

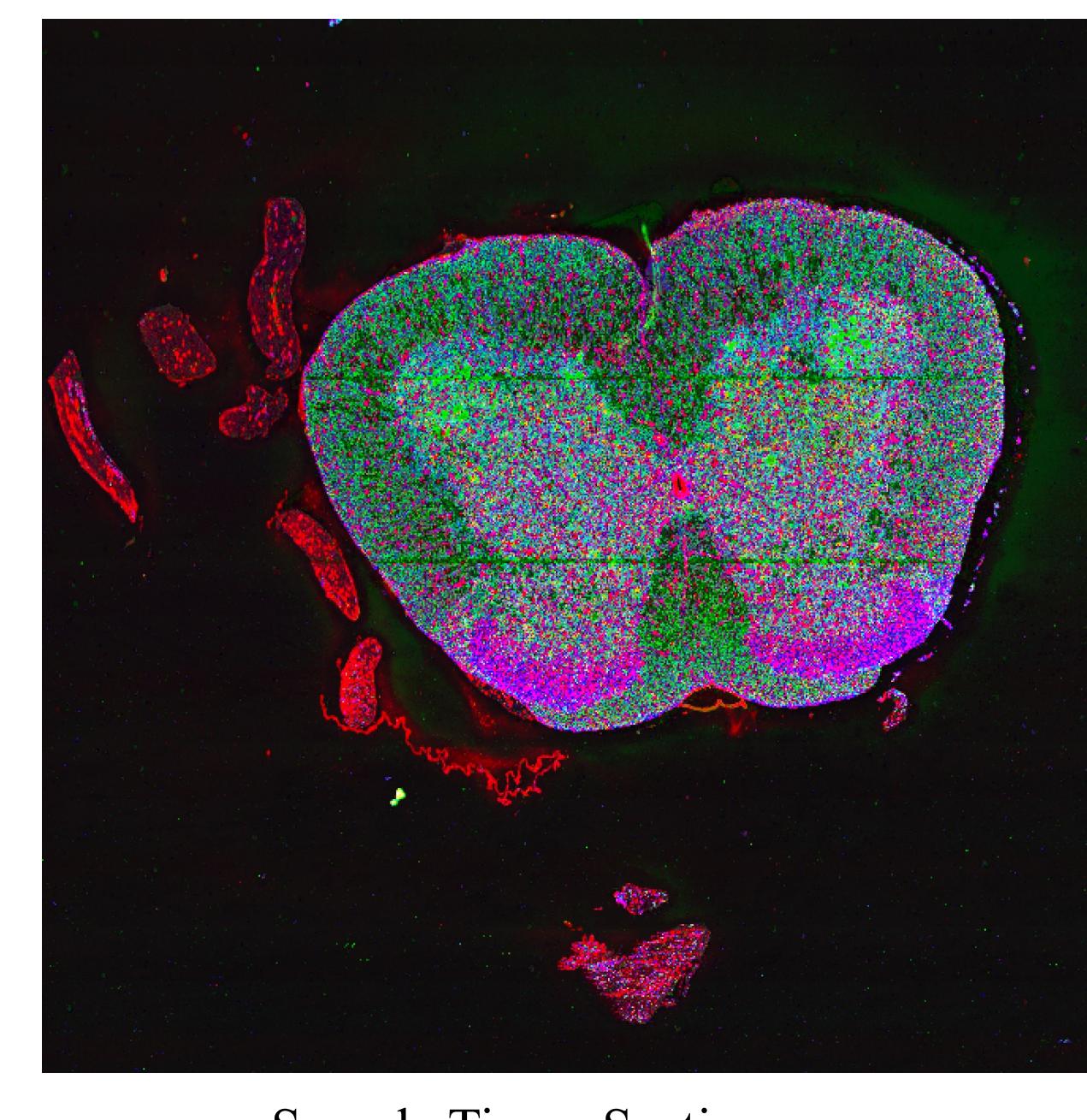
Hudson: A Computational Pipeline For Spatial Analysis Of Multiplexed Images

Jagjit Singh^{1,2}, Kunal Pandit², Rui Fu², Sanja Vickovic^{1,2}

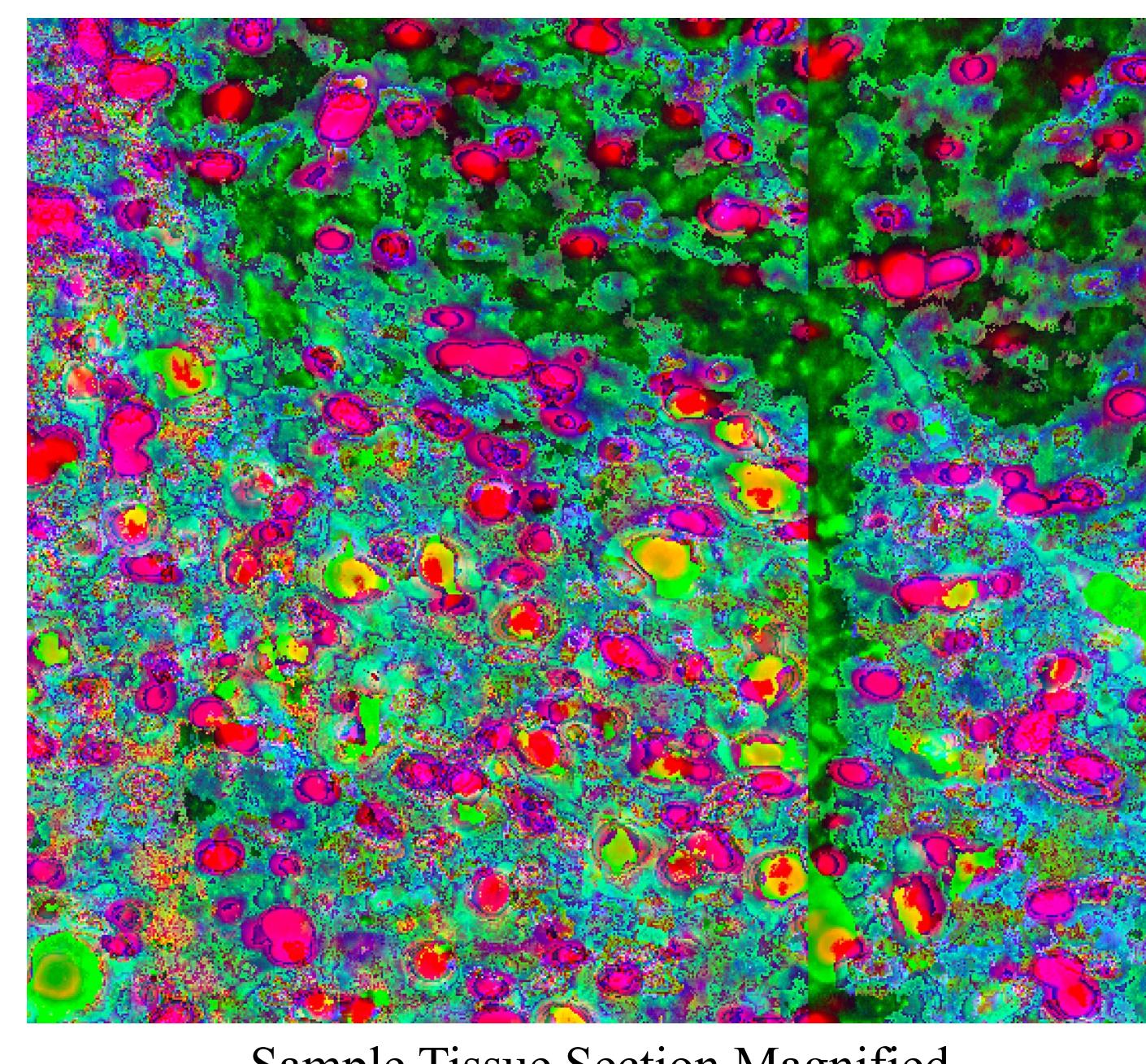
Columbia University in the City of New York¹, New York Genome Center²

ABSTRACT

Multiplexed imaging technologies are amongst the fastest growing areas of multi-omics. These technologies capture many parameters of single cells while preserving their spatial location. As such, there is a need for a flexible, robust, and easily deployed end-to-end computational pipeline for processing and analysis of the data such technologies produce. The goal of such a pipeline is to capture cell-type compositions in a localized shape-preserving manner. To help researchers achieve this goal, we have developed Hudson. It is a fully automated computational pipeline for spatial analysis of multiplexed images obtained primarily via PySeq2500, which is an open-source toolkit for repurposing the HiSeq2500 sequencing system as a versatile fluidics and imaging platform. Hudson preprocesses and segments the multiplexed image to identify micro-environments within the tissue section and performs spatial analysis upon the computed micro-environments. Hudson uses the Snakemake workflow management system and can be scaled up from a personal computer to a research computing cluster for heavy workloads. The pipeline begins with the essential pre-processing of the image. This includes background correction, registering channels, stitching the image tiles together, and removing overlapping regions. As part of preprocessing, Hudson deploys the PICASSO algorithm to perform unmixing of signals from dyes with overlapping emission spectra. Once pre-processing is complete, the spatial analysis phase begins with instance segmentation to detect image labels and computation of mean intensity per label. The mean intensities for all available markers in each label are then used for cell type identification via clustering and comparison with a reference single cell sequencing dataset. The cell type identifiers and other image properties including the labels and label centroids are output as an Anndata object. Finally, Hudson extracts data from the Anndata object to build local spatial environments, compute relevant spatial statistics, and build a cell connectivity graph for researchers to use in their own downstream analysis. Importantly, the end result is consistent with the input data regardless of the underlying system architecture on which Hudson is deployed. Given the computationally intense and programmatically advanced nature of analyzing highly multiplexed images, Hudson saves significant time for researchers wanting to incorporate spatial information and serves the overarching goal of making spatial analysis more accessible.



Sample Tissue Section

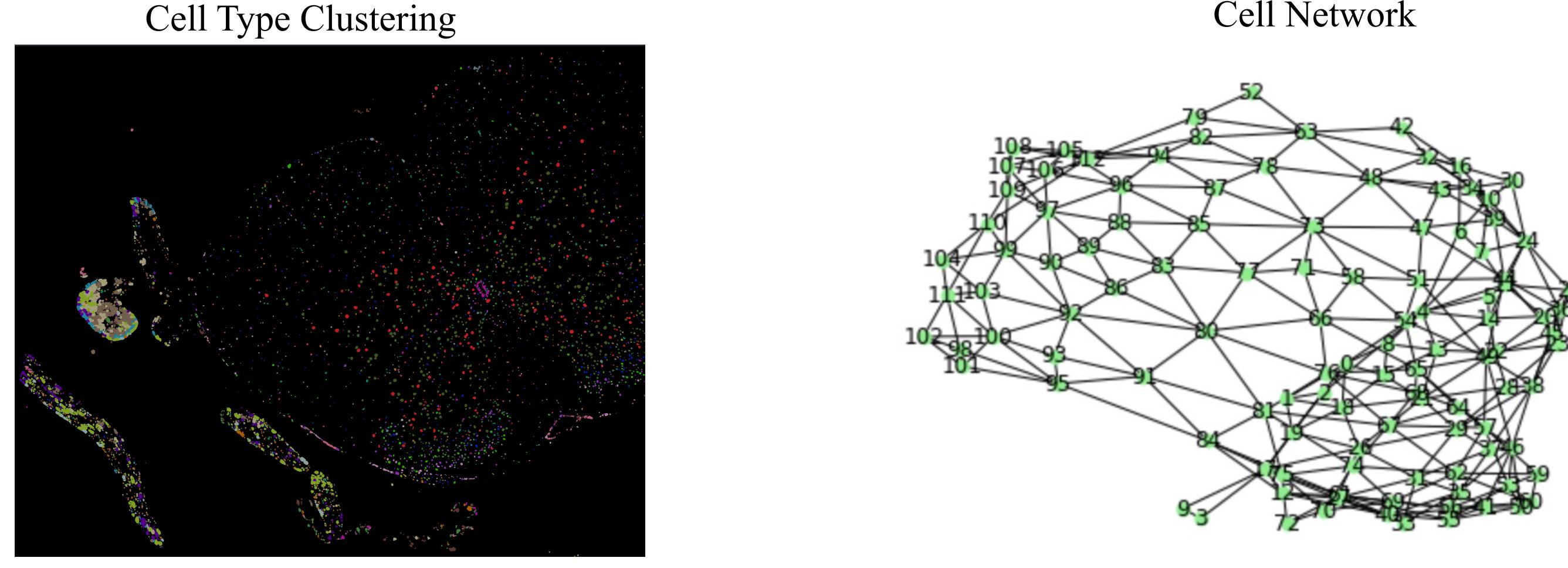
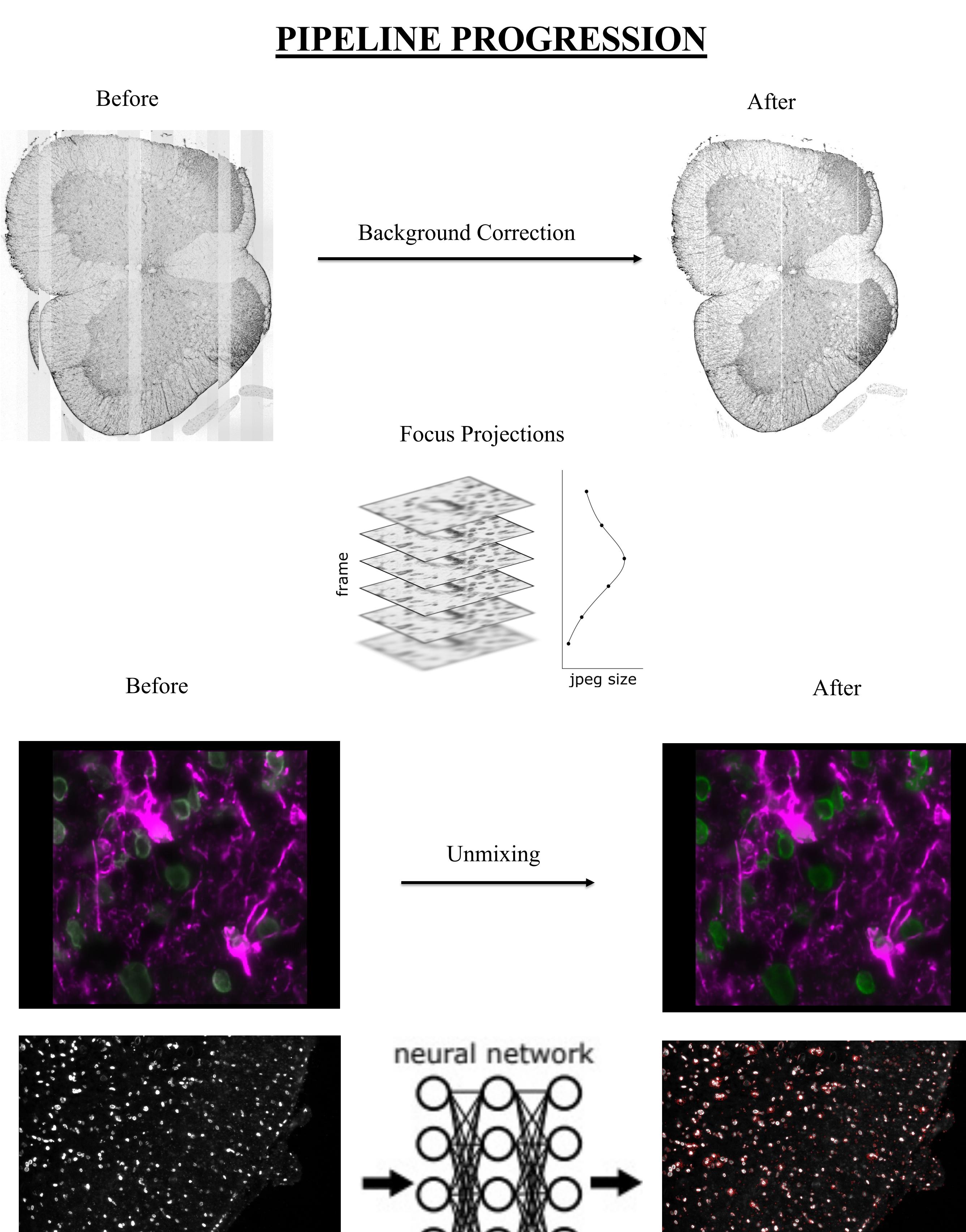


Sample Tissue Section Magnified

WORKFLOW

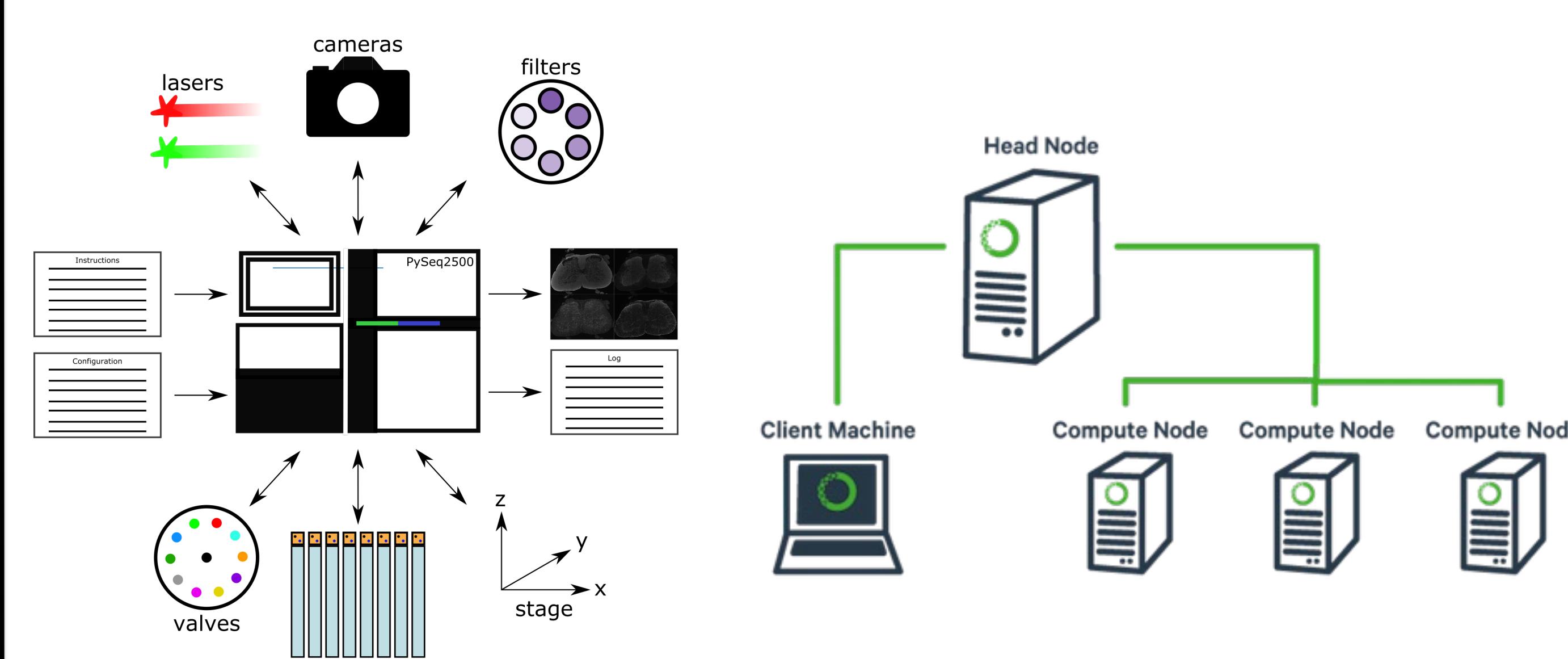
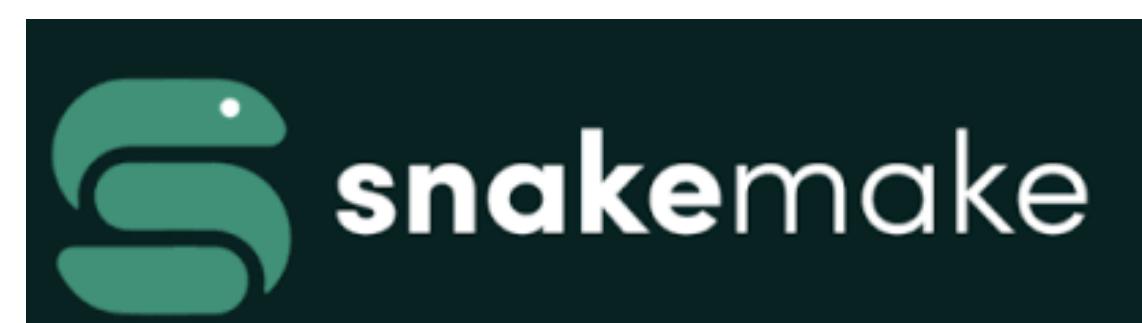
The main objective of the pipeline is to develop microenvironments over the observed tissue which is achieved through multiple individual steps, the steps being:

- 1) **Save Raw:** The preprocessing begins with converting the unprocessed images to a Zarr store. It can also be used as a jumping point to the segmentation step.
- 2) **Fix Lighting:** This step performs background correction, registering channels, stitching the image tiles together, and removing overlapping regions.
- 3) **PICASSO:** PICASSO is an algorithm to remove spillover fluorescence by minimizing the mutual information (MI) between sink and source images.
- 4) **Segmentation:** The pipeline performs instance segmentation on the image to generate a mask array of the same dimension as the input. Hudson performs segmentation based on both cell and nucleic markers.
- 5) **Intensity Computation:** This step in the pipeline computes mean intensities for all available markers in each label.
- 6) **Data Object:** In this step, the image labels along with their cell type composition and location are condensed and stored in an Anndata object.
- 7) **Cell Type Identification:** From protein mean intensity measurements, cells are clustered via the PhenoGraph algorithm and classified into basic cell types with Canonical Correlation Analysis in Seurat.
- 8) **Network Graph:** Hudson builds an undirected graph from cell connectivity based on a Delaunay triangulation.
- 9) **Microenvironments:** Finally, Hudson creates micro-environments based on the Delaunay Triangulation by condensing nodes to get larger triangulations and recording cell-type compositions within those regions.

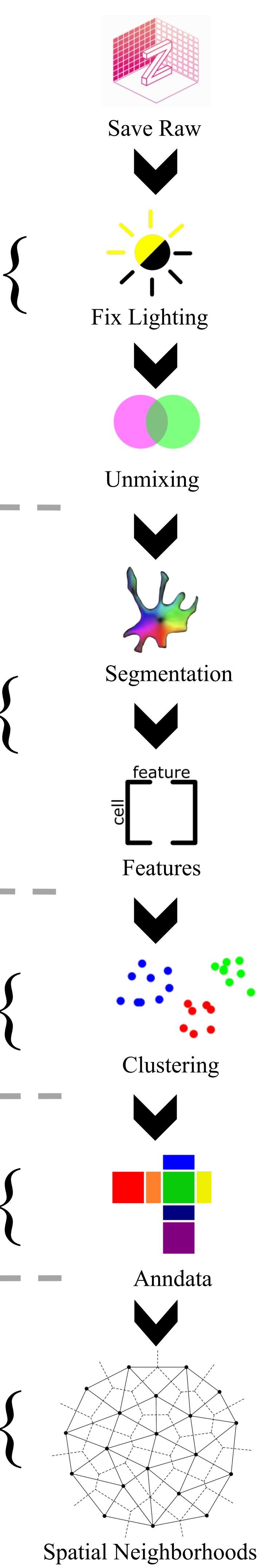
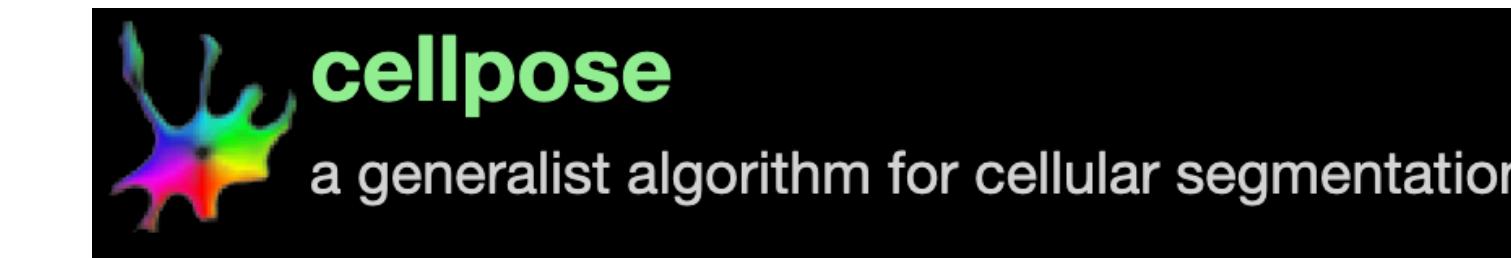
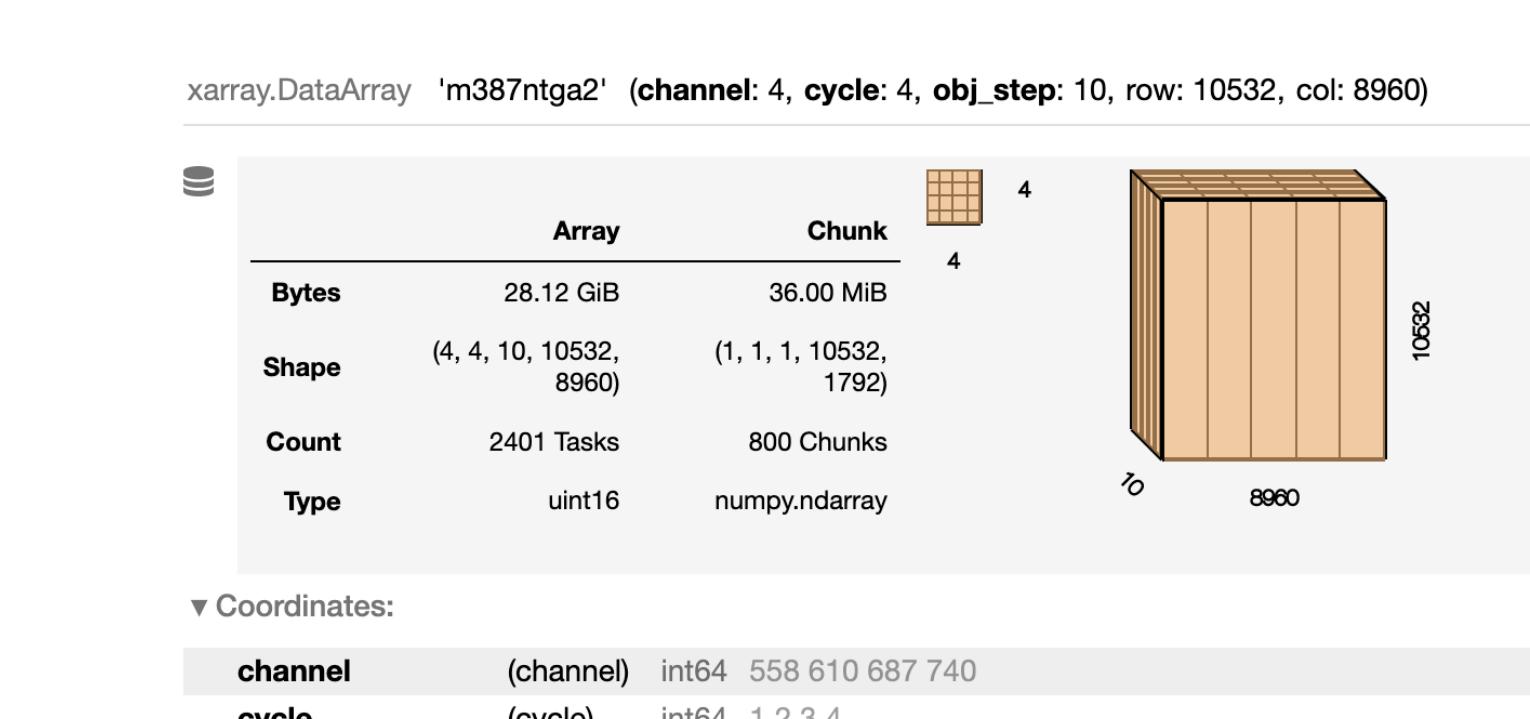


Microenvironment Compositions							
	Astocytes	Cholinergic neurons	Endothelial cells	Excitatory neurons	Inhibitory neurons	Microglia/Macrophages	Oligodendrocytes
XUNCMB	2	0	0	0	0	0	2
QXLBYZ	2	0	1	0	1	0	0
ZHIGIN	3	0	2	0	2	0	1

SYSTEM REQUIREMENTS



METHODOLOGY



REFERENCES

- Pandit K. & Petrescu, J. et al DOI: 10.1038/s41598-022-08740-w (2022)
- Stringer, C. & Pachitariu, M DOI: 2022.04.01.486764 (2022)
- Hao Y & Hao S, et al DOI: 10.1016/j.cell.2021.04.048
- Uwitonze A, et al DOI: 10.1371/journal.pone.0193350

ACKNOWLEDGEMENTS

- Silas Maniatis
- Nicolas Robine

- Joana Petrescu
- Hemali Phatani



Chan
Zuckerberg
Initiative