Cell segmentation free analysis of spatially resolved transcriptomics data

Naveed Ishaque
Berlin Institute of Health at Charité University Hospital, Germany
ELIXIR-Germany
ELIXIR SPODA 2025

BIH Berlin Institute of Health @Charité

Resolution revolution – transcriptomics

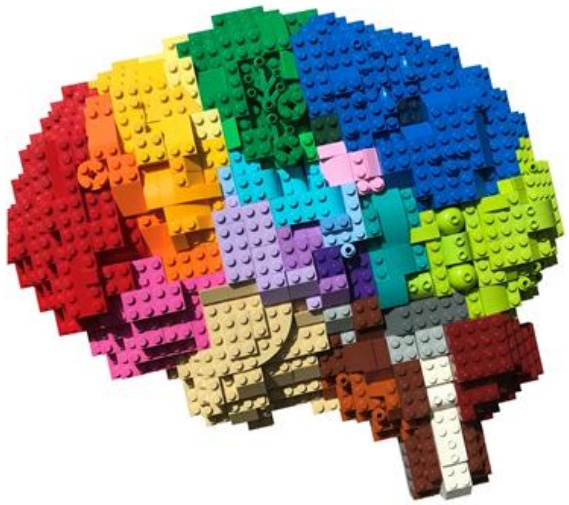
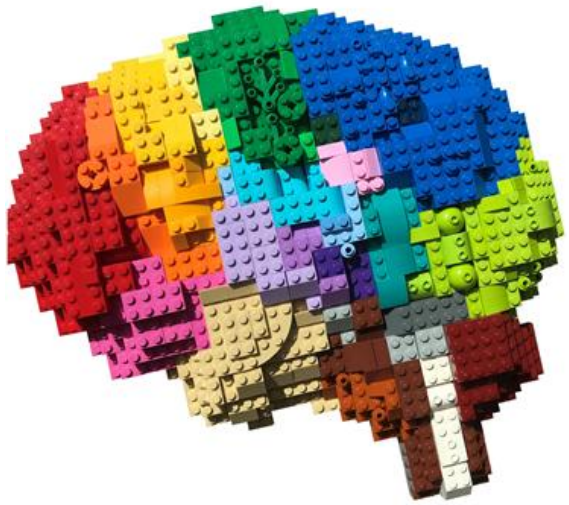


Image credit: Bo Xia (@BoXia7)

Resolution revolution – transcriptomics: bulk



Bulk

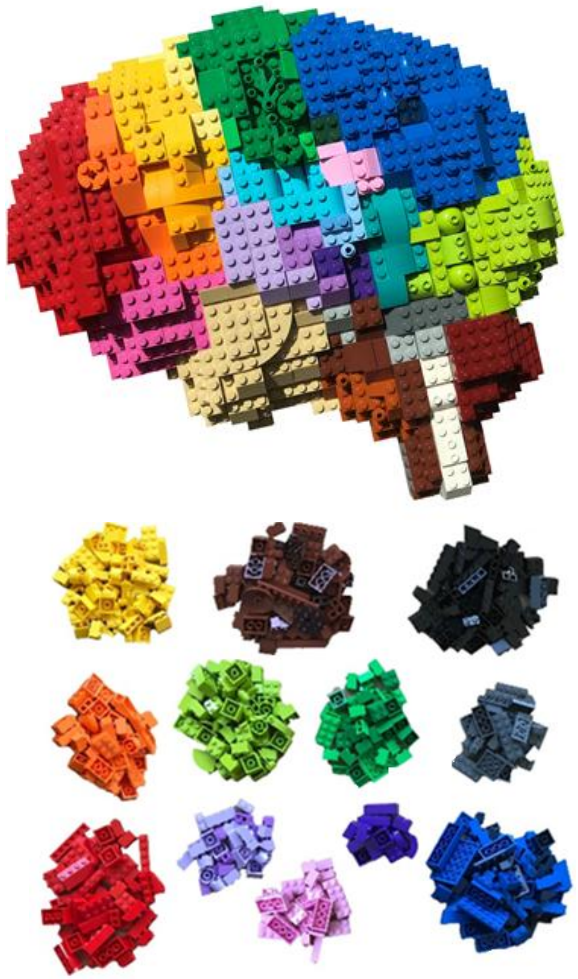
Image credit: Bo Xia (@BoXia7)

elixir

IBIH Berlin Institute of Health @Charité

Resolution revolution – transcriptomics: single cell

Bulk



Single-cell

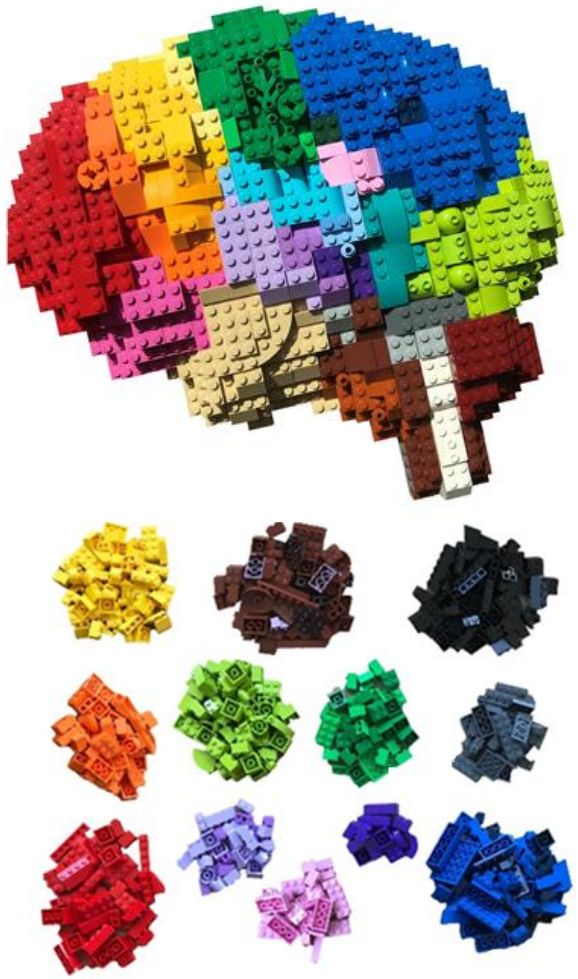
Image credit: Bo Xia (@BoXia7)

elixir

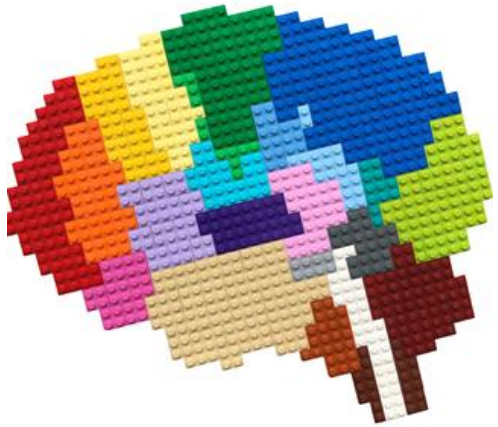
IBIH Berlin Institute of Health @Charité

Resolution revolution – transcriptomics: spatial

Bulk



Single-cell



Spatial

Image credit: Bo Xia (@BoXia7)

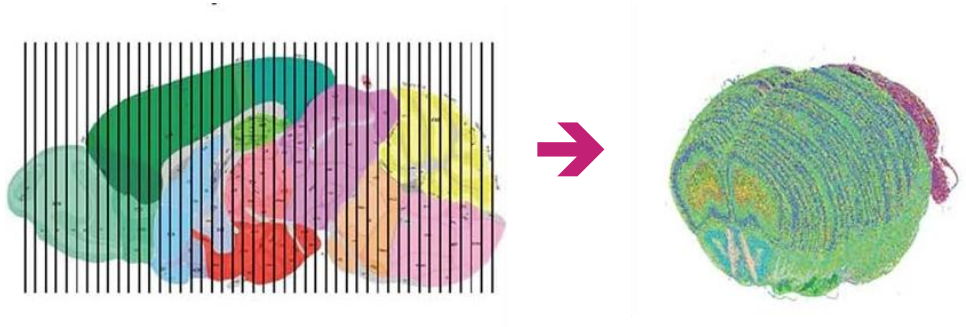
elixir

BIH Berlin Institute of Health @Charité

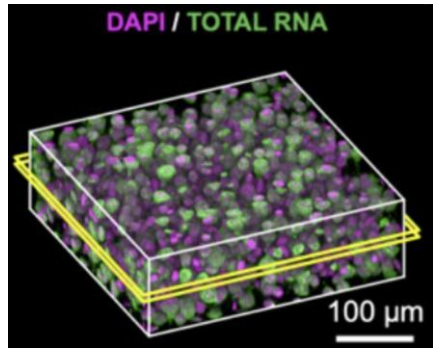
Emerging state of the art for spatial transcriptomics

Size and resolution

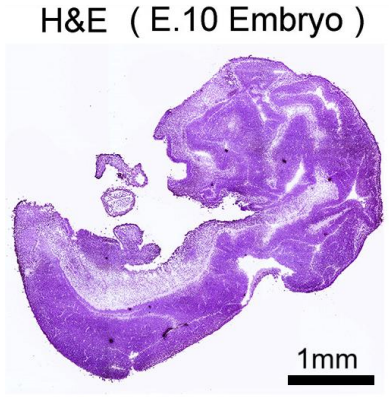
Pseudo 3D

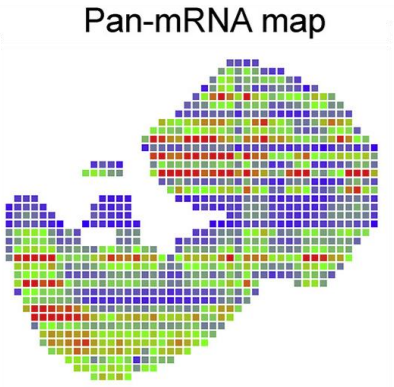


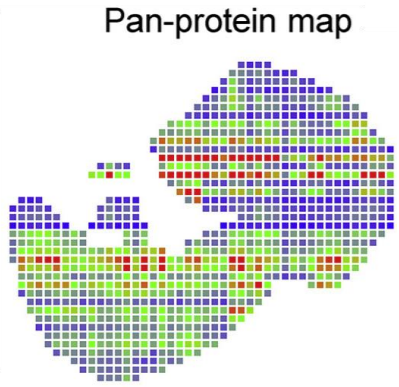
Real 3D



Multi-omics

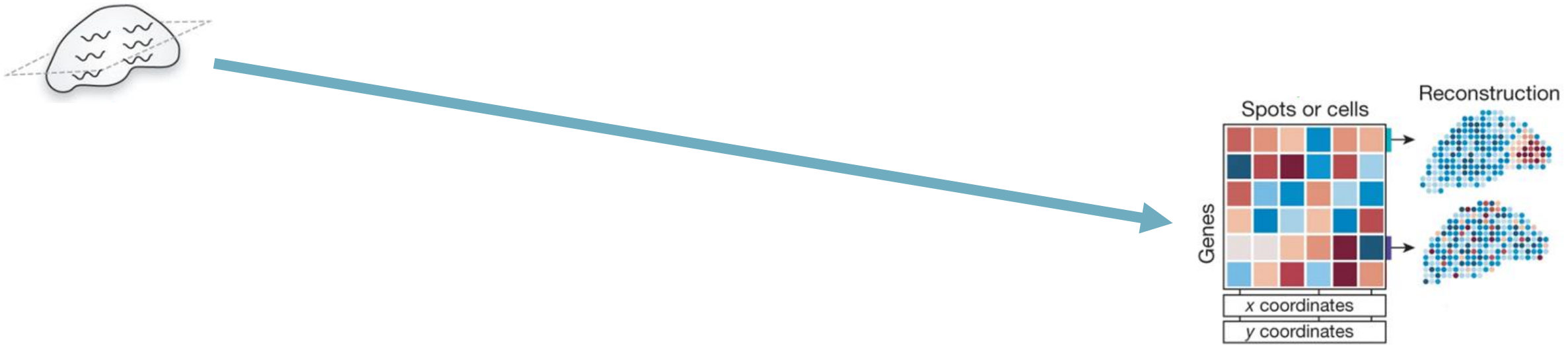






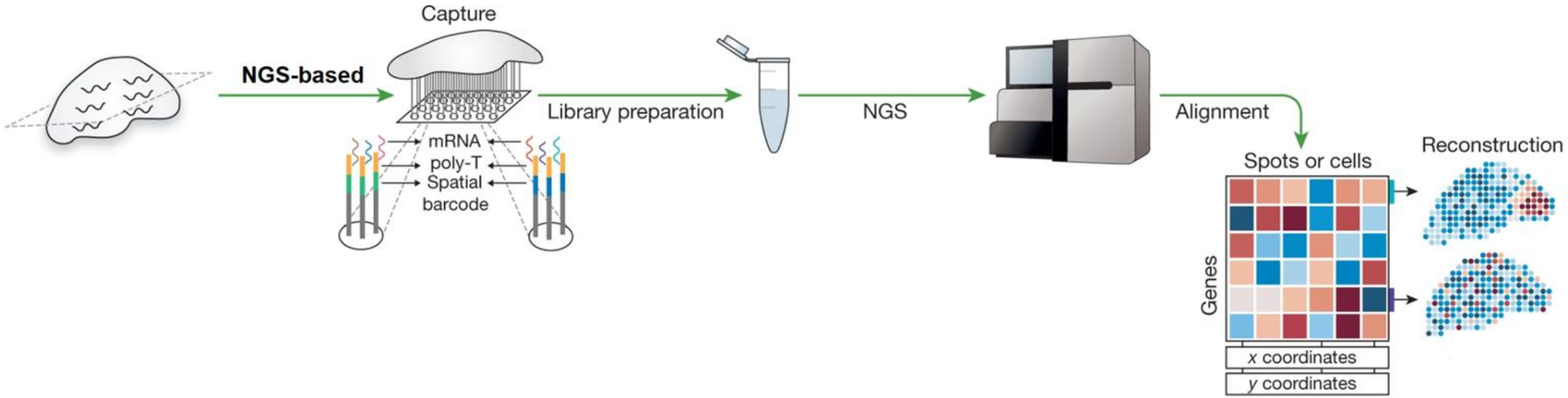
Liu, Yang, Deng et al. Cell 183, 1665-1681 (2020); Chen et al, Cell 185, 1777-1792 (2022); Fang et al, eLife12:RP90029 (2023);
Yao, van Velthoven, Kunst et al. Nature 624, 317–332 (2023); Mueller-Boetticher et al. bioRxiv (2024)

Spatial transcriptomics methodologies



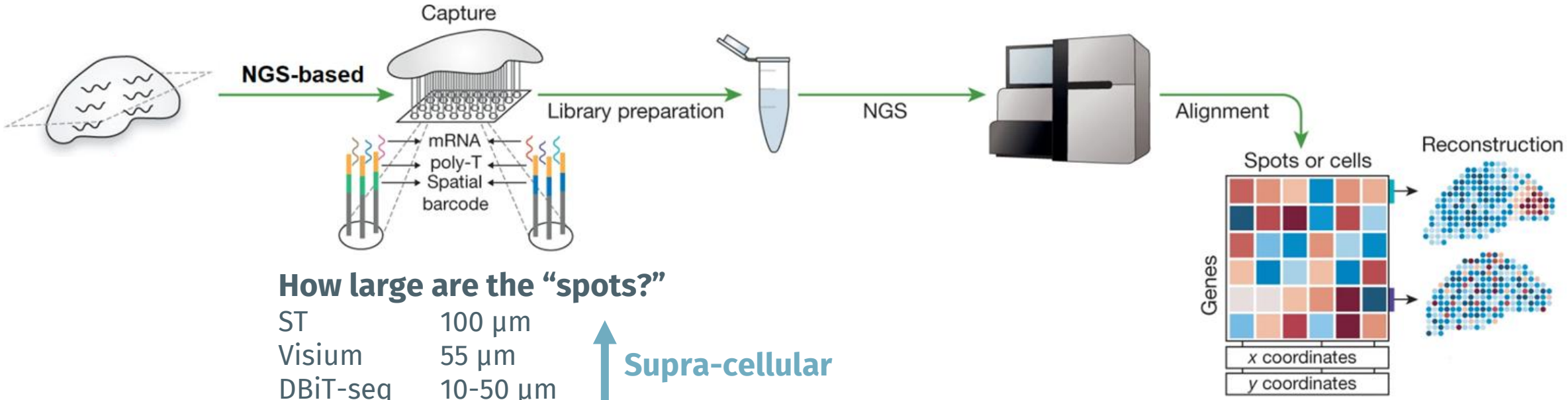
Adapted from Rao et al (2021) Nature

Spatial transcriptomics methodologies: NGS-based



Adapted from Rao et al (2021) Nature

Spatial transcriptomics methodologies: NGS-based



How large are the "spots?"

ST	100 µm
Visium	55 µm
DBiT-seq	10-50 µm
Slide-Seq	10 µm
HDST	2 µm
PIXEL-seq	1.22 µm
Seq-Scope	0.5-0.8µm
Stereo-seq	0.22 µm

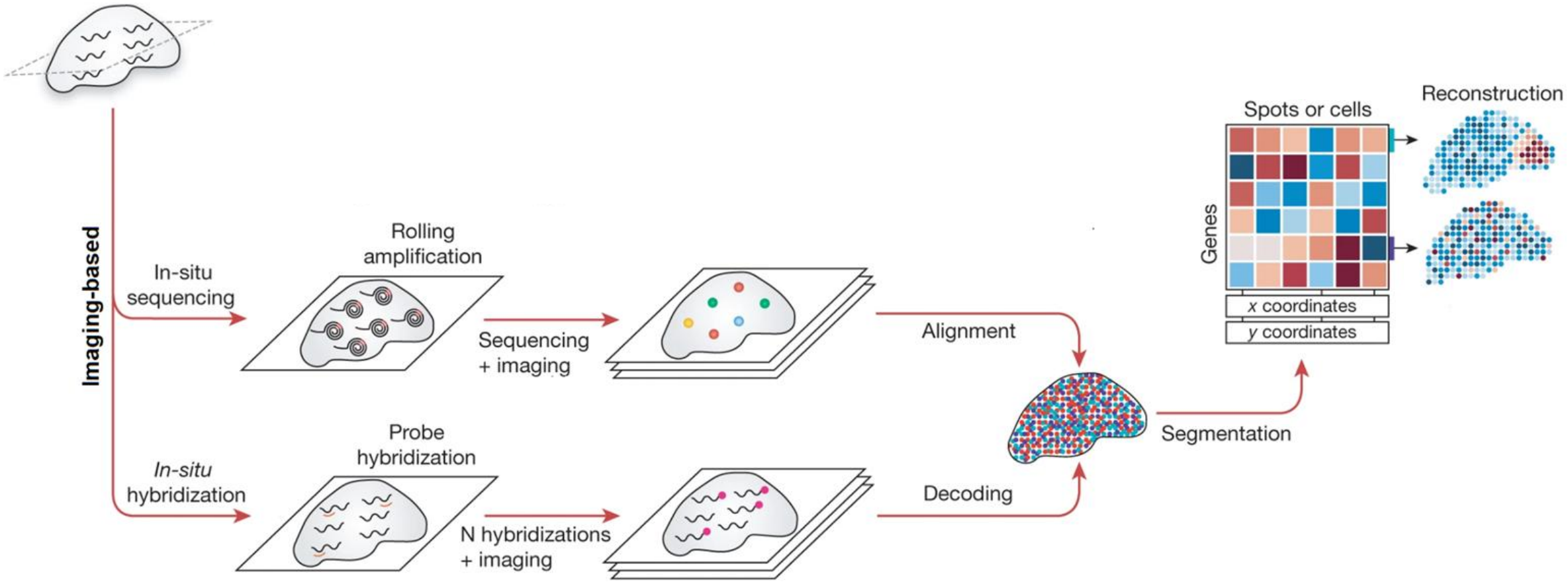
Supra-cellular

Cellular

Sub-cellular

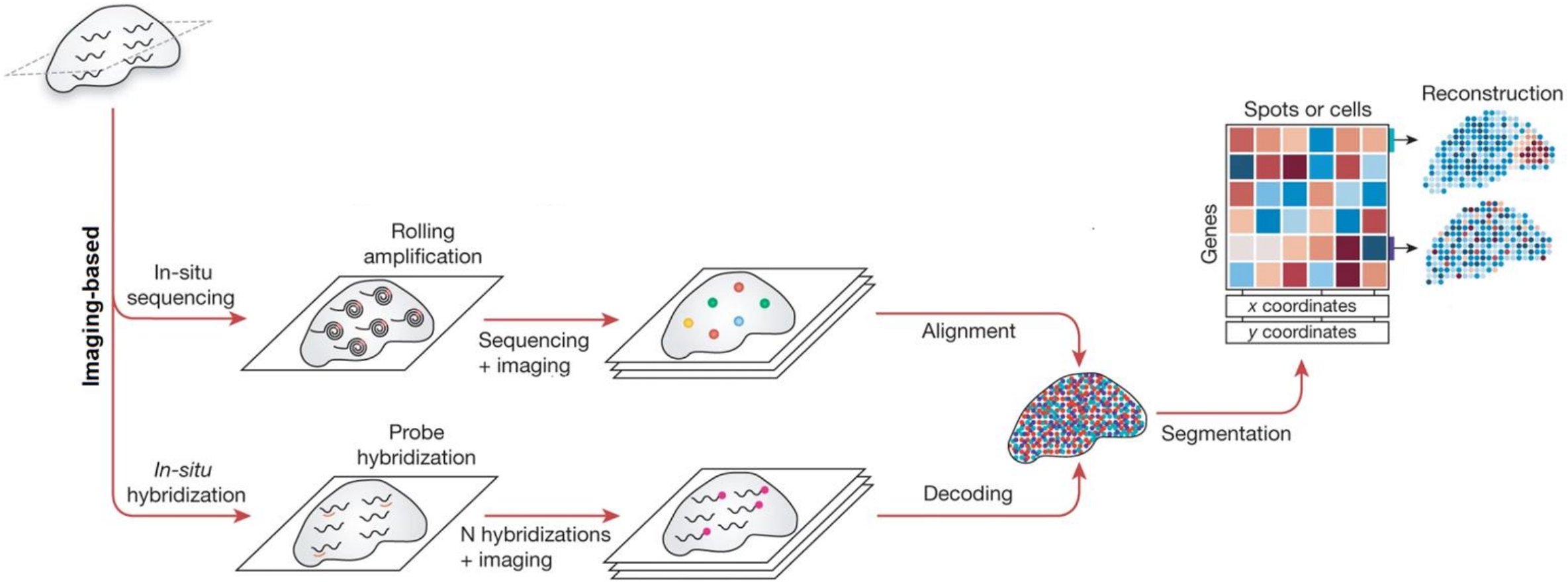
Adapted from Rao et al (2021) Nature

Spatial transcriptomics methodologies: imaging-based



Adapted from Rao et al (2021) Nature

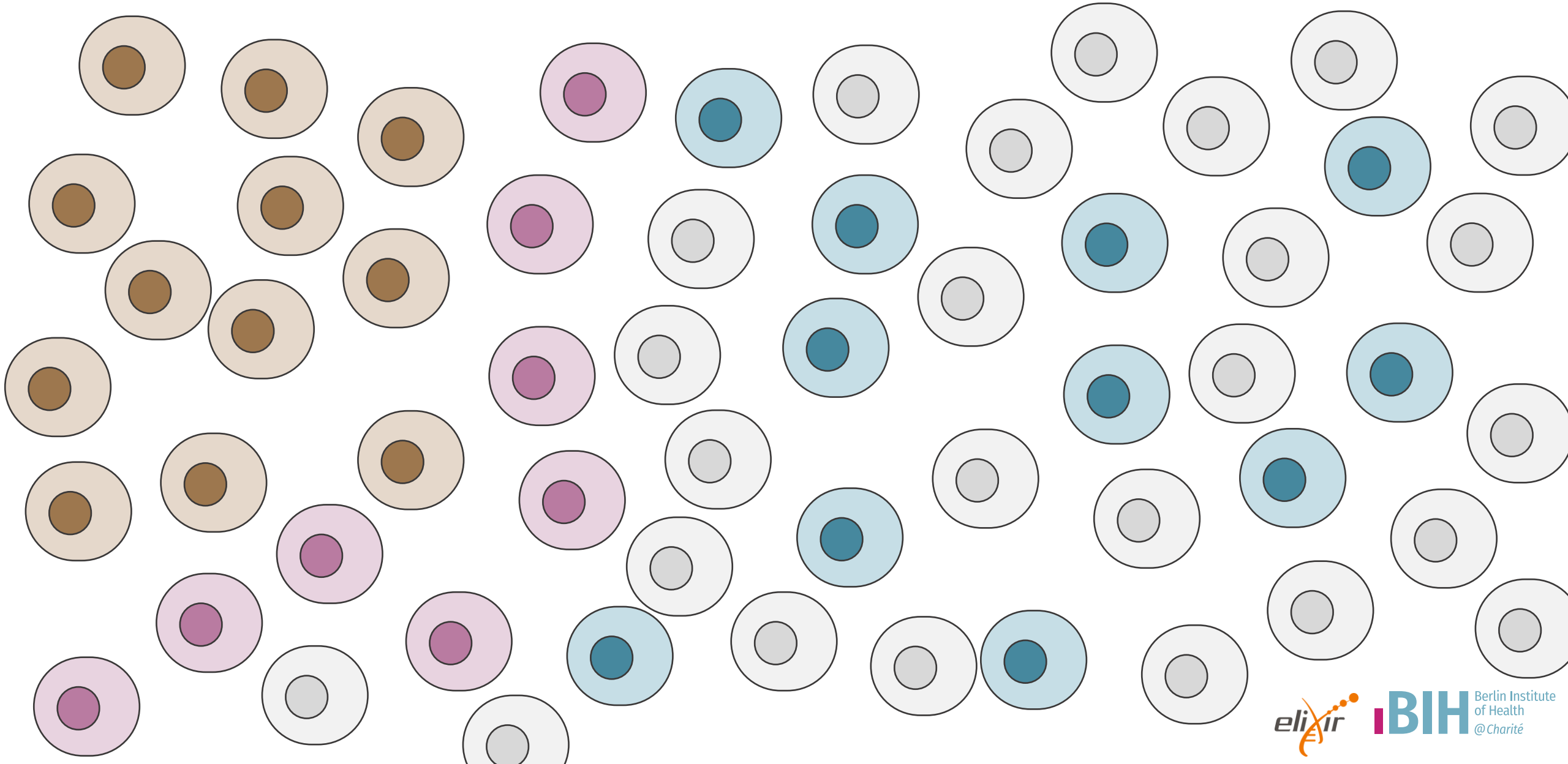
Spatial transcriptomics methodologies: imaging-based



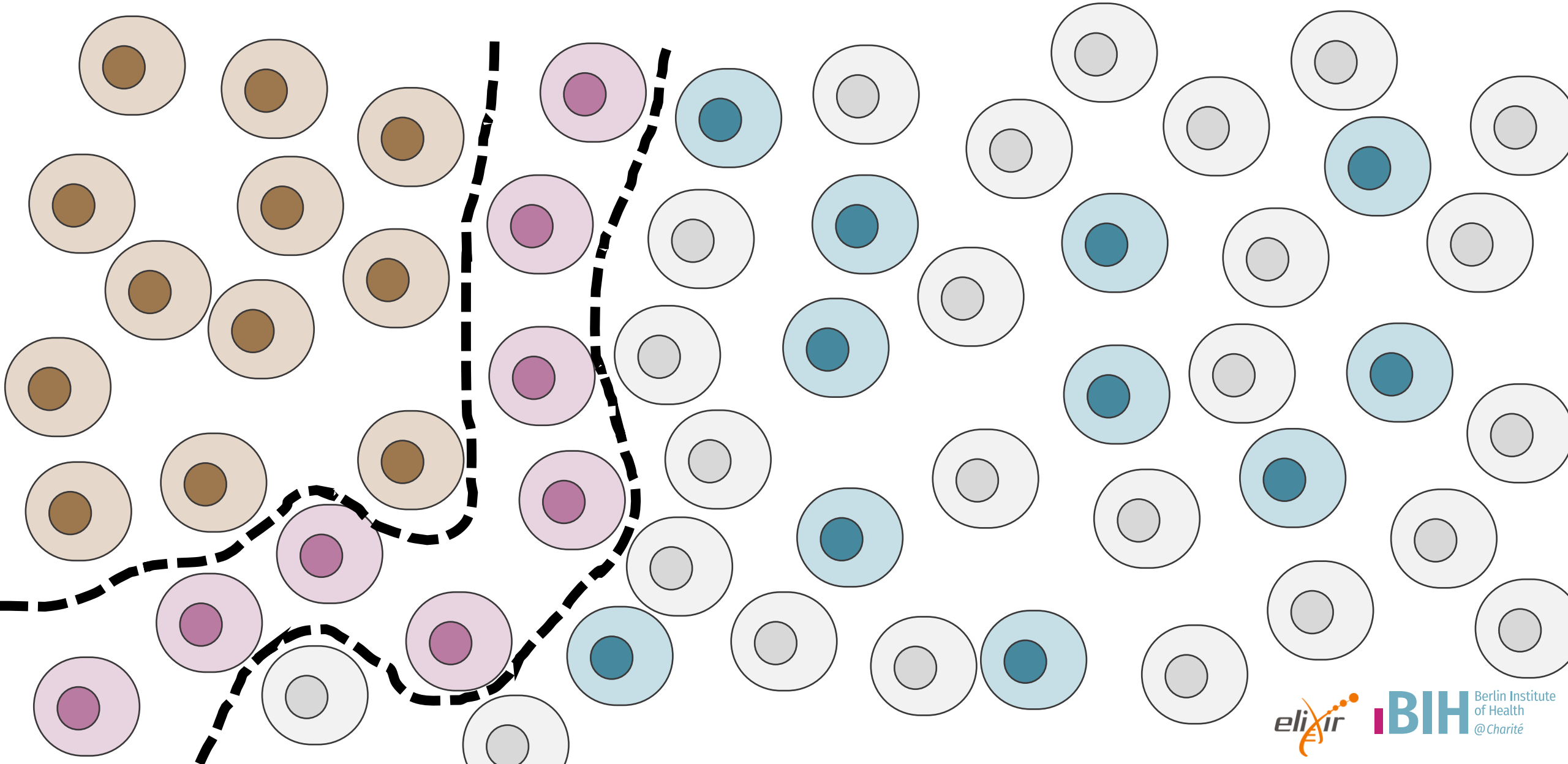
MERFISH/VizGen, ISS/Xenium, Molecular Cartography, CosMx, Barista-seq, seqFISH, STARmap, EEL-FISH, …

Adapted from Rao et al (2021) Nature

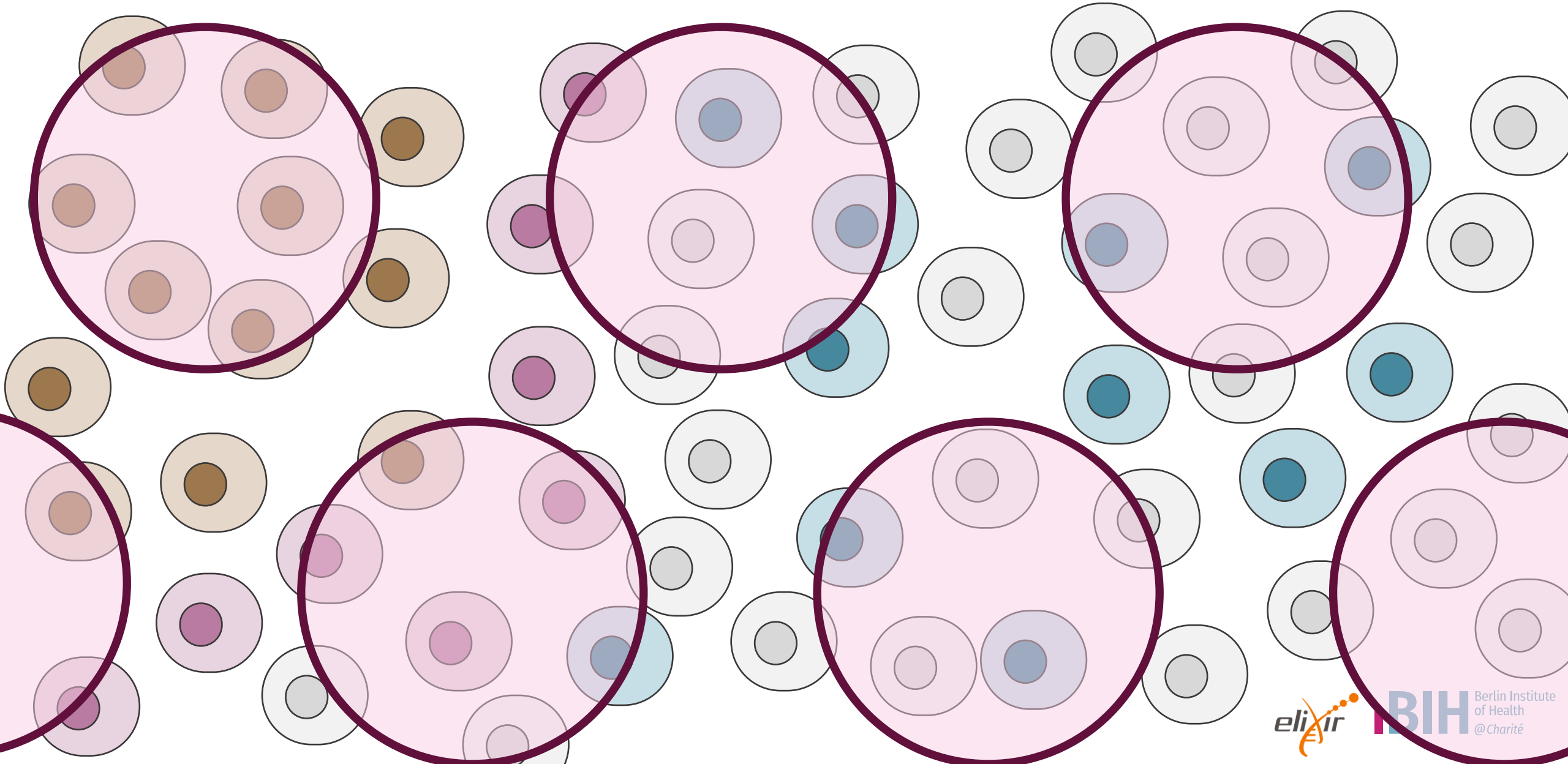
Spatial resolution



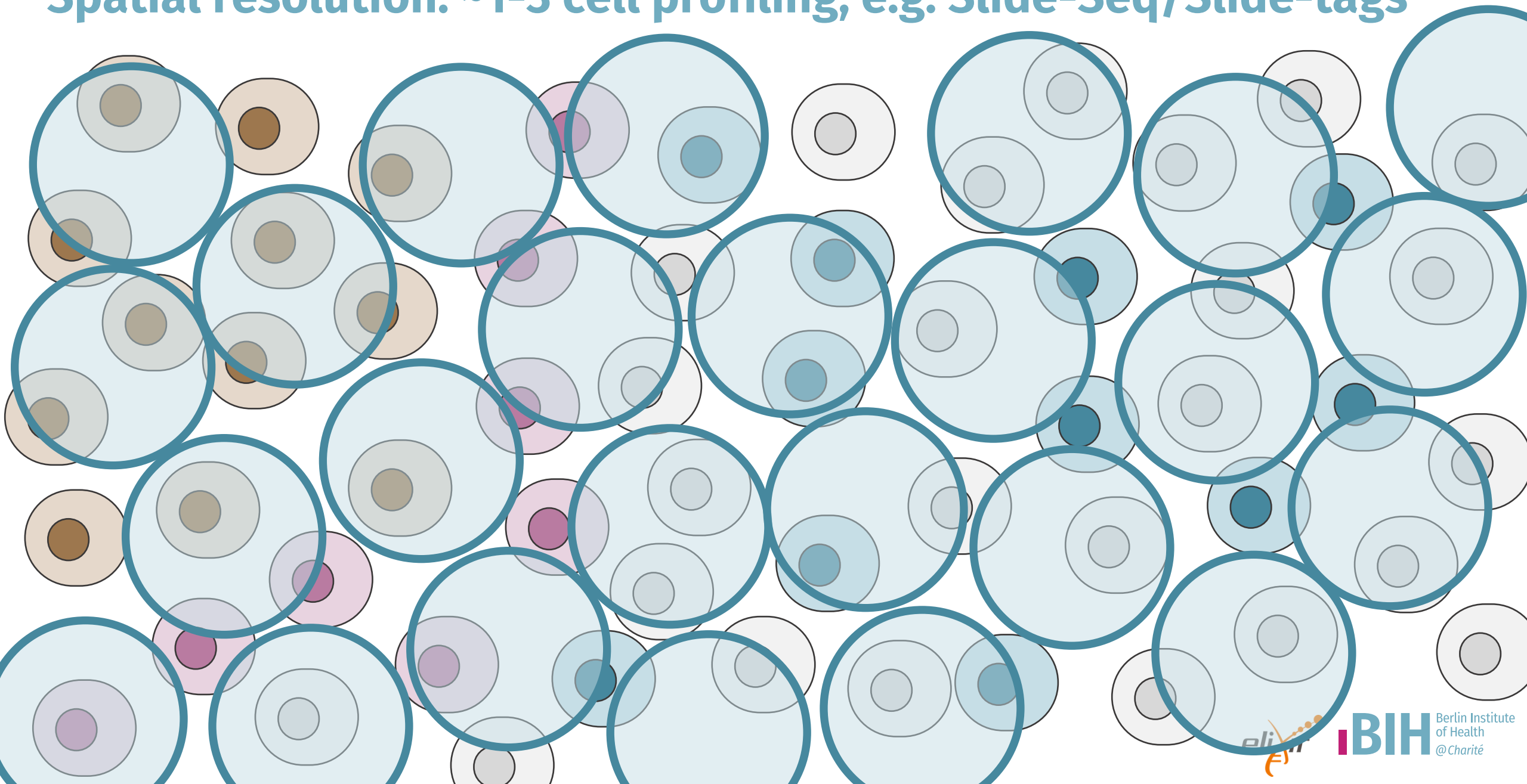
Spatial resolution: microdissection, e.g. TIVA, Geo-seq, etc



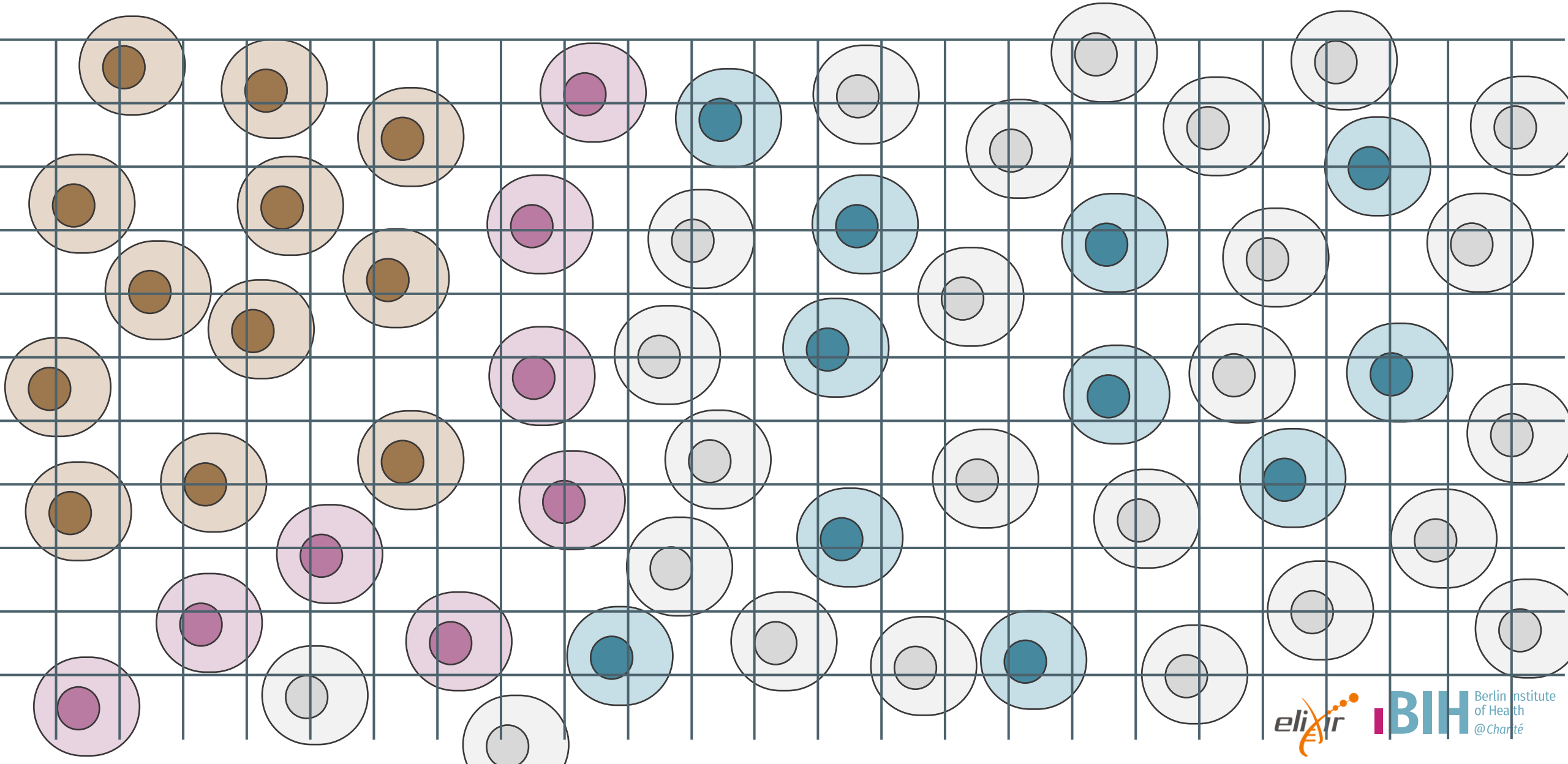
Spatial resolution: supracellular grid, e.g. Visium



Spatial resolution: ~1-3 cell profiling, e.g. Slide-Seq/Slide-tags

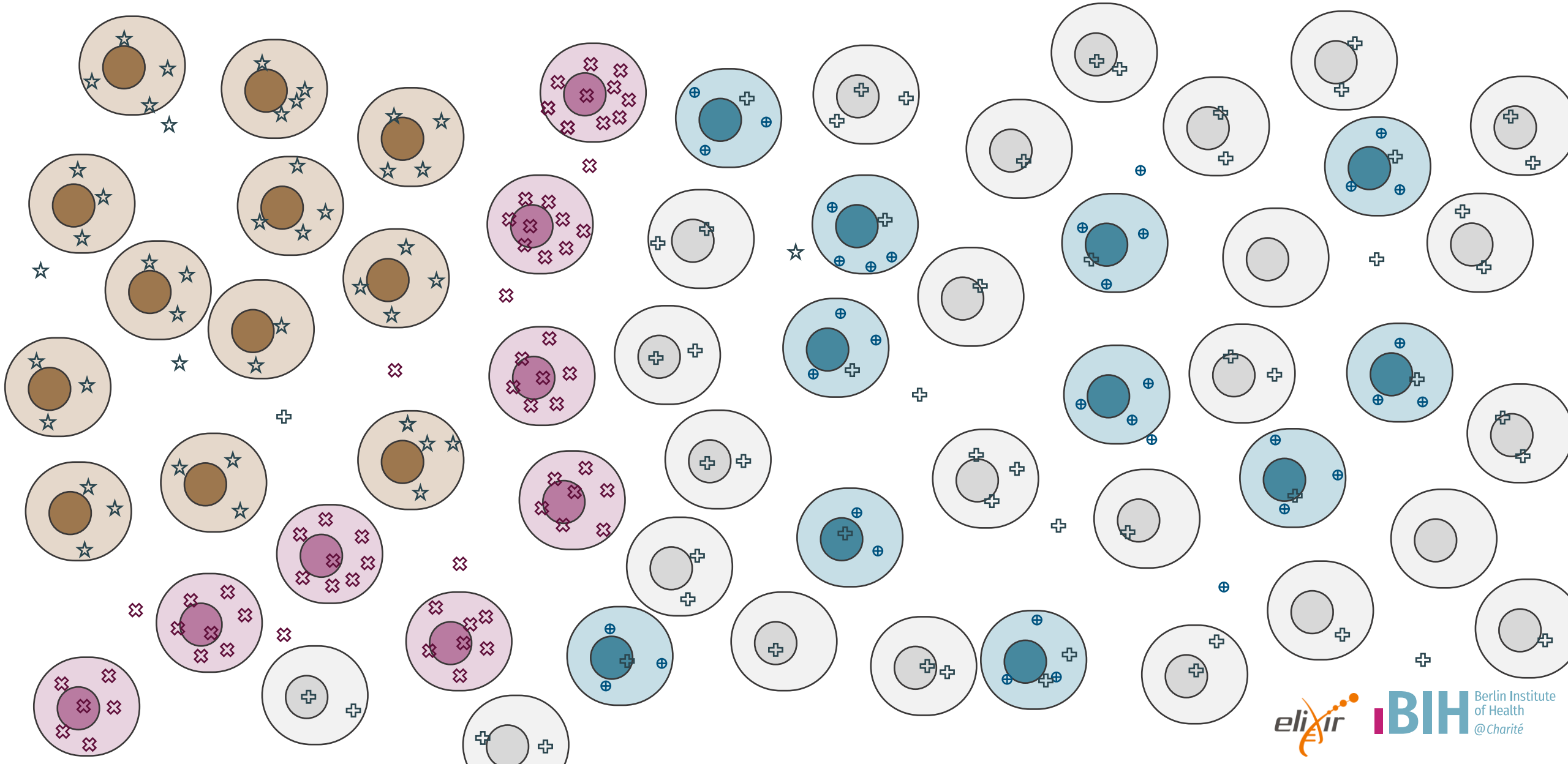


Spatial resolution: sub-cellular, e.g. VisiumHD



elixir BIH Berlin Institute of Health @Charité

Spatial resolution: single molecule, e.g. MERSCOPE, cosMX, Xenium



elixir

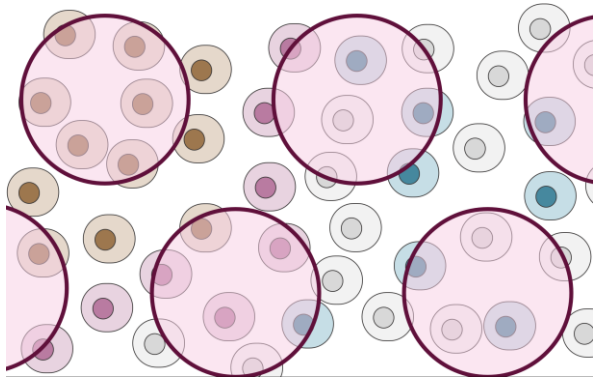
BIH Berlin Institute of Health @Charité

Resolution of spatial transcriptomics technologies

Spatial resolution	Example Technologies
Microdissection	TIVA, Geo-seq, etc
~3-20 Cells	ST, Visium, etc
~1-3 Cell	Curio seeker, VisiumHD, etc
sub cellular	VisiumHD, openST, stereo-seq
Single Molecule	Xenium, MERSCOPE, molecular cartography, STARmap, etc

Resolution of spatial transcriptomics and challenges

Spatial resolution	Example Technologies	Computational challenge w.r.t gene expression signals
Microdissection	TIVA, Geo-seq, etc	Deconvoluting mixed signals in spots
~3-20 Cells	ST, Visium, etc	
~1-3 Cell	Curio seeker, VisiumHD, etc	
sub cellular	VisiumHD, openST, stereo-seq	
Single Molecule	Xenium, MERSCOPE, molecular cartography, STARmap, etc	



elixir BIH Berlin Institute of Health @Charité

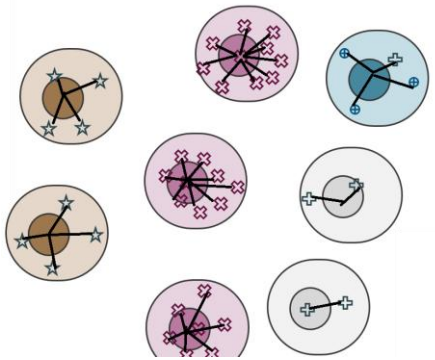
Resolution of spatial transcriptomics and challenges

Low-res

High-res

Spatial resolution	Example Technologies	Computational challenge w.r.t gene expression signals
Microdissection	TIVA, Geo-seq, etc	
~3-20 Cells	ST, Visium, etc	
~1-3 Cell	Curio seeker, VisiumHD, etc	
sub cellular	VisiumHD, openST, stereo-seq	
Single Molecule	Xenium, MERSCOPE, molecular cartography, STARmap, etc	

Aggregating signals into cells

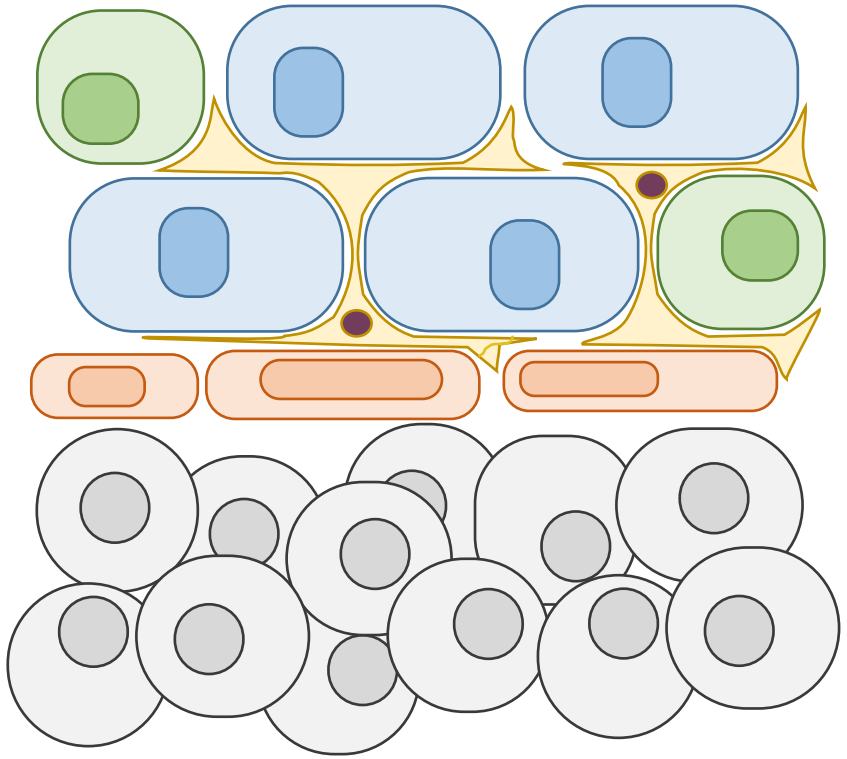


elixir

BIH Berlin Institute of Health @Charité

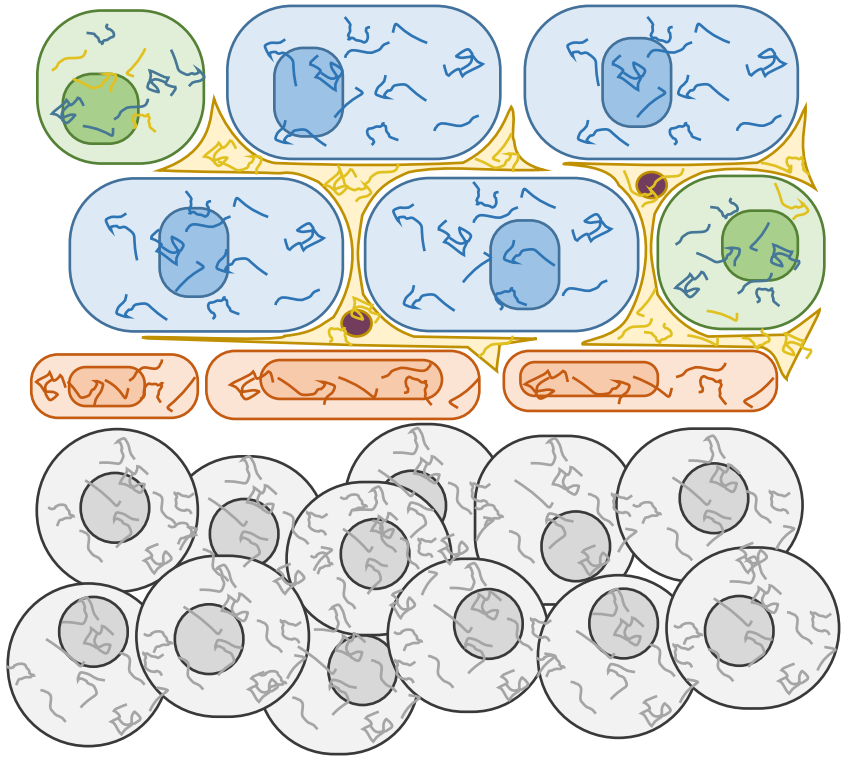
Imaging-based spatial transcriptomics

- Cells in tissue



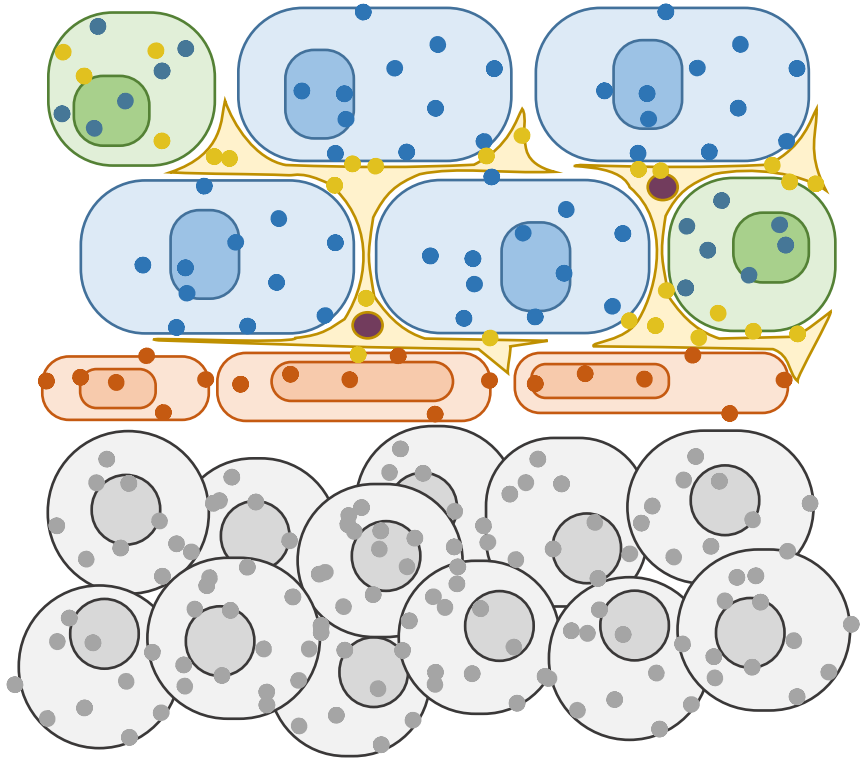
Imaging-based spatial transcriptomics

- Cells in tissue
- Transcripts locations



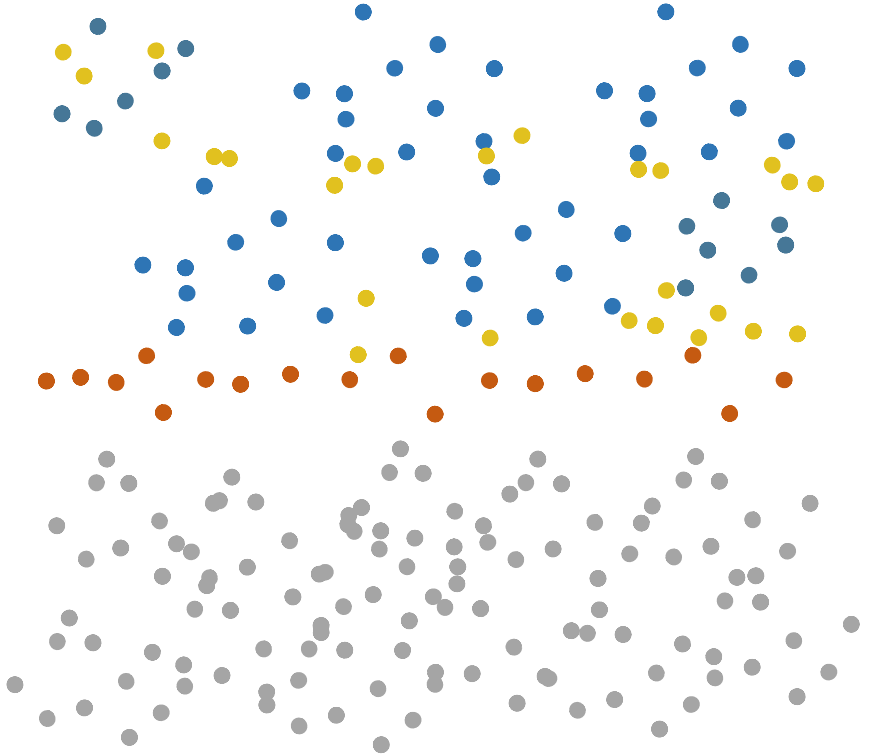
Imaging-based spatial transcriptomics

- Cells in tissue
- Transcripts locations
- Image transcript locations



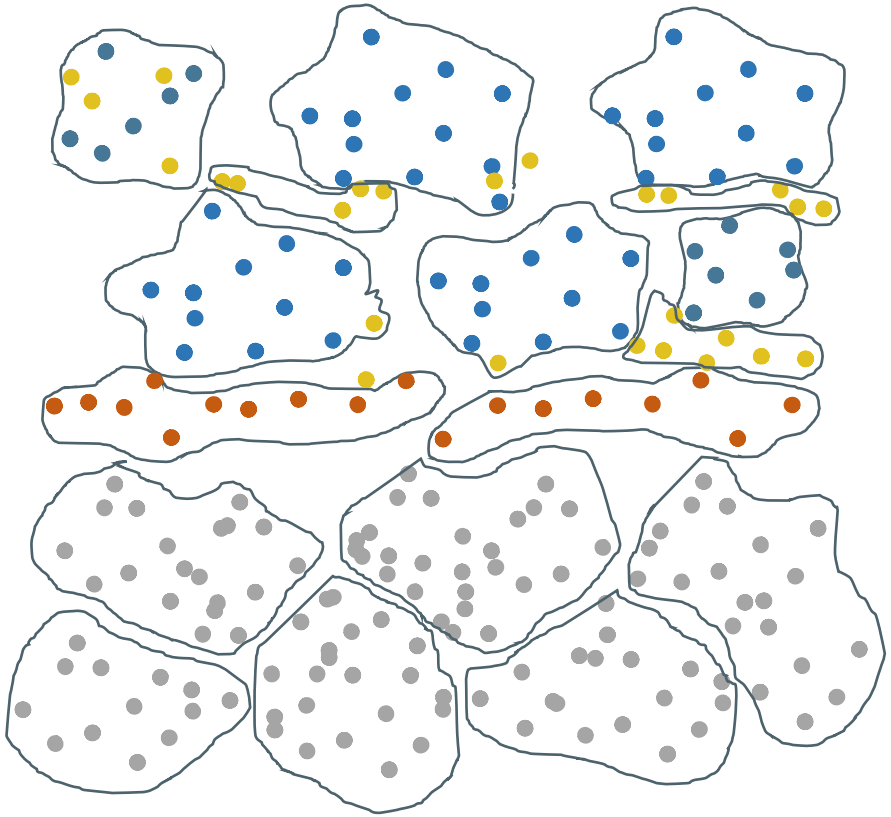
Imaging-based spatial transcriptomics

- Cells in tissue
- Transcripts locations
- Image transcript locations
- Report transcript locations



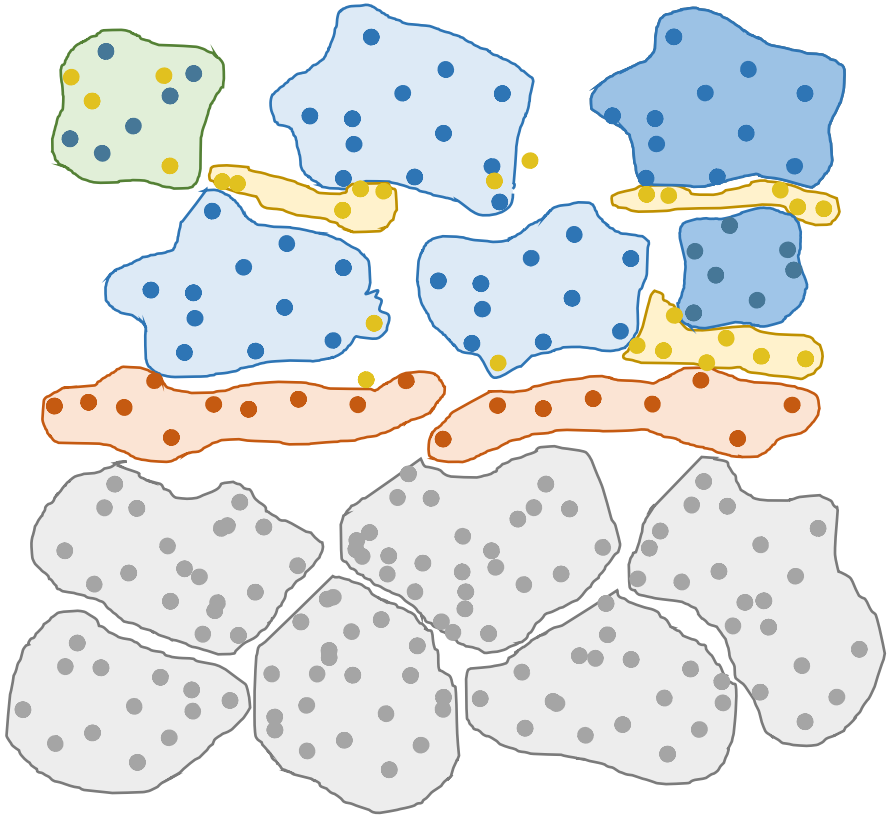
Imaging-based spatial transcriptomics

- Cells in tissue
- Transcripts locations
- Image transcript locations
- Report transcript locations
- Identify cells (segmentation)



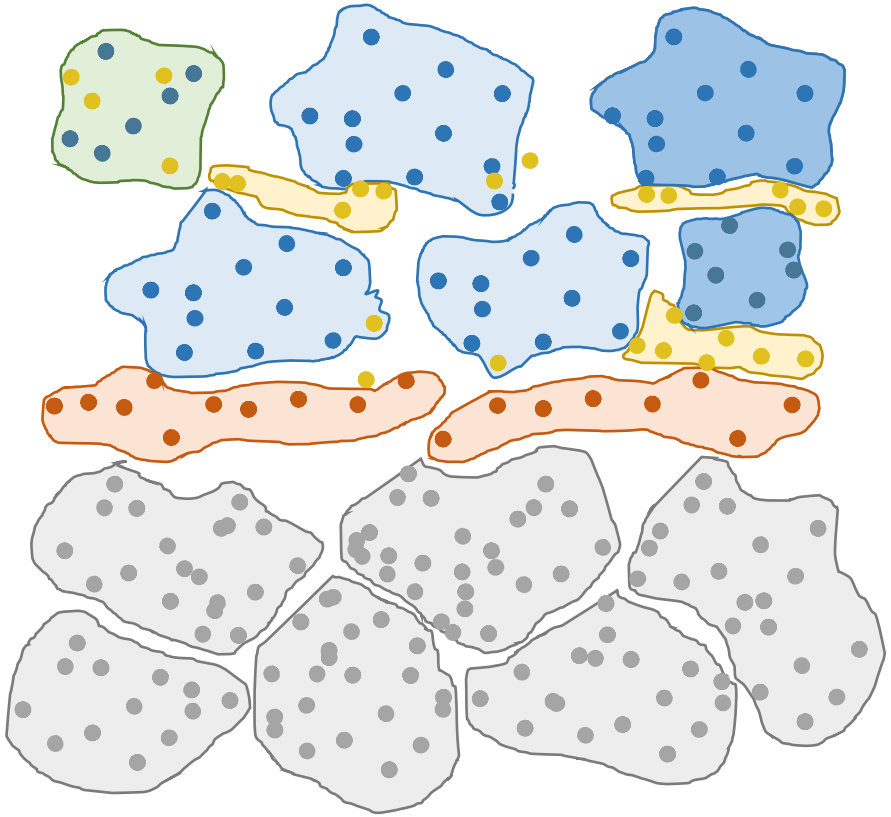
Imaging-based spatial transcriptomics

- Cells in tissue
- Transcripts locations
- Image transcript locations
- Report transcript locations
- Identify cells (segmentation)
- Analyse cells

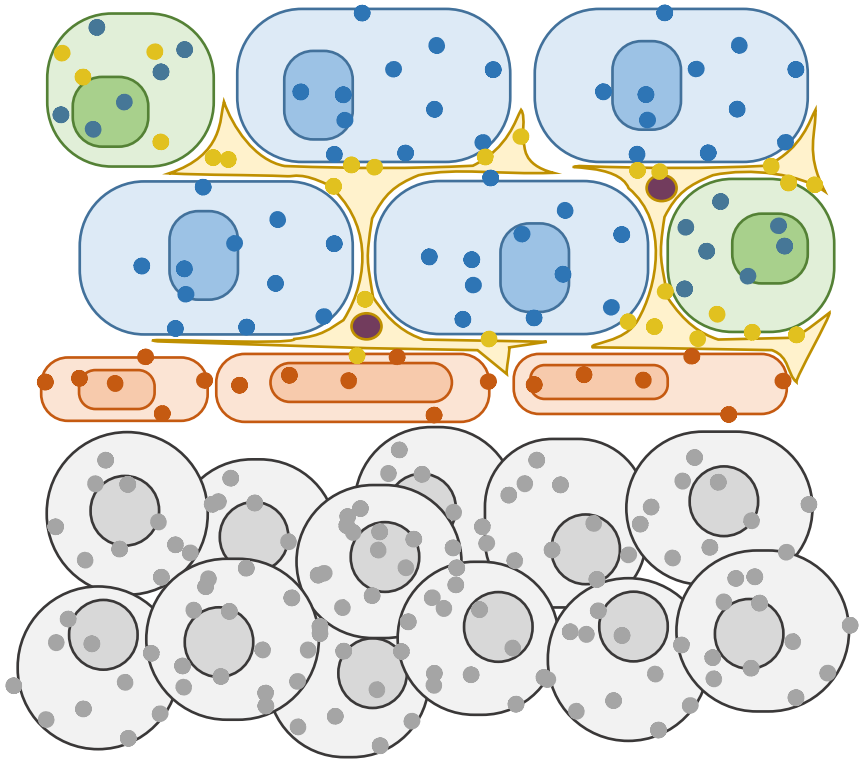


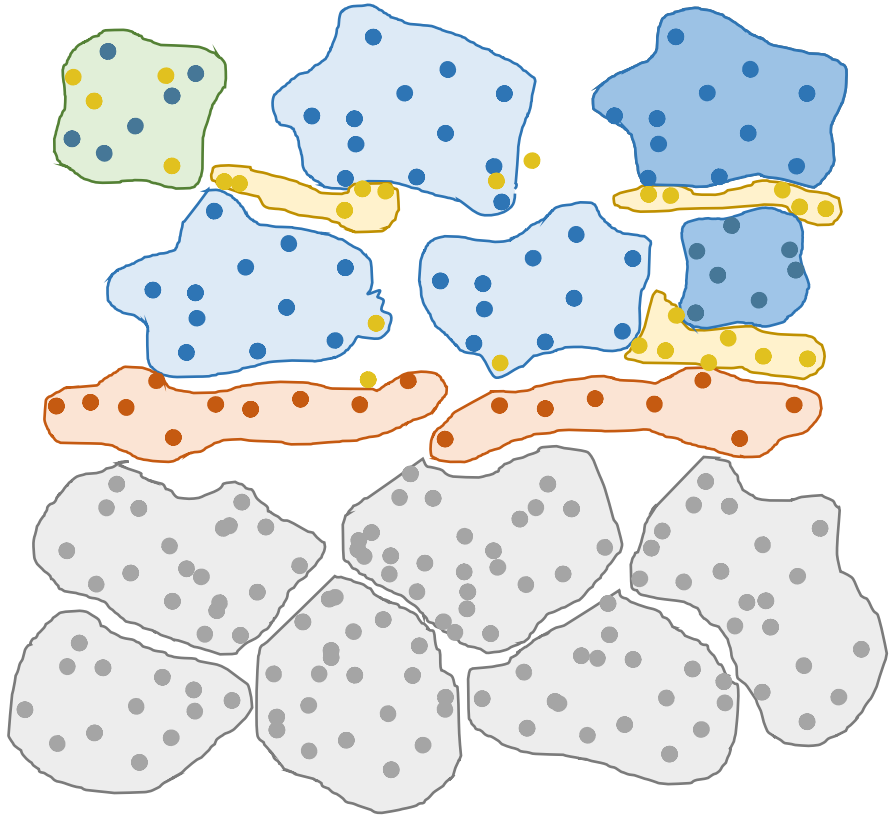
Imaging-based spatial transcriptomics

- Cells in tissue
- Transcripts locations
- Image transcript locations
- Report transcript locations
- **Identify cells (segmentation)**
- Analyse cells

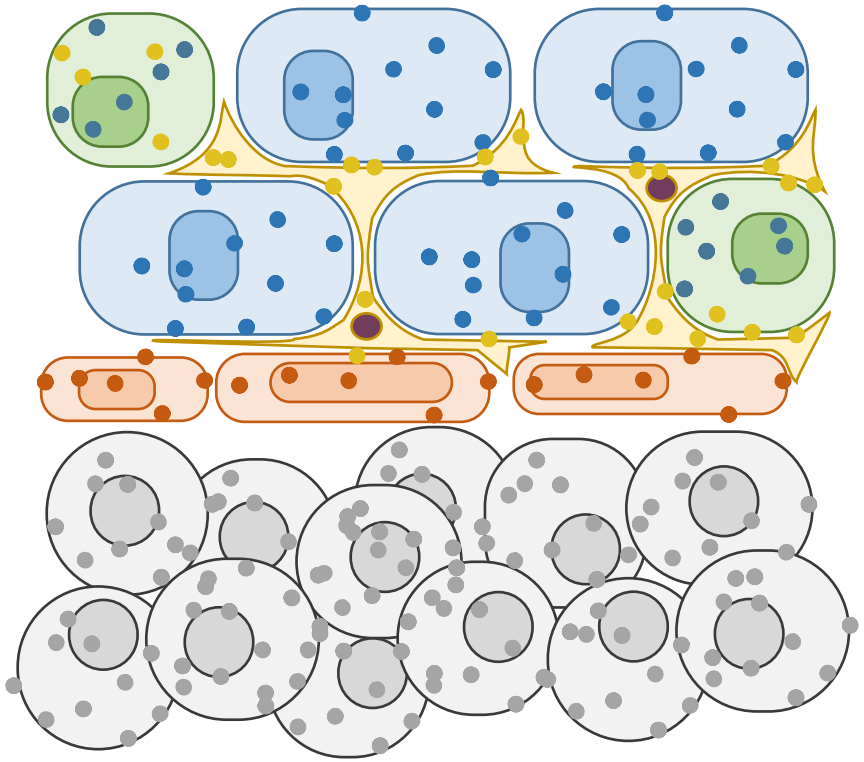


Evaluation - spot the difference!

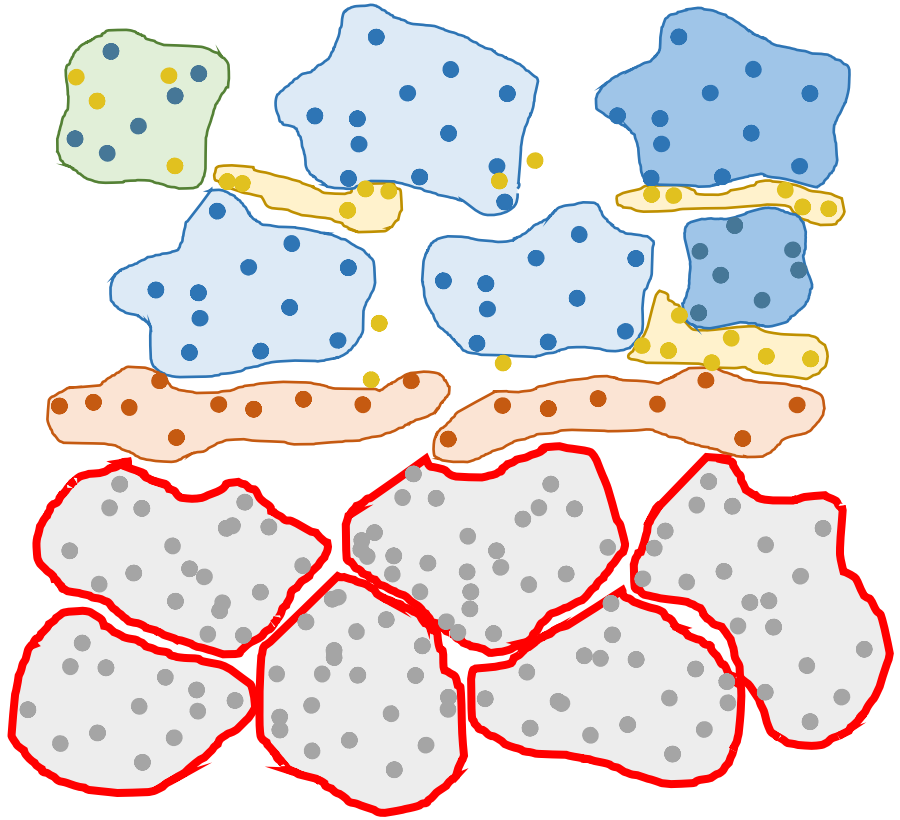




Evaluation – cell proportions

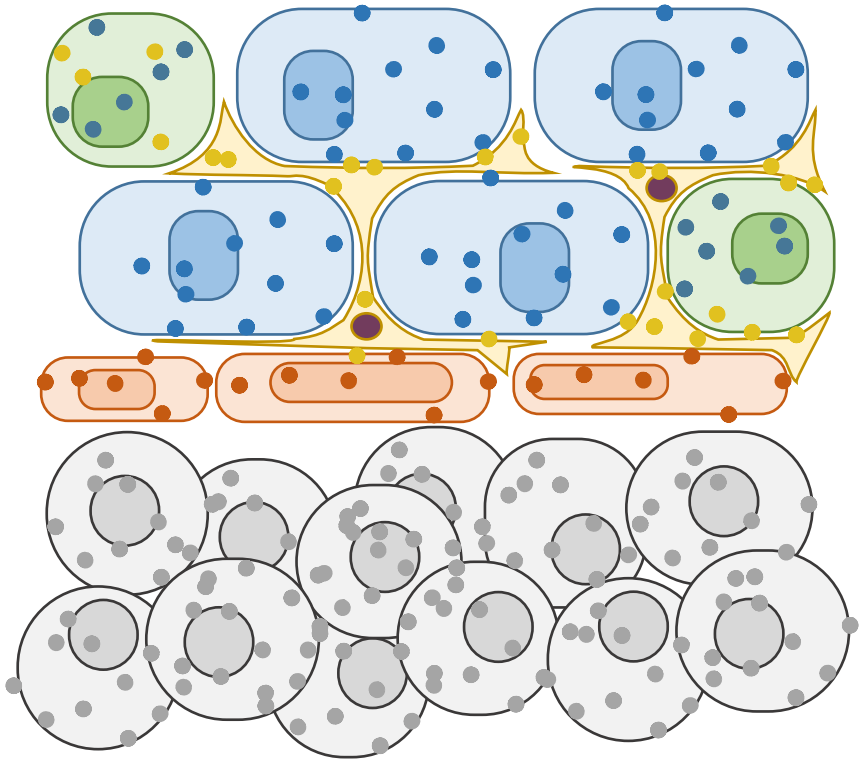


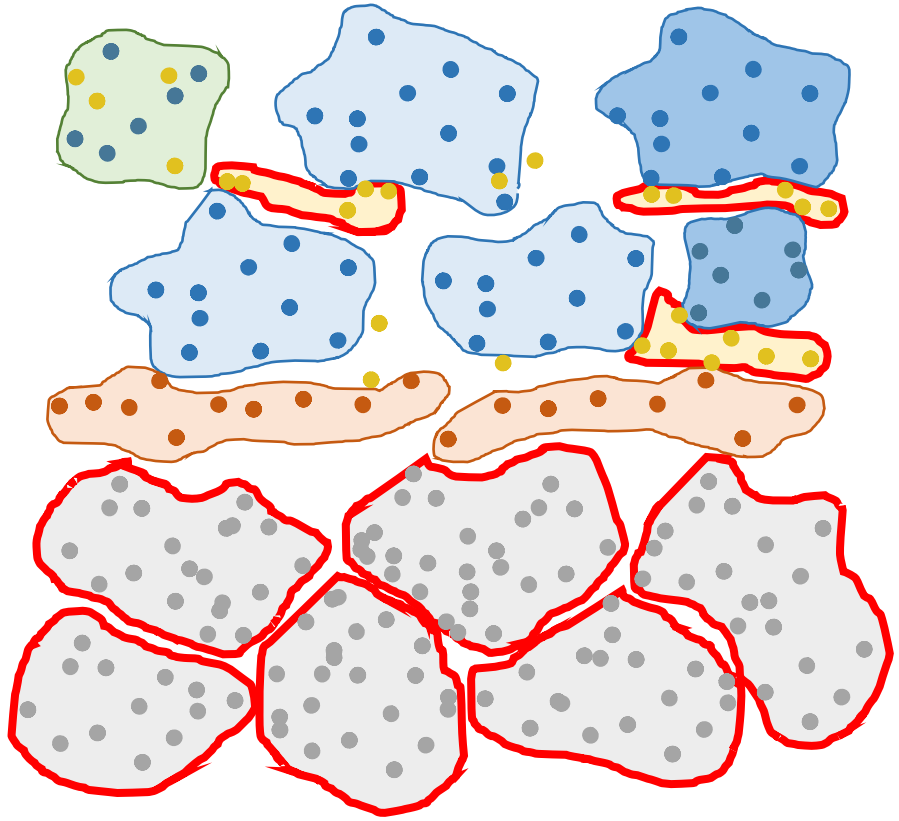
Green, blue, orange, yellow cells all +/- 1



Highly dense grey cells are hard to separate

Evaluation – cell size

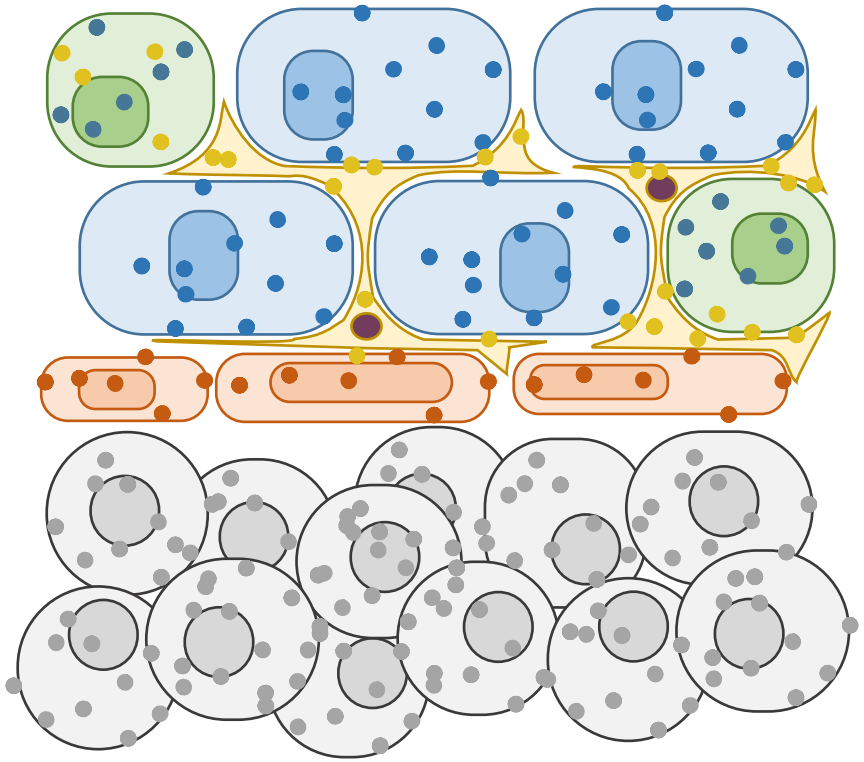


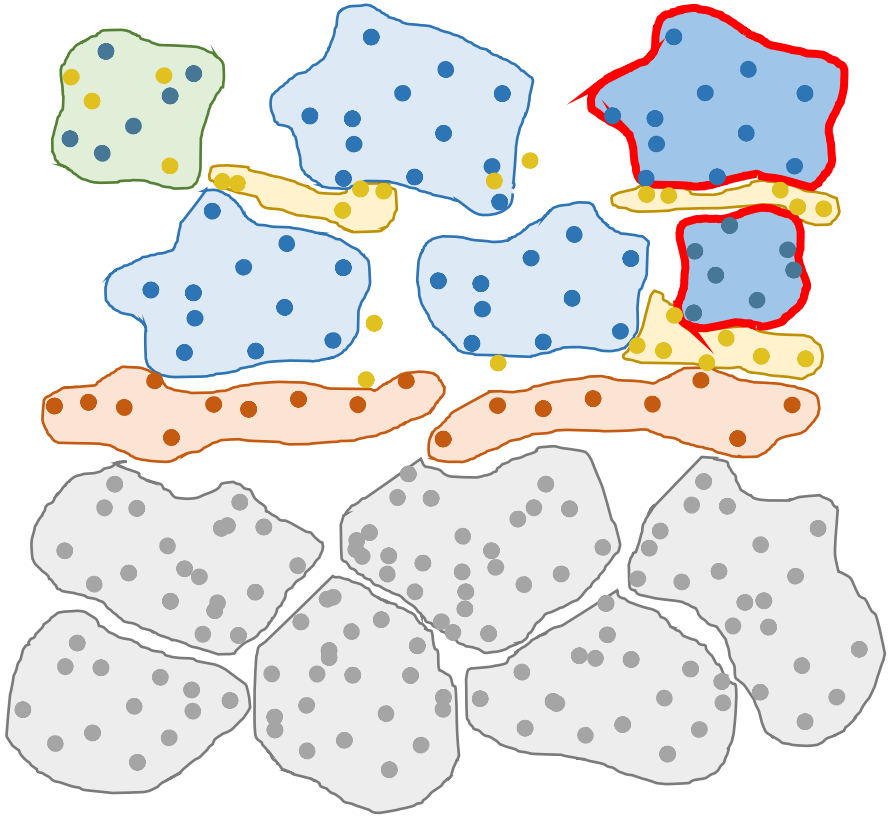


yellow cells are split or missed

Highly dense grey cells are too large

Evaluation – cell type annotation

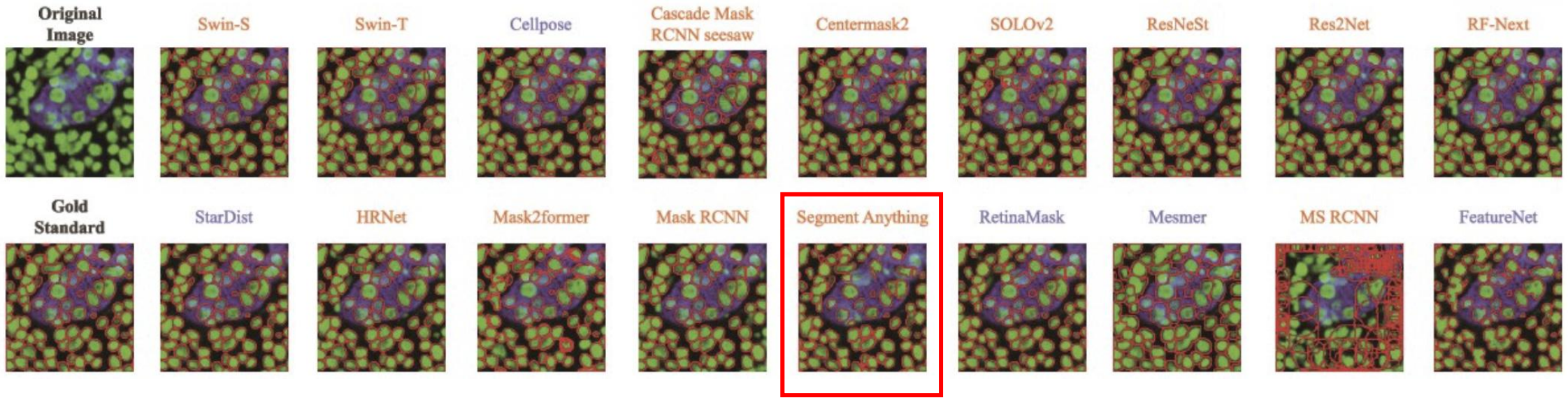




Cell segmentation algorithms

Mainly demonstrated on segmenting DAPI (a nucleus stain)

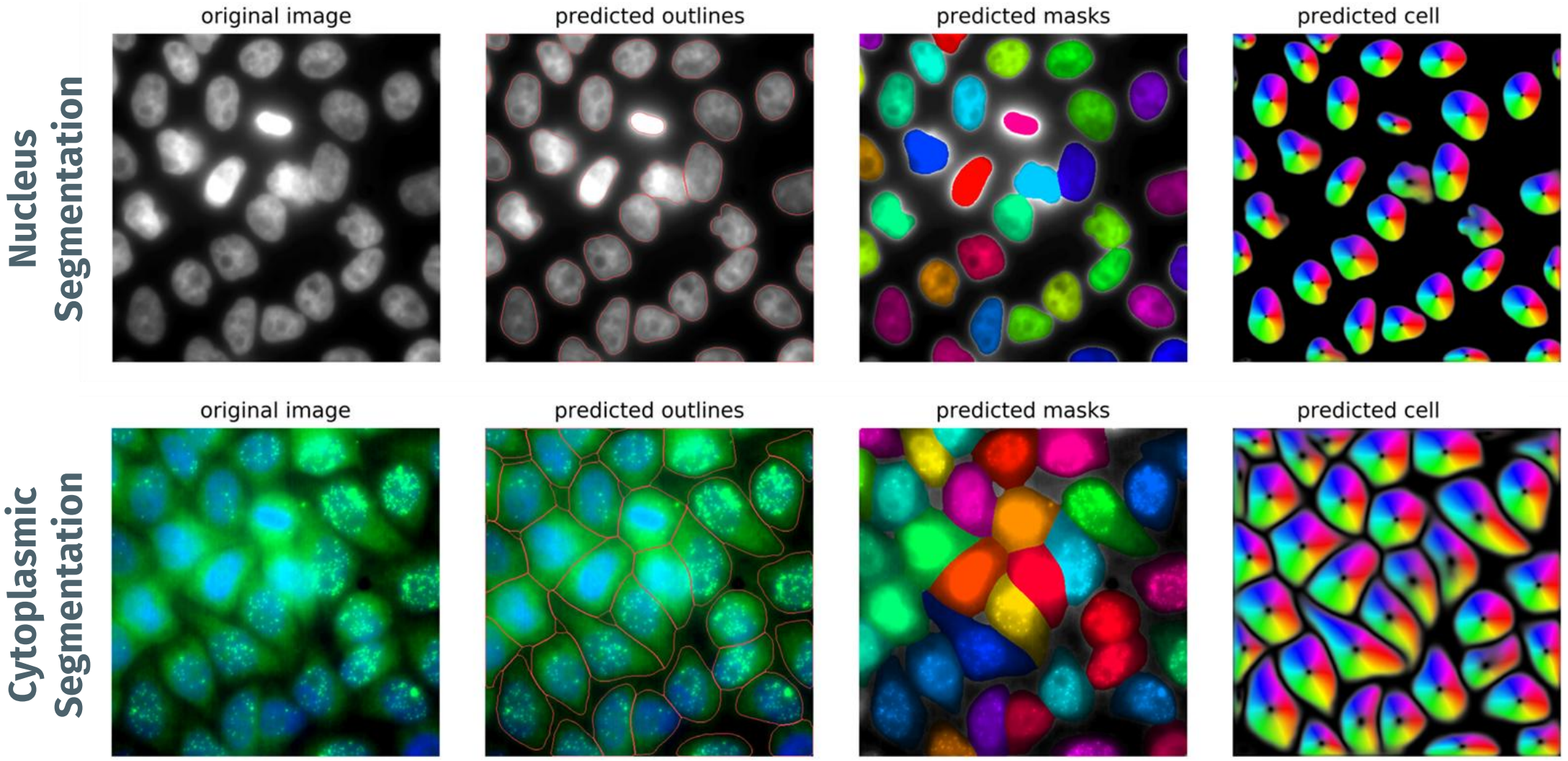
Watershed is considered a reference algorithm, but there are many others...



Benchmark: Wang et al (2024) Briefings in Bioinformatics, https://doi.org/10.1093/bib/bbae407

[1] Wang et al (2024) Briefings in Bioinformatics

Cell segmentation by staining cell landmarks

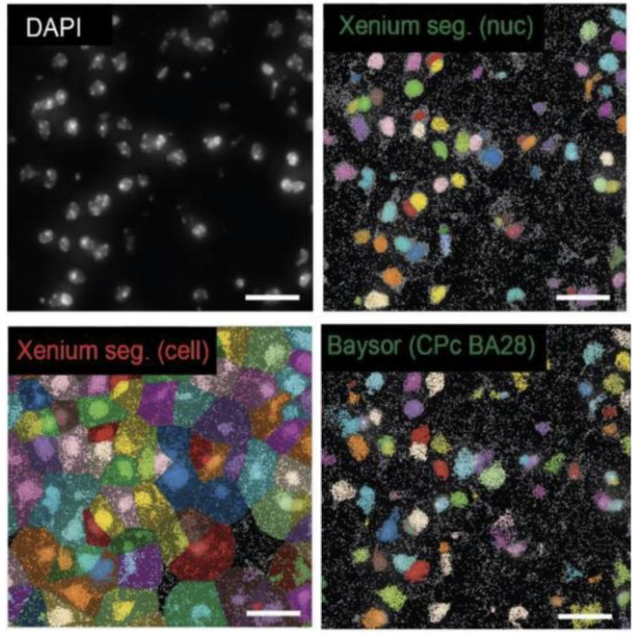


https://fq-segmentation.readthedocs.io/

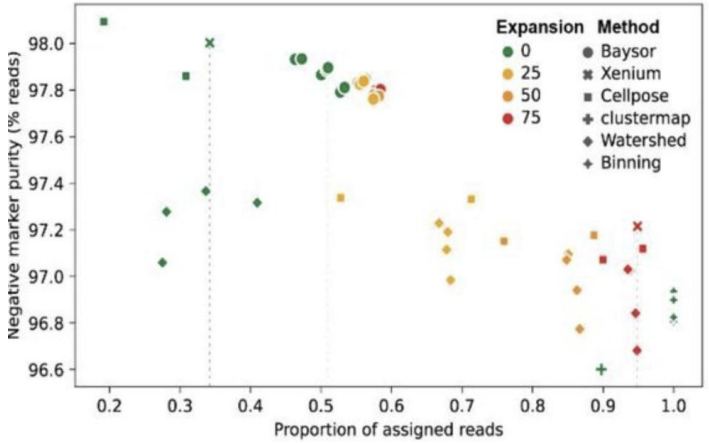
Cell segmentation is typically DAPI (nucleus) + expansion

Sweeping assumptions:

- nucleus at the centre of the cell
- cell shapes are roundish (or square-ish when they are close to others)
- cells are all the same size (unless they are close to others)
- Users want to optimise % of transcript in cells



Incorrect segmentation leads to incorrect assignment of transcripts



[1] Salas et al (2023) biorXiv

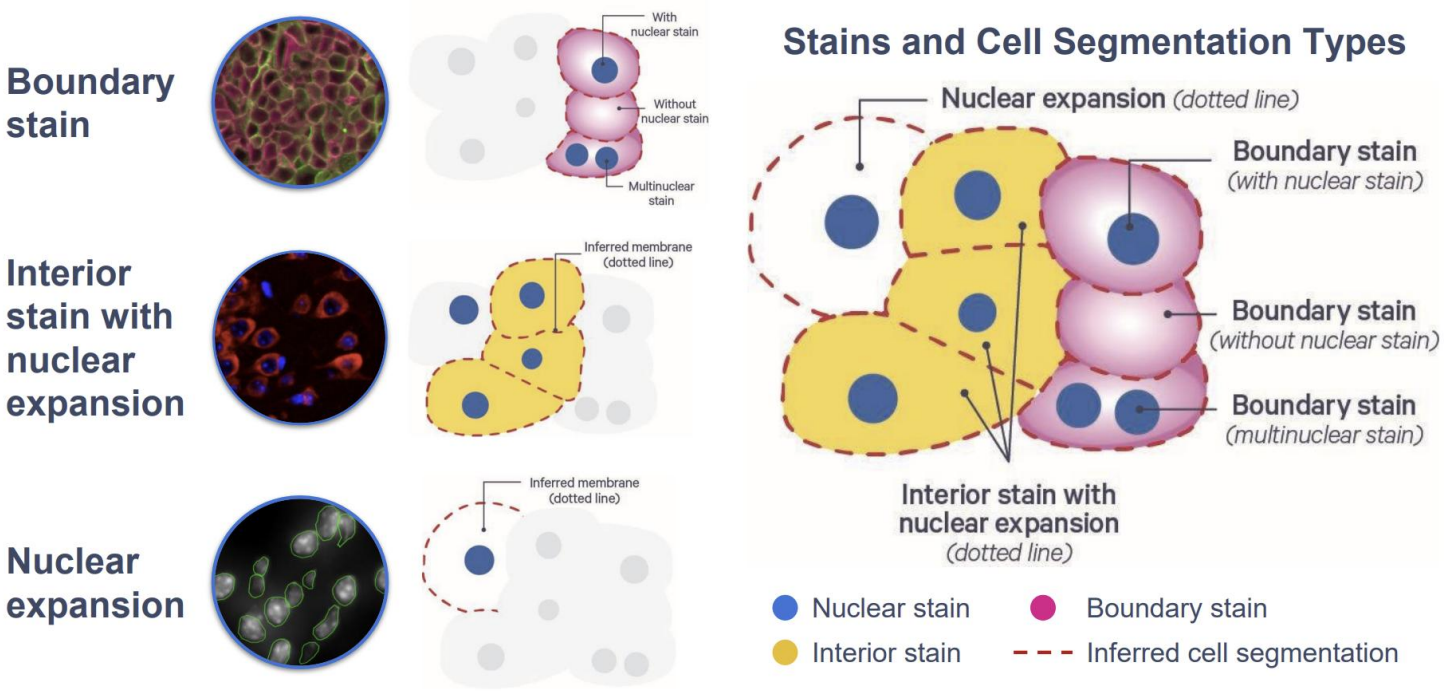
elixir

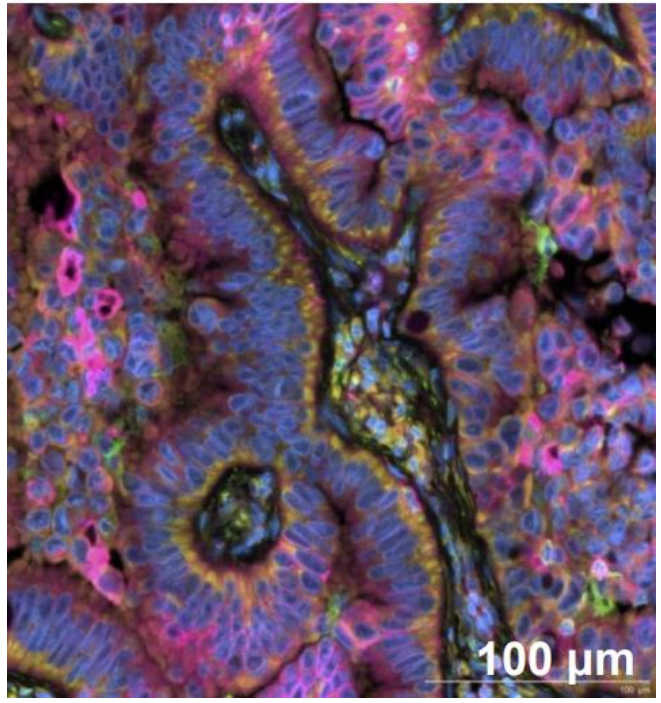
IBIH Berlin Institute of Health @Charité

Improving cell segmentation - staining multiple cell landmarks

Vendors now offer staining of multiple cell landmarks for “multi-modal” cell segmentation

(... looks beautiful compared to just DAPI)





[1] https://pages.10xgenomics.com/rs/446-PBO-704/images/AGBT_2024_Cell_Segmentation_Poster.pdf

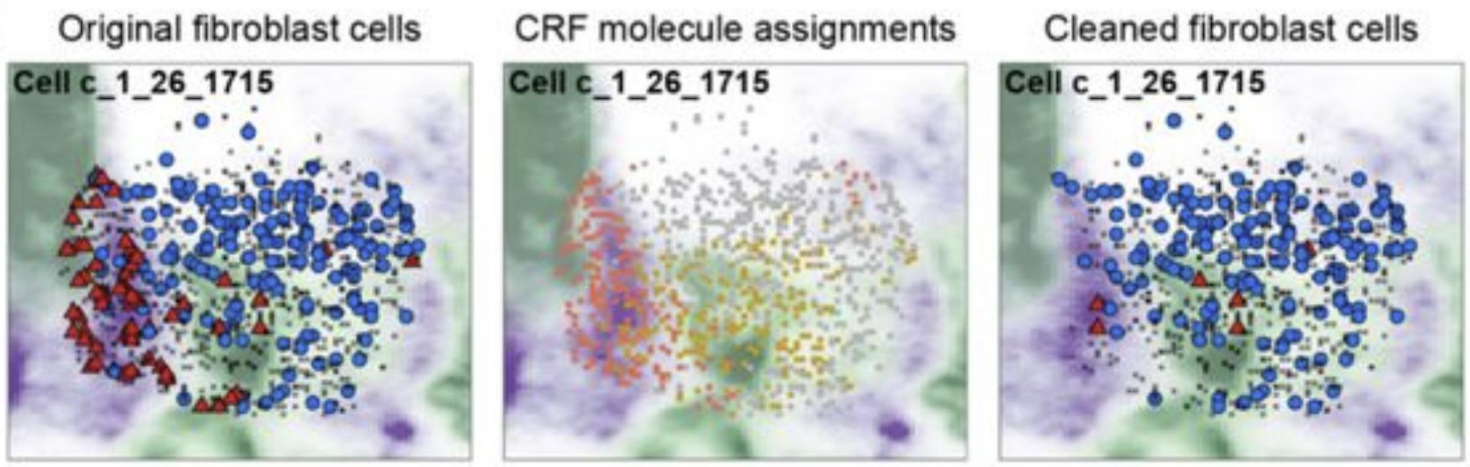
elixir

BIH Berlin Institute of Health @Charité

Emerging post-segmentation quality control: spatial doublets

Missegmentation incorrectly assign transcripts from adjacent cells

- Referred to as “spatial doublets”
- … we are trying to call these “x-y spatial doublets”…



Impact of Segmentation Errors in Analysis of Spatial Transcriptomics Data

Jonathan Mitchel, Teng Gao, Eli Cole, Viktor Petukhov, Peter V. Kharchenko

doi: https://doi.org/10.1101/2025.01.02.631135

This article is a preprint and has not been certified by peer review [what does this mean?].

Abstract | Full Text | Info/History | Metrics | Preview PDF

Abstract

Spatial transcriptomics aims to elucidate cell coordination within biological tissues by linking the state of the cell with its local tissue microenvironment. Imaging-based assays are particularly promising for exploring such interdependencies, as they can resolve molecular and cellular features with subcellular resolution in three dimensions. Quantification and analysis of cellular state in such data, however, ultimately depends on the ability to recognize which molecules belong to each cell. Despite computational and experimental progress, this cell segmentation task remains challenging. Here we re-analyze data from multiple tissues and platforms and find that segmentation errors currently confound most downstream analysis of cellular state, including analysis of differential expression, inference of neighboring cell influence, and ligand-receptor interactions. The extent to which mis-segmented molecules impact the results can be striking, often dominating the set of top hits. We show that factorization of molecular neighborhoods can be effective at isolating such molecular admixtures and minimizing their impact on downstream analysis, analogous to doublet filtering of scRNA-seq data. As applications of spatial transcriptomics assays become more widespread, we expect corrections for the confounding effect of segmentation errors to become increasingly important for being able to resolve molecular mechanisms of tissue biology.

Mitchel et al (2025) bioRxiv

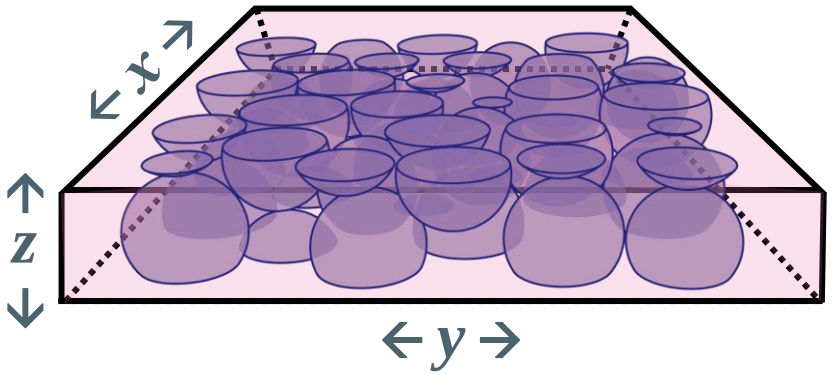
elixir

BIH Berlin Institute of Health @Charité

Overlapping cells in tissue sections (... spatial doublets)

Even though tissues sections are verrry thin, they are still 3D

- "Z-type spatial doublets"
- How many cells do you expect to overlap?



BONUS PRESENTATION!

- Check out slides 82 onwards

2D, or not 2D? Investigating Vertical Signal Integrity of Tissue Slices

Sebastian Tiesmeyer, Niklas Müller-Bötticher, Alexander Malt, Brian Long, Sergio Marco-Salas, Paul Kiessling, Paul Horn, Adrien Guillot, Louis B Kuemmerle, Leyao Ma, Frank Tacke, Fabian Theis, Christoph Kuppe, Mats Nillson, Roland Eils, Naveed Ishaque

doi: https://doi.org/10.1101/2025.01.13.632601

This article is a preprint and has not been certified by peer review [what does this mean?].

Abstract | Full Text | Info/History | Metrics | Preview PDF

Abstract

Imaging-based spatially resolved transcriptomics can localise transcripts within cells in 3D. Cell segmentation precedes assignment of transcripts to cells and annotation of cell function. However, cell segmentation is usually performed in 2D, thus unable to deal with spatial doublets arising from overlapping cells, resulting in segmented cells containing transcripts originating from multiple cell-types. Here we present a computational tool called ovrlpy that identifies overlapping cells, tissue folds and inaccurate cell-segmentation.

Competing Interest Statement

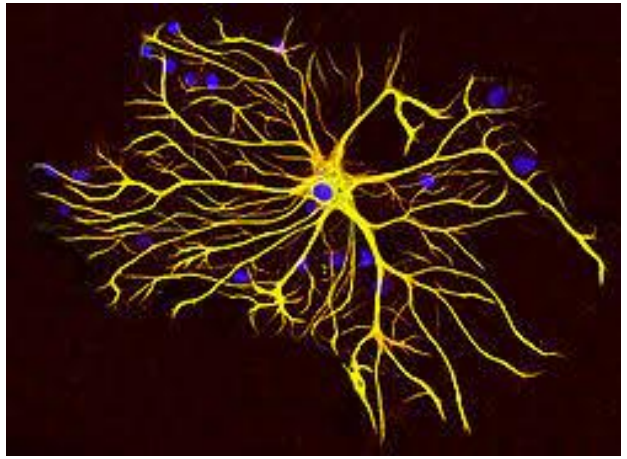
The authors have declared no competing interest.

Tiesmeyer et al (2025) bioRxiv

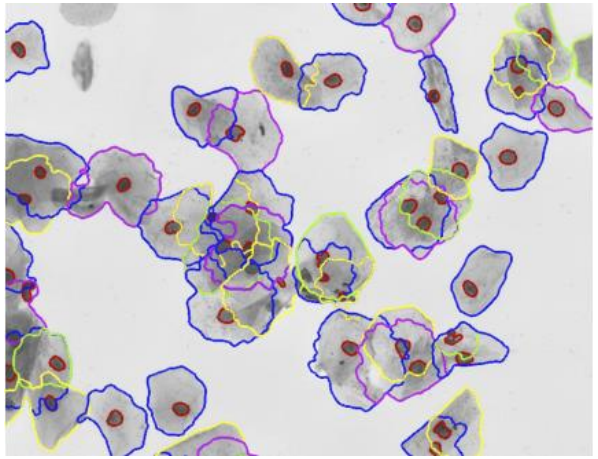
elixir

BIH Berlin Institute of Health @Charité

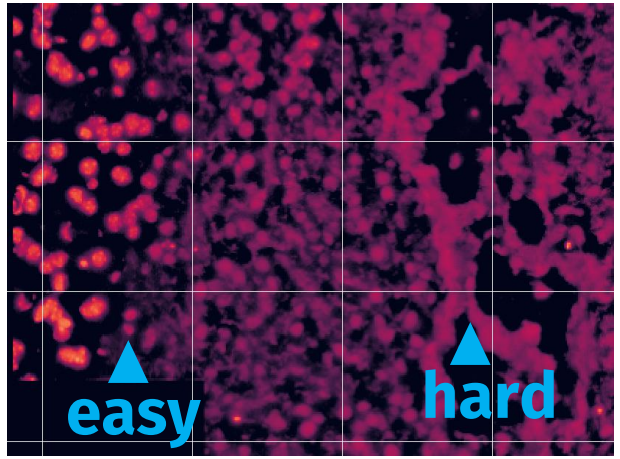
Cell segmentation isn't always easy



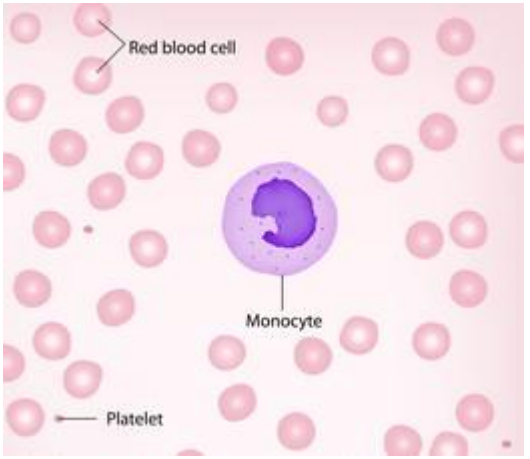
Complex shapes



Overlapping cells



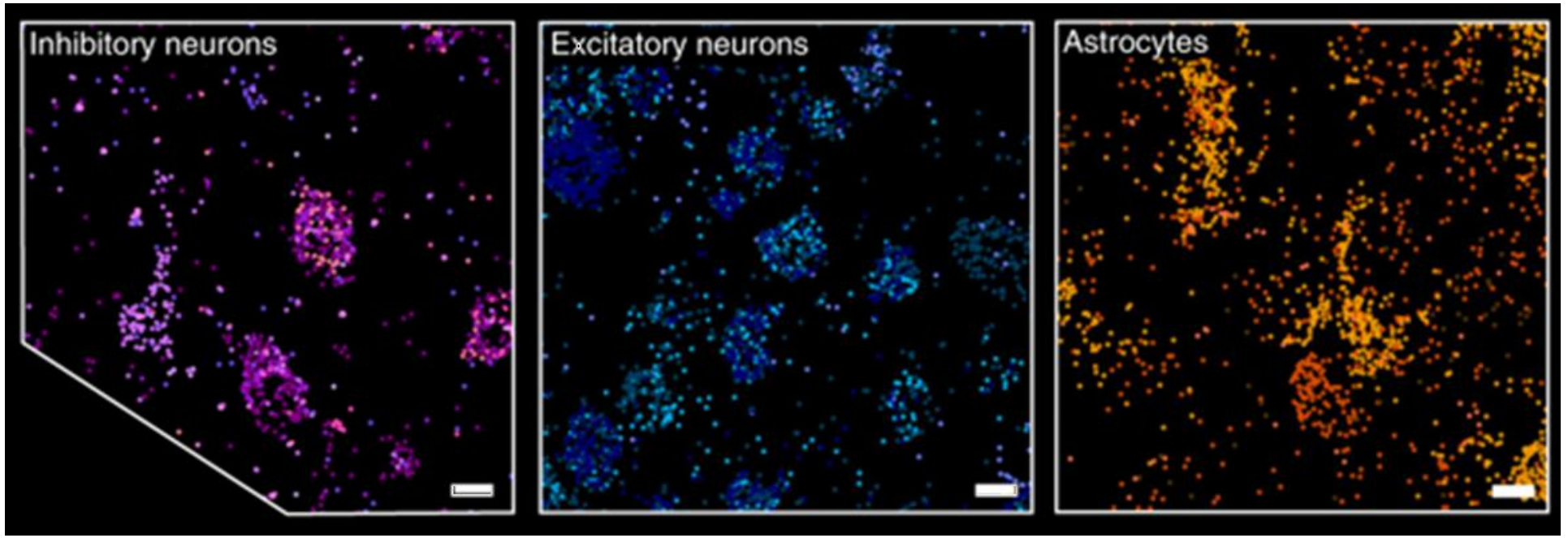
Cell stains might have issues



Cells might not be stained correctly (e.g. red blood cells)

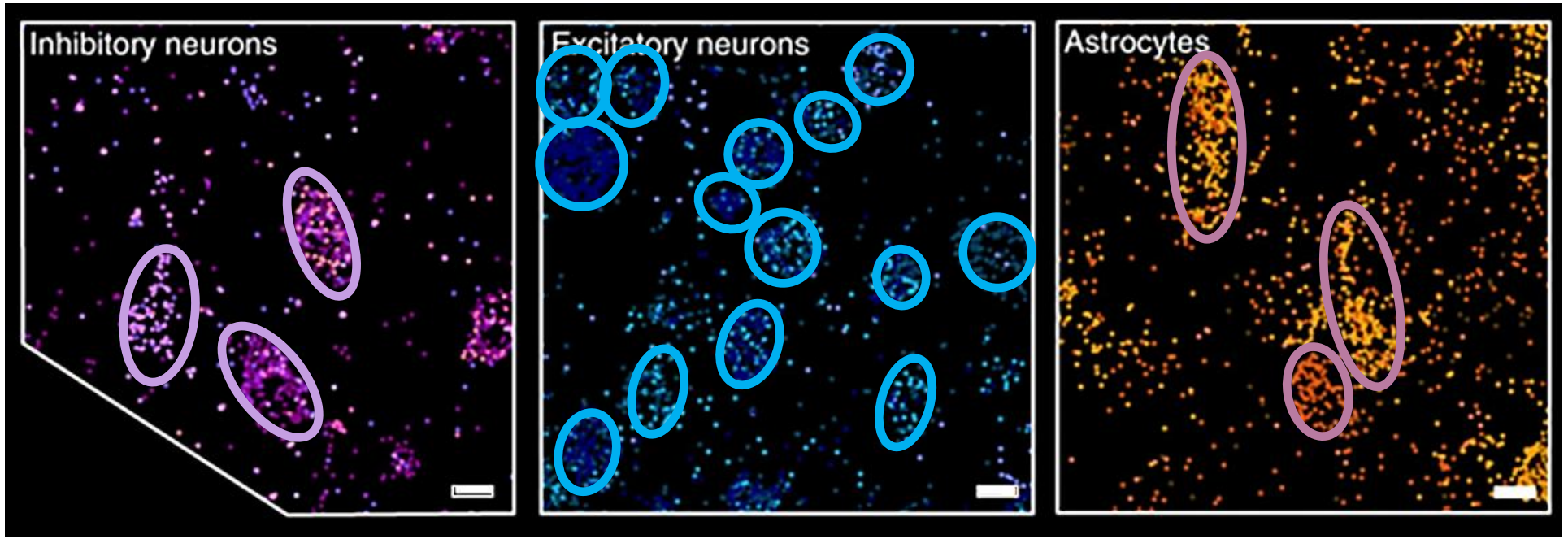
[1] Lu et al (2017) IEEE Trans Med Imaging. [2] Molnar et al (2016) Scientific Reports
[3] Codeluppi et al (2019) Nature Methods. [4] Shutterstock.com 161560277

mRNA molecule organisation patterns are not random



Codeluppi, Borm et al (2018) Nature Methods

mRNA molecule organisation patterns are not random

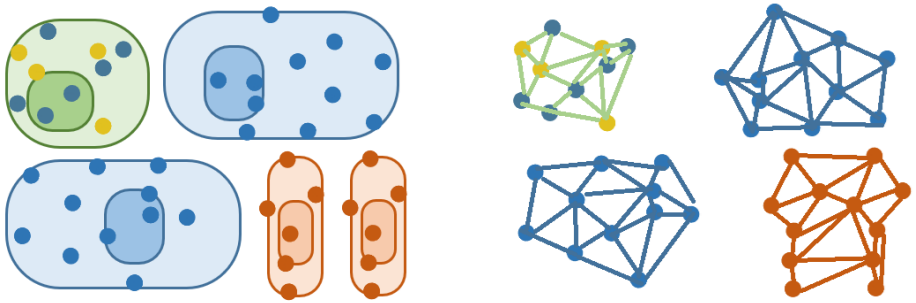


Codeluppi, Borm et al (2018) Nature Methods

Modelling mRNA distribution

Spatial model – how are mRNA molecules organised when they come from the same cell?

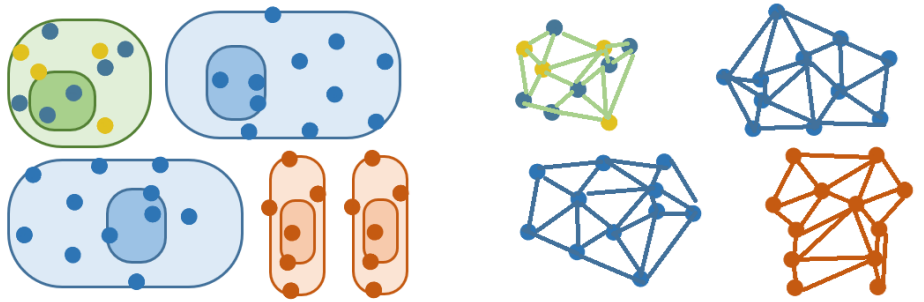
- Graph-based models



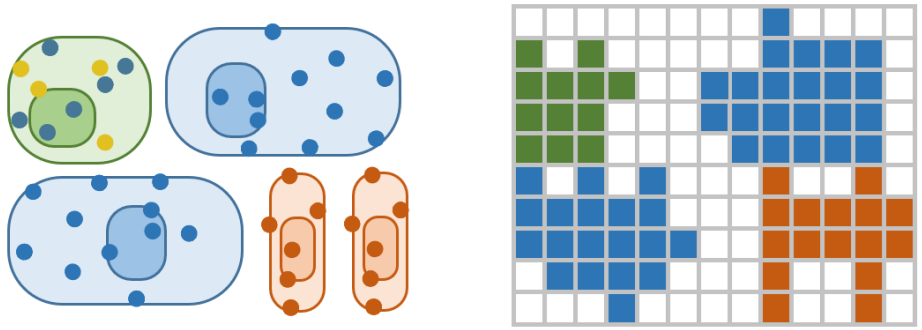
Modelling mRNA distribution

Spatial model – how are mRNA molecules organised when they come from the same cell?

- Graph-based models



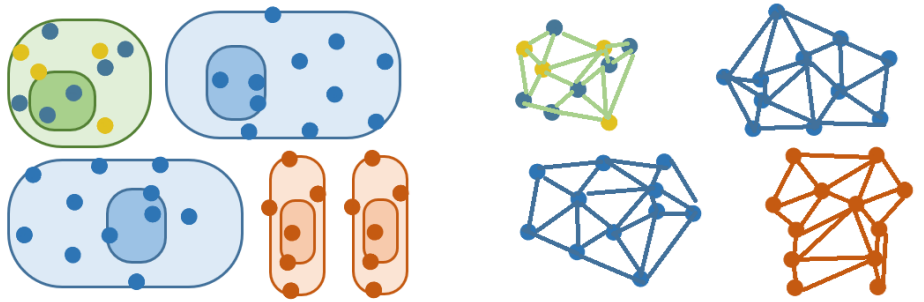
- Density-based models



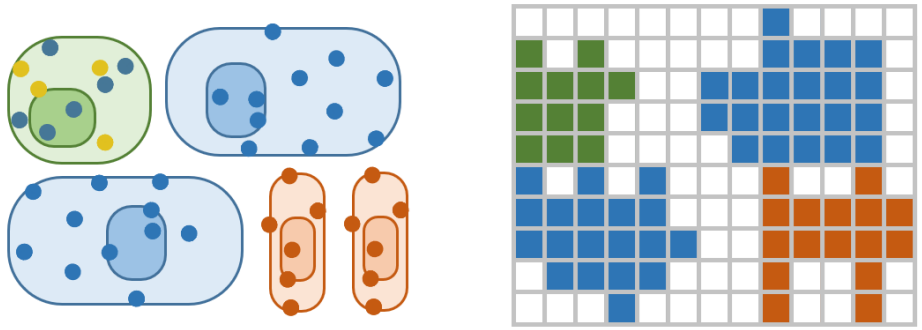
Modelling mRNA distribution

Spatial model – how are mRNA molecules organised when they come from the same cell?

- Graph-based models



- Density-based models



Cell type model – how is spatial gene expression associated to different cell types?

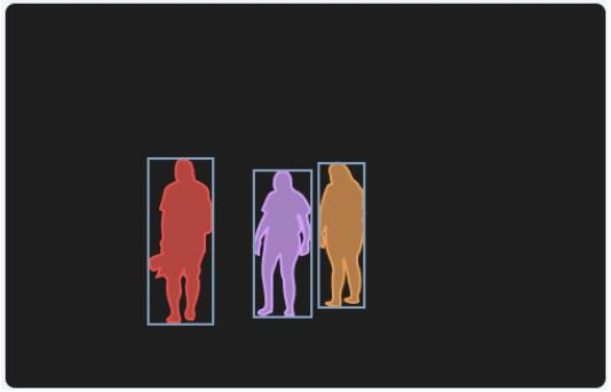
- Prior cell type specific expression signatures e.g. from single cell RNA sequencing data

Cell-segmentation free analysis is still segmentation...

Image

(imagine people = cells)

Instance segmentation



Traditional cell segmentation

+

Graph-based methods

Semantic segmentation

Transcript density-based methods

https://huggingface.co/blog/mask2former

elixir

IBIH Berlin Institute of Health @Charité

Cell segmentation free analysis tools (… there many more!)

- **Graph-based models** (is transcript aggregation/clustering different from cell-segmentation?)
  - spage2vec (Partel and Wählby, FEBS J, 2020)
  - Baysor (Petukhov et al, Nat Biotechnol, 2021)
  - Points2Regions (Andersson et al, Cytometry A, 2024)
- **Density-based models**
  - SSAM (Park et al, Nat Commun, 2021) *
  - SSAM-lite (Tiesmeyer et al, Front Genet, 2022) *
  - FICTURE (Si et al, Nat Methods, 2024)
  - TopACT (Benjamin et al, Nature 2024)
  - SAINSC (Mueller-Boetticher et al, Small Methods, 2024) *
- **Augmented Cell Segmentation methods (using scRNA-seq data to improve segmentation)**
  - Baysor – can work with a DAPI prior
  - pciSeq (Qian et al, Nat Methods, 2019) – Poisson point process + negative binomial
  - JSTA (Littmann et al, MSB, 2021) – joint segmentation and typing applying ML on top of Watershed segmentation
  - Segger (unpublished) – GNN that utilises nucleus segmentation and transcript graphs

Cell segmentation free analysis – pro's and cons

Pros:

- Generally require less computational resources
- Not limited to stains (e.g. red blood cells have no nucleus, so DAPI isn't useful)
- Analysis of measured transcripts

Cons:

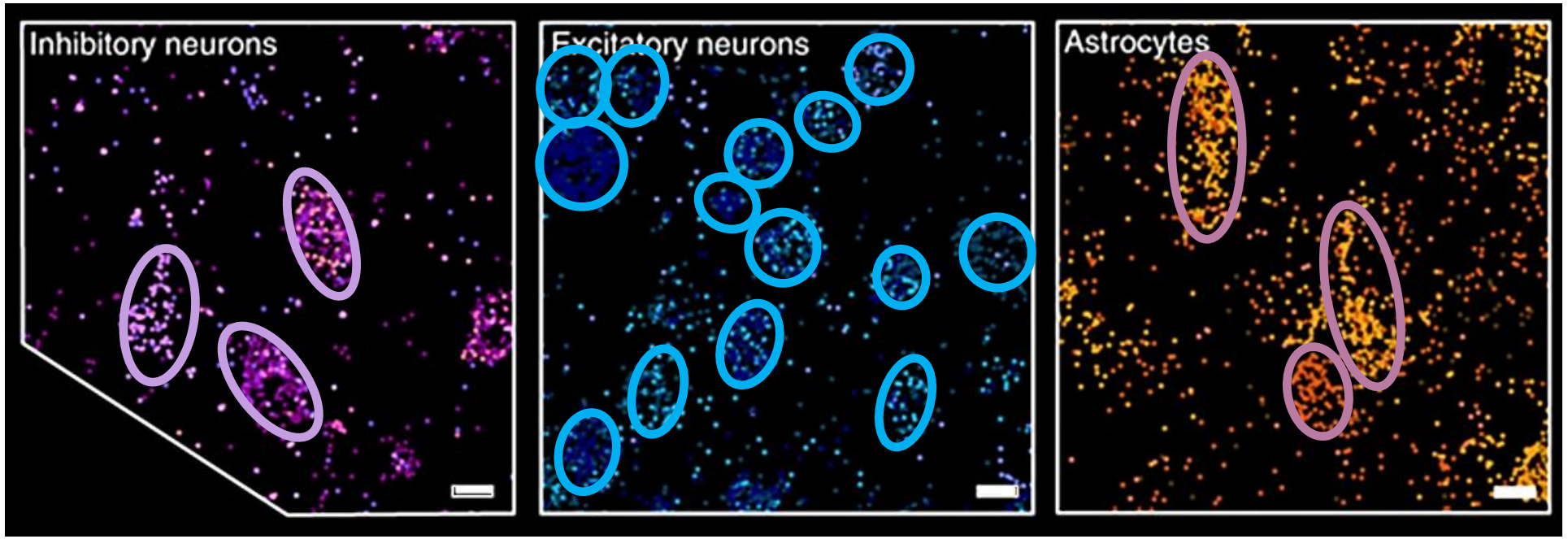
- Cannot identify cells without transcripts (e.g. if a cell-type marker didn't work)
- Conceptual interpretation of results - where are my cells?
- Limited downstream analysis options

Modelling transcript density using SSAM and Sainsc

Park, Jeongbin et al. “Cell segmentation-free inference of cell types from in situ transcriptomics data.” Nature communications vol. 12,1 3545. 10 Jun. 2021, doi:10.1038/s41467-021-23807-4

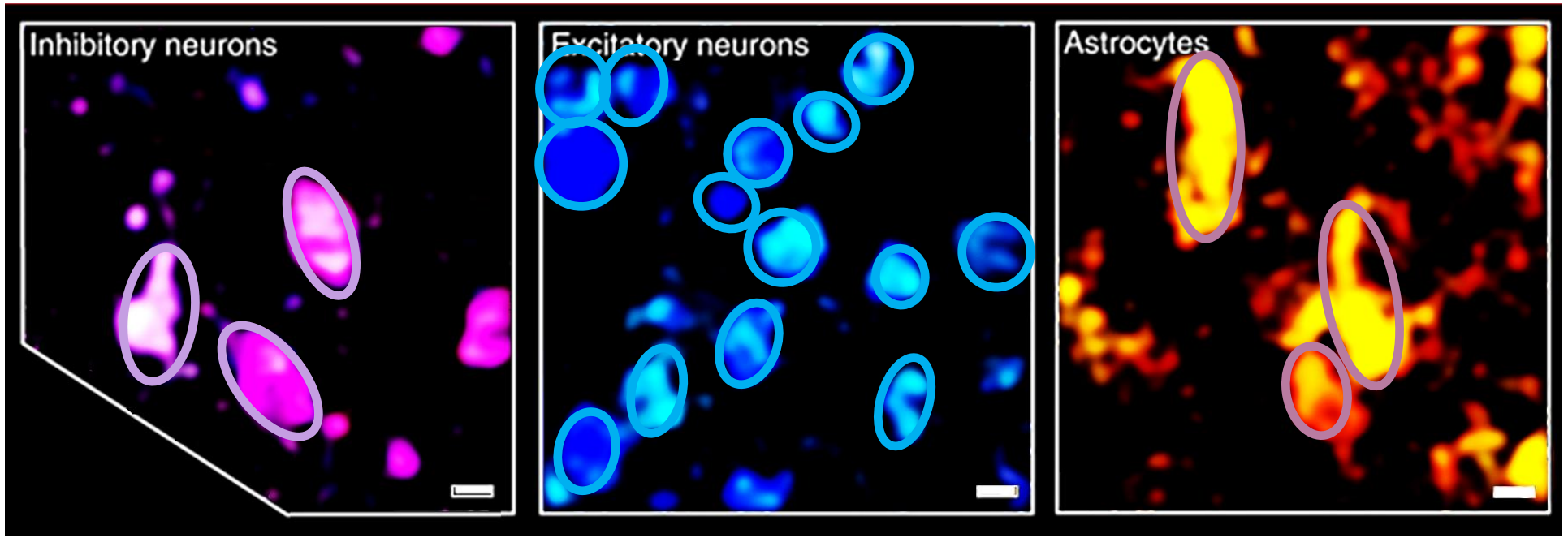
Müller-Bötticher, Niklas et al. “Sainsc: A Computational Tool for Segmentation-Free Analysis of In Situ Capture Data.” Small methods, e2401123. 12 Nov. 2024, doi:10.1002/smtd.202401123

Transcript density indicate likely cell locations



Codeluppi, Borm et al (2018) Nature Methods

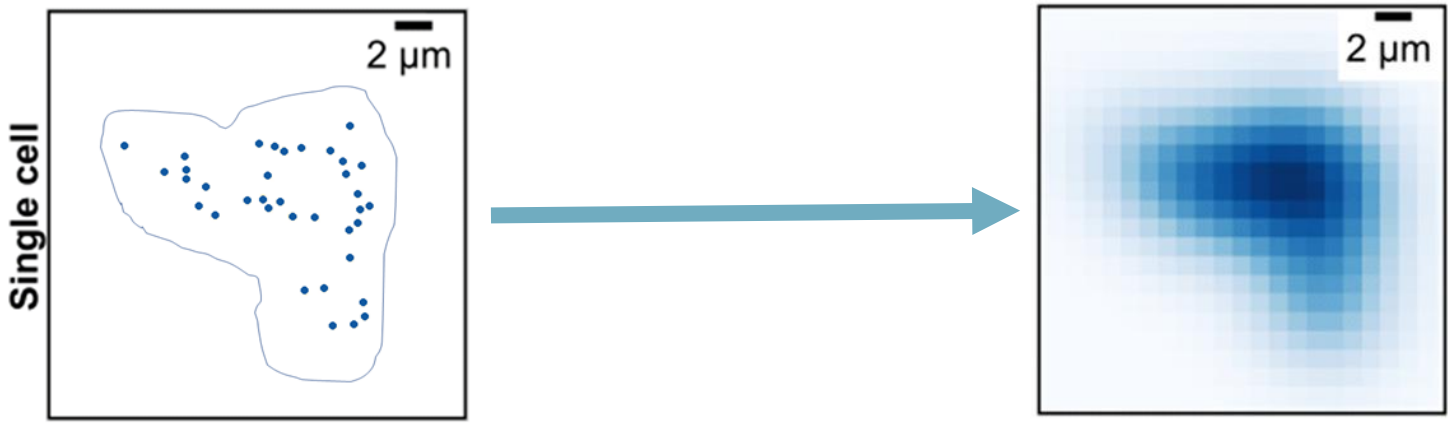
Transcript density indicate likely cell locations



Codeluppi, Borm et al (2018) Nature Methods

SSAM: cell segmentation free analysis of spatial data

Analyse spatial gene expression density, not cells



Park et al (2022) Nature Communication

The SSAM algorithm in 3 steps

1) Smooth gene expression
2) Provide/identify cell type signatures
3) Generate the cell-type map (sematic segmentation)

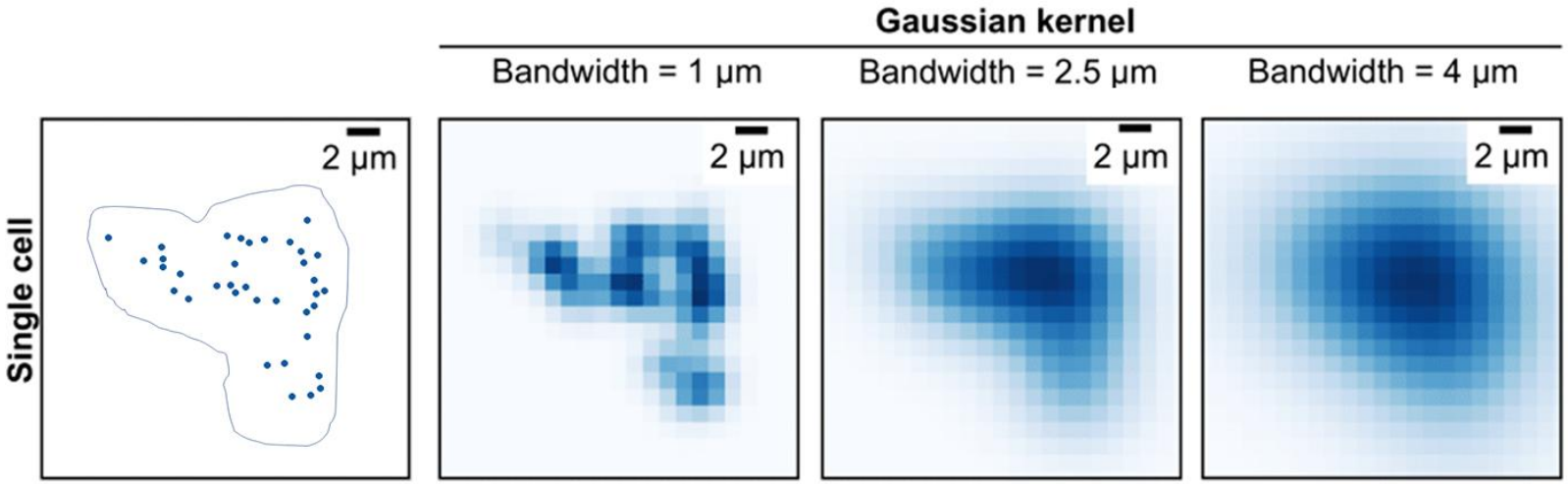
Step 1.1: calculate spatial mRNA density

Apply Kernel Density Estimation (KDE) with Gaussian kernel

Resulting image represents the probability density of mRNA existence

- ***From discrete molecules to cloud of gene expression ("gene expression per pixel")***

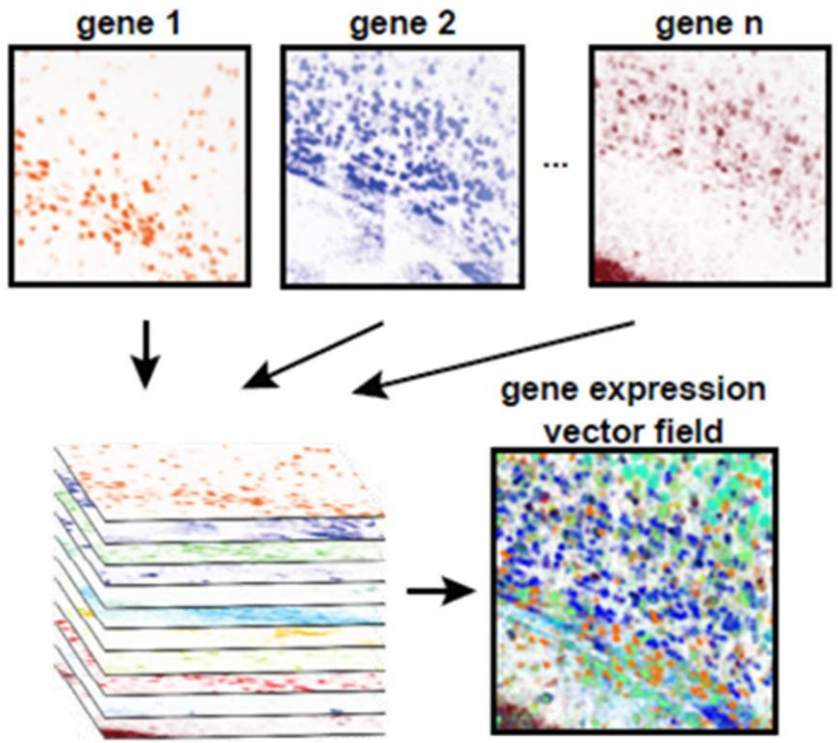
Bandwidth (sigma) should smooth between mRNA, but not outside of cells



Step 1.2: creating the “vector field” of gene expression

Stacking the KDE of each profiled gene creates the gene expression “vector field”

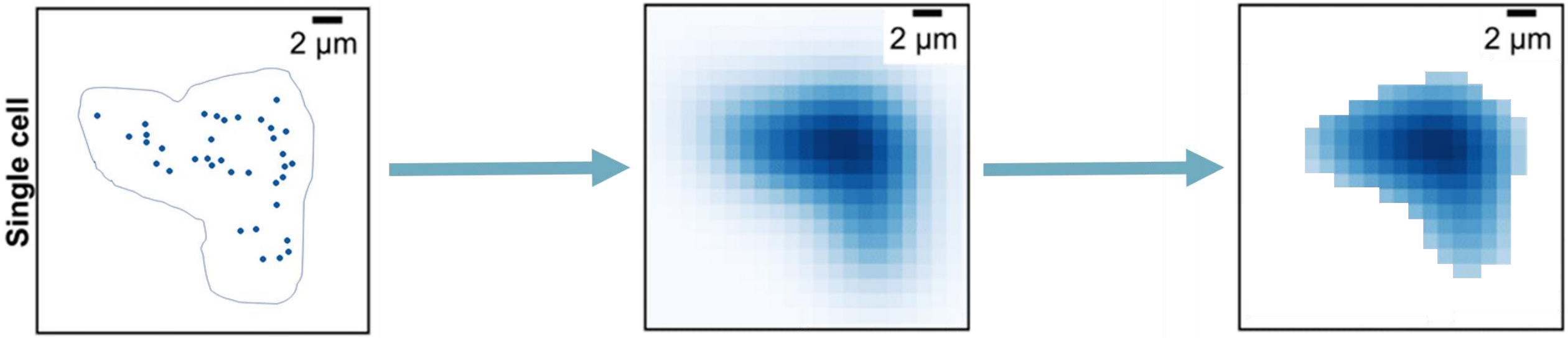
Each pixel in the vector field can be thought to have its own expression profile



Step 1.3: define gene expression threshold

If total gene expression density is too low then the signal likely originates from outside the cell

- Filtering low gene expression regions prevents classification of “low-quality” areas

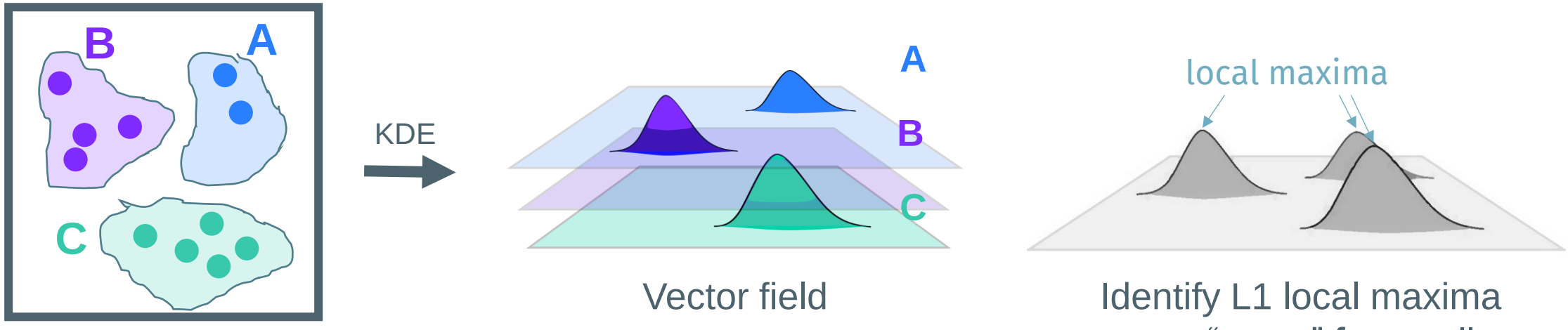


Step 2.1: identify cell-type signatures

For many cell types, cell-type signatures are known: e.g. single-cell RNA sequencing

If cell-type signatures are not known then they need to be computed from the data

Selecting local maxima of mRNA signal as representatives of “cells”



mRNA locations

Vector field

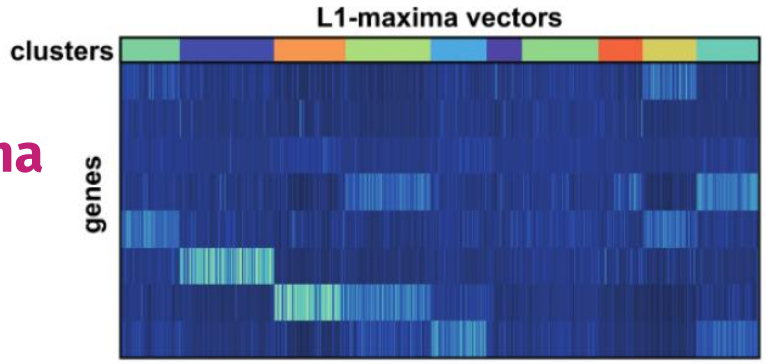
Identify L1 local maxima as a “proxy” for a cell

Tiesmeyer et al (2024) Front. Genet.

elixir

BIH Berlin Institute of Health @Charité

Step 2.2: identify cell-type signatures – cluster local maxima



Cluster gene expression profiles of scRNAseq data or local L1 maxima

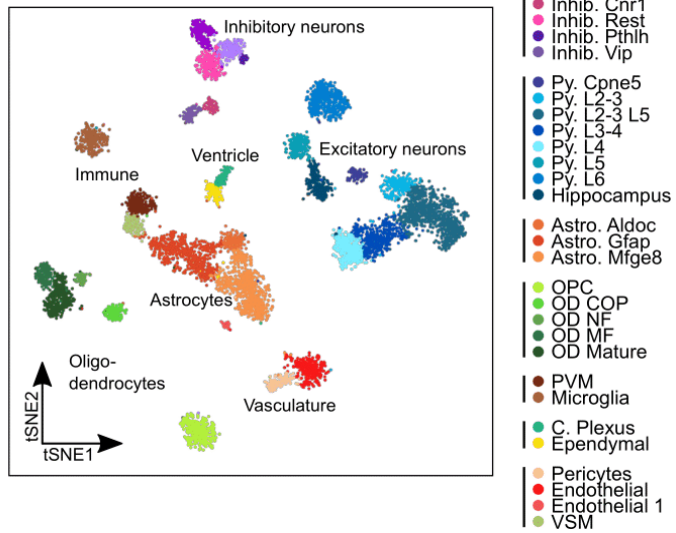
Different cluster = different cell type = different function!

SSAM adopts a Louvain algorithm clustering approach

- This can be exchanged with your favourite clustering method!

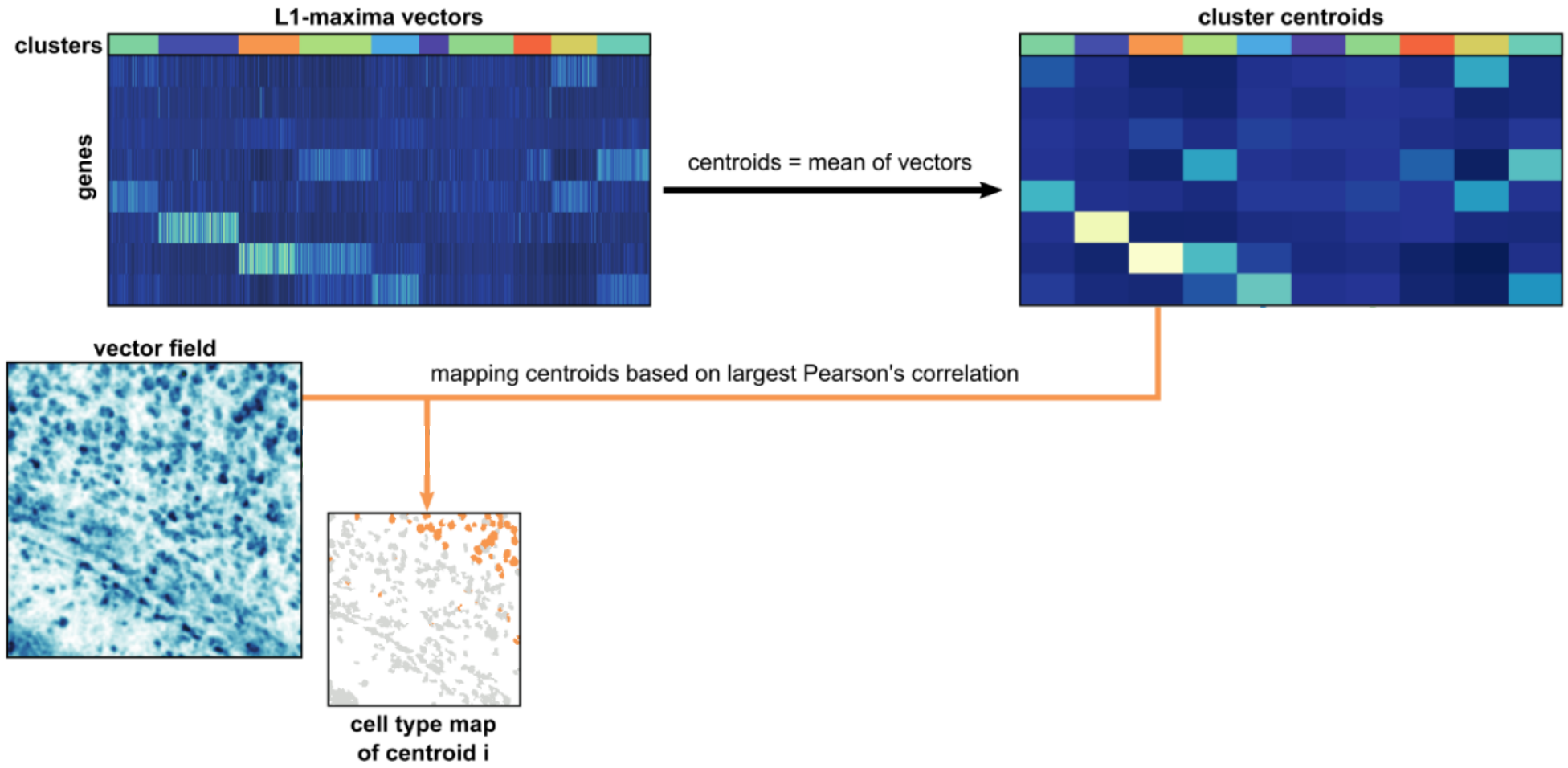
Median cluster expression = cell-type gene expression signature

Visualise using UMAP

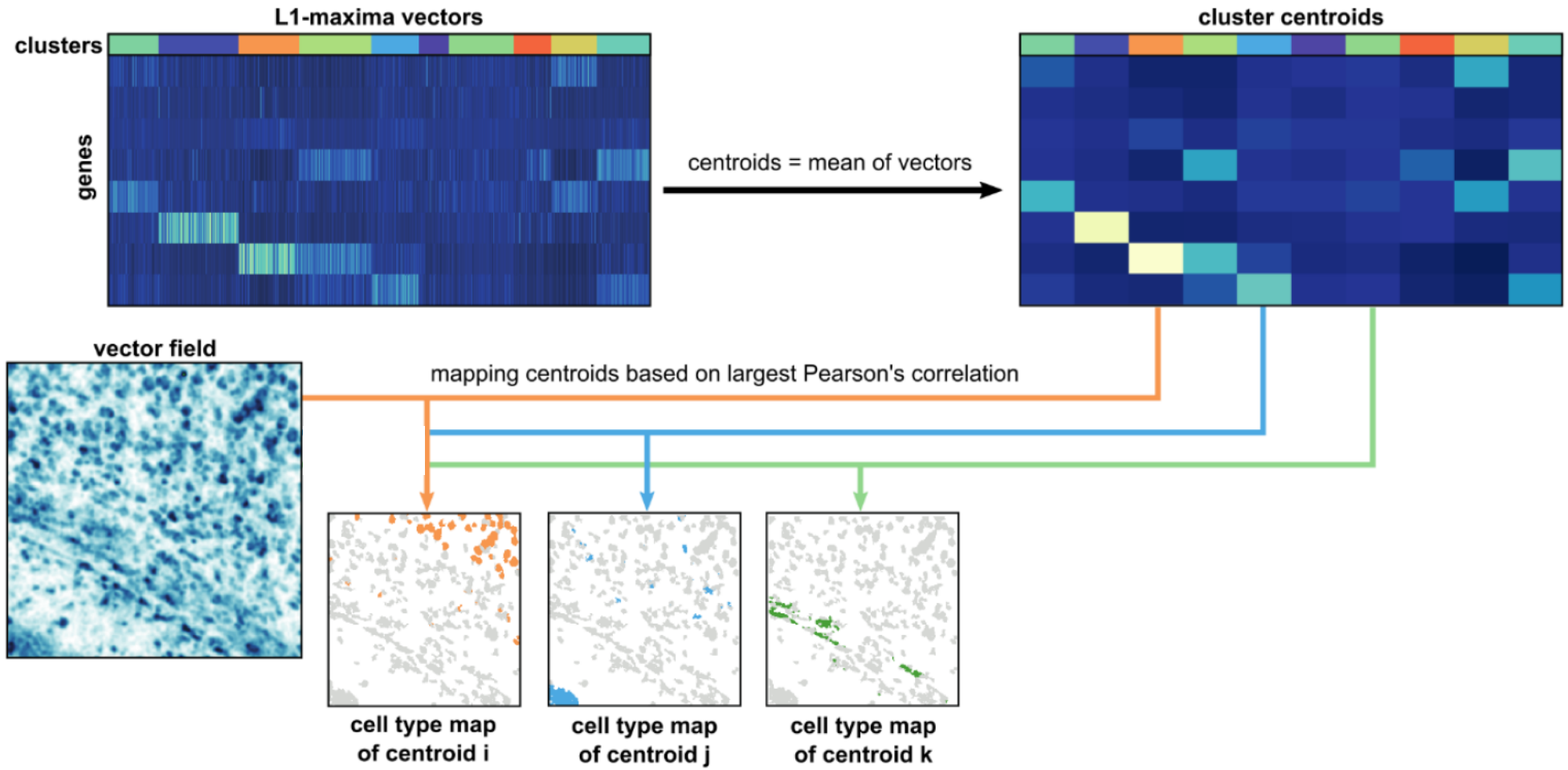


Park et al (2022) Nature Communication

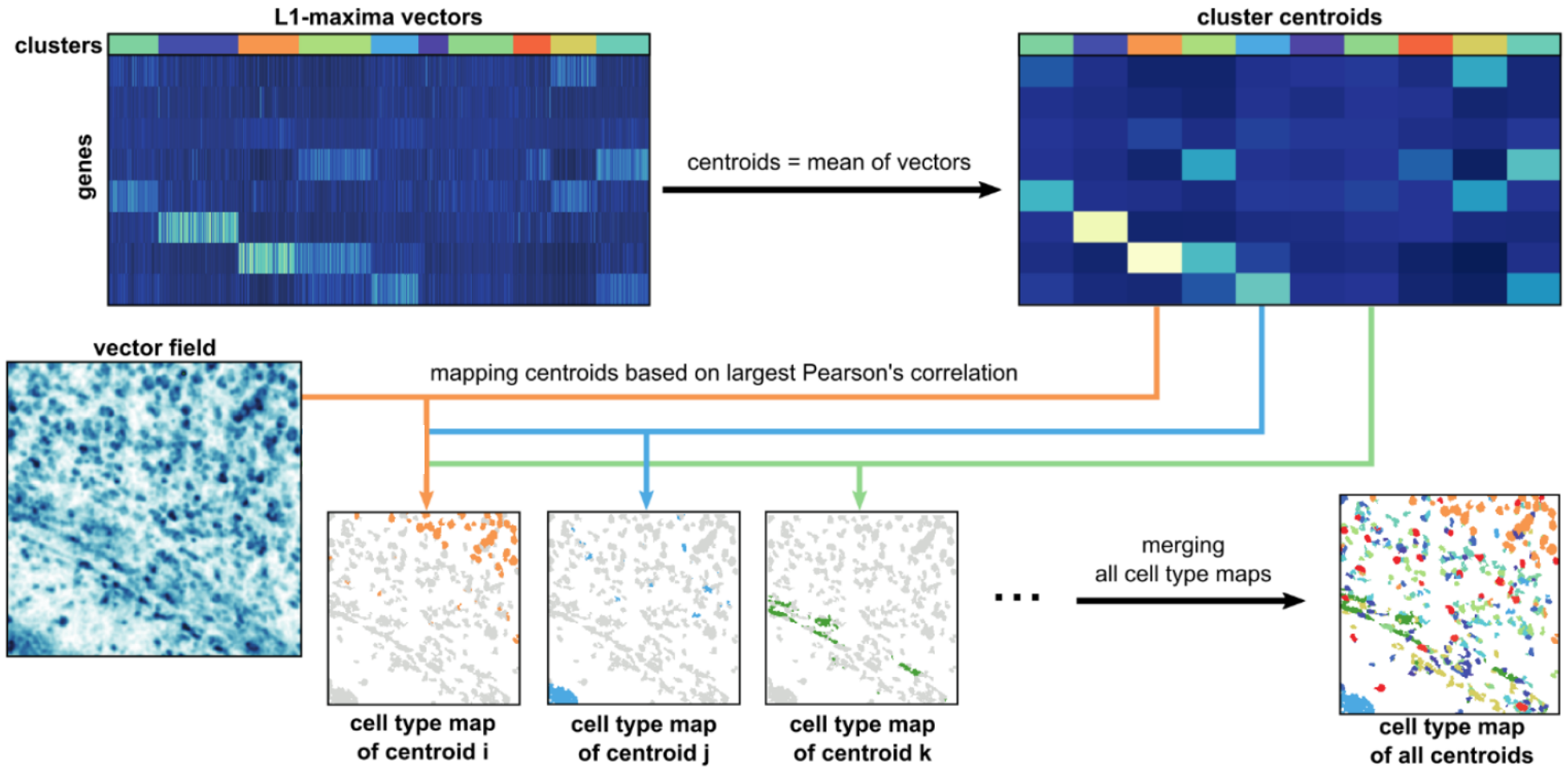
Step 3: generate the cell type map



Step 3: generate the cell type map



Step 3: generate the cell type map



Step 3: generate the cell type map

Classifying pixels

Pixels are classified based on a Pearson correlation

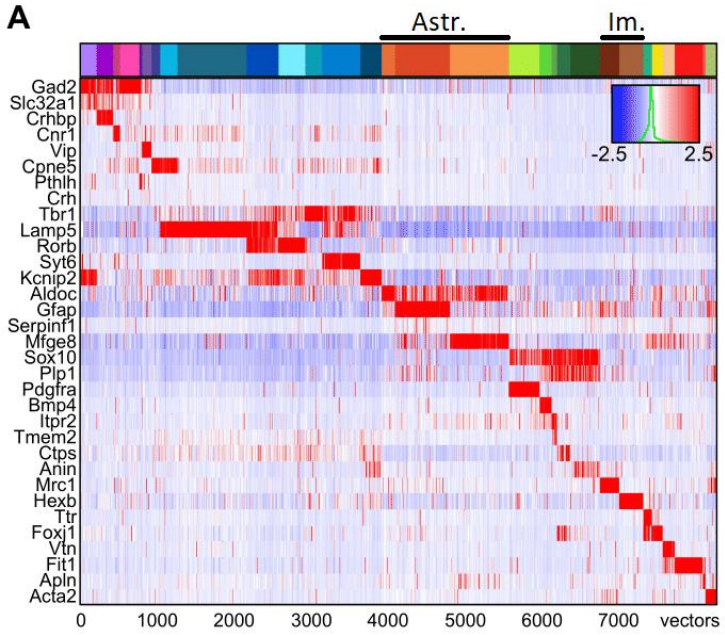
- Pixel gene expression VS cell-type gene expression signature
- Simple but effective
  - Works well when genes are robust cell type markers (i.e. low plex cell typing panels)
  - Doesn’t work well when genes are not cell type specific (i.e. high plex gene panels)
- This step can be exchanged with your favourite ML classification method!

How does SSAM perform?

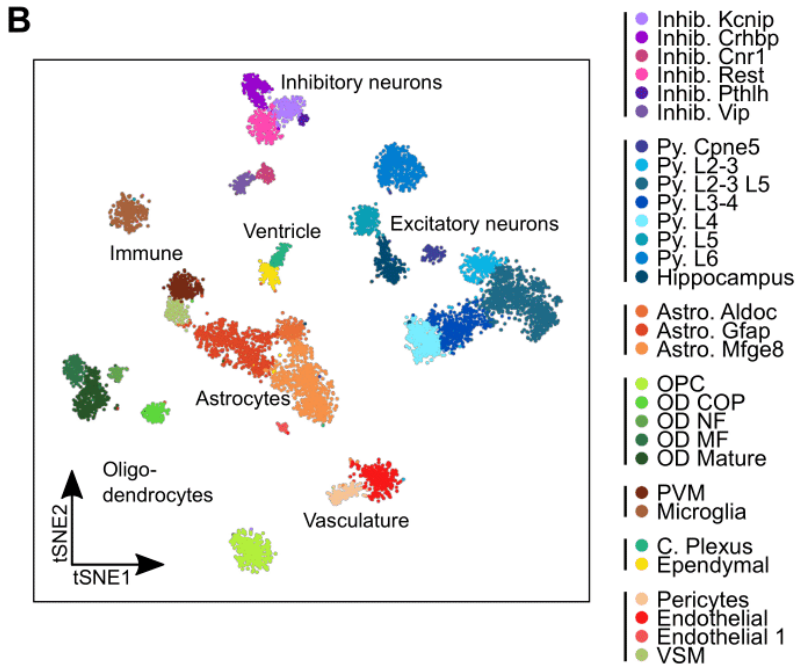
Adult mouse brain somatosensory cortex (SSp)
osmFISH, 35 genes
Codeluppi, Borm et al (2018), Nature Methods

SSAM identifies cell types accurately

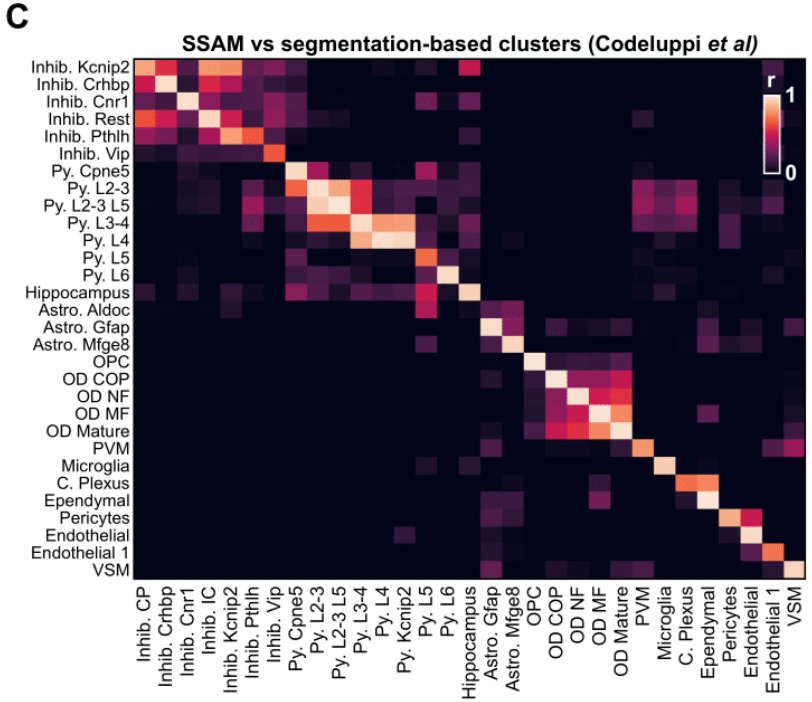
De novo cell-type signatures



Clustering



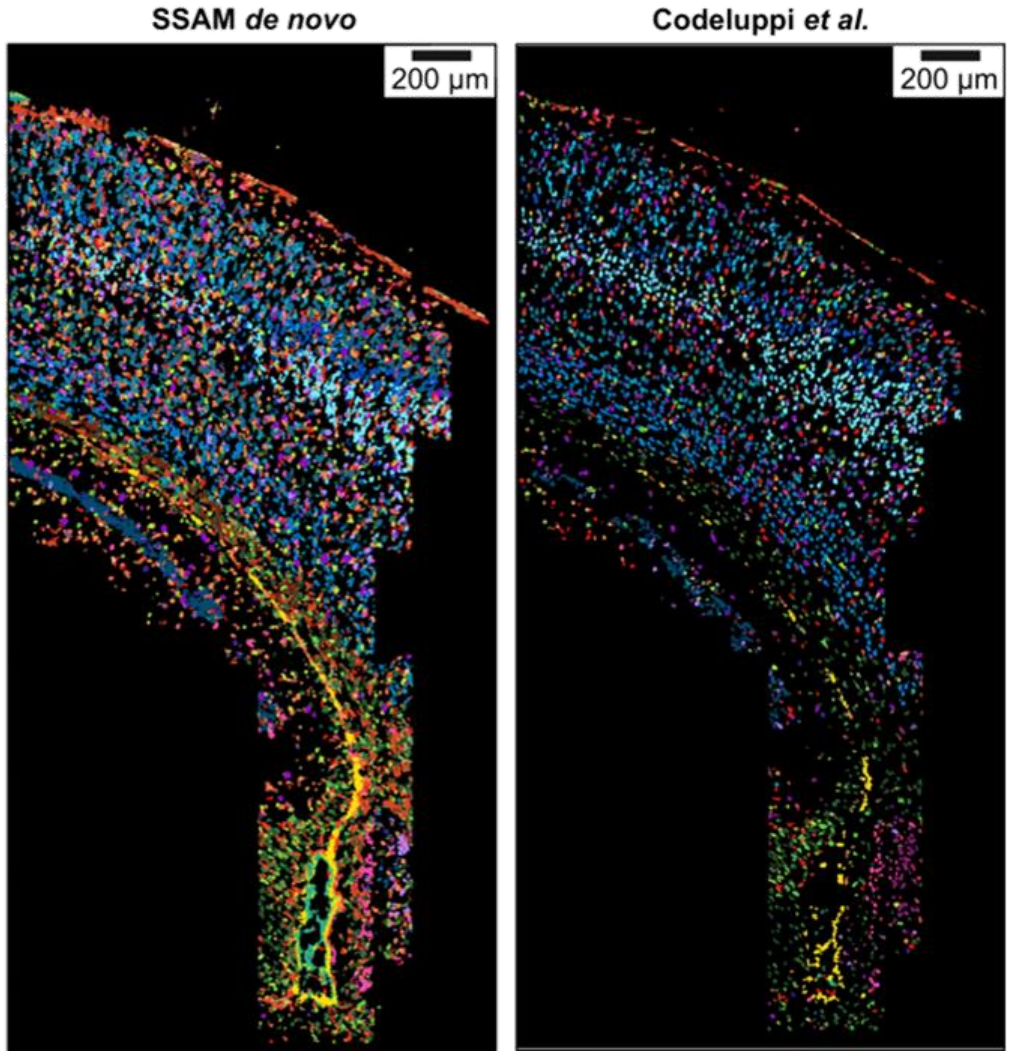
Signature comparison



Park et al (2022) Nature Communication

SSAM reconstructs the mouse SSp cell-type map

SSAM reconstructs mouse brain somatosensory cortex



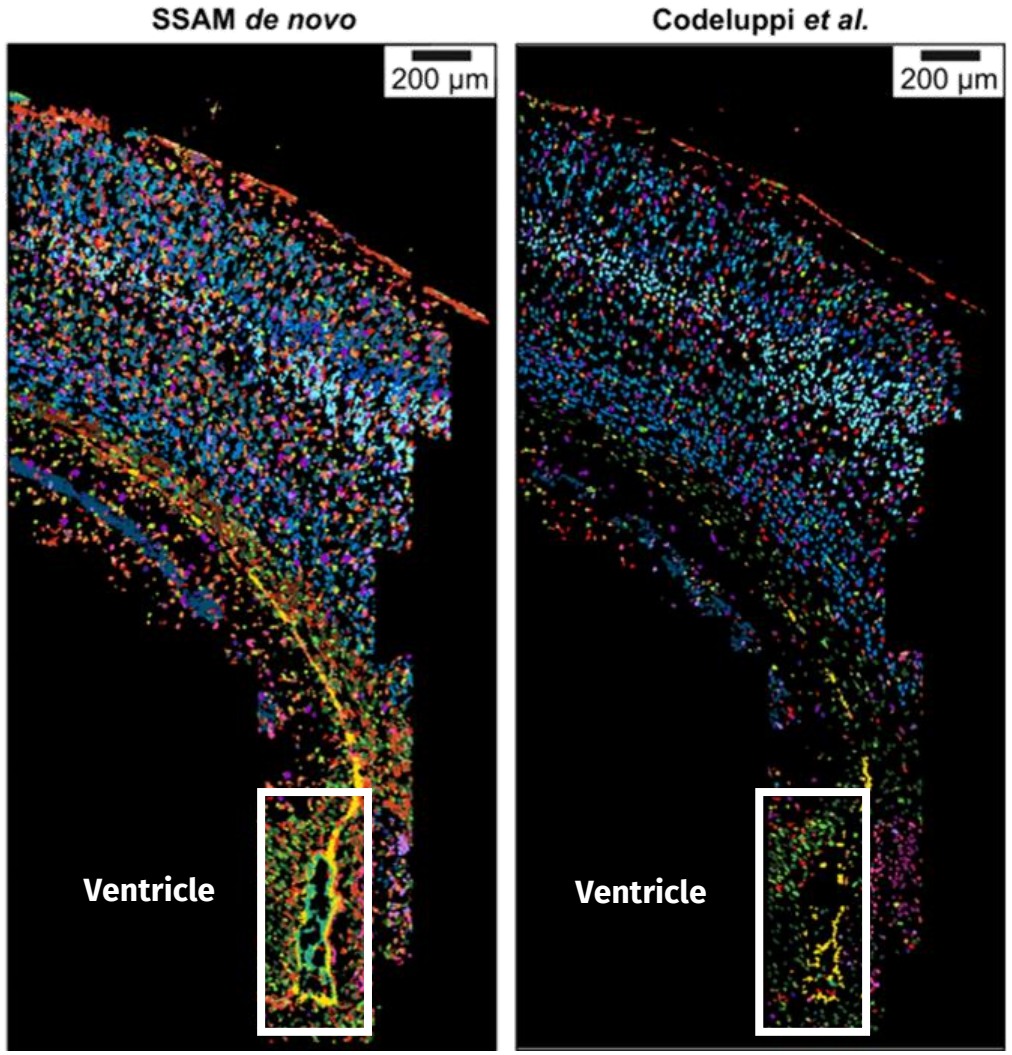
Park et al (2022) Nature Communication

elixir

IBIH Berlin Institute of Health @Charité

SSAM: how well does it work?

SSAM reconstructs mouse brain somatosensory cortex



Park et al (2022) Nature Communication

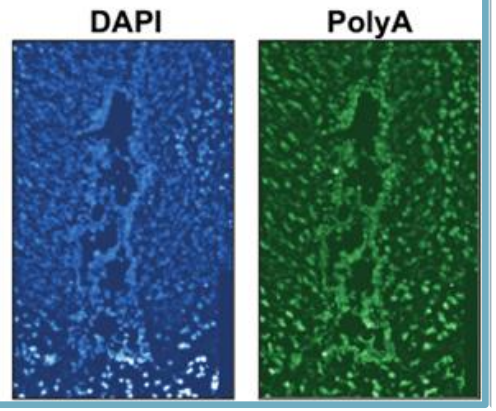
elixir

IBIH Berlin Institute of Health @Charité

SSAM improves mapping of the ventricle region

Problem: low DAPI/Poly-A signal and occlusion

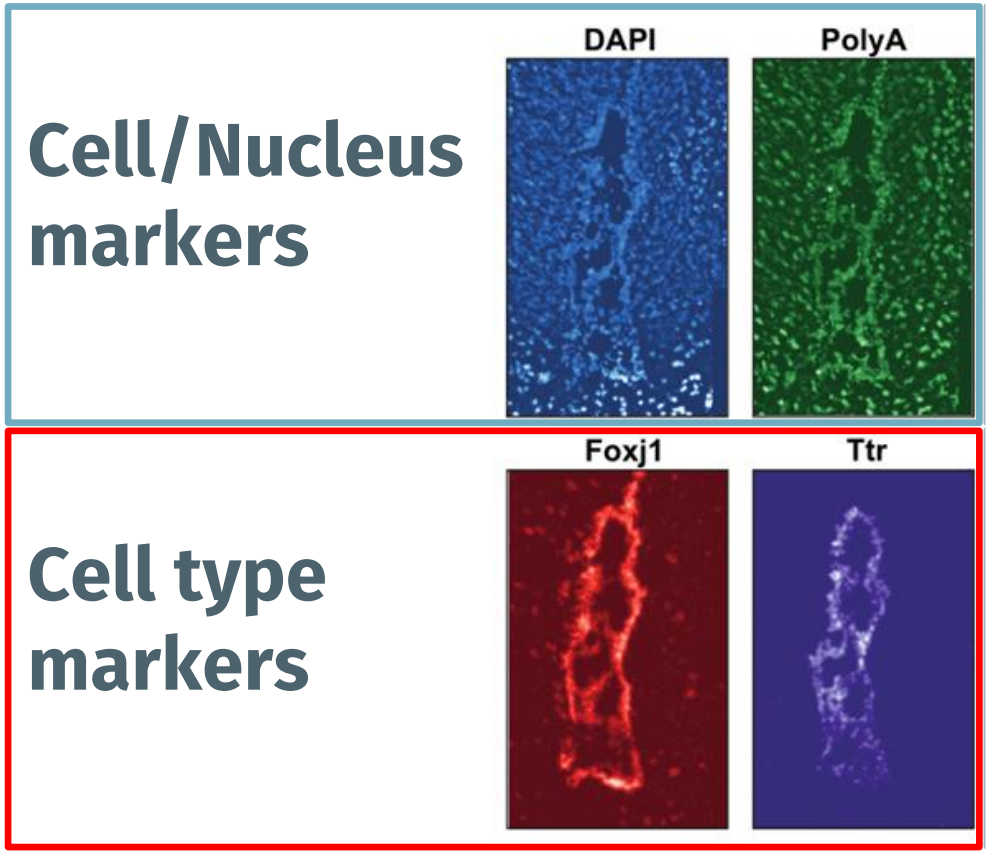
Cell/Nucleus markers



Park et al (2022) Nature Communication

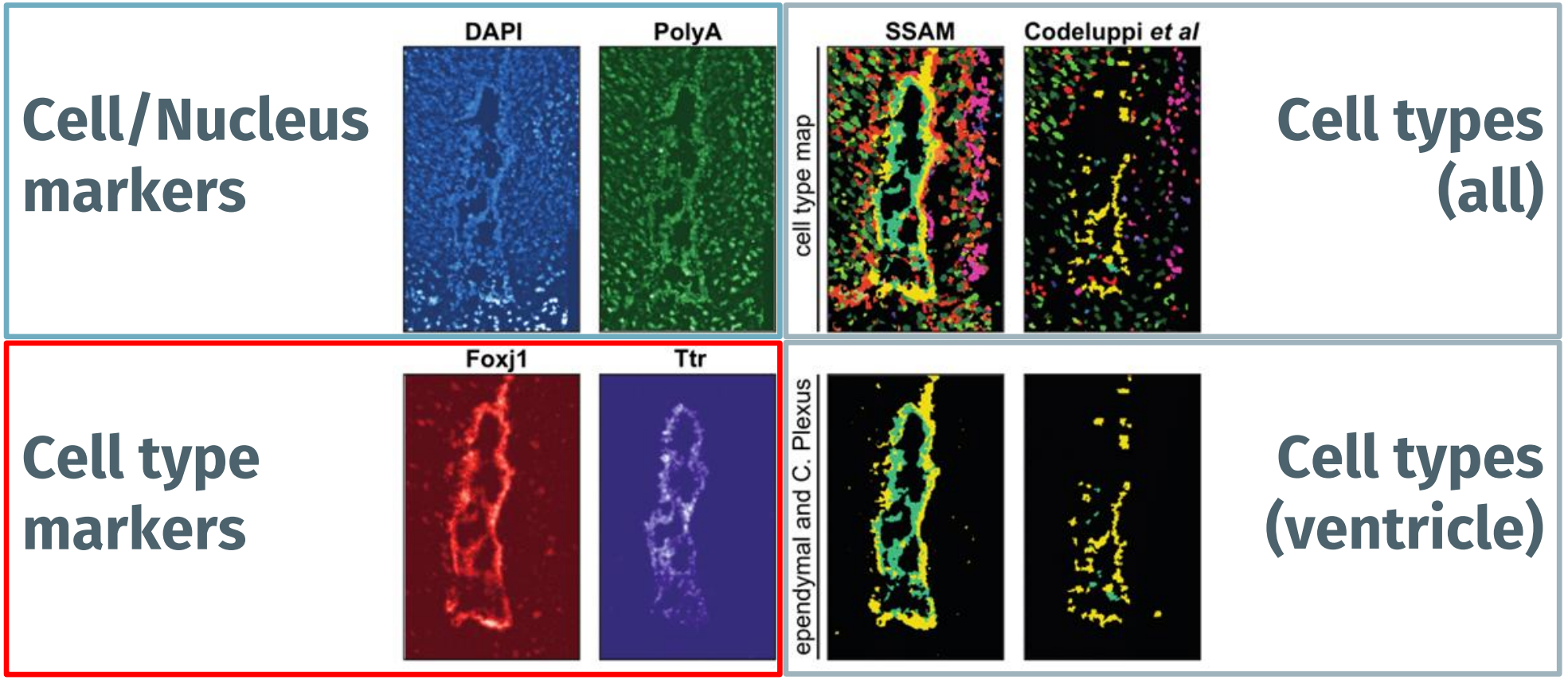
SSAM improves mapping of the ventricle region

Problem: low DAPI/Poly-A signal and occlusion, but high marker gene expression



SSAM improves mapping of the ventricle region

Problem: low DAPI/Poly-A signal and occlusion



Emerging state of the art for spatial transcriptomics

Size and resolution

Pseudo 3D

Real 3D

Multi-omics

Liu, Yang, Deng et al. Cell 183, 1665-1681 (2020); Chen et al, Cell 185, 1777-1792 (2022); Fang et al, eLife12:RP90029 (2023);
Yao, van Velthoven, Kunst et al. Nature 624, 317–332 (2023); Mueller-Boetticher et al. bioRxiv (2024)

Sainsc: optimising SSAM for millions of cells & organism-scale

Segmentation-free Analysis of IN Situ Capture data

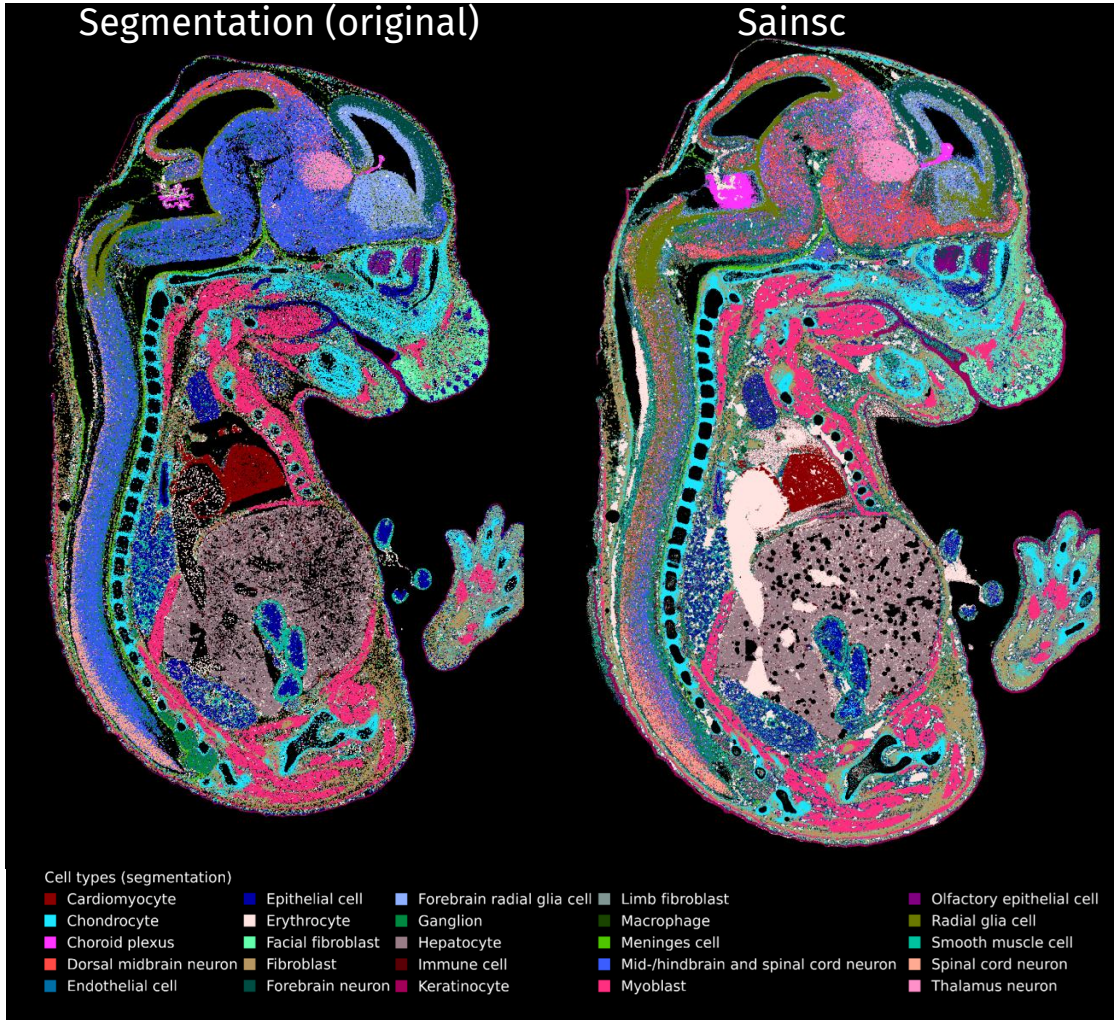
- Segmentation-free identifies red blood cells in the spleen and umbilical cord

Optimisation for organism scale analysis

- 10,000 times faster than SSAM
- 100 times less memory usage than SSAM

Suitable for imaging and sequencing-based spatial transcriptomics

- E.g. Stereo-seq, Open-ST, Nova-ST, VisiumHD



Mueller-Boetticher et al (2024) Small Methods

Sainsc: optimising SSAM for millions of cells & organism-scale

Segmentation-free Analysis of IN Situ Capture data

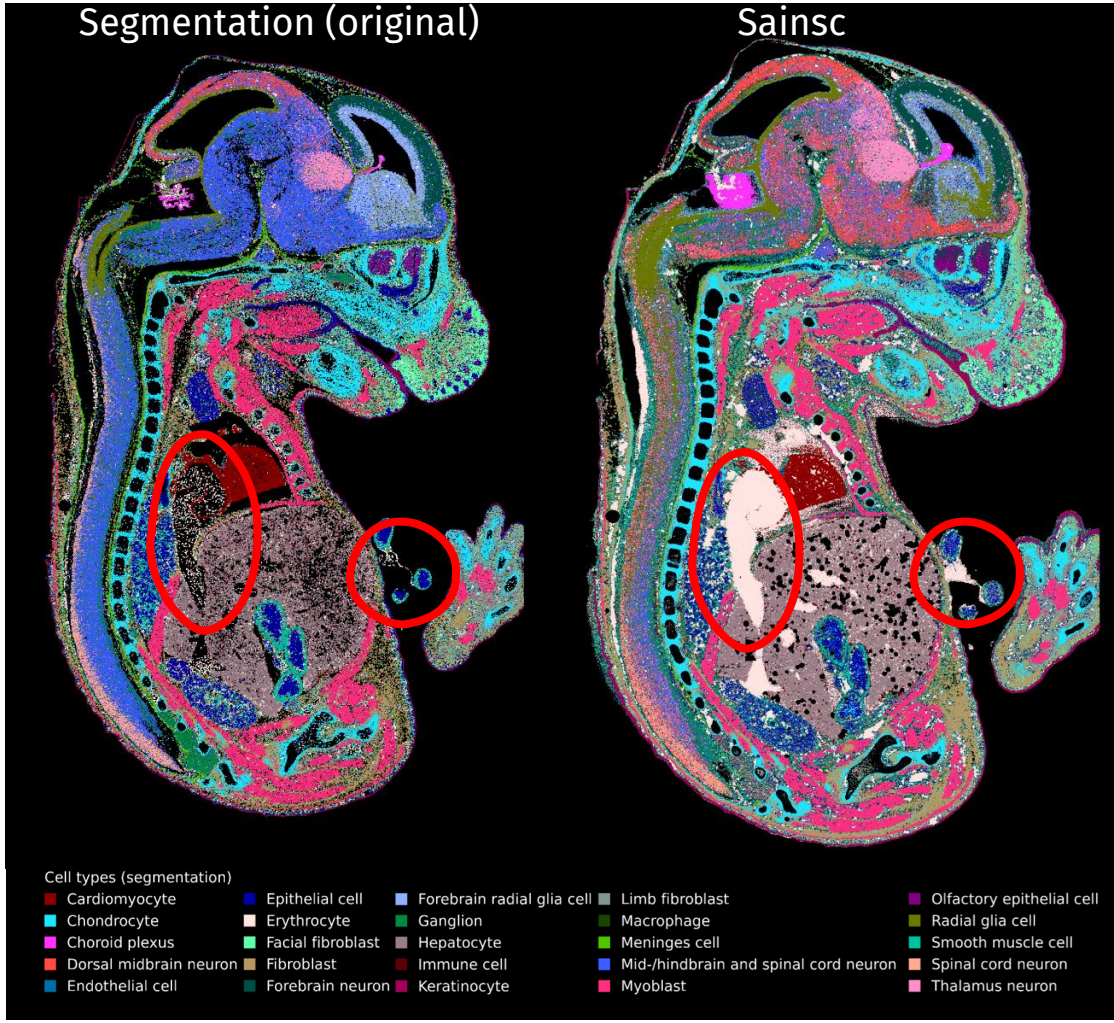
- Segmentation-free identifies red blood cells in the spleen and umbilical cord

Optimisation for organism scale analysis

- 10,000 times faster than SSAM
- 100 times less memory usage than SSAM

Suitable for imaging and sequencing-based spatial transcriptomics

- E.g. Stereo-seq, Open-ST, Nova-ST, VisiumHD



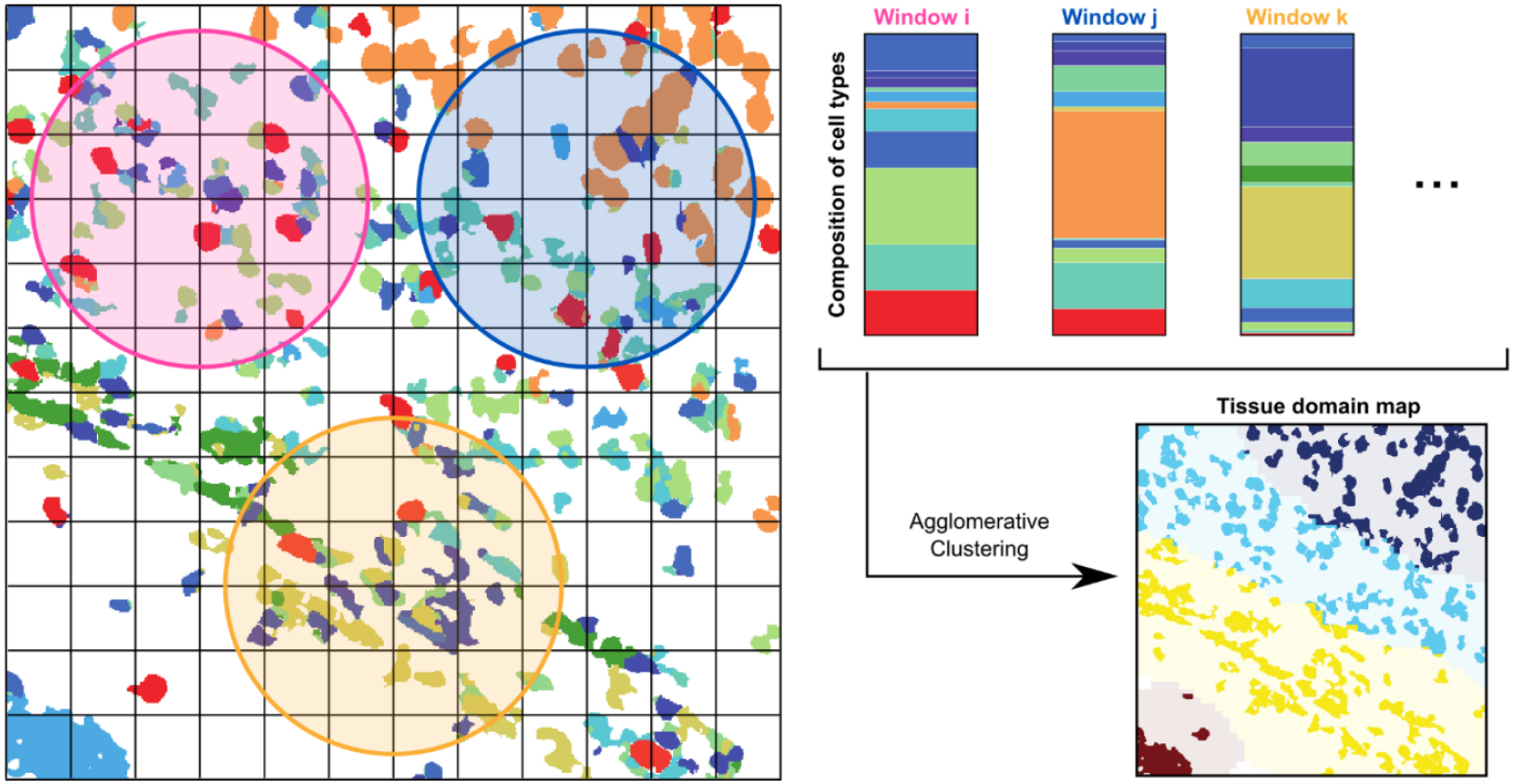
Mueller-Boetticher et al (2024) Small Methods

elixir

BIH Berlin Institute of Health @Charité

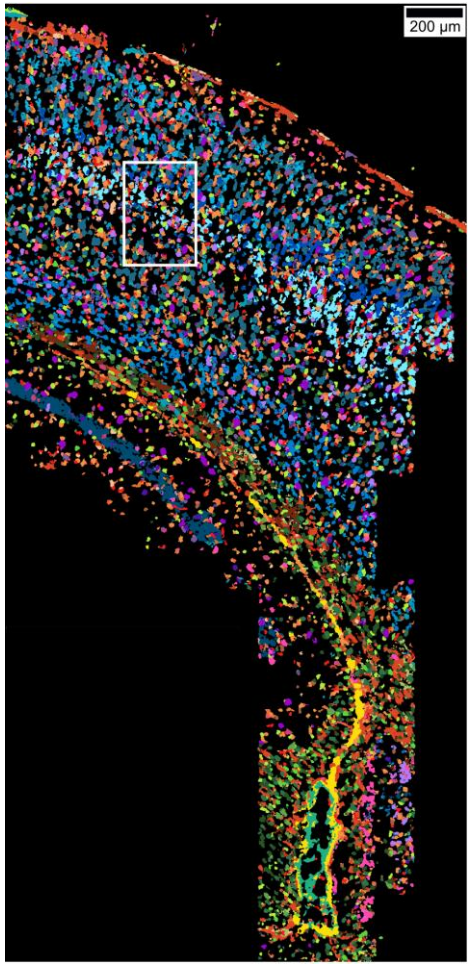
But what can you do without cells?

Downstream analysis – spatial domains

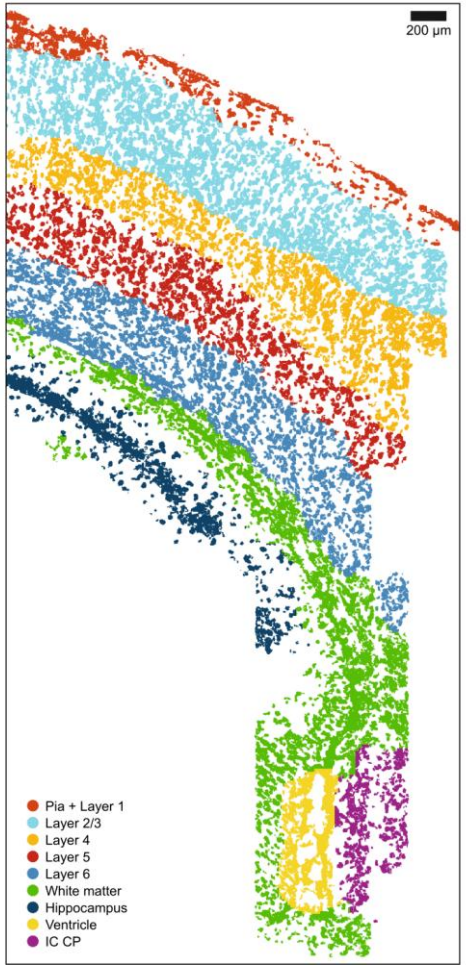


SSAM identified mouse SSp cortical layers

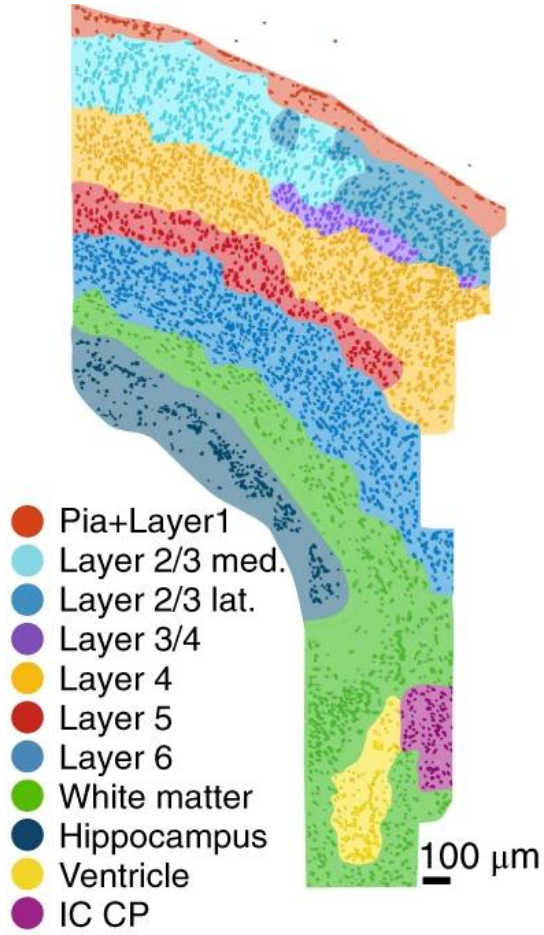
SSAM Cell type map



SSAM Domain map

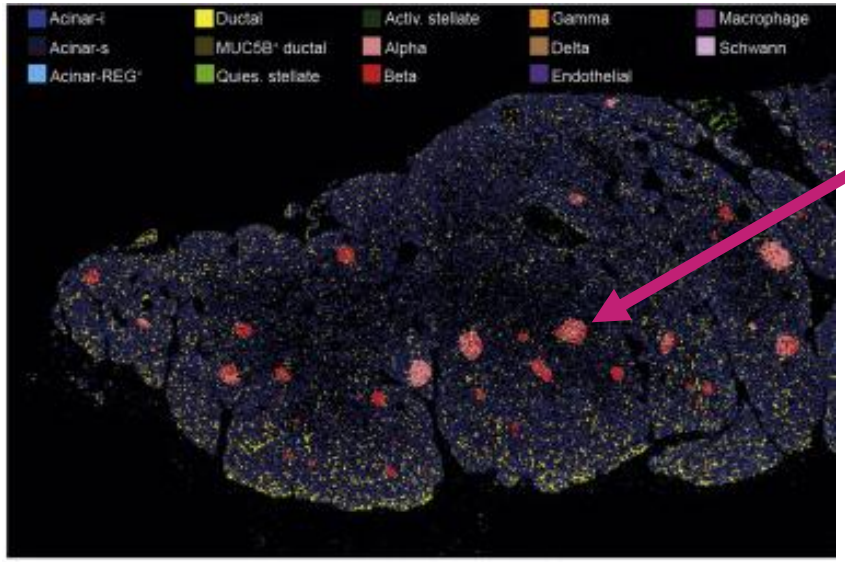


Original



CellSonar: modelling of spatial relationships between cell types

- **Statistical modelling of spatial relationships in the pancreas**

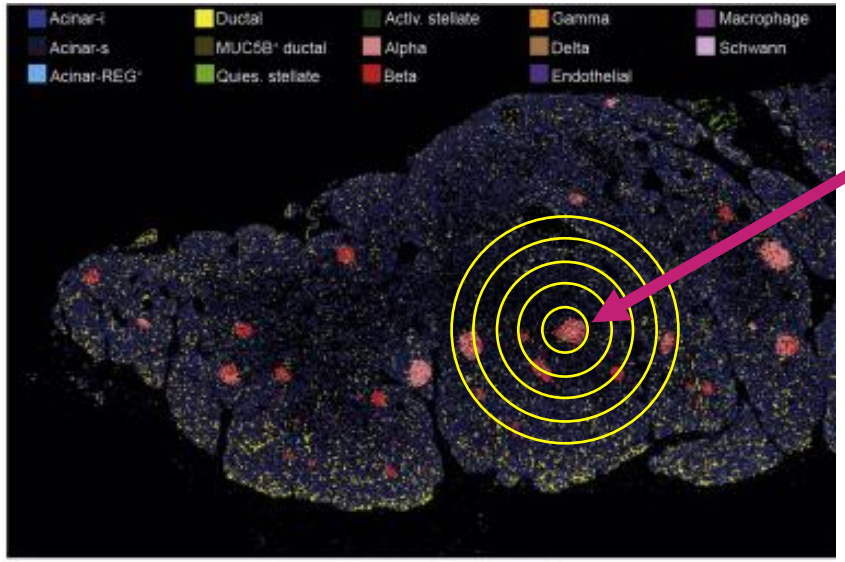


Pancreatic islet

Tosti et al (2021) Gastroenterology

CellSonar: modelling of spatial relationships between cell types

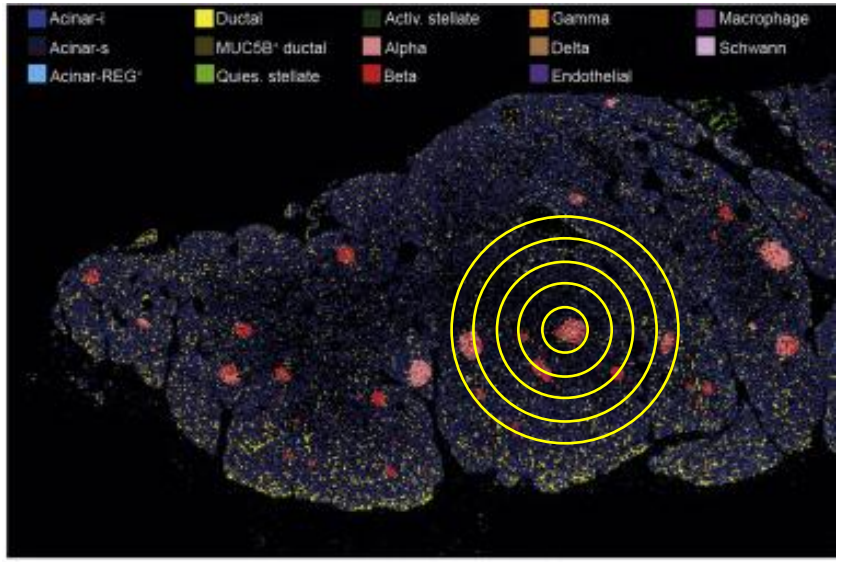
- **Statistical modelling of spatial relationships in the pancreas**

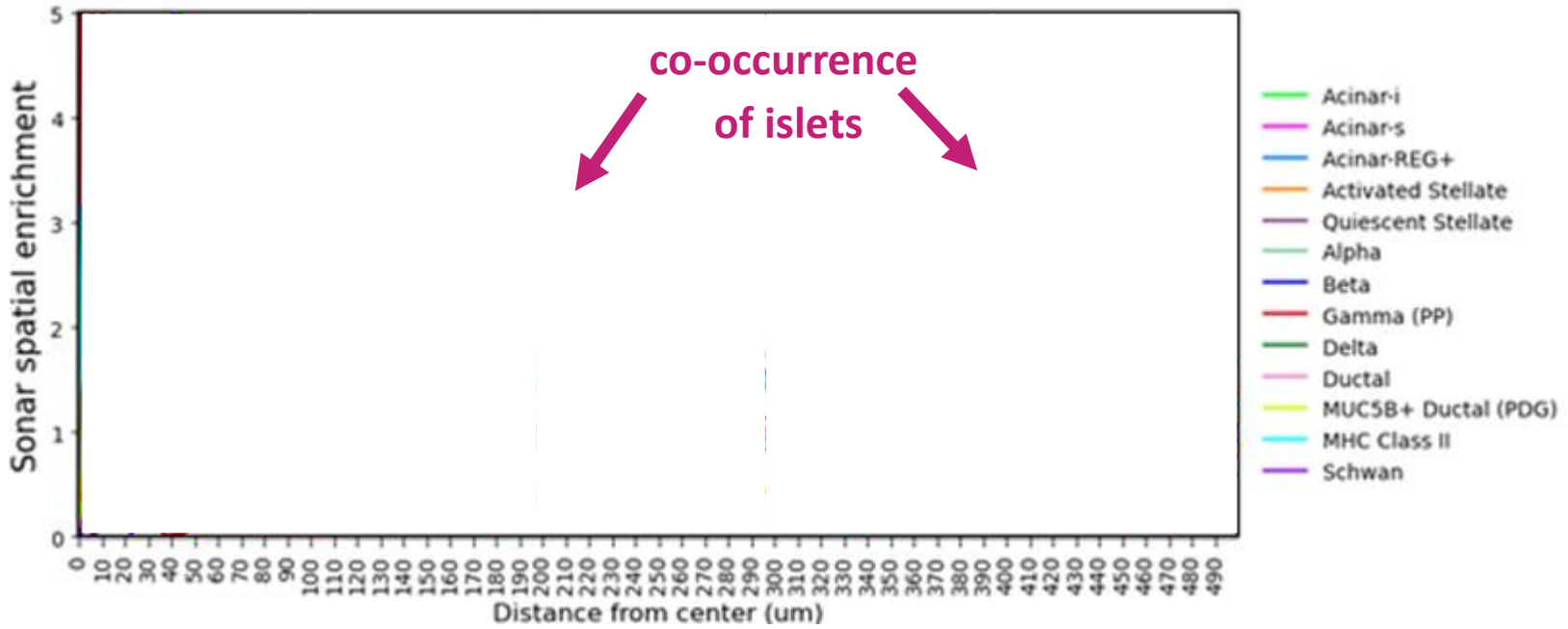


Tosti et al (2021) Gastroenterology

CellSonar: modelling of spatial relationships between cell types

- **Statistical modelling of spatial relationships in the pancreas**

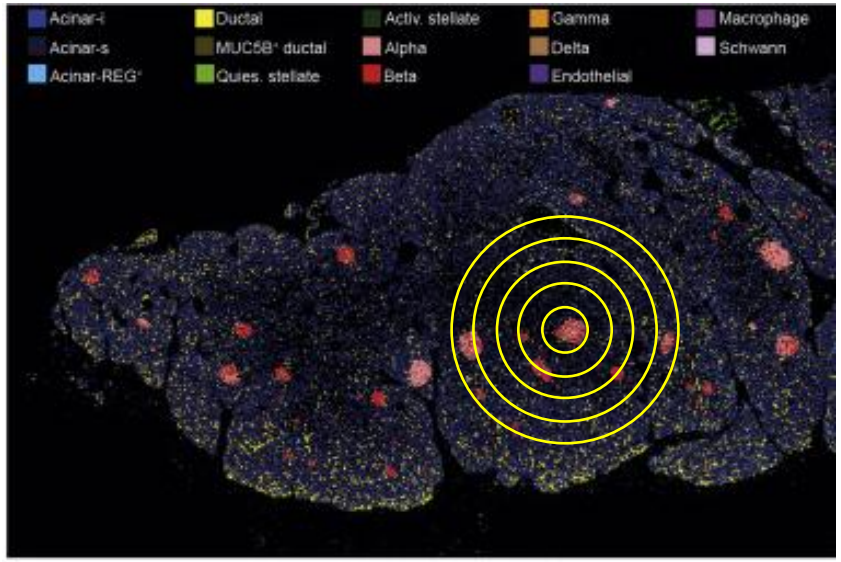


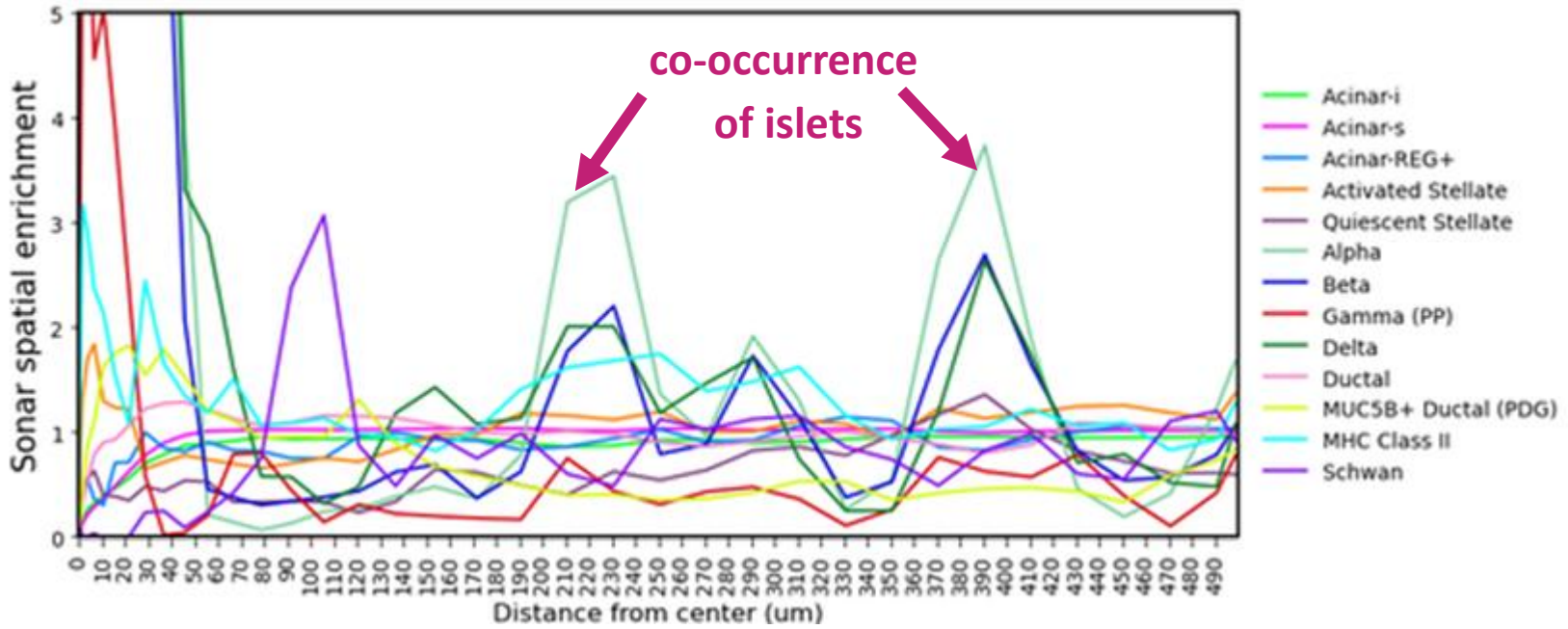


Tosti et al (2021) Gastroenterology

CellSonar: modelling of spatial relationships between cell types

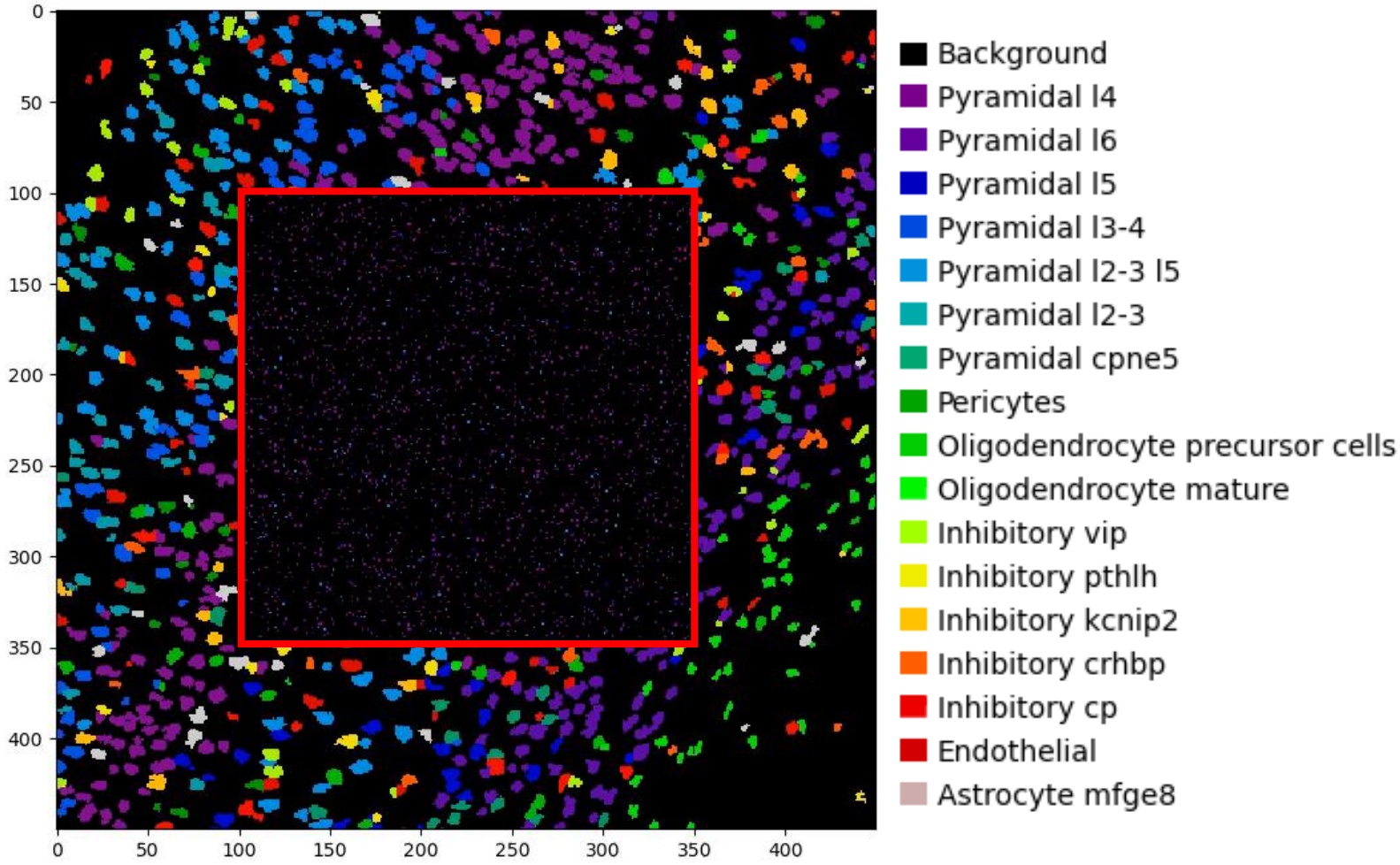
- **Statistical modelling of spatial relationships in the pancreas**



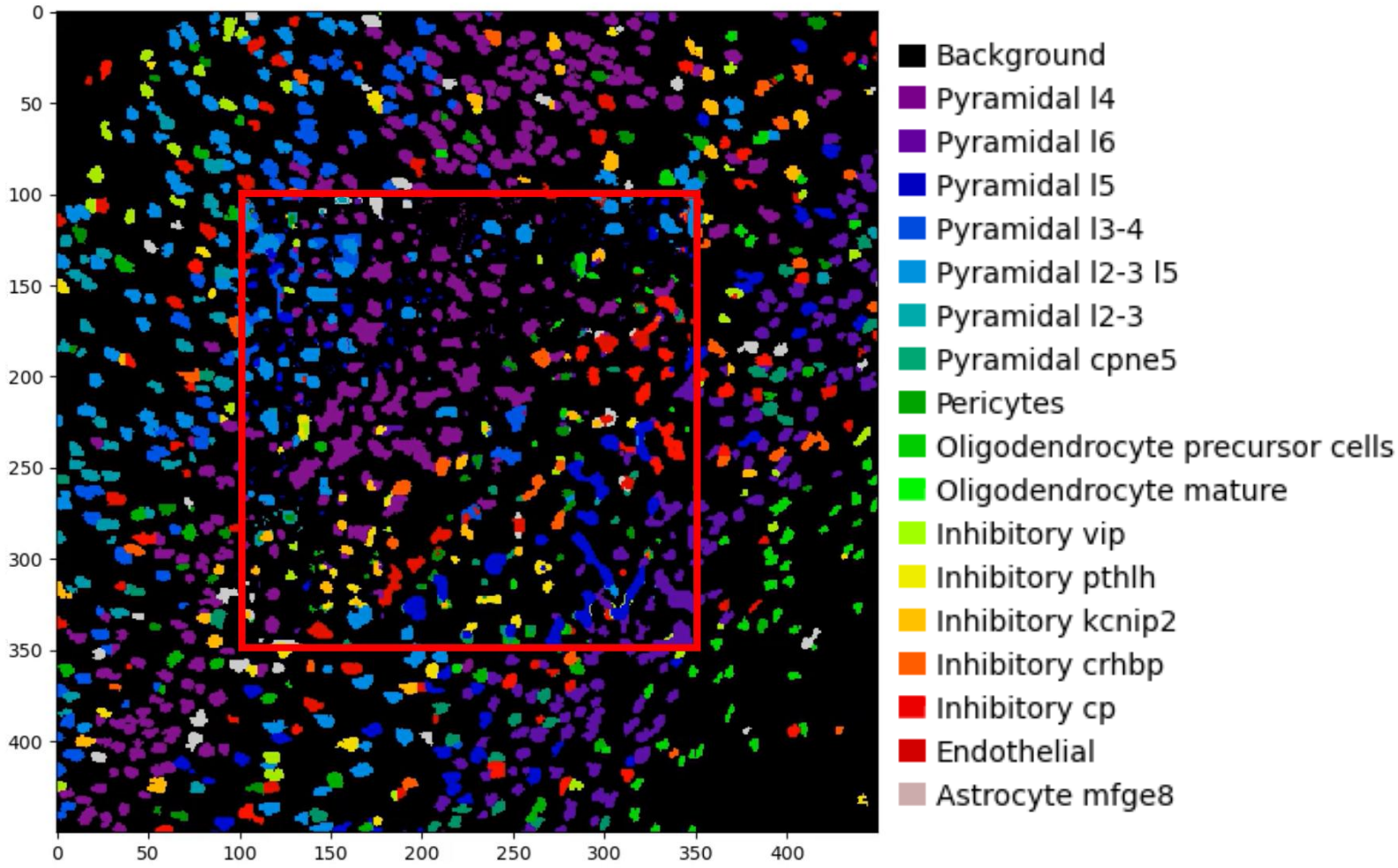


Tosti et al (2021) Gastroenterology

CellSonar: generative capabilities (click play!)



CellSonar: generative capabilities (click play!)

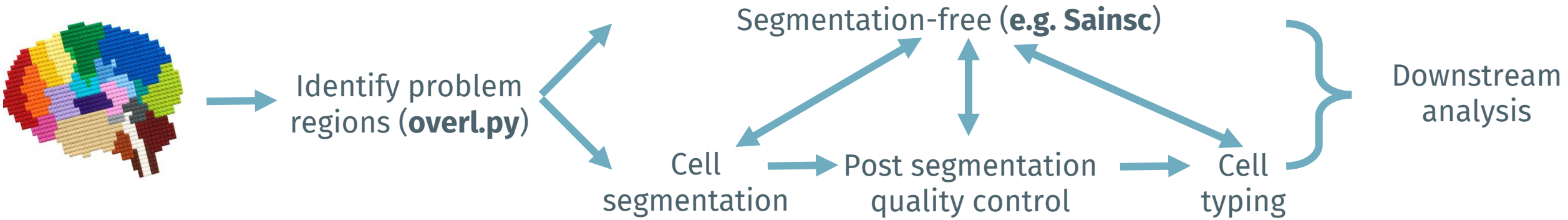


Summary

Spatial transcriptomics goes beyond single cells

- Early (bad) cell segmentation can lead to inaccuracies and missing important signal
- Cell-segmentation free approaches are powerful parallel analysis avenues

Proposed workflow



Tutorial

1. Learn to use the Sainsc tool
2. Analyse a Xenium dataset of a mouse brain coronal section
3. Identify cell type gene expression patterns
4. Define minimal gene expression thresholds
5. Create a cell-type map

Acknowledgements

The patients and their families

BIH Centre of Digital Health, Charite, Berlin
- Roland Eils
- Sebastian Tiesemeyer
- Jeongbin Park (Pusan National University)
- Shashwat Sahay
- Niklas Mueller-Boetticher
- Sebastian Mackowiak
- Tobias Graf

ECRC, Charite, Berlin
- Olivia Nonn
- Florian Herse
- Alikerim Secener

Med Uni Graz, Graz
- Amin El-Heliebi
- Katja Sallinger

SciLifeLab, Sweden
- Sergio Marco Salas
- Paulo Czarnewski
- Saga Helgadottir
- Christoffer Langseth
- Katarina Tiklova
- Maria Chatzinikolaou
- Christophe Avenel
- Axel Andersson
- Carolina Wählby
- Mats Nilsson

DKFZ, Heidelberg
- Matthias Schlesner
- Florian Heyl
- Oliver Stegle

Linnarsson Lab, Karolinska Inst.
- Lars Borm
- Simone Codeluppi

INAB/CERTH, Greece
- Georgios Gavriilidis
- Fotis Psomopoulos

Allen Brain Inst, Seattle
- Brian Long
- Meghan Turner
- Ed Lein
- Bosiljka Tasic

Helmholtz Munich
- Louis Kummerle
- Malte Luecken
- Fabian Theis

RWTH Aachen
- Paul Kiessling
- Christoph Kuppe

CZI SpaceTX Consortium

Bonus material!

`ovrl.py` – a tool to identify overlapping cells in imaging-based spatial transcriptomics data

2D, or not 2D? Investigating Vertical Signal Integrity of Tissue Slices

Sebastian Tiesmeyer, Niklas Müller-Bötticher, Alexander Malt, Brian Long, Sergio Marco-Salas, Paul Kiessling, Paul Horn, Adrien Guillot, Louis B Kuemmerle, Leyao Ma, Frank Tacke, Fabian Theis, Christoph Kuppe, Mats Nillson, Roland Eils, Naveed Ishaque

doi: https://doi.org/10.1101/2025.01.13.632601

This article is a preprint and has not been certified by peer review [what does this mean?].

Abstract | Full Text | Info/History | Metrics | Preview PDF

Abstract

Imaging-based spatially resolved transcriptomics can localise transcripts within cells in 3D. Cell segmentation precedes assignment of transcripts to cells and annotation of cell function. However, cell segmentation is usually performed in 2D, thus unable to deal with spatial doublets arising from overlapping cells, resulting in segmented cells containing transcripts originating from multiple cell-types. Here we present a computational tool called ovrlpy that identifies overlapping cells, tissue folds and inaccurate cell-segmentation.

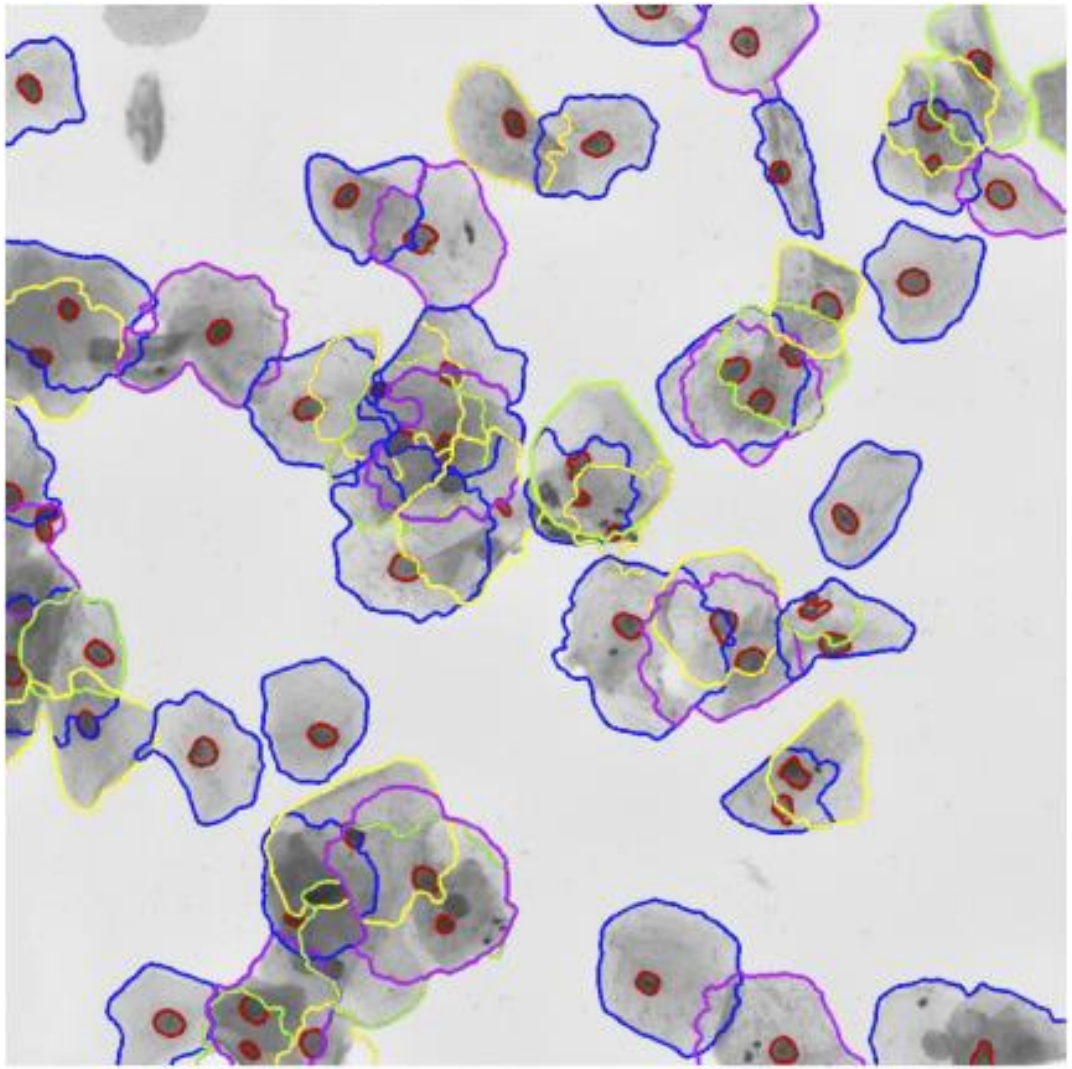
Pre-processing and quality control

We go too quickly into downstream analysis (e.g. annotation, spatial relationships)

A lack of early pre-processing and quality guidelines of imaging based spatial transcriptomics

An thus far ignored aspect of spatial transcriptomics: **overlapping cells**

Overlapping cells affects various cell types in practice



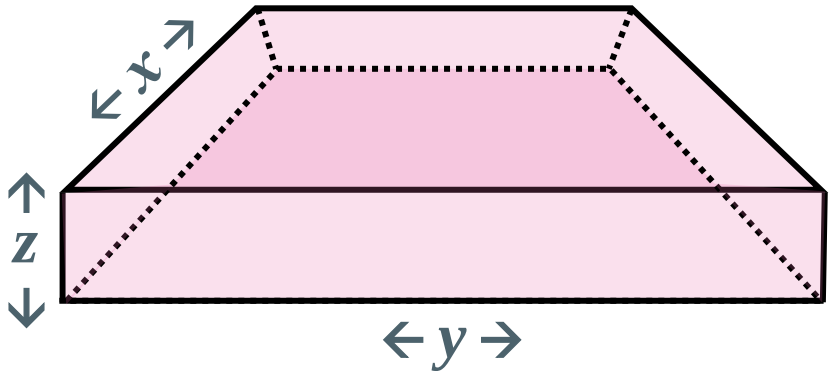
Lu et al, 2017. IEEE Trans Med Imaging

Imaging-based spatial transcriptomics is 3-D

… but how 3-D is it?

A typical section would be up to 1 cm x 1 cm x 10 µm (x, y, z)

- 10,000 x 10,000 x 10 µm (x, y, z)



Marco-Salas et al (2023), Nature Methods (accepted)
Tiesmeyer et al (2025) bioRxiv

elixir

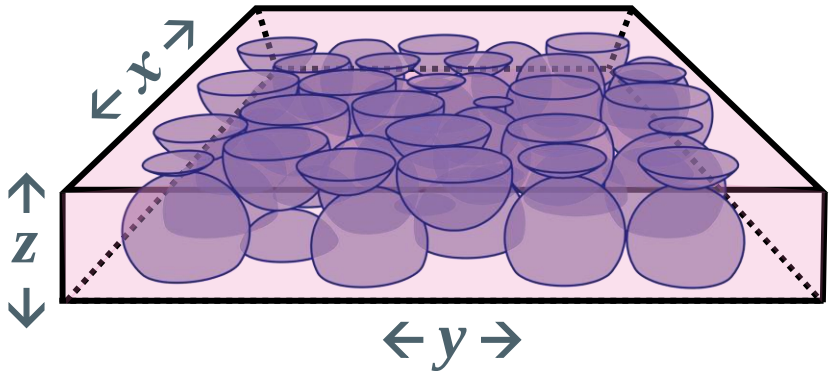
BIH Berlin Institute of Health @Charité

Imaging-based spatial transcriptomics is 3-D

… but how 3-D is it?

A typical section would be up to 1 cm x 1 cm x 10 µm (x, y, z)

- 10,000 x 10,000 x 10 µm (x, y, z)



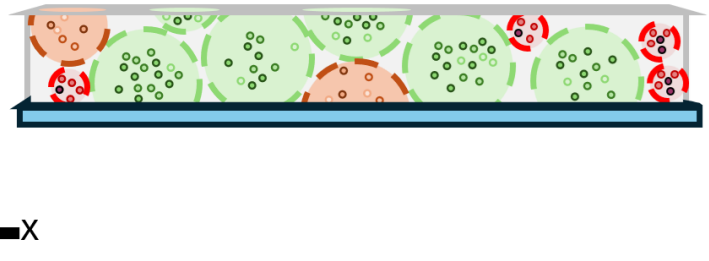
Marco-Salas et al (2023), Nature Methods (accepted)
Tiesmeyer et al (2025) bioRxiv

elixir

IBIH Berlin Institute of Health @Charité

Overl.py – a python tool for identifying cell overlaps

A light-weight python tool to identify regions with 3D overlapping cells



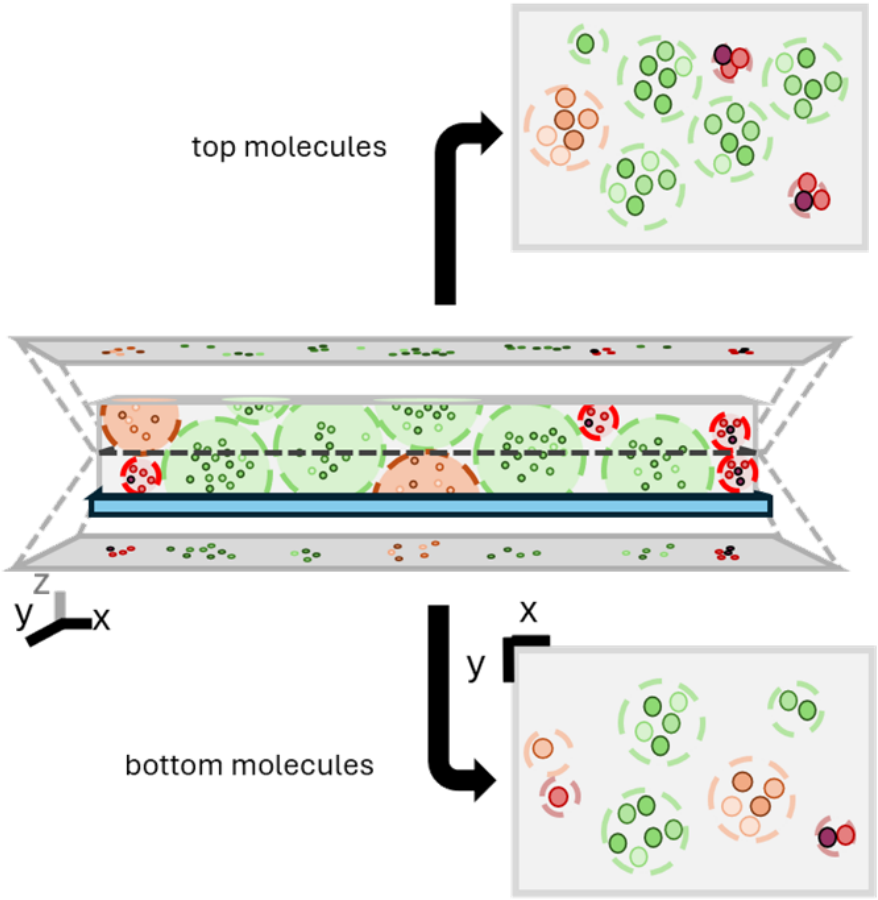
Marco-Salas et al (2023), Nature Methods (accepted)
Tiesmeyer et al (2025) bioRxiv

elixir

BIH Berlin Institute of Health @Charité

Overl.py – a python tool for identifying cell overlaps

A light-weight python tool to identify regions with 3D overlapping cells



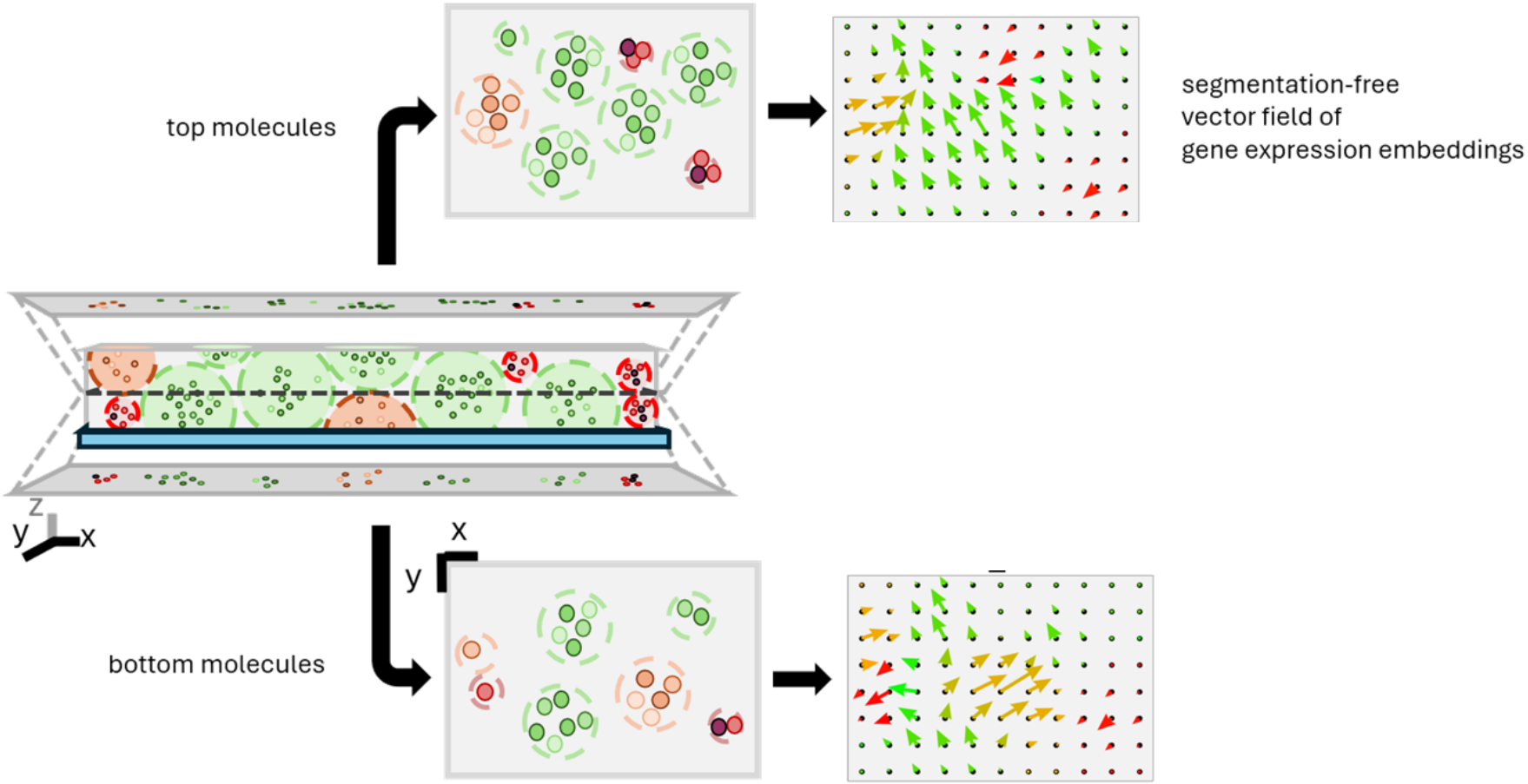
Marco-Salas et al (2023), Nature Methods (accepted)
Tiesmeyer et al (2025) bioRxiv

elixir

BIH Berlin Institute of Health @Charité

Overl.py – a python tool for identifying cell overlaps

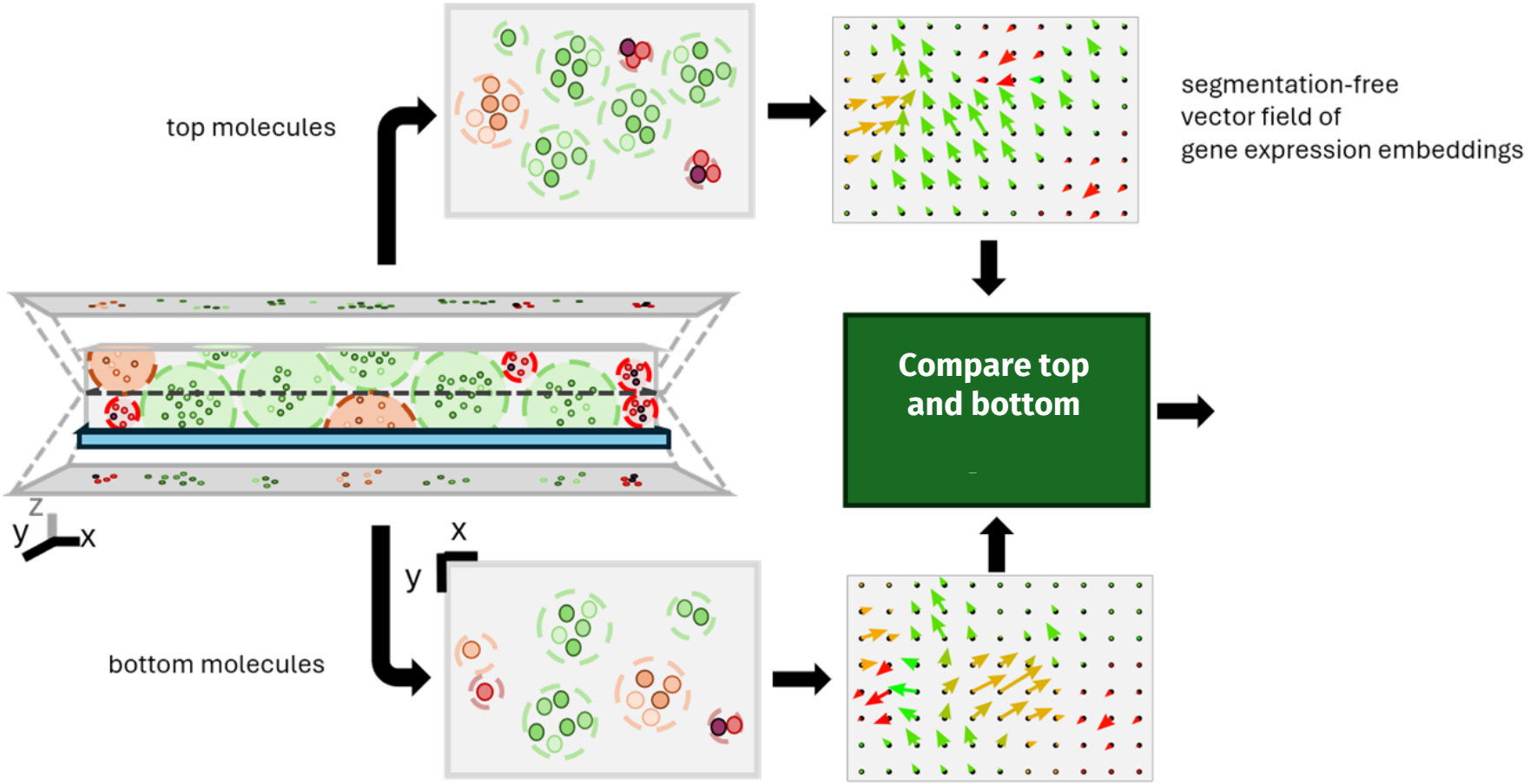
A light-weight python tool to identify regions with 3D overlapping cells



Marco-Salas et al (2023), Nature Methods (accepted)
Tiesmeyer et al (2025) bioRxiv

Overl.py – a python tool for identifying cell overlaps

A light-weight python tool to identify regions with 3D overlapping cells



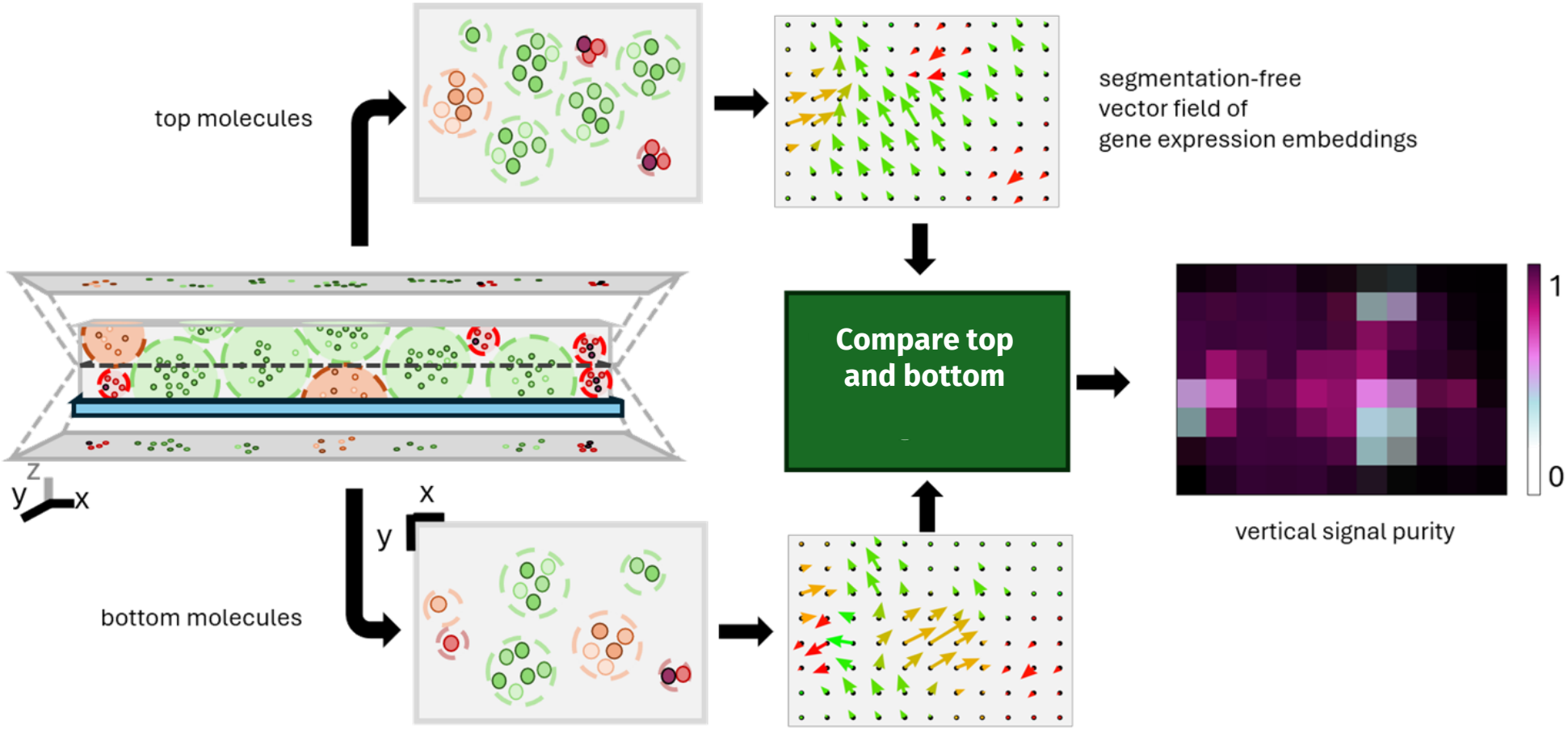
Marco-Salas et al (2023), Nature Methods (accepted)
Tiesmeyer et al (2025) bioRxiv

elixir

BIH Berlin Institute of Health @Charité

Overl.py – a python tool for identifying cell overlaps

A light-weight python tool to identify regions with 3D overlapping cells

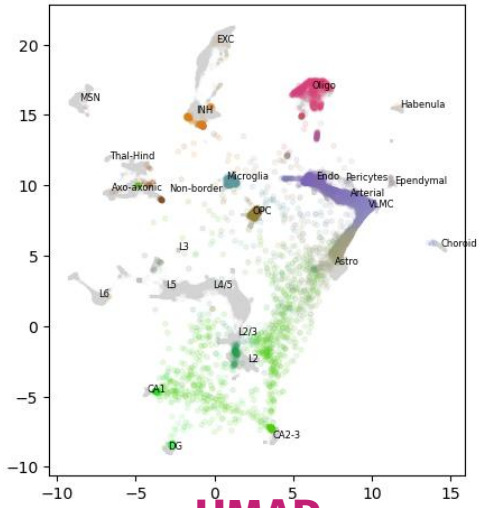


Marco-Salas et al (2023), Nature Methods (accepted)
Tiesmeyer et al (2025) bioRxiv

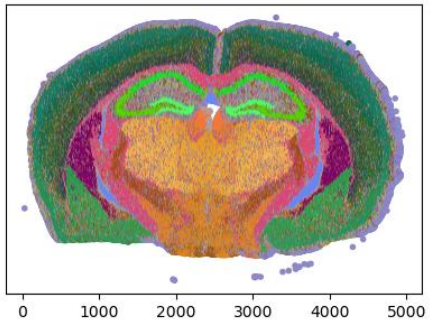
elixir

BIH Berlin Institute of Health @Charité

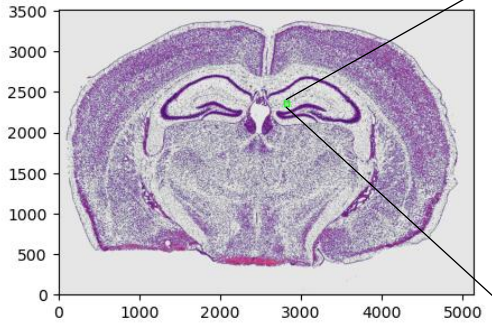
Overl.py visualisation of a region of interest (ROI) in mouse brain



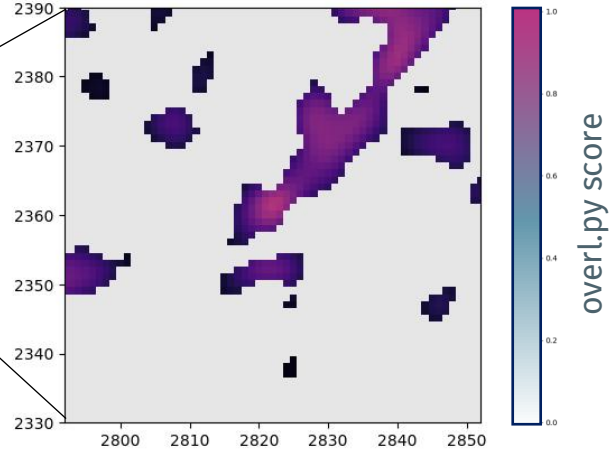
UMAP



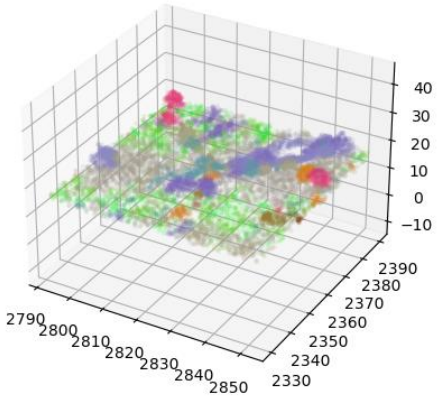
Cell-type map



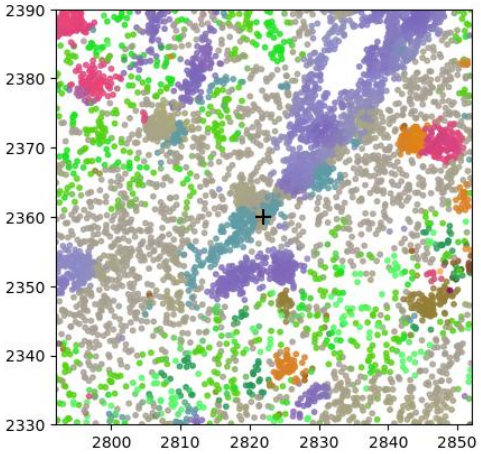
Overl.py map



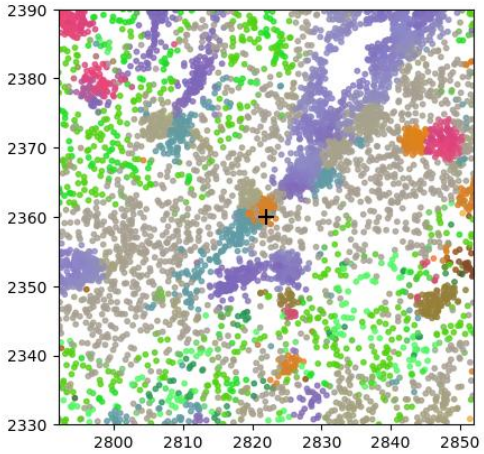
ROI: Overl.py map



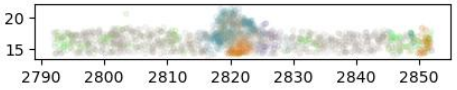
ROI: 3D



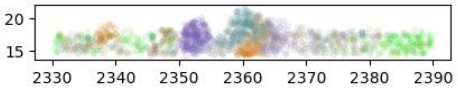
ROI: top view



ROI: bottom view



ROI: side view, *x*



ROI: side view, *y*

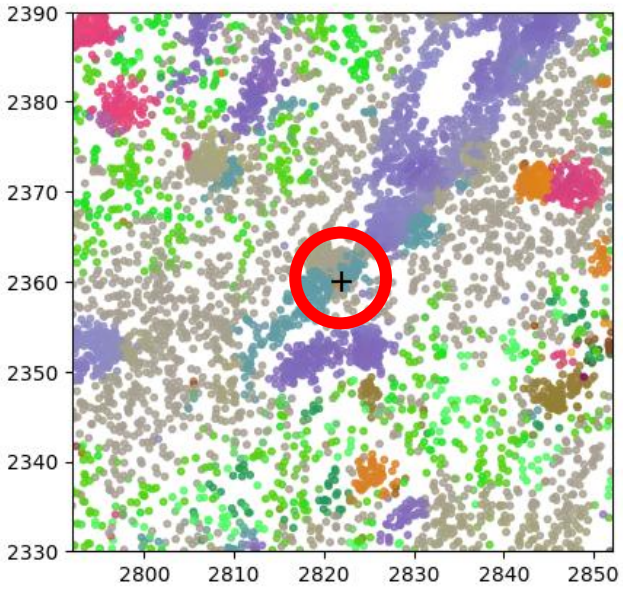
Marco-Salas et al (2023), Nature Methods (accepted)
Tiesmeyer et al (2025) bioRxiv

elixir

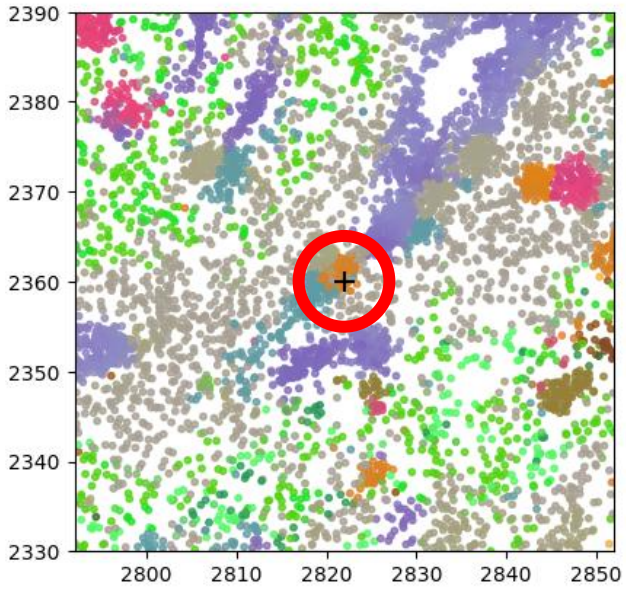
BIH Berlin Institute of Health @Charité

Example: 3-way cell overlap!

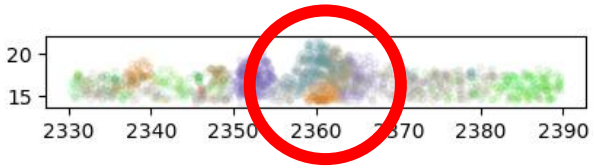
Microglial cell (blue) and astrocytes (khaki) cell on top of an inhibitory neuron (orange)



ROI: top view



ROI: bottom view

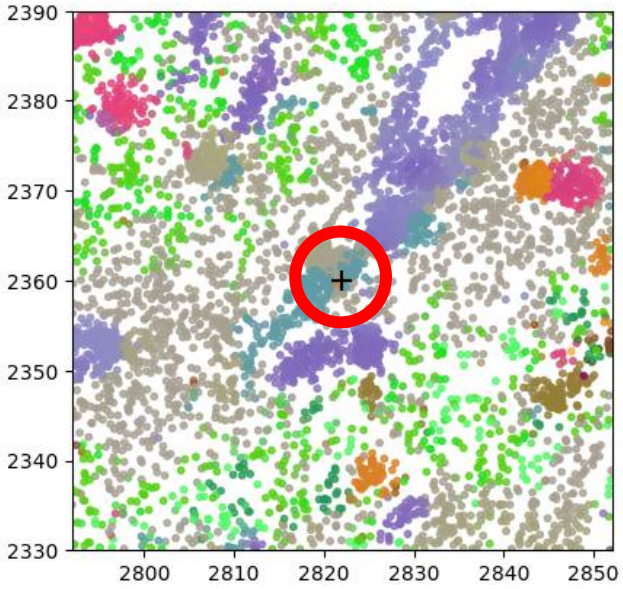


ROI: side view, *y*

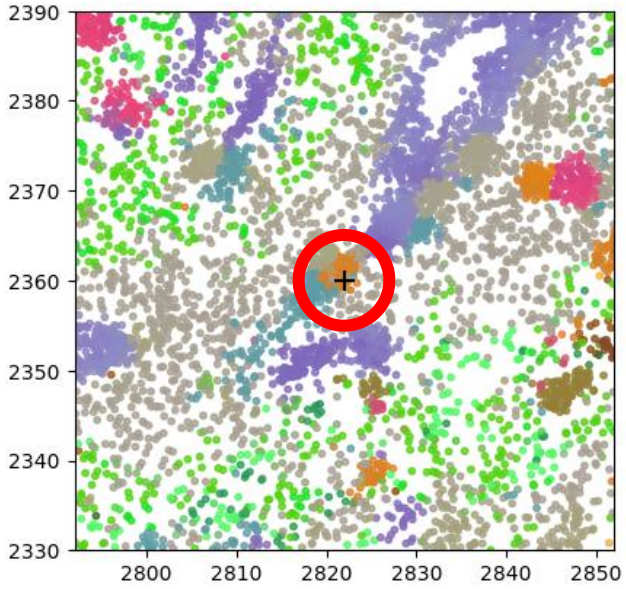
Marco-Salas et al (2023), Nature Methods (accepted)
Tiesmeyer et al (2025) bioRxiv

Example: 3-way cell overlap!

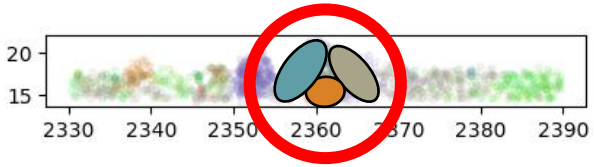
Microglial cell (blue) and astrocytes (khaki) cell on top of an inhibitory neuron (orange)



ROI: top view



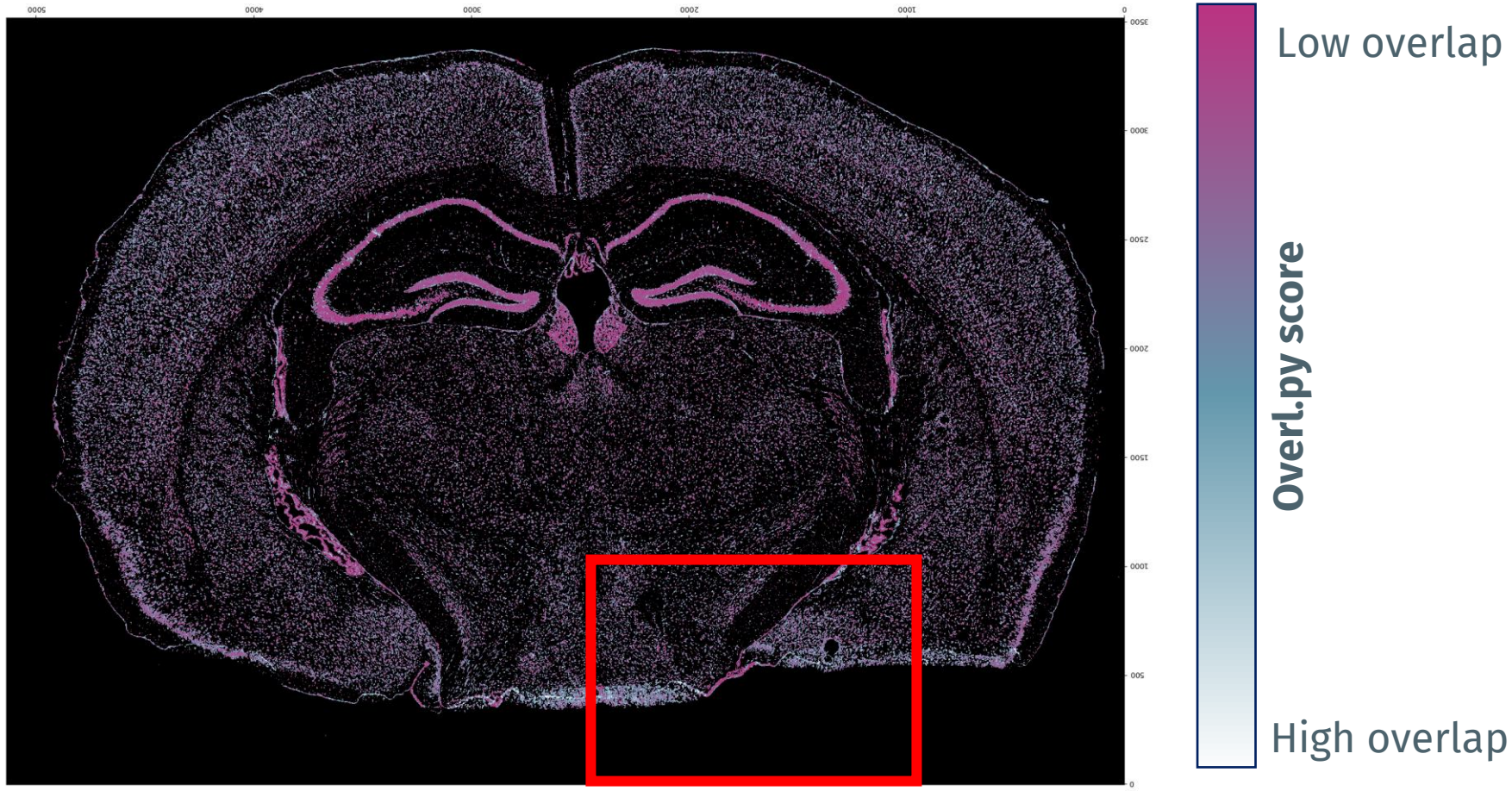
ROI: bottom view



ROI: side view, *y*

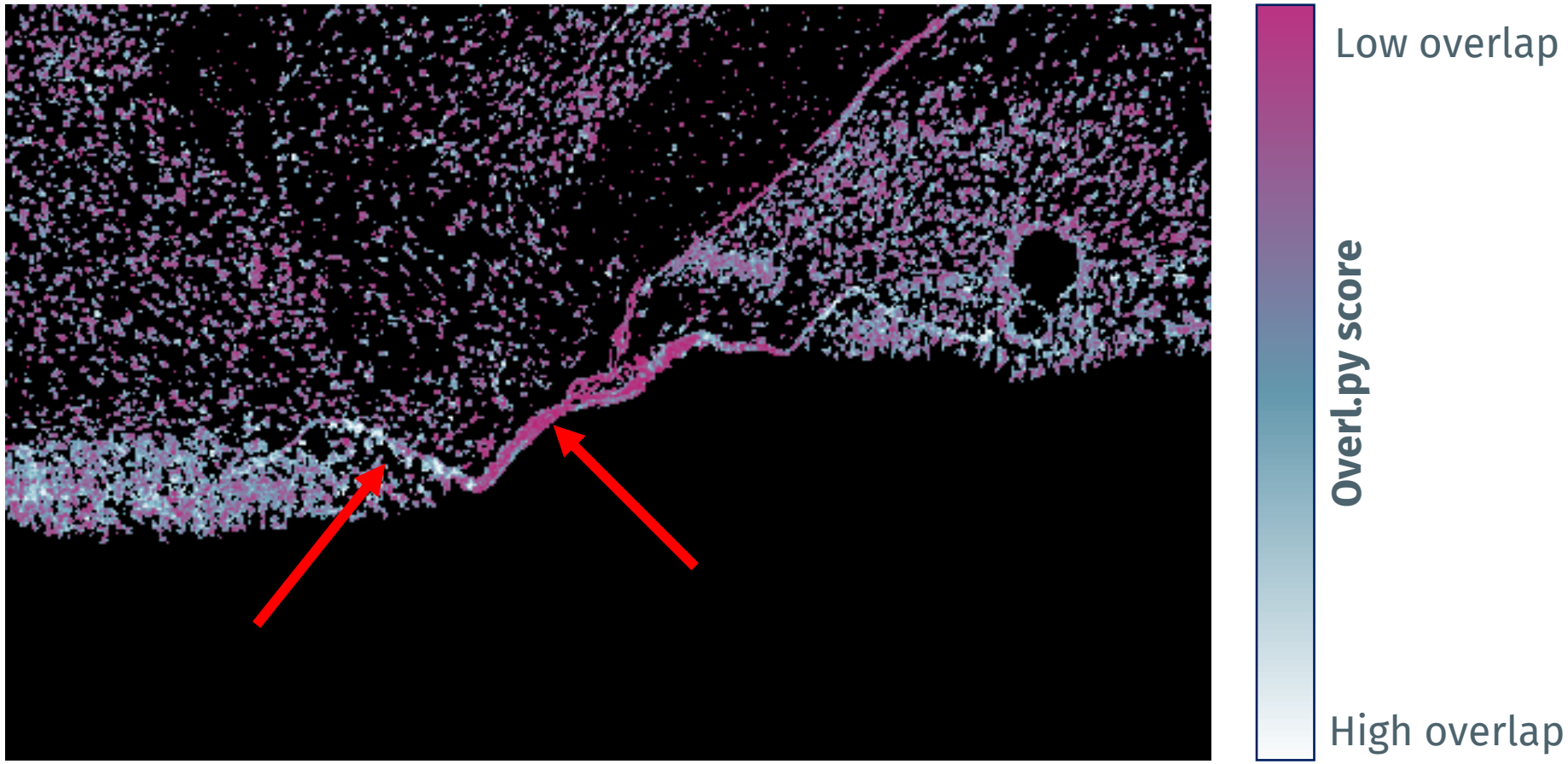
Marco-Salas et al (2023), Nature Methods (accepted)
Tiesmeyer et al (2025) bioRxiv

Overl.py detects folds in the tissue sample



Marco-Salas et al (2023), Nature Methods (accepted)
Tiesmeyer et al (2025) bioRxiv

Overl.py detects folds in the tissue sample

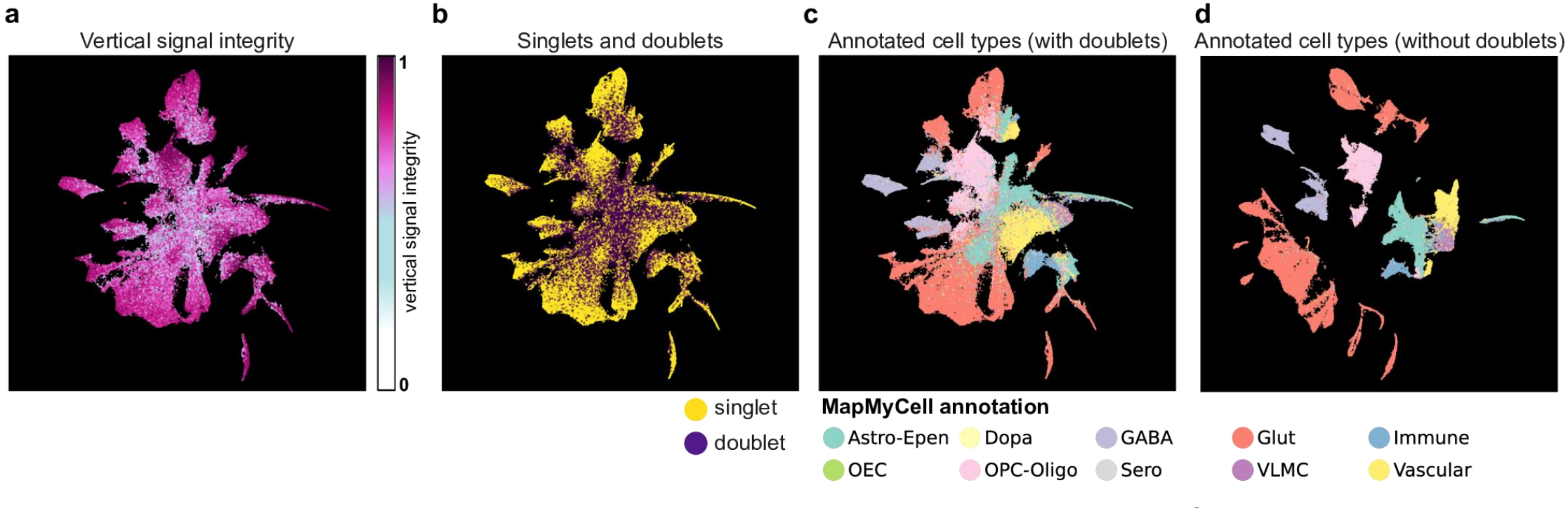


Marco-Salas et al (2023), Nature Methods (accepted)
Tiesmeyer et al (2025) bioRxiv

elixir

BIH Berlin Institute of Health @Charité

Removing overlapping cells improves cell-type clustering



Marco-Salas et al (2023), Nature Methods (accepted)
Tiesmeyer et al (2025) bioRxiv