

# ColiCoords: A Python package for the analysis of bacterial fluorescence microscopy data

Jochem H. Smit<sup>1,\*</sup>, Yichen Li<sup>1</sup>, Eliza M. Warszawik<sup>1</sup>, Andreas Herrmann<sup>1,3</sup>, and Thorben Cordes<sup>1,2</sup> 

<sup>1</sup>Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

<sup>2</sup>Physical and Synthetic Biology, Faculty of Biology, Ludwig-Maximilians-Universität München, Großhadernerstr. 2-4, 82152 Planegg-Martinsried, Germany

<sup>3</sup>Institute of Technical and Macromolecular Chemistry, RWTH Aachen University, Worringervogel 2, 52074, Aachen, Germany

\* Current affiliation: Department of Microbiology and Immunology, Rega Institute for Medical Research, Laboratory for Molecular Bacteriology, KU Leuven, Herestraat 49, Gasthuisberg Campus, B-3000 Leuven, Belgium

## Abstract

Single-molecule fluorescence microscopy studies of bacteria provide unique insights into the mechanisms of cellular processes and protein machineries in ways that are unrivalled by any other technique. With the cost of microscopes dropping and the availability of fully automated microscopes, the volume of microscopy data produced has increased tremendously. These developments have moved the bottleneck of throughput from image acquisition and sample preparation to data analysis. Furthermore, requirements for analysis procedures have become more stringent given the requirement of various journals to make data and analysis procedures available. To address this we have developed a new data analysis package for analysis of fluorescence microscopy data of rod-like cells. Our software ColiCoords structures microscopy data at the single-cell level and implements a coordinate system describing each cell. This allows for the transformation of Cartesian coordinates of both cellular images (e.g. from transmission light or fluorescence microscopy) and single-molecule localization microscopy (SMLM) data to cellular coordinates. Using this transformation, many cells can be combined to increase the statistical significance of fluorescence microscopy datasets of any kind. ColiCoords is open source, implemented in the programming language Python, and is extensively documented. This allows for modifications for specific needs or to inspect and publish data analysis procedures. By providing a format that allows for easy sharing of code and associated data, we intend to promote open and reproducible research.

The source code and documentation can be found via the project's [GitHub](#) page.

Open Source Software | Fluorescence Microscopy | Super-Resolution Microscopy | Bacterial Imaging

Correspondence: [jhsmit@gmail.com](mailto:jhsmit@gmail.com), [cordes@bio.lmu.de](mailto:cordes@bio.lmu.de)

## Introduction

Fluorescence microscopy has become a crucial tool in studying bacterial cell biology(1–8). It is minimally invasive and allows for the study of living bacteria in a controlled environment and to monitor the motion and sub-cellular topologies of any proteinaceous factor(9–24) or nucleic acids(25–28). Through the multitude of genetically programmable fluorescent protein probes(29–36) and commercially available dyes(37–41) with different conjugation capabilities(42–46), virtually all components of the bacterial machinery can be studied with high specificity and spatio-temporal resolution. Even the detection of single fluorescent probes has become a

routine experiment, revealing both dynamics and heterogeneity.

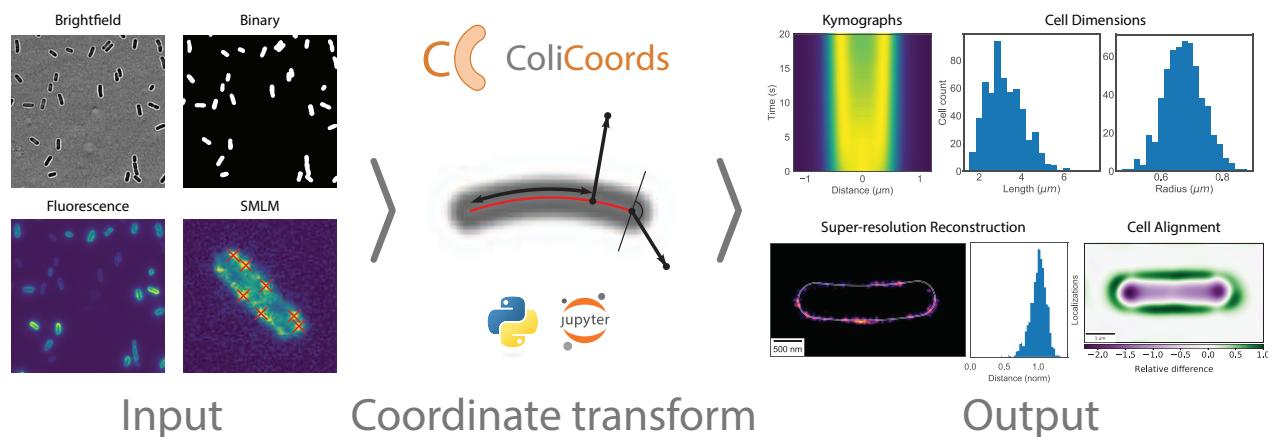
However, with the introduction of fully automated(47) and autonomously operating microscopes(48, 49), images can be acquired independently of experimentalist interference at high acquisition rates(50–52). A typical experiment conducted overnight spanning the dimensions of position, time and channels can easily generate thousands of images. This explosion of available multidimensional data has made analysis the new bottleneck in single-cell fluorescence microscopy studies. Furthermore, in the spirit of reproducibility and open availability of scientific results, more and more peer-reviewed journals require deposition of source data and evaluation methods – a process that is hampered by non-standardized evaluation routines and lab-specific software, often based on commercial platforms.

The exception is current options for bacterial image analysis such as SuperSegger(62, 63), Oufti(64) and MicrobeJ(65). Although these data analysis tools are able to tackle a lot of the problems inherent to (bacterial) live-cell image analysis, we have identified several drawbacks which we aimed to address with ColiCoords. The main points are related to the platform, structure and philosophy of the data analysis routine rather than the functionality for analysing images.

First, with available tools, it is difficult for users themselves to either inspect the exact mathematical procedures applied to their data or customize them. The source code of these analysis packages are freely available and licensed under the GNU General Public License. Modification of Oufti and SuperSegger, however, require a Matlab license, which is not available in many institutions. MicrobeJ, on the other hand, can be freely modified, but users face a daunting task due to the sheer size of the project and the lack of docstrings.

ColiCoords is written in the freely available language Python and the source code has been released under the MIT license. The code is available on GitHub and is extensively documented with both docstrings, online [documentation](#) and its basic principles are described in this paper. Being hosted on GitHub together with Continuous Integration for testing makes ColiCoords an ideal platform for other users to contribute and modify its code.

Second, most other currently available analysis options based on a graphical user interface (GUI). We would argue here that GUIs are inherently limiting and obfuscate the exact data analysis procedure. While GUI-based analysis is intuitive it



**Fig. 1.** Overview of workflow pipeline using ColiCoords. Input data is either image data or sparse data (localizations). Input images need to be segmented to identify cell location and orientation. Third party options for image segmentation include *lilastik*(53), *CellProfiler*(54, 55) or *Keras*(56)/*Tensorflow*(57). SMLM data have to be reconstructed by external software such as DAOSTORM(58, 59), ThunderSTORM(60), QuickPALM(61) or others prior to use. ColiCoords can then be used to transform the Cartesian coordinates of the input data to cellular coordinates. The transformed data can be generated output graphs, such as kymographs, histograms of the cell's dimensions, axial distributions or to align cells.

limits the use to the available graphical elements, with little flexibility. Importantly, due to the large number of permutations to the order of operations that can be executed via a GUI, the exact data analysis procedure is typically not documented in detail and thus hard to reproduce.

In contrast, ColiCoords offers a basic foundation for single cell data analysis where users can apply all functionality from Python itself, the SciPy(66) ecosystem and community developed image analysis procedures. The workflow of ColiCoords is based on Jupyter Notebooks(67, 68), which act as a lab journal page, where all steps of data analysis and its associated parameters can be executed and described and results are plotted in interactive graphs. This type of interactive analysis workflow is being adopted rapidly in the scientific community(69). An example from the microscopy community is the smFRET analysis package FRETBursts(70).

In addition, ColiCoords is open-ended, meaning it can be used in a pipeline of analysis where users can freely choose pre- and post-processing options (Figure 1). The input data can be either image-based (brightfield, phase contrast, fluorescence) or sparse data (super-resolution localizations). Users are therefore free to choose preprocessing options for tasks such as segmentation(53, 54, 56, 57, 64) and super-resolution reconstruction(58–61). Colicoords is out of the box compatible with High-performance computing (HPC) as operations on a per-cell basis can be computed in parallel.

The central feature of bacterial image analysis in ColiCoords is transforming input Cartesian coordinates (image pixel coordinates or super-resolution localizations) to cellular coordinates. The coordinate system can be derived from different sources, including binary images, brightfield images or even super-resolution membrane markers. With the coordinate system in place, ColiCoords offers several different visualization and analysis methods, including alignment of cells, radial, longitudinal or angular (poles) distributions, kymographs and super-resolution reconstruction. (Figure 1).

## ColiCoords principles and features

**A. Coordinate definitions.** The general principle of the coordinate system is shown in Figure 2. Here, a brightfield image of an *E. coli* cell is shown with the coordinate system overlaid. By establishing a per-cell coordinate system every pair of Cartesian coordinates  $(x_p, y_p)$  from either pixels or fluorescent foci can be mapped to cell coordinates  $(l_c, r_c, \phi)$ ; Figure 2A. The transformation is based on a second-degree polynomial  $p(x)$ ,  $x_l \leq x \leq x_r$  (red line) which corresponds to the midline of the cell. The parameters describing  $p(x)$  are initially based on guess values which can be refined after. To transform a point  $(x_p, y_p)$ , the first step is to find the point  $(x_c, y_c)$  on  $p(x)$  which is closest to  $(x_p, y_p)$ , perpendicular to  $p(x)$ . This is done by minimizing the squared distance between both points:

$$r^2 = (x_c - x_p)^2 + (p(x_c) - y_p)^2 \quad (1)$$

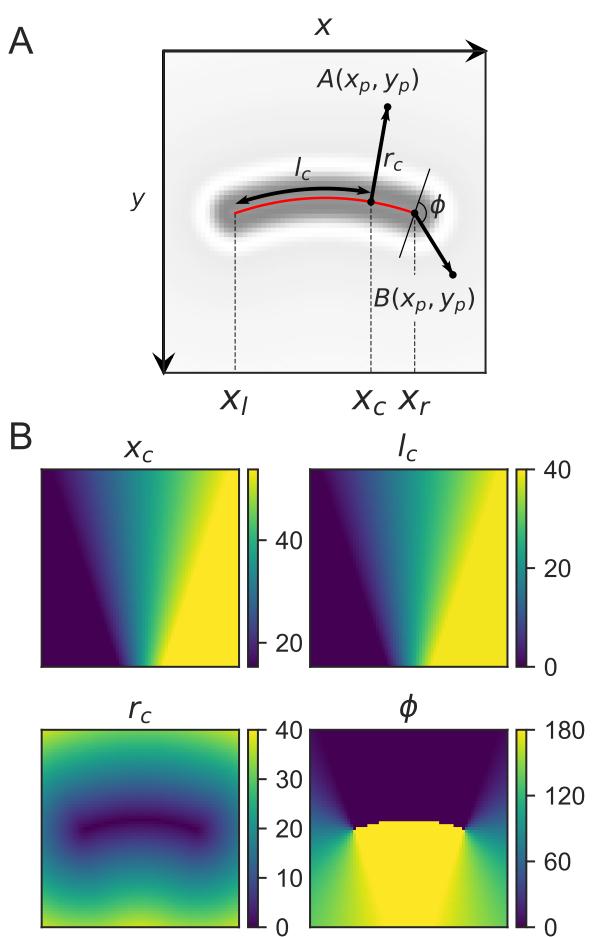
$$\frac{\partial r^2}{\partial x_c} = 0 \quad (2)$$

Solving equation 2 gives a cubic equation which is solved analytically to ensure fast transformation of many data points. To account for points B at the poles of the cells the coordinate  $x_c$  is restricted to the domain  $[x_l, x_r]$ . The cellular coordinates  $(l_c, r_c, \phi)$  can then be calculated from  $x_c$  and  $p(x)$ . The longitudinal coordinate  $l_c$  is given by the arc length along  $p(x)$  from  $x_l$  to  $x_c$ :

$$\int_{x_l}^{x_c} \sqrt{1 + \left( \frac{dp(x)}{dx} \right)^2} dx \quad (3)$$

This expression also gives the full length of the cell when  $x_c$  is substituted by  $x_r$ . The radial coordinate  $r_c$  is simply the euclidean distance between  $(x_c, p(x_c))$  and  $(x_p, y_p)$ :

$$r = \sqrt{(x_c - x_p)^2 + (p(x_c) - y_p)^2} \quad (4)$$



**Fig. 2.** General description of the in-cell coordinate system. **A** Brightfield image of an *E. coli* cell with coordinate system overlayed. Every point with coordinates  $x_p, y_p$  can be transformed to coordinates  $l_c, r_c, \phi$ . **B** Images showing the values of  $x_c$  as well as cellular coordinates values  $l_c, r_c, \phi$ .

The third coordinate is the angle  $\phi$  (in degrees) and it uniquely defines points A along the body of the cell by distinguishing between the top and bottom of the cell, as well as defining the position of points B at the poles. The value of  $\phi$  is 0 at the top of the cell runs from 0 to 180 along the right pole, where  $\phi$  is defined as the angle between the line perpendicular to  $p(x)$  at  $x_r$  and the line from  $(x_c, p(x_c))$  to  $(x_p, y_p)$ . For points below the midline of the cell the value of  $\phi$  is 180, which then runs from 180 back to 0 along the left pole.

Note that the top area of the cell, as it is displayed in Panel A of Figure 2, is given by  $y_p < p(x_p)$ , since the origin of the Cartesian coordinates is at the top left, where the positive y-axis is running down. By this definition, the coordinate of the midpoint of the top-left pixel is (0.5, 0.5), in line with the coordinate definitions in the image analysis software ImageJ(71, 72).

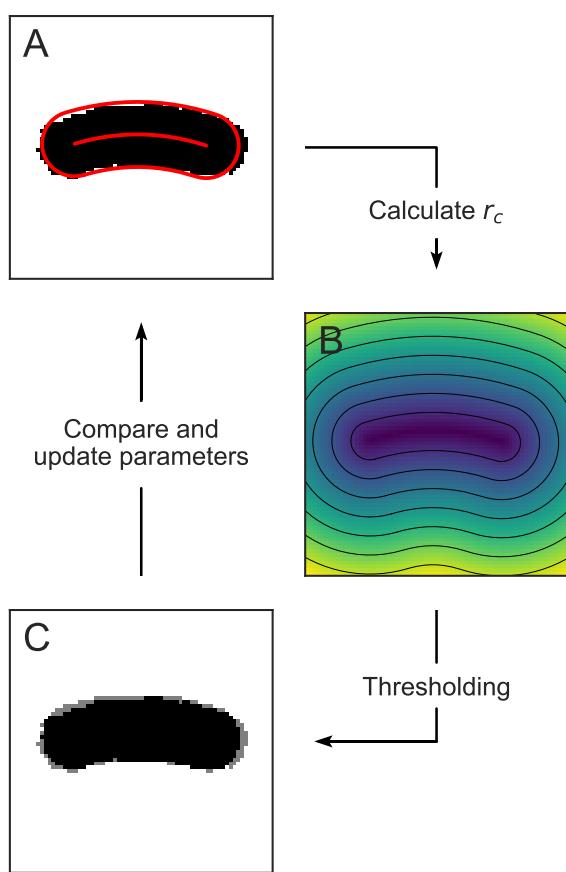
**B. Preprocessing and optimization.** To analyse fluorescence microscopy data with ColiCoords some preprocessing of the input data is required. Notably, ColiCoords is not designed to segment or detect cells in the raw microscopy images. Binary (labelled) input images are required for Coli-

Coords processing to identify the location of cells in the images. However, these binaries are only needed to localize cells and initialize the coordinate system, and they are not needed to form the final optimized coordinate system, consequently a high accuracy of segmentation is not required. Binary or segmented images can be derived from brightfield, differential interference contrast (DIC), phase contrast or fluorescence images by established tools such as CellProfiler (54), Ilastik(53), Oufti(64) or Convolutional Neural Networks (e.g. implemented by TensorFlow(57) or Keras(56)). An implementation of the U-Net segmentation architecture(73) is included in ColiCoords together with examples on how to use it.

Fluorescence images should be (optionally) background or illumination corrected and different channels such as brightfield or DIC should be aligned if they are acquired with different optics. ColiCoords allows the processing of any kind of image data as well as sparse data, such as single-molecule tracking and single-molecule localization microscopy (SMLM) data. Based on the provided binary images, single-cells are automatically cut out from all image data, the positions of the localizations are selected and transferred to the coordinates of the cropped image. The orientation of the cell is calculated using the binary image by calculating the image moments(74). All data elements are then rotated to orient the cells horizontally. The result of this process is a collection of single-cell Python-objects where all data are organized on a cell-by-cell basis, and every cell has its own coordinate system which can be used to perform calculations, analysis or visualizations on the data element of choice. These sets of cells can then be indexed and selected based on user-defined criteria (shape, size, fluorescent signal etc.) and further analysed or inspected through interactive plots within the Jupyter Notebook environment.

After initialization of the cell objects, the coordinate system needs to be refined to allow it to more accurately describe the cell's shape. ColiCoords implements the optimization of the coordinate system based on different data sources by providing several objective functions which, when minimized, give the best matching coordinate system for the respective data source. Any image-type data (e.g. brightfield) can be used provided there is an anisotropic signal along the length of the cell, as well as localization (SMLM) data provided the localizations describe the outline of the cell.

Figure 3 shows the iterative optimization process. In Panel A the ground truth binary image (derived from measurement) is shown together with the cell's coordinate system. The coordinate system is described by a total of 6 parameters, 3 coefficients  $a_0, a_1, a_2$  which define the second degree polynomial  $p(x)$ , the left and right endpoints  $x_l$  and  $x_r$ , and the cell's radius  $r$ . In the figure, the polynomial  $p(x)$  is shown together with the isodistance line from  $p(x)$  with distance  $r$ . As can be seen from the figure, the initial coordinate system does not accurately describe the cell's shape. To refine the coordinate system, first a radial distance image is calculated (Panel B). This image is then thresholded with the initial guess value of  $r$ , which gives a new binary image (Panel C, black). Compar-

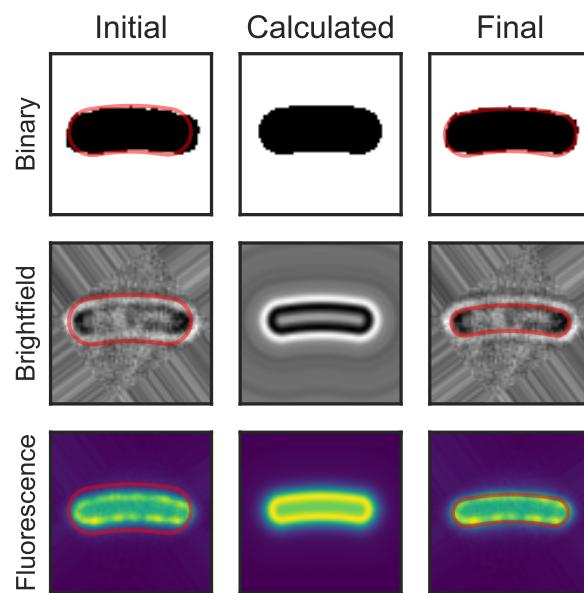


**Fig. 3.** Optimization of the cell's internal coordinate system. **A** Ground-truth binary image with cell midline and outline calculated from initial guess parameters. **B** Radial distance image calculated from initial guess coordinates. **C** Calculated binary image (black) obtained by thresholding the distance image superimposed on the ground truth image (grey).

ison of this image with the ground truth binary image gives the  $\chi^2$  value, which is minimized to optimize the fit of the coordinate system. The fitting in ColiCoords is done via the package Symfit(75), which provides an API to the minimizers implemented by `scipy.optimize`(66).

Other microscopy images can be used to optimize the coordinate system, where the only constraint is that the signal needs to be isotropic along the length of the cell. This is shown for binary, brightfield and fluorescence images in Figure 4. The figure shows the three different channels measured for a single cell. The binary is derived from the brightfield and is used for initial guesses of the coordinate system. This initial guess coordinate system is shown in red in the left column. This coordinate system is used to calculate a synthetic image, averaged along the angular coordinate (middle). A comparison between this image and the measurement gives a measure for how well the current parameters describe the shape of the cell. The optimal parameters can be found via iterative optimization (right). For brightfield- or fluorescence data-based optimization the radius  $r$  of the cell is determined in a second step by determining the half-maximum point of pixel intensity.

Finally, ColiCoords can optimize the coordinate system



**Fig. 4.** Iterative optimization of the coordinate system based on binary, brightfield and fluorescence images.

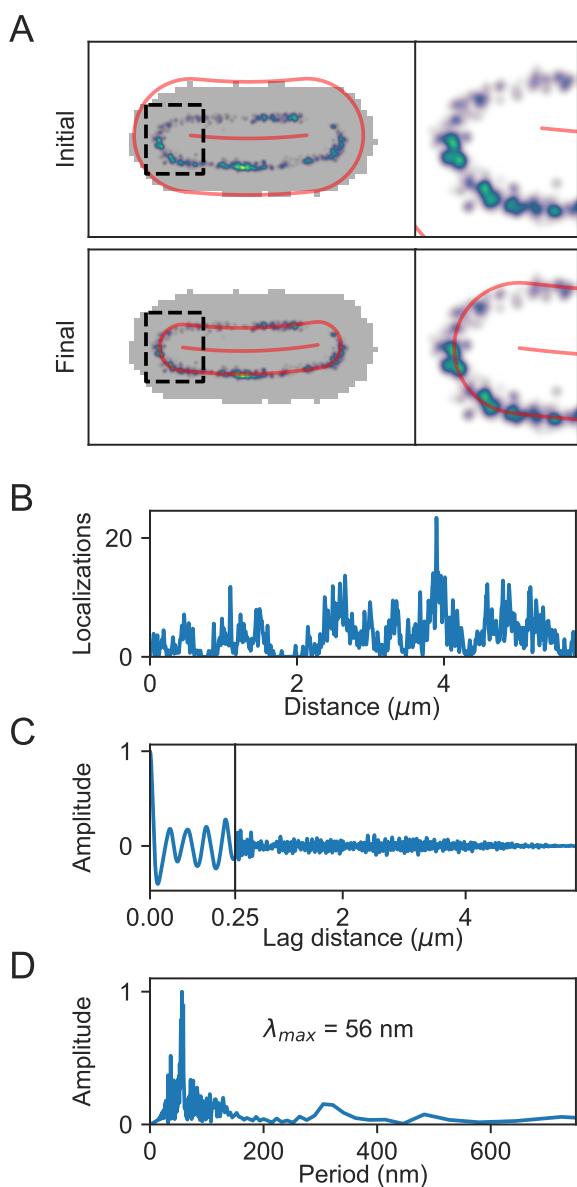
based on SMLM data of a membrane marker. This type of data can be obtained through various super-resolution techniques such as Point Accumulation for Imaging in Nanoscale Topography (PAINT)(27, 76–78), Stochastic Optical Reconstruction Microscopy (STORM)(79, 80) or PhotoActivated Localization Microscopy (PALM)(81, 82) imaging.

The optimization process and a possible application thereof is illustrated in Figure 5. Panel A shows a STORM super-resolution reconstruction of the membrane protein marker LacY-eYFP. In the top panels, the coordinate system based on initial guesses (derived from binary image; grey) is shown. The bottom panel shows the result upon optimization based on the STORM-localizations data. Here, the radial distance  $r_c$  for all localizations is calculated and by comparison to the value for  $r$  the  $\chi^2$  (squared differences) is calculated. Minimization of the  $\chi^2$  gives the parameters where the coordinate system that most appropriately describes the bacterial membrane. Interestingly, the localizations along the membrane show a periodic fine structure. To further characterize this, the position along the perimeter of the cell was plotted in panel B. The zero-point position is defined as the beginning of the ‘top’ part of the membrane ( $l_c = 0, \phi = 0, r_c = r$ , panel A) and the positive directions runs clockwise along the cell outline.

To extract the size of the periodic structures observed, at first a spatial auto-correlation(83) function was calculated and its low-frequency components were subtracted by means of a sliding window, to reveal clear oscillations (panel C). By Fourier transforming this signal (Panel D) we found a periodicity of the signal of 52 nm, which we attribute to an artefact resulting from the YFP fusion, as previously reported for MreB-YFP constructs(84).

In conclusion, optimization based on a SMLM super-resolution membrane marker is expected to provide the most accurate approach for creating a coordinate system within

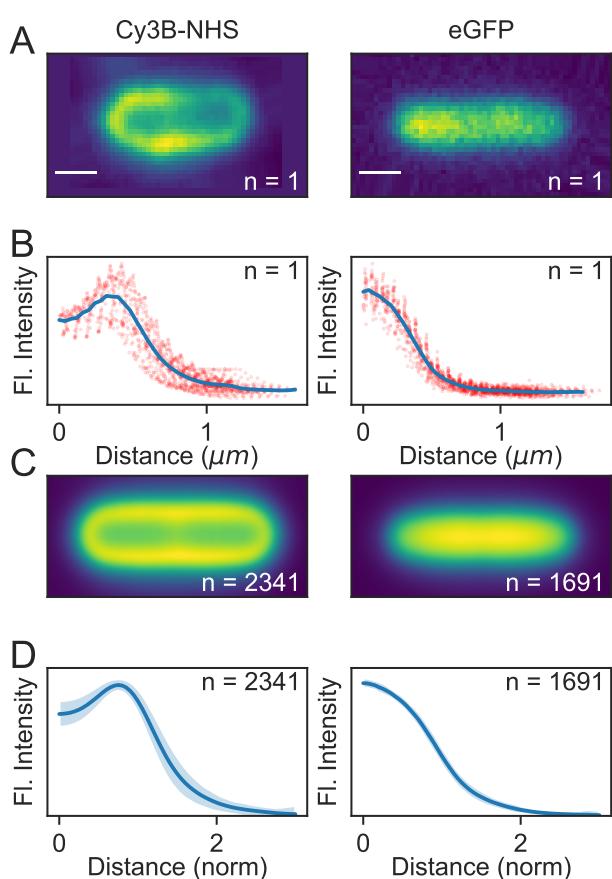
the cell (vide infra, section D). This allows for aligning and combination of multiplexed localization data from many cells with high accuracy.



**Fig. 5.** Iterative optimization of the coordinate system based STORM super-resolution data of the membrane marker LacY-eYFP. **A** Top: STORM reconstruction on top of the *E. coli* binary image (grey) with initial guesses coordinate system in red. Bottom: Coordinate system after optimization based on STORM localizations. **B** STORM reconstruction along the perimeter of the cell showing localizations as a function of distance along the membrane. **C** Spatial autocorrelation function of **B**. **D** Fourier transform of **C** showing the largest amplitude at a periodicity of 56 nm.

**C. Batch processing.** To demonstrate the ability of ColiCoords to analyse a datasets consisting of many cells, we applied a typical analysis procedure to two sets of fluorescence microscopy images. The first set of *E. coli* cells were incubated with the reactive dye Cy3B-NHS to label the outer membrane, while the second set of cells expressed eGFP in the cytosol.

For each condition, a subset of 100 images were manually annotated through the use of a custom GUI element. These an-



**Fig. 6.** Batch processing of several thousand cell objects. Two datasets were analyzed, one where cells are labelled on the outer membrane with Cy3B-NHS and one with cell expressing eGFP in the cytosol. **A** Fluorescence images of single *E. coli* cells, scale bar 750 nm. **B** Radial distribution profiles for the cells in **A**. Individual datapoints from every pixel in the image are shown as red points, together with the resulting radial distribution (blue line). **C** Aligned and average fluorescence images for  $n=2341$  and  $n=1691$  cells, respectively. **D** Average radial distribution profiles. The standard deviation is shown as a shaded region.

notated binary images were then combined and used to train a convolutional neural network. An implementation of the U-Net architecture(73) neural network was used, based on the Keras(56) API using Tensorflow(57) as backend.

The brightfield images were first scaled down to 256x256 to preserve graphics card memory. A custom implementation of Keras' Sequence was used for preprocessing and augmentation. All input images were normalized using tanh-estimators(85) and augmented 8-fold through permutations of horizontal mirroring, vertical mirroring and transposing the images. This augmented data was randomly split into a set validation data and training data.

After training the neural network, the network was applied to the whole set of images to generate binary images. The images were then scaled back to 512x512 pixels. Next, binary shapes were filtered from the segmentation mask by their shape and size. The resulting binary images were used as input for ColiCoords, together with the corresponding brightfield and fluorescence images. The cell objects were filtered based on the results of optimizations by binary images and brightfield images as well as a measurement of the cell's radius from the brightfield image. This yielded a total of 2341

and 1691 cells for the Cy3B and eGFP datasets, respectively. The analysis output is shown in Figure 6. In panel **A**, two individual cells for each condition are shown. Their radial distribution profile is shown in panel **B** (blue line), together with the individual datapoints from each pixel as red points. In panel **C**, all cells from each dataset were aligned by transforming all pixel coordinates in the fluorescence images from each cell to cellular coordinates. These cellular coordinates are then transformed back onto a model cell with user-specified dimensions. The resulting point cloud is then used to generate an aligned fluorescence image by convolution with a 2D gaussian. The final image is displayed with spline interpolation. In panel **D**, the average radial distribution for all cells is shown. The curves are individually normalized in the x-direction by setting the brightfield-measured radius to 1 whereas the y-direction is normalized by the maximum of each curve. The shaded region (standard deviation) therefore only reports on variations in the shape of the radial distribution profile.

**D. Synthetic benchmarks.** In order to evaluate ColiCoords' performance a synthetic dataset was generated and subsequently analyzed. By comparing the results for different conditions and optimization methods the accuracy can be benchmarked.

First, a set of 10000 synthetic cells were generated using ColiCoords' synthetic\_data module. The geometric parameters describing these synthetic cells were chosen based on a set of measured *E. coli* cells. These synthetic cells have a binary image data element by default. A brightfield data element was added based on normalized brightfield radial distributions measured from *E. coli*. The ratio between the background intensity (no cells) of the brightfield and the maximum of the scattered light from the bacterial membrane was measured and kept constant to replicate realistic signal-to-noise ratios. The background light was set to one and then multiplied by the chosen photon count number (500, 1000 and 10000), after which the resulting values were drawn from a Poisson distribution to simulate shot noise. Finally, normally distributed noise was added with a standard deviation of 20 photon counts.

To each cell two homogeneously distributed membrane-localized STORM data elements were added. The first data element describes localizations at the inner membrane. The mean radial distance of the localizations was set according to where the inner membrane should be with respect to the radius of the cell measured from the brightfield image based on measurements with the construct LacY-eYFP. The standard deviation from the mean radius was set to 0.25 pixels (20 nm). A second STORM data element was added 100 nm further out with the same standard deviation.

From this set of synthetic cells, images were reconstructed. An average of  $10 \pm 3$  cells were taken from the set and randomly rotated and placed in each image of 512x512 pixels ( $40 \mu\text{m} \times 40 \mu\text{m}$ ), ensuring at least 5 pixels distance between the cells' binary images. The coordinates of the cell's STORM data element were combined into one big STORM localizations table.

These processes yielded 1000 ground-truth binary images together with 3 sets of 1000 brightfield images with different photon numbers (500, 1000, 10000 per pixel) and therefore different signal-to-noise ratios (Example images in Figure 7A). The whole ColiCoords' data analysis pipeline was performed on this dataset to benchmark its performance.

First, the brightfield images were segmented. For all 3 signal-to-noise conditions, the first 400 images were used to train the neural network. After training, the network was applied to the whole set of 1000 images to generate binary images.

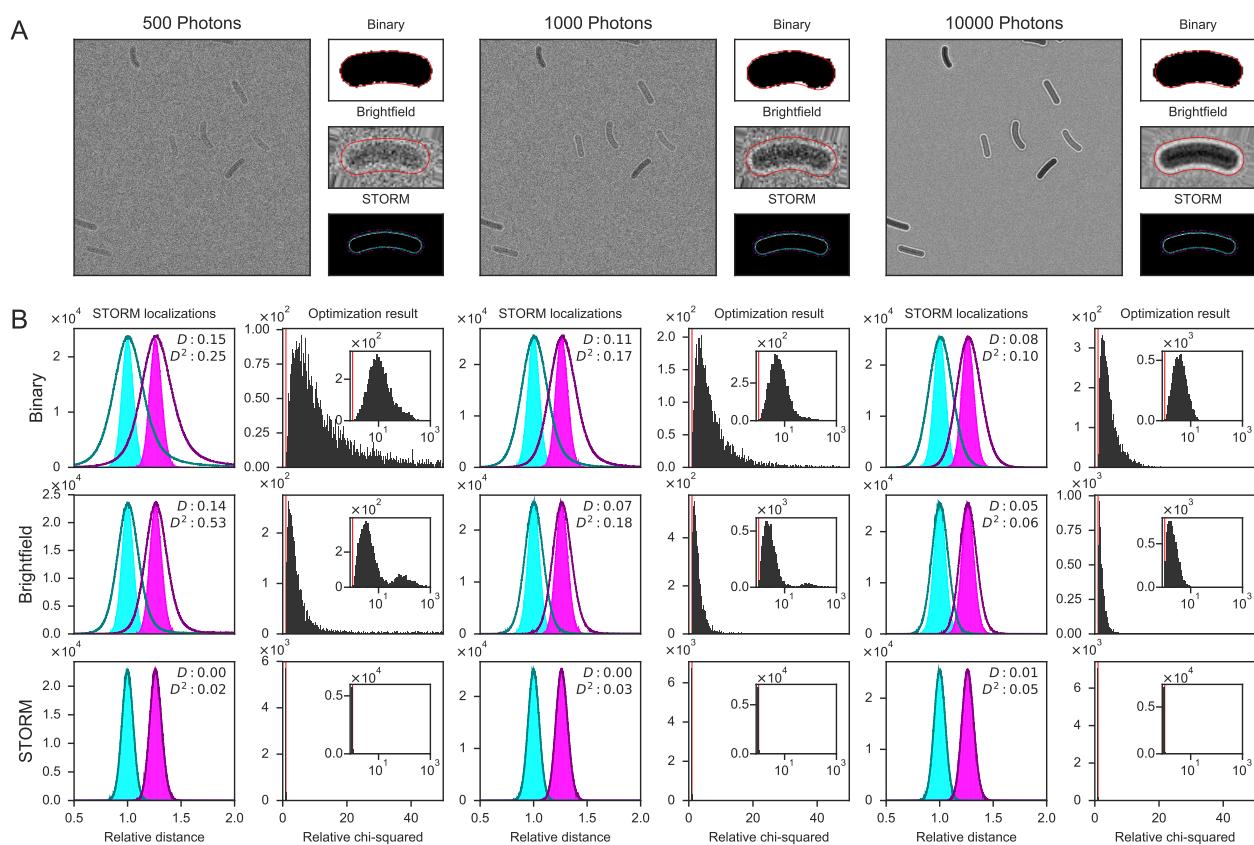
Next, the binary images were filtered. Bordering cells were removed and the remaining binary objects were selected based on their size and ellipticity where the selection criteria were obtained from the ground-truth binary images. The filtered and labelled binary images were then used by ColiCoords together with the brightfield and STORM data to generate a set of 'measured' cell objects.

The 'measured' cell objects obtained are correlated to the original ground-truth cell objects by first performing a cluster analysis on the input STORM table(86). The ground-truth cells are then identified by encoding the combined STORM intensity and comparing to the ground-truth intensity of the STORM clusters. The corresponding measured cell was then found by comparing the centre-of-mass positions of the STORM clusters and the filtered binary images used as input. This yielded two sets of cells, one with ground-truth cells and one with corresponding 'measured' cells for comparison.

Cells which had too few STORM localizations compared to their ground-truth counterparts were discarded due to them being too poorly segmented. Several cells had too many STORM localizations due to localizations from neighbouring cells. These additional localizations were also removed. This yielded a total of 6569, 7129 and 7245 cells out of maximum 7245 cells for brightfield images with 500, 1000 and 10000 photons, respectively. This number is lower for input images with lower signal-to-noise because some cells are segmented poorly by the neural network.

Finally, the coordinate system of the cells was optimized using different approaches to compare their accuracy. The cells in each condition were initialized with guesses derived from segmented binary images and subsequently optimized based on the data elements binary, brightfield and storm (inner membrane). This yielded a total of 9 sets of cells which are summarized in panel **B** of Figure 7. In the left column of each condition the relative radial distances of all STORM localizations in the dataset are plotted as calculated by its coordinate system. The radial distances are normalized to the ground truth inner membrane distance to eliminate cell-to-cell radius variations. The ground truth radial distances of the inner (cyan) and outer (magenta) are shown in light colours and filled histogram. The calculated radial distances are shown on top with dark colours where the outer line of the histogram is shown only.

From these graphs the effect of the accuracy of the coordinate system can be shown on the resulting radial distribution graph. As can be seen, in the condition with the lowest number of photons (500) and a coordinate system optimized on



**Fig. 7.** Benchmarking of ColiCoords software with a synthetic dataset. **A** Examples of generated input data brightfield images for different photon counts (500, 1000 and 10000 average per pixel) and example of one processed single cell. Different data elements (Binary, Brightfield, STORM) are shown together with the outline of the coordinate system (red/white line) where the coordinate system is optimized based on that data element. **B** Evaluation of the coordinate system of  $n=6569, 7127, 7245$  cells out of 7245. Left: relative radial distance of all inner (cyan) and outer (magenta) membrane STORM localizations. The radial position is normalized to the ground-truth inner membrane position. The ground-truth radial distance (light colours, filled histogram) is compared with radial distances calculated with the coordinate system as calculated for the different data elements (binary, brightfield, STORM) for different brightfield image photon counts (dark colours, line only). The absolute mean deviation ( $D$ ) and the root mean squared deviation ( $D^2$ ) are shown in the graph. Right: Relative minimization objective function  $\chi^2$ . The  $\chi^2$  is calculated for the obtained coordinate system for each cell for the STORM data element for each condition. The obtained value is divided by the ground-truth  $\chi^2$  value, therefore a value of 1 indicates a perfectly fitted coordinate system (red line).

the binary data element, the resulting radial distribution histogram is significantly wider than the ground-truth radial distribution. Compared to binary images derived from higher signal-to-noise images, the result more accurately describes the underlying ground-truth. This is further reflected by comparing the obtained  $\chi^2$  values from the optimization process. In the histogram on the right (second column) the relative  $\chi^2$  of each cell derived from its optimized coordinate system is plotted. The values are obtained by calculating the  $\chi^2$  value for STORM optimization (mean squared difference in distance between coordinate system outline and STORM localizations) and dividing this value by the ground-truth  $\chi^2$  value to get a relative  $\chi^2$ . This value is an indication for the goodness of the fit, where a value of one indicates a perfect fit (red line in histograms).

As can be seen in the histogram, for 500 photons and optimization based on the binary image (Panel B, top left), the median  $\chi^2$  value of optimization is roughly 10-fold higher compared to the ground-truth values. The inset (log scale x-axis and logarithmically increasing bin size) shows a long tail in the distribution to higher relative  $\chi^2$  values up to 500.

When comparing the binary optimization result for 1000 and

10000 photons, it can be seen that the radial distribution histograms become narrower and more accurately match the ground truth radial distribution. The distribution of relative  $\chi^2$  values also reflect this with median values of 5.8 and 3.8, respectively.

Moving one row down the results are plotted for optimization based on the brightfield image. The initial guesses for optimization derived from the binary images and therefore the quality of the segmentation influences the final result. As can be seen from the radial distribution histograms, the coordinate system obtained more accurately describes the cell compared to binary optimization. For 10000 photons the radial distribution graph very closely matches the ground truth. In the relative  $\chi^2$  distributions a second population can be seen which is mostly present for lower photon numbers. This is a consequence of poor initial guesses and the optimization therefore returns a different local minimum. The median relative  $\chi^2$  values are 4.1, 2.6 and 1.9, respectively.

The third row shows the result from optimization based on the inner membrane STORM localization dataset. Since the STORM localizations are identical in all cells and are independent of the brightfield image, any difference in the opti-

mization result is due to the initial guesses. As expected, the obtained radial distribution histograms almost perfectly overlap with the ground truth histograms. The obtained median relative  $\chi^2$  values are all 0.99, however some small outliers are found, where the population sizes of values >50 are 2, 3 and 9 respectively.

In Figure 7 for 10000 photons and brightfield optimization it can be seen that the ground-truth result is matched closely but not completely. This is unexpected since the cells were generated from a perfect second-degree polynomial shape and therefore it was expected that the optimization result would more accurately describe the cell's shape. The  $\chi^2$  value of the obtained result was lower compared to the ground-truth, thus the problem is not that the algorithm returned a local minimum instead of a global minimum. Therefore, switching to a global optimization algorithm such as differential evolution(87) will not yield the desired improved result.

Instead, it was found that in the optimizer's implementation a choice of speed over accuracy is the origin of the discrepancy. Specifically, in the bootstrapping process to calculate a simulated brightfield image to compare with the measured image, the radial distribution of the brightfield image is calculated by normal binning (histogram) compared to convolution with a Gaussian kernel. This choice was made to keep the calculation times manageable, and optimizing of this procedure, possibly through a two step (course and fine) optimization process is expected to yield improved results.

The STORM optimization results (lower row, Figure 7) do closely match the ground-truth result. Any deviation is mostly originating from a small number of cells which do not converge to the correct solution. Since the population is larger for lower photon brightfield images, and thus less accurate segmentation, this is be attributed to poor initial guesses for the coordinate system.

Finally, using an optimizer that is restricted by bounds on the optimization parameters is likely to reduce the subpopulations in both brightfield and STORM optimization. These features and the issues mentioned above will be addressed in a future performance update of ColiCoords.

## Conclusions

ColiCoords is an open-source software package to analyse fluorescence microscopy data of rod-shaped cells. It allows for the transformation of Cartesian coordinates from any data source to cellular coordinates. This transformation can then be used to obtain distributions of fluorescence along the cell long or short axes or its perimeter and align the whole cell to allow for the combination of data from many cells.

The open source nature and Jupyter-notebook based workflow together with an open and compact file format promotes open and reproducible analysis of microscopy data. As illustrated in this paper, ColiCoords facilitates the publication in the form a of a 'reproducible article'(88), as recently featured [eLife](#), where data and code to generate the figures are bundled together with the publication. The code and data to generate the figures in this article is available online (see Methods). ColiCoords' code is well documented and the main features

are illustrated by example notebooks, which can be tested live on [Binder](#). The software itself can be installed directly from both the [Anaconda](#) and [PyPi](#) package managers.

Analysis of synthetic data has shown that we can determine the cellular coordinates of SMLM super-resolution datapoints based on only the brightfield image of the cell with high accuracy (5%). If the coordinate system is derived from super-resolution measurements of a membrane marker, the accuracy is further increased (1%).

Although the current state of ColiCoords is a mature project, future updates are planned to be released, either by us or in collaboration with the community. Three major updates are planned; first a performance update to increase speed and accuracy of the optimization process, second a metadata update allowing the storage of image-associated metadata, and finally an update to allow the organization of cell objects into a hierarchical structure to process lineage and temporal information.

## Methods

**Cy3B-NHS staining** *E. coli* MG1655 cells were grown overnight in LB medium and subsequently diluted 1000 times in EZ rich medium (Teknova) with 0.2% glucose. The cells were grown until the culture reached an OD600 value of 0.3 after which the cells were centrifuged at 3000 rcf for 5 minutes and resuspended in PBS buffer. The amino-reactive dye Cy3B-NHS (GE Healthcare) was added to a final concentration of 1 mM and incubated while protected from light at room temperature for 1 hour(89). Next, the cells were centrifuged at 3000 rcf for 10 minutes and the cell pellet was resuspended in PBS.

**eGFP staining** *E. coli* BL21D3 cells were grown overnight in LB medium and subsequently diluted 1000 times 1000 times in EZ rich medium (Teknova) with 0.2% glucose. The growth media are supplemented with the appropriate antibiotics. When the cell culture reached an OD600 of 0.3 eGFP expression was induced with IPTG at a final concentration of 200  $\mu$ M for 30 minutes at 37 °C.

For imaging, 3  $\mu$ L of the bacterial culture was transferred to a microscope coverslip and covered by an agarose pad(90, 91). Imaging was done on an Olympus IX83 inverted microscope with an Olympus UAPON 100x NA 1.49 TIRF oil immersion objective. The excitation light (514 nm, Coherent) was coupled into the objective via the ET - 442/514/561 Laser Triple band set (69904, Chroma) and the fluorescence was collected on a electron multiplying charge-coupled CCD camera (512x512 pixel, C9100-13, Hamamatsu). To focal position was held constant by the Olympus ZDC2 and images on multiple positions were collected automatically using Olympus' CellSens software.

**Super-resolution imaging** LacY-eYFP-expressing *E. coli* C41 cells were grown overnight in LB medium with appropriate antibiotics and subsequently diluted 1000 times in EZ rich medium (Teknova) with 0.4% glycerol. When the cell culture reached an OD600 of 0.3, Lacy-eYFP expression was induced with 0.01%(v) L-arabinose for 30 minutes at 37 °C.

The cells were centrifuged at 3000 rcf for 5 minutes and resuspended in EZ rich medium with 0.2% glucose. 3  $\mu$ L of the bacterial culture was transferred onto a microscope coverslide and covered by an agarose gel pad.

Images were captured with an exposure time of 50 ms using the open source software  $\mu$ Manager(92, 93).

To obtain the final STORM reconstruction 2000 frames were collected which were processed with the ImageJ plugin ThunderSTORM(60).

All data and code to used generate the figures in this article can be found at [Zenodo](#) (DOI: 10.5281/zenodo.2637790) and [GitHub](#), respectively.

#### ACKNOWLEDGEMENTS

This work was financed by an ERC Starting Grant (ERC-STG638536 – SM-IMPORT to T. C.) and an ERC Advanced Grant (ERC-ADG694610–SUPRABIOTICS to A.H.). We thank T. Economou for critically reading the manuscript, M. Roelfs for insightful discussions on fitting best practises and R. Wind for logo design.

## Bibliography

1. X. Sunney Xie, Paul J. Choi, Gene-Wei Li, Nam Ki Lee, and Giuseppe Lia. Single-Molecule Approach to Molecular Biology in Living Bacterial Cells. *Annual Review of Biophysics*, 37 (1):417–444, 2008. doi: 10.1146/annurev.biophys.37.092607.174640.
2. Andreas Gahlmann and W. E. Moerner. Exploring bacterial cell biology with single-molecule tracking and super-resolution imaging. *Nature Reviews Microbiology*, 12(1):9–22, January 2014. ISSN 1740-1534. doi: 10.1038/nrmicro3154.
3. Zhizhong Yao and Rut Carballido-López. Fluorescence Imaging for Bacterial Cell Biology: From Localization to Dynamics, From Ensembles to Single Molecules. *Annual Review of Microbiology*, 66(1):459–476, September 2014. ISSN 0066-4227. doi: 10.1146/annurev-micro-091213-113034.
4. Johannes P. Schneider and Marek Basler. Shedding light on biology of bacterial cells. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1707):20150499, November 2016. ISSN 0962-8436, 1471-2970. doi: 10.1098/rstb.2015.0499.
5. Van Gijtenbeek, Lieke A, and Jan Kok. Illuminating Messengers: An Update and Outlook on RNA Visualization in Bacteria. *Frontiers in Microbiology*, 8, 2017. ISSN 1664-302X. doi: 10.3389/fmicb.2017.01161.
6. Johan Elf and Irmeli Barkfors. Single-Molecule Kinetics in Living Cells. *Annual Review of Biochemistry*, 88(1):null, 2019. doi: 10.1146/annurev-biochem-013118-110801.
7. Achillefs N. Kapanidis, Alessia Lepore, and Meriem El Karoui. Rediscovering Bacteria through Single-Molecule Imaging in Living Cells. *Biophysical Journal*, 115(2):190–202, July 2018. ISSN 0006-3495. doi: 10.1016/j.bpj.2018.03.028.
8. Achillefs N. Kapanidis, Stephan Uphoff, and Mathew Stacy. Understanding Protein Mobility in Bacteria by Tracking Single Molecules. *Journal of Molecular Biology*, 430(22):4443–4455, October 2018. ISSN 0022-2836. doi: 10.1016/j.jmb.2018.05.002.
9. Mark C. Leake, Jennifer H. Chandler, George H. Wadham, Fan Bai, Richard M. Berry, and Judith P. Armitage. Stoichiometry and turnover in single, functioning membrane protein complexes. *Nature*, 443(7109):355–358, September 2006. ISSN 1476-4687. doi: 10.1038/nature05135.
10. Anjana Badrinarayanan, Rodrigo Reyes-Lamothe, Stephan Uphoff, Mark C. Leake, and David J. Sherratt. In Vivo Architecture and Action of Bacterial Structural Maintenance of Chromosome Proteins. *Science*, 338(6106):528–531, October 2012. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1227126.
11. Seamus J. Holden, Thomas Pengo, Karin L. Meibom, Carmen Fernandez Fernandez, Justine Collier, and Suliana Manley. High throughput 3d super-resolution microscopy reveals Caulobacter crescentus in vivo Z-ring organization. *Proceedings of the National Academy of Sciences*, 111(12):4566–4571, March 2014. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.131368111.
12. Matthieu Palayret, Helen Armes, Srinjan Basu, Adam T. Watson, Alex Herbert, David Lando, Thomas J. Etheridge, Ulrike Endesfelder, Mike Heilemann, Ernest Laue, Antony M. Carr, David Klenerman, and Steven F. Lee. Virtual-'Light-Sheet' Single-Molecule Localisation Microscopy Enables Quantitative Optical Sectioning for Super-Resolution Imaging. *PLOS ONE*, 10(4):e0125438, April 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0125438.
13. Yi Liao, Jeremy W. Schroeder, Burke Gao, Lyle A. Simmons, and Julie S. Benteen. Single-molecule motions and interactions in live cells reveal target search dynamics in mismatch repair. *Proceedings of the National Academy of Sciences*, 112(50):E6898–E6906, December 2015. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1507386112.
14. Feng Yang, Tai-Yen Chen, Łukasz Krzeminski, Ace George Santiago, Won Jung, and Peng Chen. Single-Molecule Dynamics of the Molecular Chaperone Trigger Factor in Living Cells. *Molecular microbiology*, 102(6):992–1003, December 2016. ISSN 0950-382X. doi: 10.1111/mmi.13529.
15. Britta Barlag, Oliver Beutel, Dennis Janning, Frederik Czarniak, Christian P. Richter, Carina Kommnick, Vera Göser, Rainer Kurre, Florian Fabiani, Marc Erhardt, Jacob Piehler, and Michael Hensel. Single molecule super-resolution imaging of proteins in living *Salmonella enterica* using self-labelling enzymes. *Scientific Reports*, 6:31601, August 2016. ISSN 2045-2322. doi: 10.1038/srep31601.
16. Thomas R Beattie, Nitin Kapadia, Emilien Nicolas, Stephan Uphoff, Adam JM Wollman, Mark C Leake, and Rodrigo Reyes-Lamothe. Frequent exchange of the DNA polymerase during bacterial chromosome replication. *eLife*, 6:e21763, March 2017. ISSN 2050-084X. doi: 10.7554/eLife.21763.
17. Andreas Diepold, Erdinc Sezgin, Miles Huseyin, Thomas Mortimer, Christian Eggeling, and Judith P. Armitage. A dynamic and adaptive network of cytosolic interactions governs protein export by the T3ss injectome. *Nature Communications*, 8:15940, June 2017. ISSN 2041-1723. doi: 10.1038/ncomms15940.
18. Jacob S Lewis, Lisanne M Spinkelink, Slobodan Jergic, Elizabeth A Wood, Enrico Monachino, Nicholas P Horan, Karl E Duderstadt, Michael M Cox, Andrew Robinson, Nicholas E Dixon, and Antoine M van Oijen. Single-molecule visualization of fast polymerase turnover in the bacterial replisome. *eLife*, 6:e23932, April 2017. ISSN 2050-084X. doi: 10.7554/eLife.23932.
19. Mathew Stacy and Achillefs N. Kapanidis. Single-molecule and super-resolution imaging of transcription in living bacteria. *Methods*, 120:103–114, May 2017. ISSN 1046-2023. doi: 10.1016/j.ymeth.2017.04.001.
20. Federico Garza de Leon, Laura Sellars, Mathew Stacy, Stephen J. W. Busby, and Achillefs N. Kapanidis. Tracking Low-Copy Transcription Factors in Living Bacteria: The Case of the lac Repressor. *Biophysical Journal*, 112(7):1316–1327, April 2017. ISSN 0006-3495. doi: 10.1016/j.bpj.2017.02.028.
21. Alexandre W. Bisson-Hilho, Yen-Pang Hsu, Georgia R. Squyres, Erkin Kuru, Fabai Wu, Calum Jukes, Yingjie Sun, Cees Dekker, Seamus Holden, Michael S. VanNieuwenhze, Yves V. Brun, and Ethan C. Garner. Treadmilling by FtsZ filaments drives peptidoglycan synthesis and bacterial cell division. *Science*, 353(626):739–743, February 2017. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aak9973.
22. Lisa E. Lamberte, Gabriele Baniulyte, Shivani S. Singh, Anne M. Stringer, Richard P. Bonocora, Mathew Stacy, Achillefs N. Kapanidis, Joseph T. Wade, and David C. Grainger. Horizontally acquired AT-rich genes in *Escherichia coli* cause toxicity by sequestering RNA polymerase. *Nature Microbiology*, 2(3):16249, March 2017. ISSN 2058-5276. doi: 10.1038/nmicrobiol.2016.249.
23. Victoria A. Lund, Katarzyna Wacnik, Robert D. Turner, Bryony E. Cotterell, Christa G. Walther, Samuel J. Fenn, Fabian Grein, Adam JM Wollman, Mark C. Leake, Nicolas Olivier, Ashley Cadby, Stéphane Mesnage, Simon Jones, and Simon J. Foster. Molecular coordination of *Staphylococcus aureus* cell division. *eLife*, 7, February 2018. ISSN 2050-084X. doi: 10.7554/eLife.32057.
24. Harshad Ghodke, Bishnu P. Paudel, Jacob S. Lewis, Slobodan Jergic, Kamya Gopal, Zachary J. Romero, Elizabeth A. Wood, Roger Woodgate, Michael M. Cox, and Antoine M. van Oijen. Spatial and temporal organization of RecA in the *Escherichia coli* DNA-damage response. *eLife*, 8, February 2019. ISSN 2050-084X. doi: 10.7554/eLife.42761.
25. Robert Crawford, Joseph P. Torella, Louise Aigrain, Anne Plochowietz, Kristofer Gryte, Stephan Uphoff, and Achillefs N. Kapanidis. Long-Lived Intracellular Single-Molecule Fluorescence Using Electroporated Molecules. *Biophysical Journal*, 105(11):2439–2450, December 2013. ISSN 00063495. doi: 10.1016/j.bpj.2013.09.057.
26. Anne Plochowietz, Ian Farrell, Zeev Smilansky, Barry S. Cooperman, and Achillefs N. Kapanidis. In vivo single-RNA tracking shows that most tRNA diffuses freely in live bacteria. *Nucleic Acids Research*, 45(2):926–937, January 2017. ISSN 0305-1048. doi: 10.1093/nar/gkw787.
27. Christoph K. Spaeth, Mathilda Glaesmann, Jonathan B. Grimm, Anthony X. Ayala, Luke D. Lavis, and Mike Heilemann. A toolbox for multiplexed super-resolution imaging of the *E. coli* nucleoid and membrane using novel PAINT labels. *Scientific Reports*, 8(1):14768, October 2018. ISSN 2045-2322. doi: 10.1038/s41598-018-33052-3.
28. Ivan L. Volkov, Martin Lindén, Javier Aguirre Rivero, Ka-Weng leong, Mikhail Metelev, Johan Elf, and Magnus Johansson. tRNA tracking for direct measurements of protein synthesis kinetics in live cells. *Nature Chemical Biology*, 14(6):618–626, June 2018. ISSN 1552-4450, 1552-4469. doi: 10.1038/s41589-018-0063-y.
29. Konstantin A. Lukyanov, Dmitry M. Chudakov, Sergey Lukyanov, and Vladislav V. Verkhusha. Photoactivatable fluorescent proteins. *Nature Reviews Molecular Cell Biology*, 6(11):885–890, November 2005. ISSN 1471-0080. doi: 10.1038/nrm1741.
30. Nathan C. Shaner, Robert E. Campbell, Paul A. Steinbach, Ben N. G. Giepmans, Amy E. Palmer, and Roger Y. Tsien. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nature Biotechnology*, 22(12):1567–1572, December 2004. ISSN 1546-1696. doi: 10.1038/nbt1037.
31. Dmitry Shcherbo, Christopher S. Murphy, Galina V. Ermakova, Elena A. Solovieva, Tatiana V. Chepurnykh, Aleksandr S. Shcheglov, Vladislav V. Verkhusha, Vladimir Z. Pletnov, Kristin L. Hazelwood, Patrick M. Roche, Sergey Lukyanov, Andrey G. Zaraisky, Michael W. Davidson, and Dmitry M. Chudakov. Far-red fluorescent tags for protein imaging in living tissues. *Biochemical Journal*, 418(3):567–574, March 2009. ISSN 0264-6021, 1470-8728. doi: 10.1042/BJ20081949.
32. Annalee W. Nguyen and Patrick S. Daugherty. Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nature Biotechnology*, 23(3):355–360, March 2005. ISSN 1546-1696. doi: 10.1038/nbt1066.
33. Mats Ormö, Andrew B. Cubitt, Karen Kallio, Larry A. Gross, Roger Y. Tsien, and S. James Remington. Crystal Structure of the *Aequorea victoria* Green Fluorescent Protein. *Science*, 273(5280):1392–1395, September 1996. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.273.5280.1392.
34. George H. Patterson and Jennifer Lippincott-Schwartz. A Photoactivatable GFP for Selective Photolabeling of Proteins and Cells. *Science*, 297(5588):1873–1877, September 2002. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1074952.
35. Bartosz Turkowyd, Alexander Balinovic, David Virant, Haruko G. Götz Carnero, Fabienne Caldana, Marc Endesfelder, Dominique Bourgeois, and Ulrike Endesfelder. A General Mechanism of Photoconversion of Green-to-Red Fluorescent Proteins Based on Blue and Infrared Light Reduces Phototoxicity in Live-Cell Single-Molecule Imaging. *Angewandte Chemie International Edition*, 56(38):11634–11639, 2017. ISSN 1521-3773. doi: 10.1002/anie.201702870.
36. Jörg Wiedenmann, Sergey Ivanchenko, Franz Oswald, Florian Schmitt, Carlheinz Röcker, Anya Salih, Klaus-Dieter Spindler, and G. Ulrich Nienhaus. EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. *Proceedings of the National Academy of Sciences*, 101(45):15905–15910, November 2004. ISSN 0027-8424,

- 1091-6490. doi: 10.1073/pnas.0403668101.
37. Jonathan B. Grimm, Brian P. English, Jiji Chen, Joel P. Slaughter, Zhengjian Zhang, Andrey Revyakin, Ronak Patel, John J. Macklin, Davide Normanno, Robert H. Singer, Timothée Lionnet, and Luke D. Lavis. A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nature Methods*, 12(3):244–250, March 2015. ISSN 1548-7105. doi: 10.1038/nmeth.3256.
38. Jonathan B. Grimm, Timothy A. Brown, Brian P. English, Timothée Lionnet, and Luke D. Lavis. Synthesis of Janeila Fluor HaloTag and SNAP-Tag Ligands and Their Use in Cellular Imaging Experiments. In Holger Erfle, editor, *Super-Resolution Microscopy: Methods and Protocols*, Methods in Molecular Biology, pages 179–188. Springer New York, New York, NY, 2017. ISBN 978-1-4939-7265-4. doi: 10.1007/978-1-4939-7265-4\_15.
39. Jonathan B. Grimm, Timothy A. Brown, Ariana N. Tkachuk, and Luke D. Lavis. General Synthetic Method for Si-Fluoresceins and Si-Rhodamines, August 2017.
40. Alexey N. Butkevich, Vladimir N. Belov, Kirill Kolmakov, Viktor V. Sokolov, Heydar Shojaei, Sven C. Sidenstein, Dirk Kamin, Jessica Matthias, Rifka Vlijm, Johann Engelhardt, and Stefan W. Hell. Hydroxylated Fluorescent Dyes for Live-Cell Labeling: Synthesis, Spectra and Super-Resolution STED. *Chemistry – A European Journal*, 23(50):12114–12119, 2017. ISSN 1521-3765. doi: 10.1002/chem.20170212.
41. Kirill Kolmakov, Elke Heßlisch, Thomas Wolfram, Lars A. Nordwig, Christian Andreas Wurm, Haisen Ta, Volker Westphal, Vladimir N. Belov, and Stefan W. Hell. Far-Red Emitting Fluorescent Dyes for Optical Nanoscopy: Fluorinated Silicon-Rhodamines (SiRF Dyes) and Phosphorylated Oxazines. *Chemistry*, 21(38):13344–13356, 2015. doi: 10.1002/chem.201501394.
42. Antje Keppler, Susanne Gendreizig, Thomas Gronemeyer, Horst Pick, Horst Vogel, and Kai Johnson. A general method for the covalent labeling of fusion proteins with small molecules *in vivo*. *Nature Biotechnology*, 21(1):86–89, January 2003. ISSN 1546-1696. doi: 10.1038/nbt765.
43. Georgyi V. Los, Lance P. Encell, Mark G. McDougall, Danette D. Hartzell, Natasha Karassina, Chad Zimprich, Monika G. Wood, Randy Learish, Rachel Friedman Ohana, Marjeta Urh, Dan Simpson, Jacqui Mendez, Kris Zimmerman, Paul Otto, Gediminas Vidugiris, Ji Zhu, Aldis Darzins, Dieter H. Klubertanz, Robert F. Bullerit, and Keith V. Wood. HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis, June 2008.
44. Tilman Plass, Sigrid Milles, Christine Koehler, Jędrzej Szymbarski, Rainer Mueller, Manfred Wießler, Carsten Schultz, and Edward A. Lemke. Amino Acids for Diels–Alder Reactions in Living Cells. *Angewandte Chemie International Edition*, 51(17):4166–4170, 2012. ISSN 1521-3773. doi: 10.1002/anie.201108231.
45. Achillefs N. Kapanidis, Yon W. Ebright, and Richard H. Ebright. Site-Specific Incorporation of Fluorescent Probes into Protein: Hexahistidine-Tag-Mediated Fluorescent Labeling with (Ni<sup>2+</sup>-Nitrilotriacetic Acid) <sub>n</sub> -Fluorochrome Conjugates. *Journal of the American Chemical Society*, 123(48):12123–12125, December 2001. ISSN 0002-7863, 1520-5126. doi: 10.1021/ja017074a.
46. Stephen R. Adams, Robert E. Campbell, Larry A. Gross, Brent R. Martin, Grant K. Walkup, Yong Yao, Juan Llopis, and Roger Y. Tsien. New Biarsenical Ligands and Tetracycline Motifs for Protein Labeling in Vitro and in Vivo: Synthesis and Biological Applications. *Journal of the American Chemical Society*, 124(21):6063–6076, May 2002. ISSN 0002-7863, 1520-5126. doi: 10.1021/ja017687n.
47. Christian Conrad and Daniel W. Gerlich. Automated microscopy for high-content RNAi screening. *The Journal of Cell Biology*, 188(4):453–461, February 2010. ISSN 0021-9525, 1540-8140. doi: 10.1083/jcb.200910105.
48. M. Zeder, E. Kohler, and J. Pernthaler. Automated quality assessment of autonomously acquired microscopic images of fluorescently stained bacteria. *Cytometry Part A*, 77A(1): 76–85, 2010. ISSN 1552-4930. doi: 10.1002/cyto.a.20810.
49. Yair Rivenson and Aydogan Ozcan. Toward a Thinking Microscope: Deep Learning in Optical Microscopy and Image Reconstruction. *Optics and Photonics News*, 29(7):34, July 2018. ISSN 1047-6938, 1541-3721. doi: 10.1364/OPN.29.7.000034. arXiv: 1805.08970.
50. Rainer Pepperkok and Jan Ellenberg. High-throughput fluorescence microscopy for systems biology. *Nature Reviews Molecular Cell Biology*, 7(9):690–696, September 2006. ISSN 1471-0080. doi: 10.1038/nrm1979.
51. Zvi Kam, Eli Zamir, and Benjamin Geiger. Probing molecular processes in live cells by quantitative multidimensional microscopy. *Trends in Cell Biology*, 11(8):329–334, August 2001. ISSN 0962-8924. doi: 10.1016/S0962-8924(01)02067-0.
52. Walter Schubert, Bernd Bonnekoh, Ansgar J. Pommer, Lars Philipsen, Raik Böckelmann, Yanina Malykh, Harald Gollnick, Manuela Friedenberger, Marcus Bode, and Andreas W. M. Dress. Analyzing proteome topology and function by automated multidimensional fluorescence microscopy. *Nature Biotechnology*, 24(10):1270–1278, October 2006. ISSN 1546-1696. doi: 10.1038/nbt1250.
53. C. Sommer, C. Strähle, U. Köthe, and F. A. Hampecht. Ilastik: Interactive learning and segmentation toolkit. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 230–233, March 2011. doi: 10.1109/ISBI.2011.5872394.
54. Anne E. Carpenter, Thouis R. Jones, Michael R. Lamprecht, Colin Clarke, In Han Kang, Ola Friman, David A. Guérin, Joo Han Chang, Robert A. Lindquist, Jason Moffat, Polina Golland, and David M. Sabatini. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology*, 7(10):R100, October 2006. ISSN 1474-760X. doi: 10.1186/gb-2006-7-10-r100.
55. Lee Kamentsky, Thouis R. Jones, Adam Fraser, Mark-Anthony Bray, David J. Logan, Katherine L. Madden, Vebjørn Ljosa, Curtis Rueden, Kevin W. Eliceiri, and Anne E. Carpenter. Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. *Bioinformatics*, 27(8):1179–1180, April 2011. ISSN 1460-2059, 1367-4803. doi: 10.1093/bioinformatics/btr095.
56. François Chollet et al. Keras. <https://keras.io>, 2015.
57. Martin Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
58. Seamus J. Holden, Stephan Uphoff, and Achillefs N. Kapanidis. DAOSTORM: an algorithm for high- density super-resolution microscopy. *Nature Methods*, 8(4):279–280, April 2011. ISSN 1548-7105. doi: 10.1038/nmeth.0411-279.
59. Hazen Babcock, Yaron M. Sigal, and Xiaowei Zhuang. A high-density 3d localization algorithm for stochastic optical reconstruction microscopy. *Optical Nanoscopy*, 1(1):6, July 2012. ISSN 2192-2853. doi: 10.1186/2192-2853-1-6.
60. Martin Ovesný, Pavel Křížek, Josef Borkovec, Zdeněk Švindrych, and Guy M. Hagen. ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics*, 30(16):2389–2390, August 2014. ISSN 1367-4803. doi: 10.1093/bioinformatics/btu202.
61. Ricardo Henriques, Mickael Lelek, Eugenio F. Fornasier, Flavia Valtorta, Christophe Zimmer, and Musa M. Mhlanga. QuickPALM: 3d real-time photoactivation nanoscopy image processing in ImageJ. *Nature Methods*, 7(5):339–340, May 2010. ISSN 1548-7105. doi: 10.1038/nmeth.0510-339.
62. Julie A. Cass, Stella Stylianidou, Nathan J. Kuwada, Beth Traxler, and Paul A. Wiggins. Probing bacterial cell biology using image cytometry. *Molecular Microbiology*, 103(5):818–828, March 2017. ISSN 1365-2958. doi: 10.1111/mmi.13591.
63. Stella Stylianidou, Connor Brennan, Silas B. Nissen, Nathan J. Kuwada, and Paul A. Wiggins. SuperSegger: robust image segmentation, analysis and lineage tracking of bacterial cells. *Molecular Microbiology*, 102(4):690–700, November 2016. ISSN 1365-2958. doi: 10.1111/mmi.13486.
64. Ahmad Paintidakis, Bradley Parry, Manuel Campos, Irnov Irnov, Johan Elf, Ivan Surovtsev, and Christine Jacobs-Wagner. Outfit: an integrated software package for high-accuracy, high-throughput quantitative microscopy analysis: Outfit: image analysis software. *Molecular Microbiology*, 99(4):767–777, February 2016. ISSN 0950-382X. doi: 10.1111/mmi.13264.
65. Adrien Ducret, Ellen M. Quardokus, and Yves V. Brun. MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. *Nature Microbiology*, 1(7):16077, July 2016. ISSN 2058-5276. doi: 10.1038/nmicrobiol.2016.77.
66. Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. [Online; accessed <today>].
67. Helen Shen. Interactive notebooks: Sharing the code. *Nature News*, 515(7525):151, November 2014. doi: 10.1038/515151a.
68. Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, Paul Ivanov, Damián Avila, Safia Abdalla, and Carol Willing. Jupyter notebooks – a publishing format for reproducible computational workflows, 2016.
69. Jeffrey M. Perkel. Why Jupyter is data scientists' computational notebook of choice. *Nature*, 563:145, October 2018. doi: 10.1038/d41586-018-07196-1.
70. Antonino Ingargiola, Eitan Lerner, SangYoon Chung, Shimon Weiss, and Xavier Michalet. FRETBURsts: An Open Source Toolkit for Analysis of Freely-Diffusing Single-Molecule FRET. *PLOS ONE*, 11(8):e0160716, August 2016. ISSN 1932-6203. doi: 10.1371/journal.pone.0160716.
71. Caroline A. Schneider, Wayne S. Rasband, and Kevin W. Eliceiri. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9:671–675, June 2012. ISSN 1548-7105. doi: 10.1038/nmeth.2089.
72. Curtis T. Rueden, Johannes Schindelin, Mark C. Hiner, Barry E. DeZonia, Alison E. Walter, Ellen T. Arena, and Kevin W. Eliceiri. ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics*, 18(1):529, November 2017. ISSN 1471-2105. doi: 10.1186/s12859-017-1934-z.
73. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.
74. Luis Pedro Coelho, Mahotas: Open source software for scriptable computer vision. *Journal of Open Research Software*, 1(1):e3, July 2013. ISSN 2049-9647. doi: 10.5334/jors.ac.
75. Martin Roelfs and Peter C Kroon. tBuLi/symfit: symfit 0.4.6, October 2018.
76. Alexey Sharonov and Robin M. Hochstrasser. Wide-field subdiffraction imaging by accumulated binding of diffusing probes. *Proceedings of the National Academy of Sciences*, 103(50):18911–18916, December 2006. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0609643104.
77. Christoph Spahn, Francesca Cella-Zannacchi, Ulrike Endesfelder, and Mike Heilemann. Correlative super-resolution imaging of RNA polymerase distribution and dynamics, bacterial membrane and chromosomal structure in Escherichia coli. *Methods and Applications in Fluorescence*, 3(1):014005, March 2015. ISSN 2050-6120. doi: 10.1088/2050-6120/3/1/014005.
78. Yong Hwee Foo, Christoph Spahn, Hongfang Zhang, Mike Heilemann, and Linda J. Kenney. Single cell super-resolution imaging of *E. coli* OmpR during environmental stress. *Integrative Biology*, 7(10):1297–1308, October 2015. ISSN 1757-9708. doi: 10.1039/C5IB00077G.
79. Michael J. Rust, Mark Bates, and Xiaowei Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods*, 3(10):793–796, October 2006. ISSN 1548-7091. doi: 10.1038/nmeth929.
80. Andrew Robinson, John P. McDonald, Victor E. A. Caldas, Meghna Patel, Elizabeth A. Wood, Christiaan M. Punter, Harshad Ghodke, Michael M. Cox, Roger Woodgate, Myron F. Goodman, and Antoine M. van Oijen. Regulation of Mutagenic DNA Polymerase V Activation in Space and Time. *PLOS Genetics*, 11(8):e1005482, August 2015. ISSN 1553-7404. doi: 10.1371/journal.pgen.1005482.
81. Eric Betzig, George H. Patterson, Rachid Sougrat, O. Wolf-Lindwasser, Scott Olenych, Juan S. Bonifacino, Michael W. Davidson, Jennifer Lippincott-Schwartz, and Harald F. Hess. Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science*, 313(5793): 1642–1645, September 2006. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1127344.
82. Julie S. Bente, Michael A. Thompson, Nicole K. Tselenitis, Grant R. Bowman, Lucy Shapiro, and W. E. Moerner. Super-resolution imaging in live *Caulobacter crescentus* cells using photoswitchable EYFP. *Nature Methods*, 5(11):947–949, November 2008. ISSN 1548-7105. doi: 10.1038/nmeth.1258.
83. A. Ingargiola. Pyccorelate. <https://github.com/OpenSMFS/pyccorelate>, 2017.

84. Matthew T. Swulius and Grant J. Jensen. The Helical MreB Cytoskeleton in Escherichia coli MC1000/pLE7 Is an Artifact of the N-Terminal Yellow Fluorescent Protein Tag. *Journal of Bacteriology*, 194(23):6382–6386, December 2012. ISSN 0021-9193, 1098-5530. doi: 10.1128/JB.00505-12.
85. Frank R Hampel, Elvezio M Ronchetti, Peter J Rousseeuw, and Werner A Stahel. *Robust statistics: the approach based on influence functions*, volume 196. John Wiley & Sons, 2011.
86. Daniel Müllner. fastcluster: Fast hierarchical, agglomerative clustering routines for r and python. *Journal of Statistical Software, Articles*, 53(9):1–18, 2013. ISSN 1548-7660. doi: 10.18637/jss.v053.i09.
87. Rainer Storn and Kenneth Price. Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *Journal of Global Optimization*, 11(4): 341–359, December 1997. ISSN 1573-2916. doi: 10.1023/A:1008202821328.
88. Jeffrey M. Perkel. Data visualization tools drive interactivity and reproducibility in online publishing. *Nature*, 554:133, January 2018. doi: 10.1038/d41586-018-01322-9.
89. Ian L. Gunsolus, Dehong Hu, Cosmin Mihai, Samuel E. Lohse, Chang-soo Lee, Marco D. Torelli, Robert J. Hamers, Catherine J. Murphy, Galya Orr, and Christy L. Haynes. Facile method to stain the bacterial cell surface for super-resolution fluorescence microscopy. *The Analyst*, 139(12):3174–3178, June 2014. ISSN 0003-2654. doi: 10.1039/c4an00574k.
90. H. Peter Lu, Luying Xun, and X. Sunney Xie. Single-Molecule Enzymatic Dynamics. *Science*, 282(5395):1877–1882, December 1998. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.282.5395.1877.
91. Samuel O. Skinner, Leonardo A. Sepúlveda, Heng Xu, and Ido Golding. Measuring mRNA copy number in individual *Escherichia coli* cells using single-molecule fluorescent *in situ* hybridization. *Nature Protocols*, 8(6):1100–1113, June 2013. ISSN 1750-2799. doi: 10.1038/nprot.2013.066.
92. Arthur D. Edelstein, Mark A. Tsuchida, Nenad Amodaj, Henry Pinkard, Ronald D. Vale, and Nico Stuurman. Advanced methods of microscope control using μManager software. *Journal of Biological Methods*, 1(2):e10, November 2014. ISSN 2326-9901. doi: 10.14440/jbm.2014.36.
93. Arthur Edelstein, Nenad Amodaj, Karl Hoover, Ron Vale, and Nico Stuurman. Computer Control of Microscopes Using μManager. *Current Protocols in Molecular Biology*, 92(1): 14.20.1–14.20.17, 2010. ISSN 1934-3647. doi: 10.1002/0471142727.mb1420s92.