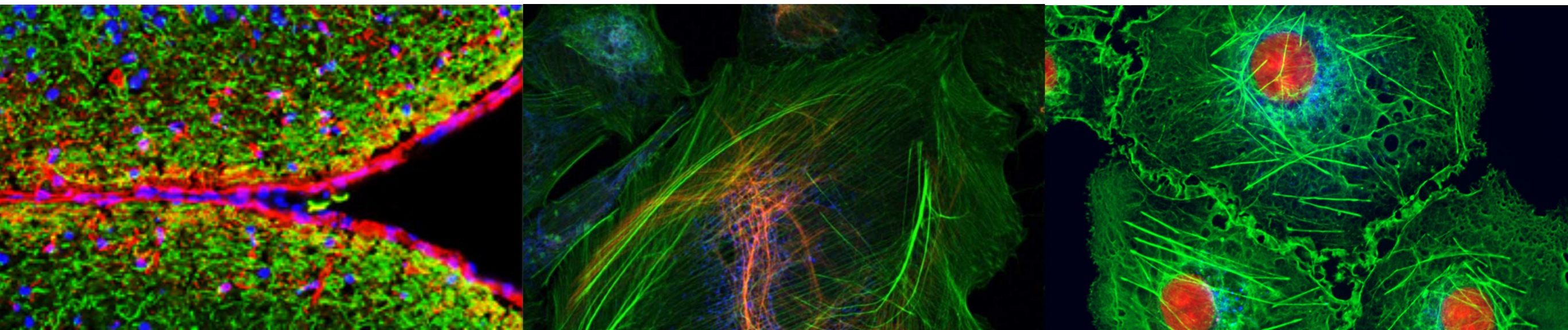


Non-parametric methods for point pattern analysis in microscopic images of protein abundance and other applications

Julia Brettschneider

Liverpool, December 1, 2021



Preamble:
Applied mathematics as bridge

*“The **instrument that mediates between theory and practice**, between thought and observation, is mathematics; it builds the **connecting bridge** and makes it stronger and stronger. Thus it happens that our entire present-day culture, insofar as it rests on intellectual insight into and harnessing of nature, is founded on mathematics.”*

David Hilbert

In Königsberg on 8 September 1930, David Hilbert addressed the yearly meeting of the Society of German Natural Scientists and Physicians (Gesellschaft der Deutschen Naturforscher und Ärzte).

Full text of the speech in English and German at url below, including audio file:

<http://math.sfsu.edu/smith/Documents/HilbertRadio/HilbertRadio.pdf>

About right and wrong

What are models for? Prediction and explanation. Answering questions...

*“Far **better an approximate answer to the right question**, which is often vague, than the exact answer to the wrong question, which can always be made precise.”*

John Tukey (1915-2000)

American statistician (FFT, various statistical tests, EDA)

How good is a model?

*“All models are wrong, some are **useful**.”*

George Box, FRS (1919-2013)

English statistician (quality control, time series, design of experiments, response surfaces, Bayesian inference etc)

Rephrase “how good”:

How good is it *at the task what you want it to do (prediction and/or explanation)?*

Outline of this talk

Microscopy

... quantifying spatial abundance of proteins

Point patterns in cellular structures

... probabilistic models for spatial point patterns

... statistical summaries and tests

... case study: microtubules formation

Dependencies between bulk movement patterns

... colocalisation

... earth movers distance and comparison statistics

... tests of spatial hypotheses

... simulations

... case study: relationship in the evolution of two proteins

Applications to digital detector QAQC

... spatial analysis of dead pixels

Microscopy to observe quantifying spatial abundance of proteins

Confocal fluorescent laser microscopy (live cells)

Electron microscopy (dead cells, higher resolution)

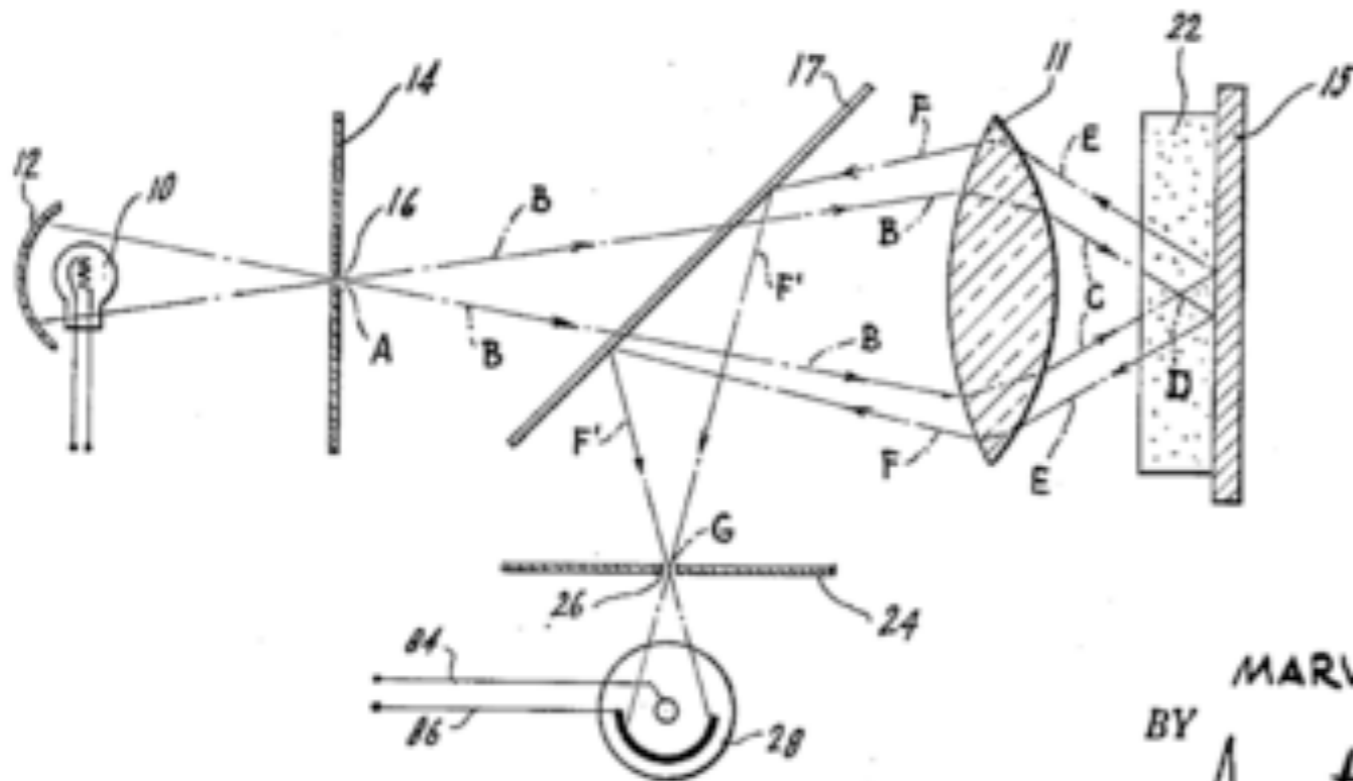
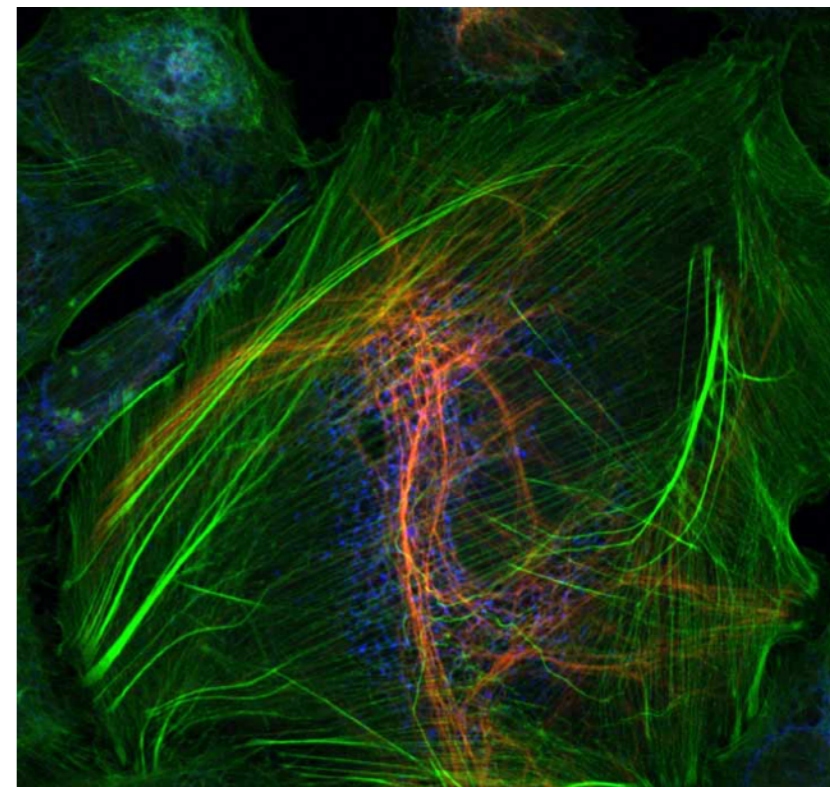
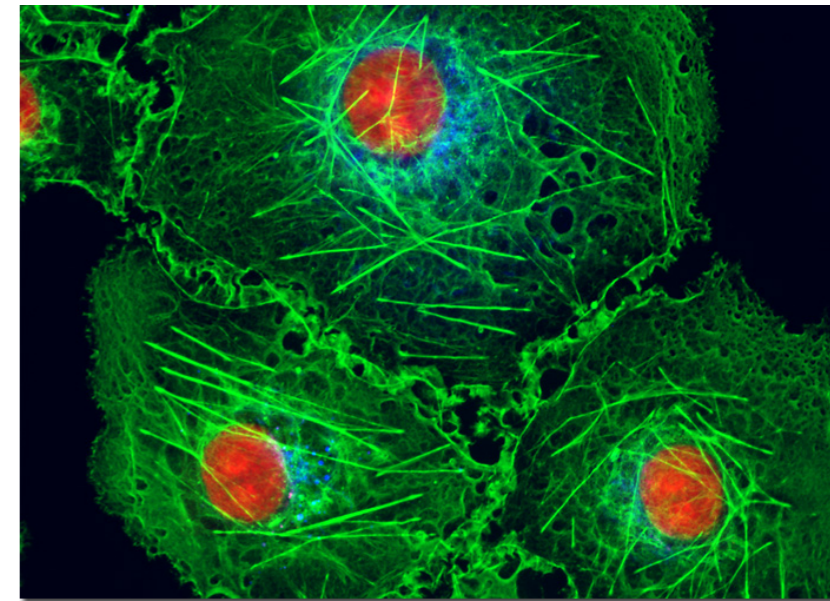
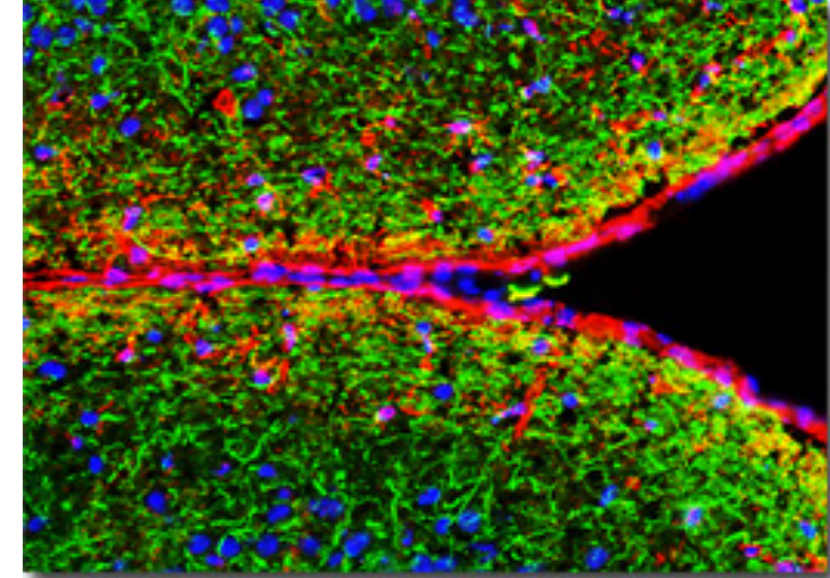


FIG. 3 .

INVENTOR.
MARVIN MINSKY
BY *Ameter & Levy*
ATTORNEYS

Confocal microscope

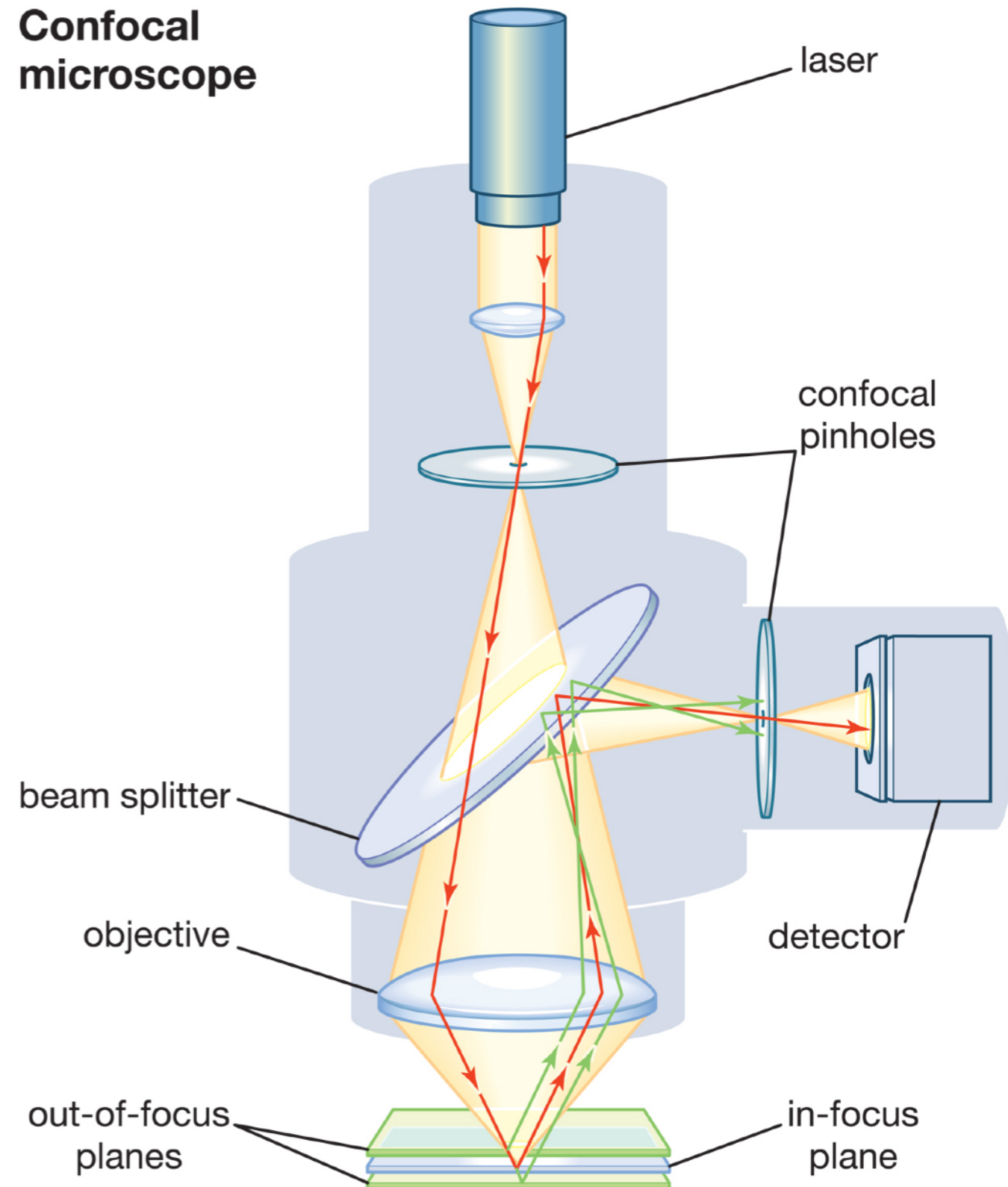
Wide-field microscopy:

- All of specimen excited at the same time
- Large unfocused background

Confocal microscope:

- Field of view limited by geometric optics
- Pinhole in optically conjugate plane in front of the detector to eliminate out-of-focus signal
- Long exposure required
- Scanning arrangement to build up image of larger region
- Better resolution

Confocal microscope



© 2012 Encyclopædia Britannica, Inc.

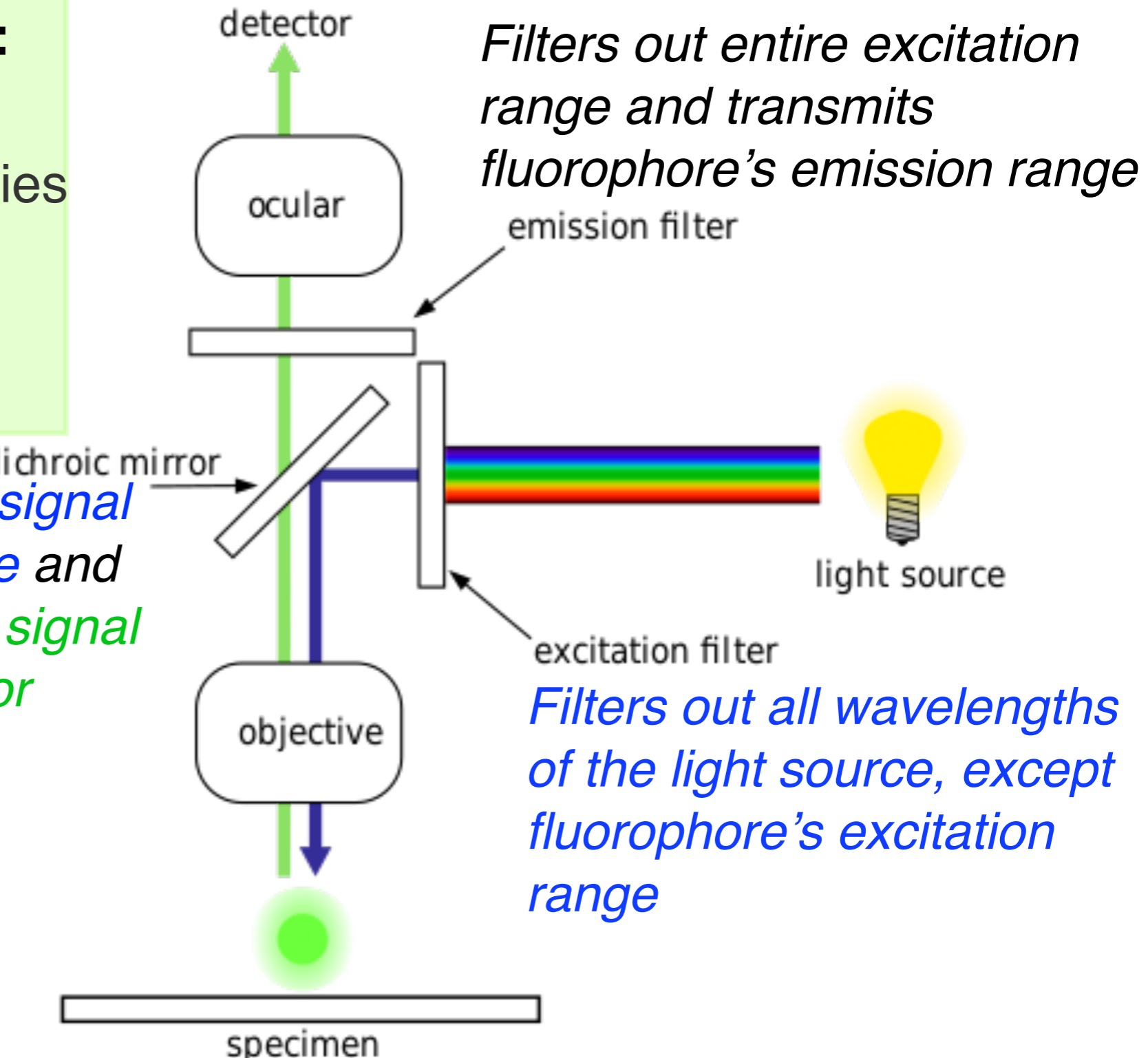
<https://www.britannica.com/technology/microscope/Confocal-microscopes>

Fluorescent microscope

Fluorescent microscope:

- high intensity light source
- excites a fluorescent species in a sample
- Sample emits different wavelength

Reflects excitation signal towards fluorophore and transmits emission signal towards the detector

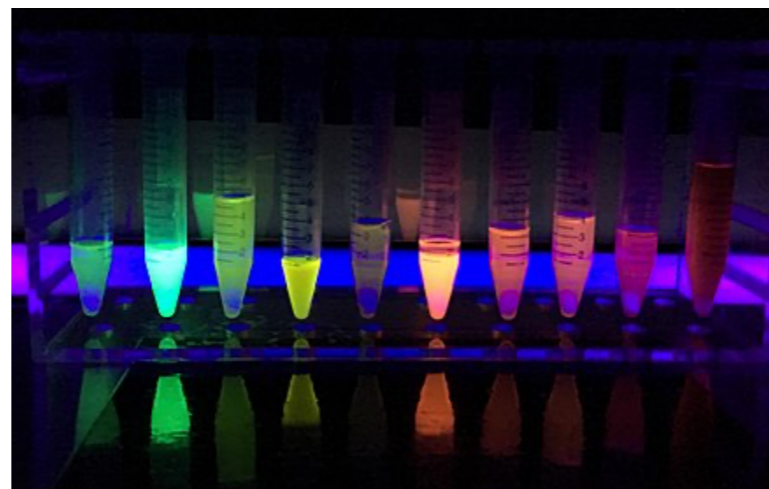
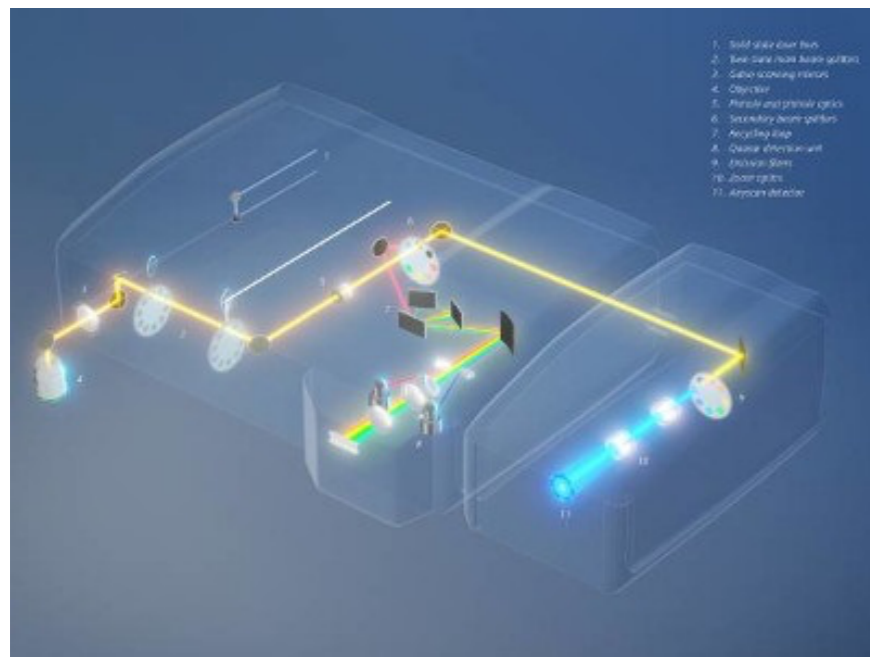


Labelled with fluorescent protein (e.g. GFP)

Confocal fluorescent laser microscope

Fluorescent confocal microscope:

- Combination of two ideas in microscopy technology
- High resolution images
- Life cells
- 2D or 3D through scanning schemes
- Multi-channel through use of range of fluorescent proteins

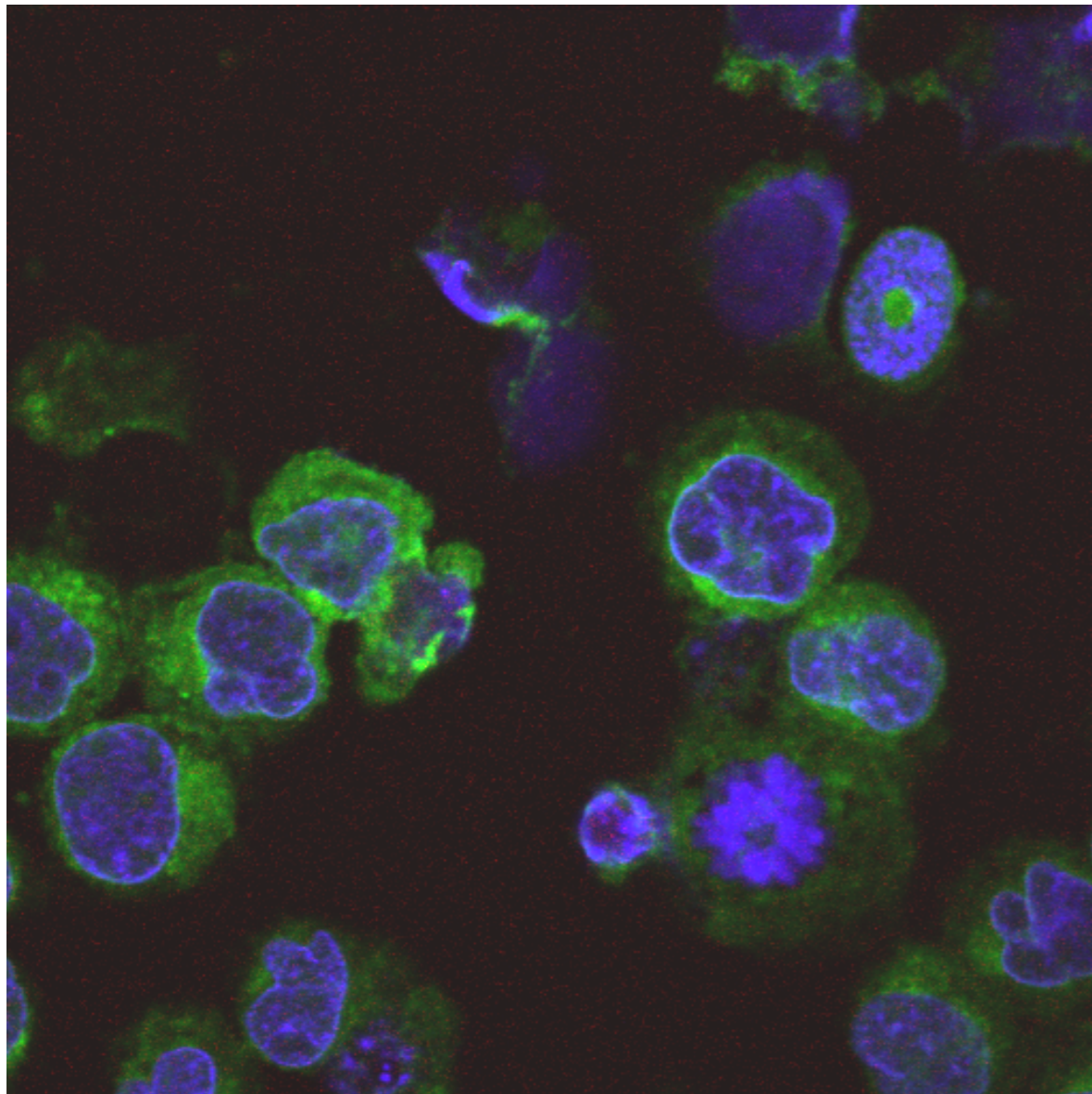


https://www.biocompare.com/25608-Microscopes-and-Cell-Imaging-Systems/14617250-ZEISS-LSM-980-Confocal-Laser-Scanning-Microscope/?pda=25608|14617250_0_1|2254289,2254327|1|&dfp=true

https://en.wikipedia.org/wiki/Green_fluorescent_protein#/media/File:Fluorescence_from_Fluorescent_Proteins.jpg

Example: Quantifying protein abundance in their actual locations in cells

Sub-cellular localisation of tumour antigen SSX2IP in leukemia cells



Green: SSX2IP expression visualised by anti-SSX2IP-fluorescein isothiocyanate on the cell's surface.

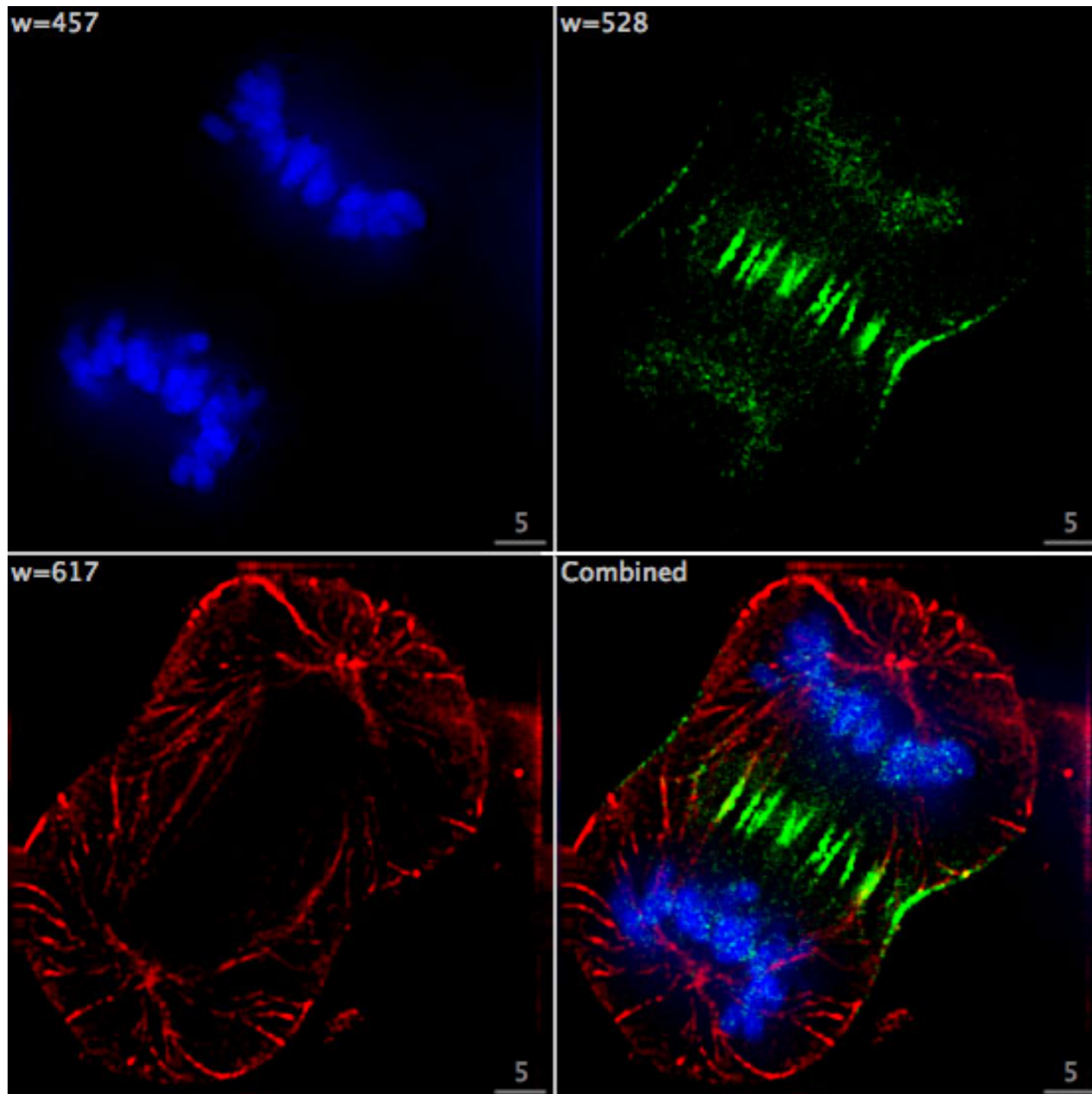
Blue: Stained Cell nuclei using 4,6'-diamino-2-phenylindole (DAPI).

Protocol of the experiment:

Leukaemia cell line K562 air dried for 4-18hours onto glass microscope slides, stored at -20°C wrapped in saranwrap, defrosted, stained with antigen specific primary, and fluorescently labelled secondary antibodies.

Example:

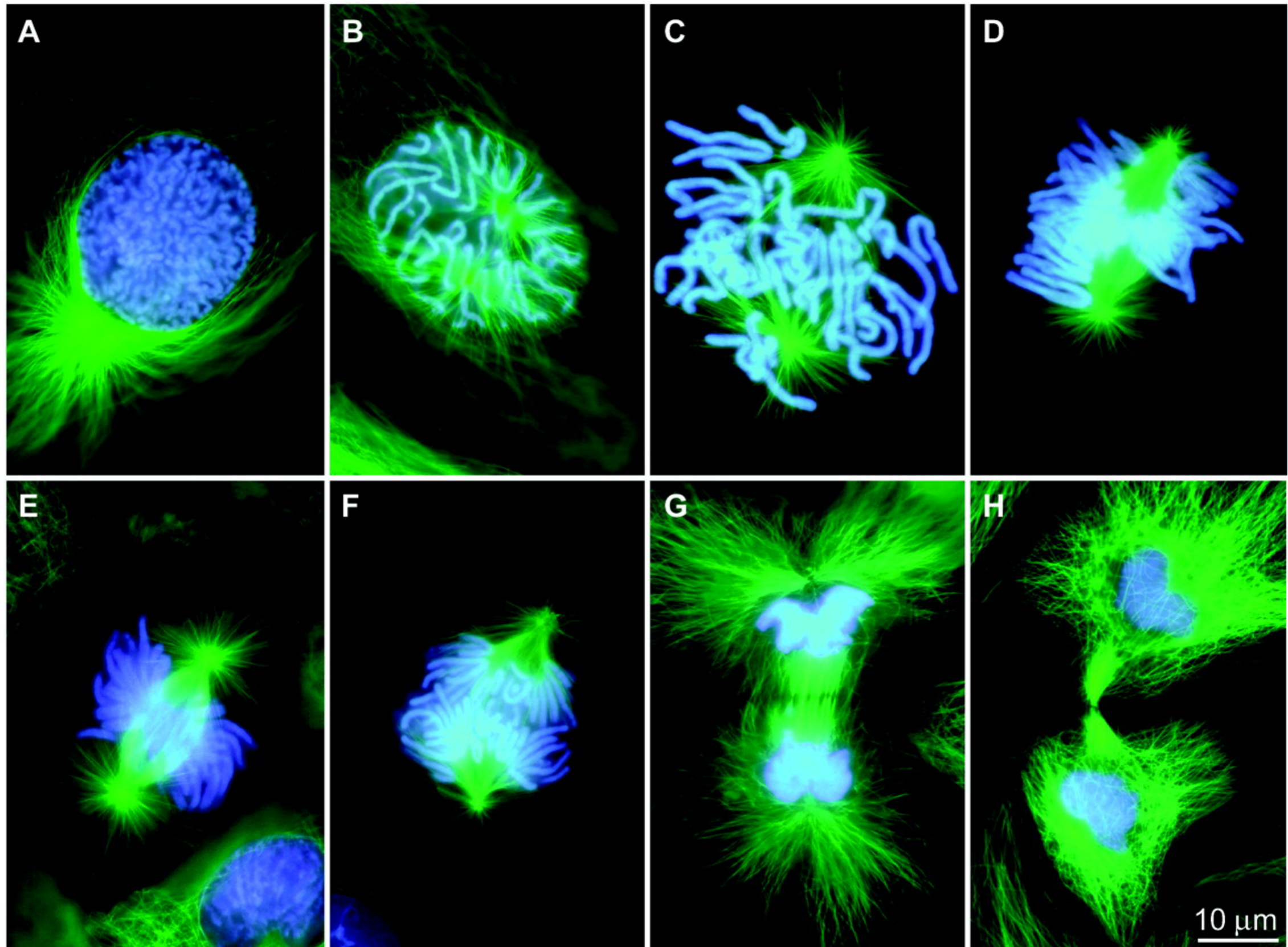
3 components in dividing human cancer cells



Scanning scheme for fluorescent imaging:

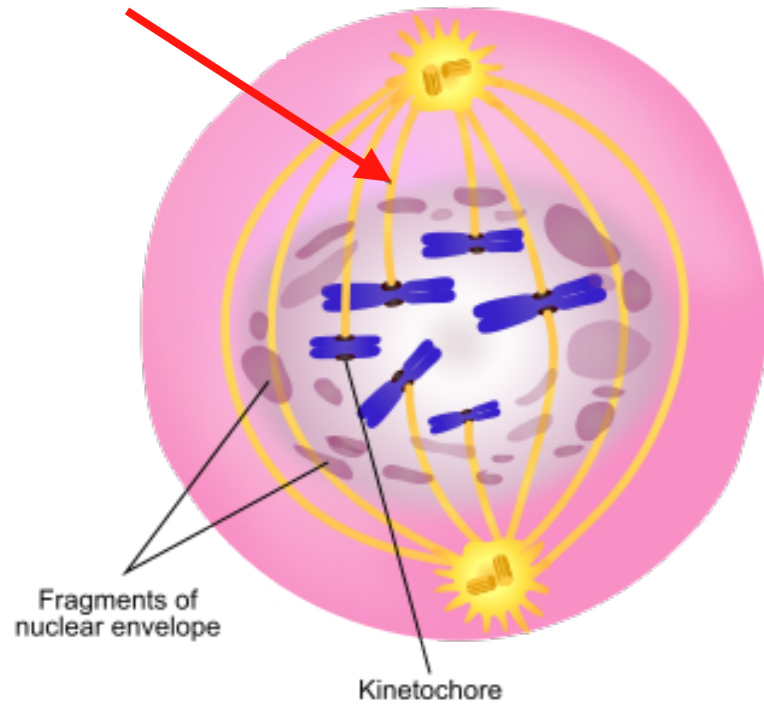
- Blue: Chromosomes (DNA)
- Green: INCENP (protein)
- Red: microtubules
- Fluorophores imaged separately using different excitation and emission filters
- Images captured sequentially
- Overlaid

Study I: Microtubules formation



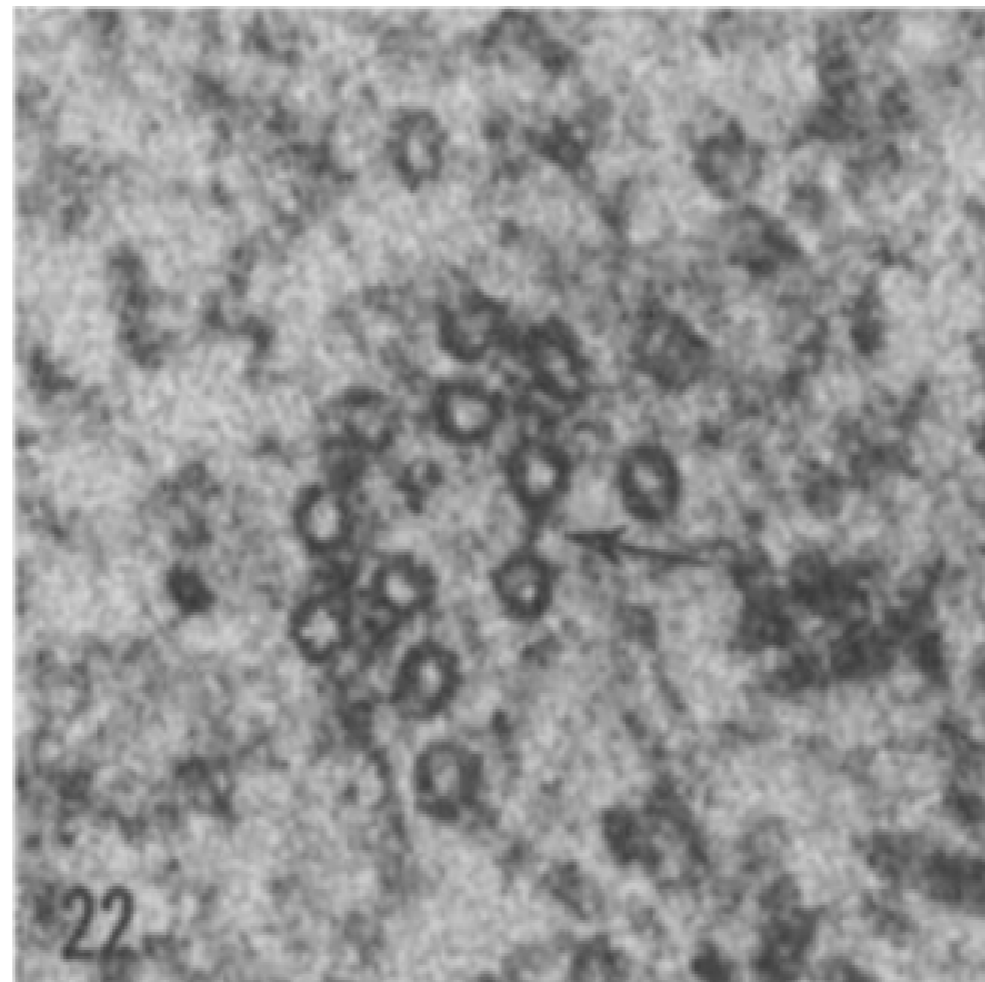
Microtubules during cell division (mitosis)

Microtubule

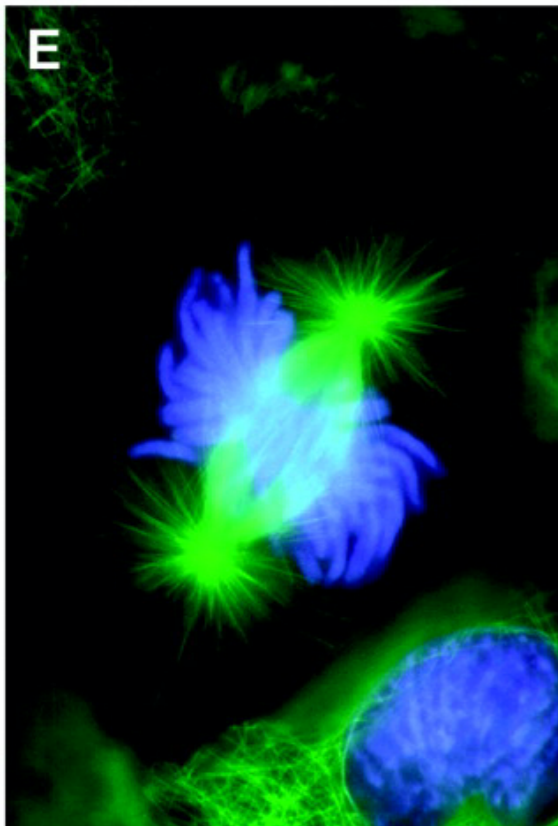
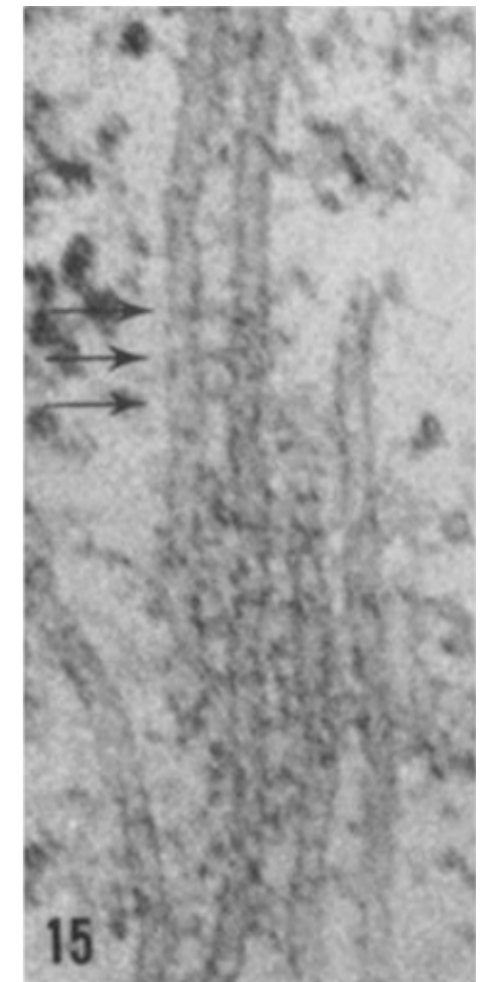


- Centrosomes = centrioles + microtubules
- Centrioles help the spindle into proper formation
- Spindle microtubules are arranged in K-fibers
- Intertubule bridges formed by mesh

Perpendicular to the microtubule axis



Parallel



Study I: Microtubules locations as point patterns

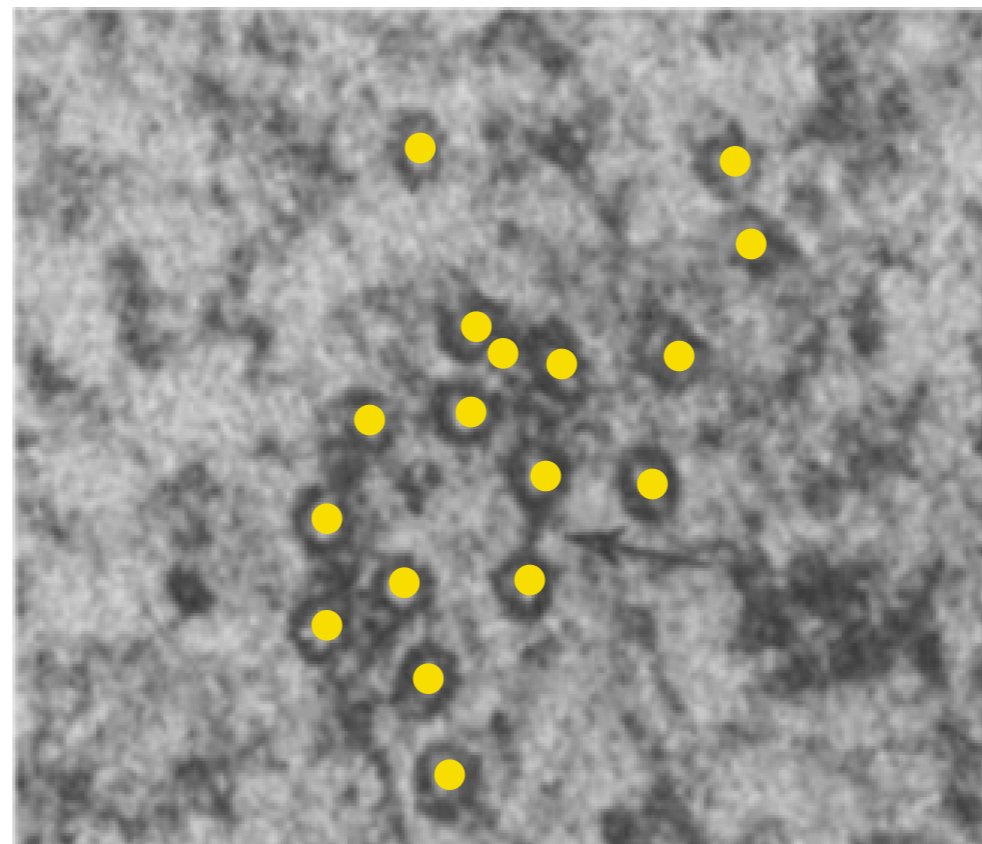
Stephen Royle's Lab (Centre for Mechanochemical Cell Biology) asks:
What is the role of TACC3 protein for the structure of microtubules within K-fibres and mesh?

Experiment: Overexpression of TACC3 through treatment versus control.

Data: Microscopic images collected in planes perpendicular to the fibre axes.

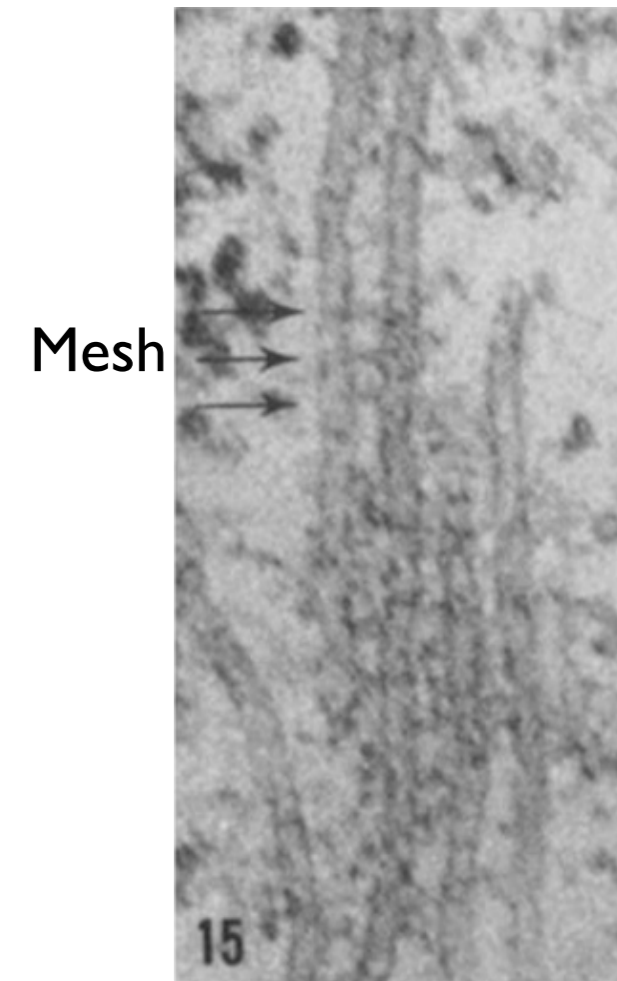


Perpendicular view



Model: locations as point pattern

Parallel



15

Describing and comparing protein abundance

Data:

Point patterns \underline{x}^I , $I = I_0 \cup I_1$.

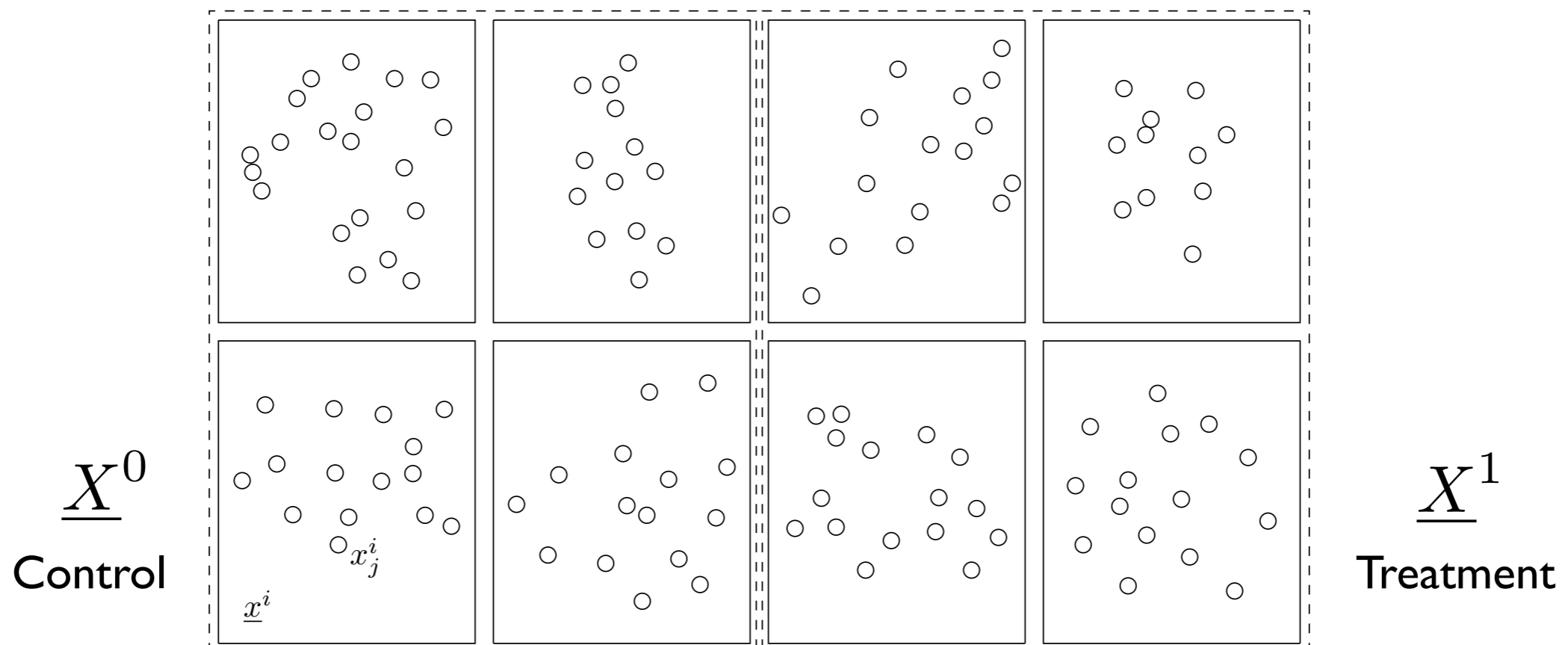
Model:

Point patterns \underline{x}^{I_0} independent realisations of point process \underline{X}^0 .

Point patterns \underline{x}^{I_1} independent realisations of point process \underline{X}^1 .

Task:

Inference on existence and form of a difference between \underline{X}^0 and \underline{X}^1 .



Point patterns models

Set of point patterns:

$$\chi_2 := \{(\underline{x} = x_1, x_2, \dots, x_{n(\underline{x})}) : n(\underline{x}) \in \mathbb{N}, x_i \in \mathbb{R}^2 \text{ for } i = 1, 2, \dots, n\}$$

Model pattern as realisations of a point process:

Random subset \underline{X} on \mathbb{R}^2 .

For B in Borel σ -algebra $\mathcal{B}(\mathbb{R}^2)$ on \mathbb{R}^2 : $\underline{X}_B = \underline{X} \cap B$

Counts (random variable): $N(B) = n(\underline{X}_B) =$ number of points of \underline{X} in B

Intensity measure μ

$$\mu(B) = \mathbb{E}[N(B)], \quad \forall B \in \mathcal{B}(\mathbb{R}^d).$$

If for some function $\rho : \mathbb{R}^2 \rightarrow [0, \infty)$

$$\mu(B) = \int_{x \in B} \rho(x) dx, \quad \forall B \in \mathcal{B}(\mathbb{R}^d),$$

then ρ is referred to as the intensity function of \underline{X} .

Summary statistics: basics

Let \underline{x} be a realisation of \underline{X} on the observation window W .

Estimator for the **intensity** of \underline{X} :

$$\hat{\rho} = \frac{n(\underline{x})}{|W|}$$

Let $\text{nn}(x_j)$ be the (set of) **nearest neighbours** of point x_j .

$$\text{nn}(x_j) = \{x_k : k = \operatorname{argmin}_l \|x_l - x_j\|\},$$

and $\text{nnd}(x_j)$ its **nearest neighbour distance**

$$\text{nnd}(x_j) = \inf_{x \in \text{nn}(x_j)} \{\|x_j - x\|\}.$$

Estimator for the **mean nearest neighbour distance** for \underline{X} :

$$\overline{\text{nnd}}(\underline{x}) = \frac{1}{n(\underline{x})} \sum_{j=1}^{n(\underline{x})} \text{nnd}(x_j)$$

Summary statistics: K-function

K-function (Ripley 1977) (scaled neighbourhood count function):

$$K(r) = \frac{1}{\rho} \mathbb{E} \left[\frac{1}{N(S)} \sum_{x_j \neq x_k \in \underline{X}} 1_{\{\|x_j - x_k\| < r\}} \right]$$

Estimate:

$$\hat{K}(\underline{x}, r) = \frac{|W|}{n(\underline{x})^2} \sum_{j \neq k} e_{j,k} 1_{\{\|x_j - x_k\| \leq r\}}$$

where $e_{j,k}$ is the proportion of the circumference of the circle with centre x_j and radius $\|x_j - x_k\|$ in W (edge correction).

$K(r) = \pi r^2$: CSR (complete spatial randomness)

$K(r) > \pi r^2$: aggregation at distances less than r

$K(r) < \pi r^2$: repulsion at distances less than r

Summary statistics: G-function

Nearest neighbour function (Diggle 2003):

$$G(r) = \frac{1}{\rho|B|} \mathbb{E} \left[\sum_{x \in \underline{X}_B} 1_{\{\underline{X} \setminus x \cap b(x,r) \neq \emptyset\}} \right]$$

for finite B in \mathbb{R}^2 , and $b(x, r)$ the disc centred at x with radius r .

(For stationary \underline{X} it is independent of B .)

Distribution of distance of randomly selected point to its nearest neighbour.

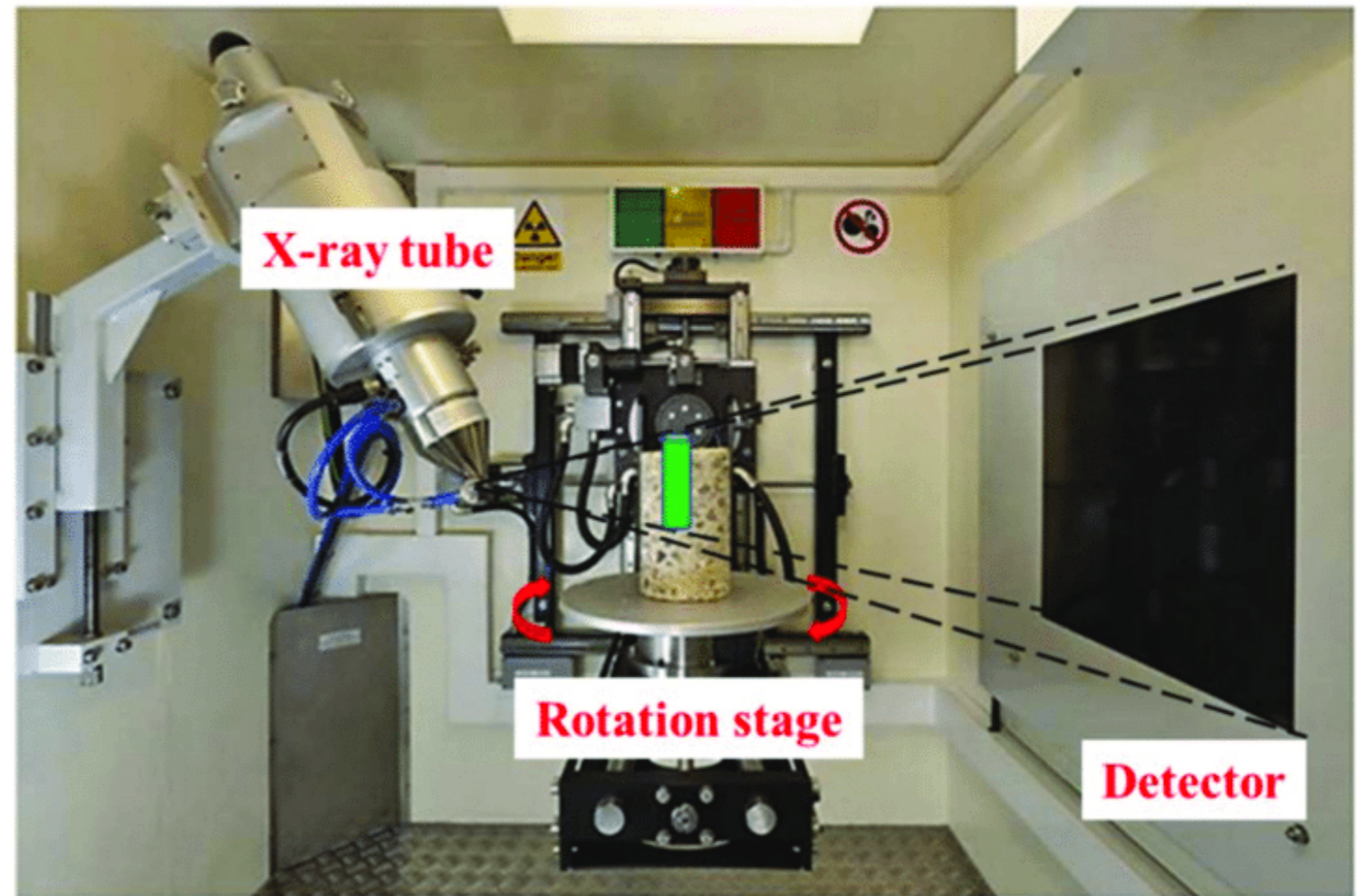
Estimate:

$$\hat{G}(\underline{x}, r) = \frac{1}{n(\underline{x})} \sum_{j=1}^{n(\underline{x})} 1_{\{\text{nnd}(x_j) \leq r\}}$$

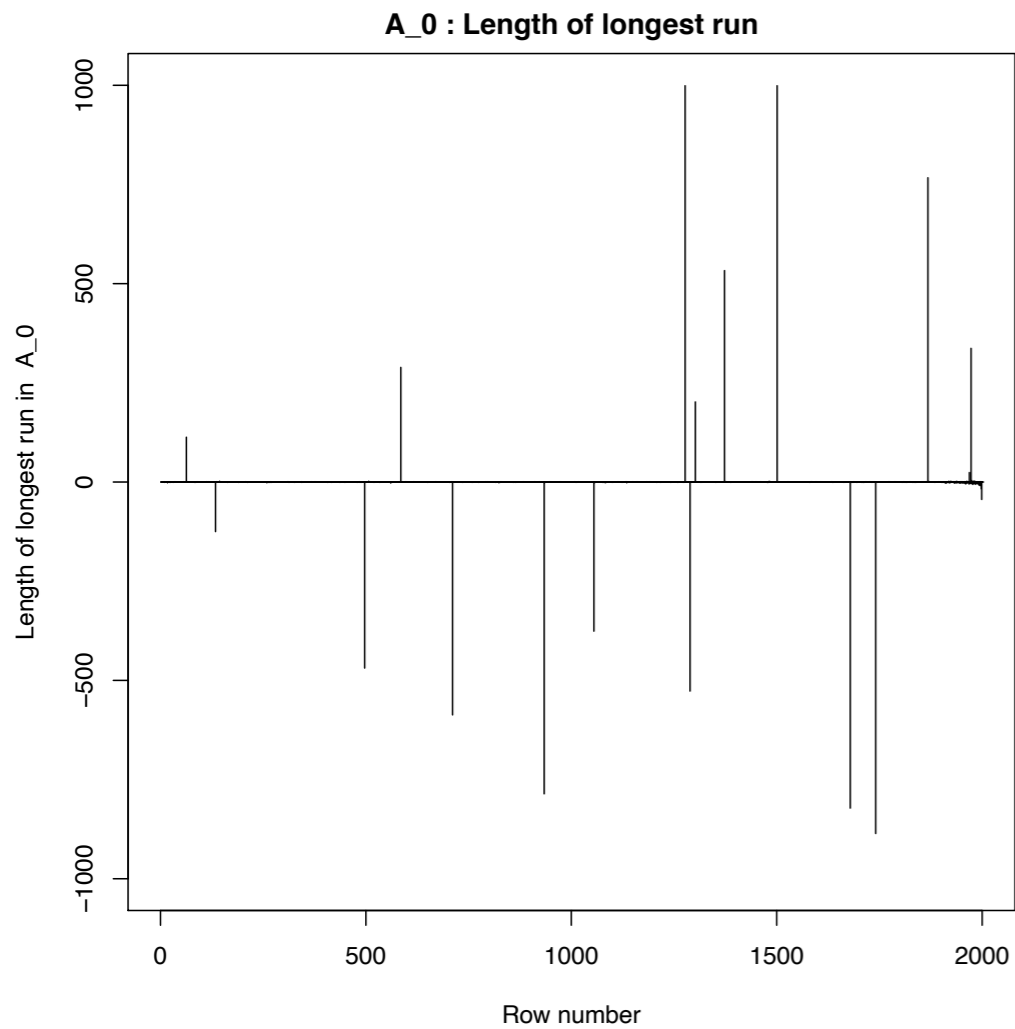
If \underline{X} is completely spatially at random then $G(r) = 1 - \exp(-\rho\pi r^2)$

Excursion: Point process models of dead pixels

Modern computed tomography uses digital detectors



Local defects

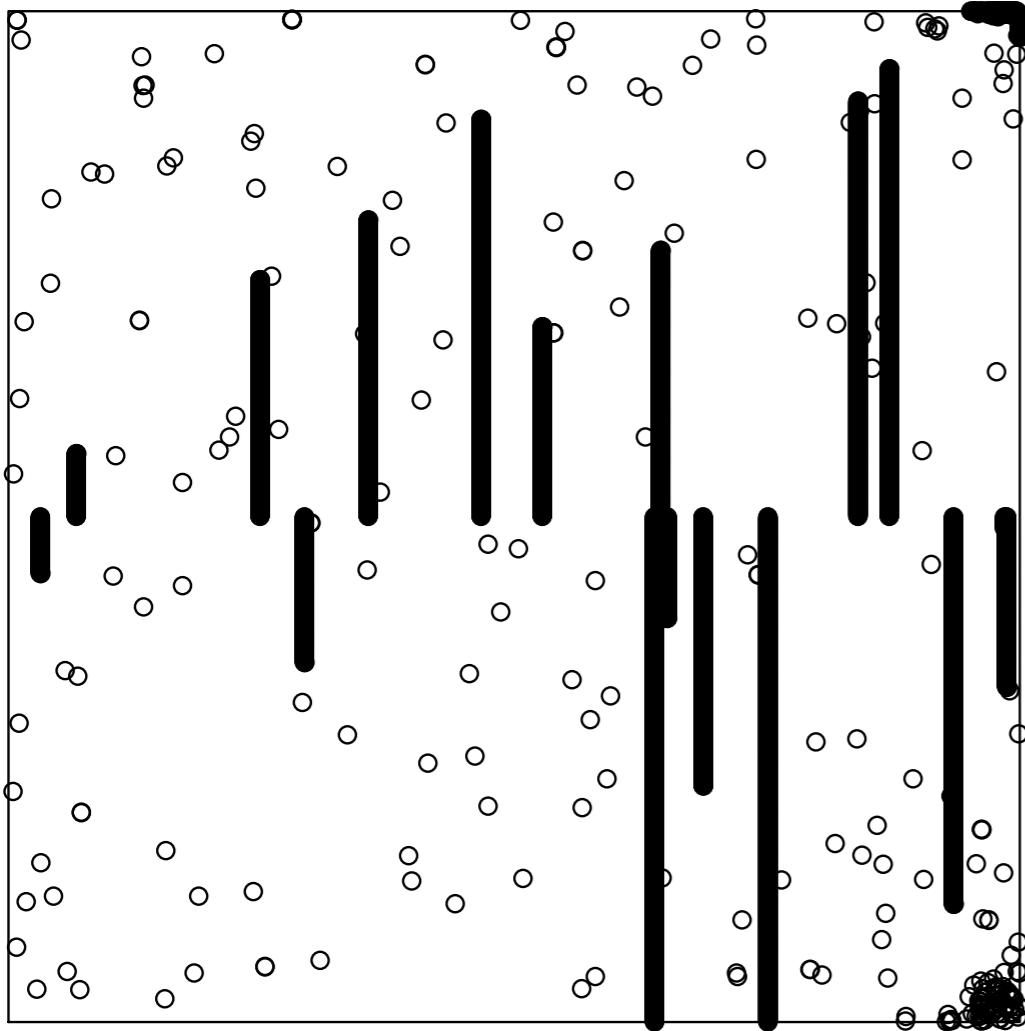


A_0: Graph of
bad pixel images

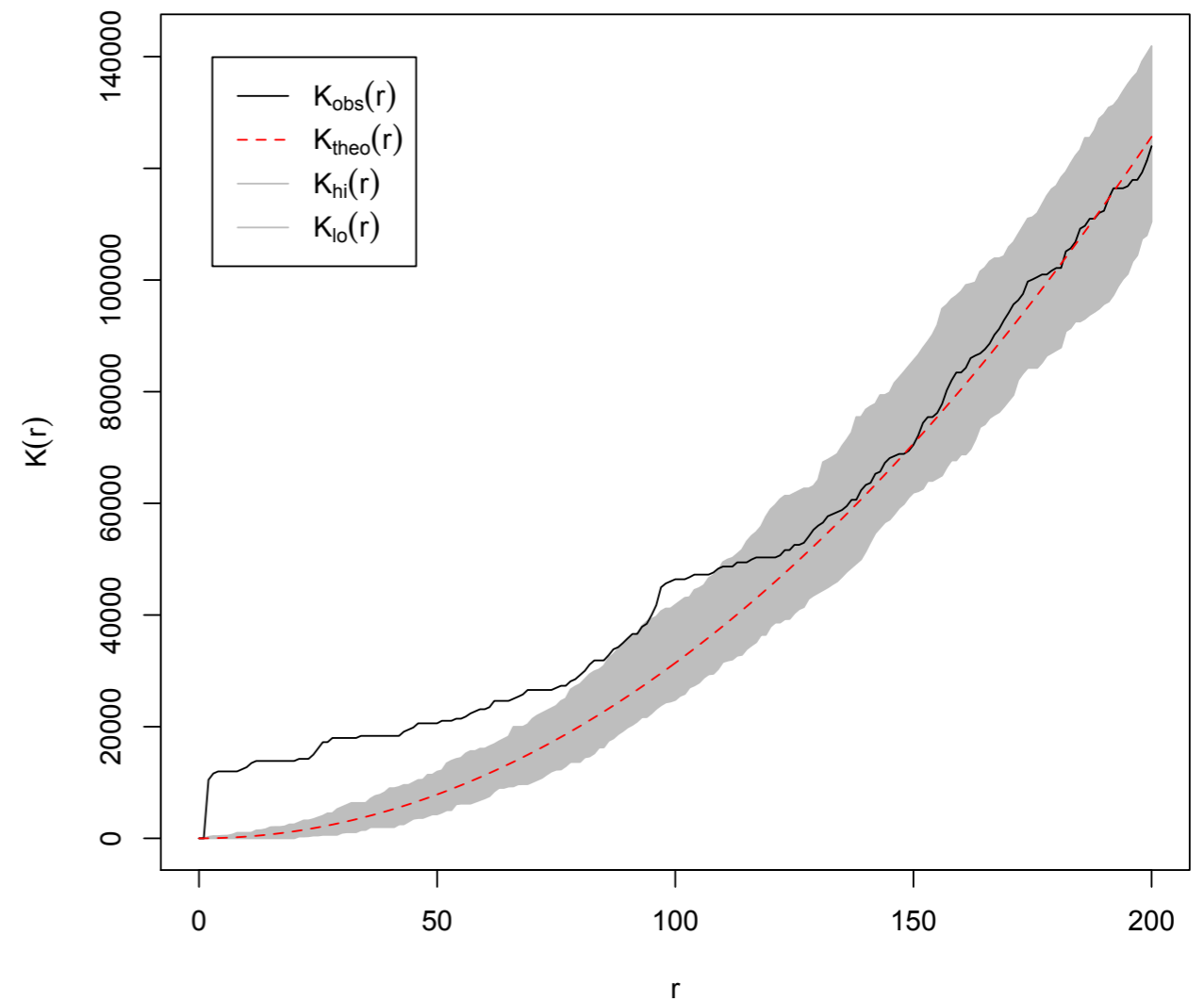
A_0: Bad pixel image

Point pattern and K-function

Point pattern A_0

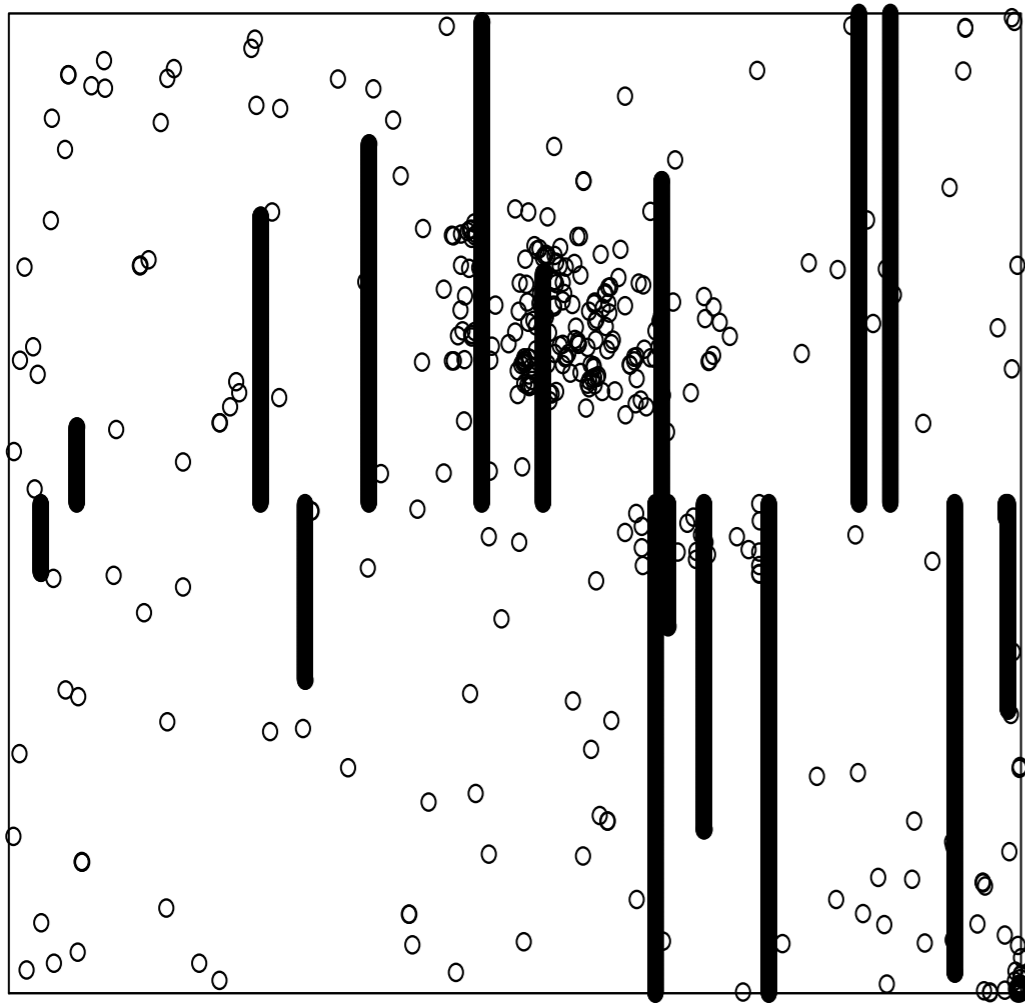


K function A_0

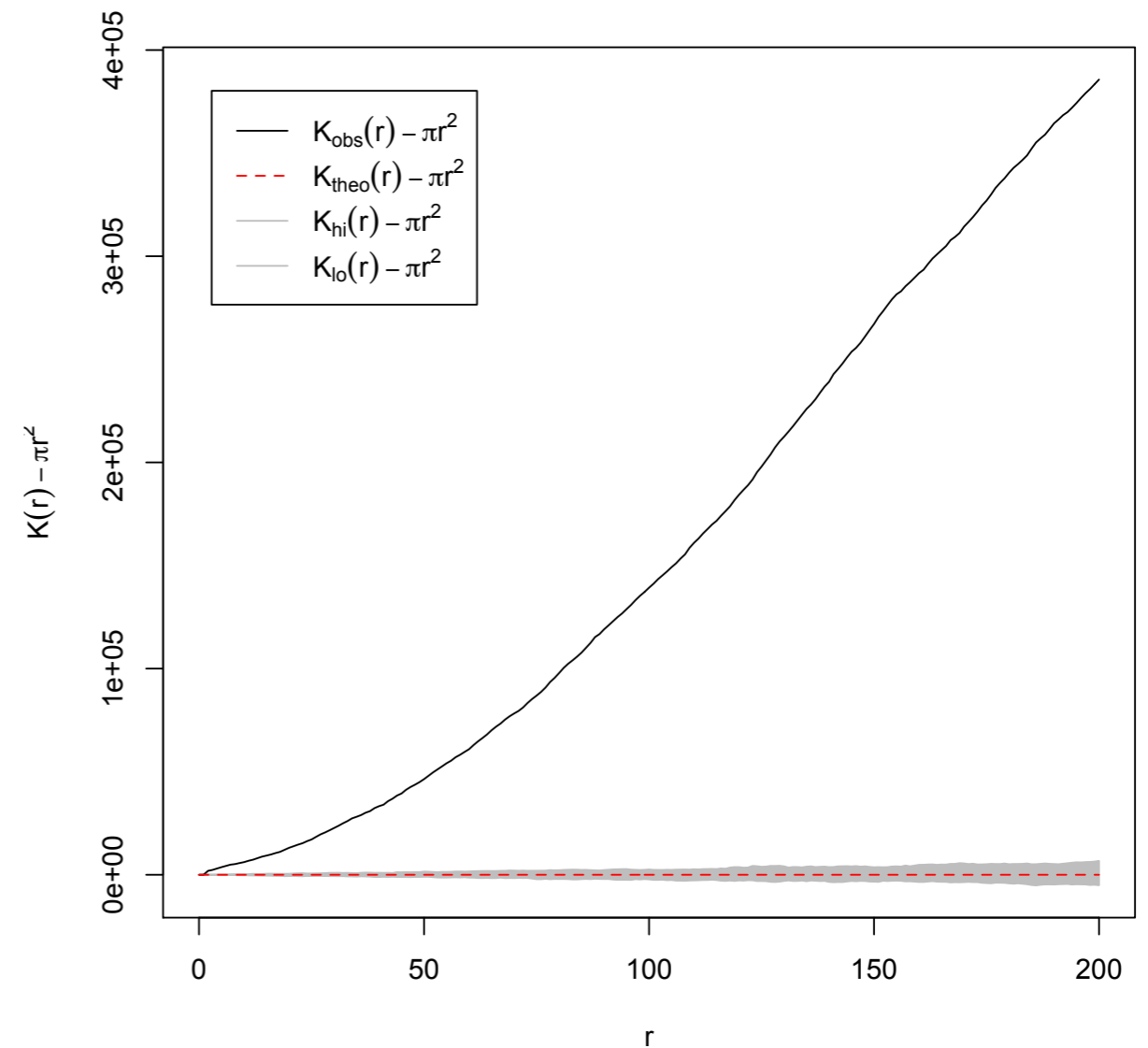


Point pattern and K-function

Point pattern E_0



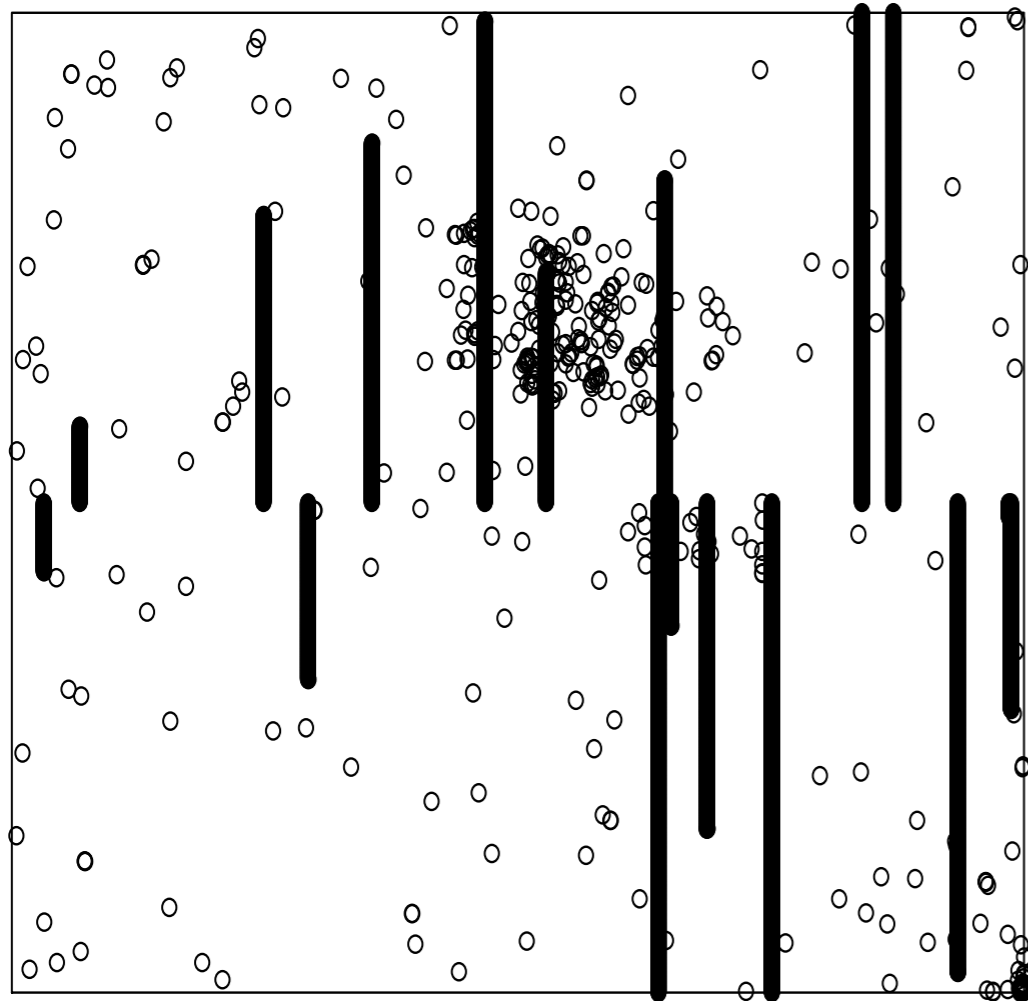
K function normed E_0



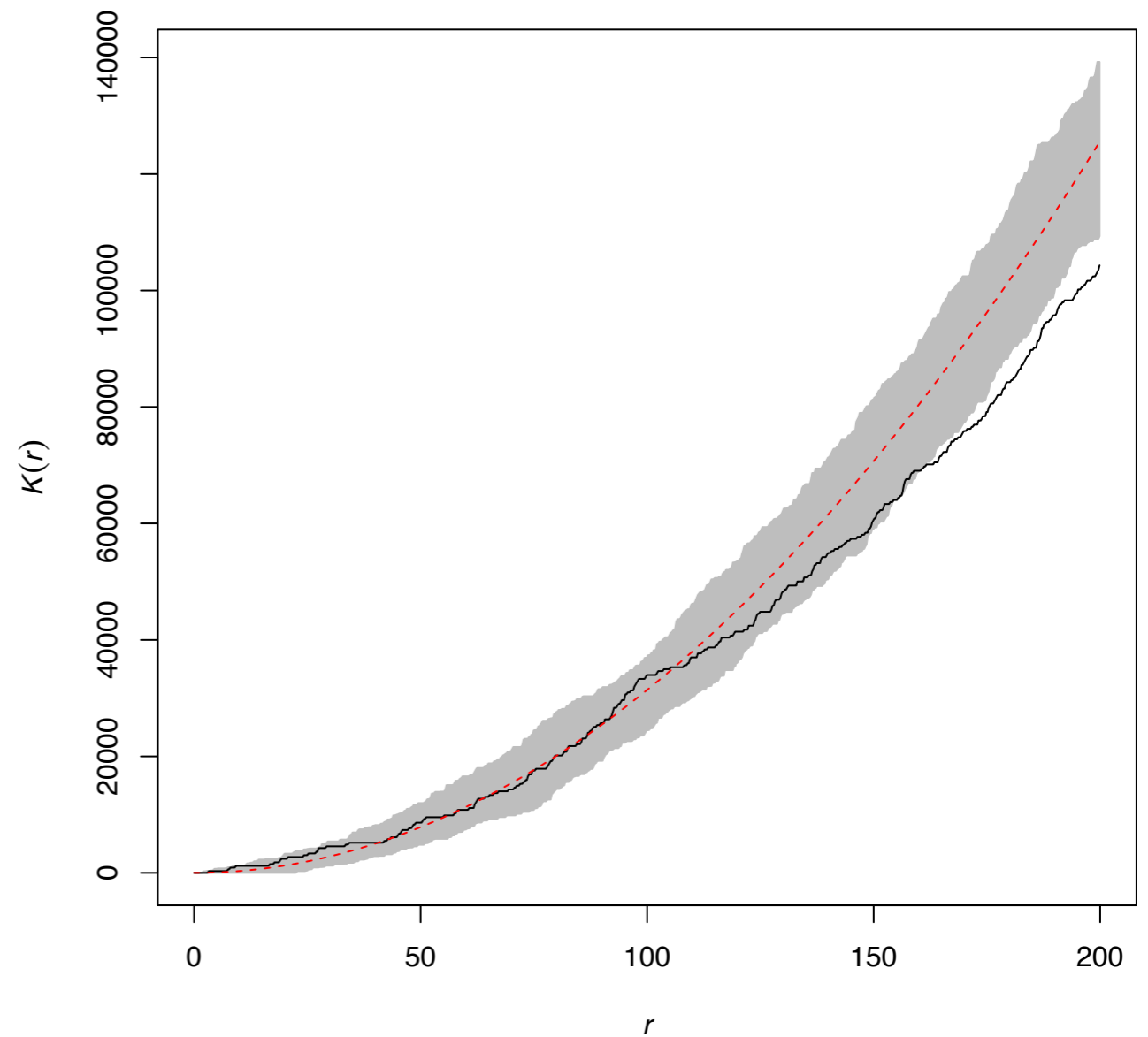
After modification: K-function

Excluding lines and area with high density

Point pattern E_0

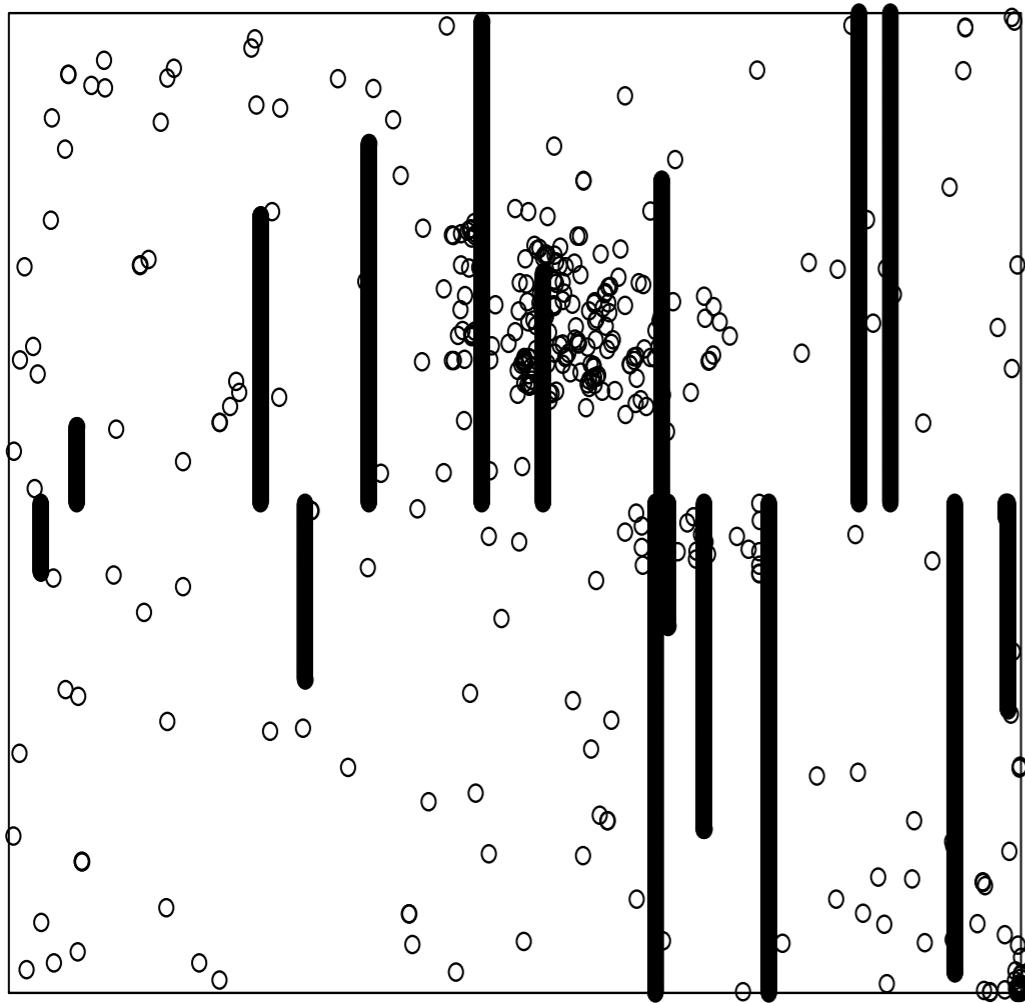


K-function, Events, nsim=100

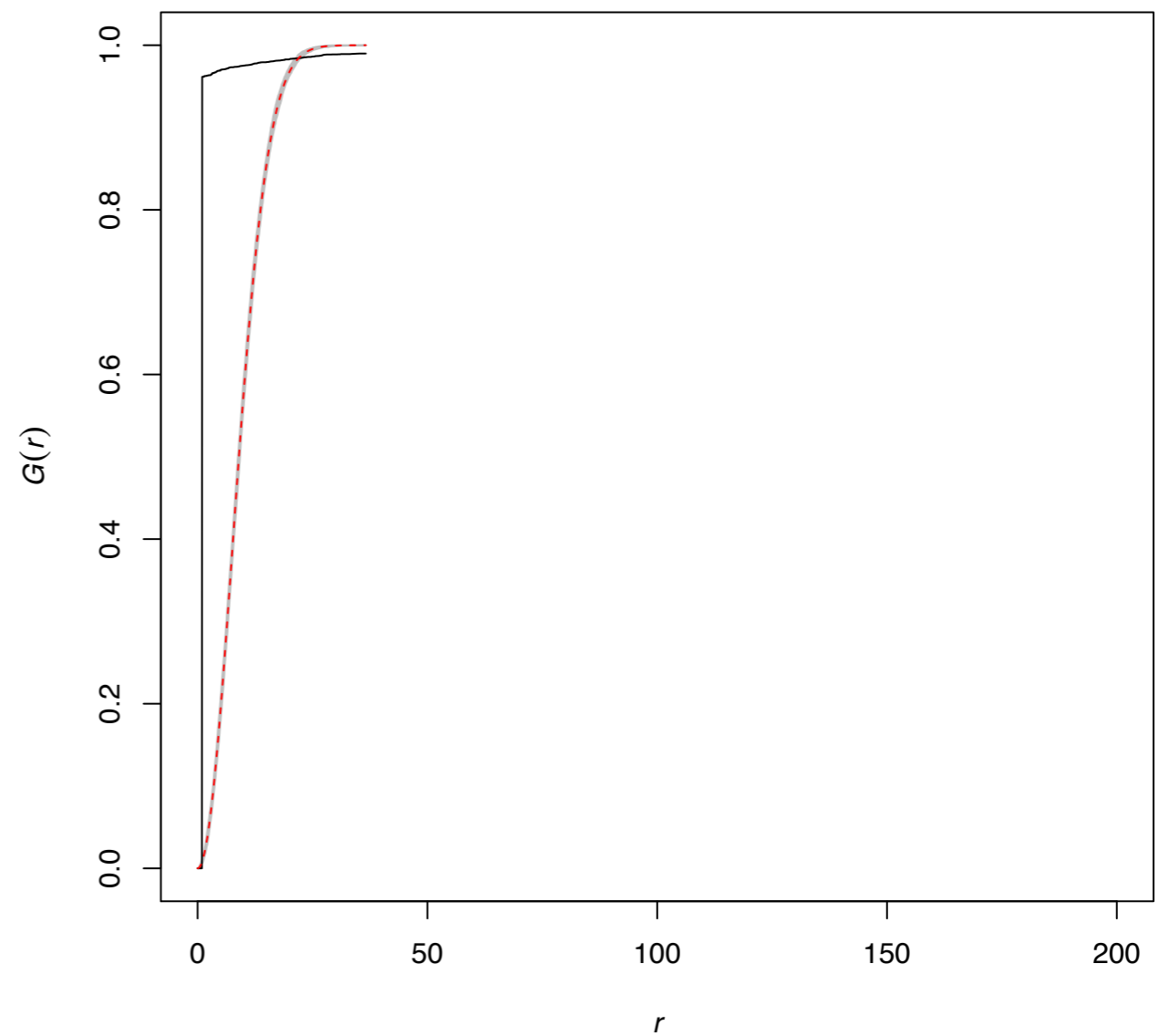


Before modification: G-function

Point pattern E_0



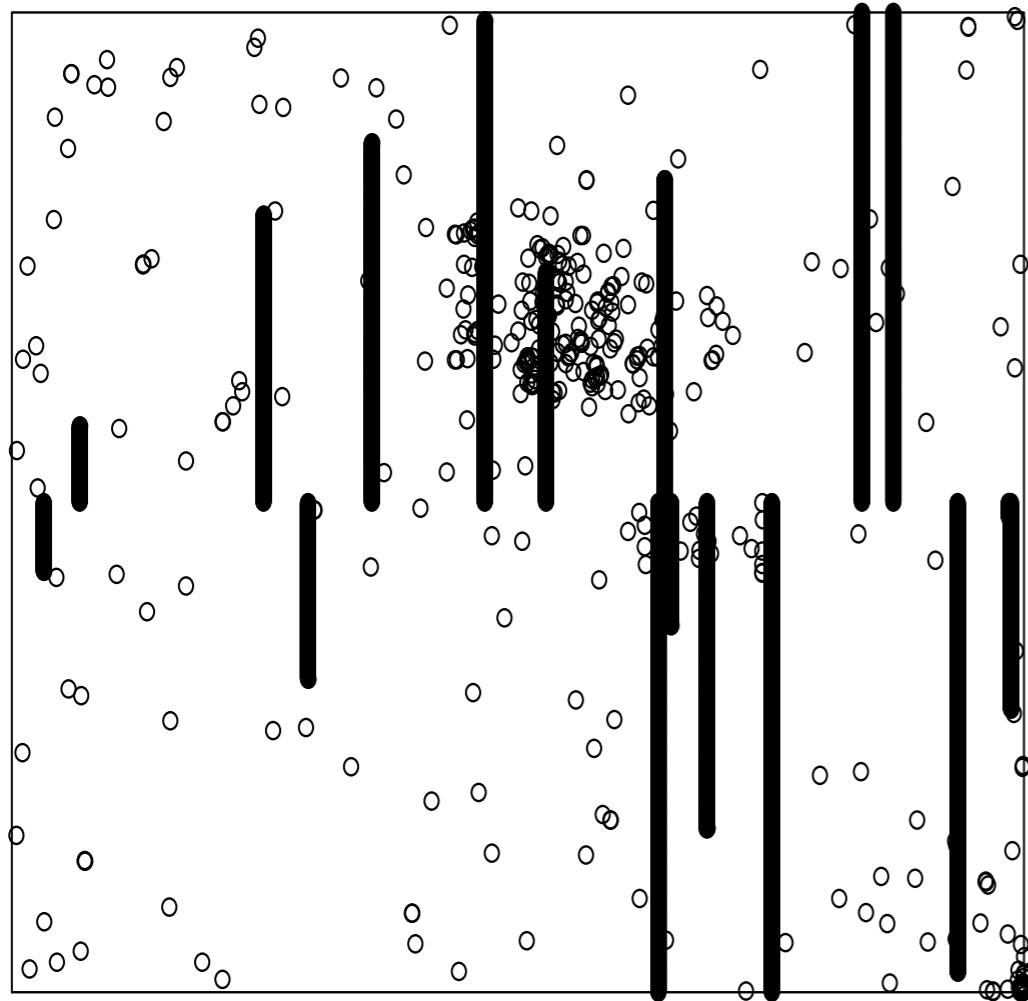
G-function, Pixels, nsim=100



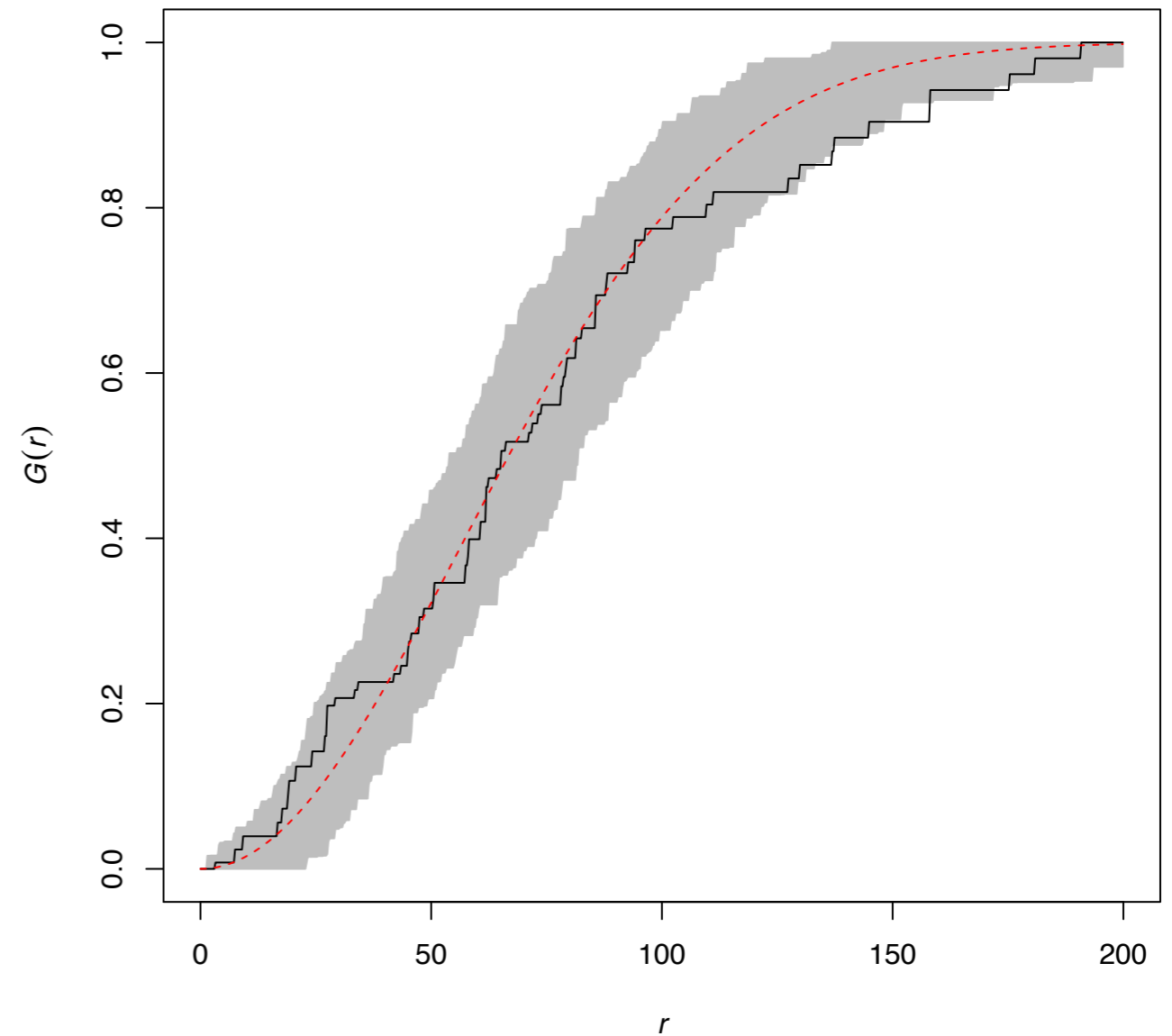
After modification: G-function

Excluding lines and area with high density

Point pattern E_0



G-function, Events, nsim=100



Software project with the Alan Turing Institute

Objectives:

Web application “DetectorChecker”

- Feedback about state of detector through pixel damage analysis
- Detector data repository

Seed funded project:

- Working with Turing Research Software Engineer Group
- *DetectorChecker* R package for statistical analysis of pixel damage in CT scanners available at <https://github.com/alan-turing-institute/DetectorChecker>
- *DetectorCheckerWebApp* for useful initial graphical/analysis, available at <https://detectorchecker.azurewebsites.net>
- Facility to upload data in different formats (crowd sourcing)

Brettschneider et al., (2020). DetectorChecker: analyzing patterns of defects in detector screens. Journal of Open Source Software, 5(56), 2474

Study I: Microtubules locations as point patterns

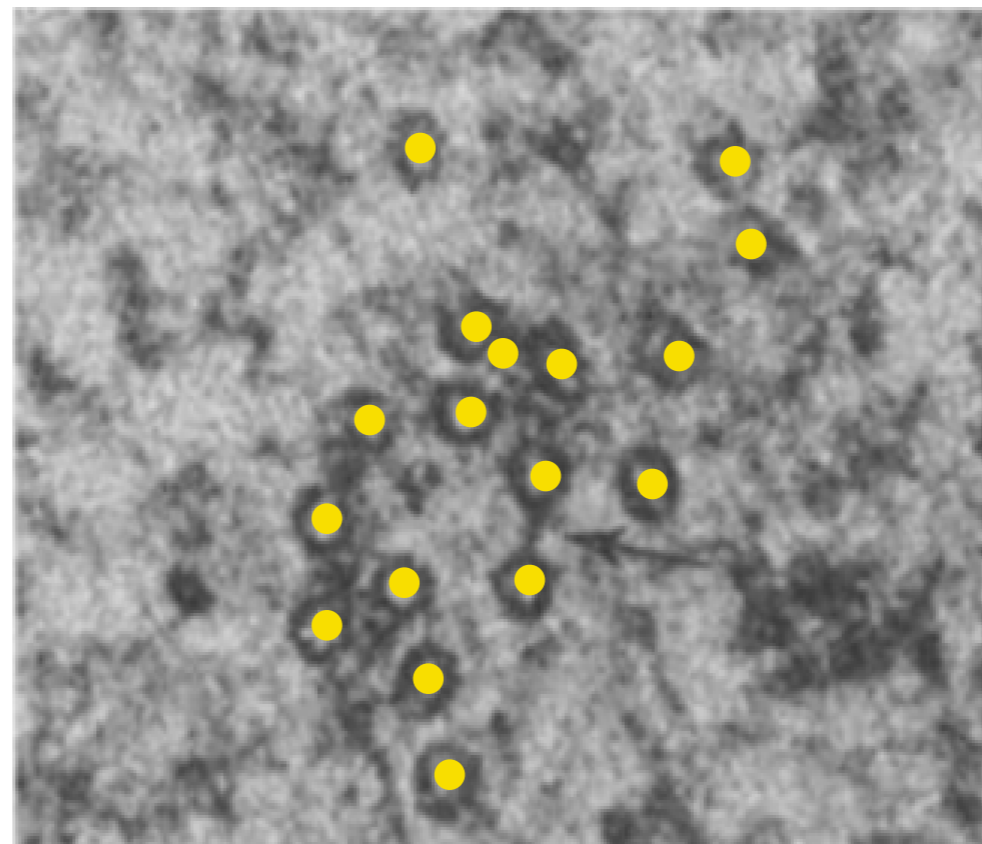
Stephen Royle's Lab (Centre for Mechanochemical Cell Biology) asks:
What is the role of TACC3 protein for the structure of microtubules within K-fibres and mesh?

Experiment: Overexpression of TACC3 through treatment versus control.

Data: Microscopic images collected in planes perpendicular to the fibre axes.

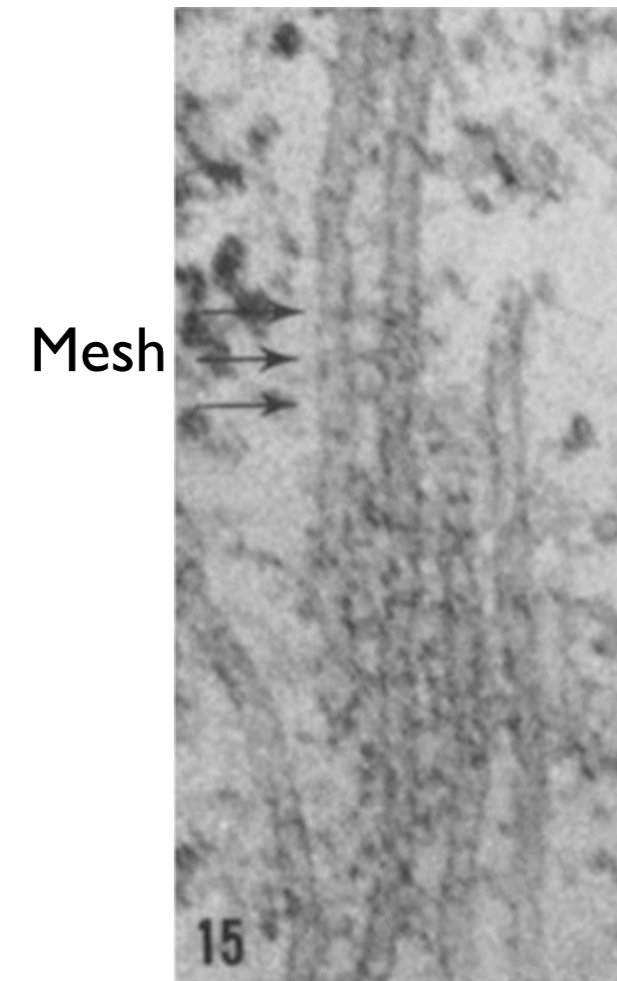


Perpendicular view



Model: locations as point pattern

Parallel



Test statistics based on basic observations

Pattern size test statistic:

$$\delta_N(I) = \frac{1}{|I_0|} \sum_{i \in I_0} n(\underline{x}^i) - \frac{1}{|I_1|} \sum_{i \in I_1} n(\underline{x}^i)$$

Observation window statistic:

$$\delta_W(I) = \frac{1}{|I_0|} \sum_{i \in I_0} |W^i| - \frac{1}{|I_1|} \sum_{i \in I_1} |W^i|$$

Intensity test statistic:

$$\sum_{i \in I_0} \omega_0(\underline{x}^i) \hat{\rho}(\underline{x}^i) - \sum_{i \in I_1} \omega_1(\underline{x}^i) \hat{\rho}(\underline{x}^i)$$

where $\delta_\rho(I)$ denotes unweighted case using $\omega_k(\underline{x}^i) = 1/|I_k|$ ($k = 0, 1$)

$\delta_{\rho,\omega}(I)$ denotes weighted case using $\omega_k(\underline{x}^i) = n(\underline{x}^i) / \sum_{j \in I_k} n(\underline{x}^j)$ ($k = 0, 1$)

Test statistics based on nearest neighbours

Mean nearest neighbour test statistic:

$$\delta_{\text{nnd}}(I) = \sum_{i \in I_0} \omega_0(\underline{x}^i) \overline{\text{nnd}}(\underline{x}^i) - \sum_{i \in I_1} \omega_1(\underline{x}^i) \overline{\text{nnd}}(\underline{x}^i)$$

Unweighted and weighted versions as above.

Further work includes mean minimum spanning test statistics.

Test statistics based on G-functions

Estimated nearest neighbour functions averaged over the collection of point patterns \underline{x}^J with weights ω_J as above:

$$\hat{G}(\underline{x}^J, r) = \sum_{i \in J} \omega_J(\underline{x}^i) \hat{G}(\underline{x}^i, r)$$

Nearest neighbour distribution test statistic statistics:

$$\delta_{G,1}(I) = \|\hat{G}(\underline{x}^{I_0}, r) - \hat{G}(\underline{x}^{I_1}, r)\|_1 = \int_0^\infty |\hat{G}(\underline{x}^{I_0}, r) - \hat{G}(\underline{x}^{I_1}, r)| dr$$

$$\delta_{G,\infty}(I) = \|\hat{G}(\underline{x}^{I_0}, r) - \hat{G}(\underline{x}^{I_1}, r)\|_\infty = \sup_r |\hat{G}(\underline{x}^{I_0}, r) - \hat{G}(\underline{x}^{I_1}, r)|$$

For comparison of $\hat{G}(\underline{x}^{I_0}, r)$ and $\hat{G}(\underline{x}^{I_1}, r)$ across the range of distances $r > 0$.

Also, scaled neighbourhood count test statistic (Diggle 2000).

Significance quantification

- Based on permutation tests (nonparametric)
- Need exchangeability under the Null under suitable set of operations
- Statistics under permutations are identically distributed
- p-values are uniformly distributed (test e.g. with KS)
- Exact or approximate (subset of operations)

Operations $\Gamma = \{\gamma_0, \gamma_1, \dots, \gamma_m\}$, where $\gamma_0 = \text{Id}$.

p-value for two-sided test of H_0 using statistic t :

$$p = \frac{1}{m+1} \sum_{\gamma \in \Gamma} 1_{\{|t(\gamma x)| \geq |t(\gamma_0 x)|\}}$$

Γ : random subsets of the symmetric group S_I

Permuted I results in subsets $I_0^{(j)}$ and $I_1^{(j)}$ satisfying

$$I_0^{(j)} \cup I_1^{(j)} = I, \quad I_0^{(j)} \cap I_1^{(j)} = \emptyset, \quad |I_0^{(j)}| = |I_0|, \quad |I_1^{(j)}| = |I_1|$$

Simulation study

- Homogeneous Poisson process density ρ
- Inhomogeneous Poisson process with density $\rho(x)$
- Cluster point pattern generated by rejection sampling algorithm

Data: d, W

Result: Point pattern \underline{x} on W with nearest neighbour distances $\{d_1\} \cup d$

$x_1 \leftarrow$ centre of W ;

for i in $1 : n$ do

 repeat

$j \sim \text{Uniform}\{1, \dots, i\}$;

$\theta \sim \text{Uniform}[0, 2\pi]$;

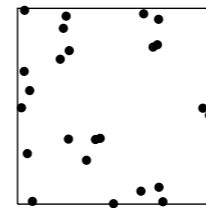
$x_{i+1} \leftarrow x_j + d_i(\cos \theta, \sin \theta)$;

 until $\min_{k \in \{1, 2, \dots, i\}} \|x_{i+1} - x_k\| \geq d_i$ and $x_{i+1} \in W$;

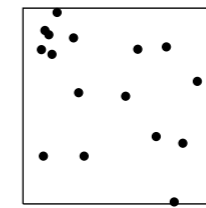
end

$\underline{x} \leftarrow (x_1, x_2, \dots, x_{n+1})$;

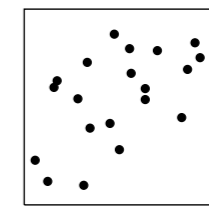
Homogeneous intensity



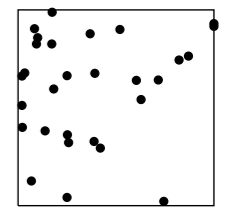
$\alpha = 1$



$\alpha = 1.1$

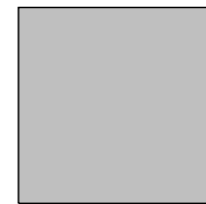


$\alpha = 1.2$

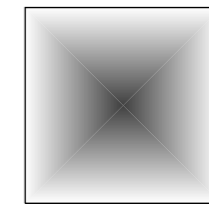
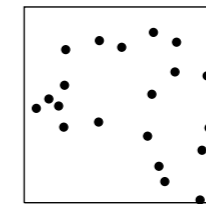


$\alpha = 1.5$

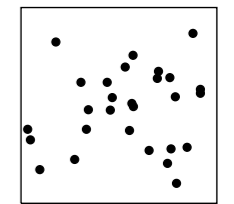
Inhomogeneous intensity



Homogeneous



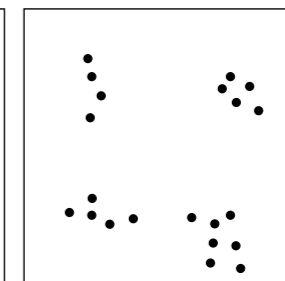
Inhomogeneous



Disjoint cluster



Single cluster

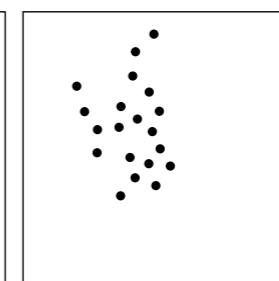


Multiple clusters

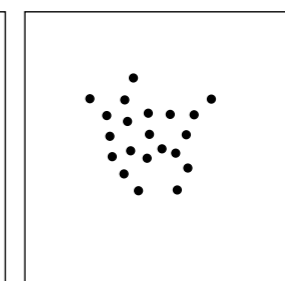
Cluster variance



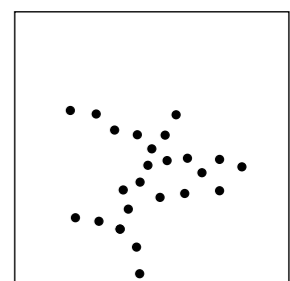
$\alpha = 1$



$\alpha = 1.1$



$\alpha = 1.2$



$\alpha = 1.5$

Study I: Microtubules

Stephen Royle's Lab (Centre for Mechanochemical Cell Biology):

What is the role of the TACC3 protein for the structure of microtubules within K-fibres and mesh?

Experiment:

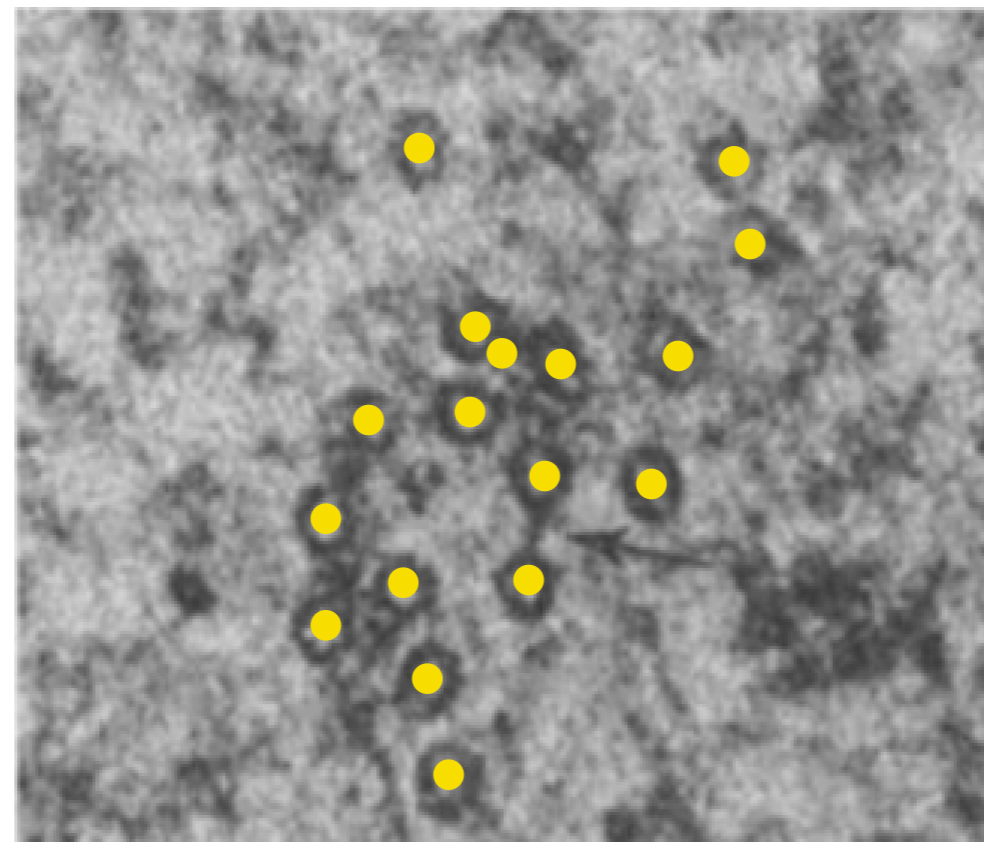
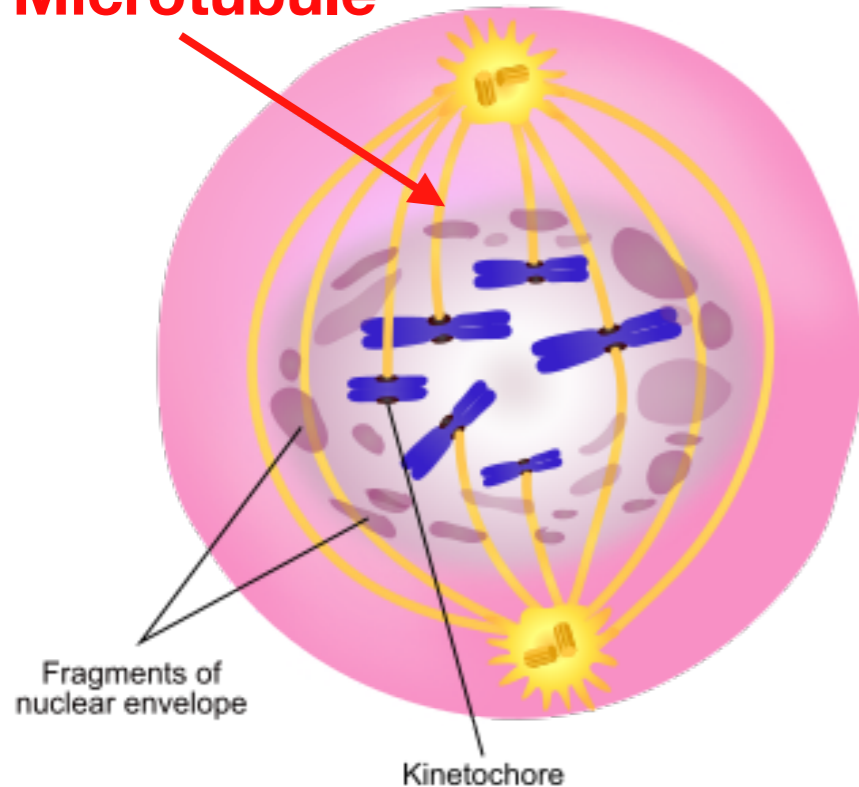
Overexpression of TACC3 through treatment versus control.

Microscopic images collected in planes perpendicular to the fibre axes.

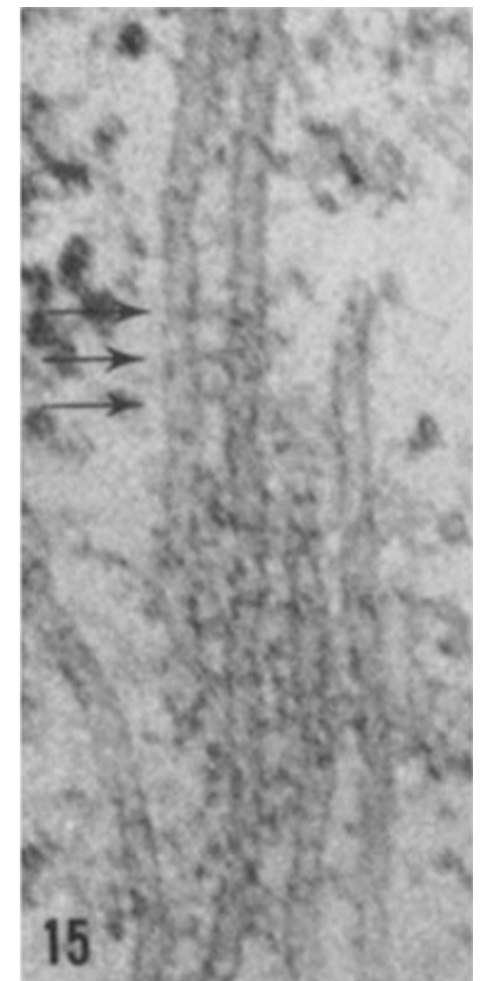
Perpendicular to the microtubule axis

Parallel showing mesh

Microtubule



Model locations as point pattern

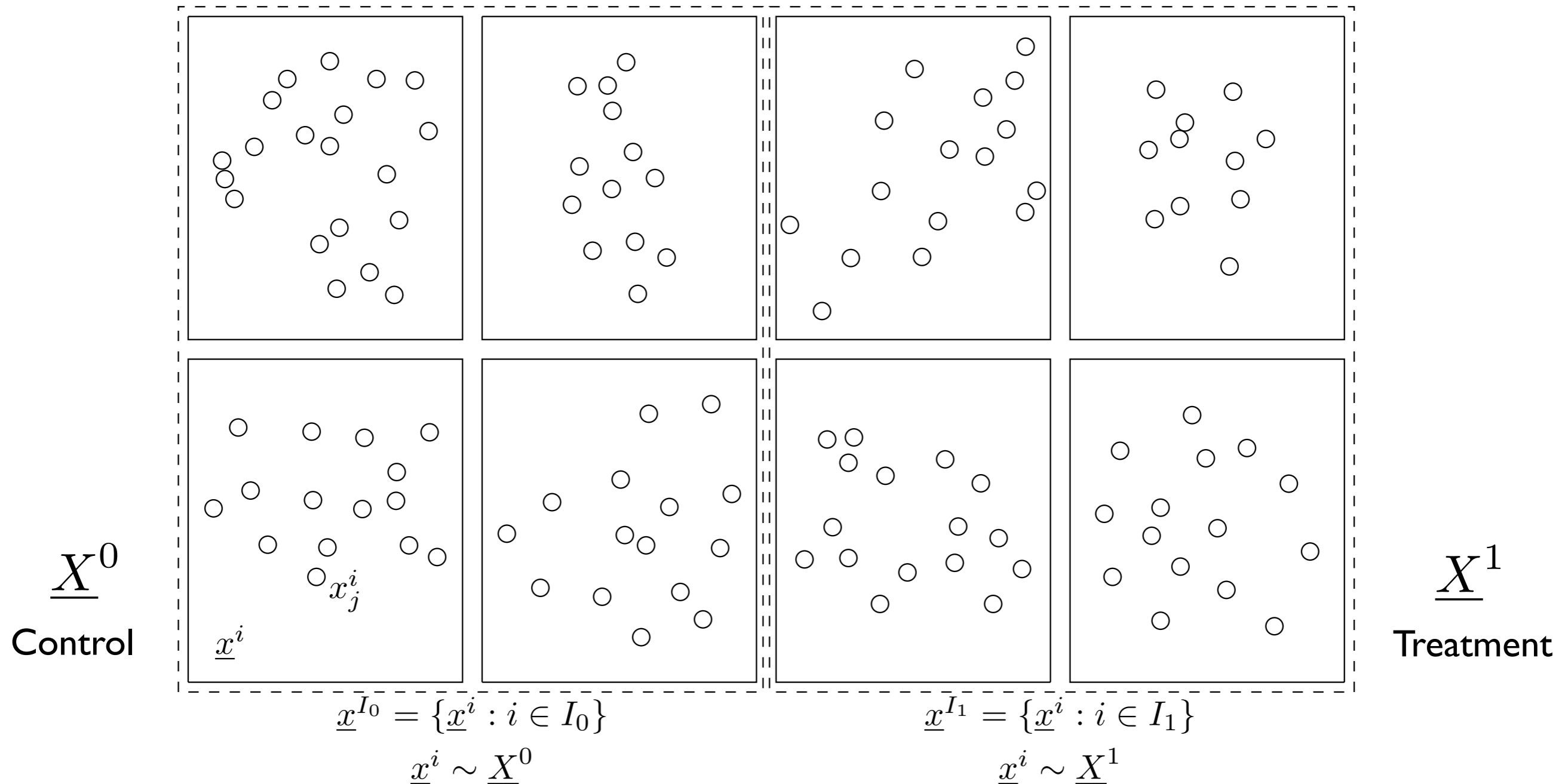


Study I: Microtubules - mathematical model

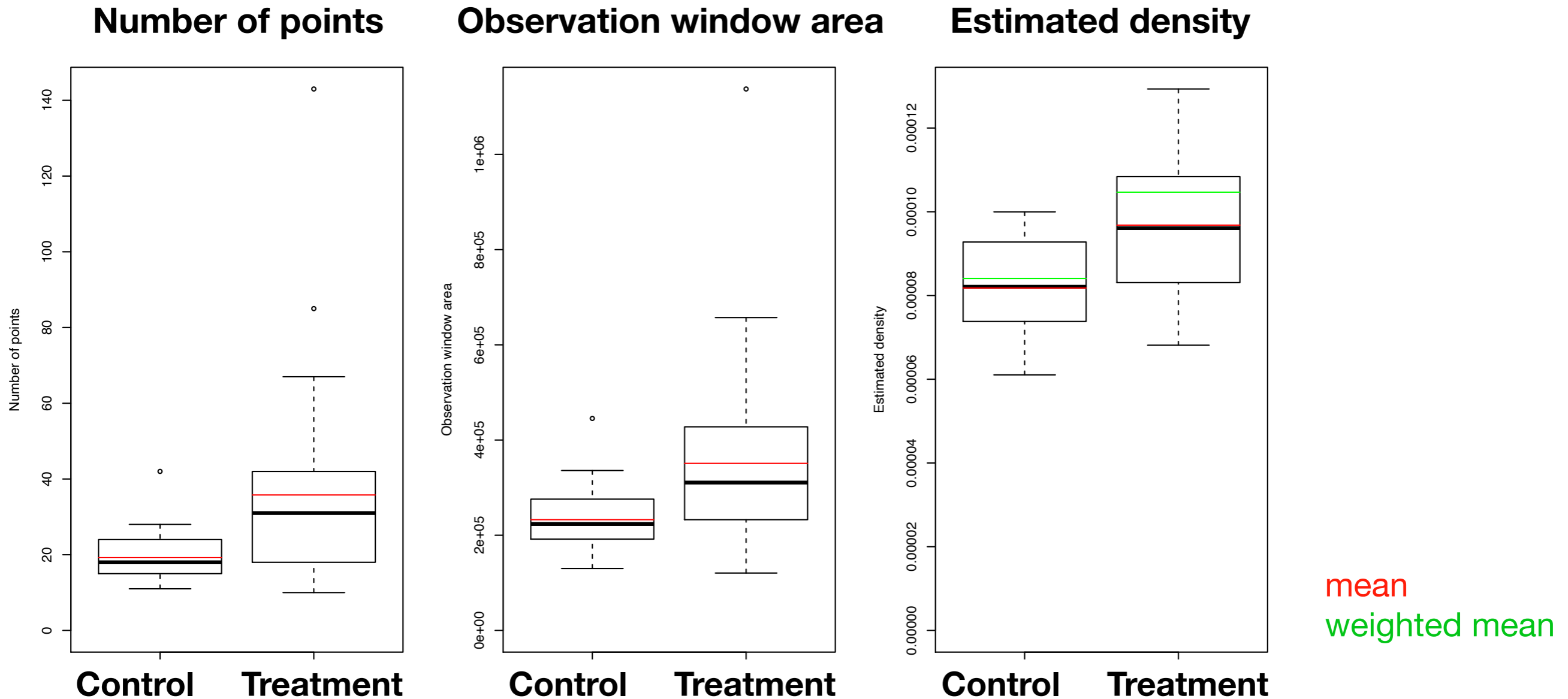
Data:

Microscopic images of treatment (n=37) versus control (n=26)

Observation window surrogate for cross sectional area of K-fibres



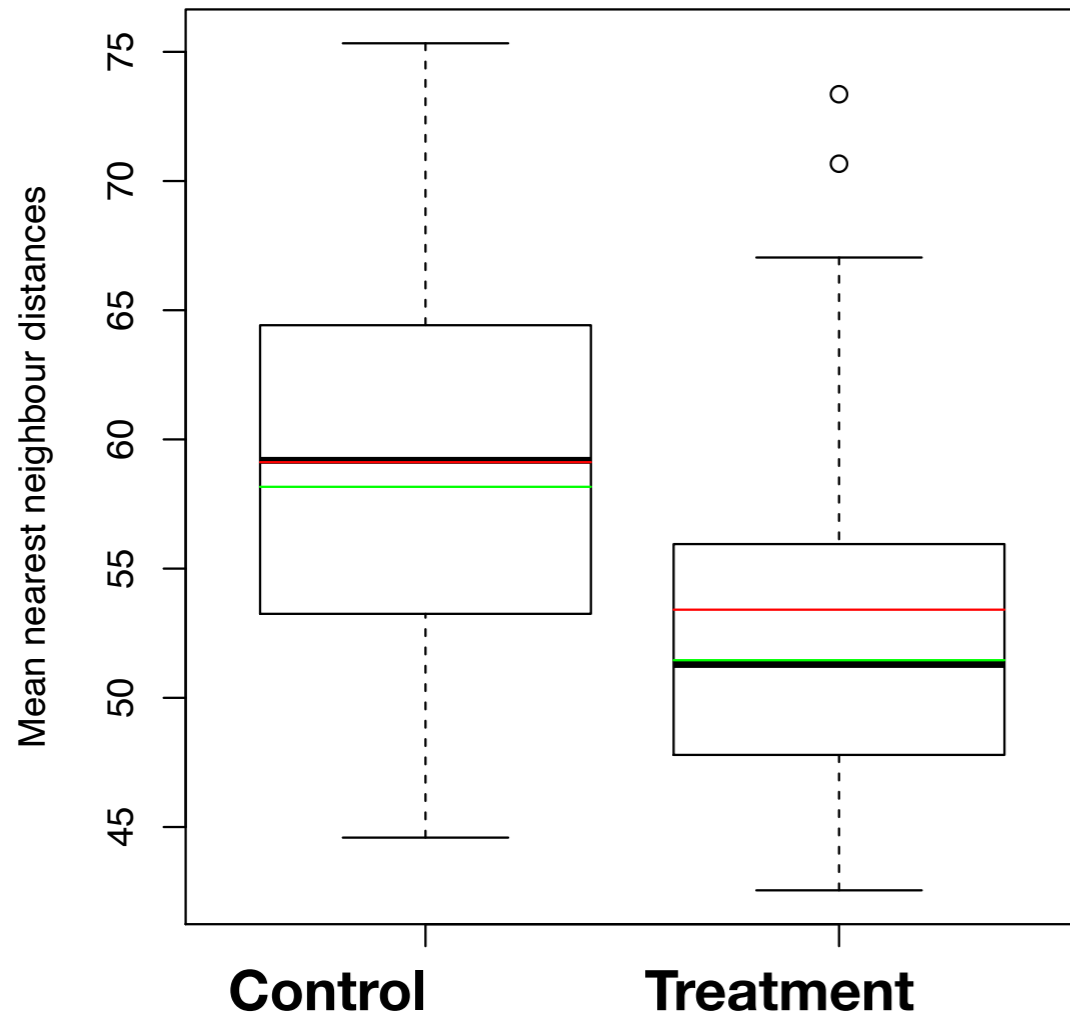
EDA: First order statistics



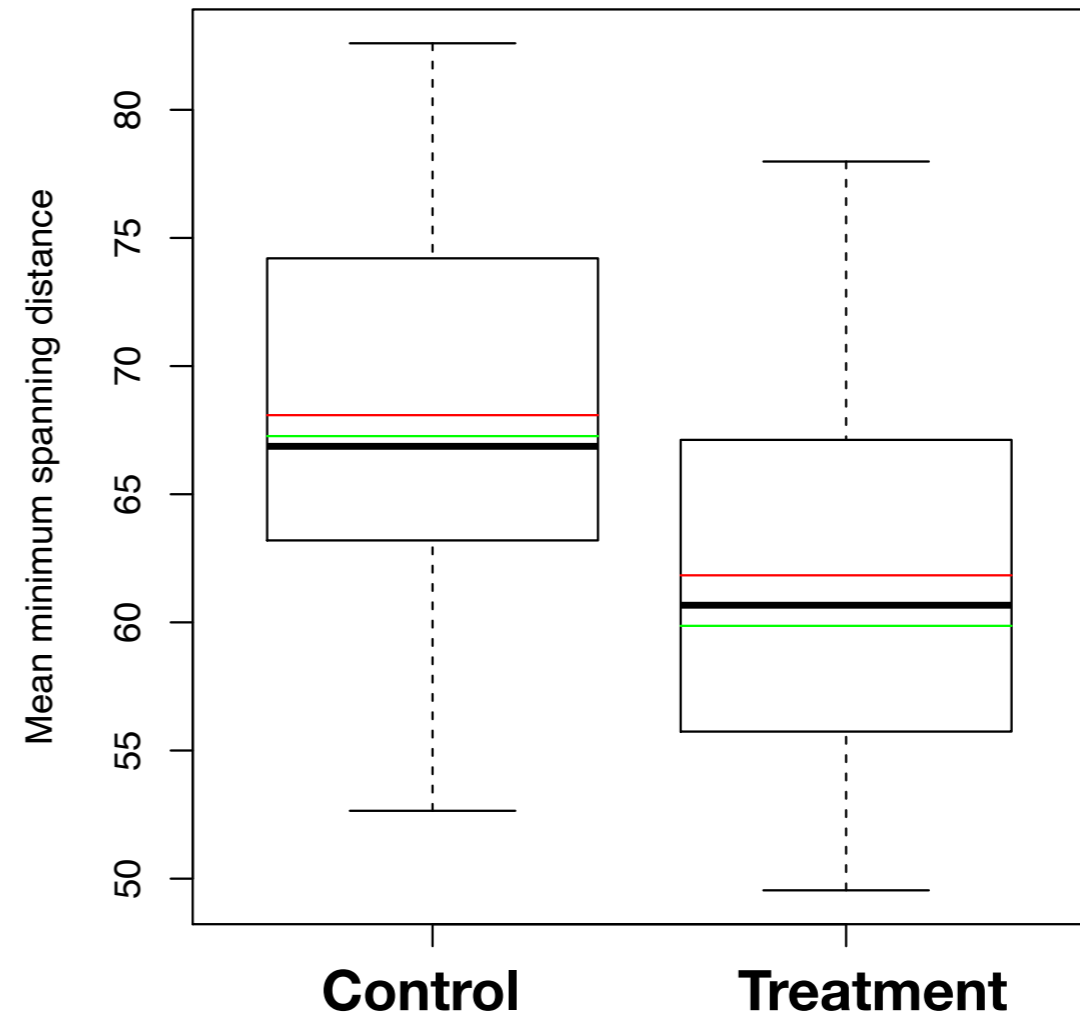
- All means/medians are greater for treatment
- Treated K-fibers are made up of a greater number of microtubules which are more closely separated within thicker K-fibers
- Weighted mean densities greater than unweighted means densities (i.e. K-fibers with greater numbers of microtubules are more tightly packed)

EDA: Second order statistics

Mean nearest neighbour distance



Mean minimum spanning distance

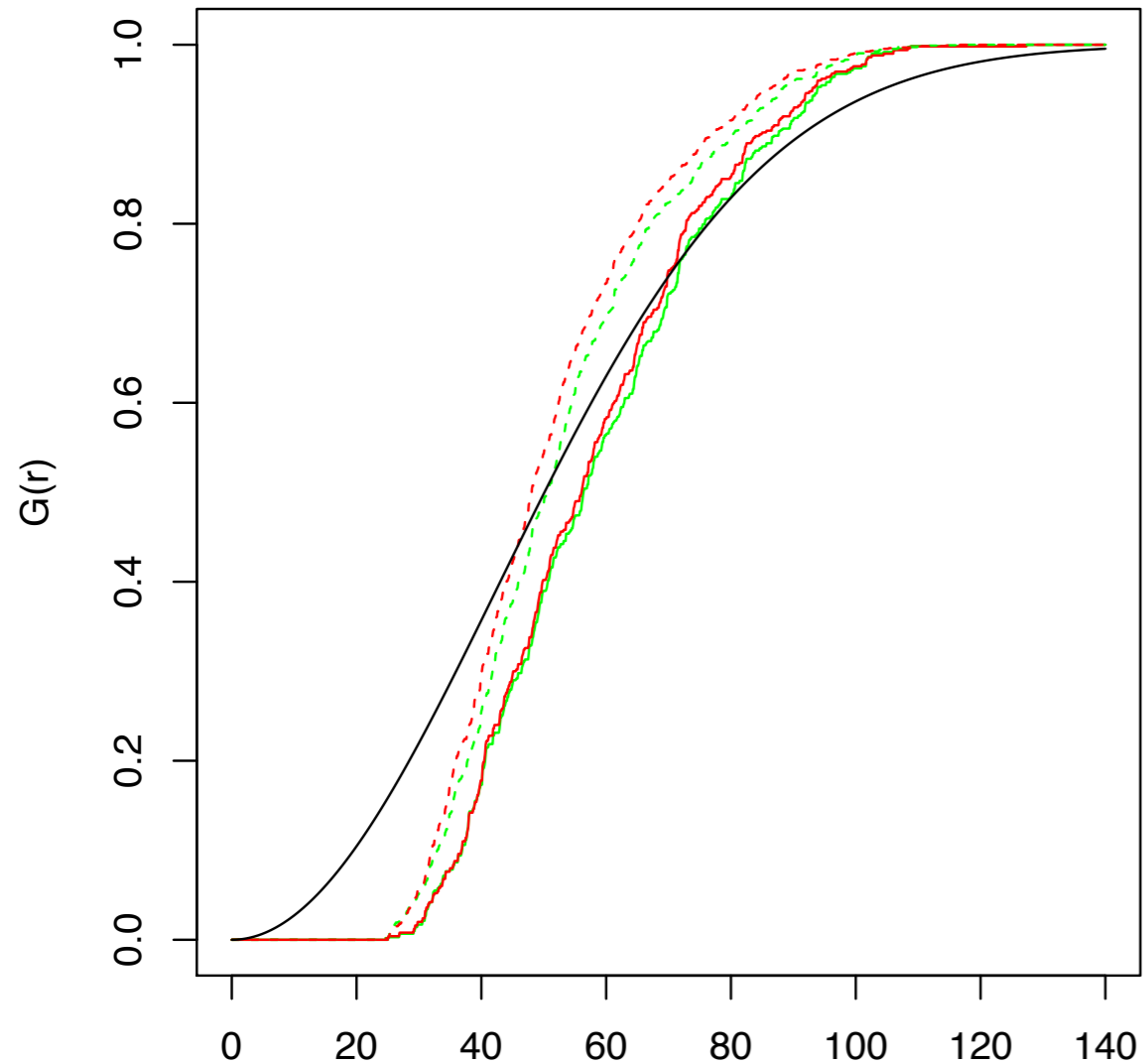


mean
weighted mean

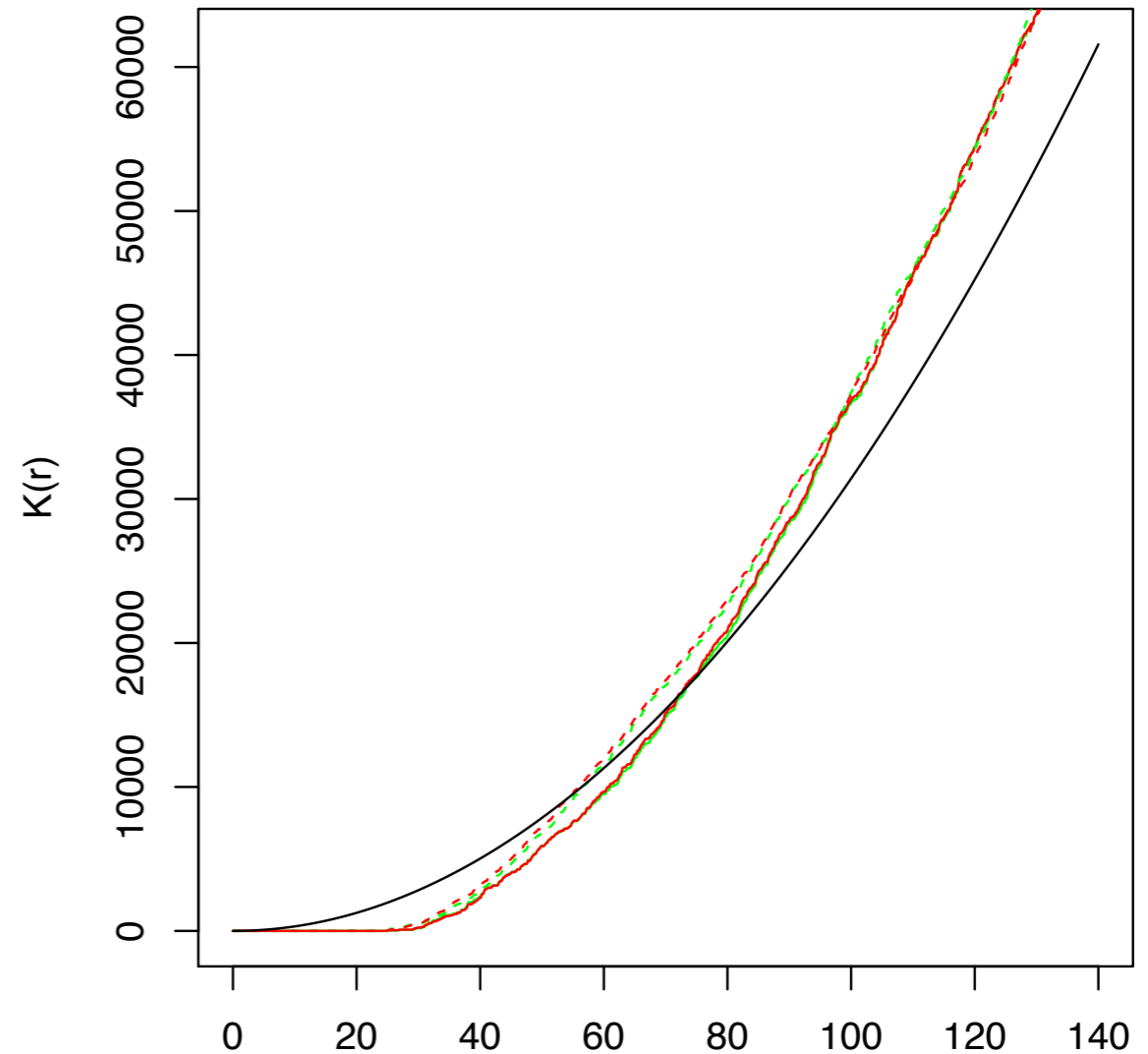
- In each case the average separation distance is reduced for treatment observations
- Weighted means smaller than unweighted means (i.e. K-fiber with more microtubules are more tightly packed).

EDA: Spatial functions

average G functions



average K functions



— control, mean r — control, weighted mean
- - - - - treatment, mean - - - - - treatment, weighted mean — Homogeneous Poisson

- Some evidence of clustering at larger length scales
- Effect of limitation of nnd in [25,105]
- Difference between weighted mean and unweighted mean negligible

Test statistics

Observations of exploratory analysis can be confirmed by formal testing.
All proposed test statistics show significant results:

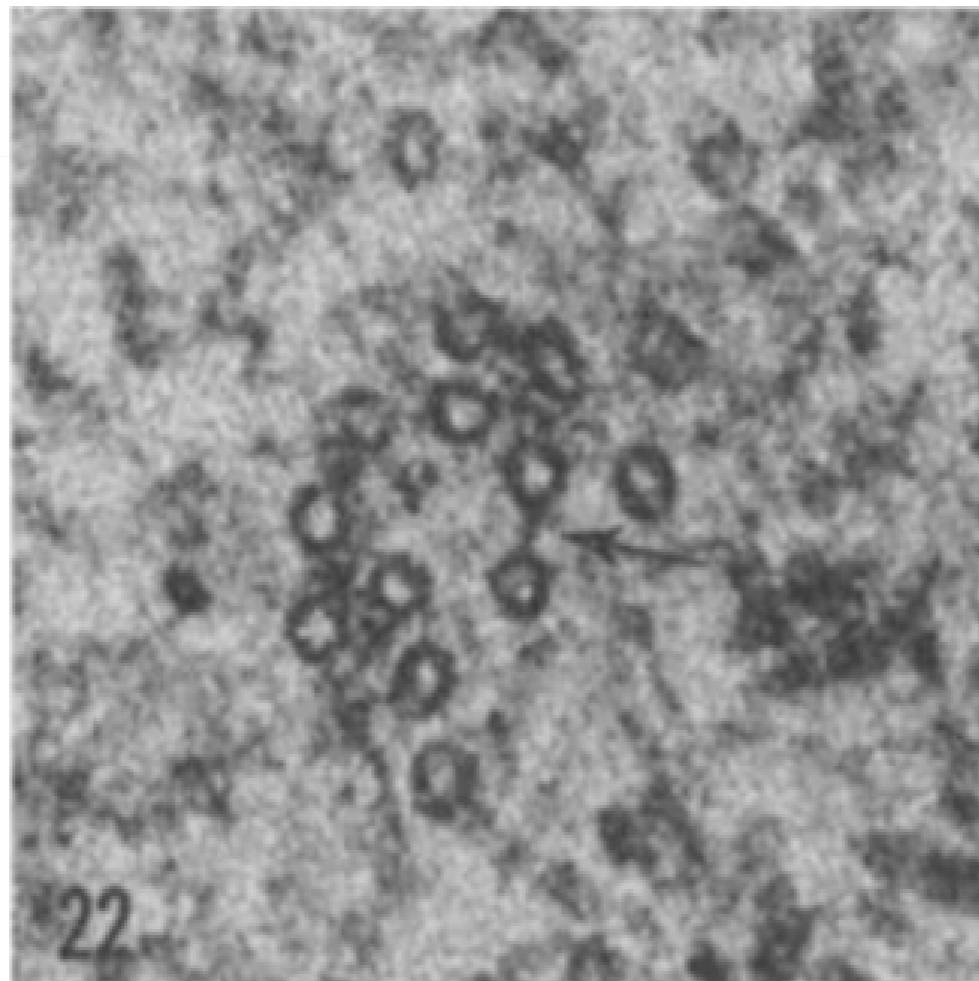
δ_N	0.0005	δ_{nnd}	0.0057	δ_K	0.1092	δ_{EFT}	0.0011
δ_W	0.0018	$\delta_{\text{nnd},\omega}$	0.0005	$\delta_{G,1}$	0.0061	$\delta_{EFT,\omega}$	0.0005
δ_ρ	0.0001	δ_{msd}	0.0019	$\delta_{G,1,\omega}$	0.0005		
$\delta_{\rho,\omega}$	0.0002	$\delta_{\text{msd},\omega}$	0.0005	$\delta_{G,\infty}$	0.0087		
				$\delta_{G,\infty,\omega}$	0.0013		

Results remain significant after multiple testing adjustment of critical p-value (using Bonferroni).

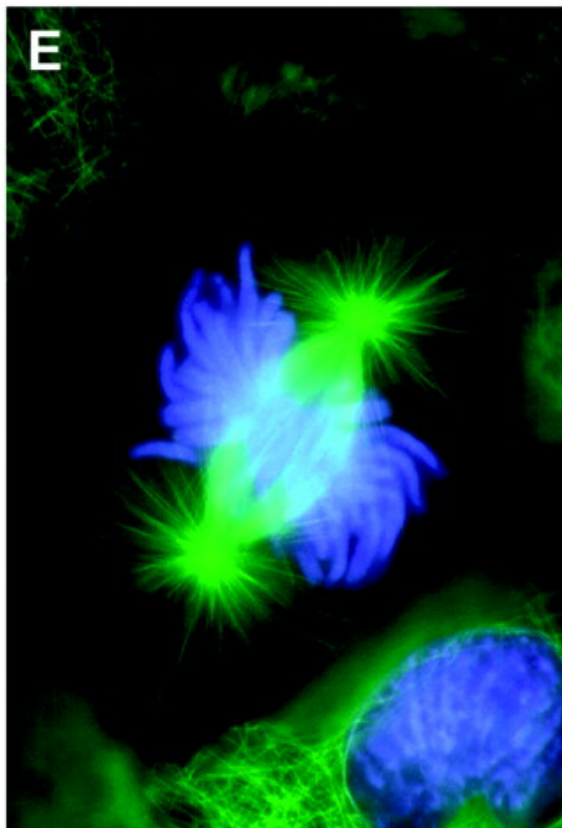
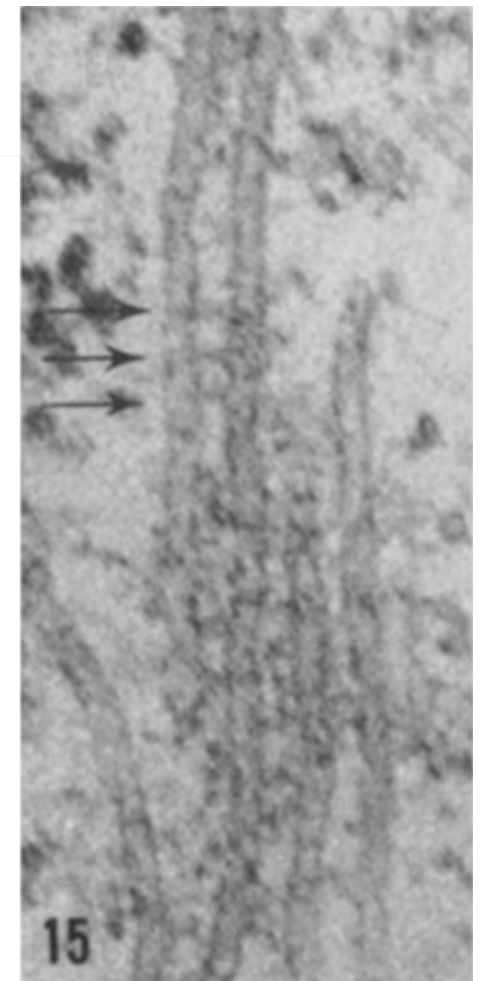
What did we find?

- Microtubules are **bound together** (in K-fibers, by mesh-like structure)
- TACC3 **overexpression** is **associated with an impact** on the mesh
- Detection of treatment effects not visible by eye

Perpendicular to the microtubule axis



Parallel view



Study II:

Relationship between two proteins over time

Background and questions

- Protein EB3 localises at the tip of growing microtubules during mitosis.
- Relationship between TACC3 and EB3?
- Role of the protein TACC3 during that same process?

Data

- Confocal fluorescence microscopy images collected across seven samples at a total number of between 47 and 57 time points.
- Images are collected of live cells during mitosis with TACC3 tagged with a green fluorescing protein and EB3 tagged with a red fluorescent protein.

Question that can be answered from images

Is the protein TACC3 present in the same locations as the protein EB3 during the process of mitosis?

Dependencies between bulk movement patterns

Mathematical formulation of case study question

What is the relationship between two bulk movement patterns?

Modelling

- Measure for closeness of two spatial protein distributions
- Comparing their evolution over time

Applications

- Animals e.g. predator and prey (Mitchell and Lima, 2002)
- Air particles, e.g. pollution
- Cellular structures (Chenouard 2014)

Tool: Colocalisation of proteins

Example for (traditional) visual detection of colocalisation

Colocalization of Actin and Vinculin in Normal Tahr Ovary Cells

Colocalization in the lateral optical plane of the cytoskeletal protein **actin** with **vinculin**, a protein associated with focal adhesion and adherens junctions.

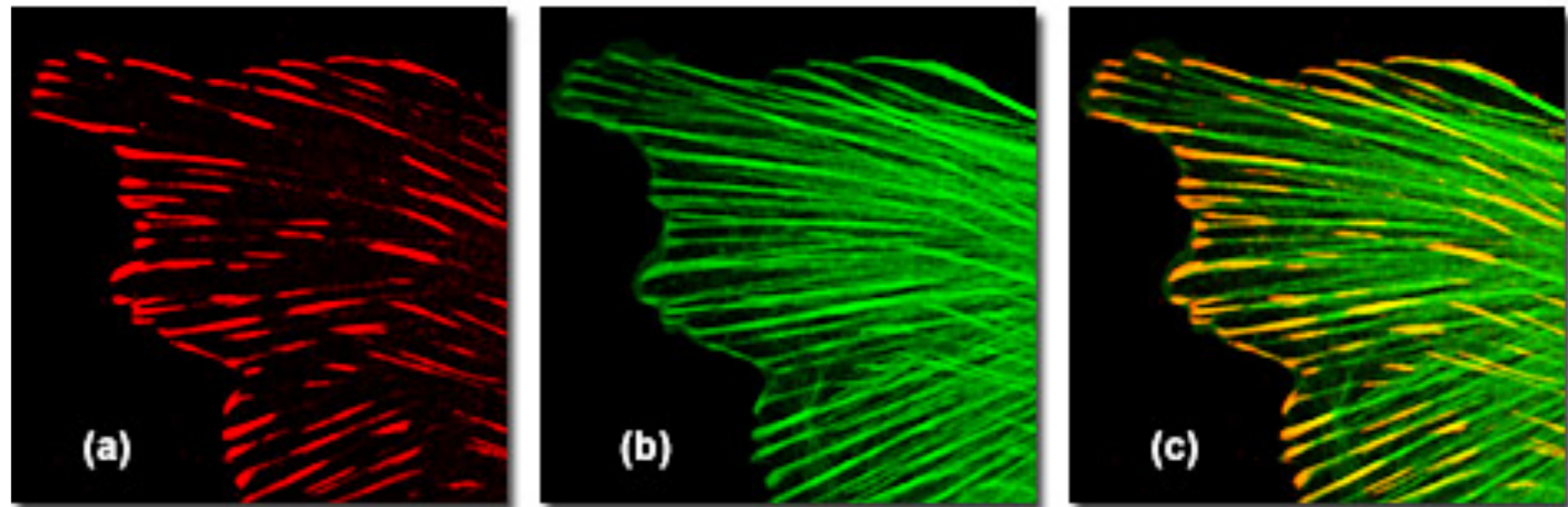


Figure 1

Applications

- Detect physical location within cell
- Uncover functions of proteins based on location
- Unravel interactions, build networks, infer function

Quantifying colocalisation (static)

Correlation

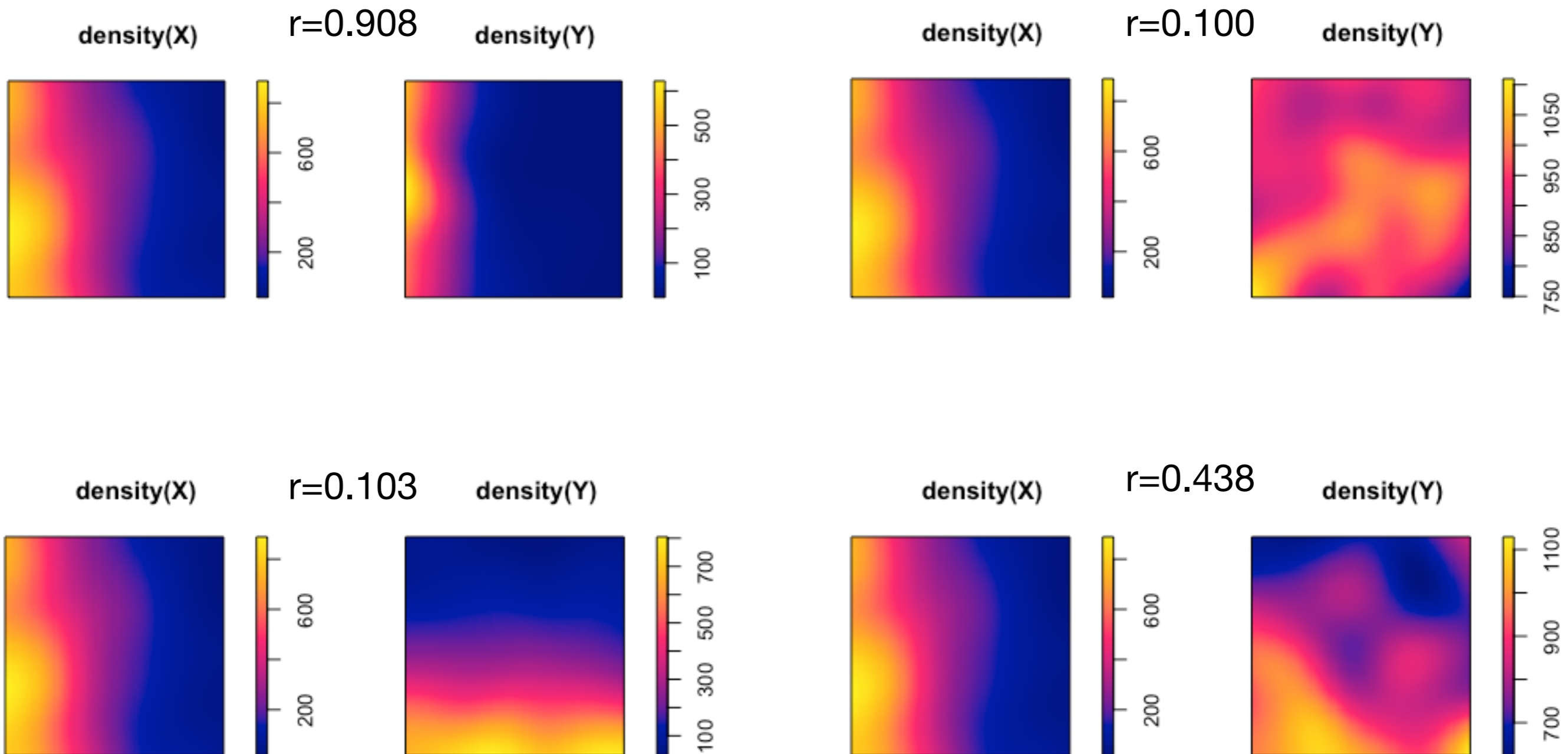
$$r_p = \frac{\sum_i (A_i - a)(B_i - b)}{\sqrt{\sum_i (A_i - a)^2 (B_i - b)^2}}$$

where A_i and B_i are the voxel or pixel intensities (also called grey values) of channels A and B, respectively, and a and b are the corresponding average intensities over the entire image.

- Scaling invariant
- Costes' threshold to deal with noise
- Ongoing area of research, e.g. Wang et al. (2018) for automatic segmentation

Quantifying colocalisation: Illustration

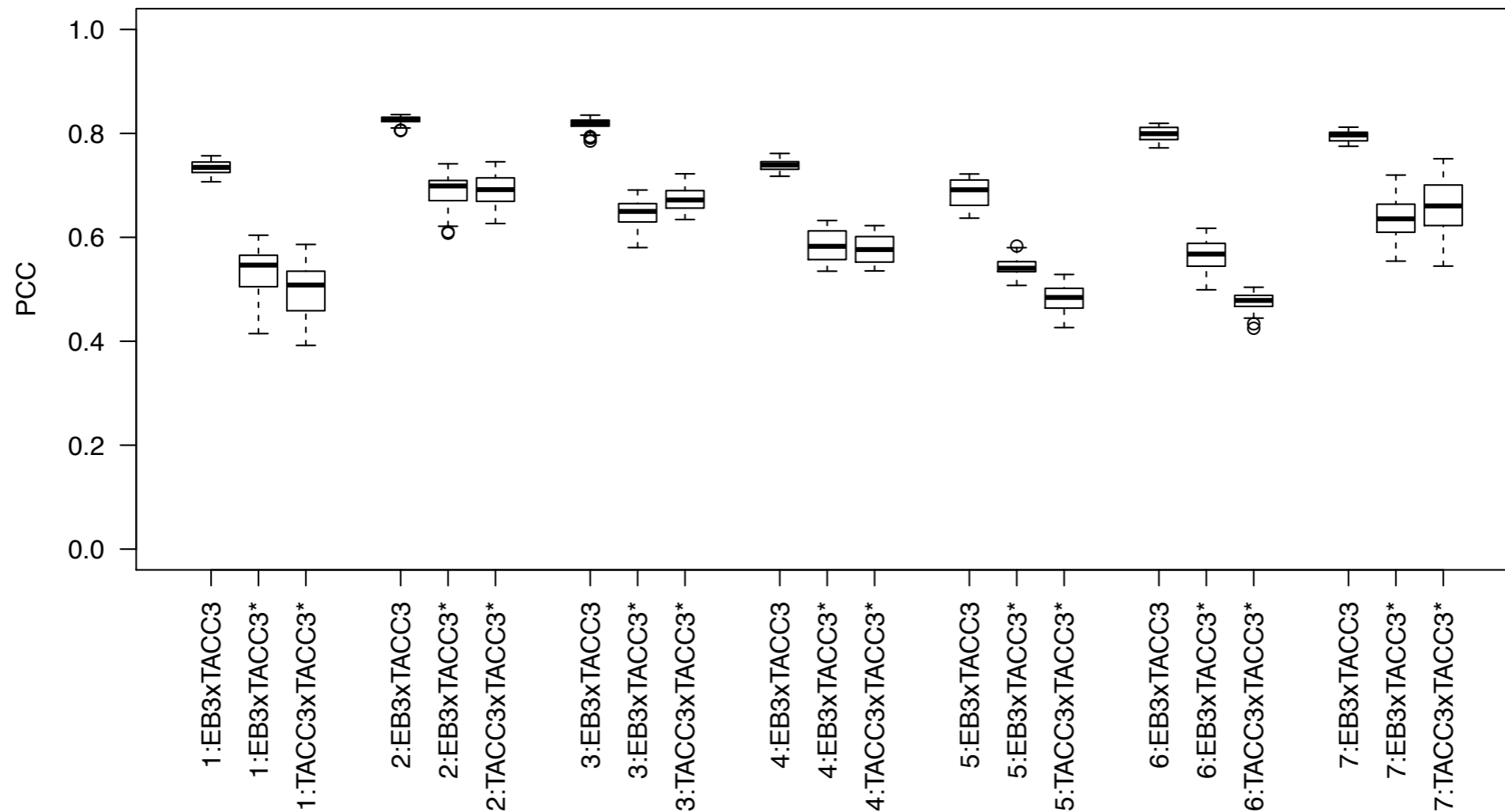
Correlation examples (simulations) for hypothetical proteins X and Y



Intuitive check: Colocalisation for consecutive time points

Colocalisation of protein distribution between consecutive time points for each sample combination of TACC3, TACC3 (vertical reflection), and EB3.

TACC3* and TACC3: only coincidental similarity expected



Colocalisation of EB3 with TACC3 is always higher than any of the other combinations.

Model for bulk movement patterns

Need to have a model that captures the evolution rather than individual time points.

Observed pixel intensity values $m^0(x)$ and $m^1(x)$ across ROI:

$$\Psi^* \subseteq \Psi = \{1, 2, \dots, n_1\} \times \{1, 2, \dots, n_2\}$$

Spatio-temporal process M denoted by $M_t(x)$ ($x \in \Psi, t \in \Upsilon$).

$F_{s,t}(x, y)$ mass moving from location x at time s to y at time t .

Direct dependency of movement patterns F^0 and F^1 :

Mass $F_{s,t}^0(x, y)$ positively associated with $F_{s,t}^1(x, y)$

(across all pairs of locations and times).

Use earth movers distance (EMD) (Kantorovich-Wasserstein metric).

Idea: Find minimal transportation costs

Earth mover's distance (EMD)

Non-negative spatial processes m^0 and m^1 over χ^0 and χ^1 .

**Work normalised
by the total flow**

$$\text{EMD}(m^0, m^1) = \frac{\sum_{x \in \chi^0, y \in \chi^1} \hat{f}(x, y) d(x, y)}{\sum_{x \in \chi^0, y \in \chi^1} \hat{f}(x, y)}$$

Flow that minimises overall cost

$$\hat{f} = \operatorname{argmin}_{f \in \eta(m^0, m^1)} \sum_{x \in \chi^0, y \in \chi^1} f(x, y) d(x, y),$$

for cost function $d(x, y)$ and $\eta(m^0, m^1)$ the set of f for which

$$f(x, y) \geq 0 \quad \forall x \in \chi^0, y \in \chi^1$$

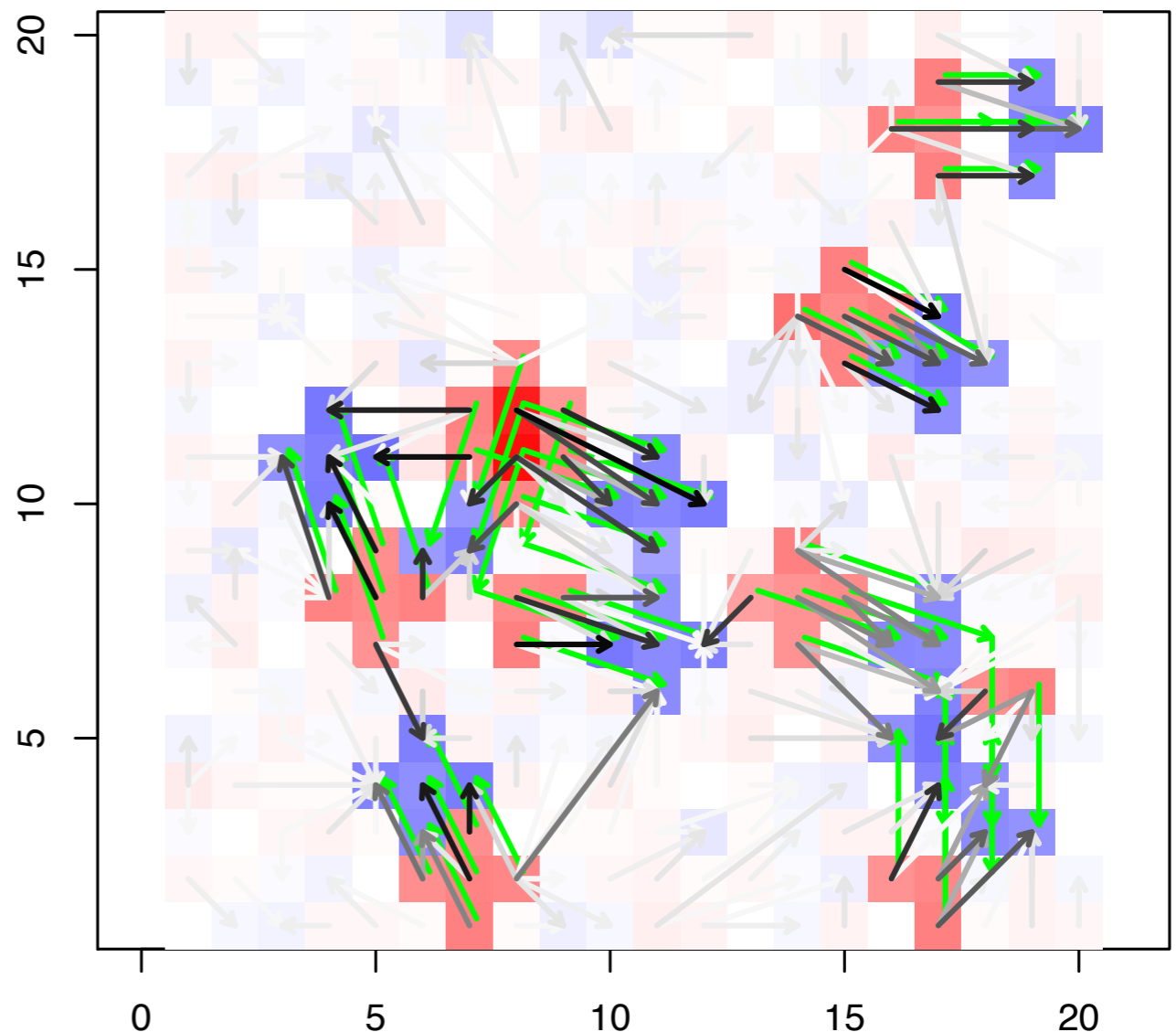
$$\sum_{x \in \chi^0} f(x, y) \leq m^1(y) \quad \forall y \in \chi^1$$

$$\sum_{y \in \chi^1} f(x, y) \leq m^0(x) \quad \forall x \in \chi^0$$

$$\sum_{x \in \chi^0, y \in \chi^1} f(x, y) = \min \left(\sum_{x \in \chi^0} m^0(x), \sum_{y \in \chi^1} m^1(y) \right).$$

Movement summary statistics

- Built on EMD
- Discretisation into 8 directions
- Subregions (to avoid unintuitive results from large spaces)
- Specification of Null hypothesis
- Permutation test set up
- Simulation study



Null comprised of 3 statements

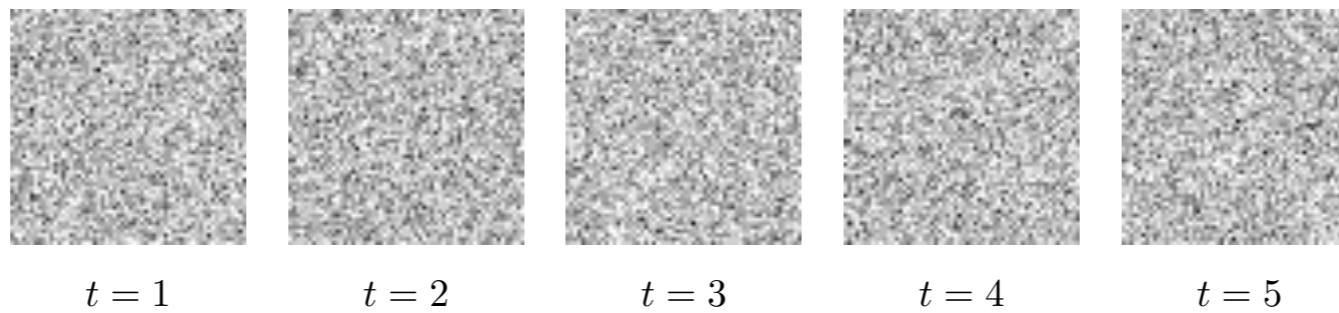
1. Between-sample independence of local bulk movement patterns:
 $\{S_{s,t}^{\psi_1,0}, S_{s,t}^{\psi_2,0}, \dots, S_{s,t}^{\psi_w,0}\}$ independent of $\{S_{s,t}^{\psi_1,1}, S_{s,t}^{\psi_2,1}, \dots, S_{s,t}^{\psi_w,1}\}$.
2. Specify set of operations Λ .
3. Within-sample independence of local bulk movement patterns:
 $S_{s,t}^{\psi_j}$ is independent of $S_{s,t}^{\psi_k}$ for $j \neq k$
(required to ensure exchangeability under the action of $\lambda \in \Lambda$).

Example hypotheses

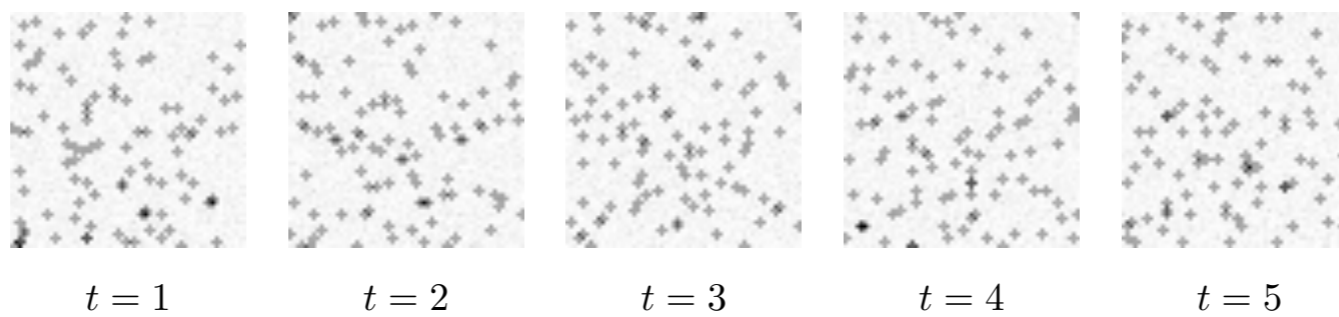
- Isotropic: rotation, reflection, reordering
- Homogeneous: reordering
- Symmetric: rotation, reflecting (limited)
- Horizontal reflection

Simulated data

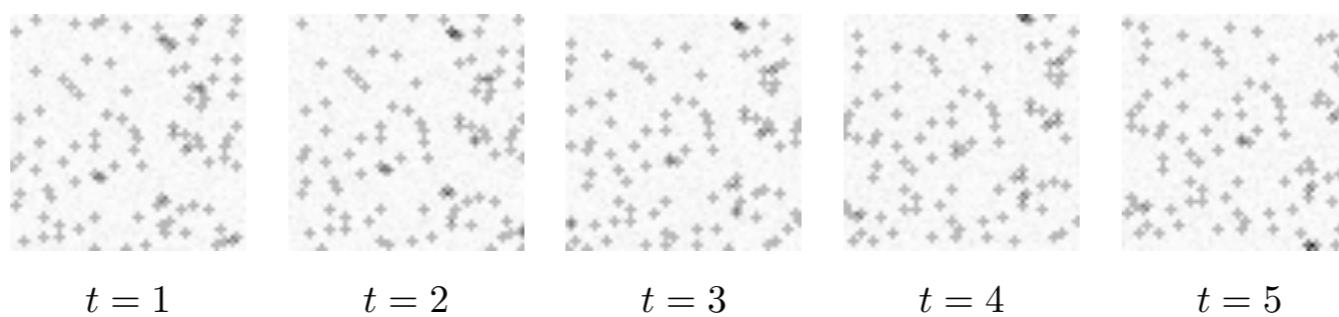
Noise



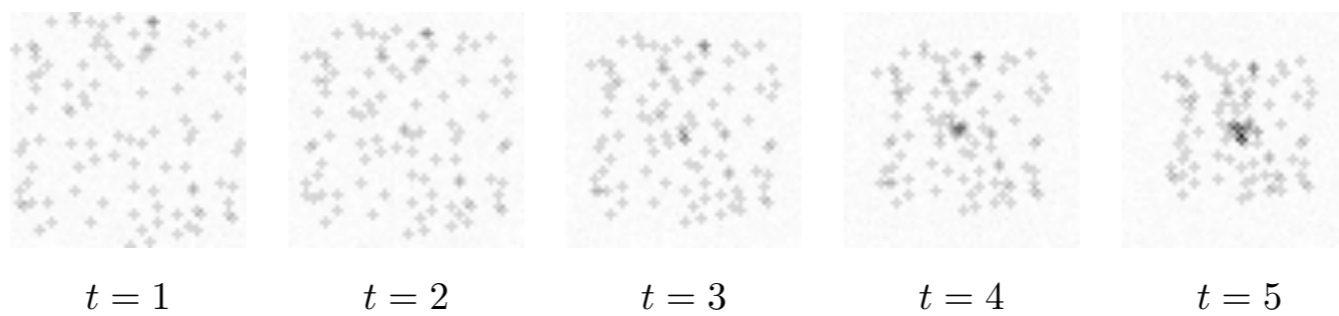
Isotropic



Homogeneous



Symmetric



Simulation results

Independent movement

- mostly confirms theoretical method
- some issues with composite hypothesis

Dependent movement

- higher rate of incorrect rejections
- evidence for validity of omnibus hypothesis approach

Study II:

Evolution of bulk movement patterns

Question

Does the protein TACC3 evolve spatially colocalised with the protein EB3 during the process of mitosis?

Analysis

Compare movement patterns of TACC3 and EB3 during mitosis using EMD.

Results

Omnibus null hypothesis consistently rejected at 5%.

Conclusions

- Movement patterns of EB3 and TACC3 are dependent
- Potentially through their localisation on the tips of growing microtubules

Thanks

Microscopy work is joint with Tom Honnor (Warwick Statistics, now UCL), Adam Johnson (Warwick Statistics), Steve Royle (Warwick Medical School)

*Thomas R. Honnor, Julia A. Brettschneider, Adam M. Johansen (2017),
Differences in spatial point patterns with application to subcellular biological structures*

*Honor TR, Johansen AM and Brettschneider JA. (2017)
A nonparametric test for dependency between estimated local bulk movement patterns*

Nixon, F.M., Honnor*, T.R., Starling, G.P., Beckett, A.J., Johansen, A.M., Brettschneider, J.A., Prior, I.A. & Royle, S.J.
J Cell Science, April 2017*

Microtubule organization within mitotic spindles revealed by serial block face scanning EM and image analysis