PSC

Carnegie Mellon University

University of Pittsburgh

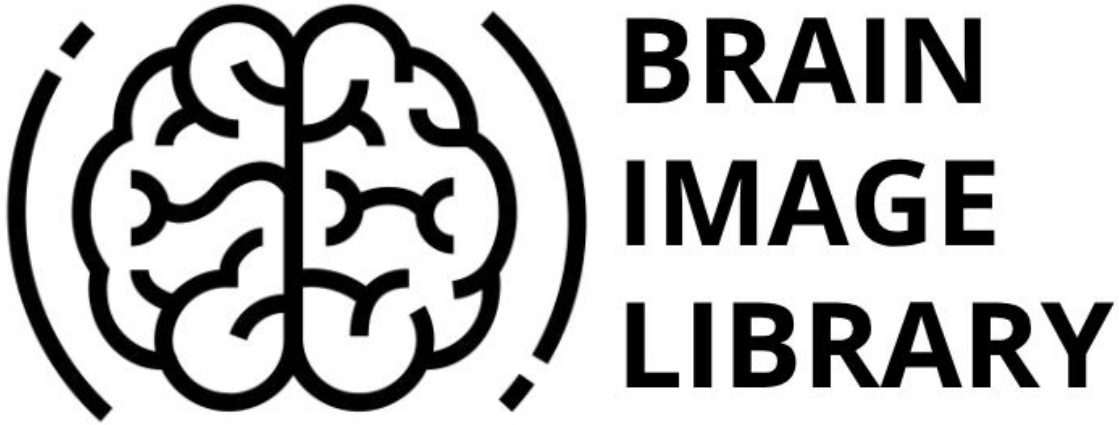
Overview of Spatial Data in BIL and HuBMAP

Alexander Ropelewski
Director of Biomedical Applications

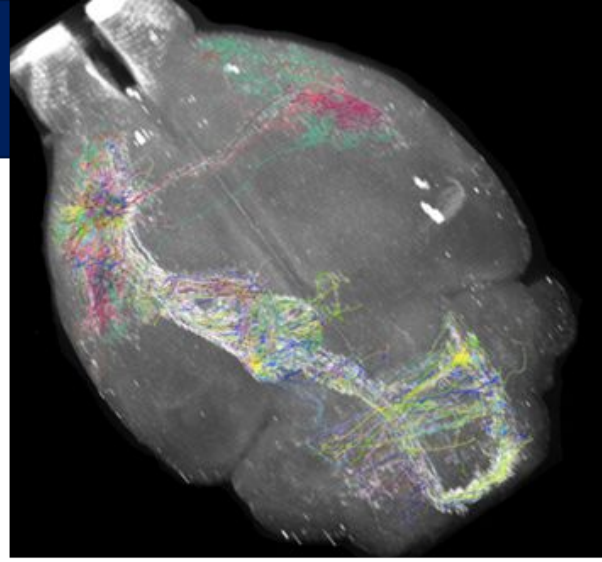
Brendan J. Honick
Senior Data Analyst/Curator, Biomedical Applications

May 12th, 2025

Part 1:
The Brain Image Library (BIL)



Brain Image Library



Mission: *National public resource enabling researchers to deposit, analyze, mine, share and interact with optical microscopy datasets of the brain.*

Preserving Data

- No depositor size limitations
- Professional curation
- Help desk support
- Network bottleneck/analysis assistance for depositors transferring large datasets
- DOI Issuance for datasets through DataCite membership
- Landing pages for datasets
- LTO Tape for disaster recovery of data
- LTO tape (7/8/9) capable for import of large datasets

Providing Access

Analysis Ecosystem

- Scalable Architecture
- Data available on PSC's HPC and AI Resources for computational analysis without downloading data
- Frequently used community open-source and commercial software tools available

Web Resources

- Metadata API with web searchable interface + File SDK
- Visualization Resources:
  - Neuroglancer (3d Datasets)
  - OpenSeaDragon (2d Datasets)
  - Napari plug-in

Engagement

Workshops

- Data Submission Workshops
- Spatial Transcriptomics Workshop for BICAN & BICAN Orientation
- CFDE Reproducible Workflow Workshop

Meetings

- Society for Neuroscience
- BRAIN Investigators
- Annual Biomedical Research Conference for Minoritized Scientists (ABRCMS)

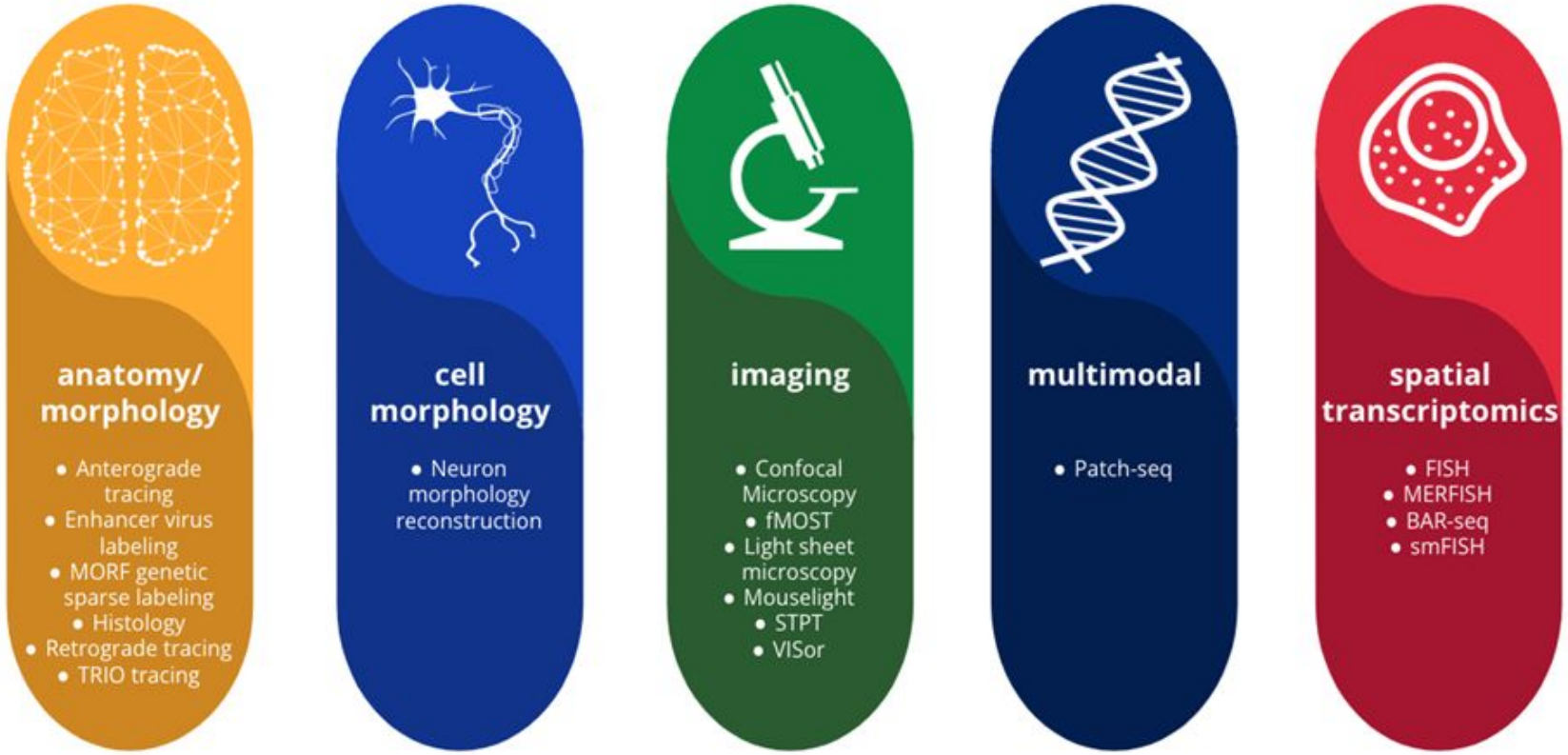
Newsletter

- The BIL Bulletin

www.brainimagelibrary.org

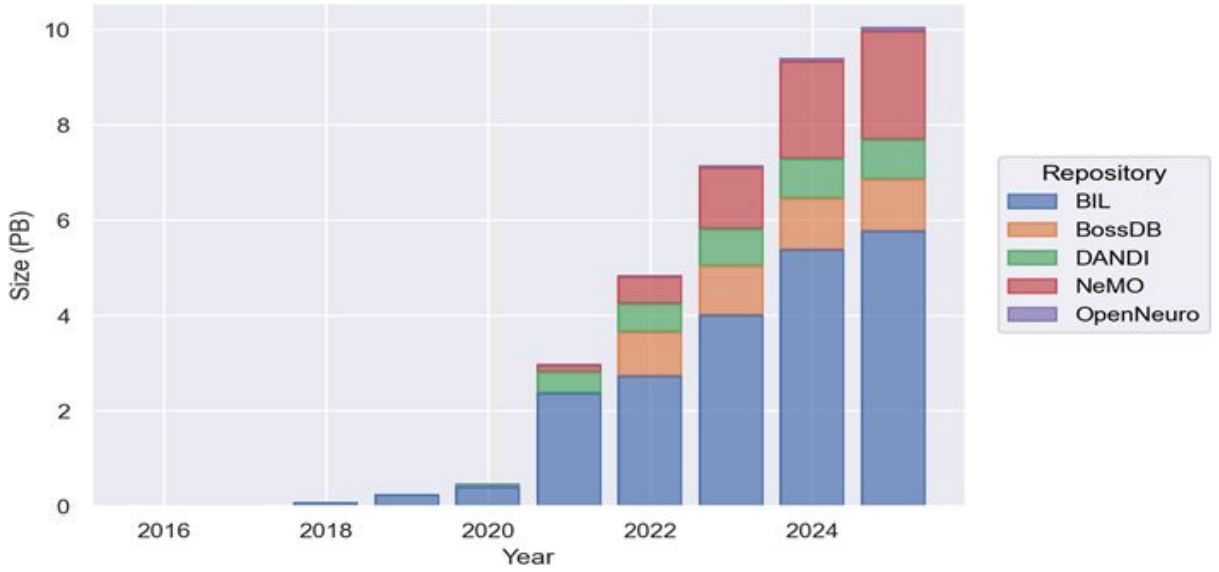
Breadth of Data

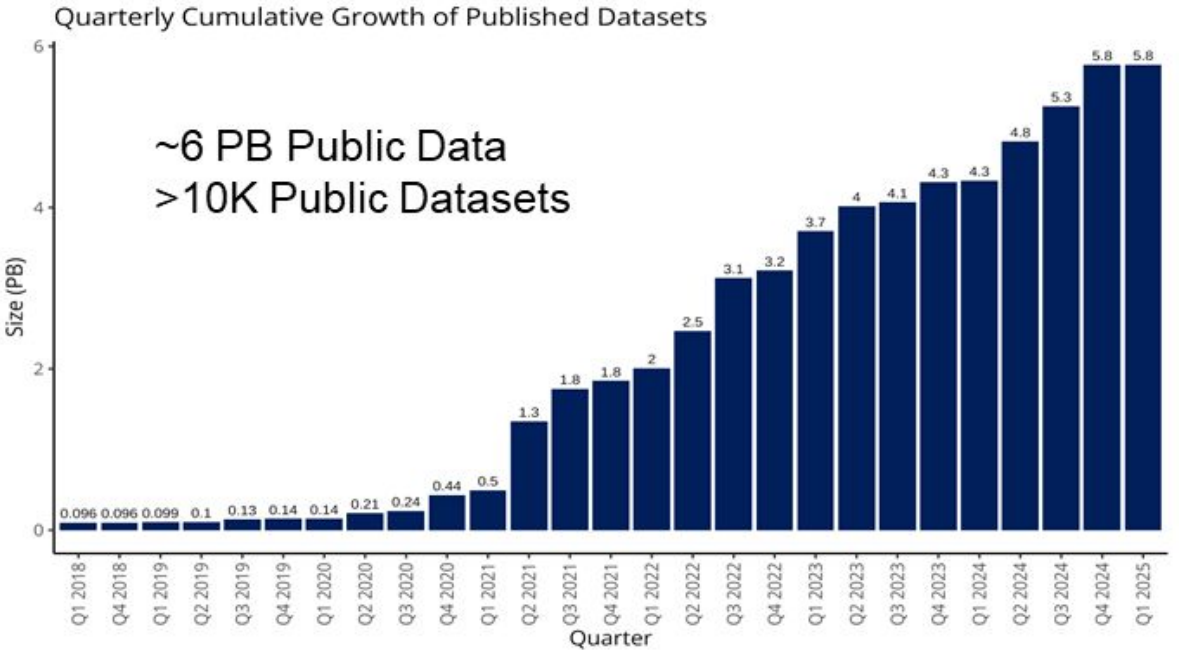
- Whole (and partial) optical microscopy brain datasets along with their higher-level data and annotations
- Targeted experiments including connectivity between cells and spatial transcriptomics
- Historical collections



Data Contributed

Data from individual Investigators plus the BICCN, BICAN, ASAP (Aligning Science Across Parkinson), SSPsyGene, CONNECTS consortia





Flagship Papers:

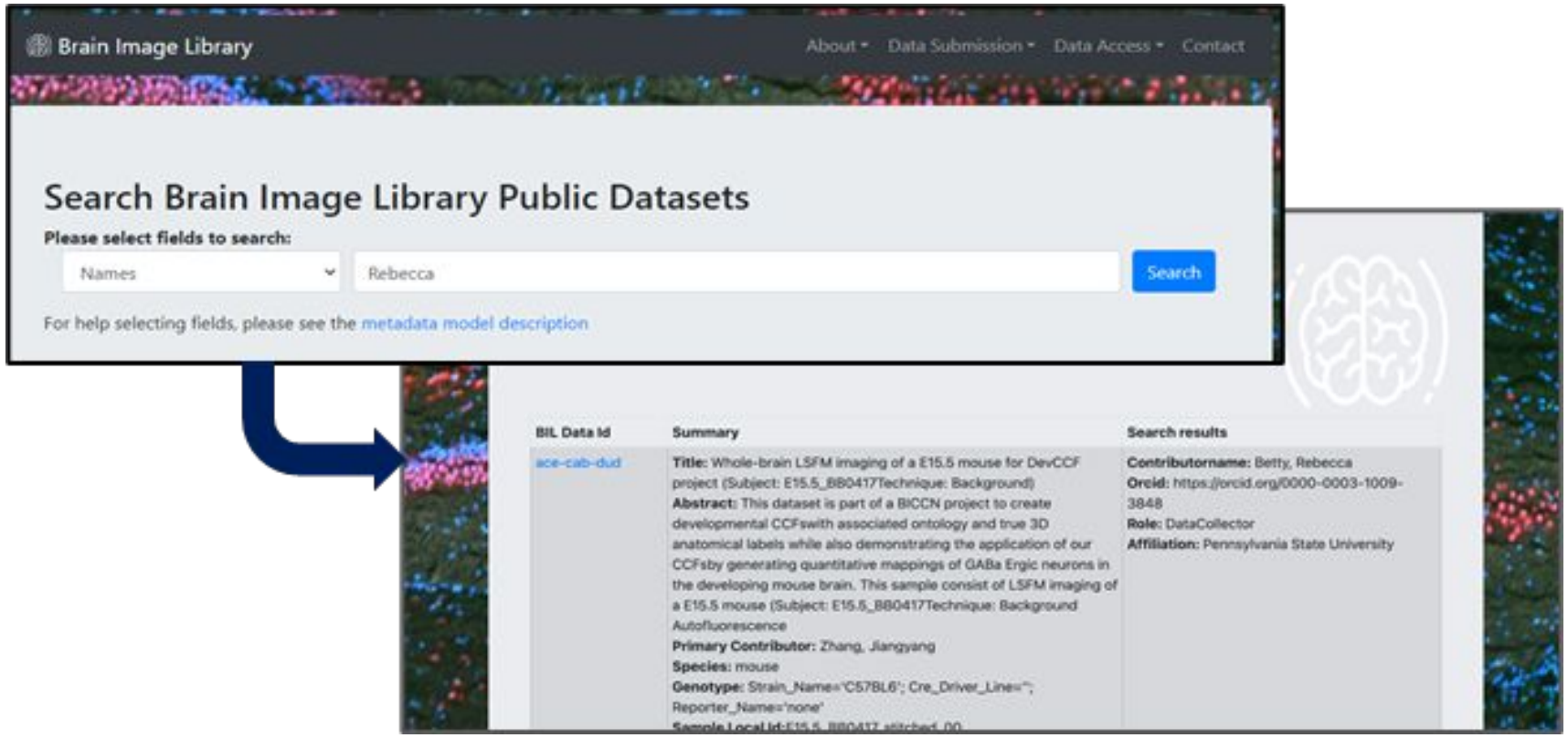
Kenney, M., Vasylieva, I., Hood, G. *et al.* The Brain Image Library: A Community-Contributed Microscopy Resource for Neuroscientists. *Sci Data* **11**, 1212 (2024). https://doi.org/10.1038/s41597-024-03761-8

Benninger, K. Hood, G..Simmel, D. *et al.* Cyberinfrastructure of a Multi-Petabyte Microscopy Resource for Neuroscience Research. In Practice and Experience in Advanced Research Computing 2020: Catch the Wave (PEARC '20). Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3311790.3396653

www.brainimagelibrary.org

Finding Spatial Data in BIL

Websearch: https://api.brainimagelibrary.org/web/



Or Metadata API (Programmatic Access): https://www.brainimagelibrary.org/metadataapi.html

Finding Spatial Data in BIL

Websearch: https://api.brainimagelibrary.org/web/

Web Search Field	Search Text	Number of Entries
All Metadata	Spatial	710
All Dataset Fields	Spatial	516
All Dataset Fields	Spatial Transcriptomics	436
All Dataset Fields	MERFISH	106
All Dataset Fields	RNA FISH	26
All Dataset Fields	smFISH	229
All Dataset Fields	seqFISH	8
All Dataset Fields	FISH	119
All Dataset Fields	CISI	12

Dataset Location - Noted on the Landing Page

Data:

Visualization Links:

Help with Visualization
https://brainapi.brainimagelibrary.org/ng/bil/assets/ace/cap/bog/asset/brainpi/mouseID_454427-192343.terafly

Data location on the Brain Image Library Analysis Ecosystem:

/bil/data/2f/12/2f12b1a3901e4181/mouseID_454427-192343

Dataset download link:

https://download.brainimagelibrary.org/2f/12/2f12b1a3901e4181/mouseID_454427-192343

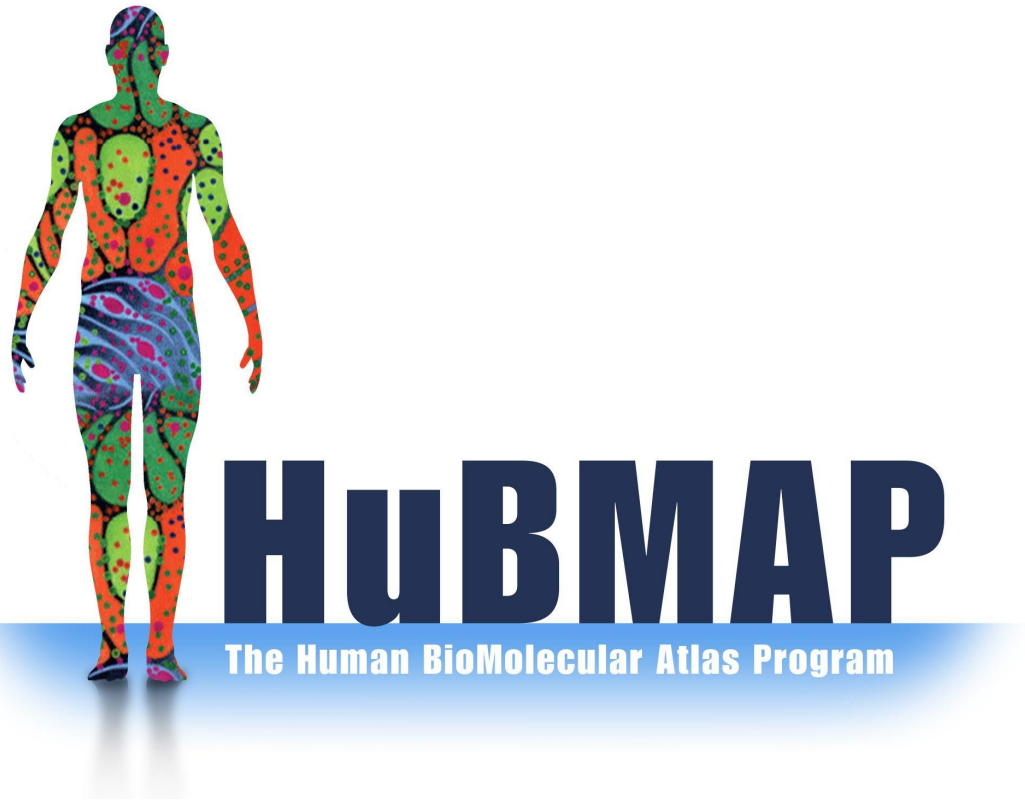
Metadata download link:

https://api.brainimagelibrary.org/retrieve?bildid=ace-cap-bog

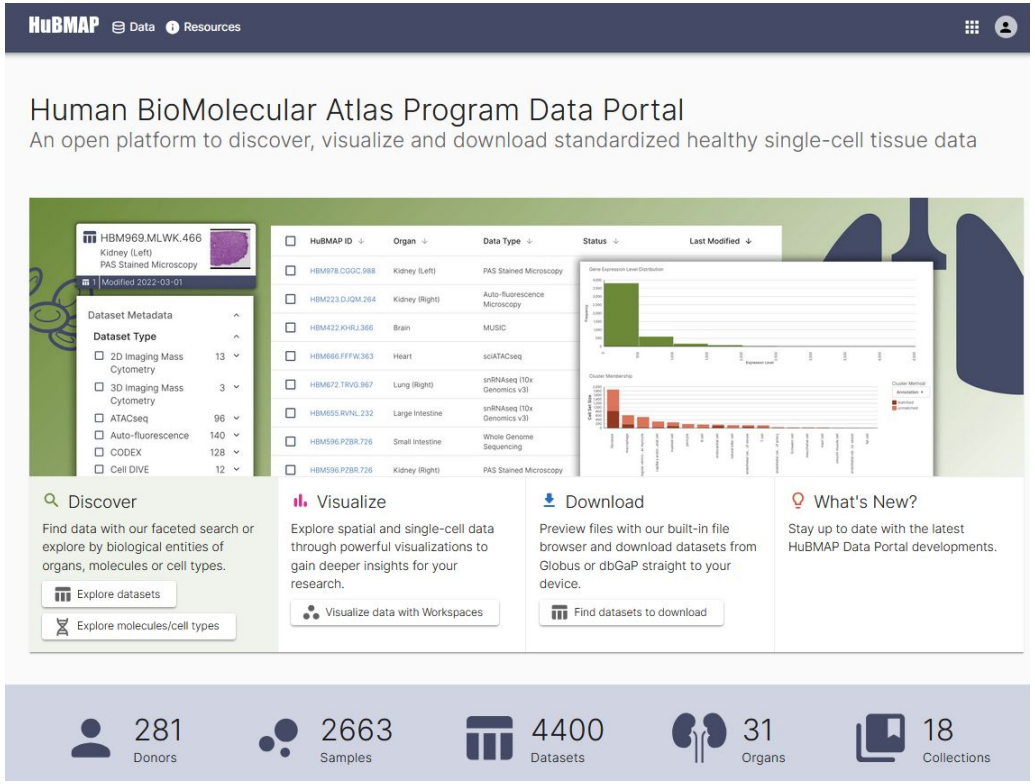
Manifest download link:

https://download.brainimagelibrary.org/inventory/datasets/c3f113f2-e61c-565d-8158-ab630bdcd03f.json

Part 2: The Human Biomolecular Atlas Program (HuBMAP)



What is HuBMAP?



The HuBMAP Data Sharing Portal.

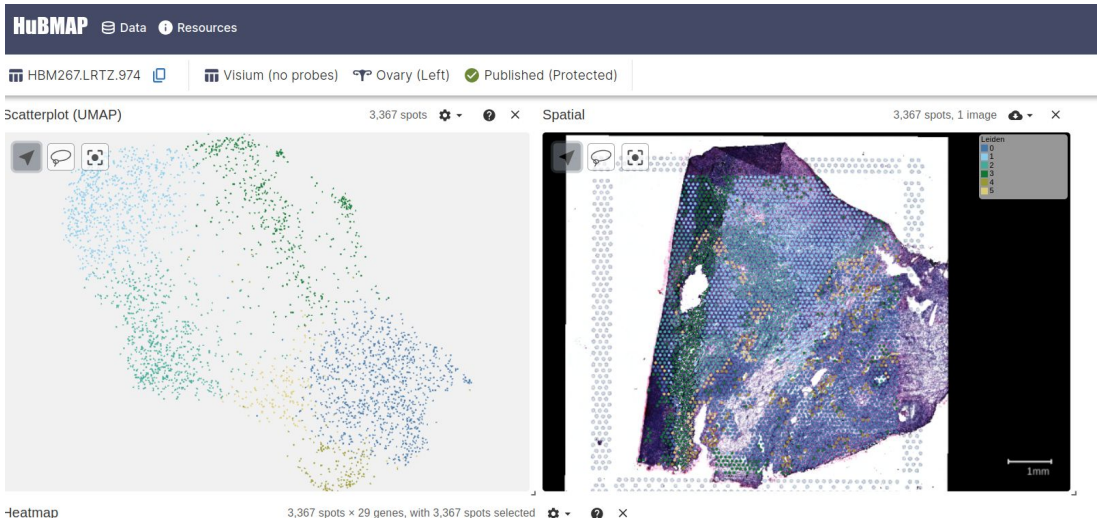
The Human Biomolecular Atlas Program (HuBMAP) is dedicated towards mapping the human body at the single-cell level. It is funded through the National Institutes of Health Common Fund.

- 42 contributing sites
  - HuBMAP Integration, Visualization, and Engagement (HIVE) group
    - Includes the Infrastructure and Engagement Component (IEC) - HuBMAP's "data core"
  - Tissue Mapping Centers (TMCs)
  - Transformative Technology Developers (TTDs)
- TMCs work with a variety of experimental assays on tissue samples from 31 organs throughout the human body

Classifying HuBMAP

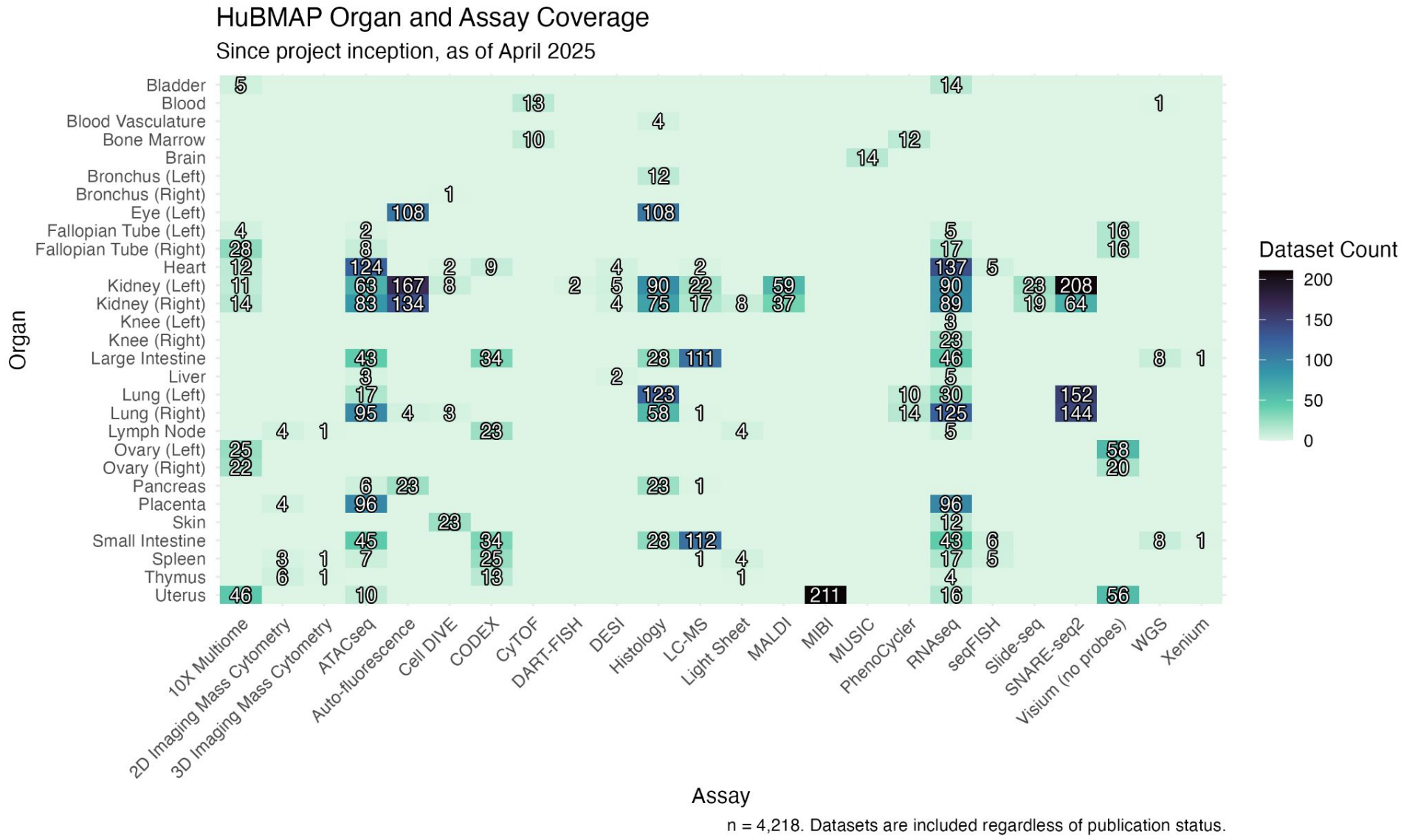
HuBMAP is...

- **A data repository**
  - HuBMAP ingests, validates, and publishes a wide array of data from different modalities. Some of the data types stored include the following:
    - Spatial transcriptomics - Xenium, Visium, GeoMX
    - Transcriptomics - sc/snRNAseq, bulk RNAseq, SNAREseq2
    - Antibody-based imaging - Phenocycler, CODEX, Cell DIVE
    - Microscopy - H&E, PAS, AB-PAS, Autofluorescence
- **A knowledge base**
  - HuBMAP develops tools and resources for public use, including metadata and directory schemas, the Human Reference Atlas, Vitessce, and Azimuth.



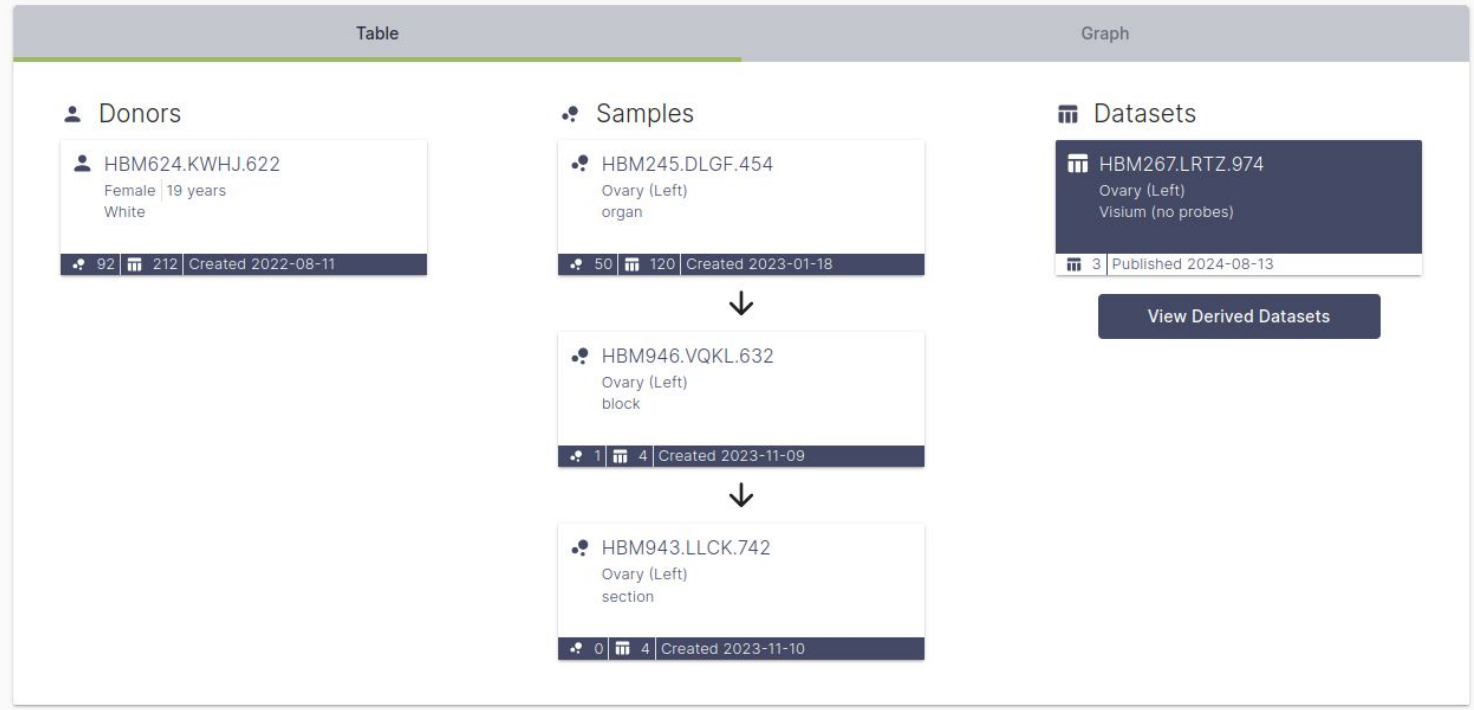
HIVE central analysis (Salmon + Scanpy) of a uterine Visium (no probes) dataset from TMC-University of Pennsylvania (HBM267.LRTZ.974)

HuBMAP Data Coverage



Entity Provenance in HuBMAP

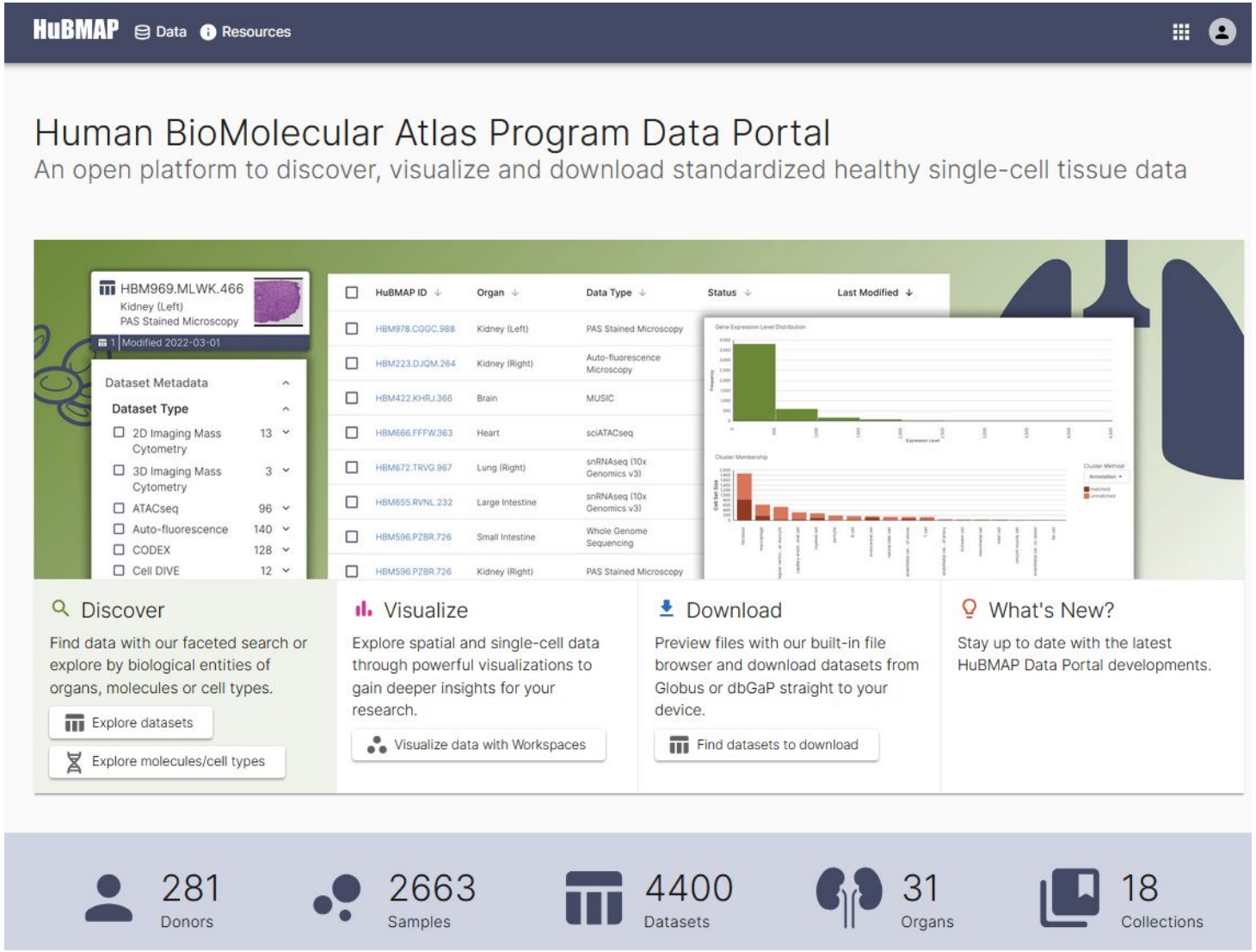
- Entities in HuBMAP have unique identifiers starting with “HBM”
- Common provenance
  - Donors registered first
  - Organs registered to donors
  - Tissue blocks registered to organs
  - Tissue sections/suspensions registered to blocks
  - Datasets registered to sections/suspensions
- Donors, tissue blocks/sections/suspensions, and datasets can all have metadata



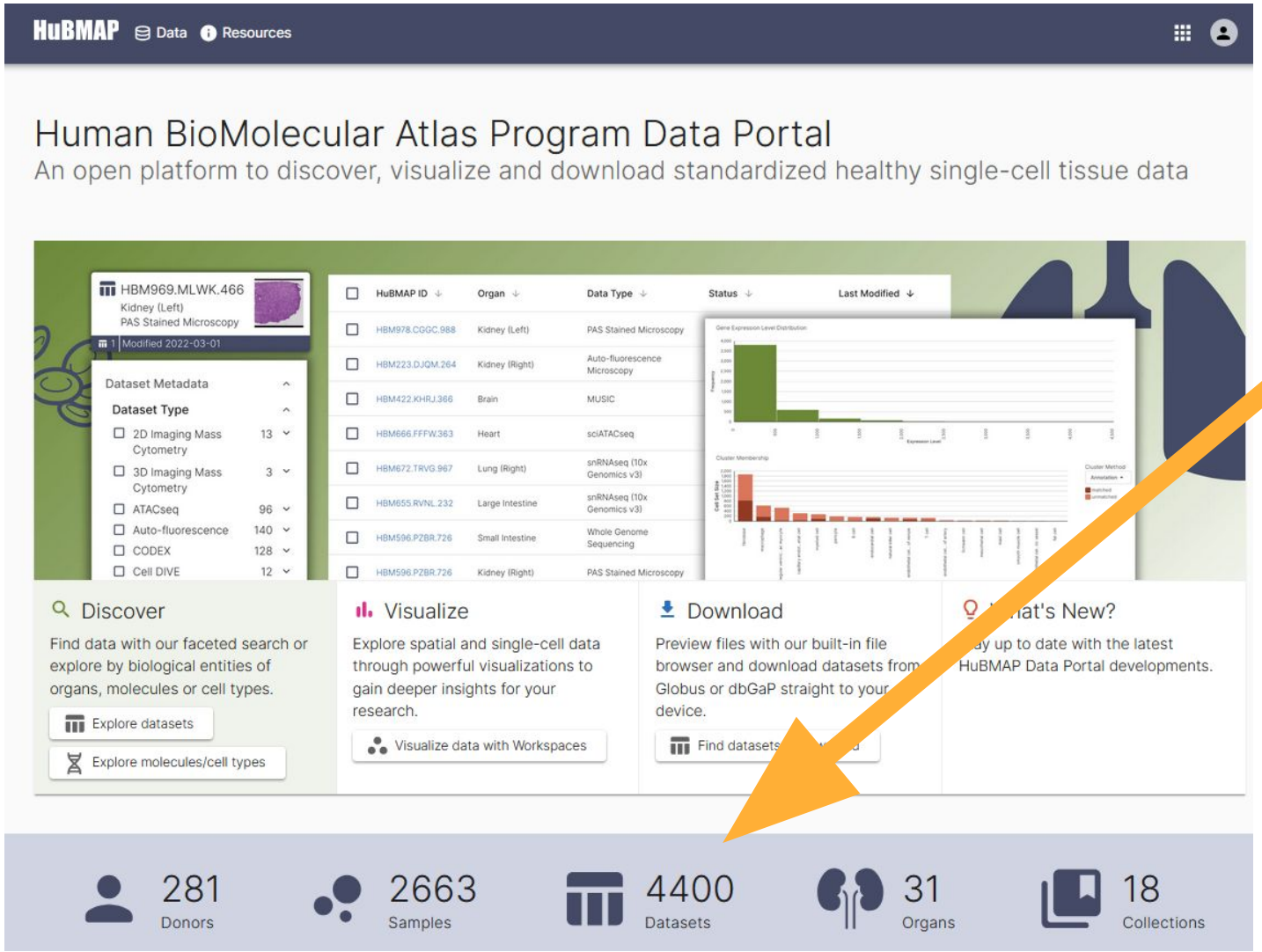
How Do I Find Spatial Transcriptomics Data in the HuBMAP Portal?

In this example, we are going to look for Visium (no probes) data in the HuBMAP Data Portal.

First, go to [portal.hubmapconsortium.org.](portal.hubmapconsortium.org)



How Do I Find Spatial Transcriptomics Data in the HuBMAP Portal?



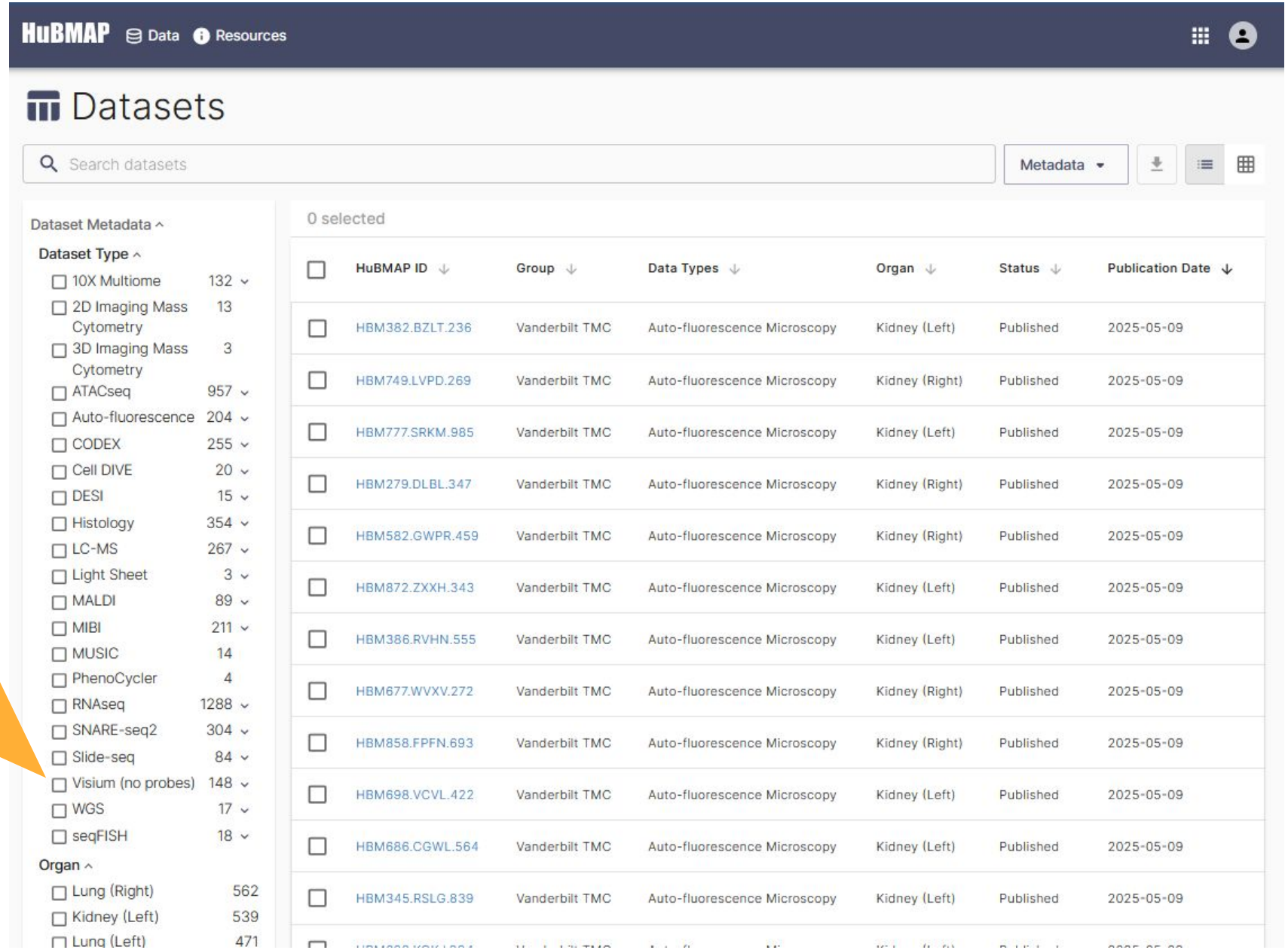
Click on “Datasets.”

How Do I Find Spatial Transcriptomics Data in the HuBMAP Portal?

The datasets search page has multiple facets with which a user can filter datasets.

Published datasets can be browsed without logging into the portal. Further details on portal access/login will be covered in the HuBMAP workspaces talk later this afternoon.

For this example, we will click on “Visium (no probes).”



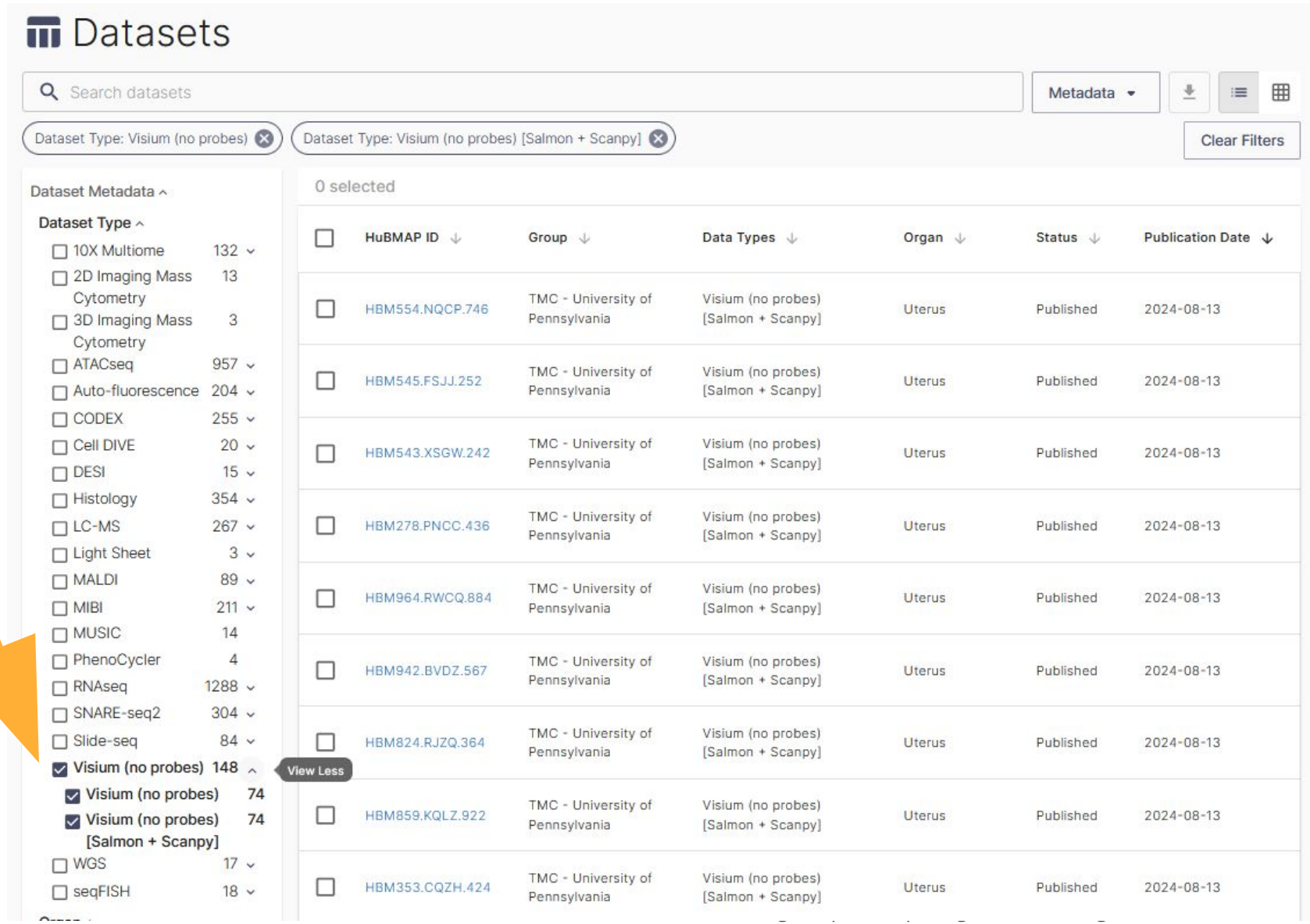
How Do I Find Spatial Transcriptomics Data in the HuBMAP Portal?

When we select “Visium (no probes),” both primary datasets and processed datasets are automatically selected.

Among the multiple data classes, there are two main ones, primary and processed.

Primary datasets contain the raw data from an experimental modality.

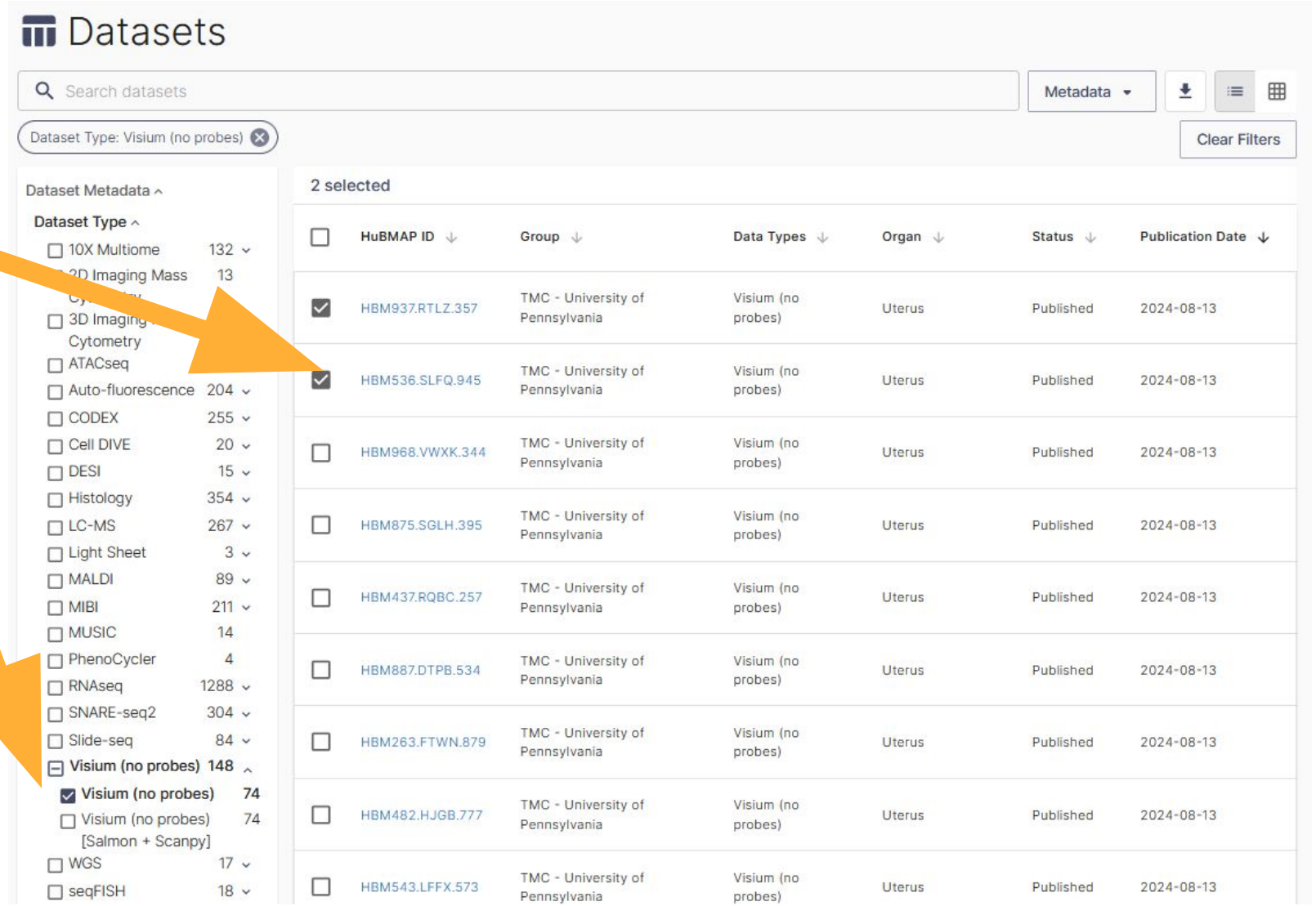
Processed datasets feature further analysis performed on primary datasets using standardized HIVE pipelines. These can be identified by the name of the pipeline in square brackets, such as “[Salmon + Scanpy].” Processed datasets are displayed on their “parent” primary datasets via a unified view.



How Do I Find Spatial Transcriptomics Data in the HuBMAP Portal?

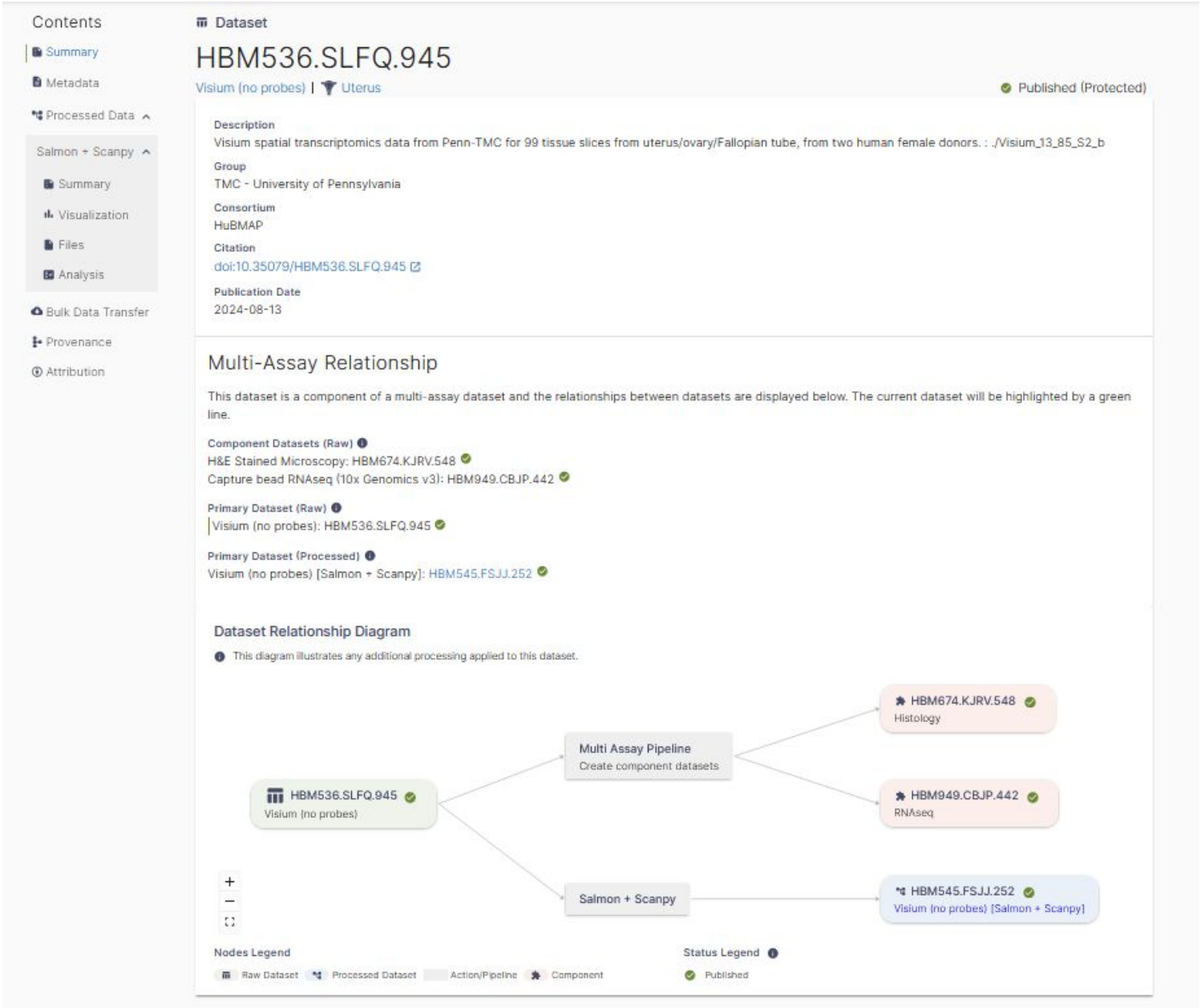
For now, let's filter by raw "Visium (no probes)" datasets.

We will take a look at the second dataset listed, HBM536.SLFQ.945.



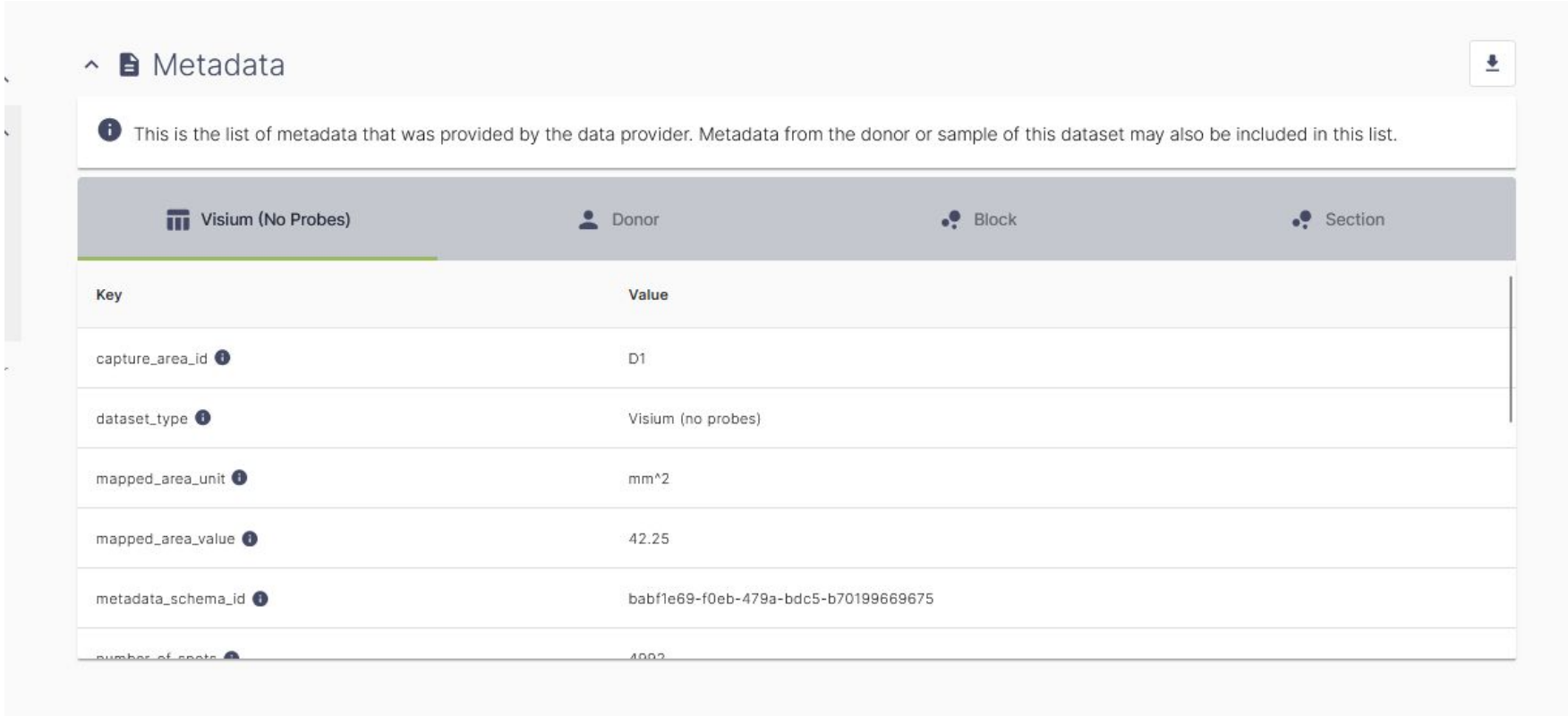
How Do I Find Spatial Transcriptomics Data in the HuBMAP Portal?

At the top of a dataset's portal page, we can see a brief description, its DOI, and an overview of the different dataset classes associated with it in HuBMAP's internal graph database.



How Do I Find Spatial Transcriptomics Data in the HuBMAP Portal?

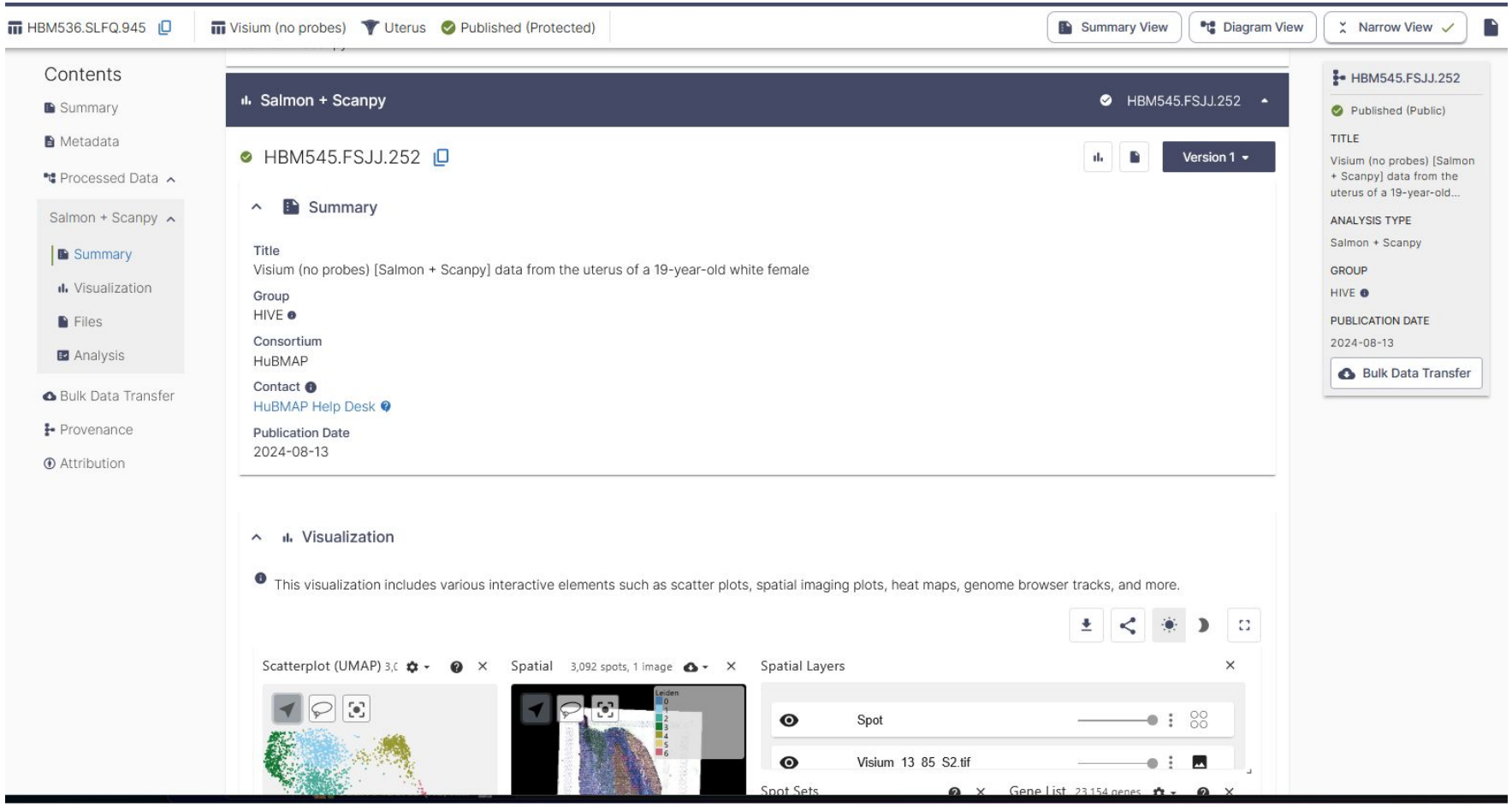
We can browse metadata for different entities in the dataset's provenance chain.



How Do I Find Spatial Transcriptomics Data in the HuBMAP Portal?

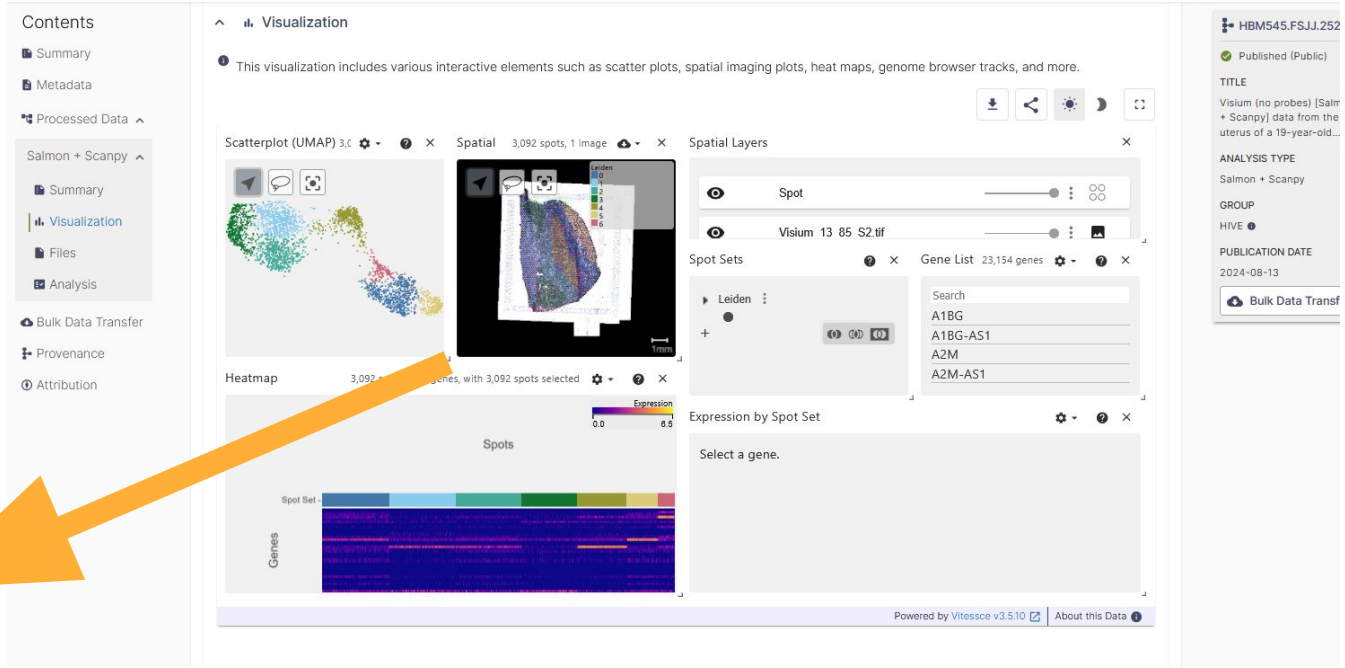
Any processed datasets are embedded on the primary dataset's page.

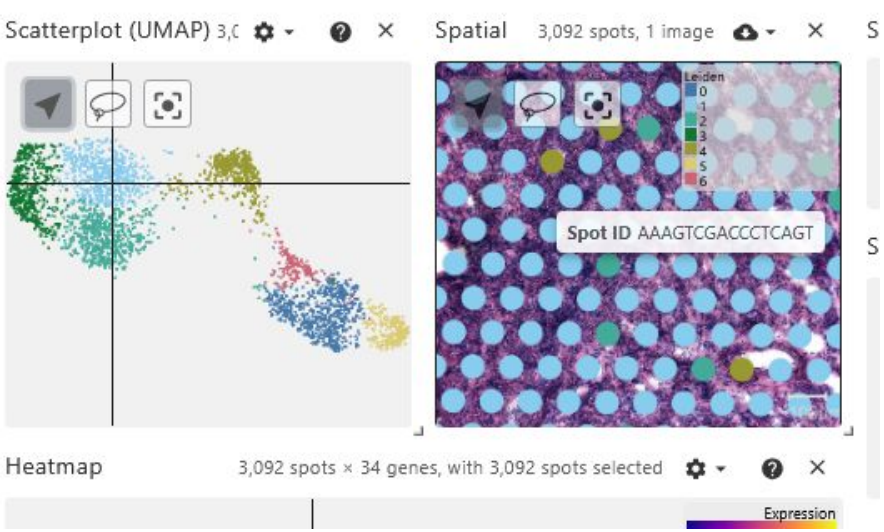
If you select a processed dataset from the search results, it will redirect you to its primary dataset's page.



How Do I Find Spatial Transcriptomics Data in the HuBMAP Portal?

An interactive visualization of the HIVE central analysis pipeline results is displayed via Vitessce.





HuBMAP Data for This Workshop

- For this workshop, HuBMAP TMC-Rochester has provided unpublished data:
  - https://app.globus.org/file-manager?origin_id=af603d86-eab9-4eec-bb1d-9d26556741bb&origin_path=%2Fspatial-data-workshop%2FTMA2%2F&two_pane=true
-

Other Helpful Resources

- List of HuBMAP metadata and directory schemas for assays
  - https://docs.hubmapconsortium.org/metadata
- HuBMAP donor LineUp
  - https://portal.hubmapconsortium.org/lineup/donors
- HuBMAP APIs + examples
  - https://docs.hubmapconsortium.org/apis