Development of an analysis tool for quantification of cellular structures by single molecule localization microscopy

Author: Pere López Gutiérrez
Advisor: Sílvia Pujals Riatós
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.*

Abstract: Super resolution microscopy techniques allows optical imaging with sub-diffraction resolution. In single molecule localization microscopy, each target molecule is imaged separately, with a lateral resolution that routinely achieves 20 nm, and the final image is reconstructed by merging coordinates of single-molecule positions. Derivation of quantitative parameters is therefore possible by the analysis of single-molecule coordinates instead of processing of pixel-based images. Combined analysis of single-molecule localization microscopy data and low-resolution pixel-based images is often challenging, even for elementary operations, due to the lack of specific software analysis tools. The aim of this project is to develop a code for quantitative analysis of single-molecule localization microscopy data, using a low-resolution pixel-based image as reference to select specific areas of interest. After testing the methodology on model samples, the analysis was employed to quantify the single-molecule signal from individual protein receptors exposed on the surface of three different cancer cell lines. Remarkably, a low-resolution image of the cells was employed as reference to restrict the quantification to the molecules exposed on the cell membrane area. An easy-to-use program, developed in MATLAB, was written for this analysis and provided robust results, comparable with those obtained with a different methodology, in a considerably shorter time and with an intuitive user interface.

I. INTRODUCTION

Super-resolution microscopy (SRM) is emerging as a new tool in nanomedicine that allows imaging of biomolecules with sub-diffraction spatial resolution. [1]

SRM techniques such as stimulated emission depletion microscopy (STED), structured illumination microscopy (SIM) or single-molecule localization microscopy (SMLM) enable direct measurement of size and morphology of sub-cellular systems and visualization of biomolecules distribution with nanoscale accuracy. [2] Among these, SMLM is endowed with one of the best spatial resolutions, routinely achieving 20 nm, and with the possibility of quantification, that is counting of biomolecules.

SMLM acquisition takes advantages of photo-induced switching of individual fluorescent probes between an off-state (not detected) and an on-state (detected). By illuminating the sample with light of appropriate wavelength and intensity, a small subset of fluorescent reporters is randomly activated to on-state. Fluorescence emission from single molecules in on-state is detected and their centre position is determined with nanometric accuracy, depending on the number of photons detected per individual emitter, by fitting the diffraction-limited fluorescent spot generated from an isolated emitter with a 2D Gaussian model [3]. The off-on cycle is repeated multiple times so that different subsets of molecules are localized in every frame of the acquisition and a final super-resolution image is reconstructed by merging together all the single-molecule positions retrieved during acquisition.

Different types of SMLM fundamentally differ by the way probes are switched between ON and OFF state. Most common techniques are: photo-activated localization microscopy (PALM) [4], that exploits photoactivation of specific fluorescent proteins expressed on the target; stochastic optical reconstruction microscopy (STORM) [5], that is based on the photoswitching of organic dyes placed on the target, typically induced by specific buffer; points accumulation for imaging in nanoscale topography (PAINT) [6]. In PAINT, single molecule localizations are associated to a specific transient interaction between the target molecules and a dye-conjugated probe diffusing in solution at low concentrations, so that only a few probes are found on the target in every acquisition frame and are localized. A common implementation of PAINT exploits the reversible annealing between short single strands of DNA: a *docking* strand, placed on the target, and a dye-conjugated *imager* strand, diffusing in solution (DNA-PAINT) [7]. Remarkably, molecule quantification [8,9] is also possible with this approach by analysing the frequency of interactions.

In this work, the data used are collected via STORM and PAINT.

The information retrieved by SMLM essentially consists in a list of XYT-coordinates of the detected single-molecule emission events (named localizations) rather than an image composed of pixels. Therefore, standard programs for image analysis can hardly be applied in a reliable way to SMLM data. Additionally, simple operations like the direct overlap and alignment of a diffraction-limited image (pixel-based) with the corresponding SMLM data (XYT-coordinates) may result very challenging for unexperienced users and often require the use of specific proprietary softwares and tedious procedures. This is an important limitation, since diffraction limited images are often used as a mask to roughly identify specific regions of interest (ROIs), e.g. specific sub-cellular structures, to which the analysis of SMLM data should be restricted.

The aim of this work is to develop an easy-to-use analysis program for the quantification of SMLM localizations within specific ROIs that are defined by the user on a corresponding pixel-based reference image. The program requires only a few inputs, and then, the quantitative data can be obtained quickly.

II. PROGRAM ARCHITECTURE

The code is written as a MATLAB function and its workflow is schematically represented in Figure 1. The code can be used to select as many ROIs as we want and obtain the coordinates of the localizations within them. For each of these

user-selected ROIs, the area, the number of single-molecule localizations and the density of localizations are reported.

The main inputs consist in a table of XYT-coordinates, that can be directly extracted from the SMLM data obtained with specific softwares, and a pixel-based image of the same field of view, typically with lower resolution, such as a diffraction limited fluorescence or bright-field image. The pixel size of this image is required as additional input parameter.

The program first displays the low-resolution image and allows the user to interactively adjust the contrast or to zoom in, for a better selection of the area of interest. These conditions can be changed during all the execution of the function to facilitate the selection of different areas within the same image.

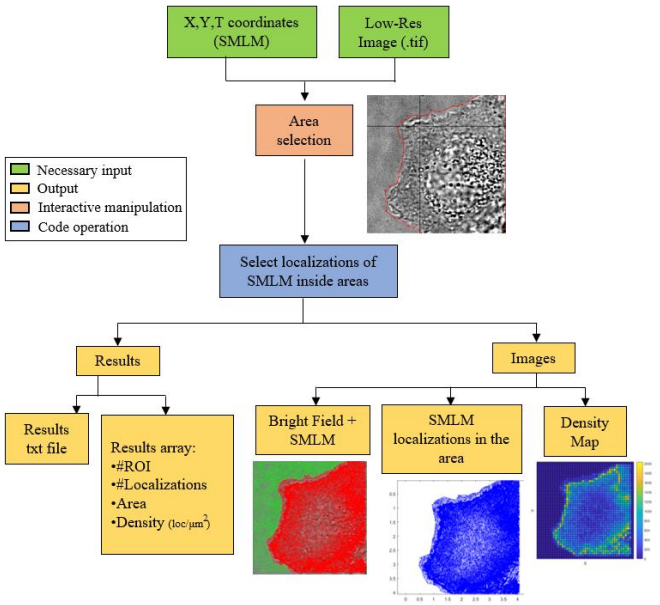

FIG. 1: Workflow of the MATLAB code.

Then, the user is asked to introduce the number of ROIs and to manually select them simply by drawing a polygonal area on the low-resolution image.

Once finished the selection of all the ROIs in the low-resolution image, the program analyse the SMLM data using the coordinates of each polygon vertices defining a ROI as reference. It automatically identifies the number of SMLM localizations within the ROIs and its area, and it calculates the localization density.

Finally, these values are returned and the XYT-coordinates of the selected localizations within each ROIs are stored for further processing.

In addition, the programme returns three output images: one figure shows the low-resolution image overlapped with the SMLM localizations displayed as points within different colour (red if they're inside the ROIs, green outside); another figure shows the coordinates of the selected SMLM localizations inside the ROIs; and last one is a density map of localizations where the SMLM localizations are binned in two-dimensions with a defined bin size.

III. RESULTS

Our analysis code was tested on different structures to prove its versatility. First, we analysed a sample of nanoparticles absorbed of a glass surface. This was exploited as a model system in order to check the accuracy of the alignment of SMLM coordinates with the pixel-based image. Then, we tested the analysis routine of several images acquired on cell samples using different SMLM methodologies like STORM and PAINT. In addition, we employ the program to analyse previously-acquired PAINT data showing the expression level of a cell membrane receptor on different type of cancer cell lines and compared the results with those obtained from a previous analysis, performed without the direct overlap of SMLM with reference images.

A. Accuracy of images alignment

We employed a model sample of nanoparticles seeded on microscope glass slide to check the accuracy of the alignment between SMLM localizations and pixel-based image, when overlapped by the program. This is crucial to understand the limit of accuracy provided by the program and correct for possible systematic errors. Due to their 300 nm size, particles are perfectly resolved by DNA-PAINT imaging but they are also visible in a corresponding low-resolution image of the same field of view, where they appear as diffraction-limited spots with a size of a few pixels.

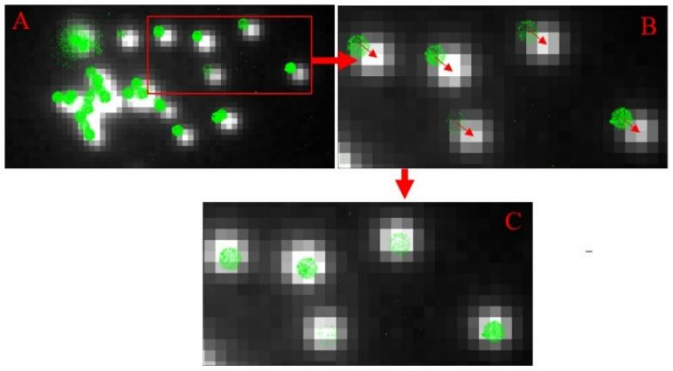

FIG. 2: A) Image of the alignment of SMLM localizations and low-resolution image of nanoparticles, we can see that we have a mismatch in the alignment. B) Correction we want to apply on the code. C) Correction done.

Analysing several samples of nanoparticles we could see that we had a systematic mismatch between particle positions in pixel-based and SMLM images. Specifically, SMLM coordinates resulted always drifted to the top left corner of the low-resolution image, as can be appreciated in Figure.3 A-B-C. This mismatch was caused by a systematic drifting of SMLM coordinates, obtained by the microscope analysis software, with respect to camera pixels and could be easily corrected.

Figure 2 shows overlapped low-resolution and SMLM image with corrected alignment. The precision of our new alignment was estimated to be of less than a pixel, in Figure 3 we see how the overlap changed after the correction.

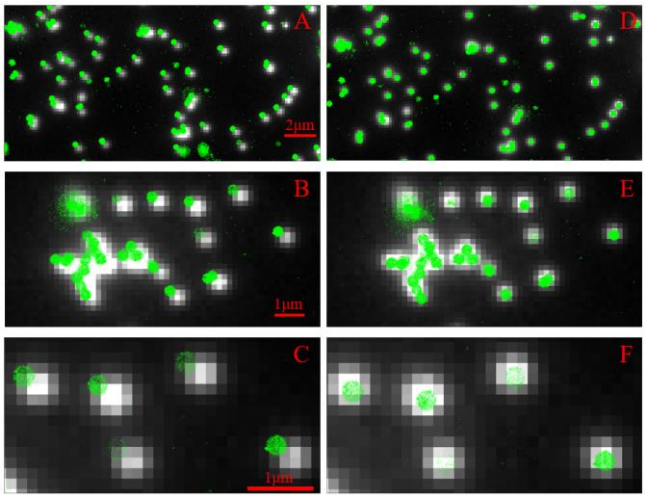

FIG. 3: A-C) Nanoparticles before the correction, there is still the mismatch between the localizations on the low-resolution image. D-F) Nanoparticles after the correction where we can appreciate a correct alignment.

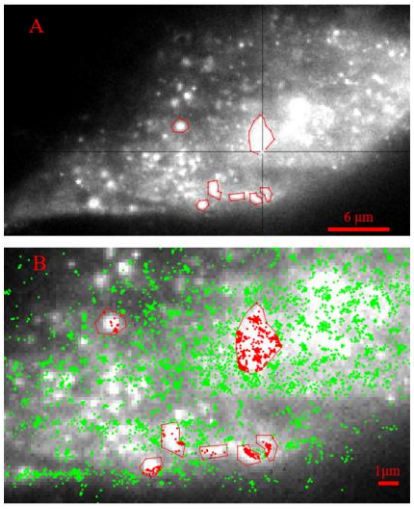

FIG. 4: A) Multiple selection of different endosomes. B) Localizations inside our ROIs after the quantification of the code.

B. Selection of multiple ROIs

In order to test the ability of the program to properly recognize SMLM localizations within multiple ROIs selected in a corresponding low-resolution image, we exploited an image of a cell exposed to dye-labelled particles that are internalized into endosomes. Cells are then chemically fixed and the endosomes are stained with a different dye. Co-localization of the SMLM localizations and low-resolution fluorescence, acquired in different channels, provides the amount of nanoparticles located on endosomes, which is a relevant information but often challenging to obtain due different nature of the images.

In this case, we focus on testing the performance of the function in the selection of more than one ROI. Representative results are shown in Figure 4. In this case, endosomes were immunostained and imaged in low-resolution. These are easy to identify in the low-resolution fluorescence image, and can be easily selected manually by drawing a few polygonal ROIs. Then, the code automatically returns you the density of SMLM localizations, associated to nanoparticles, inside each selected ROIs. This endosomes contain different amounts of SMLM localizations and in many cases co-localization is poor.

The obtained results highlight that density of nanoparticles inside some endosomes are not significant due the bad co-localization.

C. Quantify membrane receptor expression

Having tested the performance of the program on two different model samples, we applied the analysis to a biologically relevant question that is the quantification of membrane protein receptors exposed on different living cell lines. These living cells express epidermal growth factor receptors (EGFR) on their membrane, although with a different expression level. [10,11] EGFR are transmembrane protein receptors of epidermal growth factor relevant for several biomedical applications since they are exploited as biomarkers for a variety of cancer types or as targets for drug delivery. In order to image the receptors, living cells were exposed to an EGFR-binding aptamer (a sequence of nucleic acids that selectively recognize a target thanks to their unique folding), labelled with a dye and diffusing in solution at nanomolar concentration. Following the principle of PAINT imaging, single-molecule localizations were associated to the transient interaction between aptamer and EGFR occurred. A non-EGFR specific aptamer was employed as control to check whether signal is associated to specific aptamer-EGFR interactions or to non-specific unwanted interactions. [12] The three cell lines are compared, A431, a human epidermal carcinoma line showing high over-expression of EGFR of the membrane; and two breast cancer cell lines: MDA-MB-231, showing moderate EGFR expression, and MCF-7, characterized by low EGFR expression.

The code process consists in the selection of a big ROI in the bright field image of the cell. This ROI is selected to roughly cover the cell membrane area, and therefore select all the localizations on the membrane. Due to the heterogeneity of the expression levels between cells, we analyse five samples of each cell lines obtained with the EGFR aptamer and five samples obtained with control aptamer. Figure 5 shows one sample of each line, where selected localizations, roughly corresponding to the area covered by the cell basal membrane are in red, while excluded localizations, mainly resulting from non-specific interaction of the aptamer with glass coverslip surface are in green.

A similar analysis was already performed previously by means of a tedious procedure with the NIS-elements software (Nikon) that allowed manual selection of multiple squared, instead of polygonal, ROIs and without using the bright field image as reference. In this way, a comparison between the procedures can be done. Remarkably, the analysis of one image with our code, takes approximately 20 seconds, while counting localizations manually using multiple square ROIs approach, which is the only option possible in many analysis softwares, requires typically 5 minutes. Additionally, such softwares often do not allow the overlap of pixel-based bright field images with SMLM localizations, impairing the quality of the selection.

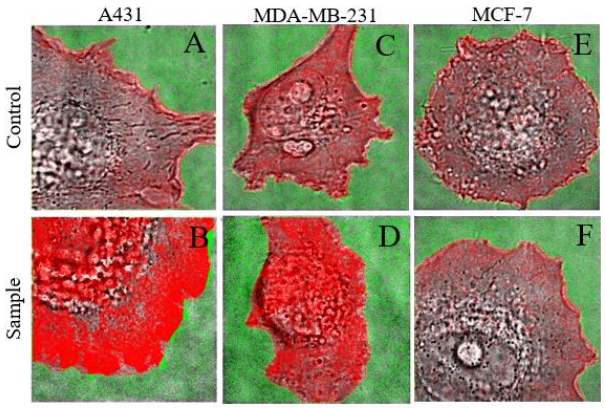

FIG. 5: A/C/E) Cells of control, we can see that the density is similar inside and outside the border of the cell, and very similar between cells. B/D/F) Low-affinity MinE07 cell, clearly in A431 the density of EGFR is very impressive, in MDA-MB-231 this density is significant but not as much as in A431 and in MCF-7 we didn't see a substantial change.

The quantification of all images, thirty in total, take us around one hour and made it possible to analyse a complete set of experiments in a shorter time and more accurately.

The average results for the localization densities on the different cell lines are shown in Table 1. A431 cells exposed to EGFR-aptamer probes displayed a significant localization density related to the very high EGFR expression on the membrane. The retrieved localization density was three times bigger than the level obtained with control aptamers. MDA-MB-231 cells exposed to the same concentration of EGFR-aptamer displayed a significant density due to EGFR on the membrane, but not clearly distinguishable than the value obtained with control aptamer. In the last cell line, MCF-7, we noticed approximately the same localization density for EGFR and control aptamers, or outside the membrane. This result can be explained with the very low EGFR expression level of these cells.

Specimen type	Membrane localization density EGFR-aptamer (loc./μm^2)	Membrane localization density Control-aptamer (loc./ μm^2)	Glass density (loc./ μm^2)
A431	512.82	183.85	197.54
MDA-MB-231	259.96	253.81	145.06
MCF-7	172.17	162.29	154.62

TABLE I: Results of our quantification of PAINT localization densities for different cells lines and using different aptamer probes.

The results were then compared with the ones obtained by previous analysis. For a better relative comparison of the results obtained on the different cell lines, we normalized the data to density values obtained on A431, which is the cell line presenting highest signal. Data are very consistent with those obtained with manual analysis as we observe in Figure 6.

Once we had the normalization, we ran a statistical test to check data between new and old analysis for each cell line. The test we chose was the Mann-Whitney-Wilcoxon test, which is non-parametric. For each line, the test resulted positive, what means that both analysis (old-new) had the same distribution.

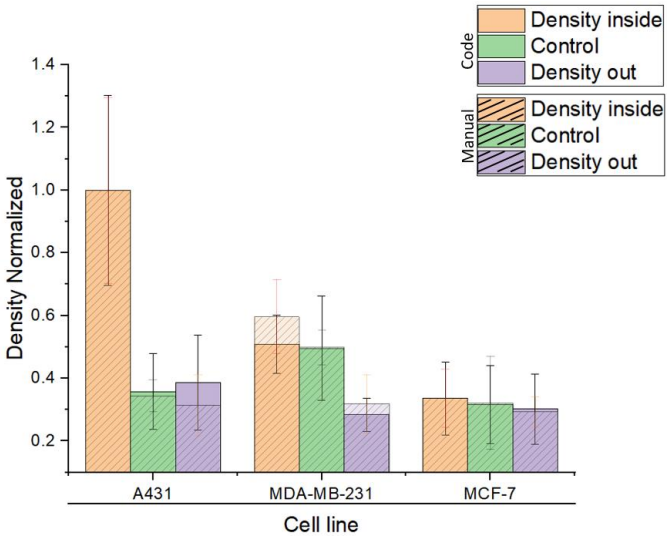

FIG. 6: Histogram showing both quantifications, we can see that we have a little discrepancy on MDA-MB-231 but A431 and MCF-7 had a quite good correlation. The error bars are calculated using standard deviation.

D. Multichannel analysis and time-traces

In DNA-PAINT imaging, it is common to work analysing multiple proteins using different labels on the same sample. One of the advantages of this technique is indeed the possibility of imaging multiple targets, e.g. proteins, on the same cell. This is done by labelling each target with a different docking DNA strand, having a different nucleobases sequence. Then, the sample is exposed sequentially to a solution with the imager complementary to each of the docking. This is typically done using a micro-fluidic device that allows exchanging the imager solutions without removing the sample from the microscope holder. [13]

In order to analyse this type of data, we extended the code to allow the quantification of SMLM localizations in different channels, located within the same ROI that is previously selected on a pixel-based image.

Our code prepared to analyse one channel was rewritten to receive two or three channels and do the quantification of the ROIs selected on all channels at the same time. Inputs needed for the analyses were x-y-t coordinates of each channel and the same bright field image.

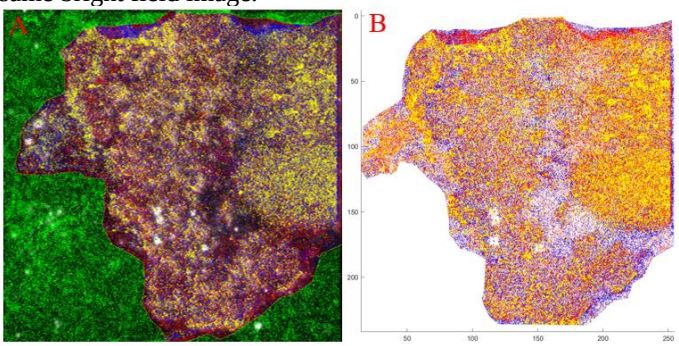

FIG. 7: A) Display of the three channels inside the cell in red, blue and yellow, green points are the outside points. We can notice the movement of the cell in each acquisition looking the density of each channel on the borders. B) Display only of the three channels, the coordinates are in pixels, where you can appreciate better the moving of the sample.

Additionally, we implemented an analysis of time traces of each channel. The time trace represents the detection of a

localization in every frame of acquisition. [14] Under ideal conditions, only one or none events are detected in each frame. Using this binary time-trace, we can determine the frequency of binding and unbinding for each docking-imager pair and ultimately count the molecules. For these reasons, to have a tool that generates a time-trace for each channel within a specific ROI is crucial in this type of measurements. This type of analysis was easily implemented within our code and a representative result is shown in Figure 8. Here, a small ROI corresponding to emission generated by a bunch of proteins, was selected in a low-resolution image and the localizations within the same area are selected the three channels, corresponding to the three viral proteins. After selection, the time traces for each of the channels are reported, which constitutes the information needed for molecule counting.

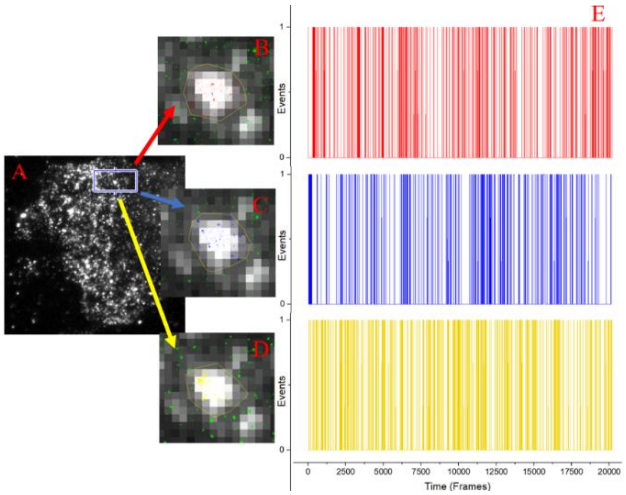

FIG. 8: A) Low-resolution image of the cell with a square on the region analysed. B-D) Selection of the same region of the cell but in each image there is a different channel. The correspondence is channel one to B, two to C and three to D. E) Time trace of each channel, one in red, two in blue and three in yellow, representing each event on each frame, there are an amount of around 2000 events on 20000 frames.

IV. DISCUSSION

Our project provided us coherent data when compared with the data already calculated using a previous analysis. In addition, we achieved the purpose to do a single software for alignment and overlap of pixel-based image and SMLM coordinates, all integrated in a simple MATLAB environment. This tool potentially makes quantification more reliable and easy to control, also due to the intuitive user interface and the limited number of inputs.

Moreover, we had some limitations, these results were obtained using data acquired using different SMLM techniques but using the same software and microscope for single-molecules localizations, it would be needed to use different systems to prove the reliability of the program.

For future prospects, it would be interesting to integrate in the program some tools for segmentation analysis that automatically recognize specific ROI.

V. CONCLUSIONS

This function has proved to obtain reliable results and accomplish the formulated purposes. We used the program to analyse SMLM data of EGFR expression level. Not only the results were very consistent with a previous analysis but also they were obtained with a considerably shorter time, opening new possibilities for thought screening. Moreover, all the method was developed in a single MATLAB code, with user interface, which makes the process of quantification accessible also to non-expert MATLAB users.

In addition, we confirmed the suitability of the code for different purposes on different kind of cell structures as we analysed images of nanoparticles, endosomes and aptamers with positive and reliable results.

Also, we had results that proves we can use the adaptation of the code for multichannel DNA-PAINT samples and the remarkable possibility of obtaining time-traces for multiple channels within the same ROI, potentially allowing counting of multiple types of molecules in a single analysis.

ACKNOWLEDGMENTS

I would like to thank my advisor Silvia and my supervisor Pietro for his great guidance on this project, in brief, for being super advisors, as the entire Nanoscopy for nanomedicine IBEC group. Also thanks a lot to my parents, sisters and all my friends of the degree, without them I wouldn't be who I am.

[1] Schermelleh L, Ferrand A, Huser T, et al. NATURE CELL BIOLOGY. **21**(1):72, (2019).

[2] Delcanale P, Miret-Ontiveros B, Arista-Romero M, Pujals S, Albertazzi L. ACS NANO. **12**(8):7629, (2018).

[3] Deschout H, Zanacchi FC, Mlodzianoski M, et al. NATURE METHODS. **11**(3):253, (2014).

[4] Peters R, Muniz MB, Griffie J, et al. BIOINFORMATICS. **33**(11):1703, (2017).

[5] Ovesný M, Křížek P, Borkovec J, Švindrych Z, Hagen GM. BIOINFORMATICS. **30**(16):2389, (2014).

[6] Stein J, Stehr F, Schueler P, et al. NANO LETTERS. **19**(11):8182, (2019).

[7] Jungmann R, Steinhauer C, Scheible M, et al. NANO LETTERS. **10**(11):4756, (2010).

[8] Jungmann R, Avendano MS, Dai M, et al. NATURE METHODS. **13**(5):439, (2016).

[9] Baker MAB, Nieves DJ, Hilzenrat G, et al. NANOSCALE. **11**(26):12460, (2019).

[10] Taïeb D, Jha A, Treglia G, Pacak K. ENDOCRINE-RELATED CANCER. **26**(11), (2019).

[11] Avutu V. Avidity effects of MinE07, University of Texas at Austin. (2010).

[12] Q.Ashton Acton, Cancer: New Insights for the Healthcare Professional, Atlanta, Scholarly Editions. (2010).

[13] Agasti SS, Wang Y, Schueder F, et al. CHEMICAL SCIENCE. **8**(4):3080, (2017).

[14] Grußmayer KS, Yserentant K, Herten D-P. Cornell University (2018).