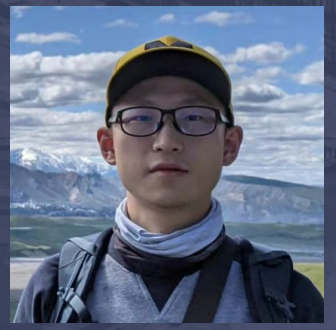
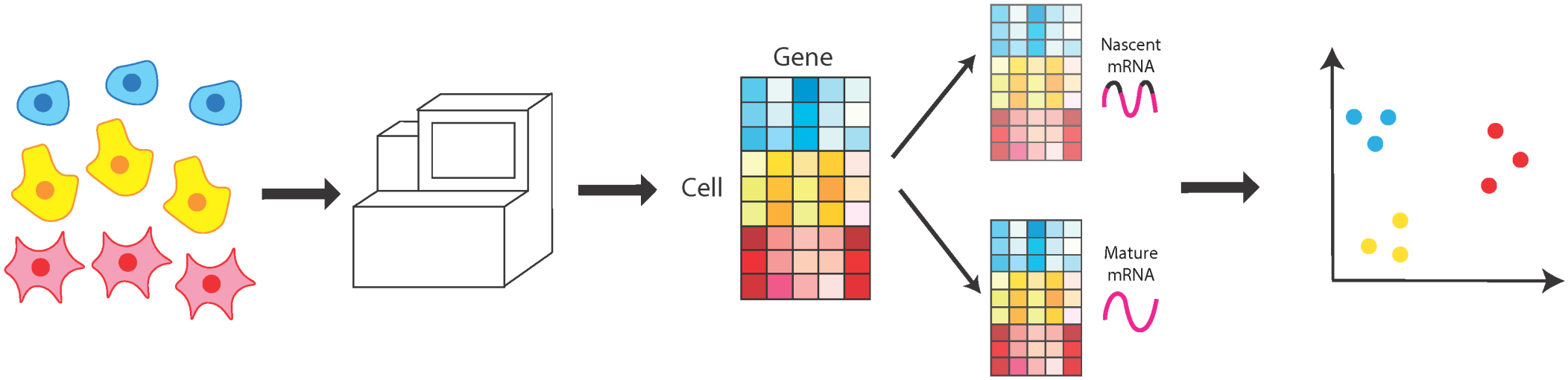
Variational Mixtures of ODEs for Inferring Cellular Gene Expression Dynamics

Yichen Gu, David Blaauw, Joshua Welch



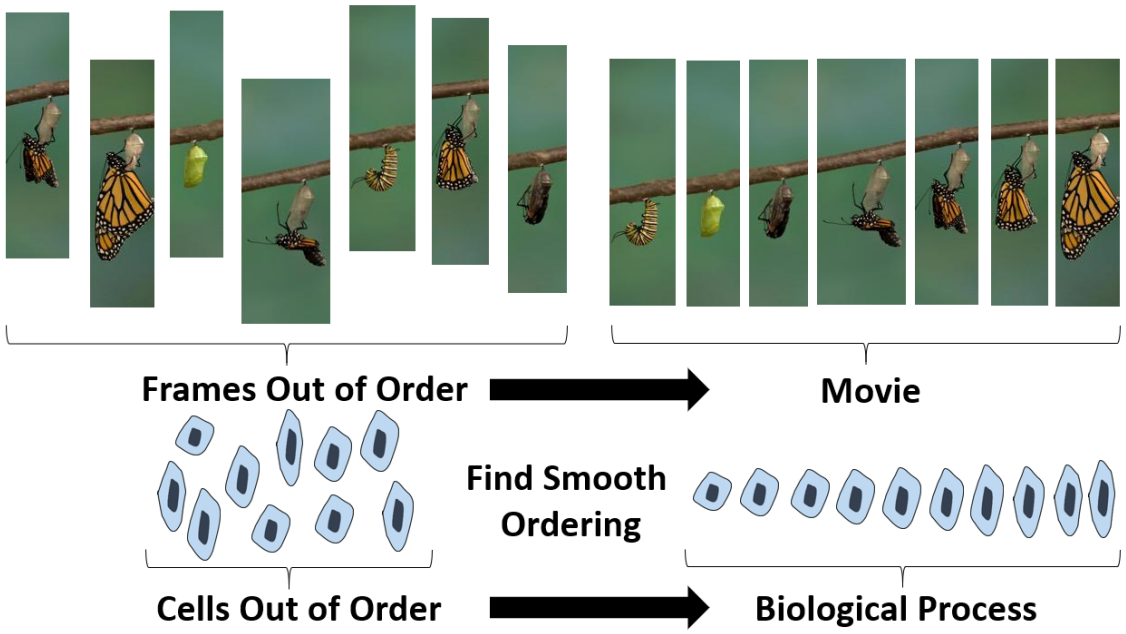
Single-Cell RNA Sequencing

- Gene expression profiling at single-cell resolution.
- Cell-by-gene mRNA count matrix (snapshots) $\mathbf{X}$.
  - $X_{ij} :=$ mRNA count of the i-th cell and j-th gene
  - Cell time is unknown



Problem Formulation

- Given: mRNA count matrices **U**, **S**
- Assumptions
  - **U**, **S** are generated by a system of ODEs plus noise
  - Temporal information is absent
  - Unique solution $F(t; \boldsymbol{\theta})$ (Kinetic Functions)
- Goal
  - Infer the latent time, t
  - Predict future states using F



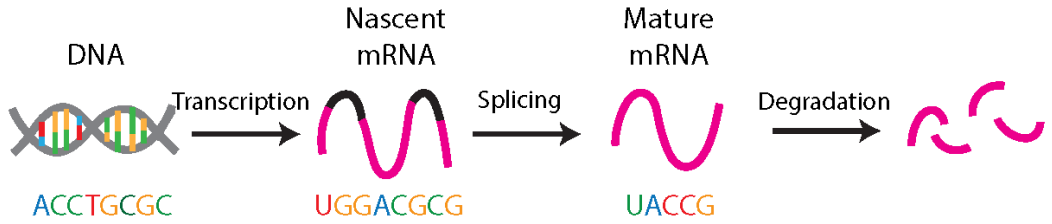
RNA Velocity

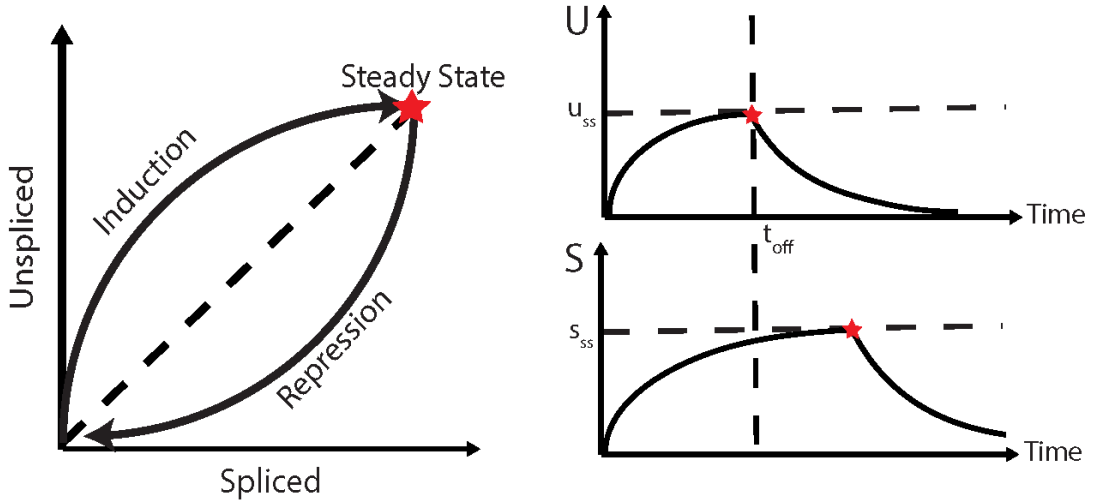
- Key Observation: Splicing
- ODE Model

$$\frac{du}{dt} = \alpha - \beta u$$

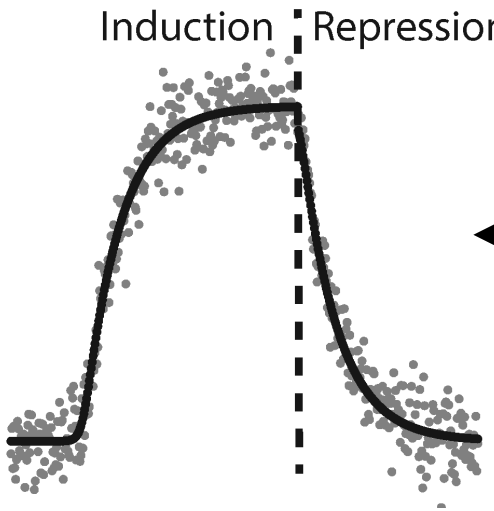
$$\frac{ds}{dt} = \beta u - \gamma s$$

- RNA Velocity $:= \frac{ds}{dt}$



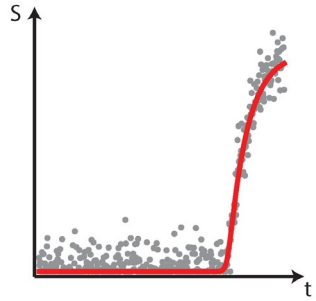


Limitations of the SOTA Method

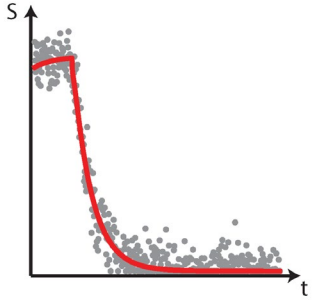


scVelo

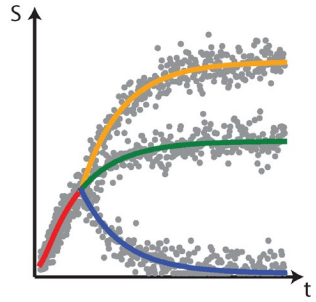
Late Induction



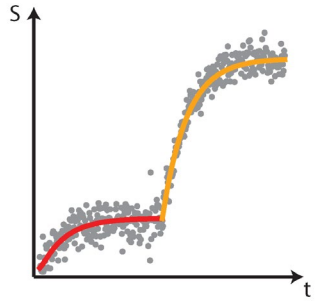
Early Repression



Cell Type Multifurcation



Variable Rates



Variational Mixtures of ODEs

Inference Model

$$q_{\phi}(\boldsymbol{c}, t|\boldsymbol{x}) \sim N\big(h(\boldsymbol{u}, \boldsymbol{s}; \phi)\big)$$

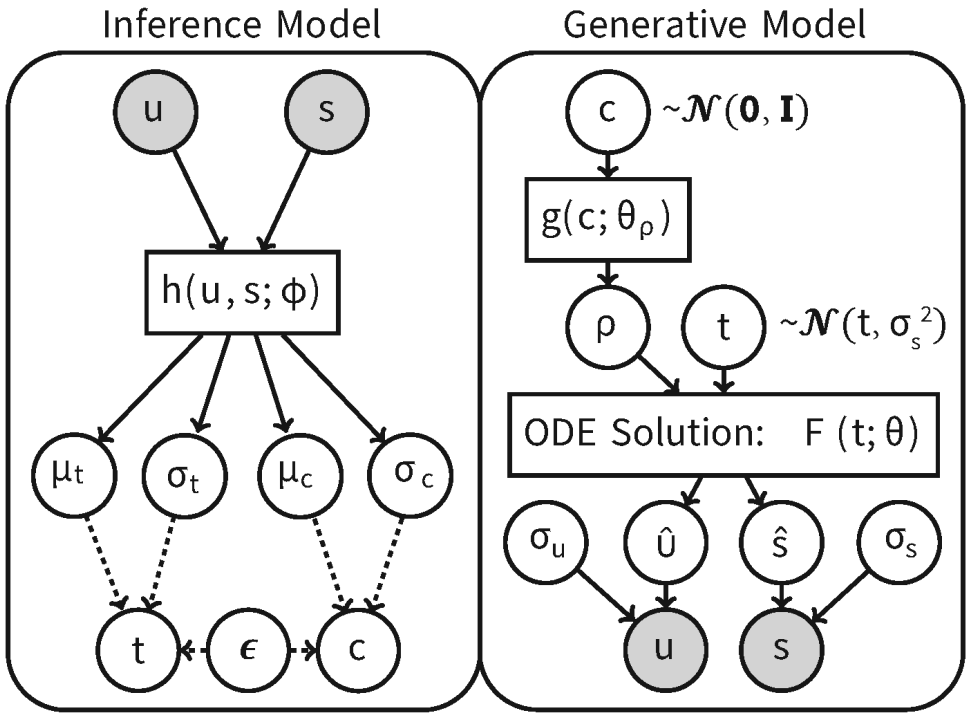
Generative Model

$$t \sim N(t_0, \sigma_0^2), \boldsymbol{c} \sim N(\mathbf{0}, \mathbf{I})$$

$$\boldsymbol{\rho} = g(\boldsymbol{c})$$

$$\tilde{\boldsymbol{\alpha}} = \boldsymbol{\rho} \odot \boldsymbol{\alpha}$$

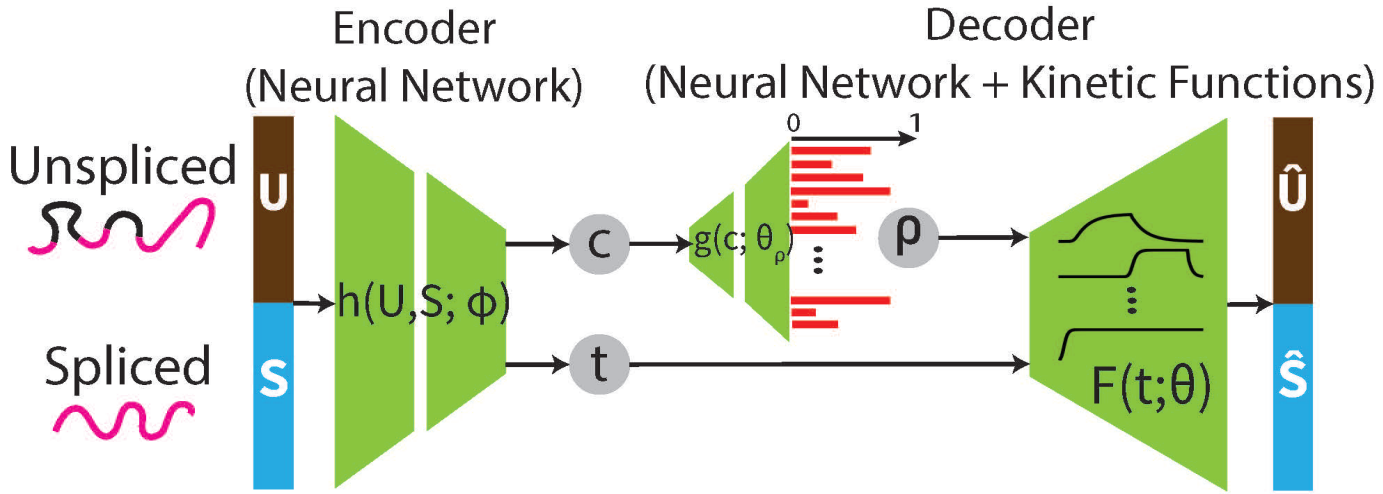
$$\boldsymbol{x} \sim N(\boldsymbol{F}(t; \boldsymbol{\theta}), \boldsymbol{\Sigma_r})$$



VeloVAE

Training Objective – ELBO

$$\sum_{i \in B} \left[-KL\left(q_{\phi}(\boldsymbol{c}, t|\boldsymbol{x_i}) || p(\boldsymbol{c}, t)\right) + \underbrace{\mathbb{E}_{q_{\phi}(\boldsymbol{c},t|\boldsymbol{x_i})}}_{\text{Approximated by Sampling}} [\log p_{\theta}(\boldsymbol{x_i}|\boldsymbol{c}, t)] \right]$$



Results

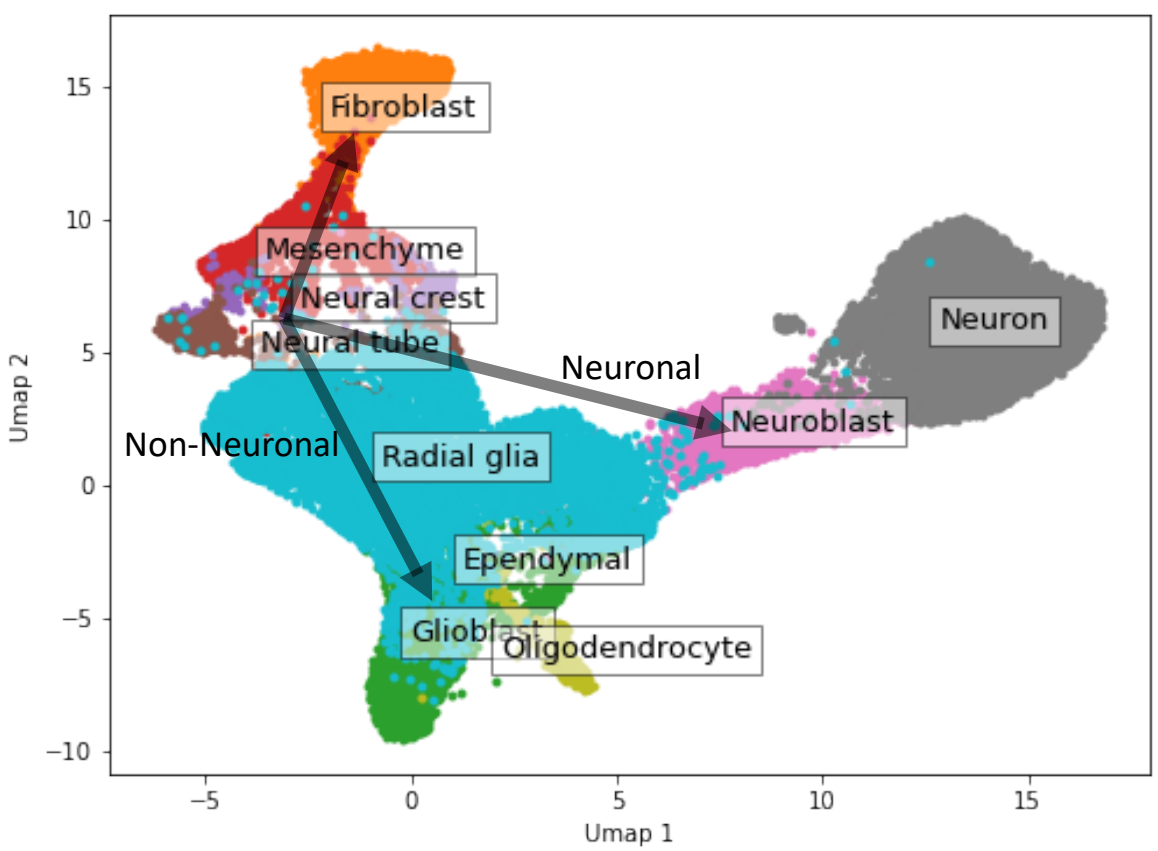
VeloVAE achieved a SOTA performance on data reconstruction.

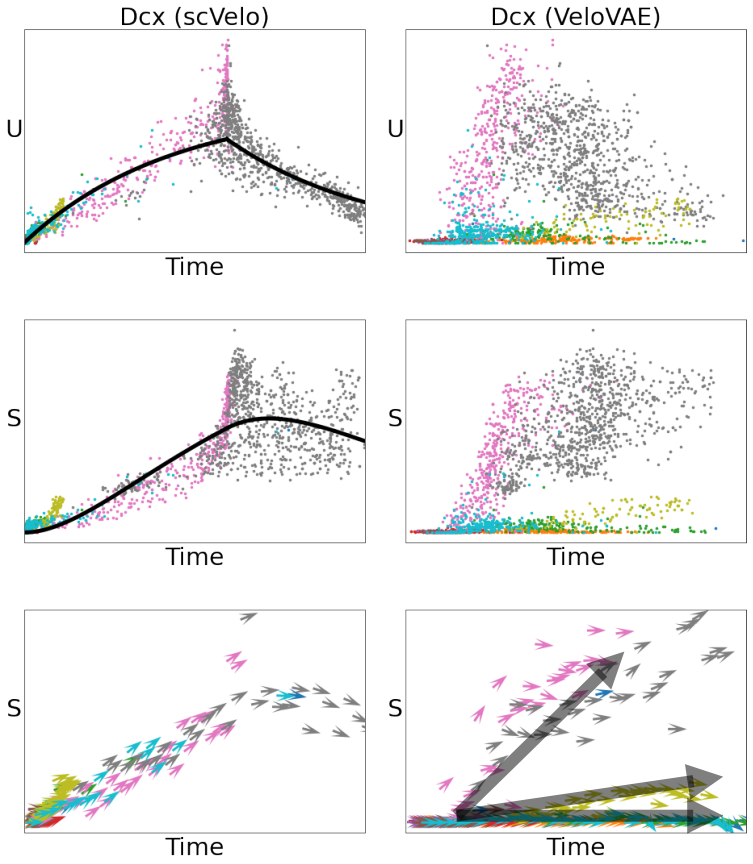
Dataset	MSE		Time Correlation	
	scVelo	VeloVAE	scVelo	VeloVAE
Pancreas	2.107	**0.823**	N/A	N/A
Dentate Gyrus #1	0.670	**0.243**	N/A	N/A
Mouse Brain #1	10.160	**1.886**	N/A	N/A
Erythroid	0.873	**0.151**	-0.707	**0.855**
Dentate Gyrus #2	1.385	**0.159**	-0.158	**0.707**
Mouse Brain #2	18.19	**0.152**	-0.777	**0.897**

Table 1: Performance Comparison

Results – Mouse Brain

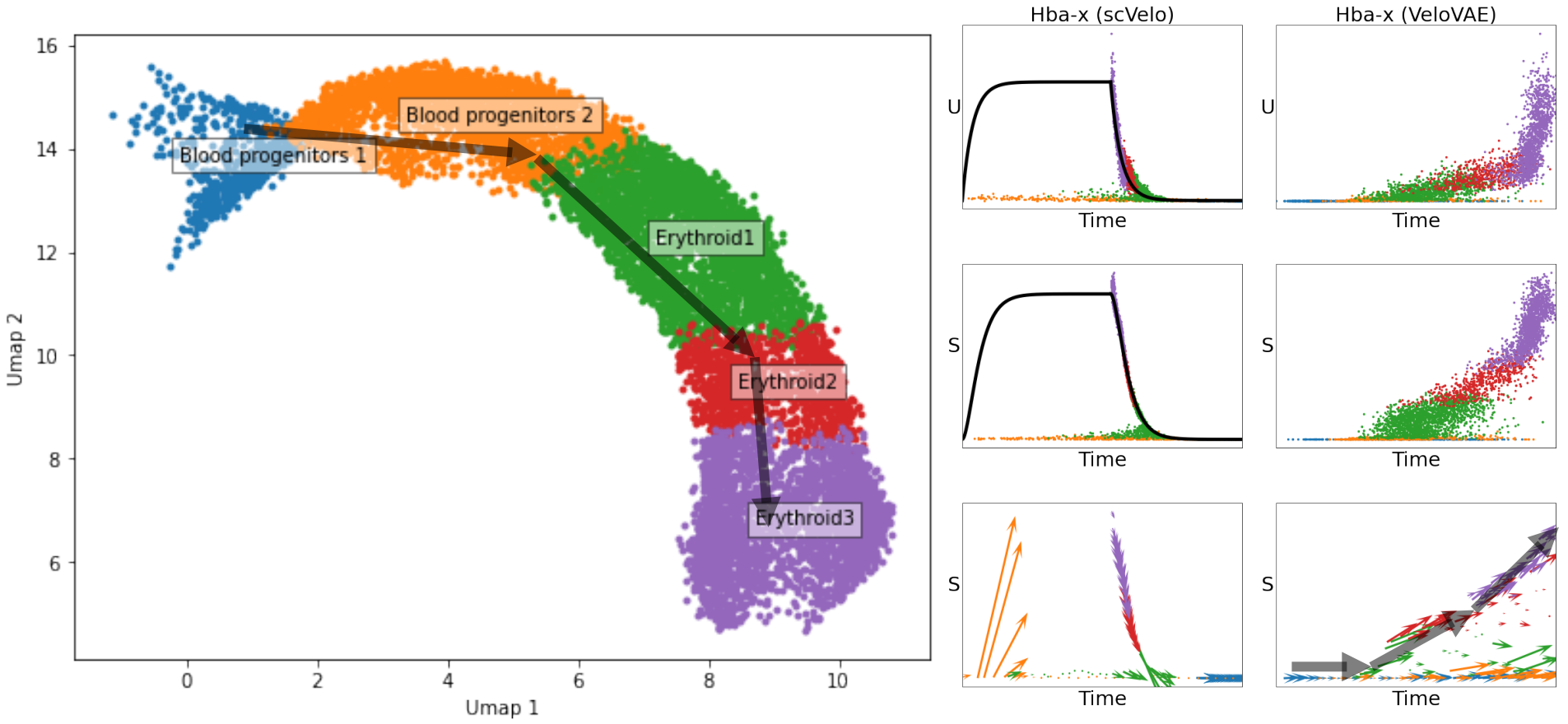
VeloVAE learns branching dynamics during differentiation.





Results – Erythroid

VeloVAE identifies transcriptional boosts in erythroid lineage.



UNIVERSITY OF
MICHIGAN

ADVANCED RESEARCH
COMPUTING
UNIVERSITY OF MICHIGAN

NIH
National Institutes
of Health

Thank you!