TRACE: TRANSCRIPTIONAL REGULATOR MODULATED REFERENCE GEOMETRY FOR TEMPORAL SINGLE-CELL FLOW MATCHING

Anonymous authors
Paper under double-blind review

ABSTRACT

Progenitor-exhausted (Tpex) CD8$^+$ T cells retain self-renewal and proliferative capacity, making their differentiation therapeutically important. Resolving this progression from temporal single-cell RNA sequencing is difficult because destructive sampling provides population snapshots without directly observed cell trajectories. Here, we introduce **TRACE**, a regulatory-informed conditional flow-matching model for predicting unobserved temporal populations. To incorporate regulatory direction into transport, we map temporal derivatives of expression-derived transcription-factor regulon activities into latent-space tangents for shape-preserving cubic-Hermite paths between optimal-transport-coupled populations. Across three T-cell exhaustion studies, the clearest improvements occurred at earlier holdouts. On the Gupta day 7 development holdout, W_2 was 21.7% lower than with linear CFM and 8.8% lower than with cubic CFM. Improvements extended to the earlier Giles and DSM holdouts, including a reduction in DSM day 8 W_2 from 11.175 ± 0.111 with linear CFM to 10.725 ± 0.140 with TRACE. At later holdouts, linear or alternative transport often matched or exceeded TRACE performance. These results support stage-dependent benefits of regulatory-informed reference paths for temporal interpolation, with the largest gains during earlier exhaustion-associated transitions.

1 INTRODUCTION

T-cell exhaustion limits durable immune responses during chronic infection and cancer (Im et al., 2016; Miller et al., 2019). Within this continuum, progenitor-exhausted CD8$^+$ T cells (Tpex) retain self-renewal and proliferative capacity and replenish more differentiated exhausted populations (Utzschneider et al., 2020; Beltra et al., 2020). Their expansion after PD-1 pathway blockade contributes to therapeutic responses (Im et al., 2016; Miller et al., 2019; Siddiqui et al., 2019). Progression toward exhaustion involves coordinated transcriptional changes and the establishment of persistent epigenetic states (Chen et al., 2019; Khan et al., 2019; Pauken et al., 2016).

Following this progression with single-cell RNA sequencing requires inferring trajectories from destructive population snapshots (Schiebinger et al., 2019). In chronic lymphocytic choriomeningitis virus (LCMV) infection, for example, transcriptionally distinct progenitor, effector-like, and exhausted populations occur at successive sampling times (Giles et al., 2022; Daniel et al., 2022; Schleicher et al., 2025). We evaluated reconstruction of intermediate populations excluded from fitting to determine how geometric and regulatory information affected prediction across developmental stages.

In conditional flow matching (CFM), a continuous velocity field is trained against velocities along reference paths between sampled states (Lipman et al., 2023; Tong et al., 2024). Multiple intermediate paths can connect the same observed temporal marginals, making reference geometry an explicit modeling choice (Albergo et al., 2024; Kapuśniak et al., 2024). Nonlinear paths allow changes in direction between sampled states, creating a corresponding need to specify local tangents.

Here, we introduce **TRACE** (Transcriptional-Regulator-Modulated Reference Geometry), a regulatory-informed framework for temporal single-cell flow matching. To connect reference-path tangents to regulatory state, we estimate transcription-factor (TF) activity from signed target-gene expression (Alvarez et al., 2016; Badia-i Mompel et al., 2022; Holland et al., 2020) and map tem-

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.

poral changes in these activities into latent directions that modulate the tangents. Using the same optimal-transport couplings, neural-field architecture, and optimization procedure, we compare linear, nonlinear, and regulatory-informed references. Across three T-cell exhaustion studies, we observed the clearest improvements at earlier holdouts, including reductions in W_2 of 21.7% relative to linear CFM and 8.8% relative to cubic CFM on the Gupta day 7 development holdout. These results support stage-dependent benefits of regulatory-informed reference geometry for temporal population prediction.

Our contributions are threefold.

1. **Regulon-derived supervision for temporal flow matching.** We introduce a reference-path construction that uses temporal changes in TF-regulon activity to steer shape-preserving cubic-Hermite tangents in latent space. By retaining the same OT-linked cell states, we can assess the predictive contribution of regulatory directions beyond geometric interpolation.
2. **Controlled evaluation of geometry and regulatory steering.** We reconstruct temporal populations excluded from all fitted preprocessing, coupling, and model training. With matched architectures and training procedures, we compare interpolation families and regulon-permutation controls to assess how reference geometry and regulatory structure affect prediction.
3. **Lower mean error at early holdouts across three studies.** After selecting the steering coefficient on Gupta day 7, we fixed it for all subsequent holdouts in the Gupta, Giles, and DSM studies (Schleicher et al., 2025; Giles et al., 2022; Fagerberg et al., 2025). Mean W_2 was lower with TRACE than with linear and cubic CFM at the earlier holdout in every study. At later holdouts, simpler transport often matched or exceeded TRACE performance, supporting stage-dependent benefits of regulatory steering.

2 Related Work

RNA velocity and trajectory inference. RNA velocity is an estimate of local transcriptional change derived from spliced and unspliced RNA under kinetic assumptions (La Manno et al., 2018; Bergen et al., 2020; Li et al., 2024). Velocity-based analyses have been used to study directed transitions during T-cell differentiation (Daniel et al., 2022; Giles et al., 2022). With CellRank and CellRank 2, terminal states and fate probabilities can be estimated from transition information derived from several data modalities, including temporal measurements (Lange et al., 2022; Weiler et al., 2024). Pseudotime and graph-based analyses further characterize lineage structure and cell-state progression (Trapnell et al., 2014; Street et al., 2018; Wolf et al., 2019; Setty et al., 2019).

Population dynamics and continuous transport. Optimal transport has been used to couple cells across sampled populations in Waddington-OT and moscot (Schiebinger et al., 2019; Klein et al., 2025). Continuous population dynamics have also been inferred through dynamic optimal transport in TrajectoryNet, manifold-aware neural ODEs in MIOFlow, and stochastic potential dynamics in PRESCIENT (Tong et al., 2020; Huguet et al., 2022; Yeo et al., 2021). Extensions include joint estimation of cell-state changes and population growth (Sha et al., 2024; Wang et al., 2025), as well as flow matching across multiple sampling times and experimental conditions (Rohbeck et al., 2025; Lee et al., 2025; Kviman et al., 2026).

Reference geometry and directional supervision. In conditional flow matching, velocity supervision is defined along specified reference paths (Lipman et al., 2023; Tong et al., 2024). Nonlinear constructions include approximate geodesics under data-dependent metrics and stochastic interpolants connecting multiple marginals (Kapuśniak et al., 2024; Albergo et al., 2024). Spline-based reference trajectories have also been investigated in SplineFlow and FlowGRN+ (Rathod et al., 2026; Tong & Pang, 2026). Beyond path geometry, inferred velocities have been incorporated into a reference process with nonzero drift in Curly Flow Matching, and reference-dynamics parameters have been estimated through iterative refinement (Petrović et al., 2025; Shen et al., 2025). Within this line of work, we use temporal regulon derivatives to modulate shape-preserving cubic-Hermite tangents between OT-linked cell states.

Regulatory information in cell-state modeling. Cell-level TF activity can be inferred from downstream target-gene expression using regulon-based methods (Alvarez et al., 2016; Badia-i Mompel et al., 2022; Holland et al., 2020). Regulatory networks have been incorporated into cell-dynamics estimation in RegVelo and in silico perturbation analysis in CellOracle (Wang et al., 2026; Kamimoto et al., 2023). Conversely, inferred trajectories and gene velocities have been used to estimate regulatory networks in GRIT, OTVelo, and FlowGRN (Lamoline et al., 2025; Zhao et al., 2025; Tong & Pang, 2025). For TRACE, we use a signed TF-target prior to derive regulatory features for reference-path supervision and assess their contribution to temporal population prediction.

Generative models for single-cell data. Conditional generation of single-cell expression profiles has been studied using diffusion in scDiffusion and flow matching with discrete likelihood decoders in CFGen (Luo et al., 2024; Palma et al., 2025). These settings include generation under specified cellular attributes and across measurement modalities. Our evaluation concerns interpolation between experimentally sampled temporal populations, with each target population excluded from fitted preprocessing, reference construction, and flow-field training.

3 Problem Setup

We seek to reconstruct an unobserved CD8$^+$ T-cell population from surrounding measured populations and determine when nonlinear reference geometry and regulon-derived directions improve temporal prediction. Evaluation concerns population distributions and does not establish individual lineages, developmental irreversibility, or causal TF effects.

Let $d_1 < \cdots < d_K$ denote the retained sampling days after an interior day $d_\star$ has been held out. We normalize biological time as $\tau(d) = (d - d_1)/(d_K - d_1)$ and write $\tau_k = \tau(d_k)$ and $\tau_\star = \tau(d_\star)$. For example, the Gupta/Schleicher study contains modeled sampling times $d \in \{2.5, 7, 14, 21\}$, with each interior time held out separately (Schleicher et al., 2025).

Let μ_{τ_k} denote the empirical distribution of retained cell states at normalized time τ_k. Each state $z \in \mathbb{R}^p$ contains PCA coordinates standardized using retained-cell means and standard deviations. We fit both the PCA basis and this latent scaling on retained cells. Predictions are returned to the corresponding unstandardized PCA coordinates for evaluation. We learn a velocity field $v_\theta : \mathbb{R}^p \times [0,1] \to \mathbb{R}^p$, parameterized by θ, that specifies the rate of change of the latent state. Starting from cells sampled at a retained source time s, the dynamics satisfy

$$\frac{\mathrm{d}z(\tau)}{\mathrm{d}\tau} = v_\theta\big(z(\tau), \tau\big), \qquad z(s) \sim \mu_s. \tag{1}$$

Thus, the latent state changes with normalized time according to the learned velocity field.

Let $\Phi^\theta_{s\to t}$ denote the solution map from source time s to target time t. The predicted population distribution is $\widehat{\mu}_t = (\Phi^\theta_{s\to t})_{\#}\mu_s$, where $\#$ denotes pushforward, or the distribution obtained by evolving source cells through the solution map. Our prediction target is the held-out distribution $\mu_{\tau_\star}$.

All fitted preprocessing, reference construction, and velocity-field training use retained populations. The steering coefficient ρ controls the combination of geometric and regulatory tangents before shape-preserving projection. We selected ρ using prediction error on the Gupta day 7 development holdout and fixed it for Gupta day 14 and all Giles and DSM holdouts. For each subsequent holdout, we refitted preprocessing and model parameters using its retained populations while keeping ρ fixed.

4 TRACE

4.1 OT-linked temporal tuples

We couple cells at consecutive retained times using balanced optimal transport (OT) in PCA space. Let $\Pi(\mu_{\tau_k}, \mu_{\tau_{k+1}})$ be the set of couplings between the empirical distributions at times τ_k and τ_{k+1}. With squared Euclidean cost, the coupling is

$$\pi_{k,k+1} = \arg\min_{\pi \in \Pi(\mu_{\tau_k}, \mu_{\tau_{k+1}})} \mathbb{E}_{(z,z')\sim\pi}\left[\|z' - z\|_2^2\right].$$

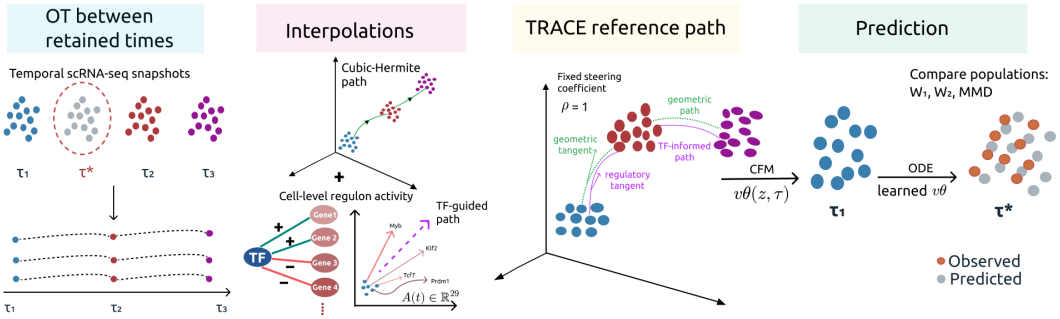


Figure 1: **TRACE overview.** We hold out the population at time $\tau_\star$ and couple cells across retained times using optimal transport. We construct cubic-Hermite reference paths and map temporal derivatives of signed TF-regulon activities into PCA space. Geometric and norm-calibrated regulatory tangents are combined using ρ before shape-preserving projection. We selected $\rho = 1$ on Gupta day 7 and fixed this coefficient for subsequent holdouts, refitting the model on each retained dataset. Using the resulting paths, we train $v_\theta(z, \tau)$ by conditional flow matching and integrate the learned dynamics to obtain $\widehat{\mu}_{\tau_\star}$. Predictions are evaluated against the observed population using W_1, W_2, and MMD.

Composing the conditional couplings gives a distribution over temporal tuples,

$$\Gamma(dz_{1:K}) = \mu_{\tau_1}(dz_1) \prod_{k=1}^{K-1} \pi_{k,k+1}(dz_{k+1} \mid z_k).$$

Each tuple $Z = (z_1, \ldots, z_K) \sim \Gamma$ links cell states across retained snapshots. The linked states describe statistical correspondences between independently sampled populations. During training, we approximate this construction with OT couplings between sampled cell batches, as detailed in Appendix B.2.

4.2 Shape-preserving cubic reference geometry

For each tuple Z, we construct a coordinatewise shape-preserving piecewise cubic Hermite interpolant (PCHIP) through the linked states. The geometric tangent at retained time τ_k is

$$m_k^{\text{geom}} = \text{PCHIPSlope}(Z; \tau_k).$$

For latent coordinate j, the secant slope between consecutive linked states is

$$\delta_{k,j} = \frac{z_{k+1,j} - z_{k,j}}{\tau_{k+1} - \tau_k}.$$

The secant is a finite difference between OT-linked knots with respect to normalized biological time. PCHIP zeros coordinate slopes at sign changes and limits overshoot on irregular intervals (Fritsch & Carlson, 1980; Fritsch & Butland, 1984). The resulting reference passes through each retained state: $H_Z^{\text{geom}}(\tau_k) = z_k$. Its tangents provide the cubic-CFM reference and the geometric component of TRACE.

4.3 Regulon-derived tangents in latent space

We infer a regulatory activity vector for each retained training cell from signed TF–target interactions. Let R be the number of retained regulons. For regulon $r \in \{1, \ldots, R\}$, let G_r denote its measurable target genes, w_{rg} the signed interaction weight for gene g, and $\widetilde{x}_{ig}$ the expression of gene g in cell i after standardization using the pooled retained training cells. The activity score is

$$A_{ir} = \frac{\sum_{g \in G_r} w_{rg} \widetilde{x}_{ig}}{\sqrt{\sum_{g \in G_r} w_{rg}^2 + \epsilon}}, \qquad \epsilon = 10^{-12}.$$

Here, stabilizer ϵ prevents numerical division by zero. Genes with training set standard deviation below 10^{-8} are assigned unit scale before activity scoring.

We standardized each regulon score by subtracting its pooled retained-training mean and dividing by its pooled retained-training standard deviation. Each cell in an OT-linked training tuple therefore carries both a latent state $z_k \in \mathbb{R}^p$ and a standardized regulon-activity vector $a_k \in \mathbb{R}^R$.

We used a T-cell-informed panel of 29 mouse TF regulons. The prior includes annotations for naive to exhausted T-cell programs and activity scoring uses only the signed TF-target interactions with CollecTRI provenance. These program annotations provide biological context to the regulatory path. Prior construction and filtering are detailed in Appendix B.1.

Let $(a_1, \ldots, a_K)$ denote the activities along a tuple. We estimate their temporal derivatives with the same shape-preserving slope construction,

$$m_k^{\mathrm{TF}} = \mathrm{PCHIPSlope}\big((a_1, \ldots, a_K); \tau_k\big).$$

To express these derivatives in latent coordinates, we fit an affine ridge map on retained cells for each run. Writing $\mathbf{A} \in \mathbb{R}^{N_{\mathrm{ret}} \times R}$ and $\mathbf{Z} \in \mathbb{R}^{N_{\mathrm{ret}} \times p}$ for the standardized activities and latent states, where N_{ret} is the number of retained training cells in that run, we obtain

$$(W, b) = \arg\min_{W,b} \|\mathbf{Z} - \mathbf{A}W - \mathbf{1}b\|_F^2 + \lambda \|W\|_F^2,$$

where $W \in \mathbb{R}^{R \times p}$ is the fitted weight matrix, $b \in \mathbb{R}^{1 \times p}$ is a row-vector intercept, $\mathbf{1} \in \mathbb{R}^{N_{\mathrm{ret}} \times 1}$ is a column of ones, and we use ridge coefficient $\lambda = 1$. We use row vectors for activities and tangents. Under this affine map, an activity derivative has the latent direction

$$m_k^{\mathrm{bio}} = m_k^{\mathrm{TF}} W.$$

We use these empirical directions to modulate the reference tangents. Let

$$g_k = \|m_k^{\mathrm{geom}}\|_2, \qquad q_k = \|m_k^{\mathrm{bio}}\|_2.$$

To account for differences in scale, we calibrate the regulatory tangent against the geometric tangent,

$$c_k = \begin{cases} 0, & g_k \le 10^{-8} \text{ or } q_k \le 10^{-8}, \\ \mathrm{clip}\left(\dfrac{g_k}{q_k + 10^{-12}}, 0.25, 4\right), & \text{otherwise}, \end{cases} \qquad \widetilde{m}_k^{\mathrm{bio}} = c_k m_k^{\mathrm{bio}}.$$

Near-zero tangents are handled using a fixed numerical tolerance; exact numerical safeguards are given in Appendix B.3.

4.4 TRACE REFERENCE PATHS

For $\rho \in [0, 1]$, we combine the geometric and calibrated regulatory tangents without changing the OT-linked knots:

$$u_k(\rho) = (1 - \rho) m_k^{\mathrm{geom}} + \rho \widetilde{m}_k^{\mathrm{bio}}, \qquad M_k(\rho) = \begin{cases} m_k^{\mathrm{geom}}, & \rho = 0, \\ \big[\Pi_{\mathrm{shape}}(u(\rho); Z, \tau)\big]_k, & 0 < \rho \le 1. \end{cases}$$

Appendix B.3 specifies the shape projection and endpoint handling. The resulting piecewise cubic-Hermite reference H_Z^{TRACE} satisfies $H_Z^{\mathrm{TRACE}}(\tau_k) = z_k$ at every retained time. At $\rho = 0$, the implementation uses the ordinary PCHIP tangents directly, so its reference path and training velocity target coincide with cubic CFM. We evaluated the prespecified values $\rho \in \{0, 0.25, 0.50, 0.75, 1.00\}$ using mean held-out W_2 across five training resamples on Gupta day 7. The selected value, $\rho = 1.0$, was fixed for Gupta day 14 and all Giles and DSM holdouts.

4.5 CONDITIONAL FLOW MATCHING AND PREDICTION

For a sampled tuple Z, let

$$\gamma_{\rho,Z}(\tau) = H_Z^{\mathrm{TRACE}}(\tau), \qquad \dot{\gamma}_{\rho,Z}(\tau) = \frac{d}{d\tau} H_Z^{\mathrm{TRACE}}(\tau)$$

be its reference position and velocity. We train the time-conditioned field v_θ to match this velocity (Lipman et al., 2023; Tong et al., 2024). To balance supervision, we allocate approximately equal

numbers of samples to the retained intervals within each mini-batch. Conditional on interval k, we sample $\tau \sim \mathrm{Uniform}(\tau_k, \tau_{k+1})$. Equal interval weights correspond to the time density

$$q(\tau) = \frac{1}{(K-1)(\tau_{k+1} - \tau_k)}, \qquad \tau \in [\tau_k, \tau_{k+1}).$$

The training objective is

$$\mathcal{L}_{\mathrm{TRACE}}(\theta) = \mathbb{E}_{\substack{Z \sim \Gamma \\ \tau \sim q}} \left[\left\| v_\theta\big(\gamma_{\rho,Z}(\tau), \tau\big) - \dot{\gamma}_{\rho,Z}(\tau) \right\|_2^2 \right].$$

Thus, we minimize the mean squared difference between the learned velocity and the reference-path derivative, weighting retained intervals equally while preserving their actual biological-time spacing. At evaluation, we integrate the learned field from cells at the immediately preceding retained time to the held-out time to obtain the predicted population.

5 Experimental Design

5.1 Datasets and temporal holdouts

Gupta/Schleicher. We pooled samples from days 1-4 into an early marginal assigned to day 2.5. The modeled times were days 2.5, 7, 14, and 21; we held out days 7 and 14 separately (Schleicher et al., 2025). Day 7 falls near the separation of exhausted and memory-like populations. By day 14, a considerable fraction of cells had progressed toward the day 21 endpoints.

Giles. We used the naive day 0 population and samples from days 8, 15, and 30, holding out days 8 and 15 separately (Giles et al., 2022). Day 8 captures the peak response. The exhausted-cell transcriptional program is established by day 15, although its chromatin landscape continues to change thereafter.

DSM. We used the chronic LCMV Clone 13 arm of the Fagerberg et al. dataset, with CD8$^+$ T-cell samples from days 4, 8, 28, and 40 (Fagerberg et al., 2025). We held out days 8 and 28 separately.

5.2 Temporal holdout and evaluation protocol

For each holdout, we excluded the target population before selecting highly variable genes or fitting expression scaling, PCA, regulon-activity standardization, the activity-to-latent ridge map, OT couplings, reference paths, and the velocity field. We transformed held-out cells only to evaluate predictions. Gupta day 7 was the development holdout used to select ρ by mean W_2 across five training resamples. We then fixed $\rho = 1.0$ for Gupta day 14 and all Giles and DSM holdouts, without using those target populations for model selection.

We compared TRACE with OT displacement, linear CFM, shape-preserving cubic CFM, MMFM (Rohbeck et al., 2025), MMSFM (Lee et al., 2025), ALI-CFM (Kviman et al., 2026), and moscot (Klein et al., 2025). To examine reference geometry under matched conditions, we also compared linear, quadratic, natural-cubic, and PCHIP paths using the same OT alignment, velocity network, training budget, and evaluation procedure. Linear CFM, cubic CFM, and TRACE additionally shared the same optimization and integration settings. For the other baselines, we used their method-specific training procedures or the benchmark adaptations documented in Appendix B.7.

For population prediction, we used the full population at the immediately preceding retained time as the source. The neural-flow models evolved these cells to the held-out time under their learned dynamics. We compared predicted and observed populations in the same holdout-specific PCA space using W_1, W_2, and the squared empirical maximum mean discrepancy, labeled MMD throughout. Wasserstein calculations used at most 1,500 predicted and 1,500 observed cells, whereas MMD used at most 1,000 cells from each population. For MMD, we used a Gaussian kernel with a shared bandwidth within each holdout and evaluation run, so all methods were evaluated under the same kernel bandwidth (Appendix B.6). We report the mean and population standard deviation across five runs per method. Linear CFM, cubic CFM, TRACE, and the regulatory controls used matched training resamples; method-specific seed settings are listed in Appendix B.7. Lower values indicate closer agreement.

Algorithm 1 TRACE training with regulatory tangent modulation

Require: Retained marginals $\{\mu_{\tau_k}\}$, retained expression matrices, signed TF-target prior, steering coefficient ρ

1: Select genes and fit expression scaling, PCA, and latent scaling on retained cells
2: Compute and standardize regulon activities on retained training cells
3: Fit the retained-training ridge map $\mathbf{Z} \approx \mathbf{A}W + \mathbf{1}b$
4: Initialize velocity-field parameters θ
5: **for** $n = 1, \ldots, N_{\text{steps}}$ **do**
6: Sample cell batches from the retained marginals and compute consecutive balanced-EMD couplings
7: Sample OT-linked tuples of latent states and their paired activity vectors
8: Compute geometric tangents m_k^{geom} and activity tangents m_k^{TF}
9: Map $m_k^{\text{bio}} = m_k^{\text{TF}} W$ and calibrate its norm against m_k^{geom}
10: Form $u_k(\rho) = (1-\rho)m_k^{\text{geom}} + \rho \widetilde{m}_k^{\text{bio}}$
11: Set $M_k(0) = m_k^{\text{geom}}$; for $\rho > 0$, set $M(\rho) = \Pi_{\text{shape}}(u(\rho); Z, \tau)$
12: Allocate tuples approximately equally across retained intervals; sample each τ uniformly within its assigned interval
13: Evaluate the Hermite-path positions $\gamma_{\rho,Z}(\tau)$ and velocities $\dot{\gamma}_{\rho,Z}(\tau)$
14: Update θ using the mini-batch mean of $\|v_\theta(\gamma_{\rho,Z}(\tau), \tau) - \dot{\gamma}_{\rho,Z}(\tau)\|_2^2$
15: **end for**
16: **return** trained velocity field v_θ

6 RESULTS

6.1 TIMEPOINT TRANSITIONS FAVORED DIFFERENT REFERENCE GEOMETRIES

We first tested whether reference geometry affected reconstruction differently across the two Gupta holdouts. Day 7 lies near the separation of exhausted and memory-like populations, whereas many day 14 cells have progressed toward the day 21 endpoints (Schleicher et al., 2025). We used day 7 for development and evaluated day 14 after fixing the model configuration.

At day 7, all three nonlinear interpolants reduced mean W_2 relative to linear CFM (12.605 ± 0.098). PCHIP cubic achieved the lowest W_1 and W_2 among the matched interpolation families, including $W_2 = 10.827 \pm 0.157$; quadratic interpolation achieved the lowest MMD (Table 1). At day 14, linear CFM achieved the lowest mean value for all three metrics in this ablation. The preferred interpolation family therefore depended on which population we reconstructed.

Table 1: Gupta interpolation-family comparison. All paths use matched OT alignment, velocity networks, training budgets, and evaluation. Values are mean ± population standard deviation across five matched training resamples. Bold indicates the lowest mean for each time and metric; lower is better.

	Day 7			Day 14		
Interpolator	$W_1 \downarrow$	$W_2 \downarrow$	MMD↓	$W_1 \downarrow$	$W_2 \downarrow$	MMD↓
Linear	12.317 ± 0.100	12.605 ± 0.098	0.3307 ± 0.0092	$\mathbf{8.680 \pm 0.064}$	$\mathbf{8.882 \pm 0.075}$	$\mathbf{0.0212 \pm 0.0014}$
Quadratic	10.897 ± 0.138	11.085 ± 0.143	$\mathbf{0.1581 \pm 0.0073}$	14.372 ± 0.101	14.675 ± 0.106	0.1872 ± 0.0070
Natural cubic	11.221 ± 0.091	11.437 ± 0.094	0.2048 ± 0.0067	12.202 ± 0.088	12.425 ± 0.090	0.1230 ± 0.0049
PCHIP cubic	$\mathbf{10.651 \pm 0.152}$	$\mathbf{10.827 \pm 0.157}$	0.1711 ± 0.0104	9.484 ± 0.028	9.644 ± 0.030	0.0458 ± 0.0012

6.2 REGULON-DERIVED STEERING IMPROVED THE GUPTA DEVELOPMENT PREDICTION

Given the day 7 benefit of cubic geometry, we tested whether regulon-derived tangents improved prediction further. TRACE reduced mean W_2 from 10.827 ± 0.157 with cubic CFM to 9.872 ± 0.159, an 8.8% reduction. Relative to linear CFM (12.605 ± 0.098), the reduction was 21.7% (Table 2). These development results used the coefficient $\rho = 1.0$ selected on the same holdout.

Table 2: Gupta population reconstruction across baseline methods and TRACE. Day 7 was the development holdout used to select ρ; day 14 was evaluated after fixing that choice. Each target population was excluded from fitted preprocessing, coupling, reference construction, and flow-field training. Values are mean $\pm$ population standard deviation across five runs; the direct CFM comparisons used matched resamples. Bold indicates the lowest mean for each time and metric; lower is better.

	Day 7			Day 14		
Method	$W_1 \downarrow$	$W_2 \downarrow$	MMD$\downarrow$	$W_1 \downarrow$	$W_2 \downarrow$	MMD$\downarrow$
OT displacement	12.183 ± 0.100	12.461 ± 0.091	0.3174 ± 0.0099	8.667 ± 0.030	8.856 ± 0.038	0.0222 ± 0.0007
Linear CFM	12.317 ± 0.100	12.605 ± 0.098	0.3307 ± 0.0092	8.680 ± 0.064	8.882 ± 0.075	0.0212 ± 0.0014
Cubic CFM	10.651 ± 0.152	10.827 ± 0.157	0.1711 ± 0.0104	9.484 ± 0.028	9.644 ± 0.030	0.0458 ± 0.0012
MMFM	11.194 ± 0.151	11.392 ± 0.156	0.1721 ± 0.0117	14.607 ± 0.240	14.814 ± 0.245	0.1948 ± 0.0135
MMSFM (SB)	11.45 ± 0.05	11.64 ± 0.05	0.175 ± 0.008	12.12 ± 0.16	12.39 ± 0.20	0.063 ± 0.007
ALI-CFM	11.058 ± 0.110	11.236 ± 0.136	0.2536 ± 0.0125	10.857 ± 0.408	11.016 ± 0.406	0.1149 ± 0.0173
moscot	11.979 ± 0.098	12.242 ± 0.093	0.3104 ± 0.0072	$\mathbf{8.374 \pm 0.065}$	$\mathbf{8.554 \pm 0.067}$	$\mathbf{0.0180 \pm 0.0016}$
TRACE	$\mathbf{9.721 \pm 0.153}$	$\mathbf{9.872 \pm 0.159}$	$\mathbf{0.1497 \pm 0.0096}$	8.953 ± 0.028	9.112 ± 0.030	0.0320 ± 0.0014

We assessed the contribution of regulatory structure with matched permutations (Appendix Table 6). On day 7, TRACE achieved lower W_2 than both controls in all five paired runs. Within-day cell permutation preserved the distribution of activities at each time while disrupting their correspondence to individual latent states. TF-channel permutation retained the fitted ridge map, so it disrupted the alignment between activity channels and that map. These comparisons support sensitivity to cell-specific regulatory features and their fitted channel organization. The TF-channel control also changes the feature-to-map interface, which limits conclusions about the TF-target prior itself.

6.3 TRACE IMPROVED EARLIER EXTERNAL PREDICTIONS WITH A FIXED STEERING COEFFICIENT

To test the development-selected steering strength across studies, we fixed $\rho = 1.0$ and evaluated the Giles and DSM holdouts without external retuning. At Giles day 8, TRACE achieved mean $W_2 = 18.097 \pm 1.358$, compared with 20.440 ± 0.521 for linear CFM and 18.386 ± 0.820 for cubic CFM. These correspond to reductions of 11.5% and 1.6%, respectively. MMSFM achieved the lowest MMD on this holdout (Table 3).

At DSM day 8, TRACE achieved mean $W_2 = 10.725 \pm 0.140$, a 4.0% reduction from linear CFM (11.175 ± 0.111) and a 9.3% reduction from cubic CFM (11.820 ± 0.151). TRACE also achieved the lowest mean W_1 and MMD among the compared methods (Table 4). On DSM day 8, TRACE had lower W_2 than each permutation control in four of five runs, with a small advantage over cell permutation (Appendix Table 8).

Table 3: Giles temporal-holdout benchmark. TRACE used $\rho = 1.0$, selected on Gupta day 7 without Giles retuning. Values are mean $\pm$ population standard deviation across five runs. Bold indicates the lowest mean for each time and metric; lower is better.

	Day 8			Day 15		
Method	$W_1 \downarrow$	$W_2 \downarrow$	MMD$\downarrow$	$W_1 \downarrow$	$W_2 \downarrow$	MMD$\downarrow$
Linear CFM	18.579 ± 0.339	20.440 ± 0.521	0.3465 ± 0.0121	11.535 ± 0.078	12.273 ± 0.087	0.0892 ± 0.0040
Cubic CFM	16.436 ± 0.509	18.386 ± 0.820	0.2125 ± 0.0135	13.934 ± 0.161	15.132 ± 0.194	0.1345 ± 0.0039
MMFM	16.954 ± 0.390	18.657 ± 0.639	0.2238 ± 0.0154	19.778 ± 0.306	21.151 ± 0.351	0.2850 ± 0.0093
MMSFM (SB)	17.608 ± 0.344	19.382 ± 0.412	$\mathbf{0.1814 \pm 0.0079}$	14.764 ± 0.266	15.038 ± 0.257	0.1852 ± 0.0129
ALI-CFM	18.667 ± 0.647	20.712 ± 0.788	0.3718 ± 0.0266	12.487 ± 1.145	12.925 ± 1.175	0.1603 ± 0.0314
moscot	18.382 ± 0.151	20.326 ± 0.193	0.3252 ± 0.0057	$\mathbf{10.950 \pm 0.133}$	$\mathbf{11.612 \pm 0.165}$	$\mathbf{0.0784 \pm 0.0042}$
TRACE	$\mathbf{16.094 \pm 0.835}$	$\mathbf{18.097 \pm 1.358}$	0.2032 ± 0.0203	13.577 ± 0.154	14.808 ± 0.189	0.1278 ± 0.0046

6.4 LATER HOLDOUTS FAVORED SIMPLER OR ALTERNATIVE TRANSPORT

To assess whether the early gains persisted, we evaluated the later holdouts with the same fixed coefficient. On Gupta day 14, TRACE achieved $W_2 = 9.112 \pm 0.030$, compared with 8.882 ± 0.075 for linear CFM and 8.554 ± 0.067 for moscot (Table 2). On Giles day 15, TRACE obtained 14.808 ± 0.189, while linear CFM obtained 12.273 ± 0.087 and moscot obtained 11.612 ± 0.165 (Table 3). Both regulatory permutation controls also achieved lower W_2 than TRACE in all five paired Giles day 15 runs (Appendix Table 7).

Table 4: DSM temporal-holdout benchmark. TRACE used $\rho = 1.0$, selected on Gupta day 7 without DSM retuning. Values are mean $\pm$ population standard deviation across five runs. Bold indicates the lowest mean for each time and metric; lower is better.

	Day 8			Day 28		
Method	$W_1 \downarrow$	$W_2 \downarrow$	MMD$\downarrow$	$W_1 \downarrow$	$W_2 \downarrow$	MMD$\downarrow$
Linear CFM	10.797 ± 0.110	11.175 ± 0.111	0.1086 ± 0.0046	13.685 ± 0.123	14.278 ± 0.152	$\mathbf{0.1742 \pm 0.0018}$
Cubic CFM	11.566 ± 0.146	11.820 ± 0.151	0.1368 ± 0.0061	13.850 ± 0.140	14.441 ± 0.169	0.1838 ± 0.0031
MMFM	12.024 ± 0.097	12.255 ± 0.096	0.1496 ± 0.0064	32.803 ± 0.472	33.713 ± 0.508	0.3896 ± 0.0099
MMSFM (SB)	12.120 ± 0.111	12.366 ± 0.112	0.1312 ± 0.0061	16.094 ± 0.465	16.478 ± 0.494	0.1770 ± 0.0146
ALI-CFM	12.068 ± 0.584	12.327 ± 0.598	0.2059 ± 0.0280	14.640 ± 0.418	15.233 ± 0.482	0.2359 ± 0.0131
moscot	10.715 ± 0.085	11.098 ± 0.077	0.1071 ± 0.0039	$\mathbf{13.137 \pm 0.142}$	$\mathbf{13.756 \pm 0.186}$	0.1845 ± 0.0030
TRACE	$\mathbf{10.458 \pm 0.135}$	$\mathbf{10.725 \pm 0.140}$	$\mathbf{0.0941 \pm 0.0062}$	13.678 ± 0.077	14.292 ± 0.104	0.1786 ± 0.0038

On DSM day 28, TRACE and linear CFM had similar mean W_2 values (14.292 ± 0.104 and 14.278 ± 0.152), while moscot achieved 13.756 ± 0.186 (Table 4). The later Gupta and Giles holdouts followed substantial progression toward the day 21 endpoints and establishment of the exhausted-cell transcriptional program, respectively (Schleicher et al., 2025; Giles et al., 2022). Across the three studies, the earlier holdout improvements support using regulatory-informed reference paths to predict populations undergoing exhaustion-associated changes.

7 Discussion

In this work, we developed TRACE to guide flow-matching reference paths with temporal changes in TF-regulon activity. Because many paths connect the same observed populations, we used coordinated TF-target expression to inform the intervening directions. With the steering coefficient fixed after Gupta day 7, this approach improved prediction at earlier holdouts across three chronic-LCMV studies. Linear CFM or moscot performed better later, motivating interval-specific steering selected using retained populations. Such an extension would allow the contribution of regulatory information to change over the course of differentiation.

Because regulon activities and latent states were derived from the same expression measurements, we used the signed TF-target prior to organize expression changes before mapping them into reference tangents. We observed positive alignment between mapped activity changes and OT-linked latent displacements across all six holdouts (Appendix Table 9), supporting the empirical map used for directional supervision. However, both permutation controls retained substantial gains over cubic CFM on DSM day 8, and the advantage of TRACE over cell permutation was small. We also held the ridge map fixed during TF-channel permutation, which disrupted the correspondence between activity features and the fitted map. Comparisons with randomized TF-target priors and separately fitted maps would more directly determine how much predictive information came from the signed regulatory relationships.

Identifying which TF-target relationships improve prediction would help prioritize candidate regulators of Tpex differentiation. These cells retain self-renewal and proliferative capacity and replenish downstream exhausted populations (Im et al., 2016; Utzschneider et al., 2020; Beltra et al., 2020), making the timing of their differentiation relevant to sustained immune responses. We could test candidate TFs through perturbations during the earlier intervals, followed by lineage measurements and quantification of cell-state composition. Such experiments would determine whether the regulatory changes used for prediction influence subsequent cell fate. Across the present studies, our findings support incorporating expression-derived regulatory direction into temporal population prediction during early exhaustion-associated transitions.

Ethics Statement

This study analyzes previously collected single-cell datasets and introduces no new human- or animal-subject experiments. TRACE models population-level statistical transport and is not interpreted as recovering individual cell lineages, causal regulatory mechanisms, or clinically actionable interventions.

Reproducibility Statement

For each holdout, all subsequently fitted quantities, including gene selection, expression and latent scaling, PCA, regulon standardization, activity-to-latent mapping, OT coupling, reference-path construction, and flow-field training, used retained cells only. Gupta day 7 target metrics were used for development selection of ρ but not for gradient training; Gupta day 14 and external target metrics were excluded from model selection.

AI Use Statement

Generative AI tools were used for manuscript language editing and for assistance with code drafting and debugging. All reported numerical results were produced by the authors' computational pipelines and checked against saved outputs. The authors reviewed and revised all AI-assisted text and code and take responsibility for the final manuscript, analyses, and artifacts.

References

Michael S. Albergo, Nicholas M. Boffi, Michael Lindsey, and Eric Vanden-Eijnden. Multimarginal generative modeling with stochastic interpolants. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/f72af04ff6852f2669b2ed620c3c86c8-Abstract-Conference.html.

Mariano J Alvarez, Yao Shen, Federico M Giorgi, Alexander Lachmann, B Belinda Ding, B Hilda Ye, and Andrea Califano. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nature genetics*, 48(8):838–847, 2016. doi: 10.1038/ng.3593.

Pau Badia-i Mompel, Jesús Vélez Santiago, Jana Braunger, Celina Geiss, Daniel Dimitrov, Sophia Müller-Dott, Petr Taus, Aurelien Dugourd, Christian H. Holland, Ricardo O. Ramirez Flores, and Julio Saez-Rodriguez. decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinformatics Advances*, 2(1):vbac016, 2022. doi: 10.1093/bioadv/vbac016. URL https://doi.org/10.1093/bioadv/vbac016.

Jean-Christophe Beltra, Sasikanth Manne, Mohamed S Abdel-Hakeem, Makoto Kurachi, Josephine R Giles, Zeyu Chen, Valentina Casella, Shin Foong Ngiow, Omar Khan, Yinghui Jane Huang, et al. Developmental relationships of four exhausted CD8+ t cell subsets reveals underlying transcriptional and epigenetic landscape control mechanisms. *Immunity*, 52(5):825–841, 2020.

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*, 38(12): 1408–1414, 2020.

Zeyu Chen, Zhicheng Ji, Shin Foong Ngiow, Sasikanth Manne, Zhangying Cai, Alexander C. Huang, John Johnson, Ryan P. Staupe, Bertram Bengsch, Caiyue Xu, Sixiang Yu, Makoto Kurachi, Ramin S. Herati, Laura A. Vella, Amy E. Baxter, Jennifer E. Wu, Omar Khan, Jean-Christophe Beltra, Josephine R. Giles, Erietta Stelekati, Laura M. McLane, Chi Wai Lau, Xiaolu Yang, Shelley L. Berger, Golnaz Vahedi, Hongkai Ji, and E. John Wherry. TCF-1-centered transcriptional network drives an effector versus exhausted CD8 T cell-fate decision. *Immunity*, 51(5):840–855.e5, 2019. doi: 10.1016/j.immuni.2019.09.013. URL https://doi.org/10.1016/j.immuni.2019.09.013.

Bence Daniel, Kathryn E Yost, Sunnie Hsiung, Katalin Sandor, Yu Xia, Yanyan Qi, Kamir J Hiam-Galvez, Mollie Black, Colin J. Raposo, Quanming Shi, et al. Divergent clonal differentiation trajectories of t cell exhaustion. *Nature immunology*, 23(11):1614–1627, 2022.

Eric Fagerberg, John Attanasio, Christine Dien, Jaiveer Singh, Emily A Kessler, Leena Abdullah, Jian Shen, Brian G Hunt, Kelli A Connolly, Edward De Brouwer, et al. KLF2 maintains lineage fidelity and suppresses CD8 t cell exhaustion during acute LCMV infection. *Science*, 387(6735): eadn2337, 2025.

Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T. H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. POT: Python optimal transport. *Journal of Machine Learning Research*, 22(78): 1–8, 2021. URL https://www.jmlr.org/papers/v22/20-451.html.

F. N. Fritsch and J. Butland. A method for constructing local monotone piecewise cubic interpolants. *SIAM journal on scientific and statistical computing*, 5(2):300–304, 1984. doi: 10.1137/0905021.

Frederick N Fritsch and Ralph E Carlson. Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17(2):238–246, 1980. doi: 10.1137/0717021.

Josephine R Giles, Shin Foong Ngiow, Sasikanth Manne, Amy E Baxter, Omar Khan, Ping Wang, Ryan Staupe, Mohamed S Abdel-Hakeem, Hua Huang, Divij Mathew, et al. Shared and distinct biological circuits in effector, memory and exhausted CD8+ t cells revealed by temporal single-cell transcriptomics and epigenetics. *Nature immunology*, 23(11):1600–1613, 2022. doi: 10.1038/s41590-022-01338-4.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773, 2012. URL https://www.jmlr.org/papers/v13/gretton12a.html.

Christian H. Holland, Jovan Tanevski, Javier Perales-Patón, Jan Gleixner, Manu P. Kumar, Elisabetta Mereu, Brian A. Joughin, Oliver Stegle, Douglas A. Lauffenburger, Holger Heyn, Bence Szalai, and Julio Saez-Rodriguez. Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. *Genome Biology*, 21:36, 2020. doi: 10.1186/s13059-020-1949-z. URL https://doi.org/10.1186/s13059-020-1949-z.

Guillaume Huguet, Daniel Sumner Magruder, Alexander Tong, Oluwadamilola Fasina, Manik Kuchroo, Guy Wolf, and Smita Krishnaswamy. Manifold interpolating optimal-transport flows for trajectory inference. *Advances in neural information processing systems*, 35:29705–29718, 2022.

Se Jin Im, Masao Hashimoto, Michael Y Gerner, Junghwa Lee, Haydn T Kissick, Matheus C Burger, Qiang Shan, J Scott Hale, Judong Lee, Tahseen H Nasti, et al. Defining CD8+ t cells that provide the proliferative burst after PD-1 therapy. *Nature*, 537(7620):417–421, 2016.

Kenji Kamimoto, Blerta Stringa, Christy M. Hoffmann, Kunal Jindal, Lilianna Solnica-Krezel, and Samantha A. Morris. Dissecting cell identity via network inference and in silico gene perturbation. *Nature*, 614:742–751, 2023. doi: 10.1038/s41586-022-05688-9. URL https://www.nature.com/articles/s41586-022-05688-9.

Kacper Kapuśniak, Peter Potaptchik, Teodora Reu, Leo Zhang, Alexander Tong, Michael Bronstein, Avishek Joey Bose, and Francesco Di Giovanni. Metric flow matching for smooth interpolations on the data manifold. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-4291. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/f381114cf5aba4e45552869863deaaa7-Abstract-Conference.html.

Omar Khan, Josephine R. Giles, Sierra McDonald, Sasikanth Manne, Shin Foong Ngiow, Kunal P. Patel, Michael T. Werner, Alexander C. Huang, Katherine A. Alexander, Jennifer E. Wu, John Attanasio, Patrick Yan, Sangeeth M. George, Bertram Bengsch, Ryan P. Staupe, Greg Donahue, Wei Xu, Ravi K. Amaravadi, Xiaowei Xu, Giorgos C. Karakousis, Tara C. Mitchell, Lynn M. Schuchter, Jonathan Kaye, Shelley L. Berger, and E. John Wherry. TOX transcriptionally and epigenetically programs CD8+ T cell exhaustion. *Nature*, 571(7764):211–218, 2019. doi: 10.1038/s41586-019-1325-x. URL https://www.nature.com/articles/s41586-019-1325-x.

Dominik Klein, Giovanni Palla, Marius Lange, Michal Klein, Zoe Piran, Manuel Gander, Laetitia Meng-Papaxanthos, Michael Sterr, Lama Saber, Changying Jing, et al. Mapping cells through time and space with moscot. *Nature*, 638(8052):1065–1075, 2025.

Oskar Kviman, Kirill Tamogashev, Nicola Branchini, Víctor Elvira, Jens Lagergren, and Nikolay Malkin. Multi-marginal flow matching with adversarially learnt interpolants. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/cb0ce861adaf6f8a93069c064733f402-Abstract-Conference.html.

Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E Kastriti, Peter Lönnerberg, Alessandro Furlan, et al. RNA velocity of single cells. *Nature*, 560(7719):494–498, 2018.

François Lamoline, Isabel Haasler, Johan Karlsson, Jorge Gonçalves, and Atte Aalto. Dynamic gene regulatory network inference from single-cell data using optimal transport. *Bioinformatics*, 41(8): btaf394, 2025. doi: 10.1093/bioinformatics/btaf394. URL https://academic.oup.com/bioinformatics/article/41/8/btaf394/8198065.

Marius Lange, Volker Bergen, Michal Klein, Manu Setty, Bernhard Reuter, Mostafa Bakhti, Heiko Lickert, Meshal Ansari, Janine Schniering, Herbert B Schiller, et al. CellRank for directed single-cell fate mapping. *Nature methods*, 19(2):159–170, 2022.

Justin Lee, Behnaz Moradijamei, and Heman Shakeri. Multi-marginal stochastic flow matching for high-dimensional snapshot data at irregular time points. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 33476–33498. PMLR, 2025. URL https://proceedings.mlr.press/v267/lee25x.html.

Shengyu Li, Pengzhi Zhang, Weiqing Chen, Lingqun Ye, Kristopher W Brannan, Nhat-Tu Le, Junichi Abe, John P Cooke, and Guangyu Wang. A relay velocity model infers cell-dependent RNA velocity. *Nature biotechnology*, 42(1):99–108, 2024.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Erpai Luo, Minsheng Hao, Lei Wei, and Xuegong Zhang. scDiffusion: conditional generation of high-quality single-cell data using diffusion model. *Bioinformatics*, 40(9):btae518, 2024.

Brian C. Miller, Debattama R. Sen, Rose Al Abosy, Kevin Bi, Yamini V. Virkud, Martin W. LaFleur, Kathleen B. Yates, Ana Lako, Kristen Felt, Girish S. Naik, Michael Manos, Evisa Gjini, Juhi R. Kuchroo, Jeffrey J. Ishizuka, Jenna L. Collier, Gabriel K. Griffin, Seth Maleri, Dawn E. Comstock, Sarah A. Weiss, Flavian D. Brown, Arpit Panda, Margaret D. Zimmer, Robert T. Manguso, F. Stephen Hodi, Scott J. Rodig, Arlene H. Sharpe, and W. Nicholas Haining. Subsets of exhausted CD8+ T cells differentially mediate tumor control and respond to checkpoint blockade. *Nature Immunology*, 20(3):326–336, 2019. doi: 10.1038/s41590-019-0312-6. URL https://doi.org/10.1038/s41590-019-0312-6.

Sophia Müller-Dott, Eirini Tsirvouli, Miguel Vazquez, Ricardo O Ramirez Flores, Pau Badia-i Mompel, Robin Fallegger, Dénes Türei, Astrid Lægreid, and Julio Saez-Rodriguez. Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities. *Nucleic acids research*, 51(20):10934–10949, 2023.

Alessandro Palma, Till Richter, Hanyi Zhang, Manuel Lubetzki, Alexander Tong, Andrea Dittadi, and Fabian Theis. Multi-modal and multi-attribute generation of single cells with CFGen. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/13079a5dd5c994a7e5f0cc691e0338e8-Abstract-Conference.html.

Kristen E. Pauken, Morgan A. Sammons, Pamela M. Odorizzi, Sasikanth Manne, Jernej Godec, Omar Khan, Adam M. Drake, Zeyu Chen, Debattama R. Sen, Makoto Kurachi, R. Anthony Barnitz, Caroline Bartman, Bertram Bengsch, Alexander C. Huang, Jason M. Schenkel, Golnaz Vahedi, W. Nicholas Haining, Shelley L. Berger, and E. John Wherry. Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. *Science*, 354(6316):1160–1165, 2016. doi: 10.1126/science.aaf2807. URL https://doi.org/10.1126/science.aaf2807.

Katarina Petrović, Lazar Atanackovic, Viggo Moro, Kacper Kapuśniak, Ismail Ilkan Ceylan, Michael Bronstein, Joey Bose, and Alexander Tong. Curly flow matching for learning non-gradient field dynamics. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-4053. URL https://papers.nips.cc/paper_files/paper/2025/hash/b00692865928625ace212e89ded2decd-Abstract-Conference.html.

Santanu Subhash Rathod, Pietro Liò, and Xiao Zhang. SplineFlow: Flow matching for dynamical systems with B-spline interpolants, 2026. URL https://arxiv.org/abs/2601.23072. arXiv preprint.

Martin Rohbeck, Edward De Brouwer, Charlotte Bunne, Jan-Christian Huetter, Anne Biton, Kelvin Chen, Aviv Regev, and Romain Lopez. Modeling complex system dynamics with flow matching across time and conditions. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/206b764b00195b8c896c5af8bde80726-Abstract-Conference.html.

Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943, 2019.

Jan T Schleicher, Revant Gupta, Dario Cerletti, Ioana Sandu, Annette Oxenius, and Manfred Claassen. Exploratory trajectory inference reveals convergent lineages for CD8 t cells in chronic LCMV infection. *Plos one*, 20(9):e0332406, 2025.

Manu Setty, Vaidotas Kiseliovas, Jacob Levine, Adam Gayoso, Linas Mazutis, and Dana Pe'er. Characterization of cell fate probabilities in single-cell data with Palantir. *Nature biotechnology*, 37(4):451–460, 2019.

Yutong Sha, Yuchi Qiu, Peijie Zhou, and Qing Nie. Reconstructing growth and dynamic trajectories from single-cell transcriptomics data. *Nature Machine Intelligence*, 6(1):25–39, 2024. doi: 10.1038/s42256-023-00763-w. URL https://www.nature.com/articles/s42256-023-00763-w.

Yunyi Shen, Renato Berlinghieri, and Tamara Broderick. Multi-marginal schrödinger bridges with iterative reference refinement. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 3817–3825. PMLR, 2025. URL https://proceedings.mlr.press/v258/shen25b.html.

Imran Siddiqui, Karin Schaeuble, Vijaykumar Chennupati, Silvia A. Fuertes Marraco, Sandra Calderon-Copete, Daniela Pais Ferreira, Santiago J. Carmona, Leonardo Scarpellino, David Gfeller, Sylvain Pradervand, Sanjiv A. Luther, Daniel E. Speiser, and Werner Held. Intratumoral Tcf1+PD-1+CD8+ T cells with stem-like properties promote tumor control in response to vaccination and checkpoint blockade immunotherapy. *Immunity*, 50(1):195–211.e10, 2019. doi: 10.1016/j.immuni.2018.12.021. URL https://doi.org/10.1016/j.immuni.2018.12.021.

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, 2018.

Alexander Tong, Jessie Huang, Guy Wolf, David Van Dijk, and Smita Krishnaswamy. TrajectoryNet: A dynamic optimal transport network for modeling cellular dynamics. In *International conference on machine learning*, pp. 9526–9536. PMLR, 2020.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024. URL https://openreview.net/forum?id=CD9Snc73AW.

Tsz Pan Tong and Jun Pang. FlowGRN: Scalable and dropout-robust gene regulatory network inference via flow matching-based trajectory reconstruction. In *Proceedings of the 16th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. Association for Computing Machinery, 2025. doi: 10.1145/3765612.3767196. URL https://doi.org/10.1145/3765612.3767196.

Tsz Pan Tong and Jun Pang. FlowGRN+: Improving gene regulatory network inference by spline fitting and manifold projection in conditional flow matching (technical report), 2026. URL https://arxiv.org/abs/2608.10407. arXiv technical report.

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, 2014.

Daniel T Utzschneider, Sarah S Gabriel, David Chisanga, Renee Gloury, Patrick M Gubser, Ajithkumar Vasanthakumar, Wei Shi, and Axel Kallies. Early precursor t cells establish and propagate t cell exhaustion in chronic infection. *Nature immunology*, 21(10):1256–1266, 2020.

Dongyi Wang, Yuanwei Jiang, Zhenyi Zhang, Xiang Gu, Peijie Zhou, and Jian Sun. Joint velocity-growth flow matching for single-cell dynamics modeling. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-5363. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/eb1bad7a84ef68a64f1afd6577725d45-Abstract-Conference.html.

Weixu Wang, Zhiyuan Hu, Philipp Weiler, Sarah Mayes, Marius Lange, Daniel M. Fountain, Julianna O. Haug, Jingye Wang, Zhengyuan Xue, Tatjana Sauka-Spengler, and Fabian J. Theis. RegVelo: Gene-regulatory-informed dynamics of single cells. *Cell*, 189(12):3773–3800.e44, 2026. doi: 10.1016/j.cell.2026.04.022. URL https://doi.org/10.1016/j.cell.2026.04.022.

Philipp Weiler, Marius Lange, Michal Klein, Dana Pe'er, and Fabian J. Theis. CellRank 2: unified fate mapping in multiview single-cell data. *Nature Methods*, 21(7):1196–1205, 2024. doi: 10.1038/s41592-024-02303-9. URL https://www.nature.com/articles/s41592-024-02303-9.

F Alexander Wolf, Fiona K Hamey, Mireya Plass, Jordi Solana, Joakim S Dahlin, Berthold Göttgens, Nikolaus Rajewsky, Lukas Simon, and Fabian J Theis. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome biology*, 20(1):59, 2019.

Grace Hui Ting Yeo, Sachit D Saksena, and David K Gifford. Generative modeling of single-cell time series with PRESCIENT enables prediction of cell trajectories with interventions. *Nature communications*, 12(1):3222, 2021.

Wenjun Zhao, Erica Larschan, Björn Sandstede, and Ritambhara Singh. Optimal transport reveals dynamic gene regulatory networks via gene velocity estimation. *PLOS Computational Biology*, 21(5):e1012476, 2025. doi: 10.1371/journal.pcbi.1012476. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1012476.

A Supplementary Analyses

A.1 Development sweep for the steering coefficient

We used Gupta day 7 as the only temporal target for selecting ρ. Mean held-out W_2 decreased monotonically across the prespecified sweep:

The minimum occurred at $\rho = 1.00$, which we froze before Gupta day 14 and all Giles/DSM evaluations. Gupta day 7 therefore served as the development-validation target.

A.2 Paired regulatory controls across temporal holdouts

B Implementation Details

B.1 Dataset preprocessing, latent representation, and regulatory prior

Dataset-specific inputs. The input matrices contain total-count-normalized, log-transformed expression. The Gupta/Schleicher development dataset contains 13,763 cells and 2000 input expression

Table 5: Gupta day-7 development sweep used to select the TRACE steering coefficient. Lower W_2 is better.

ρ	Mean W_2
0.00	10.827
0.25	10.391
0.50	10.061
0.75	9.909
1.00	**9.872**

Table 6: Paired W_2 values across five matched training runs for the Gupta temporal holdouts. Within each row, all methods use the same training resample and random seed. Lower is better.

Holdout	Seed	Linear	Cubic	TRACE	Cell-perm	TF-perm
d7	0	12.520	11.073	**9.935**	10.480	11.964
d7	1	12.745	10.842	**9.556**	10.228	12.780
d7	2	12.470	10.576	**9.931**	10.187	11.670
d7	3	12.653	10.824	**9.961**	10.427	12.447
d7	4	12.636	10.818	**9.978**	10.220	12.115
d14	0	**8.971**	9.652	9.125	9.113	9.136
d14	1	**8.774**	9.644	9.131	9.045	9.043
d14	2	**8.942**	9.591	9.057	9.008	8.997
d14	3	**8.816**	9.647	9.107	9.035	9.076
d14	4	**8.904**	9.686	9.142	9.130	9.004

Table 7: Paired W_2 values across the five matched Giles training resamples. Lower is better.

Holdout	Seed	Linear	Cubic	TRACE	Cell-perm	TF-perm
d8	0	20.790	**18.806**	18.916	19.342	20.259
d8	1	20.510	18.976	18.930	**18.920**	19.014
d8	2	19.435	17.050	**15.430**	17.247	17.864
d8	3	20.894	19.255	**18.943**	18.964	19.672
d8	4	20.570	**17.841**	18.268	19.229	20.174
d15	0	**12.376**	15.339	14.942	13.995	13.730
d15	1	**12.311**	15.028	14.772	13.890	13.617
d15	2	**12.214**	15.215	14.881	14.023	14.070
d15	3	**12.134**	14.804	14.459	13.682	13.378
d15	4	**12.329**	15.276	14.988	14.058	14.125

Table 8: Paired W_2 values across the five matched DSM training resamples. Lower is better.

Holdout	Seed	Linear	Cubic	TRACE	Cell-perm	TF-perm
d8	0	11.319	11.987	**10.851**	10.988	11.157
d8	1	11.047	11.609	**10.536**	10.650	10.722
d8	2	11.298	11.994	10.901	10.883	**10.809**
d8	3	11.096	11.789	**10.733**	10.804	11.049
d8	4	11.116	11.718	**10.603**	10.761	11.108
d28	0	14.242	14.149	**14.092**	14.334	14.457
d28	1	14.326	14.431	14.301	**14.278**	14.490
d28	2	14.337	14.675	14.371	**14.299**	14.354
d28	3	**14.014**	14.487	14.319	14.298	14.646
d28	4	14.473	14.464	**14.376**	14.436	14.487

Table 9: Directional alignment of the activity-to-latent map on retained cells. Cosine similarity compares the mapped regulatory displacement $(\Delta A)W$ with the corresponding OT-linked latent-state displacement.

Dataset	Holdout	Median cosine ↑	Fraction > 0 ↑
Gupta	d7	0.335	0.932
Gupta	d14	0.316	0.940
Giles	d8	0.260	0.919
Giles	d15	0.285	0.944
DSM	d8	0.270	0.925
DSM	d28	0.241	0.905

features at modeled days $\{2.5, 7, 14, 21\}$. Original activation samples from days 1-4 are pooled into the day-2.5 marginal. The modeled marginals contain 2,338, 3,687, 4,137, and 3,601 cells, respectively. Giles contains 17,920 cells and 2000 features at days $\{0, 8, 15, 30\}$; the modeled input includes naive and chronic Cl13 cells and excludes Armstrong cells. DSM contains 20,478 cells and 2000 features at days $\{4, 8, 28, 40\}$.

Feature panels and holdout-specific fits. Input gene panels were constructed excluding the temporal holdout evaluation: highly variable genes were selected using retained study matrix, and available targets from the TF-target prior in the regulatory representations. For each temporal holdout h, let $\mathcal{T}_{-h}$ denote the retained marginals. Expression standardization, estimates gene means and standard deviations from $\mathcal{T}_{-h}$ was applied. A 30-component PCA is fitted using retained cells only. OT alignment, reference construction, and flow training use those standardized latent coordinates; held-out cells do not contribute to these fitted quantities. For regulon scoring, gene-level means and standard deviations are fitted on retained training cells only. Raw regulon activities are standardized using their pooled retained-training mean and standard deviation.

Regulatory prior. The mouse T-cell prior contains program metadata, but only rows marked `provenance=collectri` supply the signed TF-target interactions used in activity scoring (Müller-Dott et al., 2023). Repeated TF-target entries across programs are collapsed by averaging their edge values. In the primary sign-weight configuration, each interaction uses its activation or repression sign; a missing or zero explicit sign falls back to the sign of the stored edge weight.

Regulons required at least five measurable target genes and at least 15% coverage of their targets in the prior, without a top-k cap. Selection on the Gupta development feature representation yielded 29 regulons, whose TF identities were fixed for subsequent Giles and DSM evaluations. The archived prior contains 2,580 CollecTRI-provenance records, collapsing to 1,642 distinct TF–target interactions.

B.2 Optimal-transport alignment

TRACE uses balanced optimal transport to align cells between consecutive retained temporal marginals. For latent batches $X \in \mathbb{R}^{n\times p}$ and $Y \in \mathbb{R}^{m\times p}$, the transport cost is squared Euclidean,

$$C_{ij} = \|X_i - Y_j\|_2^2,$$

with uniform mass vectors $a = (1/n, \ldots, 1/n)$ and $b = (1/m, \ldots, 1/m)$. Here, $\Pi(a, b)$ denotes the nonnegative transport matrices with row sums a and column sums b. We solve

$$P^\star = \operatorname*{argmin}_{P\in\Pi(a,b)} \langle P, C\rangle$$

using the Python Optimal Transport implementation `ot.emd` (Flamary et al., 2021). During neural-flow training, at most 128 cells are sampled without replacement from each retained marginal. Consecutive batches are coupled sequentially. For each source cell, the corresponding row of the OT plan is normalized,

$$\widehat{P}_{ij} = \frac{P_{ij}}{\sum_j P_{ij} + 10^{-12}}$$

and a destination index is sampled from this conditional distribution. Regulon-activity vectors are carried with the same sampled cells, so latent state and regulatory state remain paired along the resulting multi-time tuple.

B.3 Reference-path construction

Numerical safeguards for norm calibration. Let

$$g_k = \|m_k^{\mathrm{geom}}\|_2, \qquad q_k = \|m_k^{\mathrm{bio}}\|_2.$$

To avoid unstable rescaling of near-zero tangents, if either $g_k \le 10^{-8}$ or $q_k \le 10^{-8}$, we set $c_k = 0$. Otherwise,

$$c_k = \mathrm{clip}\left(\frac{g_k}{q_k + 10^{-12}}, 0.25, 4\right), \qquad \widetilde{m}_k^{\mathrm{bio}} = c_k m_k^{\mathrm{bio}}.$$

The threshold 10^{-8} treats near-zero tangent norms as numerically zero, while 10^{-12} prevents division by zero in the calibration ratio.

Shape-preserving tangent projection. For an OT-linked tuple $Z = (z_1, \dots, z_K)$ at retained times $\tau_1 < \cdots < \tau_K$, let

$$\delta_{k,j} = \frac{z_{k+1,j} - z_{k,j}}{\tau_{k+1} - \tau_k}, \qquad k = 1, \dots, K-1,$$

be the secant slope in latent coordinate j. The operator $\Pi_{\mathrm{shape}}(u; Z, \tau)$ acts coordinatewise on candidate knot tangents $u = (u_1, \dots, u_K)$. Define the endpoint rule

$$E(q, \delta) = \begin{cases} \mathrm{sgn}(\delta)\min\{|q|, 3|\delta|\}, & q\delta > 0, \\ 0, & \text{otherwise.} \end{cases}$$

The first and final knot tangents are

$$[\Pi_{\mathrm{shape}}(u)]_{1,j} = E(u_{1,j}, \delta_{1,j}),$$
$$[\Pi_{\mathrm{shape}}(u)]_{K,j} = E(u_{K,j}, \delta_{K-1,j}).$$

At an interior knot $k = 2, \dots, K-1$, the rule is

$$[\Pi_{\mathrm{shape}}(u)]_{k,j} = \begin{cases} \mathrm{sgn}(\delta_{k-1,j})\min\big\{|u_{k,j}|, 3\min(|\delta_{k-1,j}|, |\delta_{k,j}|)\big\}, & \substack{\delta_{k-1,j}\delta_{k,j} > 0, \\ u_{k,j}\delta_{k-1,j} > 0}, \\ 0, & \text{otherwise.} \end{cases}$$

Zero secants, changes of secant sign, and candidate tangents with an inconsistent sign therefore yield zero interior tangents. For $K = 2$, both endpoints use the single secant. The bounds limit overshoot coordinatewise on each retained interval.

Reduction to cubic CFM. When $\rho = 0$, the implementation assigns $M_k(0) = m_k^{\mathrm{geom}}$ directly. These are the ordinary PCHIP tangents used by cubic CFM. With the same OT-linked knots, the reference path and its time derivative are therefore exactly the cubic-CFM reference and target velocity.

Hermite reference. For a tuple Z and projected tangents $M_k(\rho)$, let $\Delta_k = \tau_{k+1} - \tau_k$ and $\xi = (\tau - \tau_k)/\Delta_k$ on retained interval k. The Hermite reference is

$$H_Z^{\mathrm{TRACE}}(\tau) = h_{00}(\xi) z_k + h_{01}(\xi) z_{k+1} + \Delta_k\big[h_{10}(\xi) M_k(\rho) + h_{11}(\xi) M_{k+1}(\rho)\big],$$

where the scalar basis functions are

$$h_{00}(\xi) = 2\xi^3 - 3\xi^2 + 1, \quad h_{01}(\xi) = -2\xi^3 + 3\xi^2,$$
$$h_{10}(\xi) = \xi^3 - 2\xi^2 + \xi, \quad h_{11}(\xi) = \xi^3 - \xi^2.$$

Thus, the path interpolates the two cell states and their projected tangents, with interval duration scaling the tangent terms. We evaluate the training velocity by differentiating with respect to normalized biological time:

$$\frac{dH_Z^{\mathrm{TRACE}}}{d\tau} = \frac{h'_{00}(\xi) z_k + h'_{01}(\xi) z_{k+1}}{\Delta_k} + h'_{10}(\xi) M_k(\rho) + h'_{11}(\xi) M_{k+1}(\rho),$$

where primes denote differentiation with respect to ξ.

B.4 VELOCITY NETWORK AND OPTIMIZATION

Linear CFM, cubic CFM, and TRACE use the same velocity-network architecture and optimization budget. The velocity field is an MLP with three hidden layers of width 256 and SiLU nonlinearities.

Normalized biological time $\tau \in [0, 1]$ is encoded using eight Fourier frequencies giving the 16-dimensional embedding. These features are concatenated with the 30-dimensional latent state.

Within each optimization mini-batch, retained intervals receive approximately equal numbers of time samples. Conditional on interval k, time is sampled as $\tau \sim \mathrm{Uniform}(\tau_k, \tau_{k+1})$. Sampling is therefore approximately uniform across intervals and does not scale with biological duration. The reference path and its derivative still use the true irregular endpoints.

Each matched linear-CFM, cubic-CFM, and TRACE model is trained for 4,000 optimization steps using Adam with learning rate 10^{-3} and mini-batch size 128. At evaluation time, the complete population at the immediately preceding retained day is used as the source population. The learned velocity field is integrated from that source time to the held-out time using 50 Euler steps. A numerical sensitivity analysis using 20, 30, 40, and 50 Euler steps showed that held-out W_2 was stable at the higher integration resolutions. The resulting predictions are transformed from the standardized latent coordinates back to the common holdout-specific PCA space before scoring.

Table 10: Gupta TRACE training and evaluation hyperparameters. Linear CFM, cubic CFM, and TRACE share the neural-field and optimization settings.

Hyperparameter	Value
Input normalization	log-normalized expression
Expression scaling	retained-only z-score, clipped to $[-10, 10]$
PCA solver	full SVD
Zero-vector threshold	10^{-8}
Norm-calibration clip	$[0.25, 4]$
Shape-projection cap	$3\times$ local secant magnitude
PCA dimension	30
Latent scaling	retained-training z-score
Temporal coordinate	biological time normalized to $[0, 1]$
Velocity hidden layers	3
Gradient norm clipping	1.0
Velocity hidden width	256
Activation	SiLU
Optimization steps	4,000
Optimizer	Adam
Learning rate	10^{-3}
Mini-batch size	128
Training resamples	5
Random seeds	$0, 1, 2, 3, 4$
Euler steps	50
OT cost	squared Euclidean
OT solver	balanced EMD
OT conditional denominator ϵ	10^{-12}
Ridge coefficient λ	1.0
Norm-calibration denominator ϵ	10^{-12}
ρ candidates	$0, 0.25, 0.50, 0.75, 1.00$
Selected ρ	1.00
ρ selection holdout	Gupta day 7

B.5 MATCHED REGULATORY CONTROLS

Cell permutation. Regulon-activity rows are permuted once within each retained day for each matched run, preserving the day-level activity distribution while breaking the correspondence between a cell's latent state and its regulatory activity. For the reported fixed-TRACE control, the activity-to-latent map is the map fitted on the true retained activities and is held fixed after permutation.

TF permutation. Regulon-activity channels are globally permuted once per matched run, preserving activity values and dimensionality while assigning them to a different channel ordering. The activity-to-latent ridge map fitted on the *true* regulatory activities is held fixed; refitting after a global channel permutation would algebraically absorb the permutation.

These controls test dependence on cell-to-activity correspondence and on the fitted feature-to-map organization. Fixed-map TF permutation also tests corruption of the activity-to-map interface. Neither permutation establishes that the authentic TF-target graph is superior to a matched random gene-set or randomized-prior representation.

B.6 Population-level evaluation metrics

All methods are evaluated in the same 30-dimensional holdout-specific PCA space. Model rollout is performed using the full available source population, and the full held-out population is retained until metric computation.

For computational tractability and to keep all methods on an identical evaluation protocol, Wasserstein metrics use at most 1,500 predicted and 1,500 target cells. The subsets are sampled without replacement using the corresponding run seed.

We compute W_1 as exact balanced transport with Euclidean ground cost,

$$W_1(P,Q) = \min_{\Gamma \in \Pi(P,Q)} \sum_{ij} \Gamma_{ij} \|x_i - y_j\|_2,$$

where $P = n^{-1} \sum_{i=1}^{n} \delta_{x_i}$ and $Q = m^{-1} \sum_{j=1}^{m} \delta_{y_j}$ are the predicted and observed empirical distributions. We evaluate this cost using `ot.emd2`. We compute W_2 using squared Euclidean ground cost,

$$W_2(P,Q) = \left[\min_{\Gamma \in \Pi(P,Q)} \sum_{ij} \Gamma_{ij} \|x_i - y_j\|_2^2 \right]^{1/2}.$$

For maximum mean discrepancy (MMD), we subsample at most 1,000 predicted cells and 1,000 observed held-out cells per method and run, using the corresponding run seed (Gretton et al., 2012). To ensure that MMD values are directly comparable across methods, we use a single shared Gaussian-kernel bandwidth for all methods within a given holdout and run.

Let $\gamma_{h,r}$ denote the shared kernel parameter for holdout h and run r. We use the Gaussian kernel

$$k_{h,r}(x,y) = \exp\left(-\gamma_{h,r} \|x - y\|_2^2\right).$$

We report the empirical squared MMD, including the diagonal kernel terms, and label this quantity MMD in the tables:

$$\widehat{\mathrm{MMD}}^2(P,Q) = \frac{1}{n^2} \sum_{i,j=1}^{n} k_{h,r}(x_i, x_j) + \frac{1}{m^2} \sum_{i,j=1}^{m} k_{h,r}(y_i, y_j) - \frac{2}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} k_{h,r}(x_i, y_j),$$

where

$$P = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}, \qquad Q = \frac{1}{m} \sum_{j=1}^{m} \delta_{y_j}$$

are the predicted and observed empirical distributions, respectively. We use the same $\gamma_{h,r}$ for all methods within each holdout and run, ensuring that MMD values are directly comparable under a common kernel bandwidth.

For each holdout, we report the mean and population standard deviation (`ddof=0`) across the five training runs. Lower values indicate closer agreement between predicted and observed held-out populations.

B.7 BASELINE-SPECIFIC CONFIGURATIONS

We use a common held-out temporal evaluation protocol across methods. Predictions are compared in the same 30-dimensional holdout-specific PCA representation and rescored with the same W_1, W_2, and MMD evaluator. Where possible, method-specific settings from the authors' implementations are retained. In particular, MMFM and MMSFM use benchmark-specific training adaptations and ALI-CFM applies author-provided optimization schedule.

OT displacement. We compute balanced OT between the retained marginals bracketing the hidden time and use McCann displacement interpolation at the held-out biological-time timepoint.

Linear CFM and cubic CFM. These baselines use exactly the same MLP architecture, optimizer, training resamples, integration procedure, and OT-linked training tuples as TRACE. Linear CFM uses piecewise linear conditional paths. Cubic CFM uses the shape-preserving cubic-Hermite reference without regulatory tangent modulation. Model capacity and optimization budget are held fixed, so differences relative to TRACE isolate the conditional reference velocity.

MMFM. We use the authors' MMFM implementation, retaining its multi-marginal spline-based reference construction and flow-matching formulation. MMFM constructs cubic-spline trajectories through samples associated with multiple temporal marginals and regresses the resulting conditional dynamics with a neural velocity field. For compatibility with our common benchmark, MMFM receives the same holdout-specific 30-dimensional PCA representation, the hidden marginal is omitted, and rollout begins from the immediately preceding retained population. Five runs use seeds 0–4 and a fixed 4,000-step training budget. The 30-dimensional representation and 4,000-step budget are benchmark adaptations rather than the validation-selected configurations of the original MMFM experiments. Predictions are rescored with the common $W_1/W_2/\mathrm{MMD}$ evaluator.

MMSFM (SB). We use the authors' stochastic MMSFM implementation with diffusion scale $\sigma = 0.15$. MMSFM aligns consecutive snapshots using a first-order Markov approximation to a multi-marginal OT plan and constructs monotone cubic-Hermite transport splines over overlapping temporal windows. Drift and score networks are trained against the corresponding stochastic spline-path targets. Biological observation times are normalized to $[0, 1]$ while preserving irregular spacing. For the common benchmark, we use the holdout-specific 30-dimensional PCA representation and five runs with seeds 0–4 and 4,000 optimization steps; these dimensionality and training-budget choices are benchmark adaptations. Predictions are scored using the common W_1, W_2, and MMD evaluator.

moscot (balanced OT). We use `moscot.problems.time.TemporalProblem` with the common PCA coordinates. The temporal OT problems are solved with entropic regularization $\epsilon = 10^{-2}$, balanced marginals, and `scale_cost="mean"`. From each resulting transport matrix, we row-normalize to obtain conditional destination probabilities, sample paired cells, and apply displacement interpolation at the held-out biological-time fraction. Predictions are evaluated using the common W_1, W_2, and MMD evaluator.

ALI-CFM. We retain the authors' method-specific staged optimization without constraining it to TRACE's configuration. Five runs use the authors' single-cell seeds 42-46. For each temporal holdout, the interpolant is initialized for 2,000 pretraining steps, trained adversarially for 70,000 steps using the R3GAN configuration, and followed by 2,000 CFM-training steps. Interpolant, discriminator, and CFM networks use width 1,024 and batch size 128. Pretraining and CFM use learning rate 10^{-3} and generator and discriminator use learning rate 5×10^{-5}.