

METHODOLOGY ARTICLE

Open Access



Modified Pearson correlation coefficient for two-color imaging in spherocylindrical cells

Sonisilpa Mohapatra^{1,2*} and James C. Weisshaar¹

Abstract

The revolution in fluorescence microscopy enables sub-diffraction-limit (“superresolution”) localization of hundreds or thousands of copies of two differently labeled proteins in the same live cell. In typical experiments, fluorescence from the entire three-dimensional (3D) cell body is projected along the z -axis of the microscope to form a 2D image at the camera plane. For imaging of two different species, here denoted “red” and “green”, a significant biological question is the extent to which the red and green spatial distributions are positively correlated, anti-correlated, or uncorrelated. A commonly used statistic for assessing the degree of linear correlation between two image matrices \mathbf{R} and \mathbf{G} is the Pearson Correlation Coefficient (PCC). PCC should vary from -1 (perfect anti-correlation) to 0 (no linear correlation) to $+1$ (perfect positive correlation). However, in the special case of spherocylindrical bacterial cells such as *E. coli* or *B. subtilis*, we show that the PCC fails both qualitatively and quantitatively. PCC returns the same $+1$ value for 2D projections of distributions that are either perfectly correlated in 3D or completely uncorrelated in 3D. The PCC also systematically underestimates the degree of anti-correlation between the projections of two perfectly anti-correlated 3D distributions. The problem is that the projection of a random spatial distribution within the 3D spherocylinder is non-random in 2D, whereas PCC compares every matrix element of \mathbf{R} or \mathbf{G} with the constant mean value \bar{R} or \bar{G} . We propose a modified Pearson Correlation Coefficient (MPCC) that corrects this problem for spherocylindrical cell geometry by using the proper reference matrix for comparison with \mathbf{R} and \mathbf{G} . Correct behavior of MPCC is confirmed for a variety of numerical simulations and on experimental distributions of HU and RNA polymerase in live *E. coli* cells. The MPCC concept should be generalizable to other cell shapes.

Keywords: Pearson correlation coefficient, Two color imaging, Fluorescence microscopy, Superresolution imaging, Bacterial imaging

Background

In widefield and superresolution fluorescence microscopy of eukaryotic and prokaryotic cells, the fluorescent species occupy a three-dimensional (3D) volume. In typical usage, the laser illuminates the entire thickness of the cell (“epi illumination”). The microscope then projects fluorescence from a 3D source along the z axis to form a two-dimensional (2D) image at the xy camera plane. For two-color imaging of two different species, herein called the “red species” and the “green species”, an important biological question is the degree to which the red and green spatial distributions are positively correlated, anti-correlated, or uncorrelated with each other. Positive correlation may suggest binding to each other

or to a common cytoplasmic element such as a membrane or the chromosomal DNA. It may also suggest common sites of production, action, or degradation. Negative correlation may suggest a physical or biochemical mechanism that sequesters red and green species from each other [1, 2]. A number of different procedures for assessing co-localization between two images are described in a recent review [3].

For super-resolution images, a family of point pattern analysis methods evaluates the spatial co-distribution of points on very short (sub-100 nm) length scales. These include Ripley’s K test [4–6] and a variety of cross-correlation methods [7–10]. These procedures provide a function of r (the inter-particle separation distance) that describes the spatial distribution of red and green molecules with respect to each other. Such methods take advantage of the sub-pixel accuracy and allow determination of whether the red and green proteins are dispersed, clustered, or

* Correspondence: smohapa2@jhmi.edu

¹Department of Chemistry, University of Wisconsin-Madison, Madison, WI 53706, USA

²Present Address: Department of Biophysics and Biophysical Chemistry, Johns Hopkins School of Medicine, Baltimore 21205, USA



randomly distributed within the region of interest. The data density must be commensurate with the length scale of interest, i.e., high data density is required to obtain information on the sub-100 nm scale.

For some time now, we have been interested in the degree to which ribosomes and the chromosomal DNA are spatially segregated from each other on a length scale of ~ 200 nm and longer in *E. coli* bacterial cells growing exponentially under different conditions [11, 12]. The cells are spherocylindrical, typically of length 3–5 μm and diameter ~ 1 μm or smaller. In rapidly growing cells, the chromosomal DNA has segregated into two nucleoid lobes that interleave three ribosome-rich regions [11], each of whose size is of the order of 0.5–1.0 μm. For this problem, sub-pixel resolution is not needed. In small bacterial cells, the coordinate based cross-correlation methods provide readily interpretable information only for r substantially smaller than the shortest cell dimension. Accordingly, we have chosen to use superresolution imaging to minimize the blurring inherent in widefield microscopy. We subsequently pixelate the red and green images and calculate a modification of the Pearson correlation coefficient (PCC) that returns a single number in the range + 1.0 to - 1.0 that measures the degree of linear correlation or anti-correlation between red and green images, averaged over the entire cell.

As described in detail below, all correlation quantification methods have limitations in the common case of 2D images projected from the 3D spatial distributions of fluorophores emitting from small bacterial cells. A reference distribution that is random in 3D within the cell boundaries produces a non-uniform 2D spatial distribution when projected onto the camera plane. Moerner and co-workers have recently applied Ripley’s K to characterize the clustering of HU proteins in the crescent-shaped bacteria *C. crescentus* and corrected the reference random distribution by methods similar to those we employ here [13]. Here we describe a detailed procedure for handling the same problem in estimates of the Pearson correlation coefficient in the case of spherocylindrical cells like *E. coli* and *B. subtilis*.

The Pearson correlation coefficient (PCC) [14, 15] is one of the most commonly used statistical tools to measure the degree of linear correlation in pixel-by-pixel intensity between two data sets X and Y:

$$PCC = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

Here (x_i, y_i) are individual paired samples from the data sets X and Y and n is the total number of pairs; \bar{x} and \bar{y} are the mean values of the samples in data sets X and Y. With the advent of two-color superresolution fluorescence microscopy, the PCC is increasingly used as

a statistic for quantifying the degree of correlation between the subcellular distributions of two distinguishable species. For image matrices **R** (red channel) and **G** (green channel), the formula for PCC becomes:

$$PCC = \frac{\sum_{i=1}^m \sum_{j=1}^n (R_{ij} - \bar{R})(G_{ij} - \bar{G})}{\sqrt{\sum_{i=1}^m \sum_{j=1}^n (R_{ij} - \bar{R})^2} \sqrt{\sum_{i=1}^m \sum_{j=1}^n (G_{ij} - \bar{G})^2}} \quad (2)$$

Here m and n are the number of rows and columns in the image matrices; there are $m \times n$ total pixels in each image. The R_{ij} and G_{ij} are the corresponding intensities of pixel ij in **R** and **G**; for superresolution images these are integers (counts/pixel). \bar{R} and \bar{G} are the mean pixel intensities of **R** and **G**. In the PCC formula, all elements of the reference matrix with which **R** or **G** is compared have the same value. The value \bar{R} (or \bar{G}) is subtracted from each individual pixel intensity R_{ij} (or G_{ij}), yielding both positive and negative difference intensities $(R_{ij} - \bar{R})$ and $(G_{ij} - \bar{G})$. Thus, the product in the PCC numerator provides information about the correlation between deviations of R_{ij} from \bar{R} and deviations of G_{ij} from \bar{G} . The denominator normalizes PCC so that it always lies in the range - 1 to + 1. Ideally, PCC = 1 indicates two perfectly linearly correlated images for which each red pixel ij deviates from the red mean in direct proportion to the deviation of the corresponding green pixel ij from the green mean. PCC = 0 indicates two linearly uncorrelated images. PCC = - 1 indicates two perfectly anti-correlated images (red and green deviations of equal magnitude but of opposite sign). A PCC value significantly different from zero is a measure of the degree to which two distributions are correlated or anti-correlated as compared with the null hypothesis of PCC = 0, corresponding to two uncorrelated, random distributions.

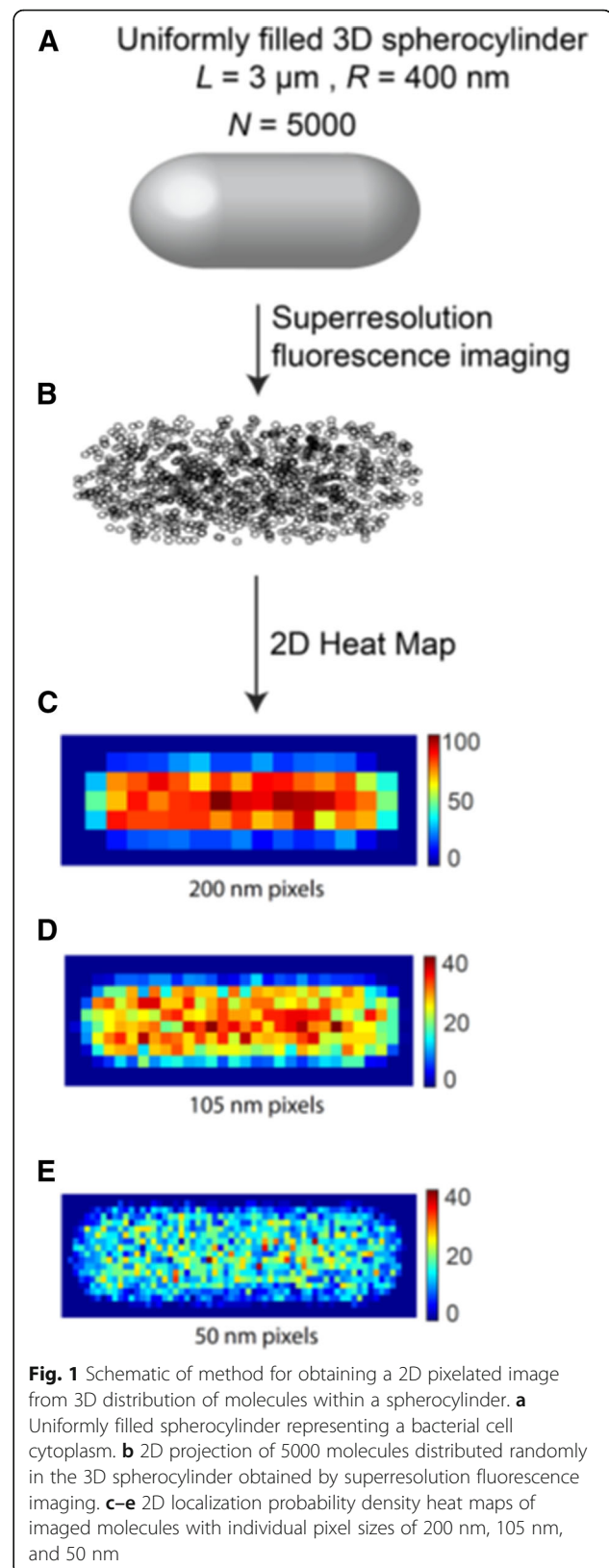
The ImageJ software [16] extensively used for image analysis in the field of fluorescence microscopy provides Coloc2 and JaCoP plugins [17] that enable the user to calculate PCC between two images. In the recent literature, PCC has been used to characterize the correlation in 2D spatial distributions of two fluorescently labeled proteins in both bacterial cells [18–20] and eukaryotic cells [21–27]. McDonald and co-workers recently catalogued some common pitfalls in the use of PCC on eukaryotic cells [23].

For the most common shapes of bacteria (spherical, rod-shaped and spiral), the standard PCC procedure applied to 2D projected images fails both qualitatively and quantitatively. We specialize to small, rod-shaped, approximately spherocylindrical bacterial cells such as *E. coli* and *B. subtilis*, whose typical length is $L_{cell} \sim 4 \mu\text{m}$ and whose diameter is $2r \sim 1 \mu\text{m}$. Spherocylinders have strong curvature at the two endcaps and in the cylindrical region. As a result, the projection of

molecules randomly distributed in a 3D spherocylindrical volume does not form a random distribution in 2D. In Fig. 1, we illustrate the 2D projection of 5000 molecules that are distributed randomly in a 3D spherocylinder with dimensions similar to that of an *E. coli* cell in good growth conditions. The endcap regions and the edges of the spherocylinder project a smaller volume onto the camera plane, and thus have fewer counts/pixel in the 2D image than the central cylindrical region. This effect is clear in the pixelated 2D localization density maps shown in Fig. 1c-e. Pixels in the 2D projection of a random 3D distribution vary in intensity by a factor of five or more, depending on the chosen pixel size. The variations are highly systematic.

Consequently, the PCC reference matrix used for comparison with \mathbf{R} and \mathbf{G} is inappropriate. The PCC difference intensities $(R_{ij}-\bar{R})$ and $(G_{ij}-\bar{G})$ for pixels at the edges and end caps are systematically negative, i.e., strongly biased towards having fewer molecules/pixel than the mean value in a 2D projection of a 3D random distribution. In those regions, the products $(R_{ij}-\bar{R})(G_{ij}-\bar{G})$ are systematically positive. Similarly, the difference intensities of the pixels in the central region of the spherocylinder are systematically positive, strongly biased towards having more molecules/pixel than the mean of a projection of a 3D random distribution. In that region, the products $(R_{ij}-\bar{R})(G_{ij}-\bar{G})$ are again systematically positive. For two uncorrelated, random distributions in 3D, this causes the traditional PCC of the 2D projection to incorrectly approach +1, not the desired result of zero. The same systematic positive bias causes the traditional PCC to underestimate the degree of anti-correlation between two perfectly anti-correlated images, as we will show.

In the following sections, we describe a procedure for calculating what we call the modified Pearson correlation coefficient (MPCC) in the special case of interest, spherocylindrical bacterial cells. The procedure could prove useful for both widefield and superresolution images, and in principle it could be adapted to other cell shapes [3]. We use numerical simulations to show that MPCC properly approaches zero for random sampling from two uncorrelated, random distributions, approaches -1 for sampling from two perfectly anti-correlated distributions, and approaches $+1$ for sampling from two perfectly correlated distributions. We also provide guidance for pixelation of superresolution images and show how to determine the probability p that a measured non-zero MPCC did not arise from two uncorrelated, random 3D distributions. We conclude with an experimental example of a significantly positive MPCC between



superresolution images of RNA polymerase and of the DNA-binding protein HU in live *E. coli*. The package of MATLAB codes required for calculating MPCC between two different molecules imaged in rod shaped cells such as *E. coli* and *B. subtilis* is available on GitHub: <https://github.com/SoniMohapatra/MPCC>.

Results

The modified Pearson correlation coefficient MPCC

The MPCC of two images \mathbf{R} and \mathbf{G} is evaluated as follows:

$$\text{MPCC} = \frac{\sum_{i=1}^m \sum_{j=1}^n (R_{ij} - \bar{U}_{ij}^R) (G_{ij} - \bar{U}_{ij}^G)}{\sqrt{\sum_{i=1}^m \sum_{j=1}^n (R_{ij} - \bar{U}_{ij}^R)^2} \sqrt{\sum_{i=1}^m \sum_{j=1}^n (G_{ij} - \bar{U}_{ij}^G)^2}}. \quad (3)$$

Here we have replaced \bar{R} and \bar{G} in Eq. 2 with the modified reference matrices \bar{U}_{ij}^R and \bar{U}_{ij}^G , respectively. \bar{U}_{ij}^R and \bar{U}_{ij}^G denote the intensity of pixel ij in the 2D projection of a large set of molecules distributed randomly in a 3D spherocylinder. The total number of molecules in \bar{U}^R and \bar{U}^G has been scaled to be the same as the total number of molecules in \mathbf{R} and \mathbf{G} , respectively.

In favorable conditions, superresolution imaging provides (x,y) spatial localization of hundreds or thousands of molecules per cell with spatial resolution of $\sigma_{x,y} \sim 20$ – 50 nm. Conversion of these single molecule locations into 2D probability density maps requires selection of a pixel size; several examples are shown in Fig. 1c–e. The intensity in each pixel equals the total number of molecules assigned to it. The dependence of the calculated MPCC on the chosen pixel size and the number of imaged molecules is described later. These pixelated 2D maps for the red and green channels are denoted by \mathbf{R} and \mathbf{G} , the image matrices in Eq. 3.

To form the numerator of Eq. 3, we then subtract \bar{U}^R and \bar{U}^G from the corresponding image matrix in the red and green channels (\mathbf{R} and \mathbf{G} , respectively) to obtain the (unnormalized) difference matrices Δ^R and Δ^G . The resultant difference matrices have pixels with positive and negative values. Finally, to constrain MPCC to lie in the range $+1$ to -1 , we normalize Δ^R and Δ^G so that the sum of the squares of individual pixel values in the difference matrix is 1. The resultant normalized 2D difference matrices are called $\hat{\Delta}^R$ and $\hat{\Delta}^G$ respectively. MPCC is obtained by taking the Frobenius inner product of the two normalized matrices $\hat{\Delta}^R$ and $\hat{\Delta}^G$ (Eq. 6 in Methods). A detailed step-by-step description of the methodology for obtaining MPCC is presented in the Methods section.

The MPCC ranges from $+1$ to -1 , as does standard PCC. The MPCC for two images is $+1$ when the normalized difference matrices are perfectly linearly related, i.e., when $\hat{\Delta}_{ij}^R = \hat{\Delta}_{ij}^G$ for every pixel ij . As a result, $\text{MPCC} = \sum_{i=1}^m \sum_{j=1}^n \hat{\Delta}_{ij}^R \hat{\Delta}_{ij}^G = \sum_{i=1}^m \sum_{j=1}^n \hat{\Delta}_{ij}^{R^2} = +1$. The MPCC is -1 when the normalized difference matrices are perfectly inversely related to each other, i.e., $\hat{\Delta}_{ij}^R = -\hat{\Delta}_{ij}^G$ for every pixel. As a result, $\text{MPCC} = \sum_{i=1}^m \sum_{j=1}^n \hat{\Delta}_{ij}^R \hat{\Delta}_{ij}^G = -\sum_{i=1}^m \sum_{j=1}^n \hat{\Delta}_{ij}^{R^2} = -1$. When the normalized difference matrices of two images are uncorrelated with each other, the MPCC is 0.

Next, we carry out numerical simulations comparing MPCC with PCC for sampling from 2D projections of three model distributions in 3D spherocylinders: perfect 3D correlation that projects into perfect 2D correlation, perfect 3D anti-correlation that projects into perfect 2D anti-correlation, and uncorrelated, random 3D distributions. For all these examples, the \mathbf{R} and \mathbf{G} image matrices have 10,000 molecules each. The spherocylinder has tip-to-tip length $L_{cell} = 3.5$ μm and diameter $2r = 0.82$ μm . The 2D pixel size in the image matrices \mathbf{R} and \mathbf{G} is chosen to be 200 nm in both dimensions, so that 75 pixels cover the 2D projection.

Perfect anti-correlation in 3D

To examine the case of two perfectly anti-correlated distributions, we have simulated 3D random distributions of 20,000 molecules confined to the spherocylindrical volume. The $\sim 10,000$ molecules located in the left half of the spherocylinder are designated red; the $\sim 10,000$ molecules located in the right half are designated green. This ensures that there is no spatial overlap of molecules in the red and green channels. We call this anti-correlation Case I. For such strong spatial anti-correlation, we should expect $\text{MPCC} = -1$. An example of the corresponding 2D image matrices \mathbf{R} and \mathbf{G} is shown in Fig. 2a. In Fig. 2b, c, we have compared the reference matrices and the key normalized difference matrices the products of whose corresponding elements enter the traditional PCC (Eq. 2) and the new MPCC (Eq. 3).

For the traditional PCC (Fig. 2b), there are $\sim 10,000$ molecules of each color distributed in a cell area covering 75 pixels. As in Eq. 2, we subtract the mean pixel intensity $\bar{R} = 133.3$ and $\bar{G} = 133.3$ from each individual pixel intensities R_{ij} and G_{ij} . The resulting normalized difference matrices, $\frac{R_{ij} - \bar{R}}{\sqrt{\sum_{i=1}^m \sum_{j=1}^n (R_{ij} - \bar{R})^2}}$ and $\frac{G_{ij} - \bar{G}}{\sqrt{\sum_{i=1}^m \sum_{j=1}^n (G_{ij} - \bar{G})^2}}$, are depicted as heat maps labeled $\sim(\mathbf{R} - \bar{R})$ and $\sim(\mathbf{G} - \bar{G})$ in Fig. 2b. These are the PCC analogues of $\hat{\Delta}_{ij}^R$ and $\hat{\Delta}_{ij}^G$ in the

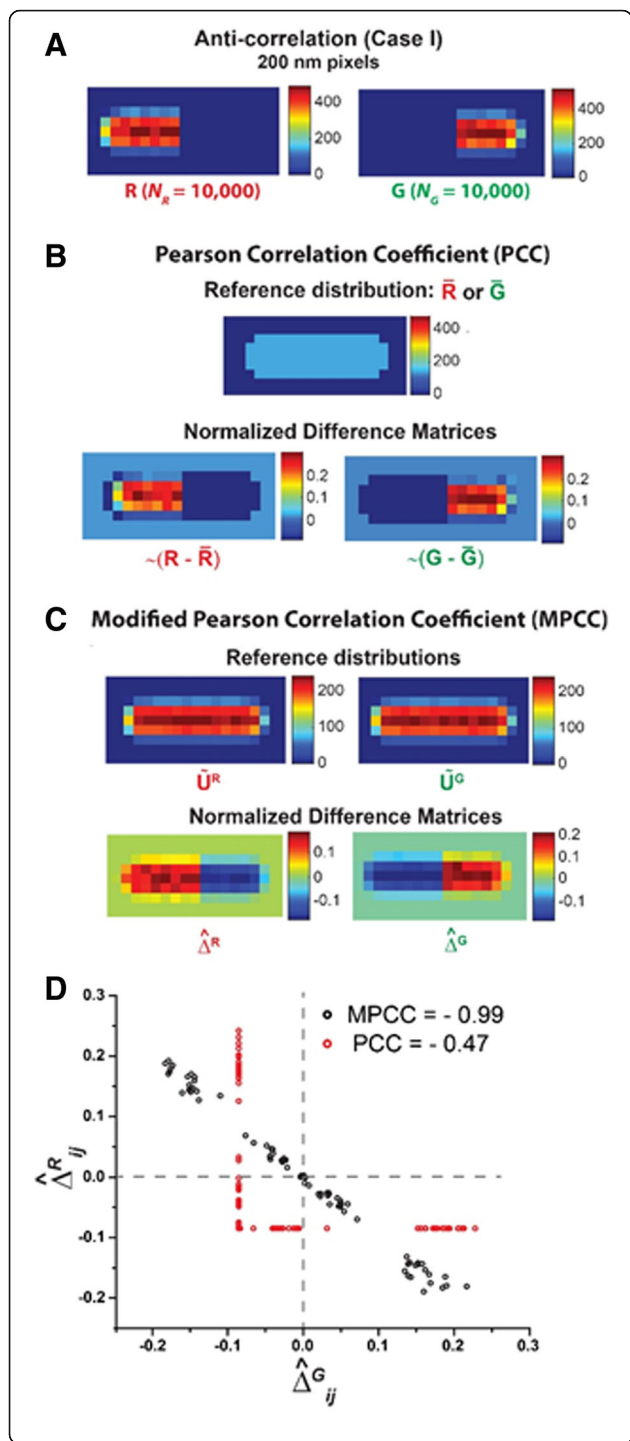


Fig. 2 Scheme for calculating PCC and MPCC for two representative images R and G sampled from distributions that are perfectly anti-correlated in both 3D and 2D. **a** Heat maps of R and G with 200 nm pixels. Each image comprises $\sim 10,000$ molecules. Color scale indicates the number of molecules in each pixel. **b** Standard PCC calculation. *Top*: The 2D uniform reference distribution \bar{R} or \bar{G} that is subtracted from images R or G . *Bottom*: Normalized difference matrices $\sim(R - \bar{R})$ and $\sim(G - \bar{G})$ obtained after subtraction. The Frobenius inner product of these two difference matrices gives the PCC. **c** Modified PCC calculation. *Top*: Reference distribution \bar{U}^R and \bar{U}^G , which are 2D projections of 3D random distributions of 100,000 molecules within the spherocylinder and normalized to have a total of 10,000 molecules. These are subtracted from images R and G , respectively. *Bottom*: Normalized difference matrices $\hat{\Delta}^R$ and $\hat{\Delta}^G$ obtained after subtraction. The Frobenius inner product of these two difference matrices gives the MPCC. **d** Scatter plot of individual normalized difference matrix elements for PCC (Red) and for MPCC (Black). The MPCC elements are negatively correlated within the noise level, while the PCC elements are not. The resulting MPCC and PCC values are -0.99 and -0.47 , respectively

MPCC equation. In the left half of the spherocylinder, the red difference matrix has a thin shell of systematically negative values (endcap and edge pixels) and a central core of systematically positive values. When multiplied by the corresponding elements of the left half of the green difference matrix, which contains all equal negative elements, the contributions to PCC will be positive and negative, respectively. The same type of systematically positive and negative contributions will arise from the right half of the spherocylinder. The resulting red and green contributions to PCC are not linearly anti-correlated. This is seen clearly in Fig. 2d, where we show a scatter plot of the individual red normalized differences vs the corresponding green normalized differences. The net result is $PCC = -0.47$, suggesting only partial anti-correlation of the two spatial distributions even though they are completely anti-correlated in both 3D and 2D.

In contrast, the MPCC formula of Eq. 3 subtracts from each pixel the proper 2D contribution of the projection of a smooth 3D random distribution (Fig. 2c). The resulting normalized difference matrices $\hat{\Delta}^R$ and $\hat{\Delta}^G$ are also depicted in Fig. 2c. The scatter plot of individual difference matrix elements $\hat{\Delta}_{ij}^R$ vs. $\hat{\Delta}_{ij}^G$ in Fig. 2d shows the expected strong linear anti-correlation for all pixels. The resulting MPCC is -0.99 , very close to the expected value of -1 .

In Additional file 1: SI Text S1, we examine two additional examples of anti-correlation. In anti-correlation Case II shown in Additional file 1: Figure S1, the two endcap regions are occupied by $\sim 10,000$ red molecules and the central region is occupied by $\sim 10,000$ green molecules. Again, the normalized difference matrix elements are linearly anti-correlated and the calculated MPCC is -0.99 . In anti-correlation Case III

(Additional file 1: Figure S1), the $\sim 10,000$ red molecules occupy the leftmost 2/3 of the spherocylinder volume while the $\sim 10,000$ green molecules occupy the rightmost 1/3. The result is the same. The advantages of MPCC vs traditional PCC are apparent.

Perfect positive correlation in 3D and 2D

When the red and green 3D spatial distributions are perfectly positively correlated, so will be their 2D projections. As described before, an MPCC value of +1 is expected for a case of perfect correlation in the 2D projections. The same is true of the traditional PCC. To examine the case of two perfectly correlated distributions, we have simulated 3D random distributions of 20,000 molecules confined to the spherocylindrical volume. The $\sim 10,000$ molecules located in the left half of the spherocylinder are designated red; the molecules in the right half are deleted. We then independently simulated another 20,000 molecules distributed randomly in a 3D spherocylinder. The $\sim 10,000$ molecules located in the left half of the spherocylinder are designated green; the molecules in the right half are again deleted. The resulting 3D distributions are projected into 2D and pixelated to yield the image matrices depicted in Additional file 1: Figure S2A. We calculate the MPCC = +0.99 between these two distributions, very close to the anticipated value of +1. The resulting normalized difference matrices $\hat{\Delta}^R$ and $\hat{\Delta}^G$ obtained during evaluation of MPCC are depicted in Additional file 1: Figure S2C. The scatter plot of individual matrix elements $\hat{\Delta}_{ij}^R$ vs. $\hat{\Delta}_{ij}^G$ in Additional file 1: Figure S2D shows the expected strong linear correlation for all pixels. Similarly, the scatter plot of individual normalized difference matrix elements analogous to $\hat{\Delta}_{ij}^R$ vs. $\hat{\Delta}_{ij}^G$ for PCC in Additional file 1: Figure S2D shows the expected strong linear correlation for all pixels. If $R_{ij} = G_{ij}$ and $\bar{R} = \bar{G}$, then PCC = 1. Therefore, for two spatial distributions that are perfectly correlated in 3D and in the 2D projection, both the MPCC and the PCC will approach +1 within the statistical noise.

Random distributions in 3D

Two independent, uncorrelated, random distributions should have a Pearson correlation coefficient of 0 within the statistical noise. In the numerical tests, we have randomly distributed 10,000 red molecules and 10,000 green molecules in 3D within the spherocylinder. The two random distributions are generated independently, so we expect them to be uncorrelated with each other. We add appropriate localization errors $\sigma_R = 50$ nm and $\sigma_G = 50$ nm and then project the “measured” positions into the xy -plane. PCC and MPCC between the two 2D projection matrices (Fig. 3a) will be compared.

The resulting reference matrices and normalized difference matrices for PCC and for MPCC are depicted in Fig. 3b and c respectively. The scatter plots of $\hat{\Delta}_{ij}^R$ vs $\hat{\Delta}_{ij}^G$ for MPCC and of their analogues for PCC are shown in Fig. 3d. The data indeed appear uncorrelated for MPCC, but they are strongly positively correlated for PCC. The resulting calculated coefficients are MPCC = +0.10 and PCC = +0.98. The cause of the large, positive PCC value between two random 3D distributions was described in the Introduction. The 2D projections have matching regions of systematically positive and systematically negative deviations from the 2D mean values.

Finally, we tested whether the distribution of calculated MPCC outcomes for two independent random distributions is appropriately centered at zero and unbiased towards positive or negative values. For 200 trials, we calculated MPCC values between two 2D projections of 3D independent, random distributions of 10,000 red and 10,000 green molecules using the same 200 nm pixel size. We fit the resulting distribution (Additional file 1: Figure S3) to a Gaussian function. The mean of the best-fit Gaussian distribution is $\langle \text{MPCC} \rangle = +0.0041$ and the standard error is $\sigma_{\text{MPCC}} = 0.13$. The mean is close to zero and the distribution is symmetric about zero, as hoped for. The probability that a particular trial would yield an MPCC of magnitude 0.10 or larger on either side of the Gaussian distribution is $p = 0.44$. The “measured” example MPCC of +0.10 (Fig. 3d) lies within 1σ of the mean; it was not a particularly unusual event.

Dependence of MPCC and its uncertainty on pixel size and total number of imaged molecules

Before evaluating MPCC between two superresolution images, the pixel size in the 2D localization density maps must be chosen. For a fixed cell size and number of detected molecules, the smaller the pixel size, the greater will be the total number of pixels N_p , the better the spatial resolution, and the smaller the mean occupancy per pixel. We have shown in SI (Additional file 1: Figure S4) that for a fixed number of localizations $N_R = N_G = 10,000$ distributed randomly in 3D, as the pixel size decreases (and N_p increases) the width of the distribution of MPCC values becomes narrower. All the MPCC distributions for uncorrelated images are symmetric and centered about 0 and well fit by a Gaussian function. For these random, uncorrelated 3D distributions, the standard deviation of the Gaussian MPCC distributions scales as $N_p^{-1/2}$. This scaling holds even for N_R and N_G as small as 500.

Narrower widths of the MPCC distribution from random 3D distributions generally provide greater statistical confidence that a non-zero measured value of MPCC is significantly different from zero. This argues for fine

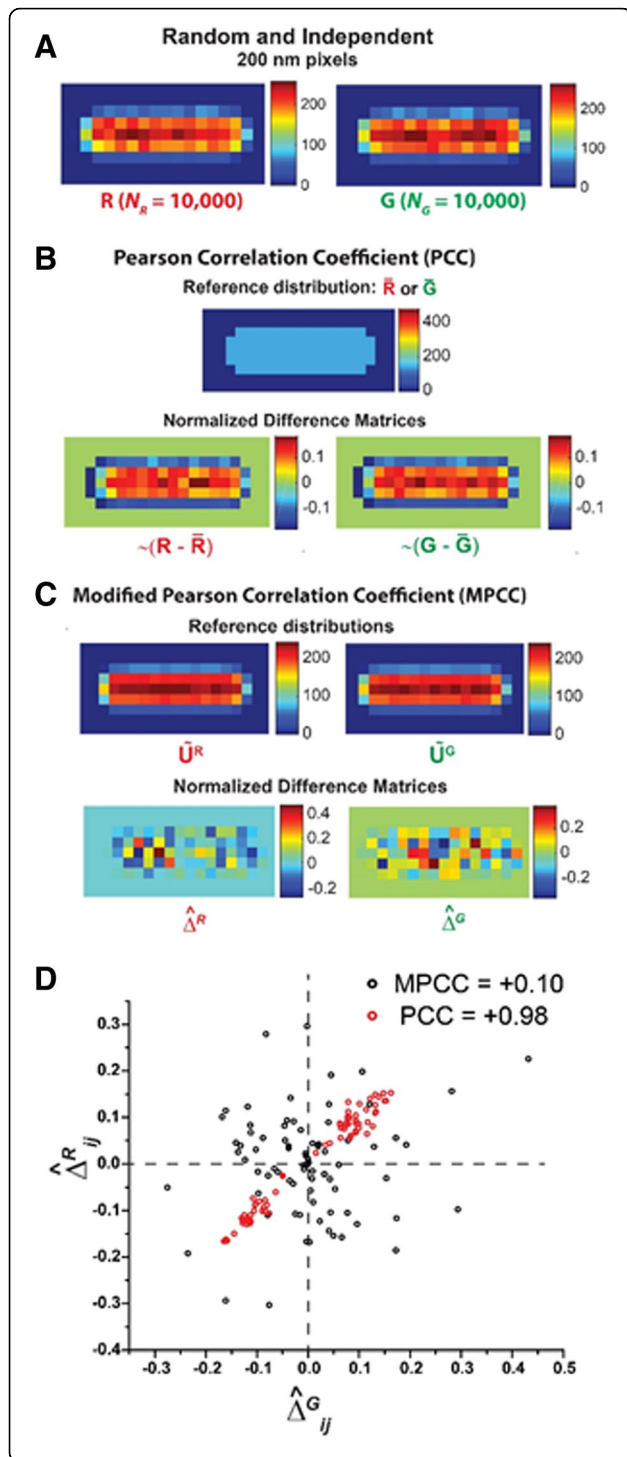


Fig. 3 Scheme for calculating PCC and MPCC for two representative projected images **R** and **G** arising from two random and independent distributions in 3D. **a** Heat maps of **R** and **G** with 200 nm pixels. Each image comprises ~ 10,000 molecules. Color scale indicates the number of molecules in each pixel. **b** Standard PCC calculation. *Top*: The 2D uniform reference distribution \bar{R} or \bar{G} that is subtracted from images **R** or **G**. *Bottom*: Normalized difference matrices $\sim(R - \bar{R})$ and $\sim(G - \bar{G})$ obtained after subtraction. **c** Modified PCC calculation. *Top*: Reference distribution \bar{U}^R and \bar{U}^G , which are 2D projections of 3D random distributions of 100,000 molecules within the spherocylinder and normalized to have a total of 10,000 molecules. These are subtracted from images **R** and **G**, respectively. *Bottom*: Normalized difference matrices $\hat{\Delta}^R$ and $\hat{\Delta}^G$ obtained after subtraction. **d** Scatter plot of individual normalized difference matrix elements for PCC (Red) and for MPCC (Black). The MPCC elements are randomly distributed, while the PCC elements are positively correlated. The resulting MPCC and PCC values are + 0.10 and + 0.98, respectively

pixelation. In practice, we suggest simulating the distribution of MPCC values between the 2D projections of 3D random distributions using the same number of molecules as were imaged in the red and green channels and the same pixel size chosen for **R** and **G**. This enables assignment of a probability p that the measured MPCC arose from two random 3D distributions. If p is unacceptably large, finer pixelation of both experimental and simulated locations may decrease p . Finer pixelation also enables detection of correlation or anti-correlation on smaller length scales.

However, for non-random 3D distributions such as the completely anti-correlated distribution of Fig. 2 or the positively correlated distribution of Additional file 1: Figure S2, it is important not to pixelate so finely that the matrices **R** and **G** become too sparse. In the case of the anti-correlated model matrices **R** and **G**, this leads to false positive linear correlations between $\hat{\Delta}_{ij}^R$ and $\hat{\Delta}_{ij}^G$. One way to think about this is that the zeroes and small-integer occupancies appearing in the left-hand region of **R** begin to positively correlate with the zeroes that fill the empty half of **G**. Similarly, the zeroes and small-integer occupancies arising due to sparseness in the right-hand region of **G** positively correlate with the zeroes in the empty half of **R**. These systematically bias the MPCC for truly anti-correlated distributions towards more positive values, underestimating the degree of linear anti-correlation. We explore this effect numerically in Additional file 1: Figure S5. For a given pixel size, the mean MPCC moves closer to the expected value of -1 for two anti-correlated images as the number of imaged molecules increases. The key controlling parameter seems to be the mean occupancy per pixel.

In practice, we suggest carrying out numerical simulations of perfectly anti-correlated distributions using values of N_R and N_G that match experiment. The pixel size

chosen for analysis of the experimental data should be the smallest pixel size for which the mean MPCC for perfectly anti-correlated distributions is acceptably close to -1 . In the numerical example of Fig. 2, with 10,000 molecules distributed over 75 pixels, the mean occupancy was 133 molecules/pixel, which yielded $\text{MPCC} = -0.99$. For these images sampled from perfectly anti-correlated model distributions, if the mean occupancy is ~ 7 copies/pixel (~ 14 copies per pixel in the occupied halves of the case in Fig. 2), then the MPCC will be about -0.9 . MPCC approaches -1 as the occupancy per pixel increases.

For similar reasons, for two perfectly positively correlated distributions we expect that MPCC will systematically underestimate the degree of positive correlation as the red and green matrices become sparse. In the case of positively correlated **R** and **G** (Additional file 1: Figure S2), the zeroes appearing in the images due to sparseness are not positively correlated. The sparseness in number of molecules due to finer pixelation leads to false negative linear correlations between $\hat{\Delta}_{ij}^R$ and $\hat{\Delta}_{ij}^G$. This leads to systematic negative deviations of the calculated MPCC from the expected value of $+1$. We investigated the mean occupancy/pixel that is required for the calculated MPCC between strongly positively correlated images to be ~ 0.9 , close to the expected value of $+1$. As shown in Fig. 4e and S5, a mean occupancy of ~ 7 copies/pixel (14 copies/pixel in the occupied regions) yields MPCC values of about $+0.9$.

While this rule of thumb seems to hold for the perfectly anti-correlated and perfectly correlated model distributions, the pixel occupancy requirement may be more stringent for less strongly anti-correlated or correlated cases. See the experimental example below. In the next section we analyze experimental RNAP and HU distributions and suggest a procedure for assessing the reliability of MPCC values more generally.

Experimental example of MPCC from superresolution images of RNAP and HU in *E. coli*

To test our MPCC concept on real experimental data, we performed two-color superresolution fluorescence imaging of RNA polymerase and HU in live *E. coli* cells. RNAP is primarily located in the nucleoid region because of its frequent specific and non-specific interactions with chromosomal DNA [28]. HU is a DNA binding protein that should also localize within the nucleoids [29, 30]. We expect significant positive correlation between the spatial distributions of RNAP and HU and therefore a positive value of MPCC.

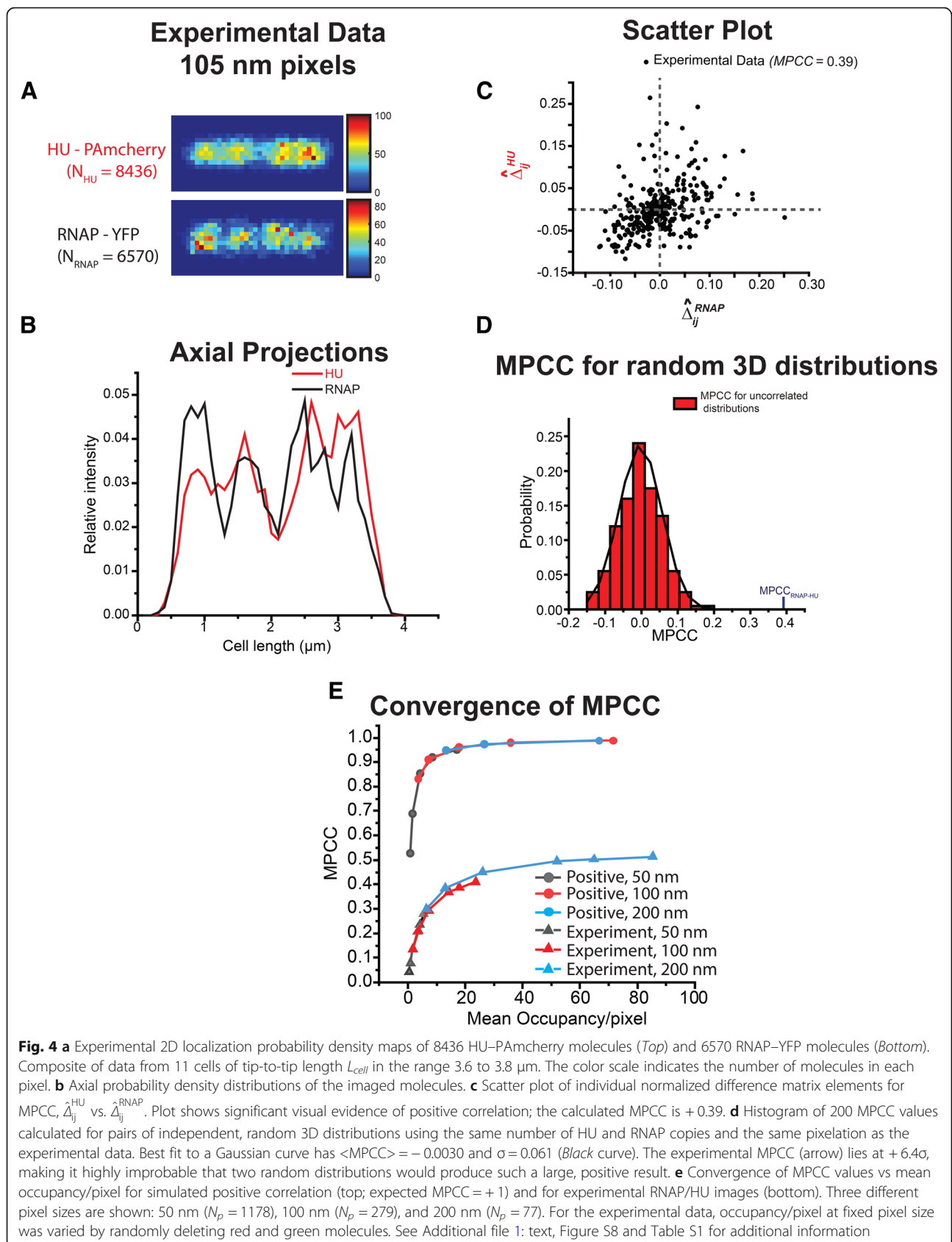
For superresolution co-imaging of RNAP and HU in live *E. coli* cells, we constructed a strain where the gene coding for the fluorescent protein YFP (observed in the green channel) [31] is fused to the C terminus of the

endogenous *rpoC* gene in *E. coli* VH1000. Single copies are imaged using the reversible photobleaching method described earlier [32]. An inducible plasmid that expresses HU labeled with the photoactivatable fluorescent protein PAmcherry [33] (observed in the red channel) was introduced into the same strain. The cells were grown in EZ rich defined medium at 30°C , plated on a glass coverslip, and imaged with 30 ms exposure time. The details of strain construction, growth conditions, and imaging conditions are described in Additional file 1: SI Text S3.

To obtain a useful number of imaged copies without inducing laser damage to the cells, we combine locations of red HU and green RNAP molecules from different cells of essentially the same length. The imaged cells were sorted by tip-to-tip length based on phase contrast images in order to avoid broadening of spatial distribution of molecules due to the range of cell lengths. For the analysis, we chose cells of length 3.6 to 3.8 μm , the bin with the highest number of imaged cells. The resulting composite distribution of spatial localizations of $N_G = 6570$ RNAP-YFP and $N_R = 8436$ HU-PAmcherry molecules from 11 cells pixelated to 105 nm (279 total pixels) is illustrated in Fig. 4a. The mean number of molecules per pixel is ~ 25 and ~ 30 for the RNAP and HU channels respectively. The corresponding 1D projected axial distributions are compared in Fig. 4b. The raw data indeed suggest significant positive correlation between the two distributions.

For evaluation of MPCC we simulated two random distributions of 100,000 molecules each, corresponding to the RNAP (green) and HU (red) channels, using a spherocylinder whose dimensions match those of the chosen cells. The resulting reference images are normalized to have same number of molecules as imaged RNAP and HU. For accurate estimation of the cytoplasmic radius r of the imaged cells in the chosen length bin, we also imaged photoactivatable Kaede molecules [34, 35], believed to distribute homogeneously in the cytoplasmic volume [36]. The detailed procedure is described in Additional file 1: SI Text S4. The resulting cell length is $L_{\text{cell}} = 3.74 \mu\text{m}$; the diameter is $2r = 0.82 \mu\text{m}$ (Additional file 1: Figure S6). The two simulated 3D random distributions incorporated localization errors $\sigma_{\text{RNAP}} = 38 \text{ nm}$ and $\sigma_{\text{HU}} = 60 \text{ nm}$, determined by the intercepts of MSD plots (Additional file 1: Figure S7). We followed the procedure described above with pixel size of 105 nm to calculate $\text{MPCC} = +0.39$. The scatter plot of $\hat{\Delta}_{ij}^R$ vs $\hat{\Delta}_{ij}^G$ (Fig. 4c) also indicates significant positive correlation.

The final step estimates the probability p that a value of $\text{MPCC} = +0.39$ or larger would be obtained from two random 3D distributions with the same number of imaged molecules and the same pixel size used for the experimental data. In Fig. 4d, we show a histogram of the outcomes of 200 such simulations.



The best-fit Gaussian distribution has a mean value $\langle \text{MPCC} \rangle = -0.0030$ and standard error $\sigma_{\text{MPCC}} = 0.061$. The measured MPCC value lies $6.4\sigma_{\text{MPCC}}$ away from zero. Under the assumption that the statistics of the simulated MPCC trials are Gaussian, the probability that two random 3D distributions would produce an MPCC value of magnitude 0.39 or larger on either side of the Gaussian curve is $p \sim 1.6 \times 10^{-10}$. Thus, we can reject the null hypothesis that $\text{MPCC} = +0.39$ arose from two random, uncorrelated 3D distributions and assert significant positive correlation between the RNAP and HU distributions with very high confidence.

The choice of pixel size does affect the calculated MPCC. For 200 nm pixels ($N_p = 77$ total pixels), the experimental MPCC is +0.51. The corresponding simulations of two random distributions gave $\langle \text{MPCC} \rangle = 0.0082$ and $\sigma_{\text{MPCC}} = 0.12$. In this case, the probability that two 3D random distributions would produce an MPCC value of magnitude 0.51 or higher on either side of the mean of the Gaussian curve is $p \sim 1.3 \times 10^{-4}$. For 50 nm pixels ($N_p = 1178$ total pixels), the experimental MPCC is +0.25. The corresponding simulations of two random distributions gave $\langle \text{MPCC} \rangle = 0.0027$ and $\sigma_{\text{MPCC}} = 0.033$. In this case, the probability that two 3D random distributions would produce an MPCC value of magnitude 0.25 or higher on either side of mean of Gaussian curve is $p \sim 3.6 \times 10^{-14}$. The estimated experimental MPCC decreases systematically as N_p increases and the same data set is pixelated more finely, but the simulated σ_{MPCC} decreases more rapidly.

The conclusion of significant positive correlation between the RNAP and HU experimental distributions is robust, but what is the best value of MPCC to report? In Fig. 4e and Additional file 1: Figure S8, we explore how the calculated value of MPCC varies with the mean occupancy per pixel. Given a limited number of experimental localizations, there are two ways to vary this parameter: we can keep all the experimental localizations and change the pixel size (50 nm, 105 nm, 200 nm), or we can fix the pixel size and randomly delete red and green copies from each image. MPCC values generated by both procedures fall on the same smooth curve in plots of calculated MPCC vs occupancy per pixel (Fig. 4e, Additional file 1: Figure S8 and Table S1). For the experimental images, the MPCC values are approaching an asymptote of ~ 0.5 as the mean occupancy/pixel approaches 100. Our best estimate is thus $\text{MPCC} = 0.50 \pm 0.05$. Because the features of interest in the images are large, 500 nm to 1 μm in size, we feel justified in including pixel sizes in the range 50–200 nm in the analysis.

As suggested by the projected axial distribution of RNAP and HU (Fig. 4b), the two species are not completely correlated in space. There are several factors that may explain why the MPCC is significantly smaller than 1. We have averaged the data over 11 cells whose nucleoids have irregular shapes in 3D that are not axially

symmetric and that vary from cell to cell. In addition, while RNAP and HU both bind to the DNA, they have different biological functions and should not be expected to have spatial distributions that correlate perfectly.

As a cautionary note, we observe that for the perfectly correlated or anti-correlated model distributions, MPCC converges towards its asymptotic value vs occupancy/pixel substantially more rapidly than the experimental images (Fig. 4e). In the model images, MPCC reached 90% of its asymptote of ± 1 when the occupied side of the image had 14 copies per pixel (7 copies/pixel averaged over the entire cell, which is half empty for both colors). For the experimental data, MPCC reaches 90% of the apparent asymptote of 0.5 only when the occupancy/pixel approaches 30. While mean occupancy/pixel appears to be the controlling parameter, the magnitude required to achieve 10% accuracy evidently depends on the image shape.

Discussion

The Pearson correlation coefficient is one of the statistics commonly used for quantifying the degree of linear correlation in pixel-by-pixel intensity between two different images [14, 37–39]. Owing to simplicity of usage and availability in most image analysis software packages (ImageJ, Colocalizer Pro), PCC is used increasingly in the literature of two-color fluorescence microscopy. Because it is pixel-based, PCC can in principle be applied to both wide-field and superresolution images [3]. The fluorescence intensity of individual pixels in widefield images is proportional to the number of emitted photons incident upon each pixel. The MPCC value can then be calculated using fluorescence intensity per pixel rather than molecules per pixel. Background subtraction to produce zero-based images is important.

For two-color, three-dimensional fluorescence microscopy [40, 41], the standard PCC would provide an accurate measure of linear correlation, assuming the 3D image matrices are sufficiently populated. However, by far the more common case of two-color microscopy projects the 3D spatial distributions onto the 2D camera plane. The central point of this work is simple. For most cell shapes, random 3D spatial distributions (no spatial correlations) do not make random 2D projections. In the particular case of spherocylindrical cells, projections of random 3D distributions are skewed to have more molecules/pixel in the central region compared to the edges and the endcap regions (Fig. 1). This renders the standard PCC reference matrices (Eq. 2), whose elements are the constant values \bar{R} and \bar{G} , highly inappropriate. As a result, the standard PCC fails both qualitatively and quantitatively to describe the nature and degree of the spatial correlation. A calculated PCC value of +1

could equally well arise from perfectly correlated 3D distributions (Additional file 1: Figure S2) or from completely random 3D distributions (Fig. 3). For strongly anti-correlated images, the degree of anti-correlation will be systematically underestimated (Fig. 2).

In the special case of spherocylindrical cells, we have described a method for calculating a modified Pearson correlation coefficient (MPCC) that uses the 2D projection matrices \tilde{U}_{ij}^R and \tilde{U}_{ij}^G derived from independent 3D random distributions as the reference matrices with which the 2D image matrices **R** and **G** are compared (Eq. 3). The resulting MPCC is normalized to lie in the range -1 to $+1$. Within noise limitations, MPCC approaches 0 for the projections of two distributions that are independent and random in 3D, approaches -1 for two distributions that are perfectly anti-correlated in both 3D and 2D, and approaches $+1$ for two distributions that are perfectly positively correlated in both 3D and 2D. Additionally, we have used the new procedure to estimate a positive value $\text{MPCC} = +0.50 \pm 0.05$ between experimentally obtained spatial localizations of individual RNAP and HU molecules in live *E. coli* cells (Fig. 4). Both RNAP and HU bind the chromosomal DNA, which occupies a subset of the cytoplasmic volume called the nucleoid. As expected, we obtain positive correlation that is significantly outside the range of model MPCC values computed for two uncorrelated distributions using the same pixel and copy number parameters as the experimental data.

While MPCC corrects a significant flaw in the standard PCC, it is important to note that for two images that are correlated or anti-correlated in 3D (i.e., not random), the MPCC applied to the 2D projections will typically underestimate the degree of correlation or anti-correlation in 3D. Projection from 3D to 2D always involves a loss of information. If the two 3D distributions are correlated or anti-correlated, their 2D projections will typically be less so. Our model correlated images (Additional file 1: Figure S2) and anti-correlated images (Fig. 2 and Additional file 1: Figure S1) are special cases in that they preserve perfect correlation or anti-correlation when projected from 3D to 2D. More irregular, less symmetric 3D distributions generally will not. This means that a 2D MPCC value that is not significantly different from zero does not imply the absence of 3D spatial correlations.

We have also shown how a small average number of molecules per pixel can cause systematic errors in MPCC values (Fig. 4, Additional file 1: Figures S5 and S8). For images sampled from both perfectly anti-correlated and perfectly correlated distributions,

this effect diminishes the magnitude of MPCC (biasing it towards zero). The minimum number of molecules per pixel required to obtain a trustworthy MPCC was ~ 7 for our model images but increased to ~ 30 for our experimental images. The MPCC user needs to measure a sufficient number of localizations and make a knowledgeable choice of pixel size based on the questions being asked. In each situation, by altering the pixelation or by randomly deleting copies, the user can determine how many copies per pixel is sufficient for the desired accuracy. As pixel size increases, spatial correlations on shorter length scales will be lost. It appears to us that MPCC will be most useful in exploring correlations on a length scale of ~ 200 nm or larger, as in our HU/RNAP example.

In earlier work applying PCC to eukaryotic cells, Dunn, et al. [23] warned against inclusion of empty extracellular regions in the image matrices **R** and **G**. Such extra zeroes alter the mean value in the references matrices and also artificially inflate the calculated PCC due to positive correlations between the empty regions in both matrices. They suggested carefully outlining only the regions of space that are occupied by the cells of interest. However, the MPCC is impervious to such extra zeroes. The mean pixel intensity over the region of interest does not participate in the calculation of MPCC. The empty regions of the image outside the cell boundary cause corresponding zeroes in the 2D projected reference matrix. They affect neither the normalization condition (Eq. 4 in Methods) nor the calculated MPCC (Eq. 3). In the MPCC procedure, one need not worry about empty regions of the image matrices that lie outside the projected cell boundaries.

The caveats outlined here apply to essentially all types of cells or organelles. In principle, the central concept of MPCC can be generalized to other cell geometries, including irregular eukaryotic shapes. However, the MPCC will probably find its greatest use in bacterial cells, whose shapes are often quite uniform for given growth conditions. It is relatively straightforward to simulate appropriate 2D reference distributions for rod-shaped bacteria like *E. coli* and *B. subtilis*, using a spherocylinder as the simplified model. The problem becomes more difficult for other shapes, such as the spiral shaped *H. pylori*.

One possible purely experimental solution would be to co-image a large population of freely diffusing fluorophores that presumably map out the 2D projection of a 3D random distribution in the cell volume of interest. To test this concept on *E. coli*, we have imaged Kaede under the same growth conditions used to image the RNAP and HU spatial distributions of Fig. 4. Kaede is a non-native tetrameric fluorescent protein that diffuses freely in *E. coli* and appears to fill the cytoplasm

uniformly [36]. We imaged Kaede in 15 cells of length 3.6 to 3.8 μm , the same length bin used for RNAP and HU. The composite distribution of 54,719 spatial localizations from 8 of the 15 cells was pixelated to give an experimental estimate of $\tilde{\mathbf{U}}^{\text{R}}$. An estimate of $\tilde{\mathbf{U}}^{\text{G}}$ was generated from the pixelated 2D projection of 66,301 Kaede copies from the other 7 cells. Using these experimentally generated matrices $\tilde{\mathbf{U}}^{\text{R}}$ and $\tilde{\mathbf{U}}^{\text{G}}$, we calculated MPCC for the same RNAP and HU spatial distributions to be 0.56, 0.42 and 0.32 for chosen pixel sizes of 200 nm, 105 nm and 50 nm respectively. These completely experimentally derived MPCC values are similar to the MPCC values of 0.51, 0.39 and 0.25 obtained from simulation of $\tilde{\mathbf{U}}^{\text{R}}$ and $\tilde{\mathbf{U}}^{\text{G}}$ for the same respective pixel sizes. For cases in which it is difficult to simulate the 3D cell geometry, the experimental approach to generation of the reference matrices may prove useful.

Conclusions

In this work, we have described a method of calculating a modified Pearson Correlation coefficient between two 2D fluorescence images of spherocylindrical cells. Calculation of traditional Pearson correlation coefficient uses a constant mean value of the image as the reference distribution. This leads to incorrect estimation of correlation coefficient both quantitatively and qualitatively. We have proposed a modified Pearson Correlation Coefficient (MPCC) that corrects this problem for spherocylindrical cell geometry by employing the proper reference matrices, 2D projection matrices derived from independent 3D random distributions in spherocylinders, for comparison with the images under analysis. MPCC can be employed for 2D superresolution as well as widefield images, conventionally acquired for wide variety of studies. The application of MPCC to irregularly shaped bacterial cells may be possible by imaging a large population of freely diffusing fluorophores that presumably serve as an experimental reference distribution. We demonstrated the applicability of MPCC to experimentally acquired superresolution images of RNAP and HU in *E. coli*, using both simulated and experimental reference distributions. MPCC will prove most useful in quantifying spatial correlation between two different fluorophore-labeled molecules on length scales comparable to the shortest cell dimensions.

Methods

As a first step towards generation of the matrices $\tilde{\mathbf{U}}^{\text{R}}$ and $\tilde{\mathbf{U}}^{\text{G}}$ in Eq. 3, a large number of molecules (here 100,000) are randomly distributed in a spherocylinder whose dimensions match those of the cells being imaged. We want $\tilde{\mathbf{U}}^{\text{R}}$ and $\tilde{\mathbf{U}}^{\text{G}}$ to have high signal-to-noise in each pixel. For a cell

of length 3.5 μm and width of 0.82 μm , the choice of 200 nm for the pixel size results in 75 pixels in the cell. 100,000 molecules makes the mean occupancy 1333 molecules/pixel. An appropriate localization error σ is applied to each particle location in both x and y coordinates by sampling a Gaussian distribution with standard deviation σ , yielding the model “measured” location of each molecule, which is binned appropriately. For generating 3D random distribution of molecules corresponding to the red and green channels, the localization error applied is the same as that measured upon imaging fluorescent molecules in red (σ_{R}) and green channels (σ_{G}) respectively. The 2D projections along the z axis of these two 3D random distributions give the matrices \mathbf{U}^{R} and \mathbf{U}^{G} . The elements U_{ij}^{R} and U_{ij}^{G} are positive integers.

Next the counts in individual pixels of \mathbf{U}^{R} and \mathbf{U}^{G} are normalized so that the total number of red and green molecules is equal to N_{R} and N_{G} , the total number of molecules imaged in each channel. This yields the normalized matrix $\tilde{\mathbf{U}}^{\text{R}}$:

$$\tilde{U}_{ij}^{\text{R}} = U_{ij}^{\text{R}} \times N_{\text{R}} / 100,000 \tag{4}$$

so that $\sum_{i=1}^m \sum_{j=1}^n \tilde{U}_{ij}^{\text{R}} = N_{\text{R}}$. Similarly, \mathbf{U}^{G} is normalized so that the sum of all elements of $\tilde{\mathbf{U}}^{\text{G}}$ is N_{G} .

We then subtracted $\tilde{\mathbf{U}}^{\text{R}}$ and $\tilde{\mathbf{U}}^{\text{G}}$ from the corresponding image matrix under analysis, \mathbf{R} and \mathbf{G} respectively to obtain the unnormalized difference matrices Δ^{R} and Δ^{G} . We normalized Δ^{R} and Δ^{G} so that the sum of the squares of individual pixel values in the difference matrix is 1:

$$\hat{\Delta}_{ij}^{\text{R}} = \frac{\Delta_{ij}^{\text{R}}}{\|\Delta^{\text{R}}\|}, \tag{5}$$

where $\|\Delta^{\text{R}}\| = \sqrt{\sum_{i=1}^m \sum_{j=1}^n \Delta_{ij}^{\text{R}2}}$. The resultant normalized 2D difference matrix $\hat{\Delta}^{\text{R}}$ has $\|\hat{\Delta}^{\text{R}}\| = 1$. The difference matrix Δ^{G} in the green channel is similarly normalized to obtain $\hat{\Delta}^{\text{G}}$ such that $\|\hat{\Delta}^{\text{G}}\| = 1$.

MPCC is obtained by taking the Frobenius inner product of the two normalized matrices $\hat{\Delta}^{\text{R}}$ and $\hat{\Delta}^{\text{G}}$:

$$\text{MPCC} = \sum_{i=1}^m \sum_{j=1}^n \hat{\Delta}_{ij}^{\text{R}} \hat{\Delta}_{ij}^{\text{G}}. \tag{6}$$

Additional file

Additional file 1: SI Text S1. Two additional cases of anti-correlation in 3D and in 2D. SI Text S2. Example of positive correlation. SI Text S3. Materials and Methods. SI Text S4. Determining cell width and length. SI Figures (S1-S8) and Captions. **Table S1.** Variation of MPCC with number of pixels N_p and with mean occupancy per pixel for the experimental RNAP (green) and HU (red) image matrices. (DOCX 2165 kb)

Abbreviations

2D: Two dimension; 3D: Three dimension; MPCC: Modified Pearson Correlation Coefficient; PCC: Pearson Correlation Coefficient

Acknowledgements

We would like to acknowledge Mr. Sayar Karmakar, University of Chicago for his insightful discussions on Pearson Correlation Coefficient.

Funding

This work was supported by a grant from the National Science Foundation (MCB-1512946 to JCW). Funding bodies did not play any role in design of the study, collection, analysis, and interpretation of the data and in writing of the manuscript.

Availability of data and materials

All data generated or analyzed during this study are included in this article and its Additional file 1. The datasets used and/or analyzed during the current study are also available from the corresponding author on reasonable request.

Authors' contributions

SM and JCW conceived and planned the project. SM performed the experiments and the data analysis. SM performed the analytical calculations. JCW supervised experimental and analytical findings. SM and JCW wrote the manuscript. Both authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 19 June 2018 Accepted: 22 October 2018

Published online: 16 November 2018

References

- Mondal J, Bratton BP, Li Y, Yethiraj A, Weisshaar James C. Entropy-based mechanism of ribosome-nucleoid segregation in *E. coli* cells. *Biophys J*. 2011;100(11):2605–13.
- Neeli-Venkata R, Martikainen A, Gupta A, Gonçalves N, Fonseca J, Ribeiro AS. Robustness of the process of nucleoid exclusion of protein aggregates in *Escherichia coli*. *J Bacteriol*. 2016;198(6):898–906.
- Aaron JS, Taylor AB, Chew T-L. Image co-localization – co-occurrence versus correlation. *J Cell Sci*. 2018;131(3):1–10.
- Owen DM, Rentero C, Rossy J, Magenau A, Williamson D, Rodriguez M, et al. PALM imaging and cluster analysis of protein heterogeneity at the cell surface. *J Biophotonics*. 2010;3(7):446–54.
- Perry GLW. SpPack: spatial point pattern analysis in excel using visual basic for applications (VBA). *Environ Model Softw*. 2004;19(6):559–69.
- Ripley BD. Tests of 'Randomness' for spatial point patterns. *J R Stat Soc Ser B Methodol*. 1979;41(3):368–74.
- Sengupta P, Jovanovic-Talisman T, Skoko D, Renz M, Veatch SL, Lippincott-Schwartz J. Probing protein heterogeneity in the plasma membrane using PALM and pair correlation analysis. *Nat Methods*. 2011;8:969.
- Veatch SL, Machta BB, Shelby SA, Chiang EN, Holowka DA, Baird BA. Correlation functions quantify super-resolution images and estimate apparent clustering due to over-counting. *PLoS One*. 2012;7(2):e31457.
- Stone MB, Veatch SL. Steady-state cross-correlations for live two-colour super-resolution localization data sets. *Nat Commun*. 2015;6:7347.
- Schnitzbauer J, Wang Y, Zhao S, Bakalar M, Nuwal T, Chen B, et al. Correlation analysis framework for localization-based superresolution microscopy. *Proc Natl Acad Sci USA*. 2018;115(13):3219–24.
- Bakshi S, Siryaporn A, Goulian M, Weisshaar JC. Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. *Mol Microbiol*. 2012;85(1):21–38.
- Mohapatra S, Weisshaar JC. Functional mapping of the *E. coli* translational machinery using single-molecule tracking. *Mol Microbiol*. 2018;110(2):262–82.
- Lee Steven F, Thompson Michael A, Schwartz MA, Shapiro L, Moerner WE. Super-resolution imaging of the nucleoid-associated protein HU in *Caulobacter crescentus*. *Biophys J*. 2011;100(7):L31–L3.
- Manders EM, Stap J, Brakenhoff GJ, van Driel R, Aten JA. Dynamics of three-dimensional replication patterns during the S-phase, analysed by double labelling of DNA and confocal microscopy. *J Cell Sci*. 1992;103(3):857.
- Pearson K. Mathematical contributions to the theory of evolution III. Regression, heredity, and panmixia. *Philos Trans R Soc Lond Ser B Biol Sci*. 1896;187:253.
- Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012;9(7):671–5.
- Boite S, Cordelières FP. A guided tour into subcellular colocalization analysis in light microscopy. *J Microsc*. 2006;224(3):213–32.
- Karunatilaka KS, Cameron EA, Martens EC, Koropatkin NM, Biteen JS. Superresolution imaging captures carbohydrate utilization dynamics in human gut symbionts. *MBio*. 2014;5(6):e02172–14.
- Männik J, Wu F, Hol FJH, Bisicchia P, Sherratt DJ, Keymer JE, et al. Robustness and accuracy of cell division in *Escherichia coli* in diverse cell shapes. *Proc Natl Acad Sci*. 2012;109(18):6957–62.
- Strahl H, Bürmann F, Hamoen LW. The actin homologue MreB organizes the bacterial cell membrane. *Nat Commun*. 2014;5:3442.
- Costes SV, Daelemans D, Cho EH, Dobbin Z, Pavlakis G, Lockett S. Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophys J*. 2004;86(6):3993–4003.
- Dedecker P, Mo GCH, Dertinger T, Zhang J. Widely accessible method for superresolution fluorescence imaging of living systems. *Proc Natl Acad Sci U S A*. 2012;109(27):10909–14.
- Dunn KW, Kamocka MM, McDonald JH. A practical guide to evaluating colocalization in biological microscopy. *Am J Physiol Cell Physiol*. 2011;300(4):C723–42.
- Earle Kristen A, Billings G, Sigal M, Lichtman Joshua S, Hansson Gunnar C, Elias Joshua E, et al. Quantitative imaging of gut microbiota spatial organization. *Cell Host Microbe*. 2015;18(4):478–88.
- Skočaj M, Resnik N, Grundner M, Ota K, Rojko N, Hodnik V, et al. Tracking cholesterol/sphingomyelin-rich membrane domains with the osteolysin A-mCherry protein. *PLoS One*. 2014;9(3):e92783.
- Wu Z, Tang M, Tian T, Wu J, Deng Y, Dong X, et al. A specific probe for two-photon fluorescence lysosomal imaging. *Talanta*. 2011;87:216–21.
- George TC, Fanning SL, Fitzgerald-Bocarsly P, Medeiros RB, Highfill S, Shimizu Y, et al. Quantitative measurement of nuclear translocation events using similarity analysis of multispectral cellular images obtained in flow. *J Immunol Methods*. 2006;311(1):117–29.
- Cabrera JE, Jin DJ. The distribution of RNA polymerase in *Escherichia coli* is dynamic and sensitive to environmental cues. *Mol Microbiol*. 2003;50(5):1493–505.
- Castaing B, Zelwer C, Laval J, Boiteux S. HU protein of *Escherichia coli* binds specifically to DNA that contains single-strand breaks or gaps. *J Biol Chem*. 1995;270(17):10291–6.
- Wang W, Li G-W, Chen C, Xie XS, Zhuang X. Chromosome organization by a nucleoid-associated protein in live bacteria. *Science*. 2011;333(6048):1445.
- Nielsen HJ, Ottesen JR, Youngren B, Austin SJ, Hansen FG. The *Escherichia coli* chromosome is organized with the left and right chromosome arms in separate cell halves. *Mol Microbiol*. 2006;62(2):331–8.
- Biteen JS, Thompson MA, Tselentis NK, Bowman GR, Shapiro L, Moerner WE. Super-resolution imaging in live *Caulobacter crescentus* cells using photoswitchable EYFP. *Nat Methods*. 2008;5(11):947–9.
- Subach FV, Patterson GH, Manley S, Gillette JM, Lippincott-Schwartz J, Verkhusa W. Photoactivatable mCherry for high-resolution two-color fluorescence microscopy. *Nat Methods*. 2009;6(2):153–9.
- Ando R, Hama H, Yamamoto-Hino M, Mizuno H, Miyawaki A. An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *Proc Natl Acad Sci U S A*. 2002;99(20):12651–6.
- Hayashi I, Mizuno H, Tong KI, Furuta T, Tanaka F, Yoshimura M, et al. Crystallographic evidence for water-assisted photo-induced peptide cleavage in the stony coral fluorescent protein Kaede. *J Mol Biol*. 2007;372(4):918–26.

36. Bakshi S, Bratton Benjamin P, Weisshaar James C. Subdiffraction-limit study of Kaede diffusion and spatial distribution in live *Escherichia coli*. *Biophys J*. 2011;101(10):2535–44.
37. Adler J, Parmryd I. Quantifying colocalization by correlation: the Pearson correlation coefficient is superior to the Mander's overlap coefficient. *Cytometry A*. 2010;77A(8):733–42.
38. McDonald JH, Dunn KW. Statistical tests for measures of colocalization in biological microscopy. *J Microsc*. 2013;252(3):295–302.
39. Lagache T, Sauvonnnet N, Danglot L, Olivo-Marin J-C. Statistical analysis of molecule colocalization in bioimaging. *Cytometry A*. 2015;87(6):568–79.
40. Fiolka R, Shao L, Rego EH, Davidson MW, Gustafsson MGL. Time-lapse two-color 3D imaging of live cells with doubled resolution using structured illumination. *Proc Natl Acad Sci*. 2012;109(14):5311–5.
41. Jones SA, Shim S-H, He J, Zhuang X. Fast, three-dimensional super-resolution imaging of live cells. *Nat Methods*. 2011;8(6):499–505.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

