

# Colocalization Analysis

by Dominic Waithe UKRI Innovation Fellow.

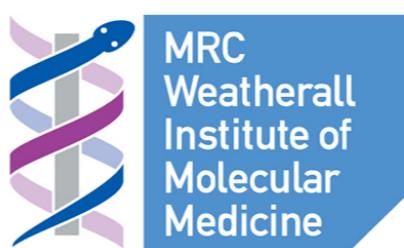
**12<sup>th</sup> December 2019**

**IAFIG-RMS - Bioimage Analysis With Python**  
**Cambridge Bioinformatics Training Centre**

## TODAYS TALK:

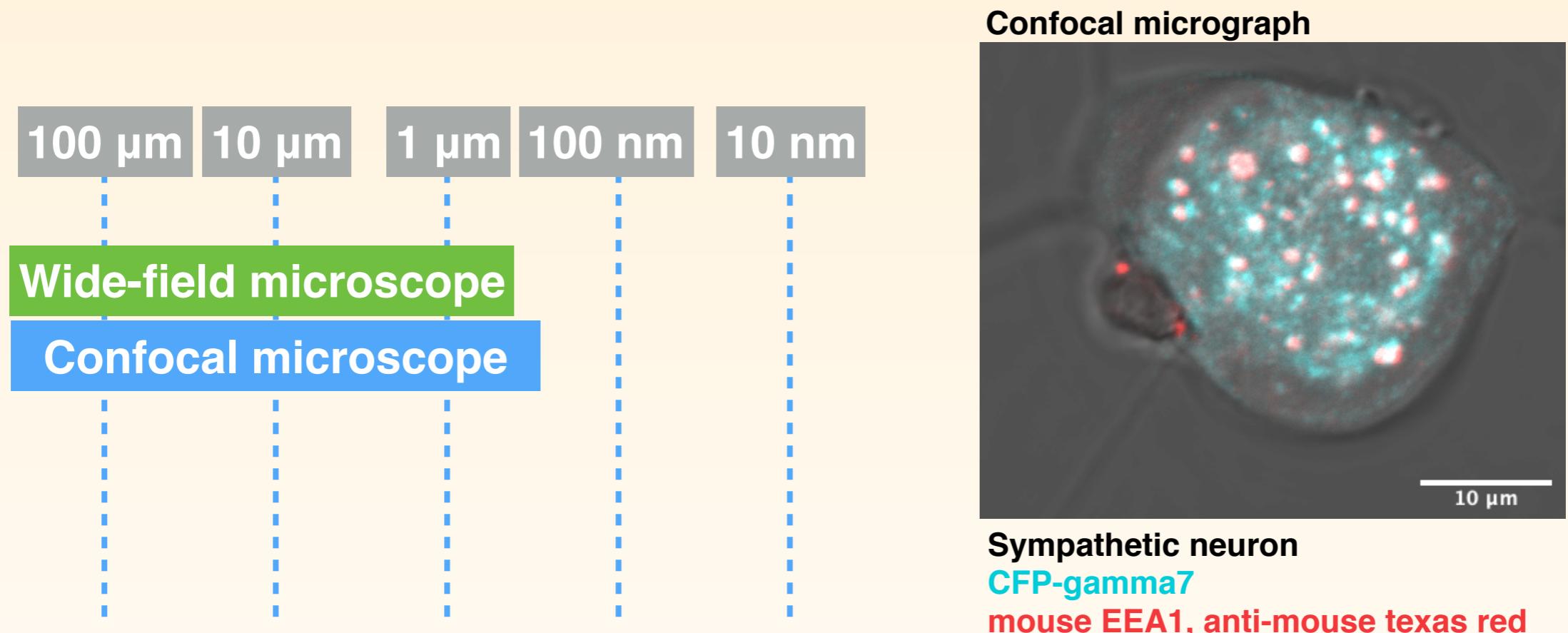
- Conventional Approaches
- Mander's test
- Pearson's test
- Object based techniques
- Super-resolution colocalization

UK Research  
and Innovation



# **Conventional Approaches**

# Light microscopy “the good old days.”



Images are diffraction limited.

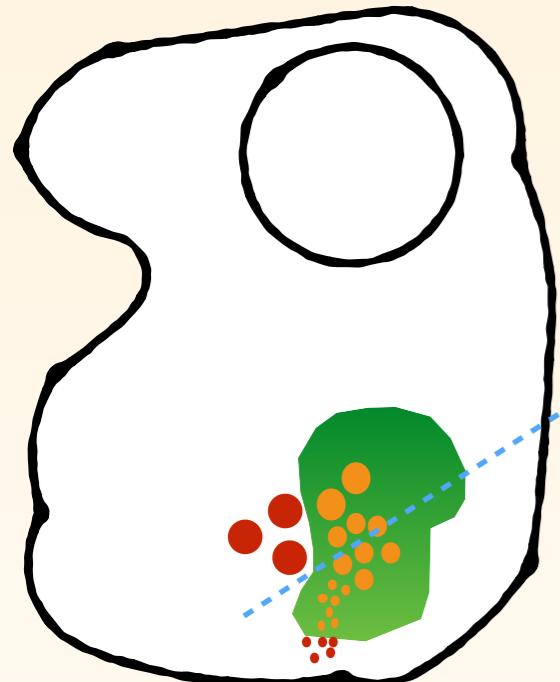
We can resolve down to the organelle level. Our pixel represent areas bigger than a single protein (>250 nm), more like a chunk of organelle.

Source: Adapted from <http://zeiss-campus.magnet.fsu.edu/articles/superresolution/introduction.html>

# Pixel level colocalization techniques.

## Experiment A

### Cooccurrence



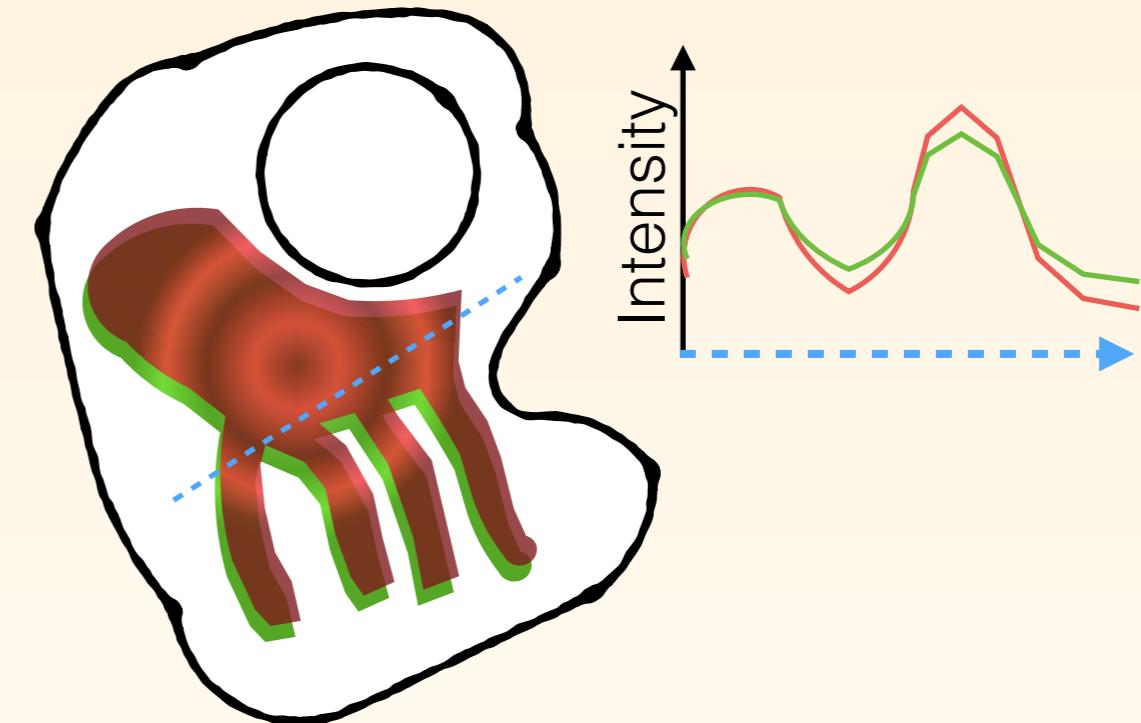
We are interested in whether pixels overlap (Cooccurrence).

E.g. Mander's test

E.g. Do these vesicles with Protein A bind this structure containing Protein B?

## Experiment B

### Correlation



Or a more powerful question, do the pixel intensities distribute in the same way (Correlation)

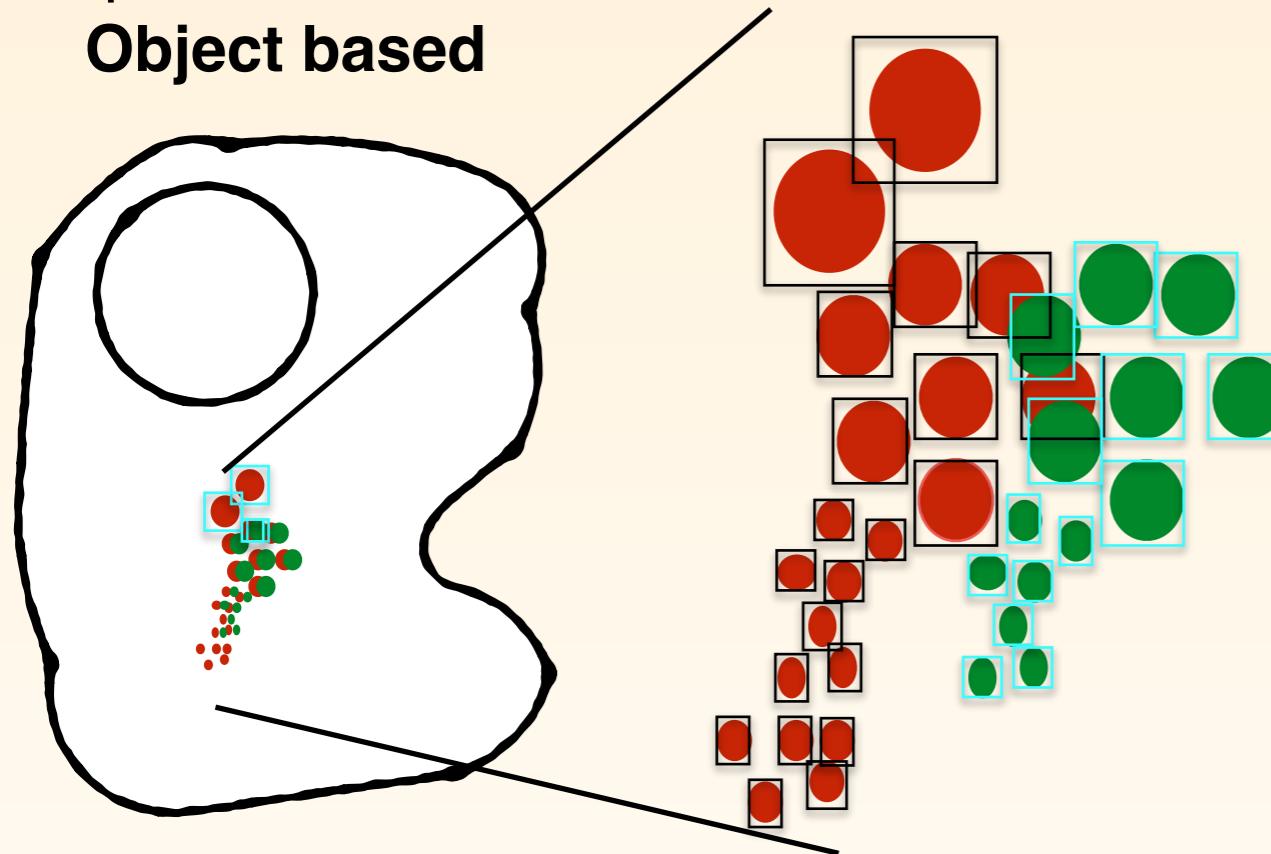
e.g. Pearson's test

E.g. Are these two proteins enriched in the same locations?

# Object-based colocalization

## Experiment C

### Object based

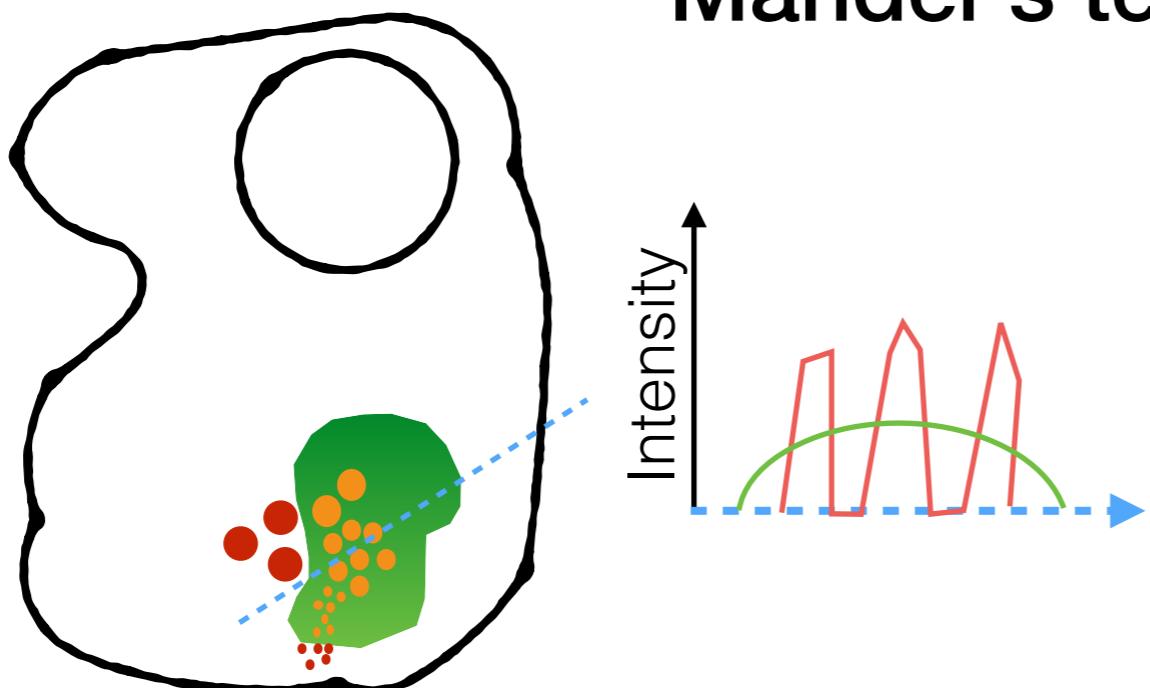


x y	x y
15,25	13,12
34,4	34,23
34,15	34,23
23,5	12,23
50,32	16,7
40,3	34,43
12,35	13,23
3,23	3,12
8,23	

We parameterise the location of the objects (e.g. segmentation, maxima finding, model fitting).

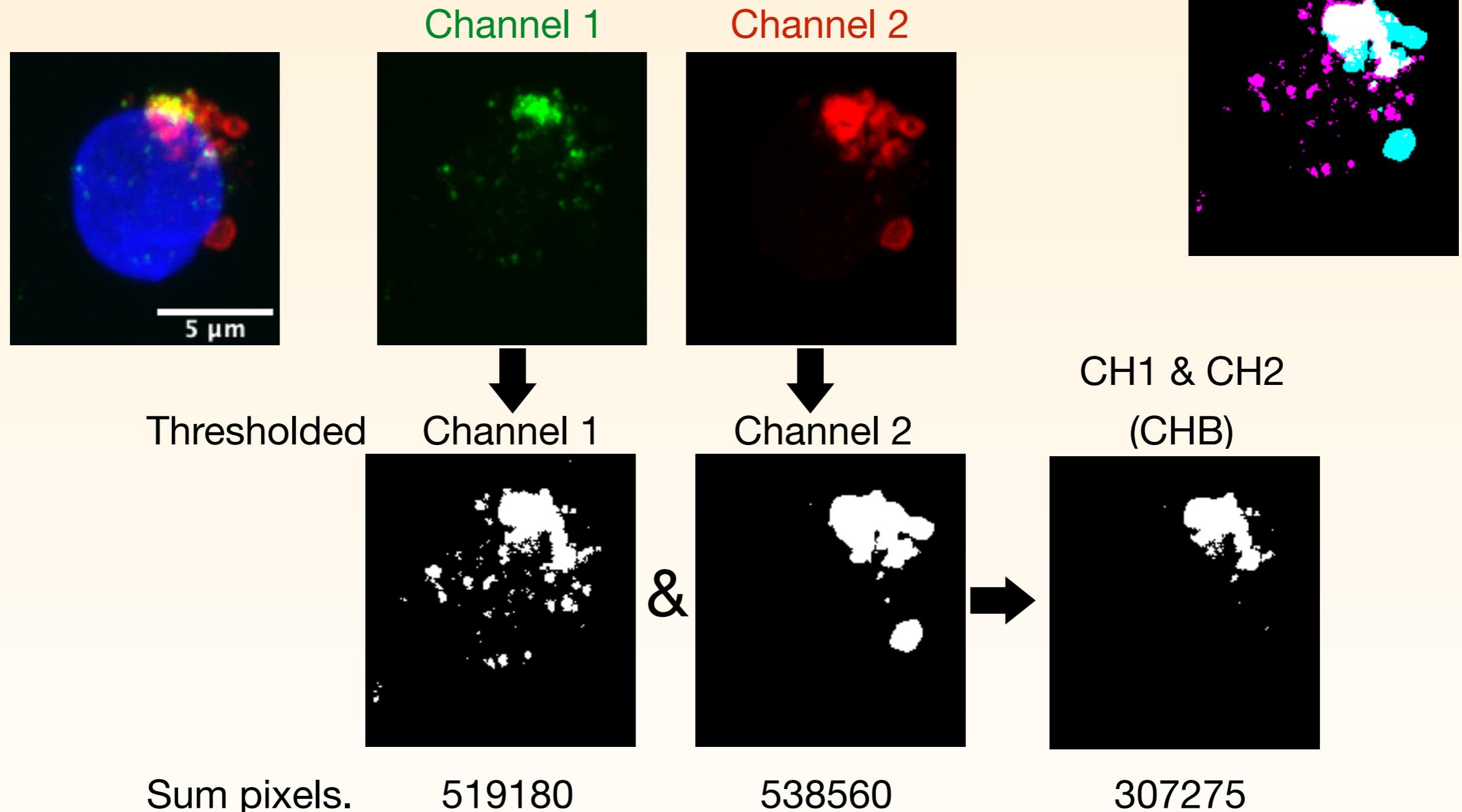
We then compare the coordinates of the points and make a decision about the resulting distribution.

## Mander's test.



# Cooccurrence example using Mander's test

Mander's test is very simple.



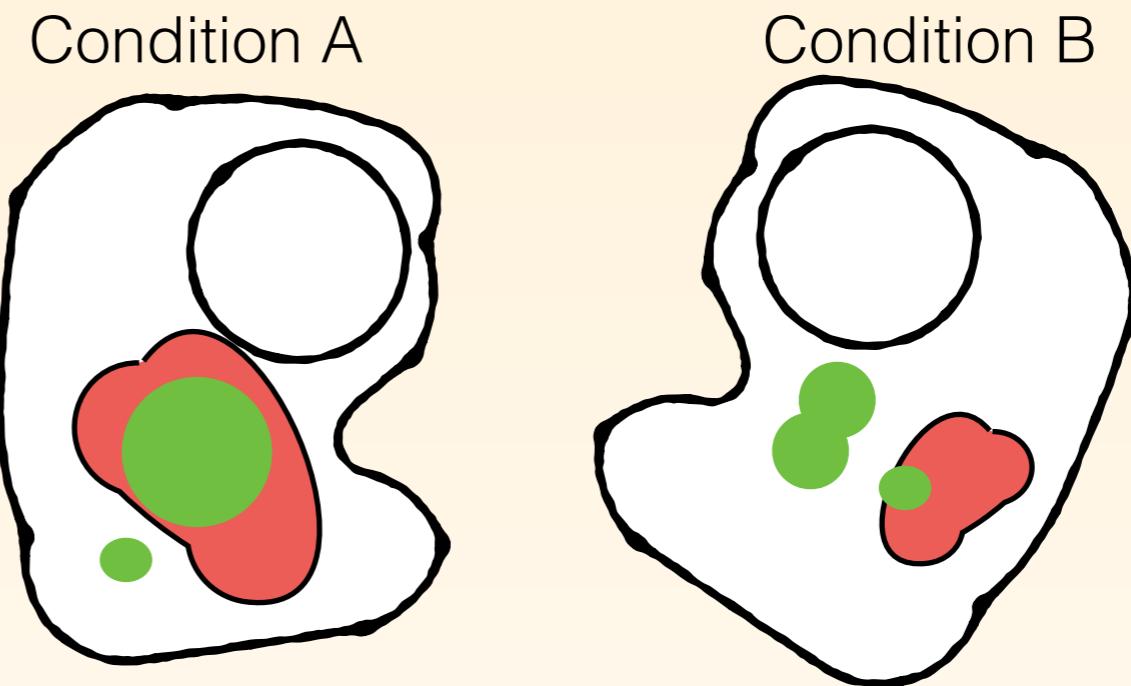
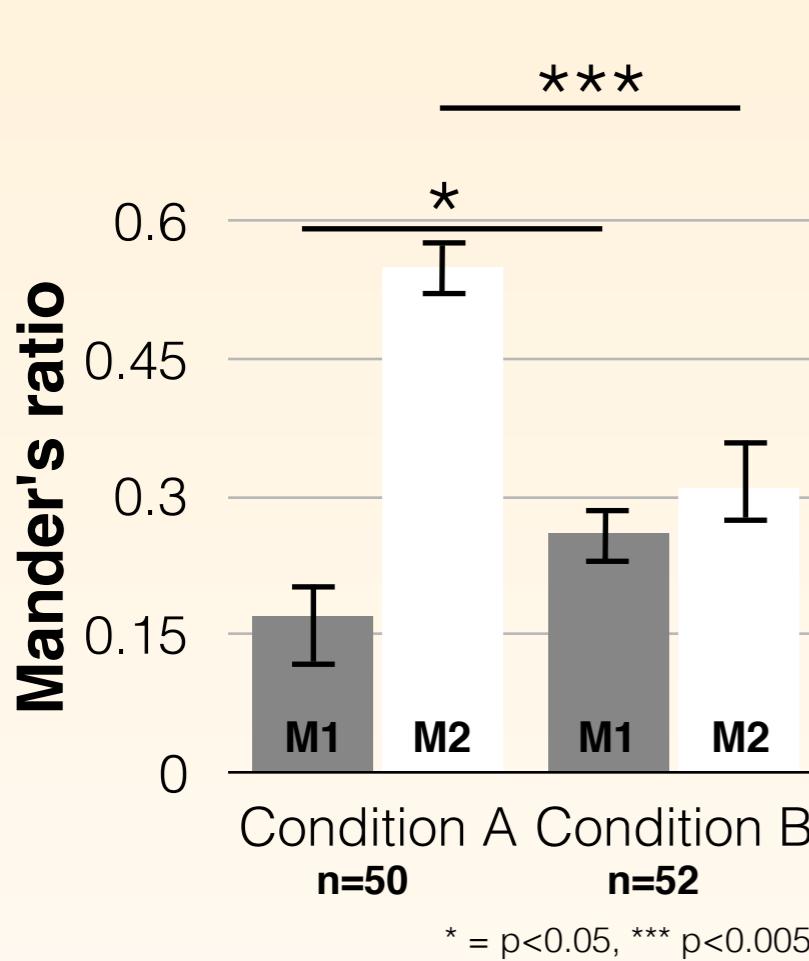
Results in two statistics:

$$M1 = \text{CHB}/\text{CH1} \text{ and } M2 = \text{CHB}/\text{CH2}$$

$$M1 = 0.519 \text{ and } M2 = 0.571$$

Source: C5aR\_LAMP-1\_cell003.czi Nazish Malik

# Mander's test interpretation



“In Condition B, the M2 ratio was significantly lower than in Condition A suggesting that protein B had translocated out of the organelle in which protein A was enriched.”

do you agree?

The Mander's test can be used as above. With the comparisons made between the conditions. You have to be careful as the interpretation can be quite complex.

M1 = CHB/CH1 and M2 = CHB/CH1

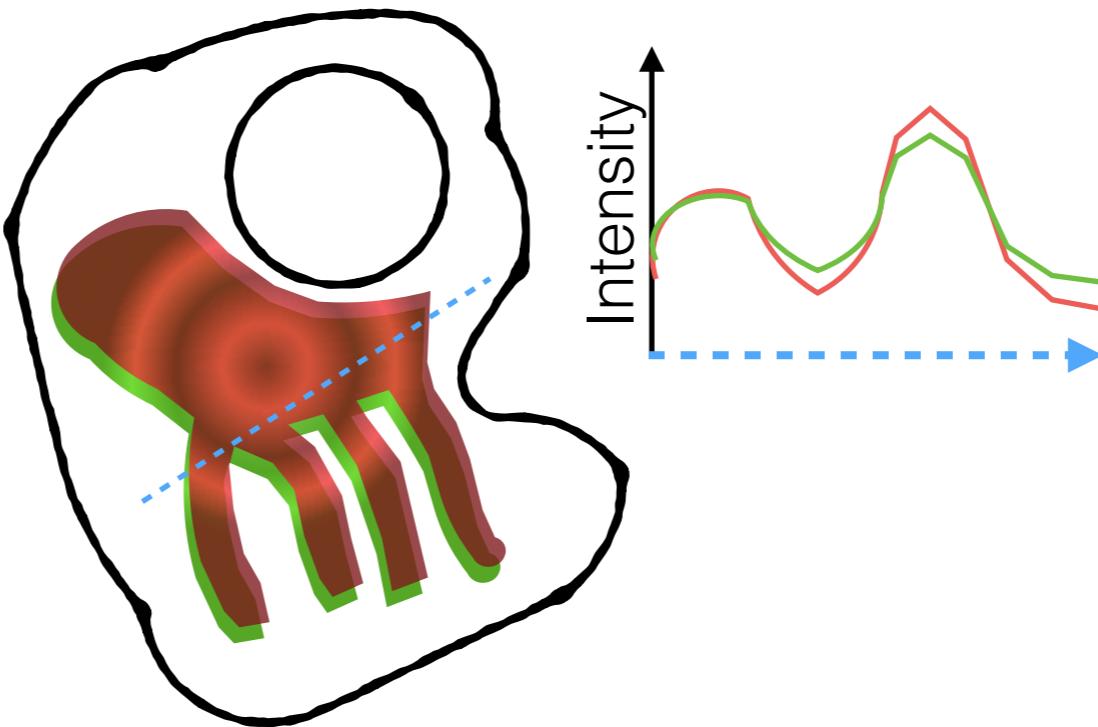
# Mander's Practical. 20 min. Try this yourself:

In the shared data drive, please goto “Colocalization” folder.

**1a)** Open C5aR\_LAMP-1\_cell003.tif. Split the channels. Threshold the green channel (GFP C5aR) with your choice of algorithm. Threshold the red channel (LAMP-1) with your choice of algorithm. Use the Image Calculator and combine the images using the “AND” operator (CHB). Measure each thresholded channel and the resulting CHB and store the “rawIntegratedDensity” in a spreadsheet. Calculate M1 and M2 statistic.  $M1 = CHB/CH1$  and  $M2 = CHB/CH2$ .

**ADVANCED: Write a script to do this automatically.**

## Dot Product to Pearson's test.



# The dot product of two vectors (algebraic)

in  $\mathbf{R}^{12}$

$a$

5,  
6,  
8,  
6,  
5,  
4,  
3,  
2,  
3,  
5,  
7,  
9,  
7,

$b$

4,  
5,  
6,  
5,  
4,  
3,  
5,  
6,  
5,  
4,  
3,  
4,

$a \bullet b$

(5 x 4) +  
(6 x 5) +  
(8 x 6) +  
(6 x 5) +  
(5 x 4) +  
(3 x 3) +  
(2 x 5) +  
(3 x 6) +  
(5 x 5) +  
(7 x 4) +  
(9 x 3) +  
(7 x 4) +

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^n a_i b_i = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$$

Scalar/dot product

= 293

Source:

# Motivation for understanding dot product

colocalisation

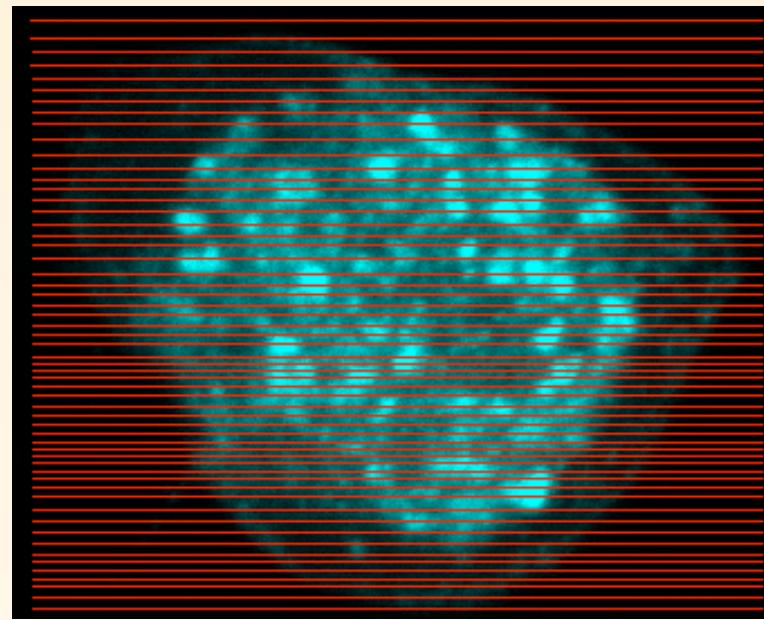
Fluorescence correlation spectroscopy

Registration

Convolution

At the core of a lot of  
techniques

# Image (2d array) to list (1d)



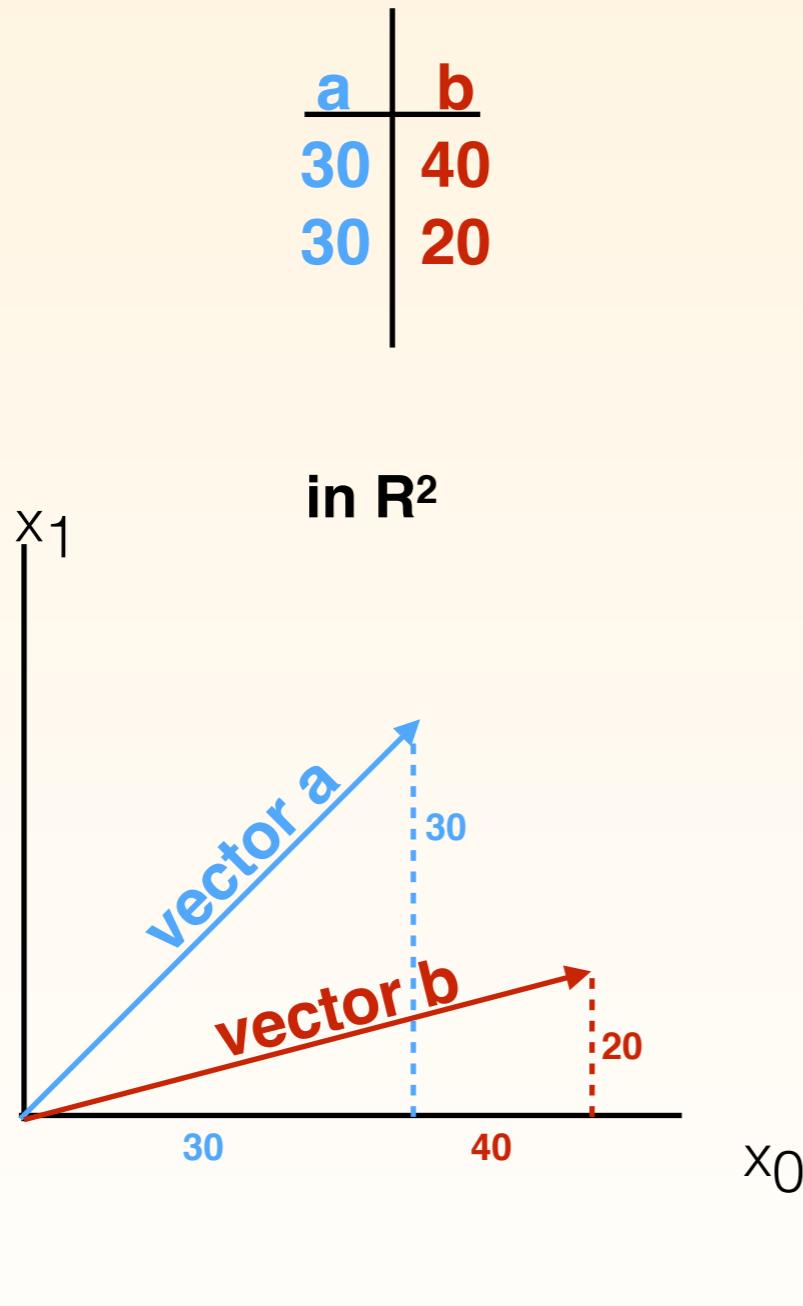
=

[ 93], [ 23], [ 23], [155], [155], [155]	155
[ 107], [ <b>198</b> ], [ <b>198</b> ], [140], [140], [140]	140
[ 121], [ 11], [ 11], [ 7], [ 7], [ 7]	7
[ 135], [ <b>235</b> ], [ <b>235</b> ], [198], [198], [198]	198
[ 149], [114], [114], [ <b>213</b> ], [ <b>213</b> ], [ <b>213</b> ]	<b>213</b>
[ 163], [ <b>187</b> ], [ <b>187</b> ], [ 9], [ 9], [ 9]	9
[ 8], [ 80], [ 80], [150], [150], [150]	150
[ 22], [ <b>187</b> ], [ <b>187</b> ], [ 20], [ 20], [ 20]	20
[ 16], [165], [165], [111], [111], [111]	111
[ 158], [ 15], [ 15], [ 34], [ 34], [ 34]	34
[ <b>200</b> ], [120], [120], [ 69], [ 69], [ 69]	69
	155
	140
	7
	198
	<b>213</b>
	9
	150
	20
	111
	34
	69
	etc
	etc

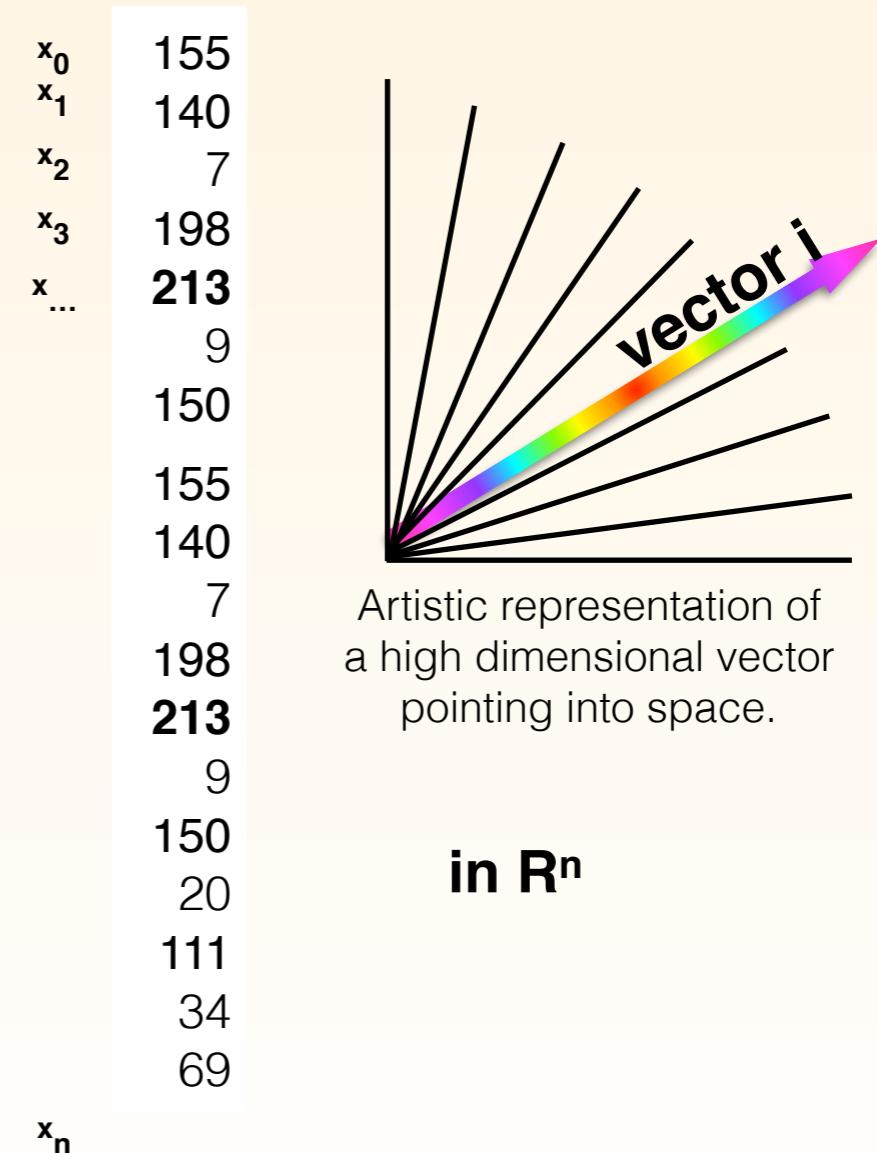
Yes. We take our image and represent it as a very long list of pixel intensities.

Source:

# Visualisation of vectors.

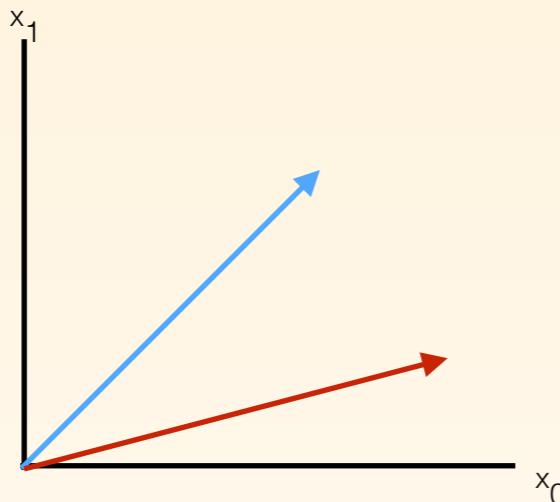


Our image would be a very high-dimensional vector, but would still point somewhere



Source:

# What does the dot product mean in this case?



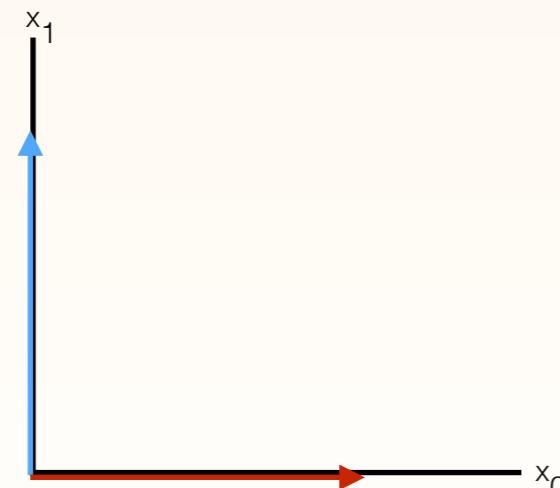
a	b
30	40
30	20

$$a \cdot b = (30*40) + (30*20) \\ = 1800$$

They have components in similar dimensions

In Euclidean space a vector has magnitude (length<sup>2</sup>) and direction.

The dot product of the two vectors represent the magnitude within a common dimensional space.



a	b
0	60
60	0

$$a \cdot b = (0*60) + (60*0) \\ = 0$$

They have no shared dimensionality

Source:

# The algebraic dot product

in  $\mathbf{R}^{12}$

$a$	$b$	$a \bullet b$
5,	4,	(5 x 4) +
6,	5,	(6 x 5) +
8,	6,	(8 x 6) +
6,	5,	(6 x 5) +
5,	4,	(5 x 4) +
3,	3,	(3 x 3) +
2,	5,	(2 x 5) +
3,	6,	(3 x 6) +
5,	5,	(5 x 5) +
7,	4,	(7 x 4) +
9,	3,	(9 x 3) +
7,	4,	(7 x 4) +

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^n a_i b_i = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$$

Scalar/dot product

$$= 293$$

dot product represents the magnitude of these two vectors into a common space.  
Its hard to imagine directly!

# Geometric dot product

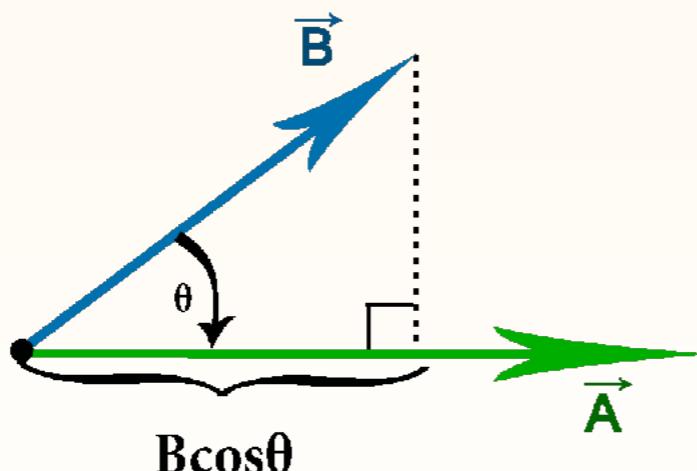
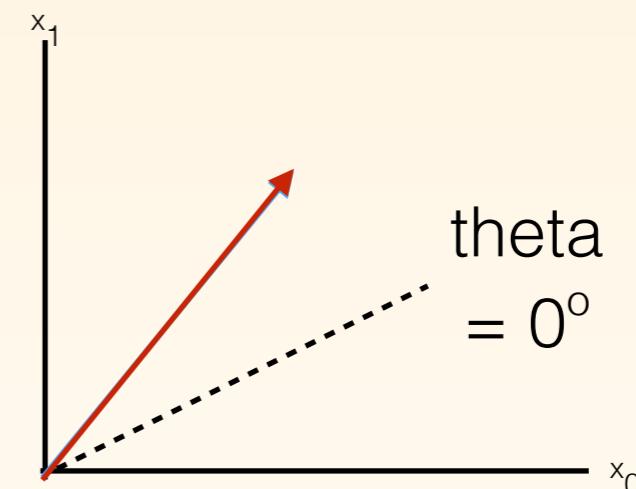
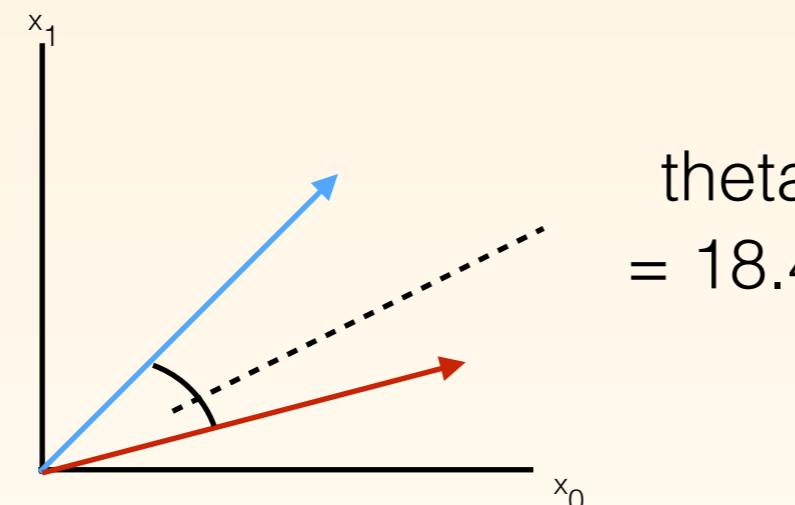
in  $\mathbb{R}^2$

$$\begin{array}{|c|c|} \hline a & b \\ \hline 30 & 40 \\ \hline 30 & 20 \\ \hline \end{array}$$

$a \cdot b = (30*40) + (30*20) = 1800$

$$\begin{array}{|c|c|} \hline a & b \\ \hline 30 & 30 \\ \hline 30 & 30 \\ \hline \end{array}$$

$a \cdot b = (30*30) + (30*30) = 1800$



We use the relationship between the algebraic and geometric dot product

$$\mathbf{A} \cdot \mathbf{B} = \|\mathbf{A}\| \|\mathbf{B}\| \cos \theta,$$

$$\|\mathbf{A}\| = \sqrt{\mathbf{A} \cdot \mathbf{A}}$$

# Pearson's product-moment correlation test.

$$\mathbf{A} \cdot \mathbf{B} = \|\mathbf{A}\| \|\mathbf{B}\| \cos \theta,$$

Pearson's equation:

$$r = \frac{\sum (R_i - \bar{R}) \times (G_i - \bar{G})}{\sqrt{\sum (R_i - \bar{R})^2 \times \sum (G_i - \bar{G})^2}}$$

r is short for cos(theta) and varies between -1.0 and 1.0

Represents the angle between two vectors, in our case the alignment of two very high dimensional vectors.

Source: [http://en.wikipedia.org/wiki/Correlation\\_coefficient](http://en.wikipedia.org/wiki/Correlation_coefficient)

# Pearson's product-moment correlation test.

Pearson's equation:

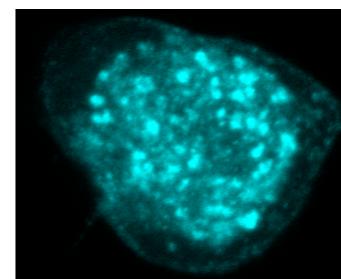
$$r = \frac{\sum (R_i - \bar{R}) \times (G_i - \bar{G})}{\sqrt{\sum (R_i - \bar{R})^2 \times \sum (G_i - \bar{G})^2}}$$

if r is 1.0 means correlation

if r is close to '0.0' no correlation.

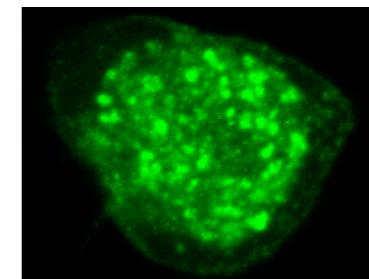
if r is -1.0 it means anti-correlation.

R refers to one channel, G refers to Green channel. G or R with a bar refers to mean intensity in that channel. 'i' refers to each pixel in image. Sigma (big E) refers to sum. So sum of all pixels minus their mean.



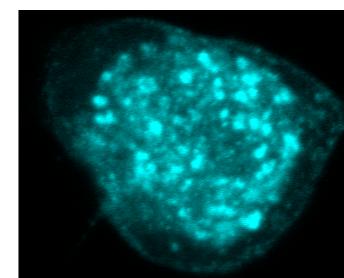
VS

high



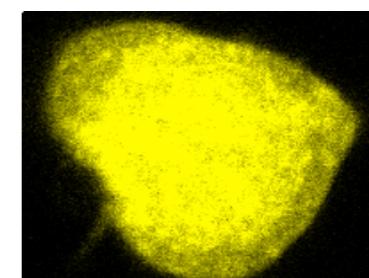
0.8

Pearson's test r value



VS

low

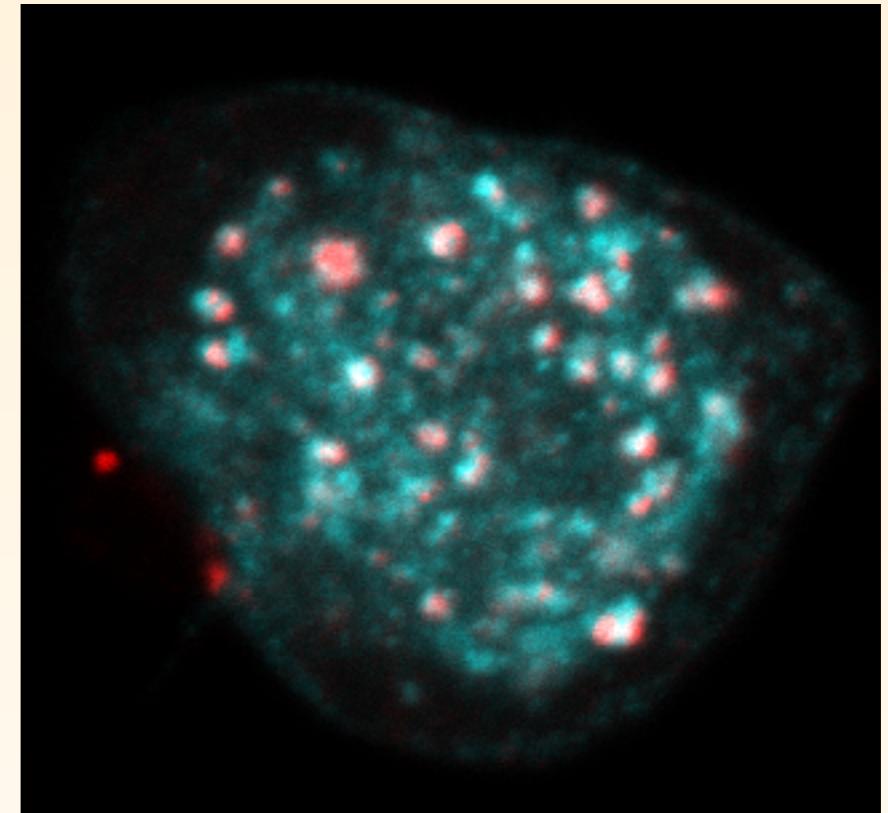
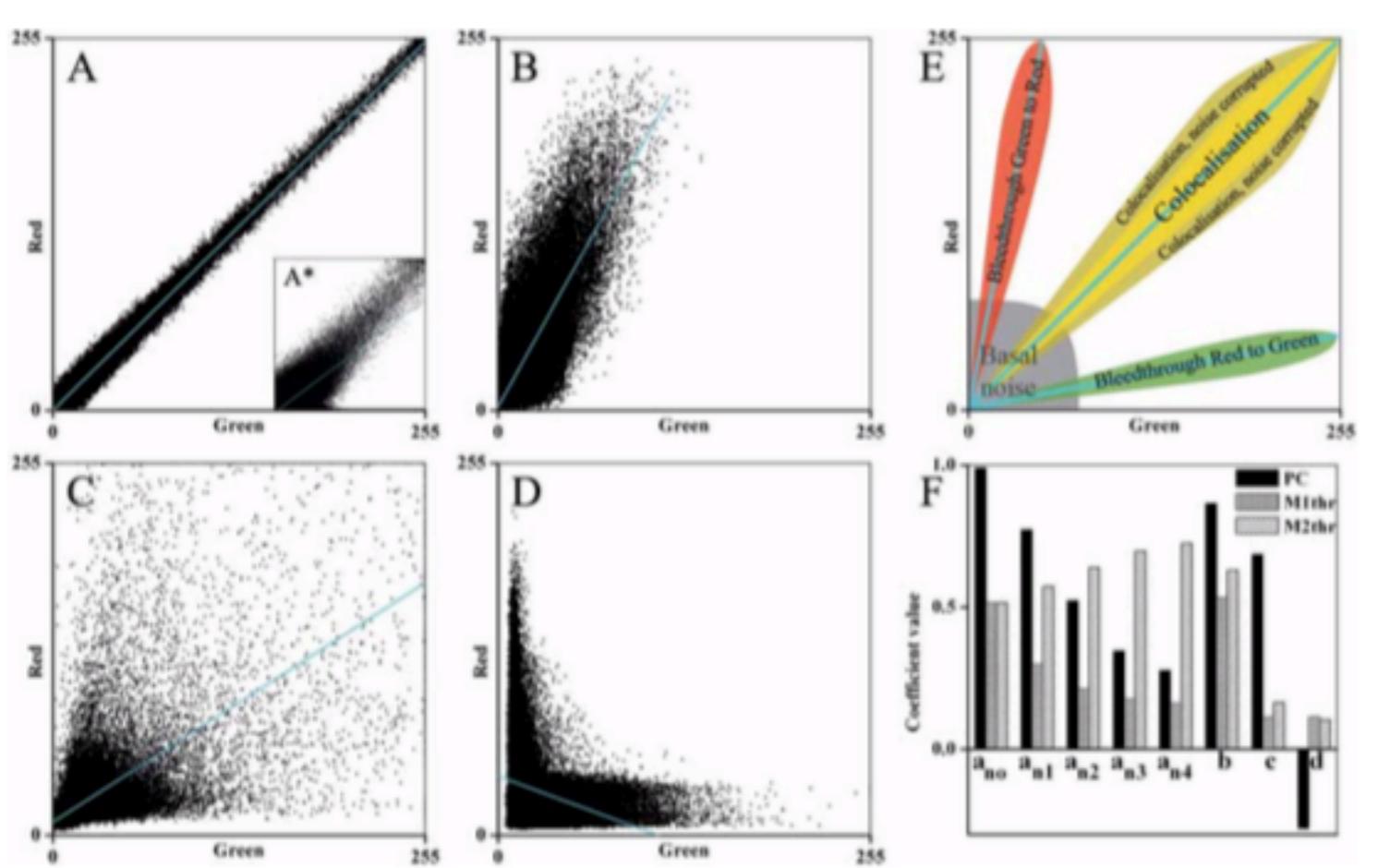


0.2

Dimensionless and normalised comparison. Can be used on any two images as long as they are the same spatial size and don't have too many black pixels

Source: [http://en.wikipedia.org/wiki/Correlation\\_coefficient](http://en.wikipedia.org/wiki/Correlation_coefficient)

# Visually inspecting your data.

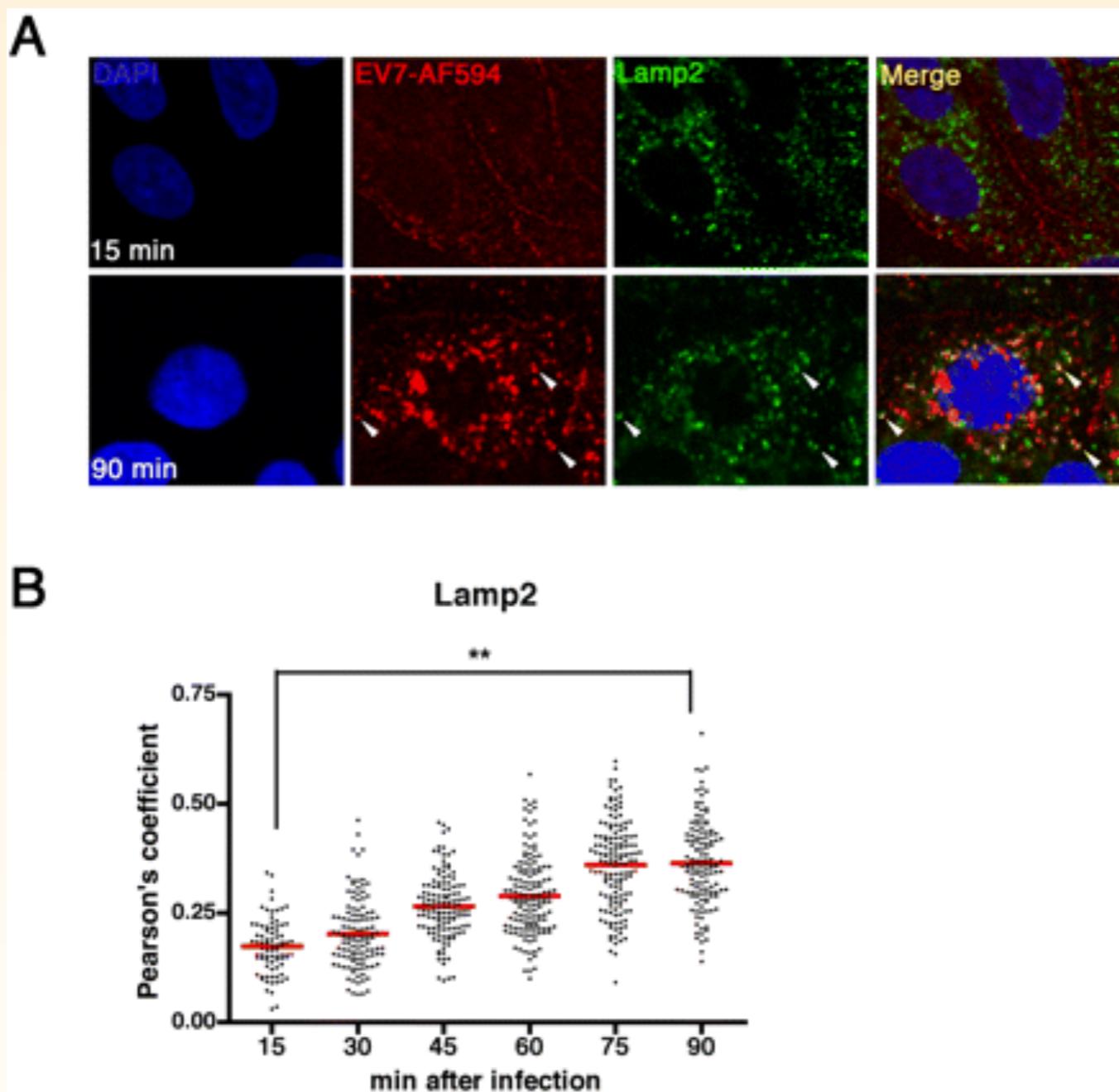


Plotting your image data channel A versus channel B, can show you the correlations in an image.

Overlaying your channels also can help you distinguish colocalization. Although choose colours sympathetic to red/green colour blind people.

Source: A guided tour into subcellular colocalization analysis in light microscopy  
S. BOLTE\* & F. P. CORDELIÈRES. Journal of Microscopy

# Pearson's test interpretation



Controls and replication are everything in colocalization analysis. Although Pearson's test has statistical meaning when applied to one image, this is meaningless in the context of biological images which suffer multiple artefacts.

Source: <http://mbio.asm.org/content/3/2/e00304-11/F5.expansion.html>

IAFIG-RMS Bioimage analysis with Python - Cambridge - 2019 - Colocalization Analysis

# Pearson's Practical A. 10min Try this yourself:

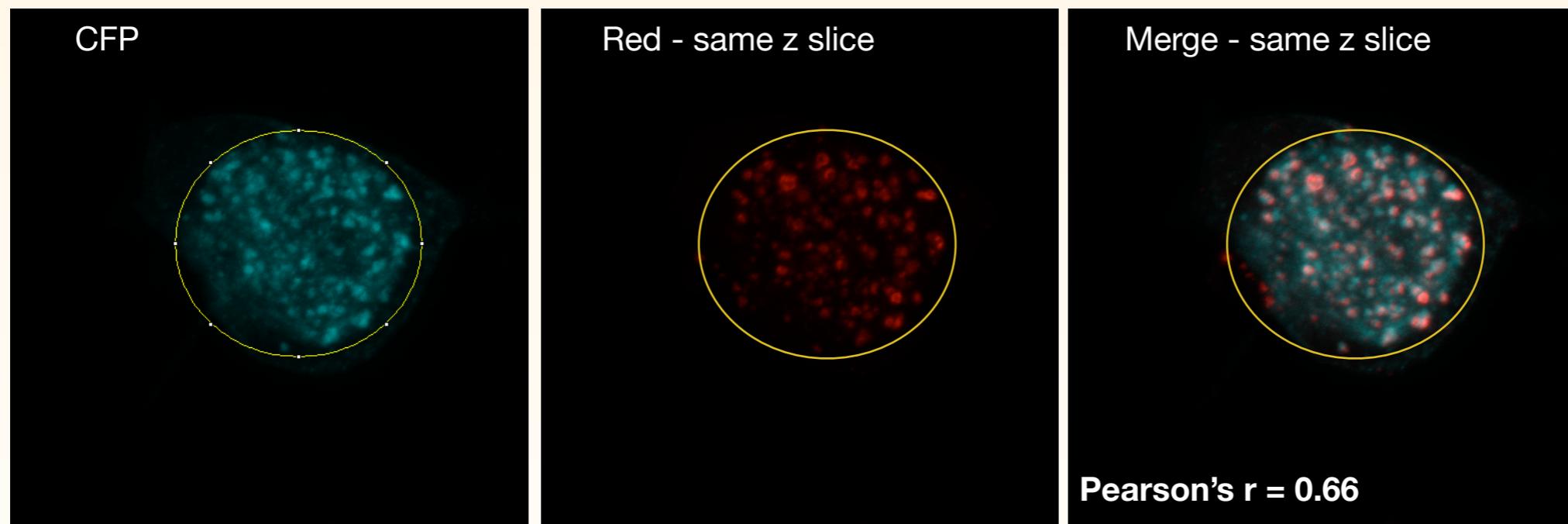
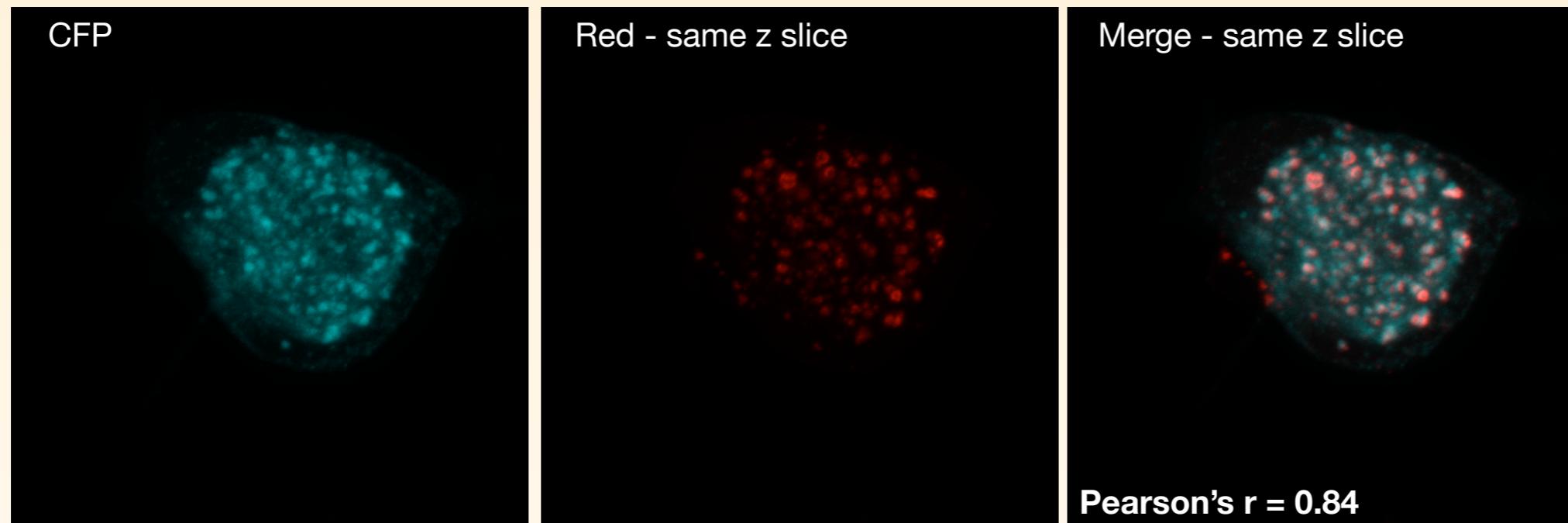
In the shared data drive, please goto “Colocalization” folder.

**1b)** Open the image Healthy.lsm. Examine the image. You should notice that there are two channels. We want to establish the colocalization between the two channels using the Pearson's correlation coefficient. To do this you will first have to split the channels: *Image->Color->Split Channels*. Note: (Green channel is filipin III (binds cholesterol) magenta is Lysotracker (as name implies stains lysosomes). Niemann Pick C is a disease where cholesterol trafficking to plasma membrane is impaired and it accumulates in lysosomes). Inspect the images. From what you have learnt from the lecture on colocalization what do you think makes this image less than ideal for Pearson's colocalization analysis? Tip (try using the Hi-Lo LUT).

**1c)** Go to, via the menu, to *Analyze->Colocalization-> Coloc 2* function. Add the two channels that you have just split, one from either channel, leave the 'ROI mask' unselected for now. Run the colocalization with all the parameters unchecked. Look for the output marked Pearson's R value (no threshold). What value do you get? Repeat this for the other image called Sick.lsm. How do the R- values differ for these two images? Do you think you could use this result as the first step towards distinguishing the visual phenotype of the two images?

Source: healthy.lsm sick.lsm Erdinc Sezgin

# P's test is sensitive zero pixels and saturation

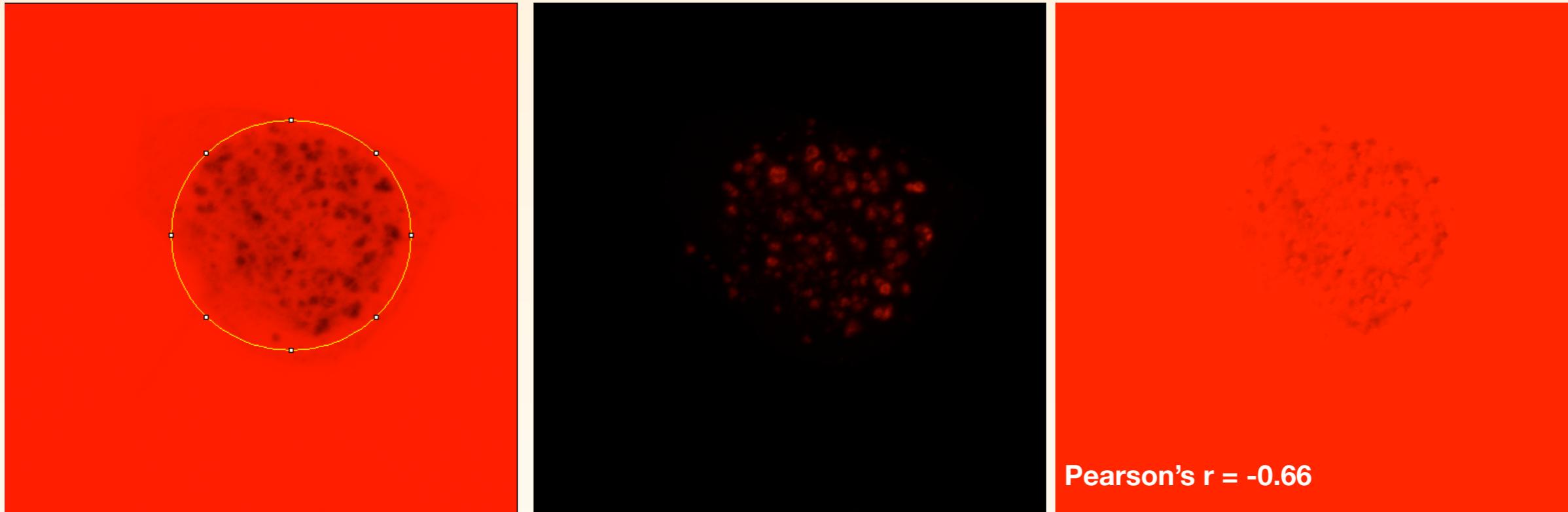


- Pearson's test doesn't ignore '0' pixels and noise within calculation.
- Coloc 2 plugin does warn you however: The ratio between zero-zero pixels and other pixels is larger 0.37. Maybe you should use a ROI.

# Pearson's Practical B. 10min.Try this yourself:

- 1d)** Open the image neuron.tif. We only want to work with the first and last channel. One way to separate them from the original image is to use image->Duplicate. Check the box marked 'Duplicate HyperStack' but type '1' in the channels box rather than having '1-4'. Click 'ok'. Repeat, but duplicate channel 4 instead. You should now have two images (one for each channel).
- 1e)** Using the Fiji menu go to *Analyze->Colocalization->Coloc 2* function. Click the first two slices that you created, one from either channel, leave the 'ROI mask' unselected for now. Run the colocalization with all the parameters checked. Look for the output marked Pearson's R-value (no threshold). What value do you get?
- 1f)** This time repeat the exercise but with an oval selection surrounding only the fluorescent region of the cell with no minimal background. Next go to *Edit- >Selection->Create Mask*. In the 'Coloc 2' dialog box you will have to select under 'mask or ROI' the newly created mask image entitled 'Mask'. Run the 'Coloc 2' plugin. What is the Pearson's R-value now? What is the effect of removing the background pixels on the R-value of the test and control condition? Why do you think this is the case?

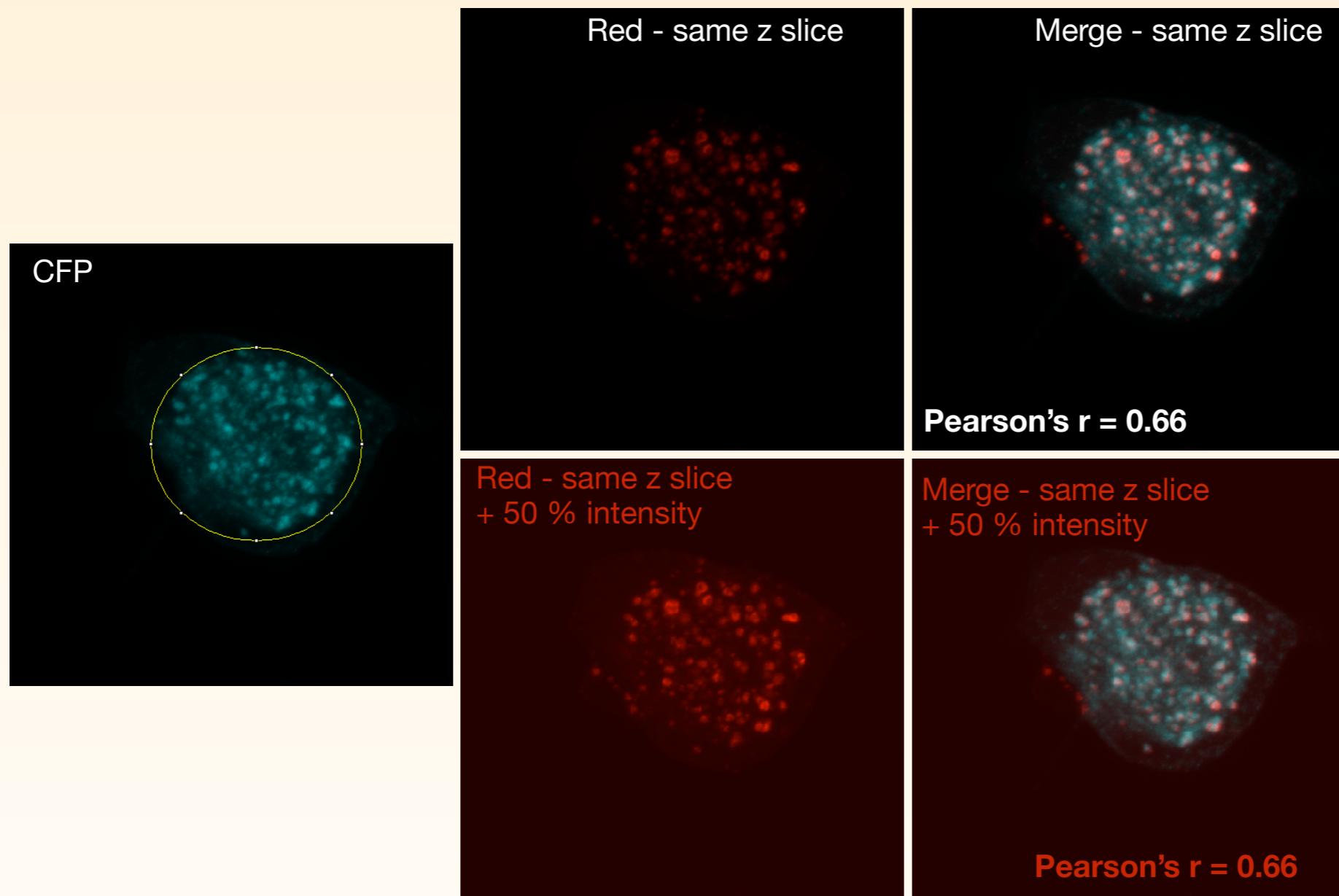
# Pearson's test and anti-correlation



- Pearson's test can also be used to establish when something is negatively colocalised. Can be shown by taking inverse of input image.
- For when something is being actively excluded from an area.

Source:

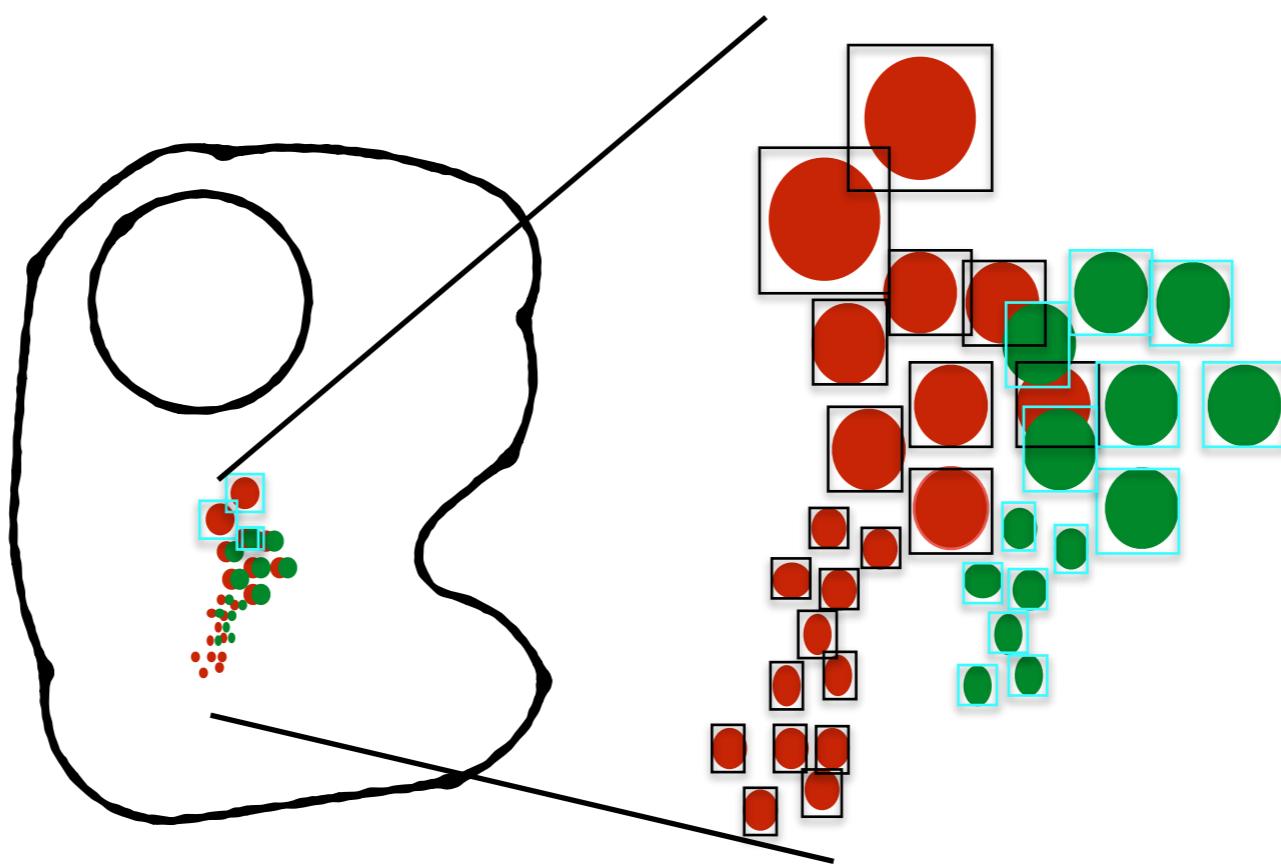
# Pearson's test is insensitive to global intensity



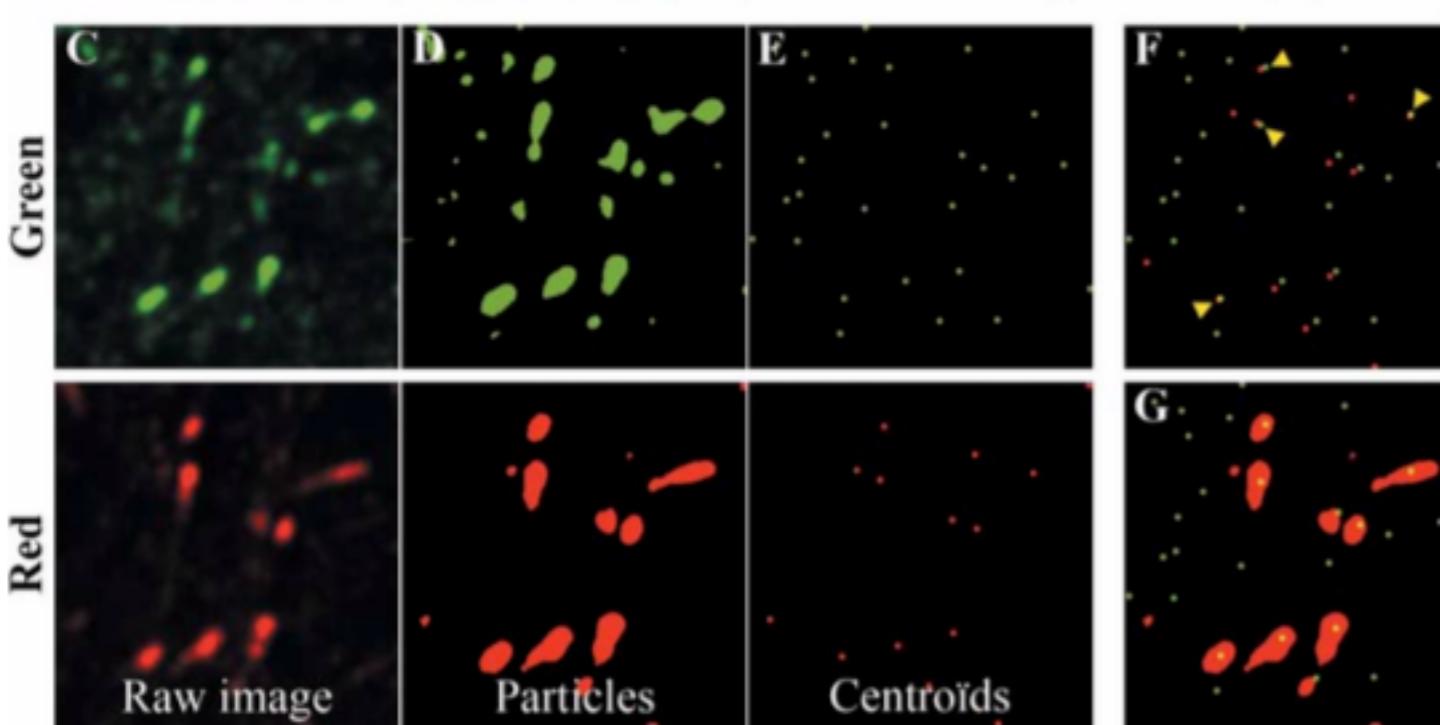
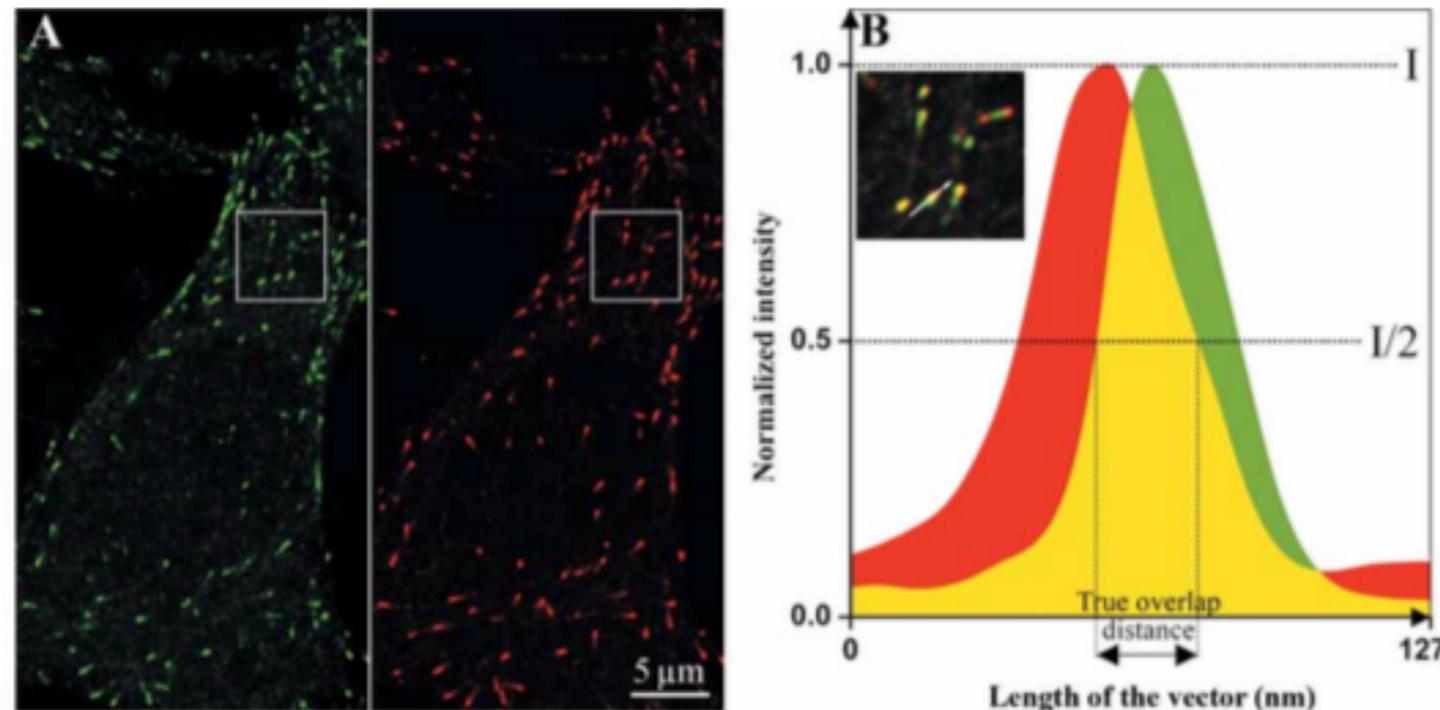
- Pearson's test is (within reason) insensitive to linear changes in intensity.
- This is good, it looks at trends rather than absolute values.
- This means expression variation between cells does not ruin experiment

Source:

# Object-based Colocalization



# Object-based colocalization



Many ways to do this.  
One way is to threshold each channel.

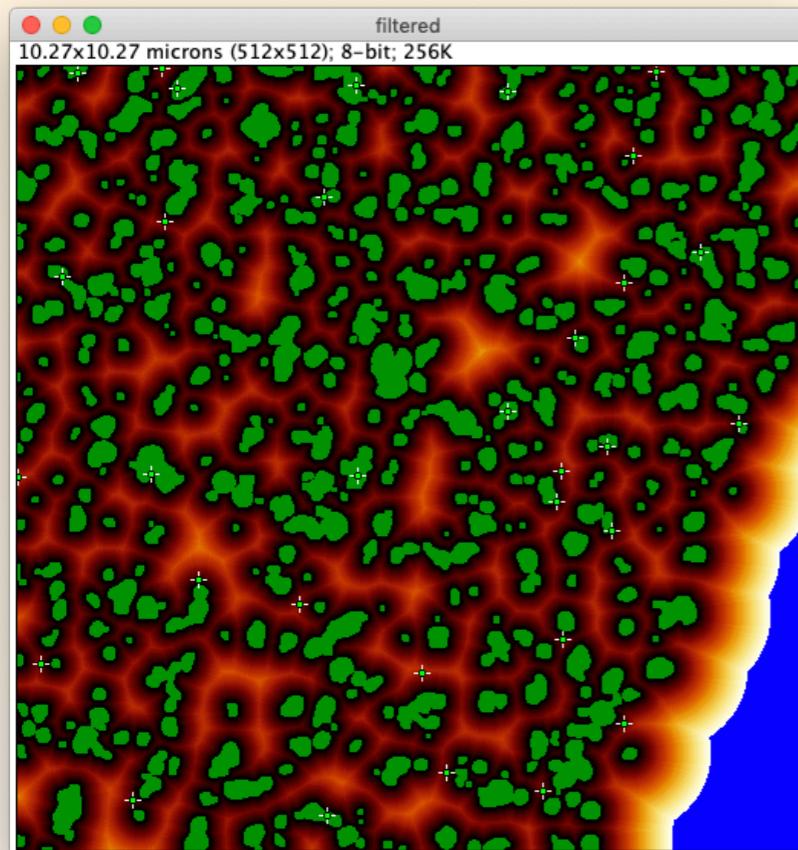
Next separate binary image into individual blobs.

Once separated, the blobs can be parameterised (e.g. centroid locations).

You can then compare the nearest neighbour distance between each channel for features.

Source: A guided tour into subcellular colocalization analysis in light microscopy  
S. BOLTE\* & F. P. CORDELIÈRES

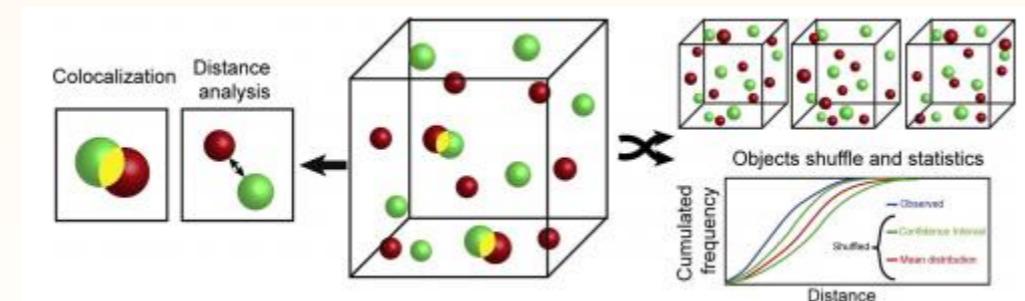
# Practical Object-based colocalization



Mostly in my experience this is solved using a custom macros or commercial software like Imaris.

Left) Screenshot from macro processing of image. The “Distance Transform” makes it straightforward to calculate nearest neighbour from parameterised locations.

Distance Analysis (DiAna) attempts to automate this with GUI.



Source: <http://bitplane.com> <https://imagejdocu.tudor.lu/plugin/analysis/>  
[distance analysis diana 2d 3d/start](#)

# **Summary of Conventional Approaches**

# Summary

## Cooccurrence (e.g. Mander's test):

- +ve Works when one species is much sparser than the other.
- +ve If there is a lot of black/empty space in analysis area.
- ve Requires thresholding of both channels.
- ve Sensitive to noise.

## Correlation (e.g. Pearson's test):

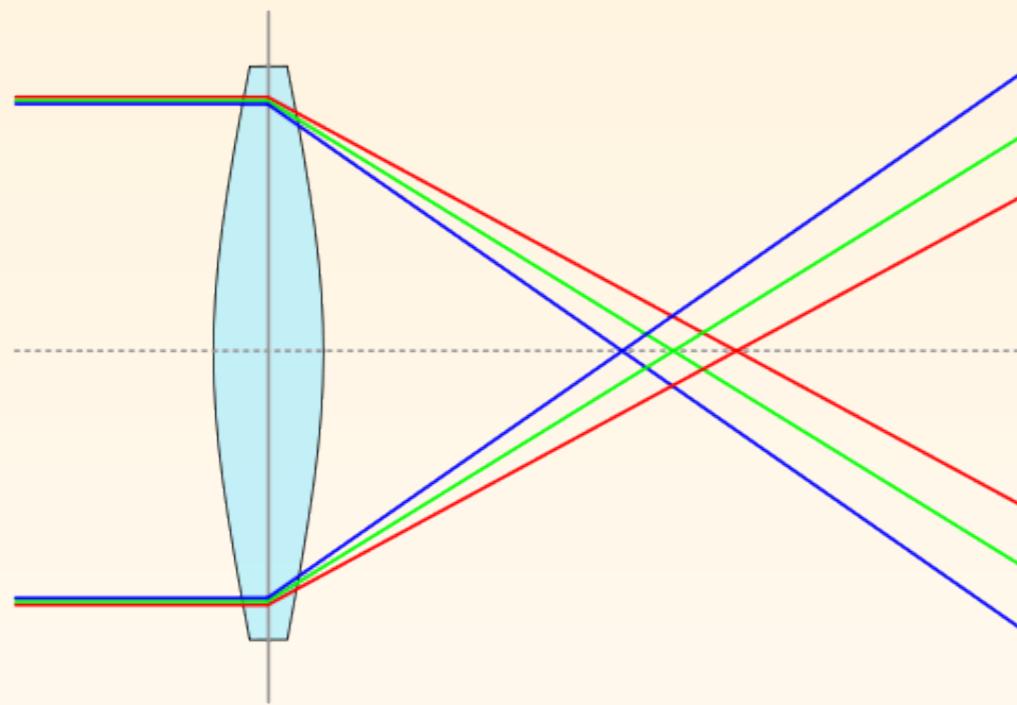
- +ve Very powerful analysis, no segmentation needed.
- +ve Highly resistant to global intensity changes.
- +ve Can be used for anti-correlation very easily.
- ve Both channel species must occupy similar pixel area.
- ve requires that black/empty space is very low.
- ve Sensitive to noise.

## Object-based (e.g. nearest neighbour analysis).:

- +ve Works when objects don't necessarily overlap.
- +ve With right object detector can be designed to be resilient to noise.
- ve Works best when objects are small and uniform.
- ve Requires thresholding of both channels or some kind of detection.

# Important notes: Chromatic abberation

Optics:

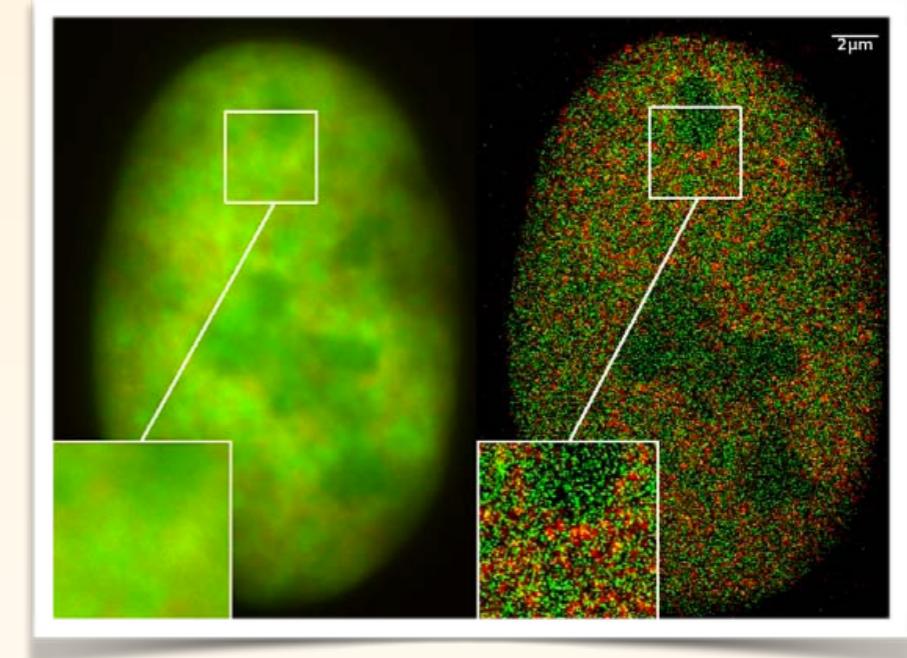
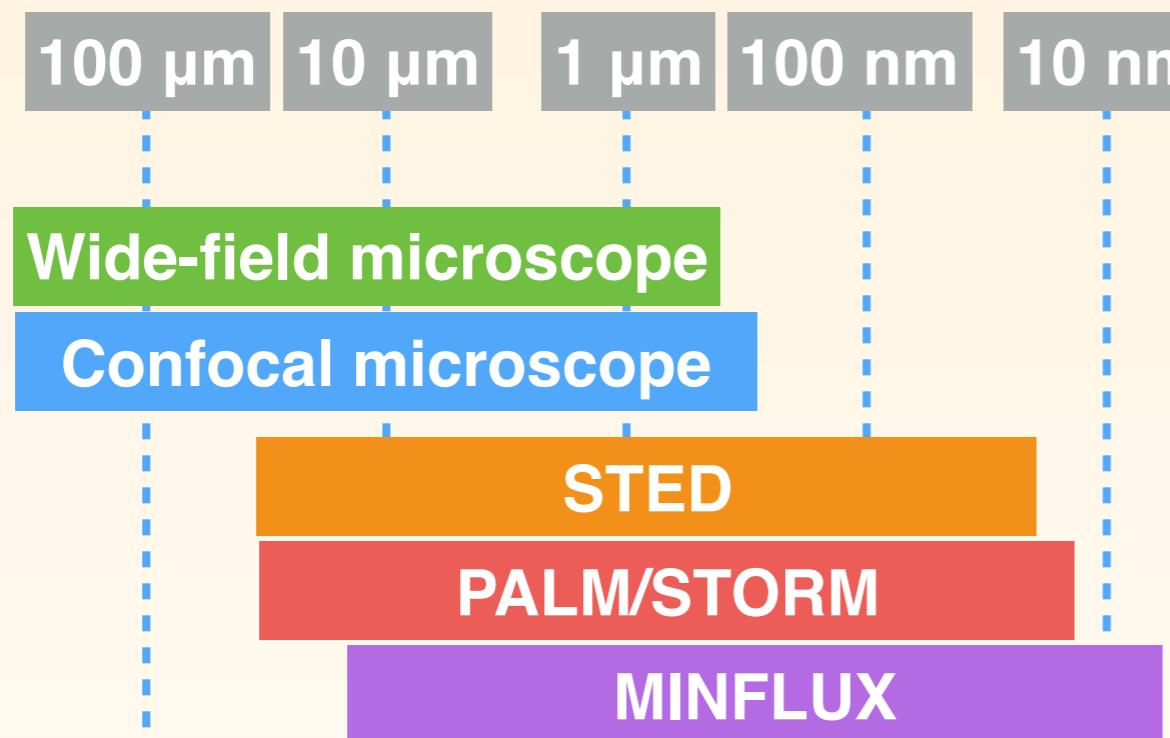


- A artefact of all optical lens is chromatic abberation.
- Important to correct for accurate colocalization.
- Refractive index of lens is wavelength dependent.
- More obvious for high-resolution imaging.
- Can be corrected for using careful bead calibration + other.

Source: [http://commons.wikimedia.org/wiki/File:Chromatic\\_aberration\\_convex.svg](http://commons.wikimedia.org/wiki/File:Chromatic_aberration_convex.svg)

# **Super-resolution colocalisation**

# Light microscopy beyond the diffraction limit

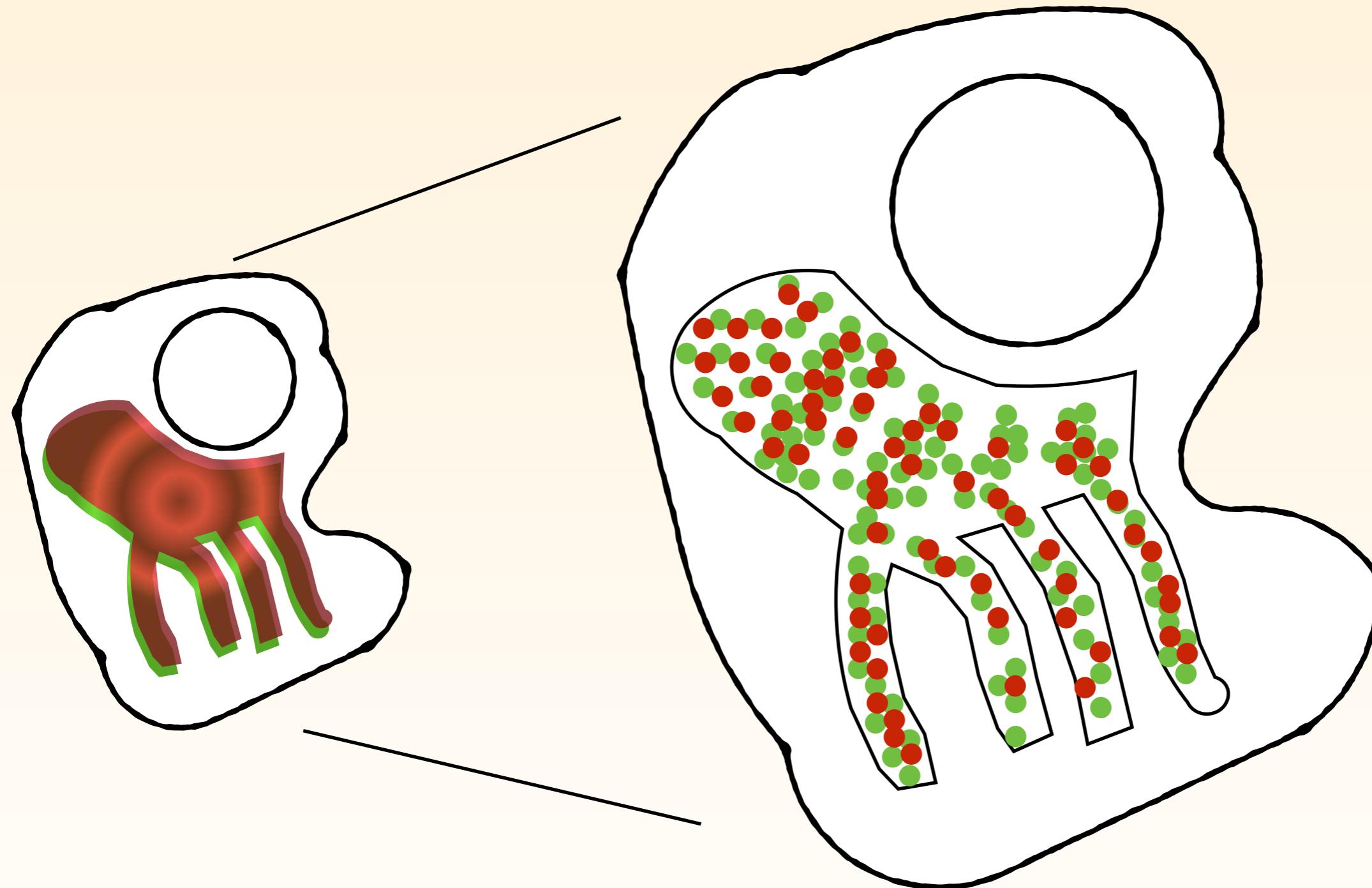


It is now fairly routine that we have access to microscopes capable of breaking the diffraction limit of light.

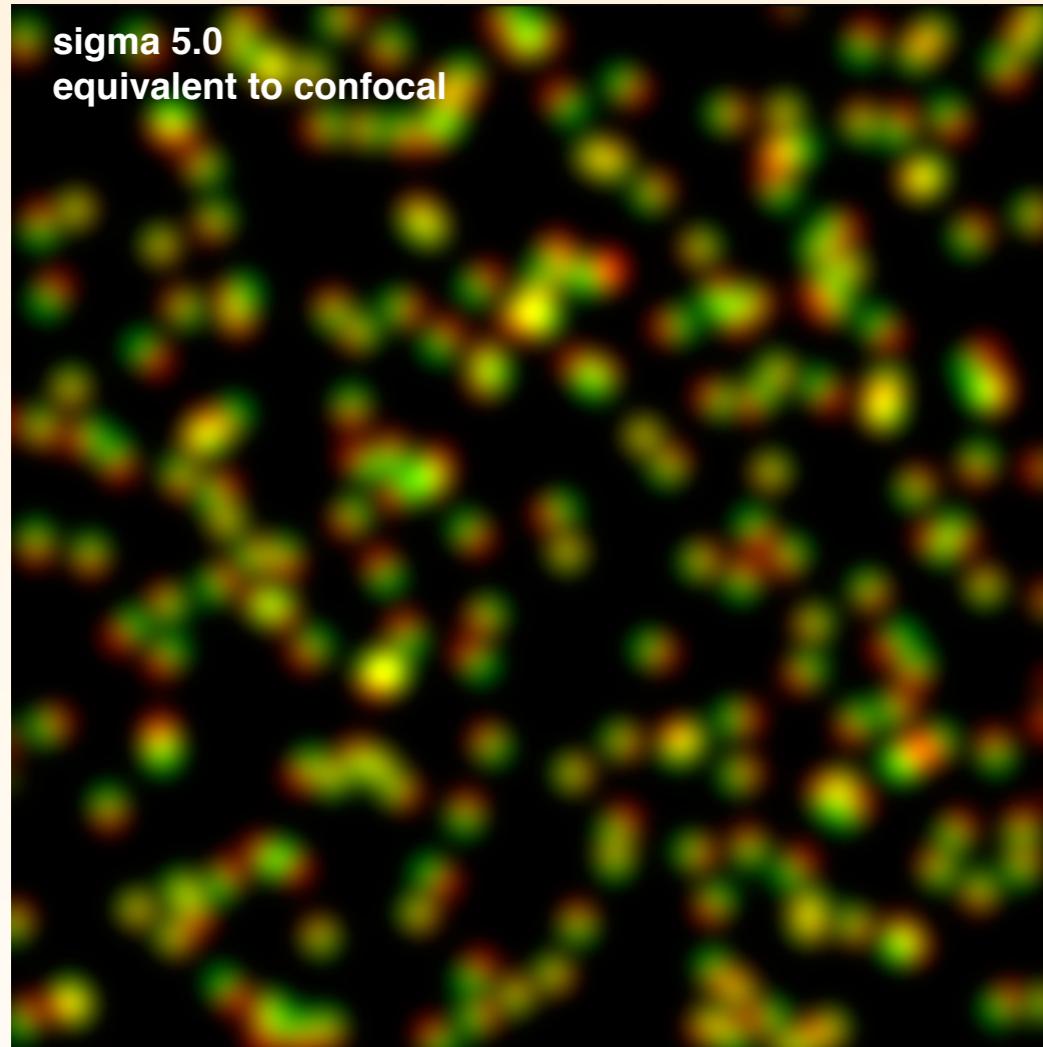
Different scales require different approaches for assessing colocalization.

Source: Adapted from <http://zeiss-campus.magnet.fsu.edu/articles/superresolution/introduction.html> [http://en.wikipedia.org/wiki/Super-resolution\\_microscopy](http://en.wikipedia.org/wiki/Super-resolution_microscopy)

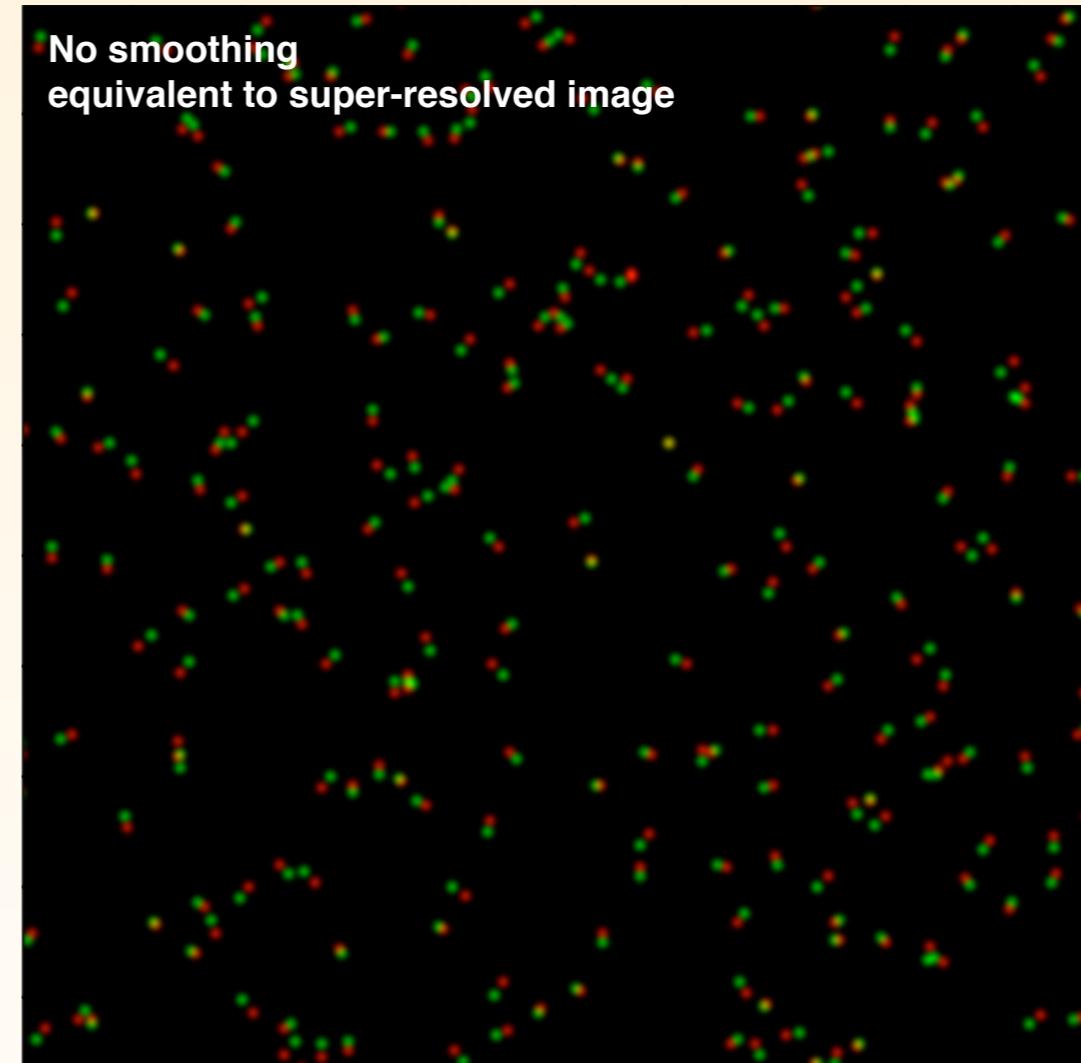
# At super-resolution protein localisation can appear sparse



# Colocalization metrics are very dependent on resolution



$r = 0.84$

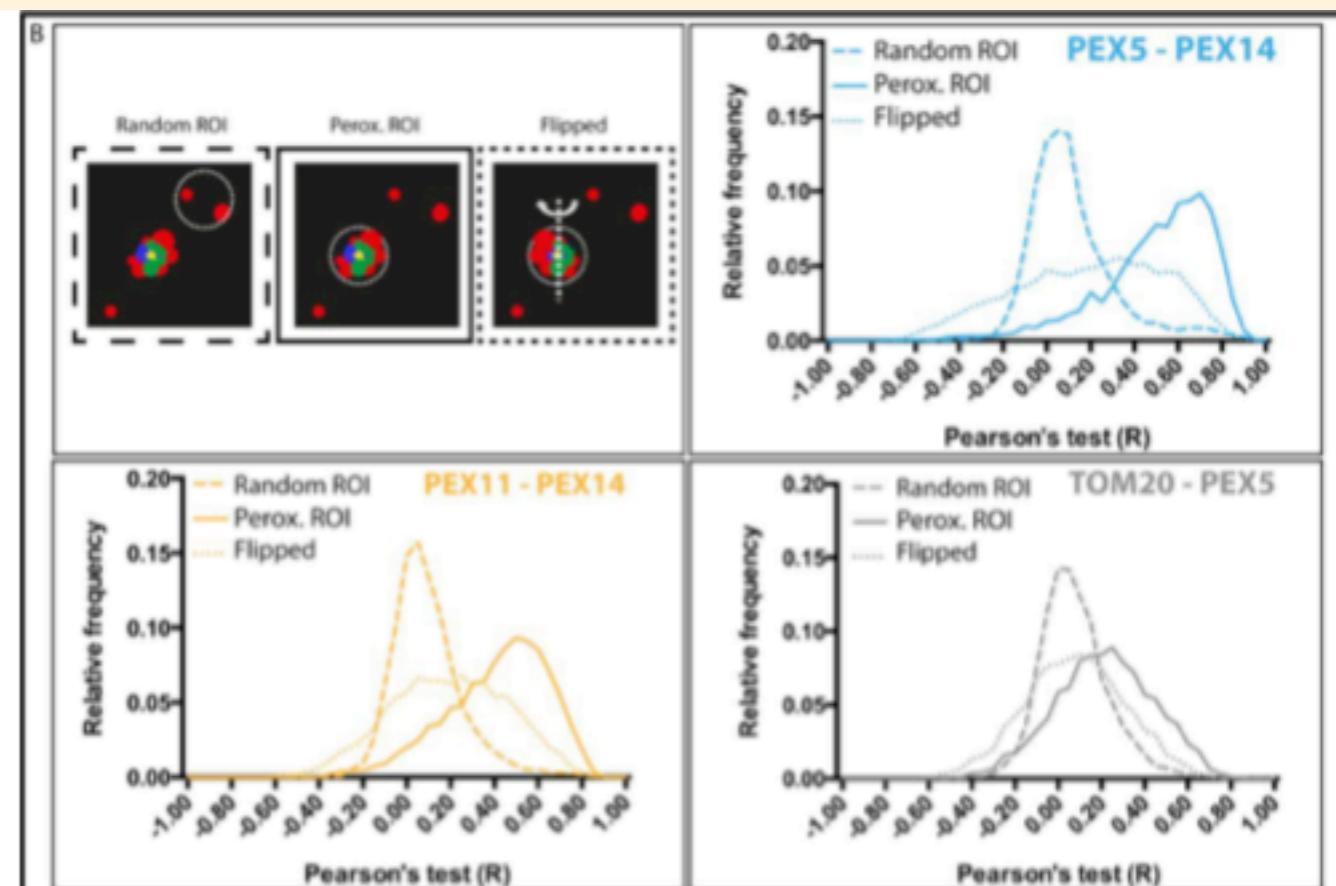
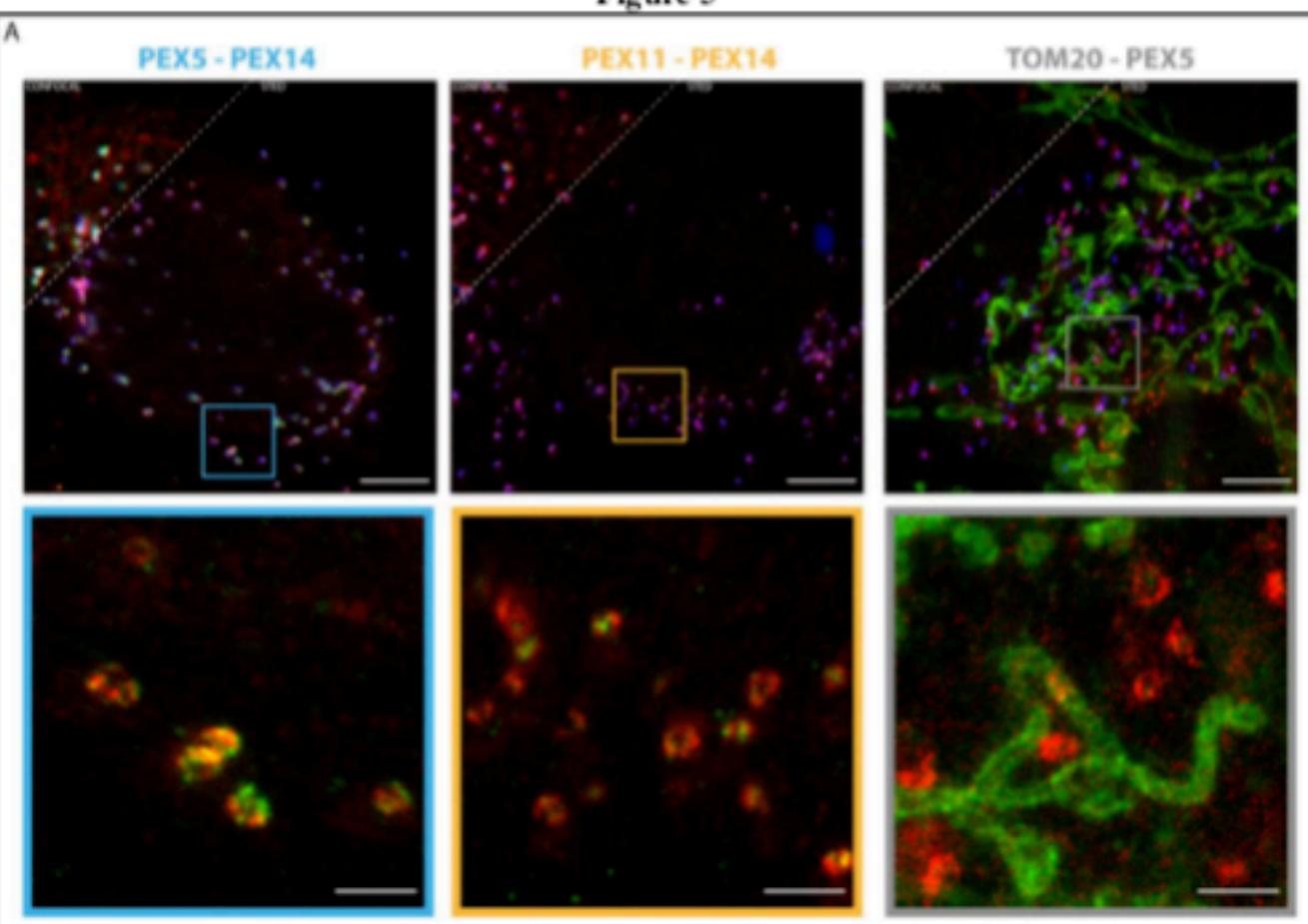


$r = 0.23$

- The Pearson's value ( $r$ ) changes with scale. This is not good. As the same underlying mechanism is present. There is also a lot of black pixels! which is bad. You have to choose your question carefully.

# With STED, sometimes you can still use the conventional metrics

Figure 5

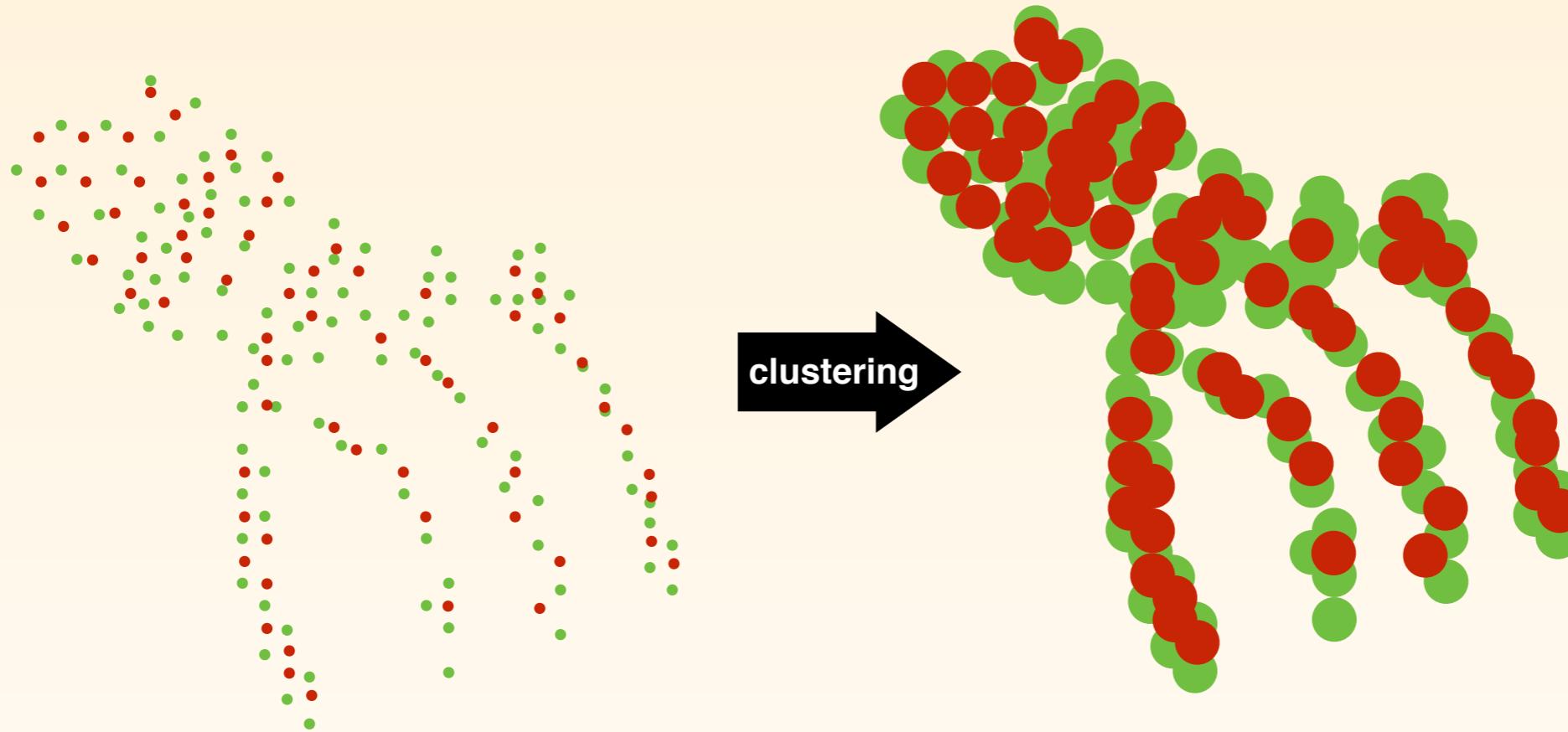


We changed the hypothesis region.  
We found that we could apply the  
Pearson's test on individual peroxisomes.

We used carefully designed controls to  
show our effect was not just due to random  
overlap.

Source: Super resolution microscopy reveals compartmentalization of peroxisomal membrane proteins Silvia Galiani1,\*  
Dominic Waithe2,\*  
Katharina Reglinski1, Luis Daniel Cruz-Zaragoza3, Esther Garcia2, Mathias P. Clausen1,4, Wolfgang  
Schliebs3, Ralf Erdmann3, Christian Eggeling1

# With STORM/PALM and high-res STED



With STORM/PALM we need to first cluster the data to find meaningful structures and then perform Object-based colocalization.

There are many possible ways as too how to do this and the approaches generally come together.

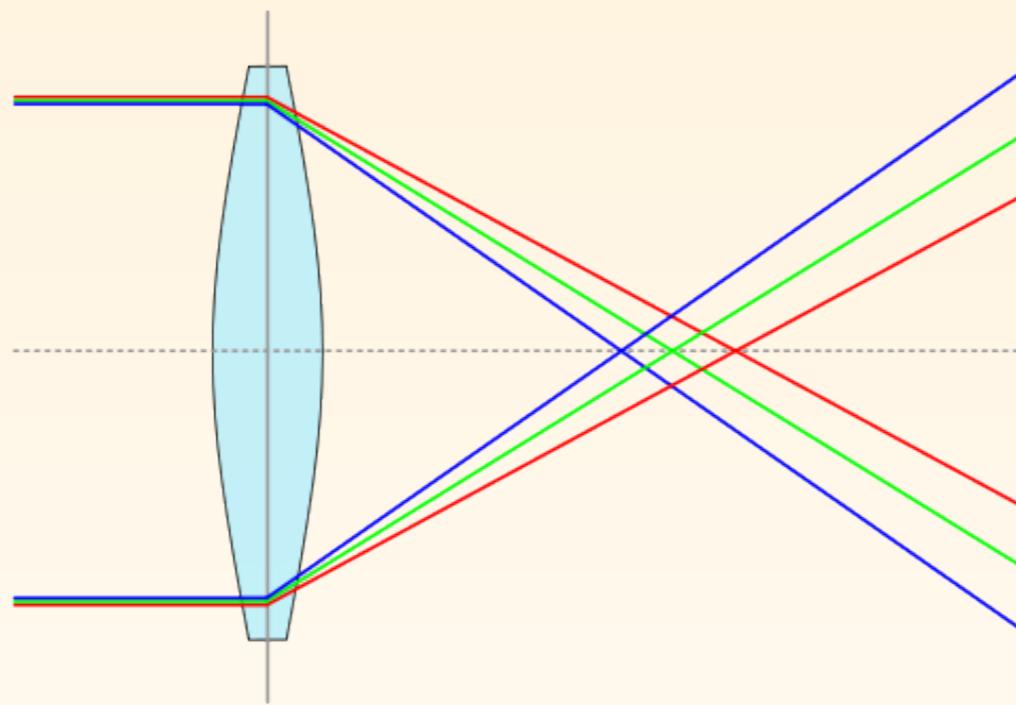
# Many papers on this subject....

- Challenges in quantitative single molecule localization microscopy. A. Shivanandan, H. Deschout, M. Scarselli, A. Radenovic FEBS Letters. Volume 588, Issue 19, 1 October 2014, Pages 3595-3602.
- ClusterViSu, a method for clustering of protein complexes by Voronoi tessellation in super-resolution microscopy. Leonid Andronov, Igor Orlov, Yves Lutz, Jean-Luc Vonesch & Bruno P. Klaholz
- True Molecular Scale Visualization of Variable Clustering Properties of Ryanodine Receptors Isuru Jayasinghe, Alexander H. Clowsley, Ruisheng Lin, Tobias Lutz, Carl Harrison, Ellen Green, David Baddeley, Lorenzo Di Michele, and Christian Soeller. Cell Reports.
- Bayesian cluster identification in single-molecule localization microscopy data Patrick Rubin-Delanchy, Garth L Burn, Juliette Griffié, David J Williamson, Nicholas A Heard, Andrew P Cope & Dylan M Owen
- Clus-DoC: a combined cluster detection and colocalization analysis for single-molecule localization microscopy data Sophie V. Pageon, Philip R. Nicovich, Mahdie Mollazade, Thibault Tabarin, and Katharina Gaus. Mol Biol Cell. 2016 Nov 7; 27(22): 3627–3636.
- MosaicIA: an ImageJ/Fiji plugin for spatial pattern and interaction analysis. Arun Shivanandan, Aleksandra Radenovic, and Ivo F Sbalzarini. BMC Bioinformatics. 2013; 14: 349.

# Registration

# Image Restoration

## Optics:

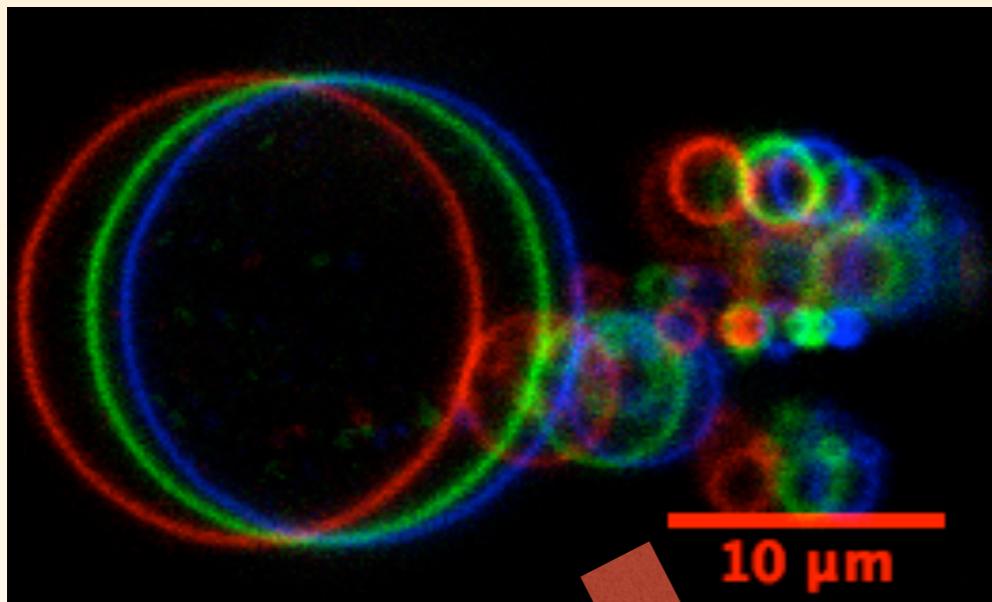


- A artefact of all optical lens is chromatic aberration.
- Refractive index of lens is wavelength dependent.
- More obvious for high-resolution imaging.
- Can be corrected for using careful bead calibration + other.

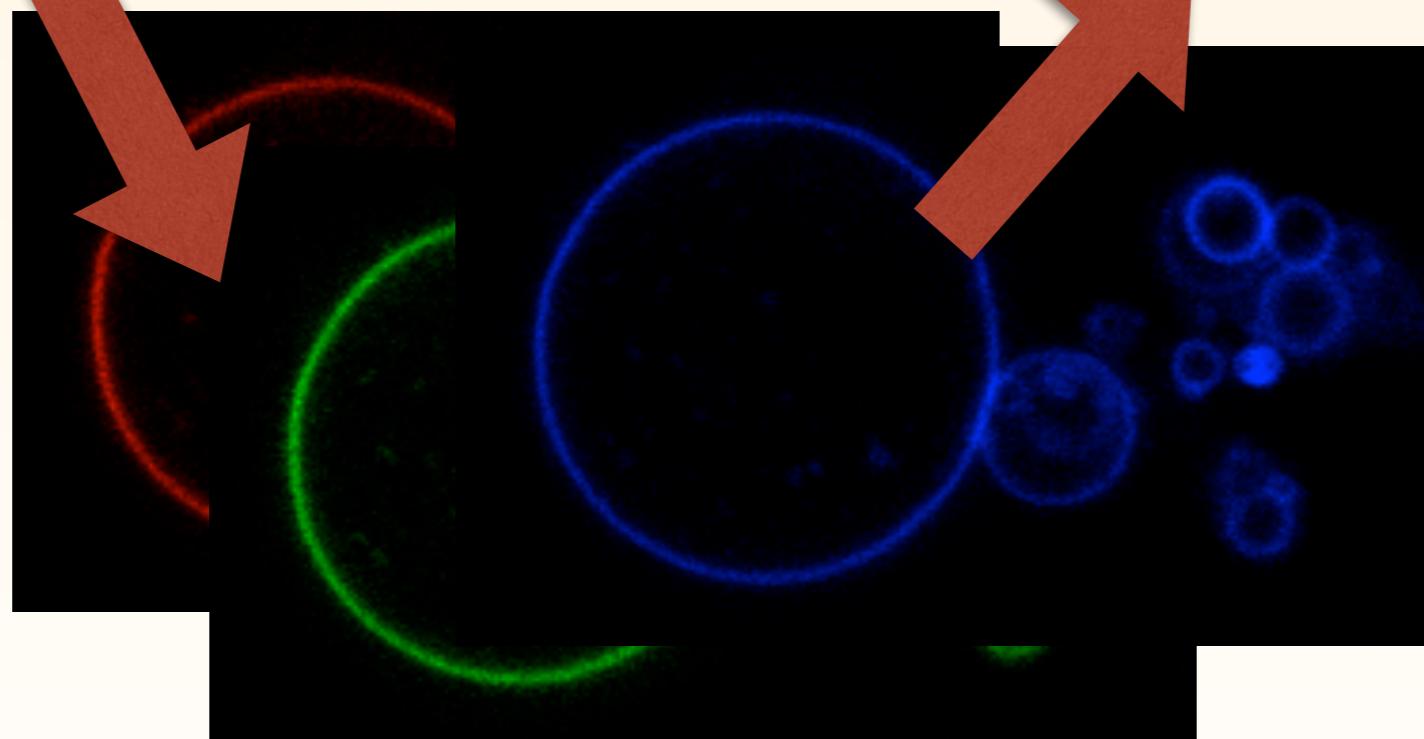
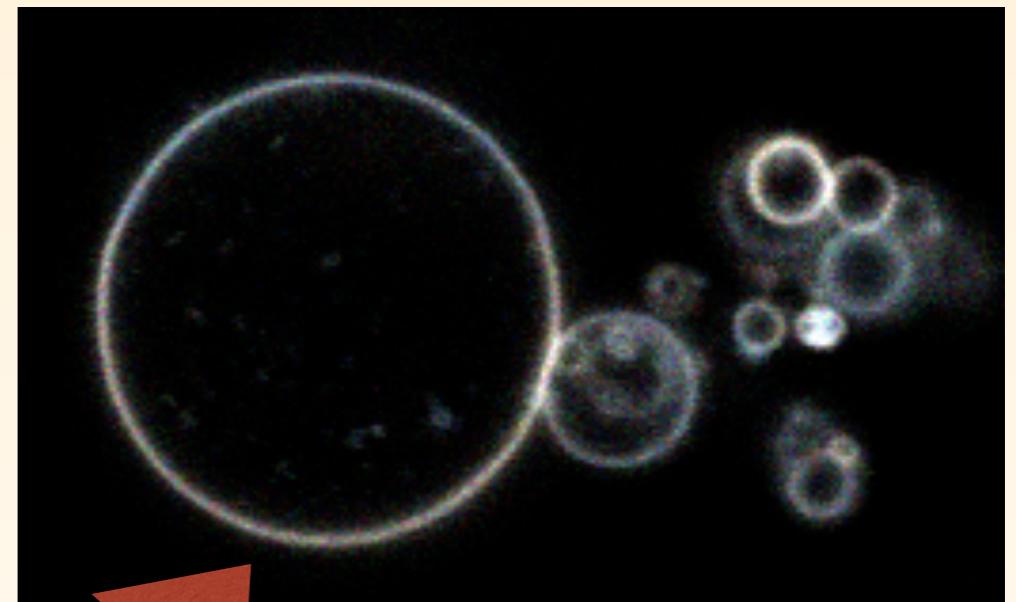
Source: [http://commons.wikimedia.org/wiki/File:Chromatic\\_aberration\\_convex.svg](http://commons.wikimedia.org/wiki/File:Chromatic_aberration_convex.svg)

# Colour alignment using colocalisation

Artificially Unaligned



Aligned

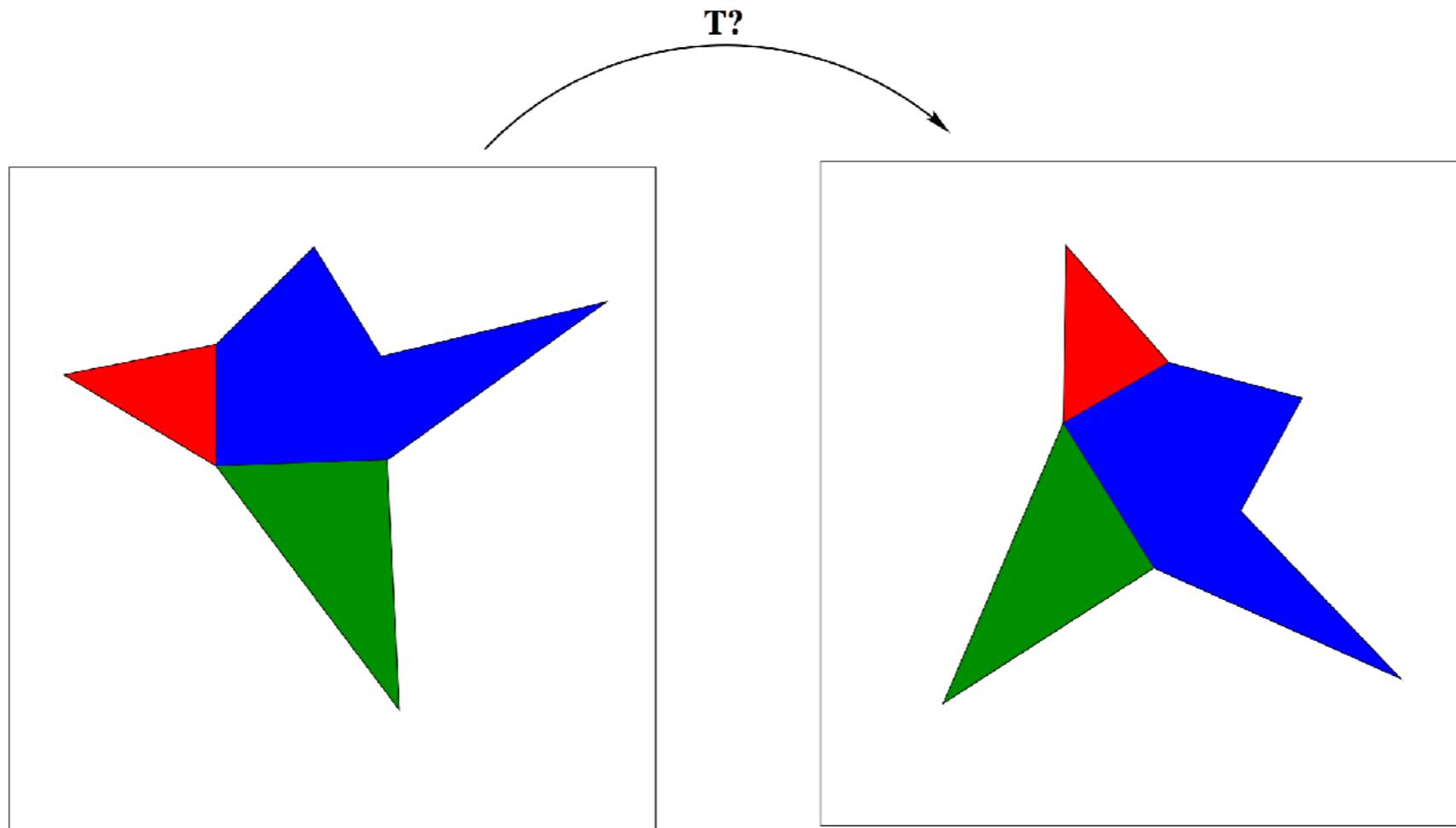


You will see this in your practical later.

Source: Image, Erdinc Sezgin.

# Pixel-based registration

## Example



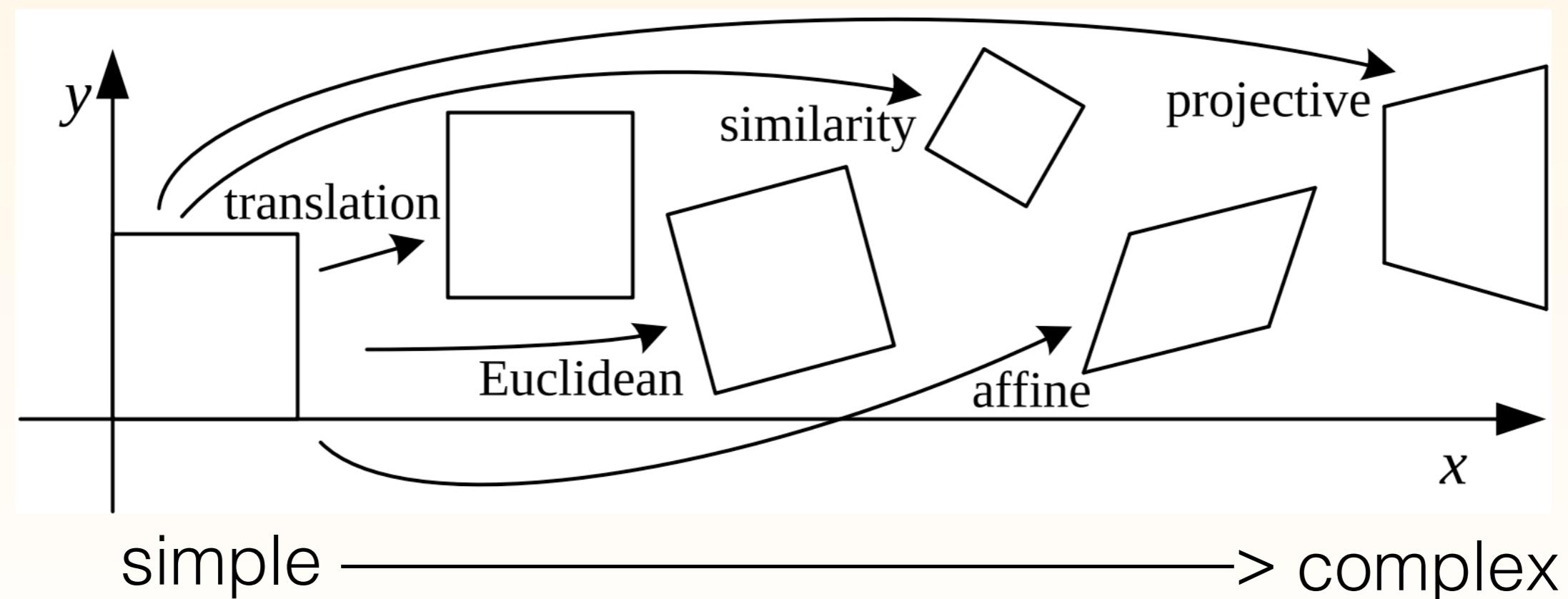
We want to register/align these two stars so that they overlap perfectly.  
To do this we want to find  $T$ , which is the transformation function.

Source: Sébastien Ourselin, UCL

# T can take a variety of forms.

Transformation	Matrix	#DoF	Preserves	Icon
translation	$[I t]_{2x3}$	2	orientation	
rigid (Euclidean)	$[R t]_{2x3}$	3	lengths	
similarity	$[sR t]_{2x3}$	4	angles	
affine	$[A]_{2x3}$	6	parallelism	
projective	$[\tilde{H}]_{3x3}$	8	straight lines	

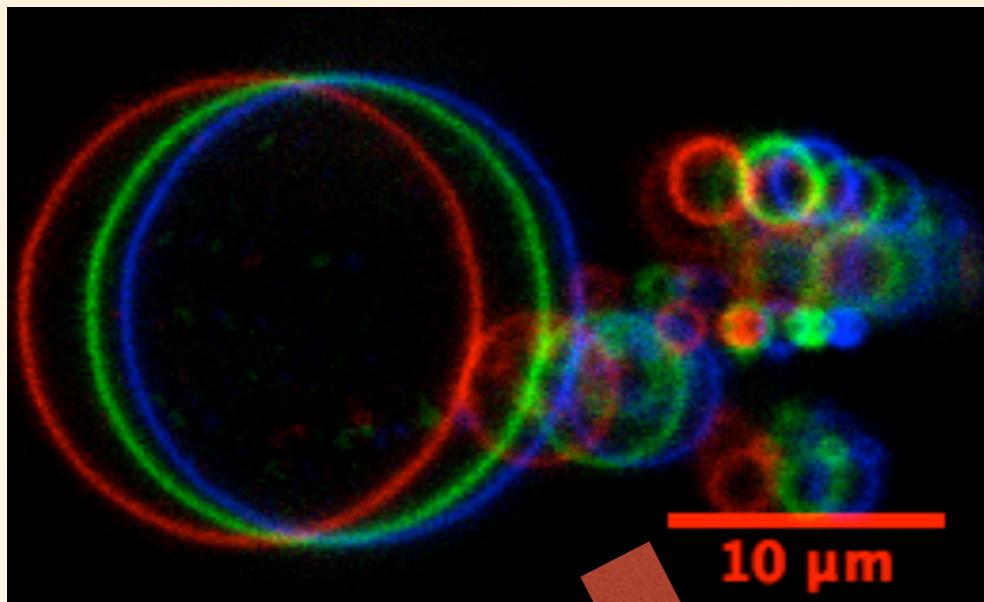
Depending on the complexity of transformation we need, we will have to optimise the parameters for increasing large matrices to find the transformation between our images.



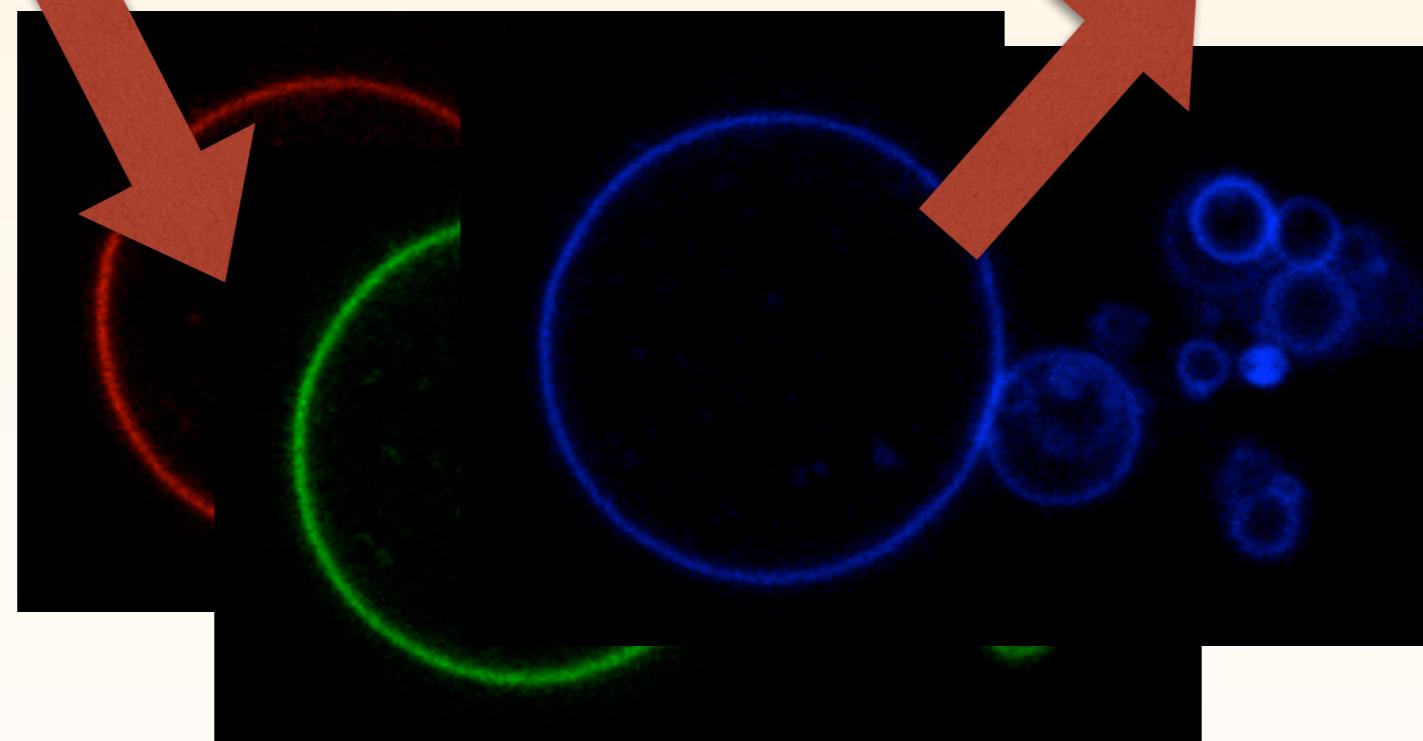
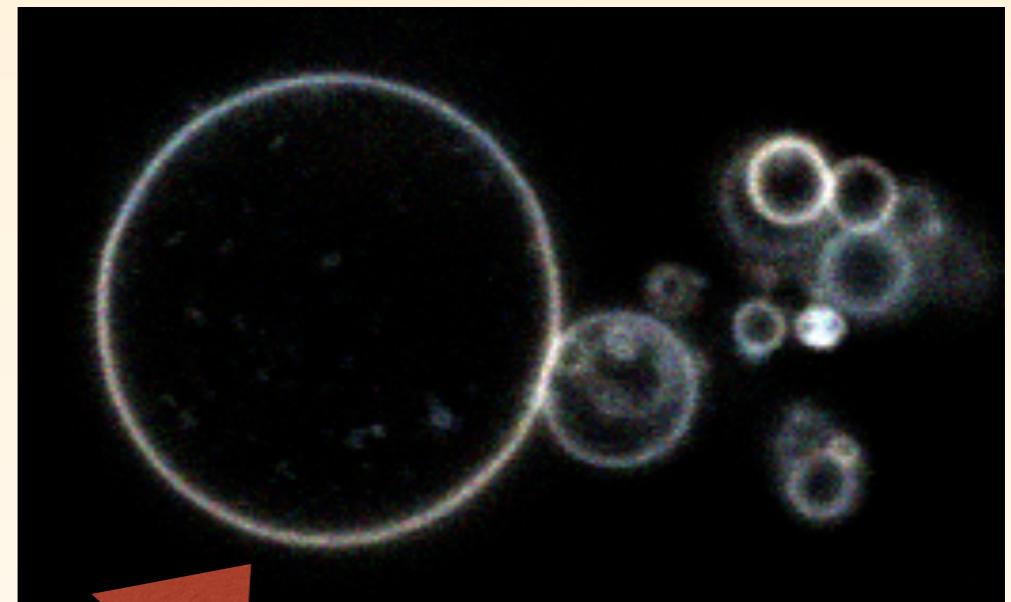
Source: <https://wiki.tum.de/display/lfdv/Spatial+Transformer+Networks>

# This is a simple translation transformation.

Artificially Unaligned



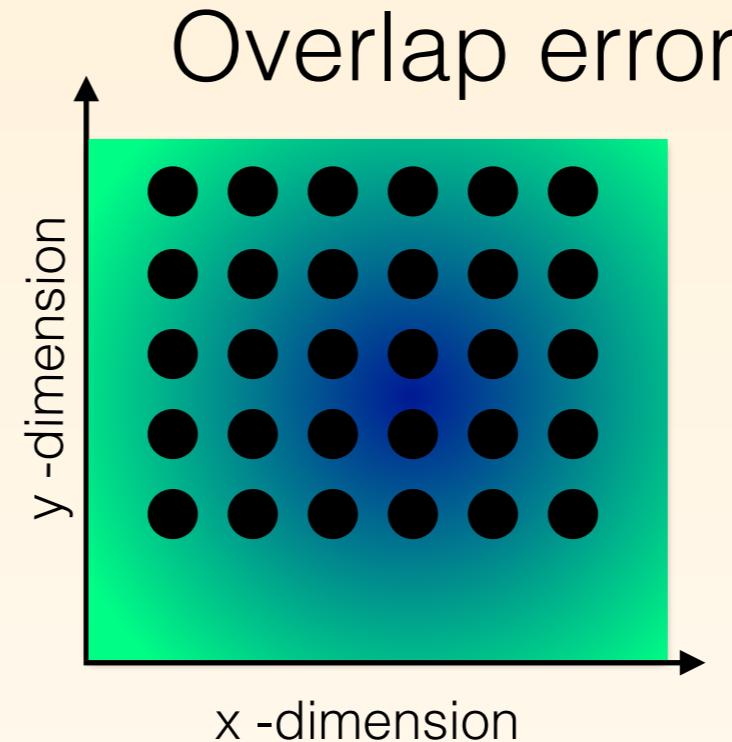
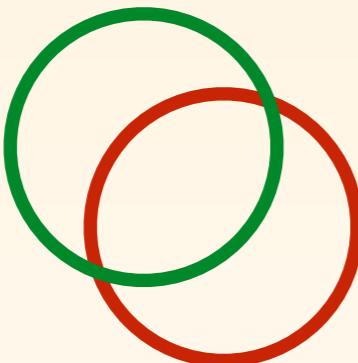
Aligned



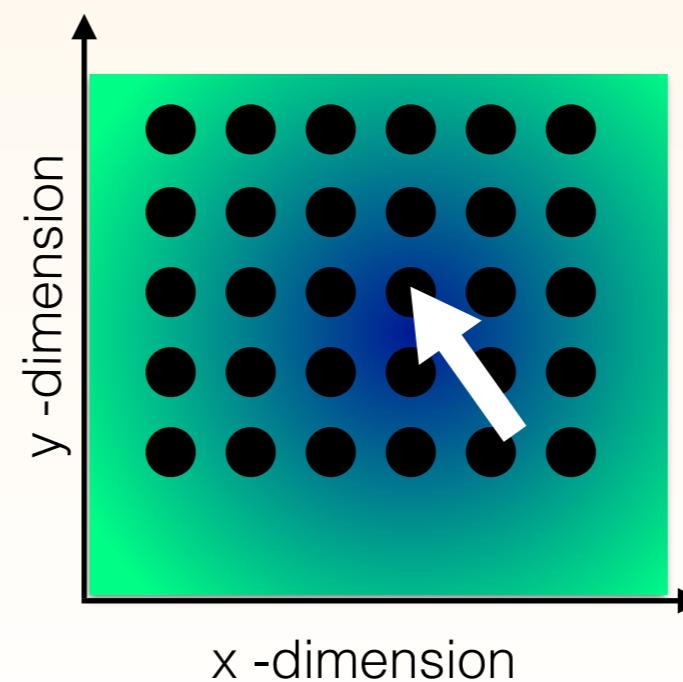
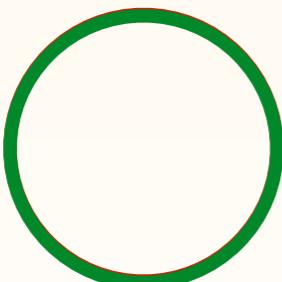
You will see this in your practical later.

Source: Image, Erdinc Sezgin.

# Optimising our transformation.



Optimum overlap. Should give us the lowest point on our error plot



# Many possible loss functions but...

## Correlation Coefficient

- Image normalisation

$$\tilde{I}(x_k) = \frac{I(x_k) - \bar{I}}{\sigma(I)}$$

with  $\bar{I} = \frac{1}{n} \sum_{k=1}^n I(x_k)$  and  $\sigma(I) = \sqrt{\frac{1}{n} \sum_{k=1}^n [I(x_k) - \bar{I}]^2}$

- We want a measure which is invariant by linear transformation of the intensity

$$J = aI + b$$

$$CC = \frac{Cov(I, J)}{\sqrt{Var(I) Var(J)}}$$

This is the same as the Pearson's test but theoretically 'I' and 'J' are 2d not 1d.

Works well, just as the Pearson's test does for assessing similarity of signal

Source:

# **Summary of Talk**

# Some take home messages.

- Colocalization is a popular technique which is used frequently in the imaging sciences.
- Learn to diagnose the needs of your experimenter to pick the right approach.
- Be aware that scale will influence your outcome.
- Understand your approaches so as to understand their weaknesses.

# Thanks for your time.

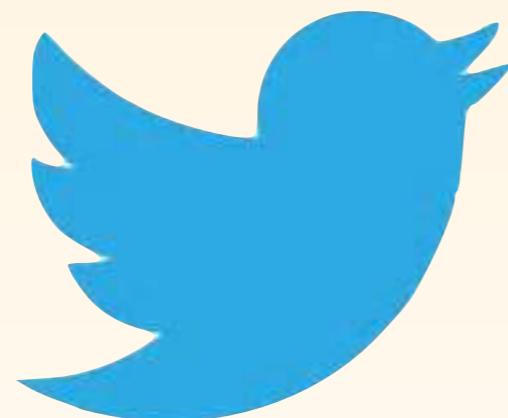
## For these slides and more:

<https://github.com/IAFIG-RMS/Bioimage-training>

<https://twitter.com/dwaithe>



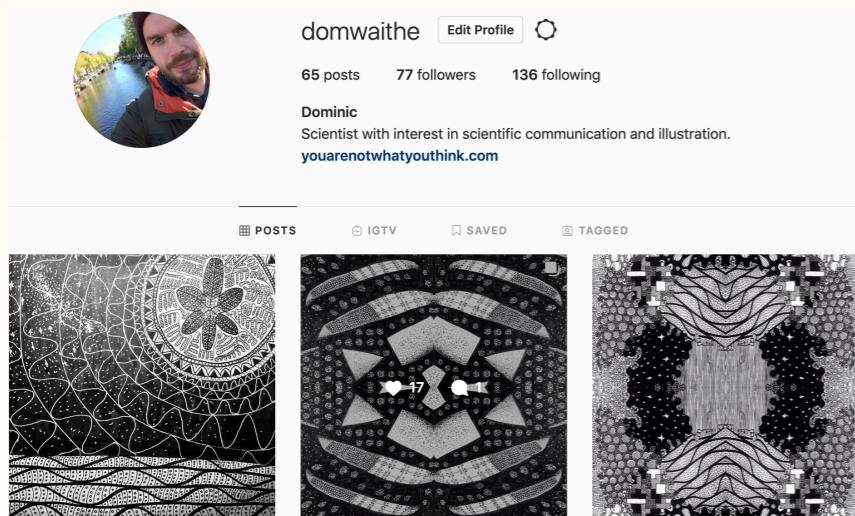
A screenshot of a Twitter profile for 'Dominic Waithe' (@dwaithe). The profile picture shows a man with a beard. The bio reads: 'Scientist with interest in scientific communication and illustration. youarenotwhatyouthink.com'. Statistics below show 401 tweets, 157 following, and 196 followers.



<https://github.com/dwaithe>



<https://instagram.com/dwaithe>



A screenshot of an Instagram profile for 'domwaithe'. The bio reads: 'Scientist with interest in scientific communication and illustration. youarenotwhatyouthink.com'. The profile picture is a circular photo of a man. Statistics show 65 posts, 77 followers, and 136 following. Below the bio are three small square images showing scientific illustrations.

[dominic.waithe@imm.ox.ac.uk](mailto:dominic.waithe@imm.ox.ac.uk)



UK Research  
and Innovation

