



# From Noisy Single-Cell Trajectories to Interpretable Gene Regulatory Dynamics

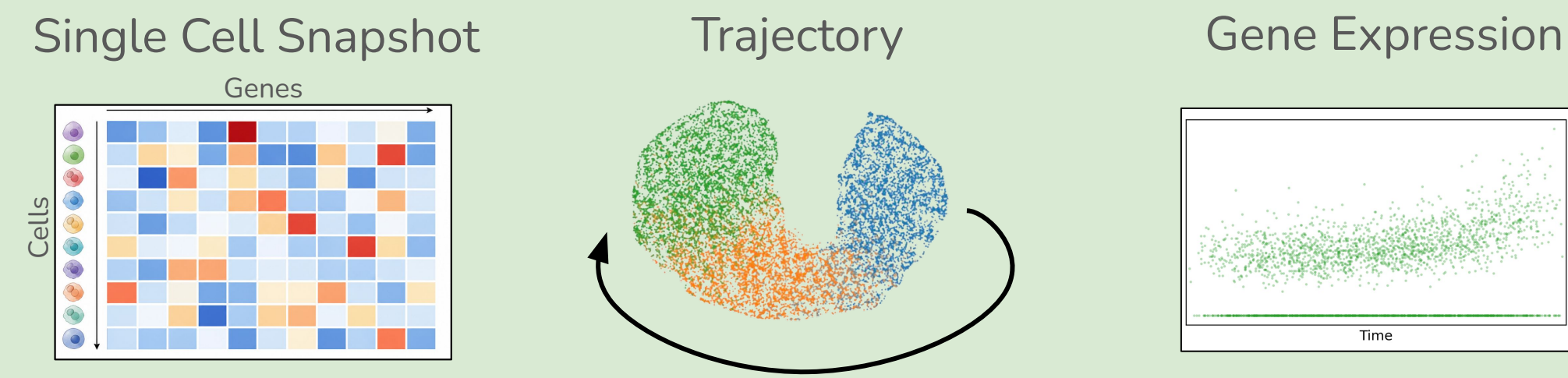
Samia Islam<sup>1,3</sup>, Zijun Cui<sup>1</sup>, Sudin Bhattacharya<sup>2,3</sup>

<sup>1</sup>Department of Computer Science and Engineering, <sup>2</sup>Departments of Biomedical Engineering and Pharmacology & Toxicology, <sup>3</sup>Institute 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 underlying dynamics of gene regulation?**

## ZINB-VAE [1]

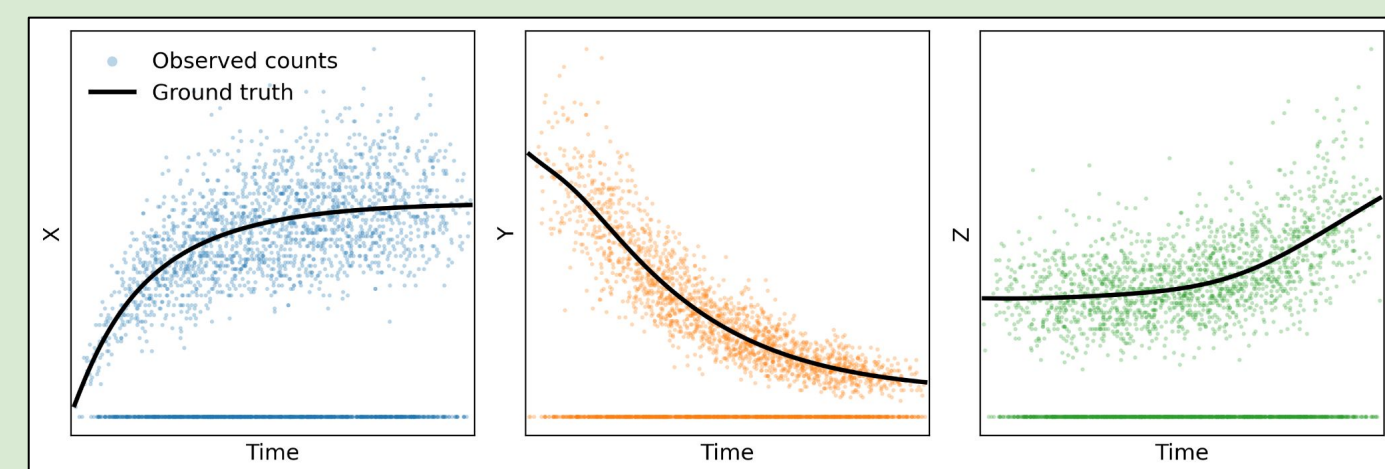
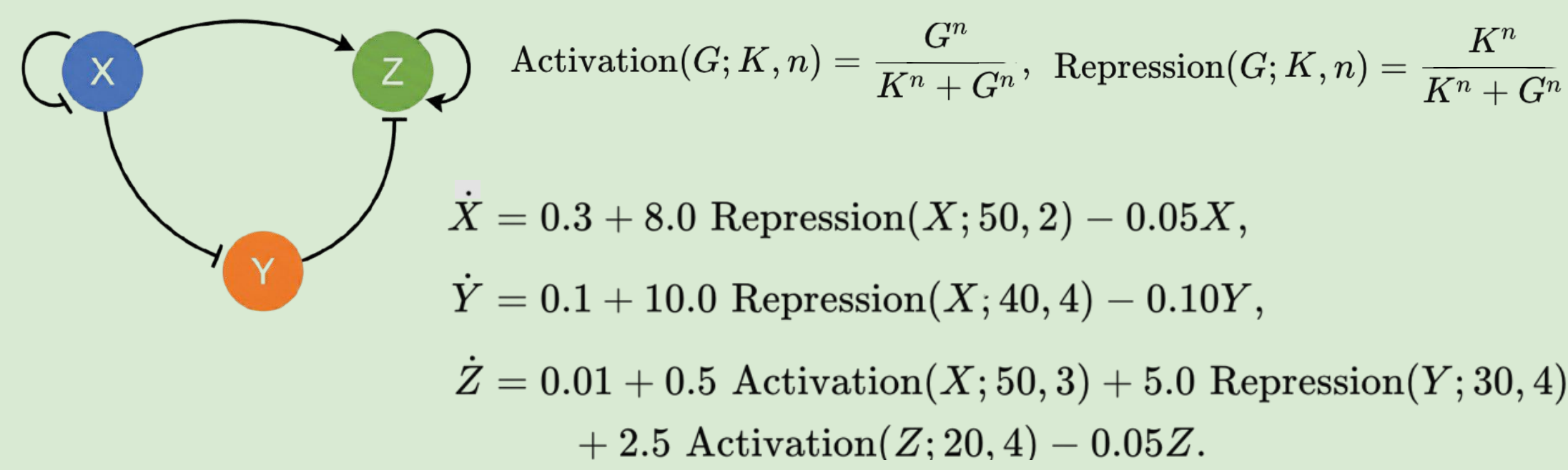
- The encoder maps high-dimensional cell profile to a low-dimensional latent distribution
- The decoder produces estimated pre-dropout mean expression, count dispersion, and excess-zero probability for each cell
- The decoder's estimated mean is used for downstream trajectory reconstruction

## Prior-Informed SINDy [2]

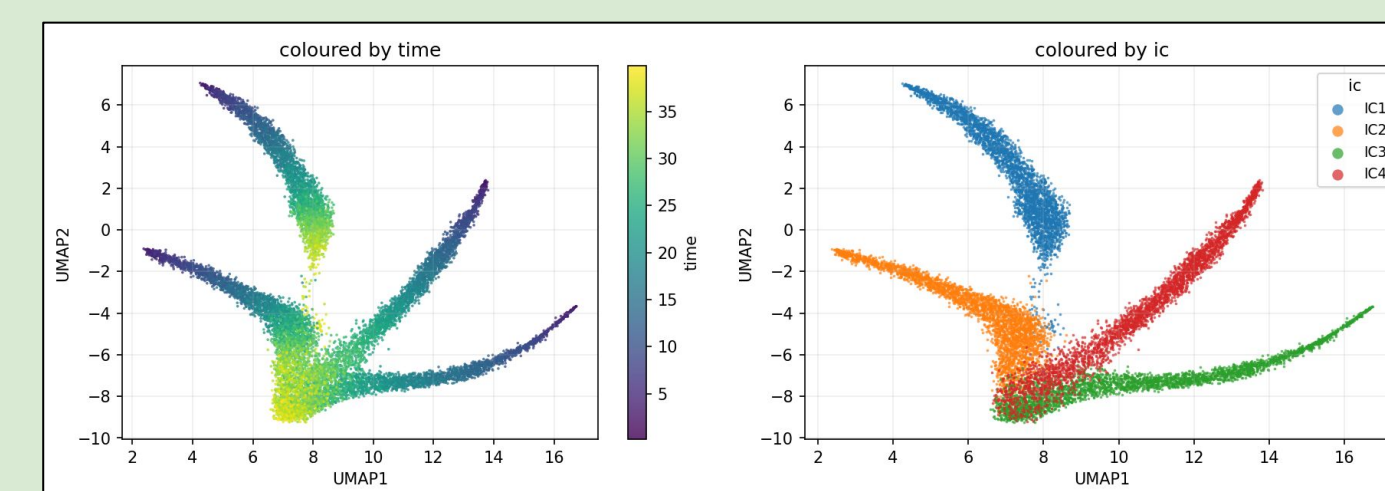
- We constructed a biologically motivated candidate library:
  - Basal production
  - Hill activation and repression
  - Linear degradation
- Optionally supplied a regulatory prior specifying the candidate activation or repression terms

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

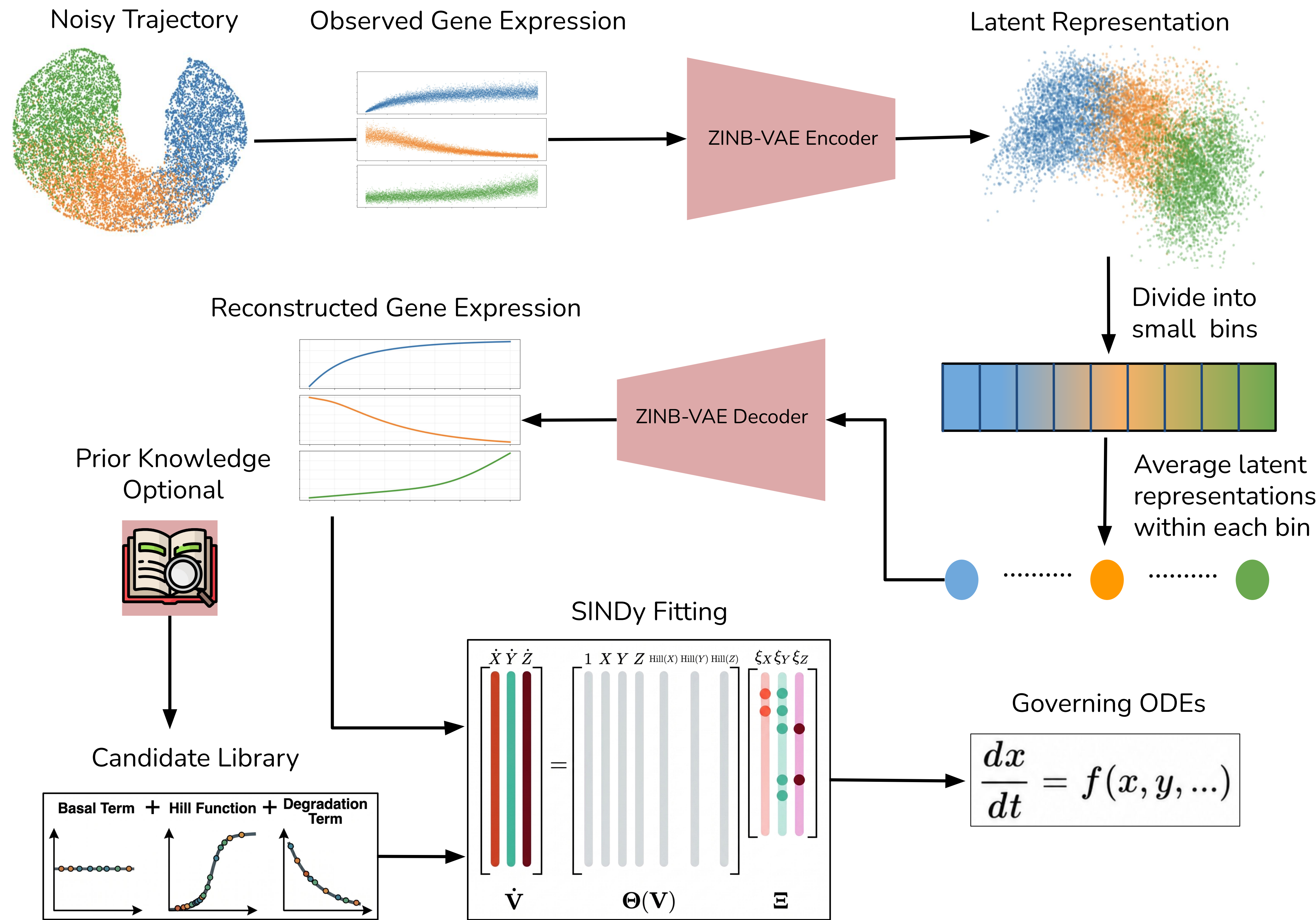
## A Synthetic Benchmark with 3 Core TFs



Ground Truth vs Zero-Inflated Negative Binomial expressions

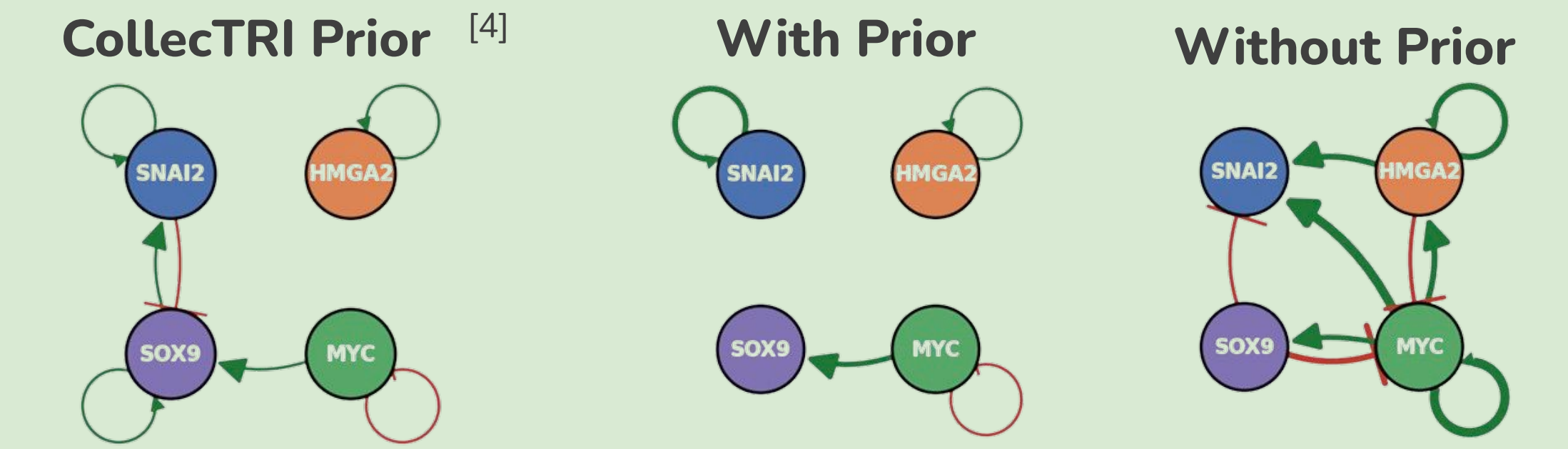


Four Generated Trajectories with 1000 genes



## Results from EMT (HuMEC) Data (Train) [3]

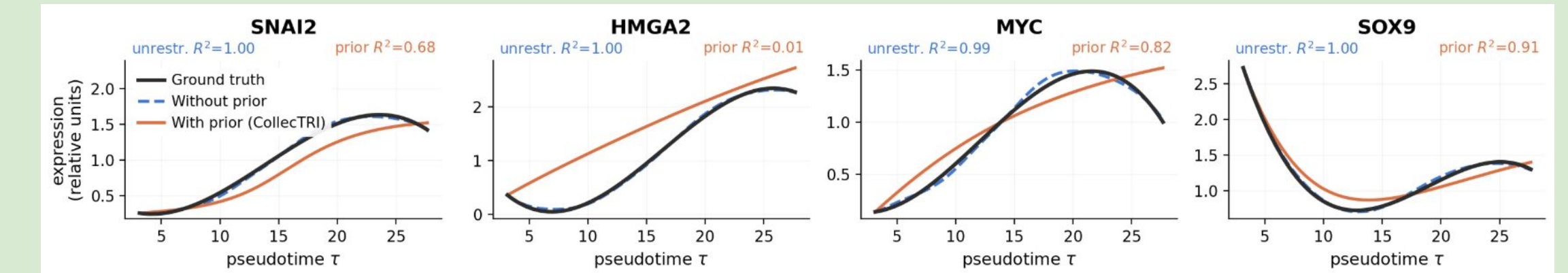
### Core TF Regulatory Interaction



### Predicted Regulatory Equations (without prior)

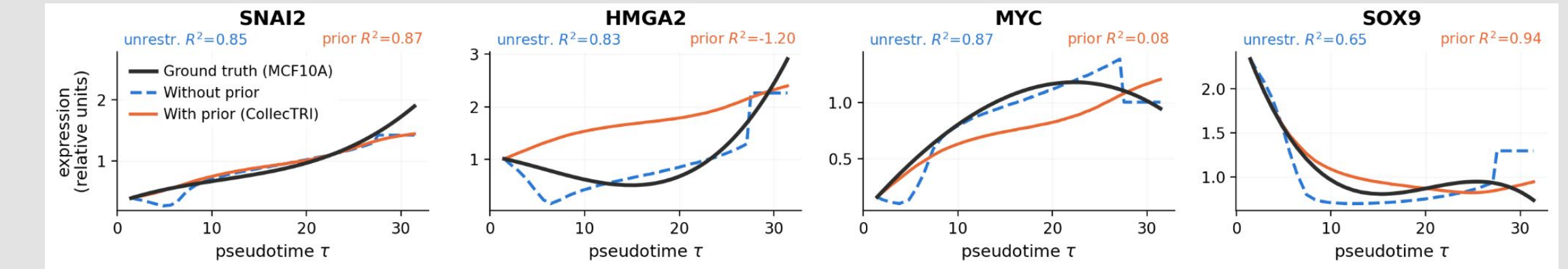
$$\begin{aligned} \frac{dSNAI2}{dt} &= 0.268 + 2.089 h^+(MYC; 1.09, 2) + 0.748 h^+(HMGA2; 1.64, 3) + 0.124 h^-(SOX9; 0.74, 2) - 1.338 SNAI2 \\ \frac{dHMGA2}{dt} &= 0.695 h^+(MYC; 0.78, 3) + 0.860 h^+(HMGA2; 1.64, 3) - 0.515 HMGA2 \\ \frac{dMYC}{dt} &= 2.156 h^+(MYC; 1.32, 3) + 0.227 h^-(HMGA2; 0.42, 3) + 0.724 h^-(SOX9; 0.74, 3) - 0.946 MYC \\ \frac{dSOX9}{dt} &= 0.024 + 0.475 h^+(MYC; 1.32, 3) - 0.191 SOX9 \end{aligned}$$

### Core TF Expressions

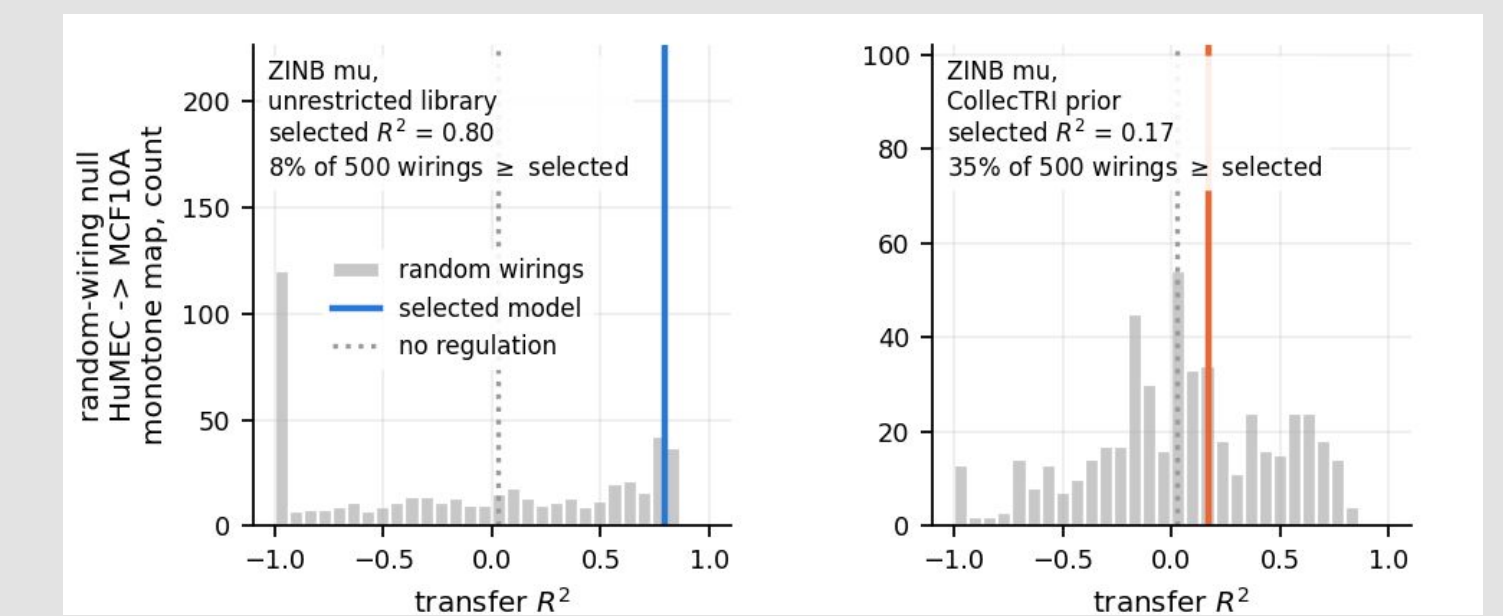


## Results from EMT (MCF10A) Data (Test) [3]

### Core TF Expressions



### Transfer score comparison with randomly wired models



## Key Findings

- ZINB-VAE successfully reconstructs underlying gene-expression trajectories from sparse, noisy single-cell counts
- Hill-constrained SINDy recovers interpretable regulatory ODEs and the system's underlying dynamics.
- For EMT, the inferred dynamics better explain and predict gene-expression trajectories than models constrained by existing regulatory prior knowledge.

## References:

- Lopez, Romain, et al. "Deep generative modeling for single-cell transcriptomics." Nature methods 15.12 (2018): 1053-1058.
- Brunton, Steven L., Joshua L. Proctor, and J. Nathan Kutz. "Discovering governing equations from data by sparse identification of nonlinear dynamical systems." Proceedings of the national academy of sciences 113.15 (2016): 3932-3937.
- McFaline-Figueroa, José L., et al. "A pooled single-cell genetic screen identifies regulatory checkpoints in the continuum of the epithelial-to-mesenchymal transition." Nature genetics 51.9 (2019): 1389-1398.
- Müller-Dott, Sophia, et al. "Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities." Nucleic acids research 51.20 (2023): 10934-10949.
- Bhattacharya, Sudin, Qiang Zhang, and Melvin E. Andersen. "A deterministic map of Waddington's epigenetic landscape for cell fate specification." BMC systems biology 5.1 (2011): 85.