Published in final edited form as:

Small Methods. 2018 September 11; 2(9): . doi:10.1002/smt.201700324.

Advanced 3D Analysis and Optimization of Single-Molecule FISH in *Drosophila* Muscle

Akiko Noma⁺,

RNA Therapeutics Institute, University of Massachusetts Medical School, Worcester, MA 01605, USA

Carlos S. Smith⁺⁺,

RNA Therapeutics Institute, University of Massachusetts Medical School, Worcester, MA 01605, USA

Maximiliaan Huisman,

RNA Therapeutics Institute, University of Massachusetts Medical School, Worcester, MA 01605, USA

Robert M. Martin,

Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, 1600-276 Lisboa, Portugal

Melissa J. Moore⁺⁺⁺,

RNA Therapeutics Institute, University of Massachusetts Medical School, Worcester, MA 01605, USA

David Grunwald

RNA Therapeutics Institute, University of Massachusetts Medical School, Worcester, MA 01605, USA

Abstract

Single-molecule fluorescence in situ hybridization (smFISH) provides direct access to the spatial relationship between nucleic acids and specific subcellular locations. The ability to precisely localize a messenger RNA can reveal key information about its regulation. Although smFISH is well established in cell culture or thin sections, the utility of smFISH is hindered in thick tissue sections due to the poor probe penetration of fixed tissue, the inaccessibility of target mRNAs for probe hybridization, high background fluorescence, spherical aberration along the optical axis, and the lack of methods for image segmentation of organelles. Studying mRNA localization in 50 μ m thick *Drosophila* larval muscle sections, these obstacles are overcome using sample-specific optimization of smFISH, particle identification based on maximum likelihood testing, and 3D

david.grunwald@umassmed.edu.

[⁺]Present Address: Casebia Therapeutics, Cambridge, MA 02139, USA

[⁺⁺]University of Oxford, OX1 3SR, Oxford, UK

[⁺⁺⁺]Moderna Therapeutics, Cambridge, MA 02141, USA

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Conflict of Interest

The authors declare no conflict of interest.

multiple-organelle segmentation. The latter allows independent thresholds to be assigned to different regions of interest within an image stack. This approach therefore facilitates accurate measurement of mRNA location in thick tissues.

Keywords

image processing; segmentation; smFISH

1. Introduction

Localization of mRNA is a central mechanism for regulating mRNA function.^[1] Single-molecule fluorescence in situ hybridization (smFISH) provides a snapshot of the distribution of individual mRNAs within cells. Initial smFISH protocols used long (50 nt) hybridization probes labeled with ≈ 5 dyes.^[2] Because it introduces fluctuations in signal strength, however, nonuniform labeling of long probes decreases the threshold of detection, especially for RNA clusters. More recently, the smFISH protocol was simplified to use shorter (20 nt) probes labeled with a single fluorophore.^[3] These probes are cheaper to make and easier to label. Moreover, multiplexing up to 48 probes against a single mRNA species yields high contrast between specific and nonspecific binding.^[3]

Typically, smFISH is used to detect individual mRNAs in cultured cells—e.g., yeast cells, monolayer cultures, or isolated neurons,^[4–7] yet many cultured cell types lack key features found in tissues. Within muscle tissue, for example, cells are large and multinucleate. The dimensions of muscle cells and nuclei make it desirable to work with thick ($>40\ \mu\text{m}$) tissue sections. However, available smFISH protocols for tissues, including *Drosophila* imaginal wing discs and *Caenorhabditis elegans* larvae and embryos,^[3,8,9] are limited to 5–10 μm sections. Performing smFISH in larger tissue volumes poses seven specific challenges: (1) fixative and probe penetration rates differ between the cytoplasm and nucleus and from sample to sample; (2) mRNAs are structured and coated with proteins, so hybridization sites can be masked;^[4] (3) intrinsic background fluorescence is aggravated by suboptimal removal of unbound probes from deep layers; (4) the refractive index can vary along the optical axis; (5) single-molecule signals must be identified in crowded environments and in varying image quality along the optical axis; (6) 3D image-reconstruction algorithms must work with anisotropic samples; and (7) nongeometric 3D objects (e.g., nuclei)—often incompletely labeled—must be reliably segmented over multiple 2D images in a discretely sampled image stack, even in otherwise very advanced analysis environments.^[10–12]

Here, we describe sample processing and imaging optimizations that have allowed us to perform smFISH in thick preparations of *Drosophila* larval body wall muscle. We optimized: (1) sample fixation and permeability conditions to improve probe penetration to deeper tissue sections and nuclei; (2) proteolytic processing to maximize mRNA accessibility to smFISH probes; and (3) time and temperature to improve probe signal and reduce background. We developed 3D-image analysis tools that allowed us to segment organelles and identify single-molecule signals in $\approx 50\ \mu\text{m}$ thick sections.

Using this optimized smFISH approach, we investigated the composition and nuclear export of megaRNPs—giant ribonucleoprotein particles associated with the nuclear budding pathway for mRNA export.^[13] We find that nuclear mRNA complexes that could be interpreted as megaRNPs contain up to 62 copies of single mRNA species. MegaRNPs were only detectable inside nuclei, but not in the cytoplasm.

Although smFISH provides a versatile and accessible method for wide ranging inquiries into mRNA localization at the tissue level, data analysis becomes a limiting factor, especially if distances to organelle surfaces need to be measured with low bias. Most bias results from oversimplified (i.e., symmetric) segmentation of irregularly shaped compartments (e.g., nuclei; Figure 1). Our new 3D Gaussian fitting of smFISH data (3DISH) algorithm closes this gap.

2. Results

To use smFISH to precisely localize individual mRNAs and megaRNPs within the cytoplasm and nuclei of *Drosophila* larval body wall muscles, we needed to optimize smFISH sample preparation for thick tissues, both to maximize signal and to minimize background throughout the tissue. Moreover, we needed to develop image analysis tools to automate spot detection, segment nuclei, and determine the positions of spots relative to one another and the nuclear membrane. We performed both in parallel, but for simplicity, we first describe our image analysis program—called 3DISH—followed by the smFISH optimization and our analysis of megaRNPs.

2.1. 3DISH Enables Automated Spot Detection and Organelle Segmentation

Available spot analysis tools (e.g., FISH-quant and ImageM) provide different levels of 3D localization and precision, but no pre-existing tool could perform 3D segmentation by defining the perimeter or border of the nucleus as an irregularly shaped surface in 3D space.^[3,8,14–17] We therefore developed a new 3D image analysis tool that enables fast, parallelized 3D Gaussian fitting of smFISH data (3DISH).

3DISH supports automated 3D segmentation of nuclei using a continuous 4D hull function to fill gaps caused by discontinuous immunofluorescence staining of nuclear lamin, a marker of the nuclear periphery (Figure 1 and Figure 2). Physically collapsed or improperly segmented nuclei can be excluded from further analysis. Segmentation based on nuclear shape allows us to apply different parameters for spot detection in nuclear and cytoplasmic compartments and to independently account for different background levels and spot sizes. 3DISH also allows us to freely define 3D regions of interest (3D ROIs) and to apply specialized image processing filters and spot detection criteria to each defined region (e.g., subtract local background, perform 3D spot detection, and verify spot location using 3D Gaussian model fitting and likelihood ratio testing). Gaussian fitting is slow, but graphics processing unit implementation on a high-end gaming computer reduced the processing time to less than 10 min for an image stack of >50 planes and 1000 × 1000 pixels per plane. Finally, 3DISH allows us to calculate distances between individual spots or between spots and edges of segmented objects. We can therefore perform colocalization analyses using absolute 3D distances.

2.2 Segmentation of Multiple Incompletely Labeled Organelles

Most image-analysis programs use box-shaped volume segmentation—i.e., a rectangular box drawn around the largest section of the nucleus and extending along the z -axis either throughout the whole image stack or to the top and bottom of the nucleus (e.g., Figure 1). However, box-shaped volume segmentation misrepresents the curved shape of the nucleus and results in inappropriate projection of signals into the nucleus, even for signals in z -planes far above or below the nucleus. To distinguish between nuclear and cytoplasmic megaRNPs, we needed to segment nuclei according to their exact envelope contours. We faced the following segmentation challenges: (1) Image stacks containing multiple nuclei; (2) individual nuclei extending over multiple z -frames; and (3) discontinuous nuclear lamin signal due to discrete z -sampling along the optical axis, nuclear invaginations, or incomplete immunofluorescence staining.

Segmentation (edge detection) by 3DISH involves the following steps:

- (1) Two uniformly filtered image stacks— I_1 ($5 \times 5 \times 1$ pixel kernel width) and I_2 ($10 \times 10 \times 1$ pixel kernel width)—are computed from the original image stack (I_0).
- (2) I_1 is subtracted from I_2 , and the gradient of $I_1 - I_2$ is computed, which results in a 4D matrix of 3D vectors representing each pixel of the image stack ($I_3 = I_1 - I_2$). The magnitude of each vector is calculated, resulting again in a 3D image stack (I_4).
- (3) In I_4 , all small magnitude values are limited to a minimum value $\max(I_4, \max(I_4)/N)$, where N is user-defined. The parameter N depends on the signal-to-background ratio of the organelles that are to be segmented. In our case, N was typically 300 (arbitrary units).
- (4) A histogram-dependent, binary threshold (T_h) is set on image I_4 using the ISODATA algorithm resulting in $I_5 = I_4 > T_h$.^[18] In the binary image I_5 , the value 1 represents volumes of the original image stack I_0 for which a significantly high gradient was found.
- (5) To reduce false edges, the voxel numbers of separate objects are counted, resulting in a list with objects and their corresponding volumes. The mean volume of all objects in the image stack is calculated and all objects smaller than the mean volume size plus $1 \times$ the standard deviation of the mean are rejected.
- (6) The remaining edges can be discontinuous, caused by inhomogeneous staining of the nuclei. To connect the boundaries of individual nuclei, a dilation followed by an erosion is performed using a stencil with a width chosen by the user, based on the visual impression of nuclear sizes. The stencil size must be big enough to connect discontinuous edges but small enough not to connect adjacent nuclei. Often a stencil size of 10 is sufficient.
- (7) To identify the number of nuclei present in the image stack, the number of continuous volumes are identified and each volume is labeled with a unique identifier.^[19]

- (8) To segment the nuclei based on the size of the identified, possibly discontinuous edges, a 3D convex hull is calculated by finding the smallest volume that includes all edges belonging to the same nucleus (prior assignment in step 6).^[20] Gaps are closed by linear connection of the end points, making it necessary that a sufficient number of possibly isolated edges are found during edge detection.
- (9) The resulting segmentation is used as an initialization point for 3D active contour optimization.^[21] In this optimization, the fitting energy of the contour is minimal when the curve is on the boundary of the object.
- (10) If the nuclear stain is found to be discontinuous, step 7 is repeated.

For an example of segmented nuclei, see Figure 2.

2.3. Spot Detection Using Local Background Subtraction

The major challenge for spot detection in thick tissues is accurate estimation of z-position. We needed to obtain an accurate z-position estimate to correctly assign RNAs as inside or outside of the nucleus and to measure the distance between RNAs and the nuclear envelope or RNAs and transcription sites. Inter-RNA distances are ultimately used to define colocalization criteria. In our experience, the most accurate methods for spot localization use a maximum likelihood-based (i.e., Gaussian) fitting approach.^[22,23] For 3D images, a center of mass search can be combined with 2D Gaussian fitting in the plane of the brightest voxel.^[24] In thick samples, however, pinpointing the brightest voxel is limited by local background. Thick tissue samples are highly scattering, so the local background signal varies as a function of position along the optical axis. Inherent inhomogeneity of the biological sample—i.e., structures like nuclei or fibers—also contributes to local background. Fitting a 3D Gaussian is therefore more precise, as the z-position of a spot is not restricted by the discrete z-positions of image planes; 3D Gaussian fitting uses the information in pixels across multiple image planes to find the best fit.

To identify smFISH signals in 3D, we applied an elliptic 3D maximum filter with parameters: $2(\sigma_{\text{PSF}_{xy}}, \sigma_{\text{PSF}_{xy}}, \sigma_{\text{PSF}_z})$ to the background-corrected image stack, (σ_{PSF_k}) being the theoretical width of the point spread function (PSF) for the k-th dimension. This filter is moved over all voxels in the 3D stack and creates a local maximum map. To find candidate positions for our Gaussian fitting routine, we compared the 3D local maximum map to a 2D tophat background-corrected image stack. If a pixel had the same intensity value in the 3D maximum map and the background-filtered 2D tophat image, the pixel was accepted as a candidate position. Dim candidate spots were rejected using an intensity threshold equal to the mean intensity plus N times the standard deviation of the background intensity in gray values. N was manually set to account for image quality—e.g., $N = 1$ to 3 typically rejected nonspecific smFISH probe signals. A different threshold can be set for each nucleus and cytoplasmic ROI, with the largest possible ROI being the whole image stack.

We centered each candidate position within a discrete candidate volume (an isolated cube) proportional in size to the theoretical expected PSF ($3(2(\sigma_{\text{PSF}_k}) + 1)$) in pixels. For each cube, we extracted the deconvolved and aligned data and fit a 3D Gaussian, yielding the

following parameters: x, y, z position, signal localization precision ($\sigma_{\text{Loc}_x}, \sigma_{\text{Loc}_y}, \sigma_{\text{Loc}_z}$), intensity (I), background (bg), signal height (σ_{PSF_z}), and width ($\sigma_{\text{PSF}_{xy}}$), the false alarm probability (P_{FA}), and the signal-to-background ratio ($I^2/\sqrt{I^2 + bg^2}$).

These parameters can be used to manually define upper and lower limits and control the number of false-positive detections. Because background is not uniform, parameter limits can be independently defined in the nucleus and in the cytoplasm for each channel. Figure 2 shows a raw image (Figure 2A), the image after nuclei segmentation and mRNA detection (Figure 2B), and the images superimposed (Figure 2C).

2.4. Single-Molecule FISH Optimization Improves Signal Detection

To maximize the functionality of our 3DISH image analysis tools for smFISH in thick tissue samples, we also needed to optimize sample preparation and probe hybridization. Using published protocols as a starting point,^[4,8,25] we systematically optimized sample preparation parameters (Table 1). To assess the performance of each parameter change, we examined the quality of antilamin immunofluorescence signal and nuclear morphology, and we quantified the ratio of double-labeled smFISH spots (i.e., using two sets of probes, each labeled with a different color) and all detected spots (i.e., both double- and single-color). Optimizing for dual-color labeling of the same target while maintaining immunofluorescence signal and nuclear morphology resulted in limited visual image quality but provided maximal information content.

Duration of fixation affected the number and size distribution of faint smFISH spots: longer fixation times resulted in more homogeneous and robust signals within and between samples. Permeabilization was essential for probe and antibody penetration throughout the tissue, but excessive permeabilization reduced the number of smFISH signals, increased background fluorescence, or damaged the sample (Table 1 and Figure 3). Protease treatment improved the accessibility of mRNA to smFISH probes in all but the harshest condition.^[4] Compared to standard protocols, our optimized sample preparation for *Drosophila* muscle tissue significantly increased the number of mRNAs that could be successfully visualized by smFISH (Figure 3). Improvements were most pronounced in the nucleus, but were also observed in the cytoplasm. We conclude that the protease treatment and permeabilization not only resulted in better signals but also in a better labeling of the mRNA population as a whole (compare Figure 3A,B to D,E). Our findings emphasize that standard smFISH protocols may not be optimal for every tissue, so we recommend careful optimization of sample preparation conditions.

Key to our optimization was the use of probes in two color spaces to label the *pak* mRNA. We designed 96 probes tiled along *pak* mRNA: 48 probes labeled in green (odd—visualized in yellow) alternating with 48 labeled in red (even—visualized in magenta). We did this in part because initial counting experiments yielded inconclusive or inconsistent results if mRNAs were labeled with probes in a single-color space (e.g., 48 probes labeled in green tiled along *pak*, *par6*, or *cask*). We reasoned that probing a single mRNA species in two color spaces would increase confidence in the data. As a test, we imaged multicolor beads and achieved 90% colocalization using a spot-to-spot distance of 3 voxels (a voxel

measures $107 \times 107 \times 200$ nm) as the colocalization cutoff, as this is sufficiently large to account for chromatic aberrations. Prior to protocol optimization, only 36% of yellow and magenta *pak* smFISH spots colocalized in nuclei ($N = 3$, range 25–49%) and 48% colocalized in cytoplasm ($N = 3$, range 35 to 58%). Our optimized procedure increased colocalization frequencies in both compartments to 59% in nuclei ($N = 4$, range 40 to 74%) and 72% in cytoplasm ($N = 4$, range 63 to 81%). A strong smFISH signal in one channel was sometimes accompanied by a weak signal in the other channel (Figure 4). Therefore, this weak signal would not be counted as a positive *pak* mRNA spot. smFISH signals detected by 3DISH were RNA-specific—ribonuclease treatment strongly reduced smFISH signals and virtually eliminated colocalization of yellow and magenta *pak* probes. Thus, despite extensive optimization of labeling conditions (see Table 1 and Table 2), probe binding affinity and probe accessibility to mRNA limits the smFISH labeling efficiency as seen by the low fraction ($\approx 80\%$) of mRNAs stained with two colors. Nevertheless, using probes in two color spaces to label an mRNA does increase confidence in detected spots.

2.5. Analysis of megaRNPs

MegaRNPs containing *pak*, *par6*, and *cask* mRNAs were previously observed at lamin C foci, in the lumen of the nuclear envelope, and at neuromuscular junctions (NMJs) using digoxigenin-11-dUTP probes and antidigoxigenin and fluorophorelabeled secondary antibody.^[13] We sought to use smFISH to directly and simultaneously visualize *pak*, *par6*, and *cask* mRNAs in *Drosophila* muscle tissue. Because we never obtained 100% colocalization of yellow and red *pak* smFISH signals, however, spot detection of probes in one-color space could not be used to definitively accept or reject spots as an individual mRNA. But because each megaRNP likely contains many mRNA molecules,^[13] we expected that our assay would robustly report on the number and distribution of megaRNPs and whether an individual megaRNP contains single or multiple mRNA species.

MegaRNPs were proposed to export from the nucleus by budding through the nuclear envelope and to deliver mRNAs to the NMJ for localized translation. In tissues stained with *pak* probes, we analyzed the intensity and distribution of megaRNPs in nuclei and cytoplasm, accepting only double-labeled spots as true signal. Based on our data (Figure 5), we defined candidate megaRNPs as spots with a minimum intensity seven times the average intensity of an isolated mRNA. Of 22 nuclei examined, 10 contained two or fewer smFISH signals that met this minimum intensity requirement. These nuclei were excluded from further analysis because transcription sites might also harbor multiple mRNA molecules. The remaining 12 nuclei contained up to 37 candidate megaRNPs. Based on the brightness of isolated *pak* mRNAs in the cytoplasm, we estimated that the brightest nuclear megaRNP in our data set contained ≈ 50 copies of *pak* mRNA (Figure 5). This intensity analysis probably underestimates the number of *pak* mRNA copies, because based on our 60% probability of *pak* mRNA detection, labeling of mRNAs (especially nuclear) was likely incomplete. In further support of likely underestimating the maximum number of mRNA in a megaRNP is that we often had to reject data due to saturation (signal intensity exceeded the dynamic range of our camera), especially signals originating from nuclei.

Plotting normalized signal intensity (i.e., number of mRNAs per spot) against distance from the nuclear or cytoplasmic side of the nuclear periphery (Figure 5), we observed a 2-voxel gap (≈ 215 nm) in the distribution of spots, coinciding with the nuclear envelope (Figure 5). Most cytoplasmic signals were between 0.5 and 1.5 normalized intensity, consistent with the detection of single mRNAs. Most of the brightest spots in the cytoplasm (1.6 to 6 normalized intensity) were within $2\text{ }\mu\text{m}$ of the nuclear periphery (Figure 5). The MegaRNPs were only identified inside the nucleus and were on average 11-fold more intense than isolated mRNA spots (range 7- to 54-fold brighter; median 9-fold brighter). The distribution of the nuclear megaRNPs peaked at $\approx 4\text{ }\mu\text{m}$ from the nuclear lamin signal, toward the center of the nucleus. Due to the limited number of the megaRNPs, however, the distribution was not statistically significant. Therefore, in agreement with the original work,^[13] we found no strong correlation between the number of nuclear megaRNPs and distance to the nuclear envelope.

To test whether individual megaRNPs contain one or multiple mRNA species, we performed smFISH experiments with differently colored probes for *pak*, *par6*, and *cask* mRNA, all of which are thought to form the megaRNPs, and *GluRII* mRNA, which was previously shown not to colocalize with the megaRNPs.^[13] For dual-color detection of two different mRNA species, we assume detection probabilities for each probe set similar to the *pak* mRNA probes ($\approx 60\%$ in nuclei and $\approx 70\%$ in cytoplasm, noted above) and treat the observed binding of the single-color probes as an increase in the false-positive detection rate for the megaRNPs. The presence of multiple mRNA species in a megaRNP will increase the probability of detecting the megaRNPs by reducing the miss rate ($\approx 40\%$) of individual mRNAs. For example, a hypothetical megaRNP containing three copies of a single mRNA species would be missed $\approx 6\%$ of the time, and 0.001% for seven copies—the miss rate for the megaRNP is given by the multiplication of the miss rates of the individual mRNAs. Thus, we expect the miss rate to be very low if the megaRNPs contain multiple copies of two mRNA species. Nevertheless, we did not observe colocalization of *pak*, *par6*, or *cask* mRNAs in the cytoplasm any more frequently than we observed colocalization of *pak* mRNA with *GluRII* mRNA. Furthermore, colocalized signals of all possible pairs of those mRNAs were always of fluorescence intensity levels consistent with one or two mRNA molecules, not larger numbers (Figure 6). Of 23 nuclei in which we detected both *pak* and *cask* signals, 9 nuclei had one observable *pak*–*cask* colocalization event and 2 nuclei had more than one. These isolated colocalization events could reflect proximity of the *pak* and *cask* transcription, which are located on the same chromosome.

3. Discussion

In this study, we used single-molecule FISH to investigate three mRNAs, *par6*, *pak*, and *cask*, that were previously identified as components of the megaRNPs in *Drosophila* larval body wall muscle.^[13] We aimed to better understand the composition and cellular location of the megaRNPs. Single-molecule FISH in thick tissue samples, however, presents numerous technical challenges. To overcome these, we optimized existing protocols and developed improved image analysis tools specific for 3D environments.

3.1. Probe Specificity and Colocalization Calibration

Motivated by: (1) initial probe colocalization rates far below our prediction (see Figure 3B); (2) large numbers of presumed smFISH signals that yielded inconclusive results; and (3) the wish to employ probe mixes rather than synthesize and optimize individual probes, we focused on finding a way to test and predict binding specificities of probe sets in situ. We worked with a mix of 20 nt probes, as called for in recent smFISH protocols.^[3] These probes are cost-efficient compared to using 50 nt probes, allow lower hybridization temperatures (presumably preferable for sample integrity), and are overall easier to use than longer probes that carry multiple dyes. In all, we designed a total of 96 probes against a single target mRNA (*Pak*). We numbered the probes sequentially in order of their location along the mRNA, and then divided them into two sets based on even and odd numbers. Odd probes were labeled with Quasar 570 and even probes with Quasar 670, yielding an alternating pattern of red and green probes along the target mRNA. The probes had similar GC content, and both sets were predicted to be of similar specificity to the target mRNA. We therefore expected that the colocalization frequency would only be limited by the optical performance of the microscope (>90% colocalization for multicolor beads), and not by the labeling protocol or probe performance.

A number of potential factors likely contributed to our initial lack of success using fixation, permeabilization, and protease digestion conditions previously developed for monolayer cells. For example, limited probe accessibility can lead to a scenario where the minimal number of probes needed for detection (estimated at ≈ 20 ;^[3]) is reached after protease treatment for one probe set but not the other. In our case, however, colocalization rates did not increase upon changes to permeabilization conditions, protease digestion, or hybridization times. If the number of colocalization events does not increase despite more protein being digested, probe accessibility is unlikely the reason the experiment failed to match the predicted colocalization frequency. Single color, noncolocalized spots could be caused by properties of the dyes. To rule this out, we treated our tissue samples with RNase and found that all signals, both double and single-color spots, disappeared (Figure 4), indicating that they were RNA-dependent. Besides higher order folding structures of the target mRNA, this left open the possibility that substantial performance differences exist within our probe mix and that a subset of probes exhibited high-level off-target binding. We are unaware of any published data on the off-target probability of probes as a function of nucleotide length. It is worth noting, however, that smFISH was originally developed using 50 nt probes to increase probe specificity.^[2] A cocktail of 96 probes makes it difficult to test the efficiency of individual probes in the mix—either one by one or by using established methods, e.g., “sense” probes.

Despite these difficulties, we ultimately achieved sufficient probe performance to address several key questions regarding the formation and nature of megaRNPs. The large number of mRNA molecules within the megaRNPs minimizes detection sensitivity limits and averages the variations in probe accessibility to different mRNA regions. For example, if our colocalization-positive detection rate is 60% for a single mRNA, the overall colocalization detection rate increases to 99% if 5 mRNAs are present in a megaRNP (i.e., miss rate = $0.4^5 = 0.01$). Though the detection probability of the megaRNPs is not limiting in our

experiments, the variance in labeling efficiency introduces a bias for the intensity analysis of megaRNPs, lowering the estimated number of mRNAs detected per megaRNP. Probe specificity for colocalization experiments could be increased if each probe sequence was tested individually and a limited set of probes selected based on specificity and background performance. Such an effort might be useful if smaller mRNA complexes are targeted, but based on our “odd–even” colocalization calibration, we predict that the use of multicolor probes against the same target is a more viable strategy for assay optimization and to test overall assay performance.

3.2. On the Localization and Composition of megaRNPs

Nuclear budding of large mRNP-granules, or megaRNPs was described as an alternative mRNA export pathway from the nucleus.^[13] Based on the mRNA binding factors found in such granules, it was hypothesized that a specific function of such megaRNPs is to promote the colocalized expression of functionally related proteins in specific locations in the cytoplasm (e.g., at a synapse). If this model is correct, then the megaRNPs should be observable in both the nuclear and cytoplasmic compartments.

First, we investigated the hypothesis that the individual megaRNPs contain multiple mRNA species by directly visualizing three different mRNA species (*par6*, *pak*, and *cask*) previously identified as megaRNP components.^[13] Based on our analysis of these three mRNA species, our results favor a model in which megaRNPs are predominantly composed of a single mRNA species. This is consistent with studies suggesting that RNA granules in dendritic processes contain only one mRNA,^[26–28] although reports of multiple mRNA species in mRNA granules do exist.^[29,30] The average number of mRNA molecules we find per megaRNP is ≈ 10 , but given our data and the physical size of megaRNPs (≈ 200 nm),^[13] this estimate is, as discussed, a lower limit. Second, we investigated the composition of megaRNPs using smFISH, but we were only able to observe megaRNPs within the nucleus. We therefore asked, third, what the localization of megaRNPs in the differentiated cell would be if probed by smFISH. We based this analysis on smFISH using dual-color probes against *Pak* mRNA. Dual-color labeling of the same mRNA offers a higher degree of confidence in identifying “true” mRNA signal, if only dual-color spots are admitted for analysis. In our data, we find a clear two-voxel-wide gap between megaRNPs located in the nucleus and the cytoplasm, as one would expect for export via nuclear pore complexes in the steady-state cells. With an antilamin IF stain as label for the nuclear periphery, we do not have direct knowledge of the position of the nuclear envelope, which in cells showing egress can be heavily displaced with the inner and outer membrane of the nuclear envelope being separated. While such separation should result in a wider range of distances, without a clear local minimum, the budding of megaRNPs could be a sporadic process possibly contributing to our results.

The use of single-molecule FISH to study megaRNPs has several advantages, as discussed in the introduction. We limited our analysis to data sets with signal intensities within the dynamic range of our camera but had to reject other data sets with higher signal intensity. We did so to be able to carefully control the technical quality of our work. For this very reason, we also optimized the labeling protocol using proteinase K and pepsin at various

concentrations suitable for the study of RNA–protein complexes in neuronal cells. Our final protocol uses only $0.01 \mu\text{g mL}^{-1}$ protease, and our visual impression is that the frequency of bright megaRNPs is similar whether the smFISH samples are prepared in the presence or absence of proteases (e.g., Figure 3A). And though our controls suggest that our method is robust, the mild protease conditions might be sufficient to disrupt weak interactions between mRNAs or between proteins and mRNAs within megaRNPs. So, we cannot rule out that smFISH is a less than ideal technology to study megaRNPs. The image processing challenges in our work, however, are inherent to any 3D sample. The workflow and solutions we present here on the imaging aspects of detecting megaRNPs can easily be applied to other 3D experiments and introduce exact and defined measures for 3D distances and colocalization.

3.3. Image Processing and Analysis

Image analysis of our 3D stacks, even after optimization of the smFISH protocol, provided challenges based on the thickness of the tissue for a number of reasons: (1) The refractive index of the sample changes with tissue depth. (2) As a result, signals from deep-tissue planes are weaker than from planes closer to the objective. The same effect was exacerbated by (3) the limited penetration of probes and reduced levels of protease digestion in deeper, more central tissue sections, (4) sample-specific background from filamentous structures within the tissue, and (5) higher levels of background than found in single layer cultured cells. Another challenge was the need to minimize bias in locating smFISH signals inside or outside of the nucleus and measuring the distance between a smFISH signal and the nuclear periphery. The origin of this challenge was the irregular shapes of nuclei and consequent bias from “box” approximations of their volume (Figure 1) on the one hand, and the discontinuous labeling of the nuclear periphery by the lamin IF stain (Figure 2) on the other. (7) The signal intensities of single mRNAs detected using a Delta Vision wide-field microscope with deconvolution were too weak to be reliably detected using confocal microscopes available to us.

Finally, robust detection of smFISH signals and estimating their 3D position turned out to be more complicated than we predicted based on the quantity of photons expected from the smFISH probes. This discrepancy results from variability in the number of probes bound per mRNA due to accessibility and specificity, and from optical aberrations and high background levels from the tissue.

Machine learning, with its caveats of training dependability, has been used very successfully for spot identification.^[10–12] Our approach differs in as much as it uses local background analysis and noise models for candidate spot identification combined with a likelihood ratio test for spot verification. Another difference is that our approach uses a full 3D Gaussian fitting model, which means that z-positions are assigned in between image slices in a stack, rather than being projected into the closest image in the stack. The ability of machine learning approaches to vary the parameter set for detection has proven very helpful with analysis of thicker tissues of $\approx 10 \mu\text{m}$. Here, we suggest methodology developed for optical reading of information from storage devices with changing refractive indexes, e.g.,

multilayer DVDs. This method allowed us to have comparable spot detection probability throughout a 50 μm thick muscle section, based on our likelihood ratio testing feature.

Here we described a method to segment nuclei by wrapping them in a hull that closely follows the shape of the lamin IF stain, to close gaps in this stain and in doing so largely eliminate any bias in compartment assignment. We found that this segmentation method also works well on other compartments; for instance, we have used it to identify neuronal junctions based on IF stains (data not shown). We combined this advanced segmentation with refractive index correction as used with optical disk drives,^[24] true 3D Gaussian fitting of particle locations,^[23,31] and a maximum-likelihood test to verify signals.^[32,33] To reduce image processing time, we implemented graphic card processing for 3D localization and likelihood testing. Despite these advances, differences between compartments (e.g., different nuclei, nucleoplasm, cytoplasm, or other ROIs) were too extreme for fully automated analysis. We therefore enabled manual filtering options that could be individually applied to specific compartments.

4. Conclusion

The ability to measure 3D distances between smFISH signals as well as distances between smFISH signals and any organelle surface is crucial to interrogate mRNA function with minimal experimental error. Our analysis tool provides new and useful features currently not available to most biological research groups. Going forward, however, it might be fruitful to explore how machine learning approaches might complement datadriven model-based solutions, like the one presented here.

5. Experimental Section

Drosophila Stocks and Antibodies:

Wild-type flies (CantonS; CS) were raised at 25 °C. Immunofluorescence staining was done using primary mouse antilamin C antibody (LC28.26, 1:20–1:30, DSHB) and secondary anti-mouse antibody from Invitrogen (A31553, 1:200).

Drosophila Larva Dissection:

Third instar larvae were obtained and rinsed with ice-cold phosphate buffer saline (PBS). Larvae were dissected in ice-cold PBS, then immediately immersed in ice-cold fixation solution (see Table 1). Larvae were incubated at room temperature with gentle agitation for different lengths of time (see Table 1). After fixation, larvae were rinsed three times with PBS, then kept in either 70% (vol/vol) ethanol at 4 °C or PBS on ice until use. For RNase A treatment, fixed larval filet was incubated with 100 $\mu\text{g mL}^{-1}$ RNase A in 2 $\times$ SSC for 1 h at 37 °C, then stored in either 70% (vol/vol) ethanol at 4 °C or PBS on ice until use.

Oligonucleotides:

Single-molecule fluorescence in situ hybridization was done using Stellaris probes purchased from Biosearch Technologies (Novato, CA, USA). All sequences are listed in

Table S1 (Supporting Information). Probe sequences were designed using the Biosearch Technologies online probe designer.

Single-Molecule Fluorescence In Situ Hybridization-Immunofluorescence Staining:

Probe pools against specific mRNAs contained up to forty eight 20 nt probes (Table S1, Supporting Information). In situ hybridization was performed either by following manufacturer's protocol or with modifications specified in Table 1. Pepsin (Sigma-Aldrich) treatment was performed in 10×10^{-3} M HCl, and proteinase K (Invitrogen) treatment was performed in $2 \times$ SSC. After proteinase treatment, proteinase K solution was removed, and samples were quenched in 2 mg mL^{-1} glycine in PBS. Larval filets were then washed three times with 2 mg mL^{-1} glycine in PBS, and postfixed as noted in Table 1. Larval filets were washed twice with in PBS containing 0.1% (v/v) Triton-X100 (0.1% PBT), and once with wash buffer for 5 min or as shown in Table 1. Hybridization was performed for 4 h at 37°C in hybridization buffer containing 0.25×10^{-6} M probe pool; 10% (wt/vol) dextran sulfate; $2 \times$ SSC; 10% (vol/vol) deionized formamide; 1 mg mL^{-1} yeast tRNA (Invitrogen), 2×10^{-3} M ribonucleoside vanadyl complex; with or without primary antibody. After hybridization, wash buffer was added to the hybridization solution. Then, larval filets were washed three times with wash buffer. For smFISH-IF, samples were incubated with secondary antibody in hybridization buffer for 1 h at 37°C . Both hybridization and secondary antibody incubation were performed in at least $50 \mu\text{L}$ of reaction solution per 1 larval filet in microtubes. Afterward, the larval filets were washed three times with wash buffer and three times with PBS, and then rinsed with water. Larval filets were mounted on glass slides with Prolong Gold (Invitrogen).

Image Acquisition:

All images were taken with a DeltaVision Elite deconvolution microscope system (GE Healthcare Life Sciences), equipped with $60\times/\text{NA } 1.40$ oil immersion objective lens (PlanApo 60XO; Olympus) and a filter sets to image fluorophores in 4',6-diamidino-2-phenylindole (DAPI) (central wavelength/bandwidth: 390/18 excitation; 435/48 emission), tetramethylrhodamine (TRITC) (542/27; 594/45), and Cy5 (632/22; 676/34) channels, and a CoolSNAP HQ camera (Photometrics). The pixel size of the camera is $6.45 \mu\text{m}$ resulting in a pixel size in image space of close to 107 nm . The z-stack images were spaced 200 nm apart along the optical axis. Image size was cut to 1024×1024 pixels. Immersion oil (Olympus immersion oil AX9602) with a refractive index of 1.516 at room temperature was used. The embedding medium (Prolong Gold, Invitrogen) has a refractive index of 1.46.

Deconvolution:

A momentum-preserving deconvolution was applied and corrected for chromatic shifts (Parameter: area radius of background estimation 700 nm , CLME algorithm, stop criteria image differences $<10^{-5}$ or 50 iterations maximum in Huygens, SVI Delft, The Netherlands). This deconvolution only relocates the photons and does not affect the actual photon counts. After deconvolution, the remaining local background signal was removed using a round (stencil) 2D tophat filter with a width of eight pixels on each 2D image within the image stack, which was followed by segmentation of nuclei and spot detection.

In samples as thick as larval *Drosophila* muscle tissue, variations in refractive index mismatch along the optical axis will contribute to a significant spherical optical aberration and wavelength-dependent defocus. The spherical aberration adds an additional source of image distortion and the z-shift creates a misalignment between the different channels. This situation is common when using high NA objectives, but due to the inherent 3D nature of our image acquisition, the spherical aberrations introduced on top of the expected defocus need to be taken into account in image processing.^[24] The focal shift is caused by the wavelength dependency of the defocus function,^[31] which can be described as follows

$$\begin{aligned} D_i &= \Delta z(k_i - k_{z_i}) \\ &= n \Delta z k_0 \left(1 - \sqrt{1 - \frac{k_x^2 + k_y^2}{k_i^2}} \right) \\ &= n_i \Delta z k_0 (1 - \cos\theta_i) \end{aligned} \quad (1)$$

where n_i is the refractive index of medium i, k_0 is the length of the wave vector in vacuum, k_{z_i} is the wave vector length of the z component in medium i, z is the thickness of the medium. The differences in sign and constant phase offset to the original publication can be ignored. In addition to the original approach, we had to compensate for the spherical aberration introduced by refractive index mismatch. Furthermore, the spherical aberrations, also inherently dependent on wavelength, is caused by index of refraction mismatch,^[31,34] and is given by

$$\begin{aligned} SA_{1,2} &= \Delta z(k_1 - k_{z_1}) - \Delta z(k_2 - k_{z_2}) \\ &= n_1 \Delta z k_0 (1 - \cos\theta_1) - n_2 \Delta z k_0 (1 - \cos\theta_2) \end{aligned} \quad (2)$$

where z is the thickness of medium 2, θ_1 is the angle in medium 1, and θ_2 is the angle in medium 2. Note that this expression not only contains spherical aberration (of all orders) but also defocus.

First, the spherical aberration is estimated and deconvolution (Huygens, SVI, The Netherlands) is performed with a z-dependent momentum preserving deconvolution to compensate for the depthdependent PSF distortion. Taking the spherical aberration into account for the deconvolution significantly reduces artifacts and increases image sharpness. Second, the different 3D color channels are aligned to adjust for the wavelength-dependent defocus, which was calibrated using TetraSpeck 0.2 μm beads (Invitrogen). Calibration was done by acquiring bead image stacks in all color-channels, deconvolving them, and estimating the chromatic shift using the corresponding function in Huygens.

Fitting of Candidate Spots:

The likelihood of a signal measurement is assumed to be a Poisson process, and is therefore given by

$$L(d|\vec{\theta}) = \prod_{i=1}^N \prod_{j=1}^M \frac{\mu_{i,j}(\vec{\theta})^{d_{i,j}} \exp[-\mu_{i,j}(\vec{\theta})]}{d_{i,j}!} \quad (3)$$

where $\mu_{i,j}(\vec{\theta})$ is the expected photon count in pixel i, j ; and $\vec{\theta} = \theta_I, \theta_x, \theta_{bg}$ are the parameters to be estimated with I as the intensity and $\vec{x}$ position (x, y, z); bg is the background, and $d_{i,j}$ is the measured photon count in pixel i, j .

The likelihood function is maximized by finding the zero of the negative log likelihood function by use of Newton–Raphson method. The derivative of the negative log likelihood is given by

$$\frac{\ln(L(d|\vec{\theta}))}{\partial \vec{\theta}} = \sum_{i=1}^N \sum_{j=1}^M \left(d_{i,j} \mu_{i,j}(\vec{\theta})^{-1} - 1 \right) \frac{\partial \mu_{i,j}(\vec{\theta})}{\partial \vec{\theta}} \quad (4)$$

The precision with which we can determine the position of the single molecule depends significantly on the intensity of the single molecule, the background model, aberrations, and the axial position; it is therefore essential to also estimate these parameters. This round of fitting is computationally expensive because a nonlinear optimization problem needs to be solved; the fitting was therefore done on a graphical processing unit similarly to that described by Smith et al.,^[23] dramatically reducing the computation time. The expected photon count depends on the choice of PSF. Here, the PSF is approximated by a Gaussian distribution

$$\text{PSF}(x, y, z) = \frac{1}{\sqrt{8\pi^3 \sigma_{xy}^2 \sigma_z}} e^{-\frac{1}{2\sigma_{xy}^2}[(x-x_0)^2 + (y-y_0)^2] - \frac{(z-z_0)^2}{2\sigma_z^2}} \quad (5)$$

This PSF must be integrated over the pixel area to arrive at the expected photon count at each pixel i, j

$$\mu_{i,j} \vec{\theta} = I \Delta E_{xy}(x_i - x_0) \Delta E_{xy}(y_i - y_0) \Delta E_z(z_j - z_0) + bg \quad (6)$$

with

$$\Delta E_k(u) = \frac{1}{2} \left[\text{erf} \left(\frac{u + \frac{1}{2}}{\sqrt{2}\sigma_k} \right) - \text{erf} \left(\frac{u - \frac{1}{2}}{\sqrt{2}\sigma_k} \right) \right] \quad (7)$$

where (x_i, y_i, z_j) are the pixel coordinates in unit ([pixel]) of pixel i, j ; (x_0, y_0, z_0) is the location of the center of the PSF in unit [pixel]; and σ_k is the PSF width in the k -th dimension, depending on the numerical aperture (NA), magnification (M), pixel size (p), and the wavelength of the light (λ).

The estimated parameters are total intensity (I), background intensity (bg), the PSF width (σ_{xy} and σ_z), signal-to-noise ratio ($1/\sqrt{I^2 + bg^2}$), the estimated Cramer–Rao lower bound ($\Sigma(\vec{\theta})$), and the false alarm probability P_{FA} .^[33]

Distance Calculation:

Spot-to-spot distances were measured using the equivalent function of the MatLab tool box DIPImage (TU Delft, The Netherlands). Initially, the distance between a spot and a compartment edge was calculated by multiplying the compartment mask with a radial image (i.e., an image with the value of the R-coordinate of each pixel as the pixel values), indexing the pixel in which an mRNA was detected. The most efficient implementation was accomplished by use of the k-nearest search algorithm presented by Friedman et al.,^[35] which finds the nearest periphery coordinate using a list of mRNA locations. This method provides us with the distance to the periphery with a maximum error of 1 diameter of a voxel.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

A.N. and C.S.S. contributed equally to this work. A.N., C.S.S., M.J.M. and D.G. conceptualized the work. A.N. and C.S.S. performed formal analysis, validation, investigation, and developed methodology with support from M.H., M.J.M., and D.G. A.N. performed all biological experiments and protocol optimization. M.J.M. and D.G. administered and acquired funding for the project. C.S.S. developed software and algorithms with support from A.N. A.N., C.S.S., M.H., R.M.M., and D.G. created figures. A.N., C.S.S., and D.G. wrote the original draft. All authors reviewed and edited the manuscript. The authors used casrai.org contributor classification. The authors thank Adina Buxbaum and Robert H. Singer for sharing and helping with smFISH protocols, Vivian Budnik for discussions, reagents and critical comments on the manuscript, Vahbiz Jokhi for help with fly maintenance, and Darryl Conte for editing and critical reading of the manuscript. The authors thank all RTI staff for their invaluable support. This work was funded by a Worcester Foundation grant and NIH grants R01 AI120902 and U01 EB021238 to D.G.; C.S.S. received support from Merton College; and M.J.M. received support from HHMI. No ethical approval was required for the work on *Drosophila* larvae.

Published online:

- [1]. Wu B, Eliscovich C, Yoon YJ, Singer RH, Science 2016, 352, 1430. [PubMed: 27313041]
- [2]. Femino AM, Fay FS, Fogarty K, Singer RH, Science 1998, 280, 585. [PubMed: 9554849]
- [3]. Raj A, Bogaard PVD, Rifkin S, Oudenaarden A, Tyagi S, Nat. Methods 2008, 5, 877. [PubMed: 18806792]
- [4]. Buxbaum AR, Wu B, Singer RH, Science 2014, 343, 419. [PubMed: 24458642]
- [5]. Kwon S, BMB Rep. 2013, 46, 65. [PubMed: 23433107]
- [6]. Shaffer SM, Wu M-T, Levesque MJ, Raj A, PLoS One 2013, 8, e75120. [PubMed: 24066168]
- [7]. Coassin SR, Orjalo AVJ, Semaan SJ, Johansson HE, Methods Mol. Biol. 2014, 1211, 189. [PubMed: 25218386]
- [8]. Lyubimova A, Itzkovitz S, Junker JP, Fan ZP, Wu X, van Oudenaarden A, Nat. Protoc. 2013, 8, 1743. [PubMed: 23949380]
- [9]. Oka Y, Sato TN, Sci. Rep. 2015, 5, 8571. [PubMed: 25711926]
- [10]. Titlow JS, Yang L, Parton RM, Palanca A, Davis I, Methods Mol. Biol. 2018, 1649, E1. [PubMed: 29450865]
- [11]. Titlow JS, Yang L, Parton RM, Palanca A, Davis I, Methods Mol. Biol. 2018, 1649, 163. [PubMed: 29130196]
- [12]. Herbert AD, Carr AM, Hoffmann E, PLoS One 2014, 9, e114749. [PubMed: 25478967]
- [13]. Speese SD, Ashley J, Jokhi V, Nunnari J, Barria R, Li Y, Ataman B, Koon A, Chang YT, Li Q, Moore MJ, Budnik V, Cell 2012, 149, 832. [PubMed: 22579286]

- [14]. Semrau S, Crosetto N, Bienko M, Boni M, Bernasconi P, Chiarle R, van Oudenaarden A, Cell Rep. 2013, 6, 1.
- [15]. Mueller F, Senecal A, Tantale K, Marie-Nelly H, Ly N, Collin O, Basyuk E, Bertrand E, Darzacq X, Zimmer C, Nat. Methods 2013, 10, 277. [PubMed: 23538861]
- [16]. Zhang C, Sun C, Vallotton P, Fenech M, Pham TD, Microsc J 2013, 252, 122.
- [17]. Zenklusen D, Larson DR, Singer RH, Nat. Struct. Mol. Biol. 2008, 15, 1263. [PubMed: 19011635]
- [18]. Ridler TW, Calvard S, IEEE Trans. Syst. Man Cybern. 1978, 8, 630.
- [19]. Samet H, Tamminen M, IEEE Trans. Pattern Anal. Mach. Intell. 1988, 10, 579.
- [20]. Barber CB, Dobkin DP, Huhdanpaa H, ACM Trans. Math. Softw. 1995, 22, 469.
- [21]. Chan TF, Vese LA, IEEE Trans. Image Process. 2001, 10, 266. [PubMed: 18249617]
- [22]. Deschout H, Cella Zanacchi F, Mlodzianoski M, Diaspro A, Bewersdorf J, Hess ST, Braeckmans K, Nat. Methods 2014, 11, 253. [PubMed: 24577276]
- [23]. Smith CS, Joseph N, Rieger B, Lidke KA, Nat. Methods 2010, 7, 373. [PubMed: 20364146]
- [24]. Smith CS, Preibisch S, Joseph A, Abrahamsson S, Rieger B, Myers E, Singer RH, Grunwald D, Cell Biol J 2015, 209, 609.
- [25]. Lécuyer E, Parthasarathy N, Krause HM, Methods Mol. Biol. 2008, 420, 289. [PubMed: 18641955]
- [26]. Batish M, van den Bogaard P, Kramer FR, Tyagi S, Proc. Natl. Acad. Sci. USA 2012, 109, 4645. [PubMed: 22392993]
- [27]. Mikl M, Vendra G, Kiebler MA, EMBO Rep. 2011, 12, 1077. [PubMed: 21869818]
- [28]. Knowles RB, Sabry JH, Martone ME, Deerinck TJ, Ellisman MH, Bassell GJ, Kosik KS, J. Neurosci. 1996, 16, 7812. [PubMed: 8987809]
- [29]. Gao Y, Tatavarty V, Korza G, Levin MK, Carson JH, Mol. Biol. Cell 2008, 19, 2311. [PubMed: 18305102]
- [30]. Trcek T, Grosch M, York A, Shroff H, Lionnet T, Lehmann R, Nat. Commun. 2015, 6, 7962. [PubMed: 26242323]
- [31]. Stallinga S, Appl. Opt. 2005, 44, 849. [PubMed: 15751673]
- [32]. Serge A, Bertaux N, Rigneault H, Marguet D, Nat. Methods 2008, 5, 687. [PubMed: 18604216]
- [33]. Smith CS, Stallinga S, Lidke KA, Rieger B, Grunwald D, Mol. Biol. Cell 2015, 26, 4057. [PubMed: 26424801]
- [34]. Braat J, Appl. Opt. 1997, 36, 8056. [PubMed: 18264336]
- [35]. Friedman JH, Bentley JL, Finkel RA, ACM Trans. Math. Software 1977, 3, 209.

Figure 1.

Relative localization of multiple signals in 3D: Left) Nuclei are segmented using a box (boxing method) that is set to enclose the structure in its mid-plane. This results in areas outside of the nucleus that are interpreted as being inside the nucleus (see panels for boxing method). If multiple nuclei exist in different z-planes of the image stack that overlap in the z-projection, boxing methods can lead to ambiguities when assigning smFISH signals to the correct nucleus. Right) If nuclei are segmented using a hull (shape method), the probability of correct assignment for smFISH signals in and close to the nucleus increases. The hull can also be used to estimate the shape of a nucleus that is incompletely labeled—often the case with lamin staining.

Figure 2.

Automated nucleus segmentation and spot finding: A) Visualization of raw data for pak mRNAs labeled with odd (yellow) and even (magenta) probes. The nucleus is labeled with antilamin antibody (blue). B) Identical view of the 3D projected segmentation and spot finding results. Cyan indicates odd probes, green indicates even probes, and red the nuclear volume. C) Overlay of (A) and (B). Green and red channels were assigned yellow and magenta lookup tables for visualization.

Figure 3.

Protocol Optimization: The high amount of background and autofluorescence as well as limited accessibility of probes and antibodies to deeper tissue regions made it necessary to adjust existing smFISH protocols. Panels (A)–(E) show maximum projections of 3D image stacks. A) Nonoptimized protocol as provided by manufacture of FISH probes. B–E) Examples of optimized samples (see Table 1 for exact conditions for all panels). These are minor variations of the ideal protocol. To exemplify the variation between replicates: the sample shown in panel (C) differs from (D) and (E) (identical) only in the fixation step; samples in panels (D) and (E) are different replicates with strong differences in background. Based on our analysis of at least three biological replicates and at least four different fields of view per sample, we find that conditions (D) and (E) offer the highest reliability. F) 3D visualizations of (A). G) 3D projection of (E). Full stacks and 3D projections in supplement. Contrast and brightness of all images were adjusted linearly to their minimum and maximum intensity, and the minimum value set above the background level. Green and red channels were assigned yellow and magenta lookup tables for visualization. Optimizations were done on >10 samples with >3 repeats for each protocol condition. The scale bar represents 5 μm , images were processed as described in the Experimental Section.

Figure 4.

Probe specificity and reliability: Two probes sets with different dyes against Pak mRNA were used to test labeling efficiency through colocalization under ideal preparation conditions. A) 3D projection of one of seven replicates showing more than 90% colocalization. Arrows highlight typical examples of smFISH signals visible in both or only one channel. B–D) Zoom out from box in panel (A), showing the B) colocalized, C) odd, and D) even smFISH signals. Arrows as in panel (A). Besides colocalization, the variance in signal between spots in each channel is noteworthy—see arrow 4 for example. Often a bright spot in one channel corresponded to a weak spot in the other channel. E) RNase treated sample processed (smFISH and IF) identically to (A) with equal intensity scaling. Contrast and brightness of A–D) images were adjusted linearly to their minimum and maximum intensity, and the minimum value set above the background level. Green and red channels were assigned yellow and magenta lookup tables for visualization.

Figure 5.

Distribution of megaRNPs: Brightness and distribution of double positive (odd and even probes) labeled Pak mRNAs. The vertical gray bar indicates the NE with a width of 2 voxels, note that in all data sets we find only one mRNA localizing inside the NE. The horizontal red line is set to 1 unit in normalized intensity with the transparent red bar indicating the 0.5 to 1.5 range of normalized intensity units classified as single mRNA. The horizontal gray line indicates an arbitrary lower threshold of 7 mRNAs per spot as cutoff between megaRNPs (>7 RNA) and single mRNAs or low copy number mRNA clusters. No megaRNPs are located in the cytoplasm. Cytoplasmic clusters of mRNAs with 5 or less mRNAs are more prevalent close to the NE. No data-points exist between 20 and 50 mRNAs. The data have been pooled from three replicates, which are indicated by color of the spots.

Figure 6.

Composition of megaRNPs: Pak, Par6, and Cask mRNAs were detected in megaRNPs and predicted to travel together in megaRNPs. Colocalization analysis of smFISH for pairs of these mRNAs gives insight into whether megaRNPs contain multiple mRNA species. Nuclei are stained with antilamin antibody (blue). A) Par6 and Cask mRNA smFISH, B) Par6 and Pak mRNA smFISH, C) Cask and Pak mRNA smFISH, and D) control using smFISH against GluRIIA and Pak two mRNA species not likely to colocalize. Green and red channels were assigned yellow and magenta lookup tables for visualization.

Table 1.

Optimization of sample preparation for combined smFISH and IF.

Fixation	Concentration	Time	Temperature	Panel Figure 1	Comment
Formaldehyde	3.7%	30 min	20 °C	B	Best condition
Formaldehyde	3.7%	45 min	20 °C	A, C, D, E	
Ethanol	70%	60 min	4 °C	C	Number of smFISH signals reduced
Ethanol	70%	480 min	4 °C	Not shown	
Ethanol & Triton-X100	70% + 0.5%	60 min	4 °C	Not shown	Sample damaged/fragile
Methanol	70%	2 min	-20 °C	Not shown	Increased background fluorescence
Methanol	70%	10 min	-20 °C	Not shown	Increased background fluorescence
Triton-X100 in PBS	0.5%	15 min	20 °C	A, B, D, E	Best condition
Protease digestion					
Pepsin	0.01 µg mL ⁻¹	30 min	4 °C	Not shown	Best condition
Pepsin	0.01 µg mL ⁻¹	30 min	20 °C	Not shown	
Pepsin	0.1 µg mL ⁻¹	15 min	20 °C	Not shown	Best condition
Proteinase K	0.01 µg mL ⁻¹	30 min	20 °C	B, C, D, E	
Proteinase K	0.02 µg mL ⁻¹	10 min	20 °C	Not shown	Double incubation, sample damaged
Proteinase K	0.02 µg mL ⁻¹	30 min	4 °C	Not shown	
Proteinase K	0.03 µg mL ⁻¹	13 + 60 min	20 °C + 4 °C	Not shown	Best condition
Paraformaldehyde	4%	15 min	20 °C	B, C, D, E	
Paraformaldehyde	4%	30 min	20 °C	Not shown	No improvement
Triton-X100 in PBS	0.1%	15 min	20 °C	Not shown	
Triton-X100 in PBS	0.2%	15 min	20 °C	Not shown	No improvement

Table 2.

Final Protocol.

	Time [min]	Temperature	Step
1	2–5 min per larvae	≈2 °C (ice cold)	Dissect larvae
2	45	20 °C	Fixation with 3.7% formaldehyde
3	15	20 °C	Permeabilization with 0.5% Triton X-100
4	30	20 °C	Protease digestion with 0.01 µg mL ⁻¹ proteinase K
5	≈3	20 °C	Wash with glycine (2 mg mL ⁻¹ in PBS)
6	15	20 °C	Postfixation with 4% paraformaldehyde
7	≈1	20 °C	Wash with 0.1% Triton X-100 in PBS
8	≈1	20 °C	Wash with 2× saline-sodium citrate buffer (SSC) containing 10% deionized formamide
9	240 (4 h)	37 °C	Hybridize smFISH probes and primary antibody, add 1 mg mL ⁻¹ tRNAs as competitor in hybridization buffer, final concentration
10	≈3	20 °C	Wash with 2× SSC containing 10% deionized formamide
11	60	37 °C	Incubate with secondary antibody in the dark
12	≈3	20 °C	Wash with 2× SSC containing 10% deionized formamide
13	≈1	20 °C	Rinse with PBS
14	≈3	20 °C	Mount with mounting medium
Additional comments:			

Penetration rate for fixative is ≈1 mm h⁻¹ (<http://www.leicabiosystems.com/pathology/leaders/fixation-and-fixatives-2-factors-influencing-chemical-fixation-formaldehydeand-glutaraldehyde/>)

Shorter initial fixation times lead to a reduction in number of detectable smFISH signals, longer postfixation times (after permeabilization) also lead to a reduction in number of detectable smFISH signals

Methanol permeabilization increase background levels to the point where no smFISH signals could be detected

IF fluorescence staining did not compromise smFISH signal, e.g., due to RNase contamination of the antibody

Hybridization specificity of smFISH probes improved in the presence of tRNAs but not with N20 DNA oligomers

Increased hybridization temperature of 45 °C did not reduce nonspecific smFISH signal, but did damage tissue

Harsher washes with reduced 0.1× or 0.5× SSC, instead of normal wash with 2× SSC, did not reduce nonspecific smFISH signal