

From Single-Cell Trajectories to Dynamic Gene Regulatory Models

Samia Islam

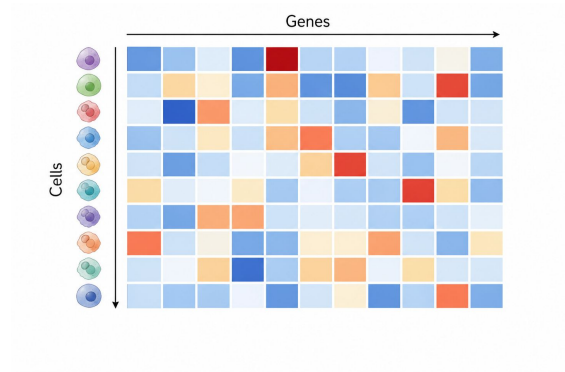
PhD Student

Bhattacharya Lab

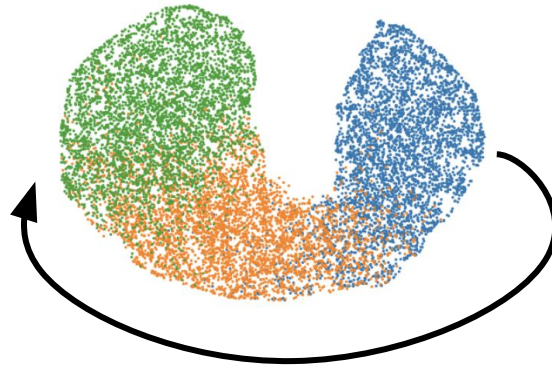
Department of Computer Science and Engineering

What single-cell data gives us

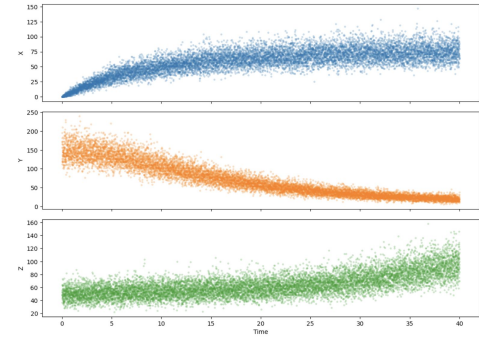
Single Cell Snapshot



Trajectory

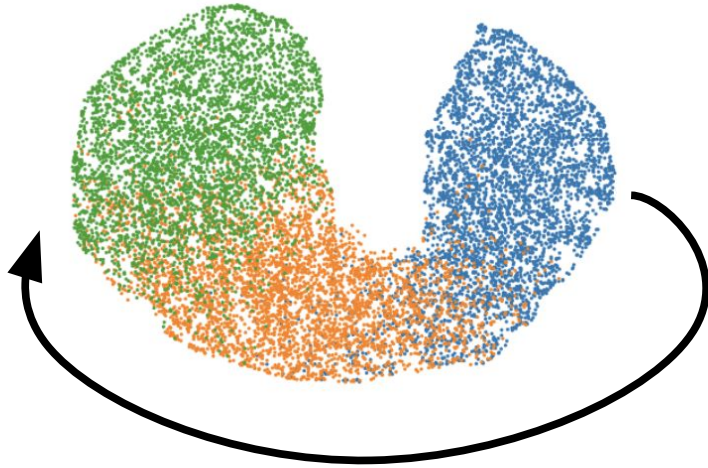


Gene Expression



Cell differentiates as the gene expression profile changes

What is missing?



What we observe:

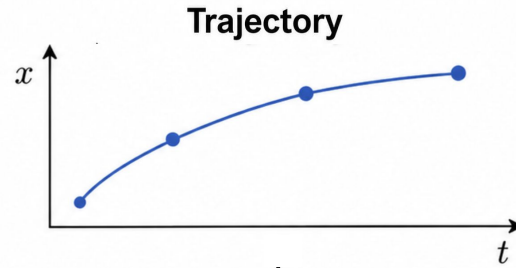
cells arranged along a trajectory

What we want:

rules that explain how states change

**Trajectories describe progression,
but not the governing rules**

Trajectory to Regulatory Dynamics

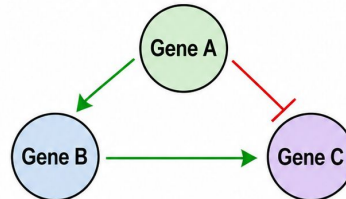


ODE model

$$\frac{dx}{dt} = f(x, y, \dots)$$



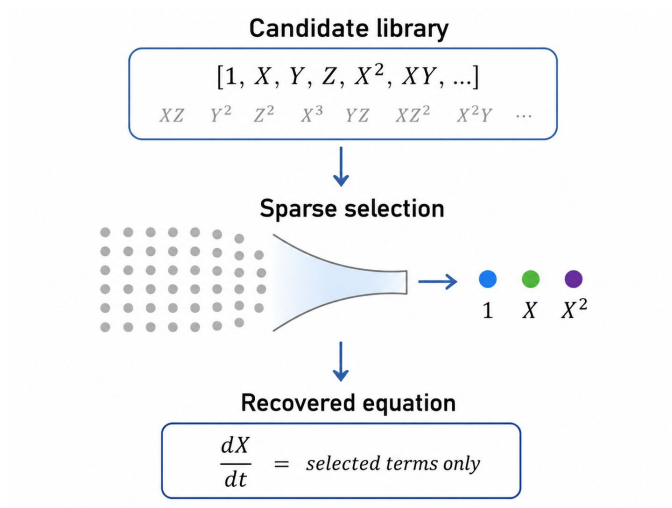
Regulatory Mechanism



How can we learn equations from observed trajectories?

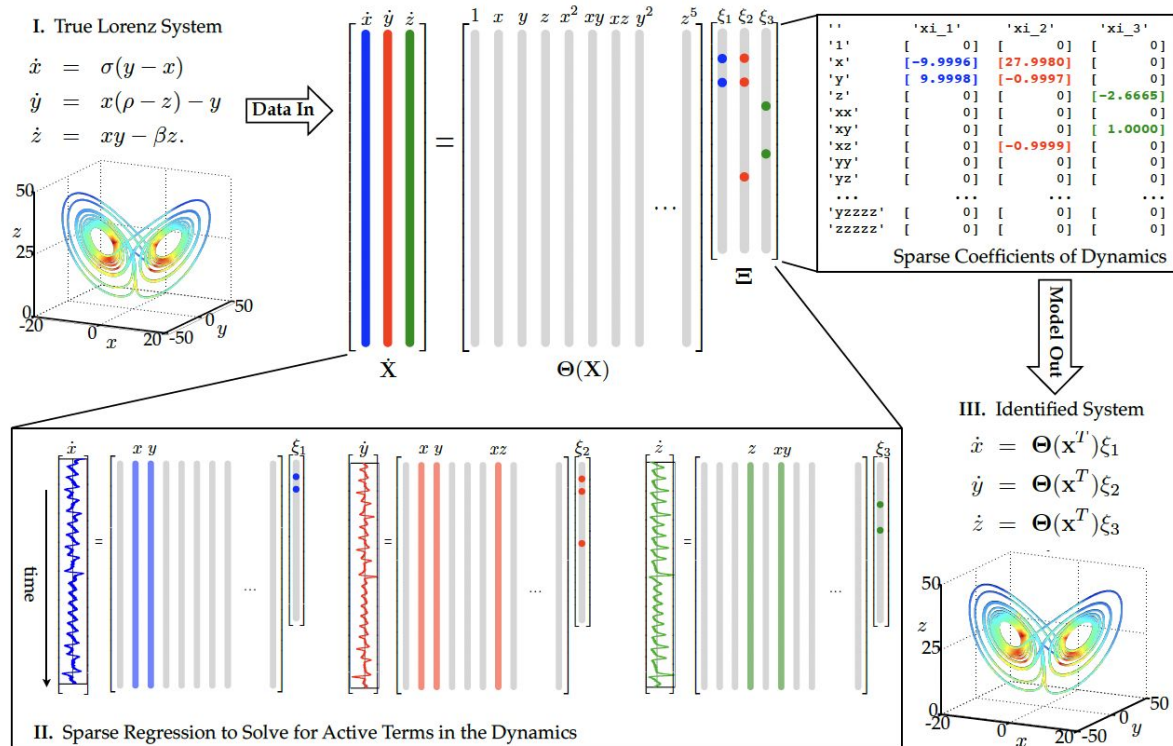
SINDy (Sparse Identification of Nonlinear Dynamics)

Core idea: choose a few terms from many candidates

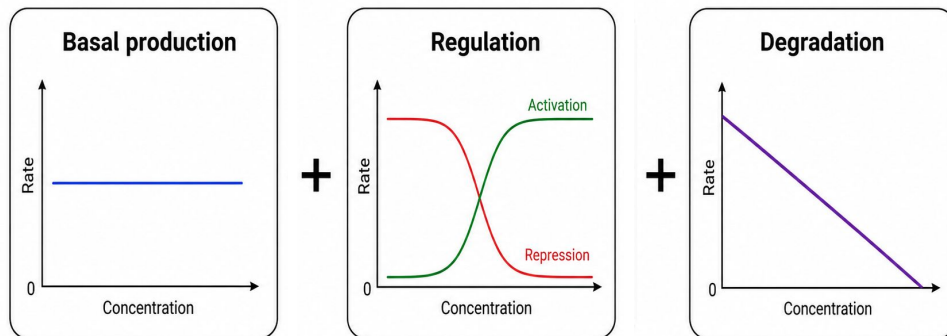


Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15), 3932-3937.

Overview of SINDy



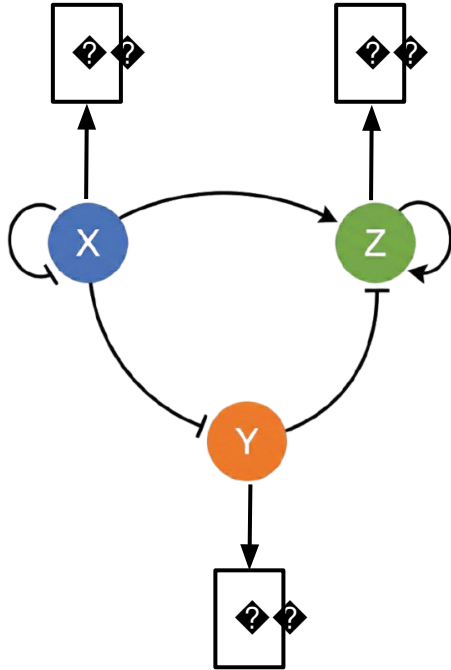
Gene Regulatory ODE



$$\frac{dX_i}{dt} = \underbrace{b_i}_{\text{basal}} + \underbrace{\sum_j \alpha_{ij} H_{ij}(X_j; K, n)}_{\text{activation/repression}} - \underbrace{\beta_i X_i}_{\text{degradation}}$$

$$H_{ij}(X_j; K, n) = \begin{cases} \frac{X_j^n}{K^n + X_j^n}, & \text{if } X_j \text{ activates } X_i \\ \frac{K^n}{K^n + X_j^n}, & \text{if } X_j \text{ represses } X_i \end{cases}$$

Gene Regulatory ODE: A Toy System

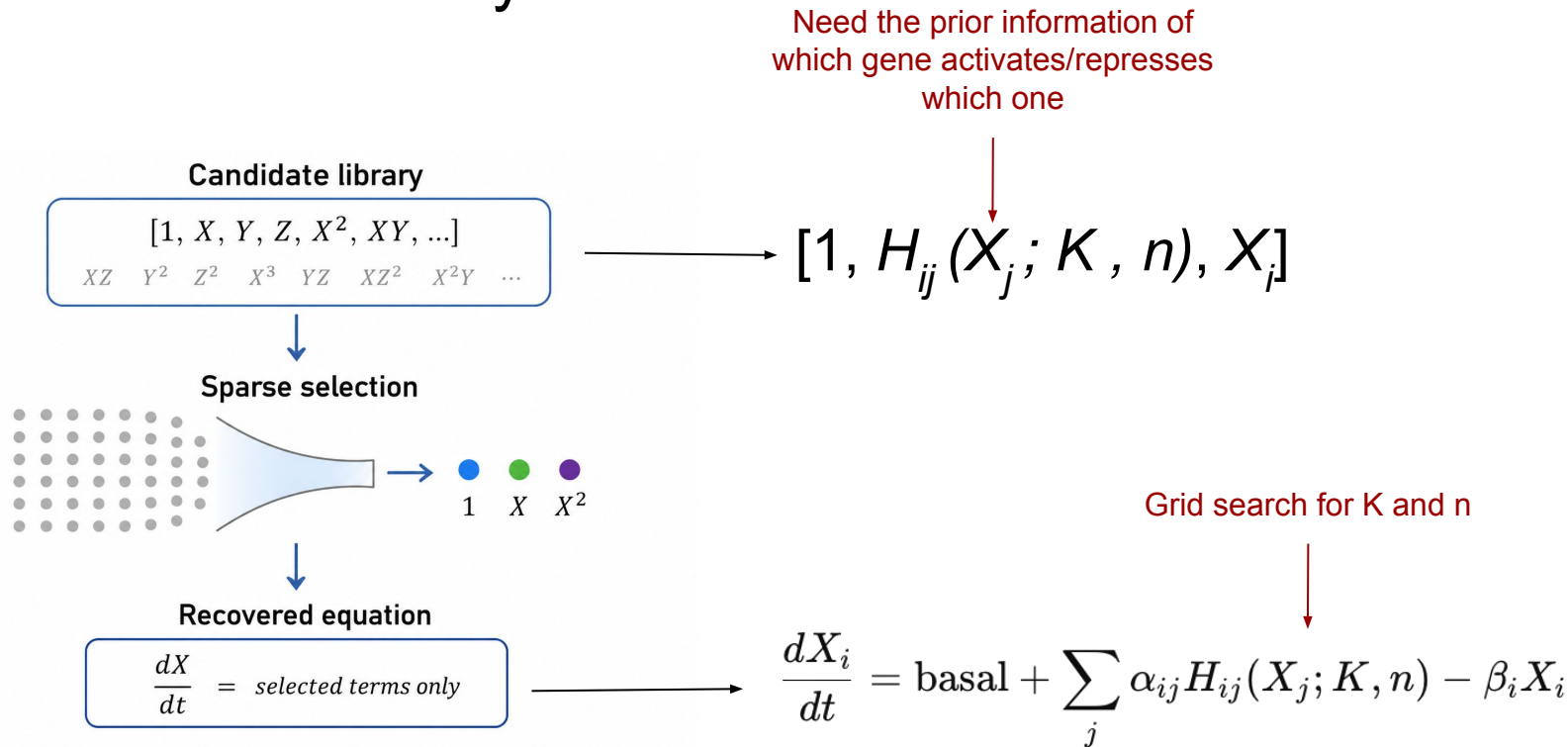


$$\frac{dX}{dt} = 0.3 + 8.14 \cdot \text{Rep}(X; 50, 2) - 0.035X$$

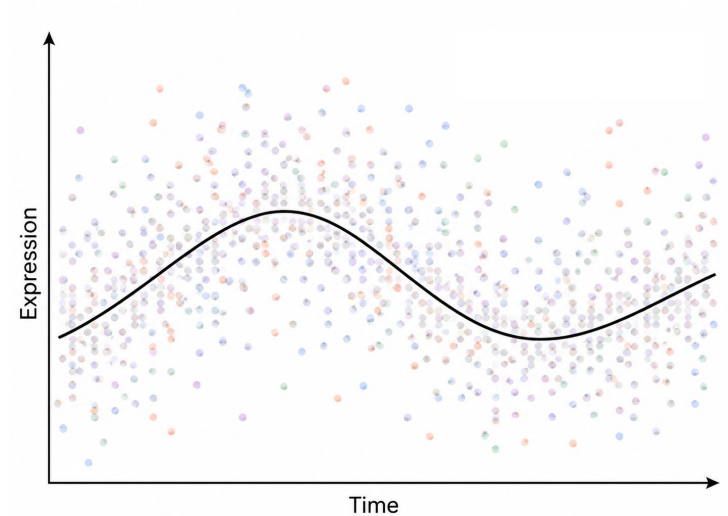
$$\frac{dY}{dt} = 0.02 + 9.26 \cdot \text{Rep}(X; 40, 4) - 0.08Y$$

$$\frac{dZ}{dt} = 0.0028 + 0.28 \cdot \text{Act}(X; 50, 3) + 5 \cdot \text{Rep}(Y; 30, 4) + 2.5 \cdot \text{Act}(Z; 20, 4) - 0.04Z$$

Prior-informed SINDy



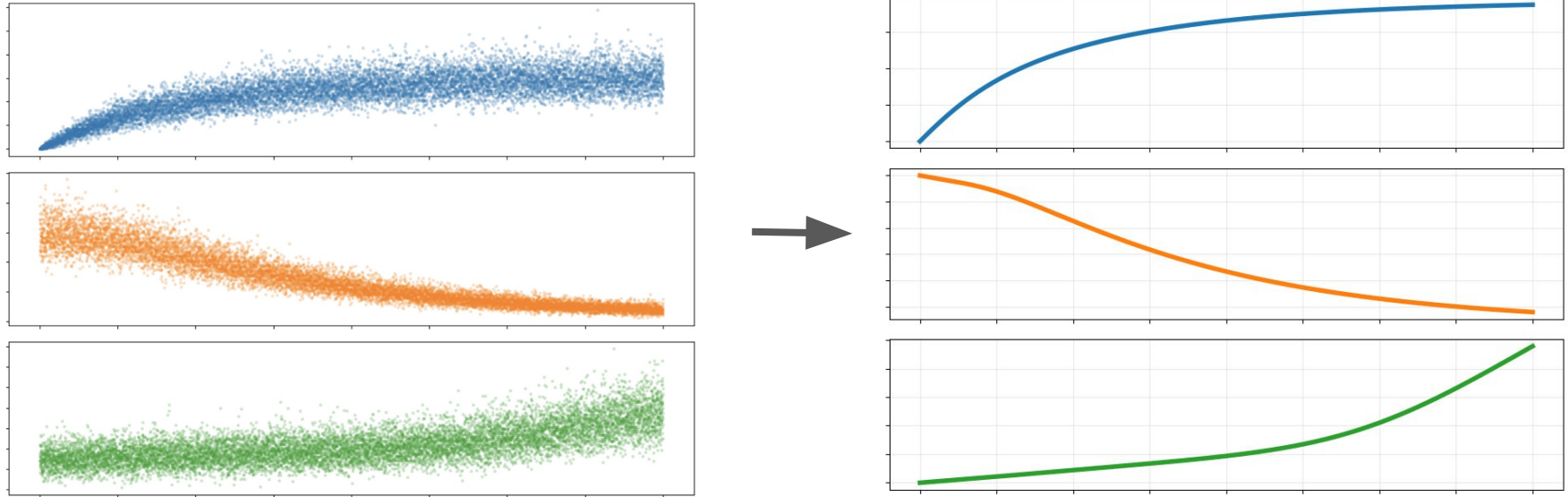
The problem: scRNA-seq is very noisy



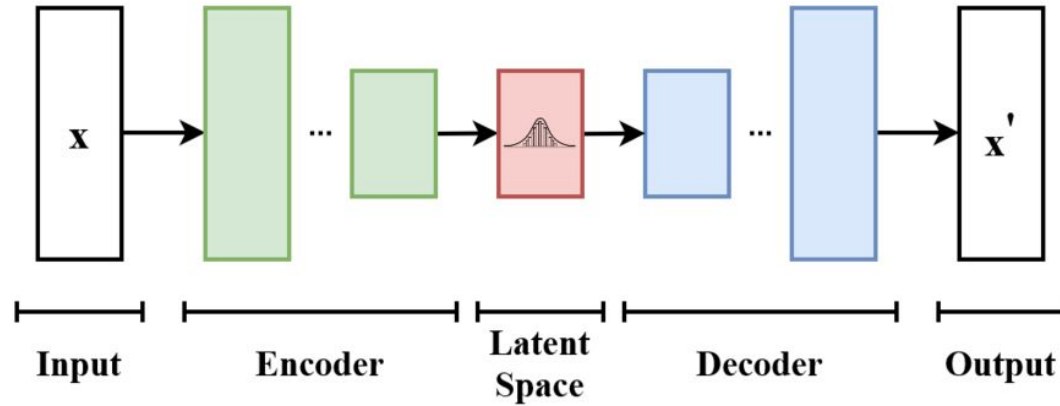
Noisy trajectory $\rightarrow$ Noisy derivatives $\rightarrow$ Wrong equations!

What if reconstruct the trajectory by generative modeling to reduce noise?

Main Goal of the Reconstruction



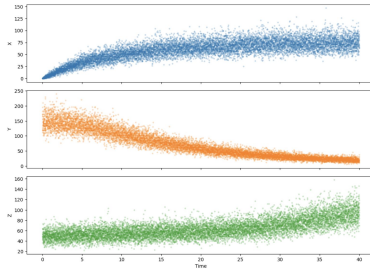
Variational Autoencoder



A VAE learns a low-dimensional representation of cell states

Step 1: Latent Space Trajectory

Noisy Gene Expression



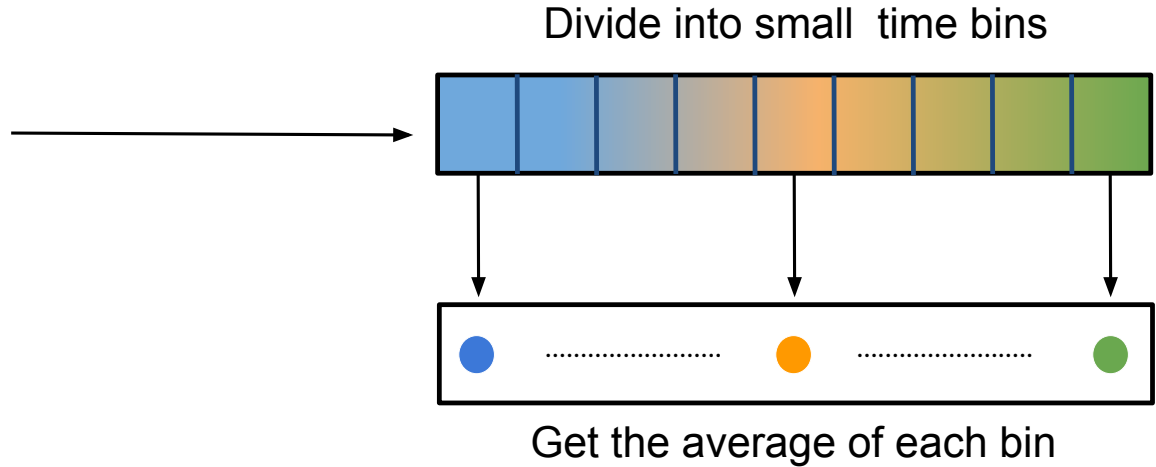
VAE Encoder

Latent Representation

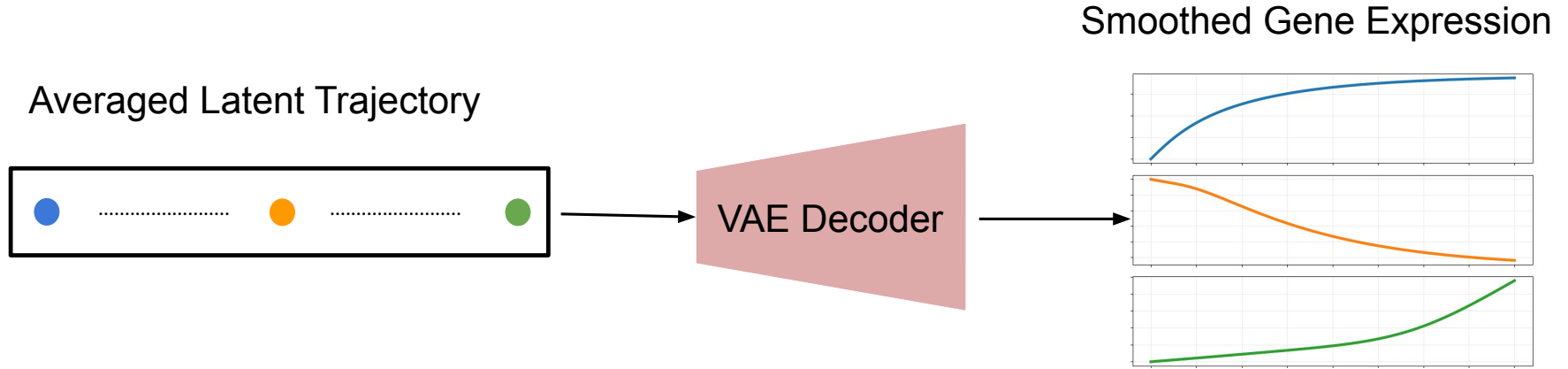


Step 2: Trajectory Reconstruction in Latent Space

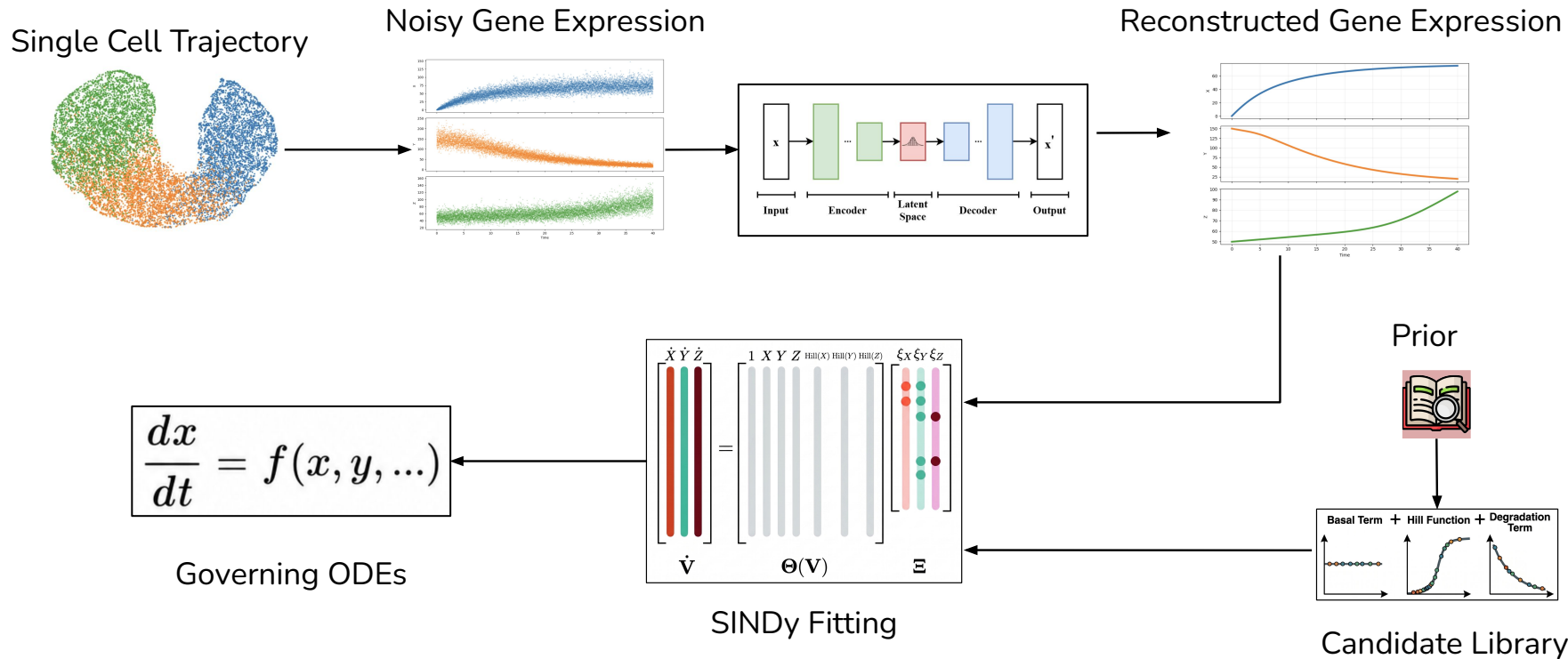
Latent Representation



Step 3: Decode Averaged Latent Points

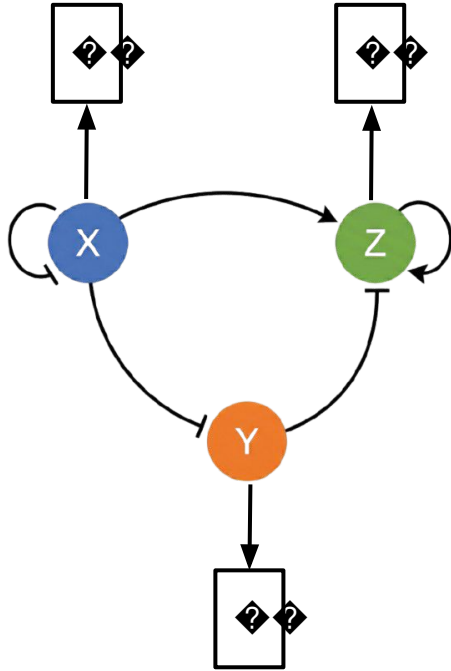


Overall Framework



Let's Verify the Framework

A Toy System: Simple Three Gene Network



$$\frac{dX}{dt} = 0.3 + 8.14 \cdot \text{Rep}(X; 50, 2) - 0.035X$$

$$\frac{dY}{dt} = 0.02 + 9.26 \cdot \text{Rep}(X; 40, 4) - 0.08Y$$

$$\frac{dZ}{dt} = 0.0028 + 0.28 \cdot \text{Act}(X; 50, 3) + 5 \cdot \text{Rep}(Y; 30, 4) + 2.5 \cdot \text{Act}(Z; 20, 4) - 0.04Z$$

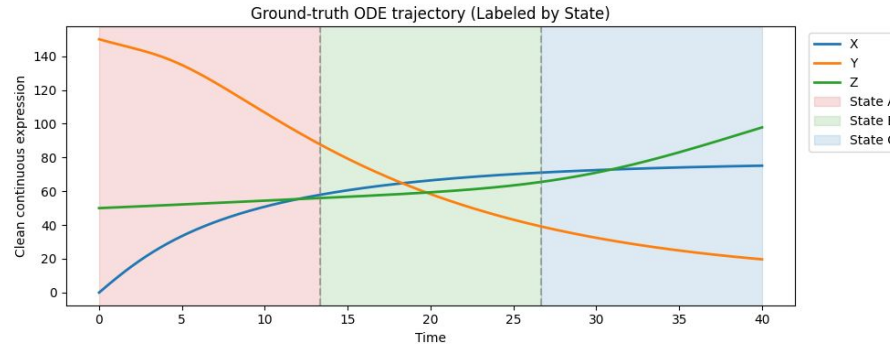
Simulate Trajectories from the ODE

$$\frac{dX}{dt} = 0.3 + 8.14 \cdot \text{Rep}(X; 50, 2) - 0.035X$$

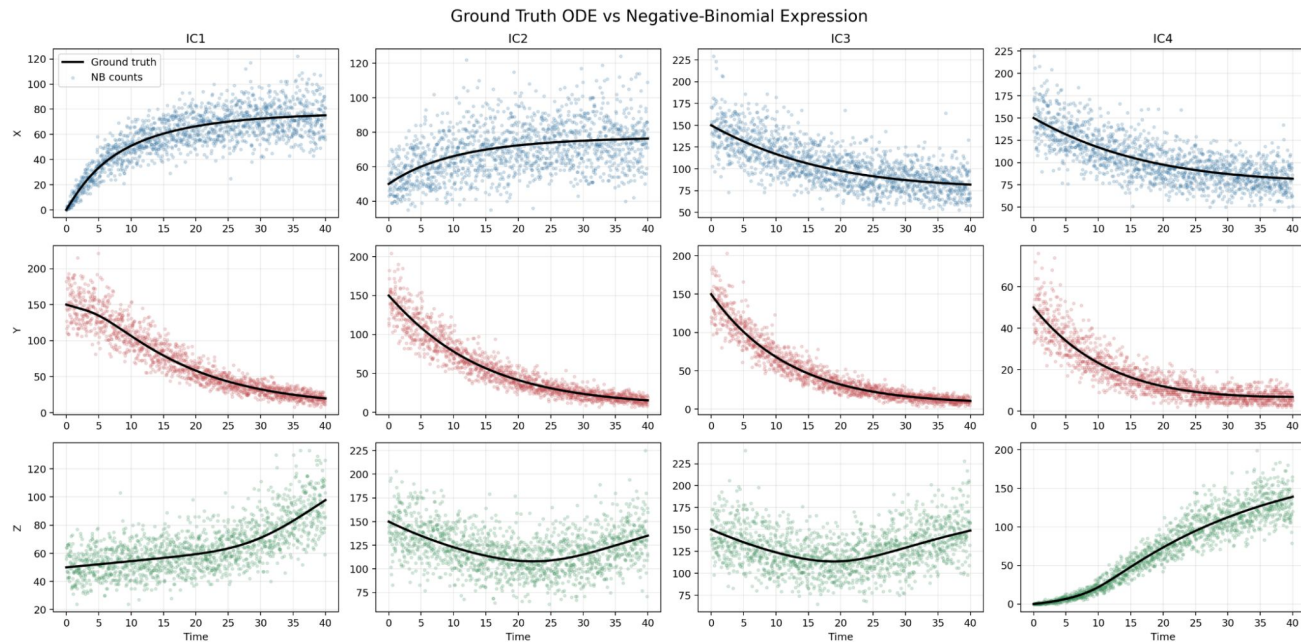
$$\frac{dY}{dt} = 0.02 + 9.26 \cdot \text{Rep}(X; 40, 4) - 0.08Y$$

$$\frac{dZ}{dt} = 0.0028 + 0.28 \cdot \text{Act}(X; 50, 3) + 5 \cdot \text{Rep}(Y; 30, 4) + 2.5 \cdot \text{Act}(Z; 20, 4) - 0.04Z$$

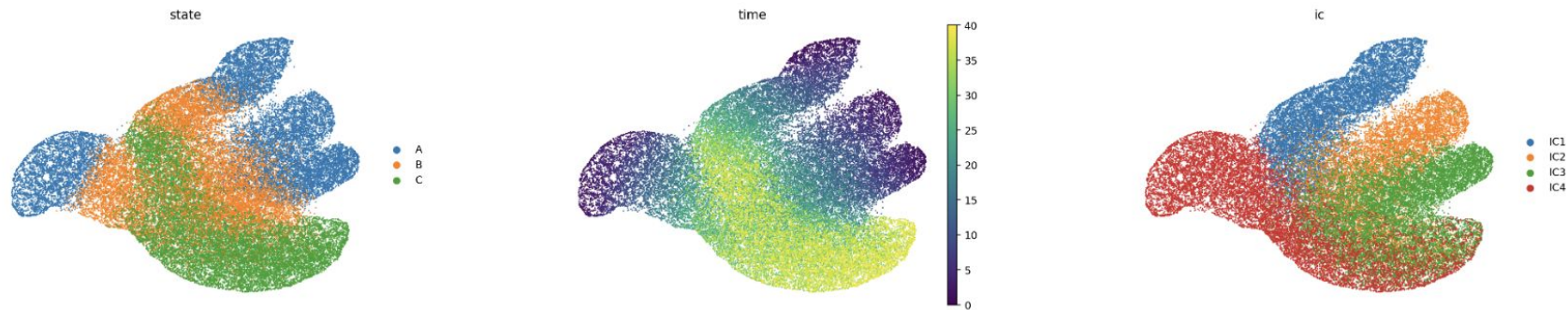
- Simulate 4 trajectories for 4 initial conditions
- Manually Assign the state labels



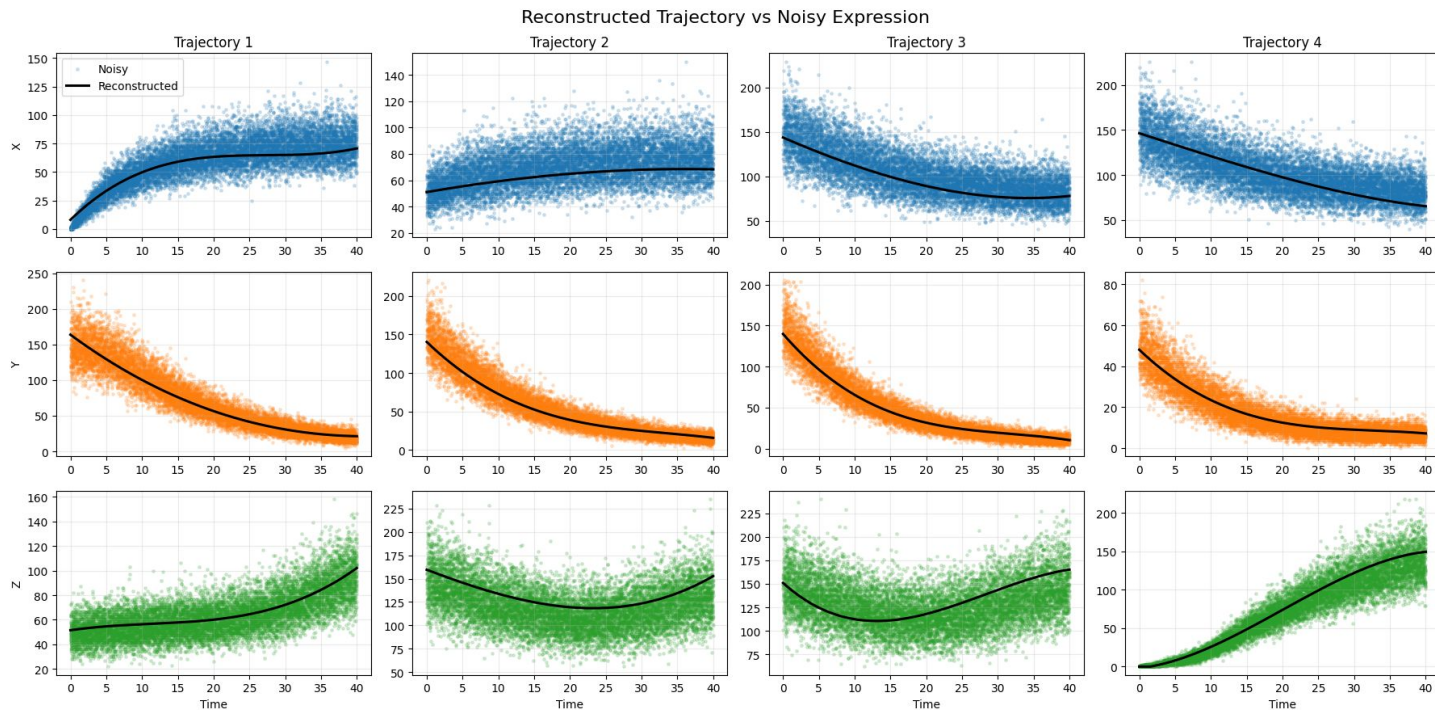
Convert Trajectories to scRNA-seq Like Data



Final Synthetic Data



Reconstructed Trajectories



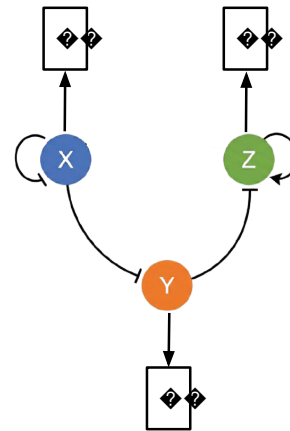
Let's apply SINDy on the reconstructed trajectories

Predicted Dynamics

$$\frac{dX}{dt} = 0.404125 - 0.0267433X + 5.79453 \text{Rep}(X; 50, 4)$$

$$\frac{dY}{dt} = 0.124928 - 0.0754461Y + 4.87418 \text{Rep}(X; 40, 4)$$

$$\frac{dZ}{dt} = 0.346258 - 0.0501905Z + 4.41363 \text{Act}(Z; 50, 2) + 5.96386 \text{Rep}(Y; 30, 4)$$

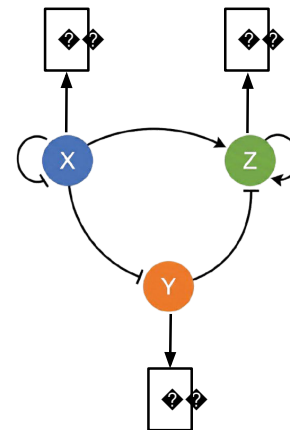


Ground Truth

$$\frac{dX}{dt} = 0.3 + 8.14 \cdot \text{Rep}(X; 50, 2) - 0.035X$$

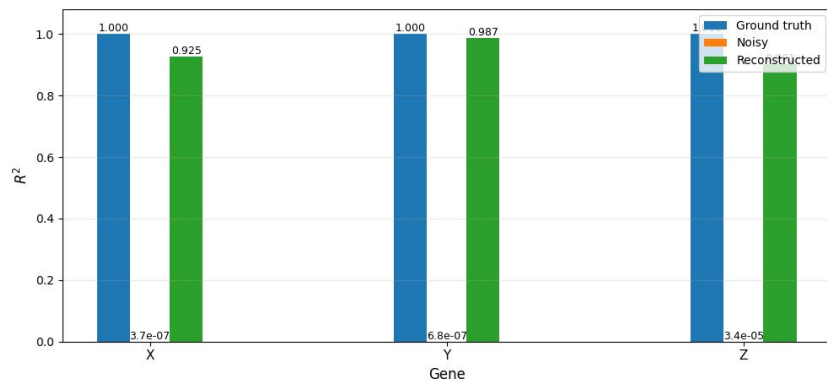
$$\frac{dY}{dt} = 0.02 + 9.26 \cdot \text{Rep}(X; 40, 4) - 0.08Y$$

$$\frac{dZ}{dt} = 0.0028 + 0.28 \cdot \text{Act}(X; 50, 3) + 5 \cdot \text{Rep}(Y; 30, 4) + 2.5 \cdot \text{Act}(Z; 20, 4) - 0.04Z$$

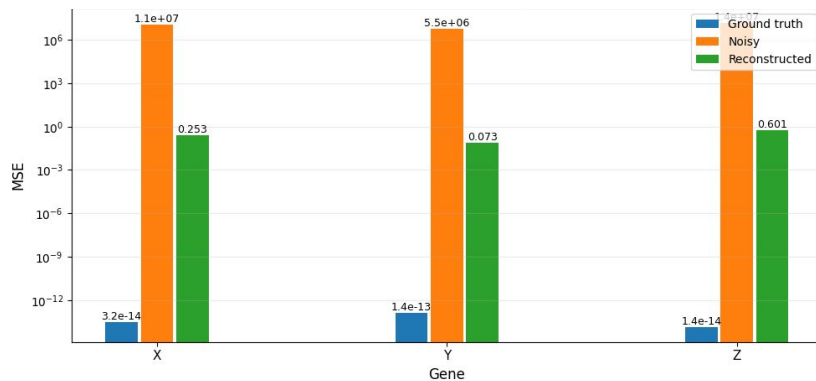


Quantitative Results

$R^2 (\uparrow)$



MSE ($\downarrow$)



Next Steps

1. Take care of drop out
2. Apply the framework on real scRNA-seq trajectory
 - EMT (Epithelial to Mesenchymal Transition)
 - HSC Differentiation

Thanks!