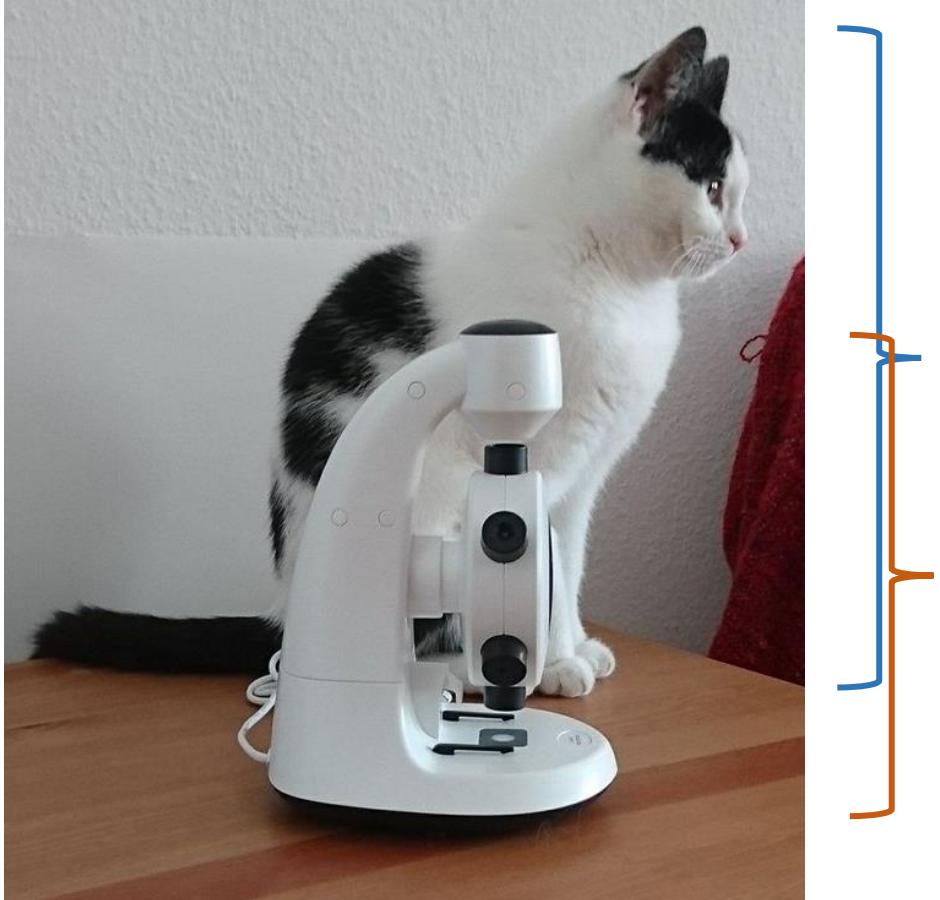


# Introduction to Bio-Image Analysis

Robert Haase

Reusing some materials made by Beth Cimini,  
Anne Carpenter & colleagues (Broad Institute)

- Deriving quantitative information from images of biological samples taken with microscopes



cat height = 1.5 x microscope height

- Deriving quantitative information from images of biological samples taken with microscopes + visualization

The figure displays three sequential steps in a quantitative bio-image analysis workflow using the napari software:

- Segmentation:** The first panel shows a grayscale image of Drosophila nuclei with a color-coded segmentation mask overlaid. The text "Drosophila, mitotic wave" is overlaid on the image. A sidebar lists processing steps: Result of top\_hat\_box [0], Result of connected, Result of threshold\_, Result of gaussian\_t, Result of top\_hat\_b, and CalibZAPWfixed\_000.
- Pixel count map:** The second panel shows a color-coded map where each nucleus is assigned a unique color based on its size. The text "Pixel count map" is overlaid. A sidebar lists: Noise removal, Background removal, Filter, Combine, Transform, Projection, Binarize, Label, Label processing, Label measurem..., Map, and Mesh.
- How many nuclei are there?** The third panel shows the original grayscale image with the text "How many nuclei are there?" overlaid. A sidebar lists: Result of top\_hat\_box [0] and Result of label\_pixel\_count\_map [172 - 13].

**How is the nuclei marker distributed?**

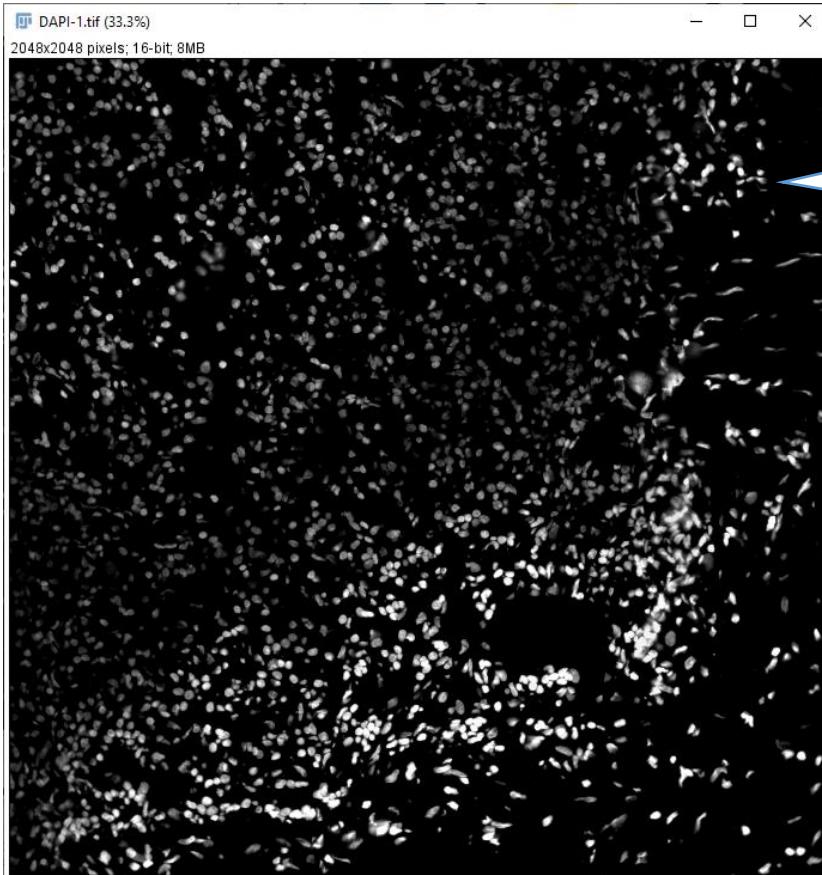
**How many nuclei are there?**

**Pixel count map**

**How large are these nuclei?**

# Objective bio-image analysis

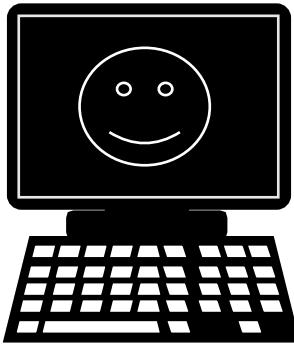
- Measurements should be objective, not influenced by human interpretation



Nuclei in this image are ...

... more dense than in this image.

Use automation for less subjective analysis.



December 2022

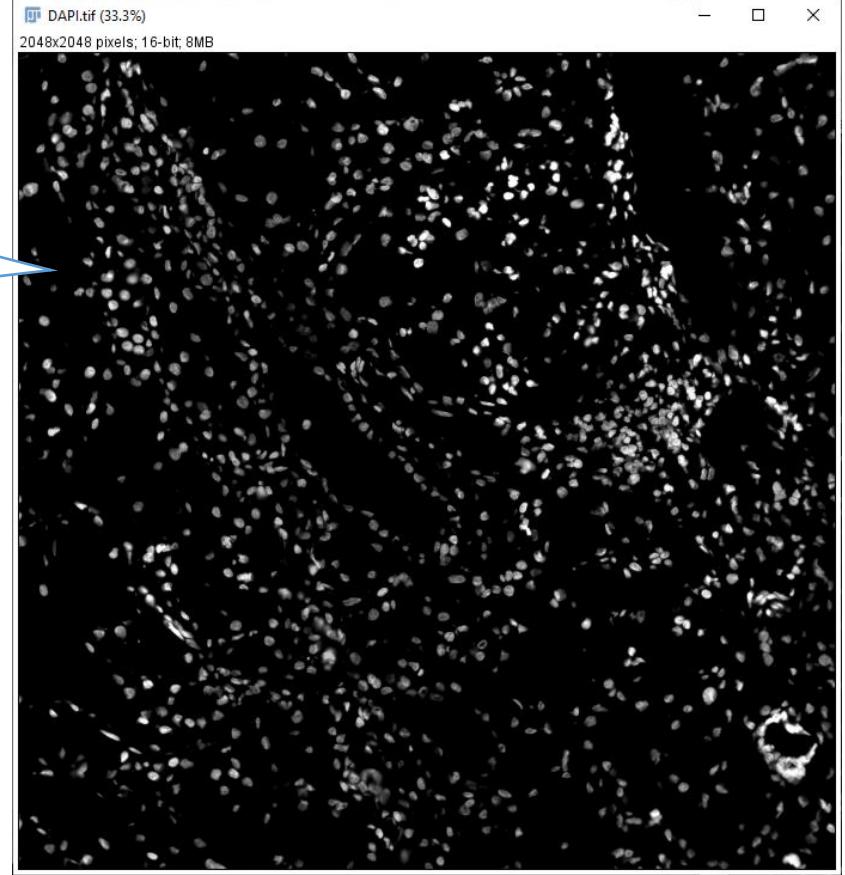
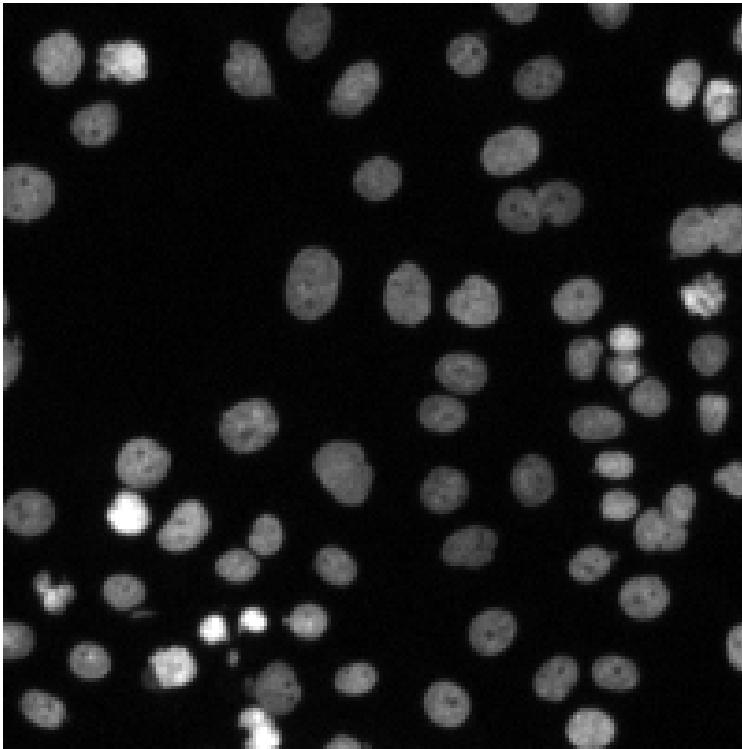


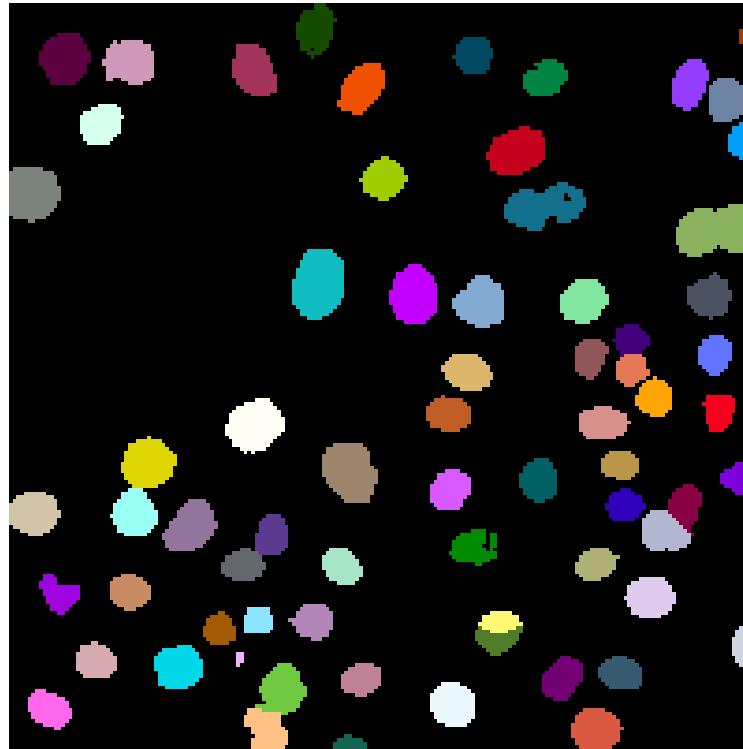
Image data source: Pascual-Reguant, Anna. (2021). Immunofluorescence staining of a human kidney (#2, peri-tumor area) obtained by MELC [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.4434462> licensed CC-BY 4.0

- Algorithms must be reliable (trustworthy). Visualization helps gaining trust in automated methods.

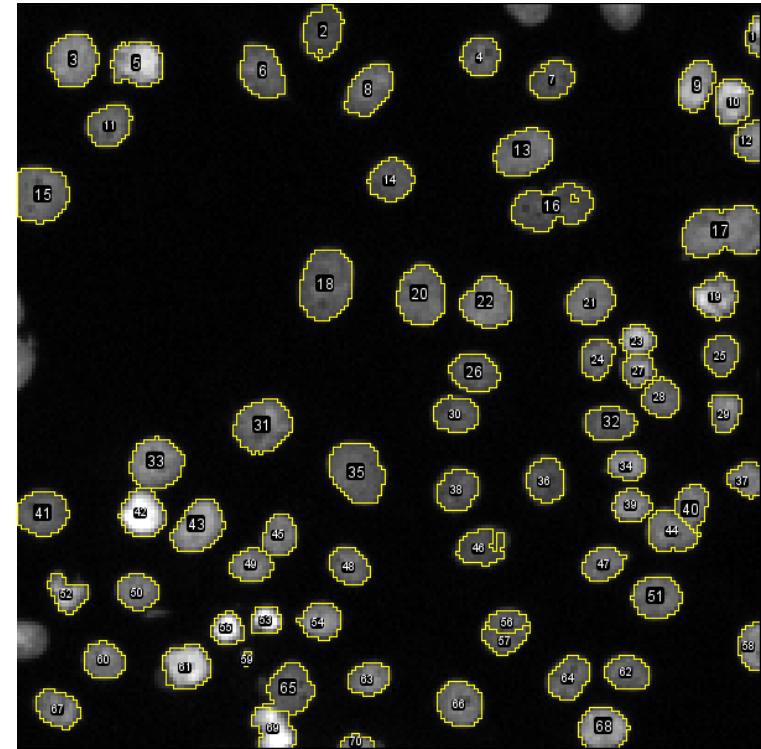
Original image



Label image



Overlay

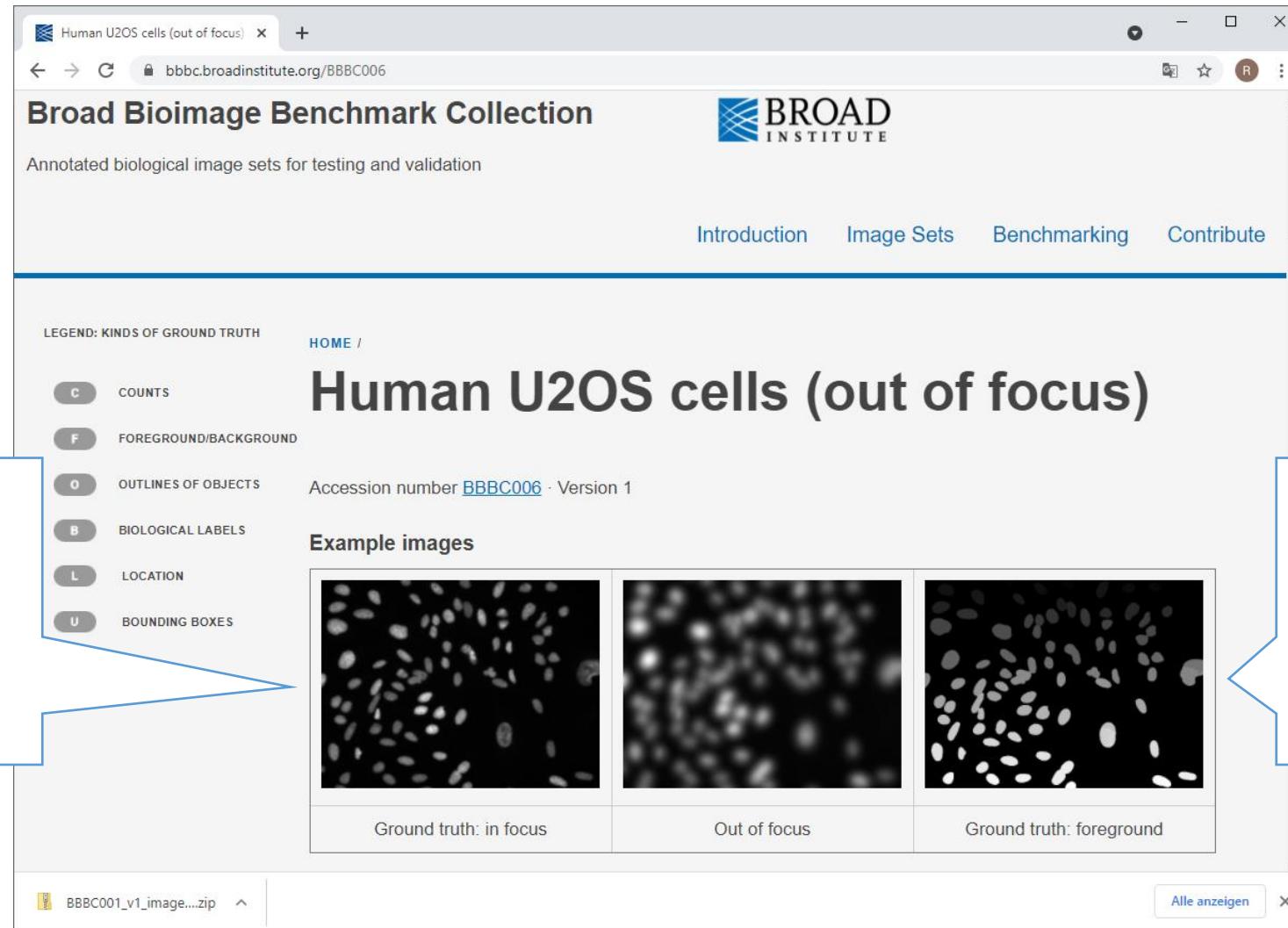


There are 70 nuclei  
in this image.

Image Data source: [BBBC001v1](#) [Carpenter et al., *Genome Biology*, 2006] from December 2022 the Broad Bioimage Benchmark Collection [[Ljosa et al., Nature Methods, 2012](#)]

# Reliable bio-image analysis

- Algorithms must be reliable (validated methods). Publicly available benchmark data sets allow to compare algorithms on common data.



Original image data

"Ground truth" label images

# Reproducible bio-image analysis

- “The image data was analyzed with ImageJ.”

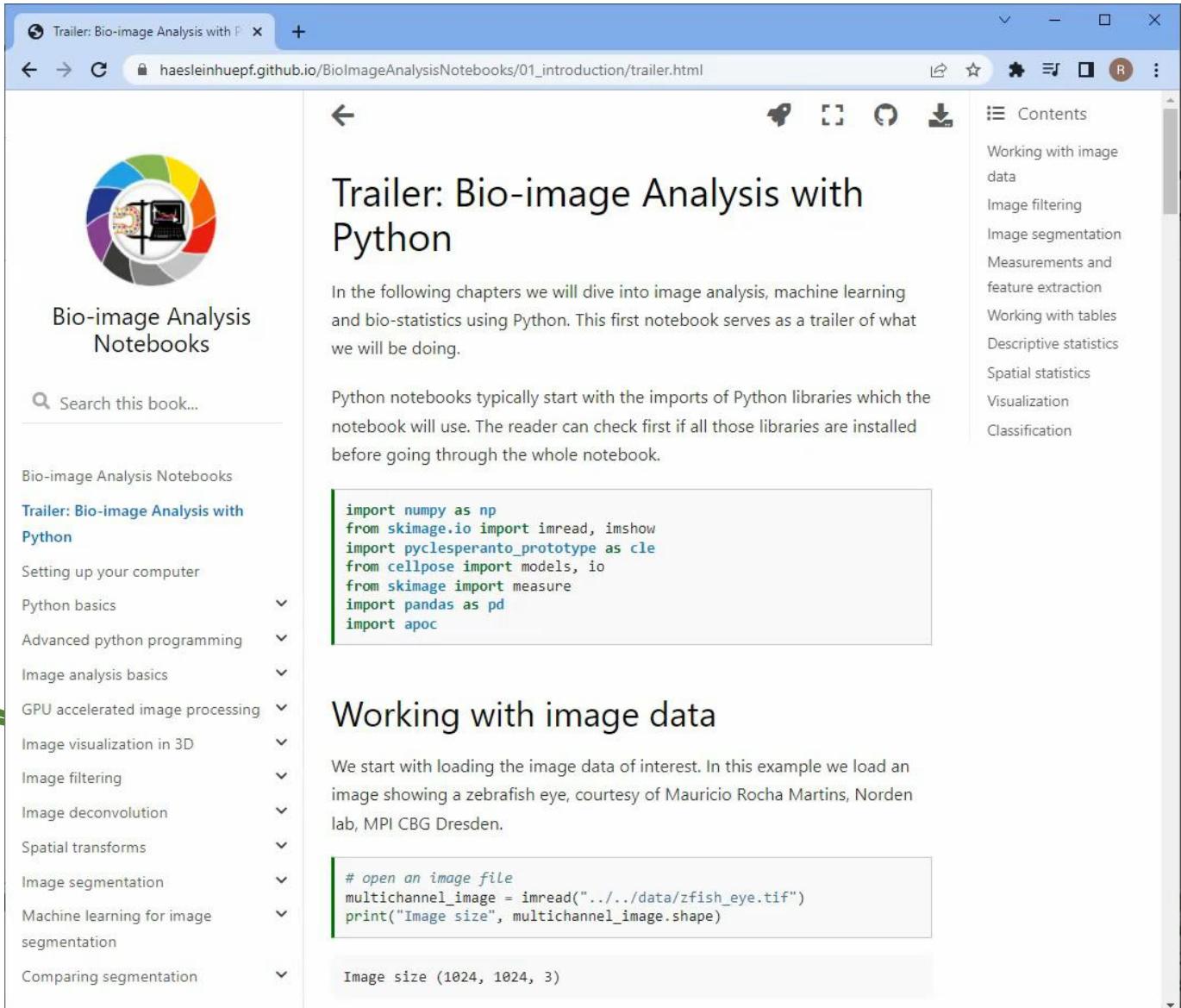
Can you reproduce  
what they did?

# Reproducible bio-image analysis

- “The image data was analyzed with ImageJ.”

Can you reproduce what they did?

Can you reproduce what they did?



The screenshot shows a web browser displaying a trailer for a bio-image analysis notebook. The title is "Trailer: Bio-image Analysis with Python". The page content includes a brief introduction, code snippets, and a sidebar with a table of contents.

**Title:** Trailer: Bio-image Analysis with Python

**Content:**

In the following chapters we will dive into image analysis, machine learning and bio-statistics using Python. This first notebook serves as a trailer of what we will be doing.

Python notebooks typically start with the imports of Python libraries which the notebook will use. The reader can check first if all those libraries are installed before going through the whole notebook.

```
import numpy as np
from skimage.io import imread, imshow
import pyclesperanto_prototype as cle
from cellpose import models, io
from skimage import measure
import pandas as pd
import apoc
```

**Working with image data**

We start with loading the image data of interest. In this example we load an image showing a zebrafish eye, courtesy of Mauricio Rocha Martins, Norden lab, MPI CBG Dresden.

```
# open an image file
multichannel_image = imread("../data/zfish_eye.tif")
print("Image size", multichannel_image.shape)
```

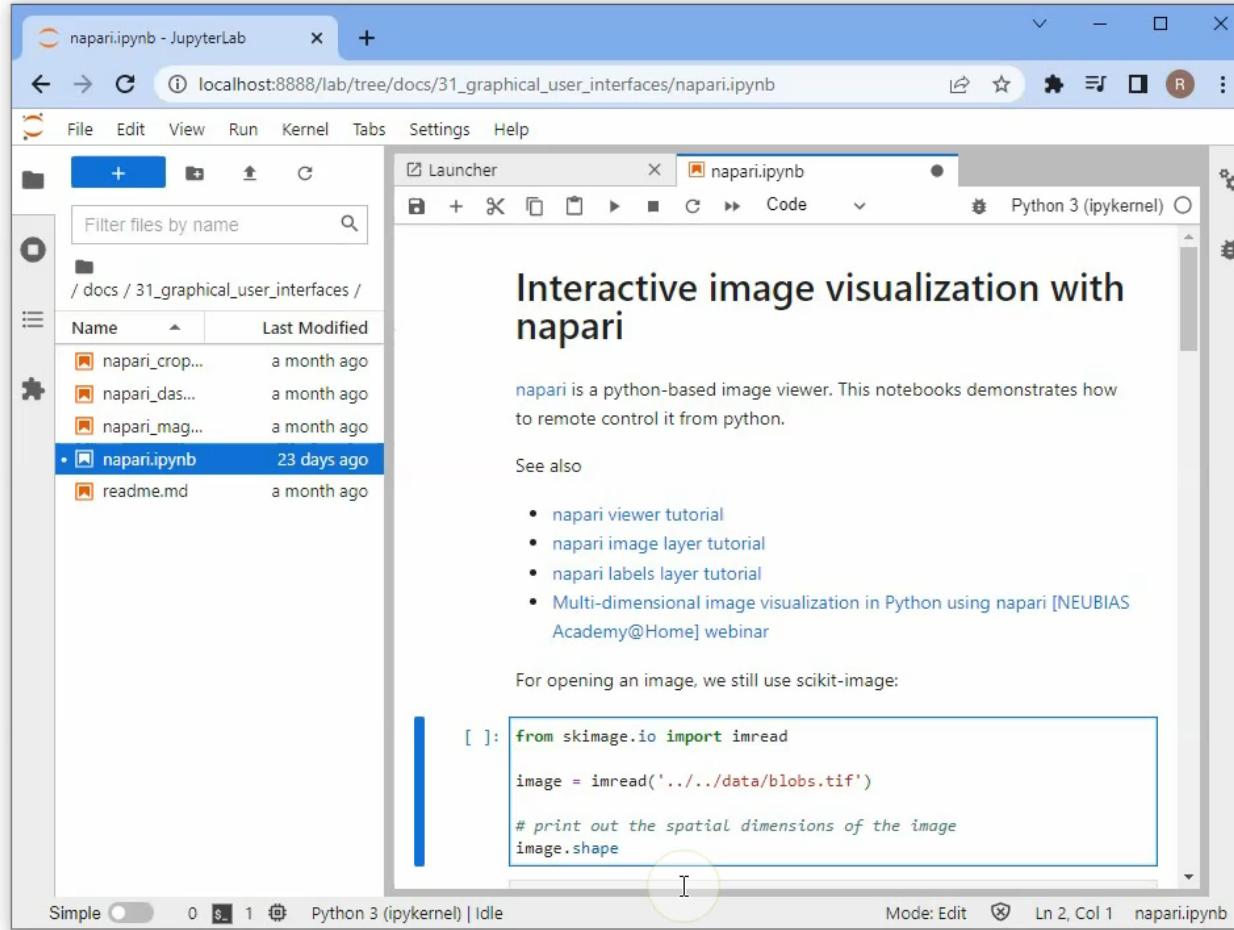
Image size (1024, 1024, 3)

**Table of Contents:**

- Bio-image Analysis Notebooks
- Trailer: Bio-image Analysis with Python
- Setting up your computer
- Python basics
- Advanced python programming
- Image analysis basics
- GPU accelerated image processing
- Image visualization in 3D
- Image filtering
- Image deconvolution
- Spatial transforms
- Image segmentation
- Machine learning for image segmentation
- Comparing segmentation

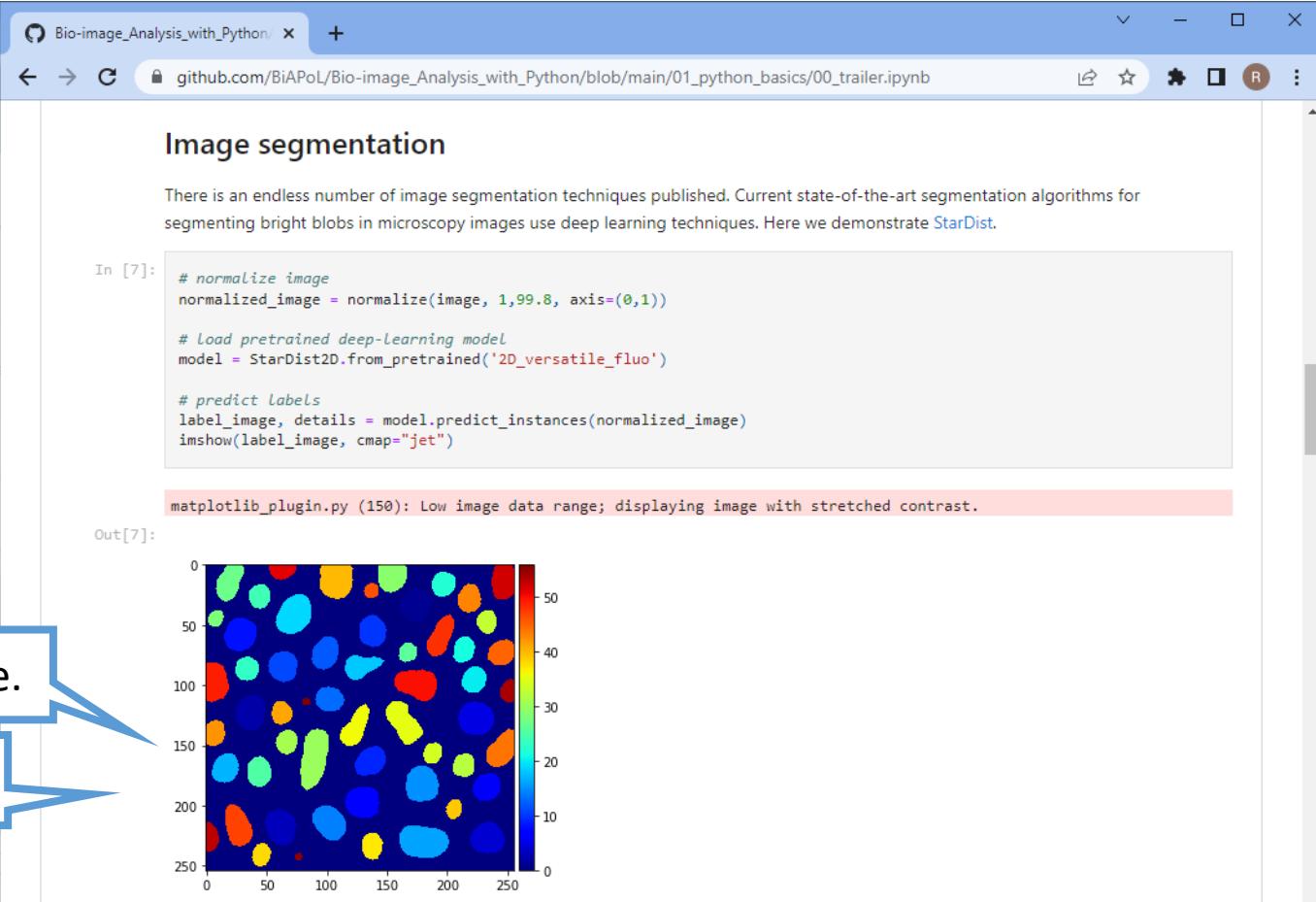
# Interactivity versus reproducibility

- Remote controlling graphical user interfaces



- Compared to wet-lab experiments, image analysis (in-silico) experiments are typically repeatable.
- In wet-lab experiments, samples may get destroyed while executing the experiment.
- Repeatability is a property of the experiment. You cannot improve repeatability by better documentation.

- However, you need to pay some extra attention to have repeatable image processing
  - Scripts need to be written in a way that their execution is repeatable.
  - Test it!



The screenshot shows a Jupyter Notebook cell titled "Image segmentation". The cell contains Python code for image segmentation using the StarDist2D library. The code includes normalization of the image, loading a pretrained model, and predicting labels. A warning message from matplotlib is displayed: "matplotlib\_plugin.py (150): Low image data range; displaying image with stretched contrast." Below the code, a heatmap visualizes the segmented image with a color scale from 0 to 50. Three blue callout boxes with arrows point from the text below to the code and the resulting image:

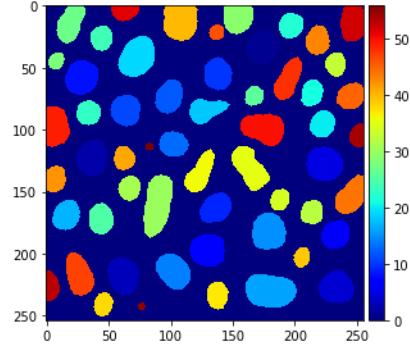
- "Run it once." points to the "# predict labels" line in the code.
- "Run it twice." points to the "imshow(label\_image, cmap="jet")" line in the code.
- "Are results identical?" points to the heatmap.

```
# normalize image
normalized_image = normalize(image, 1, 99.8, axis=(0,1))

# load pretrained deep-learning model
model = StarDist2D.from_pretrained('2D_versatile_fluo')

# predict labels
label_image, details = model.predict_instances(normalized_image)
imshow(label_image, cmap="jet")

matplotlib_plugin.py (150): Low image data range; displaying image with stretched contrast.
```

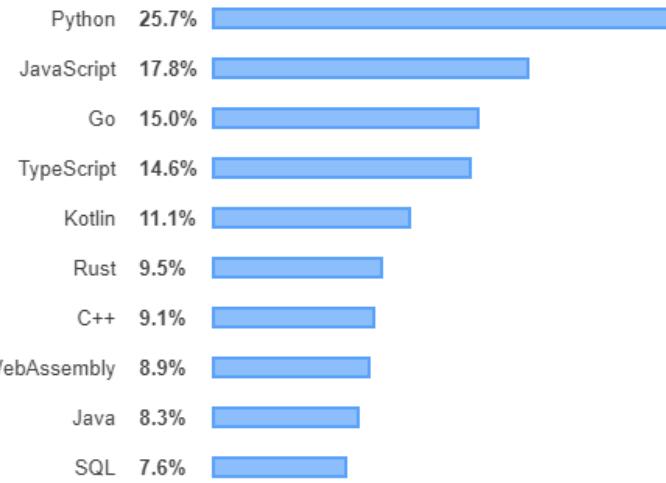


# Why Python?

- We will focus on Python programming.
- “Python is the most wanted language for the third year in a row, meaning that developers who do not yet use it say they want to learn it.” Stackoverflow survey, 2019

## Most Loved, Dreaded, and Wanted

### Most Loved, Dreaded, and Wanted Languages



```
b = 3  
c = a + b
```

```
print(c)
```

```
8
```

```
d = 6  
e = 7  
f = a * d  
g = f / e  
h = 1 + g
```

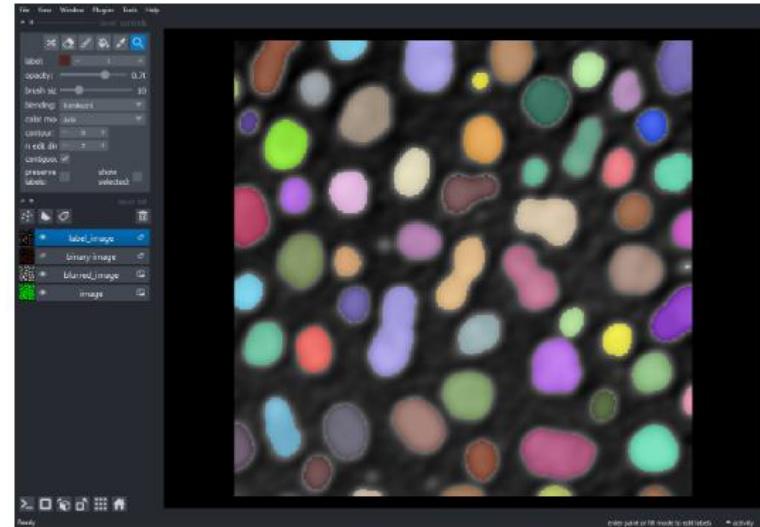
```
print(h)
```

```
5.285714285714286
```

```
from skimage.measure import label  
label_image = label(binary_image)  
  
# add labels to viewer  
label_layer = viewer.add_labels(label_image)
```

You can visualize labelled objects as overlay (per default)

```
napari.utils.nbscreenshot(viewer)
```



# Introduction to bio-image analysis

- Bio-image analysis is supposed to be
  - **Quantitative**
    - We derive numbers from images which describe physical properties of the observed sample.
  - **Objective**
    - The derived measurement does not depend on who did the measurement. The measurement is free of interpretation.
  - **Reliable (trustworthy / validated)**
    - We are confident that the measurement is describing what it is supposed to describe.
  - **Reproducible**
    - Somebody else can do the experiment under *different conditions* and gets similar measurements. For this, documentation is crucial!
  - **Repeatable**
    - We can do the same experiment twice under the *same conditions* and get similar measurements.

# Image analysis is part of the experiment



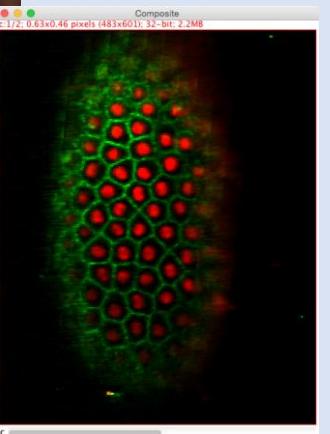
**PoL**  
Physics of Life  
TU Dresden



Observation

$$p_{ij}(t) = \frac{[\tau_{ij}(t)]^\alpha \cdot [\eta_{ij}]^\beta}{\sum_{j=1}^n [\tau_{ij}(t)]^\alpha \cdot [\eta_{ij}]^\beta}$$

Modeling



Imaging

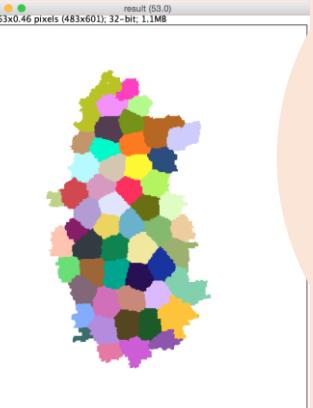


Image processing

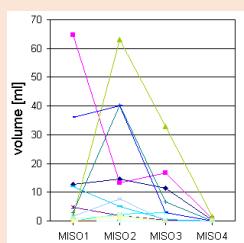
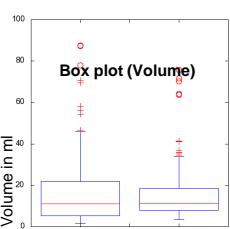


Image analysis  
Bio-statistics



ember 2022

- Think about how to analyze your images before starting the experiment.
  - Consider adapting your experiment so that quantitative image analysis can be performed easily.
  - Do small test-experiments.
- Think about controls, counter-proves, an easy to falsify null-hypothesis.
  - Be a lazy scientist. Do simple experiments.
- How can you exclude yourself from the experiment?
  - Think of blinding yourself or fully automate analysis.
- One experiment usually answers one question. Or less.

- To measure cell volume, we need cell segmentation.

True

False

- To *count cells*, we need cell segmentation.

True

False

- To count cells over time, we need *cell tracking*.

True

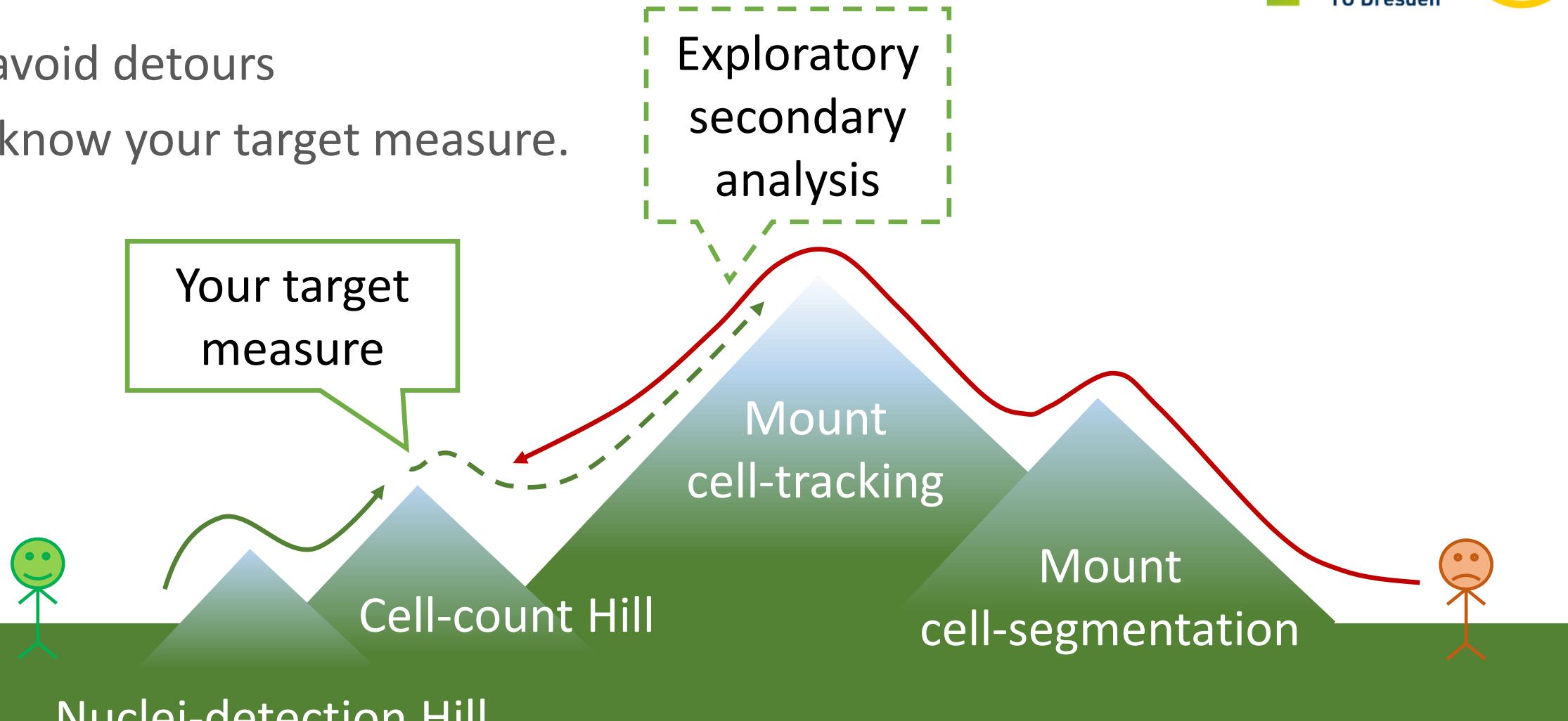
False

- For *cell tracking*, we need cell segmentation.

True

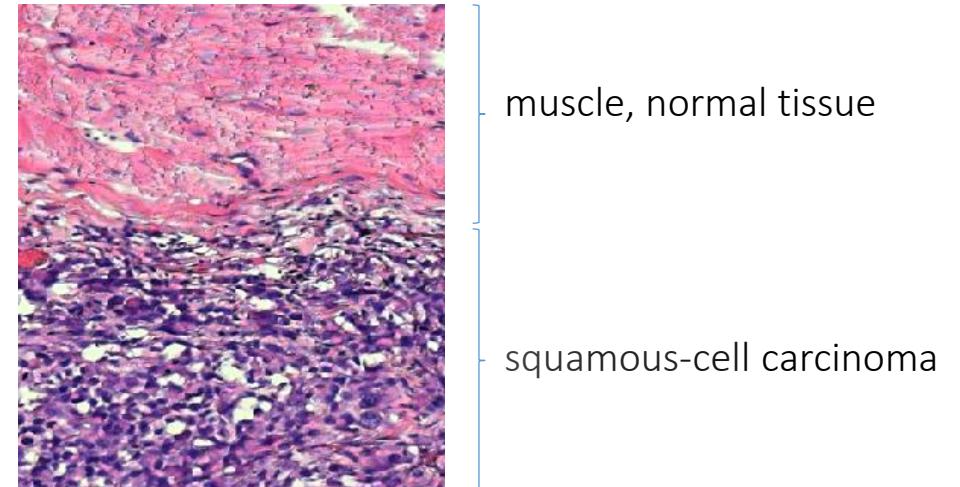
False

- ... to avoid detours
- Hint: know your target measure.

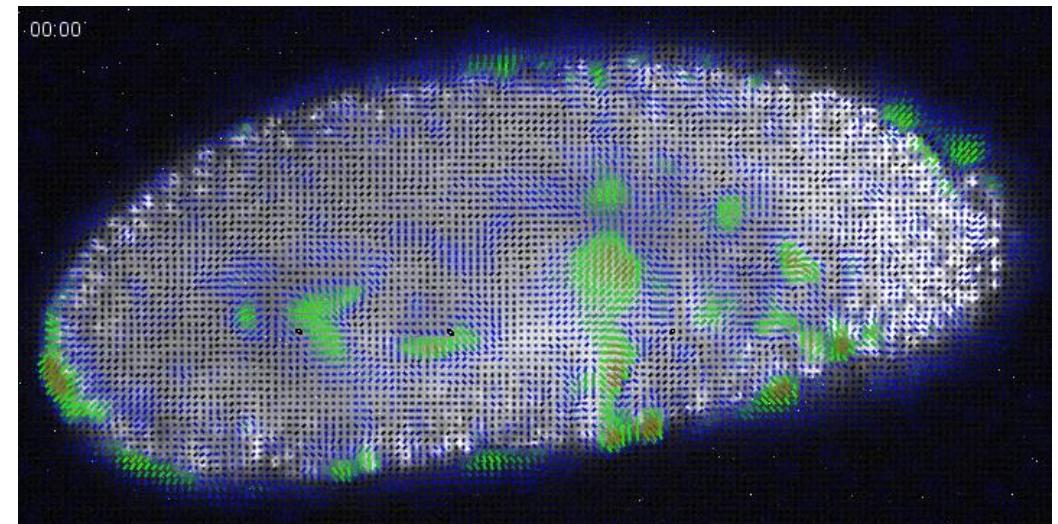


# Common questions

- Typical questions bio-image analysts deal with
  - Is signal intensity different under varying conditions?
  - How many cells are in my image?
  - How high is cell density?
    - Bio-statistics / medicine
  - How are different tissues characterized?
    - Machine learning



- Typical questions bio-image analysts struggle with
  - What force drives the observed processes?
  - What is the lineage tree of one particular cell?
  - Are observation A and observation B related?
  - Are structures observed in different color channels colocalized?



# Hypothesis-driven quantitative biology

- Hypothesis: Cell shape can be influenced by modifying X.

Should we use a different segmentation algorithm?

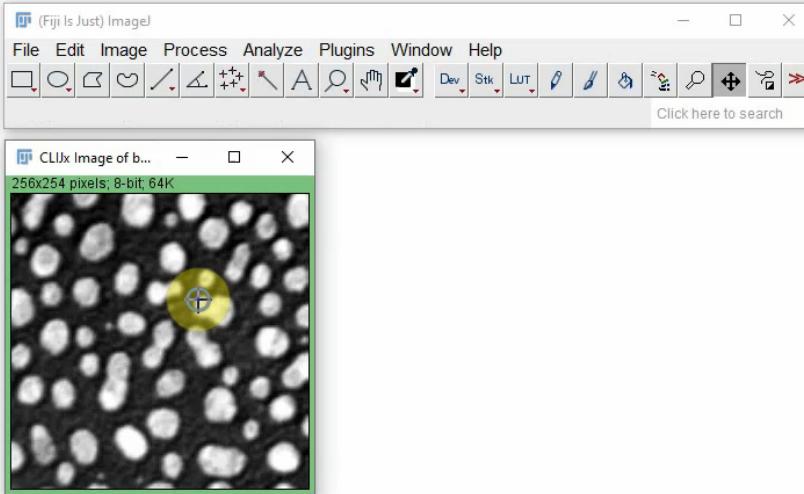
- Sample preparation
- Imaging
- Cell segmentation
- Circularity measurement
- Statistics

Shall we use a different microscope?

Is circularity the right parameter to measure?

# Hypothesis-driven quantitative biology

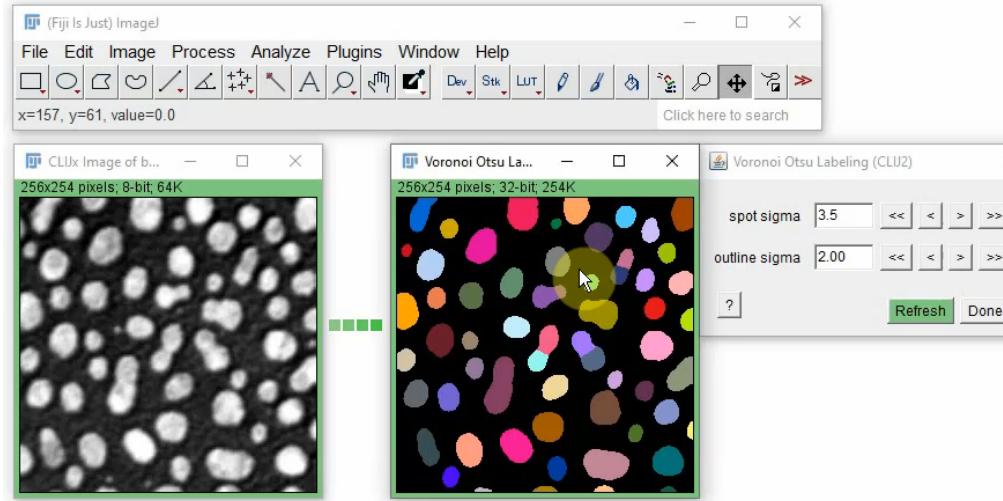
- Should we use a different segmentation algorithm?



<https://clij.github.io/>

# Hypothesis-driven quantitative biology

Is circularity the right parameter to measure?



# Hypothesis generating quantitative biology

- Hypothesis: Cell shape can be influenced by modifying X.
- Question: Which image-derived parameter is influenced when modifying X? Why?
  - Sample preparation
  - Imaging
  - Cell segmentation algorithm A, algorithm B, algorithm C
- Measurement of circularity, solidity, elongation, extend, texture, intensity, topology ...
  - Statistics

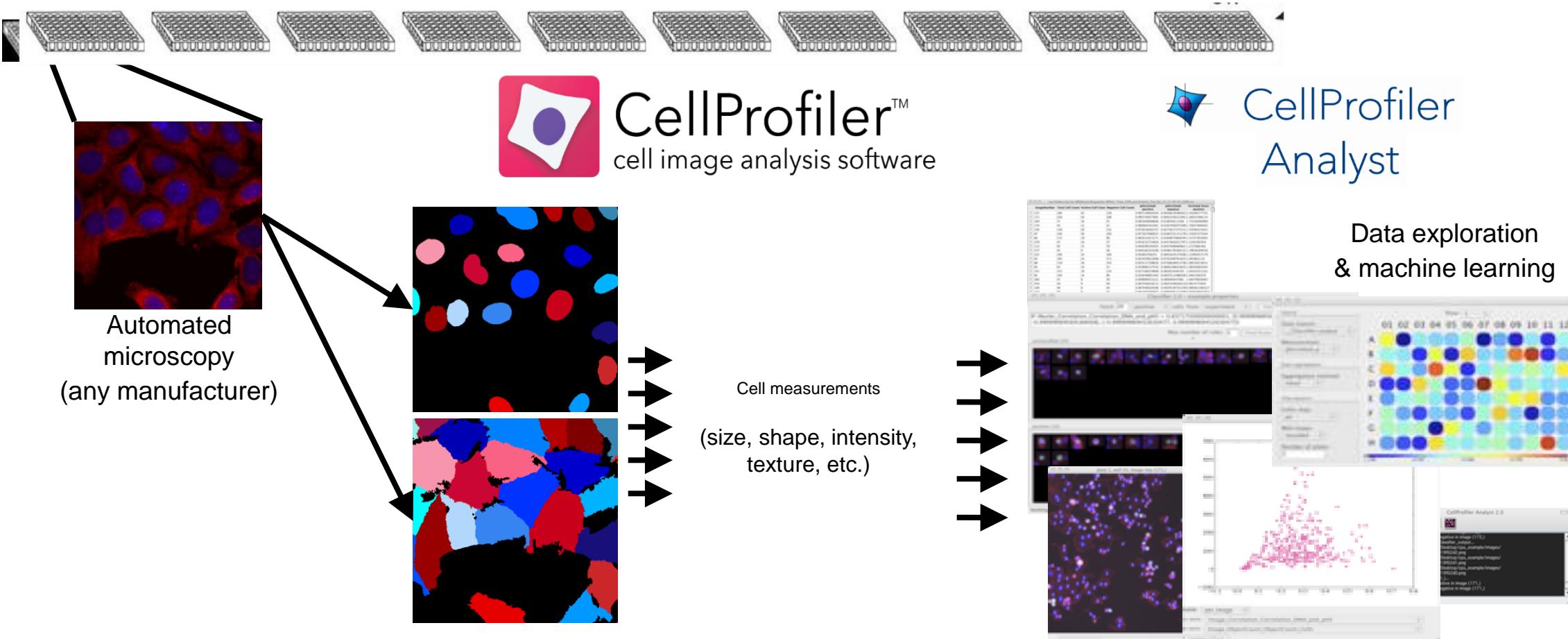
Which segmentation algorithms allow measurements that show a relationship with X?

Why?

Which parameter shows any relationship with X?

# Cell Profiler

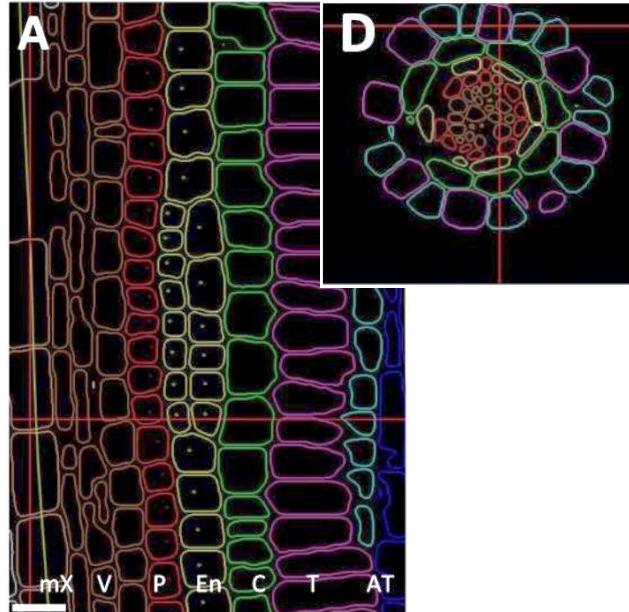
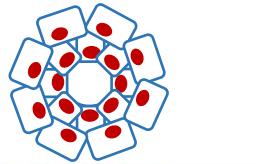
- Cells or organisms in multiwell plates, each well treated with a gene or chemical perturbant



# Image analysis beyond segmentation

- Spatial pattern / image analysis for cell biology

Membranes  
Nuclei



Structural biology of plants<sup>1</sup>

15 years ago!

[Joint Pattern Recognition Symposium](#)  
DAGM 2006: [Pattern Recognition](#) pp 182-191 | [Cite as](#)

Fast Scalar and Vectorial Grayscale Based Invariant Features for **3D Cell Nuclei Localization and Classification**

Authors

Authors and affiliations

Janina Schulz, Thorsten Schmidt, Olaf Ronneberger, Hans Burkhardt, Taras Pasternak, Alexander Dovzhenko, Klaus Palme

Conference paper

10

1.8k

Citations Downloads

Part of the [Lecture Notes in Computer Science](#) book series (LNCS, volume 4174)

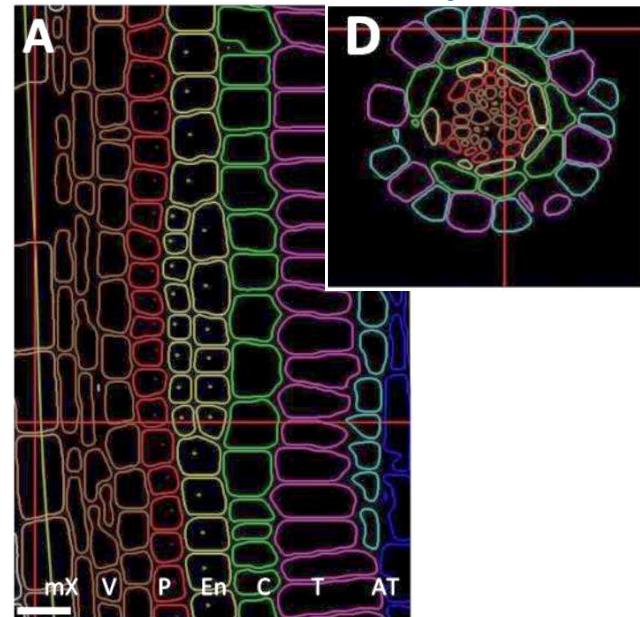
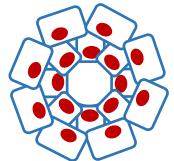


1) Image data source: Pasternak et al. ([CC-BY 4.0](#)) <https://www.biorxiv.org/content/10.1101/2021.01.01.425043v2>  
@Haesemhuepr  
@PoLDresden

# Image analysis beyond segmentation

- Spatial pattern / image analysis for cell biology

Membranes  
Nuclei



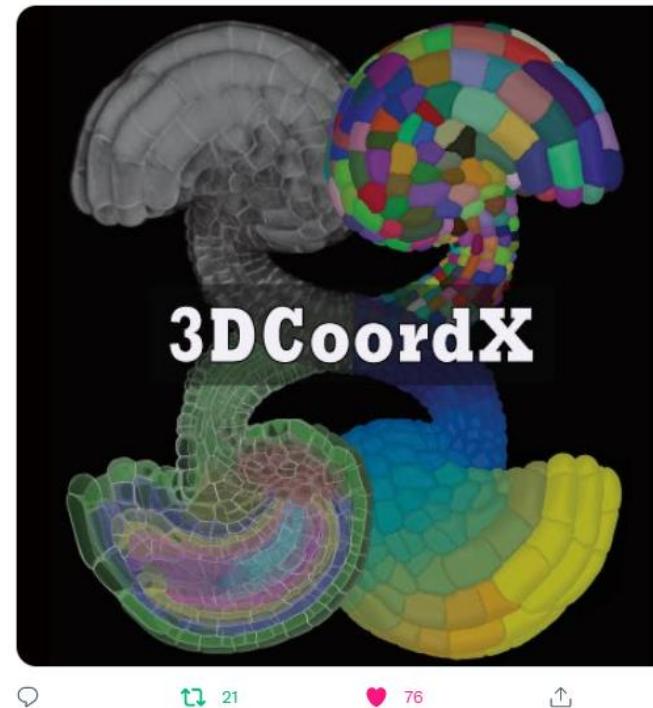
Structural biology of plants<sup>1</sup>



Kay Schneitz @PlantDevTUM · Mar 29

Just in case you hadn't yet noticed our supercool new toolbox for MorphoGraphX. Provides organ-centric spatial context to cellular features in 3D digital plant organs.  
[tinyurl.com/2n6rb6ds](http://tinyurl.com/2n6rb6ds)  
@athul\_r\_v @StraussSoe @tejasvinee\_m @kareninglelee @RichardSmithLab @PlantPhys

...



*Plant Physiology*®

Issues

Advance Articles

More Content ▾

Submit ▾

Purchase

About ▾

All Plant Physiology

Advanced Search

#### Article Contents

Abstract

Author notes

Supplementary data

#### ACCEPTED MANUSCRIPT

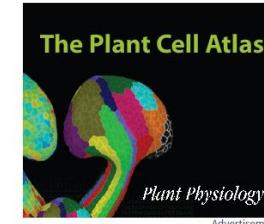
The annotation and analysis of complex 3D plant organs using 3DCoordX

Athul Vijayan, Soeren Strauss, Rachele Tofanelli, Tejasvinee Atul Mody, Karen Lee, Miltos Tsiantis, Richard S Smith, Kay Schneitz Author Notes

*Plant Physiology*, kiac145, <https://doi.org/10.1093/plphys/kiac145>

Published: 28 March 2022 Article history ▾

PDF Split View Cite Permissions Share ▾



American Society of Plant Biologists

Plant Physiology

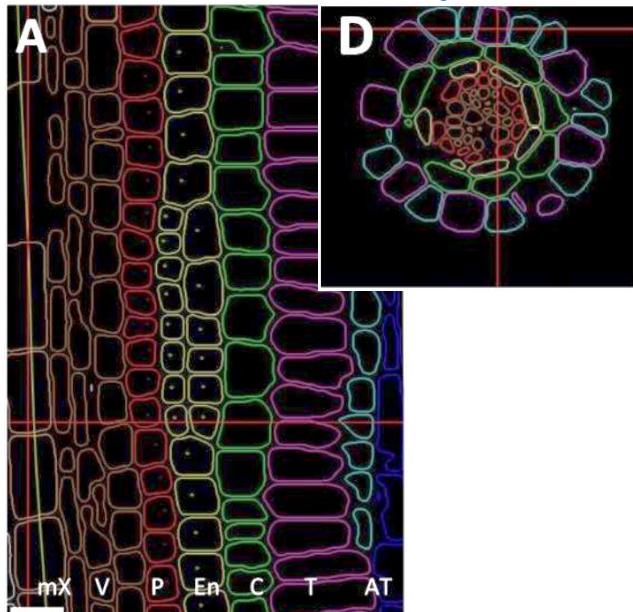
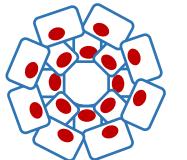
Advertisement



# Image analysis beyond segmentation

- Spatial pattern / image analysis for cell biology

Membranes  
Nuclei



Structural biology of plants<sup>1</sup>

**Cell Reports Resource**

**CytoMAP: A Spatial Analysis Toolbox Reveals Features of Myeloid Cell Organization in**

**nature methods**

Caleb R. Stoltzfus,<sup>1</sup> Jakub Filip, Miranda R. Lyons-Cohen,<sup>1</sup> Jes...  
<sup>1</sup>Department of Immunology, University of Washington, Seattle, Washington, USA.  
<sup>2</sup>Seattle Children's Research Institute, Seattle, Washington, USA.  
<sup>3</sup>Department of Pediatrics, University of Washington, Seattle, Washington, USA.  
<sup>4</sup>School of Medicine, University of Washington, Seattle, Washington, USA.  
<sup>5</sup>Roche Innovation Center Munich, Penzberg, Germany.  
<sup>6</sup>Roche Innovation Center Zurich, Fällanden, Switzerland.  
<sup>7</sup>Lead Contact.  
\*Correspondence: geremy@uw.edu.  
<https://doi.org/10.1101/2021.01.01.425043v2>

Explore content ▾ About the journal ▾ Publish with us ▾

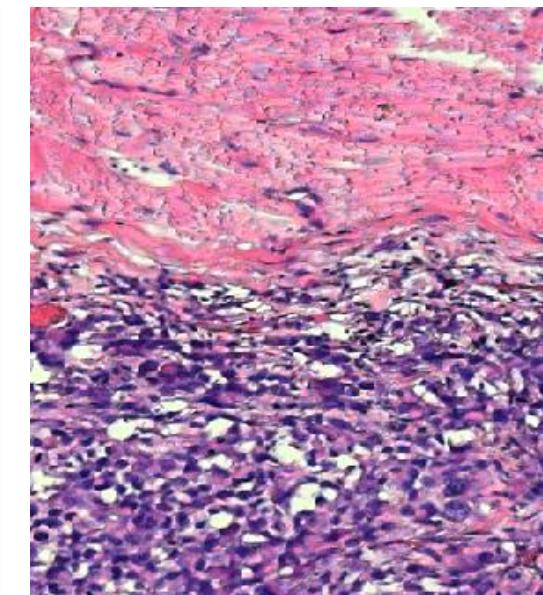
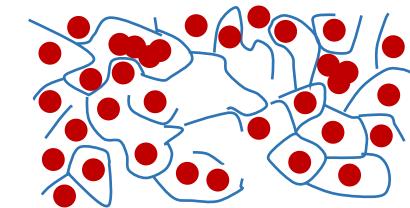
nature > nature methods > articles > article

Article | Open Access | Published: 31 January 2022

**Squidpy: a scalable framework for spatial omics analysis**

Giovanni Palla, Hannah Spitzer, Michal Klein, David Fischer, Anna Christina Schaar, Louis Benedikt Kuemmerle, Sergei Rybakov, Ignacio L. Ibarra, Olle Holmberg, Isaac Virshup, Mohammad Lotfollahi, Sabrina Richter & Fabian J. Theis

*Nature Methods* 19, 171–178 (2022) | [Cite this article](#)  
19k Accesses | 3 Citations | 375 Altmetric | [Metrics](#)



Cancer research / histology

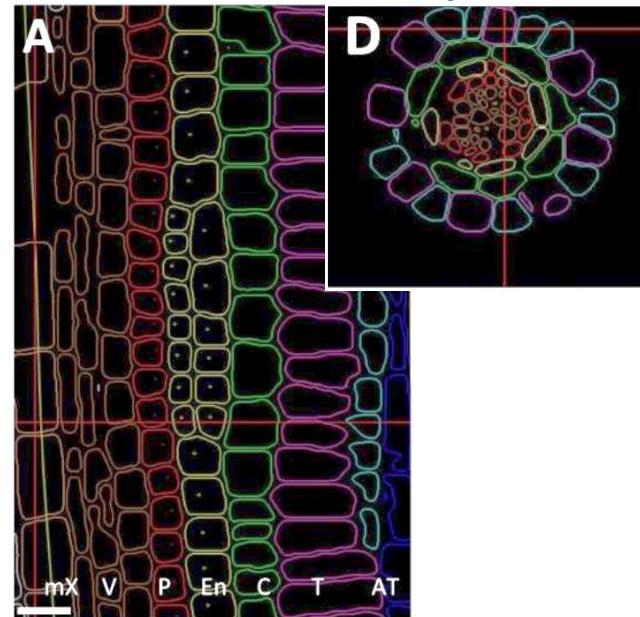
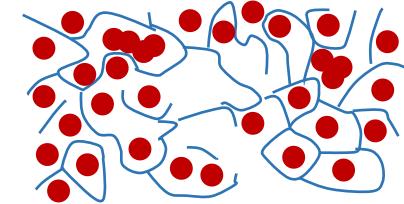
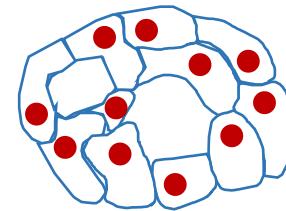
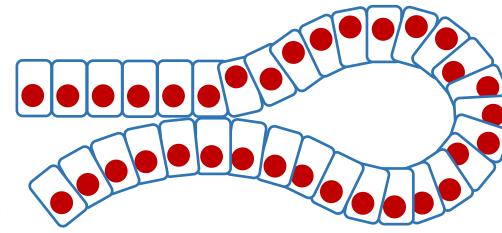
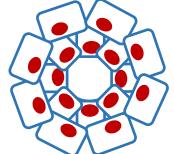


- @haesleinhuepf<sup>1</sup>) Image data source: Pasternak et al. (CC-BY 4.0) <https://www.biorxiv.org/content/10.1101/2021.01.01.425043v2>  
@PoLDresden<sup>2</sup>) [https://www.cell.com/cell-reports/fulltext/S2211-1247\(20\)30423-X](https://www.cell.com/cell-reports/fulltext/S2211-1247(20)30423-X)  
3) Solorzano et al. <https://www.nature.com/articles/s41592-021-01358-2>

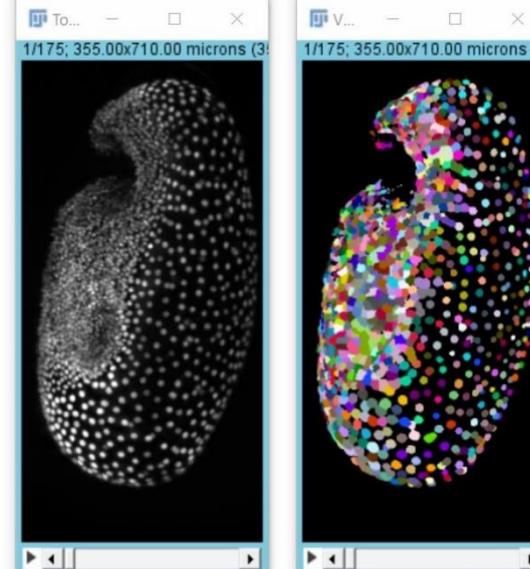
# Image analysis beyond segmentation

- Spatial pattern / image analysis for cell biology

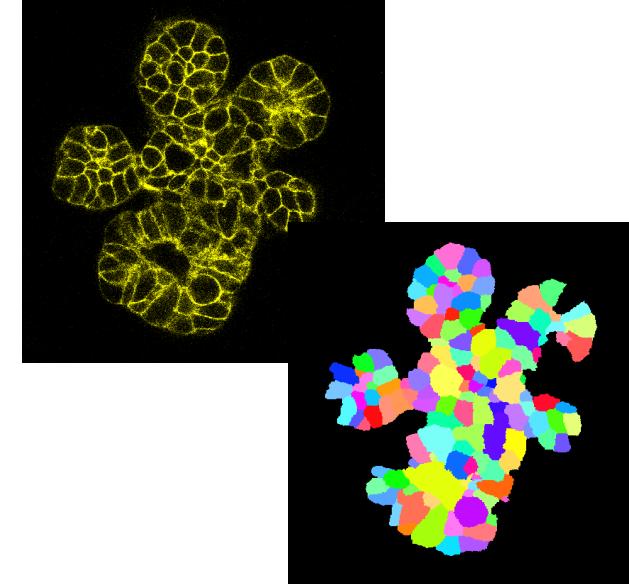
Membranes  
Nuclei



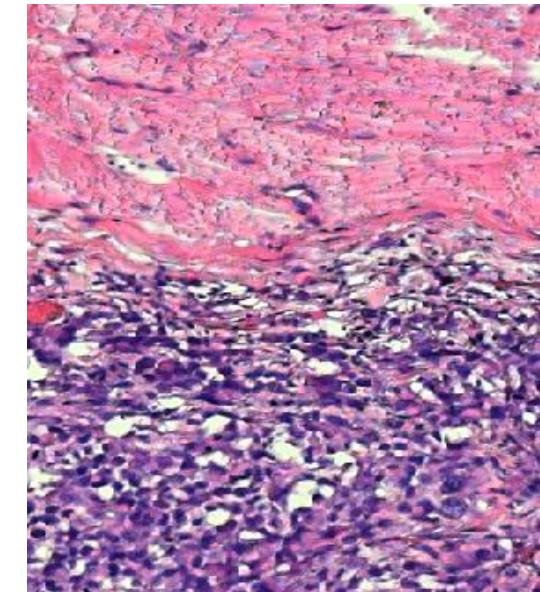
Structural biology of plants<sup>1</sup>



Early embryo development



Organoid formation<sup>3</sup>



Cancer research / histology

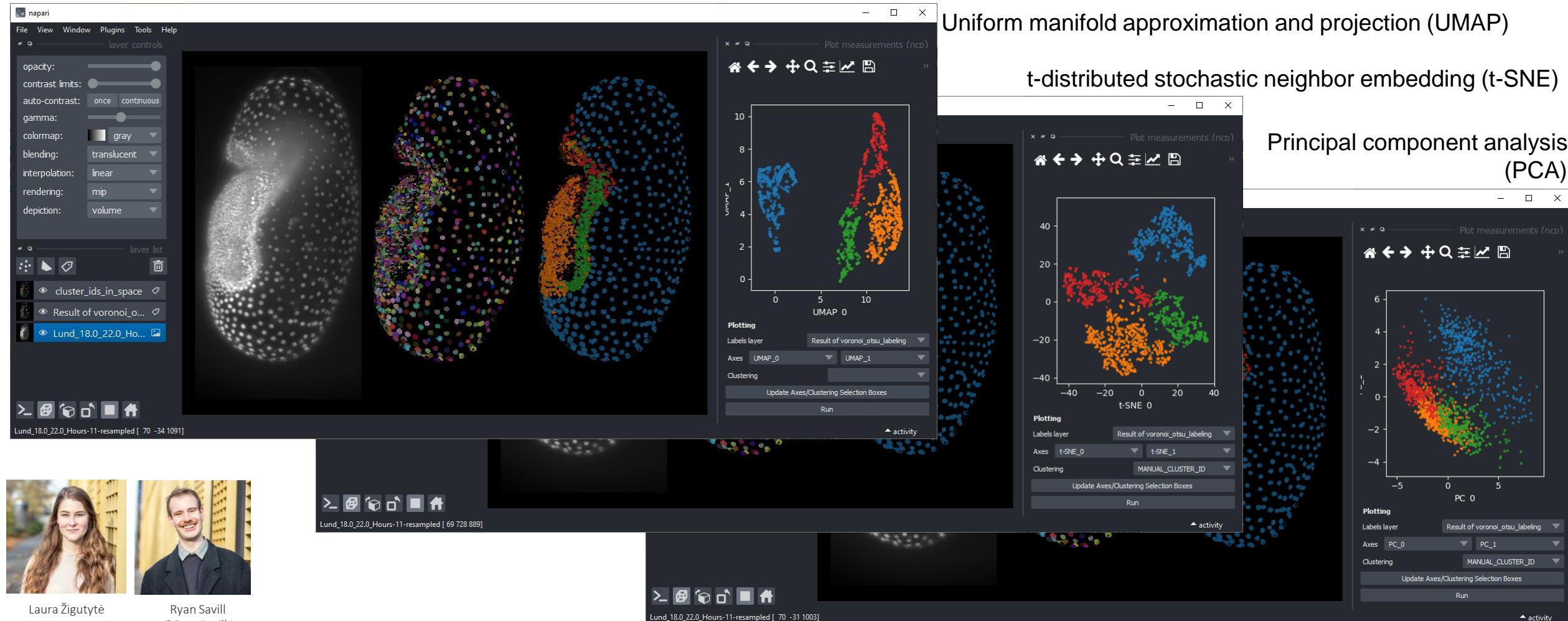


@haesleinhuepf  
@PoLDresden

1) Image data source: Pasternak et al. (CC-BY 4.0) <https://www.biorxiv.org/content/10.1101/2021.01.01.425043v2>

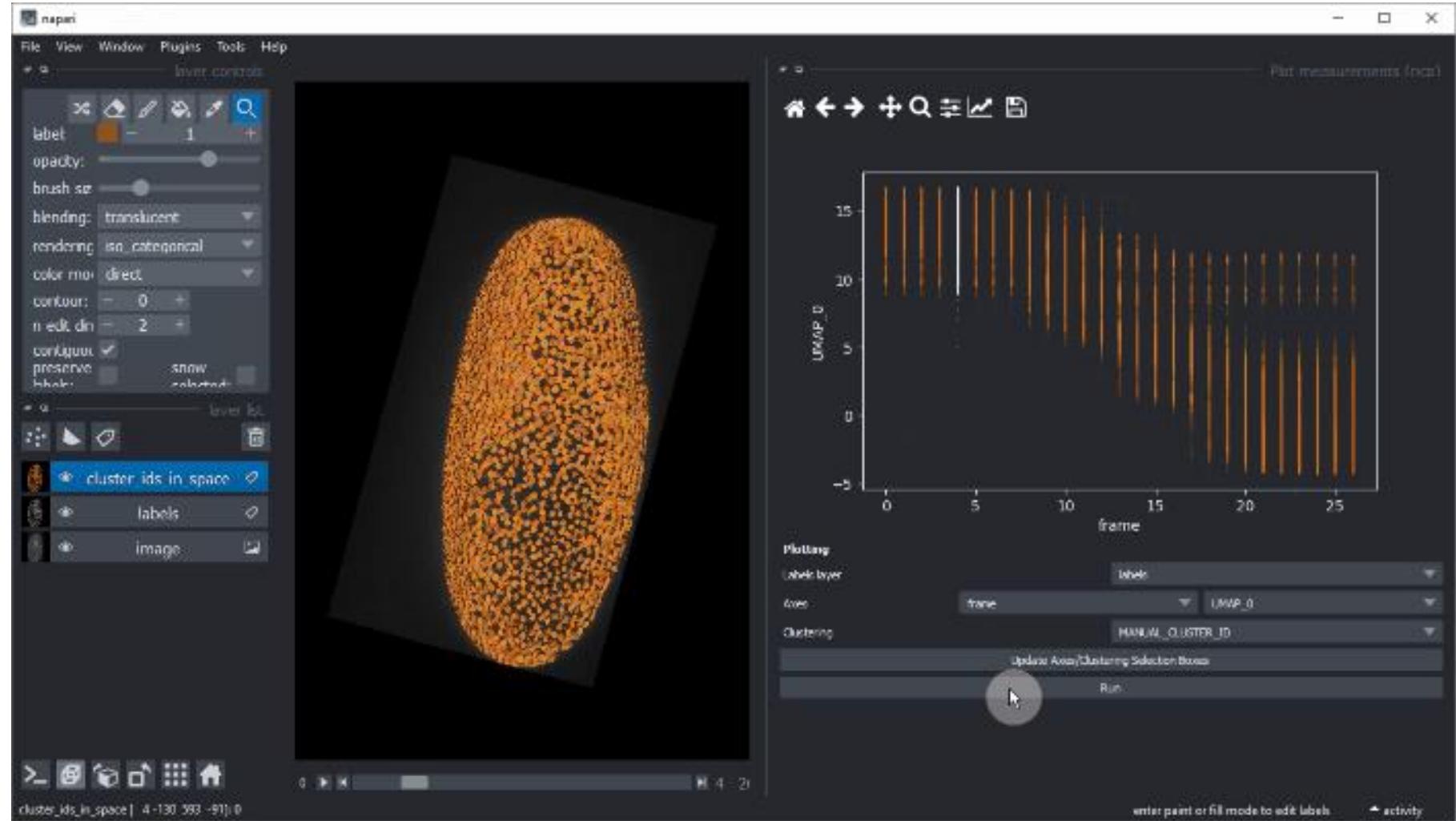
3) Image data source: ByungHo Lee, Grapin-Botton Lab

# Dimensionality reduction



# Data exploration: Hypothesis generation

- Manual clustering to gain deeper insights in relationships between measured parameters



Laura Žigutytė  
@zigutyte

Ryan Savill  
@RyanSavill4

Marcelo Zoccoler  
@zoccolermarcelo

# Image analysis is part of the experiment



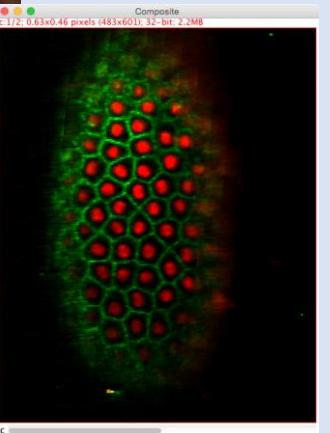
**PoL**  
Physics of Life  
TU Dresden



Observation

$$p_{ij}(t) = \frac{[\tau_{ij}(t)]^\alpha \cdot [\eta_{ij}]^\beta}{\sum_{j=1}^n [\tau_{ij}(t)]^\alpha \cdot [\eta_{ij}]^\beta}$$

Modeling



Imaging

ember 2022

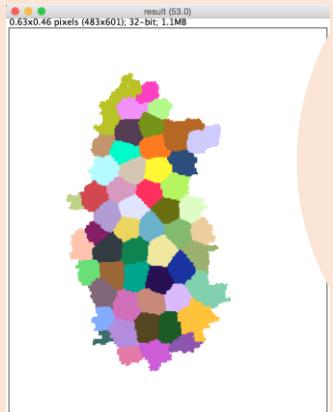


Image processing

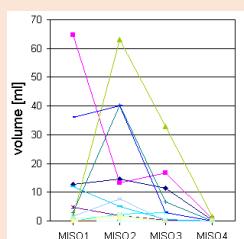
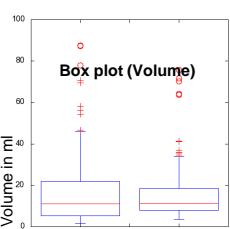


Image analysis  
Bio-statistics



Start talking to image-analysis / data-science experts  
early during your project.

So far, you learned

- *Bio-image analysis*
  - Quantitative
  - Objective
  - Reproducible
  - Repeatable
  - Reliable
- Hypothesis-driven versus hypothesis generating image data science
- When to talk to an expert

Coming up next

- Python basics