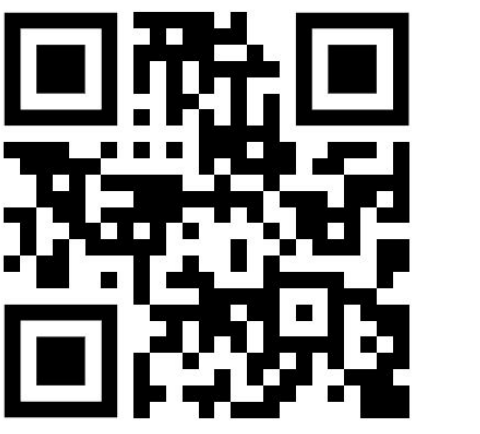




From Snapshots to Dynamics: Discovering Gene Regulatory Equations from Single-Cell Data

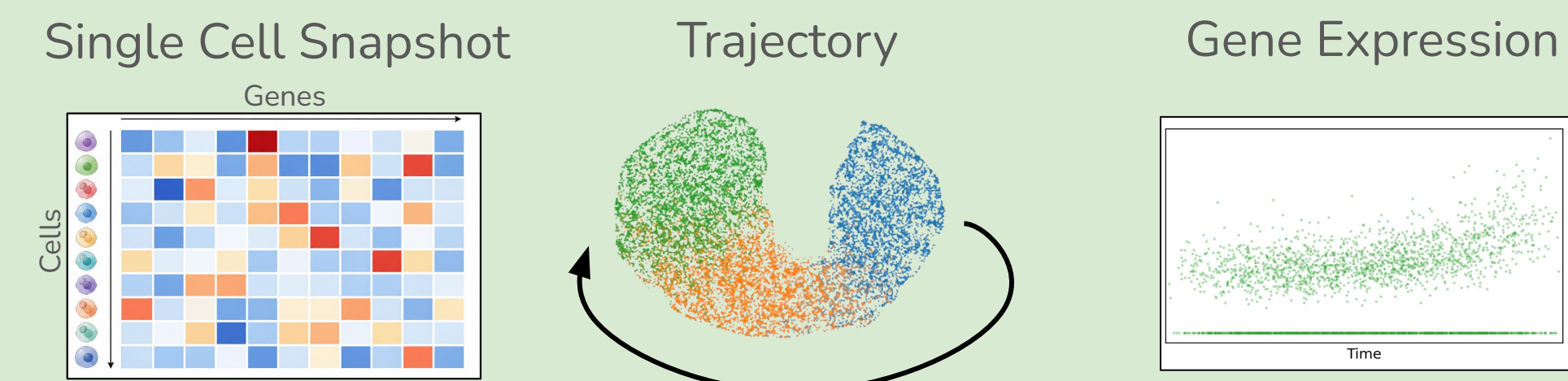
Samia Islam^{1,3}, Zijun Cui¹, Sudin Bhattacharya^{2,3}

¹Department of Computer Science and Engineering, ²Departments of Biomedical Engineering and Pharmacology & Toxicology, ³Institute for Quantitative Health Science and Engineering, Michigan State University



Bhattacharya Lab @ MSU

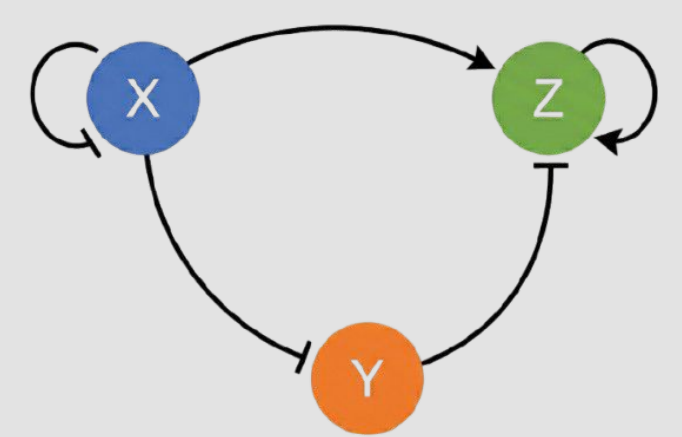
What scRNA-seq Data Provides



Trajectories describe progression, but not the governing rules.

Can sparse single-cell trajectories reveal the ODEs governing gene regulation?

A Toy System

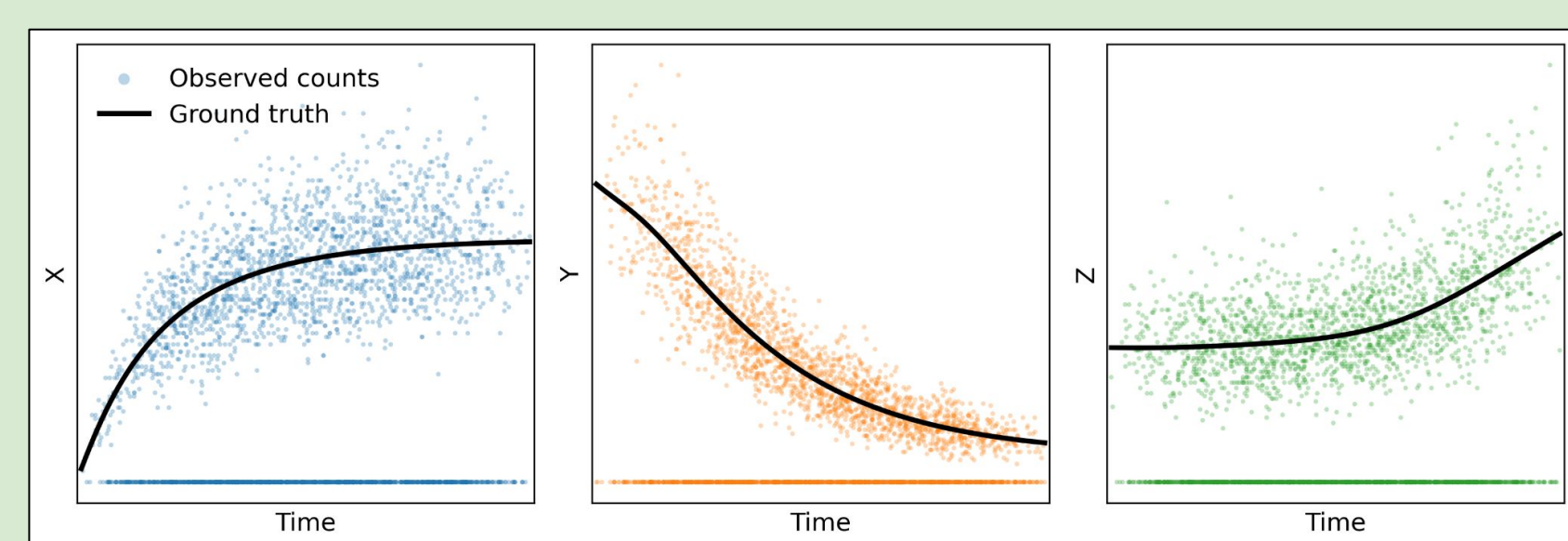


$$\begin{aligned}\dot{X} &= 0.3 + 8.0 \text{ Repression}(X; 50, 2) - 0.05X, \\ \dot{Y} &= 0.1 + 10.0 \text{ Repression}(X; 40, 4) - 0.10Y, \\ \dot{Z} &= 0.01 + 0.5 \text{ Activation}(X; 50, 3) + 5.0 \text{ Repression}(Y; 30, 4) \\ &\quad + 2.5 \text{ Activation}(Z; 20, 4) - 0.05Z.\end{aligned}$$

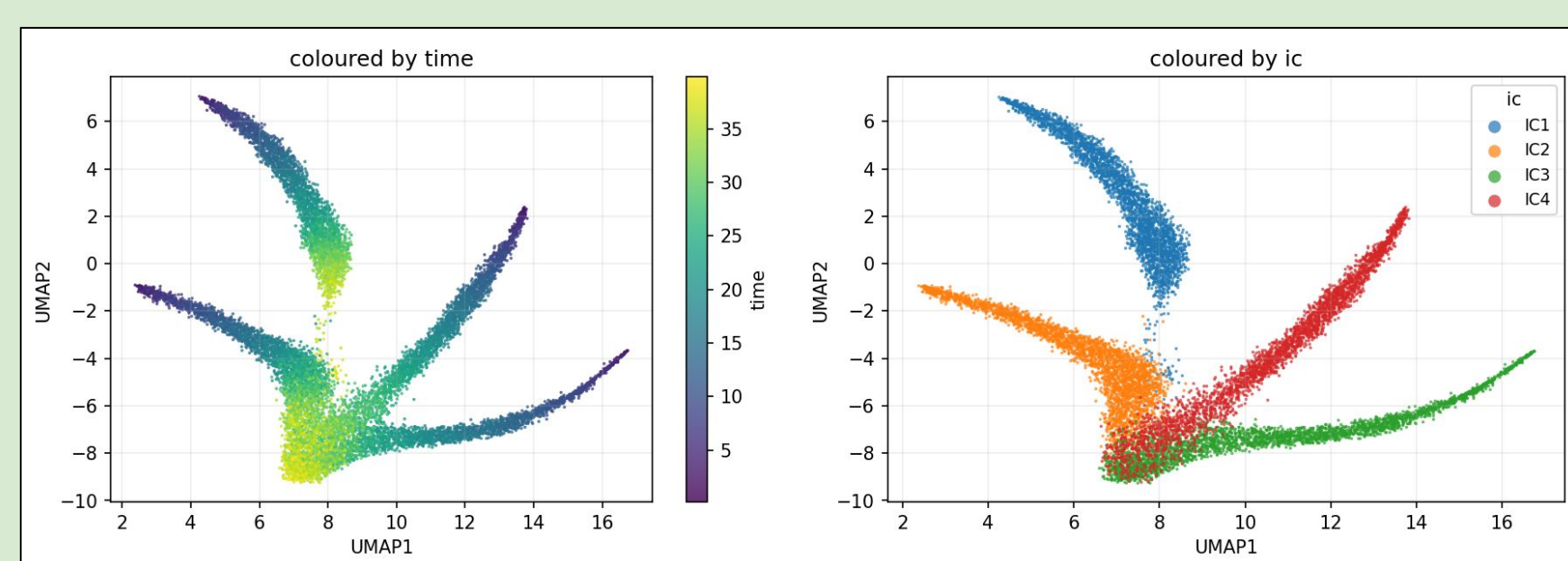
$$\text{Activation}(G; K, n) = \frac{G^n}{K^n + G^n}, \quad \text{Repression}(G; K, n) = \frac{K^n}{K^n + G^n}$$

Data Generation from the Toy System

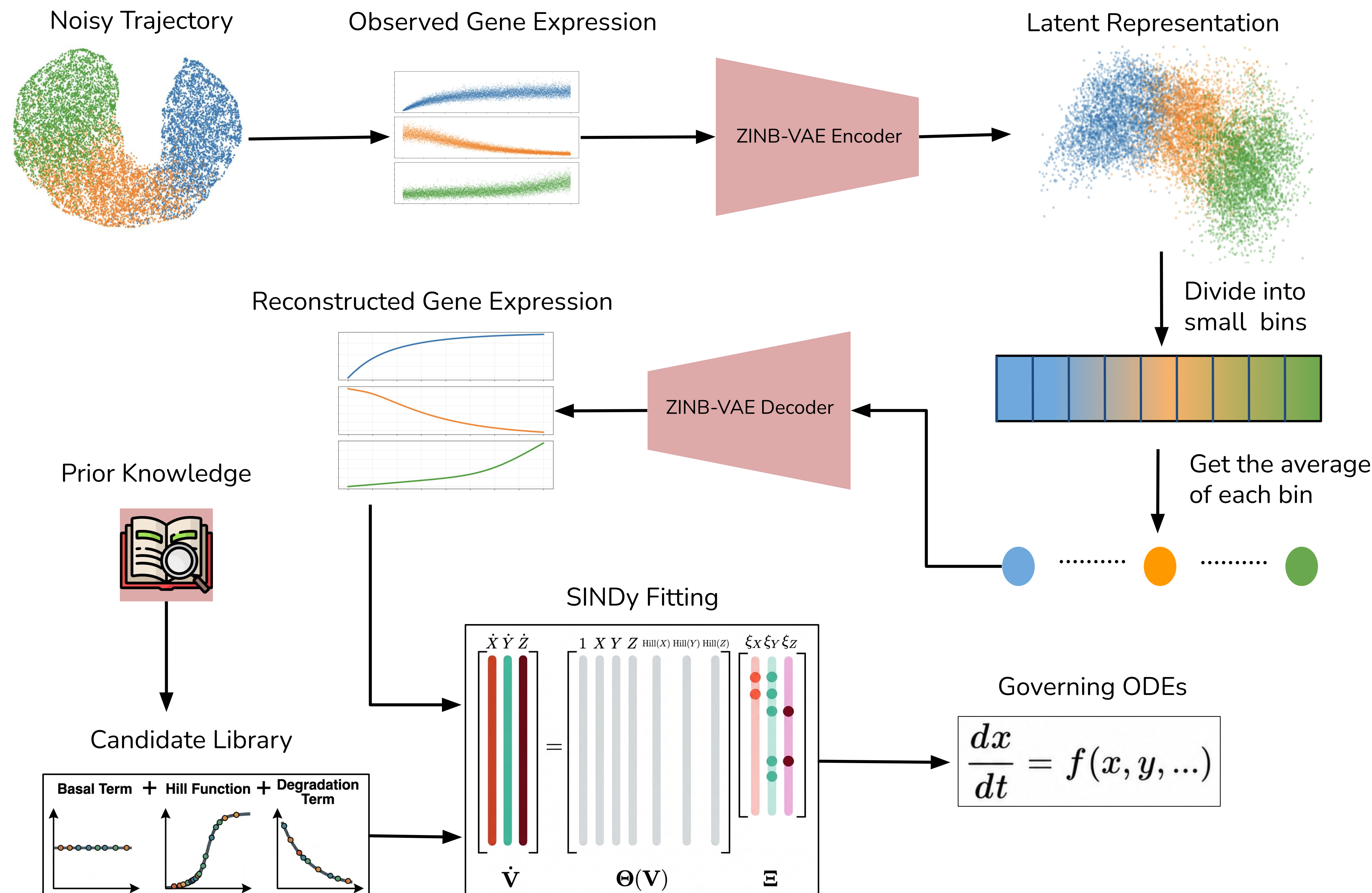
- Simulated four trajectories from different initial condition over a fixed time interval
- Sampled nonuniform time points from each trajectory
- Projected the three regulatory states into a 1,000-gene expression panel to represent a higher-dimensional transcriptomic measurement
- Converted the expected expression values into discrete counts using a zero-inflated negative-binomial (ZINB) observation model



Ground Truth vs Zero-Inflated Negative Binomial expressions



Four Generated Trajectories



ZINB-VAE

- A zero-inflated negative-binomial likelihood explicitly represents:
 - Negative-binomial sampling noise
 - Additional expression-dependent dropout
- The encoder maps high dimensional cell profile to a low-dimensional latent distribution
- The decoder produces estimated pre-dropout mean expression, count dispersion, and excess-zero probability for each cell
- The decoder's estimated mean is used for downstream trajectory reconstruction

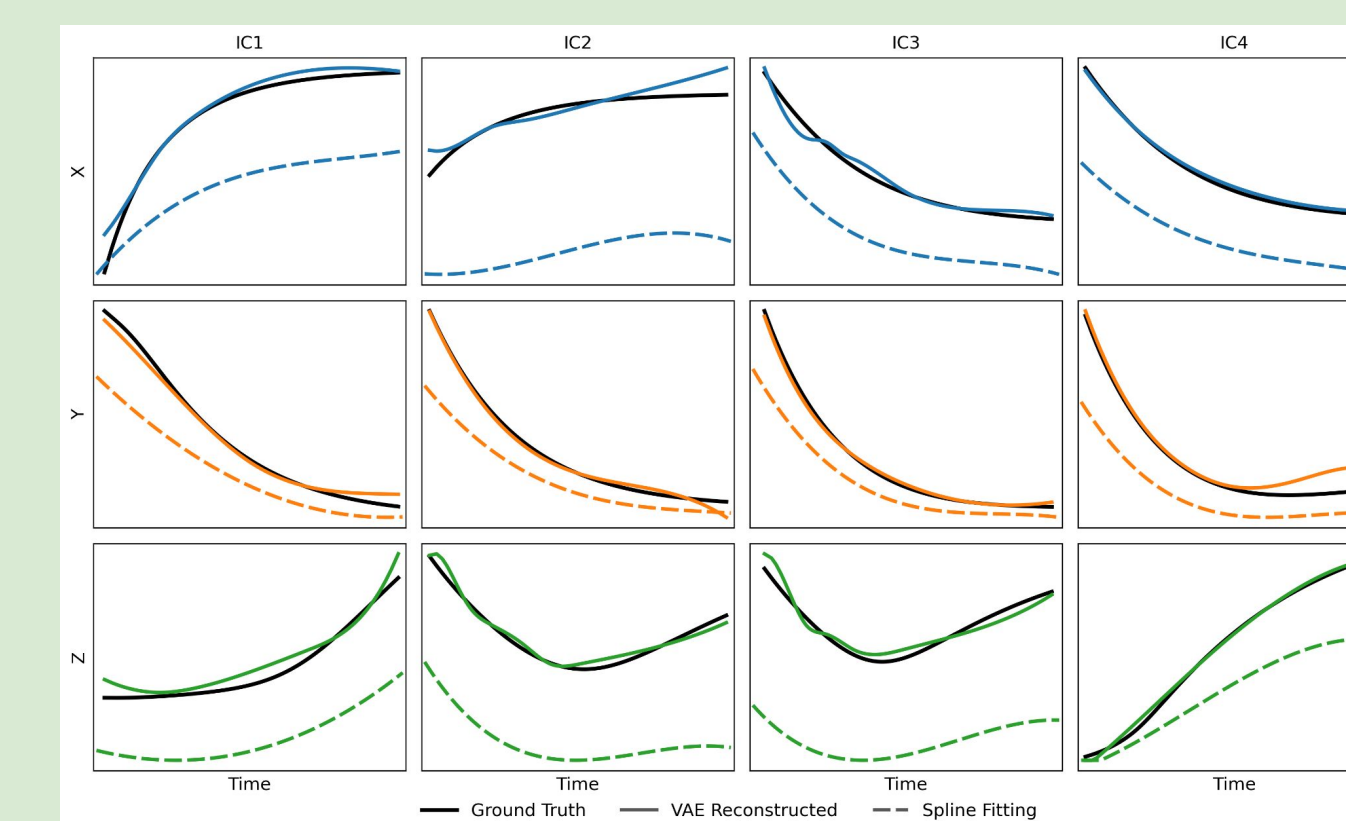
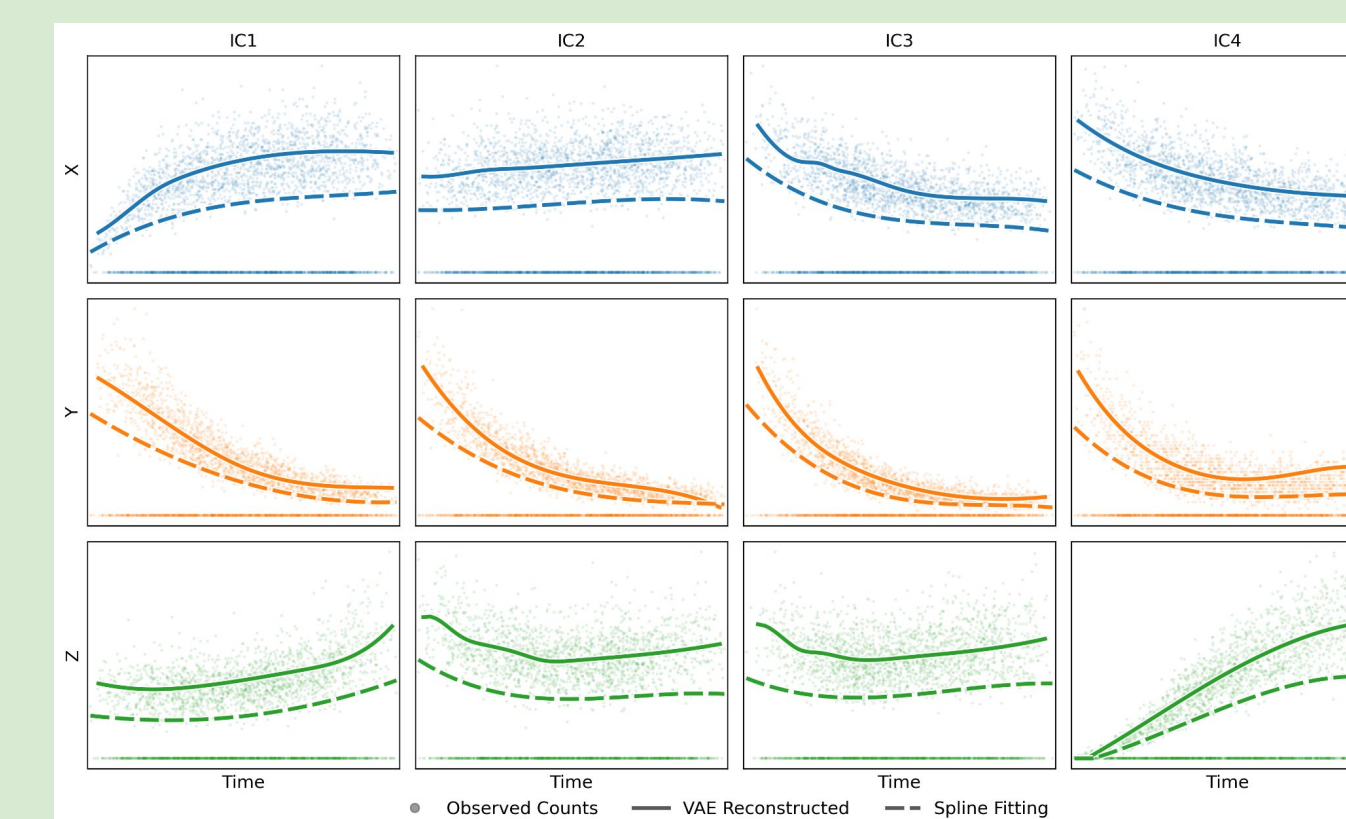
Prior-Informed SINDy

- Constructed a biologically motivated candidate feature library containing:
 - Basal production
 - Hill activation and repression
 - Linear degradation
- Supplied a regulatory prior specifying the candidate activation, or repression term

$$\frac{dG_i}{dt} = B_i + \sum_j \alpha_{ij} \mathcal{R}_{ij}(G_j; K_{ij}, n_{ij}) - \beta_i G_i$$

Results: Trajectory Reconstruction

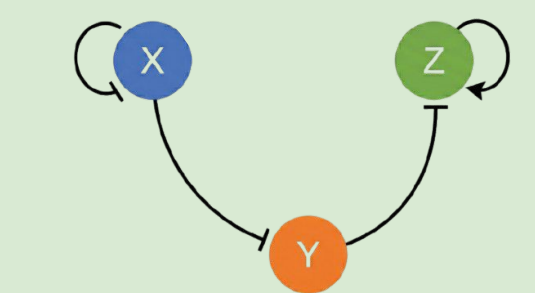
Extracting information across the 1,000-gene panel recovered smooth trajectories that closely followed the underlying three-gene dynamics.



Results: Equation Discovery

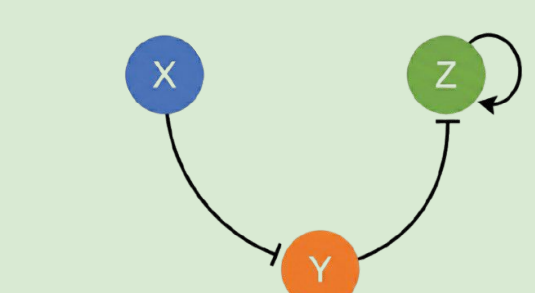
With Prior Gene Regulatory Information:

$$\begin{aligned}\dot{X} &= 3.9236 + 1.8915 \text{ Repression}(X; 60, 4) - 0.0697X, \\ \dot{Y} &= 0.3947 + 10.1314 \text{ Repression}(X; 40, 4) - 0.1028Y, \\ \dot{Z} &= 2.7805 + 3.9568 \text{ Activation}(Z; 60, 4) + 4.6371 \text{ Repression}(Y; 30, 4) - 0.0795Z.\end{aligned}$$

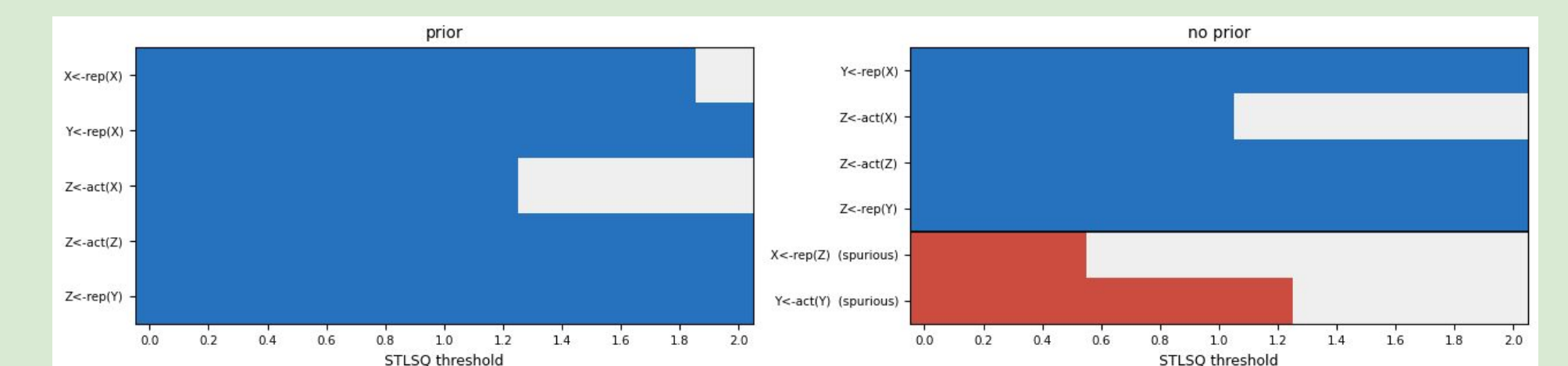


Without Prior Gene Regulatory Information:

$$\begin{aligned}\dot{X} &= 5.9249 - 0.0878X, \\ \dot{Y} &= 0.4992 + 27.4791 \text{ Repression}(X; 20, 3) - 0.0941Y, \\ \dot{Z} &= 2.7805 + 3.9568 \text{ Activation}(Z; 60, 4) + 4.6371 \text{ Repression}(Y; 30, 4) - 0.0795Z.\end{aligned}$$



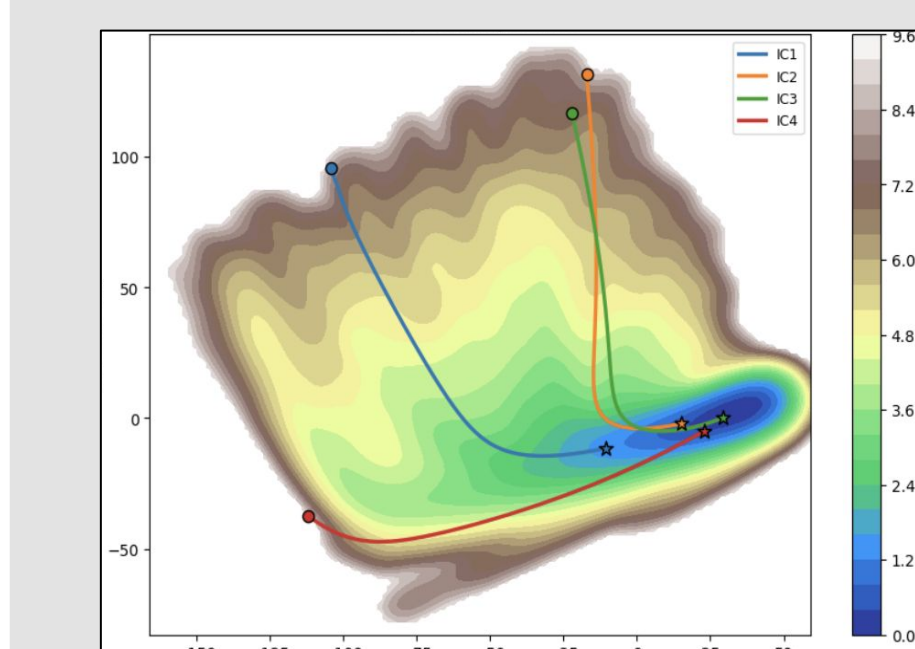
Edge persistence across sparsity thresholds:



The regulatory prior restricted the candidate space and prevented spurious interactions, while the no-prior model recovered most, but not all of the network.

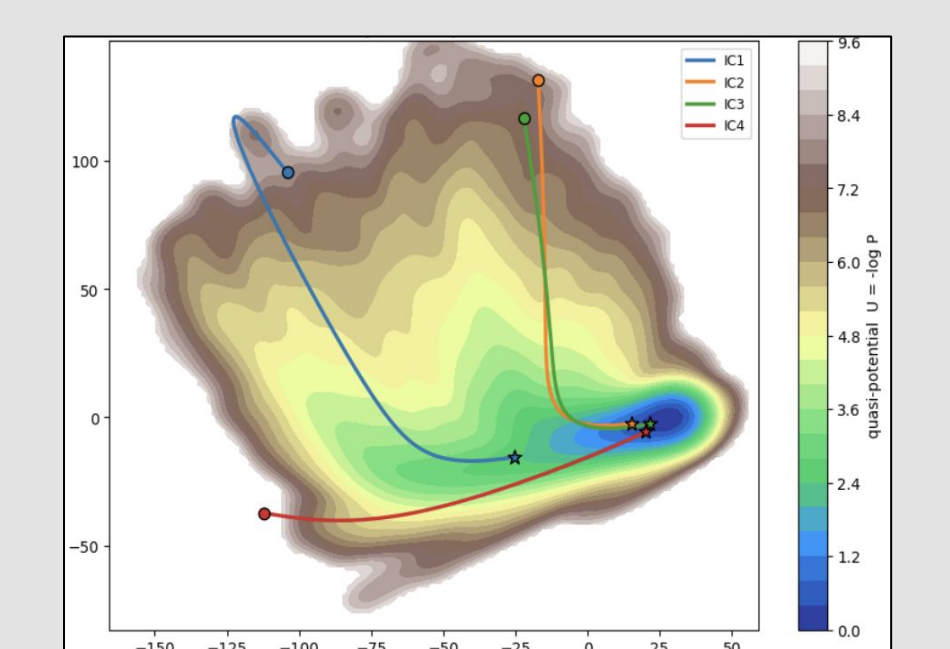
Waddington Landscape

Ground Truth:



Attractor state:
65.26, 13.39, 139.40

Predicted (without prior):



Attractor state:
67.24, 13.44, 132.33

The inferred ODEs reproduced the global organization of the dynamical landscape.

Even with 70.7% zeros, latent trajectory reconstruction enabled compact regulatory ODEs that reproduced and preserved the global behavior of the underlying gene-regulatory system.