

Spatial proteomics analysis that makes you happy

Slides courtesy of
Benjamin Rombaut



Flanders AI
Research Program

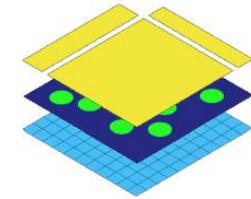
25th of June 2025

Scalable spatial transcriptomics processing



SPArrOW

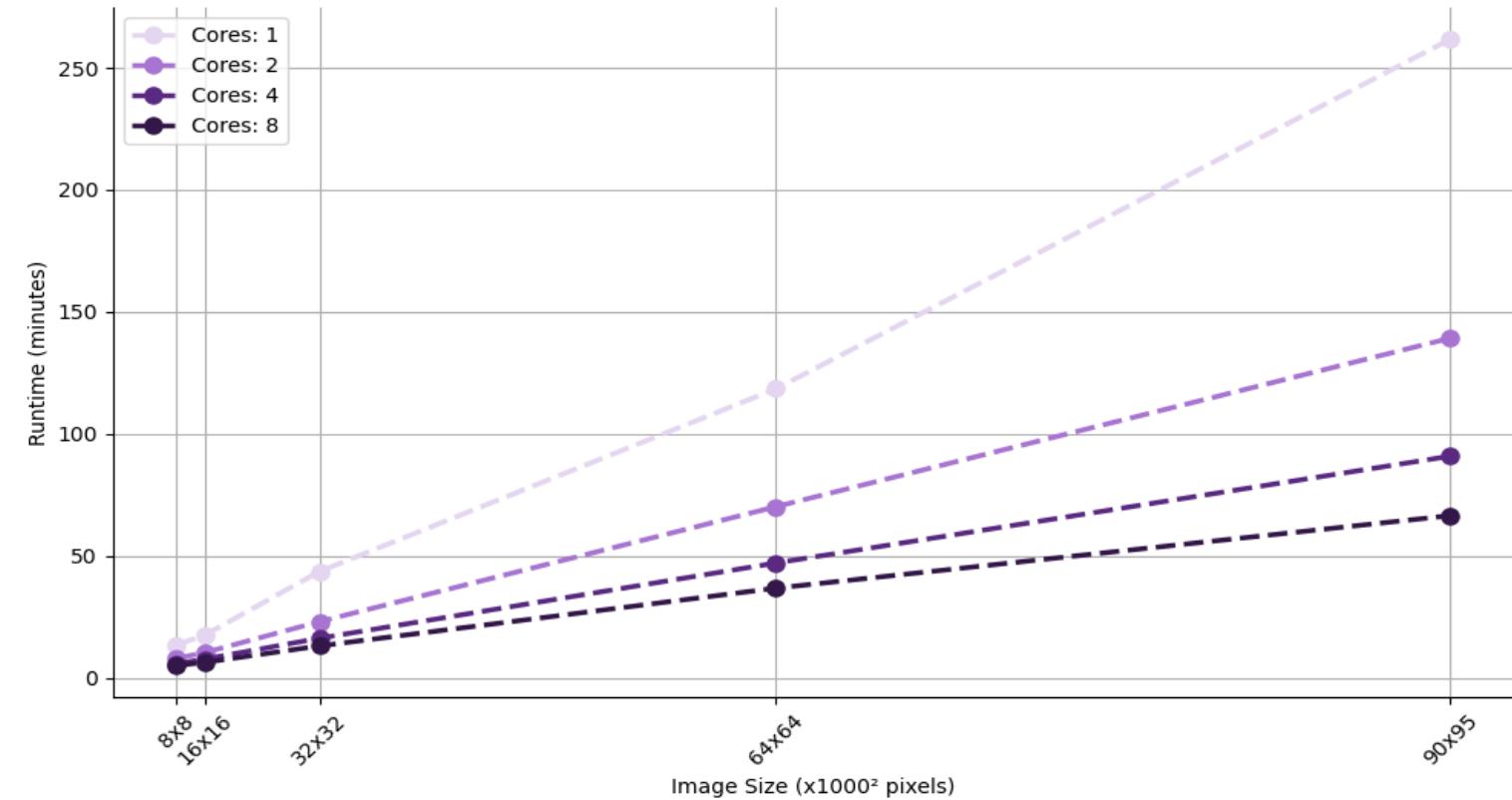
A flexible, interactive and scalable pipeline
for spatial transcriptomics analysis



SpatialData



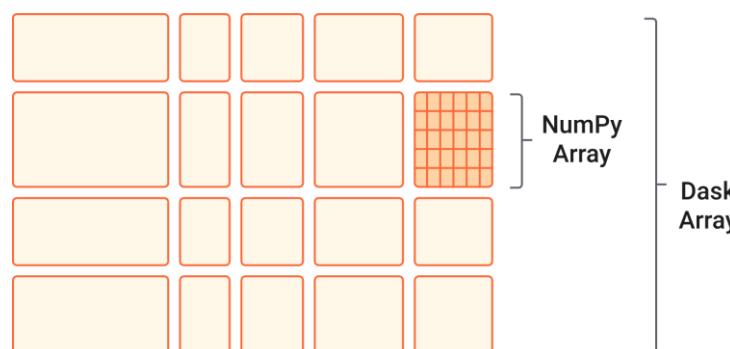
Lotte
Polaris



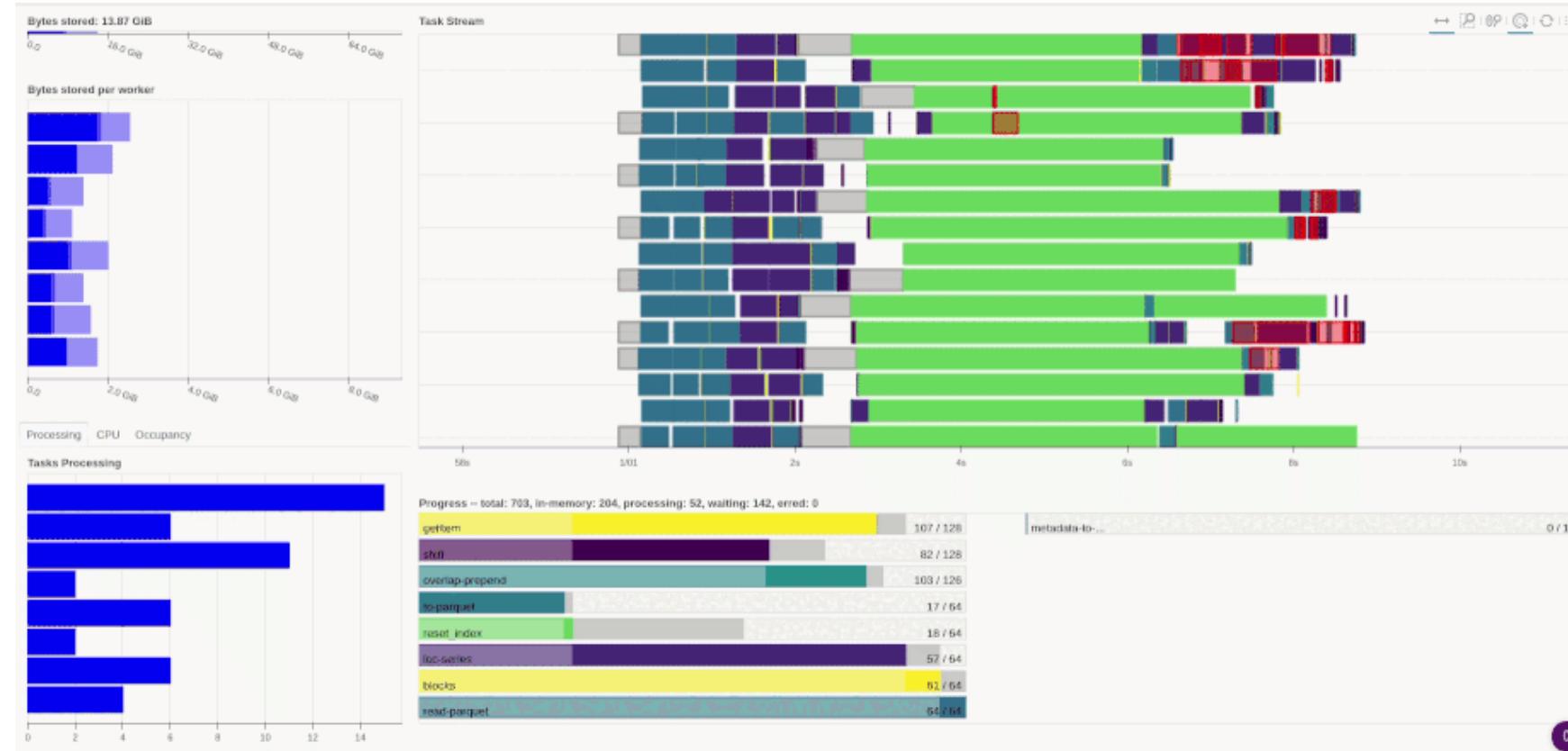
Runtime of whole SPArrOW pipeline on large MERSCOPE dataset
with Dask and GPU acceleration

Polaris, 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>

Dask is a flexible parallel computing library



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

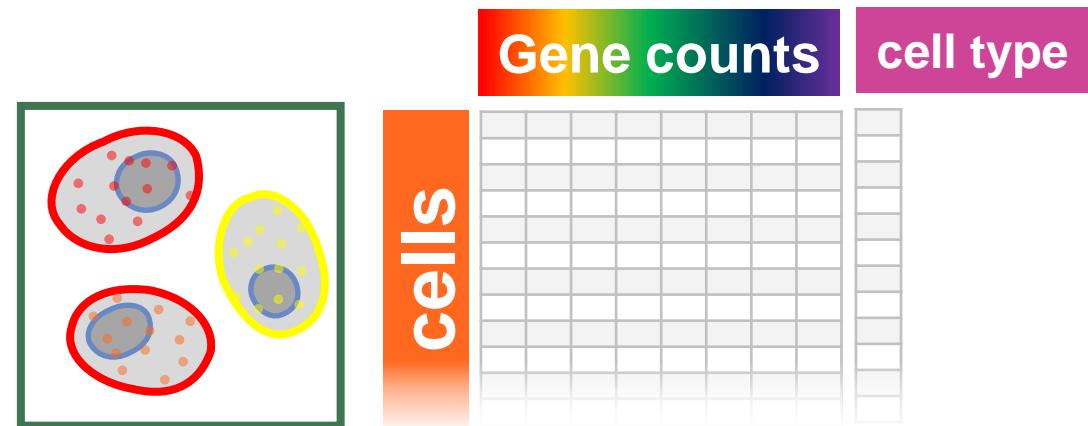
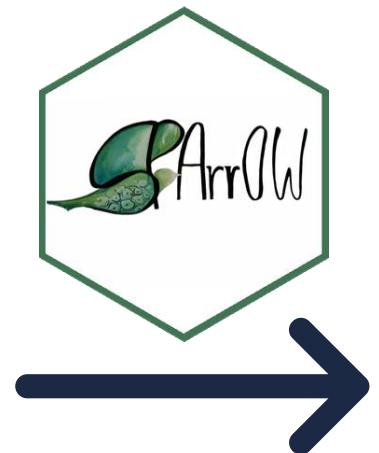
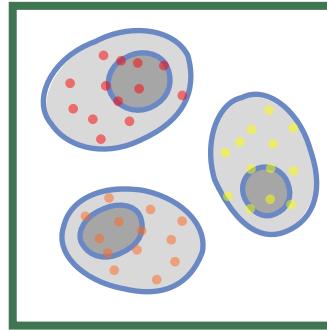


*An analysis run visualized with the Dask Dashboard:
docs.dask.org/en/latest/dashboard.html*

Image-based spatial proteomics analysis is a lot of images

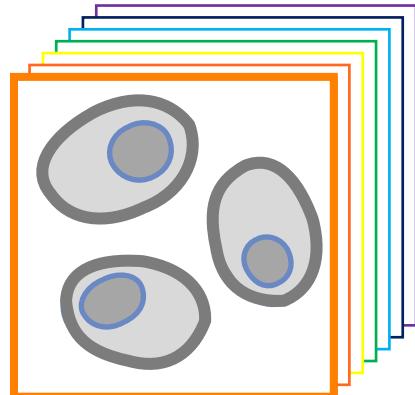
Spatial transcriptomics

Transcript locations +
whole-slide cell image



Spatial proteomics

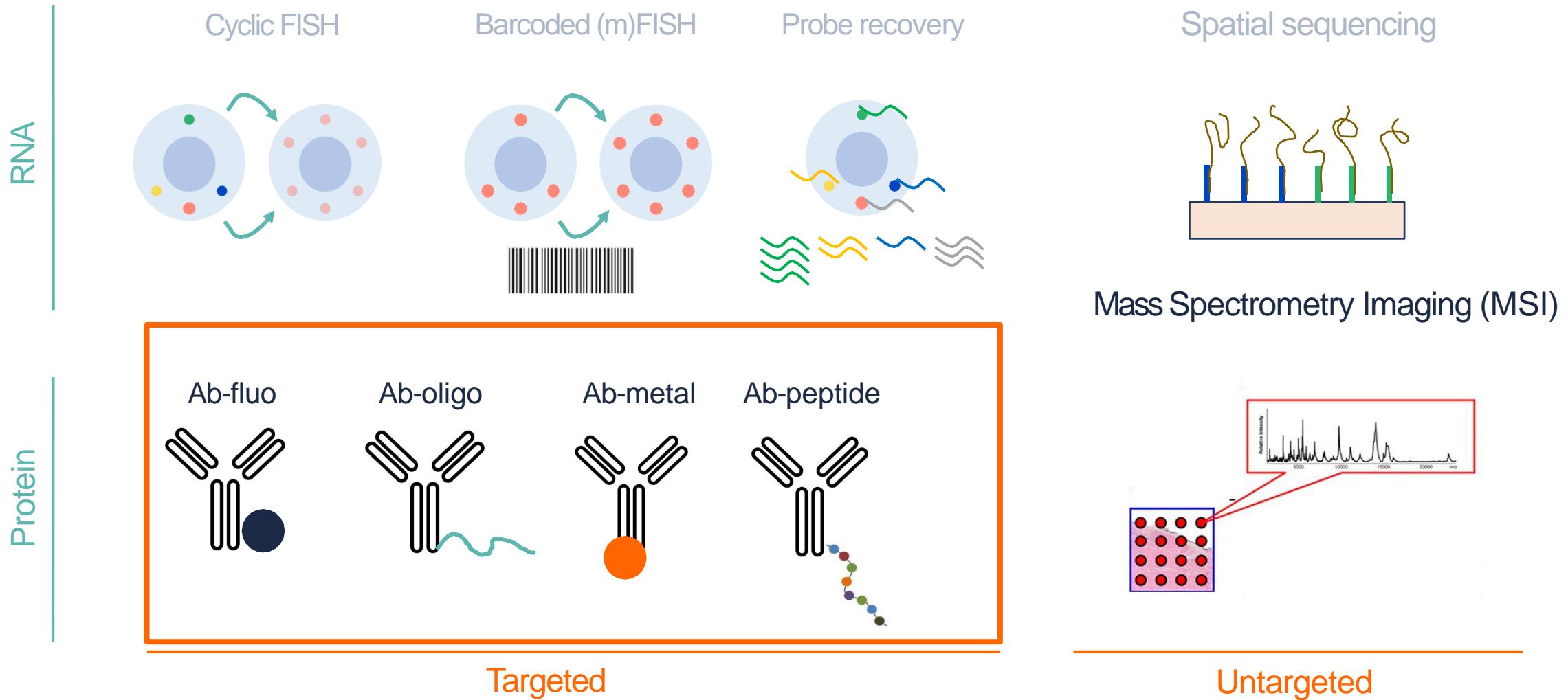
whole-slide
highly-multiplexed images



annotated single-cell data with spatial context



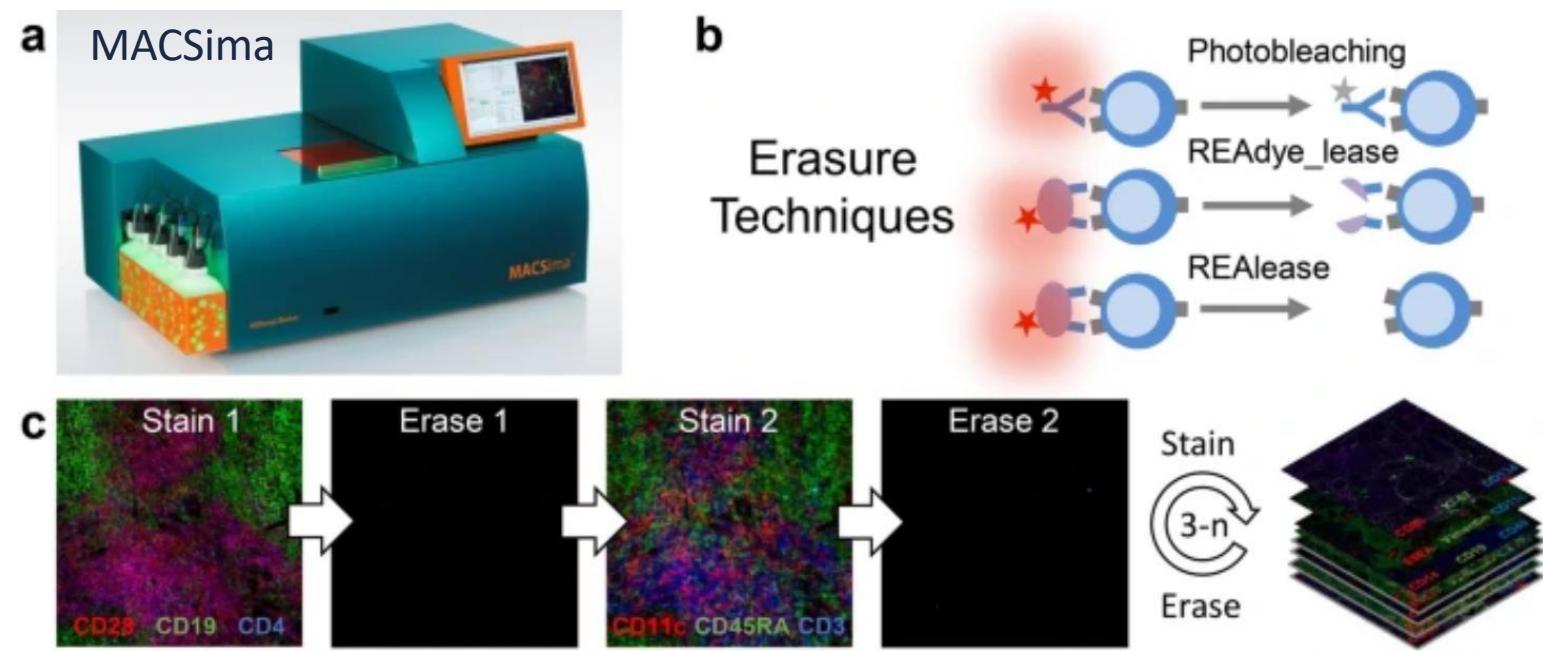
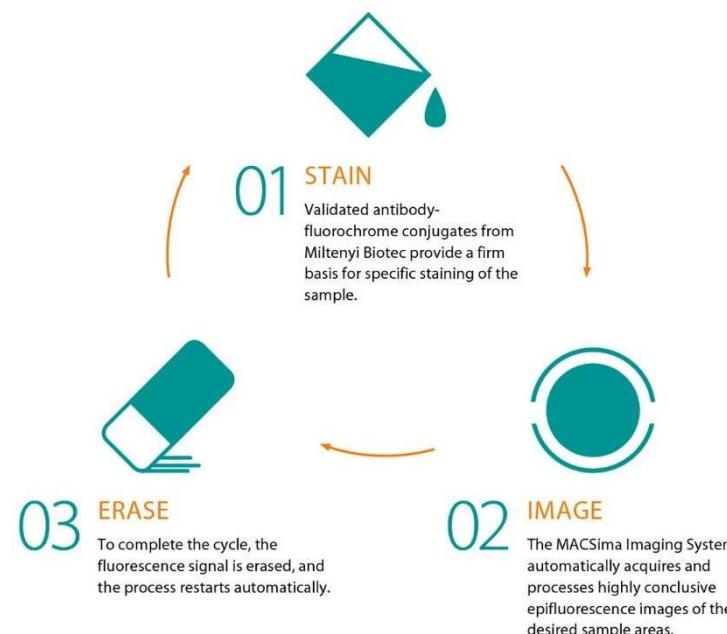
Targeted spatial proteomics



Targeted (imaging-based) spatial proteomics

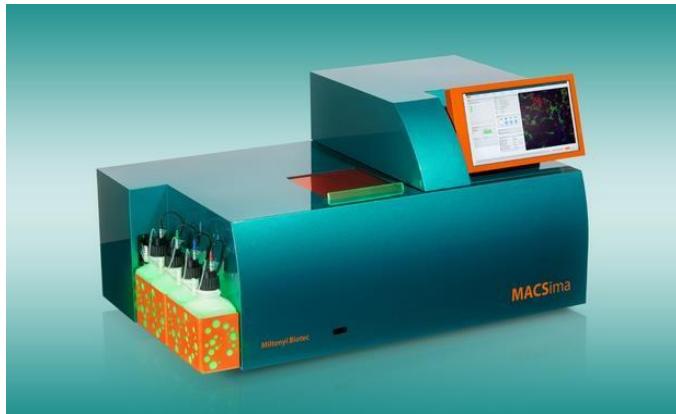
Platforms: MACSima (Miltenyi Biotec), PhenoCycler Fusion (Akoya), OPAL (Akoya), COMET (Lunaphore), CellScape...

Plexity: 6- to 200-plex

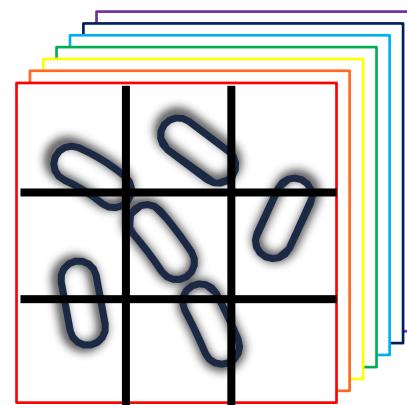


Raw image tiles need to be stitched

Multiplexed imaging



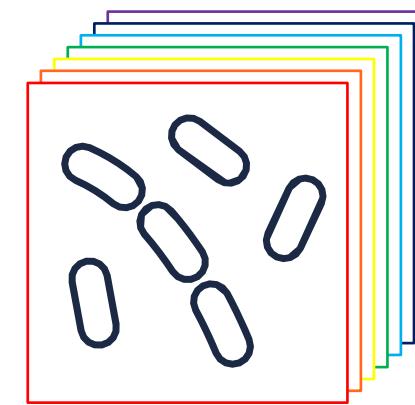
Level 1 Data
Raw image tiles



**REGISTRATION
STITCHING**
→
Commercial
preprocessing
ASHLAR

e.g. multiple multichannel
 1000×1000 tiles

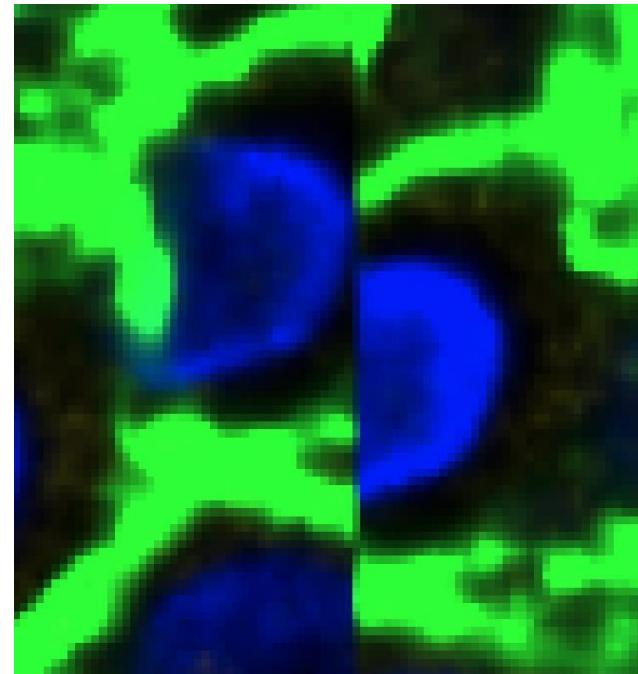
Level 2 Data
Whole-slide images



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

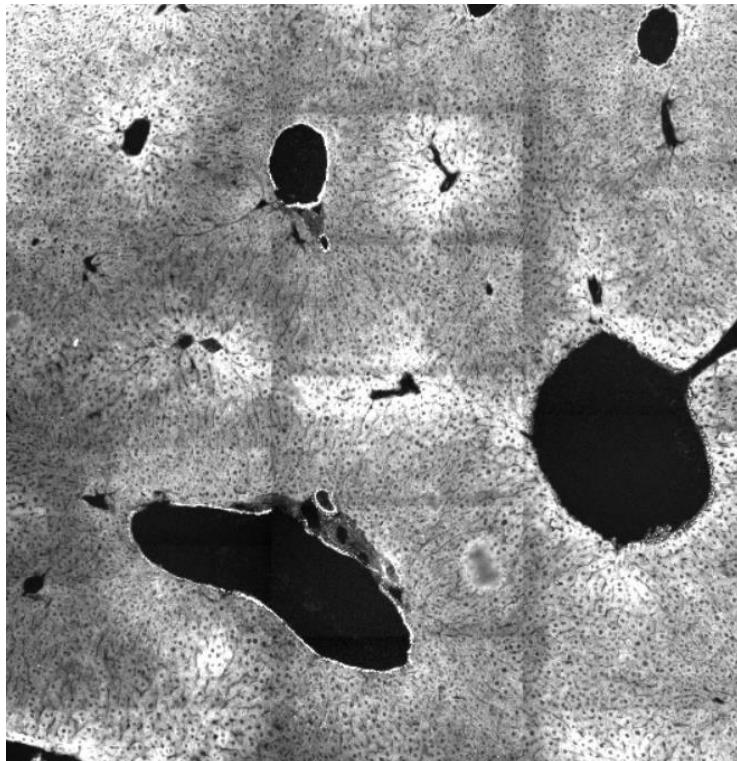
Difficulties of multiplexed bioimaging

Stitching artifacts

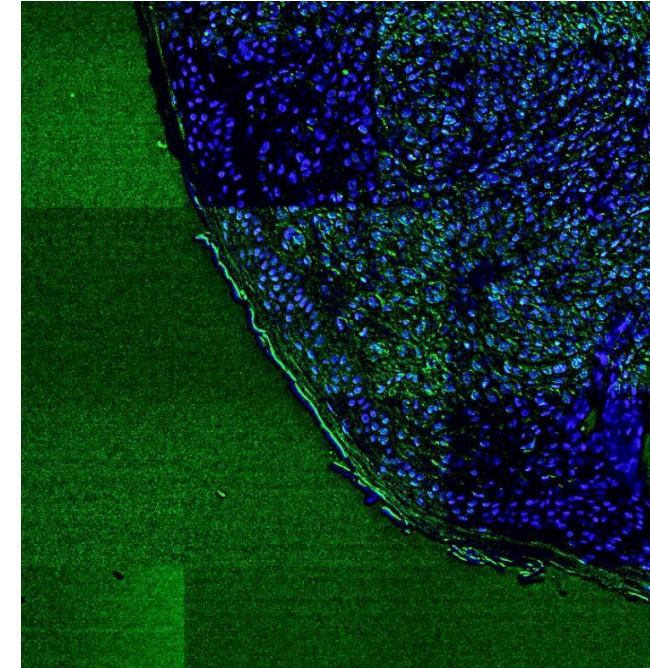


Difficulties of multiplexed bioimaging

Autofluorescence



Background noise



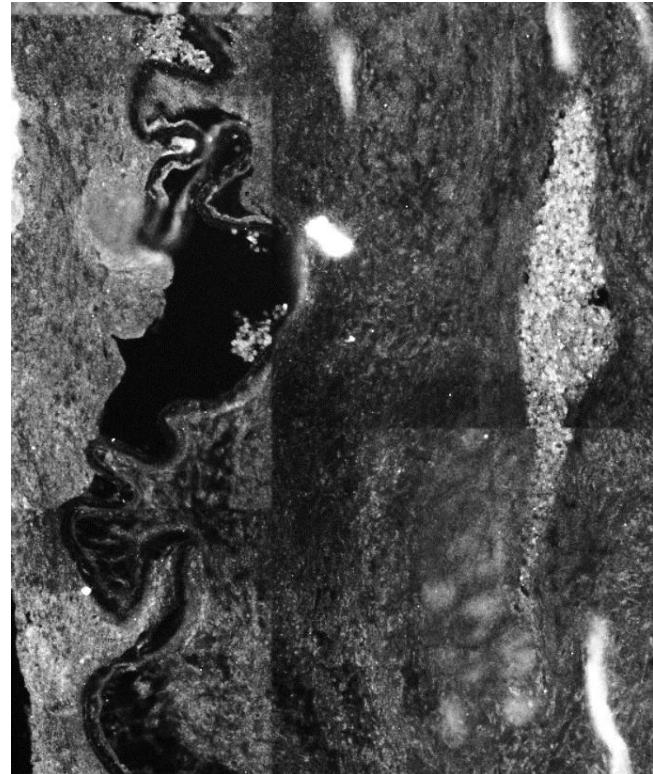
Difficulties of multiplexed bioimaging

Staining artifacts



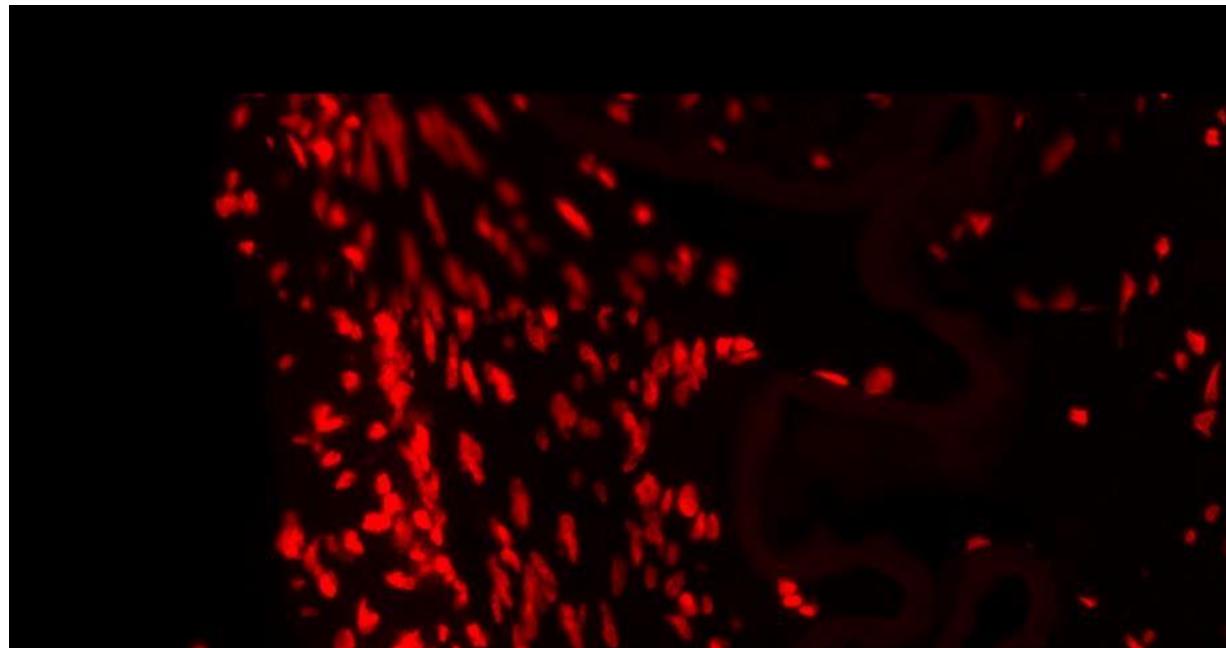
Difficulties of multiplexed bioimaging

Acquisition
bleaching



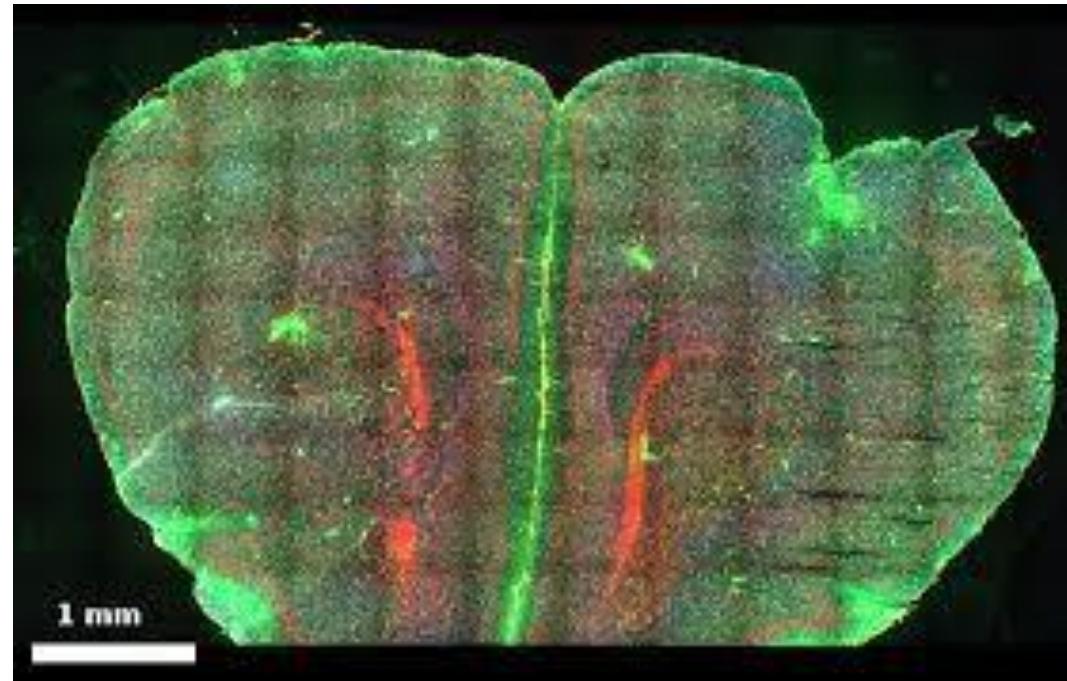
Difficulties of multiplexed bioimaging

Bad focus



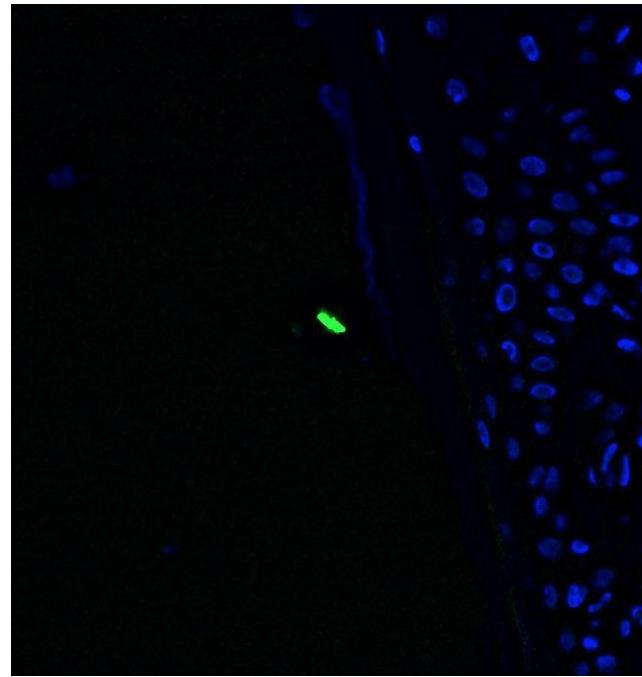
Difficulties of multiplexed bioimaging

Uneven illumination



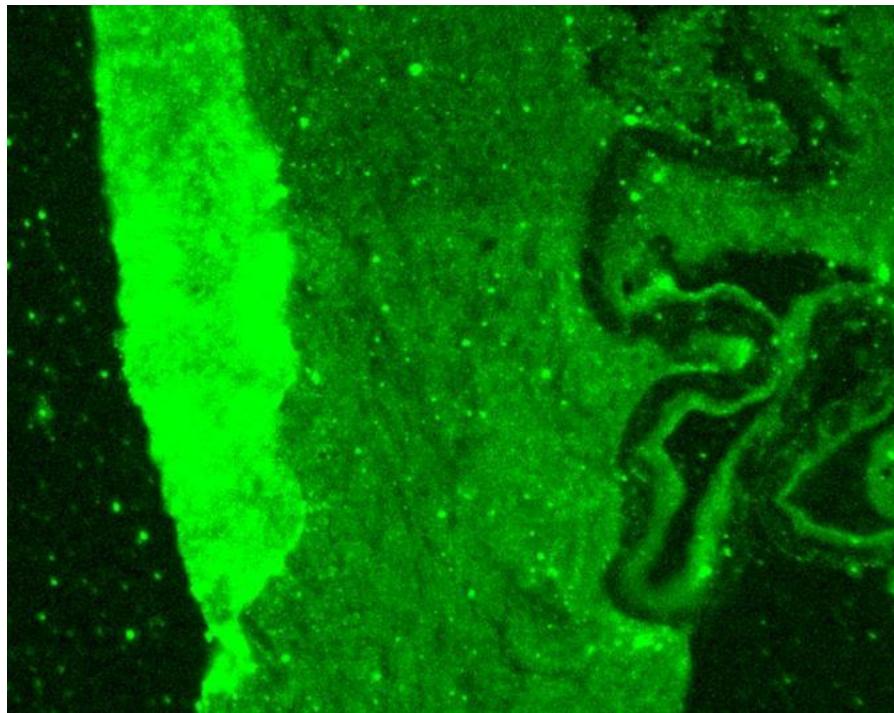
Difficulties of multiplexed bioimaging

Outliers, debris, ...



Difficulties of multiplexed bioimaging

Tissue folds, detachments,



Difficulties of multiplexed bioimaging

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

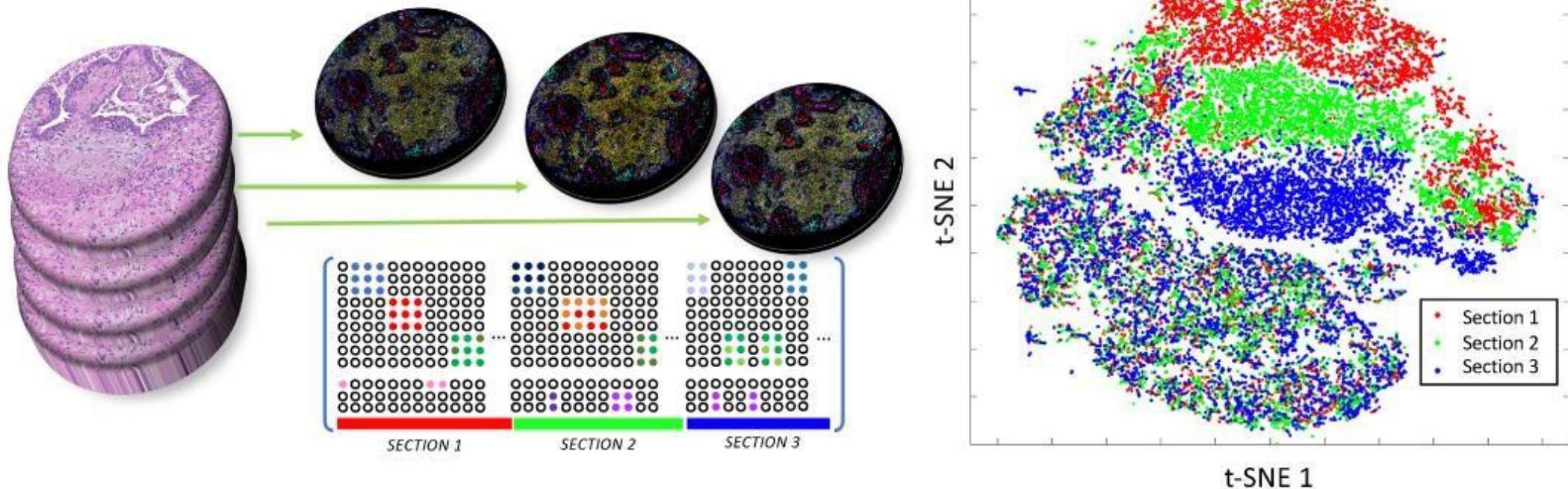
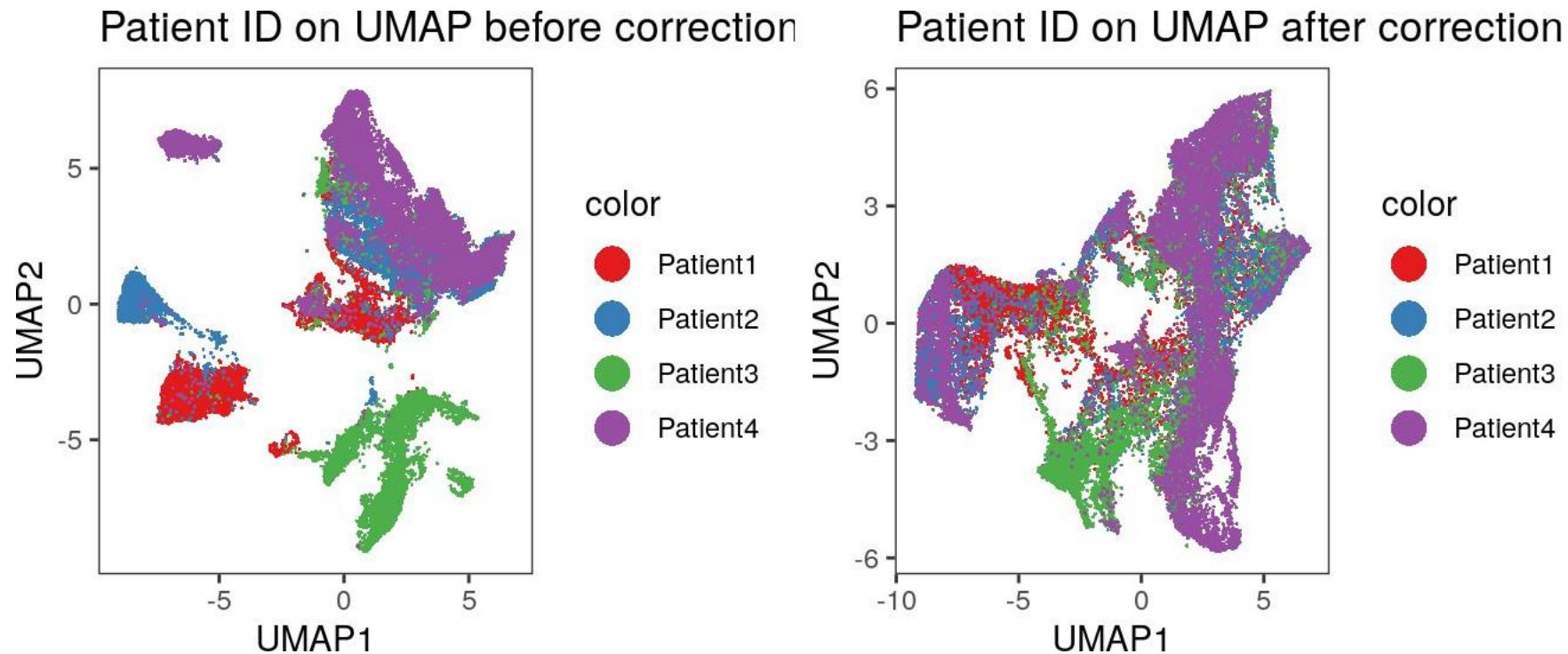


Fig. 1 An illustration of immunostaining intensity variation. Multiplexed immunostaining intensity varies across samples; intensity profiles from three adjacent sections ($5\ \mu\text{m}$) show intensity variations. t-SNE embedding of individual single-cell intensity features from three adjacent sections 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/>

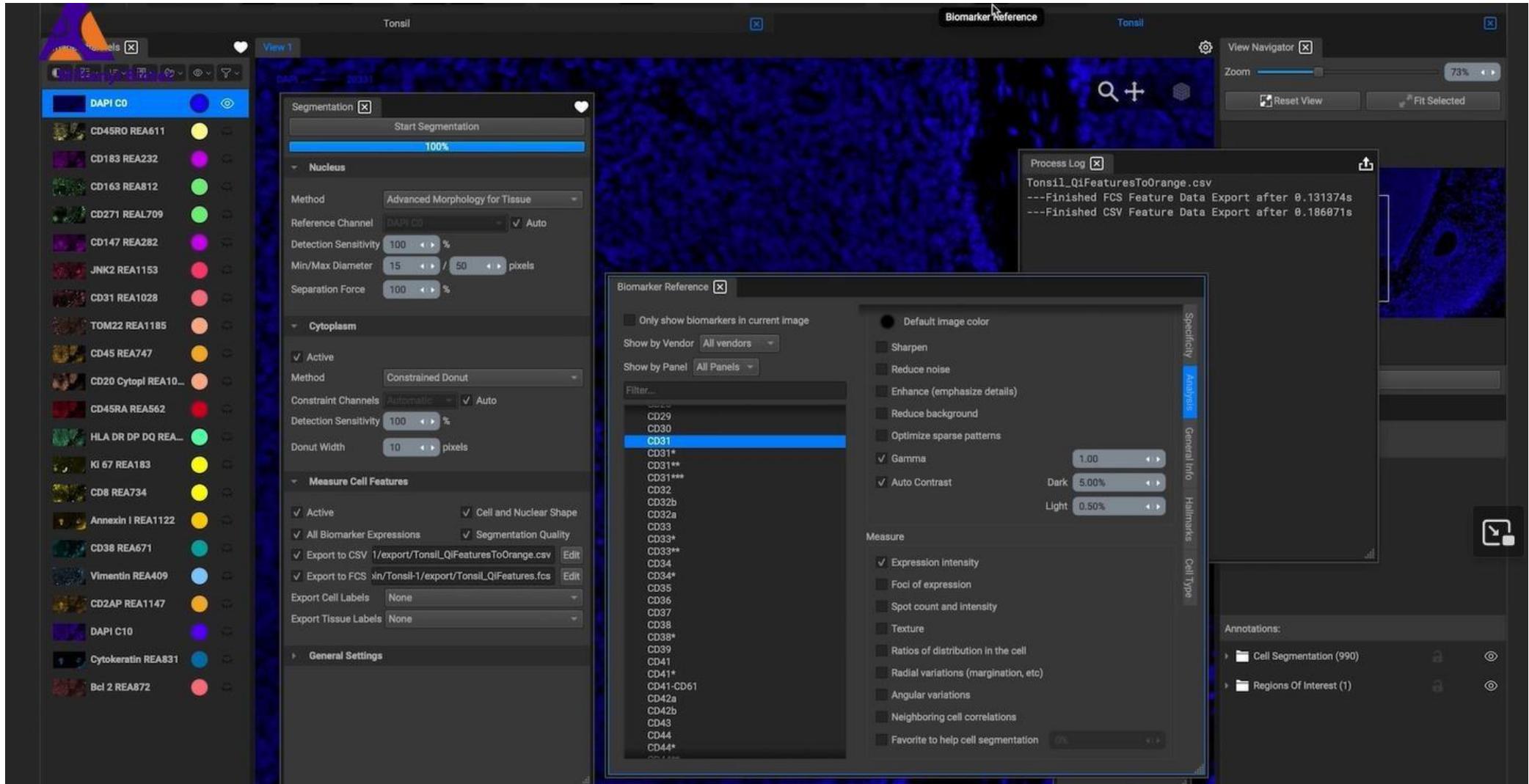


See also ark-analysis: Robust phenotyping of highly multiplexed tissue imaging data using pixel-level clustering

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

Overview of different spatial proteomics workflows

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

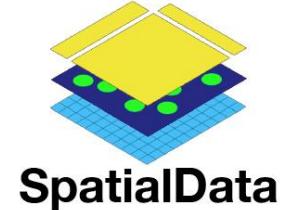
Expensive license needed

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...

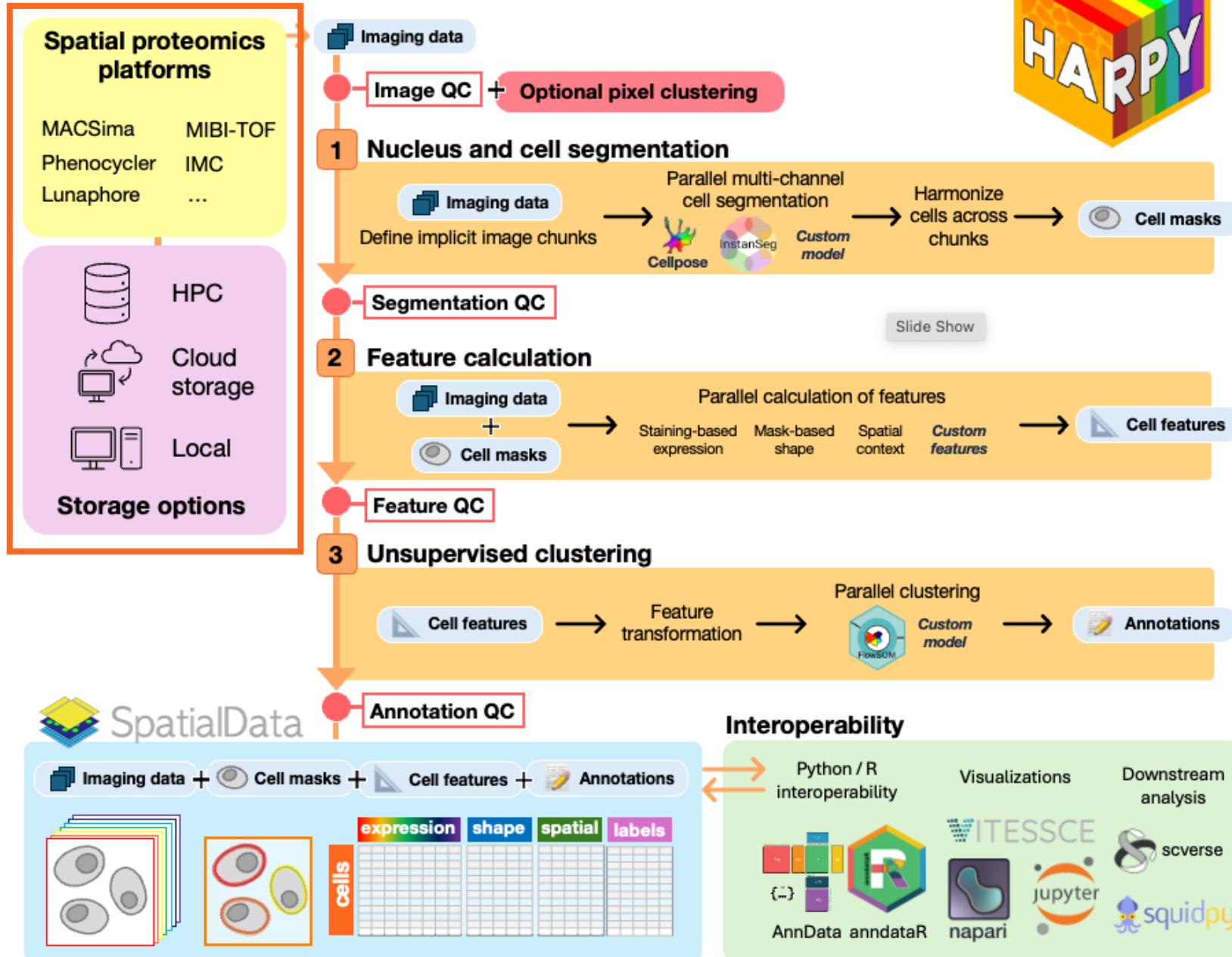
4 Harpy highlights

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

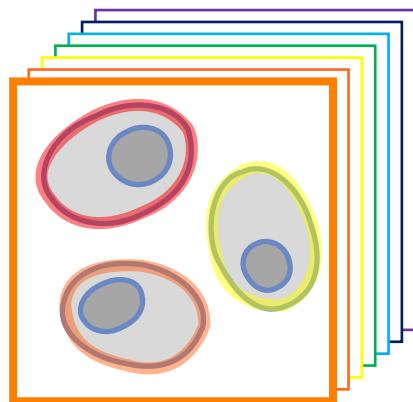


<https://github.com/saeyslab/harpy>

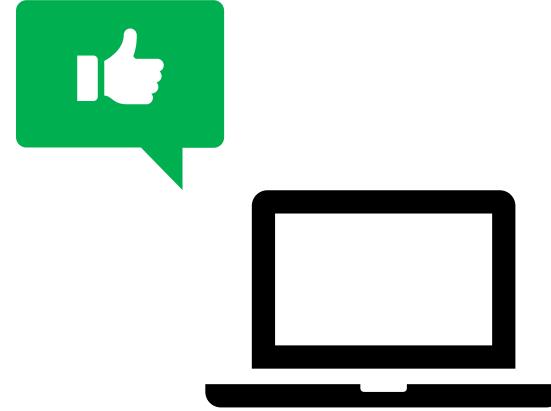
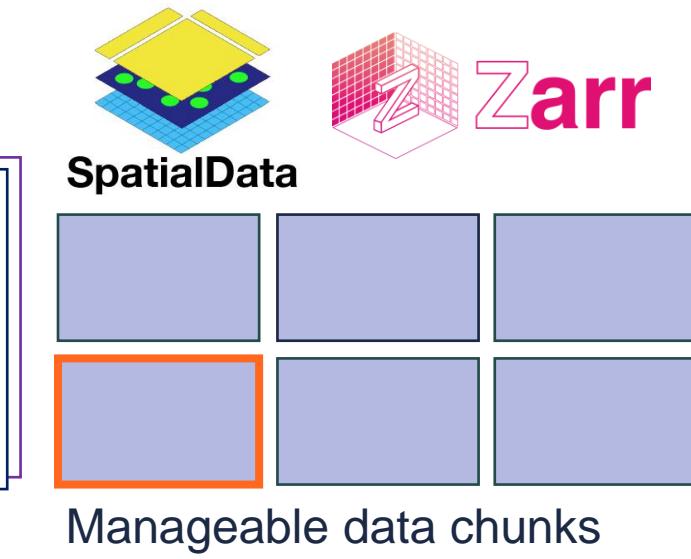
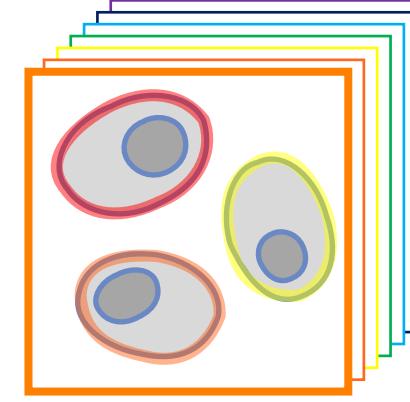
Harpy analysis workflow



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

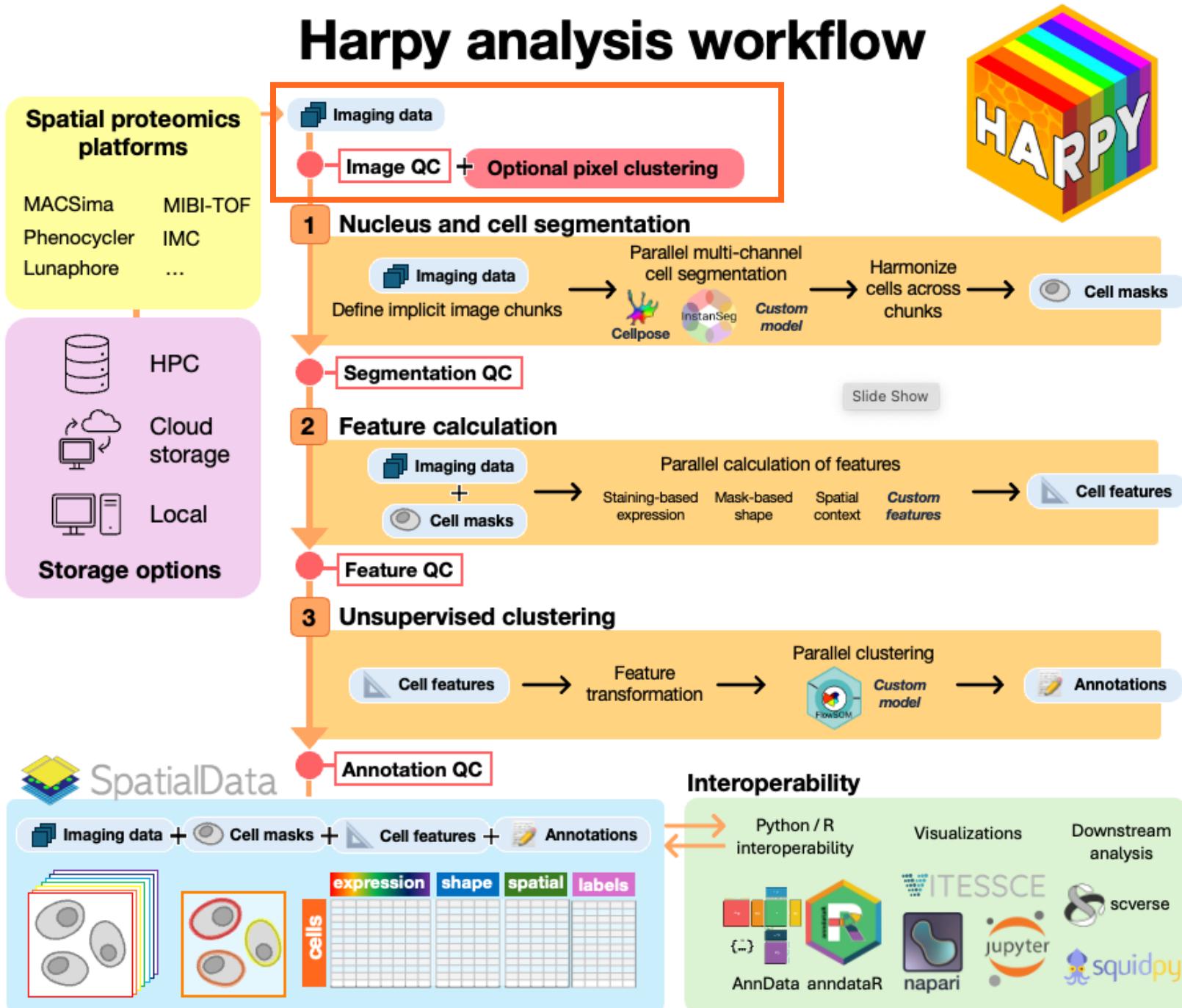


One sample: >10 Gigabytes!



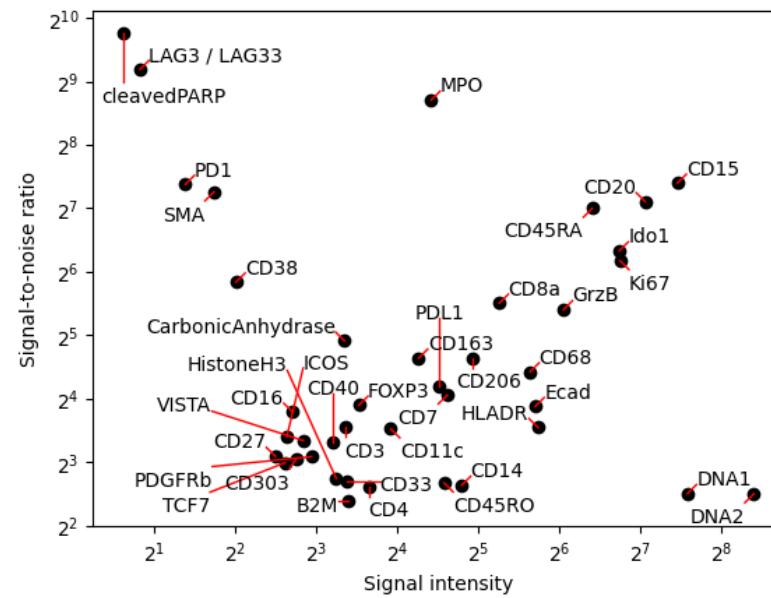
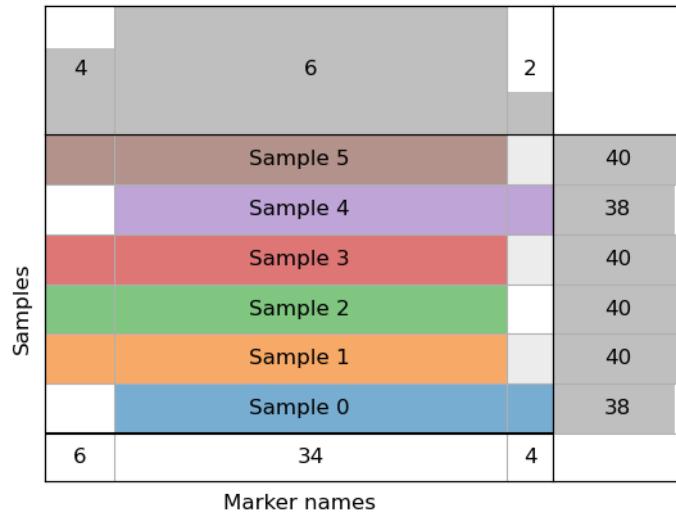
PR to add native support in SpatialData: <https://github.com/scverse/spatialdata/pull/842>

Harpy analysis workflow

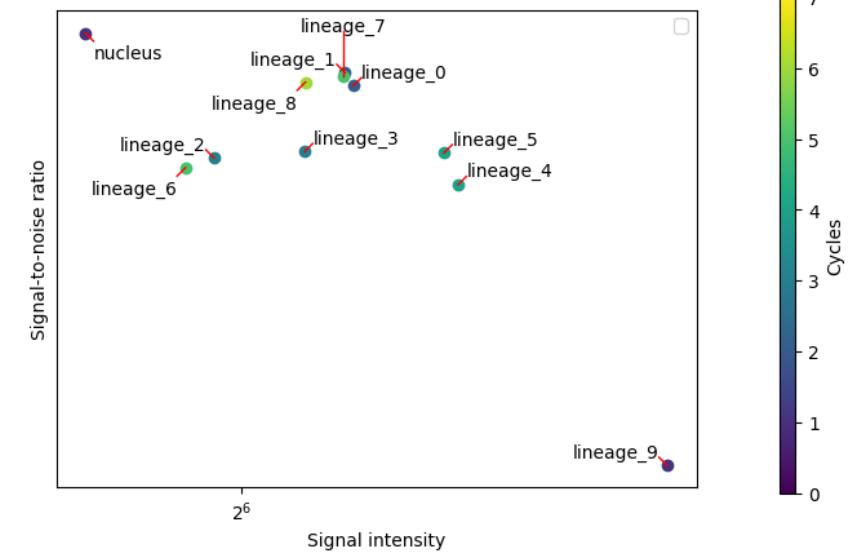


Quality control that scales to your whole dataset

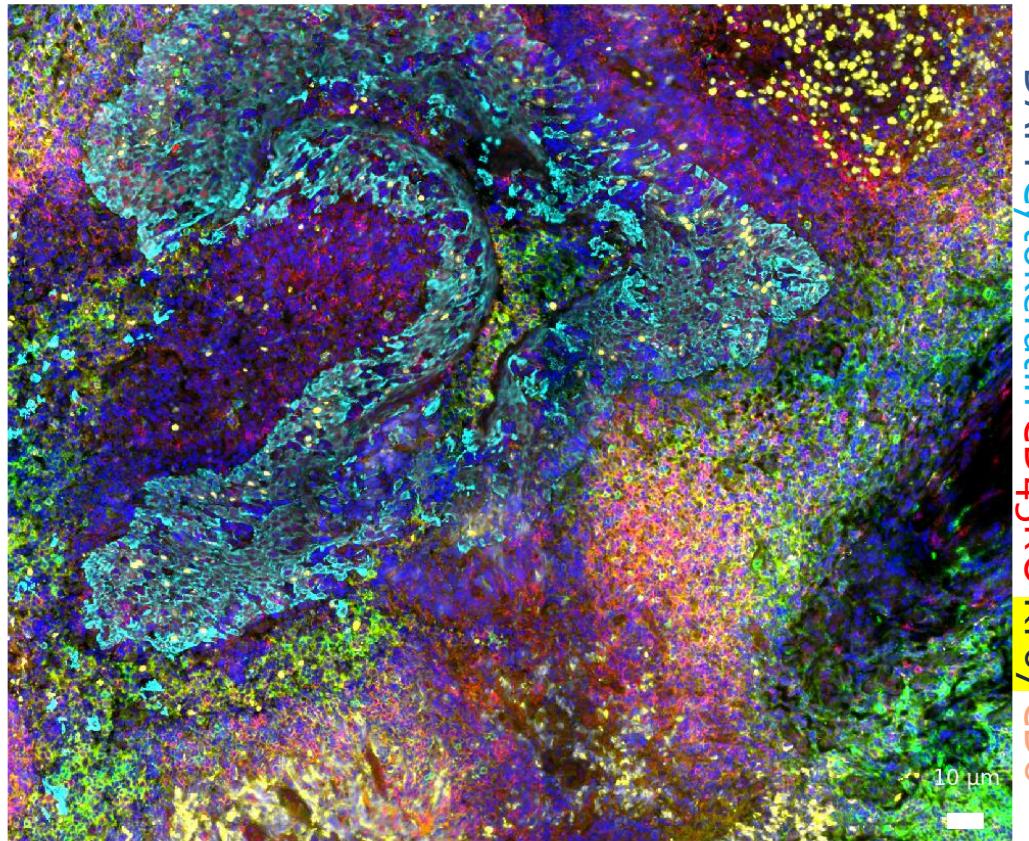
Similar to IMCDataAnalysis, but in **Python**



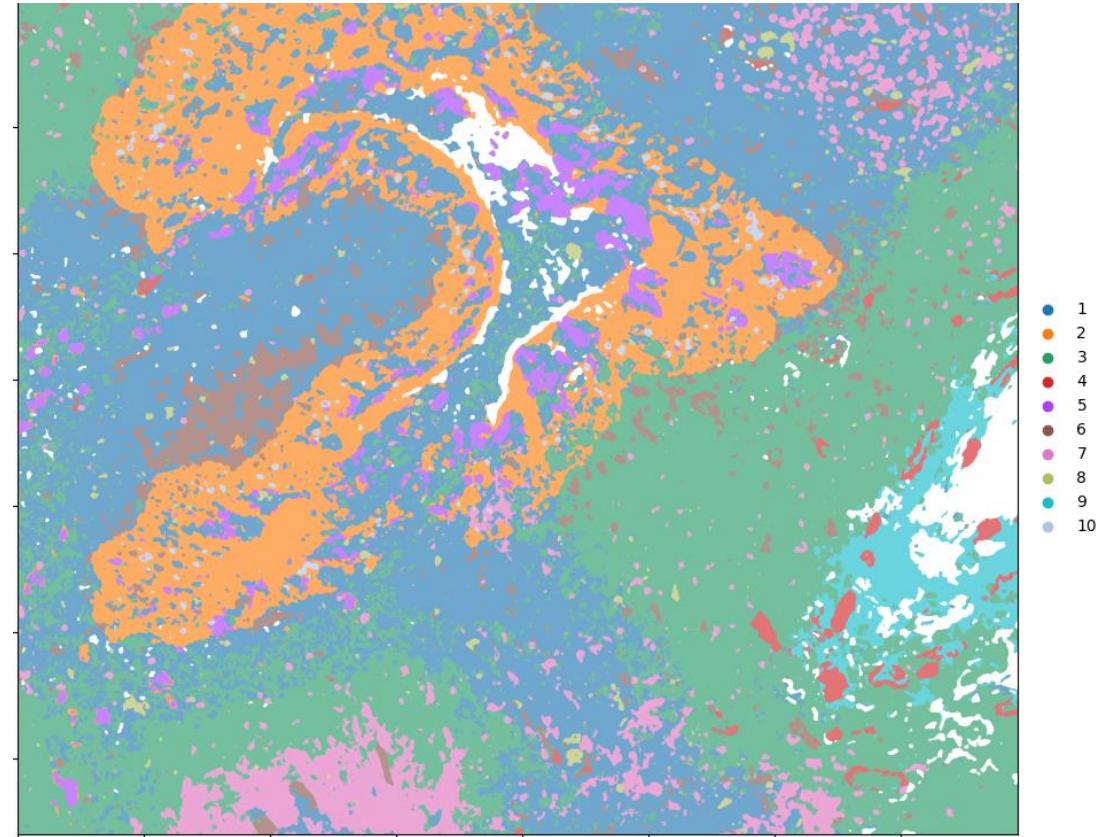
New plots with support for metadata e.g. **cycle order**



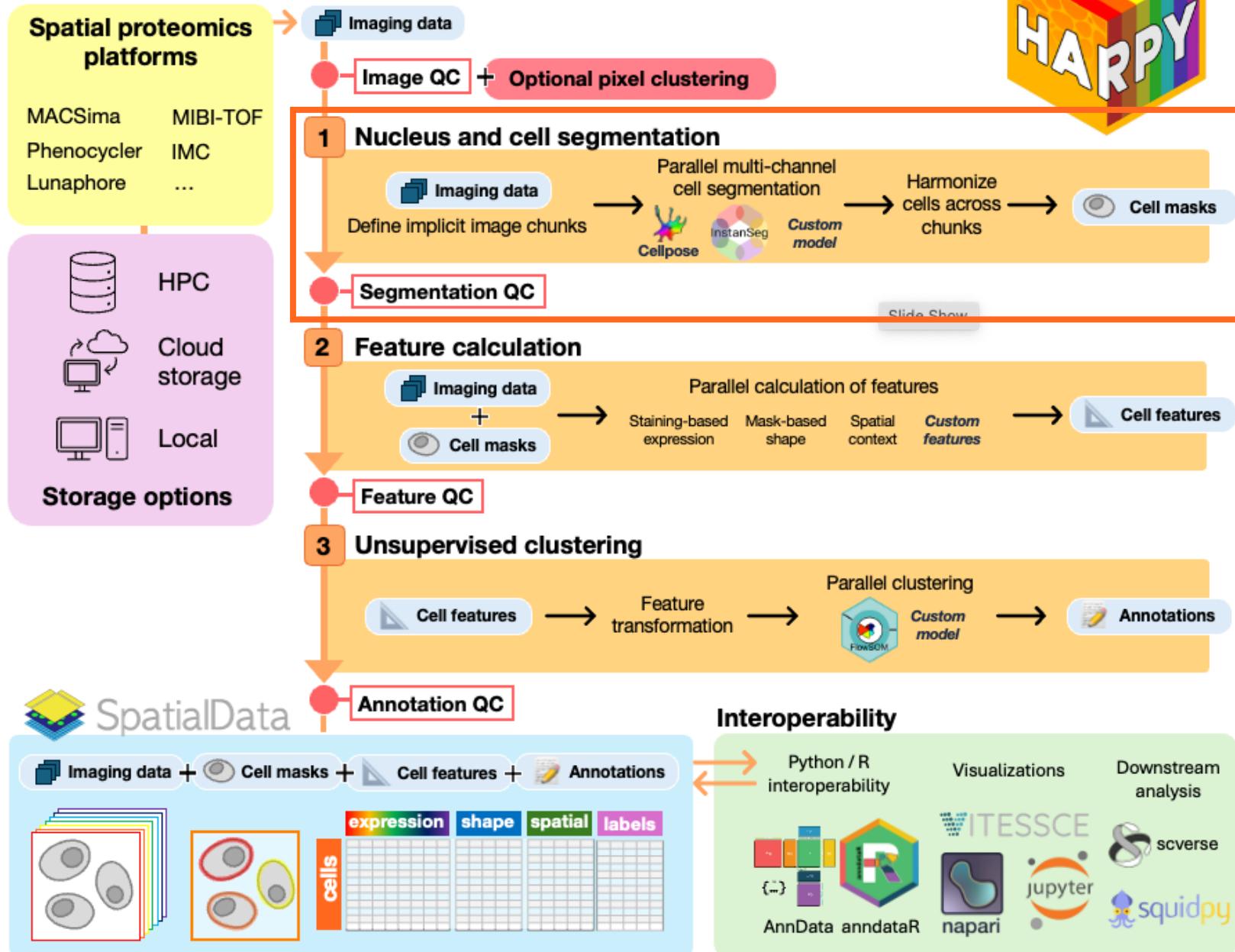
Pixel clustering allows tissue overview without marker and range selection



Pixel clustering with FlowSOM



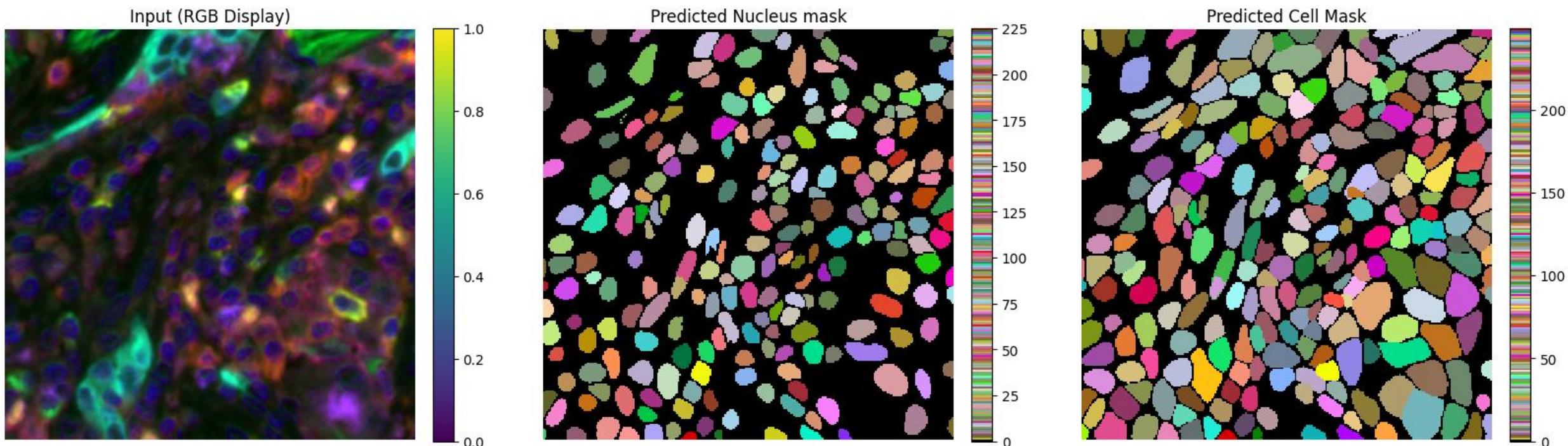
Harpy analysis workflow



Harpy can apply the latest cell segmentation models



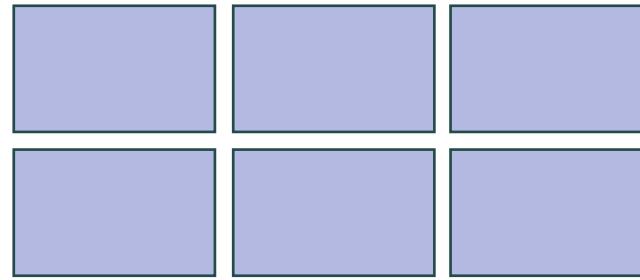
Works on **any combination** of image channels in **any order**
Outputs both predicted **nucleus** and **cell** mask



Harpy is fast because of Dask parallelisation



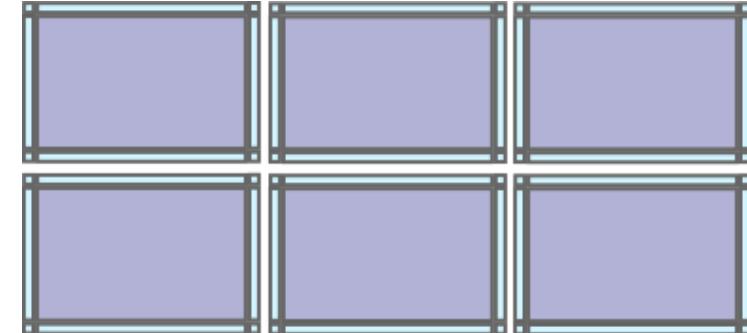
Whole-slide image



Model apply
on WSI



Out-of-memory
Very slow



Model apply on
chunks



Fast
Border artefacts

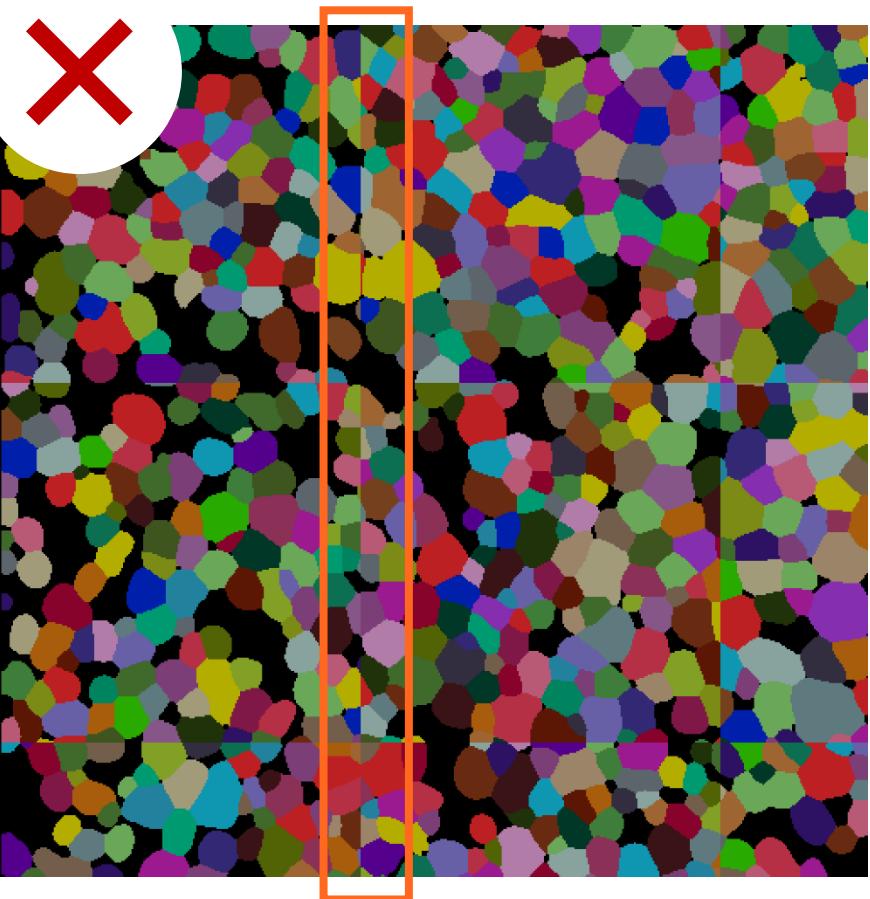
Model apply on
chunks via
Dask `map_overlap`



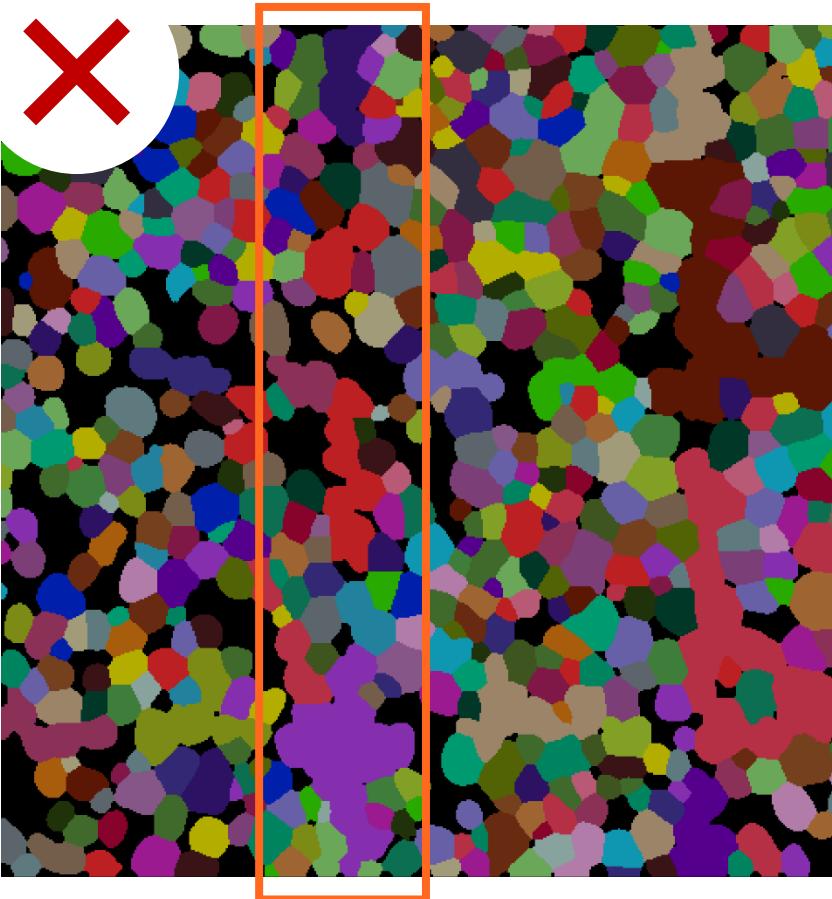
Fast, no memory issues
No artefacts
No manual chunks

Good label merging is essential

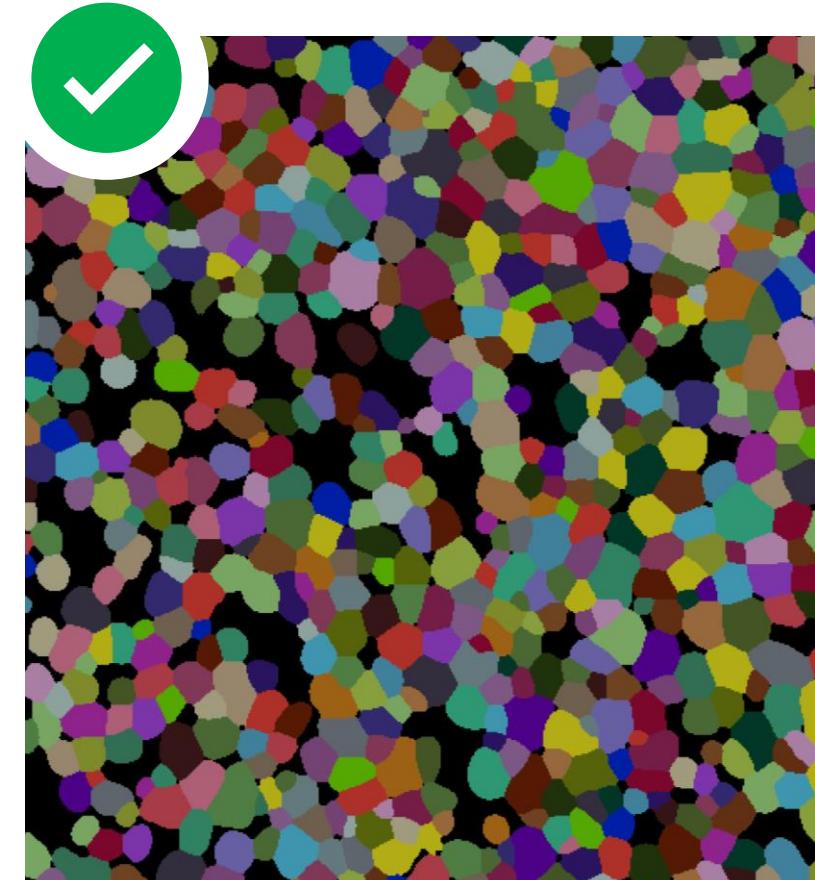
Naive dask.map_overlap



Naive connected_components or squidpy.segment

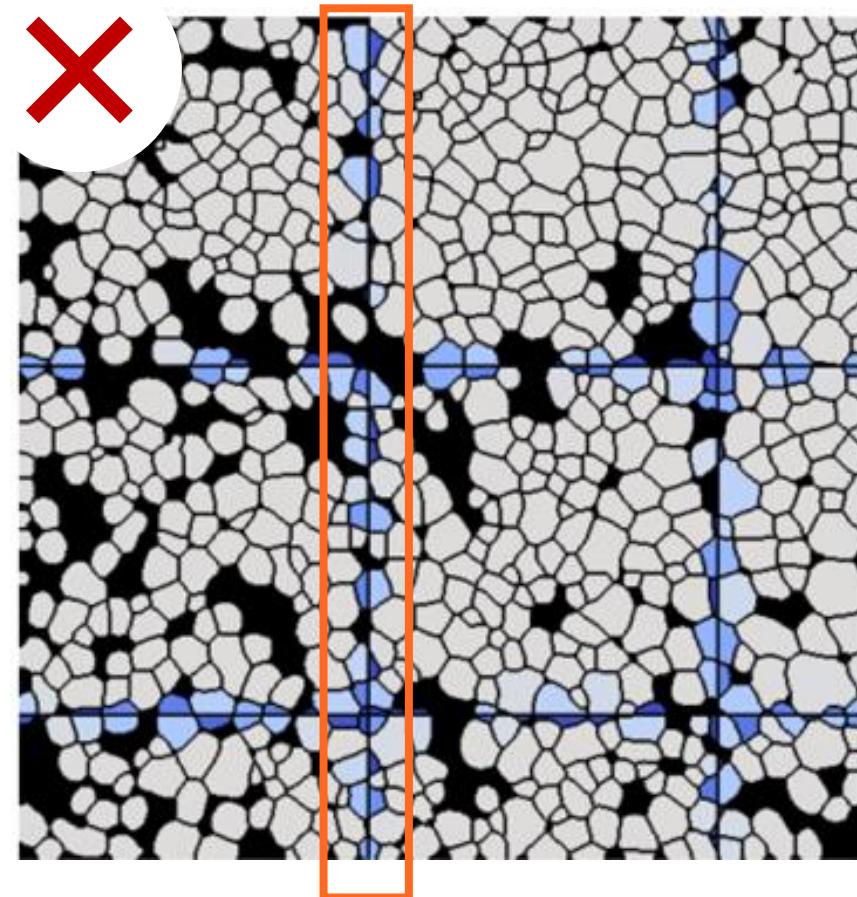


harpy.im.segment



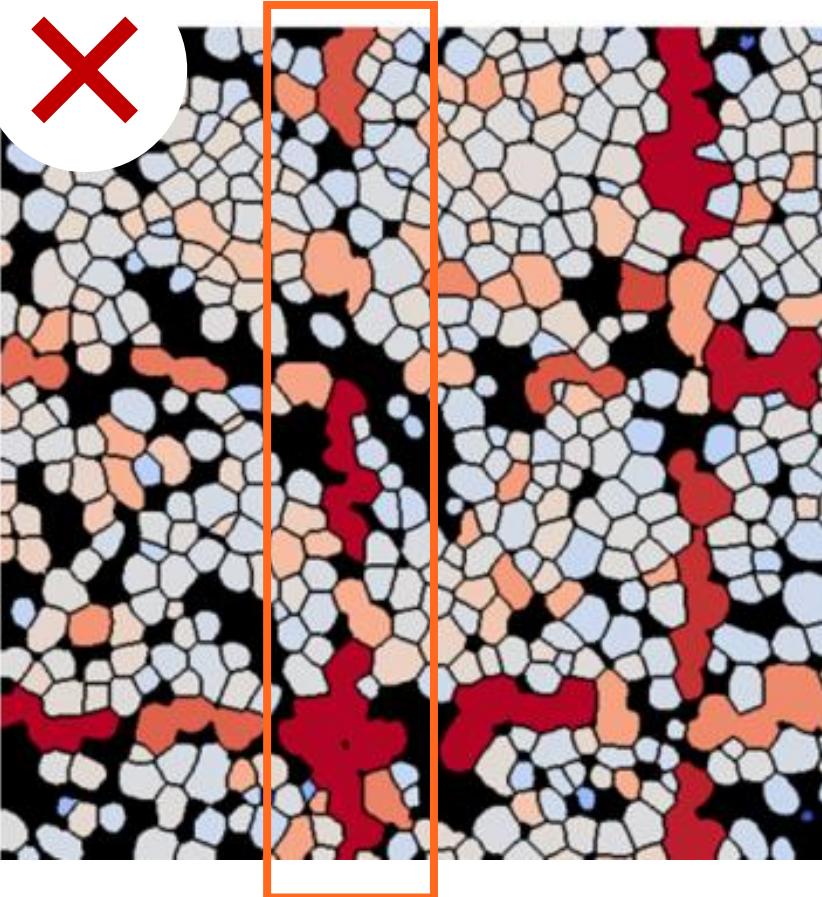
Good label merging is essential

Naive dask.map_overlap



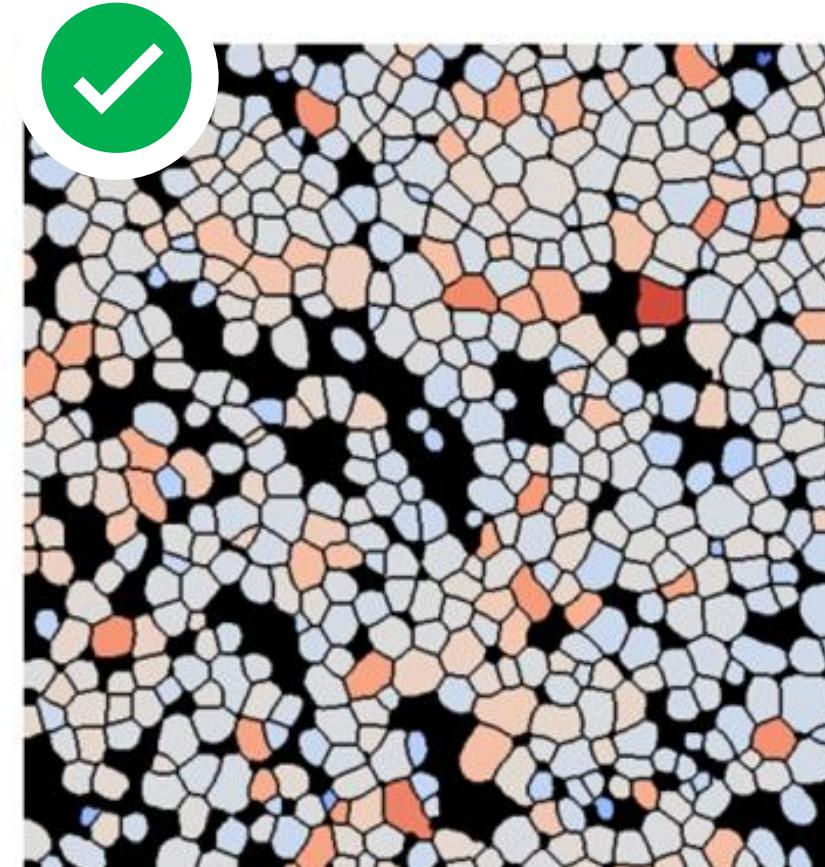
No label merges at tile borders

Naive connected_components or squidpy.segment



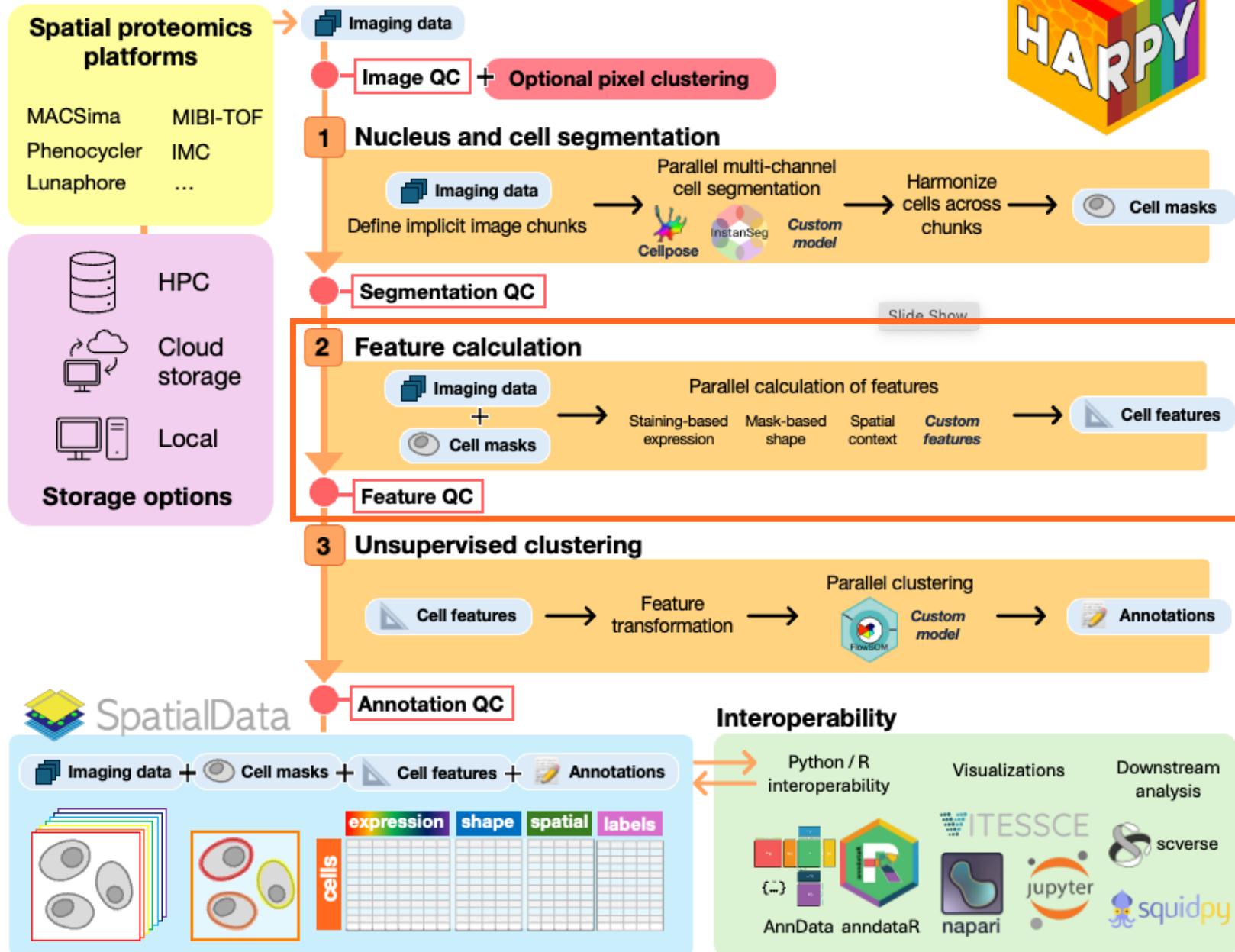
Too many label merges at tile borders

harpy.im.segment

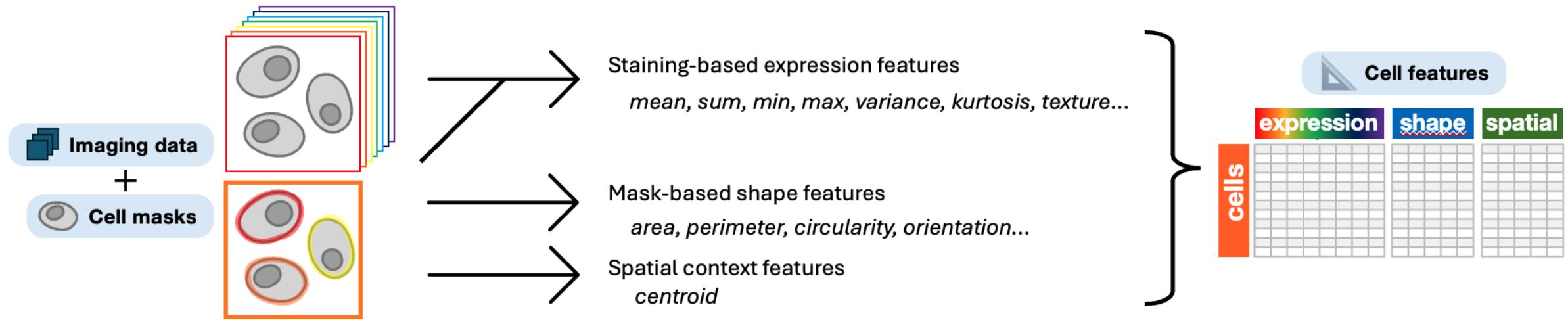


Correct label merges

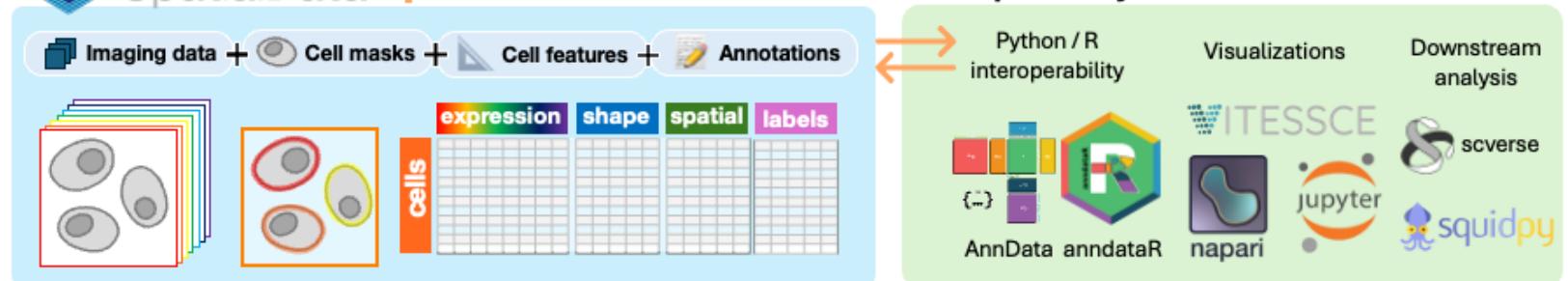
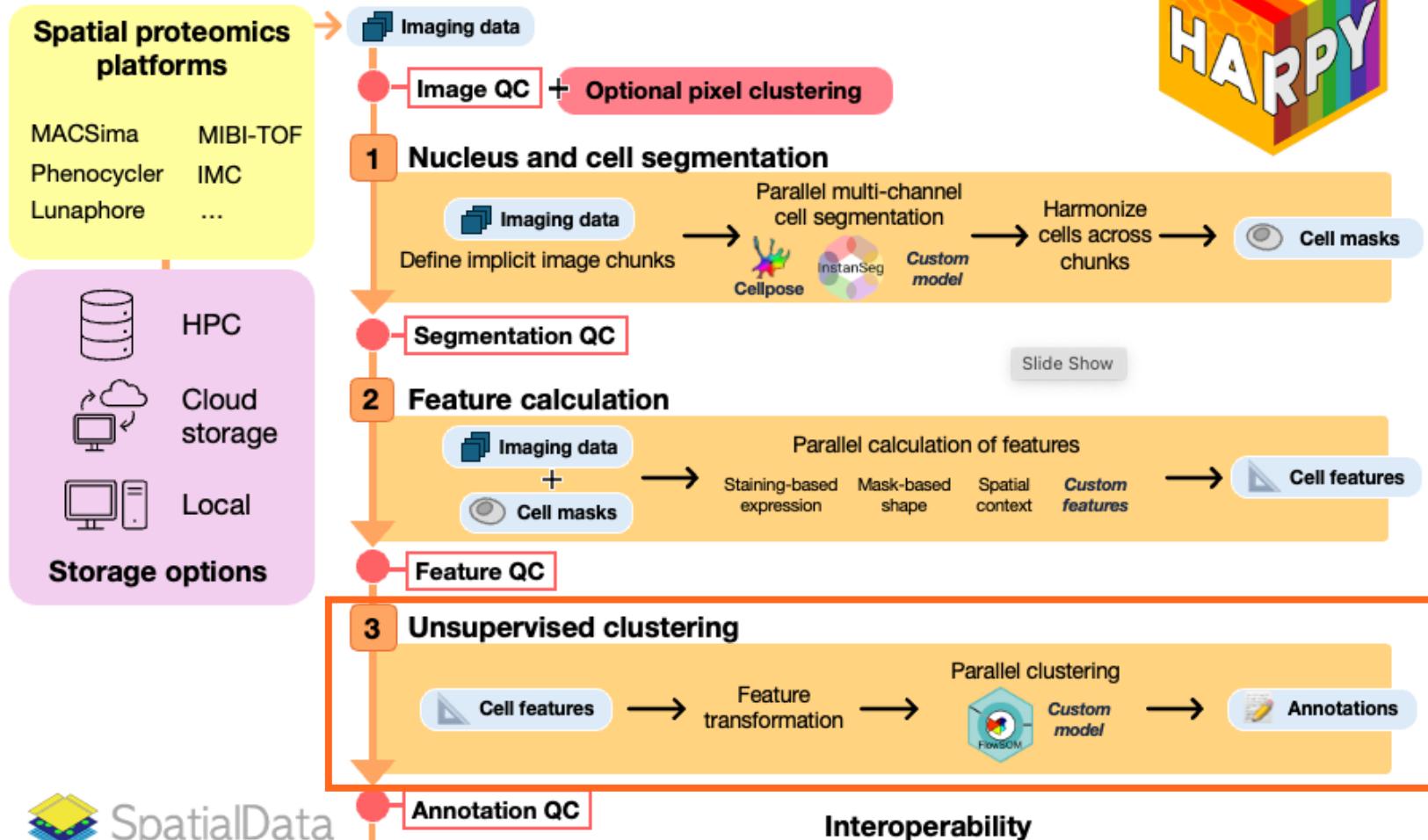
Harpy analysis workflow



Different features from images and masks

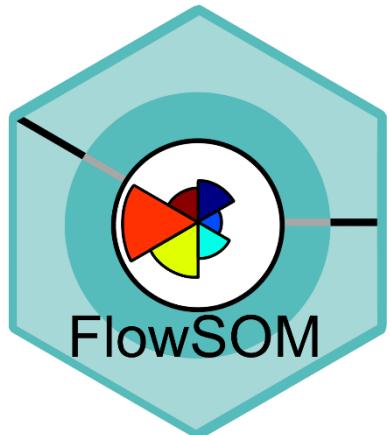


Harpy analysis workflow

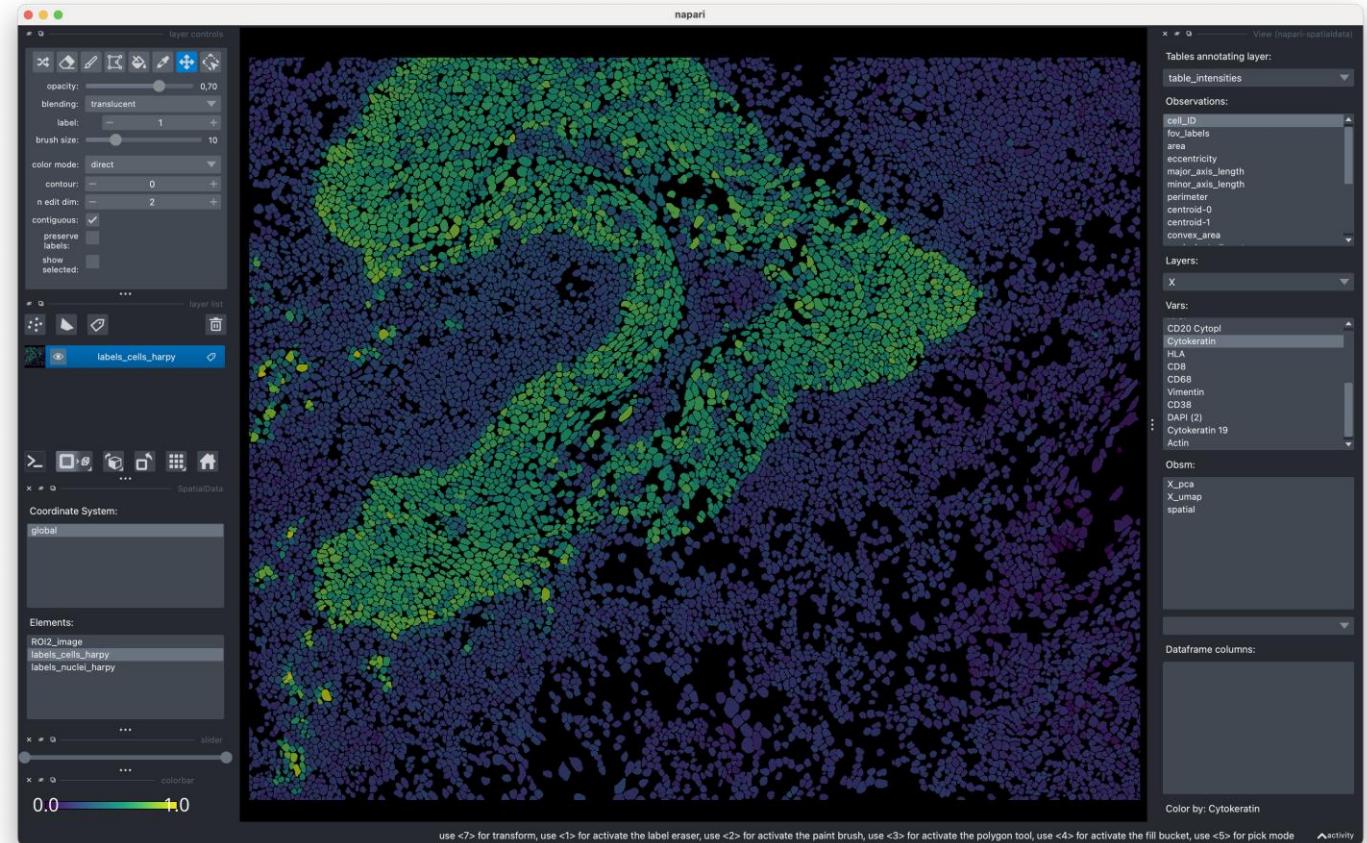


Interactive unsupervised clustering

Cell and **pixel** clustering workflow
using FlowSOM in Python



napari

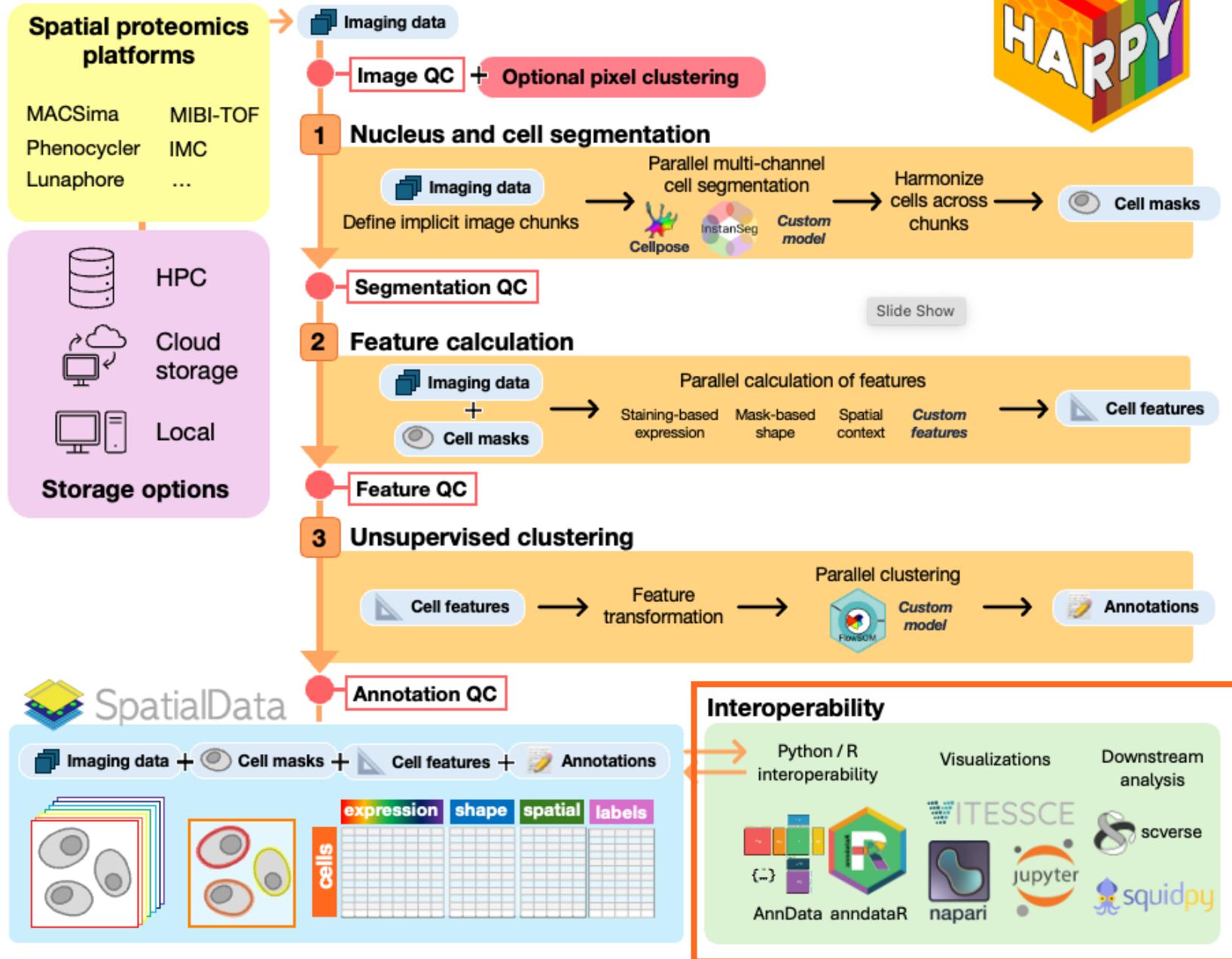


Visualize results in notebook, with napari or
with other SpatialData plugins like e.g. napari-spatialdata

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>

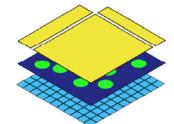
Harpy analysis workflow



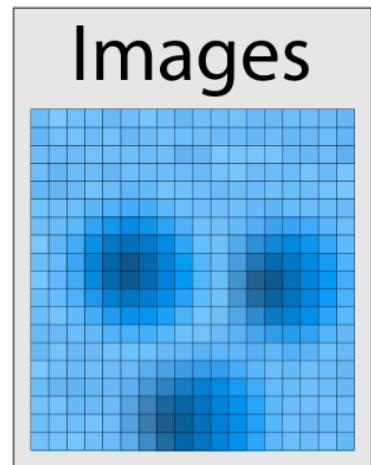
Harpy output is interoperable with R



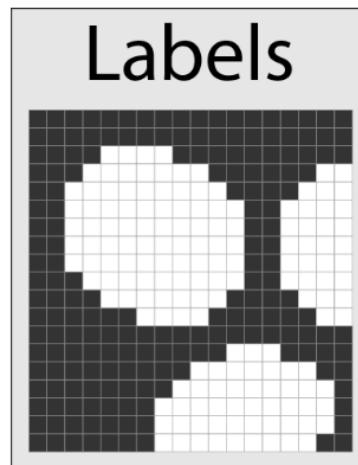
Louise Deconinck



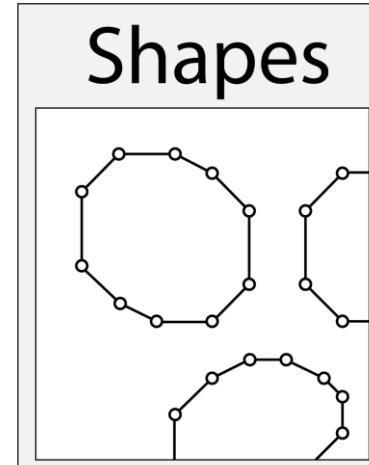
SpatialData



Microscopy
images ...

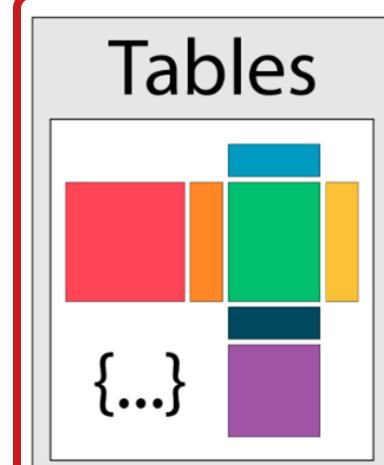


Segmentation
mask

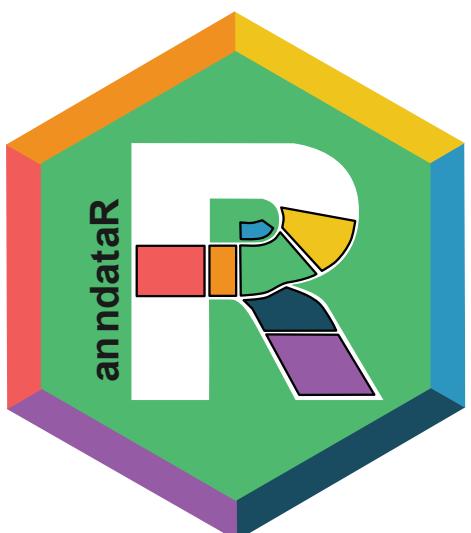


Cells, ROIs
...

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>

Acknowledgments

VIB-UGent Center for Inflammation Research

SaeysLab

Benjamin Rombaut

Lotte Pollaris

Ruth Seurinck

Sofie Van Gassen

...

Yvan Saeys

VIB Spatial Catalyst

Arne Defauw

Frank Vernaillen

Julien Mortier

Evelien Van Hamme



Flanders AI
Research Program