

Everything as Code

David Van Valen MD, PhD

SITC Computational Immuno-Oncology Webinar Series

07/18/2023



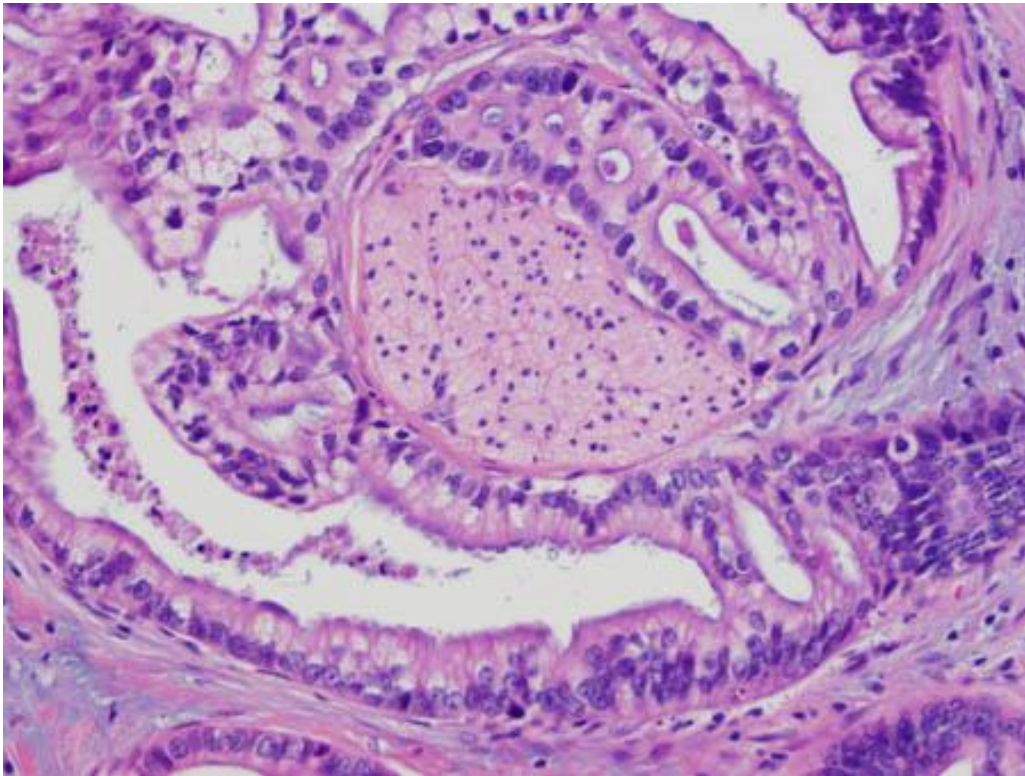
Disclosures

- I'm currently the Chief Scientist for Barrier Biosciences
- Many of the research results I will discuss today are unpublished

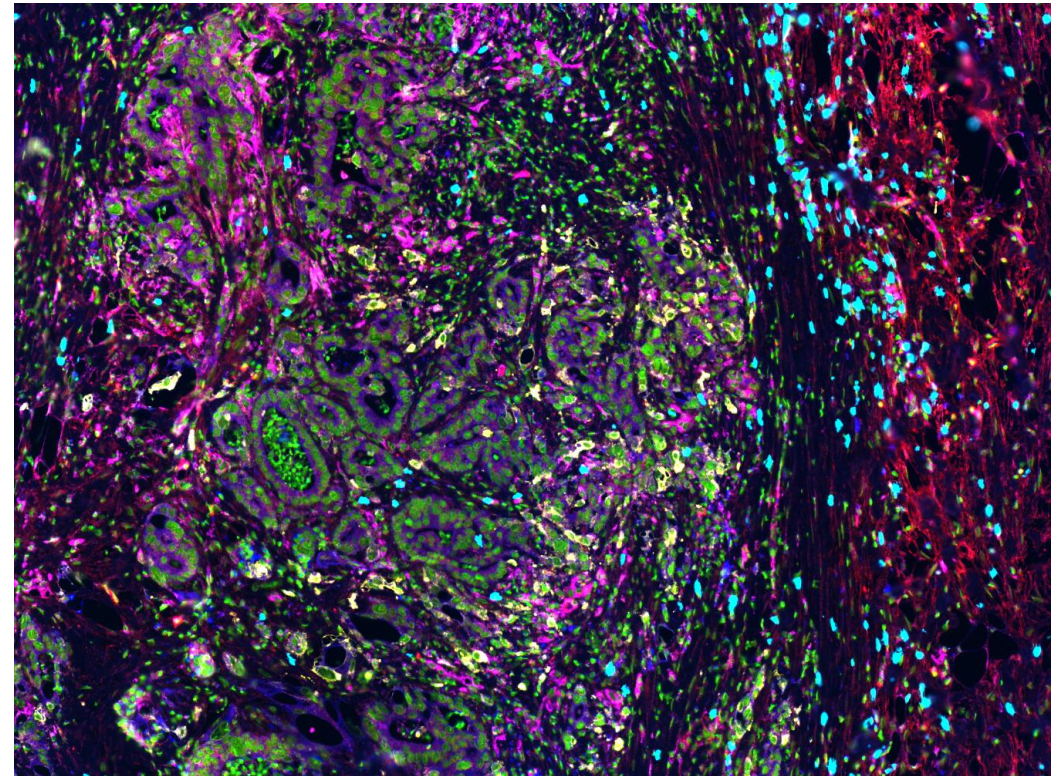


Technological revolutions are changing how we study living systems

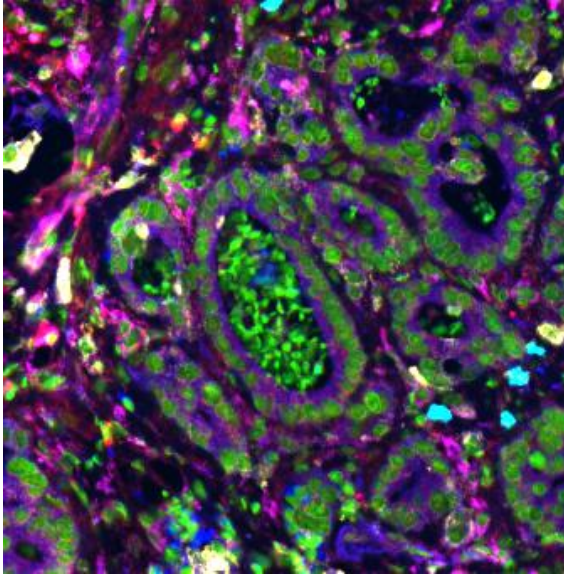
Classical pathology stain



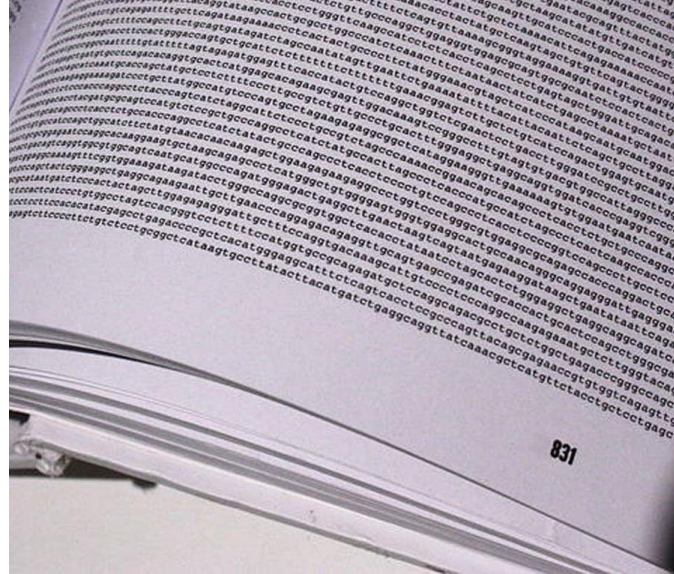
Spatial genomics



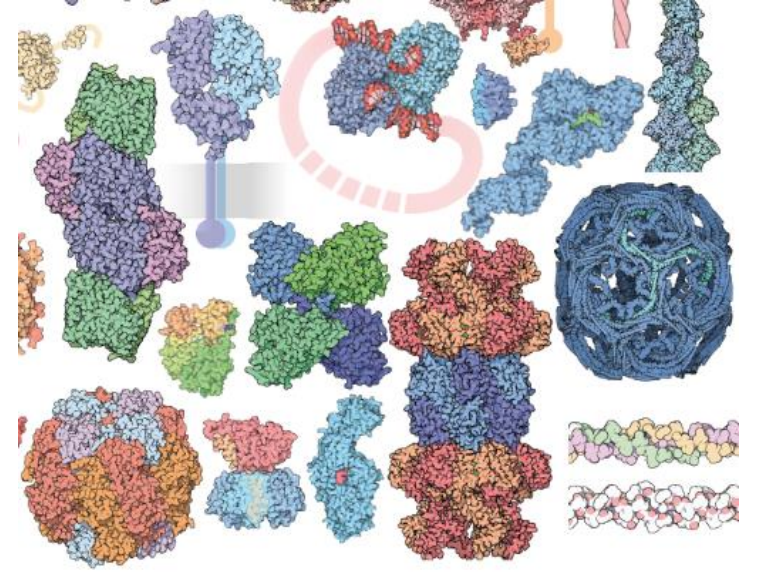
Deep learning is changing our relationship with biological data



Images

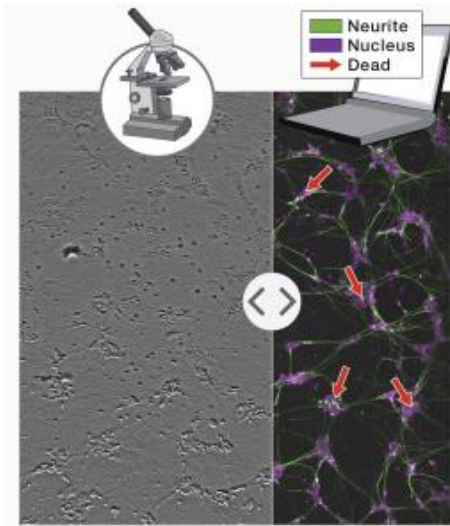


Sequences

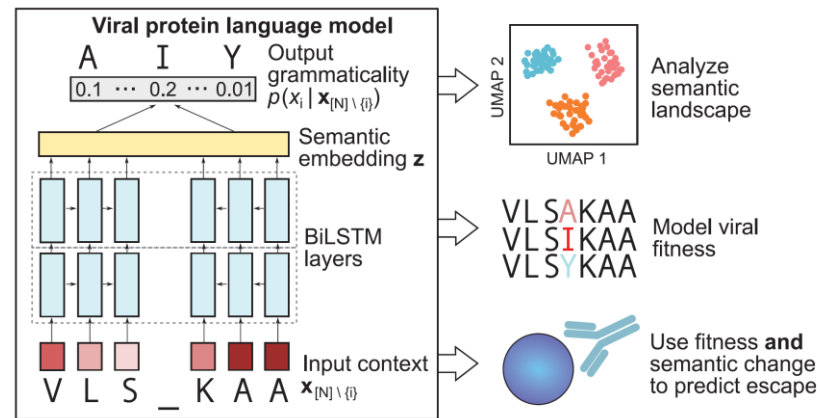


Structures

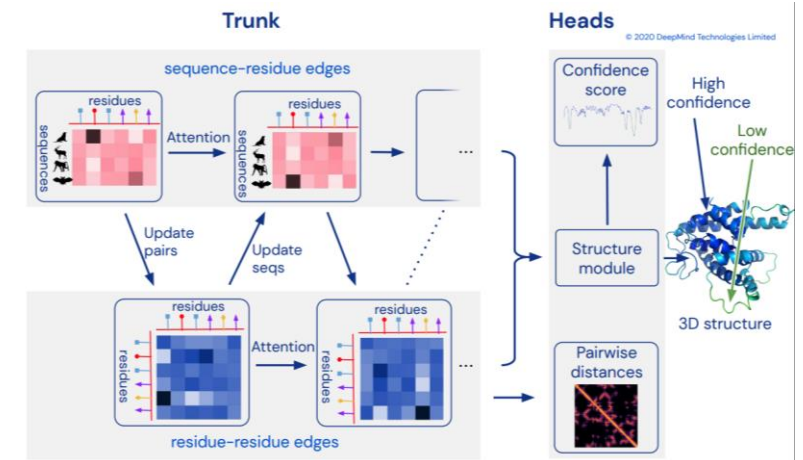
Deep learning is changing our relationship with biological data



Convolutional
Neural
Networks

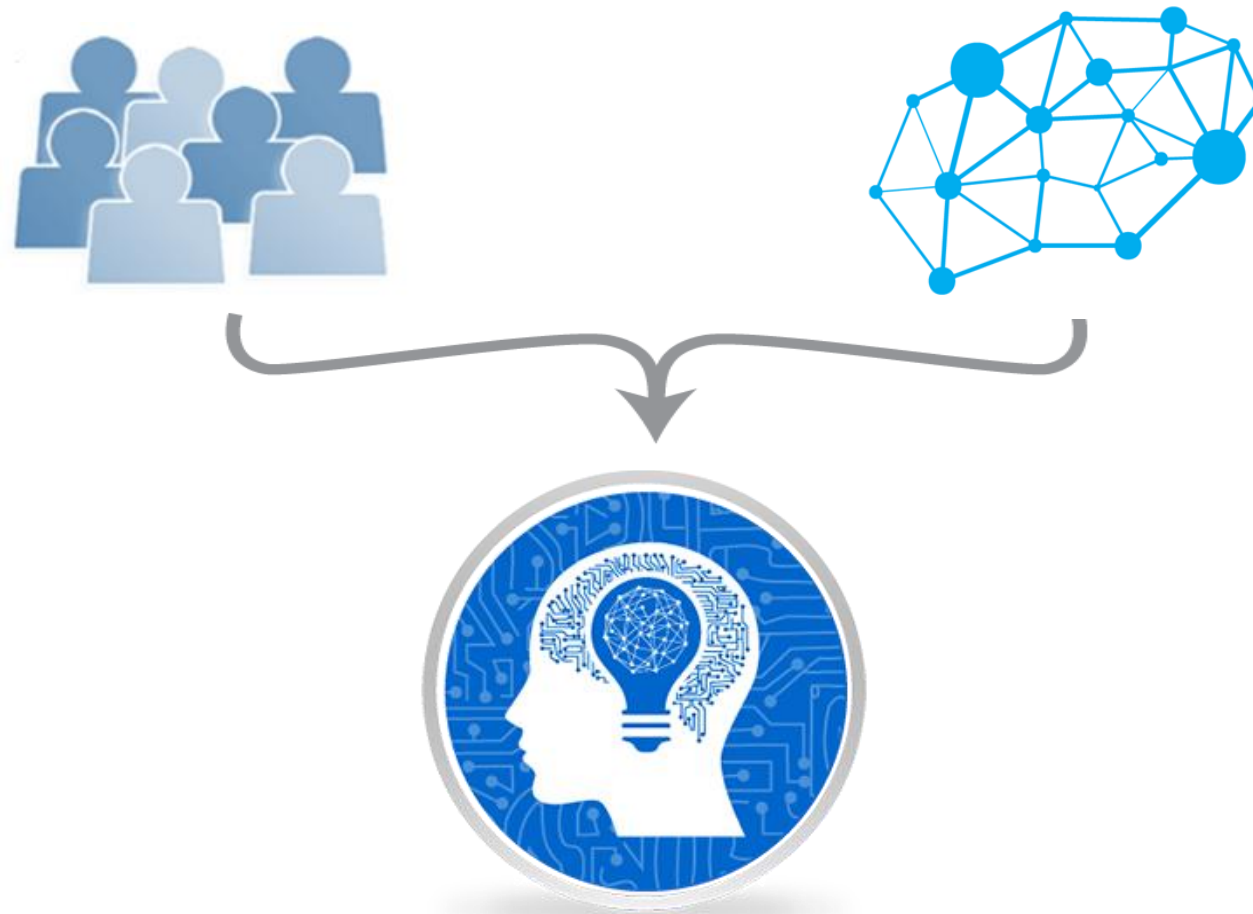


Language Models



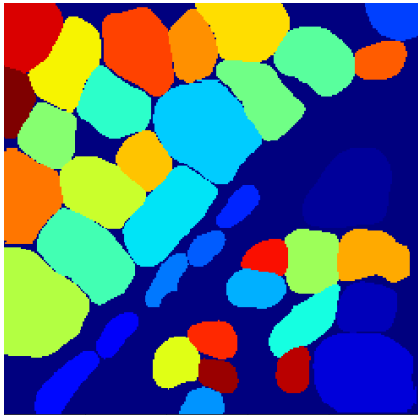
Equivariant Graph
Neural Networks

Deep learning is data-driven - there is no revolution without data and labels

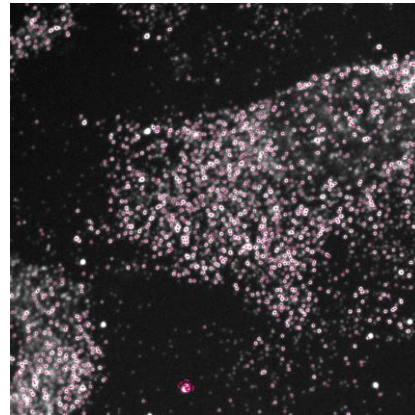


Deep learning is changing how we interpret biological images

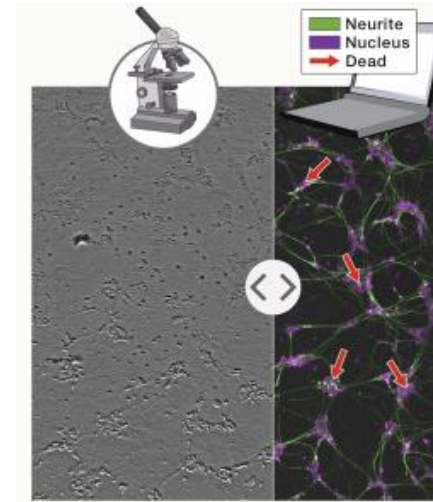
Deep learning



Segmentation



smFISH Analysis



Augmented Microscopy

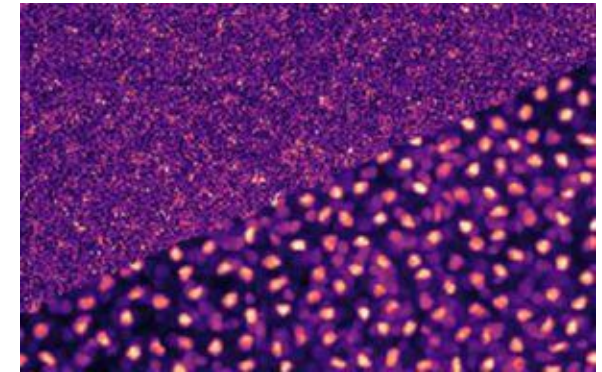
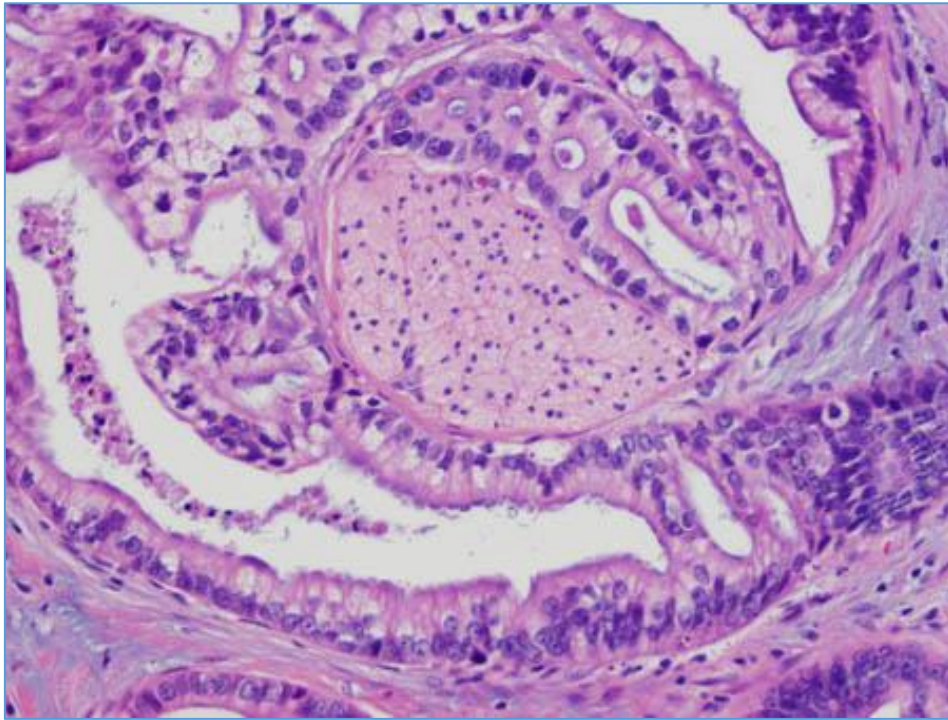


Image Restoration

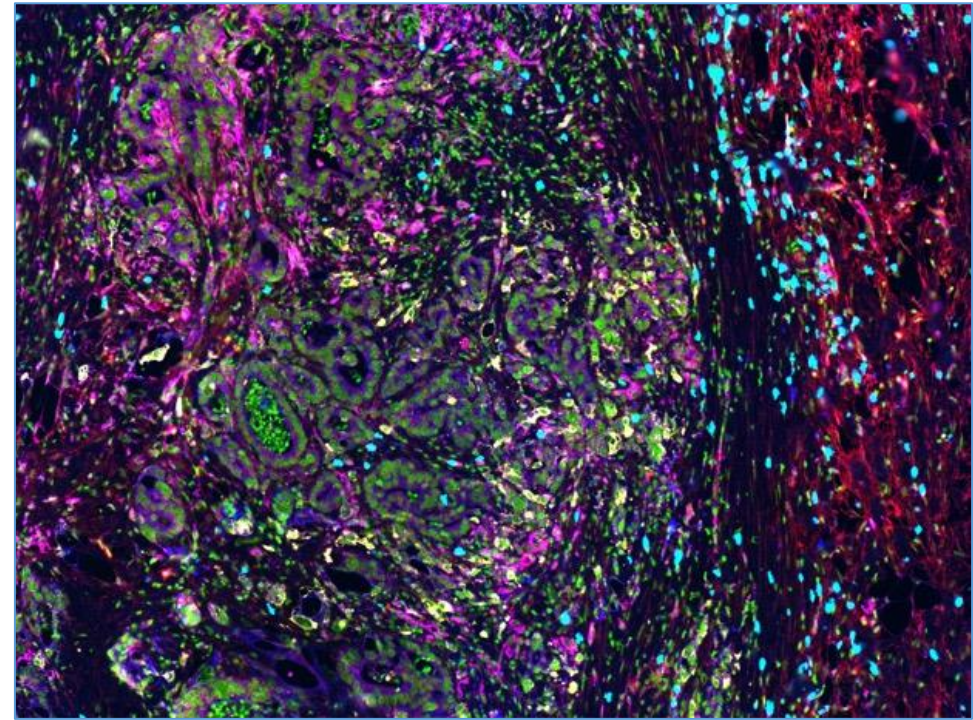
Angelo et al., 2014
Cai et al, 2020, 2021
Johnson et al. 2018
Greenwald, Miller et al. 2021
And many others!

How will spatial biology and AI transform our understanding of living systems?

Classic Imaging Data



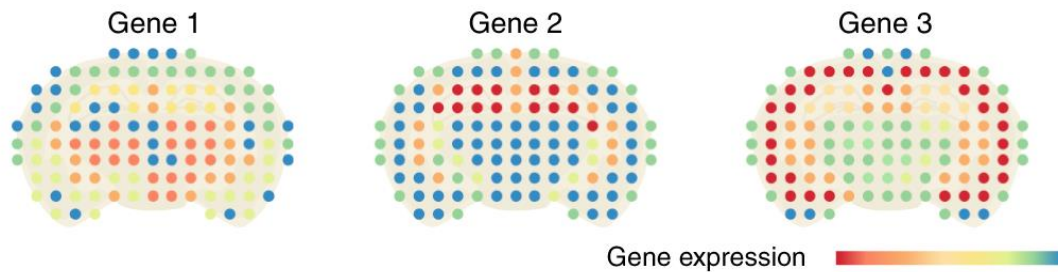
Modern Imaging Data



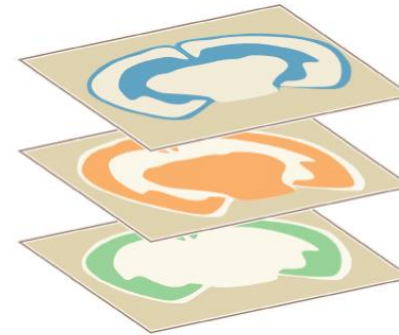
AI can render modern spatial data interpretable

How will spatial biology and AI transform our understanding of living systems?

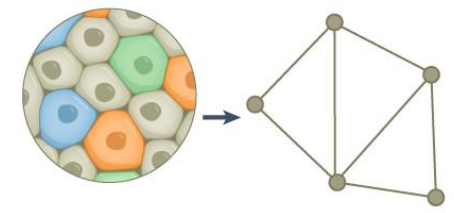
Spatial variation



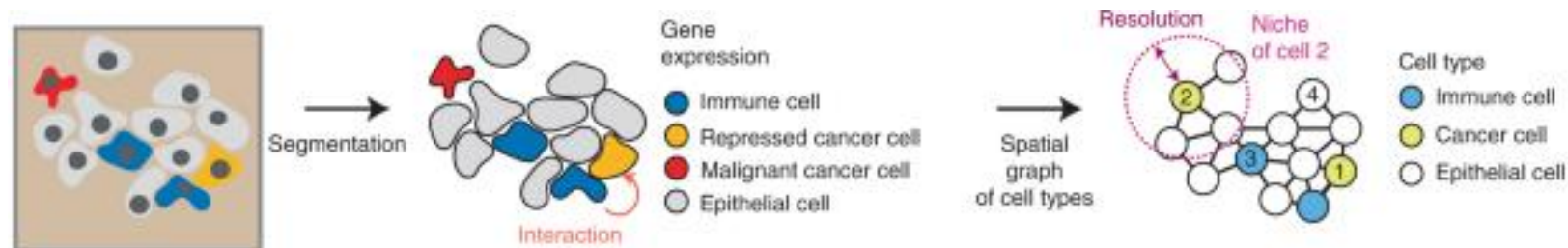
Parts list variation



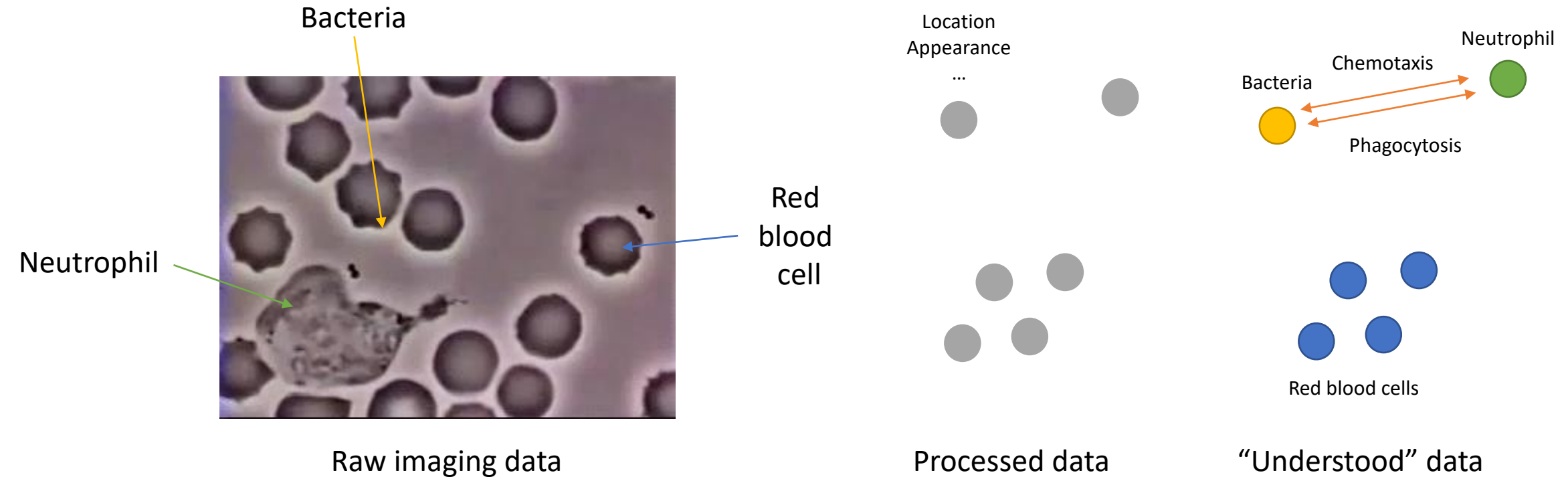
Organization



Target discovery and spatial biomarkers

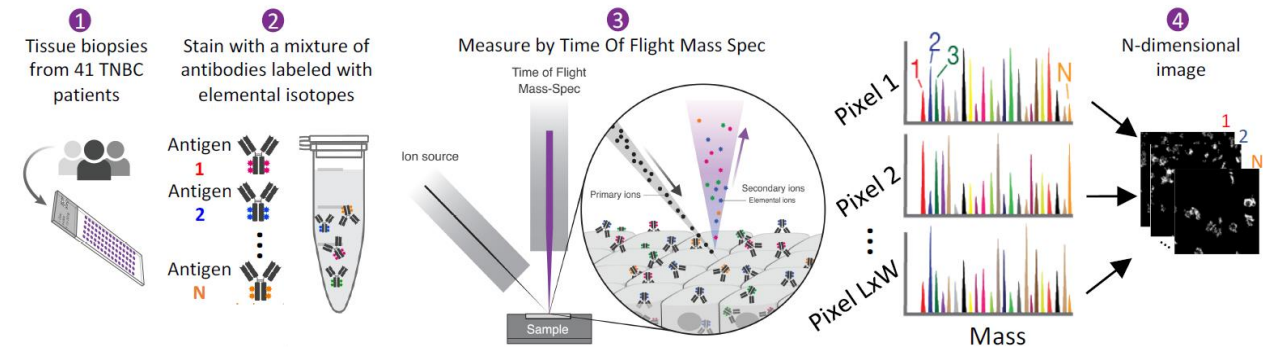
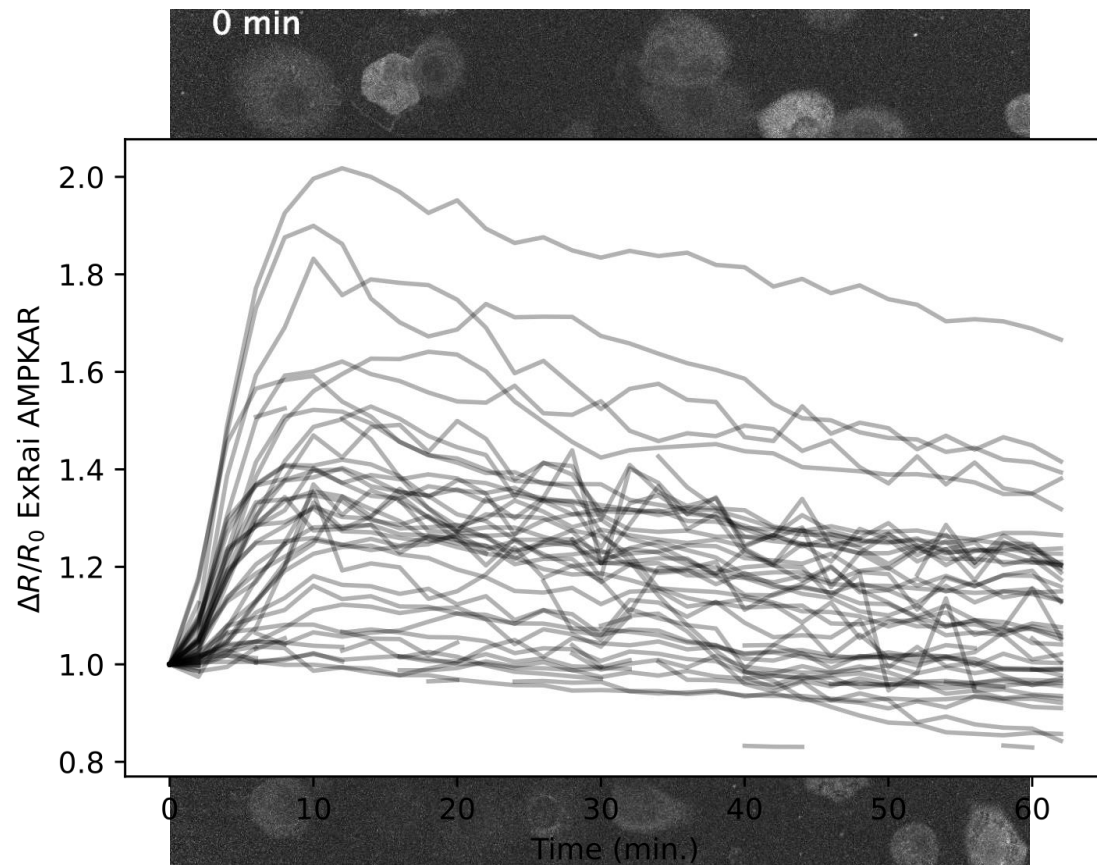


Computational primitives for cellular imaging

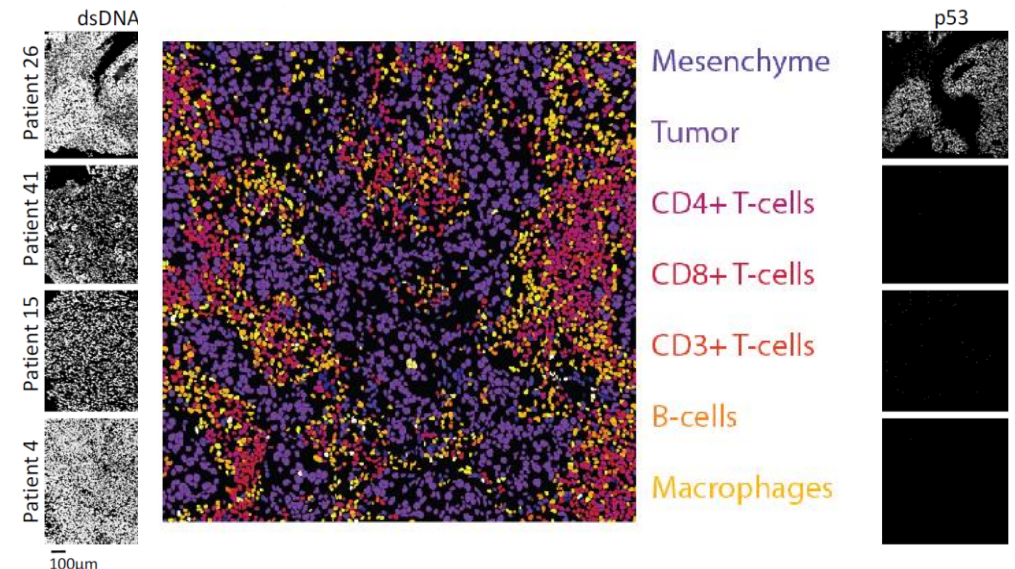


Computational primitives are necessary to convert raw imaging data into a more interpretable format like spatio-temporal graphs

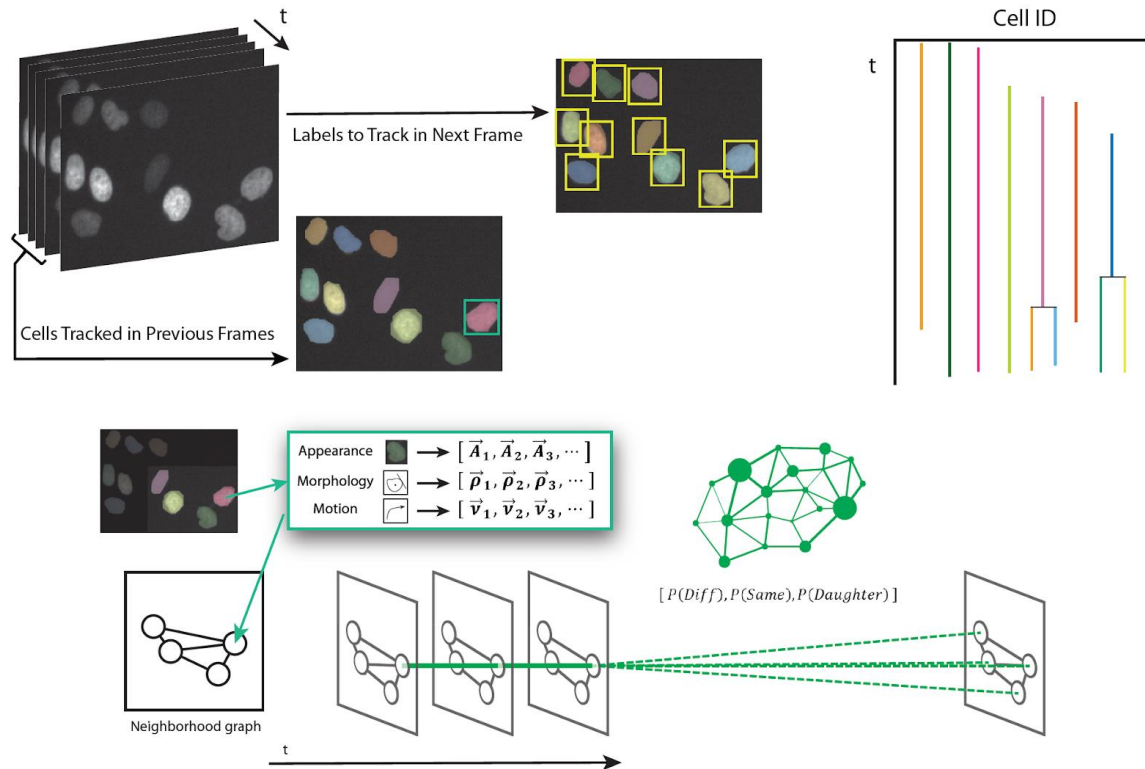
Single-cell analysis is a challenge for cellular imaging experiments



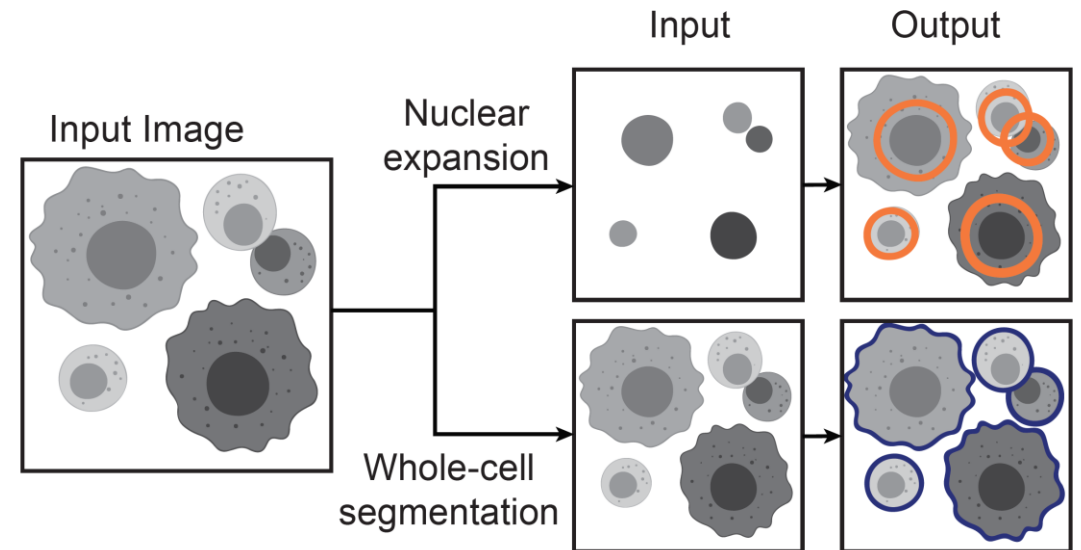
Multiplexed Ion Beam Imaging (MIBI)



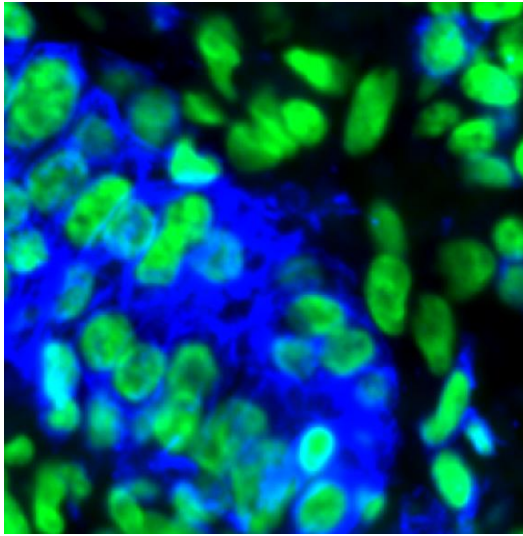
Single cell analysis is a common challenge for biological imaging experiments



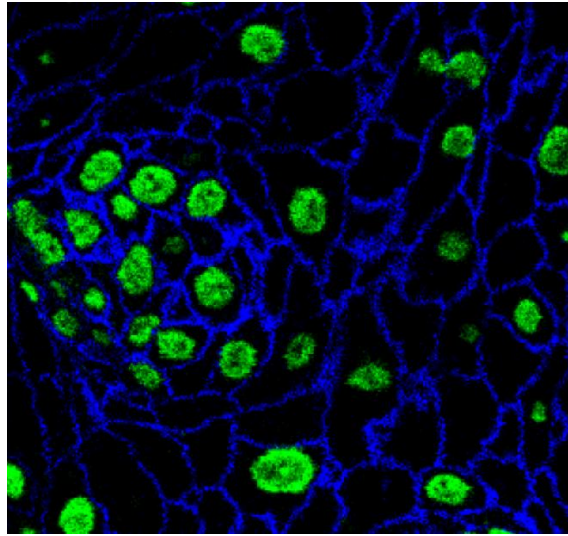
Cell segmentation and tracking
in live-cell imaging movies



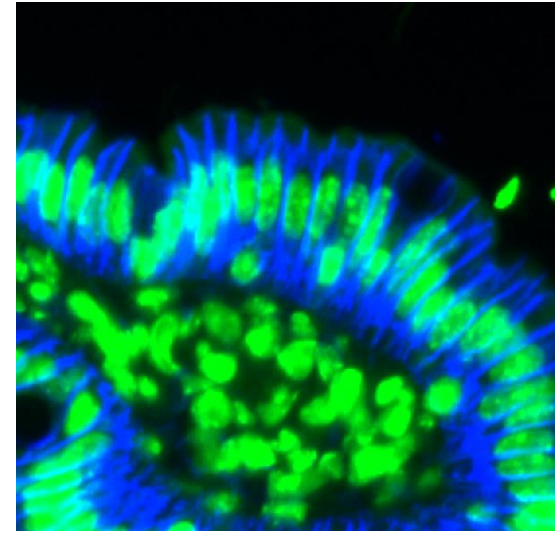
Whole cell and nuclear segmentation
in tissues



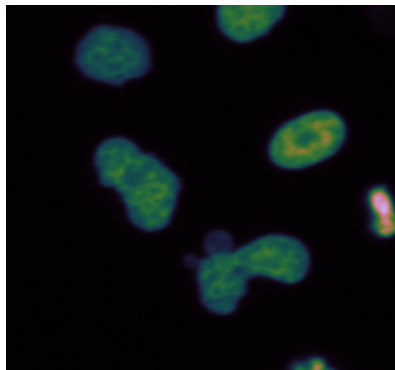
SCLC (CyCIF)



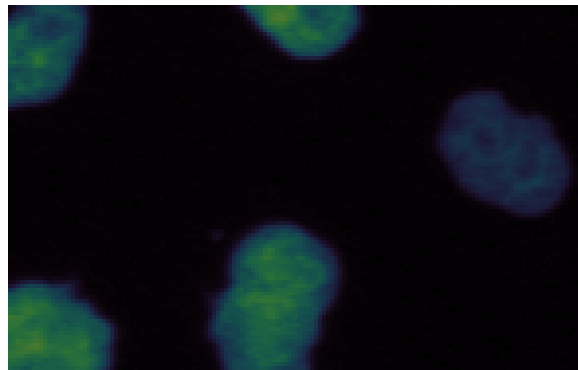
GI (MIBI)



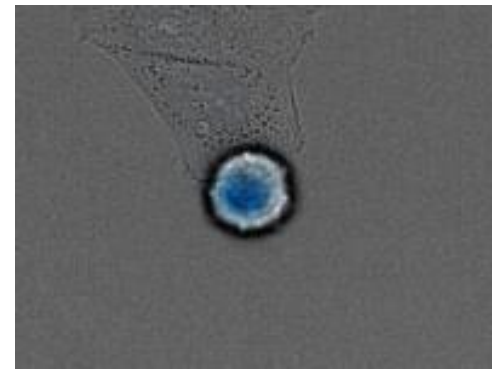
Colon (MxIF)



HeLa-S3 H2B-mCherry

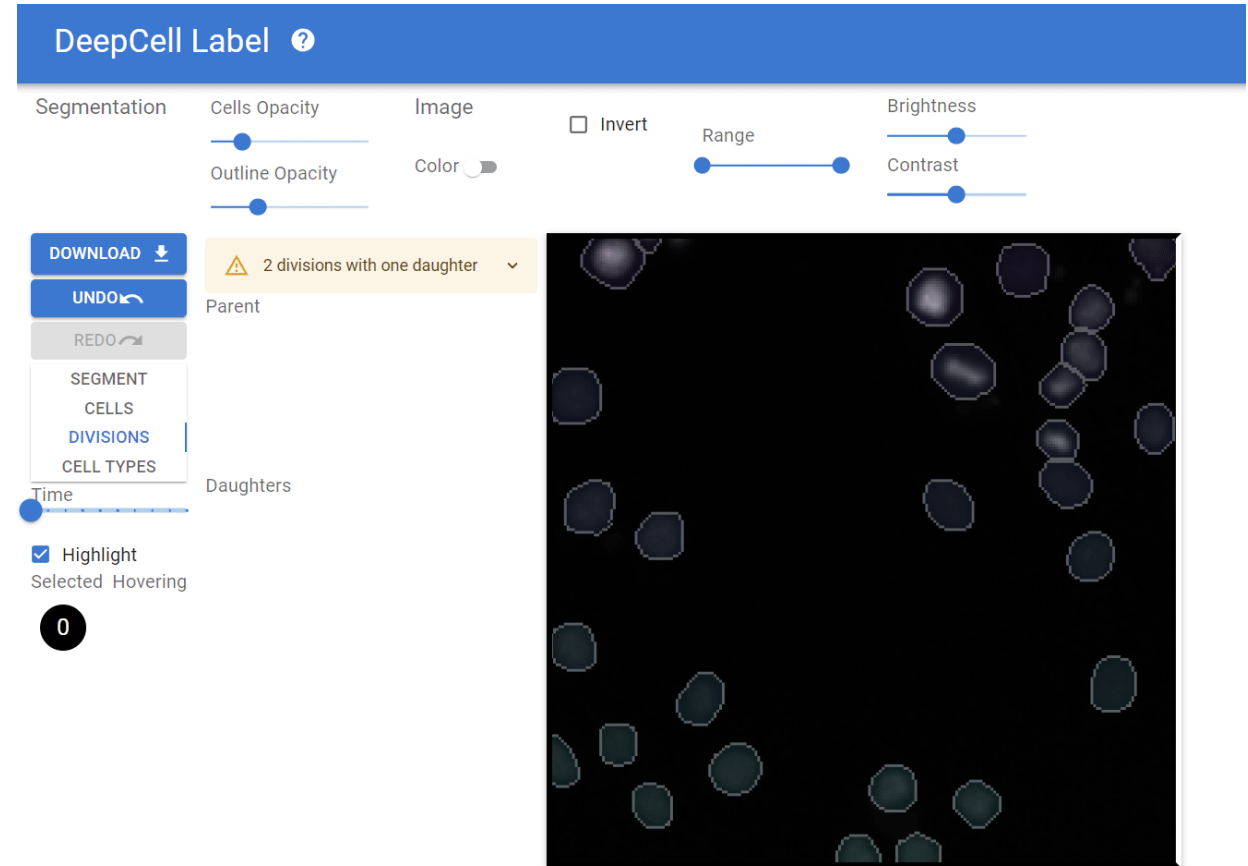
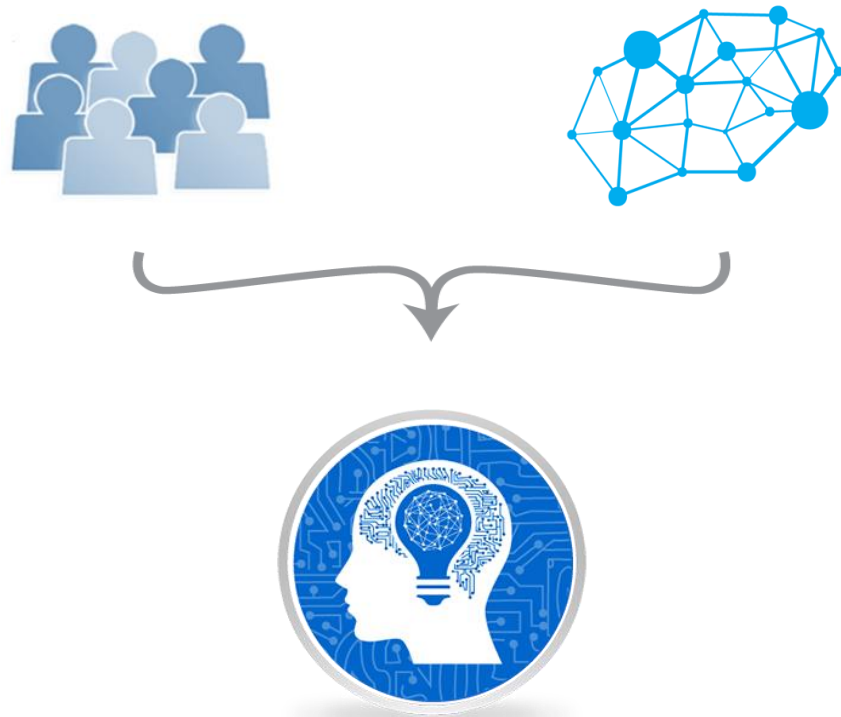


HeLa-S3 H2B-mCherry



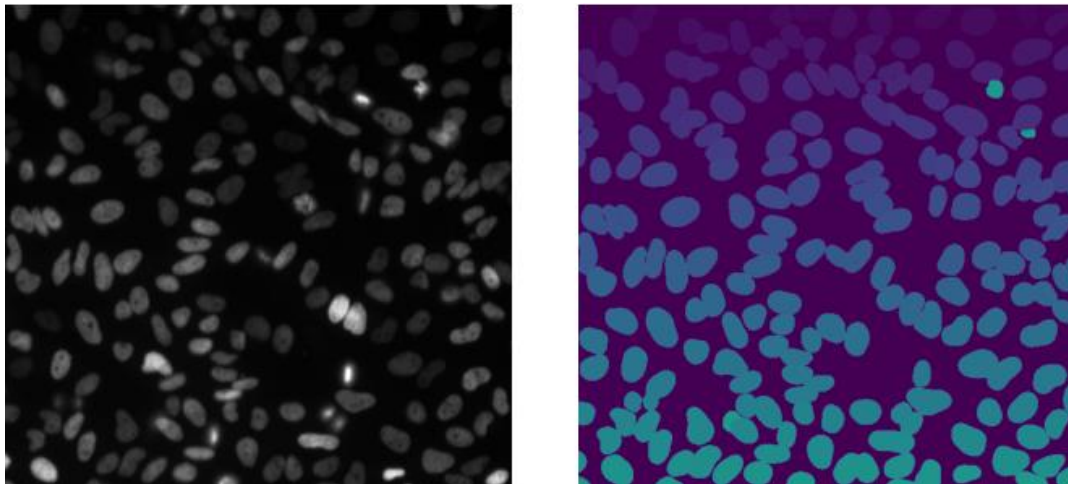
HeLa-S3 Phase

Image labeling at scale through human-AI collaboration

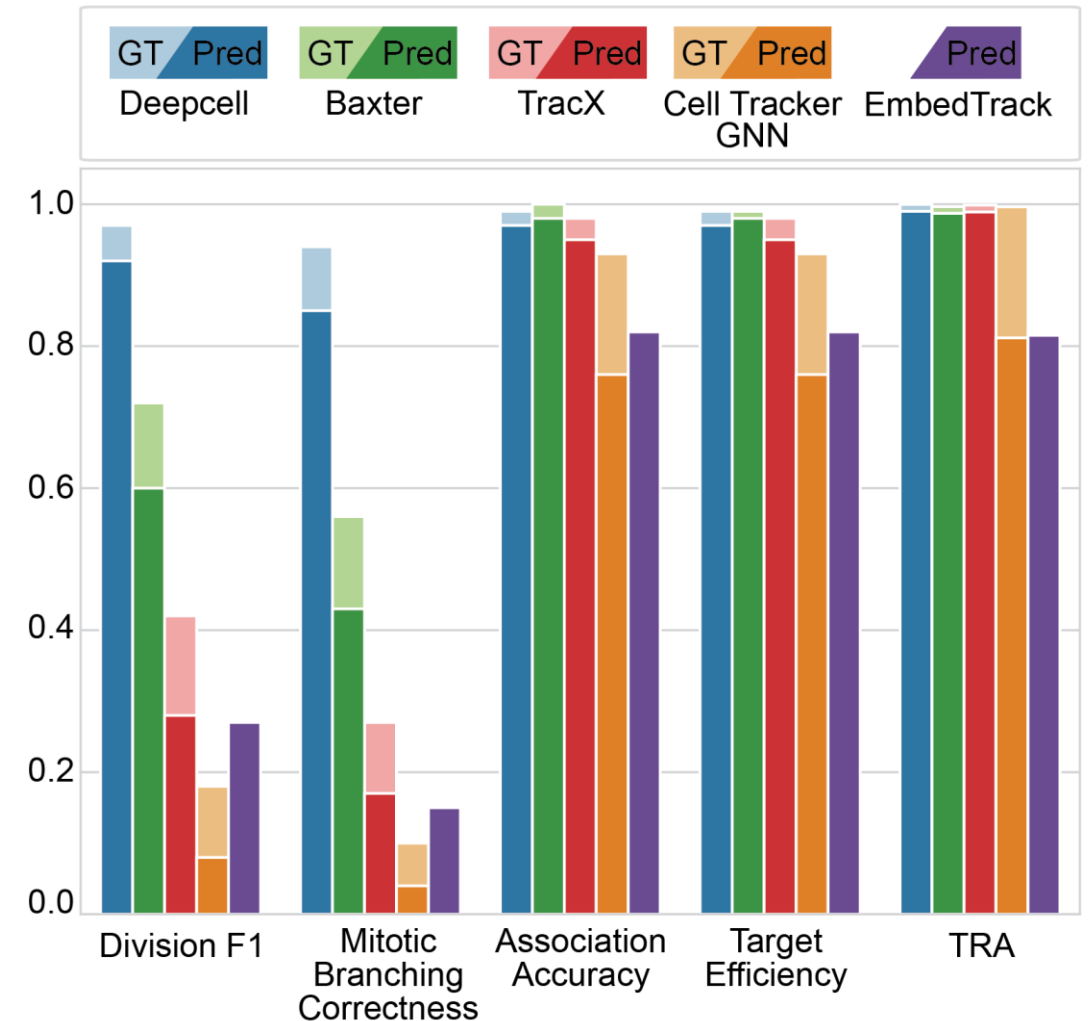


© 2016-2023 The Van Valen Lab at the California Institute of Technology (Caltech). All rights reserved.
For any questions or issues, please post on our [GitHub Issues](https://github.com/vanvalenlab/deepcell-label/issues) page.

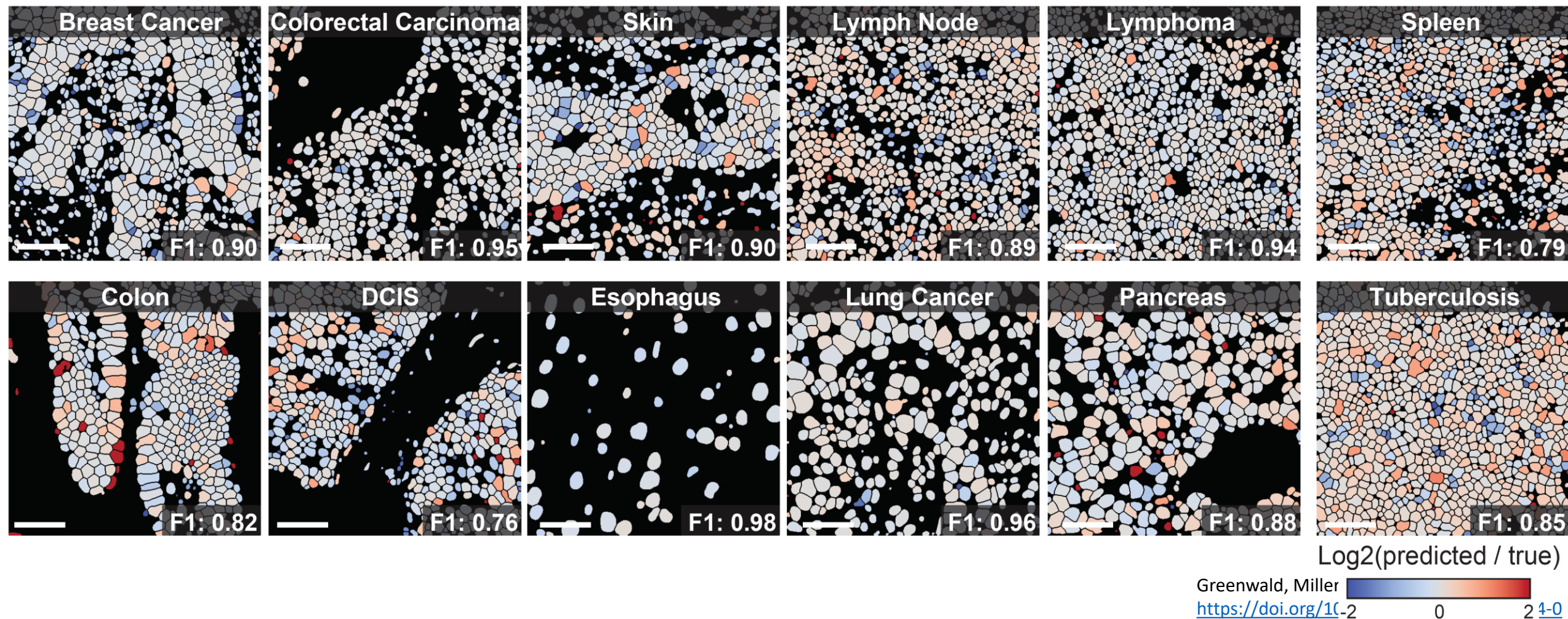
Caliban: Single-cell analysis for live-cell imaging



Cell Type	Objects	Tracks	Divisions
RAW264.7	335,049	8,642	1,322
Hek293	127,192	2,727	456
HeLa-S3	77,180	2,234	189
NIH-3T3	62,983	1,413	250
PC3	5,051	159	5
Total	606,455	15,175	2,222

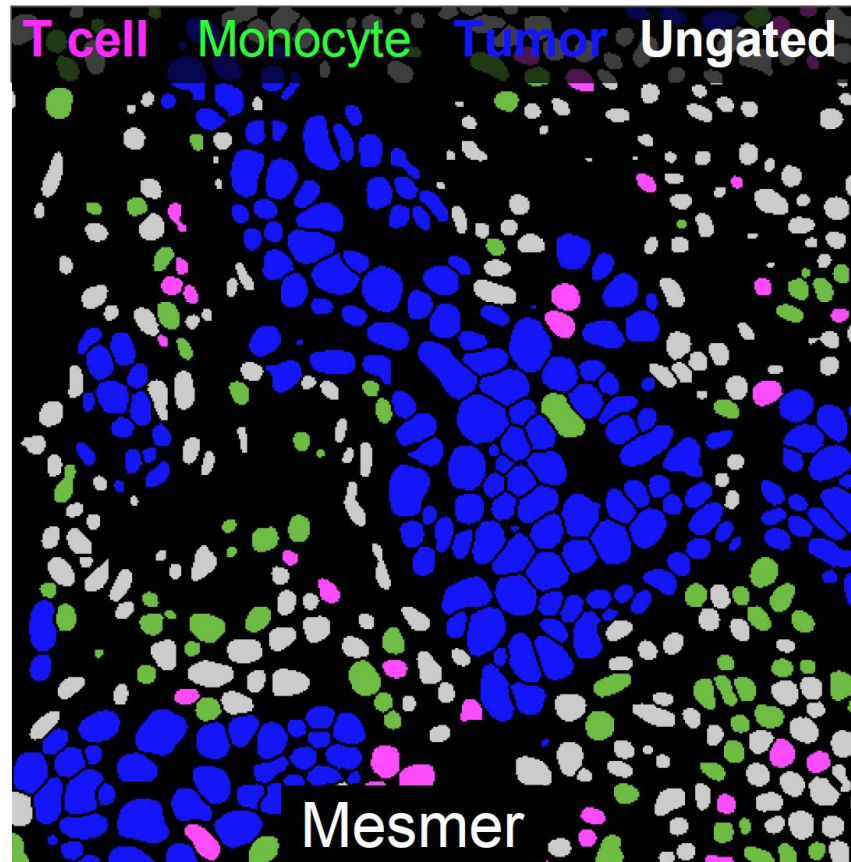


Mesmer: Cell segmentation for tissue imaging

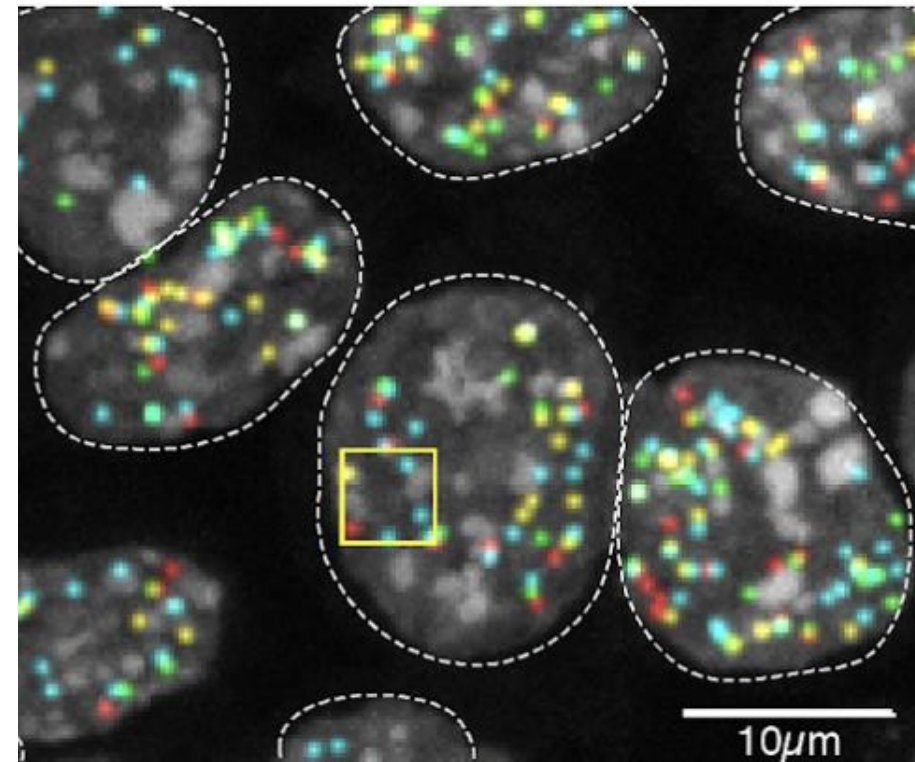


What happens after cell segmentation?

Cell phenotyping for
spatial proteomics

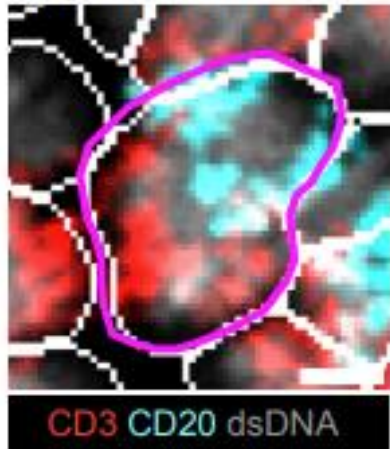


Spot phenotyping for
spatial transcriptomics

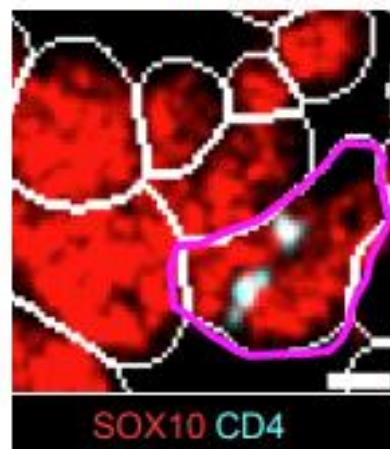


Why is cell phenotyping hard?

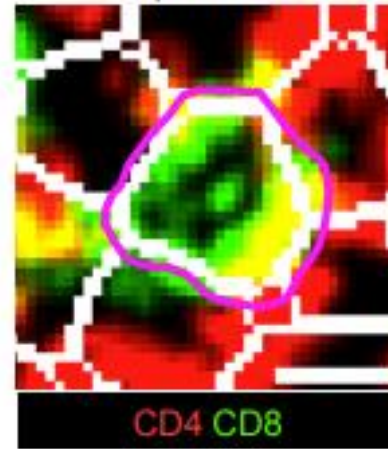
Segmentation errors



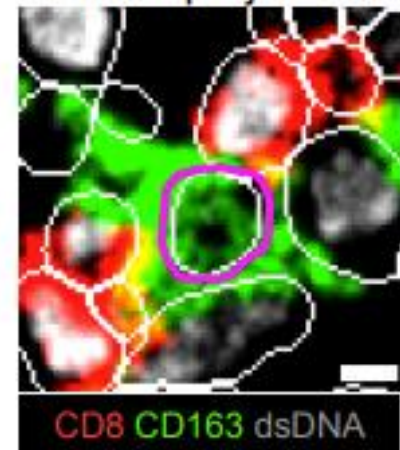
Noise



Spillover




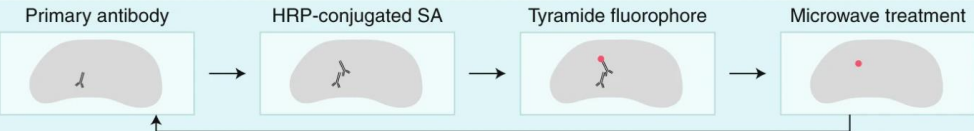
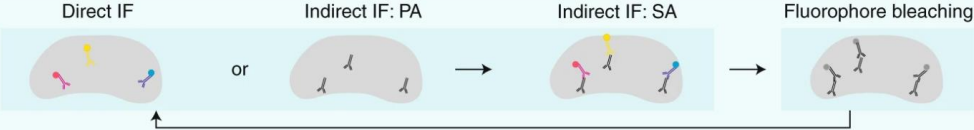

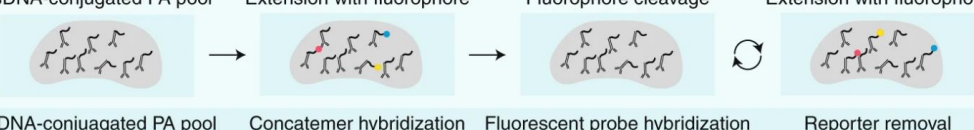
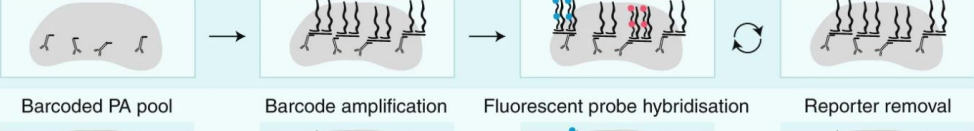




Cellular projections



Generalization

Cell phenotyping: Generalization across imaging platforms

Spatial proteomics

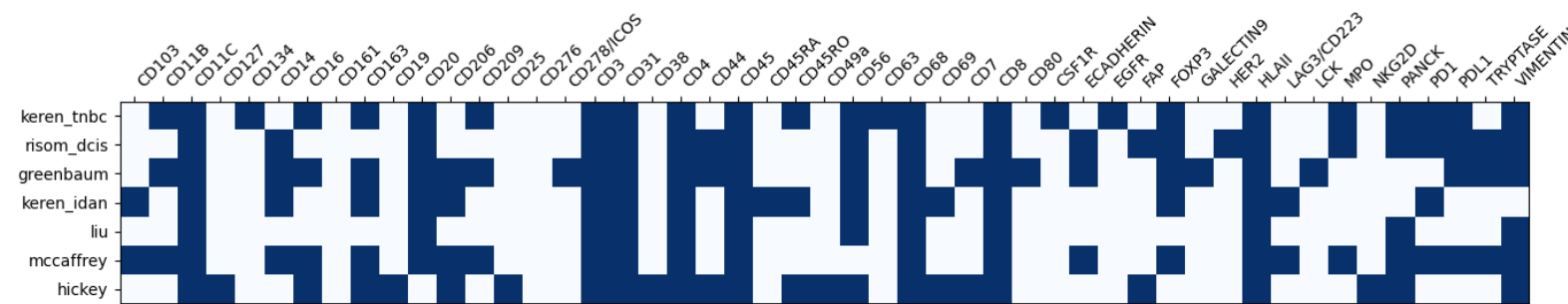
				No. of targets	Tissue prep.
Iterative	mIHC		Primary antibody Secondary antibody	30	FFPE
	OPAL		Primary antibody HRP-conjugated SA Tyramide fluorophore Microwave treatment	10	FFPE
	CyclIF		Direct IF Indirect IF: PA Indirect IF: SA Fluorophore bleaching	60	FFPE
Iterative (fluorescence)	REAdye_lease and REAfinity		Primary antibody Fluorophore release Fluorophore bleaching	100 (400)	FFPE
	CODEX		dsDNA-conjugated PA pool Extension with fluorophore Fluorophore cleavage Extension with fluorophore	60	FF* FFPE
	Immuno-SABER		ssDNA-conjugated PA pool Concatemer hybridization Fluorescent probe hybridization Reporter removal	10 (50)	Whole-mount FF* FFPE
	InSituPlex		Barcoded PA pool Barcode amplification Fluorescent probe hybridisation Reporter removal	10	FFPE
TOF-mass pectrometry	IMC		Metal-conjugated PA pool UV laser ablation TOF mass spectrometry	40 (100)	FF FFPE
	MIBI		Metal-conjugated PA pool Ion beam gun TOF mass spectrometry	40 (100)	FF FFPE
Sequencing	DSP		Stain + oligonucleotide-conjugated PA pool Oligonucleotide cleavage Quantitative analysis	44 (100)	FF* FFPE

Cell phenotyping: Generalization across datasets

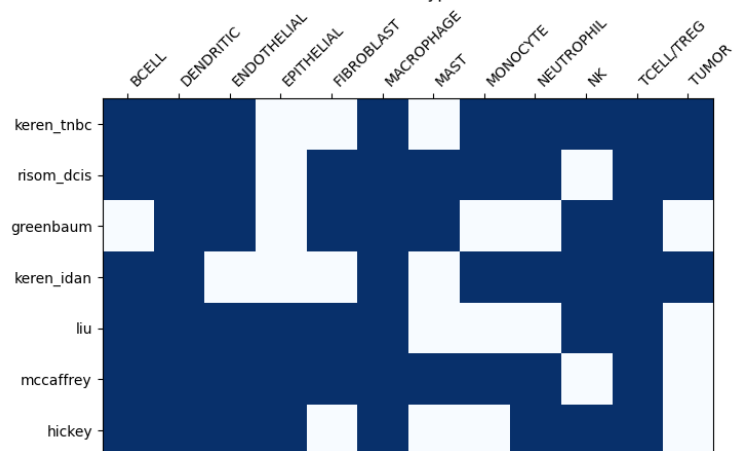
Currently ingested all publicly available spatial proteomics data from the literature

- 17 datasets
- Curated panel of 51 markers
- 4 million labeled cells from existing data
 - Significant label noise in all labeled datasets
- Estimated 20 million cells overall

Dataset - Channel Matrix



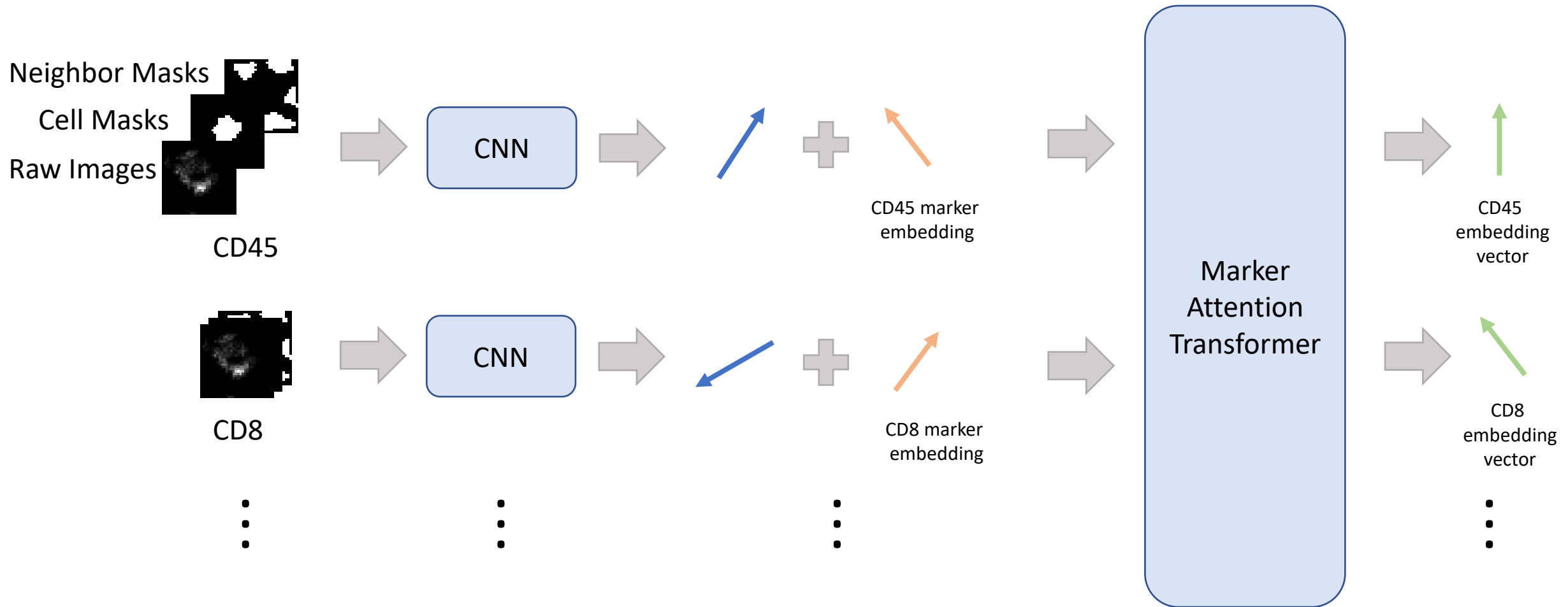
Dataset - Cell Type Matrix



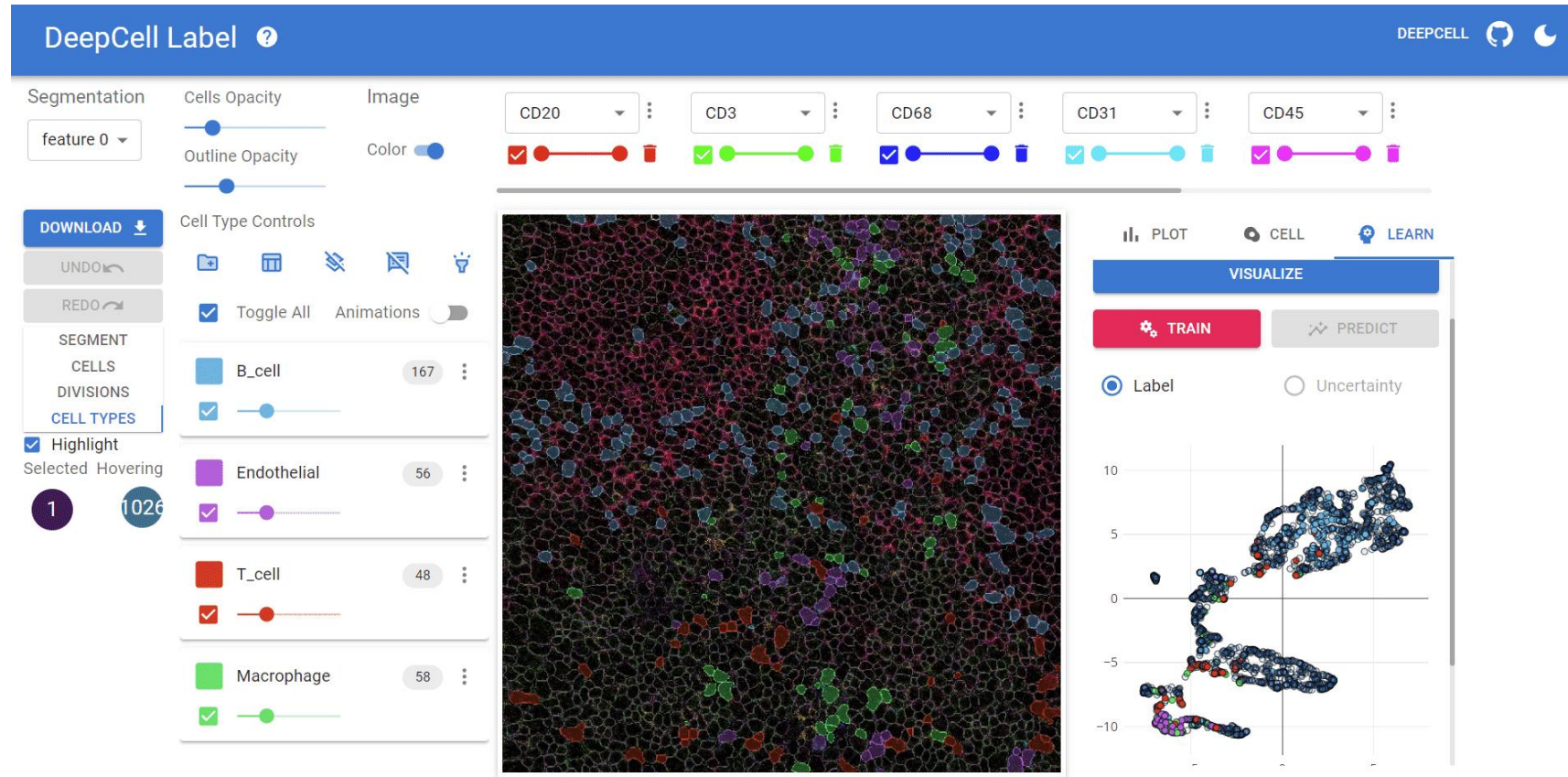
For generalization, we need:

- Normalization
- Access to the full information content of images
- Deep learning models that are aware of the marker panel
- Human-in-the-loop labeling

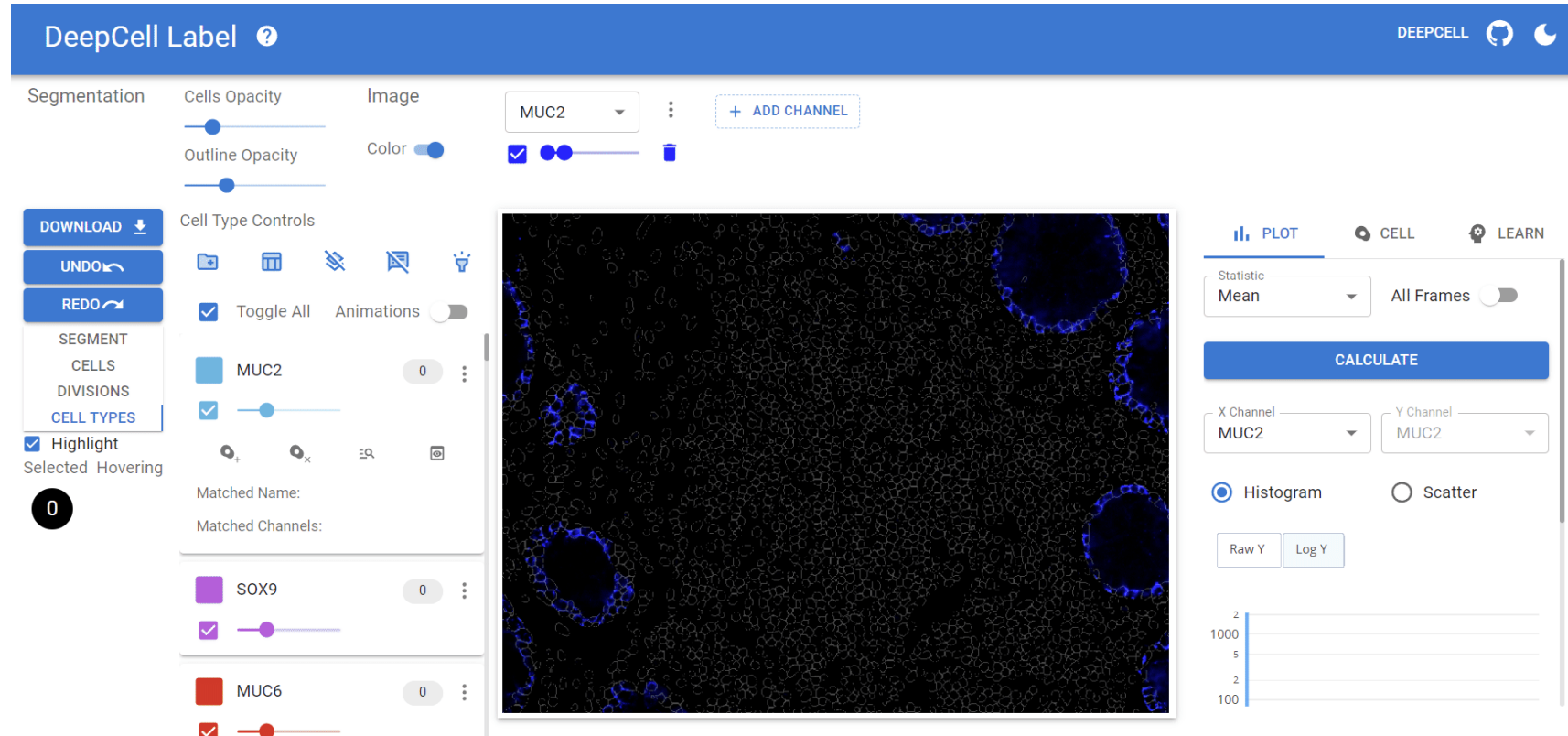
Marker-wise attention allows generalization across datasets



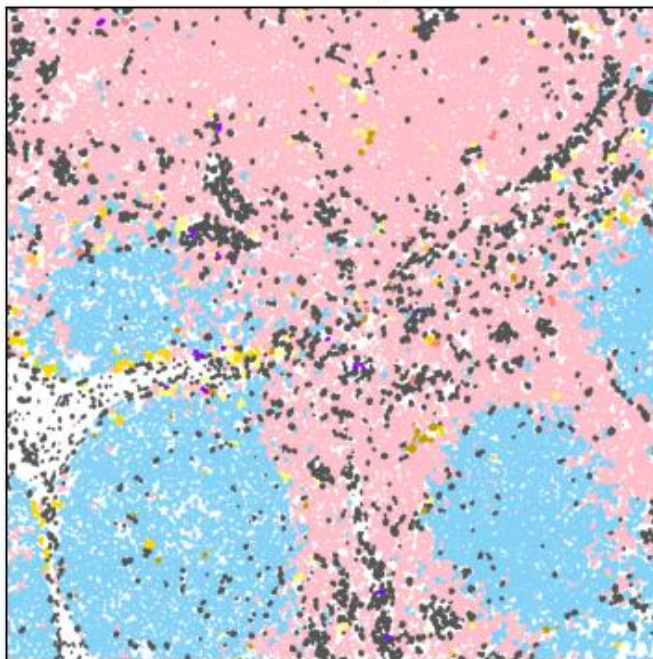
Human-in-the-loop labeling of cell type and marker positivity with DeepCell Label



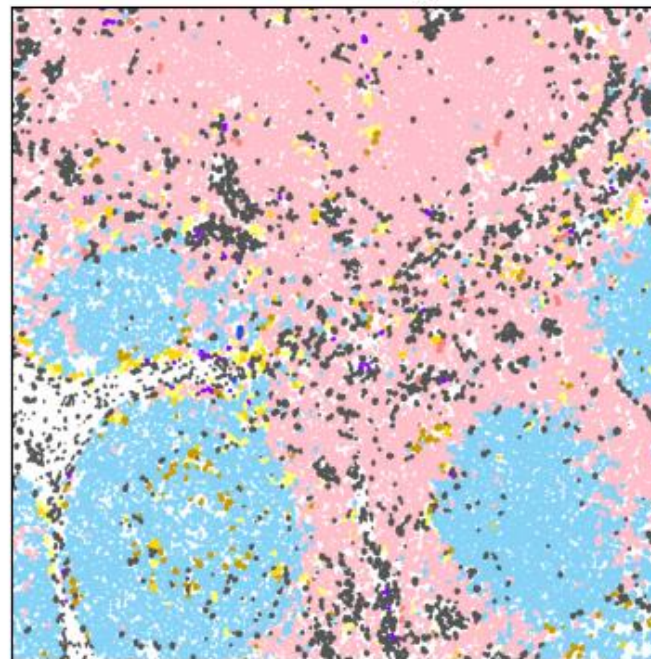
Human-in-the-loop labeling of cell type and marker positivity with DeepCell Label



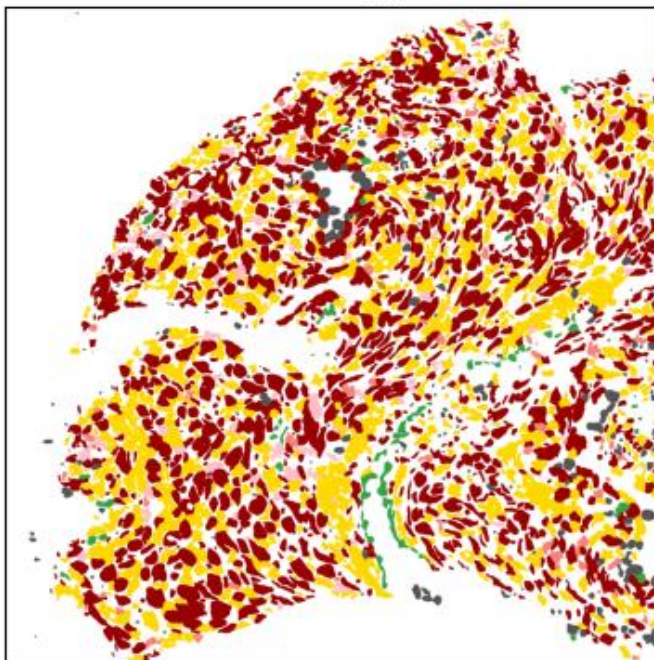
True cell types



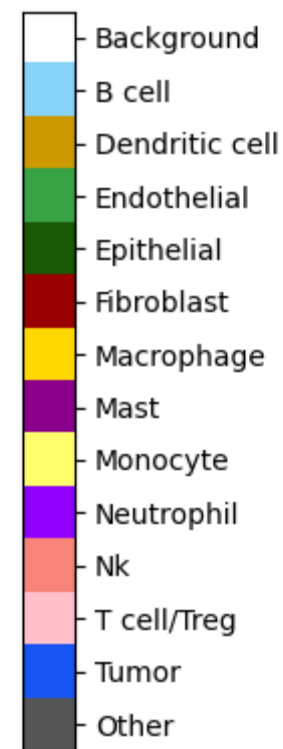
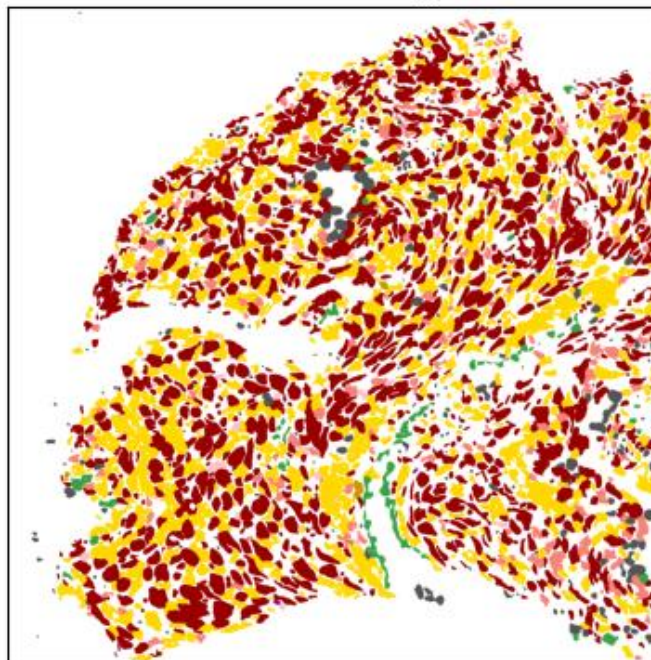
Predicted cell types



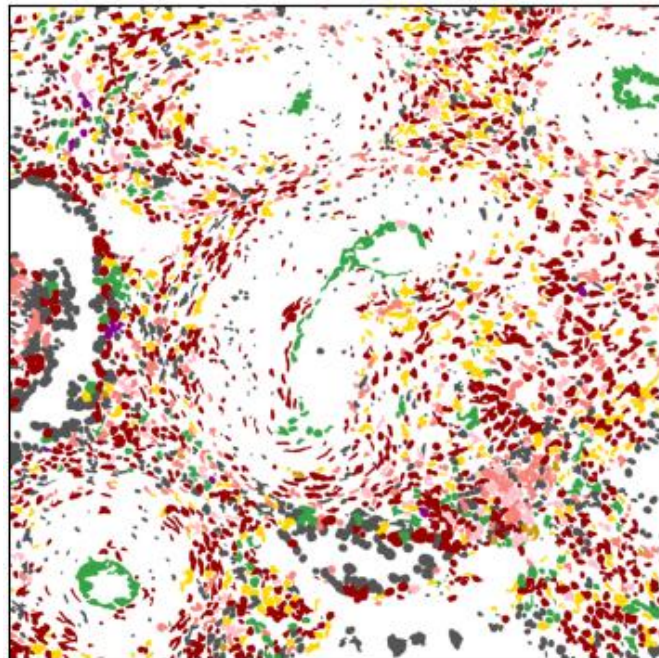
True cell types



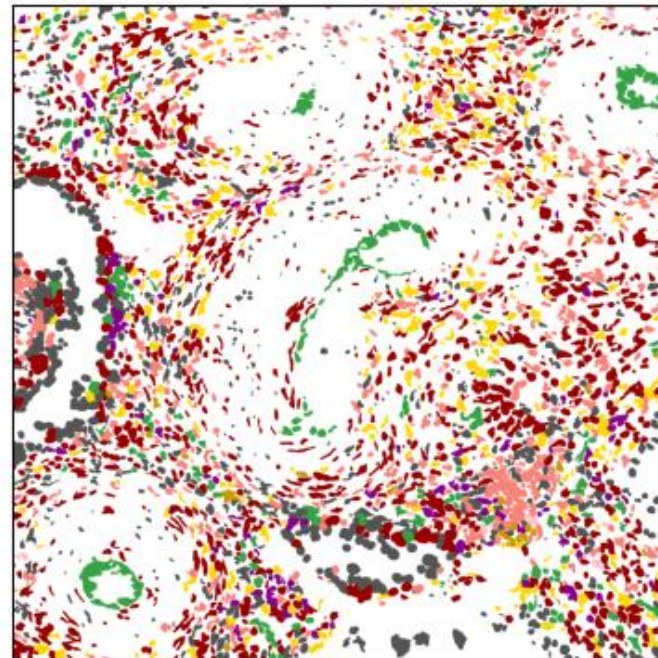
Predicted cell types



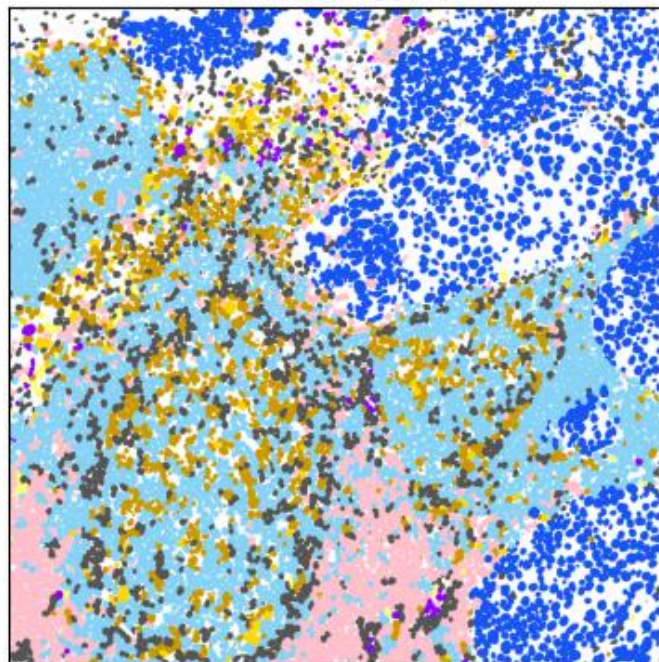
True cell types



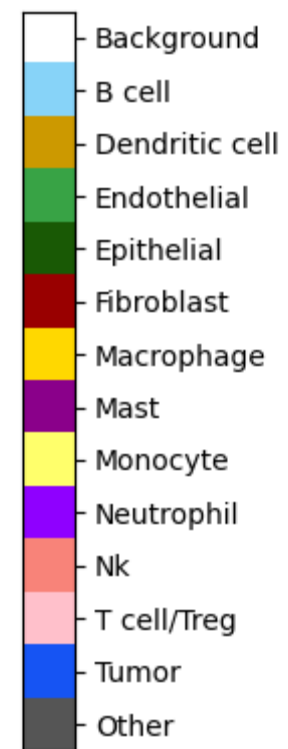
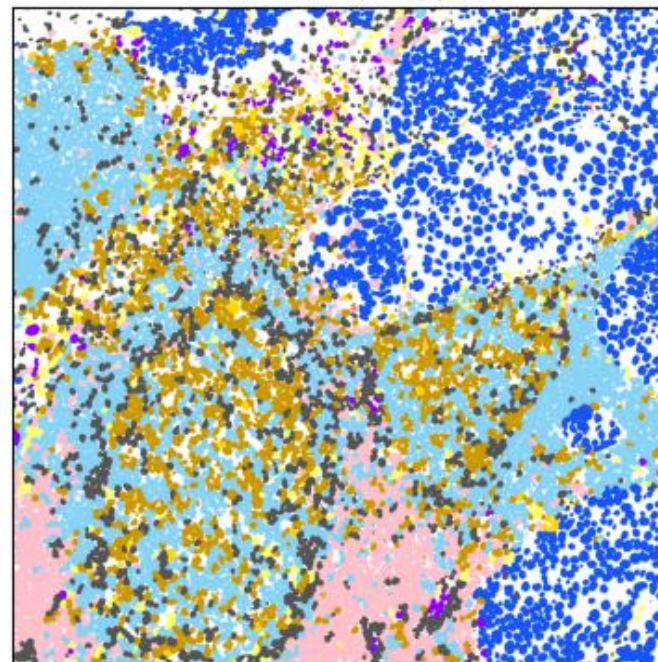
Predicted cell types



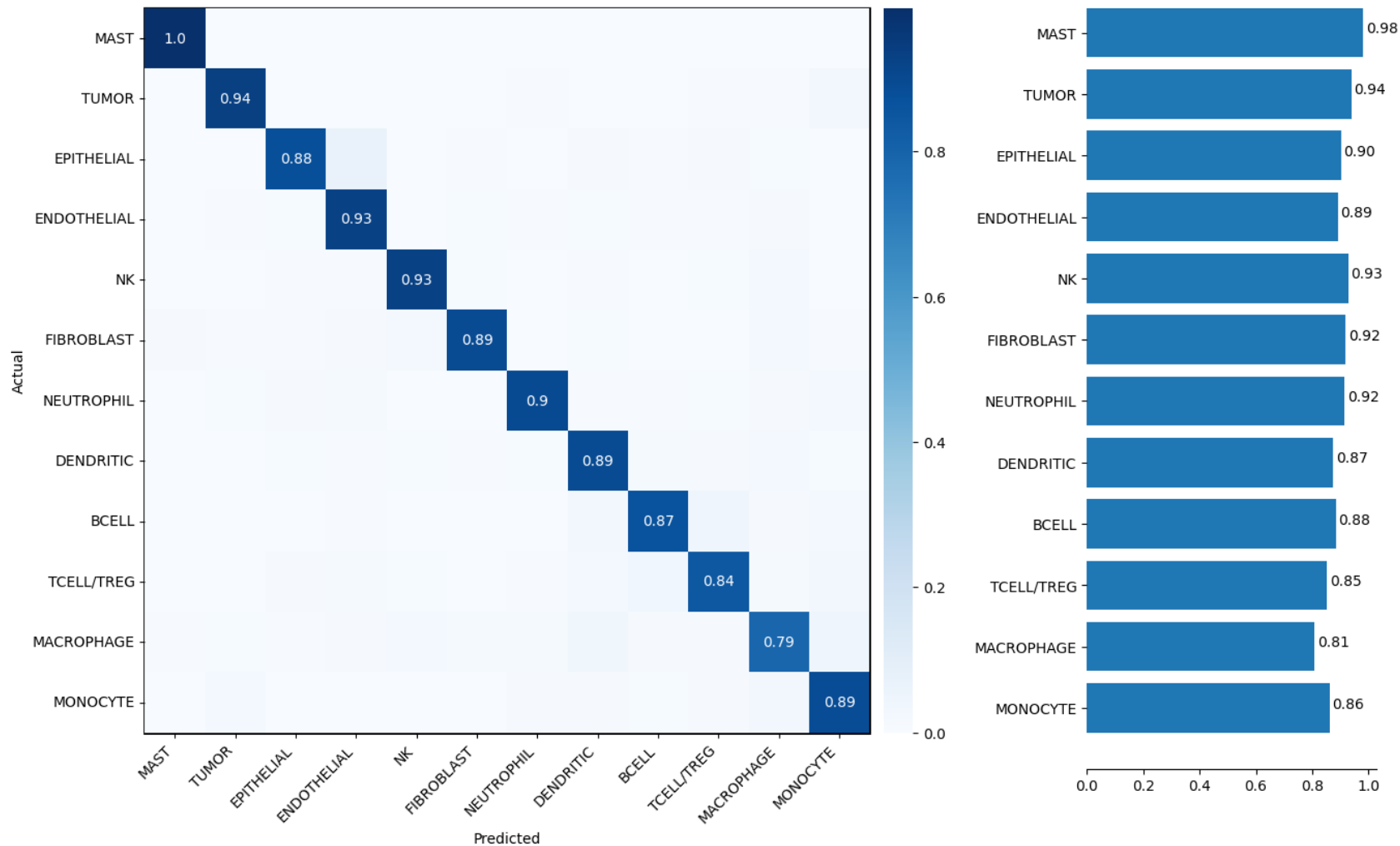
True cell types



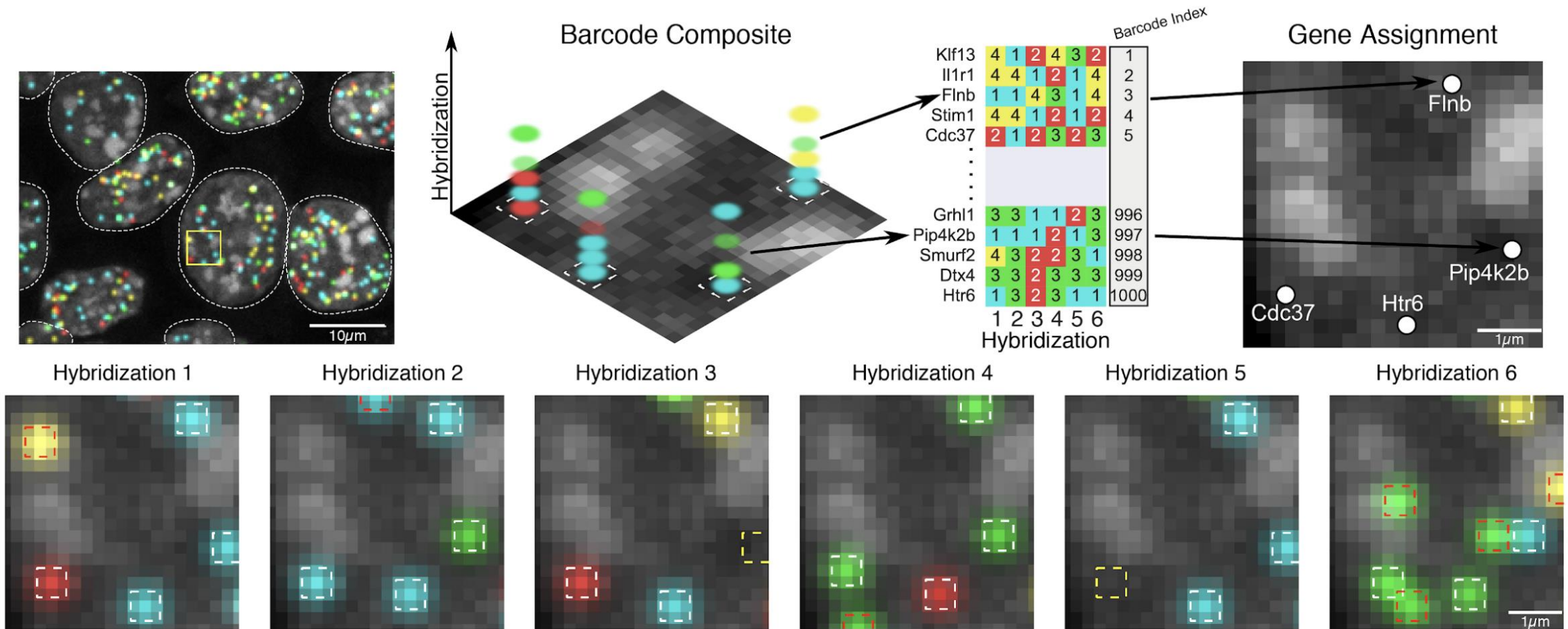
Predicted cell types



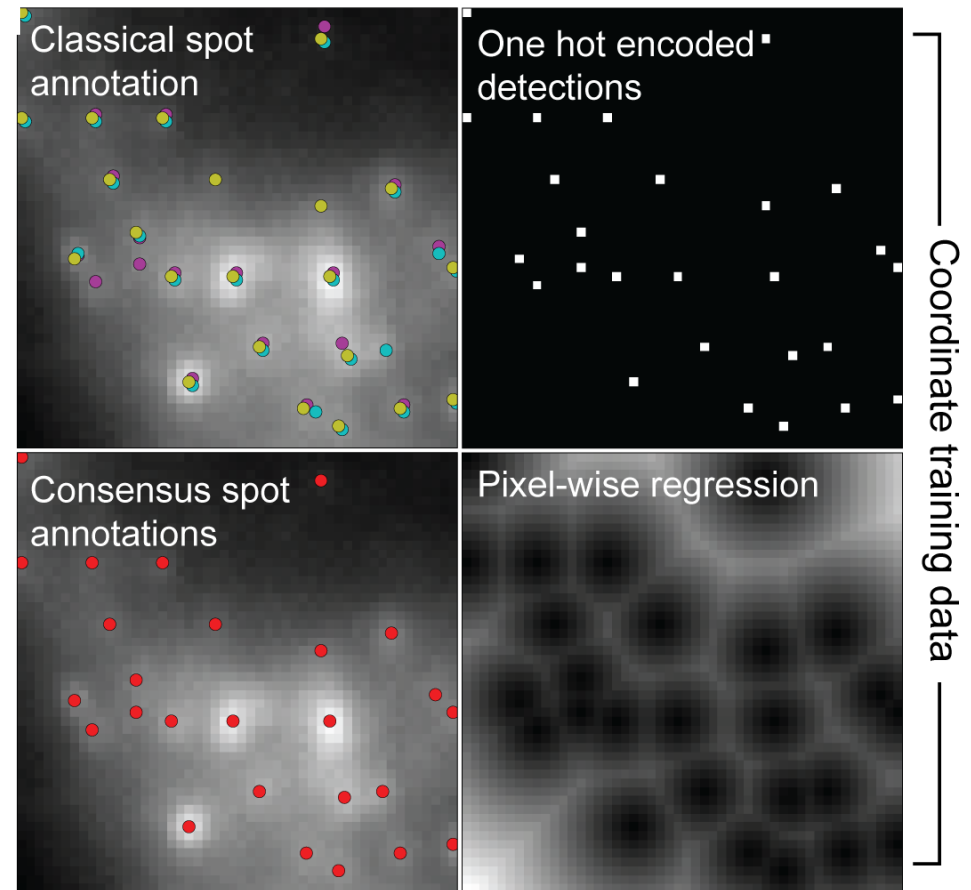
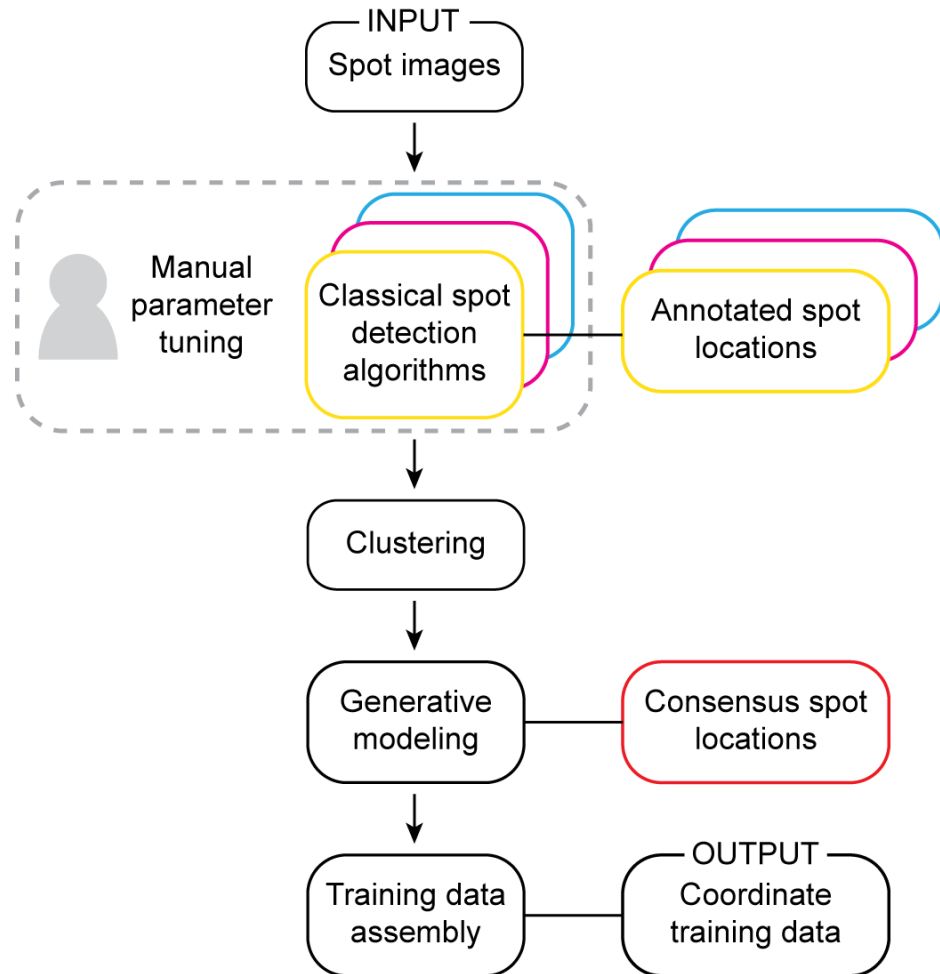
DeepCellTypes enables accurate cell phenotyping for spatial proteomics



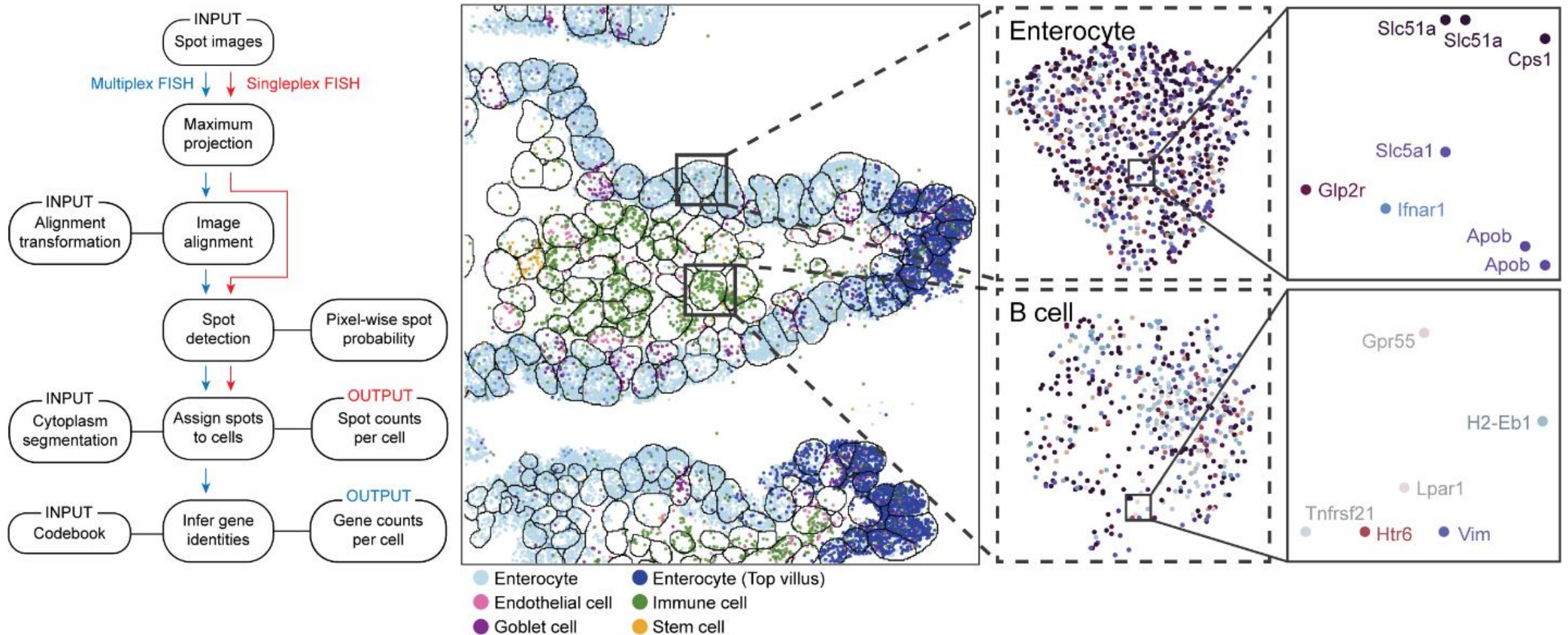
Multiplexed FISH measures gene expression while preserving spatial information



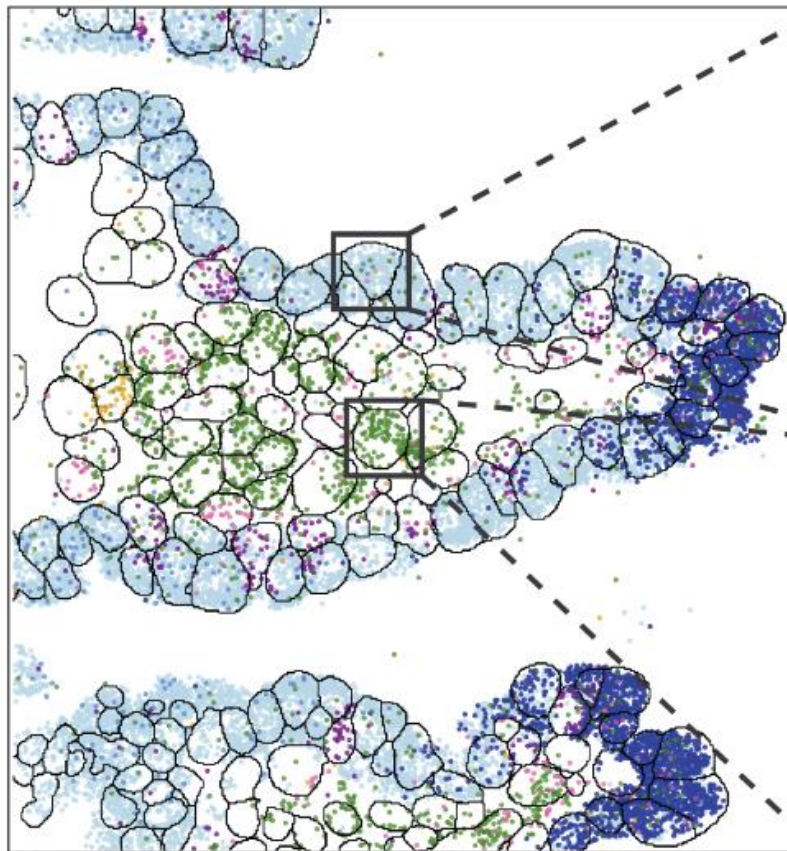
Polaris: An integrated pipeline for spatial transcriptomics



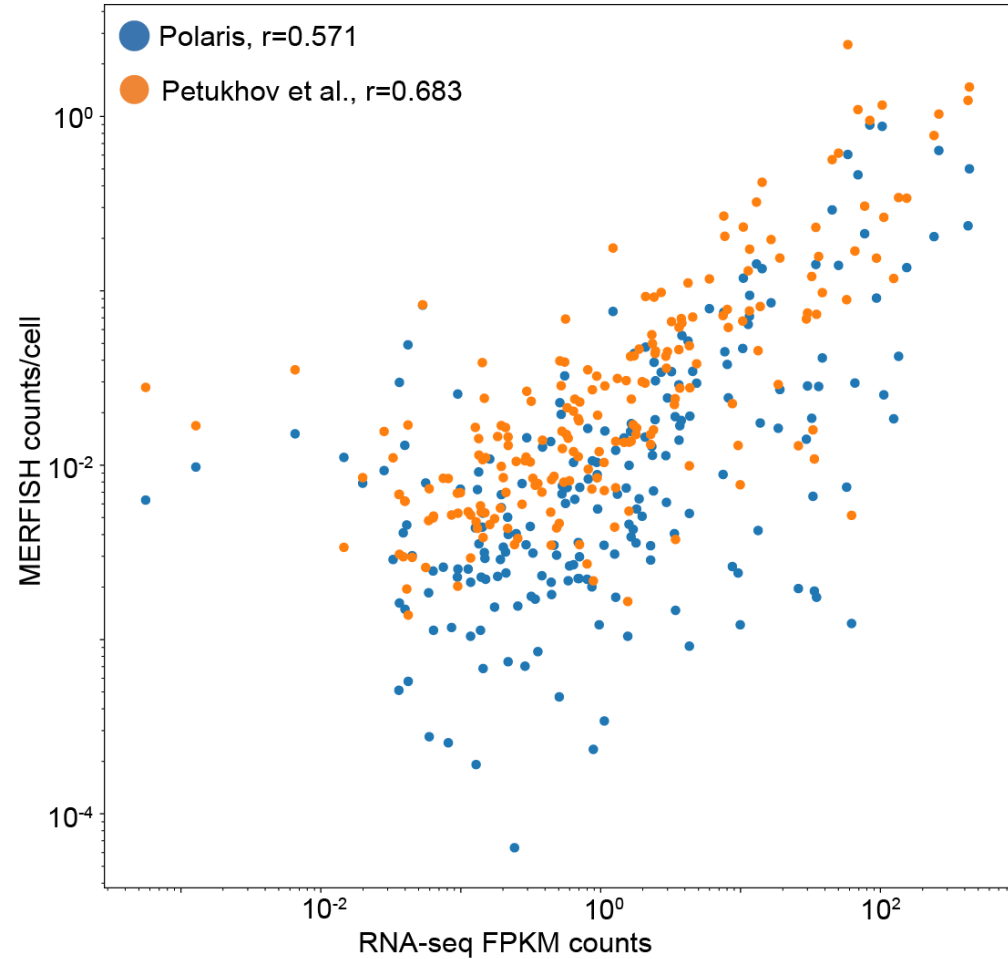
Polaris: A unified pipeline for spatial transcriptomics



Polaris: A unified pipeline for spatial transcriptomics



Enterocyte
Endothelial cell
Goblet cell
Enterocyte (Top villus)
Immune cell
Stem cell



Thanks!

The Van Valen Lab

Edward Pao

Ross Barnowski

Uriah Israel

Emily Laubscher

Morgan Schwartz

Changhua Yu

Xuefei Wang

Rohit Dilip

Qilin Qu

Jonathan Soro

Elora Pradhan

Ada Ates

Kevin Yu

Tiny Nanazian

Tom Dougherty

Geneva Miller

Will Graf

Erick Moen

Lincoln Ombelets

Ekta Patel

The Angelo Lab

Noah Greenwald

Michael Angelo

The Keren Lab

Leeat Keren

The Cai Lab

Long Cai

Noushin Koulana

Nico Pierson

The Yue Lab

Uriah Israel

Cloud Posse

Eric Osterman

