

scverse

Foundational tools for single-cell omics data analysis

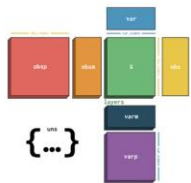
[GitHub](#)[Discourse](#)[Zulip](#)[Twitter](#)[YouTube](#)

```
import scanpy as sc
```



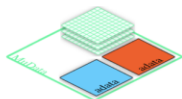
```
sc.pl.pca(adata)
```





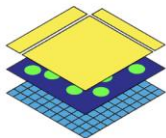
anndata

Anndata is a Python package for handling annotated data matrices in memory and on disk, positioned between pandas and xarray. anndata offers a broad range of computationally efficient features including, among others, sparse data support, lazy operations, and a PyTorch interface.

[GitHub](#)[Documentation](#)[PyPI](#)[Conda](#)

mudata

MuData is a format for annotated multimodal datasets where each modality is represented by an AnnData object. MuData's reference implementation is in Python, and the cross-language functionality is achieved via HDF5-based .h5mu files with libraries in R and Julia.

[GitHub](#)[Documentation](#)[PyPI](#)[Conda](#)[Muon.jl](#)

spatialdata

SpatialData is a data framework that comprises a FAIR storage format and a collection of python libraries for performant access, alignment, and processing of uni- and multi-modal spatial omics datasets. This repository contains the core spatialdata library. See the links below to learn more about other packages in the SpatialData ecosystem.

[GitHub](#)[Documentation](#)[PyPI](#)[spatialdata-io](#)

scanpy

Scanpy is a scalable toolkit for analyzing single-cell gene expression data built jointly with anndata. It includes preprocessing, visualization, clustering, trajectory inference and differential expression testing. The Python-based implementation efficiently deals with datasets of more than one million cells.

[GitHub](#)[Documentation and tutorials](#)[PyPI](#)[Conda](#)

scvi-tools

scvi-tools is a library for developing and deploying machine learning models based on PyTorch and AnnData. With an emphasis on probabilistic models, scvi-tools streamlines the development process via training, data management, and user interface abstractions. scvi-tools also contains easy-to-use implementations of more than 14 state-of-the-art probabilistic models in the field.

[GitHub](#)[Documentation and tutorials](#)[PyPI](#)[Website](#)

squidpy

Squidpy is a tool for the analysis and visualization of spatial molecular data. It builds on top of scanpy and anndata, from which it inherits modularity and scalability. It provides analysis tools that leverages the spatial coordinates of the data, as well as tissue images if available.

[GitHub](#)[Documentation and tutorials](#)[PyPI](#)

Introduction to the SpatialData framework

Tim Treis
2025-11-13

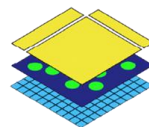
November 13th, 2024, Boston, MA



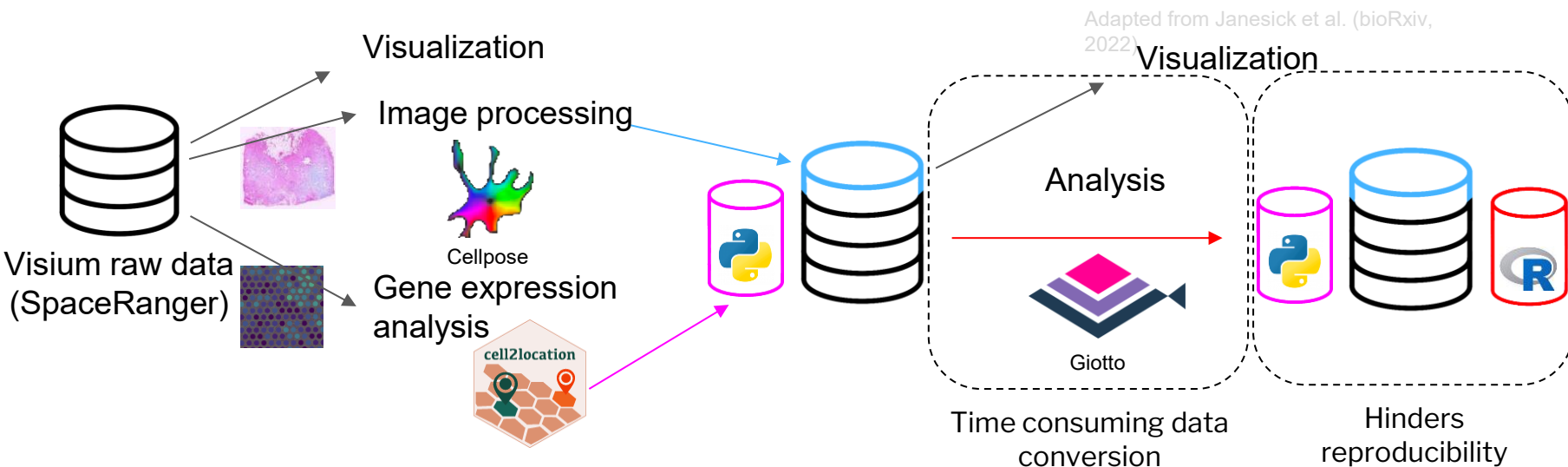
scverse

tim.treis@helmholtz-munich.de

SpatialData



The analysis of spatial omics data presents representation challenges



A downside of the fast-paced technological innovation is file format heterogeneity



RARECYTE

BaristaSeq
Cartana IIS
DBIT-seq
FISSEQ
HDST
IMC/Hyperion

Rebus Biosystems

SABER-FISH

seqFISH+

Seq-scope

Slide-DNA-seq

Slide-seq v2

smFISH

SpaceM

Spatial Transcriptomics

StarMAP

Tomo-seq

Visium (SpaceRanger 1.0.0, 1.1.0, 1.2.0, 1.3.0, 1.3.1, 2.0.0, 2.0.1, 2.1.0, ...)

Visium HD

Xenium (Xenium Analyzer 1.0.2, 1.3.0, 1.4.0, 1.5.0, 1.6.0, 1.7.0, 2.0.0, ...)

XYZeq

...



nanoString

LCM

MALDI

MERFISH

MIBI-TOF

Molecular Cartography

NanoString CosMx SMI

NanoString GeoMx DSP

osmFISH

PLISH

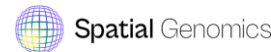
Pixel-seq

RNA SPOTs

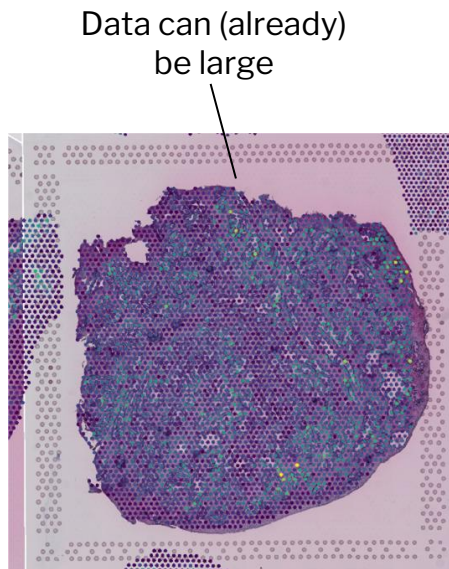
RNAscope

vizgen

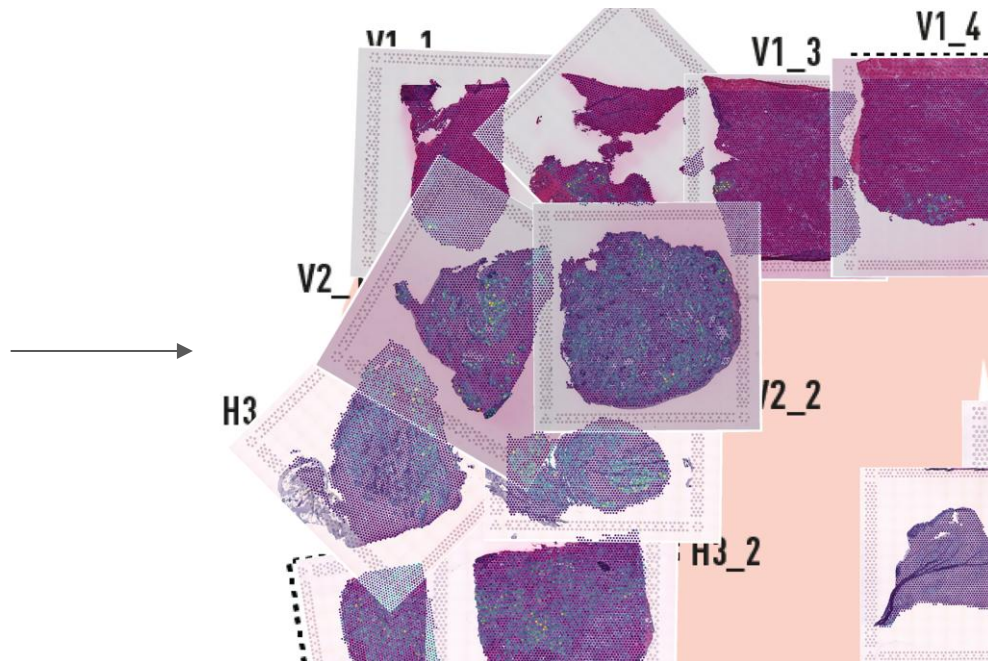
curio



Spatial omics experiments are increasingly complex and disk-heavy







Single slide



Dataset from Erickson, Lundeberg et al. (Nature, 2022)

Multiple (overlapping) slides in space

The emerging landscape of spatial profiling technologies

Jeffrey R. Moffitt ^{1,2}, Emma Lundberg ^{3,4,5,6} and Holger Heyn ^{7,8} 

NATURE REVIEWS | GENETICS

VOLUME 23 | DECEMBER 2022 | 741

Data and performance standards. A key requirement on the roadmap towards broad use and clinical implementation will be the standardization of data generation and file formats¹⁷⁸

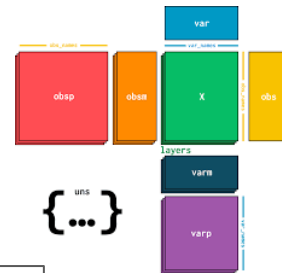
Not only will such standards continue to drive computational innovations, they would also facilitate the large-scale integration of spatial omics and existing pathology imaging data.

SpatialData is a solution that bridges existing communities

scverse core



Data analysis



AnnData



Relies on existing Python GIS technologies

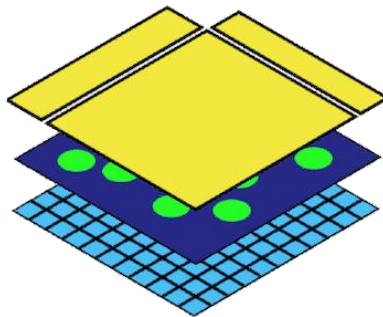
Napari core



napari

Interactive visualization

SpatialData



Marconato*, Palla*, Yamauchi*, Virshup* et al.
(Accepted in Nature Methods, 2023)

SpatialData:

- infrastructure for data storage, manipulation and visualization
- non-goal: not an analysis library

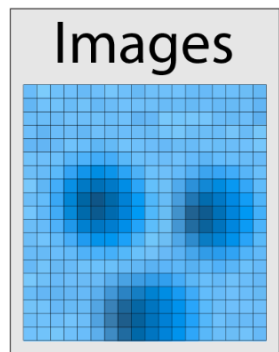
OME (Open Microscopy Environment)



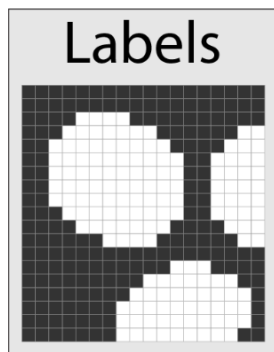
OMEZarr

Large images, standard formats

Data representation is abstracted as a modular combination of reusable elements



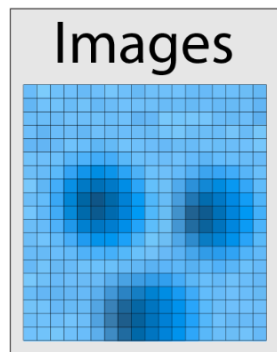
Microscopy
images, ...



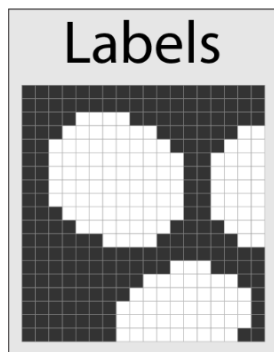
Segmentation
mask

Raster
geometries

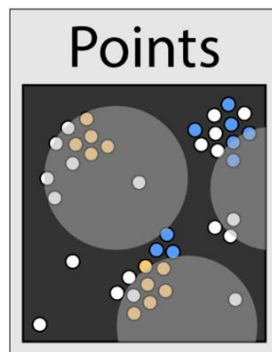
Data representation is abstracted as a modular combination of reusable elements



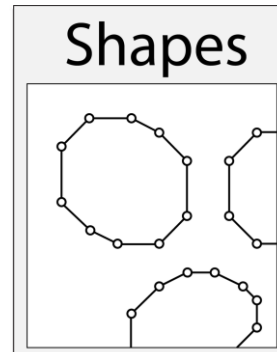
Microscopy
images, ...



Segmentation
mask



Transcripts
locations

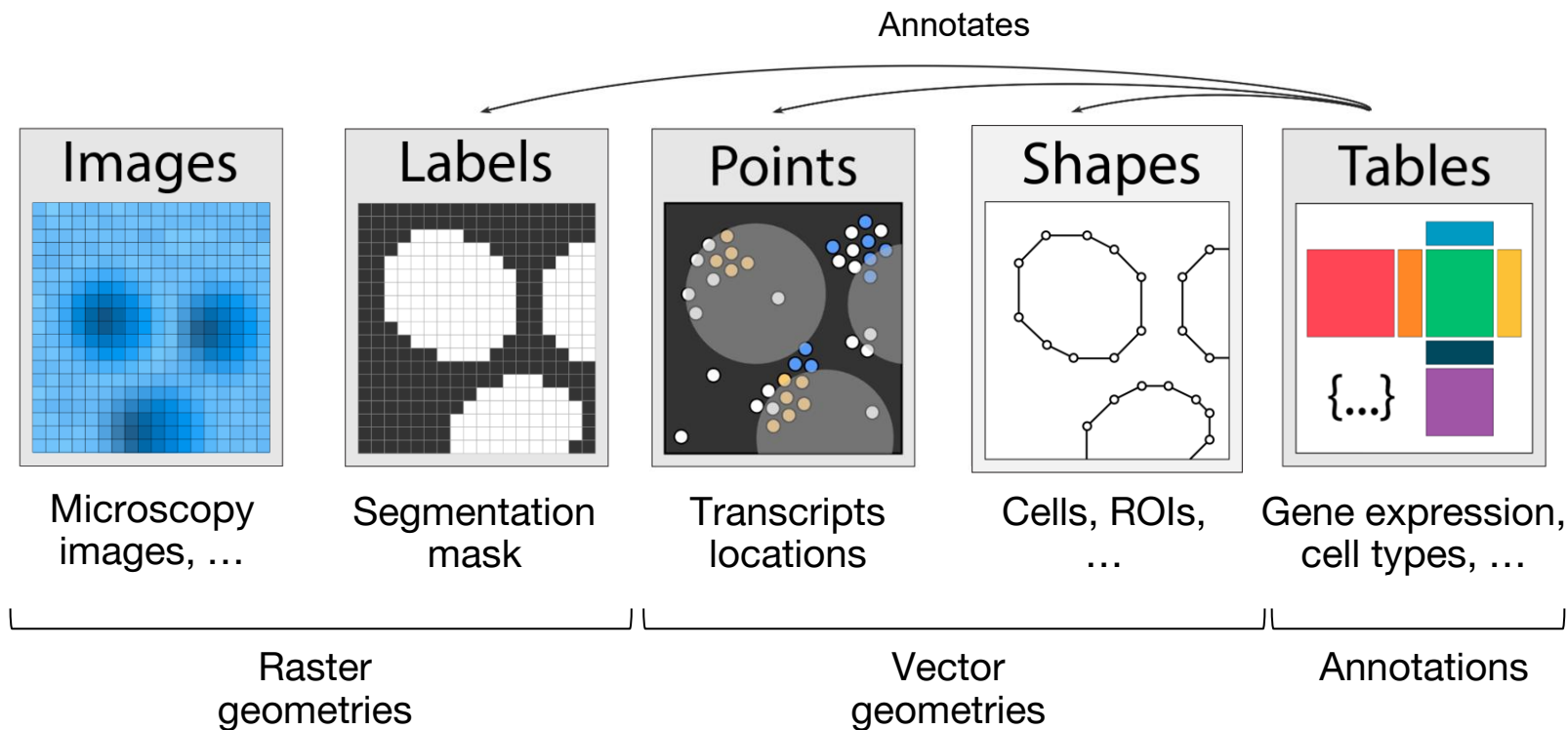


Cells, ROIs,
...

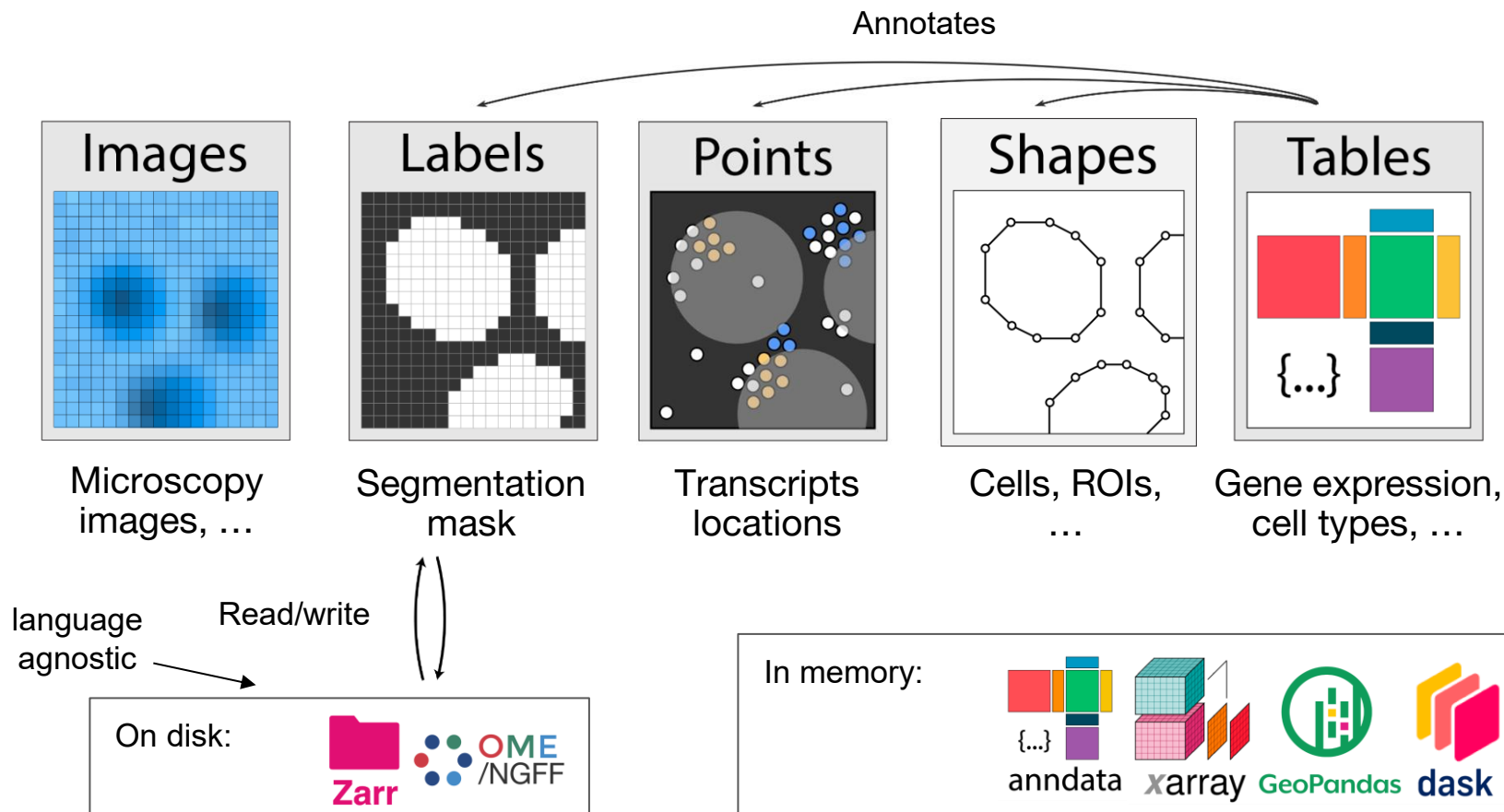
Raster
geometries

Vector
geometries

Data representation is abstracted as a modular combination of reusable elements



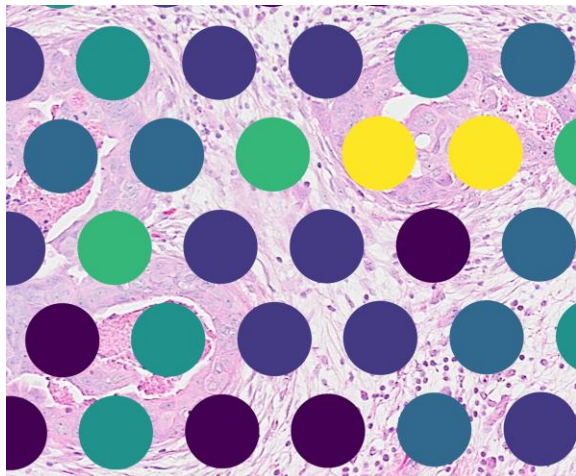
Data representation is abstracted as a modular combination of reusable elements



SpatialData unifies the representation of spatial omics across technologies

Data at different scales

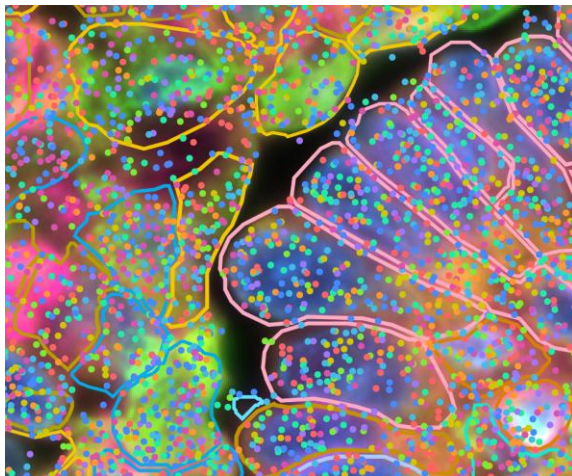
Visium



Data: <https://www.10xgenomics.com/products/xenium-in-situ/preview-dataset-human-breast>

Resolution: 55µm
Transcriptome-wide

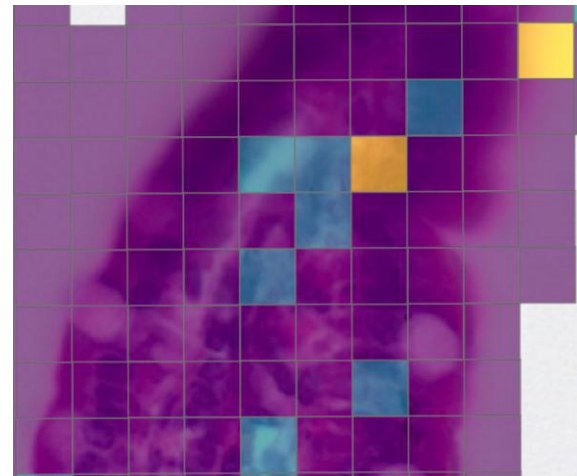
Xenium



Preview Data: FFPE Human Lung Cancer with Xenium Multimodal Cell Segmentation

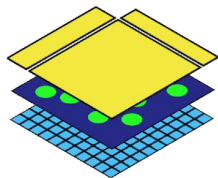
Resolution: single-molecule
Up to 5K genes

Visium HD



Visium HD Spatial Gene Expression Library, Mouse Small Intestine (FFPE)

Resolution: 2µm, 8µm, 16µm, ...
Transcriptome-wide



SpatialData

- Simple read/write
- Flexible representation
- Object manipulation

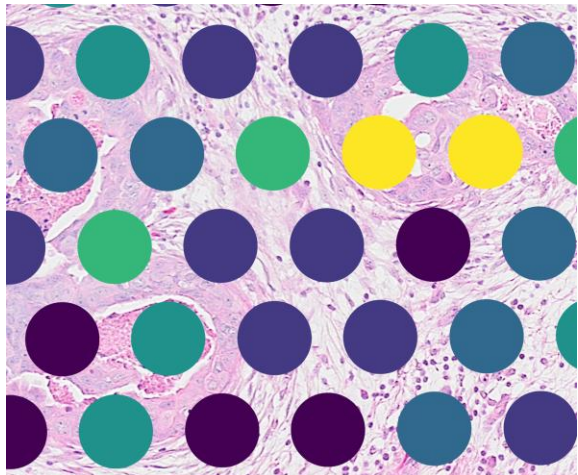
- Interoperability across analysis methods



SpatialData unifies the representation of spatial omics across technologies

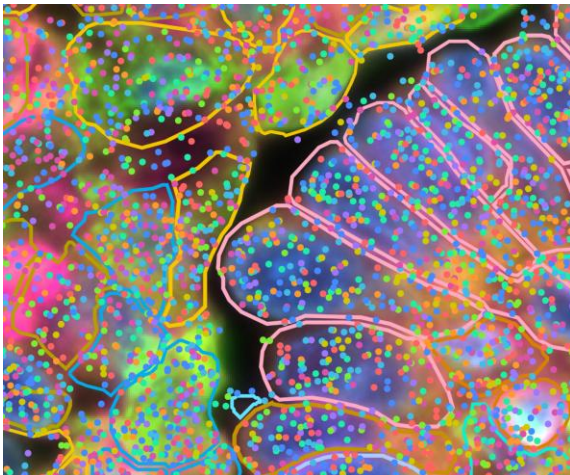
Data at different scales

Visium



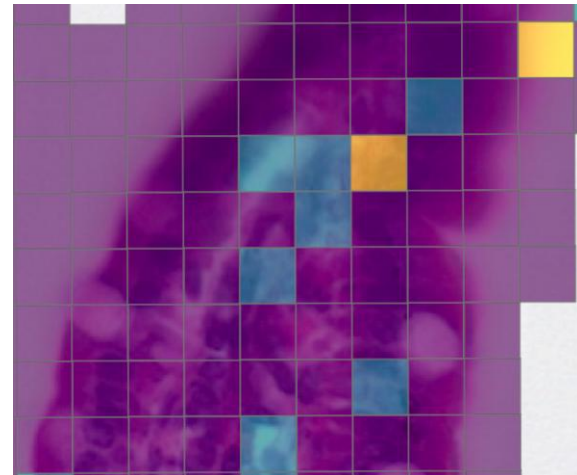
Data: <https://www.10xgenomics.com/products/xenium-in-situ/preview-dataset-human-breast>

Xenium



Preview Data: FFPE Human Lung Cancer with Xenium Multimodal Cell Segmentation

Visium HD



Visium HD Spatial Gene Expression Library, Mouse Small Intestine (FFPE)

Reading the data from disk:

```
from spatialdata_io import visium  
  
sdata = visium('my_space_ranger_output')
```

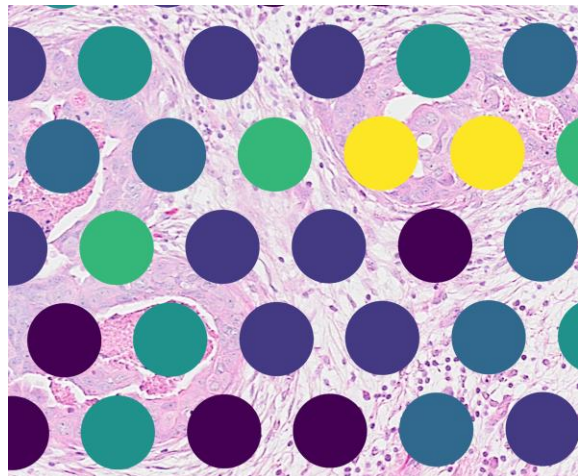
```
from spatialdata_io import xenium  
  
sdata = xenium('my_xenium_analyzer_output')
```

```
from spatialdata_io import visium_hd  
  
sdata = visium_hd('my_space_ranger_output')
```

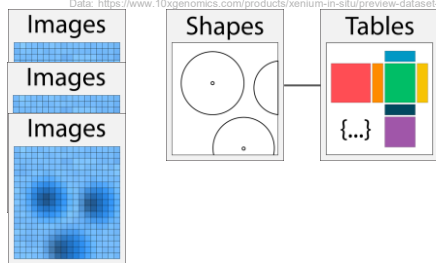

SpatialData unifies the representation of spatial omics across technologies

Data at different scales

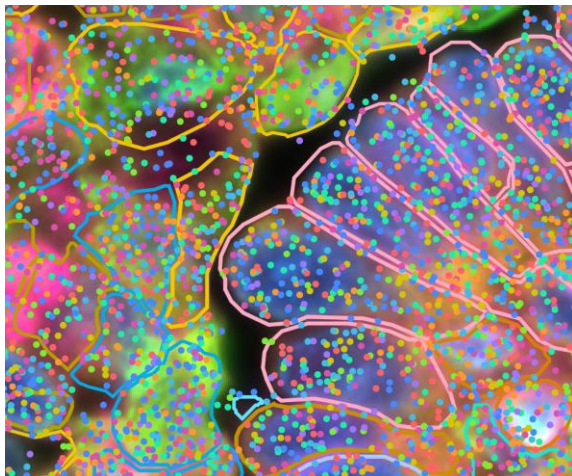
Visium



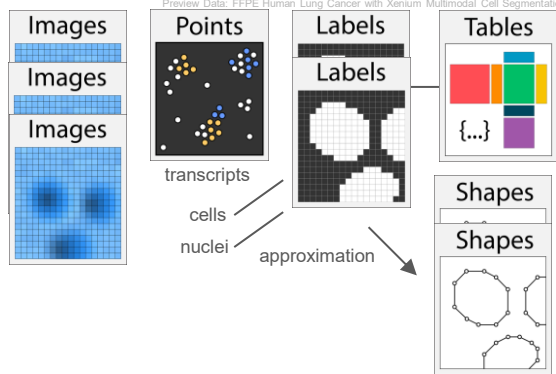
Data: <https://www.10xgenomics.com/products/xenium-in-situ/preview-dataset-human-breast>



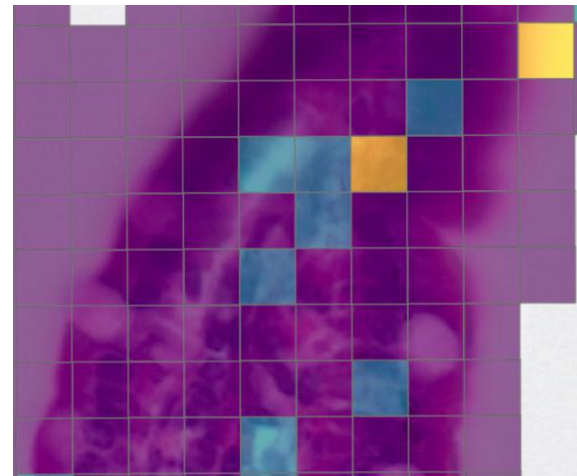
Xenium



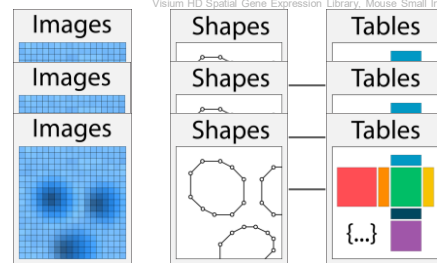
Preview Data: FFPE Human Lung Cancer with Xenium Multimodal Cell Segmentation



Visium HD

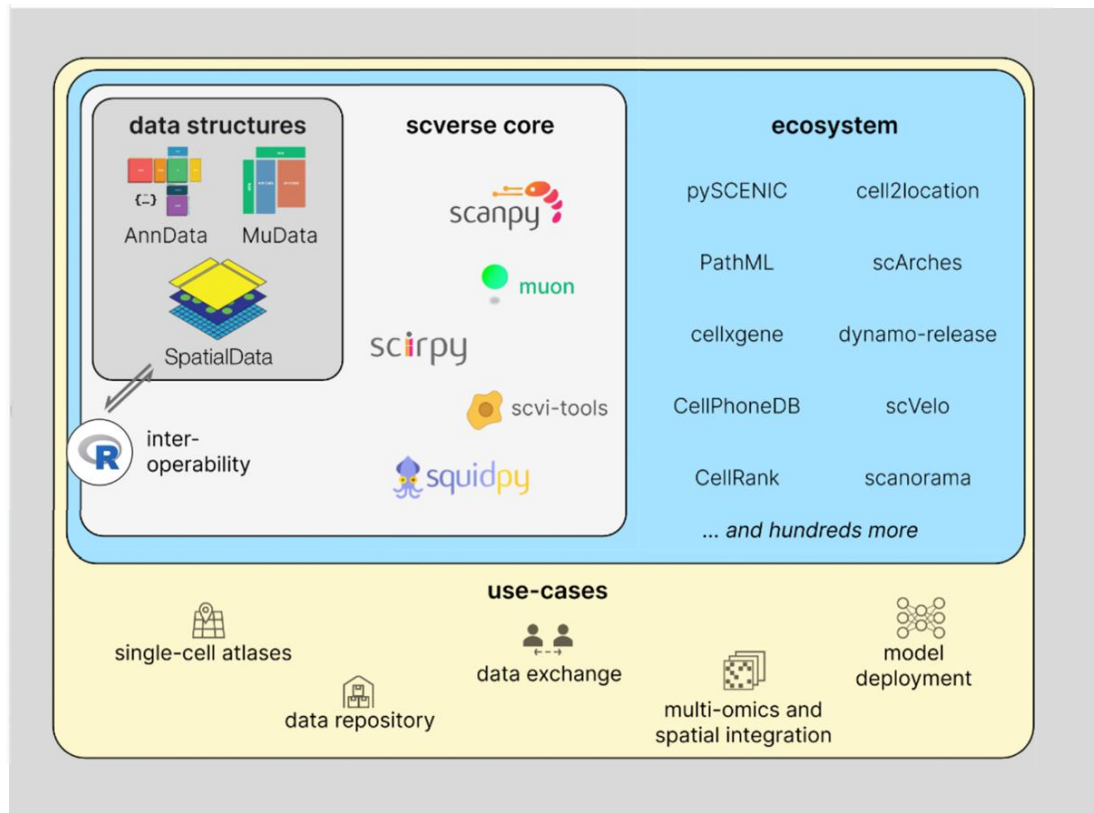


Visium HD Spatial Gene Expression Library: Mouse Small Intestine (FFPE)

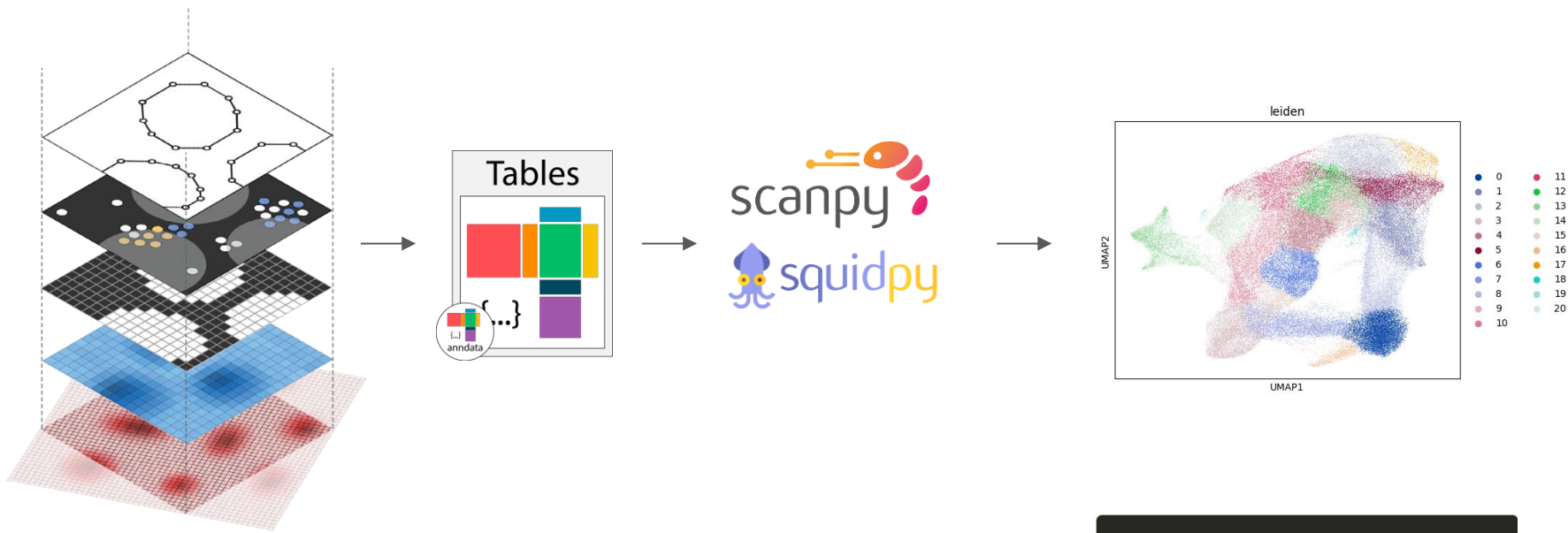


2µm, 8µm, 16µm bins

SpatialData is a core data structure for the scverse ecosystem



The modular structure of SpatialData objects integrates into existing analysis workflows



```
import spatialdata

sdata = spatialdata.read_zarr('my_visium_hd_data')
adata = sdata['16um_bins']
```

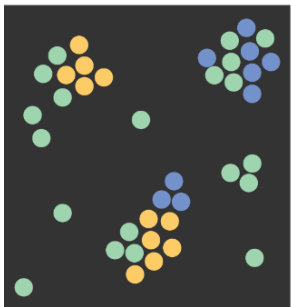
```
import scanpy as sc

sc.pp.calculate_qc_metrics(adata, ...)
...
sc.pp.pca(adata)
...
sc.tl.umap(adata)
```

Aggregation of signals between modalities

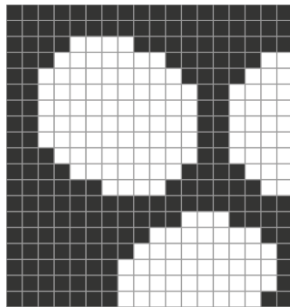
Source values
(signal)

Transcripts

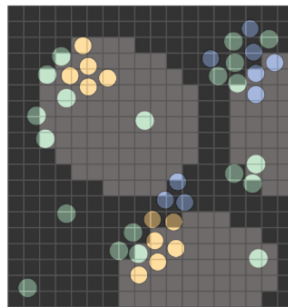


Target
geometries

Cells



Aggregate
transcripts by cells
count



6	5	2
---	---	---

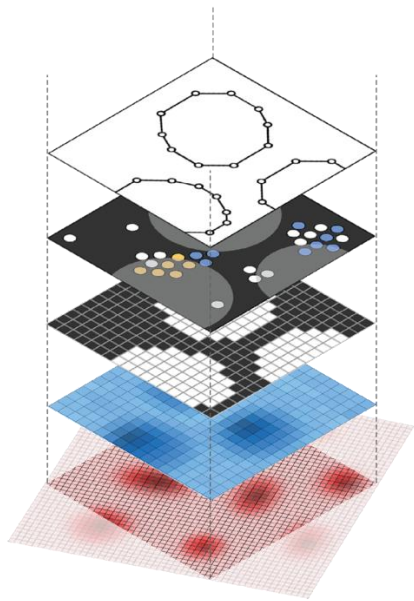
3	0	3
---	---	---

2	5	0
---	---	---

Gene expression

6	5	2
3	0	3
2	5	0

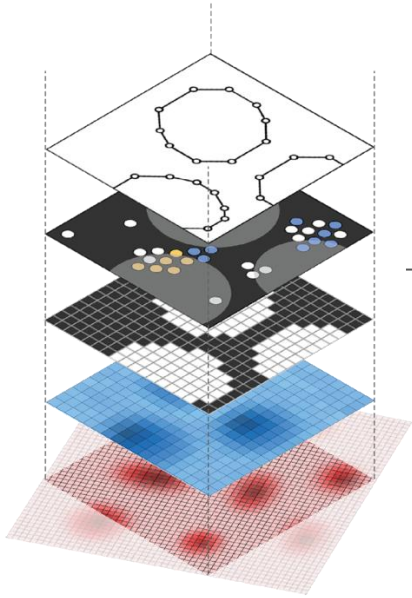
We provide methods for interactive visualization and annotation



```
import spatialdata
from napari_spatialdata import Interactive

sdata = spatialdata.read_zarr('my_visium_hd_data')
Interactive(sdata)
```

We provide methods for interactive visualization and annotation

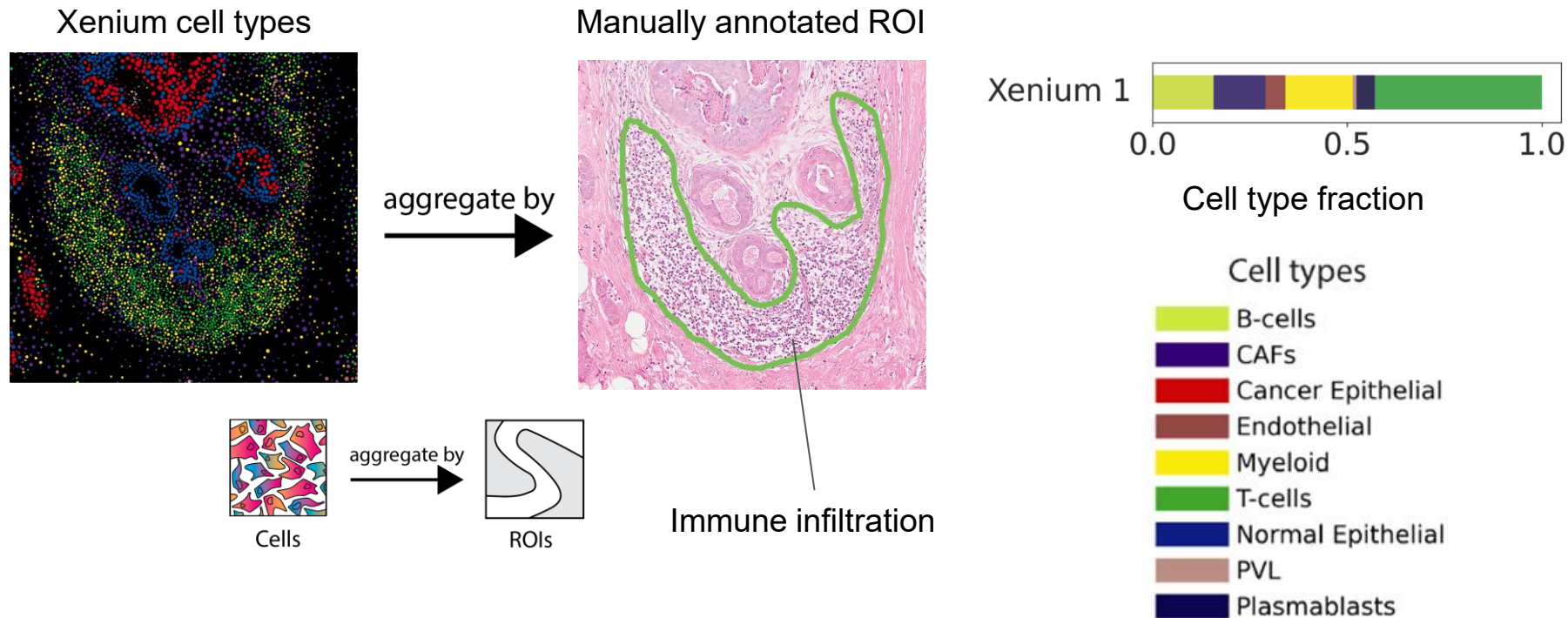


Visium HD Spatial Gene Expression Library, Mouse Small Intestine (FFPE)

```
import spatialdata
from napari_spatialdata import Interactive

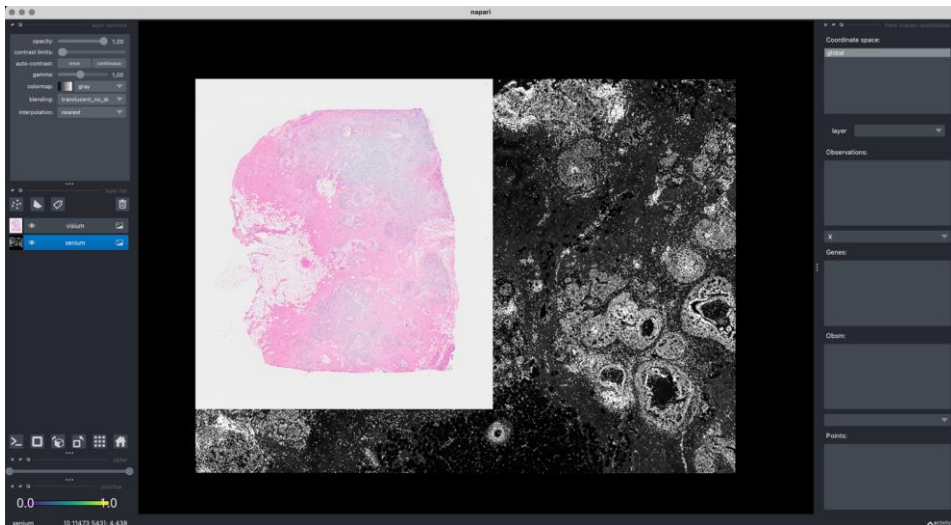
sdata = spatialdata.read_zarr('my_visium_hd_data')
Interactive(sdata)
```

Aggregation example: computing cell types fractions within regions of interest

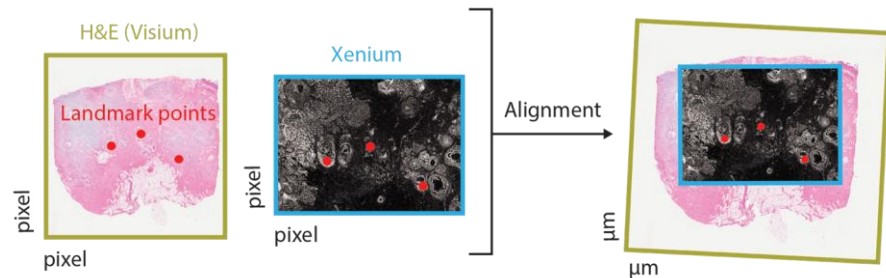
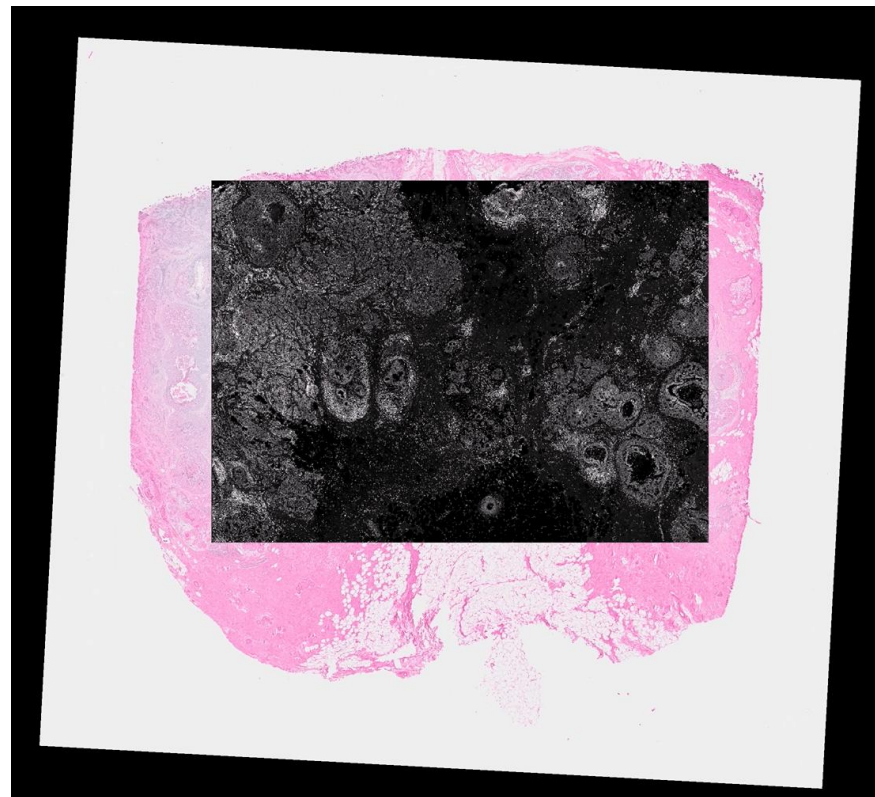


Example: joint analysis of Visium + Xenium datasets require spatial alignment

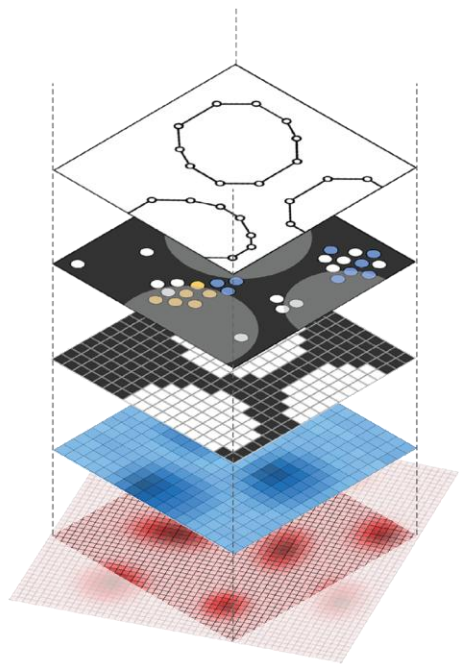
Before alignment (raw data)



After alignment

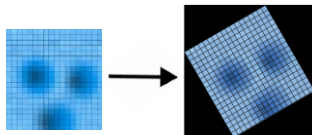


Generalized, reusable operations are defined for SpatialData objects



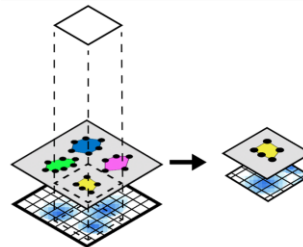
Coordinate transformations

e.g. rotate an image



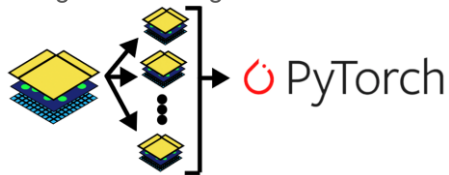
Spatial queries

e.g. crop the data



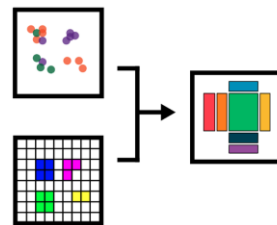
Deep learning interface

e.g. create image tiles around cells

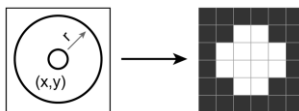


Spatial aggregations

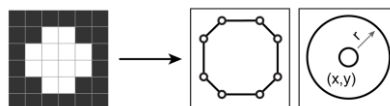
e.g. count transcripts inside cells



Rasterize



Vectorize



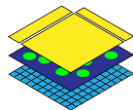
Documentation:

<https://spatialdata.scverse.org/en/stable/>

Conclusions and acknowledgements



- established interoperable format for spatial omics based on OME-NGFF



- in-memory multimodal representation
- processing, visualization
- scales to large datasets



Luca Marconato



Giovanni Palla



Kevin Yamauchi



Isaac Virshup



Tim Treis



Wouter-Michiel Vierdag



Josh Moore



Sonja Stockhaus



Elias Heidari



Marcela Toth



Quentin Blampey



Laurens Lehner



Rahul B. Shrestha



Benjamin Rombaut



Lotte Pollaris



Harald Vöhringer



Oliver Stegle



Fabian Theis



Moritz Gerstung



Sinem Saka



Yvan Saeys



Wolfgang Huber



Nature Methods, 2024

First authors are underlined

EMBL

Oliver Stegle
Luca Marconato
 Sinem Saka
 Wouter-Michiel Vierdag
 Wolfgang Huber
 Harald Vöhringer
 Constantin Ahlmann-Eltze
 Mike Smith

Helmholtz Munich

Fabian Theis
Giovanni Palla
Isaac Virshup
 Tim Treis
 Sonja Stockhaus
 Laurens Lehner
 Marcella Toth
 Rahul Shrestha

DKFZ

Ilia Kats
 Tobias Graf
 Moritz Gerstung
 Elias Heidari

Josh Moore (OME, GerBi)
Kevin Yamauchi (ETH)
 Yvan Saeys (UGhent)

Lotte Pollaris (UGhent)
Benjamin Rombaut (UGhent)
 Christian Tischer (EMBL)
 Andreas S. Eisenbarth (EMBL)
 Omer Bayraktar (Sanger)
 Tong Li (Sanger)
 Ilan Gold (Helmholtz)
 Helena Crowell (CNAG)

Nils Eiling (UZH)
 Will Moore (OME, U Dundee)
 Quentin Blampey (UParis Saclay)
 Florian Wünnemann (UKHD)
 Mark Keller (HMS)
 10x Genomics team
 CZI cellxgene team
 ...

EMBL



dkfz.



HELMHOLTZ
MUNICH



UNIVERSITÄT
HEIDELBERG
ZUSAMMEN
SEIT 1386



université
PARIS-SACLAY

spatialdata.scverse.org

Funded by



CHAN
ZUCKERBERG
INITIATIVE