang, B., Zerubia, J. & Olivo-Marin, J.-C. Gaussian approximations of fluorescence microscope point-spread function models. *Appl. Opt.* **46**, 1819–1829 (2007).
5. Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* **4**, e309 (2006).
6. Raj, A., Van den Bogaard, P., Rifkin, S. A., Van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Meth* **5**, 877–879 (2008).
7. Parthasarathy, R. Rapid, accurate particle tracking by calculation of radial symmetry centers. *Nat. Methods* **9**, 724–726 (2012).
8. Ober, R. J., Ram, S. & Ward, E. S. Localization accuracy in single-molecule microscopy. *Biophys. J* **86**, 1185–1200 (2004).
9. Kirshner, H., Sager, D. & Unser, M. in *Proceedings of the Twelfth International Conference on Methods and Applications of Fluorescence Spectroscopy, Imaging and Probes* 154 (2011).
10. Zenklusen, D., Larson, D. R. & Singer, R. H. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol* **15**, 1263–1271 (2008).
11. Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of single RNA transcripts in situ. *Science* **280**, 585–590 (1998).
12. Boireau, S. *et al.* The transcriptional cycle of HIV-1 in real-time and live cells. *J. Cell Biol.* **179**, 291–304 (2007).
13. Decker, C. J. & Parker, R. P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. *Cold Spring Harb Perspect Biol* **4**, a012286 (2012).
14. Müller, W. G., Walker, D., Hager, G. L. & McNally, J. G. Large-scale chromatin decondensation and recondensation regulated by transcription from a natural promoter. *J. Cell Biol* **154**, 33–48 (2001).
15. Darzacq, X. *et al.* In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **14**, 796–806 (2007).
16. Lawrence, J. B., Singer, R. H. & Marselle, L. M. Highly localized tracks of specific transcripts within interphase nuclei visualized by in situ hybridization. *Cell* **57**, 493–502 (1989).
17. Bellemer, C. *et al.* Microprocessor dynamics and interactions at endogenous imprinted C19MC microRNA genes. *J. Cell. Sci.* **125**, 2709–2720 (2012).

18. Vitali, P., Royo, H., Marty, V., Bortolin-Cavaillé, M.-L. & Cavaillé, J. Long nuclear-retained non-coding RNAs and allele-specific higher-order chromatin organization at imprinted snoRNA gene arrays. *J. Cell. Sci.* **123**, 70–83 (2010).
19. Royo, H. *et al.* Bsr, a nuclear-retained RNA with monoallelic expression. *Mol. Biol. Cell* **18**, 2817–2827 (2007).
20. Wegel, E. & Shaw, P. Gene activation and deactivation related changes in the three-dimensional structure of chromatin. *Chromosoma* **114**, 331–337 (2005).
21. Belmont, A. S. & Straight, A. F. In vivo visualization of chromosomes using lac operator-repressor binding. *Trends Cell Biol.* **8**, 121–124 (1998).
22. Tilgner, H. *et al.* Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* **22**, 1616–1625 (2012).
23. Schmidt, U. *et al.* Real-time imaging of cotranscriptional splicing reveals a kinetic model that reduces noise: implications for alternative splicing regulation. *J. Cell Biol.* **193**, 819–829 (2011).
24. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025–1036 (2005).
25. Fraefel, C. *et al.* Spatial and temporal organization of adeno-associated virus DNA replication in live cells. *J. Virol.* **78**, 389–398 (2004).
26. Tsukamoto, T. *et al.* Visualization of gene activity in living cells. *Nat. Cell Biol* **2**, 871–878 (2000).
27. Darzacq, X. *et al.* Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. *J. Cell Biol.* **173**, 207–218 (2006).
28. Fusco, D. *et al.* Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. *Curr. Biol.* **13**, 161–167 (2003).