Co-localization analysis of microscopy images:

Manders, Costes, Ripley, and spatial statistics

2023-11-29 postdoctoral training (T32)

Simon Flyvbjerg Nørrelykke

HARVARD MEDICAL SCHOOL | BLAVATNIK INSTITUTE SYSTEMS BIOLOGY

Image Analysis Collaboratory

CENTER FOR COMPUTATIONAL BIOMEDICINE
HARVARD MEDICAL SCHOOL

Image Analysis Collaboratory

CCB Seminar Series 2023

Special: The Center for Computational Biomedicine (CCB) is hosting a joint seminar series with Image Analysis Collaboratory (IAC) at Harvard Medical School with a focus on best practices and leading tools for quantitative analysis of biomedical images.

iac.hms.harvard.edu

Speaker: Beth Cimini, Ph.D.

Associate Director for Bioimage Analysis and a CZI Imaging Scientist in the Imaging Platform at the Broad Institute

Topic: Using high content analysis and deep learning to make the most of your microscopy

Date: Monday, December 11, 2023

Where: Gordon Hall, 106 Waterhouse Conference Room

Time: 10:00 AM – 11:00 AM ET

Virtual: Zoom link

In the age of the digital camera, microscopy images constitute a fantastically rich source of quantitative data. Yet, it currently remains difficult for most scientists to mine quantitative data from these images easily such that they can answer their important biological questions. In this talk, we will discuss open source tools that make quantitative image analysis both easier and more reproducible, as well as bioinformatic approaches allowing users to extract novel connections from their data.

SCAN ME

HARVARD MEDICAL SCHOOL

Image Analysis Collaboratory

About

- **Harvard Medical School**, Lecturer, 2022—present
  - *Director of the Image Analysis Collaboratory*, 2022—present
- **ETH Zurich**, Lecturer, 2015-2022
  - *Head of Image and Data Analysis Group*, 2012–2022
- **Europe**, s*elf-employed*, 2011–2012
- **Princeton University**, Department of Molecular Biology, *visiting fellow*, 2007–2010
- **Max-Planck Institute** for the Physics of Complex Systems, *visiting scientist*, 2004–2007
- **LENS** (European Laboratory for Non-Linear Spectroscopy), *post doc*, 2003–2004
- **Niels Bohr Institute**, Denmark, *PhD in bio-physics*, 2002

Nationality: Danish. Languages: Danish, English, German, some Italian/Spanish

Content

After these ~90 minutes you will have a better idea about

1. What the Image Analysis Collaboratory is
2. **Manders'** coefficients and **Costes'** randomization
3. **Object** based and **Spatial statistics** beyond colocalization
4. Some **software** you can use, free & commercial

What is the Image Analysis Collaboratory?

Group of Bioimage Analysts

Hosted by department of Systems Biology

Located in Cell Biology (LHRRB 105) and Sys Bio (Arm 531D)

Works closely with local microscopy facilities

Collaborates with any department *on the Quad*

What is a Bioimage Analyst?

Expert in **image analysis** and **machine learning**

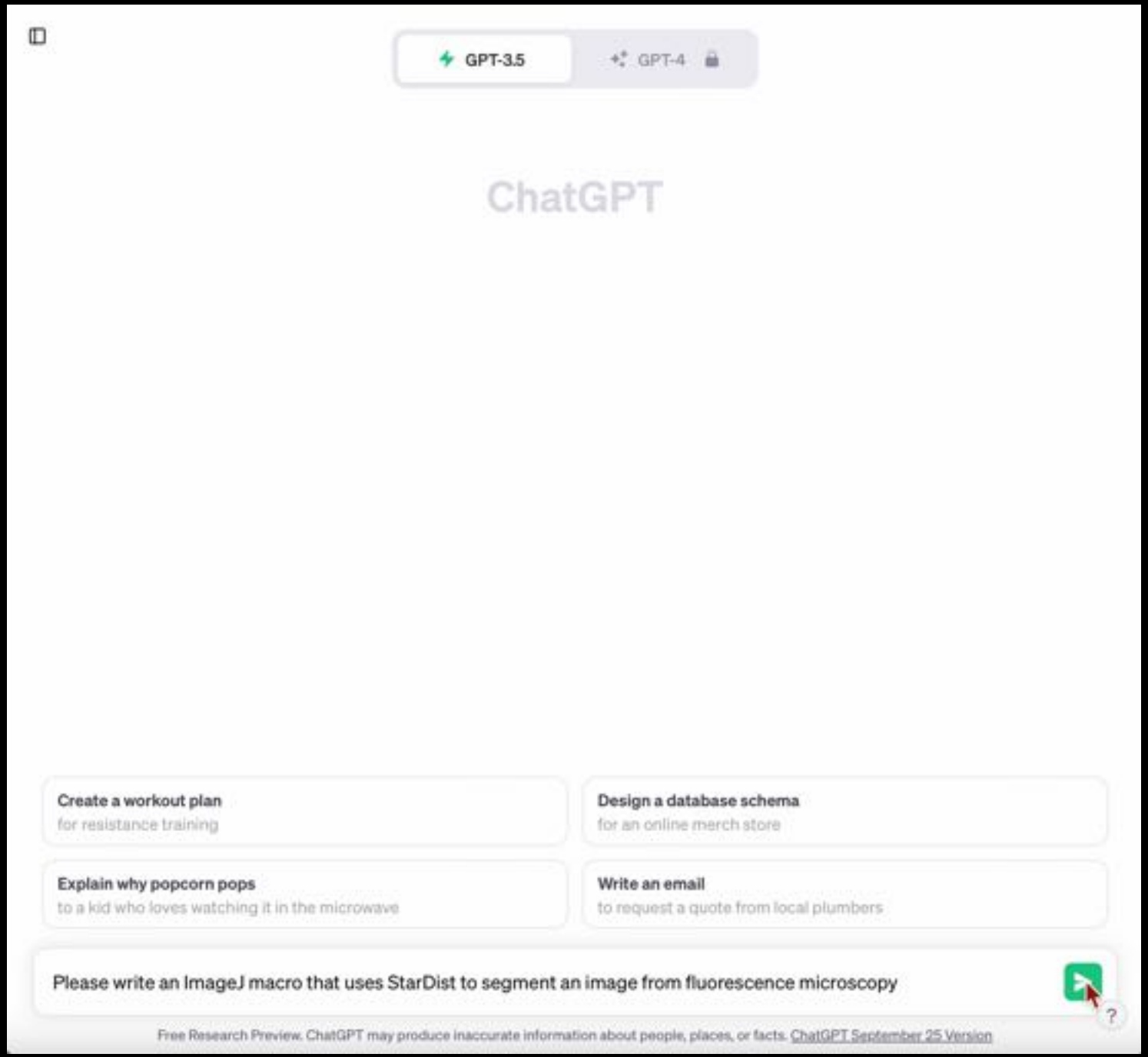
Fluent in one or more **programming languages**

Knowledgeable about **microscopy** and **statistics**

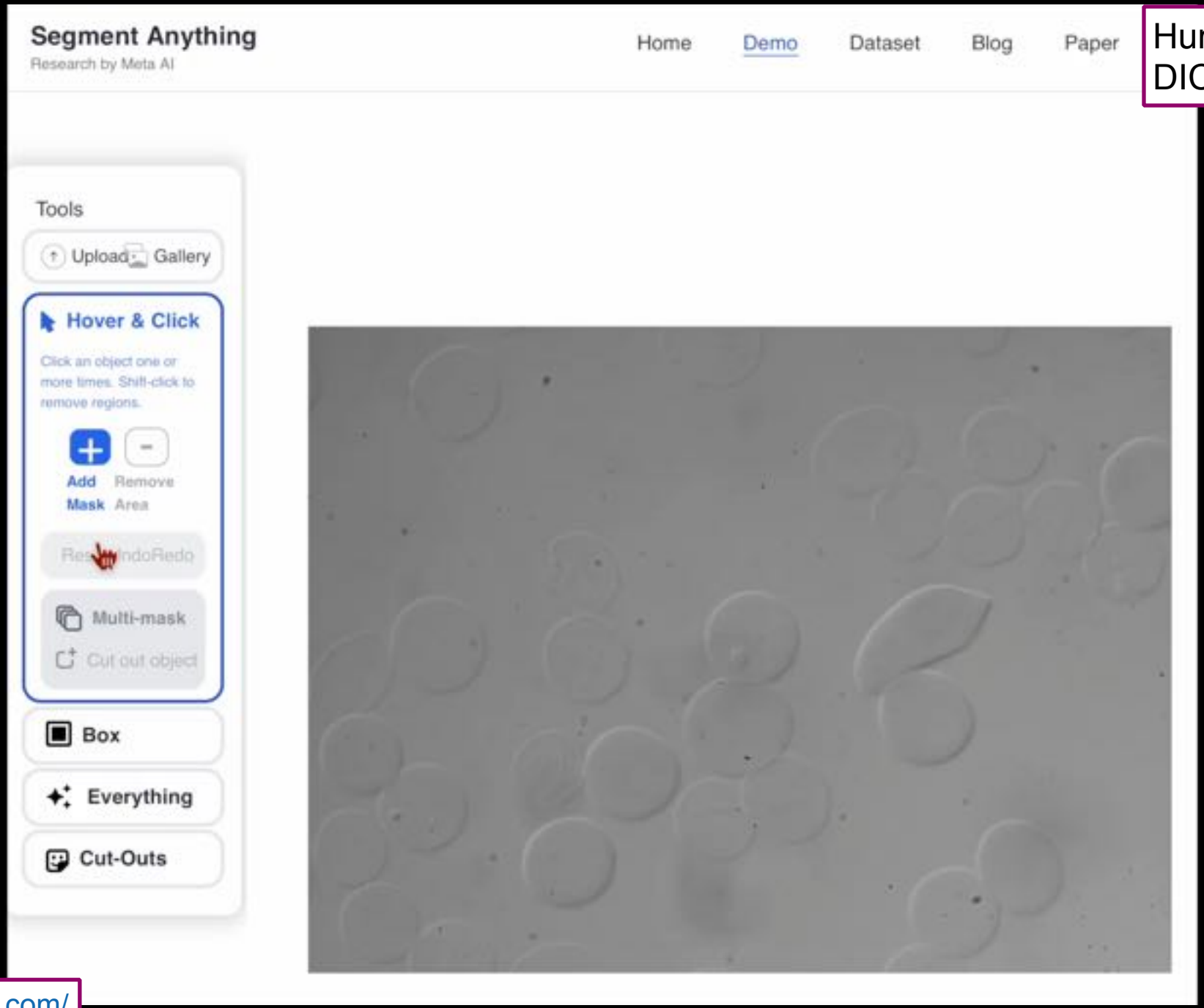
Conversational in **biology**

Intuitive feeling for **data**

(Didn’t go to Facebook, Google, Apple, biotech)



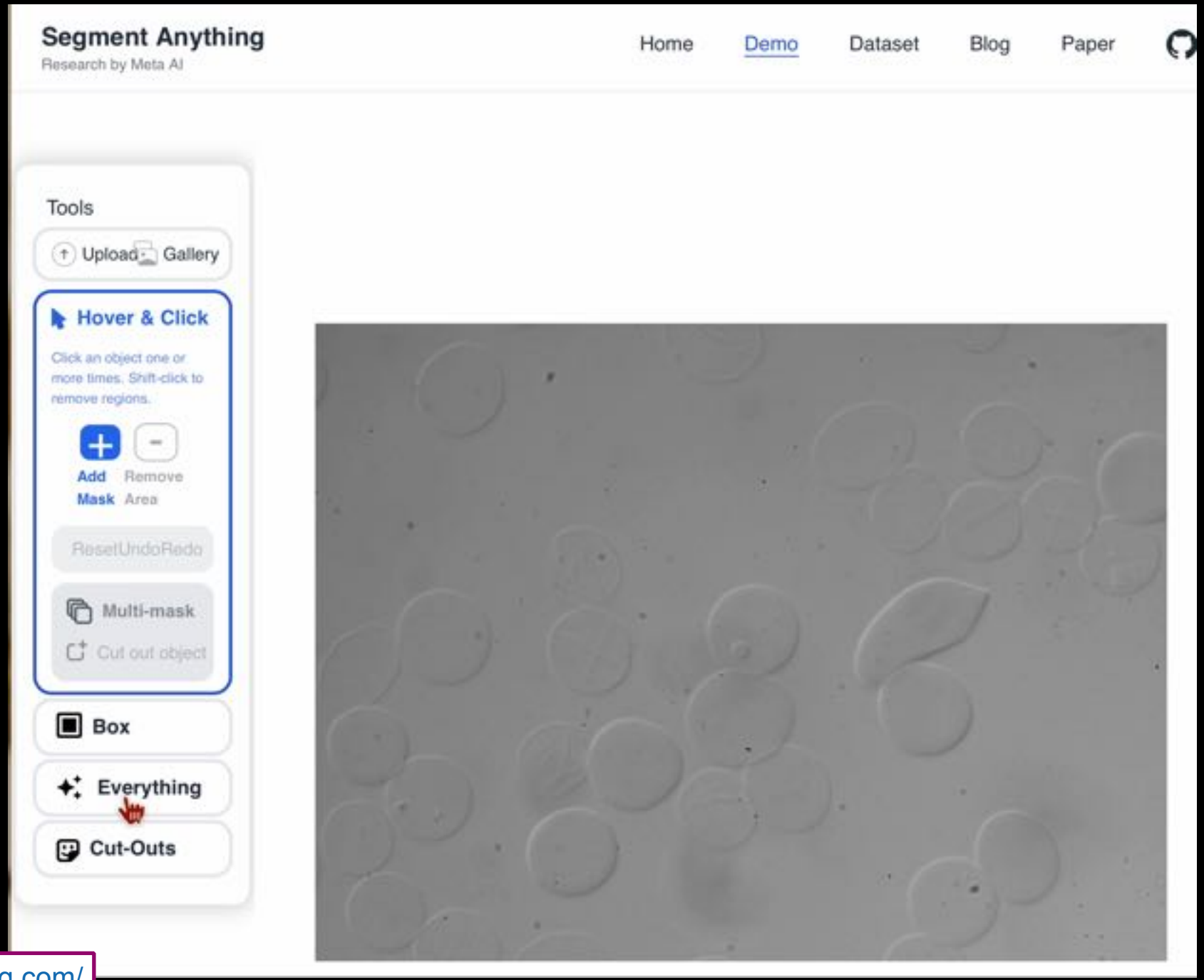
https://chat.openai.com/



Human red blood cells
DIC microscopy

"BBBC009v1 from the Broad Bioimage Benchmark Collection [Ljosa et al., *Nature Methods*, 2012]."

https://segment-anything.com/



https://segment-anything.com/

Mission

*Reproducible extraction of **meaningful** information from images*

Collaborate on projects with research labs (and hospitals)

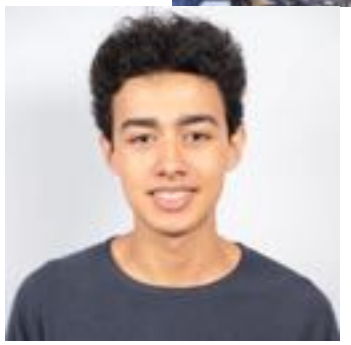
Train the next generation of bioimage analysts

Teach bioimage analysis to life scientists

Build and support bioimage analysis communities

***Objective:** Make as many Quad-groups as happy as possible!*

Who we are, currently



Simon, PhD, director, lecturer

Physicist (theory & experiments); bioimage analyst (past 10 years)

Ranit Karmakar, PhD, specialist postdoc

Computer Engineer

Joined August 2023

Antoine Ruzette, MSc, researcher

Bioengineer, bioinformatician (shared with Sean Megason)

Joined ~June 2023

Assil Achour, research intern

Computer Scientist

Joined September 2023

Who we are, incoming

Named, PhD, specialist postdoc

2024-Q1

Named, PhD, staff

2024-Q1

Unnamed, specialist postdoc(s) or staff

CS, physics, comp bio, …

5-year funding through ARPA-H

IAC Founding and Funding

Started: Operating since mid-September 2022

Now: Supported by internal HMS-Foundry grant

Sean Megason and Sahand Hormoz (~2019)

Future: Working on it

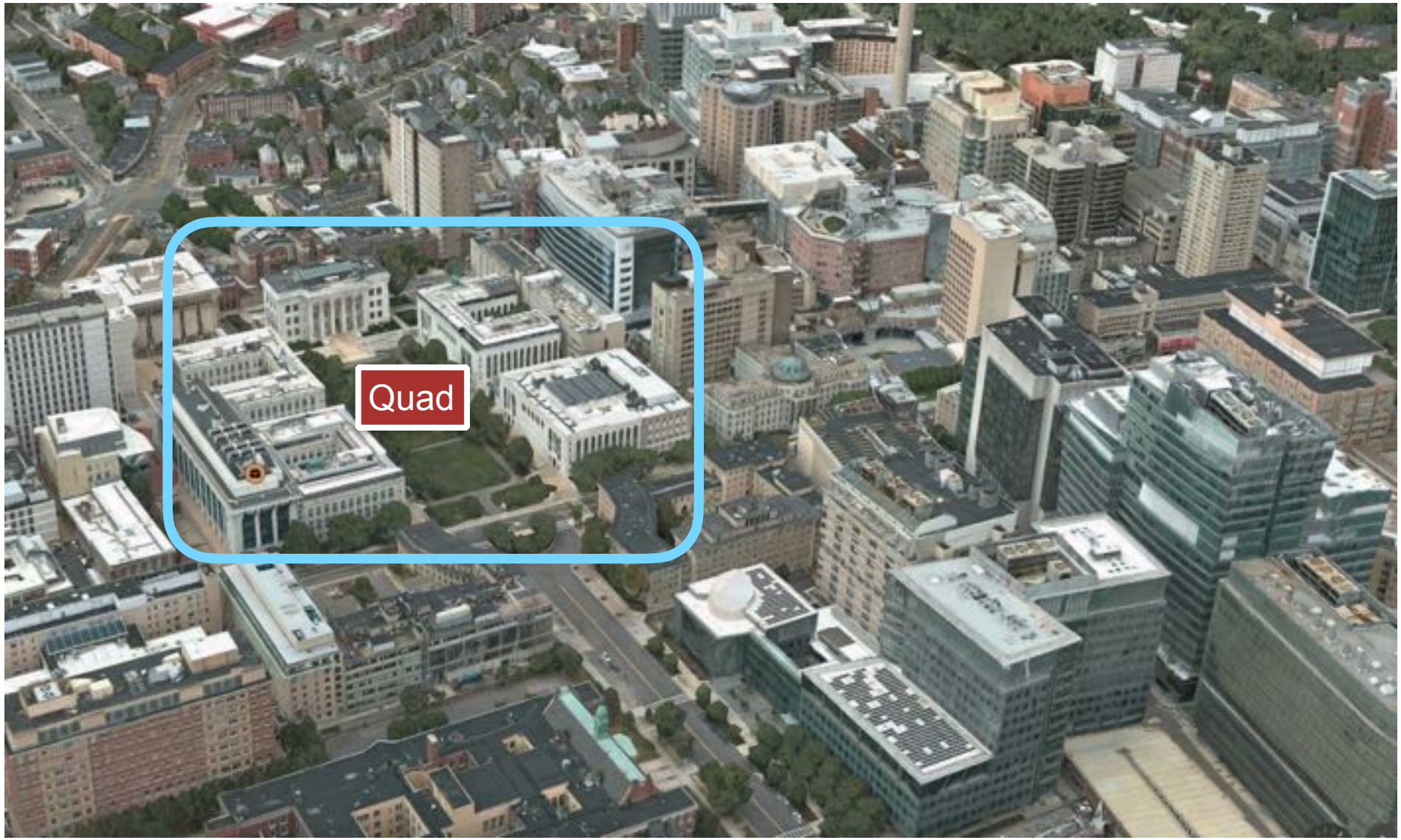
grants, departmental buy-ins, …

ARPA-H (small slice of $104 million awarded to DARTS)

Core facility or **research group**?

Neither and both = *Collaboratory (we don't currently charge for work)*

Harvard Medical School & Friends

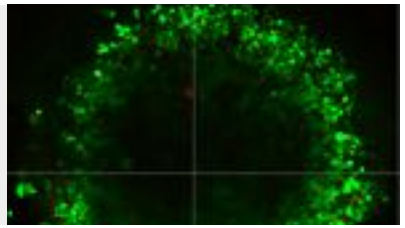


Project Overview (sample)

Segmentation and Quantification of Cells and Patterns in a Sorting Assay

Sean McGeary, PhD
PI: Allon Klein, PhD

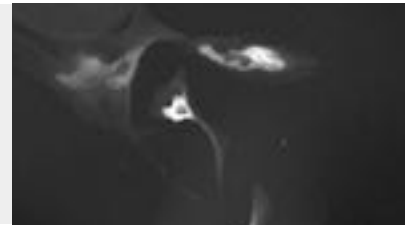
Sys Bio



Detection and Classification of Cell Aging in Chemically Induced Cells

Thomas Dixon-McDougall, PhD
PI: David Sinclair, AO, PhD

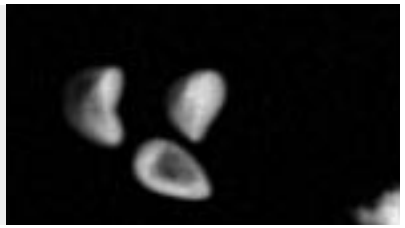
Genetics



Measuring the Polymerized Mass and Classifying Cell Type

Daniel De Souza, PhD
PI: John Higgins, PhD

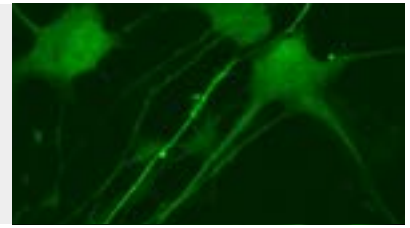
Sys Bio



Measuring the Level of ER-Mito Stabilizers in Cell Body/Soma

PI: Raja Bhattacharyya, PhD

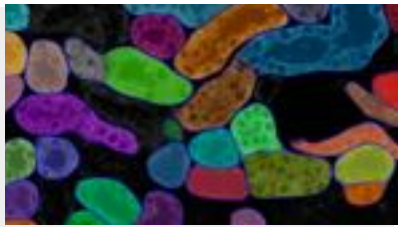
Non-Quad



Determining Protein and Lipid Contents in Raman Imaged Organs

Will Trim, PhD
PI: Marc Kirschner, PhD

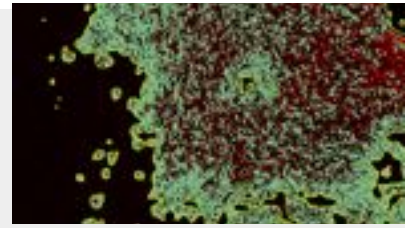
Sys Bio



Spatial Analysis of Cancer Cell Distributions in Stromae

Nina Kozlova, PhD
PI: Taru Muranen, PhD

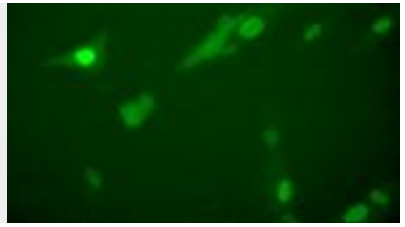
Non-Quad



Tracking and Identification of Cell State

Noelle Ozimek
PI: Randy King, MD, PhD

Cell Bio



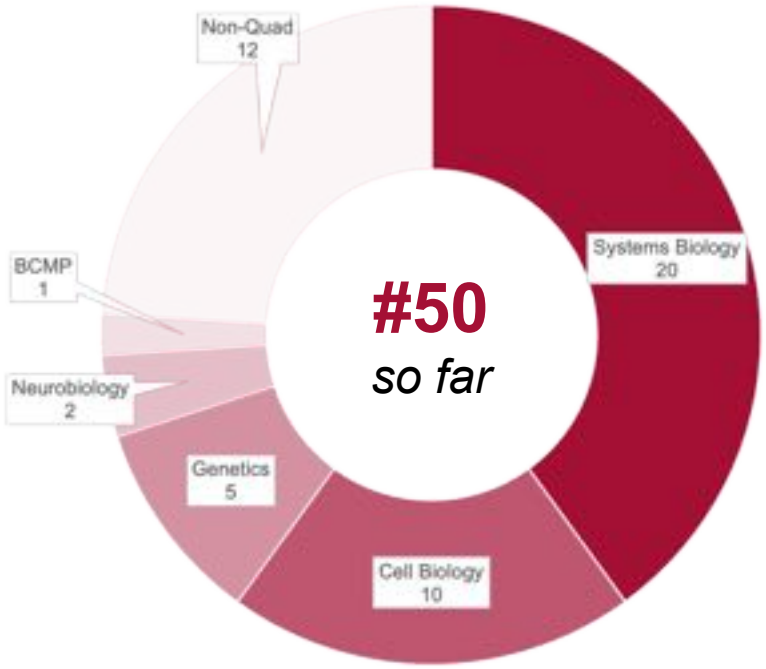
Project DIOS

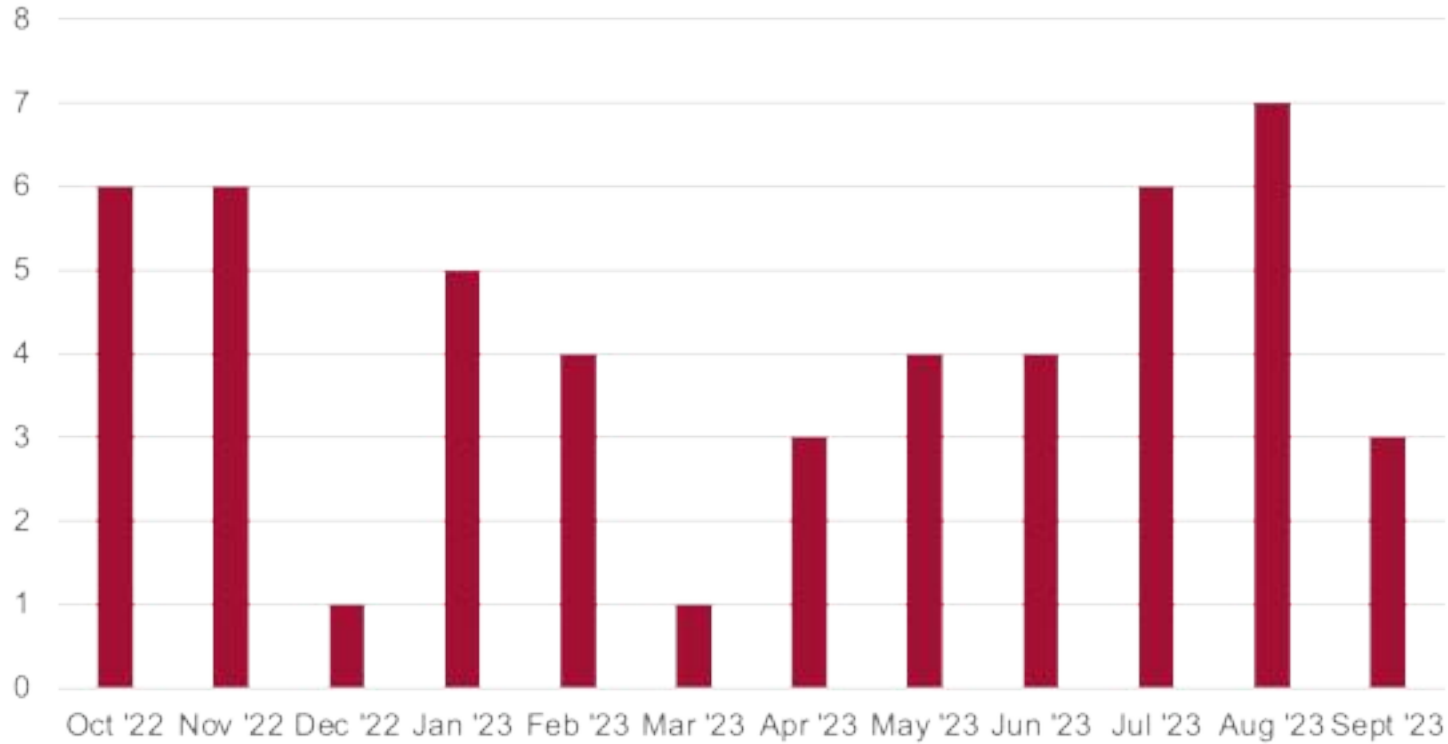
Ranit Karmakar, PhD
PI: Simon Nørrelykke, PhD

All-Quad

Image Clinics / Consultations

"Image Clinics are consultations where we look at your data and discuss solutions to your image analysis needs"





HARVARD MEDICAL SCHOOL

Image Analysis Collaboratory

Teaching at/from HMS

With NIC
Federico Gasparoli
Anna Jost

Jennifer's course

THE UNIVERSITY OF CHICAGO MARINE BIOLOGICAL LABORATORY

Analytical and Quantitative Light Microscopy

A comprehensive and intensive course in light microscopy for researchers in biology, medicine, and material sciences.

Apr, '23

CSH Cold Spring Harbor Laboratory

Quantitative Imaging: From Acquisition to Analysis

CSHL Courses are intensive, running all day and often including evenings and weekends; students are expected to attend all sessions and reside on campus for the duration of the course.

Apr, '23

Introduction to Image Analysis using ImageJ/Fiji

Two-day intro to quantitative bioimage analysis for life-scientists. No preparation, no homework.

With the Nikon Imaging Center.

Nov, '23

Dec, '22 Nov, '22 Mar, '23

HARVARD MEDICAL SCHOOL

https://iac.hms.harvard.edu/teaching/

Image Analysis Collaboratory

Teaching at/from ETH: 50+ lectures, courses, and schools

EMBO Practical Course: **Advanced Methods** in BioImage Analysis (2021)

Deep Learning for Image Analysis [EMBL Course] (2020–2022)

Zurich/Switzerland's Image and Data Analysis School, ETH/EPFL (2017–2022)

Introduction to Image Analysis using Fiji/ImageJ, ETH (2013–2022)

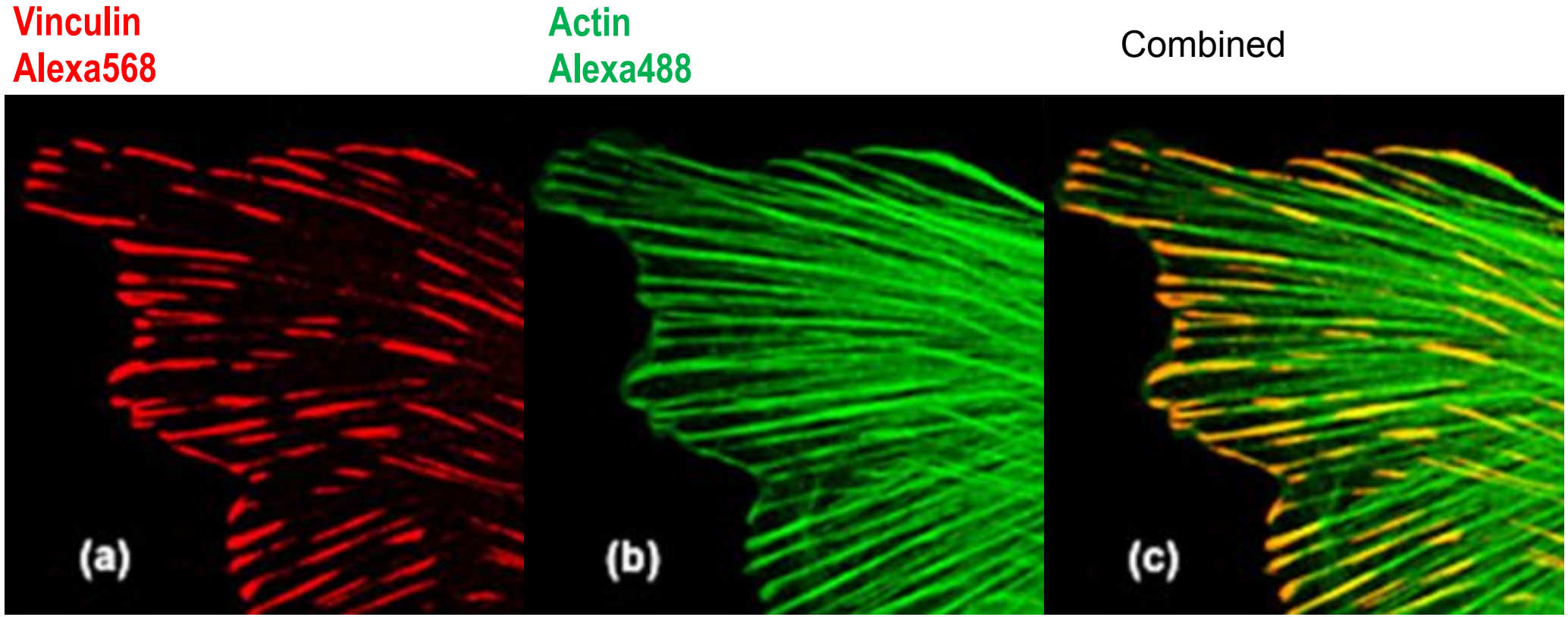
HARVARD MEDICAL SCHOOL

https://www.let-your-data-speak.com/#teaching

Image Analysis Collaboratory

What is colocalization?

Protein Colocalization



http://www.olympusconfocal.com/applications/colocalization.html

Colocalization: The presence of two or more fluorophores on the same physical structure (in a cell).

What Problems are we trying to Solve?

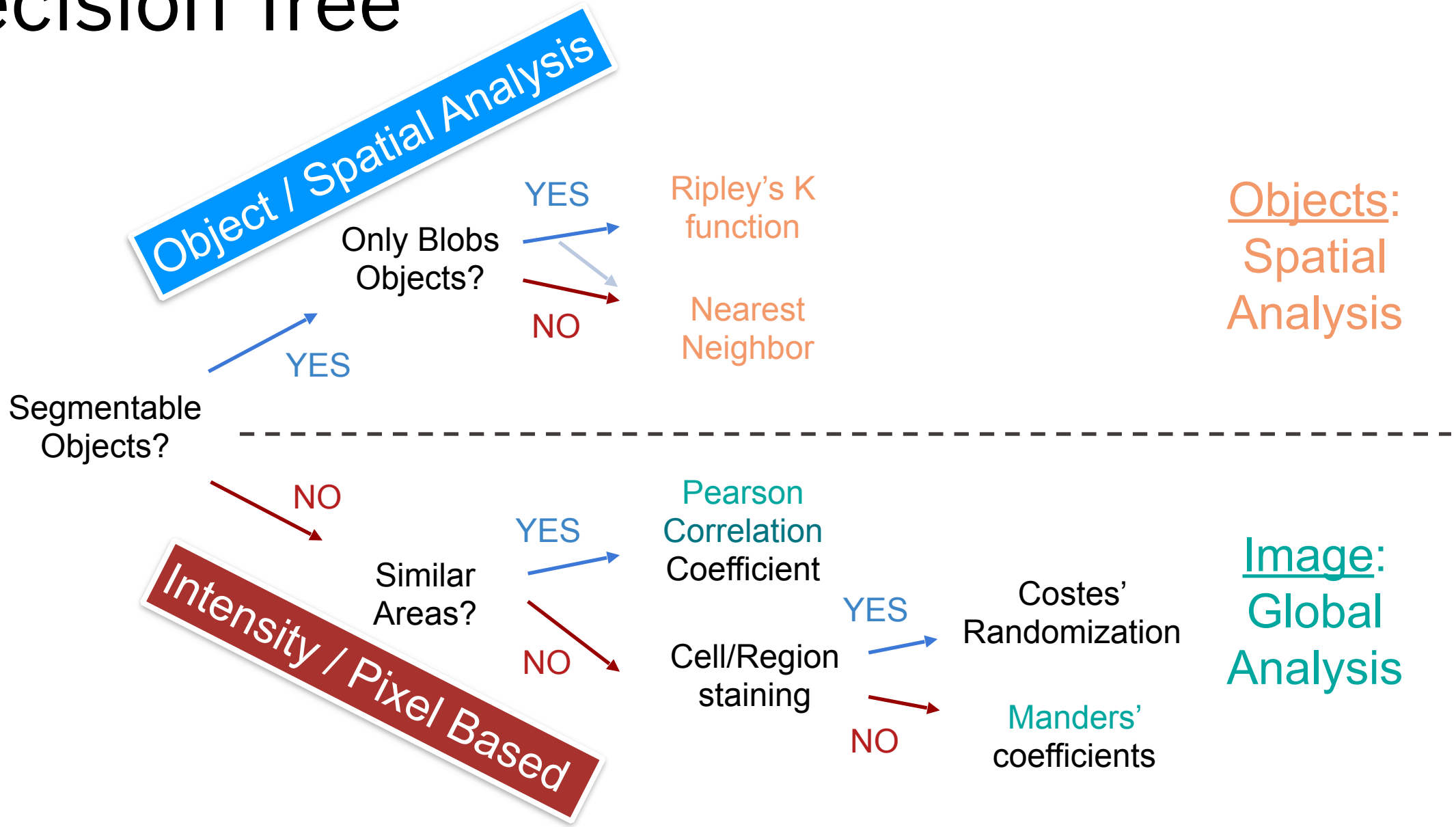
Want: Show that one protein *cause* the presence of another

Have: Images from various experimental conditions

Do: *Quantify* the degree to which information about one image allows us to make predictions about another image (mutual information, very loosely interpreted)

Limits: Typically cannot answer *causal* questions, only *correlative* ones

Decision Tree



From slide by Romain Guit, BIOP, EPFL

Dear Child has many Names

Co-localization

Co-expression

Co-variation

Co-distribution

Co-occurence

Concomitance

Coincidence Analysis

Overlap Analysis

Spatial Correlation

Proximity Analysis

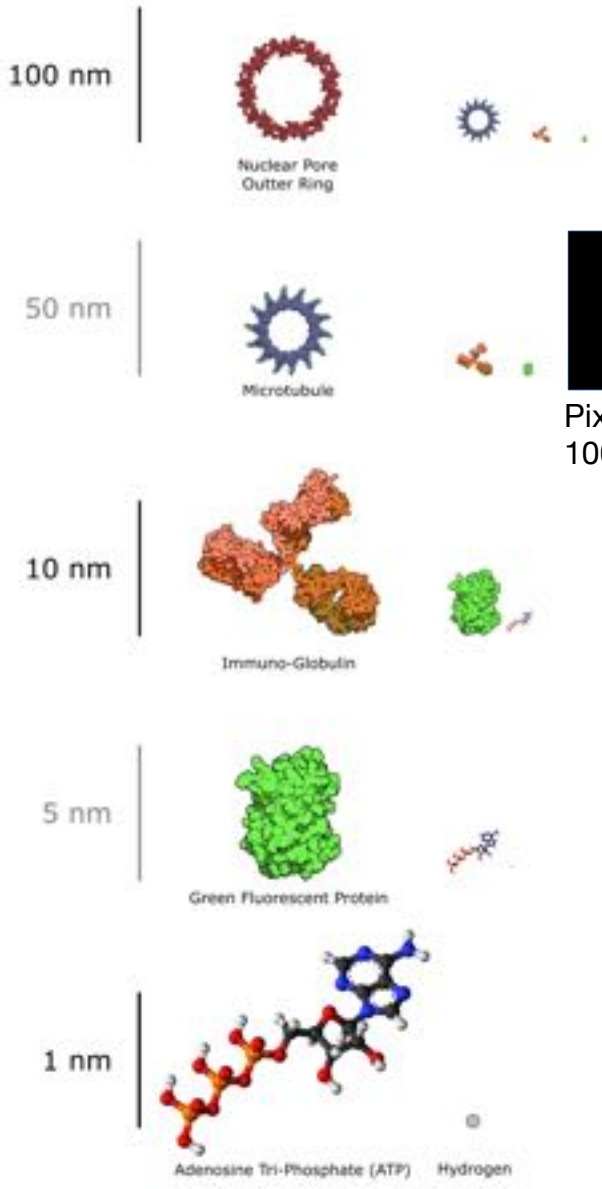
Simultaneous Localization

Intersection Analysis

What's in a name? That which we call a rose
By any other name would smell as sweet;

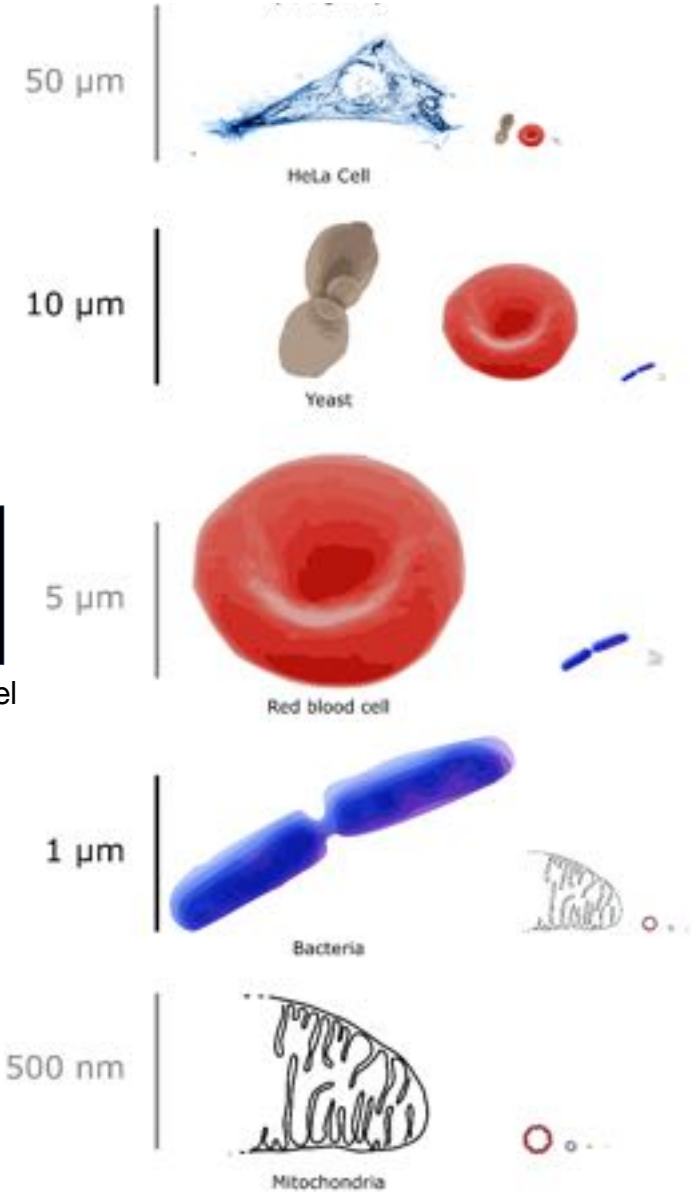
Imaging & Scales

Length Scales



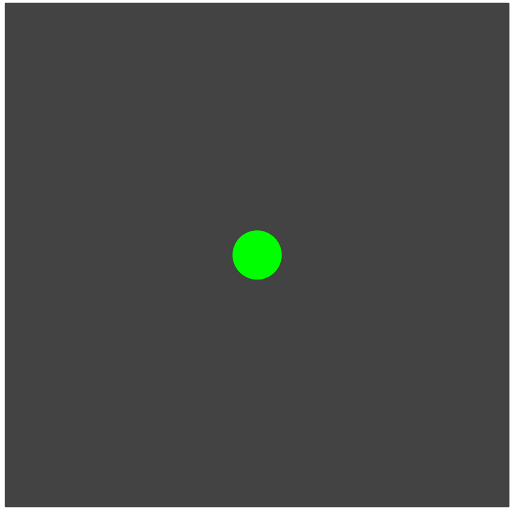
Pixel using a
100x objective

Camera pixel



https://c4science.ch/w/bioimaging_and_optics_platform_biop/teaching/probes/

GFP



Pixel of a camera
at 100X

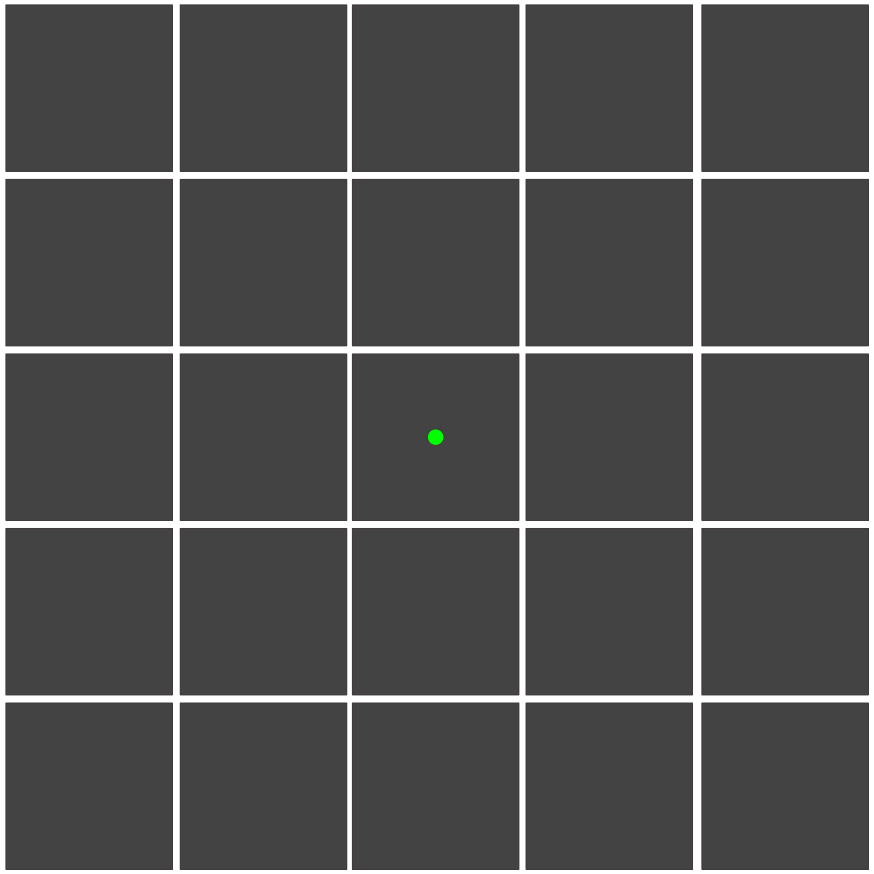
Biology scales

VS

Observation scales

Romain Guiet, BIOP, EPFL

GFP



Pixels grid of a camera at 100X

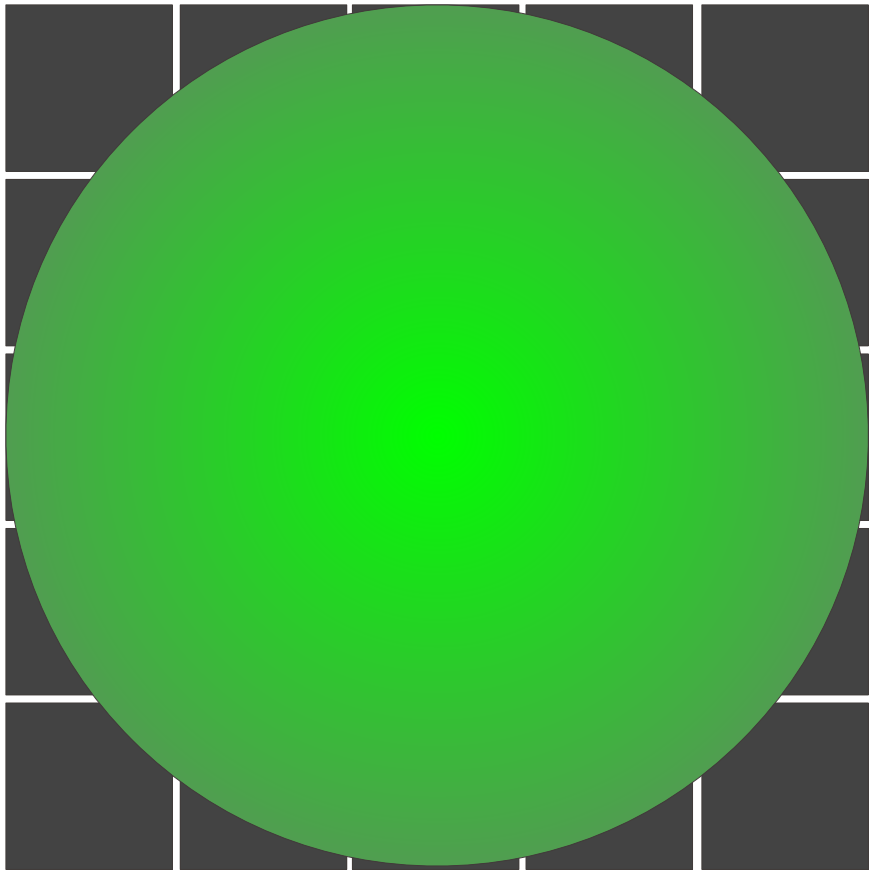
Biology scales

VS

Observation scales

Romain Guiet, BIOP, EPFL

GFP-diffraction limited signal



Pixels grid of a camera at 100X

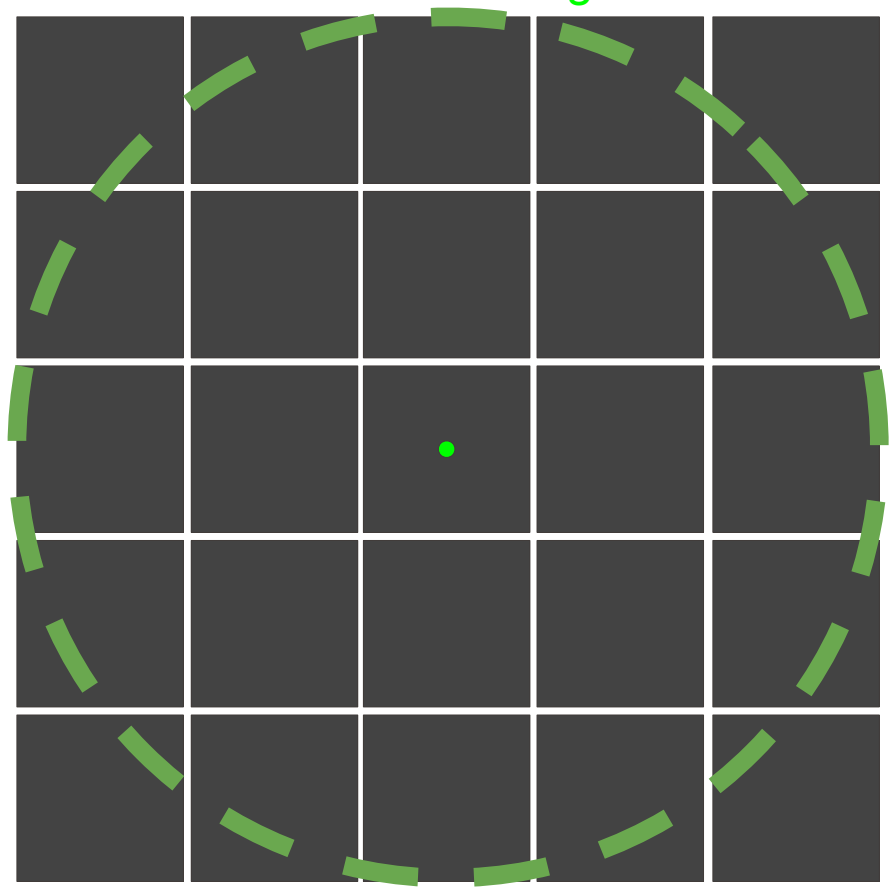
Biology scales

VS

Observation scales

Romain Guiet, BIOP, EPFL

GFP-diffraction limited signal



Pixels grid of a camera at 100X

Biology scales

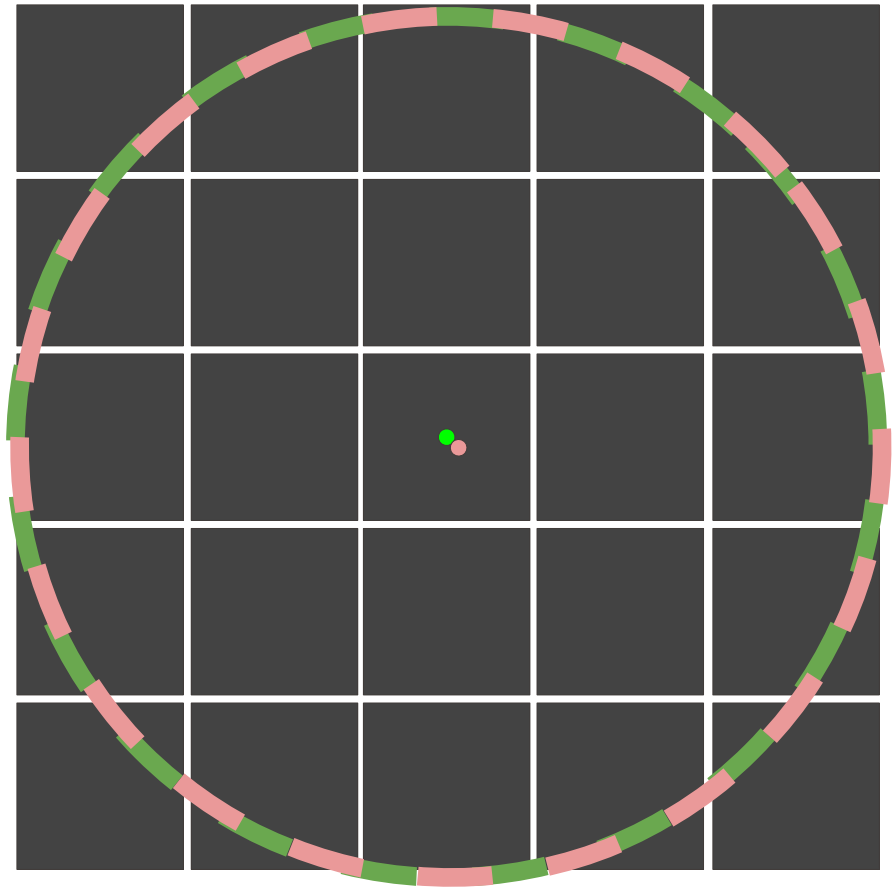
VS

Observation scales

Romain Guiet, BIOP, EPFL

XFP-diffraction limited signal

GFP-diffraction limited signal



Pixels grid of a camera at 100X

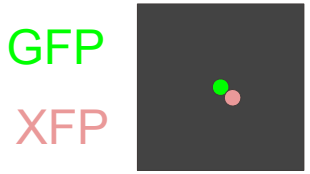
Biology scales

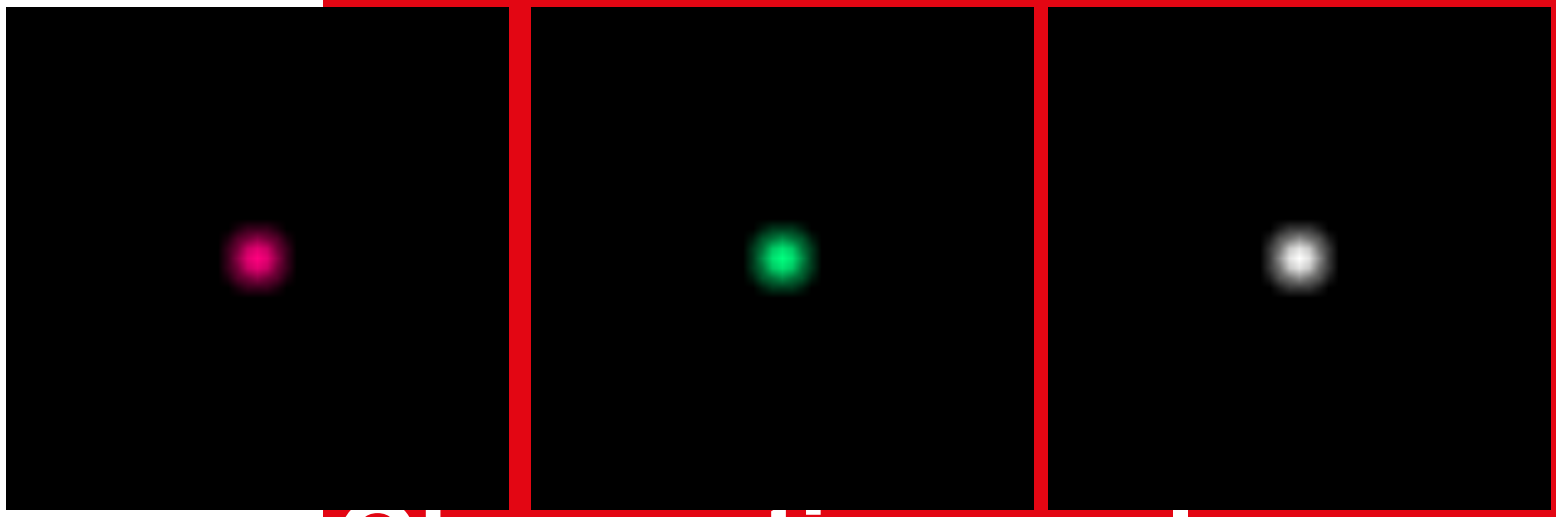
VS

Observation scales

Romain Guiet, BIOP, EPFL

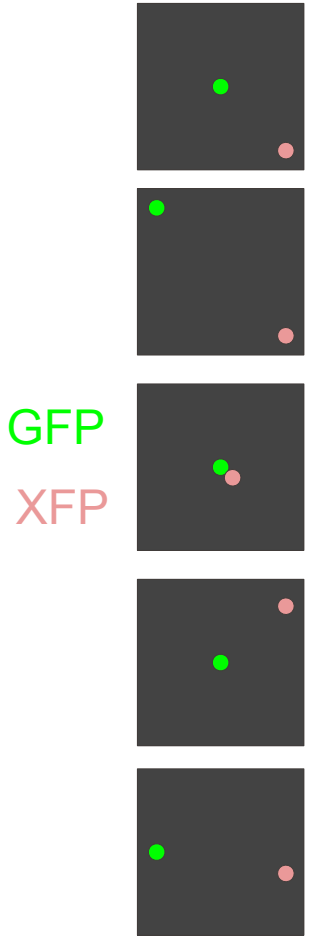
Biology scales



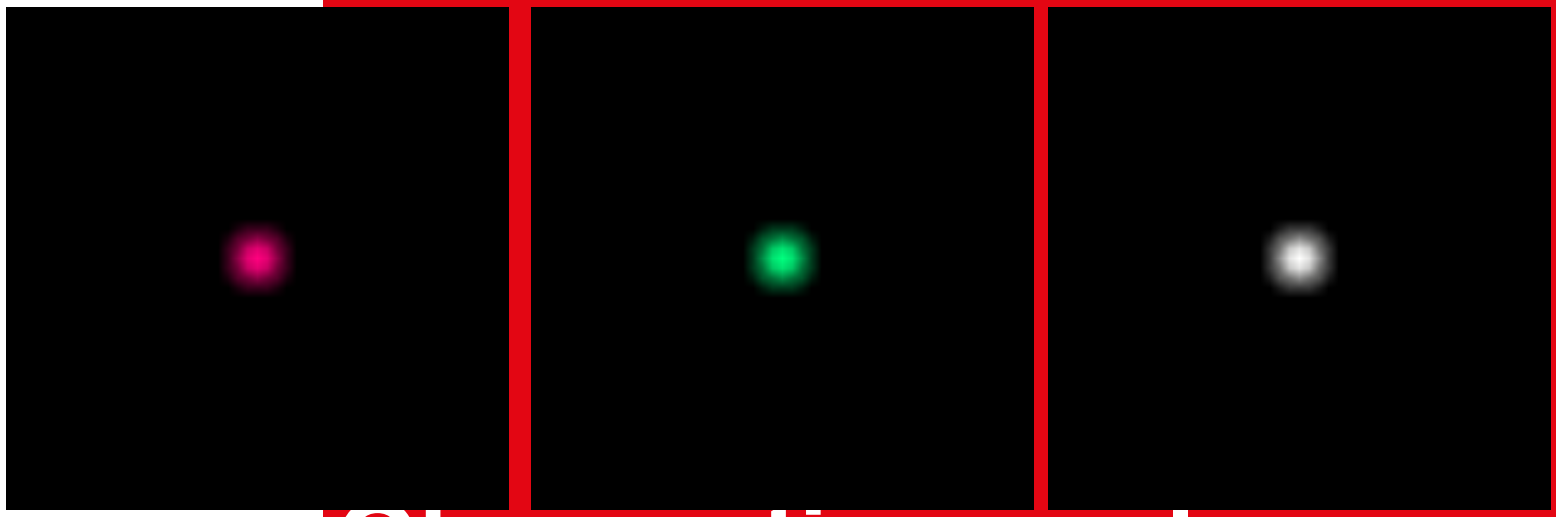


Observation scales

Romain Guiet, BIOP, EPFL



Biology scales

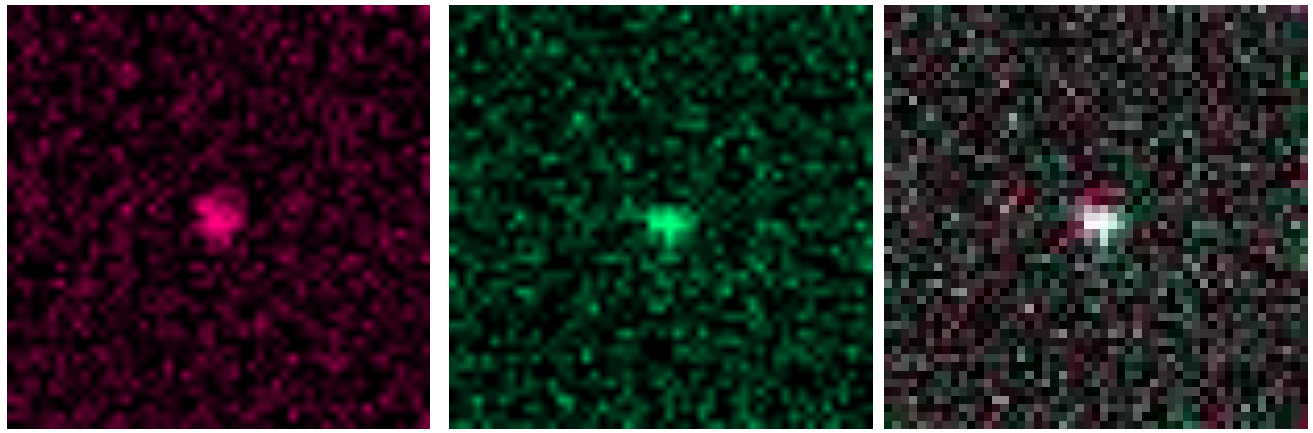


Observation scales

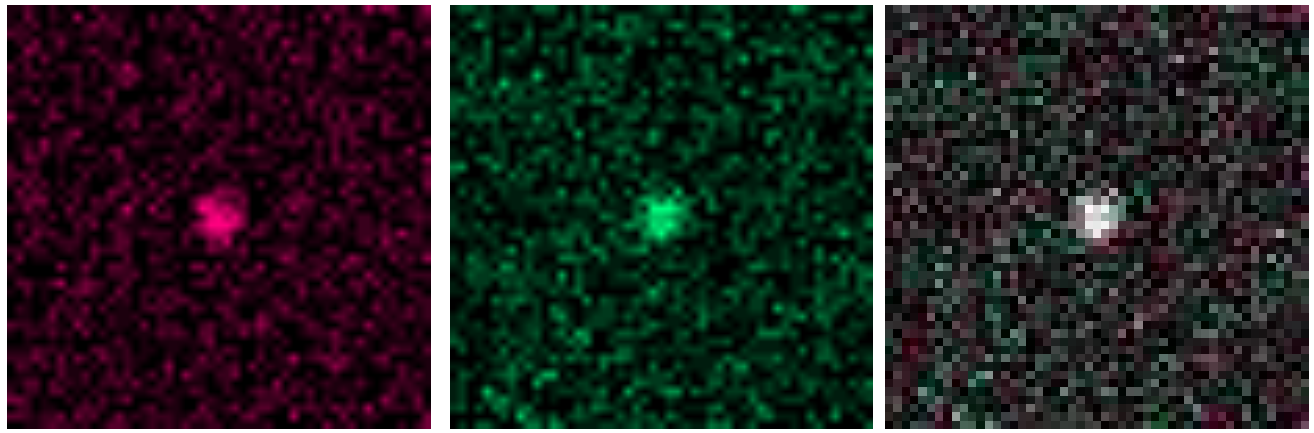
Romain Guiet, BIOP, EPFL

Noise Influence

A

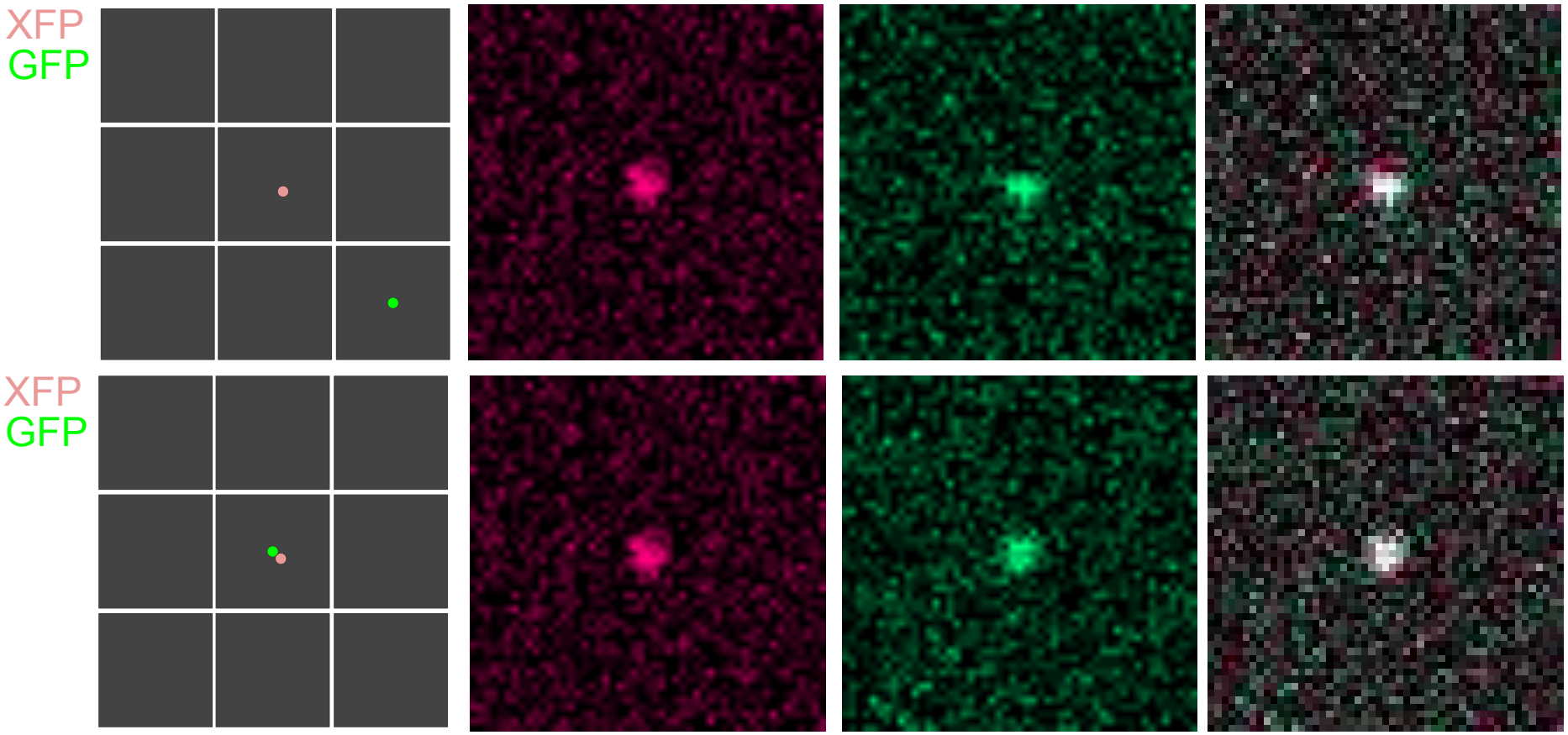


B



Romain Guiet, BIOP, EPFL

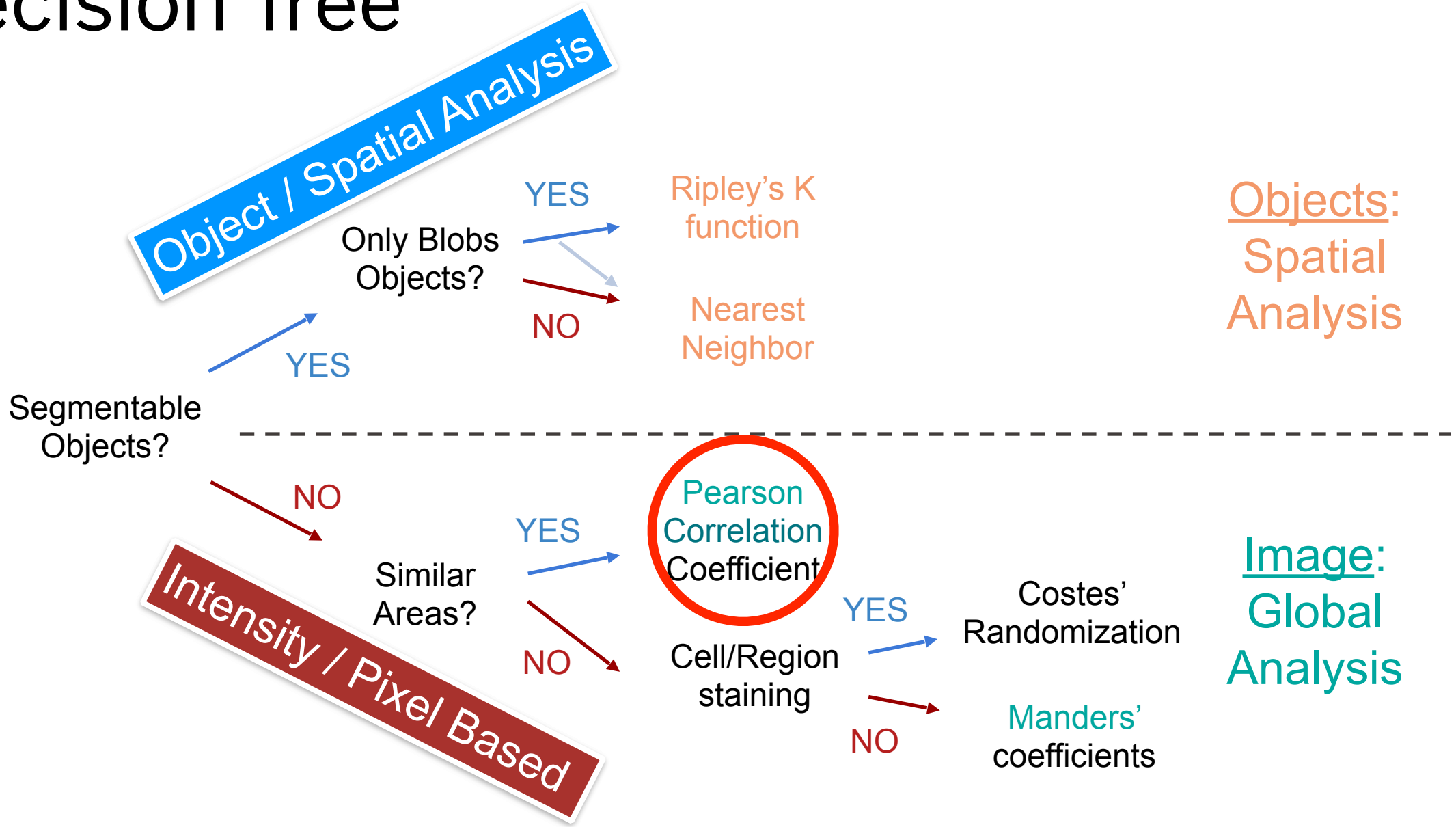
Noise Influence



Romain Guiet, BIOP, EPFL

Pearson’s

Decision Tree



From slide by Romain Guiet, BIOP, EPFL

Pearson's Correlation Coefficient

$$r_P = \frac{\mathrm{cov}(R,G)}{\sigma(R)\sigma(G)} = \frac{\sum_i (R_i - R_{avg})(G_i - G_{avg})}{\sqrt{\sum_i (R_i - R_{avg})^2 \sum_i (G_i - G_{avg})^2}}$$

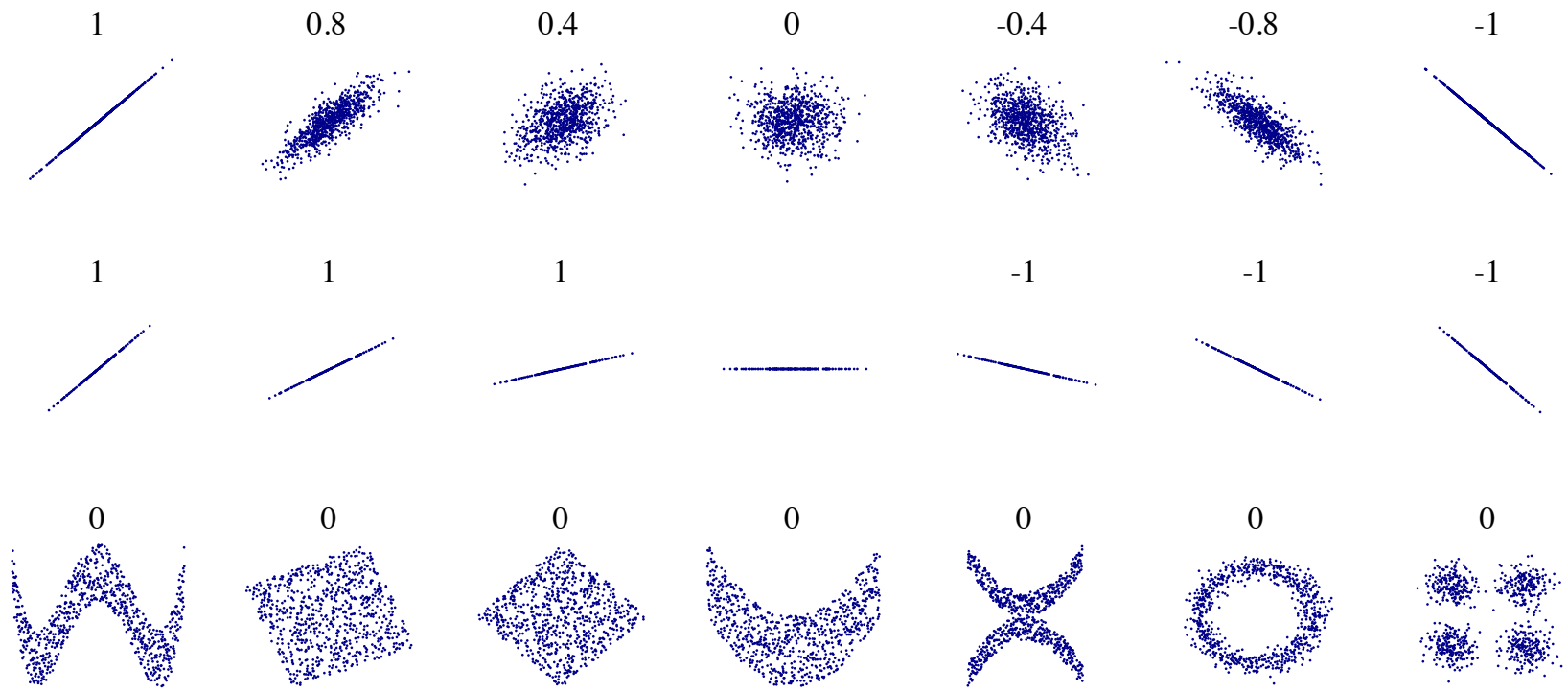
Linear correlation coefficient, unlike Spearman's rank etc

Invariant to *affine* intensity transformations

$$R_i^{new} = aR_i + b$$

Gain and offset, not quite *exposure time* and *background*

Pearson's Correlation Coefficient



Not sensitive to patterns (*non-linear* relations)

https://en.wikipedia.org/wiki/Correlation

Pearson's Correlation Coefficient

Channel 1 | Channel 2 | Combined

$$r_P = \frac{\sum_i (R_i - R_{avg})(G_i - G_{avg})}{\sqrt{\sum_i (R_i - R_{avg})^2 \sum_i (G_i - G_{avg})^2}} = 0.94$$

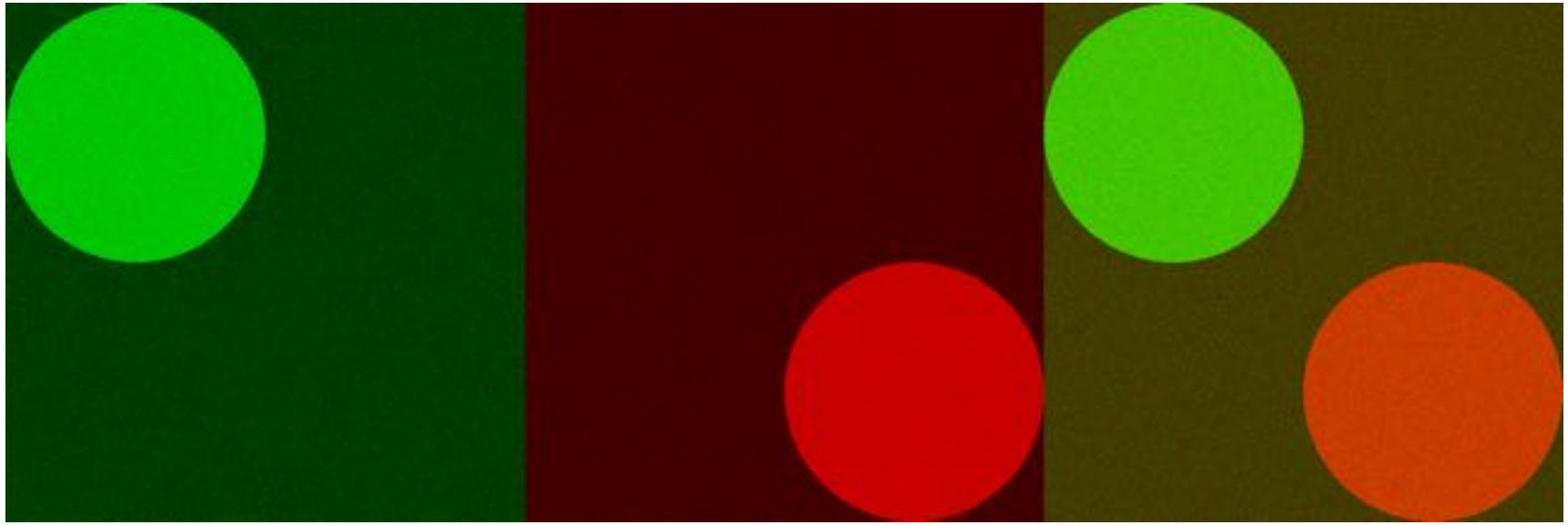
Anti-Correlation

Channel 1 Channel 2 Combined

$$r_P = \frac{\sum_i (R_i - R_{avg})(G_i - G_{avg})}{\sqrt{\sum_i (R_i - R_{avg})^2 \sum_i (G_i - G_{avg})^2}} = -0.94$$

Exclusion

Channel 1 Channel 2 Combined



$$r_P = \frac{\sum_i (R_i - R_{avg})(G_i - G_{avg})}{\sqrt{\sum_i (R_i - R_{avg})^2 \sum_i (G_i - G_{avg})^2}} = -0.29$$

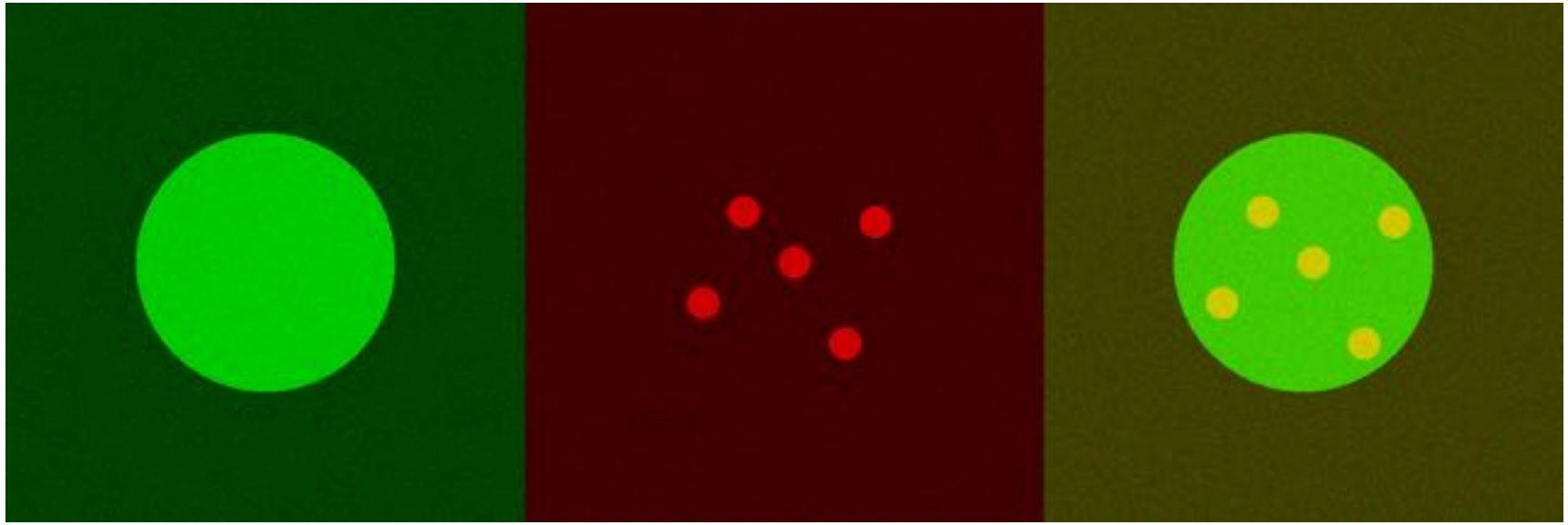
Partial Overlap

Channel 1 Channel 2 Combined

$$r_P = \frac{\sum_i (R_i - R_{avg})(G_i - G_{avg})}{\sqrt{\sum_i (R_i - R_{avg})^2 \sum_i (G_i - G_{avg})^2}} = -0.016$$

Inclusion of small Objects

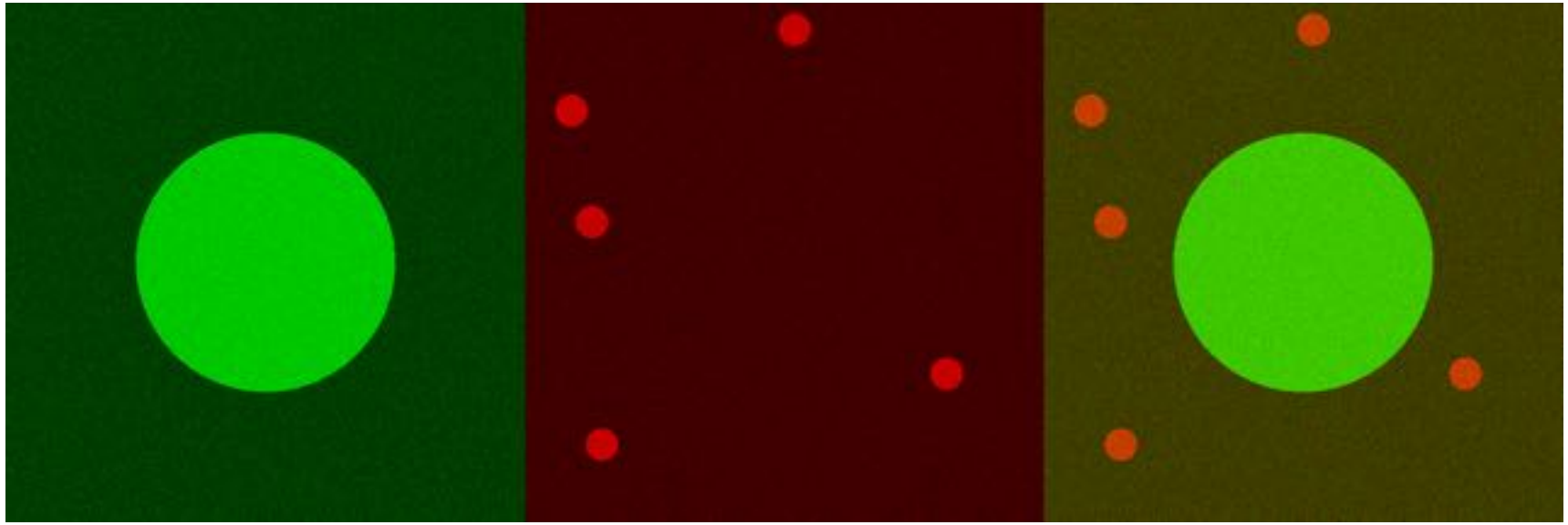
Channel 1 Channel 2 Combined



$$r_P = \frac{\sum_i (R_i - R_{avg})(G_i - G_{avg})}{\sqrt{\sum_i (R_i - R_{avg})^2 \sum_i (G_i - G_{avg})^2}} = 0.19$$

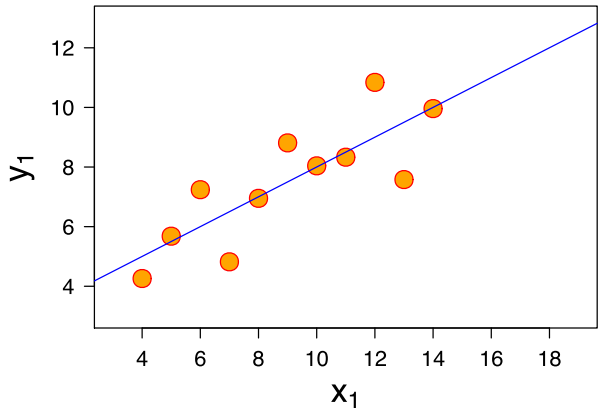
Exclusion of small Objects

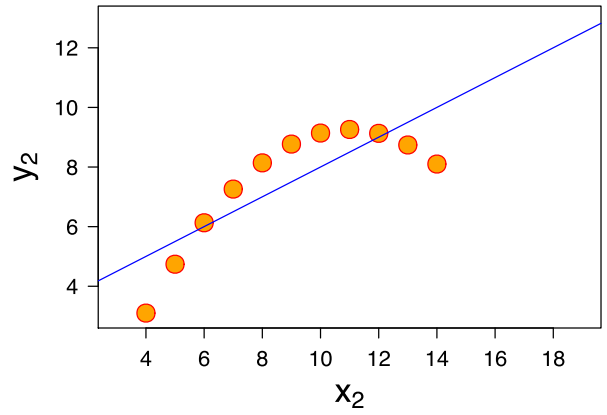
Channel 1 Channel 2 Combined

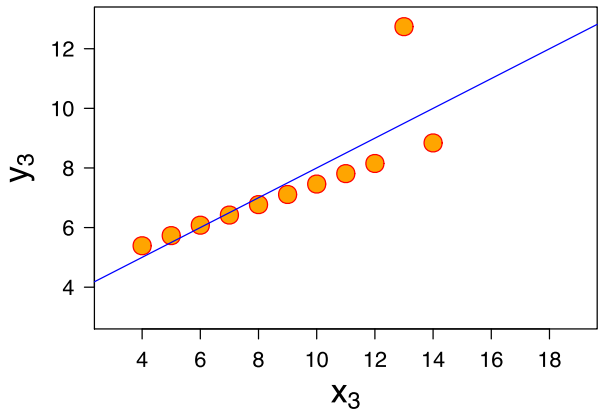


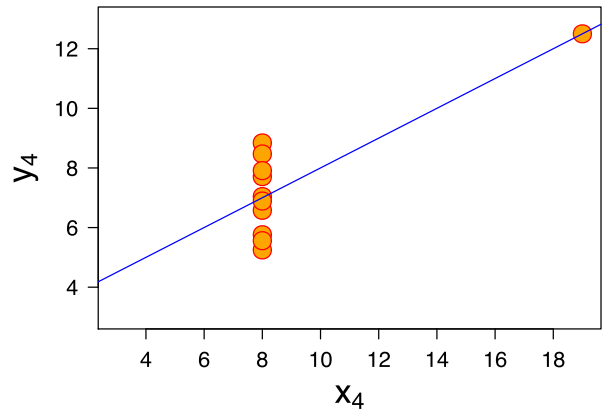
$$r_P = \frac{\sum_i (R_i - R_{avg})(G_i - G_{avg})}{\sqrt{\sum_i (R_i - R_{avg})^2 \sum_i (G_i - G_{avg})^2}} = -0.047$$

Ascombes Quartet









Property	Value
Mean of x	9
Sample variance of x: s_x^2	11
Mean of y	7.50
Sample variance of y: s_y^2	4.125
Correlation between x and y	0.816
Linear regression line	$y = 3.00 + 0.500x$
Coefficient of determination of the linear regression: R^2	0.67

https://en.wikipedia.org/wiki/Anscombe's_quartet

Pearson's Correlation Coefficient

Clear interpretation (only one number), somewhat robust

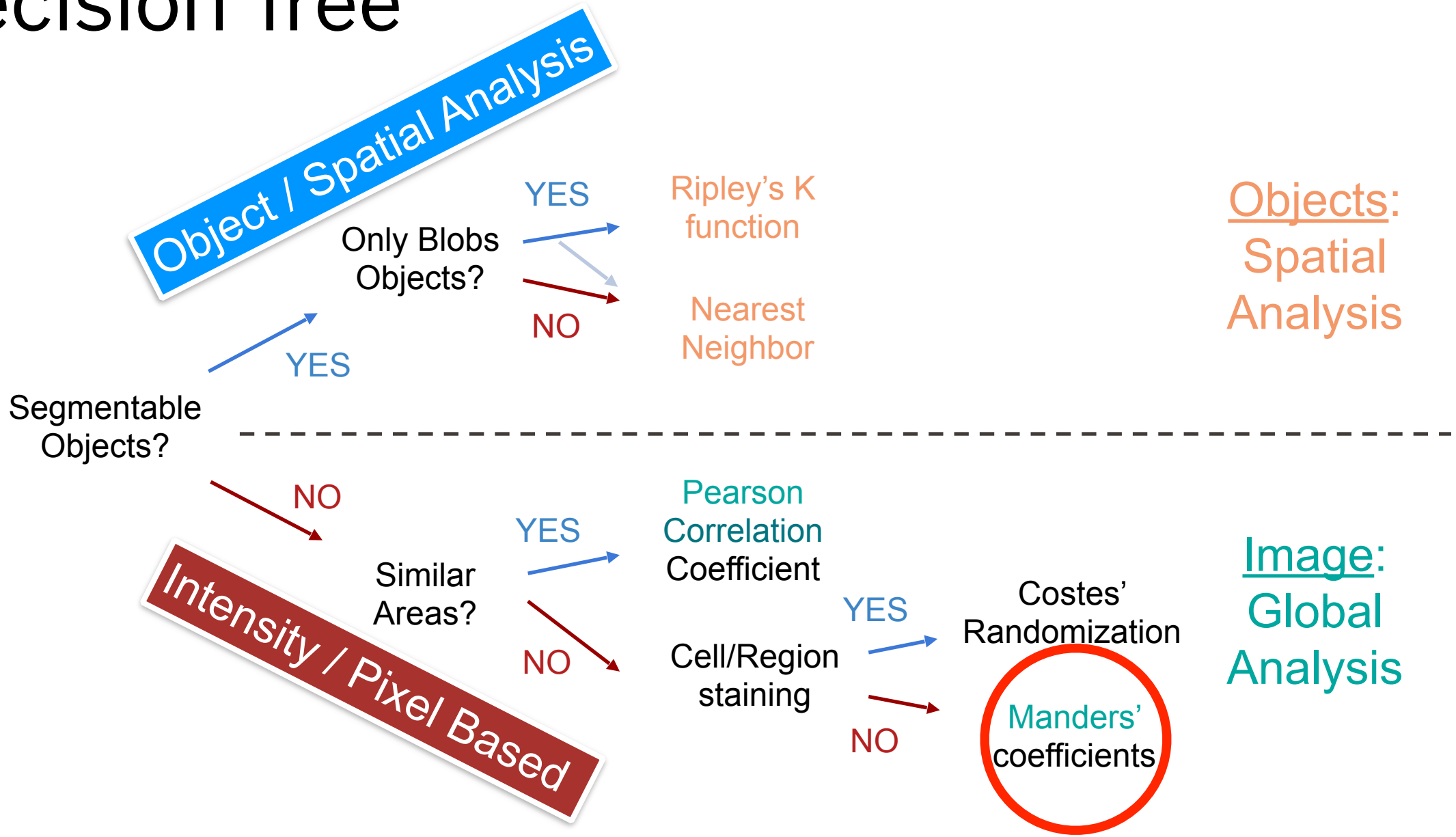
Doesn't return statistical significance

Fails: unequal number of objects in images compared

Solution: Manders' coefficients (or object based coloc)

Manders

Decision Tree



From slide by Romain Guiet, BIOP, EPFL

Manders' coefficients

Eric Manders (then at University of Amsterdam) introduced the use, in confocal analysis, of Pearson's coefficient in 1992

Then came up with his own coefficients in 1993

Implemented in Imaris (commercial software)

Manders, E. M., Stap, J., Brakenhoff, G. J., Driel, R. van & Aten, J. A. Dynamics of three-dimensional replication patterns during the S-phase, analysed by double labelling of DNA and confocal microscopy. *J. cell Sci.* **103 (Pt 3)**, 857–62 (1992).

MANDERS, E. M. M., VERBEEK, F. J. & ATEN, J. A. Measurement of co-localization of objects in dual-colour confocal images. *J. Microsc.* **169**, 375–382 (1993).

Mander's Overlap Coefficient (MOC), r_M

Introduced to avoid negative values; very similar to Pearson's

$$r_M = \frac{\sum_i R_i G_i}{\sqrt{\sum_i R_i^2 \sum_i G_i^2}}; \quad r_P = \frac{\sum_i (R_i - R_{avg})(G_i - G_{avg})}{\sqrt{\sum_i (R_i - R_{avg})^2 \sum_i (G_i - G_{avg})^2}}$$

Invariant to *linear* intensity transformations, e.g. $R_i^{new} = aR_i$

Ambiguous results when number of objects in R and G differs

Mander's Split Overlap Coefficients

Addresses ambiguity in the Overlap Coefficient, r_M

$$k_1 = \frac{\sum_i R_i G_i}{\sum_i R_i^2} \; ; \; k_2 = \frac{\sum_i R_i G_i}{\sum_i G_i^2} \text{; so that } r_M^2 = k_1 k_2$$

Each depends *linearly* on the intensity of the other channel

Mander's Colocalization Coefficients

Addresses linear dependence in the Split Overlap Coefficients

$$M_1 = \frac{\sum_i R_i^{coloc}}{\sum_i R_i} \text{ and } M_2 = \frac{\sum_i G_i^{coloc}}{\sum_i G_i}$$

$$\text{where } R_i^{coloc} = \begin{cases} 0, & G_i = 0 \\ R_i, & G_i > 0 \end{cases} \text{ and } G_i^{coloc} = \begin{cases} 0, & R_i = 0 \\ G_i, & R_i > 0 \end{cases}$$

Each is now "*independent*" of intensity in the other channel

Or rather, now the dependence is *non-linear*

Mander's Colocalization Coefficients

$$M_1 = \frac{\sum_i R_i \mathbf{1}_{G_i>0}}{\sum_i R_i} \text{ and } M_2 = \frac{\sum_i G_i \mathbf{1}_{R_i>0}}{\sum_i G_i}$$

where $\mathbf{1}_{X>0} = \begin{cases} 0, & X = 0 \\ 1, & X > 0 \end{cases}$ is the indicator function

Note the *non-linearity* and *mixed dependence:*

M_1 depends on *red intensity* in the *area in green* with positive intensity

$\sum_i \mathbf{1}_{G_i>0}$ = area in green channel used to "mask" red channel

Some Observations

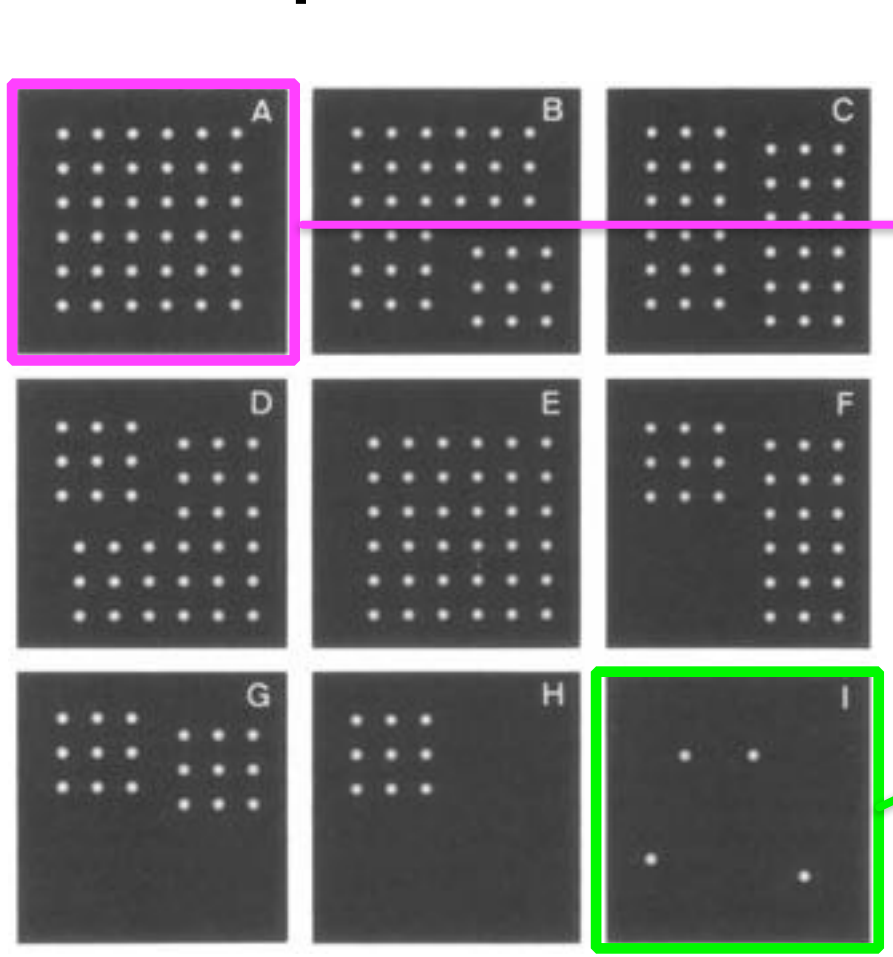
$M_1, M_2 \in [0,1]$, but tends to have values close to 1:

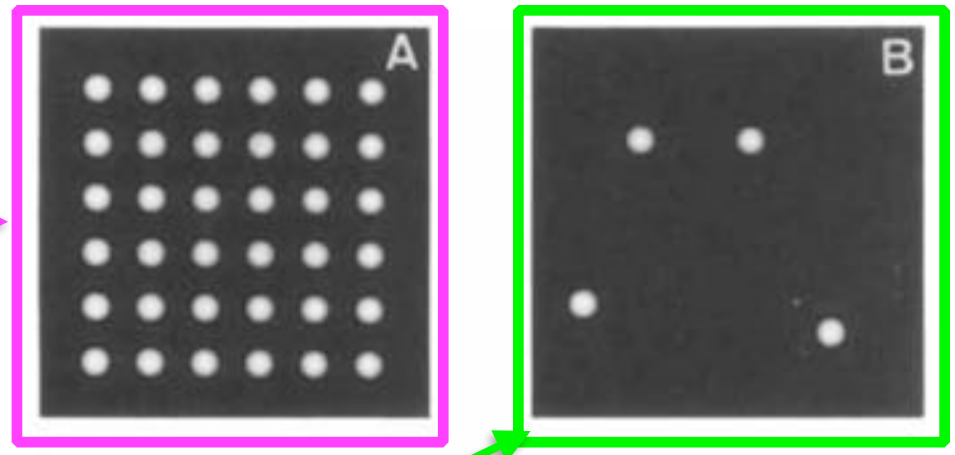
If there are no black pixels in the green channel $M_1 = 1$, and vice versa

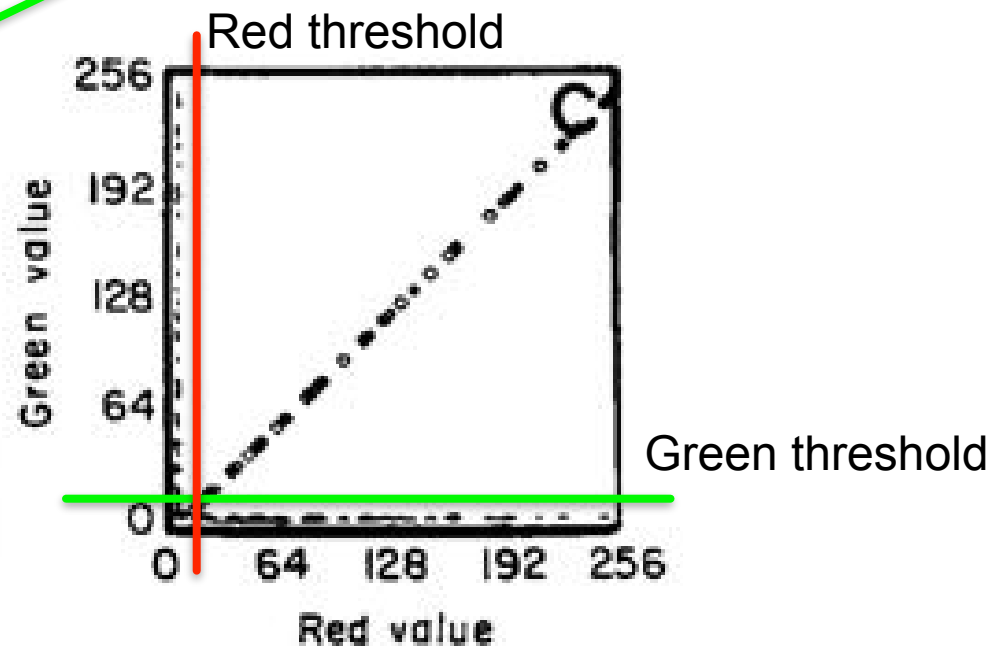
M_1 depends on *red intensity* and *area in green* with positive intensity; and vice versa for M_2

M_1 and M_2 "*are proportional to the amount of fluorescence of the co-localizing objects in each component [channel] of the image, relative to the total fluorescence in that component [channel]*"

Example





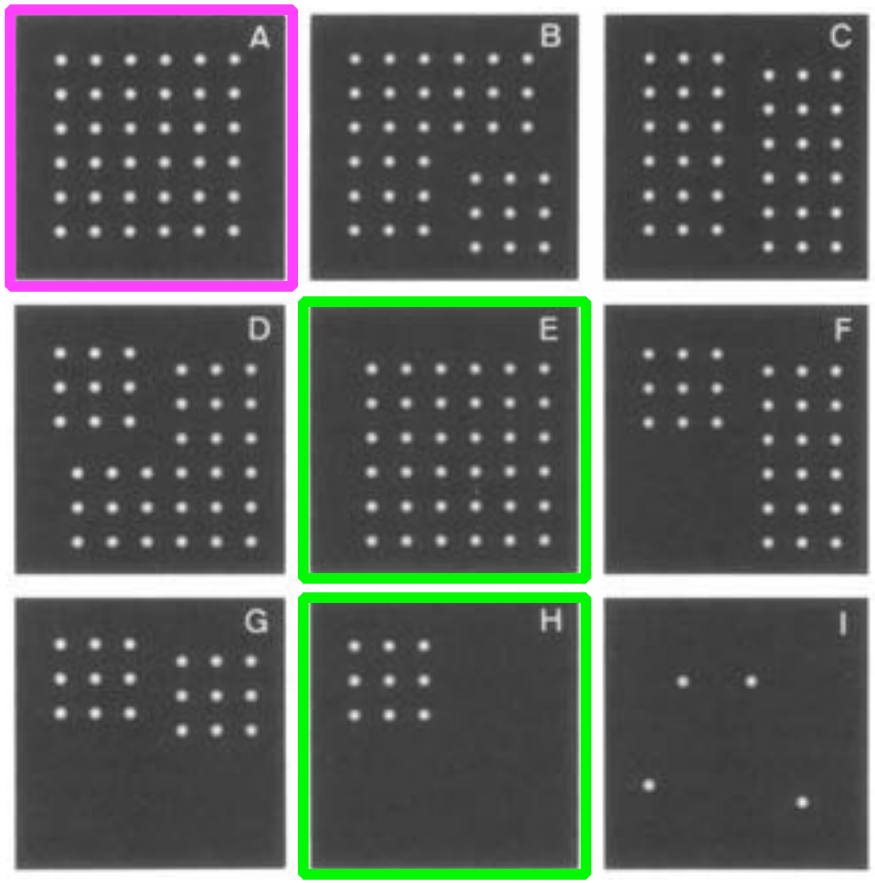


MANDERS, E. M. M., VERBEEK, F. J. & ATEN, J. A. Measurement of co-localization of objects in dual-colour confocal images. *J. Microsc.* **169**, 375–382 (1993).

Workflow

1. Preprocess images (noise reduction, illumination correction)
2. *Manually* set thresholds so “background” is black
3. Calculate r_M, M_1, and M_2 (for all pixels above thresholds)

Example



MANDERS, E. M. M., VERBEEK, F. J. & ATEN, J. A. Measurement of co-localization of objects in dual-colour confocal images. *J. Microsc.* **169**, 375–382 (1993).

Figures	Number of objects: Red	Green	Co-localization	r_P	r	M_1	M_2
A A	36	36	36	1·00	1·00	1·00	1·00
A B	36	36	27	0·72	0·75	0·75	0·75
A C	36	36	18	0·44	0·50	0·50	0·50
A D	36	36	9	0·16	0·25	0·25	0·25
A E	36	36	0	−0·12	0·00	0·00	0·00
A F	36	27	9	0·22	0·29	0·25	0·33
A G	36	18	9	0·30	0·35	0·25	0·50
A H	36	9	9	0·48	0·50	0·25	1·00
A I	36	4	3	0·23	0·25	0·08	0·75

Number of objects

Figures	*Red*	*Green*	*Colocalizing*	r_P	r_M	M_1	M_2
AA	*36*	*36*	*36*	*1.00*	*1.00*	*1.00*	*1.00*
AE	*36*	*36*	*0*	*-0.12*	*0.00*	*0.00*	*0.00*
AH	*36*	*9*	*9*	*0.48*	*0.50*	*0.25*	*1.00*
AI	*36*	*4*	*3*	*0.23*	*0.25*	*0.08*	*0.75*

Here: $M_1 \approx \frac{\text{Colocalizing}}{\text{Red}}$ and $M_2 \approx \frac{\text{Colocalizing}}{\text{Green}}$

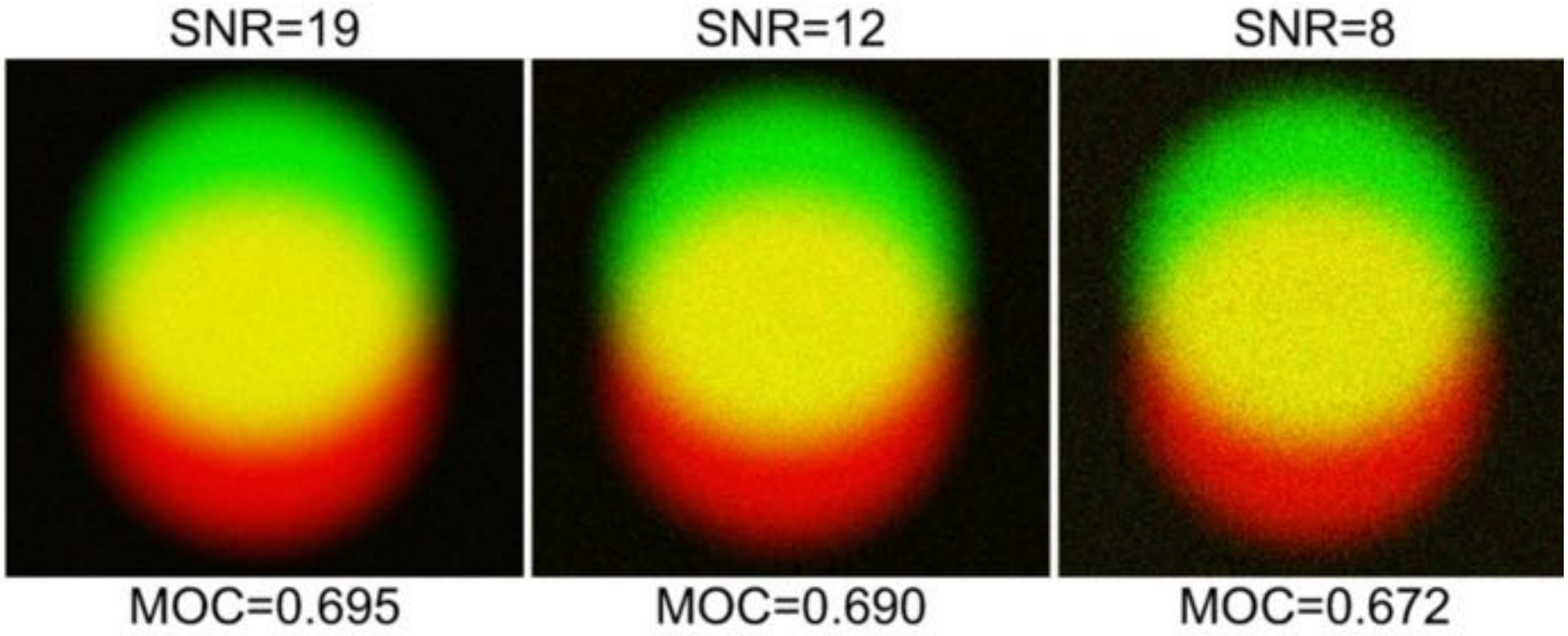
Comments on Example

Images need to be *processed* before analysis, to turn the background into black pixels (denoise, bgr subtract, threshold)

Here, simply counting the number of spots gives M_1 and M_2, but this is *misleading*—we measure *area* overlap, not *object* matching

A single large blob could completely outweigh the many small ones

Manders is insensitive to Signal to Noise Ratios

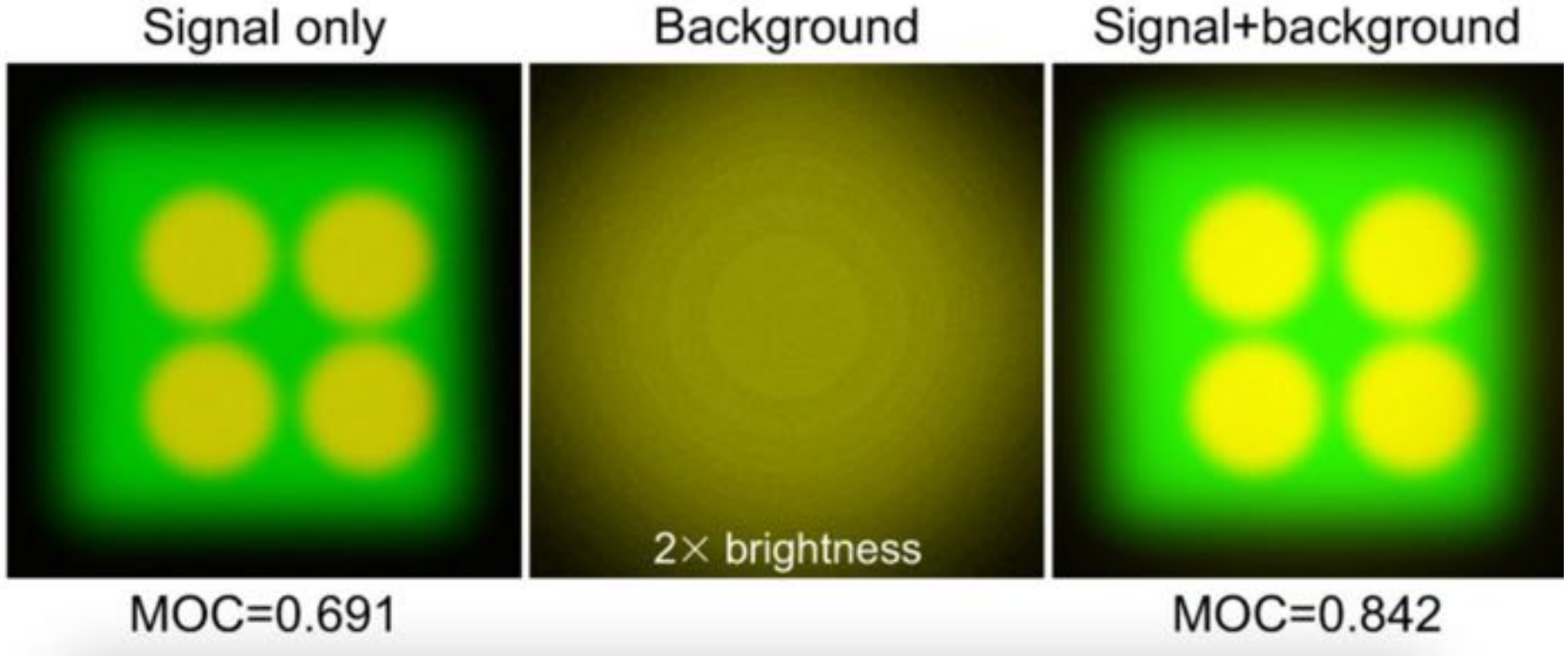


For high SNRs, changes in SNR doesn't change r_M

For low SNR it becomes harder to threshold the background

Aaron, J. S., Taylor, A. B. & Chew, T.-L. Image co-localization – co-occurrence versus correlation. *J. Cell Sci.* **131**, jcs211847 (2018).

Manders is sensitive to offset

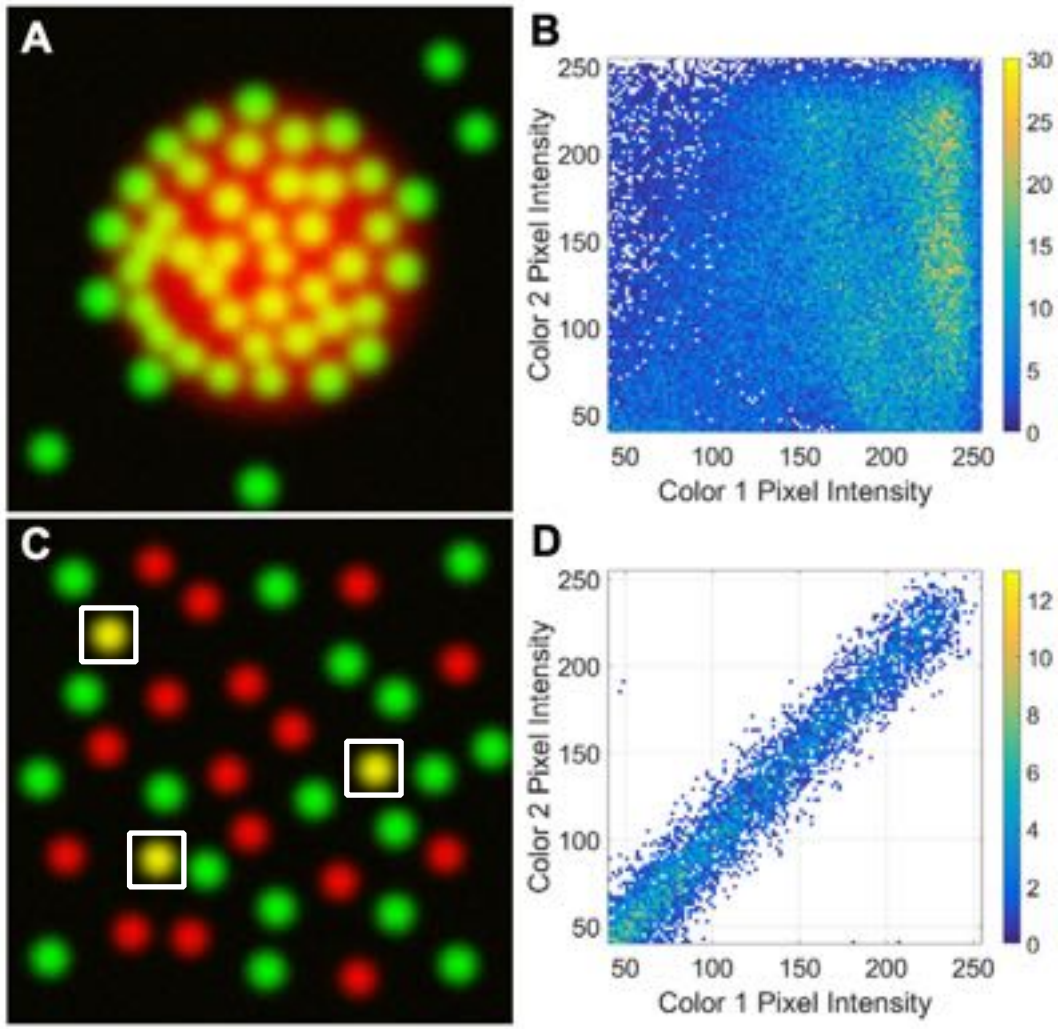


Adding a bit of non-uniform background changes r_M by >20%

Unless illumination correction is performed of course

Aaron, J. S., Taylor, A. B. & Chew, T.-L. Image co-localization – co-occurrence versus correlation. *J. Cell Sci.* **131**, jcs211847 (2018).

Co-occurence versus Correlation



A: High co-occurence ($r_M = 0.89$) but low correlation ($r_P = 0.11$). Pixel-intensities do not co-vary

B: Low co-occurence ($r_M = 0.14$) but high correlation ($r_P \simeq 1$ in *overlapping* regions)

Aaron, J. S., Taylor, A. B. & Chew, T.-L. Image co-localization – co-occurrence versus correlation. *J. Cell Sci.* **131**, jcs211847 (2018).

Manders' Coefficients

Straight forward interpretation, in some cases

Address several shortcomings of Pearson's coefficient

Doesn't provide algorithm for setting thresholds

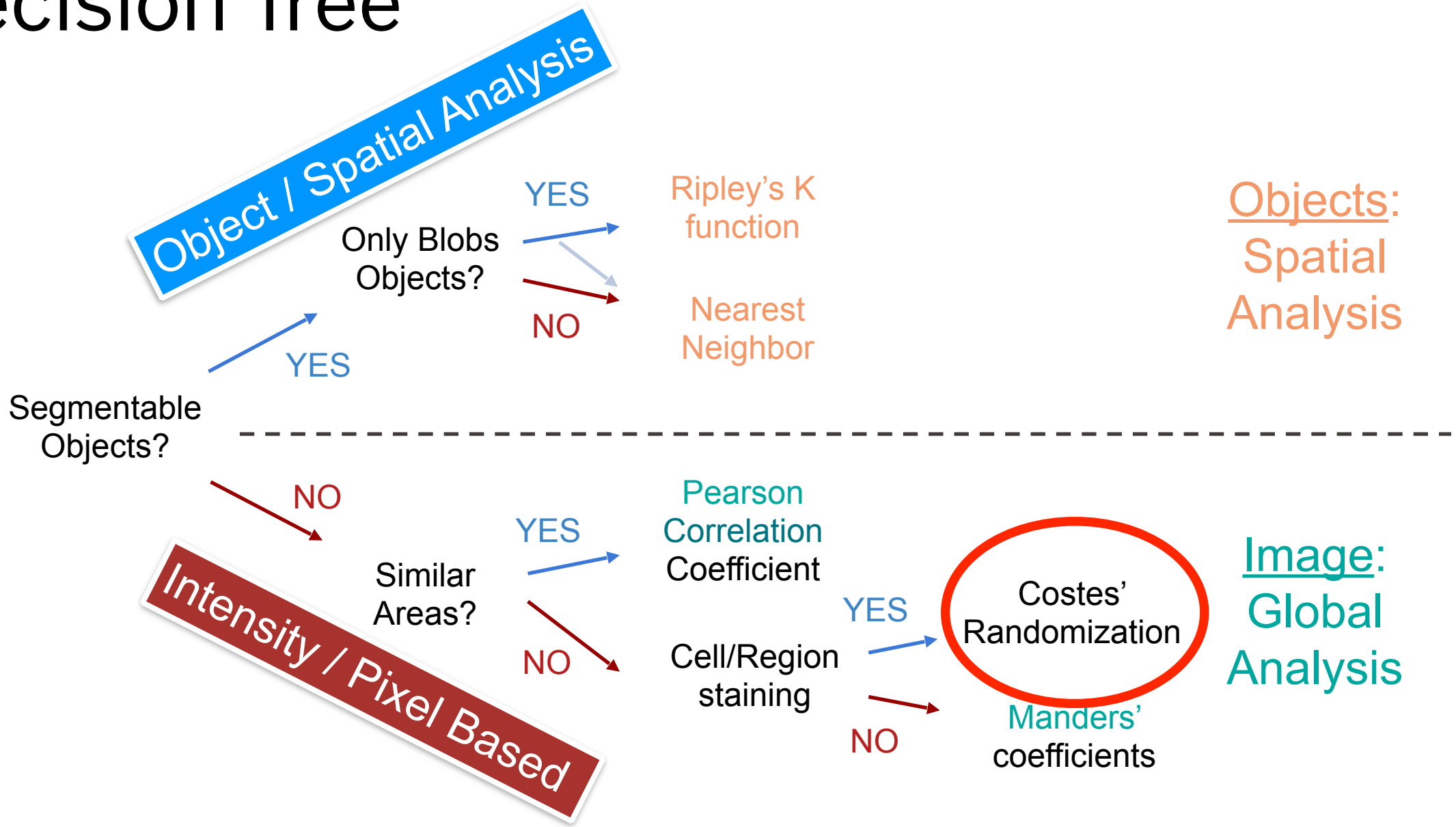
Doesn't return statistical significance

Fails: When there is random overlap (and in other ways)

Solution: Costes method

Costes

Decision Tree



From slide by Romain Guiet, BIOP, EPFL

Rationale for Costes

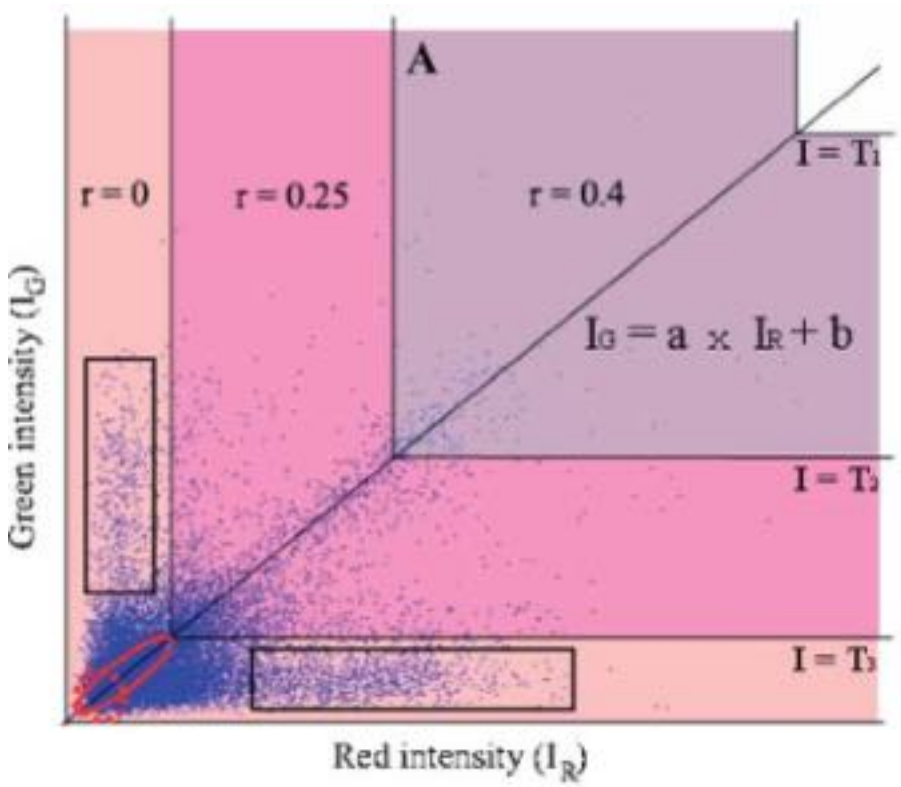
Address shortcomings of Manders' approach by

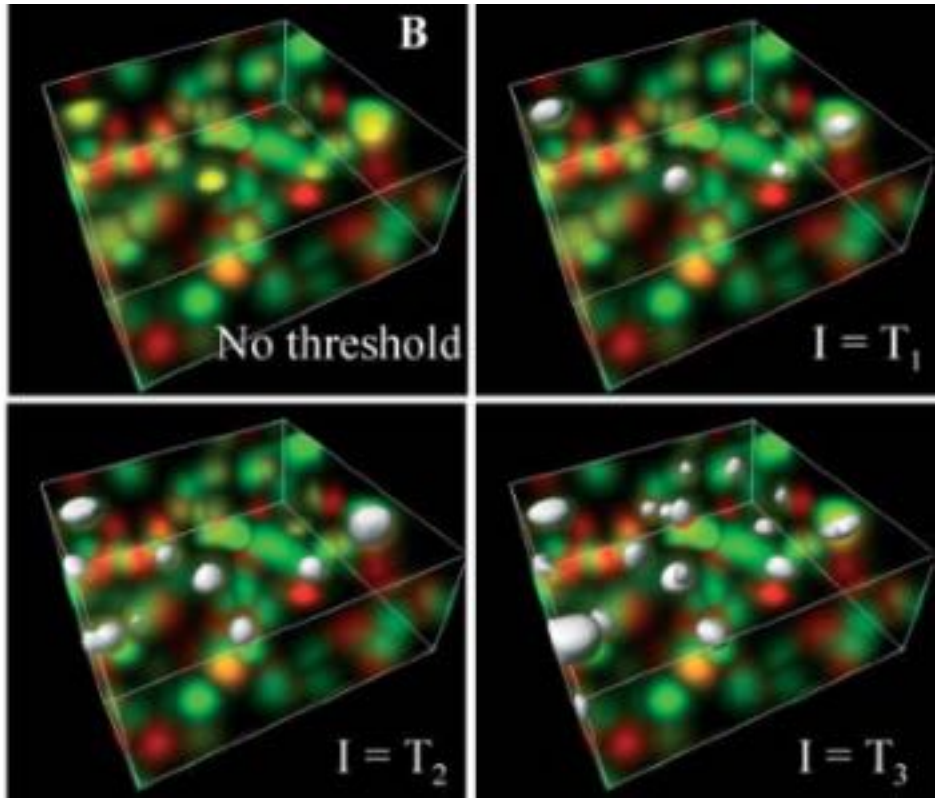
1. Providing *correlation-based* algorithm for image thresholds
2. "Shuffle" one image to control for random overlap
3. Return *p*-values (statistical significance) for overlap

Workflow for Thresholds

1. Preprocess images (noise reduction, illumination correction)
2. Fit straight line (least squares) to red-green scatter plot
3. Iterate thresholds until $r_P = 0$
4. Calculate M_1^{Costes} and M_2^{Costes} for all pixels above thresholds

Algorithmic Threshold Determination





Costes, S. V. *et al*. Automatic and Quantitative Measurement of Protein-Protein Colocalization in Live Cells. *Biophys. J.* **86**, 3993–4003 (2004).

Definition of Manders-Costes Coefficients

$$M_1^{Costes} = \frac{\sum_{R_i>T} R_i}{\sum R_i} \simeq M_1 \text{ and } M_2^{Costes} = \frac{\sum_{G_i>aT+b} G_i}{\sum G_i} \simeq M_2$$

Note the difference in *which* pixels are included in the nominator

The thresholds T and $aT + b$ depend on both R and G, through the straight-line fit $G = aR + b$

Controlling for Random Overlap

Densely packed objects tend to have *random* overlap

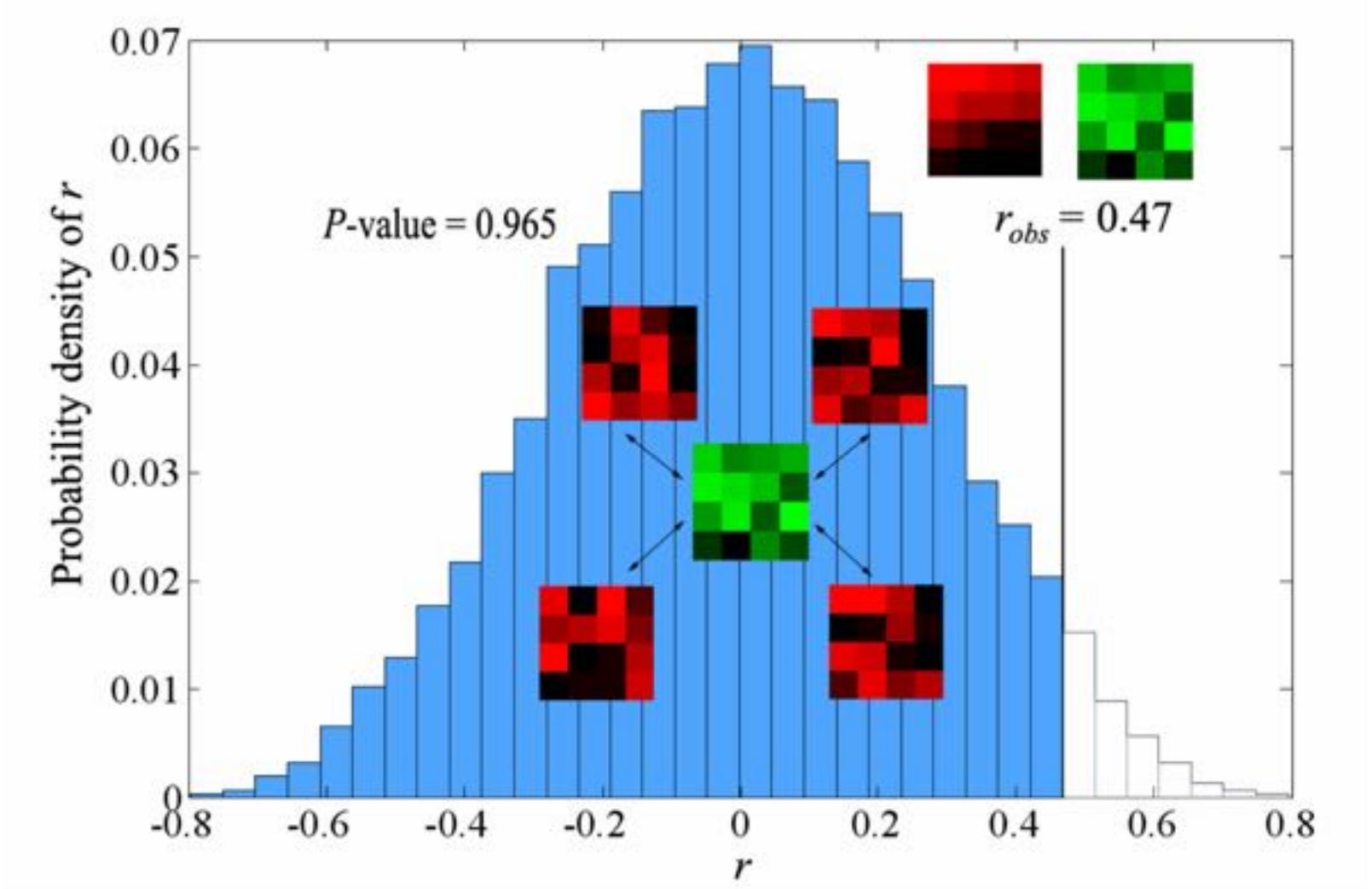
To *control* for this create images *without* true colocalization

Simplest approach: rotate one channel 90 degrees

Costes' approach: shuffle image blocks

Image blocks: Size of *typical object* of interest, but not smaller than size of *point spread function* (PSF)

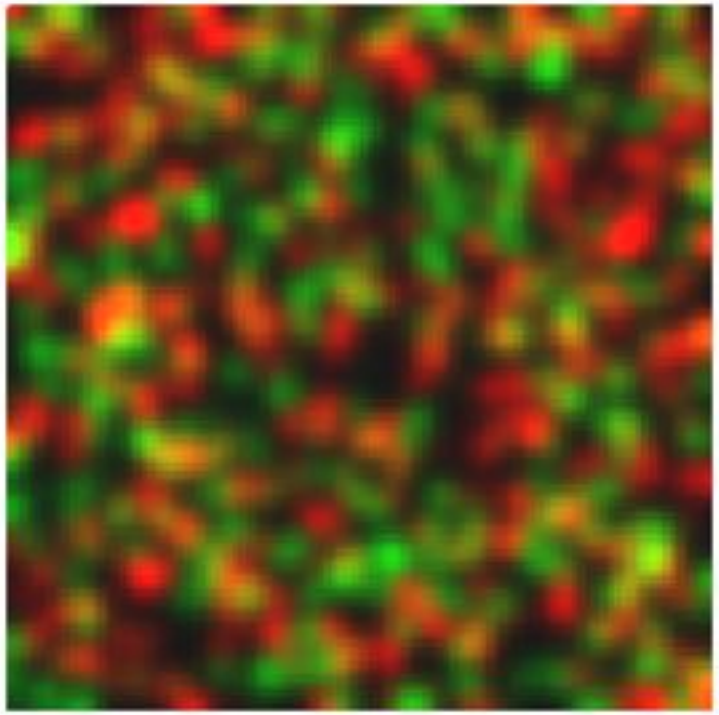
Pixel Scrambling

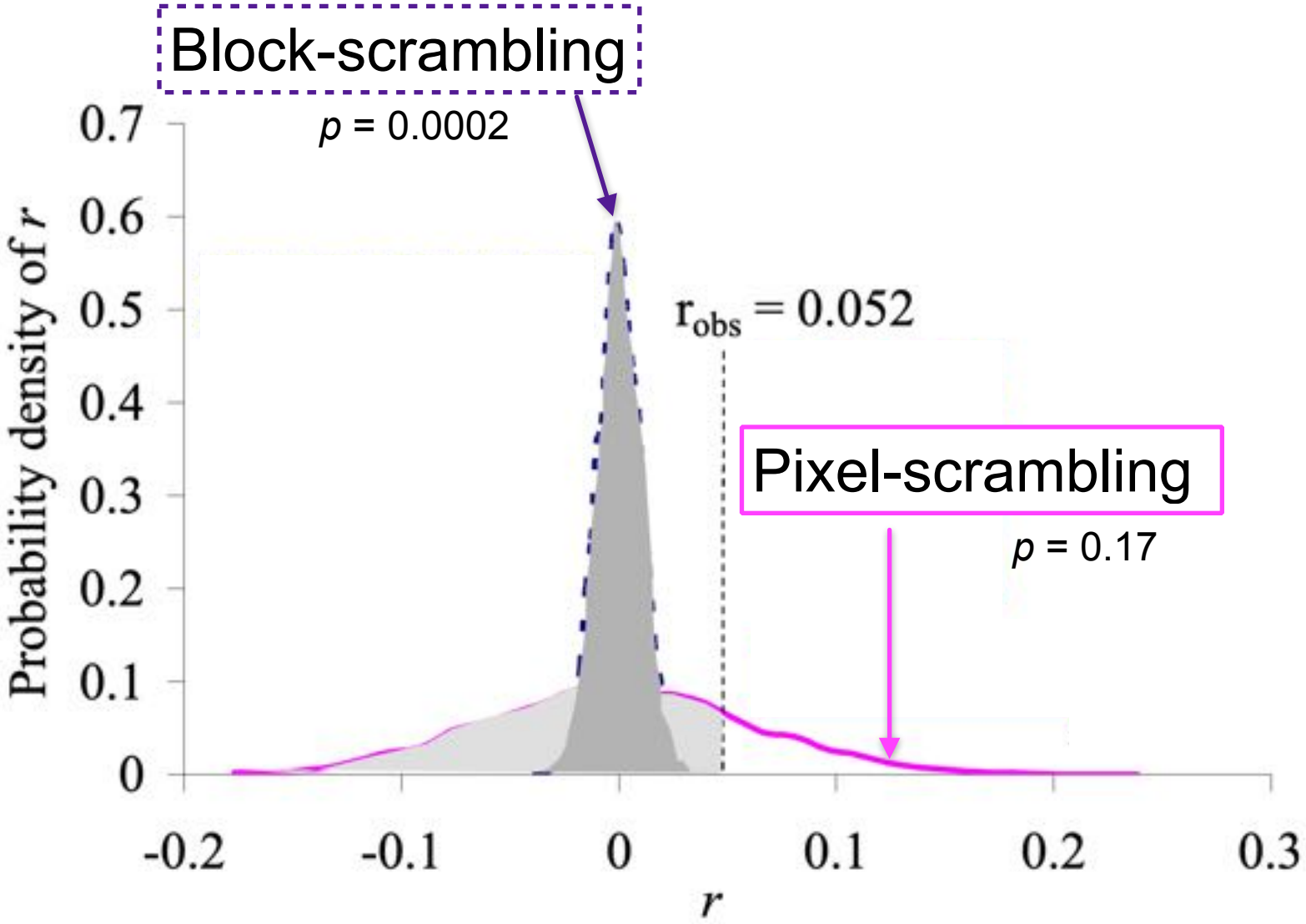


p=0.035

Costes, S. V. *et al*. Automatic and Quantitative Measurement of Protein-Protein Colocalization in Live Cells. *Biophys. J.* **86**, 3993–4003 (2004).

Block-scrambling in 5% Overlap Image

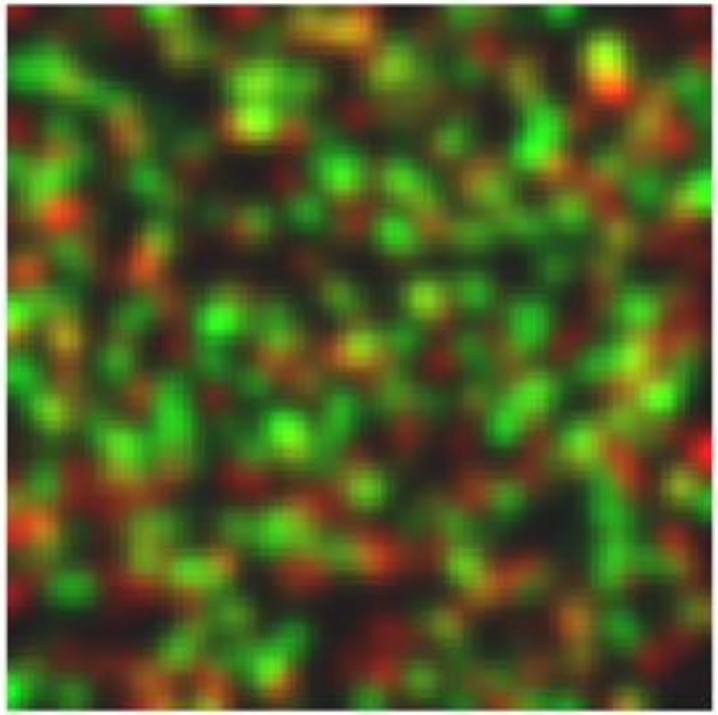




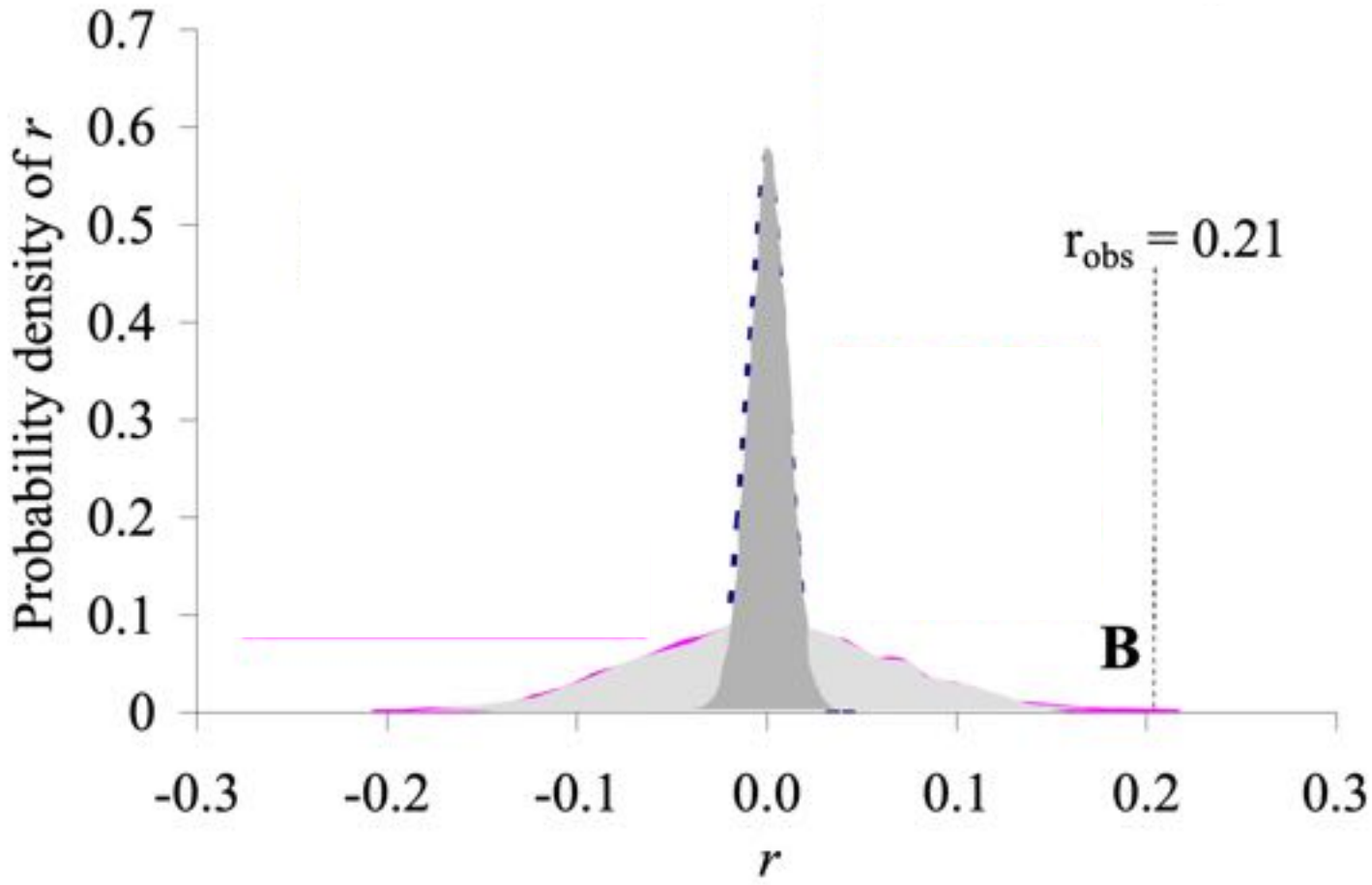
Pixel-scrambling: Wrong conclusion!
(that overlap is not stat. significant)

Costes, S. V. *et al*. Automatic and Quantitative Measurement of Protein-Protein Colocalization in Live Cells. *Biophys. J.* **86**, 3993–4003 (2004).

Block-scrambling in 20% Overlap Image



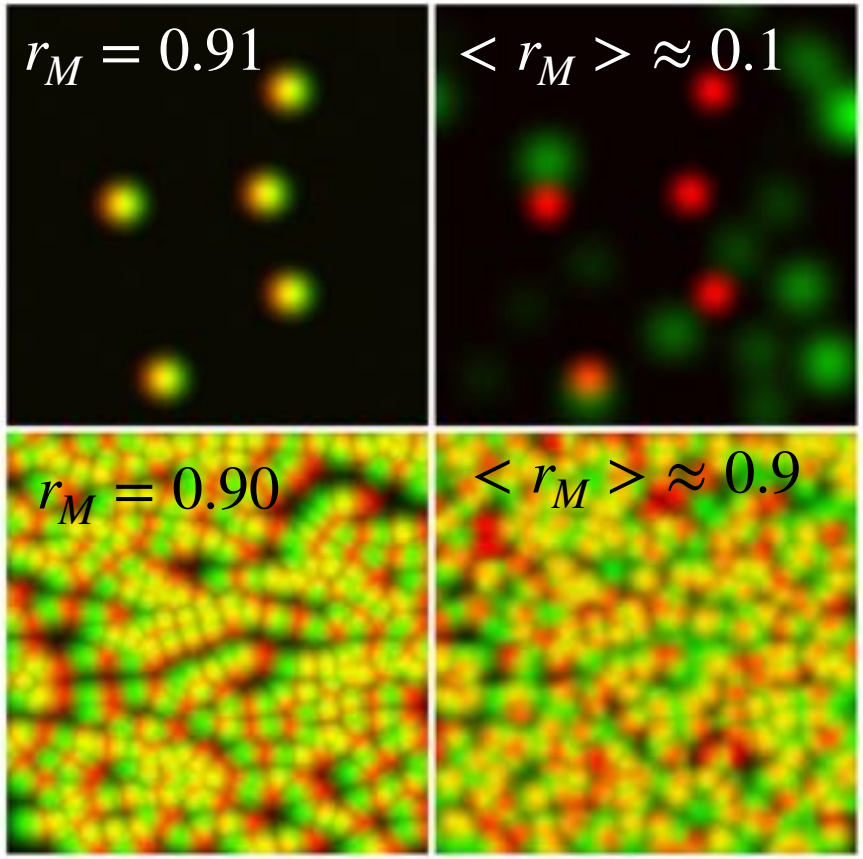
Scrambling of *either* pixels or blocks lead to conclusion of statistical significance of the $r_M = 0.21$ value with *p*=0.0004 and *p*=0.0002 respectively



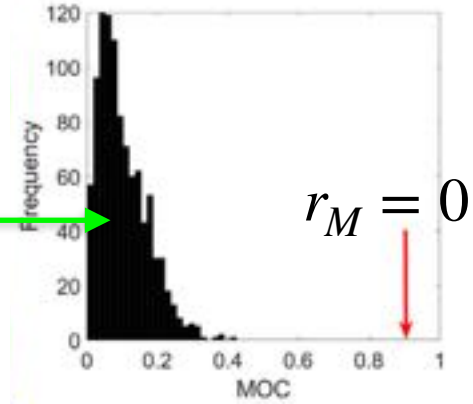
Costes, S. V. *et al*. Automatic and Quantitative Measurement of Protein-Protein Colocalization in Live Cells. *Biophys. J.* **86**, 3993–4003 (2004).

Random Overlap and "Real" Colocalization

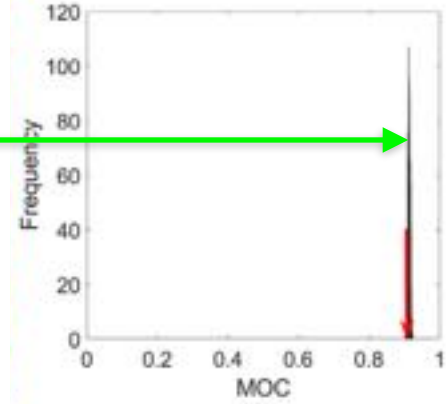
Green scrambled



Scramble green 1000 times



Scramble green 1000 times



Aaron, J. S., Taylor, A. B. & Chew, T.-L. Image co-localization – co-occurrence versus correlation. *J. Cell Sci.* **131**, jcs211847 (2018).

Costes' Method

Automatic calculation of thresholds and control for random overlap

Returns statistical significance (*p*-value)

Requires careful preprocessing of image, like Manders

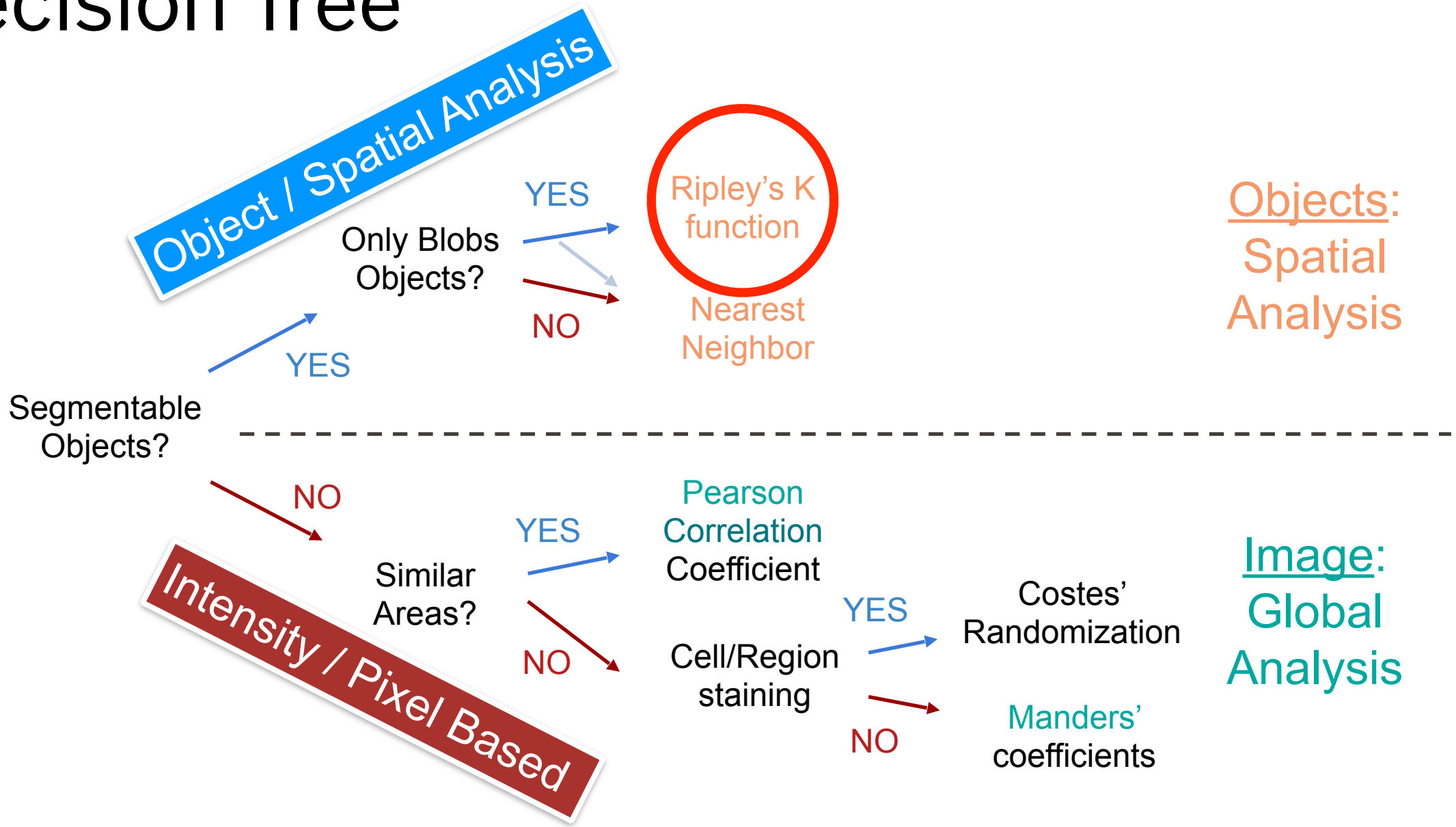
The two thresholds are not independent (linear dependence)

Fails: When object don't overlap or background hard to filter out

Solution: Object Based and Spatial Statistics

Object based

Decision Tree



From slide by Romain Guiet, BIOP, EPFL

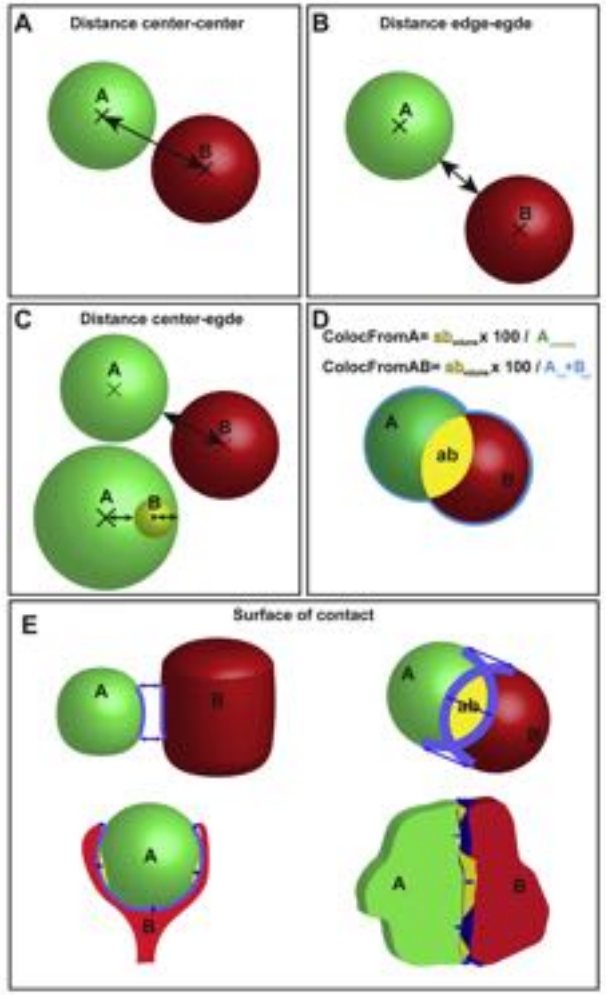
Beyond Pixels: Object based Analysis

What if you are only interested in the *number* of interacting objects, irrespective of size, shape, and intensity?

Determine each object and decide if it *interacts* with another

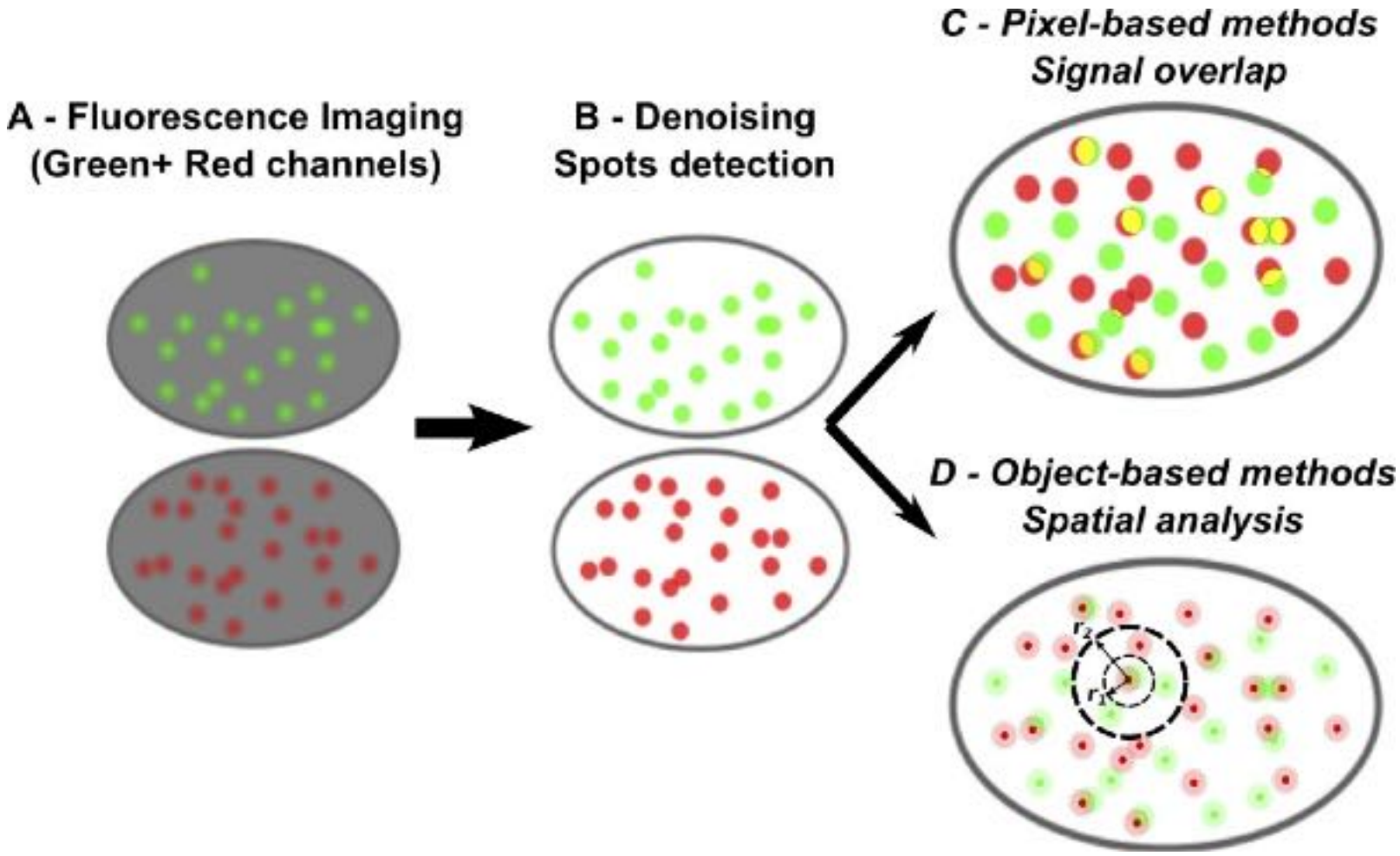
Either segment and look for *overlap*

Or, detect and measure *distances*



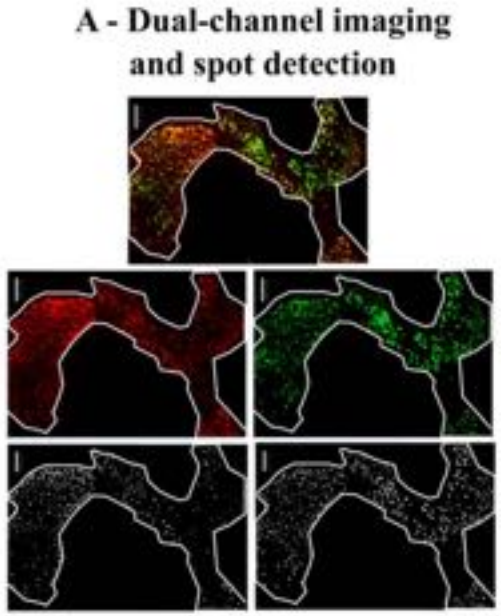
Gilles, J.-F., Santos, M. D., Boudier, T., Bolte, S. & Heck, N. DiAna, an ImageJ tool for object-based 3D co-localization and distance analysis. *Methods* **115**, 55–64 (2017).

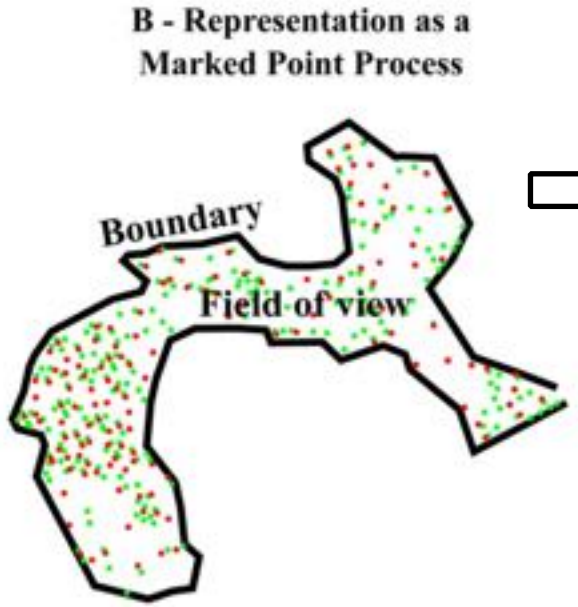
Pixel versus Object based Analysis

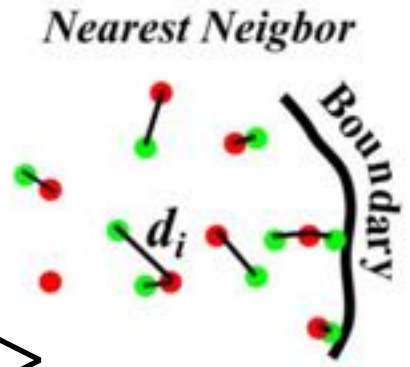


Lagache, T., Sauvonnet, N., Danglot, L. & Olivo-Marin, J.-C. Statistical analysis of molecule colocalization in bioimaging. *Cytometry* **87**, 568–579 (2015).

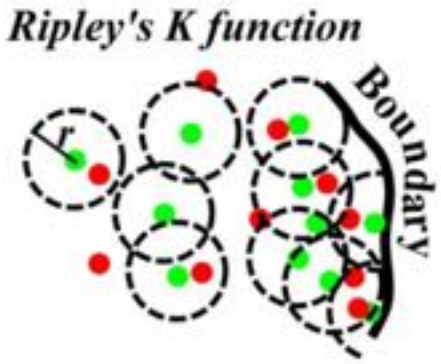
Treating Objects as Points





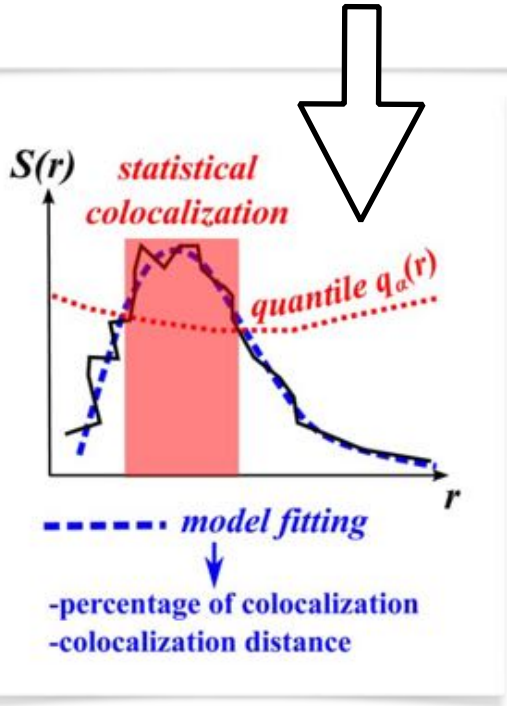


$$S(r) = \frac{1}{n_1}\sum_{i=1}^{n_1} \mathbf{1}\,(d_i < r)$$



$$S(r) = \frac{|\Omega|}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \mathbf{1}\,(d_{ij} < r)\, b(i,j,r)$$

boundary correction



Lagache, T., Sauvonnet, N., Danglot, L. & Olivo-Marin, J.-C. Statistical analysis of molecule colocalization in bioimaging. *Cytometry* **87**, 568–579 (2015).

Ripley's K and L Functions

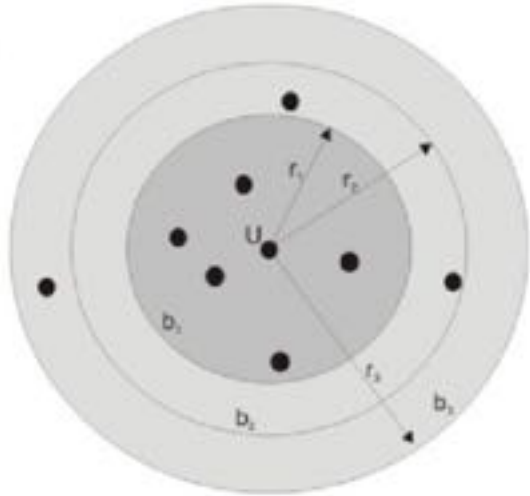
Descriptive statistics for detecting *deviations* from spatial homogeneity

$$K(r) = (n\lambda)^{-1} \sum_{i \neq j} \mathbf{1}_{d_{ij} < r} \quad \text{and} \quad L(r) = \sqrt{\frac{K(r)}{\pi}}$$

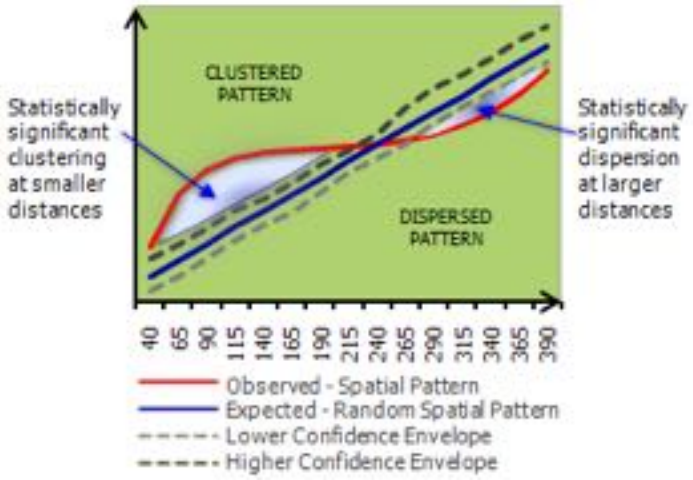
For given j, the sum gives number of points closer than r

For homogeneous 2D distribution $K(r) = \pi r^2$ and $L(r) = r$

(n: total number of points, λ: average density)



Tonini, Marj & Pedrazzini, Andrea & Penna, Ivanna & Jaboyedoff, Michel. (2012). Spatial pattern of landslides in Swiss Rhone Valley. Natural Hazards. 73. 10.1007/s11069-012-0522-9.



https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-multi-distance-spatial-cluster-analysis-ripl.htm

Is is really that simple? No!

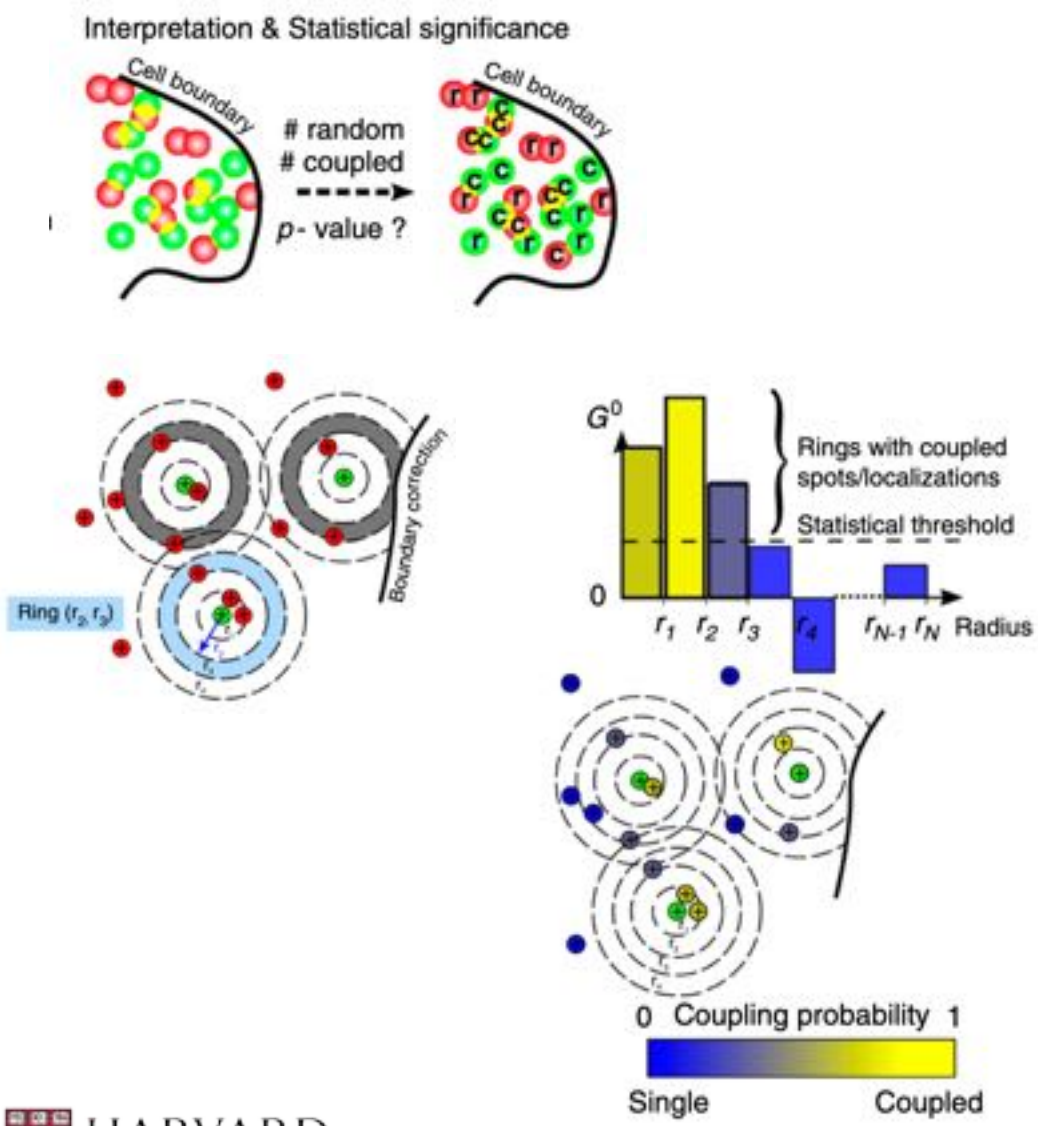


Table 2 Mathematical variables

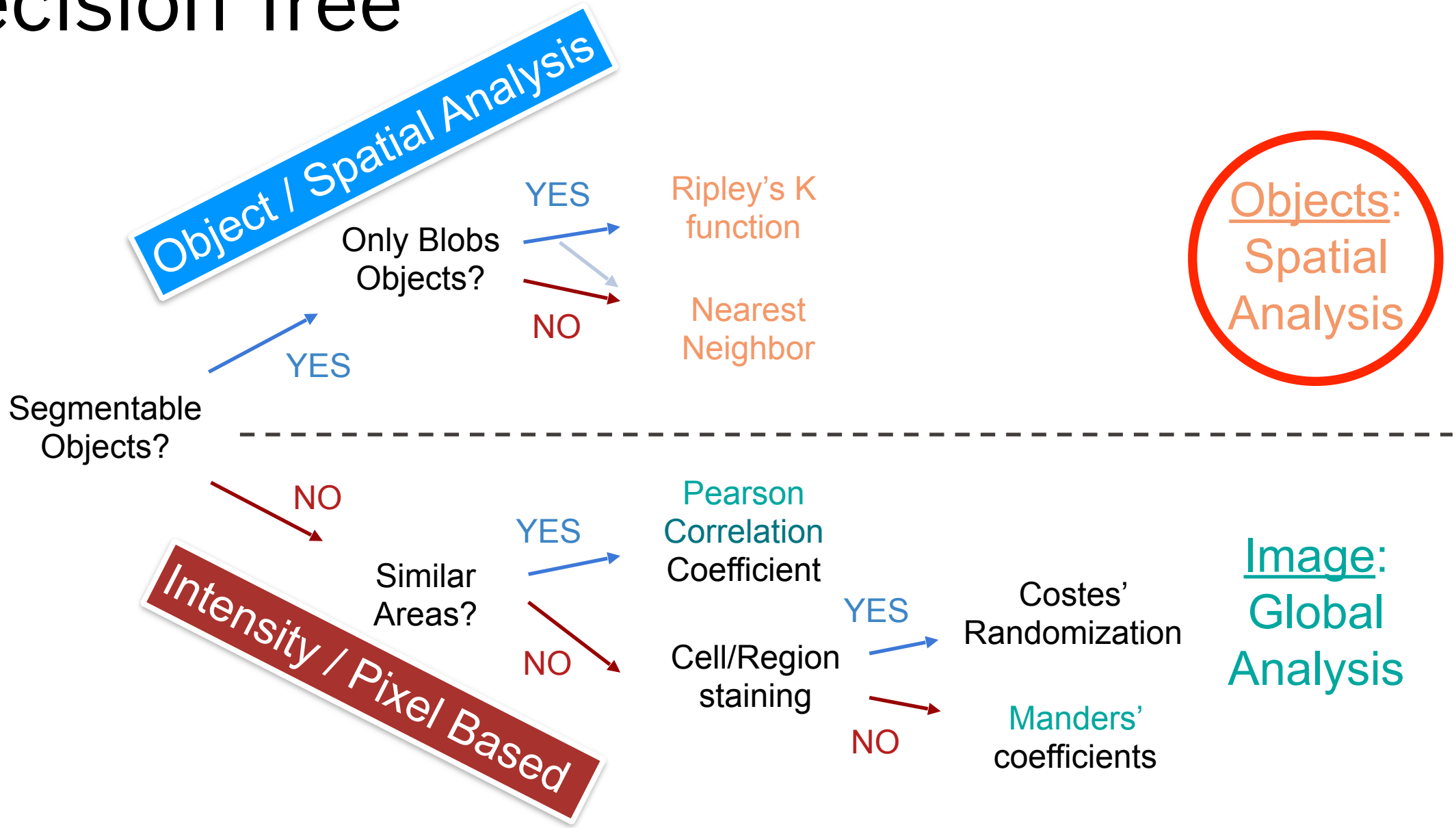
Name	Mathematical Expression	Meaning
Point-process $i = 1, 2$	$A_{i=1,2}$	Positions of all the objects (spots or localisations) $i = 1, 2$
Number of objects $i = 1, 2$	$n_{i=1,2}$	Number of objects in $A_{i=1,2}$
Distance between objects	$d(\mathbf{x}, \mathbf{y})$	Distance between (green) object located at position **x** and (red) object located at **y**
Boundary correction	$k(\mathbf{x}, \mathbf{y})$	Corrects the under-estimation of object's neighbors near the ROI boundary (Supp. Methods)
Ripley's K function	$K(r) = \frac{\text{Volume}(ROI)}{n_1 n_2} \sum_{x,y} 1_{[d(x,y)\le r]} k(x,y)$	Counts the number of (red) objects at a distance below r from (green) objects
Searching distances	$0 = r_0 < r_1 < \cdots < r_N$	Increasing distances around (green) objects where the K function is computed
Rings	$Ring(r_i, r_{i+1})$	Sub-region of the ROI that contains points (**y**) located at a distance $r_i \le d(\mathbf{x}, \mathbf{y}) \le r_{i+1}$ from a (green) object (**x**)
Ripley-based vector	$G = [K(r_{i+1}) - K(r_i)]_{0 \le i \le N-1}$	Counts the number of (red) objects inside concentric rings around (green) objects
Number of rings	N	Number of rings and length of the vector G
Mean of **G**	$\mu = [\mu_i]_{0 \le i \le N-1}$ with $\mu_i = \pi(r_{i+1}^2 - r_i^2)$ (2D) or $\mu_i = \frac{4}{3}\pi(r_{i+1}^3 - r_i^3)$ (3D)	Expected mean of G under the null hypothesis of A_2 randomness
Standard deviation of **G**	$\sigma = [\sigma_i]_{0 \le i \le N-1}$	Standard deviation of G under the null hypothesis of A_2 randomness (see Supplementary Methods)
Rings' overlapping matrix	$\mathbf{A} = [\alpha_{ij}]_{0 \le i,j \le N-1}$ with, $\alpha_{ij} = \frac{\text{Volume}\{Ring(r_i, r_{i+1}) \cap Ring(r_j, r_{j+1})\}}{\text{Volume}\{Ring(r_i, r_{i+1})\}}$	Proportion of the volume of $Ring(r_i, r_{i+1})$ that overlaps with $Ring(r_j, r_{j+1})$
Reduced Ripley-based vector	$G^0 = \frac{1}{\sigma} A^{-1}.\lvert G - \mu \rvert$	Reduced Ripley-based vector with zero mean and unit variance (under the null hypothesis of A_2 randomness)
Statistical threshold	$T(N) = \sqrt{2\log(N)}$	Statistical threshold to extract rings with coupled (red) objects.
Number of couples per ring	$C = \left[1_{G_i^0 \ge T(N)} \frac{n_1 n_2}{\text{Volume}(ROI)} (G_i - \mu_i)\right]_{0 \le i \le N-1}$	Statistical estimate of the number of couples per ring.
Couples without overlapping	$\bar{C} = A^{-1}.C = \left[1_{G_i^0 \ge T(N)} \frac{n_1 n_2}{\text{Volume}(ROI)} G_i^0\right]_{0 \le i \le N-1}$	Number of couples corrected for rings' overlapping.
Number of pairs	$\frac{n_1 n_2}{\text{Volume}(ROI)} G$	Total number of object pairs inside rings.
Coupling probability	$P(x,y) = \sum_{i=0}^{N-1} 1_{r_i \le d(x,y) < r_{i+1}} \frac{1_{G_i^0 > T(N)} \pi_i G_i^0}{G_i}$	Probability that a (green) object located at position x is coupled with a (red) object located at y
Coupling index	$Coupling\ Index(A_i) = \frac{1}{n_1} \sum_{x,y} P(x,y)$	Mean number of coupled objects (i.e., probability-weighted) in each population $A_{i=1,2}$
Mean coupling distance	$Mean\ Coupling\ Distance = \frac{\sum_{x,y} P(x,y) d(x,y)}{\sum_{x,y} P(x,y)}$	Probability-weighted distance between coupled objects

Lagache, T. *et al*. Mapping molecular assemblies with fluorescence microscopy and object-based spatial statistics. *Nat. Commun.* **9**, 698 (2018).

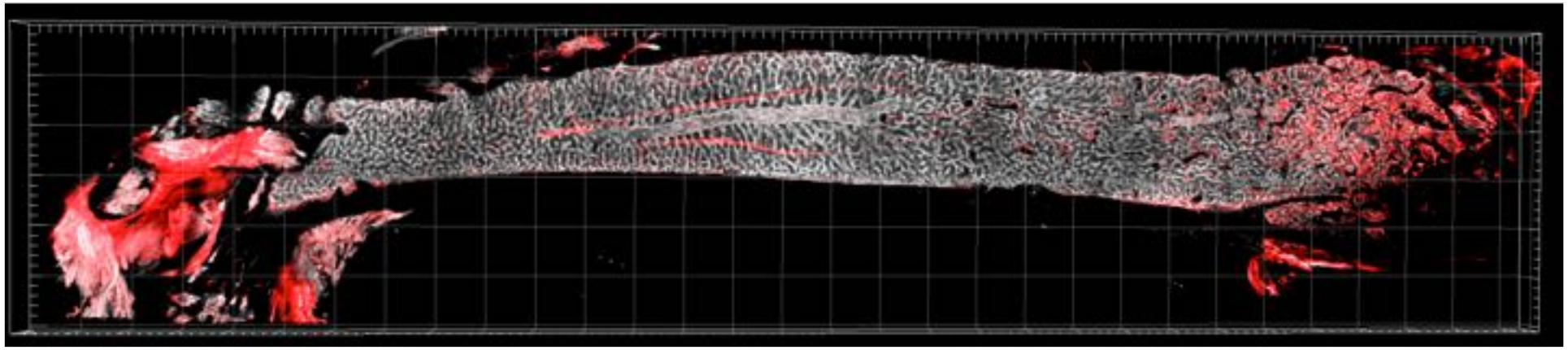
Example

Treating Objects as Objects

Decision Tree



3D Microarchitecture of Bone Marrow Vascular System

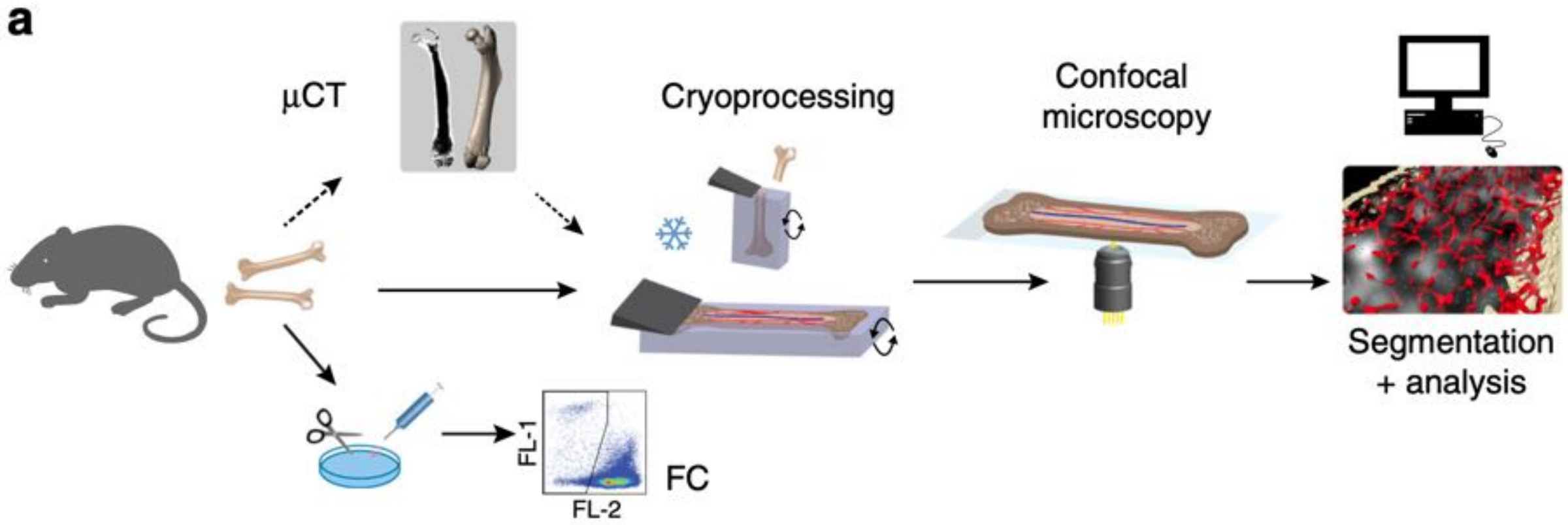


Prof. Cesar Nombela-Arrieta
Alvaro Gomariz

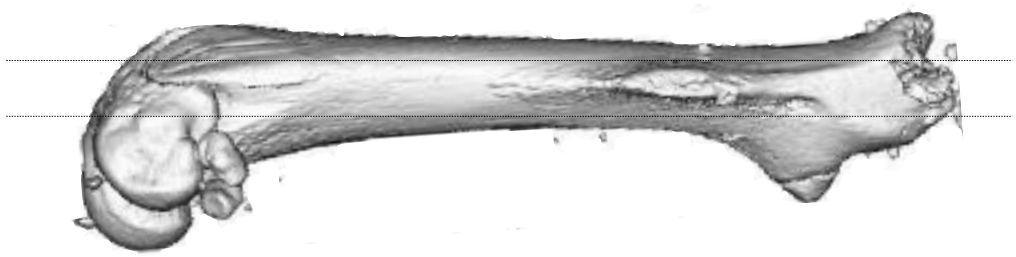
Universität Zürich[UZH]

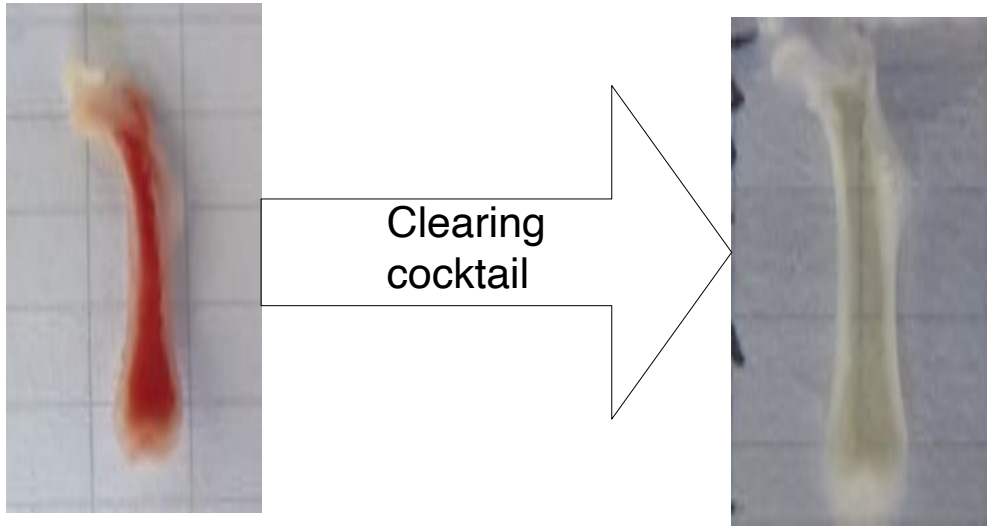
Results obtained without DL (DL happened while in review):
"Quantitative spatial analysis of haematopoiesis-regulating stromal cells in the bone marrow microenvironment by 3D microscopy"
Gomariz et al
Nature Communications, volume 9, Article number: 2532 (2018)

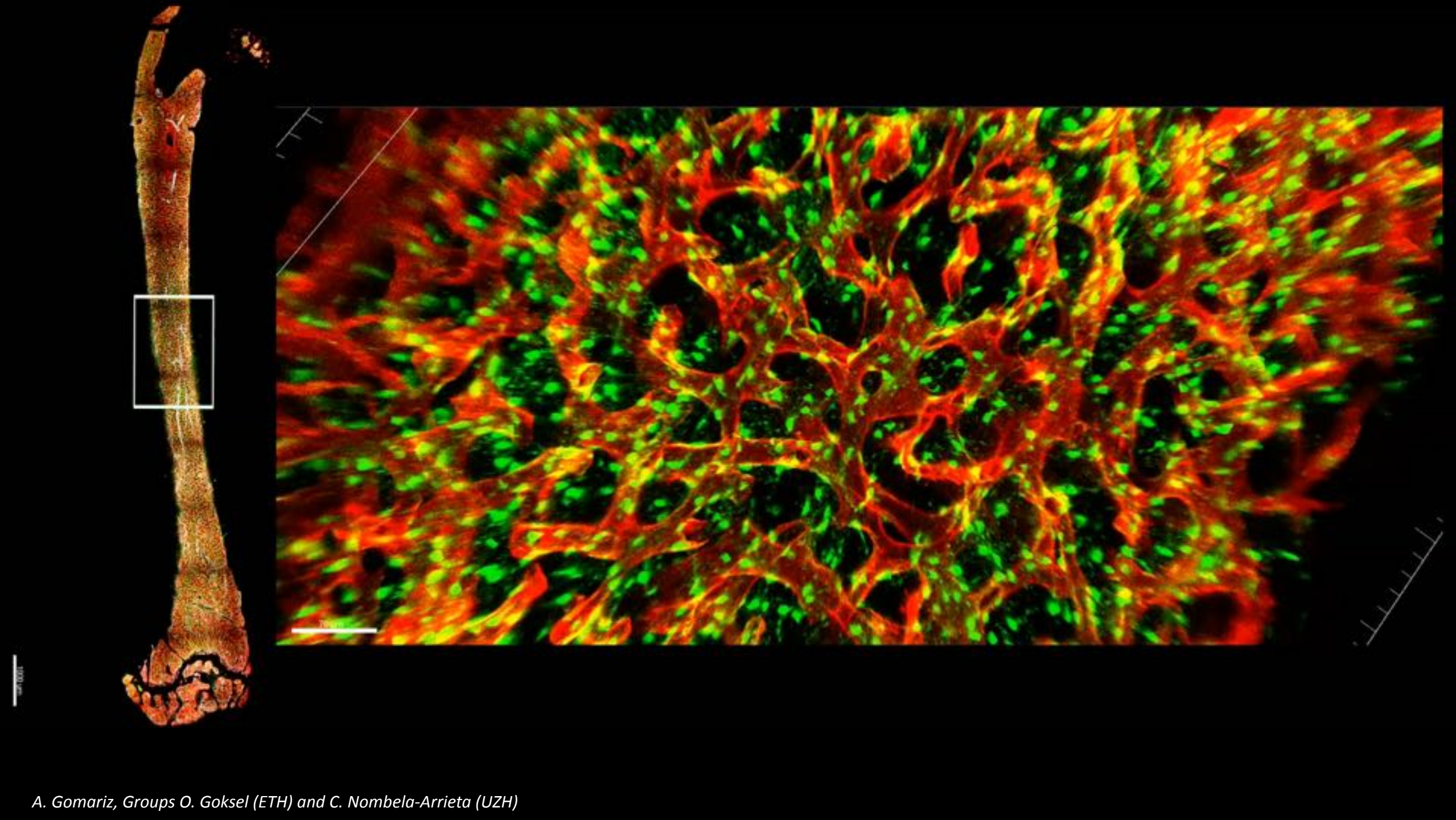
Deep Tissue 3D imaging of thick Bone Marrow Slices



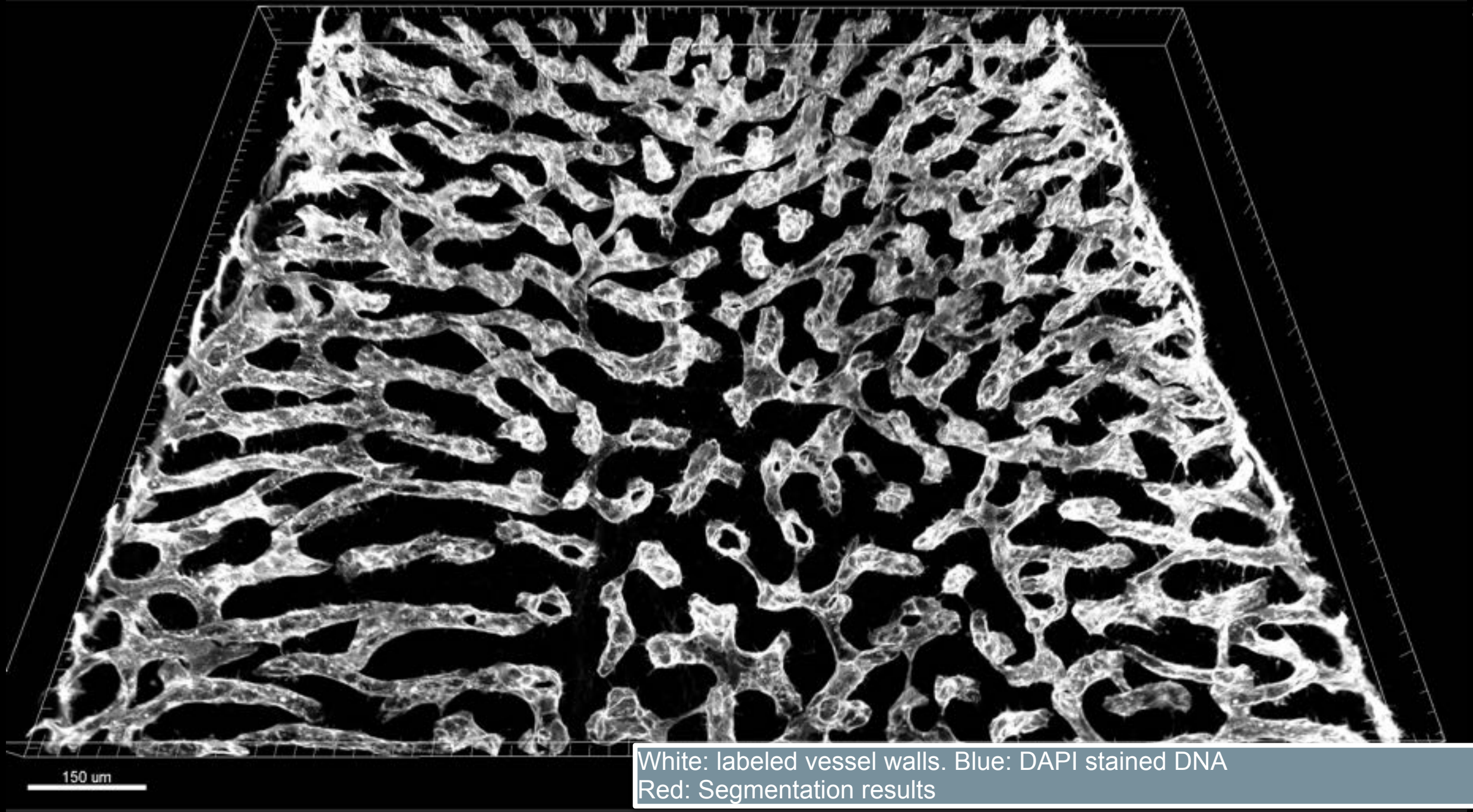
Deep Tissue 3D imaging of thick Bone Marrow Slices







A. Gomariz, Groups O. Goksel (ETH) and C. Nombela-Arrieta (UZH)

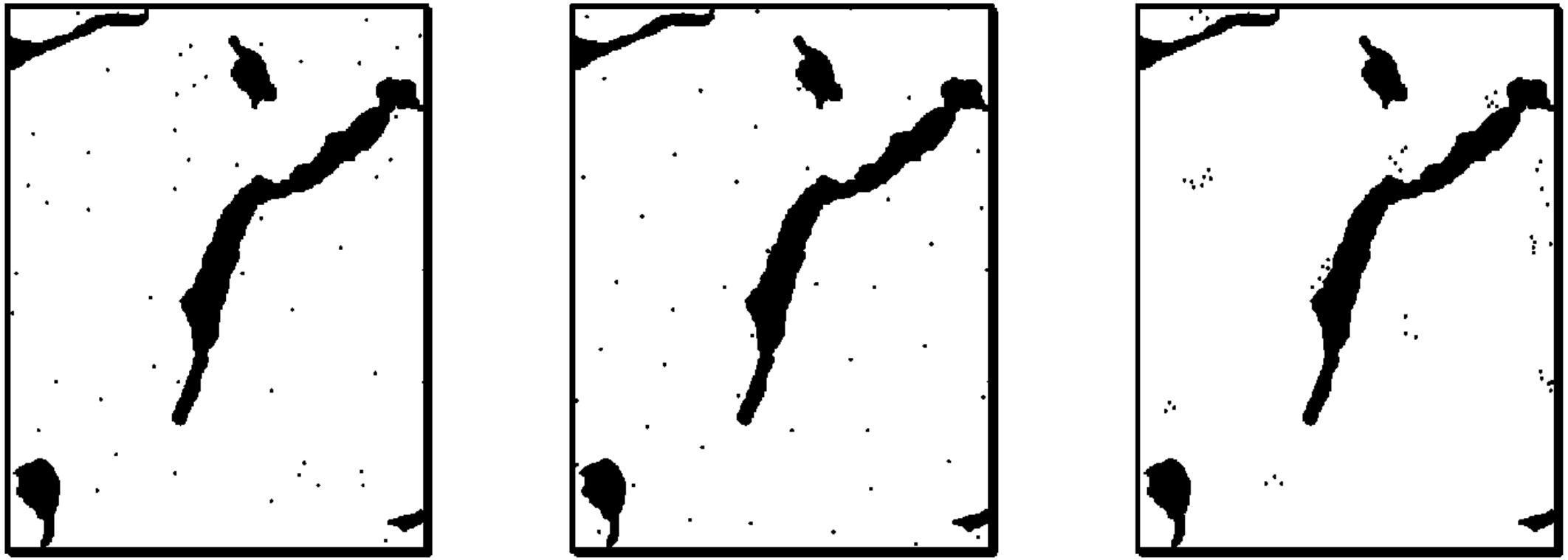


White: labeled vessel walls. Blue: DAPI stained DNA
Red: Segmentation results

A. Gomariz, Groups O. Goksel (ETH) and C. Nombela-Arrieta (UZH)

Spatial Statistics

Examining effect of cells interacting with *each other*

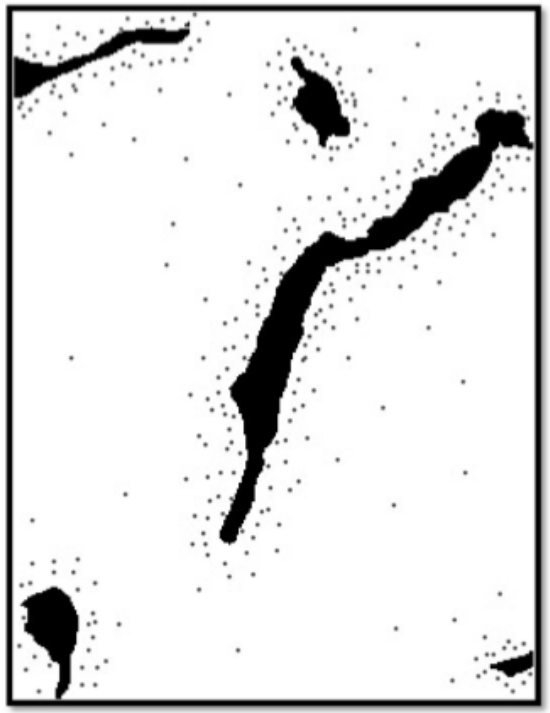


(a) Independence *(b) Avoidance* *(c) Clustering*

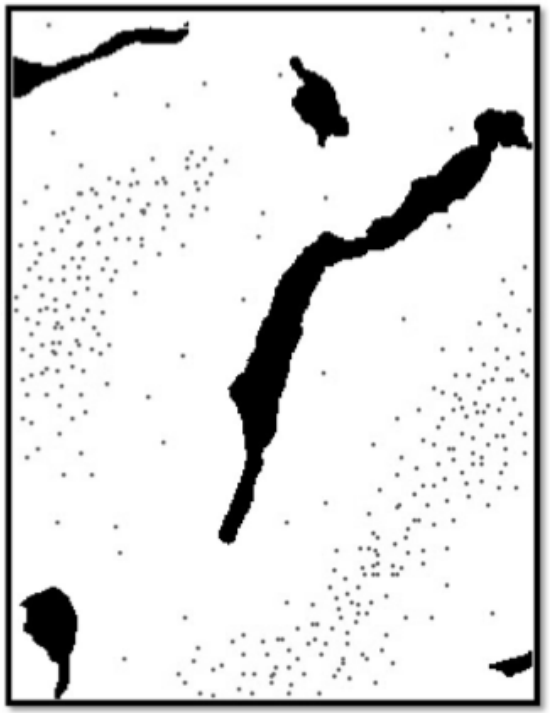
Spatial Statistics

Examining effect of cells interacting with fixed *spatial structure*

(a) Poisson



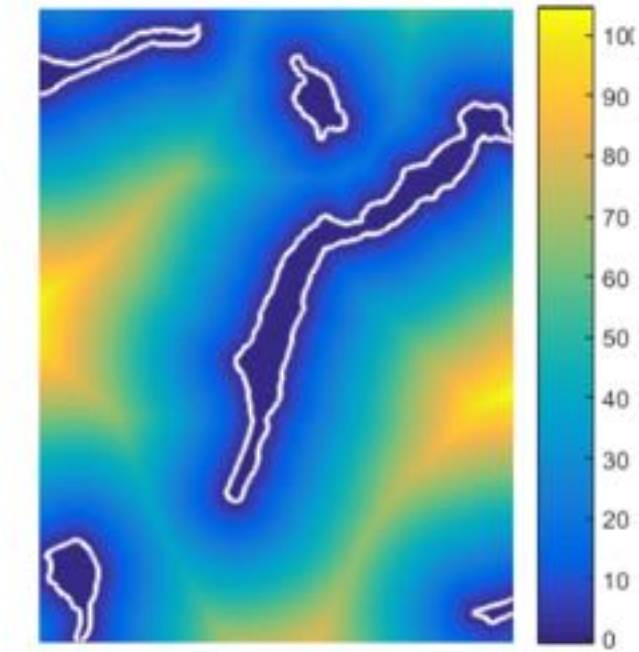
(b) Attraction



(c) Repulsion

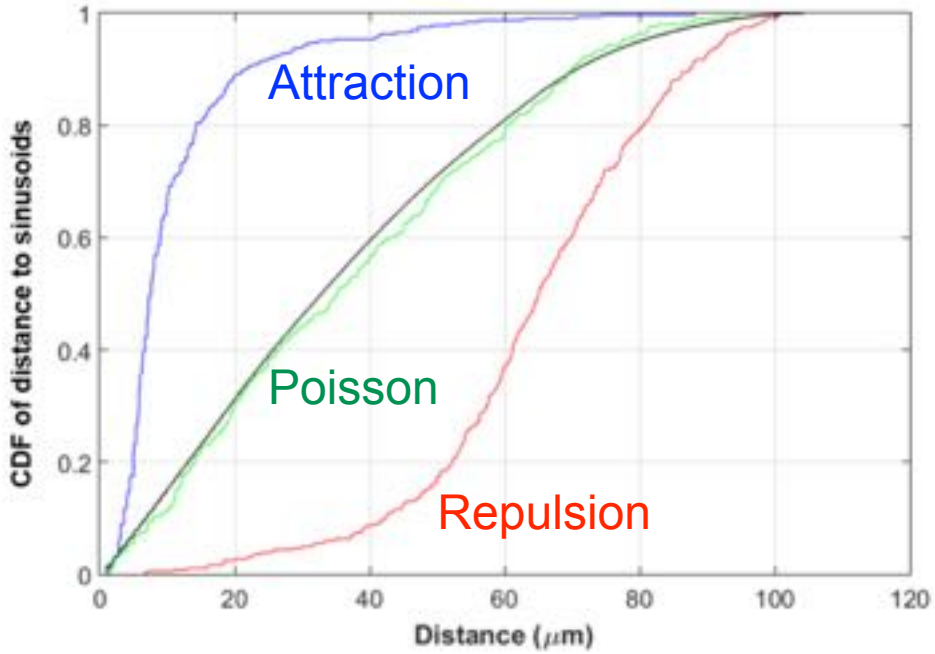
The volume is confined by the segmented sinusoidal network (Small World Model).

Spatial Statistics



(d) Distance transform

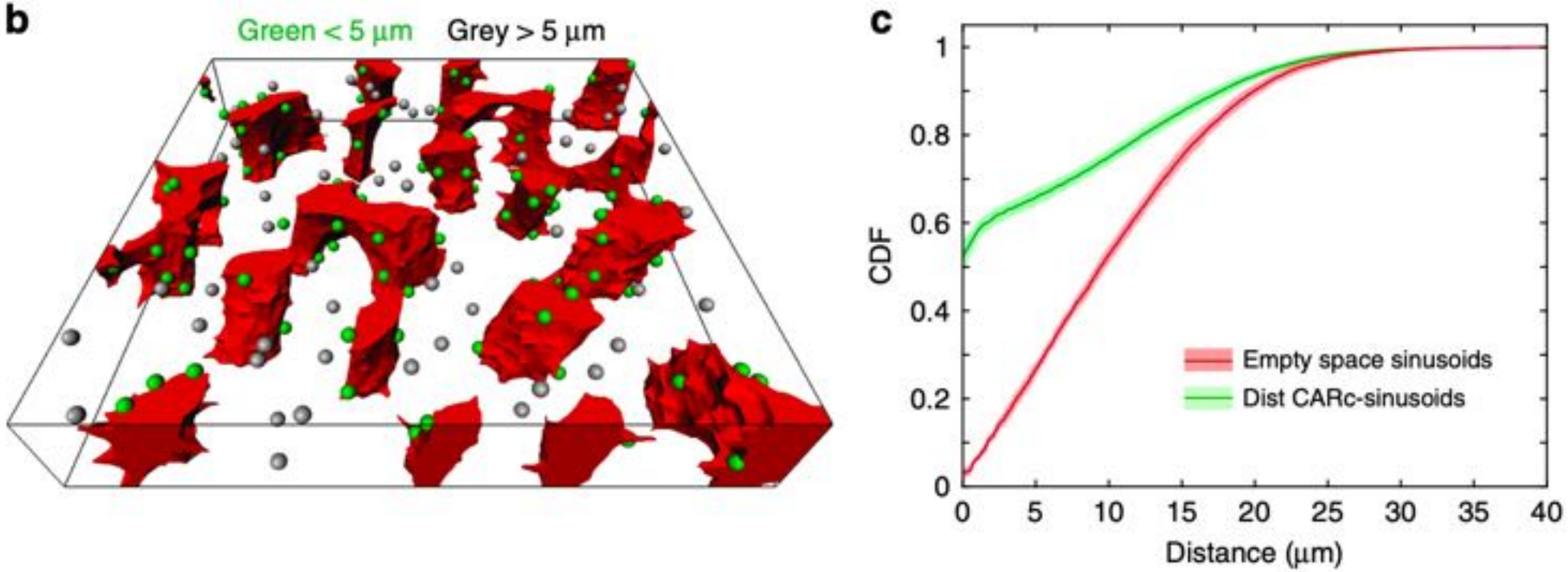
Empty space distance transform



(e) CDF distance transform

Panel (d) gives the “x-axis” (abscissa) in panel (e)

Spatial Statistics - CAR cells relative to sinusoids



***CARcs accumulate in close physical contact with sinusoidal vessel walls*.**
b Rotated 3D view of a rendered volume from the segmented image
c Side-by-side comparison of the CDF of the distance to nearest sinusoid evaluated at all positions, as well as evaluated at CARc centroids. Solid lines represent mean distance and envelopes indicate standard deviations.

A. Gomariz et al, Nature Communications 9, 407 (2018).

From Research to Software

This analysis was done in MATLAB with Imaris (Bitplane)

Bitplane then worked directly with Alvaro Gomariz

Now, a spatial statistics module is available in Imaris

- Measures distances, performs 3D randomizations, returns probability distributions for experimental and simulated results

Object based and Spatial Statistics

Very powerful and flexible

Goes beyond standard colocalization

Mathematically demanding to do right

Requires segmentation or localization of objects

Fails: When objects cannot be defined/segmented

Solution: Manders, Costes, better image analysis or data

Pixels versus Object- and Spatial-Analysis

Pixel based

Requires images to be carefully corrected first

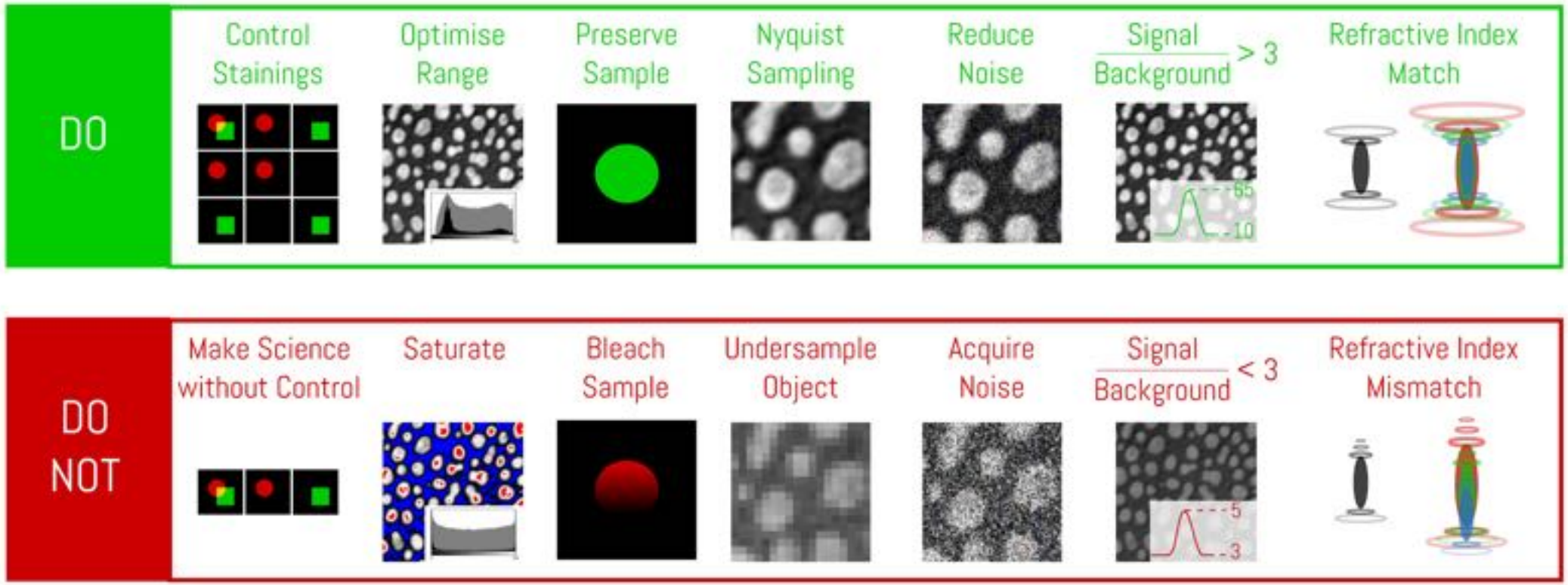
(Deceptively) easy and flexible to apply

Object based

Requires segmentation or detection

Allows for full statistical analysis w. significance testing

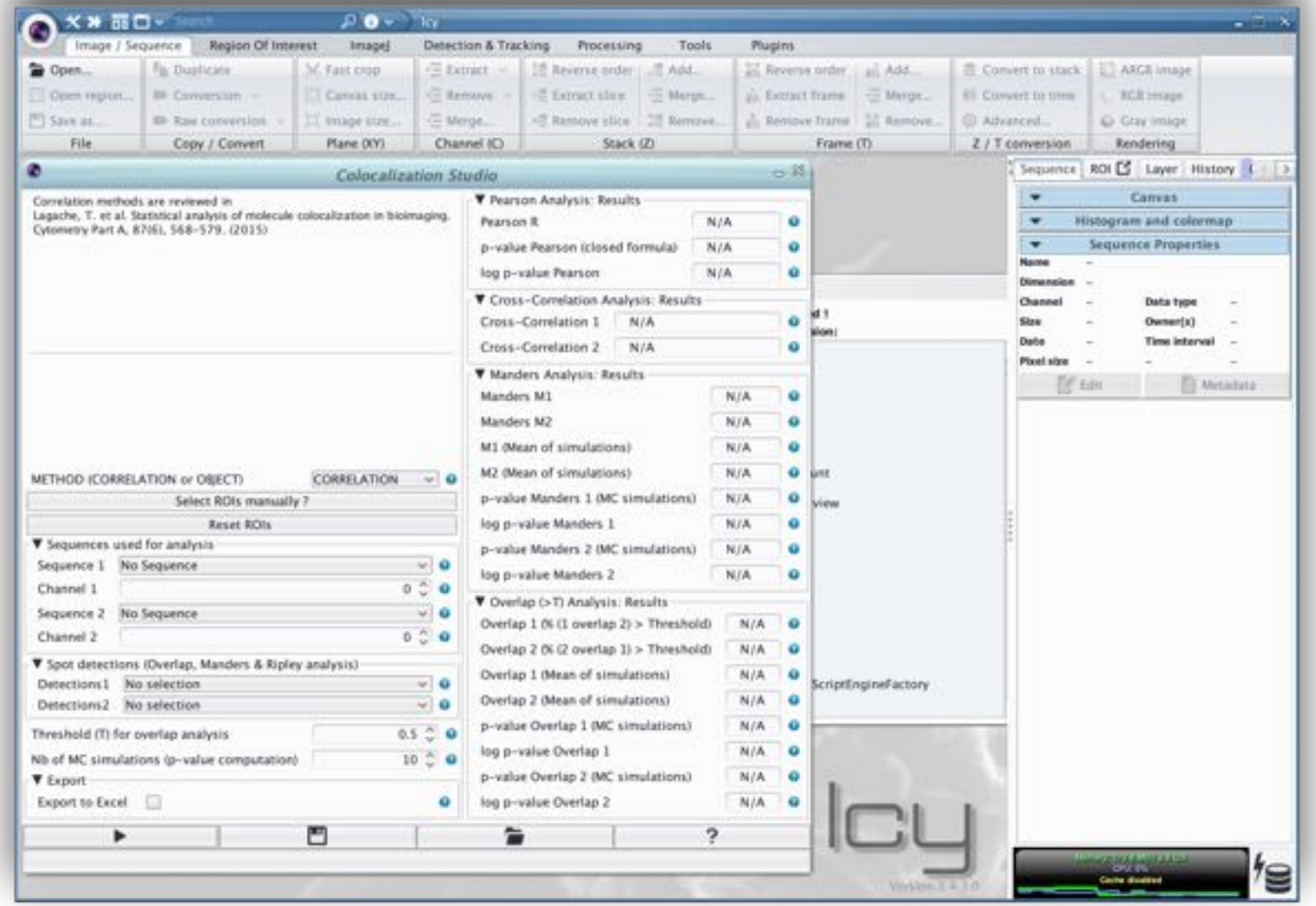
Minimizing Imaging Artifacts

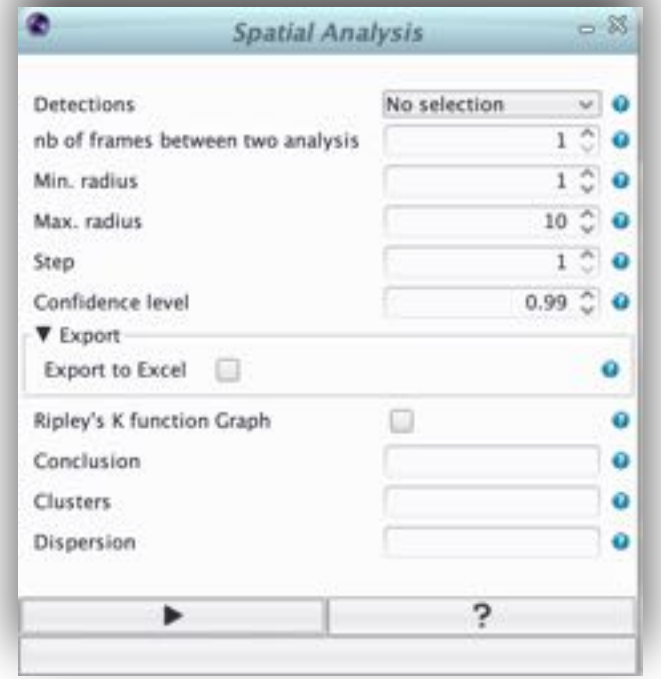


From slides by Romain Guiet, BIOP, EPFL

Software & Resources

ICY: Colocalization Studio & Spatial Analysis



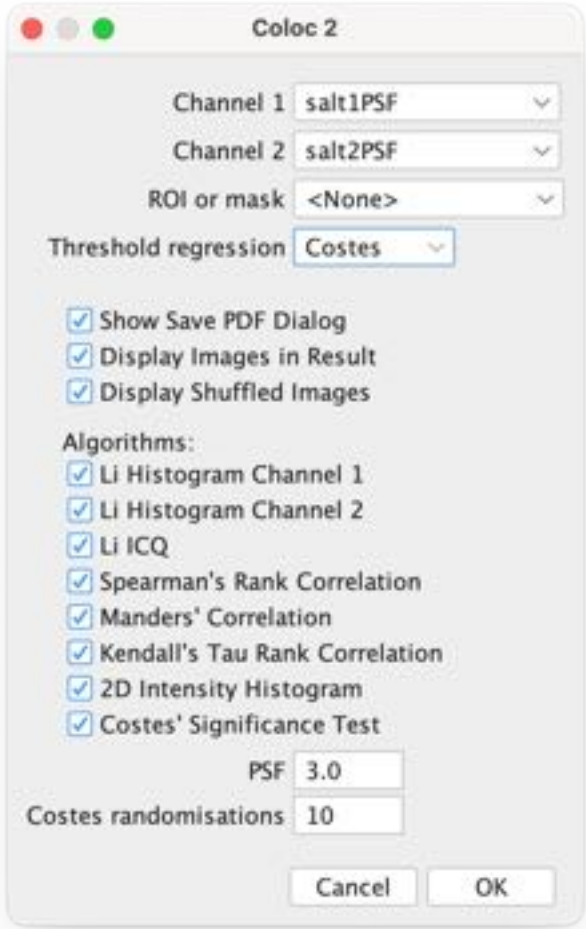


HARVARD MEDICAL SCHOOL

https://icy.bioimageanalysis.org/

Image Analysis Collaboratory

ImageJ/Fiji: Coloc 2 , DiAna, JACoP



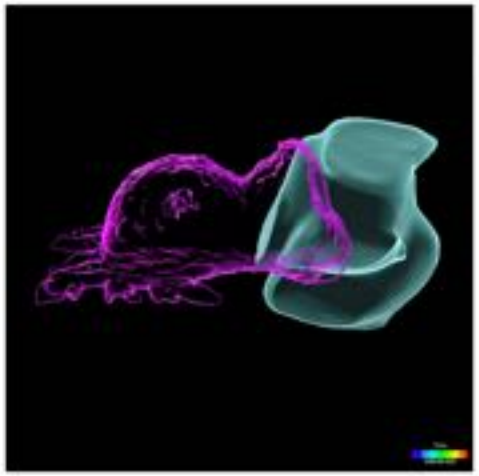
https://imagej.net/plugins/distance-analysis

https://imagej.net/plugins/coloc-2

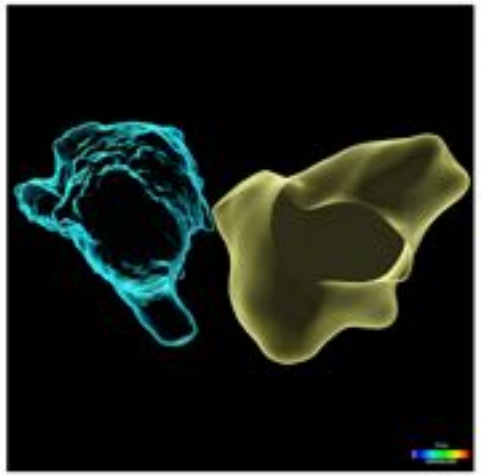
https://imagej.net/plugins/jacop

Imaris: Spatial Statistics

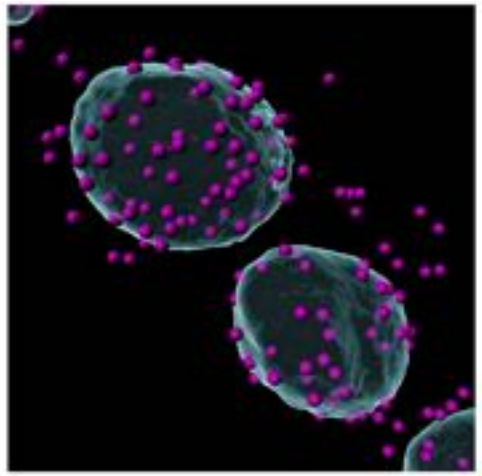
Volume Overlap

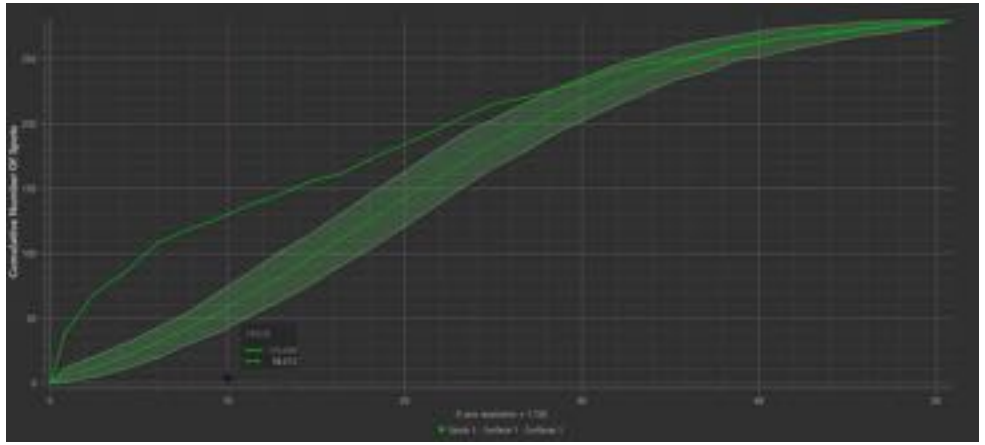


Distance/Proximity



Attraction to Surface





OXFORD INSTRUMENTS IMARIS

Expensive software. IAC has a full license

https://imaris.oxinst.com/

Final Words

"*The first principle is that you must not fool yourself and you are the easiest person to fool*."

— Richard P. Feynman

User-friendly software doesn't mean fool-proof results!

You will always get numbers, but what do they mean?

Hopefully you have a better idea now!

Further Learning (https://iac.hms.harvard.edu/resources/)

Forum: Knowledge exchange and support

- https://forum.image.sc/

Online book with code: Introduction to Bioimage Analysis

- https://bioimagebook.github.io/

Online training: NEUBIAS Academy

- https://www.youtube.com/c/NEUBIAS
  - Deconstructing co-localisation workflows: A journey into the black boxes
  - Introduction to 3D Analysis with 3D ImageJ Suite

Inference

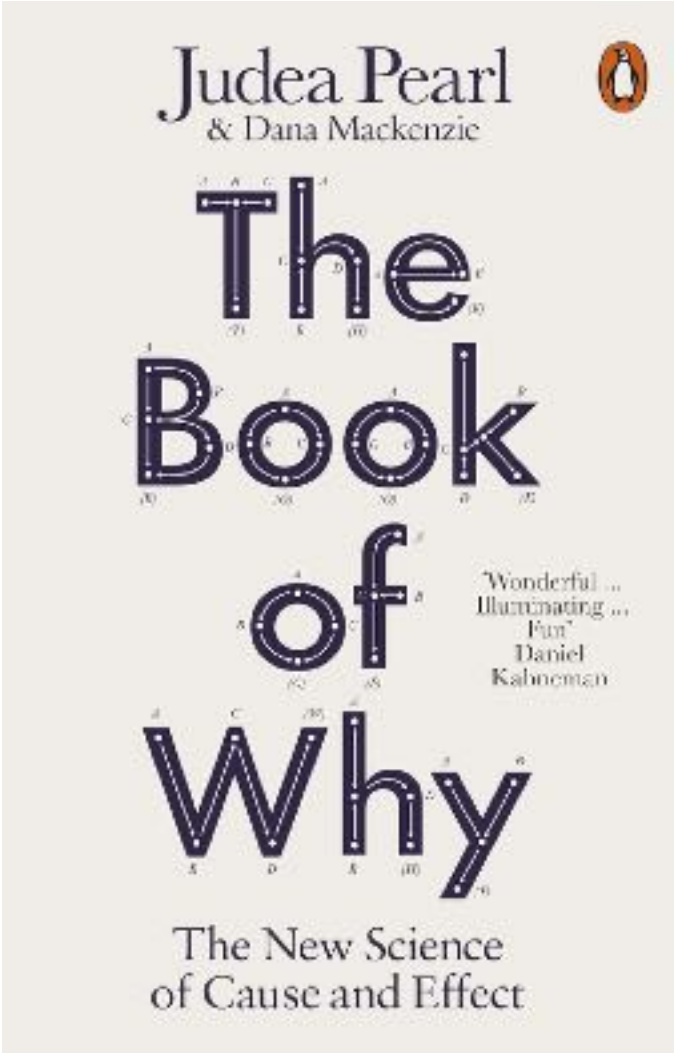
(plausible reasoning)

It is not Causal Analysis

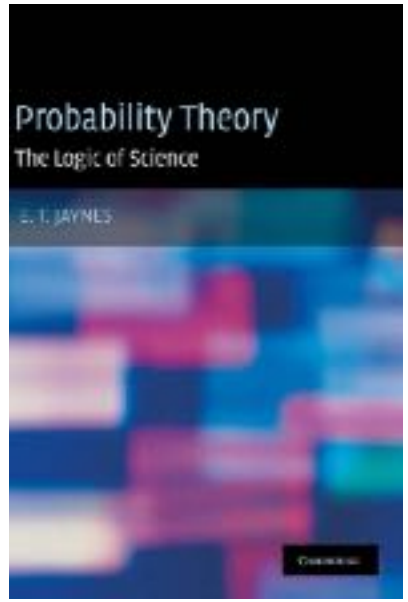
Doesn't address interactions directly

Cannot say if A causes B or vice versa

At best says how different from random the signal co-variation (intensity/location) is



Aristotelian Deductive Reasoning



Two strong syllogisms

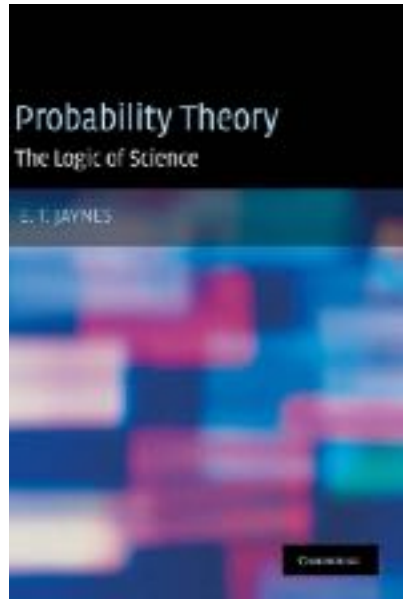
Major premise: if A is true, then B is true

Minor premise: A is true (B is false)

Conclusion: therefore, B is true (A is false)

Condensed form: $A \implies B \iff \bar{B} \implies \bar{A}$

Aristotelian Deductive Reasoning



Major: If it rains (A), the pavement will get wet (B)

Minor: It rains (A)

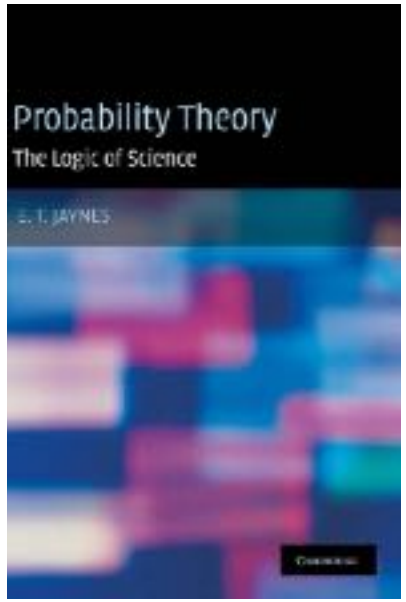
Conclusion: Therefore, the pavement will get wet (B)

Major: If it rains (A), the pavement will get wet (B)

Minor: The pavement is **not** wet ($\bar{B}$)

Conclusion: Therefore, it **didn't** rain ($\bar{A}$)

Plausible Reasoning



Two **weak** syllogisms

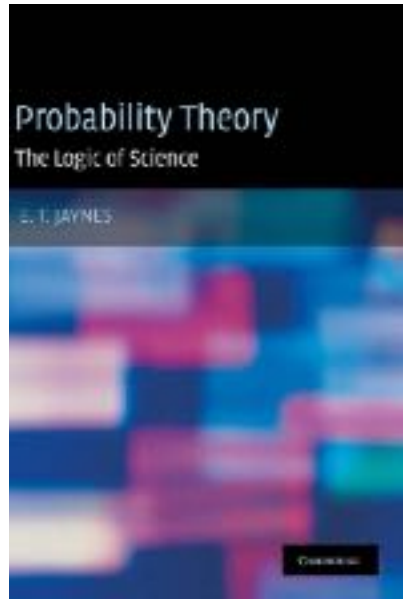
Major premise: if A is true, then B is true

Minor premise: B is true (A is false)

Conclusion: therefore, A becomes more plausible

(B becomes less plausible)

Plausible Reasoning



Major: If it rains (A), the pavement will get wet (B)

Minor: The pavement is wet (B)

Conclusion: Therefore, it is more plausible that it rained (A)

But it could have been a garden sprinkler and not rain

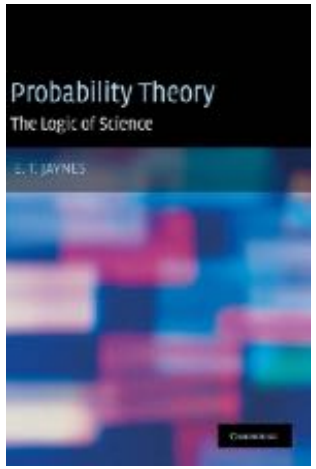
Major: If it rains (A), the pavement will get wet (B)

Minor: It **didn't** rain ($\bar{A}$)

Conclusion: Therefore, it is **less** plausible that the pavement will be wet (B)

But the pavement could have been made wet by other means

Plausible Reasoning



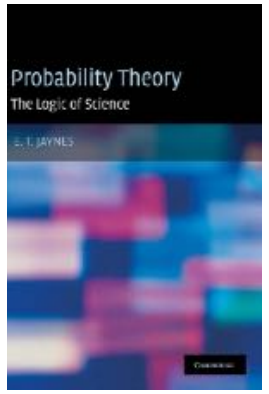
An even **weaker** syllogism

Major premise: If A is true, then B becomes more plausible

Minor premise: B is true

Conclusion: therefore, A becomes more plausible

Plausible Reasoning



Major: If it rains (A), the pavement is **more likely** to get wet (B)

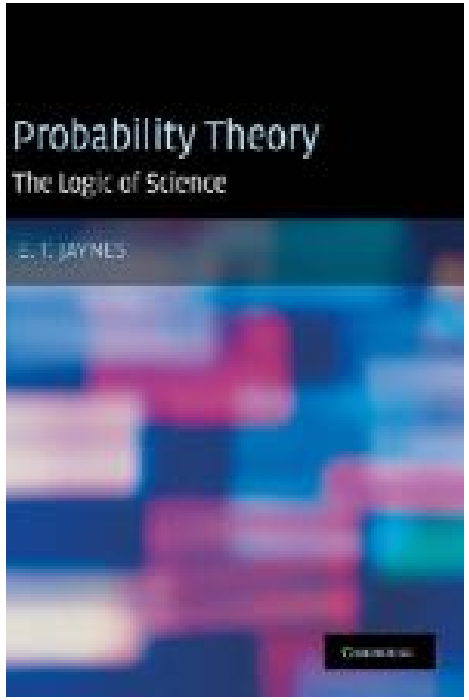
Minor: The pavement is wet (B)

Conclusion: Therefore, it is more plausible that it rained (A)

Maybe the pavement is covered by a tarp

Maybe a garden sprinkler was active nearby

Probability Theory



Jaynes derives two fundamental equations

From which follow Bayes' and just about all else

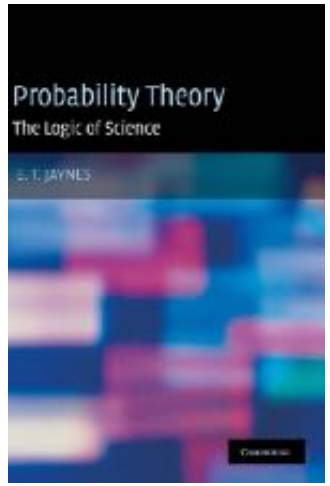
Product rule

$$p(AB \mid C) = p(A \mid C)p(B \mid AC) = p(B \mid C)p(A \mid BC)$$

Sum rule

$$p(A \mid B) + p(\bar{A} \mid B) = 1$$

Aristotelian Logic and the Product Rule



Let C stand for the **major** premise

$$C \equiv A \implies B$$

Then, note that, according to the **strong** syllogisms

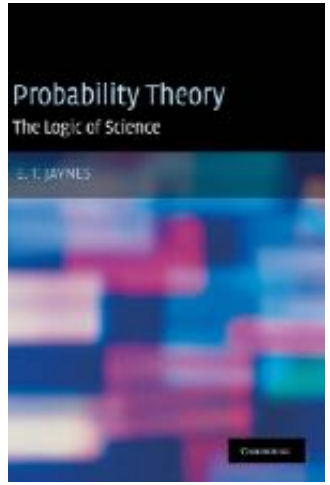
$$p(AB \mid C) = p(A \mid C) \text{ and that } p(A\bar{B} \mid C) = 0$$

Insert in product rule and get:

Strong syllogism 1: $$p(B \mid AC) = \frac{p(AB \mid C)}{p(A \mid C)} = 1$$

Strong syllogism 2: $$p(A \mid \bar{B}C) = \frac{p(A\bar{B} \mid C)}{p(\bar{B} \mid C)} = 0$$

Plausible Reasoning and the Product Rule



Let C stand for the **major** premise

$$C \equiv A \implies B$$

The product rule states

$$p(A \mid BC) = p(A \mid C)\frac{p(B \mid AC)}{p(B \mid C)}$$

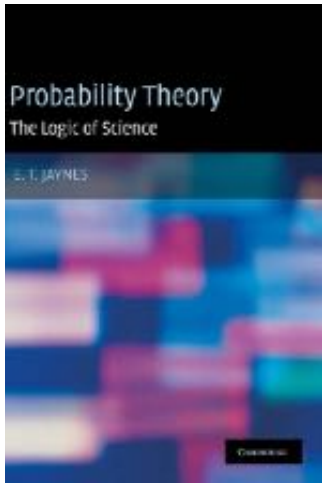
According to the first **weak** syllogism and overall definition of p

$$p(B \mid AC) = 1 \; (A \implies B) \text{ and } p(B \mid C) \leq 1 \text{ (generally)}$$

Insert in product rule and get:

Weak syllogism 1: $p(A \mid BC) \geq p(A \mid C)$

Probability Theory



From the these follow the very useful

Extended sum rule

$$p(A + B \mid C) = p(A \mid C) + p(B \mid C) - p(AB \mid C)$$

Directly applicable to colocalization questions