

Colocalization Analysis in Fluorescence Micrographs: Verification of a More Accurate Calculation of Pearson's Correlation Coefficient

Andrew L. Barlow,¹ Alasdair MacLeod,¹ Samuel Noppen,² Jeremy Sanderson,^{2,†} and Christopher J. Guérin^{2,*}

¹PerkinElmer, Viscount Centre II, Millburn Hill Road, Warwick University Science Park, Coventry CV4 7HS, UK

²Microscopy Core Facility, Department for Molecular Biomedical Research, Flanders Institute of Biotechnology (VIB), and University of Ghent, Fiers-Schell-Van Montagu Building, Technologiepark 927, B-9052 Ghent, Belgium

Abstract: One of the most routine uses of fluorescence microscopy is colocalization, i.e., the demonstration of a relationship between pairs of biological molecules. Frequently this is presented simplistically by the use of overlays of red and green images, with areas of yellow indicating colocalization of the molecules. Colocalization data are rarely quantified and can be misleading. Our results from both synthetic and biological datasets demonstrate that the generation of Pearson's correlation coefficient between pairs of images can overestimate positive correlation and fail to demonstrate negative correlation. We have demonstrated that the calculation of a thresholded Pearson's correlation coefficient using only intensity values over a determined threshold in both channels produces numerical values that more accurately describe both synthetic datasets and biological examples. Its use will bring clarity and accuracy to colocalization studies using fluorescent microscopy.

Key words: colocalization, fluorescence microscopy, Pearson's correlation, algorithm, quantitation

INTRODUCTION

Demonstrating a relationship between two biological molecules is a common research question that is tackled by numerous techniques including coprecipitation, yeast two hybrid assays, and fluorescent microscopy of labeled molecules. Detecting colocalization microscopically involves collecting two fluorescent signals, one from each putatively-related molecule and in some way correlating them. Researchers, even in high impact journals, primarily rely on the artifact-prone demonstration of yellow structures where red and green structures overlap (see Supplementary Table 1).

Supplementary Table 1

Supplementary Table 1, which shows the results of a survey of quantitative colocalization in three high impact journals, can be found online. Please visit journals.cambridge.org/jid_MAM.

More rigorous studies use statistical analysis (Bolte & Cordelières, 2006). No single technique has been established as

a standard method; indeed, the discipline has been described as being afflicted with ambiguity and inconsistency (Bolte & Cordelières, 2006). Lack of understanding of the limitations of fluorescence microscopy may account for the frequency of the red-plus-green-gives-yellow technique in publications. Nonetheless, this is a source of frustration to those who are aware of this technique's shortcomings (Bolte & Cordelières, 2006; North, 2006; Adler et al., 2008; French et al., 2008; Scriven et al., 2008). These include a lack of objectivity, failure to identify covariance of intensities in the paired datasets, and possible overemphasis of trivial relationships (Costes et al., 2004; Li et al., 2004; Adler et al., 2008). While the present methods of mathematical analysis have their faults, it is a matter of much graver concern that so few scientists attempt to calculate fluorescence correlation at all (see Supplementary Table 1).

Information can be gleaned about the relationship between intensities in a pair of images by plotting pixels according to their intensity in each channel in a two-dimensional (2D) scatter plot. In such plots spatial information is lost in favor of a means of assessing the degree of covariance of the image pair, as well as clues to the degree of bleed-through and noise in the images (Bolte & Cordelières, 2006). Such scatter plots are commonly generated by image analysis programs.

Manders et al. (1992) improved fluorescent colocalization by introducing Pearson's correlation coefficient (PCC) to microscopists. PCC is a standard statistical analysis de-

Received April 15, 2010; accepted July 6, 2010

*Corresponding author. E-mail: chris.guerin@dmb.vib-ugent.be

†Current address: MRC Harwell, Harwell Science and Innovation Campus, Oxford OX11 0RD, UK

signed to measure the strength of a linear relationship between two variables, in this case fluorescent intensities from two images. The rationale is that a genuine physical relationship between the molecules can be described by a statistical relationship. To calculate PCC for a pair of fluorescence images, all of the pixels having the same image coordinates are paired. The mean pixel intensity of an image is subtracted from the intensity of each pixel within the image, and the value generated for each pixel is multiplied by the equivalent value from the pixel's partner in the counterpart image to generate the product of the difference from the mean. The product of the difference from the mean is summed for the entire dataset and divided by the maximum possible sum of the product of the difference from the mean (Adler et al., 2008). PCC generates a range of values from 1, a perfect positive correlation, to -1 , a perfect but inverse correlation, with 0 representing a random distribution. Manders et al., analyzing double labeled DNA, calculated PCC for all of the pixels in their image pairs, regardless of whether they contained data from positively labeled structures or background intensities. Such calculations generate "global" values for PCC (Manders et al., 1992, 1993).

$$PCC = \frac{\sum (R_i - R_{aver}) \cdot (G_i - G_{aver})}{\sqrt{\sum (R_i - R_{aver})^2 \cdot \sum (G_i - G_{aver})^2}}$$

Pearson's correlation coefficient

R_i = intensity in red channel

R_{aver} = average intensity in red channel

G_i = intensity in green channel

G_{aver} = average intensity in green channel

Concerned that the negative values that can be generated by PCC would be difficult to interpret when the degree of overlap is of interest, Manders et al. (1993) dropped the subtraction of mean values from the calculation of PCC, generating Manders' overlap coefficient, which generates only positive numbers between 0 and 1.

$$MOC = \frac{\sum (R_i) \cdot (G_i)}{\sqrt{\sum (R_i)^2 \cdot \sum (G_i)^2}}$$

Further descriptors of colocalization were introduced including Manders' colocalization coefficients, M_1 and M_2 . These coefficients provide a measure of the total colocalized intensity divided by the total intensity for each channel. M_1 and M_2 require a threshold that demarcates signal from background, with only the signal included in the calculations of these coefficients. Manders et al. worked exclusively with

background corrected data, with a fixed threshold of 0. Manders et al. demonstrated the utility of these coefficients with a series a test images composed of patterns of Gaussian spots, in which various degrees of correlation could be clearly illustrated (Table 1).

$$M_1 = \frac{\sum R_{i,coloc}}{\sum R_i}$$

Overlap coefficient M_1

$R_{i,coloc} = R_i$ if $R_i > T_1$ and $G_i > T_2$

R_i = red intensity if $R_i > T_1$

$$M_2 = \frac{\sum G_{i,coloc}}{\sum G_i}$$

Overlap coefficient M_2

$G_{i,coloc} = G_i$ if $G_i > T_2$ and $R_i > T_1$

G_i = green intensity if $G_i > T_2$

R_i = pixel intensity of red channel

G_i = pixel intensity of green channel

T_1 = threshold for red channel

T_2 = threshold for green channel.

M_1 and M_2 can also be calculated by setting a threshold to match the estimated value of background instead of zero (Bolte & Cordelières, 2006).

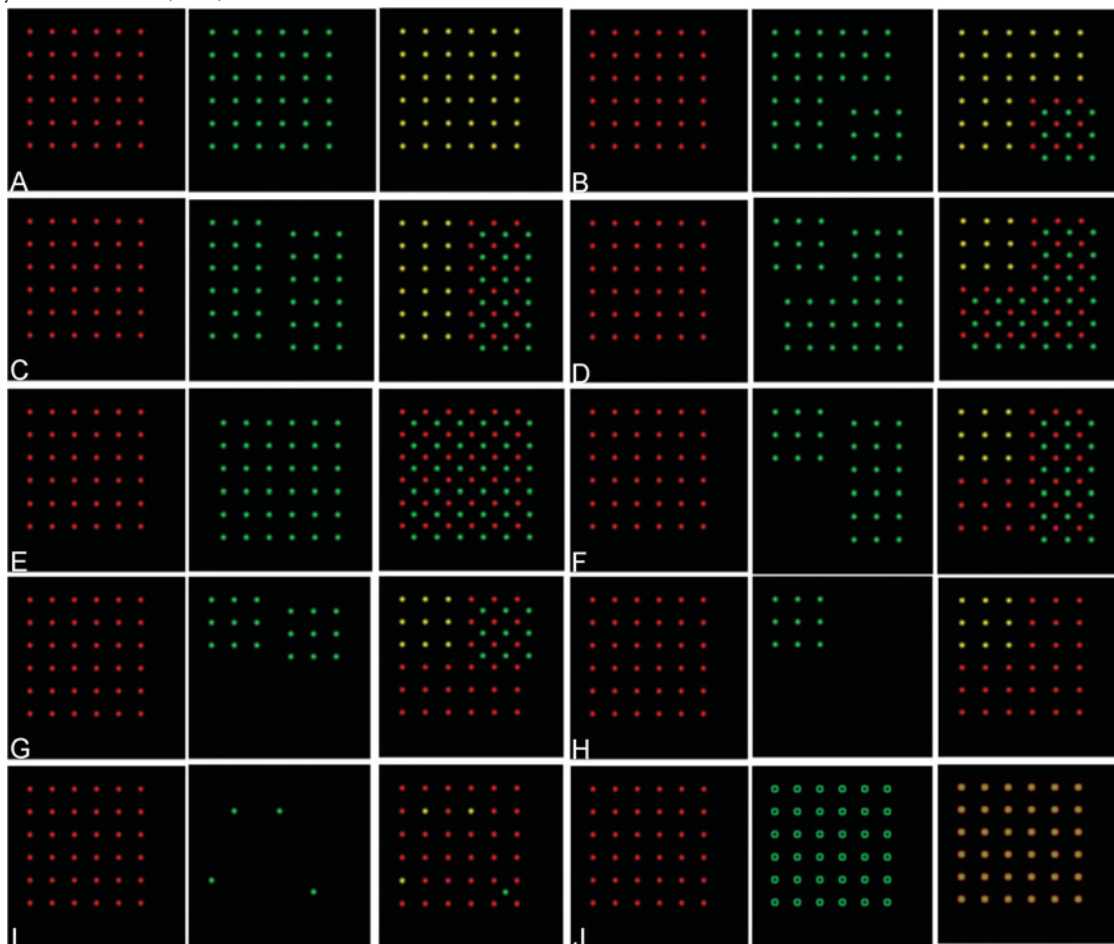
Further developments with PCC M_1 and M_2 were made in the widely cited study of Costes et al. (2004). Costes et al. (2004) stated that PCC is a quantitative estimate of colocalization, but prefer M_1 and M_2 , which they consider to be a more biologically meaningful set of coefficients. They therefore developed an algorithm that uses PCC to objectively set thresholds from which variants of M_1 and M_2 are calculated.

$$M_1 = \frac{\sum_{I_1 > T_1} I_1}{\sum_{All I_1} I_1} \quad M_2 = \frac{\sum_{I_2 > T_2} I_2}{\sum_{All I_2} I_2}$$

Costes et al. variants of M_1 and M_2

T_1 = threshold set by the automated algorithm for channel 1

T_2 = threshold set by the automated algorithm for channel 2

Table 1. Results of a Global PCC Colocalization Analysis of Synthetic Gaussian Spot Patterns Similar to Those Used by Manders et al. (1993).*

Test Pattern	M_1	M_2	Global PCC Velocity 5.3	PCC JACoP
A	1	1	1.0	1.0
B	0.75	0.75	0.742	0.741
C	0.5	0.5	0.483	0.483
D	0.25	0.25	0.225	0.224
E	0	0	-0.034	-0.033
F	0.25	0.33	0.268	0.268
G	0.25	0.5	0.339	0.338
H	0.25	1.0	0.494	0.493
I	0.083	0.75	0.244	0.243
J	0.992	1.0	0.495	0.495

*Spot patterns simulating two image pairs were analyzed with JACoP (and Velocity 5.3). Global PCC yields a positive PCC for image pair J that is perfectly anticorrelated for all intensities above the background.

I_1 = pixel intensity for channel 1

I_2 = pixel intensity for channel 2

All I_1 = all pixel intensities for channel 1

All I_2 = all pixel intensities for channel 2.

First, a least-squares fit is calculated for the 2D scattergram based on orthogonal regression. The algorithm progressively lowers intensity thresholds for both axes of the scatterplot and calculates the PCC between the pixels beneath the intensity thresholds of each channel. Once a pair of thresholds is established beneath which no positive PCC exists, the established thresholds are considered to have

been set in an objective fashion, and the variants of M_1 and M_2 are calculated. Costes' variants of M_1 and M_2 show the proportion of intensities above background (assumed to be zero) to which the positive correlation applies. Therefore, PCC is used as a tool for making objective measurements of Costes' variant of M_1 and M_2 .

The approach of Costes et al. therefore only considers positively correlated populations of pixel pairs to be of interest—uncorrelated and anticorrelated pixel pairs are ignored by their approach. In the same paper, they also stated that “negative values of PCC are not used for colocalization because they indicate an anticorrelation situation where a pixel is bright in one channel and dim in the other.” They also introduced a statistical significance test based on image randomization. Image 1 is randomized by shuffling pixel blocks, the size of which is based on the imaging system's point spread function. PCC between the unscrambled image 2 and the scrambled image 1 is compared over 200 different sets of randomizations so that the statistical significance of PCC can be calculated (Costes et al., 2004; Bolte & Cordelières, 2006).

In an extensive review of colocalization methods, Bolte and Cordelières (2006) cautioned that PCC rarely discriminates differences between partial colocalization or exclusion, especially if images contain noise, and that midrange values (−0.5 to 0.5) do not allow conclusions to be drawn. These assertions are the subject of some debate (Adler & Parmryd, 2007; Bolte & Cordelières, 2007), but it is of interest that both Costes et al. (2004) and Bolte and Cordelières (2006) effectively abandoned half the range of PCC, although the element they chose to abandon was not the same.

The work of Manders et al. (1992, 1993) and Costes et al. (2004) set a standard that has been implemented in numerous image analysis software packages commonly found in academic imaging facilities. As part of their extensive review, Bolte and Cordelières (2006) implemented these and other colocalization methods in a public domain tool JACoP (just another colocalization plugin), a plugin for the public domain Image J software (Rasband, 1997–2010).

By repeating Manders' work with test patterns of Gaussian spots, augmented with additional test patterns and biological datasets, we have clearly illustrated the shortcomings in current implementations of PCC calculation; therefore, we propose a modified method of calculating PCC. Our results demonstrate our method corrects defects in current implementations of PCC. This change generates a broader range of values for PCC, including negative correlations, while preventing dilution of positive relationships thus preserving important biologically relevant conclusions.

MATERIALS AND METHODS

Synthetic Spot Patterns

Approximate replicates of the 2D test patterns produced by Manders et al. (1993) were made using GraphicConverter

(available at www.lemkesoft.com/content/188/graphic_converter.html). Each image has dimensions of 256×256 pixels. Gaussian spots with a radius of 6 pixels were added in the positions described by Manders et al. (1993). All images are 8-bit with a background value of 0 and a maximum intensity of 245. To demonstrate exclusion, the base pattern of 36 spots was inverted using Graphic Converter, and the gray level of the area outside the spots was reset to zero using the fill tool. This pattern has a maximum intensity of 241. This pattern is a perfectly inverted version of the base pattern for all intensities above the background gray level of zero.

To explore the sensitivity of PCC, the red image of combination A was progressively shifted by increments of 1 pixel to the left from 0 to 13 pixels to generate a series of shifted images using Graphic Converter's picture shift function. Both thresholded PCC and global PCC were calculated for combinations of the green image and each of the shifted versions of the red image.

Test image pairs with a single Gaussian spot with a radius of 6 pixels were made. These images were made at a range of sizes to test the effect of background on PCC; image sizes were 20×20 , 60×60 , and 120×120 pixels. All images were 8 bit with a background value of 0 and a maximum intensity of 245. One of each of these test images was progressively shifted by increments of 1 pixel to the left from 0 to 11 pixels to generate a series of shifted images using Graphic Converter's picture shift function. Both global and thresholded values for PCC and Manders' overlap coefficient were calculated for these patterns in their original and each of the shifted positions.

Image Thresholding

The commercial image analysis program Velocity 5.3 calculates variants of M_1 and M_2 with a user defined threshold that separates signal from background (as described by Bolte & Cordelières, 2006). These coefficients are described as M_x and M_y throughout this article and are shown below.

$$M_x = \frac{\sum R_{i,coloc}}{\sum R_i}$$

Overlap coefficient M_x

$$R_{i,coloc} = R_i \text{ if } R_i > T_1 \text{ and } G_i > T_2$$

$$R_i = \text{red intensity if } R_i > T_1$$

$$M_y = \frac{\sum G_{i,coloc}}{\sum G_i}$$

Overlap coefficient M_y

$$G_{i,coloc} = G_i \text{ if } G_i > T_2 \text{ and } R_i > T_1$$

$$G_i = \text{green intensity if } G_i > T_2$$

R_i = pixel intensity of red channel

G_i = pixel intensity of green channel

T_1 = threshold for red channel

T_2 = threshold for green channel.

These coefficients are designed to work with a background level that is not zero, in which case a threshold to exclude background can be set. When a threshold of 1 is set, these coefficients behave exactly as M_1 and M_2 described by Manders et al. (1993). With a threshold of 1 set, these coefficients are described as M_1 and M_2 throughout this article. M_1 and M_2 were also calculated with JACoP to confirm the values generated by Volocity.

In the synthetic datasets described here, all pixels outside of the Gaussian spots have a gray level of zero. The threshold was therefore set to 1 for each image.

Global Measurements of PCC

Global PCC values were calculated in JACoP. All pixel pairs are included in this calculation. These coefficients are referred to as global PCC throughout this article.

Thresholded PCC

PCC was calculated as a thresholded coefficient in Volocity 5.3 by ignoring pixel pairs with a gray level beneath the threshold of either image. Thresholded PCC uses the same thresholds as M_x and M_y . PCC was then calculated only from those pixels that are above the threshold in *both* channels. To calculate PCC for this subset of pixels, the mean must be calculated from the same subset of pixels. These coefficients are referred to as thresholded PCC throughout this article.

The test images were analyzed using Volocity 5.3 (Perkin-Elmer, Coventry, UK) and FIJI (available at <http://pacific.mpi-cbg.de>) using the “JACoP” plugin (available at http://imagejdocu.tudor.lu/doku.php?id=plugin:analysis:jacop_2.0:just_another_colocalization_plugin:start). In all cases the threshold value was set to 1, except for when the method of Costes et al. (2004) was used to automatically calculate thresholds. Global PCC, M_1 and M_2 were calculated using JACoP. Thresholded PCC and Manders’ overlap coefficient were calculated using Volocity 5.3.

Multispeck Bead Data

An Invitrogen Multispeck multispectral fluorescence microscopy standards kit was used to prepare test slides composed

of beads with 4 μm diameter (M-7901 www.invitrogen.com). First, the Multispeck component of the kit was dried down onto a microscope slide. Multispeck beads are labeled with a mixture of blue (Ex 365 Em 405), green (Ex 520 Em 525), and red (Ex 580 Em 600) emitting dyes. Then the RGB mix suspension was dried down on top of the Multispeck beads. The RGB mix contains a mixture of singly labeled beads with the same excitation and emission properties as the multispeck beads, but in separate beads. Each stock solution was diluted 1:5 in water before being dried down. Images of this bead mixture were acquired with a Leica SP5 laser scanning confocal microscope (Leica Microsystems, Wetzlar, Germany; www.leica-microsystems.com). Red and green images of the beads were acquired. Images of the green dyes were acquired with a 488 nm laser and an acousto-optic tunable filter (AOTF) filter set to 520–550 nm. Images of the red dyes were acquired with a 543 nm laser and an AOTF filter set to 590–655 nm.

Cell Culture

HEK 293T cells (gift from Dr. M. Hall) and HeLa cells (ATCC, Manassas, VA, USA; www.atcc.org) were maintained in Dulbecco’s modified Eagle’s medium, enriched with 10% heat inactivated fetal calf serum, 2 mM L-glutamine, 1 mM sodium pyruvate, 1% penicillin, and 1% streptomycin at 37°C in 5% CO_2 . The cells were transferred and grown in Lab-tek 8 well chambers (Nalge Nunc International, Rochester, NY, USA) in the same medium and atmospheric condition (5,000 cells in 500 μL medium per well). After reaching the desired confluence, cells were either imaged live in RPMI without phenol red, or fixed in 4% paraformaldehyde in 0.1 M sodium cacodylate buffer at pH 7.4.

MitoTracker and Dapi

Cultured cells were stained with MitoTracker orange CMT-MRos (Invitrogen) at concentrations of 50 nM for 20 min. Cultures were fixed in 4% paraformaldehyde and stained with DAPI added at a final concentration of 5 nM (Invitrogen Corporation, Carlsbad, CA). Live cell imaging was performed with cells in RPMI without phenol red and imaged between 5 and 60 min after labeling.

Transferrin Green and Orange

Cultured cells were stained with a mixture of equal amounts of transferrin AlexaFluor 488 and transferrin AlexaFluor 568 (both Invitrogen, final concentration 10 $\mu\text{g}/\text{mL}$) in RPMI without phenol red and imaged between 10 and 60 min after labeling.

Histone H2B-GFP and Hoechst

Cultured cells were transduced with Cellular Lights Histone 2B green fluorescent protein (GFP) baculovirus vector (Invitrogen) according to the supplied protocol. The following day cells were stained with Hoechst 33342 (Invitrogen) at a

final concentration of 1 $\mu\text{g}/\text{mL}$ in RPMI without phenol red for 15 min then imaged on the confocal microscope.

Confocal Fluorescence Microscopy

All cells were imaged using a Leica SP5 laser scanning confocal microscope and a $63\times$ 1.4 oil immersion planapochromatic lens. Images of fluorochromes were obtained sequentially using a 488 nm laser line and an AOTF with a detection gate of 505–530 nm for green, a 543 nm laser line and a detection gate of 560–620 nm for orange, and a 405 nm laser line and a detection gate of 410–440 nm for Hoechst/DAPI. When obtaining Z stacks, Nyquist sampling criteria were followed. Colocalization analysis of fluorescence microscopic images of cells was conducted using Volocity 5.3. A 3×3 median filter was applied to images before threshold analysis to reduce the influence of noise (Landmann & Marbet, 2004; Agnati et al., 2005). Threshold values for each image were set to the background mean intensity plus three standard deviations.

RESULTS

Synthesized Test Patterns

When global PCC was calculated for approximate copies of the Manders test patterns, correlation coefficients that were similar to those found in Manders et al. (1993) were generated (Table 1).

When a negatively correlated dataset was analyzed (see Table 1J), a global PCC close to 0.5 was generated by JACoP and Volocity 5.3 (as well as other commercial software packages, data not shown). While this accurately reflects the genuine good correlation between the two images, it does not correctly describe the inverse correlation of the intensities within the images (greater than the background). The reason that the two images yield a positive correlation is that inclusion of background pixels in the calculation of the mean generates an unrealistically low number (4.8 for each channel in test pattern J, Table 1). All of the pixels in the image are included in the calculation of the mean, and the threshold values are not used in the calculation of PCC. This creates a situation in which almost all of the intensities brighter than the background are above the mean, so almost exclusively positive correlation values are generated. In addition, the intensities of background pixels, which are below the mean in both channels and therefore positively correlated, contribute further positive values to the global PCC. This results from the inclusion of background pixels in the analysis that produces a bias toward positive values.

We attempted to calculate a value for PCC that is a more accurate reflection of the relationship between the intensities above the background in the images by only calculating thresholded PCC from pixel pairs that are above the threshold in both images. This requires that the mean for each image is set from this same subset of pixels. This

generates significantly larger means (90 for each image in test pattern J). Many of the intensities above the threshold in the test images are below the larger means, so that large negative correlation values can be correctly generated. The thresholded PCC for test pattern combination J is -1 , the expected value for a perfectly inversely correlated dataset (Table 2).

When this method of calculating thresholded PCC is applied to the Manders' test patterns A to I, a PCC of 1 is generated for all the images with the exception of combination E, in which there are no pixel pairs over the threshold in either image (Table 2). This may be considered an unhelpful result, until it is interpreted with the benefit of M_1 and M_2 , which clearly illustrate the proportion of nonbackground intensities to which this perfect correlation applies (Table 2). A perfect correlation does exist between a proportion of the intensities in all of the test patterns (with the exception of pattern E), but this perfect correlation is "diluted out" of measurements of global PCC. When only pixel pairs that are above the threshold in both channels are used for the calculation of thresholded PCC, the perfect correlation is preserved. Pearson's correlation coefficient is generally used to demonstrate either a positive correlation, an inverse correlation, or a random relationship between two images. Thresholded PCC reveals a fourth possibility in test pattern E, namely a complete absence of any relationship between the two channels. This nonrelationship is exposed when no pixels are above the thresholds in both channels.

Displaced Spot Patterns

We manipulated Manders' spot pattern A by displacing the red image in 1 pixel increments, and calculated both global and thresholded PCC at each increment (Table 3). Thresholded PCC detects these displacements, becoming increasingly negative after a displacement of four pixels. Global PCC reports positive values for displacements up to 7 pixels, greater than the radius of the Gaussian objects. With displacements of 8 pixels or more, global PCC values become slightly negative but fail to detect objects that are strongly anticorrelated. Thresholded PCC generates a slightly curious result with a displacement of 11 pixels, for which the correlation coefficient is 1. Close inspection of the dataset reveals that this result is generated when only 4 peripheral pixels from the edges of the Gaussian objects overlap, and the four pixels have the same gray levels in both the red and green channels (2, 5, 5, 2). As these four pixels per pair of Gaussian objects are the only pixels over the threshold in both channels and therefore the only pixels analyzed, the correlation coefficient is correctly calculated as 1. M_1 and M_2 at 0.002 reveal that this perfect correlation applies to only a tiny proportion of the dataset (0.2% of the total intensities above the threshold in each image).

We further analyzed the displaced spot patterns with the "JACoP" plugin after selecting the "Costes automatic thresholding" option (see Supplementary Table 2).

Table 2. Results of a Thresholded Analysis of the Synthetic Spot Patterns in Table 1.*

Test Pattern Combination	M_1	M_2	Thresholded PCC	Description	Interpretation
A	1	1	1.0	36 green spots 36 red spots	Perfect correlation for all intensities
B	0.75	0.75	1.0	36 green spots 36 red spots. 9 green spots displaced	Perfect correlation for 75% of green and red intensities
C	0.5	0.5	1.0	36 green spots 36 red spots 18 green spots displaced	Perfect correlation for 50% of green and red intensities
D	0.25	0.25	1.0	36 green spots 36 red spots 27 green spots displaced	Perfect correlation for 25% of green and red intensities
E Uncorrelated	0	0	No relationship	36 green spots 36 red spots All 36 green spots displaced	No relationship
F	0.25	0.33	1.0	27 green spots 36 red spots 18 green spots displaced	Perfect correlation between 33% of the green intensities and 25% of the red intensities
G	0.25	0.5	1.0	18 green spots 36 red spots 9 green spots displaced	Perfect correlation between 50% of the green intensities and 25% of red intensities
H	0.25	1.0	1.0	9 green spots 36 red spots	Perfect correlation between all of the green intensities and 25% of red intensities
I	0.083	0.75	1.0	4 green spots 36 red spots 1 green spot displaced	Perfect correlation between 75% of green intensities and 0.08% of red intensities
J Anticorrelated	0.922	1.0	-1	36 green spots 36 inverse red spots	A perfect inverse correlation between all green and 99.2% of red intensities

*Thresholded PCC calculated in Volocity 5.3 yields results that correctly describe the relationship between nonbackground intensities in the image pairs, in particular yielding the expected value of -1 for the anticorrelated image pair J.

Supplementary Table 2















Supplementary Table 2, which shows the results of an analysis of displaced spot pattern A, can be found online. Please visit journals.cambridge.org/jid_MAM.

This plugin sets thresholds for each image using the algorithm of Costes et al. (2004) and calculates PCC for pixels brighter than the thresholds, and PCC for pixels fainter than one or both of the thresholds. The Costes algorithm calculates threshold values for image pairs with a displacement up to 7 pixels, beyond which no positive correlations exist within the images and the algorithm (which is only designed to work with positively correlated data) sets the thresholds to the maximum gray level within the images. When the same threshold values generated by the Costes' automated thresholding were used to calculate thresholded PCC with Volocity 5.3, the same values were generated as those generated by JACoP for the pixels greater than the thresholds (see Supplementary Table 2).

The Effect of Background

A single pair of Gaussian objects (radius 6) was added to the center of image pairs of various sizes. One of the pair of test images was moved to the left in increments of 1 pixel from between 0 and 11 pixels, and global PCC, thresholded PCC, and Manders' overlap coefficient (MOC) were calculated (Table 4). MOC was unaffected by the change in the amount of background in the images (this is only the case for MOC when nonsignal pixels have an intensity of 0); however, global PCC varies depending on the amount of background. At high degrees of offset, global PCC is able to generate negative values only when the amount of background in the images is low. As the amount of background pixels increase relative to the degree of nonbackground pixels, global PCC starts to converge with MOC. Thresholded PCC is always identical, regardless of the number of background pixels in the images. This convergence of global PCC and MOC occurs as increasing numbers of background pixels are included in the calculation of the mean for each channel. This reduces the mean and makes its

Table 3. Results of Thresholded and Global PCC Analysis of an Experiment Where Gaussian Spots Were Progressively Displaced.*

	Number of Pixels Shifted	Thresholded PCC Velocity 5.3	Global PCC JACoP	M_1	M_2
	0	1.0	1.0	1.0	1.0
	1	0.886	0.951	0.994	0.994
	2	0.549	0.817	0.957	0.957
	3	0.121	0.632	0.877	0.877
	4	-0.235	0.433	0.752	0.752
	5	-0.397	0.257	0.595	0.595
	6	-0.465	0.123	0.425	0.425
	7	-0.500	0.037	0.286	0.286
	8	-0.561	-0.009	0.144	0.144
	9	-0.544	-0.028	0.060	0.060
	10	-0.914	-0.033	0.017	0.017
	11	1.00	-0.034	0.002	0.002
	12	No relationship	-0.034	0.000	0.000
	13	No relationship	-0.034	0.000	0.000

*Thresholded and global analysis of a pattern where Gaussian spots were progressively displaced demonstrates the greater sensitivity of thresholded PCC. After a displacement of only 4 pixels, thresholded PCC becomes negative, but global PCC remains positive until a pixel shift of 8 is reached. Even at extreme displacements, global PCC is only slightly negative showing the strong positive bias in this method of analysis.

subtraction from the intensities of each channel in the calculation of PCC have little effect. Eventually the mean can become so small that its subtraction from image intensities during the calculation of global PCC becomes insignificant, at which point global PCC (with a theoretical range of 1 to -1) becomes indistinguishable from MOC (with a range of 1 to 0).

Multispeck and RGB Bead Combinations

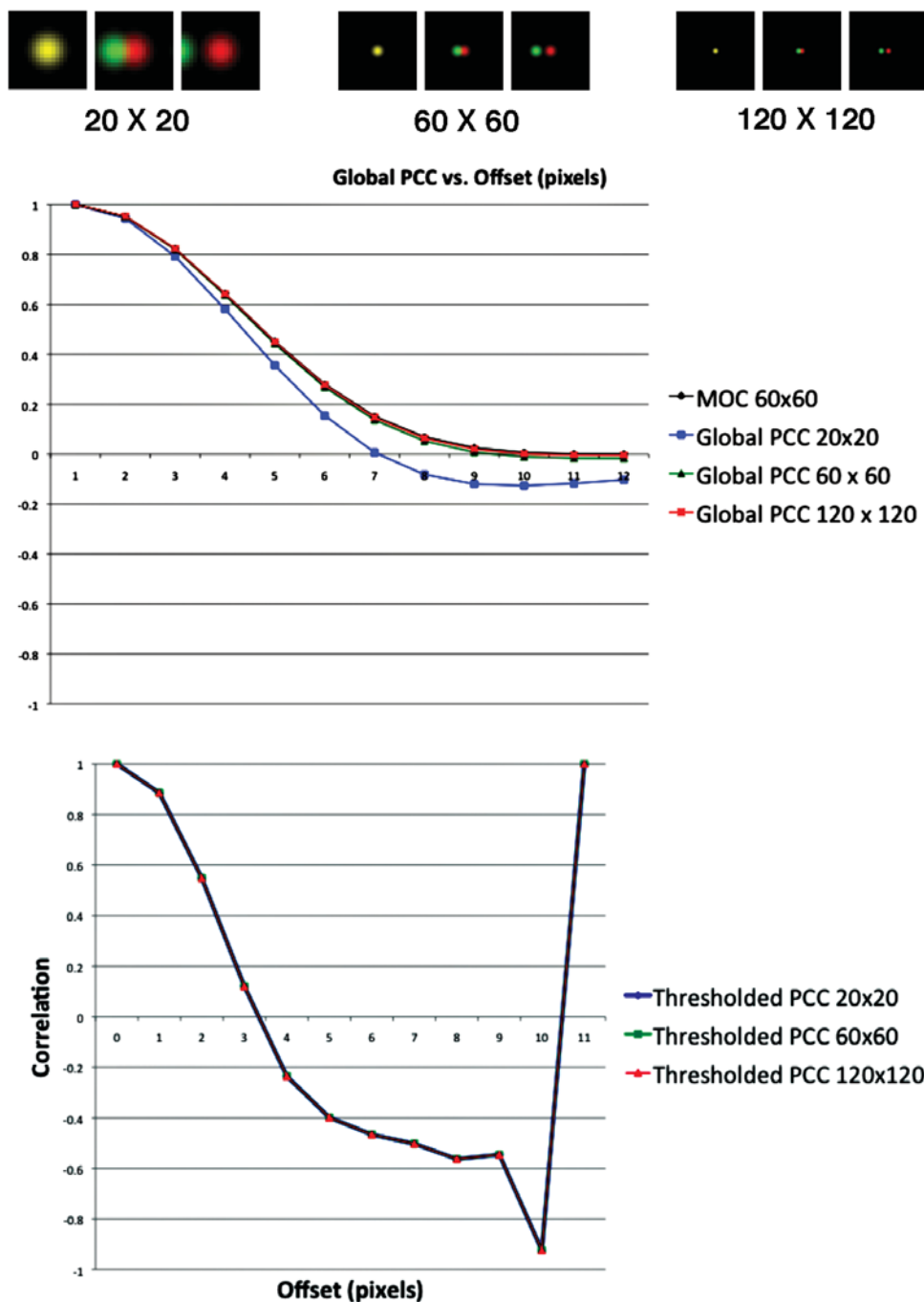
To test the effects of uncorrelated data on PCC, we used mixtures of multispeck beads that are triple-labeled with red, green, and blue dyes mixed with beads labeled with only red or green (Table 5). Multispeck beads contribute pixels of well-correlated intensities, whereas individually

labeled beads contribute red pixels that do not coincide with green pixels and vice versa. Global PCC and thresholded PCC were calculated for these fields. In all of the fields imaged, the thresholded PCC was >0.9 , whereas the global PCC values ranged from 0.098 to 0.923. Global PCC diminishes as the ratio of multispeck to singly labeled beads decreases. Thresholded PCC is unaffected and M_x and M_y clearly illustrate to which proportions of green and red intensities the thresholded PCC applies.

Live Cell Imaging

To check the accuracy of our method in biological specimens, we compared thresholded with global PCC on confocal micrographs of cells (Figs. 1, 2). Figure 1A–C shows the

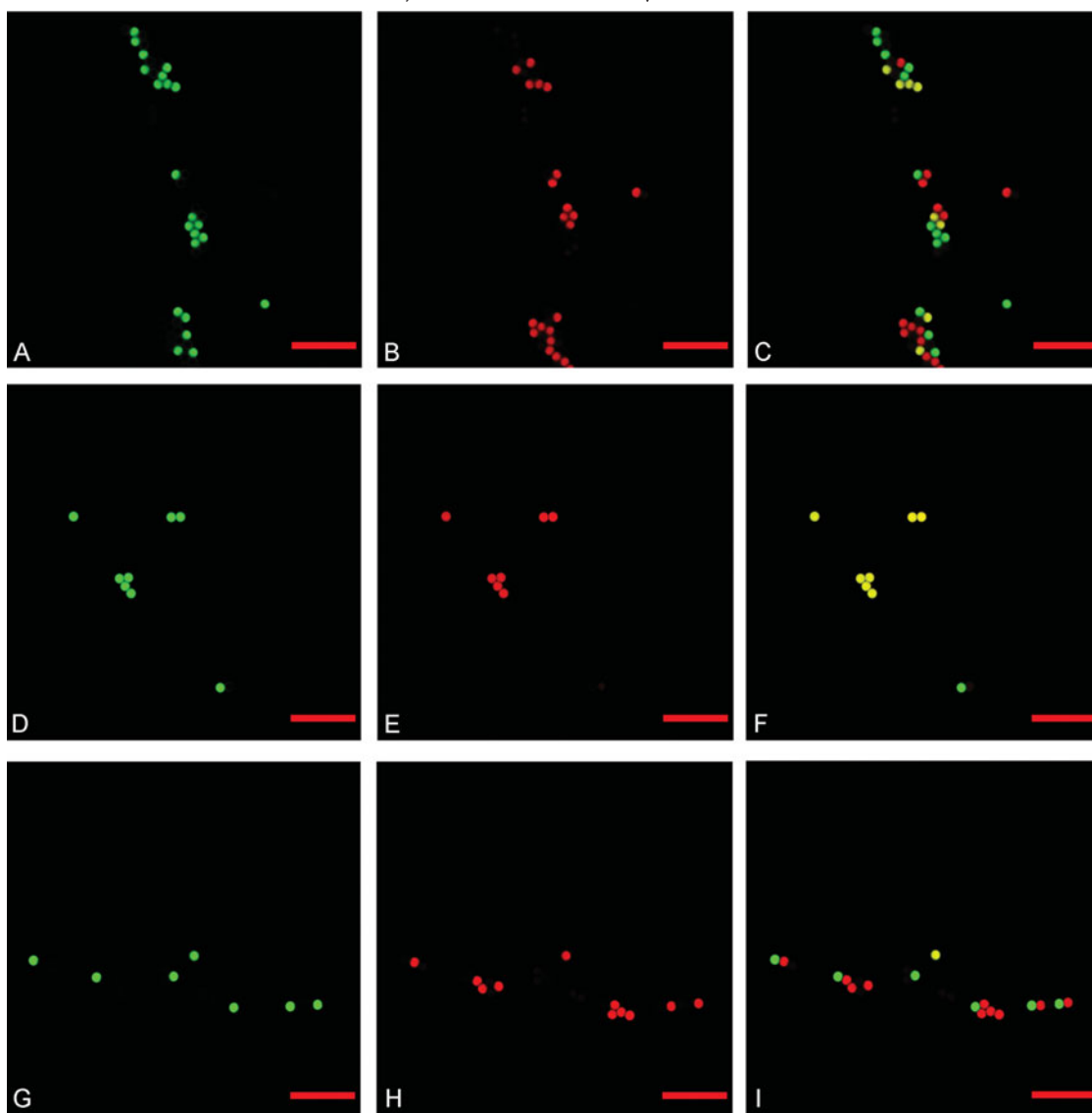
Table 4. The Effect of Background on Global PCC, MOC, and Thresholded PCC.*



*Identical Gaussian objects with a radius of 6 were drawn in identical positions in green and red images. Outside of the Gaussian objects all pixels had a gray level of 0. Adding additional pixels with a value of 0 does not affect MOC, which is identical for all image sizes. Global PCC varies as the additional background decreases the mean of the images. As the image gets larger, global PCC starts to converge with MOC. As the mean decreases, the ability of global PCC to generate negative values diminishes. Thresholded PCC does not vary as the size of images changes.

result of an experiment in which nuclei were labeled with Hoechst (pseudocolored blue) and mitochondria were labeled with MitoTracker orange (pseudocolored red). It is expected that these two structures would not colocalize and our thresholded PCC results confirm this. A global PCC of

0.122 was calculated, whereas the thresholded PCC was -0.231 . The global PCC indicates that the relationship between the two dyes is slightly positive whereas the thresholded PCC indicates exclusion, a more accurate description of the relationship between the molecules.

Table 5. The Effect of Noncolocalized Objects on Correlation Analysis.*

Field	Green Threshold	Red Threshold	Number of Multispeck	Red Only Beads	Green Only Beads	M_x	M_y	Global PCC	Thresholded PCC
A–C	29	29	8	13	14	0.37	0.36	0.354	0.924
D–F	1	15	7	0	1	1	0.86	0.923	0.954
G–I	28	26	1	10	5	0.09	0.14	0.098	0.939

*Scale bars measure 30 μm . Comparison of thresholded PCC and global PCC using mixtures of singly and multiply labeled beads show that with a global PCC analysis, mixtures containing progressively more and more uncorrelated objects dilute out the positively correlated ones. Conversely, thresholded PCC maintains its correct analysis despite the number of unlabeled objects in the image.

To demonstrate partial colocalization, cells were incubated with equal amounts of transferrin conjugated to green or red fluorochromes (Fig. 1D–F), both global (0.756) and thresholded (0.749) PCC ($M_x = 0.543$, $M_y = 0.964$) indicated the relationship expected.

Transient Transduction

We transduced HEK cells with a histone H2B-GFP construct and counterstained with Hoechst, both label the nucleosome, so a good correlation between Hoechst and

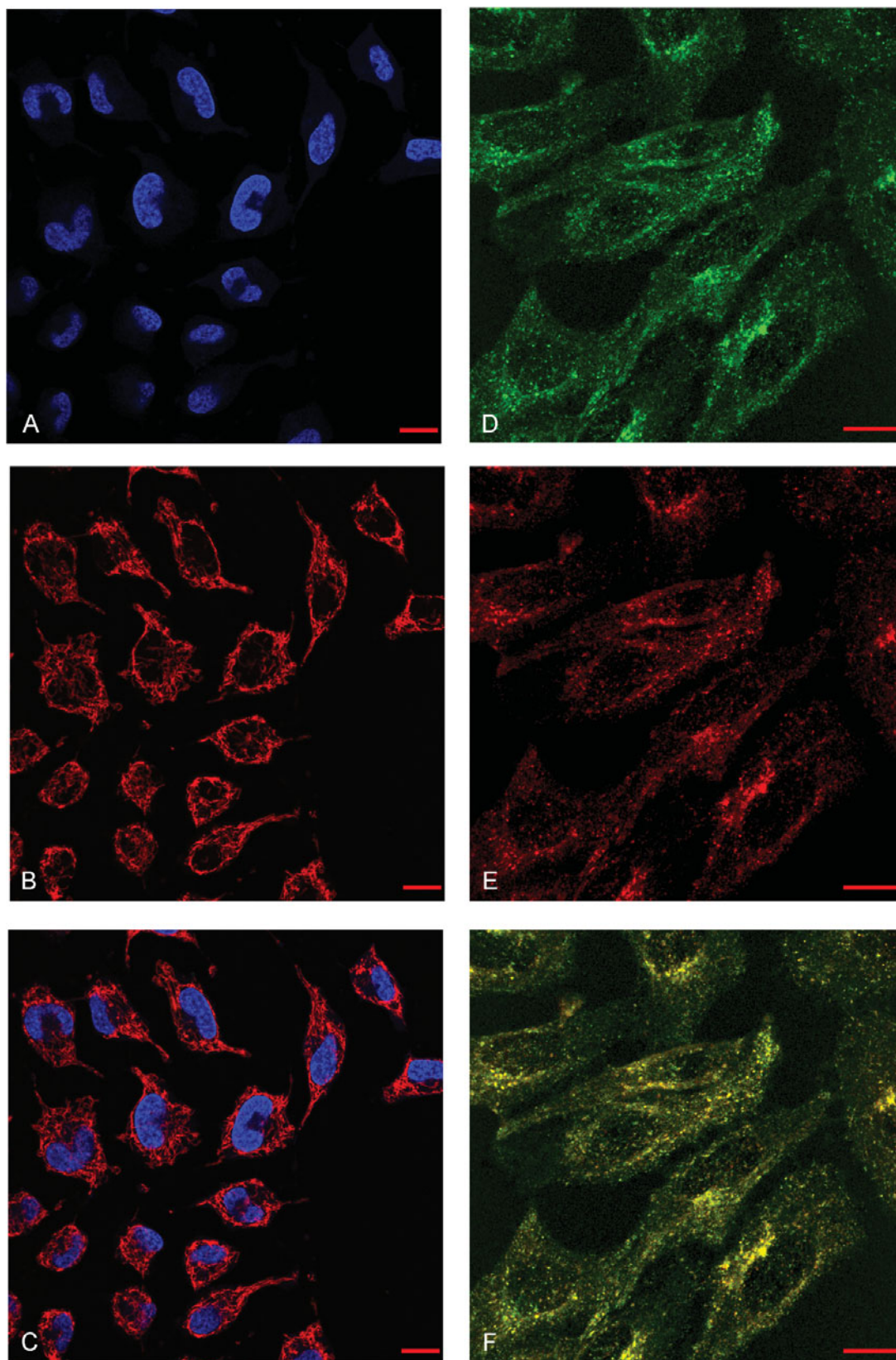


Figure 1. Colocalization and exclusion in biological samples. Cultured HEK cells (A–C) stained with MitoTracker and DAPI; the labeled structures do not colocalize. HeLa cells (D–F) stained simultaneously with transferrin AlexaFluor 488 and transferrin AlexaFluor 568; the endosomes demonstrate a high degree of colocalized staining. Scale bars measure 30 μm .

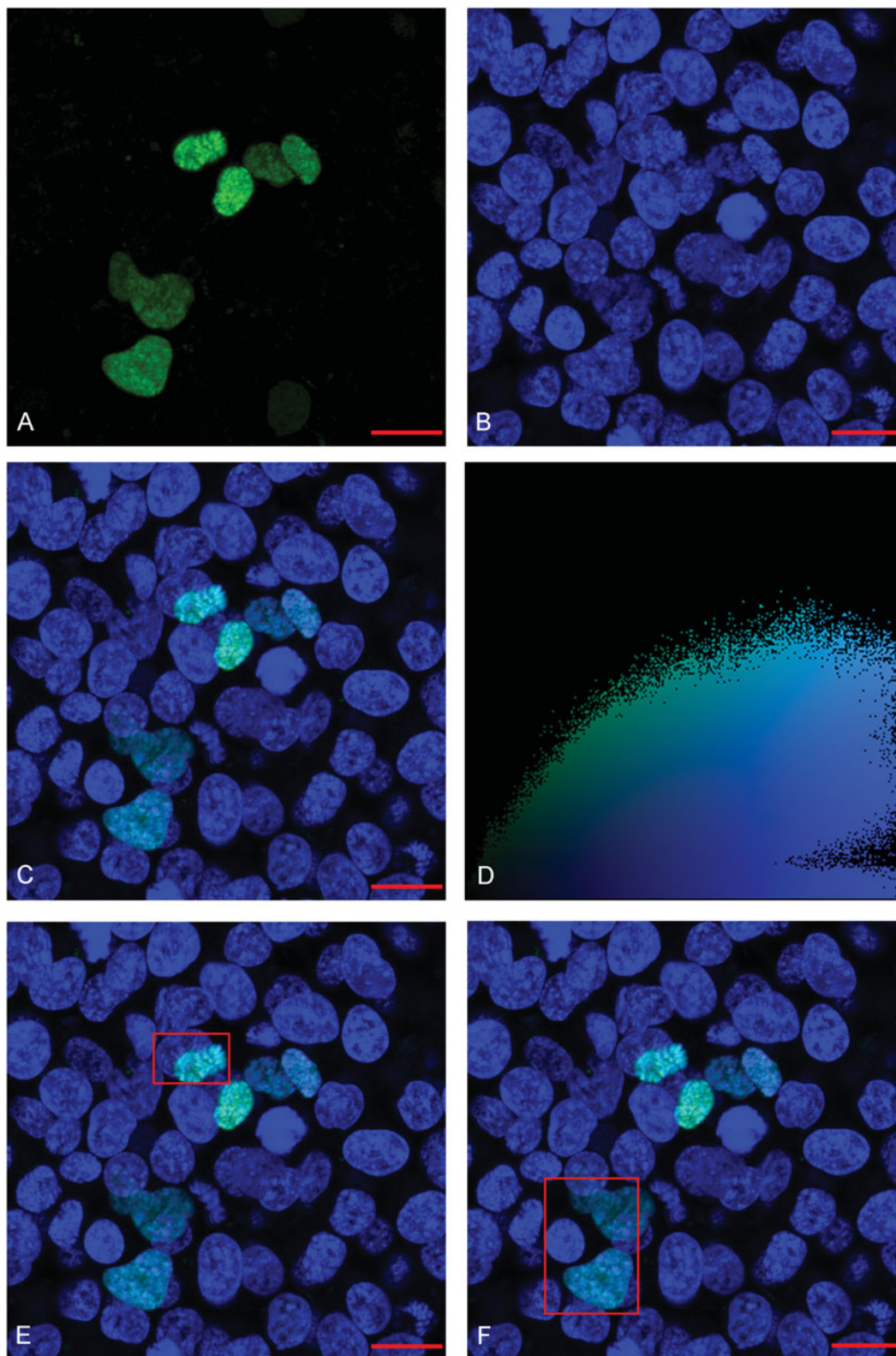


Figure 2. Colocalization in a subset of transduced cells. Cultured HEK cells transduced with Histone 2B-GFP and Hoechst. **A:** Single channel fluorescence micrograph of nuclei transduced with Histone 2B-GFP. As expected, expression levels of the transduced protein vary. **B:** Single channel fluorescent micrograph of the same field of cells shown in panel A with all nuclei stained with Hoechst. **C:** Overlay of the two channels shown in panels A and B. **D:** Scatter plot of pixel intensities in the two channels. **E, F:** Channel overlay as seen in panel C with two regions of interest defined. Scale bars measure 30 μm .

GFP is expected. Typically in transient transductions, only a subset of cells expresses the recombinant protein (Fig. 2). Thresholded PCC proved more effective at exposing this good correlation than global PCC. When the entire field of view shown in Figure 2C was analyzed, a thresholded PCC of 0.678 was generated, whereas the global PCC was 0.227. Scatter plots (Fig. 2D) contained more than one intensity distribution (exposed as more than one “cloud” of pixels in the scatter plot), indicating different expression levels of histone H2B. When only regions of interest containing a single intensity were analyzed, both thresholded and global PCC values increased (Fig. 2E: thresholded PCC = 0.821, global PCC = 0.57; Fig. 2F: thresholded PCC = 0.801, global PCC = 0.515). The higher thresholded values are a more accurate representation of the correlation.

DISCUSSION

The use of two fluorochromes to demonstrate colocalization or coincidence of two molecules is arguably the most common use of fluorescence microscopy. If used properly it can be a powerful tool to indicate the potential for molecular associations. Therefore, it is critically important that it is analyzed appropriately, quantitatively and with the most accurate mathematical tools. In both our theoretical and practical experiments, the use of a thresholded PCC of pixels above the mean in both image channels proved more accurate than current analytical algorithms.

Both theoretical and biological datasets have illustrated several shortcomings in the use of global PCC. The inclusion of background in the calculation of the mean generates unrealistically low means. This is of particular concern in biological datasets that frequently contain background. Low means result in a disproportionately large number of pixels above the mean, and a disproportionately low number of pixels below the mean. Positive differences from the mean become very large, and negative differences from the mean become very small. A feature of PCC is that intensities distant from the mean carry more weight than those close to the mean (van Steensel et al., 1996; Adler et al., 2008). This generates a strong bias toward positive values and makes generation of negative values unlikely. We argue that the inclusion of pixels that are not positively labeled by both fluorophores is an error that generates PCC values that are inaccurate. The calculation of a correlation coefficient that describes the relationship between a pair of labels (thresholded PCC) is much more likely to be of biological relevance than the calculation of the correlation coefficient that describes the relationship between two images (global PCC). This problem is starkly illustrated by pattern combination J (Table 1), which is perfectly inversely correlated in all intensities above the background yet generates a global PCC of 0.5. Likewise, in the biological data in Fig. 1A–C, global PCC reports a slight positive correlation even for a dataset

where no colocalization exists. Positive global PCC coefficients must be considered with suspicion if it is possible to generate them from inversely correlated datasets. It is essential that all positive PCC values be derived from situations of genuine positive correlation.

Our method of calculating thresholded PCC is not the same as the PCC of pixels above the thresholds set by applications that implement the approach of Costes et al. (2004) such as JACoP. Software implementing the Costes' method often display Pearson's correlation coefficient for two subsets of pixels, those with intensities above the thresholds and those with intensities beneath (this differs from the original implementation of Costes et al. in which only PCC for the subset of pixels fainter than either threshold was calculated). This is done to reassure users that the thresholds have been set in positions that separate positively correlated intensities from uncorrelated intensities, and to set thresholds that give objective measurements of Costes' variants of M_1 and M_2 . Given that PCC is used to determine these thresholds, the same thresholds cannot then be used to determine thresholded PCC objectively. Our approach requires thresholds that separate signal from background to be set, and then thresholded PCC, M_x , and M_y are generated using all pixels above both thresholds. While the values calculated for pixels above the automatically set thresholds are the same when calculated by JACoP and Velocity 5.3, the method of setting the threshold for the calculation of thresholded PCC must differ from the approach of Costes et al. in order to make an objective measurement of thresholded PCC. Additionally, the algorithm of Costes et al. (2004) is designed to set thresholds that separate positively correlated pixels from uncorrelated or negatively correlated pixels. Thresholded PCC, which requires thresholds that separate signal from background, is effective at describing uncorrelated and negatively correlated datasets.

It is likely that the Bolte and Cordelières (2006) observation that PCC rarely discriminates differences between partial colocalization and exclusion stems from the bias in global PCC, a bias that is not present in thresholded PCC. Restriction of PCC calculation to only those pixels above the threshold values for both images was suggested by Adler et al. (2008). We agree strongly with this approach, and our study has demonstrated that failure to do this reduces the mean to unrepresentatively low values, which generates a bias toward positive results. Microscopists should make a distinction between calculating a correlation coefficient between a pair of images (which contain background that is of no interest), and calculating a correlation coefficient between a pair of signals. We stress that failure to set the mean to only the subset of pixels for which PCC is being calculated represents a statistical error. It has been stated that global PCC is invariant to image background due to the influence of the background on the mean (Costes et al., 2004; Landmann & Marbet, 2004); however, the invariant influence of background is not the equivalent of the exclusion of background from the calculation of PCC. We have

demonstrated that to generate a mean capable of producing accurate PCC able to accurately describe the relationship between a pair of signals, the background must be ignored to avoid false positive results.

Some regard negative PCC values to be of no importance (Costes et al., 2004) or point out that they should be interpreted with caution (Zinchuk et al., 2007). We suggest that viewing negative PCC values as unimportant is flawed, and that this point of view may be a consequence of the implementation of global PCC, which makes the occurrence of negative values a rarity. Reevaluating the situation with thresholded PCC may reveal that inverse correlations are more common than previously thought. Zinchuk et al. (2007) stated that negative values for PCC should be interpreted with caution, and that when values less than zero are generated one should consider switching to Manders' overlap coefficient. We take the view that switching to another coefficient when an apparently inconvenient value is generated is not scientifically valid. A single coefficient should be used to measure all degrees of correlation. Negative thresholded PCC values are valid and informative.

We have demonstrated that thresholded PCC is capable of generating negative correlation coefficients for both synthetic and real biological datasets, and that in biological datasets the negative values generated were better descriptors of the relationship than the global PCC. This ability of our algorithm to generate negative correlations has not been achieved at the expense of detecting genuine positive correlations. This is demonstrated clearly in the examples of fluorescent micrographs we analyzed. Thresholded PCC is not affected by the presence of uncorrelated intensities within a pair of images, unlike global PCC in which good correlations are diluted out by areas of the image in which no relationship exists. This was demonstrated by our observations of multispeck beads, in which a correlation of >0.9 between the green and red elements of the multispeck beads was not "diluted out" by the presence of singly labeled green and red beads within the same field of view (Table 5). This was further demonstrated by analysis of HEK cells transiently transduced with GFP tagged histone H2B (Fig. 2). Thresholded PCC proved more effective at demonstrating a good correlation in this experiment. We did note that the thresholded PCC could be diminished by the presence of mixed intensity distributions within a single field of view (the product of including cells with different expression levels of the recombinant protein in a single analysis). Accurate correlations require restricting the analysis to areas in which labeling intensity is similar. For this reason we caution that mixed distributions should be avoided with thresholded PCC by restricting analysis to regions-of-interest. Global PCC is less prone to this effect because the unrealistically low mean ensures that few intensities are ever below the mean. This difference between the two methods illustrates the far higher sensitivity of thresholded PCC.

It has been suggested that in unbalanced situations, where the proportion, number, or intensity of biological

molecules colocalized with one another is not the same, only Costes' variants of M_1 and M_2 can fully describe the relationship (Costes et al., 2004). Thresholded PCC, by definition, only describes the relationship between images in pixel pairs that are above the threshold in both channels and so works perfectly in unbalanced situations. While we agree with Costes et al. that M_1 and M_2 are vital in describing such a relationship, only M_x and M_y combined with thresholded PCC can fully describe this relationship. When analyzing biological material during this study, thresholds were set to background intensity plus three standard deviations, so objective measurements of thresholded PCC, M_x , and M_y are generated. This is perfectly illustrated by Manders' test pattern I (Table 1), where a perfect correlation exists between 75% of green intensities and 8% of the red intensities. M_1 and M_2 alone lack a descriptor of the quality of the relationship and are only descriptors of the extent of the relationship. Global PCC tends to generate unrealistically low means. These low means tend to make global PCC "Manders' coefficient-like," e.g., difficult to generate negative values. This problem is avoided with thresholded PCC, making PCC, M_x , and M_y truly distinct and informative.

It has been suggested that midrange values for PCC are difficult to interpret (Costes et al., 2004; Bolte & Cordelières, 2006). We agree that this is the case for global PCC. Global PCC is insensitive to shifts deliberately introduced into combinations of Gaussian spot patterns. Midrange positive values of global PCC can be generated from cases of perfect anticorrelation (see Table 1), making midrange positive values of global PCC impossible to interpret. The greater sensitivity of thresholded PCC, the consequent avoidance of dilution of positive values, and the avoidance of positive values from cases of exclusion make thresholded PCC much more exact than global PCC.

CONCLUSIONS

This article is not intended to be an exhaustive survey of every image analysis software available to measure colocalization, nor a criticism of those who have put in significant efforts to quantitate fluorescent microscopic results; indeed, they should be commended. Software packages other than the ones we have tested may be available that correctly calculate a thresholded Pearson's correlation coefficient as we have described. However, we believe that this article is the first to clearly present and test thresholded PCC and to demonstrate why it is more accurate than global PCC implementations.

Bolte and Cordelières (2006) suggested that colocalization is afflicted with ambiguity and inconsistency. We absolutely concur and feel that it arises not only from a lack of quantitative analysis but, as we have demonstrated, the problems with currently used statistical measures such as global PCC. The use of a measure that cannot reliably

discriminate between positively and negatively correlated datasets, and in which positive correlations are subject to dilution by uncorrelated intensities, is yielding both false positive and false negative results. Thresholded PCC is not ambiguous and will serve to bring clarity to colocalization studies in fluorescent microscopy.

ACKNOWLEDGMENTS

We acknowledge that A. Barlow and A. MacLeod are employees of Perkin Elmer who sell the Volocity image analysis software; however, no commercial advantage is gained as this publication serves to put the method in the public domain. We thank Dr. A. Bredan for his excellent editorial assistance and Eef Parthoens, Dr. Geoffrey Lewis, and Profs. Kris Vleminckx and Jody Haigh for critical reading of the manuscript.

REFERENCES

- ADLER, J., PAGAKIS, S.N. & PARMRYD, I. (2008). Replicate based noise corrected correlation for accurate measurements of colocalization. *J Microsc* **230**, 121–133.
- ADLER, J. & PARMRYD, I. (2007). Letter to the Editor. *J Microsc* **227**, 83.
- AGNATI, L.F., FUXE, K., TORVINEN, M., GENEDANI, S., FRANCO, R., WATSON, S., NUSSDORFER, G.G., LEO, G. & GUIDOLIN, D. (2005). New methods to evaluate colocalization of fluorophores in immunocytochemical preparations as exemplified by a study on A_{2A} and D₂ receptors in Chinese hamster ovary cells. *J Histochem Cytochem* **53**, 941–953.
- BOLTE, S. & CORDELIÈRES, F.P. (2006). A guided tour into sub-cellular colocalization analysis in light microscopy. *J Microsc* **224**, 213–232.
- BOLTE, S. & CORDELIÈRES, F.P. (2007). Reply to Letter to the Editor. *J Microsc* **227**, 84–85.
- COSTES, S.V., DAELEMANS, D., CHO, E.H., DOBBIN, Z., PAVLAKIS, G. & LOCKETT, S. (2004). Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophys J* **86**, 3993–4003.
- FRENCH, A.P., MILLS, S., SWARUP, R., BENNETT, M.J. & PRIDMORE, T.P. (2008). Colocalization of fluorescent markers in confocal microscope images of plant cells. *Nat Protoc* **3**(4), 619–628.
- LANDMANN, L. & MARBET, P. (2004). Colocalization analysis yields superior results after image restoration. *Microsc Res Techniq* **64**, 103–112.
- LI, Q., LAU, A., MORRIS, T.J., GUO, L., FORDYCE, C.B. & STANLEY, E.F. (2004). A syntaxin 1, Gα_o, and N-type calcium channel complex at a presynaptic nerve terminal: Analysis by quantitative immunolocalization. *J Neurosci* **24**, 4070–4081.
- MANDERS, E.M.M., STAP, J., BRAKENHOFF, G.J., VAN DRIEL, R. & ATEN, J.A. (1992). Dynamics of three-dimensional replication patterns during the S-phase, analysed by double labelling of DNA and confocal microscopy. *J Cell Sci* **103**, 857–862.
- MANDERS, E.M.M., VERBEEK, F.J. & ATEN, J.A. (1993). Measurement of co-localisation of objects in dual-colour confocal images. *J Microsc* **169**, 375–382.
- NORTH, A.J. (2006). Seeing is believing? A beginners' guide to practical pitfalls in image acquisition. *J Cell Biol* **172**, 9–18.
- RASBAND, W.S. (1997–2010). ImageJ. Bethesda, MD: U.S. National Institutes of Health; available at <http://rsb.info.nih.gov/ij/>.
- SCRIVEN, D.R.L., LYNCH, R.M. & MOORE, E.D.W. (2008). Image acquisition for colocalization using optical microscopy. *Am J Physiol Cell Physiol* **294**, C1119–C1122.
- VAN STEENSEL, B., VAN BINNENDIJK, E.P., HORNSBY, C.D., VAN DER VOORT, H.T.M., KROZOWSKI, Z.S., DE KLOET, E.R. & VAN DRIEL, R. (1996). Partial colocalization of glucocorticoid and mineralocorticoid receptors in discrete compartments in nuclei of rat hippocampus neurons. *J Cell Sci* **109**, 787–792.
- ZINCHUK, V., ZINCHUK, O. & OKADA, T. (2007). Quantitative colocalization analysis of multicolor confocal immunofluorescence microscopy images: Pushing pixels to explore biological phenomena. *Acta Histochem Cytochem* **40**, 101–111.