Quantitative Microbiology with Microscopy: Effects of Projection and Diffraction

Georgeos Hardo[1], Ruizhe Li[1], Somenath Bakshi[1]*

[1]Department of Engineering, University of Cambridge, Cambridge, United Kingdom

***Correspondence:**

Corresponding Author

sb2330@cam.ac.uk

Keywords: image analysis, fluorescence microscopy, single molecule imaging, image segmentation, microcolony on agar, deep learning of image processing, gene-expression noise, cell size regulation

Abstract

Live cell imaging of microbial cells with microscopy has revolutionised quantitative microbiology. Micrographs are one of the most information-rich data types captured about a microbe, allowing quantification of the size and morphology of individual cells and their gene-expression over time. However, an optical microscope is a diffraction-limited system, and the comparable size of the system's point spread function to the size of a microbial cell can lead to imaging artefacts which corrupt and bias the data. Additionally, the comparable thickness of a microbe to the microscope's depth of field means that the 2D image contains compressed, projected 3D information. This makes it difficult to extract the underlying 3D distribution of photon emitters. For unknown distributions, the problem can be as ill-posed as a deconvolution problem, usually not having a unique solution. Together, the diffraction and projection effects affect our ability to accurately quantify the size and shape of microbial cells from their images and their contents from intensity measurements. In this paper, we use a mixture of simulations and experiments of microscopic image formation of microbial cells to illustrate the effects of diffraction and projection on cell segmentation and signal quantification. We use targeted experiments to validate the predictions where possible. Finally, we use the knowledge of these effects to design experiments which can help to reduce the errors and biases in our analysis. Awareness of these effects and the approaches towards alleviating them will help to accurately quantify microbiology from microscopy data.

Introduction

Optical microscopy with visible light has become one of the main tools for microbiology, as it enables one to see and follow living microbial cells over time [1], [2]. However, beyond just the ability to see, microscopy has enabled us to quantify the lives of microbes, transforming the field of microbiology from qualitative to quantitative [3]. The ability to image individual living

cells over time has enabled us to quantify their size, shape, and content in a noninvasive, timelapse manner. High-throughput microscopic investigation of bacterial cells under controlled environments have recently supercharged the field of quantitative microbiology. These include numerous studies on analysing modes of size regulation and division control in bacteria [4]–[7] and a range of investigations quantifying the regulation of gene-expression noise and timing in individual cells [8]–[10]. On the other hand, single molecule measurements within cells have enabled us to study the dynamics and organisation of individual genes, mRNAs, and proteins and to count their copies to construct an accurate distribution of their abundance [11]–[13]. In summary, live-cell microscopy is helping us to develop a multiscale picture of the organisation, dynamics, and regulation of various processes within microbes. To this end, several dedicated software packages have been developed to extract quantitative information of microbial cells from microscopy data [14]–[17].

However, caution is needed for the informed use of microscopy data to extract quantitative information about the underlying microbiology. An optical microscope is a diffraction-limited system, where, due to the wave-like nature of light, a single point-source (the object) will appear as a blurred point spread function (PSF) (the image) [18]. In practice this means that the photons from a real object may not appear to originate from that object's true position in the resultant diffraction-limited image. This affects our abilities to quantify the intensity content of individual cells in order to construct abundance distributions across a population. On the other hand, most microbes, especially bacterial cells, are comparable in size to the PSF [19]. This means that any image taken with traditional wide-field fluorescence microscopes will be significantly blurred. Additionally, bacterial cells have comparable thickness as the depth of field (DoF) of a microscope objective [20], [21]. If one does not use techniques such as confocal or two-photon microscopy (which are less suitable for time-resolved live-cell imaging) then one also has to contend with the effects of projection (out of focus light). Due to the finite depth of field in regular widefield imaging, out of focus light contributes to the image, as much of the depth of the sample is illuminated by the excitation light, and photons from all sample planes are collected. Together these effects affect our ability to estimate the size, shape, and intensity of cells from their 2D images. Furthermore, the combination of these effects affects single molecule microscopy, making it difficult to separate spots from individual labelled molecules, as spots from various depths get defocussed to different extents, overlap with each other in the 2D projection and get blurred into a single spot due to diffraction, making single molecule counting a challenge. This leads to undercounting errors in quantifying the abundance of low-copy number molecules, such as mRNAs and transcription factors. Put together, these effects seriously limit our ability to extract quantitative information about microbiological systems from their microscopic images. These effects are more significant for bacteria, which are a key focus of most of the quantitative microbiology studies, and are also small enough to experience significant diffraction and projection issues.

In this paper we model bacterial image formation in the fluorescence widefield microscope to explore and illustrate the effects of diffraction and projection on cell segmentation, size estimation, and signal quantification. We use the recently developed SyMBac pipeline to generate synthetic images with various imaging settings and properties of microbial cells and their organisation in the image [22]. These digital experiments enable us to have easy and precise control over a wide range of parameters, including optics, physics, and cellular properties, and to analyse their effect on image formation and data extraction. The setup also enables a wide coverage of parameters that are otherwise difficult or impossible to achieve in real experimental setups but play a critical role to identify important variables by amplifying their effects. Furthermore, the simulated images are accompanied by accurate ground truth information, enabling easy performance benchmarking of various algorithms for different tasks.

We primarily focus on fluorescence images of bacterial cells in the analysis of cell segmentation performance. However, the results and conclusions are equally valid for the brightfield and phase contrast images, as they also suffer from projection and diffraction effects. For signal quantification, we analyse both the total intensity based measurements and single molecule counting to estimate abundance of tagged molecules. Our analysis shows that the effects of projection and diffraction are significant in the estimates of cell size, intensity, and molecule copy number, and are difficult to correct for. We have also performed targeted experiments to examine the predictions from simulated data. Finally, for each section, we have illustrated possible ways to design experiments and analysis methodologies to reduce the extent of such effects. Awareness of these effects and the approaches towards alleviating them will help to accurately quantify microbiology from microscopy data.

Computational Methods

Virtual fluorescence microscopy using SyMBac

All image simulations in this study were performed using the SyMBac Python library [22]. We primarily focus on simulating fluorescence images in this paper, but some of the analysis also involves simulated phase contrast images of cells growing in microfluidic devices. Unless otherwise stated, all virtual fluorescence microscopy images are generated using the same workflow: The 3D spherocylindrical hull of a digital cell is placed within an environment (isolated or with other neighbouring cells, if position data from colony growth simulations is present), and fluorescent emitters are uniformly sampled within its volume. A 3D array keeps track of the indices where fluorescent emitters are located, with the value at each position indicating the number of emitters. For homogeneously distributed fluorescence in each cell, we define a “density” value, which is defined as the average number of emitters per volume element in a cell, and so the number of molecules within a cell is given by the density multiplied by the cell volume. Diffraction and projection effects are then simulated (generating a synthetic image) by convolution of this data with a PSF. Convolution was performed using both a theoretical PSF and an empirically measured PSF. SyMBac’s

`PSF_generator` class was used to generate synthetic PSFs according to the Gibson-Lanni model [23].

To simulate the effect of projection and out-of-focus light typical of a widefield fluorescence microscope, it is assumed that the centre of the PSF is located at the centre of the cell. Each slice of the PSF is therefore convolved with its corresponding slice of the cell as shown in Figure 1a. Therefore, to simulate the non-physical effect of reducing projection, the depth of field of the microscope can be artificially changed by truncating the number of PSF planes convolved with the cell. For example, a 1 micron wide cell simulated at a 0.0038 micron pixel size will have 263 slices in the Z direction. To simulate an image with no projection artefacts, only the middle Z-slice of the cell is convolved with the middle Z-slice of the 3D PSF. This technique applies to using both simulated and measured PSFs. To remove the effect of diffraction entirely, the PSF can be replaced with the identity ("do-nothing") kernel, which upon convolution will return the original data. To scale the effect of diffraction continuously, the frequency of light used to simulate the kernel can be increased arbitrarily.

Simulations of fluorescence images of individual cells

Digital cells of a constant length of 6 μm were used to avoid introducing errors from length underestimation into the width estimations. Cell width was varied between 0.5 and 3.0 μm, and the simulated depth of field was varied to between 3 μm (widest cell) and 0.0038 μm (a single Z slice). The PSF generator was set to the "3d fluo" mode, which uses the Gibson-Lanni PSF model [23] for 3D convolution. Additionally, we included the identity kernel as the PSF for a hypothetical undiffracted microscope with an imaging wavelength of 0 μm.

To simulate images of cells with cytoplasmic fluorescent markers, 3D spherocylindrical cells were rendered, and emitters sampled as described above. To simulate membrane fluorescent cells, fluorescent emitters were sampled only within a single pixel layer corresponding to the outermost cell volume. We ensured to render images at a high resolution, allowing accurate drawing of the cell membrane. Rendering at a lower resolution, even a single pixel would be significantly thicker than the true thickness of the cell membrane.

Simulated images of fluorescent single cells, either with cytoplasmic markers or membrane markers, were generated at different wavelengths, widths, and depths of focus, a sample of which can be seen in figure 1b and figure 2a. It is important to note here that when the depth of field is changed, this is a simulation of a non–physical effect. In reality the amount of out of focus light collected by the microscope is defined by the objective lens. By tuning the number of Z-PSF layers that a 3D cell volume is convolved with, we are simulating the non-physical effect of artificially reducing the amount of out of focus light being collected by the microscope objective. This is still important, as it allows us to identify the source of bias and error in our measurement. The same can be said for diffraction, when using the identity kernel, we are simulating an image without diffraction. While this is also non-physical, it is critically useful to observe how an image is corrupted by the microscope's PSF. In real experiments, both these effects are taking place at the same time, and therefore the

experimental images are only compared here with the simulated images that contain both effects.

Cell size quantification and analysis

After simulating single cells, the size estimation errors due to diffraction and projection can be quantified in two different ways. For single fluorescent cells, a simple binary mask of the resultant synthetic image is created using Otsu threshold [24], and the length and width are calculated by taking the dimensions of the two principal axes of the binary object. The width and length of synthetic fluorescent membrane images, however, is calculated by finding the inter-peak distance of the 1D intensity profile drawn down the two principal axes of the cell.

Simulations of fluorescence images of microcolonies of cells

The agar pad simulation feature of SyMBac was used to generate microcolony simulations [22], each of which was terminated when reaching a size of 250 cells . This pipeline uses the cellmodeller platform (which has been integrated into SyMBac) for generating the ground truth microcolonies of cells [25]. For the purpose of these simulations, cell width was fixed at 1 μm, and cells were set to divide at a target length of 3.5 μm +/- *Uniform*(-0.25μm, 0.25μm). This yielded image sequences of growing densely packed colonies.

Individual cells in each colony simulation contain uniformly distributed fluorescent emitters. To control the coefficient of variation (CV) of intensity among the cells within a microcolony, we fixed the mean density of emitters within a cell, but varied the variance of a truncated (at 0) normal distribution in order to sample single cell intensities with a desired CV. The CV was sampled between 0 and 0.3. Synthetic images from these colonies were generated with 3D convolution in the same manner as described in the previous section, with the PSF being rendered for a 1.49NA objective, a refractive index of 1.518, a working distance of 0.17 mm, and with imaging wavelengths of 0.4 μm, 0.55 μm, and 0.7 μm for separate simulations. Additionally, the same ground-truth data was convolved with an empirical PSF captured from a real 100x objective lens with the same parameters described, imaged at 0.558 μm wavelength light. This yielded synthetic microscope images of microcolonies along with corresponding ground truth masks under a variety of imaging conditions. The relevant codes are provided in the GitHub repository (https://github.com/georgeoshardo/projection_diffraction_frontiers_in_physics) associated with this project.

Colony image intensity quantification and analysis

Analysis of synthetic microcolonies

Calculating the intensity of each cell within the synthetic microcolony images did not require segmentation because the ground truth mask positions are available from the simulations. Thus, for each ground truth mask, the average intensity in the corresponding position in the

synthetic microscope image was enumerated and used to calculate the CV, which could be compared to the ground truth CV. To assess whether the ground truth CV can be recovered under ideal circumstances, we performed Wiener deconvolution using the original PSF [26]. The deconvolved CV can then be compared to the ground truth CV by enumerating the deconvolved image's average intensity within each ground truth mask position.

The distance of each cell from the centre of the colony was calculated by calculating each cell's Euclidean distance to the mean position of all cells within the colony (the colony centroid). The distance was normalised by dividing by the maximum Feret's radius of the colony as calculated by Scikit-image's regionprops function [27]. The number of neighbours for each cell was calculated by first dilating all cell masks by 4 pixels to ensure that neighbouring cell masks touch. The mask image was then converted into a region adjacency graph, where each node is a cell, and edges represent a neighbour connection between two cells (cells that touch). The graph is then traversed and each node's degree (corresponding to that cell's neighbour number) is enumerated (Figure 4d).

Analysis of experimental microcolonies

Experimental microcolony data were analysed using the same methods as synthetic microcolonies, but were first segmented in phase contrast using the Omnipose model to generate masks. Datasets generated this way have no ground truth intensity estimation. Fluorescence images were first background subtracted by subtracting the mean of the lowest 5% of pixel intensities within the image. Mean cell intensity was defined as the sum of the pixel intensities within each mask divided by the cell mask area. Deconvolution was performed using the empirical PSF, and the intensity of deconvolved cells recorded. Since no ground truth data exists, we did not estimate the CV of real data, but rather focussed on showing the effects of colony size, cell neighbour number, and cell position within a colony on the observed cell intensity. These values were quantified using the same methods as for synthetic colonies.

Manual annotation platform and analysis

To assess the effects of projection and diffraction on the performance of manual (human) annotation of cells (for the purposes of training data preparation), five individuals were asked to annotate 600 cells each. The cells had an average width of 1 micron, and were partitioned into 4 possible groups: projection "off" cells (cells where the imaging DoF is 1 pixel wide)/projection "on" cells (where the depth of field contains the entire 3D cell), and cells imaged with a fluorescent emitter capable of emitting 0.2 micron wavelength light, and a 0.6 micron wavelength light emitter, distributed with a density of 0.4 emitters per volume element within each cell. 150 of each cell type were scattered with uniform random position and orientation on a 2048x2048 plane with a pixel size of 0.065. Convolution for the two wavelengths was performed once again with the Gibson-Lanni PSF model. Camera noise was added after convolution using SyMBac's camera class, using a baseline intensity of 100, a

sensitivity of 2.9 ADUs/electron, and a dark noise variance of 8. Annotators ran a Python script which presented them with a Napari window [28] and a layer upon which to draw on cells. The order in which the images with various effects were displayed was randomised.

Deep-learning models for image segmentation

Size estimation

In addition to comparing human annotation accuracy, we sought to test the accuracy of a pretrained model (in this case Omnipose's bact_fluor_omni model) on simulated images where the perfect ground truth dimension of each cell is known. We generated 200 images containing on average 200 cells per image (approximately 40,000 total cells) according to the same method described in the previous section, but with cell areas varying between 0.15 square micron to 3.5 square micron. The PSF model used an imaging wavelength of 600 nm. Ground truth mask and image pairs were saved. Images were then segmented according to the recommended Omnipose documentation with the bact_fluor_omni model. Ground truth cells were matched with their prediction and the IoU for each cell calculated.

We then assessed the benefit of training a model on synthetic data matching one's experimental setup exactly. The performance gained by training a model on an independent set of synthetic images with perfect ground truth available was checked by generating a test set of 200 images according to the aforementioned method. These were used to train a new model according to the Omnipose documentation for 4000 epochs. The 4000th epoch model was then used to segment the original test data, and ground truth cells were matched with their prediction and IoUs calculated once again.

Division time estimation

We also assessed a deep learning model's accuracy in estimating the point at which a cell divides into two. Achieving accurate cell division time is critical for accurately assessing a cell's model of size regulation. For this task we also chose to use Omnipose, since it learns to create instance masks purely from the provided training data. Omnipose also has the added benefit of allowing cells to touch, and so no loss of precision is incurred due to single pixel separation requirements given by semantic masks (which in microscopy experiments can account for up to 10% of a cell's width).

We used SyMBac to generate synthetic training and test data (1200 and 5000 samples respectively) images of fluorescent cells in the mother machine undergoing division at very high time resolution. We extended SyMBac by implementing a full septum pinching model to accurately represent cellular division dynamics. An Omnipose model was trained on this data using the default arguments given in the documentation. We compared the segmented output of both our trained model, and the pretrained "bact_fluor_omni" model provided by Omnipose to compare synthetic data with a model pretrained on human annotated data.

Analysis of cell wall labelled cells on agar pads

After acquisition of cells labelled with fluorescent D-amino acids (FDAAs), images were segmented with Omnipose. Since size estimation of the cell by segmentation is not important, the pre-trained bact_phase_omni model was sufficient to segment the cells. To ensure that all the signal from the fluorescently labelled cell wall was captured, cell masks were binary dilated. After this, all individual cells were cropped, and Scikit-image's regionprops function [27] was used to calculate the orientation of the cells and rotate them

Simulations of fluorescence images of individual molecules within cells

Fluorescent single molecule images of single cells were generated in the same manner as described before, but with very few emitters per cell (1-30). These low density fluorescent cells were convolved with a synthetic PSF using the layer-by-layer technique previously described. All analyses were performed on 3 cell types: 1) A typical 1 µm wide, 1 µm deep, 5 µm long cell, 2) An enlarged 2 µm wide, 2 µm deep, 5 µm long cell, 3) A cell trapped by the MACS [11] platform, 2 µm wide, 0.6 µm deep, 5.5 µm long.

We employed two techniques to count the single molecules in these cells. The first approach, which we term the naive approach, involved sampling fluorescent emitters within the cell and partitioning the emitters into 3 groups: 1) Molecules lost to depth of field; these were defined as molecules more than 0.25 µm away from the centre of the cell. 2) Molecules lost to diffraction; these were defined as molecules residing within 1 Rayleigh diffraction limit of at least one other molecule. 3) Resolved molecules; these are the sum of any remaining resolvable single molecules and clusters of molecules within 1 Rayleigh diffraction limit of another (appearing as a single molecule).

The second approach we applied was Deep-STORM [29], a deep learning method for super resolution single molecule localisation. A more sophisticated method such as this should perform better than the naive method since it can learn to use defocus, changes in local intensity, and local spatial patterning information to better estimate the number of molecules in a region. We trained Deep-STORM by downloading and modifying the ZeroCostDL4Mic [30] implementation for local use. While Deep-STORM is typically trained on simulated data, and comes with its own simulator, it does not take into account thick samples such as the depth of entire bacterial cells where defocus is appreciable. Therefore, we generated our own synthetic training data by reducing the number of fluorescent emitters in each cell to between 1-30. Individual models were trained for the regular cell and the MACS cell, both with PSNRs (peak signal to noise ratios) of 5, which is typical for a bacterial single molecule experiment. To test how much SNR affects localisation and counting accuracy, we generated datasets which simulate two non-physical effects: A dataset with infinite SNR (no background noise), and a dataset in which every emitter is in perfect focus.

Experimental Methods

Strain preparation

For imaging cells labelled with membrane stains, we used the strain *E. coli* MG1655 7740 ΔmotA with no fluorescent reporter. Cells were grown overnight from a glycerol stock in LB medium at 37°C with 250 RPM shaking. The following day, cells were diluted by 100x into 1 mL of fresh LB. The fresh LB was supplemented simultaneously with both HADA and RADA to a final concentration of 1 mM each. HADA and RADA are fluorescent D-amino acids (FDAAs) which incorporate into the bacterial cell wall during growth, allowing imaging of only the cell outline using fluorescence microscopy [31], [32]. Cells were allowed to grow in the presence of the FDAAs for 2 hours, after which a 300 µL aliquot was spun down and washed with PBS according to the protocol in [33], taking care to ensure that cells were kept on ice between washes and when not being used.

For imaging microcolonies of fluorescently tagged cells, we used the strain *E. coli* MG1655 7740 ΔmotA with constitutively produced cyan fluorescent protein (SCFP3A, FPbase ID: HFE84) under the control of prpsL. Cells were grown overnight from a glycerol stock in LB medium at 37°C. The following day, cells were diluted by 200x in fresh LB and grown to an OD or 0.1-0.2 to ensure large cell size. Once the desired OD was reached, 1mL of cells were spun down at 4000x g for 5 minutes, and the pellet resuspended in phosphate buffered saline (PBS) for imaging.

Single cell imaging on agar pad

Agar pads were prepared according to the protocol described in [2]. Since only snapshot microscopy was to be performed, agar pads were instead prepared with PBS instead of growth medium, and were kept as thin as possible. Agar pads were cut to approximately 22 x 22mm, and placed upon a 22 x 40mm coverslip. Cells on the agar pad were imaged using a Nikon ECLIPSE Ti2 inverted microscope using a 100x (NA=1.49) objective with F-type immersion oil (n=1.518) with a second 1.5x post-magnification lens inserted, for an effective magnification of 150x. The camera used was an ORCA-Fusion C14440-20UP digital CMOS camera from Hamamatsu, with a pixel size of 6.5 µm x 6.5 µm. Cells stained with HADA were imaged with excitation light: 365 nm, 435 nm filter, 100% power, 1 second exposure. RADA was imaged with excitation light: 561 nm, 595 nm filter, 100% power, 0.5 second exposure. Focussing and FOV selection was again done using phase contrast, but special care was taken to account for chromatic aberration by adjusting the Z-plane offset between the focussed phase contrast image and the RADA and HADA images. This was crucial to ensuring that the cell wall was in focus in each image.

Imaging microcolonies on agar pads

Since, cells can change their intensities during the growth on agar pad, we image preformed colonies to image the effects of diffraction on the cells. In order to generate preformed microcolonies, a higher OD of cells (0.1-0.2) was preferred. 3 µL of cell suspension was pipetted directly onto the agar pad and allowed to “dry” for 5 minutes, after which a second 22 x 40mm was placed upon it. Agar pads were then immediately imaged using the ECLIPSE Ti2 inverted microscope using a 100x (NA=1.49) objective. This enabled us to collect samples of cell clusters (preformed colonies) of various sizes. To avoid photobleaching, well separated fields of view were first selected and focussed in phase contrast. Fluorescent images were captured by excitation with 440 nm light with an LED power of 50% for 900 ms, and with a filter wavelength of 475 nm. Images were captured as multidimensional 16-bit ND2 files for further analysis.

Timelapse imaging of cell division at high frame-rate

Timelapse imaging of single cells at high frame rate (0.1Hz) was performed in the microfluidic device known as the mother machine. To prepare the strain for loading, 2mL of LB was inoculated from a glycerol stock of *E. coli* MG1655 7740 ΔmotA with constitutively produced RFP (mCherry, FPbase ID: ZERB6) under the control of prpsL. The cells were grown overnight at 37°C and 250 RPM shaking. After growth, the cells were spun down at 4000xg to make a dense pellet. The mother machine was prepared using the density loading method whereby the pellet is resuspended in <20 µL of LB, drawn into a gel loading tip, and injected into the device. The cells were left to load into the trenches for 30 minutes. After cell loading, blunt-end syringe needles were connected to either end of the mother machine feeding lane. A syringe pump was connected with silicone tubing to the mother machine inlet and used to continuously flow LB at 15 µL/min through the mother machine. The other end of the feeding lane was connected to a waste bottle. Enough time was given for the cells to exit the stationary phase and start dividing in the exponential phase. Fluorescence microscopy was performed on a Nikon ECLIPSE Ti2 with 594 nm excitation light (provided by a Lumoncore SPECTRA Light Engine, LED power at 100%) with a 632 nm emission filter imaging every 3 minutes. A 100x 1.49NA objective with n=1.518 immersion oil was used for imaging and images were collected using the ORCA-Fusion camera.

Point spread function acquisition

Our microscope’s (a Nikon ECLIPSE Ti2) point spread function was captured using fluorescent 0.1 µm TetraSpeck Microspheres from Invitrogen. Slides with fluorescent microspheres were prepared according to [34], with the only change being a bead dilution of 1000x, and the use of Fluoromount-G Mounting Medium from Invitrogen. PSFs were captured using 0.70 NA, 0.95 NA, and 1.49 NA objective lenses, with magnifications of 20x,

40x, and 100x respectively. PSFs were captured with and without the addition of a 1.5x post-magnification lens. Z-stacks were taken of the beads with 0.05 µm spacing. The most in-focus Z stack was determined by taking the radial profile of the PSF and finding the Z-slice with the highest peak intensity and narrowest FWHM.

Results and Discussion

Projection and diffraction effects in widefield fluorescence microscopy

In this paper we primarily focus our analysis on conventional widefield fluorescence microscopy, which is widely used for quantitative microbiology research. In this imaging setup, the cells from the sample are held between a glass coverslip and biocompatible material (e.g agarose, PDMS, etc), and are imaged through the cover glass by the microscope's objective lens. The configuration of the imaging setup can be upright or inverted [35]. For both the upright and inverted imaging setup, the microscope objective is used to both focus the excitation beam to illuminate the sample and to collect the emitted photons from the entire sample to form the two dimensional image on the camera. This is often termed as an epifluorescence microscopy setup. The rest of the paper discusses the artefacts in this setup, but the conclusions about diffraction effects are also valid for Köhler illumination setup, which is typically used for brightfield imaging and phase-contrast microscopy.

In the epifluorescence setting, since the illumination light from the objective is not perfectly focussed down to a point, it bathes the entire sample in excitation light, and thus the entire sample fluoresces. At the same time, as neither the sample nor the focal plane of the objective are infinitesimally thin, the emitted light is collected from multiple planes along the sample's Z-axis that are within the depth of field of the objective. These planes are all projected onto one another which form the projected final 2D image (Figure 1a). We term both of these as the "projection effects" in this paper, and are a result of the three dimensional nature of the excitation and collection setup used to acquire the 2D images of 3D samples. Projection effects would be largely minimised in the case of alternative imaging modalities, such as confocal or two-photon microscopy, where illumination of sample and collection of photons could be limited to a narrow Z depth. However, we note that such methods do not offer sufficient time-resolution for high throughput live-cell imaging capabilities.

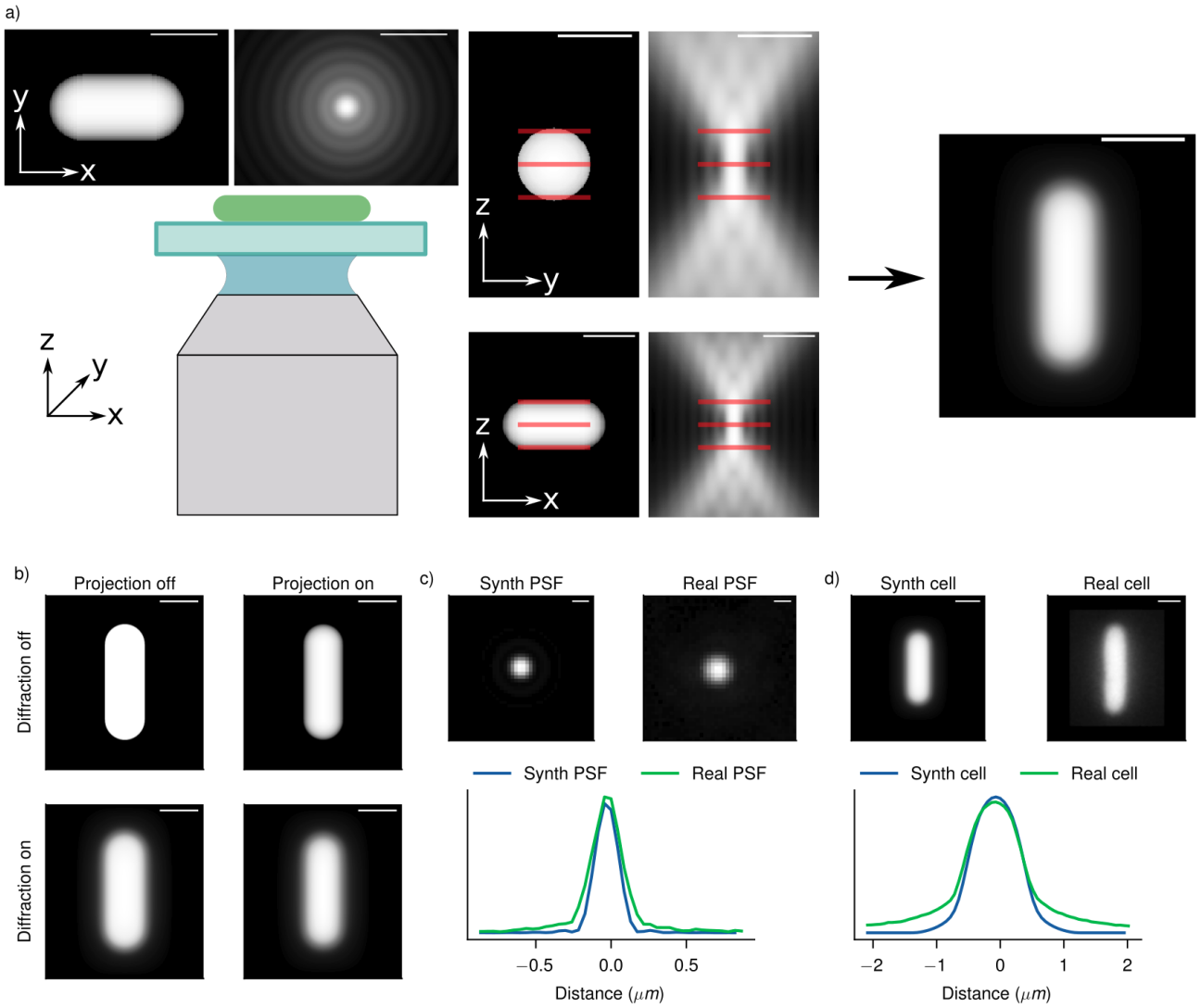

Figure 1 - Widefield fluorescence Image formation:

a) The 3D nature of the sample and the PSF leads to different extent of diffraction effects at different Z depths. The projection of a bacterium and corresponding cross sections of the 3D PSF are shown for three different viewing angles. The final 2D image contains out of focus light from all emitting Z planes. (All scale bars = 1 micron).

b) The SyMBac image simulation platform enables turning the effects of projection or diffraction selectively on or off. (Scale bars = 1 micron).

c) The theoretical point spread function model used in the simulations is narrower than the empirical point-spread function. (Scale bars = 200 nm).

d) As a result of the wider tail of the empirical point spread function, experimental images of cells have much wider tails in their intensity profile and will produce stronger and long-range diffraction blur. (Scale bars = 1 micron).

"Diffraction effects" come from the point spread function (PSF) of the microscope, which blurs the image. Since the size of individual cells of bacteria is within the same order of magnitude as that of a typical PSF (Figure 1a), while individual cells are easily resolvable, accurate quantification of their dimensions and intensity content is difficult. The blurring by the PSF makes it difficult to draw a pixel-perfect mask around a cell, as the convolution of the

PSF with the object leads to a loss of information in the resultant image. This problem will scale with the size of the PSF, whose size is determined by the imaging wavelength and resolution and numerical aperture of the objective lens. Therefore, for a constant objective the signature of diffraction effects is their wavelength dependence. Accordingly, in our digital simulations, we have used PSFs of varying wavelengths to identify to what extent the errors and biases in the estimates originate from, and scale with diffraction. On the other hand, varying the DoF of the imaging setup enables convenient identification of projection effects. The SyMBac image simulation platform also enables us to selectively turn the effects of projection or diffraction on/off, allowing us to observe the effects of each of them separately (Figure 1b). Such non-physical experiments are useful to develop intuitions about the direction of each effect.

It is important to note here that the simulation experiments carried out in the rest of the paper use a theoretical model PSF [23] (unless otherwise stated), which appeared narrower than the empirical PSF, and with shorter tails (Figure 1c). Therefore the effects of diffraction are likely to be much stronger and long-range in real experiments than the simulated results. As expected, we found that the radial intensity profile of individual cells displayed much wider tails in the experimental images, compared to their simulated equivalents (Figure 1d).

Identifying the boundary of a cell and quantifying its dimensions

First, we analyse the effects of projection and diffraction on our ability to accurately quantify the size and shape of a cell from its 2D fluorescence image. Quantifying the shape and dimensions of a bacteria from their microscopic images involves drawing boundaries on the images to identify regions that are within cells. Here, we use synthetic images of bacteria to benchmark the estimates of cell dimensions as a function of imaging depth and the wavelength used for imaging. These synthetic images enable us to perform digital experiments with control and coverage of the parameters that are otherwise difficult if not impossible to access in an experimental setting. Here, we have primarily focussed our analysis on the estimation of cell width, where, being the shortest dimension, these effects are felt most significantly. Accurate width measurement is also crucial because the volume of a cell has squared dependence on the width making the errors in width estimation more significant towards size estimation.

To identify cell outlines, quantitative microbiologists often use a membrane stain, which is a dye that labels the lipid membranes of cells and gives an image of a glowing boundary. Algorithms have been developed to identify the cell boundary from thresholding the images of membrane stained cells or to identify the brightest contour as the cell boundary [36], [37]. By turning off the effects of projection and diffraction in the synthetic images of a membrane stained cell, we can see that projection causes a significant proportion of the 3D intensity to be allocated towards the centre of the cell in its projected image (Figure 2a). This is further illustrated in the radial profile of the intensity in the inset in Figure. 2b. The black profile (projected intensity without any diffraction effect) is shifted inward compared to the dashed line that marks the cell edge. This effect will cause peak finding algorithms or thresholding

algorithms to underestimate the cell dimensions, such as width estimate in this case. Diffraction effects further shift the intensities towards the centre and cause more underestimation, in a wavelength dependent manner (Figure. 2b inset). Interestingly, the diffraction effects are only significant when there is a significant projection effect to shift the intensity distribution. As seen in the comparison of width estimates at various depths, the difference in various wavelengths only becomes apparent when the depth of field is close to the full Z depth of the cell (Figure. 2c). When the imaging depth is comparable to the Z-depth of the cell, the width is maximally underestimated and the extent of underestimation scaled linearly with the wavelength (Figure. 2f). These results suggest that in a typical widefield imaging setup with a dye that fluoresces at 600 nm, one underestimates the width of a cell by 30% for a 1 um wide cell and for a 0.5 um wide cell, one underestimates the width by 50%.

Alternatively, thresholding algorithms are used to segment images of bacterial cells from their fluorescence images, where the cellular cytoplasm contains fluorescently tagged molecules [2], [15], [38]–[41]. The fluorescence from the uniformly distributed emitters in the cell cytoplasm can be used to approximate the 3D volume of the cell. However, the 2D images, that are typically thresholded to outline cells, affect our estimates due to projection effects. In this case, diffraction effects act in the opposite direction and cause the light from the cell to bleed out and make the cells appear wider (Figure. 1b and 2d). To quantitatively assess the effects of projection and diffraction on our ability to extract cell dimension information from these images, we used Otsu's method to threshold synthetic images collected at various DoFs and with varying wavelengths for the PSF. As seen in Figure. 2e, increasing the depth of projection leads to an underestimation of the cell dimensions, while on the other hand, increasing the imaging wavelengths leads to increased blur of the image and causes the dimensions to be overestimated (Figure 2f). With increasing depth of fields the projection error dominates and causes underestimation.

In order to experimentally test the predictions from this analysis, we labelled the peptidoglycan layer of individual cells with two different stains with different emission wavelengths (HADA = 450 nm and RADA = 580 nm). Both of these dyes are fluorescent D-amino acids that are incorporated into the growing peptidoglycan and should have the same physical location and spatial distribution for each cell. However, the radial intensity profiles for these two dyes show that the dye with the longer wavelength (RADA) has the intensity peaks significantly shifted inwards compared to that of the dye with lower wavelength (HADA), as expected from the simulated profiles shown in Figure 2b. We analysed n = 137 cells and found that using the dye with the longer wavelength underestimated the cell widths more than the dye with shorter wavelength (Figure 2j, additional examples in Supporting information section 1). Furthermore, the extent of underestimation increased with the increased thickness of the cell, as expected from the trends shown in Figure 2c. In a similar fashion, we found that estimating the width of the cell from phase contrast microscopy was wavelength dependent, and using blue light emission filters provided sharper images with smaller widths compared red emission filters, minimising the extent of

overestimation from diffraction, as predicted in Figure 2e (Supporting information section 2).

It is important to reiterate that the extent of such biases towards cell dimension estimation in both of these imaging approaches are also strongly dependent on the dimension of the cell itself. As shown in Figure. 2g, the relative width estimation bias from membrane stains decreases as a function of the cell width. The estimates from the cytoplasmic marker show a similar but weaker dependence. The size dependence of these biases makes it difficult to use a calibration for such effects to computationally correct the estimates. In the section below, we discuss possible ways of incorporating the knowledge of these effects to train deep-learning models for accurate estimation of cell dimensions from their 2D images.

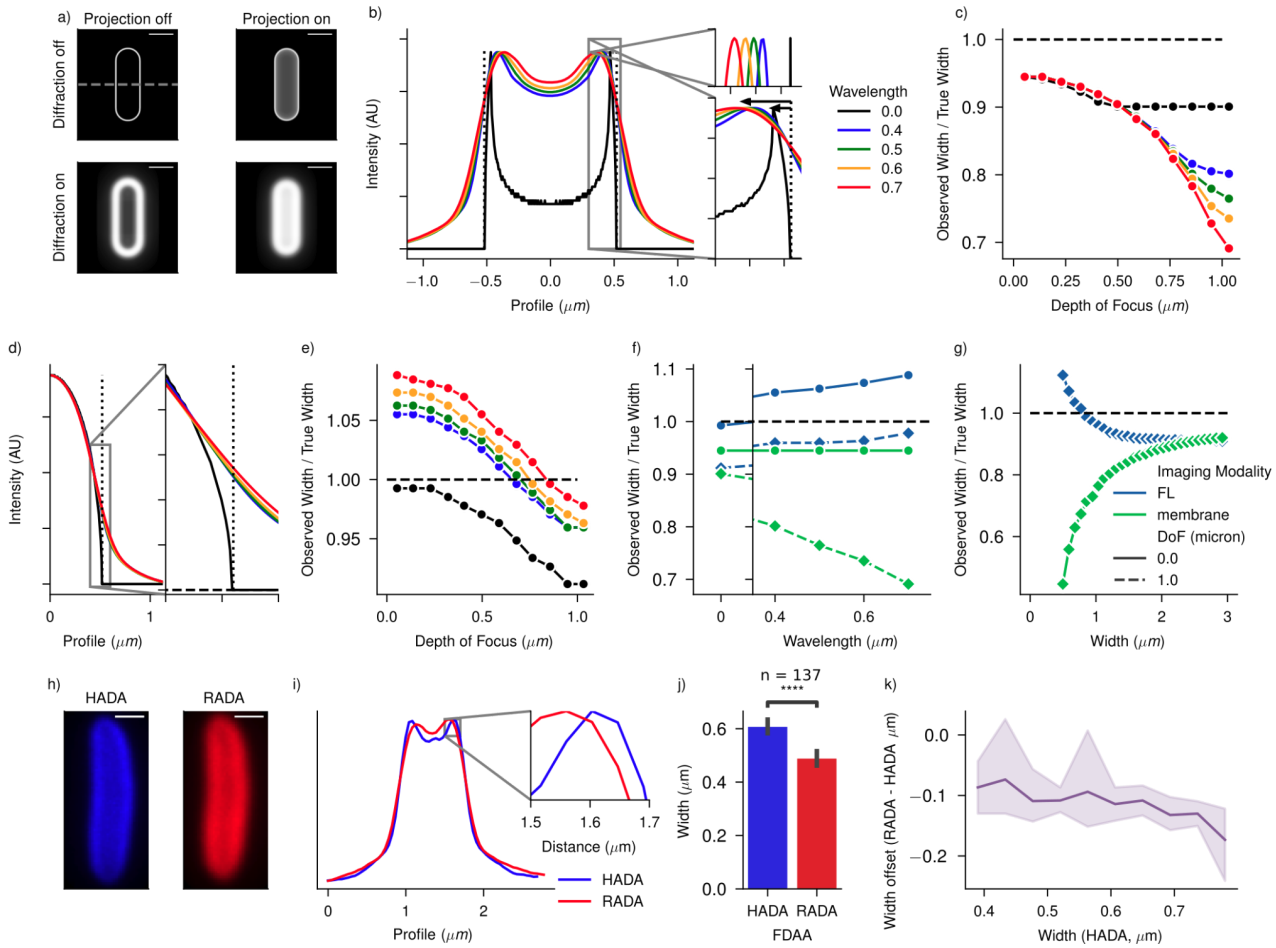

Figure 2 - Effects on cell dimension estimation:

a) Synthetic images of a membrane stained digital cell shows the effects of diffraction and projection on the 2D image (scale bar = 1 μm).
b) We compare the radial profile of the intensity (across the cell width) without diffraction (black) and with diffraction from PSFs at various wavelengths (coloured appropriately). The insets show the direction of shift relative to the true position of the cell edge.
c) The ratio of estimated width to true width is plotted as a function of the depth of focus of the imaging system. Increasing depths leads to an increased proportion of

underestimation caused by the projection effect. When the projection error is significant, the diffraction effects cause further underestimation, which are largest when the depth is comparable to the sample thickness (1 µm here).

d) The radial intensity profile of a cell uniformly filled with emitters shows how projection shifts the light inward and diffraction creates wider tails.
e) Images of uniformly filled cells were simulated with and without the projection and diffraction effects. Increasing projection causes cells to appear narrower, while increasing diffraction causes blurring, making them appear wider (scale bar = 1 µm).
f) Estimated widths from Otsu's threshold divided by the true widths are shown for various wavelengths at various imaging depths. (legend shared with g)
g) Width estimation error as a function of the true width shows opposite but converging effects for membrane stained and uniformly fluorescent cells.
h) The same cell is stained with fluorescent D-amino acids HADA and RADA.
i) i) Intensity traces across the cell's width show qualitatively different structure as a result of different amounts of diffraction. RADA (red) emission is more diffracted than HADA (blue) emission, and the diffraction is biased towards the centre of the cell.
j) Plot showing the average measured width of a single population of cells stained with HADA and RADA (error-bar = 99% CI). The choice of membrane staining colour can affect the mean measured width not only within a single cell, but across a population.
k) Taking the width as measured in the HADA channel as a basis, the discrepancy due to diffraction between the two colours grows as the width decreases (error-bar = 99% CI).

Deep-learning approaches for informed cell segmentation

Deep-learning approaches are rapidly gaining popularity in image segmentation of bacterial cells [42]–[46]. However, the performance of these deep-learning models is strictly dependent on the quality of the training data. Typically these training data are generated through manual annotation of images or using computational segmentation tools followed by expert curation by human users. The analysis presented in the previous section illustrates how computational thresholding algorithms for annotating bacterial cells and generating training data can suffer from size dependent biases from projection and diffraction effects. We further set up a test experiment to benchmark human annotators in terms of their ability to accurately outline the boundaries of cells from their fluorescence images. In this case the users are provided with synthetic images where the ground truth is known and can be used to benchmark their performance. The results from this experiment are presented in Figure 3a, and show that human annotators have variable performance but have a similar bias, which will propagate to deep learning models. Consistent with a similar test done in our earlier work [22], we find that users tend to underestimate cell dimensions. In the current study, using nonphysical experiments, we could identify that the extent of underestimation was correlated with the degree of projection, which is expected from Figure 1b and the analysis presented in Figure 2d. Diffraction effects tend to offset the extent of underestimation, as evidenced from the reduced offset at higher wavelengths.

Imprecision and biases in the training data, from both computationally thresholded or human annotated data, results in misinterpretation of the object-image relationships and translates into imprecise and inaccurate performance of deep learning models. As shown in the analysis below (Figure. 3c and supporting information 3), the highly versatile Omnipose model demonstrates suboptimal performance in segmenting cells from the synthetic fluorescence images, when trained with human annotated images. This effect is reminiscent of the analysis presented in our recent paper [22], where we compared the performance of pretrained deep-learning segmentation models towards phase contrast images of bacterial cells. Segmentation outputs from pretrained deep-learning models retained the biases from their training data and led to large variation in the segmentation output.

However, the deep learning framework and our ability to generate realistic synthetic images with the SyMBac pipeline provides an opportunity to use the information about the projection and diffraction to improve segmentation performance. Since the synthetic image generation pipeline provides access to the accurate ground truth, models trained with these data are able to learn the object-to-image relations more accurately and precisely. To test this hypothesis, we retrained Omnipose with the SyMBac generated synthetic fluorescence images of bacteria with comparable coverage of cell size and shapes. The retrained model segmented cells with higher IoUs and the resultant distribution of estimated cell sizes more closely matched the ground truth cell size distribution (Figure. 3c). Example images and segmentation results are presented in the supporting information section 3. This digital experiment illustrates the potential use of synthetic images to encode the effects of projection and diffraction in the deep-learning model such that it can accurately infer the underlying shape and size of bacterial cells from their images.

Cell-size regulation analysis has recently gained popularity and precise analysis of the size estimates from the analysis approach illustrated above should help in drawing more accurate conclusions [36]. Size regulation analysis often involves comparing the size of cells right before and after cell division, which requires accurate identification of the time-point when a cell divides to become two cells. Bacterial cells start to pinch before division and the projection effects cause cells to appear separated even before they are physically separated (Figure 3e). As a result, pretrained deep-learning segmentation models identify cells as divided much earlier than their actual division times (Fig. 3g). We generated synthetic kymographs of cells during the division process and benchmarked the performance of pretrained omnipose models and omnipose model trained with SyMBac generated data that contain the accurate ground truth information about the time of cell division (Figure 3f). Synthetic data was able to correct the bias of the model, but retained some of the large variability (Figure 3g). This large variability in deciding when a cell has divided stems from the subtle differences in contrasts as a result of the opposing effects of projection and diffraction in such images.

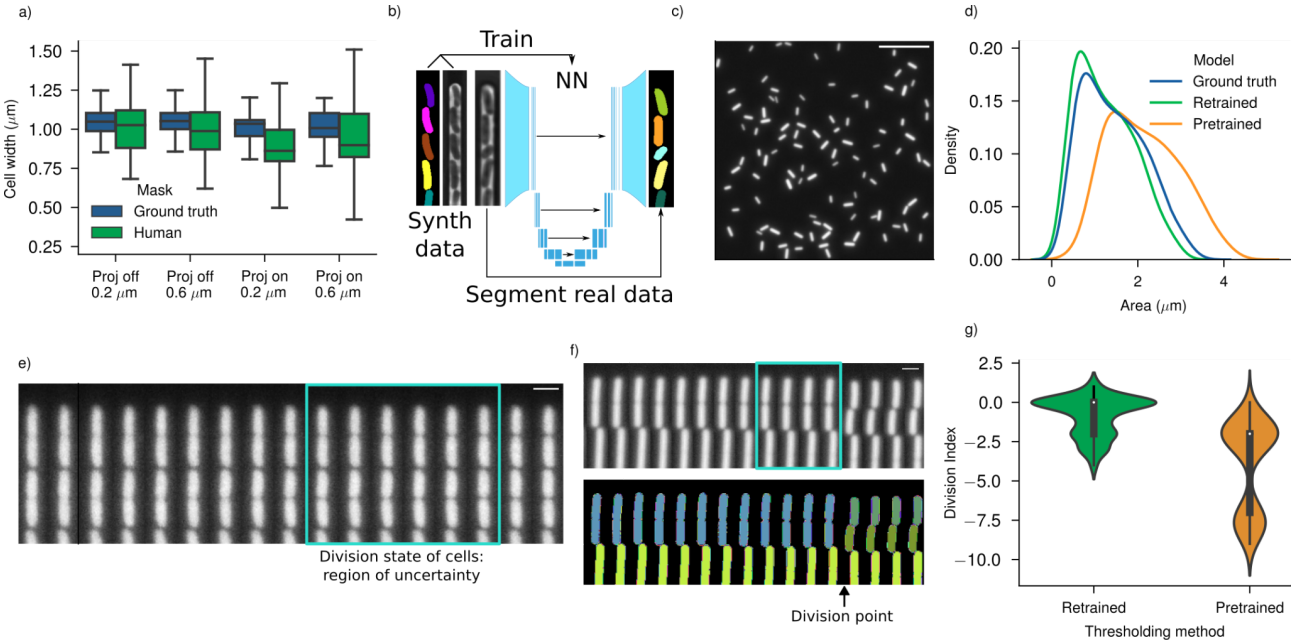

Fig 3 - Cell segmentation with deep-learning models trained on synthetic data:

a) Five human annotators were asked to carefully label cells on a simulated high resolution 100x (0.065 µm/pix) image of fluorescent cells scattered across an agarose pad. 4 datasets were provided, modulating both projection and diffraction. Human annotators consistently underestimated the cell width across all datasets. Such annotation error is propagated to deep learning models when generating training data.

b) To alleviate the effects of human annotation quality, we retrained a popular image segmentation tool, Omnipose, on samples of simulated agar pad data generated using SyMBac. Simulated data comes with perfect ground truth,

c) Example of the agar pad data supplied to the model (scale bar = 15 µm).

d) Comparison of the output distribution of cell sizes shows that the pre-trained model fails to reconstruct the underlying ground truth distribution, whereas the output distribution from the SyMBac trained model closely mimics the underlying ground truth distribution (scale bar = 2 µm).

e) kymographs from high time resolution images of cell division show the process of cells separating (scale bar = 2 µm). The cyan box shows the ambiguous case of cells where it is unclear if they have divided.

f) A synthetic kymograph generated by SyMBac shows similar features and the corresponding ground truth is shown below. The cyan box shows cells which may be mistakenly identified as having divided, while the ground truth shows them as still pre-division.

g) The pretrained Omnipose model detects division points earlier than expected, possibly due to the projection affecting the performance of human annotators. The pretrained model also shows widely variable performance, possibly due to lack of consistent labelling of cells near the division point. Omnipose models retrained with

SyMBac can correct for the bias (achieving the mean ground truth division point) and reduce the variability.

Quantifying the distribution of cellular intensities from widefield images

Fluorescently tagged proteins or transcriptional reporters are used to monitor the expression dynamics of cells over time or across a population of cells. The fluctuations and heterogeneities of the intensity from the fluorescent reporter is used to analyse the sources and consequences of phenotypic heterogeneity. Historically, and for the simplicity of the experimental setup, microcolonies of cells are grown on agar pads to follow the intensity of individual cells over time and to compare the intensities of cells tightly packed in the microcolony at various points of time [2], [47]. However, while simple to set up, such experimental designs, including microfluidic devices where cells are closely packed together [48], [49], cause a serious artefact in gene-expression measurements due to the diffraction effects of the imaging system.

The point spread function (PSF) of an imaging system causes light to be diffracted away from the source. For a cell full of fluorescent emitters, the emitted light detection spreads far beyond the boundaries of the cell and into nearby cells if they are close enough (Figure. 4a-c). The effect of such light bleedthrough due to diffraction is more significant with the empirical point spread function (Figure. 4d). For an isolated cell, this causes the intensity within a cell to be underestimated as a function of the PSF dimension (depending on the wavelength of emitted light and numerical aperture of the objective). However, for tightly packed cells in microcolonies, as shown in Figure. 4d and 4e, it causes misallocation of light to nearby cells. We previously coined the term 'light bleedthrough' to describe this effect [50]. While the diffraction limit of the optical system may not be large enough to make it difficult to resolve a single-cell in a fluorescence image of cell clusters, the light bleedthrough effect from diffraction makes it difficult to resolve their true intensity (the empirical PSF is still sharp, but has long tails).

As a result of the light bleedthrough, cells should appear brighter in larger colonies, as more neighbour cells contribute to the intensities of every individual. As shown in the images of simulated microcolonies, individual isolated cells are much dimmer than cells that are part of a larger microcolony (Figure 4e). In fact, the mean intensity of cells in a microcolony monotonically increases as a function of colony size. At the limit of a colony size of an infinite number of cells, the mean intensity of an individual cell should reach the true intensity, since any intensity lost from a cell is simply allocated to another cell (which itself lost intensity, cancelling the effect out at long ranges). This is clearly seen in the simulated trends of intensity vs colony size plot, where we plot the fraction of recovered intensity (observed intensity/ground truth intensity) on the right axis (Figure 4f). Beyond the overall mean cell intensity of an entire colony, the intensity trends of individual cells within a colony can be predicted from their position within the colony and the number of neighbour cells around them. The closer a cell is to the centre of a large colony and the more contribution it receives from all the neighbours, the brighter it appears in an image (Figure 4g and 4h). Such effects

can be very misleading as it might appear that the intensity of an individual cell was dependent on its interaction with the neighbouring cells and point to plausible biological mechanisms. Studies where similar patterns of intensity distribution were used as an evidence for cell-cell interactions should consider the effects of optics and perform necessary controls to separate biological mechanisms from optical illusions [51]. Digital control experiments using virtual microscopy platforms could provide a key advantage in such circumstances.

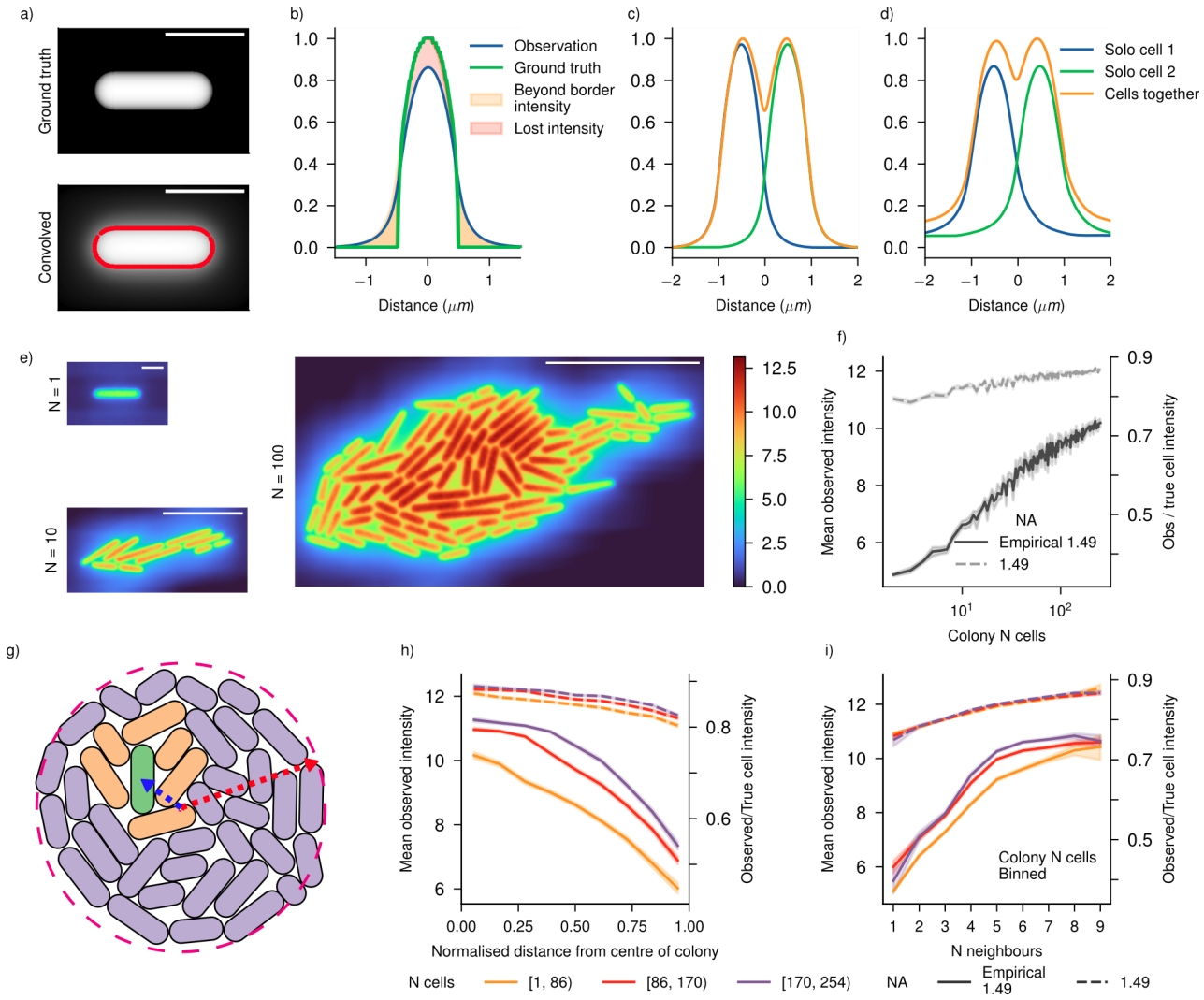

Figure 4 - Diffraction effects on intensity quantification - simulated data:

a) A cell with and without convolution with a synthetic point spread function. Intensity is seen to “leak” or “bleed” through the true outline (shown in red).
b) Comparing the intensity of the raw image against the convolved image shows where the intensity is re-allocated. The maximum intensity is reduced and the trace is given tails which extend beyond the border of the cell.
c) Two cells side by side convolved with the synthetic PSF show somewhat higher mean peak intensity than if the cells were positioned alone (legend shared with d)

d) When convolved with an empirical PSF, the effect of increased peak intensity is amplified.
e) Simulation of a growing microcolony using an empirically measured 100x 1.49NA objective's PSF. All cells in the simulation have identical ground truth intensities. Small microcolony sizes show low cell intensities, while as the colony grows larger, cells in dense regions, such as at the centre of the colony begin receiving more light contribution from surrounding cells, artificially increasing their perceived intensity.
f) As colony size increases the mean observed intensity of each cell in the colony also increases. (Error bars = 99% CI)
g) Schematic representation of a cell (green) some distance from the centre of the microcolony, with its neighbouring cells labelled (orange).
h) Cells at the centre of a colony appear brighter than cells at the periphery due PSF diffraction bleedthrough. (Error bars = 99% CI, data sample averaged for colonies of size 20-250)
i) Cells with more neighbours appear brighter (Error bars = 99% CI, data sample averaged for colonies of size 20-250)

We have performed experiments on microcolonies on agar pads to examine to what extent such effets affect the intensity measurements in typical widefield experiments. Instead of allowing colonies to form through continuous growth from individual cells on agar pads, where the intensity distribution can change over time, we placed a high density of cells on agar pads and imaged them immediately, such that all the cells are sampled from the same intensity distribution, irrespective of the colony they belong to. We picked a transcription reporter of the ribosomal gene expression promoter rpsL, as it has a strong expression level and one of the lowest gene expression noises [52] (CV = 15%, quantified from comparing isolated individual cells). The example snapshots containing cell clusters of various sizes show that as expected from the computational models, the cells in larger colonies appear brighter than the isolated individuals (Figure 5a). To quantify the intensity of individual cells from the colonies, we segmented their images and extracted the intensities of individual cell masks (as described in methods, segmentation example in supporting information 4). When we compare the distribution of intensities of cells from different cluster sizes (here referred as 'preformed colonies'), we find that there is no overlap between the intensity distributions from isolated individuals and cells from larger colonies (Figure 5b). The plot of mean intensity of cells as a function of the colony size they belong to produces a trend (Figure 5c) that is quantitatively comparable to the trend predicted from the computational models of diffraction effects on microcolony images. Similarly, the spatial pattern of intensity dependences on the number of neighbours and position within a colony is consistent with the model predictions (Figure 5d and e). The effects are stronger for larger colonies but are also significantly strong for relatively small colonies.

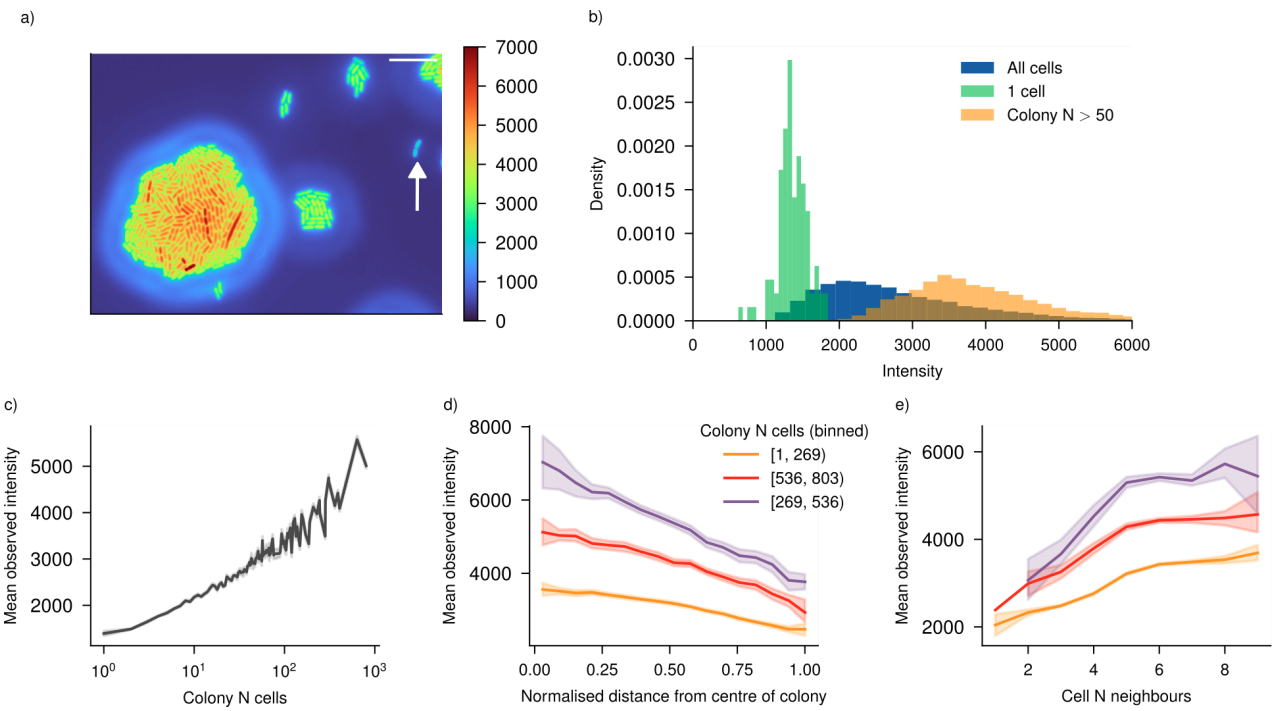

Figure 5: Diffraction effects on intensity quantification - experimental data

a) An image taken of a real agarose pad showing microcolonies of various sizes, along with a single cell (white arrow). Important features to note are the similarity to the simulated data shown in figure 4.
b) The intensity distribution of cells depends on which cells are sampled. Single cells have low mean intensities, cells from colonies greater than 50 in size have higher mean intensities.
c) Experimental data shows average cell intensity increases with colony population size on large scales.
d) Average cell intensity decreases towards the periphery of the colony in real data. The effect is also amplified for large colony sizes.
e) Increased neighbour number causes observed intensity to increase in experimental data.

The effect of the PSF on the variance of the intensity distribution of cells in a patch can be even more counter-intuitive. We used SyMBac to generate synthetic images of cells on agar-pad type microcolonies, where each cell contains randomly sampled emitters within their convex hulls. We then compared the resultant intensity distribution of cells before and after the PSF convolution. When there is no ground truth variability across the population (coefficient of variation, CV = 0), the bleedthrough contribution to individual cells depends on their position in the microcolony and the number of neighbour cells around them (as seen in Figure. 4h and i). Since this is variable across the colony, the bleedthrough effect causes the cells to appear artificially variable, and leads to an estimated CV of 12% in this case (Figure 6a - top row). On the other hand, the PSF can act as a smoothing filter to reduce the variability when the variation among the cells is large (CV = 22 %), by taking intensity away from bright cells to the dimmer cells (Figure 6a - bottom row). This leads to underestimating the

underlying variability (CV estimated = 14%). These two extremes highlight the problem of diffraction effects in the timelapse quantification of gene-expression variability, as it leads to similar results from largely variable underlying distributions and makes it difficult to correct for (additional examples presented in supporting information 5).

To evaluate this CV dependent effect from the light bleedthrough, we evaluated thousands of synthetic bacterial microcolonies and compared the estimated CVs to their underlying true CVs. We see that below a threshold CV of 0.15, the bleedthrough effect always overestimated the CV, while above this threshold CVs were always underestimated. While this suggests that estimating the true underlying CV from the measured CV from a microcolony is near impossible, one could take advantage of the knowledge of the colony's shape and the organisation of the cells within it to improve the estimates. Since the light bleedthrough contribution depends on the organisation of the cells within the microcolony and the dimensions of the PSF, we could use a model PSF to deconvolve the image to recover the underlying CV. As shown in the supporting information figure section 6, deconvolving the image of a simulated microcolony improves our performance to estimate the true CV when we consider a 2D PSF. This example illustrates the potential use of the knowledge of the diffraction effect to mitigate the artefacts. However this demonstrates a best case scenario, as we have perfect knowledge of the PSF. Most importantly, the deconvolution based approach for removing diffraction effects only seems to work when we use 2D PSF and not so well when we use full 3D PSF model. In our experience using the empirical PSF model for deconvolving experimental colony images to correct for the diffraction effect is usually not effective. Instead, we recommend that based on this knowledge, one could design their experiment differently, such as using a structured microfluidic platform where cells are kept apart from each other to minimise the bleedthrough effects [50]. Picking the right spacing between cells will effectively eliminate this effect and enable users to accurately estimate the heterogeneity and fluctuation dynamics of intensities from individual cells [53].

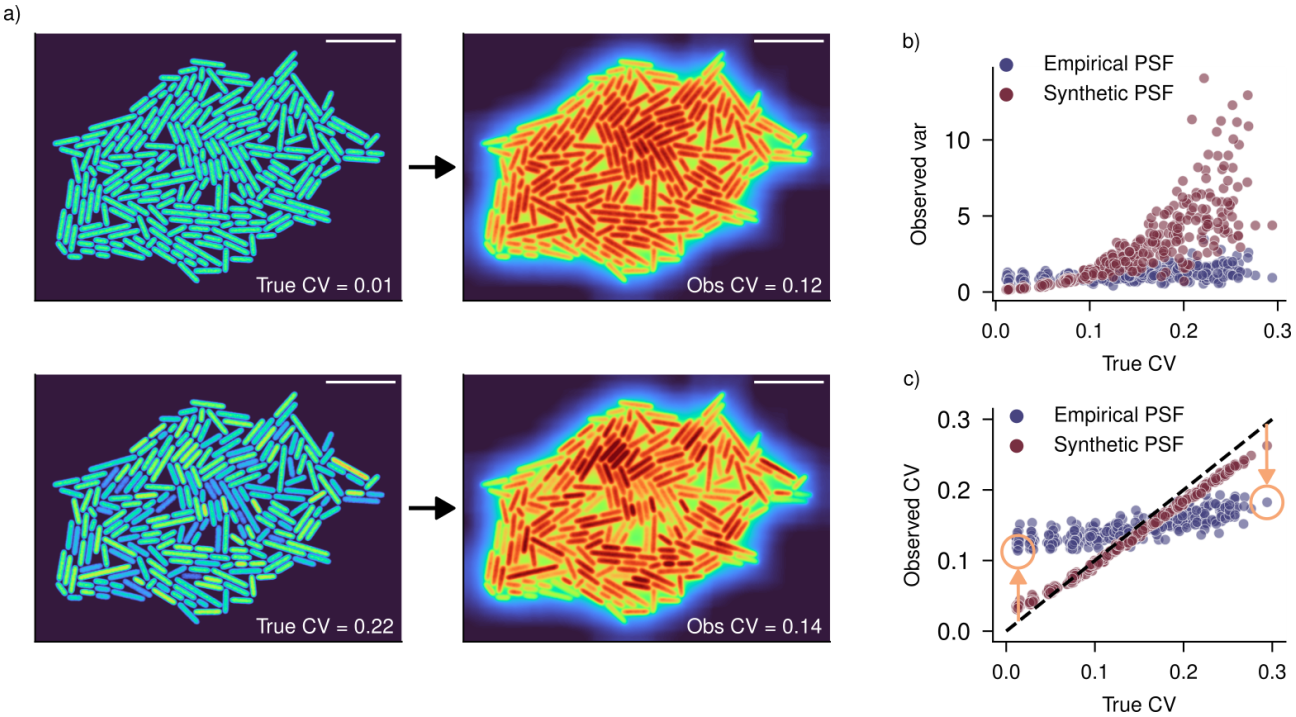

Figure 6: Miscalculation of gene expression noise (CV) due to PSF diffraction bleedthrough

a) Two simulated colonies with a true (cell intensity) CV of 0.01 and 0.22 respectively, along with their convolution with the empirical PSF showing that their observed CVs are now very similar, despite markedly different underlying cell intensity distributions.
b) When the true CV of a simulated microcolony is high, the variance appears unpredictable when simulated with a synthetic PSF, most likely due to effects from colony shape. The empirical PSF has long tails and so spreads intensity to such an extent as to reduce this effect.
c) PSF convolution acts in two directions when estimating the CV of a microcolony. At low CVs, where the underlying cell intensity distribution is uniform, the PSF causes central cells to appear bright and peripheral cells to appear dim (increasing CV), while when the CV is high, the PSF is a blurring filter, lowering the variance in the population.

Counting individual molecules from fluorescence imaging

If the artefacts from bleed-through effects are minimised through the use of appropriate imaging platform and an accurate model of PSF is used for deconvolution, quantifying signals from high-copy fluorescent reporters should be relatively straightforward when the signal is orders of magnitude higher than the cellular autofluorescence. However, when the signal is comparable to the autofluorescence background, translating the signal to abundances of proteins becomes quite challenging. For counts below 100 copies per cell, methods such as background deconvolution have been used, but such methods do not have single cell resolution as the entire distribution is deconvolved with the distribution of

autofluorescence levels [54]. The deconvolution methods also do not work for low-copy number proteins (copy number <10) that are often the most important regulators of gene-expression, such as gene copy number, transcription factors (TFs), mRNAs. etc.

To reliably quantify the abundances of low-copy proteins, we need to count individual fluorescently tagged molecules in a cell [55]. While it is possible to image single molecules of any copy number protein by making only a subset of them visible at a time using techniques which leverage photoactivatable proteins such as PALM or STORM, counting the true copy number of individual proteins is not as straightforward. The problem of counting stems from a combination of the diffraction limit and 2D projection of the particles distributed in 3D. Individual tagged molecules show up as a diffraction limited blurred spot in the image depending on their position with respect to the focal plane (Figure 7a). This means, any number of spots that are within a distance less than the diffraction limit will get combined into one spot and any spot that is out of focus will be difficult to detect. The combination of projection and diffraction effects makes it difficult to count even a small number of molecules in a bacterial cell.

In Figure. 7b, we illustrate the extent of undercounting from these two effects. We employ a 'naive counting' criteria, detailed in the Methods section, which counts: the number of molecules that are out of focus, the number that gets undercounted due to being within diffraction limit of another spot, and the sum of any remaining resolvable single molecules and clusters of molecules (appearing as a single molecule due to diffraction limit) as countable fraction. This approach enables us to identify the sources of error in our measurements as we change the cell dimensions or experimental settings. As shown in Figure 7b, when the count is small (n = 5), a small fraction is lost due to diffraction, as the spots are typically far apart from each other. However, if the thickness of the cell is increased, a large fraction of spots are undetectable due to the defocused PSF. Conversely, if microfluidic platforms can be used to flatten the cells, a larger number of spots can be brought into focus. Additionally, the increased cross-section in the imaging plane further reduces the undercounting effect caused by the diffraction limit. This is consistent with the previous work by Okumus et al (Microfluidics-Assisted Cell Screening (MACS)) [11], which demonstrated that flattening the cells using microfluidics enables more accurate counting of higher copies of molecules in individual cells. As the number of molecules increases, the proportions of undercounting from the diffraction limit increases, while the proportion lost due to defocus remains fixed as it is simply proportional to the volume fraction of the cell outside the focal plane (Figure 7b). One could imagine that the randomisation of spot positions caused by their diffusion would enable one to pick the snapshot with the configuration where most molecules are separated from each other (supporting information section 7). We note that the timescales of such measurements must be faster than the copy-number fluctuation (possibly with stroboscopy) while enabling sufficient diffusion of molecules between frames. However, for techniques such as MACS or fixed cells, the diffusive motion of molecules are largely minimised, and renders this approach difficult to implement.

Due to the combination of projection and diffraction effects, in a normal sized bacterium, a single snapshot can even mis-count 2 molecules as 1 for 10 % of the time (Figure S7 in supporting information section 7). The performance rapidly declines as the number of molecules in the cell are increased (Figure 7c). However, one could design an image analysis pipeline which exploits the knowledge of defocus and spatial patterning and thus improves the performance of counting. To explore this possibility, we trained a popular convolutional neural network estimator used for detecting single molecule images called Deep-STORM. When trained with simulated synthetic images containing appropriate samples of defocused images, the Deep-STORM model was able to reliably detect accurate counts of molecules for up to 15 copies per cell (Figure 7c). This is a remarkable feat, when considering the naive counting estimates and the empirical performance demonstrated in previous work in this area [11]. Similar to the deep-learning models for cell segmentation (discussed in the section on Deep-learning approaches for informed cell segmentation), when trained with realistic synthetic data, deep-learning models for single-molecule detection are able to vastly improve the performance. However, we note that the performance of the model declined with the reduction of signal-to-noise ratio (SNR). For SNR = 5, the Deep-STORM model consistently undercounted the number of molecules for even lower copy numbers, even though the extent of undercounting was lower than the naive estimates. To investigate the source of the errors from low SNR, we performed a non-physical simulation experiment by placing all the molecules in focus and found that the performance at such low SNR (SNR = 5) became comparable to the performance in very high SNR. From this experiment, we concluded that DeepSTORM was robust to increased density of spots but requires high SNR to achieve better performance on defocused spots. The ability of these models to exploit the shape of a defocused spot to use as a signature of their Z position appears to be strongly dependent on the SNR. At low SNR, a defocused spot, which has its intensity spread across a larger area, becomes difficult to detect beyond the background. As demonstrated in the previous analysis (Figure 7b), one can reduce the fraction of defocused spots by flattening the cells, using platforms such as MACS, and can emulate the situation where most spots are in focus. Indeed, Deep-STORM models, when implemented in a simulated MACS type scenario, performed reliably up to a very large number of molecules per cell (Figure 7d). The lower error rates in this scenario is also a result of the wider area of the flattened cells, which reduces the diffraction errors from spots overlapping with each other in the image plane.

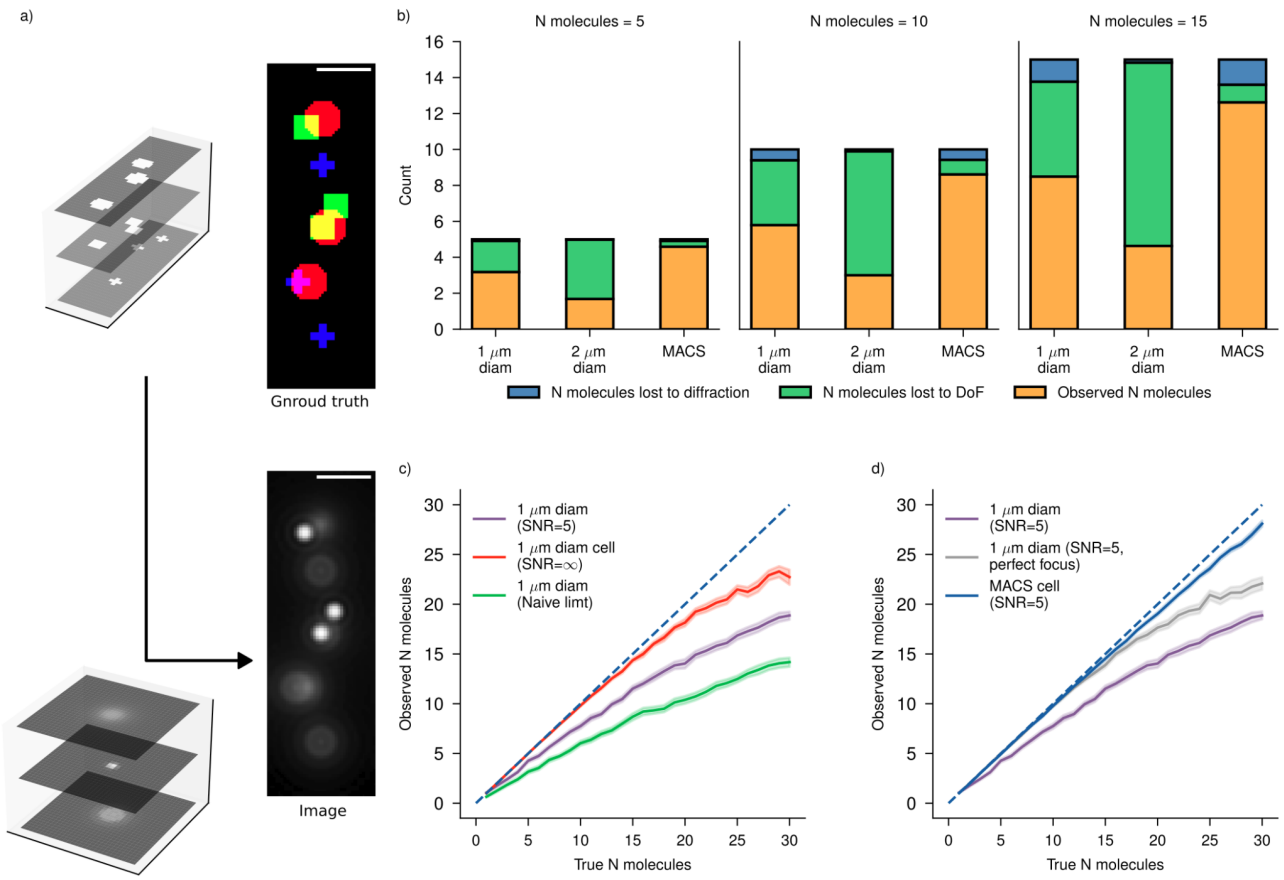

Figure 7 - undercounting errors from the combined effect of projection and diffraction:

a) Molecules in a cell exist in three dimensions. Shown are 9 molecules in a digital cell and their ground truth positions projected in x-y. Upon convolution with the PSF, the resultant image is a projection of all 9 molecules, of which only 3 are in focus. The remaining 6 molecules are dim and hard to detect.

b) Simulated sampling of points within a digital cell shows naive detection bounds. If two points were more than 250 nm beyond the focal plane, they were considered too dim to localise, and so are lost due to depth of field. Additionally, if two molecules were within 1 Rayleigh diffraction limit (λ = 550nm, NA = 1.49) of one another, they are considered a single object, and so undercounting occurs, thus these molecules are lost due to diffraction. The remaining population of molecules are considered countable. 3 cell types are simulated, a typical spherocylindrical cell with a diameter 1 µm, a large cell with a 2 µm diameter, and a cell trapped in the Microfluidics Assisted Cell Screening (MACS) platform, which is squashed down to a thickness of 660 nm, and a width of 2 µm. All cells are 5 µm long.

c) Advances in deep learning have pushed single molecule microscopy image analysis beyond these naive bounds. 3 cell types are simulated with increasing numbers of uniformly distributed fluorescent emitters and then processed using Deep-STORM for single molecule localisation. The naive limit is compared with a typical experiment: a cell being imaged with a peak signal to noise ratio of 5 (typical of single molecule

experiments in bacteria), and this is compared with the ideal case where there is no noise (SNR = ∞).

d) Two additional cell types are compared, a cell trapped in the MACS platform, and the hypothetical case of a cell where all molecules are in perfect focus. Based on the naive limits, one would expect a larger difference between perfect focus case, and the base case. It is possible that the Deep-STORM model is learning to use defocus as a hallmark to improve localisation performance.

Summary and outlook

Quantitative interpretations of micrographs are crucial for the use of microscopy in quantitative microbiology. In this paper, we have used digital experiments to explore the effects of projection and diffraction towards image formation, which limits our ability to accurately interpret the results. The predictions from our model were also validated using appropriate targeted experiments.

The analysis presented here shows that cell size estimation by both traditional segmentation techniques and machine learning are biased and difficult to correct for. The direction of bias depends on the type of labelling approach, imaging settings, and even on the size of the cell itself. Therefore, it is difficult to use calibration to correct for the resultant bias in size estimation. However, such biases can be partially alleviated through the use of synthetic training data which captures both these effects. In this paper, we have primarily focused on fluorescence imaging of bacterial cells in the analysis of cell segmentation. However, the results and conclusions are equally valid for the brightfield and phase contrast images of bacteria, as they suffer from similar projection and diffraction effects.

Fluorescence microscopy is an excellent tool for quantifying the dynamics and heterogeneity of gene-expression in live bacterial cells from intensity measurements. Timelapse imaging of microcolonies on agar pads and microfluidic devices is often used to study gene expression dynamics in living cells [47]–[49], [56]. Using digital image simulations and experimental fluorescence imaging of samples with multiple cells, we show that quantification of the true fluorescence signal of cells in clusters ('microcolonies') is corrupted by diffraction due to misallocation of light intensity from neighbouring cells. This can affect the estimation of expression variation and correlation analysis using such platforms [9], [51], [57], [58]. To reduce (but not entirely eliminate) this artefact, the image of the sample can be deconvolved with an appropriate model of the point spread function. However, it will be unknown how close the deconvolution output is to the ground truth, and this will be heavily dependent on the accuracy of the deconvolution kernel. Alternatively, structured imaging platforms, such as microfluidic devices, where cells can be maintained at a fixed distance from each other, can also help to reduce such artefacts [50], [53]. The quantification of the number of particles (such as counting mRNA or proteins) within individual cells is also affected for similar reasons, making it difficult to accurately count n>10 individual molecules. We have suggested and illustrated combinations of alternative experimental designs and

deep-learning based analysis protocols that harness the knowledge of these effects to alleviate the resultant bias and errors.

Put together, our results show that knowledge of the effects from projection and diffraction is of critical importance to extract meaningful information about the underlying microbiology, whether it is shape and size of cells or their content from intensity measurements or single-molecule counting. We believe that digital experiments using virtual microscopy platforms should be routinely used to test the limitations of the experimental design and possible optical illusions to ensure correct interpretations of the imaging data. The analysis and discussion presented in this paper should guide improved experiment design and help with quantitative interpretation of microscopy experiments in microbiology.

Acknowledgements

We thank Prof. Ricardo Henriques for his feedback on this work and the members of Bakshi Lab for their help with the image annotation test. All figures included in the present paper are original and contain no third party material.

Funding

This research in the SB laboratory was supported by the Wellcome Trust Award (grant number RG89305), a University Startup Award for Lectureship in Synthetic Biology (grant number NKXY ISSF3/46) and a seed fund from the School of Technology. GH was supported by the United Kingdom Biotechnology and Biological Sciences (BBSRC) University of Cambridge Doctoral Training Partnership 2 (BB/M011194/1).

Data and code availability

All code written for this paper, and used to generate figures is uploaded to https://github.com/georgeoshardo/projection_diffraction_frontiers_in_physics. For backwards compatibility, the version of SyMBac used in this paper has been frozen and included in the code repository for this paper.

The most updated version of SyMBac is available at: https://github.com/georgeoshardo/SyMBac

Experimental data analysed in this paper can be found at [59].

References

[1] D. C. Fung and J. A. Theriot, ‘Imaging techniques in microbiology’, *Curr. Opin. Microbiol.*, vol. 1, no. 3, pp. 346–351, Jun. 1998, doi: 10.1016/S1369-5274(98)80040-4.

[2] J. W. Young *et al.*, ‘Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy’, *Nat. Protoc.*, vol. 7, no. 1, Art. no. 1, Jan. 2012, doi: 10.1038/nprot.2011.432.

[3] Z. Gitai, ‘New fluorescence microscopy methods for microbiology: sharper, faster, and quantitative’, *Curr. Opin. Microbiol.*, vol. 12, no. 3, pp. 341–346, Jun. 2009, doi: 10.1016/j.mib.2009.03.001.

[4] S. Iyer-Biswas *et al.*, ‘Scaling laws governing stochastic growth and division of single bacterial cells’, *Proc. Natl. Acad. Sci.*, vol. 111, no. 45, pp. 15912–15917, Nov. 2014, doi: 10.1073/pnas.1403232111.

[5] S. Taheri-Araghi *et al.*, ‘Cell-Size Control and Homeostasis in Bacteria’, *Curr. Biol.*, vol. 25, no. 3, pp. 385–391, Feb. 2015, doi: 10.1016/j.cub.2014.12.009.

[6] I. Soifer, L. Robert, and A. Amir, ‘Single-Cell Analysis of Growth in Budding Yeast and Bacteria Reveals a Common Size Regulation Strategy’, *Curr. Biol.*, vol. 26, no. 3, pp. 356–361, Feb. 2016, doi: 10.1016/j.cub.2015.11.067.

[7] F. Si, G. Le Treut, J. T. Sauls, S. Vadia, P. A. Levin, and S. Jun, ‘Mechanistic Origin of Cell-Size Control and Homeostasis in Bacteria’, *Curr. Biol.*, vol. 29, no. 11, pp. 1760-1770.e7, Jun. 2019, doi: 10.1016/j.cub.2019.04.062.

[8] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox, ‘Real-Time Kinetics of Gene Activity in Individual Bacteria’, *Cell*, vol. 123, no. 6, pp. 1025–1036, Dec. 2005, doi: 10.1016/j.cell.2005.09.031.

[9] M. J. Dunlop, R. S. Cox, J. H. Levine, R. M. Murray, and M. B. Elowitz, ‘Regulatory activity revealed by dynamic correlations in gene expression noise’, *Nat. Genet.*, vol. 40, no. 12, Art. no. 12, Dec. 2008, doi: 10.1038/ng.281.

[10] N. D. Lord, T. M. Norman, R. Yuan, S. Bakshi, R. Losick, and J. Paulsson, ‘Stochastic antagonism between two proteins governs a bacterial cell fate switch’, *Science*, vol. 366, no. 6461, pp. 116–120, Oct. 2019, doi: 10.1126/science.aaw4506.

[11] B. Okumus *et al.*, ‘Mechanical slowing-down of cytoplasmic diffusion allows in vivo counting of proteins in individual cells’, *Nat. Commun.*, vol. 7, no. 1, Art. no. 1, May 2016, doi: 10.1038/ncomms11641.

[12] S.-H. Lee, J. Y. Shin, A. Lee, and C. Bustamante, ‘Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy (PALM)’, *Proc. Natl. Acad. Sci.*, vol. 109, no. 43, pp. 17436–17441, Oct. 2012, doi: 10.1073/pnas.1215175109.

[13] G. C. Rollins, J. Y. Shin, C. Bustamante, and S. Pressé, ‘Stochastic approach to the molecular counting problem in superresolution microscopy’, *Proc. Natl. Acad. Sci.*, vol. 112, no. 2, pp. E110–E118, Jan. 2015, doi: 10.1073/pnas.1408071112.

[14] A. Ducret, E. M. Quardokus, and Y. V. Brun, ‘MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis’, *Nat. Microbiol.*, vol. 1, no. 7, Art. no. 7, Jun. 2016, doi: 10.1038/nmicrobiol.2016.77.

[15] A. Paintdakhi *et al.*, ‘Oufti: an integrated software package for high-accuracy, high-throughput quantitative microscopy analysis’, *Mol. Microbiol.*, vol. 99, no. 4, pp. 767–777, 2016, doi: 10.1111/mmi.13264.

[16] S. Stylianidou, C. Brennan, S. B. Nissen, N. J. Kuwada, and P. A. Wiggins, ‘SuperSegger:

robust image segmentation, analysis and lineage tracking of bacterial cells', *Mol. Microbiol.*, vol. 102, no. 4, pp. 690–700, 2016, doi: 10.1111/mmi.13486.

[17] R. Hartmann *et al.*, 'Quantitative image analysis of microbial communities with BiofilmQ', *Nat. Microbiol.*, vol. 6, no. 2, Art. no. 2, Feb. 2021, doi: 10.1038/s41564-020-00817-4.

[18] K. Rossmann, 'Point Spread-Function, Line Spread-Function, and Modulation Transfer Function', *Radiology*, vol. 93, no. 2, pp. 257–272, Aug. 1969, doi: 10.1148/93.2.257.

[19] M. C. Portillo, J. W. Leff, C. L. Lauber, and N. Fierer, 'Cell Size Distributions of Soil Bacterial and Archaeal Taxa', *Appl. Environ. Microbiol.*, vol. 79, no. 24, pp. 7610–7617, Dec. 2013, doi: 10.1128/AEM.02710-13.

[20] Y. Zhang and H. Gross, 'Systematic design of microscope objectives. Part I: System review and analysis', *Adv. Opt. Technol.*, vol. 8, no. 5, pp. 313–347, Oct. 2019, doi: 10.1515/aot-2019-0002.

[21] S. Bakshi, B. P. Bratton, and J. C. Weisshaar, 'Subdiffraction-Limit Study of Kaede Diffusion and Spatial Distribution in Live Escherichia coli', *Biophys. J.*, vol. 101, no. 10, pp. 2535–2544, Nov. 2011, doi: 10.1016/j.bpj.2011.10.013.

[22] G. Hardo, M. Noka, and S. Bakshi, 'Synthetic Micrographs of Bacteria (SyMBac) allows accurate segmentation of bacterial cells using deep neural networks', *BMC Biol.*, vol. 20, no. 1, p. 263, Nov. 2022, doi: 10.1186/s12915-022-01453-6.

[23] S. F. Gibson and F. Lanni, 'Experimental test of an analytical model of aberration in an oil-immersion objective lens used in three-dimensional light microscopy', *JOSA A*, vol. 9, no. 1, pp. 154–166, Jan. 1992, doi: 10.1364/JOSAA.9.000154.

[24] N. Otsu, 'A Threshold Selection Method from Gray-Level Histograms', *IEEE Trans. Syst. Man Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979, doi: 10.1109/TSMC.1979.4310076.

[25] T. J. Rudge, P. J. Steiner, A. Phillips, and J. Haseloff, 'Computational Modeling of Synthetic Microbial Biofilms', *ACS Synth. Biol.*, vol. 1, no. 8, pp. 345–352, Aug. 2012, doi: 10.1021/sb300031n.

[26] S. Prigent *et al.*, 'SPITFIR(e): a supermaneuverable algorithm for fast denoising and deconvolution of 3D fluorescence microscopy images and videos', *Sci. Rep.*, vol. 13, no. 1, Art. no. 1, Jan. 2023, doi: 10.1038/s41598-022-26178-y.

[27] S. van der Walt *et al.*, 'scikit-image: image processing in Python', *PeerJ*, vol. 2, p. e453, Jun. 2014, doi: 10.7717/peerj.453.

[28] N. Sofroniew *et al.*, 'napari: a multi-dimensional image viewer for Python'. Zenodo, Nov. 03, 2022. doi: 10.5281/zenodo.7276432.

[29] 'Deep-STORM: super-resolution single-molecule microscopy by deep learning'. https://opg.optica.org/optica/fulltext.cfm?uri=optica-5-4-458&id=385495 (accessed May 13, 2023).

[30] L. von Chamier *et al.*, 'Democratising deep learning for microscopy with ZeroCostDL4Mic', *Nat. Commun.*, vol. 12, no. 1, Art. no. 1, Apr. 2021, doi: 10.1038/s41467-021-22518-0.

[31] E. Kuru *et al.*, 'In Situ Probing of Newly Synthesized Peptidoglycan in Live Bacteria with Fluorescent D-Amino Acids', *Angew. Chem. Int. Ed.*, vol. 51, no. 50, pp. 12519–12523, 2012, doi: 10.1002/anie.201206749.

[32] Y.-P. Hsu *et al.*, 'Full color palette of fluorescent D-amino acids for in situ labeling of bacterial cell walls', *Chem. Sci.*, vol. 8, no. 9, pp. 6313–6321, Aug. 2017, doi: 10.1039/C7SC01800B.

[33] E. Kuru, S. Tekkam, E. Hall, Y. V. Brun, and M. S. Van Nieuwenhze, 'Synthesis of fluorescent D-amino acids and their use for probing peptidoglycan synthesis and bacterial growth in situ', *Nat. Protoc.*, vol. 10, no. 1, Art. no. 1, Jan. 2015, doi: 10.1038/nprot.2014.197.

[34] 'Measuring a Point Spread Function', *iBiology*. https://www.ibiology.org/talks/measuring-a-point-spread-function/ (accessed May 10, 2023).

[35] K. Thorn, 'A quick guide to light microscopy in cell biology', *Mol. Biol. Cell*, vol. 27, no. 2, pp. 219–222, Jan. 2016, doi: 10.1091/mbc.e15-02-0088.

[36] T. Ursell *et al.*, 'Rapid, precise quantification of bacterial cellular dimensions across a genomic-scale knockout library', *BMC Biol.*, vol. 15, no. 1, p. 17, Feb. 2017, doi: 10.1186/s12915-017-0348-8.

[37] J. H. Smit, Y. Li, E. M. Warszawik, A. Herrmann, and T. Cordes, 'ColiCoords: A Python package for the analysis of bacterial fluorescence microscopy data', *PLOS ONE*, vol. 14, no. 6, p. e0217524, Jun. 2019, doi: 10.1371/journal.pone.0217524.

[38] J. M. Guberman, A. Fay, J. Dworkin, N. S. Wingreen, and Z. Gitai, 'PSICIC: Noise and Asymmetry in Bacterial Division Revealed by Computational Image Analysis at Sub-Pixel Resolution', *PLOS Comput. Biol.*, vol. 4, no. 11, p. e1000233, Nov. 2008, doi: 10.1371/journal.pcbi.1000233.

[39] T. R. Jones *et al.*, 'CellProfiler Analyst: data exploration and analysis software for complex image-based screens', *BMC Bioinformatics*, vol. 9, no. 1, p. 482, Nov. 2008, doi: 10.1186/1471-2105-9-482.

[40] I. Mekterović, D. Mekterović, and željka Maglica, 'BactImAS: a platform for processing and analysis of bacterial time-lapse microscopy movies', *BMC Bioinformatics*, vol. 15, no. 1, p. 251, Jul. 2014, doi: 10.1186/1471-2105-15-251.

[41] J. M. H. Goudsmits, A. M. van Oijen, and A. Robinson, 'A Tool for Alignment and Averaging of Sparse Fluorescence Signals in Rod-Shaped Bacteria', *Biophys. J.*, vol. 110, no. 8, pp. 1708–1715, Apr. 2016, doi: 10.1016/j.bpj.2016.02.039.

[42] C. Spahn *et al.*, 'DeepBacs for multi-task bacterial image analysis using open-source deep learning approaches', *Commun. Biol.*, vol. 5, no. 1, Art. no. 1, Jul. 2022, doi: 10.1038/s42003-022-03634-z.

[43] J. T. Sauls *et al.*, 'Mother machine image analysis with MM3'. bioRxiv, p. 810036, Oct. 18, 2019. doi: 10.1101/810036.

[44] O. M. O'Connor, R. N. Alnahhas, J.-B. Lugagne, and M. J. Dunlop, 'DeLTA 2.0: A deep learning pipeline for quantifying single-cell spatial and temporal dynamics', *PLOS Comput. Biol.*, vol. 18, no. 1, p. e1009797, Jan. 2022, doi: 10.1371/journal.pcbi.1009797.

[45] J. Ollion and C. Ollion, 'DistNet: Deep Tracking by displacement regression: application to bacteria growing in the Mother Machine'. arXiv, Sep. 19, 2020. doi: 10.48550/arXiv.2003.07790.

[46] K. J. Cutler *et al.*, 'Omnipose: a high-precision morphology-independent solution for bacterial cell segmentation', *Nat. Methods*, vol. 19, no. 11, Art. no. 11, Nov. 2022, doi: 10.1038/s41592-022-01639-4.

[47] J. C. W. Locke and M. B. Elowitz, 'Using movies to analyse gene circuit dynamics in single cells', *Nat. Rev. Microbiol.*, vol. 7, no. 5, Art. no. 5, May 2009, doi: 10.1038/nrmicro2056.

[48] A. Prindle, P. Samayoa, I. Razinkov, T. Danino, L. S. Tsimring, and J. Hasty, 'A sensing array of radically coupled genetic "biopixels"', *Nature*, vol. 481, no. 7379, Art. no. 7379,

Jan. 2012, doi: 10.1038/nature10722.
[49] G. Ullman, M. Wallden, E. G. Marklund, A. Mahmutovic, I. Razinkov, and J. Elf, 'High-throughput gene expression analysis at the level of single proteins using a microfluidic turbidostat and automated cell tracking', *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 368, no. 1611, p. 20120025, Feb. 2013, doi: 10.1098/rstb.2012.0025.
[50] G. Hardo and S. Bakshi, 'Challenges of analysing stochastic gene expression in bacteria using single-cell time-lapse experiments', *Essays Biochem.*, vol. 65, no. 1, pp. 67–79, Apr. 2021, doi: 10.1042/EBC20200015.
[51] T. Julou *et al.*, 'Cell–cell contacts confine public goods diffusion inside Pseudomonas aeruginosa clonal microcolonies', *Proc. Natl. Acad. Sci.*, vol. 110, no. 31, pp. 12577–12582, Jul. 2013, doi: 10.1073/pnas.1301428110.
[52] A. Urchueguía, L. Galbusera, D. Chauvin, G. Bellement, T. Julou, and E. van Nimwegen, 'Genome-wide gene expression noise in Escherichia coli is condition-dependent and determined by propagation of noise through the regulatory network', *PLOS Biol.*, vol. 19, no. 12, p. e3001491, Dec. 2021, doi: 10.1371/journal.pbio.3001491.
[53] S. Bakshi, E. Leoncini, C. Baker, S. J. Cañas-Duarte, B. Okumus, and J. Paulsson, 'Tracking bacterial lineages in complex and dynamic environments with applications for growth control and persistence', *Nat. Microbiol.*, vol. 6, no. 6, Art. no. 6, Jun. 2021, doi: 10.1038/s41564-021-00900-4.
[54] Y. Taniguchi *et al.*, 'Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells', *Science*, vol. 329, no. 5991, pp. 533–538, Jul. 2010, doi: 10.1126/science.1188308.
[55] S. O. Skinner, L. A. Sepúlveda, H. Xu, and I. Golding, 'Measuring mRNA copy number in individual Escherichia coli cells using single-molecule fluorescent in situ hybridization', *Nat. Protoc.*, vol. 8, no. 6, Art. no. 6, Jun. 2013, doi: 10.1038/nprot.2013.066.
[56] I. G. de Jong, K. Beilharz, O. P. Kuipers, and J.-W. Veening, 'Live Cell Imaging of Bacillus subtilis and Streptococcus pneumoniae using Automated Time-lapse Microscopy', *JoVE J. Vis. Exp.*, no. 53, p. e3145, Jul. 2011, doi: 10.3791/3145.
[57] S. van Vliet, A. Dal Co, A. R. Winkler, S. Spriewald, B. Stecher, and M. Ackermann, 'Spatially Correlated Gene Expression in Bacterial Groups: The Role of Lineage History, Spatial Gradients, and Cell-Cell Interactions', *Cell Syst.*, vol. 6, no. 4, pp. 496-507.e6, Apr. 2018, doi: 10.1016/j.cels.2018.03.009.
[58] M. B. Elowitz and S. Leibler, 'A synthetic oscillatory network of transcriptional regulators', *Nature*, vol. 403, no. 6767, Art. no. 6767, Jan. 2000, doi: 10.1038/35002125.
[59] G. Hardo, 'Experimentally analysed data for Quantitative Microbiology with Microscopy: Effects of Projection and Diffraction'. Zenodo, May 15, 2023. doi: 10.5281/zenodo.7938932.

Copyright

Copyright © 2023 Hardo, Li, and Bakshi . This is an open-access article distributed under the terms of the Creative Commons Attribution Licence (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance

with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.