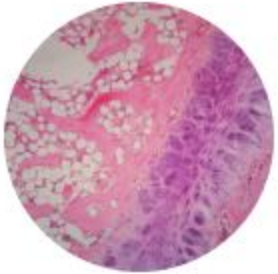


10PM

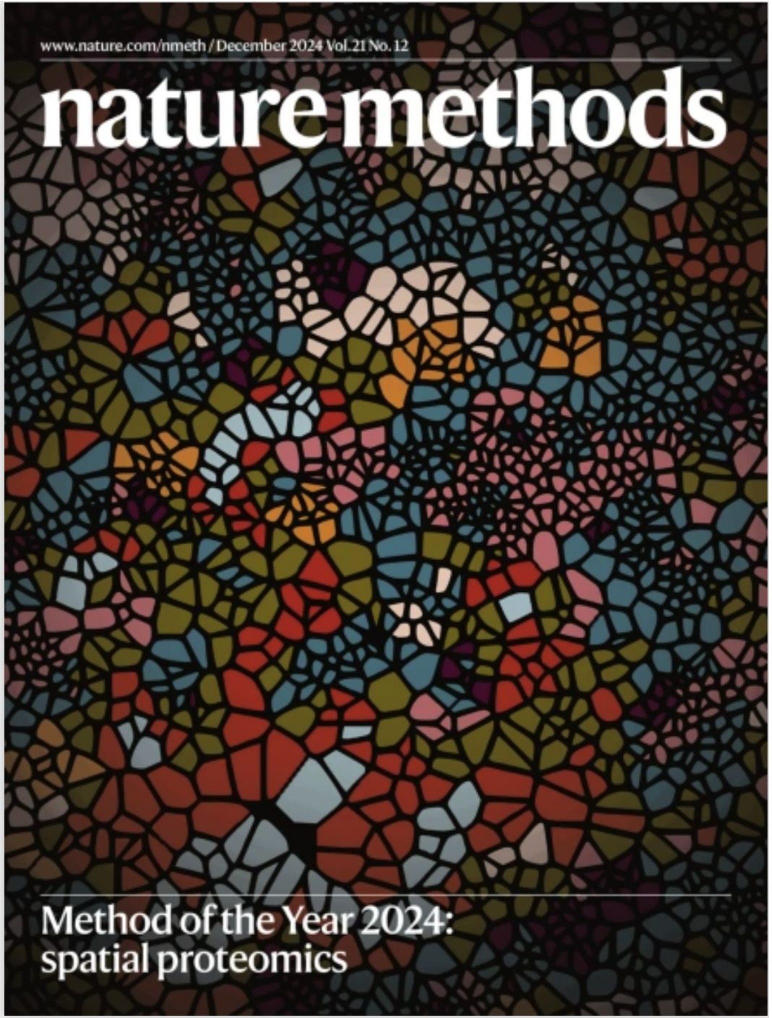
3AM in London (GMT), 12PM in Tokyo (GMT+9)

STELLAR & Cell Neighborhoods

Moderator: Ellen Quardokus, *Indiana University*

Presenters:

- John Hickey, *Duke University*
- Jure Leskovec, *Stanford University*

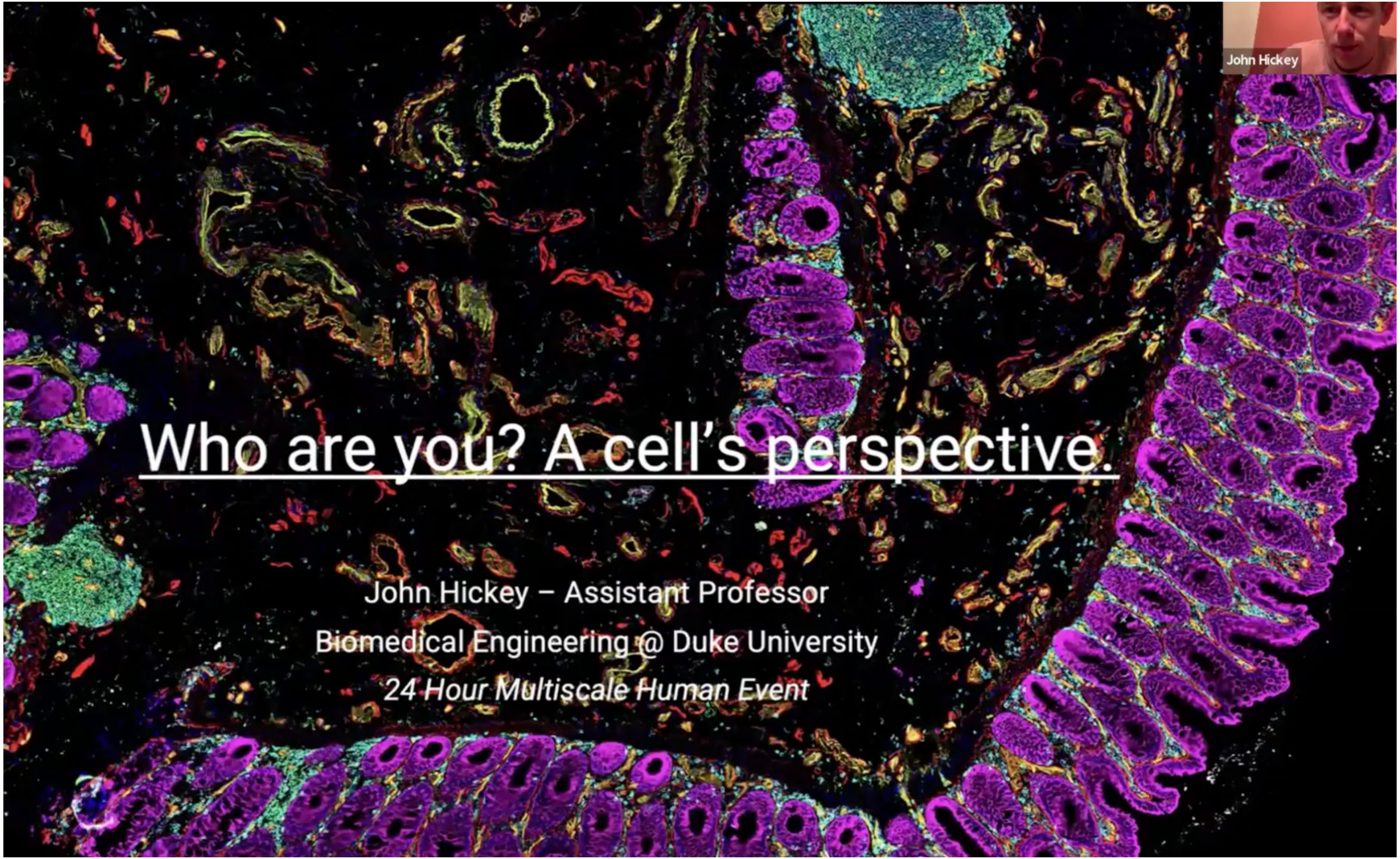


Method of the Year 2024: spatial proteomics

Editorial | 06 Dec 2024

https://www.nature.com/nmeth/volumes/21/issues/12

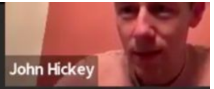
John Hickey, *Duke University*



Who are you? A cell's perspective.

John Hickey – Assistant Professor
Biomedical Engineering @ Duke University

24 Hour Multiscale Human Event



Who am I?

It Depends on What You Analyze

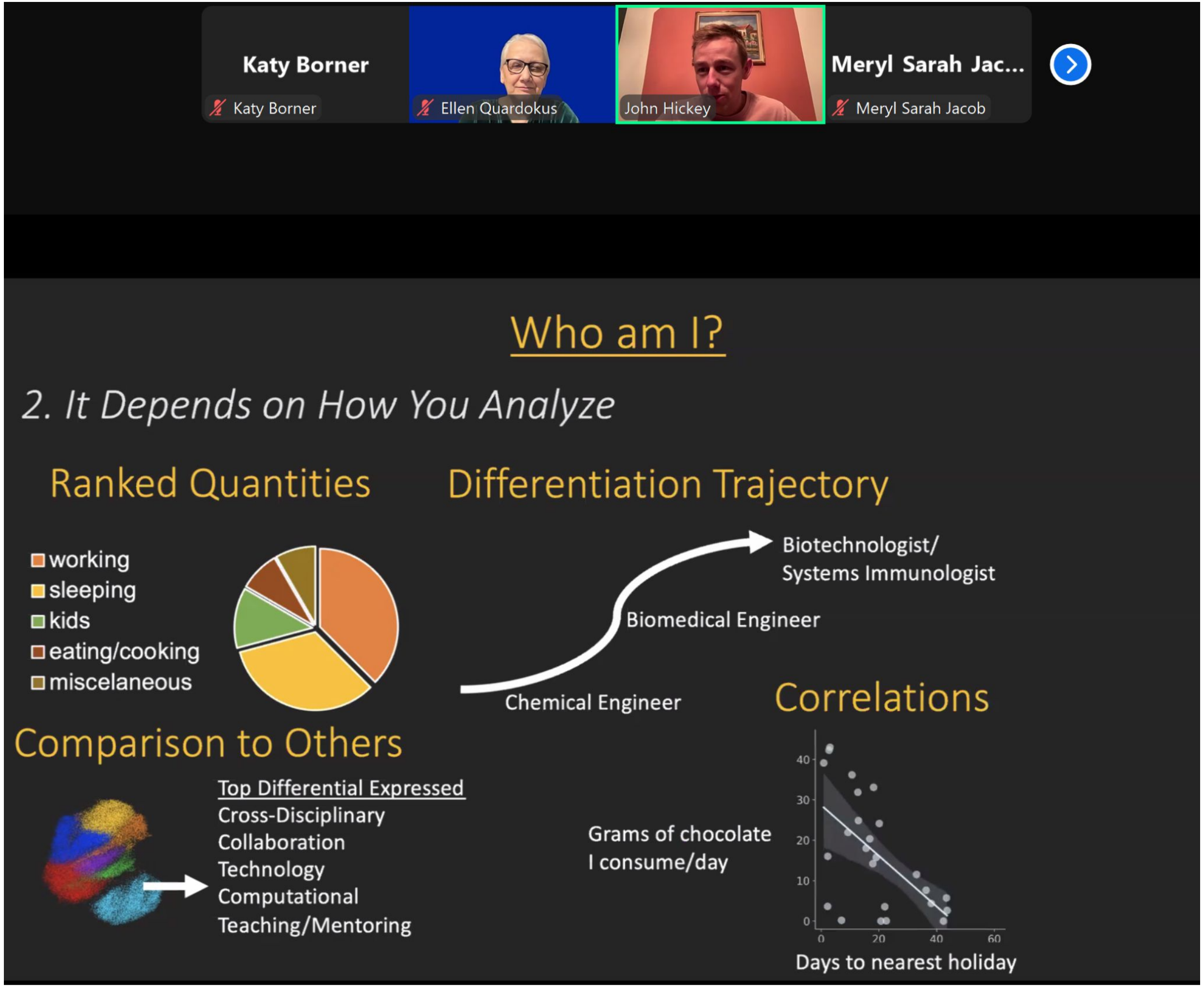
My "origin"

My "genotype"

My "phenotype"

My "state"

What I "do"



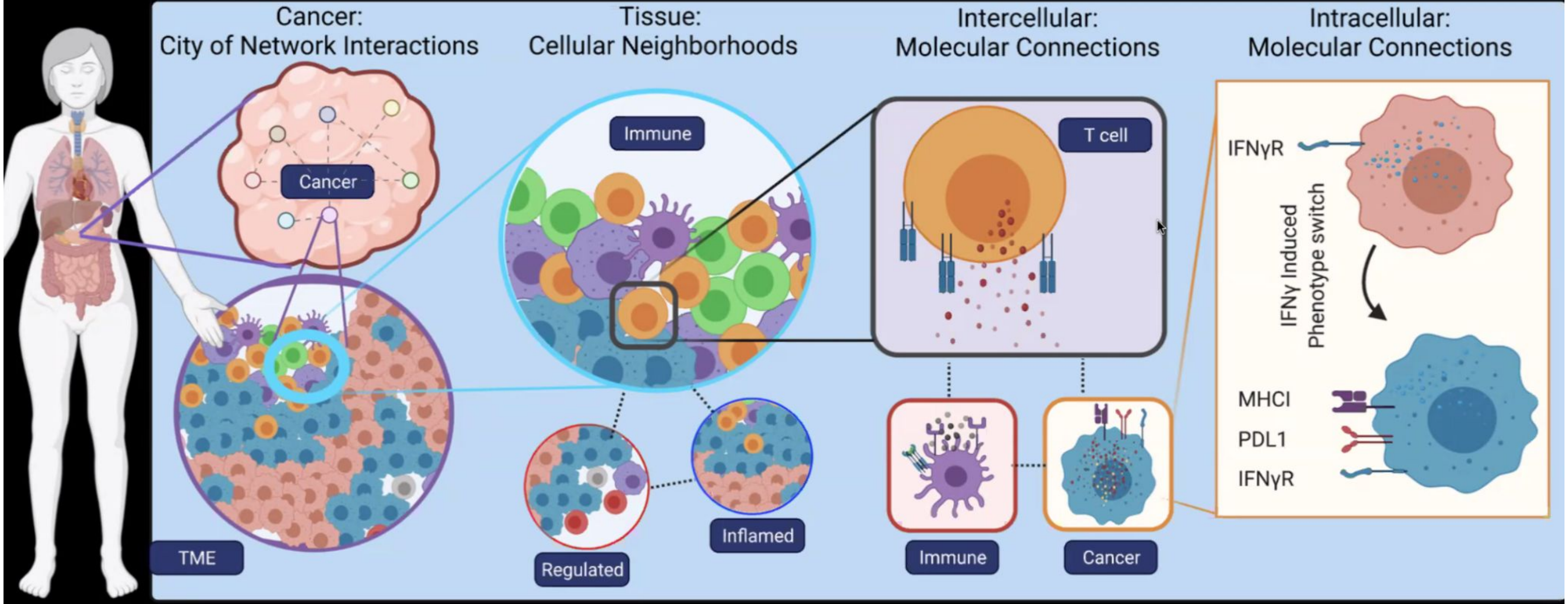
Katy Borner
Katy Borner
Ellen Quardokus
John Hickey
Meryl Sarah Jac...
Meryl Sarah Jacob
Who am I?
2. It Depends on How You Analyze
Ranked Quantities
working
sleeping
kids
eating/cooking
miscelaneous
Differentiation Trajectory
Biotechnologist/
Systems Immunologist
Biomedical Engineer
Chemical Engineer
Comparison to Others
Top Differential Expressed
Cross-Disciplinary
Collaboration
Technology
Computational
Teaching/Mentoring
Correlations
Grams of chocolate
I consume/day
40
30
20
10
0
0
20
40
60
Days to nearest holiday

How would you define yourself if you were a cell?

As a human cell, I am a tiny, yet essential unit of life, working tirelessly within a vast, interconnected community to maintain balance and support the body's functions. I carry the blueprint of existence within my nucleus, adapt to my environment, and collaborate with neighboring cells to ensure the survival and well-being of the organism I inhabit.

ChatGPT 2024

Tissues: Multi-scale Network of Interactions

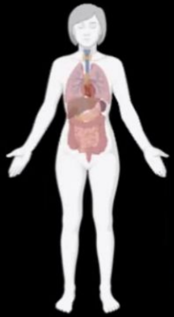


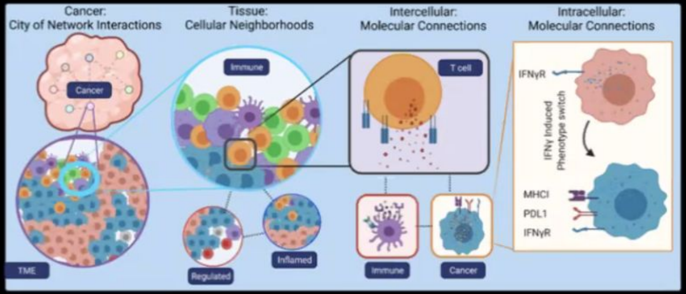
HICKEY LAB

(Hickey, *Cell Systems*, 2024)

Duke

Tissues: Multi-scale Network of Interactions



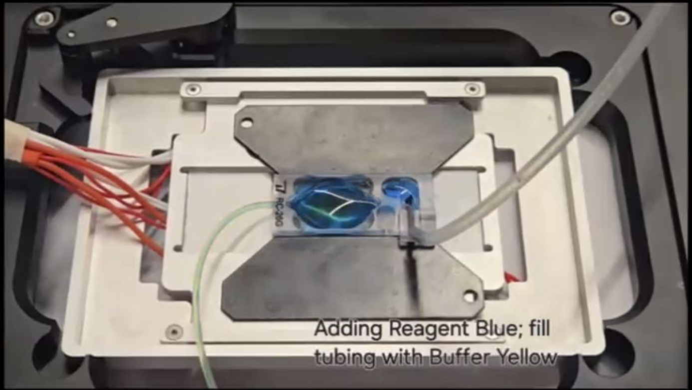


HICKEY LAB

(Hickey, *Cell Systems*, 2024)

Duke

Microscopes & Robotics

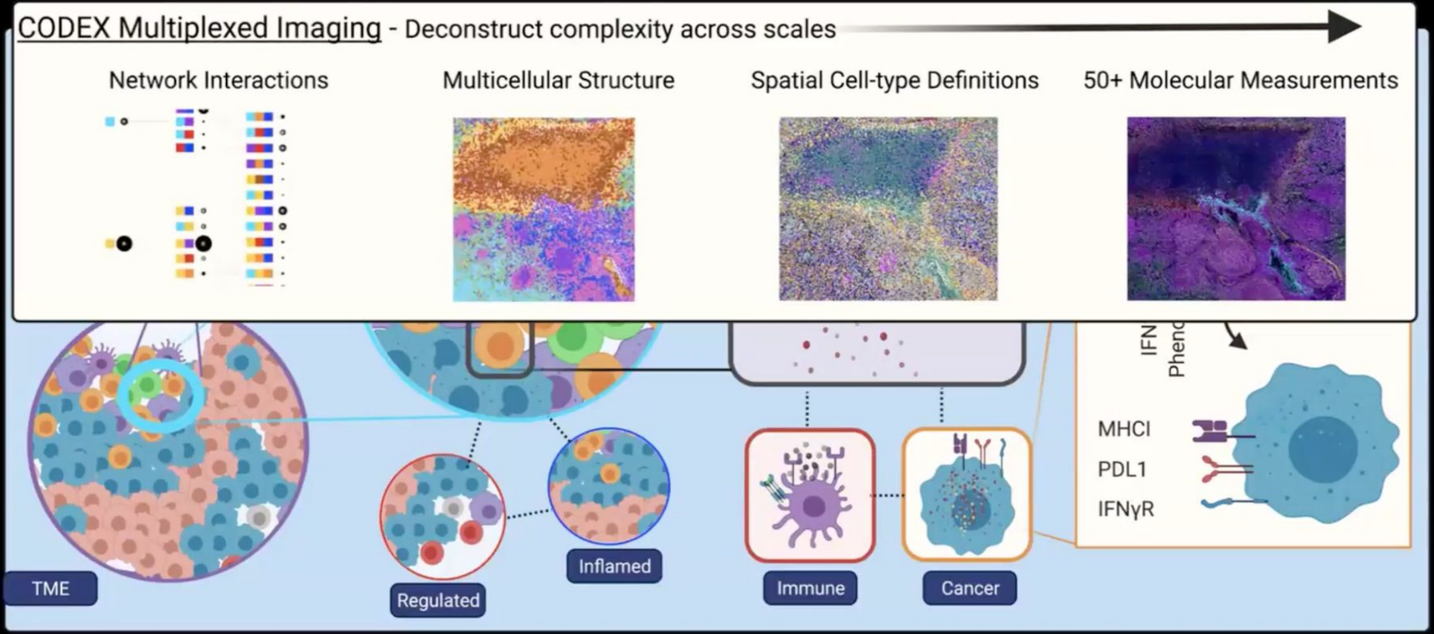


HICKEY LAB

Cephla Inc.

Duke

Deconstructing Complexity Across Scales: Spatial-o



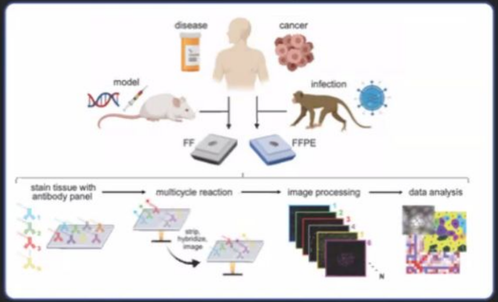
HICKEY LAB

(Hickey, *Cell Systems*, 2024)

Duke

Multiplexed Imaging Standardization & Pipelines

Protocols



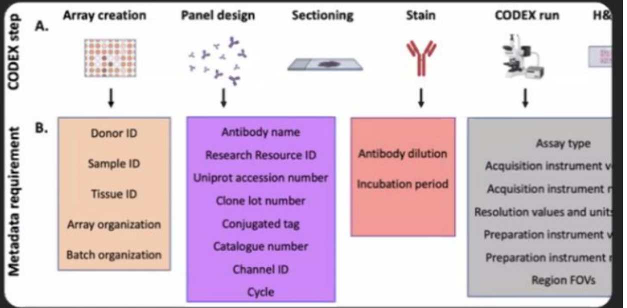
(Black*, Phillips*, Hickey*, *Nature Protocols*, 2021)

Data Processing & Storage

State	Level of Processing
0 (Raw Data)	Raw Data
1 (Processed)	Stitching, tiling, thresholding, background subtra deconvolution, alignment, and extended depth of

(Hickey*, Neumann*, Radtke* *Nature Methods*, 2021)

Data Metadata Standardization



(Caraccio, Hickey, *Elsevier*, in press)

Antibody Panel Development

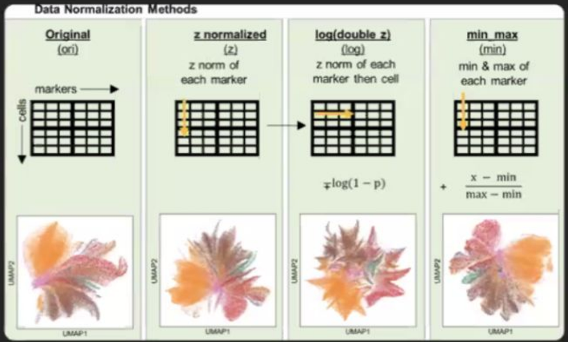
Organ Mapping Antibody Panel (OMAP) for Multiplexed Antibody-Based Imaging of Human Intestine with CODEX

Author Nam John Hickey
Author ORCI 0000-0001-9961-7673
Reviewer(s): Teri Longacre
Reviewer OR 0000-0002-0935-7300
General Publication(s):
Data DOI: https://doi.org/10.48539/HBM373.HQCB.363
Date: 6-May-22
Version Num v1.0

uniprot_acce	HGNC_ID	target_name	antibody_nai	host_organis	clonality	vendor	catalog_num	lot_number	recomb
P46013	HGNC:7107	Ki67	Anti-Ki67 ant	mouse	B56	BD Bioscienc	556003	NA	No
P01589	HGNC:6008	CD25	Anti-CD25 ar	mouse	7G7B6	Bio X Cell	BE0014	NA	No
P08195	HGNC:11026	CD98	Anti-CD98 ar	mouse	MEM-108	Biolegend	315602	NA	No
P08670	HGNC:12692	Vimentin	Anti-Vimenti	mouse	RV202	BD Bioscienc	550513	NA	No
Q9P2W7	HGNC:921	CD57	Anti-CD57 ar	mouse	HCD57	Biolegend	322302	NA	No
P26718	HGNC:18788	NKG2D (CD3	Anti-NKG2D	mouse	1D11	Biolegend	320802	NA	No
P09564	HGNC:1695	CD7	Anti-CD7 ant	mouse	CD7-6B7	Biolegend	343102	NA	No
P01730	HGNC:1678	CD4	Anti-CD4 ant	rat	A161A1	Biolegend	357402	NA	No
P15391	HGNC:1633	CD19	Anti-CD19 ar	mouse	HIB19	BD Bioscienc	555410	NA	No
P04233	HGNC:1697	HLA-DR	Anti HLA-DR	mouse	L243	Biolegend	307651	NA	No
P20702	HGNC:6152	CD11c	Anti-CD11c a	mouse	B-ly6	BD Bioscienc	555391	NA	No
Q12918	HGNC:6373	CD161	Anti-CD161 a	mouse	HP-3G10	Biolegend	339902	NA	No
P08247	HGNC:11506	Synaptophys	Anti-Synaptc	mouse	7H12	Novus Bio	NBP1-47483	NA	No
P28906	HGNC:1662	CD34	Anti-CD34 ar	mouse	561	Biolegend	343602	NA	No
P08637	HGNC:3619	CD16	Anti-CD16 ar	mouse	3G8	BD Bioscienc	555404	NA	No
P22083	HGNC:4015	CD15	Anti-CD15 ar	mouse	HI98	BD Bioscienc	555400	NA	No
Q12884	HGNC:3590	FAP	Anti-FAP ant	rat	2.2B10	Thermo Fish	13-0300	NA	No

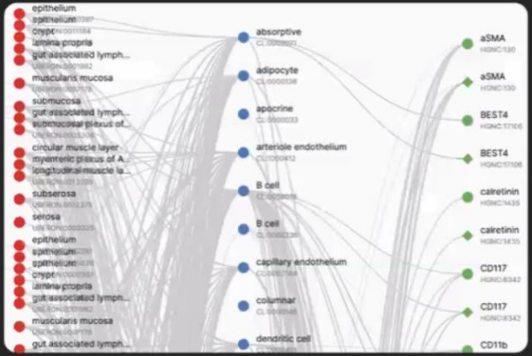
(Quordokus, *Nature Methods*, 2023)

Data Normalization & Cellular Analysis



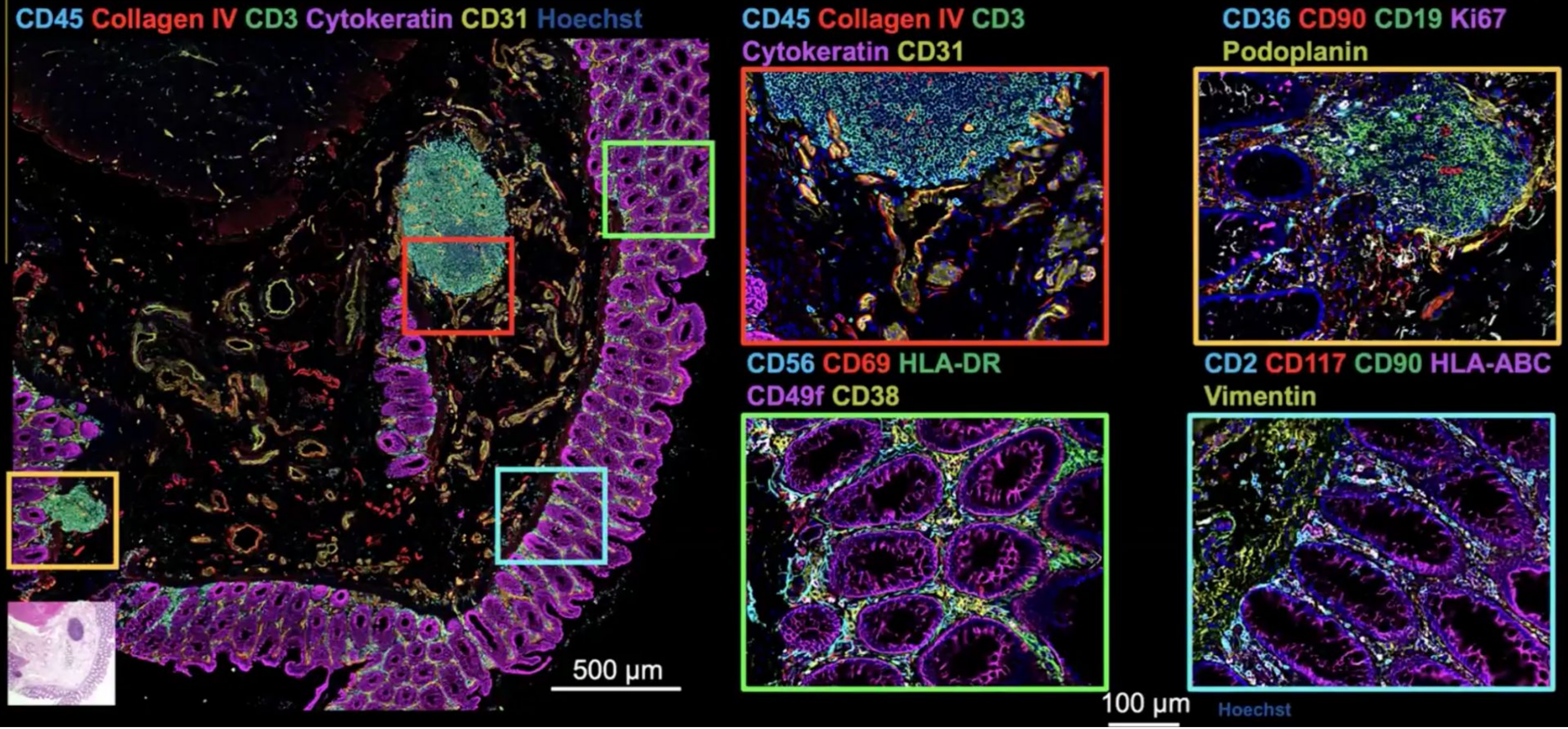
(Hickey, *Frontiers Immunology*, 2021)

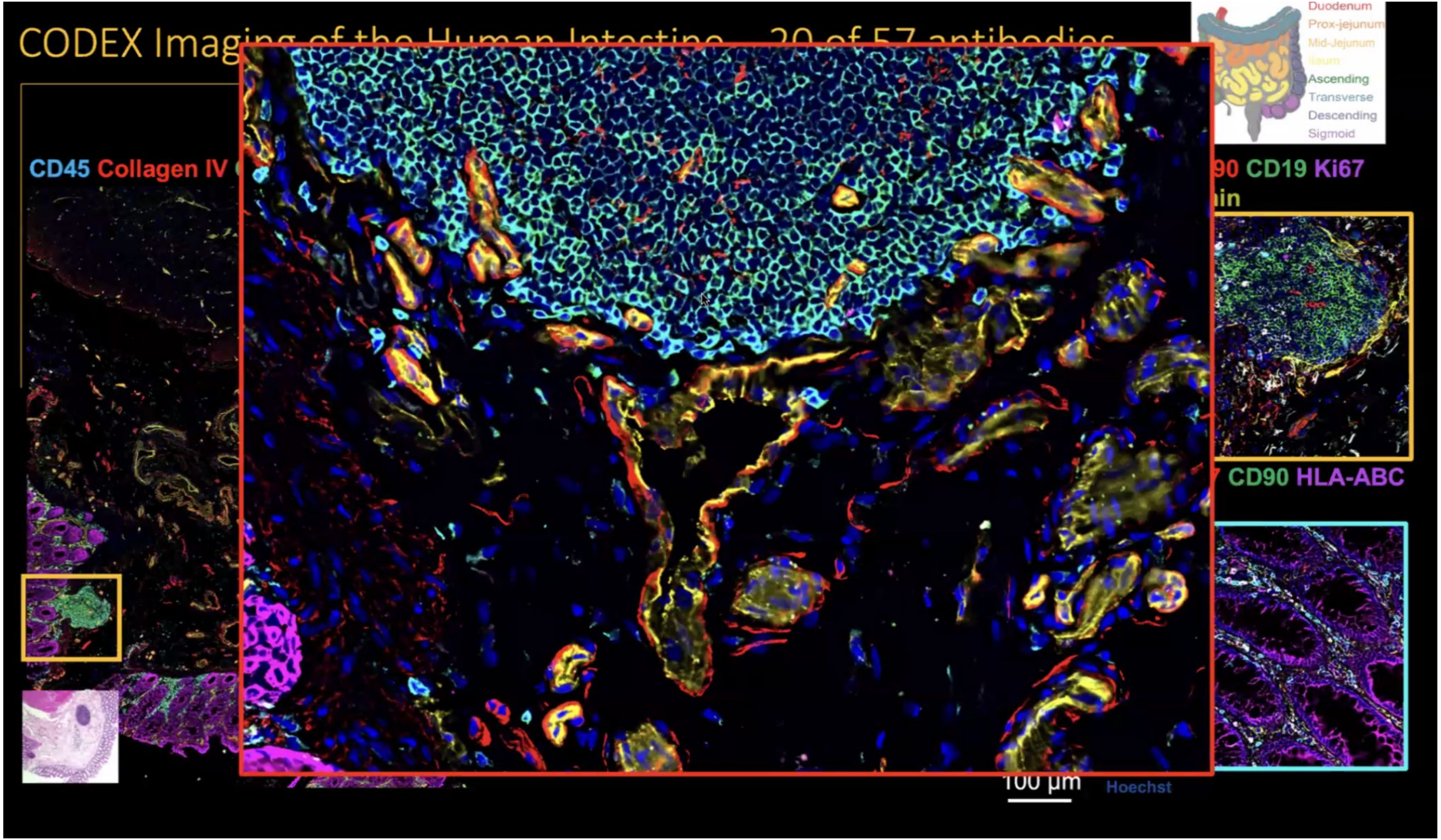
Marker, Cell, Tissue Unit Ontology



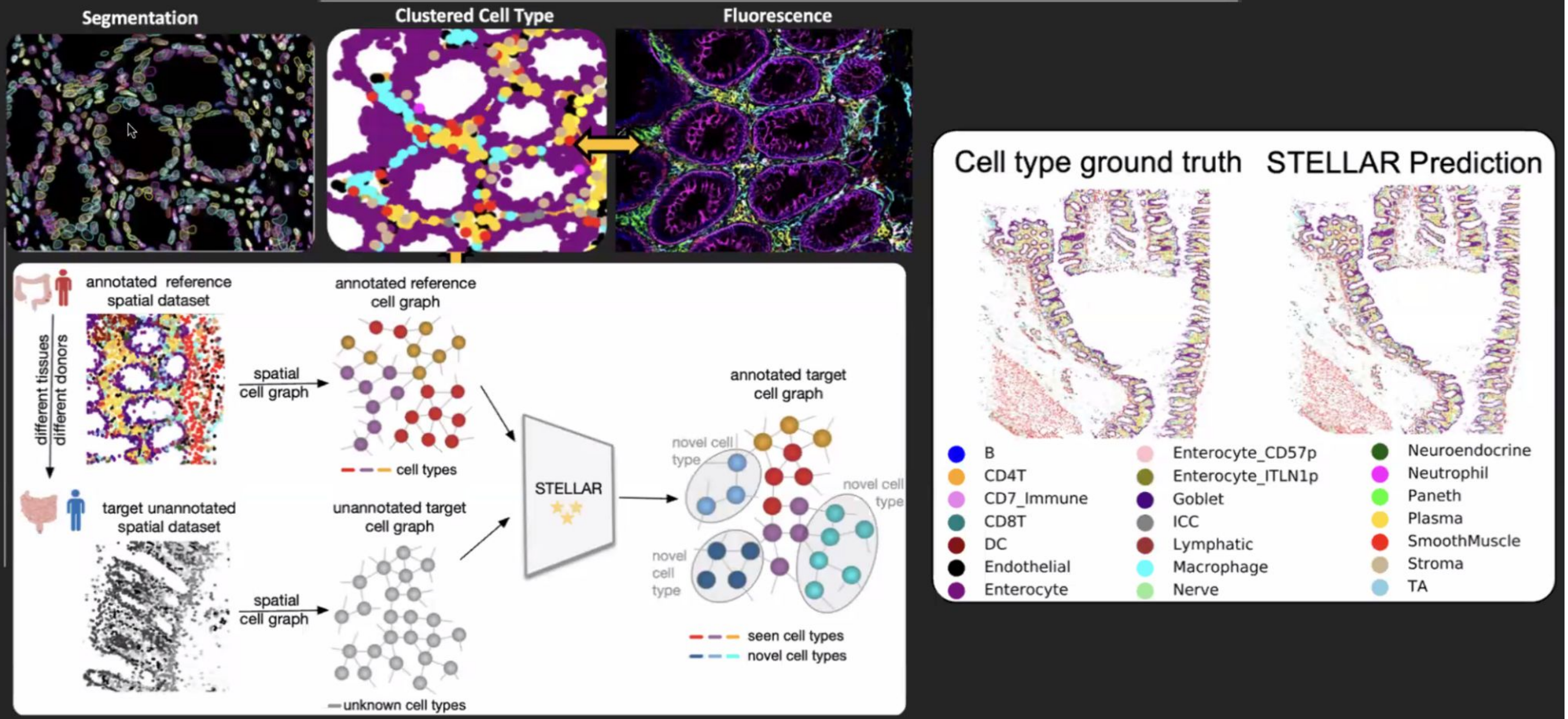
(Borner, *Nature Cell Biology*, 2021)

CODEX Imaging of the Human Intestine – 20 of 57 antibodies



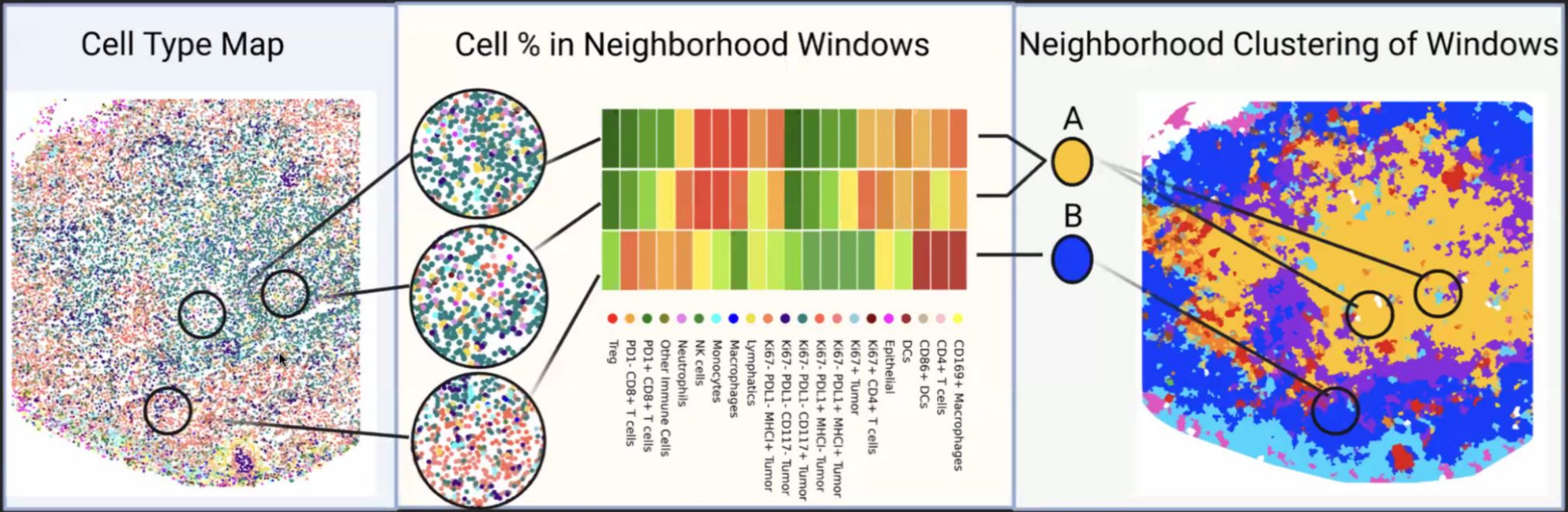

CODEX Imag
Duodenum
Prox-jejunum
Mid-Jejunum
Ascending
Transverse
Descending
Sigmoid
CD45 Collagen IV
90 CD19 Ki67
in
CD90 HLA-ABC
100 µm
Hoechst

AI to help with computing and analyzing this big data



(Brbic*, Cao*, Hickey*, *Nature Methods*, 2022)

Multi-cellular Neighborhood Identification

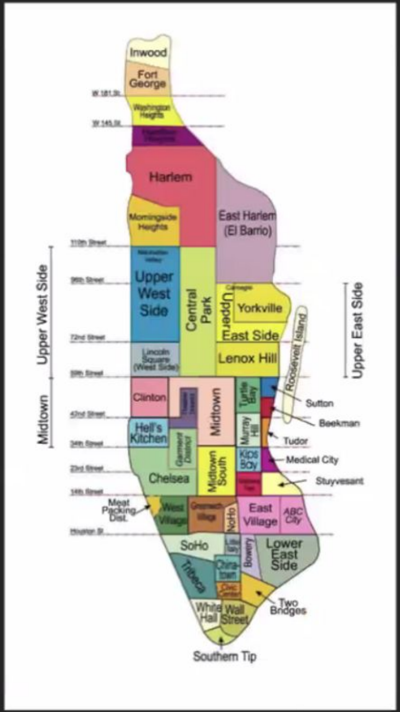


Neighborhood Analysis Reveals Conserved Multicellular Structures of Intestine

(Hickey*, Nevins*, Becker*, *Nature*, 2023)

How Can We Analyze Hierarchical Spatial Domains?

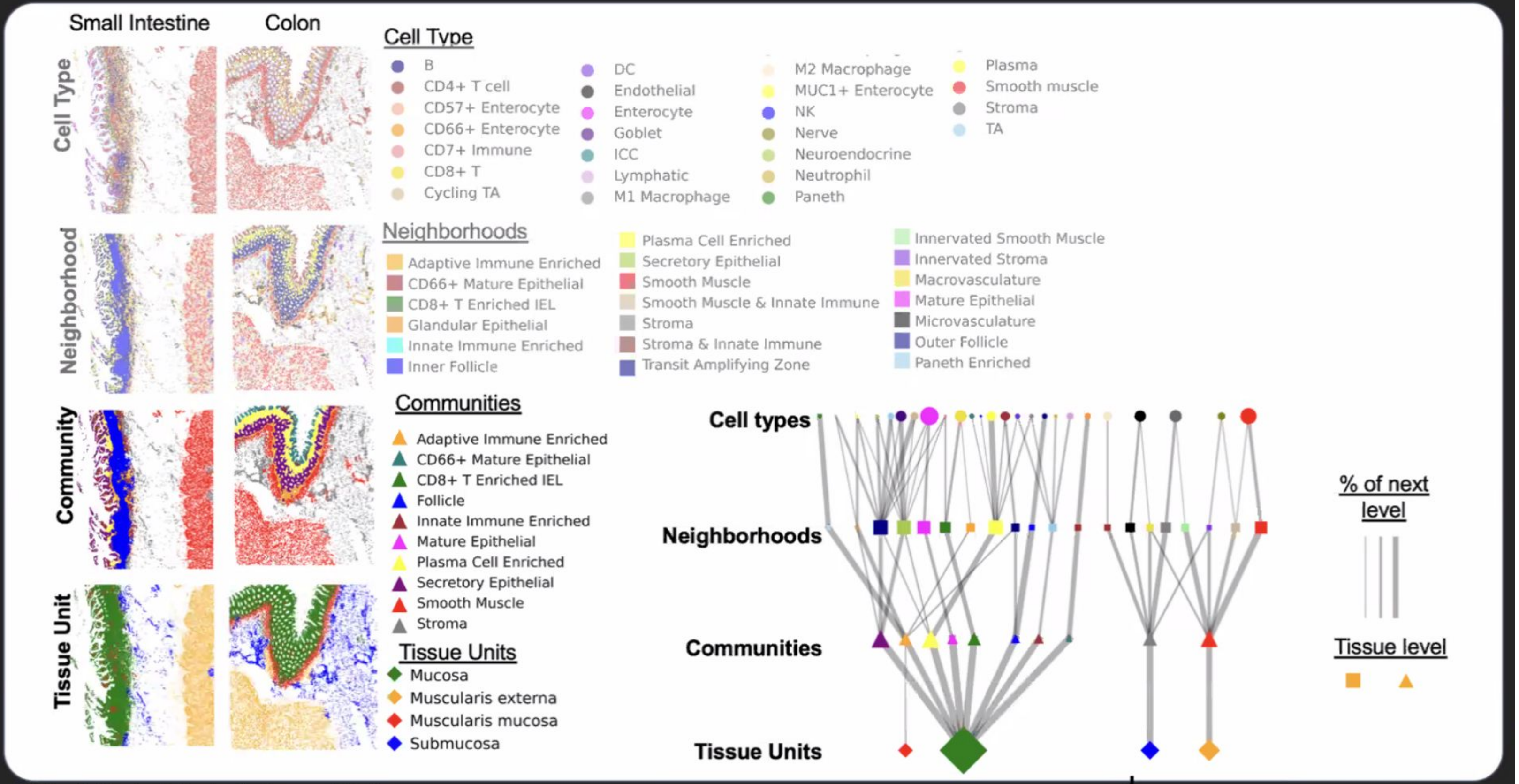
Neighborhood - County



Counties - State

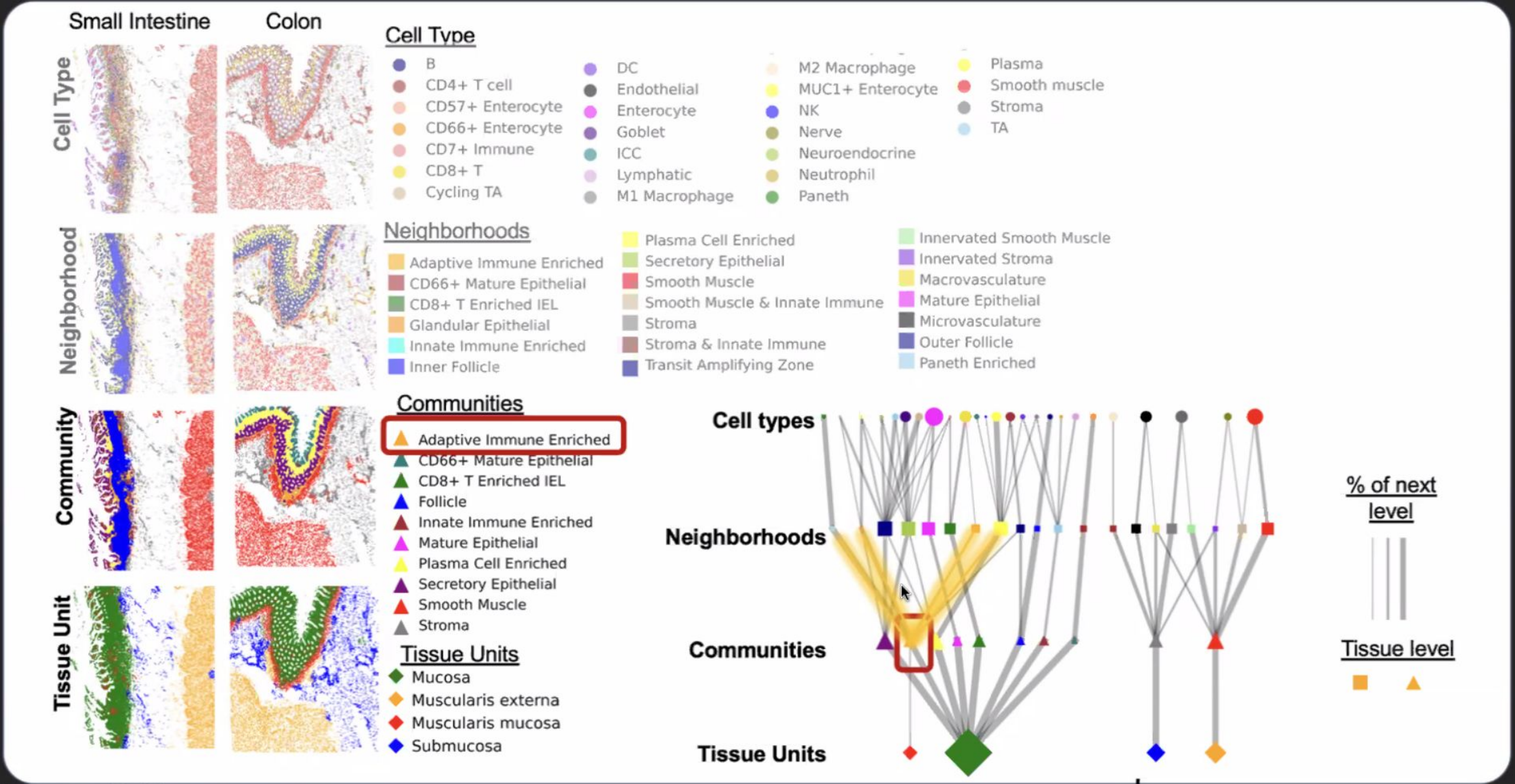
States - Country

Overall structure of intestine by multi-level analysis of functional units



(Hickey*, Nevins*, Becker*, *Nature*, 2023)

Overall structure of intestine by multi-level analysis of functional units

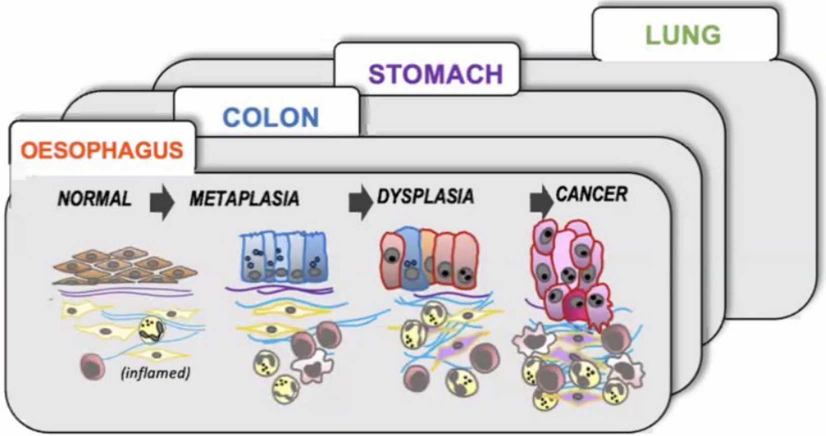


(Hickey*, Nolan*, Becker*, *Nature*, 2023)

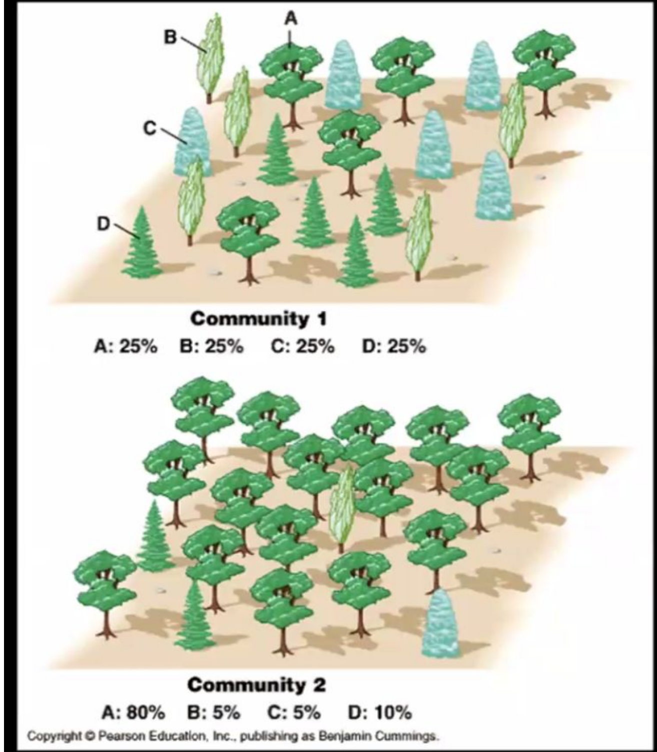
Chronic Inflammation-Associated Cancers (CIACs)

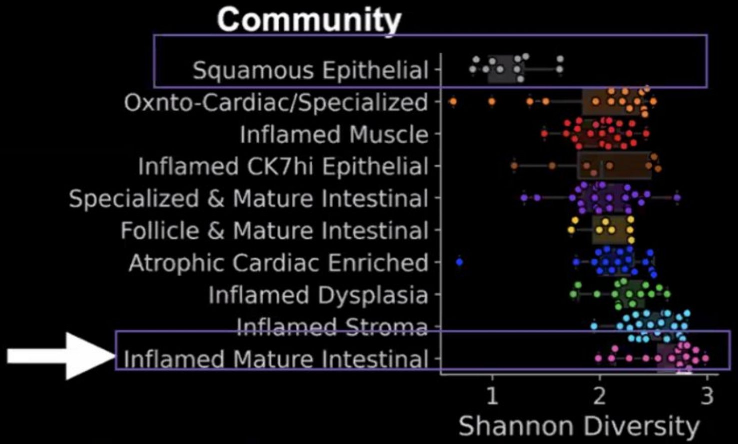
A Shared Pattern of Tissue Disruption

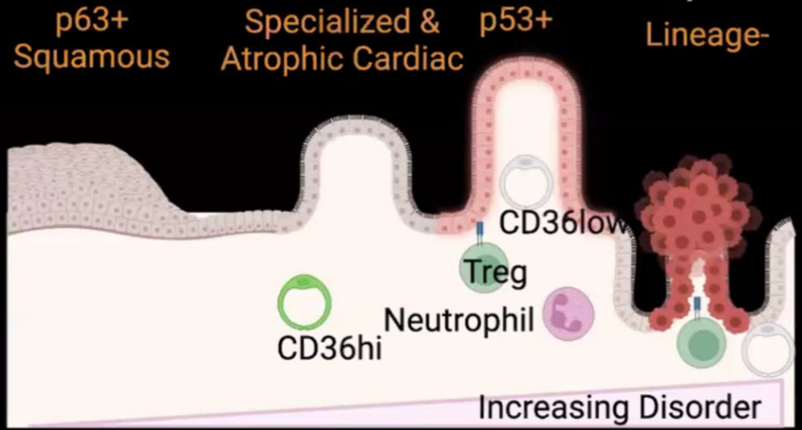
Slides – Thea Tlsty
Team:
STORMING CANCER



Communities of Neighborhoods Highlight Lack of Organization in Tumor



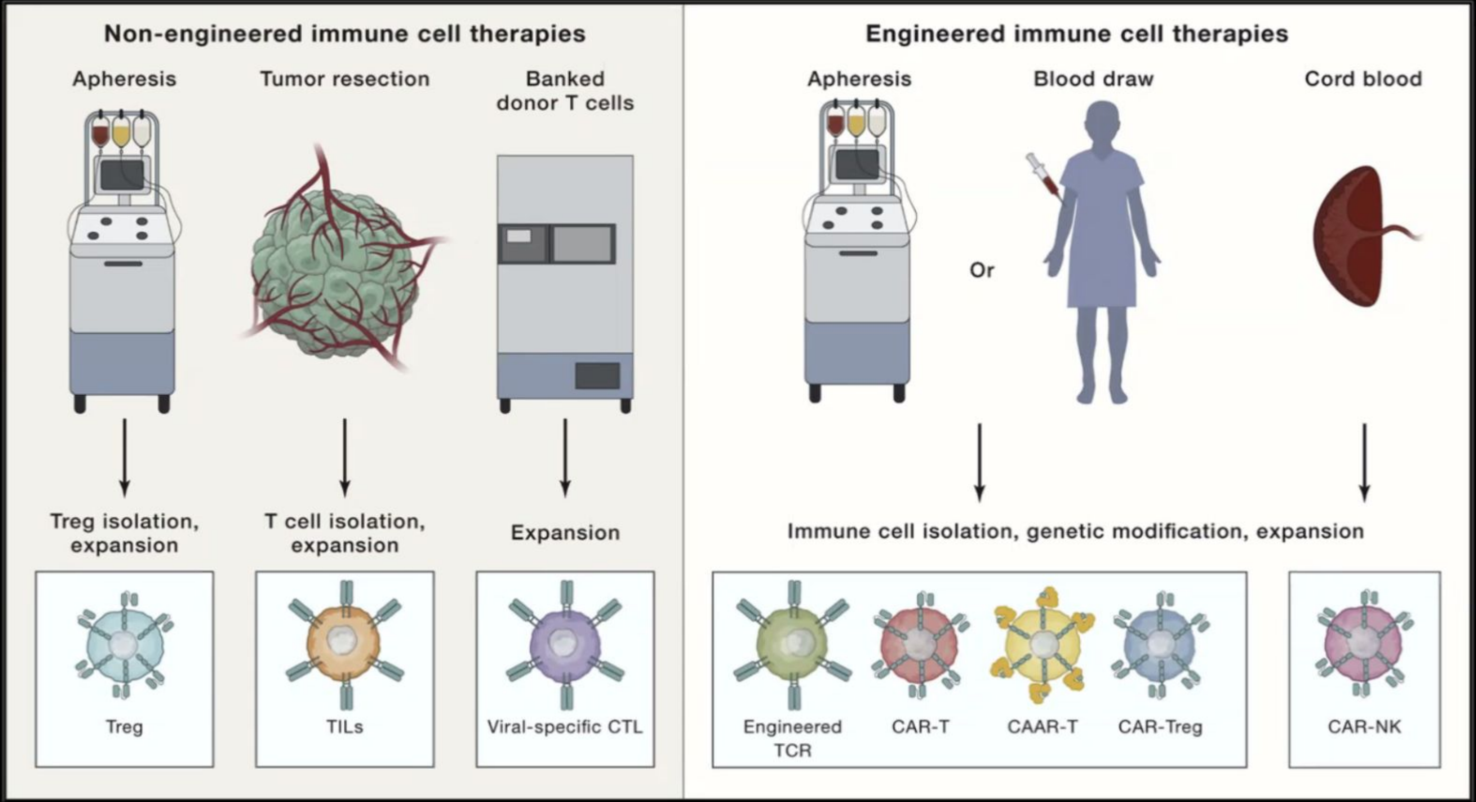




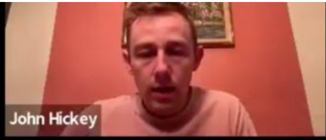
HICKEY LAB

Duke

Immune Cell Therapies for Cancer

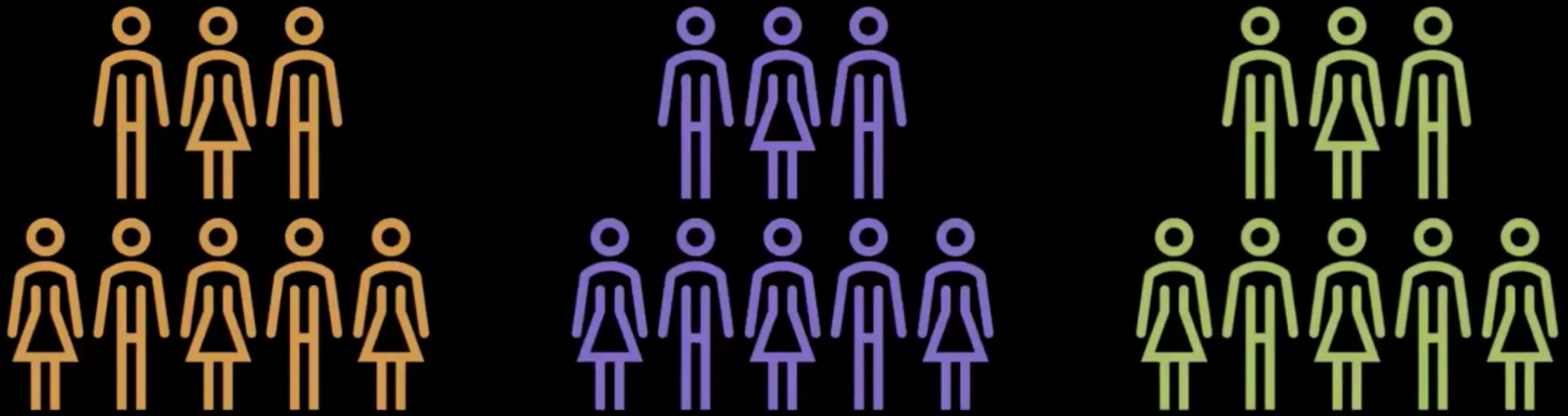


(Weber, *Cell*, 2020)

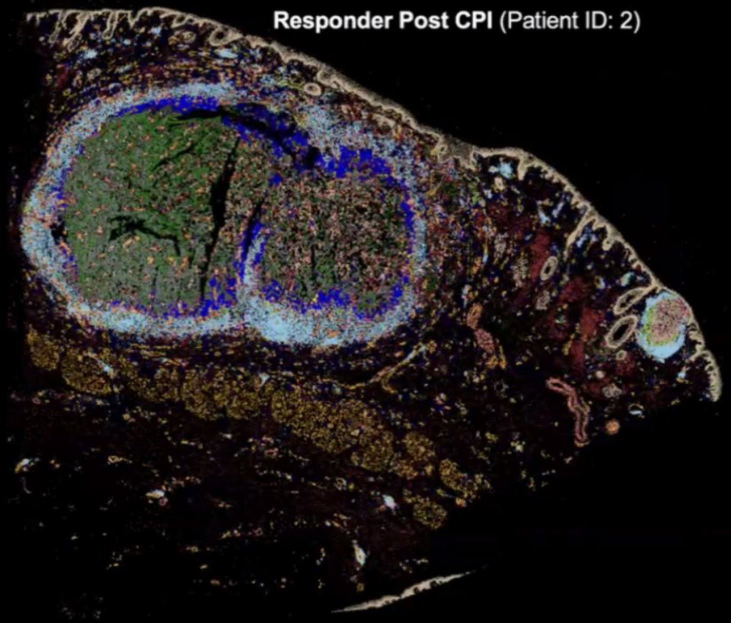


Teams of people working together

Engineering | Sales | R&D



Cellular Neighborhood Organization in Human Tumors



Human Cellular Neighborhood (hCN)

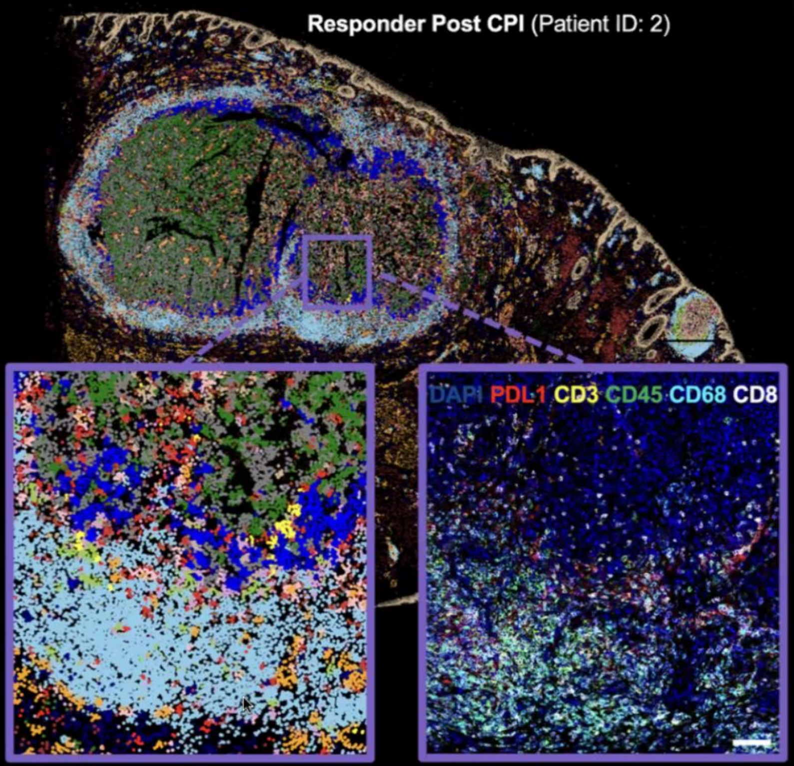
- DC Enriched Immune
- Epithelial/Skin Appendages
- Follicle
- Immune Infiltrate
- Inflamed Tumor
- Macrophage Enriched Immune
- Neutrophil Enriched
- PDPN+ Stromal Enriched
- Perivascular
- Productive T cell & Tumor
- Proliferating Tumor
- Resting Tumor
- Stromal Enriched
- Tumor & Immune
- Vasculature
- Vasculature & Immune

HICKEY LAB

(Hickey, *Cell Reports*, 2023)

Duke

Cellular Neighborhood Organization in Human Tumors



Human Cellular Neighborhood Map

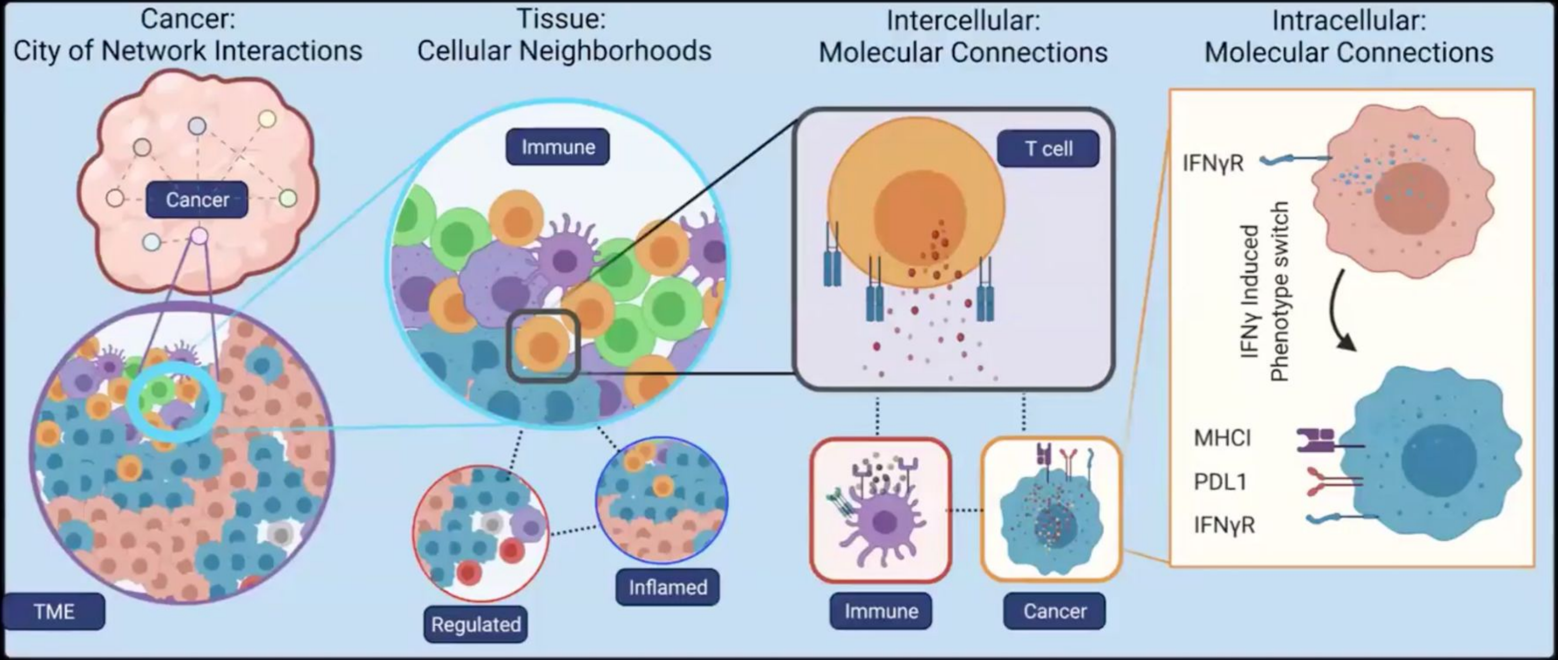
Overlay of CODEX imaging

Human Cellular Neighborhood (hCN)

- DC Enriched Immune
- Epithelial/Skin Appendages
- Follicle
- Immune Infiltrate
- Inflamed Tumor
- Macrophage Enriched Immune
- Neutrophil Enriched
- PDPN+ Stromal Enriched
- Perivascular
- Productive T cell & Tumor
- Proliferating Tumor
- Resting Tumor
- Stromal Enriched
- Tumor & Immune
- Vasculature
- Vasculature & Immune

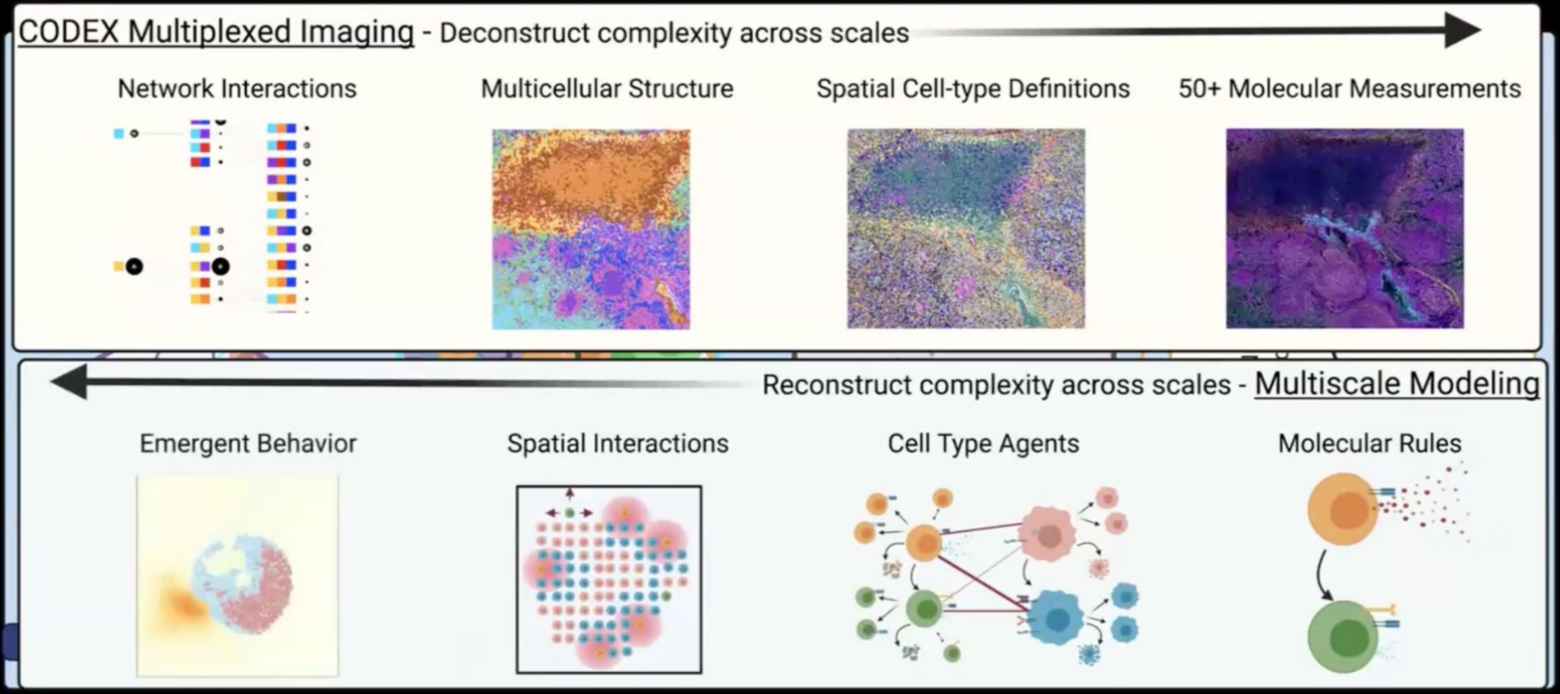
(Hickey, *Cell Reports*, 2023)

Recreating Complexity Across Scales Compliments Deconstructing



(Hickey, *Cell Systems*, 2024)

Recreating Complexity Across Scales Compliments Deconstructing

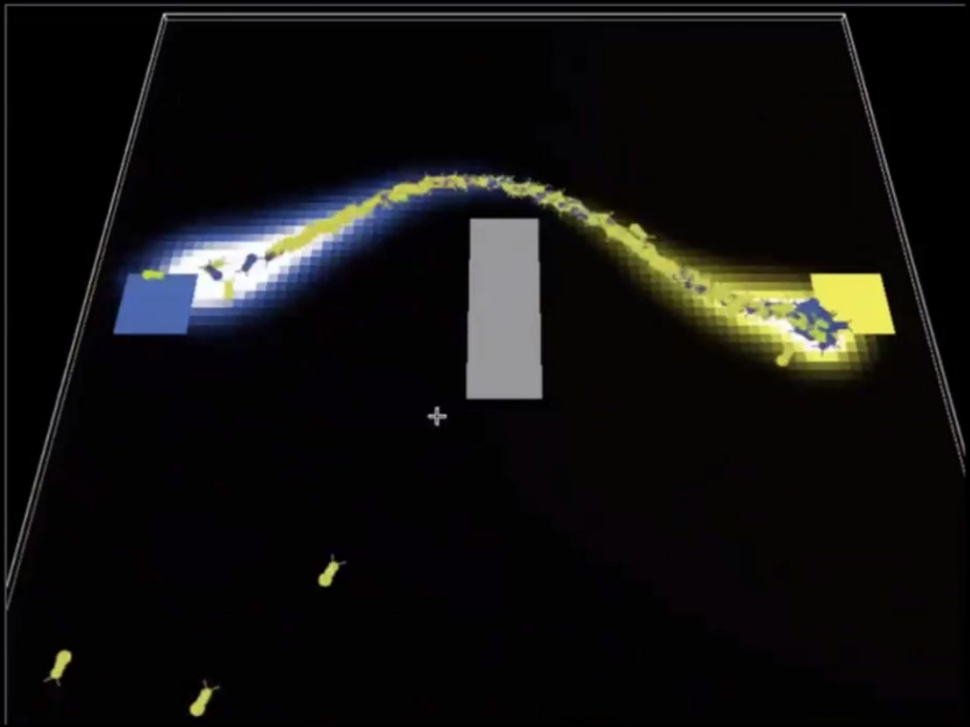


HICKEY LAB

(Hickey, *Cell Systems*, 2024)

Duke

Ant Agent-based Model

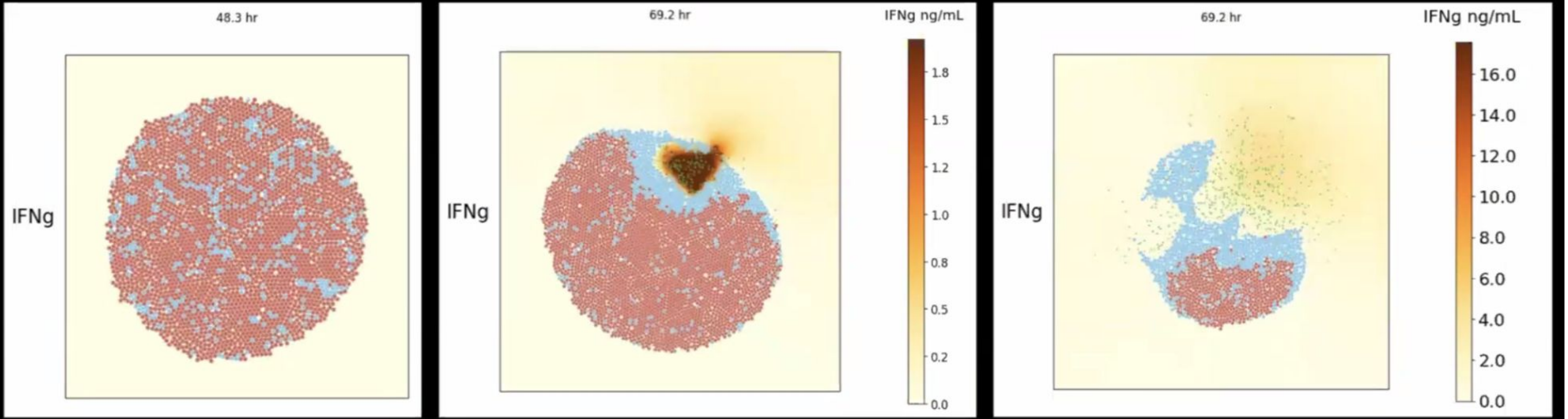


HICKEY LAB

Duke

Multi-scale Agent-based Modeling of the T Cell Immunotherapy

No T cells | T cells | Transformed T cells



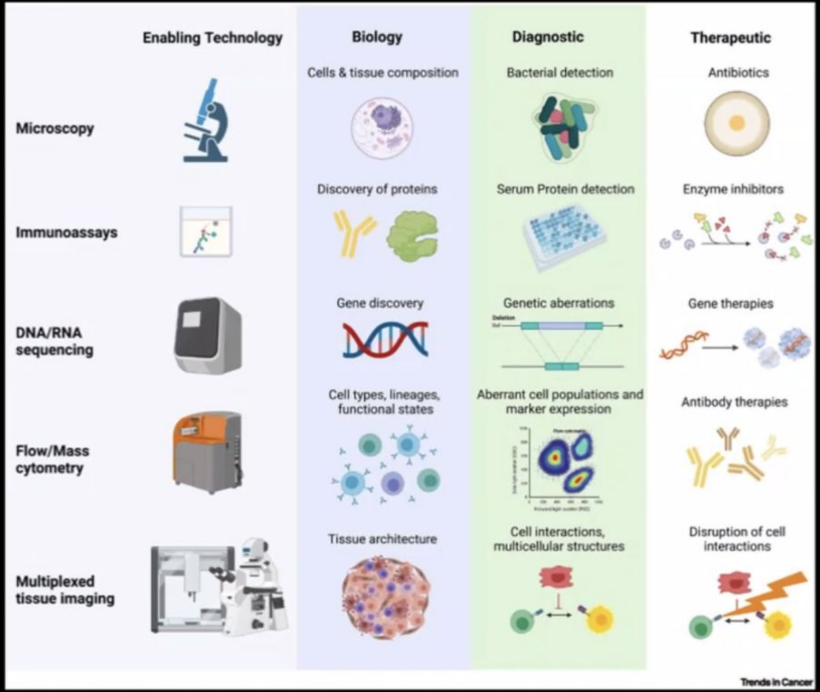
PDL1+ tumor PDL1- tumor PD1+ T cell PD1- T cell

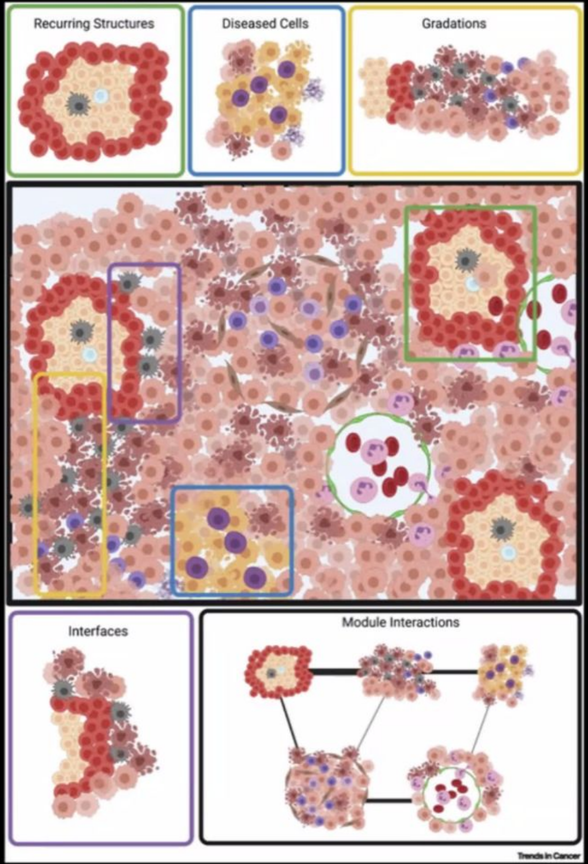
HICKEY LAB

(Hickey, *Cell Systems*, 2024)

Duke

A rethinking of therapies





(Baertsch, Nolan, Hickey, *Trends Cancer*, 2022)

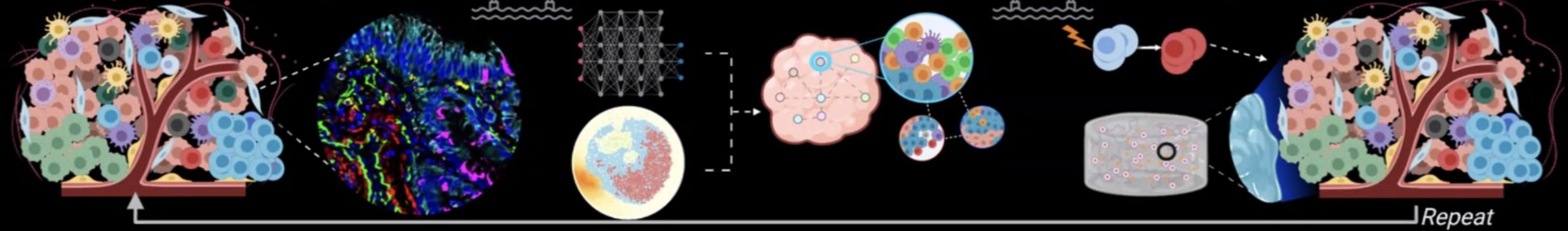
John Hickey – john.hickey@duke.edu
Assistant Professor Biomedical Engineering

Hickey Lab: Synergy and Bridging Omics, Computation, and Engineering

Multiscale Tissue Structure: Single Cell Spatial-Omics Technologies

Multiscale Computational Approaches for Mechanism

Tissue Structure Control: Cell Reprogramming and Biomaterials



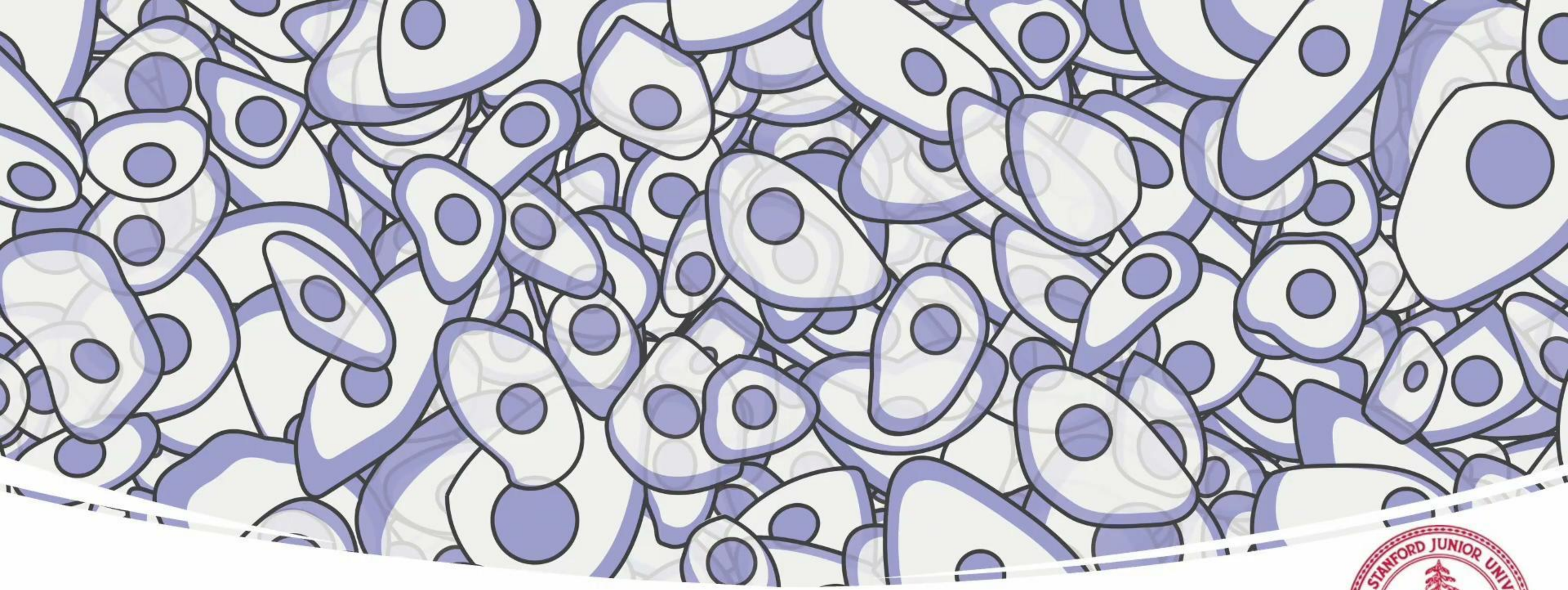
Our expertise.

Engineering · Bioinformatics
Immunology · Systems Biology

Tools we use.

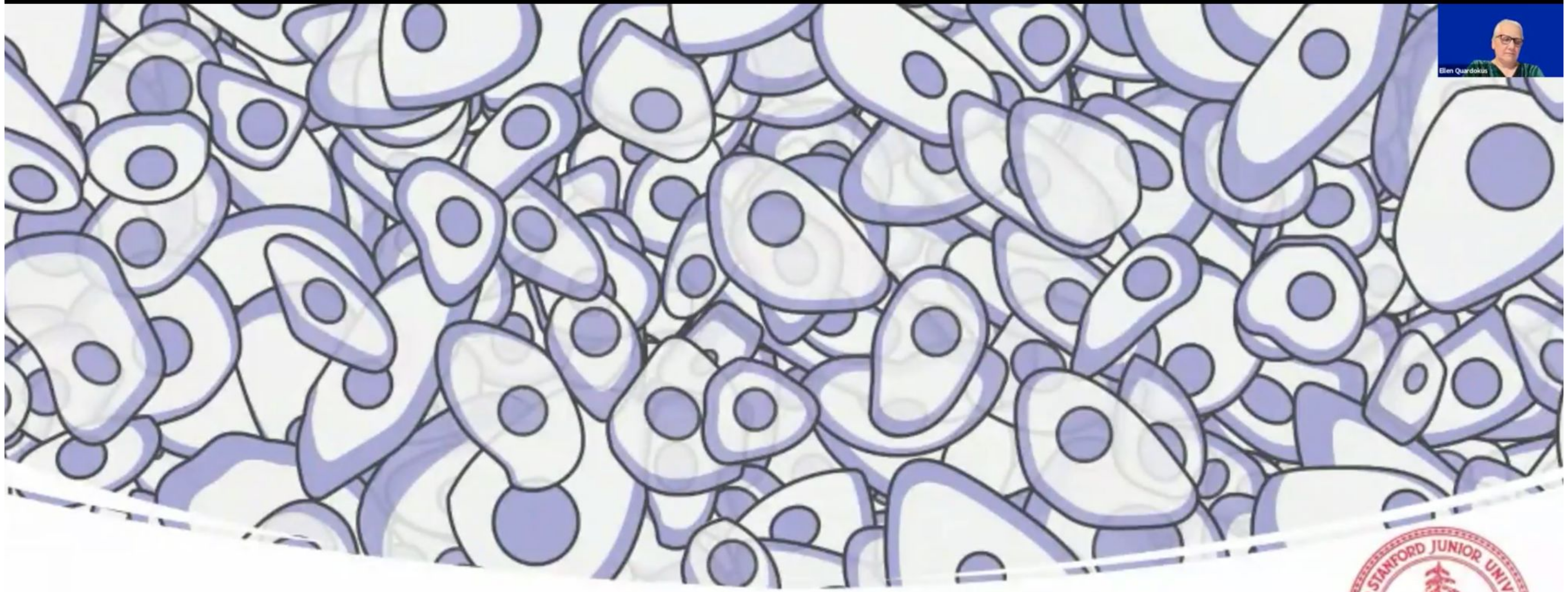
Spatial Omics · Biomaterials
Machine Learning · Modeling

Jure Leskovec, *Stanford University*



Building Foundation Models for the AI Virtual Cell

Jure Leskovec
Stanford University



Building Foundation Models for the AI Virtual Cell

Jure Leskovec
Stanford University

Cell is a Fundamental Unit of Life

- Cells are essential for health and disease
- Advances in AI and omics offer new opportunities to rethink traditional models

Cell
Leading Edge

CellPress
OPEN ACCESS

Perspective

How to build the virtual cell with artificial intelligence: Priorities and opportunities

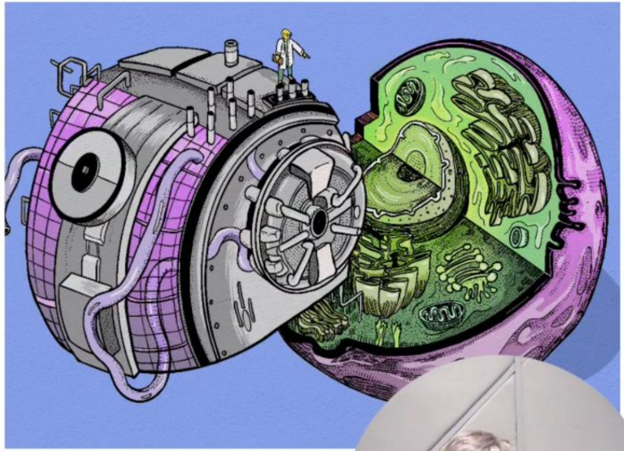
Simulating Biology with the AI Virtual Cell

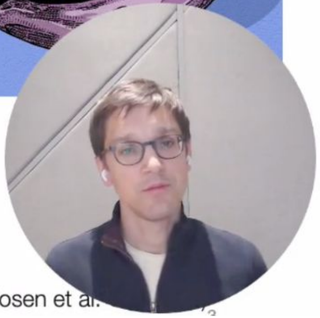
How can we simulate biology?

- Multi scale, non-linear, stochastic/noisy, measured in different ways, incomplete data

We create an AI Virtual Cell:

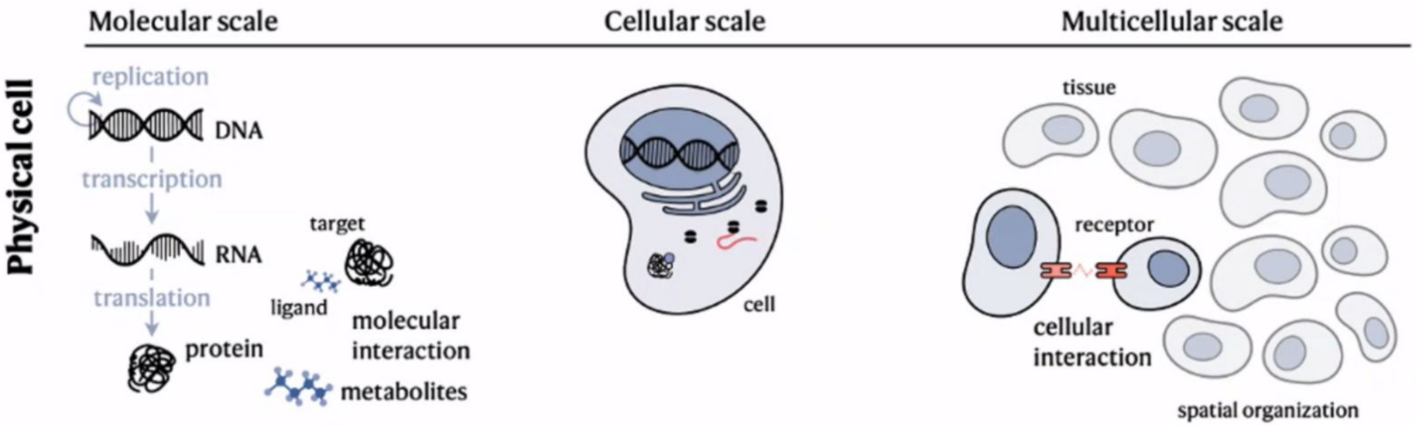
A connected framework of AI models that simulate increasingly complex and dynamic biological systems.



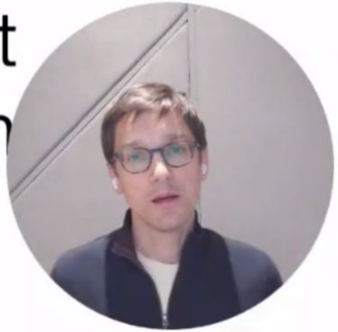


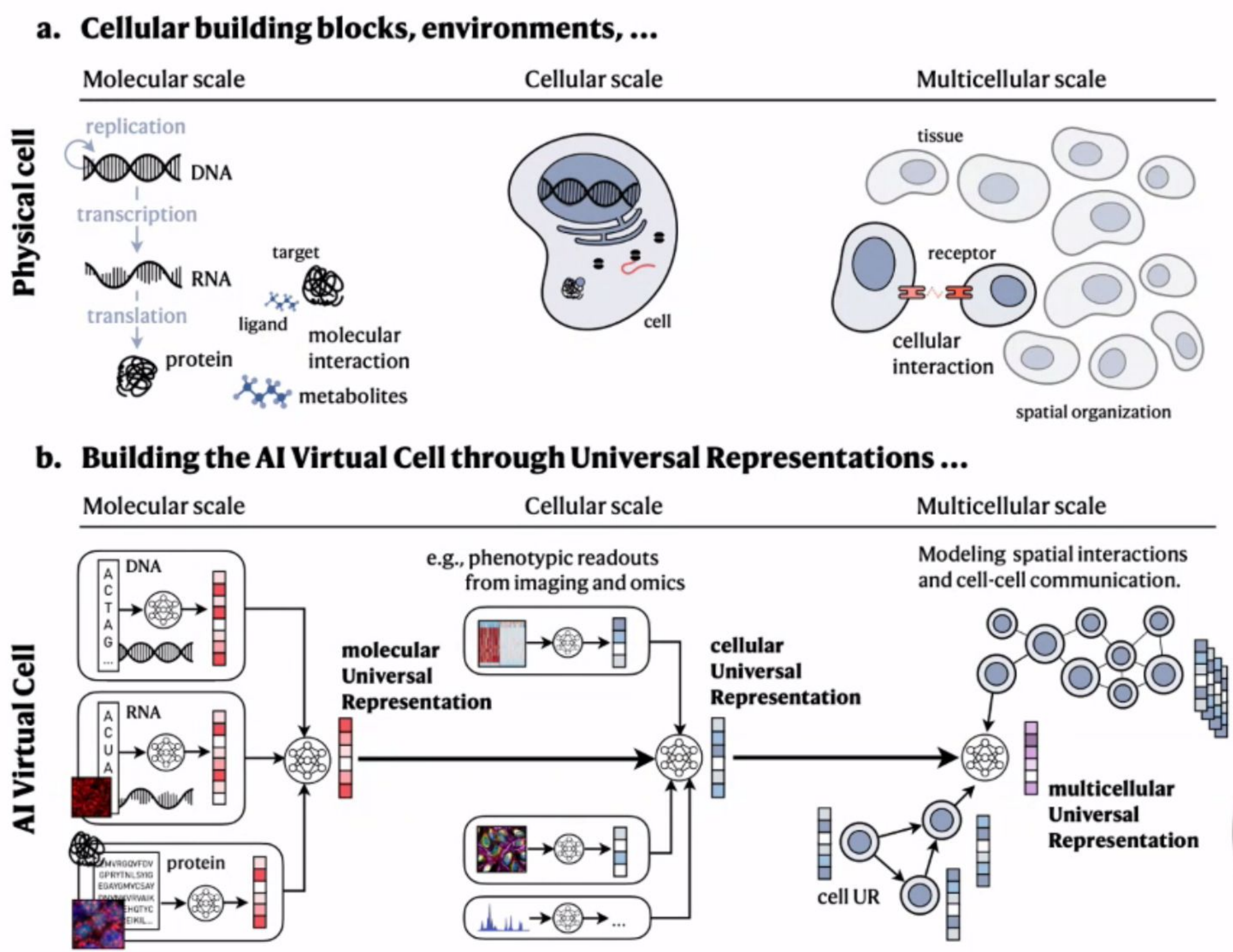
How to Build the Virtual Cell with Artificial Intelligence: Priorities and Opportunities (Bunne, Roohani, Rosen et al.

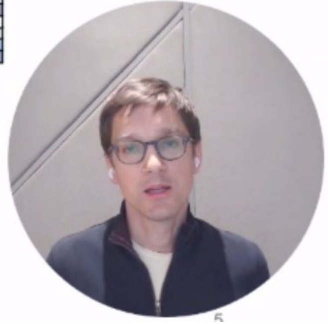
Connecting Biology's Physical Scales



A connected framework of AI models that simulate increasingly complex and dynam
biological systems.

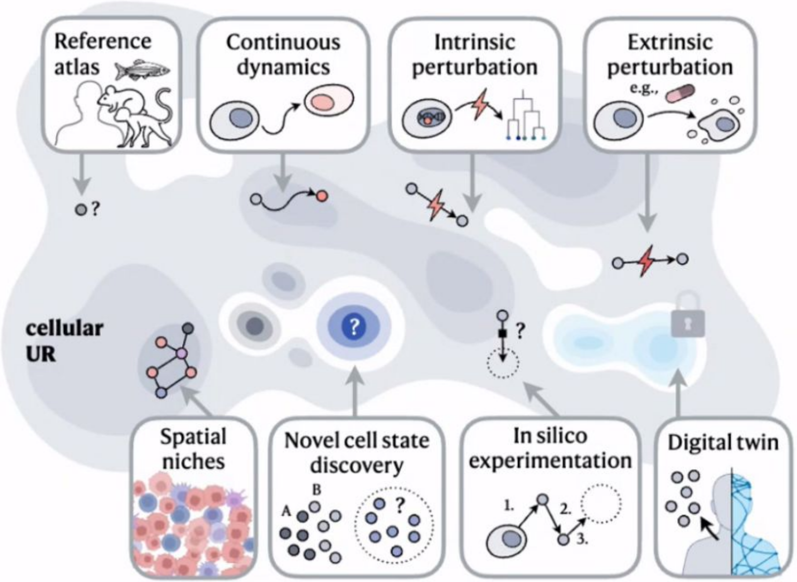


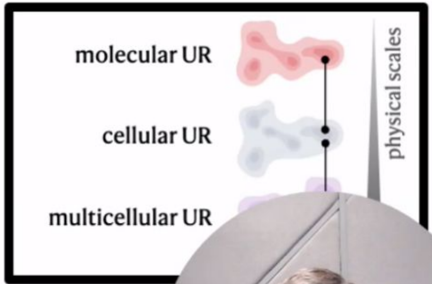

a. Cellular building blocks, environments, ...
Molecular scale
Cellular scale
Multicellular scale
Physical cell
replication
DNA
transcription
RNA
translation
protein
target
ligand
molecular
interaction
metabolites
cell
tissue
receptor
cellular
interaction
spatial organization
b. Building the AI Virtual Cell through Universal Representations ...
Molecular scale
Cellular scale
Multicellular scale
AI Virtual Cell
DNA
RNA
protein
molecular
Universal
Representation
e.g., phenotypic readouts
from imaging and omics
cellular
Universal
Representation
Modeling spatial interactions
and cell-cell communication.
cell UR
multicellular
Universal
Representation

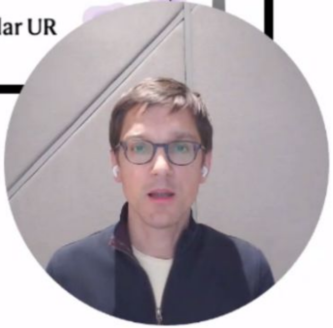


The Power of Representation Learning

Learning universal representation spaces unlocks fundamental capabilities for biomedicine.







Virtual Instruments

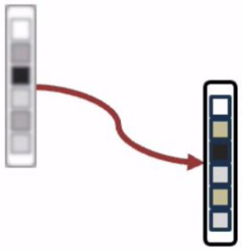
1. **Manipulate:**
   Embedding -> Embedding

   What happens to a cell after a drug is applied?

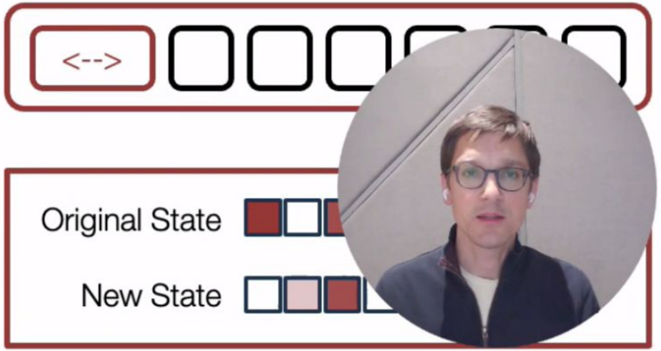
2. **Decode:**
   Embedding -> Readout

   What is the 3d structure of a protein?

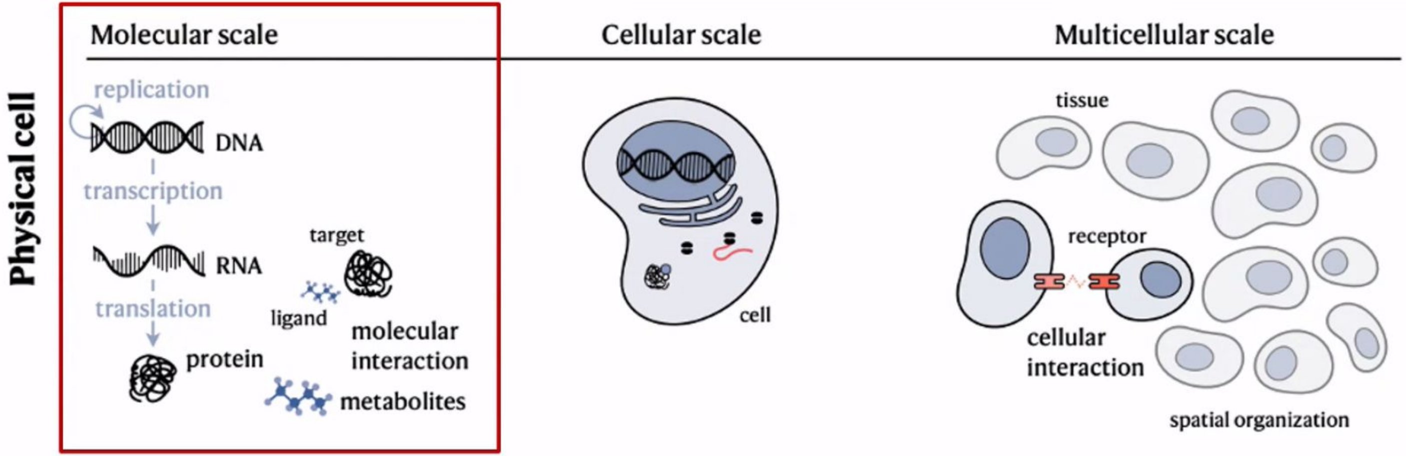
Diffusion Models

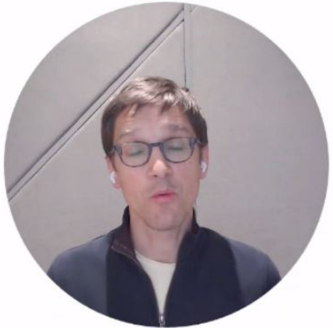


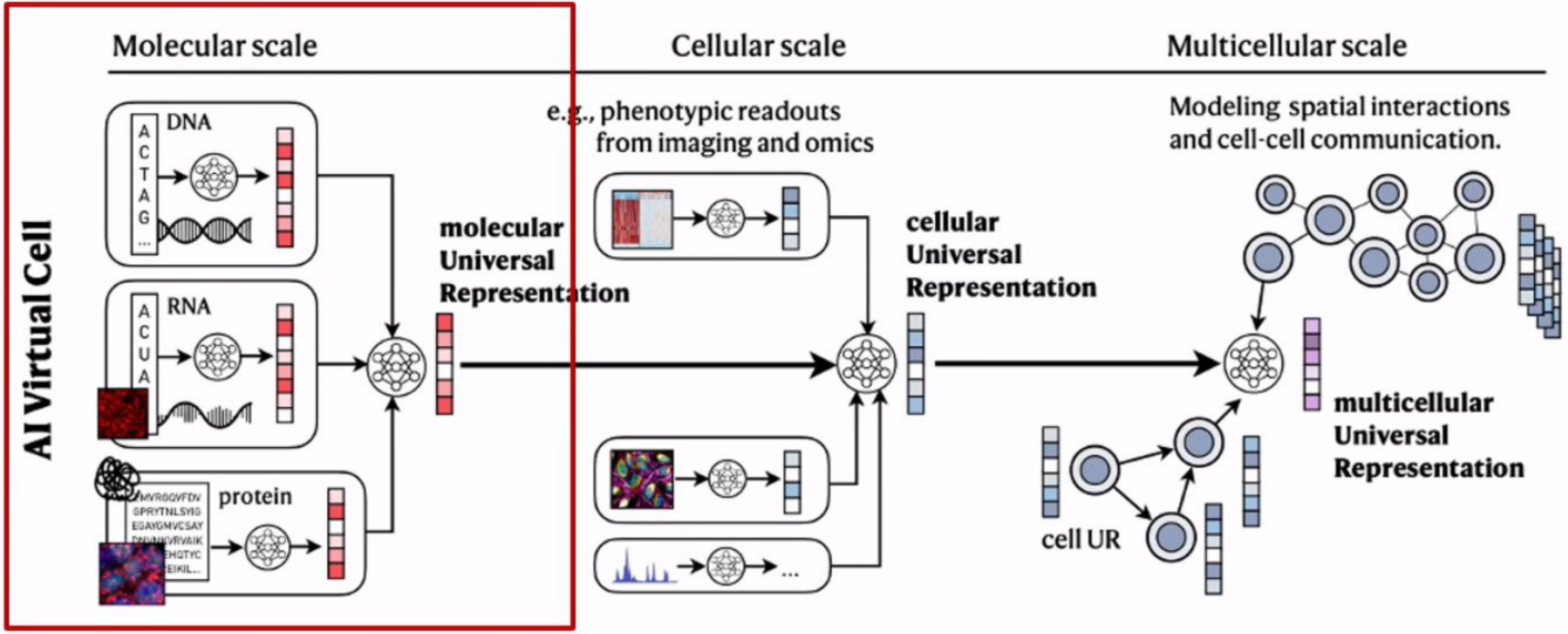
Prompting



This talk: AI Virtual Cell







AI Virtual Cell
Molecular scale
Cellular scale
Multicellular scale
DNA
RNA
protein
molecular Universal Representation
e.g., phenotypic readouts from imaging and omics
cellular Universal Representation
Modeling spatial interactions and cell-cell communication.
multicellular Universal Representation
cell UR

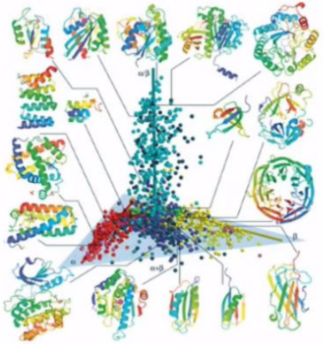
Protein Language Models

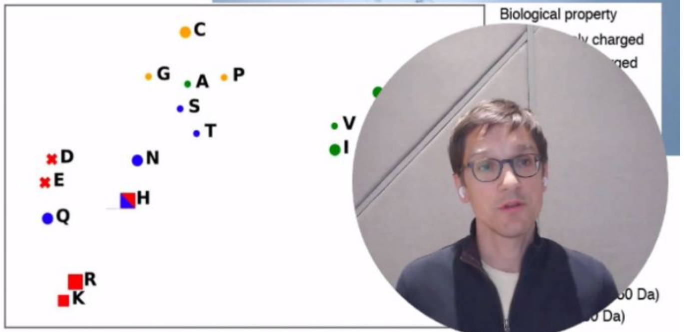
<u>Protein Language Models:</u> ProtT5, ESM encode <u>the whole protein universe</u>

- Motivated by ChatGPT & AlphaFold
- Trained on 250M+ proteins

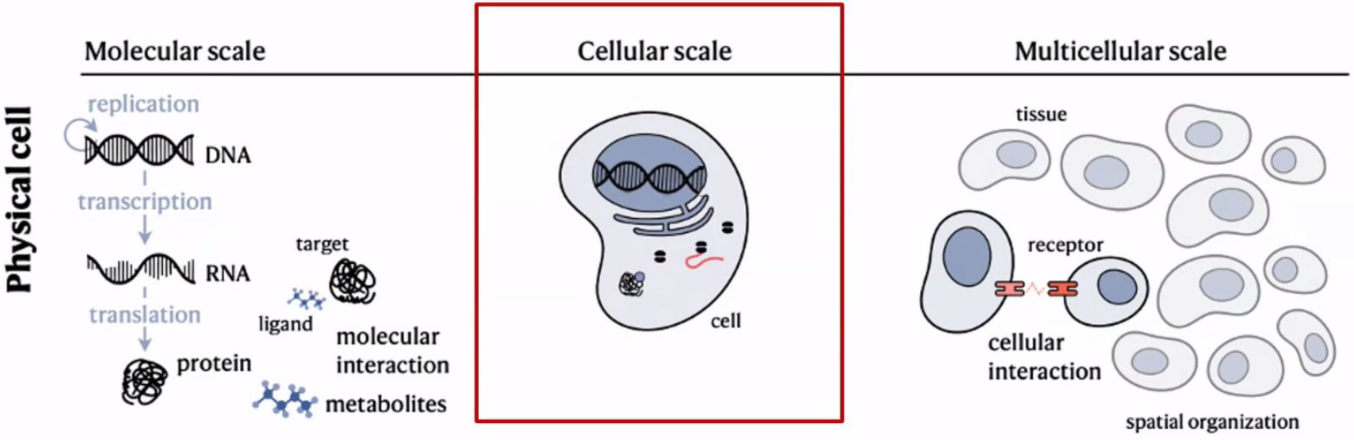
Protein Embeddings encode:

- Structure
- Molecular propt.
- Orthology





Rives et al., *PNAS (2021)*



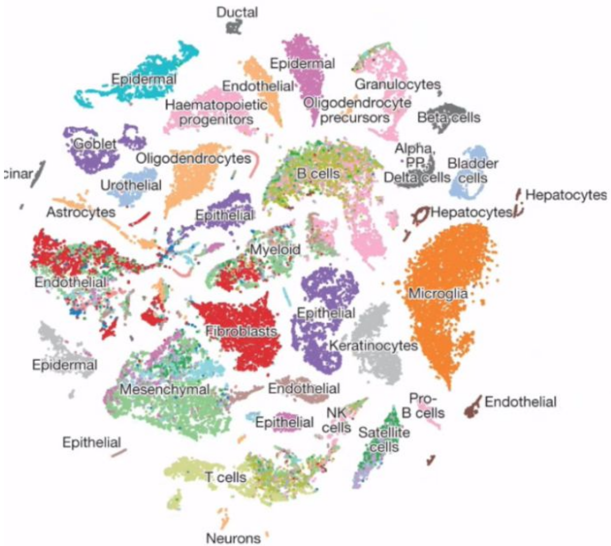
Universal Cell Embeddings

Yanay Rosen*, Yusuf Roohani*, Ayush Agrawal,
Leon Samotorčan,
Stephen Quake, Jure Leskovec

Universal Cell Embeddings: A Foundation Model for Cell Biology (Rosen, Roohani et al. Preprint)

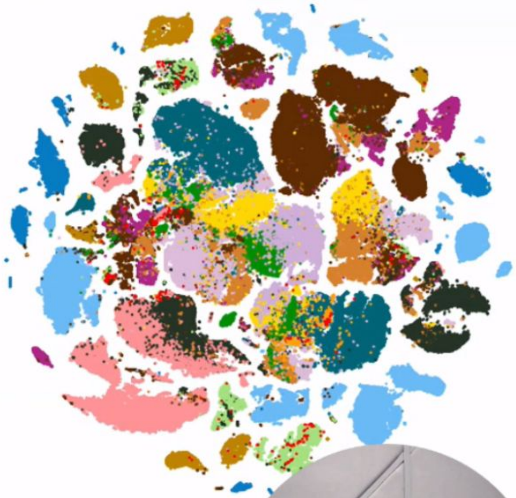
Towards Universal Cell Embeddings: Integrating Single-cell RNA-seq Datasets across Species with SATURN (Rosen, Brbic, Roohani et al. Nature M

Cell Atlas Datasets



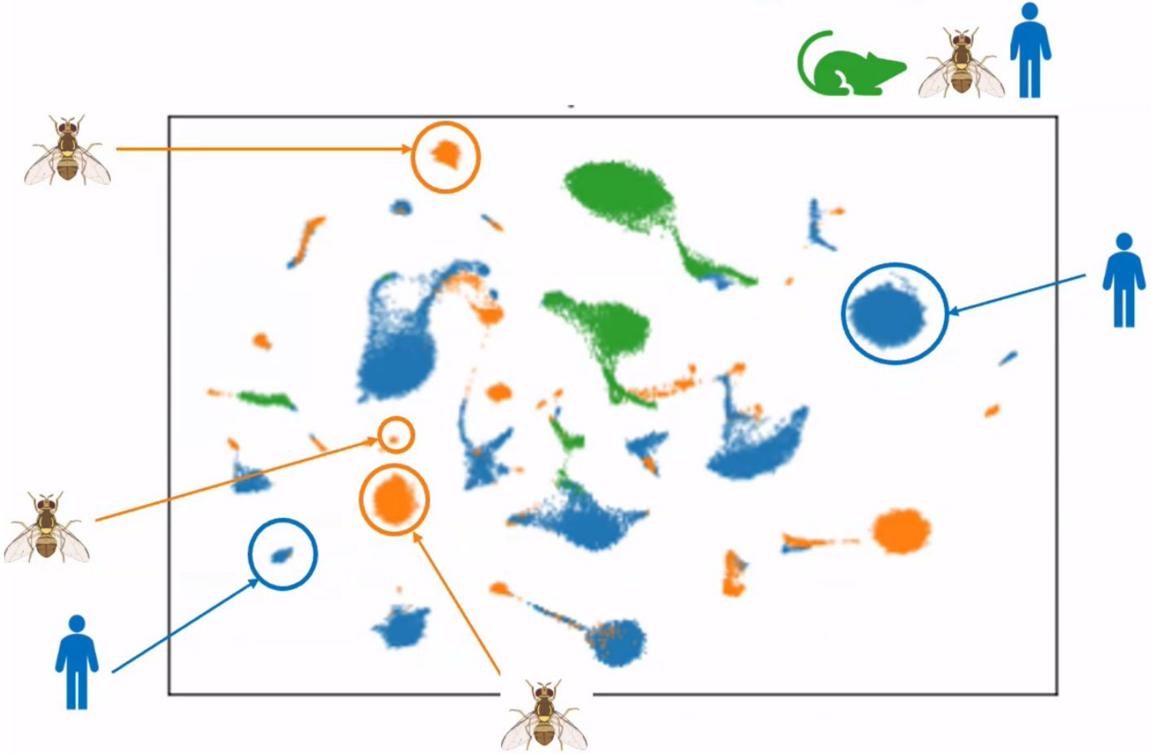
Tabula Muris
(Nature '18, '20)

Tabula Sapiens
bioRxiv '22



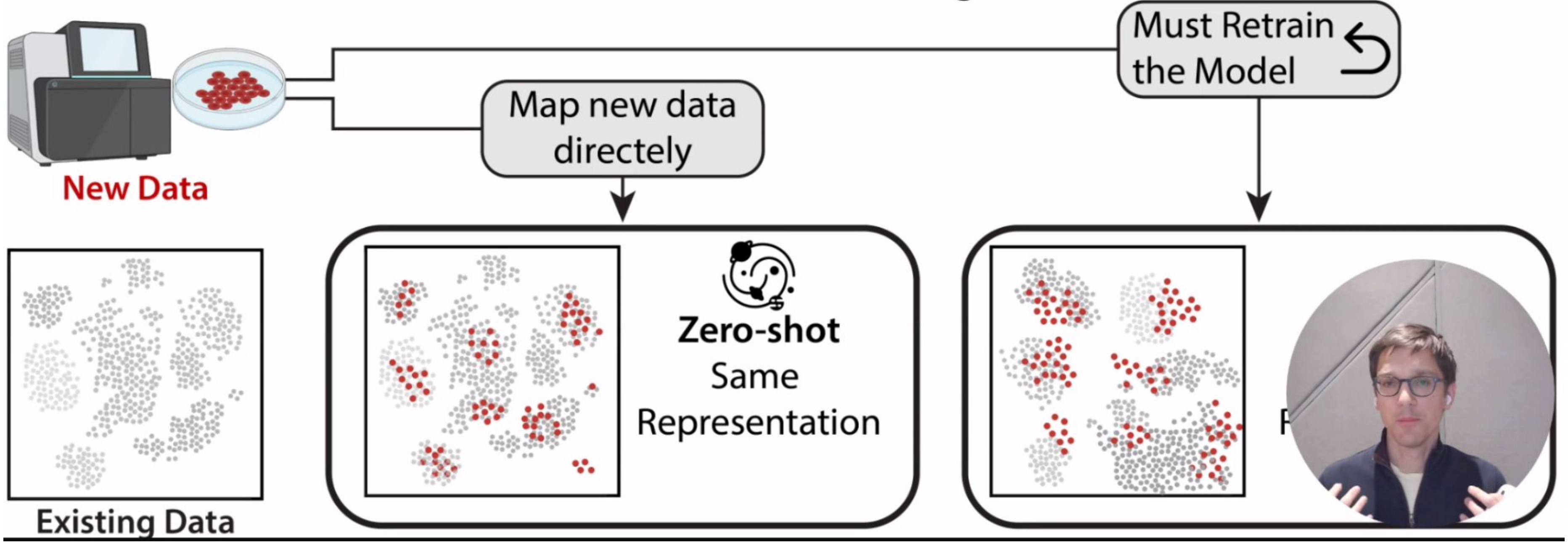
Fly
Sc

Goal: Cross-Species Cell Embedding Space

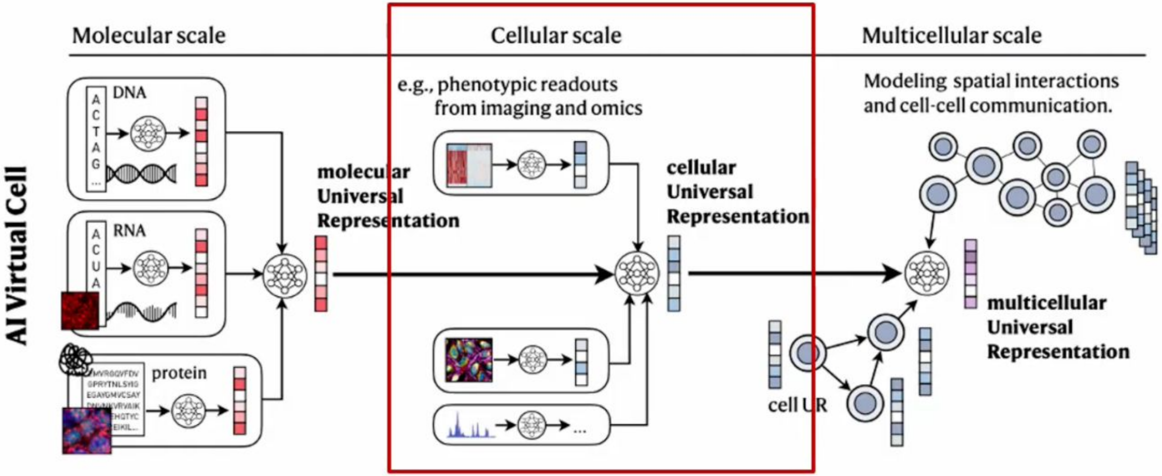


Universal Spaces are *Fixed*

- Representations of data should be consistent and fixed
  - ChatGPT works without finetuning



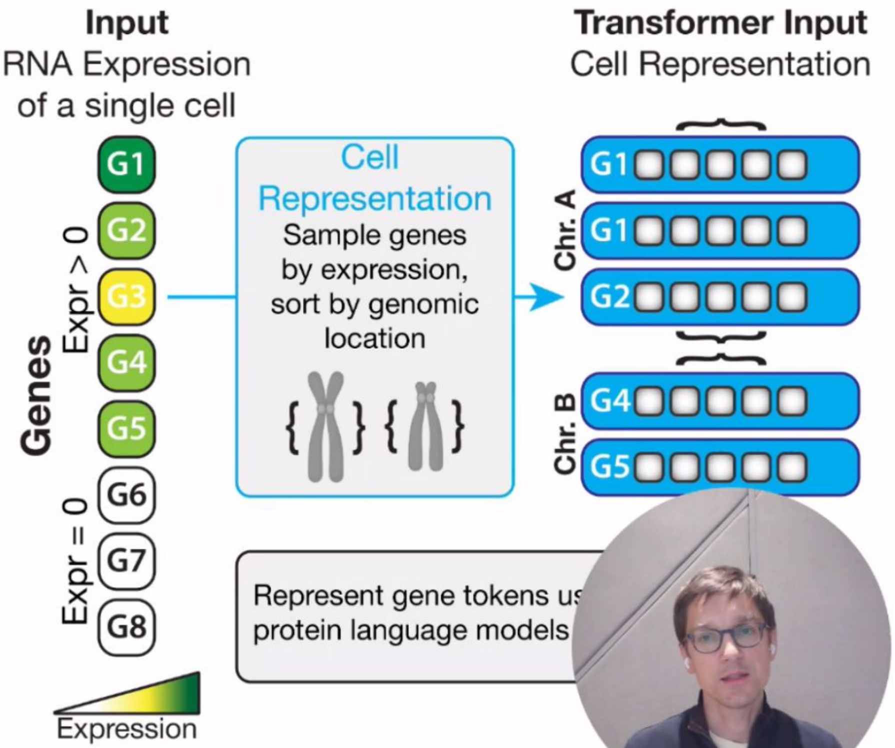
We built a Cellular-Scale Foundation Model



The Universal Cell Embedding (UCE) Model

Key Model Choices

- Gene expression is not natural language
- A universal space is a fixed space
  - Foundation Models are zero-shot
- Self-supervised
  - Organization is emergent



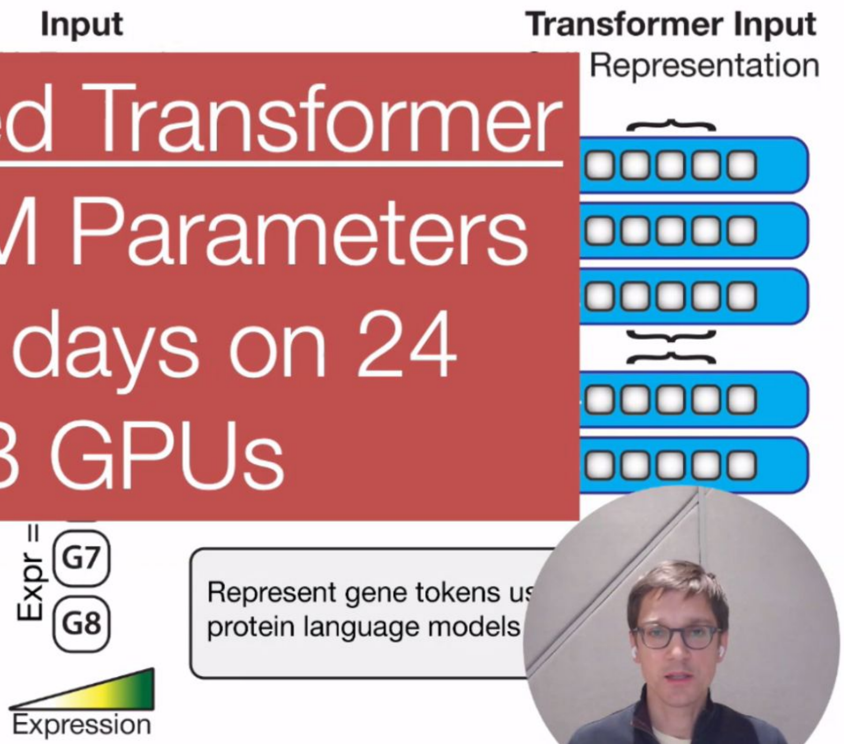
The Universal Cell Embedding (UCE) Model

Key Model Choices

- Ge
  nat
- A u
  fixe
  - a
- Self-supervised
  - Organization is emergent

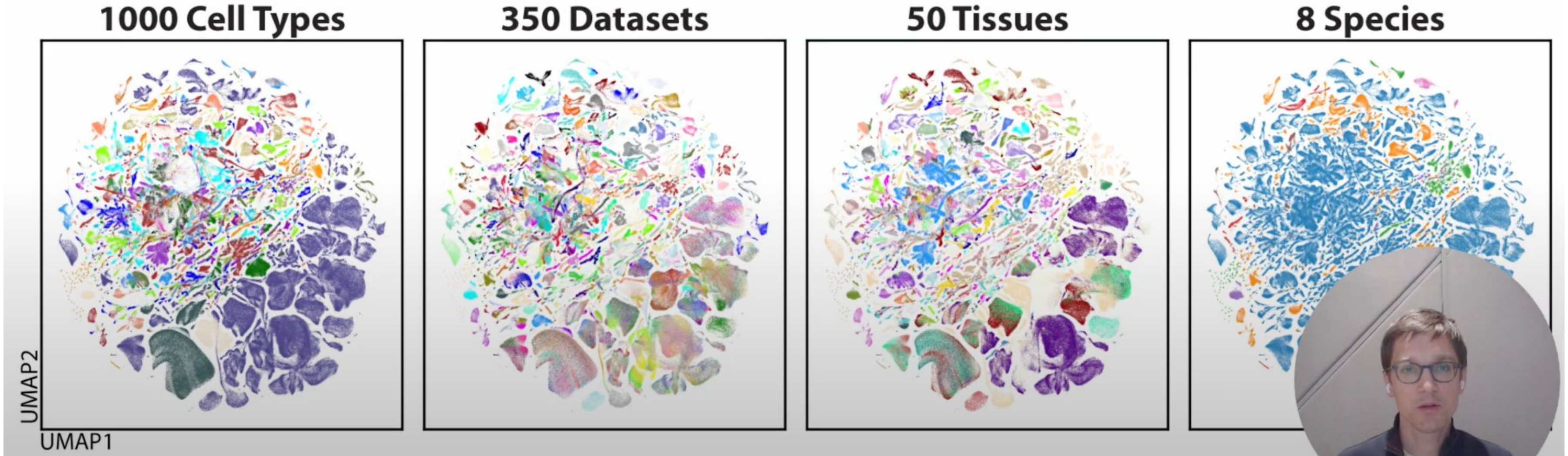
Biologically Inspired Transformer

- 33 Layers, 650M Parameters
- Trained for 40 days on 24 A100 80GB GPUs



Integrated Mega-scale Atlas (IMA)

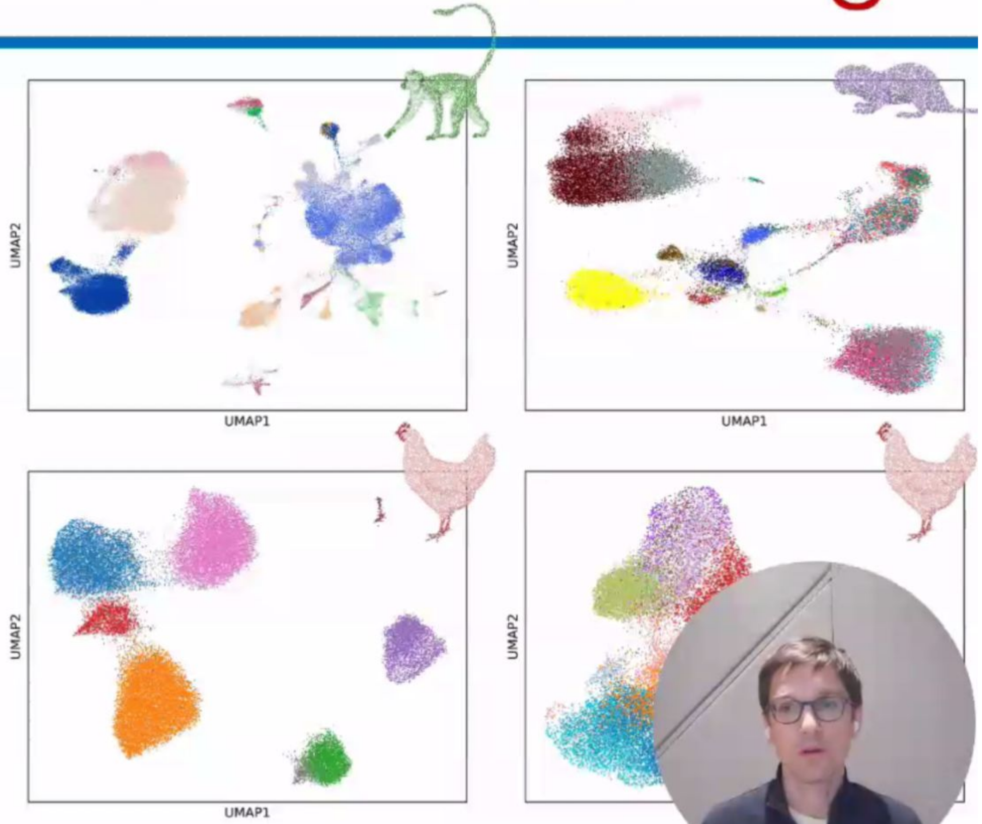
Emergent organization of 36M cells



Map new data with no fine-tuning

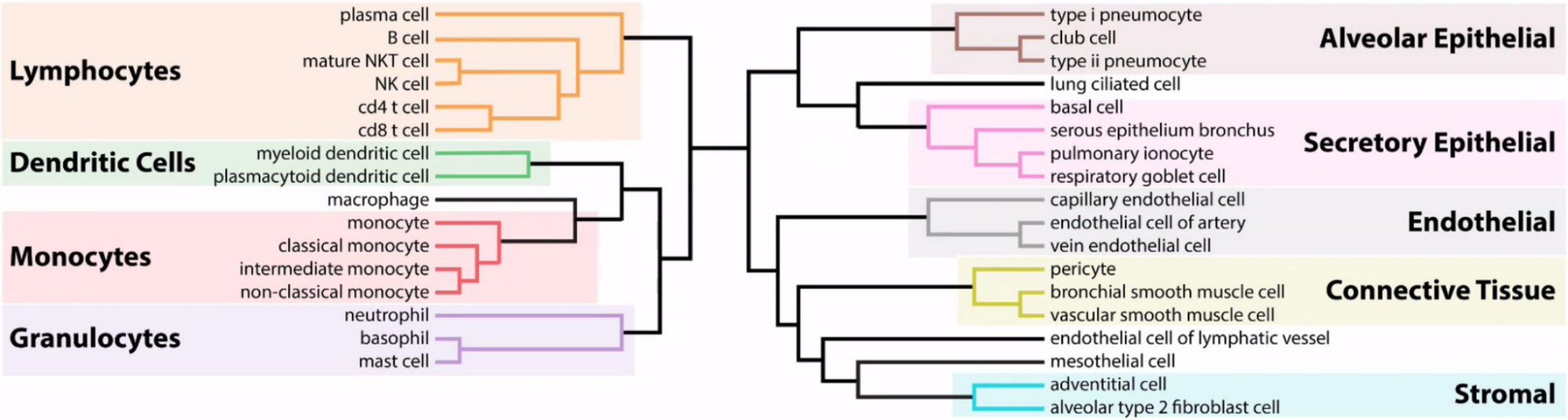
Map new data to same fixed space

- Even novel species! (No BLAST)



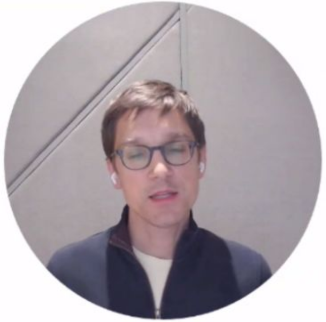
Colored by cell type

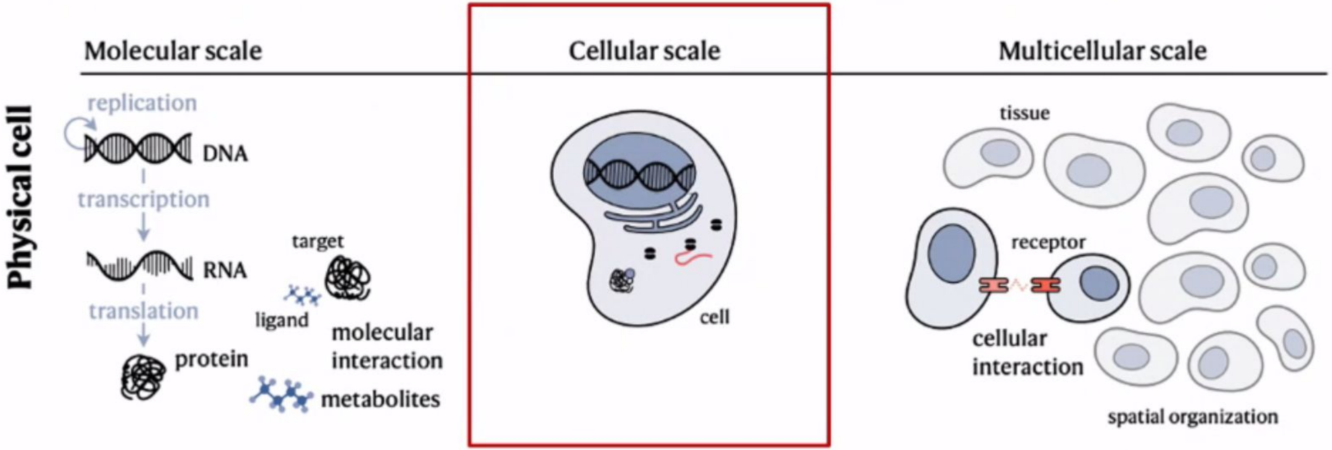
Cell Type Organization Naturally Emerges



Tabula Sapiens v2 Lung

Inferred Cell Hierarchy



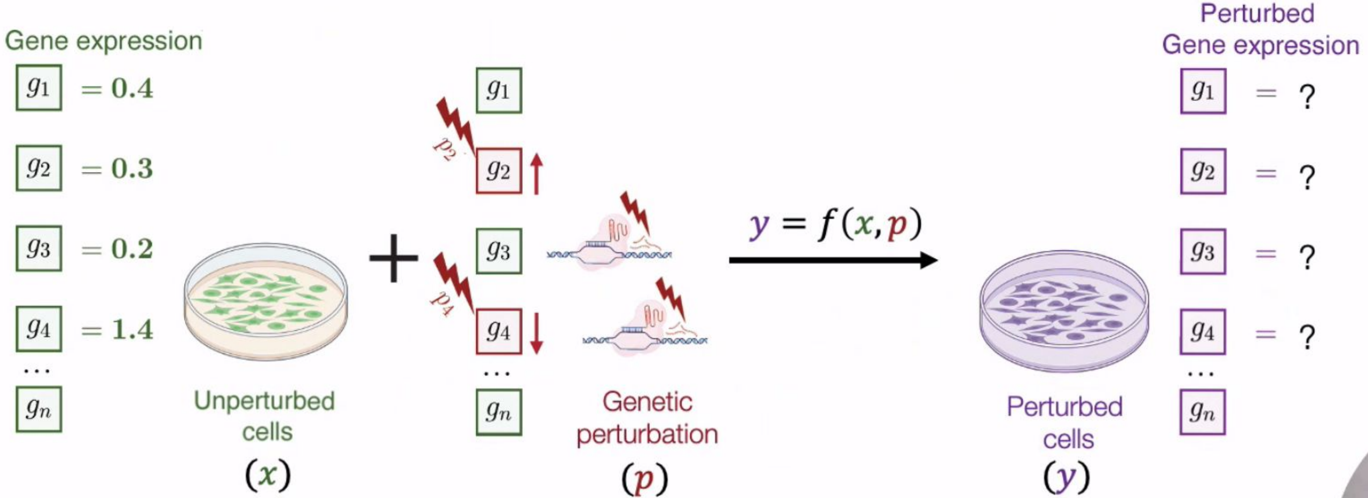


GEARS: Predicting transcriptional outcomes of novel multi-gene perturbations

Yusuf Roohani, Kexin Huang, Jure Leskovec
Stephen Quake, Jure Leskovec

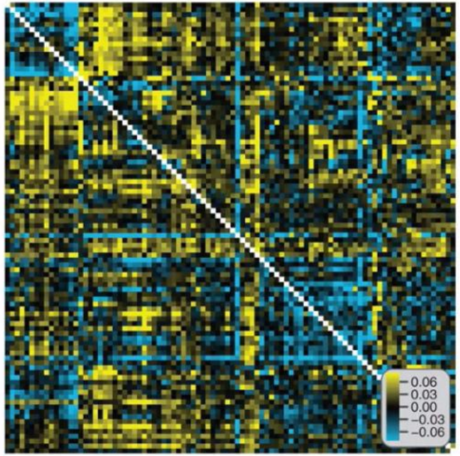
Problem Formulation

Predict the outcome of a genetic perturbation



What is the gene expression response of perturbing a combination of genes not seen experimentally perturbed

Why Predict Perturbation



Perturbational space is too vast to be explored experimentally

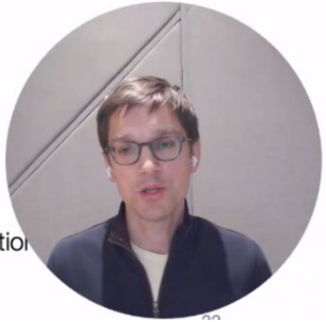
There are 4×10^8 pairwise combinations of all known protein coding genes

Pick 100 genes

5000 unique combinations of 2 gene perturbations

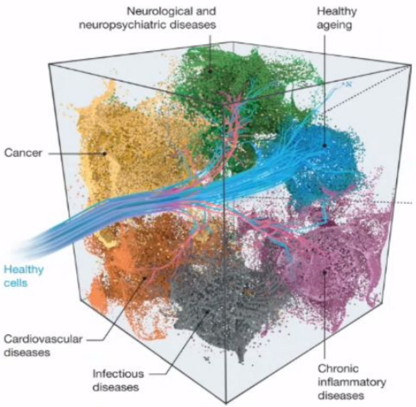
161700 unique combinations of 3 gene perturbations

3921225 unique combinatio of 4 gene perturbations



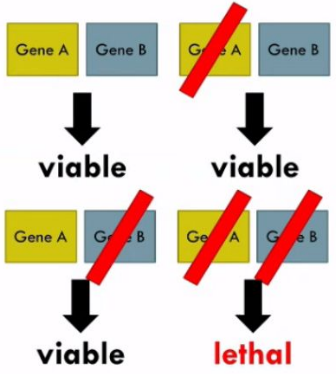
Why is this useful?

1. Drug target discovery



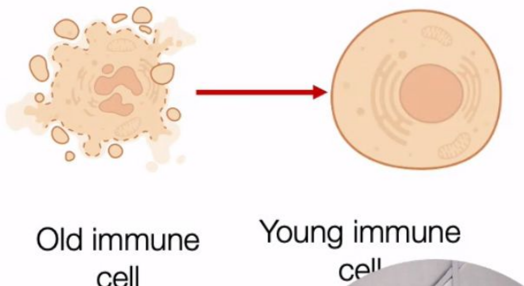
Identify therapeutic targets that can reverse disease phenotypes

2. Identifying genetic interactions

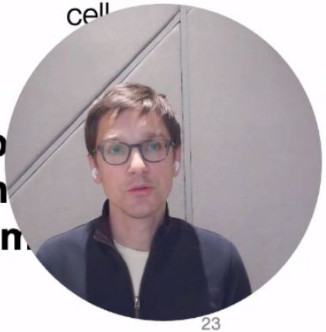


Predict genetic interactions

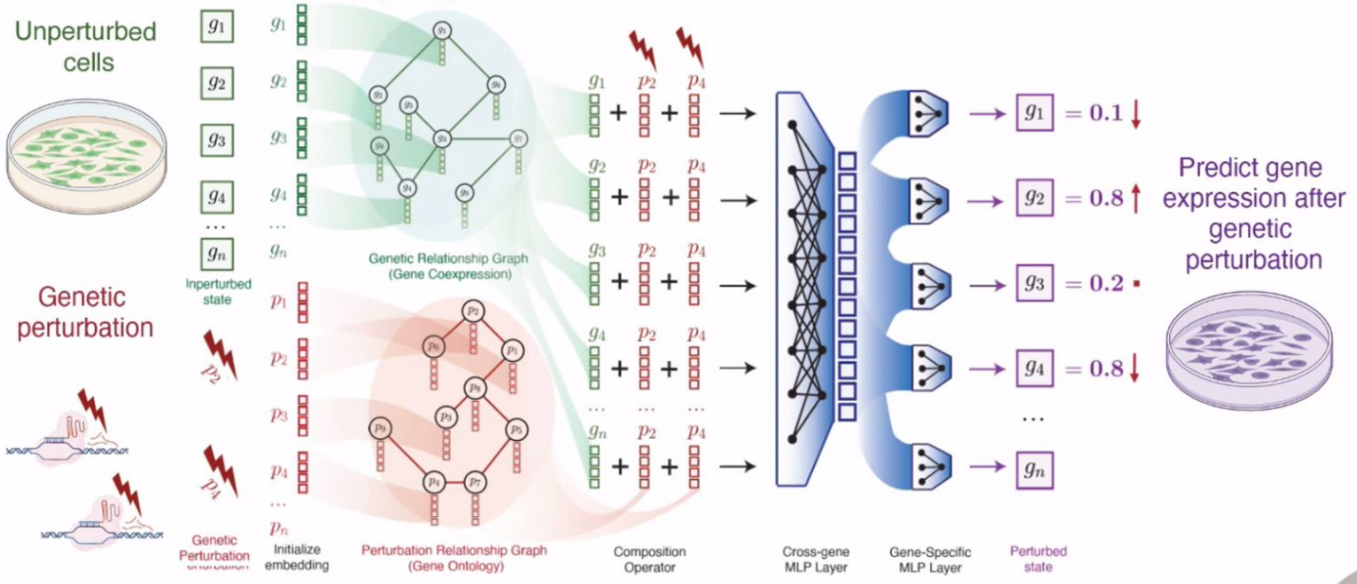
3. Re-engineering cells



Enhance p… and efficien… reprogram…



Our Approach: GEARS



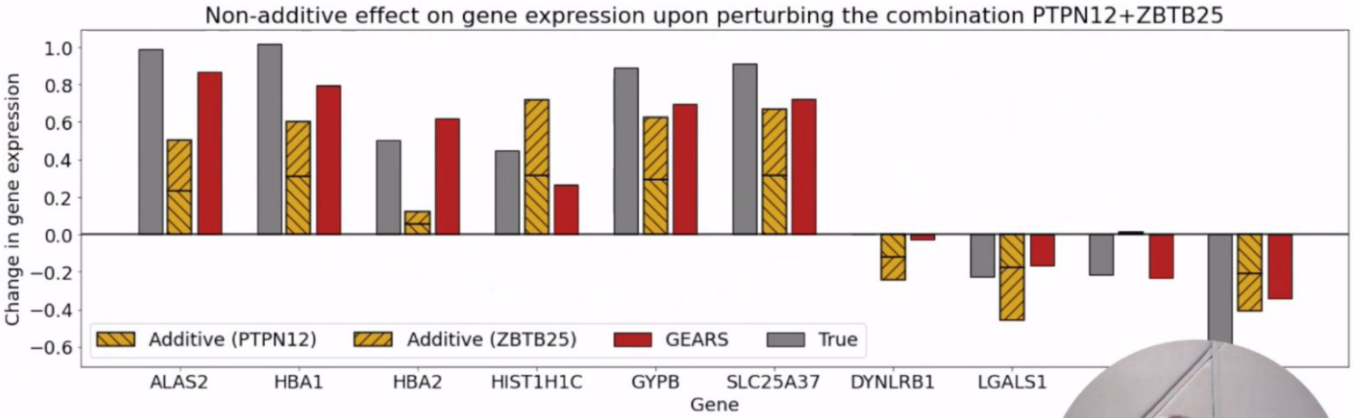
A deep learning model constrained by pri
knowledge of gene-gene relationships

[GEARS: Predicting transcriptional outcomes of novel multi-gene perturbations]

Yusuf Roohani, Kexin Huang, Jure Leskovec, Nature Biotech, 2023

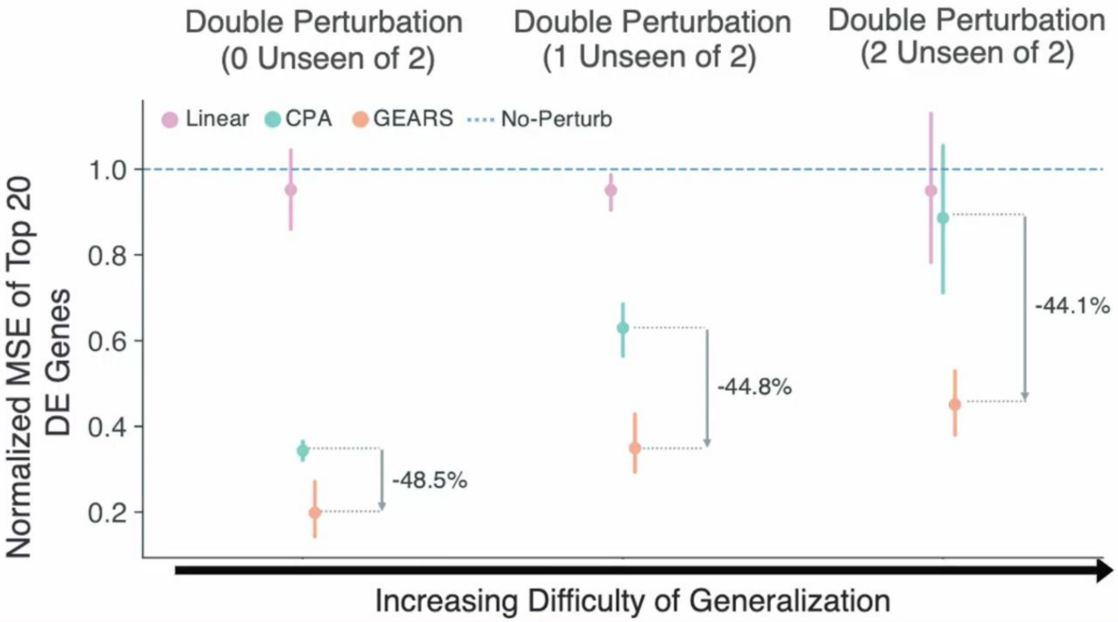
Results: Predicting non-additive genetic interactions

PTPN12 + ZBTB25

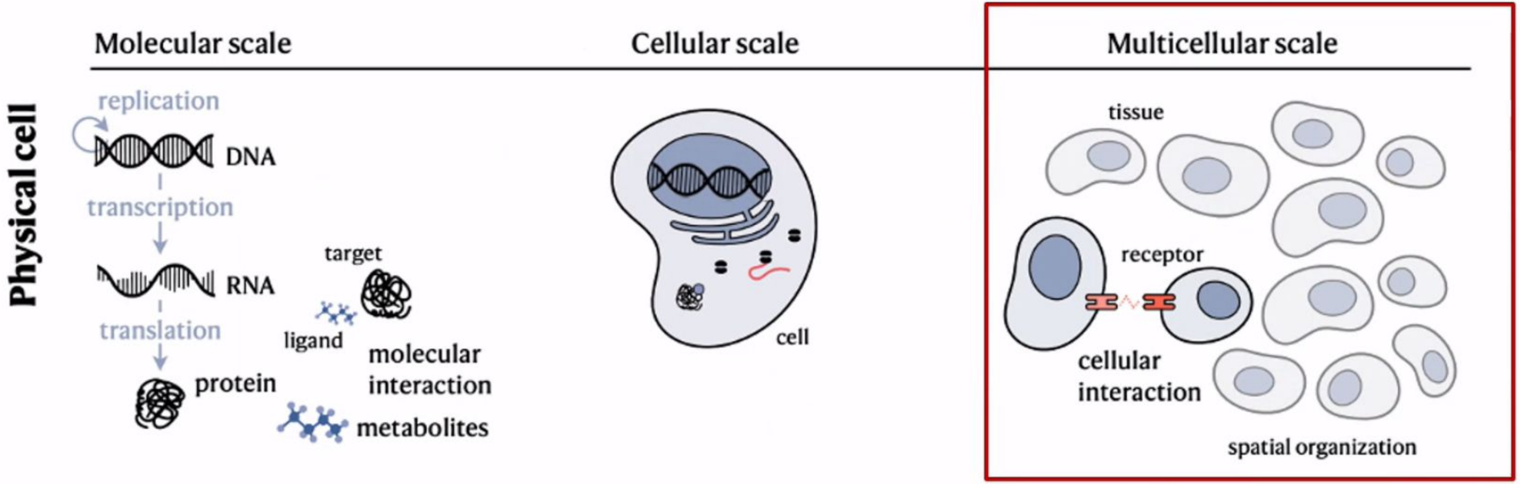


GEARS had 50% higher precision in detecting ge
interactions. 2x as accurate in predicting strongest inte

Generalizing to unseen genes

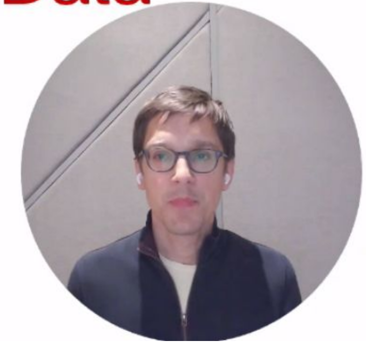


GEARS outperforms other approaches in predic
outcomes of genetic perturbation by 45%.



Annotation of Spatially Resolved Single-cell Data with STELLAR.

Brbic*, Cao*, Hickey*, Tan, Snyder, Nolan, Leskovec
Nature Methods 2022.



Spatially Resolved Single-Cell Data

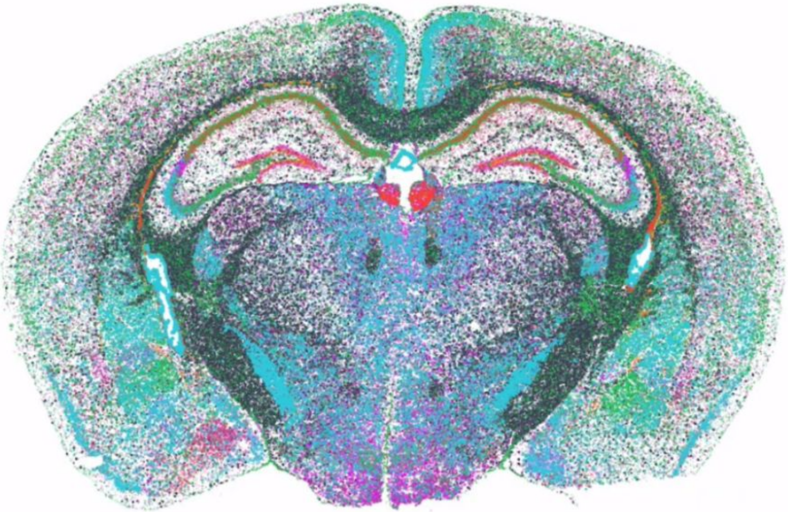
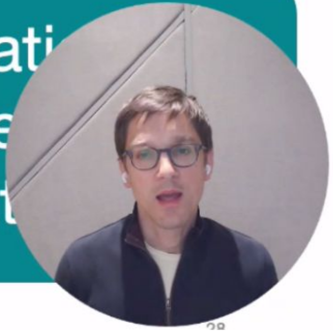


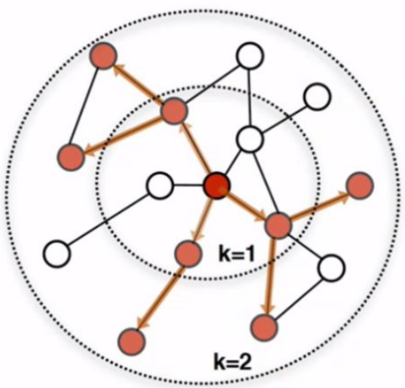
Figure from Vizgen MERFISH Mouse Brain Receptor Map dataset

- Captures spatial context of cells
- Each cell is represented with a smaller number gene/protein expressions and cell coordinates

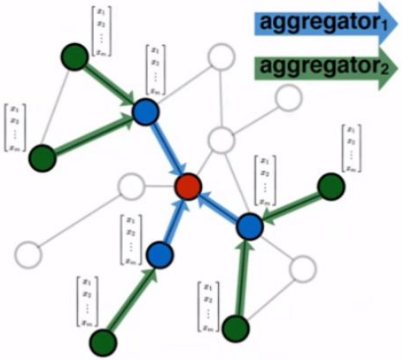
Goal: capture spati
organization of the ce
their molecular feat



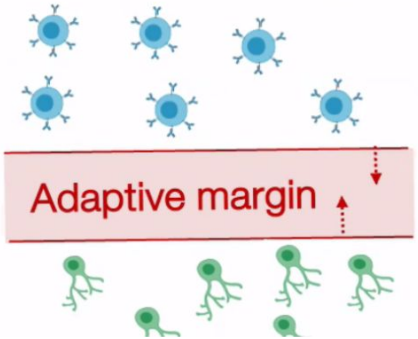
Solution: Graph Neural Networks



Determine node computation graph

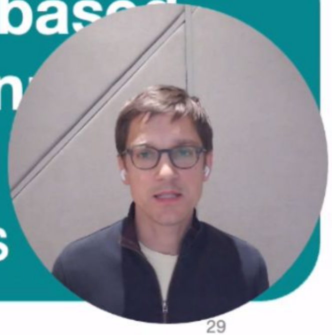


Propagate and transform information

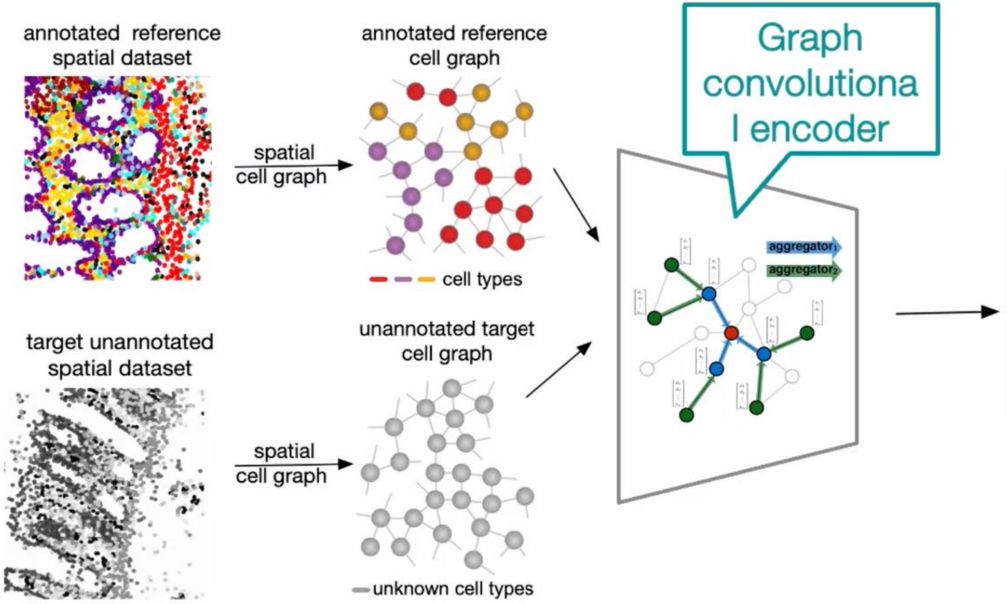


Capture spatial and molecular context of cells using graph convolutional neural networks

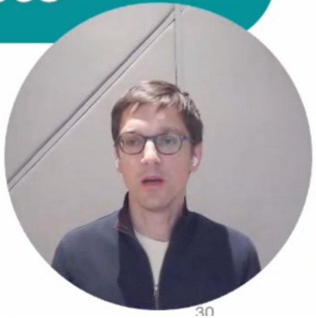
Uncertainty bas
adaptive margin
learning speed
classes



Our Method: STELLAR

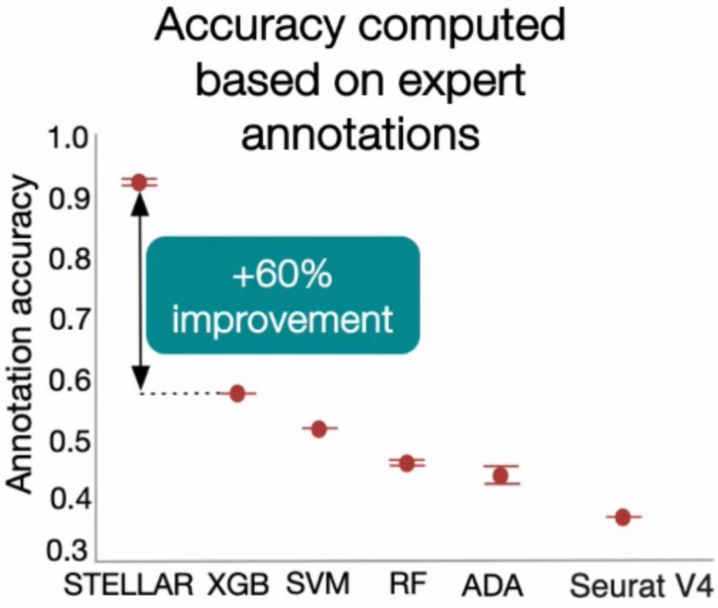


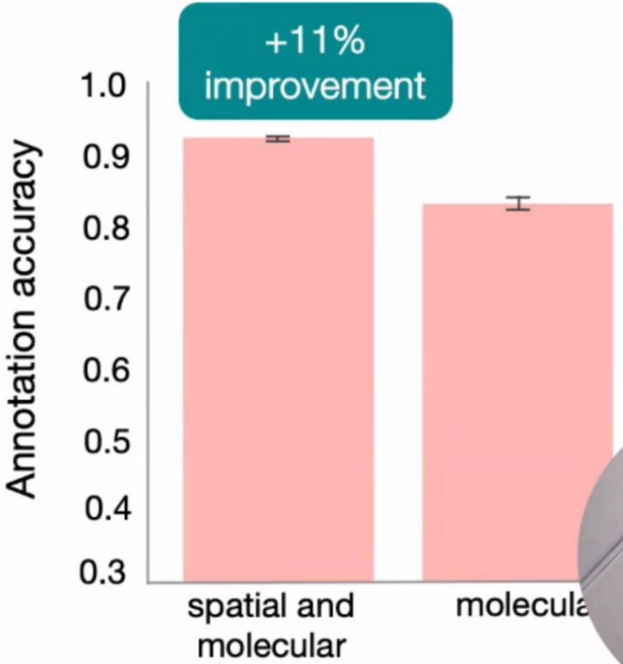
Assign cells in the unannotated dataset to either cell types in the reference dataset, or discover novel cell types



Annotation of Spatially Resolved Single-cell Data with STELLAR.
Brbic*, Cao*, Hickey*, Tan, Snyder, Nolan, Leskovec. *Nature Methods 2022.*

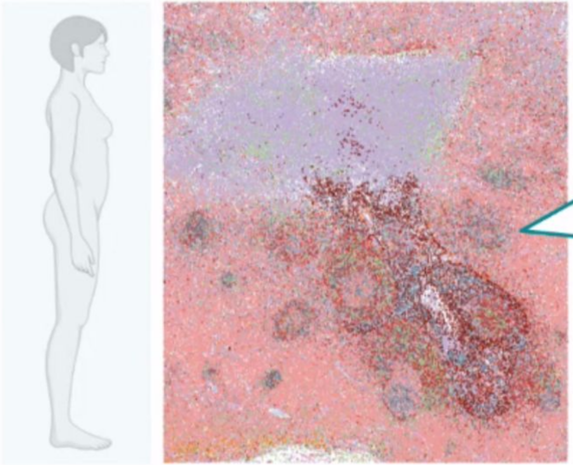
STELLAR: Cell labeling





Can We Annotate Cancer Donor Tissue using Healthy Donor Tissue?

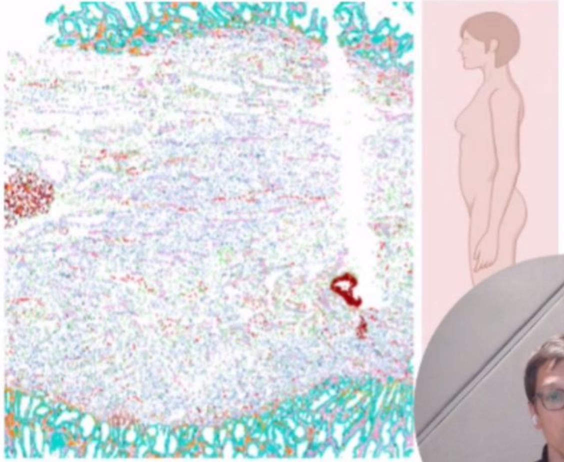
Reference labeled data:
Healthy tonsil tissue



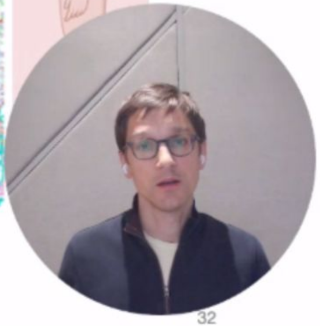
CODEX multiplexed imaging data

Colors denote different cell types

Unlabeled data:
Esophageal cancer

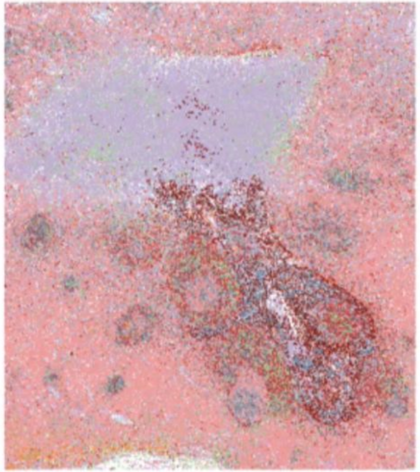


CODEX multiplexed imaging data



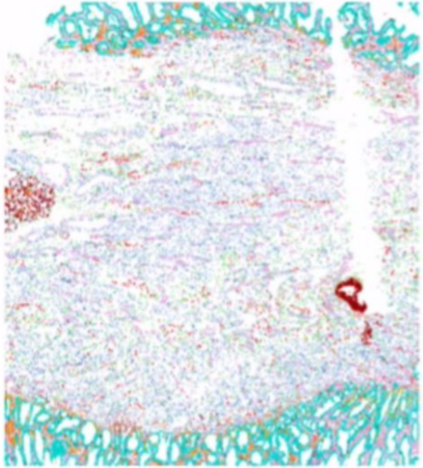
Can We Annotate Cancer Tissue using Healthy Tissue?

Healthy tonsil tissue



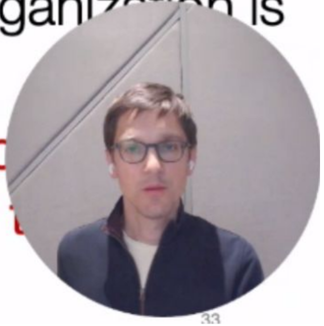
CODEX multiplexed imaging data

Esophageal cancer



CODEX multiplexed imaging data

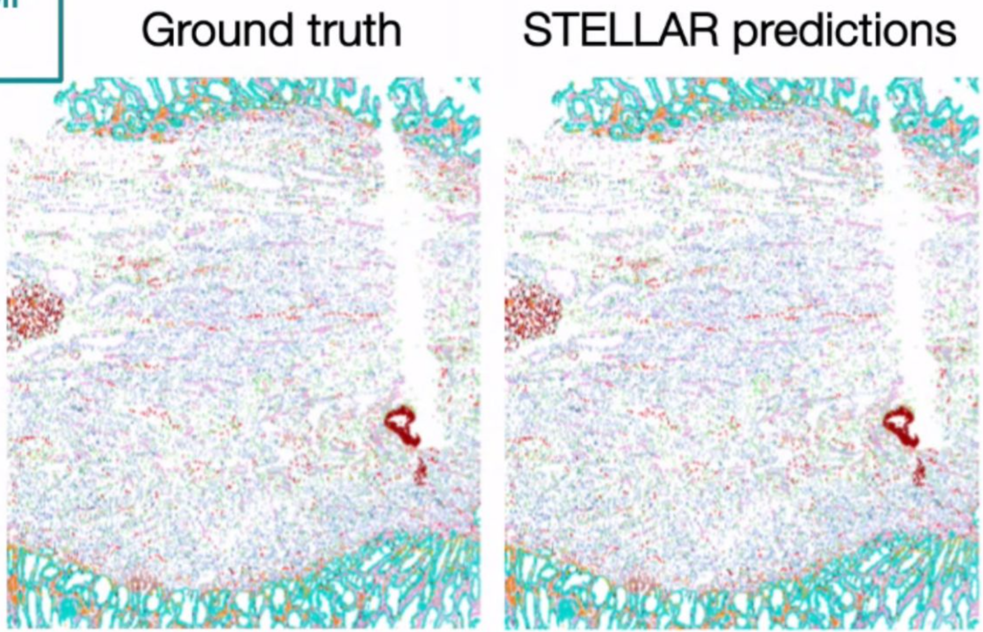
- Distribution between cell types and their spatial organization is different
- 3 out of 12 cell typ… only in the cancer t…

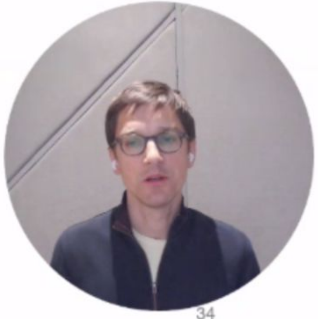


⭐ STELLAR Correctly Annotates Cancer Tissue

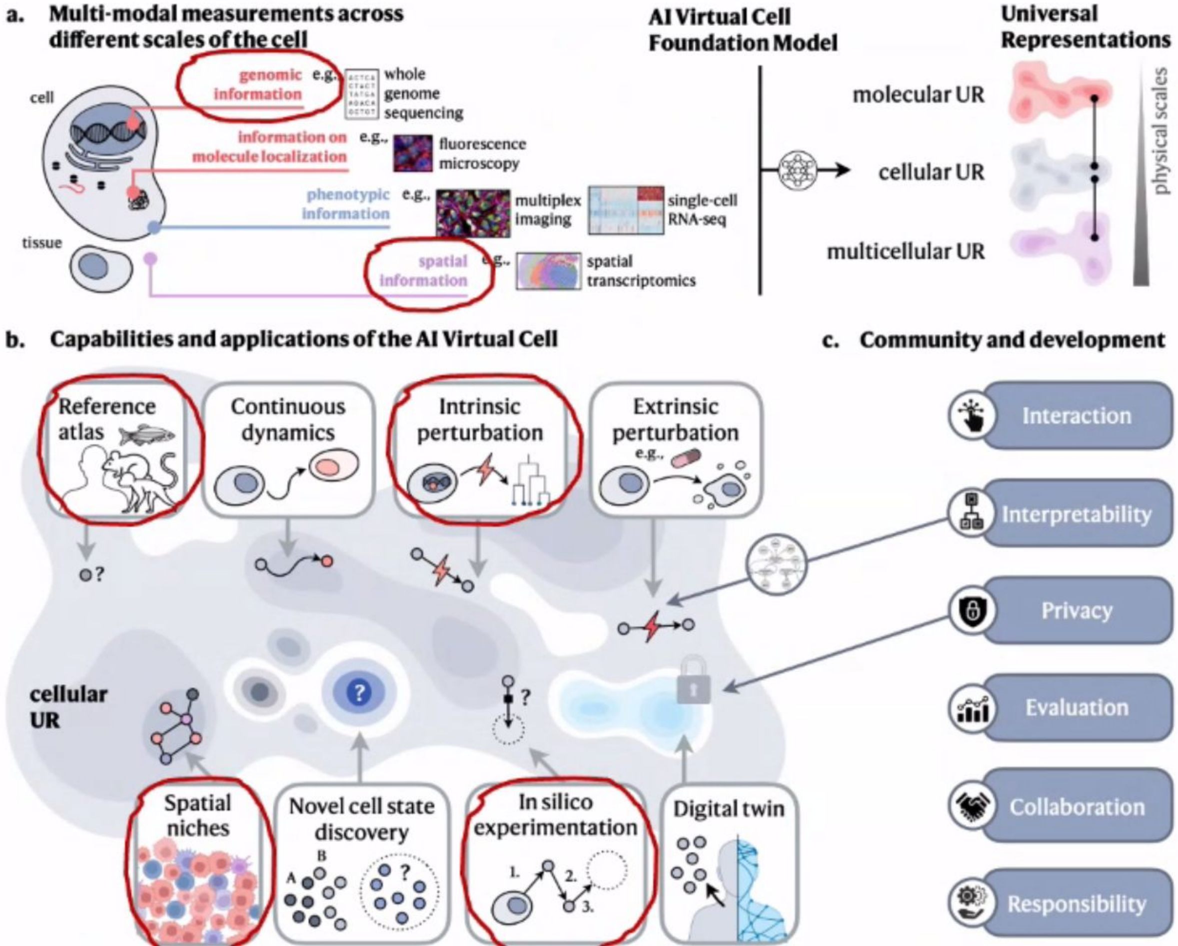
Cells are colored according to their cell types

Novel cell type



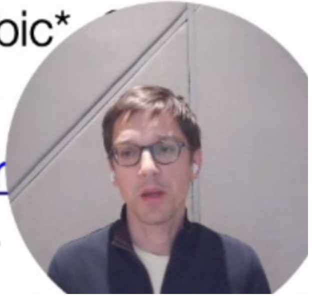


Conclusion: AI Virtual Cell



Papers

- [How to Build the Virtual Cell with Artificial Intelligence: Priorities and Opportunities](), Bunne, Roohani, Rosen et al. Cell '24.
- [Universal Cell Embeddings: A Foundation Model for Cell Biology](). Rosen, Brbić, Samotorčan, Roohani, Quake, Leskovec, '24.
- [MARS: Discovering Novel Cell Types across Heterogenous Single-cell Experiments.]() Brbic, Zitnik, Wang, Pisco, Altman, Darmanis, Leskovec. Nature Methods '20
- [Annotation of Spatially Resolved Single-cell Data with STELLAR](). Brbic* Hickey*, Tan, Snyder, Nolan, Leskovec. Nature Methods 2022.
- [GEARS: Predicting transcriptional outcomes of novel multi-gene per]() Yusuf Roohani, Kexin Huang, Jure Leskovec, Nature Biotech, 2023.



Acknowledgements

Jure Leskovec | Yusuf Roohani | Maria Brbić | Charlotte Bunne | Stephen Quake | Emma Lundberg | Aviv Regev | Theofanis Karaletsos

Zoe Piran | Hanchen Wang | Kexin Huang | Minkai Xu | Marcel Rød | Michael Bereket | Owen Queen | Kuan Pang | Lata Nair | Charilaos Kanatsou | Rok Sosic

Ayush Agrawal | Arnuv Tandon | Gashon Hussein | Aditya Tadimeti | Leon Samotorčan | Krish Parikh

Stanford ARTIFICIAL INTELLIGENCE

STANFORD BIO-X

Wu Tsai Neurosciences Institute Stanford University

NSF

Genent

A Member of the Ro

PhD Students


Camilo Ruiz

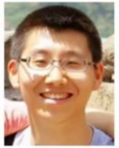

Hongyu Ren


Hamed Nilforoshan


Kaidi Cao


Kexin Huang


Jared Davis


Michi Yasunaga


Qian Huang


Serina Chang


Weihua Hu


Yanay Rosen


Yusuf Roohani

Post-Doctoral Fellows


Michael Moor


Tailin Wu

Industrial Visitors


Yanan Wang


Takashi Maruyama


Yoshitaka Sugita

Industry Partnerships

Pinterest
أرامكو السعودية
saudi aramco
amazon
DELL
VISA
UnitedHealth Group
Wikipedia The Free Encyclopedia
TOSHIBA
BOEING
NVIDIA
gsk GlaxoSmithKline
intel
NTT docomo
Viaduct
JD.COM
HITACHI
KDDI
CHASE

Funding

Office of Naval Research Science & Technology ONR
NIH
DARPA
MURI ARO
NSF
CHAN ZUCKERBERG INITIATIVE
IARPA

Research Staff


Rok Sosic

Stanford Artificial Intelligence

SNAP GROUP

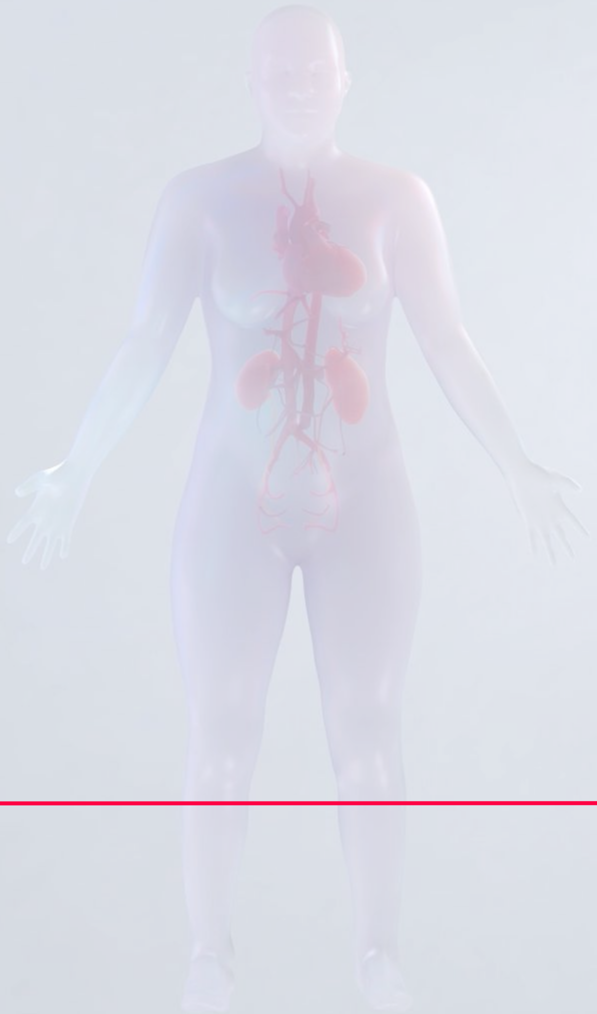
STANFORD INFOLAB

Stanford | Statistical Machine Learning Group

Collaborators

Dan Jurafsky, Linguistics, Stanford U
David Gleich, CS, Purdue U
David Grusky, Sociology, Stanford U
James Zou, Medicine, Stanford U
Jochen Profit, Medicine, Stanford U
Jon Kleinberg, CS, Cornell U
Madhav Marathe, CS, U of Virginia
Marinka Zitnik, Medicine, Harvard U
Russ Altman, Medicine, Stanford U
Scott Delp, Bioengineering, Stanford U
Chis Manning, CS, Stanford U
Sendhill Mullainathan, Economics, U Chicago
Stephen Boyd, EE, Stanford U
VS Subrahmanian, CS, U of Maryland

Q&A



https://humanatlas.io/events/2024-24h

Questions

How do we define a Multiscale Human?

How do we map a Multiscale Human?

How do we model a Multiscale Human?

How can Large Language Models (LLMs) or Retrieval-Augmented Generation (RAGs) be used to advance science and clinical practice?

Thank you