Multi-Image Colocalization and Its Statistical Significance

Patrick A. Fletcher,[△] David R. L. Scriven,[△] Meredith N. Schulson, and Edwin D. W. Moore*
Department of Cellular and Physiological Sciences, University of British Columbia, Vancouver, British Columbia, Canada

ABSTRACT Accurately localizing molecules within the cell is one of main tasks of modern biology, and colocalization analysis is one of its principal and most often used tools. Despite this popularity, interpretation is often uncertain because colocalization between two or more images is rarely analyzed to determine whether the observed values could have occurred by chance. To address this, we have developed a robust methodology, based on Monte Carlo randomization, to measure the statistical significance of a colocalization. The method works with voxel-based, intensity-based, object-based, and nearest-neighbor metrics. We extend all of these to measure colocalization in images with three colors. We also introduce three new metrics; blob colocalization, where the blob consists of a local maximum surrounded by a three-dimensional group of voxels; cluster diameter, to measure the clustering of fluorophores in three or more images; and the intercluster distance to measure the distance between these clusters. The robustness of these metrics was tested by varying the image thresholds over a broad range, which produced no change in the statistical significance of the colocalizations. A comparison of blob colocalization with voxel and Manders colocalization metrics shows that the different measures produce consistent results with similar values for significance and nonsignificance. Using our methodology, we are able to determine not only whether the labeled molecules colocalize with a probability greater than chance, but also whether they are sequestrated into different compartments. The program, written in C++, is freely available as source, as well as in a Linux version.

INTRODUCTION

Colocalization using fluorescence imaging is one of the most commonly used tools for determining whether molecules are located in positions where they can interact with each other, and numerous methods for quantifying it have been published (1). A major flaw, which has been made worse with the advent of automated imaging systems, is that authors rarely attempt to determine whether the observed colocalization is significant, although such a method has been available for some time (2), albeit with limitations (3). Further, none of the existing methods has been extended to more than two fluorophores, a concern given the advent of multispectral microscopes and the magnitude and complexity of the distribution maps required by areas such as proteomics.

Colocalization between three or more labels offers the possibility of identifying different associative subgroups for each of the labeled molecules. Biologically, a triple colocalization may represent a different situation from multiple doubles, because such a grouping may operate in a way that is functionally different from the dual groupings. For this reason, given the three proteins A, B, and C, we distinguish among the three dual combinations A-B, A-C, and B-C, and the triple colocalization A-B-C, with this latter grouping being enumerated and treated separately. The presence or absence of triple (or higher) colocalization cannot be derived from doing multiple double colocalizations, because one cannot determine whether they are due to distinct or overlapping subgroups of the molecules in question. We have extended two existing methodologies to deal with triple colocalization: voxel colocalization (4) and the intensity-based Manders' *M* metric (5). A number of other measures, such as the Pearson's correlation coefficient (6) and its various derivatives (1), either cannot be easily extended, or produce measures that are very difficult to interpret; issues that we will discuss in detail later in this article.

A characteristic of all standard confocal and wide-field microscopes is that subresolution point sources, even after deconvolution, produce three-dimensional punctuate objects that consist of multiple voxels, spread along the *x*, *y*, and *z* axes. The size of these objects depends on the wavelength at which the source was measured as well as the parameters of the microscope and the recording system. Images can be regarded as the sum of many of these objects which overlap to form the observed intensity distribution. Measuring colocalization between two or more point sources emitting at different wavelengths using a voxel-by-voxel comparison, whatever the metric used, may give an inaccurate result (7). This is exacerbated by the fact that membrane-bound proteins are often distributed in clusters, and it is the cluster, not the individual molecule, that is the functional unit. We therefore treat three-dimensional groups of voxels centered on local maxima within the image as a single object of interest. Using this definition, we have extended a two-dimensional method of determining object overlap (8) to a three-dimensional object-based colocalization.

To determine the statistical significance of double, triple, and higher order colocalizations, we have developed a robust

Submitted January 7, 2010, and accepted for publication July 2, 2010.

[△]Patrick A. Fletcher and David R. L. Scriven contributed equally to this article.

*Correspondence: edmoore@interchange.ubc.ca

Editor: Petra Schwille.

0006-3495/10/09/1996/10 $2.00

doi: 10.1016/j.bpj.2010.07.006

methodology, which estimates the colocalization expected by chance directly from the data. This is achieved by creating simulated images in which the spatial locations of objects present in the data are randomized. The methodology can be applied to a wide variety of colocalization metrics including voxel-based, intensity-based Manders *M* coefficient, and object-colocalization types. In addition, our methodology allows us to provide extensive information about the spatial organization of the fluorophores—data which is important for tasks such as mathematical modeling. In particular, we extend the concept of nearest-neighbor distance (9) for groups of three or more fluorophores by introducing the cluster diameter, which gives a measure of how closely associated groups of molecules are; and the intercluster distance, which measures how far apart the clusters are.

METHODS

Immunofluorescent labeling

Atrial cells from adult Wistar rats were isolated using the methodology described in Scriven et al. (10). Cells were then labeled using three antibodies: monoclonal anti-ryanodine receptor (RyR; MA3-916; Thermo Fisher Scientific, Rockford IL); monoclonal anti-caveolin-3 (Cav3; 610420; BD Biosciences, Franklin Lakes, NJ); and the polyclonal anti-L-type Ca^{2+} channel ($Ca_v1.2$; gift of Dr. W. Catterall, Seattle, WA). The polyclonal anti-$Ca_v1.2$ was incubated overnight at 4°C and then incubated with the secondary, Alexa 488-anti rabbit (A11055; Invitrogen, Burlington, ON, Canada). The two monoclonals were then labeled using Zenon anti-mouse IgG_1 kits (Invitrogen, Carlsbad, CA) to form conjugates: anti-RyR with Alexa350 (Z25000) and anti-Cav3 with Alexa 594 (Z25007). The cells were incubated with the conjugates for 1 h, after which they were fixed in 4% paraformaldehyde and then washed in glycine followed by phosphate-buffered saline. Image stacks with voxel sizes of 95 nm in *X* and *Y* and 250 nm in *Z* were captured on a wide-field microscope, deconvolved (11), and a region of interest (ROI) was isolated.

Images from six different cells were collected, analyzed, and tested for significance; the results shown here are from one of these cells and are typical of what we found.

Defining the region of interest

Our program allows the ROI to be defined in one of two ways; by segmenting the image so the ROI covers only those regions in which the molecules of interest were known to be distributed, or by defining a mask, with ones defining the ROI and zeros elsewhere. This latter approach, which was used for the data shown here, is useful when the areas in the ROI are either not contiguous, or where the ROI has a complex shape; for example, when excluding the nuclear region from the three-dimensional image of two or more cytosolic proteins (12). It is important that, however it is defined, the ROI represent the biologically relevant compartment, so that the randomly generated images distribute the fluorophores into the same volume as the original image.

Fig. 1 *A* shows a full three-dimensional data set, collected as described above, showing all three fluorophores, with RyR labeled in blue, Cav3 in green, and $Ca_v1.2$ in red. Because the surface and interior of atrial cells have different distributions for the molecules studied (13,14), the mask chosen selected the upper and lower surfaces of the cells, and excluded the edges (Fig. 1 *B*) to avoid false colocalization due to reduced resolution in the *Z* axis (15). The data within the ROI defined by the mask is shown in Fig. 1 *C*. A detailed description of how we isolated the cell surface is given elsewhere (10).

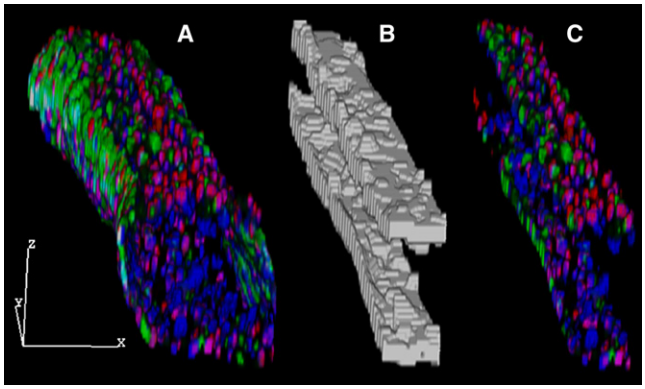

FIGURE 1 (*A*) Three-dimensional image of an atrial cell labeled with RyR in blue, Cav3 in green, and $Ca_v1.2$ in red. (*B*) Mask applied to the image so as to extract the top and bottom surfaces from image *A*. (*C*) Data within the ROI isolated by multiplying the image *A* by the mask *B*. Scalebars are 5 μm.

Deconstructing the image

To calculate the significance of the colocalizations we generated a series of random images, with the same properties as the original image. We then measured the distribution of the colocalizations within these random images (the colocalization that occurs by chance) and determined where the measured colocalizations fell on these distributions. To generate the random images we used our program to deconstruct the original images into their components parts. We first used a block-search algorithm to find local maxima, after which each lit (i.e., above threshold) voxel was assigned to a maximum using a hill-climbing algorithm. The result was that the images were split along the valleys between the maxima into separate parts or blobs, each containing a single intensity maximum. In those rare instances where the local maximum had adjacent voxels with identical values, their geographic center was taken as the maximum. We recorded the position of this maximum to provide a reference point for the metrics requiring spatial analysis (blob colocalization, cluster diameter, and intercluster distances). The isolated three-dimensional blobs were stored in a library for later use. A separate library was generated for each image, i.e., at each wavelength.

Measures of colocalization with more than two colors

There are numerous ways of measuring colocalization between fluorophores emitting at two wavelengths. Some of these, such as voxel-based colocalization (4) and the intensity-based Manders' *M* coefficients (5), can be easily extended to measure colocalization between three or more fluorophores (see the Supporting Material). In cases where we had three molecules R, G, and B (say), the triple colocalization R-G-B was regarded as being separate and not contributing to the doubles R-G, R-B, and G-B.

Blob colocalization

Another useful metric measuring the colocalization between objects has been proposed by Lachmanovich et al. (8), in which one object is colocalized with another when its reference point (in their case, the center of mass) is contained within the other object. As defined by them, the colocalization is not commutative, so that if A is colocalized with B, B may or may not be colocalized with A. A further implication of this is that the nearest-neighbor distance between clusters of colocalizing blobs will depend on which fluorophore we use as a basis for comparison; a situation which is exacerbated

when the fluorophores have differing densities. For these reasons, we imposed a stricter criterion for colocalization between blobs: Two or more objects are colocalized when their reference points are contained within the other object(s). This ensures that the colocalization is commutative and that there is only a single value when measuring the nearest-neighbor distance between colocalizing clusters.

This can be stated formally as follows: Colocalization between one blob and another occurs when the reference points, P, of each lies within the extent of the other. This extent, given by the coordinates of its n above-threshold voxels, is described by n xyz triplets,

$$T = \{xyz_1,\ xyz_2,\ \ldots,\ xyz_n\}.$$

Thus, given two blobs recorded at wavelengths R and G with reference points P_R and P_G and extents T_R and T_G, colocalization occurs when

$$P_R \in T_G \text{ and } P_G \in T_R.$$

If we have three wavelengths R, G, and B, a blob collected at the R wavelength is colocalized with the G and B blobs if all of these criteria are satisfied:

$$\{P_G, P_B\} \subseteq T_R;\ \{P_R, P_B\} \subseteq T_G;\ \{P_R, P_G\} \subseteq T_B.$$

Two measures can be derived for this form of colocalization. First, the percentage of blobs that colocalize is given by

$$BC_{RGB} = 100\,\frac{B_{RGB_{coloc}}}{B_R},$$

where $B_{RGBcoloc}$ is the number of blobs satisfying the colocalization criterion and B_R the total number of blobs at the R wavelength, with similar formulae for the triple colocalization at the G and B wavelengths:

$$BC_{GBR} = 100\,\frac{B_{GBR_{coloc}}}{B_G};\ BC_{BRG} = 100\,\frac{B_{BRG_{coloc}}}{B_B}.$$

Similar formulae can be derived for the dual colocalizations, e.g.:

$$BC_{RG} = 100\,\frac{B_{RG_{coloc}}}{B_R};\ BC_{GR} = 100\,\frac{B_{RG_{coloc}}}{B_G}.$$

The second measure calculates the number of voxels within the colocalizing blobs. Assuming that k blobs colocalize, triple (RGB) colocalization with respect to R is calculated by first summing the number of voxels in the colocalizing blobs,

$$n_{RBG_{coloc}} = \sum_{i=1}^{k} n_{R_i},$$

where n_{Ri} is the number of voxels in the i^{th} colocalizing blob at the R wavelength. The percentage colocalization with respect to the R wavelength is thus

$$BV_{RGB} = 100\,\frac{n_{RGB_{coloc}}}{n_R},$$

where n_R is the total number of nonzero voxels at the R wavelength, while for the G and B wavelengths:

$$BV_{GBR} = 100\,\frac{n_{GBR_{coloc}}}{n_G};\ BV_{BRG} = 100\,\frac{n_{BRG_{coloc}}}{n_B}.$$

The multiple dual colocalizations (R-G, G-B, and R-B) satisfy the following criteria:

$$\text{R} - \text{G}:\ P_G \subseteq T_R;\ P_R \subseteq T_G;\ \{P_R, P_G\} \notin T_B,$$

$$\text{G} - \text{B}:\ P_G \subseteq T_B;\ P_B \subseteq T_G;\ \{P_G, P_B\} \notin T_R,$$

$$\text{R} - \text{B}:\ P_B \subseteq T_R;\ P_R \subseteq T_B;\ \{P_R, P_B\} \notin T_G.$$

The formulae are similar in form to that of the triple colocalization, i.e.,

$$n_{RG_{coloc}} = \sum_{i=1}^{k} n_{R_i};\ BV_{RG} = 100\,\frac{n_{RG_{coloc}}}{n_R},$$

$$n_{GR_{coloc}} = \sum_{i=1}^{k} n_{G_i};\ BV_{GR} = 100\,\frac{n_{GR_{coloc}}}{n_G},$$

with similar formulae for the G-B and R-B colocalizations.

If a blob satisfies the colocalization criterion, the entire blob, not just the portions that overlap, is assumed to be colocalized (Fig. 2). The reference point was chosen to be the maximum of the blob so that the brightest point in each blob would overlap with the colocalizing blob. In our data, the difference between the intensity-weighted center of mass and the maximum of a blob was $\leq$50 nm, which is less than the optical resolution of our system.

Cluster diameter: an extension of the nearest-neighbor metric

Nearest-neighbor distances are difficult to interpret with the multiple interactions that are possible with three or more images. We introduce a new colocalization metric D_C, the cluster diameter, which is the diameter of a sphere that just contains the local maxima of the colocalizing blobs. D_C represents the size of the colocalizing cluster; when there are only two images, it is equivalent to the nearest-neighbor distance between the maxima, whereas with three or more it is the diameter of the smallest sphere that contains all the maxima. There are four measures: D_{CRGB}, D_{CRG}, D_{CRB}, and D_{CGB}, for three colors. As before, molecules that contributed to the triple colocalizations could not contribute to the doubles. The intercluster distance was defined as the distance between the centers of the colocalizing clusters. A two-dimensional example of both cluster diameter and intercluster distance is shown in Fig. 2 *D*.

Calculating D_C requires that we first identify the blobs from each image that are involved in the colocalization of interest. We replace each colocalizing blob in the images by a single three-dimensional reference point (either the intensity-weighted center of mass or the local maximum). Nearest-neighbor distances between these reference points were measured using Kd-Tree algorithms from the Computational Geometry Algorithms Library (CGAL; http://www.cgal.org). If there are two images, D_C is equivalent to the nearest-neighbor distance, whereas for three or more images we calculate the diameter of the minimum bounding circle or sphere using the algorithm min_sphere from the CGAL library. If needed, we can also calculate D_C for those blobs that do not colocalize with any other any blob in the other image(s).

Determining the significance of colocalization

To determine whether a colocalization was significant, we needed to estimate the distribution of colocalization values that could occur by chance. Blobs were selected from the library at random and their centers placed at a random coordinate in the ROI, so that their placement was completely spatially random. Voxels from the blob that fell outside the ROI were excluded. If any part of a blob overlaid another, the intensities of the overlaid voxels were summed.

Randomization was achieved by using the Mersenne Twister (16) and the selection of blobs was done with replacement, i.e., a particular blob could be chosen more than once or not at all. Populating the image with blobs was

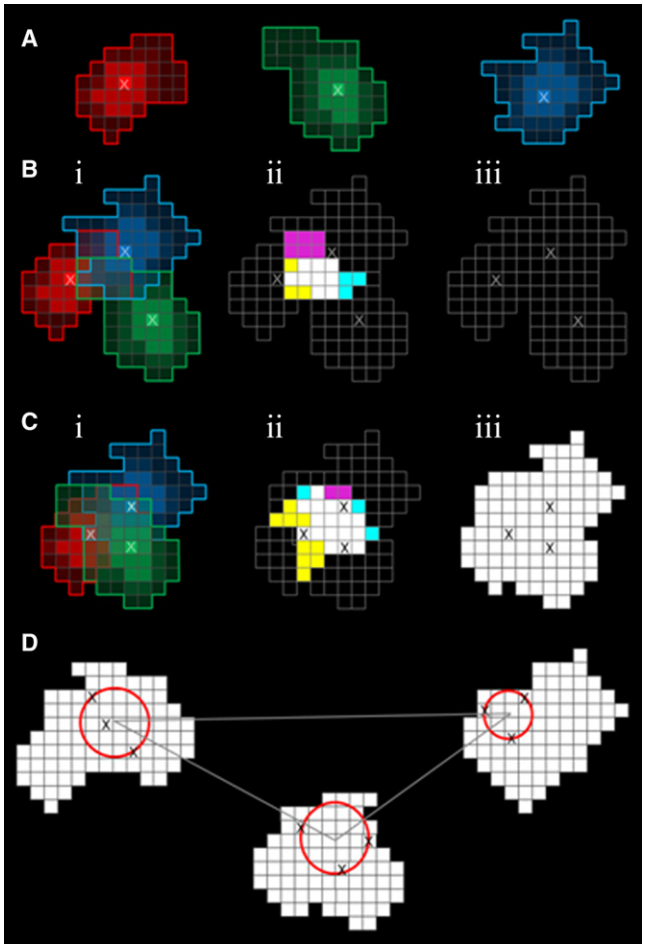

FIGURE 2 Voxel colocalization, blob colocalization, cluster diameter, and the intercluster distance. (*A*) Three blobs from each of the wavelengths, R, G, and B. The maximum intensity of each is shown with a cross. (*B*) (*i*) Blob overlap where none of maxima are within the other. (*ii*) Voxel colocalization. (RGB, *white*; RG, *yellow*; GB, *cyan*; and RB, *magenta*.) (*iii*) Blob colocalization is blank. (*C*) (*i*) Overlap of the three blobs where each of the maxima is within the other two blobs. (*ii*) Voxel colocalization with the same color coding as before. (*iii*) The blobs are triply colocalized and blob colocalization now covers all three blobs. (*D*) Cluster diameter (*red circles*) and intercluster distance (length of the *gray lines*). Intercluster distance is measured from the center of one cluster to the other.

stopped when the generated image had the same number of lit voxels as the original. The number of local maxima in the generated images varied, but clustered about the observed value with a normal distribution, so that properties of the random blob colocalization were representative of the observed image. Although it may seem more intuitive to use the sum of the intensities (integrated optical density; IOD) as the criterion for the Manders' coefficients, the intensity appears both in numerator and denominator of the coefficients so rescaling the image is not necessary. In addition, because the blobs could overlay each other, matching the IOD in the generated image to that of the original often resulted in too-few voxels being placed in the image. This situation got worse as the density of lit voxels increased, because the probability of overlaying a blob increased—resulting in the IOD increasing faster than the number of lit voxels. The randomization process was repeated for each blob library producing three random images, one for each fluorophore. Fig. 3 shows a single plane from the image stack, comparing the original data (Fig. 3, *A–D*) with random images generated for the same plane (Fig. 3, *E–H*). RyR is shown in blue (Fig. 3, *A* and *E*), Cav3 in green (Fig. 3, *B* and *F*), and $Ca_v1.2$ in red (Fig. 3, *C* and *G*). The overlap of all three images (Fig. 3, *D* and *H*) shows the colocalization of RyR-Cav3 in cyan, RyR-$Ca_v1.2$ in magenta, Cav3-$Ca_v1.2$ in yellow, and RyR-Cav3-$Ca_v1.2$ in white. The desired colocalization metrics were measured and stored and the procedure repeated until the stop criteria described below were satisfied.

In the images we have analyzed here, the mask thickness was comparable to the size of the blob in *Z* (Fig. 1 *B*). This created a problem, because random placement in *Z* would have a high probability of placing portions of the blob outside the mask, placing few voxels but increasing the number of local maxima beyond that of the original image. To circumvent this, we restricted the randomization of the blob position to the *X* and *Y* axes, keeping the *Z* position the same as that of the original blob. This allowed the generated images to retain their randomness while ensuring that the number of local maxima in the randomized image, on average, matched that of the original.

After generating a number of random image sets, each colocalization measure from the generated data was ranked, and the position of the observed data in these ranking was used to determine its probability of occurrence by chance,

$$p = 1.0 - \frac{R}{N_T},$$

where R is the rank of the observed data and N_T the number of iterations.

Stop criteria for the Monte Carlo simulations

A limitation of Monte Carlo simulations is that it is usually not possible to calculate beforehand how many iterations are necessary for an accurate answer. In our case we need to do only as many as would be required to determine whether the colocalization was significant or not. We did this by performing a bootstrap on the results from the randomized data at fixed intervals to determine the confidence limits of the ranking of the observed data (17). This interval has a default of 500 iterations, although it could be reduced if the images were large, or the identification of the blobs proved to be computationally intensive. The minimum number of iterations was set as five times the reciprocal of the significance level (0.05) to account for the loss of power in statistical tests that occurs with Monte Carlo simulations (9). When the confidence limits of all the parameters of a particular metric were unequivocally within either a region of significance or nonsignificance, the simulation for that metric was stopped. An upper limit for the number of iterations (default: 3000) was placed in case the confidence limit for one or more parameters remained straddling the two regions. It is possible, given the nature of such simulations, that there would be no definitive answer no matter how many iterations are done. In these cases, we would treat the values as nonsignificant.

RESULTS

Voxel colocalization

Fig. 4 shows voxel colocalizations from a single plane in the original data set (Fig. 4 *A*) and the same plane from one of the randomly generated images (Fig. 4 *B*); for simplicity, only the colocalized voxels are shown. The overlaps of RyR and $Ca_v1.2$ (*purple*) and Cav3-$Ca_v1.2$ (*yellow*) are clearly different in the two images, an observation confirmed by the findings of significance listed in Table 1. Blob colocalization for the same plane is shown for the original data set (Fig. 4 *C*) and the same random image is shown in Fig. 4 *B* (and Fig. 4 *D*) as a comparison.

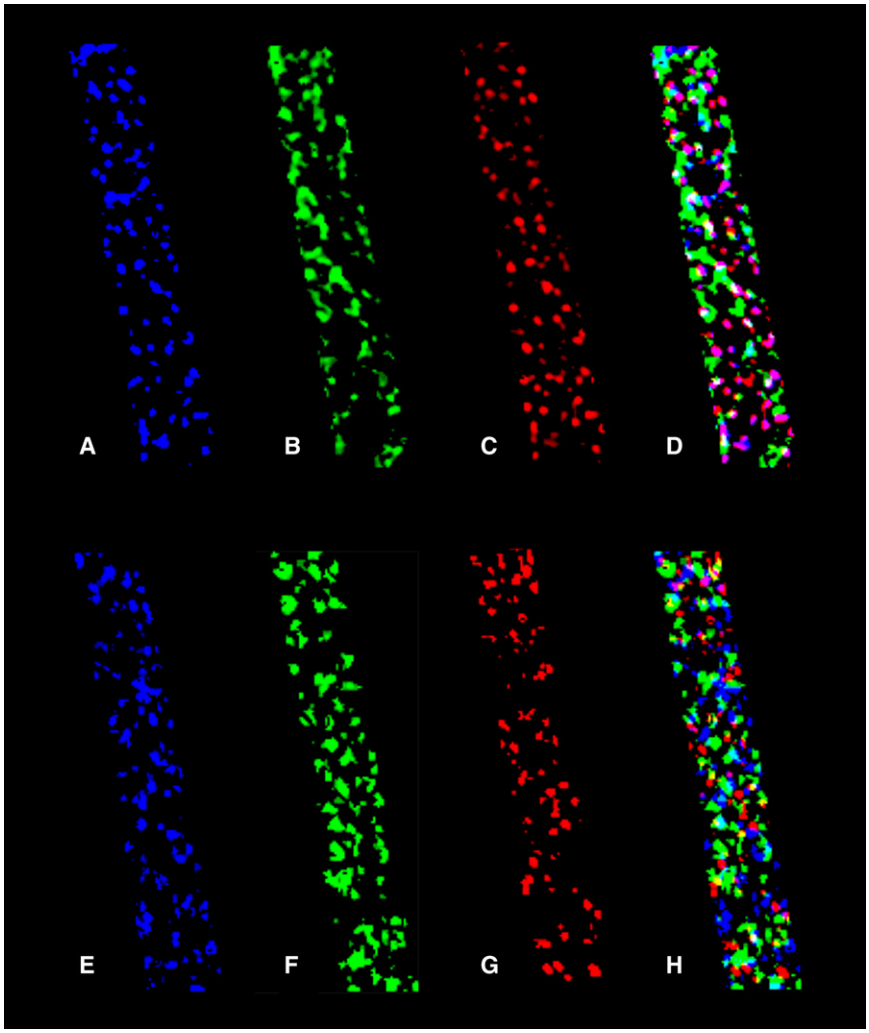

FIGURE 3 A single plane from the surface of an atrial myocyte stained with antibodies against (*A*) RyR; (*B*) Cav-3; and (*C*) $Ca_v1.2$. The three planes are superimposed in panel *D* with the colocalized voxels shown as follows: RyR-Cav3 (in *cyan*), RyR-$Ca_v1.2$ (in *magenta*), Cav3-$Ca_v1.2$ (in *yellow*), and RyR-Cav3-$Ca_v1.2$ (in *white*). (*E–H*) The same plane from a simulated image with the same ROI showing the equivalent random distributions of (*E*) RyR; (*B*) Cav3; and (*G*) $Ca_v1.2$. (*H*) Superposition of the three planes with the colocalized voxels labeled as before.

Manders' *M* colocalization

The intensity-based Manders' coefficient and voxel colocalization are closely related and produce identical results if the mean intensity of the colocalizing voxels is the same as the mean intensity of the voxels in that channel (see the Supporting Material). We repeated the above simulations using the Manders' coefficient as the measure of colocalization and found that the significance of the Manders' colocalizations were identical to that of the voxel (Table 1).

Blob colocalization

Blob colocalization has stricter criteria than either the voxel or Manders' metrics and produces different values, because it requires that the brightest spot, the maximum, overlaps the corresponding blob. In particular, there are fewer triple colocalizations which, depending on the arrangement of the blobs, can produce an increase in the number of doubles (RyR with $Ca_v1.2$ and vice versa; Table 1 and Fig. 4 *C*). In all cases, the significance is the same for both the number of colocalizing blobs as well as the number of voxels within the blobs. It is worth noting that although the percentage of colocalizing blobs is either similar to or less than the voxel and Manders' metrics, probably because of the stricter colocalization requirements, the values for the voxels within the blobs for RyR-$Ca_v1.2$ are higher. Fig. 5 shows the distribution of the colocalizations of voxel within blobs generated by the Monte Carlo simulations; for clarity each histogram is the result of 3000 simulations, which is in excess of what was needed to get a definitive result for this data set. Fig. 5 *A* shows the results for the triple colocalization among RyR, Cav3, and $Ca_v1.2$. The rank of the observed value is >3000, which indicates that the colocalization of 7.3% is highly significant ($p < 0.0003$). Fig. 5 *B* shows the distribution for Cav3 and $Ca_v1.2$. In this case, the rank of the observed colocalization (5.6%) is 4 ($p = 0.9987$), indicating that the observed value is highly significant, but less than would be expected by chance. As for the voxel colocalization, a comparison of panels *C* and *D* of Fig. 4 shows that the amount of Cav3-$Ca_v1.2$ overlap (*yellow*) is clearly less in the measured image than in the randomly generated one. Fig. 5 *C* shows the results for the colocalization

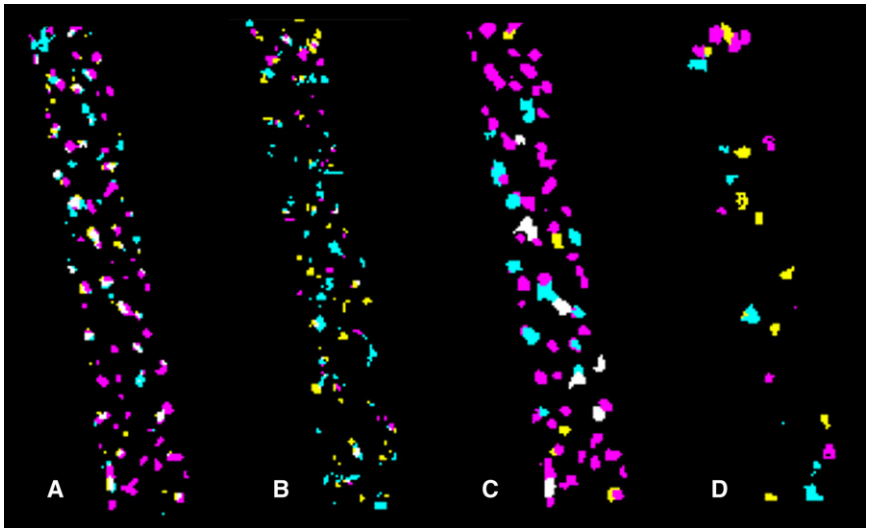

FIGURE 4 Voxel and blob colocalization. For simplicity, only the colocalized voxels are shown. (*A*) Voxel colocalization: a single plane from the measured image (Fig. 1 *D*) compared with (*B*) the same plane from a simulated image. (*C* and *D*) Blob colocalization: the same planes displayed using blob colocalization. RyR-Cav3 (in *cyan*), RyR-$Ca_v1.2$ (in *magenta*), Cav3-$Ca_v1.2$ (in *yellow*), and RyR-Cav3-$Ca_v1.2$ (in *white*).

between RyR and $Ca_v1.2$ which is highly significant; once again, reference to Fig. 4 supports this contention. Lastly, we simulated the interaction between RyR and Cav-3 (Fig. 5 *D*). In this case the rank of 855 ($p = 0.715$) shows that the colocalization is not significant. Graphs similar to Fig. 5 can be generated for both the voxel and Manders colocalizations.

Intercluster distances and the cluster diameters

Fig. 6 shows the cumulative distribution curves for the intercluster distances of the same colocalizations shown in Fig. 5. The solid line on each graph represents the observed distribution function while the dotted lines show the 95% confidence limits of the simulated data. Observed data that falls outside these limits is assumed not to be spatially random. The colocalized clusters of RyR-$Ca_v1.2$-Cav3 (Fig. 6 *A*) and RyR-Cav3 (Fig. 6 *D*) are randomly distributed. However, the RyR-$Ca_v1.2$ clusters (Fig. 6 *C*) are closer together and the $Ca_v1.2$-Cav3 clusters (Fig. 6 *B*) further apart than would be expected if their distribution in the cell were random (Table 2), suggesting that the placement of these clusters is determined by some underlying structure.

Examination of the cluster diameters (Fig. S1) show that the cluster diameter of the RyR-$Ca_v1.2$ colocalization (Fig. S1 *c*) is clearly not random with a significant left shift; the median cluster diameter of 95 nm is also found to be less than chance (Table 2). The cluster diameter for all of the other colocalizations (Fig. 6, *a*, *b*, and *d*) lie within the range that accords with spatial randomness except for a region at ~200 nm in the Cav3-$Ca_v1.2$ colocalization. Given the small number of colocalizations that occur between these two molecules, it is not clear whether this has any significance.

TABLE 1 Voxel, Manders, and blob colocalization values

Colocalization	RyR with Cav3	RyR with $Ca_v1.2$	RyR with Cav3 and $Ca_v1.2$
Voxel	17.8*	31.7†	12.0†
Manders' *M*	14.8*	36.8†	16.3†
Blobs	7.3	37.4†	4.4†
Voxels within blobs	10.5	54.8†	9.2†
	Cav3 with RyR	Cav3 with $Ca_v1.2$	Cav3 with RyR and $Ca_v1.2$
Voxel	16.1*	3.1†	10.8†
Manders' *M*	16.6*	3.2†	10.9†
Blobs	8.6	3.2*	5.2†
Voxels within blobs	14.3	5.7*	7.3†
	$Ca_v1.2$ with RyR	$Ca_v1.2$ with Cav3	$Ca_v1.2$ with RyR and Cav3
Voxel	45.7†	4.9*	17.3†
Manders' *M*	52.3†	3.5*	19.8†
Blobs	50.0†	3.5*	5.8†
Voxels within blobs	71.1†	3.2*	9.9†

Values are %; Manders' *M* values have been multiplied by 100 for ease of comparison with voxel and blob values.
*$p \geq 0.99$ (significantly less colocalization than chance).
†$p < 0.01$ (significantly more colocalization than chance).

DISCUSSION

The statistical power and utility of this method is shown by comparing the values of the colocalizations; if we consider the blob colocalization values (Fig. 5 and Table 1), it is clear that although the RyR-Cav3 colocalization is greater than that of the RyR-Cav3-$Ca_v1.2$ triple colocalization, and less than that of the RyR-$Ca_v1.2$ colocalization, the former is not significant whereas both of the latter two are highly significant. This shows that the value of a colocalization is no guide as to its significance. Another notable finding of these tests is that the colocalization between $Ca_v1.2$ and caveolin-3 is significantly less than one would expect by chance, which suggests that the two molecules are

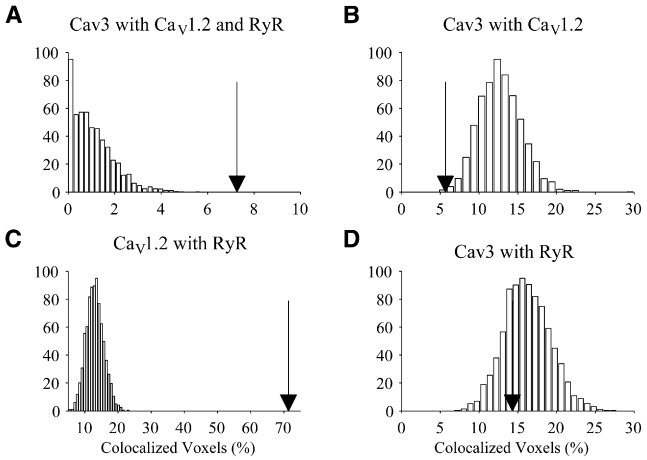

FIGURE 5 Distribution of results from simulations showing the colocalization of the voxel within blobs. (*Histograms*) Distributions obtained from the simulated data. (*Arrows*) Values of the observed data. Colocalizations shown are (*A*) RyR-Cav3-$Ca_v1.2$; (*B*) Cav3-$Ca_v1.2$; (*C*) RyR-$Ca_v1.2$; and (*D*) RyR-Cav3.

sequestered into separate compartments. This is, as far as we are aware, the first time that any colocalization method has been able to show this. It is a powerful result, because unlike colocalization, which shows that the molecules are close enough to interact (without proving that they do), this result shows that the molecules are unlikely to have any direct interaction at all. This finding is different from the anticolocalization described by Costes et al. (2)—which implies that the intensities of voxels with the same coordinates are negatively correlated. Calculating a Pearson's correlation coefficient for all the voxels in the ROI gives a positive value

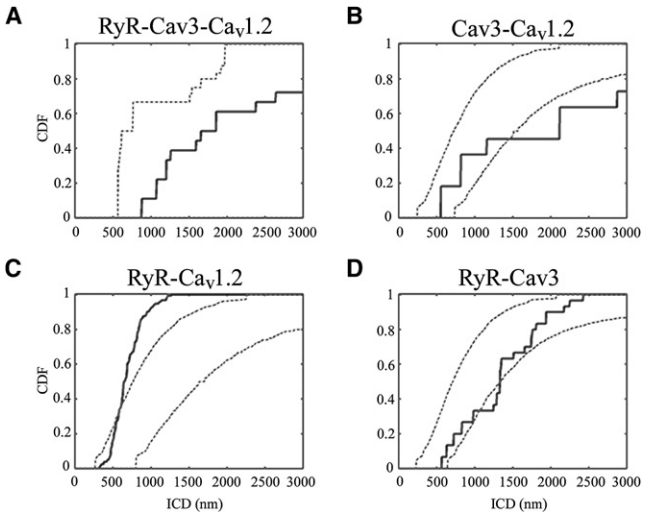

FIGURE 6 Cumulative distribution functions for the intercluster distances. (*Solid line*) Measured data. (*Dotted lines*) The 95% limits of the simulated distribution functions. Solid lines that fall outside these limits are not spatially random. Shown are (*A*) RyR-Cav3-$Ca_v1.2$; (*B*) Cav3-$Ca_v1.2$; (*C*) RyR-$Ca_v1.2$; and (*D*) RyR-Cav3.

TABLE 2 Cluster diameters and intercluster distances for colocalized blobs

Colocalization	Cluster diameter (nm)	ICD (nm)
RyR with Cav3	212	1334
RyR with $Ca_v1.2$	95*	646*
Cav3 with $Ca_v1.2$	134†	2118†
RyR with Cav3 and $Ca_v1.2$	259	1757

Values shown are the median of the measured distribution. ICD, intercluster distances.
*Significantly closer (−) than predicted by chance.
†Significantly farther apart (+) than predicted by chance.

of 0.22 for the $Ca_v1.2$-caveolin-3 colocalization, which is greater than that for RyR-Cav3 (0.17), which has neither significant colocalization nor separation.

We have introduced blob colocalization as a measure of how groups of associated voxels relate to one another. Because even a point source appears as a group of voxels in the observed image, individual voxels should not be considered in isolation. Blob colocalization describes the interaction between the blobs, which are voxel groups, each centered about a local maximum. The metric provides two useful measures; the number of blobs that overlap and the number of voxels within those blobs that are part of the colocalizing group. The closeness of the interaction is shown by the cluster diameter; the smaller D_C is, the tighter the association.

The values for the RyR-Cav3 colocalization differ depending on the metric: The blob colocalizations are nonsignificant while the voxel and Manders' *M* values are significantly less than chance. Fig. 4 shows that the voxel colocalization of RyR-Cav3 (*cyan*) produces multiple small colocalized spots (Fig. 4 *B*), far more than in the original image (Fig. 4 *A*), while blob images (Fig. 4, *C* and *D*) have far fewer, but comparable, numbers of colocalizations. This leads us to believe that the apparent separation of voxel and Manders colocalizations is due to random overlap at the periphery of the blobs and that the two molecules are not significantly colocalized or separate.

The usefulness of measuring group associations via blob colocalization can be shown by considering the relationship between $Ca_v1.2$ and RyR in the dyad of the cardiomyocyte. These molecules are known to be highly colocalized (18) and within 20 nm of each other (13,14). There is considerable evidence that the RyR exist in clusters and that the density of $Ca_v1.2$ is much lower than that of the RyR (19). The experiments described here show that the blob size of the $Ca_v1.2$ is smaller than that of the RyR, so even complete voxel colocalization may not show how these molecules interact, because Ca^{2+} released from the channels may activate RyR over a wider range than their physical overlap. This is highlighted by the RyR-$Ca_v1.2$ colocalizations listed in Table 1, where voxels within the blobs have considerably higher values than do either the voxel or Manders metrics.

This work extends that of Costes et al. (2) who published a method in which the image was split into blocks and then shuffled to determine the significance of the Pearson's correlation coefficient (6) and the Manders' M coefficients (5). The size of these blocks is determined from autocorrelation measurements, and is fixed for the whole image. Because many of our objects exceed this size in one dimension or the other, we were concerned that the characteristics of the random images, in particular the number and size of the blobs, would not be preserved. This was especially important for determining the significance of the two blob colocalization measurements as well as the cluster diameter and intercluster distance. Further, the Costes' approach excludes the possibility of using irregularly-shaped masks.

We have omitted Pearson's correlation coefficient (r_p) from our analysis. The reasons for this are multiple:

First, the coefficient measures both the correlation in voxel intensity as well as the linearity of the relationship between the copy numbers of the two molecules, so that deviations from linearity ($|r_p^2| < 1$) can be due either to a nonlinear relationship between the copy numbers of the two molecules or to a reduction in colocalization, and it is impossible to separate the two effects. A recent article (20) has supported this contention by finding that the nonlinear Spearman's rank coefficient may be a better measure of intensity correlation than the Pearson's coefficient.

Second, for this coefficient to be meaningful, the form of the relationship between copy number and intensity must be the same in both wavelengths measured, possible when both molecules are transfected with a fluorescent protein, but unlikely otherwise; indeed with indirect labeling, some secondary fluorophores have weak or highly nonlinear relationships between intensity and the number of bound fluorophores (21,22).

Third, the interpretation of r_p is difficult, because it is not always clear what intermediate values mean (8)—a task made even more difficult when considering correlations between three or more variables.

Effects of varying the threshold

Our methodology assumes that the image has been appropriately thresholded before it is processed. For the reasons given above, we did not use Costes method (2) for automatic thresholding because it depends on the Pearson's correlation coefficient. Further, the method seems to break down when the point density in the images are different (3). We prefer to use controls in which no primary antibody is present and which are treated identically to the experimental data, to set a level at which 99% of the control image is absent (10,18).

Thresholding has often been highlighted as a weakness in image processing, because different thresholds can produce different values of colocalization. However, as the data in Table 1 and Fig. 5 clearly demonstrate, the value of the colocalization is of limited usefulness if we do not determine its significance. While the value of a colocalization can change with threshold (Fig. 7), of far greater relevance for drawing biological conclusions is that changing the threshold produced no change in the statistical significance of the colocalizations. All of the molecular groupings examined, irrespective of whether the colocalization was significantly greater (RyR-Cav3-$Ca_v1.2$ and RyR-$Ca_v1.2$), significantly less (Cav3-$Ca_v1.2$), or no different (RyR-Cav3) than expected by chance, remained so despite wide changes in the threshold value.

The conclusions we draw from these statistical findings are:

1. RyR and $Ca_v1.2$ are tightly, though not completely, colocalized.
2. A subset of RyR and $Ca_v1.2$ are also colocalized with caveolin-3.
3. In the absence of RyR, $Ca_v1.2$ and caveolin-3 are located in separate membrane domains.

These inferences are not possible from the raw colocalization numbers themselves, regardless of the metric used or the threshold applied.

Thresholding is not unique to wide-field microscopy; confocal data is effectively thresholded by changing the black level setting before the data is even acquired, and these

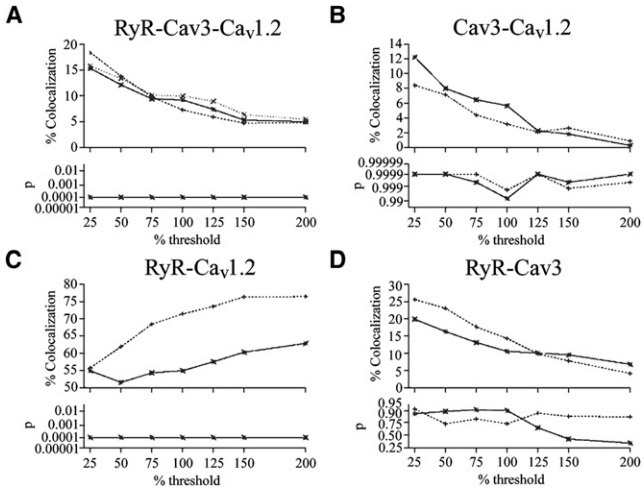

FIGURE 7 The effect of threshold on colocalization and its significance. The top graph in each segment plots colocalization of the voxels in blobs versus threshold, with 100% being the value used for the colocalizations listed in the tables, whereas the lower graph in each segment shows the probability that the colocalization could have occurred by chance; the Y scale follows the normal distribution. Many of the values in this graph fall on top of each other, so that there are fewer lines than in the corresponding colocalization graph. Shown are (*A*) RyR with Cav3 and $Ca_v1.2$ (*); Cav3 with RyR and $Ca_v1.2$ (+); and $Ca_v1.2$ with RyR and Cav3 (×). (*B*) Cav3 with $Ca_v1.2$ (*) and $Ca_v1.2$ with Cav3 (+). (*C*) RyR-$Ca_v1.2$; RyR with $Ca_v1.2$ (*) and $Ca_v1.2$ with RyR (+). (*D*) RyR with Cav3 (*) and Cav3 with RyR (+).

values are seldom reported. Although this can have profound effects on the absolute value of the colocalizations, it may have little impact on their statistical significance.

Limitations of the method

A fundamental assumption of our method is that the image can be broken down into isolated blobs and that each object can be isolated from the other, which depends on the data being punctate in nature. This is characteristic of high-frequency (i.e., rapidly changing) data such as the immunofluorescence of membrane proteins, but not of low frequency data such as calcium concentration measurements, or of molecules that are widely dispersed throughout the cytosol. Using colocalization for low frequency data of this type is inappropriate, because any other image that overlays this region would have a colocalization nearing 100%, which would be the same value that would occur by chance (15). Failure can also occur when using punctate data where the intensity of the image exceeds the maximum range of the sensor. This results in large objects with wide central plateaus which need to be reacquired before they can be used.

Our methodology also depends on an estimation of the space into which the molecules can be placed. For example, if the image encompasses the cell interior, the molecules may be limited to either the cytoplasm or the nucleus, in which case a digital mask needs to be placed over the image, reducing the actual volume into which the blobs can be placed. Overestimating the volume of distribution will have the effect of decreasing the voxel density and thus the probability of colocalization, shifting the histograms in Fig. 4 to the left and increasing the probability of a false-positive (Type 1) error. The completely spatially random envelopes are shifted rightward, although the movement is relatively small unless there is a large ($\geq 50\%$) overestimate in volume. It may be that obtaining an accurate measure of the size of the compartments in which the molecules of interest are distributed is not feasible. In this case, one could set upper and lower bounds for the volume and run the simulations twice, once for the upper bound and once for the lower. If the significance is consistent between simulations, then this value is robust and can be used, but if it changes, then no conclusion can be reached.

Although we have used three fluorophores in this example, the method works equally well with two, because the methodology is independent of the modality of measurement and could be extended to other metrics if one should so choose. All of the metrics discussed can be extended to four or more fluorophores, although it rapidly becomes difficult to track all of the combinations; 11 for four fluorophores (one quadruple, four triples, and six doubles) and 26 for five (one quintuple, five quadruples, 10 triples, and 10 doubles).

The code, written in C++, is freely downloadable from our website at http://crg.ubc.ca/moore/downloads. While we believe that triple colocalization represents a different biological situation from the double, and should be counted separately, the program contains an option that allows triples to be counted as doubles, if the user so desires.

SUPPORTING MATERIAL

One figure and 17 equations are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(10)00850-7.

P.A.F. and D.R.L.S. wrote the code and performed the simulations; D.R.L.S., P.A.F., and E.D.W.M. developed the theoretical basis for the work; D.R.L.S. and P.A.F. analyzed the data; M.N.S. performed the immunofluorescence experiments; and D.R.L.S., E.D.W.M., and P.A.F. wrote the article.

We thank Prof. Lawrence Lifshitz of Biomedical Imaging Group at the University of Massachusetts Medical Center for helpful discussions and Dr. W. Catterall for the gift of the CNC antibody (National Institutes of Health grant No. R01 HL085372). This work was supported by grants from the Canadian Institutes of Health Research (grant No. MOP12875), the Heart and Stroke Foundation of British Columbia and the Yukon, and the Natural Scientific and Engineering Research Council of Canada to E.D.W.M.

REFERENCES

1. Bolte, S., and F. P. Cordelières. 2006. A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* 224:213–232.
2. Costes, S. V., D. Daelemans, …, S. Lockett. 2004. Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophys. J.* 86:3993–4003.
3. Comeau, J. W., S. Costantino, and P. W. Wiseman. 2006. A guide to accurate fluorescence microscopy colocalization measurements. *Biophys. J.* 91:4611–4622.
4. Lynch, R. M., K. E. Fogarty, and F. S. Fay. 1991. Modulation of hexokinase association with mitochondria analyzed with quantitative three-dimensional confocal microscopy. *J. Cell Biol.* 112:385–395.
5. Manders, E. M. M., F. J. Verbeek, and J. A. Aten. 1993. Measurement of colocalization of objects in dual-color confocal images. *J. Microsc. (Oxford).* 169:375–382.
6. Manders, E. M., J. Stap, …, J. A. Aten. 1992. Dynamics of three-dimensional replication patterns during the S-phase, analyzed by double labeling of DNA and confocal microscopy. *J. Cell Sci.* 103:857–862.
7. Wu, Y., M. Eghbali, …, E. Stefani. 2010. Quantitative determination of spatial protein-protein correlations in fluorescence confocal microscopy. *Biophys. J.* 98:493–504.
8. Lachmanovich, E., D. E. Shvartsman, …, A. M. Weiss. 2003. Co-localization analysis of complex formation among membrane proteins by computerized fluorescence microscopy: application to immunofluorescence co-patching studies. *J. Microsc.* 212:122–131.
9. Diggle, P. J. 2003. Statistical Analysis of Spatial Point Patterns, 2nd Ed. Arnold, London, UK.
10. Scriven, D. R., A. Klimek, …, E. D. Moore. 2005. Caveolin-3 is adjacent to a group of extradyadic ryanodine receptors. *Biophys. J.* 89: 1893–1901.
11. Carrington, W. A., R. M. Lynch, …, F. S. Fay. 1995. Superresolution three-dimensional images of fluorescence in cells with minimal light exposure. *Science.* 268:1483–1487.
12. Dan, P., J. C. Cheung, …, E. D. Moore. 2003. Epitope-dependent localization of estrogen receptor-α, but not -β, in en face arterial endothelium. *Am. J. Physiol. Heart Circ. Physiol.* 284:H1295–H1306.

13. Carl, S. L., K. Felix, …, D. G. Ferguson. 1995. Immunolocalization of sarcolemmal dihydropyridine receptor and sarcoplasmic reticular triadin and ryanodine receptor in rabbit ventricle and atrium. *J. Cell Biol.* 129:673–682.

14. Sun, X. H., F. Protasi, …, C. Franzini-Armstrong. 1995. Molecular architecture of membranes involved in excitation-contraction coupling of cardiac muscle. *J. Cell Biol.* 129:659–671.

15. Scriven, D. R., R. M. Lynch, and E. D. Moore. 2008. Image acquisition for colocalization using optical microscopy. *Am. J. Physiol. Cell Physiol.* 294:C1119–C1122.

16. Matsumoto, M., and T. Nishimura. 1998. Mersenne Twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simul.* 8:3–30.

17. DiCiccio, T. J., and B. Efron. 1996. Bootstrap confidence intervals. *Stat. Sci.* 11:189–212.

18. Scriven, D. R. L., P. Dan, and E. D. W. Moore. 2000. Distribution of proteins implicated in excitation-contraction coupling in rat ventricular myocytes. *Biophys. J.* 79:2682–2691.

19. Bers, D. M., and V. M. Stiffel. 1993. Ratio of ryanodine to dihydropyridine receptors in cardiac and skeletal muscle and implications for E-C coupling. *Am. J. Physiol.* 264:C1587–C1593.

20. Adler, J., S. N. Pagakis, and I. Parmryd. 2008. Replicate-based noise corrected correlation for accurate measurements of colocalization. *J. Microsc.* 230:121–133.

21. Berlier, J. E., A. Rothe, …, R. P. Haugland. 2003. Quantitative comparison of long-wavelength Alexa Fluor dyes to Cy dyes: fluorescence of the dyes and their bioconjugates. *J. Histochem. Cytochem.* 51: 1699–1712.

22. Panchuk-Voloshina, N., R. P. Haugland, …, R. P. Haugland. 1999. Alexa dyes, a series of new fluorescent dyes that yield exceptionally bright, photostable conjugates. *J. Histochem. Cytochem.* 47:1179–1188.