BEYOND DISTRIBUTION MATCHING: KINETIC CONSTRAINTS FOR UNPAIRED MULTI-OMICS ALIGNMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Unpaired static multi-omics data does not clearly answer which cells or molecular states correspond across modalities as distinct multi-omics mappings can produce the same target distribution with potentially biologically impossible states. We introduce Kinetics induced Optimal Transport (KOT), a framework that combines distribution matching with first-order constraints on the differential of the cross-modal mapping itself. KOT pushes an observed source vector field through the mapping's own Jacobian, requiring the induced target dynamics to be in agreement with a mechanistic target law, without the need for paired cells or observed target velocities. We show theoretically that this constraint reduces distribution-preserving alignments, that ignore the dynamics of single-cell biology, resulting in identifiability up to shared dynamical symmetries. For RNA–protein transport, biochemical anchors resolve a global kinetic-scale ambiguity. On CITE-seq datasets, KOT reaches competitive performance for cell-level RNA–protein alignment in an unpaired setting. Moreover, a mapping trained only on wild-type cells also transfers to out-of-distribution CRISPR perturbations. Mechanistic ablations show that state correspondence, internal kinetic consistency, and biological dynamics are distinct properties of KOT. Lastly, a chromatin–RNA extension exposes the boundaries of KOT. First-order constraints affect the learned differential, but cannot provide source-field-specific ambiguity resolution when the available chromatin directional field does not contain enough information. These experiments further establish first-order dynamical consistency as an inductive bias for learning unpaired cross-modal correspondences beyond static distribution matching. Code is provided in `https://anonymous.4open.science/r/KOT/README.md`

1 INTRODUCTION

Single-cell multi-omics (SCMO) measurements provide complementary views of cellular states from different modalities, but these views are often obtained from different cells at random states. In this unpaired setting, the correspondence between modalities is unknown and paired supervision is often difficult to obtain. Recovering SCMO correspondence is fundamentally underdetermined as matching populations does not determine which source state maps to which target state. In ribonucleic acid (RNA)-to-protein integration, e.g., multiple mappings can reproduce the observed protein distribution, while associations of individual transcriptomic states with biologically incompatible protein states are possible. Hence, the central challenge is not only to match cross-modal populations, but also to identify a mapping that is compatible with the molecular process (e.g. central dogma of molecular biology) underlying the target-space dynamical laws.

Unpaired SCMO integration algorithms tackle the mapping ambiguity by utilizing geometric and distributional structures such as SCOT, which leverages Gromov–Wasserstein optimal transport for alignment, while MOSCOT extends SCOT to space and time. Other methods such as GLUE use guidance graphs to integrate prior regulatory pathways and scDART utilizes cross modal gene-activity functions with a trajectory regularization to achieve SCMO alignment (Demetci et al., 2022; Klein et al., 2023; Cao et al., 2022; Cao & Gao, 2022; Zhang et al., 2022). Cellular dynamics provide an additional source of information about the cell state (Chen et al., 2025; Karbalayghareh et al., 2025). RNA velocity estimates a local vector field in transcriptomic state space from spliced and unspliced RNA counts (La Manno et al., 2018; Bergen et al., 2020). Protein velocity uses jointly

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.

measured RNA and protein to infer temporal relationships between cell states (Gorin et al., 2020), MultiVelo couples chromatin, transcription, and splicing dynamics in paired multi-omic data (Li et al., 2023). The prior works establish that dynamical information can be transferred or at least modeled across representations and modalities. Our setting is different as the target-modality velocity is *unobserved* while the source and target cells are also *unpaired*, and the cross-modal state map itself is unknown. We investigate whether single cell dynamics can identify the SCMO mapping rather than just describing dynamics after a SCMO correspondence has been established. We introduce **Kinetics induced Optimal Transport (KOT)**, combining distributional alignment, via Sinkhorn divergence to match the distribution of mapped RNA states to the observed protein population (Feydy et al., 2019), with a first-order mechanistic constraint on the cross-modal mapping. The Sinkhorn term constrains where source states are mapped, while the kinetic term constrains how the learned mapping transforms the local modality specific dynamics. Crucially, KOT does not require paired RNA–protein cells or empirical protein velocities which are unobserved. We use the observed source vector fields with a target-space mechanistic law, modeling the translation of source to target modality, in order to constrain the differential of a potentially ambiguous unpaired transport map. This construction is related to physics-informed learning through the use of differentiable dynamical residuals (Raissi et al., 2019), but the residual is used to identify a cross-modal correspondence rather than a directly observed state trajectory.

Our main contributions are fourfold: (1) We introduce a differentiable constraint that pushes an observed source-modality vector field through the Jacobian of a learned cross-modal map and matches the induced, unobserved target-modality dynamics to a mechanistic target-space law. (2) We characterize how the dynamical constraint reduces distributional alignment ambiguity, identify the shared dynamical symmetries, and prove that biochemical degradation-rate anchors select a global kinetic scale. (3) We show that KOT improves unpaired RNA–protein alignment and that the learned mapping generalizes to unseen CRISPR perturbations. (4) We show that the same dynamical ablations that removes or disrupts source dynamics, also degrade the prediction of out-of-distribution protein effects, while the Jacobian itself accurately maps intervention-induced RNA changes to downstream protein changes.

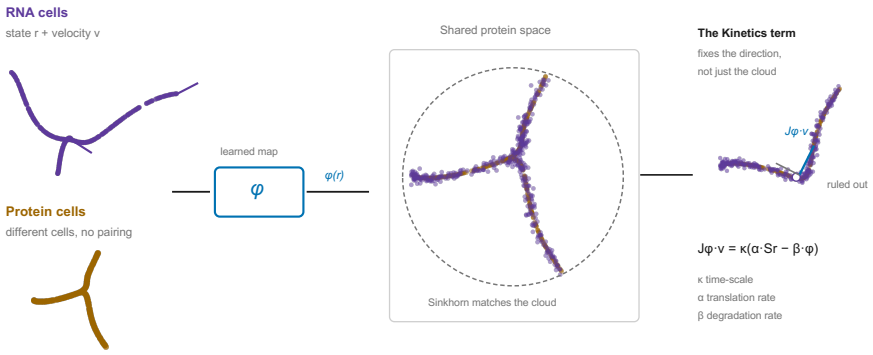


Figure 1: **Overview of the proposed Kinetics induced Optimal Transport model.** KOT constrains both where an unpaired mapping sends cells and how it transports their dynamics. RNA states r are observed together with a source vector field approximation v_r, while protein states are sampled from a separate, unpaired population. We learn a mapping ϕ, whose distribution matches the observed protein population via Sinkhorn divergence. The same map induces unobserved target-space dynamics through the Jacobian–vector product $J_\phi(r)v_r$, required to agree with the effective translation–degradation law $\kappa(r)\alpha(r) \odot Sr - \beta \odot \phi(r)]$.

2 METHOD

2.1 KINETIC ALIGNMENT OF UNPAIRED MODALITIES

KOT combines unpaired distribution matching with a first-order constraint requiring the differential of the cross-modal mapping to transport an observed source dynamics into a biologically plausible

(admissible) target dynamics. Let $\mathbf{x} \sim \mu_X$ and $\mathbf{y} \sim \mu_Y$ be the source and target states. We observe the unpaired samples $\{\mathbf{x}^{(i)}\}_{i=1}^N$ and $\{\mathbf{y}^{(j)}\}_{j=1}^M$ and empirical measures $\widehat{\mu}_X = \frac{1}{N}\sum_{i=1}^N \delta_{\mathbf{x}^{(i)}}$, $\widehat{\mu}_Y = \frac{1}{M}\sum_{j=1}^M \delta_{\mathbf{y}^{(j)}}$, where δ denotes a Dirac measure, and learn a continuously differentiable map $\phi_\theta : \mathcal{X} \to \mathcal{Y}$.

Assumption 2.1 (State sufficiency)**.** *On a manifold represented by the source modality, the corresponding target state is approximately described by a continuously differentiable function of the source state,* $\mathbf{y} \simeq \phi^\star(\mathbf{x})$*. The approximation might fail when unobserved lagged effects or relevant state variables are absent from the source modality.*

Distributional alignment. Let $\mathcal{I}_{\text{align}}$ denote target coordinates for population alignment and Π_{align} the corresponding projection. We match the observed target populations using the debiased Sinkhorn divergence (Feydy et al., 2019),

$$\mathcal{L}_{\text{align}} = \mathcal{S}_\varepsilon\left((\Pi_{\text{align}} \circ \phi_\theta)_{\#}\widehat{\mu}_X,\ (\Pi_{\text{align}})_{\#}\widehat{\mu}_Y\right). \tag{1}$$

Distribution matching alone is non-identifying since distinct mapping can induce the same target distribution while still assigning different cross-modal correspondences.

Jacobian push-forward of dynamics. Assume the source modality is further annotated by a vector field $\mathbf{v}_X(\mathbf{x}) \approx \dot{\mathbf{x}}$ describing the dynamics of x. Along a trajectory $\dot{\mathbf{x}} = \mathbf{v}_X(\mathbf{x})$, the chain rule provides the target-space velocity given by a candidate mapping, $\widehat{\mathbf{v}}_Y(\mathbf{x}) = \mathbf{J}_{\phi_\theta}(\mathbf{x})\mathbf{v}_X(\mathbf{x})$. Note, that we evaluate $\widehat{\mathbf{v}}_Y(\mathbf{x})$ directly as a Jacobian–vector product (JVP), without explicitly constructing the whole Jacobian. Further, let $f_\eta : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^{D_y}$ describe a mechanistic target-space law. KOT needs to fulfill:

$$\mathbf{J}_{\phi_\theta}(\mathbf{x})\mathbf{v}_X(\mathbf{x}) \approx f_\eta(\mathbf{x}, \phi_\theta(\mathbf{x})). \tag{2}$$

The empirical target velocities are not needed. Equation (Eq.) 2 compares target dynamics implied by the differential of the learned mapping with dynamics implied by the target law.

Dynamic consistency. Suppose $\widetilde{\mathbf{v}}_X^{(i)}$ denotes the preprocessed source velocity of cell i and let i's velocity be $\widehat{\mathbf{v}}_Y^{(i)} = \mathbf{J}_{\phi_\theta}(\mathbf{x}^{(i)})\widetilde{\mathbf{v}}_X^{(i)}$ and $\mathbf{v}_{\text{kin}}^{(i)} = f_\eta\big(\mathbf{x}^{(i)}, \phi_\theta(\mathbf{x}^{(i)})\big)$. Then, let $\mathcal{I}_{\text{kin}} \subseteq \mathcal{I}_{\text{align}}$ contain coordinates with a link between source and target and $c^{(i)} \geq 0$ be a cell-level velocity-confidence weight (=1 unless specified differently), and $s_j > 0$ a target scale. We minimize

$$\mathcal{L}_{\text{dyn}} = \frac{1}{N}\sum_{i=1}^N c^{(i)} \frac{1}{|\mathcal{I}_{\text{kin}}|} \sum_{j \in \mathcal{I}_{\text{kin}}} \left(\frac{\widehat{v}_{Y,j}^{(i)} - v_{\text{kin},j}^{(i)}}{s_j}\right)^2. \tag{3}$$

The marginal term regularizes where cells map to, while the kinetic term constrains how the derivative of the mapping transforms local dynamics from source to target.

The total optimization objective is given by $\mathcal{L}_{\text{KOT}}^{(t)} = \mathcal{L}_{\text{align}} + \lambda_{\text{dyn}}^{(t)}\mathcal{L}_{\text{dyn}} + \mathcal{R}(\Theta)$, where $\mathcal{R}$ contains modality-specific biochemical priors, weak kinetic-scale regularization, and network regularization. $\lambda_{\text{dyn}}^{(t)}$ controls the strength of kinetic consistency. All details of hyperparameters are given in Appendix A.6 while figure 1 summarizes the KOT workflow.

Notice that Eq. 2 is modality-agnostic. Cross-modal translation enters KOT via target law f_η. Appendix A.13 provide an application for chromatin to unspliced RNA to spliced RNA.

2.2 RNA-TO-PROTEIN KINETIC TRANSPORT

For RNA-to-protein alignment, $\mathbf{x} = \mathbf{r}$ and $\phi_\theta(\mathbf{r}) = \widehat{\mathbf{p}}$. Let $\mathbf{S} \in \mathbb{R}^{D_p \times D_r}$ be a sparse correspondence matrix indicating the transcripts encoding the measured proteins. Protein production and degradation are modeled via $f_{\text{prot}}(\mathbf{r}, \mathbf{p}) = \kappa_\psi(\mathbf{r})\left[\boldsymbol{\alpha}_\omega(\mathbf{r}) \odot (\mathbf{S}\mathbf{r}) - \boldsymbol{\beta} \odot \mathbf{p}\right]$, where $\boldsymbol{\alpha}_\omega(\mathbf{r}) > 0$ is a state-dependent production rate, $\boldsymbol{\beta} > 0$ a vector of protein degradation rates, and $\kappa_\psi(\mathbf{r}) > 0$ the RNA-velocity time scale. Substitution into Eq. 2 gives

$$\mathbf{J}_{\phi_\theta}(\mathbf{r})\mathbf{v}_r(\mathbf{r}) \approx \kappa_\psi(\mathbf{r})\left[\boldsymbol{\alpha}_\omega(\mathbf{r}) \odot (\mathbf{S}\mathbf{r}) - \boldsymbol{\beta} \odot \phi_\theta(\mathbf{r})\right]. \tag{4}$$

KOT constrains an unpaired RNA–protein mapping using RNA dynamics without requiring paired cells or unobserved protein velocities.

Biochemical rate calibration. For a protein h, a measured half-life $t_{1/2,h}$ implies $\beta_h^{\text{phys}} = \frac{\log 2}{t_{1/2,h}}$. Since KOT operates in an inferred velocity time scale, physical rates are converted by one dataset-level factor a_β, such that $\beta_h^\star = a_\beta \beta_h^{\text{phys}}$ preserves the relative structure. We treat these measures as a soft log-rate prior $\mathcal{L}_\beta = \frac{1}{2}\big(\sum_{h\in\mathcal{H}} w_h \left(\log \beta_h - \log \beta_h^\star\right)/\tau_h\big)^2\big)/(\sum_{h\in\mathcal{H}} w_h)$ with protein half-lives from (Mathieson et al., 2018). The construction of a_β can be found in Appendix A.9. Additionally, We use the log-scale regularizer $\mathcal{L}_\kappa = \frac{1}{N}\sum_{i=1}^N \left(\log(\kappa_\psi(\mathbf{r}^{(i)})/\kappa_0)\right)^2$.

In the RNA-protein setting, we define $\mathcal{R}(\Theta) = \rho_\beta \mathcal{L}_\beta + \rho_\kappa \mathcal{L}_\kappa + \rho_{\text{net}} \sum_{\vartheta\in\Theta_{\text{net}}} \|\vartheta\|_2^2$. For synthetic data generated directly from f_{prot}, the kinetic variables can be directly interpreted from the ordinary differential equation (ODE). In the case of normalized CITE-seq measurements, we interpret f_{prot} as an effective translation–degradation vector field in the measurement coordinates.

2.3 SYMMETRY CHARACTERIZATION FROM FIRST-ORDER INFORMATION

We analyze KOT at the population level considering the exact distributional and dynamical constraints. The results relate to the identifiability of the cross-modal mapping and its global kinetic scale. They do not imply recovery in finite-samples or unique identification of every kinetic parameter.

Definition 2.2 (Admissibility)**.** *A continuously differentiable $\phi : \mathcal{X} \to \mathcal{Y}$ is* distributionally admissible *if $\phi_\# \mu_X = \mu_Y$, and* dynamically admissible *if*

$$\mathbf{J}_\phi(\mathbf{x})\mathbf{v}_X(\mathbf{x}) = f_\eta(\mathbf{x}, \phi(\mathbf{x})) \qquad \text{, for } \mu_X \text{ - almost every } \mathbf{x} \tag{5}$$

Lemma 2.3 (Static distributional ambiguity)**.** *If ϕ is distributionally admissible and $T : \mathcal{Y} \to \mathcal{Y}$ satisfies*

$$T_\# \mu_Y = \mu_Y, \tag{6}$$

then $T \circ \phi$ is also distributionally admissible.

Theorem 2.4 (Dynamic compatibility of target-modality symmetries)**.** *Let ϕ be distributionally and dynamically admissible and let T be continuously differentiable with $T_\#\mu_Y = \mu_Y$. Then $\widetilde{\phi} = T \circ \phi$ is dynamically admissible if and only if*

$$\mathbf{J}_T(\phi(\mathbf{x}))f_\eta(\mathbf{x}, \phi(\mathbf{x})) = f_\eta(\mathbf{x}, T(\phi(\mathbf{x}))) \qquad \mu_X\text{-a.e.} \tag{7}$$

KOT circumvents target transformations that preserve the observed marginal but do not preserve the underlying dynamics.

Corollary 2.5 ((Uniqueness within a target-modality symmetry orbit))**.** *Let $\mathcal{O}(\phi) = \{T \circ \phi : T_\#\mu_Y = \mu_Y\}$. The mappings satisfying both KOT constraints are generated by distribution preserving transformations satisfying Eq. 7. If the identity is the only transformation that satisfies both KOT constraints, ϕ is unique within $\mathcal{O}(\phi)$.*

Corollary 2.6 (Broken dynamical symmetry)**.** *Let*

$$\mathbf{\Delta}_T(\mathbf{x}) = \mathbf{J}_T(\phi(\mathbf{x}))f_\eta(\mathbf{x}, \phi(\mathbf{x})) - f_\eta(\mathbf{x}, T(\phi(\mathbf{x}))). \tag{8}$$

If ϕ satisfies the dynamical constraint, then

$$\mathcal{E}_{\text{dyn}}(T \circ \phi) = \mathbb{E}_{\mathbf{x}\sim\mu_X}\left[\|\mathbf{\Delta}_T(\mathbf{x})\|_2^2\right]. \tag{9}$$

Thus, if $\|\mathbf{\Delta}_T(\mathbf{x})\|_2 \geq \delta$ on a set $\mathcal{A}$ with $\mu_X(\mathcal{A}) \geq q$, then $\mathcal{E}_{\text{dyn}}(T \circ \phi) \geq q\delta^2$.

Theorem 2.7 (Global kinetic-scale and anchor calibration)**.** *For a RNA-to-protein field, with*

$$\kappa_a(\mathbf{r}) = a\kappa(\mathbf{r}), \qquad \boldsymbol{\alpha}_a(\mathbf{r}) = a^{-1}\boldsymbol{\alpha}(\mathbf{r}), \qquad \boldsymbol{\beta}_a = a^{-1}\boldsymbol{\beta}, \tag{10}$$

f_{prot} stays constant for every $a > 0$. The kinetic residual itself cannot identify the global rate scale. For any non-empty anchor set with prior in $\mathcal{L}_\beta$ and $w_h > 0$ and $\tau_h > 0$, the anchor loss is strictly convex in $z = \log a$ along this gauge orbit and therefore selects a unique global scale.

Theorem 2.7 only affects the global rate-scale symmetry without implying unique identification of $\boldsymbol{\alpha}_\omega$, κ_ψ, or the individual kinetic parameters. Proofs of Lemma 2.3, Theorem 2.4, Corollary 2.6, and Theorem 2.7 are given in Appendix A.1.

3 EXPERIMENTAL SETTING

3.1 DATASETS AND EVALUATION TASKS

KOT is evaluated on four regimes to investigate distributional alignment, first-order identifiability, perturbational generalization, and transfer to another molecular process. A mechanistically generated dataset provides known cross-modal correspondences and ground-truth dynamics (see Appendix A.2.1 for more details). Two CITE-seq datasets (PBMC Hao et al. (2021) and BMMC Luecken et al. (2021)) evaluate unpaired RNA–protein alignment in real cells (see Appendix A.2.2). A CRISPR screen (Papalexi et al., 2021), (details in Appendix A.2.4) tests whether a mapping trained only on unperturbed cells transfers to unseen interventions. A paired ATAC–RNA multiome setting from BMMC are treated as unpaired to test whether the KOT construction generalizes from translation–degradation to a transcription–splicing–degradation relay. Unless explicitly stated otherwise, known cell pairing is used only for evaluation and not training. Detailed dataset descriptions and task definitions are provided in Appendices A.2 and A.3.

3.2 IMPLEMENTATION DETAILS

KOT is implemented in PyTorch (Paszke et al., 2019) with simple MLP networks for ϕ_θ, α_ω, and κ_ψ. In RNA–protein experiments, source dynamics are inferred with scVelo Bergen et al. (2020) in the same spliced-RNA coordinates from the RNA state. Velocities are confidence weighted and normalized by a dataset-level scale. The Sinkhorn, kinetic, biochemical-prior, and network-regularization terms are optimized jointly with Adam with independently sampled source and target minibatches for the alignment. Hyperparameters are selected on development splits and frozen for the reported multi-seed evaluations. Unless stated otherwise 12 seeds were used. Preprocessing, architectures, optimization schedules, baseline settings, splits, anchors, and experiment-specific configurations are detailed in Appendix A.6.

3.3 RESULTS

Table 1: RNA-protein alignment performance on unpaired synthetic and CITE-seq datasets. Synthetic experiments are of increasing difficulty. TotalVI is a paired reference. Fraction of samples closer than the true match (FOSCTTM) is used for evaluation and lower is better. FOSCTTM is reported as mean ± standard deviation over 12 seeds unless the method is deterministic (SCOT, moscot and linear ODE).

		Synthetic (FOSCTTM ↓)			**CITE-seq (FOSCTTM ↓)**	
	Method	Oracle	Clean	Branch	BMMC	PBMC
KOT (Ours)	KOT	**0.030**±0.002	**0.039**±0.001	**0.026**±0.003	**0.129**±0.004	**0.135**±0.006
	KOT ($\lambda_{\text{dyn}} = 0$)	0.037±0.002	0.051±0.004	0.388±0.002	0.588±0.061	0.443±0.165
Unpaired baselines	GLUE	–	0.424±0.007	–	0.392±0.094	0.429±0.100
	Linear ODE	0.444	0.446	–	0.442	0.251
	MaxFuse	–	–	–	0.183±0.002	0.226±0.003
	SCOT	0.174	0.515	–	0.402	0.400
	moscot	0.544	0.480	–	0.337	0.211
	uniPort	0.480±0.122	0.499±0.006	–	0.495±0.087	0.472±0.082
References	totalVI (paired)	–	0.000±0.000	0.000±0.000	–	0.000±0.000
	KOT kinetic oracle	0.020±0.001	–	–	–	–

Kinetic constraints resolve ambiguous correspondences and improve real alignment. First, we test the main premise of KOT, which is whether first-order dynamics can resolve ambiguous correspondences inherent in static manifold matching. In the synthetic Branch benchmark, exchanging the two post-bifurcation branches preserves the target distribution by design. Therefore, removing the kinetic constraint doesn't change the same marginal objective but increases FOSCTTM from 0.026 ± 0.003 to 0.388 ± 0.002 (Table 1). But KOT recovers the branch correspondence, providing a controlled realization of the distribution-preserving ambiguity characterized in Sec. 2.3. The performance on the Oracle and Clean settings show similar empirical evidence.

The same constraint improves alignment in measured cells. On BMMC and PBMC, KOT shows FOSCTTM of 0.129 ± 0.004 and 0.135 ± 0.006, compared to 0.183 ± 0.002 (MaxFuse on BMMC and 0.226 ± 0.003 in PBMC) and 0.211 (moscot on PBMC and 0.337 on BMMC) for the strongest evaluated unpaired competitor. Removing the dynamics within the same KOT setting worsens FOSCTTM to 0.588 ± 0.061 and 0.443 ± 0.165, respectively. Hence, the first-order constraint that resolves the constructed synthetic ambiguity generalizes to improving unpaired correspondence in measured real single-cell data.

Table 2: Performance on out-of-distribution (unseen) CRISPR protein-effect prediction from observed perturbed RNA data. KOT is trained on non-targeting (NT) cells only, with the RNA–protein pairings staying hidden. KOT is then frozen before evaluation on unseen CRISPR perturbations. The standard KOT prediction uses a MLP, while KOT (local Jacobian) uses a first-order response defined in Appendix A.3. We evaluate with Spearman rank correlation (ρ) and Pearson correlation (r) comparing predicted and observed perturbation-level protein effects. Additionally normalized root-mean-square error (NRMSE) measures the effect-size of the errors. Higher correlations and lower NRMSE are better. Note, that paired NT translators class models are trained on supervised paired RNA–protein controls, and MultiPert receives stronger perturbational supervision as well.

Method	Supervision	Spearman ↑	Pearson ↑	NRMSE ↓
Unpaired KOT (Ours)				
KOT	unpaired NT	0.426±0.027	0.594±0.033	0.918±0.015
KOT (local Jaccobian)	unpaired NT	0.410±0.015	0.615±0.022	0.938±0.018
KOT ($\lambda_{dyn} = 0$)	unpaired NT	-0.158±0.042	-0.366±0.091	1.674±0.090
KOT (v shuffled)	unpaired NT	-0.064±0.118	0.006±0.095	1.139±0.046
KOT ($v \to -v$)	unpaired NT	0.167±0.027	0.351±0.076	0.963±0.042
KOT ($v \to 0$)	unpaired NT	-0.220±0.032	-0.168±0.076	1.404±0.077
KOT (S perturbed)	unpaired NT	0.448±0.025	0.619±0.027	0.908±0.017
Paired NT translators				
Ridge	paired NT	0.493	0.702	0.826
MLP	paired NT	0.465±0.047	0.582±0.066	6.050±0.866
scButterfly	paired NT	0.338±0.084	0.528±0.070	0.918±0.042
sciPENN	paired NT	0.401±0.049	0.704±0.036	0.755±0.030
References / additional supervision				
Cognate mRNA	reference	0.415	0.610	0.985
Zero change	reference	–	–	1.067
Paired multipert.	paired perturbation	-0.107	-0.033	1.221

The learned differential retains predictive signal on unseen CRISPR perturbation predictions. Next, we test whether the first-order structure learned from observational cells is useful for out of distribution tasks on unseen crispr perturbation predictions. KOT is trained only on non-targeting (NT) Papalexi cells, with RNA–protein pairing hidden and without observing knockout protein measurements. The model is then frozen before zero-shot perturbation evaluation (Appendix A.3).

The frozen map, trained only on wild type cells, predicts unseen protein effects with Spearman 0.426 ± 0.027 (Table 2). The local first-order response obtained from the learned Jacobian retains nearly the same signal (Spearman 0.410 ± 0.015, Pearson 0.615 ± 0.022), directly linking the differential constrained during KOT training to out-of-distribution intervention effects. In contrast, removing the kinetic objective or shuffling the source velocity reduces Spearman to -0.158 ± 0.042 and -0.064 ± 0.118, respectively. That leads us to conclude, that the learned first-order structure remains predictive beyond observational population on which the mapping was fitted on. A fully in-silico extension, in which the perturbed RNA state itself must be predicted, can be found in Appendix A.12.

3.4 MECHANISTIC AND BOUNDARY ANALYSES

3.4.1 KOT MECHANISM

The mechanistic ablations show that static alignment, satisfaction of the kinetic equation, and biological correctness of the induced dynamics are distinct quantities (Table 3). Several corrupted vector

Table 3: Mechanistic ablation of KOT, distinguishing cell-level correspondence, biological dynamics, and internal kinetic consistency. BMMC and PBMC are evaluated. We evaluate via Fraction of samples closer than the true match (FOSCTTM) measuring cell-level cross-modal correspondence, static ρ measuring the per-protein Spearman correlation between predicted and observed protein states, biological dynamics r which re-scores the fitted effective dynamics using the inferred biological RNA velocity and the correct gene–protein correspondence and the Jacobian–vector-product to right-hand-side agreement (JVP–RHS) measuring internal consistency between $J_\phi v$ and the mechanistic kinetic field. Higher correlations and lower FOSCTTM and residual are better.

Dataset	Condition	FOSCTTM ↓	Static ρ ↑	Bio. dyn. r ↑	JVP–RHS ↑	Residual ↓
BMMC	KOT	0.129±0.004	0.369±0.009	0.832±0.011	0.952±0.003	0.180±0.007
	KOT (v shufled)	0.122±0.002	0.396±0.006	0.640±0.019	0.899±0.008	0.257±0.008
	KOT ($v \to -v$)	0.126±0.003	0.373±0.006	-0.736±0.017	0.911±0.002	0.235±0.004
	KOT ($v \to 0$)	0.136±0.003	0.350±0.006	0.110±0.014	0.000±0.000	1.000±0.000
	KOT (S perturbed)	0.510±0.090	-0.017±0.055	0.813±0.019	0.952±0.005	0.180±0.009
	KOT ($\lambda_{dyn} = 0$)	0.588±0.061	-0.067±0.047	0.179±0.045	0.064±0.043	0.996±0.001
PBMC	KOT	0.135±0.006	0.328±0.011	0.640±0.040	0.728±0.029	0.433±0.020
	KOT (v shufled)	0.138±0.008	0.320±0.021	0.330±0.052	0.488±0.042	0.613±0.021
	KOT $v \to -v$	0.142±0.010	0.318±0.017	-0.535±0.040	0.631±0.041	0.502±0.026
	KOT ($v \to 0$)	0.139±0.010	0.322±0.011	0.093±0.039	0.000±0.000	1.000±0.000
	KOT (S pertubed)	0.428±0.151	0.058±0.108	0.612±0.026	0.735±0.053	0.435±0.030
	KOT ($\lambda_{dyn} = 0$)	0.443±0.165	0.029±0.092	0.167±0.073	0.064±0.074	0.991±0.001

fields preserve strong static alignment. E.g. on BMMC shuffling, reversing, or zeroing the source velocity has little effect on static correspondence as on BMMC these variants give FOSCTTM between 0.122 and 0.136, close to 0.129 for the observed source field, and the corresponding PBMC values are similarly close. In contrast, perturbing the transcript–protein correspondence S substantially degrades state correspondence, increasing FOSCTTM to 0.510 on BMMC and 0.428 on PBMC. Thus, in the measured CITE-seq experiments, the static correspondence gain is substantially more sensitive to the feature-level mechanistic link than to the orientation of the supplied source vector field.

In contrast, a model can satisfy its kinetic equation while representing the wrong biological direction. Reversing the source velocity retains high JVP–RHS cosine (0.911 ± 0.002 on BMMC and 0.631 ± 0.041 on PBMC), while the independently re-scored biological dynamics become negative (-0.736 ± 0.017 and -0.535 ± 0.040). We conclude, that a small training residual or high JVP–RHS agreement can be seen as evidence of internal consistency, but not evidence that the imposed dynamics are biologically faithful.

Shuffling the source velocity provides a complementary analysis. Cell-level correspondence stays almost constant but biological dynamic agreement reduces from 0.832 to 0.640 on BMMC and from 0.640 to 0.330 on PBMC. Zero velocity removes first-order information entirely and reduces biological agreement toward zero. These results show that the dynamic diagnostics detect information not captured by FOSCTTM alone.

The ablations also expose limitations of single diagnostics. Permuting the multi-omics correspondence matrix S preserves JVP–RHS and the external dynamic score while diminishing state correspondence and protein reconstruction, suggesting that dynamic agreement does not identify the correct cross-modal feature correspondence by itself. On the other hand, removing the kinetic objective produces poor state alignment and weak biological dynamics despite retaining the same mapping architecture and marginal objective. Consequently, the evidence for KOT comes from the conjunction of distributional correspondence and first-order consistency rather than from either metric in isolation.

3.4.2 Chromatin–RNA transfer exposes a boundary of kinetic disambiguation

Lastly, we analyze whether KOT transfers to the distinct ordering chromatin → nascent RNA → mature RNA. Unlike the RNA–protein experiments, BMMC Multiome provides no measured chromatin velocity. The source field is therefore an ATAC-derived directional pseudotime derivative, while independently inferred RNA velocity is reserved entirely for post-training evaluation. Across 12 seeds, adding the chromatin kinetic constraint does not improve static cell-level correspondence. Spliced-RNA FOSCTTM (s-FOSCTTM) is 0.2210 ± 0.0009 without dynamics and

0.2225 ± 0.0008 for the full KOT. The kinetic term influences the learned differential. Agreement of the pushed-forward ATAC field with independently inferred spliced-RNA velocity increases from 0.0054 ± 0.0077 without dynamics to 0.0210 ± 0.0039 with KOT, and JVP–RHS agreement increases from -0.0345 ± 0.0048 to 0.0647 ± 0.0085 (see Table 4).

However, the first-order signal is largely preserved when the source directional field is shuffled as JVP–scVelo remains at 0.0200 ± 0.0046, with FOSCTTM 0.2223 ± 0.0010. Therefore, the BMMC chromatin experiment does not provide the same evidence for source-field-specific ambiguity resolution as the RNA–protein experiments. Instead, the kinetic objective alters the learned differential in a direction more consistent with independent RNA dynamics, although this effect is not specific to the biological ordering of the present ATAC field.

The failure analysis clarifies this boundary condition. Chromatin representations suitable for static transport are not inherently suitable for first-order transport. Among the screened source coordinates, TF–IDF/LSI is the only representation that provides positive absolute agreement with the independent RNA-velocity field. However, the same representation for the transcription-rate head, drives around half of the production rates to their lower bound. Linear CP10K regulatory features counteracts the degeneracy while maintaining the differential signal. The geometry used to define trajectories, the coordinates differentiated by the map, and the features utilized by the mechanistic rate model do not serve the same role.

Table 4: Performance of the chromatin–RNA alignment on BMMC Multiome. KOT treats the experimentally paired chromatin-accessibility and RNA measurements as unpaired during training. We evaluate the performance via spliced-RNA fraction of samples closer than the true match (s-FOSCTTM) for static correspondence, JVP–scVelo as the median centred cosine similarity between the chromatin-field Jacobian–vector product (JVP) and independently inferred spliced-RNA velocity, as well as Δ_{null}, measuring te observed JVP–scVelo score minus the 95th percentile of its permutation null. Lastly, JVP–RHS measures agreement with the mechanistic right-hand side (RHS) is used.. Target RNA velocity is used only for post-training evaluation. Values are mean $\pm$ standard deviation over 12 seeds.

KOT variant	s-FOSCTTM ↓	JVP–scVelo ↑	Δ_{null} ↑	JVP–RHS ↑
KOT	0.2225 ± 0.0008	0.0210 ± 0.0039	$+0.0014 \pm 0.0009$	0.0647 ± 0.0085
KOT ($\lambda_{\text{dyn}} = 0$)	0.2210 ± 0.0009	0.0054 ± 0.0077	-0.0001 ± 0.0014	-0.0345 ± 0.0048
KOT (shuffled v_z)	0.2223 ± 0.0010	0.0200 ± 0.0046	$+0.0009 \pm 0.0010$	0.0558 ± 0.0074

This extension is not a static alignment win. scGLUE and MaxFuse obtain lower FOSCTTM in their respective alignment spaces, so we interpret the chromatin experiment primarily as a stress test of first-order transfer rather than state-of-the-art correspondence.

Additional ablations show that stronger internal kinetic fits can be misleading as removing the residual gate increase the raw JVP–RNA cosine but simultaneously also increases its permutation null above the observed value, whereas omitting the unspliced auxiliary constraint removes the positive null margin. These diagnostics motivates us to report both the absolute first-order score as well as its permutation null rather than internal consistency alone. More details and further chromatin ablations are provided in Appendix A.13.1 while Appendix A.13 contains the details for the analysis.

4 Discussion

KOT is motivated by a simple distinction between zeroth- and first-order information in unpaired multi-omics alignment. Matching marginal distributions constrains, where the source population is sent via OT, but not determine which source state corresponds to which target state. The kinetic constraint adds information about the differential of a mapping by requiring observed source directions to be transported into a biologically admissible target field. The population-level analysis makes the distinction formally explicit since distribution-preserving transformations are only admissible when they simultaneously preserve the target modalities dynamics. The synthetic Branch experiments provides a controlled realization of the symmetry-breaking mechanism of KOT, while the real data results show that the inductive bias can improve correspondence in measured data. We therefore do

not see KOT as a replacement for distributional alignment, but more a way to restrict the family of distributionally plausible mapping using information that static snapshots do not utilize.

Moreover, the CRISPR experiment provides a complementary analysis of what is exactly learned by KOT. The RNA–protein map is fitted only on wild type (non-targeting) cells and frozen before unseen interventions are predicted. Both the multi-modal mapping and its local Jacobian maintain predictive signals for perturbation-induced protein effects. In particular, the first-order response of the frozen map performs close to the full prediction, directly connecting the differential regularized during training to the out-of-distribution prediction behavior. Importantly, the perturbation experiments should not be interpreted as unique identification of a complete biochemical mechanism as perturbing the transcript–protein correspondence matrix S does not reduce the performance, which indicates that the intervention results validate the usefulness of the learned differential more directly than the recovery of the exact gene-specific kinetic decomposition. In this sense, the mechanistic law acts as a structured inductive bias on the mapping, while individual kinetic components require separate interpretation and calibration.

The ablation studies also clarify when first-order consistency is informative. Internal agreement between the mapping's JVP and the fitted kinetic right-hand side does not establish biological faithfulness by itself but a corrupted source field can be fitted consistently, while a corrupted feature-level mechanism can preserve dynamical agreement but inhibit state correspondence. The chromatin extension further exposes the same limitation in a molecular relay. There, the kinetic objective alters the learned differential in a direction consistent with independent RNA dynamics, although shuffling the ATAC directional field maintains almost the same signal strength. Hence, we can conclude that first-order constraints only disambiguate when the supplied source field and mechanistic representation contain information that breaks the static symmetry. This boundary is consistent with the theory rather than an exception to it since KOT reduces distributional alternatives to the extent that they fail to preserve the imposed dynamical law only.

Several limitations follow directly from the above perspective. The analysis assumes that the relevant target state is approximately a differentiable function of observed source state. But unobserved hysteresis, or missing state variables can break this assumption. The source vector field by itself is estimated and may contain systematic error, and the mechanistic law is obviously an approximation in normalized single-cell measurement coordinates instead of a mathematical certainty. Furthermore, the RNA–protein kinetic decomposition contains a multiplicative gauge as biochemical anchors select a rate scale along a direction but do not uniquely identify α, κ, and individual degradation rates. Finally, the theoretical results are population-level statements under exact constraints and do not guarantee finite-sample recovery. Natural extensions include uncertainty-aware source fields, richer latent-state or history-dependent mappings, and mechanistic laws that combine an arbitrary amount of complementary first-order and second-order signals.

5 Conclusion

We introduced Kinetics induced Optimal Transport, which augments unpaired distribution matching with a first-order constraint on the differential of the cross-modal mapping. KOT uses an observed source vector field and a mechanistic target law to find distributionally valid correspondences that transport local dynamics, without requiring observed target velocity. Theory characterizes the remaining shared dynamical symmetries, while synthetic, RNA-protein, and CRISPR experiments suggest that the resulting differential can improve correspondence and the transfer to out-of-distribution interventions. The mechanistic and chromatin analyses further show that kinetic consistency is informative only when the first-order structure is sufficiently specific. Altogether, the results establish dynamical consistency as a principled complement to static distribution matching for unpaired cross-modal learning.

REFERENCES

John Aitchison. The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):139–160, 1982.

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.

Kai Cao, Yiguang Hong, and Lin Wan. Manifold alignment for heterogeneous single-cell multi-omics data integration using pamona. *Bioinformatics*, 38(1):211–219, 2022.

Yichuan Cao, Xiamiao Zhao, Songming Tang, Qun Jiang, Sijie Li, Siyu Li, and Shengquan Chen. scbutterfly: a versatile single-cell cross-modality translation method via dual-aligned variational autoencoders. *Nature Communications*, 15(1):2973, 2024.

Zhi-Jie Cao and Ge Gao. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nature biotechnology*, 40(10):1458–1466, 2022.

Yuhao Chen, Yan Zhang, Jiaqi Gan, Ke Ni, Ming Chen, Ivet Bahar, and Jianhua Xing. Graphvelo allows for accurate inference of multimodal velocities and molecular mechanisms for single cells. *Nature Communications*, 16(1):7831, 2025.

Haotian Cui, Chloe Wang, Hassaan Maan, Kuan Pang, Fengning Luo, Nan Duan, and Bo Wang. scgpt: toward building a foundation model for single-cell multi-omics using generative ai. *Nature methods*, 21(8):1470–1480, 2024.

Pinar Demetci, Rebecca Santorella, Björn Sandstede, William Stafford Noble, and Ritambhara Singh. Scot: single-cell multi-omics alignment with optimal transport. *Journal of computational biology*, 29(1):3–18, 2022.

Stefan Elfwing, Eiji Uchibe, and Kenji Doya. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural networks*, 107:3–11, 2018.

Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-ichi Amari, Alain Trouvé, and Gabriel Peyré. Interpolating between optimal transport and mmd using sinkhorn divergences. In *The 22nd international conference on artificial intelligence and statistics*, pp. 2681–2690. PMLR, 2019.

Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249–256. JMLR Workshop and Conference Proceedings, 2010.

Gennady Gorin, Valentine Svensson, and Lior Pachter. Protein velocity and acceleration from single-cell multiomics experiments. *Genome biology*, 21(1):39, 2020.

Yuhan Hao, Stephanie Hao, Erica Andersen-Nissen, William M Mauck, Shiwei Zheng, Andrew Butler, Maddie J Lee, Aaron J Wilk, Charlotte Darby, Michael Zager, et al. Integrated analysis of multimodal single-cell data. *Cell*, 184(13):3573–3587, 2021.

Alireza Karbalayghareh, Benedikt Pelzer, Christopher R Chin, Ari Melnick, Darko Barisic, and Christina S Leslie. Deep dynamical models of single-cell multiomic velocities predict loss-of-function and rescue perturbations in b cells. *bioRxiv*, pp. 2025–04, 2025.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Dominik Klein, Giovanni Palla, Marius Lange, Michal Klein, Zoe Piran, Manuel Gander, Laetitia Meng-Papaxanthos, Michael Sterr, Aimée Bastidas-Ponce, Marta Tarquis-Medina, et al. Mapping cells through time and space with moscot. biorxiv. *2023*, 2005(2011):540374, 2023.

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.

Justin Lakkis, Amelia Schroeder, Kenong Su, Michelle YY Lee, Alexander C Bashore, Muredach P Reilly, and Mingyao Li. A multi-use deep learning method for cite-seq and single-cell rna-seq data integration with cell surface protein prediction and imputation. *Nature machine intelligence*, 4(11):940–952, 2022.

Chen Li, Maria C Virgilio, Kathleen L Collins, and Joshua D Welch. Multi-omic single-cell velocity models epigenome–transcriptome interactions and improves cell fate prediction. *Nature biotechnology*, 41(3):387–398, 2023.

Jie Liu, Yuanhao Huang, Ritambhara Singh, Jean-Philippe Vert, and William Stafford Noble. Jointly embedding multiple single-cell omics measurements. In *Algorithms in bioinformatics:... International Workshop, WABI..., proceedings. WABI (Workshop)*, volume 143, pp. 10, 2019.

Malte D Luecken, Daniel Bernard Burkhardt, Robrecht Cannoodt, Christopher Lance, Aditi Agrawal, Hananeh Aliee, Ann T Chen, Louise Deconinck, Angela M Detweiler, Alejandro A Granados, et al. A sandbox for prediction and integration of dna, rna, and proteins in single cells. In *Thirty-fifth conference on neural information processing systems datasets and benchmarks track (Round 2)*, 2021.

Toby Mathieson, Holger Franken, Jan Kosinski, Nils Kurzawa, Nico Zinn, Gavain Sweetman, Daniel Poeckel, Vikram S Ratnu, Maike Schramm, Isabelle Becher, et al. Systematic analysis of protein turnover in primary cells. *Nature communications*, 9(1):689, 2018.

Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral Normalization for Generative Adversarial Networks. In *International Conference on Learning Representations*, 2018. URL https://mlanthology.org/iclr/2018/miyato2018iclr-spectral/.

Efthymia Papalexi, Eleni P Mimitou, Andrew W Butler, Samantha Foster, Bernadette Bracken, William M Mauck III, Hans-Hermann Wessels, Yuhan Hao, Bertrand Z Yeung, Peter Smibert, et al. Characterizing the molecular regulation of inhibitory immune checkpoints with multimodal single-cell screens. *Nature genetics*, 53(3):322–331, 2021.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

Yusuf Roohani, Kexin Huang, and Jure Leskovec. Predicting transcriptional outcomes of novel multigene perturbations with gears. *Nature biotechnology*, 42(6):927–935, 2024.

Hiroaki Sakoe and Seibi Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1):43–49, 1978.

Tim Stuart, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M Mauck, Yuhan Hao, Marlon Stoeckius, Peter Smibert, and Rahul Satija. Comprehensive integration of single-cell data. *cell*, 177(7):1888–1902, 2019.

Weixu Wang, Zhiyuan Hu, Philipp Weiler, Sarah Mayes, Marius Lange, Daniel M Fountain, Julianna O Haug, Jingye Wang, Zhengyuan Xue, Tatjana Sauka-Spengler, et al. Regvelo: gene-regulatory-informed dynamics of single cells. *Cell*, 189(12):3773–3800, 2026.

Ziqi Zhang, Chengkai Yang, and Xiuwei Zhang. scdart: integrating unmatched scrna-seq and scatac-seq data and learning cross-modality relationship simultaneously. *Genome biology*, 23(1):139, 2022.

Mengyuan Zhao, Xinyue Tang, Jiawei Li, Cheng Liang, Jijun Tang, and Fei Guo. Multipert: An adversarial alignment and dual attention framework for single-cell multi-omics perturbation prediction. *PLOS Computational Biology*, 22(3):e1014054, 2026.

AI USE STATEMENT

ChatGPT has been used during the preparation and revision of this manuscript. It was used for writing and aid in the communication of the research. We used ChatGPT for:

- Condensing the text and message,
- Language and grammar improvement,
- General scientific communication, especially in communicating the novelty, contributions and relationships internal between the main body and the appendix,
- And proof reading.

The assistance was used to clear inconsistencies within the presentation and communicating the message mainly. Throughout the main manuscript and the appendix, we utilized LLMs for helping with the writing. The usage of LLM concerns only the presentation of the results.

All AI/LLM-assisted text and suggestions were reviewed by the authors. Suggested edits, consistency issues, and scientific statements were checked against the underlying derivations, implementations, experimental outputs, and cited literature as appropriate. The authors take full responsibility for the methodology, mathematical results, experiments, claims, citations, and final contents of the paper.

A Appendix

A.1 Theoretical Proofs

A.1.1 Proof of Lemma 2.3

Proof. Push-forward measures compose, hence

$$(T \circ \phi)_{\#}\mu_X = T_{\#}\left(\phi_{\#}\mu_X\right) = T_{\#}\mu_Y = \mu_Y.$$

□

A.1.2 Proof of Theorem 2.4

Proof. By the chain rule,

$$\mathbf{J}_{\widetilde{\phi}}(\mathbf{x})\mathbf{v}_X(\mathbf{x}) = \mathbf{J}_T(\phi(\mathbf{x}))\mathbf{J}_\phi(\mathbf{x})\mathbf{v}_X(\mathbf{x}). \tag{11}$$

Because ϕ is dynamically admissible,

$$\mathbf{J}_\phi(\mathbf{x})\mathbf{v}_X(\mathbf{x}) = f_\eta(\mathbf{x}, \phi(\mathbf{x}))$$

for μ_X-almost every $\mathbf{x}$. Hence

$$\mathbf{J}_{\widetilde{\phi}}(\mathbf{x})\mathbf{v}_X(\mathbf{x}) = \mathbf{J}_T(\phi(\mathbf{x}))f_\eta(\mathbf{x}, \phi(\mathbf{x})) \qquad \mu_X\text{-a.e.} \tag{12}$$

The transformed map $\widetilde{\phi} = T \circ \phi$ is dynamically admissible if and only if

$$\mathbf{J}_T(\phi(\mathbf{x}))f_\eta(\mathbf{x}, \phi(\mathbf{x})) = f_\eta(\mathbf{x}, T(\phi(\mathbf{x})))$$

for μ_X-almost every $\mathbf{x}$, which is precisely Eq. 7. □

A.1.3 Proof of Corollary 2.6

Proof. Theorem 2.4 and Eq. 12 imply that the dynamical residual of $T \circ \phi$ is exactly $\mathbf{\Delta}_T(\mathbf{x})$. Therefore

$$E_{\mathrm{dyn}}(T \circ \phi) = \mathbb{E}_{\mathbf{x}\sim\mu_X}\left[\|\mathbf{\Delta}_T(\mathbf{x})\|_2^2\right],$$

which gives Eq. 9. If $\|\mathbf{\Delta}_T(\mathbf{x})\|_2 \geq \delta$ on a measurable set $\mathcal{A}$ with $\mu_X(\mathcal{A}) \geq q$, then

$$E_{\mathrm{dyn}}(T \circ \phi) \geq \int_{\mathcal{A}} \|\mathbf{\Delta}_T(\mathbf{x})\|_2^2 \, \mathrm{d}\mu_X(\mathbf{x}) \geq \delta^2 \mu_X(\mathcal{A}) \geq q\delta^2.$$

□

A.1.4 Proof of Theorem 2.7

Proof. Substituting Eq. 10 gives

$$f_{\mathrm{prot}}^{(a)}(\mathbf{r}, \mathbf{p}) = a\kappa(\mathbf{r})\left[a^{-1}\boldsymbol{\alpha}(\mathbf{r}) \odot (\mathbf{S}\mathbf{r}) - a^{-1}\boldsymbol{\beta} \odot \mathbf{p}\right] \tag{13}$$

$$= f_{\mathrm{prot}}(\mathbf{r}, \mathbf{p}), \tag{14}$$

which proves the kinetic-scale invariance.

Let $z = \log a$ and define

$$d_h = \log \beta_h - \log \beta_h^\star.$$

Assume $\mathcal{H} \neq \varnothing$, with $w_h > 0$ and $\tau_h > 0$ for every active anchor $h \in \mathcal{H}$. Under the gauge transformation,

$$\log \beta_{a,h} - \log \beta_h^\star = d_h - z.$$

Hence the anchor loss along the gauge orbit is

$$\mathcal{L}_\beta(z) = \frac{1}{2} \frac{\sum_{h \in \mathcal{H}} \frac{w_h}{\tau_h^2} (d_h - z)^2}{\sum_{h \in \mathcal{H}} w_h}. \tag{15}$$

Its second derivative is

$$\frac{\mathrm{d}^2 \mathcal{L}_\beta}{\mathrm{d}z^2} = \frac{\sum_{h \in \mathcal{H}} w_h / \tau_h^2}{\sum_{h \in \mathcal{H}} w_h} > 0. \tag{16}$$

Therefore $\mathcal{L}_\beta(z)$ is strictly convex and possesses a unique minimizer,

$$z^\star = \frac{\sum_{h \in \mathcal{H}} \frac{w_h}{\tau_h^2} (\log \beta_h - \log \beta_h^\star)}{\sum_{h \in \mathcal{H}} w_h / \tau_h^2}. \tag{17}$$

Thus the soft biochemical anchors remove the flat global-scale direction by selecting a unique point along the gauge orbit. □

A.2 Dataset Details

Table 5: **Datasets and evaluation regimes.** Paired measurements are treated as unpaired during KOT training unless explicitly marked otherwise.

Dataset	Modalities	Scale	Primary task
Synthetic linked ODE	RNA → protein	5,000 cells/stage	Identifiability
BMMC CITE-seq	RNA + 134 ADTs	90,261 cells	RNA–protein alignment
PBMC CITE-seq	RNA + 32 ADTs	5,527 cells	RNA–protein alignment
Papalexi ECCITE-seq	RNA + 4 ADTs + CRISPR	20,729 cells	CRISPR perturbation
BMMC Multiome	ATAC + RNA	69,249 cells	Chromatin–RNA transport

A.2.1 Synthetic linked-ODE benchmark

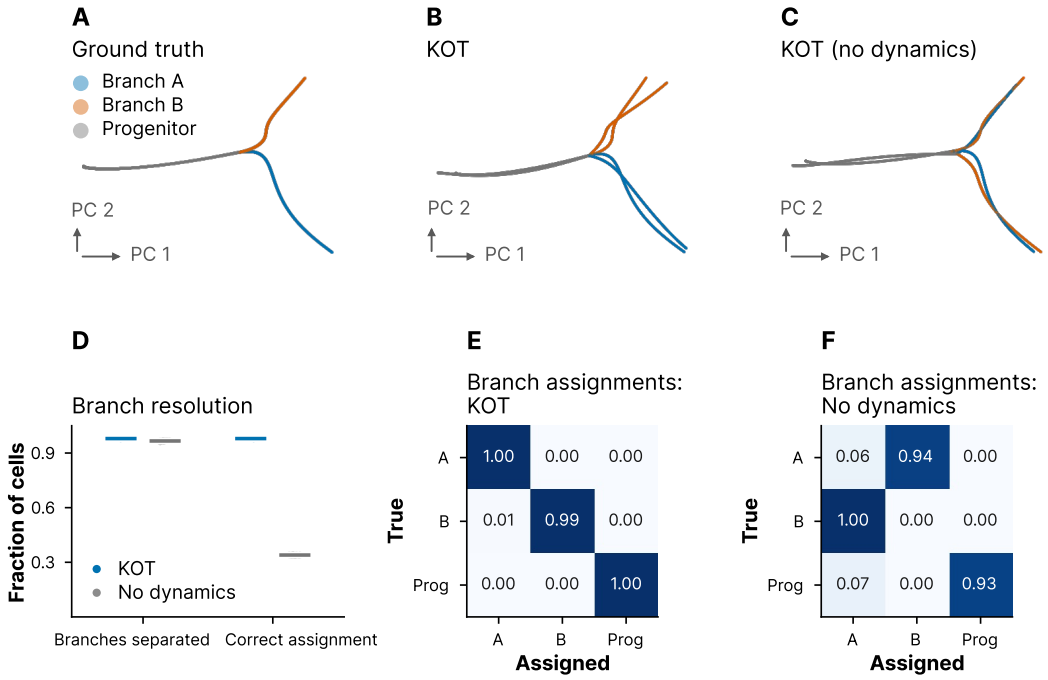


Figure 2: **Dynamics resolve a branch correspondence that is ambiguous to static distribution matching.** (A) Ground-truth protein states for the synthetic Branch benchmark, colored by the progenitor population and the two post-bifurcation branches. (B–C) Cross-modal states recovered with the kinetic constraint and with distributional alignment alone. (D) Branch-resolution summary across 12 seeds. (E–F) Confusion matrices between the true and recovered branch identities for KOT with and without kinetic constraints. Cell-level pairing is hidden during training and is only used for evaluation.

We generate paired RNA and protein states directly from the translation–degradation process in f_{prot}, creating ground-truth state correspondences, source velocities, and kinetic variables. Each stage contains 5,000 cells, 12 RNA features, and 10 linked protein features. We consider three stages of increasing difficulty. The *Oracle* contains a single noiseless trajectory with fixed kinetic coefficients.. The *Clean* version retains a single trajectory and introduces observation noise as well as state-dependent production and clock rates. Lastly, the *Branch* introduces a mirror-symmetric bifurcation constructed such that swapping the two post-bifurcation branches preserves the pooled static target distribution. The resulting branch correspondence is therefore ambiguous for distribution matching but distinguishable by the sign and direction of the dynamics. Although the simulator generates paired cells, all unpaired methods are trained after discarding without ground truth correspondence. The pairing is only used for evaluation.

Figure 2 visualizes the ambiguity built into the Branch benchmark. Exchanging the two post-bifurcation branches leaves the pooled target geometry almost unchanged. The alignment objective can recover a plausible target distribution while assigning the branches incorrectly. The source vector field breaks this symmetry, hence, requiring its push-forward to agree with the target kinetic law recovers the correct branch identity across seeds. The benchmark therefore separates successful marginal matching from recovery of the underlying cross-modal correspondence.

The per-cell diagnostics further show that these evaluation quantities capture different properties of the recovered map (Figure 3). FOSCTTM is strongly related to temporal-order error and moderately related to the directional agreement between the pushed-forward field and the kinetic law, but it is not a proxy for every dynamical diagnostic. In particular, the magnitude-based ODE residual can vary almost independently of per-cell retrieval error. We therefore report state correspondence and first-order kinetic diagnostics separately throughout the paper.

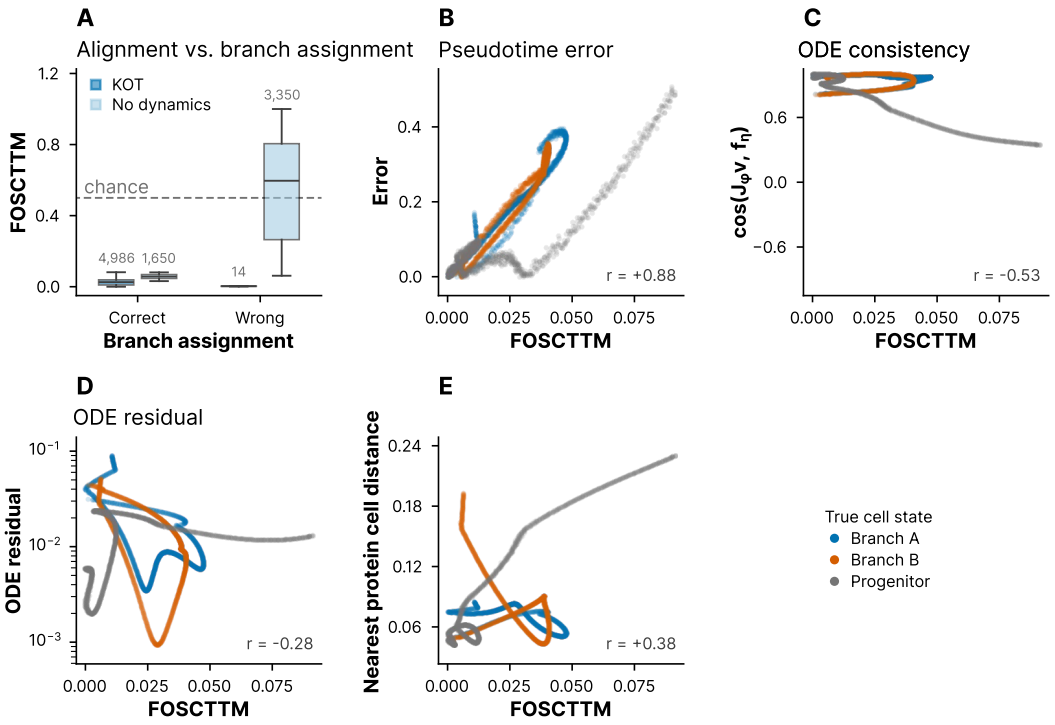


Figure 3: **Cell-level correspondence error is related temporal and kinetic diagnostics on the synthetic benchmark.** (A) FOSCTTM stratified by recovered branch assignment. (B) Correspondence error association with pseudotime error. (C) JVP–RHS directional agreement associated with correspondence error. (D) Magnitude-based ODE residual associated with per-cell FOSCTTM. (E) Correspondence error associated with distance to the nearest observed protein state. Reported r values are Pearson correlations with per-cell FOSCTTM.

A.2.2 RNA–PROTEIN CITE-SEQ DATASETS

We evaluate real-data RNA–protein alignment on two paired CITE-seq datasets, while withholding their cell-level correspondence during training.

BMMC. We use the NeurIPS 2021 multimodal bone-marrow mononuclear-cell dataset (GSE194122), comprising 90,261 cells from 12 healthy donors collected across four experimental sites (Luecken et al., 2021). The processed CITE-seq object contains 13,953 gene-expression features and 134 antibody-derived tags (ADTs). The multi-site, multi-donor design provides a heterogeneous hematopoietic population with substantial biological and technical variation. We retain the transcripts corresponding to measured ADTs during RNA preprocessing so that the mechanistic RNA–protein links remain represented even when they are not selected as highly variable genes.

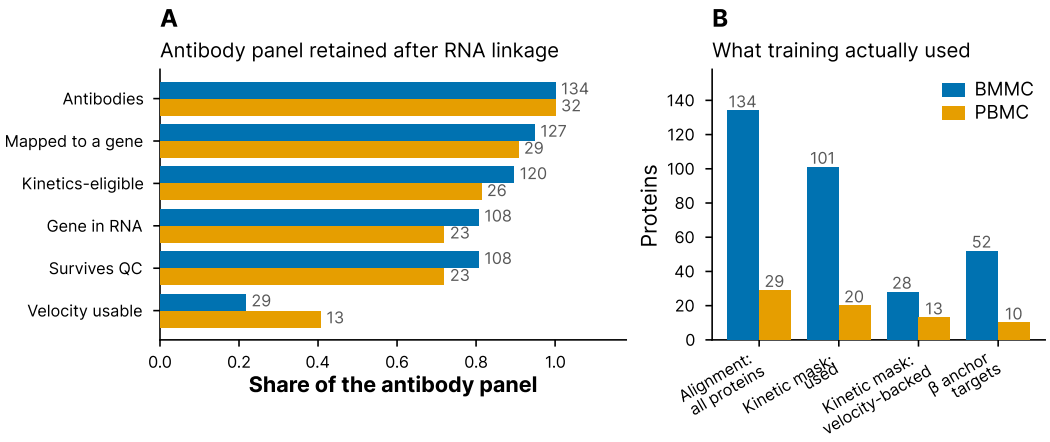


Figure 4: **Measured antibody panels and features used for RNA–protein kinetic constraint.** (A) Attrition of the BMMC and PBMC antibody panels as proteins are linked to encoding transcripts. (B) Protein subsets used by different parts of KOT.

PBMC. As a second RNA–protein benchmark, we use the 10x Genomics 5k PBMC CITE-seq dataset (Hao et al., 2021), containing 5,527 peripheral-blood mononuclear cells from a healthy donor measured with gene expression and a 32-protein TotalSeq-B antibody panel. This dataset is substantially smaller than BMMC and provides a complementary test on a conventional peripheral-immune-cell mixture.

For both datasets, the RNA state supplied to KOT is the smoothed spliced-RNA state used by the velocity model, and the corresponding source vector field is the inferred RNA velocity in the same coordinates. Protein states are the normalized ADT measurements. The known RNA–protein barcode correspondence is removed from KOT and from all unpaired baselines and is used only to compute cell-level retrieval metrics. Protein degradation measurements are used only through the soft biochemical calibration described in Sec. 2.

Not every measured ADT can contribute to the first-order constraint (Figure 4). Population alignment uses the retained protein panel, whereas the kinetic and first-order evaluation subsets are progressively restricted by the availability of a gene–protein link, retention of the encoding transcript in the processed RNA state, and support from the fitted RNA-velocity model. Biochemical half-life anchors form a separate subset and provide rate calibration only.

A.2.3 BMMC CHROMATIN–RNA MULTIOME DATASET

We use the paired 10x Multiome component of the BMMC benchmark (GSE194122), comprising 69,249 bone-marrow mononuclear cells with joint chromatin-accessibility and RNA measurements. The data originate from the same multi-donor, multi-site study as the BMMC CITE-seq benchmark. Although ATAC and RNA are experimentally paired, these correspondences are removed during KOT training and retained only for evaluation. The construction of the ATAC-derived source field and RNA measurement coordinates is described in Appendix A.13.

A.2.4 PAPALEXI ECCITE-SEQ CRISPR PERTURBATION DATASET

Papalexi ECCITE-seq perturbation dataset. For perturbational evaluation we use the pooled ECCITE-seq screen of (Papalexi et al., 2021), which combines CRISPR perturbations with single-cell transcriptome and surface-protein measurements in THP-1 cells (GSE153056). The pooled experiment contains 20,729 cells across three viral-transduction replicates and 25 targeted-gene classes in addition to non-targeting controls. We evaluate the four surface proteins retained in our processed panel: CD86, PD-L1 (CD274), PD-L2 (PDCD1LG2), and TIM-3/CD366 (HAVCR2).

Before any model evaluation, we freeze an eligibility rule requiring at least 20 cells for a perturbation in at least two biological replicates. Our primary evaluation contains all sufficiently sampled *non-*

self perturbation–protein effects. Self-effects are analyzed separately because perturbing the gene encoding the assayed protein directly targets that gene product and therefore does not test the same cross-gene regulatory transfer as the primary non-self effects. After this filtering, the primary pooled benchmark contains 94 perturbation–protein effects

A.3 EVALUATION TASKS

Unpaired cross-modal alignment. For synthetic data, BMMC CITE-seq, and PBMC CITE-seq, the known cell correspondence is withheld during optimization. Each method receives its method-specific admissible inputs but no cell-level correspondence. KOT additionally receives the source vector field and specified target kinetic law, but never an empirical target-modality velocity. After training, the withheld pairing is used to evaluate cell-level correspondence using FOSCTTM. For RNA–protein KOT we additionally report static protein reconstruction, agreement of the fitted effective dynamics with the biological source field, and internal agreement between the Jacobian push-forward and the mechanistic right-hand side.

Protein-effect prediction from observed perturbed RNA. Here, we test whether the RNA–protein map learned in an observational control population transfers to interventions. All preprocessing used by KOT, feature selection, normalization, neighborhood construction, RNA-velocity fitting, velocity confidence, and scaling statistics, is fitted using non-targeting (NT) cells only. KOT is then trained only on the NT RNA and protein populations, with their cell-level pairing hidden, and without protein half-life anchors. All model parameters are frozen before any knockout cell is evaluated.

At test time, the observed RNA state of a knockout cell is passed through the frozen map. For perturbation g in replicate b, the nonlinear protein-effect prediction is

$$\Delta\widehat{p}^{\,\phi}_{g,b} = \frac{1}{|K_{g,b}|}\sum_{i\in K_{g,b}}\phi(r_i) - \frac{1}{|N_b|}\sum_{i\in N_b}\phi(r_i),$$

where $K_{g,b}$ and N_b denote knockout and NT cells, respectively. We compare this quantity with the corresponding observed ADT effect $\Delta p_{g,b}$ and aggregate replicate effects with equal weight.

As a complementary first-order test, we also evaluate the local response predicted by the differential of the frozen map,

$$\Delta\widehat{p}^{\,J}_{g,b} = J_\phi(\bar{r}_{N_b})\left(\bar{r}_{K_{g,b}} - \bar{r}_{N_b}\right).$$

This directly tests whether the differential constrained during observational training transfers to an intervention that was never seen during fitting.

KOT therefore receives neither knockout protein measurements nor paired RNA–protein controls. For context, we additionally evaluate RNA–protein translators fitted on *paired* NT cells (ridge regression, an MLP, sciPENN (Lakkis et al., 2022), and scButterfly (Cao et al., 2024)). These methods receive **strictly more supervision** and are reported as a separate comparison group rather than as like-for-like unpaired baselines.

Fully in-silico perturbation prediction. In-silico perturbation prediction (ISP) removes access to the experimentally observed knockout RNA while retaining the same held-out ECCITE-seq protein effects for evaluation. Starting from NT RNA and a perturbation identity g, an upstream ISP model predicts a counterfactual knockout RNA state,

$$\widehat{r}_{\mathrm{KO},g} = \Psi(r_{\mathrm{NT}}, g),$$

which is then passed through the same frozen NT-trained KOT map,

$$\widehat{r}_{\mathrm{KO},g} \xrightarrow{\ \phi\ } \widehat{p}_{\mathrm{KO},g}.$$

We evaluate the resulting perturbation-level protein effects against the same observed ECCITE-seq effects used in protein effect prediction.

We consider a linear perturbation model, GEARS (Roohani et al., 2024), RegVelo, scGPT (Cui et al., 2024), and MultiPert (Zhao et al., 2026). Because these models do not share a common supervision regime, the ISP task is treated as an exploratory end-to-end benchmark rather than a like-for-like model ranking. In particular, the linear model predicts a held-out replicate of a perturbation seen in other replicates; RegVelo is fitted using NT cells and represents supported regulator knockouts mechanistically. GEARS and scGPT are trained or fine-tuned using observed perturbed RNA and MultiPert receives still stronger perturbational supervision. We therefore report the supervision regime with every ISP comparison.

Unpaired chromatin–RNA kinetic transport. KOT maps chromatin state to a joint unspliced–spliced RNA state using the relay law in Appendix A.13. The final configuration uses ATAC LSI for trajectory geometry, TF–IDF/LSI-derived gene activity for the map and JVP, and linear CP10K gene activity for the transcription-rate head. Unspliced and spliced RNA are represented using a common per-cell normalization followed by the corresponding log-coordinate transformation. The spliced-RNA marginal defines the primary Sinkhorn alignment term, while the unspliced block is constrained only through unpaired per-gene moments.

Most importantly, RNA velocity inferred from the target RNA modality is never used to train the chromatin model. After training, we compare the spliced component of $J_\Phi(z)v_z$ with independently inferred spliced-RNA velocity, providing an external first-order evaluation of the learned correspondence.

Unlike RNA velocity in the RNA–protein experiments, BMMC provides no measured chromatin velocity. We therefore construct an ATAC-derived directional trajectory field: diffusion pseudotime and forward neighborhoods are estimated from ATAC LSI geometry, and forward displacements are expressed in the chromatin coordinates on which the learned map is differentiated. We consequently interpret this field as a directional pseudotime derivative, not as a physical chromatin velocity.

For the final BMMC model we use TF–IDF/LSI map coordinates, linear CP10K regulatory features, shared-log RNA coordinates, 32 alignment dimensions, and $\lambda_{\mathrm{dyn}} = 30$; all representation and kinetic-weight sensitivity experiments are reported in Appendix A.13.1.

A.4 ADDITIONAL ALIGNMENT DIAGNOSTICS.

The qualitative embeddings in Figure 5 provide a complement to FOSCTTM. Several methods can produce apparently well-mixed RNA–protein populations in two dimensions, despite large differences in cell-level retrieval accuracy. Thus visual modality mixing is not sufficient evidence of a correct correspondence. KOT is evaluated primarily using the withheld cell pairing and the independent first-order diagnostics described above.

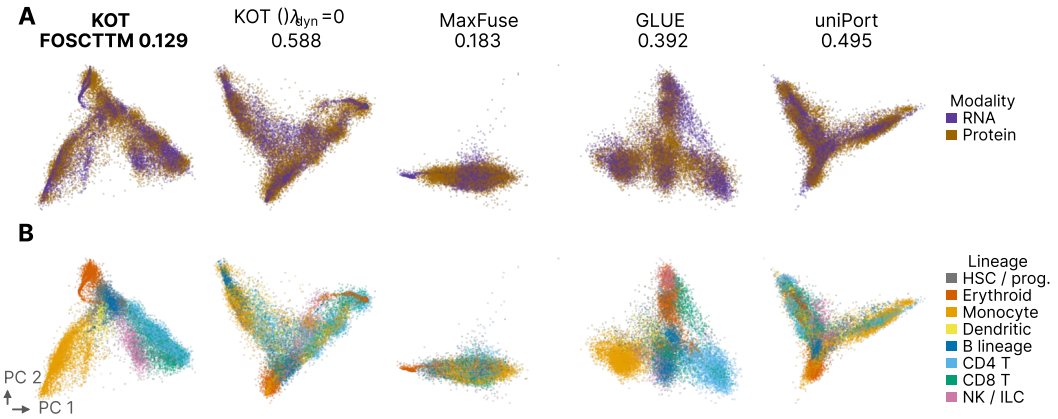


Figure 5: **Joint-embedding comparison for unpaired RNA–protein alignment on the BMMC.** (A) Joint embeddings colored by modality. (B) Embeddings colored by biological lineage.

Figure 6 expands the aggregate benchmark by stratifying retrieval error across biological lineages. The improvement is not restricted to a single large population, although the absolute difficulty varies substantially across hematopoietic compartments. The accompanying co-embedding should be interpreted only as a visualization of population structure. The quantitative correspondence metric remains the withheld-pair FOSCTTM.

Rank-based retrieval also contains information beyond exact nearest-neighbor recovery. Figure 7a reports the fraction of true partners recovered as the accepted target neighborhood is enlarged, providing a view of how concentrated the learned correspondence is around the withheld match. Figure 7b reports empirical runtime on the BMMC and PBMC benchmark runs, because these measurements were collected on mixed hardware, they are used only to document computational order of magnitude and not to rank method efficiency.

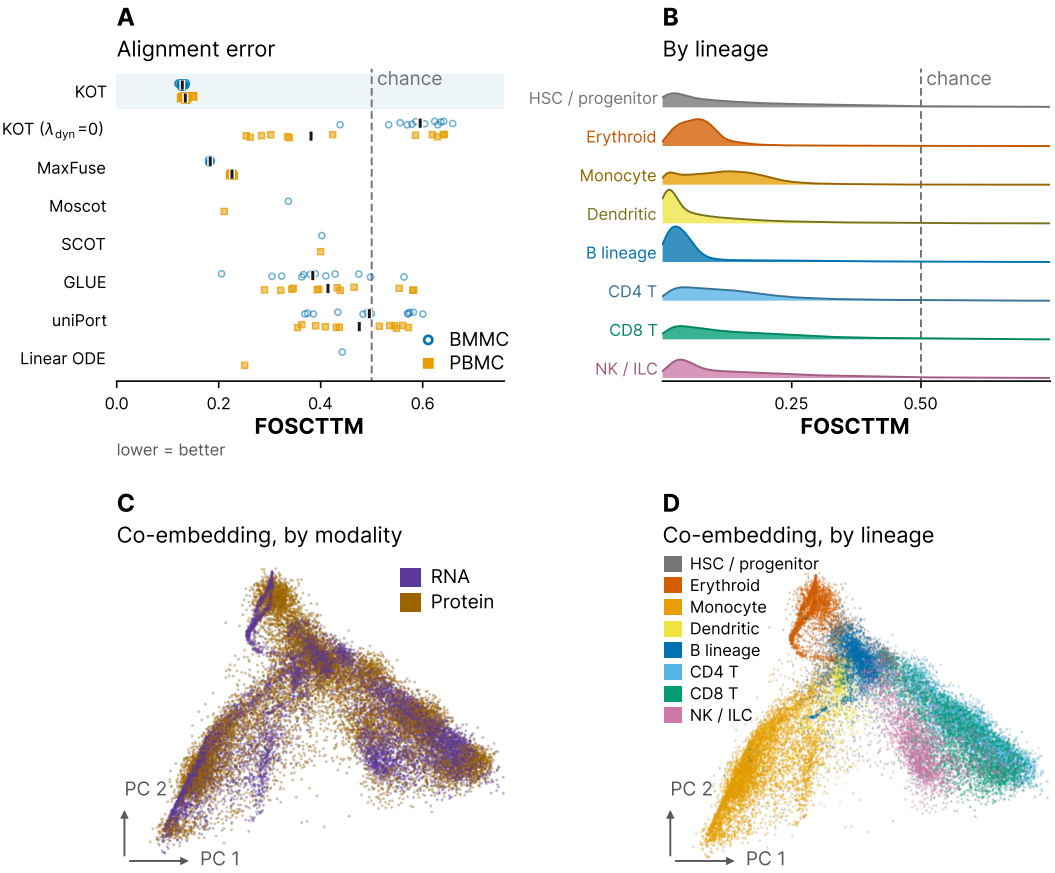


Figure 6: **Alignment accuracy across methods and biological lineages.** (A) FOSCTTM on the BMMC and PBMC. (B) Lineage-stratified FOSCTTM. (C–D) Joint embedding of mapped RNA and observed protein states colored by modality and by lineage.

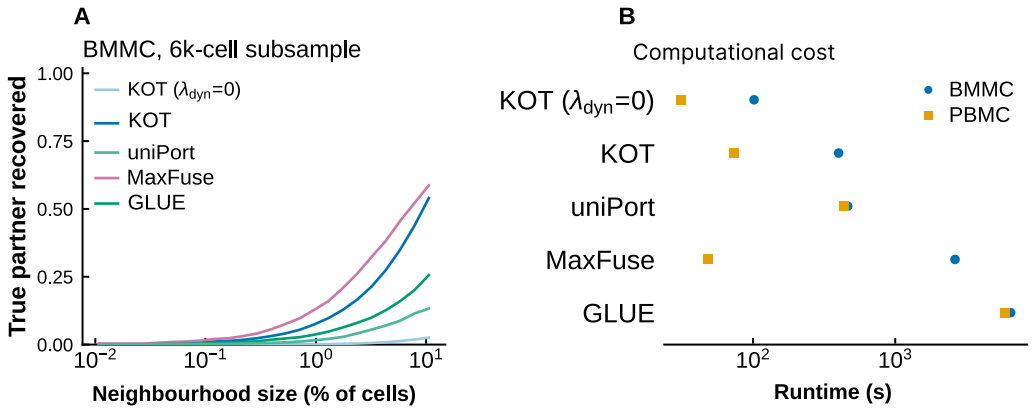


Figure 7: **Neighborhood retrieval and computational cost.** (A) Fraction of withheld true RNA–protein partners recovered from an increasing target-neighborhood size on a 6,000-cell BMMC subsample. (B) Empirical runtime on the BMMC and PBMC benchmark.

A.5 EVALUATION METRICS

Cell-level cross-modal correspondence. Our primary alignment metric is the fraction of samples closer than the true match (FOSCTTM), introduced for paired evaluation of single-cell multi-omic alignments by (Liu et al., 2019) and subsequently used in cross-modal integration benchmarks such as SCOT (Demetci et al., 2022). Known pairing never enters the training loss or gradient updates.

Let $\hat{\mathbf{y}}_i$ denote the mapped source state and $\mathbf{y}_i$ its paired observed target. For the eligible comparison set $\mathcal{C}_i$, either the complete evaluated population or the corresponding biological batch, we compute the two directional ranks

$$F_i^{\hat{Y}\to Y} = \frac{1}{|\mathcal{C}_i| - 1} \sum_{\substack{j \in \mathcal{C}_i \\ j \neq i}} \mathbf{1}[d(\hat{\mathbf{y}}_i, \mathbf{y}_j) < d(\hat{\mathbf{y}}_i, \mathbf{y}_i)] \,,$$

and

$$F_i^{Y\to \hat{Y}} = \frac{1}{|\mathcal{C}_i| - 1} \sum_{\substack{j \in \mathcal{C}_i \\ j \neq i}} \mathbf{1}[d(\mathbf{y}_i, \hat{\mathbf{y}}_j) < d(\mathbf{y}_i, \hat{\mathbf{y}}_i)] \,,$$

using Euclidean distance in the evaluation space. The reported per-cell score is

$$\mathrm{FOSCTTM}_i = \frac{1}{2}\left(F_i^{\hat{Y}\to Y} + F_i^{Y\to\hat{Y}}\right),$$

and we average over evaluated cells. Lower values indicate better cell-level retrieval, while random correspondence gives values near 0.5. Where indicated, ranking is performed separately within biological batches. For chromatin–RNA experiments, s-FOSCTTM denotes the same statistic evaluated on the predicted spliced-RNA block.

Static target reconstruction. For RNA–protein experiments, we additionally assess whether mapped protein values covary with their paired observed measurements. For each evaluated protein h, we compute

$$\rho_h^{\mathrm{static}} = \mathrm{Spearman}_i\left(\hat{p}_{ih}, p_{ih}\right)$$

across evaluated cells and report the median across proteins. This metric measures feature-wise static reconstruction and is distinct from cell-level retrieval. A map can preserve population-level protein trends without recovering the correct cell correspondence.

Internal and external first-order diagnostics. For each evaluated cell, let

$$\mathbf{a}_i = J_\phi(\mathbf{x}_i)\mathbf{v}_X(\mathbf{x}_i), \qquad \mathbf{b}_i = f_\eta(\mathbf{x}_i, \phi(\mathbf{x}_i)) \,,$$

restricted to coordinates participating in the kinetic law. Internal JVP–RHS agreement is the median cell-wise cosine similarity

$$\text{JVP-RHS} = \mathrm{median}_i \, \frac{\mathbf{a}_i^\top \mathbf{b}_i}{\|\mathbf{a}_i\|_2 \|\mathbf{b}_i\|_2 + \epsilon}.$$

Because cosine similarity is insensitive to vector magnitude, we additionally report the median relative kinetic residual

$$R_{\mathrm{kin}} = \mathrm{median}_i \, \frac{\|\mathbf{a}_i - \mathbf{b}_i\|_2}{\|\mathbf{a}_i\|_2 + \|\mathbf{b}_i\|_2 + \epsilon},$$

where $\epsilon > 0$ is a numerical stabilizer. JVP–RHS therefore measures directional agreement, whereas R_{kin} is sensitive to both direction and magnitude. Both are internal diagnostics. They quantify satisfaction of the kinetic equation supplied to the fitted model and are not, by themselves, independent evidence of biological correctness.

For the RNA–protein mechanistic ablations, we therefore additionally re-score each trained map against an uncorrupted biological reference. Irrespective of the velocity or transcript–protein correspondence used during training, the map JVP is recomputed using the biological RNA velocity and the kinetic right-hand side is evaluated using the correct transcript–protein mapping S. For each kinetic protein h, we compute

$$r_h^{\mathrm{bio}} = \mathrm{Pearson}_i\left(\left[J_\phi(\mathbf{r}_i)\mathbf{v}_r(\mathbf{r}_i)\right]_h, \left[f_\eta(\mathbf{r}_i, \phi(\mathbf{r}_i))\right]_h\right)$$

across cells and report the median across proteins. This biological-dynamics score is external to the corrupted training condition and tests whether the learned differential remains aligned with the biological first-order direction.

Independent chromatin first-order evaluation. Target RNA velocity is never used during chromatin-model training. After fitting, let

$$\mathbf{a}_i = [J_\Phi(\mathbf{z}_i)\mathbf{v}_{z,i}]_{\text{spliced}}, \qquad \mathbf{u}_i = \mathbf{v}^{\text{scVelo}}_{i,\text{spliced}}$$

denote the pushed-forward chromatin direction and independently inferred spliced-RNA velocity, respectively. We remove the population mean direction from each field,

$$\widetilde{\mathbf{a}}_i = \mathbf{a}_i - \bar{\mathbf{a}}, \qquad \widetilde{\mathbf{u}}_i = \mathbf{u}_i - \bar{\mathbf{u}},$$

and define JVP–scVelo as the median cell-wise cosine similarity

$$C_{\text{scVelo}} = \text{median}_i \, \frac{\widetilde{\mathbf{a}}_i^\top \widetilde{\mathbf{u}}_i}{\|\widetilde{\mathbf{a}}_i\|_2 \|\widetilde{\mathbf{u}}_i\|_2 + \epsilon}.$$

To construct a chance reference, we independently permute the cell correspondence of the target RNA-velocity field 20 times and recompute the median centred cosine. Let $q^{\text{perm}}_{0.95}$ denote the 95th percentile of these permutation scores. We report

$$\Delta_{\text{null}} = C_{\text{scVelo}} - q^{\text{perm}}_{0.95},$$

so positive values indicate first-order agreement exceeding the corresponding permutation threshold.

Perturbation-effect prediction. For CRISPR experiments, replicate-level perturbation effects are aggregated with equal replicate weight. Predicted and observed pooled protein effects are then compared across the eligible perturbation–protein pairs using Spearman rank correlation, Pearson correlation, and normalized root-mean-square error (NRMSE),

$$\sigma_{\Delta p} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(\Delta p_k - \overline{\Delta p}\right)^2}, \qquad \text{NRMSE} = \frac{\sqrt{K^{-1}\sum_{k=1}^{K}(\widehat{\Delta p}_k - \Delta p_k)^2}}{\sigma_{\Delta p}}.$$

where K is the number of evaluated perturbation–protein effects. Thus, NRMSE measures prediction error relative to the observed spread of biological effects. A predictor fixed at the observed mean has NRMSE one. Higher correlations and lower NRMSE indicate better prediction. Correlation is undefined for a constant prediction, such as the zero-change reference.

Temporal ordering and trajectory reconstruction. For each source cell, we identify the nearest observed protein state to $\phi(\mathbf{r}_i)$ in protein space. For evaluation only, each observed protein row inherits the pseudotime of its experimentally paired RNA cell. Time Spearman is the Spearman correlation between source-cell pseudotime and the pseudotime of its nearest observed protein target, while Time MAE is the corresponding mean absolute pseudotime difference.

Global trajectory reconstruction is evaluated using dynamic time warping (DTW) (Sakoe & Chiba, 1978). Mapped and observed protein states are each grouped into 100 equal-width evaluation-pseudotime bins, retaining only occupied bins, and represented by their within-bin centroids. DTW uses Euclidean distance between centroids with standard horizontal, vertical, and diagonal transitions. If the resulting trajectories contain T_a and T_b occupied bins, respectively, we report

$$D_{\text{DTW}} = \frac{\text{DTWcost}\left(\{\hat{\mathbf{p}}_t\}_{t=1}^{T_a}, \{\mathbf{p}_t\}_{t=1}^{T_b}\right)}{T_a + T_b}.$$

Lower values indicate more similar aggregate trajectories. These trajectory metrics are intentionally complementary to FOSCTTM. Coarse temporal ordering or global trajectory shape can be preserved even when individual cell correspondences are incorrect.

A.6 IMPLEMENTATION DETAILS

RNA and protein preprocessing. For the real RNA–protein experiments, RNA velocity is estimated with the dynamical scVelo model (Bergen et al., 2020). We select 2,000 highly variable genes using the Seurat-v3 criterion (Stuart et al., 2019) and force-retain genes linked to measured ADTs so that the kinetic projection matrix S is not restricted by HVG selection. Velocity preprocessing uses 30 principal components, 30 nearest neighbors, and a minimum of 20 shared spliced/unspliced counts. Dynamics are fit for all retained genes rather than HVGs alone.

KOT uses the scVelo smoothed spliced state M_s as its RNA state and the corresponding `velocity` layer as $\mathbf{v}_X$, ensuring that the state and vector field use the same coordinates. Protein measurements are CLR-normalized for the BMMC and PBMC alignment experiments. For the Papalexi CRISPR benchmark, where a four-protein CLR transform (Aitchison, 1982) would induce strong compositional effects, protein targets are instead normalized by the corresponding total-RNA size factor before log transformation. A RegVelo-derived RNA field (Wang et al., 2026) is used only in the velocity-provider ablation.

Network parameterization. The RNA–protein model contains three neural components:

$$\phi_\theta : \mathbb{R}^{D_r} \to \mathbb{R}^{D_p}, \qquad \alpha_\omega : \mathbb{R}^{D_r} \to \mathbb{R}^{D_p}, \qquad \kappa_\psi : \mathbb{R}^{D_r} \to \mathbb{R}.$$

The mapping network uses hidden dimensions $(1024, 512, 256)$, the production head $(256, 128)$, and the clock head $(64, 32)$. Hidden layers use SiLU activations (Elfwing et al., 2018). Weights are Xavier-normal (Glorot & Bengio, 2010) initialized with zero biases. The mapping network uses gain 0.5, whereas the kinetic heads use gain 0.1. For real RNA–protein experiments, spectral normalization (Miyato et al., 2018) is applied to every linear layer of ϕ_θ.

The clock and production heads are positive by construction. On real RNA–protein datasets,

$$10^{-3} \le \kappa_\psi(r) \le 1.5, \qquad 10^{-5} \le \alpha_{\omega,h}(r) \le 3,$$

implemented by bounded sigmoid parameterizations. Protein degradation rates are parameterized as

$$\beta_h = \mathrm{softplus}(\widetilde{\beta}_h) + 10^{-6}.$$

For the synthetic benchmark, spectral normalization (Miyato et al., 2018) is disabled and α is restricted to $[10^{-3}, 1.5]$.

Velocity support and normalization. For optional gene-level reliability weights $\mathbf{w}_v \in [0,1]^{D_x}$,

$$\mathbf{u}^{(i)} = \mathbf{w}_v \odot \mathbf{v}_X^{(i)}. \tag{18}$$

On real datasets we remove one dataset-level velocity gauge using

$$g_v = \mathrm{median}_{i:\|\mathbf{u}^{(i)}\|_2>0} \|\mathbf{u}^{(i)}\|_2, \qquad \widetilde{\mathbf{v}}_X^{(i)} = \frac{\mathbf{u}^{(i)}}{g_v + \epsilon}. \tag{19}$$

The same scalar is applied to every cell and remains fixed throughout training.

Feature scaling. The fixed scale in Eq. equation 3 is

$$s_j = \max\left\{ \mathrm{Std}(\{y_j^{(\ell)}\}_{\ell=1}^M), \mathrm{median}_k \,\mathrm{Std}(\{y_k^{(\ell)}\}_{\ell=1}^M) + \epsilon \right\}. \tag{20}$$

This down-weights high-variance coordinates without amplifying near-constant features.

Sinkhorn divergence. We use

$$\mathcal{S}_\varepsilon(\mu,\nu) = \mathrm{OT}_\varepsilon(\mu,\nu) - \frac{1}{2}\,\mathrm{OT}_\varepsilon(\mu,\mu) - \frac{1}{2}\,\mathrm{OT}_\varepsilon(\nu,\nu). \tag{21}$$

For datasets exceeding the Sinkhorn memory budget, RNA and target cells are sampled independently at every epoch. The two index sets are drawn separately, so the true paired row order is never exposed to the alignment loss.

Table 6: **KOT implementation and training configuration.** Shared settings are inherited across experiments unless an experiment-specific override is listed.

Setting	Synthetic	BMMC / PBMC	Papalexi	BMMC chromatin
ϕ hidden widths	$1024, 512, 256$	$1024, 512, 256$	$1024, 512, 256$	$1024, 512, 256$
α hidden widths	$256, 128$	$256, 128$	$256, 128$	$256, 128$
κ hidden widths	$64, 32$	$64, 32$	$64, 32$	$64, 32$
Activation	SiLU	SiLU	SiLU	SiLU
Initialization	Xavier	Xavier	Xavier	Xavier
ϕ spectral norm	no	yes	yes	yes
κ range	$[10^{-3}, 1.5]$	$[10^{-3}, 1.5]$	$[10^{-3}, 1.5]$	$[10^{-3}, 1.5]$
α range	$[10^{-3}, 1.5]$	$[10^{-5}, 3]$	$[10^{-5}, 3]$	$[10^{-5}, \infty)$
Positive rate transform	softplus	softplus	softplus	softplus
Maximum epochs	500	500	500	500
Kinetic batch size	full / 8192	8192	8192	8192
λ_{dyn}	10	1000	1000	30
Kinetic residual	scaled L2	scaled L2	scaled L2	scaled L2
Direction-only kinetic term	0	0	0	0
Velocity gauge normalization	no	yes	yes	yes
ρ_κ	0.01	0.01	0.01	0.01
κ_0	$\log 2$	$\log 2$	$\log 2$	$\log 2$
ρ_{net}	10^{-4}	10^{-4}	10^{-4}	10^{-4}
Sinkhorn blur	0.1	0.1	0.1	0.1
Sinkhorn backend	tensorized	tensorized	tensorized	tensorized
Max. Sinkhorn cells	full	BMMC: 8192; PBMC: full	8192	2048
Biochemical anchors	manual simulator calibration	protein half-lives	none	RNA half-lives
Anchor weight	$\rho_\beta = 1$	$\rho_\beta = 1$	0	$\rho_\gamma = 1$
Validation fraction	0	BMMC: 0.1; PBMC: 0.2	0.2 of NT	0.1
Test fraction	–	–	KO population	0.2
Validation stratification	–	BMMC: cell type	replicate	fixed split
Seeds	12	12	12	12 main / 3 sensitivity

Optimization and checkpointing. All loss terms are differentiated before a single optimizer update. We use Adam with separate parameter groups for the mapping network, the joint (α, κ) heads, and degradation-rate parameters. A single multiplicative learning-rate schedule is shared by all groups, preserving their relative step-size ratios throughout optimization. The schedule linearly warms from a fixed starting factor to the nominal learning rate and subsequently follows a cosine decay to a fixed minimum factor.

When kinetic warm-up is used,

$$\lambda_{\mathrm{dyn}}^{(t)} = q_t \lambda_{\mathrm{dyn}}, \qquad 0 \le q_t \le 1, \tag{22}$$

where q_t increases linearly and equals one after the configured warm-up. We train for at most 500 epochs and clip the global gradient norm to 1. The network L_2 penalty is applied to ϕ, α, and κ, but not to the biochemical rate vectors.

For RNA–protein runs, checkpoints are recorded for the lowest alignment loss, lowest held-out alignment loss when available, lowest kinetic loss, lowest total objective, and the final state. Hyperparameter search uses fixed held-out development splits. Unpaired FOSCTTM is used only to rank development configurations and never enters the optimization objective. For the final chromatin runs, checkpoint selection uses the unpaired validation Sinkhorn divergence (Feydy et al., 2019) rather than paired FOSCTTM.

Known pairing is never used in the training loss or gradient updates or during development. Paired validation FOSCTTM is only used for evaluation.

The final chromatin configuration additionally uses TF–IDF/LSI coordinates for the map and JVP, linear CP10K gene activity for the transcription head, shared-$log(1+x)$ RNA coordinates, a scalar-zero nonlinear residual gate, 32 Sinkhorn alignment dimensions, a spliced-only primary alignment loss, and an auxiliary unspliced moment loss with weight $\lambda_u = 1$. The gamma-anchor weight is $\lambda_\gamma = 1$.

Table 6 shows an overview of the specific implementation for each setting.

Optimizer settings. For the real RNA–protein experiments, the selected optimizer configuration uses learning rate 10^{-3} for the cross-modal map ϕ_θ, 10^{-4} for the production and clock heads $(\alpha_\omega, \kappa_\psi)$, and 10^{-2} for the degradation-rate parameters β. Adam (Kingma & Ba, 2014) is used

for all parameter groups. The dynamics coefficient is $\lambda_{\mathrm{dyn}} = 1000$ for the RNA–protein experiments and is increased linearly from zero during the kinetic warm-up. The learning-rate scheduler acts multiplicatively on all parameter groups, thereby preserving these relative learning-rate ratios.

A.7 Optimization and training sensitivity

All optimization choices were selected on fixed development splits before the reported multi-seed runs. These experiments are intended to distinguish sensitivity to optimization from the mechanistic effects studied in the main text. Paired evaluation FOSCTTM is used only for evaluation.

Table 7: **Dynamics-weight warm-up sensitivity.** The kinetic coefficient is increased linearly from zero to $\lambda_{\mathrm{dyn}} = 10^3$. Values are mean $\pm$ standard deviation over four seeds.

			BMMC		PBMC	
η_ϕ	$\eta_{\alpha,\kappa}$	λ_{dyn}	FOSCTTM ↓	JVP–RHS ↑	FOSCTTM ↓	JVP–RHS ↑
10^{-3}	10^{-4}	10^3	0.134 ± 0.003	0.927 ± 0.004	0.136 ± 0.006	0.743 ± 0.019
3×10^{-3}	10^{-4}	10^3	0.137 ± 0.004	0.906 ± 0.011	0.136 ± 0.006	0.702 ± 0.039
10^{-3}	10^{-4}	10^2	0.124 ± 0.004	0.728 ± 0.019	0.151 ± 0.018	0.652 ± 0.032
3×10^{-4}	10^{-4}	10^3	0.131 ± 0.002	0.925 ± 0.002	0.158 ± 0.004	0.782 ± 0.015
3×10^{-3}	10^{-4}	10^2	0.139 ± 0.008	0.576 ± 0.091	0.156 ± 0.014	0.515 ± 0.033
3×10^{-4}	10^{-4}	10^2	0.129 ± 0.006	0.750 ± 0.012	0.169 ± 0.017	0.714 ± 0.020
10^{-3}	10^{-3}	10^3	0.137 ± 0.002	0.954 ± 0.002	0.202 ± 0.006	0.809 ± 0.014
3×10^{-3}	3×10^{-3}	10^3	0.161 ± 0.044	0.942 ± 0.007	0.194 ± 0.031	0.776 ± 0.031
3×10^{-3}	10^{-3}	10^3	0.138 ± 0.004	0.941 ± 0.004	0.225 ± 0.005	0.715 ± 0.031
3×10^{-4}	10^{-3}	10^3	0.140 ± 0.009	0.958 ± 0.002	0.226 ± 0.010	0.866 ± 0.018

Learning rates and kinetic weight. We first jointly screen the learning rate of the cross-modal map η_ϕ, the learning rate of the production and clock heads $\eta_{\alpha,\kappa}$, and the final kinetic weight λ_{dyn} (Table 7). The degradation-rate learning rate is fixed at $\eta_\beta = 10^{-2}$. The selected configuration, $(\eta_\phi, \eta_{\alpha,\kappa}, \lambda_{\mathrm{dyn}}) = (10^{-3}, 10^{-4}, 10^3)$, provides a balanced operating point across BMMC and PBMC, with low cell-level alignment error while retaining substantial agreement between the Jacobian–vector product and the mechanistic right-hand side. We therefore freeze this rate triple before subsequent schedule ablations.

Table 8: **Dynamic weight ramp sensitivity.** FOSCTTM of different dynamics-weight ramps for settings B and C with B using a β learning rate of 0.001 with a 300-epoch learning-rate warmup and C uses 0.01 with a 500-epoch warmup.

Setting	Ramp	BMMC test ↓	PBMC test ↓
B	0	0.423±0.029	0.334±0.037
B	50	0.132±0.003	**0.135**±0.002
B	150	0.129±0.003	0.138±0.003
B	300	0.124±0.007	0.141±0.004
C	0	0.423±0.029	0.334±0.035
C	50	0.121±0.002	0.146±0.008
C	150	0.119±0.002	0.146±0.007
C	300	**0.118**±0.003	0.141±0.002

4 seeds throughout.

Dynamics warm-up and kinetic calibration. Applying the full kinetic weight from the beginning of training is unstable. With no dynamics warm-up, FOSCTTM increases to approximately 0.42 on BMMC and 0.33 on PBMC under both tested optimization settings (Table 8). Introducing a gradual dynamics-weight ramp restores cell-level correspondence, while longer ramps produce a modest dataset-specific trade-off rather than a uniformly optimal schedule. At the 300-epoch ramp, for example, setting C gives slightly lower BMMC FOSCTTM than setting B (0.118 ± 0.003 versus 0.124 ± 0.007), while the two settings are indistinguishable on PBMC (0.141 in both cases).

Table 9 shows that similar alignment performance can nevertheless correspond to substantially different kinetic solutions. Setting B achieves stronger internal agreement between the pushed-forward source field and the fitted mechanistic right-hand side, with JVP–RHS 0.969 ± 0.001 versus 0.930 ± 0.006 on BMMC and 0.816 ± 0.022 versus 0.803 ± 0.013 on PBMC. In contrast, setting C matches the imposed protein degradation-rate anchors far more closely. β MAE decreases from

0.296 ± 0.004 to 0.010 ± 0.001 on BMMC and from 0.121 ± 0.006 to 0.003 ± 0.000 on PBMC. Thus tighter biochemical calibration does not automatically increase internal dynamical agreement, and higher JVP–RHS does not imply more accurate recovery of the anchored degradation rates.

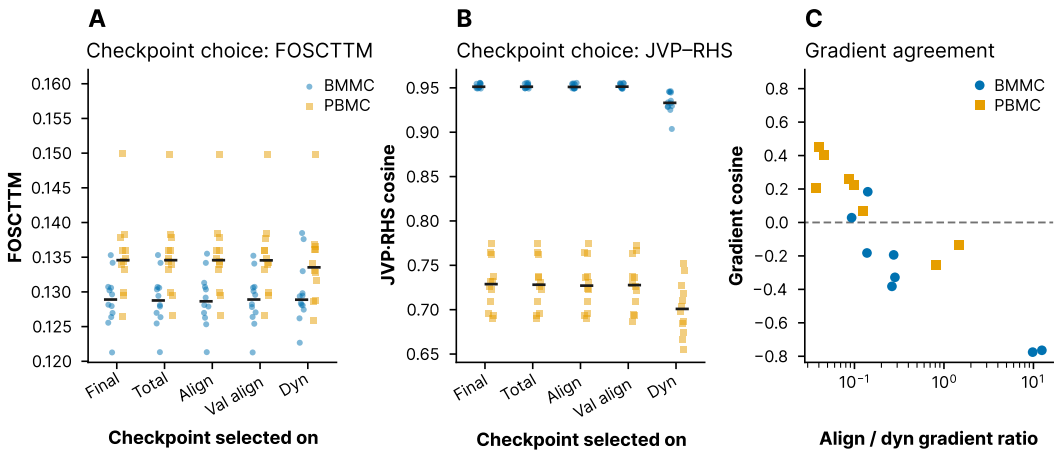


Figure 8: **Checkpoint selection based on state alignment and kinetic consistency.** (A) FOSCTTM from different training checkpoints. (B) JVP–RHS agreement across different checkpoints. (C) Relationship between the relative alignment/dynamics gradient magnitude and their gradient cosine across BMMC and PBMC optimization runs.

Figure 8 makes the optimization trade-off explicit. Checkpoint selection based only on marginal alignment can retain good FOSCTTM while sacrificing first-order consistency, whereas selection by the dynamics objective can favor the converse. The gradient diagnostic further shows that the two objectives are not uniformly cooperative, depending on their relative magnitude, their updates can be nearly orthogonal or locally antagonistic. This motivates both the dynamics-weight warm-up and our use of separate alignment and kinetic diagnostics during model selection.

This trade-off reflects the fact that the mechanistic right-hand side is jointly parameterized by α, κ, and β. A schedule that constrains β more tightly to its external targets reduces the freedom of the fitted kinetic field to adapt to the Jacobian push-forward, whereas a looser degradation-rate fit can be partially compensated by the remaining kinetic components to obtain higher JVP–RHS agreement. Because settings B and C differ jointly in the β learning rate and its learning-rate warm-up, this experiment should be interpreted as a comparison of kinetic optimization regimes rather than as an isolated effect of a single hyperparameter. Taken together, the three metrics show that cell-level alignment, internal kinetic consistency, and biochemical rate calibration are related but non-equivalent objectives.

Table 9: **Alignment–calibration trade-off at 300-epoch dynamics warm-up.** Settings B and C: B uses a β learning rate of 0.001 with a 300-epoch learning-rate warmup and C uses 0.01 with a 500-epoch warmup. Values are mean $\pm$ standard deviation over four seeds.

	BMMC test			PBMC test		
Setting	FOS ↓	JVP–RHS ↑	β MAE ↓	FOS ↓	JVP–RHS ↑	β MAE ↓
B	0.124±0.007	**0.969**±0.001	0.296±0.004	**0.141**±0.004	**0.816**±0.022	0.121±0.006
C	**0.118**±0.003	0.930±0.006	**0.010**±0.001	0.141±0.002	0.803±0.013	**0.003**±0.000

Training stability under mechanistic corruptions. Finally, we examine whether the failure modes in the main mechanistic ablations can be explained by trivial optimization collapse (Table 10). The mapped-state variance remains of comparable magnitude across the tested velocity and correspondence corruptions, indicating that their poor biological behavior is not simply caused by a collapsed output representation. In contrast, several corrupted kinetic configurations drive a substantial fraction of production-rate outputs to their lower bound, providing a direct diagnostic of

kinetic-head saturation. Gradient-norm ratios and gradient cosines further characterize the interaction between the distributional and kinetic objectives rather than replacing the biological diagnostics reported in the main text.

Table 10: **Training stability under mechanistic corruptions.** The variance ratio measures mapped-state dispersion relative to observed target distribution, while the α-floor fraction reports the fraction of production-rate outputs at lower parameter bound. The gradient ratio compares norms of the dynamics and alignment gradients, and the gradient cosine measures their directional agreement. Values are mean $\pm$ standard deviation over 12 seeds.

Condition	Var. ratio	α floor	Grad. ratio	Grad. cos.
PBMC				
KOT	0.874±0.005	–	0.080±0.017	0.111±0.314
KOT ($v \to -v$)	0.864±0.009	0.334±0.009	0.082±0.019	0.072±0.347
KOT ($v \to 0$)	0.876±0.007	0.342±0.008	0.066±0.010	0.194±0.294
KOT (S perturbed)	0.848±0.015	0.517±0.031	0.078±0.022	0.153±0.353
KOT ($\lambda_{\text{dyn}} = 0$)	0.905±0.052	–	0.000±0.000	0.085±0.117
BMMC				
KOT	0.675±0.008	–	0.317±0.070	-0.361±0.271
KOT ($v \to -v$)	0.674±0.010	0.580±0.028	0.380±0.068	-0.370±0.255
KOT ($v \to 0$)	0.687±0.007	0.434±0.012	0.287±0.046	-0.311±0.195
KOT (S perturbed)	0.652±0.014	0.740±0.014	0.337±0.055	-0.221±0.213
KOT ($\lambda_{\text{dyn}} = 0$)	0.693±0.019	–	0.000±0.000	0.057±0.067

The interaction is also strongly time dependent (Figure 9). Early in training the weighted kinetic contribution can be comparable to or larger than the alignment gradient, while later the balance changes as the two losses approach different parts of their optimization trajectories. The gradient cosine is frequently negative, showing that the kinetic term is not merely a redundant regularizer for the marginal objective. Instead, it can actively redirect the learned map toward a solution satisfying the first-order constraint.

The two-dimensional sweep in Figure 10 shows that kinetic weight and degradation-rate optimization interact rather than acting as independent knobs. Configurations with similar alignment error can produce substantially different JVP–RHS agreement, and the favorable region differs between BMMC and PBMC. We therefore treat the selected values as a frozen joint optimization configuration rather than attributing performance to a single hyperparameter in isolation.

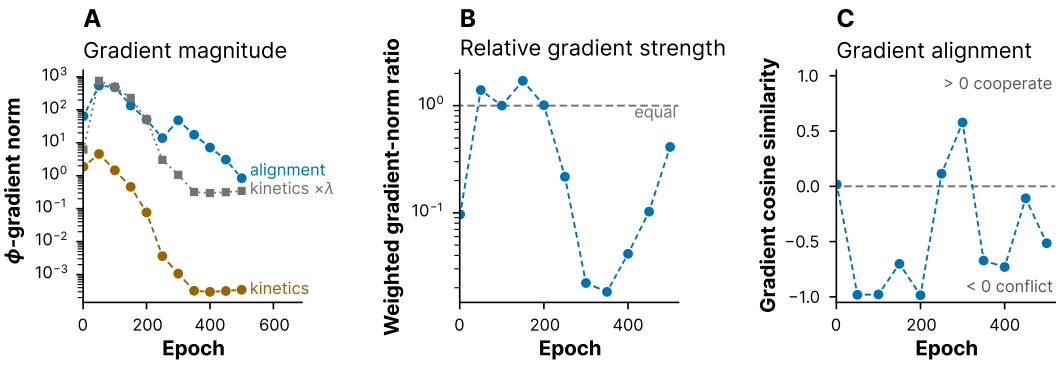


Figure 9: **Balance between marginal alignment and kinetic consistency over training duration.** (A) Gradient norms on the transport map ϕ. (B) Ratio of weighted kinetic to alignment gradient magnitude. (C) Cosine similarity between two mapping gradients. Positive values indicate locally cooperative updates.

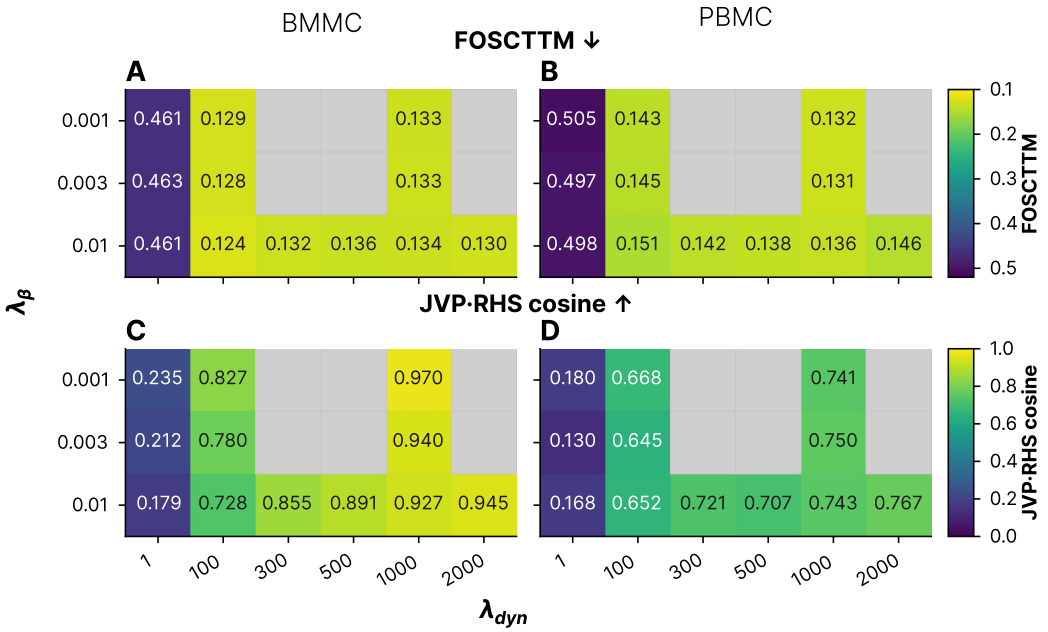


Figure 10: **Joint sensitivity to kinetic-loss weight and degradation-rate optimization.** FOSCTTM (A,B) and JVP–RHS agreement (C,D) across a tuning slice over λ_{dyn} and learning rate for β on BMMC and PBMC. Hatched cells were not evaluated.

A.8 SOURCE-VELOCITY ROBUSTNESS AND TRAJECTORY DIAGNOSTICS

The source vector field is the principal additional observation supplied to KOT beyond the two unpaired state distributions. Consequently, errors in this field can affect the learned differential even when the marginal alignment remains well specified.We therefore examine source-velocity dependence at several complementary levels. First, the synthetic benchmark allows direct corruption of a known source vector field. Second, on CITE-seq data we examine how velocity and mechanism corruptions alter the fitted effective kinetic decomposition. Third, we replace the RNA-velocity estimator while holding the remaining KOT configuration fixed. Finally, we test whether the resulting maps preserve cell-level temporal ordering and global protein-trajectory geometry. These analyses complement the local first-order diagnostics in Table 3. JVP–RHS measures consistency with the field supplied during training, whereas the trajectory metrics below evaluate properties of the recovered cross-modal correspondence.

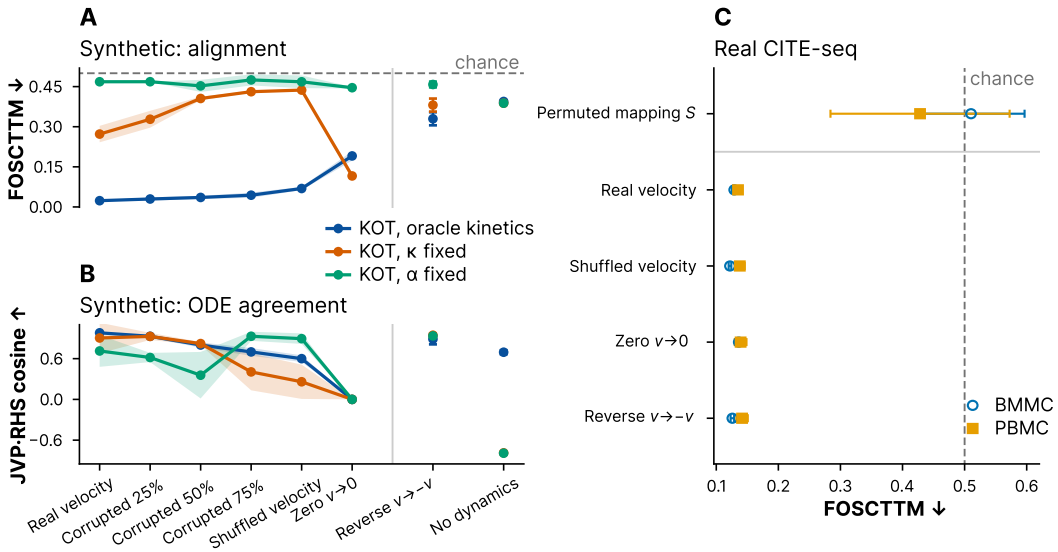


Figure 11: **Source-field perturbations.** (A) Synthetic Branch FOSCTTM under increasing corruption of the known source velocity. (B) Corresponding JVP–RHS agreement. (C) FOSCTTM on BMMC and PBMC under the observed velocity, shuffled, zeroed, or reversed velocity, permuted S.

Figure 11 summarizes the resulting separation between static and first-order behavior. In the synthetic system, where the source field is known exactly, progressive corruption reduces both correspondence quality and agreement with the target law. On real CITE-seq data, however, shuffled, reversed, and zero velocity can leave FOSCTTM close to the observed-velocity solution. Static retrieval is therefore insufficient to diagnose whether the learned differential uses the supplied source field correctly.

Ground-truth synthetic velocity ablations. The synthetic Branch setting allows us to perturb the source vector field without uncertainty from RNA-velocity estimation (see Table 11). Under privileged oracle kinetics, KOT recovers both the cell correspondence and branch identity with the true velocity (FOSCTTM 0.023 ± 0.001, branch accuracy 0.986 ± 0.008). Shuffling or partially corrupting the field progressively degrades alignment, while branch identity remains comparatively robust under moderate corruption. Reversing the velocity produces a much larger correspondence error (0.330 ± 0.026), despite retaining high branch accuracy, showing that coarse branch identity alone does not determine the correct within-branch correspondence.

The comparison across kinetic parameterizations also shows that velocity information is useful only when the target kinetic law is sufficiently compatible with the simulator. Fixing different subsets of the kinetic components produces markedly different sensitivity to the same source field, whereas the privileged oracle model provides the cleanest controlled test of velocity corruption. Removing

Table 11: **Branch-mechanism ablation.** The three column groups differ only in which synthetic kinetic components are fixed to simulator values versus learned. "Oracle" uses the privileged simulator kinetics throughout.

	KOT fixed α		KOT fixed κ		KOT oracle	
Velocity input	FOSCTTM ↓	Branch acc. ↑	FOSCTTM ↓	Branch acc. ↑	FOSCTTM ↓	Branch acc. ↑
Real velocity	0.468±0.002	0.000±0.000	0.273±0.032	0.330±0.060	**0.023**±0.001	**0.986**±0.008
Velocity corrupted						
KOT (v shuffled)	0.468±0.025	0.000±0.000	0.437±0.004	0.000±0.000	0.069±0.006	0.921±0.023
KOT ($v \to -v$)	0.458±0.011	0.014±0.024	0.381±0.025	0.232±0.075	0.330±0.026	0.954±0.036
KOT ($v \to 0$)	0.446±0.003	0.000±0.000	**0.116**±0.014	**0.840**±0.173	0.191±0.018	0.970±0.021
KOT (v corruption 25%)	0.468±0.001	0.000±0.000	0.328±0.033	0.237±0.073	0.030±0.006	0.973±0.034
KOT (v corruption 50%)	0.453±0.024	0.000±0.000	0.406±0.007	0.030±0.024	0.035±0.006	0.947±0.071
KOT (v corruption 75%)	0.475±0.022	0.000±0.000	0.431±0.004	0.000±0.000	0.044±0.010	0.930±0.038
Ablation						
KOT ($\lambda_{dyn} = 0$)	**0.388**±0.002	**0.032**±0.001	0.388±0.002	0.032±0.001	0.394±0.004	0.035±0.002

12 seeds throughout. Lower FOSCTTM (↓) and higher branch accuracy (↑) are better.

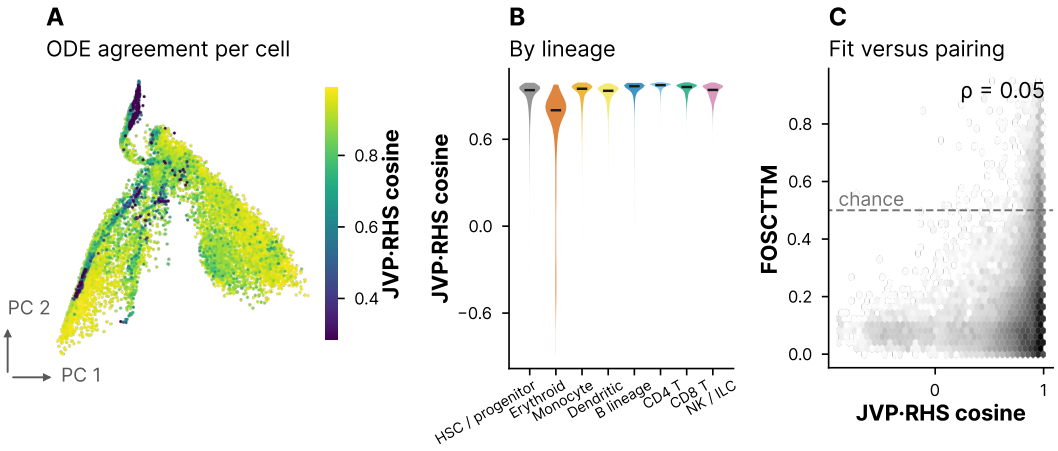


Figure 12: **Internal kinetic consistency across cells and lineages.** (A) Per-cell JVP–RHS cosine projected onto the low-dimensional cell-state geometry. (B) Distribution of per-cell JVP–RHS cosine across major hematopoietic lineages. (C) Joint distribution of per-cell JVP–RHS agreement and FOSCTTM. The dashed FOSCTTM line denotes random-scale retrieval.

the kinetic objective entirely collapses branch recovery to near chance. Thus the synthetic experiment isolates the effect of the source field itself, while the real-data experiments below additionally include uncertainty from velocity estimation and effective kinetic modeling.

Fitted kinetic response to mechanistic corruption. We next examine how the learned kinetic decomposition responds when the first-order information supplied to KOT is altered. Because the translation–degradation field is parameterized jointly through α_ω, κ_ψ, and β, a corrupted source field need not manifest only as a larger kinetic residual; the model can also redistribute the fitted effective rates. We therefore report summary statistics of the learned production, clock, and degradation components under the same mechanistic ablations used in the main text.

The fitted degradation scale is comparatively stable across the kinetic variants (see Table 12). On BMMC, the mean degradation rate $\bar{\beta}$ remains between 0.664 and 0.671 for the real-velocity, shuffled-velocity, reversed-velocity, zero-velocity, and permuted-S conditions. On PBMC it remains between 0.665 and 0.674. In contrast, the effective production and clock components respond substantially to changes in the supplied first-order structure. For example, on PBMC the median production rate α_{med} changes from 0.783 with the observed velocity to 1.111 under shuffled velocity and 1.040 under reversed velocity, while the median clock scale κ_{med} decreases from 0.017 to 0.011 and

Table 12: **Effect of mechanistic corruptions on fitted kinetic parameters.** Reported values are the median effective production rate α_{med}, median clock scale κ_{med}, and mean degradation rate $\bar{\beta}$.

	BMMC			PBMC		
Condition	α_{med}	κ_{med}	$\bar{\beta}$	α_{med}	κ_{med}	$\bar{\beta}$
Real velocity	.074	.028	.669	.783	.017	.666
Shuffled velocity	.140	.021	.668	1.111	.011	.665
Reverse velocity	.106	.024	.668	1.040	.015	.666
Zero velocity	.287	.011	.664	.945	.007	.665
Permuted S	.009	.027	.671	.136	.014	.674
No dynamics	1.500	.697	.706	1.500	.698	.708

0.015, respectively. Zeroing the velocity produces a further shift in the decomposition, particularly in κ.

The same distinction is visible at the level of individual cells (Figure 12). JVP–RHS agreement is highly structured across the cell-state manifold and differs by lineage, but it has almost no monotone association with per-cell FOSCTTM. Internal satisfaction of the kinetic equation and recovery of the correct partner are therefore complementary properties of the map rather than interchangeable measures of success.

Permuting the transcript–protein correspondence S gives a different response. The fitted degradation scale again changes little, whereas the production component moves substantially, consistent with the model being forced to explain the same target-space dynamics through an incorrect cross-modal feature correspondence. This complements the main mechanistic ablation. A model can retain strong internal kinetic agreement under an incorrect S while reallocating the effective kinetic parameters and destroying the state correspondence.

Removing the dynamical objective entirely produces the clearest qualitative change. The median production rate reaches its upper parameter bound of 1.5 on both datasets, while κ_{med} rises to approximately 0.7, more than an order of magnitude above its value in the full kinetic model. Once the Jacobian push-forward is no longer required to match the mechanistic right-hand side, the kinetic heads are therefore no longer constrained to represent the observed source dynamics and settle into a qualitatively different parameter regime.

These quantities should not be interpreted as uniquely identified biochemical rates. As shown in Theorem 2.7, the RNA–protein kinetic model contains a global multiplicative scale gauge coupling α, κ, and β. The parameter summaries in Table 12 are therefore diagnostics of how the learned effective kinetic decomposition responds to mechanistic perturbation. Their value is in revealing where the model absorbs a corrupted dynamical constraint, rather than in assigning absolute biochemical meaning to the individual fitted rates.

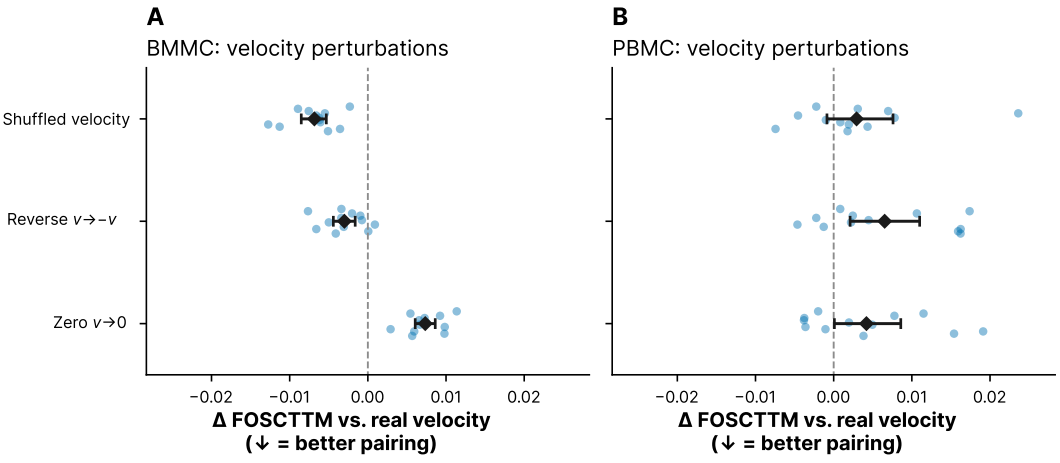


Figure 13: **Velocity corruption effect on static correspondence.**

The zoomed view in Figure 13 emphasizes how small these static effects are. Shuffling or reversing the source field changes FOSCTTM by only a few hundredths, and the sign of the change is not uniformly worse across datasets. These ablations should therefore not be interpreted as a simple claim that the observed velocity always minimizes static retrieval error. Their importance emerges from the first-order and perturbational diagnostics, where the consequences of corrupting the field are substantially larger.

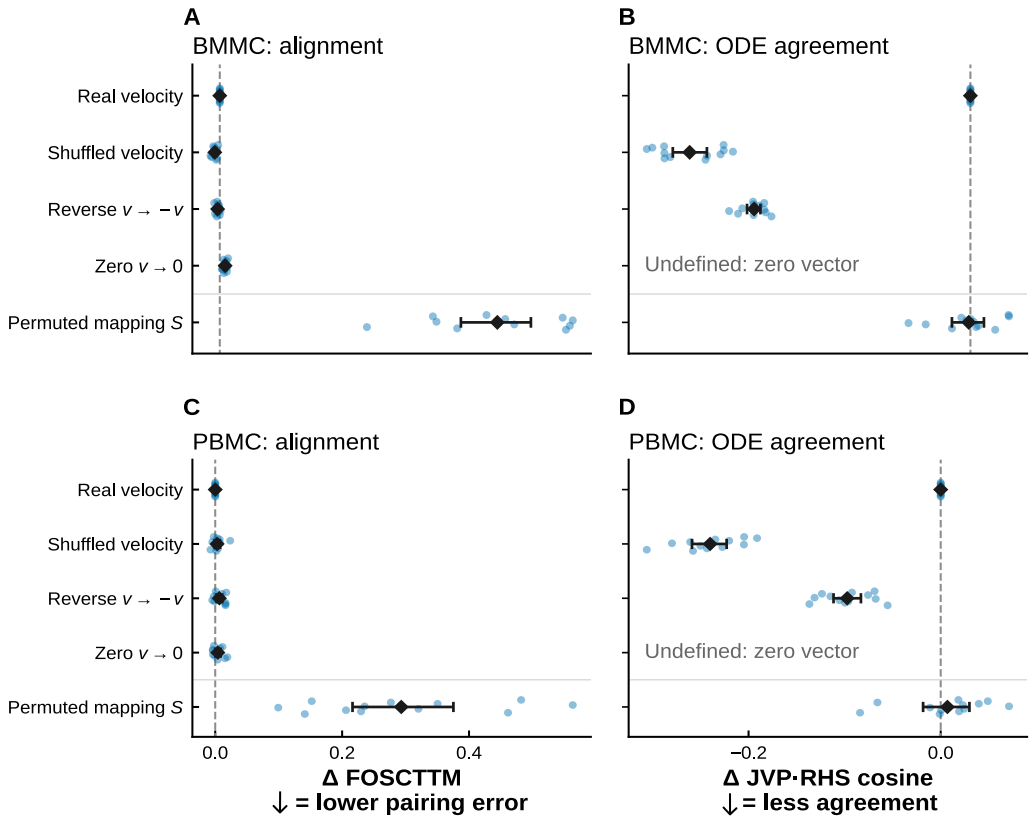


Figure 14: **Mechanistic perturbations effects.** Changes relative to the observed-velocity KOT model in (A,C) FOSCTTM and (B,D) JVP–RHS agreement on BMMC and PBMC. Light points denote individual runs acorss 12 seeds, error bars summarize repeated runs in black.

Figure 14 makes the orthogonality of these failure modes explicit. Velocity corruptions primarily alter the first-order relationship while preserving much of the static correspondence. Permuting S produces the converse pattern. The model can retain strong internal agreement between its JVP and fitted right-hand side while mapping cells to the wrong protein states. This provides a visual counterpart to Table 3 and reinforces why neither FOSCTTM nor JVP–RHS can be interpreted in isolation.

Robustness to the velocity estimator. Our primary experiments use dynamical scVelo (Bergen et al., 2020). To test whether KOT depends on this particular velocity model, we repeat the fit using RegVelo (Wang et al., 2026) while retaining the same cells and remaining model configuration (Table 13). Both estimators support successful KOT training. With scVelo, FOSCTTM is 0.121 ± 0.006 on BMMC and 0.137 ± 0.006 on PBMC, compared with 0.135 ± 0.003 and 0.148 ± 0.007 using RegVelo. Conversely, RegVelo produces higher internal JVP–RHS agreement on both datasets (0.993 ± 0.001 and 0.890 ± 0.018).

This comparison should not be interpreted as a ranking of biological velocity accuracy. JVP–RHS measures how well the learned differential agrees with the kinetic field fitted using the supplied source velocity. A higher value therefore establishes stronger internal compatibility, not necessarily a more accurate biological vector field. Rather, the experiment shows that KOT can operate with distinct source-velocity estimators and that static correspondence and internal kinetic fit need not improve together.

Table 13: **Robustness to the source RNA-velocity estimator.** Values are mean ± standard deviation over 12 seeds.

	BMMC		PBMC	
Backend	FOSCTTM ↓	JVP–RHS ↑	FOSCTTM ↓	JVP–RHS ↑
scVelo	**0.121**±0.006	0.969±0.002	**0.137**±0.006	0.790±0.022
RegVelo	0.135±0.003	**0.993**±0.001	0.148±0.007	**0.890**±0.018

Both velocity backends are compared after the same post-selection all-cell refit. These values therefore differ slightly from the canonical training-split results in the main text.

Temporal ordering and trajectory reconstruction. We next ask whether the learned cross-modal map preserves temporal structure beyond nearest-neighbor correspondence. For each source cell, we assign the pseudotime of its nearest observed protein target under the learned map and compare it with the source RNA-velocity pseudotime. Time Spearman therefore measures preservation of cell-level temporal ordering, while Time MAE measures the corresponding absolute pseudotime displacement.

We separately evaluate the global trajectory geometry. Predicted and observed protein states are ordered by evaluation pseudotime, summarized by centroids over 100 pseudotime bins, and compared using normalized dynamic time warping (DTW). DTW reconstruction therefore measures similarity of the aggregate protein trajectory and does not require that individual cells occupy the same temporal position. The two diagnostics are intentionally complementary. A map may reconstruct the coarse trajectory while assigning individual cells to incorrect positions along it.

The standard KOT map preserves temporal ordering strongly, with Time Spearman 0.795 ± 0.010 on BMMC and 0.832 ± 0.014 on PBMC (Table 14). Removing the kinetic objective causes a large loss of temporal organization: the corresponding correlations fall to -0.325 ± 0.348 and 0.256 ± 0.390, while DTW reconstruction error increases from 0.515 ± 0.017 to 1.869 ± 0.116 on BMMC and from 2.032 ± 0.057 to 3.720 ± 1.193 on PBMC. Thus the difference between KOT and the distribution-only model is visible not only in paired cell retrieval, but also in the temporal organization of the recovered cross-modal map.

Importantly, these coarse trajectory metrics are much less sensitive to the *orientation* of the supplied velocity than the independent biological dynamics score in the main text. Shuffling, reversing, or zeroing v leaves Time Spearman and DTW reconstruction close to the standard model on both datasets. In particular, reversing v preserves Time Spearman (0.784 ± 0.014 on BMMC and 0.816 ± 0.019 on PBMC), even though the independent biological-dynamics score in Table 3 becomes strongly negative.

This is not contradictory as pseudotemporal ordering and local vector-field orientation probe different properties of the map. A model can recover the coarse state manifold and its temporal ordering while transforming local directions incorrectly. Consequently, trajectory preservation alone cannot validate the orientation of the learned first-order correspondence.

Permuting the transcript–protein correspondence matrix S produces the complementary failure mode. Although the main-text ablation shows that this model can retain high internal JVP–RHS agreement, its temporal correspondence collapses since Time Spearman falls to -0.102 ± 0.450 on BMMC and 0.264 ± 0.294 on PBMC, with DTW errors approaching those of KOT ($\lambda_{\mathrm{dyn}} = 0$). Thus internal satisfaction of a kinetic equation does not preserve the correct global trajectory when the cross-modal feature correspondence itself is wrong.

The velocity-provider comparison further separates cell-level temporal assignment from aggregate trajectory reconstruction. On PBMC, the RegVelo all-cell model has substantially weaker Time Spearman than the matched scVelo refit, yet a lower DTW reconstruction error. It therefore better

Table 14: **Temporal ordering and trajectory reconstruction under velocity and mechanism ablations.** Time Spearman is the Spearman correlation between source-cell RNA-velocity pseudotime and the pseudotime assigned to the nearest observed protein target under the learned map. Time MAE is the corresponding mean absolute pseudotime error. DTW reconstruction is the normalized dynamic-time-warping distance between 100-bin pseudotime centroid trajectories of the mapped and observed protein states, with lower values indicating better reconstruction of the global protein trajectory. Rows use scVelo. Values are mean $\pm$ standard deviation over 12 seeds.

Condition	Time $\rho \uparrow$	Time MAE $\downarrow$	DTW recon. $\downarrow$
PBMC			
KOT (full)	0.832 ± 0.014	0.023 ± 0.002	2.032 ± 0.057
KOT (all-cell refit)	0.810 ± 0.018	0.025 ± 0.002	2.125 ± 0.097
KOT (RegVelo, all-cell refit)	0.401 ± 0.070	0.161 ± 0.015	1.184 ± 0.044
KOT (shuffled v)	0.824 ± 0.015	0.026 ± 0.001	2.049 ± 0.068
KOT ($v \to -v$)	0.816 ± 0.019	0.025 ± 0.002	2.042 ± 0.089
KOT ($v \to 0$)	0.821 ± 0.017	0.024 ± 0.001	2.044 ± 0.110
KOT (permuted S)	0.264 ± 0.294	0.166 ± 0.102	3.688 ± 0.921
KOT ($\lambda_{\text{dyn}} = 0$)	0.256 ± 0.390	0.166 ± 0.124	3.720 ± 1.193
BMMC			
KOT (full)	0.795 ± 0.010	0.065 ± 0.003	0.515 ± 0.017
KOT (all-cell refit)	0.574 ± 0.039	0.044 ± 0.004	0.495 ± 0.034
KOT (RegVelo, all-cell refit)	0.823 ± 0.014	0.082 ± 0.004	0.495 ± 0.012
KOT (shuffled v)	0.772 ± 0.016	0.070 ± 0.007	0.522 ± 0.027
KOT ($v \to -v$)	0.784 ± 0.014	0.071 ± 0.009	0.469 ± 0.016
KOT ($v \to 0$)	0.793 ± 0.015	0.071 ± 0.006	0.507 ± 0.014
KOT (permuted S)	-0.102 ± 0.450	0.204 ± 0.068	1.732 ± 0.190
KOT ($\lambda_{\text{dyn}} = 0$)	-0.325 ± 0.348	0.240 ± 0.050	1.869 ± 0.116

reproduces the coarse shape of the protein trajectory while assigning individual cells less accurately along that trajectory. On BMMC, by contrast, RegVelo retains strong temporal ordering and comparable DTW reconstruction. These results reinforce that no single trajectory statistic is sufficient to assess the learned correspondence.

A.9 BIOCHEMICAL ANCHOR CONSTRUCTION

KOT is trained in an inferred dynamical time scale rather than in physical time. External half-life measurements are therefore used as soft calibration anchors. They preserve relative degradation-rate structure while selecting a common conversion between physical and model time. They are not treated as exact cell-type-specific kinetic measurements.

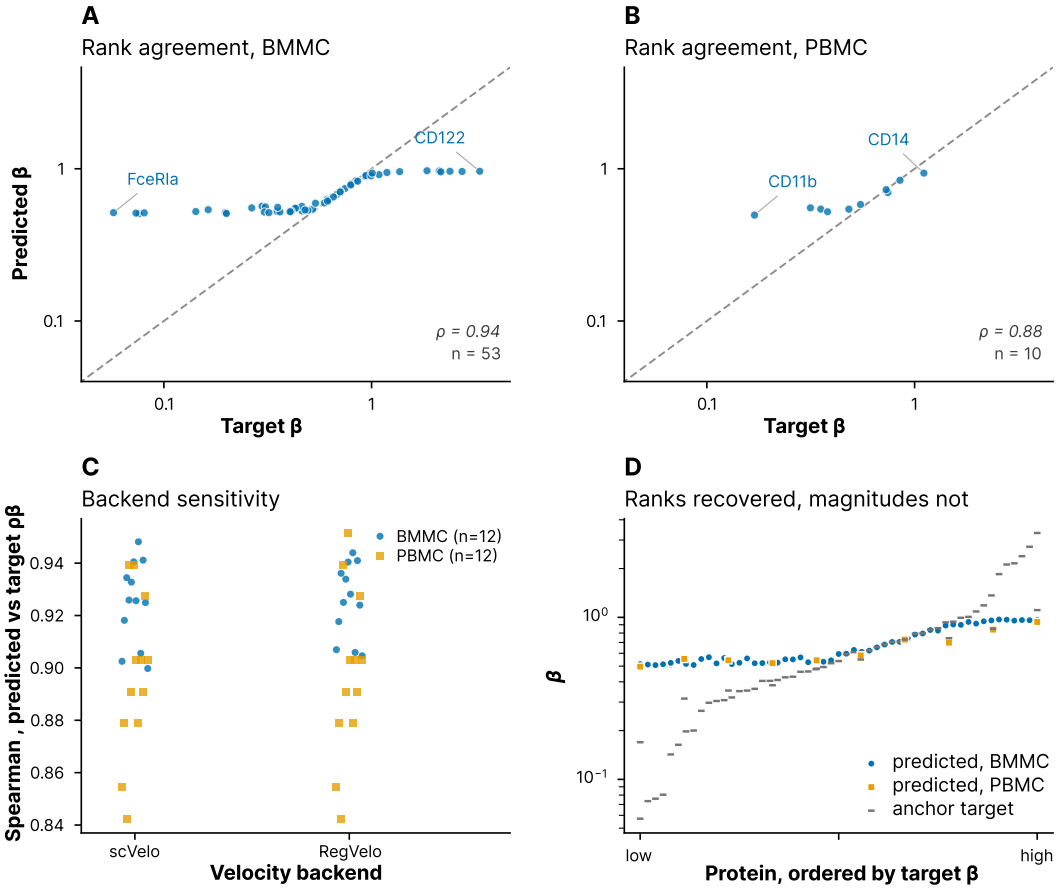


Figure 15: **Biochemical half-life anchors and degradation-rate ordering.** (A,B) Learned protein degradation rates β versus converted half-life targets on BMMC and PBMC. (C) Rank agreement between fitted and anchored degradation rates under scVelo and RegVelo source fields. (D) BMMC fitted and target β values after ordering proteins by the anchor target. External half-life measurements calibrate kinetic scale and relative degradation structure. They are not cell-level correspondence supervision.

The empirical behavior of the soft anchor prior is illustrated in Figure 15. The learned degradation parameters preserve the relative ordering of the external half-life references, but they do not collapse onto the reference values exactly. This is expected as the references enter as a soft calibration term inside a jointly optimized effective kinetic model, while the remaining kinetic components can compensate within the constraints discussed in Theorem 2.7. The anchors should therefore be read as scale and relative-rate information rather than direct supervision of the cross-modal map.

Protein degradation anchors. For protein h, a measured physical half-life $t^{\mathrm{prot}}_{1/2,h}$ defines the physical degradation rate

$$\beta^{\mathrm{phys}}_h = \frac{\log 2}{t^{\mathrm{prot}}_{1/2,h}}. \tag{23}$$

Let $\mathcal{H}_{\text{full}}$ denote the full trusted protein-anchor set and let β_0 be a nominal degradation rate in model time. We convert physical rates to the KOT time scale using a single dataset-level multiplicative factor

$$a_\beta = \frac{\beta_0}{\exp\left(\frac{1}{|\mathcal{H}_{\text{full}}|}\sum_{h\in\mathcal{H}_{\text{full}}}\log\beta_h^{\text{phys}}\right)}, \qquad \beta_h^\star = a_\beta\beta_h^{\text{phys}}. \tag{24}$$

Thus the geometric mean of the converted reference rates is β_0, while the relative fast–slow structure of the external measurements is preserved.

The conversion factor a_β is computed once from the full trusted anchor set and is then held fixed for all experiments on that dataset, including the anchor-subset ablations in Appendix A.8. Consequently, varying the number of active anchors changes only how much degradation-rate information is supplied to the model. It does not redefine the physical-to-model time conversion.

As described in Sec. 2.2, the converted targets enter the RNA–protein model through the soft log-rate prior

$$\mathcal{L}_\beta = \frac{1}{2}\frac{\sum_{h\in\mathcal{H}_\beta} w_h\left(\frac{\log\beta_h-\log\beta_h^\star}{\tau_h}\right)^2}{\sum_{h\in\mathcal{H}_\beta} w_h}. \tag{25}$$

Here $\mathcal{H}_\beta \subseteq \mathcal{H}_{\text{full}}$ denotes the active anchor subset, w_h is the anchor weight, and τ_h controls the tolerance assigned to anchor h.

RNA degradation anchors for chromatin–RNA KOT. For the chromatin–RNA experiments, RNA half-lives are converted in the same manner. For gene h,

$$\gamma_h^{\text{phys}} = \frac{\log 2}{t_{1/2,h}^{\text{RNA}}}. \tag{26}$$

Let $\mathcal{H}_{\gamma,\text{full}}$ denote the full trusted RNA half-life reference set and let γ_0 be the nominal RNA degradation rate in model time. We define

$$a_\gamma = \frac{\gamma_0}{\exp\left(\frac{1}{|\mathcal{H}_{\gamma,\text{full}}|}\sum_{h\in\mathcal{H}_{\gamma,\text{full}}}\log\gamma_h^{\text{phys}}\right)}, \qquad \gamma_h^\star = a_\gamma\gamma_h^{\text{phys}}. \tag{27}$$

The chromatin implementation uses the weighted mean squared error in log-rate space

$$\mathcal{L}_\gamma = \frac{\sum_{h\in\mathcal{H}_\gamma} w_h\left(\log\gamma_h-\log\gamma_h^\star\right)^2}{\sum_{h\in\mathcal{H}_\gamma} w_h}. \tag{28}$$

As for the protein anchors, a_γ is fixed once from the complete trusted reference set and is not recomputed when the active anchor subset changes.

Interpretation. For RNA–protein KOT, the degradation anchors select one global rate scale along the multiplicative kinetic gauge described in Theorem 2.7. They do not remove all kinetic non-identifiability and should not be interpreted as uniquely identifying α_ω, κ_ψ, or every individual degradation parameter. The chromatin RNA-half-life anchors serve the analogous role of softly calibrating degradation structure in the transcription–splicing–degradation model. In both cases, the anchors provide external scale information without entering the cross-modal correspondence supervision itself.

A.10 EFFECT OF ANCHOR COUNT

The biochemical half-life anchors are intended to calibrate the kinetic rate scale rather than to provide cross-modal correspondence information. We therefore vary the fraction of active protein degradation-rate anchors from zero to the complete trusted reference set while holding the remaining model and optimization configuration fixed. For every fraction, the physical-to-model conversion factor a_β is computed from the full trusted set as described in Appendix A.9 and is not recomputed for the selected subset. The ablation therefore changes the amount of biochemical rate information supplied to the model without changing the underlying time-scale conversion.

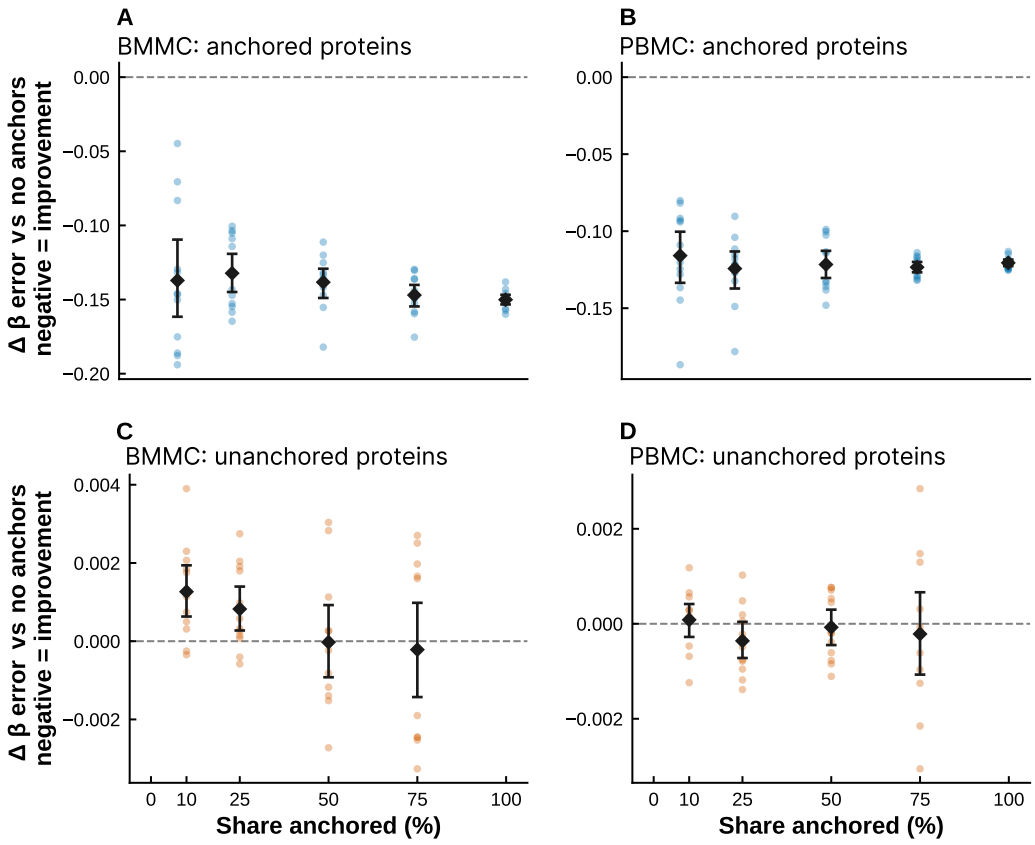


Figure 16: **Ablation on biochemical anchors used in KOT.** (A,B) Mean absolute β error on anchored and unanchored proteins as the number of active half-life anchors increases on BMMC and PBMC. (C,D) Change in error on the proteins that remain unanchored relative to the zero-anchor model. Negative values denote improvement. Points and error bars summarize 12 repeated runs.

The effect is localized primarily to the parameters that are actually constrained (Figure 16). Increasing the number of half-life references improves agreement on the active anchor set, but produces little systematic improvement in degradation-rate error on proteins that remain unanchored. This is consistent with the intended role of the prior, it supplies external rate information rather than creating a learned cell-level correspondence signal or uniquely determining every degradation parameter.

Cell-level alignment is essentially insensitive to the number of active anchors (see Table 15). Across the full sweep, BMMC FOSCTTM remains between 0.128 and 0.129, while PBMC remains between 0.135 and 0.136. Internal kinetic agreement is similarly stable: JVP–RHS varies only from 0.951 to 0.957 on BMMC and from 0.725 to 0.736 on PBMC. The learned clock scale κ also changes little across anchor fractions. Thus the improvement in cross-modal alignment is not driven by providing a large collection of protein half-life measurements.

The β MAE column should be interpreted differently from the alignment metrics. It measures agreement with the degradation-rate targets imposed on the *active* anchor subset and is therefore a calibration diagnostic, not an independent measure of kinetic truth. Because the evaluated set of proteins changes with the anchor fraction, its values across rows do not form a monotone error curve over a fixed collection of proteins. In particular, increasing the number of anchors introduces additional rate targets rather than simply adding repeated constraints on the same parameters.

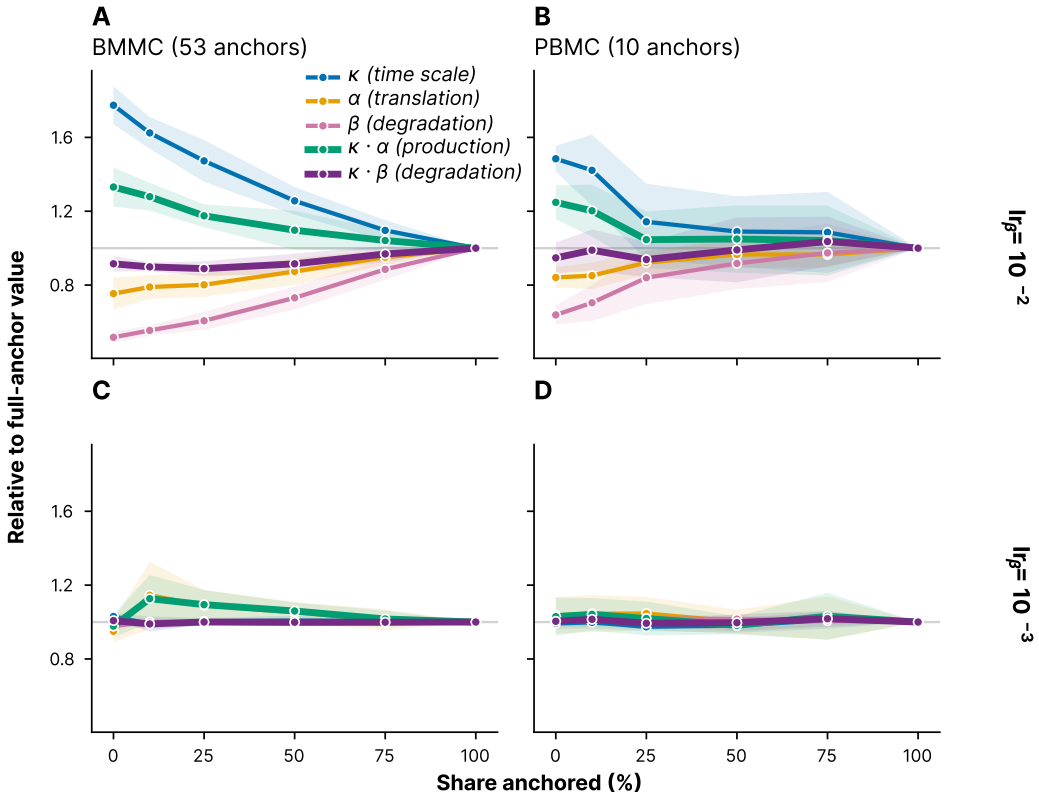


Figure 17: **The effect of kinetic components.** Median κ, α, β, and effective products $\kappa\beta$ and $\kappa\alpha$ are shown as a function of the retained protein-anchor fraction, normalized to their values with the complete anchor set. (A,B) BMMC and PBMC under the optimization regime with degradation prior. (C,D) Corresponding sweeps under the weaker-β optimization regime. Shaded bands summarize across 12 seeds.

These results support a separation between dynamical information and biochemical scale calibration. The source vector field and kinetic consistency constraint provide the first-order information used to disambiguate the cross-modal map, whereas the half-life measurements constrain the otherwise ambiguous kinetic scale and relative degradation structure. This is consistent with Theorem 2.7, which assigns the biochemical anchors the role of selecting a scale along the global multiplicative gauge. Empirically, the nearly unchanged FOSCTTM across the anchor-count sweep shows that this calibration is not responsible for the observed correspondence gain.

The 0%-anchor condition does not contradict the scale ambiguity in Theorem 2.7. Without biochemical anchors, the kinetic residual alone still cannot identify an absolute rate scale; however, the implemented model also contains the weak clock-scale regularizer $\mathcal{L}_\kappa$ from Sec. 2.2, which keeps optimization numerically well behaved along this direction. Accordingly, the stable alignment obtained without half-life anchors should be interpreted as evidence that biochemical calibration is not responsible for the correspondence gain, rather than as evidence that the individual kinetic parameters become uniquely identifiable without anchors.

Figure 17 provides a complementary view of the same experiment. When anchor information is weakened, the individual fitted components κ, α, and β can move substantially even though the state alignment and internal kinetic agreement remain nearly unchanged. The effective rate combination

Table 15: **Protein half-life anchor calibration.** Alignment, kinetic consistency, anchor error, and learned clock scale as the fraction of degradation-rate anchors changes. Anchor error measures agreement with imposed targets, not independent kinetic truth.

	BMMC				PBMC			
Anchors	FOS ↓	JVP–RHS ↑	β MAE ↓	κ	FOS ↓	JVP–RHS ↑	β MAE ↓	κ
0%	0.129±0.004	0.957±0.003	–	0.028±0.001	0.136±0.006	0.736±0.024	–	0.017±0.002
10%	0.128±0.004	0.951±0.005	0.227±0.180	0.027±0.001	0.135±0.005	0.731±0.023	0.142±0.124	0.017±0.002
25%	0.128±0.004	0.952±0.003	0.241±0.117	0.028±0.001	0.135±0.006	0.733±0.024	0.092±0.064	0.016±0.002
50%	0.129±0.003	0.953±0.003	0.259±0.077	0.028±0.001	0.135±0.005	0.730±0.028	0.103±0.042	0.016±0.002
75%	0.129±0.004	0.952±0.003	0.285±0.040	0.028±0.001	0.136±0.004	0.731±0.027	0.116±0.018	0.017±0.002
100%	0.129±0.004	0.952±0.002	0.298±0.005	0.028±0.001	0.135±0.006	0.725±0.026	0.123±0.005	0.017±0.002

12 seeds throughout. Anchor rows correspond to the nearest fractions of each panel's full reference list (53 BMMC; 10 PBMC).

entering the ODE is considerably more stable. This is the finite-sample behavior expected from a parameterization with compensatory kinetic freedom. The data constrain the effective field more directly than they constrain a unique decomposition into its individual rate components.

A.11 SYNTHETIC KINETIC-ORACLE HIERARCHY

The synthetic Oracle setting provides access to the exact kinetic quantities used by the simulator, allowing us to separate limitations of the learned cross-modal map from limitations introduced by estimating the kinetic model. We therefore compare standard KOT with a hierarchy of privileged variants in which increasingly more components of the translation–degradation law are supplied from the simulator rather than learned from data. These variants are diagnostic experiments and should not be interpreted as comparable baselines, since they receive information unavailable in the real-data setting.

Table 16: **Synthetic kinetic-oracle hierarchy.** Oracle-stage FOSCTTM as progressively more components of the translation–degradation model are learned rather than supplied from the simulator. The kinetic-oracle variants receive privileged simulator information and are included as controlled diagnostics rather than comparable baselines. Values are mean ± standard deviation over 12 seeds.

Variant	Oracle FOSCTTM ↓
KOT (anchored=full)	0.030±0.002
KOT kinetic oracle	0.020±0.001
KOT oracle, learned α	**0.019**±0.001
KOT oracle, learned α, κ	0.053±0.004

Providing the simulator kinetics improves alignment relative to the standard learned model, reducing FOSCTTM from 0.030 ± 0.002 to 0.020 ± 0.001 (Table 16). Allowing the production-rate component α to be learned while retaining the remaining privileged kinetic information does not degrade alignment. FOSCTTM is 0.019 ± 0.001. Thus, in this controlled setting, the state-dependent production component can be estimated without introducing a substantial correspondence error.

The behavior changes when both the production component α and the clock κ are learned. FOSCTTM increases to 0.053 ± 0.004, despite the underlying simulator remaining unchanged. This result is consistent with the additional flexibility of the kinetic decomposition making the joint estimation problem harder. In particular, α and κ participate in the multiplicative rate-scale ambiguity discussed in Theorem 2.7, so the experiment should not be interpreted as identifying either component individually.

The hierarchy therefore distinguishes two sources of error. With privileged kinetics, the cross-modal map itself can recover the Oracle correspondence to approximately 0.02 FOSCTTM. The additional error of the standard model is introduced by jointly learning the mapping and effective kinetic law, rather than by an inability of the map architecture to represent the underlying correspondence. At the same time, the degradation observed when both α and κ are learned shows that greater kinetic flexibility is not automatically beneficial: it can enlarge the finite-sample estimation problem even when the dynamical model class contains the true simulator.

A.12 Additional CRISPR perturbation diagnostics and fully in-silico prediction

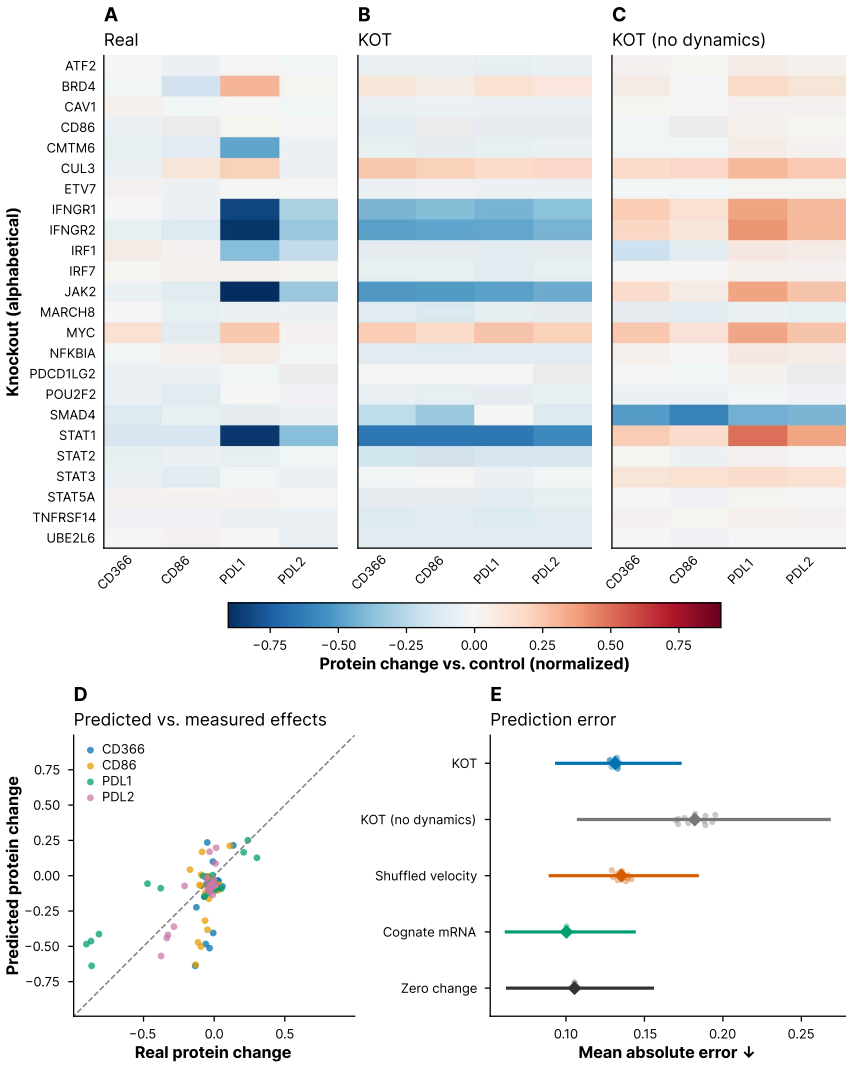


Figure 18: **Protein responses to unseen CRISPR perturbations.** (A) Observed perturbation-level changes in the four measured surface proteins across the eligible non-self knockouts. (B) Protein effects predicted by KOT, trained only on non-targeting cells. (C) Predictions after removing the kinetic objective. (D) Predicted versus observed effects for the 94 primary perturbation–protein pairs. (E) Mean absolute error on the effects on KOT.

Observed-RNA perturbation transfer. The aggregate correlations reported in the main text are also visible at the level of individual perturbation–protein effects (Figure 18). KOT recovers structured positive and negative responses across the four assayed proteins, whereas removing the kinetic objective produces a qualitatively different effect pattern and substantially larger magnitude error. The scatter and mean-absolute-error diagnostics further show that the result is not solely a consequence of the rank ordering of a small number of large effects.

The primary CRISPR experiment in Sec. 3 evaluates whether a KOT map learned only on non-targeting (NT) cells can translate an *observed* perturbed RNA state into the corresponding protein effect. We additionally consider a more demanding fully in-silico setting in which the perturbed RNA state must first be predicted. Starting from an NT RNA state and perturbation identity g, an

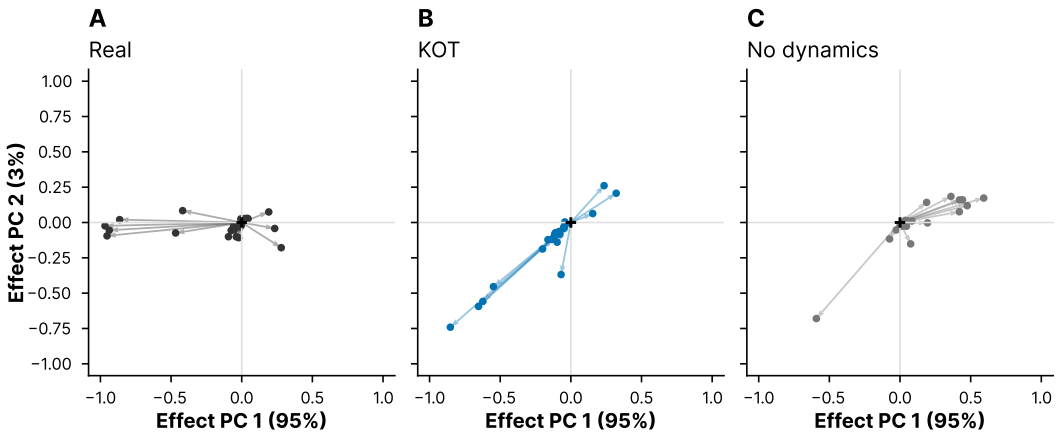


Figure 19: **Low-dimensional geometry of CRISPR protein-effect predictions.**

upstream in-silico perturbation model Ψ predicts

$$\widehat{r}_{\mathrm{KO},g} = \Psi(r_{\mathrm{NT}}, g),$$

which is then passed through the same frozen NT-trained KOT map,

$$\widehat{r}_{\mathrm{KO},g} \xrightarrow{\phi} \widehat{p}_{\mathrm{KO},g}.$$

The resulting perturbation-level protein effects are evaluated against the same held-out ECCITE-seq effects used in the observed-RNA experiment.

Figure 19 provides a complementary multivariate view. The kinetic and no-dynamics maps induce different organizations of the perturbation-effect vectors, showing that the dynamical objective changes more than individual scalar effect predictions. We treat this projection as a qualitative diagnostic and rely on the original-coordinate effect metrics for quantitative comparison.

Heterogeneous upstream supervision. We evaluate linear prediction, GEARS, RegVelo, scGPT, and MultiPert as upstream perturbation models. These methods do not receive comparable supervision. The linear model predicts a held-out replicate of a perturbation observed in other replicates, RegVelo is fitted on NT cells and supports a restricted set of regulator perturbations mechanistically, GEARS and scGPT use observed perturbed RNA during training or fine-tuning, and MultiPert receives still stronger perturbational supervision. We therefore treat Table 17 as an exploratory end-to-end diagnostic rather than as a ranking of perturbation-prediction methods.

End-to-end protein-effect prediction. Among the evaluated pipelines, the linear upstream model gives Spearman 0.456 ± 0.031, Pearson 0.604 ± 0.025, and NRMSE 0.803 ± 0.017 after its predicted RNA states are propagated through the frozen KOT map. GEARS and MultiPert retain positive rank association (0.324 ± 0.139 and 0.400 ± 0.007, respectively), whereas RegVelo and scGPT perform substantially worse in this end-to-end setting. Importantly, these differences conflate two sources of error, failure of the upstream model to predict the perturbation-induced RNA state and error in translating that predicted state through the frozen RNA–protein map. ISP therefore does not isolate KOT itself in the way that the observed-RNA experiment does.

The zero-change reference further illustrates this distinction. Several upstream models obtain protein-effect errors worse than predicting no perturbation at all, even though the same frozen KOT map performs well when the experimentally observed perturbed RNA is supplied. Thus, in the fully in-silico pipeline, the dominant bottleneck can occur before cross-modal translation. Once the upstream model fails to produce the correct perturbation-specific RNA displacement, KOT cannot reconstruct information that is absent from its input.

Figure 20 provides a controlled view of the cross-modal component of this end-to-end pipeline. Holding the upstream RNA perturbation predictor fixed, the full KOT map retains positive rank association with the measured protein effects and lower normalized error than the corresponding

Table 17: **Exploratory end-to-end CRISPR perturbation prediction.** (A) Protein effects obtained by passing predicted perturbation RNA states through the frozen NT-trained KOT map. Upstream methods receive heterogeneous perturbational supervision and are therefore not directly comparable. (B) Between-perturbation diversity of representative predicted RNA profiles. Small variance ratios and pairwise correlations near one indicate collapse toward a common response.

A. Common perturbation–protein pairs

Method	Spearman ↑	Pearson ↑	NRMSE ↓
GEARS	0.324±0.139	0.370±0.262	1.237±0.173
Linear	**0.456**±0.031	**0.604**±0.025	**0.803**±0.017
Zero change	–	–	1.075±0.000
Multipert	0.400±0.007	0.553±0.013	1.749±0.115
RegVelo	-0.036±0.046	0.213±0.030	1.376±0.046
scGPT	0.048±0.070	0.303±0.014	4.024±0.298

B. Predicted-profile diversity (representative intact files)

Method	Between-perturbation ratio ↑	Pairwise profile r ↓
GEARS	**0.033**	**0.990**
Linear	0.011	0.997
Multipert	0.010	0.994
RegVelo	0.001	1.000
scGPT	0.001	1.000

Panel B uses one representative profile file per method and is diagnostic rather than a seed-average estimate.

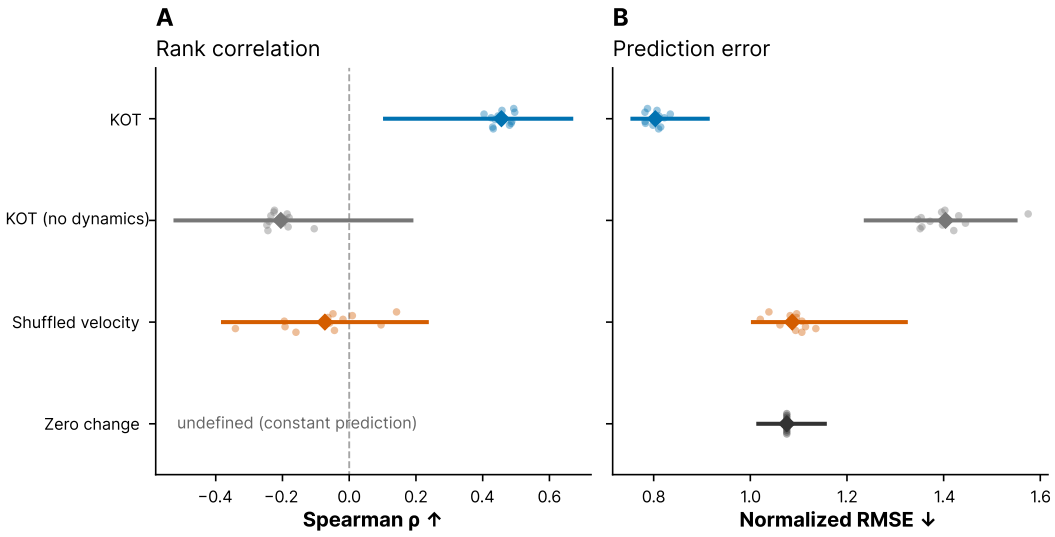


Figure 20: **Kinetic ablation on cross-modal translation.** Leave-one-replicate-out evaluation on perturbations observed. (A) Spearman correlation between predicted and observed protein effects. (B) Normalized RMSE.

no-dynamics and shuffled-velocity variants. The no-dynamics model loses the positive perturbation ranking, while shuffling the source velocity also degrades downstream protein prediction. Thus the performance differences in this comparison cannot be attributed to a different upstream estimate of the perturbed RNA state: they arise after that state is passed through differently trained RNA–protein maps.

This result also clarifies the relationship between the observed-RNA and fully in-silico experiments. Predicting the perturbed RNA state upstream introduces an additional source of error, so the end-to-end task is necessarily harder than supplying experimentally observed perturbed RNA. Nevertheless, the same qualitative dependence on the kinetic constraint is retained. The first-order structure learned from non-targeting cells therefore continues to influence intervention-response transfer even when the input to the frozen cross-modal map is itself predicted rather than observed.

Predicted-response collapse. The profile-diversity diagnostics in Table 17 reveal an additional limitation that is obscured by aggregate correlation metrics. Across the tested perturbation models, the between-perturbation variance is only 0.1–3.3% of the corresponding within-profile scale, while pairwise correlations between predicted perturbation profiles range from 0.990 to 1.000. The predicted RNA responses are therefore highly similar across perturbations. Even the linear model, which gives the strongest downstream protein-effect metrics in this experiment, has a between-perturbation ratio of only 0.011 and pairwise profile correlation 0.997.

This collapse changes how the end-to-end correlations should be interpreted. A model may recover a common perturbational direction or average response and thereby obtain positive aggregate effect correlation without accurately recovering perturbation-specific transcriptomic states. We therefore regard ISP as evidence about the current boundary of the complete perturbation-to-RNA-to-protein pipeline, rather than as evidence that the upstream intervention models recover detailed perturbation-specific trajectories. Panel B is based on representative intact prediction files and is consequently diagnostic rather than a seed-averaged performance estimate.

Additional Papalexi diagnostics. Table 18 provides two complementary checks on the primary observed-RNA experiment. First, permutation-null correlations are close to zero at both replicate and pooled aggregation, whereas the observed cognate-mRNA and paired-translator associations remain positive. This confirms that the reported perturbation-effect correlations are not produced by the evaluation aggregation alone.

Table 18: **Additional CRISPR transfer and NT-alignment diagnostics.** (A) Permutation-null checks for observed-RNA perturbation-effect prediction at replicate and pooled aggregation. (B) Cell-level alignment and kinetic diagnostics on the non-targeting Papalexi population. Papalexi NT alignment is reported as a diagnostic and is not part of the BMMC/PBMC CITE-seq benchmark.

A. Permutation-null check

Method	Aggregation	Pairs	Spearman	Pearson	Null mean
Cognate mRNA	Replicate	282	0.398	0.424	-0.001
Ridge	Replicate	282	0.342	0.453	0.002
MLP	Replicate	282	0.209	0.408	0.001
Cognate mRNA	Pooled	94	0.415	0.610	-0.002
Ridge	Pooled	94	0.462	0.714	-0.002
MLP	Pooled	94	0.268	0.573	-0.002

B. Papalexi NT alignment

Method	FOSCTTM ↓	JVP–RHS ↑	ODE residual ↓
KOT	0.489±0.003	1.000±0.000	0.028±0.002
KOT (permuted S)	0.475±0.003	0.997±0.001	0.112±0.014
KOT (reverse v)	0.496±0.003	0.997±0.001	0.089±0.009
KOT (shuffled v)	0.504±0.004	0.988±0.004	0.133±0.017
KOT (zero v)	0.509±0.006	0.000±0.000	1.000±0.000
KOT no dynamics	0.515±0.008	0.969±0.015	0.904±0.014
MLP (paired NT)	0.182±0.055	–	–
Ridge (paired NT)	0.301	–	–
scButterfly (paired NT)	0.076±0.003	–	–
sciPENN (paired NT)	0.489±0.001	–	–
totalVI (paired)	0.000±0.000	–	–

Second, Papalexi NT cell-level alignment is substantially more difficult than the BMMC and PBMC CITE-seq benchmarks. Unpaired KOT obtains FOSCTTM 0.489 ± 0.003, only modestly below the random-scale value of 0.5, whereas paired NT translators can exploit cell-level supervision to obtain considerably lower retrieval error. The Papalexi experiment should therefore not be interpreted as a strong global cell-alignment result. Its primary role is instead perturbational. Despite weak global NT retrieval, the frozen map and its local differential retain signal for predicting population-level protein responses to unseen interventions.

The kinetic diagnostics reinforce the need to separate these notions. Standard KOT attains JVP–RHS agreement 1.000 ± 0.000 with normalized ODE residual 0.028 ± 0.002, whereas removing the dynamical objective leaves a high directional cosine (0.969 ± 0.015) but a much larger residual (0.904 ± 0.014). Directional agreement alone therefore does not establish a well-matched kinetic field. Magnitude-sensitive residuals and downstream perturbational behavior provide complemen-

tary information. Together, these diagnostics show that global state retrieval, internal kinetic consistency, and intervention-response transfer probe distinct properties of the learned map.

A.13 CHROMATIN-TO-RNA KINETIC TRANSPORT

We instantiate the same construction from subsection 2.1 for the causal ordering chromatin → nascent RNA → mature RNA. Let

$$\mathbf{z} = h_{\mathrm{map}}(\mathbf{c}), \qquad \mathbf{q} = h_{\mathrm{reg}}(\mathbf{c}),$$

denote respectively the chromatin coordinates used by the cross-modal map and the features supplied to the transcription-rate model. The source vector field $\mathbf{v}_z$ is always expressed in the same coordinates $\mathbf{z}$ on which the map is differentiated. We predict

$$\Phi_\theta(\mathbf{z}) = [\widehat{\mathbf{u}}, \widehat{\mathbf{s}}], \tag{29}$$

where $\mathbf{u}$ and $\mathbf{s}$ denote unspliced and spliced RNA.

We use the transcription–splicing–degradation relay

$$\dot{\mathbf{u}} = \kappa_\psi(\mathbf{z}) \left[\boldsymbol{\alpha}_\omega(\mathbf{q}) - \boldsymbol{\beta} \odot \mathbf{u}\right], \tag{30}$$

$$\dot{\mathbf{s}} = \kappa_\psi(\mathbf{z}) \left[\boldsymbol{\beta} \odot \mathbf{u} - \boldsymbol{\gamma} \odot \mathbf{s}\right], \tag{31}$$

with positive splicing rates $\boldsymbol{\beta}$, RNA-degradation rates $\boldsymbol{\gamma}$, and state-dependent transcription $\boldsymbol{\alpha}_\omega$. KOT therefore requires

$$\mathbf{J}_{\Phi_\theta}(\mathbf{z})\mathbf{v}_z \approx \kappa_\psi(\mathbf{z}) \begin{bmatrix} \boldsymbol{\alpha}_\omega(\mathbf{q}) - \boldsymbol{\beta} \odot \widehat{\mathbf{u}} \\ \boldsymbol{\beta} \odot \widehat{\mathbf{u}} - \boldsymbol{\gamma} \odot \widehat{\mathbf{s}} \end{bmatrix}, \tag{32}$$

with the corresponding measurement-coordinate transformation applied when RNA states are represented on a transformed scale.

The spliced-RNA marginal provides the primary distributional constraint,

$$\mathcal{L}^{(s)}_{\mathrm{align}} = \mathcal{S}_\varepsilon\left((\Pi_s \circ \Phi_\theta)_{\#}\widehat{\mu}_C,\, \widehat{\mu}_S\right), \tag{33}$$

while an unpaired auxiliary loss $\mathcal{L}_u$ constrains the unspliced state. The objective is

$$\mathcal{L}_{\mathrm{chromatin}} = \mathcal{L}^{(s)}_{\mathrm{align}} + \lambda_u \mathcal{L}_u + \lambda_{\mathrm{dyn}} \mathcal{L}^{(\mathrm{chr})}_{\mathrm{dyn}} + \rho_\gamma \mathcal{L}_\gamma + \rho_\kappa \mathcal{L}_\kappa + \rho_{\mathrm{net}} \mathcal{L}_{\mathrm{net}}, \tag{34}$$

where $\mathcal{L}^{(\mathrm{chr})}_{\mathrm{dyn}}$ is Eq. equation 3 applied to the joint (u, s) field and $\mathcal{L}_\gamma$ softly anchors RNA degradation using external mRNA half-lives. Independently estimated target RNA velocities are never used in the primary KOT objective and are reserved for evaluation.

The trajectory geometry, JVP coordinates, and regulatory representation need not coincide because they serve different geometric and biochemical roles. We construct trajectory geometry in ATAC LSI space, require the map and source vector field to share coordinates, and provide gene-associated chromatin activity separately to the transcription head. The exact coordinate transformations, source-field construction, RNA measurement convention, auxiliary loss L_u, and RNA anchors are detailed below.

Role-specific representations. We distinguish

$$\mathbf{z}_{\mathrm{geom}} = h_{\mathrm{geom}}(\mathbf{c}), \qquad \mathbf{z} = h_{\mathrm{map}}(\mathbf{c}), \qquad \mathbf{q} = h_{\mathrm{reg}}(\mathbf{c}),$$

used respectively for trajectory geometry, the learned map/JVP, and transcription-rate prediction. The geometry representation determines local neighbors and trajectory direction but is not required to coincide with the coordinates in which the map is differentiated.

For BMMC, let τ_i denote ATAC-derived pseudotime and $\mathcal{N}_i^+$ the forward neighbors obtained in $\mathbf{z}_{\mathrm{geom}}$ space. The vector field supplied to the JVP is computed in the actual map coordinates,

$$\mathbf{v}_{z,i} = \frac{\sum_{j \in \mathcal{N}_i^+} w_{ij}(\mathbf{z}_j - \mathbf{z}_i)/(\tau_j - \tau_i)}{\sum_{j \in \mathcal{N}_i^+} w_{ij}}. \tag{35}$$

This guarantees coordinate consistency between $\mathbf{J}_{\Phi_\theta}(\mathbf{z})$ and $\mathbf{v}_z$.

RNA measurement coordinates. Unspliced and spliced RNA are normalized using a shared per-cell factor. For the primary log-transformed representation,

$$y_u = \log(1+u), \qquad y_s = \log(1+s),$$

and the corresponding kinetic derivatives satisfy

$$\dot{y}_u = \frac{\dot{u}}{1+u}, \qquad \dot{y}_s = \frac{\dot{s}}{1+s}. \tag{36}$$

Equation equation 32 is evaluated in these measurement coordinates using Eq. equation 36. Alternative coordinate conventions and the explicit compositional correction are evaluated as sensitivity analyses.

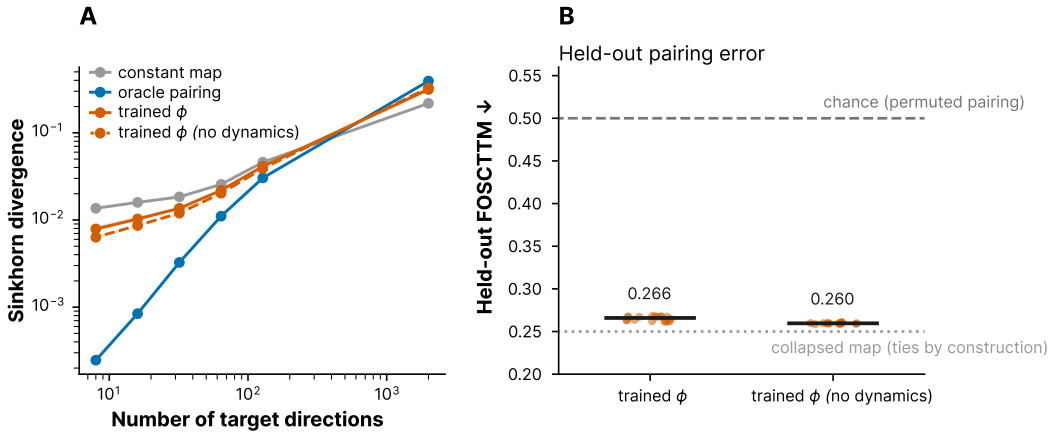


Figure 21: **Marginal distribution on cell-level chromatin–RNA alignment.** (A) Sinkhorn divergence between mapped and target RNA states dependent on the number of target directions used to evaluate the divergence. (B) Held-out spliced-RNA FOSCTTM for KOT and no-dynamics mappings.

Marginal fit and correspondence can decouple. Figure 21 illustrates a failure mode that is particularly relevant in the chromatin setting. Increasing the number of target directions clearly distinguishes the trained map from a constant-map reference under the Sinkhorn objective, showing that the model does learn nontrivial target-distribution structure. However, this improvement in marginal fit does not produce a corresponding improvement in held-out cell-level retrieval as FOSCTTM changes only modestly across alignment dimensions and sample sizes and remains much closer to the constant-map reference than to an informative paired correspondence.

This distinction motivates the evaluation strategy used below. A low distributional loss is evidence that the mapped population reproduces the target marginal in the chosen representation, but it is not evidence that individual chromatin states have been assigned to their correct RNA states. We therefore evaluate static correspondence separately from the independent first-order JVP–scVelo diagnostic and use the representation screens below to determine whether the differentiated map supports biologically oriented target-space dynamics.

Auxiliary unspliced constraint. Because the primary Sinkhorn term aligns the spliced block, we constrain the unpaired unspliced marginal through per-gene moments,

$$\mathcal{L}_u = \frac{1}{D_u}\sum_{j=1}^{D_u}\left[\left(\frac{\mu(\widehat{u}_j)-\mu(u_j)}{s_{u,j}}\right)^2 + \left(\frac{\sigma(\widehat{u}_j)-\sigma(u_j)}{s_{u,j}}\right)^2\right]. \tag{37}$$

Predicted and observed moments are computed over separate source and target populations and therefore do not introduce cell-level pairing.

Independent kinetic evaluation. RNA velocities inferred from the target RNA modality are never supplied to the primary chromatin KOT objective. After training, we compare

$$\left[\mathbf{J}_{\Phi_\theta}(\mathbf{z})\mathbf{v}_z\right]_s$$

against independently inferred spliced-RNA velocities. This therefore tests the learned differential rather than supervising it. The left-hand side is therefore not an estimated target velocity but the target-space dynamics implied by the candidate correspondence.

A.13.1 CHROMATIN EXPERIMENTAL RESULTS

Tables 20–22 report the controlled representation and configuration screen used during training, whereas Table 19 reports the frozen final configuration over the full 12 seed evaluation. Their absolute values should therefore not be interpreted as repeated estimates of the same run.

Table 19: **Full BMMC chromatin–RNA alignment and kinetic diagnostics.** The fraction of samples closer than the true match (s-FOSCTTM) is computed for the spliced-RNA component. Gene r denotes the paired gene-level Pearson correlation, and partner diversity measures the diversity of selected target partners, with larger values indicating less concentrated matching. For KOT, JVP–scVelo is the median centred cosine between the pushed-forward chromatin field and independently inferred spliced-RNA velocity. Null p_{95} is the 95th percentile of the permutation null, Δ_{null} is their difference, and JVP–RHS measures agreement between the Jacobian–vector product and the mechanistic right-hand side. Target RNA velocity is never supplied to KOT during training. Static competitor metrics are evaluated in method-specific matching spaces and should therefore be interpreted as alignment diagnostics rather than as a like-for-like kinetic comparison. The reported scGLUE result uses the input-parity adapter; the canonical peak-level implementation is discussed separately.

Method	s-FOSCTTM↓	Gene r↑	Partner diversity ↑	JVP–scVelo ↑	Null p_{95}	Δ_{null}↑	JVP–RHS↑
KOT ($\lambda_{dyn} = 0$)	0.2210 ± 0.0009	0.0589 ± 0.0007	0.0231 ± 0.0019	0.0054 ± 0.0077	0.0054 ± 0.0064	-0.0001 ± 0.0014	-0.0345 ± 0.0048
KOT	0.2225 ± 0.0008	0.0582 ± 0.0008	0.0189 ± 0.0011	0.0210 ± 0.0039	0.0196 ± 0.0031	$+0.0014 \pm 0.0009$	0.0647 ± 0.0085
KOT (v shuffled)	0.2223 ± 0.0010	0.0582 ± 0.0006	0.0187 ± 0.0010	0.0200 ± 0.0046	0.0191 ± 0.0035	$+0.0009 \pm 0.0010$	0.0558 ± 0.0074
scGLUE†	0.1294 ± 0.0023	0.0863 ± 0.0010	0.4916 ± 0.0090	–	–	–	–
MaxFuse	0.1328 ± 0.0016	0.0824 ± 0.0011	0.4822 ± 0.0046	–	–	–	–
scDART	0.3007 ± 0.1265	-0.0016 ± 0.0004	0.5771 ± 0.0117	–	–	–	–
moscot	0.3522 ± 0.0034	0.0111 ± 0.0018	0.0597 ± 0.0026	–	–	–	–

†Current input-parity scGLUE adapter. A canonical peak-level scGLUE implementation is discussed separately. JVP–scVelo denotes median centred cosine against independently inferred spliced-RNA velocity; target RNA velocity is never used for KOT training.

Static alignment and first-order transfer. The chromatin experiment exposes a qualitatively different regime from the RNA–protein setting. Adding the kinetic objective does not improve static cell-level correspondence. Spliced-RNA FOSCTTM is 0.2210 ± 0.0009 without dynamics and 0.2225 ± 0.0008 for the full model (Table 19). Gene-level reconstruction is likewise nearly unchanged. Moreover, static integration methods such as scGLUE and MaxFuse obtain substantially lower FOSCTTM in their respective matching spaces. We therefore do not interpret the chromatin extension as a static alignment improvement.

The kinetic objective nevertheless changes the learned differential. Agreement between the pushed-forward chromatin field and independently inferred spliced-RNA velocity increases from 0.0054 ± 0.0077 without dynamics to 0.0210 ± 0.0039 with the full model, while JVP–RHS agreement changes from -0.0345 ± 0.0048 to 0.0647 ± 0.0085. The observed external JVP–scVelo score lies slightly above its permutation null, yielding $\Delta_{\mathrm{null}} = +0.0014 \pm 0.0009$. Since target RNA velocity is never used during training, this comparison provides an independent first-order evaluation of the differential induced by the learned map.

However, the same signal is largely retained after shuffling the ATAC-derived source field: JVP–scVelo remains 0.0200 ± 0.0046, with $\Delta_{\text{null}} = +0.0009 \pm 0.0010$. Thus the present experiment does not establish that the improvement is specific to the biological ordering encoded by the constructed chromatin field. The kinetic objective reproducibly induces a differential that is more aligned with independent RNA dynamics than the no-dynamics model, but the shuffled-field control shows that this effect cannot yet be attributed specifically to the supplied ATAC direction. This is the principal boundary of the current chromatin result.

The source-coordinate screen shows that representations suitable for static chromatin–RNA transport are not automatically suitable for first-order transport (Table 21). Raw, CP10K, L2-normalized, and TF–IDF gene-space representations all give non-positive absolute JVP–scVelo agreement. A positive permutation-null margin by itself is not sufficient in these cases. For example, TF–IDF gene space yields $\Delta_{\text{null}} = +0.0059$ while the absolute JVP–scVelo cosine remains negative (-0.0042). We therefore require both correctly oriented absolute agreement and comparison against the permutation null.

Among the single-representation configurations, TF–IDF/LSI is the only screened representation to produce positive independent first-order agreement, with JVP–scVelo 0.0242 and $\Delta_{\text{null}} = +0.0018$. Using this same representation for the transcription-rate head, however, drives approximately half of the α outputs to their lower bound (49.86%), indicating severe kinetic-head saturation. Supplying linear CP10K gene-activity features to the transcription head while retaining TF–IDF/LSI for the map and JVP removes this degeneracy: the α-floor fraction falls to 0.02%, while JVP–scVelo remains 0.0244 and the null margin increases slightly to $+0.0023$. This motivates the final role-specific construction in which trajectory geometry, differentiable map coordinates, and regulatory features are allowed to differ.

The representation diagnostics in Table 20 further show why no single representation criterion is sufficient. Raw accessibility is dominated by depth ($|\text{corr}(\text{PC1}, \text{depth})| = 0.9944$), whereas both CP10K-log and TF–IDF/LSI reduce this dependence. At the same time, the representation with the strongest paired chromatin-to-RNA velocity ceiling is not the one that gives the strongest unpaired KOT result. These paired quantities are therefore useful diagnostic ceilings, but they do not by themselves determine which coordinates will support a stable unpaired first-order transport model.

Table 20: BMMC representation diagnostics before the final sweep. The paired columns are diagnostic ceilings and are not used for KOT training.

Representation	Unpaired $F_s \downarrow$	PC1–depth $\lvert r \rvert$	Paired state ceiling $F_s \downarrow$	Paired $v_c \to v_s$ cosine $\uparrow$
Raw / as-is	0.2938	0.9944	0.1328	0.1013
CP10K log1p	0.2479	0.7478	0.1345	**0.1854**
TF–IDF/LSI	**0.2073**	0.7471	0.1472	0.0692

Insufficiency of internal kinetics. Targeted ablations provide several examples in which the mechanistic equation can be fit internally without producing the desired external first-order behavior (Table 22). Removing the residual gate, for example, increases the absolute JVP–RNA cosine to 0.0406 and gives JVP–RHS 0.0909, but its permutation-null 95th percentile rises even higher to 0.0423, producing a negative null margin. Similarly, a strong κ prior and fixed-κ model obtain JVP–RHS values of 0.1363 and 0.1423, respectively, while providing little or no positive separation from the external permutation null.

The most extreme example is the global-linear RNA-coordinate variant. It attains JVP–RHS 0.2433, substantially higher than the selected model, yet its independent JVP–RNA agreement is negative (-0.0087) with a strongly negative null margin (-0.0082). Thus high agreement between $J_\Phi v_z$ and the model's own mechanistic right-hand side can result from the two learned quantities adapting to one another without recovering the correct target-space direction. This mirrors the distinction observed in the RNA–protein ablations. Internal satisfaction of the kinetic equation is a necessary training diagnostic, but is not by itself evidence that the induced dynamics are biologically correct.

Instead, the selected configuration balances these diagnostics. It gives positive external JVP–RNA agreement (0.0244), a positive permutation-null margin ($+0.0023$), positive JVP–RHS agreement (0.0727), and essentially no α or κ floor saturation. We therefore select models using the conjunc-

Table 21: BMMC source-coordinate screen at $\lambda_{\mathrm{dyn}} = 30$. The same representation is used for the map/JVP and regulatory head except in the final row. A positive null margin with a negative absolute cosine is not considered evidence of correctly oriented target dynamics.

Map/JVP representation	$F_s \downarrow$	JVP–scVelo	Null p_{95}	Δ_{null}	κ floor (%)	α floor (%)
Raw / as-is	0.3000	-0.0085	-0.0088	+0.0002	91.25	0.00
CP10K linear	0.3833	-0.0144	-0.0061	-0.0084	0.00	0.00
CP10K log1p	0.2590	-0.0128	-0.0163	+0.0035	81.77	3.73
L_2 per cell	0.2455	-0.0069	-0.0064	-0.0005	0.00	0.00
TF–IDF gene space	0.2784	-0.0042	-0.0101	+0.0059	0.00	55.47
TF–IDF/LSI	0.2238	0.0242	0.0223	+0.0018	0.00	49.86
TF–IDF/LSI + linear regulatory input	**0.2177**	**0.0244**	0.0221	**+0.0023**	0.00	**0.02**

tion of external first-order orientation, null separation, internal kinetic consistency, and kinetic-head stability rather than maximizing any single quantity.

Table 22: Targeted BMMC chromatin to RNA failure analyses. F_s is spliced-RNA FOSCTTM, JVP–RNA is centred cosine against independent scVelo. JVP–RHS measures agreement with the chromatin–RNA transcription–splicing–degradation mechanistic field. Floor fractions diagnose kinetic collapse.

Variant	$F_s \downarrow$	JVP–RNA ↑	Null p_{95}	Δ_{null}	JVP–RHS ↑	κ floor (%)	α floor (%)	γ err. (%)
noDyn	0.2200	0.0125	0.0115	+0.0010	-0.0295	0.00	0.00	0.00
Selected full	**0.2177**	0.0244	0.0221	+0.0023	0.0727	0.00	0.02	7.28
Same representation for α	0.2238	0.0242	0.0223	+0.0018	0.0771	0.00	49.86	7.06
Legacy stabilization ($\lambda = 9$)	0.2226	0.0084	0.0080	+0.0004	-0.0098	3.71	59.92	0.23
Per-gene residual gate	0.2776	0.0114	0.0119	-0.0006	0.0838	0.00	0.00	12.55
No residual gate	0.3053	0.0406	0.0423	-0.0017	0.0909	0.00	0.00	7.92
Strong κ prior	0.2423	0.0185	0.0188	-0.0002	0.1363	0.00	0.001	55.61
Fixed $\kappa = \log 2$	0.2327	0.0231	0.0226	+0.0005	0.1423	0.00	0.006	87.61
$\lambda_u = 0.1$	0.2276	0.0242	0.0229	+0.0013	0.1135	0.00	0.01	9.97
$\lambda_u = 0$	0.2279	0.0144	0.0144	-0.0000	0.0862	0.00	0.003	8.06
$\lambda_\gamma = 0.1$	0.2288	0.0241	0.0217	+0.0023	0.0757	0.00	0.02	70.00
$\lambda_\gamma = 10$	0.2283	0.0241	0.0224	+0.0017	0.0747	0.00	0.02	0.71
Compositional RNA correction	0.2251	0.0296	0.0278	+0.0018	0.0118	0.00	0.0001	4.43
Global-linear RNA coordinates	0.2821	-0.0087	-0.0005	-0.0082	0.2433	0.00	0.009	6.68
$d_{\mathrm{align}} = 8$	0.2295	0.0143	0.0139	+0.0003	0.0590	0.00	0.002	7.75
$d_{\mathrm{align}} = 16$	0.2295	0.0242	0.0223	+0.0020	0.0750	0.00	0.019	7.41
$d_{\mathrm{align}} = 64$	0.2291	0.0246	0.0227	+0.0018	0.0697	0.00	0.020	7.18

Auxiliary constraints and coordinate sensitivity. The auxiliary unspliced constraint contributes specifically to first-order transfer. Removing it ($\lambda_u = 0$) lowers JVP–RNA from 0.0242 to 0.0144 and eliminates the positive null margin, whereas $\lambda_u = 0.1$ restores a positive margin. This supports retaining the unspliced block as an auxiliary unpaired constraint even though the primary distributional alignment objective is defined on spliced RNA.

By contrast, changing the RNA-degradation anchor weight has little effect on the external differential over the tested range: $\lambda_\gamma = 0.1$ and $\lambda_\gamma = 10$ both give JVP–RNA approximately 0.0241 with positive null margins. Their degradation-rate errors differ dramatically, however, showing that the anchor weight primarily controls kinetic calibration rather than the presence of the first-order transfer signal. This parallels the separation between biochemical calibration and correspondence observed in the RNA–protein experiments, although the chromatin relay contains a different kinetic parameterization.

The RNA measurement convention is also consequential. A compositional correction retains positive first-order transfer and increases the absolute JVP–RNA cosine to 0.0296, whereas global-linear RNA coordinates reverse the external direction despite producing a very strong internal kinetic fit. Finally, the result is not highly sensitive to moderate changes in the number of alignment dimensions. $d_{\mathrm{align}} = 16$ and 64 give similar external JVP agreement and positive null margins, while reducing the alignment dimension to 8 weakens the signal.

Taken together, these experiments identify the current scope of the chromatin extension. KOT can impose a reproducible first-order constraint on an unpaired chromatin–RNA map and produce a

differential with greater agreement to independently inferred RNA dynamics than the corresponding no-dynamics model. Achieving this behavior requires coordinate-consistent differentiation, role-specific chromatin representations, and auxiliary constraints that prevent kinetic degeneracy. At the same time, the persistence of the external signal under source-field shuffling shows that the present ATAC-derived directional field does not yet provide the source-specific disambiguation obtained in the RNA–protein experiments. We therefore view this experiment as evidence that the KOT construction transfers to a distinct molecular relay, together with a concrete characterization of the representation and source-field requirements that must be satisfied for stronger source-field-specific dynamical disambiguation.