

# **Zoom & Enhance: Enabling Subcellular Analysis of Spatial Transcriptomics with Bento**

Clarence Mah | Postdoctoral Researcher  
Sanford Stem Cell Innovation Center  
Yeo Lab | UC San Diego

# About me

UC San Diego



Current  
Postdoc, Yeo Lab UCSD



**Gene Yeo, UCSD**

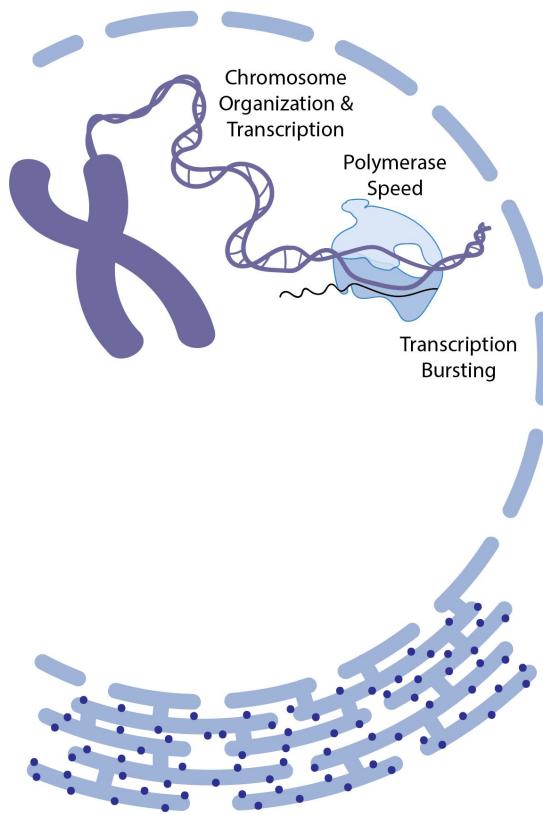
- RNA binding proteins
- Neurodegenerative diseases
- Spatial transcriptomics



**Hannah Carter, UCSD**

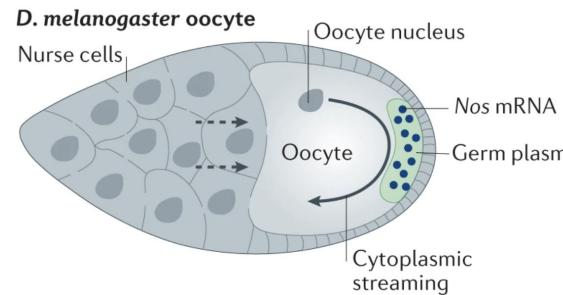
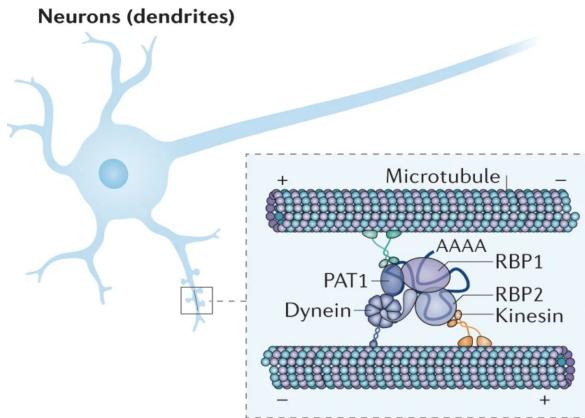
- Cancer biology
- Human genetics
- Variant interpretation



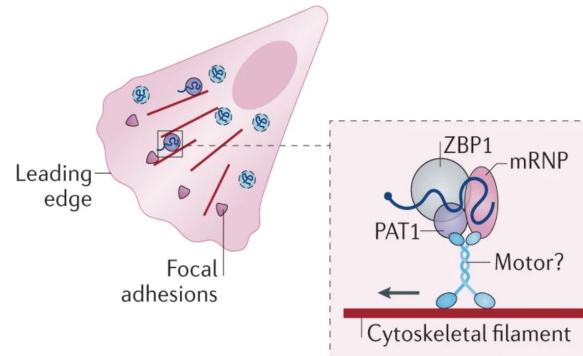


RNA is regulated in  
both **space & time**

# RNA localization is a conserved mechanism modulating translation

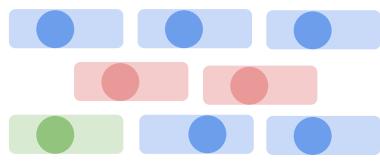


Fibroblast leading edge

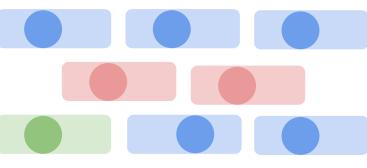


# Genomic technologies for studying RNA biology at scale

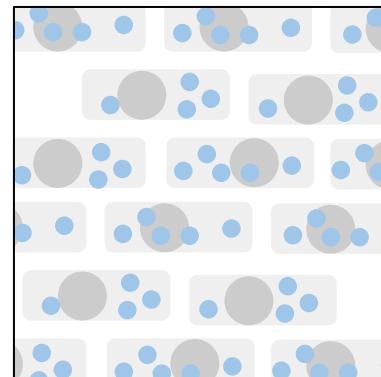
RNA sequencing



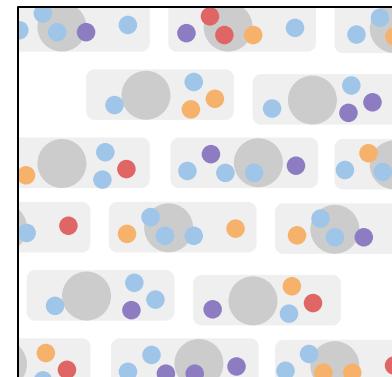
Single-cell  
RNA-seq



Spatial transcriptomics



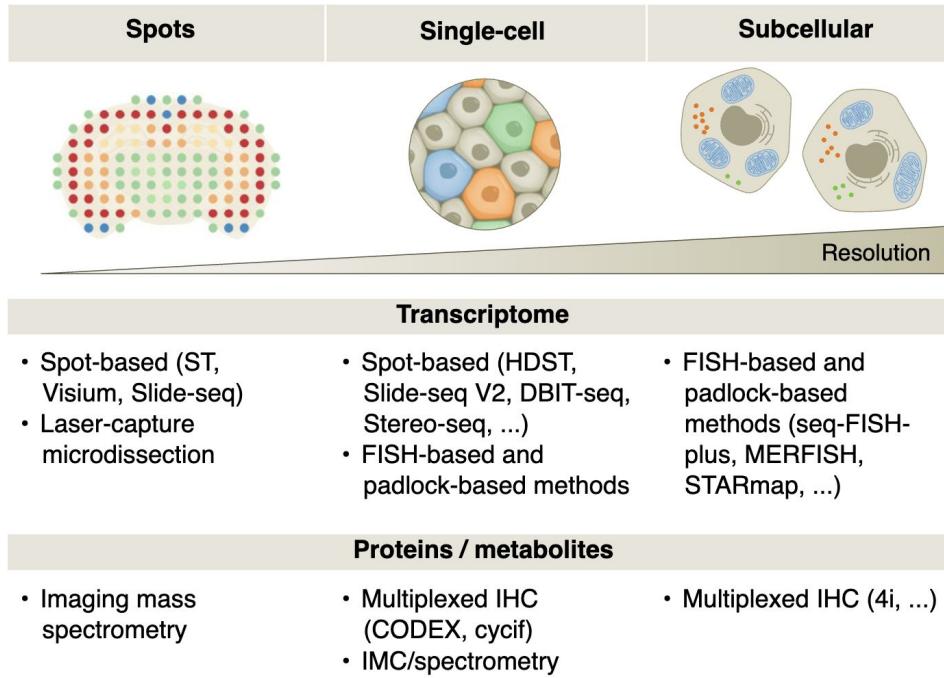
Spatial transcriptomics  
(multiplexed)



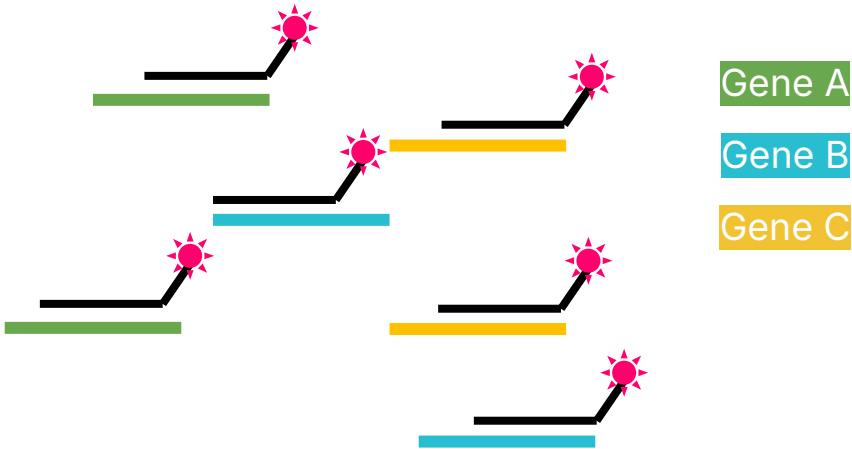
Granularity  
Less More

Genes  
10s 1000s

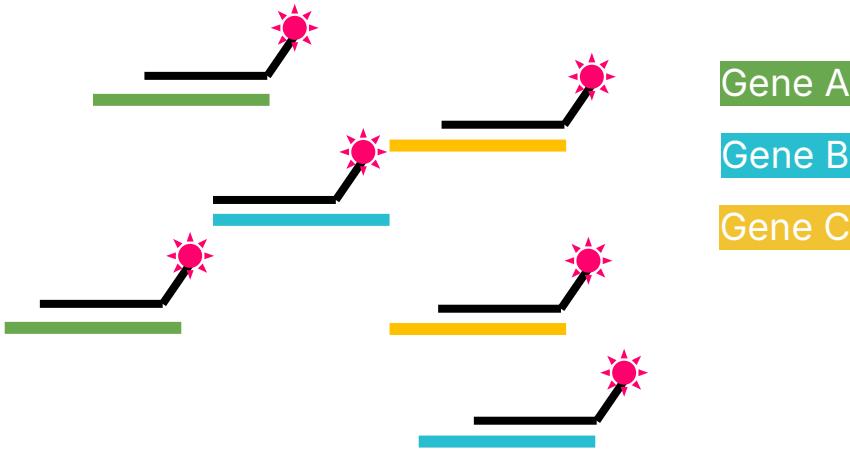
# Technologies span a range of resolutions



# FISH/probe imaging methods capture single-molecule resolution data



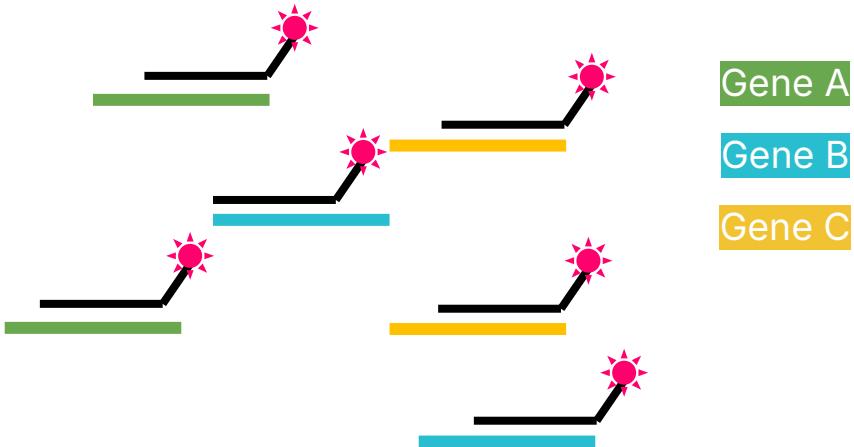
# FISH/probe imaging methods capture single-molecule resolution data



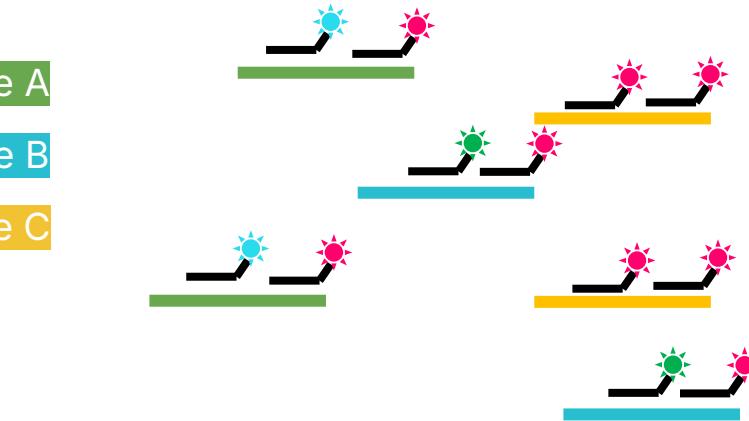
**Single-molecule FISH**

1-10 genes

# FISH/probe imaging methods capture single-molecule resolution data



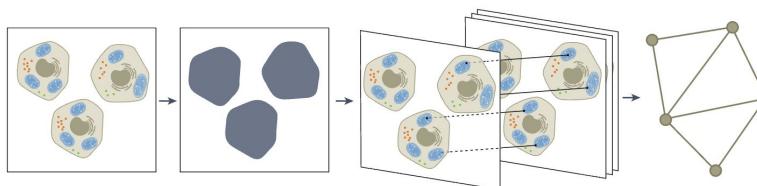
**Single-molecule FISH**  
1-10 genes



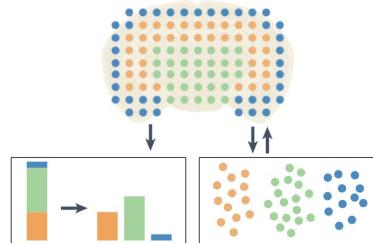
**Multiplexed FISH/probes**  
e.g. MERFISH, seqFISH, STARmap etc.  
100-10,000 genes

# Tissue/single-cell level analysis are accessible

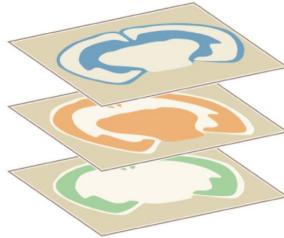
**a** Image processing, segmentation, coordinate registration and data structure



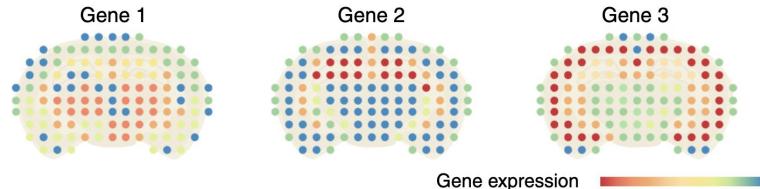
**b** Deconvolution and data integration



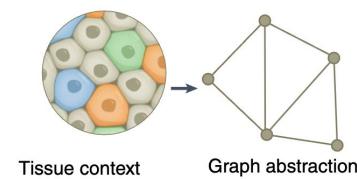
**c** Multimodal analysis



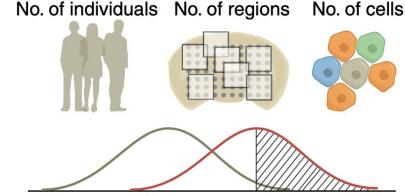
**d** Spatially variable features



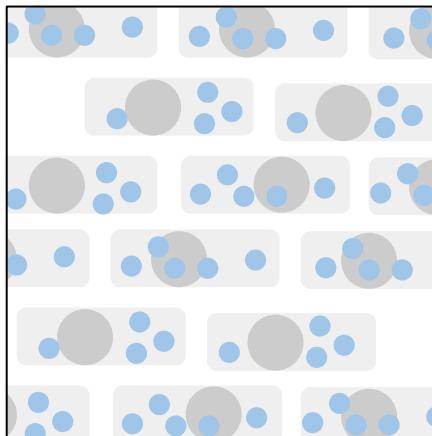
**e** Cellular neighborhood



**f** Spatial power analysis



# Spatial transcriptomics enables measuring RNA localization at scale



Spatial transcriptomics



- Gene A
- Gene B
- Gene C
- Gene D
- Gene E

# Bento: A toolkit for subcellular analysis of spatial transcriptomics data

## RNA Imaging Data



seqFISH  
MERFISH  
STARmap  
...



```
pip install bento-tools
```



Github



Paper



Clarence Mah  
[@mah\\_clarence](https://twitter.com/mah_clarence)



Noorsher Ahmed  
[@NoorsherAhmed](https://twitter.com/NoorsherAhmed)

# Bento uses SpatialData to standardize spatial analysis



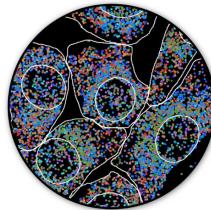
Dylan Lam

## RNA Imaging Data

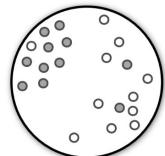


## Standardized Data

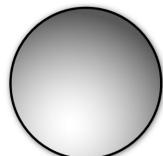
- AnnData
- Segmentation masks
- Molecular coordinates



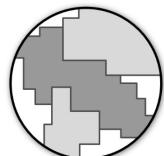
## Machine Learning & Statistical Analysis



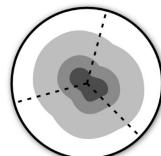
Geospatial  
Statistics



Molecular  
Gradients

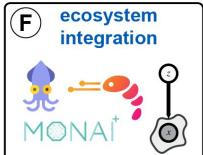
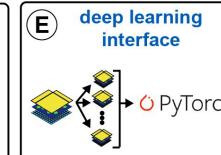
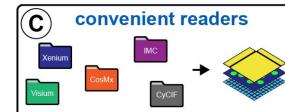
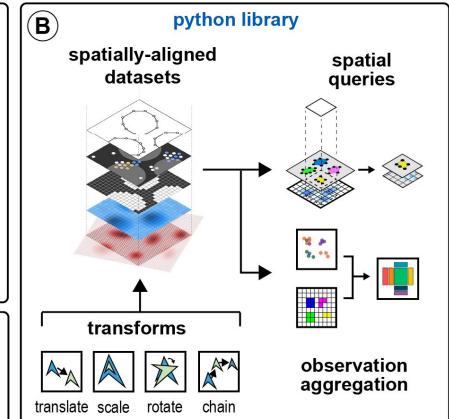
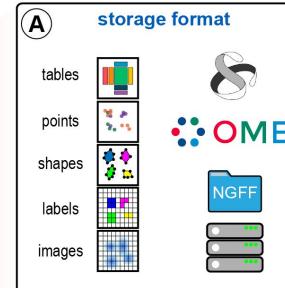


Subcellular  
Compartments



Data  
Visualization

## SpatialData



# Workshop Topics

How can we conceptualize and model single-molecule data?

1.  **Spatial features:** measuring physical relationships
2.  **Rasters:** image-based representations

# Feature spaces can get complex

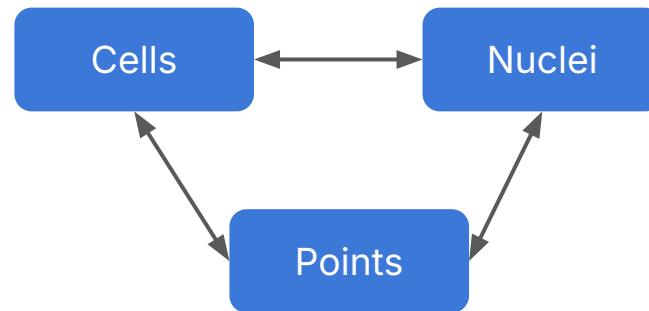
Simple features **easy**

- Single object measures
- Examples: cell area, perimeter, diameter, solidity, etc.

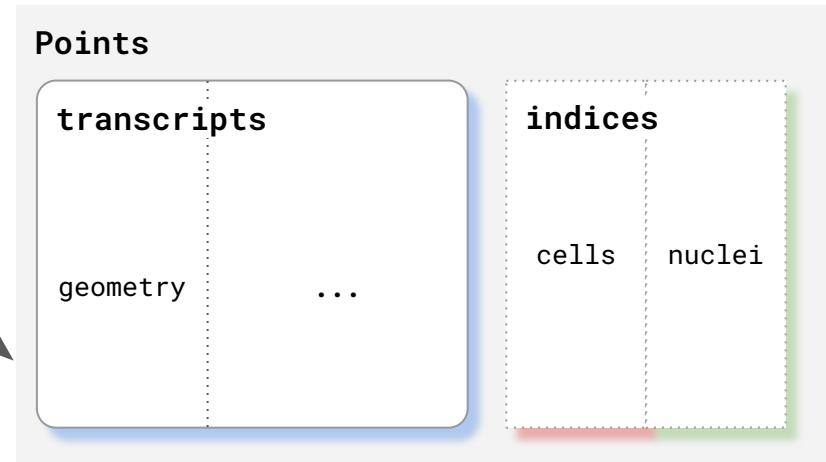
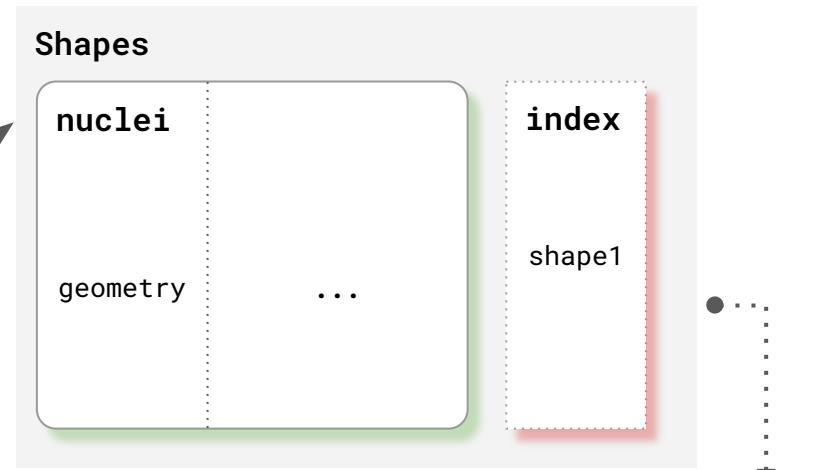
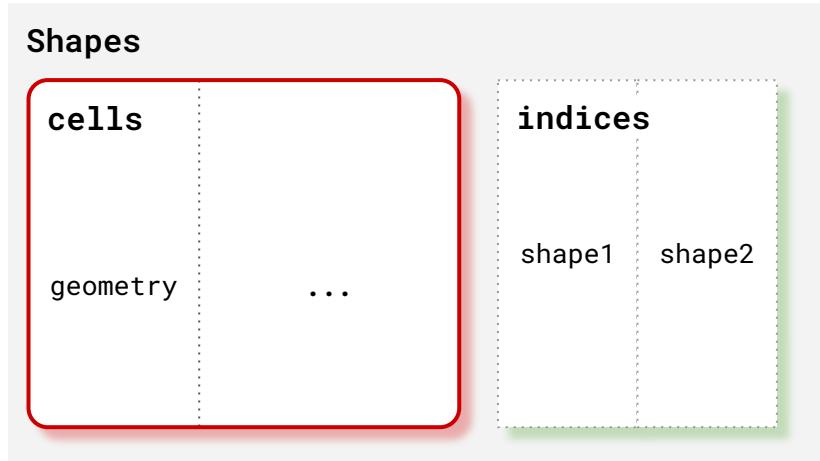


Complex features **hard**

- Multi-object measures
- Example: distance between points and cell boundary

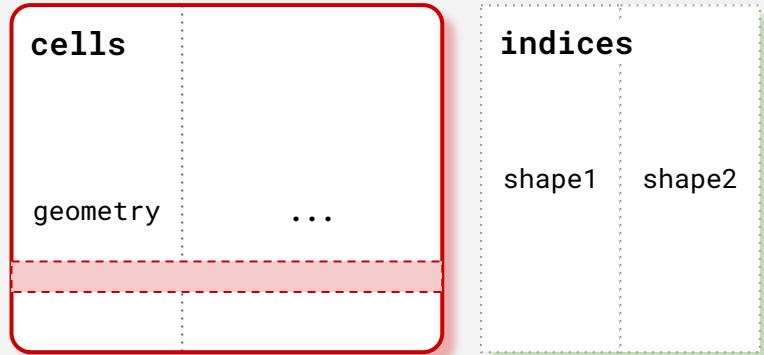


# Indexing enables fast spatial querying



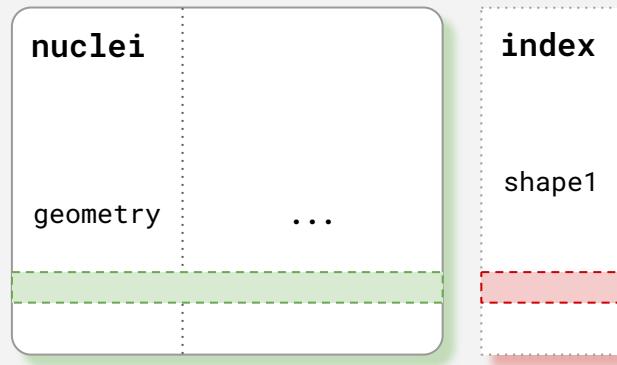
# Now we can compute features on spatial combos of objects

## Shapes



$$f(\text{red box}, \text{green box}, \text{blue box})$$

## Shapes



## Points



# RNAforest: Supervised learning to identify localization patterns

## Spatial transcriptomics dataset

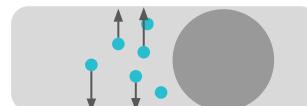
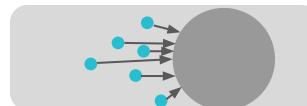
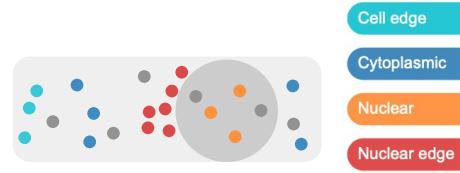
	Gene A	Gene B	Gene C	...
Cell D				
Cell E				
Cell F				
...				

$$f(\textcolor{red}{\square}, \textcolor{green}{\square}, \textcolor{blue}{\square})$$

Encode spatial features (13)

- Proximity
- Asymmetry
- Dispersion
- Density

Predict



ML models



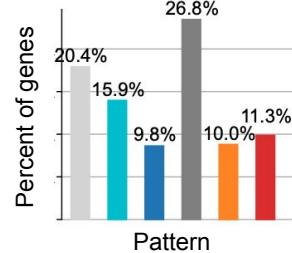
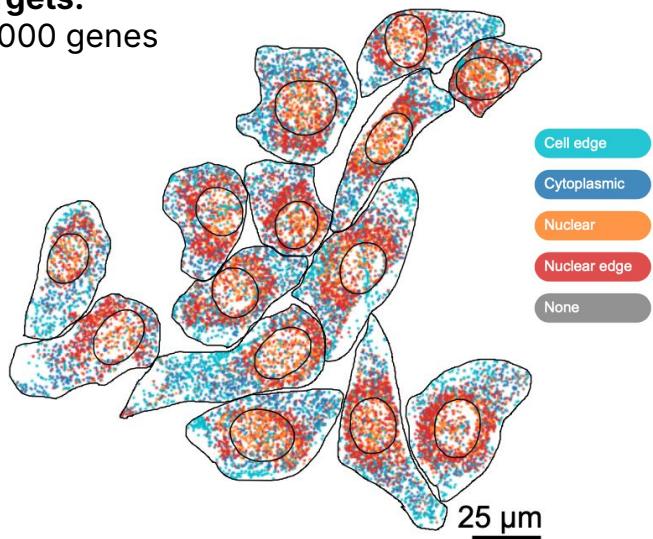
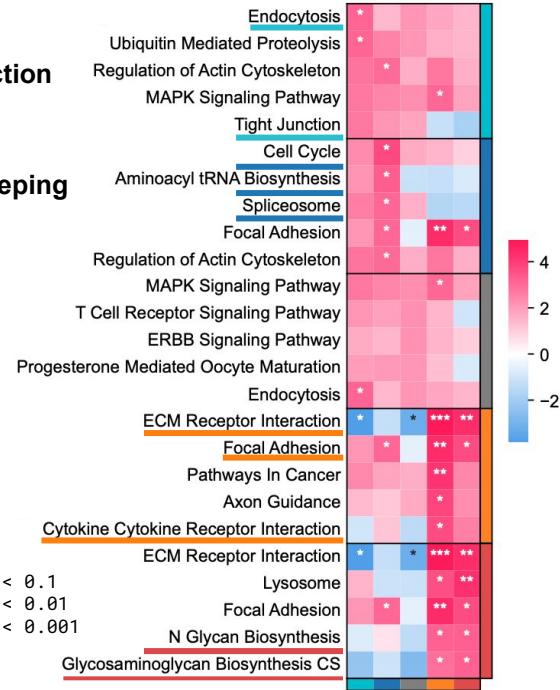
# RNA transcript localization is associated with cellular functions

**Cell type:**

Embryonic mouse fibroblast cell line (3T3)

**Targets:**

10,000 genes


**Local function**
**Housekeeping**
**Gene set enrichment**


# Spatial features data model

## Pros

- Physically constrained
- Highly interpretable features
- Descriptive feature space

## Cons

- Reliant on quality of segmentation data
- Requires feature engineering

## Ideas

- cell x subcell. spatial features →
  - High resolution cell typing
  - Representation learning of single-cells
  - Supervised learning tasks



# Bento API (parallelized with Dask!)

## Spatial indexing of SpatialData object

```
bt.io.prepdata)
```

## Compute spatial features

```
# Shape features
bt.tl.analyze_shapes(
    sdata,
    shape_keys=["cell_boundaries", "nucleus_boundaries"],
    feature_names=["area", "radius", "perimeter"]
)

# Point features
bt.tl.analyze_points(
    sdata,
    shape_keys="cell_boundaries",
    feature_names=["distance", "asymmetry"],
    groupby="gene"
)
```

```
bt.tl.list_shape_features()
```

```
[  
    'area',  
    'aspect_ratio',  
    'bounds',  
    'density',  
    'opening',  
    'perimeter',  
    'radius',  
    'raster',  
    'second_moment',  
    'span'  
]
```

```
bt.tl.list_point_features()
```

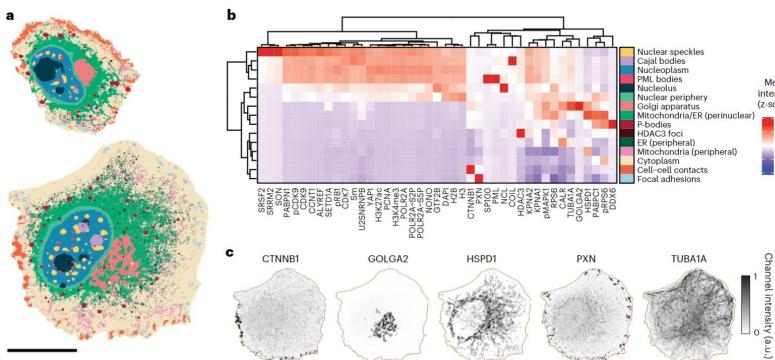
```
[  
    'proximity',  
    'asymmetry',  
    'point_dispersion_norm',  
    'shape_dispersion_norm',  
    'distance',  
    'offset',  
    'point_dispersion',  
    'shape_dispersion',  
    'ripley',  
    'shape_enrichment',  
]
```

API for adding your custom features **WIP**  
(talk to me if interested!)



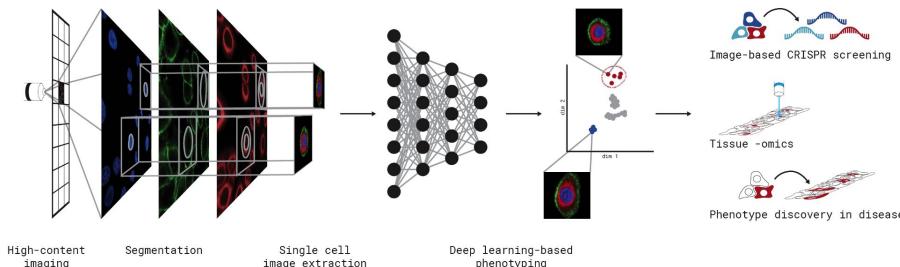
# Protein imaging is a computer vision problem

CAMPA



Hannah et al. 2023, *Nature Methods*

# scPortrait



Schmake et al. 2023. *BioRxiv*



# Spatial tx as a computer vision problem

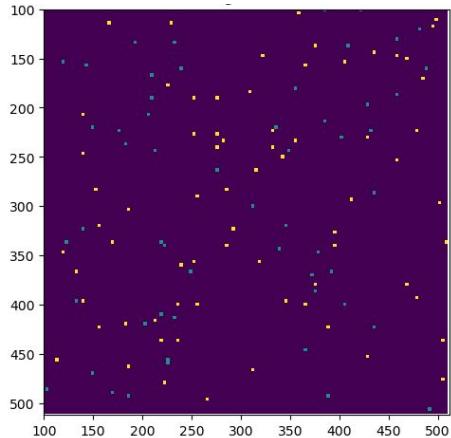
Coordinates

	x	y	gene_id
0	1	1	gene_a
1	2	3	gene_a
2	4	5	gene_b

Shape: (3, N)

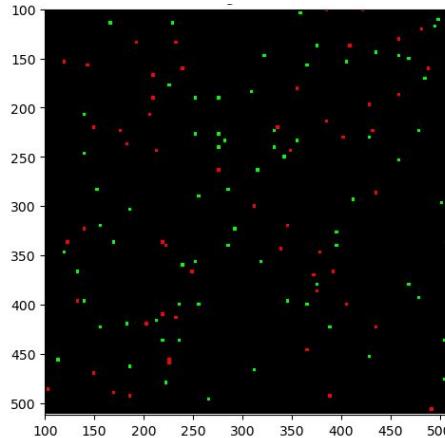
raster

Single-channel



Shape: (1, yrangle, xrange)

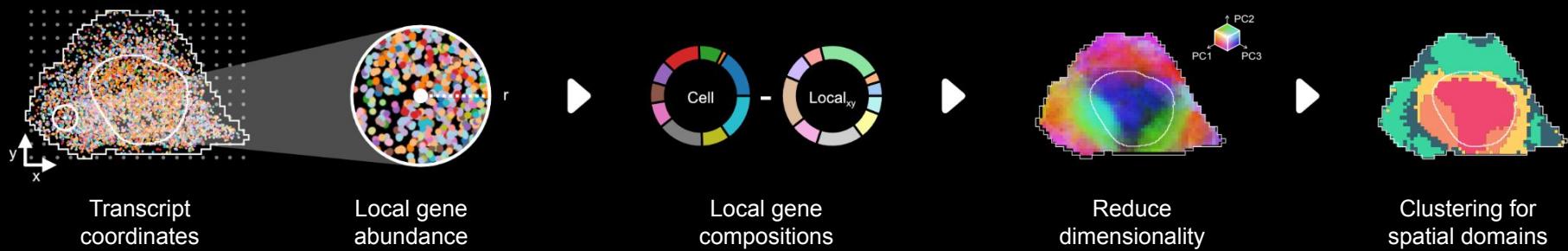
Multi-channel (per gene)



Shape: (G, yrangle, xrange)



# RNAflux: an unsupervised approach to quantify spatial transcriptomic gradients

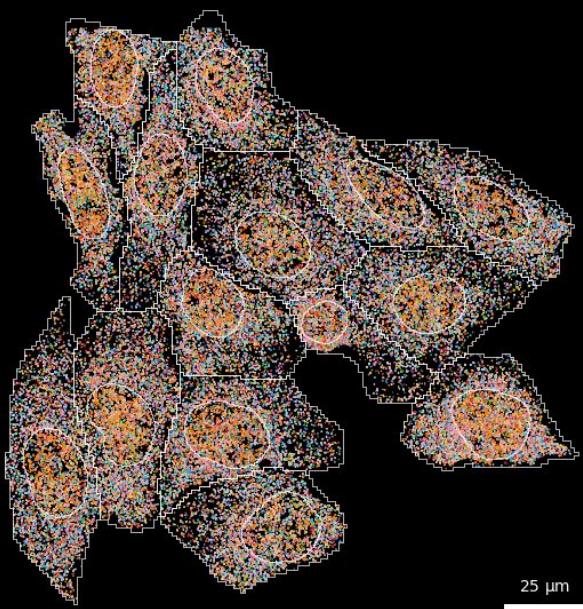


# RNAflux embeddings are consistent across cells



Dr. Quan Zhu

Transcripts



## MERFISH

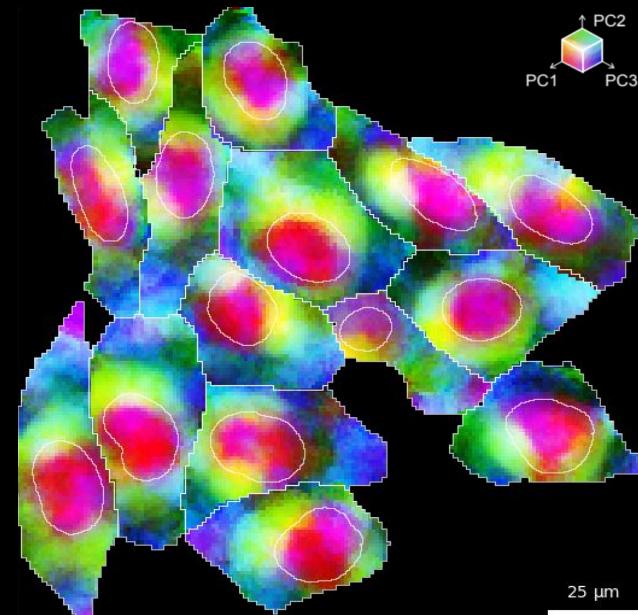
1153 U2OS cells, 130 genes



Embed + PCA

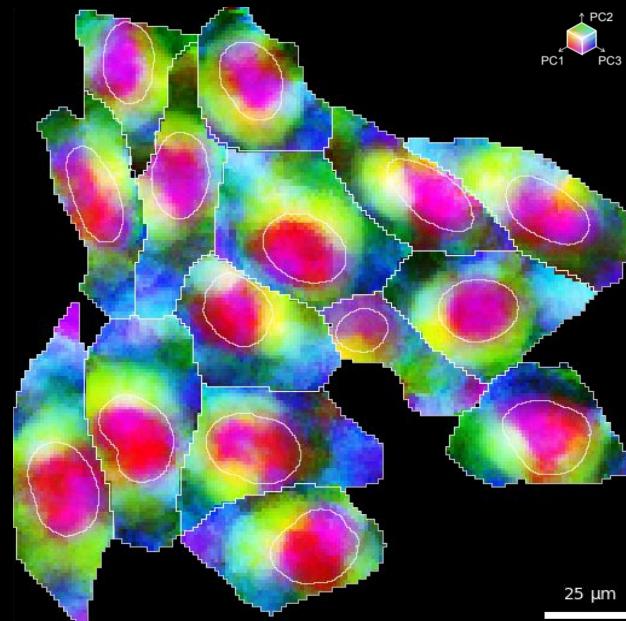


RNAflux embedding



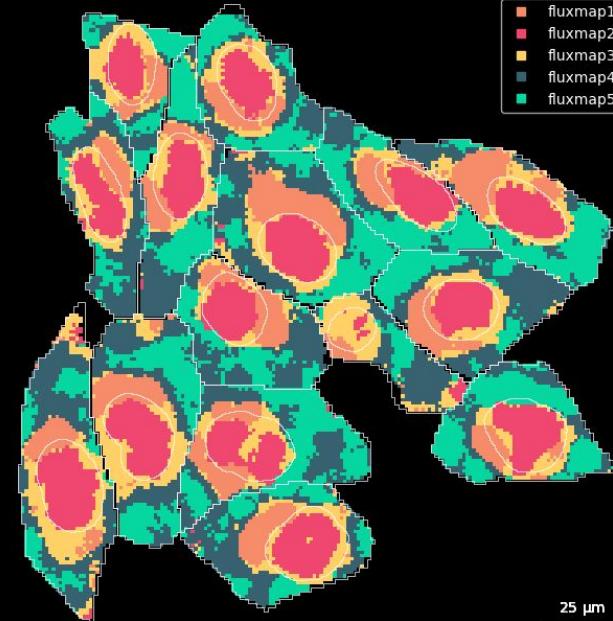
# Unsupervised clustering identifies consistent subcellular domains

RNAflux embedding



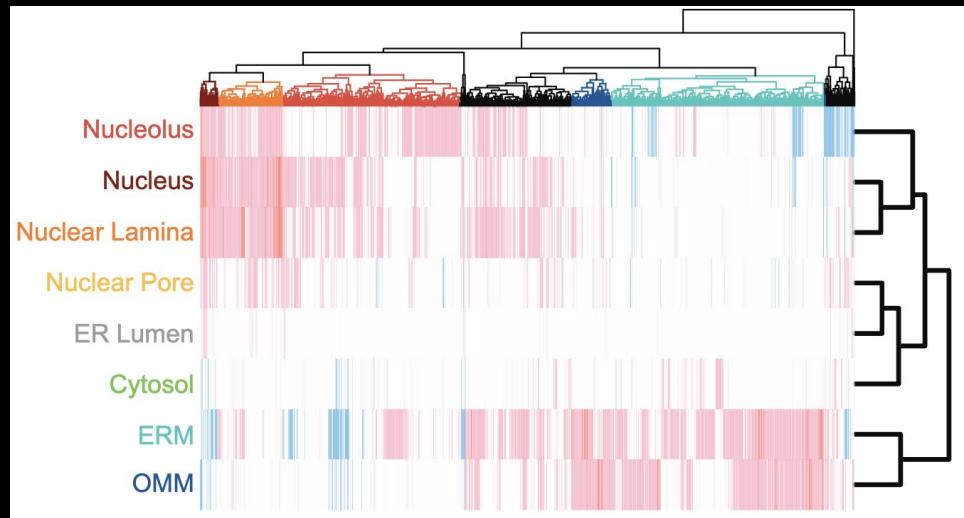
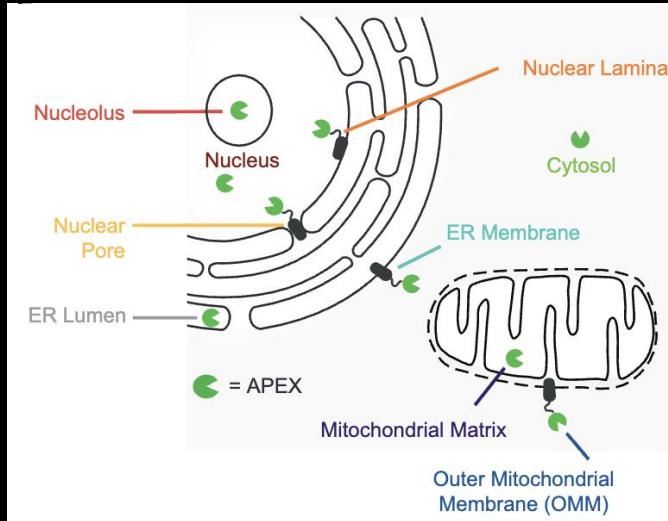
SOM cluster embeddings

Fluxmap domains



# Integrating known local expression signatures

8 local expression signatures



Differentially  
expressed genes  
n = 3288

MERFISH panel overlap  
60/130 overlap

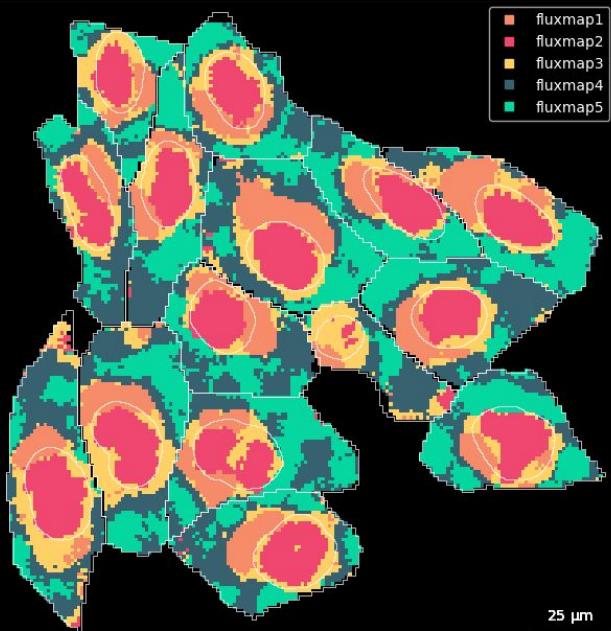


# Fluxmap domains are enriched for organelle expression

Enrichment of local expression signatures

	1	2	3	4	5
Cytosol	0.19	0.10	0.43	0.82	0.94
ER Lumen	0.68	0.75	0.60	0.28	0.23
ERM	0.90	0.61	0.70	0.28	0.09
Lamina	0.20	0.94	0.51	0.25	0.28
Nuclear Pore	0.63	0.78	0.53	0.32	0.20
Nucleolus	0.14	0.84	0.48	0.38	0.50
Nucleus	0.10	0.96	0.50	0.24	0.34
OMM	0.99	0.51	0.68	0.26	0.04

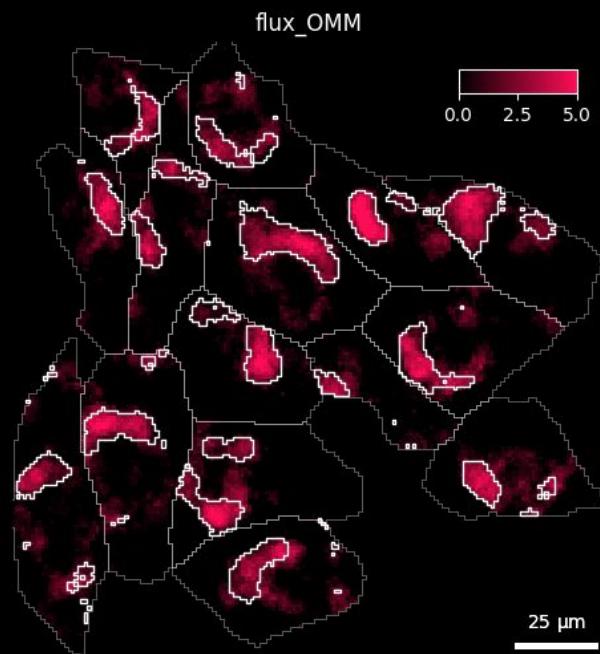
Fluxmap domains



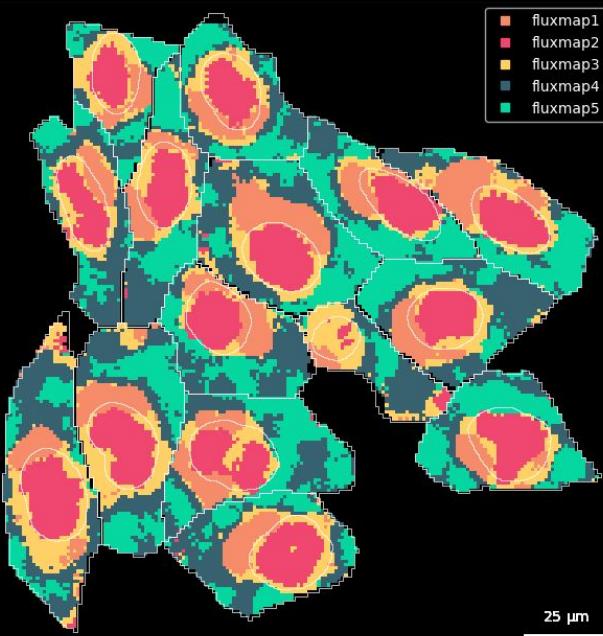
# Fluxmap 1 ~ Mitochondria signature

Enrichment of local expression signatures

	1	2	3	4	5
Cytosol	0.19	0.10	0.43	0.82	0.94
ER Lumen	0.68	0.75	0.60	0.28	0.23
ERM	0.90	0.61	0.70	0.28	0.09
Lamina	0.20	0.94	0.51	0.25	0.28
Nuclear Pore	0.63	0.78	0.53	0.32	0.20
Nucleolus	0.14	0.84	0.48	0.38	0.50
Nucleus	0.10	0.96	0.50	0.24	0.34
OMM	0.99	0.51	0.68	0.26	0.04



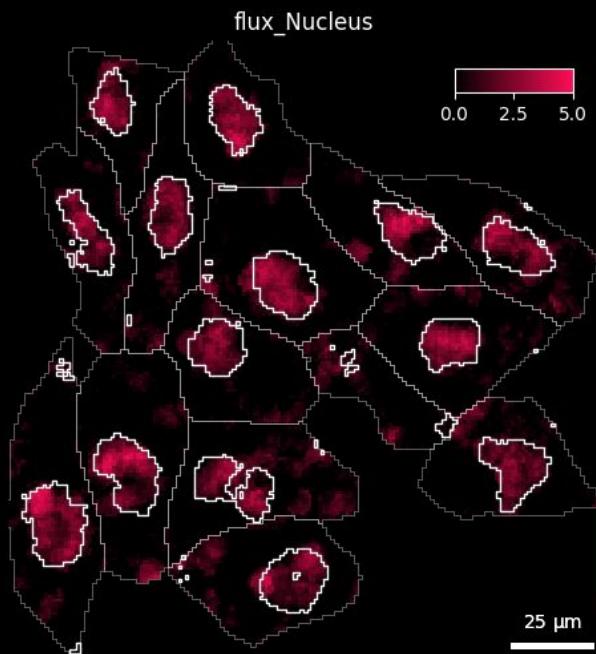
Fluxmap domains



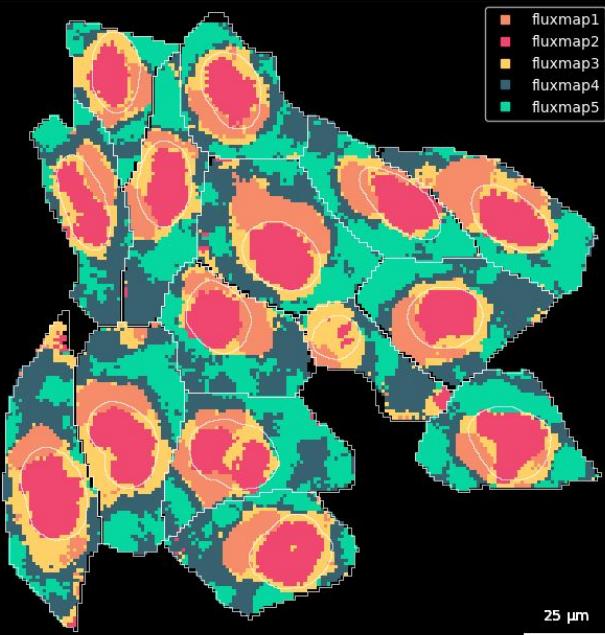
# Fluxmap 2 ~ Nucleus, nuclear pore, nuclear lamina, etc.

Enrichment of local expression signatures

	1	2	3	4	5
Cytosol	0.19	0.10	0.43	0.82	0.94
ER Lumen	0.68	0.75	0.60	0.28	0.23
ERM	0.90	0.61	0.70	0.28	0.09
Lamina	0.20	0.94	0.51	0.25	0.28
Nuclear Pore	0.63	0.78	0.53	0.32	0.20
Nucleolus	0.14	0.84	0.48	0.38	0.50
Nucleus	0.10	0.96	0.50	0.24	0.34
OMM	0.99	0.51	0.68	0.26	0.04



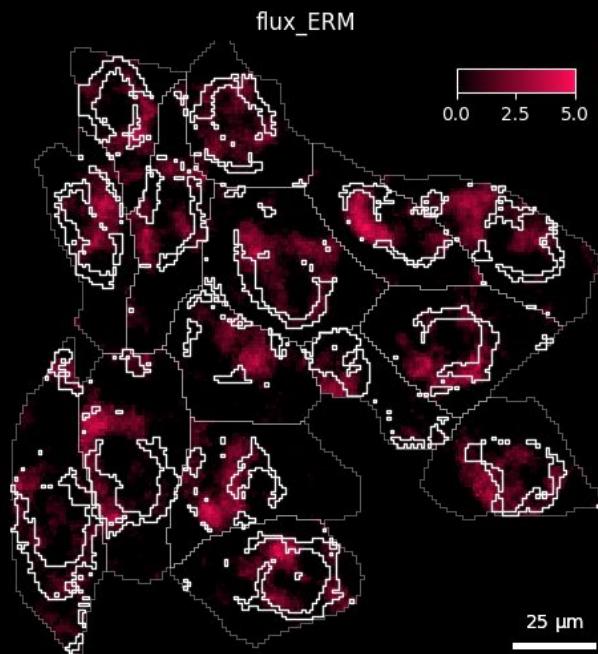
Fluxmap domains



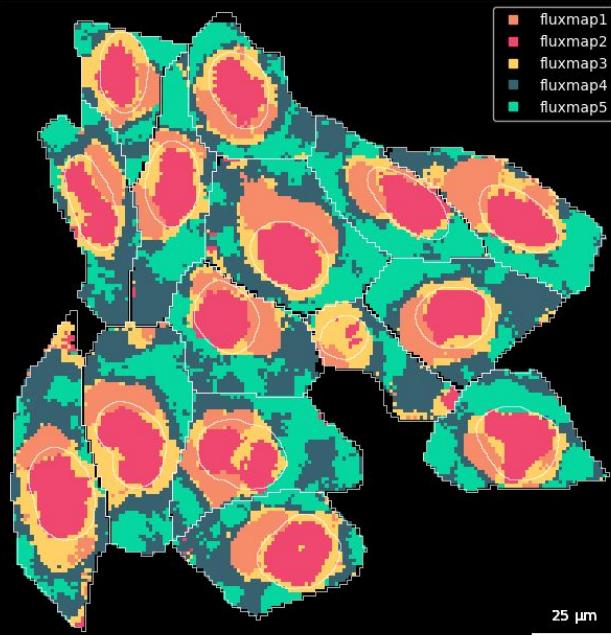
# Fluxmap 3 ~ Endoplasmic reticulum signature

Enrichment of local expression signatures

	1	2	3	4	5
Cytosol	0.19	0.10	0.43	0.82	0.94
ER Lumen	0.68	0.75	0.60	0.28	0.23
ERM	0.90	0.61	0.70	0.28	0.09
Lamina	0.20	0.94	0.51	0.25	0.28
Nuclear Pore	0.63	0.78	0.53	0.32	0.20
Nucleolus	0.14	0.84	0.48	0.38	0.50
Nucleus	0.10	0.96	0.50	0.24	0.34
OMM	0.99	0.51	0.68	0.26	0.04



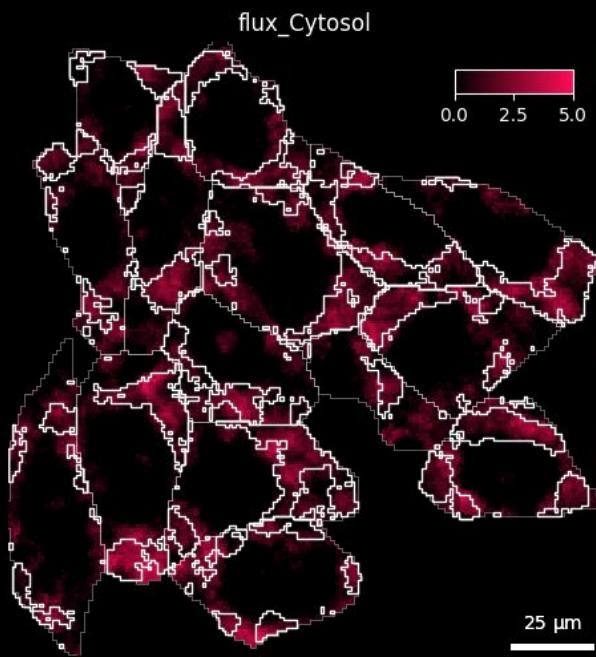
Fluxmap domains



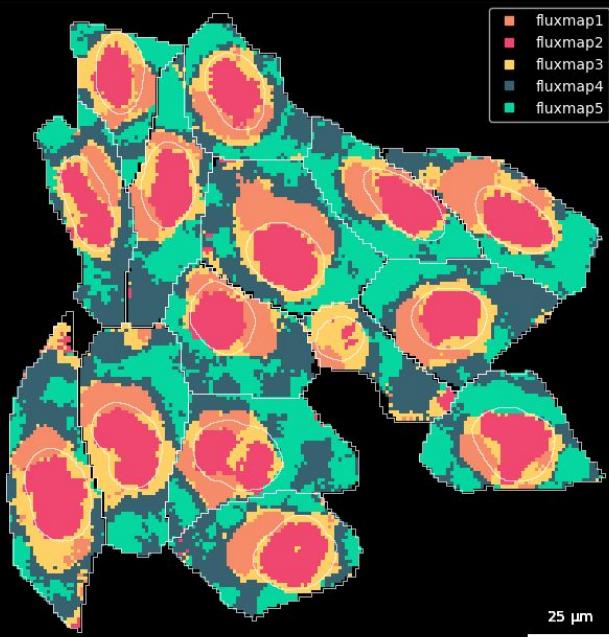
# Fluxmap 4 ~ Cytosol signature

Enrichment of local expression signatures

	1	2	3	4	5
Cytosol	0.19	0.10	0.43	0.82	0.94
ER Lumen	0.68	0.75	0.60	0.28	0.23
ERM	0.90	0.61	0.70	0.28	0.09
Lamina	0.20	0.94	0.51	0.25	0.28
Nuclear Pore	0.63	0.78	0.53	0.32	0.20
Nucleolus	0.14	0.84	0.48	0.38	0.50
Nucleus	0.10	0.96	0.50	0.24	0.34
OMM	0.99	0.51	0.68	0.26	0.04



Fluxmap domains



# Bento API

Currently performs custom  
rasterization...

```
import bento as bt

# RNAflux embedding
bt.tl.flux(sdata, res=1)

# Predict subcellular compartments
bt.tl.fluxmap(sdata, n_clusters=5, res=1)

# Gene set enrichment analysis
bt.tl.fe(sdata, genesets)
```

Integration with SpatialData  
rasterization coming soon!

```
import spatialdata as sd

sd.rasterize(
    sdata,
    axes=['x', 'y'],
    min_coordinate=[0, 0],
    max_coordinate=[1000, 1000],
    target_coordinate_system='global'
)
```



# Rasterization data model

## Pros

- Flexible across length scales
- Robust to segmentation quality
- Amenable to computer vision inspired analysis
- Dimensions are in gene space

## Cons

- Sparse signal compared to fluorescent imaging for protein
- Dependent on technology detection efficiency



# Future Work

- Expand spatial feature sets
- Improving integration with SpatialData
- Support for 3D analysis

What applications are you interested in?



# Thanks for listening! Questions?

Yeo Lab (UCSD)

Gene Yeo

Noorsher Ahmed

Nicole Lopez

Dylan Lam

Alex Monell

Avery Pong

Gino Prasad

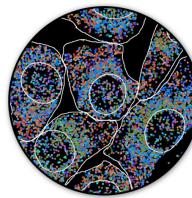
## RNA Imaging Data



seqFISH  
MERFISH  
STARmap  
...

## Standardized Data

- AnnData
- Segmentation masks
- Molecular coordinates



ZULIP #bento

BENTO



pip install bento-tools

## Carter Lab (UCSD)

Hannah Carter

Adam Klie

David

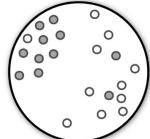
## Center of Epigenomics (UCSD)

Quan Zhu

Colin Kern

Yuanyuan Han

## Machine Learning & Statistical Analysis



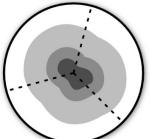
Geospatial Statistics



Molecular Gradients



Subcellular Compartments



Data Visualization



Github



Paper

## Lundberg Lab (Stanford)

Emma Lundberg

Anthony Cesnik



Bioinformatics & Systems Biology



Clarence Mah

@mah\_clarence