



Counting mRNA Copies in Intact Bacterial Cells by Fluctuation Localization Imaging-Based Fluorescence In Situ Hybridization (fliFISH)

Dehong Hu, Yi Cui, Lye M. Markillie, William B. Chrisler, Qian Wang, Roland Hatzenpichler, and Galya Orr

Abstract

A method for measuring mRNA copies in intact bacterial cells by fluctuation localization imaging-based fluorescence in situ hybridization (fliFISH) is presented. Unlike conventional single-molecule FISH, where the presence of a transcript is determined by fluorescence intensity, fliFISH relies on On-Off duty cycles of photo-switching dyes to set a predetermined threshold for distinguishing true signals from background noise. The method provides a quantitative approach for detecting and counting true mRNA copies and rejecting false signals with high accuracy.

Key words Single-molecule FISH, Super-resolution fluorescence imaging, Gene expression, mRNA copies, Bacterial cells, Photo-blinking

1 Introduction

Building on single-molecule localization-based super-resolution fluorescence imaging as STORM or PALM [1, 2], we have developed fluctuation localization imaging-based fluorescence in situ hybridization (fliFISH), where individual transcripts are detected and counted in intact cells with high resolution (20–30 nm) and exceptional accuracy [3]. Unlike conventional single-molecule FISH (smFISH), where the presence of a transcript is determined by fluorescence intensity, fliFISH relies on On-Off duty cycles of photo-switching dyes to determine the presence of a transcript. The method uses On-time fraction values, measured over a series of exposures, which are distinct for transcripts bound to a known number of oligonucleotide FISH probes, compared to nonspecifically bound stray probes, as well as autofluorescence. Thus, On-time fraction values can guide the setting of a threshold for

distinguishing true from false signals. More details about the fliFISH concept and method can be found in Cui et al. [3].

The application of fliFISH to count mRNA copies for multiple genes in intact eukaryotic cells has been also demonstrated [4]. Here, we provide the method for applying fliFISH to count mRNA copies in intact bacterial cells, where additional challenges are presented. First, the number of mRNA copies in a bacterial cell can be quite low [5]. Consequently, any nonspecific binding by stray probes can mask the presence of a few true signals and introduce significant counting errors. Second, due to the small volume of bacterial cells, transcripts for genes that are expressed at higher levels tend to show within tightly packed clusters, rather than distinct copies or loosely packed clusters as they mostly appear in larger eukaryotic cells. Such packing within clusters presents a challenge for accurate quantification of copy numbers. Here, we provide a protocol optimized for counting copies at low-expression levels, as well as a computational approach for counting transcript copies within tight clusters. We demonstrate the application of fliFISH by targeting aspartate aminotransferase (AAT) mRNA expressed in *E. coli* at varying expression levels, from low to high transcript copies per cell.

2 Materials

2.1 FISH Probe Preparation

1. FISH probes design: Each primary FISH probe contains a sequence of about 20 oligonucleotides (NTs) that are complementary to the target mRNA, and a sequence of 28 NTs overhang that will be hybridized with the secondary probe. Multiple 20 NT-long probes are designed to hybridize with different segments of the target mRNA, each includes the same overhang sequence. Here, we used 16 probes to target AAT mRNA (Table 1). Suggestions for probe design and selection can be found in **Note 1**.
2. Secondary probe: The use of the secondary probe follows the approach described in Tsanov et al. [6]. The secondary probe is tagged with two Alexa 647 dye molecules, one at each end, and is designed to hybridize with the 28 NT-long overhang sequence. Alexa 647 has excellent single-molecule blinking property but other dyes for super-resolution microscopy (PALM/STORM) could be used [7]. The secondary probe sequence we used here was: CACTGAGTCCAGCTC GAAACTTAGGAGG . It was ordered commercially with HPLC purification. *See Note 1* for more details.
3. 10× NEB3 buffer: The buffer is used for preparing the FISH probe solution for hybridization and it contains 1 M NaCl, 0.5 M Tris-HCl and 0.1 M MgCl₂.

Table 1
The sequences of the 16 FISH probes used to target the AAT transcripts

5-GCCTTGGCCACTTCCATCACCCCTCCTAAGTTTTCGAGCTGGACTCAGTG-3
5-AAATCATCGGCTCGCGCTGCCTCCTAAGTTTTCGAGCTGGACTCAGTG-3
5-CAGTGCAGCTTCTGCACCACCTCCTAAGTTTTCGAGCTGGACTCAGTG-3
5-TACCAGCCGCTGATGCGTTCCCTCCTAAGTTTTCGAGCTGGACTCAGTG-3
5-GTTCACGCCAAAGCGGCTCTCCTCCTAAGTTTTCGAGCTGGACTCAGTG-3
5-AAACACGCCAGCTGCAGTGCCCTCCTAAGTTTTCGAGCTGGACTCAGTG-3
5-TTGTCGGCGCTCAGCTGGTACCTCCTAAGTTTTCGAGCTGGACTCAGTG-3
5-TTCATTCCATGCGGCACGCACCTCCTAAGTTTTCGAGCTGGACTCAGTG-3
5-TCATCGTGATGCCGTCGTGCCCTCCTAAGTTTTCGAGCTGGACTCAGTG-3
5-TGTTGTCTCGATGGCAAGCCCTCCTAAGTTTTCGAGCTGGACTCAGTG-3
5-CTGAAGCTGTTGATGCTGATCCTCCTAAGTTTTCGAGCTGGACTCAGTG-3
5-GCCGGTCATGTTGAAGTACTCCTCCTAAGTTTTCGAGCTGGACTCAGTG-3
5-ATGTTGCGACACCGTGCTGGCCCTCCTAAGTTTTCGAGCTGGACTCAGTG-3
5-GTACTCGGCAATGCTCTCGGCCCTCCTAAGTTTTCGAGCTGGACTCAGTG-3
5-TCAGCTCGAACGCACACTCCCCTCCTAAGTTTTCGAGCTGGACTCAGTG-3
5-AAAGCGCACGAAACGTTTCGGCCCTCCTAAGTTTTCGAGCTGGACTCAGTG-3

2.2 Cell Fixation

1. Phosphate buffered saline (PBS): 137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄ in water.
2. 50% ethanol and 50% Phosphate-buffered saline (PBS).
3. 4% paraformaldehyde (PFA) in PBS.
4. Centrifuge: Eppendorf 5430R or an equivalent centrifuge.

2.3 Hybridization and Imaging

1. 2× saline-sodium citrate (SSC): 0.3 M NaCl in 0.03 M sodium citrate. This buffer is used to prepare the wash and hybridization buffers below.
2. Wash buffer: 10% formamide, 0.01% sodium dodecyl sulfate (SDS), 2× SSC and 0.6 M NaCl.
3. Hybridization buffer: 10% dextran sulfate, 10% formamide, 0.01% SDS, 2× SSC and 0.6 M NaCl.
4. Imaging buffer: 10% glucose in 10 mM Tris buffer, 0.5 mg/ml glucose oxidase, 12% (V/V) of catalase (1 mg/ml), 1% beta-mercaptoethanol. The buffer must be freshly prepared right before imaging (Originally described in [2]).
5. Poly-L-Lysine solution: 0.1% W/V in H₂O.

6. Hybridization oven (any small oven that provides the required temperature and space).
7. Glass coverslip: 0.17 mm thick, variable sizes (must fit the microscope sample holder).
8. Nail polish (any brand and color).
9. Glass-bottom petri dish or chamber.

2.4 Microscope

Any wide-field fluorescence microscopy system with single-molecule detection sensitivity can be used. The system used in this work was a motorized wide-field inverted fluorescence microscope (AxioObserver Z1) from Zeiss. The laser source was 660 nm wavelength, 100 mW power, solid state laser from Crystalaser. The power was modulated by an acousto-optic tunable filter (AOTF). The laser beam was coupled through a single-mode optical fiber and was brought to the microscope through the total internal reflection fluorescence (TIRF) port. A 100× objective lens (NA 1.46, alpha Plan-Apochromat oil DIC) was used. The fluorescence filters were a dichroic (670 DCXR) and a long pass (HQ680LP) from Chroma. An electron-multiplied CCD (EMCCD) camera (iXon 888) from Andor was used in this work.

3 Methods

3.1 FISH Probe Preparation

1. Mix the probes for the target gene (in our case 16 probes for AAT) by taking 1 μl of 100 μM in water from each probe. Add water to a total of 80 μl .
2. FISH probe solution for hybridization: Take 4 μl from the above mixture, add 1 μl 100 μM Secondary probe, 2 μl 10× NEB3 buffer, and 13 μl water. Heat the mixture to 85 °C for 3 min in a water bath and gradually cool down to room temperature to allow the primary and secondary probes to melt and hybridize. This step was adapted from [6].

3.2 Bacterial Cell Fixation

1. Suspend the cells in 200 ml 4% PFA in PBS for 20 min at room temperature. Preferably, enough cells are available to form a pellet at the bottom of the tube, which can be seen by the naked eye. Otherwise, a magnifying glass or a microscope should be used to verify the presence of the pellet.
2. Wash the fixative using PBS by spinning down the suspension and resuspending the pellet in PBS. Repeat two more times.
3. Resuspend the washed cells in 50% ethanol/50% PBS and store at −20 °C. The cells can be stored for weeks at this point.

3.3 Cell Hybridization with FISH Probes

1. Mix 10 μl of the above bacterial suspension with 50 μl PBS. Spin down at $5000 \times g$ for 4 min. Remove supernatant.
2. Resuspend the cells in 50 μl PBS. Spin down at $5000 \times g$ for 4 min. Remove supernatant.
3. Resuspend the cells in 50 μl Wash buffer. Spin down at $5000 \times g$ for 4 min. Remove supernatant.
4. Add 30 μl Hybridization buffer and 6.6 μl FISH probe solution for hybridization (prepared in Subheading 3.1, step 2). Place in a hybridization oven at 37°C overnight.
5. Spin down at $5000 \times g$ for 4 min. Remove supernatant and resuspend in 50 μl Wash buffer. Keep in the hybridization oven at 37°C for 15 min.
6. Repeat the step above.
7. Spin down at $5000 \times g$ for 4 min. Remove supernatant and resuspend in 50 μl PBS buffer. Repeat twice. This is the cell suspension for imaging.
8. Probe concentrations, buffer components, hybridization time, and temperature were adjusted to target challenges presented in the application of fliFISH in bacterial cells as discussed in **Note 2**.

3.4 Fluorescence Imaging

1. Mix 1 μl of the cell suspension for imaging (prepared in Subheading 3.3, step 7) with 5 μl Imaging buffer.
2. Place this mixture on a glass coverslip. Cover with another glass coverslip and seal the edge with nail polish.
3. Imaging setup is similar to single molecule-based super-resolution fluorescence microscopy such as STORM or PALM. The cells are imaged using an inverted fluorescence microscope with $100\times$ oil immersion objective. Images are acquired using a CCD camera with single-molecule detection sensitivity. The laser power and camera setting, such as exposure time and gain, are hardware-dependent and should be adjusted based on the user's microscope. The laser power density should be set at several hundred to several thousand watts per square centimeter. In our case, the microscope power density was $500\text{ W}/\text{cm}^2$. The camera exposure time should be shorter than a typical blinking or On-time of the fluorescent dye molecule. In our case, the exposure time was set at 0.5 s. The total number of camera exposures should be at least a thousand or until the dye molecules are photobleached to ensure that enough blinking events are available for fliFISH analysis. Under our microscopy setting, about 2500 images were collected in each series at a continuous mode. The settings should be consistent for all samples to allow quantitative comparisons between images.

4. Optional. If the microscope is equipped with a focus drift correction hardware (for example, Zeiss Definite Focus), enabling it would increase method accuracy.
5. The average On-time fraction of a single FISH probe is required for fliFISH analysis. To quantify this value ($F_{\text{single-on}}$), dilute 0.1 nM dye-labeled probes in Imaging buffer. Place 5 μl of the solution in a glass-bottom petri dish or chamber and leave for 20 min. The use of glass-bottom coated with poly-L-lysine is recommended to increase the attachment of the dye-labeled probes. Use the exact same microscope settings to acquire images of the labeled probes.

3.5 fliFISH Analysis

1. Overview: The first step in fliFISH analysis is the localization of single-molecule blinking events, which is similar to the analysis of single-molecule localization super-resolution imaging such as STORM/PALM. Thus, freely available analysis software for single-molecule localization microscopy [8], such as Quick-PALM [9] or DAOSTORM [10] can be used for this initial step. The following steps are specific to fliFISH analysis in bacterial cells and include: quantifying the blinking properties of single dye molecules, identifying clusters, and counting transcripts. These steps are described below. Also *see* **Note 3** for a discussion of approaches to identify clusters. The analysis program used by us is available upon request.
2. Analysis of the acquired time series of single-molecule images, using any available software for single-molecule localization microscopy [8], calculates the center and intensity of each blinking event in each image in the series. The software produces a list of single-molecule blinking events with X, Y locations and frame numbers. The X, Y location units can be presented either as the camera pixel numbers or in nanometers. The two units can be converted to each other by a scale factor: location (nm) = location (pixel) \times camera sensor pixel size (nm)/microscope magnification. In our experimental setup, the camera's pixel size was 13 μm and the microscope magnification was 160 \times . Thus, the scale factor was 81 nm/pixel. Note that locations presented as pixel numbers may be decimals as the analysis software identifies the center location of single molecules with sub-pixel resolution. The list of single-molecule blinking events with X, Y locations and frame numbers is used to reconstruct the super-resolution fluorescence image.
3. The average On-time fraction of single FISH probes ($F_{\text{single-on}}$) is required for fliFISH analysis and is quantified by imaging fluorescent probes on glass-bottom dishes as prepared in Sub-heading 3.4, **step 5**. Images are acquired under the exact same conditions that are used to image the cells. The On-time fraction and location (with nm resolution) for each diffraction-

limited fluorescent spot in the image series are calculated. This list is used to calculate the average on-time fraction of a single probe.

4. The average on-time fraction of a single probe ($F_{\text{single-on}}$), found in the above step, together with the number of FISH probes (N) hybridized to each target transcript, are used to calculate the average On-time fraction expected from a transcript bound to multiple probes ($F_{\text{ensemble-on}}$). This value is calculated according to the equation below, which is also found in Cui et al. [3].

$$F_{\text{ensemble-on}} = 1 - (1 - F_{\text{single-on}})^{N_{\text{probe}}} \quad (1)$$

In the reality, the hybridization efficiency is usually only around 70% [10, 11]. Since the protocol provided here uses two sequential hybridization steps, using primary and secondary probes, the hybridization efficiency is expected to be lower. The exact hybridization efficiency, in this case, has not been measured. Here, we estimate it at 60%. Thus, only 60% of the probes designed to target a transcript should be used as N in the equation above. For example, here we designed 16 probes for targeting each transcript. Thus, in our case, N equals 9.6. $F_{\text{ensemble-on}}$ will be used in the steps below to find the number of transcripts within clusters of fluorescent spots, which are selected by a minimum On-time fraction threshold determined as described below.

On-time fraction values for single fluorophores are widely distributed due to the stochastic nature of the process. Thus, On-time fractions for single probes, as well as single transcripts, have a wide distribution as well (*see* Figs. 2B and 4A in Cui et al. [3]). To minimize bias against transcripts bound with a small number of probes, a minimum On-time fraction threshold is set to select for fluorescent spots that should be further evaluated as true transcripts. In this protocol, we suggest the use of the geometric middle point to determine the minimum On-time fraction threshold. For example, here we used 16 probes to target a gene (considering 60% hybridization efficiency: $N = 16 \times 60\% = 9.6$), and experimentally found that On-time fraction of a single probe ($F_{\text{single-on}}$) was 0.36%. Using these values in Eq. 1 above, the average On-time fraction for a transcript ($F_{\text{ensemble-on}}$) was 3.5%. Similarly, using the value for four probes (the geometric middle point between 16 and 1 probes), the minimum On-time fraction that was used here as the lower threshold for selecting fluorescent spots for further consideration was 0.9%. An upper threshold can be set as well (in our case using four times the actual number of probes to calculate the maximum On-time fraction for a transcript).

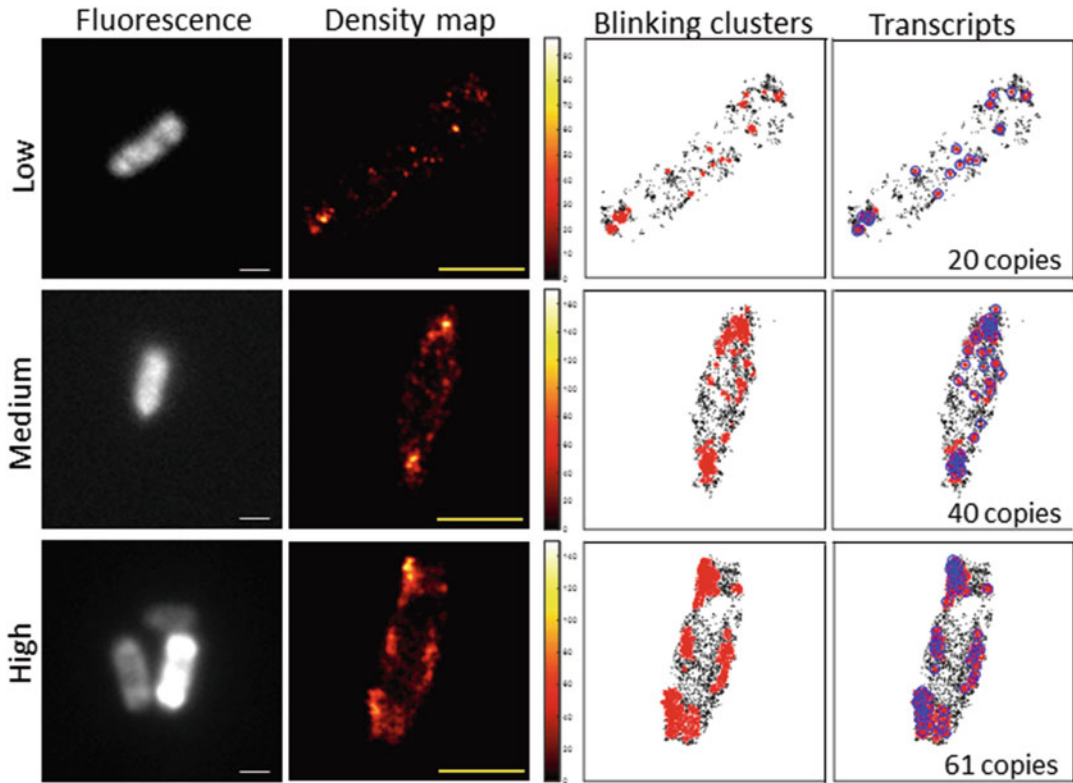


Fig. 1 Examples for fliFISH analysis of three selected *E. coli* cells expressing low, medium, and high levels of AAT transcript copies. The fluorescence images for each cell are shown on the left, followed by a density map where the colors correspond to the number of blinking within each pixel. Blinking events are shown next, where events that are clustered within 24 nm radius are marked in red, while non clustered events are marked in black. In the final step, clustered events are further analyzed for the number of transcripts, indicated with blue circles, as described in the text. Scale bars equal 1 μ m. In the example for high expression (bottom row), only the bright cell in the fluorescence image is pursued

However, this value (13.6%) was never reached within the radius R of a fluorescent spot, and therefore, is not needed.

- The next step in the fliFISH analysis is to group blinking events into clusters in the image series of cells hybridized with the FISH probes. To achieve this, we introduce the density map (Fig. 1), where the value of each pixel (represented by the color map) indicates the number of blinking events within distance R of the pixel. R is the accuracy of localizing a single molecule in the raw images (**step 2**), which is equivalent to the resolution achieved in images acquired by PALM/STORM and is impacted by the fluorescence intensity signal to background ratio [11]. In our case, R was found to be 24 nm. Details about how to calculate R can be found in **Note 4**. The density map shares the same pixel size and pixel number with the reconstructed super-resolution image.

6. Using the minimum On-time fraction value, calculated in **steps 3** and **4**, to set a threshold, patchy areas above the threshold are identified in the density map and the boundaries of these areas are expanded by distance R . The blinking events within the expanded areas are therefore considered to originate from true transcripts. The blinking events outside the expanded areas are considered to come from nonspecifically bound stray probes. The number of blinking events in each expanded area is counted and is translated to transcript copy numbers by dividing it by the actual average On-time fraction that is expected from a transcript, as calculated in **steps 3** and **4**. The sum of transcripts found in each expanded area in the cell provides the total transcript copies for that cell. The computation of this step is performed by a MATLAB routine that is posted online (<https://github.com/hudehong/flifish>).
7. Optional lateral drift correction. If the microscope stage has significant lateral drift during data acquisition, the drift can be corrected by the analysis software without experimentally introducing fiducial markers. The method is described in detail in [12]. Briefly, the lateral drift is found by the cross-correlation of single-molecule locations in the image series and is subtracted from the locations of all molecules in each frame. This step is performed before a super-resolution image is reconstructed and before the fliFISH analysis is applied.

4 Notes

1. A useful online tool for designing a set of oligonucleotide FISH probes from an mRNA coding sequence is the Stellaris Probe Designer, which can be found at: <https://www.biosearchtech.com/stellaris-designer>. The properties of the probes, such as T_m , DG, and hybridization efficiency, should be evaluated. A useful online tool for evaluating these parameters is MathFISH, which can be found at: <http://mathfish.cce.wisc.edu/index.html>. All the probes within a set should have similar properties at a given hybridization condition, including temperature, salt, and formamide concentrations. A probe with properties that are different from the rest of the probes should be excluded. The probes that we used here were ordered from Integrated DNA Technologies with desalt purification.
2. As mentioned earlier, one of the challenges in quantitative mRNA FISH in bacterial cells is presented by the low-expression levels of some genes, where only few copies could be found in a cell. Using high concentrations of fluorescent probes can, therefore, lead to a high level of nonspecifically bound stray probes and significantly bias the counting. Here,

we provide a protocol where the probe concentration has been optimized to be in the nM level, and the hybridization time and temperature, as well as buffer components, have been adjusted.

3. In our earlier publications describing fliFISH [3, 4], we have used a density-based spatial clustering of applications with noise (DBSCAN) program [13] to identify clusters of single-molecule blinking events. However, DBSCAN requires the nearest neighbor distance between molecules as an input parameter, which is affected by the data acquisition length of time, as well as by random noise, and therefore, is poorly defined. Additionally, the DBSCAN program is computationally inefficient and analyzing a whole image may take many hours. Here, we use a different approach, described in Subheading 3.5, steps 5 and 6. The analysis uses reliably measured input parameters with physical meaning.
4. There are two methods to estimate R . The first method estimates R from experimental data acquired in Subheading 3.4, step 5 and analyzed in Subheading 3.5, step 3. The blinking events of single molecules on a glass surface are fitted to obtain the center location of each event. The locations are clustered into a nanometer-sized circular region with radius R . As the glass surface has a lower fluorescence background compared with cells, a slightly larger R may be used for cellular data. The second method to estimate R is based on theoretical calculations as described in Thompson et al. [11]. In this reference, an equation was derived for calculating the accuracy of single-molecule location fitting from the signal level, background level, and camera pixel size. The camera pixel size is a known value, while the signal and background levels are obtained from the single-molecule images. With these values, R can be calculated using Eq. 17 in that ref. [11]. The resulted number from this equation is in terms of standard deviation. R is set to be two times the standard deviation, which means 95% of the single-molecule blinking locations are inside the circle with radius R . We prefer the first method because it is more convenient and more relevant to the experimental data.

Acknowledgments

We thank Dr. Timothy McDermott at the Land Resources & Environmental Sciences, Montana State University, for providing us with *E. coli* expressing AAT. The majority of the work was done at the Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by DOE's Office of Biological and Environmental Research (BER) and located at Pacific Northwest National Laboratory. Part of the work was

done under the Facilities Integrating Collaborations for User Science (FICUS), using resources at EMSL and the Joint Genome Institute (JGI). Both are DOE scientific user facilities sponsored by BER and operated under Contract Nos. DE-AC05-76RL01830 (EMSL) and DE-AC02-05CH11231 (JGI).

References

1. Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, Davidson MW, Lippincott-Schwartz J, Hess HF (2006) Imaging intracellular fluorescent proteins at nanometer resolution. *Science* 313 (5793):1642–1645
2. Rust MJ, Bates M, Zhuang X (2006) Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods* 3(10):793
3. Cui Y, Hu D, Markillie LM, Chrisler WB, Gaffrey MJ, Ansong C, Sussel L, Orr G (2018) Fluctuation localization imaging-based fluorescence in situ hybridization (fliFISH) for accurate detection and counting of RNA copies in single cells. *Nucleic Acids Res* 46(2):e7
4. Cui Y, Melby ES, Mensch AC, Laudadio ED, Hung MN, Dohnalkova A, Hu D, Hamers RJ, Orr G (2019) Quantitative mapping of oxidative stress response to lithium cobalt oxide nanoparticles in single cells using multiplexed in situ gene expression analysis. *Nano Lett* 19 (3):1990–1997
5. Skinner SO, Sepulveda LA, Xu H, Golding I (2013) Measuring mRNA copy number in individual *Escherichia coli* cells using single-molecule fluorescent in situ hybridization. *Nat Protoc* 8(6):1100–1113
6. Tsanov N, Samacoits A, Chouaib R, Traboulsi AM, Gostan T, Weber C, Zimmer C, Zibara K, Walter T, Peter M, Bertrand E, Mueller F (2016) smiFISH and FISH-quant - a flexible single RNA detection approach with super-resolution capability. *Nucleic Acids Res* 44 (22):e165
7. Dempsey GT, Vaughan JC, Chen KH, Bates M, Zhuang X (2011) Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. *Nat Methods* 8(12):1027–1036
8. Sage D, Pham T-A, Babcock H, Lukes T, Pengo T, Chao J, Velmurugan R, Herbert A, Agrawal A, Colabrese S, Wheeler A, Archetti A, Rieger B, Ober R, Hagen GM, Sibarita J-B, Ries J, Henriques R, Unser M, Holden S (2019) Super-resolution fight club: assessment of 2D and 3D single-molecule localization microscopy software. *Nat Methods* 16:387–395
9. Henriques R, Lelek M, Fornasiero EF, Valtorta F, Zimmer C, Mhlanga MM (2010) QuickPALM: 3D real-time photoactivation nanoscopy image processing in ImageJ. *Nat Methods* 7(5):339–340
10. Holden SJ, Uphoff S, Kapanidis AN (2011) DAOSTORM: an algorithm for high-density super-resolution microscopy. *Nat Methods* 8 (4):279–280
11. Thompson RE, Larson DR, Webb WW (2002) Precise nanometer localization analysis for individual fluorescent probes. *Biophys J* 82 (5):2775–2783
12. Geisler C, Hotz T, Schonle A, Hell SW, Munk A, Egner A (2012) Drift estimation for single marker switching based imaging schemes. *Opt Express* 20(7):7274–7289
13. Ester M, Kriegel H-P, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: Simoudis E, Han J, Fayyad UM (eds) *Proceedings of the second international conference on Knowledge Discovery and Data Mining (KDD-96)*