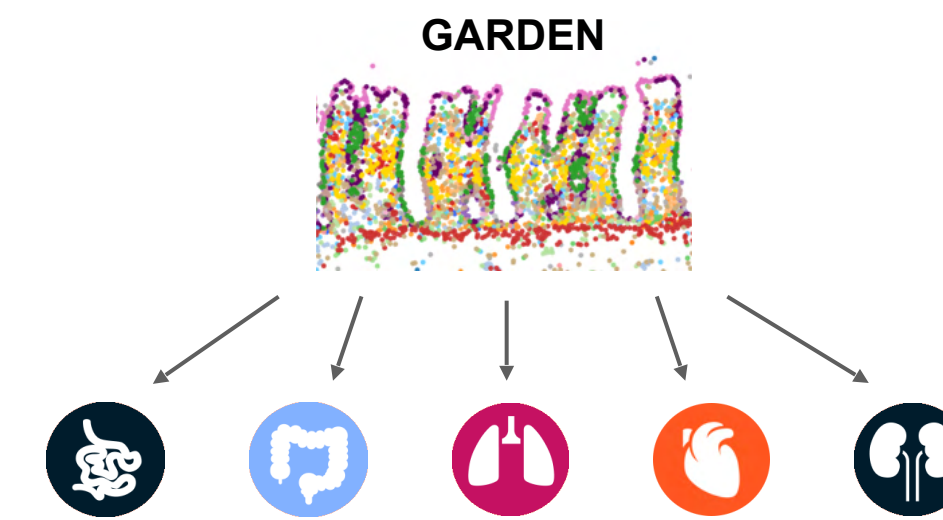


INTRODUCTION

To date, HuBMAP has produced over 4,300 datasets across 31 organs. However, the lack of standardized cell type annotations remains a major barrier to downstream biological interpretation and multi-organ analysis. To address this, we developed GARDEN (Generalized Annotation pipeline for Resolving DiversE cell types in multi-organ spatial omics datasets), a framework for data preprocessing, clustering, visualization, and cell type annotation.

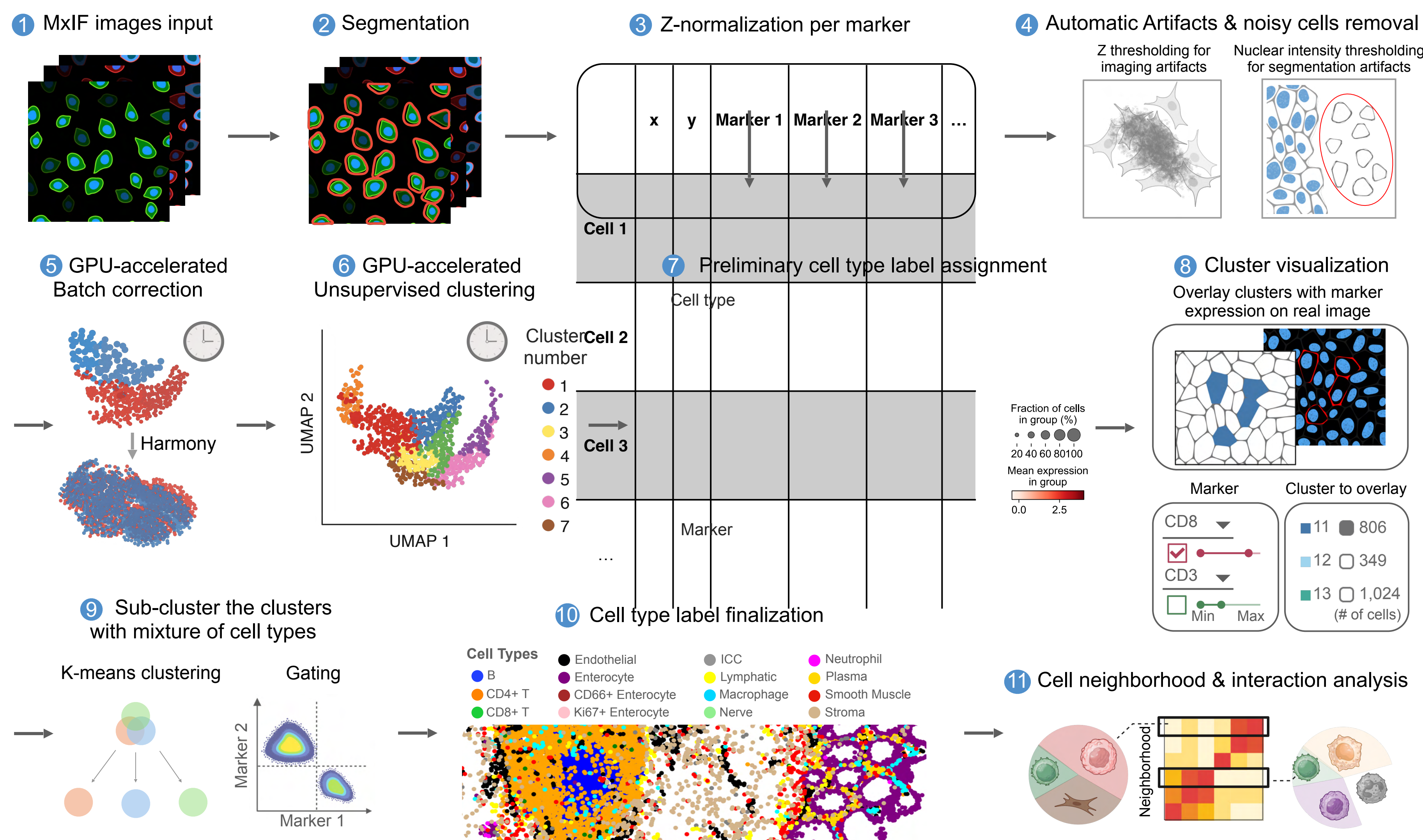
We aim to:

- (1) To provide a user-friendly Workspace template for the community.
- (2) To generate cell type annotations for multiple organs from both data portal and SWAT, to support integrative, multi-organ spatial analyses.



ANNOTATION PIPELINE

GARDEN - a Generalized Annotation pipeline for Resolving DiversE cell types in multi-organ spatial omics datasets



Workspace

Template notebook outline

- Workspace Setup
 - Check GPU and import libraries
 - Utility functions
- Download data
 - Download and save data locally
 - Download the data from SWAT
- Merge and cleanup datasets
 - Merge the files from same HuBMAP dataset
 - Standardize the marker names across all datasets
 - Get the unique names and identify duplicate marker names
 - Standardize the marker names
 - Identify the markers that are present in all datasets
- Z-normalization and preprocessing
 - Z-normalization
 - Pre-processing after merge and Z-norm
 - Step #1: total Z-score thresholding
 - Step #2: nuclear intensity filter
 - (Optional) Check lost cells in pre-processing
- Rapids Clustering
 - Save the dataframe in anndata
- Clustering
 - Batch Correction
 - Leiden Clustering
 - Plot UMAP
- Annotating Clusters
 - Dotplot visualization
 - Plot clusters
 - Cell type annotation

HuBMAP Datasets Specific Steps

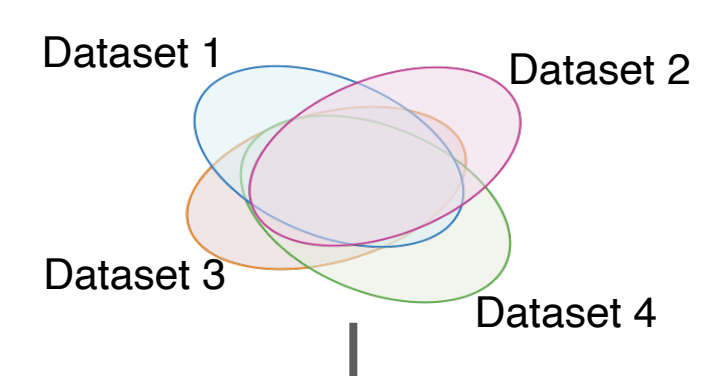
- Download HuBMAP spatial CODEX datasets (data portal / SWAT)



- Standardize marker names

"PDPN": "Podoplanin"
"Synapto": "Synaptophysin"
"aDef5": "aDefensin5"
"Col1IV": "Collagen IV"

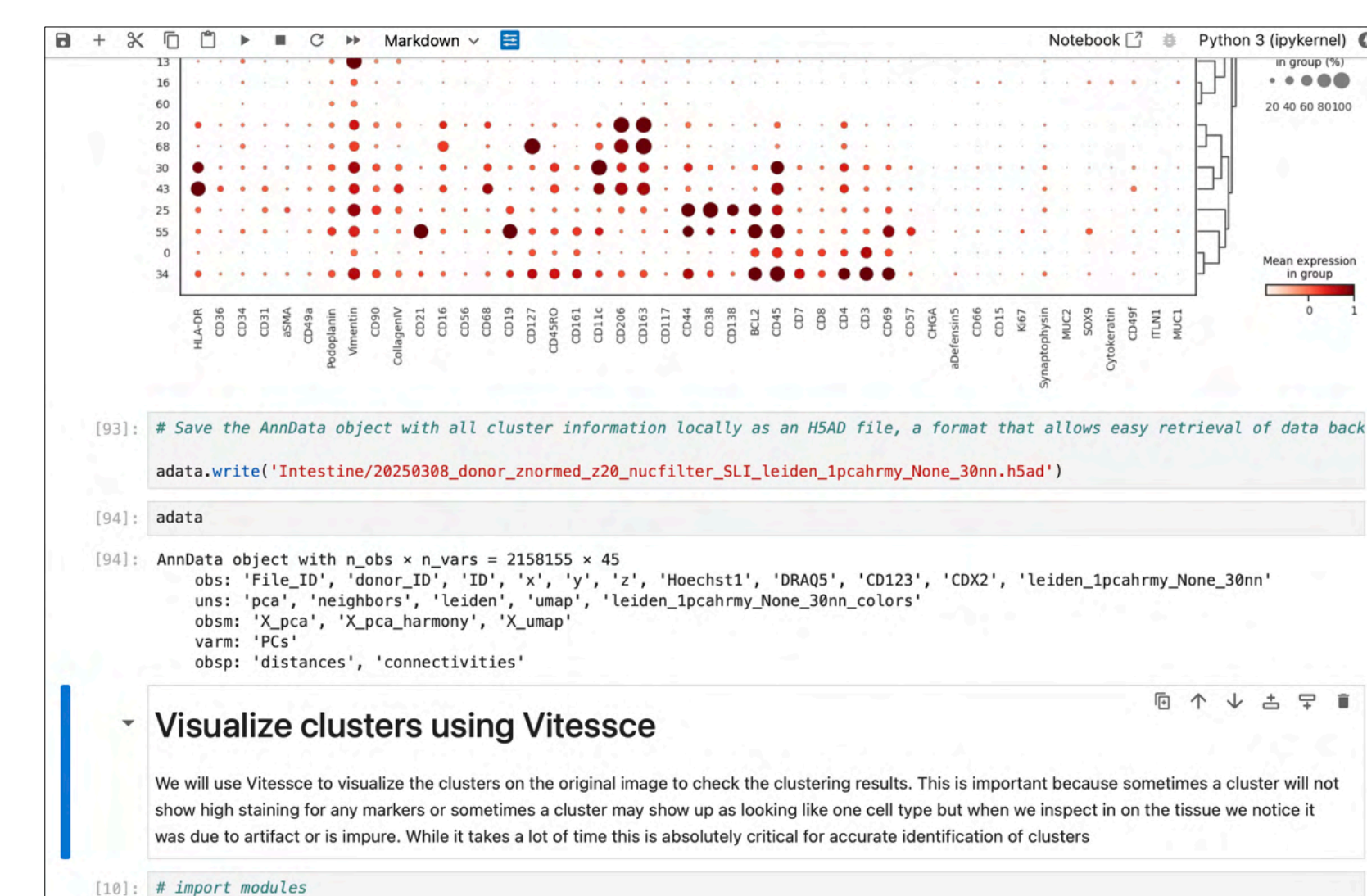
- Common marker identification



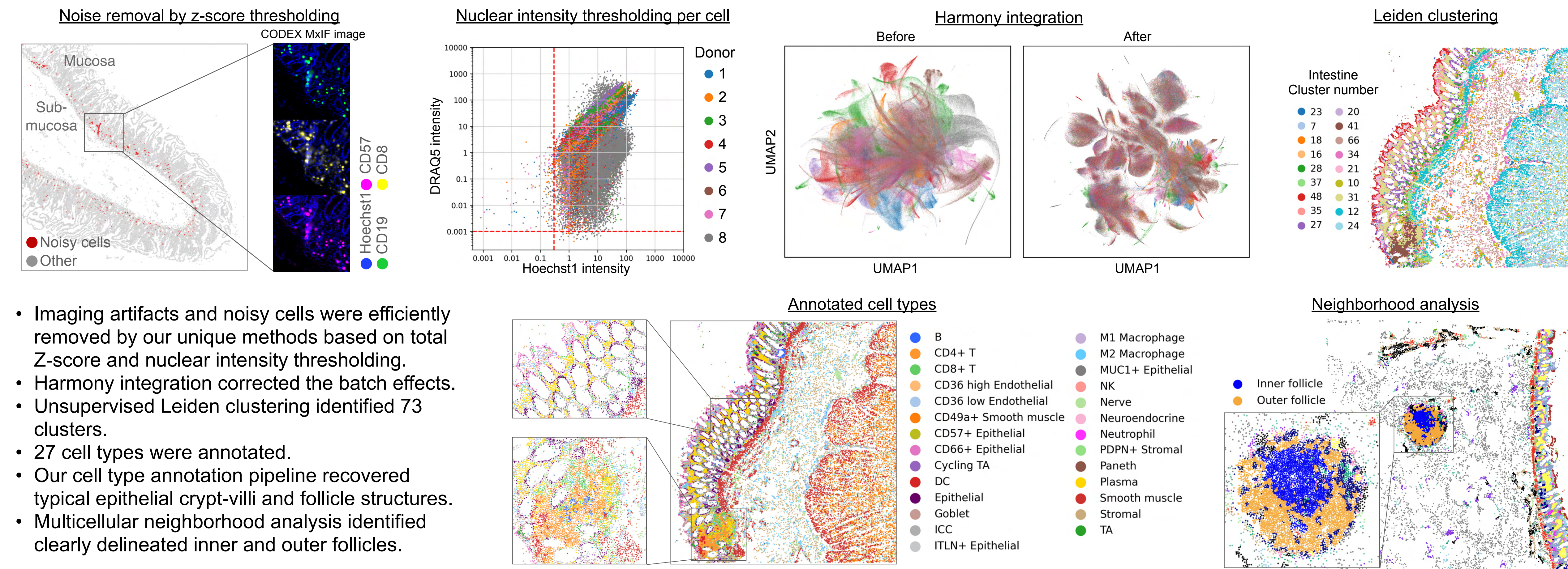
GARDEN annotation pipeline

- We developed a notebook template for cell type annotation that supports direct dataset extraction within the HuBMAP Workspace and is also compatible with datasets outside the data portal.
- Annotating cell types across multiple tissues requires additional steps beyond the general GARDEN pipeline, including standardizing marker names and defining core shared marker panels.

Snapshot of notebook template

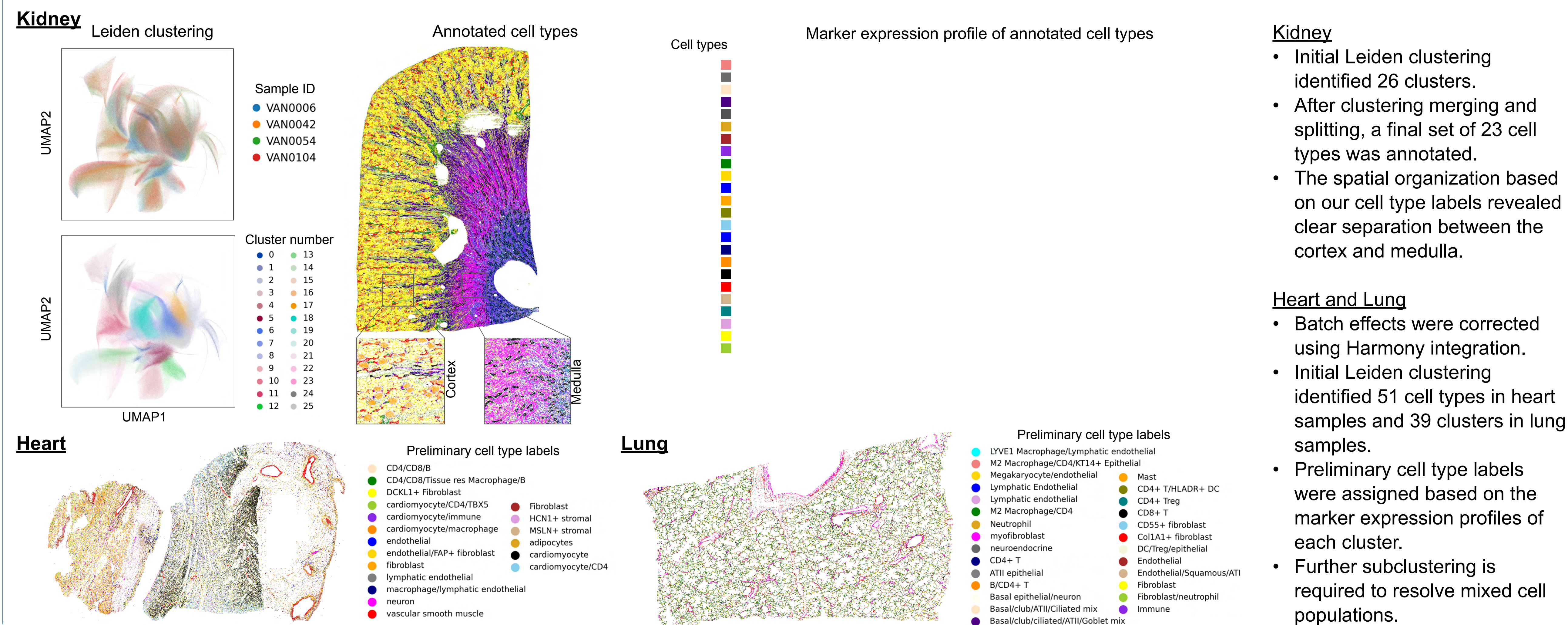


APPLICATION 1 – INTESTINE FROM DATA PORTAL



- Imaging artifacts and noisy cells were efficiently removed by our unique methods based on total Z-score and nuclear intensity thresholding.
- Harmony integration corrected the batch effects.
- Unsupervised Leiden clustering identified 73 clusters.
- 27 cell types were annotated.
- Our cell type annotation pipeline recovered typical epithelial crypt-villi and follicle structures.
- Multicellular neighborhood analysis identified clearly delineated inner and outer follicles.

APPLICATION 2 – SWAT DATASETS



Kidney

- Initial Leiden clustering identified 26 clusters.
- After clustering merging and splitting, a final set of 23 cell types was annotated.
- The spatial organization based on our cell type labels revealed clear separation between the cortex and medulla.

Heart and Lung

- Batch effects were corrected using Harmony integration.
- Initial Leiden clustering identified 51 cell types in heart samples and 39 clusters in lung samples.
- Preliminary cell type labels were assigned based on the marker expression profiles of each cluster.
- Further subclustering is required to resolve mixed cell populations.

FUTURE WORK

- We will adapt our cell type annotation pipeline to additional organs and generate standardized cell type labels.
- These standardized labels will be used to train STELLAR (SpaTial cEIL LeARning), a deep learning tool for transferring cell type annotations to other HIVE-processed datasets.
- We will implement multi-organ hierarchical cellular neighborhood analysis to explore cellular organization and tissue architecture across the human body.
- We will collaborate with the HRA visualization team to integrate hierarchical tissue labels into 3D visualizations.

ACKNOWLEDGMENTS

This project was supported by NIH 3OT2OD033759-01S4 and NIH 3U54-AG075936-03S1. We gratefully acknowledge all our collaborators, Matt Ruffalo, Phil Blood, Gloria Pryhuber, Tarek Ashkar, Angela Sabo, Sanjay Jain, Thai Pham, Melissa Farrow, Nils Gehlenborg, Peiran Lu, Liming Pei, David Van Valen, Katy Borner, Jean Fan, and all SWAT team members.



REFERENCES

- [1] Hickey, J. W. et al. Organization of the human intestine at single-cell resolution. Nature 619, 572–584 (2023).
- [2] Hickey, J. W., Tan, Y., Nolan, G. P. & Goltsev, Y. Strategies for Accurate Cell Type Identification in CODEX Multiplexed Imaging Data. Front. Immunol. 12, 727626 (2021).