MAPPING SINGLE CELL DIFFERENTIATION FATES WITH FATEODE

Anonymous authors
Paper under double-blind review

ABSTRACT

Single-cell transcriptomics provides a powerful lens for deciphering cellular differentiation dynamics and understanding how phenotypic diversity emerges. However, because sequencing is inherently destructive, temporal progression cannot be tracked longitudinally within individual physical cells. Existing paradigms therefore diverge into two separate regimes: recovering macroscopic differentiation trends via trajectory inference, or approximating instantaneous dynamics through RNA velocity. Fundamentally, velocity and trajectory are the differential (local derivative) and integral (global path) manifestations of the same underlying dynamical system. While recent continuous dynamical frameworks (such as Neural ODEs) attempt to unify these perspectives, they exhibit three critical deficiencies: 1) failing to strictly enforce bidirectional consistency between velocity fields and trajectory integrals; 2) insufficiently accommodating technical measurement noise inherent to sequencing; and 3) inaccurate depiction of intrinsic stochasticity in cell fate bifurcations. Consequently, inferred dynamics frequently drift along biologically implausible off-manifold paths or overfit local noise, yielding erratic and overly tortuous trajectories. To overcome these limitations, we propose **FateODE**, a stochastic Neural ODE framework that enforces rigorous derivative-integral consistency while capturing manifold-constrained stochastic fate transitions. FateODE harmonizes macroscopic trajectories with splicing velocities by anchoring continuous integration across discrete waypoint observations and regularizing point-wise derivatives with high-confidence velocity vectors. To capture bifurcation stochasticity without sacrificing manifold geometry, FateODE introduces an inference-time diffusion term parameterized by a decoder-only variational architecture under Maximum Mean Discrepancy (MMD) supervision, faithfully simulating Brownian motion directly on the transcriptomic manifold. Beyond reconstructing smooth developmental vector fields, FateODE functions as a continuous "world model" capable of projecting cellular fates, prioritizing differentiation driver genes, and forecasting fate-landscape shifts under in-silico genetic perturbations. Extensive benchmarks across synthetic and real lineage-traced datasets demonstrate FateODE's superior dynamic modeling fidelity, accurately predicting terminal fate distributions under both unperturbed and perturbed conditions. `https://anonymous.4open.science/r/FateODE-83D8/`

1 INTRODUCTION

Cellular differentiation is a dynamic and stochastic process in which progenitor cells acquire specialized identities through changes in gene regulation Stumpf et al. (2017); Weinreb et al. (2020). Reconstructing these dynamics is essential for understanding lineage commitment and predicting cell fate under various perturbations. Single-cell RNA sequencing (scRNA-seq) provides snapshots of the underlying transcriptomic states, but destructive sampling prevents longitudinal observation of the same cell, while technical measurement noise obscures continuous biological transitions.

Despite these hurdles, computational efforts to reconstruct differentiation kinetics from static snapshot data have largely coalesced around two distinct paradigms: i) **Trajectory inference and pseudotime analysis** Trapnell et al. (2014); Qiu et al. (2017); Street et al. (2018): These approaches leverage geometric or topological similarities across transcriptomic manifolds to infer macroscopic differentiation paths, ordering cells along continuous pseudotemporal coordinates to chart global

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

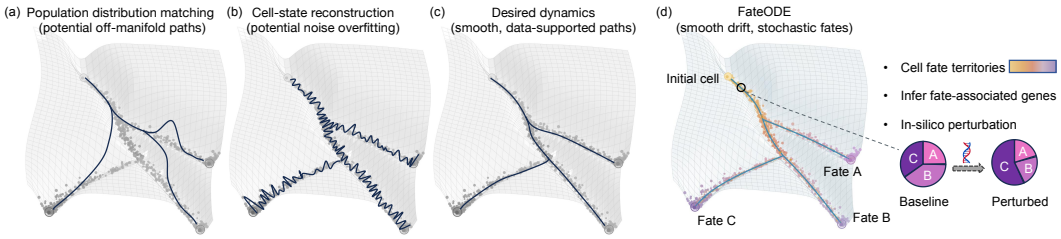

Figure 1: **Motivation of FateODE.** (a,b) NeuralODE failure modes: off-manifold drift and overfitting technical noise. (c) Ideal dynamics: smooth, on-manifold paths. (d) FateODE combines smooth drift with geometry-guided stochastic transitions to estimate cell fate territories, identify fate-associated genes, and simulate perturbation-induced shifts in fate probabilities.

progressions from stem cell pools to mature endpoints. ii) **RNA splicing kinetics and velocity** Bergen et al. (2020); Weiler et al. (2024); Li et al. (2024); Gao et al. (2022): These methods simultaneously quantify unspliced (pre-mRNA) and spliced (mature mRNA) transcripts within single cells. By fitting kinetic ordinary differential equation (ODE) models of transcription, splicing, and degradation, they approximate instantaneous directional velocities in gene expression space.

Drawing an analogy to classical dynamical systems in physics, velocity and trajectory are fundamentally the differential and integral manifestations of the same underlying dynamical process: instantaneous velocity is the local time derivative of the trajectory ($\mathbf{v}(t) = \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}$), while the trajectory is the path integral of the velocity field ($\mathbf{x}(t) = \mathbf{x}(0) + \int_0^t \mathbf{v}(\tau)\,\mathrm{d}\tau$). These two aspects are naturally connected through Calculus. Constructing a continuous dynamical framework that unifies these two perspectives is therefore of paramount importance, offering the following capabilities: 1) **Elucidating molecular cascades:** Resolving continuous, fine-grained temporal expression profiles across fleeting intermediate cellular states. 2) **Pinpointing fate bifurcations:** Identifying the exact tipping points where cells irreversibly commit to specific lineages and ranking the upstream drivers governing each transition. 3) **Delineating fate landscapes:** Quantifying the residual differentiation plasticity of progenitor populations directly on the transcriptomic manifold. 4) **Serving as an in-silico "world model" for perturbation analysis:** Simulating the transcriptomic outcomes of genetic perturbations (e.g., targeted knockouts or overexpressions) and forecasting shifts in terminal lineage proportions, establishing a testbed for directed cell reprogramming simulations.

Recognizing that cellular differentiation can be naturally formulated as a continuous dynamical system, several recent studies have applied differential equations and deep learning to model transcriptomic trajectories Gu et al. (2026); Farrell et al. (2023); Burdziak et al. (2023); Qiu et al. (2022); Boudjelthia et al.. For instance, scKINETICS employs a linear dynamical formulation ($\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{A}\mathbf{x}$) constrained by prior gene regulatory networks (GRNs) to infer transcriptional couplings Burdziak et al. (2023). However, linear formulations lack the capacity to capture the curved geometries characteristic of cellular manifolds. In contrast, scTOUR leverages Neural ODEs to parameterize continuous vector fields in a low-dimensional latent space, fitting a continuous trajectory through pseudotime-ordered cells Li (2023). Further extending continuous modeling, frameworks such as CELLMNN, scDIFFEQ, and DYNNET utilize deep neural networks to map transcriptional velocities, primarily focusing on aligning cross-sectional datasets sampled across discrete developmental time points Vinyard et al. (2025); Dou et al. (2026). Among these, scDIFFEQ and DYNNET explicitly incorporate stochastic terms to capture process variance, with DYNNET additionally integrating graph neural networks to embed transcriptional regulatory relationships.

Despite their conceptual promises, existing approaches fail to reconcile the fundamental derivative–integral duality, leading to three pronounced failure modes: **Off-manifold trajectory drift:** Methods such as CELLMNN, scDIFFEQ, and DYNNET optimize boundary-matching objectives, frequently via Maximum Mean Discrepancy (MMD) or optimal transport between discrete time points, without constraining intermediate trajectories to the empirical developmental manifold in the transcriptomic space. Consequently, simulated cell trajectories easily drift through biologically implausible, off-manifold expression regimes. Similarly, the linear assumptions of scKINETICS fail to trace non-linear manifold geometry during integration. **Noise overfitting and velocity divergence:** While scTOUR fits continuous paths through individual cells, it optimizes trajectory interpolation

without accounting for technical measurement noise. This causes the parameterized vector field to overfit local observation noise, producing overly tortuous trajectories that diverge substantially from empirical splicing velocities. **Latent space abstraction vs. genome-wide interpretability:** Frameworks such as scTour, CellMNN, and scDiffEq operate strictly within latent space embeddings, precluding direct tracking of explicit gene-level kinetics. Conversely, while DynNet operates in gene expression space, it is restricted to preselected subgraphs of known TF-target interactions, failing to deliver a transcriptome-wide view of developmental regulation.

To resolve these limitations, we introduce **FateODE**, a stochastic Neural ODE framework that enforces rigorous derivative-integral consistency while modeling manifold-constrained stochastic fate transitions. FateODE reconciles macroscopic trajectories with microscopic splicing velocities through four key principles: 1) **Waypoint-anchored path integration:** Rather than forcing rigid point-by-point interpolation through noisy individual cells, FateODE anchors continuous integration across multiple well-spaced waypoints sampled along macroscopic trajectories. To accommodate technical measurement noise, integrated paths are constrained only to remain within a probabilistic vicinity of each waypoint rather than fitting exact snapshot coordinates. 2) **Splicing velocity derivative regularization:** FateODE regularizes the instantaneous derivatives of its Neural ODE vector field against high-confidence splicing velocities derived from quality-filtered genes, ensuring that continuous trajectory derivatives remain grounded in empirical splicing kinetic evidence. 3) **Bidirectional consistency across multi-waypoint integration:** To ensure dynamic stability and avoid overfitting to specific initializations, FateODE introduces a multi-round optimization scheme. By starting at an arbitrary waypoints in each turn to initiate forward and backward temporal rollouts, the framework enforces bidirectional consistency across integration passes. 4) **Manifold-constrained stochastic diffusion at inference:** To capture non-deterministic fate bifurcation without sacrificing manifold geometry, FateODE incorporates an inference-time diffusion term parameterized by a decoder-only variational architecture supervised by Maximum Mean Discrepancy (MMD). This formulation faithfully simulates localized Brownian motion directly along the transcriptomic manifold, accurately reproducing branching probabilities without drifting into unobserved regions.

2 Preliminary

2.1 Estimating RNA velocity via modeling the splicing dynamics with ODE

Pioneered by La Manno et al. (2018); Bergen et al. (2020), RNA velocity has become a foundational framework for analyzing single-cell differentiation dynamics. It infers differentiation dynamics by distinguishing unspliced (u) and spliced (s) transcripts via intronic reads. Their temporal evolution is modeled by a first-order kinetic system:

$$\frac{\mathrm{d}u(t)}{\mathrm{d}t} = \alpha(t) - \beta u(t), \quad \frac{\mathrm{d}s(t)}{\mathrm{d}t} = \beta u(t) - \gamma s(t), \tag{1}$$

where $\alpha(t)$, β, and γ denote transcription, splicing, and degradation rates, respectively. From static snapshots, kinetic frameworks estimate these gene-specific parameters alongside a cell-specific latent time t, mapping discrete observations onto continuous phase trajectories.

Despite advancing the study of cellular dynamics, RNA velocity models remain constrained by the fragile quantification of unspliced (u) and spliced (s) transcripts. Low sequencing depth amplifies technical dropout and Poisson noise, while end-biased short reads systematically misclassify unspliced pre-mRNAs lacking captured intronic junctions as mature transcripts. Consequently, robust velocity estimation is largely **confined to a restricted subset of highly expressed genes**.

2.2 Dynamic system modeling with NeuralODE

Neural Ordinary Differential Equations (Neural ODEs) have emerged as a powerful paradigm for modeling complex dynamical systems (Chen et al., 2019). In Neural ODE, a neural network parameterizes the velocity field $\mathbf{v}_\theta : \mathbb{R}^n \to \mathbb{R}^n$. For each initial state x_0, the associated ODE flow Φ_t maps x_0 to the state reached after an elapsed time t, satisfying

$$\mathrm{d}\Phi_\mathrm{t}(\mathrm{x}_0)/\mathrm{dt} = \mathbf{v}_\theta(\Phi_t(x_0)), \quad \Phi_0(x_0) = x_0.$$

For a fixed initial state x_0, the mapping $t \mapsto \Phi_t(x_0)$ defines a continuous-time trajectory, evaluated by an ODE solver. Instead of backpropagating through internal solver operations, training uses the

adjoint sensitivity method to compute gradients by solving an auxiliary reverse-time ODE, requiring only constant $\mathcal{O}(1)$ memory.

With real time-series observations, a Neural ODE is trained by integrating the vector field across an ordered sequence of observation timestamps $\{t_1, \ldots, t_M\}$ and minimizing the cumulative discrepancy against empirical measurements. In scRNA-seq, however, because sequencing is destructive, longitudinal measurements of individual physical cells are fundamentally unavailable, rendering standard trajectory-fitting paradigms inapplicable. Instead, existing frameworks rely on surrogate supervisory signals, which inevitably distort the dynamics vector field. Designing supervisory signals that simultaneously accommodate technical sequencing stochasticity, while enforcing strict geometric adherence to the transcriptomic manifold, is, therefore, the primary objective of this work.

3 Mapping Fate with Deterministic-and-Stochastic Nerual ODE

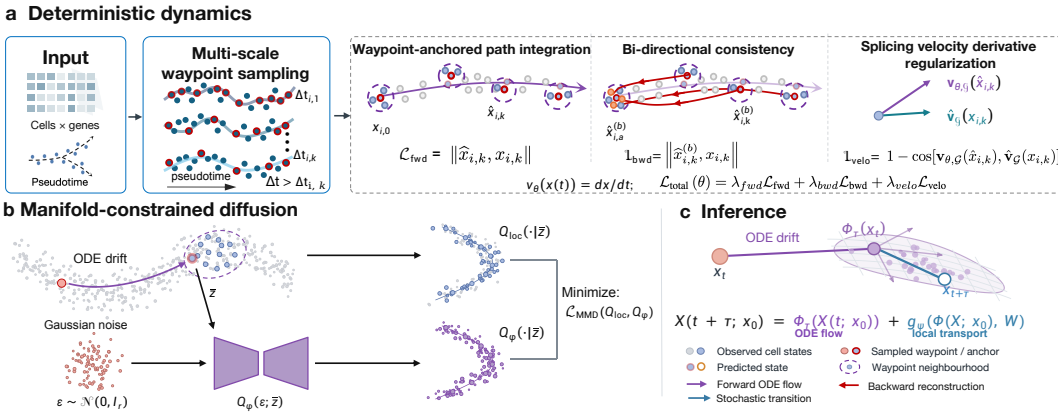

Figure 2: **FateODE combines deterministic dynamics and manifold-constrained stochastic diffusion.** (a)Waypoint-anchored static dynamics with bidirectional consistency and splicing velocity regularization. (b) Manifold-constrained stochastic diffusion in local neighborhood with variational decoder. (c) The combined drift and diffusion models bifurcating fate decisions at inference time.

FateODE models cellular dynamics as a deterministic ODE flow coupled to state-dependent stochastic variation. Over a time step τ, the evolution from an initial state x_0 is expressed as

$$X(t+\tau; x_0) = \Phi_\tau(X(t; x_0)) + g_\psi(\Phi(X; x_0), W), \qquad W \sim \mathcal{N}(0, 1),$$

Here, Φ_τ denotes the ODE flow induced by the velocity field $\mathbf{v}_\theta$, which describes the intrinsic directional tendencies of transcriptional change. The diffusion g_ψ aims to capture stochastic displacements of each cell constrained by the local manifold, characterized by the distribution of its neighboring cells. Cellular differentiation is thus formulated as directed but stochastic transition through the transcriptomic state space.

3.1 Learn Deterministic Dynamics by Integrating Trajectory and Velocity

3.1.1 Stabilizing Dynamical Learning through Multi-scale waypoint

In single-cell data, short temporal displacements are often dominated by measurement noise, while excessively long displacements may traverse unrelated biological branches and obscure geometric detail. We therefore supervise neural ODE trajectories using randomly sampled waypoint sequences spanning multiple temporal intervals. Longer intervals emphasize macroscopic developmental trends, while shorter intervals attend to local cell-state geometry.

At each training iteration, we sample an initial cell and an associated waypoint sequence $\mathcal{S}_i = \{(x_{i,k}, t_{i,k})\}_{k=0}^{K_i}$, where $t_{i,k}$ denotes inferred pseudotime. Adjacent time lags are sampled from a truncated distribution supported on $[\Delta t_{\min}, \Delta t_{\max}]$, with $\Delta t_{i,k} = t_{i,k} - t_{i,k-1}$. The lower bound reduces emphasis on potentially noise-dominated short transitions, while the upper bound limits gaps between consecutive supervision points.

Starting from $x_{i,0}$, the neural ODE predicts each subsequent waypoint state through integration:

$$\hat{x}_{i,k} = \Phi(X(t_{i,k}; x0)) = x_i + \int_{t_{i,0}}^{t_{i,k}} v_\theta(x(\tau))d\tau$$

where $\hat{x}$ denotes the predicted state. The forward loss is then defined as

$$\mathcal{L}_{\text{fwd}} = \mathbb{E}_{S_i}\left(\frac{1}{K_i}\sum_{k=1}^{K_i}\|\widehat{x}_{i,k}, x_{i,k}\|\right),$$

The observed waypoint states provide surrogate supervision for these integrated predictions, jointly constraining the shared velocity field across sampled starting states and temporal intervals and promotes the shared field to capture developmental changes without excessively tracking individual local fluctuations.

3.1.2 Enforcing transition consistency with bidirectional supervision

Noise in single-cell measurements introduces uncertainty into the observed states used to initialize ODE rollouts. For smooth dynamics, the flow Jacobian characterizes the first-order propagation and potential amplification of small initial-state deviations (Teschl, 2012). To reduce this sensitivity, we therefore complement forward reconstruction with backward rollouts from selected later observed waypoints, supervising the reconstruction of earlier observations. This bidirectional supervision encourages local consistency around sampled waypoints across integration origins and temporal intervals, while treating observations as soft reconstruction targets that accommodate nonzero fitting residuals without requiring exact interpolation of noisy measurements.

For each sampled waypoint sequence S_i, we select backward intervals $\mathcal{P}_i \subseteq \{(a,b) : 0 \le a < b \le K_i\}$. For each $(a,b) \in \mathcal{P}_i$, we integrate the shared velocity field v_θ backward from the observed state $x_{i,b}$ to reconstruct every waypoint in the interval:

$$\widehat{x}_{i,k}^{(b)} = \Phi_{t_{i,k}-t_{i,b}}(x_{i,b}), \qquad k = b, b-1, \ldots, a. \tag{2}$$

and define

$$\mathcal{L}_{\text{bwd}} = \mathbb{E}_{S_i,\mathcal{P}_i}\left(\frac{1}{|\mathcal{P}_i|}\sum_{(a,b)\in\mathcal{P}_i}\frac{1}{b-a}\sum_{k=a}^{b}\left\|\widehat{x}_{i,k}^{(b)}, x_{i,k}\right\|\right), \tag{3}$$

Backward reconstruction constrains the time-integrated velocity along trajectories initialized at multiple observed waypoints, thereby extending coverage of supervision across local waypoint neighborhoods.

3.1.3 Reliability-aware derivative supervision with splicing kinetics

To constrain velocity-field learning in high-dimensional gene-expression space, FateODE aligns predicted trajectory derivatives with splicing-derived reference velocities using high-confidence splicing-derived velocities, on a quality-filtered gene set $\mathcal{G}$. Gene selection uses a splicing quality score (SQS), detailed in Appendix A.1. For sampled waypoint sequences $\mathcal{S}_i$, we minimize the directional alignment loss

$$\mathcal{L}_{\text{velo}} = \mathbb{E}_{\mathcal{S}_i}\left[\frac{1}{K_i}\sum_{k=1}^{K_i}(1 - \cos[\mathbf{v}_{\theta,\mathcal{G}}(\hat{x}_{i,k}), \hat{\mathbf{v}}_{\mathcal{G}}(x_{i,k})])\right]. \tag{4}$$

Here, $\hat{x}_{i,k}$ is the predicted state corresponding to observed waypoint $x_{i,k}$, and $\hat{\mathbf{v}}_{\mathcal{G}}$ denotes the reference velocity restricted to $\mathcal{G}$.

We integrate the neural ODE with an adaptive Runge–Kutta solver and train it by stochastic gradient descent on

$$\mathcal{L}_{\text{total}}(\theta) = \lambda_{\text{fwd}}\mathcal{L}_{\text{fwd}} + \lambda_{\text{bwd}}\mathcal{L}_{\text{bwd}} + \lambda_{\text{velo}}\mathcal{L}_{\text{velo}}. \tag{5}$$

The weights λ_{fwd}, λ_{bwd}, and λ_{velo} balance forward reconstruction, backward reconstruction, and velocity alignment, respectively. Together, these objectives constrain the deterministic neural ODE through both integrated state transitions and splicing-informed derivative directions.

Table 1: Benchmarking terminal-fate classification and fate-distribution prediction in HSCs.

Method	Precision ↑	Recall ↑	Specificity ↑	F1 ↑	AUPRC ↑	Fate cosine ↑	Brier ↓	JS similarity ↑
FateODE	0.202	**0.275**	**0.768**	0.224	**0.507**	**0.676**	**0.401**	**0.714**
DynNet	0.162	0.250	0.750	0.196	0.462	0.559	0.807	0.495
scDiffEq	0.165	0.214	0.735	0.186	0.457	0.588	0.589	0.578
CellMNN	**0.250**	0.245	0.752	**0.227**	0.464	0.411	0.984	0.389
scTour	0.162	0.250	0.750	0.196	0.461	0.559	0.808	0.494
scKINETICS	0.244	0.261	0.746	0.221	0.458	0.441	0.962	0.409

3.2 Model Stochastic Differentiation with Manifold-constrained diffusion

Building on the learned deterministic velocity field, FateODE models stochastic differentiation by learning a transport map from a Gaussian noise distribution to a local distribution of cell states. Neighbours of the drift-evolved state provide an empirical reference for plausible local variation, guiding stochastic transitions according to local cell-state geometry. This neighbourhood-based construction provides a soft manifold constraint, encouraging stochastic updates to remain compatible with the transcriptomic geometry supported by observed cells.

For each drift-evolved state $\bar{z}$, we identify neighbours around its latent feature. Neighbour sampling and interpolation define a local reference distribution $Q_{\mathrm{loc}}(\cdot \mid \bar{z})$. The neural transport map generates cell-state samples $Q_\phi(\epsilon; \bar{z})$ from Gaussian noise $\epsilon \sim \mathcal{N}(0, I_r)$. We align the generated distribution $Q_\phi(\cdot \mid \bar{z})$ with $Q_{\mathrm{loc}}(\cdot \mid \bar{z})$ by minimizing the squared maximum mean discrepancy (MMD) with a radial basis function (RBF) kernel (Appendix A.2). Matching the local distribution encourages generated samples to reflect the shape and extent of variation supported by neighbouring cells. This incorporates local manifold geometry into stochastic state generation.

To connect latent geometry with high-dimensional expression states, we learn latent features using a variational autoencoder (VAE). Reconstruction encourages the latent representation to retain high-dimensional gene-expression information. Neighbourhood consistency further encourages consistent local cell connectivity across the original and latent spaces. The complete VAE objective, including standard KL regularization, is

$$\mathcal{L}_{\mathrm{VAE}} = \mathcal{L}_{\mathrm{reconstruction}} + \alpha\, \mathcal{L}_{\mathrm{reg}} + \beta \mathcal{L}_{\mathrm{neighbor}}.$$

Further details are provided in Appendix A.3.

Overall, MMD shapes stochastic variation around each drift-evolved state while reconstruction and neighbourhood consistency link the geometry used for neighbour selection to gene-expression states.

At inference, deterministic drift and stochastic transport generate alternative developmental trajectories. Applying local geometric constraint successively encourages sampled paths to remain compatible with the transcriptomic manifold while allowing varied developmental routes and fate outcomes.

4 Results

4.1 Benchmarking fate prediction and developmental dynamics

We compared each evaluable method's predicted terminal-fate proportions with barcode-derived distributions across four hematopoietic lineage (HSC): erythroid (Ery), megakaryocytic (Meg), monocytic (Mon) and neutrophilic (Neu) Li et al. (2023). FateODE achieves the best fate cosine similarity, JS similarity and the Brier score (Table 1), with substantial improvements over scDiffEq, the next-best method. FateODE's predicted terminal-fate proportions showed the closest agreement with the barcode-derived reference compositions among the evaluated methods.

We assigned each HSC the terminal fate with the largest predicted proportion. FateODE achieved the highest AUPRC, recall and specificity, and the second-highest F1 score. Despite modest differences in classification performance, fate compositions differed substantially (Fig. 3b). Only

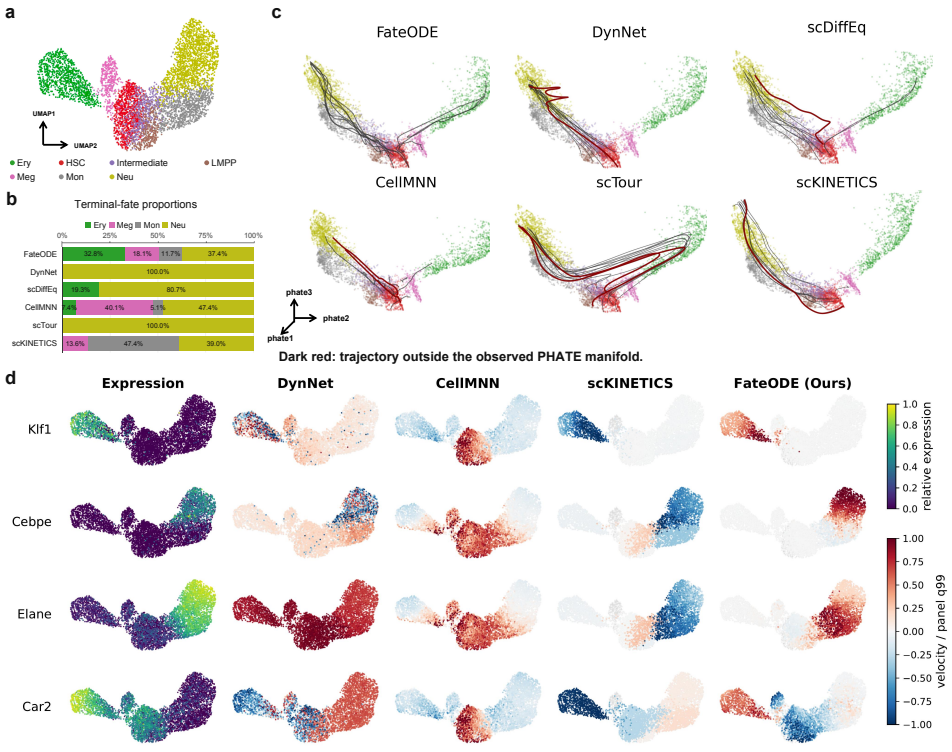

Figure 3: Benchmarking hematopoietic fates and gene-level dynamics. (a) UMAP of hematopoietic cells coloured by cell type. (b) Terminal-fate proportions across methods. (c) HSC-initialized trajectories in three-dimensional PHATE space; thick dark-red paths indicate departures from the observed manifold.(d) Relative expression and predicted velocities of Klf1, Cebpe, Elane and Car2.

FateODE and CellMNN assigned HSCs to four terminal populations, although CellMNN's assignments favoured Meg and Neu. DynNet, scTour and scDiffEq failed to distinguish Mon from Neu. We next examined whether inferred trajectories adhered to the developmental manifold. For each method, trajectories were initialized from all HSCs, with paths departing from the PHATE manifold highlighted in dark red (Fig. 3c). FateODE trajectories largely followed the observed branches, supporting the consistency of its inferred dynamics with the empirical manifold. By comparison, CellMNN trajectories took shortcuts through sparsely populated regions, bypassing the curvature of the manifold. scTour produced trajectories that crossed between the erythroid and myeloid branches, connecting otherwise distinct lineages. DynNet and scDiffEq showed excursions away from the observed manifold, with sharp turns along intermediate portions of their trajectories. These patterns illustrate how matching population-level observations at discrete time points can leave intermediate paths insufficiently constrained. scKINETICS also produced shortcuts near the HSC-to-LMPP transition, potentially reflecting limited flexibility of its linear regulatory model in this setting.

Replacing DynNet's drifting component with that of FateODE improves fate diversity, with trajectories no longer collapsing toward a single lineage (Fig. 4), supported by the increased Fate cosine. Nevertheless, since DynNet's diffusion process lacks an explicit manifold constraint, some trajectories still make large jumps across lineages, resulting in biologically implausible transitions.

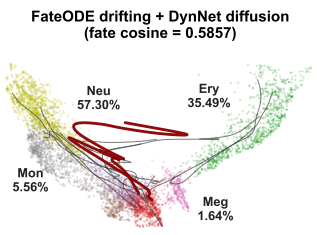

Figure 4: Ablation study of the drift component in FateODE.

FateODE estimates gene-resolved velocities directly in high-dimensional gene-expression space. Predictions for Car2, Klf1, Elane and Cebpe matched lineage-associated expression patterns, whereas DynNet, CellMNN and scKINETICS yielded opposing or spatially discordant patterns (Fig. 3d). Car2 velocities were posi-

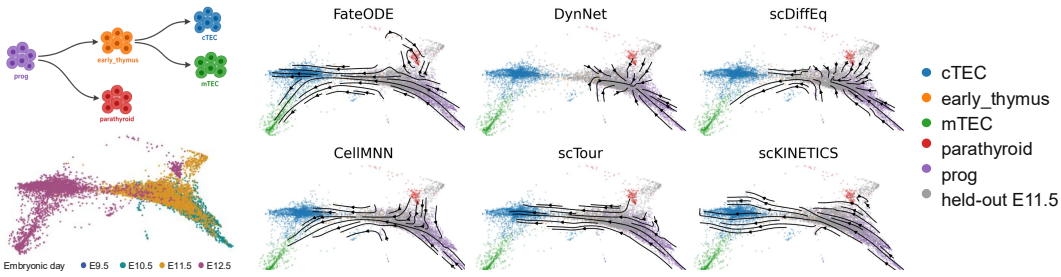

Figure 5: Recovery of held-out developmental states and branching trajectories in mTEC/BC.

tive along the erythroid branch and strongly negative in undifferentiated cells, with negative velocities weakening towards LMPP. Klf1 and Elane velocities rose sharply during cell-state transitions and appeared to precede pronounced expression patterns, suggesting rapid gene activation during fate commitment.

4.2 Recovering held-out cell states and developmental branches

To assess trajectory reconstruction and recover unobserved cell with a missing time point, we excluded E11.5 cells from the pharyngeal endoderm dataset (mtec/BC) Magaletta et al. (2022) and integrated forward from earlier progenitors (Fig. 5b). FateODE maintained trajectories through observed and held-out regions while preserving all developmental branches. By contrast, scTour and scKINETICS predominantly directed flow towards cTEC, whereas CellMNN showed limited recovery of the parathyroid and cTEC branches. DynNet and scDiffEq showed local convergence near progenitor and held-out regions, with limited continuation along differentiated branches.

To quantify recovery of the withheld E11.5 population, we evaluated distributional agreement, coverage across branches and preservation of branch proportions. FateODE achieved the lowest energy distance and branch-proportion Jensen-Shannon divergence among the evaluated methods (Table 2). It also attained the highest coverage of different branches. Although scTour achieved the second-lowest energy distance to the observed E11.5 distribution, it substantially underrepresented the less abundant parathyroid branch. These results support FateODE's ability to reconstruct branching developmental dynamics and recover intermediate cell-state distributions absent from training, with improved branch coverage and global distributional agreement.

Table 2: Recovery of the held-out E11.5 population and branch structure in mTEC/BC. Energy distance measures the overall discrepancy between predicted and true cell distributions. Mean cov. and Min cov. measure the mean and minimum coverage across branches respectively. Branch JS measures the discrepancy between predicted and true branch proportions.

Method	Energy distance ↓	Mean cov. ↑	Min cov. ↑	Branch JS ↓
DynNet	0.829	0.385	0.229	0.059
scDiffEq	0.542	0.506	0.307	0.054
scTour	0.340	0.429	0.132	0.064
CellMNN	0.371	0.457	0.254	0.070
scKINETICS	1.033	0.320	0.193	0.106
FateODE	**0.281**	**0.582**	**0.365**	**0.015**

4.3 Mapping fate territories and guiding in-silico perturbation

In the Darlin dataset, we estimated the probability distribution over terminal lineages for each progenitor by forward integration. Mapping these probabilities across HSCs revealed graded fate territories within the HSC compartment. These territories exhibited opposing gradients of neutrophil/monocyte and megakaryocytic/erythroid fate probabilities (Fig. 6a). This pattern indicated a continuum of inferred lineage biases within the undifferentiated compartment, reflecting structured heterogeneity in predicted developmental potential.

To identify molecular features potentially involved in HSC fate decisions, we computed Spearman correlations between gene expression and inferred lineage-specific fate probabilities across HSCs.

This analysis nominated *Nfe2*, *Gata2*, *Nfia* and *Sox6* as candidate fate regulators (Fig. 6b), consistent with their established hematopoietic functions. Gata2 is required for adult HSC maintenance, whereas Nfia promotes erythroid over granulocytic differentiation, Nfe2 and Sox6 regulate downstream differentiation programmes, supporting megakaryocyte maturation and platelet production, and erythroid cell survival and terminal maturation, respectively Shivdasani et al. (1995); Menendez-Gonzalez et al. (2019); Starnes et al. (2009); Dumitriu et al. (2006). The associations provided molecular context for the inferred territories and a basis for testing whether these factors influence HSC lineage fate decisions.

Finally, we used the reconstructed transcriptomic vector field to simulate the effects of jointly upregulating these four factors on HSC differentiation. Forward integration under this perturbation shifted predicted lineage outcomes towards the megakaryocytic and erythroid branches (Fig. 6c). This shift was reflected in a displacement of the fate boundary, with the MkP/Ery-biased territory expanding into regions previously biased towards Neu/Mon outcomes. The proportion of simulated terminal outcomes represented an increase of 11% in combined megakaryocytic and erythroid output, accompanied by reduced neutrophil and monocyte, which was qualitatively consistent with experimental studies Starnes et al. (2009). These results illustrate how the reconstructed transcriptomic vector field can predict future cell states and simulate how genetic perturbations reshape fate territories.

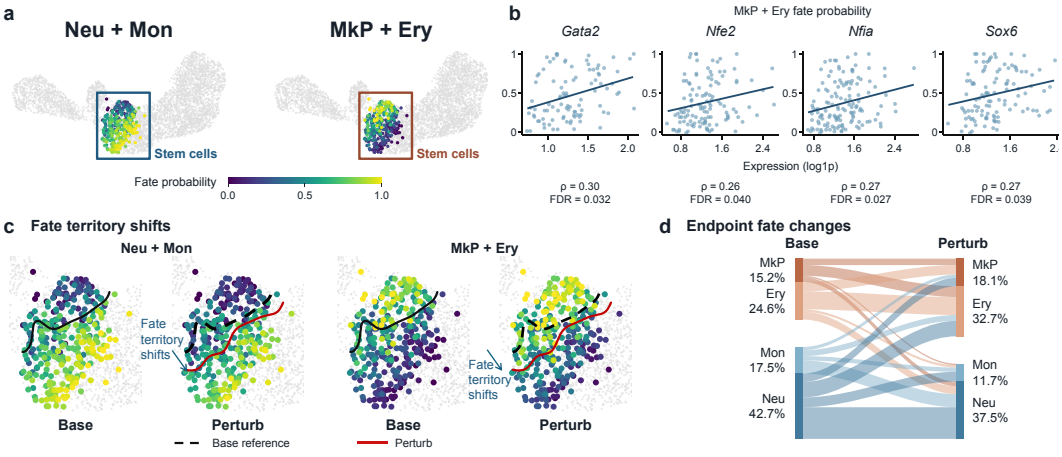

Figure 6: Hematopoietic fate territories and their shifts under in silico perturbation.(a) Lineage-biased fate territories in HSCs. (b) Associations between expression of our selected regulators for Ery fate probability.(c) Fate territories before (Base) and after (Perturb) simulated joint upregulation of these four genes. Black (dashed and solid) and red curves mark original and perturbed boundaries, respectively.(d) Terminal-fate redistribution between base and perturb.

5 Discussion

FateODE combines trajectory supervision, splicing-informed velocities and neighborhood-guided stochastic transitions to reconstruct differentiation dynamics. Strong accordance with lineage-barcode-derived HSC fate distributions and recovery of held-out E11.5 cells demonstrate its ability to capture lineage diversity and intermediate states. Graded HSC fate territories linked transcriptional heterogeneity to predicted lineage biases, while its in silico perturbations further support the exploration of regulatory influences on lineage outcomes and the prioritization of candidate regulators. Despite these strengths, FateODE currently focuses on modeling phenotypic differentiation trajectories without incorporating prior gene regulatory network (GRN) knowledge. Nevertheless, its modular dynamical formulation is naturally amenable to integrating mechanistic GRN priors, which would enable joint inference of cell state transitions and causal gene regulation, a promising direction for future work. Overall, FateODE represents an important step forward in transcriptomic modeling, recapitulating lineage diversity and identifying fate-determining driver genes. It also serves as an in-silico "world model" to predict terminal fate redistribution under targeted genetic perturbations.

AI USE STATEMENT

The research idea, problem formulation, and overall methodological direction of this work were independently developed by the authors. Generative AI tools were not used to identify the research problem, originate the main research idea, or formulate the overall method.

Generative AI tools were used to assist with the implementation and refinement of specific technical components. In particular, they were used to provide suggestions on model architecture and loss design, to assist with the selection of evaluation metrics and the interpretation of experimental results, and to generate software implementations of the proposed method, including data augmentation procedures used for VAE training.

Generative AI tools were also used to assist with drafting and editing the manuscript and with the creation or modification of scientific figures and images.

We did not use generative AI tools to generate synthetic datasets, develop theoretical models or conceptual frameworks, formulate mathematical claims or proofs, propose scientific hypotheses, perform dataset cleaning or reformatting, conduct qualitative or thematic data analysis, or assist with translation.

The authors evaluated AI-assisted technical suggestions before incorporating them into the final implementation. AI-generated code was tested and validated by the authors through the experiments reported in this paper. AI-assisted interpretations were checked against the experimental results, and AI-assisted text and figures were reviewed and revised by the authors. The authors take full responsibility for the methodology, implementation, experimental analysis, text, claims, and other artifacts in the final work.

REFERENCES

Kathleen Abadie, Elisa C. Clark, Rajesh M. Valanparambil, Obinna Ukogu, Wei Yang, Riza M. Daza, Kenneth K. H. Ng, Jumana Fathima, Allan L. Wang, Judong Lee, Tahseen H. Nasti, Avinash Bhandoola, Armita Nourmohammad, Rafi Ahmed, Jay Shendure, Junyue Cao, and Hao Yuan Kueh. Reversible, tunable epigenetic silencing of TCF1 generates flexibility in the T cell memory decision. *Immunity*, 57(2):271–286.e13, 2024. doi: 10.1016/j.immuni.2023.12.006. URL https://doi.org/10.1016/j.immuni.2023.12.006.

Volker Bergen, Marius Lange, Stefan Peidli, F. Alexander Wolf, and Fabian J. Theis. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature Biotechnology*, 2020. doi: 10.1038/s41587-020-0591-3. URL https://doi.org/10.1038/s41587-020-0591-3.

Idris Kouadri Boudjelthia, Salvatore Milite, Yuanhua Huang, Andrea Sottoriva, and Guido Sanguinetti. NeuroVelo: interpretable learning of temporal cellular dynamics from single-cell data.

Cassandra Burdziak, Chujun Julia Zhao, Doron Haviv, Direna Alonso-Curbelo, Scott W. Lowe, and Dana Pe'er. scKINETICS: inference of regulatory velocity with single-cell transcriptomics data. *Bioinformatics*, 2023. doi: 10.1093/bioinformatics/btad267. URL https://doi.org/10.1093/bioinformatics/btad267.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural Ordinary Differential Equations, December 2019. URL http://arxiv.org/abs/1806.07366. arXiv:1806.07366 [cs.LG].

Jingyu Dou, Wentao Lyu, Feng Chen, Qing Nie, and Chunhe Li. Inferring stochastic dynamics by biophysical Neural ODE using single-cell transcriptomics. *Nature Communications*, 2026. doi: 10.1038/s41467-026-73257-z. URL https://doi.org/10.1038/s41467-026-73257-z.

B. Dumitriu, M. R. Patrick, J. P. Petschek, S. Cherukuri, U. Klingmuller, P. L. Fox, and V. Lefebvre. Sox6 cell-autonomously stimulates erythroid cell survival, proliferation, and terminal maturation and is thereby an important enhancer of definitive erythropoiesis during mouse development. *Blood*, 108(4):1198–1207, 2006. doi: 10.1182/blood-2006-02-004184.

Spencer Farrell, Madhav Mani, and Sidhartha Goyal. Inferring single-cell transcriptomic dynamics with structured latent gene expression dynamics. *Cell Reports Methods*, 3(9):100581, September 2023. ISSN 2667-2375. doi: 10.1016/j.crmeth.2023.100581. URL https://www.sciencedirect.com/science/article/pii/S2667237523002254.

Mingze Gao, Chen Qiao, and Yuanhua Huang. UniTVelo: temporally unified RNA velocity reinforces single-cell trajectory inference. *Nature Communications*, 13(1):6586, November 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-34188-7. URL https://www.nature.com/articles/s41467-022-34188-7.

Yichen Gu, Yuxuan Song, David Blaauw, and Joshua D. Welch. Bayesian inference of RNA velocity incorporating timepoints, lineage bifurcations, and count data. *PLOS Computational Biology*, 22(3):e1014060, March 2026. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1014060. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1014060.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations*, 2014. URL https://arxiv.org/abs/1312.6114.

Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E. Kastriti, Peter Lönnerberg, Alessandro Furlan, Jean Fan, Lars E. Borm, Zehua Liu, David van Bruggen, Jimin Guo, Xiaoling He, Roger Barker, Erik Sundström, Gonçalo Castelo-Branco, Patrick Cramer, Igor Adameyko, Sten Linnarsson, and Peter V. Kharchenko. RNA velocity of single cells. *Nature*, 560(7719):494–498, August 2018. ISSN 1476-4687. doi: 10.1038/s41586-018-0414-6. URL https://www.nature.com/articles/s41586-018-0414-6. Number: 7719.

Li Li, Sarah Bowling, Sean E. McGeary, Qi Yu, Bianca Lemke, Karel Alcedo, Yuemeng Jia, Xugeng Liu, Mark Ferreira, Allon M. Klein, Shou-Wen Wang, and Fernando D. Camargo. A mouse model with high clonal barcode diversity for joint lineage, transcriptomic, and epigenomic profiling in single cells. *Cell*, 186(23):5183–5199.e22, November 2023. ISSN 0092-8674, 1097-4172. doi: 10.1016/j.cell.2023.09.019. URL https://www.cell.com/cell/abstract/S0092-8674(23)01040-1.

Qian Li. scTour: a deep learning architecture for robust inference and accurate prediction of cellular dynamics. *Genome Biology*, 2023. doi: 10.1186/s13059-023-02988-9. URL https://doi.org/10.1186/s13059-023-02988-9.

Shengyu Li, Pengzhi Zhang, Weiqing Chen, Lingqun Ye, Kristopher W. Brannan, Nhat-Tu Le, Jun-ichi Abe, John P. Cooke, and Guangyu Wang. A relay velocity model infers cell-dependent RNA velocity. *Nature Biotechnology*, 42(1):99–108, January 2024. ISSN 1546-1696. doi: 10.1038/s41587-023-01728-5. URL https://www.nature.com/articles/s41587-023-01728-5.

Margaret E. Magaletta, Macrina Lobo, Eric M. Kernfeld, Hananeh Aliee, Jack D. Huey, Teagan J. Parsons, Fabian J. Theis, and René Maehr. Integration of single-cell transcriptomes and chromatin landscapes reveals regulatory programs driving pharyngeal organ development. *Nature Communications*, 13(1):457, January 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-28067-4. URL https://www.nature.com/articles/s41467-022-28067-4.

J. B. Menendez-Gonzalez, M. Vukovic, A. Abdelfattah, L. Saleh, A. Almotiri, L. A. Thomas, A. Agirre-Lizaso, A. Azevedo, A. C. Menezes, G. Tornillo, S. Edkins, K. Kong, P. Giles, F. Anjos-Afonso, A. Tonks, A. S. Boyd, K. R. Kranc, and N. P. Rodrigues. Gata2 as a Crucial Regulator of Stem Cells in Adult Hematopoiesis and Acute Myeloid Leukemia. *Stem Cell Reports*, 13(2):291–306, 2019. doi: 10.1016/j.stemcr.2019.07.005.

Xiaojie Qiu, Qi Mao, Ying Tang, Li Wang, Raghav Chawla, Hannah A Pliner, and Cole Trapnell. Reversed graph embedding resolves complex single-cell trajectories. *Nature Methods*, 14(10):979–982, October 2017. ISSN 1548-7091, 1548-7105. doi: 10.1038/nmeth.4402. URL https://www.nature.com/articles/nmeth.4402.

Xiaojie Qiu, Yan Zhang, Jorge D. Martin-Rufino, Chen Weng, Shayan Hosseinzadeh, Dian Yang, Angela N. Pogson, Marco Y. Hein, Kyung Hoi Joseph Min, Li Wang, Emanuelle I. Grody, Matthew J. Shurtleff, Ruoshi Yuan, Song Xu, Yian Ma, Joseph M. Replogle, Eric S. Lander, Spyros Darmanis, Ivet Bahar, Vijay G. Sankaran, Jianhua Xing, and Jonathan S. Weissman. Mapping transcriptomic vector fields of single cells. *Cell*, 2022. doi: 10.1016/j.cell.2021.12.045. URL https://doi.org/10.1016/j.cell.2021.12.045.

R. A. Shivdasani, M. F. Rosenblatt, D. Zucker-Franklin, C. W. Jackson, P. Hunt, C. J. Saris, and S. H. Orkin. Transcription factor NF-E2 is required for platelet formation independent of the actions of thrombopoietin/MGDF in megakaryocyte development. *Cell*, 81(5):695–704, 1995. doi: 10.1016/0092-8674(95)90531-6.

L. M. Starnes, A. Sorrentino, E. Pelosi, M. Ballarino, O. Morsilli, M. Biffoni, S. Santoro, N. Felli, G. Castelli, M. L. De Marchis, G. Mastroberardino, M. Gabbianelli, A. Fatica, I. Bozzoni, C. Nervi, and C. Peschle. NFI-A directs the fate of hematopoietic progenitors to the erythroid or granulocytic lineage and controls beta-globin and G-CSF receptor expression. *Blood*, 114(9): 1753–1763, 2009. doi: 10.1182/blood-2008-12-196196.

Kelly Street, Davide Risso, Russell B. Fletcher, Diya Das, John Ngai, Nir Yosef, Elizabeth Purdom, and Sandrine Dudoit. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*, 19(1):477, December 2018. ISSN 1471-2164. doi: 10.1186/s12864-018-4772-0. URL https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-018-4772-0.

Patrick S. Stumpf, Rosanna C. G. Smith, Michael Lenz, Andreas Schuppert, Franz-Josef Müller, Ann Babtie, Thalia E. Chan, Michael P. H. Stumpf, Colin P. Please, Sam D. Howison, Fumio Arai, and Ben D. MacArthur. Stem cell differentiation as a Non-Markov stochastic process. *Cell Systems*, 5(3):268–282.e7, 2017. doi: 10.1016/j.cels.2017.08.009. URL https://doi.org/10.1016/j.cels.2017.08.009.

Gerald Teschl. *Ordinary Differential Equations and Dynamical Systems*, volume 140 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, Rhode Island, 2012. ISBN 978-0-8218-8328-0. URL https://www.mat.univie.ac.at/~gerald/ftp/book-ode/.

Cole Trapnell, Davide Cacchiarelli, Jonna Grimsby, Prapti Pokharel, Shuqiang Li, Michael Morse, Niall J Lennon, Kenneth J Livak, Tarjei S Mikkelsen, and John L Rinn. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*, 32(4):381–386, April 2014. ISSN 1087-0156, 1546-1696. doi: 10.1038/nbt.2859. URL https://www.nature.com/articles/nbt.2859.

Michael E. Vinyard, Anders W. Rasmussen, Ruitong Li, Allon M. Klein, Gad Getz, and Luca Pinello. Learning cell dynamics with neural differential equations. *Nature Machine Intelligence*, 2025. doi: 10.1038/s42256-025-01150-3. URL https://doi.org/10.1038/s42256-025-01150-3.

Jan-Philipp von Bassewitz, Adeel Pervez, Marco Fumero, Matthew Robinson, Theofanis Karaletsos, and Francesco Locatello. Learning Explicit Single-Cell Dynamics Using ODE Representations. In *Proceedings of the Fourteenth International Conference on Learning Representations*, 2026. URL https://arxiv.org/abs/2510.02903.

Philipp Weiler, Marius Lange, Michal Klein, Dana Pe'er, and Fabian Theis. CellRank 2: unified fate mapping in multiview single-cell data. *Nature Methods*, 2024. doi: 10.1038/s41592-024-02303-9. URL https://doi.org/10.1038/s41592-024-02303-9.

Caleb Weinreb, Alejo Rodriguez-Fraticelli, Fernando D. Camargo, and Allon M. Klein. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science*, 2020. doi: 10.1126/science.aaw3381. URL https://pubmed.ncbi.nlm.nih.gov/31974159/.