Quantifying receptor trafficking and colocalization with confocal microscopy

Pike, Jeremy; Styles, Iain B; Rappoport, Joshua Z; Heath, John K

DOI:
10.1016/j.ymeth.2017.01.005

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Pike, J, Styles, IB, Rappoport, JZ & Heath, JK 2017, 'Quantifying receptor trafficking and colocalization with confocal microscopy', *Methods*, vol. 115, pp. 42–54. https://doi.org/10.1016/j.ymeth.2017.01.005

Methods 115 (2017) 42–54

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Quantifying receptor trafficking and colocalization with confocal microscopy

Jeremy A. Pike [a,b,c,*], Iain B. Styles [d], Joshua Z. Rappoport [b,e], John K. Heath [b]

[a] PSIBS Doctoral Training Centre, School of Chemistry, University of Birmingham, Birmingham B15 2TT, UK
[b] School of Biosciences, University of Birmingham, Birmingham B15 2TT, UK
[c] Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Cambridge CB2 0RE, UK
[d] School of Computer Science, University of Birmingham, Birmingham B15 2TT, UK
[e] Center for Advanced Microscopy and the Nikon Imaging Center, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA

ARTICLE INFO

Article history:
Received 14 July 2016
Received in revised form 20 January 2017
Accepted 23 January 2017
Available online 26 January 2017

Keywords:
Colocalization
Trafficking
Endocytosis
Confocal
Receptor
EGFR

ABSTRACT

Confocal microscopy is a powerful tool for the study of cellular receptor trafficking and endocytosis. Unbiased and robust image analysis workflows are required for the identification, and study, of aberrant trafficking. After a brief review of related strategies, identifying both good and bad practice, custom workflows for the analysis of live cell 3D time-lapse data are presented. Strategies for data pre-processing, including denoising and background subtraction are considered. We use a 3D level set protocol to accurately segment cells using only the signal from fluorescently labelled receptor. A protocol for the quantification of changes to subcellular receptor distribution over time is then presented. As an example, ligand stimulated trafficking of epidermal growth factor receptor (EGFR) is shown to be significantly reduced in both AG1478 and Dynasore treated cells. Protocols for the quantitative analysis of colocalization between receptor and endosomes are also introduced, including strategies for signal isolation and statistical testing. By calculating the Manders and Pearson coefficients, both co-occurrence and correlation can be assessed. A statistically significant decrease in the level of ligand induced co-occurrence between EGFR and rab5 positive endosomes is demonstrated for both the AG1478 and Dynasore treated cells relative to a control. Finally, a strategy for the visualisation of co-occurrence is presented, which provides an unbiased alternative to colour overlays.

1. Introduction

The signalling and trafficking of cellular receptors are highly interlinked processes [1–3]. Ligand induced signalling regulates endocytosis and receptor trafficking within the endocytic network, which in turn attenuates receptor signalling. Moreover, the hypothesis of signalling endosomes, for which there is now extensive evidence, implies that the subcellular location of activated receptor triggers distinct signalling responses [3–8]. Homeostatic receptor trafficking is essential for organism development [9,10], and aberrant activity is implicated in numerous diseases [11,12].

Fluorescence microscopy is commonly used to study ligand induced changes to the quantity of receptor located at the plasma membrane [13], and also colocalization with subcellular structures, such as endosomes [14]. Developing a proper understanding of these experiments requires quantitative, unbiased, and reproducible analysis protocols. In this paper, with these requirements in mind, we describe fully automated image analysis workflows for analysing live cell 3D time-lapse data. Confocal microscopy is used exclusively, but all protocols are equally applicable to deconvolved widefield images [15]. A HeLa cell line expressing fluorescent protein tagged constructs for both epidermal growth factor receptor (EGFR) and rab5 is used as a model system [16]. rab5 is an early endosome associated GTPase and key regulator of receptor trafficking [17]. Inhibitors for EGFR kinase (AG1478) [18] and dynamin (Dynasore) [19] are used to perturb the ligand (EGF) induced trafficking response. We demonstrate the effectiveness of the described workflows, and show that both drug treatments perturb EGFR trafficking and colocalization with rab5 positive endosomes.

The rest of the paper is structured as follows. In Section 2 a brief review of related approaches, identifying both good and bad practice, is presented. Section 3 describes, and shows the use of, the proposed protocols. Finally, Sections 4 and 5 provide a discussion and conclusion.

* Corresponding author at: Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Cambridge CB2 0RE, UK.
E-mail address: jeremy.pike@cruk.cam.ac.uk (J.A. Pike).

http://dx.doi.org/10.1016/j.ymeth.2017.01.005
1046-2023/

2. Related approaches

2.1. Workflows to quantify the subcellular distribution of receptor

Confocal microscopy data can be acquired in either two, or three, spatial dimensions. In a 3D approach multiple axial slices are acquired at different focal planes through the sample. A 3D approach is inherently superior to 2D acquisition as the entire cellular volume can be sampled [5]. In a 2D approach only a single plane through the cell is acquired, hence key information can be missed [13]. If the axial position is not set automatically, for example at a set distance from the coverslip, or at the widest nuclear plane, then user bias is introduced to the acquisition [20]. Note that for live samples light exposure should be kept low. Therefore for time-lapse imaging there is a practical trade-off between the number of axial slices, and the frame-rate.

Post-acquisition, data can be processed to isolate biologically relevant regions of interest (ROIs) such as the plasma membrane. Subsequently, the fluorescence intensity of a specific marker can be quantified within each ROI. Selection of ROIs can either be manual or automated. Manual selection should be avoided as it is prone to user bias and error, and time-consuming and difficult to implement in 3D. With time-lapse data the change in (normalised) intensity, within each ROI, over time can be calculated. For example Fortian and Sorkin (2014) acquired 3D time-lapse data with spinning disk confocal microscopy, and used an automated 3D edge based segmentation protocol to identify the cellular ROI [5]. The segmentation was eroded by a set number of pixels to identify ROIs for the intra-cellular region and plasma membrane. This was used to calculate the normalised percentage, of both EGF and Grb2, associated with the plasma membrane over time. This is an excellent example of an automated 3D strategy for the quantification of temporal changes to the subcellular distribution of a fluorescent construct. However, the pre-processing and segmentation protocols are not fully defined, only the software package and associated components are cited. As the specific image processing algorithms are not referenced, reproduction of this methodology has not been possible in an alternative software application.

ROI intensities, and colocalization measures, can be calculated using either the raw or pre-processed data. Raw data refers to the unprocessed data as acquired by the microscope. There is extensive literature on both image denoising and deconvolution [15,21,22]. These techniques respectively aim to remove corruption and out of focus contributions within image data. Although these approaches can be inaccessible for biological researchers, due to either lack of knowledge or user-friendly tools, working with raw data cannot be considered best practice. However, the use of unjustified or poorly specified methods is worse as results cannot be reproduced. Following the initial pre-processing steps, data can be further processed to enhance, or isolate, biologically meaningful components. Note there is no generalised workflow for image pre-processing and care should be taken to match the approaches used to both the data and the biological context. For example, Dunn et al. (2011) suggest that background subtraction, as calculated with a median filter, is appropriate for the quantification of signal within endosomes [14]. When the width of the filter is at least twice as large as the endosomal structures a reliable estimation of local background is produced.

2.2. Colocalization analysis

Colocalization analysis is typically used to determine if labelled proteins colocalize, or cluster, to the same subcellular structures. High quality analysis relies on high quality data, and particular care must be taken to avoid detector saturation and cross-talk between channels [23]. The spatial sensitivity of colocalization analysis is limited by the resolution limit of the microscope, which is determined by the point spread function (PSF) [24]. According to the Nyquist criterion the pixel size, and the axial spacing, should be less than approximately half this limit to accurately represent the sample at this resolution [25]. However, sampling at this optimal rate may be practically infeasible for live experiments, or large scale screens. When using larger pixels, or axial spacing, artefacts can be introduced and it is the pixel size, not the resolution of the microscope, which limits the spatial sensitivity of the colocalization analysis. For example, consider the imaging of endosomes using a pixel size of 0.25 μm, and axial spacing of 0.5 μm. When using standard imaging wavelengths, and an objective with a numerical aperture of 1.4, this is larger than the size defined by the Nyquist criterion. Two endosomes, can only be distinguished if they are separated by more than approximately 0.5 μm laterally, or 1 μm axially. Therefore colocalization analysis, even using super-resolution techniques, is poorly suited to the identification of direct protein-protein interaction [26]. Techniques such as Förster Resonance Energy Transfer (FRET) are more appropriate for this purpose [27]. Conversely, when there is no direct interaction between the proteins but association within subcellular structures, such as endosomes, FRET cannot be used.

In studies of receptor trafficking, quantitative analysis is often neglected, and colour merges are used to provide qualitative evidence for colocalization [6,28–30]. This can leave interpretation and presentation of results open to user bias, either through the image display settings, or the choice of *representative* images. Visualisation of correlation is better performed using joint-histograms, not colour merges [14]. There are two distinct strategies for colocalization analysis. The first is based on the overlap, or correlation, between pixels [14,23,25]. The second detects objects within the data and uses the centre of mass for each object to determine clustering statistics such as Ripley's K-function [25,26,31]. Object based methods have shown promising results for localization and TIRF microscopy, where the data is well modelled by point, or spot like, objects [26]. However, for the application of receptor trafficking using confocal microscopy, the receptor is typically localised to either the plasma membrane or endosomal structures, the former of which is not well represented by a point distribution. Therefore we will focus on pixel based measures which can be split into two categories; co-occurrence and correlation [32]. Co-occurrence measures quantify how often, or how much, signal from each channel overlaps with the other channel based only on the presence, or absence, of signal. For example, 50% of channel 1 signal overlaps with channel 2 signal. Correlation measures assess the extent of a relationship between the signals from each fluorophore. For example, if there is high positive linear correlation a pixel with high intensity in channel 1 would typically also have high intensity in channel 2. For high negative linear correlation a pixel with high intensity in channel 1 would typically have low intensity in channel 2.

The Manders Coefficients (MCs) (M1 and M2) are well-established co-occurrence measures which simply calculate the percentage of total signal from one channel which overlaps with signal from the other, such that [33],

$$M1 = \sum_i \frac{C1_{i,coloc}}{C1_i} \qquad M2 = \sum_i \frac{C2_{i,coloc}}{C2_i} \tag{2.2.1}$$

where $C1_i$ and $C2_i$ represent the intensities of individual pixels for channels 1 and 2 respectively. $C1_{i,coloc}$ and $C2_{i,coloc}$ represent the colocalizing pixels such that $C1_{i,coloc} = C1_i$ when $C2_i > 0$ and $C1_{i,coloc} = 0$ otherwise. Similarly $C2_{i,coloc} = C2_i$ when $C1_i > 0$ and $C2_{i,coloc} = 0$ otherwise. The Pearson coefficient (PC), R is a well-established measure of linear correlation, defined such that [34],

$$R = \frac{\sum_i (C1_i - C1_{av}) \times (C2_i - C2_{av})}{\sqrt{\sum_i (C1_i - C1_{av})^2 \times \sum_i (C2_i - C2_{av})^2}} \quad (2.2.2)$$

where $C1_{av}$ and $C2_{av}$ are the average intensities for channel 1 and channel 2 signal respectively. In this work we have avoided the use of combined correlation and co-occurrence measures such as the Manders overlap coefficient (distinct from the MCs), and the more recent measure introduced by Singan et al. (2011), as it is advantageous to maintain the ability to distinguish between co-occurrence and correlation [33,35]. The Manders overlap coefficient is also sensitive to small variations in background signal and detector offset [32]. A promising new colocalization measure was introduced by Humpert et al. (2015) which is robust at both low signal to noise ratios (SNR) and varying background levels [36]. It can also be used to simultaneously analyse more than two fluorescent channels. However the interpretation of this measure is not yet clear.

To calculate the MCs it is necessary to isolate the pixels containing biologically relevant signal from both channels. Although signal isolation is not an essential step for the calculation of the PC, Adler et al. (2010) argue that, to prevent artificial inflation, only the pixels containing isolated signal from both channels should be used to calculate the PC [32]. When calculated using only pixels containing isolated signal from both fluorophores, the PC is easily interpreted as a measure of linear correlation within pixels containing probes from both channels [14,26]. This will often have a clear biological interpretation, for example the correlation in endosomes positive for both rab5 and EGFR. Therefore, if both the PC and the MCs are calculated, the level of overlap between the two channels, and the correlation within that overlap can be simultaneously assessed. If all pixels in the ROI, for example a cell, are used to calculate the PC then the extent of linear correlation between the two signals across the whole cell is evaluated. In this case we consider the biological interpretation to be less clear. Signal isolation should be performed with an automated approach to avoid user bias and error. In the popular approach introduced by Costes et al. (2004) a linear fit is found for the joint histogram of the data [37]. The point on this line of best fit below which there is no correlation ($R \leqslant 0$) is used to define the threshold values for both channels. A critical discussion of this approach is presented in Section 4.2.

Colocalization analysis is typically performed to test one of two hypotheses; that the level of colocalization is higher than that predicted for randomly distributed signal (within a ROI), or that there is a difference in the level of colocalization between test conditions. In the former case each ROI can be considered independently, but care must be made to avoid auto-correlation effects (see [26] for a recent review). However, for most biological studies it is more informative to consider if the distribution of measurements taken across biologically independent replicates is significant. To do this McDonald and Dunn (2013) showed, using simulated data, that a distribution of PC measurements can be compared to the expected value of R = 0 using a *t*-test. Similarly, a distribution of either M1, or M2, measurements can be compared to the fractional volume of the ROI occupied by either channel 2, or channel 1, respectively [38]. The difference between the MCs and the expected value can be expressed as,

$$M1_{diff} = M1 - \frac{V_2}{V_{ROI}} \qquad M2_{diff} = M2 - \frac{V_1}{V_{ROI}} \quad (2.1.1)$$

where V_{ROI} is the total volume of the ROI and V_1, V_2 are the volumes of the isolated signal from channel 1 and channel 2 respectively. Note that an accurate segmentation of the ROI containing the signal from both channels is essential for this approach. For the application of receptor trafficking, the ROI should contain the cytoplasm and plasma membrane, but not intracellular structures such as the nucleus. Moreover it is not clear if an expected value of R = 0 is suitable for real data where auto-correlation, or imperfect ROI selection, can lead to inflation of the PC. McDonald and Dunn also showed that a two way *t*-test can be used to test the null-hypothesis that two distributions of either the PC or the MCs, have the same mean. This approach is robust, and also easy to implement, as the effects of ROI selection, signal isolation and auto-correlation will be reproduced for both conditions. Therefore, when possible, we consider it highly desirable to design experiments with a negative control and to test for *changes* in the level of colocalization.

3. Methods and results

In this section the proposed analysis workflows, designed specifically for live cell 3D time-lapse datasets, are presented (Fig. 1). The use of these workflows is demonstrated on confocal microscopy data of cells expressing both EGFR-EGFP, and rab5-mRFP constructs. The cells were treated with either AG1478, Dynasore or a DMSO control. The cell culture and microscopy methodology is described in Appendix A. Unless otherwise stated, the algorithms were implemented in 3D using Matlab (v2015a, The Math-Works, Inc., Natick, MA, USA).[1]

3.1. Data pre-processing and cellular segmentation

Image pre-processing and segmentation are essential components of the analysis. The first step is the manual cropping of each time-series such that the cropped data contains a single cell. This quality control measure was the only manual component of the workflow. This was done blindly and efficiently using a custom-designed interface, where the only criterion for selection is that the cell be alive and non-mitotic. A maximal projection of the first time-point is used to define a region for the cropping of the entire time-lapse.

Each cropped time-lapse was then denoised using an ImageJ plugin implementing the PURE-LET scheme described by Luisier et al. (2010) [39]. ImageJ was run within a Matlab script using MIJ and the plugin was set to automatically estimate noise parameters (4 spin cycles, 3 multi-frame) [40,41]. The PURE-LET scheme is designed for the removal of Poisson noise, it is simple to use, relatively fast, and has been shown to have similar performance to other state of the art methods. Noise with a Poisson distribution is produced in the imaging process by the inherent uncertainty in arrival time of photons at the detector [42]. Briefly, the PURE-LET scheme estimates and minimises the error between the unknown noiseless image and the processed image based on an assumption of Poisson noise.

To identify the cellular volume (segmentation), the denoised EGFR-EGFP channel was processed using a 3D level set segmentation protocol. Specifically, we used a 3D implementation of the edge based distance regularized level set evolution (DRLSE) framework described by Li et al. (2010) (parameters listed in Table A.2) [43,44]. This powerful approach facilitated the segmentation of the cellular boundary using only the EGFR-EGFP signal, where conventional thresholding approaches would fail. The DRLSE term allows for a simple finite difference implementation without the need for re-initialising sub-routines. The computational costs of a level set framework are high so it is advantageous to implement a fast protocol to obtain an initial estimate of the segmentation, and to use this estimate as the starting point for the level set algorithm. To do this an algorithm based on K-means clustering was used (described in Appendix A, example shown in Fig. 2B). After both

[1] Matlab code available at https://github.com/JeremyPike/receptor-trafficking-toolbox

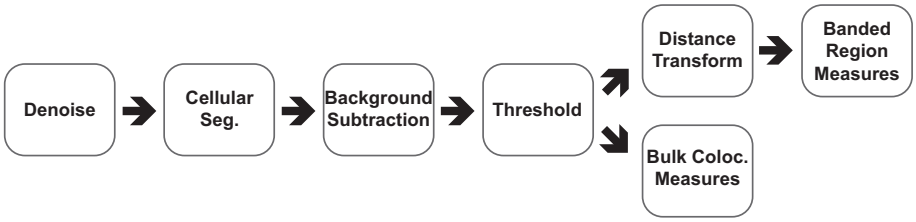

Fig. 1. Flowchart summary of image analysis workflows. The first step is data denoising which is performed to reduce the corruption introduced during image acquisition. This is followed by automated segmentation of the cellular boundary using the membrane bound receptor signal. This is done to produce an accurate region of interest (ROI) for the subsequent analysis protocols. Background subtraction is followed by automated thresholding to isolate signal in both channels. At this point, bulk colocalization statistics can be calculated for each cell, or the cellular ROI can be split into banded volumes using a distance transform. The percentage of receptor (without thresholding), or colocalization measures, can then be calculated for each band. Together these statistics and measures provide a thorough description of the subcellular receptor distribution and colocalization with endosome sub-populations.

the initial and level set based segmentation the largest connected component was selected, and any *holes* were filled.

Ligand treatment triggers receptor internalisation and a decrease in the SNR for the EGFR-EGFP channel at the plasma membrane. Therefore this is a complex segmentation problem. Validation of the segmentation protocol was performed using a reference produced by blind manual segmentation of EGFR-EGFP and transmission images. The SNR at the plasma membrane was sufficient at all time-points to perform manual validation, although a separate membrane stain could have been used if this was not the case. A key limitation of this approach is that manual segmentation will contain bias and not be perfect, but is currently a widely accepted approach to segmentation validation. The Jaccard similarity index, J, was used as a performance measure [45]. The Jaccard index is defined as the intersection divided by the union of the manual segmentation, M, and the automated segmentation, A, such that a perfect match would give $J = 1$;

$$J(M,A) = \frac{|M \cap A|}{|M \cup A|} \tag{3.1.1}$$

The results of this comparison are shown in Fig. 2D. Importantly, a high mean performance ($J = 0.87$) and a statistically significant improvement from the level set method over the K-means based estimate was demonstrated.

The final pre-processing step was background subtraction with a 3D rolling ball approach [46]. In this approach, a background volume was obtained by morphological opening (erosion followed by dilation) of the denoised data using a spherical structure element (1 μm radius). This background was then subtracted from the denoised data to produce background subtracted data. When performing rolling ball background subtraction it is important to set the radius to be at least as large as the width of the largest biologically relevant structures, in this case, endosomes. Note, for the denoising, segmentation and background subtraction steps, each 3D time-point, and each channel, were processed independently.

3.2. Quantifying subcellular receptor distribution over time

In this subsection, a protocol for quantifying the subcellular distribution of receptor over time is presented. The data was first pre-processed, and the cellular ROI was segmented, as described in the previous subsection. Subsequently, the cellular ROI (for each time-point) was split into banded volumes of equal width (0.5 μm), based on Euclidean distance from the segmentation edge (Fig. 2E). This was done using the computationally fast 3D distance transform described by Mishchenko (2015) [47]. Importantly, the transform calibrates for differences in the lateral and axial spacing.

The percentage of the total receptor signal (after pre-processing) contained within each band was calculated to characterise the subcellular distribution at each time-point (Fig. 2F). These measurements were then volume corrected by subtracting the fractional volume of the band. This was done to calibrate for differences in cell size and shape resulting in bands of varying volume (Section 4.1). Note, bands further than 5 μm from the segmentation edge are not shown as only larger cells will exceed this depth. When analysing time-lapse data the change in receptor distribution, for each cell, can be calculated by subtracting the measurements from the first time-point. This is advantageous and justified as it isolates the ligand induced change in receptor distribution, and corrects for cellular variation in the receptor distribution before ligand treatment.

The plots in Fig. 3A show the mean change in percentage EGFR as a function of distance from the segmentation edge across all time-points and conditions. These plots are useful for identifying condition-dependent changes to (ligand induced) receptor trafficking. Fig. 3B shows the initial distribution of receptor at the start of the time-lapse. This can be used to check for any condition dependent variation of receptor distribution before the ligand induced trafficking response. It is informative to perform a statistical analysis to determine if the effects of a specific treatment are significant, relative to the control population. However, when considering distributions of single cell measurements the assumption of a normal distribution is unlikely to be valid, and there will be outliers. Non-parametric tests equivalent to the two-way ANOVA are not well established, so for simplicity we restricted the statistical testing to the change between the first and final time-points (Fig. 3C). For each cell, the mean (absolute) percentage EGFR change across all bands (up to 5 μm) was calculated. This characterises the total change in subcellular receptor distribution 30 min post ligand treatment. Using this measure, a non-parametric Kruskal-Wallis analysis of variance was performed. If significant ($p < 0.05$), post hoc-testing using a Mann-Whitney U test, with Bonferroni correction for multiple hypothesis testing ($n = 2$), was performed for each treatment relative to the control. With this approach, a statistically significant reduction in the magnitude of the trafficking response, for both the AG1478 and Dynasore treatments, was identified.

3.3. Signal isolation for colocalization analysis

In the previous subsection, the EGFR-EGFP signal was used to characterise changes in the subcellular distribution of receptor. In this section, we suggest and validate a strategy for thresholding both the EGFR and rab5 channels. It is necessary to threshold the data to calculate the Manders coefficients and good practice for the Pearson coefficient (Section 2.2) [32]. After data denoising and background subtraction, as described in Section 3.1, an automated global thresholding approach is typically sufficient. Specifically, we used the Otsu approach where it is assumed that the data histogram is bi-modal, consisting of background and signal peaks

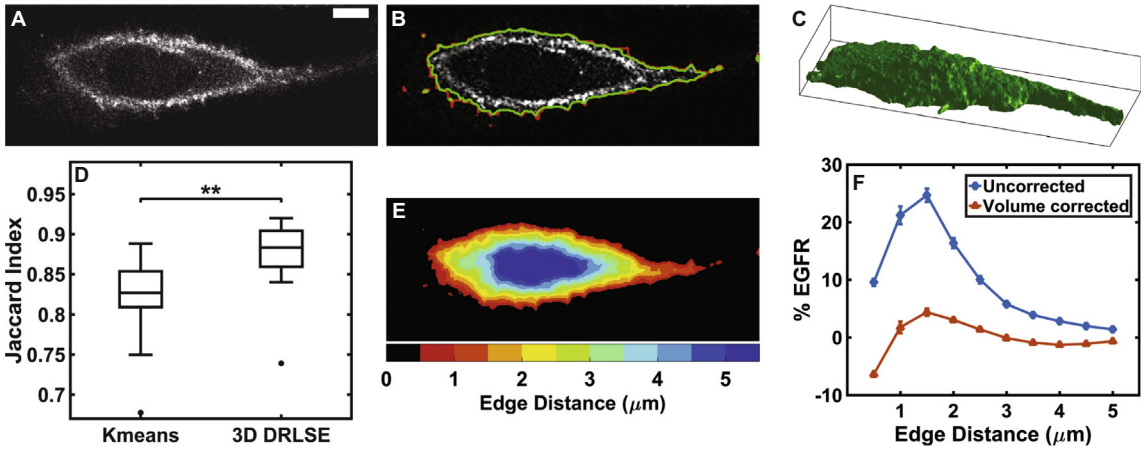

Fig. 2. Pre-processing, cellular segmentation and band based analysis of 3D time-lapse data. (A) Representative raw EGFR-EGFP image slice from a 3D stack. Scale bar set at 5 μm and contrast enhanced such that the display range is between zero and half the maximum pixel intensity. (B) Pre-processed data after denoising and background subtraction. The results from the K-means based (red) and level set (green) segmentation protocols are shown. (C) Surface rendering of the level set segmentation result for a single time-point. (D) The segmentation performance of both the K-means estimation and level set algorithm (3D DRLSE) was quantified using the Jaccard index. This was done using 10 datasets where 14 evenly spaced slices per time-point were manually segmented. The mean Jaccard index for the K-means and level set protocols were 0.82 and 0.87 respectively. A sign test was used to determine that the level set protocol produced a significant increase in performance ($p = 0.002$). Central mark on boxplot represents the median, and the edges of the box are the 25th and 75th percentiles. (E) The cellular ROI is split into banded volumes based on distance from the segmentation boundary. Each band has a width of 0.5 μm. (F) Uncorrected plot shows percentage of total cellular EGFR signal contained within each band. For the volume corrected plot, the fractional volume of each band has been subtracted. This was calculated using data from the control population for the first time-point (immediately after EGF treatment). Error bars given by the SEM ($n = 12$). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

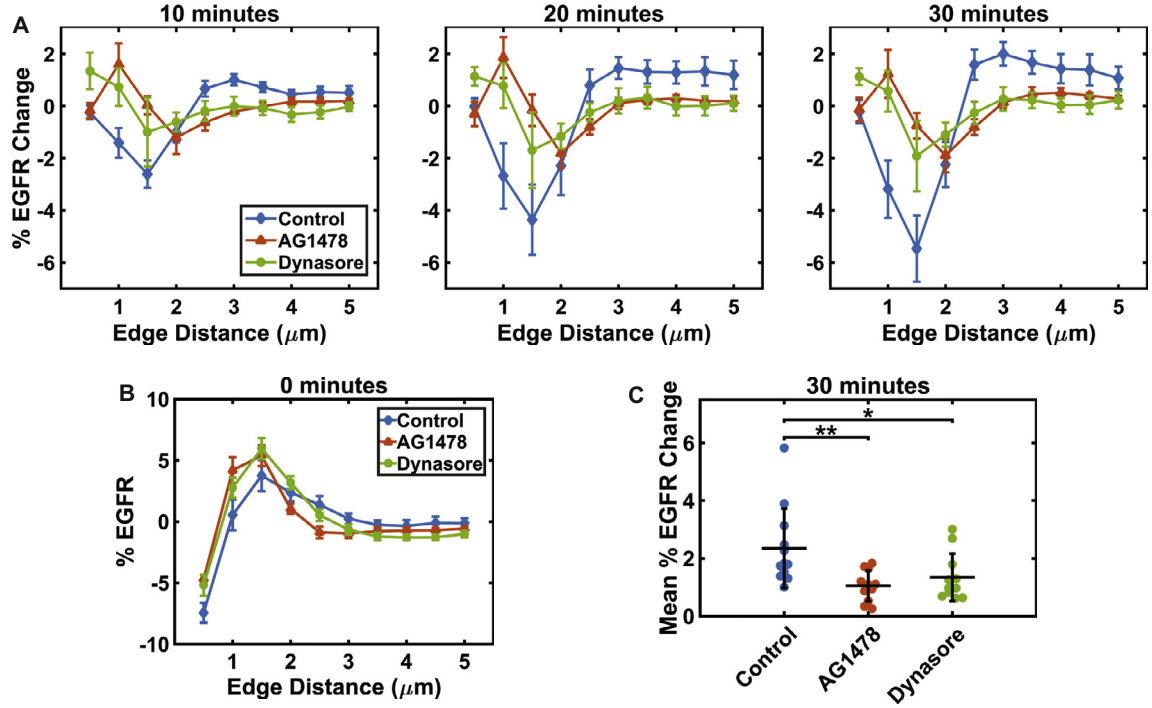

Fig. 3. Band based spatio-temporal analysis of subcellular receptor distribution. (A) Plots of the volume corrected change in percentage EGFR signal over time, for each banded volume (up to 5 μm), in response to EGF treatment. The data for the DMSO control and the two drug treatments are shown (AG1478 and Dynasore). Error bars given by the SEM and $n > 10$ for all treatments. (B) Plot showing the volume corrected subcellular distribution of EGFR for the first time-point. (C) Plot of mean (absolute) change in percentage EGFR signal (volume corrected) between the first and final time-points (30 min). A Kruskal-Wallis one-way analysis of variance returned $p = 0.005$, indicating that there are statistically significant differences between treatments. Post-hoc testing of both drug treatments relative to the control by the Mann-Whitney *U* test (corrected by the Bonferroni method), returned $p = 0.007$, $p = 0.03$ for the AG1478 and Dynasore treatments respectively. The central band represents the mean, and the error bars are the standard deviation.

[48]. The threshold value is defined such that the intra-class variance between the two peaks is minimised. This value best separates the signal and background components of the data. Only pixels contained within the cellular ROIs were included in calculations, and by combining the data from all time-points, a single threshold value was found for each cell.

To validate this approach on real data, threshold values were manually set for 15 cells at a single time-point. This was done

blindly, using a script which randomly selected cells, and time-points, from the entire data-set. Manual thresholds were subsequently set using Fiji [49]. The threshold values were chosen to isolate both the membrane and endosomal associated signal. The resulting binary data was compared to that generated by automated Otsu thresholding using the Jaccard index (Eq. (3.1.1)). The approach proposed by Costes et al. (2004) was also tested (Section 2.2) [37]. Both approaches demonstrated a comparable mean performance for the EGFR channel, but the performance of the Otsu approach was significantly higher for the rab5 channel (Fig. 4).

A comparison of the Otsu and Costes methods was also performed on synthetic data. 3D two channel image stacks of spots (with Gaussian profile to approximate point spread function) were generated in the open image analysis platform, Icy, using a mixed Poisson-Gaussian noise model (Fig. 5A) (detailed in Appendix A) [50,51]. Synthetic data was generated for low, medium and high levels of noise. The level of colocalization was varied from 100% colocalized (spot overlap) to 100% anti-colocalized (spot avoidance). Spots not specified as colocalized (or anti-colocalized) were distributed randomly. 3D Gaussian filtering (width = 1 pixel) was used as a simple pre-processing step. Threshold values were then calculated using either a Costes or Otsu approach on both the raw and pre-processed data. The MCs ($M1$ and $M2$) were calculated before subtracting the expected value to obtain $M1_{diff}$ and $M2_{diff}$ (Eq. (2.1.1)). Fig. 5B shows the rate of failure for the Costes approach in the low noise test data. Failure is defined as extreme over segmentation resulting in 100% signal overlap and a MC equal to one. The Costes approach has non-zero fail rate when there is either no colocalization, or anti-colocalization, indicating that it is not appropriate under these conditions. The Otsu approach does not fail for any tested condition. Fig. 5C shows $M1_{diff}$ for all levels of colocalization and noise. $M1_{diff}$ is shown only where the failure rate is zero. In all cases, pre-processing increases the performance. Otsu thresholding (with pre-processing) outperforms the Costes approach across all noise levels. The implications of this analysis are discussed in more detail in Section 4.2.

3.4. Quantitative colocalization analysis for 3D time-lapse data

Section 3.1 describes the cellular segmentation and pre-processing steps of the proposed workflows. Section 3.3 introduces a strategy for signal isolation. Using only the isolated and pre-processed signal, the Manders (M1 and M2) and Pearson (R) coefficients were calculated for all time-points. To assess the ligand induced change in colocalization over time, the change in all coefficients was calculated, for each cell, by subtracting the measurement from the first time-point. Fig. 6A–C show plots for each of the coefficients across all time-points, and for all conditions.

To identify statistically significant differences in the colocalization response for a specific condition relative to the control, the colocalization coefficients were processed as conventional measurement variables. As in Section 3.2 we restrict our statistical analysis to the change between the first and final time-points. This characterises the change in either co-occurrence, or correlation, for the Manders and Pearson coefficients. A Kruskal-Wallis analysis of variance with post hoc testing of each treatment relative to the control (Mann-Whitney U test with Bonferroni correction) was then used to identify statistically significant changes in the colocalization response. Fig. 6D– F show the results of this analysis for the Manders and Pearson coefficients. A statistically significant decrease in the level of EGF induced co-occurrence, as measured by the Manders coefficients, was shown for both the AG1478, and Dynasore treatments. No significant change in correlation, as measured by the Pearson coefficient, was established for either treatment.

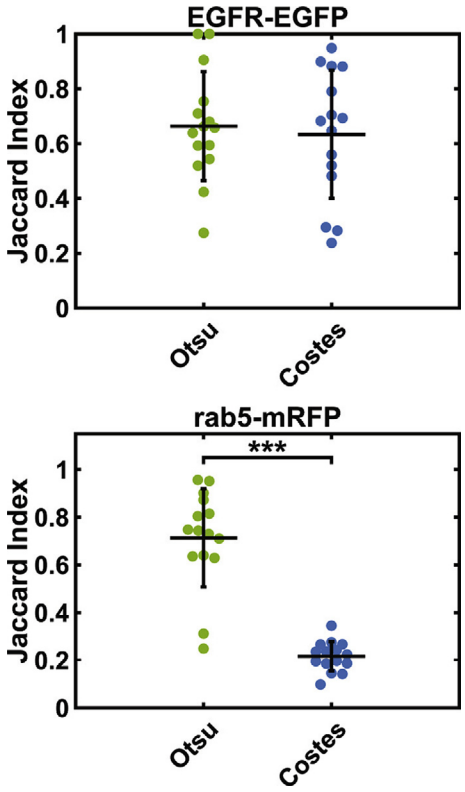

Fig. 4. Comparison of the Otsu and Costes thresholding approaches for signal isolation using real data. Before calculating the threshold values, denoising and background subtraction was performed. Both methods were compared to manually set threshold levels from 15 randomly selected cells and time-points. This was done by calculating the Jaccard index, J, using the resulting binary images. For the EGFR-EGFP channel, the mean performance was $J = 0.66$ and $J = 0.63$, for the Otsu and Costes methods respectively. A sign test returned $p = 0.6$ indicating that no significant difference in performance was detected. For the rab5-mRFP channel, the mean performance was $J = 0.71$ and $J = 0.22$ for the Otsu and Costes methods respectively. A sign test returned $p = 9.8 \times 10^{-4}$ indicating that the Otsu method performed significantly better than the Costes approach.

3.5. Visualising co-occurrence in 3D

In this subsection an unbiased strategy for the visualisation of colocalization is presented. The proposed method produces a spatial map of co-occurrence. For a specific cell and time-point, the relative contribution of each pixel to one of the Manders coefficients (either M1 or M2) is used to determine the intensity of the co-occurrence map. For M1 this is the isolated signal from channel 1, which overlaps with isolated signal in channel 2. This 3D map is then scaled between the maximum and minimum display values, and normalised by multiplying by M1 (or M2) for the given time-point. This normalisation allows for unbiased visual comparison between multiple cells.

The co-occurrence maps (and cellular ROIs) can be visualised interactively in 3D (Fig. 7A). To do this, the Matlab vol3d v2 script[2] was used, but any software capable of 3D data rendering could be used for this purpose. If 2D visualisation is required then either the maximal projection, or single slices, of the map can be shown. In Fig. 7B, maximal projections of M1 co-occurrence maps are shown using a false colour lookup table (LUT) for representative cells from the control, AG1478 and Dynasore treatments (20 min post EGF treatment). Finally, recall that it is informative to include joint-

[2] available at http://uk.mathworks.com/matlabcentral/fileexchange/22940-vol3d-v2, accessed 26/06/16

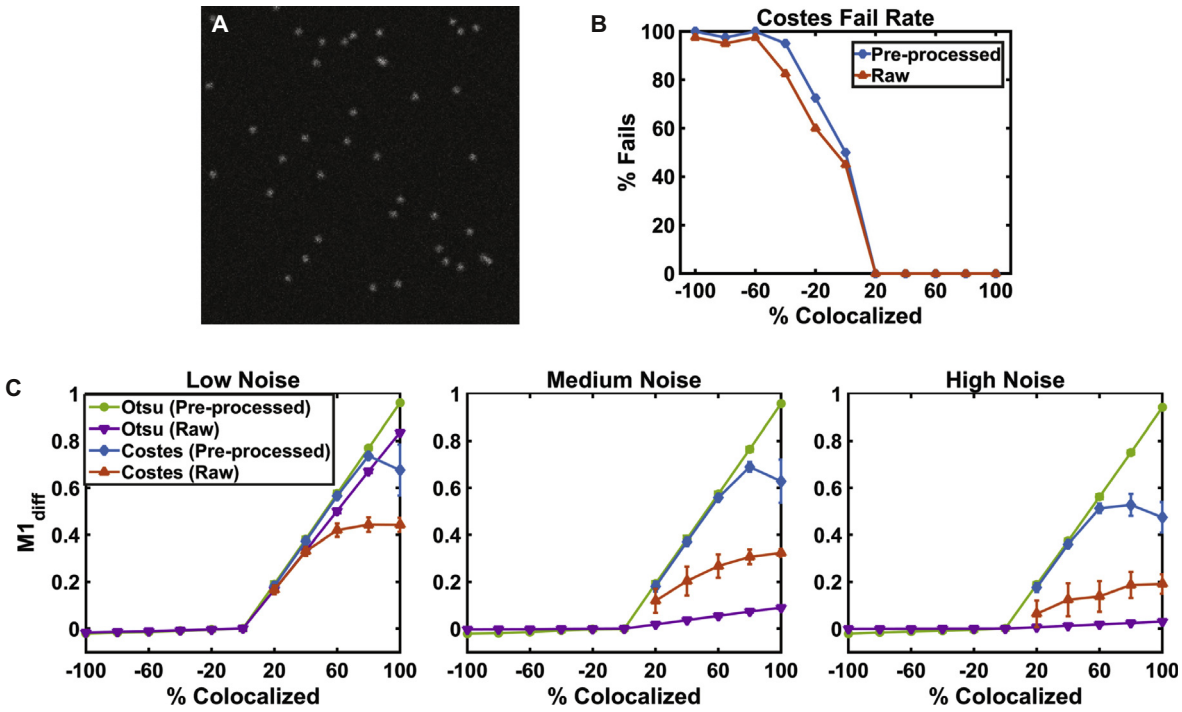

Fig. 5. Comparison of the Otsu and Costes thresholding approaches for signal isolation using synthetic data. (A) 3D test data of *spots* was generated for varying levels of colocalization (and anti-colocalization), and for three levels of noise. 50 stacks were generated per condition. A single slice from the medium noise level is shown. (B) In some instances the Costes approach fails to find appropriate threshold values resulting in Manders coefficients equal to one. The percentage of stacks for which the Costes approach fails is shown for each level of colocalization (low noise). (C) $M1_{diff}$ is calculated by subtracting the expected value from the Manders Coefficient M1. This is shown for all methods, levels of colocalization and noise levels. Note $M1_{diff}$ is shown only when the failure rate is zero.

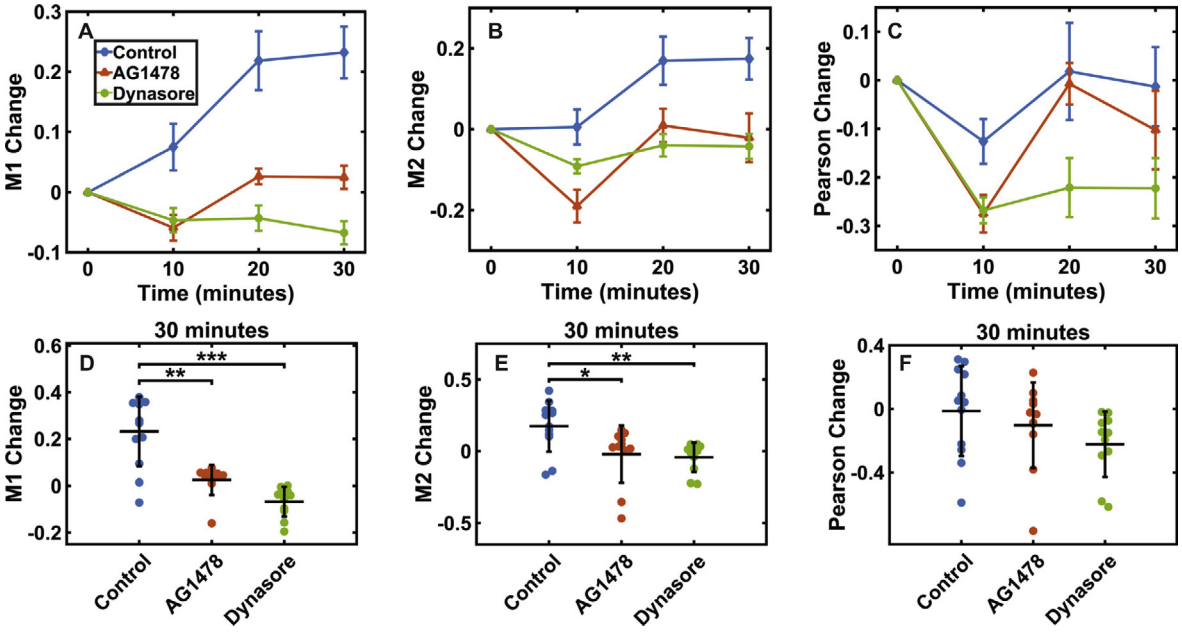

Fig. 6. Quantifying colocalization with 3D time-lapse data. (A–C) Plots of change in either the Manders (M1 and M2), or Pearson coefficients. Measurements for a DMSO control, AG1478 and Dynasore treatments are shown. These plots characterise the change in co-occurrence (Manders) or correlation (Pearson) between EGFR-EGFP and rab5-mRFP, in response to EGF treatment. Error bars are given by the SEM and n > 10 for all treatments. (D, E) Plots of the change between the first and final time-points for the Manders and Pearson coefficients. Central band represents the mean, and the error bars are the standard deviation. A Kruskal-Wallis one-way analysis of variance returned, $p = 4 \times 10^{-5}$, p = 0.002, p = 0.08 for M1, M2 and the Pearson coefficient respectively. This indicates that there are statistically significant differences between treatments for both M1 and M2. (D) For M1 post-hoc testing of both drug treatments, relative to the control, by the Mann-Whitney *U* test (corrected by the Bonferroni method) returned p = 0.008, $p = 6 \times 10^{-4}$ for the AG1478 and Dynasore treatments respectively. (E) For M2 Post-hoc testing returned p = 0.02, p = 0.007 for the AG1478 and Dynasore treatments respectively. (F) For the Pearson coefficient the analysis of variance was not significant so no post-hoc testing was performed.

histograms as a visualisation of correlation [14]. Fig. 7 C shows the corresponding joint-histograms where only pixels within the cellular ROI have been included.

3.6. Quantifying subcellular colocalization distribution over time

In Section 3.2 a method was proposed which split the cellular segmentation into banded regions using a 3D distance transform. In Sections 3.2–3.5 workflows for quantifying and visualising colocalization were presented. Here these two strategies are merged to quantify the subcellular distribution of colocalization. We propose two approaches. In the first we consider the signal contained within each band individually and calculate both the Manders coefficients and the Pearson coefficient. This quantifies the level of co-occurrence and correlation for each band. In this approach, if we compare the measurements from two different bands, we cannot say if there is more, or less, co-occurring signal, only that a greater or smaller percentage of that band's signal is co-occurring. In the second approach, the percentage of signal from the co-occurrence map (Section 3.5) contained within each band is calculated and volume corrected. This approach follows the same protocol as that described for the EGFR signal in Section 3.2 and quantifies where the co-occurring signal is located. When comparing two bands we can say that there is more co-occurrence in one band but we cannot state what percentage of the band's signal is co-occurring, or comment on correlation.

The results from the first approach are shown in Fig. 8. Fig. 8A shows the change in M1 with respect to the first time-point across all banded regions up to 5 μm from the cellular segmentation. This characterises the ligand induced changes in co-occurrence on a subcellular level. Note, in the control population, the dominant increase in co-occurrence occurs less than 3 μm from the edge. This is as expected since we are quantifying colocalization between receptor and early endosomes. Fig. 8B shows the same analysis for the change in the Pearson coefficient. Fig. 8C and D show M1 and the Pearson coefficient at the first time-point. Fig. 9 shows the results from the second approach. Note from Fig. 9A that there is no clear change in how the co-occurring signal is distributed in response to ligand. Fig. 9B shows the distribution of the co-occurrence map at the first time-point. Fig. 9C shows the mean (absolute) change in percentage signal between the first and final time-point across all bands up to 5 μm. No statistically significant change was found.

4. Discussion

4.1. Band based analysis of subcellular receptor distribution

In the band-based approach introduced in Section 3.2, the cellular ROI is split into volumetric regions based on distance from the segmentation edge. The relative distribution of receptor signal across the banded volumes can then be calculated. This approach facilitates the quantification of both the surface (membrane associated) and intracellular signal. Importantly, the analysis is automated and performed in 3D, hence avoiding any user bias. The level set based cellular segmentation protocol has high performance when using the receptor signal as the input (Fig. 2D). This is advantageous for live cell microscopy as the use of a secondary marker for the plasma membrane, or cytoplasm, would increase sample exposure and experimental complexity. This increase in light exposure, and associated photo-toxicity, would reduce the sampling and/or frame-rate achievable.

The measurement for each band was volume corrected by subtracting the expected value assuming a homogenous distribution of signal, specifically the fractional volume of each band. This is justified because cells have varying size and shape which in turn, change over time. By subtracting the expected value, we calibrate for varying band volume which facilitates direct comparison between cells over time. The bands were chosen to be of equal width (0.5 μm) for ease of interpretation. In this approach the number of bands per cell will vary and therefore we chose to restrict plots to a maximum of 5 μm from the segmentation edge. An alternative approach would be to fix the number of bands but vary the band width between cells. This would allow for quantitative comparison of the entire cellular volume, but the interpretation would be less clear.

In the presented workflows, the nucleus is not excluded from the analysis ROI. Due to cellular size and shape variation, the proportion of nuclear region contained in a specific band will vary. Since the nuclear region contains less receptor, and no endosomes, this will introduce noise to bands in the cellular interior. Therefore using a nuclear stain to segment the nucleus would be advantageous. However, as with the addition of a membrane stain, the use and imaging of a nuclear marker would require additional light exposure. If a nuclear stain is used, the band based analysis could be repeated within the nuclear region, or alternatively extending outwards from the nucleus to the plasma membrane. Although this is not particularly relevant for the EGFR and rab5 data presented, this is an excellent example of how the proposed workflows could be adapted for a variety of applications.

When using live-cell time-lapse data, the change in a given measurement can be isolated at a single cell level, by subtracting the measurement value at the first time-point. This is advantageous as it corrects for cellular variation in the receptor distribution prior to ligand stimulation. Therefore the ligand induced change in receptor distribution is isolated (Fig. 3A). This approach can also be applied to colocalization measurements as shown in Figs. 6A, 8A, B and 9A.

The mean absolute change in percentage signal (between the first and last time-points) across all bands provides a single measurement per cell. This quantifies the magnitude of the, ligand induced, trafficking response (Fig. 3C). This summary measurement is useful as non-parametric hypothesis testing can be employed to identify statistically significant changes to the trafficking response between two conditions (typically a control and treatment) (Fig. 3C). However, information is lost by the reduction of the analysis to a single measurement per cell.

In summary, the proposed band-based analysis provides researchers with a powerful tool to identify, and quantify, perturbations to receptor trafficking in live cells. Moreover this approach to quantifying the subcellular distribution of a fluorophore is adaptable and could be applied to applications outside this field.

4.2. A critical review of Costes thresholding

Costes thresholding is an automated strategy for signal isolation in colocalization analysis (Section 2.2) [37]. Briefly, the linear line of best fit for the joint-histogram of the data is found, and the point on this line below which the data has a Pearson coefficient of less than zero is used to define global threshold values. This strategy is cited in review articles as good practice [14,25], and implemented in many popular image processing applications including Fiji and Imaris (Bitplane AG, Zurich, Switzerland) [49]. Adler and Parmryd (2013) note that a Costes approach will fail if there is no correlation in the data, and Dunn et al. (2011) suggest the approach may fail if the SNR is too low, the labelling density is too high, or if there are too many structures in one channel [14,23]. We also note that with a Costes approach there is an assumption that the data is well represented by a single linear fit, and therefore has strong correlation. In Fig. 5 B we show that a Costes approach has non-zero fail rate when tested on simulated data when there is either no colocaliza-

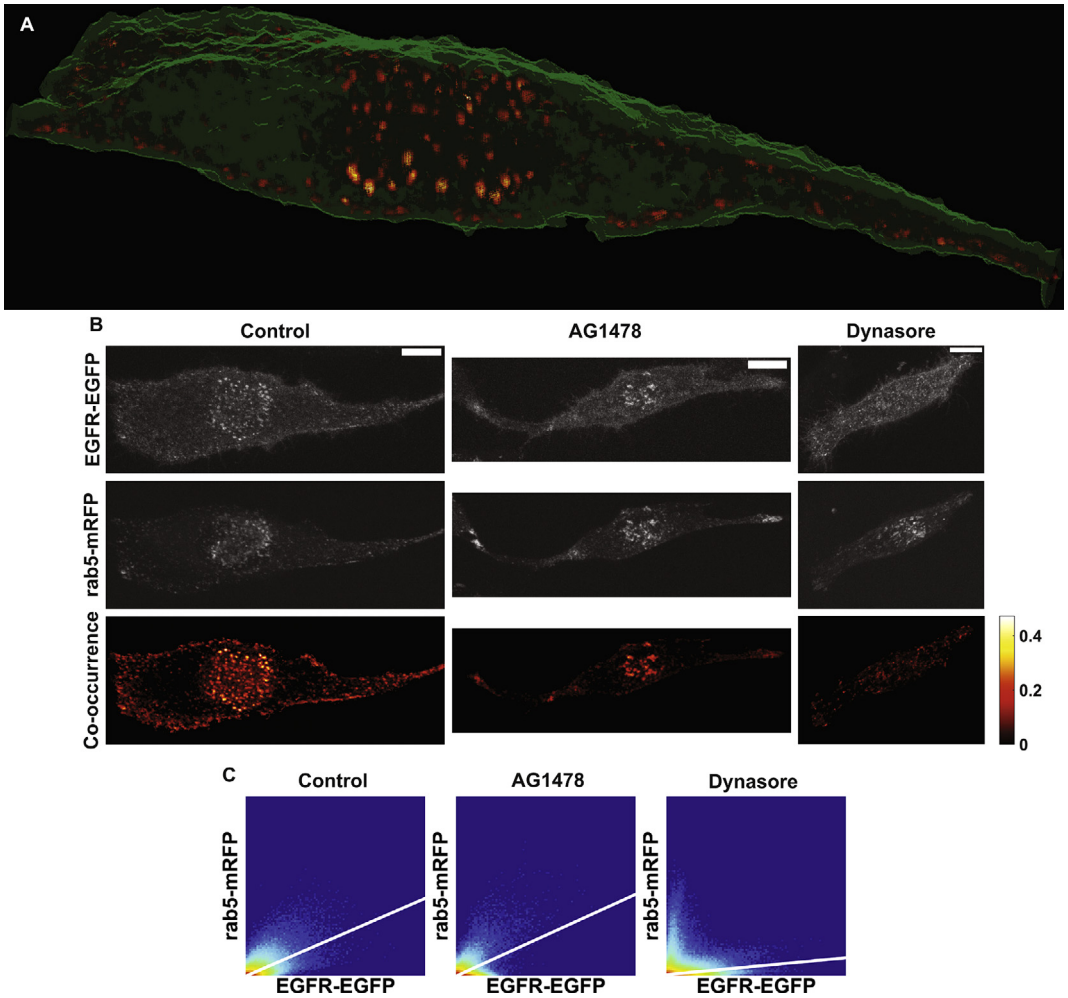

Fig. 7. Unbiased visualisation of colocalization. (A) Snapshot of a M1 based map of co-occurrence which has been visualised by 3D rendering. The surface of the cellular ROI is shown in green. (B) Maximal projections of representative cells (20 min post EGF treatment) for the control, AG1478 and Dynasore treatments respectively. Raw unprocessed data for the EGFR-EGFP and rab5-mRFP channels are shown in the first two rows respectively. The third row shows the normalised M1 based map of co-occurrence. Each map has been scaled by the corresponding M1 calculation to facilitate visual comparison. Scale bar set at 5 μm. (C) Joint-histograms for the corresponding cells and time-point (logarithmic scale used). Only pixels within the cellular ROI were included. The line of best fit is shown in white. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

tion, or anti-colocalization. This analysis was performed on a very simple test data set, where the intensities of spots in both channels were non-varying and equal (Appendix A). The Costes approach fails under these conditions as the gradient of the line of best fit can by less than or equal to zero. Therefore we conclude that a Costes approach is appropriate only if it can be assumed before the study that well correlated colocalization is present. Since the Costes approach is used to calculate measures of colocalization this should clearly not be assumed in the majority of cases.

Consider the specific case of receptor colocalization with endosomal sub-populations. In such a case there are three key structures; membrane bound receptor, receptor positive endosomes and receptor negative endosomes. Each of the structures will have very different levels of correlation. Therefore the joint histogram representation will not be well represented by a single linear fit, and a Costes approach is not applicable (Fig. 10). Moreover it should not be assumed that there will be correlation between receptor and early endosomes in the drug treated populations or for all time-points. Note from Fig. 4 that the Costes approach fails to reliably isolate signal in the rab5 channel, when compared to manually set threshold values.

We consider well defined pre-processing of data, followed by global thresholding strategies, such as Otsu or minimum cross entropy thresholding, to be a more appropriate strategy for signal isolation in colocalization studies [48,52]. Fig. 4 demonstrates that Otsu thresholding outperforms Costes thresholding on the real dataset. For the simulated data, Fig. 5C demonstrates that when calculated on data pre-processed with a Gaussian filter, the Otsu approach outperforms the Costes approach for all tested levels of colocalization and noise parameters. The Otsu approach performs substantially better at higher noise levels when performed on pre-processed data. This emphasises the importance of the pre-processing steps. Finally note, we are not proposing that the Otsu method will be appropriate for all applications and datasets. Our conclusion is that custom design and testing of pre-processing and thresholding steps is necessary to reliably segment biologically relevant signal for colocalization analysis. Importantly the developed strategy should be automated and applied consistently to all conditions to avoid user bias and variability.

4.3. Quantitative colocalization analysis

In Section 3.3, a workflow was presented to quantify colocalization for 3D time-lapse data. Both the Manders (M1 and M2) and Pearson coefficients are calculated (Fig. 4A–C). Note the Pearson coefficient is calculated only across pixels containing isolated sig-

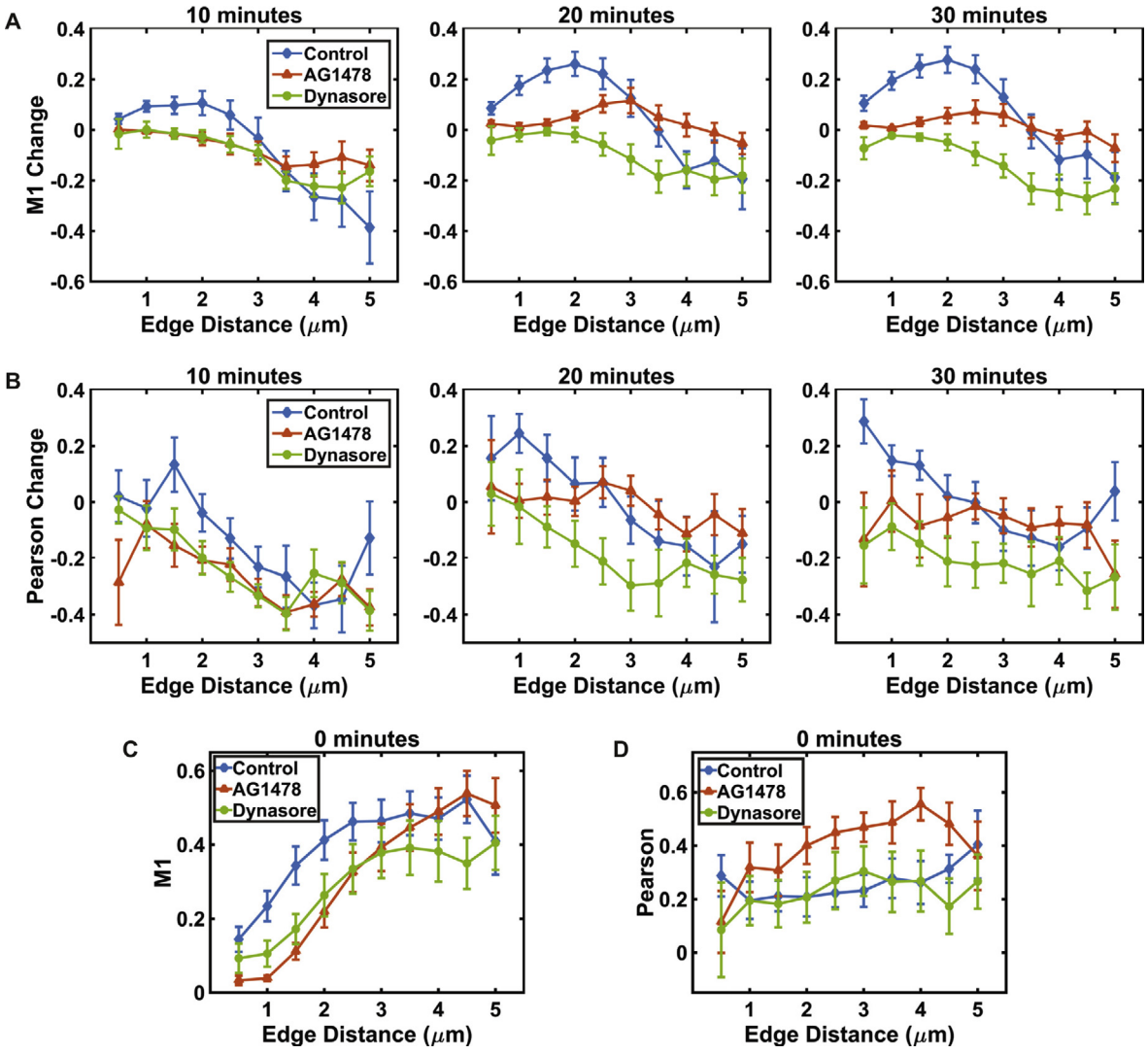

Fig. 8. Band based analysis using colocalization measures. (A) Plot of the change in the first Manders coefficient, M1, over time, for each banded volume (up to 5 μm), in response to EGF treatment. The data for the DMSO control and the two drug treatments are shown (AG1478 and Dynasore). Error bars given by the SEM and n > 10 for all treatments. (B) Plot of the charge in the first Pearson coefficient, R, over time, for each banded volume, in response to EGF treatment. (C, D) Plots of M1 and the Pearson coefficient for each banded volume at the first time-point.

nal in both channels. Recall this pairing of colocalization measures is very informative, and has a clear interpretation; the Manders coefficients quantify signal co-occurrence, and the Pearson coefficient quantifies the correlation within the co-occurring signal. For the test data used in this study, M1 represents the percentage of EGFR signal overlapping with rab5 positive endosomes, and is therefore the more relevant of the two Manders coefficients.

Non-parametric statistical testing can identify statistically significant differences in either co-occurrence, or correlation, between conditions (Fig. 4D–F). In this approach the statistical analysis is performed across multiple, as opposed to individual, ROIs which is advantageous as population based conclusions are essential. We consider the null hypothesis that the ligand induced change in co-occurrence, or correlation, is the same in the control and treated samples. The null hypothesis that the signal is randomly distributed within the cellular ROI is not considered. There are two reasons for this; firstly we consider rejecting the former null-hypothesis to be more informative and useful for the study of receptor trafficking. Secondly, the technical complications of auto-correlation and accurate ROI detection must be considered to reject the latter null-hypothesis. For example the nuclear region should be isolated and removed from calculations. The proposed approach is therefore robust, unbiased and comparatively simple to implement.

In Section 3.6 the band based analysis from Section 3.2 is used to characterise the subcellular distribution of colocalization. Two distinct approaches were taken. Firstly, the level of co-occurrence and correlation was calculated for each band independently (Fig. 8). Secondly, the distribution of total cellular co-occurrence was characterised (Fig. 9). These analyses were included to demonstrate how the workflows presented in this paper can be adapted and combined for new applications. To clarify the difference between these two approaches consider a simple example where a cell is split into just two bands, with equal volume. The outer band contains 90% of the signal from channel 1 which has 50% overlap with channel 2 signal. The inner band only has 10% of channel 1 signal but 100% overlap with channel 2 signal. The first approach would return M1 = 0.5 and M1 = 1 for the outer and inner bands respectively, but this could be misleading as 82% of the co-occurring signal is contained in the outer band. This distribution of co-occurring signal is calculated by the second approach.

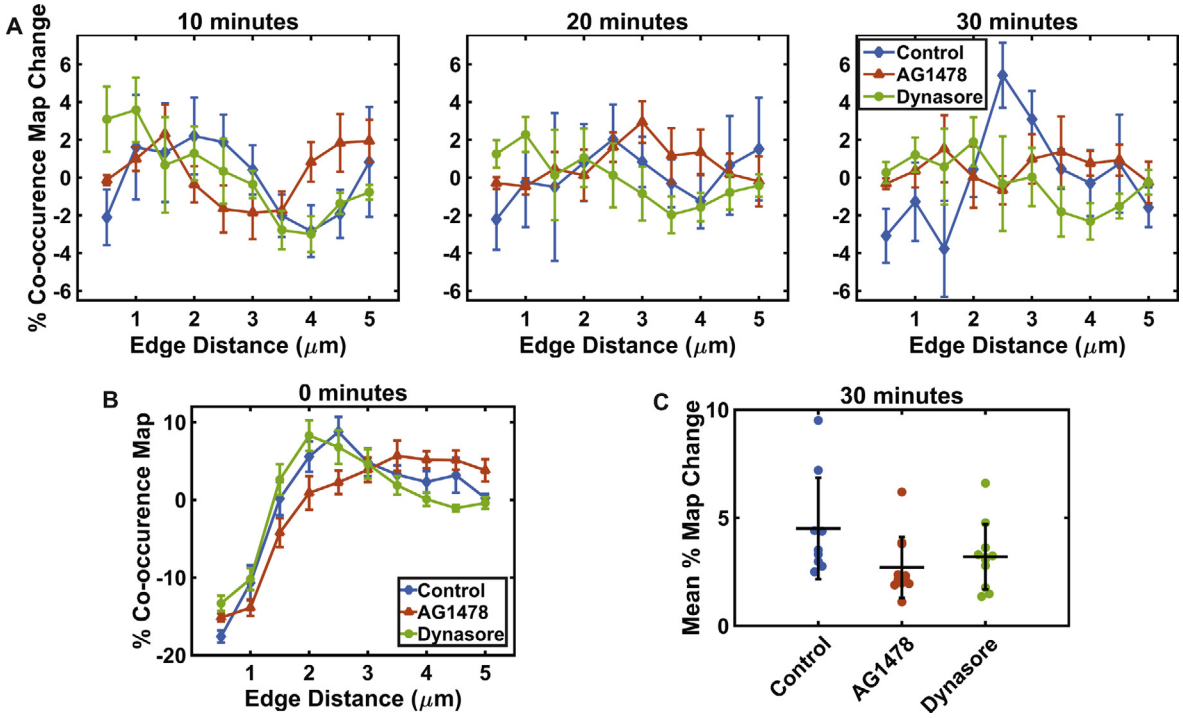

Fig. 9. Band based spatio-temporal analysis of subcellular co-occurrence distribution. (A) Plots of the change in (volume corrected) percentage co-occurrence map signal over time, for each banded volume (up to 5 μm), in response to EGF treatment. The data for the DMSO control and the two drug treatments are shown (AG1478 and Dynasore). Error bars given by the SEM and n > 10 for all treatments. (B) Plot showing the (volume corrected) subcellular distribution of the co-occurrence map for the first time-point. (C) Plot of mean (absolute) change in (volume corrected) percentage co-occurrence map between the first and final time-points (30 min). A Kruskal-Wallis one-way analysis of variance returned p = 0.07, indicating that there are no statistically significant differences between treatments. Therefore no post-hoc testing was performed.

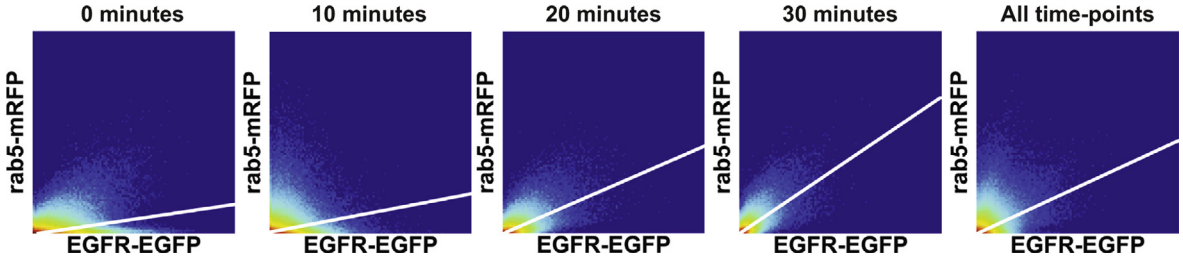

Fig. 10. Receptor and endosomal two channel data is not well represented by a single linear fit. Joint histograms (logarithmic scale) for a representative cell expressing EGFR-EGFP and rab5-mRFP. Immediately prior to imaging the cell was treated with EGF. Only pixels within the cellular ROI were included in the plots. The line of best fit for each plot is shown in white. Note the gradient of the linear fit changes over time as the receptor is internalised. Therefore, if a Costes thresholding approach is used the ratio between the two threshold values will also change. Also note that for the first two time-points, where the majority of the receptor is membrane localised, the data is weakly correlated. This is also the case if all time-points are combined to produce a single joint-histogram.

4.4. Unbiased visualisation of colocalization

Colocalization is typically visualised using colour overlays (Section 2.2). This approach is open to bias through the choice of display parameters and contrast. Joint-histograms provide an effective visualisation of correlation, but do not preserve spatial information. In Section 3.4 we introduced a spatial map for the visualisation of co-occurrence. Importantly, this allows for unbiased visualisation of the colocalizing signal, where the spatial information is preserved. Using 3D rendering techniques, researchers can interactively view the co-occurrence in 3D (Fig. 5A). However 2D images are often required, and as an alternative to colour overlays, a maximal projection (or single slice) of the co-occurrence map can be used. The use of maximal projections can be misleading. If there is more than one pixel with co-occurring signal along the projection axis then only the maximal value will be displayed. Therefore visualisation in 3D is preferable. Note that these visualisation strategies should be performed in parallel with, and not as an alternative to, quantitative analysis (Section 3.4). Finally note that, as with all visualisation strategies, the choice of the representative cell can bias the interpretation, emphasising the importance of quantitative analysis of a population.

5. Conclusion

All of the workflows introduced in this study are unbiased and described in sufficient detail as to be reproducible. These two key requirements should always be fulfilled by an image analysis protocol. Custom image analysis solutions are typically required for the analysis of subcellular signal distribution and colocalization. Therefore the aim of this work was not to provide step by step protocols, which researchers should follow exactly. Instead, illustrative examples were used to demonstrate the implementation of custom workflows which should be adapted by researches for different datasets. The described protocols are specifically designed

for 3D live cell time-lapse data, and the statistical analysis is constructed to identify differences between treated and control samples. Such an approach is robust and ideal for confirmatory studies from larger screens. For example, the workflows presented in this work could be used to validate and further investigate hits from a RNAi screen for regulators of endocytosis and trafficking [53,54]. Finally, we note that the workflows could easily be adapted for applications other than receptor trafficking.

Acknowledgments

This work was supported by the EPSRC through a studentship to JP from the PSIBS Doctoral Training Centre (EP/F50053X/1). The authors also acknowledge the Birmingham Advanced Light Microscopy (BALM) facility and Abdullah Khan. Data supporting this research is openly available from the University of Birmingham data archive at http://findit.bham.ac.uk/.

Appendix A. Supplementary materials and methods

A.1. Cell culture and transfection

The HeLa cell line was cultured in DMEM (with 4.5 g/l L-glutamine) (Invitrogen) supplemented with 10% FBS (Biosera) and 1% Penicillin-Streptomycin solution (Invitrogen) at 37 °C and 5% CO_2. The EGFR-EGFP construct was a gift from Alexander Sorkin (University of Pittsburgh, USA) and the rab5-mRFP construct was a gift from Ari Helenius (Institute of Biochemistry, ETH Zurich). 24 h prior to transfection, 2×10^6 cells/well were seeded into 35 mm glass bottomed dishes (MaTek) with antibiotic free media. Transient dual transfection of both constructs was performed using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions (10 µl of Lipofectamine 2000 and 2 µg of both constructs were used per well). 3 h after the addition of the constructs the media was changed (antibiotic free). Cells were imaged 24 h after expression.

A.2. Drug treatment and confocal microscopy

Prior to imaging, cells were washed with 1 ml DPBS (Lonza) and serum starved for 30 min in serum/antibiotic free media (37 °C and 5% CO_2). Media was then replaced with 2.5 ml cell imaging media (10 mM HEPES–Hank's balanced salt solution (Sigma), pH 7.4) and cells were treated with either 5 µg/ml AG1478 (Sigma) (1 µl/ml DMSO:MeOH 1:1 vehicle), 80 µM Dynasore hydrate (Sellek Chemicals) (1 µl/ml DMSO vehicle) or 1 µl/ml DMSO (Sigma). Cells were then incubated for a further 30 min before transfer to the microscope where cells were treated with 2 ng/ml EGF (Bachem) immediately prior to initiating the time-lapse. Confocal laser scanning microscopy was performed with an inverted microscope (Eclipse Ti, Nikon A1R) at 37 °C and 5% CO_2 using a 1.49 NA 100X NA oil-immersion objective. Note a water immersion lens, which closer matches the refractive index of living cells, could be used to reduce artefacts and spherical aberration. An argon-ion 457–514 nm laser was used to excite the EGFR-EGFP construct and a green diode 561 nm laser was used for the rab5-mRFP construct. Images were acquired, or the data was scaled down, such that the pixel size was 0.25 µm. Z-stacks with 0.3 µm slice spacing were acquired every 10 min for a total duration of 30 min. Each stack took approximately 3 min to acquire.

A.3. K-means based segmentation estimate

To produce the segmentation estimate for the initialisation of the level set function, the data from the EGFR-EGFP channel was clustered using a K-means approach (4 clusters) [55]. The 3 clusters with the highest means were combined and subsequently dilated, filled and eroded with a spherical structural element (with a 1.5 µm radius). Holes in each axial slice were filled separately.

Table A.1
Parameters for generation of simulated colocalization data using the Colocalization Simulator plugin from the software application Icy. The colocalization percentage was varied and three noise levels were generated (low/medium/high).

Parameter	Value
Sequence width/height	256
Number of slices	50
Number of spots (for both channels)	100
Minimum spot intensity (for both channels)	20
Maximum spot intensity (for both channels)	20
Mean colocalization distance	0
Std colocalization distance	0
Maximum number of iterations	50
Poisson noise (low/medium/high)	0/2/4
Mean Gaussian noise (low/medium/high)	0/2/4
Std Gaussian noise	1

Table A.2
Parameters for the distance regularized level set evolution (DRLSE) segmentation protocol. The full details of the protocol can be found in Li et al. (2010) [43]. A stopping condition is employed such that the segmentation protocol is stopped if the volume contained within the zero level set changes less than 0.01% between two iterations. For the 3D extension the axial and spatial derivatives are scaled using the pixel size and axial spacing.

Parameter	Value
Edge term constant, λ	8
Constant force term, α	0
DRLSE constant, μ	0.15
Binary step size, c_0	2
Update step, Δt	1
Dirac delta width, ε	1.5
Stopping condition	0.01%
Maximum number of iterations	50

A.4. Generation of synthetic data for colocalization analysis

To produce the simulated data with positive, or no, colocalization the Colocalization Simulator plugin (version 0.0.1.2), from the open image analysis package Icy, was used [50,51]. To generate the anti-colocalized data the Colocalization Simulator plugin was modified. First, the spots in both channel 1 and 2 were randomly placed. If a spot from channel 2 was less than a set distance (20 pixels) from a channel 1 spot, then it was randomly re-positioned. This re-positioning was repeated until all anti-colocalized spots from channel 2 were at least the minimum distance from all spots in channel 1. Finally, the remaining spots from channel 2 (neither colocalized or anti-colocalized) were randomly distributed. The plugin parameters for both the colocalized and anti-colocalized data were set as specified in Table A.1. 50 stacks were generated per colocalization percentage and noise level.

References

[1] S. Jones, J.Z. Rappoport, Interdependent epidermal growth factor receptor signalling and trafficking, Int. J. Biochem. 51 (2014) 23–28.
[2] A. Sorkin, M. von Zastrow, Endocytosis and signalling: intertwining molecular networks, Nat. Rev. Mol. Cell Biol. 10 (2009) 609–622.
[3] R. Irannejad, N.G. Tsvetanova, B.T. Lobingier, M. von Zastrow, Effects of endocytosis on receptor-mediated signaling, Curr. Opin. Cell Biol. 35 (2015) 137–143.
[4] R. Villaseñor, H. Nonaka, P. Del Conte-Zerial, Y. Kalaidzidis, M. Zerial, Regulation of EGFR signal transduction by analogue-to-digital conversion in endosomes, Elife 4 (2015) e06156.

[5] A. Fortian, A. Sorkin, Live-cell fluorescence imaging reveals high stoichiometry of Grb2 binding to the EGF receptor sustained during endocytosis, J. Cell Sci. 127 (2014) 432–444.
[6] D. Teis, N. Taub, R. Kurzbauer, D. Hilber, M.E. de Araujo, M. Erlacher, M. Offterdinger, A. Villunger, S. Geley, G. Bohn, Others, p14-MP1-MEK1 signaling regulates endosomal traffic and cellular proliferation during tissue homeostasis, J. Cell Biol. 175 (2006) 861–868.
[7] C.L. Howe, W.C. Mobley, Signaling endosome hypothesis: a cellular mechanism for long distance communication, J. Neurobiol. 58 (2004) 207–216.
[8] G. Di Guglielmo, P. Baass, W. Ou, B. Posner, J. Bergeron, Compartmentalization of SHC, GRB2 and mSOS, and hyperphosphorylation of Raf-1 by EGF but not insulin in liver parenchyma, EMBO J. 13 (1994) 4269.
[9] B.D. Grant, J.G. Donaldson, Pathways and mechanisms of endocytic recycling, Nat. Rev. Mol. Cell Biol. 10 (2009) 597–608.
[10] A. Schenck, L. Goto-Silva, C. Collinet, M. Rhinn, A. Giner, B. Habermann, M. Brand, M. Zerial, The endosomal protein Appl1 mediates Akt substrate specificity and cell survival in vertebrate development, Cell 133 (2008) 486–497.
[11] L. Munsie, A. Milnerwood, P. Seibler, D. Beccano-Kelly, I. Tatarnikov, J. Khinda, M. Volta, C. Kadgien, L. Cao, L. Tapia, et al., Retromer-dependent neurotransmitter receptor trafficking to synapses is altered by the Parkinson's disease VPS35 mutation p. D620N, Hum. Mol. Genet. 24 (2015) 1691–1703.
[12] A. Tomas, C.E. Futter, E.R. Eden, EGF receptor trafficking: consequences for signaling and cancer, Trends Cell Biol. 24 (2014) 26–34.
[13] G. Auciello, D.L. Cunningham, T. Tatar, J.K. Heath, J.Z. Rappoport, Regulation of fibroblast growth factor receptor signalling and trafficking by Src and Eps8, J. Cell Sci. 126 (2013) 613–624.
[14] K.W. Dunn, M.M. Kamocka, J.H. McDonald, A practical guide to evaluating colocalization in biological microscopy, Am. J. Physiol. Cell Physiol. 300 (2011) C723–C742.
[15] P. Sarder, A. Nehorai, Deconvolution methods for 3-D fluorescence microscopy images, IEEE Signal Process. Mag. 23 (2006) 32–45.
[16] M.A. Lemmon, J. Schlessinger, K.M. Ferguson, The EGFR family: not so prototypical receptor tyrosine kinases, CSH Perspect. Biol. 6 (2014) a020768.
[17] H. Stenmark, Rab GTPases as coordinators of vesicle traffic, Nat. Rev. Mol. Cell Biol. 10 (2009) 513–525.
[18] A. Levitzki, A. Gazit, Tyrosine kinase inhibition: an approach to drug development, Science 267 (5205) (1995) 1782–1788.
[19] E. Macia, M. Ehrlich, R. Massol, E. Boucrot, C. Brunner, T. Kirchhausen, Dynasore, a cell-permeable inhibitor of dynamin, Dev. Cell 10 (2006) 839–850.
[20] J.-M. Hsu, C.-T. Chen, C.-K. Chou, H.-P. Kuo, L.-Y. Li, C.-Y. Lin, H.-J. Lee, Y.-N. Wang, M. Liu, H.-W. Liao, et al., Crosstalk between Arg 1175 methylation and Tyr 1173 phosphorylation negatively modulates EGFR-mediated ERK activation, Nat. Cell Biol. 13 (2011) 174–181.
[21] A. Buades, B. Coll, J.-M. Morel, A review of image denoising algorithms, with a new one, Multiscale Model. Simul. 4 (2005) 490–530.
[22] C. Kervrann, C.Ó.S. Sorzano, S.T. Acton, J.-C. Olivo-Marin, M. Unser, A guided tour of selected image processing and analysis methods for fluorescence and electron microscopy, IEEE J. Sel. Top. Signal Process. 10 (2016) 6–30.
[23] J. Adler, I. Parmryd, Colocalization analysis in fluorescence microscopy, Methods Mol. Biol. 931 (2013) 97–109.
[24] E. Abbe, Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung, Arch. Mikrosk. Anat. 9 (1873) 413–418.
[25] S. Bolte, F. Cordelieres, A guided tour into subcellular colocalization analysis in light microscopy, J. Microsc. 224 (2006) 213–232.
[26] T. Lagache, N. Sauvonnet, L. Danglot, J.-C. Olivo-Marin, Statistical analysis of molecule colocalization in bioimaging, Cytometry A 87 (6) (2015) 568–579.
[27] R.B. Sekar, A. Periasamy, Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations, J. Cell Biol. 160 (2003) 629–633.
[28] M. Miaczynska, S. Christoforidis, A. Giner, A. Shevchenko, S. Uttenweiler-Joseph, B. Habermann, M. Wilm, R.G. Parton, M. Zerial, APPL proteins link Rab5 to nuclear signal transduction via an endosomal compartment, Cell 116 (2004) 445–456.
[29] M. Ward, A. Cunningham, Developmental expression of vascular endothelial growth factor receptor 3 and vascular endothelial growth factor C in forebrain, Neuroscience 303 (2015) 544–557.
[30] A. Frederick, H. Yano, P. Trifilieff, H. Vishwasrao, D. Biezonski, J. Mészáros, E. Urizar, D. Sibley, C. Kellendonk, K. Sonntag, et al., Evidence against dopamine D1/D2 receptor heteromers, Mol. Psychiatry (2015).
[31] B.D. Ripley, The second-order analysis of stationary point processes, J. Appl. Probab. (1976) 255–266.
[32] J. Adler, I. Parmryd, Quantifying colocalization by correlation: the Pearson correlation coefficient is superior to the Mander's overlap coefficient, Cytometry A 77 (2010) 733–742.
[33] E. Manders, F. Verbeek, J. Aten, Measurement of co-localization of objects in dual-colour confocal images, J. Microsc. 169 (1993) 375–382.
[34] E. Manders, J. Stap, G. Brakenhoff, R. Van Driel, J. Aten, Dynamics of three-dimensional replication patterns during the S-phase, analysed by double labelling of DNA and confocal microscopy, J. Cell Sci. 103 (1992) 857–862.
[35] V.R. Singan, T.R. Jones, K.M. Curran, J.C. Simpson, Dual channel rank-based intensity weighting for quantitative co-localization of microscopy images, BMC Bioinformatics 12 (2011) 407.
[36] F. Humpert, I. Yahiatene, M. Lummer, M. Sauer, T. Huser, Quantifying molecular colocalization in live cell fluorescence microscopy, J. Biophotonics 8 (2015) 124–132.
[37] S.V. Costes, D. Daelemans, E.H. Cho, Z. Dobbin, G. Pavlakis, S. Lockett, Automatic and quantitative measurement of protein-protein colocalization in live cells, Biophys. J. 86 (2004) 3993–4003.
[38] J.H. McDonald, K.W. Dunn, Statistical tests for measures of colocalization in biological microscopy, J. Microsc. 252 (2013) 295–302.
[39] F. Luisier, C. Vonesch, T. Blu, M. Unser, Fast interscale wavelet denoising of Poisson-corrupted images, Signal Process. 90 (2010) 415–427.
[40] D. Sage, D. Prodanov, J.-Y. Tinevez, J. Schindelin, MIJ: making interoperability between ImageJ and Matlab possible, ImageJ User & Developer Conference, 2012, pp. 24–26.
[41] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis, Nat. Methods 9 (2012) 671–675.
[42] J.B. Pawley, Sources of noise in three-dimensional microscopical data sets, Three-Dimensional Confocal Microscopy: Volume Investigation of Biological Specimens, 2012, pp. 47–94.
[43] C. Li, C. Xu, C. Gui, M.D. Fox, Distance regularized level set evolution and its application to image segmentation, IEEE Trans. Image Process. 19 (2010) 3243–3254.
[44] V. Caselles, R. Kimmel, G. Sapiro, Geodesic active contours, Int. J. Comput. Vision 22 (1997) 61–79.
[45] M. Maška, V. Ulman, D. Svoboda, P. Matula, P. Matula, C. Ederra, A. Urbiola, T. España, S. Venkatesan, D.M. Balak, et al., A benchmark for comparison of cell tracking algorithms, Bioinformatics 30 (2014) 1609–1617.
[46] S.R. Sternberg, Biomedical image processing, IEEE Comput. 16 (1983) 22–34.
[47] Y. Mishchenko, A fast algorithm for computation of discrete Euclidean distance transform in three or more dimensions on vector processing architectures, Signal Image Video Process. 9 (2015) 19–27.
[48] N. Otsu, A threshold selection method from gray-level histograms, Automatica 11 (1975) 23–27.
[49] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, et al., Fiji: an open-source platform for biological-image analysis, Nat. Methods 9 (2012) 676–682.
[50] F. De Chaumont, S. Dallongeville, N. Chenouard, N. Hervé, S. Pop, T. Provoost, V. Meas-Yedid, P. Pankajakshan, T. Lecomte, Y. Le Montagner, et al., Icy: an open bioimage informatics platform for extended reproducible research, Nat. Methods 9 (2012) 690–696.
[51] N. Chenouard, Advances in probabilistic particle tracking for biological imaging (thesis), 2011.
[52] C.H. Li, C. Lee, Minimum cross entropy thresholding, Pattern Recognit. 26 (1993) 617–625.
[53] C. Collinet, M. Stöter, C.R. Bradshaw, N. Samusik, J.C. Rink, D. Kenski, B. Habermann, F. Buchholz, R. Henschel, M.S. Mueller, et al., Systems survey of endocytosis by multiparametric image analysis, Nature 464 (2010) 243–249.
[54] P. Kozik, N.A. Hodson, D.A. Sahlender, N. Simecek, C. Soromani, J. Wu, L.M. Collinson, M.S. Robinson, A human genome-wide screen for regulators of clathrin-coated vesicle formation reveals an unexpected role for the V-ATPase, Nat. Cell Biol. 15 (2013) 50–60.
[55] S.P. Lloyd, Least squares quantization in PCM, IEEE Trans. Inf. Theory 28 (1982) 129–137.