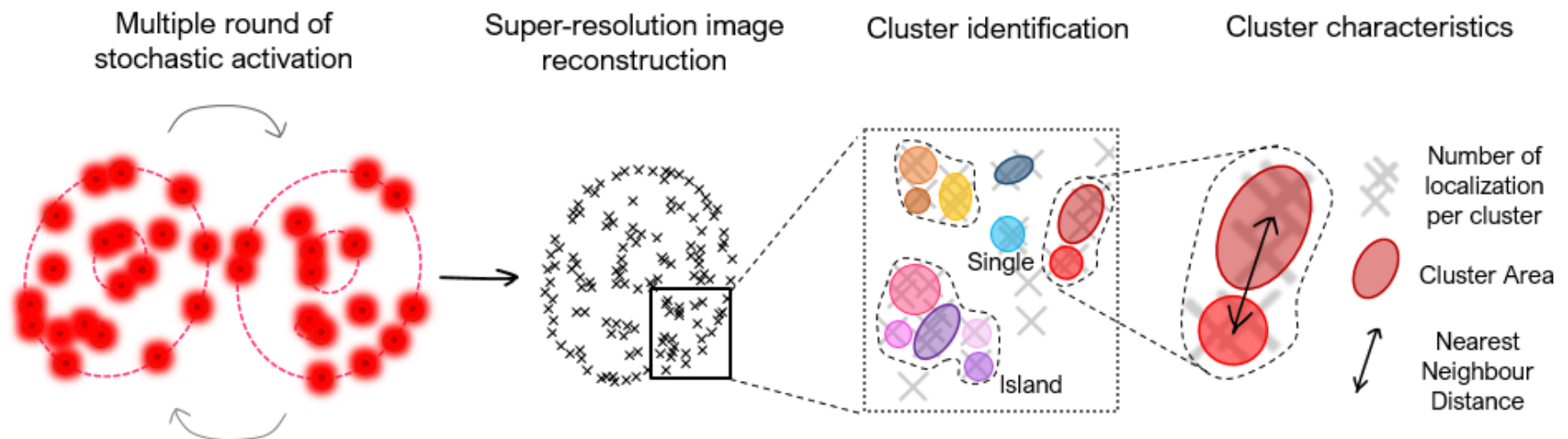# CLUSTER ANALYSIS

Visual instructions by Laura Martin, February 2022 1st 2022

**Methods in Molecular Biology 1480**

**Springer Protocols**

STORM microscopy and cluster analysis for PcG studies

Laura Martin, Álvaro Castells-García, Maria Pia Cosma, and Maria Victoria Neguembor
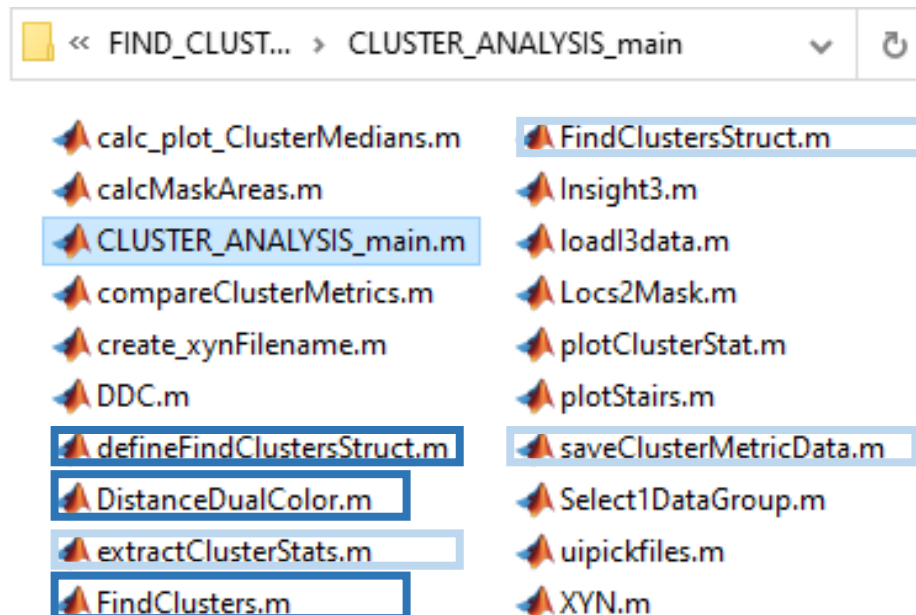
Multiple round of stochastic activation

Super-resolution image reconstruction

Cluster identification

Single

Island

Cluster characteristics

Number of localization per cluster

Cluster Area

Nearest Neighbour Distance

# CLUSTER ANALYSIS

**Original script from Carlo Manzo**
Modified in the years by
J Otterstorm, J Borbely, A Castells, & me

20 functions

# CLUSTER ANALYSIS

*Hierarchical code organization*

> **CLUSTER_ANALYSIS_main**

> **defineClustersStruct** $\longrightarrow$ Set analysis **parameters**
>> **>FindClusterStruct**

> **FindClusters** $\longrightarrow$ Identifies Islands and clusters
**Calculates clusters' features**:
- n° localizations
- Areas
- NND (of In-Island only)

> **saveClusterMetricData** $\longrightarrow$ Calculates the **global NND** between ALL clusters
>> **>extractClusterStats**
>>> **>DistanceDualColor**

**Saves** output files
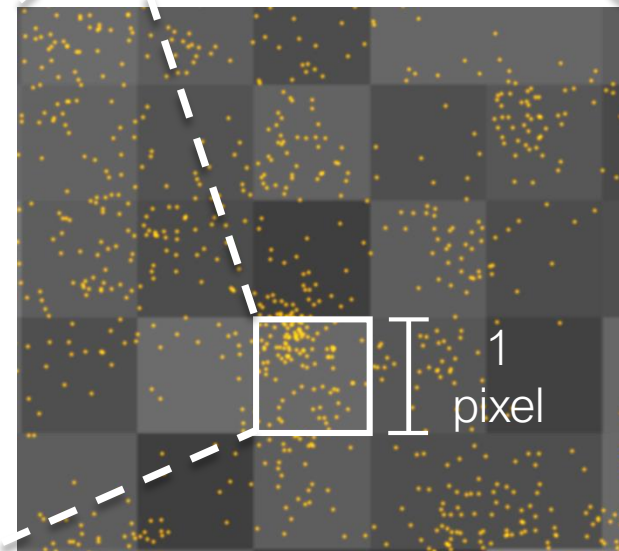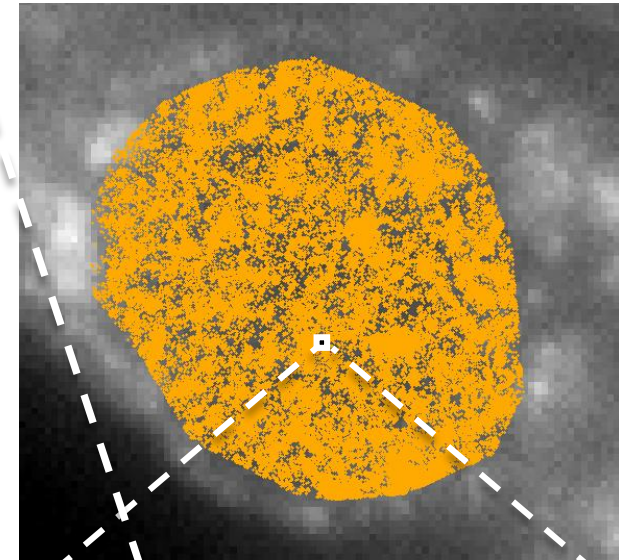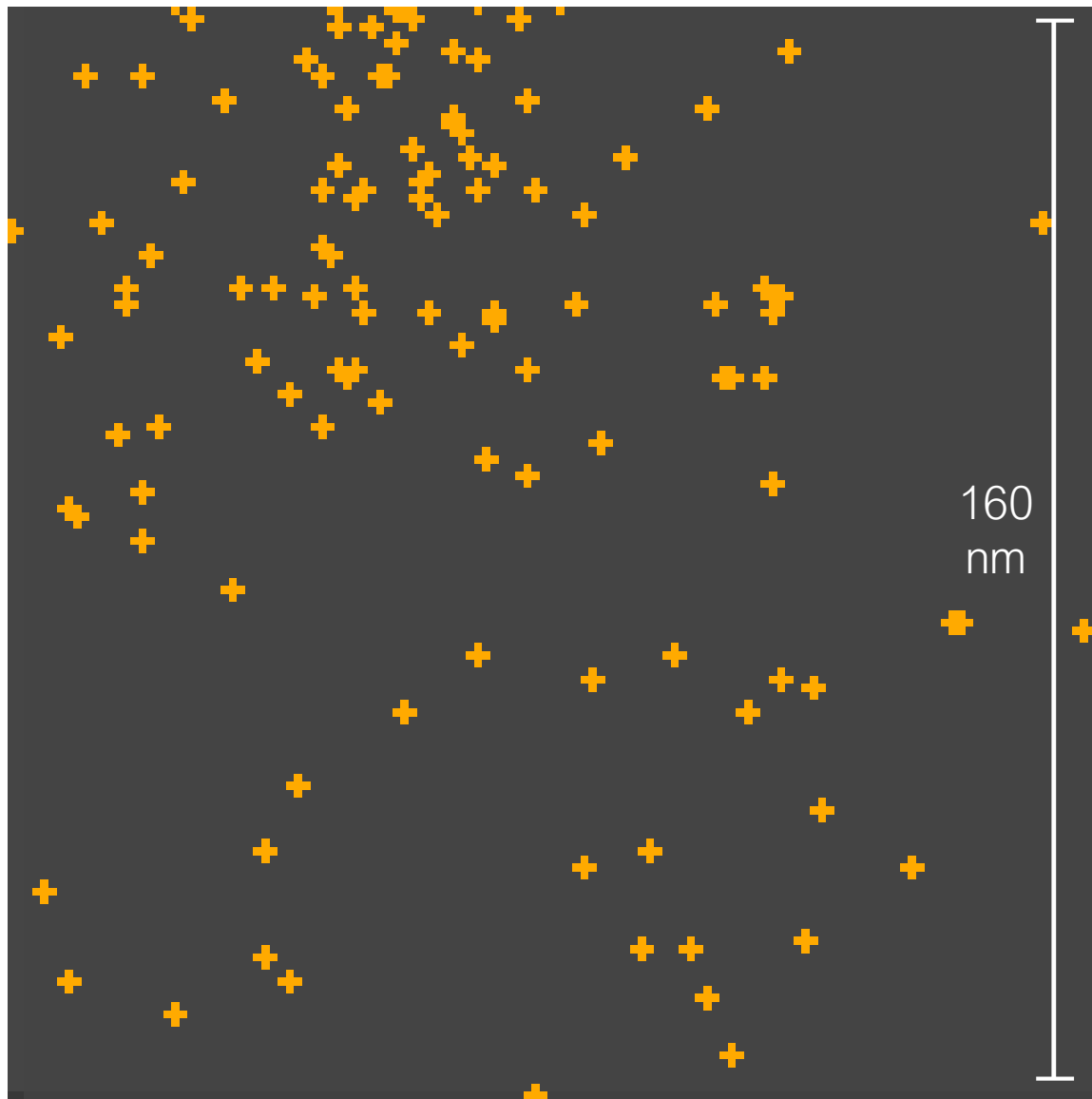(.xlsx, .mat, .bin, .dcc, .fig, .png)

# CLUSTER ANALYSIS



Dependent on Microscope/Camera settings

CRUCIAL FOR CLUSTER ANALYSIS

Optionals
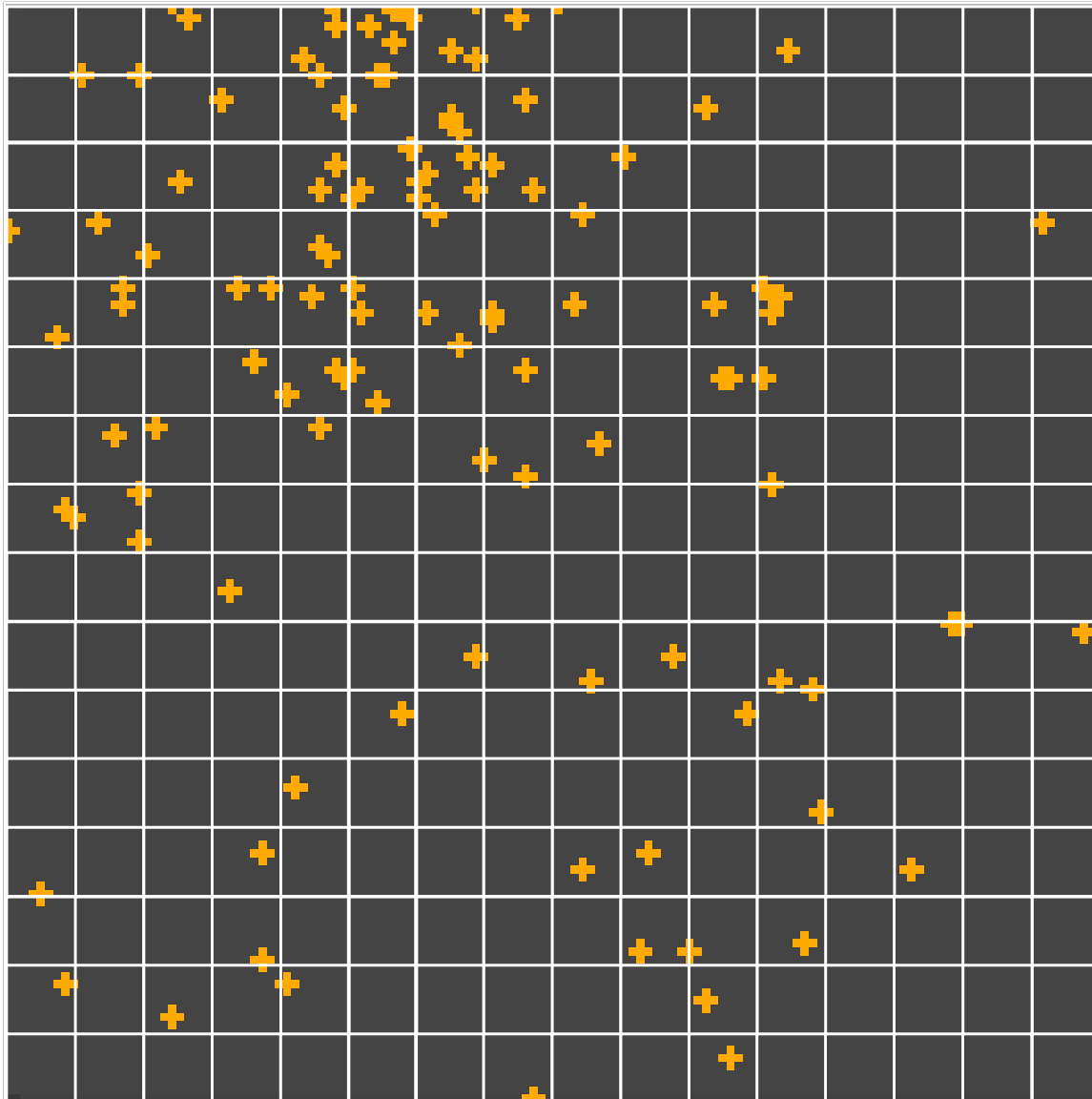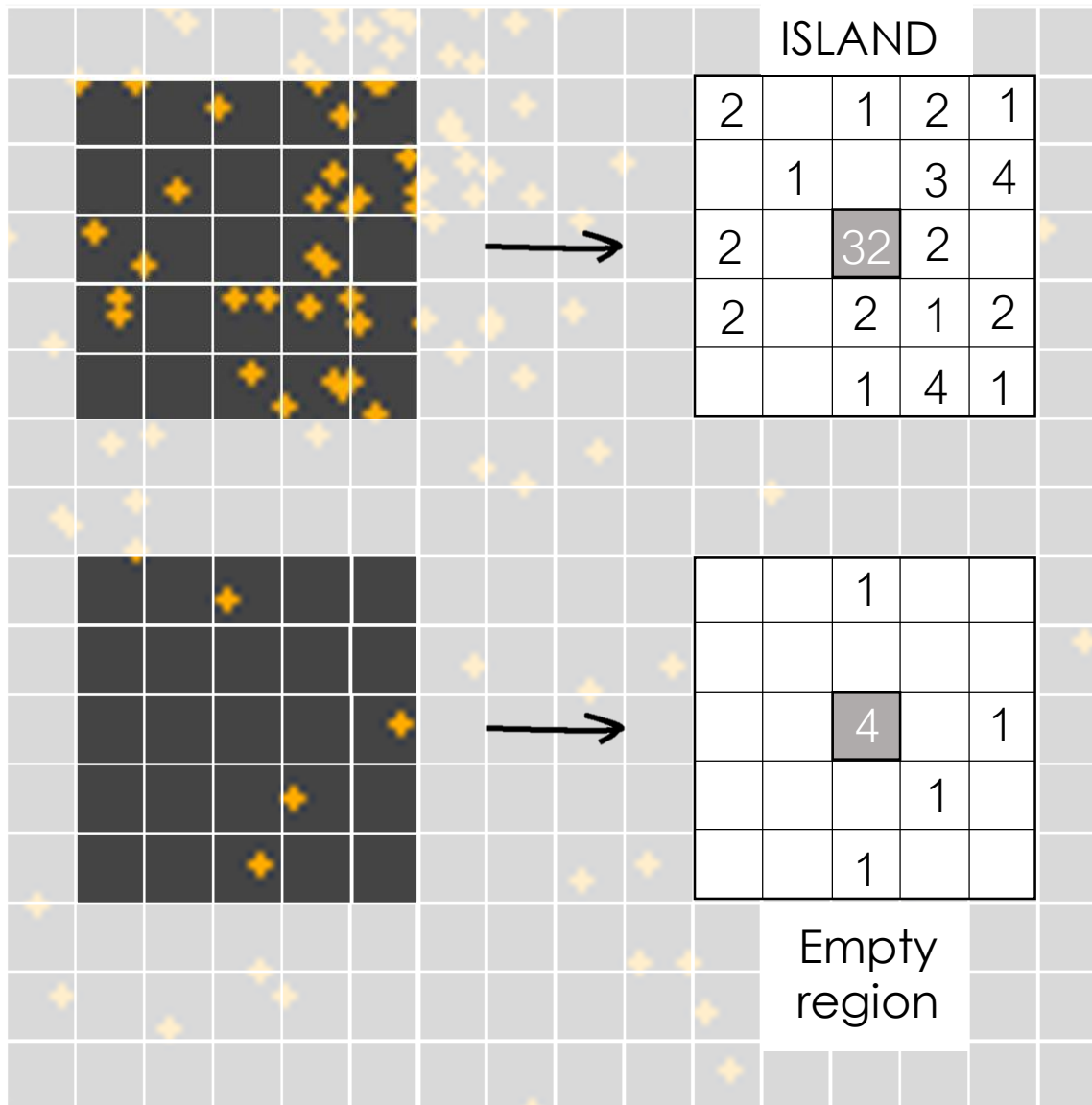(Particular conditions, plots…)

# CLUSTER ANALYSIS



160 nm

1 pixel

# CLUSTER ANALYSIS

10 nm



original_pixel_size [nm]

160

analysis_pixel_size [nm]

10

# CLUSTER ANALYSIS



ISLAND

| 2 |   | 1 | 2 | 1 |
|---|---|---|---|---|
|   | 1 |   | 3 | 4 |
| 2 |   | 32 | 2 |   |
| 2 |   | 2 | 1 | 2 |
|   |   | 1 | 4 | 1 |

|   |   | 1 |   |   |
|---|---|---|---|---|
|   |   |   |   |   |
|   |   | 4 |   | 1 |
|   |   |   | 1 |   |
|   |   | 1 |   |   |

Empty region

original_pixel_size [nm]

160

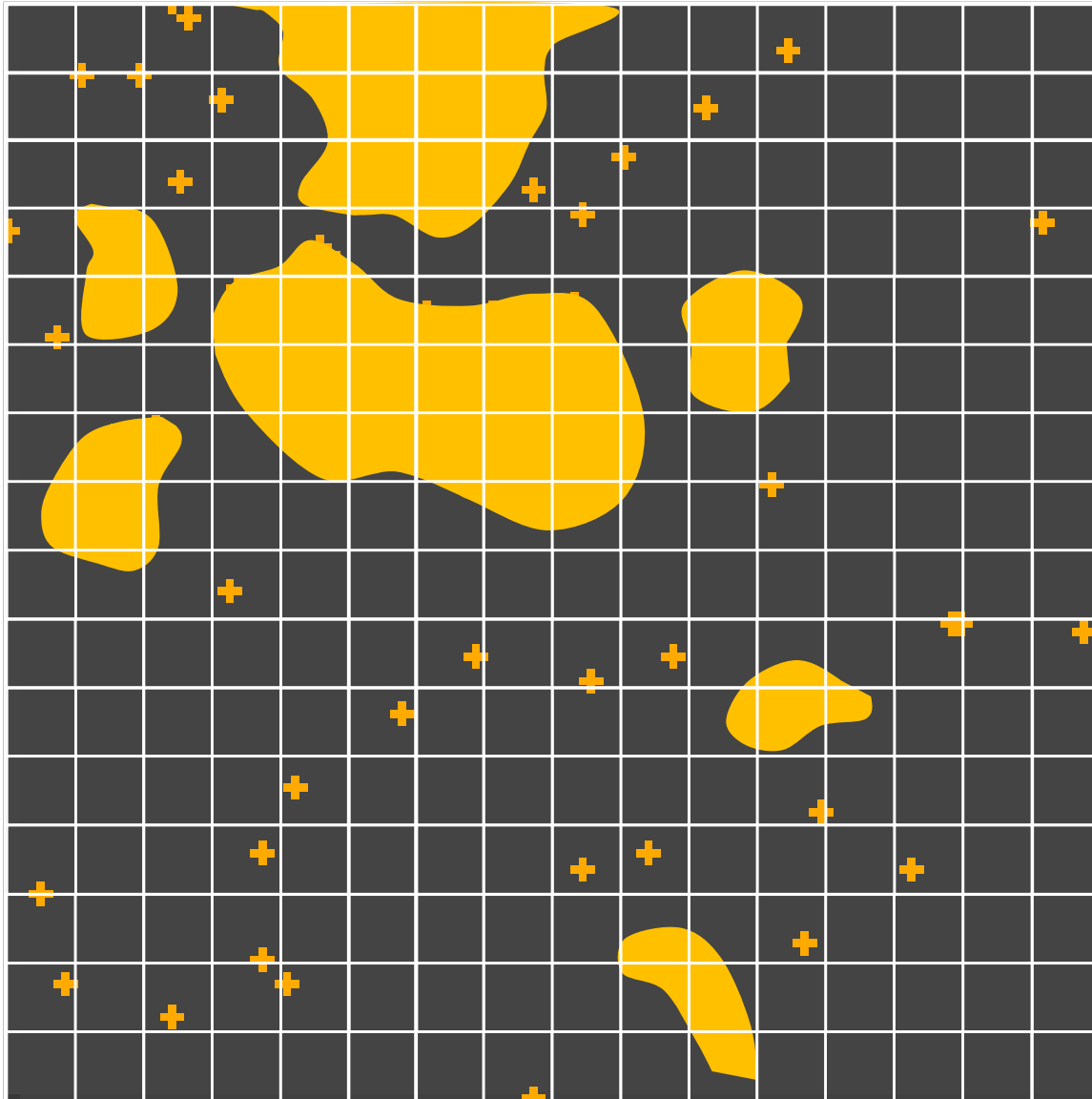analysis_pixel_size [nm]

10

sum_roi_size [pixels]

5

sum_threshold

5

# CLUSTER ANALYSIS
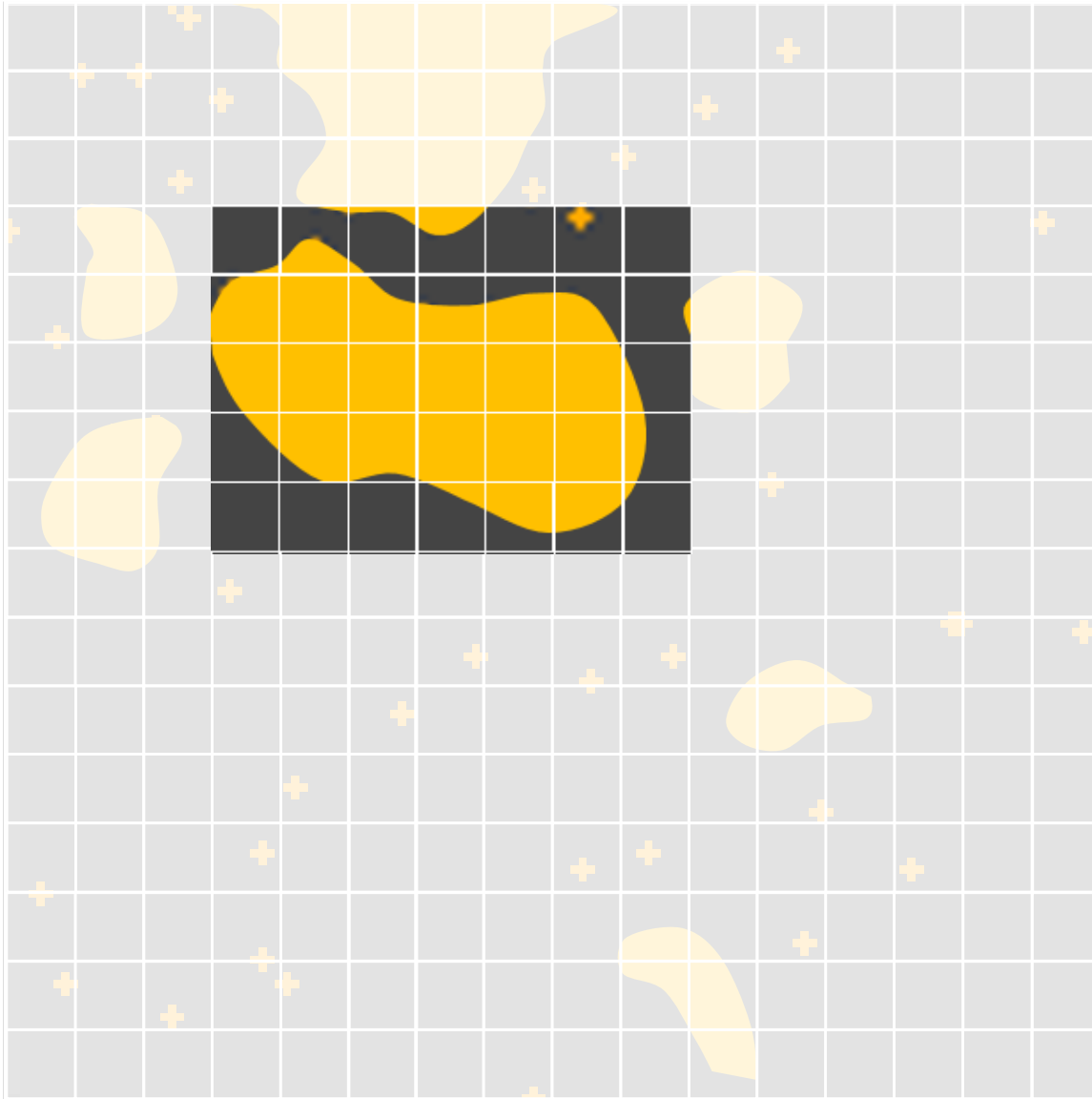
10 nm
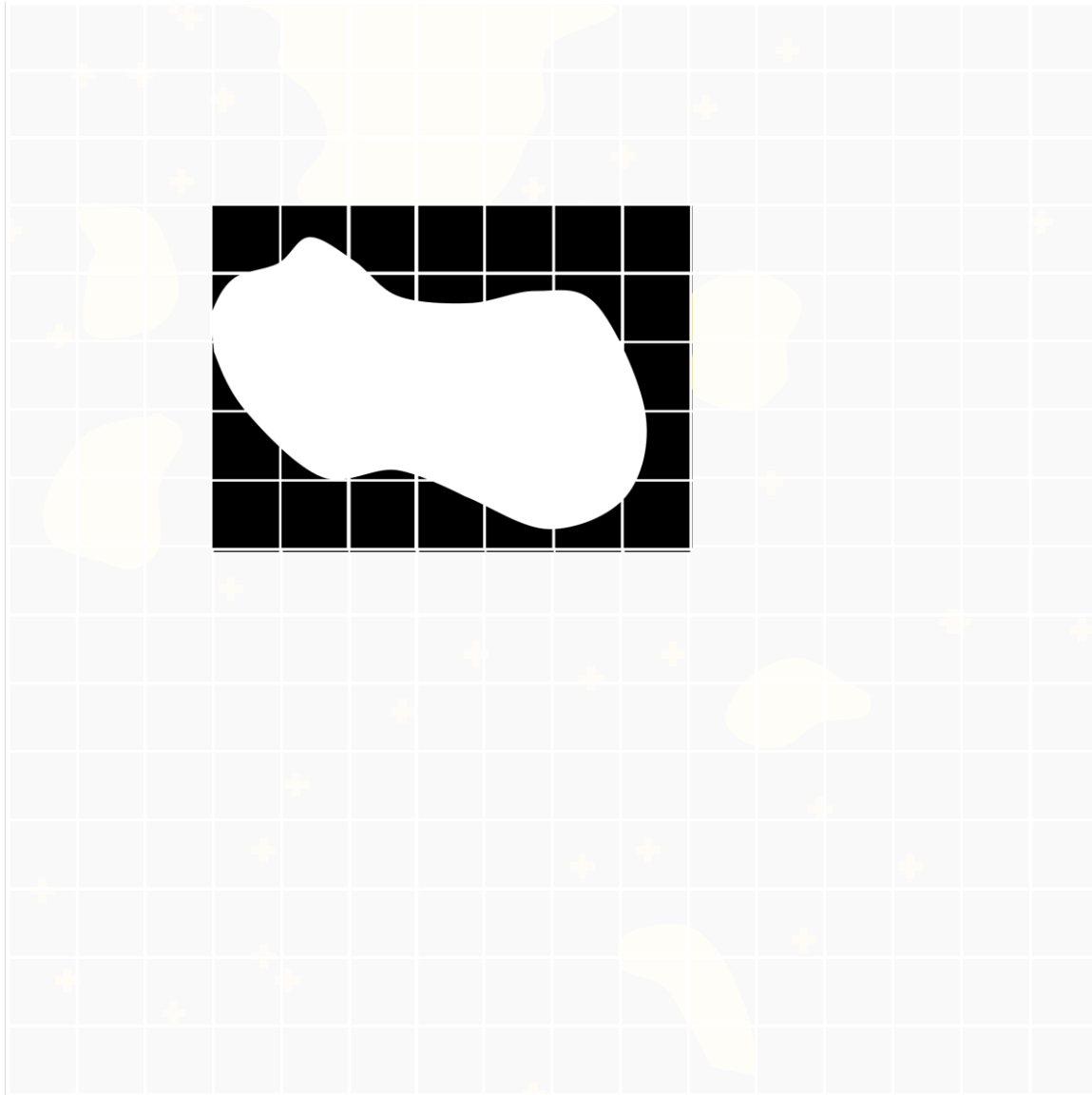


ISLANDs identification

# CLUSTER ANALYSIS

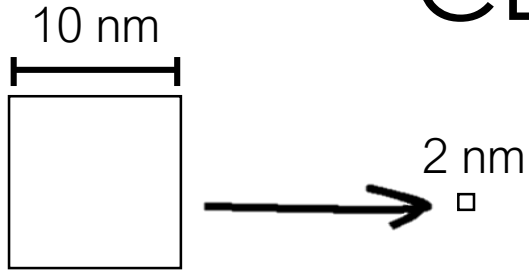10 nm

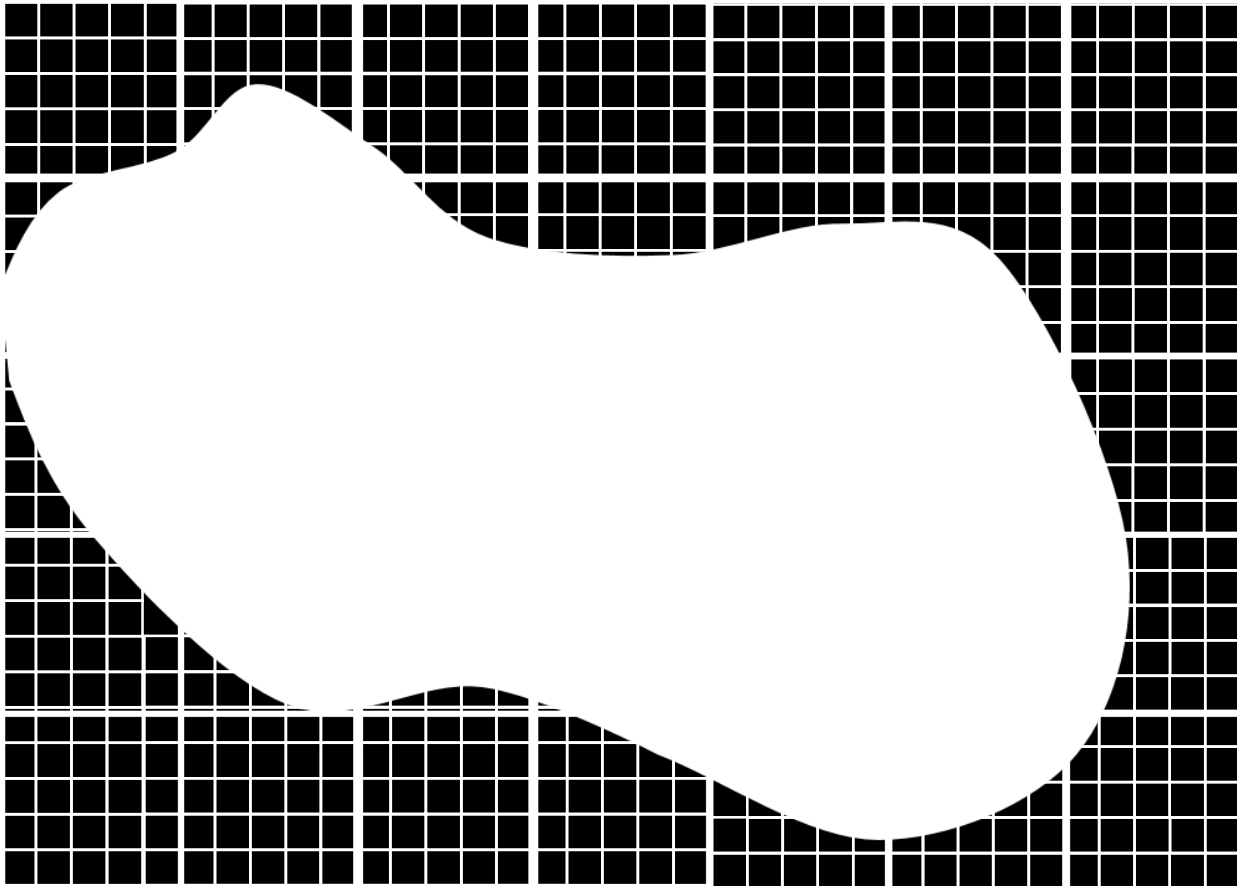It picks one Island

# CLUSTER ANALYSIS

10 nm



It picks one Island

# CLUSTER ANALYSIS

10 nm

2 nm

It picks one Island

It resizes it
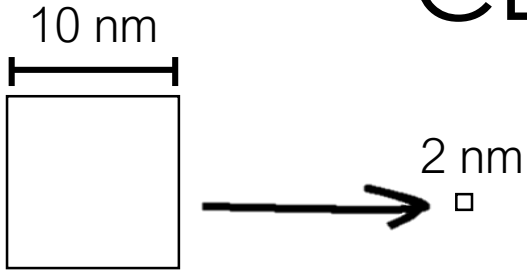
analysis_pixel_size [nm]

10

factor [use: analysis_pixel / factor]
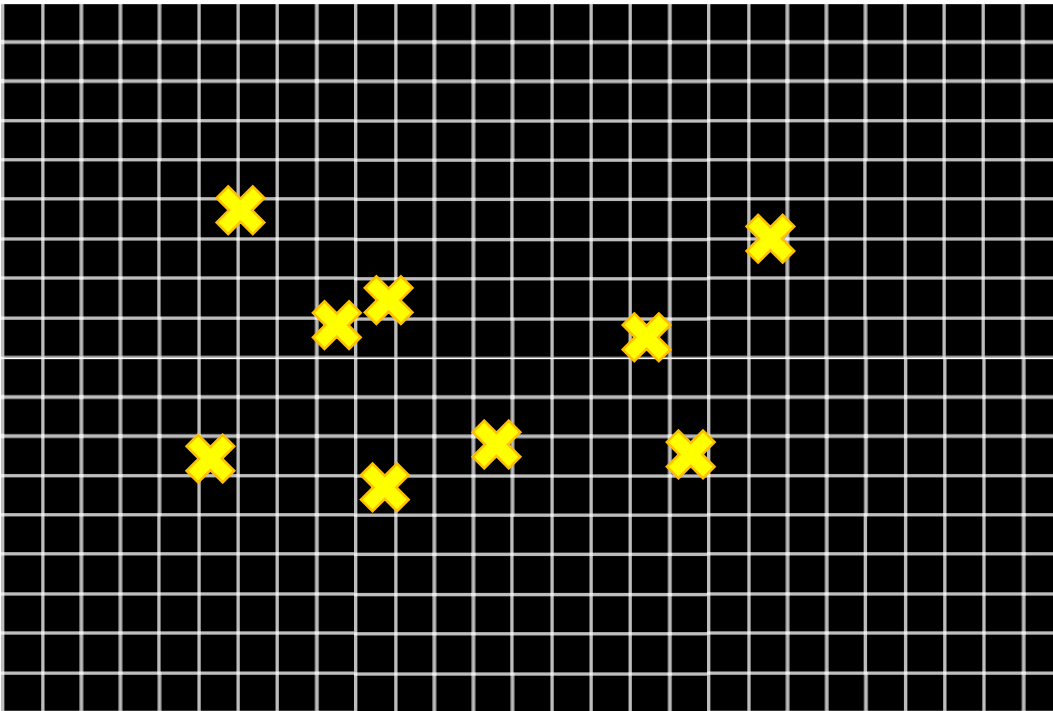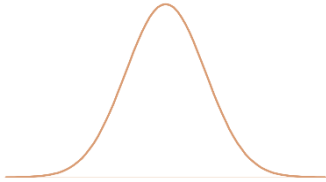
5

10 nm : **5** = 2 nm

# CLUSTER ANALYSIS

10 nm
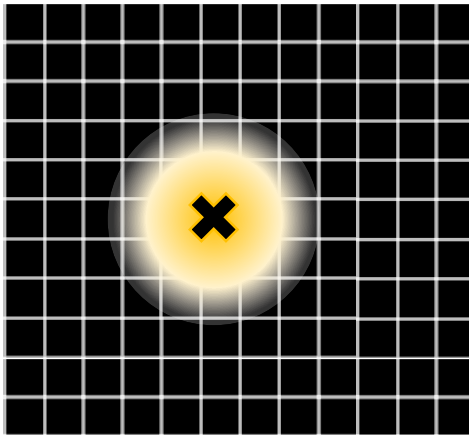
2 nm

It picks one Island

It resizes it

For each localization:

# CLUSTER ANALYSIS

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

It picks one Island

It resizes it

For each
localization:
It associates
a Gaussian curve

# CLUSTER ANALYSIS



$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

x, y localization coordinates

$$\sigma = \frac{\text{loc precision (9nm)}}{\text{new analysis px (2nm)}}$$
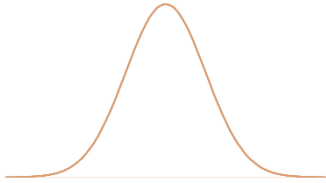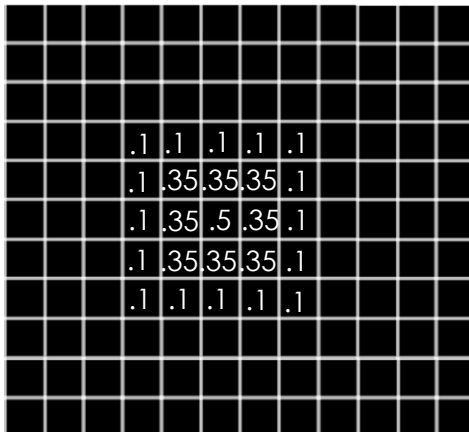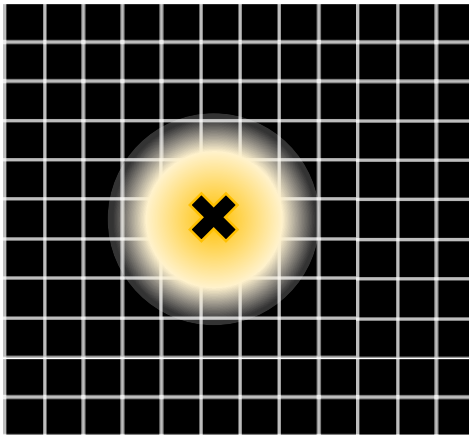
3σ

roi size = numSigma*σ

# CLUSTER ANALYSIS



It picks one Island

It resizes it
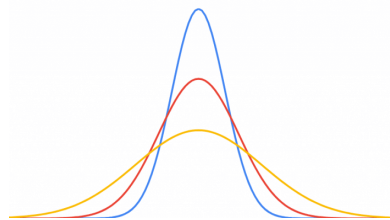
For each
localization:
It associates
a Gaussian curve

It sums the
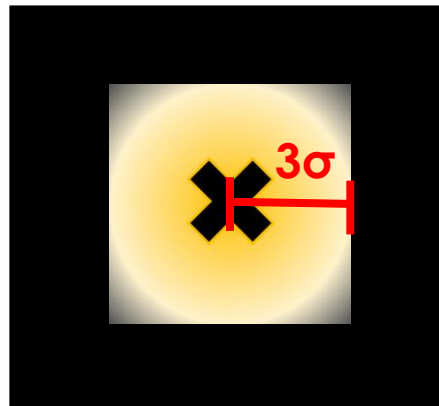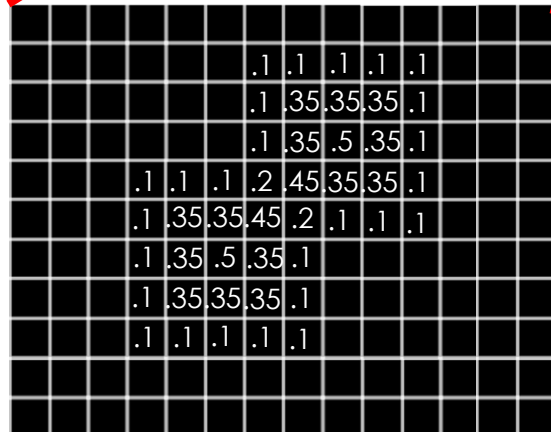Gaussians to
generate a
density map
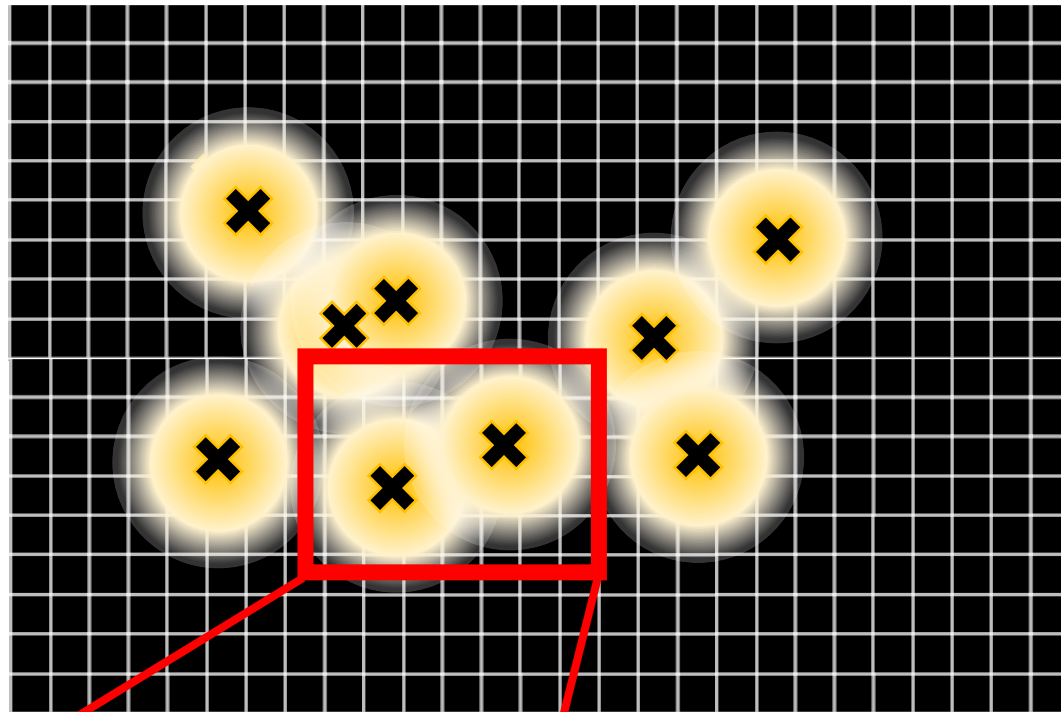
New matrix of the
summation of
Gaussian values
belonging to
localizations

# CLUSTER ANALYSIS
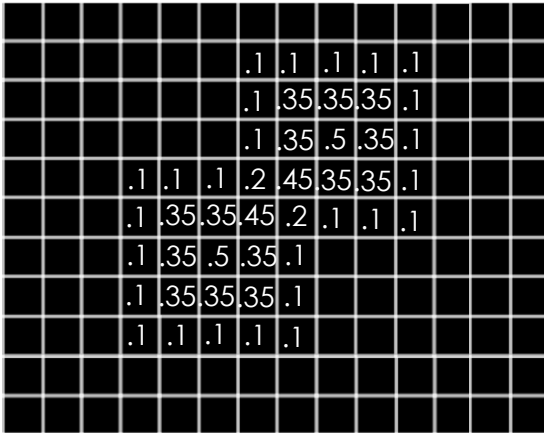


In each Island,
from the density matrix
the Matlab function **regionprops**
'magically' identifies centroids
of clusters

Then, the function **lik_sig** retrieves the Number of localizations belonging to each centroid, and the coordinates of the centroids in nm.

If the Number of localizations
belonging to each centroid is
≥ minimum_molecules_per_cluster

a cluster is defined.

# CLUSTER ANALYSIS

If region has

$n° cluster > 1$

$n° cluster = 1$

ISLAND

SINGLE cluster

**N° of localizations** per cluster are:
the sum of localizations belonging to the cluster

For each cluster, the algorithm calculates the coordinates **X, Y of the CENTROID** as:
the mean of localizations coordinates (x, y)

# CLUSTER ANALYSIS

**Area** of the cluster is approximated to a circle, of which

$$r = \frac{sd\ X + sd\ Y}{2} \qquad A = \pi r^2$$

*To convert px to nm, simply \*160 (original pixel size)*

Inside each Island, distance among all cluster centroids (X, Y) is calculated, and the smallest distance between each pair of clusters is given in nm as **NND**

$$distance = \sqrt{(X_b - X_a)^2 + (Y_b - Y_a)^2}$$

Of course, for Single clusters, NND will be 'Infinite' since for definition they have no close neighbours

# CLUSTER ANALYSIS

>**DistanceDualColor** $\longrightarrow$ To calculate NND among ALL clusters

NND is calculated using Centroids coordinates (X, Y) from EVERY cluster (InIsland and Single)

This script was originally made to calculate distances among different proteins (dual color).

For single color, the list of centroids is "doubled" and distances are calculated between the two identical lists. (distances = 0 are discarded.)