

Model-Free Quantification and Visualization of Colocalization in Fluorescence Images

Aaron B. Taylor,^{1*} Maria S. Ioannou,² Jesse Aaron,¹ Teng-Leong Chew¹

¹Advanced Imaging Center, Janelia Research Campus, Ashburn, Virginia, 20147

²Howard Hughes Medical Institute, Janelia Research Campus, Ashburn, Virginia, 20147

Received 26 October 2016; Revised 9 December 2016; Accepted 28 December 2016

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Aaron B. Taylor, Advanced Imaging Center, Janelia Research Campus, Ashburn, Virginia 20147. E-mail: taylora10@janelia.hhmi.org

Maria S. Ioannou and Jesse Aaron contributed equally to this work.

Published online 00 Month 2018 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/cyto.a.23356

© 2018 International Society for Advancement of Cytometry

• Abstract

The spatial association between fluorescently tagged biomolecules in situ provides valuable insight into their biological relationship. Within the limits of diffraction, such association can be measured using either Pearson's Correlation Coefficient (PCC) or Spearman's Rank Coefficient (SRC), which are designed to measure linear and monotonic correlations, respectively. However, the relationship between real biological signals is often more complex than these measures assume, rendering their results difficult to interpret. Here, we have adapted methods from the field of information theory to measure the association between two probes' concentrations based on their statistical dependence. Our approach is mathematically more general than PCC or SRC, making no assumptions about the type of relationship between the probes. We show that when applied to biological images, our measures provide more intuitive results that are also more robust to outliers and the presence of multiple relationships than PCC or SRC. We also devise a display technique to highlight regions in the input images where the probes' association is higher versus lower. We expect that our methods will allow biologists to more accurately and robustly quantify and visualize the association between two probes in a pair of fluorescence images. © 2018 International Society for Advancement of Cytometry

• Key terms

colocalization; mutual information; correlation; image analysis; quantification; fluorescence microscopy

THE spatial association between the concentration of biomolecules in situ provides insight into their biological relationship and can be measured from fluorescence images of molecule-specific tags or probes. For example, the degree to which two biomolecules are similarly distributed across a cell may give insight into their co-trafficcking or co-compartmentalization. Although biophysical interactions can be inferred from images with the use of specialized probes, such as in Forster resonance energy transfer and bimolecular fluorescence complementation (1,2), in all other diffraction-limited images, pixel intensity represents only total fluorescence emissions and so is proportional to probe concentration. In these cases, spatial association can only be defined phenomenologically and is quantified based on measures of image similarity colloquially termed colocalization measures.

Several distinct measures of image similarity have been developed, each of which measures a different property of the images. Most simply, the two fluorescence signals are treated as a binary readout of probe presence or absence and the areal overlap or inter-object distances between the binarized signals are computed (3). When the distribution of probe concentrations matters, Mander's Overlap Coefficient (MOC) (4), and H_C (5) measure the fraction of the signals' total gray-scale intensity that co-occupies the same pixels. Finally, the correlation between the gray-scale intensities can be measured using either Pearson's Correlation Coefficient (PCC) or Spearman's Rank Coefficient (SRC), which are measures of linear and monotonic

correlation, respectively, (6,7) that quantify how the probe concentrations co-vary. A related measure, termed ICQ, sums over only the sign of the product of each normalized deviation from a line (8).

PCC and SRC are most closely related to an intuitive understanding of what it means for signals to be spatially associated, because they measure how the signals' intensities co-vary on a pixel-by-pixel basis [Eqs. (1) and (2)]. However, the assumptions of PCC and SRC can be limiting when applied to biological images, often making their results difficult to interpret in practice. First, PCC and SRC are model-based, in that they assume the signals are associated according to a linear or monotonic function. However, as can be seen in a scatter plot, biomolecular associations often appear more complex than a single, monotonic functional relationship. Second, these measures score deviations based on either the signals' intensity (PCC) or rank (SRC), making them sensitive to bright and dark outliers, as may be caused by probe precipitates/aggregates (bright pixels) or inadvertent inclusion of background during segmentation (dark pixels). Finally, these measures are global in that they sum (or "average") over all deviations, making it difficult for them to accurately measure a small number of correlated intensities within a large population of uncorrelated intensities (such as only within a particular organelle).

To mitigate these limitations of PCC and SRC, while maintaining an intuitive understanding of what it means for two probes' concentrations to be associated, we have developed ways to measure probes' association based on their statistical dependence. Statistical dependence quantifies the frequency at which pairs of outcomes are observed to occur together relative to chance alone and is mathematically more general than correlation. Linearly correlated outcomes are statistically dependent, but statistical dependence does not imply linear correlation. Graphically, statistical dependence quantifies how well pairs of observations (pixel intensities) conform to a relationship, without having to specify the type of relationship (such as a "linear" or "monotonic").

Our measures are based on the formulation of mutual information (MI), which was first developed in the field of information theory to compare the content of sent and received messages after transmission over a distorting channel (9). Point-wise mutual information (PMI) measures the statistical dependence between pairs of message contents, such as the letter present at corresponding message positions (10) [Eq. (3)], while mutual information measures the average dependence between contents over the entire message (9) [Eq. (4)]. Here, we envision pairs of probe concentrations represented in each pixel as message contents and signal-containing regions of the images as the entire message.

We term our two measurement methods normalized point-wise mutual information (NPMI) and adjusted mutual information (AMI) (see Methods). AMI is an image-wide measure that quantifies the conformity of scatter plot data to an underlying relationship and so is analogous to PCC and SRC. However, AMI is model-free, making no assumption about the type of relationship between the probes, and so is equally applicable to any pair of signals, a property termed

equitability (11). AMI is also defined in terms of probabilities (intensity is only an outcome label), and so is not strongly impacted by outlying values that are large in magnitude but of low probability. AMI is the sum of the NPMI values, which can themselves be used to highlight in the input images the locations where the probe concentrations are more or less strongly associated. We show that when applied to biological image sets, our measures provide more intuitive, robust, and biologically meaningful quantification of the association between two signals than PCC or SRC.

METHODS

Model-Based Measures of Correlation

Pearson's correlation coefficient (PCC) [Eq. (1)] and Spearman's rank correlation coefficient (SRC) [Eq. (2)] have been widely used in the cell biology literature to measure the correlation between pixel intensities in two images.

Pearson's correlation coefficient (PCC):

$$r = \frac{\sum (I_{IA} - I_{avgA})(I_{IB} - I_{avgB})}{\sqrt{\sum (I_{IA} - I_{avgA})^2 \sum (I_{IB} - I_{avgB})^2}} \quad (1)$$

Spearman's rank correlation coefficient (SRC):

$$\rho = \frac{\sum (I_{IA} - I_{avgA})(I_{IB} - I_{avgB})}{\sqrt{\sum (I_{IA} - I_{avgA})^2 \sum (I_{IB} - I_{avgB})^2}} \quad (2)$$

I_{IA} is the pixel intensity at location l in image A , while I_{IB} is the pixel intensity at the corresponding location in image B . In Eq. (2), I represents each intensity's rank. Summands sum over all pixels within the signals' intersection.

Point-Wise Measurement of Statistical Dependence

Pointwise mutual information. Given two random variables A and B , outcomes A_i and B_j are strongly dependent if, given that A_i has occurred, the probability that B_j will also occur is high relative to chance alone. The dependence between any pair of outcomes A_i and B_j can be quantified according to their point-wise mutual information (10):

$$\text{PMI}(A_i, B_j) = P(A_i, B_j) \log_2 \left(\frac{P(A_i, B_j)}{P(A_i)P(B_j)} \right) \quad (3)$$

Here, A_i and B_j are interpreted as the probe concentration present in a given pixel in image A and image B , respectively. $P(A_i)$ and $P(B_j)$ are the probabilities of each probe concentration occurring within each image (marginal probabilities), while $P(A_i, B_j)$ is the probability that the given probe concentrations occur together (joint probability). The weighting factor $P(A_i, B_j)$ is used so that summing over the PMI values for all joint outcomes equates to mutual information (MI), a global measure of the average dependence between outcomes [Eq. (5) below] (9). Positive PMI values indicate dependence ($P(A_i, B_j) > P(A_i)P(B_j)$), zero indicates independence ($P(A_i, B_j) = P(A_i)P(B_j)$), and negative PMI values indicate anti-dependence ($P(A_i, B_j) < P(A_i)P(B_j)$). Anti-dependence is not considered further here. (Anti-dependence indicates less association than expected by chance alone and is unrelated to

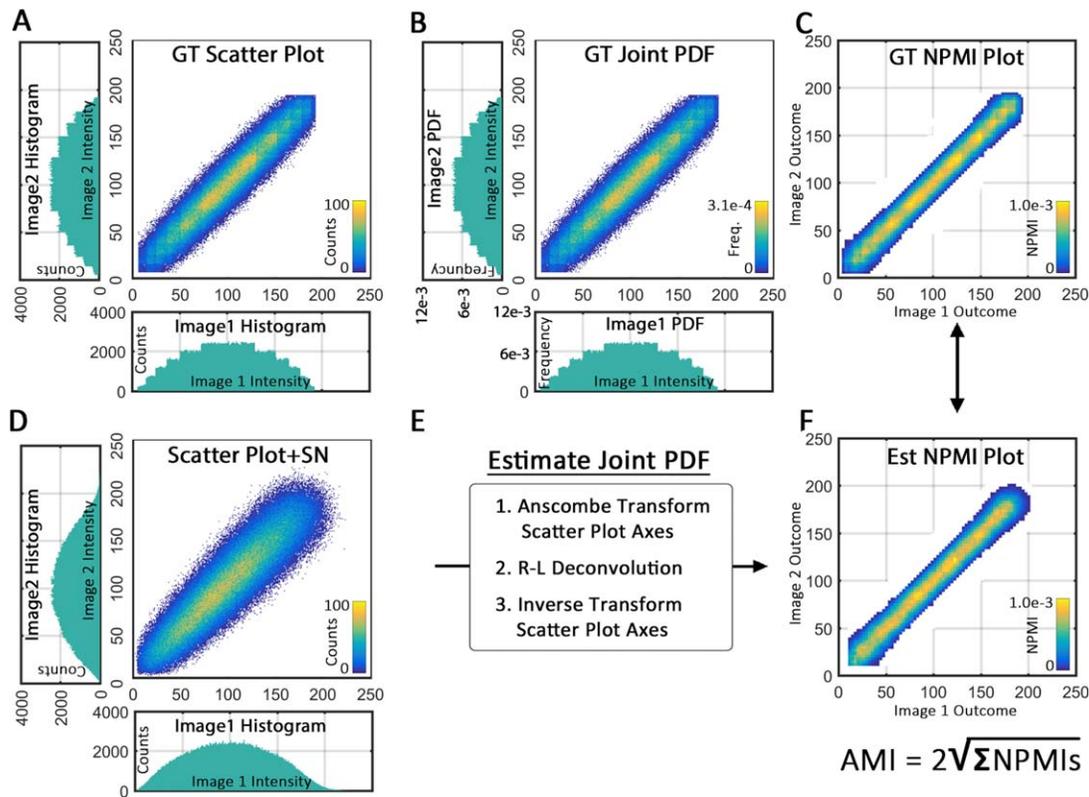


Figure 1. (A) The histograms and scatter plot between two ground truth (GT) (noise-free) images where probe concentrations exhibit linear correlation. (B) Normalization of the histograms defines the images' marginal probability density functions (PDFs), while normalization of the scatter plot defines their joint PDF. These PDFs are used to calculate the NPMI values between every pair of outcomes, which can be displayed as (C) an NPMI Plot (negative values corresponding to anti-dependence are not shown). In the presence of shot noise, the true joint and marginal distributions must be estimated before the NPMI values are calculated. (D) The histograms and scatter plot between the ground truth images in A after addition of modulatory shot noise (SN). (E) An overview of our procedure for estimating the images' true joint distribution. (F) The estimated NPMI plot (negative values not shown) calculated from the estimated joint and marginal distributions. AMI is calculated from all estimated NPMI values (both positive and negative) based on the equation shown. [Color figure can be viewed at wileyonlinelibrary.com]

negative linear correlation, which indicates a linear relationship with a negative slope.

Normalized pointwise mutual information. We adapted PMI to make it more useful in the context of comparing fluorescence images. First, we normalized the PMI values by the smallest marginal information ($H(x)$) present in either image alone, which ensures that the sum of all PMI values cannot exceed 1.0 (12). The result is a measure we term normalized point-wise mutual information (NPMI):

$$\text{NPMI}(A_i, B_j) = \frac{\text{PMI}(A_i, B_j)}{\min[H(A), H(B)]} \quad (4)$$

where $H(X) = -\sum_{x_i \in X} p(x_i) \log[p(x_i)]$

The collection of NPMI values between all pairs of probe concentrations can be displayed in a matrix, analogous to a scatter plot (see Fig. 1C). We term this matrix an NPMI plot and further describe its use below.

Global Measurement of Statistical Dependence

Normalized mutual information. Normalized Mutual Information (NMI) is the sum over all NPMI values and so is

an image-wide measure of the average statistical dependence between the probes' concentrations, analogous to PCC or SRC:

$$\text{NMI} = \sum_{B_i \in B} \sum_{A_j \in A} \text{NPMI}(A_i, B_j) \quad (5)$$

The NMI between any pair of probe concentrations ranges from 0.0 (no dependence) to 1.0 (complete dependence). NMI is suitable for quantifying the average dependence between discrete, known probe concentrations given a large sample size.

In practice, when applied to fluorescence images, probe concentrations must be estimated from noisy intensity measurements and limited sample sizes. To account for these technical issues, we define our global measure of dependence as Adjusted Mutual Information (AMI):

$$\text{AMI} = 2\sqrt{\text{NMI}} \quad (6)$$

AMI is designed to be applied to an estimate of the probes' true joint distribution as represented on a discrete graph with a fixed number of bins (as described below). When applied in this way, AMI ranges from >0.0 (no dependence) to 1.0

(complete dependence). The factor of 2 compensates for the fixed number of bins (see below). The square root ensures that AMI will respond similarly to PCC in the presence of linear correlation, as AMI is related to the areal spread of the joint distribution while PCC is related to the spread's deviation measured along only one dimension.

Estimating Probes' Joint Distribution in the Presence of Shot Noise

Modern detectors used in fluorescence microscopy are approximately shot-noise limited, meaning that pixel intensity is proportional to the number of photons detected during the exposure time. Probe concentration is statistically related to pixel intensity via a Poisson process, where the Poisson average (λ) is proportional to the probe concentration. Thus, probe concentration can only be estimated from many samples (pixels). Our joint distribution estimation procedure is based on the following observation: given many samples of a pair of fixed probe concentrations labeled with fluorescent tags, a scatter plot aggregates the samples into a cluster of joint outcomes. The position of the cluster's peak corresponds to the probes' true concentrations, while the cluster's spread is due mainly to shot noise. Thus, given a shot noise level λ , the relationship between two probes' true concentrations can be estimated by deconvolving the scatter plot with a kernel of radius $\sqrt{\lambda}$ that represents the spread due to the shot noise. The shot noise level is found through a one-time microscope calibration as described below. Although many denoising algorithms have been designed for images (13), we were unable to identify any that, in general, provided as robust of an estimate of the true joint distribution between two images as is achieved by denoising the scatter plot directly.

To implement the denoising, the scatter plot's coordinate system was first Anscombe transformed to convert the shot-noise-induced spread into a Gaussian spread of constant standard deviation (14). The transformed scatterplot was then deconvolved with a Gaussian kernel describing the shot noise spread using Richardson-Lucy deconvolution (15). The number of iterations was determined automatically based on count frequencies in the scatter plot. Finally, the coordinate system was inverse transformed, resulting in an estimate of the two probes' true joint distribution in intensity space. Outcomes corresponding to $\lambda < 10$ photons were excluded from analysis, because the Anscombe transform is not accurate for $\lambda < 10$. For very low count applications ($\lambda < 10$), an exact Anscombe transform could be implemented (16).

Due to the limited sample size encountered in real images, next, the estimated joint distribution was down-sampled into 64×64 bins, which represents a judicious trade-off between robustness to smaller sample sizes and the strength of the dependence in the joint distribution that can be accurately represented. Fewer (larger) bins not only reduce sparsity and increase counts per bin but also low-pass filter the estimated joint distribution. To understand the impact of the low-pass filtering on the estimate of the joint distribution, consider a noise-free joint distribution with spread along one dimension described by a normal curve $f(x)$. The spread's

Fourier Transform $F[f(x)](k) = e^{-2\pi^2 k^2}$ contains non-negligible ($>1\%$) power out to $k < \sim 0.5$ cycles per standard deviation. Thus, by Nyquist's Theorem, the joint distribution must be sampled at $\geq 2x \times 0.5 = 1x$ per standard deviation for its spread to be accurately represented.

We found from simulations that a linearly correlated joint distribution with a Gaussian spread whose standard deviation is equal to 1 bin width has a PCC value of ~ 0.98 . The NMI at this limiting spread is ~ 0.5 . Thus, we introduced the factor of 2 in Eq. (6) so that AMI maintains a range of 0.0–1.0 when applied to joint distributions whose spread can be accurately represented after re-binning. If a correlation > 0.98 is encountered, AMI will exceed 1.0, and the user will know that the result is inaccurate. We show in Figure 3 that even when the same protein is co-expressed except with different tags, a PCC > 0.95 was never measured so the low-pass filtering caused by re-binning does not impact our measurements in practice.

Relatedly, given a noisy, independent joint distribution of limited sample size, the lower bound of 0.0 is never reached due to slight amplification of high frequencies during the deconvolution process. We found that simulated, independent joint distributions measure at ~ 0.05 (not shown). In practice, perfect independence between the probes is not observed (see Fig. 4).

Creation of an NPMI Display

Given an NPMI plot and the two, corresponding input images, each pair of probe concentrations can be labeled according to their NPMI value. We term such a labeled image an NPMI display (e.g., see Fig. 5). Due to the limited sample sizes encountered in real images, the raw NPMI plot is smoothed prior to creating an NPMI display. Thus, an NPMI display is an approximation of the relative dependence between outcomes. We have also chosen to display only positive NPMI values, and thus, an NPMI display illustrates the level of dependence only. Locations in an NPMI display that have a negative or no NPMI value are displayed as black. A pixel intensity combination may not have an NPMI value if the combination only occurred in a region of the input images that fell outside the signals' intersection or if the combination was due to noise.

MATLAB Application

The algorithms described above including estimation of the probes' true joint distribution in the presence of shot noise, calculation of the NPMI plot, calculation of AMI, and creation of an NPMI display have been implemented in MATLAB and are easily accessible via a user-friendly graphical user interface. A stand-alone application and source code are available at https://www.dropbox.com/sh/f79geb4esz8b3wa/AADKwJAI10kkwJaC9ds3d_sBa?dl=0. The user's only responsibility is to define the signals through thresholding and calculate the microscope system's gain as described below.

Generation of Simulated Images

We used a wide range of simulated, 8-bit image sets to test the fidelity and robustness of our measures and joint distribution estimation procedure. Ground truth (exactly known

probe concentrations) pixel intensity distributions were generated using Beta functions. A given intensity distribution was then translated into an image by considering the image's x -axis to be proportional to count frequency and assigning each intensity to the appropriate number of image pixel columns, as described by the distribution.

We created image sets with different levels of association based on a scrambling procedure defined by a scramble level (SL). Our scrambling procedure was designed to match an intuitive understanding of what it means for two signals to be associated and is equally applicable to any joint relationship. Scrambling means that each pixel intensity in the input image was replaced with a similar intensity with some probability. Similar was defined by the scramble level, which is a contiguous intensity range centered on the initial intensity, from which alternative pixel intensities are drawn. The range was weighted according to a Gaussian function of the distance (in intensity) from the initial intensity. To preserve the image's overall intensity distribution and avoid creating intensities that did not originally exist, the intensity distance weights were further multiplied by each intensity's image-wide probability of occurrence. For each pixel location, a replacement intensity was then drawn from this probability distribution using inverse transform sampling.

To simulate shot noise, ground truth images were contaminated with modulatory Poisson noise, meaning that each initial intensity is considered to be the mean of a Poisson process. The resulting SNR of ~ 10 for an 8-bit image, is comparable to the noise level in real fluorescence images. For each pair of ground truth images, modulatory shot noise was added in five, independent replicates, resulting in five pairs of noisy images so that the statistics of our estimators could be measured.

System Gain Calibration

The joint distribution estimation procedure described above requires knowing the number of photoelectrons per pixel intensity (ADU) in an image, which we calculated as previously described (18). Briefly, uniformly fluorescent samples roughly matched to the dye concentrations encountered in fluorescently stained samples (5–0.5 $\mu\text{g}/\text{ml}$ free dye) were imaged using acquisition settings appropriate for an actual experiment (do not use averaging). Note that on a point-scanning microscope, both PMT voltage and dwell time determine system gain. Laser power was adjusted so that the image's average intensity fell near the middle of the detector's dynamic range and an image was recorded. Laser power (or dye concentration) was then adjusted $\sim 2\times$ higher and $\sim 2\times$ lower to change the average number of photons emitted, while avoiding saturation and again images were recorded. From the mean and standard deviation of these images, the average number of photoelectrons per average pixel intensity can be found according to: $1/(\text{stdev intensity}/\text{mean intensity})^2$. For added flexibility, the calibration procedure can be repeated using other system gains as well.

Preparation and Imaging of Biological Samples

Sample preparation. U2OS cells were plated on poly-D-lysine coated coverslips and transiently transfected using Jet-Prime Reagent (Polyplus) according to the manufacturer's instructions. Constructs used were: EGFR-GFP (19), mCh-Rab13 (20), and EGFP-N1, mCh-C1, mtagBFP-N1, transferrin receptor (TfR)-mCh, mEmerald-Rab9, Actin-mCh, and mCh-Rab9 from the Michael Davidson Fluorescent Protein Collection. Twenty-four hour post-transfection, cells were fixed with 4% paraformaldehyde and mounted in Vectashield hardset antifade mounting media (Vector Laboratories).

Confocal imaging. Samples were imaged on a Zeiss 880 LSM using a 63x oil 1.4 NA objective with the following settings: Dwell time, 5.3 μs ; Pixel size, 90 nm; Pin hole, 1.2 AU (for 488 nm); PMT Ch1 (green channel) gain: 720; PMT Ch2 (red channel) gain: 800. These settings were chosen so that the average number of photoelectrons per pixel intensity was equal to one, which matches the noise level used in our simulated image sets. Laser powers were adjusted based on protein expression levels to fill each detector's dynamic range while avoiding saturation. Laser powers typically ranged from 1 to 6% for 488 nm, and 5 to 40% for 594 nm. When imaging a blue fluorescent protein, PMT Ch1 was used and collimation of the 405 nm light path was adjusted to minimize longitudinal chromatic aberration. We do not recommend attempting to measure association based on wide-field or low numerical aperture confocal images.

Segmentation of signal-containing regions. Signal masks were defined based on spatially filtered versions of each image, as our signals corresponded to spatially definite structures (e.g., organelles). Each image was low-pass filtered by convolution with a standard deviation = 6 radius Gaussian kernel and subtracted from the original, resulting in a high-pass image that was then manually thresholded and median filtered with a one-pixel radius to produce the final mask. The black region of the mask (not signal) was pasted into the original image, leaving only the signal of interest. No processing of any kind was applied to the signal-containing regions. Even linear contrast adjustments will alter the number of photons per pixel intensity found during microscope calibration.

RESULTS

Concepts and Calculations

We first illustrate our statistical approach to measuring the spatial association between probe concentrations using simulated images where pixel intensity is exactly proportional to probe concentration (i.e., no shot noise). Two such images' histograms, as well as their scatter plot, are shown in Figure 1A. The scatter plot consists of a single linear correlation. Normalizing the graphs in Figure 1A by the total number of outcomes results in the probes' marginal and joint probability distributions, respectively, as shown in Figure 1B. These three probability distributions are then used to calculate the NPMI

value between every pair of probe concentrations in the images according to Eq. (4) (see Methods).

The NPMI values between the probe concentrations in these images can be displayed in a matrix as shown in Figure 1C. We term this matrix of NPMI values an NPMI plot. Although the pattern of NPMI values is grossly similar to the pattern of values in the images' scatter plot, the graphs differ in three notable ways: 1) NPMI values incorporate data from both the images' marginal and joint distributions, while the scatter plot is proportional to the joint distribution alone. 2) NPMI values are real numbers that measure statistical dependence, while the scatter plot values are positive integers that measure numbers of outcomes. 3) The axes of an NPMI plot are defined by unit-less labels, while in a scatter plot the axes have units of pixel intensity. We show below how these differences are beneficial when measuring the association between probes in biological images.

The calculation of NPMI values just described cannot be directly applied to shot-noise-containing fluorescence images, as pixel intensity is no longer exactly proportional to probe concentration. Thus, we developed a deconvolution-based approach to estimate the probes' true joint distribution, prior to calculating NPMI values. Figure 1D shows the scatter plot between the same images used in Figure 1A, except now, modulatory shot noise has been added (resulting in an average signal to noise standard deviation ratio (SNR) of ~ 10). Figure 1E outlines our joint distribution estimation procedure, which requires a one-time, microscope gain calibration (see Methods for details). NPMI values are then calculated from the estimated joint distribution exactly as described above for noise-free joint distributions. Figure 1F shows the NPMI values calculated from this estimate, which we term an estimated NPMI plot.

To produce an image-wide measurement of the average dependence between the probes' concentrations, we then calculated the images' AMI by summing over all of the NPMI values, as shown by the equation in Figure 1F [and Eq. (6)]. As derived in the Methods section, AMI ranges from 0 to 1, with 0 indicating no dependence (no association) and 1 indicating strongest measurable dependence (association). AMI is analogous to PCC and SRC but differs from them in important ways: 1) AMI is based on the probabilities of the outcomes, rather than on their magnitudes (or magnitude of deviations). 2) AMI incorporates data from both the images' pixel intensity histograms as well as their scatterplot, while PCC and SRC are calculated from the scatter plot alone. 3) AMI does not assume that the joint outcomes conform to a monotonic functional relationship. As we show below, these properties make AMI a more robust and meaningful measure of the association between probe concentrations when applied to biological images.

Comparison of AMI to PCC and SRC in the Presence of Linear Joint Relationships

We next used simulated datasets to compare the behavior of AMI relative to PCC and SRC given a linear correlation between the probes' concentrations, as PCC and SRC were

designed for this context. Figure 2A shows representative scatter plots of shot-noise containing signals that exhibit various levels of linear correlation as defined by a scramble level (SL). A larger scramble level reduces the association, and thus also the correlation, between the signals (and vice versa; see Methods). The line graph in Figure 2B compares the estimated (Est) AMI values (calculated from noisy signals), ground truth (GT) AMI values (calculated from corresponding noise-free signals), PCC values, and SRC values over a range of scramble levels. All three measures (PCC, SRC, Est AMI) responded in a similar way, although Est AMI exhibited a somewhat more linear response. The average Est AMI also largely matched the GT AMI, demonstrating the fidelity (unbiasedness) of our estimation procedure. In the case of extremely strong correlation ($PCC > 0.95$; SL3), the Est AMI slightly ($\sim 5\%$) underestimated the GT AMI, because AMI was not designed to measure levels of co-association unlikely to be encountered in biological images (see Methods). The standard deviation of all measures was smaller than the graph markers in every case. (As shown in Supporting Information Figure 1, applying replicate-based noise correction (21) scales the PCC and SRC values by a few percent, but does not change the shape of the response curves).

Because AMI makes no assumptions about the joint distribution, its calculations are entirely based on the observed data and so it requires a far larger sample size than PCC or SRC. The simulated images used here contain 320,000 samples (800×400 pixels) so that sample size is not limiting. As shown in Supporting Information Figure 2, $>10,000$ pixels are required for accurate (within $<2\%$) AMI measurements given biologically-relevant levels of association. This sample size corresponds to an area of 100×100 signal-containing pixels, which is readily achieved in most biological images (see Discussion).

Comparison of AMI to PCC and SRC in the Presence of Multiple Relationships

We next assessed the performance of AMI, PCC, and SRC using scatter plots that violate the assumption of a single, monotonic correlation. Scatter plots derived from biological images often exhibit multiple clouds of points suggestive of more than one type of association (see below). For example, two proteins may associate to different degrees in distinct sub-cellular compartments. Figure 2C shows examples of simulated scatter plots where the probes' joint distribution exhibits multiple relationships, but the levels of association are the same as in Figure 1A. Figure 2D shows the response of PCC, SRC, GT AMI, and Est AMI over the same range of scramble levels as in Figure 1B. Est AMI again responds in a similar way to the probes' scramble level (association). However, PCC and SRC produce dramatically lower results (40–50% decrease) relative to the case of linear correlation (Fig. 1B) at every scramble level. This discrepancy occurs because PCC and SRC are now responding to the lack of a single, monotonic relationship rather than to the probes' level of association. Thus, AMI is a more accurate and intuitive measure of probe association when multiple relationships are present.

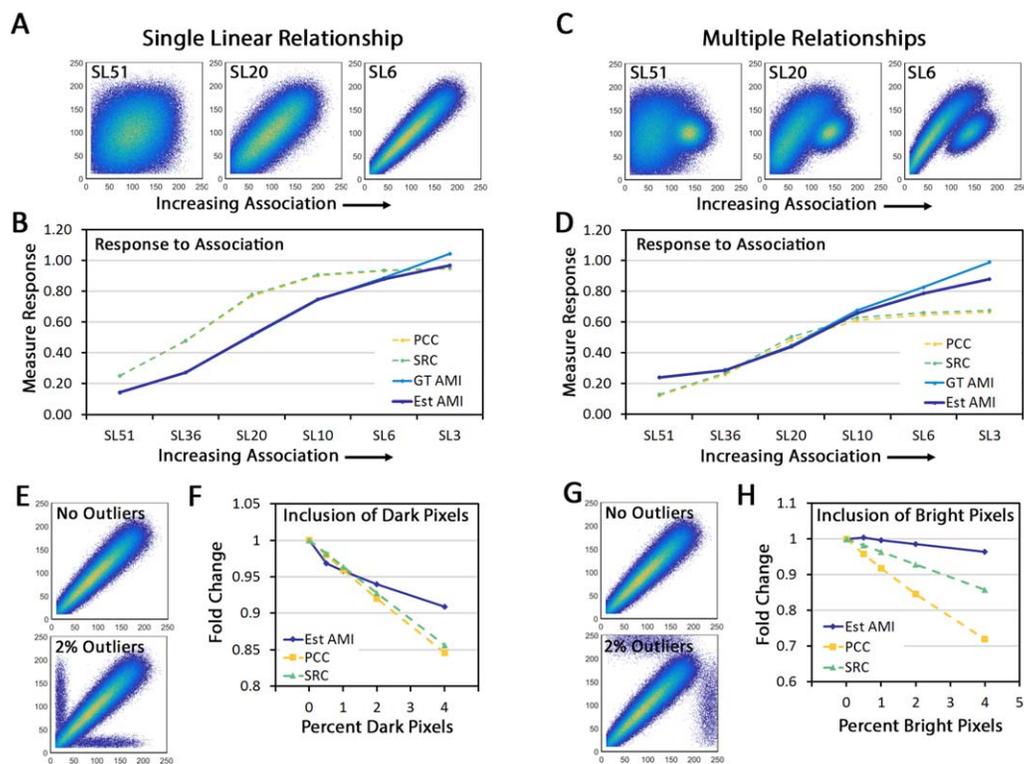


Figure 2. (A) Examples of shot-noise-containing scatter plots exhibiting increased linear correlation (and association) from left to right, as defined by a scramble level (SL) (see Methods). (B) A line graph showing the response of PCC, SRC, GT AMI, and Est AMI to various scramble levels. The Est AMI closely matches the GT AMI, except when the correlation is extremely strong (corresponding to a PCC value of ~ 0.98). (C) Examples of shot-noise-containing scatter plots exhibiting multiple relationships and increasing levels of association from left to right. (D) A line graph showing the response of PCC, SRC, GT AMI, and the Est AMI to various scramble levels. AMI reports a similar range of responses to scramble level as in B, while PCC and SRC show a dramatically decreased response. (E) (top) A shot-noise-containing scatter plot exhibiting linear correlation and (bottom) a scatter plot of the same images after inclusion of 2% dark pixels. (F) A line graph showing the response of Est AMI, PCC, and SRC in the presence of dark pixels relative to the outlier-free case. (G) (top) A shot-noise-containing scatter plot exhibiting linear correlation and (bottom) a scatter plot of the same images after inclusion of 2% bright pixels. (H) A line graph showing the response of the Est AMI, PCC, and SRC in the presence of bright pixels relative to the outlier-free case. Est AMI is least affected by outliers. [Color figure can be viewed at wileyonlinelibrary.com]

Comparison of AMI to PCC and SRC in the Presence of Outliers

Another practical issue in real fluorescence images is the presence of pixel intensity outliers, to which PCC and SRC are highly sensitive, as mentioned above. Therefore, we compared all measures' responses to scatter plots exhibiting linear correlation alone versus the same linear correlation with a small fraction of either bright or dark outliers added (by random replacement).

We first addressed the inclusion of dark (background) pixels into the signal regions, which can occur due to inaccuracies during segmentation. The upper panel of Figure 2E shows a scatter plot exhibiting linear correlation and no outliers. The lower panel shows the same dataset, except 2% of pixels have been randomly replaced with dark values. Figure 2F shows the response of Est AMI, PCC and SRC to the addition of various percentages of dark pixels, expressed as a fold change relative to their response when outliers are absent. For $<2\%$ dark outliers, all measures perform similarly. For 2–4% background, Est AMI decreases by $\sim 10\%$, while PCC and SRC decrease by $\sim 15\%$. (The response of AMI and SRC are

comparable when background pixels only overlap signal pixels of lower intensities (not shown)). Therefore, AMI is 33% less biased than PCC or SRC by the presence of dark outliers.

We next addressed the inclusion of bright pixels to simulate bright, non-specific staining (such as antibody precipitates). The upper panel of Figure 2G shows a scatter plot exhibiting linear correlation and no outliers, while the lower panel shows the same dataset, except 2% of pixels have been randomly replaced with bright values. Figure 2H shows the response of Est AMI, PCC, and SRC to various percentages of bright pixels, expressed as a fold change relative to their response when outliers are absent. AMI measures within $\sim 1\%$ of the outlier-free data in every case. PCC and SRC are much more strongly affected, decreasing by up to $\sim 35\%$ and $\sim 15\%$, respectively, when 4% of pixels are bright outliers. (Saturated pixels were excluded from measurement in every case). Thus, AMI is almost unbiased by the presence of bright outliers.

AMI is more robust to both bright and dark outliers than PCC or SRC, because these outcomes are probabilistically rare and their magnitudes do not factor into the calculation. As outliers are common in images of biological samples, these

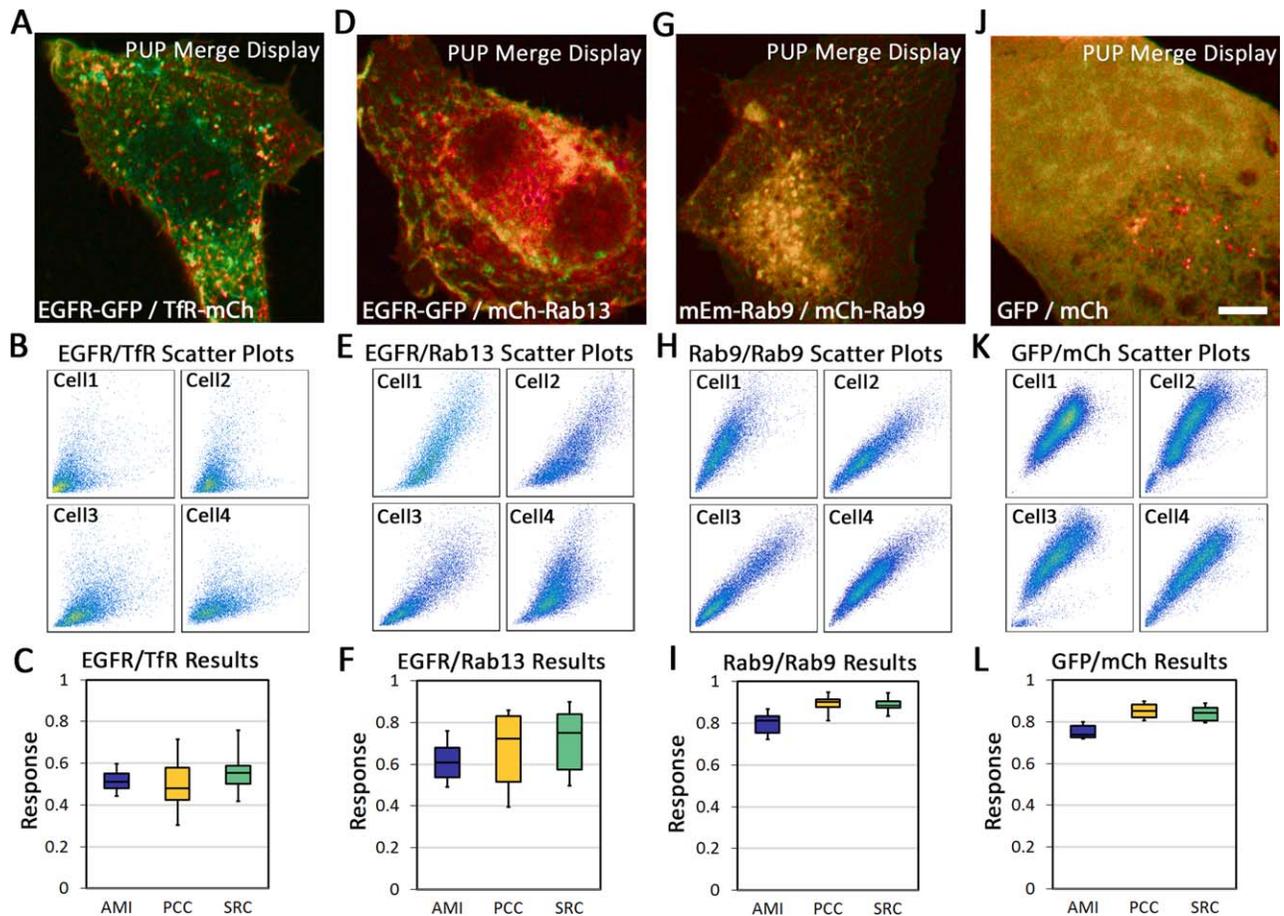


Figure 3. The response of AMI versus PCC and SRC when applied to proteins transiently expressed in U2OS cells that exhibit linear correlation. For all images shown, a gamma was applied and the intensities were scaled to similar levels of saturation prior to generating a PUP Merge display. A PUP Merge display is similar to a red-green merge display except the brightness of each channel is independent of the color and a larger range of hues is used such that Ch1 alone is displayed as greenish blue and Ch2 alone as reddish blue (17). (A) A cell expressing EGFR-GFP and TfR-mCh. (B) Representative scatter plots showing joint relationships between EGFR-GFP and TfR-mCh. Axes range from 0–255. (C) A box plot showing the distribution of AMI, PCC, and SRC results from all 10 pairs of EGFR-GFP \times TfR-mCh images. (D) A cell expressing EGFR-GFP and mCh-Rab13. (E) Representative scatter plots showing joint relationships between EGFR-GFP and mCh-Rab13. (F) A box plot showing the distribution of AMI, PCC, and SRC results from all EGFR-GFP \times mCh-Rab13 images. (G) A cell expressing mEm-Rab9 and mCh-Rab9 as a positive control for maximal association. (H) Representative scatter plots showing joint relationships between mEm-Rab9 and mCh-Rab9. (I) A box plot showing the distribution of AMI, PCC, and SRC results from all mEm-Rab9 \times mCh-Rab9 images. (J) A cell expressing free GFP and free mCh. (K) Scatter plots showing joint relationships between free GFP and free mCh. (L) A box plot showing the distribution of AMI, PCC, and SRC results from all free GFP \times free mCh images. Two experiments were performed per condition. In all box plots, whiskers represent minimum and maximum values, while bars represent first to third quartile range and center line is the median. Scale bar is 5 μ m and applies to all images. [Color figure can be viewed at wileyonlinelibrary.com]

results suggest that AMI will provide more accurate measurement of probe association.

Application of AMI to Fluorescence Images Exhibiting Linear Correlation

We next tested if AMI's performance on simulated data would translate into more robust measurements of the spatial association between probe concentrations in real images. We transiently expressed various pairs of proteins fused to fluorescent tags in U2OS cells and imaged them on a confocal microscope using pre-calibrated settings.

Signal-containing regions were defined in each image and calculations were then performed on the raw pixel intensities within the signal's intersection (see Methods).

We first compared AMI to PCC and SRC on pairs of proteins whose association appeared to approximately exhibit linear correlation. Figure 3A shows a representative example of a cell expressing EGFR-GFP and TfR-mCh, two receptors who are both expressed on the plasma membrane but otherwise trafficked through largely distinct pathways. TfR is constitutively recycled to the plasma membrane while EGFR is largely targeted for lysosomal degradation (22,23). Representative scatter plots shown in Figure 3B suggest that these proteins exhibit a weak but mostly linear joint relationship. As shown in Figure 3C, we then measured the AMI, PCC, and SRC for each pair of images and displayed the population of results using a box plot. (The measurements were sometimes not normally distributed). All measures yielded a similar median

response ($AMI = 0.51$, $PCC = 0.48$, and $SRC = 0.55$), but AMI was less variable, exhibiting an inter-quartile range (Q3-Q1) of 0.072, compared to 0.16 and 0.09 for PCC and SRC, respectively. AMI also showed the smallest difference between maximum and minimum values.

Figures 3D–3F show results of a similar experiment comparing the joint distribution of EGFR-GFP to mCh-Rab13, a small GTPase that influences EGFR trafficking and co-traffics with EGFR on endosomes (24). Again, the scatter plots suggested a linear correlation. The measures' median response was $AMI = 0.61$, $PCC = 0.72$, and $SRC = 0.75$, while AMI again had the smallest inter-quartile range (0.14 vs. 0.31 and 0.27 for PCC and SRC, respectively). The larger median values for $EGFR \times Rab13$ compared to $EGFR \times TfR$ are expected as EGFR and Rab13 co-localize on both the plasma membrane and endosomes, while EGFR and TfR are also both found on the plasma membrane but otherwise traffic to distinct endosomal compartments. A Mann-Whitney U test showed that each measure detected a significant difference between the $EGFR \times TfR$ and $EGFR \times Rab13$ groups, with AMI being most significant due to its lower variability ($AMI: P = 0.01$, $PCC: P = 0.04$, $SRC: P = 0.02$).

Finally, we co-expressed mEm-Rab9 with mCh-Rab9 as a positive control for strong association as shown in Figures 3G–3I. This experiment likely defines the upper limit of co-association that will be observed in practice when measuring proteins that are not covalently linked. All measures detected a strong association, with median responses of $AMI = 0.81$, $PCC = 0.90$, and $SRC = 0.88$. In this case, PCC's and SRC's interquartile range was lower than AMI's, while the max-min differences were comparable. The reduced variability of PCC and SRC for $PCC > 0.9$ may be in part due to their reduced sensitivity to changes in association when $PCC > 0.9$, as shown in Figure 2B.

Taken together, these data show that when applied to fluorescence images where the joint relationship is approximately linear, AMI often exhibits less variability than PCC and SRC, except when the association is very high and almost perfectly linear ($PCC > 0.9$).

Use of Biological Controls to Measure Specific Association

We first examined if raw association (or correlation) values are inherently meaningful, that is, reflect the level of specific biological association between the probes. For example, non-specific (biologically irrelevant) association could also occur due to cellular geometry or the low-pass filtering of the microscope's point spread function. To test this, we co-expressed free GFP with free mCh, which are not expected to exhibit any specific biological association. As shown in Figures 3J–3L, all measures' median response was high, with $AMI = 0.74$, $PCC = 0.85$, and $SRC = 0.84$. (Strong levels of association were also observed when comparing non-specific labeling with two secondary antibodies; not shown.) Although statistically less associated than mEm-Rab9 \times mCh-Rab9 ($P < 0.02$ for all measures), these large values none-the-less show that large association values do not imply the existence of a biologically-specific relationship.

Thus, we wanted to develop biological controls that could be used as a baseline for the level of non-specific association. We reasoned that this baseline level could then be subtracted from the raw association between proteins of interest, leaving only the specific, biologically-relevant, level of association. The association between any protein of interest and a free fluorescent protein should provide such a non-specific baseline. As a special case, if we consider free GFP and free mCh to be the proteins of interest, we should then express free GFP (or free mCh) with another free fluorescent protein and subtract this baseline level of non-specific association. Here, this procedure amounts to performing the same experiment twice and subtracting the results. Thus, the specific association between free GFP and free mCh is zero, consistent with our knowledge that free fluorescent proteins do not associate in a biologically meaningful way. Following this same logic, the specific correlation (as measured using PCC and SRC) between free GFP and free mCh is also zero. We apply this same logic below to measure the specific association between several other pairs of proteins.

Application of AMI to Fluorescence Images Exhibiting Multiple Relationships

We next measured the specific association between EGFR and TfR, whose raw association levels were measured above and have been duplicated in Figures 4A–4C for comparison. As a non-specific control, we expressed EGFR-GFP with free mCh as shown in Figures 4D–4F. As shown in Figure 2E, representative scatter plots of the $EGFR-GFP \times$ free mCh relationship indicate that these proteins exhibit multiple relationships, similar to the simulated data in Figure 2D. PCC and SRC are not applicable to this type of relationship. This inability to measure the non-specific control is a previously unrecognized but serious practical limitation of PCC and SRC. However, AMI will still measure the association accurately in the presence of multiple relationships.

To further emphasize the inapplicability of PCC and SRC, we proceeded to apply all three measures to this non-specific-control data. The median responses were $AMI = 0.42$, $PCC = 0.02$, and $SRC = 0.09$, with AMI having a much smaller inter-quartile range. We then subtracted these non-specific control medians from the previous $EGFR \times TfR$ median results to measure the level of specific association. We find that $PCC_{Cor} = PCC_{EGFR \times TfR} - PCC_{EGFR \times mCh} = 0.46$, and $SRC_{Cor} = SRC_{EGFR \times TfR} - SRC_{EGFR \times mCh} = 0.46$ but $AMI_{Cor} = AMI_{EGFR \times TfR} - AMI_{EGFR \times mCh} = 0.09$. Thus, the (wrongly) corrected PCC and SRC results lead to the conclusion that EGFR substantially associates with TfR, while the corrected AMI result leads to the conclusion that little specific association occurs. The AMI-based conclusion is most consistent with orthogonal experiments, which find that EGFR and TfR occupy largely distinct cellular compartments (22,23).

We also co-expressed EGFR-GFP with mCh-Actin, which are also expected to show little specific association, as shown in Figure 4G. (While EGFR may bind F-actin on occasion, recall that for spatial association or correlation to be measurable, the concentration distribution of the two biomolecules must be

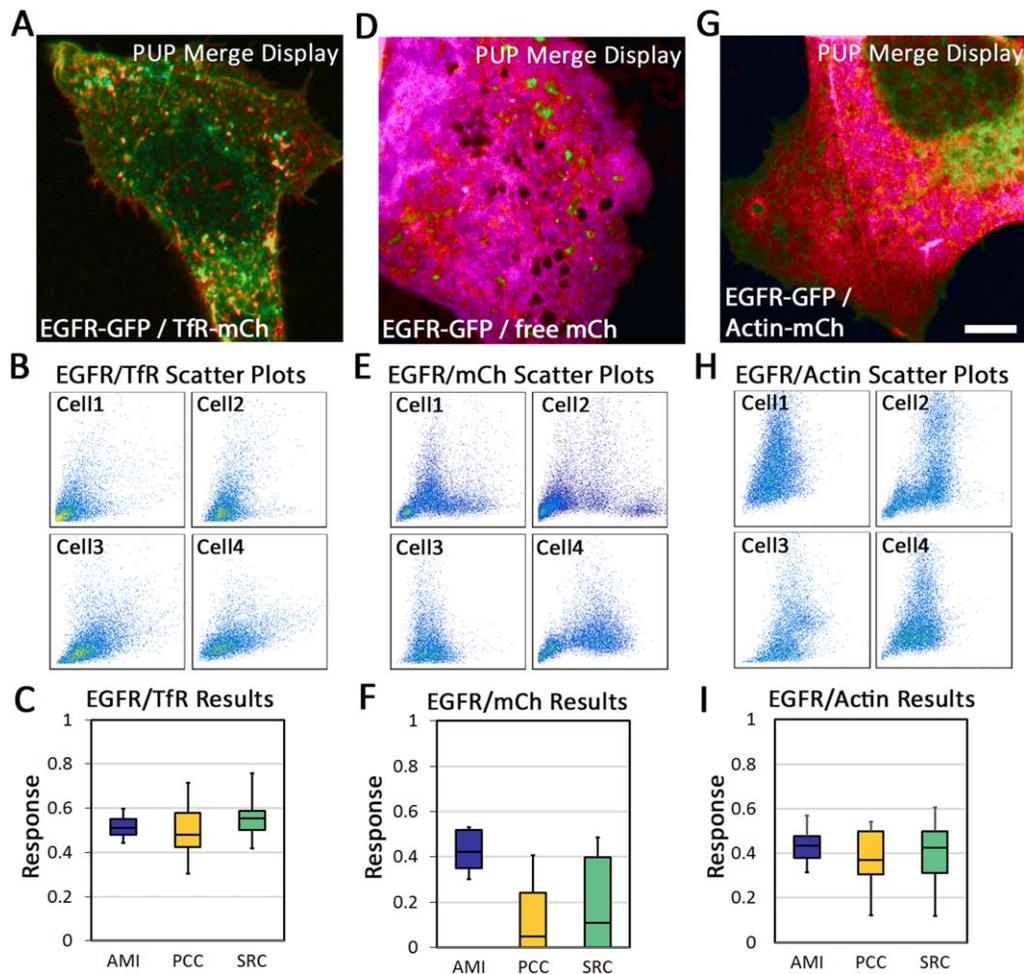


Figure 4. The response of AMI versus PCC and SRC when applied to proteins transiently expressed in U2OS cells that exhibit multiple relationships. **(A–C)** Data for cells expressing EGFR-GFP and Tfr-mCh previously shown in Figure 3 and repeated here for comparison. **(D)** A cell expressing EGFR-GFP and free mCh as a control for non-specific association. **(E)** Representative scatter plots showing joint relationships between EGFR-GFP and free mCh, which exhibit multiple relationships. Axes range from 0–255. **(F)** A box plot showing the distribution of AMI, PCC, and SRC results from all 10 pairs of EGFR-GFP \times mCh images. Negative PCC and SRC measurements occurred but fall below the graph's y-axis. **(G)** A cell expressing EGFR-GFP and Actin-mCh. **(H)** Representative scatter plots showing joint relationships between EGFR-GFP and Actin-mCh, which also exhibit multiple relationships. Axes range from 0–255. **(I)** A box plot showing the distribution of AMI, PCC, and SRC results from all 10 pairs of EGFR-GFP \times Actin-mCh images. Two experiments were performed per condition. Scale bar is 5 μ m and applies to all images. [Color figure can be viewed at wileyonlinelibrary.com]

similar, which is not expected in this case given that actin is everywhere and should be unrelated to the spatial distribution of EGFR). As shown in Figure 4H, the scatter plots again exhibit multiple relationships. Thus, PCC and SRC are again inapplicable, but if we do apply them and perform the corrections, we find that $PCC_{Cor} = PCC_{EGFR \times Actin} - PCC_{EGFR \times mCh} = 0.36$, and $SRC_{Cor} = SRC_{EGFR \times Actin} - SRC_{EGFR \times mCh} = 0.35$ but $AMI_{Cor} = AMI_{EGFR \times Actin} - AMI_{EGFR \times mCh} = 0.01$. The corrected PCC and SRC results suggest a modest (and statistically significant by M-W U test) specific correlation between EGFR and Actin while the corrected AMI result suggests no specific (and not statistically significant) association.

Visualizing the Spatial Distribution of Association

So far, we have demonstrated the properties of AMI versus PCC and SRC, all of which are image-wide, summary

measures. A weakness of image-wide measures is that a small region of associated pixels is not readily detectable when mixed within a much larger region of unassociated pixels, such as may occur if biomolecules associate within only one sub-cellular compartment but are present ubiquitously. Additionally, the location of the associated region(s) cannot be visualized, although this spatial information may be biologically significant. PCC and SRC could be adapted to supply more local results, for example by computing them over a sliding window, but the window would need to contain dozens of pixels, resulting in a loss of resolution. Coste's thresholding method can also segment a small population of linearly correlated pixels within a larger, unassociated population (25), but does not provide a graded score for every pixel intensity combination.

As NPMI values are a measure of the dependence between every pair of probe concentrations in two images (see

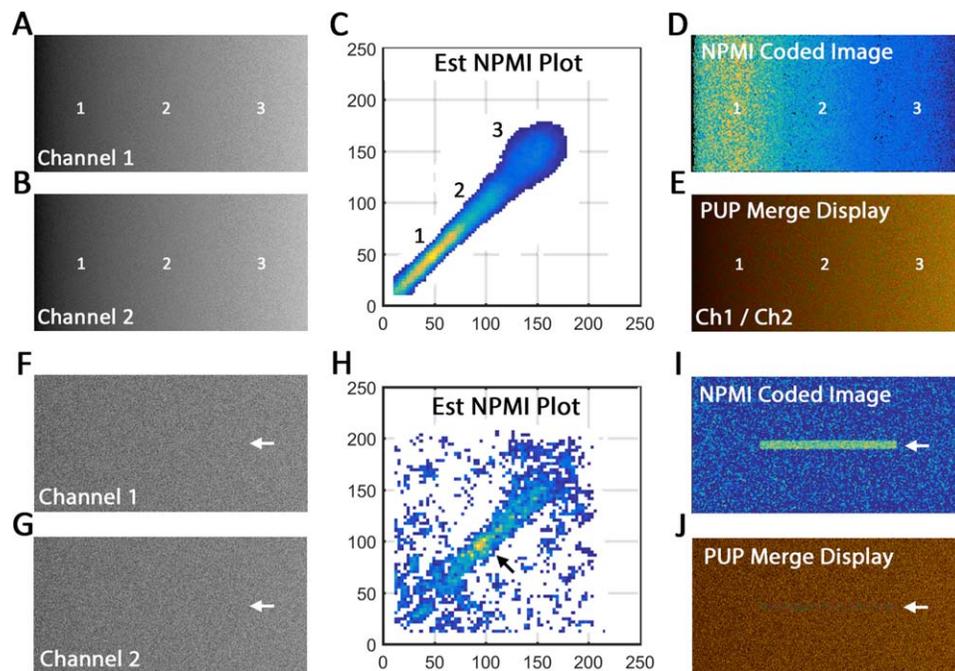


Figure 5. (A–B) A pair of simulated images evenly divided into three levels of linear correlation (SL6, SL10, and SL20). The distinct regions of association are not visible but are marked with numbers (1–3). (C) The Est NPMI plot between the images in A and B, where the three levels of association are now apparent (1–3). (D) The corresponding NPMI display, where the locations of the different regions are now clearly visible. (E) A PUP merge image display of the same images. (F–G) A pair of simulated images containing a small (4% of total) population of linear correlated pixels within an unassociated population. The region containing associated pixels is marked with an arrow. (H) The Est NPMI plot between the images in F and G, where the associated pixel intensities are now visible. (I) The corresponding NPMI display where the small region of associated intensities is clearly visible. (J) A merge image display of the same images. [Color figure can be viewed at wileyonlinelibrary.com]

Fig. 1), the data in an NPMI plot can be used to quantify the probes' level of association within individual pixels. We term this mapping an NPMI display (see Methods for details).

We first illustrate the creation of an NPMI display using simulated images. Figures 5A and 5B show two, shot-noise-containing, gray scale images that contain three regions with different levels of linear correlation (marked 1, 2, 3). Figure 5C shows the estimated NPMI plot between these images, where each level of association is now visible. These NPMI values are then used as a look-up table to label each location in the input images according to the probes' level of association. The resulting NPMI display is shown in Figure 5D, where each region of different association is now visible. Figure 5E shows that these regions are not readily visible in a merged image display.

Figures 5F–5G show another example of a pair of images that contain a small region (4% of total) of strongly linearly correlated probe concentrations (arrow) within a larger region where the probes are unassociated. Figure 5H shows the NPMI plot between these images and Figure 5I the corresponding NPMI display. The associated region is again difficult to visualize in a merged image display, as shown in Figure 5J. Thus, NPMI displays make visible the level of association between the probes at each location in the images.

Visualizing the Spatial Distribution of Association in Fluorescence Images

We finally applied the NPMI display to real images. Rather than display raw NPMI values, as above, we again used

a biological control to correct for the level of non-specific association, except here the control must be internal as the control correction is applied on a pixel by pixel basis. To display only the specific component of the association, we expressed two proteins of interest (A and B) along with a third, free fluorescent protein (C) in the same cell. We then created an NPMI display between every pair of signals ($A \times B$; $A \times C$; $B \times C$). The pixel-wise maximum of the $A \times C$ and $B \times C$ NPMI displays was defined as non-specific association and then subtracted from the $A \times B$ NPMI display, resulting in a corrected NPMI display that shows the level of specific association in each pixel (cNPMI display).

Figure 6 illustrates the creation and interpretation of an cNPMI display. Figures 6A–6C show images of a U2OS cell expressing EGFR, TfR, and free BFP. Figures 6D–6F show the NPMI display for each pair of images. The pixel-wise maximum of the control NPMI displays ($EGFR \times BFP$ and $TfR \times BFP$) was calculated and then subtracted from the $EGFR \times TfR$ NPMI display. The resulting cNPMI display is shown in Figures 6G and 6H. This display shows that EGFR specifically associates with TfR at the cell periphery and in some, but not all, round internal vesicles. Figures 6I and 6J show the same $EGFR \times TfR$ signals displayed using a previously published method termed an nMDP display that was intended to visualize “colocalization” (26). nMDP assigns larger values to pixel intensity combinations that are far from each signal's mean and has no inherent connection to linear correlation or

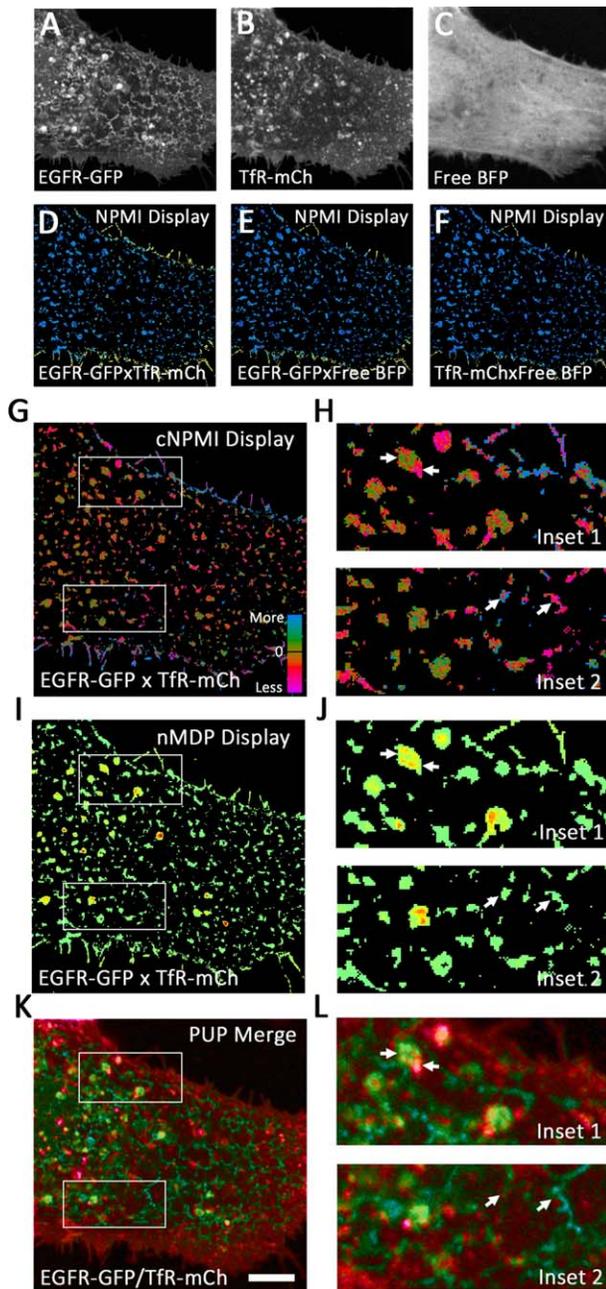


Figure 6. A U2OS cell transiently expressing (A) EGFR-GFP, (B) TfR-mCh, and (C) free BFP. NPMI displays based on the (D) EGFR \times TfR images, (E) EGFR \times BFP images, and (F) TfR \times BFP images all displayed using the same scale. To correct for non-specific association on a local basis, the maximum of the EGFR \times BFP and TfR \times BFP NPMI displays was calculated and subtracted from the EGFR \times TfR NPMI display on a pixel-wise basis. Black regions were defined as not signal during the segmentation process. (G) The corrected NPMI display shows where and to what degree EGFR and TfR are more (and less) specifically associated relative to the free BFP internal control. (H) Insets from G showing structures with a range of association levels. (I) The same signals displayed using an nMDP display, which scores pixel intensity combinations according to their relative intensity value. (J) Insets with arrows indicating structures where the cNPMI display and nMDP display disagree. (K) A PUP merge display of the same signals to provide cellular context. (L) Merge insets. [Color figure can be viewed at wileyonlinelibrary.com]

statistical dependence. Figures 6K and 6L show a merge display to provide cellular context.

DISCUSSION

We have presented new measures of the association between fluorescent probe concentrations based on the concept of statistical dependence. Adjusted Mutual Information (AMI) is a global measure consistent with an intuitive understanding of what it means for probes to be associated that remains equally quantitative (equitable) regardless of the type of joint relationship. AMI is also relatively unaffected by the presence of bright and/or dark outliers such as are commonly encountered in fluorescence images of biological samples. As a result, AMI is often a more statistically powerful measure of probe association than PCC or SRC. We also showed how AMI results can be corrected for non-specific association through use of biological controls, which is not routinely possible using PCC and SRC. Finally, we showed how the NPMI values on which AMI is based can be used to visualize the level of association between probe concentrations, independent of the concentrations themselves, a biologically relevant property that was often previously invisible.

Due to the presence of shot noise in fluorescence images, our measurements are based on an estimate of the true (denoised) joint distribution between the probes' concentrations. An analytical shot noise correction factor termed replicate-based noise correction has been previously derived for monotonic joint relationships (21). However, this technique assumes zero sample movement between replicate images, which is difficult to achieve in practice, increases the variability of the "corrected" measurements, because the correction factor is itself an estimate, and it is inaccurate when the distribution of probe concentrations in the two images is very different. In contrast, our denoising procedure estimates the noise level from one-time, single images of dye in solution, removes the spread due to noise directly from the scatter plot, and is equally applicable to any joint relationship.

One requirement of our measures is a large sample size. As PCC and SRC make strong assumptions about the type of joint distribution, just a few dozen samples are typically enough for a statistically meaningful measurement. As our measures make no assumptions, the shape of the joint distribution must be inferred from the data. For AMI, we found that $\geq 10,000$ signal-containing pixels are sufficient to accurately measure biologically-relevant joint distributions, which corresponds to the signals covering $>100 \times 100$ pixels. NPMI displays may require even larger samples, as the dependence between every pair of intensity outcomes is displayed separately. The intersection of the signals that we measured here typically contained 10,000–20,000 pixels. When sample sizes modestly $< 10,000$ are encountered, oversampling can be used during acquisition (which was not done here). For very small sample sizes, our measures cannot be applied and SRC should be used instead.

AMI values always and only represent the average strength of association between outcomes, which makes AMI results always meaningful and straight forward to interpret. However, AMI values do not describe a scatter plot's "shape."

For example, AMI gives the same range of responses (0–1) to negative linear correlation as it does to positive linear correlation, as the level of association is the same in both cases. If the main goal is to quantify the direction of the slope of a linear relationship, PCC or SRC could be used instead. Relatedly, if the scatter plot contains an equal mixture of positive and negative linear correlations, AMI would measure the strength of the association, while PCC and SRC would measure zero, but none of these measures will indicate that a mixture of positive and negative linear correlations is present. Other measures such as maximal information nonparametric exploration (MINE) (27) and the earth movers distance metric (EMD) (17) have been specifically designed to measure scatter plot shapes and are generally useful for this purpose. The sign on PCC and SRC values is useful when a monotonic correlation is present, but otherwise confounds measurement of association with measurement of “shape,” which is why these measures often provide misleading results when applied to real biological images.

Beyond the merits of any one measure, we cannot over-emphasize the need for biological controls to correct for non-specific association. We showed that over AMI’s theoretical range of 0.0–1.0, the average association between ~any target protein and free fluorescent protein is about 0.4, far from the 0.0 expected if the proteins were randomly co-distributed. The causes for this non-specific association were not investigated but likely include the 3 D convolution of optical and cellular factors. We showed how to correct AMI and NPMI results for this non-specific association by subtracting control results from the association of interest, either at the population level (as shown for AMI) or on a pixel-by-pixel basis (as shown for NPMI). Although here we have focused on the mathematical aspects control corrections, expression levels and the cellular location of the control protein should be designed to match that of the proteins of interest. For example, if the proteins of interest are both membrane targeted, the control should also be membrane targeted. Others have recently proposed purely mathematical corrections for non-specific (non-local) association within the image plane (28), but our results show that pixel intensity-based measurements are mostly influenced by factors acting in the axial direction and so are best addressed using experimental controls as described here.

We have developed new ways to quantify and visualize the specific association between fluorescence probes that are more equitable, robust to outliers, and biologically controlled than PCC or SRC. cNPMI displays can be used to visualize for the first time where within the cell the specific association is higher versus lower, irrespective of the abundance of either fluorescent probe. We expect these tools will provide new avenues for discovery in cell biology and provide a user-friendly MATLAB GUI to make these algorithms widely available.

AUTHOR CONTRIBUTIONS

A.T. conceived the measures, coded the algorithms, performed the analysis, and wrote the article. M.I. prepared

biological samples and provided cell biology insights. J.A. coded the GUI and recommended the microscope calibration procedure. T.C. provided discussion and suggested improvements. All authors edited the manuscript and have approved the final version.

LITERATURE CITED

- Kodama Y, Hu CD. Bimolecular fluorescence complementation (BiFC): A 5-year update and future perspectives. *Biotechniques* 2012;53:285–298.
- Shrestha D, Jenei A, Nagy P, Vereb G, Szöllösi J. Understanding FRET as a research tool for cellular studies. *Int J Mol Sci* 2015;16:6718–6756.
- Helmuth JA, Paul G, Sbalzarini IF. Beyond co-localization: Inferring spatial interactions between sub-cellular structures from microscopy images. *BMC Bioinformatics* 2010; 11:372.
- Manders E, Verbeek F, Aten J. Measurement of co-localization of objects in dual-colour confocal images. *J Microsc* 1993;169:375–382.
- Herce HD, Casas-Delucchi CS, Cardoso MC. New image colocalization coefficient for fluorescence microscopy to quantify (bio-)molecular interactions. *J Microsc* 2013;249:184–194.
- Manders EM, Stap J, Brakenhoff GJ, van Driel R, Aten JA. Dynamics of three-dimensional replication patterns during the S-phase, analysed by double labelling of DNA and confocal microscopy. *J Cell Sci* 1992;103(Pt 3):857–862.
- French AP, Mills S, Swarup R, Bennett MJ, Pridmore TP. Colocalization of fluorescent markers in confocal microscope images of plant cells. *Nat Protoc* 2008;3: 619–628.
- Li Q. A syntaxin 1, Galpha(o), and N-type calcium channel complex at a presynaptic nerve terminal: Analysis by quantitative immunocolocalization. *J Neurosci* 2004;24: 4070–4081.
- van der Lubbe J. *Information Theory*. Cambridge, England: Cambridge University Press; 1997.
- Church K, Hanks P. Word association norms, mutual information, and lexicography. *Comput Linguist* 1990;16:22–29.
- Kinney JB, Atwal GS. Equitability, mutual information, and the maximal information coefficient. *Proc Natl Acad Sci USA* 2014;111:3354–3359.
- Watanabe S. Information theoretical analysis of multivariate correlation. *IBM J Res Dev* 1960; 4:66–82.
- Buades A, Coll B, Morel J. A review of image denoising algorithms, with a new one. *Multiscale Model Simul* 2005;4:490–530.
- Anscombe FJ. The transformation of poisson, binomial and negative-binomial data. *Biometrika* 1948;35:246–254.
- Lucy LB. An iterative technique for the rectification of observed distributions. *Astron J* 1974;79:745.
- Makitalo M, Foi A. Optimal inversion of the Anscombe transformation in low-count Poisson image denoising. *IEEE Trans Image Process* 2011;20:99–109.
- Rubner Y, Tomasi C, Guibas LJ. The Earth Mover’s Distance as a metric for image retrieval. *Int J Comput Vis* 2000;40:99–121.
- Mortara L, Fowler A. Evaluations of Charge-Coupled Device (CCD) performance for astronomical use. *Proc SPIE* 1981; 290:28–33.
- Carter RE, Sorokin A. Endocytosis of functional epidermal growth factor receptor-green fluorescent protein chimera. *J Biol Chem* 1998;273:35000–35007.
- Ioannou MS, Bell ES, Girard M, Chaineau M, Hamlin JNR, Daubaras M, Monast A, Park M, Hodgson L, McPherson PS, et al. DENND2B activates Rab13 at the leading edge of migrating cells and promotes metastatic behavior. *J Cell Biol* 2015;208:629–648.
- Adler J, Pagakis SN, Parmryd I. Replicate-based noise corrected correlation for accurate measurements of colocalization. *J Microsc* 2008;230(Pt 1):121–133.
- Driskell OJ, Mironov A, Allan VJ, Woodman PG. Dynein is required for receptor sorting and the morphogenesis of early endosomes. *Nat Cell Biol* 2007;9:113–120.
- Leonard D, Hayakawa A, Lawe D, Lambright D, Bellve KD, Standley C, Lifshitz LM, Fogarty KE, Corvera S. Sorting of EGF and transferrin at the plasma membrane and by cargo-specific signaling to EEA1-enriched endosomes. *J Cell Sci* 2008;121(Pt 20): 3445–3458.
- Abou-Zeid N, Pandjaitan R, Sengmanivong L, David V, Le Pavec G, Salamero J, Zahraoui A. MICAL-like1 mediates epidermal growth factor receptor endocytosis. *Mol Biol Cell* 2011;22:3431–3441.
- Costes SV, Daelemans D, Cho EH, Dobbin Z, Pavlakis G, Lockett S. Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophys J* 2004;86:3993–4003.
- Jaskolski F, Mülle C, Manzoni OJ. An automated method to quantify and visualize colocalized fluorescent signals. *J Neurosci Methods* 2005;146:42–49.
- Reshef DN, Reshef YA, Finucane HK, Grossman SR, McVean G, Turnbaugh PJ, Lander ES, Mitzenmacher M, Sabeti PC. Detecting novel associations in large data sets. *Science* 2011;334:1518–1524.
- Zaritsky A, Obolski U, Gan Z, Reis CR, Kadlecova Z, Du Y, Schmid SL, Danuser G. Decoupling global biases and local interactions between cell biological variables. *Elife* 2017; 6:e22323.