

Spatial proteomics analysis that makes you happy

Benjamin Rombaut



Flanders AI
Research Program



16th of May 2025

Dry lab computational method development:

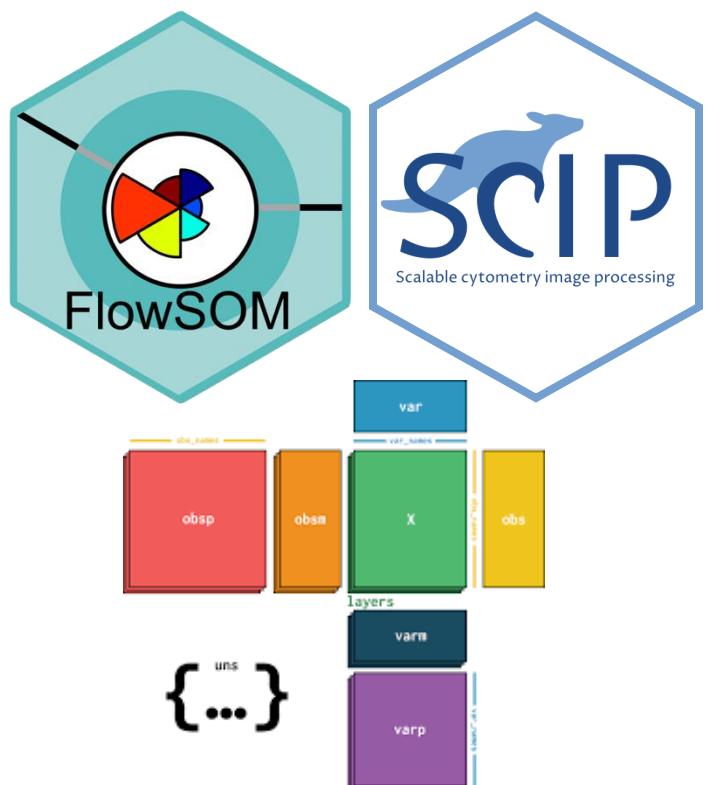
Trustworthy Models
of Cells and Tissues
4 subteams



Our tools integrate well with Python single cell tools in the scverse

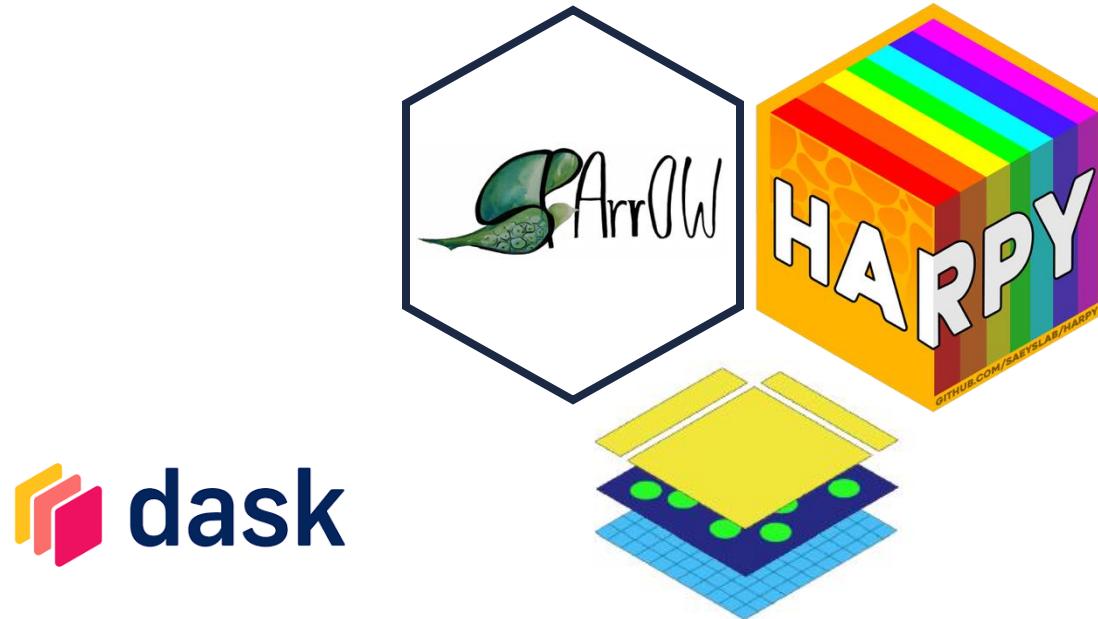


(Imaging) flow cytometry

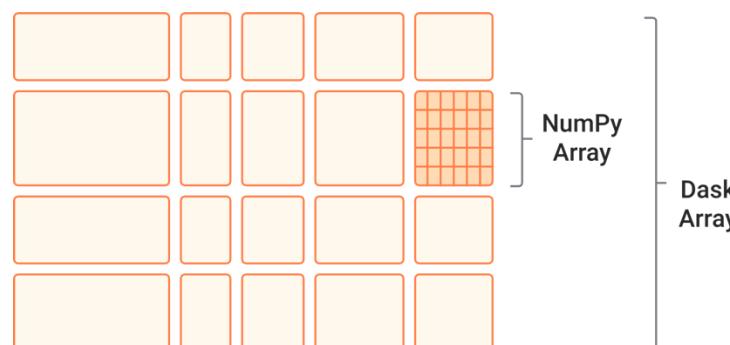


AnnData

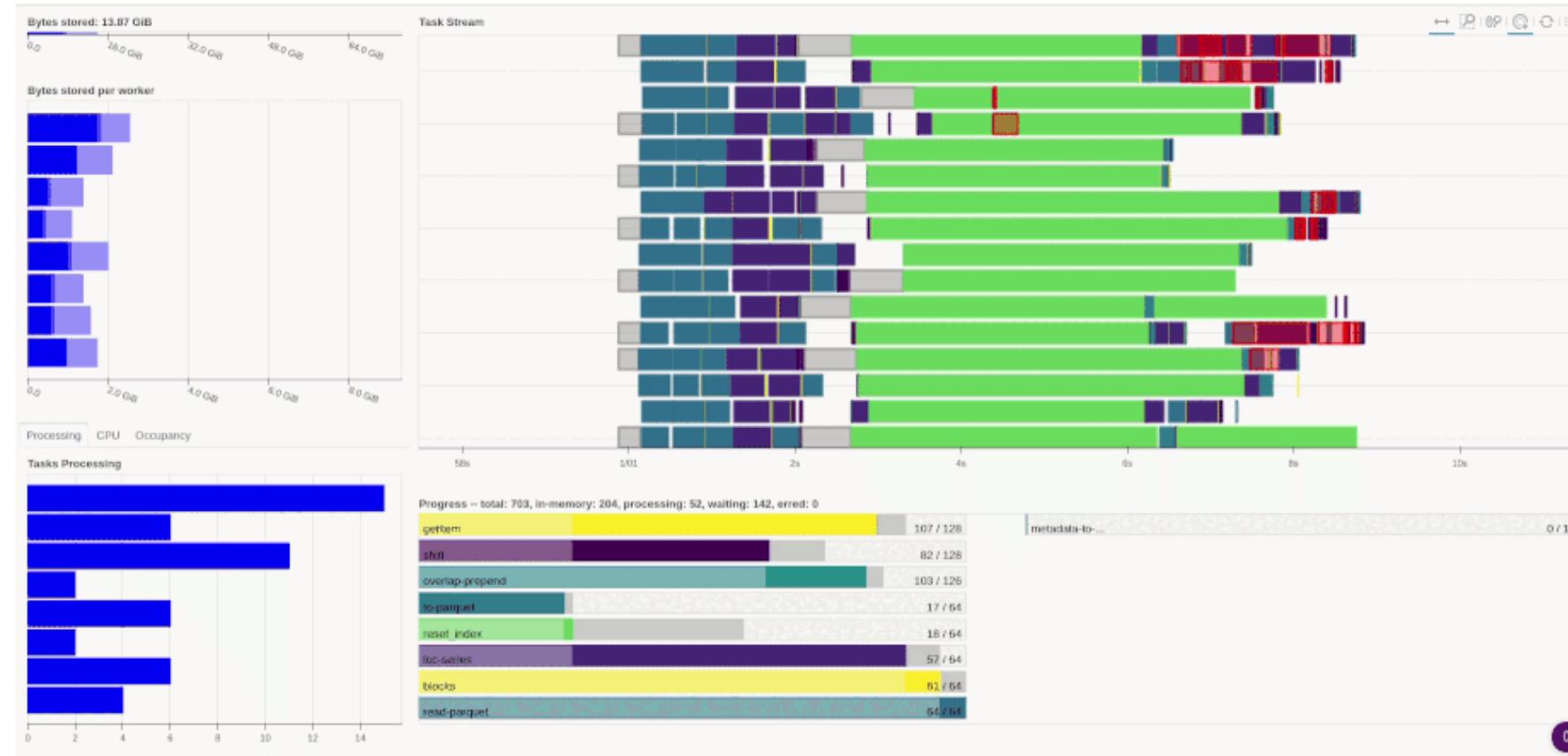
Spatial transcriptomics and proteomics



Dask is a flexible parallel computing library



*A Dask Array is
just a collection of NumPy Arrays*



*An analysis run visualized with the Dask Dashboard:
docs.dask.org/en/latest/dashboard.html*

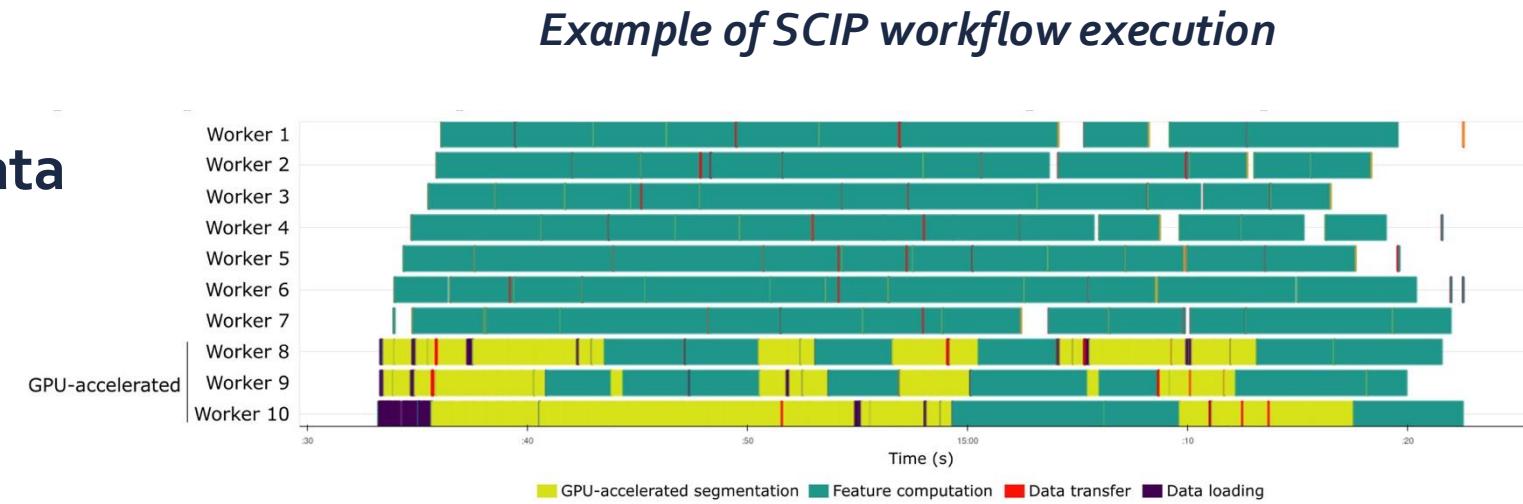
Scalable imaging flow cytometry processing



Morphological profiling of
image cytometry and
microscopy data

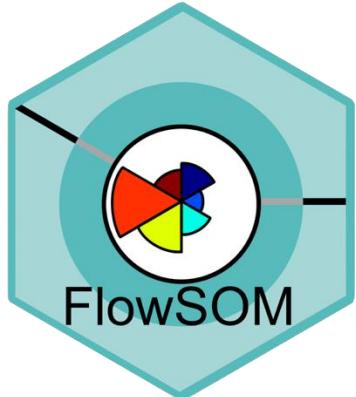
Support for storing IFC in **AnnData**

Support for parallel processing
with **Dask**



Maxim
Lippeveld

Unsupervised clustering of flow cytometry

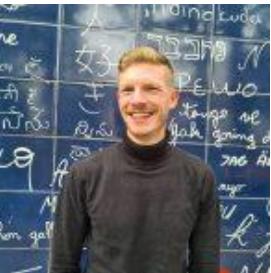
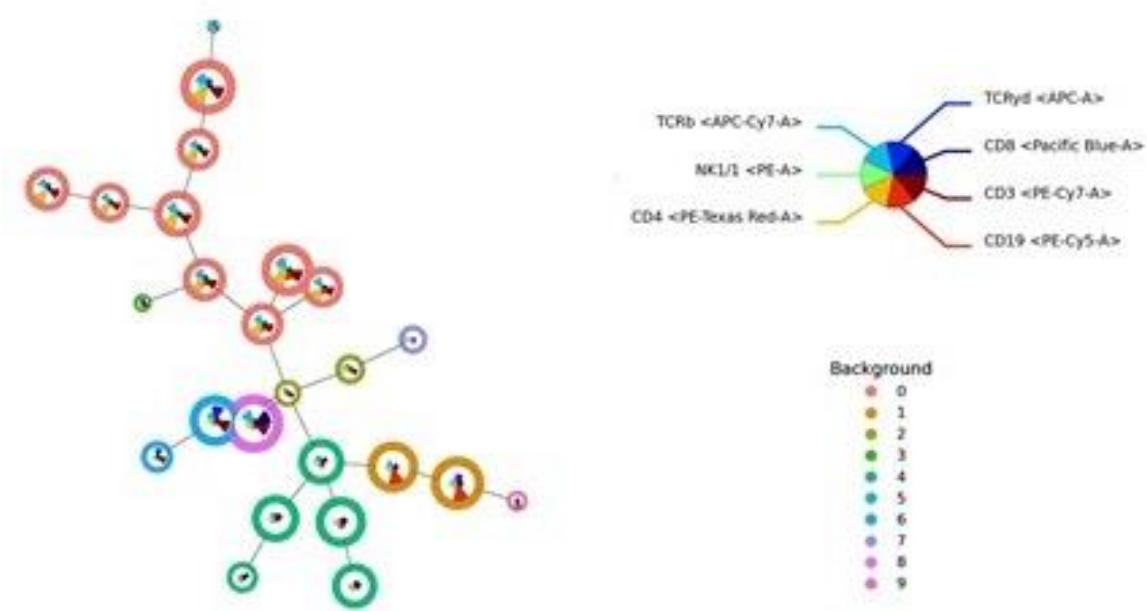


Self-organizing maps for visualization and interpretation of cytometry data

Python and R version

Support for AnnData

Support for batch parallel processing
with Dask and Numba



Artuur
Couckuyt

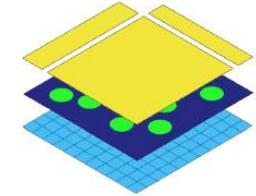
Couckuyt, A., Rombaut, B., Saeys, Y., Van Gassen, S., 2024.
Efficient cytometry analysis with FlowSOM in Python boosts interoperability with other single-cell tools.
<https://doi.org/10.1093/bioinformatics/btae179>

Scalable spatial transcriptomics processing



SPArrOW

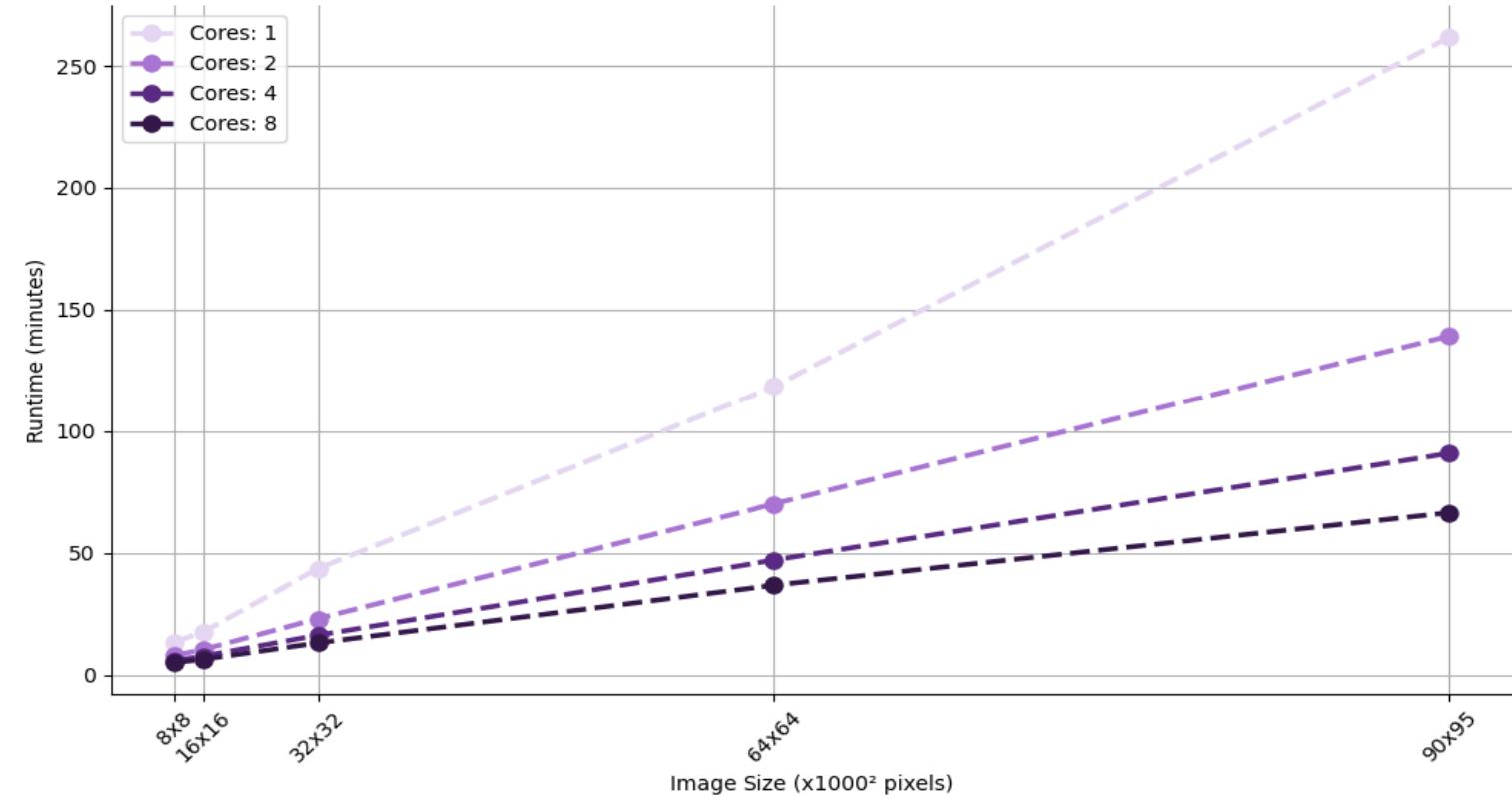
A flexible, interactive and scalable pipeline
for spatial transcriptomics analysis



SpatialData



Lotte
Polaris



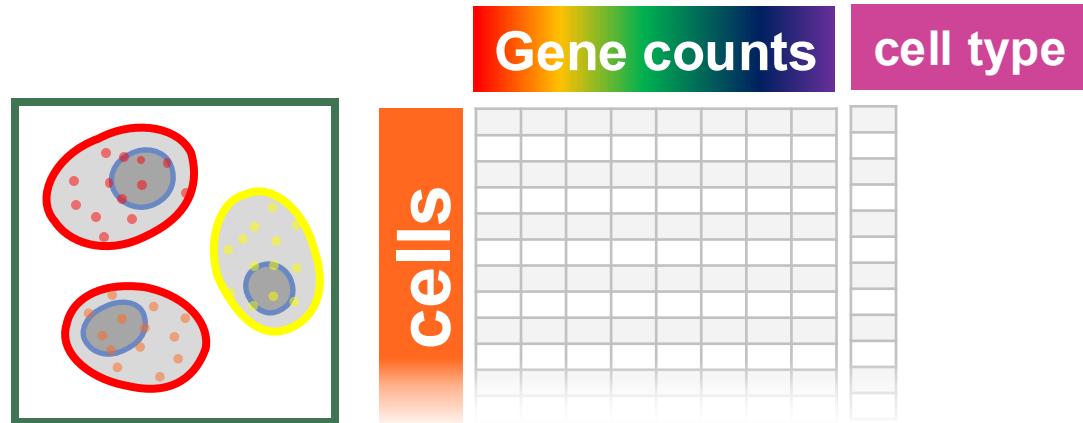
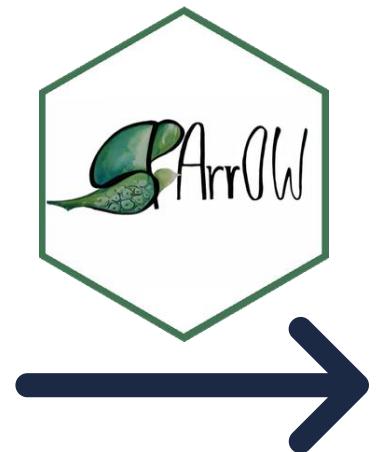
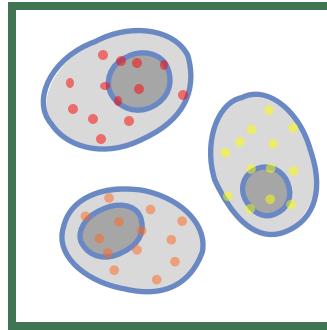
Runtime of whole SPArrOW pipeline on large MERSCOPE dataset
with Dask and GPU acceleration

Pollaris, L., Vanneste, B., Rombaut, B., Defauw, A., Vernaillen, F., Mortier, J., Vanhenden, W., Martens, L., Thoné, T., Hastir, J.-F., Bujko, A., Saelens, W., Marine, J.-C., Nelissen, H., Hamme, E.V., Seurinck, R., Scott, C.L., Guilliams, M., Saeys, Y., 2024. SPArrOW: a flexible, interactive and scalable pipeline for spatial transcriptomics analysis. <https://doi.org/10.1101/2024.07.04.601829>

Image-based spatial proteomics analysis is a lot of images

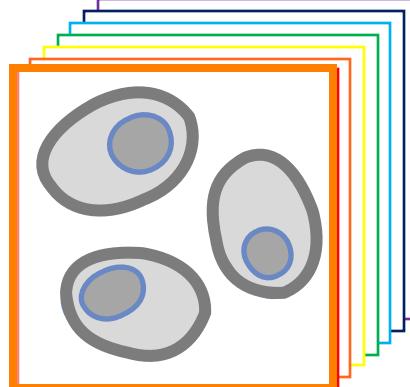
Spatial transcriptomics

Transcript locations +
whole-slide cell image



Spatial proteomics

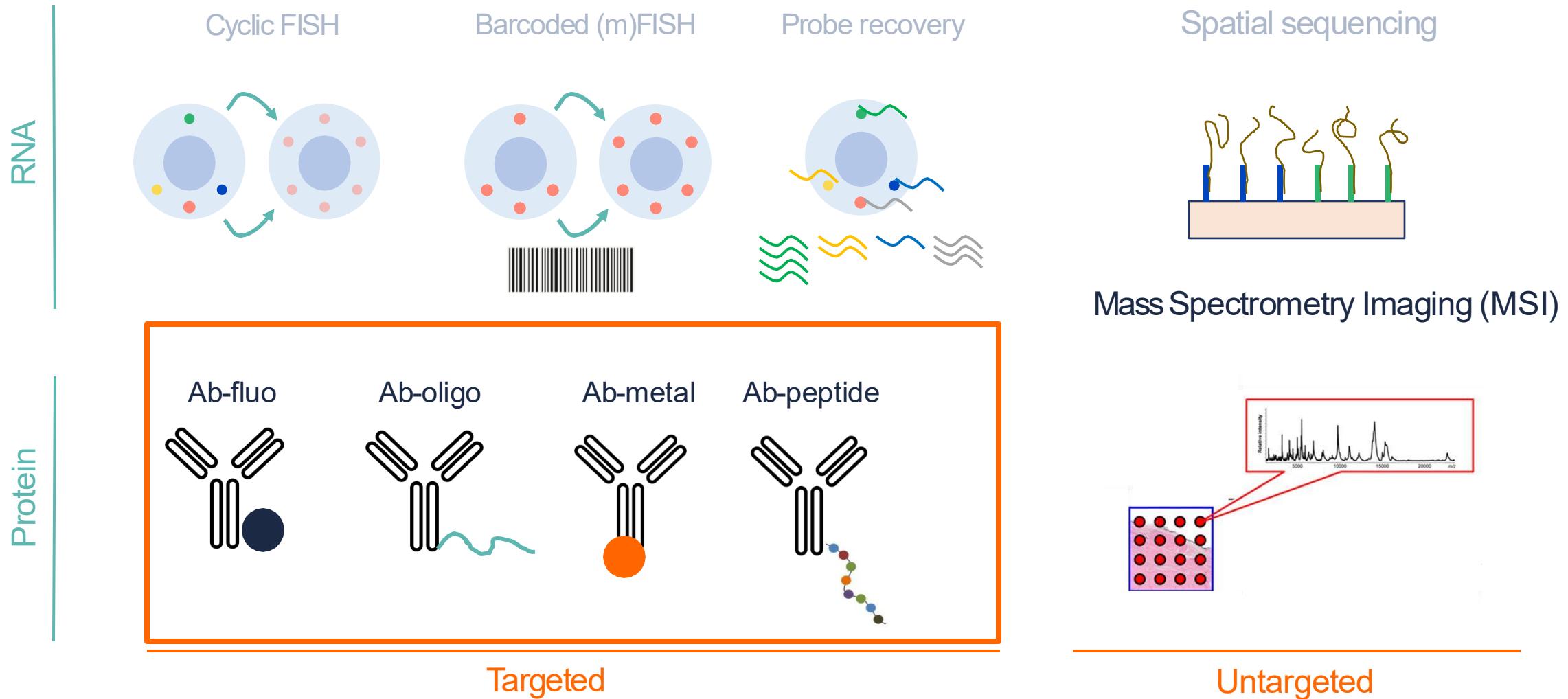
whole-slide
highly-multiplexed images



annotated single-cell data with spatial context



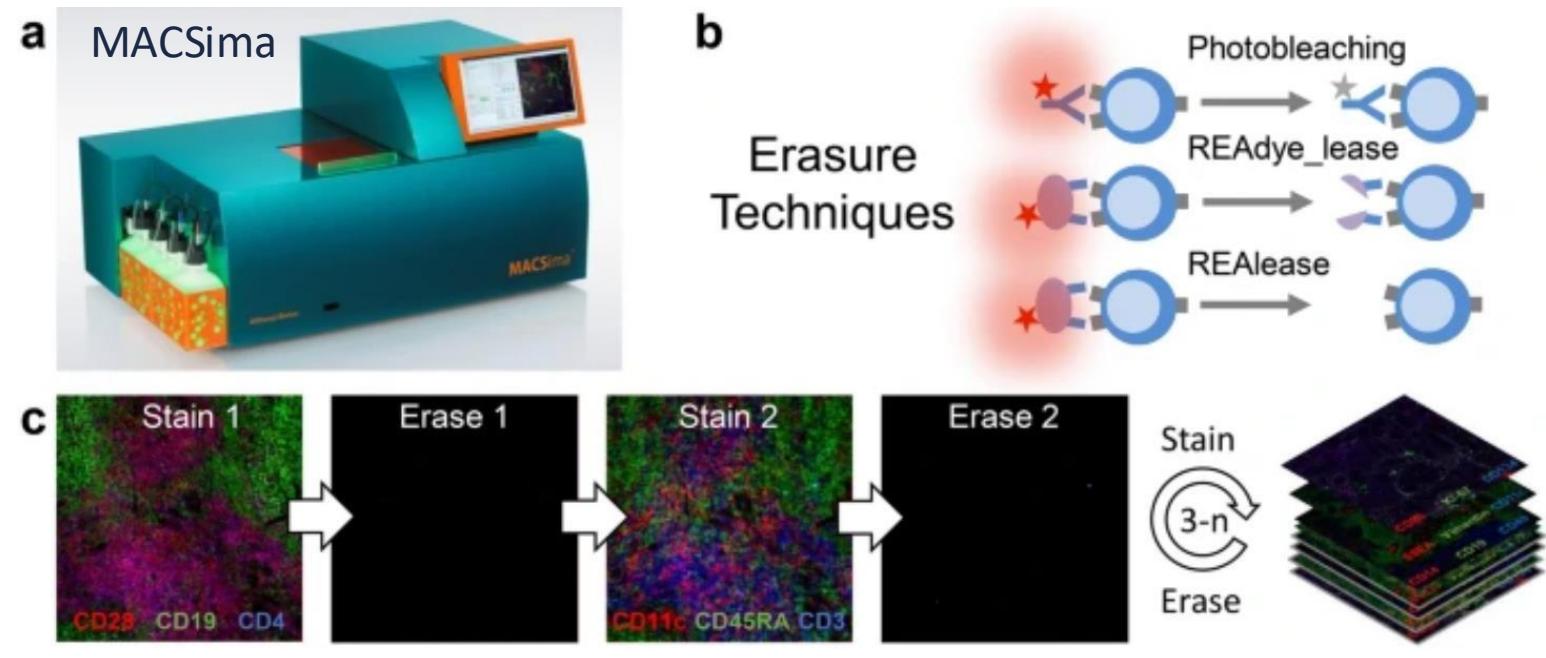
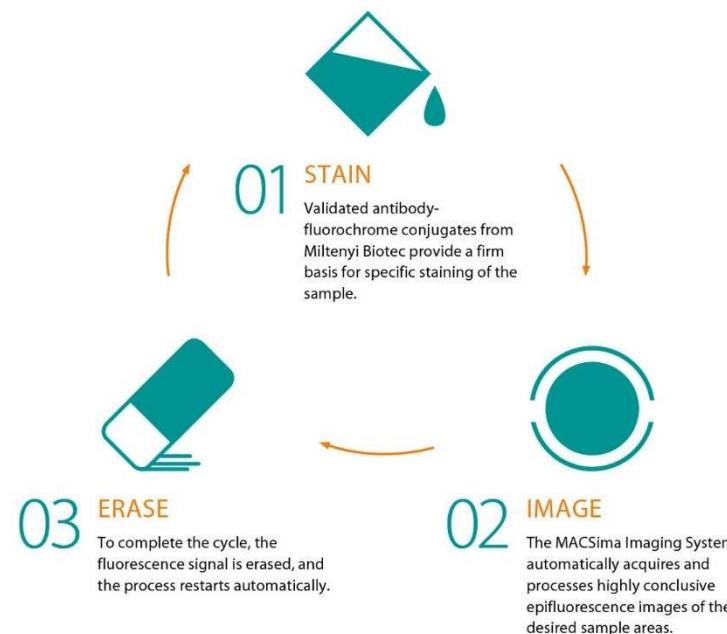
Targeted spatial proteomics



Targeted (imaging-based) spatial proteomics

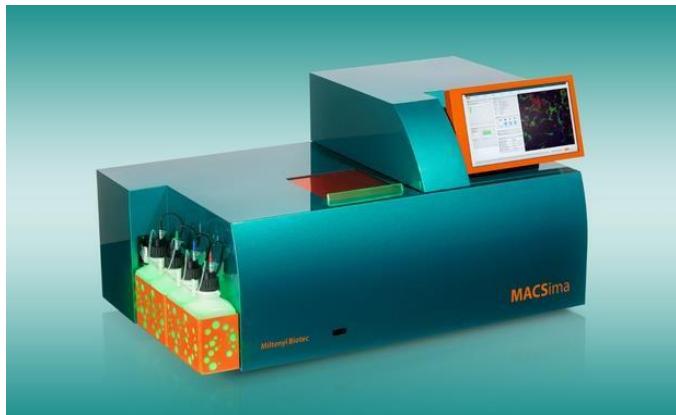
Platforms: MACSima (Miltenyi Biotec), PhenoCyder Fusion (Akoya), OPAL (Akoya), COMET (Lunaphore), CellScape...

Plexity: 6- to 200-plex

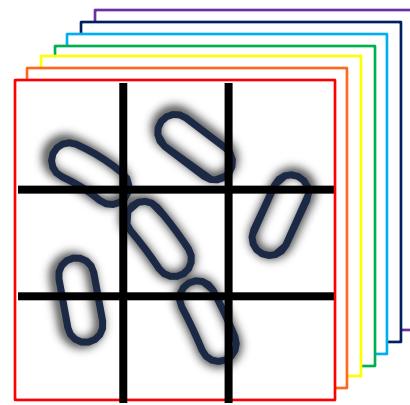


Raw image tiles need to be stitched

Multiplexed imaging



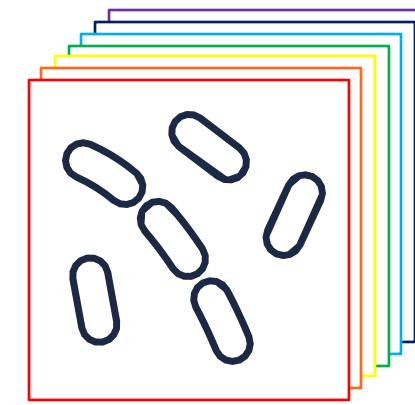
Level 1 Data
Raw image tiles



**REGISTRATION
STITCHING**
→
Commercial
preprocessing
ASHLAR

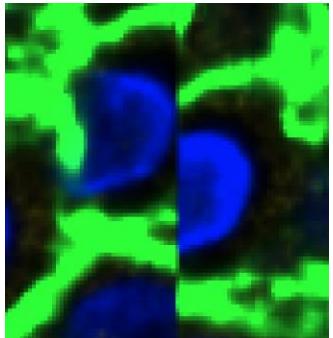
e.g. multiple multichannel
 1000×1000 tiles

Level 2 Data
Whole-slide images

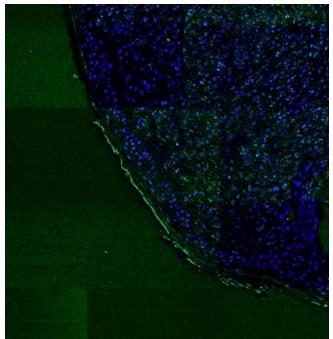


e.g. 20 samples, each:
 $20\,000 \times 20\,000$ pixels
~150 protein channels

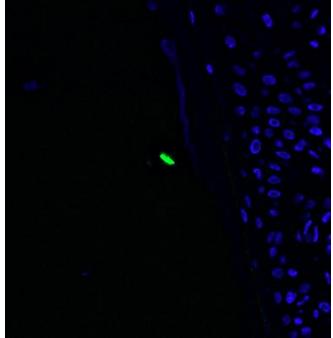
Difficulties of spatial omics



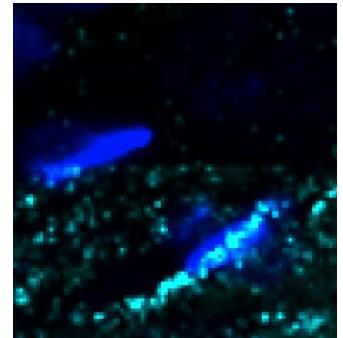
Stitching artefacts



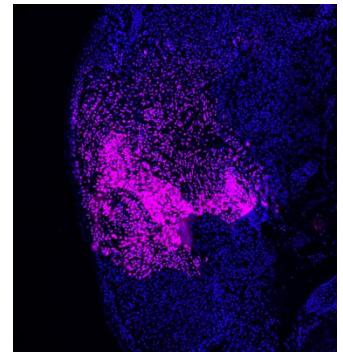
Background noise



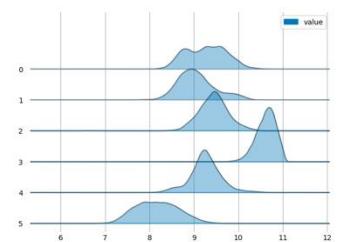
Outliers



Tiling artefacts



Staining artefacts



Batch effects

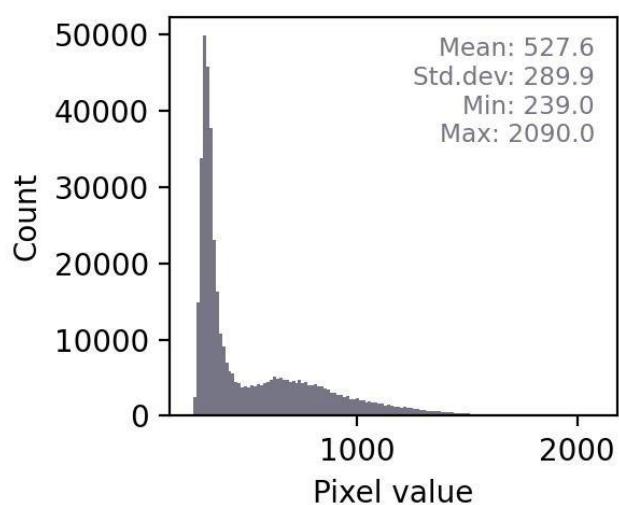
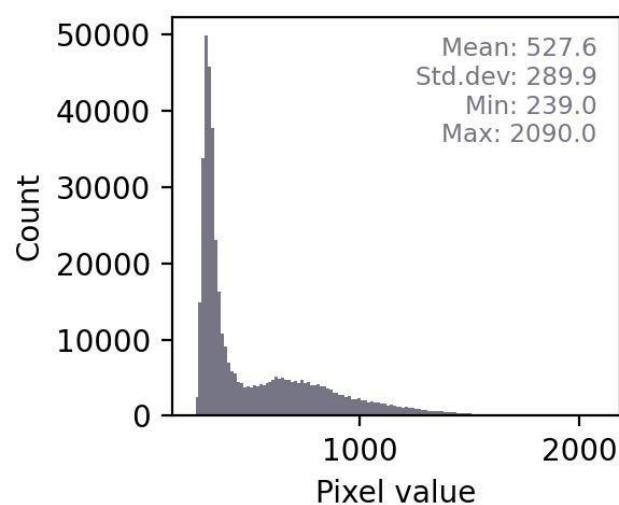
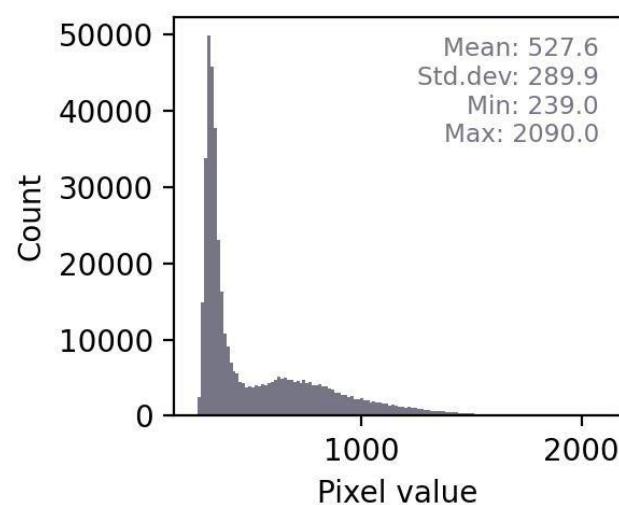
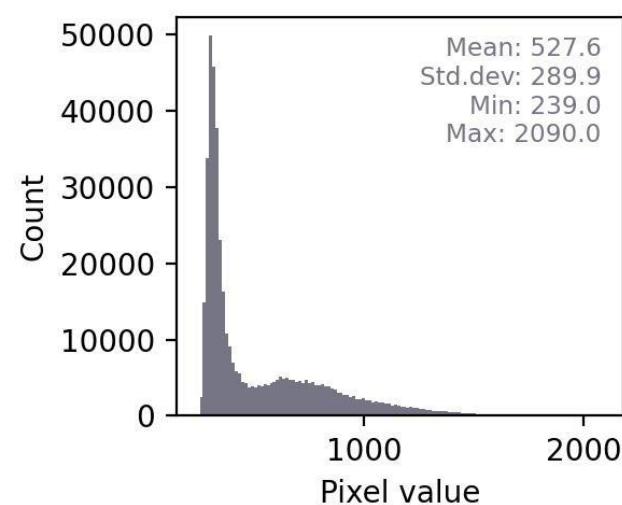
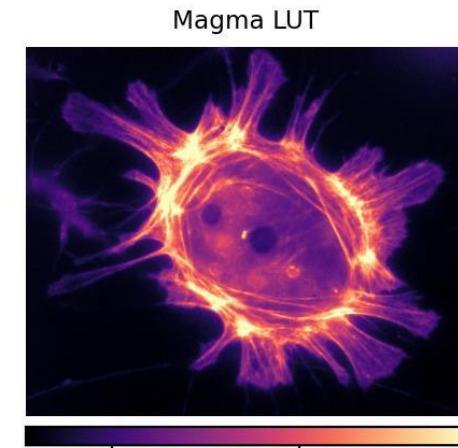
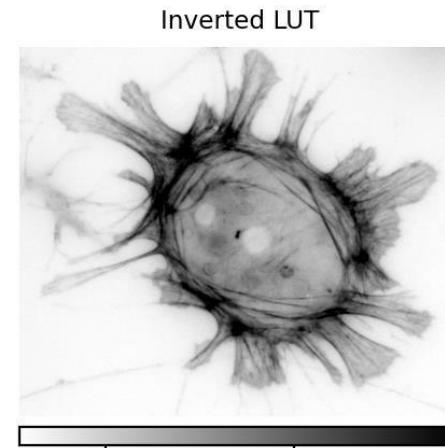
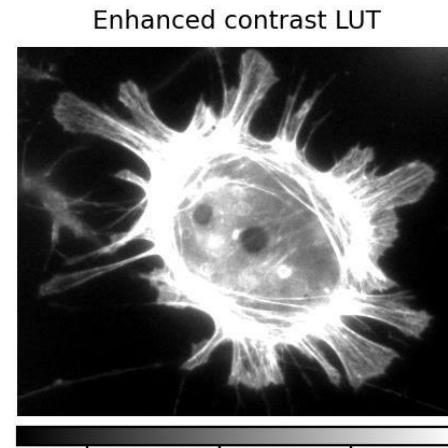
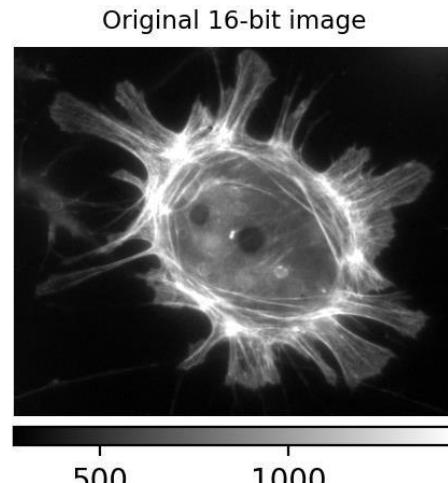
Intro into bioimage analysis

Images=pixels

30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30
30	30	30	30	30	30	30	30	30	35	40	45	50	45	40	35	30	
30	35	40	45	50	45	40	35	30	40	120	135	150	135	120	40	30	
30	40	120	135	150	135	120	40	30	45	135	230	240	230	135	45	30	
30	45	135	230	240	230	135	45	30	50	150	240	250	240	150	50	30	
30	50	150	240	250	240	150	50	30	45	135	230	240	230	135	45	30	
30	45	135	230	240	230	135	45	30	40	120	135	150	135	120	40	30	
30	40	120	135	150	135	120	40	30	35	40	45	50	45	40	35	30	
30	35	40	45	50	45	40	35	30	30	30	30	30	30	30	30	30	
30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	

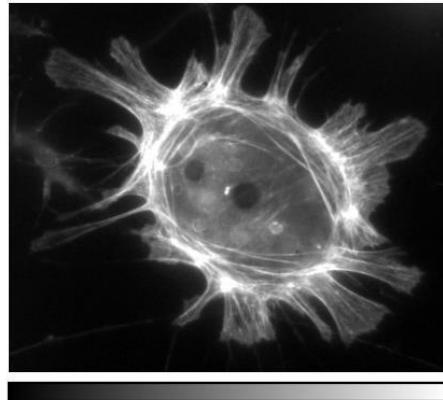
8bit images have 2^8 (256) values, ranging from 0 to 255.

Different images, same pixels



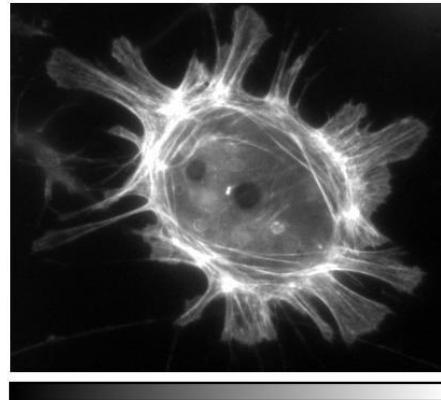
Same images, different pixels

Original 16-bit image



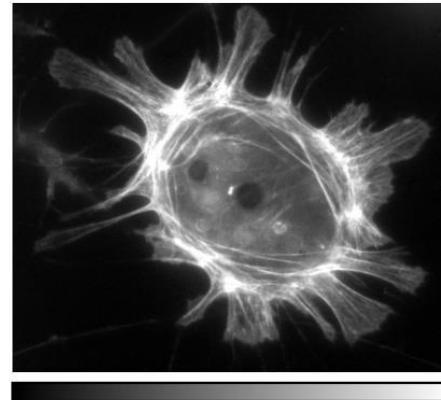
500
1000

32-bit processed



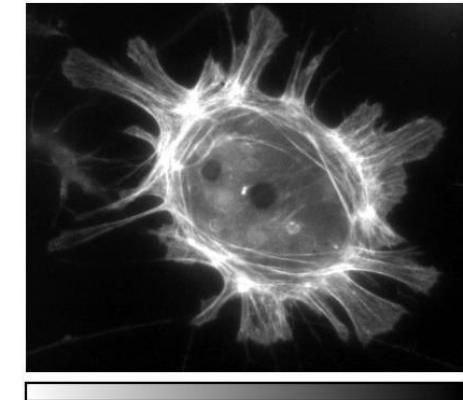
0
100

8-bit clipped

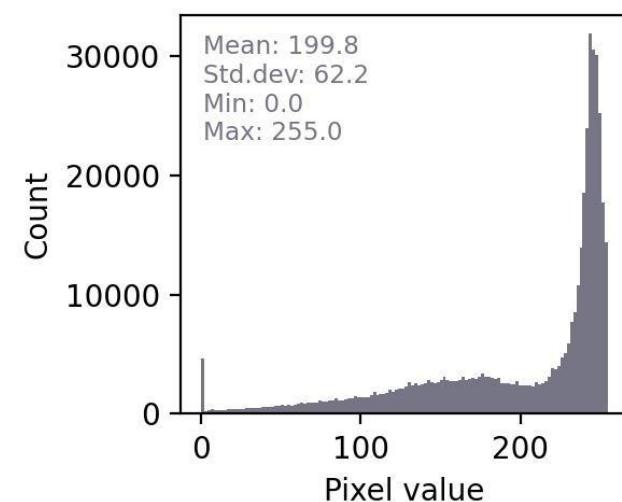
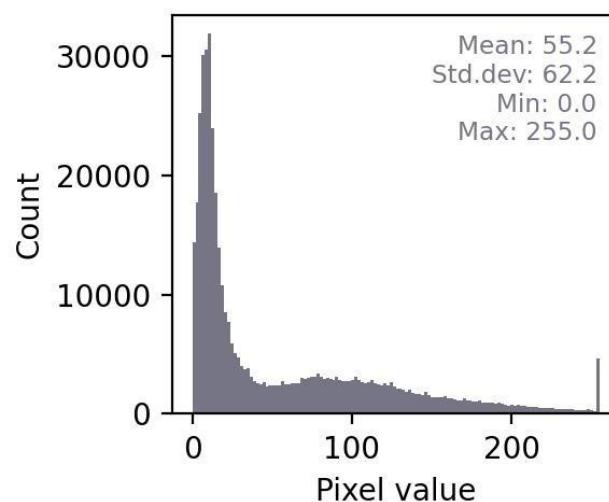
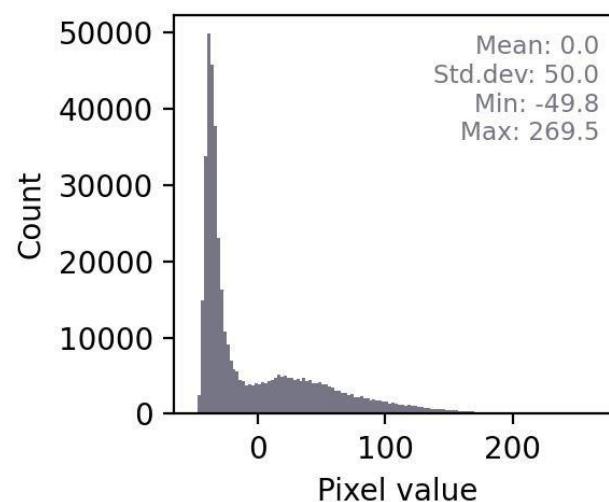
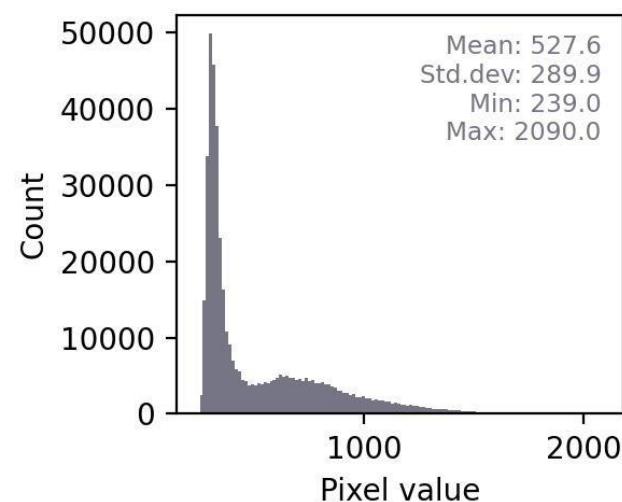


0
100
200

Inverted with inverted LUT

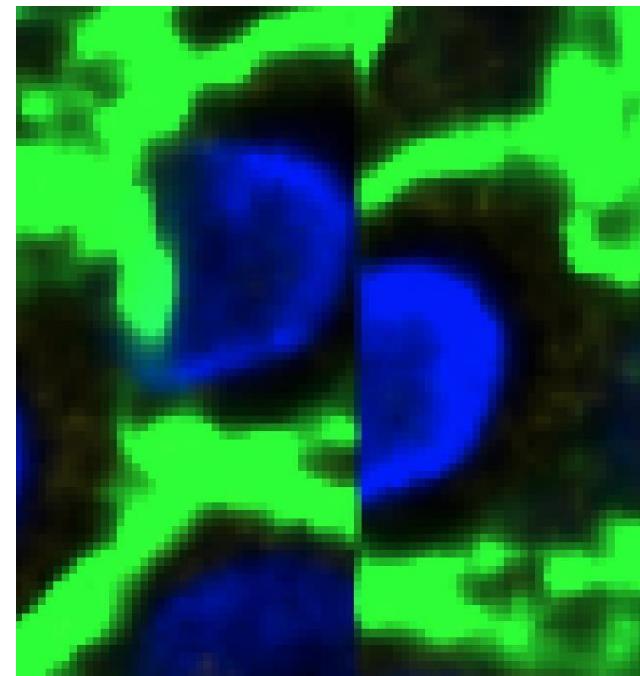


0
100
200



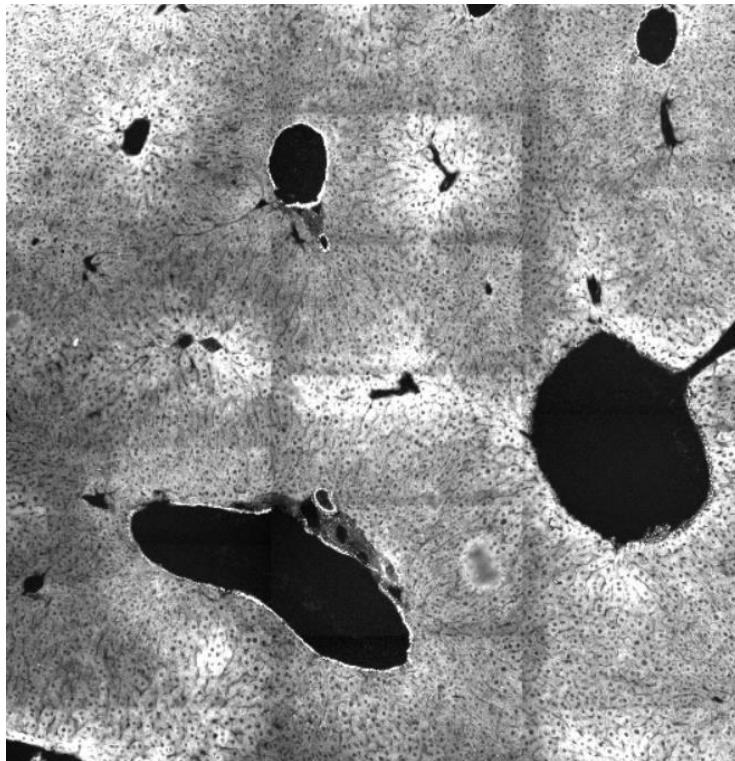
Difficulties of multiplexed bioimaging

Stitching artifacts

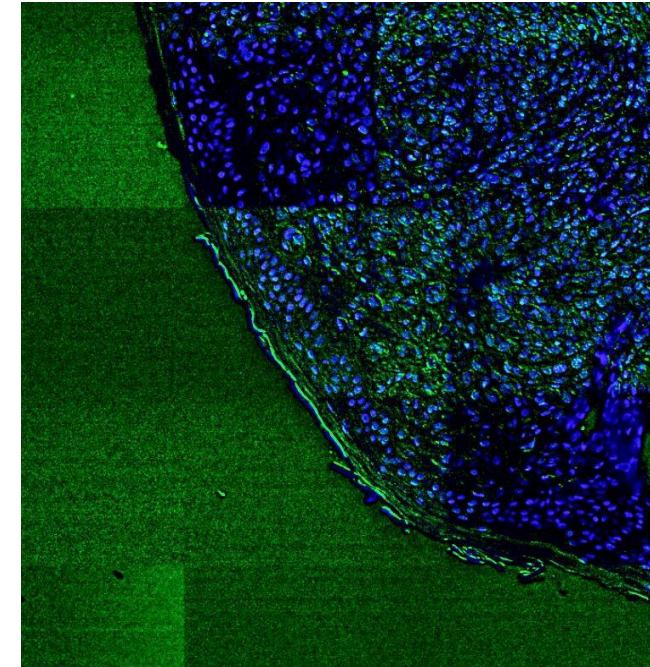


Difficulties of multiplexed bioimaging

Autofluorescence

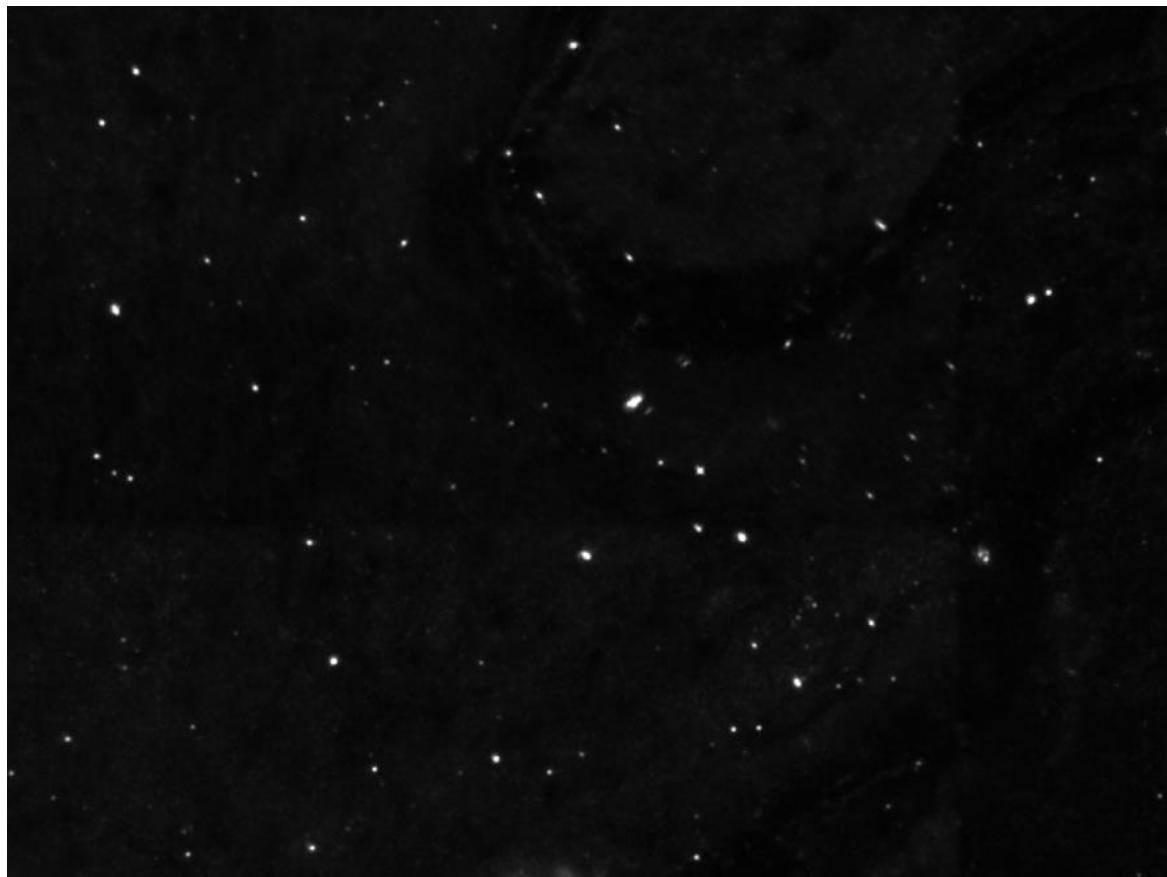


Background noise



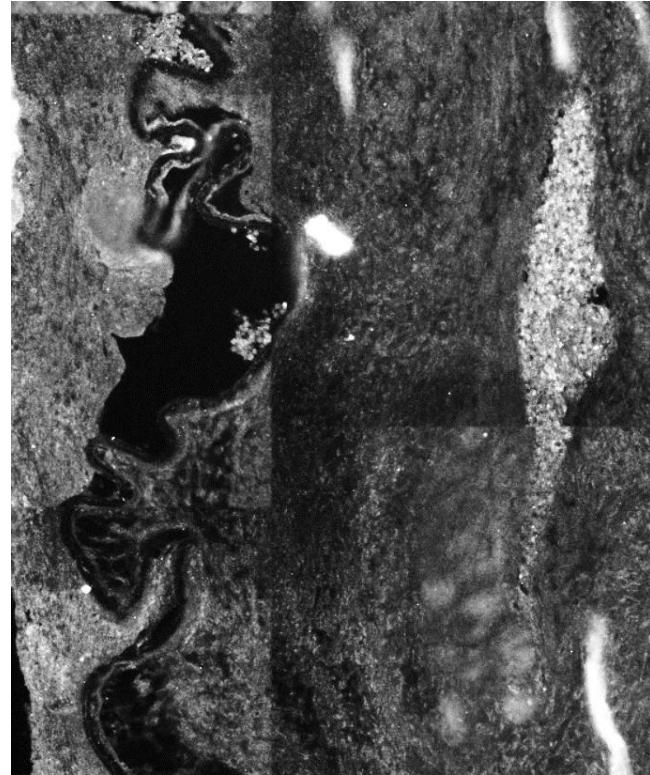
Difficulties of multiplexed bioimaging

Staining artifacts



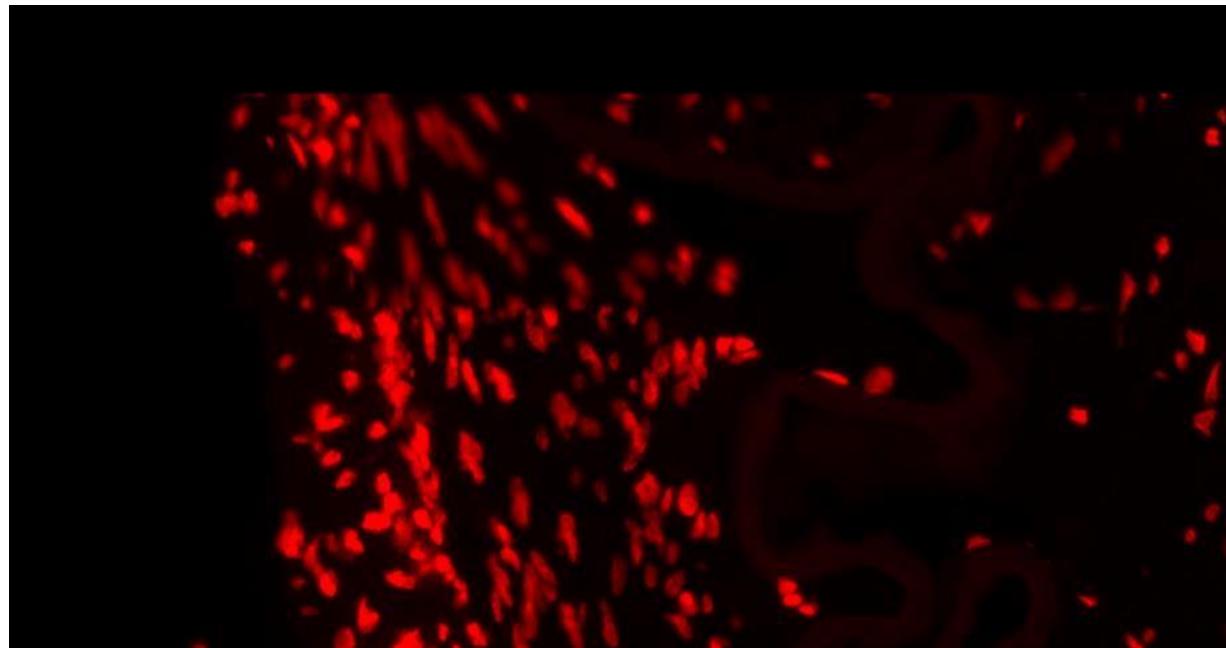
Difficulties of multiplexed bioimaging

Bleaching



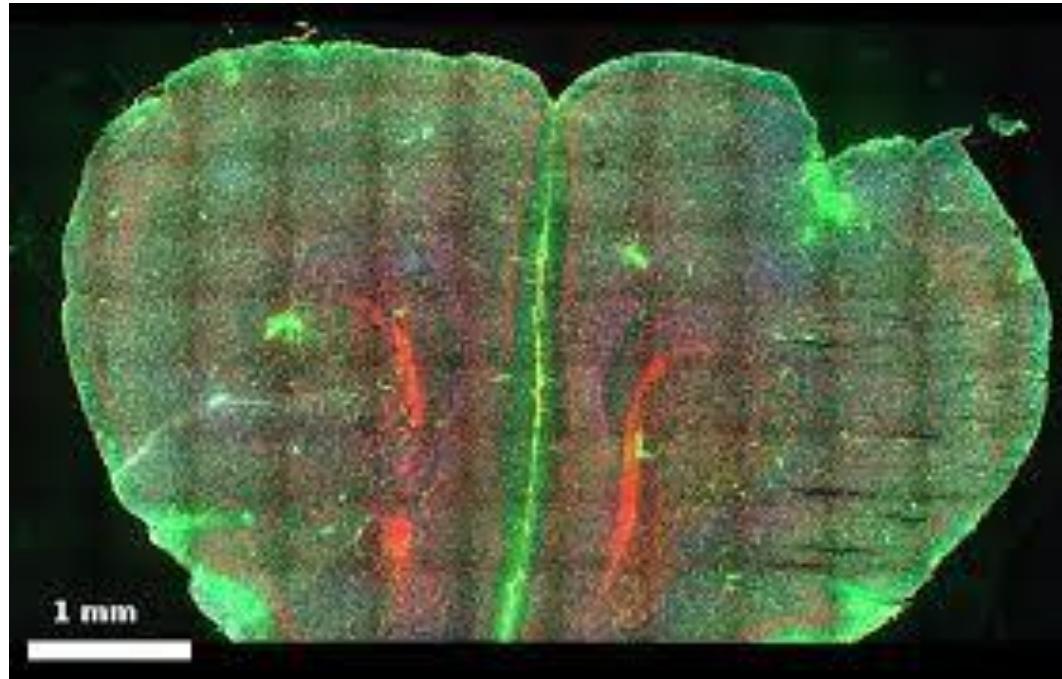
Difficulties of multiplexed bioimaging

Bad focus



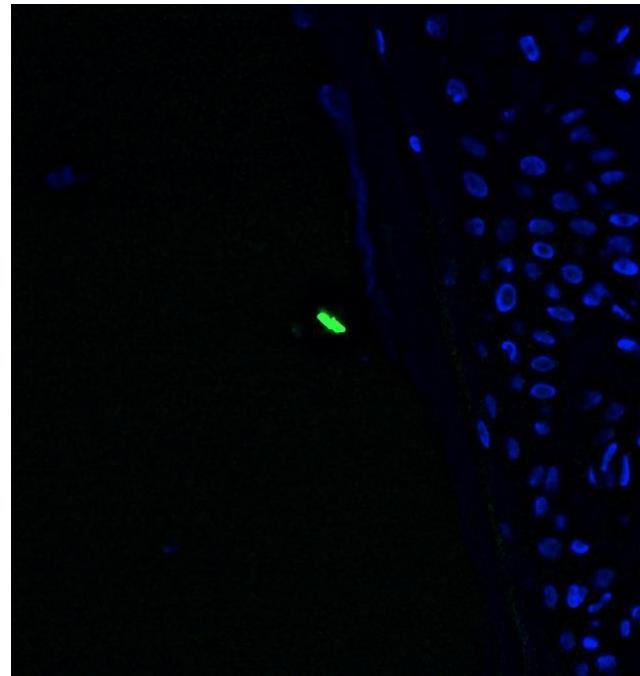
Difficulties of multiplexed bioimaging

Uneven illumination



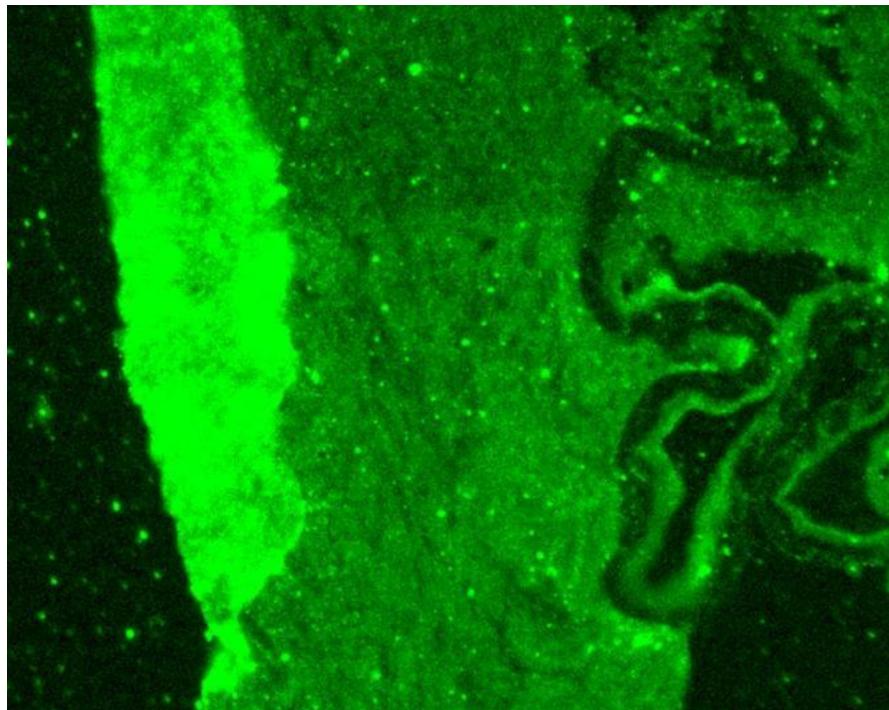
Difficulties of multiplexed bioimaging

Outliers, debris, ...



Difficulties of multiplexed bioimaging

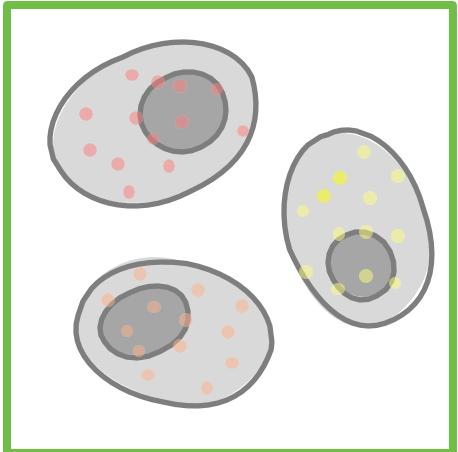
Tissue folds, detachments,



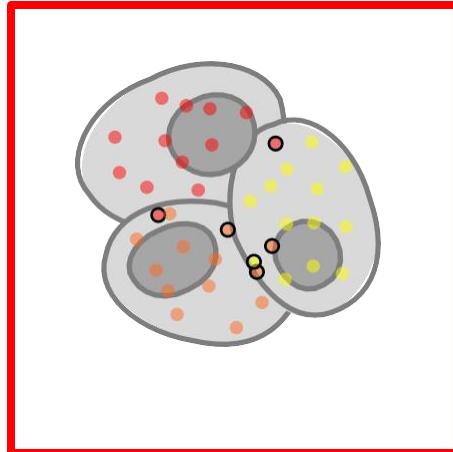
Difficulties of multiplexed bioimaging

TRANSCRIPTOMICS

No spillover

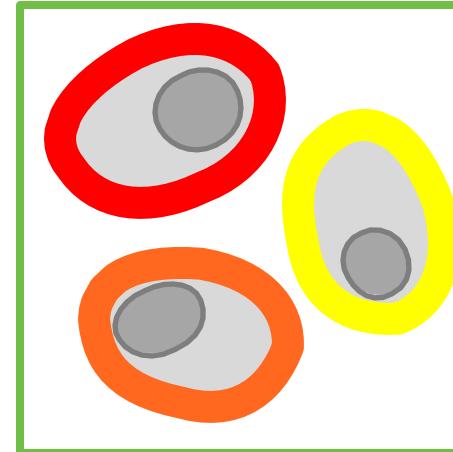


Low spillover

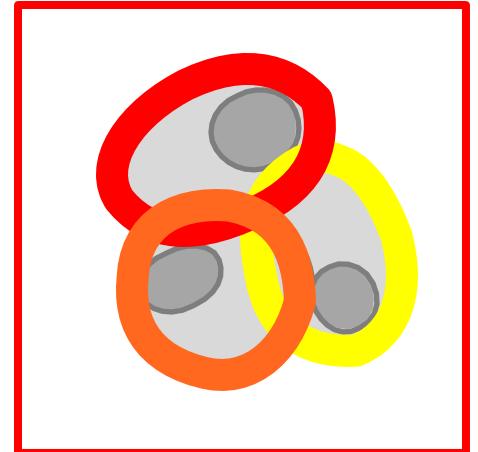


PROTEOMICS

No spillover



High spillover



Difficulties of multiplexed bioimaging

Per cell intensity feature:

Average intensity per cell (i.e. normalized for cell size)

<https://scimap.xyz/Functions/pp/rescale/>

Desired output:

bimodal distribution per channel

easy threshold for negative/positive cells

Difficulties of multiplexed bioimaging

COMMUNICATIONS BIOLOGY | <https://doi.org/10.1038/s42003-020-0828-1>

ARTICLE

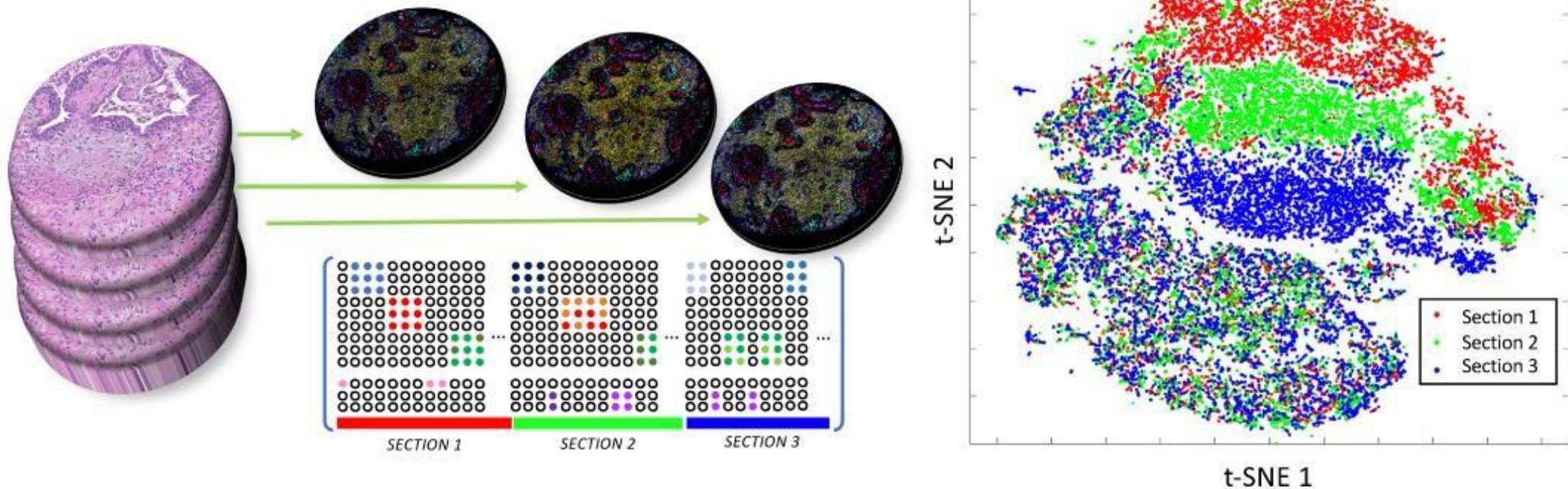
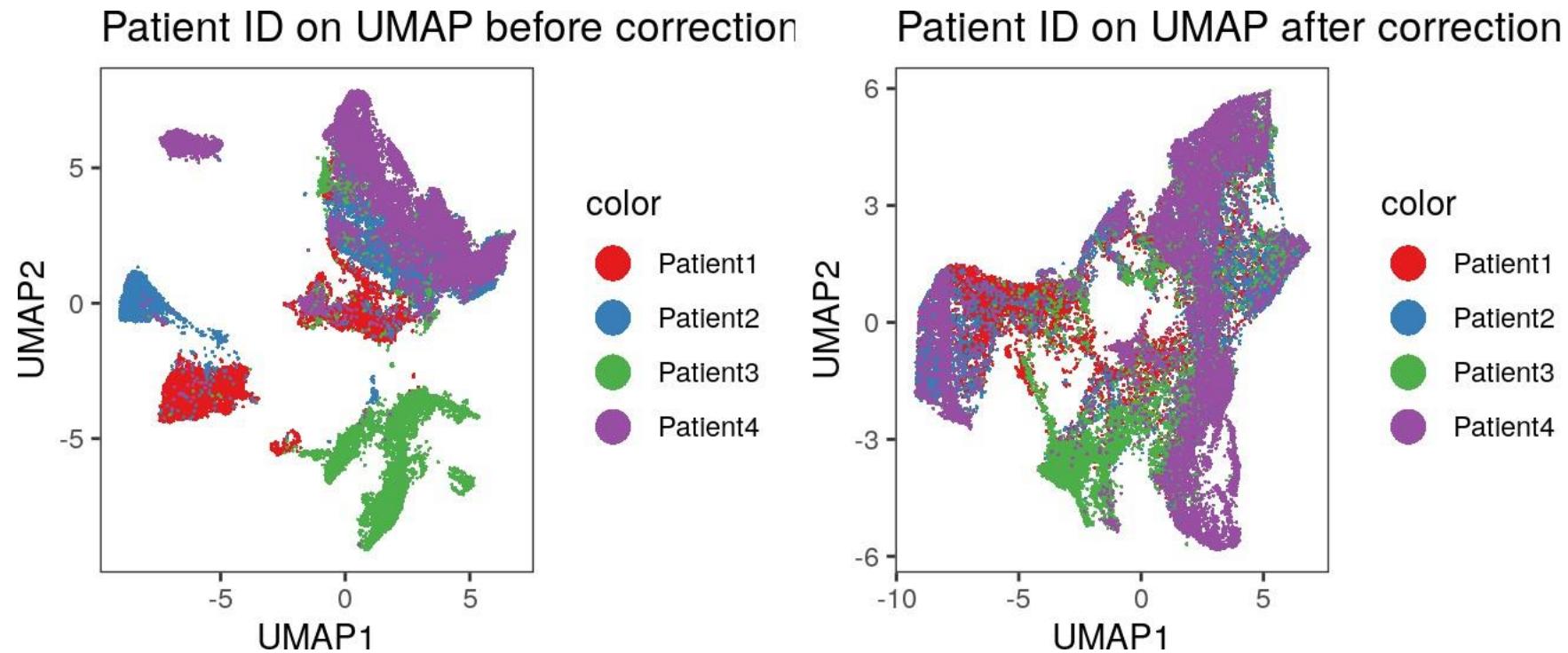


Fig. 1 An illustration of immunostaining intensity variation. Multiplexed immunostaining intensity varies across samples; intensity profiles from three adjacent sections ($5\text{ }\mu\text{m}$) show intensity variations. t-SNE embedding of individual single-cell intensity features from three adjacent sections show uniformly distributed in immune cell types (bottom half) but clustered group in cancer cell type (top half) due to the intensity variation in cytokeratin (CK) markers.

Batch effect correction methods are available, but this is still an open problem

<https://bodenmillergroup.github.io/IMCDataAnalysis/batch-effects.html>

<https://scimap.xyz/Functions/pp/combat/>



See also ark-analysis: Robust phenotyping of highly multiplexed tissue imaging data using pixel-level clustering

<https://doi.org/10.1038/s41467-023-40068-5>

Practical resources for bioimaging

- A biologist's guide to planning and performing quantitative bioimaging experiments
- <https://www.bioimagingguide.org>
- Introduction to Bioimage Analysis: <https://bioimagebook.github.io>
- Scientific Community Image Forum: <https://forum.image.sc/>
- Training modules: <https://neubias.github.io/training-resources/index.html>
- <https://imagej.net/software/fiji/>
- <https://qupath.github.io/>



Overview of different spatial proteomics workflows

Commercial data workflows

MACS iQ View

<https://www.miltenyibiotec.com/DE-en/products/mac-s-iq-view-analysis-software.html>

Phenocycler Software Suite (CODEX processor) <https://help.codex.bio/codex/processor/overview>

Indica labs – HALO Hiplex FL

<https://indicalab.com/halo/halo-modules/highplex-fl/>

Visiopharm – Phenoplex muliplex software

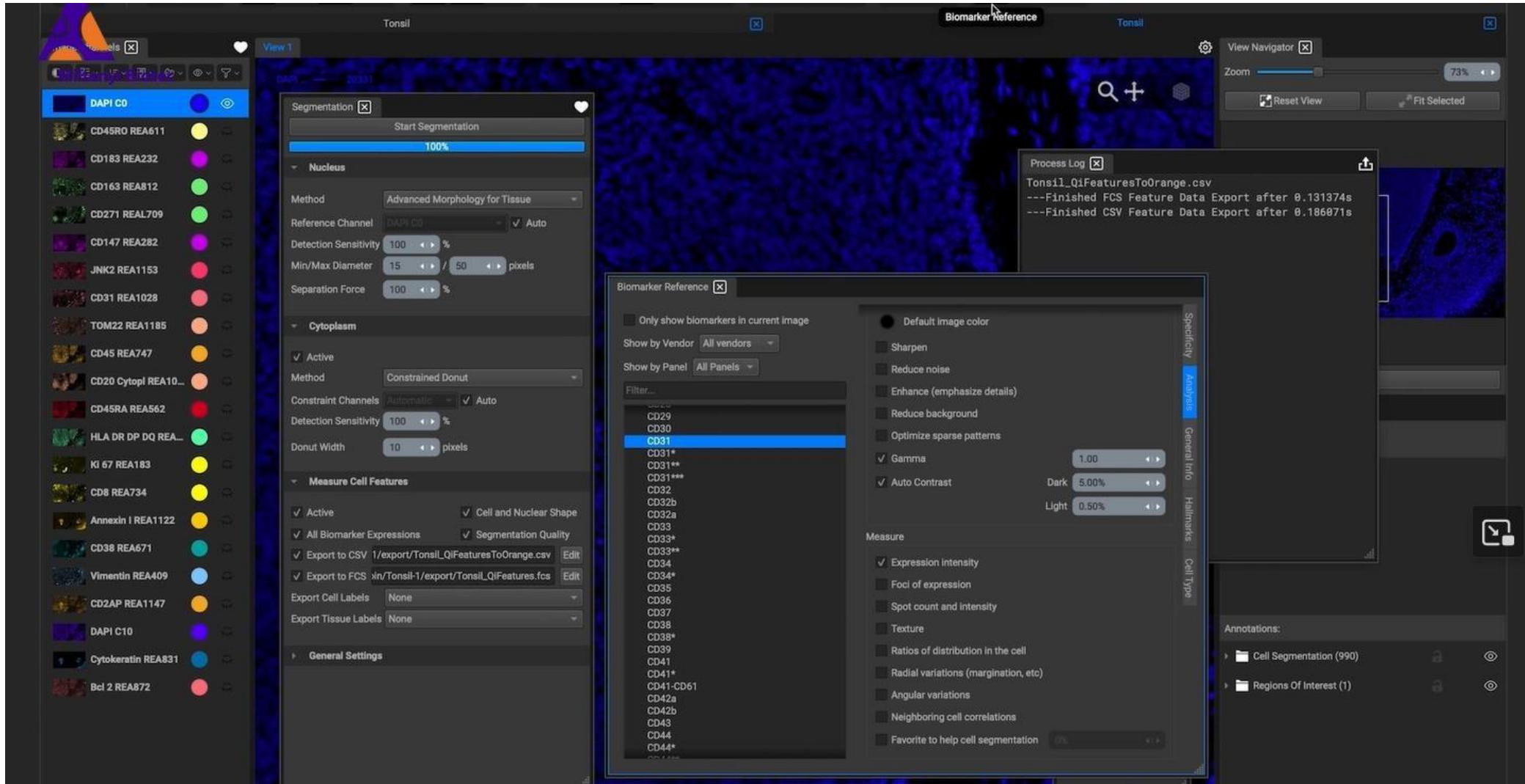
<https://visiopharm.com/phenoplex-workflow/>

AspectAnalytics – Nexus multimodal analysis

<https://www.aspect-analytics.com/platform/multimodal>

OracleBio, PathAI, Enable Medicine...

MACS iQView



MACS iQView

Strengths

End-to-end graphical and user-friendly workflow

Advanced cell type annotation with visual feedback

Python API for scripts and batch sample processing

Good for quality assessment and ground truth annotations

Weaknesses

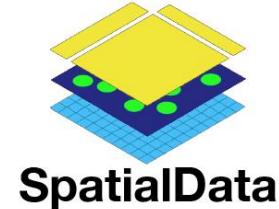
Expensive license needed

Not tested on non-MACSim data

Limited analysis options (lacks state-of the art tools)

Open-source spatial proteomics workflows

Harpy: focus on **MACSima** data and uses **SpatialData**
Python <https://github.com/saeyslab/harpy>



MCMICRO (+ CyLinter): broad focus and uses **Nextflow**

Nextflow/Python <https://nf-co.re/mcmicro/dev> <https://labsyspharm.github.io/cylinter/>

IMCDataAnalysis: focus on **IMC** data

Python/R <https://bodenmillergroup.github.io/IMCDataAnalysis/>

ark-analysis: focus on **MIBI** data

Python <https://github.com/angelolab/ark-analysis>

Other **scverse** workflows:

squidpy, sopa, spatialproteomics...

Other **R** workflows:

Giotto, VoltRon...

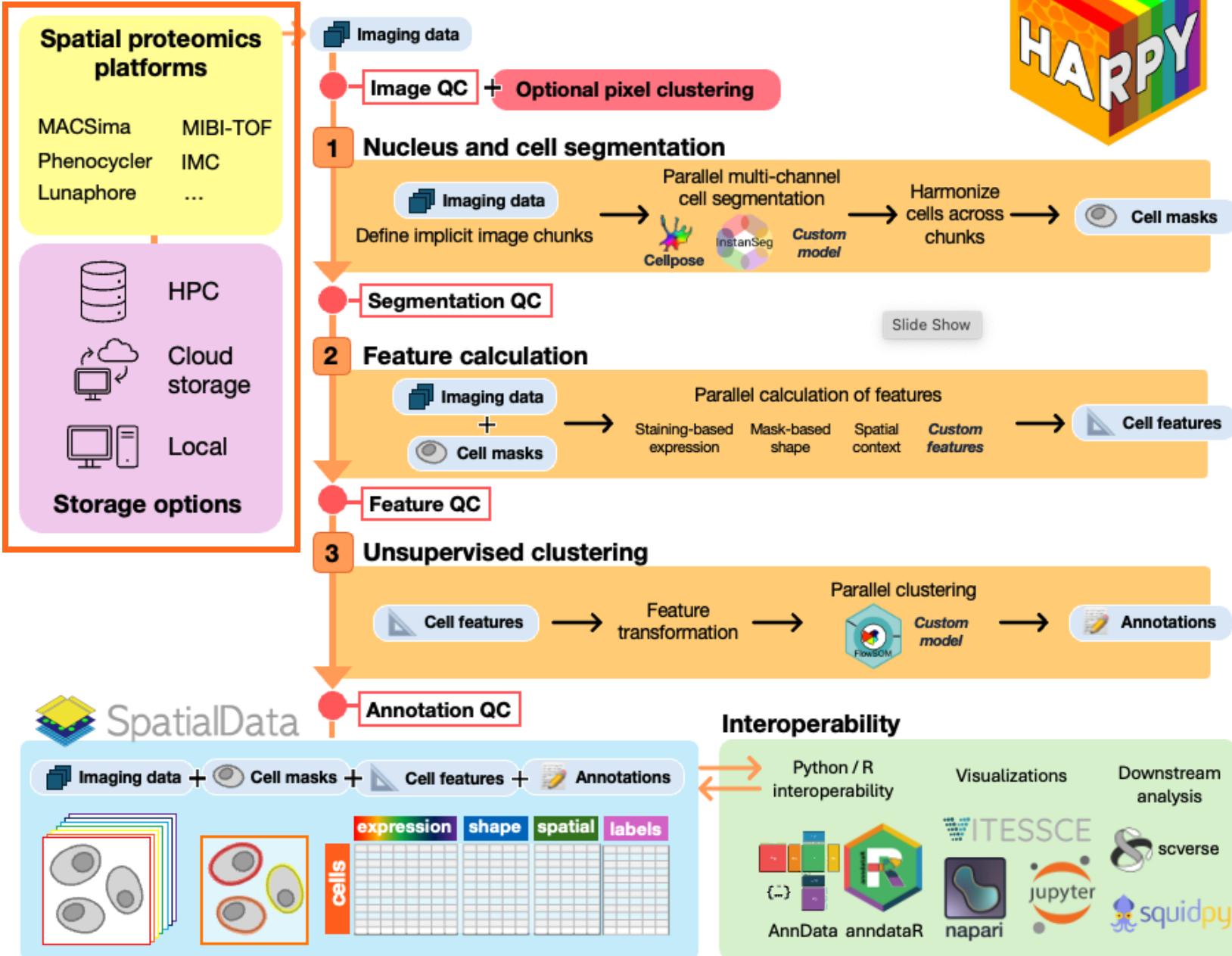
4 Harpy highlights

1. Dataset-wide quality control
2. Scalable whole-slide image processing
3. Interactive unsupervised clustering
4. Interoperability with R

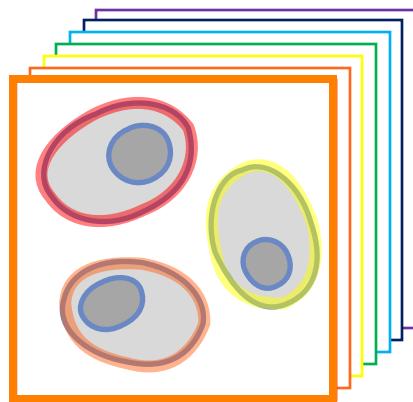


<https://github.com/saeyslab/harpy>

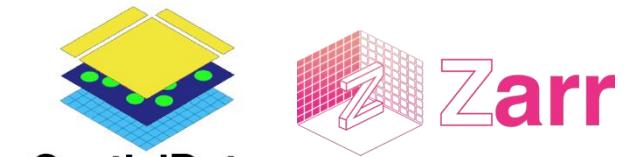
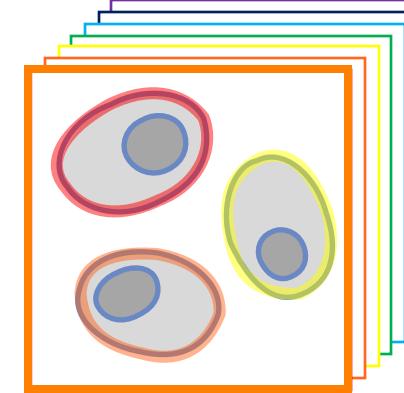
Harpy analysis workflow



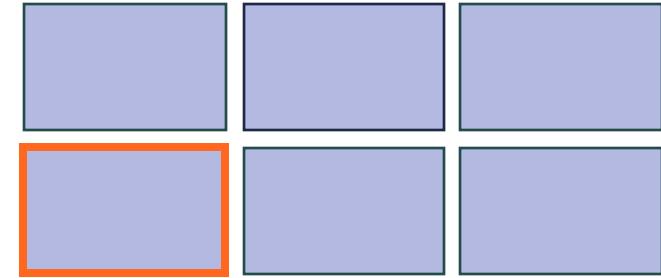
Download only the **data chunk** you need from remote object storage



One sample: >10 Gigabytes!



SpatialData



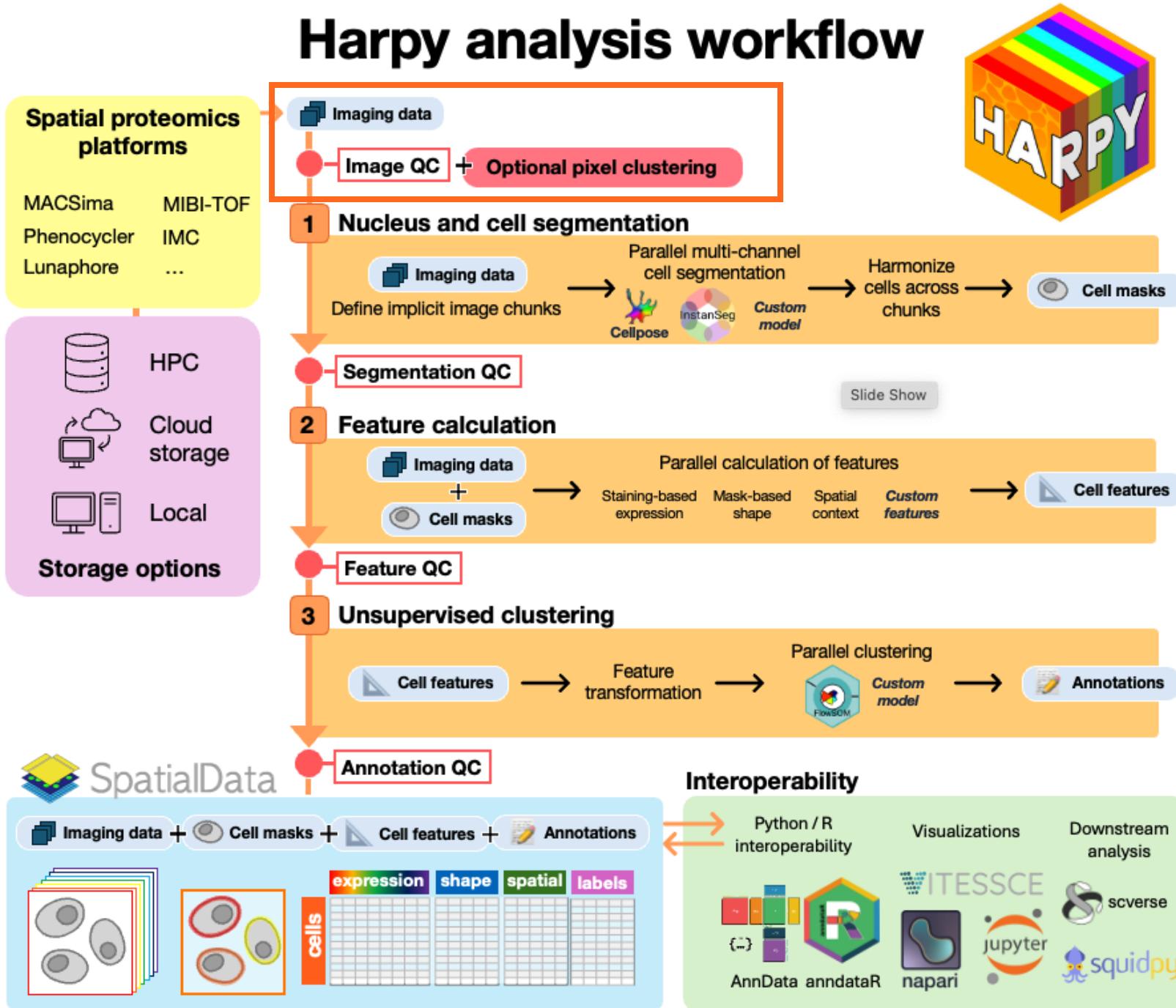
Manageable data chunks



VIB
DATA
CORE

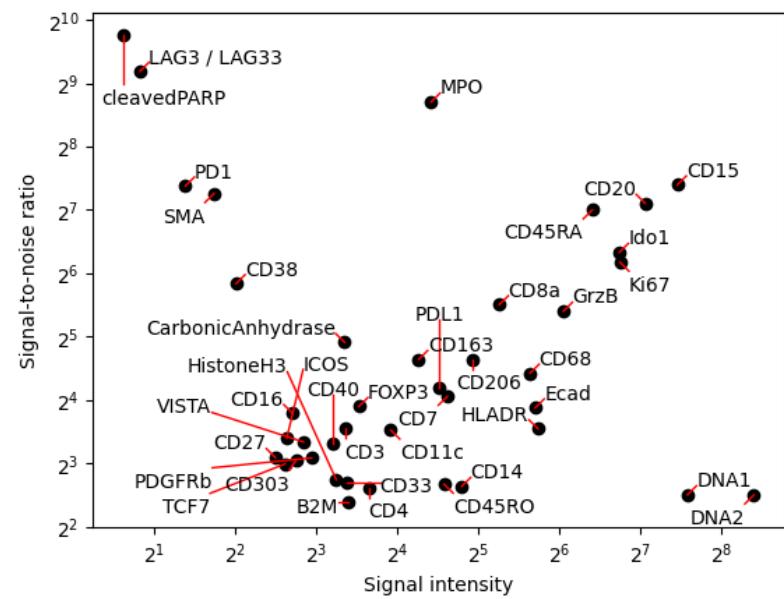
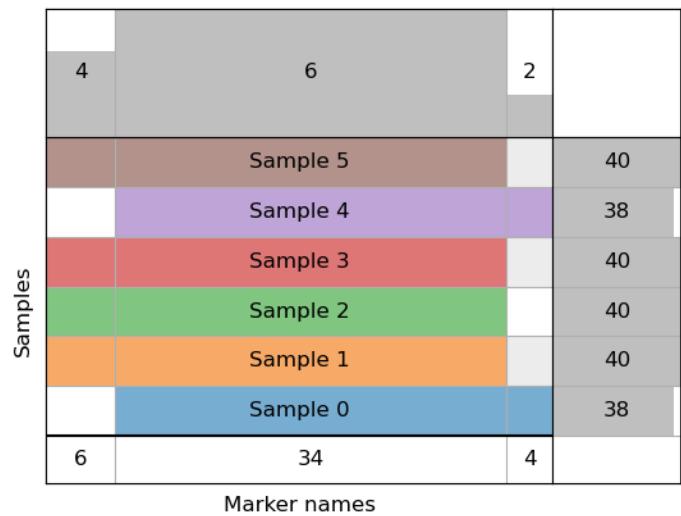
PR to add native support in SpatialData: <https://github.com/scverse/spatialdata/pull/842>

Harpy analysis workflow

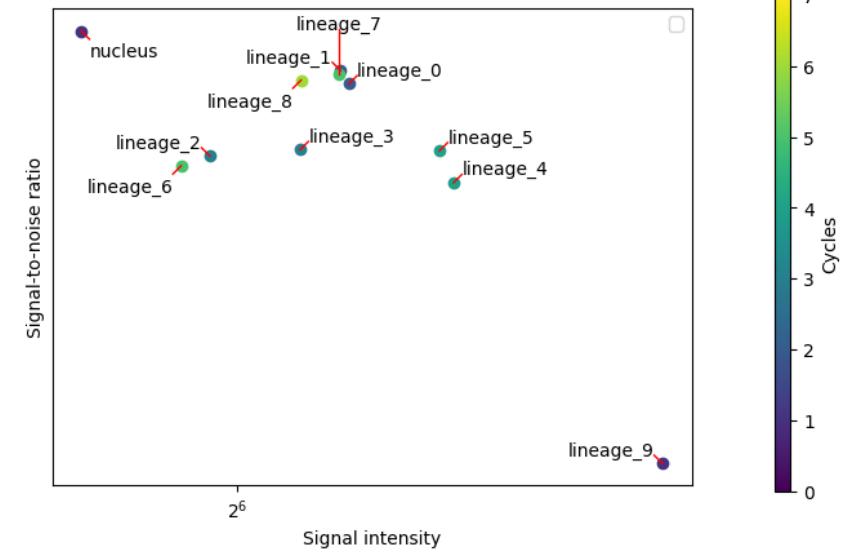


Quality control that scales to your whole dataset

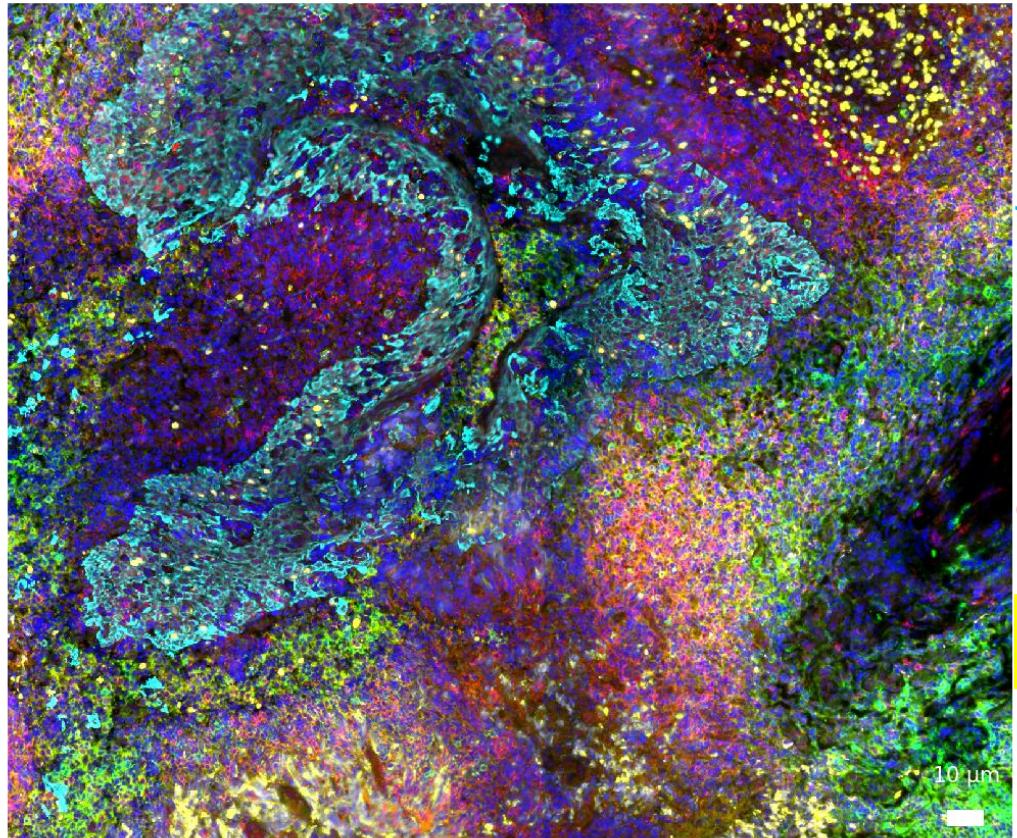
Similar to IMCDataAnalysis, but in **Python**



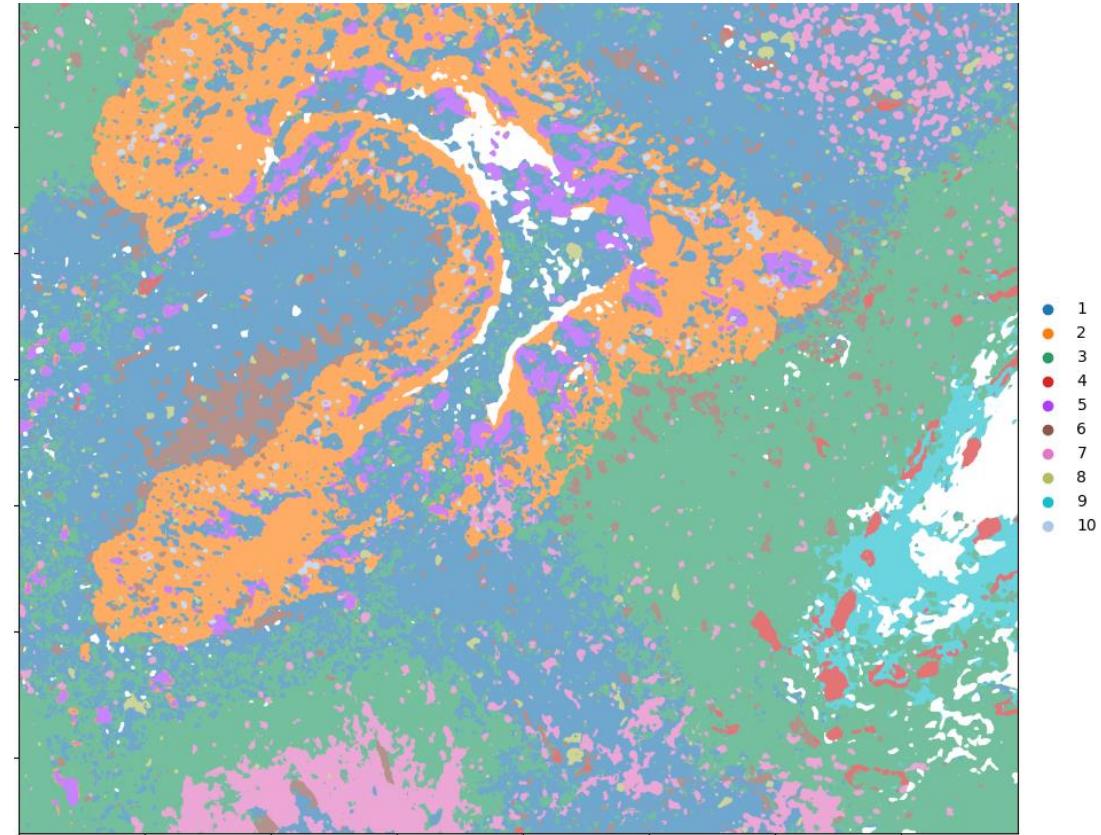
New plots with support for metadata e.g. **cycle order**



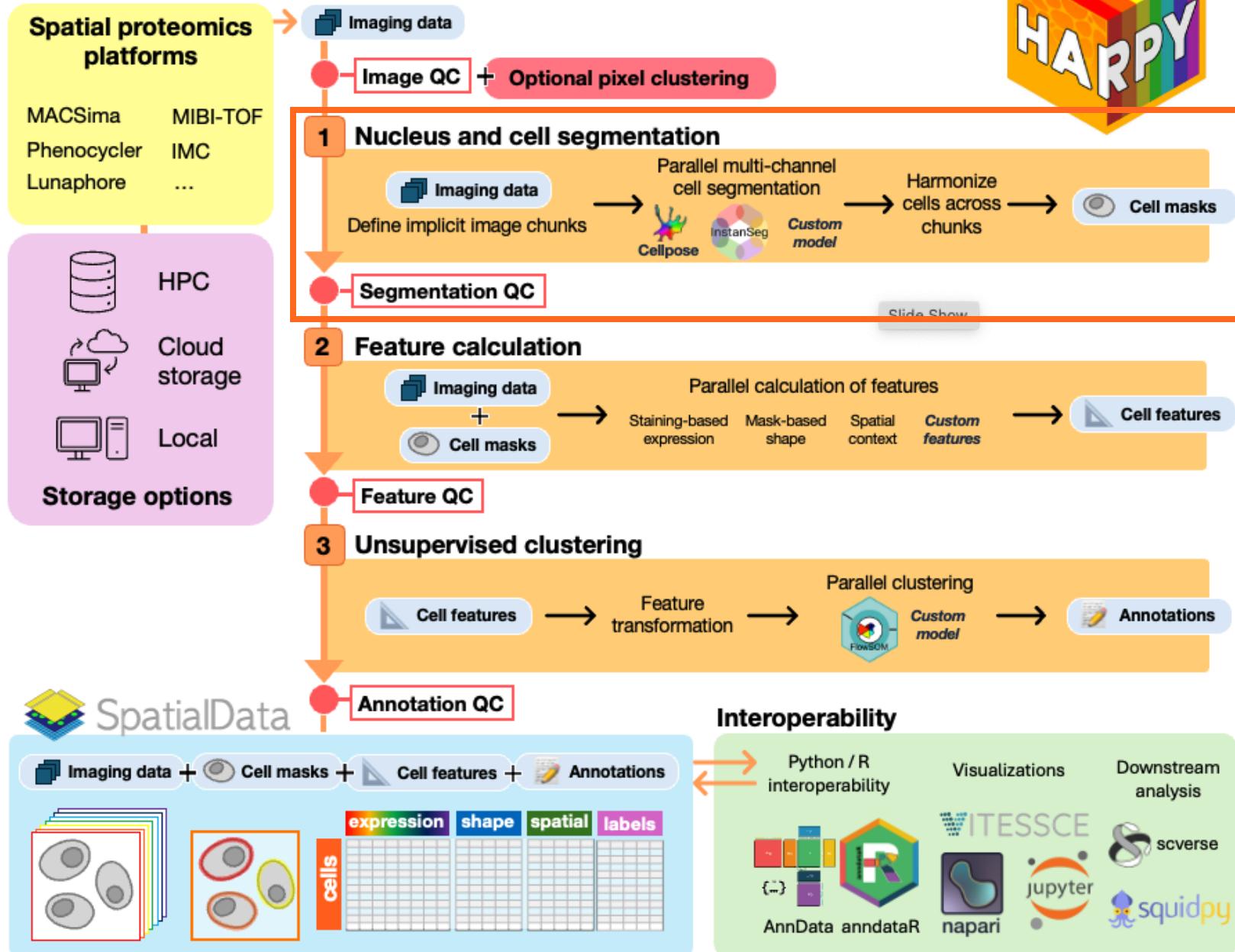
Pixel clustering allows tissue overview without marker and range selection



Pixel clustering with FlowSOM



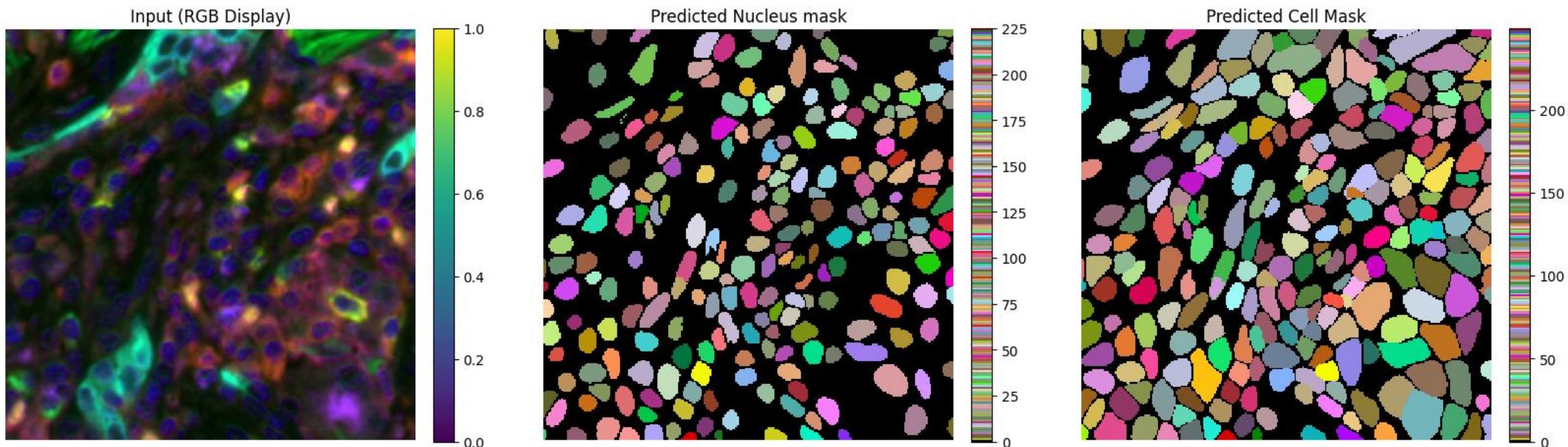
Harpy analysis workflow



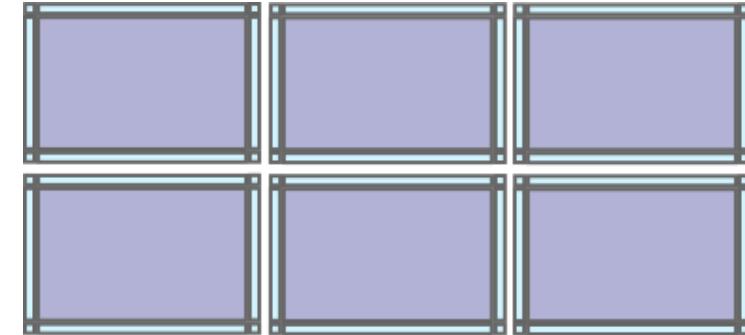
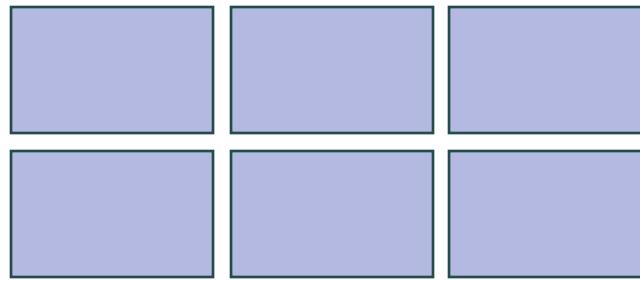
Harpy can apply the latest cell segmentation models



Works on **any combination** of image channels in **any order**
Outputs both predicted **nucleus** and **cell** mask



Harpy is fast because of Dask parallelisation



Out-of-memory
Very slow

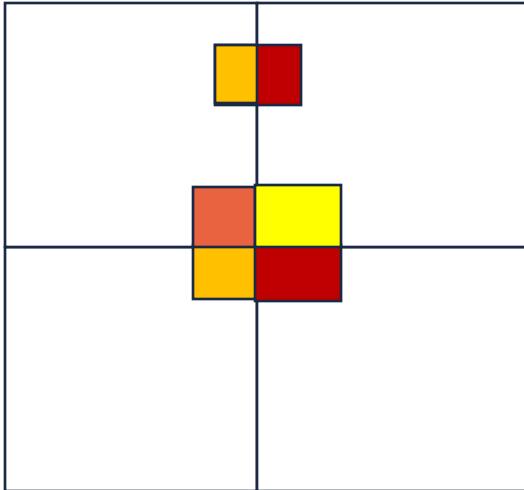


Fast
Border artefacts

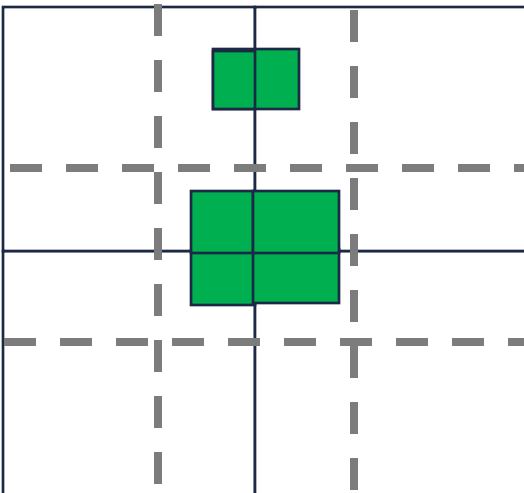


Fast, no memory issues
No artefacts
No manual chunks

Overlapping tiles are correct and quick



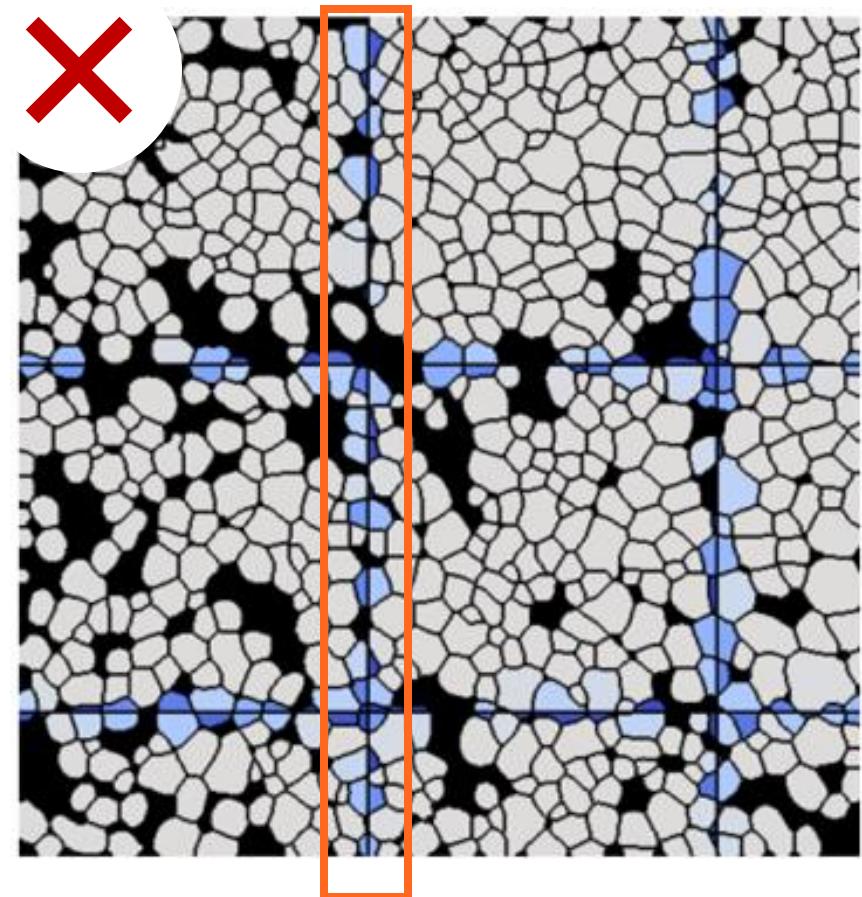
6 cells segmented using
naive apply over 4 separate tiles
= tile artefacts



2 cells segmented using
apply over 4 tiles **with Dask map_overlap**
= no tile artefacts... if you merge labels correctly!

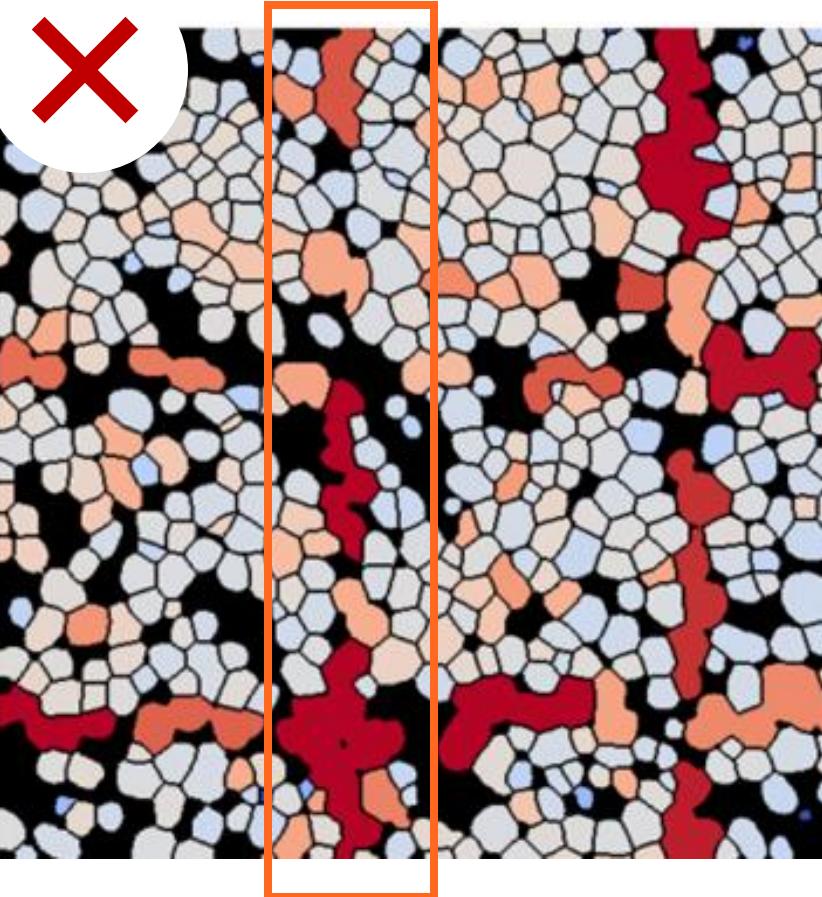
Good label merging is essential

Naive dask.map_overlap



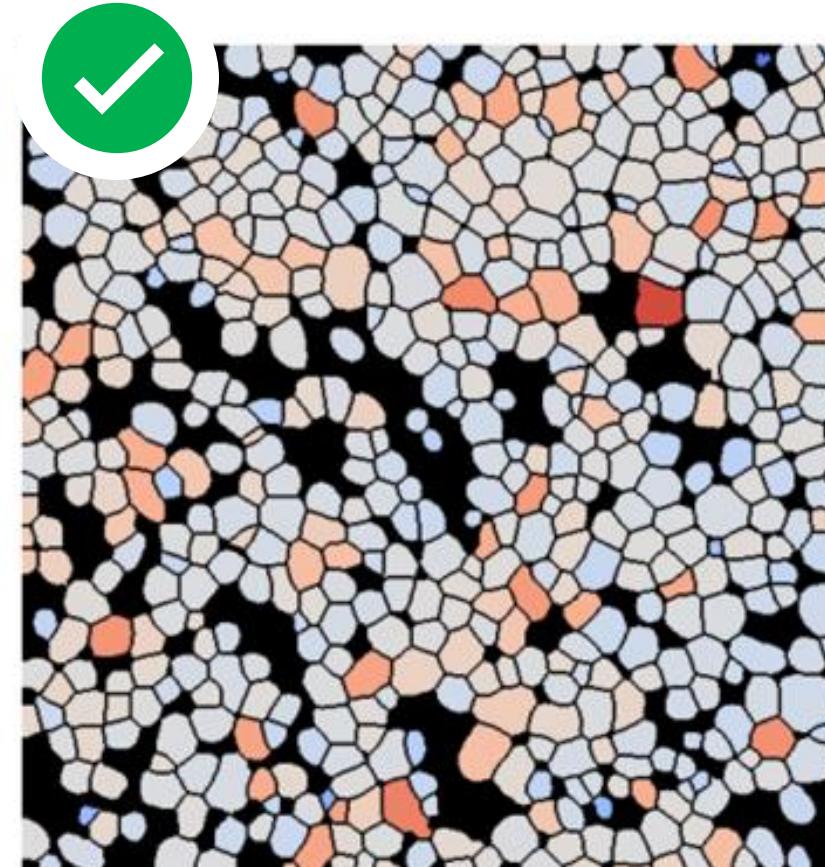
No label merges at tile borders

Naive connected_components or squidpy.segment



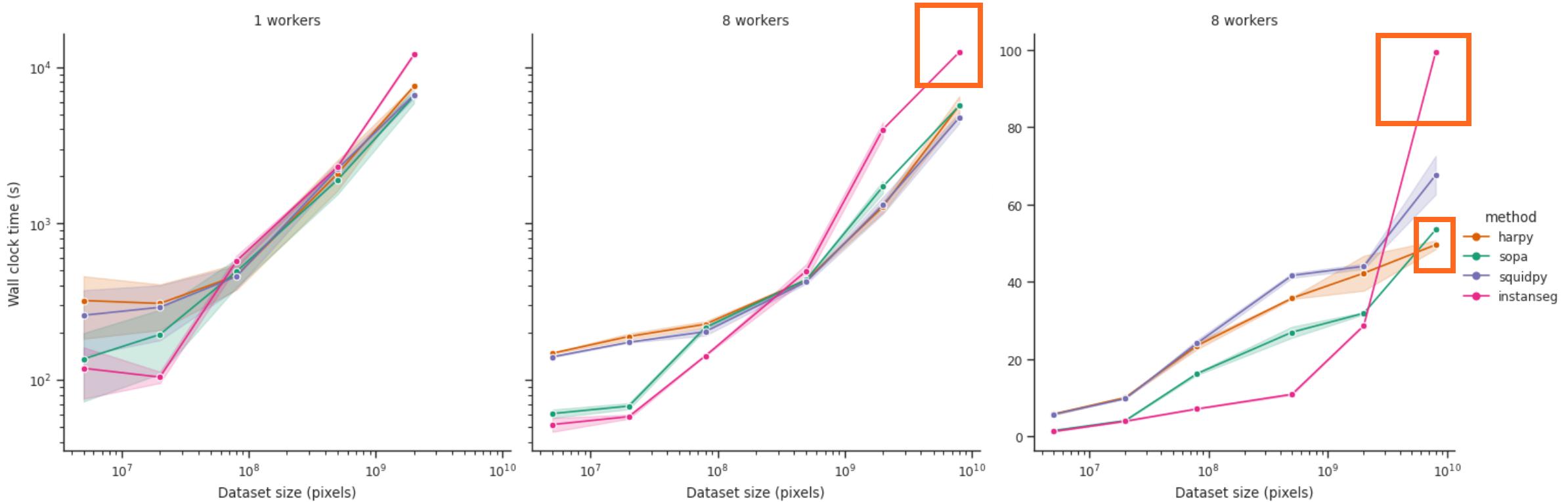
Too many label merges at tile borders

harpy.im.segment

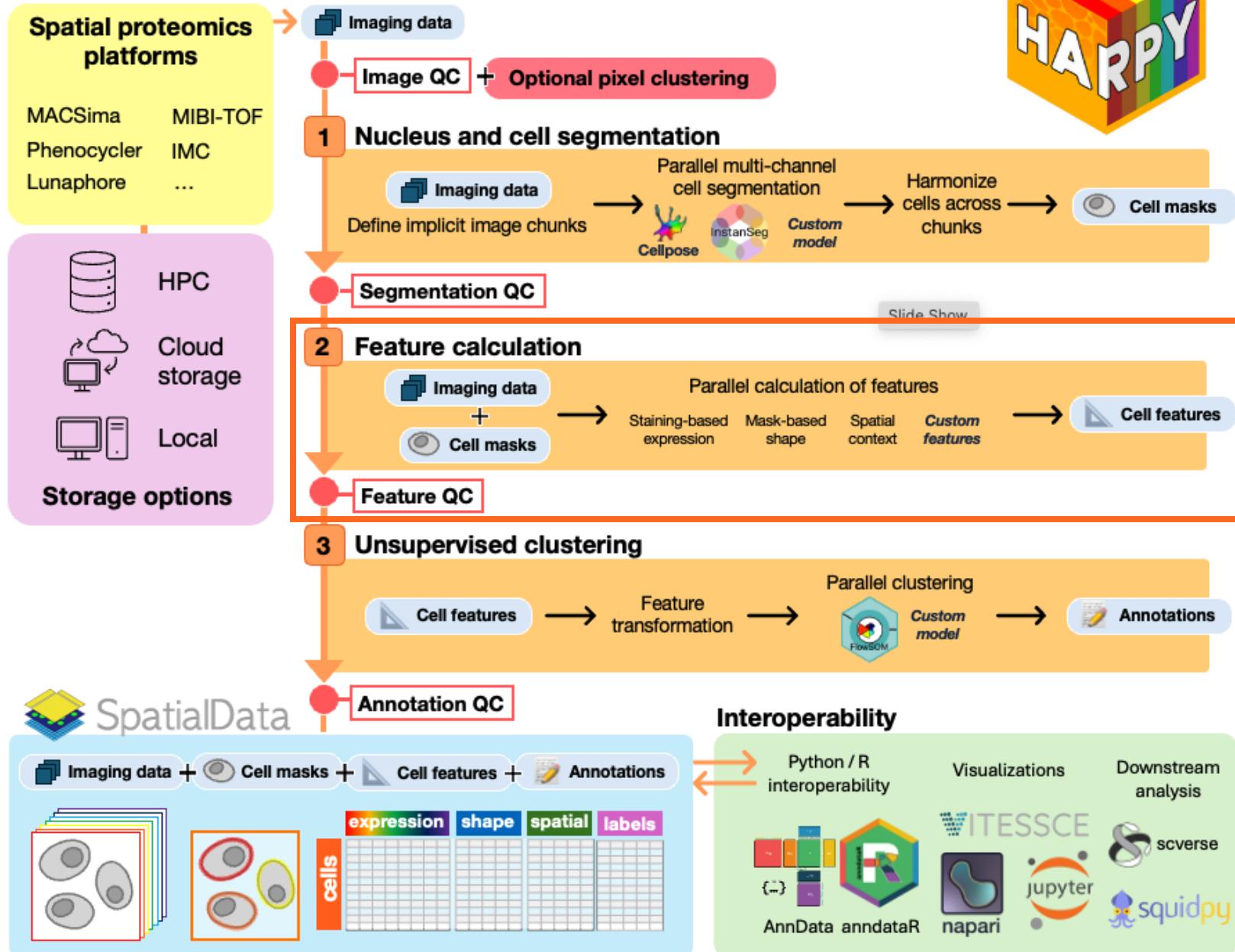


Correct label merges

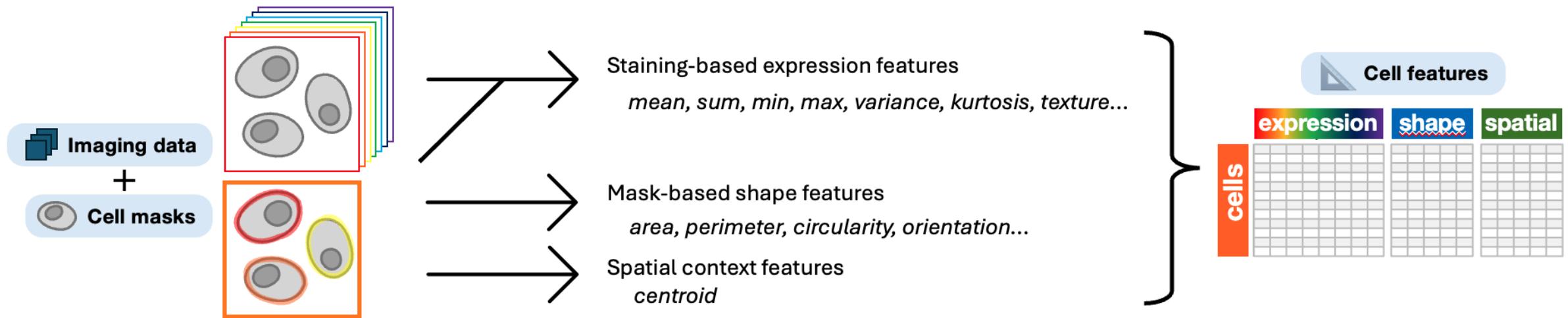
Correct and scalable segmentation



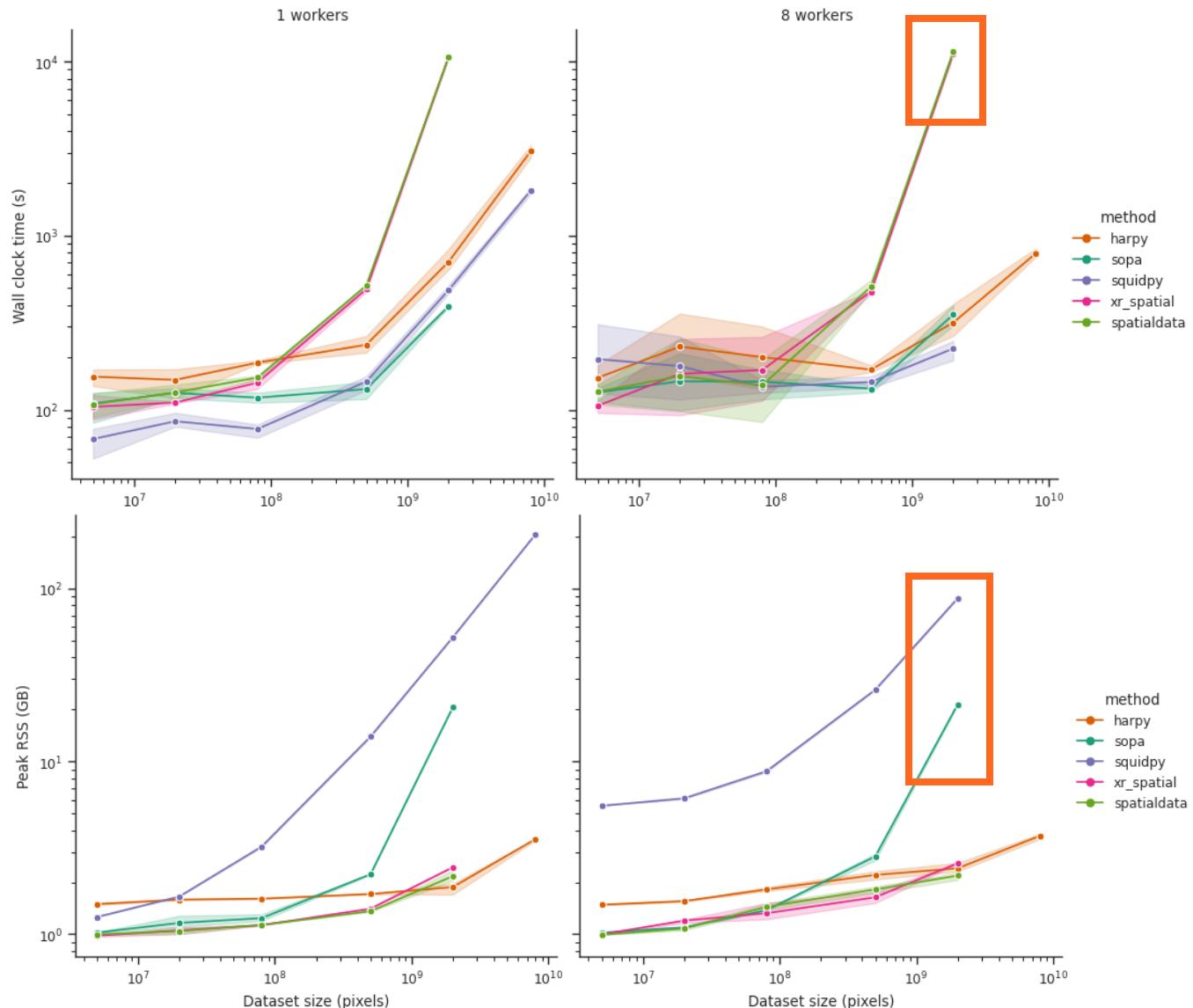
Harpy analysis workflow



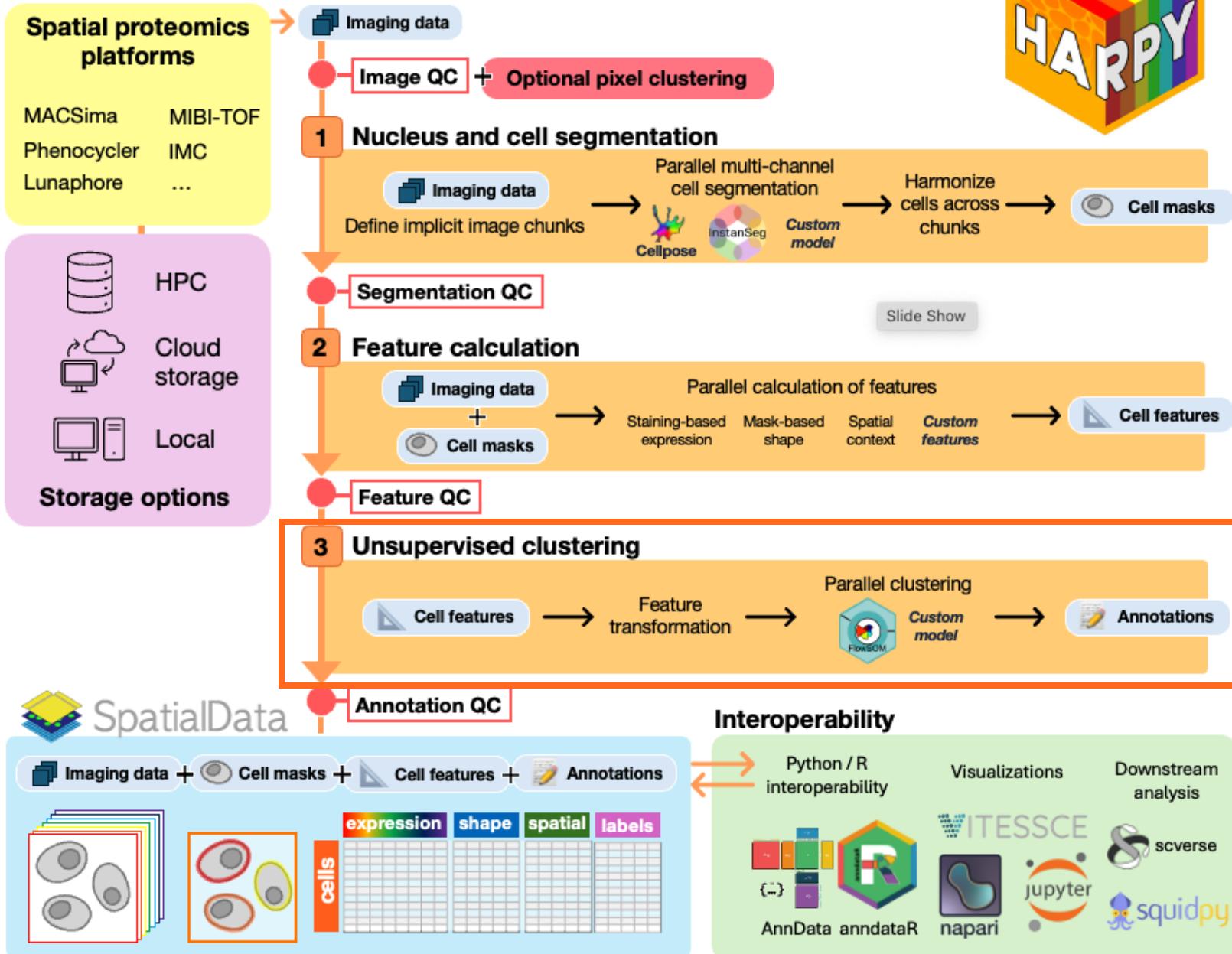
Different features from images and masks



Whole slide feature calculation that works

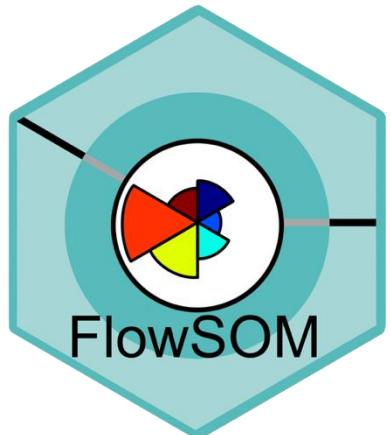


Harpy analysis workflow

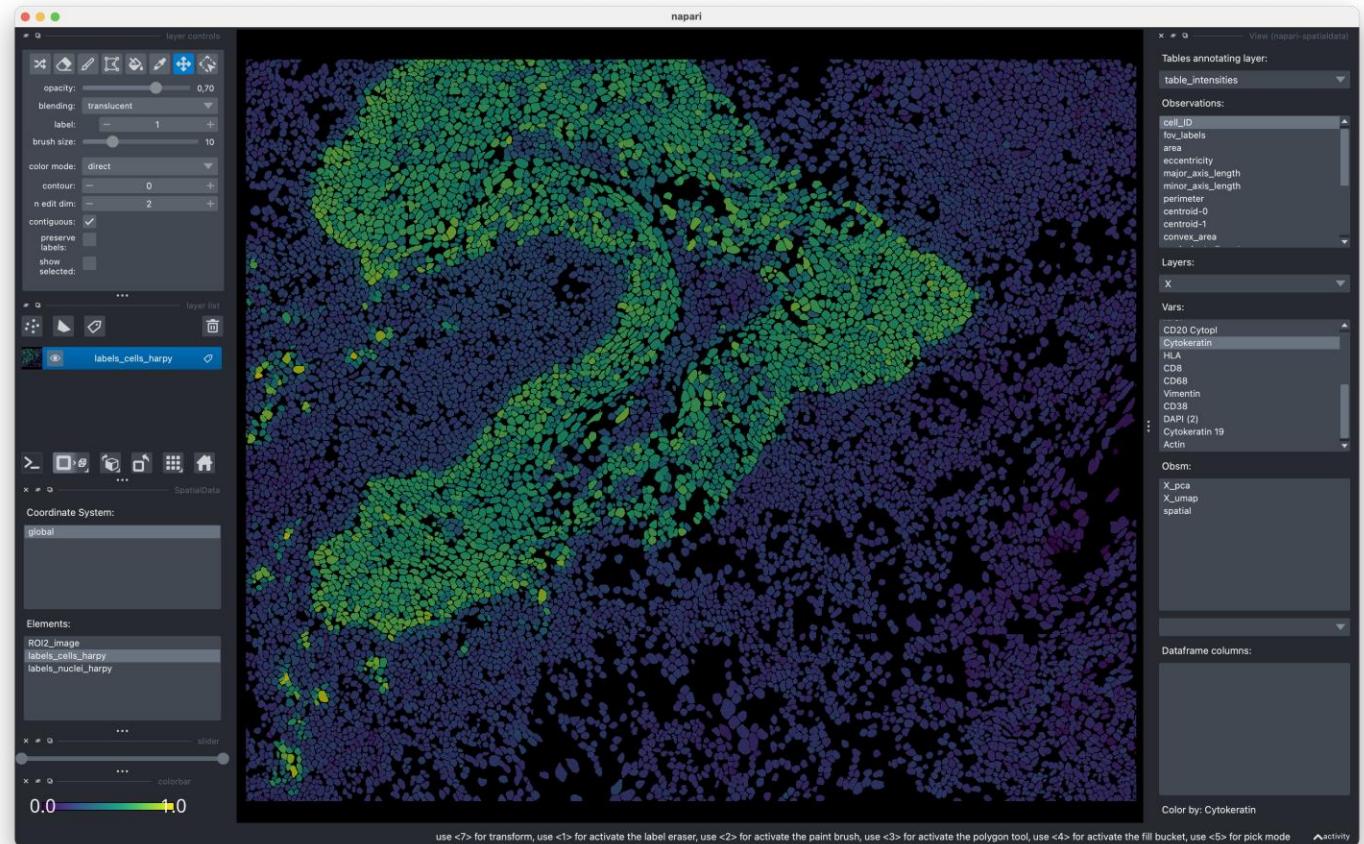


Interactive unsupervised clustering

Cell and pixel clustering workflow
using FlowSOM in Python



napari

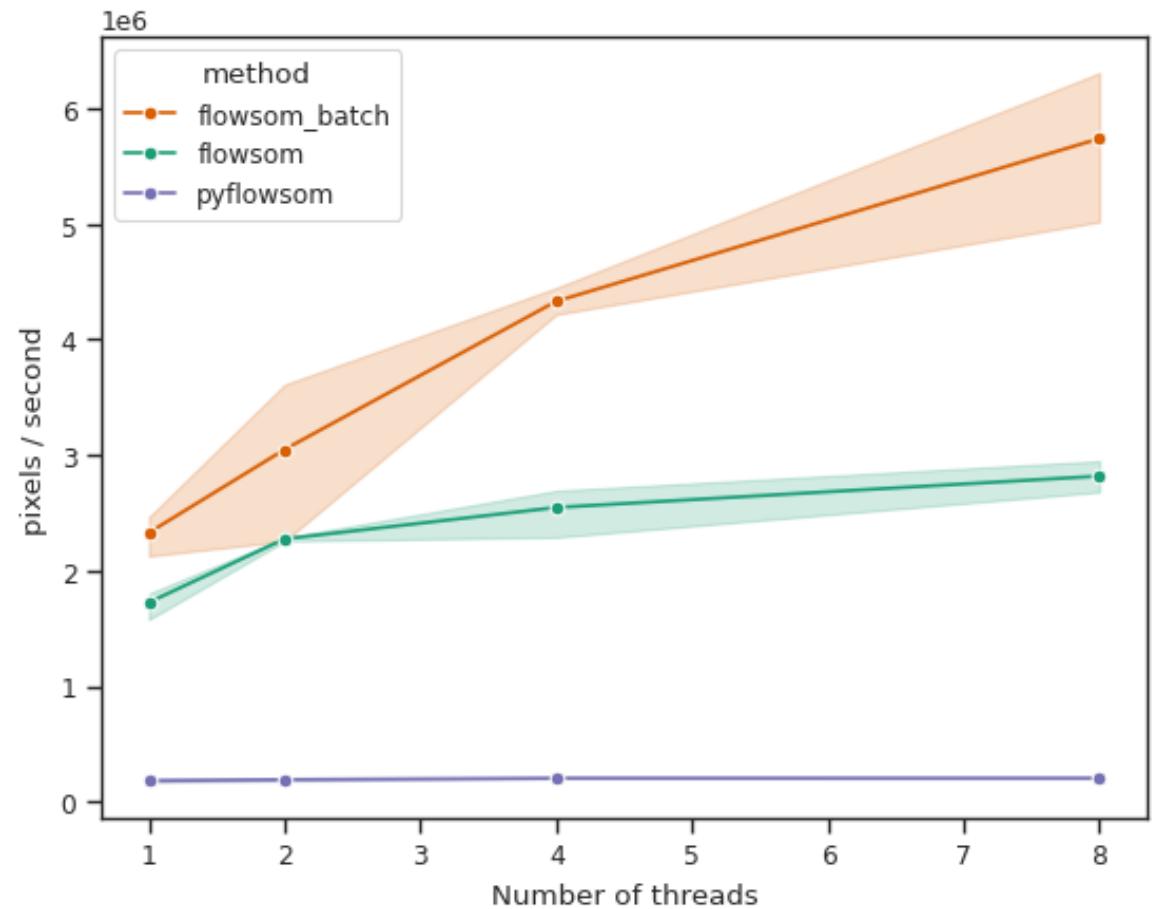
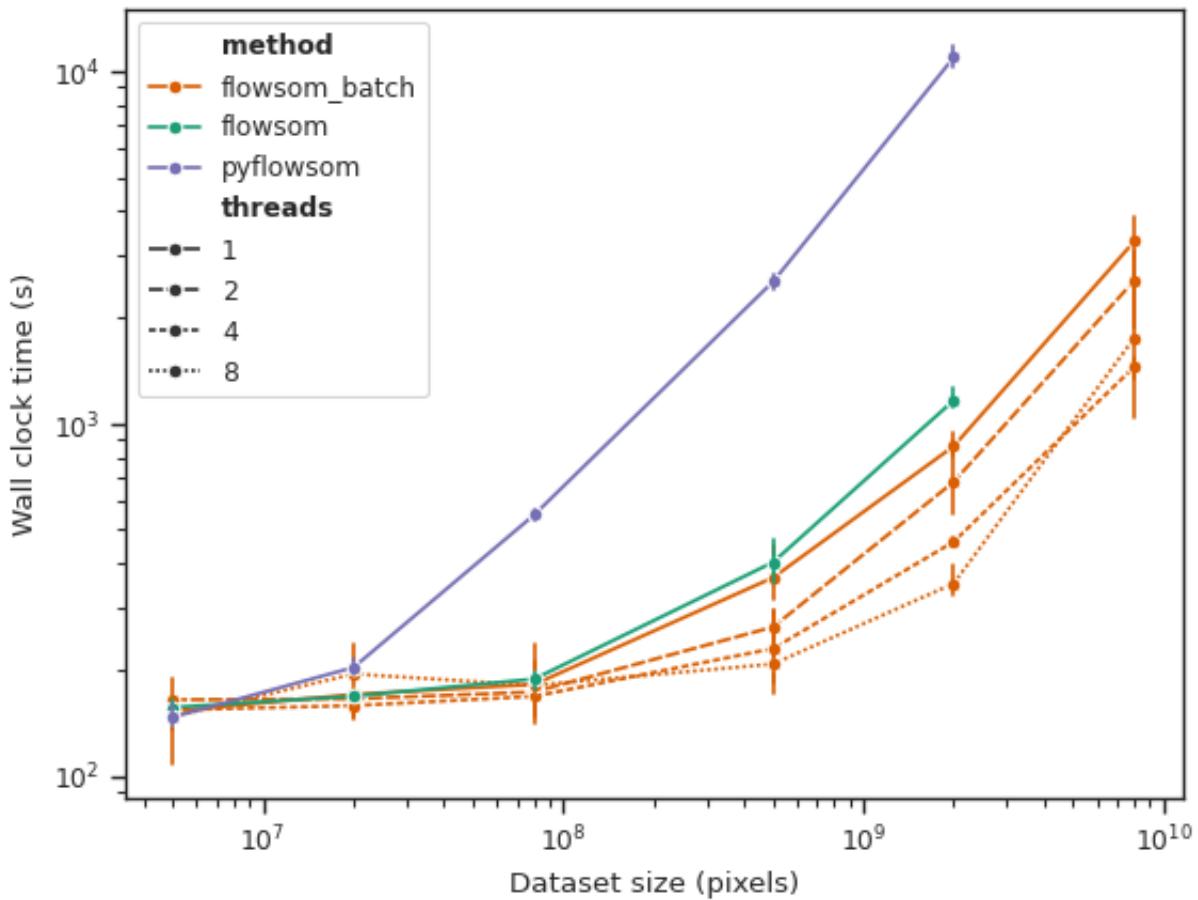


Visualize results in notebook, with napari or
with other SpatialData plugins like e.g. napari-spatialdata

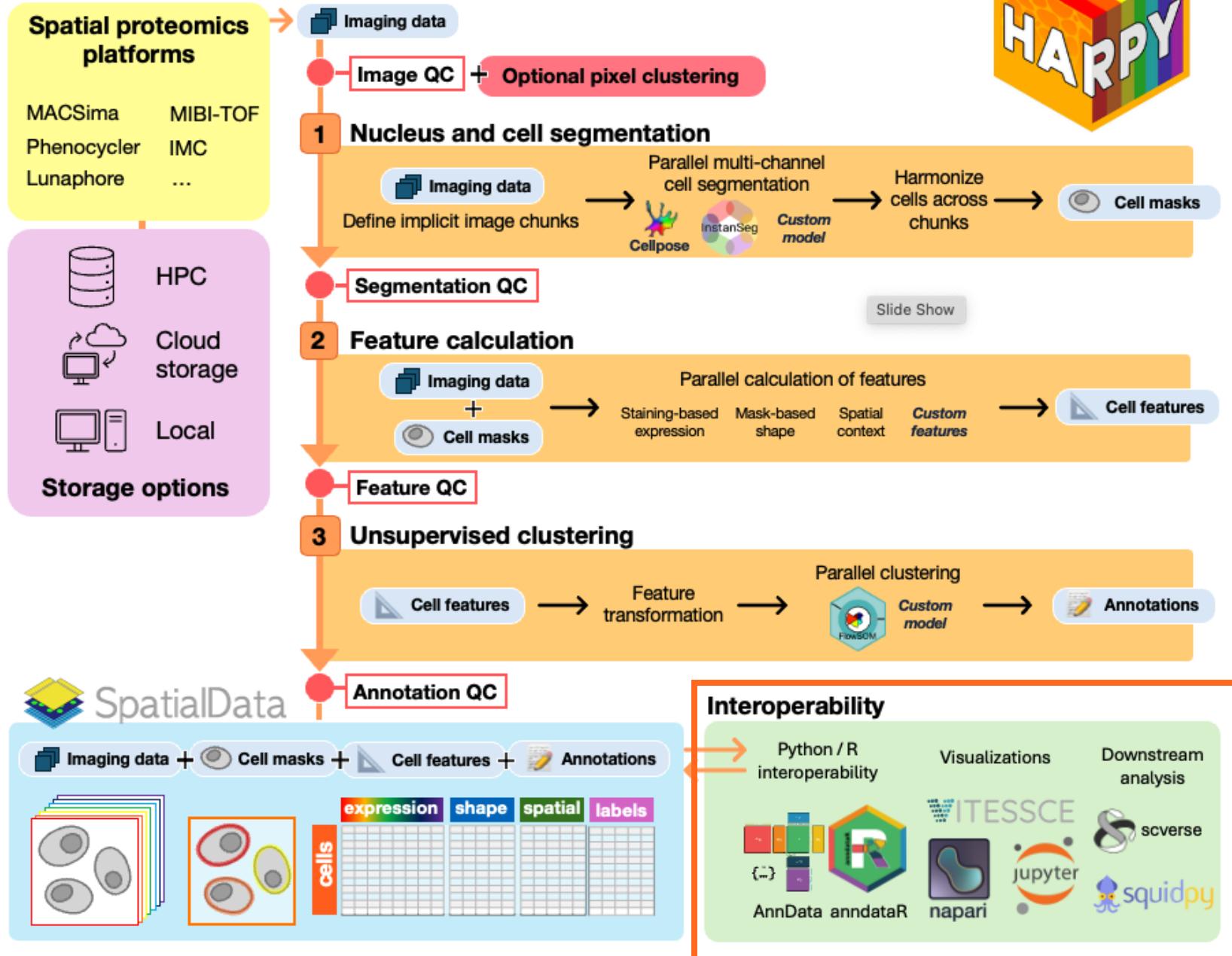
Liu, C.C., Greenwald, N.F., Kong, A., McCaffrey, E.F., Leow, K.X., Mrdjen, D., Angelo, M., 2022.

Robust phenotyping of highly multiplexed tissue imaging data using pixel-level clustering. <https://doi.org/10.1101/2022.08.16.504171>

Scalable FlowSOM in Python



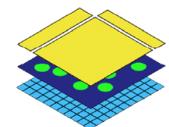
Harpy analysis workflow



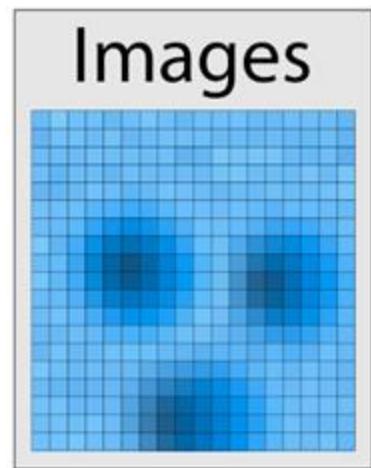
Harpy output is interoperable with R



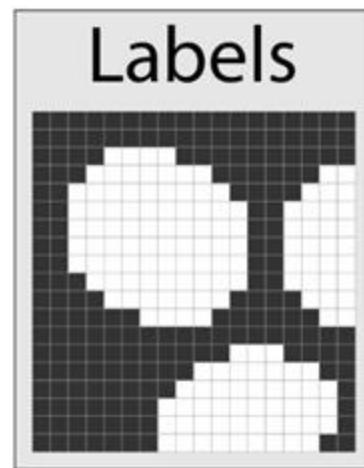
Louise Deconinck



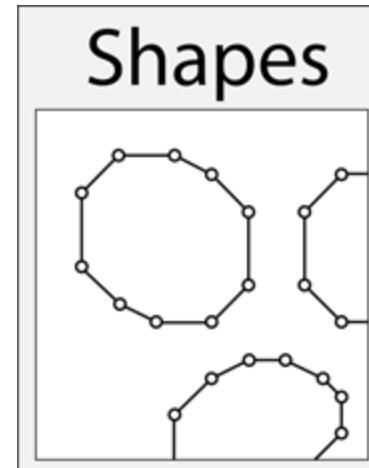
SpatialData



Microscopy
images ...

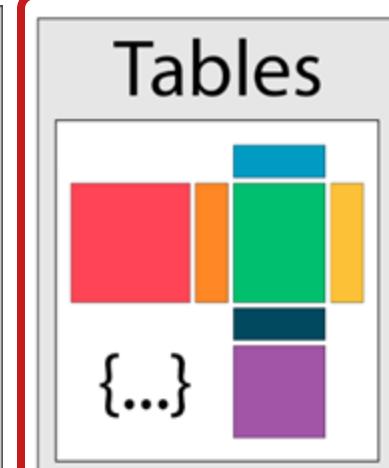


Segmentation
mask



Cells, ROIs
...

Annotates



Cell expression,
cell types ...



SeuratObject
SingleCellExperiment

Cannoodt R, Zappia L, Morgan M, Deconinck L (2025). *anndataR: AnnData interoperability in R*.

R package version 0.99.0, <https://github.com/scverse/anndataR>, <https://anndatar.data-intuitive.com/>.

Future support for complete SpatialData object in R: <https://github.com/HelenaLC/SpatialData>

More on interoperability in our workshop: Polyglot programming for single-cell analysis

AUTHORS

Benjamin Rombaut  

Louise Deconinck  

Robrecht Cannoodt  

PUBLISHED

September 12, 2024

AFFILIATIONS

Saeys lab 

VIB Center for Inflammation Research

Ghent University

Saeys lab 

VIB Center for Inflammation Research

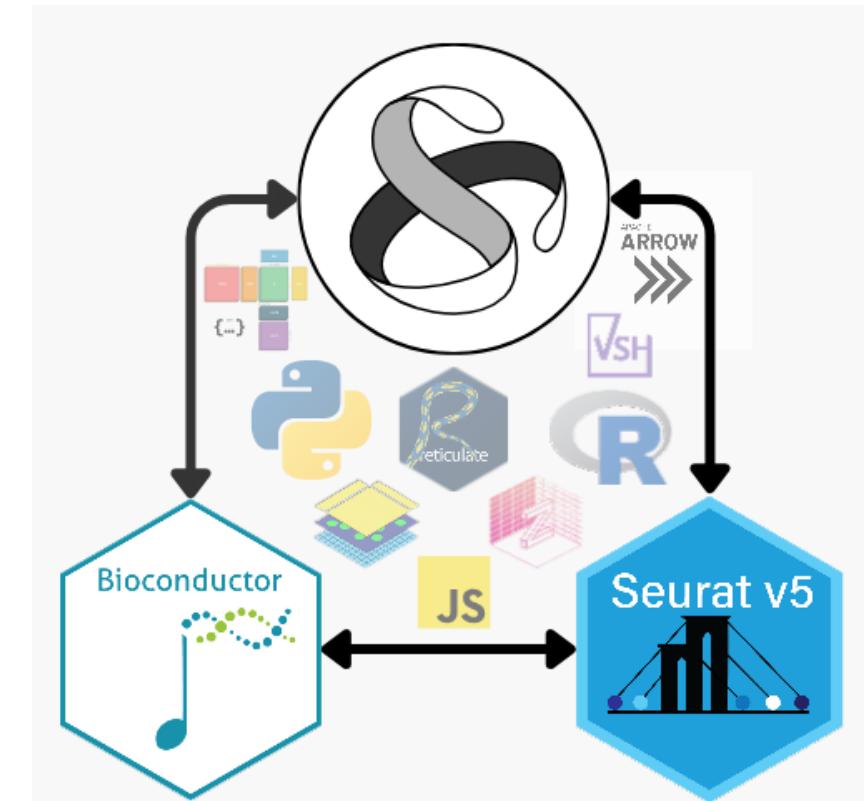
Ghent University

Data Intuitive 

Overview of the **different levels** of interoperability.

Integrating them in a **single workflow**.

<https://saeyslab.github.io/polygloty/>



[Scverse Conference 2024 workshop](#)

4 Harpy highlights

1. Dataset-wide quality control
2. Scalable whole-slide image processing
3. Interactive unsupervised clustering
4. Interoperability with R



<https://github.com/saeyslab/harpy>

Acknowledgments

VIB-UGent Center for Inflammation Research

SaeysLab

Lotte Pollaris

Ruth Seurinck

Sofie Van Gassen

...

Yvan Saeys

VIB Spatial Catalyst

Arne Defauw

Frank Vernaillen

Julien Mortier

Evelien Van Hamme

Questions?

Benjamin.Rombaut@UGent.be

or contact the **VIB Spatial Catalyst**

spatial@vib.be

Benjamin Rombaut

16th of May 2025



Flanders AI
Research Program