Interdiscip Sci Comput Life Sci (2012) 4: 27–37
DOI: 10.1007/s12539-012-0117-x

# Critical Evaluation of Quantitative Colocalization Analysis in Confocal Fluorescence Microscopy

Yong WU[1*], Vadim ZINCHUK[2], Olga GROSSENBACHER-ZINCHUK[3], Enrico STEFANI[1]
[1](Department of Anesthesiology, David Geffen School of Medicine, University of California, Los Angeles, California 90095-7115, USA)
[2](Department of Anatomy and Cell Biology, Faculty of Medicine, Kochi University, Kochi 780-8520, Japan)
[3](Institute of Anatomy, University of Fribourg, Fribourg 1700, Switzerland)

Received 6 July 2011 / Revised 21 August 2011 / Accepted 26 December 2011

**Abstract:** Spatial colocalization of fluorescently labeled proteins can reveal valuable information about protein-protein interactions. Compared to qualitative visual interpretation of dual color images, quantitative colocalization analysis (QCA) provides more objective evaluations to the degree of colocalization. However, the finite resolution power of microscopes and the spatial patterns of intracellular structures may compromise the reliability of many classical QCA methods. In this paper, we discuss the strength and weakness of some mostly used QCA methods. By studying their applications on computer-simulated images and biological images, we show that classical pixel intensity based QCA methods are often vulnerable to coincidental overlapping among resolution elements (resel) distributions and thus not suitable to images with high molecular density or with low resolution. Also, many QCA methods can mistakenly regard long range correlation as colocalization due to protein localization in intracellular structures. The newly developed protein-protein index (PPI) approach is able to reduce the influence from resel overlapping and spatial intracellular pattern compared to previous methods, significantly improving the reliability of QCA.
**Key words:** fluorescence microscopy, confocal microscopy, image analysis, colocalization.

## 1 Introduction

Spatial colocalization analysis is a standard tool widely used in fluorescence microscopy to study protein-protein interactions. By labeling two proteins with two different of fluorophores emitting at different wavelengths, dual color (we will name the two colors "red" and "green" throughout this article, despite their real color in reality) images can be generated, and highly colocalized red and green signals might indicate close spatial correlations between the proteins, which in turn signifies their probable interactions. Visual interpretation of the overlaid dual color image pair or their intensity scatter plot is an intuitive way to assess colocalization and can be very helpful (Dutartre *et al.*, 1996; Fox *et al.*, 1991). However, it is only qualitative and often vulnerable to human bias. Quantitative colocalization analysis (QCA) calculates various numerical indices to more objectively evaluate the degree of colocalization (Zinchuk and Zinchuk, 2008). Intensity based indices are calculated through the pixel intensity values; examples of those approach are Pearson's correlation coefficient (Demandolx and Davoust, 1997; Manders *et al.*, 1992), overlap coefficients (Manders *et al.*, 1993), Manders' colocalization coefficients (Manders *et al.*, 1993) and intensity correlation quotient (Li *et al.*, 2004). More advanced intensity based approaches such as the automatic threshold method (Costes *et al.*, 2004), image cross-correlation spectroscopy (ICCS) (Comeau *et al.*, 2006; Comeau *et al.*, 2008), and protein proximity index (PPI) (Wu *et al.*, 2010) were also developed. There are also object based methods, in which images are processed with image segmentation or edge detection algorithms to identify objects, and these objects rather than individual pixels are used to assess the degree of colocalization (Bolte and Cordelieres, 2006; Boutte *et al.*, 2006; Lachmanovich *et al.*, 2003).

It is obvious that colocalization analysis, and even the concept of "colocalization", is closely related to the resolution of the microscopes used to create the images. Optical microscopes have finite resolution power: A point light source produces a spatial distribution light intensity at the image plane according to the point spread function (PSF), and two point sources lo-

*Corresponding author.
E-mail: wuyong@ucla.edu

cated within the size of PSF are difficult or impossible to resolve. Conventional confocal fluorescence microscopes are subject to the diffraction limit and cannot exceed the resolution at ∼200 nanometers. Recently, super-resolution techniques such as photo-activated localization microscopy (PALM) (Betzig *et al.*, 2006), structured illumination microscopy (Gustafsson, 2000), and stimulated emission depletion (STED) microscopy (Hell, 2003), may be able to improve the resolution down to 20-50nm. Since the physical size of a proteins and their interaction distance may be on the order of several nanometers, for most existing optical microscopes, two associated proteins will produce almost coincident resolution elements (resel) on the image and are thus considered as "colocalized". On the other hand, two overlapping but not coincident resels usually should not be considered as colocalized, because the actual distance between the two proteins is too far away for them to be actually associated. On digital microscopic images, the full width half maximum (FWHM) of PSF should be sampled with at least 4 pixels according to the Nyquist sampling theorem. Therefore, a resel consists of quite a number of pixels on images, which results problems for many intensity based QCA methods: These classical QCA methods are based on counting colocalized red and green pixels, and always tend to overestimate the degree of colocalization by regarding overlapping pixels in resels as colocalization. Obviously this overestimation becomes more problematic when the degree of colocalization is low or when there is excessive resel overlapping, which happens when the resolution is too low or the density of molecules is very high.

Another significant matter to consider is that proteins in most cases are not homogenously distributed in the cell, but reside in intracellular structures and form a pattern due to their organized spatial distribution. Thus, both types of proteins under colocalization analysis usually appear in the same intracellular structures, meaning that their spatial distribution is correlated, but they are not necessarily colocalized. It is worth noting that this spatial correlation due to intracellular structures often has longer spatial range than colocalization, in other words, the intracellular structures have longer length scales than PSF. If this was not the case, the structures could not be resolved and could not be differentiated from colocalized proteins in principle. Many existing QCA approaches mistakenly count this long range spatial correlation due to intracellular structures as colocalization and yield overestimation. This issue has been discussed in a recent publication (Wu *et al.*, 2010), in which the PPI approach is presented. In this approach, spatial correlations are decomposed according to their range, and colocalization is identified only with correlation whose range is comparable to the PSF size, which effectively reduces the influence from long range correlation due to intracellular pattern.

In this article, we review a number of most widely used QCA methods in comparison to the PPI approach. We apply these approaches to computer-simulated images and biological images and compare their results. In computer-simulated images the degree of colocalization can be precisely controlled, which makes it easy for us to evaluate the reliability of the QCA methods under review. We also use two sets of biological images of the same type of cells to study the dynamic change in colocalization. We discuss the strengths and weaknesses of the QCA methods and show that the PPI approach is more reliable than other methods in minimizing the overestimation caused by both overlapping resels and spatial patterns of intercellular structures. Further improvement on the reliability of QCA will mostly rely on the advancement of microscopic technology with enhanced resolution, since higher resolution power will reduce resel overlapping and better distinguish colocalization from intracellular patterns.

## 2 Methods

The central issue of this paper is to test the accuracy and reliability of various QCA methods. QCA methods discussed in this paper include:

- Pearson's correlation coefficient $R_r$
- Manders' overlap coefficient $R$
- Overlap coefficients $k_1$ and $k_2$
- Colocalization coefficients $m_1$ and $m_2$
- Li's intensity correlation quotient (ICQ)
- Costes' automatic threshold method
- Costes' randomization test
- Image cross-correlation spectroscopy (ICCS) with scrambled images
- Protein-protein proximity index (PPI)
- Object based methods (distance between centers of mass)

Manders' overlap coefficient $R$, overlap coefficients $k_1$ and $k_2$, and colocalization coefficients $m_1$ and $m_2$ were calculated using CoLocalizer Pro 2.5 (CoLocalization Research Software) (Zinchuk and Grossenbacher-Zinchuk, 2009). Li's intensity correlation quotient, Pearson's correlation coefficient $R_r$ , Costes' randomization test and object based methods were calculated or implemented with an ImageJ plugin JACOP (Bolte and Cordelieres, 2006) and some of them were double-checked by our own program. We also programmed our own software to implement Costes' randomization test, ICCS and to calculate PPI.

To test the accuracy of these methods, we used computer simulated images, in which the degree of colocalization is known precisely, so that after applying the

QCA methods, the results can be compared with the known value. Computer simulated images were generated using MatLab 2009a (The MathWorks, Natick, MA). An image with size of m × n is represented by an m × n matrix where each matrix element stores the intensity value at the corresponding pixel. In confocal fluorescence microscopy, each fluorophore will generate a resel (according to PSF) on the resultant image. Therefore, two steps were followed to generate a computer-simulated image: First, the central pixels of resels, each corresponding to a fluorophore, were determined according to a certain spatial distribution, and at each central pixel the intensity value was set to one; secondly, the matrix was then convoluted with a Gaussian PSF. The spatial distribution of resel centers can be chosen in two ways. The first one is that they are distributed purely randomly, or, in other words, the fluorophores are distributed according to a uniform distribution, as seen in Fig. 1 (A-C). This is of course only an ideal case and can rarely be a good approximation to real biological images. The second choice is that the spatial distribution is taken from a biological image, to which the computer-simulated image will visually resemble. In this case, the central pixels are distributed according to the pixel distribution of a biological image, and the intracellular structures are thus kept in the computer simulated images, whose influences to QCA may then be assessed. It is worth noting that images in two channels cannot share the same biological image as the spatial intracellular pattern provider, otherwise the images will always have a certain degree of colocalization. We have used an image pair of a cardiomyocyte from a mouse where ryanodine receptor (RyR) and estrogen receptor $\alpha$ ($ER_\alpha$) were independently labeled (See Fig. 1 (D) and Fig. 1 (G), respectively. Images were cropped and processed with median filters to

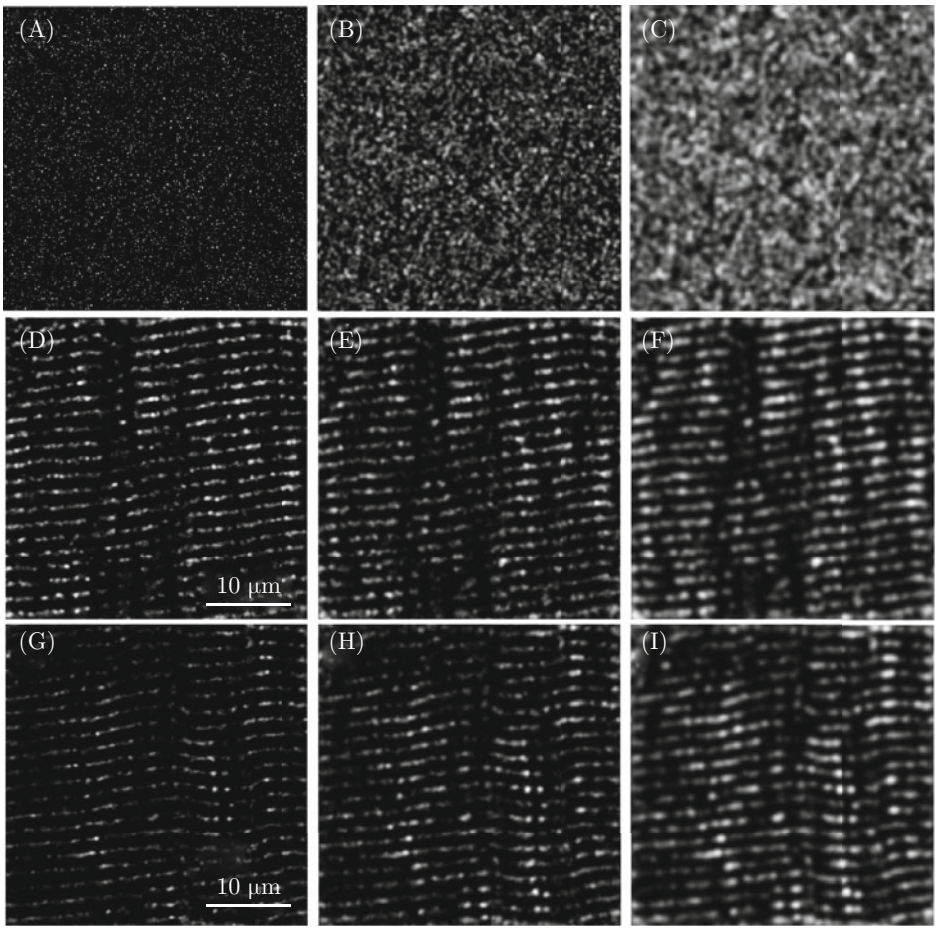


Fig. 1 Images generated by computer simulations to test QCA methods. (A) Computer-simulated image in which molecules are randomly distributed according to a uniform distribution. Each molecule is represented by a single pixel. (B and C) Convoluting the image in (A) with a Gaussian PSF with standard deviation $\sigma = 3$ and 5 pixels, respectively. (D) Cropped image of a cardiomyocyte from mouse where ryanodine receptor (RyR) were labeled. (E and F) Computer-simulated images based on the image in (D), convoluted with a Gaussian PSF with $\sigma = 3$ and 5 pixels, respectively. (G) Cropped image of a cardiomyocyte from mouse where estrogen receptor $\alpha$ (ER$\alpha$) was labeled. (H and I) Computer-simulated images based on the image in (G), convoluted with a Gaussian PSF with $\sigma = 3$ and 5 pixels, respectively.

reduce shot noises and background) as the blueprint for computer simulations. It has been shown that these two proteins do not associate (Wu *et al.*, 2010; Ropero *et al.*, 2006), and thus we could control the degree of colocalization in computer-simulated images by adding artificial colocalized molecules.

Real microscopic images are more complex than our model images due to complications such as dark current and shot noise of detectors, out of focus light, nonspecific labeling of antibodies, and bleed-through between the two channels. However, in order to concentrate our study on influence from resel overlapping and spatial intracellular pattern to QCA, these factors were not included in our computer simulations.

We also test the QCA methods on biological images, which were taken for isolated heart myocytes from mice. Proteins were labeled with specific monoclonal (anti-mouse) and polyclonal antibodies (anti-rabbit). Isolated cells were fixed with 4% paraformaldehyde in 0.1 M $Na_2HPO_4$ and 23 mM $NaHPO_4$ (pH 7.4) at room temperature for 20 min, and permeabilized with 0.2 % Triton-X 100. Nonspecific binding was blocked for 30 minutes at room temperature using 10% goat or donkey serum in PBS pH 7.4 containing 0.2% Triton X-100 to permeabilize the cells. Double labeling was achieved incubating the cells with polyclonal and monoclonal antibodies (5-10 μg/ml) incubated overnight at 4 ℃). Cells were washed, incubated (1 h, room temperature) with secondary Abs Alexa 488 anti-rabbit IgG and Alexa 594 anti-mouse IgG1 (2 μg/ml), washed again and mounted with Prolong (Molecular Probes). Stack of images were typically acquired by optically sectioning cells every 0.1 μm at 0.058 μm per pixel with a confocal microscope using 60X, 1.4 NA, oil immersion objective. PMT sensitivity was adjusted to avoid saturation.

## 3 Results and discussions

We use computer simulated images to test the reliability of QCA methods in the following aspects:

First, the result of a reliable QCA method should not count randomly overlapping resels as colocalization. One way to test this is to see whether the results vary significantly when the resolution of the microscope or the image pixel size changes. To this end we use a set of computer simulated images in which the locations of molecules are kept same, but the standard deviation $\sigma$ (or width for simplicity) of the Gaussian PSF varies from 0 to 5 pixels, as illustrated in Fig. 1 (PSF $\sigma = 0$ means that the convolution with PSF is not performed; $\sigma = 3$ is close to proper sampling according to Nyquist's theorem; $\sigma = 5$ is oversampling). Note that less resolution power at a fixed pixel size or a smaller pixel size at a given resolution (oversampling) will both result a greater PSF width measured by the number of pixels. A reliable QCA method should yield same or similar results for image pairs with different PSF widths.

Secondly, inauthentic "colocalization" introduced by the spatial pattern of intracellular structures should be excluded. To test this, we apply QCA methods on computer simulated images whose spatial distribution was taken from a mouse cardiomyocyte, as shown in Fig. 1 (D) and (G). Both species are distributed along the T-tubules, which might be incorrectly regarded as colocalization by some QCA methods.

Lastly, QCA methods should be able to differentiate significant changes in the degree of colocalization. This will be tested by using two sets of simulated or biological images that have the different known degrees of colocalization.

The testing results are summarized in Table 1, Table 2 and Fig. 2. Note that some QCA methods use only one numerical value to quantify colocalization while others use two values for two channels respectively. When comparison is made among them and a method yields two values, their geometric mean is used for comparison. Figure 2 shows the calculated values of each QCA method with growing PSF width, also their comparison to the real colocalization values predefined by the computer simulation.

Below we will discuss the reliability of each QCA methods in details.

### 3.1 Pearson's correlation coefficient

Pearson's coefficient is defined as

$$R_r = \frac{\sum_i (R_i - R_{avg})(G_i - G_{avg})}{\sqrt{\sum_i (R_i - R_{avg})^2 \sum_i (G_i - G_{avg})^2}}, \quad (1)$$

where $R_i$ and $G_i$ are intensity at the $i$th pixel in the red and green channel, respectively, and the subscript "*avg*" stands for the average intensity value. Pearson's coefficient has been criticized for its nonlinear dependence on intensity values (Costes *et al.*, 2004), and for that it only has a single numerical value which is usually not sufficient to describe colocalization in two channels (Comeau *et al.*, 2006). However, Pearson's correlation coefficient has the advantage that its value is not sensitive to varying PSF width, as shown in Table 1 and Table 2. On the other hand, the intracellular structures do greatly affect its value and produce significant overestimation to the degree of colocalization when the degree of colocalization is low. In Table 2 and Fig. 2, we see that at nonzero PSF width, $R_r \approx 0.35$ even if the image pairs do not have colocalization. This positive value of $R_r$ merely signifies correlation due to the closely related spatial pattern of T-tubules of the cardiomyocyte in both red and green channel.

**Table 1 Comparison of QCA methods applied on computer simulated images where molecules were randomly distributed**

| | PSF width (pixels) | | | | | |
|---|---|---|---|---|---|---|
| | PSF $\sigma = 0$ | | PSF $\sigma = 3$ | | PSF $\sigma = 5$ | |
| Real colocalization | 0%, 0% | 33%,25% | 0%, 0% | 33%,25% | 0%, 0% | 33%, 25% |
| Peasrson's correlation coefficient $R_r$ | 0.00 | 0.29 | 0.00 | 0.29 | 0.00 | 0.29 |
| Manders' overlap coefficient $R$ | 0.02 | 0.16 | 0.30 | 0.39 | 0.36 | 0.43 |
| Overlap coefficient $k_1$ and $k_2$ | 0.04, 0.01 | 0.39,0.06 | 1.1, 0.08 | 1.6, 0.10 | 1.9, 0.07 | 2.3, 0.08 |
| Manders' coefficient $m_1$ and $m_2$ | 0.02, 0.03 | 0.20,0.26 | 0.97, 0.24 | 0.99,0.31 | 1.0, 0.24 | 1.0, 0.31 |
| ICQ | 0.41 | 0.41 | 0.01 | 0.10 | 0.00 | 0.10 |
| Automatic threshold | 4.9%, 3.2% | 38%,29% | 70%, 44% | 87%,75% | 61%, 46% | 88%, 82% |
| Costes' randomization test | Failed | Passed | Failed | Passed | Failed | Passed |
| Objected based method | N/A | N/A | 14%, 15% | 17%,19% | 4.3%, 5.2% | 6.1%,6.0% |
| ICCS (image scrambled) | 0.00%, 0.00% | 33%,25% | 0.16%,0.11% | 33%, 25% | 0.21%, 0.14% | 33%, 26% |
| PPI | 0%, 0% | 35%,25% | 0%, 0% | 33%,25% | 0%, 0% | 34%,26% |

**Table 2 Comparison of QCA methods applied on computer simulated images based on biological images with a spatial intracellular pattern**

| | PSF width (pixels) | | | | | |
|---|---|---|---|---|---|---|
| | PSF $\sigma = 0$ | | PSF $\sigma = 3$ | | PSF $\sigma = 5$ | |
| Real colocalization | 0%, 0% | 15%, 33% | 0%, 0% | 15%, 33% | 0%, 0% | 15%, 33% |
| Peasrson's correlation coefficient $R_r$ | 0.04 | 0.25 | 0.35 | 0.51 | 0.34 | 0.50 |
| Manders' overlap coefficient $R$ | 0.07 | 0.28 | 0.35 | 0.69 | 0.45 | 0.77 |
| Overlap coefficient $k_1$ and $k_2$ | 0.08, 0.06 | 0.18, 0.44 | 0.22, 0.54 | 0.35, 1.37 | 0.29, 0.72 | 0.38, 1.59 |
| Manders' coefficient $m_1$ and $m_2$ | 0.09, 0.07 | 0.18, 0.42 | 0.40, 0.81 | 0.69, 0.87 | 0.50, 0.93 | 0.81, 0.94 |
| ICQ | 0.42 | 0.44 | 0.21 | 0.25 | 0.18 | 0.22 |
| Automatic threshold | Failed | 0%, 0% | 87%, 70% | 82%, 96% | 84%, 71% | 88%, 96% |
| Costes' randomization test | Passed | Passed | Passed | Passed | Passed | Passed |
| Objected based method | N/A | N/A | 13%, 12% | 22%, 20% | 9.5%, 8.8% | 19%, 17% |
| ICCS (image scrambled) | 4.9%, 3.6% | 17%, 36% | 34%, 36% | 42%, 61% | 33%, 35% | 42%, 59% |
| PPI | 0%, 0% | 15%, 33% | 1.1%, 1.4% | 21%, 38% | 1.5%, 1.6% | 20%, 34% |

### 3.2 Manders' overlap coefficient

Manders' overlap coefficient is defined as

$$R = \frac{\sum_i R_i G_i}{\sqrt{\sum_i R_i^2 \sum_i G_i^2}} \tag{2}$$

Compared to Pearson's coefficient, the elimination of the average intensity values avoids negative values of $R$, but it also makes $R$ vulnerable to resel overlapping, which always exaggerates its value. Other researchers also found this coefficient is less reliable compared to Pearson's correlation coefficient (Adler and Parmryd, 2010). In Table 1, Table 2 and Fig. 2, one can see that Mander's overlap coefficient grows along with increasing PSF width. Also, it is difficult for Manders' overlap coefficient to differentiate the degree of colocalization: $R$ is insensitive to the change in real colocalization values. In addition, this coefficient can also produce inauthentic colocalization due to spatial pattern of intercellular structures. Therefore, one should be very careful when applying this coefficient to images with high molecular density or with poor resolution that causes much resel overlapping, and also to images with a clear spatial intracellular pattern.

### 3.3 Overlap coefficients $k_1$ and $k_2$

In order to characterize colocalization in both channels, it is necessary to use two coefficients. Overlap coefficients $k_1$ and $k_2$ are defined as

$$k_1 = \frac{\sum_i R_i G_i}{\sum_i R_i^2}$$
$$k_2 = \frac{\sum_i R_i G_i}{\sum_i G_i^2} \tag{3}$$

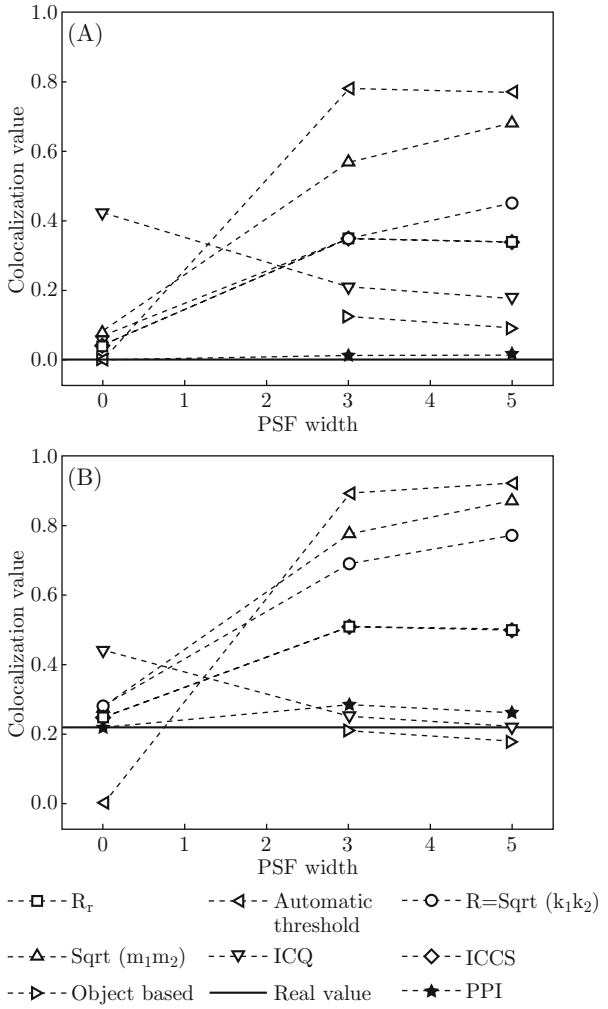


Fig. 2 Comparing the results of QCA methods applied on computer simulated images illustrated in Fig. 1, as a function of PSF width. For methods generate two values, their geometric mean is used in the plot. (A) Preset colocalization value is zero for both images. (B) Preset colocalization value is 15% for one image and 33% for the other, and the geometric mean is 22%. In both (A) and (B), PPI yields result very close to real value despite the growing PSF width.

$k_1$ and $k_2$ have the same drawbacks that Manders' Overlap Coefficient $R$ has due to the similarity between their definitions. Another obvious disadvantage is that $k_1(k_2)$ scales proportionally to the intensity value of the green (red) channel, and thus the coefficients depend on the relative strength of the two channels. These coefficients are almost only applicable to image pairs with similar intensity level in two channels. In our computer simulations we set different absolute intensity values in two channels deliberately, resulting very much uneven overlap coefficient values that are hard to interpret (see Table 1 and Table 2), because one cannot determinate whether the high $k_1(k_2)$ values are due to colocalization or simply because the green (red) channel has much higher intensity than its counterpart.

### 3.4 Manders' coefficients $m_1$ and $m_2$

Compared to $k_1$ and $k_2$, the advantage of Manders' coefficients is that they do not scale with the absolute intensity in both channels. The definition of Manders' coefficients is

$$m_1 = \frac{\sum_i R_{i,col}}{\sum_i R_i}$$
$$m_2 = \frac{\sum_i G_{i,col}}{\sum_i G_i}, \quad (4)$$

where $R_{i,col}$ is the intensity of the $i$th red pixel that has a nonzero $G_i > 0$, and $G_{i,col}$ is defined similarly. In simple terms, Manders' coefficients are defined as the fraction of coincident pixels. Manders coefficients are widely used but they could be misleading because one is actually interested in the fraction of colocalized molecules rather than coincident pixels, and again, overlapping resel and intracellular structures may produce coincident pixels without the presence of real colocalization of molecules. Therefore, Manders' coefficient share the same drawbacks that overlap coefficients have, except that they can be applied to image pairs with a relative intensity strength difference in two channels.

### 3.5 Intensity correlation quotient

This index is defined as

$$ICQ = \frac{\#_i\ (R_i - R_{avg})\,(G_i - G_{avg}) > 0}{N} - 0.5, \quad (5)$$

where $\#_i P$ stands for counting number of pixels satisfying condition $P$, and $N$ is the total number of pixels (Li *et al.*, 2004). In other words, ICQ basically measures the fraction of pixels that the red intensity value and the green intensity value are either both greater or both less than the average. Noticing the similarity between this definition and Eq. 1 in the numerators, it is not surprising that ICQ is also insensitive to overlapping resels but exaggerates colocalization because of spatial pattern of intracellular structures. Also, in Table 2 one can see that, when colocalization goes from zero to 15% (red) and 33% (green), the increment in ICQ values is very little. Therefore, ICQ can be sloppy in signifying the varying degree of colocalization, making it less suitable to study dynamic changes in cell biology.

### 3.6 Costes' randomization test

The Costes' randomization test is not an independent QCA method but is an effective way to detect inauthentic colocalization due to randomly coincident pixels (Costes *et al.*, 2004). To perform the test, one of the images is divided into small squares that are then randomly scrambled. The size of the squares should be large enough to cover a single PSF. It is highly probable that the scrambled images will have less colocalized

pixels, if colocalization between the original images is not purely coincidental. In practice the Pearson's correlation coefficient $R_r$ is used as the measure of colocalization, and the test is passed if 90% of the randomized images result a smaller $R_r$. In practice, image randomization is performed at least 200 times.

From Table 1 one can see that colocalized pixels due to resel overlapping cannot pass Costes' randomization test. Therefore, this test is essential when QCA methods that are sensitive to resel overlapping are used. However, this test gives only "yes-or-no" answers. When there is partial colocalization, this test can hardly be helpful for quantifying colocalization. Also, Table 2 shows that Costes' randomization test fails to distinguish spatial correlation due to intracellular structure from real colocalization. Passing this test is a thus necessary but not sufficient condition for the existence of colocalization.

### 3.7 Object based methods

All the methods discussed above are "pixel intensity based", in the sense that the basic element considered in these methods is individual pixels. Object based methods, on the other hand, are based on counting objects in the images. They utilize edge detection and image segmentation algorithms to identify individual objects in images, and then apply various criterions to determine colocalized objects rather than pixels. The specific approach used in this article measures the distance between the centers of mass of two objects, and if the two centers coincident within the uncertainty due to sampling, then they are counted as colocalized (Bolte and Cordelieres, 2006; Boutte *et al.*, 2006).

Object based methods obviously are very sensitive to resel overlapping: Two PSF can either be identified as two objects or as one object, depending on whether and how much they overlap. One could reduce resel overlapping by applying threshold, but there is no objective way to set the threshold values. In our analysis we used the threshold values provided automatically by the JACOP software. In order to remove influence from shot noises, only objects with size larger than 9 pixels are recognized by the software. The performance of the object based method on high density images is still not satisfactory, as shown in Table 1. Image pairs without colocalization were mistakenly measured to have significant degrees of colocalization, and for partially colocalized image pairs, the resulting values were very different from real values. Also, the result of the object based method can be unstable when PSF width changes. For example, in Table 1 one can see that at the same real colocalization value (33% and 25% for two channels, respectively), calculated colocalization values change from about 17% and 19% to 6.1% and 6.0% when PSF width goes from 3 pixels to 5 pixels.

### 3.8 Automatic threshold

The idea behind this method is fairly simple. All fluorescence signals may be decomposed into a random component and a colocalized component. It is assumed that the random components has lower intensity in general, and they are then identified by setting threshold: The signals in two channels below threshold values should have zero correlation (measured by Pearson's correlation coefficient) and are thus considered as the random components, while signals above thresholds are deemed as the colocalized components. The colocalization coefficients are simply the ratio of the integrated intensity of the colocalized component to the integration of all signals.

This simplistic method is obviously highly empirical. In the original paper where it was proposed (Costes *et al.*, 2004), it was tested with computer-simulated images but the images used were very different from real microscopic images. It has been reported that this method yields inaccurate result for images with high density of molecules. When tested with our computer simulation, it was also shown that this method is very vulnerable to both resel overlapping and spatial intracellular patterns, constantly producing very large overestimation to colocalization, and because of the exaggeration, this method is insensitive to the change in colocalization, too (See Table 1, Table 2 and Fig. 2).

### 3.9 Image cross-correlation spectroscopy (ICCS)

The essential equations in ICCS are

$$P_{red} = \frac{Cov(R,G)}{Var(G)}\frac{G_{avg}}{R_{avg}}$$
$$P_{green} = \frac{Cov(R,G)}{Var(R)}\frac{R_{avg}}{G_{avg}}, \quad (6)$$

where $Cov()$ and $Var()$ are used to denote covariance and variance, respectively. In ideal situation (without background, noises, and spatial pattern), the coefficient $P_{red}$ ($P_{Green}$) is equal to the fraction of colocalized molecules in the red (green) channel. But in practice, background and noise are always present. Therefore, covariance and variance are not calculated directly, but are derived through the correlation function defined as

$$G_{r,g}(u,v) = \frac{Cov(R, G_{u,v})}{R_{avg}G_{avg}}, \quad (7)$$

where $G_{u,v}$ stands for the image in the green channel being shifted by $u$ pixels horizontally and $v$ pixels vertically. Similarly one can define $G_{r,r}(u,v)$ and $G_{g,g}(u,v)$. The use of correlation functions to validate colocalization dates back 1996, when Van Steensel and coworkers proposed to measure the Pearson's correlation coefficient as a function of pixel shift (van *et al.*, 1996). Usually PSF can be well approximated by a Gaussian

function, and so are these correlation functions in ideal situation. ICCS calculates the correlation functions and then fit them into three-dimensional Gaussian functions, and the fitting parameters were used to calculate colocalization coefficients to reduce the influence from background and noises.

The main difficulty of the original ICCS approach is that the three-dimensional fit often fails, especially when colocalization is low. An "improved" version of ICCS was then developed (Comeau *et al.*, 2008), which scrambles the image pairs to produce a sharp Gaussian peak in the landscape of the correlation functions. Though the three-dimensional fit made easy, the process of image scrambling includes background due to out of focus light and spatial pattern of intracellular structures into the correlation functions. This essentially because by dividing image into small squares the length scales of spatial correlations are completely destroyed. In Table 2 we show that this approach overestimated the degree of colocalization and mistakenly reported the existence of colocalization for images actually with zero colocalization.

### 3.10 Protein proximity index (PPI)

This approach is also a successor of the original ICCS method (Wu *et al.*, 2010). It was realized that the correlation functions actually contain spatial correlations in several different length scales: At the shortest length scale, there is autocorrelation of shot noise, whose range is at about one pixel. At the length scale of resolution distance (PSF size at focus), there are autocorrelation of each fluorophores and cross-correlation due to colocalization of molecules, whose range is at about several hundred nanometers for diffraction-limited systems. There are also long range correlations due to the spatial pattern of intracellular structures and out of focus light, whose ranges are typically much longer than resolution distance. Obviously, the estimation of colocalization for confocal images should only use correlations at the length scale of PSF at focus. The shot noise can be effectively removed by image processing techniques such as deconvolution or median filter. A two-dimensional "double Gaussian fit" was used to differentiate correlation at the resolution length scale and the longer range correlations. The direction at which the two-dimensional fit is performed should be chosen such that the distinct between the short range correlations and the long range correlations is as obvious as possible, and usually it follows where the long range correlations decreases slowest. This technique allows one to extract even small degree of colocalization from long range correlations. In Table 1, Table 2 and Fig. 2 we show that the PPI method yields excellent results for computer-simulated images, despite the existence of both resel overlapping and spatial pattern of molecule distribution.

### 3.11 Biological images

Unlike computer-simulated images, the actual degree of colocalization is not known in biological images. It is easier to test QCA methods under changes in colocalization under different conditions. We applied QCA methods to images of mouse cardiomyocytes, where $\alpha$1C channel and ryanodine receptors (RyR) were separately labeled. The distribution and colocalization of these two kinds of proteins in a cell from a failing mouse heart were compared to those in a healthy heart cell as the control. It is expected that in the process of heart failure, the proteins not only lose the organized pattern, but also undergo a decrease in degree of colocalization. The original images have considerable amount of background due to out of focus light, detector noise, and nonspecific labeling. Thus, background subtraction and (or) deconvolution techniques were used to reduce their influences. Deconvolution is a controversial technique in colocalization analysis. While it usually produces sharper and cleaner images thus recommended by many researchers, it was also concerned that it introduces too drastic changes in the images and may result inaccurate assessment of colocalization. We adopted two ways to process the image:

1) Images were cropped, deconvoluted, and then threshold were applied. The resulting images are shown in Fig. 3.

2) Images were cropped, processed with a 3×3 median filter to remove shot noises, and then threshold were applied. The resulting images are shown in Fig. 4.

In each case, the preset threshold values automatically given by CoLocalization Pro 2.5 or JACOP are used. The aim is to test performance and consistency of QCA methods when different approaches of background reduction are used. Different background reduction approach may result different levels of resel overlapping, and a reliable QCA method is expected to be insensitive to different background reduction approaches. In Table 3 one can see that only some methods yield expected results for deconvoluted images, but most methods sensitive to resel overlapping, such as overlap coefficients and Manders' coefficients, are unable to differentiate the control and the failing heart sample. When applied on images without deconvolution, almost all methods except the PPI method fail to clearly signify the decreasing degree of colocalization from the control to the sick sample. The PPI method is the only approach that provides expected results under both image processing procedures. The reason why other approaches fail for undeconvoluted images may lie in fact that the image without deconvolution has more resel overlapping, since deconvolution techniques can effectively shrink PSF to deblur images. At least for this specific example, deconvolution is proven to be helpful in QCA and the PPI method is demonstrated to be the most reliable

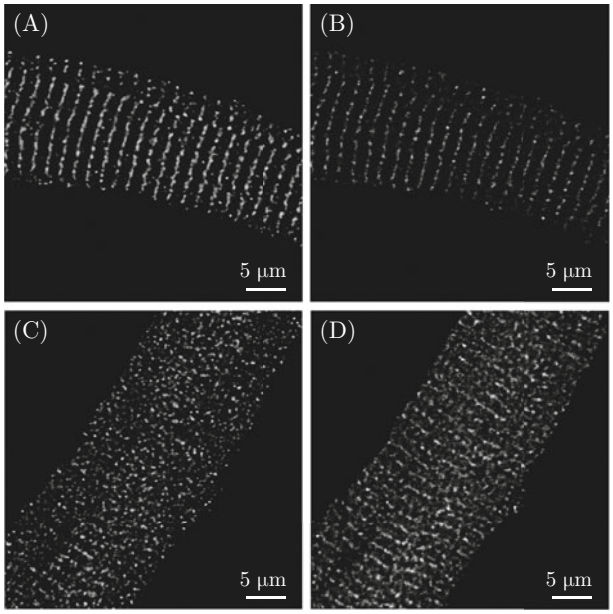


Fig. 3 Deconvoluted images (also cropped and thresholded) of ventricular cardiomyocytes from a control mouse and a mouse undergoing pressure overload 3 weeks after transaortic constriction (TAC). (A and B) α 1C calcium channel ($\alpha$1C) and ryanodine receptor (RyR) channel in the control cell, respectively. (C and D) $\alpha$1C channel and RyR channel in the failing heart cell, respectively.

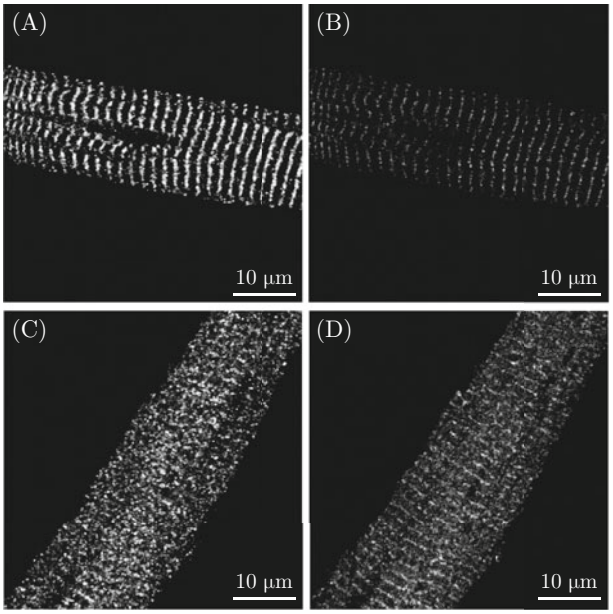


Fig. 4 Images (cropped and thresholded, but without deconvoluted) of the same cells as in Fig. 2. (A and B) $\alpha$1C calcium channel ($\alpha$1C) and ryanodine receptor (RyR) channel in the control cell, respectively. (C and D) $\alpha$1C channel and RyR channel in the failing heart cell, respectively.

**Table 3 Comparison of QCA methods applied on images of ventricular cardiomyocytes from a control mouse and a mouse undergoing pressure overload**

| | Not Deconvoluted | | Deconvoluted | |
|---|---|---|---|---|
| | Control | Failing Heart | Control | Failing Heart |
| Peasrson's correlation coefficient $R_r$ | 0.56 | 0.50 | 0.49 | 0.30 |
| Manders' overlap coefficient $R$ | 0.70 | 0.77 | 0.63 | 0.71 |
| Overlap coefficient $k_1$ and $k_2$ | 0.37, 1.30 | 0.76, 0.79 | 0.26, 1.57 | 0.65, 0.78 |
| Manders' coefficient $m_1$ and $m_2$ | 0.69, 0.92 | 0.92, 0.91 | 0.49, 0.84 | 0.89, 0.76 |
| ICQ | 0.39 | 0.33 | 0.40 | 0.26 |
| Automatic threshold | 55%, 87% | 60%, 74% | 48%, 69% | 39%, 29% |
| Costes' randomization test | Passed | Passed | Passed | Passed |
| Objected based method | 13%, 2.1% | 12%, 9.9% | 8.3%, 4.3% | 0.6%, 3.6% |
| ICCS (image scrambled) | 38%, 81% | 55%, 47% | 40%, 60% | 42%, 22% |
| PPI | 20%, 67% | 14%, 9.1% | 19%, 32% | 17%, 6.5% |

one among QCA methods we have tested.

## 4 Conclusions

In Table 4 we summarize the results we have obtained in testing the QCA methods. For QCA methods sensitive to resel overlapping, one should be cautious about its results for images with high molecular density, with low resolution, or with oversampling. For most QCA approaches except the PPI method, the result is not reliable if the images contain a clear intracellular structure.

Reliable QCA is difficult, because various QCA methods all have their drawbacks, and also because there is no universal background reduction procedure, especially when the degree of colocalization is intermediate or low. In order to enhance the reliability of QCA, it is strongly recommended that one apply several QCA methods and/or background reduction approaches to see whether consistent results can be achieved.

**Table 4 Summary of the testing for QCA methods**

| | Sensitive to resel overlapping | Sensitive to intracellular pattern | Comments |
|---|---|---|---|
| Peasrson's correlation coefficient $R_r$ | No | Yes | |
| Manders' overlap coefficient $R$ | Yes | Yes | |
| Overlap coefficient $k_1$ and $k_2$ | Yes | Yes | Only for images with similar intensity strength in two channels |
| Manders' coefficient $m_1$ and $m_2$ | Yes | Yes | |
| ICQ | No | Yes | Intermediate values insensitive to changes in colocalization value |
| Automatic threshold | Yes | Yes | |
| Costes' randomization test | No | Yes | Only gives "yes-or-no" answers, cannot quantify colocalization |
| Objected based method | Yes | Yes | |
| ICCS (image scrambled) | No | Yes | Image scrambling worsen the influence from background |
| PPI | No | No | |

The discussion of this article is based on the assumption that two associated molecules cannot be resolved by the microscope, which currently remains true in most cases. The advancement of super-resolution microscope and image processing techniques, however, may falsify this assumption, and by that time the mind set of QCA will be dramatically changed.

**Acknowledgements** This work was supported by National Institutes of Health grants No. HL088640 (E.S.) and American Heart Association Fellowship 10POST4230081 (Y.W.).

## References

[1] Adler, J., Parmryd, I. 2010. Quantifying colocalization by correlation: The Pearson correlation coefficient is superior to the Mander's overlap coefficient. Cytometry A 77, 733–742.

[2] Betzig, E., Patterson, G.H., Sougrat, R., Lindwasser, O.W., Olenych, S., Bonifacino, J.S., Davidson, M.W., Lippincott-Schwartz, J., Hess, H.F. 2006. Imaging intracellular fluorescent proteins at nanometer resolution. Science 313, 1642–1645.

[3] Bolte, S., Cordelieres, F.P. 2006. A guided tour into subcellular colocalization analysis in light microscopy. J Microsc 224, 213–232.

[4] Boutte, Y., Crosnier, M.T., Carraro, N., Traas, J., Satiat-Jeunemaitre, B. 2006. The plasma membrane recycling pathway and cell polarity in plants: Studies on PIN proteins. J Cell Sci 119, 1255–1265.

[5] Comeau, J.W.D., Costantino, S., Wiseman, P.W. 2006. A guide to accurate fluorescence microscopy colocalization measurements. Biophys J 91, 4611–4622.

[6] Comeau, J.W.D., Kolin, D.L., Wiseman, P.W. 2008. Accurate measurements of protein interactions in cells via improved spatial image cross-correlation spectroscopy. Mol Biosyst 4, 672–685.

[7] 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.

[8] Demandolx, D., Davoust, J. 1997. Multicolour analysis and local image correlation in confocal microscopy. J Microsc 185, 21–36.

[9] Dutartre, H., Davoust, J., Gorvel, J.P., Chavrier, P. 1996. Cytokinesis arrest and redistribution of actin-cytoskeleton regulatory components in cells expressing the Rho GTPase CDC42Hs. J Cell Sci 109, 367–377.

[10] Fox, M.H., Arndt-Jovin, D.J., Jovin, T.M., Baumann, P.H., Robert-Nicoud, M. 1991. Spatial and temporal distribution of DNA replication sites localized by immunofluorescence and confocal microscopy in mouse fibroblasts. J Cell Sci 99, 247–253.

[11] Gustafsson, M.G. 2000. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. J Microsc 198, 82–87.

[12] Hell, S.W. 2003. Toward fluorescence nanoscopy. Nat Biotechnol 21, 1347–1355.

[13] Lachmanovich, E., Shvartsman, D.E., Malka, Y., Botvin, C., Henis, Y.I., Weiss, A.M. 2003. Colocalization analysis of complex formation among membrane proteins by computerized fluorescence microscopy: Application to immunofluorescence copatching studies. J Microsc 212, 122–131.

[14] Li, Q., Lau, A., Morris, T.J., Guo, L., Fordyce, C.B., Stanley, E.F. 2004. A syntaxin 1, G alpha(o), and N-type calcium channel complex at a presynaptic nerve terminal: Analysis by quantitative immunocolocalization. J Neurosci 24, 4070–4081.

[15] Manders, E.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.

[16] Manders, E.M.M., Verbeek, F.J., Aten, J.A. 1993. Measurement of colocalization of objects in dual-color confocal images. J Microsc 169, 375–382.

[17] Ropero, A.B., Eghbali, M., Minosyan, T.Y., Tang, G., Toro, L., Stefani, E. 2006. Heart estrogen receptor alpha: Distinct membrane and nuclear distribution patterns and regulation by estrogen. J Mol Cell Cardiol 41, 496–510.

[18] van Steensel, B., van Binnendijk, E.P., Hornsby, C.D., van der Voort, H.T., 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.

[19] Wu, Y., Eghbali, M., Ou, J., Lu, R., Toro, L., Stefani, E. 2010. Quantitative determination of spatial protein-protein correlations in fluorescence confocal microscopy. Biophys J 98, 493–504.

[20] Zinchuk, V., Grossenbacher-Zinchuk, O. 2009. Recent advances in quantitative colocalization analysis: Focus on neuroscience. Prog Histochem Cytochem 44, 125–172.

[21] Zinchuk, V., Zinchuk, O. 2008. Quantitative colocalization analysis of confocal fluorescence microscopy images. In: Current Protocols in Cell Biology. Vol 39. John Wiley & Sons, 4.19.1–4.19.16.