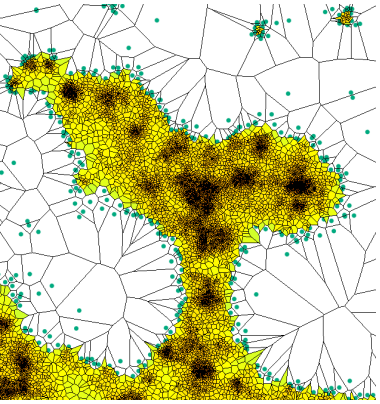
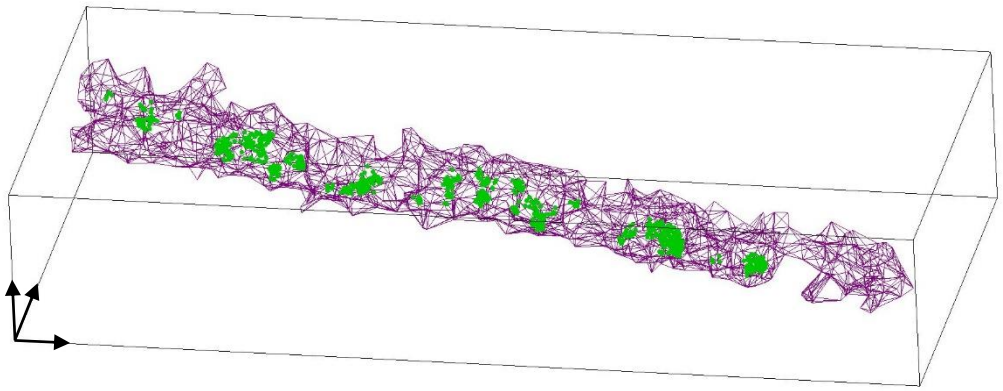
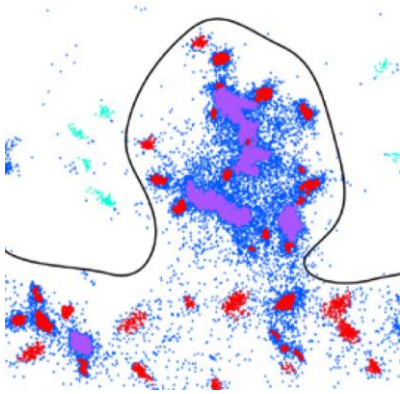
A guided tour for analyzing and quantifying single-molecule localization microscopy data

Part two: quantification strikes back!







Florian Levet

Quantitative Imaging of the Cell – Sibarita's team

Interdisciplinary Institute for Neuroscience

CNRS UMR 5297, University of Bordeaux

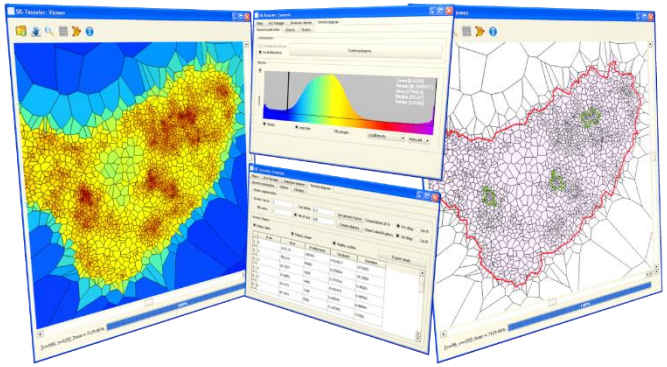
@f_levet

Part two's outline

- Intro
- Clustering methods
- Segmentation methods
- Interpretation of the quantifications
- Demos

Link to SR-Tesseler one-click installer (windows):
https://github.com/flevet/SR-Tesseler/releases

Link to datasets and slides:
https://github.com/flevet/NEUBIAS_Academy



The speaker and moderator team:

Siân Culley

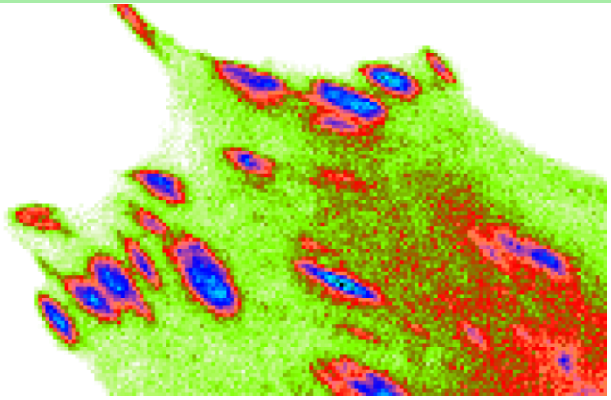
Daniel Sage

Bram van den Broek

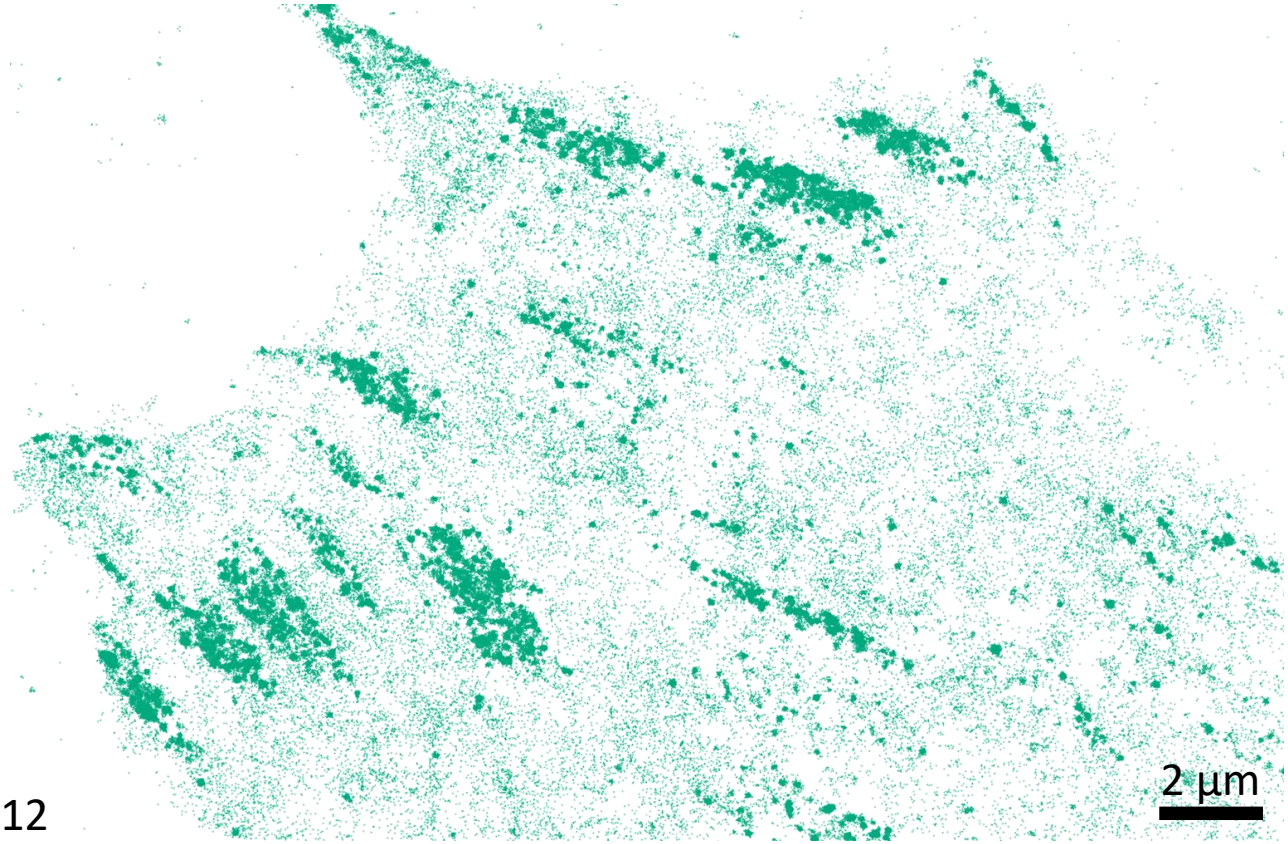
Pedro Matos Pereira

How to use the localizations ?

Fibroblast expressing integrin-β3-mEOS2

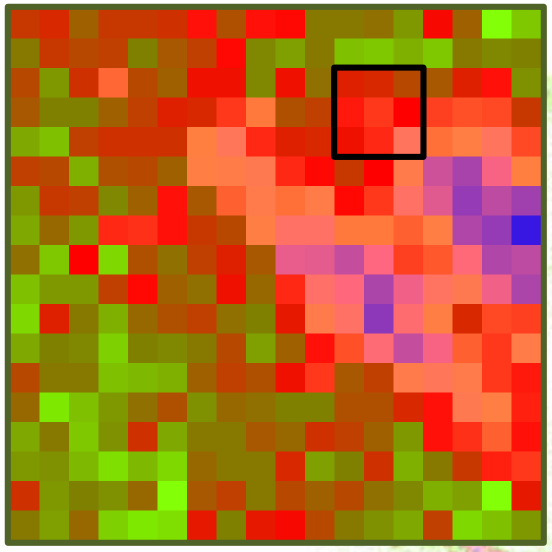


What is the organization of the adhesion sites ?

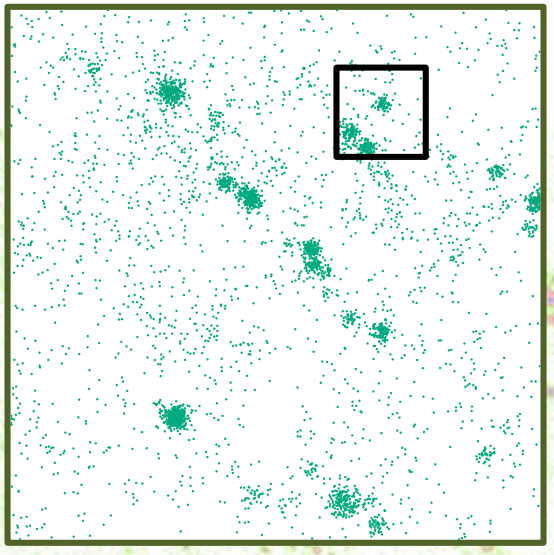


Rossier et al., 2012

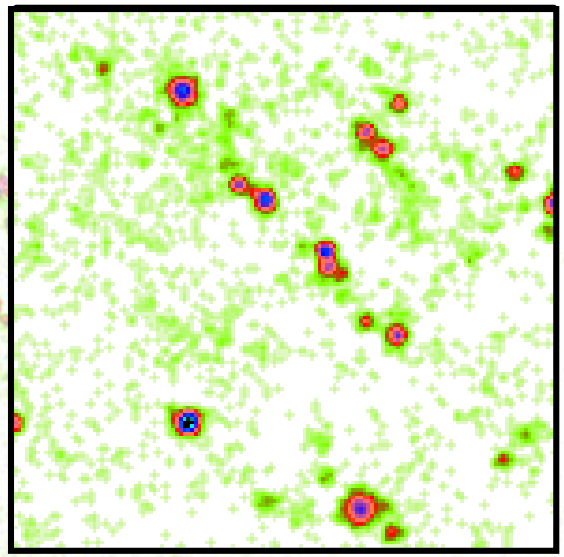
Image reconstruction in SMLM



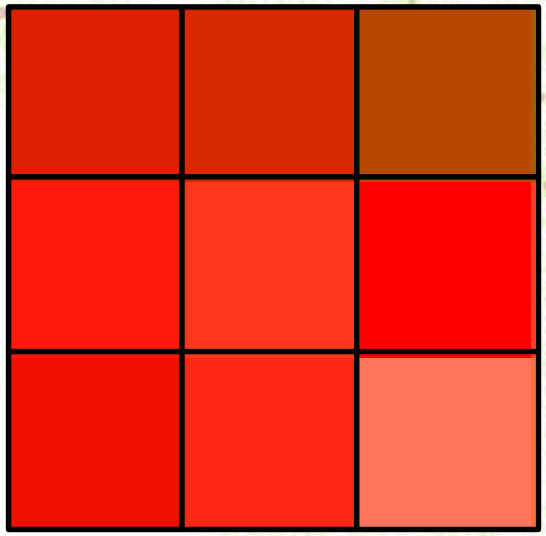
Pixel size = 160 nm

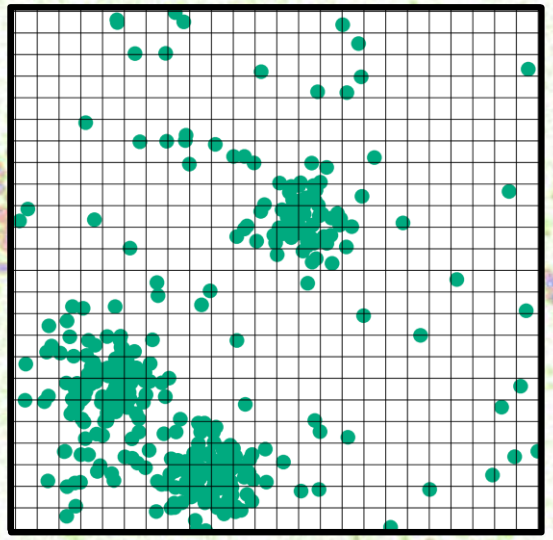


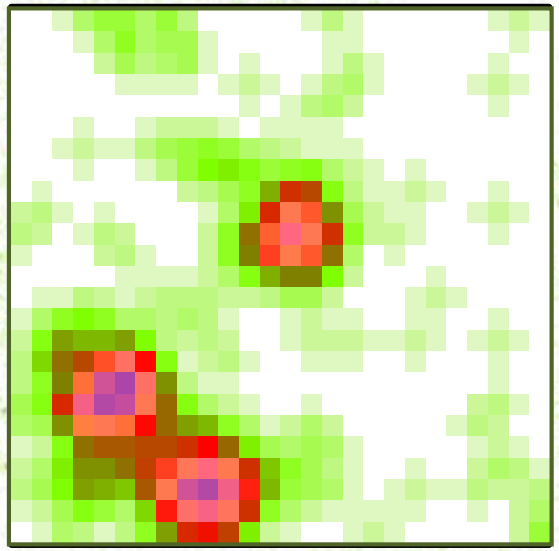
Sub-pixel localization



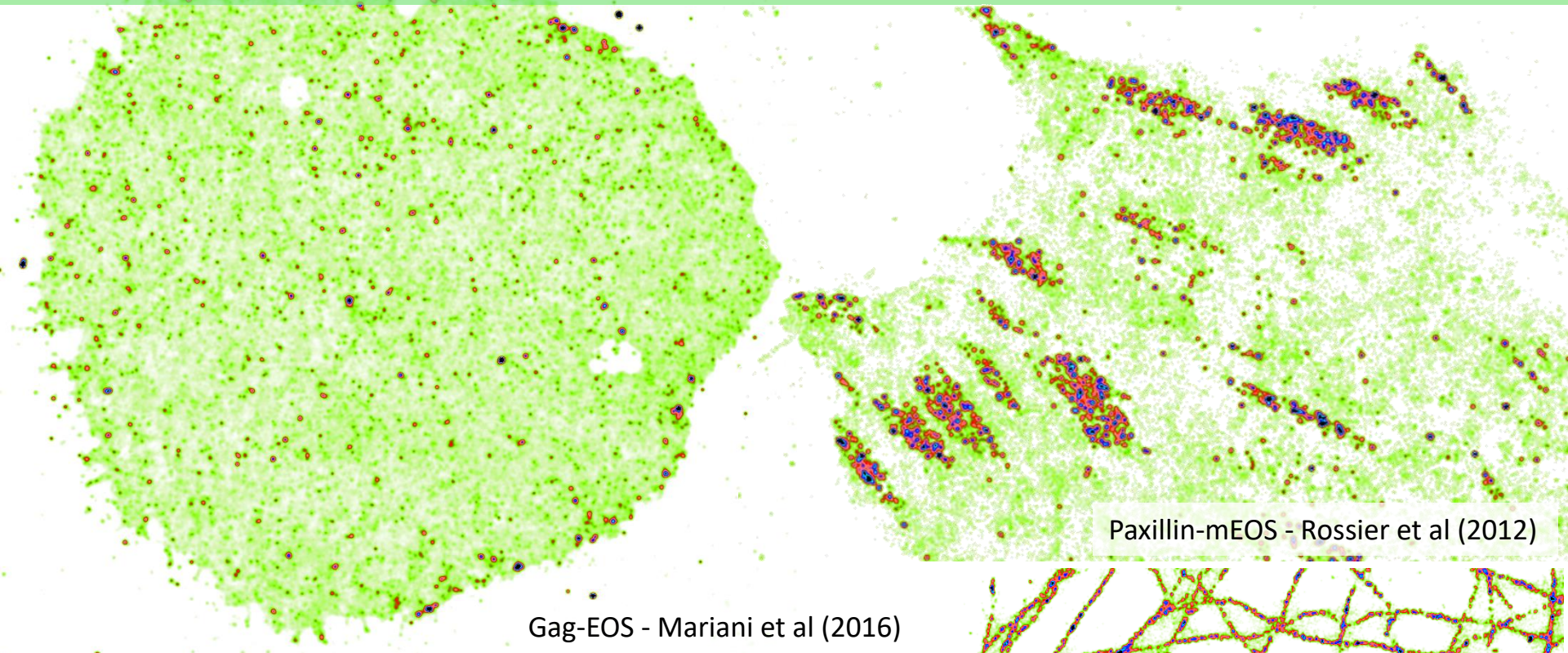
Projection





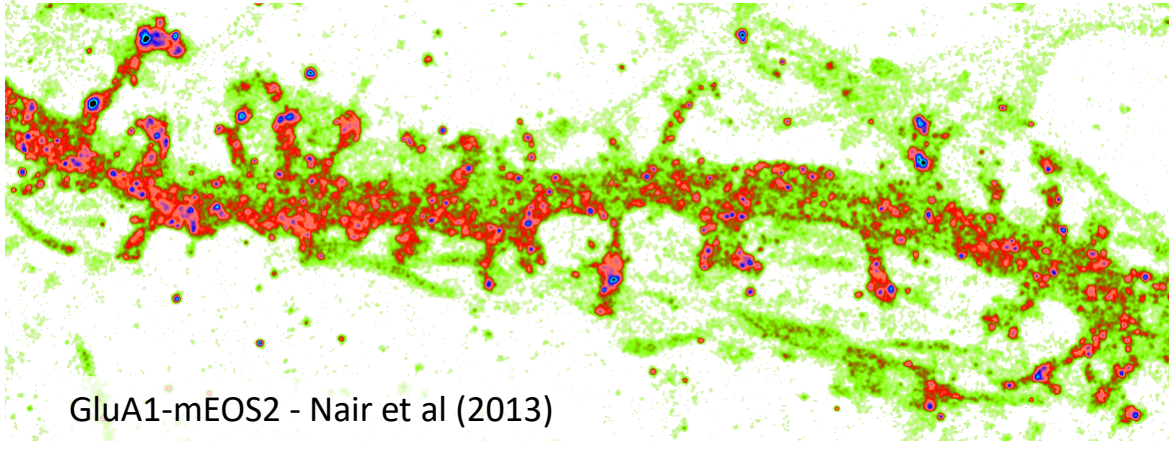


Why is pixel-based image quantification complex?

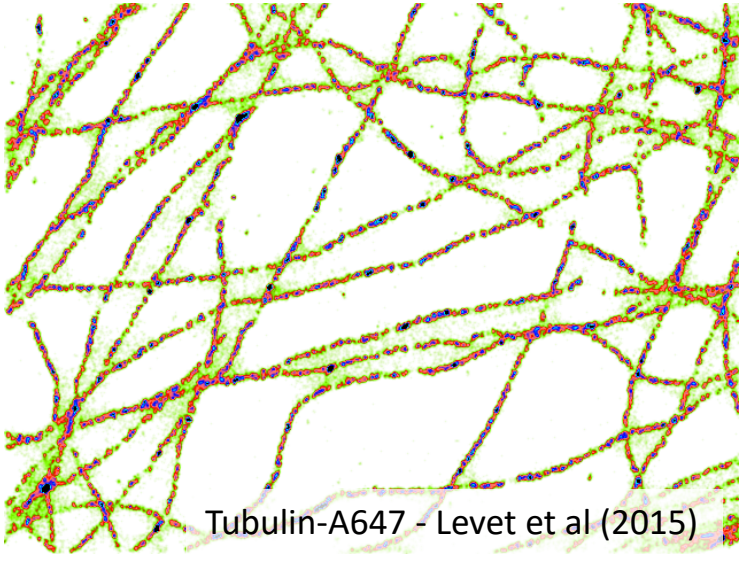


Paxillin-mEOS - Rossier et al (2012)

Gag-EOS - Mariani et al (2016)

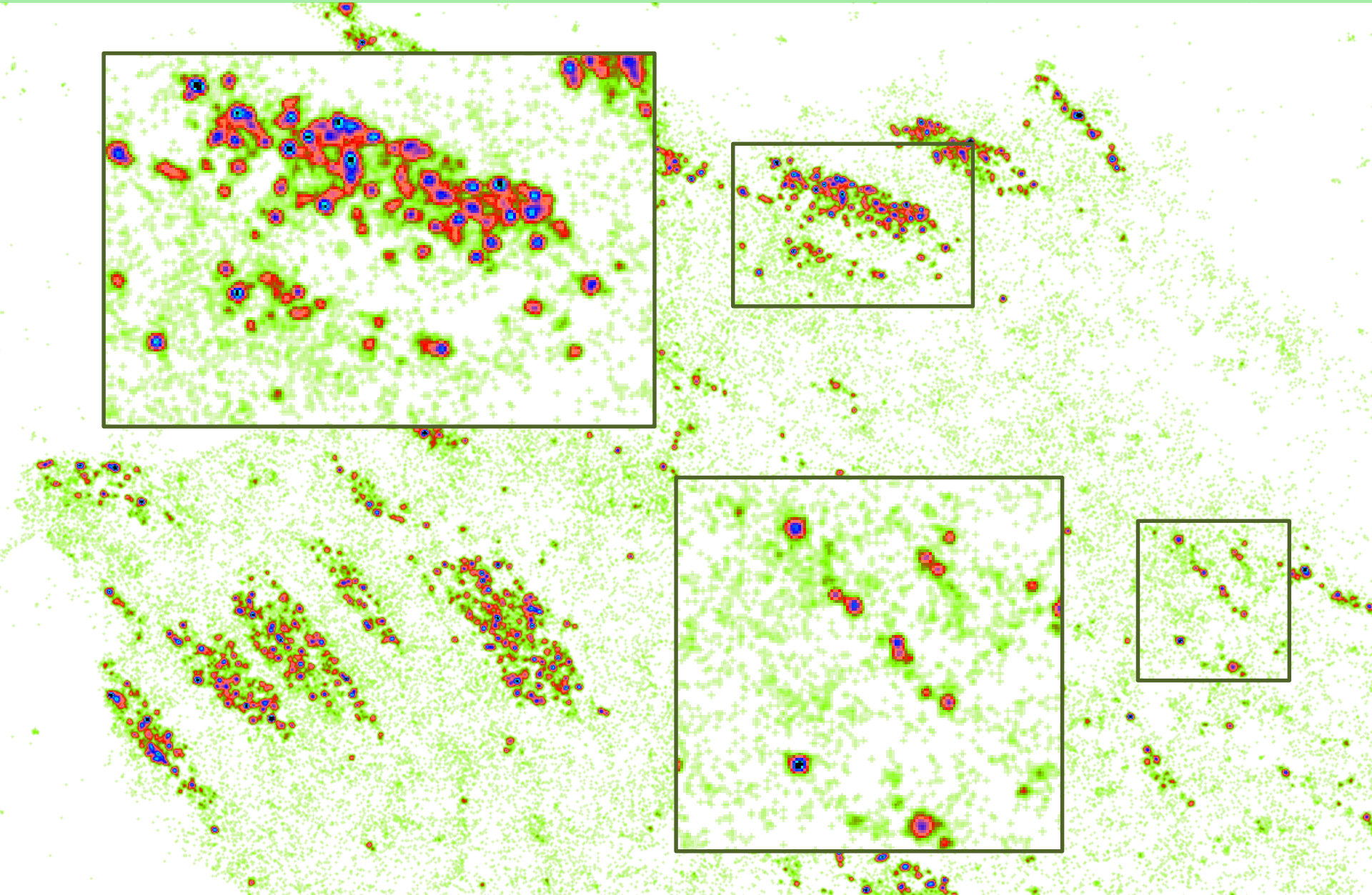


GluA1-mEOS2 - Nair et al (2013)



Tubulin-A647 - Levet et al (2015)

Why is pixel-based image quantification complex?

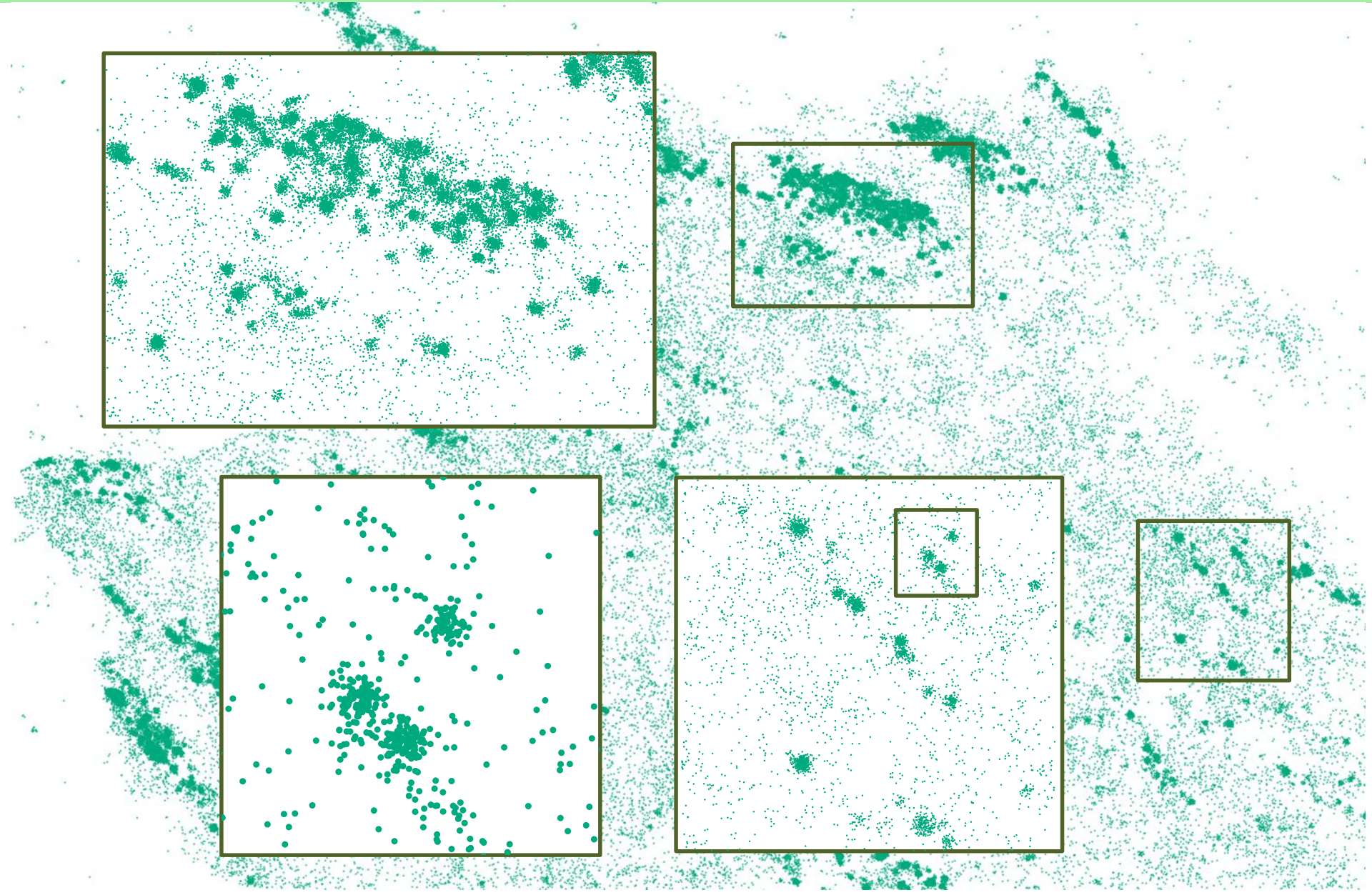


The difficulty of normalization

- Pixel size may affect the quantification
- Pixel dimension is fixed
  - Oversampling of the sparser regions
  - Undersampling of the denser ones
- Combination of different techniques
  - Complex to reproduce and generalize

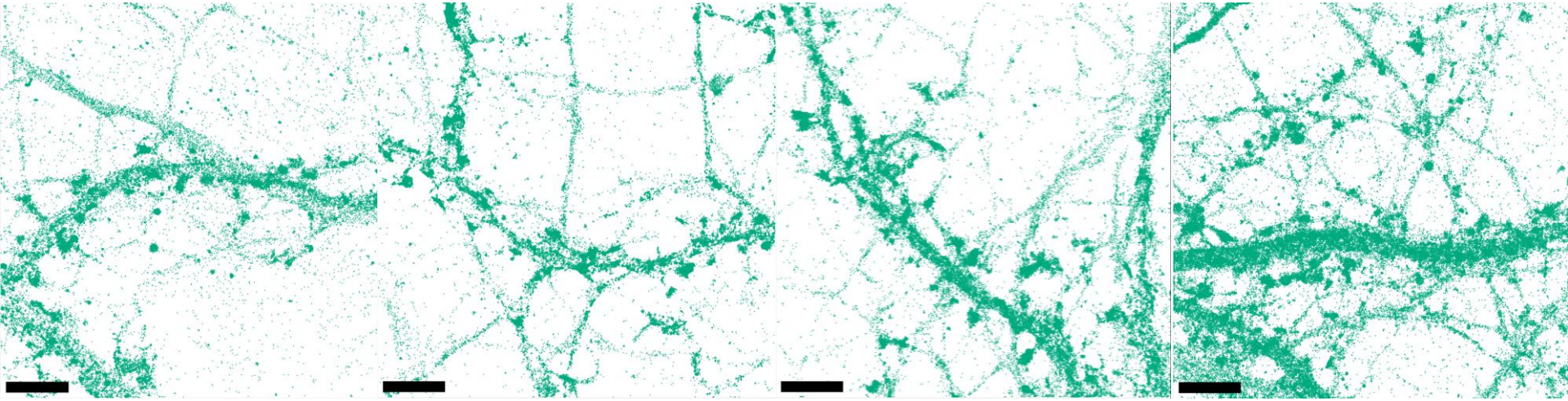
Baddeley et al. (2010)

How pointillistic nature of SMLM data can help?



Why is Quantification Complex?

- Experimental parameters
  - Fluorophore photophysics, labeling density
  - Acquisition time



1 1.43 2.78 4.49

Clustering VS Segmentation

- **Clustering**: statistically characterize small aggregates of molecules usually by comparison with a random distribution of the molecules (Ripley, PCF).

- **Segmentation**: classify molecules in different classes wrt some defined attributes (DBSCAN, Tesselation).

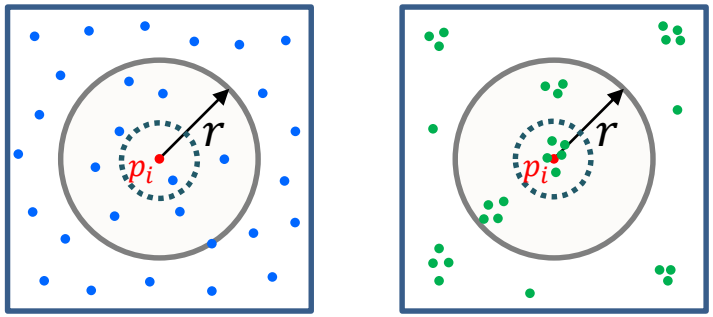
  Clustering is one application of segmentation.

Deschout et al. (2014), Nicovich et al. (2017)

K-Ripley function

- Describes the average number of molecules that exist near another molecule within different radius (***r***)
- Computes the deviations from spatially uniform distribution
- Computed iteratively for varying radius ***r***

$$K(r) = \frac{1}{n}\sum_{i=1}^{n} Np_i(r)/\lambda$$



- Statistically determines the radius of maximal aggregation of the molecules

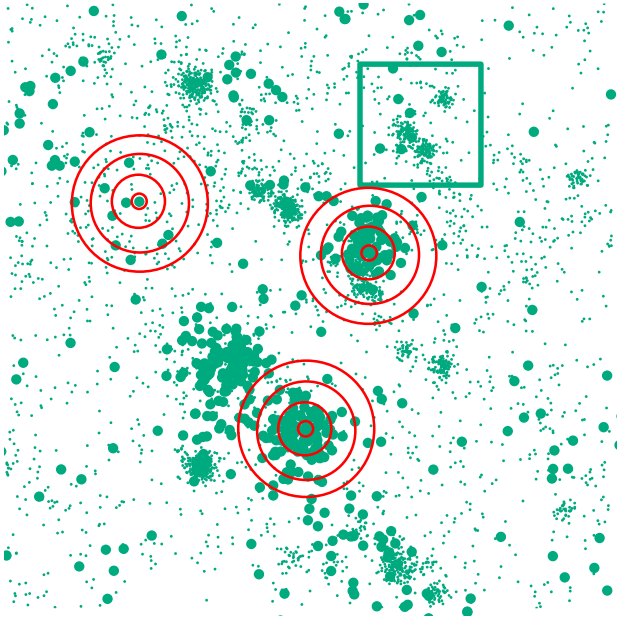
B. D. Ripley, *J Appl Probability* **13**, 255-66 (1976)

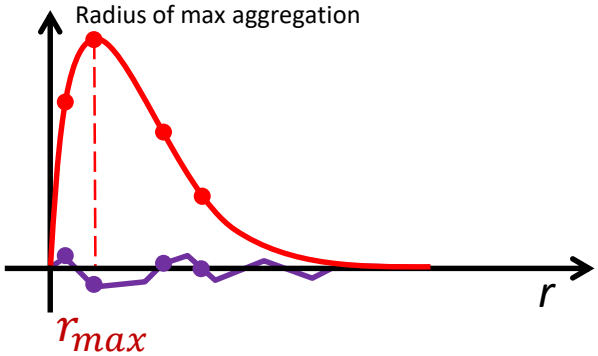
Penttinen et al. (2000) ; Lagache et al. (2013)

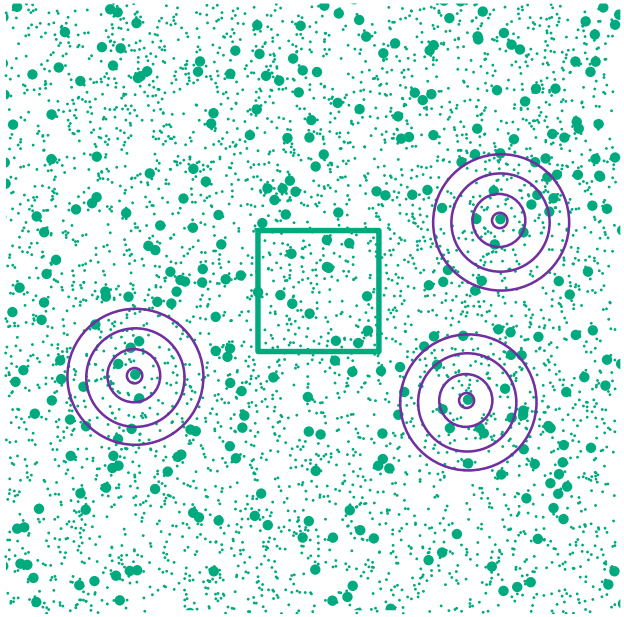
H-Ripley

- In the case of a random distribution $K(r) \propto \pi r^2$
- H-Ripley function:
  - Normalization of the K-Ripley function
  - $H(r) = 0$ for a normal distribution

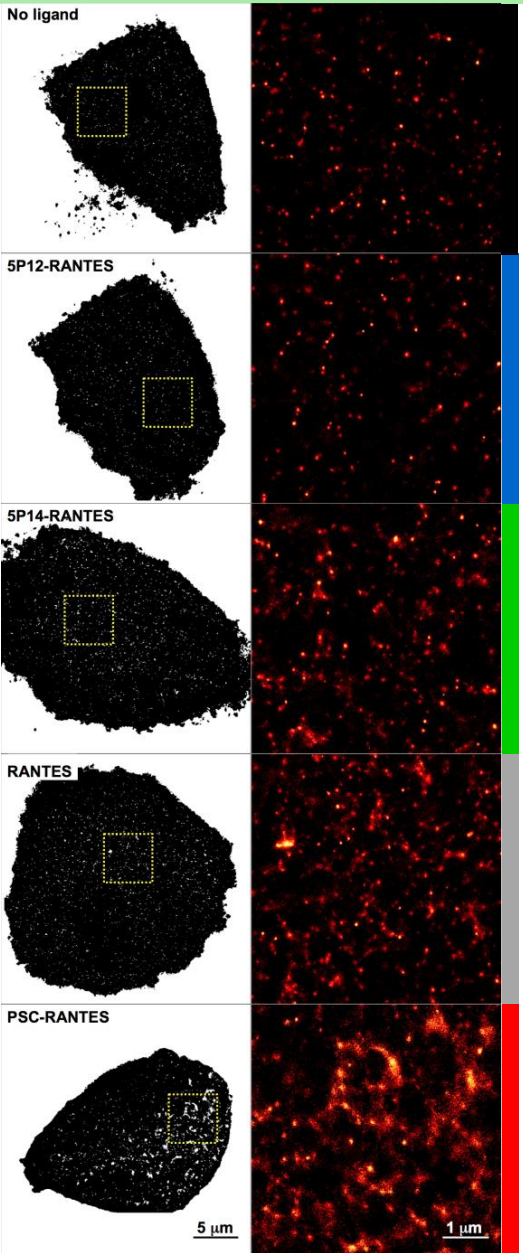
$$H(r) = \sqrt{K(r)/\pi} - r$$





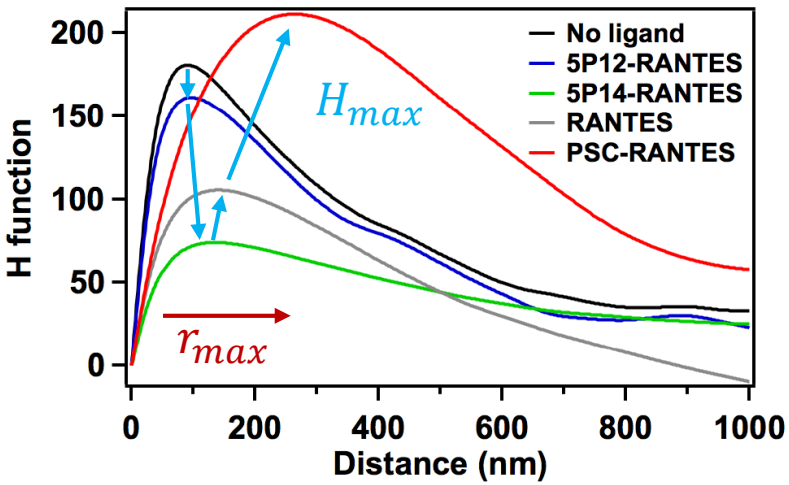


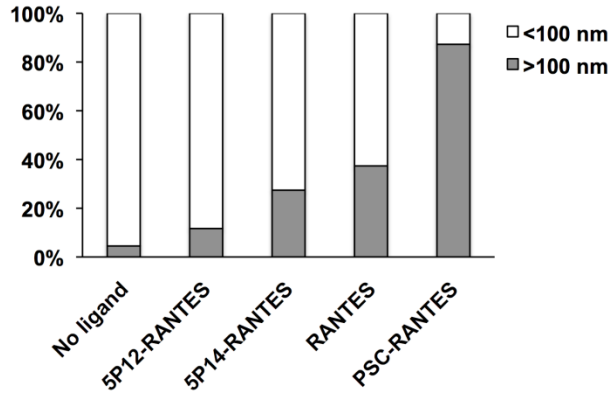
H-Ripley



Cluster Analysis with Ripley's H-Function

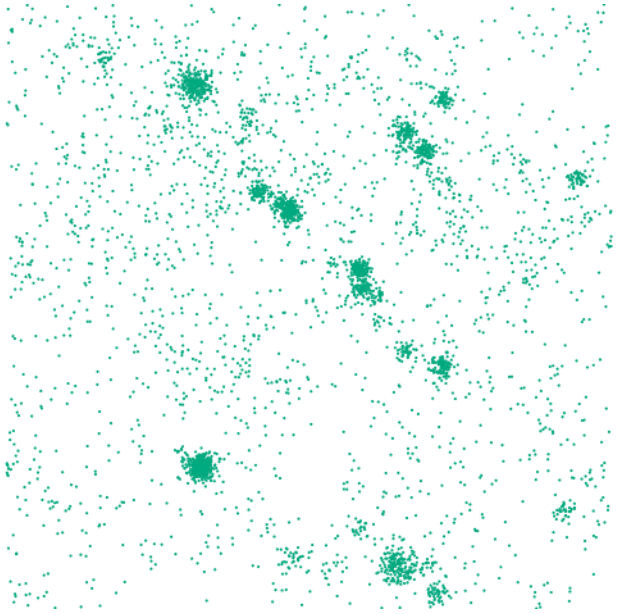
Arrestin2 clustering upon G protein-coupled receptor (CCR5) stimulation

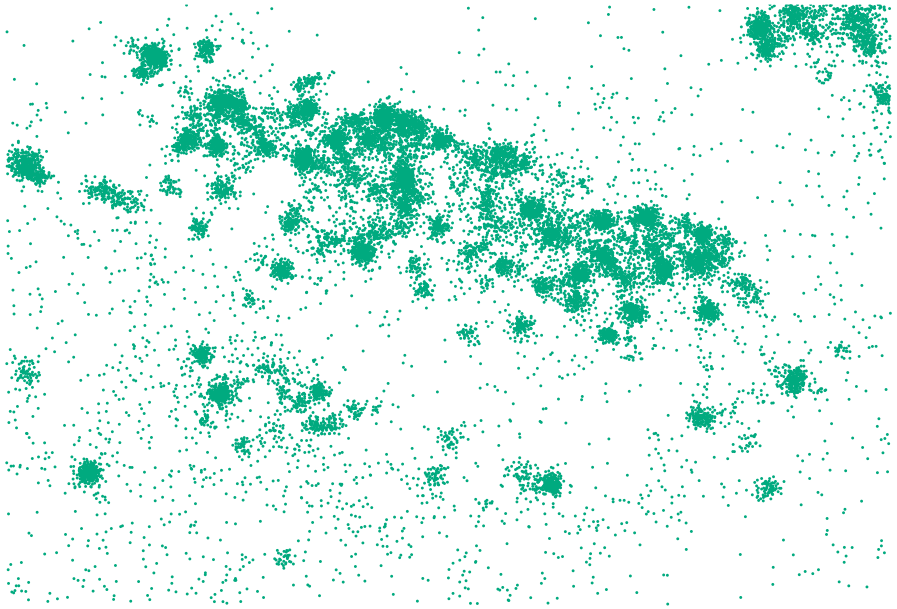


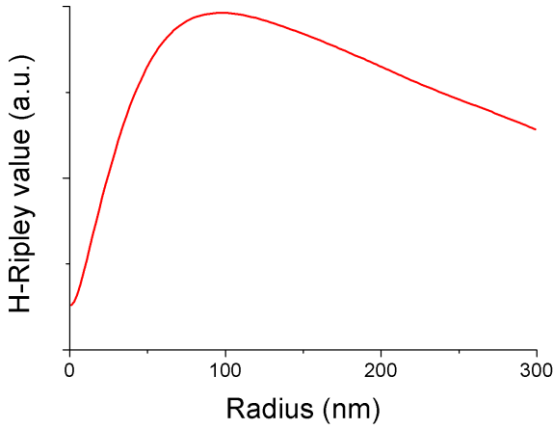


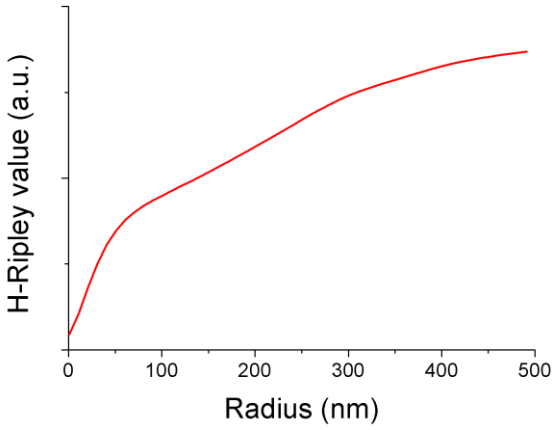
Truan et al., J Struct Biol 2013

H-Ripley






H-Ripley value (a.u.)
0
100
200
300
Radius (nm)

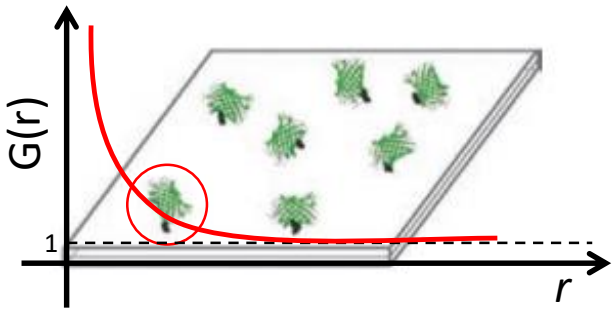

H-Ripley value (a.u.)
0
100
200
300
400
500
Radius (nm)

H-Ripley

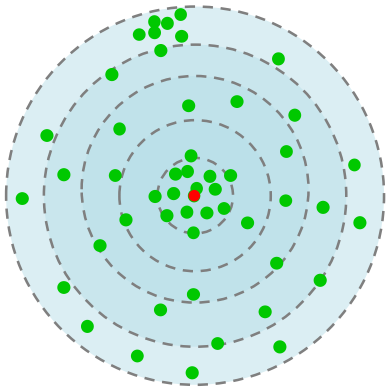
- Intrinsically normalized
- Only one value
  - Radius of maximum aggregation
- Sensitive to multiple levels of organization

Pair-correlation

- Protection against over-estimation of clustering
  - Multiple-blinking of fluorophores
  - Over-counting doesn't propagate to higher length-scales

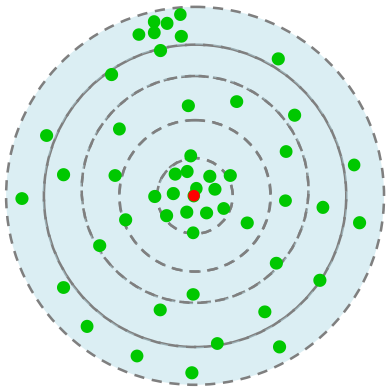


K-Ripley



12	17	26	34	51

Pair-correlation

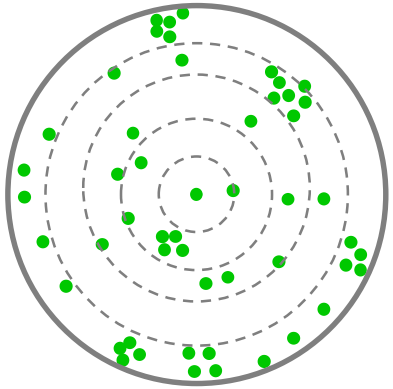


12	5	9	8	17

Sengupta et al. (2011), Veatch et al. (2012)

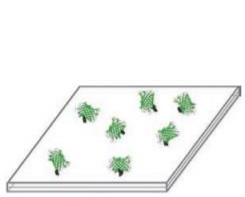
Pair-correlation

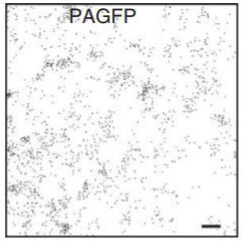
- Protection against over-estimation of clustering
  - Multiple-blinking of fluorophores
  - Over-counting doesn't propagate to higher length-scales

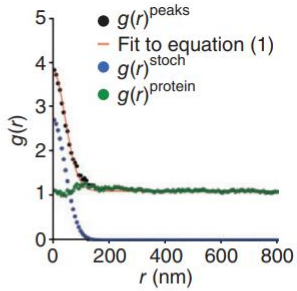


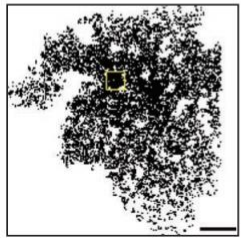
Cluster radius (ξ)
Molecules / cluster ($N^{cluster}$)
Cluster density (ψ)

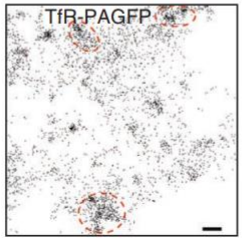
$$g(r)^{peaks} = \frac{1}{4\pi{\sigma_s}^2\rho^{average}}\exp\left(\frac{-r^2}{4{\sigma_s}^2}\right) + \left(A\exp\left(\frac{-r}{\xi}\right)+1\right) * g(r)^{PSF}$$

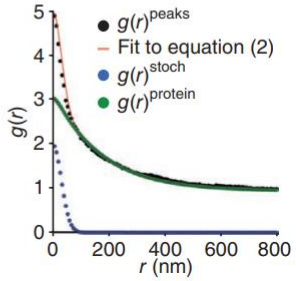










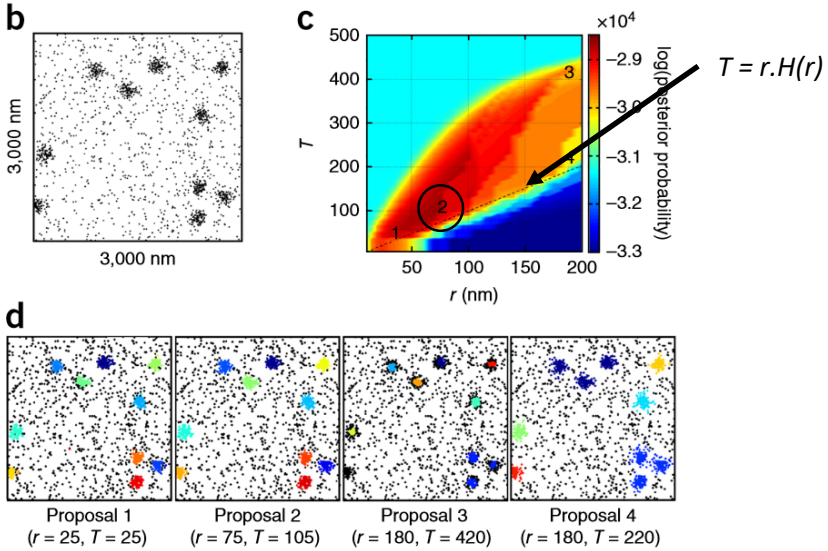


Sengupta et al. (2011), Veatch et al. (2012)

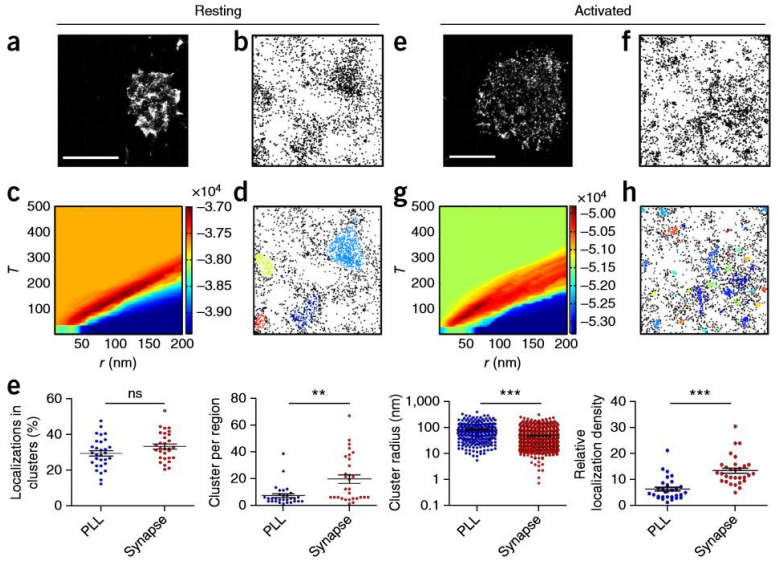
Bayesian clustering

Rubin-Delanchy et al. (2015)

- Model-based clustering
  - Prior distribution → user supplied cluster radii
  - Posterior → probability of any given assignment of localizations to clusters (cluster proposal)
  - Background if H(r) < T (threshold)



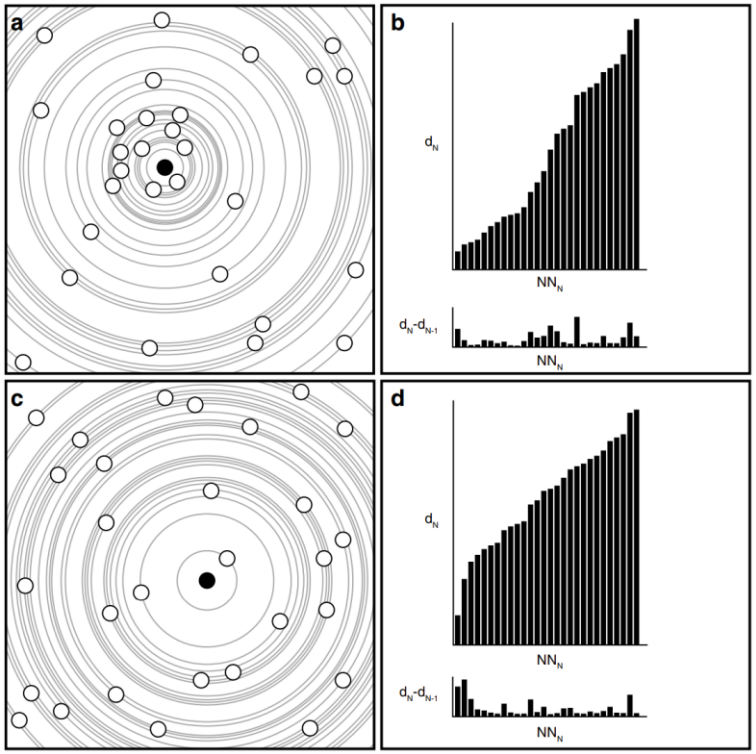
Clustering of CD3ζ-mEos3.2



Machine-learning

Williamson et al. (2020)

- Use difference in distances between consecutive neighbors
  - Transform the original coordinates
- Train neural networks on simulations



Input 100 × 1

↓

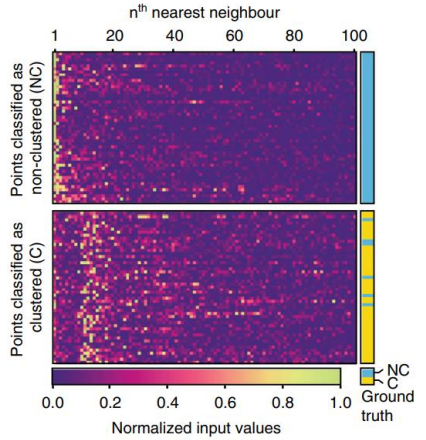
Input: 100 × 1
Dense (32 layers)
Output: 100 × 32

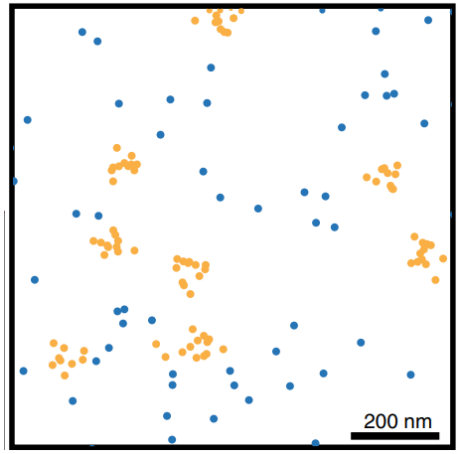
↓

Input: 100 × 32
Flatten
Output: 3200

↓

Input: 3200
Dense
Output: 1



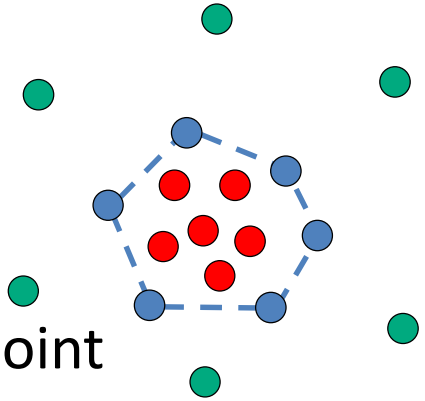


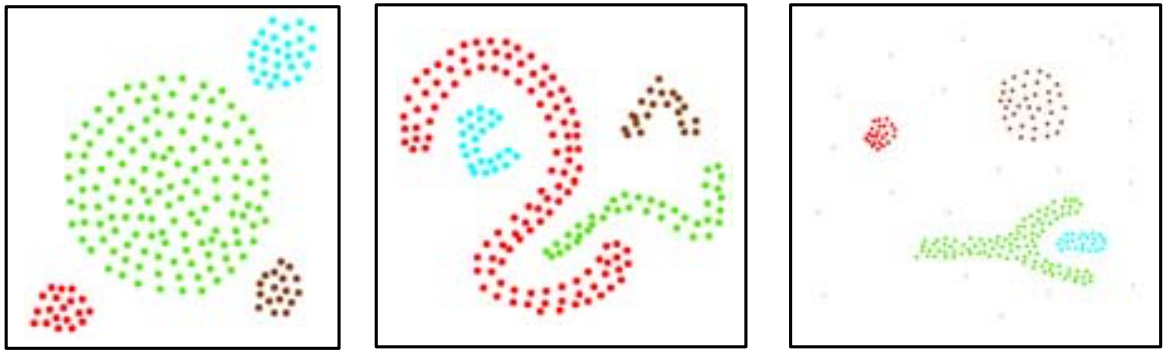
DBSCAN

Density-Based Spatial Clustering Analysis with Noise

- Segmentation technique
  - Suitable for objects and clusters
- Organizes the localizations wrt density in 3 classes
  - Core, density-reachable or outlier points
- Two parameters
  - Radius ***r*** → neighborhood size
  - ***MinPts*** → min nb of points in ***r*** to be a core point

  with ***MinPts*** ≠ min number of points of a cluster

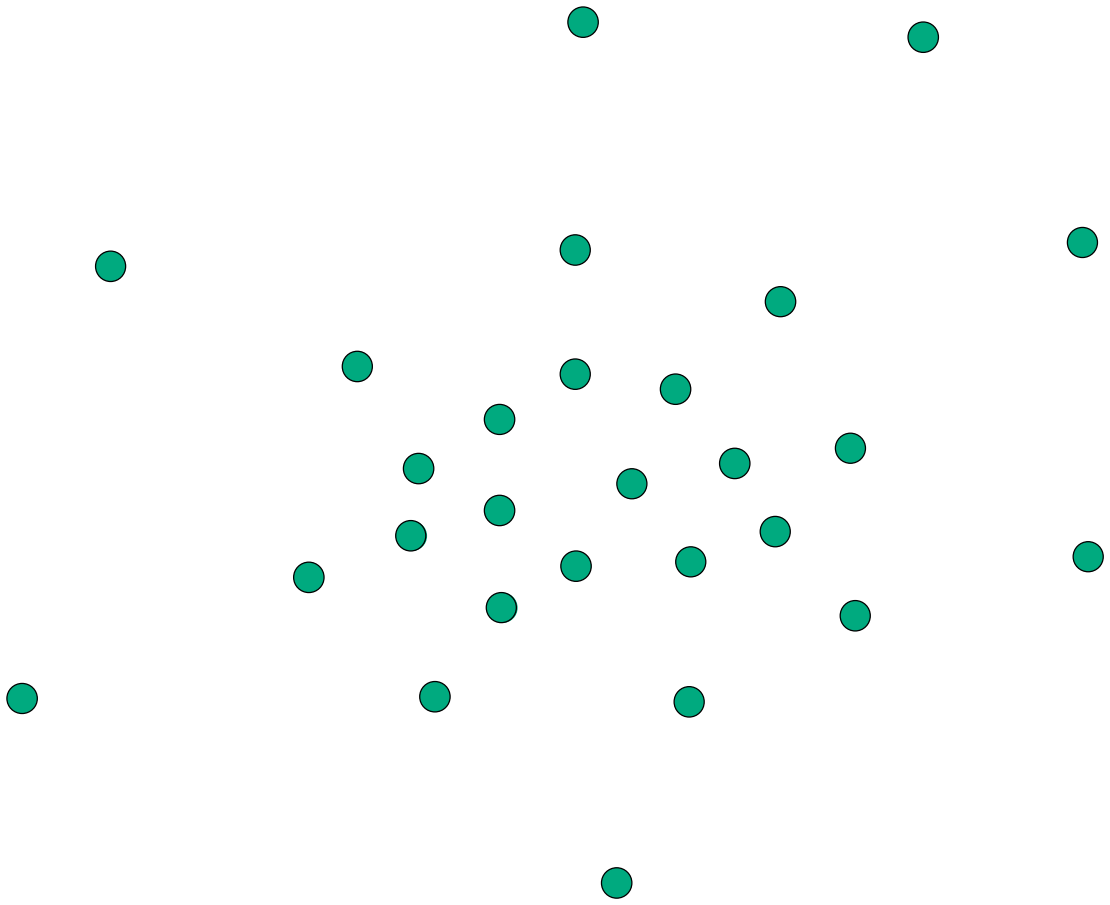




Ester et al. (1996)

DBSCAN

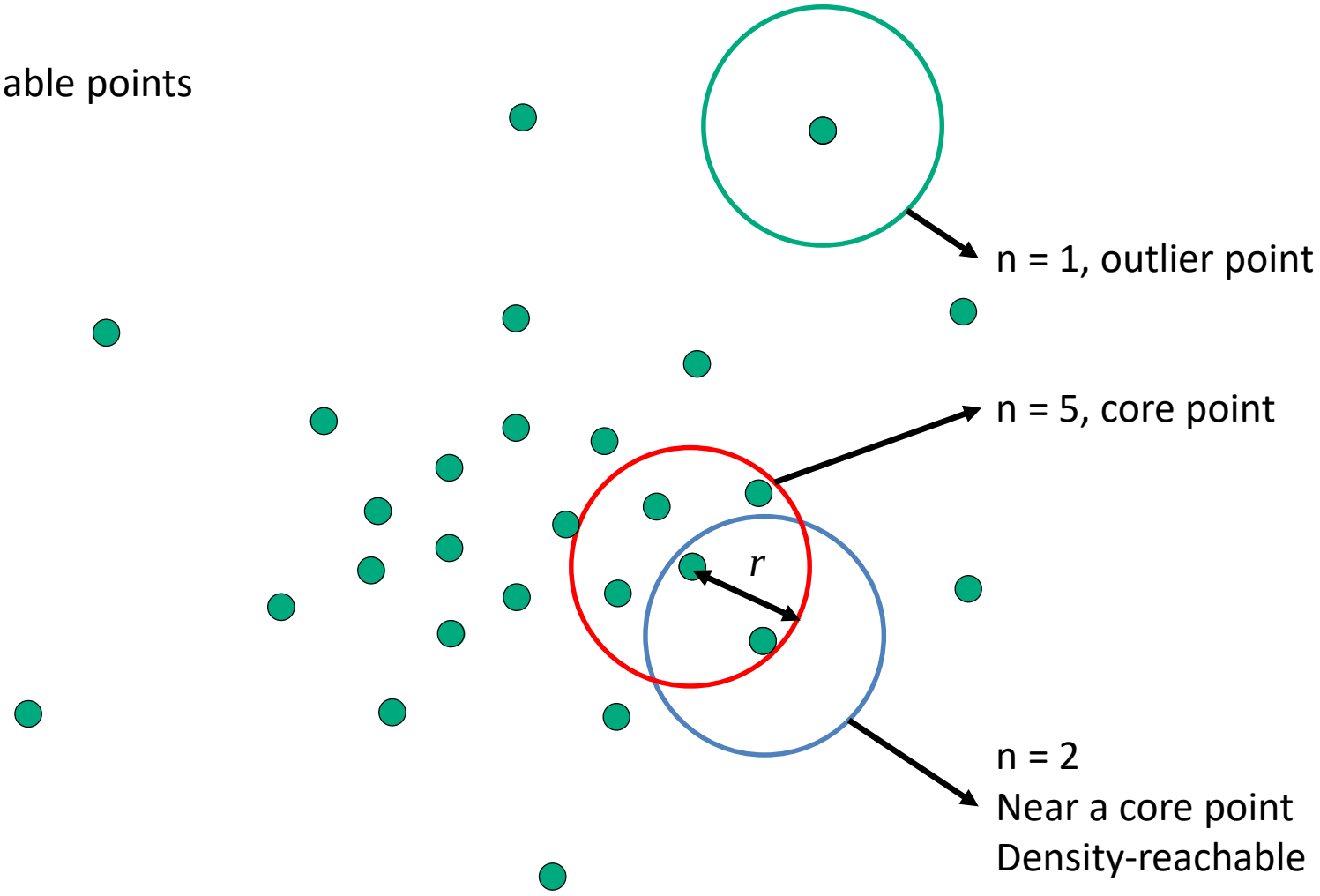
MinPts = 4 (including current)



DBSCAN

MinPts = 4 (including current)

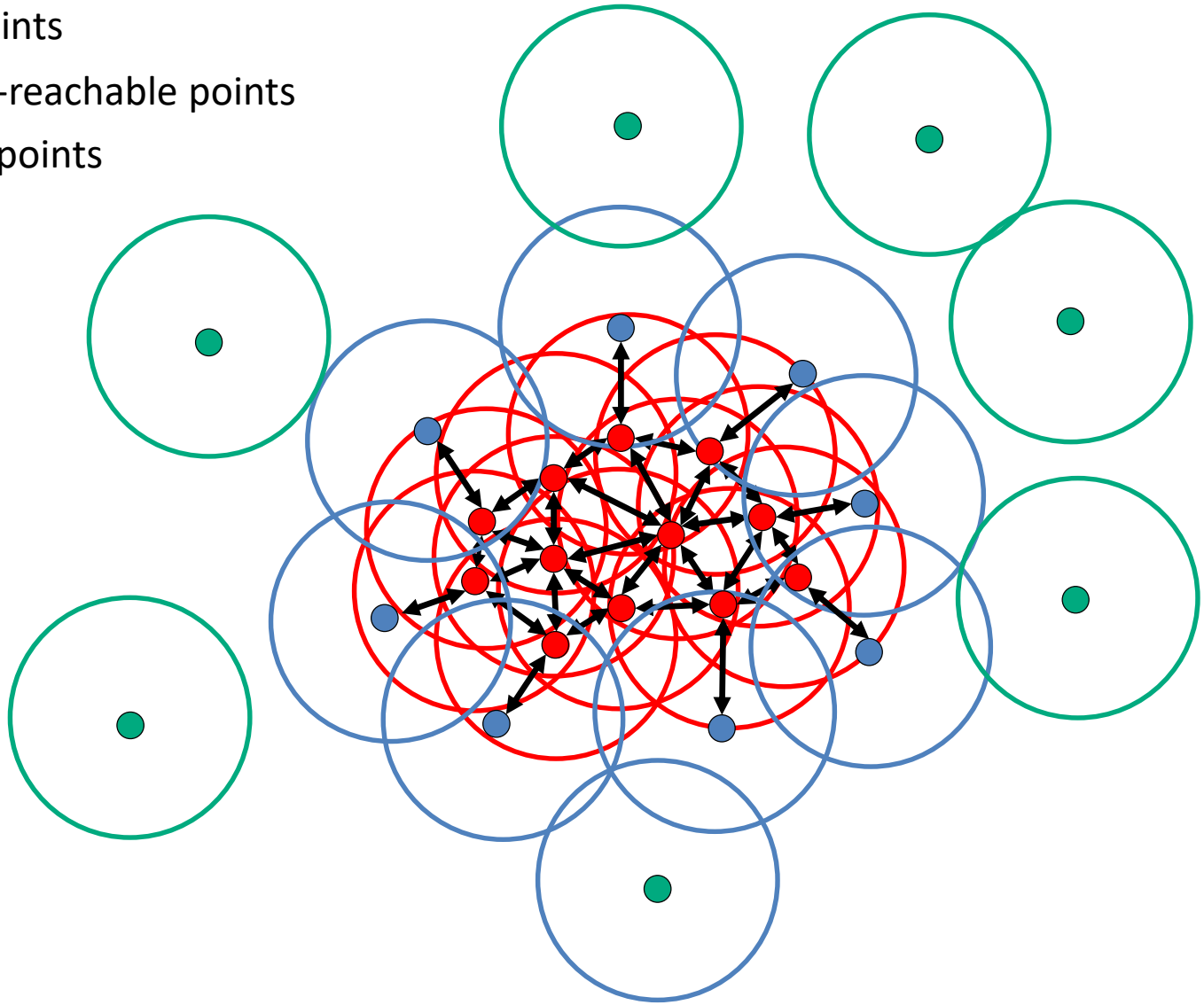
- Core points
- Density-reachable points
- Outlier points



DBSCAN

MinPts = 4 (including current)

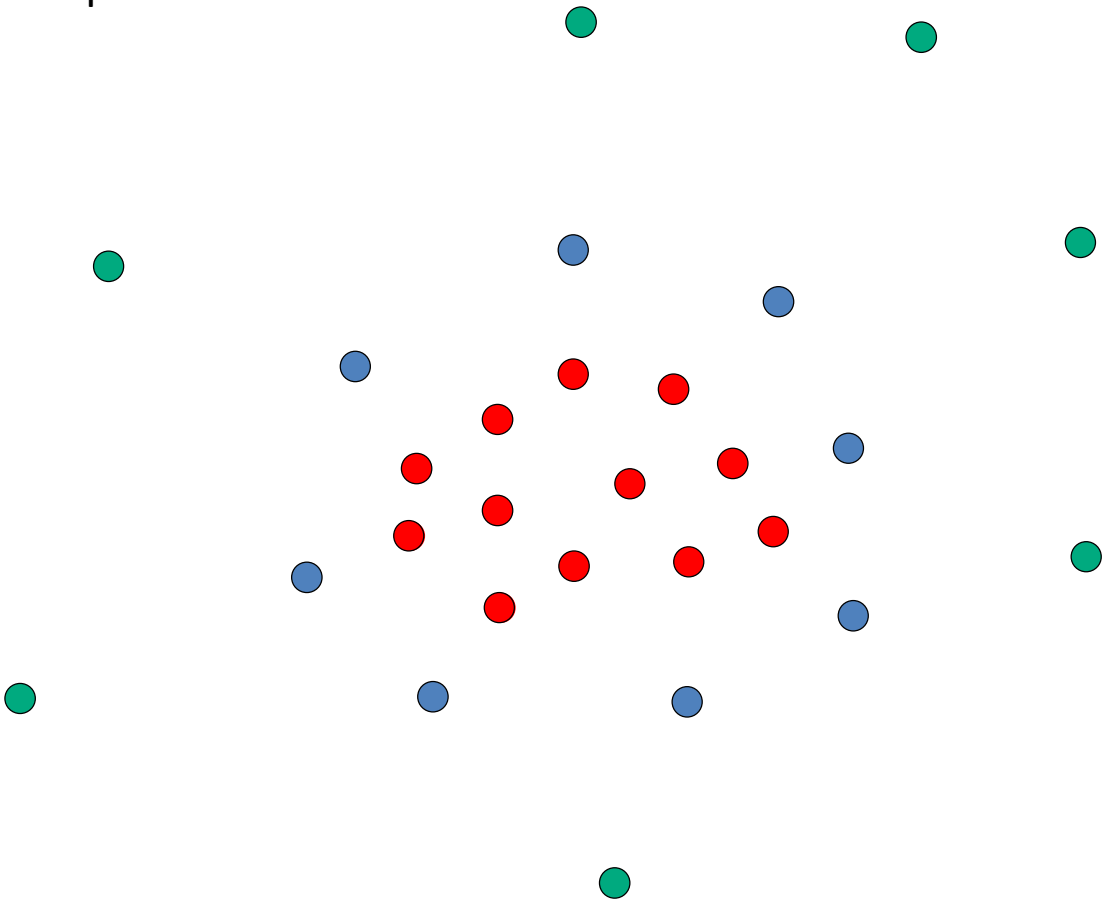
- Core points
- Density-reachable points
- Outlier points



DBSCAN

MinPts = 4 (including current)

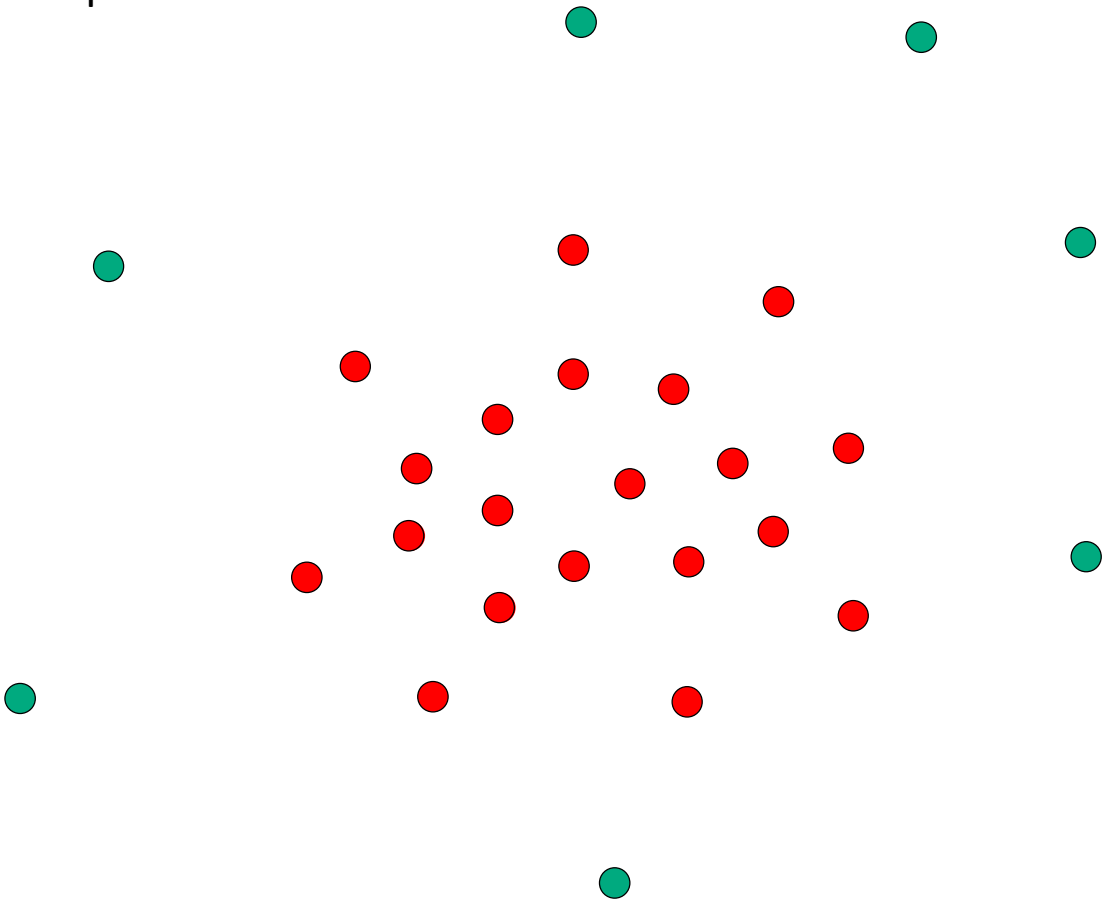
- Core points
- Density-reachable points
- Outlier points



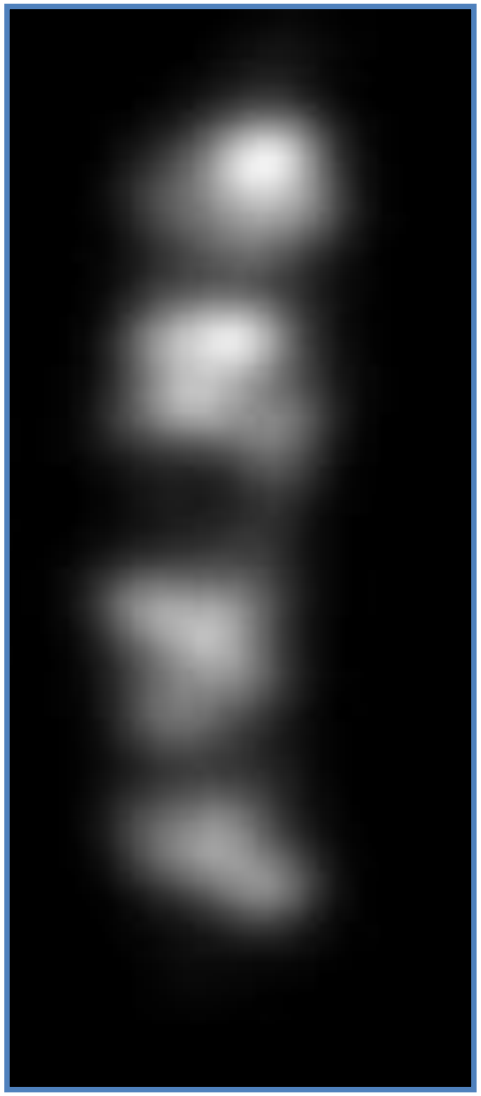
DBSCAN

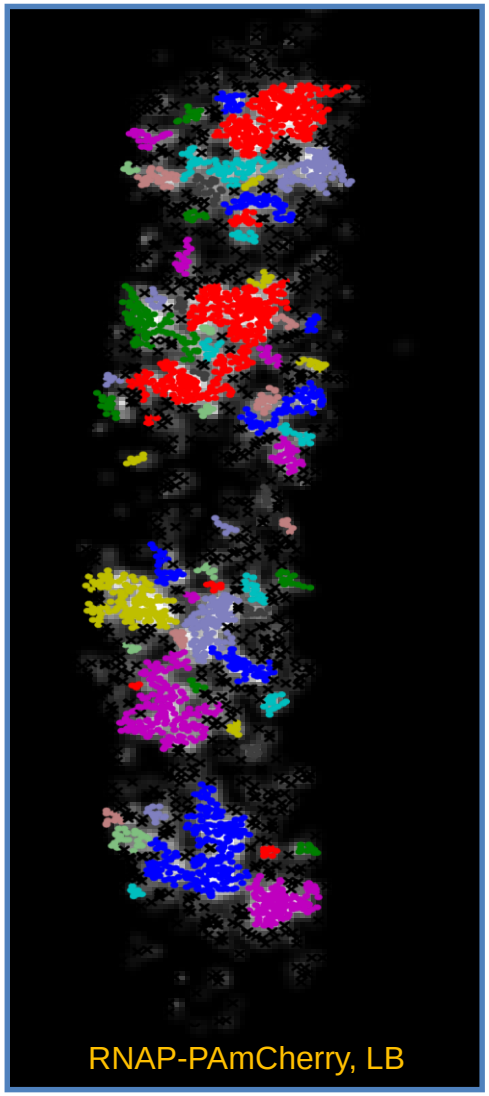
MinPts = 4 (including current)

- Core points
- Density-reachable points
- Outlier points



DBSCAN





***3288** detected molecules*

Endesfelder et al, 2013

Counting proteins in bacteria

- Stoichiometric labelling by integrating paFP into genome
- PAmCherry: low probability of repetitive blinking in *E. coli*
- Detection efficiency: ~ 50 % (calibrated by biochemical data)
- **~ 3600 copies of RNA polymerase** (detected) (LB medium)

LB medium

90-100 RNAP per cluster

6.1 clusters per chromosome

82.3 ± 15.4 nm cluster radius

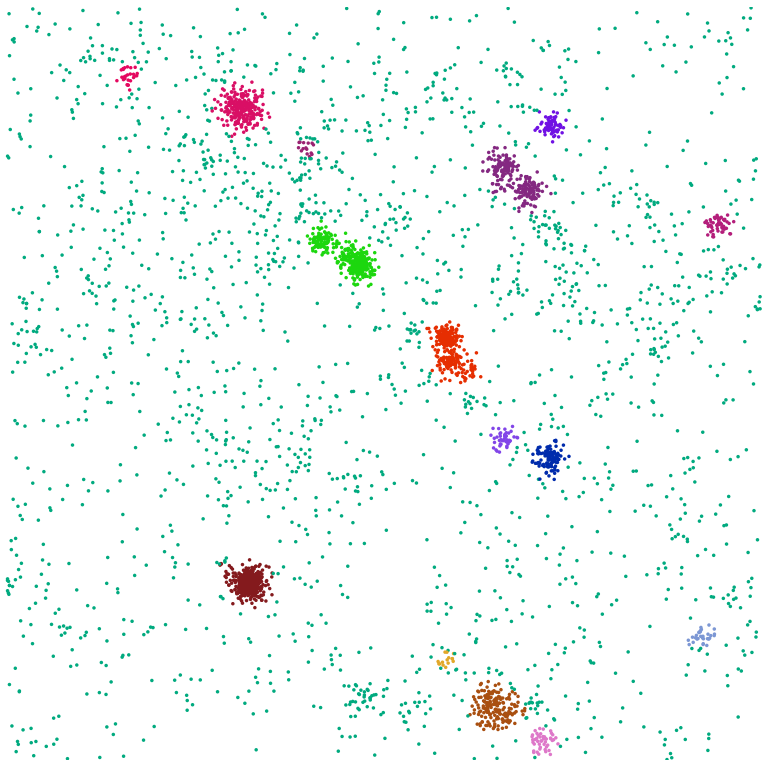
~ 35% of RNAP are found in large clusters

DBSCAN

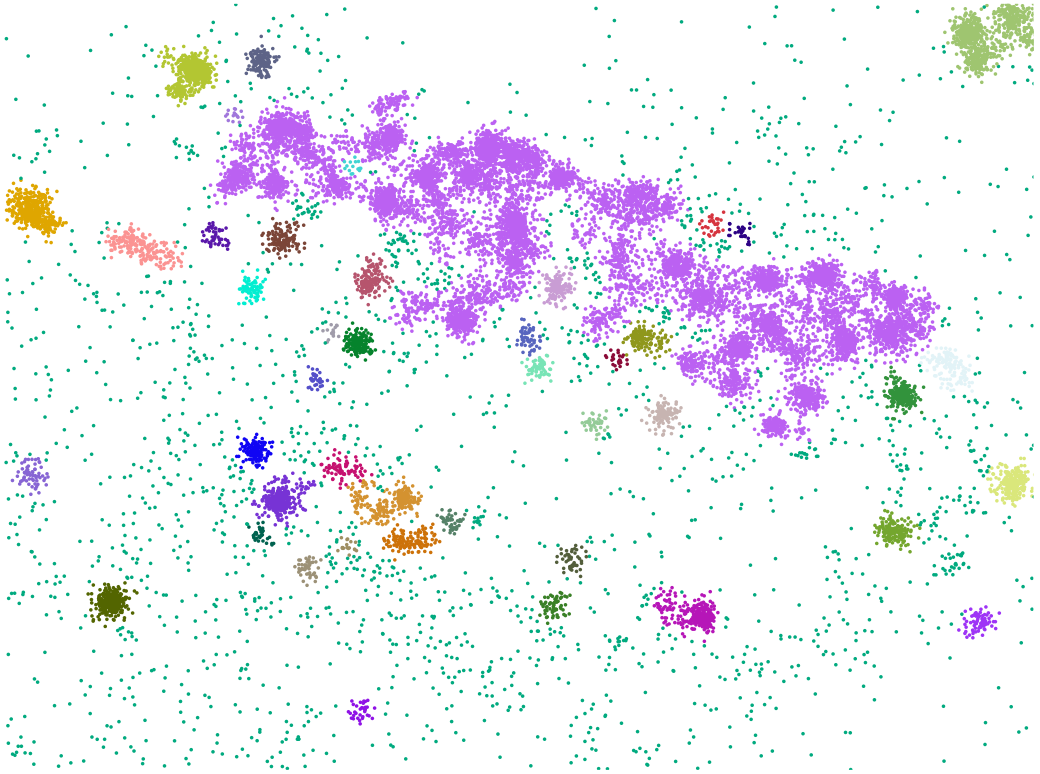
r = 35 nm

$MinPts$ = 15

Density = 400 mol/µm²



Density = 1900 mol/µm²

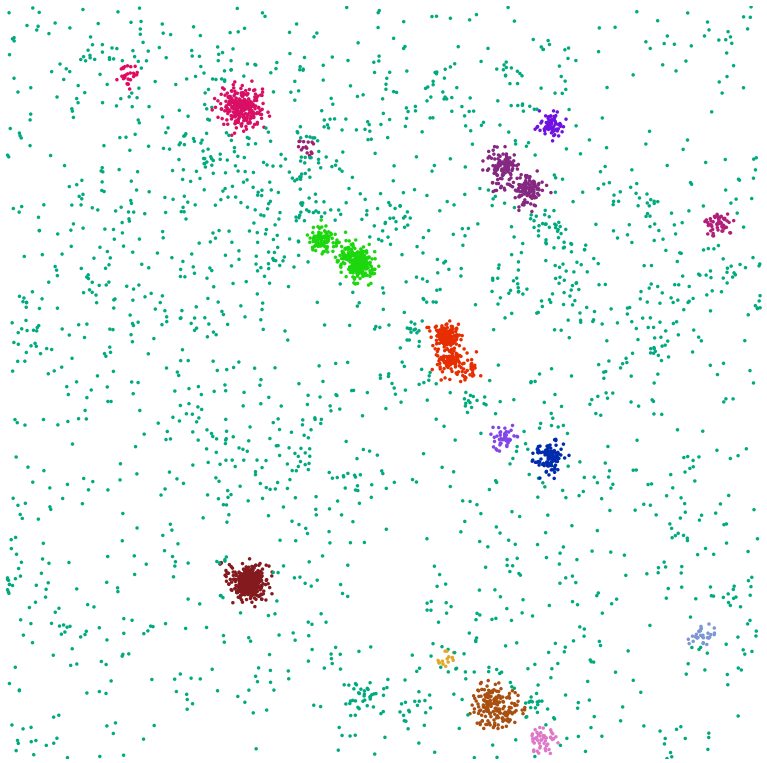


DBSCAN

r = 35 nm

$MinPts$ = 15

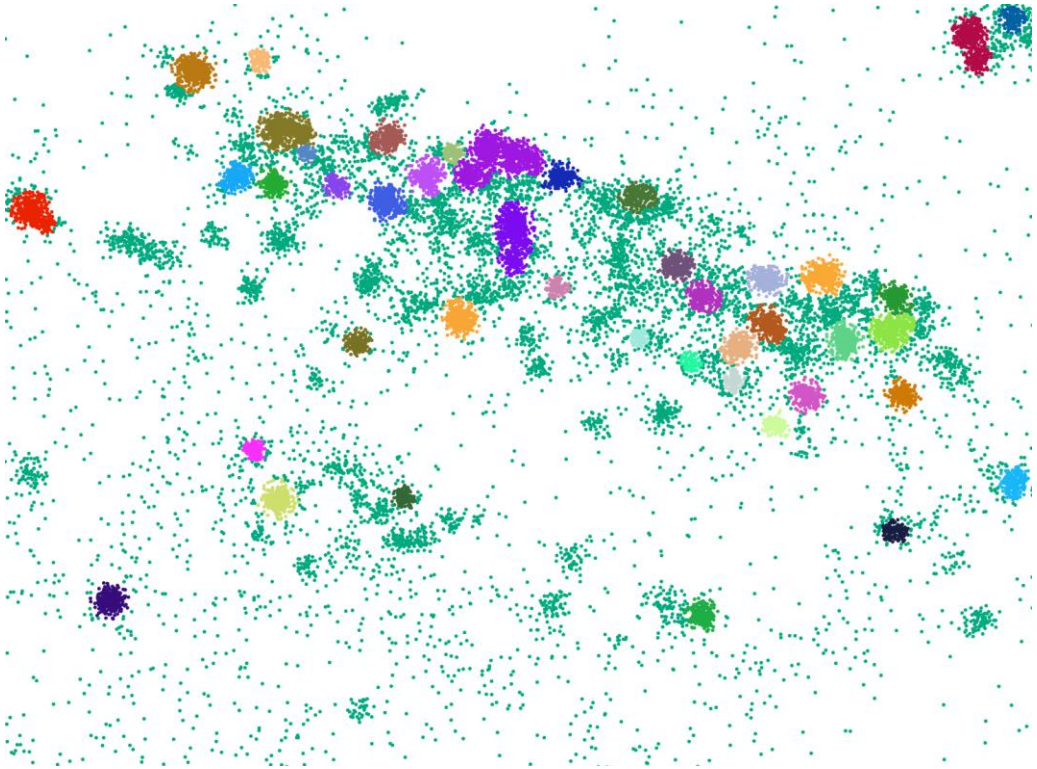
Density = 400 mol/µm²



r = 35 nm

$MinPts$ = 80

Density = 1900 mol/µm²

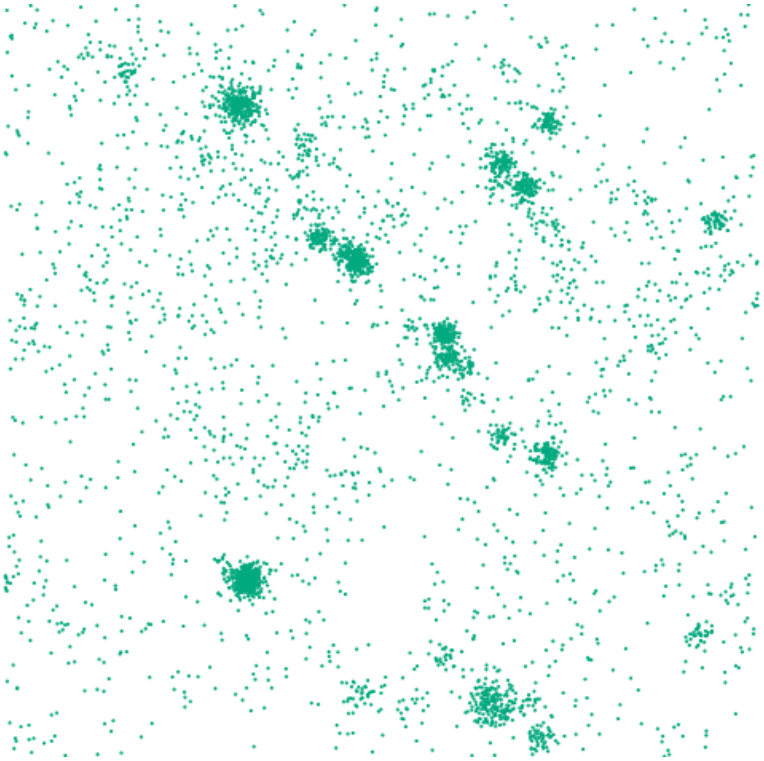


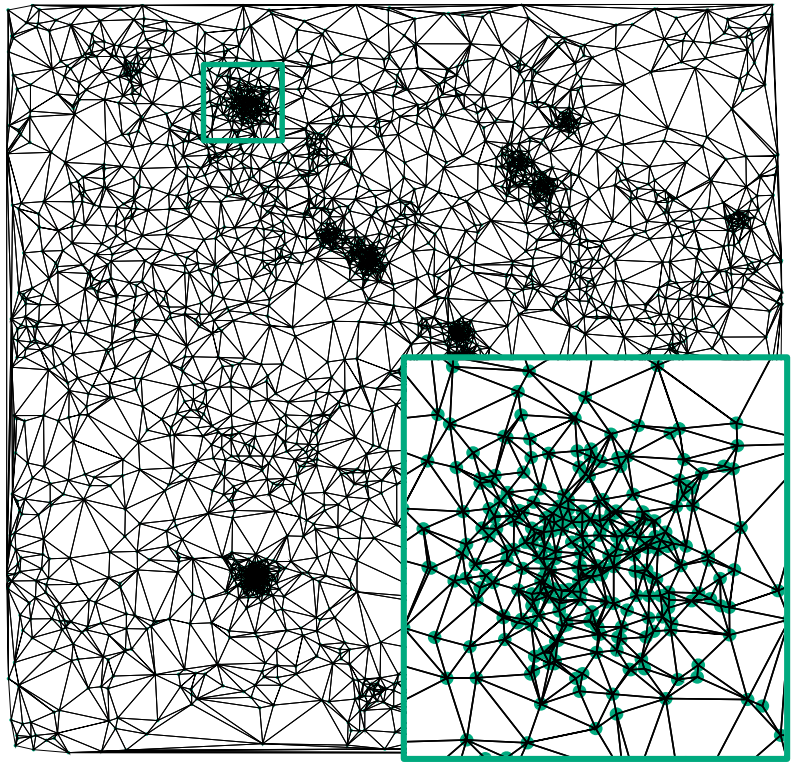
DBSCAN

- Two parameters
- Not normalized
  - Need to change the parameters if the density changes

Tesselation-based techniques

- Delaunay triangulation
  - Subdividing-space technique using triangles constructed from molecule coordinates

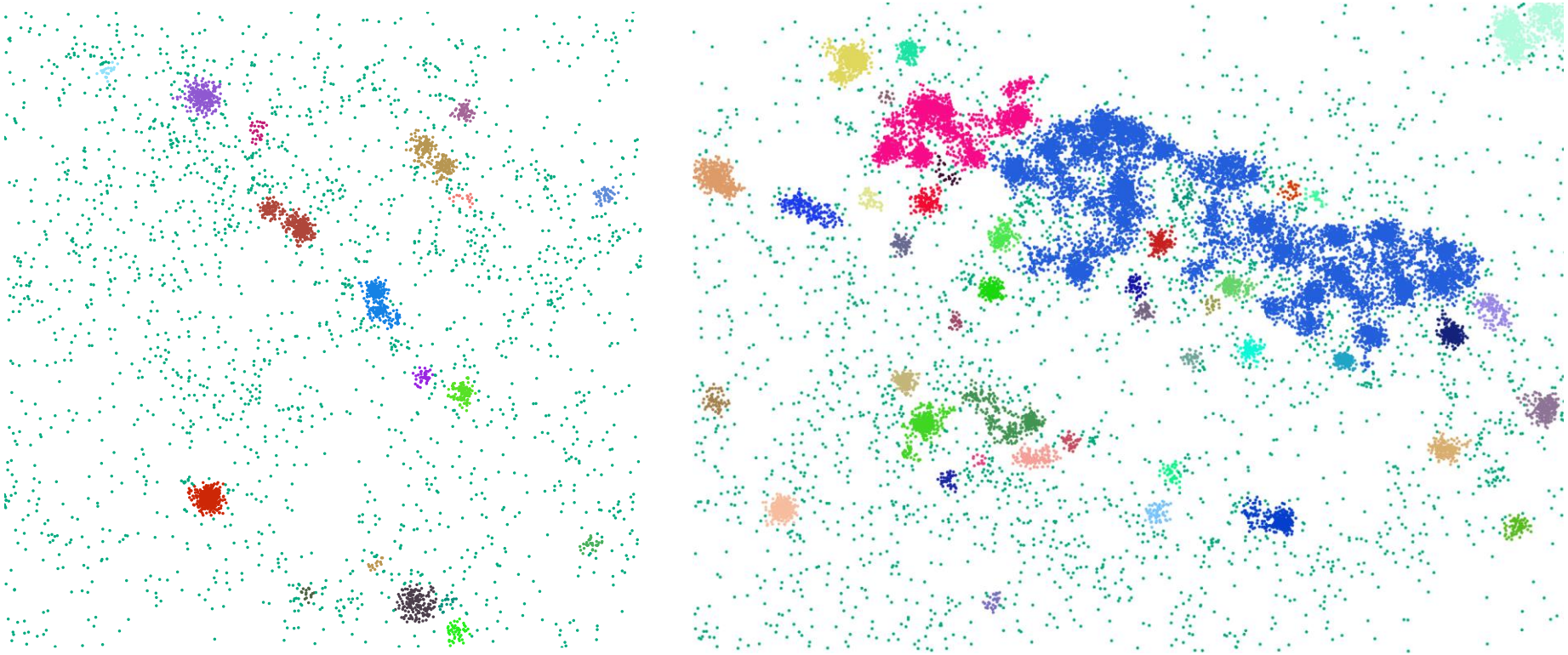




Tesselation-based techniques

- Global definition of a cut distance d_c

d_c = 25 nm



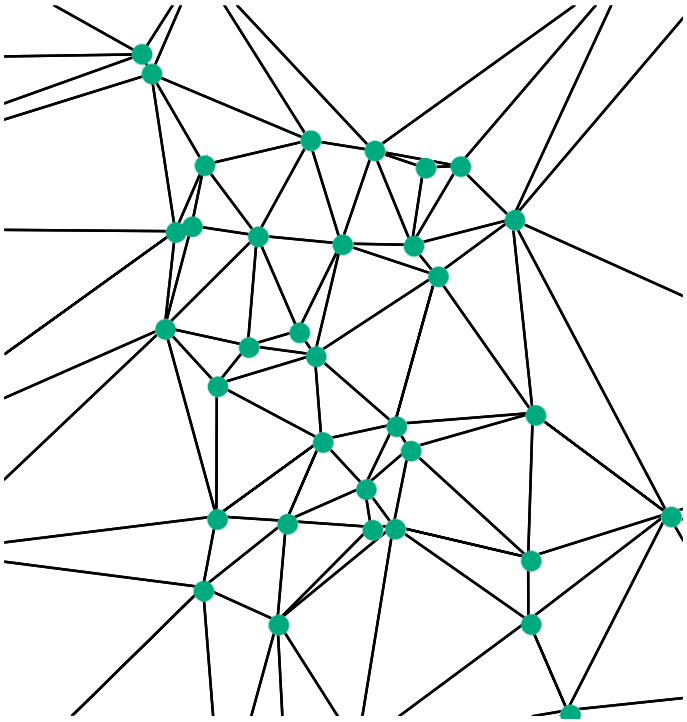
Baddeley et al. (2009)

Tesselation-based techniques

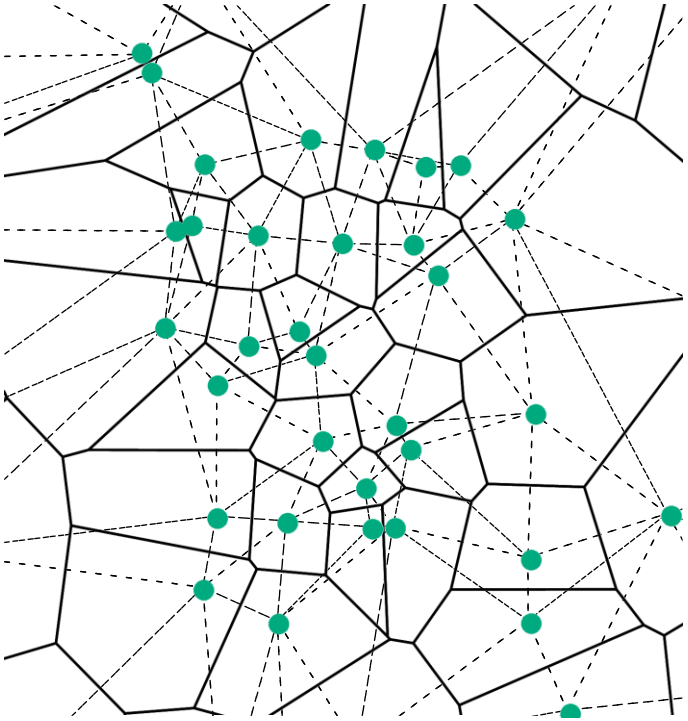
- Non normalized
- The relation between the triangles and the localizations is difficult to envision
  - How to define relevant features ?

Tesselation-based techniques

- Voronoï diagram
  - Dual of the Delaunay triangulation
  - One can be constructed from the other



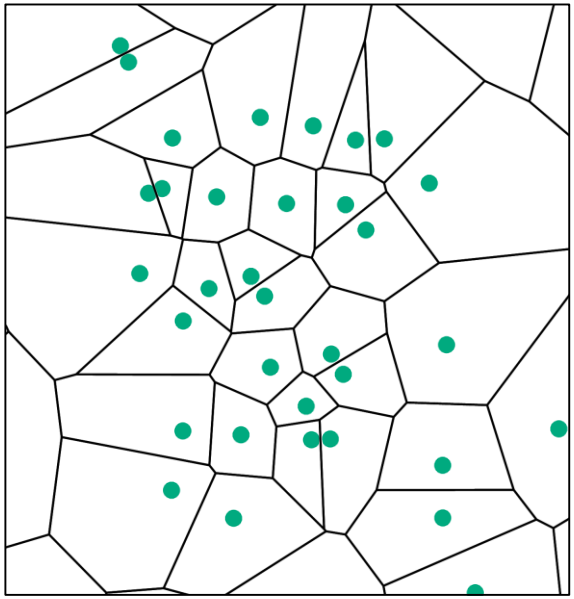
Delaunay



Voronoï

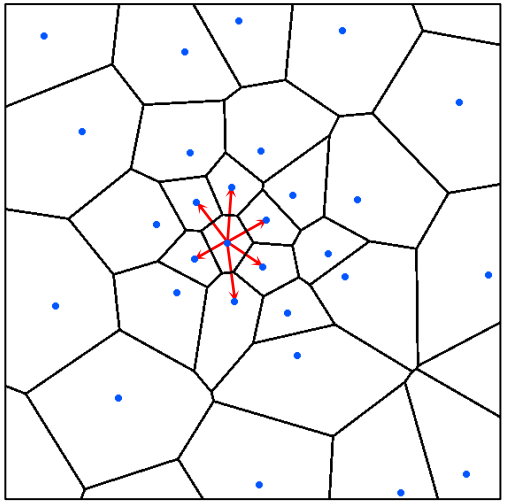
Voronoï diagram

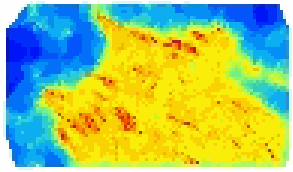
- Space-subdividing technique
- Anisotropic by nature
- One loc/polygon → bijective representation



Voronoï diagram

- Connectivity
  - Direct neighbors are known
- Scalability
  - The denser the locs, the smaller the polygons are

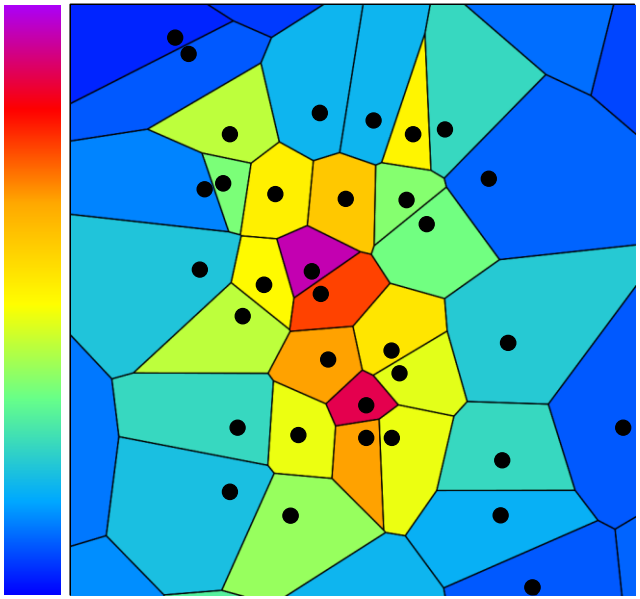




Voronoï diagram

- Space-subdividing technique
- One loc/polygon → bijective representation
- Statistics computed on the polygons (area, density...)

Polygon P
# of neighbors n
Polygon area A
Local density δ



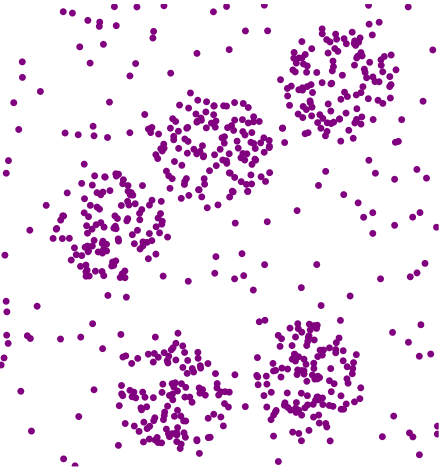
Rank 1 Density map

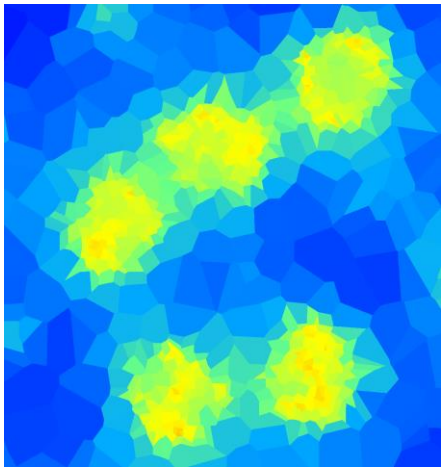
Rank i
$n_i = \#P_i + n_{i-1}$
$A_i = \sum(P_i) + A_{i-1}$
$\delta_i = n_i / A_i$

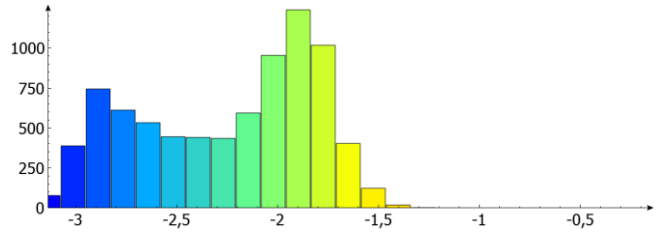
Voronoï normalization

Ratio 1:5

0.013 mol/nm²

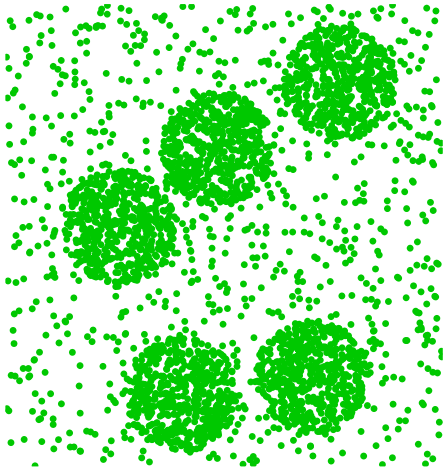


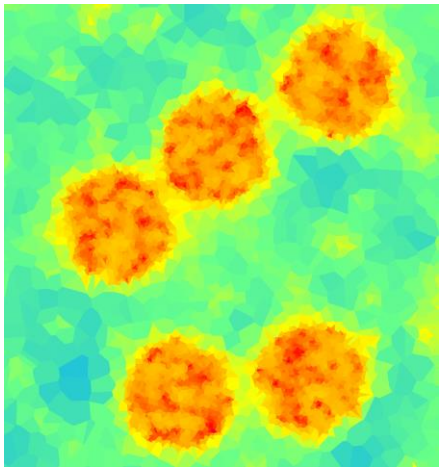


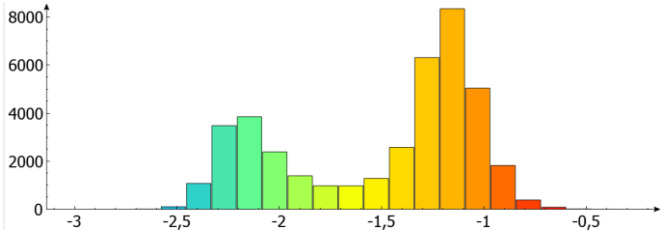


Log(δ)

0.065 mol/nm²

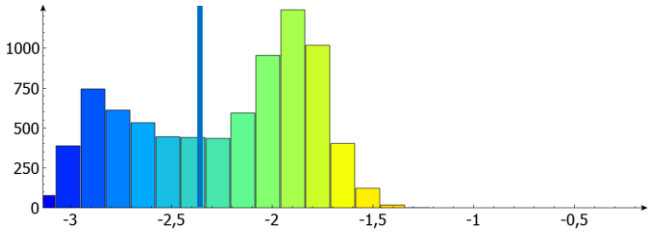


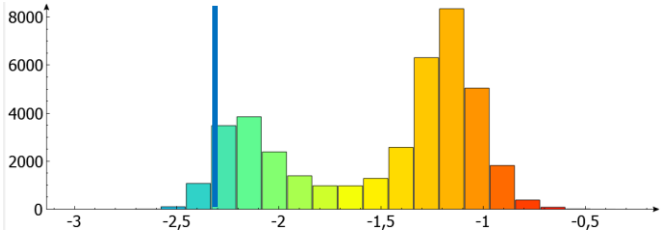


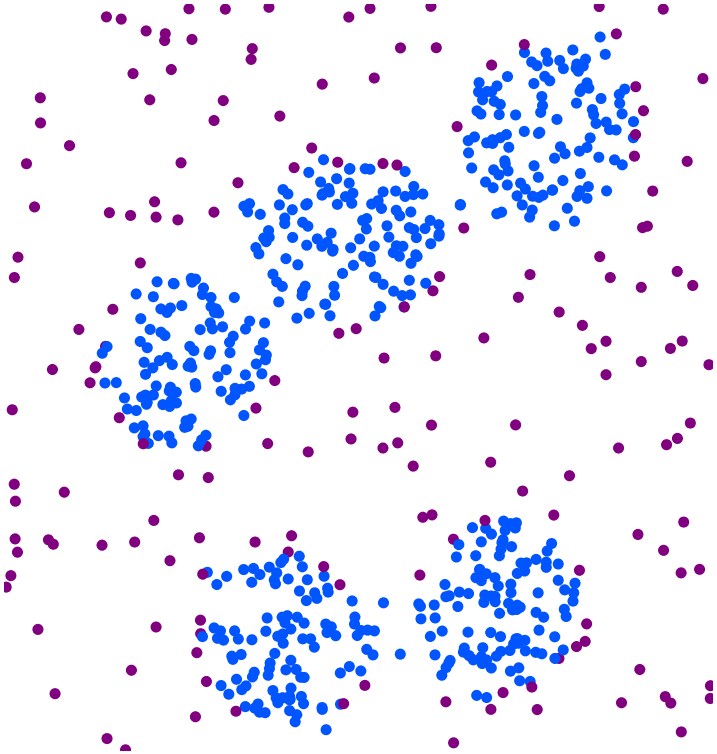


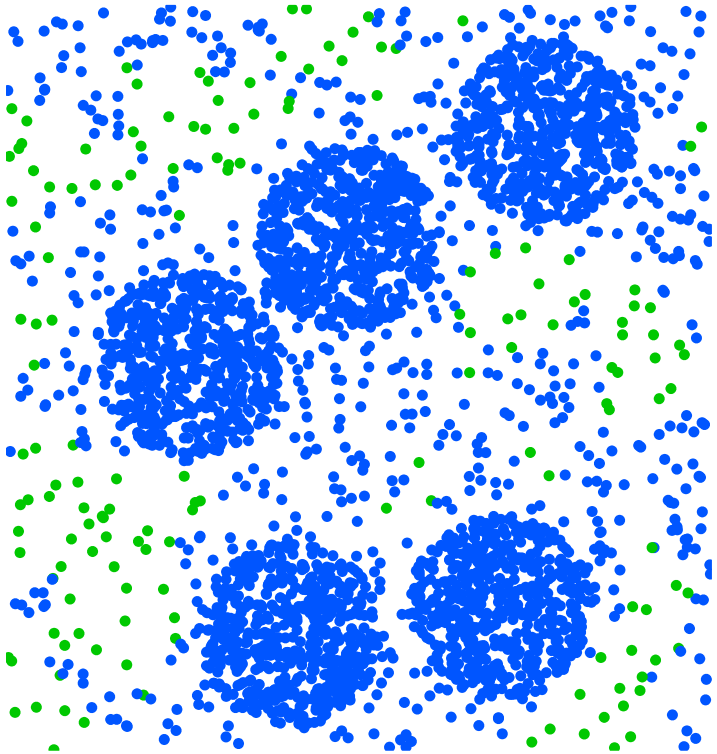
Log(δ)

Voronoï normalization





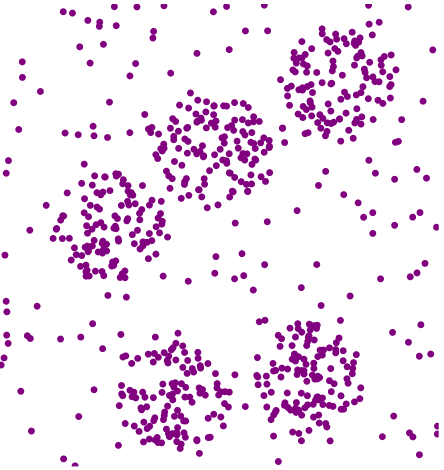


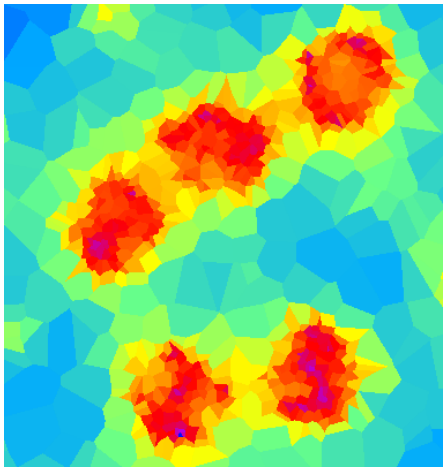


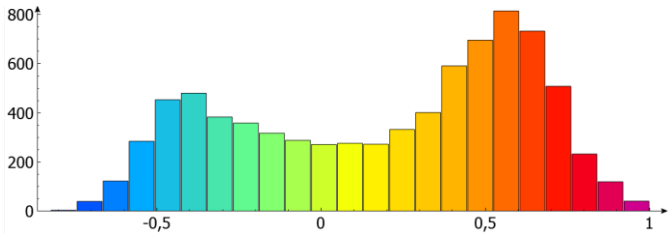
Voronoï normalization

Ratio 1:5

0.013 mol/nm²

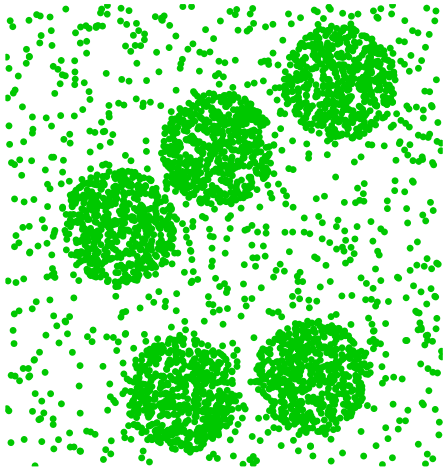


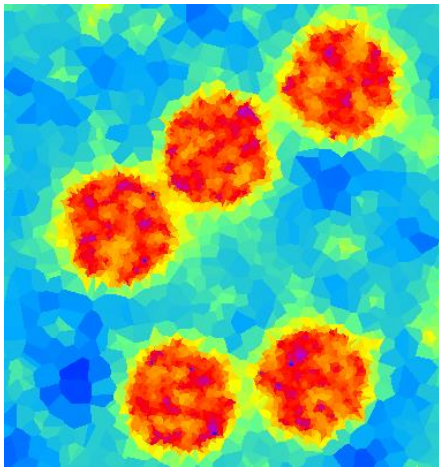


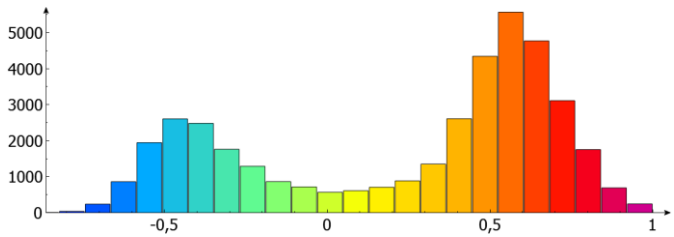


Log($\hat{\delta}$)

0.065 mol/nm²

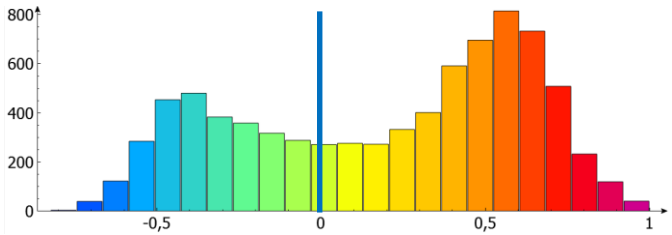


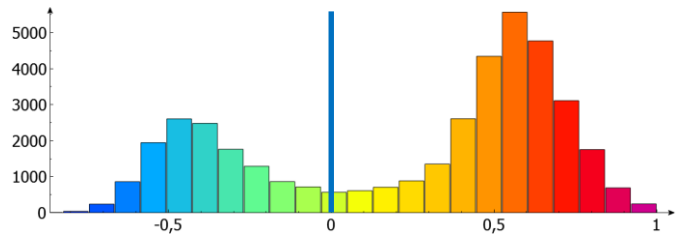


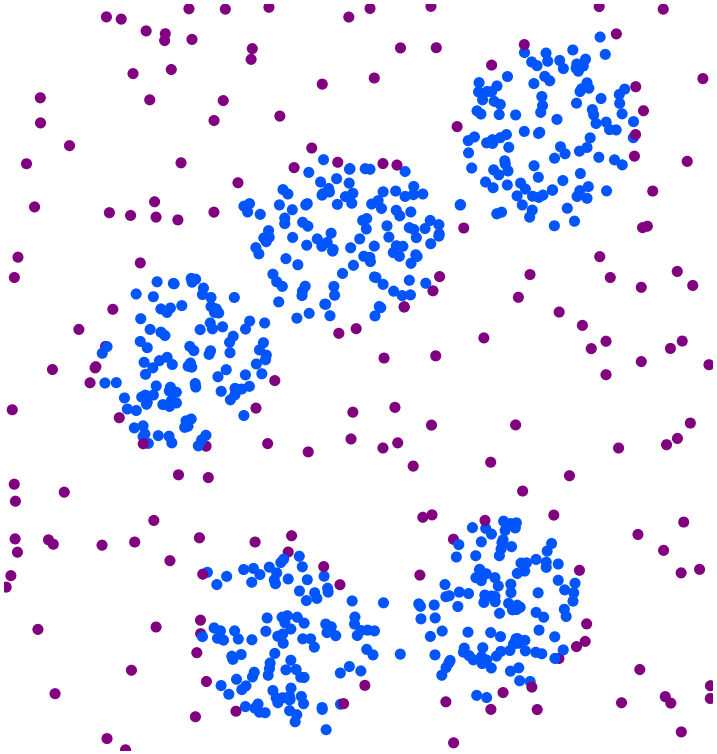


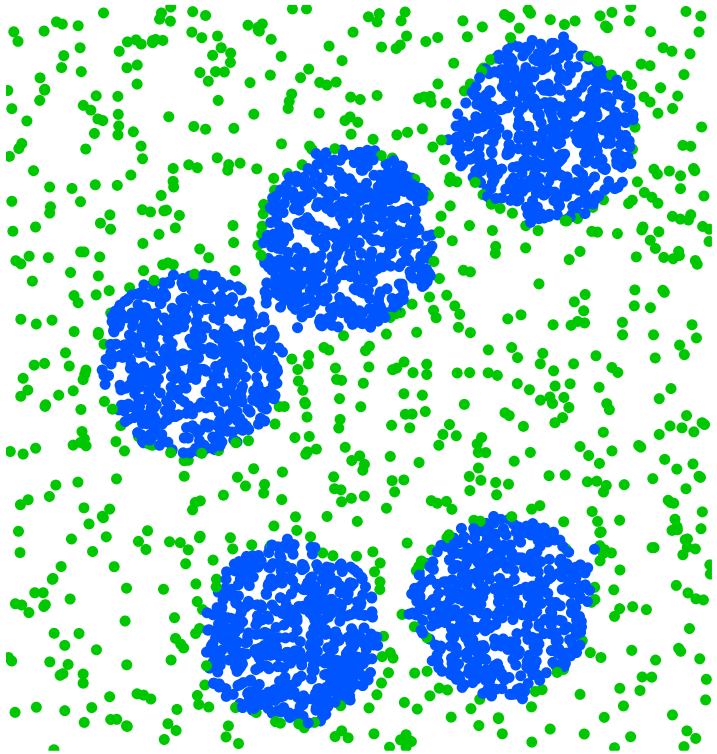
Log($\hat{\delta}$)

Voronoï normalization



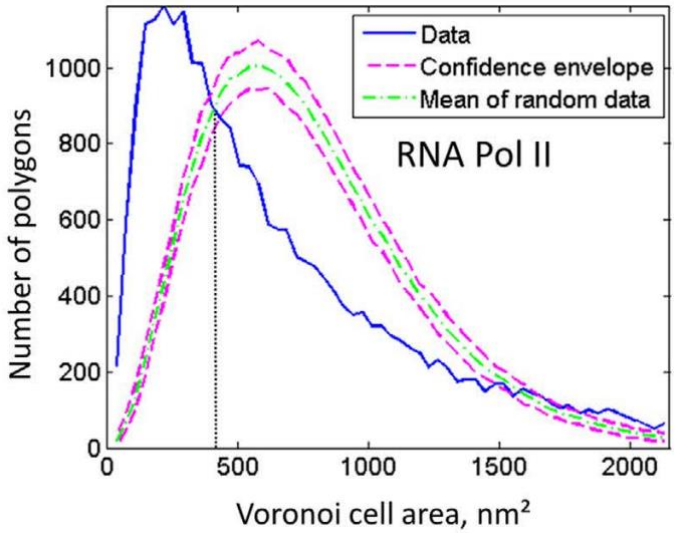






Voronoï segmentation

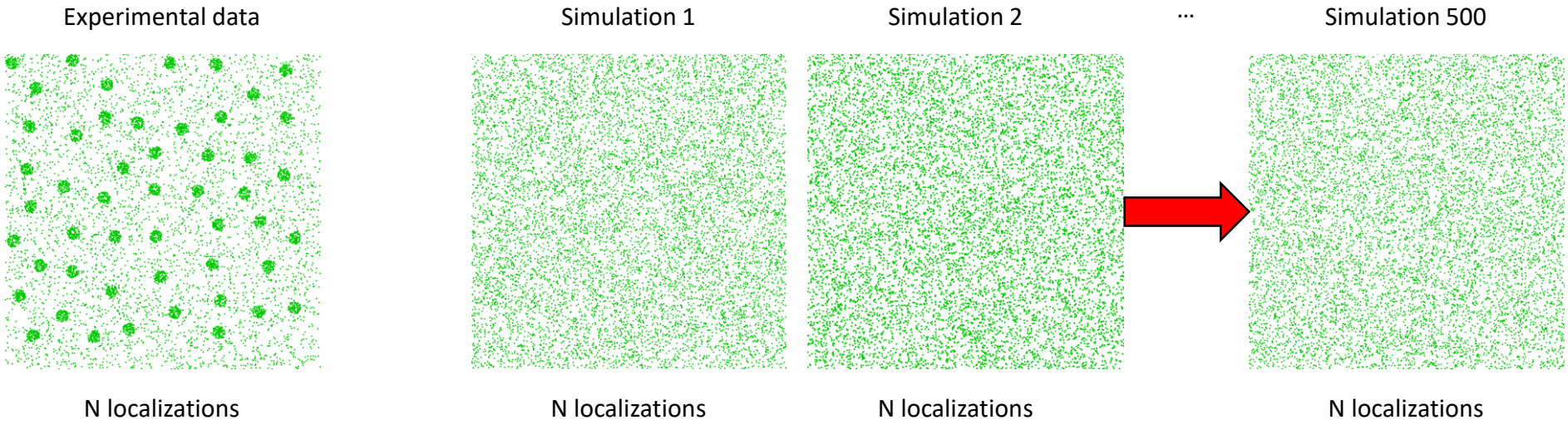
- Voronoï diagram is an intrinsic multi-resolution structure that can be normalized
- Compare the area distribution to Monte Carlo simulations



Levet et al. (2015) ; Andronov et al. (2016)

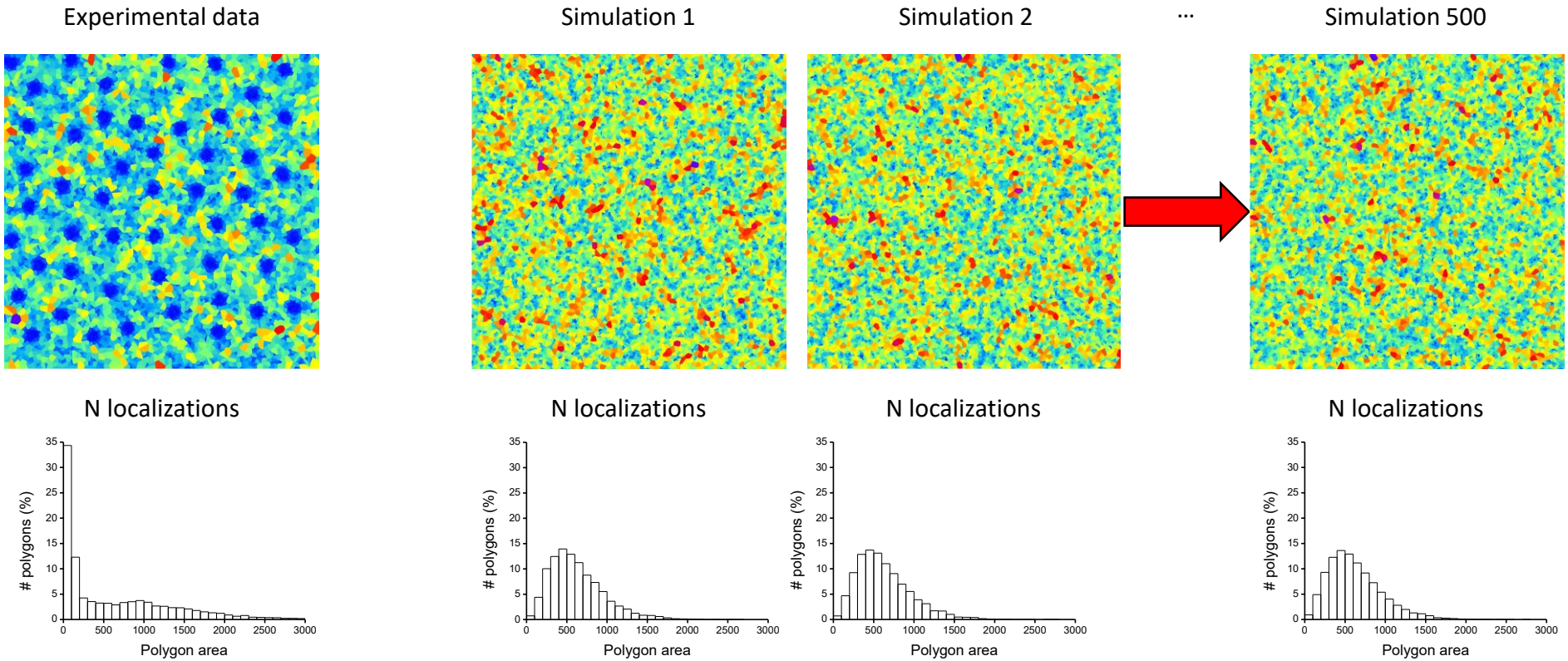
Voronoï: Monte Carlo simulations

- Statistical determination of clustering
- Simulations of complete spatiallly random data



Voronoï: Monte Carlo simulations

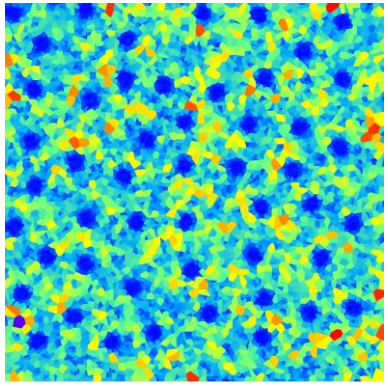
- Statistical determination of clustering
- Simulations of complete spatiallly random data



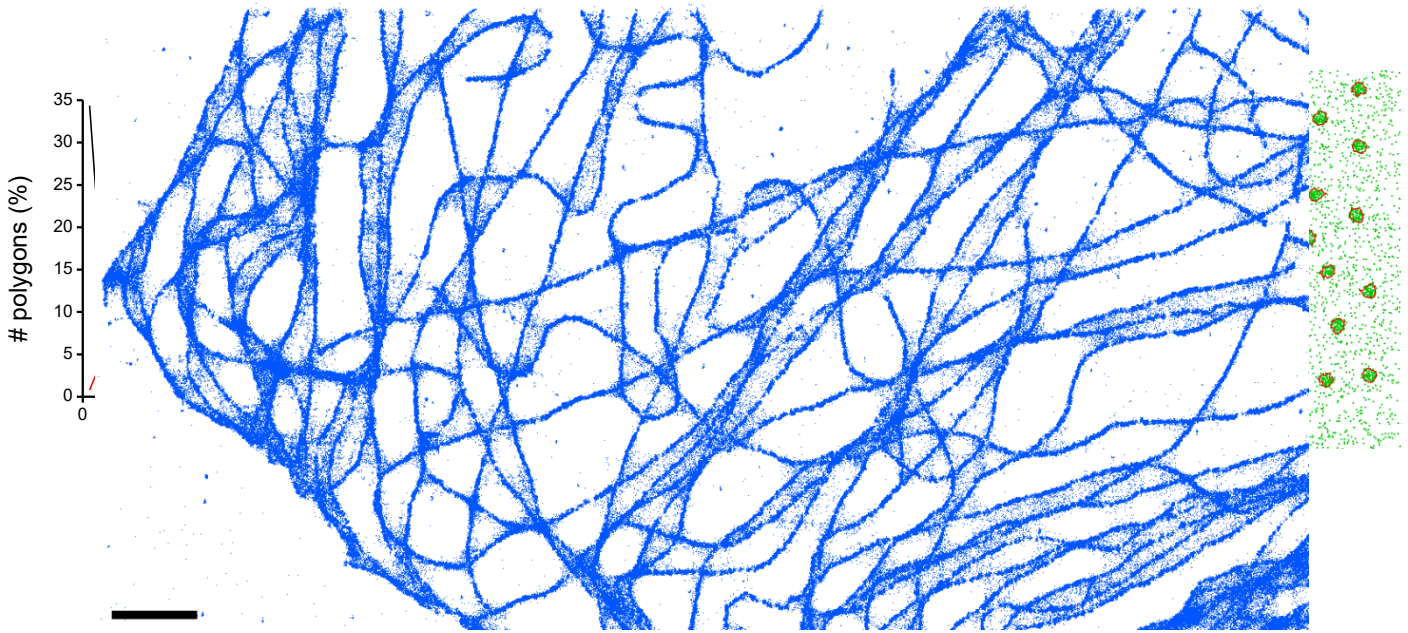
Voronoï: Monte Carlo simulations

- Statistical determination of clustering
- Simulations of complete spatiallly random data

Experimental data

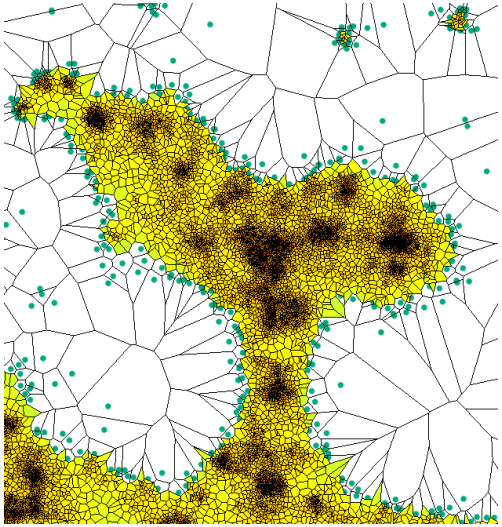


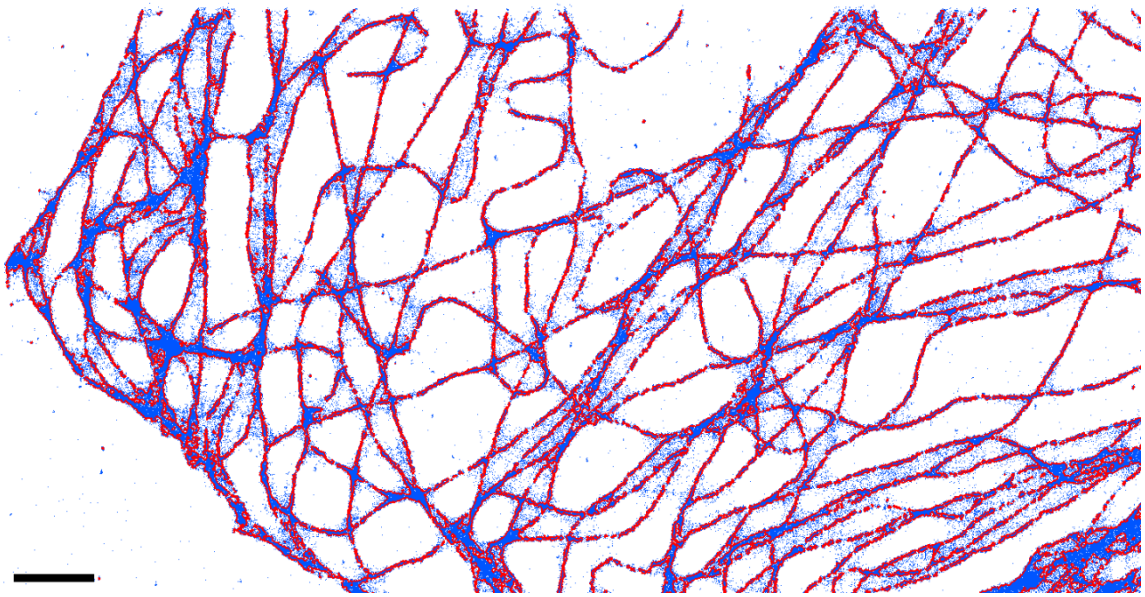
N localizations



Voronoï segmentation

- Voronoï diagram is an intrinsic multi-resolution structure that can be normalized
- Compare the local density to the average density δ

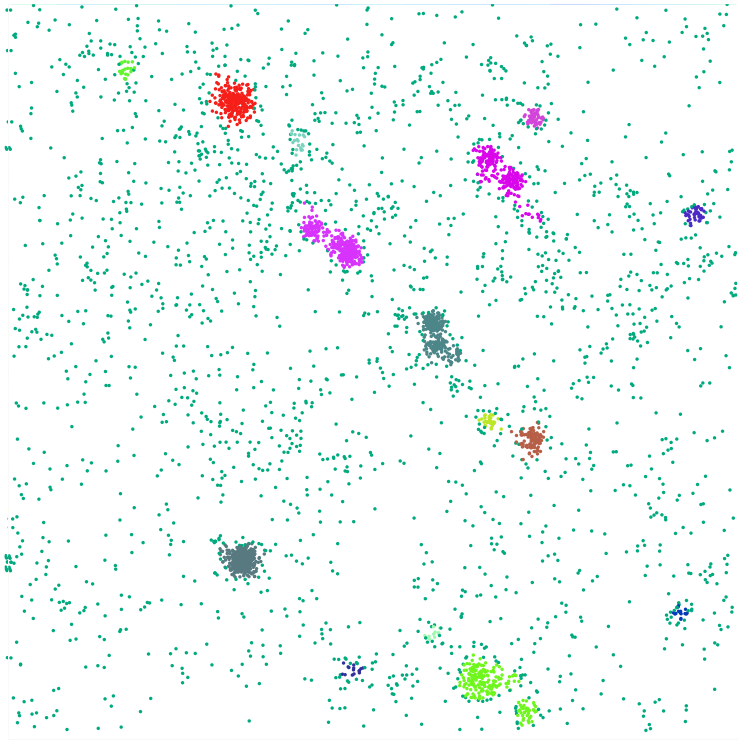


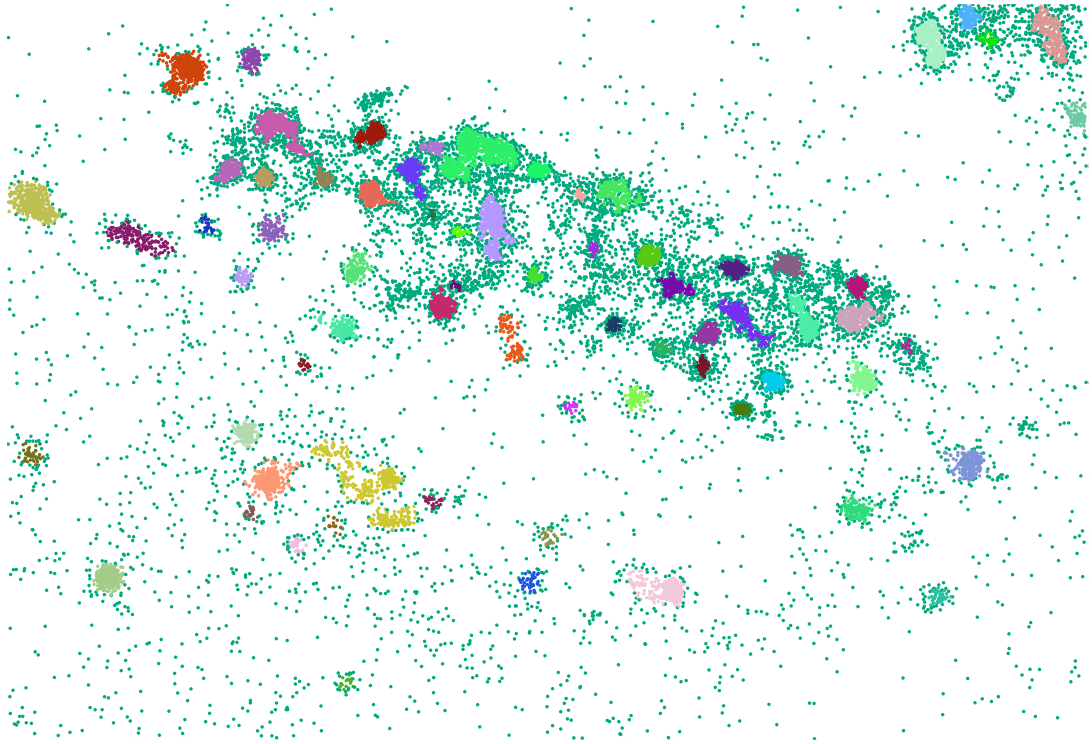


Levet et al. (2015)

SR-Tesseler software, freely available, www.iins.fr/team-sibarita

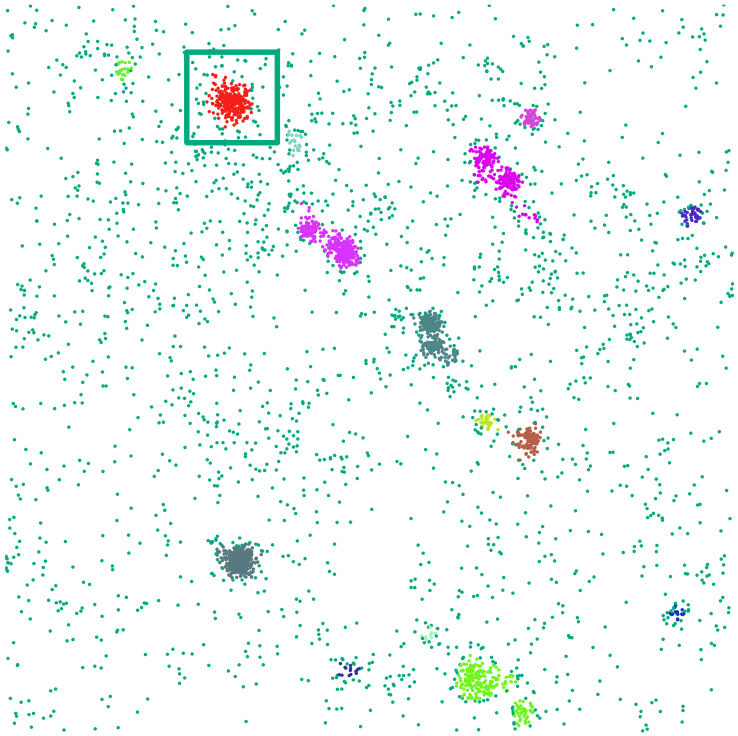
Local density normalization

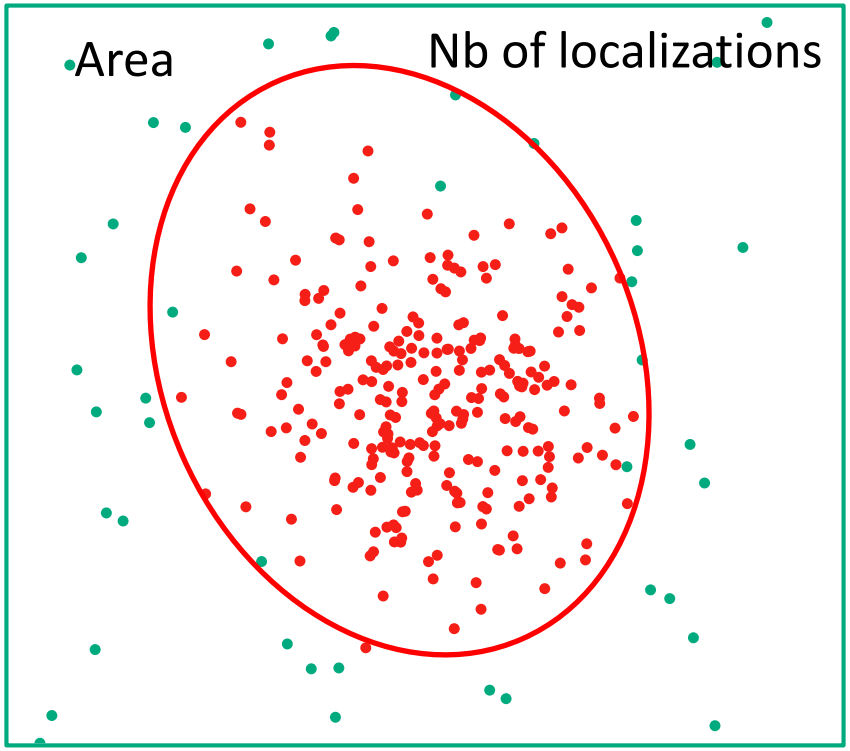




Quantifications on segmented objects

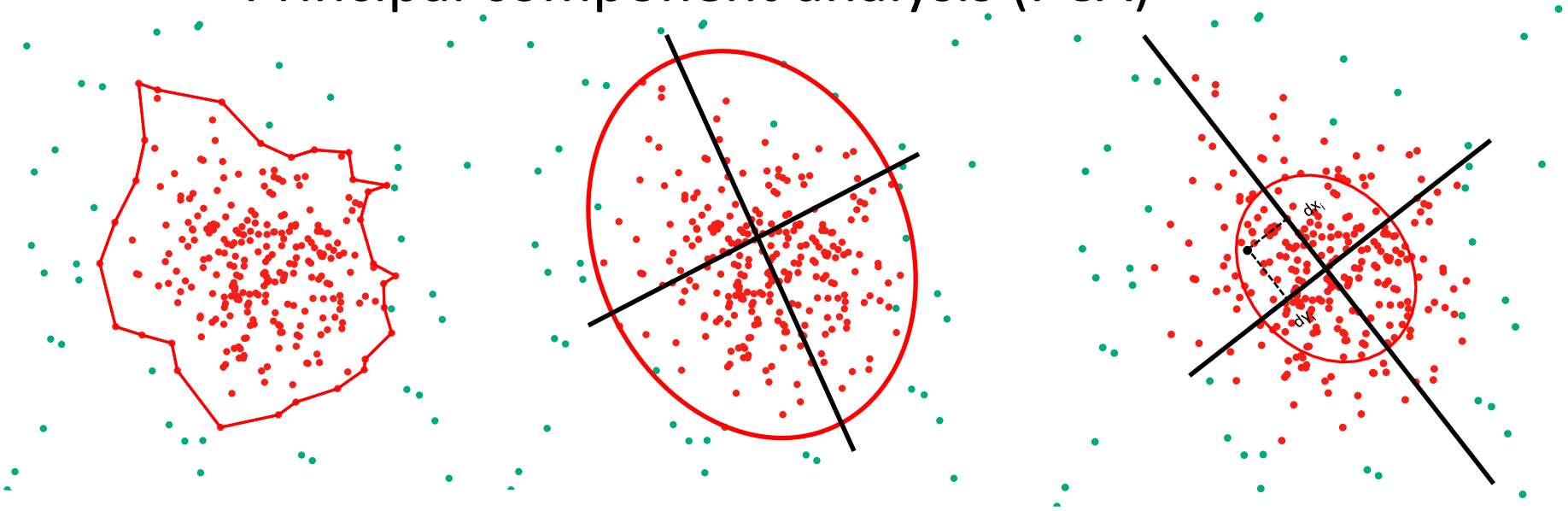
- **Size** and **Number of localizations** of segmented objects





Quantifications on segmented objects

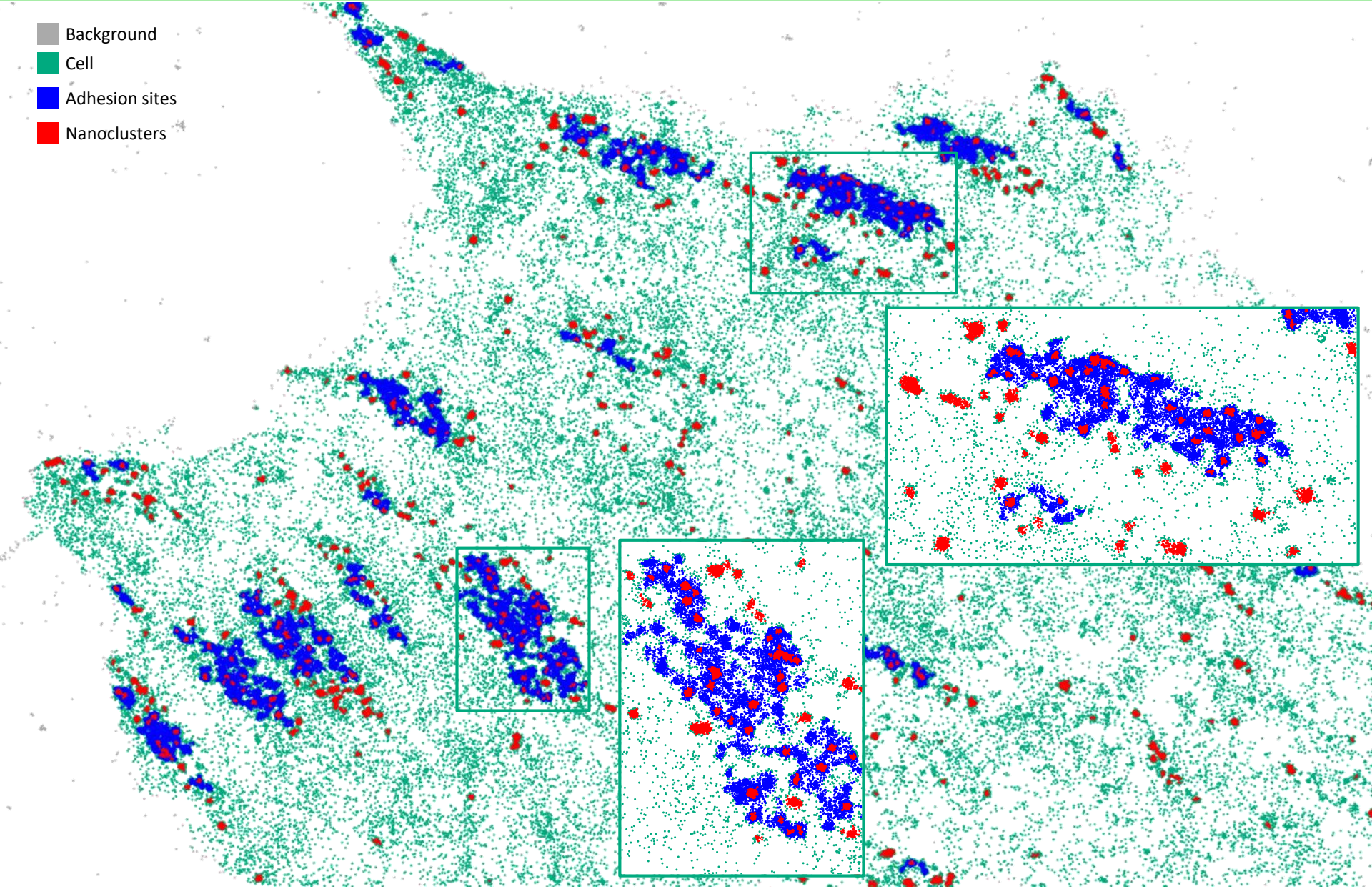
- Size of segmented objects
  - Outline → computation of the area
  - Bounding ellipse → over-estimates the object size
  - Principal component analysis (PCA)



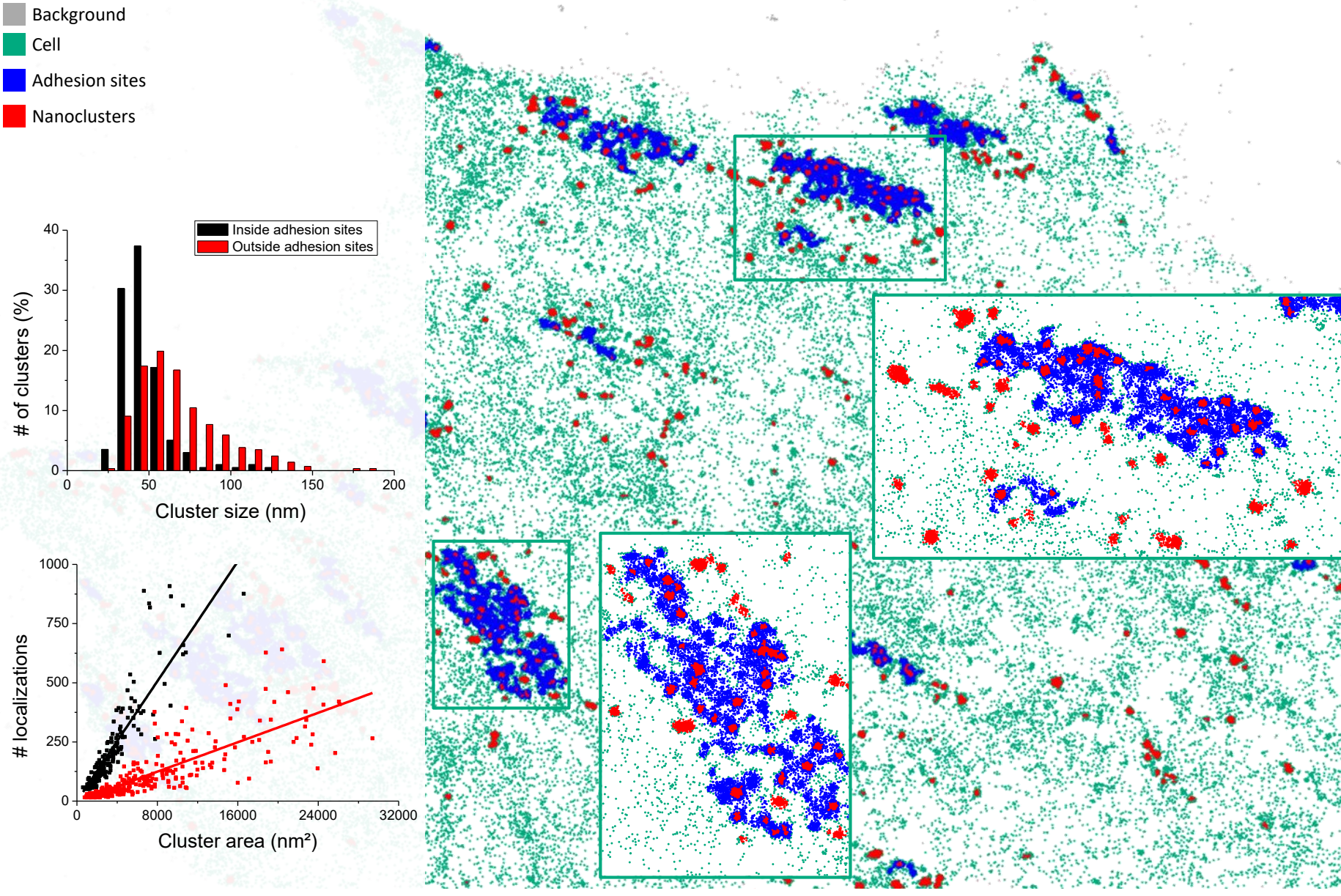
$W = 2.35 * \text{Std dev}(dx_i)$

$H = 2.35 * \text{Std dev}(dy_i)$

Multiscale segmentation

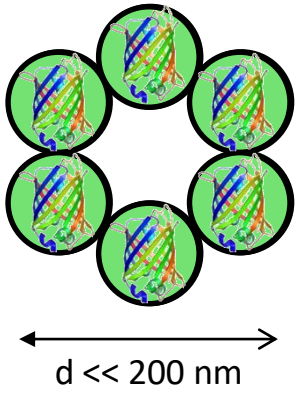


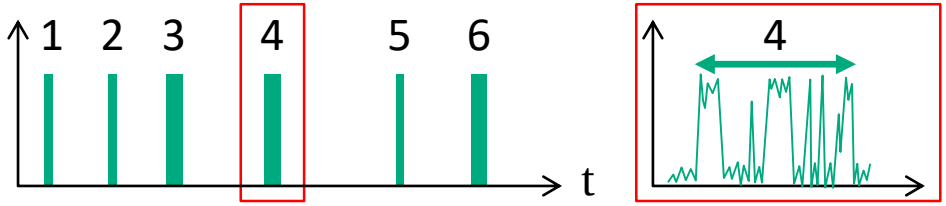
Multiscale segmentation



Quantifications on segmented objects

- Counting molecules using fusion protein



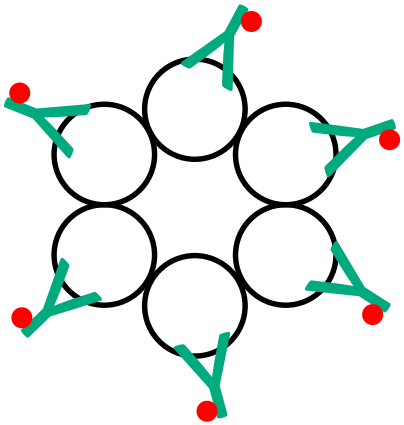


- **Problems**: **Overcounting** and **false clustering** due to **multiple localizations** per fluorescent molecule.
- **Solutions**: **Low level of activation** to temporally separate the peaks and use a **dedicated analysis program** to regroup localizations in a spatio-temporal window.

Annibale et al (2010)

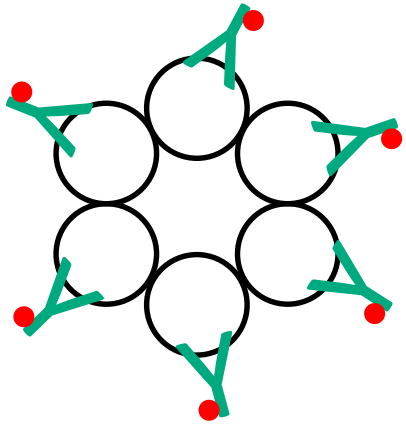
Quantifications on segmented objects

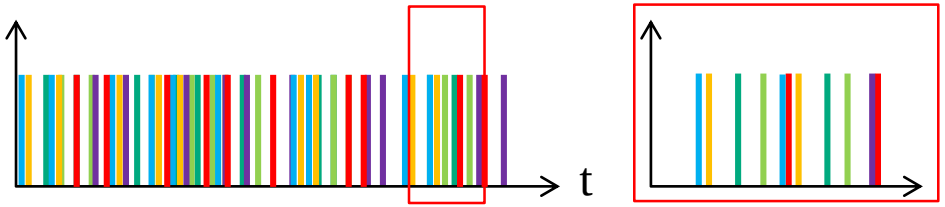
- Counting molecules using antibodies



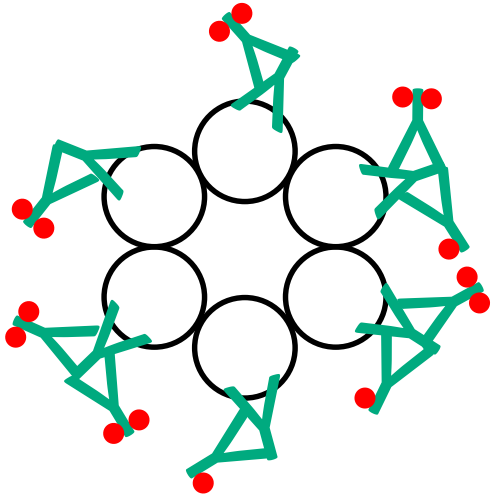
Quantifications on segmented objects

- Counting molecules using antibodies



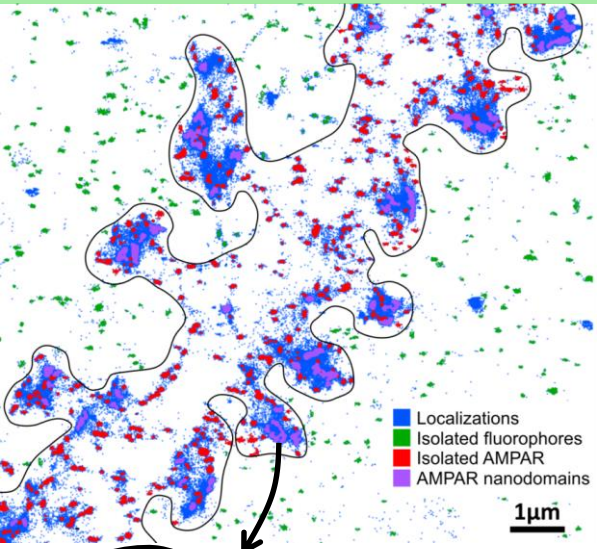


- **Ratiometric quantifications**



- **Problems**: **Labelling density** and **Over/Downcounting**.
- **Solutions**: **Direct labelling** with **short linkers** (Nanobodies, …) and tune the blinking medium to the labelling density.

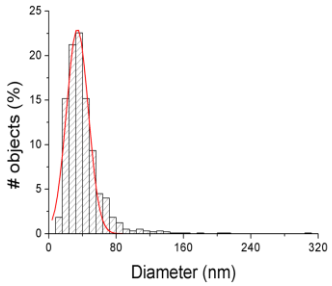
Ratiometric quantification



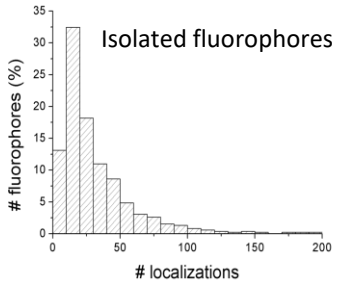
Internal control

Structure (size)

Blinking statistics (# locs)



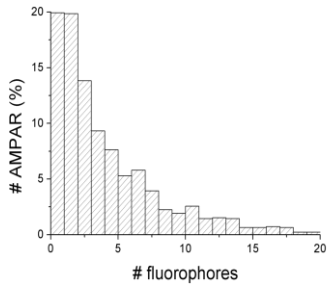
Experimental resolution
33.95 nm (FWHM)



22 localizations

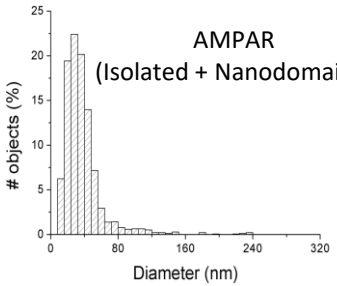
How many AMPA receptors?

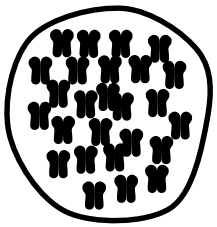
How many fluorophores?

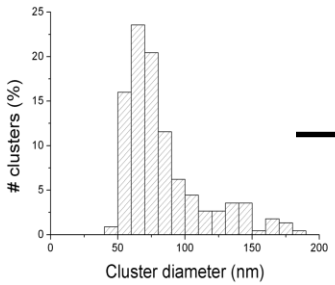


2.72 fluorophores

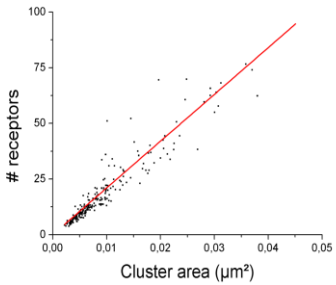
How to separate populations?







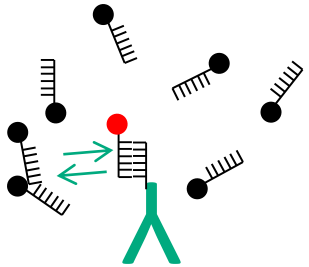
73.24 nm

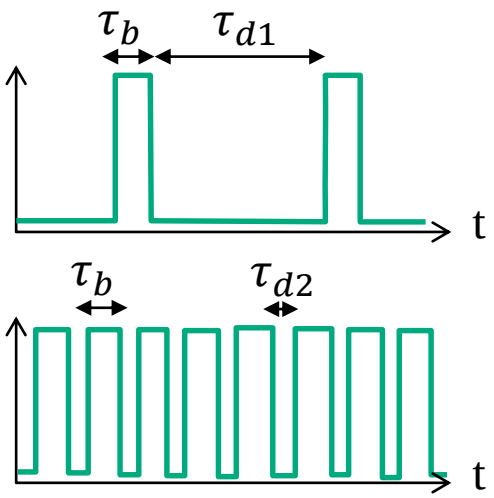


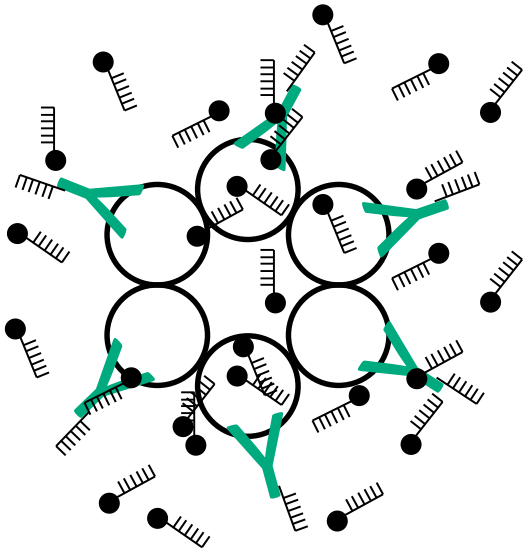
2,100 AMPAR/µm²

qPAINT

- DNA-PAINT

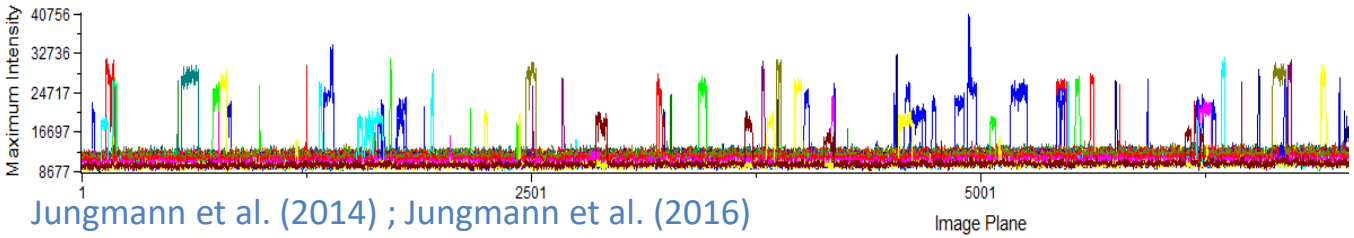


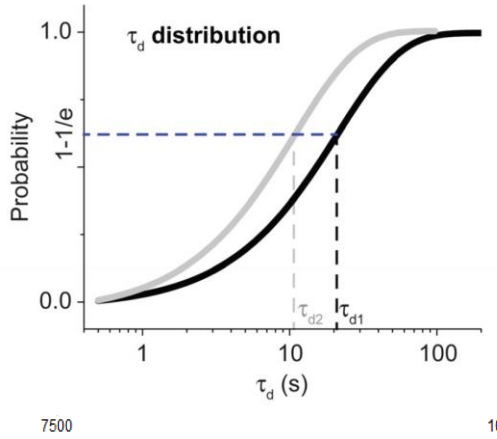




$\tau_b = k_{off}^{-1}$ $\tau_d = (c_i\, k_{on})^{-1}$

- **Quantification of the # of Binding Sites:**
  1. Determine mean dark-time τ_d from cumulative distribution function of all dark times: τ_d*
  2. Calculate # binding sites (after calibration with structure of known stochiometry)

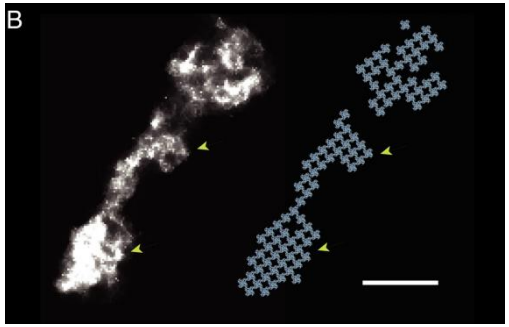




Jungmann et al. (2014) ; Jungmann et al. (2016)

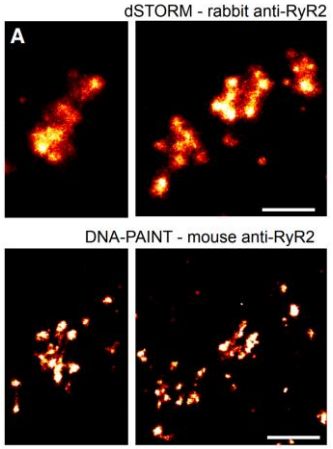
A case study: the RyRs receptors

2009: dSTORM, 30 nm resolution

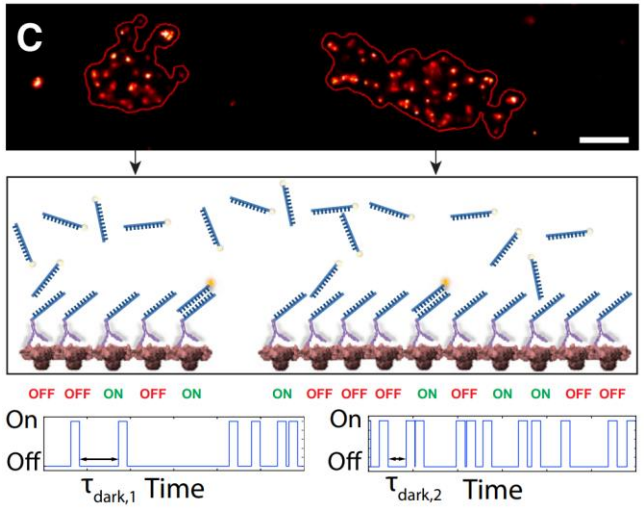


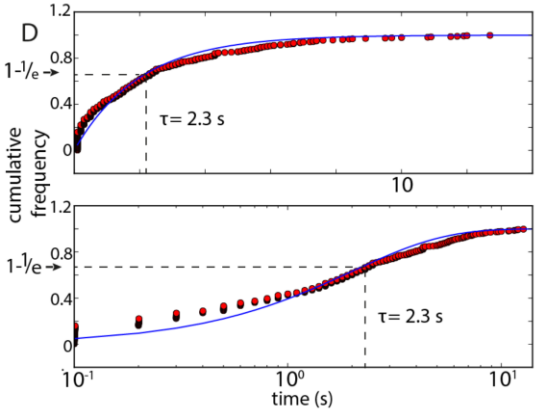
Regular 30 nm * 30 nm lattice

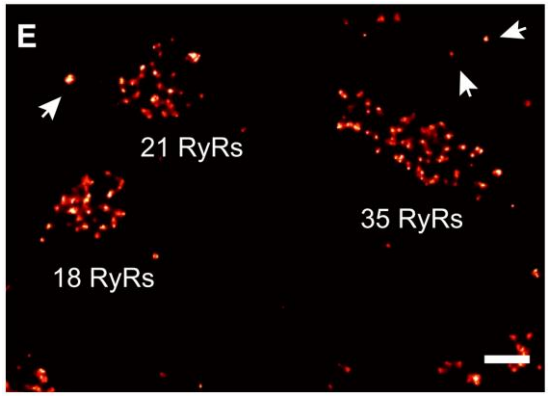
2014: Electron tomography challenged this assumpation

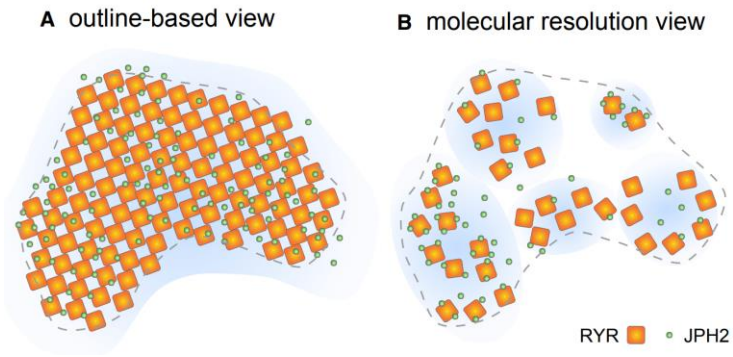


Resolution: 5 nm





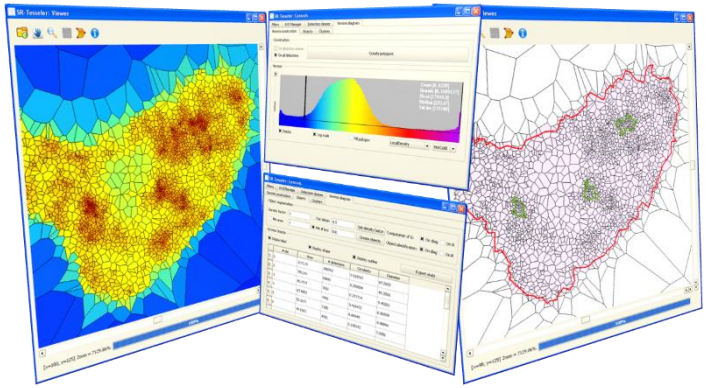




Baddeley et al. (2009), Asghari et al. (2014), Jayasinghe et al. (2018)

Misc.

Quantitative Imaging of the Cell, IINS, Bordeaux



- Source-code : https://github.com/flevet/SR-Tesseler
- Windows installer: https://github.com/flevet/SR-Tesseler/releases
- Issues: https://github.com/flevet/SR-Tesseler/issues
- Forum: https://forum.image.sc/ (tag: « tesseler »)