CIFAR Proposal Development Workshop

Mapping the Human Body Across Scales

May 24 & 25

Mapping the Human Body Across Scales: Letter of Intent Recap 1/2

- Our proposal brings together an international team of researchers to map one of science's last great frontiers: the human body

- How do our molecules and cells collaborate to make a working body?

- Traditional maps are flat and static, but the human body is a complex 3D structure changing dynamically through time

- To map the body in 4D, we will use the latest techniques in microscopy (from X-Rays to electron beams to lasers) and genomics (thanks to the single cell revolution)

Mapping the Human Body Across Scales: Letter of Intent Recap 1/2

- We will discover new ways of categorising and interpreting data at different scales

- We want to make an atlas of the human body that is as useful as Google Maps, allowing the user to zoom in from organ to tissue to cell to molecule, at each level answering questions about how these structures change in development and disease

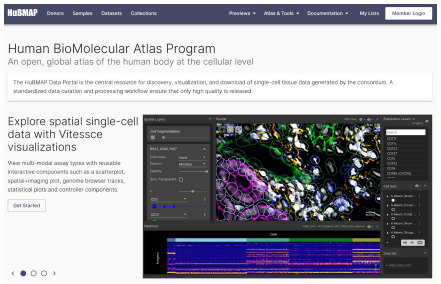
- ***Our multiscale atlas promises to change the future of being human***

Interconnected research areas

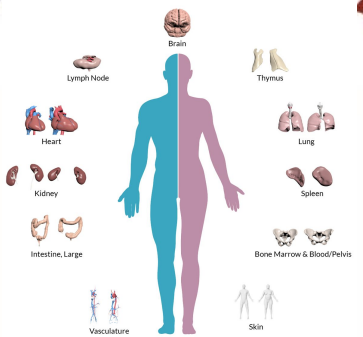
- Data visualisation
- Multiscale imaging
- From cells to molecules (anatomical structure > molecular ID/function)
- Studying mechanisms in model systems
- Computation at every level

Please add your perspective and what you hope to bring to this project to the slides below

Katy Börner



Data Generation, Harmonization and Integration



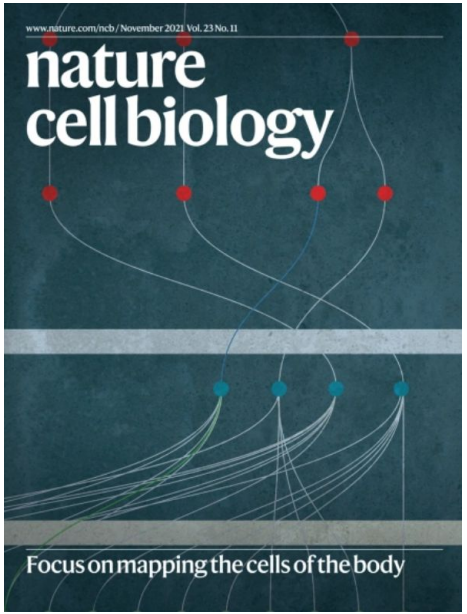
Human Reference Atlas

Advance Research Improve Human Health

Define Human Reference Atlas

The Human Reference Atlas (HRA)

1. defines the 3D space and shape of anatomical structures and cell types that are of biomedical relevance plus the biomarkers used to characterize them. Anatomical structures, cell types and biomarkers are validated and represented in/added to ontologies (Uberon/FMA, CL, HGNC).
2. defines how new datasets can be mapped to the HRA, e.g., spatially using the Visible Human CCF or Vasculature CCF (or both, see next slide), via ASCT+B ontology terms/IDs, or via gene expression data as in Azimuth.
3. it is
    - authoritative (there exists expert agreement and it was validated by data),
    - computable (supports API queries, UIs),
    - published as LOD (connected to gene, disease, and other ontologies and data),
    - open (anyone can use the HRA data and code), and
    - continuously evolving (e.g., as new technologies become available).

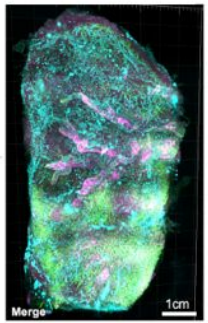


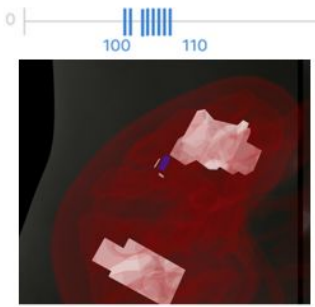
https://www.nature.com/articles/s41556-021-00788-6

Constructing a Human Reference Atlas

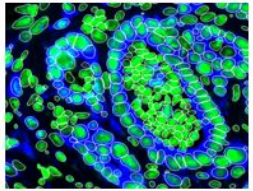
Tissue Data

Whole organ data,
e.g., light-sheet data

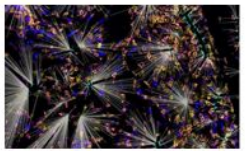




Registered tissue blocks
multiple sections per block



Cell, FTU, AS
segmentation/annotation



Cell type populations
cell distances to vessels
diverse biomarkers per cell

1 Human

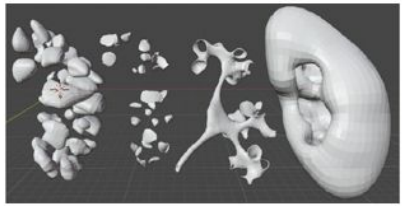
Organs
(kidney shown here)

10,000+
Anatomical
Structures

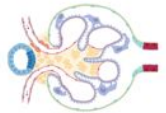
37T Cells

Reference Atlas

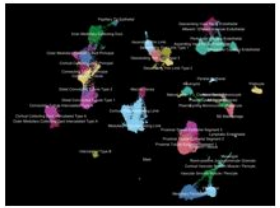
3D coordinate system,
geometric mapping process



2D Reference FTUs

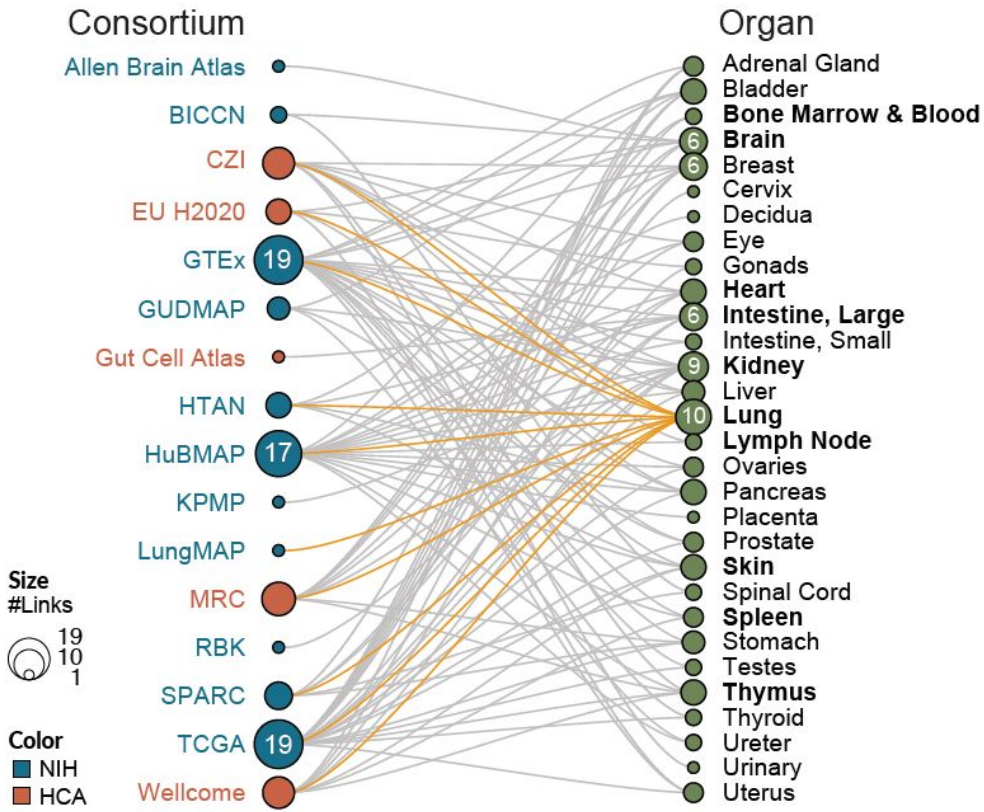


Reference object library with crosswalks to ASCT+B tables,
Synthetic HRA 3D model with cell type populations



2&3D single cell models
cell graphs

Constructing a Human Reference Atlas - Together!



Anatomical Structures (AS)

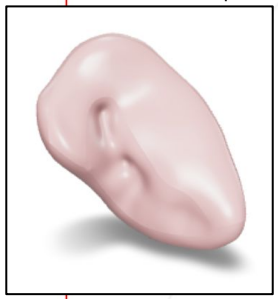
Cell Types (CT)

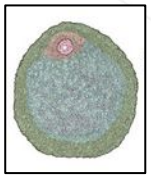
Biomarkers (B)

Partonomy Tree
part_of

- Spleen
  - Splenic Pulp
    - Red Pulp
      - Venous Sinuses
        - Pulp Arteries
        - Penicillar Arterioles
        - Sheathed Arterioles
        - Arterial Capillaries
        - Sinuses
        - Veins
      - Stroma
      - Splenic Cords
    - White Pulp
      - Secondary Follicles
        - Germinal Centers
      - Primary & Secondary Folli...
        - Mantle Zone
        - Superficial (Marginal) Zone
      - Central Arteriole (in follicl...
      - PALS and Follicles
      - PALS
      - Perifollicular Zone
  - Splenic Artery

VHP F Spleen





White pulp of spleen

Bimodal network describing which CT are *located_in* what AS

Typology Tree
is_a

- adventitial stromal cell
- B cell
- Dendritic cell
- Endothelial
- Endothelial cell
- Erythrocytes
- fibroblast
- Fibroblastic reticular cell
- Follicular Dendritic cell
- Granulocytes
- Littoral cell
- Lymphatic endothelium
- macrophage
- Monocytes
- Myofibroblast
- neurons
- NK cell
- Plasma cell
- Plasmablasts
- Platelets

BG - Genes
BP - Proteins

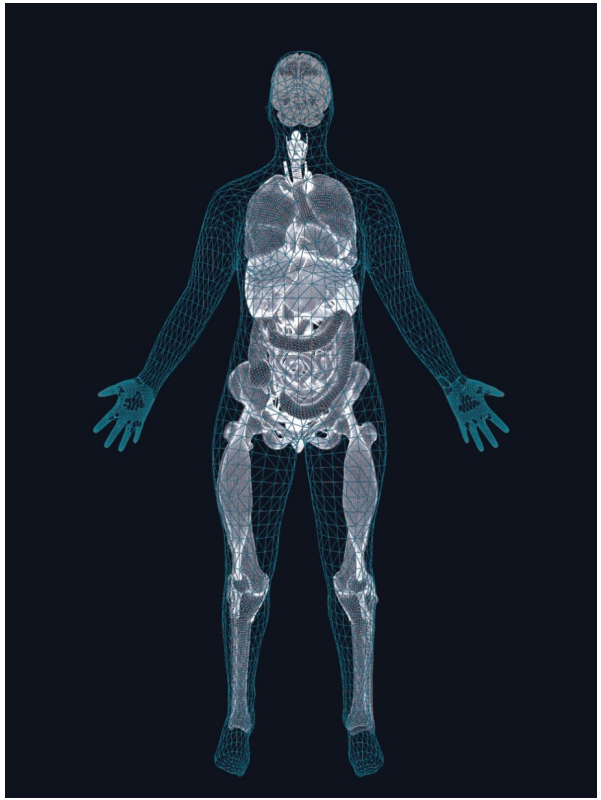
Bimodal network describing which B *characterize* what CT

- CD10
- CD11b
- CD11c
- CD138
- CD14
- CD141
- CD15
- CD163
- CD19
- CD20
- CD21
- CD22
- CD23+
- CD235a
- CD27
- CD27-
- CD271
- CD271-
- CD3
- CD3-
- CD31
- CD34
- CD4
- CD4 (helper)
- CD41

Choose **version** of ASCT+B datasets: 2nd Release, December '21

Organ	#AS	#CT	#B Total	#BG	#BP	#AS-AS	#AS-CT	#CT-B
Blood	1	30	159	112	47	1	30	506
Blood Vasculature	841	2	1	1	0	869	606	2
Bone Marrow	1	47	262	198	64	1	47	838
Brain	183	127	257	257	0	183	127	346
Eye	26	53	136	61	75	27	58	404
Fallopian Tube	55	22	25	13	12	72	65	32
Heart	50	23	45	45	0	60	183	74
Kidney	61	62	150	150	0	62	60	257
Knee	32	19	14	0	14	32	8	17
Large Intestine	54	57	167	84	83	287	1,156	352
Liver	17	30	62	16	46	17	31	75
Lung	146	83	180	174	6	909	1065	267
Lymph Node	34	45	223	106	117	43	86	499
Lymph Vasculature	4	1	1	1	0	4	2	1
Ovary	71	7	13	7	6	109	12	5
Pancreas	32	32	44	42	2	162	229	101
Peripheral Nervous System	782	1	2	1	1	803	609	2
Prostate	4	12	31	31	0	4	12	36
Skin	15	36	70	0	70	17	19	101
Small Intestine	38	48	13	13	0	69	185	13
Spleen	37	61	194	85	109	50	129	424
Thymus	17	52	394	318	76	28	39	620
Ureter	7	14	30	30	0	7	14	61
Urinary Bladder	16	15	30	30	0	16	16	63
Uterus	58	19	45	39	6	73	28	65
Totals:	**2,582**	**898**	**2,548**	**1,814**	**734**	**3,905**	**4,816**	**5,161**

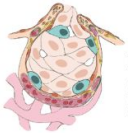
https://hubmapconsortium.github.io/ccf/pages/ccf-anatomical-structures.html



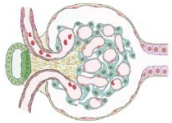
https://hubmapconsortium.github.io/ccf/pages/ccf-3d-reference-library.html (NLM VH organs)
https://community.brain-map.org/t/allen-human-reference-atlas-3d-2020-new/ (brain)
https://www3.cs.stonybrook.edu/~ari/ (male colon)

2D FTUs: Small

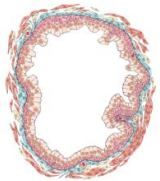
Pancreas islets
of Langerhans
(0.1 mm)



Pulmonary alveolar
parenchyma
(0.2 mm)



Renal corpuscle
(0.2 mm)

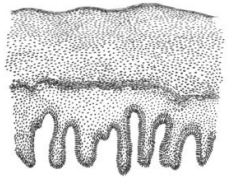


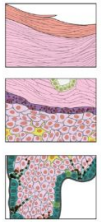
Crypt of Lieberkuhn
(0.5-0.7 mm)

Prostatic glandular acinus
(0.2 mm)

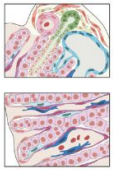
0.2 mm

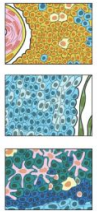
Medium





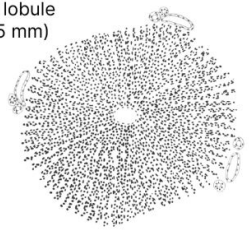
Epithelium
(0.6 mm)

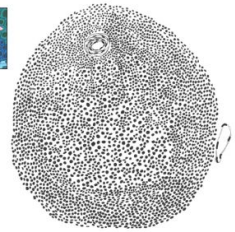




White pulp of
spleen (0.5-1 mm)

Liver lobule
(1-2.5 mm)

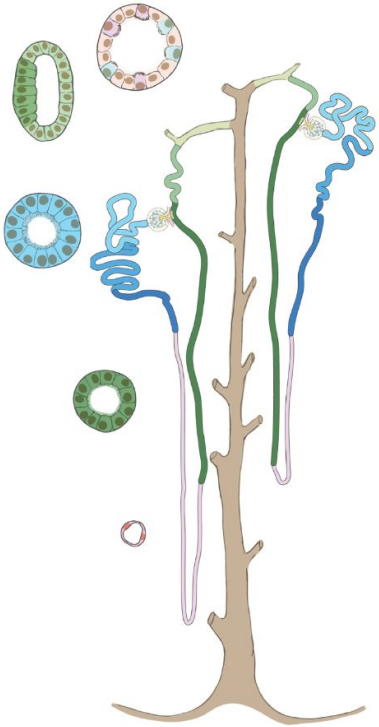




Thymus lobule
(0.5-2 mm)

0.5 mm

Large



Nephron tubule
(3-5 mm)

1 mm

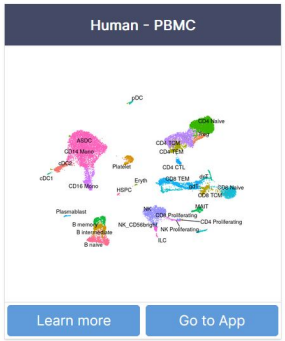
Construct a Human Reference Atlas

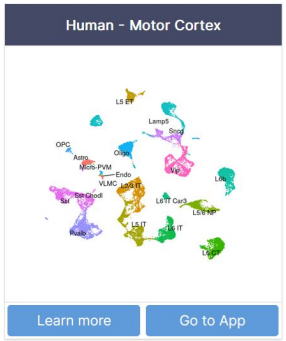
https://azimuth.hubmapconsortium.org

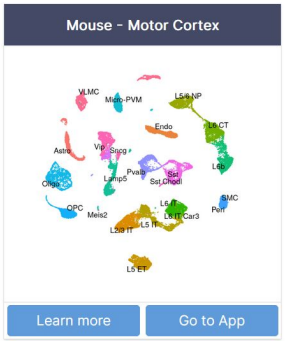
9 references and 1,036 cell types

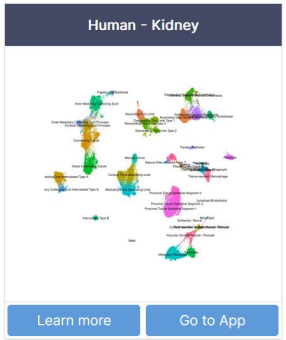
> 12,000 datasets uploaded and mapped from the community

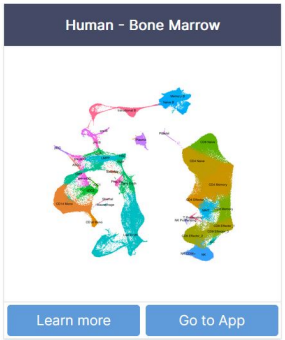
> 187,000,000 cells uploaded and mapped from the community

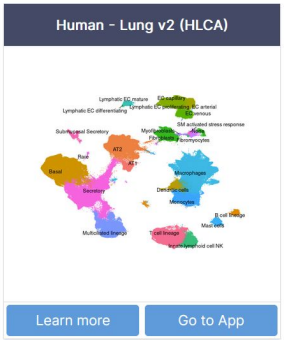








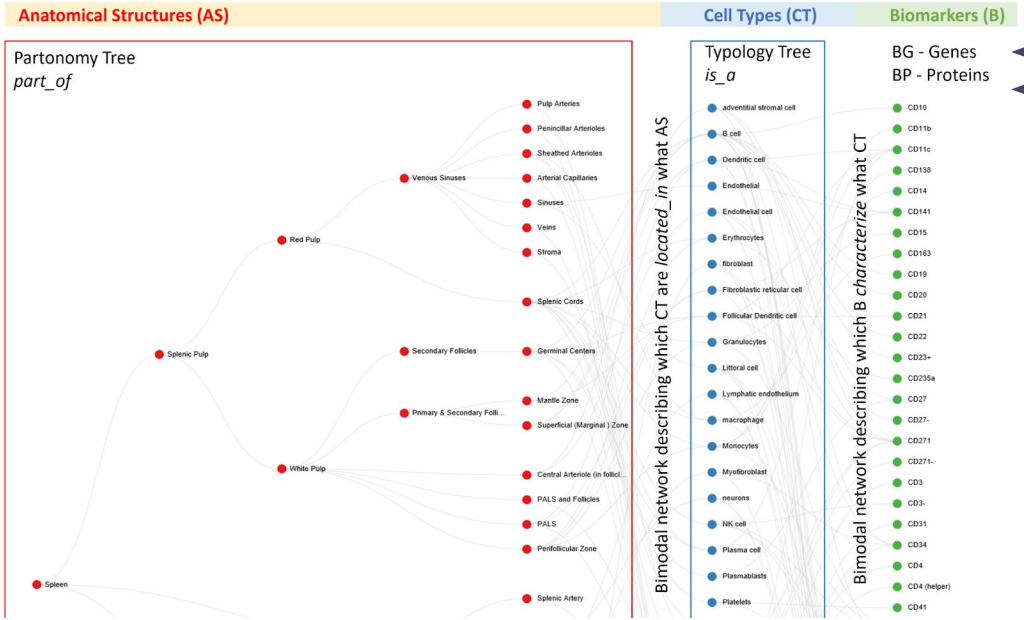




HRA Validation/ Expansion

New ATLAS publications

2D/3D Maps & Ontology Crosswalks



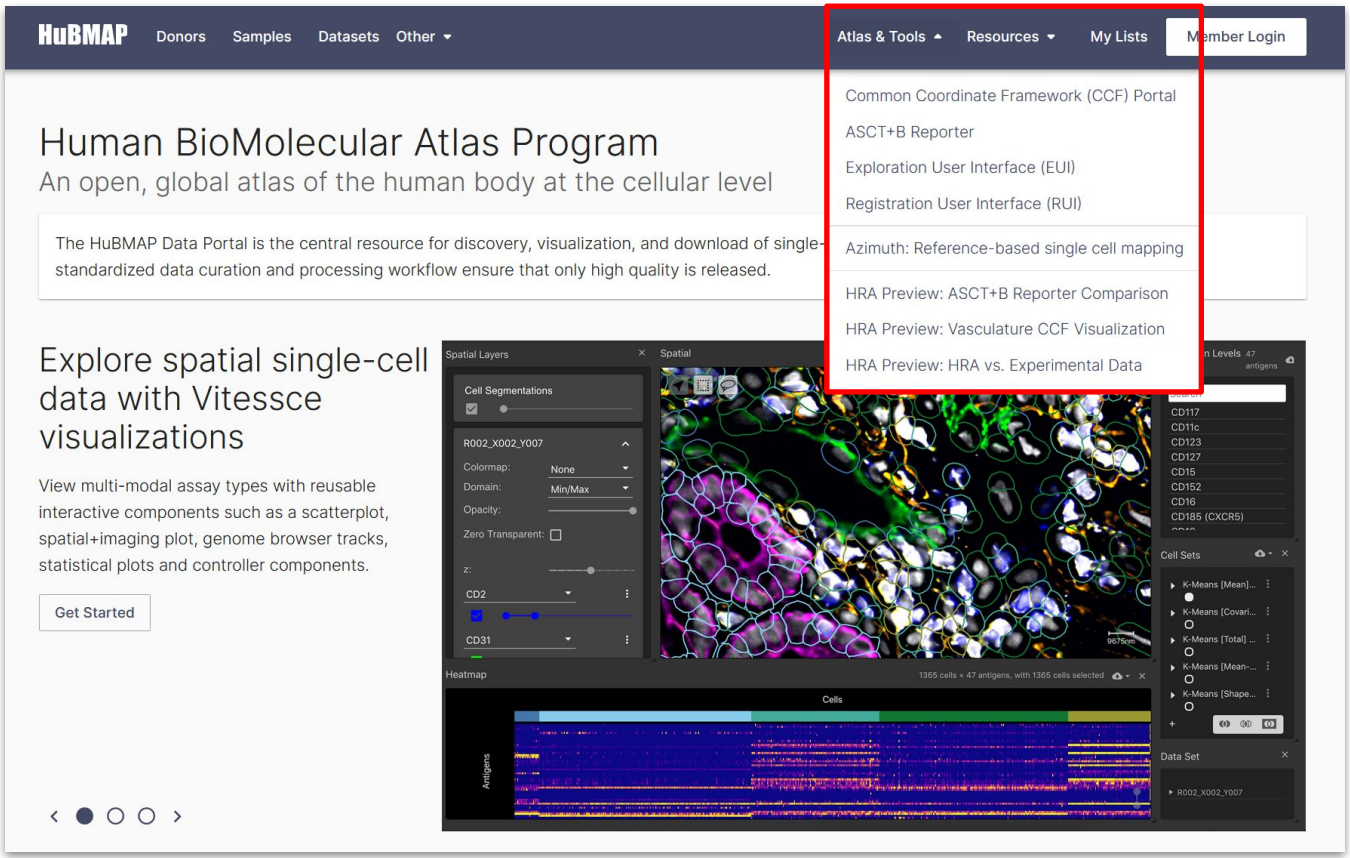
Azimuth Maps

OMAPs

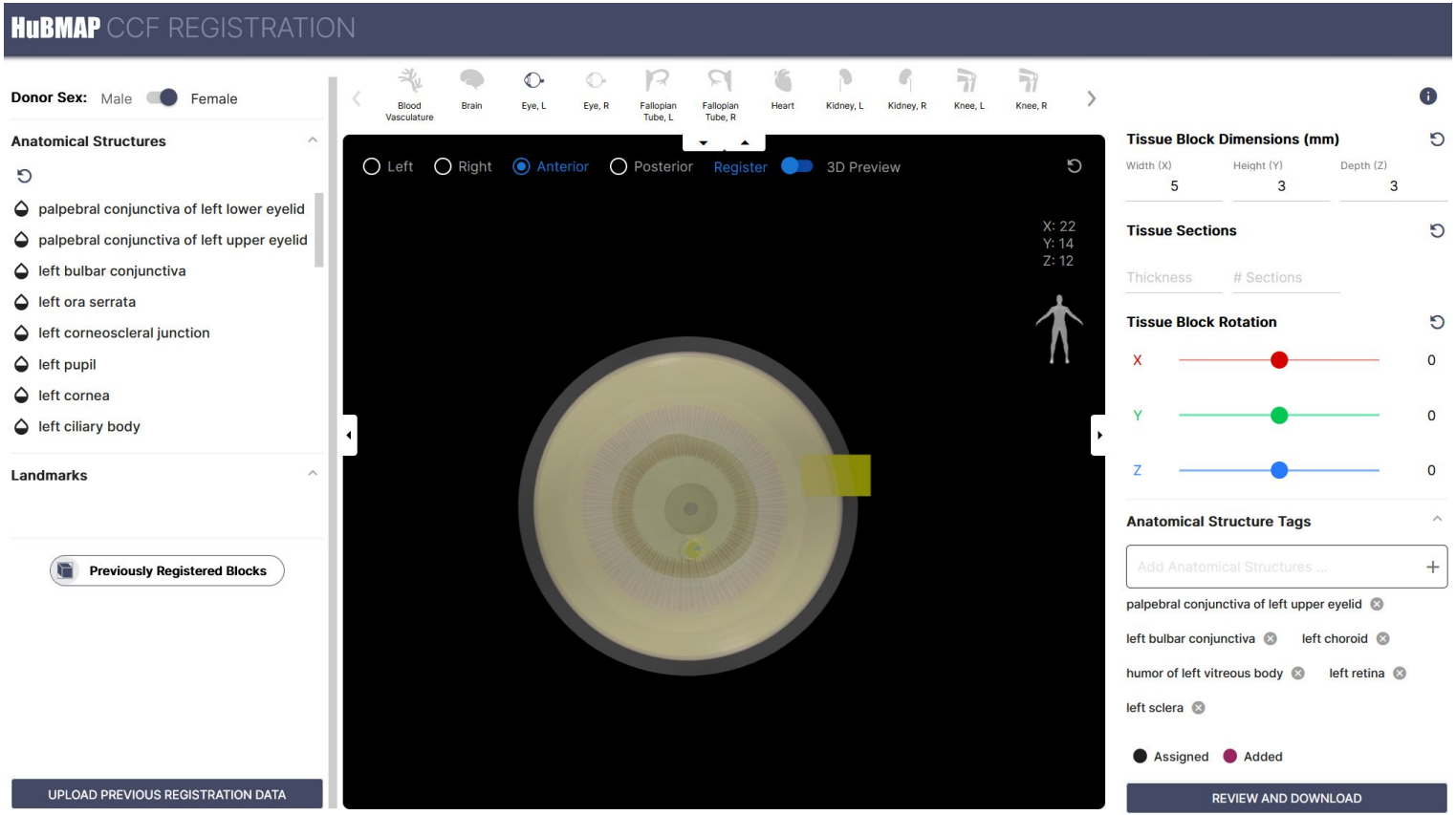
New ATLAS datasets

BF – Proteoforms
BL – Lipids
BM – Metabolites

Using the Human Reference Atlas



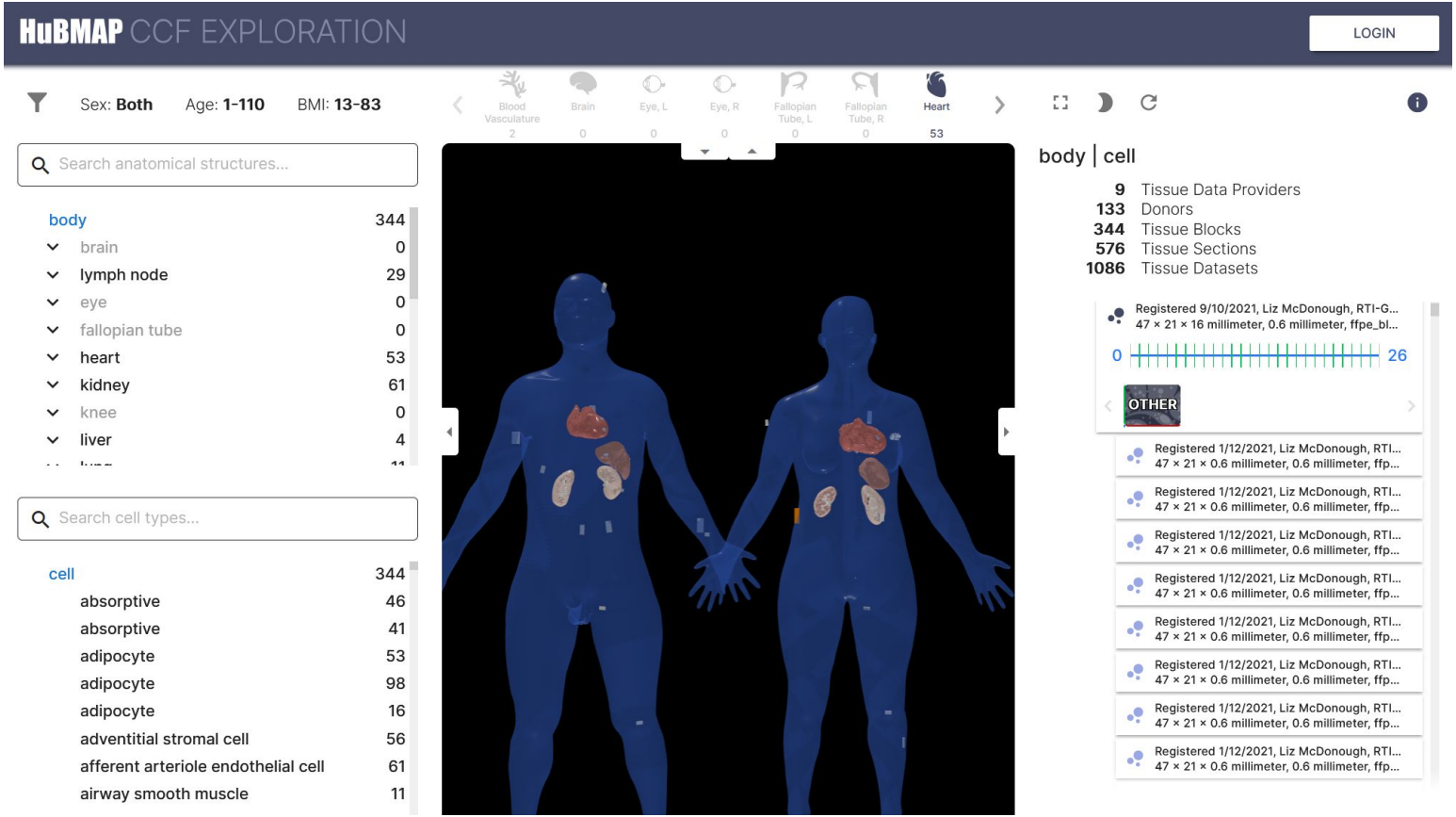
CCF Registration User Interface (RUI)



https://hubmapconsortium.github.io/ccf-ui/rui/

bioRxiv doi: 10.1101/2021.12.30.474265

CCF Exploration User Interface (EUI)



https://portal.hubmapconsortium.org/ccf-eui

bioRxiv doi: 10.1101/2021.12.30.474265

Configure Spatial Search

Donor Sex: **Male** Organ: **Heart** Edit

Probing Sphere Radius 50 cm

Tissue Blocks 20

- **Aortic Valve Apex Male**
  Entered 3/16/2021, Peter Hanna, SPARC/UCLA
- **Left Venticle Apex Male**
  Entered 3/16/2021, Peter Hanna, SPARC/UCLA
- **Right Venticle Apex Male**
  Entered 3/16/2021, Peter Hanna, SPARC/UCLA
- **Septum Apex Male**
  Entered 3/16/2021, Peter Hanna, SPARC/UCLA

Unique Anatomical Structures 6

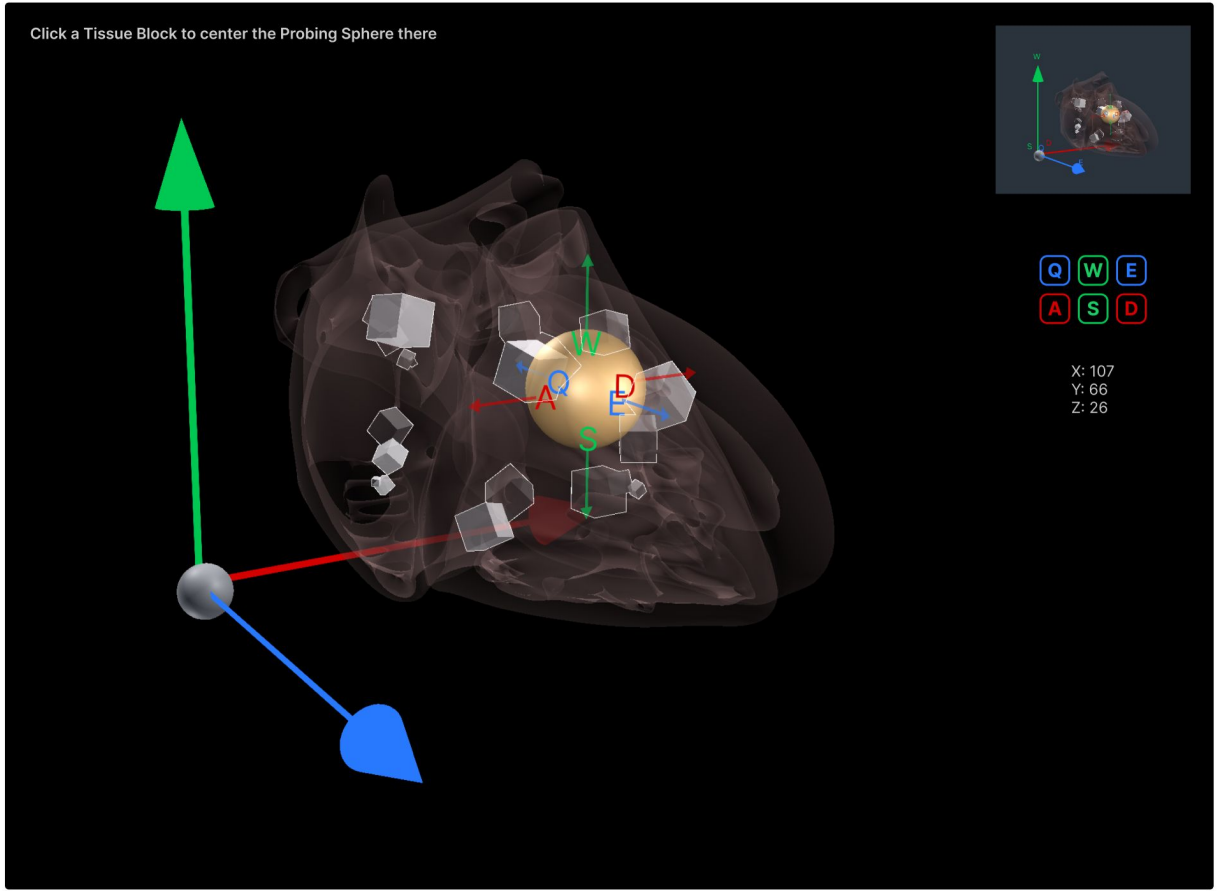
- Aortic Valve 1
- Left Atrium 1
- Left Ventricle 4
- Septum 5
- Right Atrium 4
- Right Ventricle 2

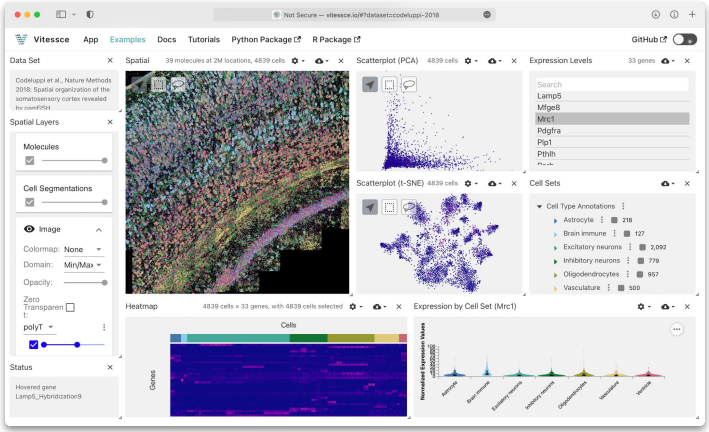
Unique Cell Type Predictions via ASCT+B Tables 9

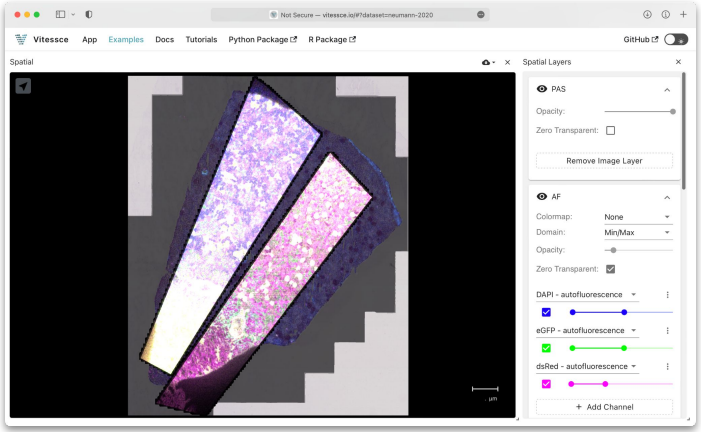
- R2D2 16
- CD8 16
- Epithelial 13
- C3PO 15
- ABCD25 16
- BB8 4

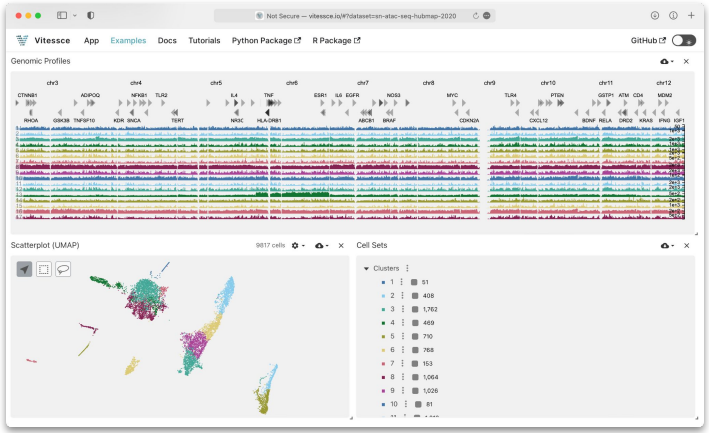
Reset Probing Sphere | Reset Camera View

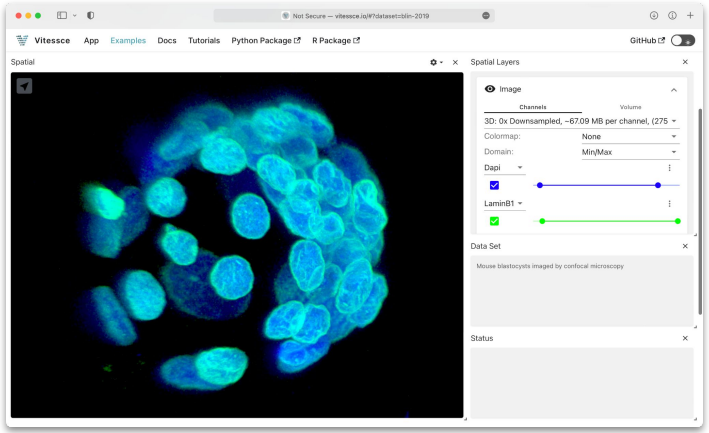
Add Spatial Search











http://vitessce.io

https://rdcu.be/cNlvp

APIs: Accessing the Human Reference Atlas

Indiana U, Stanford U, and EBI are collaborating closely on using Linked Open Data/Semantic Web Standards in support of ontology development and reasoning. Linked open data compatible with the Semantic Web is used as the ground truth. The CCF.OWL is published on Bioportal, https://bioportal.bioontology.org/ontologies/CCF/

All CCF UIs (e.g., RUI, EUI, ASCT+B Reporter) and APIs are using the CCF.OWL 1.8 data (2.0 coming in June, alpha available now). Queries can be expressed in SPARQL and exposed as standard HTTP APIs to support a whole ecosystem of collaborative and compatible APIs, libraries, UIs.

ASCT+B API Links:

- API Endpoint (includes interactive documentation): https://asctb-api.herokuapp.com
- API Documentation: https://hubmapconsortium.github.io/ccf-asct-reporter/docs/api
- OpenAPI specification: https://asctb-api.herokuapp.com/asctb-api-spec.yaml

CCF-API Links:

- API Endpoint (includes interactive documentation): https://ccf-api.hubmapconsortium.org
- API Documentation and OpenAPI specification: https://ccf-api.hubmapconsortium.org
- API Database backend is n3.js: https://github.com/rdfjs/N3.js
- Code to instantiate/use CCF Database: https://github.com/hubmapconsortium/ccf-ui/tree/main/projects/ccf-database
- SPARQL Endpoint: https://ccf-api.hubmapconsortium.org/#/operations/sparql-post
- Published Python, TypeScript, JavaScript, and Angular libraries are available via PyPi and NPM respectively

Outreach/Training

https://expand.iu.edu/browse/sice/cns/courses/hubmap-visible-human-mooc

HuBMAP Visible Human MOOC (VHMOOC)

Started Aug 4, 2020

GO TO CANVAS COURSE

You are enrolled.

INDIANA UNIVERSITY

Course Introduction

This 10h course introduces the HuBMAP project which aims to create an open, global reference atlas of the human body at the cellular level. Among others, the course describes the compilation and coverage of HuBMAP data, demonstrates new single-cell analysis and mapping techniques, and introduces major features of the HuBMAP portal.

Delivered entirely online, all coursework can be completed asynchronously to fit busy schedules. If you have questions or experience issues during registration, please email cnscntr@indiana.edu.

Learning Outcomes

- Theoretical and practical understanding of different single-cell tissue analysis techniques.
- Expertise in single-cell data harmonization used to federate data from different individuals analyzed using different technologies in diverse labs.
- Hands-on skills in the design and usage of semantic ontologies that describe human anatomy, cell types, and biomarkers (e.g., marker genes or proteins).
- Knowledge on the design and usage of a semantically annotated three-dimensional reference system for the healthy human body.
- An understanding of how the HuBMAP reference atlas might be used to understand human health but also to diagnose and treat disease.

Module Topics Include

- HuBMAP Overview: Project Goals, Setup, and Ambitions
- Tissue Data Acquisition and Analysis
- Biomolecular Data Harmonization
- Ontology, 3D Reference Objects, and User Interfaces
- HuBMAP Portal Design and Usage

Meet the Instructors

Katy Börner, Victor H. Yngve Distinguished Professor of Engineering and Information Science. Founding Director of the Cyberinfrastructure for Network Science Center at Indiana University.

Ellen M. Quardokus, staff in the Chemistry Department and research scientist, Cyberinfrastructure for Network Science Center, SICE with expertise in molecular biology, microscopy, anatomy, and interdisciplinary communication.

Andreas Bueckle, PhD Candidate in Information Science, performing research on information visualization, specifically virtual and augmented reality.

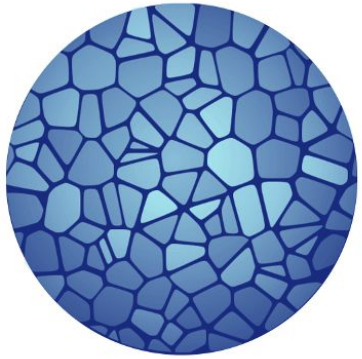
Length: 10 hours

Department: Cyberinfrastructure Network Science

Credit: None

Audience: Biomedical students and professionals interested in single-cell tissue analysis and visualization

Sarah Teichmann



HUMAN
CELL
ATLAS

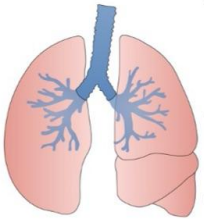
MISSION: To create a comprehensive reference map of the types and properties of all human cells, the fundamental unit of life, as a basis for understanding, diagnosing, monitoring, and treating health and disease

Teichmann lab adventures in organ mapping

Maternal–fetal interface[1]

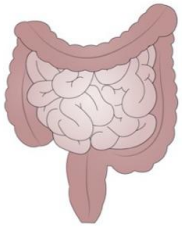
- Sampled across the placenta, plus matched maternal blood and decidua
- First human organ mapped at scale
- CellPhoneDB for cell-cell communication inference

Lung[2,3]



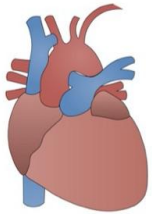
- Sampled nose/bronchi/ parenchyma, compared with spleen to identify resident T cells
- Extended multiomic study identified new immune niche
- Assayed cell changes in asthma

Gut[4,5,6]



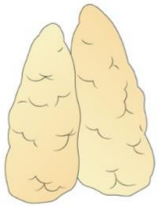
- Sampled across the GI tract, through development and inc microbiota
- Identified secondary lymphoid tissue development
- Assayed cell changes in Crohn's disease

Heart & skeletal muscle[7,8]



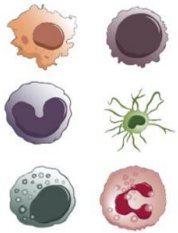
- Sampled across the heart
- Compared to skeletal muscle and kidney to identify new states and niches
- Assayed muscle ageing

Thymus[9]



- Thymus assayed in embryonic, foetal, paediatric and adult stages
- T cell developmental pathways elucidated

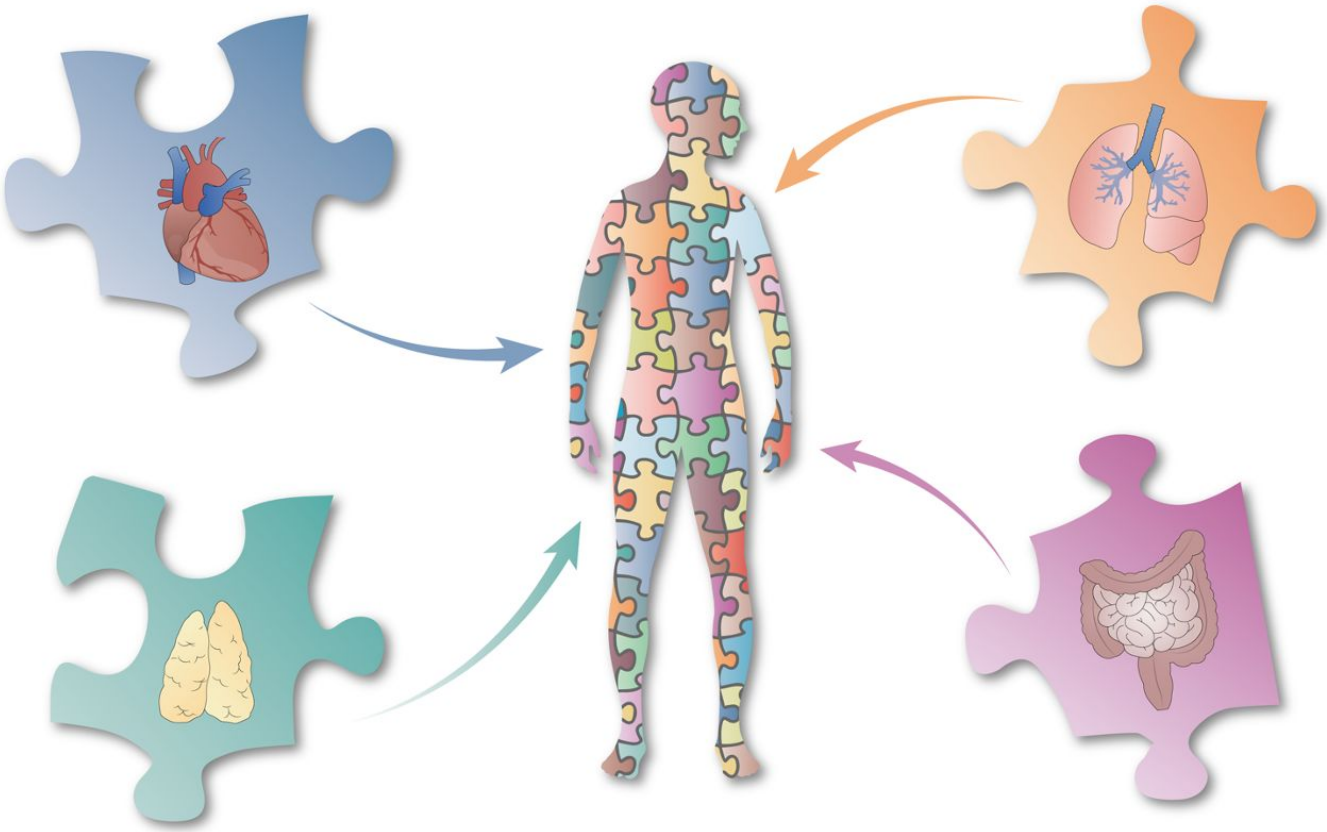
Cross-tissue immunity in embryo and adult[10,11]



- Cross-tissue characterisation of immune cell types/ states
- ~1 million cells surveyed in development
- Creating a 'cell type encyclopaedia' and automated annotation with CellTypist

Figure 1. Cell mapping of human organs from the Teichmann Lab. References: (1) Vento-Tormo 2018 *Nature.* (2) Vieira Braga 2019 *Nat Med.* (3) Madissoon, Oliver 2022 *bioRxiv.* (4) James 2020 *Nat Immunol* (5) Elmentaite 2020 *Dev Cell* (6) Elmentaite 2021 *Dev Cell* (7) Litviňuková 2020 *Nature.* (8) Kedlian 2022 *bioRxiv.* (9) Park 2020 *Science* (10) Suo, Dann 2022 *bioRxiv* (11) Dominguez-Conde, Xu, Jarvis 2022 *Science*

We want to put the pieces together for a full body atlas



Wish list

- 4D mapping: incorporating changes in time at short (physiological) and long (development / ageing) scales

- True 3D spatial transcriptomics - currently limited to slices. Larger samples in x, y and z. Cell/nucleus-level resolution

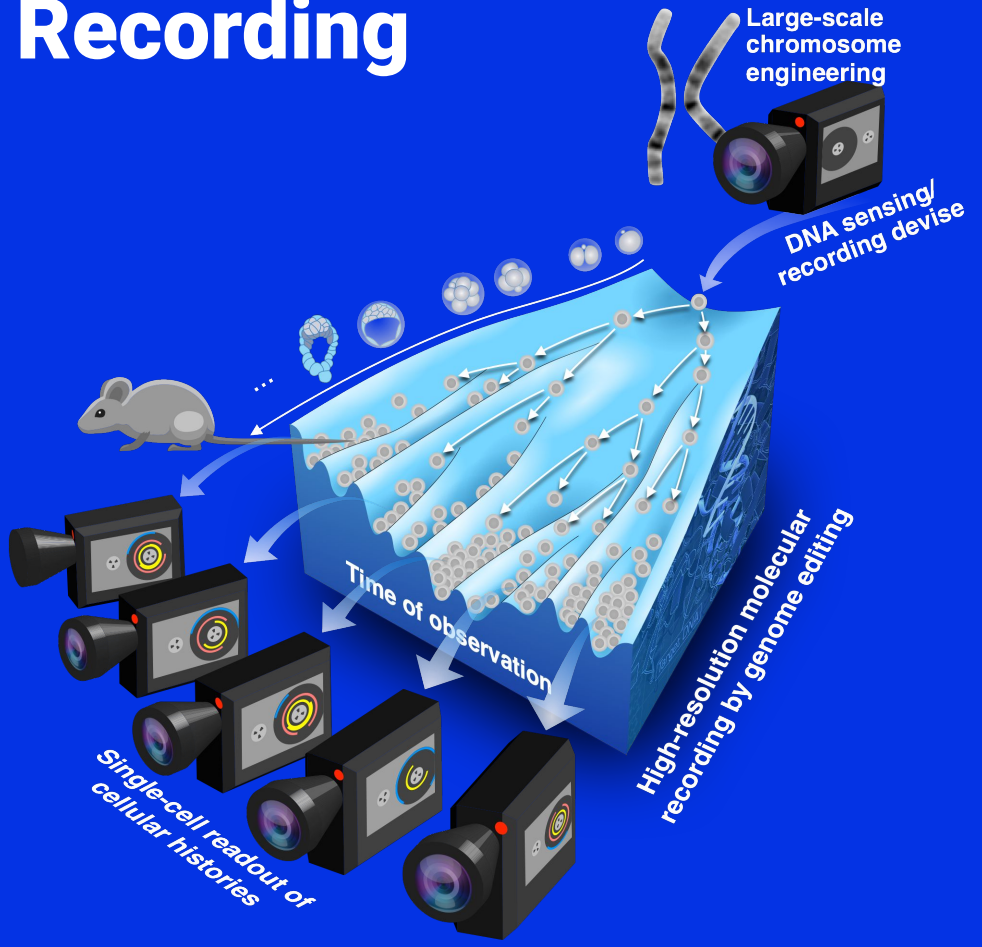
- Effective integration of data from spatial and single cell modalities (e.g. cell2location)

- Capturing cell characteristics (e.g. morphology) and linking this to molecular identity

- Going from molecular > cellular > tissue > organ scale

Nozomu Yachie

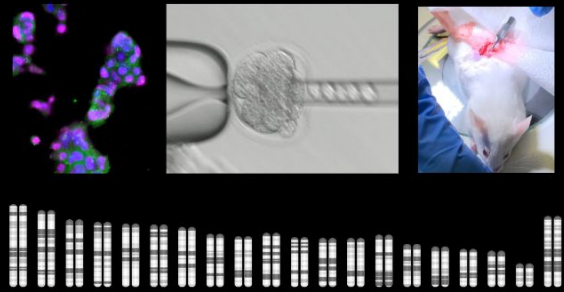
DNA Event Recording



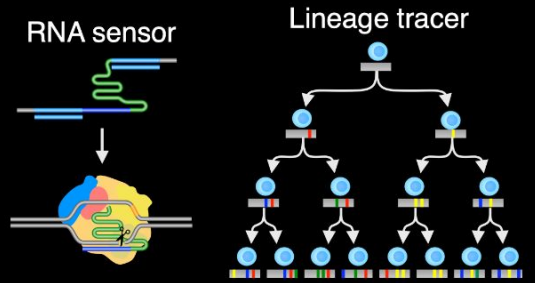
Four pillars for DNA Event Recording

Molecular Systems Biology 2016
Curr Opin in Chem Biol 2019a & b

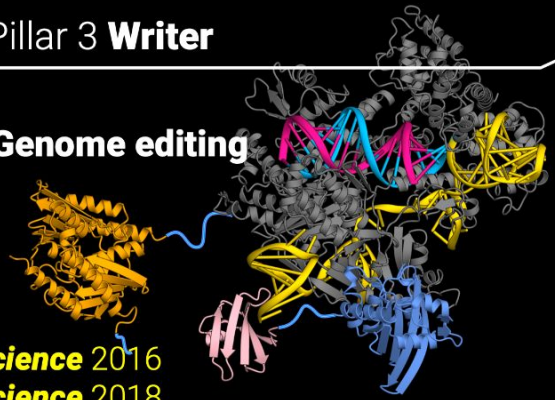
Pillar 1 **Memory**



Pillar 2 **Sensor**

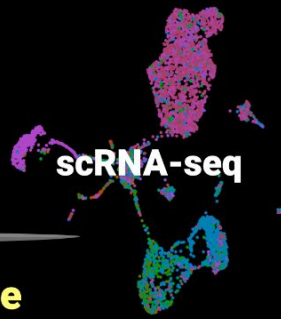


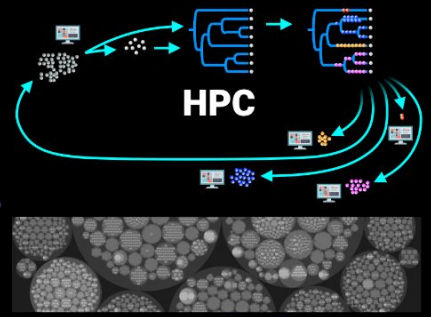
Pillar 3 **Writer**



Science 2016
Science 2018
Nature Biotechnology 2020

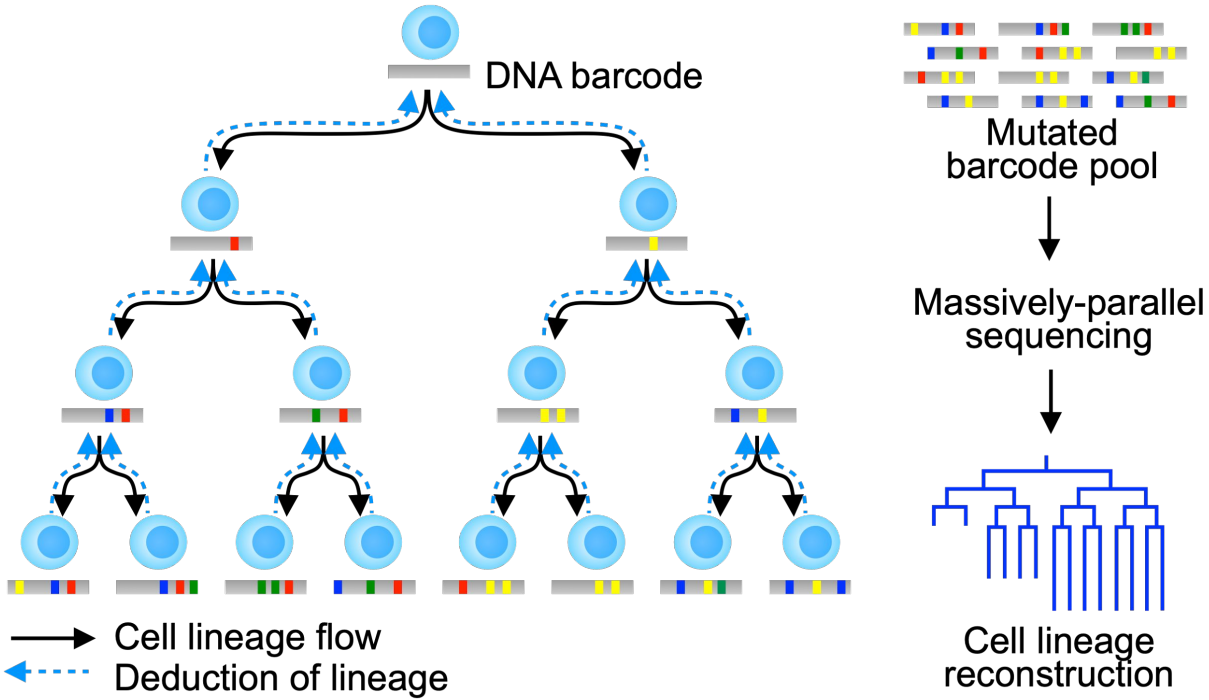
Pillar 4 **Reader**

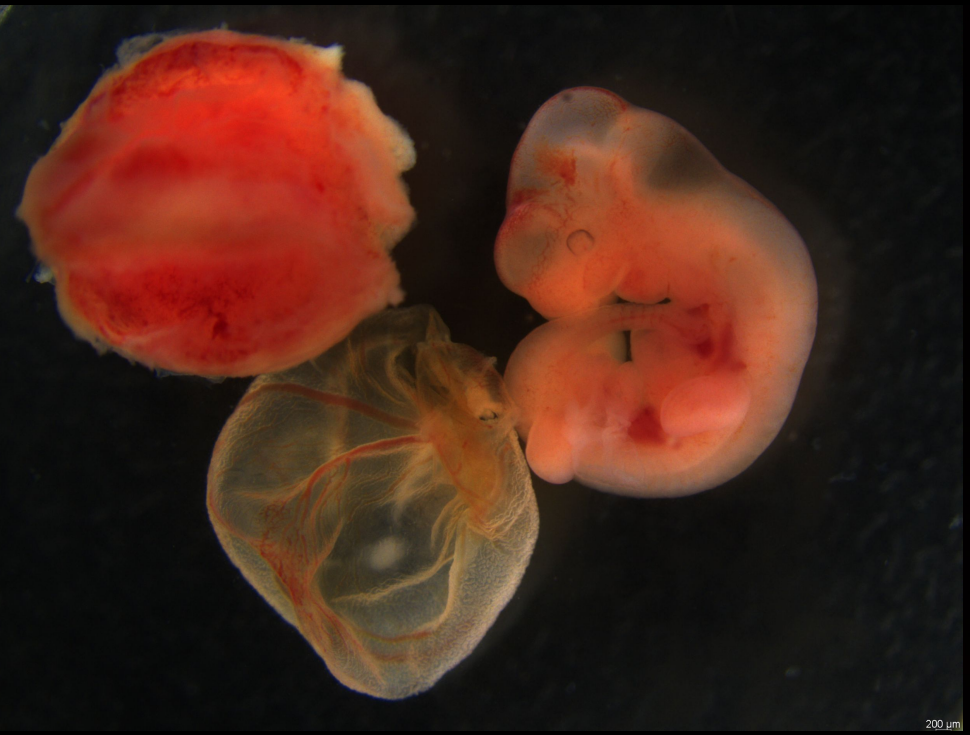


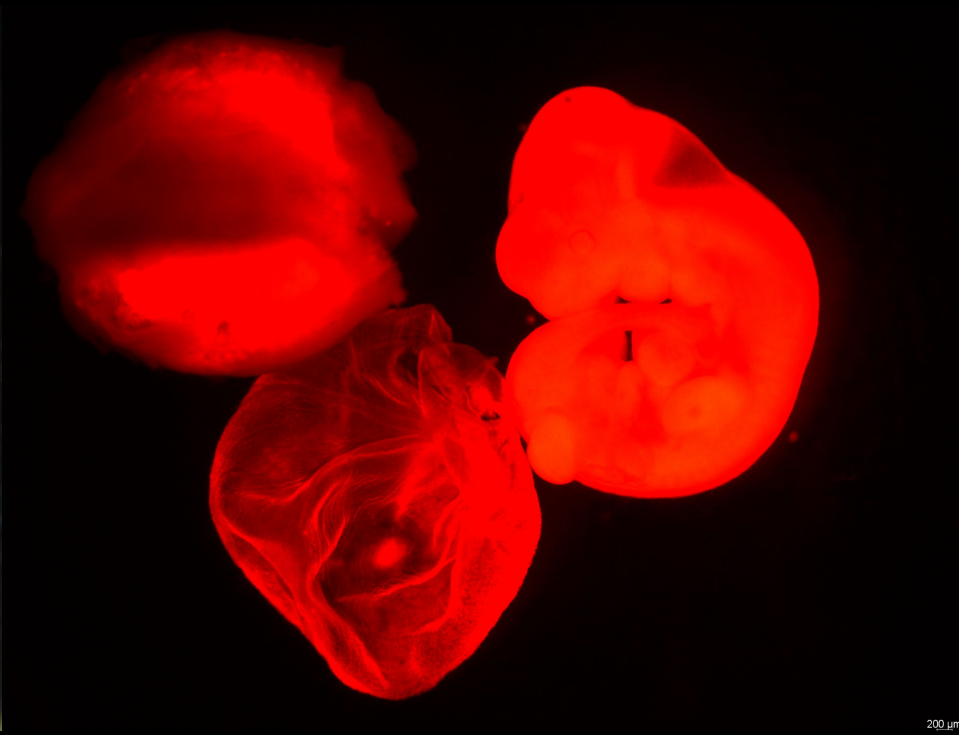


Nature Biotechnology 2022

Lineage tracing by evolving DNA barcodes

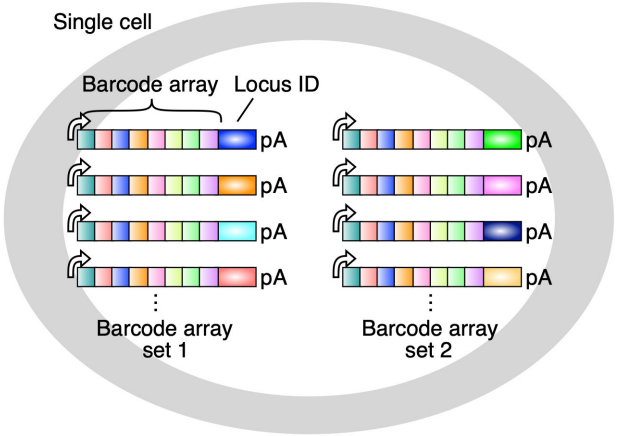


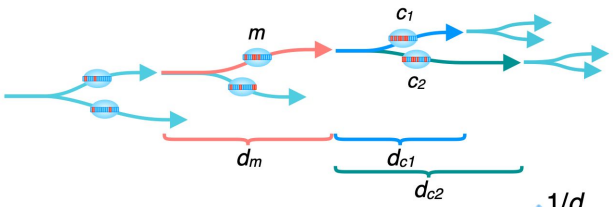


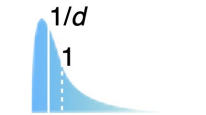


With Masa Ema@Shiga Med & Seiya Mizuno@Tsukuba U

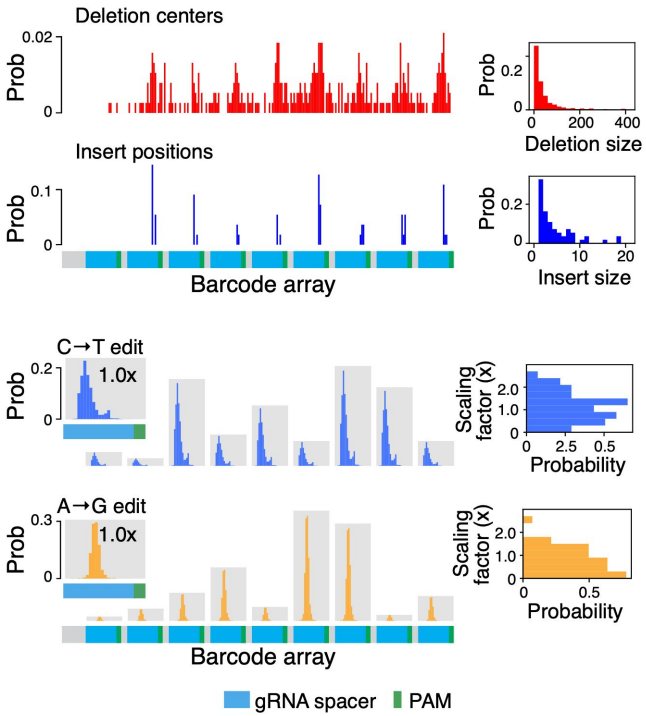
Simulating high-content cell lineage tracing

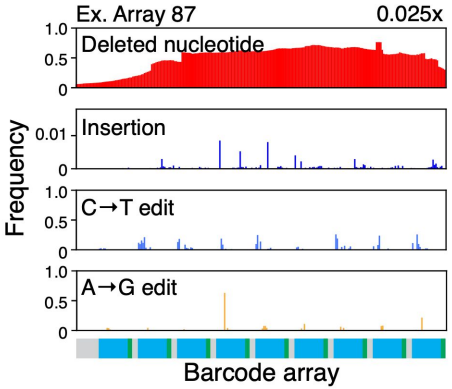






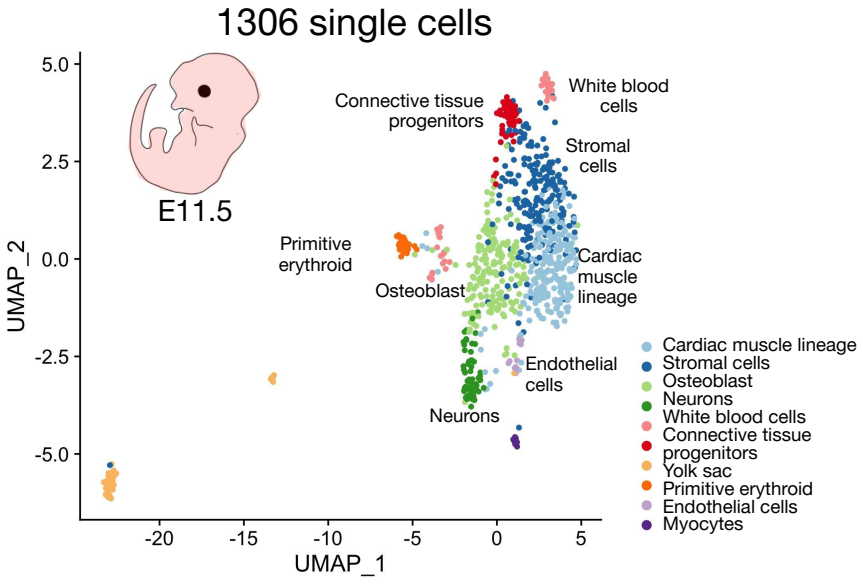
$1/d \sim \Gamma(1/\sigma^2, 1/\sigma^2)$

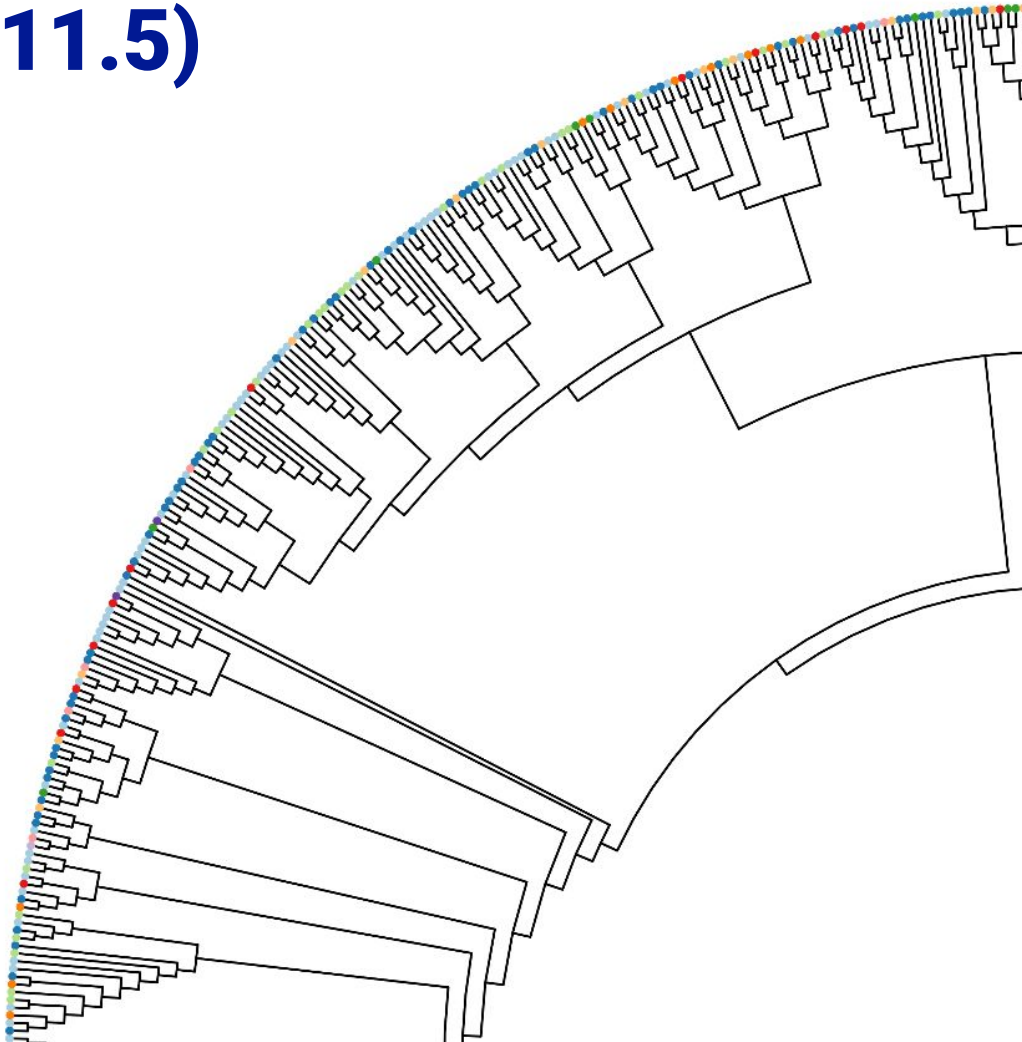




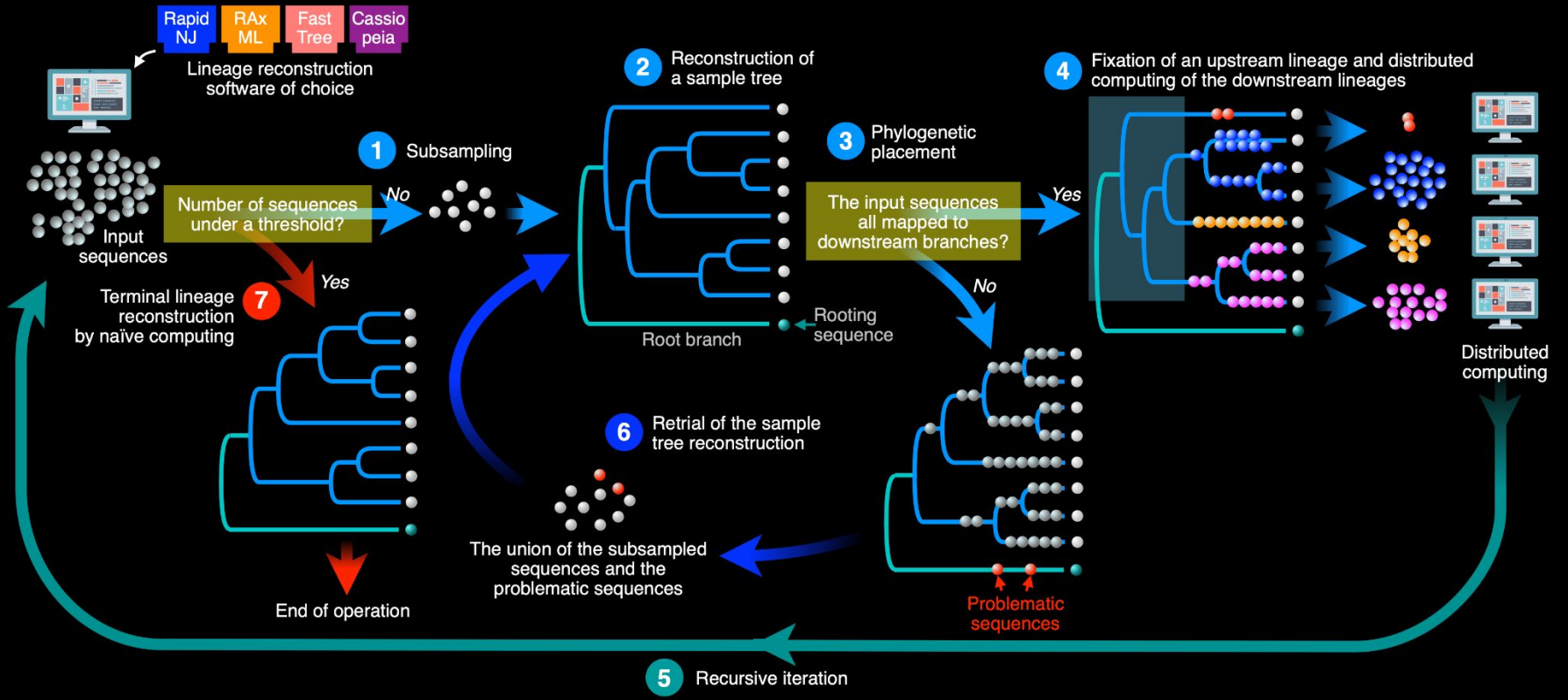
Konno+ 2022 ***Nature Biotechnology***

A very pilot lineage (E11.5)



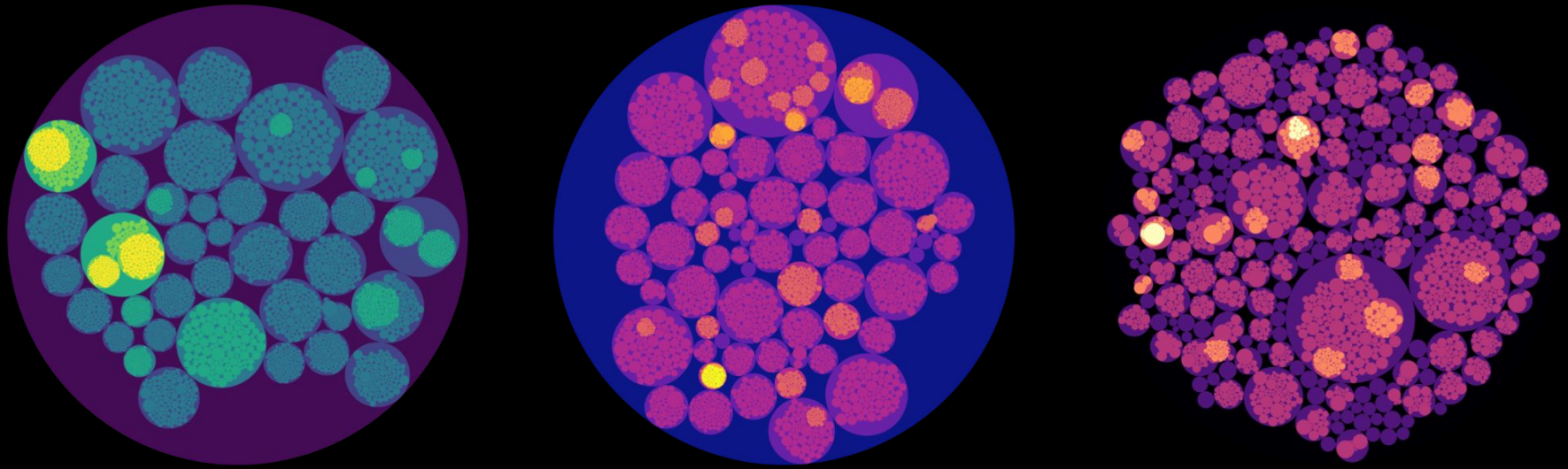


Deep distributed computing framework



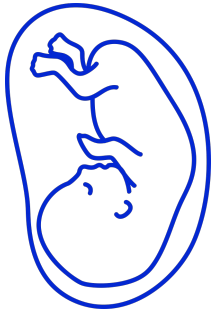
Konno+ 2022 ***Nature Biotechnology***

Lineage estimation of 235 million sequences



Konno+ 2022 ***Nature Biotechnology***

The Human Embryo Simulator Consortium



An international ECR/MCR consortium.

The Human Embryo Simulator will have the ability to predict the outcome of multicellular developmental systems and will enable in silico experimentations of human development, revolutionizing stem cell and developmental biology.

The consortium will serve as a virtual, large, intellectual sandbox for international ECR/MCR leaders and their groups to understand and build biology together with this vision.

Leads
Nika Shakiba (UBC), Nozomu Yachie (UBC), Maria Abou Chakra (UToronto)

Advisory Board Members
Gary Bader (UToronto), Peter Zandstra (UBC), Magdalena Zernicka-Goetz (Caltech), James Glazier (Indiana U), Adriana Dawes (OSU), Hiroaki Kitano (Sony)

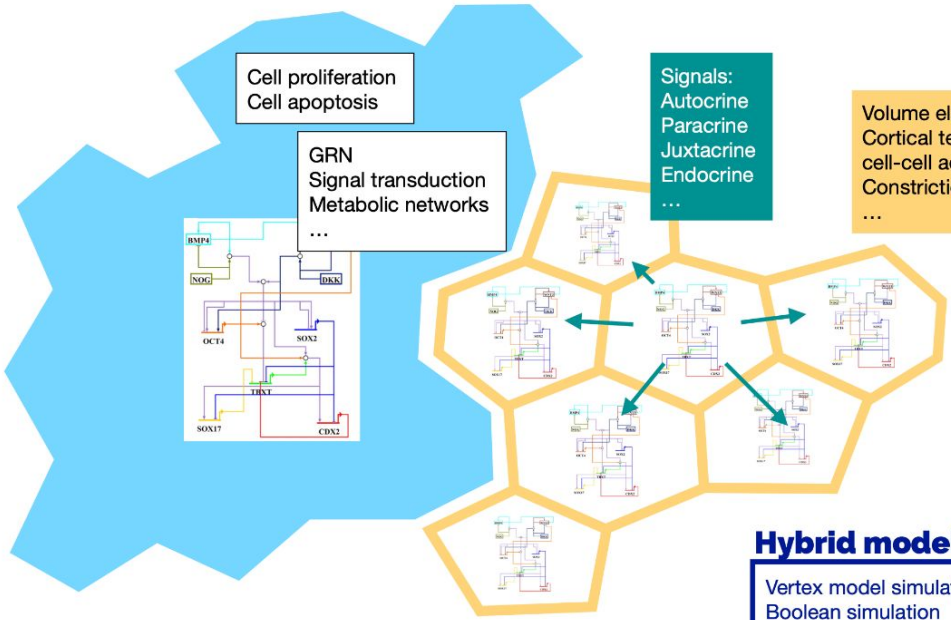
Cluster 1: Molecular and cellular systems
Experimental: Nozomu Yachie (UBC)*, Knut Woltjen (Kyoto U), Carl de Boer (UBC), Katie Galloway (MIT)
Computational: Geoffrey Schiebinger (UBC), Verônica A. Grieneisen (UCardiff), Adam MacLean (USC)

Cluster 2: Cell populations
Experimental: Nika Shakiba (UBC)*, Ivana Barbaric (USheffield), Samer Hussein (Laval U), Sadao Ota (UTokyo), Leonardo Morsut (UCSD)
Computational: Morgan Craig (UMontreal)*, Linus Schumacher (UEdinburgh), Berta Verd (UOxford)

Cluster 3: Developmental systems (tissues, embryogenesis, whole body, emergence)
Experimental: Jun Wu (UT Southwestern)*, Takanori Takebe (Cincinnati Children's Hospital)*, Miki Ebisuya (EMBL Barcelona)*, Naoki Irie (UTokyo), Mo Ebrahimkhani (UPitt)
Computational: Ruth Baker (UOxford)*, Guillaume Blin (UEdinburgh), Dagmar Iber (ETH Zurich), Ruben Perez-Carrasco (ICL), Alex Fletcher (USheffield)

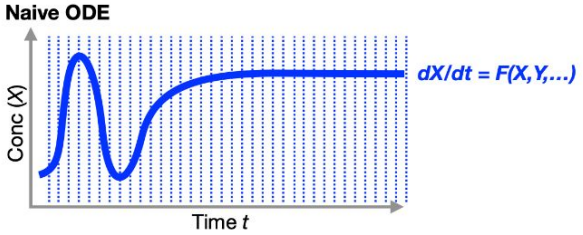
Cluster 4: Integrative simulation platforms
Maria Abou Chakra (UToronto)*, Alex Fletcher (USheffield)*, Satoru Okuda (Kanazawa U), Lutz Brusch (UDresden), Sidhartha Goyal (UToronto), Noemi Picco (USwansea)

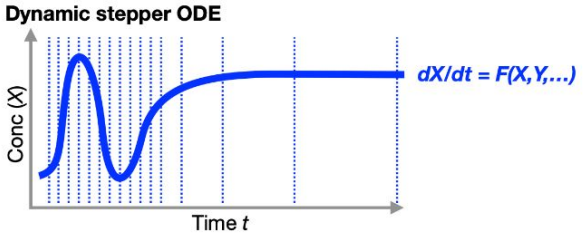
Cluster 5: Ethics, legal and social implications of embryo simulation
Vardit Ravitsky (UToronto)



Hybrid modeling platform

Vertex model simulation
Boolean simulation
ODE simulation
Reaction diffusion simulation
Fluid mechanics simulation
…





Multi-scape dynamic stepper modeling

QM/MM in molecular dynamics simulation
Mutiscale stepper in ODE simulation of metabolic networks
…

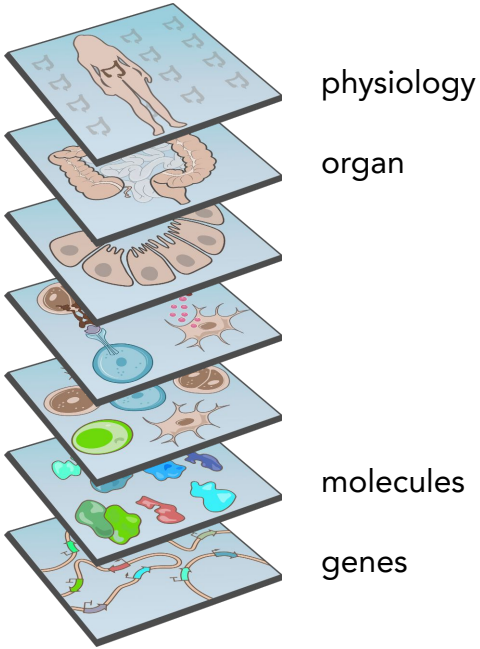
Aviv Regev

Note: As I will be able to attend only some of the sessions (I am solo with the kids this week!), I tried to include more material here, which I hope will be useful.

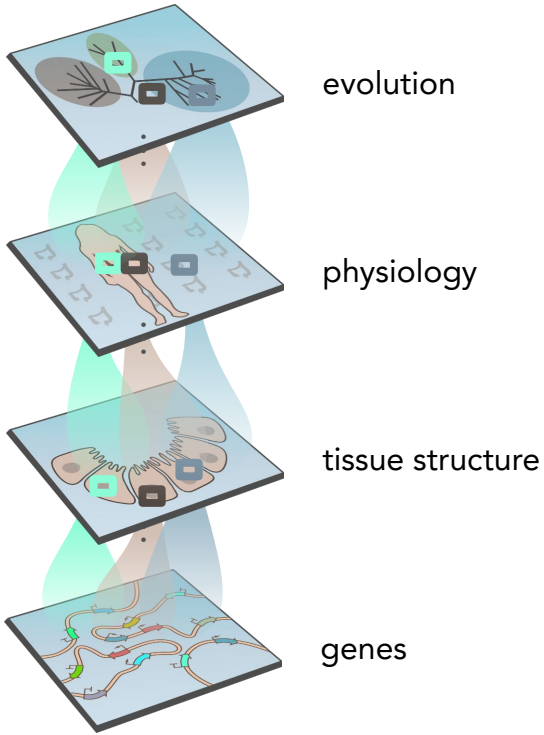
Learning the convolutions of the human body

Aviv Regev

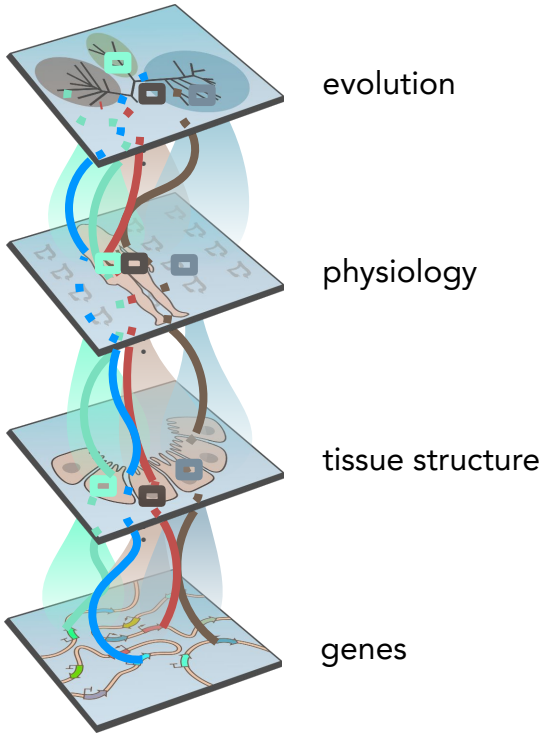
The many “convolutions” of biology



Models can map between the levels

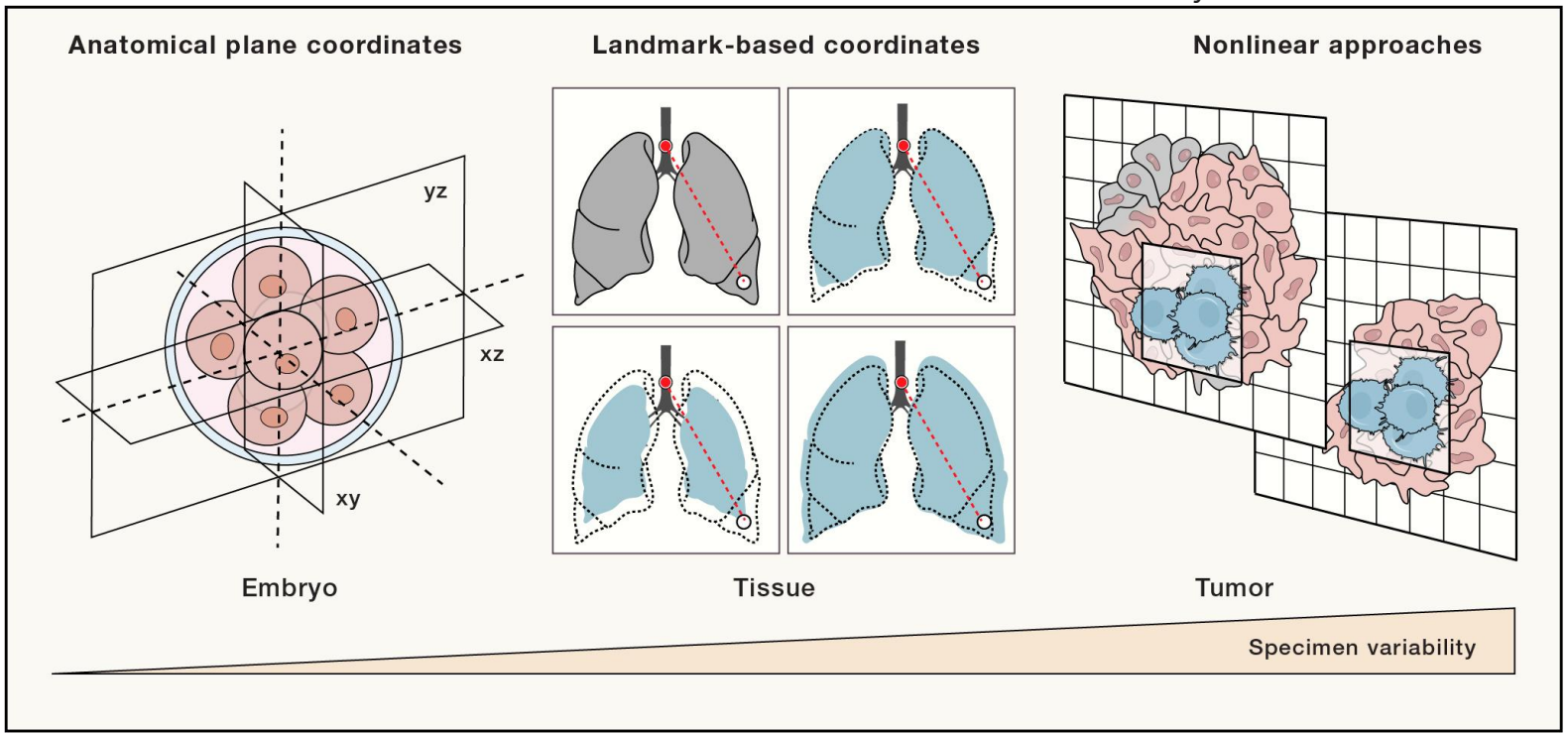


Models can decipher the mechanisms underlying the mapping



Goal: Build an atlas from cells to organs to body

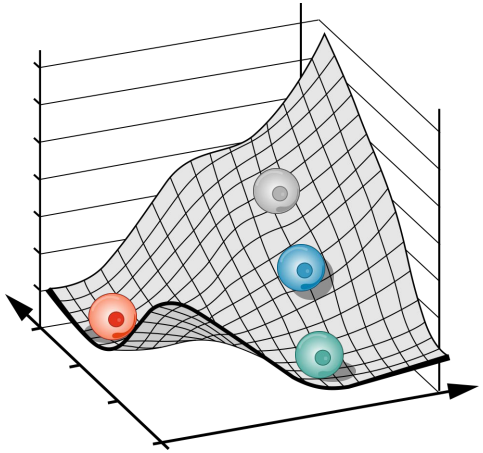
Toward a Common Coordinate Framework for the Human Body



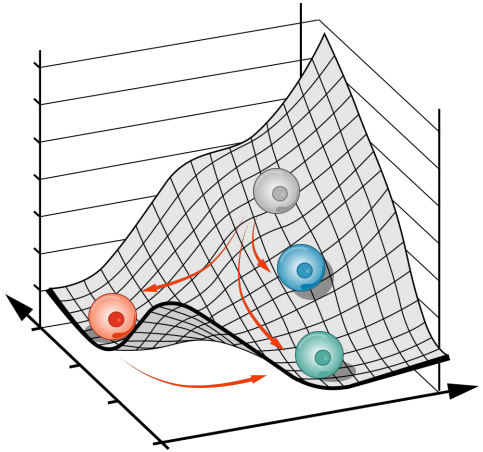
Rood et al., *Cell* 2019: With John Marioni and Rahul Satija labs

What is a Human Cell Atlas 1.0? Concepts

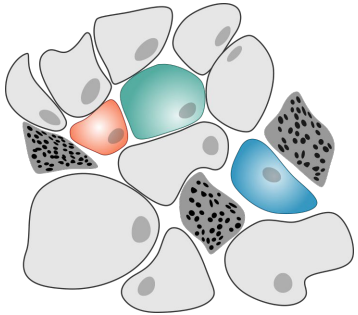
types / states



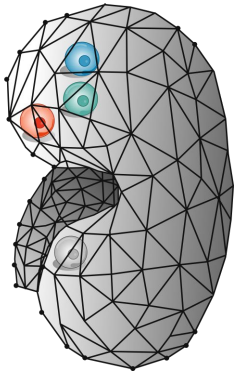
trajectories/transitions



histological modules

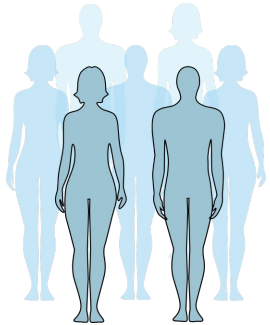


positions

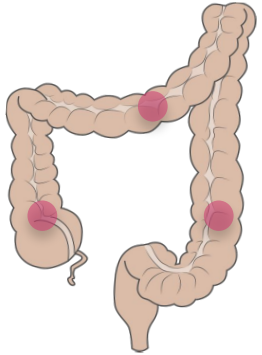


Collection roadmap for the atlas

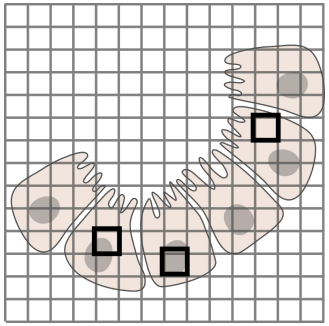
No of individuals



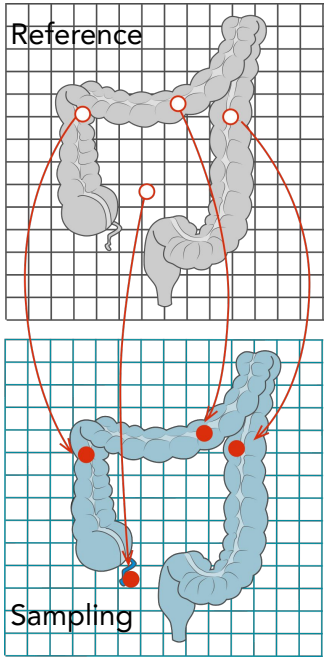
Anatomical sampling



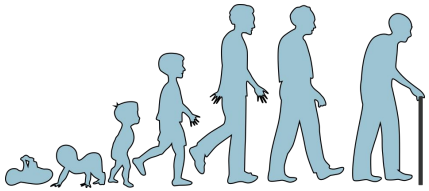
Number of regions



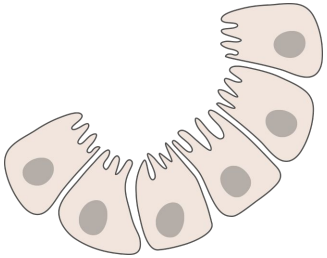
Coordinate framework



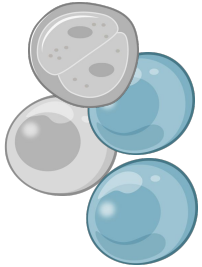
Development and aging



Histological sampling

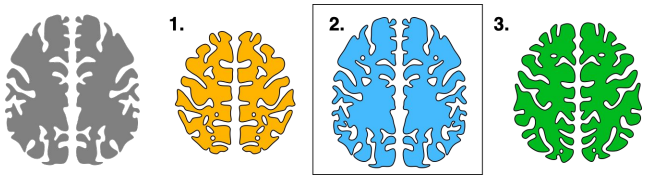


Number of cells

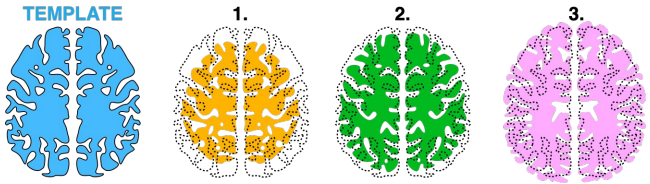


Methods to build a Common Coordinate Framework for HCA

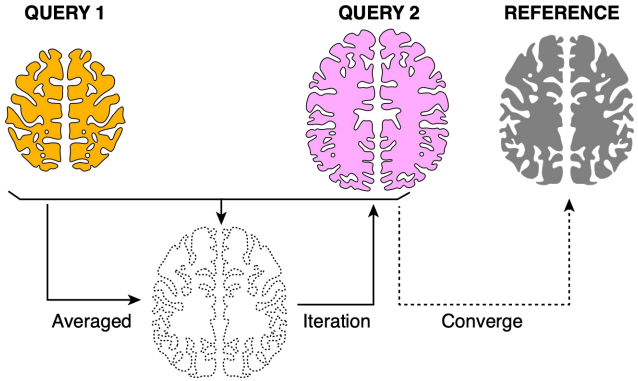
1. Calculate best template



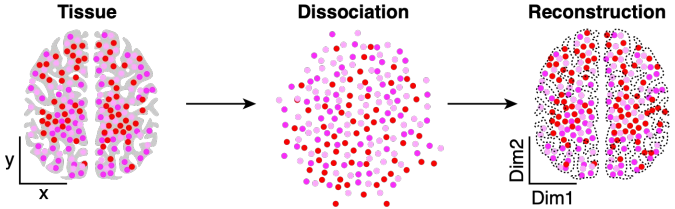
2. Approach 1: Map all to one template



3. Approach 2: Iteratively align and average



4. Reconstructing an atlas from its features

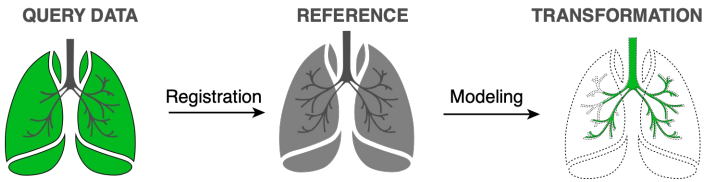


Support: HubMAP HIVE, BRAIN Initiative

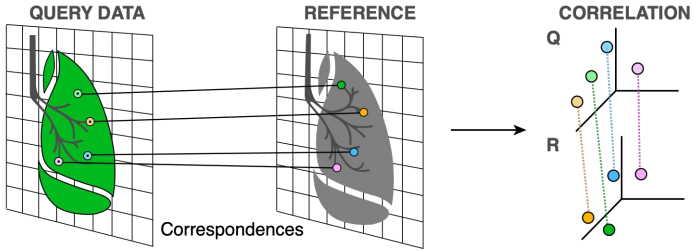
Satija, Regev, Marioni

Mapping new datasets to a CCF reference

Within modality mapping

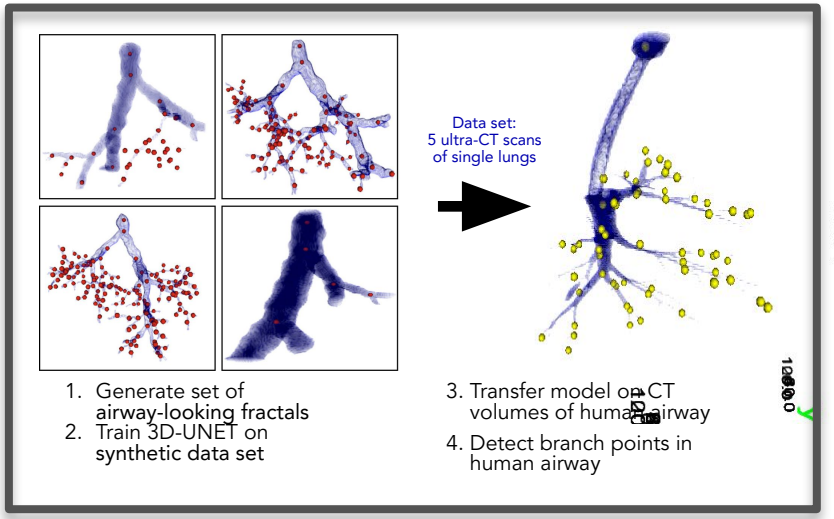


Cross modality mapping



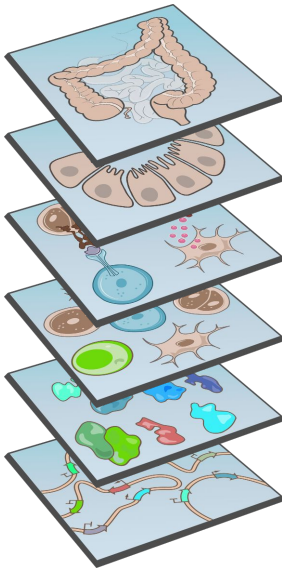
Satija, Regev, Marioni

Deep learning model IDs lunch branch points
Biancalani, Heimberg, Regev



Support: HubMAP HIVE, BRAIN Initiative

Algorithms are essential for building and using atlases

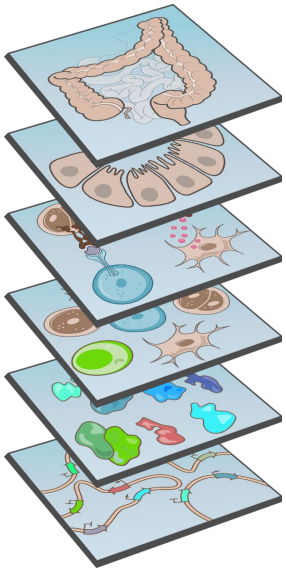


I. Integrate, enhance and construct atlas

II. Define new experimental modalities

III. Enable new biological discoveries and concepts

Algorithms are essential for building and using atlases

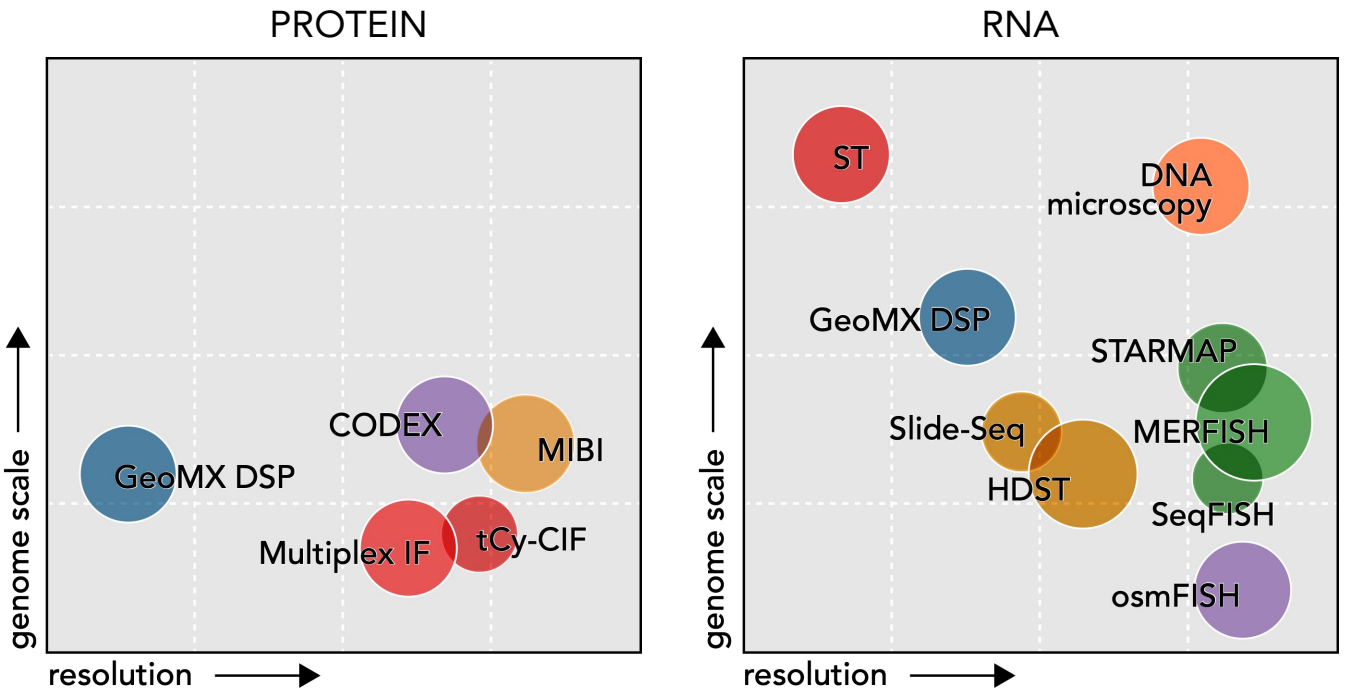


I. Integrate, enhance and construct atlas

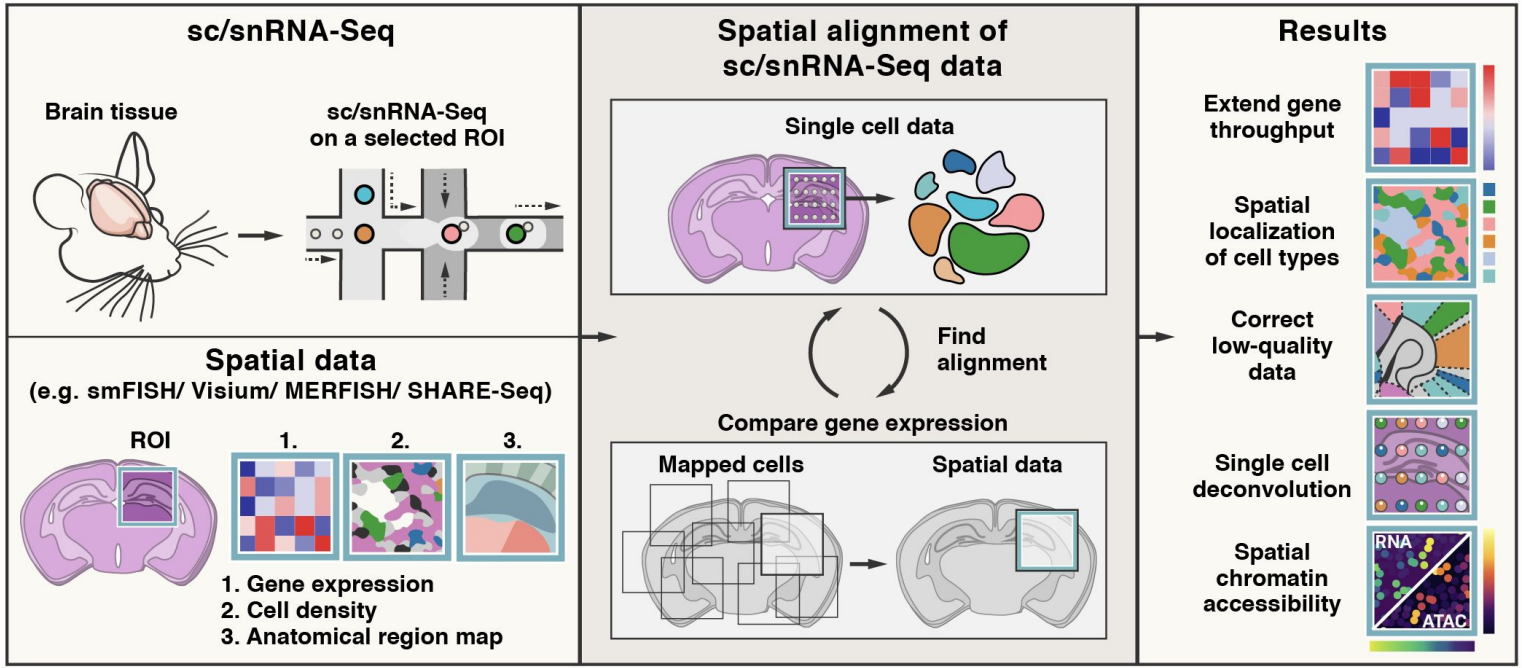
II. Define new experimental modalities

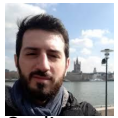
III. Enable new biological discoveries and concepts

Integration addresses limitations of measurement methods



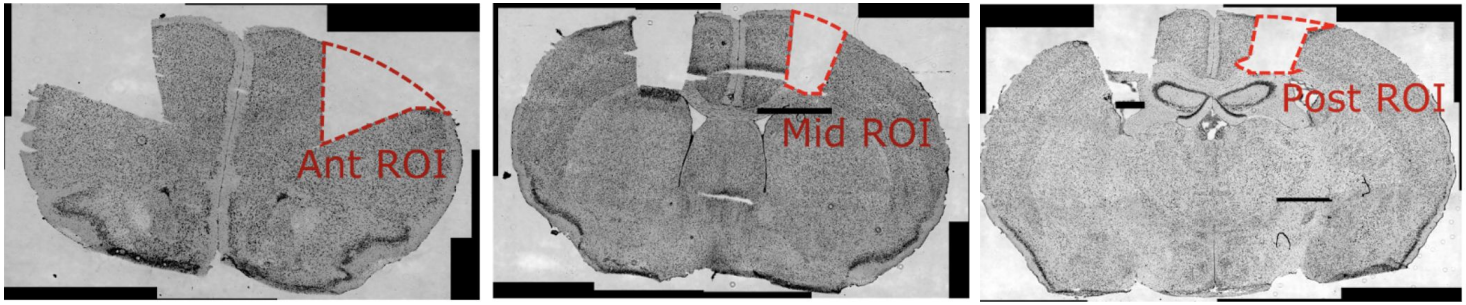
Atlas integration: molecules, cells, histology and anatomy



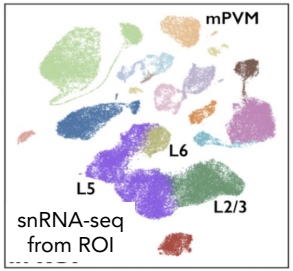



Bianalanni, Scalia *et al.*, *Biorxiv* 2020, *Nature Methods* 2021.

snRNA-Seq data is derived from imaged sections



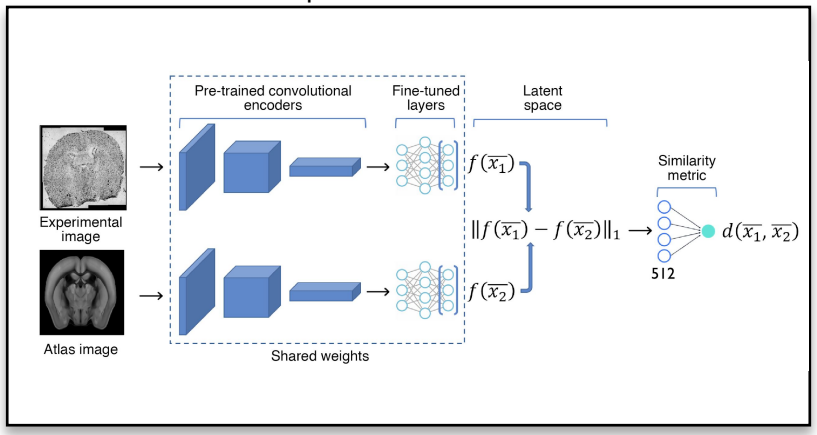
3 ROIs, 160,000 nuclei, 22 cell subsets (Charles Vanderburg, Evan Macosko lab)



Bianalanni, Scalia *et al.*, *Biorxiv* 2020, *Nature Methods* 2021.

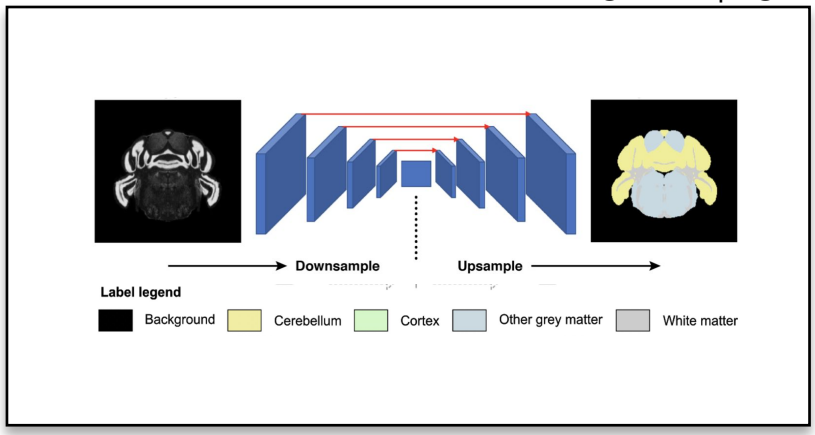
Automated atlas integration with anatomy and histology

Anatomical depth model (Lorenzo Buffoni)



Outcome: learned latent space where the point distance represents the anatomical distance between slices

Anatomical label model (Aman Sanger, Ziqing Lu)

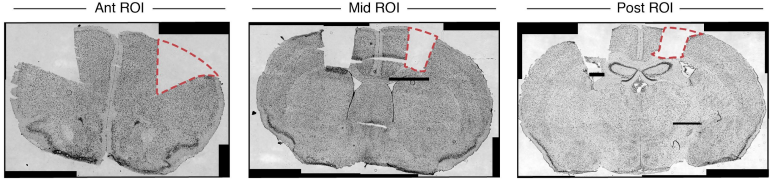


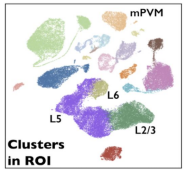
Outcome: A semantic segmentation model to call main anatomical regions

Bianalanni, Scalia *et al.*, *Biorxiv* 2020, *Nature Methods* 2021.

Tangram built anatomical, histological, cellular and molecular atlas of the somatomotor mouse cortex

Histology

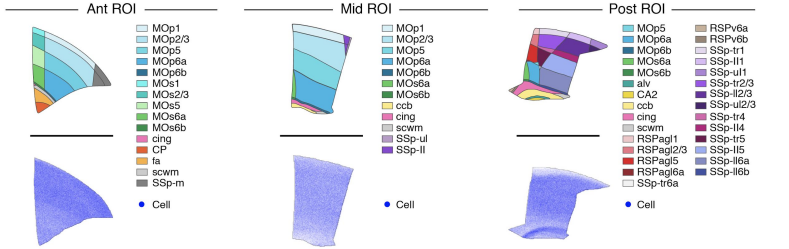




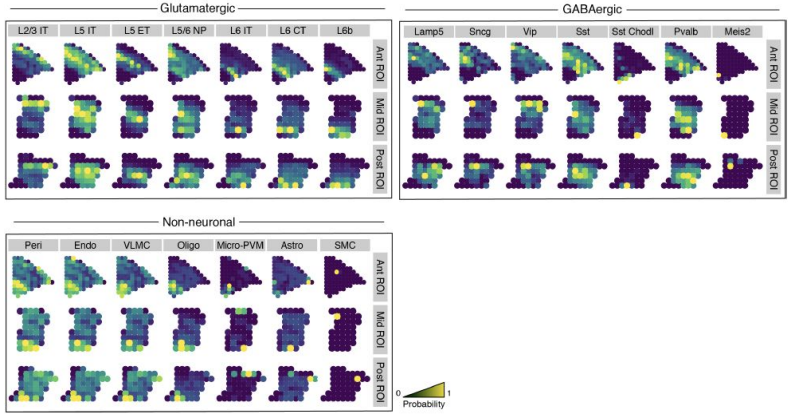
snRNA-Seq

Anatomical region map
(+ Allen CCF3)

Cell density map
(+ Blue Brain Cell Atlas)



Cellular profiles
(+ Allen ISH)



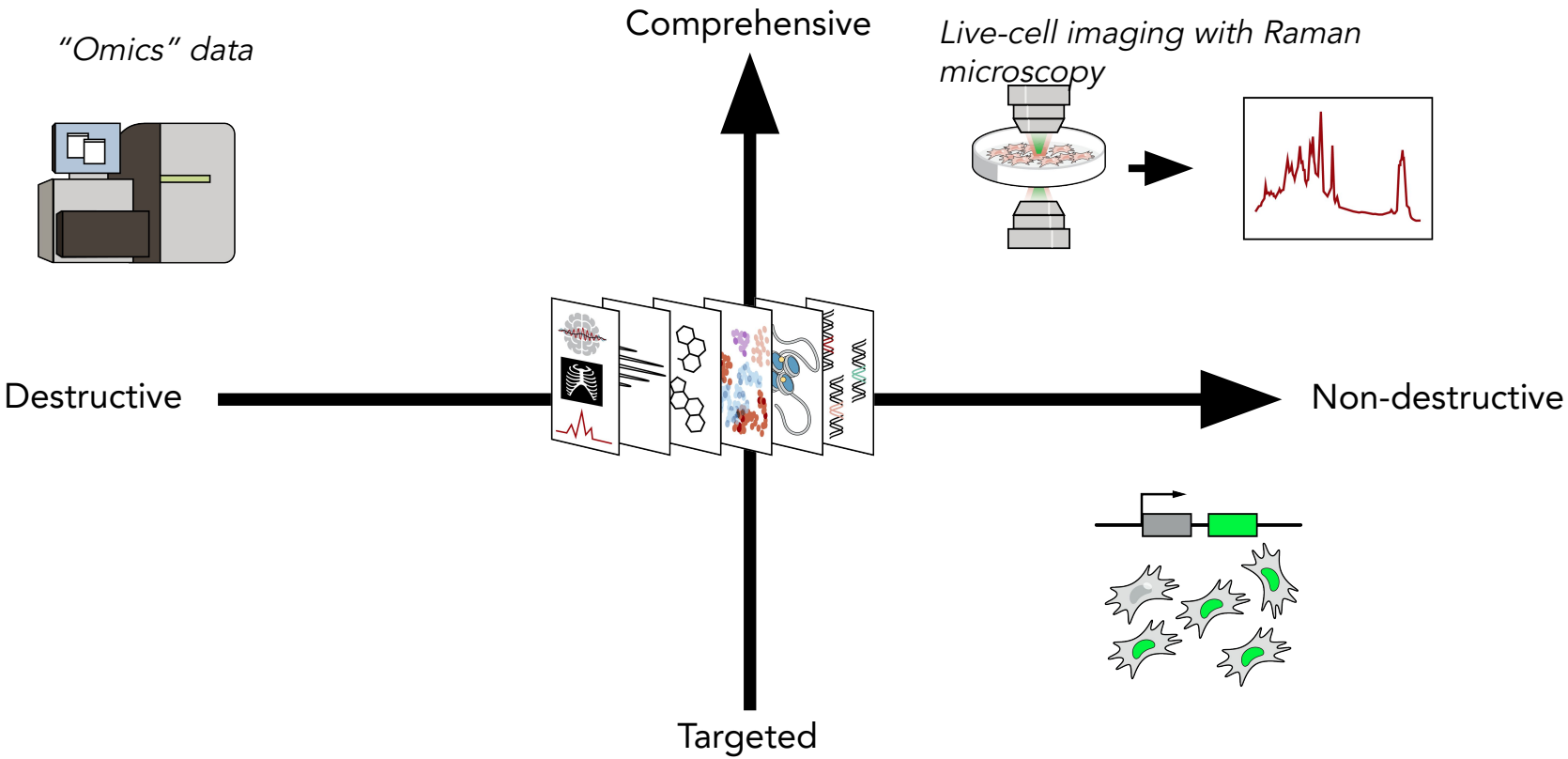
Bianalanni, Scalia *et al.*, *Nature Methods* 2021.

Neriman Tokcan

1. Can we profile live cells and organisms at genomic scale?

2. Can we map even if we have no spatial molecular measurements?

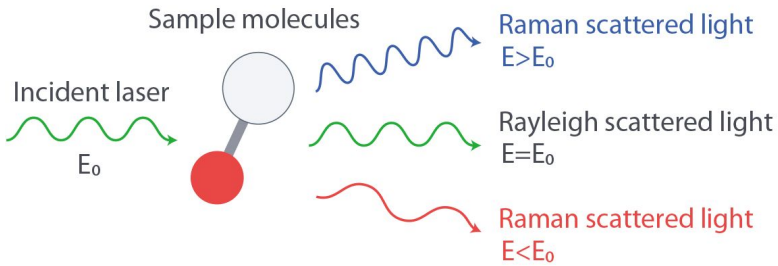
Challenge: comprehensive live profiling of cells and tissues



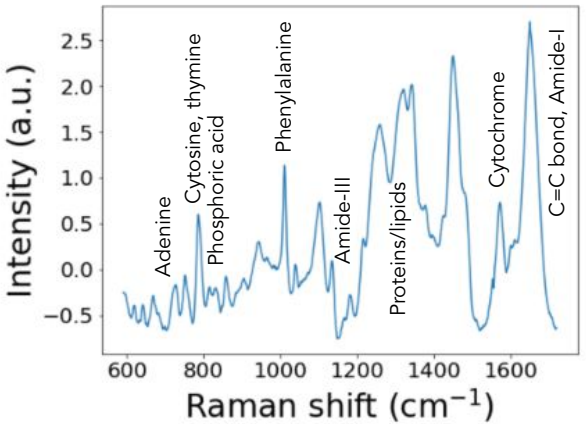
Koseki Kobayashi

Raman microscopy: molecular “fingerprint” of live cells

Label-free, non-destructive measurement of vibrational energy levels of molecules at subcellular spatial resolution in live cells

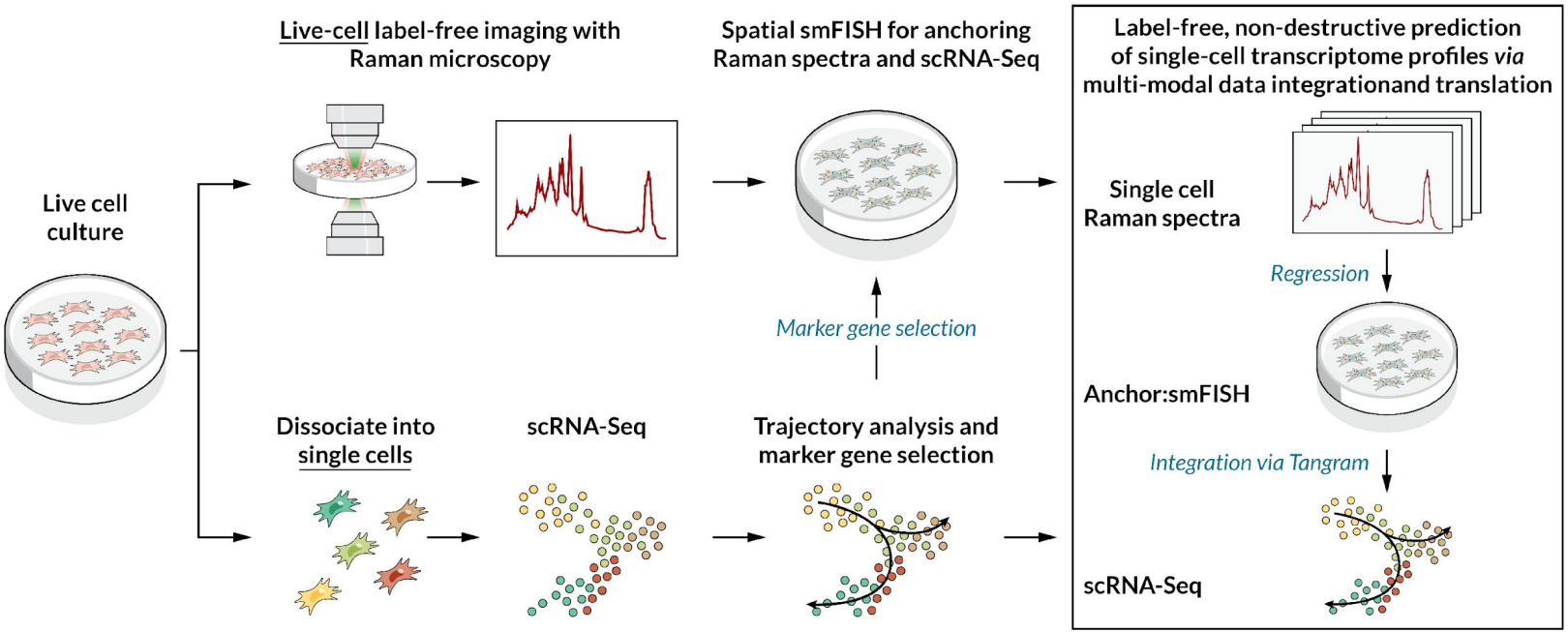


Nanophoton, Inc.



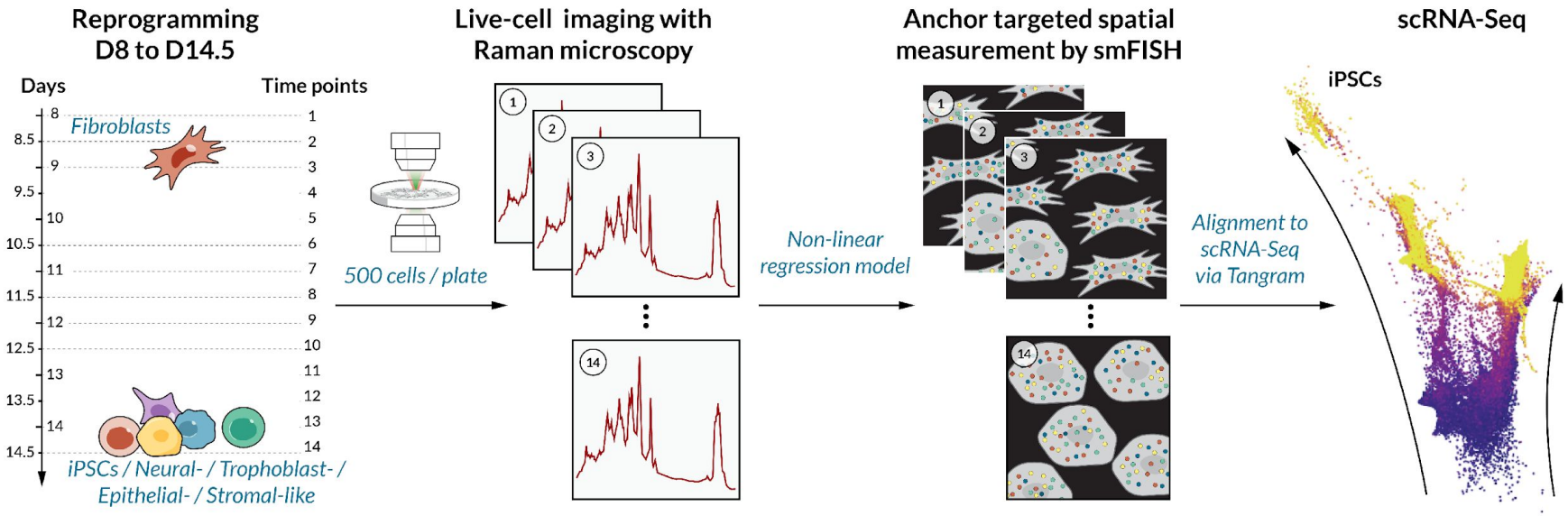
Koseki Kobayashi, in collaboration with Peter So, Tommaso Bianchalani and Jian Shu; Kobayashi et al, Biorxiv 2021

Raman2RNA (R2R): Train models to predict RNA profiles from Raman spectra of live cells



Koseki Kobayashi, in collaboration with Peter So, Tommaso Bianchalani and Jian Shu; Kobayashi et al, Biorxiv 2021

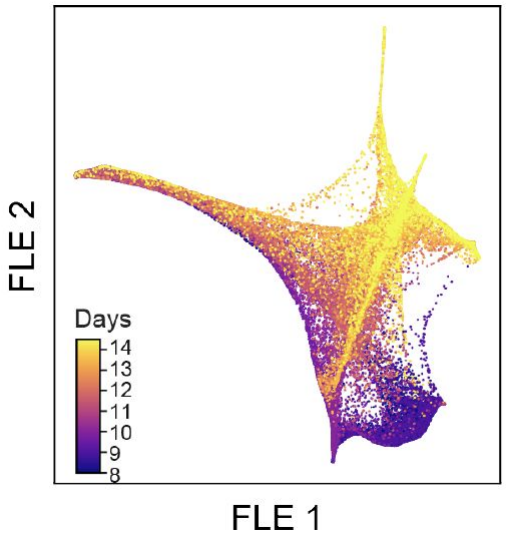
Raman2RNA (R2R) proof of concept in iPSC reprogramming



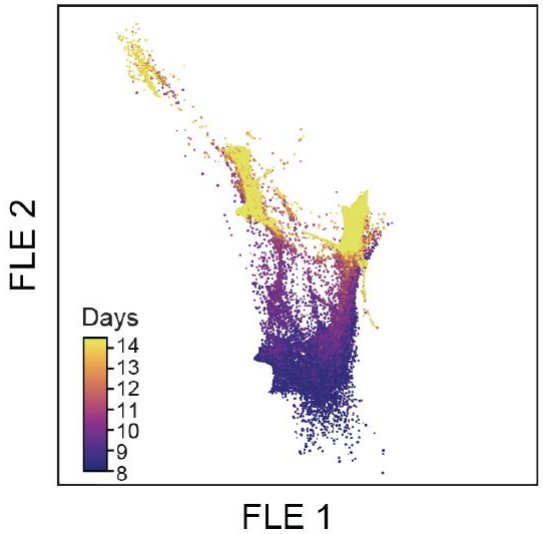
Koseki Kobayashi, in collaboration with Peter So, Tommaso Bianchalani and Jian Shu; Kobayashi et al, Biorxiv 2021

Raman spectra capture temporal progression analogous to RNA

Raman spectra, day 8-14

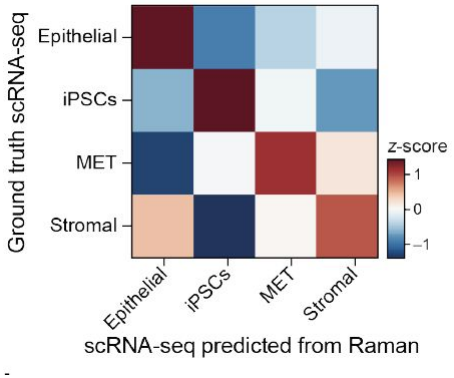


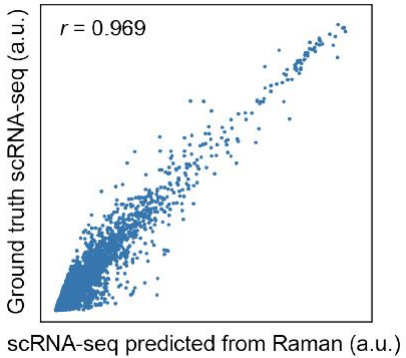
scRNA-Seq, day 8-14

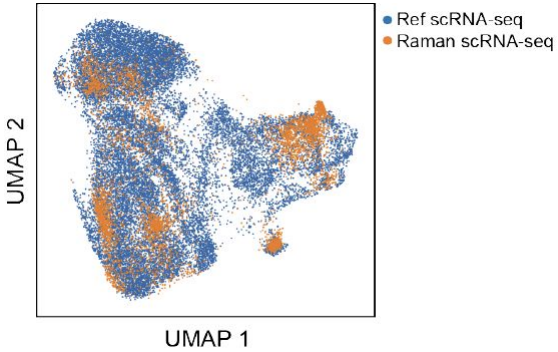


Koseki Kobayashi, in collaboration with Peter So, Tommaso Bianchalani and Jian Shu; Kobayashi et al, Biorxiv 2021

Agreement between single cell profiles measured by scRNA-Seq and predicted from Raman spectra by trained model



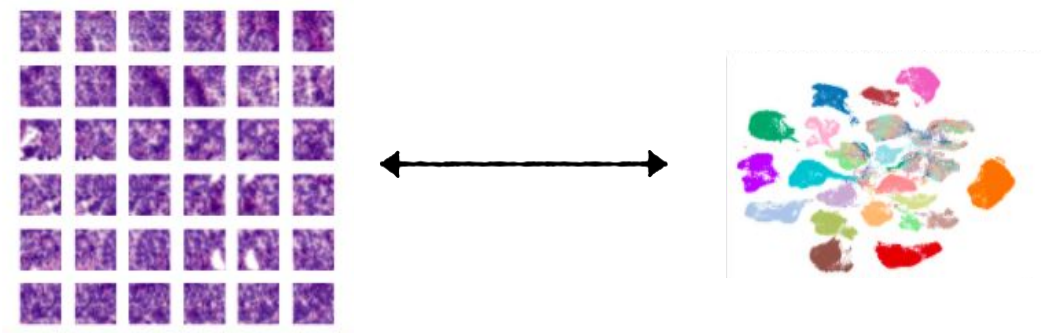




Using brightfield instead of Raman gives poor results

Koseki Kobayashi, in collaboration with Peter So, Tommaso Bianchalani and Jian Shu; Kobayashi et al, Biorxiv 2021

Challenge: relate molecular profiles and histology / cell biology *without* spatial molecular measurements



Charles Comiter, Eeshit Vaishnav

SCHAF: Adversarial autoencoder to generate scRNA profiles from histology without spatial molecular measurements

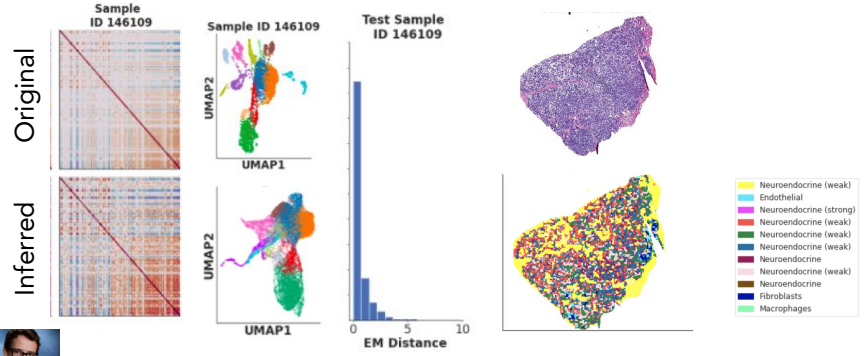
Input: paired histology and single cell profiles (multiple tumors)

Model: Adverserial autoencoder (and within domain normalization)

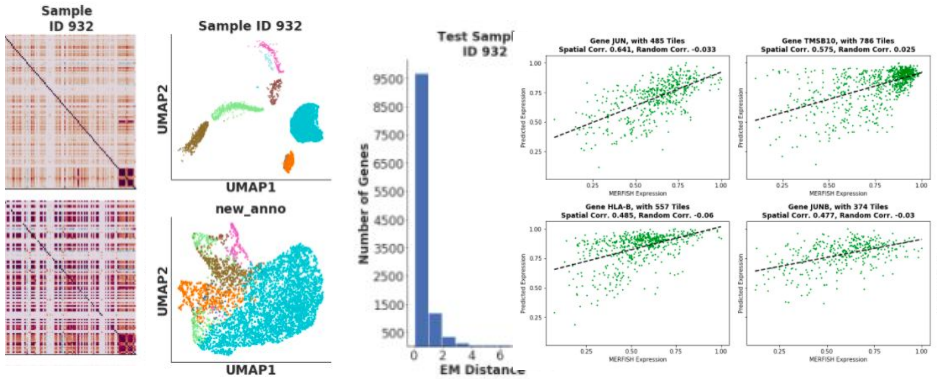
Translation: Encode with source domain encoder (e.g, histology tiles); decode with target domain decoder (eg scRNA-seq)

Result: Predicted, spatial (tiled) scRNA-Seq atlas for an H&E

Squamous Cell Carcinoma

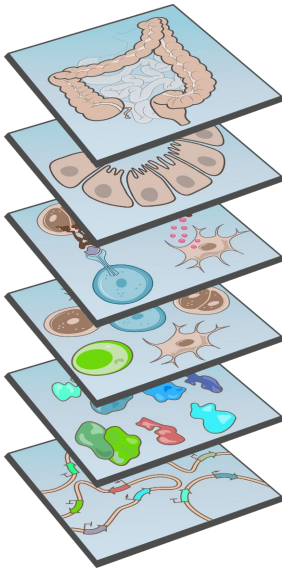


Metastatic Breast Cancer



Charles Comiter, Eeshit Vaishnav

Algorithms are essential for building and using atlases



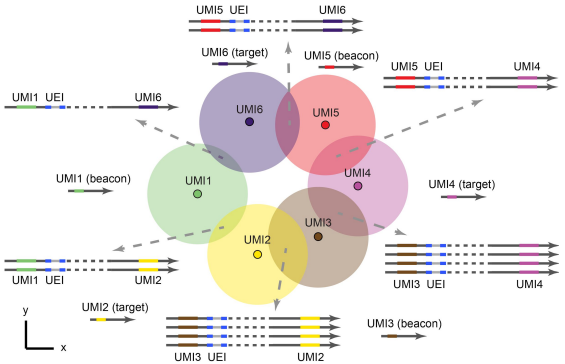
I. Integrate, enhance and construct atlas

II. Define new experimental modalities

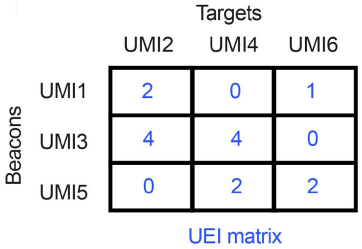
III. Enable new biological discoveries and concepts

DNA microscopy: optics-free imaging by chemical reaction

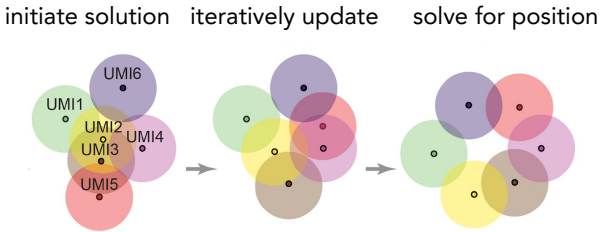
1. Encode proximity by Unique Event Identifiers



2. Sequence



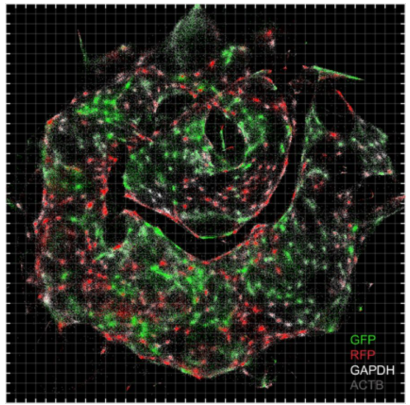
3. Decode image by inference



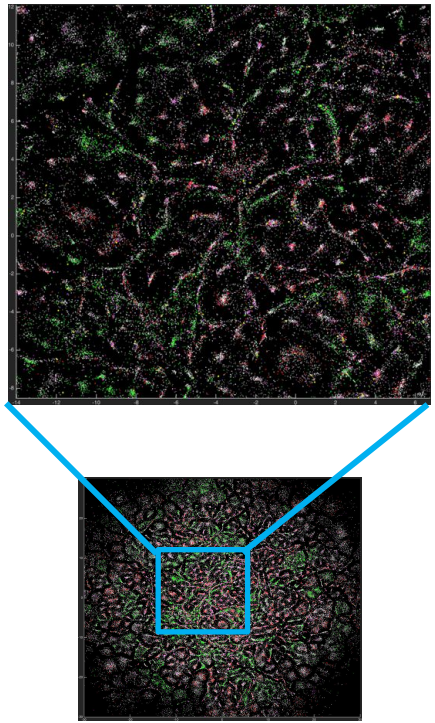
Joshua Weinstein, Aviv Regev and Feng Zhang; *bioRxiv* 2018; *Cell* 2019

DNA microscopy of signatures and whole transcriptomes

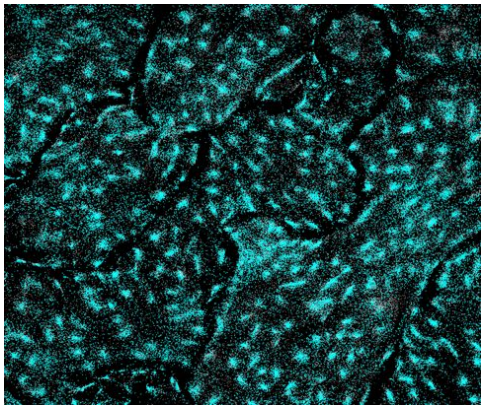
4 genes (GFP and RFP lines co-culture)



20 cell type specific genes (10+10)

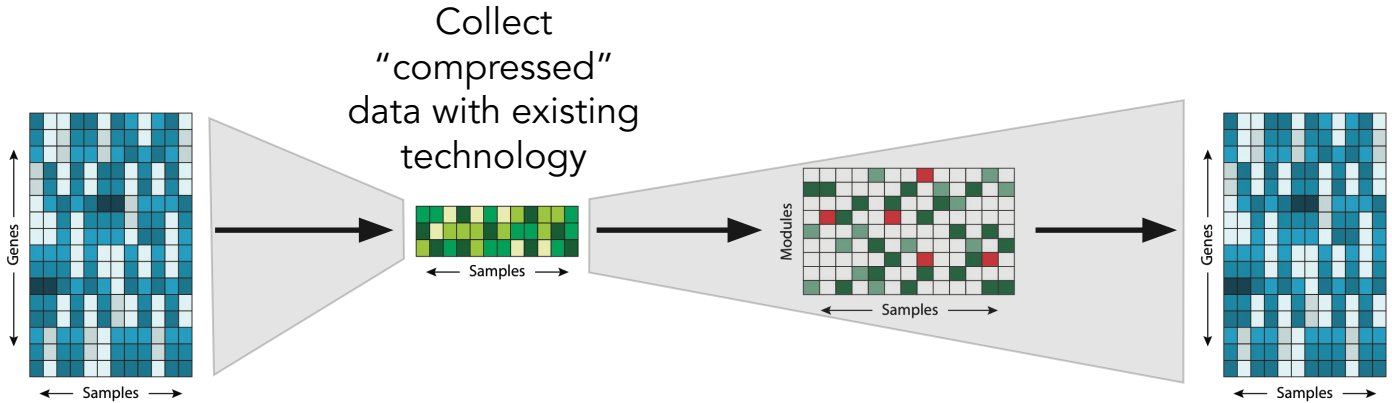


Transcriptome



Joshua Weinstein, Aviv Regev and Feng Zhang; *bioRxiv* 2018; *Cell* 2019

How to generate more data without a better instrument?



Brian Cleary, Cleary et al., Cell 2017

Using mathematics of: Random projections into low-dimensional space, random feature learning and compressed sensing (Johnson-Lindenstrauss lemma, Dasgupta and Gupta, Donoho, Candes, Tao, Eldar, etc)

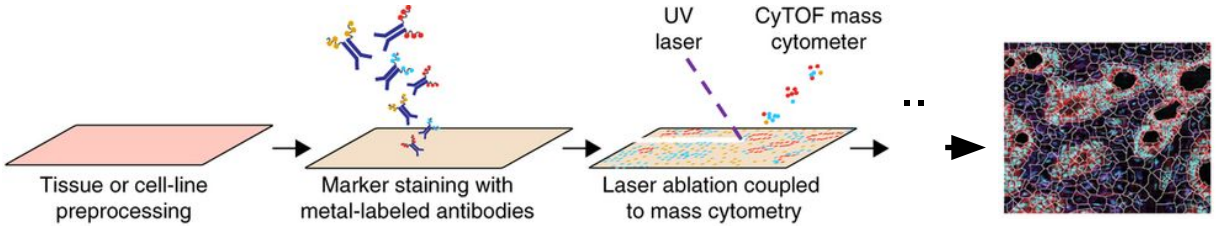
Use case: Same number of measurements; more information

Existing method: MERFISH/MIBI/CODEX

create image with abundance of
100 proteins/RNAs

Compressed version

create image with abundance of
10,000 proteins/RNAs



Giesen, et. al. 2014; also see: Angelo et al., 2014

What does it mean to acquire compressed expression data?

Instead of individual genes, measure abundance of composite genes**

Number of composite genes (m) much smaller than number of genes (g)

**Composite gene: a linear combination of abundances:

Mathematical:

$$cg_1 = \sum_{i=1}^{20000} w_{1,i} x_i$$

x_i ← abundance of gene$_i$

cg_1 ↑ composite gene$_1$

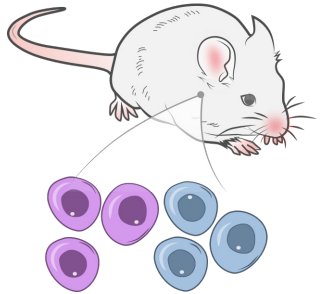
$w_{1,i}$ ↑ weight of gene$_i$ in cg$_1$

The weights can be random (and can be binary)

Cleary et al., Cell 2017

Johnson-Lindenstrauss Lemma; Indyk-Motwani

Compressed measurements in the lab with composite *in situ* imaging (CISI)

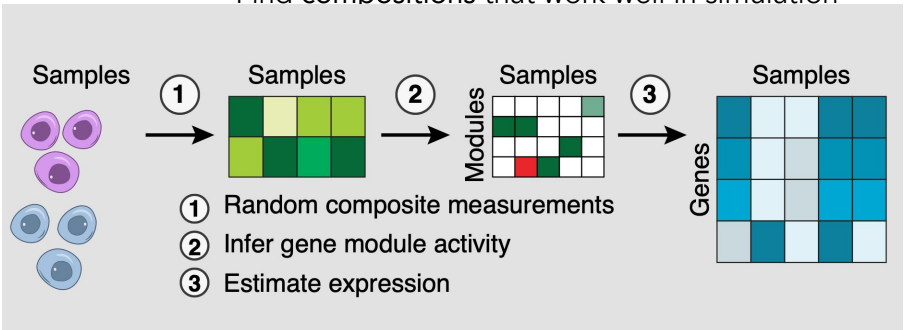


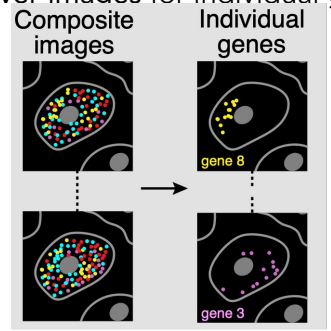
(1) scRNA-seq training data from ROI

- Find gene modules
- Find compositions that work well in simulation

(2) Composite imaging of ROI (in situ)

- Fit gene module activities (spatially)
- Recover images for individual genes (decompress)

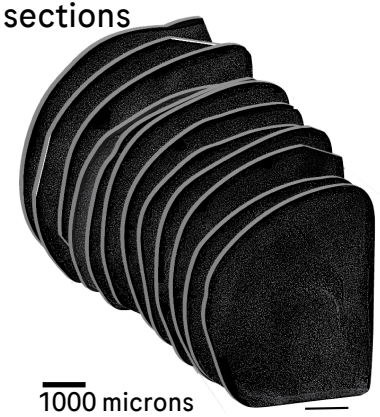




Cleary et al., Biorxiv 2019, Nature Biotechnology 2021

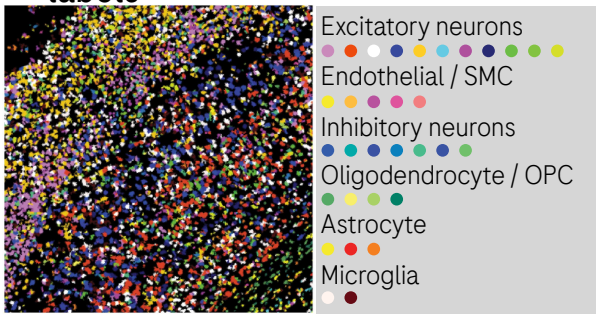
Decompressed CISI measurements to map cell type and states

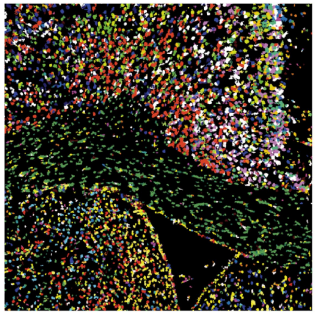
12 bisected coronal sections



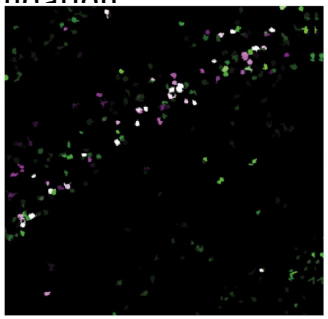
- 180mm^2, ~500,000 cells, 25x imaging
- 37 genes, 11 compositions, 4 genes / composite
- Genes: 5 IEGs, 5 classical type markers, 27 model driven
- 25x imaging; **overall gain of 537-fold increase in efficiency**

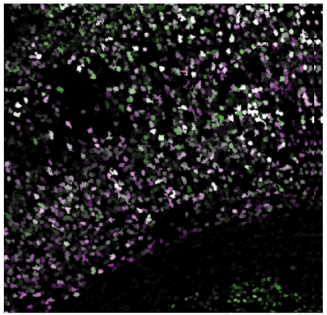
Cluster labels





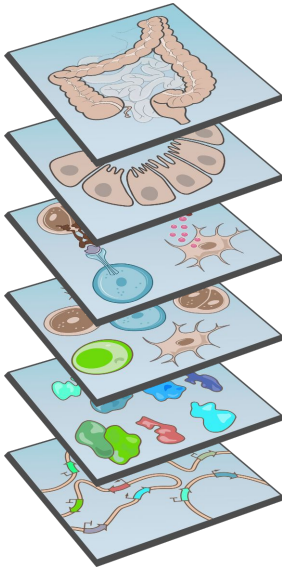
Individual gene validation





Cleary et al., *Nature Biotechnology* 2021; with Evan Murray, Anu Sinha, Fei Chen, Broad/MIT

Algorithms are essential for building and using atlases



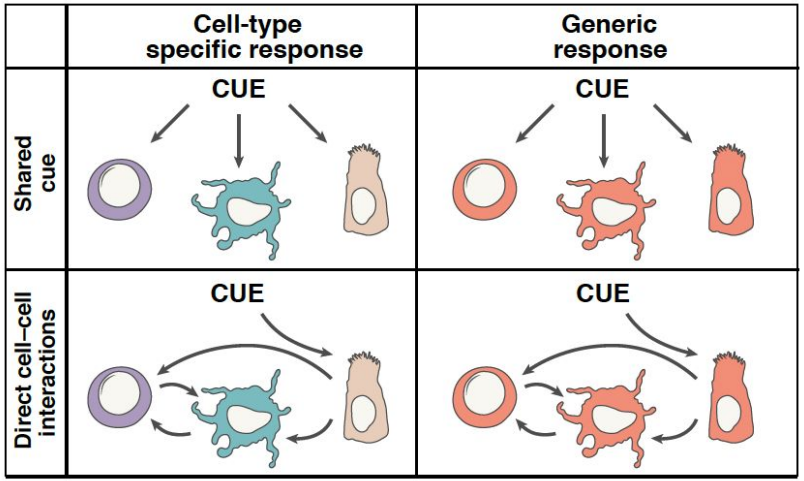
I. Integrate, enhance and construct atlas

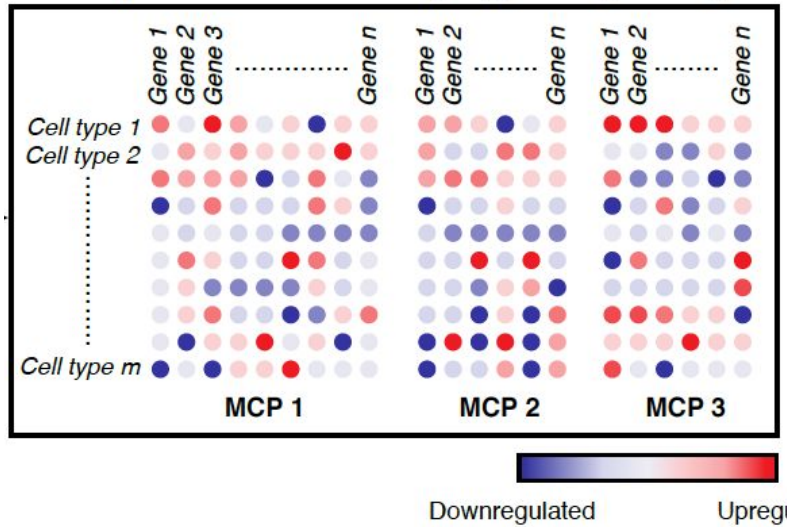
II. Define new experimental modalities

III. Enable new biological discoveries and concepts

Example: From cells to multi-cellular programs

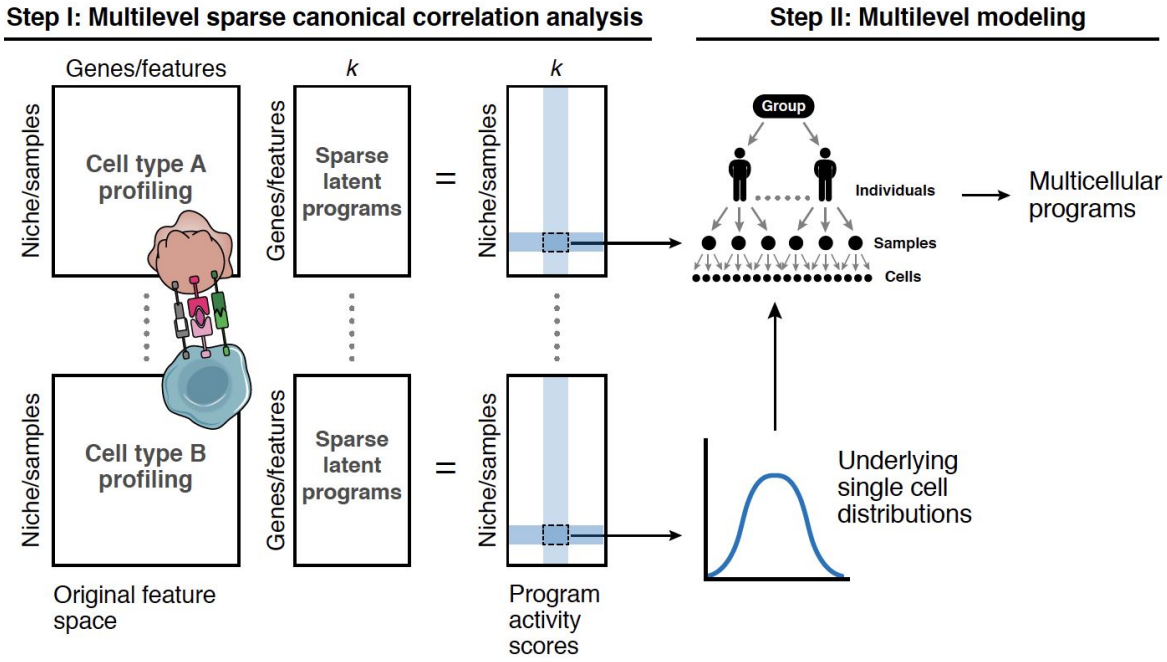
Tissue biology





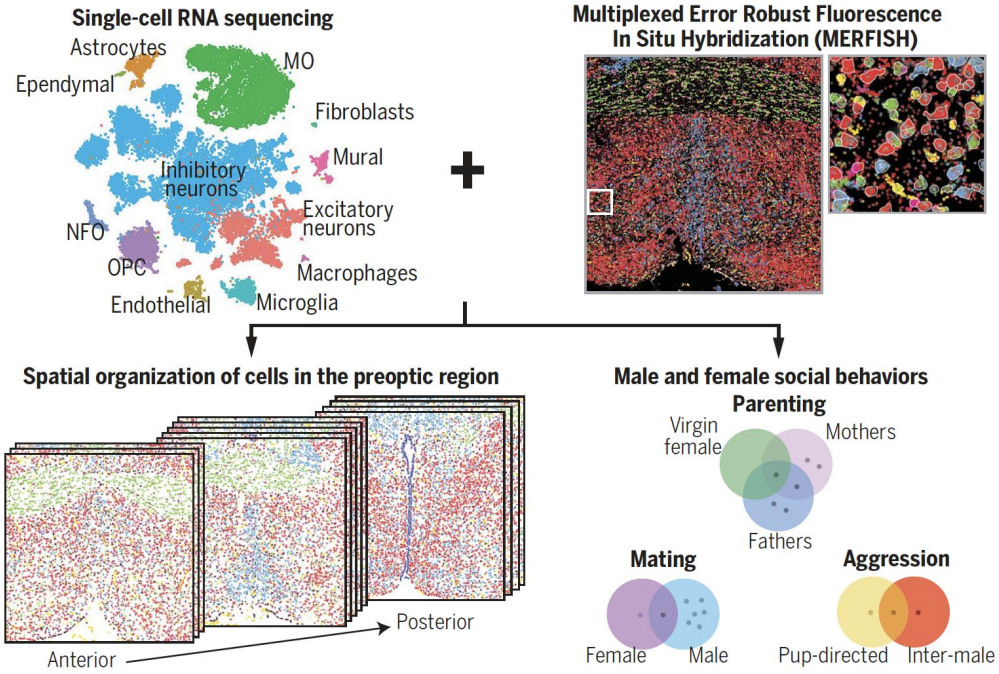
Livant Jerby, Jerby and Regev, Biorxiv 2020, *Nature Biotechnology* 2022

DIALOGUE: Inferring multi-cellular programs



Livant Jerby, Jerby and Regev, Biorxiv 2020, *Nature Biotechnology* 2022

Physical niches: MERFISH of the hypothalamic preoptic region

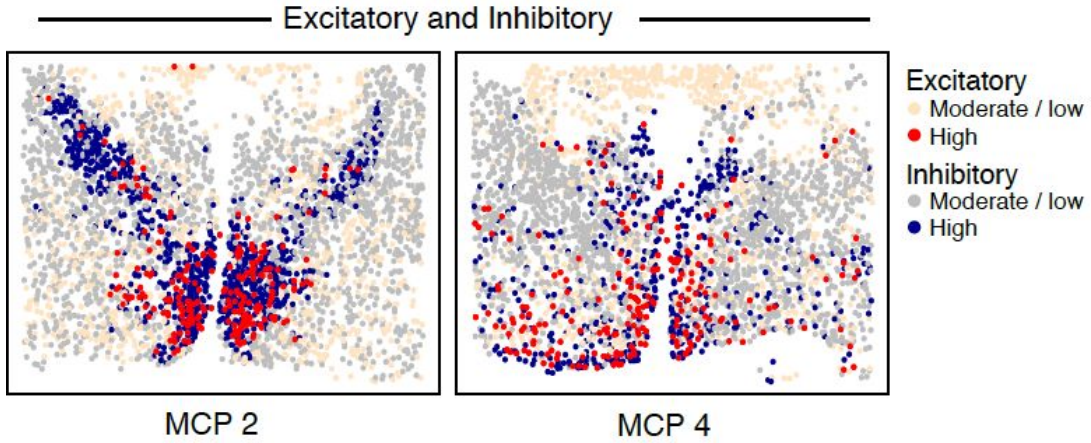


Moffitt*, Mambah-Mukkut et al., Science 2018: Xiaowei Zhuang and Catherine Dulac labs

Livant Jerby, Jerby and Regev, Biorxiv 2020, *Nature Biotechnology* 2022

Multi cellular programs: Coordinated and proximal

Excitatory neuron - inhibitory neuron MCPs (MERFISH, hypothalamus data)

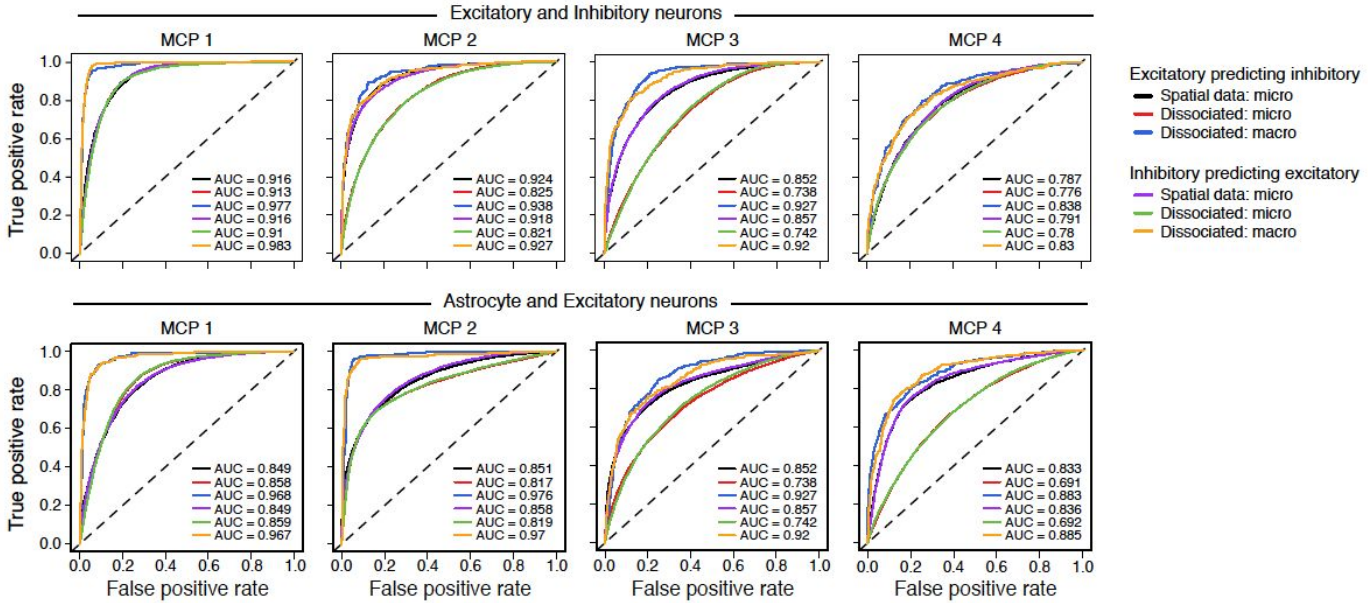


Livant Jerby, Jerby and Regev, Biorxiv 2020, *Nature Biotechnology* 2022

MCPs learned from ‘coarsed’ data can predict unseen neighbors

Learn MCPs from ‘coarse in silico dissociated’ data (~500 “in silico dissociated” cells; ~50-100μm2

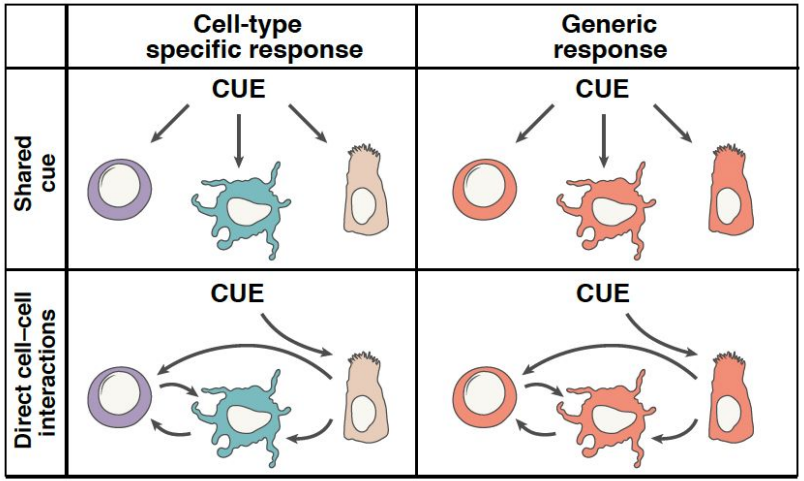
Use learned MCP and expression of a cell to **predict expression of its neighbors** in test data in radius of 15 or 500 cells



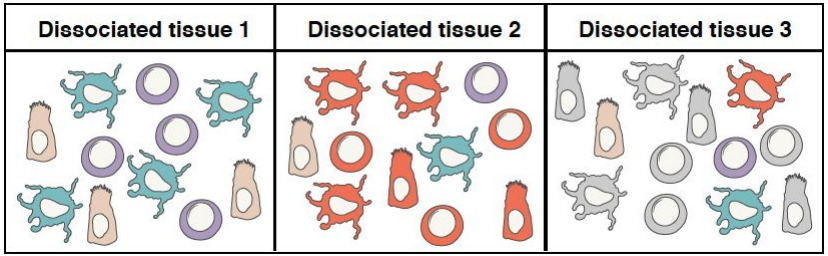
Livant Jerby, Jerby and Regev, Biorxiv 2020, *Nature Biotechnology* 2022

Multicellular programs can be learned from single cell data

Tissue biology



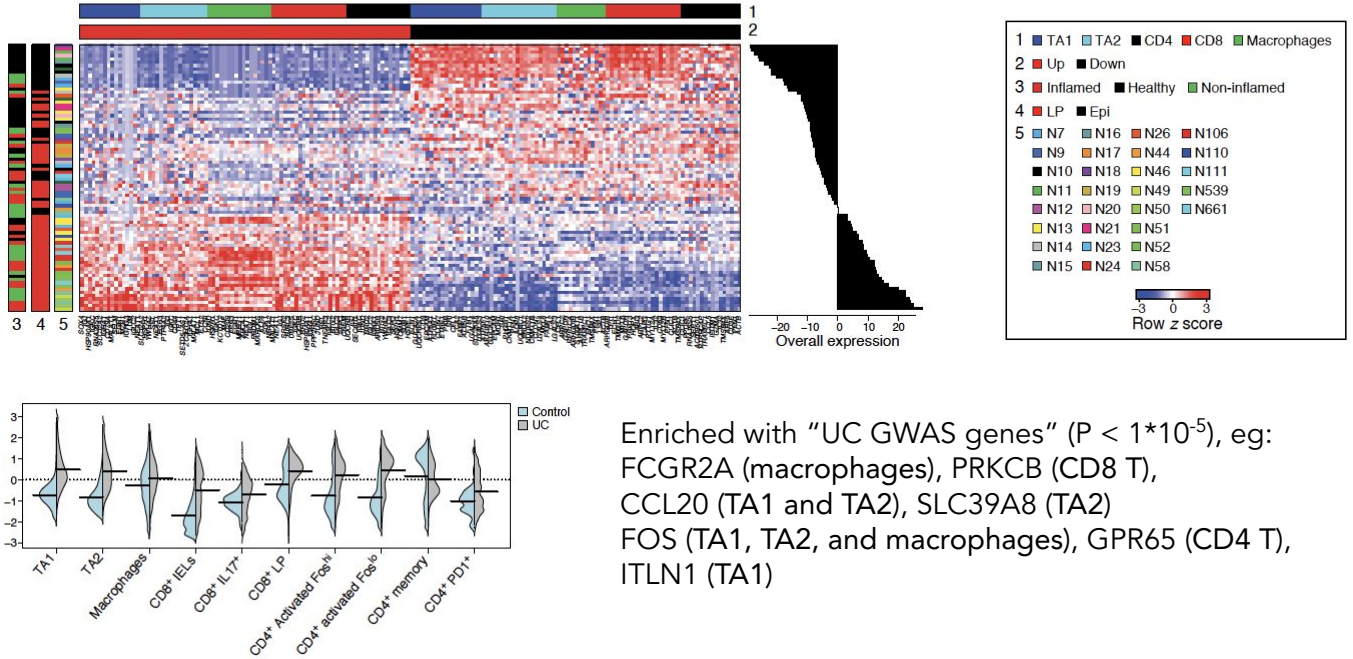
Single cell profiles



Livant Jerby, Jerby and Regev, Biorxiv 2020, *Nature Biotechnology* 2022

An IBD-associated multi-cellular program across T cells, epithelial cells and macrophages learned from scRNA-seq

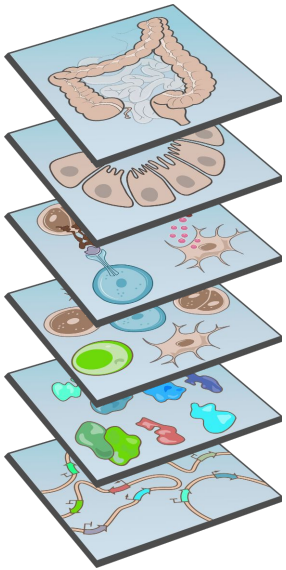
MCP2: UC associated



Enriched with "UC GWAS genes" ($P < 1*10^{-5}$), eg:
FCGR2A (macrophages), PRKCB (CD8 T),
CCL20 (TA1 and TA2), SLC39A8 (TA2)
FOS (TA1, TA2, and macrophages), GPR65 (CD4 T),
ITLN1 (TA1)

Livant Jerby, Jerby and Regev, Biorxiv 2020, *Nature Biotechnology* 2022

Algorithms are essential for building and using atlases



I. Integrate, enhance and construct atlas

II. Define new experimental modalities

III. Enable new biological discoveries and concepts

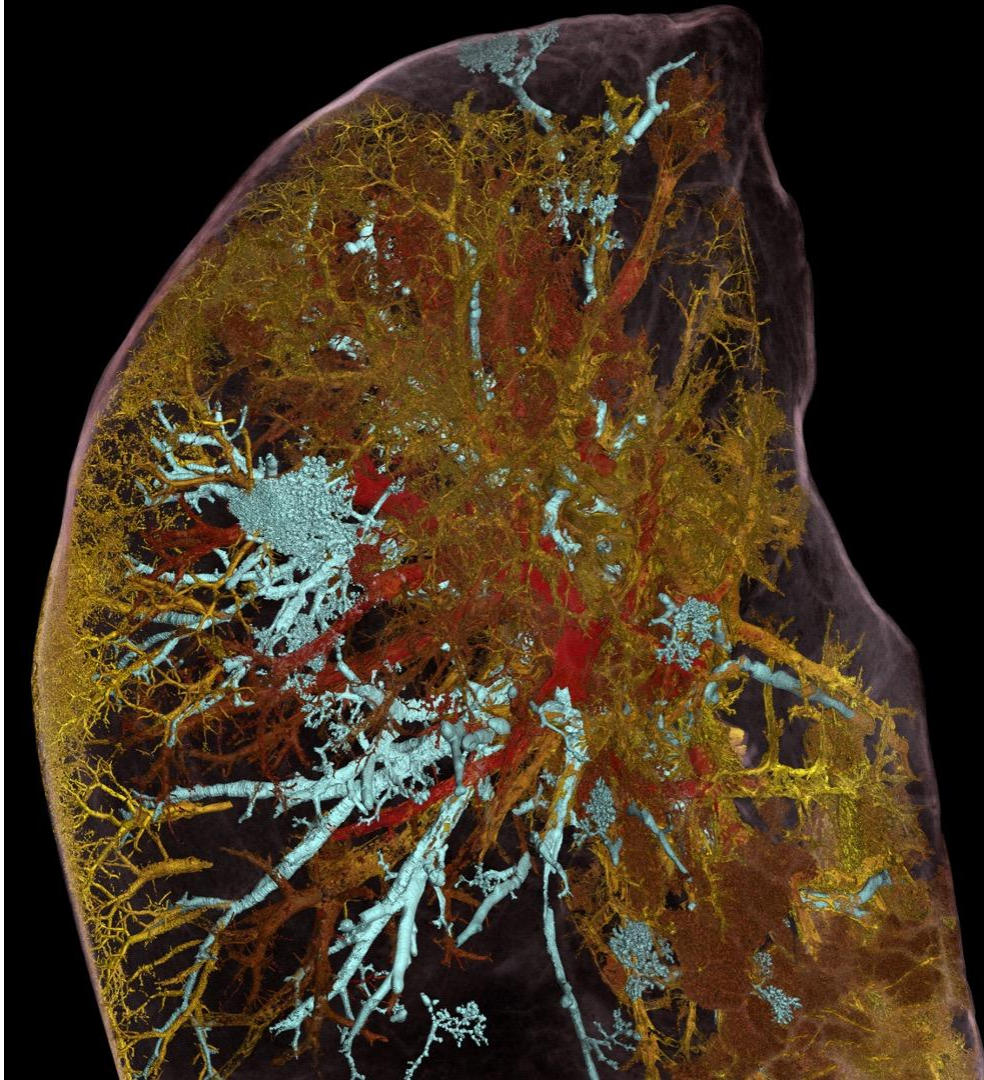
Peter Lee

Hierarchical Phase-Contrast Tomography (HiP-CT)

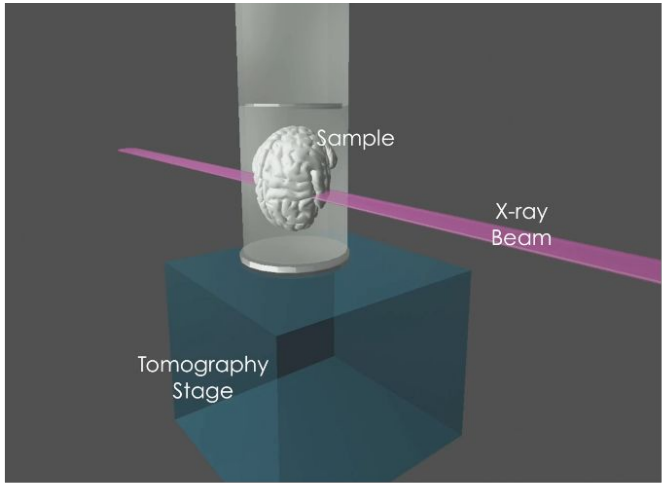
To Create

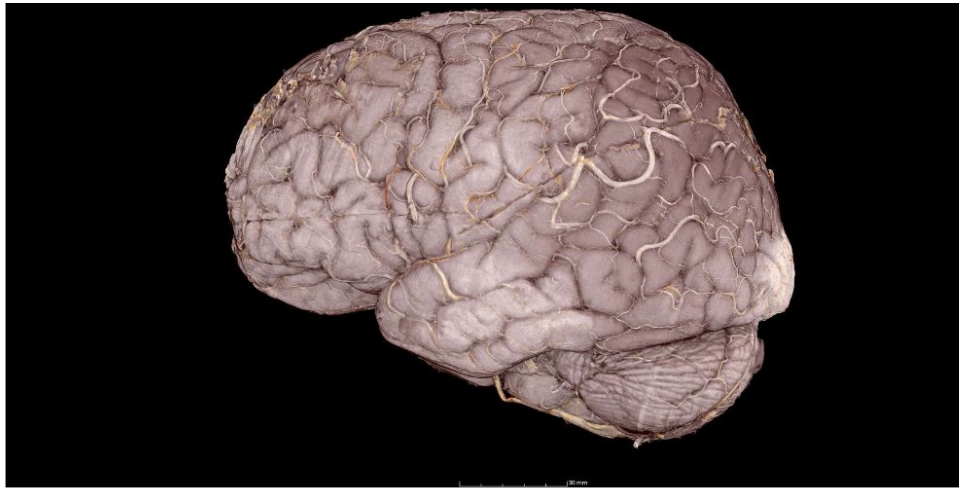
The Human Organ Atlas

Peter.Lee@ucl.ac.uk



Hierarchical Phase-Contrast Tomography





- HiP-CT is a propagation-based phase contrast local tomography technique
- Organ overview at 25µm/voxel
- Zoom regions anywhere at higher resolution (ca. 1 µm voxels)
- We can resolve some types of single cells in intact human organs

By bridging the gap from clinical radiology to histology obtained new insights into COVID-19's impact on angiogenesis

Macroscale (cm) | Mesoscale (mm) | Microscale (µm) | Nanoscale (nm)

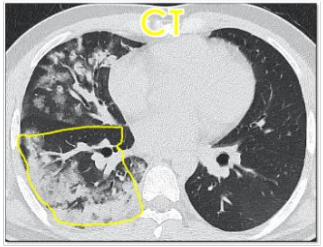
Ex vivo clinical imaging

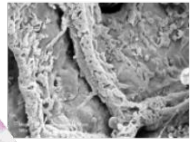
In vivo clinical (CT,MRI)

light microscopy

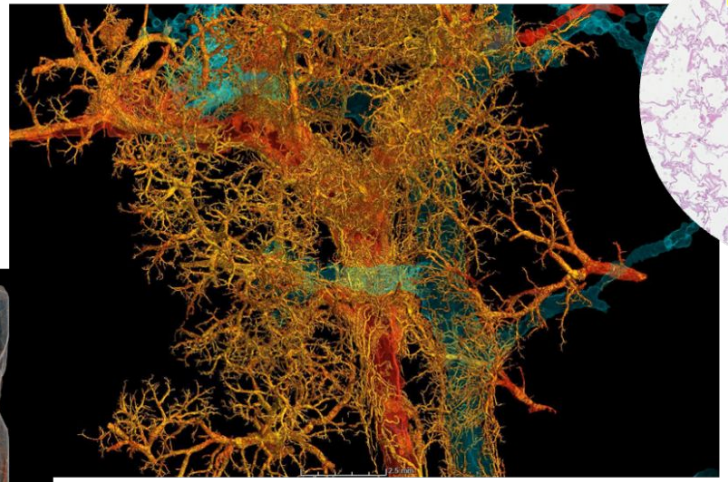
Electron microscopy

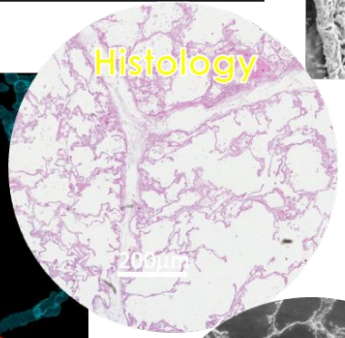
HiP-CT

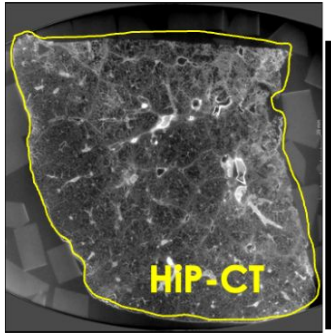


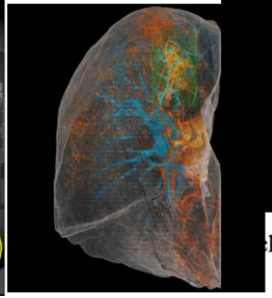


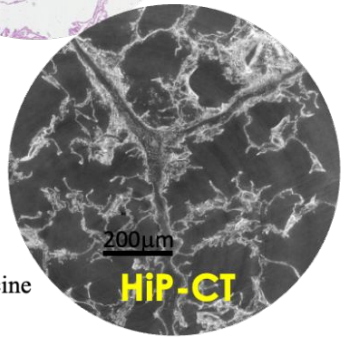
Ackermann et al., 2020









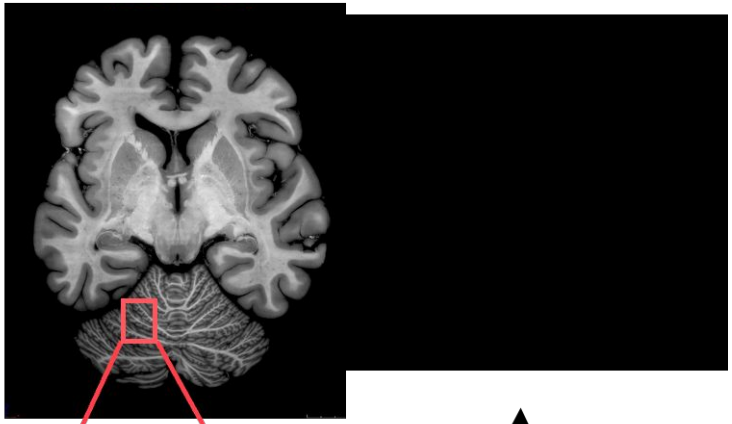


ckermann et al, American Journal of Respiratory and Critical Care Medicine

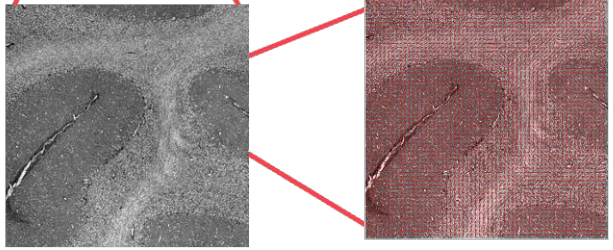
Correlation up- and down-scale

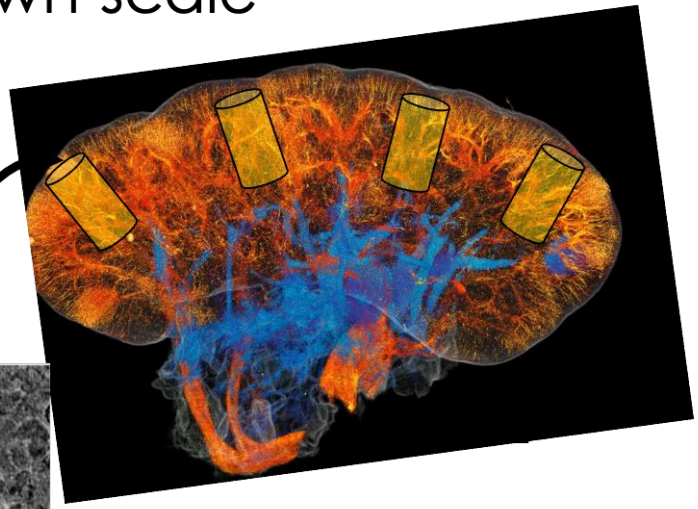
HiP-CT

MRI

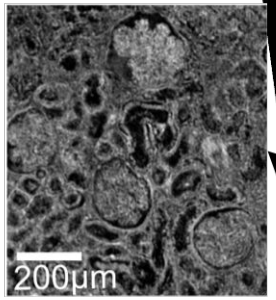


Validation

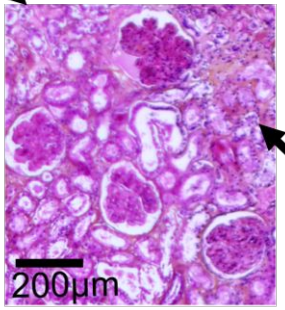




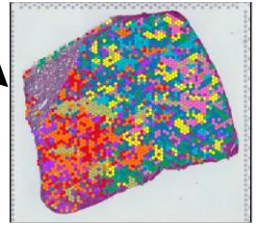
HiP-CT



Histology



Spatial transcriptomics

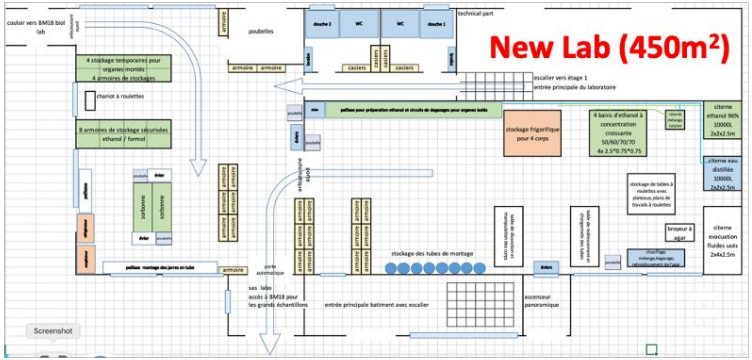


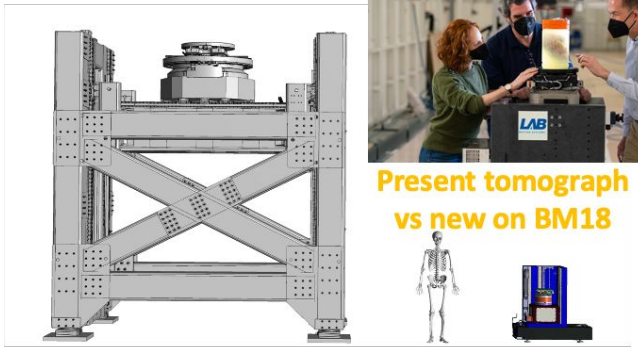
Whole human body HiP-CT Challenges

Existing 17m²

Ethical, dignified, and safe:

- Preparation
  - Requires new lab, new scanning setup, new handling
- Huge Data sizes (100Tb@20μm, petabytes with zooms)
  - Sharing
  - Quantifying (“Google Earth” vs “Google Maps”)

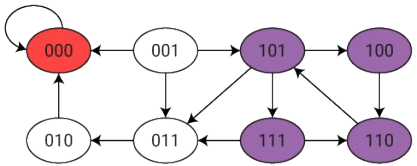




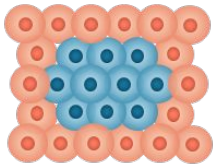
Present tomograph vs new on BM18

Peter Zandstra

Note: As I will be able to attend only some of the sessions, I tried to include more material here, which I hope will be useful.



GRN modeling of single cell fate

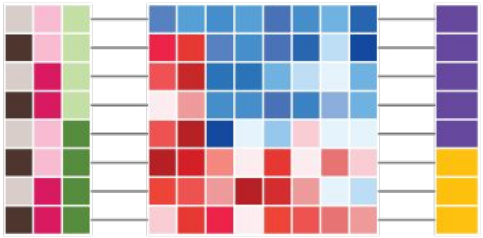


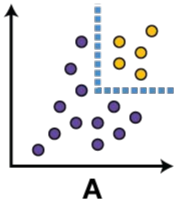
How does single cell fate map to tissue form and function?

Can we program tissue fate and function?

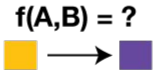
GENE CIRCUITS CODE INFORMATION ON CELLULAR DECISION MAKING

Input Processing Function



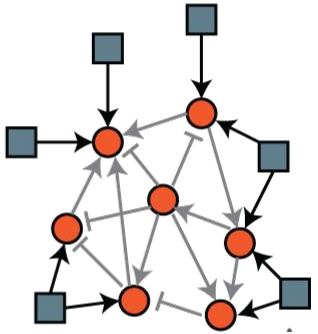


Rule-based machine learning



Satisfiability

Intracellular



Yachie-Kinoshita et al, Mol System Biol 2017
Ng. et al. Nature 2017
Lechnam et al. Cancer Cell 2016

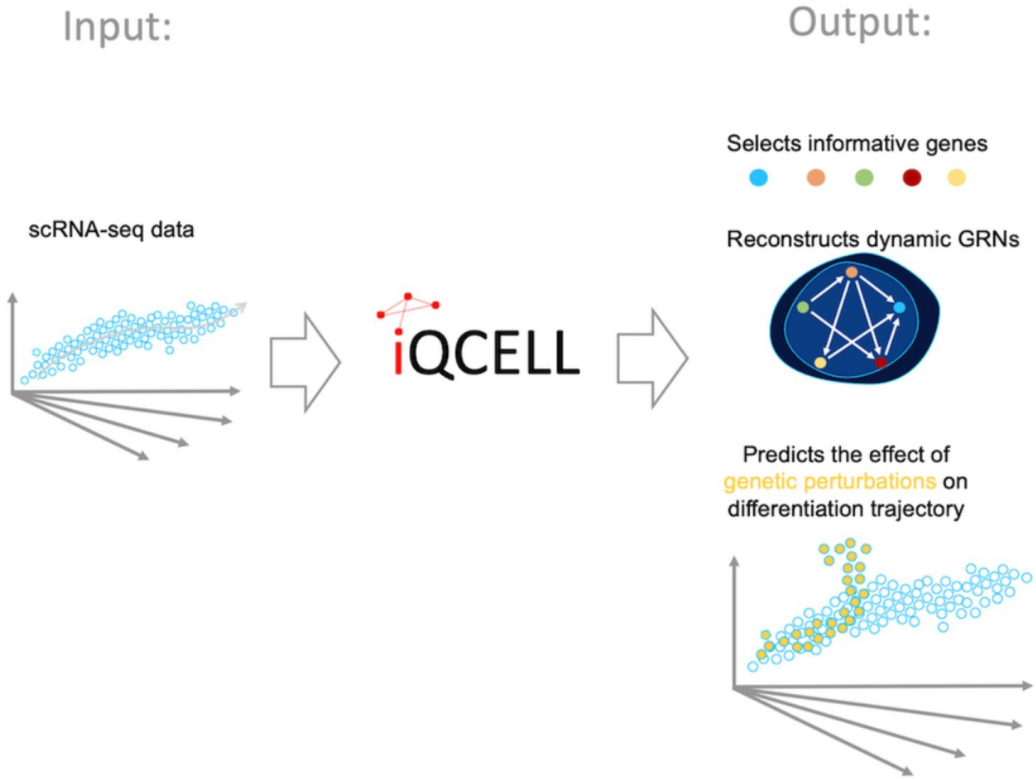
Collaboration with: Microsoft Research SBI The Systems Biology Institute

IQCELL: PREDICTING THE EFFECT OF GENE PERTURBATIONS ON DEVELOPMENTAL TRAJECTORIES

Tiam Heydari

Matthew Langley

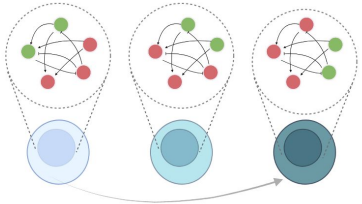
Ayako Yachie- Kinoshita



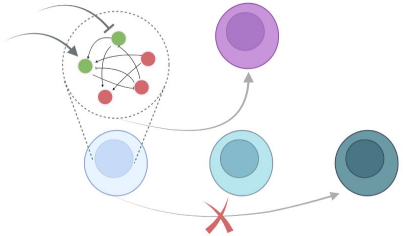
Heydari et al. *PLOS Computational Biology* 2022

https://gitlab.com/stemcellbioengineering/iqcell

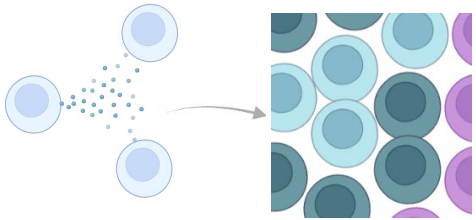
NEXT STEPS:



High throughput data-driven modeling of Boolean GRNs can predict the gene-gene interactions and the effect of gene perturbation at the 'system-level'

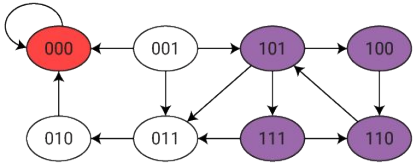


This system-level knowledge can lead to model informed intervention strategies for engineering development towards specific cell types

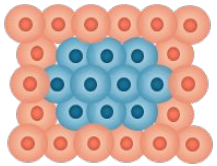


Now, we are looking at the effect of cell-cell communication on the intracellular GRNs dynamics with the aim of controlling the development of diverse multicellular tissues

THREE INTEGRATED STRATEGIES TO PROGRAM TISSUE DEVELOPMENT

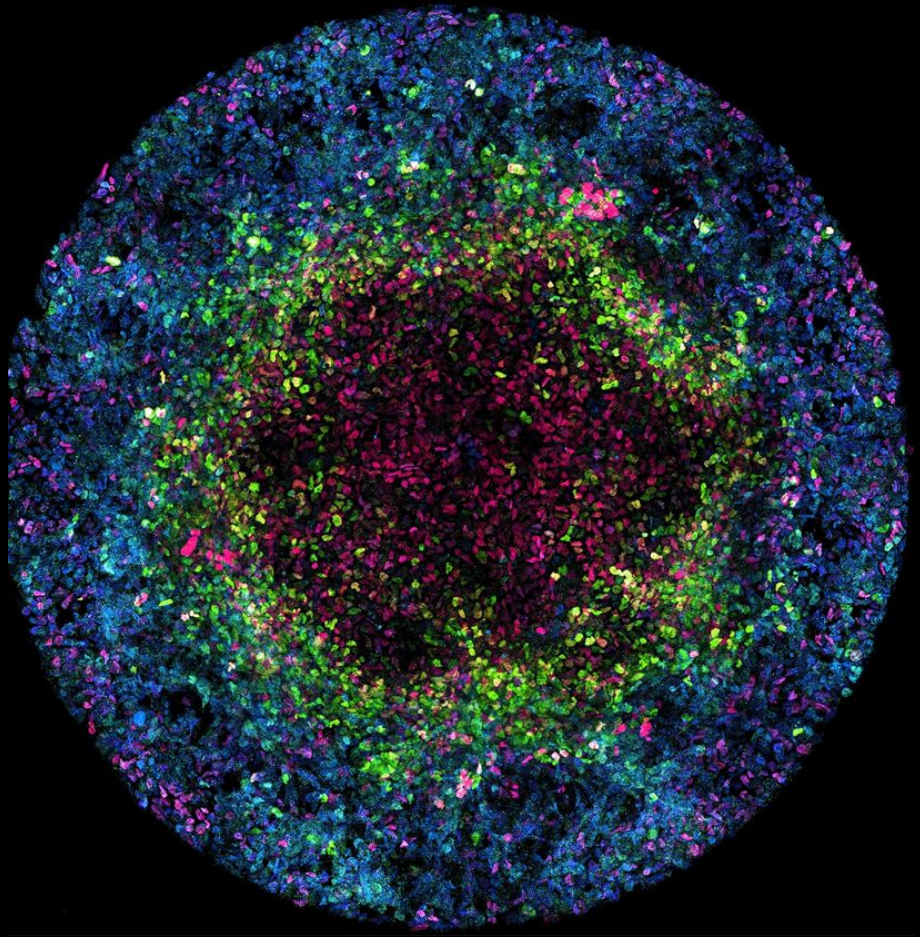


GRN modeling of PSC differentiation

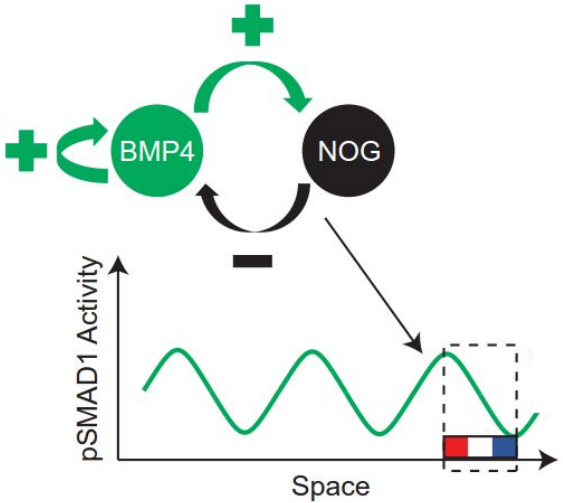


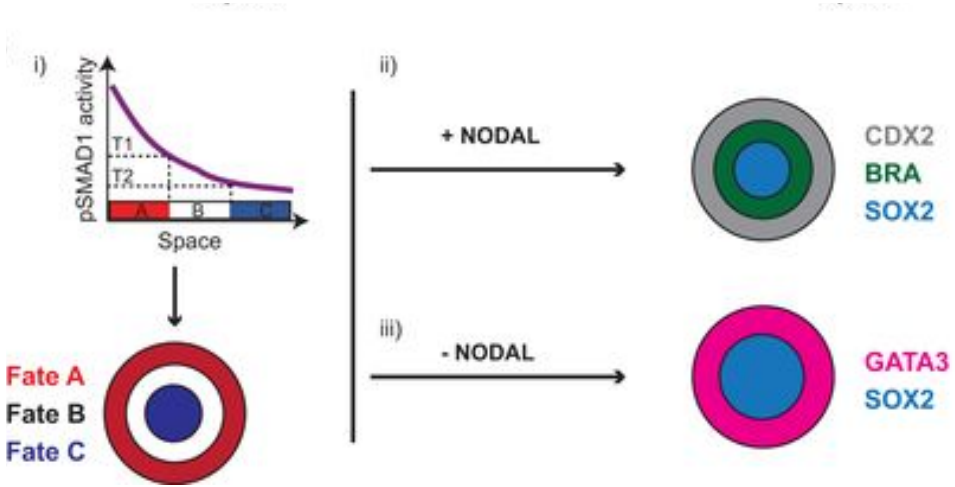
How does single cell fate map to tissue form and function?

Can we program tissue fate and function?



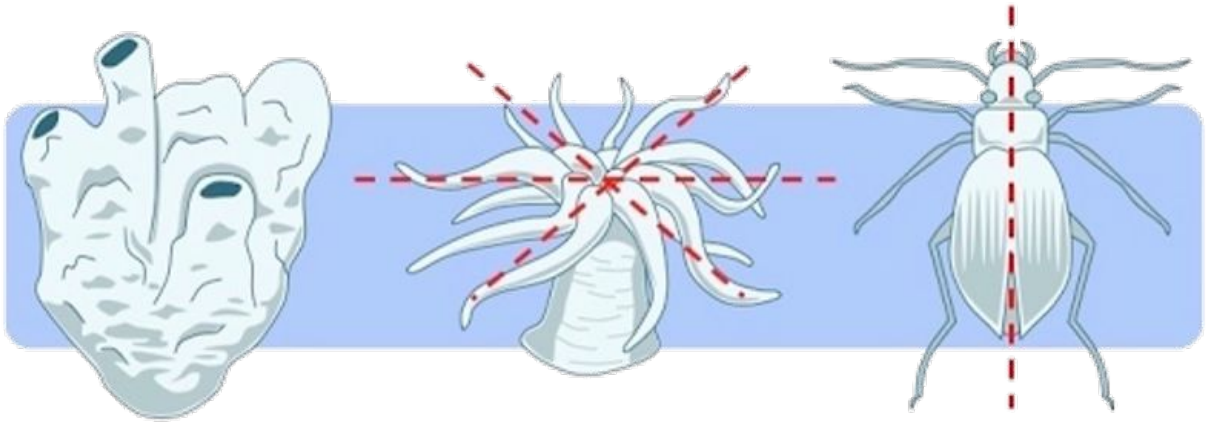
NODAL REGULATES A GASTRULATION AND NEURULATION SWITCH: CENTRO-SYMMETRIC PATTERNING



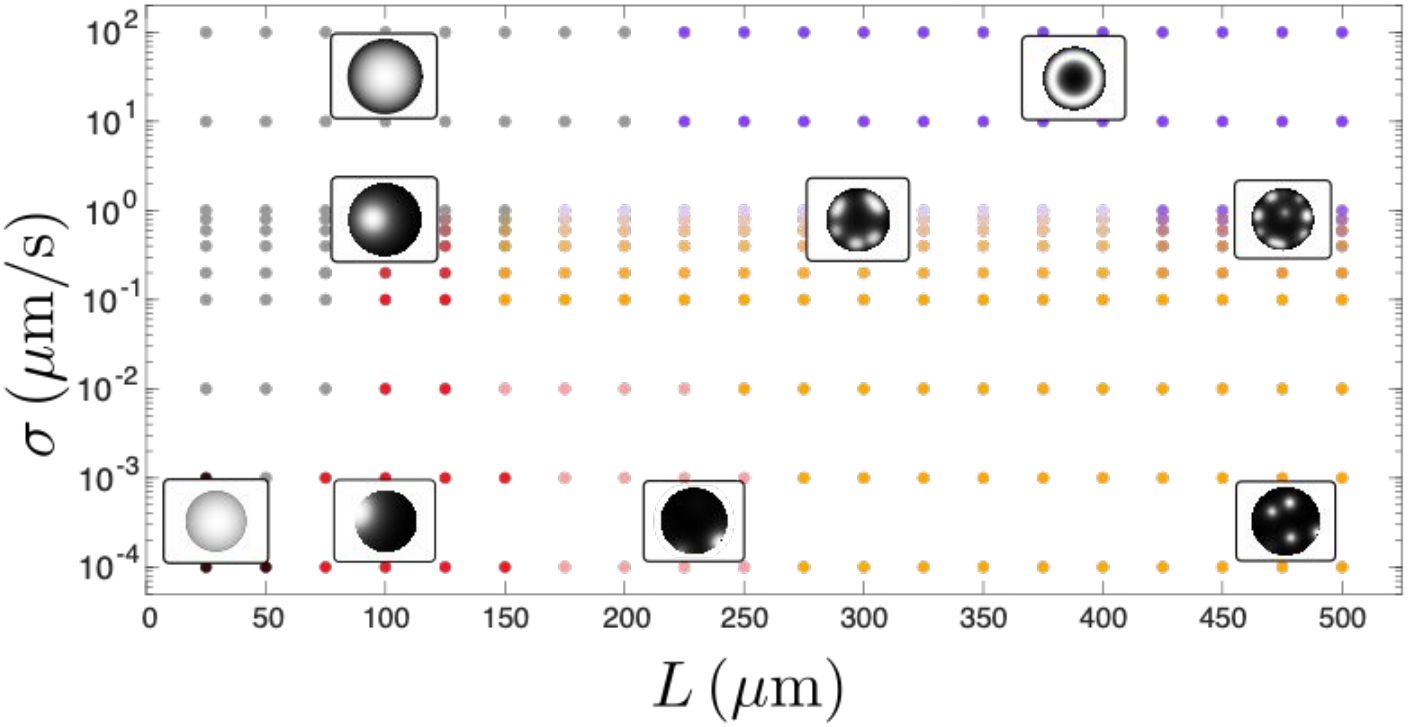


Tewary, M, et al. 2017. Development
Tewary, M, et al. 2019. *PLOS Biology*

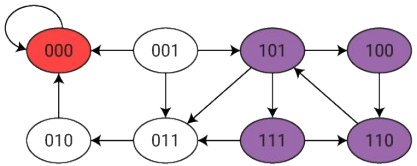
HOW DOES THE EMBRYO BREAK SYMMETRY?

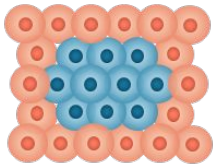


FROM ASYMMETRY TO SYMMETRY AS A FUNCTION OF SIZE



NEXT STEPS: GROWTH CONTROL AND SCALING

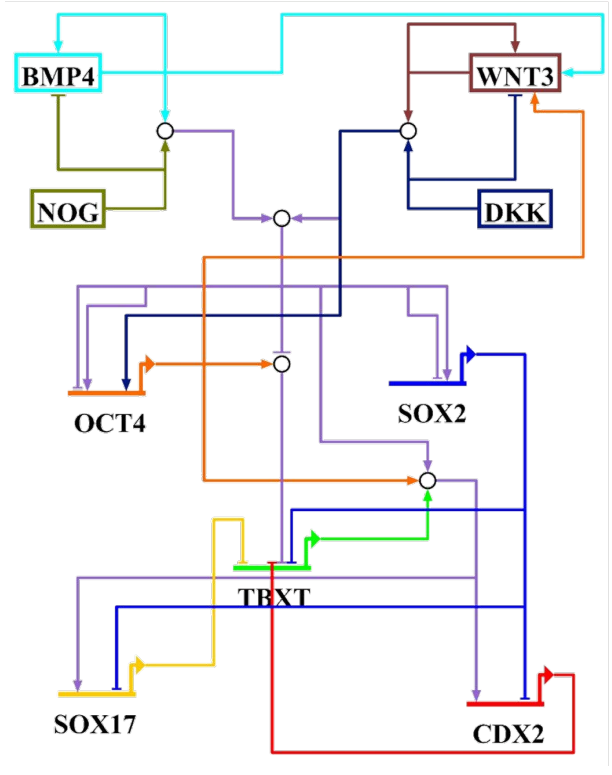


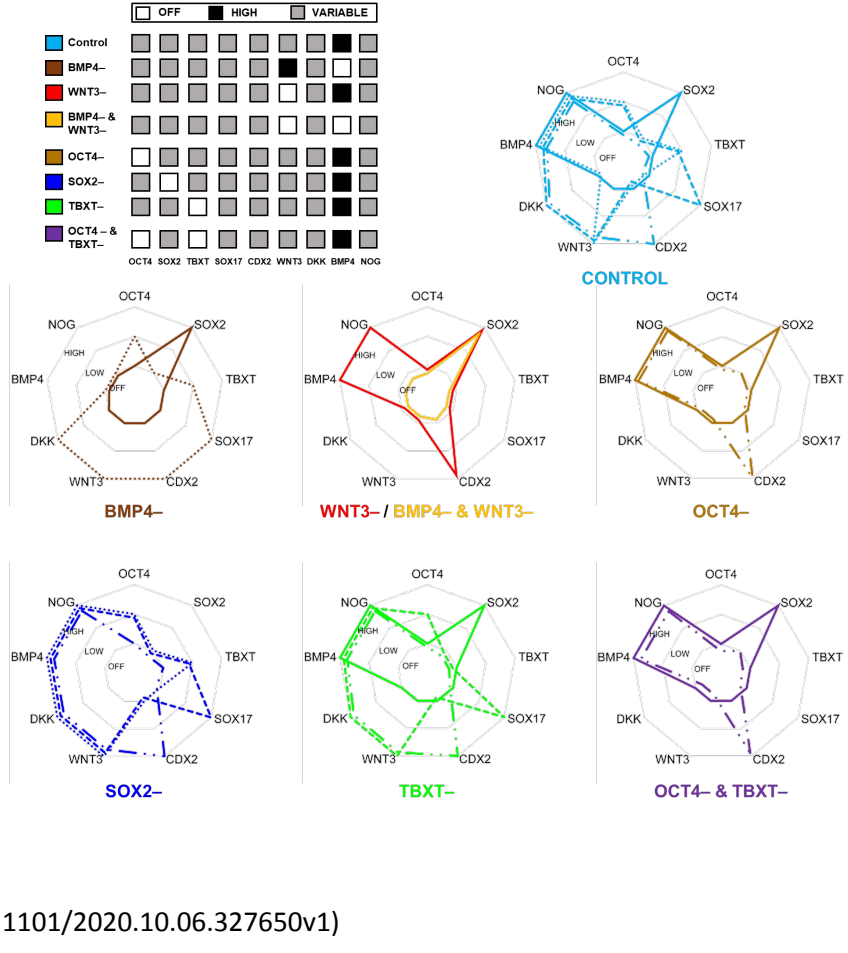


How does single cell fate map to tissue form and function?

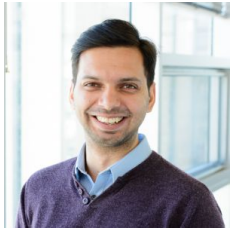
Can we program tissue fate and function?

A MINIMIZED EXECUTABLE NETWORK MODEL OF EARLY HUMAN PSC FATE RESPONSES

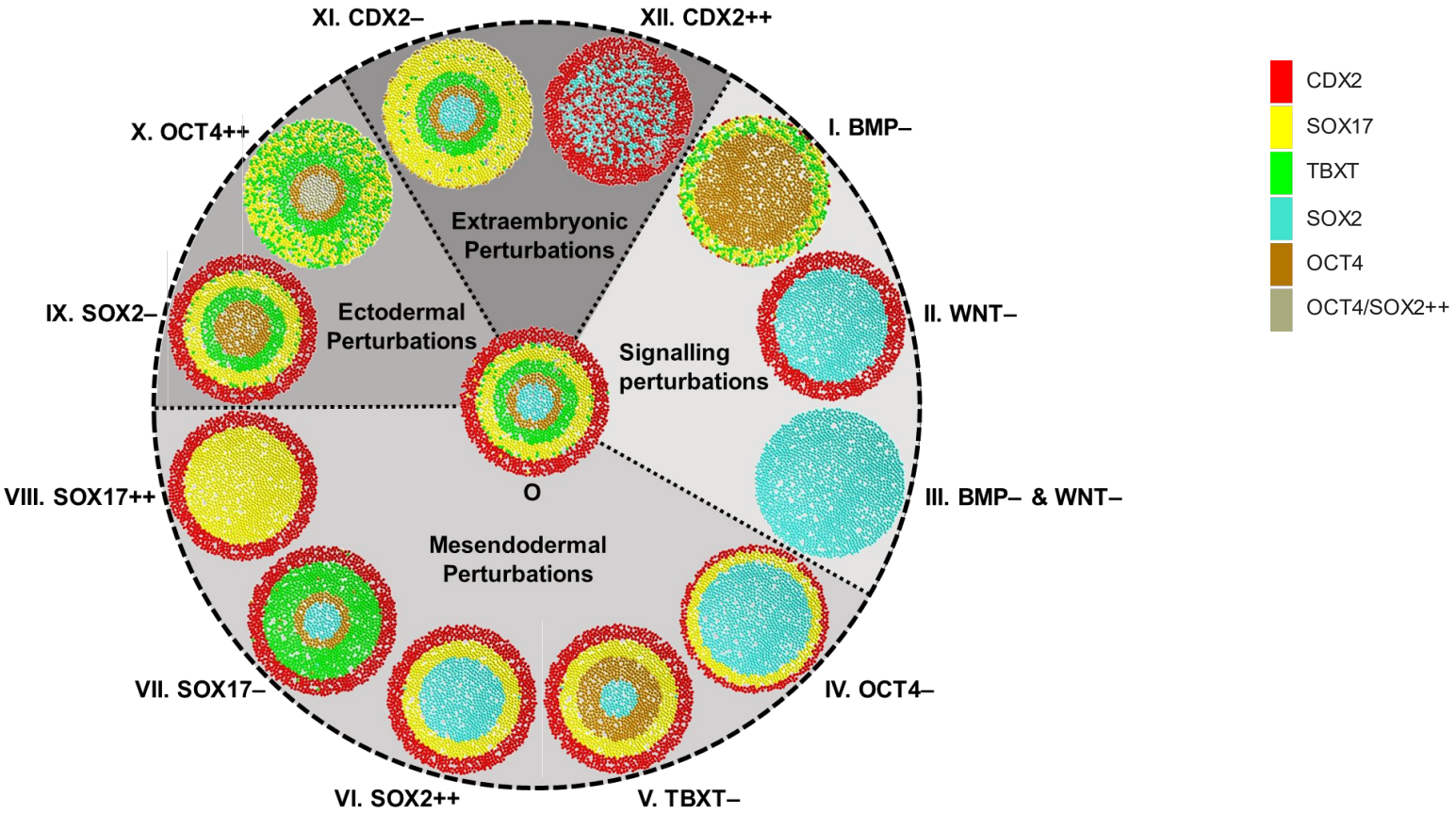




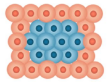
Kaul et al (https://www.biorxiv.org/content/10.1101/2020.10.06.327650v1)



CONNECTING NETWORK WIRING TO TISSUE-LIKE PATTERN FORMATION



How do single cells make decisions?

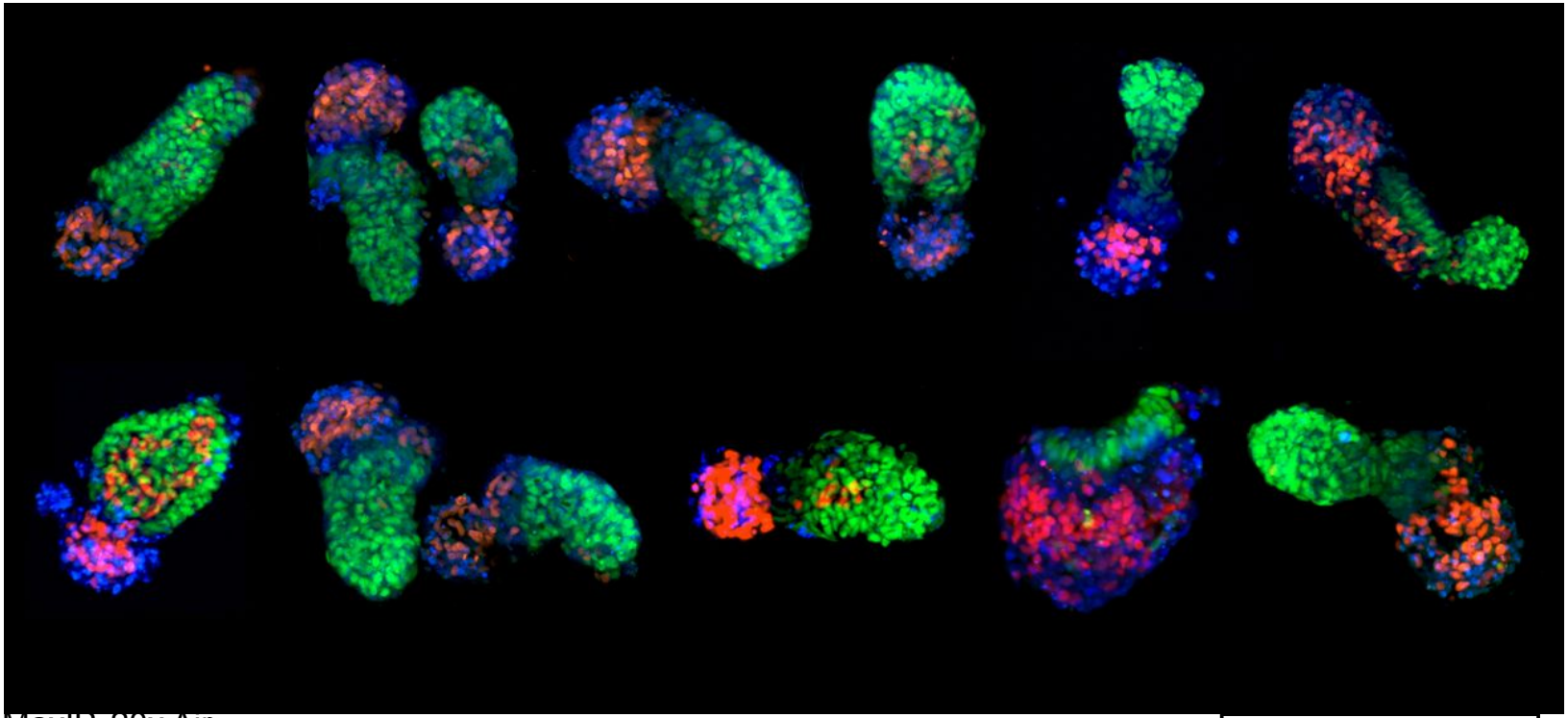


How does single cell fate map to tissue form and function?

Can we program tissue fate and function?

3D tissue development

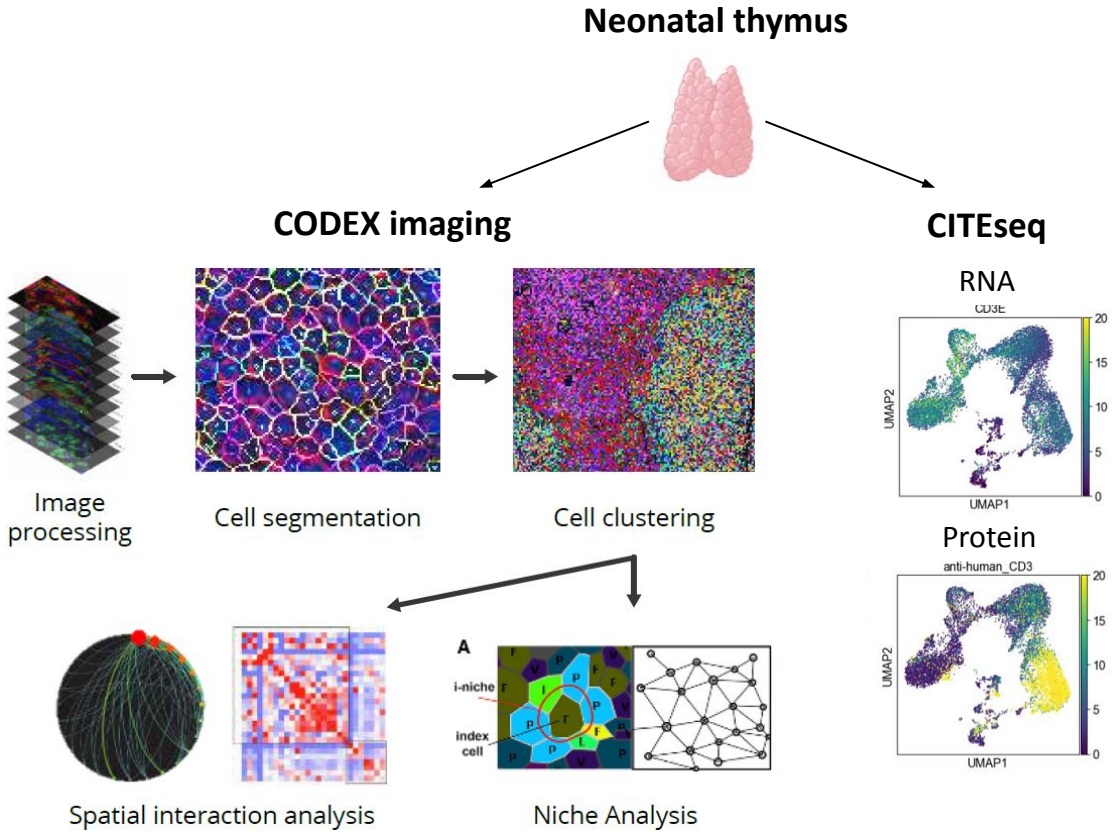
PLATFORM MOVES CELLS TOWARDS AN IN-VIVO LIKE AXIAL GERM LAYER ARRANGEMENT IN 3D



MaxIP, 20x Air

SOX2 = ectoderm
Brachyury = mesoderm
SOX17 = endoderm

A spatial-temporal molecular map of human thymus architecture




Rossi

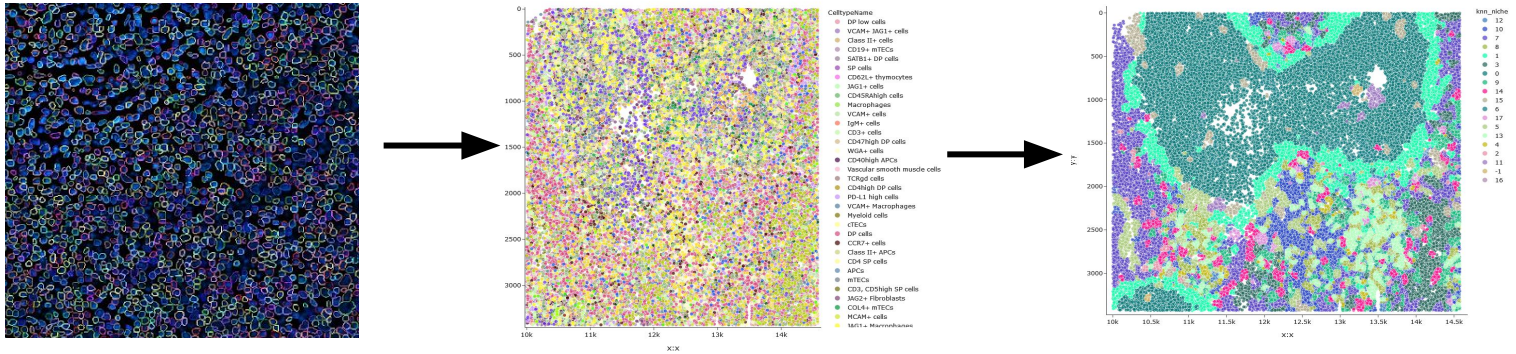

Levings

Goals

- Define spatial patterns of signaling and adhesion ligands that guide T lineage branch points
- Quantify and dissect neonatal thymic developmental niches at single cell and sub-cell resolution
- Screen stage-specific signals *in vitro* to guide lymphoid lineage commitment

Spatial multiomics identifies key developmental niches in human thymus

- Using custom image analysis tools we cluster cells into proximity-based niches
- We describe heterogenous cortical niches defined by multiple macrophage subsets, perivascular cells, and thymocyte subsets
- These cell subset phenotypes are confirmed with CITEseq



Segmentation

Cell Clustering

Neighbourhood Clustering

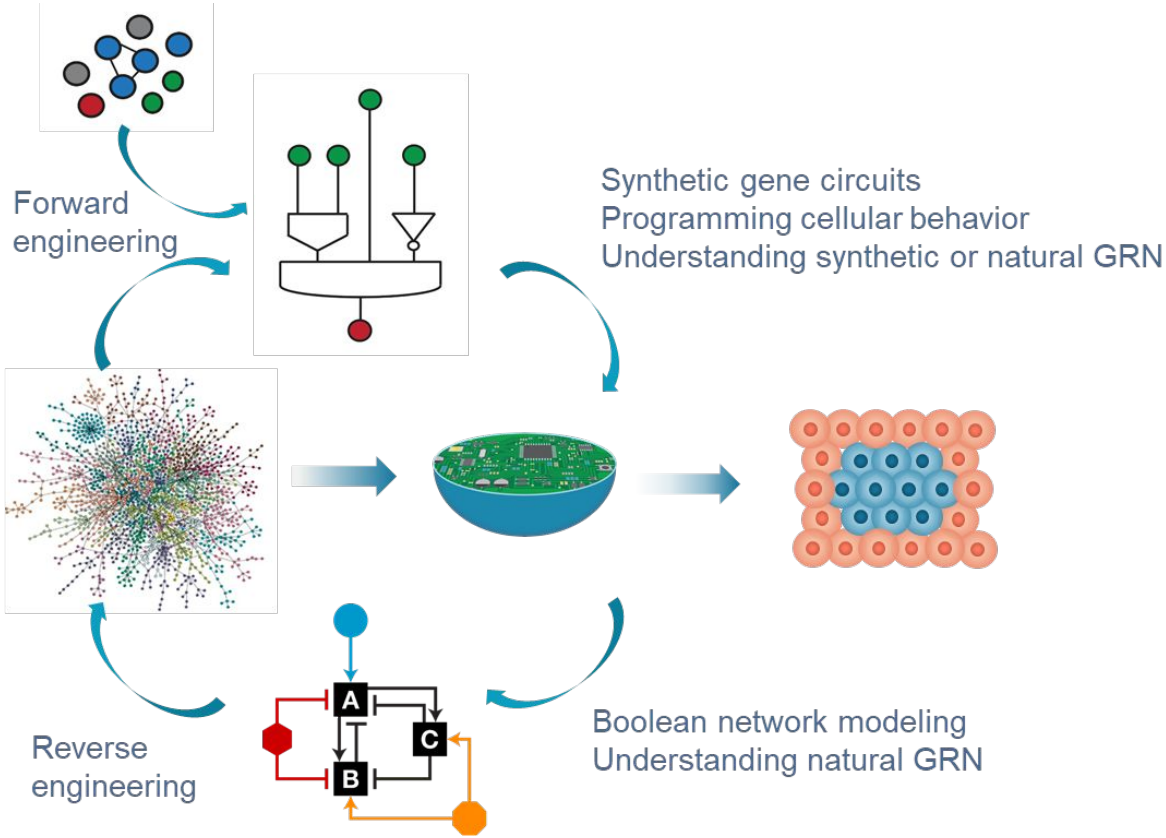
Key questions

1. Time?

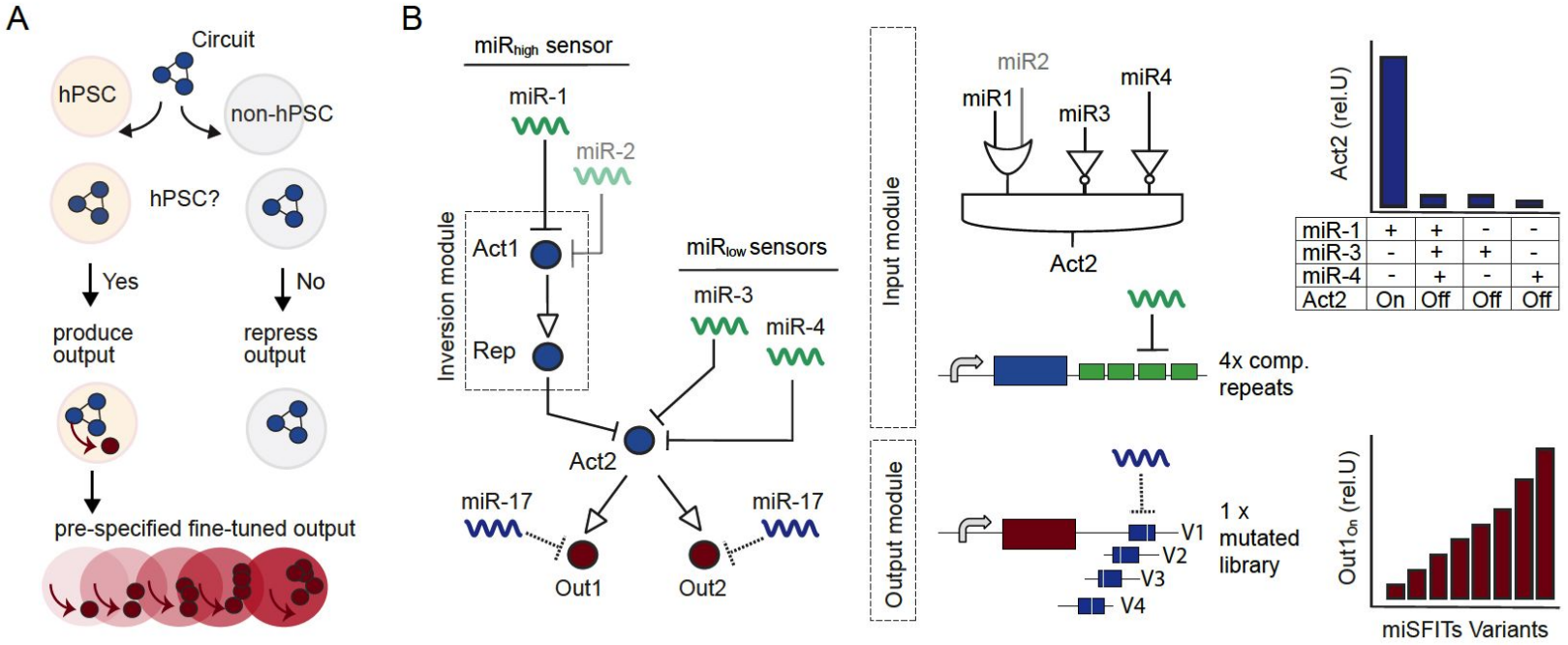
2. Learning and perturbation – how do we use the information to engineering “synthetic” tissue architecture?

3. What synthetic tissue do we want and what can we use it for?

Next steps: Adding programmable design to 3-D tissue architecture

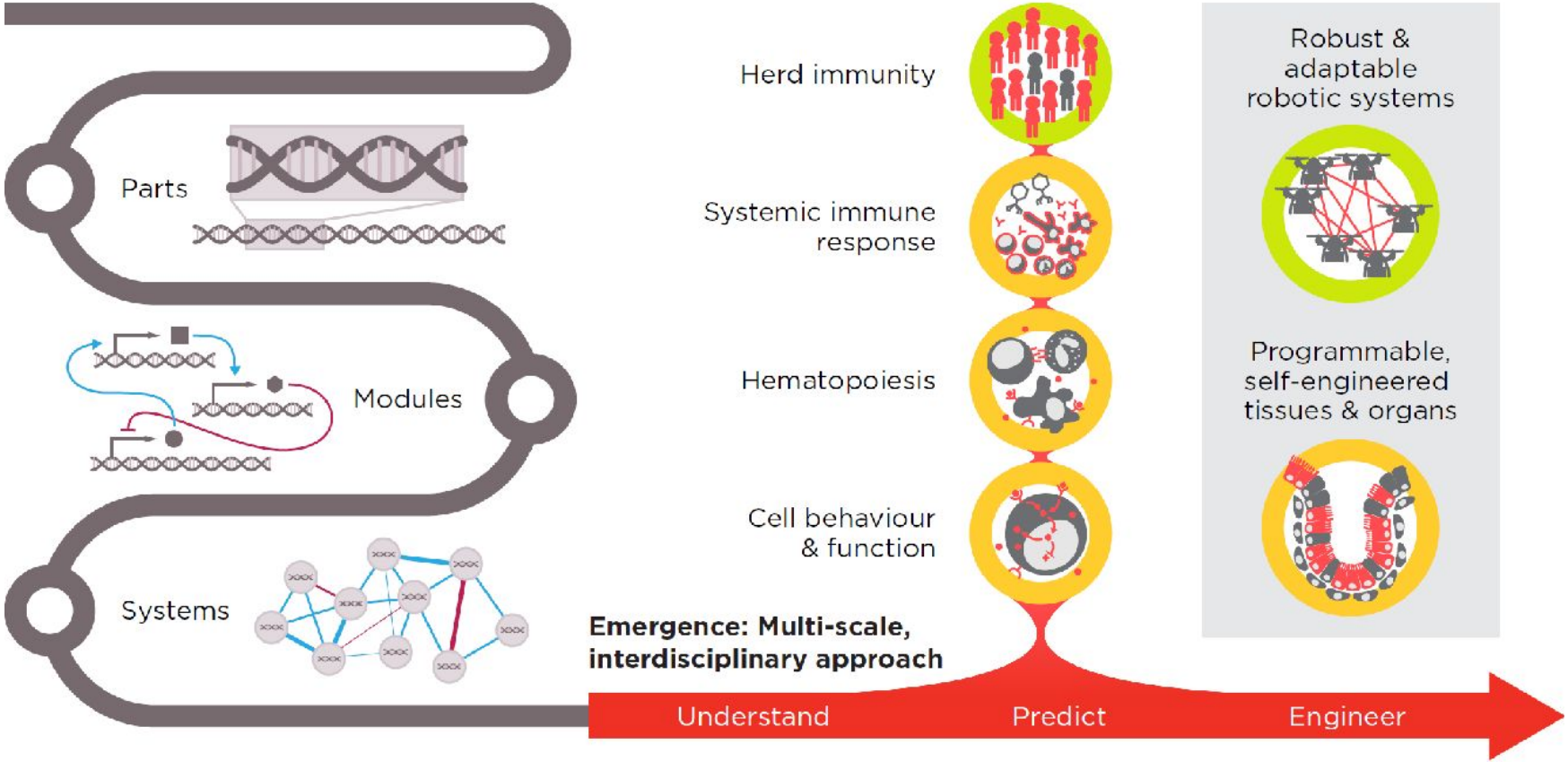


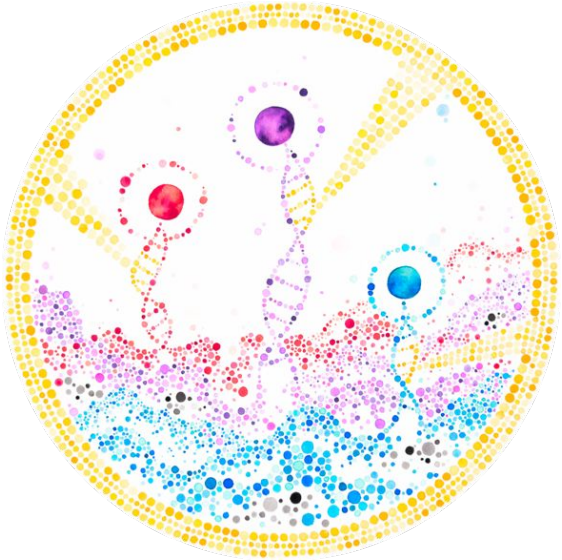
DESIGN OF HPSC-SPECIFIC DIGITAL TO ANALOG CONVERTER



Michaels et al. Nature Communications volume 10, 818 (2019)
Xie et al. Science 2011 Sep 2;333(6047):1307-11. doi: 10.1126/science.1205527
Prochazka et al. Nat Commun . 2014 Oct 14;5:4729. doi: 10.1038/ncomms5729.
Prochazka and Michaels Manuscrit in Preparation (2022)

CIFAR Emergence in Living and Non-Living Systems





Jennifer Ma, PhD
Zandstra lab, Donnelly Centre
University of Toronto

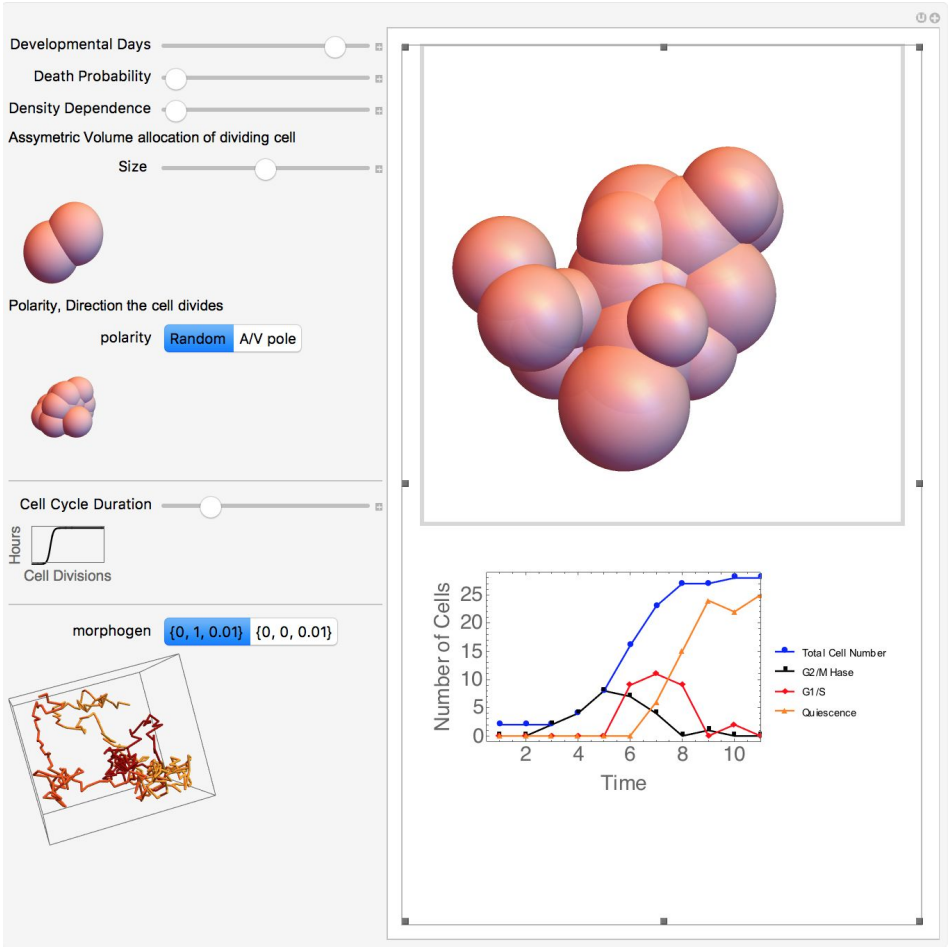
Gary Bader

Cell development simulation model

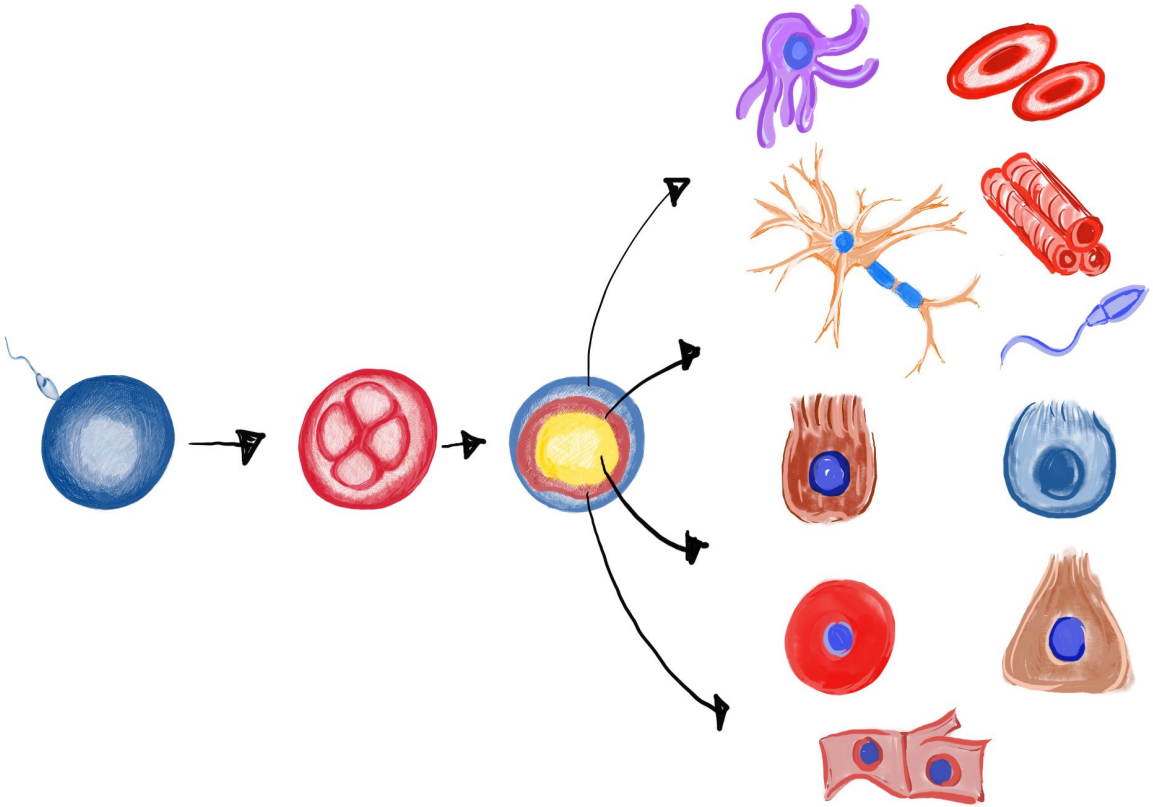
Maria Abou Chakra

Cell development simulation model

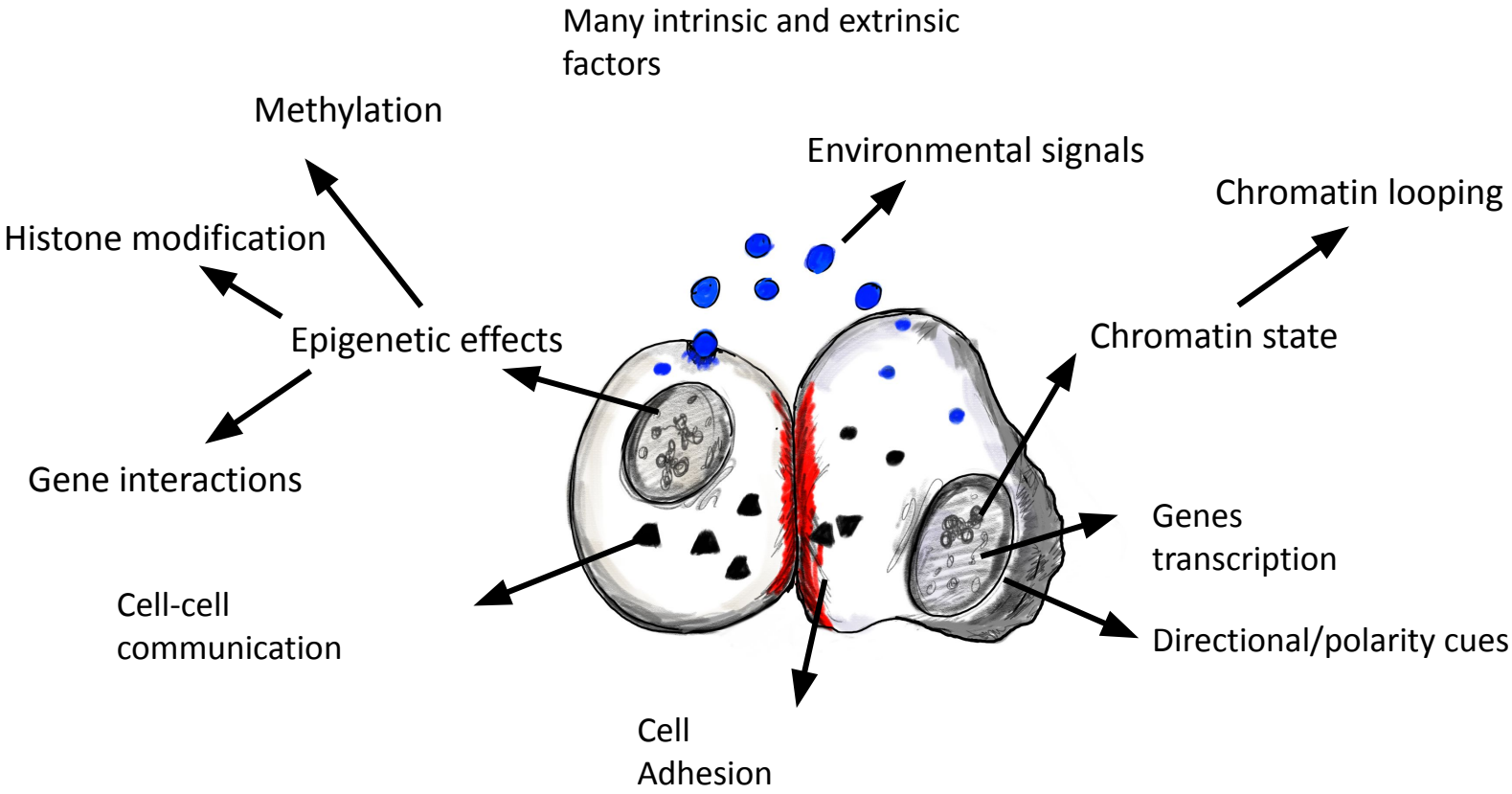
2D and 3D cell modeling



Cell diversity arises from a single cell



What controls cell diversification?

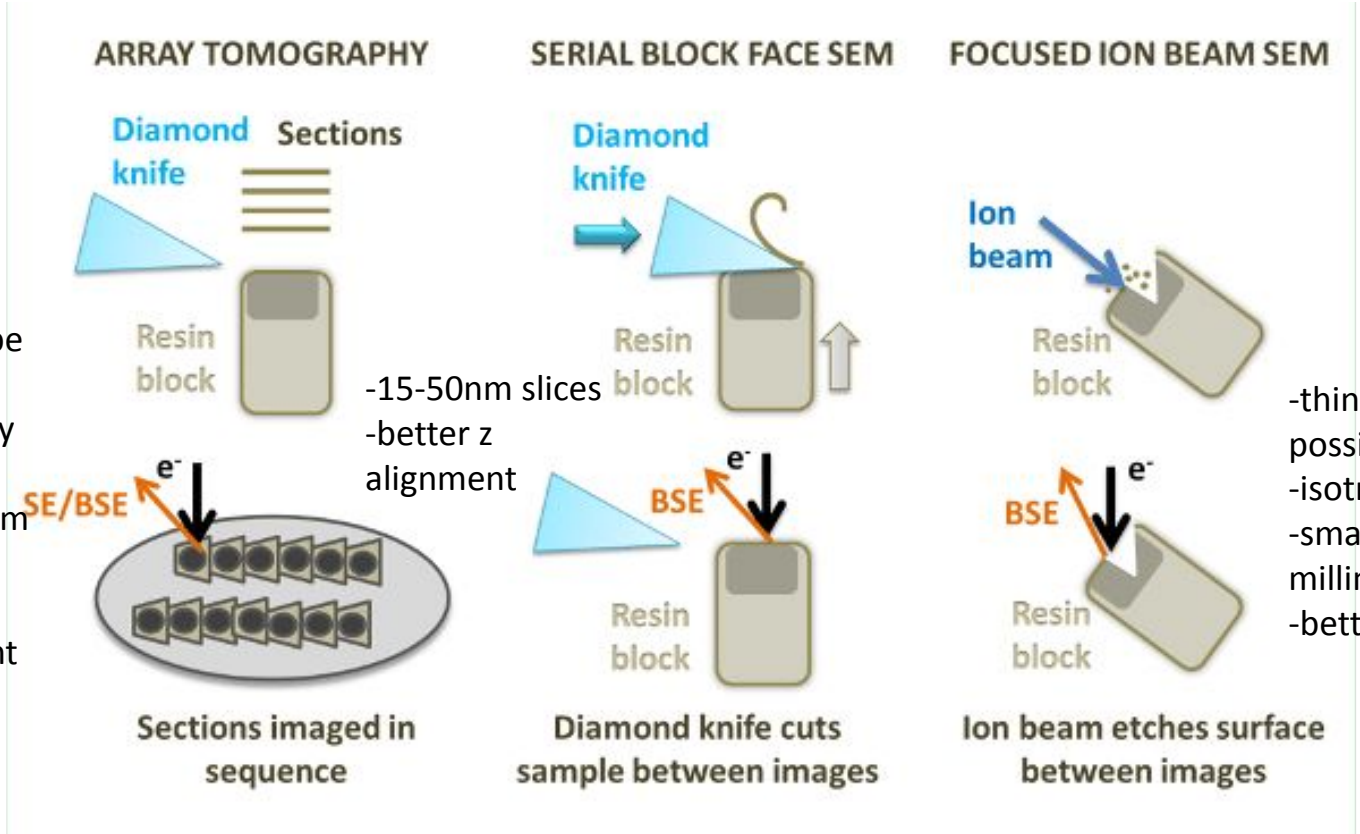


Volumetric nanoscale imaging

Mei Zhen

ATUM-SEM
automated tape
-collection
ultramicrotomy

-Slices: 30-40nm limit
-easy staining
-slice alignment problems

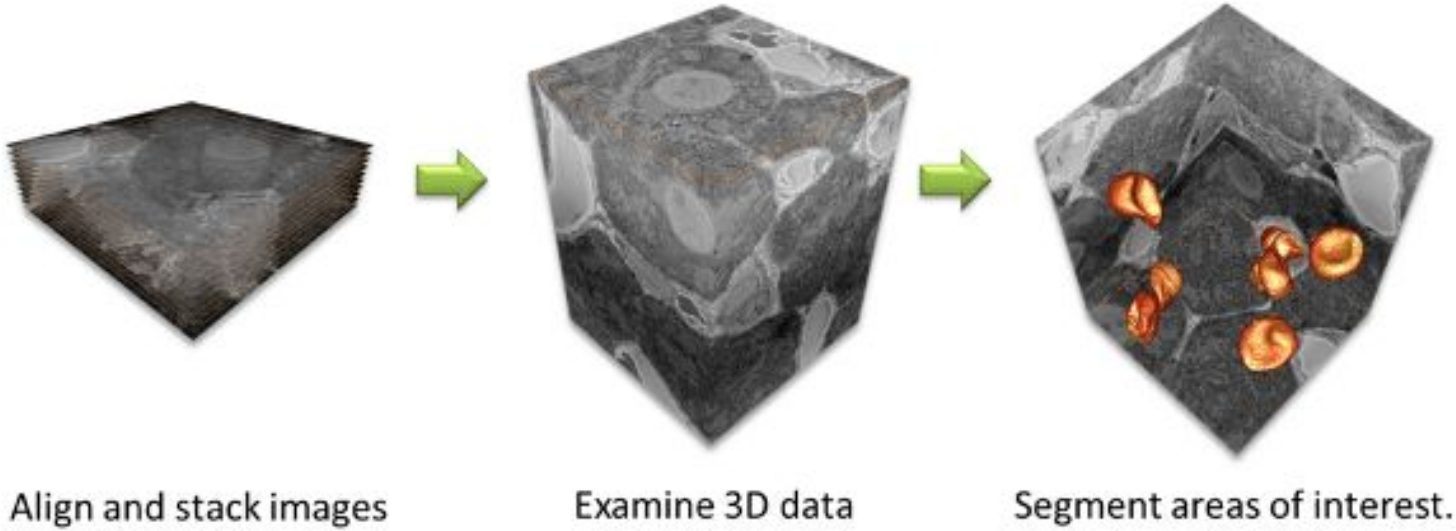


-15-50nm slices
-better z alignment

-thin slices: 3-5nm possible
-isotropic voxels
-small volume due to milling time limits
-better

https://bitesizebio.com/30796/three-dimensional-scanning-electron-microscopy-for-biology/

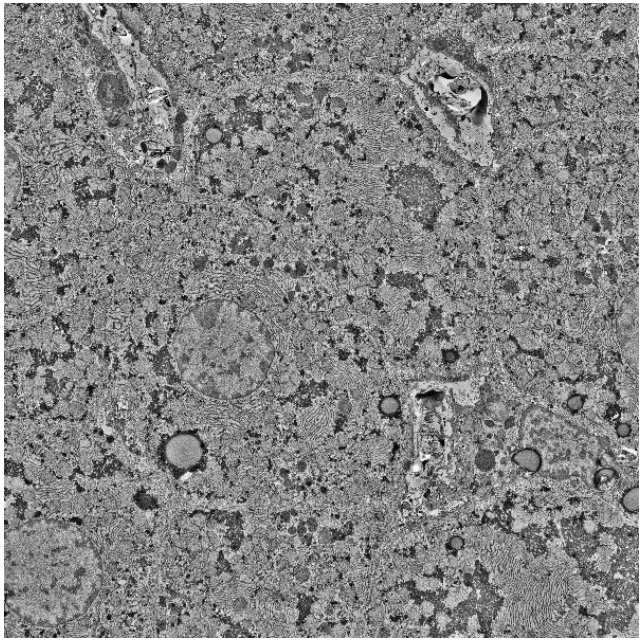
VEM data processing



https://bitesizebio.com/30796/three-dimensional-scanning-electron-microscopy-for-biology/

Ultrastructural mapping of the liver

- Generated at LTRI in September 2020 to optimize staining methods
- 32x32x10um tissue derived from Rat liver lobule
- 4000x4000x219 pixels
- 8x8x50nm resolution
- 10.7GB in size



Ali Darbandi, Ben Mulcahy, Mei Zhen, Ronald Xie, Sonya MacParland

Ideas for discussion - what to include in the proposal?

Mapping the body: What scales and modes are missing/under-represented? (e.g. tissues)

What modes/perspectives exist and how should we integrate them?

Ecosystem analogy - cellular ecosystem, systems analysis, dynamic systems, etc.

How can we define ecosystem functions, like wound response? Analogy with the protein sequence world - network view. How are cellular ecosystems adaptable?

Cooperation to achieve goals e.g. the immune system, brain - could this influence society?

Interdisciplinary approach: which ones to include?

How do we map (define) a multi-scale system? parts, interactions, temporal behaviour - what else? E.g. Do we want to map response to perturbation?

Sid Goyal

Levels of abstraction - different models at different scales

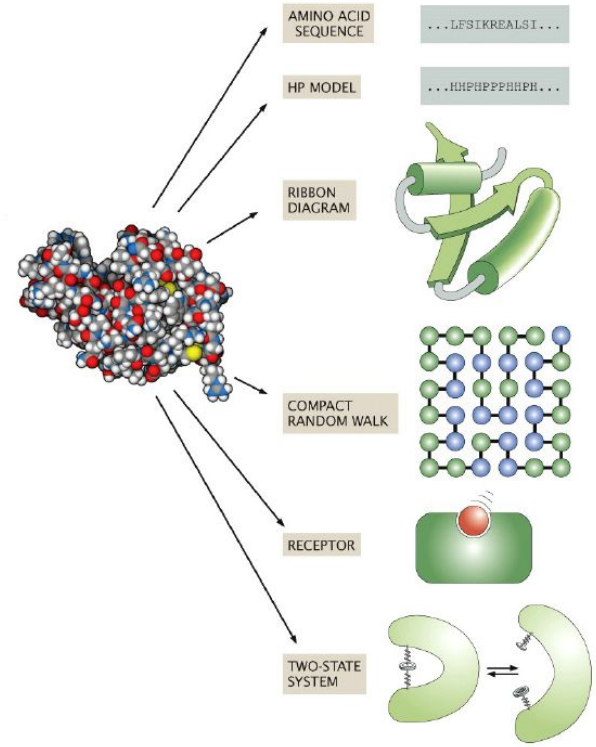


Figure 1.6 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Protein folding

Structure

?

Reactions

Function

Can we develop models at different scales for cells, tissues, organs…

Structure

Cell types requires molecular underpinning

- Metabolic **states** [1]
- Genetic drivers [2]
- Epigenetic **markers**…

Clustering

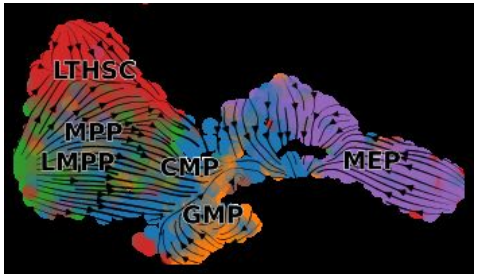
[1] Jatav, BioRxiv 2021

[2] Freedman, BioRxiv 2022

Tissues requires coordination among its constituent cell types to define “tissue states”

- **Flux** of cell types
- **Dynamics** of cell fate

Hydrodynamic theory of tissues?



??

Function

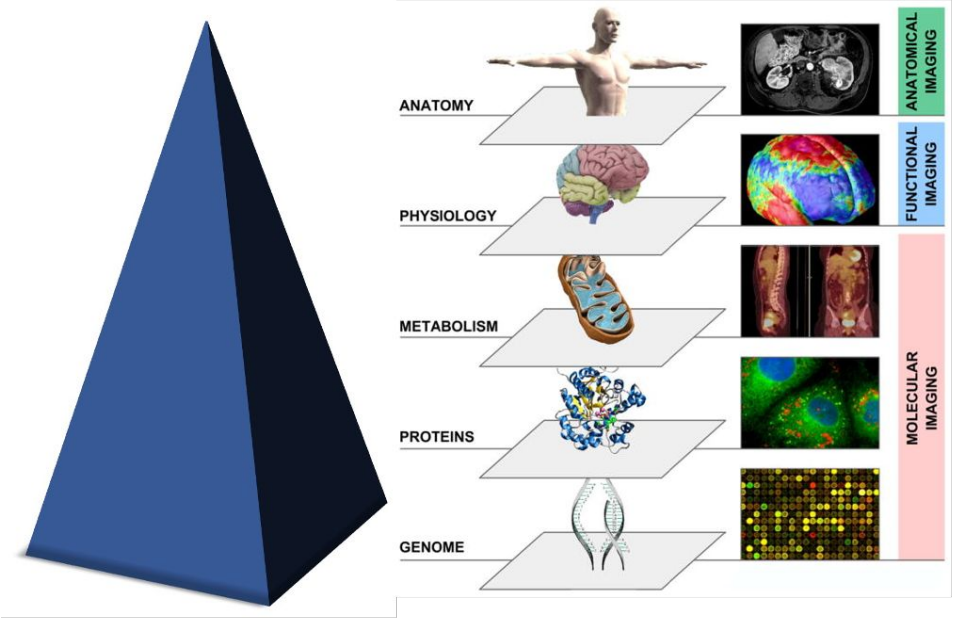
Organs combine different tissues in functionally diverse ways

- Same tissue in different contexts, e.g. epithelial?
- Function specific tissues, like insulin producing cells, or common structures such as branching in lungs, kidneys and glands?

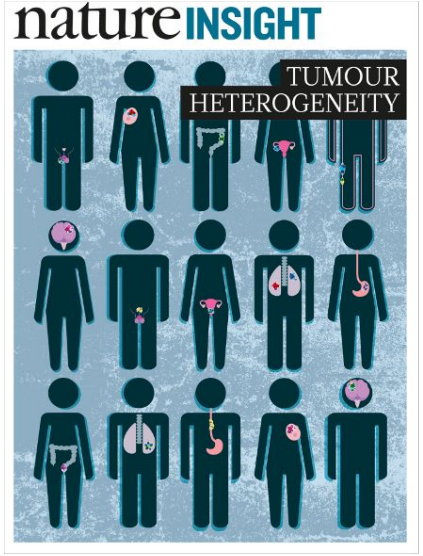
Ferdia Gallagher

fag1000@cam.ac.uk

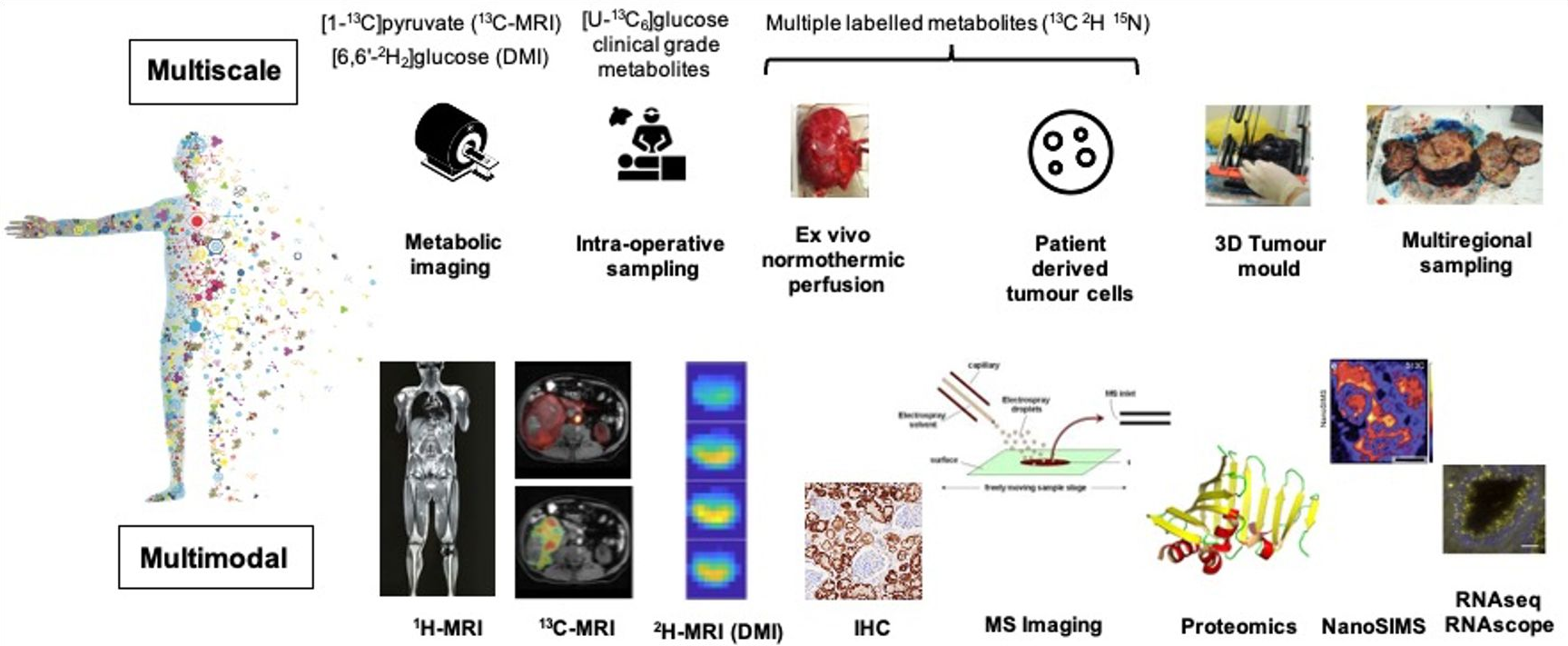
Imaging as a non-invasive mapping tool to bridge scales



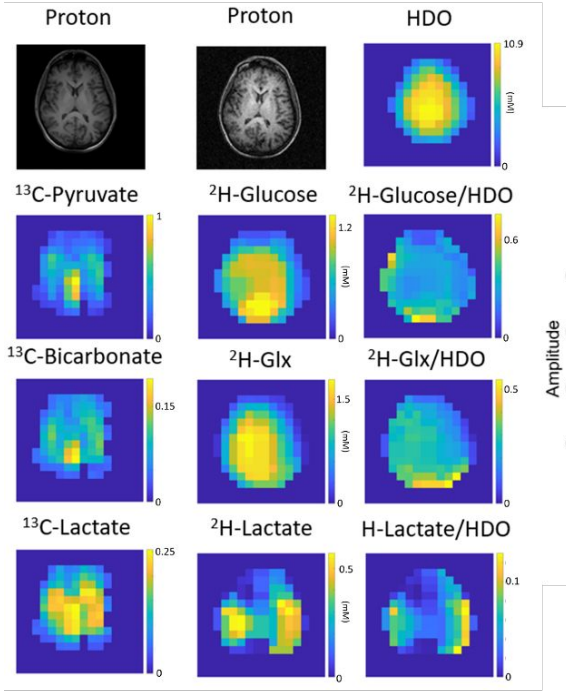
Imaging as a tool to assess heterogeneity

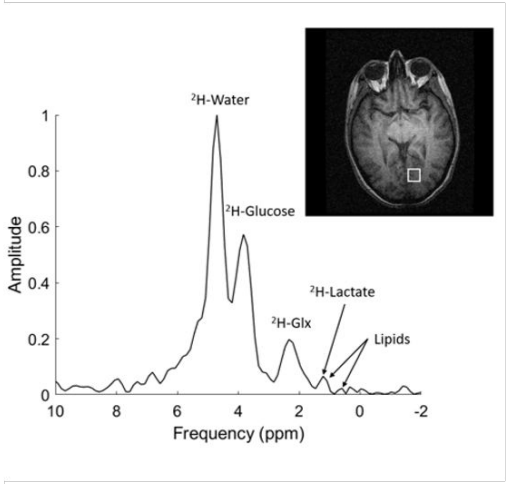


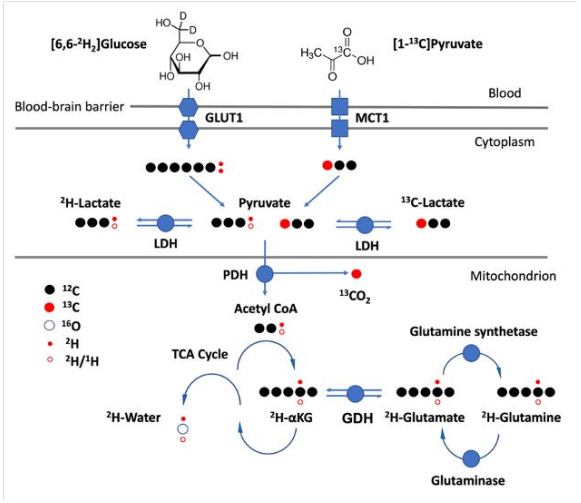
Multiscale and multimodal approaches: linking imaging to other modalities



New imaging methods to phenotype tissue e.g. metabolism







Carbon-13 labelled pyruvate and deuterium labelled glucose to probe differential oxidative and non-oxidative metabolism in the normal human brain

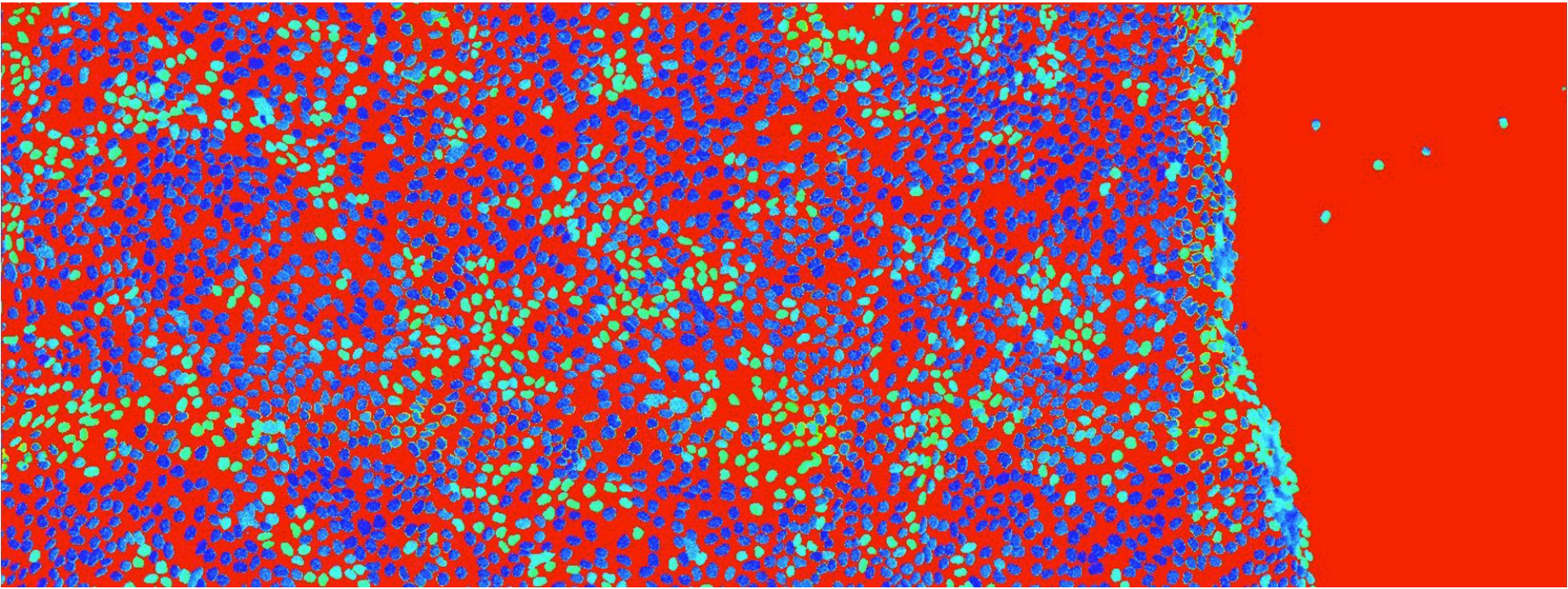
Non-invasive imaging methods at a whole organism level are key to linking biology to structure and function across a wide range of scales

Barbara Engelhardt

Statistics and Machine Learning
Gladstone Institutes
Stanford University
barbarae@stanford.edu

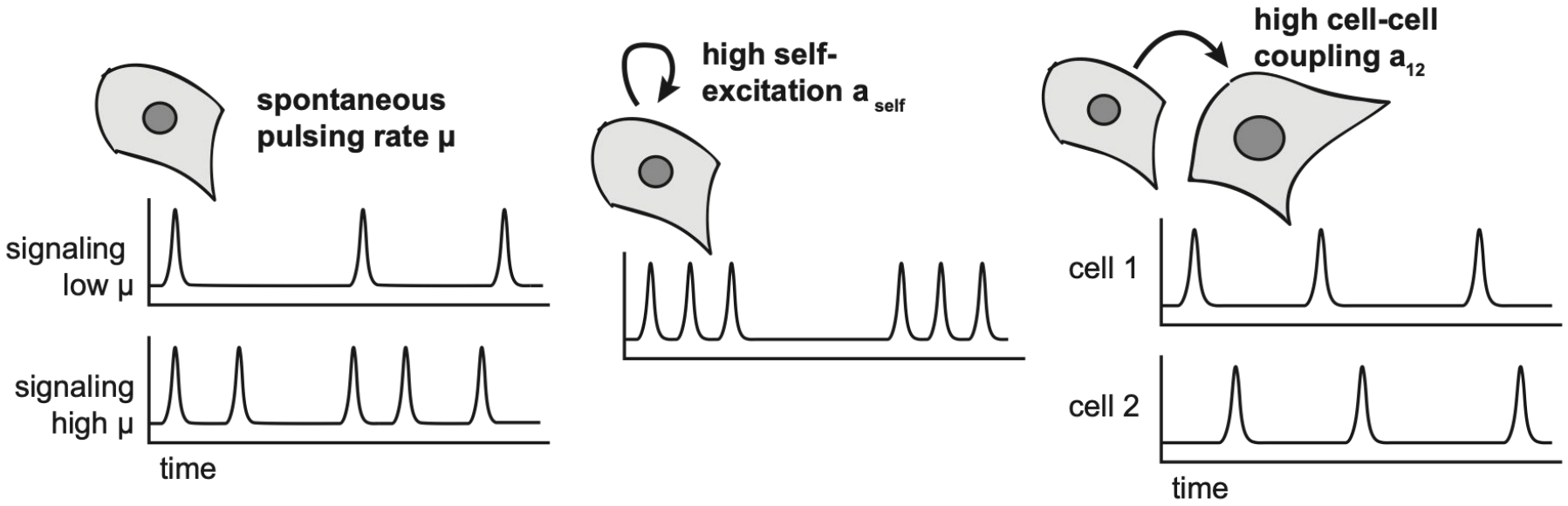
Spatiotemporal models of cell-cell signaling

Blue: Low ERK signaling; Green: High ERK signaling; Red: background



[Verma et al. 2021], work with Toettcher Lab

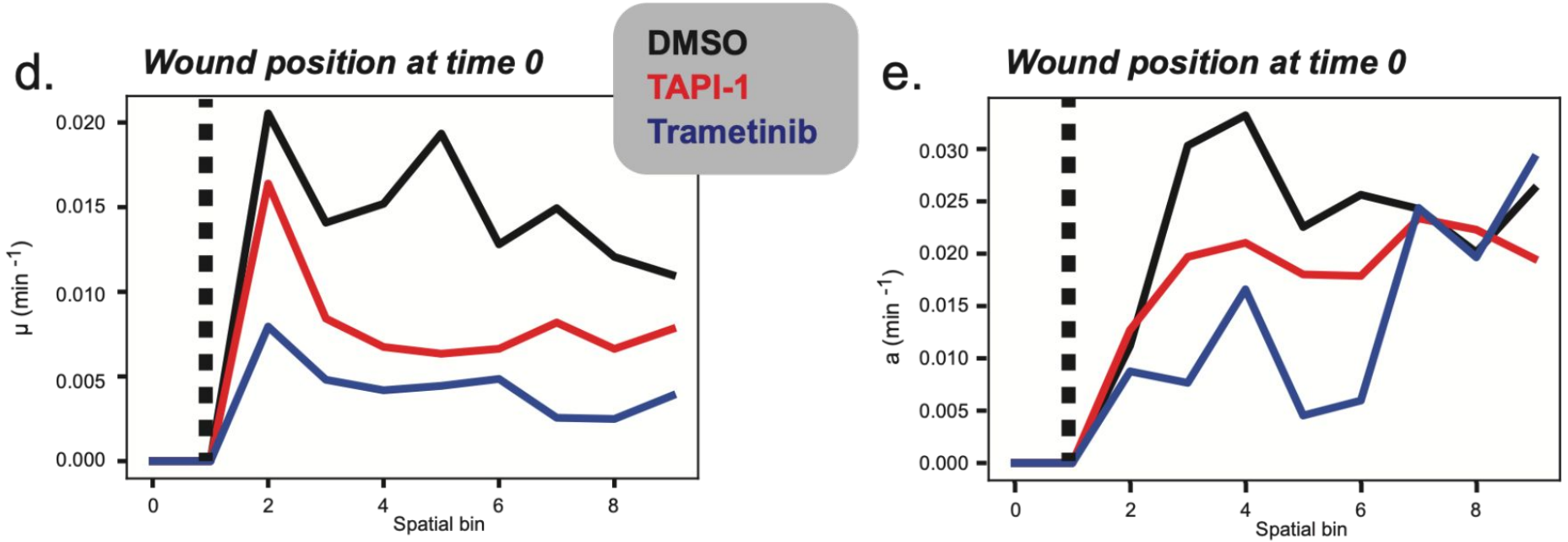
Deconvolving the three sources of pulsing



Hawkes process model to deconvolute three separate rates of pulsing; can consider multiple markers and cell types

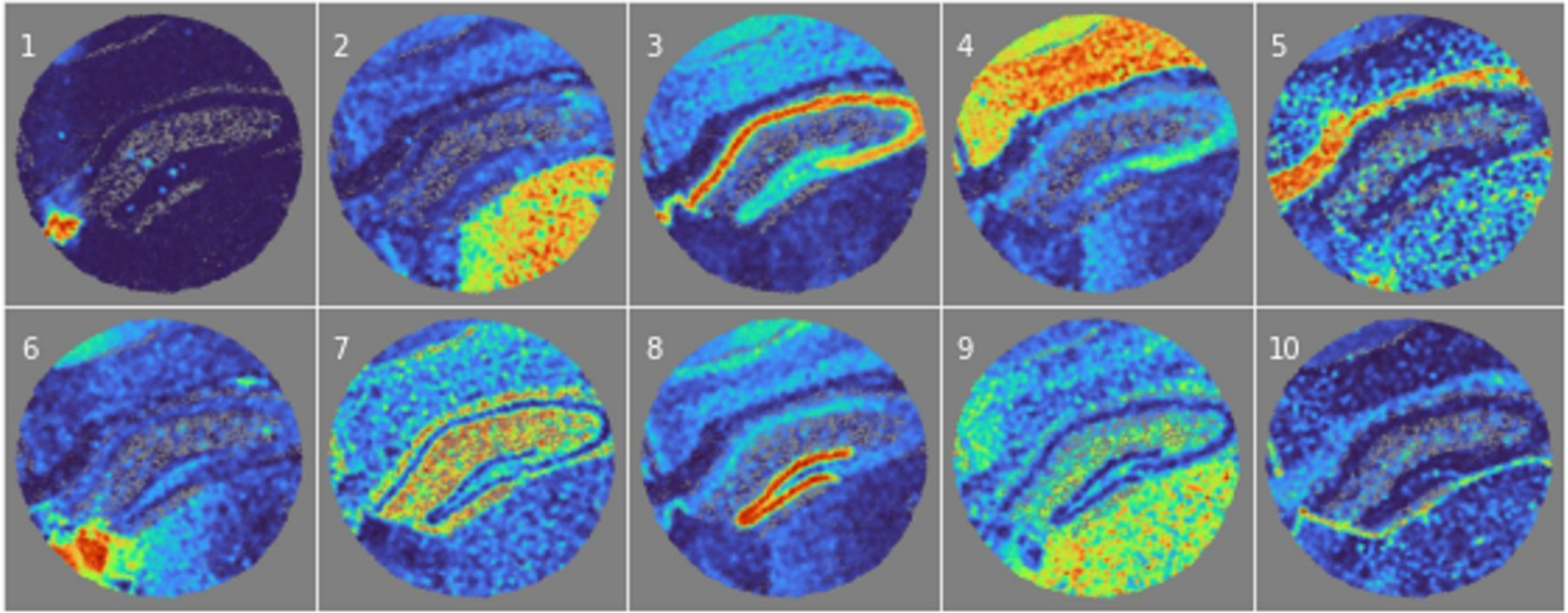
Joint work with Jared Toettcher's lab

Cell signaling: distance from wound, interventions



Joint work with Jared Toettcher's lab

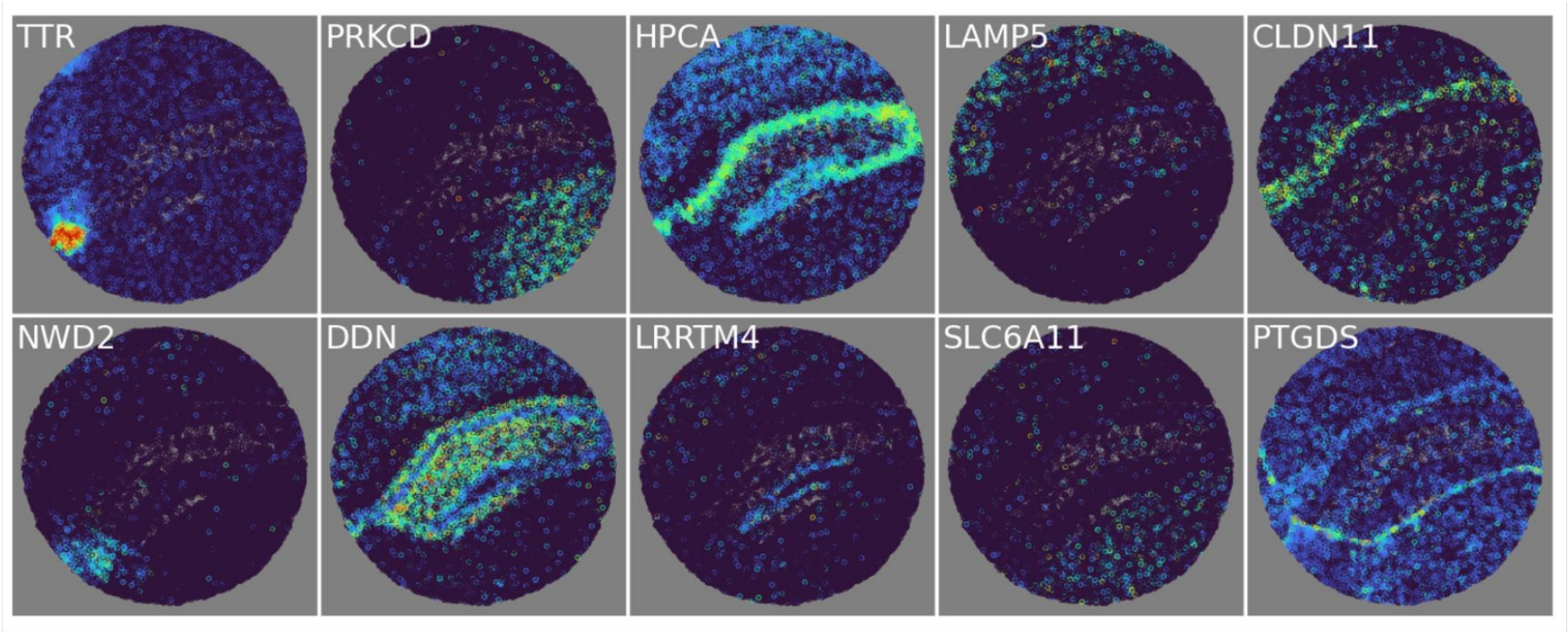
Spatial dimension reduction on spatial transcriptomics data



1. Choroid plexus of third ventricle
2. Thalamus
3. CA1-3 pyramidal layer
4. Cerebral cortex
5. Fiber tracts/corpus callosum
6. Medial habenula
7. CA strata and dendrite gyrus
8. Dentate gyrus granule layer
9. Multiple layers
10. Meninges

Slide-seqV2 on mouse hippocampus sample *[Stickels et al. 2021]*

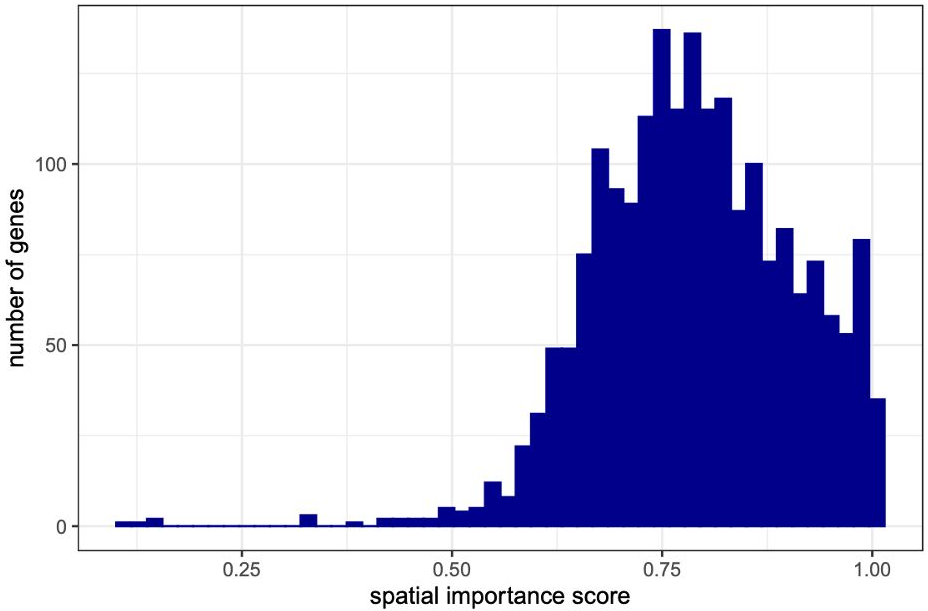
Top genes for each of the factors – many markers

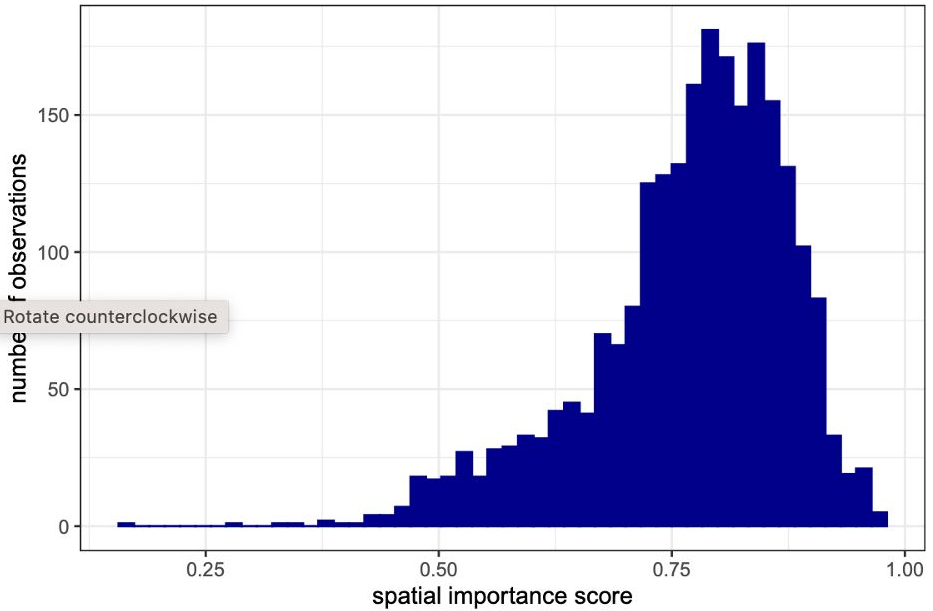


1. Choroid plexus of third ventricle
2. Thalamus
3. CA1-3 pyramidal layer
4. Cerebral cortex
5. Fiber tracts/corpus callosum
6. Medial habenula
7. CA strata and dendrite gyrus
8. Dentate gyrus granule layer
9. Multiple layers
10. Meninges

Slide-seqV2 on mouse hippocampus sample *[Stickels et al. 2021]*

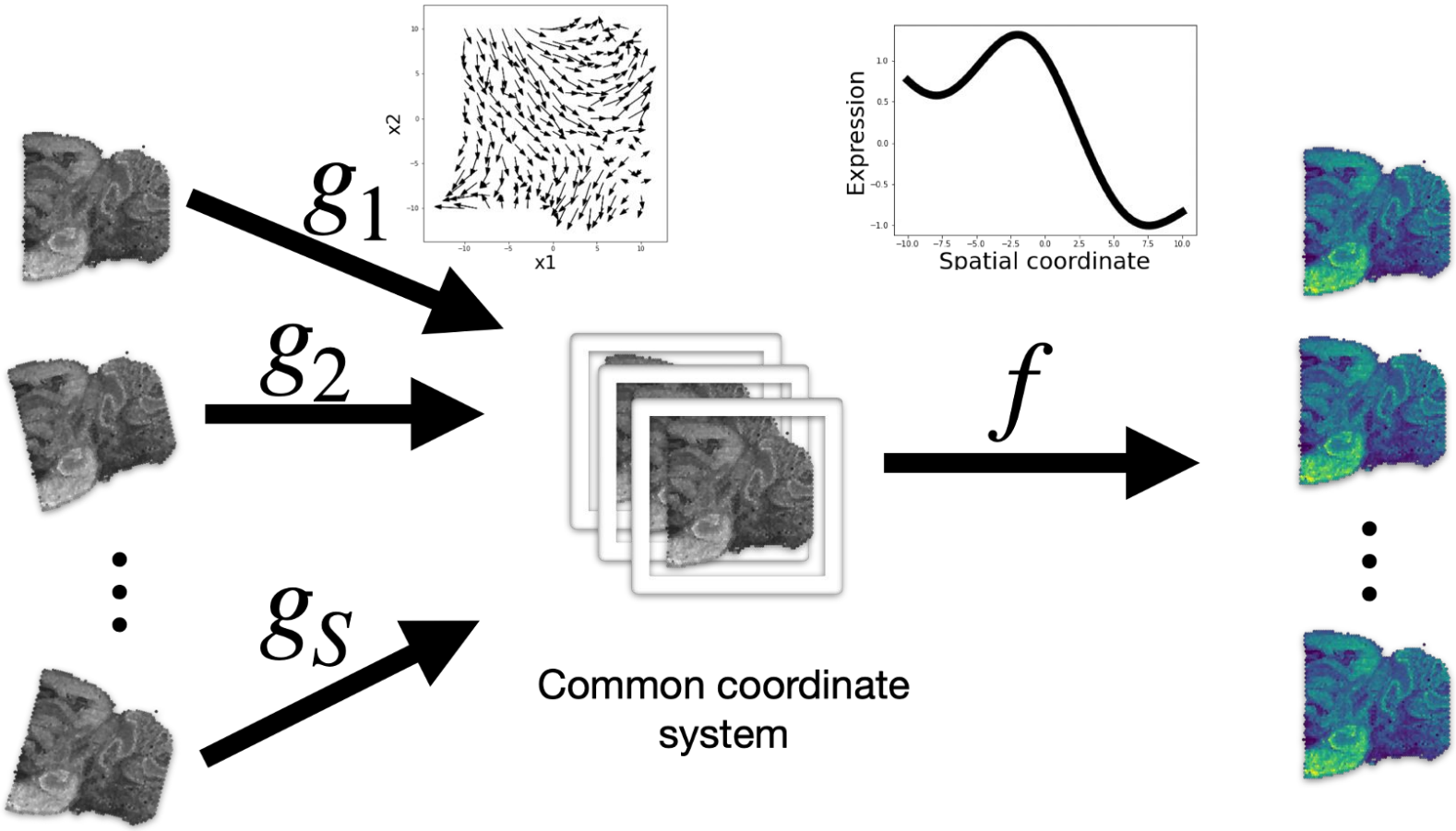
Intrinsic vs extrinsic cell states and gene expression



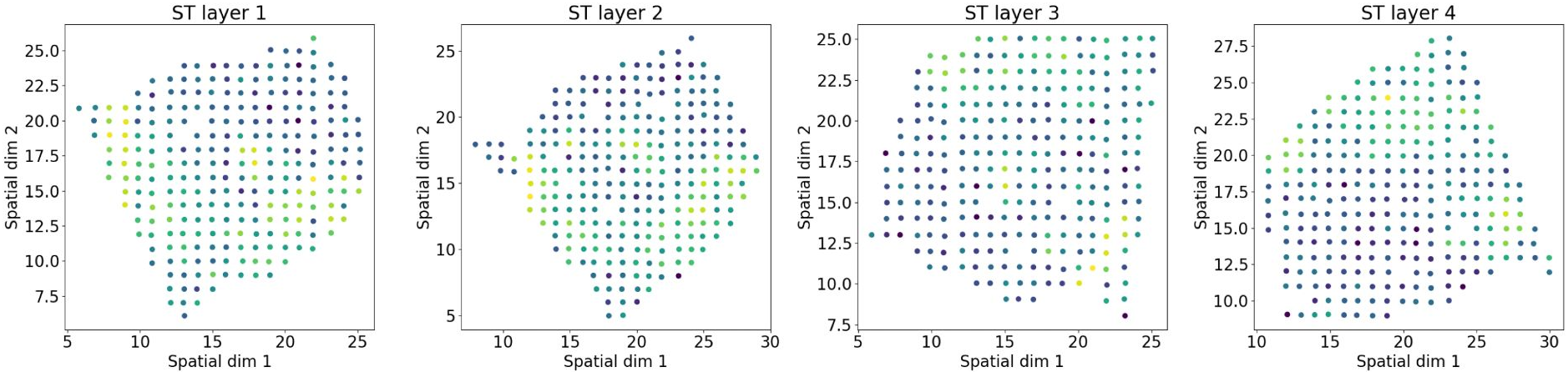


What proportion of variation in gene expression (left) or cell state (right) is explained by spatial variation?

Alignment: warping plus expression. Atlas?

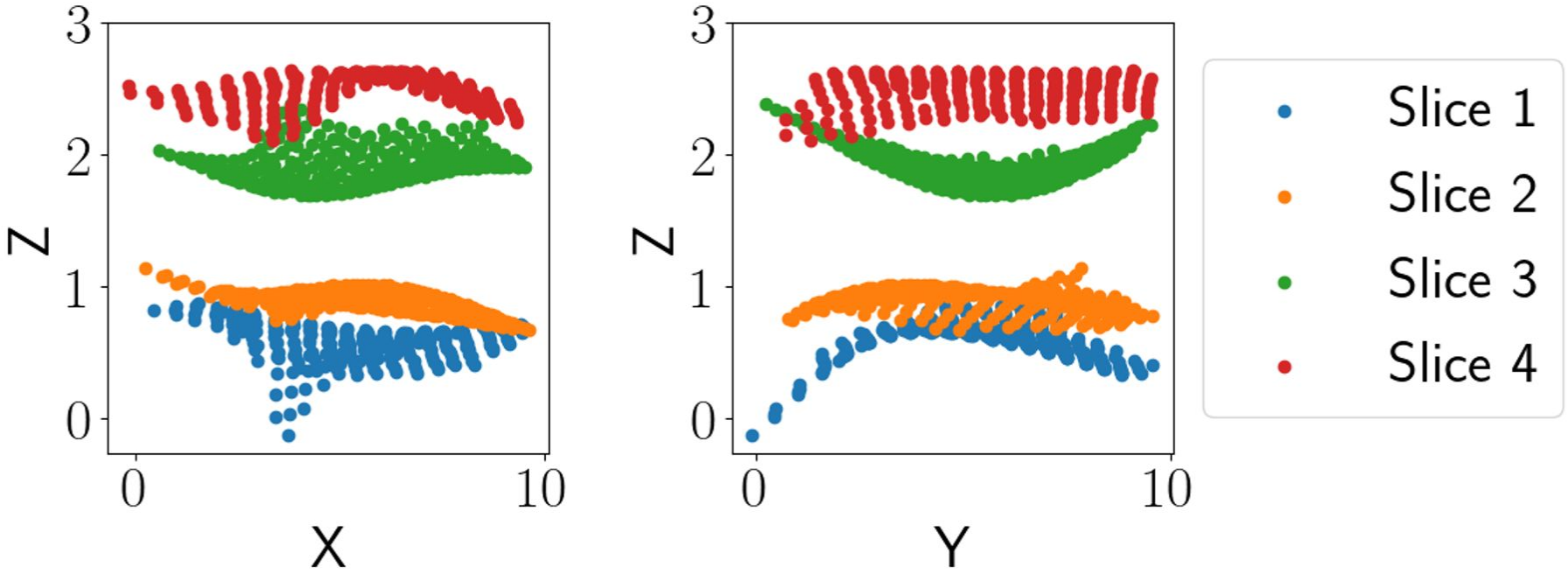


Aligning four parallel slices of a breast cancer biopsy



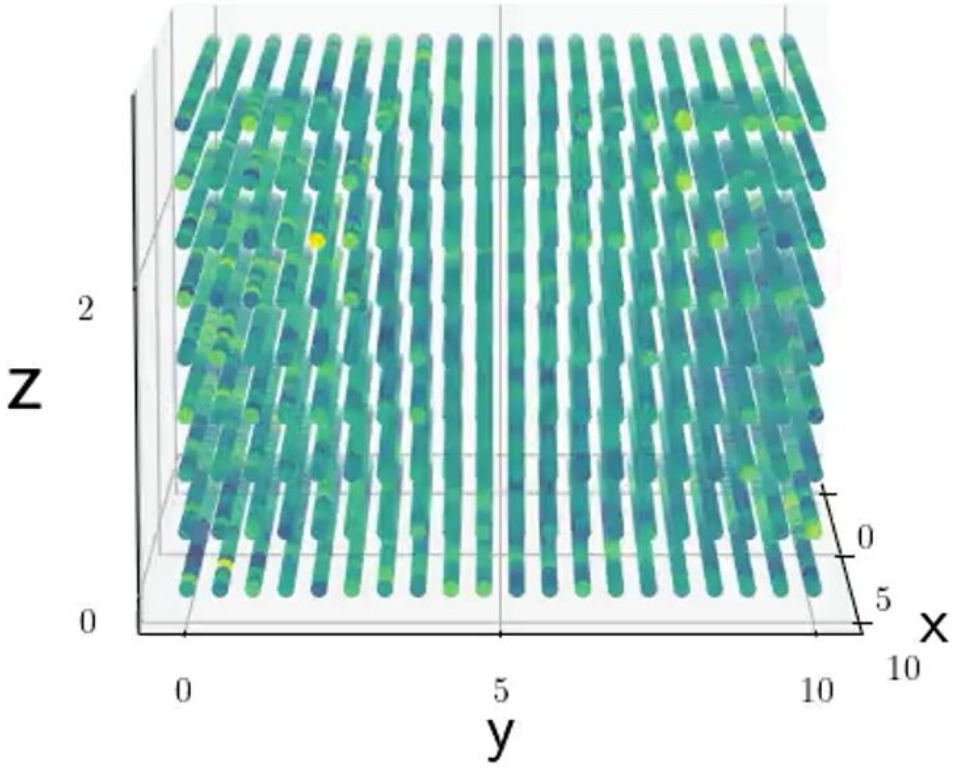
Four parallel slices of a breast cancer biopsy with Spatial Transcriptomics *[Stahl et al. 2016]*

Warping four slices in three dimensions with GPSA



Four parallel slices of a breast cancer biopsy with Spatial Transcriptomics *[Stahl et al. 2016]*

Single 3D breast tumor; *FN1* expression

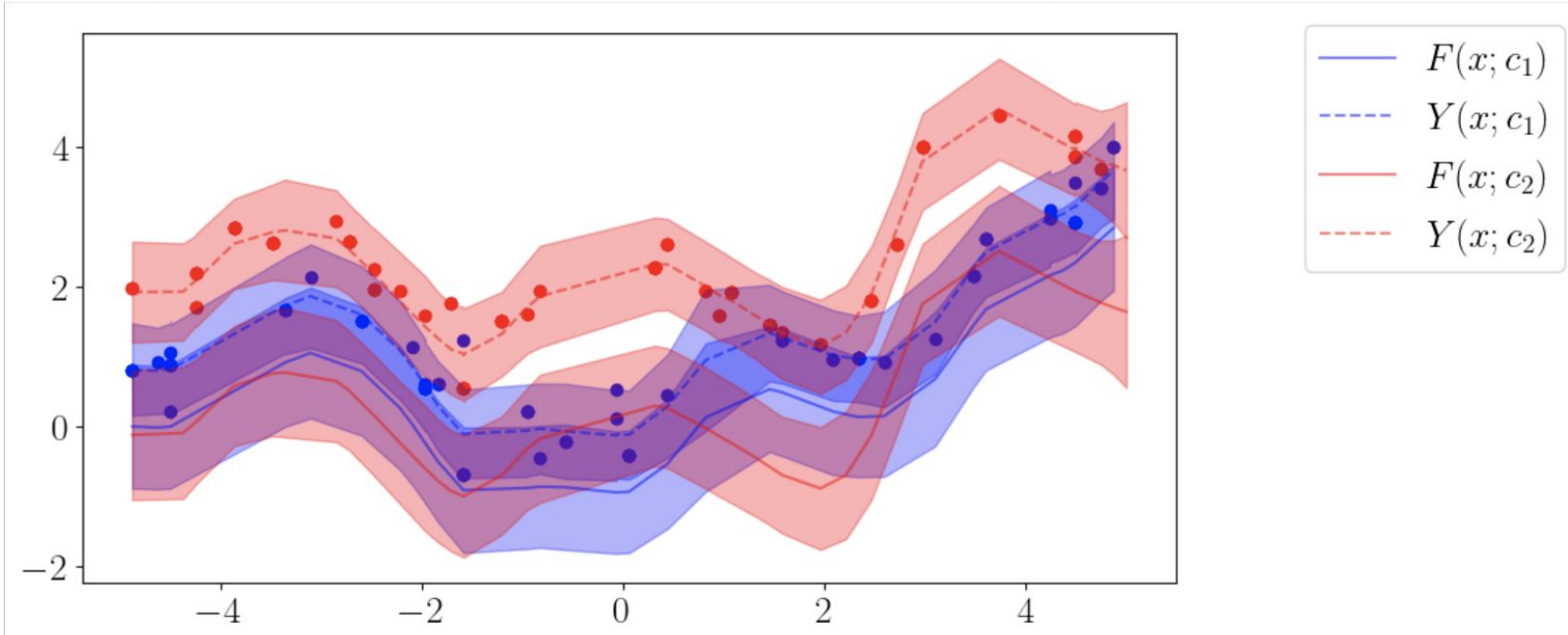


Multigroup Gaussian Processes (MGGP)

Goal: to jointly model continuous and categorical variables, capturing complex dependencies between both.

Examples: CRISPR perturbations with single cell RNA-seq; single-cell RNA-seq data with cell type, batch labels.

Model: Carefully constructed covariance function combining real and the categorical variable domains.



Flexible, robust, structured, and interpretable statistical modeling is one key to multi-scale, multi-modality, multi-technology, and noncanonical mapping and atlassing of the human body.