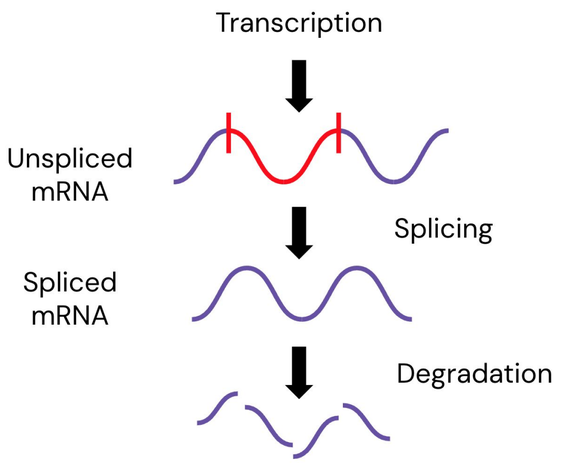
CellFlows: Inferring Splicing Kinetics from Latent and Mechanistic Cellular Dynamics

Sei Chang[1,2*], Zaiqian Chen[1,2*], Bianca Dumitrascu[1,3], David A. Knowles[1,2]

Columbia University[1], New York Genome Center[2], Irving Institute for Cancer Dynamics[3]

Introduction to RNA Velocity in Single Cell Biology

mRNA molecules are created (**transcription**), partially removed (**splicing**), then degraded (**decay**)

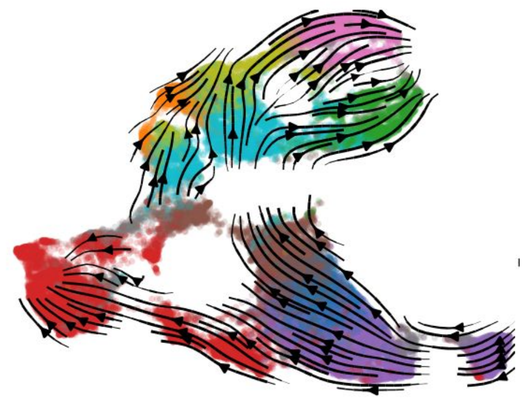


RNA regulatory dynamics are modeled using **ordinary differential equations** with kinetic rates to represent the **RNA Velocity** from unspliced to spliced *(du/dt)*

$$\frac{du}{dt} = \alpha(t) - \beta(t)u(t)$$
$$\frac{ds}{dt} = \beta(t)u(t) - \gamma(t)s(t)$$

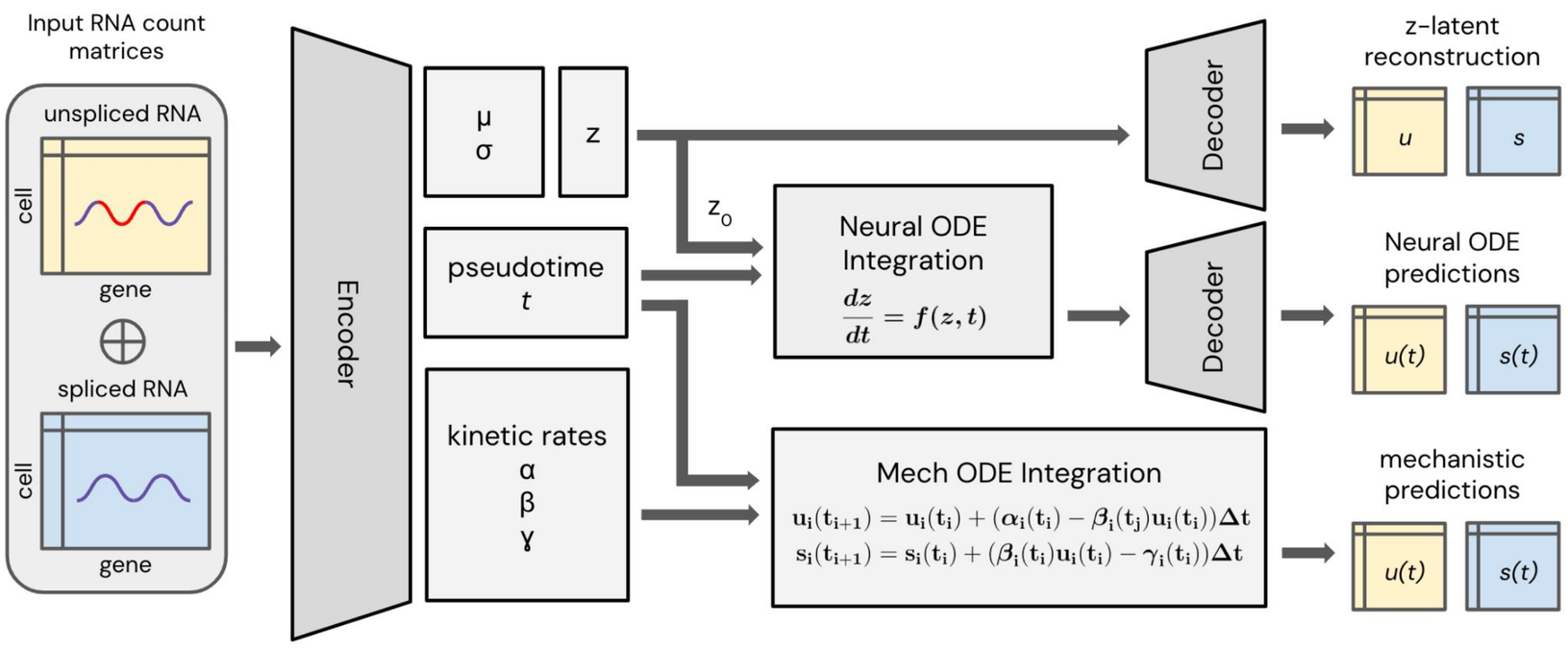
$u(t)$	unspliced counts	$\alpha(t)$	transcription rate
$s(t)$	spliced counts	$\beta(t)$	splicing rate
t	pseudotime	$\gamma(t)$	gamma rate

RNA velocity models learn an ODE from single-cell unspliced and spliced counts to estimate **cell trajectories**, **kinetic rates**, and cell-specific developmental **pseudotime**



CellFlows Methods and Architecture

CellFlows learns ODEs across both VAE-based latent and original data representations using a **shared encoded pseudotime** to derive a **neural and mechanistic vector field** for estimation of cell-gene-specific kinetic rates



$$\mathcal{L}(\theta) = \underbrace{-\log p(x|z_{\text{VAE}}) + D_{KL}(q(z_{\text{VAE}}|x)||p(z_{\text{VAE}}))}_{\text{Variational Autoencoder (VAE) Loss}} \underbrace{+ ||x - \hat{x}_{\text{ODE}}||_2^2}_{\text{Neural ODE Loss}} \underbrace{+ ||x - \hat{x}_{\text{MECH}}||_2^2}_{\text{Mechanistic ODE Loss}}$$

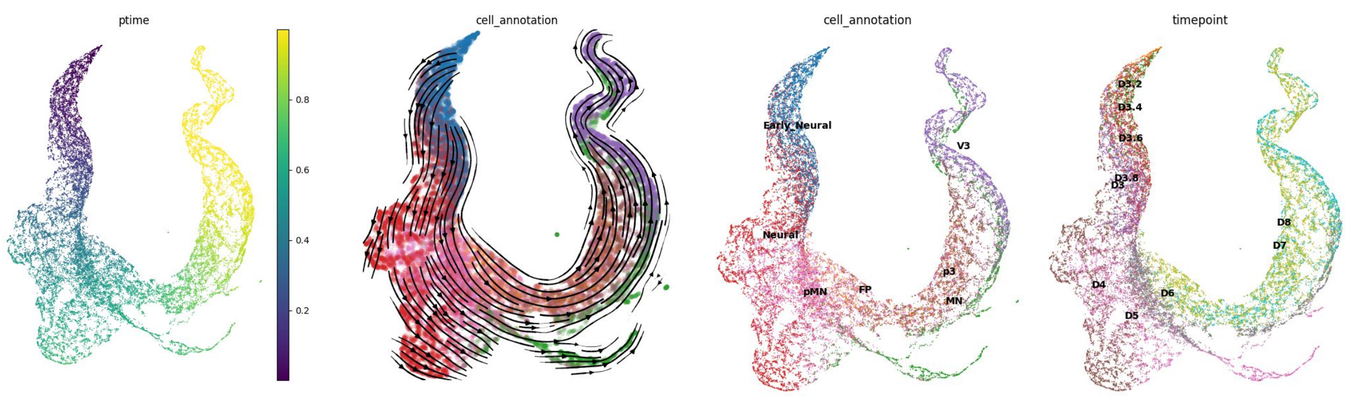
CellFlows offers **key architectural and feature improvements** over prior methods

1. **Robust Inference**. RNA velocity methods rely on extensive data preprocessing which can distort model outputs. CellFlows only requires raw RNA counts.
2. **Joint inference of pseudotime and kinetic rates.** Some RNA Velocity methods learn both separately or eschew kinetics entirely. CellFlows learns the pseudotime and kinetics jointly in a unified framework.
3. **Cell and gene-specific kinetics.** Some methods infer uniform gene kinetics across cells. CellFlows infers cell and gene-specific kinetic rates to obtain genes that drive development of cell subpopulations.

Method	Cell-gene	Raw Counts	Joint Inf.
CellFlows*	√	√	√
cellDancer	√	×	×
veloVI	×	×	√
scTour	×	√	√

Application: Trajectory Inference on Time-Series Single-Cell Data

As shown on UMAP 2D visualizations, CellFlows maps **accurate trajectories** and **pseudotimes** for development of early neural cell stem cells to V3 interneurons, motor neurons (MN), and floor plate (FP) cells

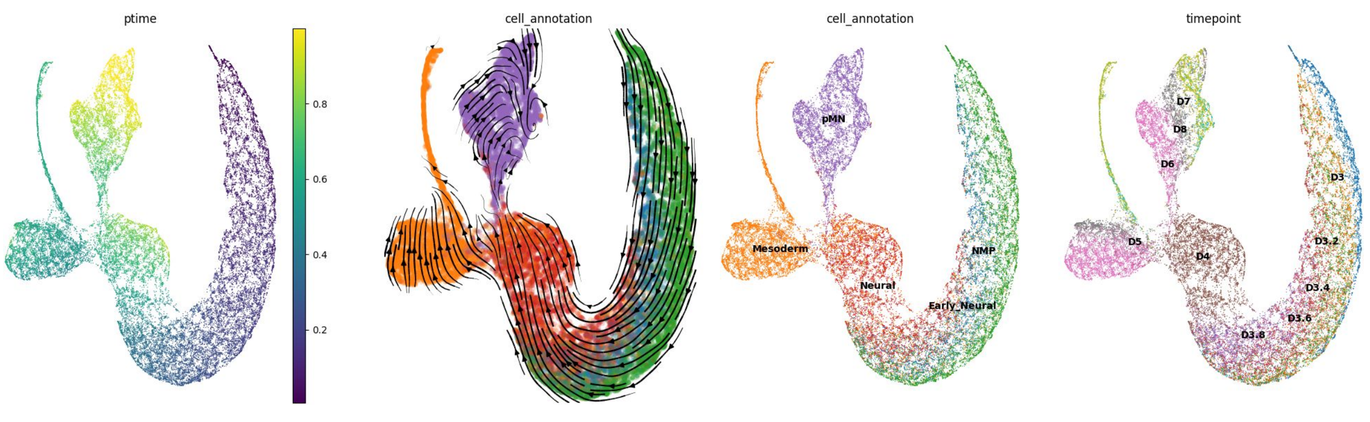


CellFlows **outperforms** published RNA velocity-based methods in estimating **pseudotimes** well correlated with ground truth **measured time**

Method	Erythroid	Neural	Mesoderm
CellFlows*	**0.816**	0.802	**0.920**
cellDancer	0.739	0.738	0.765
veloVI	0.605	0.279	0.662
scTour	0.791	**0.846**	0.917

Spearman correlation between measured time and model-estimated pseudotime

CellFlows maps **accurate trajectories** and **pseudotimes** for development of neural mesoderm progenitors (NMP) to pMN and mesoderm cells



CellFlows estimates pseudotimes consistent with RNA expression of **key marker genes** for cell types in the lineage

	Neural					
Method	E-Neural	pMN	p3	FP	MN	V3
CellFlows*	**0.301**	0.124	0.371	0.094	**0.425**	**0.489**
cellDancer	0.198	**0.280**	0.370	0.090	0.253	0.371
veloVI	0.128	0.006	0.056	0.087	0.127	0.234
scTour	0.230	0.164	**0.398**	**0.132**	0.416	0.248

Spearman correlation between pseudotime and marker gene expression for each cell type in neural and mesoderm lineage

Conclusion: CellFlows offers a powerful and adaptable framework for studying latent dynamical systems beyond RNA velocity and single-cell genomics. The model combines mechanistic equations with flexible neural networks, making it suitable for any field with complex dynamics. Future research could extend CellFlows to more complex systems, include additional data modalities, or enable transfer learning across related dynamical systems.