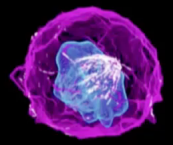
ALLEN INSTITUTE *for*
CELL SCIENCE



Toward a holistic and dynamic stem cell state landscape

Susanne Rafelski

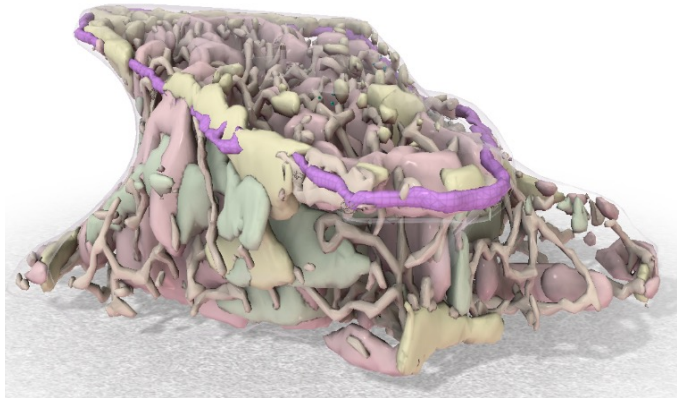
Developing Regenerative Medicine Therapies withAI
November 2025

Disclosure slide: no disclosures

Our vision: *understand how cells organize, communicate and change*

Look at a cell and know what it is doing

…what it did



what it will do…

cell → **tissue** → **organ** → **organ system** → **organism**

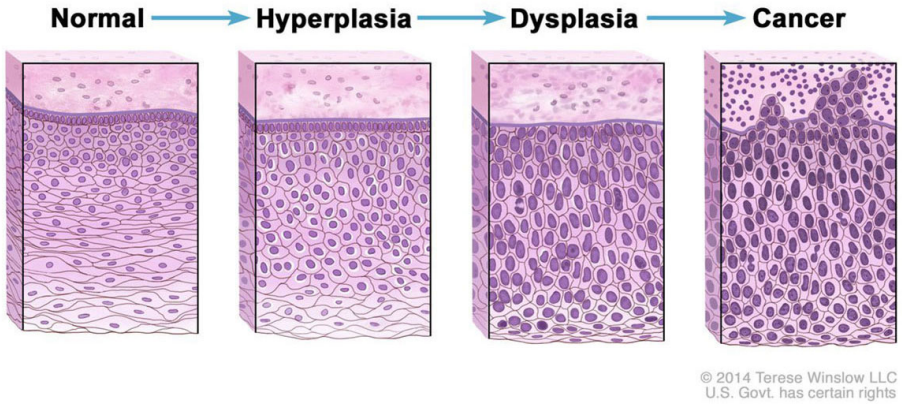
“know”

beyond prediction only – we want understanding!

Our vision: *understand how cells organize, communicate and change*

Look at a cell and know what it is doing

…what it did



what it will do…

cell → **tissue** → **organ** → **organ system** → **organism**

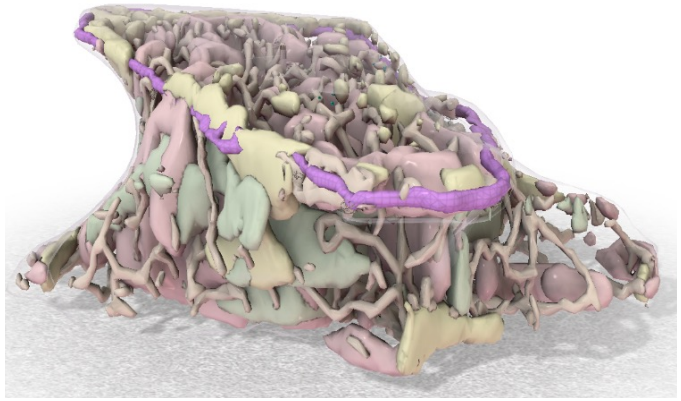
normal vs. abnormal cellular basis of damage and disease

variability within cell populations

Our vision: *understand how cells organize, communicate and change*

Look at a cell and know what it is doing

…what it did



what it will do…

How?

images

through time & across scales

Which images?

Talk outline

1: Bringing together spatiotemporal and molecular representations of cell state in the current era of big cell image data and AI

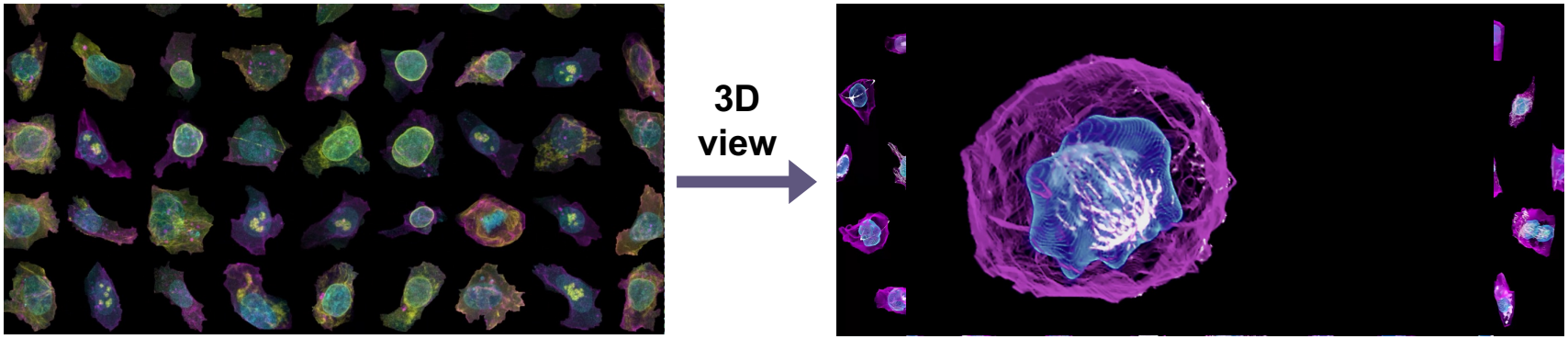
perspective: *Establishing a conceptual framework for holistic cell states and state transitions*
*Rafelski and Theriot, Cell **187**, 2633–2651 (2024)*

2: Where we have been: 10 years of the Allen Institute for Cell Science

3: Where we are heading: our new initiative – CellScapes + a case study

Single cell imaging and analysis: the next step in the post-genomic era

We are in an era of large-scale cell image datasets
and we want to make sense of them through quantitative (image) data science methods



What do we need to "know" a cell?

cell "observable": anything we can observe (measure, see) about a cell

what state a cell is in

a range of values for each observable that describes a cell in a relatively stable state

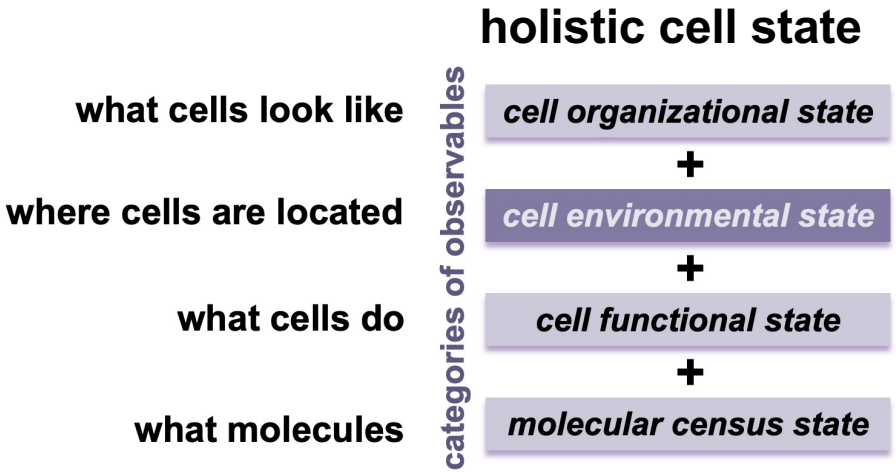
how and why a cell changes state

the dynamics of how these observables change as a cell leaves one state and settles into a new state

What do we need to "know" a cell?

Now is the time to update the concept of a *holistic cell state*

quantitative, spatiotemporal observables + molecular observables

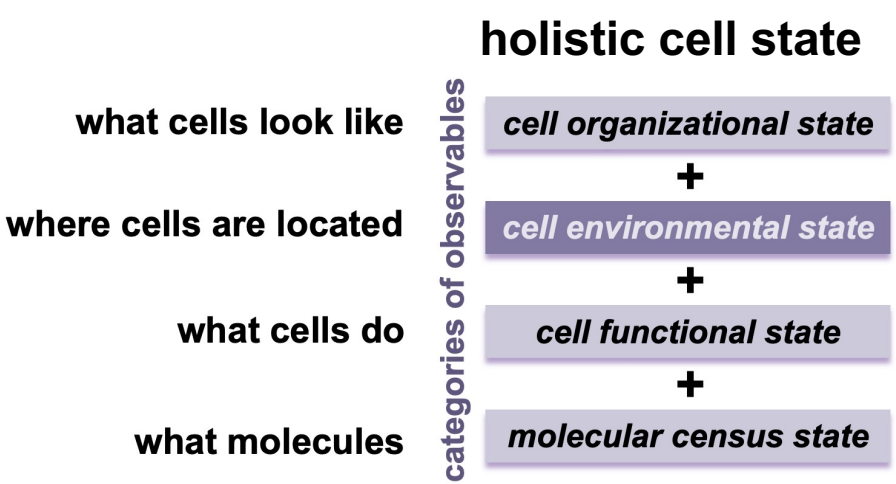


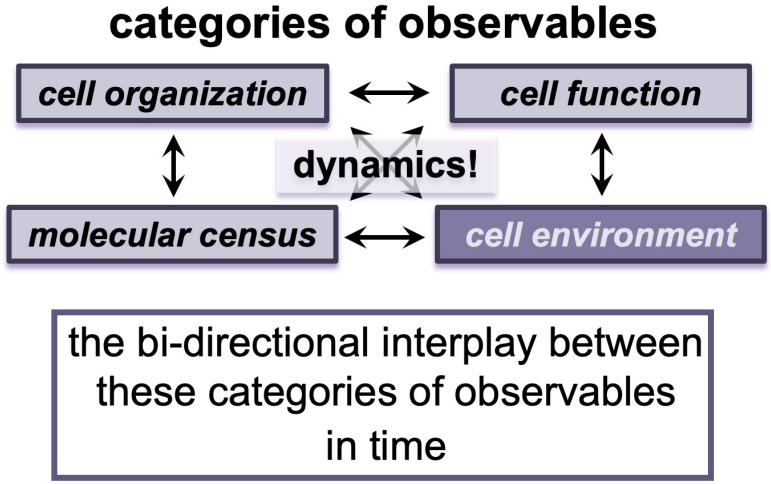
Perspective: Rafelski and Theriot, Cell 187, 2633–2651 (2024)

What do we need to "know" a cell?

Now is the time to update the concept of a *holistic cell state*

quantitative, spatiotemporal observables + molecular observables

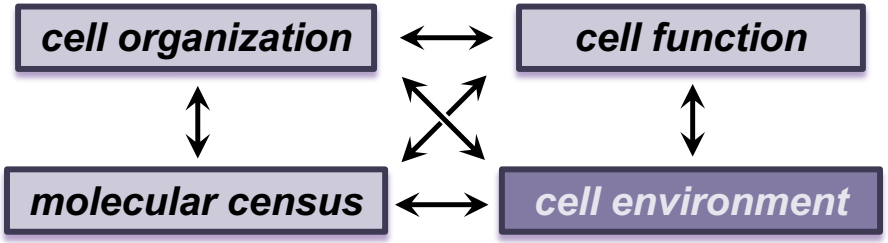




Perspective: Rafelski and Theriot, Cell 187, 2633–2651 (2024)

A conceptual *holistic cell state* framework

categories of observables

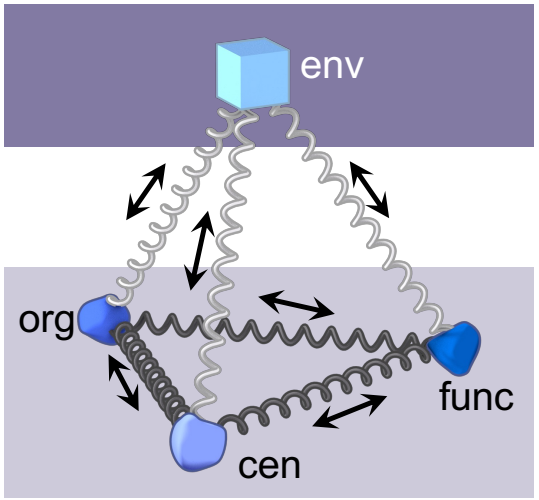


representations:

- ***vertices***: the four categories of observables
- ***springs***: bi-directional feedback between categories
- ***spring-induced movement***: the "push and pull" of the values of one observable category in response to another

conceptual visualization

spring-connected tetrahedron

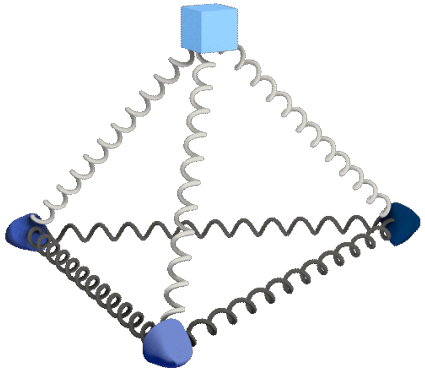


cellular phenotype (*cell-intrinsic)*

cellular environment (cell-extrinsic)

A conceptual *holistic cell state* framework – stable state

In a **stable holistic cell state**
the vertices may fluctuate…



for example:

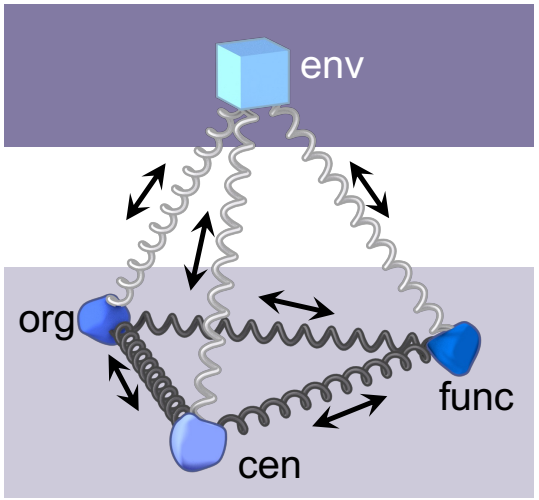
- *levels of a particular protein*
- *position of a particular organelle*
- *speed of a migrating cell*

can vary

… but ultimately the nature of the relationships between the vertices ***mutually reinforces*** them to maintain a stable holistic cell state

conceptual visualization

spring-connected tetrahedron

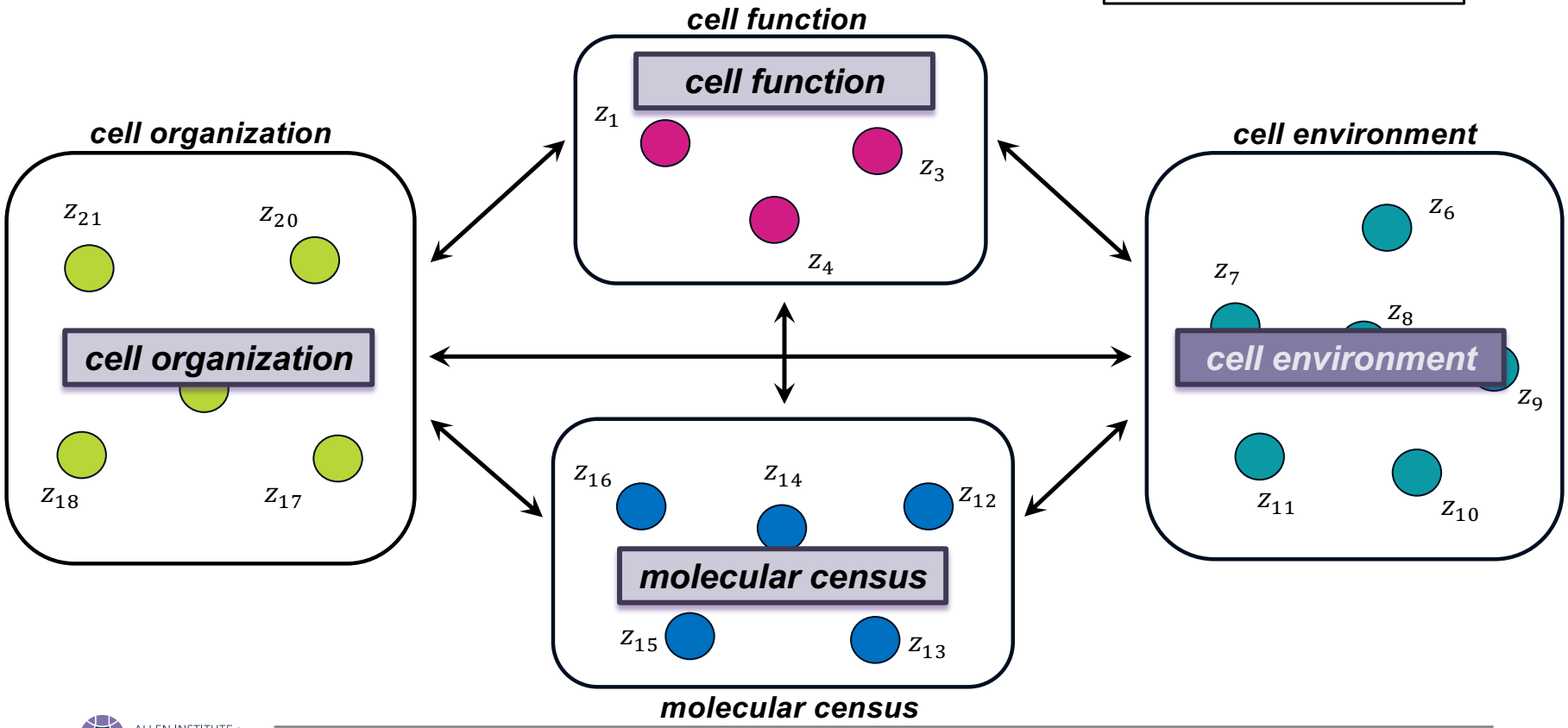


cellular phenotype (*cell-intrinsic)*

cellular environment (cell-extrinsic)

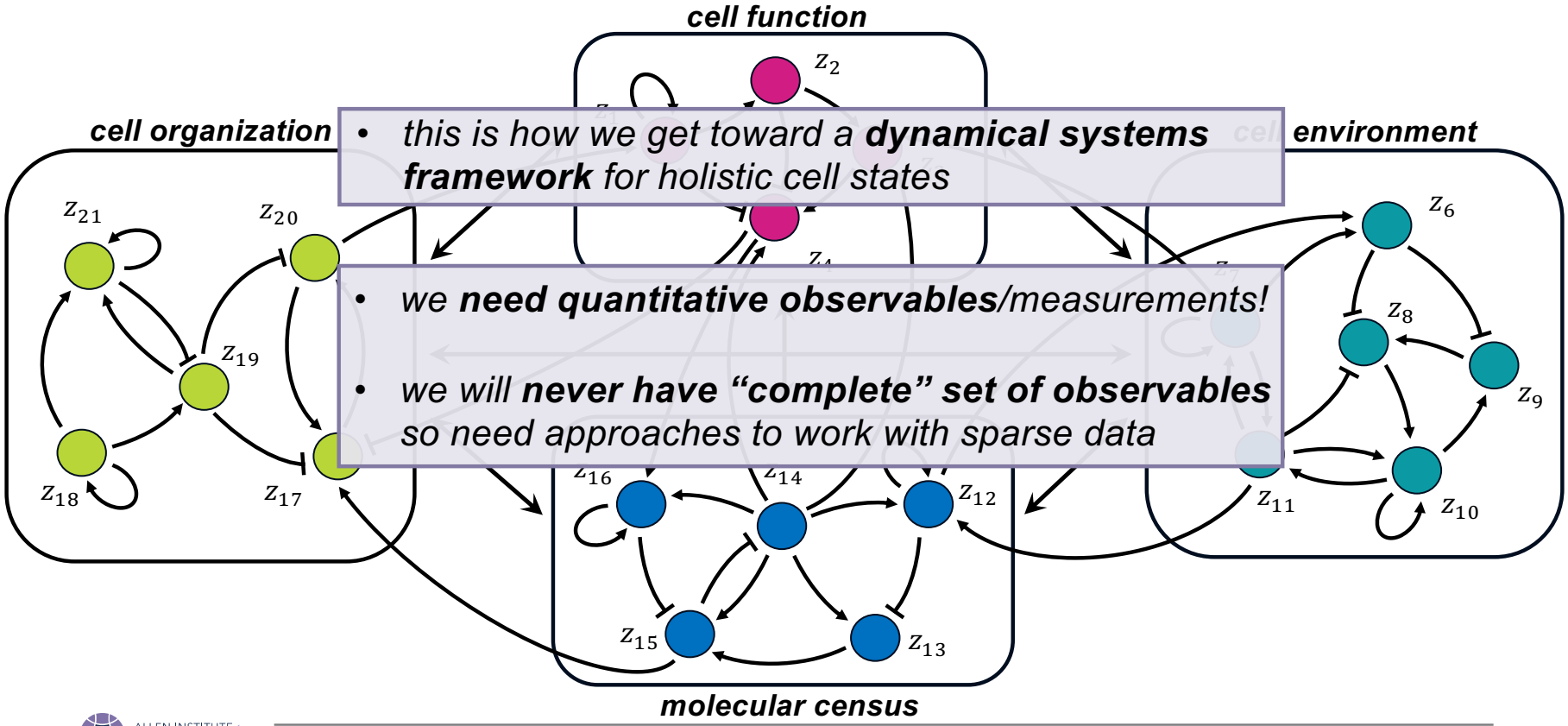
A conceptual *holistic cell state* framework – vector of all observables

$$\boldsymbol{z} = (z_1, z_2, \cdots, z_N)$$



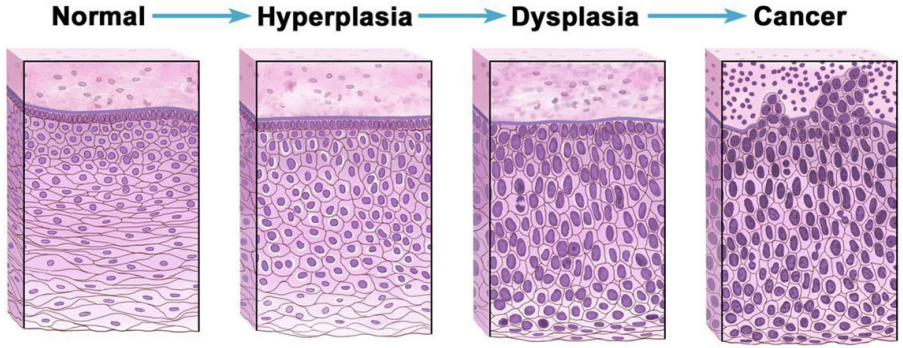
A conceptual *holistic cell state* framework – bidirectional feedback

(network diagram)



- *this is how we get toward a **dynamical systems framework** for holistic cell states*
- *we **need quantitative observables**/measurements!*
- *we will **never have "complete" set of observables** so need approaches to work with sparse data*

To get at the mechanisms underlying normal cell states, cellular damage, and disease we need to integrate across categories of observables…



… via quantitative frameworks

molecular representations:
e.g., bulk or single cell genomics

all the molecules (and how many) inside a cell

organizational representations?

the spatio-temporal, multi-scale ***arrangement***
of the molecules inside a cell

holistic cell representations

we will never have **complete** data → **partial** data

How does AI fit in?

- **dimensionality reduction:** detecting/learning patterns in highly complex data and creating ***holistic cell representations***, including over time
- **generative AI:** recreating patterns from these representations toward ***interpretation*** *and* ***insights***
- **goal:** ***predictive understanding***, not black-box prediction alone

Talk outline

1: Bringing together spatiotemporal and molecular representations of cell state in the current era of big cell image data and AI

2: Where we have been: 10 years of the Allen Institute for Cell Science

Developing quantitative frameworks for cell organization and beyond

- *overview of (some of) our publications and Open Science resources*

3: Where we are heading: our new initiative – CellScapes + a case study

The genomic era transformed our access to molecular sequence data

Single cell imaging and analysis: the next step in the post-genomic era

Human genome project launched in 1990
30 years of tool and method development
data is very straightforward to analyze

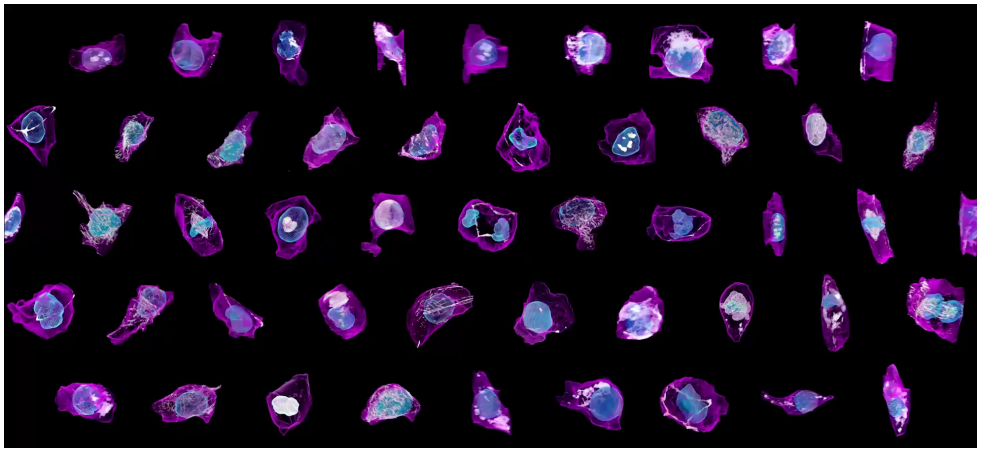
sequence alignment to find mutations

```
5′ACCTGTCTTGAAACTGTGTC 3′
5′ACGTGTCTTGACACTGTGTC 3′
```

"Cell x Gene" table for RNA expression

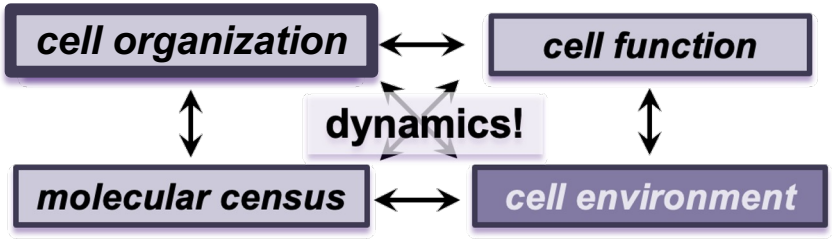
	Cell 1	Cell 2
Gene A	574	581
Gene B	23	15
Gene C	1068	23
Gene D	75	69
Gene E	1654	1702

Grand challenge: to establish a quantitative framework to compare the shapes and positions and their changes over time for 3D cellular structures in a comprehensible and generalizable manner



Single cell imaging and analysis: the next step in the post-genomic era

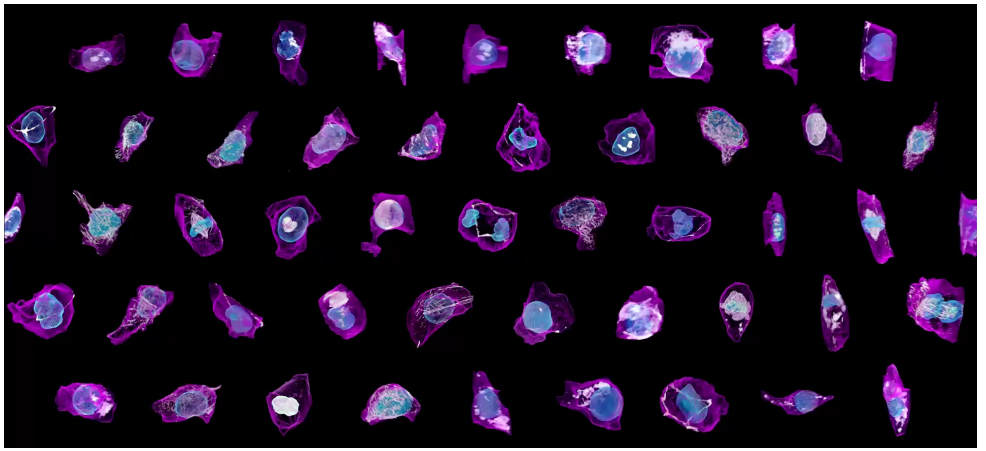
the field needs a quantitative framework for *cell organization*



to integrate with other categories of cell observables such as:

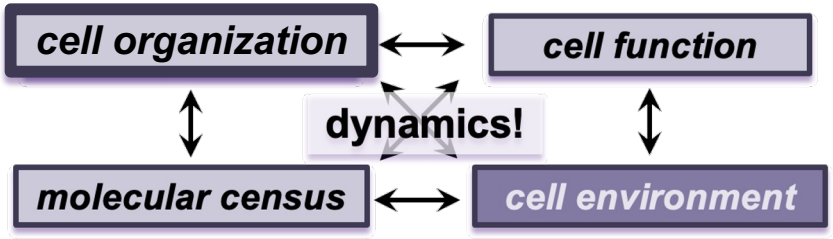
expression signatures
protein states (signaling)
chromatin organization

Grand challenge: to establish a quantitative framework to compare the shapes and positions and their changes over time for 3D cellular structures in a comprehensible and generalizable manner



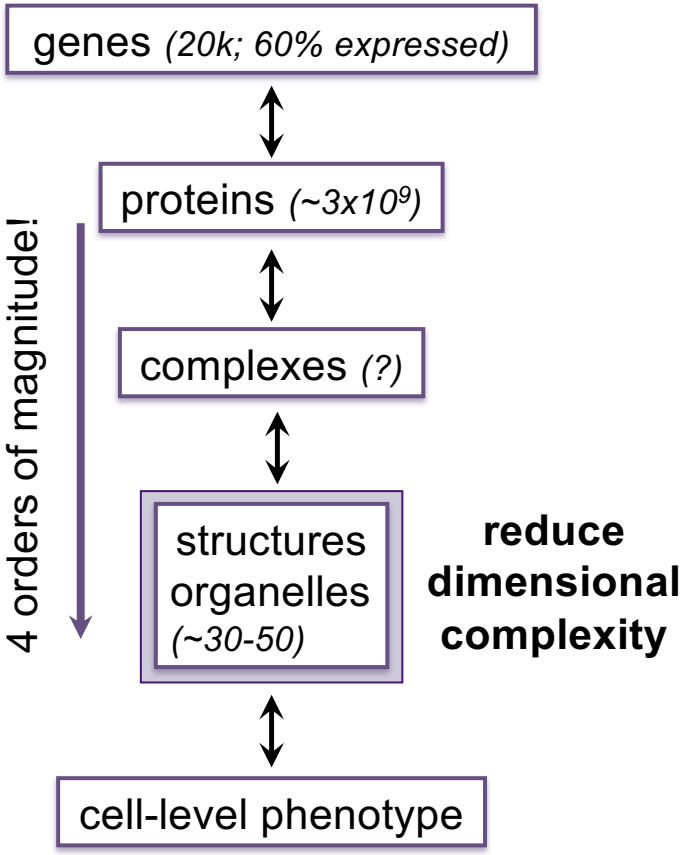
Single cell imaging and analysis: the next step in the post-genomic era

the field needs a quantitative framework for *cell organization*



to integrate with other categories of cell observables such as:

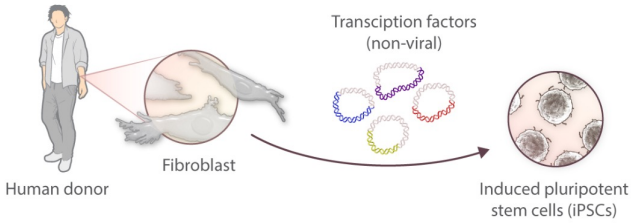
expression signatures
protein states (signaling)
chromatin organization



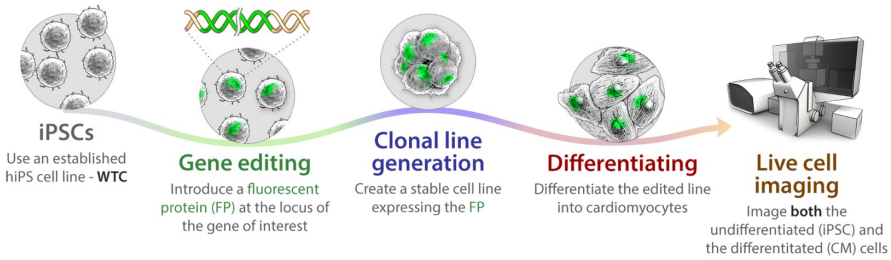
First 10 years – developing frameworks for cell organization and beyond

overview of our approach

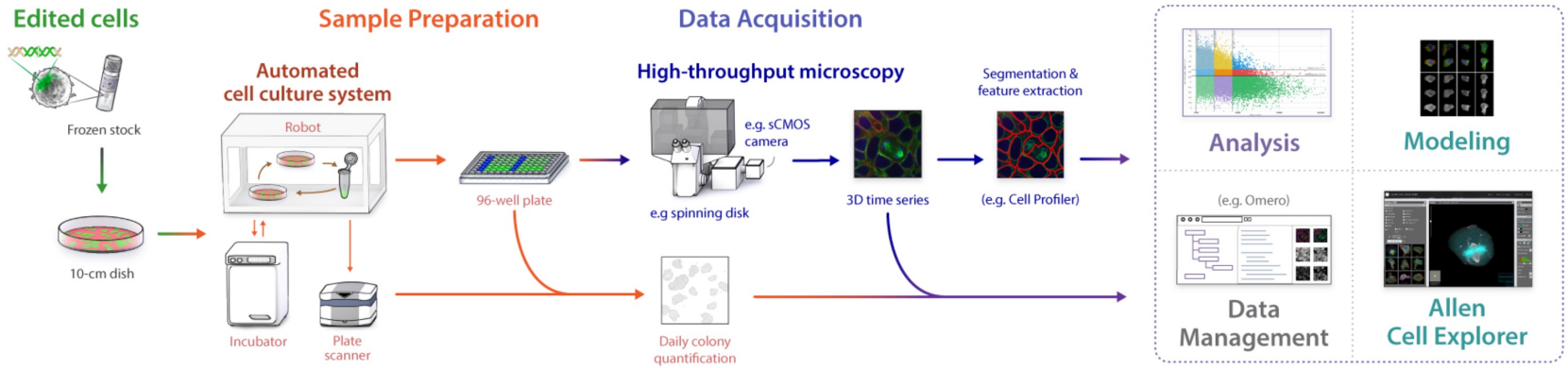
pre-clinical model: hiPS cells (WTC-11)



illuminating cellular structures in hiPS cells



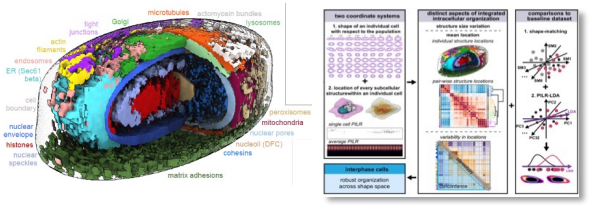
creating and analyzing large 3D image datasets of hiPS cells



First 10 years – developing frameworks for cell organization and beyond

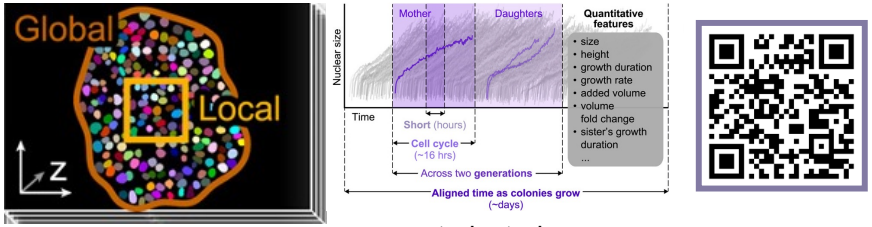
overview of (some of) our scientific publications

integrated hiPSC organization and variations (Nature 2023)



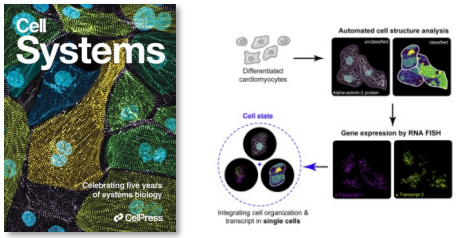
coordinate system

variations in nuclear growth dynamics across space and time (Cell Systems 2025)



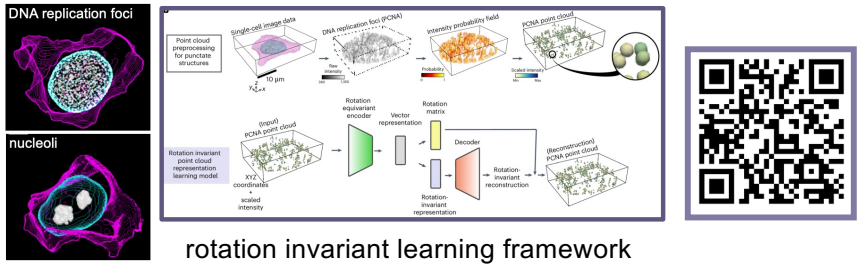
trajectories

correlating organization with gene expression in cardiomyocytes (Cell Systems 2021)



single cells and time points

interpretable representations for multipiece cellular structures (Nature Methods 2025)



rotation invariant learning framework

Tools & resources overview

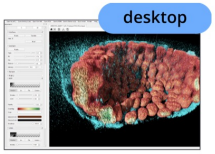
The Allen Institute is committed to open science, providing researchers, educators, and the public with access to our data, tools, methods, and analyses for reproduction and reuse.

allencell.org

Jump to section: Image analysis & visualization | Data support | Code | Cell lines

Sign up for our newsletter or join our forum to stay connected.

Image analysis & visualization



visualization

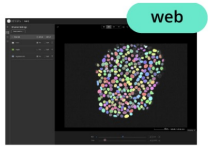
AGAVE
Powerful volume viewer with path-trace rendering to display multi-channel TIFF and OME-Zarr images

analysis visualization

Allen Cell & Structure Segmenter
napari plugin & open-source Python-based toolkit for 3D image segmentation

analysis visualization

Timelapse Feature Explorer (TFE)
Online tool to interactively visualize & analyze segmented time-series microscopy data

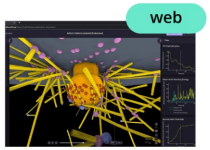


visualization

Vol-E
Online volume viewer to visualize your own volumetric data including OME-Zarr format

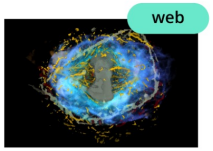
visualization education

Cell Feature Explorer (CFE)
Online tool composed of a scatterplot & 3D viewer to visualize segmented and processed hiPSC images from our open datasets



visualization education

Simularium
Online analysis tool to visualize, share and interrogate biological simulations



visualization education

Integrated Mitotic Stem Cell (IMSC)
Online educational tool that uses real-world data to provide a holistic view of human cell division

visualization education

Visual Guide to Human Cell
Online educational tool to interact with 3D cell models and learn about individual cell structures and functions

Cell lines

cell lines

Allen Cell Catalog
Quality control database of gene-edited hiPS cell lines

cell lines

Disease Cell Collection
Quality control database of gene-edited cell lines that carry mutations in genes known to cause disease

cell lines

Lab Plasmids
Deposited plasmids at Addgene for distribution to the research community

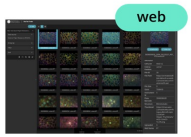
cell lines

Standard operating procedures (SOPs)
Cell culture, differentiation, and microscopy protocols for gene-edited hiPSCs

Data support

python data support

BioIO
Image reading, metadata conversion, and image writing for microscopy images in pure Python



data management

BioFile Finder (BFF)
Online tool to curate, interrogate, and share datasets through rich metadata exploration & image viewing

data download

Data Download
Open-access datasets for download

Code

machine learning

CytoDL
Framework to unify and simplify deep learning approaches for analyzing biological data

python analysis

Cell Variance Analysis
Processing of FOVs and cells for the Cell Variance Analysis

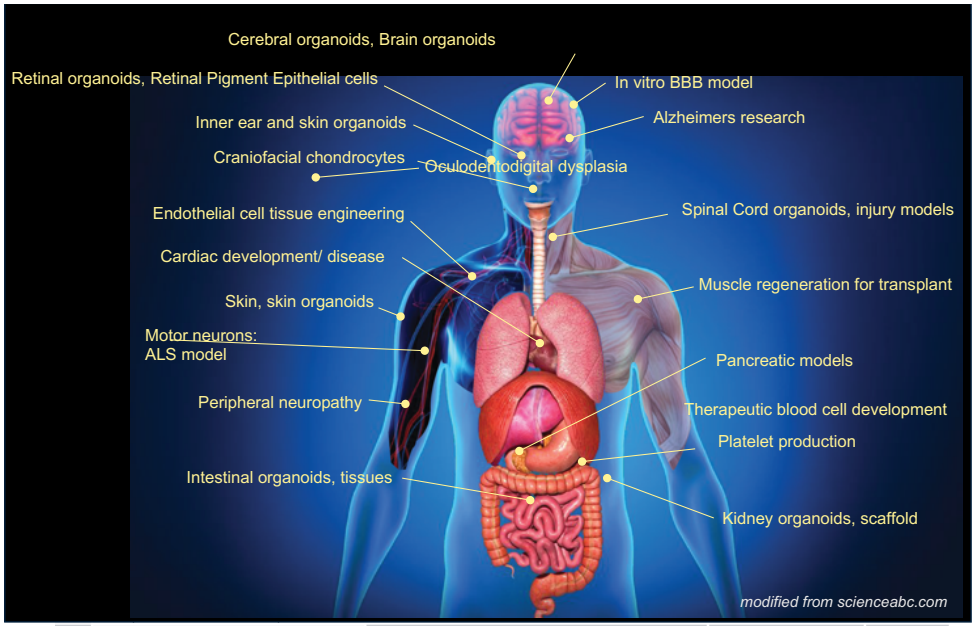
analysis

Other Codebases
Open-access code developed at the institute

The Allen Cell Collection – isogenic FP tagged hiPSC's for live cell imaging

	Structure	Protein	FP	Status
1	Matrix adhesions	Paxillin	EGFP	Available
2	Microtubules	Alpha-tubulin	mEGFP	Available
3	Nuclear envelope	Lamin B1	mEGFP	Available
4	Mitochondria	Tom20	mEGFP	Available
5	Desmosomes	Desmoplakin	mEGFP	Available

>1600 vials distributed to 25 countries
established at major universities for basic and applied research



49	Sarcomeric thick filaments	MLC-2v (late)	mEGFP	Available
50	Sarcomeric z-discs	Alpha-actinin-2	mEGFP	Available
51	Costameres	Dystrophin	mEGFP	Available

~58 FP-tagged isogenic iPSC lines (WTC-11)

- major structures, nuclear, cardio-specific, multi-edits
- 44 cellular structures
- mostly mono-allelic and single edits
- extensive QC

sharing cell lines, plasmids, and methods

- cell lines – Coriell Institute
- plasmids – Addgene
- distribution to stem cell cores at major institutes
- methods and tutorials - MBoC, Stem Cell Reports, JoVE

cell catalogs Coriell Institute

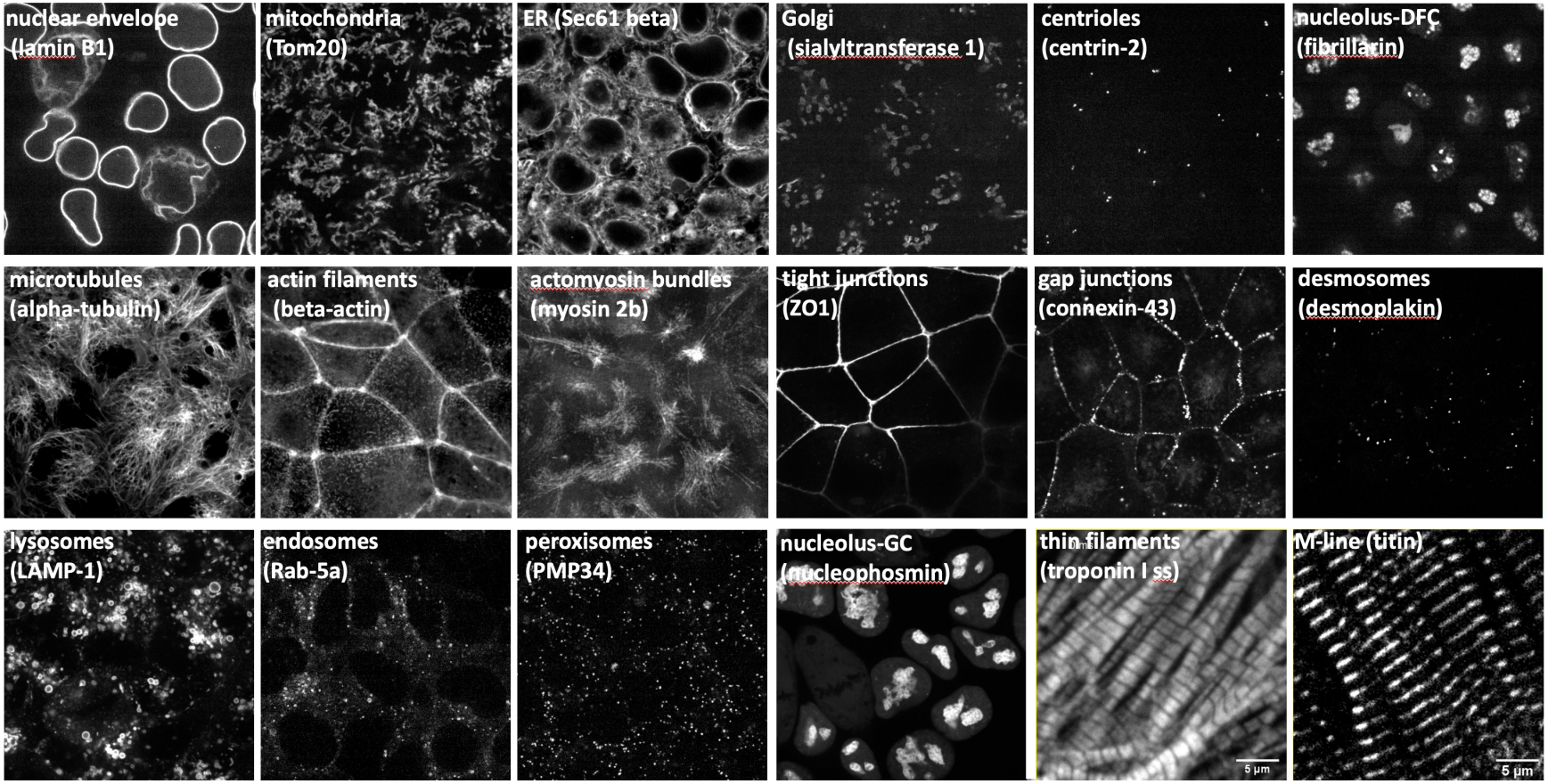
lab plasmids addgene

Allen cell collection

ALLEN INSTITUTE for CELL SCIENCE

(current as of 12/2022)

Example cell lines from the Allen Cell Collection



Talk outline

1: Bringing together spatiotemporal and molecular representations of cell state in the current era of big cell image data and AI

2: Where we have been: 10 years of the Allen Institute for Cell Science

3: Where we are heading: our new initiative – CellScapes

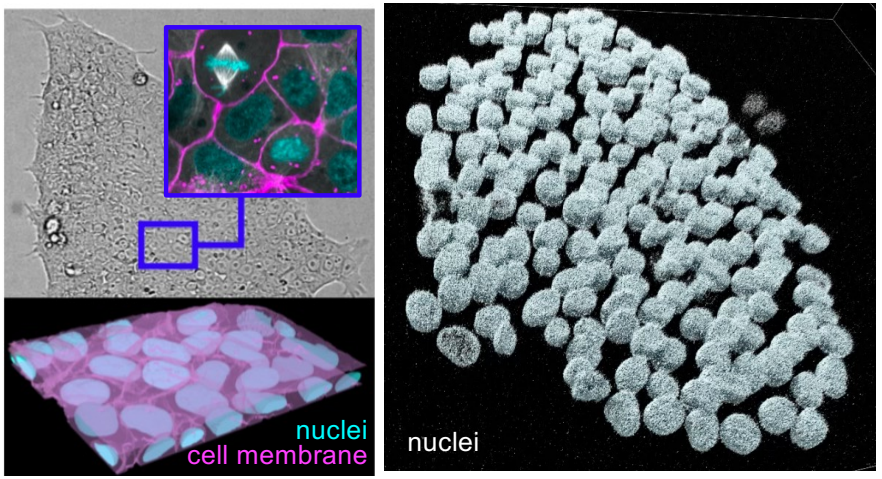
Using hiPSC models and AI toward predictive understanding of multiscale multicellular morphogenesis

Case study: *hiPSC-derived endothelial cells response to shear stress*

CellScapes: How do cells organize themselves across scales to form complex cell communities and tissues?

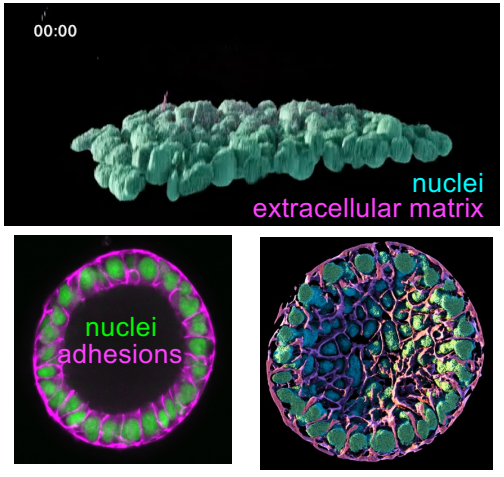
nuclei inside of human induced pluripotent stem (hiPS) cells

cells grown in 2D sheets

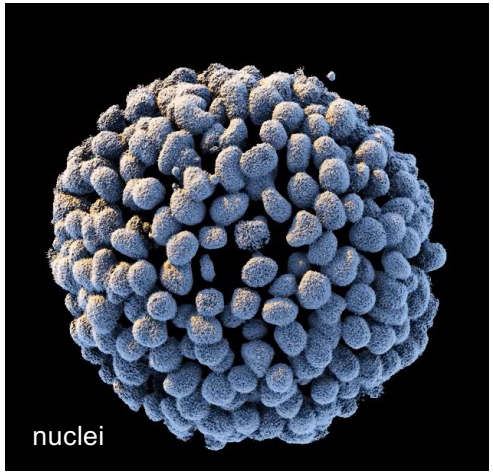


lumenoid formation

from 2D sheet to 3D lumenoid

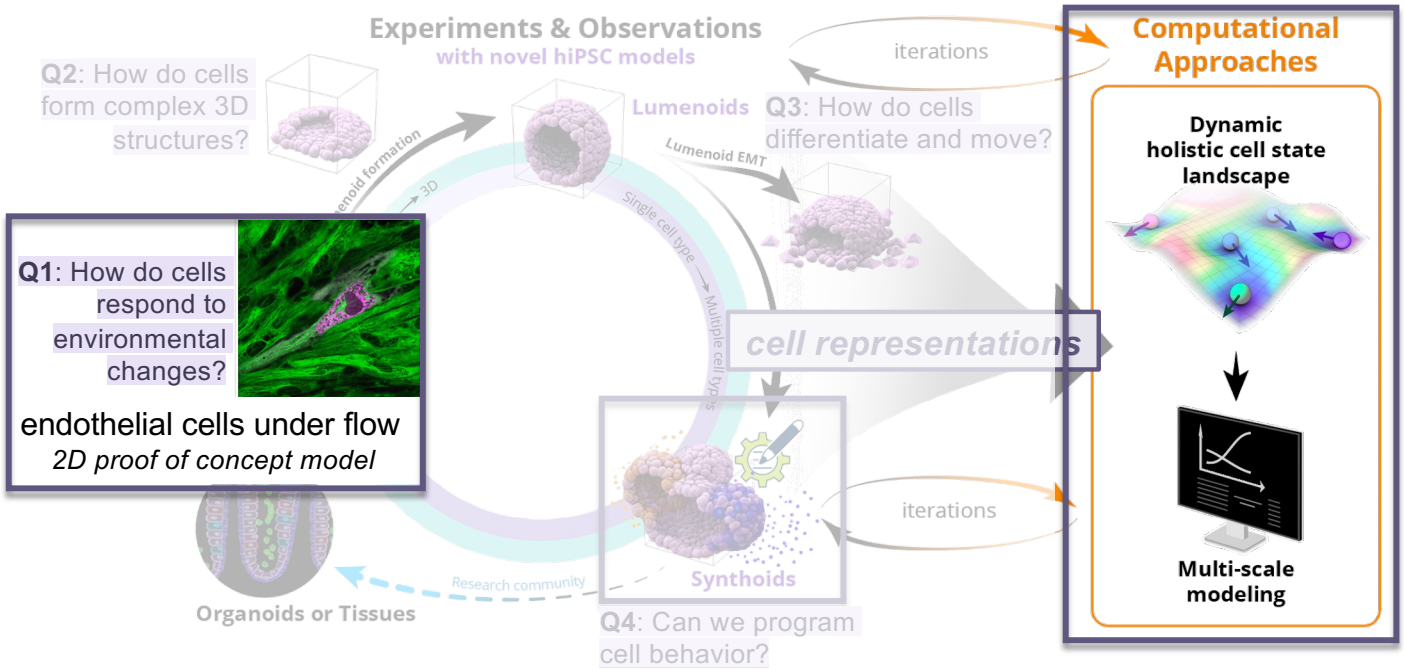


cells grown in 3D lumenoids



Goal: achieve a predictive understanding of cell state transitions during multicellular morphogenesis

Toward *predictive understanding* of multiscale multicellular morphogenesis



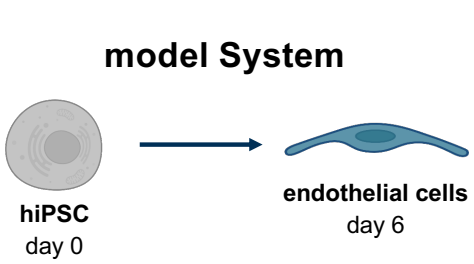
Objectives:

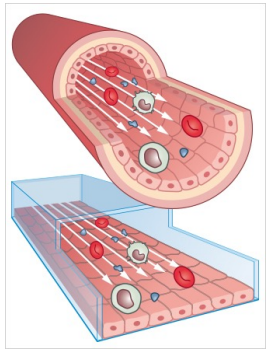
1. Quantify cell state transitions in stem cells
2. Build data-driven and dynamic cell models for discovery
3. Disseminate to amplify impact

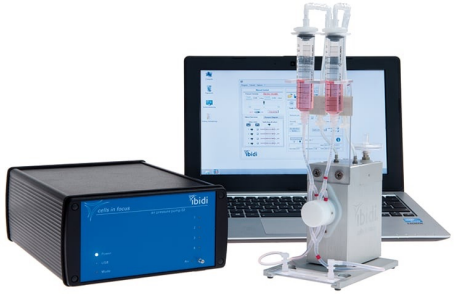
testing our "understanding":

- ***explanatory/mechanistic models & theory***
- ***"build to understand"*** **inspired by Richard Feynman's "What I cannot build. I do not understand"*

hiPSC-derived endothelial cells (hiPSC-ECs) under flow as a model system to study cell state changes

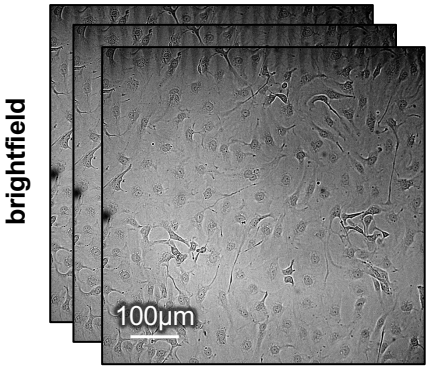


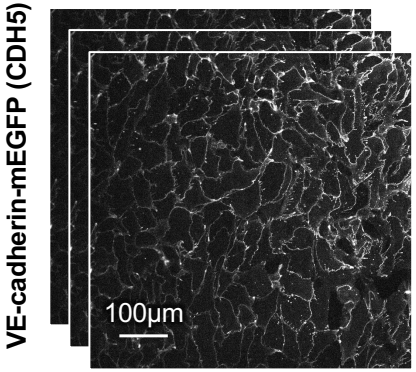


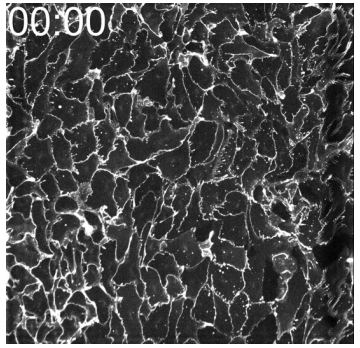


Ibidi.com

imaged in 3D at 20x every 5 minutes on 3i spinning disk confocal

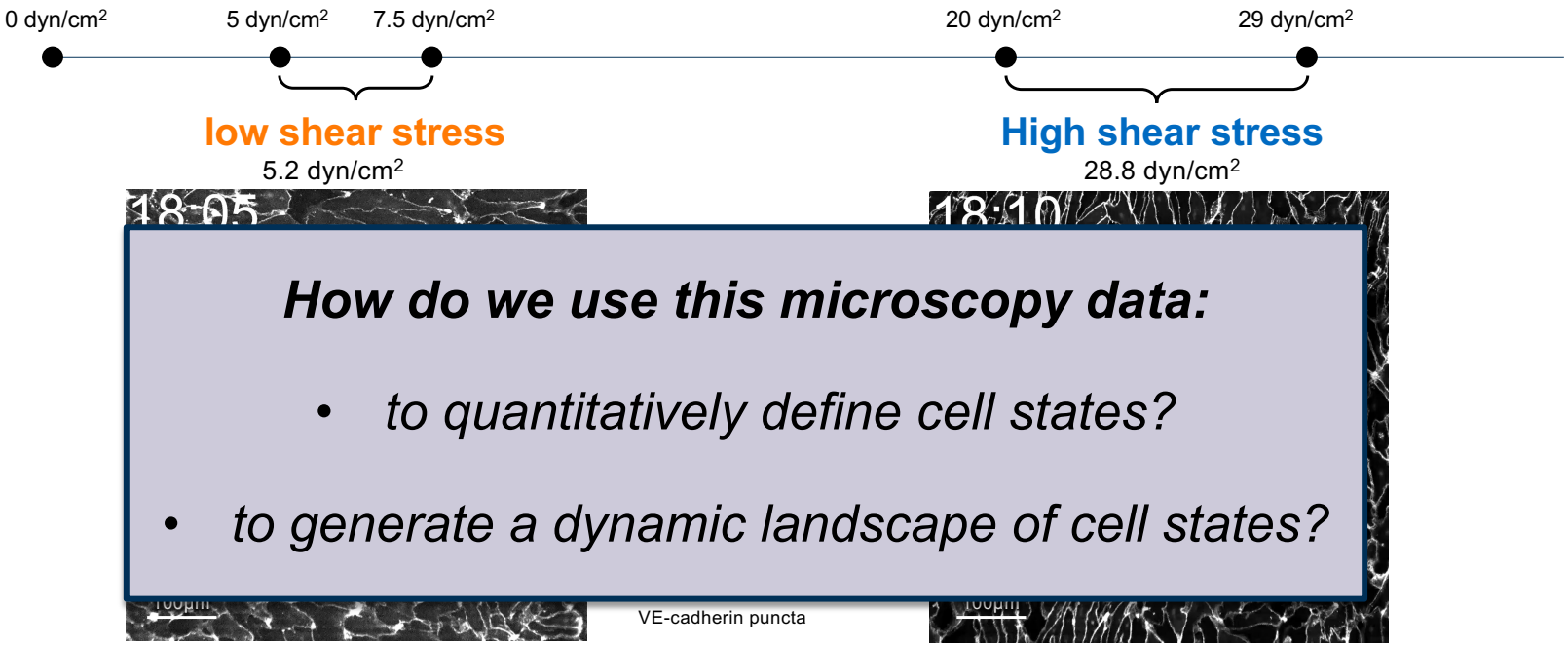






VE-cadherin is crucial for EC response to shear stress

Cells relax into distinct steady at different magnitudes of shear stress



How do we use this microscopy data:

- *to quantitatively define cell states?*
- *to generate a dynamic landscape of cell states?*

- **Morphology:** elongated & aligned parallel to direction of shear stress
- **Migration:** upstream relative to direction of shear stress
- **VE-cadherin:** puncta between adjacent cells persist

Flow →

- **Morphology:** elongated & aligned perpendicular to direction of shear stress
- **Migration:** many directions, many perpendicular to the direction of flow
- **VE-cadherin:** puncta rarely observed between adjacent cells

Toward a quantitative framework for studying cell state transitions

Where do we start? → **feature extraction**

Classic image processing workflow:
- single cell segmentation
- tracking + curation
- measurements (orientation, shape, etc.)

Can we "learn" the features that are important?

Challenges:
- naïve success rate of segmentation is low
- segmentation accuracy not consistent over time
- bias in choice of observables
  - What are the identifiable axes of variation?
- **Which features are necessary and sufficient for "useful and interpretable" characterization?**

Deep learning for unsupervised feature extraction

machine learning
deep learning / AI

*unsupervised**

*(*unsupervised ≠ naïve, uninformed)*

Can we "learn" the features that are important?

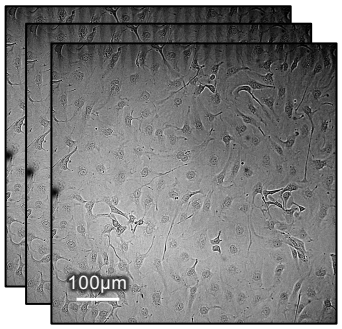
cell representations that are:
- *complete*
- *compact*
- *interpretable*

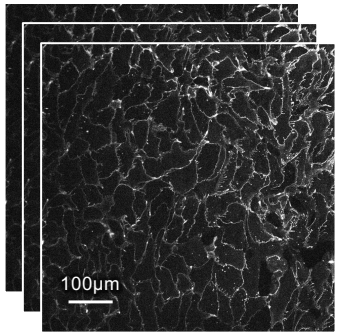
quantitative anchors for integrating data across categories of observables

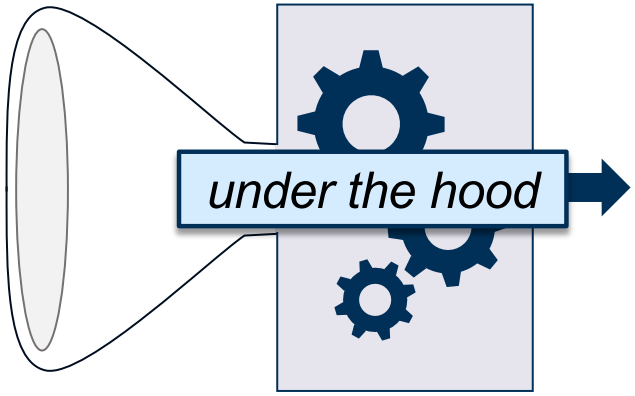
From steady state movies to a dynamic “landscape” at low sheer stress

unsupervised deep learning for image featurization

dimensionality reduction of an interpretable and generative latent space

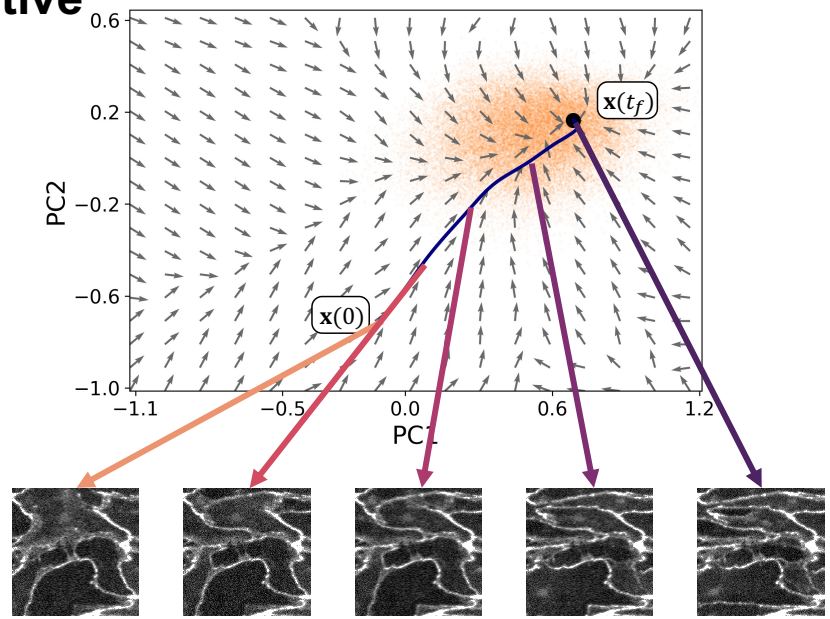






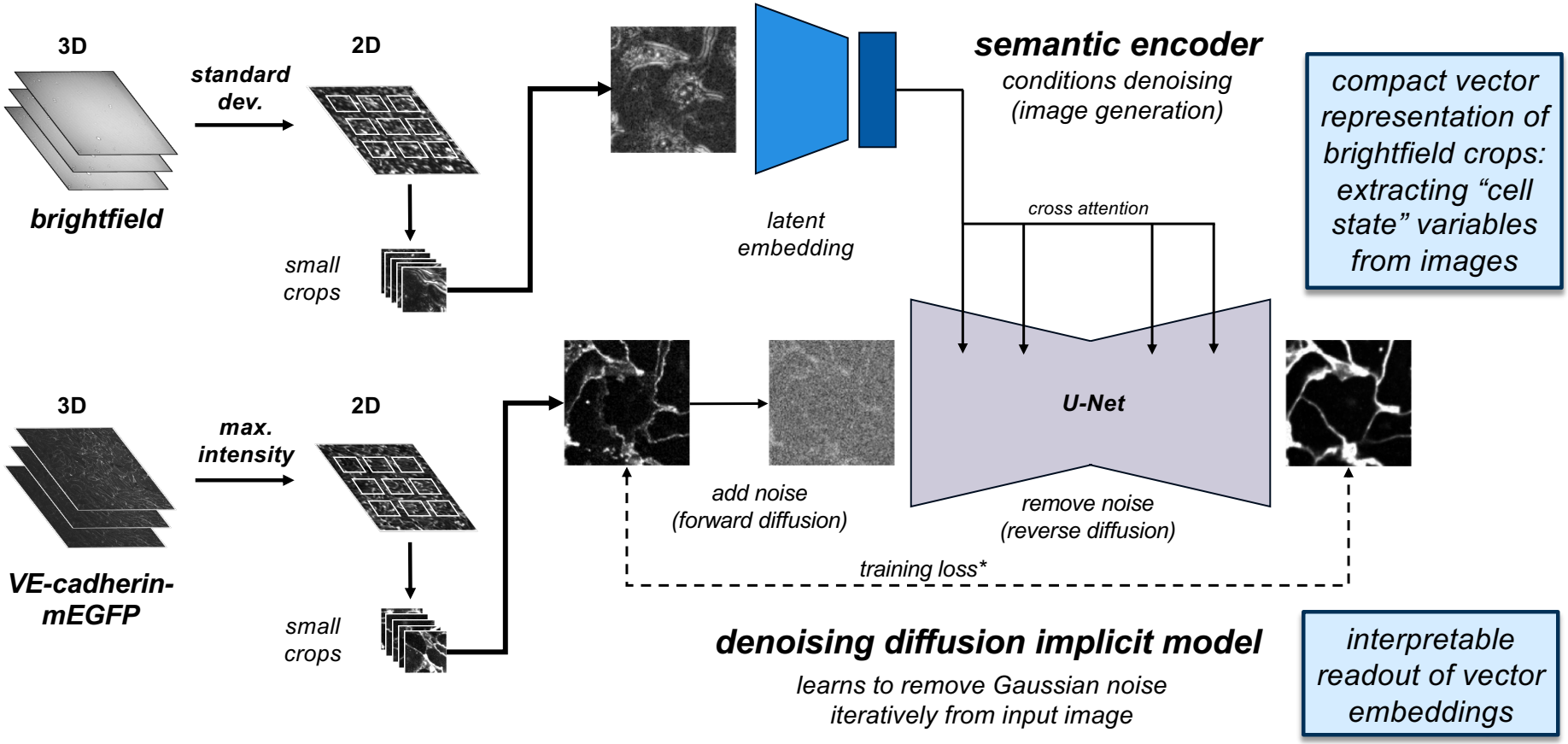
inferring single crop dynamics from feature displacements

dynamics for low shear stress:

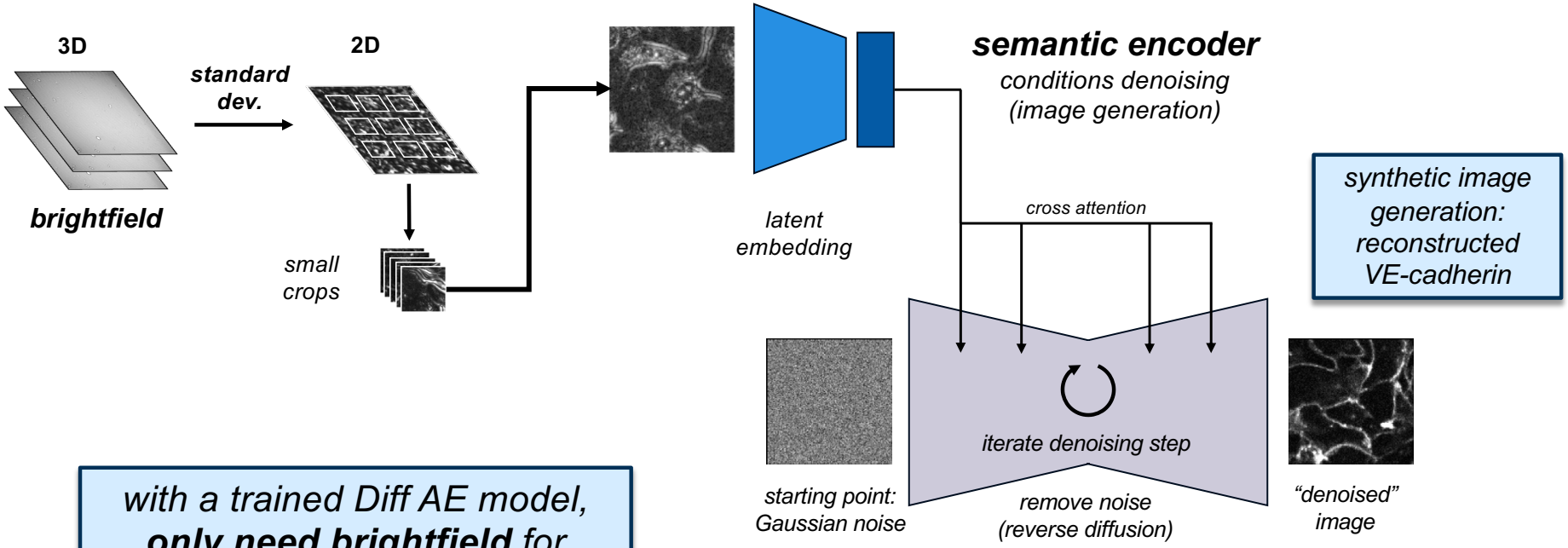


reconstructed crops generated by the model (not “real” images)

Diffusion autoencoder: learning high-level features from brightfield



Diffusion autoencoder: extraction and interpretation of latent features

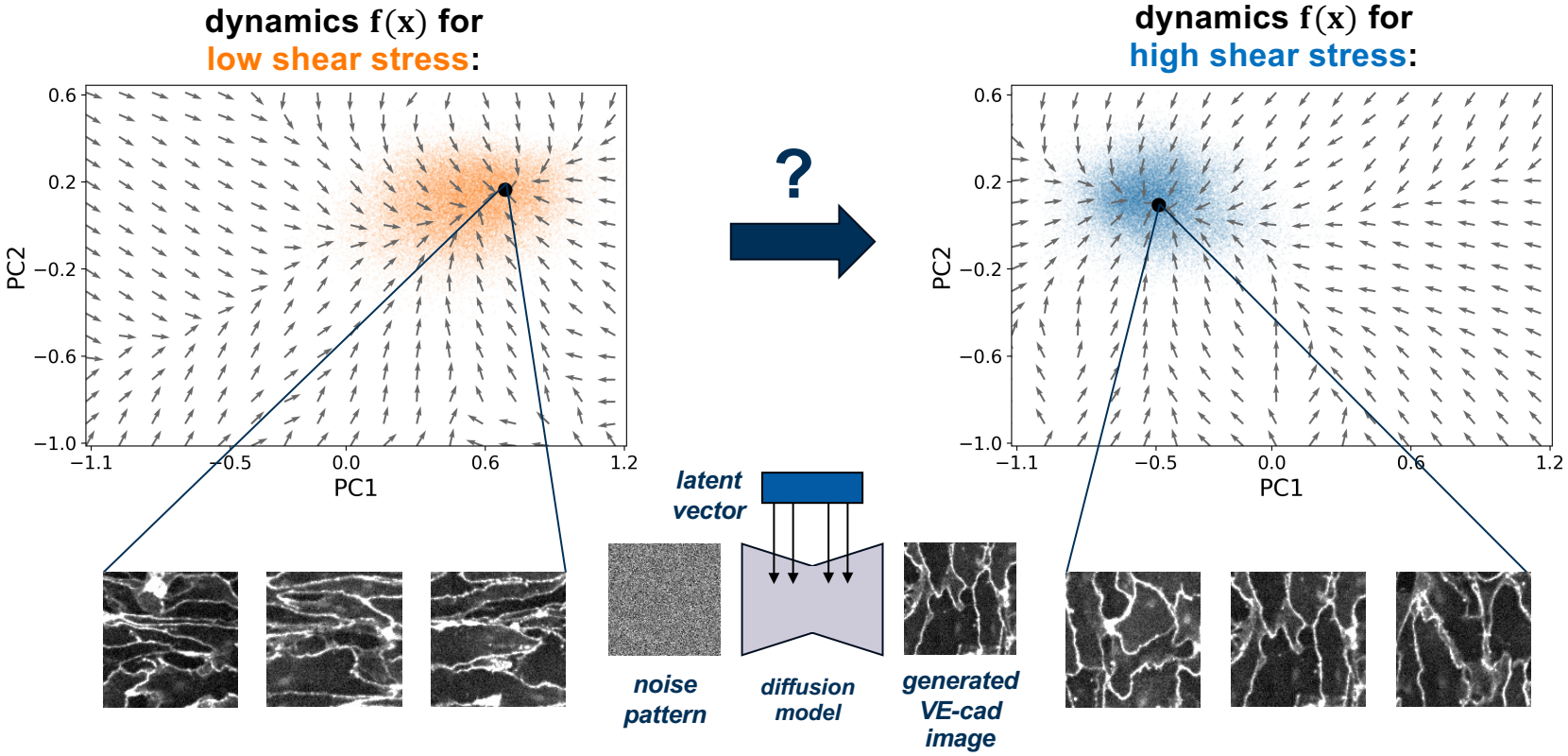


with a trained Diff AE model, ***only need brightfield*** *for interpretable cell representation*

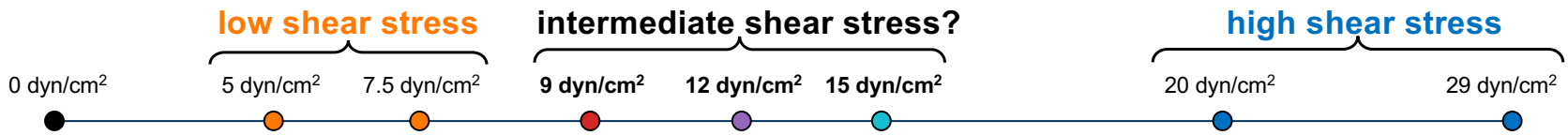
denoising diffusion implicit model

learns to remove Gaussian noise iteratively from input image

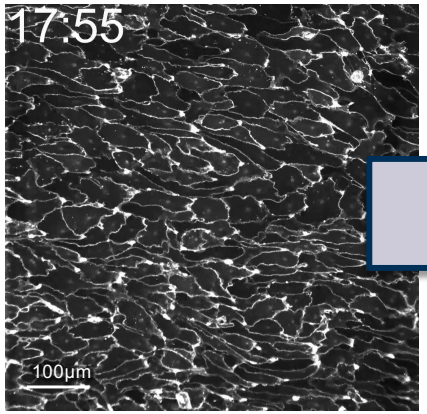
Stable fixed points at low and high shear stress represent distinct endothelial cell morphologies



Intermediate shear stress induces mixed cell alignment

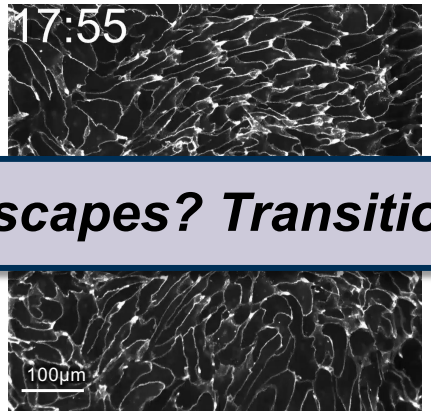


9 dyn/cm²



typical **low shear stress** response

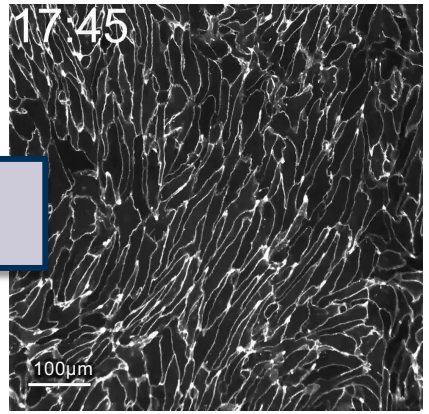
12 dyn/cm²



- **morphology remains elongated & locally aligned in different directions**
- cells exhibit local collective migration
- VE-cadherin puncta between some adjacent cells persist

Landscapes? Transitions?

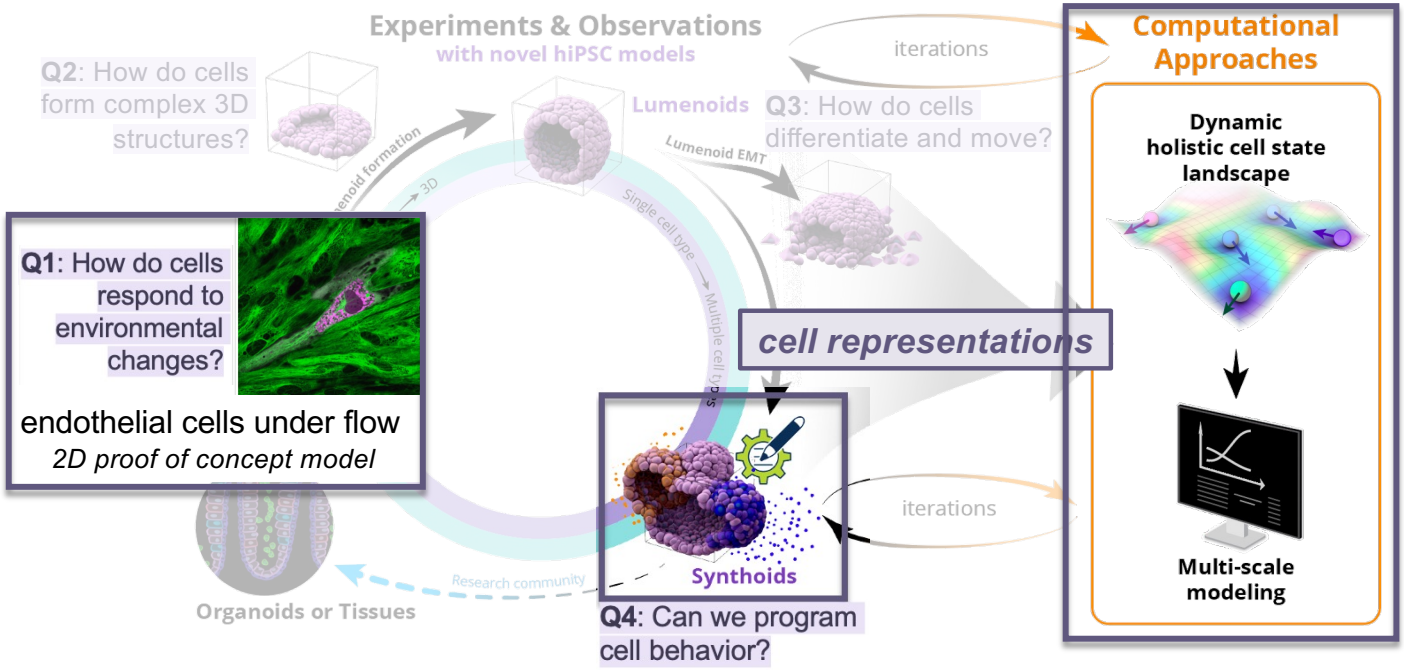
15 dyn/cm²



typical **high shear stress** response

CellScapes

Toward ***predictive understanding*** of multiscale multicellular morphogenesis



Pre-clinical research toward in vivo cell therapies needs a fundamental understanding of

what cells do

CAN

for engineering

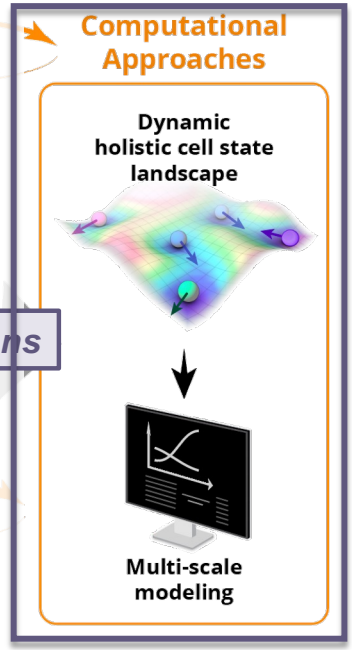
CellScapes

Toward ***predictive understanding*** of multiscale multicellular morphogenesis

- holistic cell representations *(across categories of observables)*
  &
- dynamic cell state landscapes

new types of data toward thinking differently about
- mechanisms of action
- treatment targets

cell representations



Pre-clinical research toward in vivo cell therapies needs a fundamental understanding of

what cells DO

CAN

for engineering

THANK YOU

The Allen Institute for Cell Science Team!

Julie Theriot

Department of Biology, University of Washington and Howard Hughes Medical Institute

Michael Elowitz

Department of Biology and Bioengineering, Caltech and Howard Hughes Medical Institute

Jin Wang

Department of Chemistry, Stony Brook University

ALLEN INSTITUTE | ALLEN INSTITUTE for CELL SCIENCE