DEFO: DECOMPOSED DYNAMICS FOR FORECASTING TEMPORAL SINGLE-CELL PERTURBATION RESPONSES

Anonymous authors
Paper under double-blind review

ABSTRACT

Forecasting future cellular trajectories for perturbations unseen during training from limited early-response observations is a fundamental challenge in single-cell perturbation modeling. Unlike conventional perturbation prediction that focuses on observed conditions or static endpoints, this task requires generalizing dynamic response patterns across perturbations while preserving shared cellular progression. We propose DeFo, a decomposed dynamics framework for held-out perturbation forecasting in temporal single-cell systems. DeFo introduces a control-referenced formulation that decomposes cellular evolution into a shared base field capturing perturbation-independent temporal progression and a perturbation-specific response component modeling deviations from the control trajectory. To infer future responses for unseen perturbations, we introduce Temporal Response Memory Attention (TRMA), a trajectory-valued memory retrieval mechanism that retrieves and softly composes temporal response patterns from known perturbations according to their early effects. Experiments on time-resolved chemical and genetic perturbation datasets demonstrate competitive recovery of population-level expression, control-referenced perturbation effects, and differential-expression directions. These results establish DeFo as a principled framework for learning transferable cellular dynamics and forecasting responses to unseen perturbations from limited temporal observations.

1 INTRODUCTION

Forecasting how cells respond to unseen perturbations is a fundamental challenge in single-cell computational modeling. Although numerous models have been developed for perturbation prediction, most focus on static or endpoint responses, whereas time-resolved single-cell perturbation experiments reveal transient responses, delayed effects, and eventual fate changes that cannot be captured by endpoint measurements alone Ishikawa et al. (2023); Xu et al. (2024); Zhang et al. (2024b). Beyond characterizing previously observed interventions, a more challenging goal is to predict how a new perturbation will reshape future cellular populations when only limited early-response observations are available. Such forecasting could enable computational prioritization of candidate interventions, mechanistic hypothesis generation, and more efficient experimental design Guo et al. (2025); Wei et al. (2026).

We consider a held-out perturbation forecasting setting, where trajectories of known perturbations are available during training, whereas for a new perturbation only its initial response is observed and future perturbed states must be forecast using time-matched controls as references (Figure 1). Unlike conventional perturbation prediction that estimates static endpoints or models future states under observed conditions, this setting requires generalizing dynamic response patterns across perturbations. The central challenge arises from the entanglement of two distinct but coupled factors: shared cellular progression that is repeatedly observed across perturbations, and perturbation-induced response dynamics that must be transferred to an unseen intervention.

Existing approaches address only parts of this challenge. Temporal single-cell models learn continuous cellular trajectories, but typically focus on reconstructing a single developmental process or observed cellular transition rather than transferring response dynamics across perturbations Tong

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

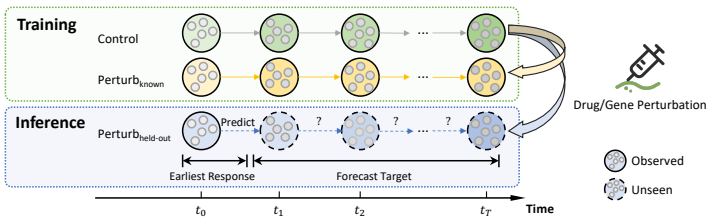


Figure 1: Task setting for forecasting held-out perturbation trajectories from early responses.

et al. (2020); Zhang et al. (2024a); Jiang & Wan (2024). In contrast, perturbation prediction methods learn perturbation-conditioned representations or control-to-perturbation transformations, but predominantly focus on static expression changes or endpoint distributions rather than forecasting future temporal responses Lotfollahi et al. (2021); Roohani et al. (2022); Cui et al. (2024). Recent generative approaches such as CellFlow and MMFM further incorporate temporal and perturbation information into continuous generative modeling Klein et al. (2025); Rohbeck et al. (2025). However, these methods are designed to generate responses for known perturbation contexts rather than infer the dynamics of a perturbation that is absent from training.

To address this challenge, we propose DeFo, which decomposes perturbed dynamics relative to control into shared base dynamics and a perturbation-specific response component. The former captures temporal progression observed in control states, whereas the latter models perturbation-induced deviations from the control trajectory.

To construct the response dynamics of a held-out perturbation, we further introduce Temporal Response Memory Attention (TRMA). TRMA represents known perturbations as a temporal memory containing their early response signatures and continuous response-velocity trajectories. Given the early observation of a new perturbation, TRMA retrieves and softly composes relevant response trajectories from known perturbations. The inferred response dynamics are then combined with the shared base dynamics to forecast future perturbed states, enabling generalization beyond perturbations observed during training.

Our main contributions are summarized as follows:

- We formulate a held-out temporal single-cell perturbation forecasting problem, where future trajectories of unseen perturbations are predicted from limited early-response observations, and introduce a control-referenced decomposition that separates shared cellular progression from transferable perturbation-induced dynamics.
- We propose Temporal Response Memory Attention (TRMA), a trajectory-valued memory retrieval mechanism that transfers time-varying response patterns across perturbations conditioned on their early effects.
- We evaluate DeFo on time-resolved chemical and genetic perturbation datasets, demonstrating competitive recovery of population-level expression and perturbation-specific responses.

2 Related Work

2.1 Temporal Single-Cell Dynamics Modeling

Temporal single-cell modeling aims to reconstruct and forecast cellular state transitions from sparsely sampled and unpaired observations. Early trajectory inference approaches recover cellular progression by estimating probabilistic transitions or population transport structures Schiebinger et al. (2019); Lange et al. (2022). Generative models such as PRESCIENT and scNODE further introduce continuous dynamical formulations by combining latent representations with energy-based or neural ODE models to predict unobserved cellular states Yeo et al. (2021); Zhang et al. (2024a). Recent approaches extend temporal forecasting using generative modeling, including diffusion-based and autoregressive frameworks for long-range prediction He et al. (2024); Luo et al. (2026); Su & Emad (2026).

Despite their ability to model cellular evolution, these methods primarily learn the temporal progression of an observed biological process. They do not explicitly consider how perturbation-induced dynamics can be transferred across interventions or inferred for perturbations whose future trajectories are unavailable.

2.2 Single-Cell Perturbation Response Prediction

Single-cell perturbation response prediction focuses on modeling how cellular states change under genetic or chemical interventions. Early representation-based methods, such as scGen and CPA, learn perturbation effects through latent transformations or condition representations Lotfollahi et al. (2019; 2021). SAMS-VAE further separates sample-specific basal variation from sparse additive perturbation effects in the latent space Bereket & Karaletsos (2023). More recent approaches, including GEARS, scGPT, STATE, PerturbNet, and ChemCPA, improve prediction across diverse perturbation contexts by leveraging relational structures, pretrained representations, or perturbation-aware latent spaces Roohani et al. (2022); Hetzel et al. (2022); Cui et al. (2024); Yu et al. (2025); Adduri et al. (2026).

Generative models further formulate perturbation prediction as distributional transformation between control and perturbed states. CellFlow and scDFM adopt flow-based formulations, while Departures and PerturbDiff model population transitions using neural Schrödinger bridges and diffusion processes, respectively Klein et al. (2025); Yu et al. (2026); Chi et al. (2026); Yuan et al. (2026). However, these approaches mainly focus on endpoint responses or static distributions and typically rely on explicit perturbation representations during inference, rather than forecasting future dynamics of perturbations absent from training.

2.3 Time-Resolved Perturbation Modeling

Recent studies have started to integrate perturbation modeling with temporal dynamics. MMFM learns shared conditional flow fields across multiple time points and experimental conditions, enabling inference of missing temporal-condition combinations Rohbeck et al. (2025). CellTempo incorporates perturbation information into autoregressive temporal forecasting, while CITE-VAE models latent causal programs underlying dynamic perturbation responses Luo et al. (2026); Jiang et al. (2026).

Although these approaches bridge perturbation prediction and temporal modeling, they address different generalization settings from ours. Existing methods generally assume access to the target perturbation condition, a learned perturbation representation, or an observed trajectory state when forecasting future responses. In contrast, DeFo considers held-out perturbation forecasting, where the target perturbation is absent during training and only its early response relative to control is available. By decomposing cellular evolution into shared temporal progression and transferable perturbation-induced dynamics, DeFo enables early-response-guided transfer of future trajectories from known perturbations.

3 Methodology

In this section, we present DeFo for held-out perturbation forecasting, where future trajectories of an unseen perturbation are inferred from its earliest observed response. The key idea is to introduce a control-referenced inductive bias: rather than learning a single perturbation-conditioned dynamical process, DeFo decomposes cellular evolution into shared temporal progression and transferable perturbation-induced response dynamics. The former is modeled by a base field learned from control trajectories, while the latter is inferred through Temporal Response Memory Attention (TRMA) from known perturbations. The overall framework is illustrated in Figure 2.

3.1 Preliminaries

Let $\mathcal{C}_{tr}$ denote the set of training perturbations and c^* a held-out perturbation. For each perturbation $c \in \mathcal{C}_{tr}$, observations are available at discrete time points $\{t_l\}_{l=0}^{L}$. Let $\mathbf{x}_{i,t_l}^{c}$ denote the state of cell i under perturbation c, and $\mathbf{x}_{i,t_l}^{ctrl,c}$ the matched control reference. Time-matched control pop-

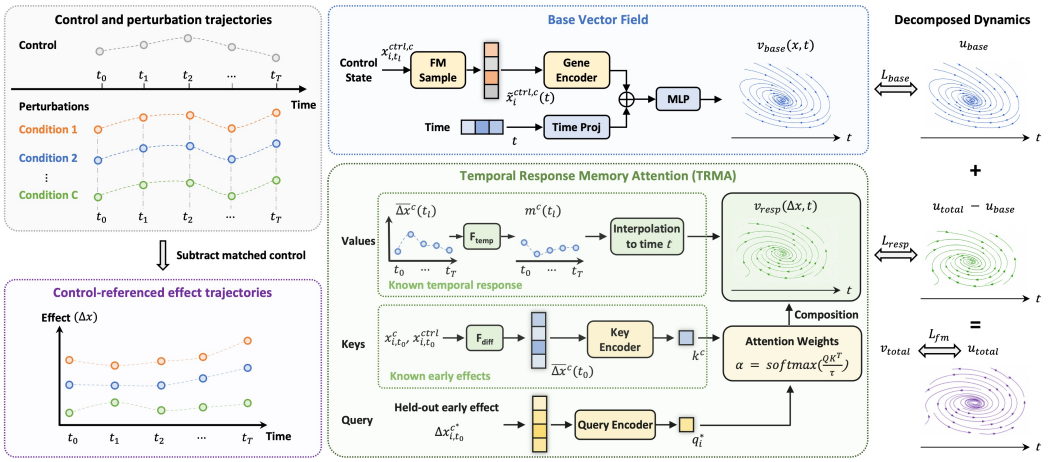


Figure 2: Overview of DeFo. DeFo models temporal single-cell perturbation responses using a decomposed vector field consisting of a shared base field and a perturbation-specific response component.

ulations are available at all sampled time points and serve as reference trajectories, whereas future populations of the held-out perturbation remain unobserved.

Because single-cell measurements are destructive and unpaired across time, we first construct pseudo-trajectories using optimal transport. The same alignment procedure is used to match perturbation and control populations at corresponding time points, producing temporally consistent control references.

We define the control-referenced perturbation effect as

$$\Delta \mathbf{x}^c_{i,t_\ell} = \mathbf{x}^c_{i,t_\ell} - \mathbf{x}^{\mathrm{ctrl},c}_{i,t_\ell}.$$

This formulation provides a reference coordinate system for characterizing perturbation-induced deviations.

3.2 Control-Referenced Decomposed Dynamics

To introduce a control-referenced decomposition of perturbation dynamics, DeFo models each perturbed trajectory relative to its matched control reference. Following the multi-marginal flow-matching formulation of Rohbeck et al. (2025), we model the aligned multi-time trajectories in continuous time, under which the state under perturbation c is decomposed as

$$\mathbf{x}^c_i(t) = \mathbf{x}^{\mathrm{ctrl},c}_i(t) + \Delta \mathbf{x}^c_i(t), \tag{1}$$

where $\mathbf{x}^{\mathrm{ctrl},c}_i(t)$ represents the control-associated temporal progression and $\Delta \mathbf{x}^c_i(t)$ denotes the perturbation-induced deviation from the control trajectory. Correspondingly, the perturbed velocity is modeled as the combination of a shared base field and a perturbation-specific response field:

$$\mathbf{v}^c_i(t) = \mathbf{v}_{\mathrm{base}}\left(\mathbf{x}^{\mathrm{ctrl},c}_i(t), t\right) + v^c_{\mathrm{resp}}\left(\Delta x^c_{i,t_0}, t\right). \tag{2}$$

The base field captures temporal dynamics associated with shared cellular progression from the MMFM-sampled control state $\widetilde{\mathbf{x}}^{\mathrm{ctrl},c}_i(t)$ and continuous time:

$$\mathbf{v}_{\mathrm{base}} = \mathrm{MLP}\left(\left[\mathrm{GeneEnc}\left(\widetilde{\mathbf{x}}^{\mathrm{ctrl},c}_i(t)\right); \mathrm{TimeProj}(t)\right]\right). \tag{3}$$

In contrast, the response field models the temporal evolution of perturbation-induced deviations relative to the control reference. This decomposition is used as a control-referenced inductive bias rather than an assumption of biological independence between the two components.

3.3 Temporal Response Memory Attention

Although the future response of a held-out perturbation is unavailable during inference, observed perturbations provide paired early effects and subsequent response dynamics. TRMA exploits this asymmetric information structure by constructing a temporal response memory from known perturbations and retrieving transferable response dynamics according to the early effect of the target perturbation.

For each known perturbation $c \in \mathcal{C}_{\mathrm{tr}}$, we summarize its control-referenced response trajectory using the population-level perturbation effect:

$$\overline{\Delta \mathbf{x}}^{c}(t_\ell) = \frac{1}{N_c} \sum_{i=1}^{N_c} \Delta \mathbf{x}^{c}_{i,t_\ell}, \tag{4}$$

where N_c denotes the number of aligned trajectories under perturbation c. The early response is used as a perturbation signature:

$$\mathbf{e}^{c} = \overline{\Delta \mathbf{x}}^{c}(t_0), \tag{5}$$

which characterizes how each perturbation initially deviates from its control reference.

To capture how the perturbation response evolves over time, we further construct a response-velocity memory. Specifically, the temporal response velocity is estimated from the mean effect trajectory:

$$\mathbf{m}^{c}(t_\ell) = \frac{\overline{\Delta \mathbf{x}}^{c}(t_{\ell+1}) - \overline{\Delta \mathbf{x}}^{c}(t_{\ell-1})}{t_{\ell+1} - t_{\ell-1}}, 1 \le \ell < L, \tag{6}$$

where forward and backward differences are used for boundary time points. Therefore, each observed perturbation is represented as a memory entry containing an early-response signature $\mathbf{e}^{c}$ and its associated temporal response velocity trajectory $\{\mathbf{m}^{c}(t_\ell)\}_{\ell=0}^{L}$.

For a held-out perturbation c^*, only its early response is available. Given the early effect of a target cell, $\Delta \mathbf{x}^{c^*}_{i,t_0}$, TRMA learns to retrieve relevant response memories rather than relying on a predefined perturbation identity. The query and memory keys are projected into a shared latent space and normalized:

$$\mathbf{q}^{*}_{i} = \mathrm{Norm}\left(\mathrm{QueryEnc}\left(\Delta \mathbf{x}^{c^*}_{i,t_0}\right)\right), \mathbf{k}^{c} = \mathrm{Norm}\left(\mathrm{KeyEnc}\left(\mathbf{e}^{c}\right)\right). \tag{7}$$

The relevance of each known perturbation is determined by the similarity between the target early response and the stored perturbation signatures:

$$\alpha_{i,c} = \frac{\exp\left((\mathbf{q}^{*}_{i})^{\top}\mathbf{k}^{c}/\tau\right)}{\sum_{c' \in \mathcal{C}_{\mathrm{tr}}} \exp\left((\mathbf{q}^{*}_{i})^{\top}\mathbf{k}^{c'}/\tau\right)}, \tag{8}$$

where τ is the temperature parameter.

Since response memories are observed only at discrete time points, each response-velocity trajectory is linearly interpolated to obtain a continuous function $\widetilde{\mathbf{m}}^{c}(t)$. The response-velocity component of the held-out perturbation is then constructed by composing the time-varying response trajectories:

$$\mathbf{v}^{c^*}_{\mathrm{resp}}\left(\Delta \mathbf{x}^{c^*}_{i,t_0}, t\right) = \sum_{c \in \mathcal{C}_{\mathrm{tr}}} \alpha_{i,c} \widetilde{\mathbf{m}}^{c}(t). \tag{9}$$

Thus, TRMA maps early perturbation effects to trajectory-valued response memories without requiring a predefined representation of the held-out perturbation.

3.4 Learning Objective and Forecasting

The decomposition introduced in DeFo provides an interpretable separation between shared cellular progression and perturbation-induced response dynamics. However, directly optimizing only the total perturbed velocity does not uniquely determine how the dynamics are allocated between the two components. Therefore, we introduce component-wise supervision to constrain the base and response fields toward their corresponding control-referenced dynamics.

The overall perturbed trajectory is first optimized using a flow-matching objective:

$$\mathcal{L}_{\mathrm{fm}} = \mathbb{E}\left[w(t)\left\|\mathbf{v}_i^c(t) - \mathbf{u}_i^c(t)\right\|_2^2\right], \tag{10}$$

where $\mathbf{v}_i^c(t)$ denotes the predicted perturbed velocity and $\mathbf{u}_i^c(t)$ denotes the conditional flow-matching target velocity constructed from the OT-aligned multi-time trajectory. The weighting function $w(t)$ increases the contribution of later stages as the normalized continuous time t increases.

To encourage the base field to capture perturbation-independent temporal progression, we align it with the velocity of the matched control trajectory:

$$\mathcal{L}_{\mathrm{base}} = \mathbb{E}\left[\left\|\mathbf{v}_{\mathrm{base}}\left(\widetilde{\mathbf{x}}_i^{\mathrm{ctrl},c}(t), t\right) - \mathbf{u}_i^{\mathrm{ctrl},c}(t)\right\|_2^2\right]. \tag{11}$$

The control target velocity $u_i^{\mathrm{ctrl},c}(t)$ is constructed analogously from the aligned control trajectory. Similarly, the response component is constrained by the temporal evolution of the control-referenced perturbation effect:

$$\mathcal{L}_{\mathrm{resp}} = \mathbb{E}\left[\left\|v_{\mathrm{resp}}^c\left(\Delta x_{i,t_0}^c, t\right) - \mathbf{u}_{\mathrm{resp},i}^c(t)\right\|_2^2\right]. \tag{12}$$

The final training objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{fm}} + \lambda_{\mathrm{base}}\mathcal{L}_{\mathrm{base}} + \lambda_{\mathrm{resp}}\mathcal{L}_{\mathrm{resp}}. \tag{13}$$

Held-out perturbation forecasting. At inference, DeFo forecasts a perturbation that is absent from training and has no observed future trajectory. The early control-referenced response of the target perturbation is first provided to TRMA, which constructs the corresponding response component by transferring temporal response dynamics from known perturbations. The future trajectory is then generated by combining the shared base field and the inferred response field:

$$\frac{\mathrm{d}\widehat{\mathbf{x}}_i^{c^*}(t)}{\mathrm{d}t} = \mathbf{v}_{\mathrm{base}}\left(\mathbf{x}_i^{\mathrm{ctrl},c^*}(t), t\right) + \widehat{\mathbf{v}}_{\mathrm{resp}}^{c^*}\left(\Delta\mathbf{x}_{i,t_0}^{c^*}, t\right), \tag{14}$$

where the trajectory is initialized from the observed early state, $\widehat{\mathbf{x}}_i^{c^*}(t_0) = \mathbf{x}_{i,t_0}^{c^*}$.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate DeFo on two time-resolved single-cell perturbation datasets, covering both chemical and genetic interventions. HSPC-Mk DeMeo et al. (2025) contains 21 chemical perturbations measured at five time points, whereas RENGE Ishikawa et al. (2023) contains 23 genetic perturbations measured at four time points. For both datasets, we retain 2,000 highly variable genes (HVGs), additionally retaining perturbation targets for genetic perturbations when absent from the HVG set.

Evaluation Protocol. We adopt condition-level holdout evaluation. HSPC-Mk and RENGE use five approximately 80/20 perturbation splits. In each setting, future populations of the held-out perturbations are excluded from training, while their earliest perturbed populations and time-matched control populations across the forecast horizon remain available.

Baselines. We compare DeFo with representative approaches from temporal single-cell dynamics modeling and perturbation response prediction, including scNODE Zhang et al. (2024a), scGPT Cui et al. (2024), CellFlow Klein et al. (2025), scDFM Yu et al. (2026), and Departures Chi et al. (2026). We additionally include two empirical baselines: Ctrl, which samples from the observed control population at each target time point, and Pert, which constructs a condition-balanced mixture from known perturbation populations. All methods are evaluated using identical data splits, feature spaces, and forecasting targets.

Evaluation Metrics. We report three complementary population-level metrics: mean squared error (MSE) between predicted and observed expression profiles, Pearson correlation of control-referenced perturbation effects (Pearson-Δ), and differential-expression direction match (DE Dir.) for gene-level response consistency.

Additional details on dataset statistics, preprocessing, baseline adaptation, and implementation settings are provided in Appendix A.

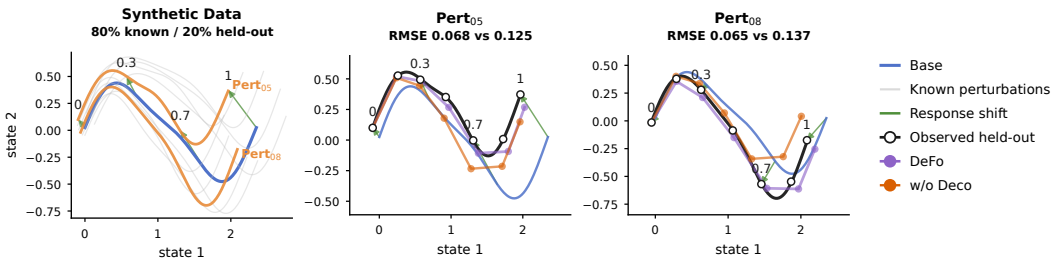


Figure 3: Synthetic validation of decomposed dynamics.

Table 1: Main results on held-out temporal perturbation forecasting. Results are reported as mean ± standard deviation across condition-level holdout splits. Best and second-best results are shown in bold and underlined, respectively.

Method	HSPC-Mk			RENGE		
	MSE ↓	Pearson Δ ↑	DE Dir. ↑	MSE ↓	Pearson Δ ↑	DE Dir. ↑
Ctrl	0.0033 ± 0.0012	0.0259 ± 0.0531	0.4952 ± 0.0174	0.0015 ± 0.0012	0.0137 ± 0.0090	0.4379 ± 0.0985
Pert	0.0021 ± 0.0008	**0.5890 ± 0.0907**	0.8008 ± 0.0443	0.0015 ± 0.0011	0.2966 ± 0.0487	0.7216 ± 0.1545
scNODE	0.0111 ± 0.0009	0.0557 ± 0.0774	0.5406 ± 0.0493	0.0026 ± 0.0012	0.2268 ± 0.0564	0.6607 ± 0.1585
scGPT	0.0035 ± 0.0010	0.4685 ± 0.0798	0.7360 ± 0.0457	0.0030 ± 0.0013	0.2221 ± 0.0291	0.7481 ± 0.0332
CellFlow	0.0019 ± 0.0007	0.5213 ± 0.1529	0.7707 ± 0.0614	0.0012 ± 0.0010	0.2655 ± 0.0938	0.6857 ± 0.1339
scDFM	0.0858 ± 0.0652	0.1894 ± 0.0629	0.5409 ± 0.0332	0.0185 ± 0.0127	0.1515 ± 0.0676	0.6612 ± 0.0994
Departures	0.0148 ± 0.0054	0.1924 ± 0.0713	0.6236 ± 0.0403	0.0036 ± 0.0008	0.0553 ± 0.0872	0.4683 ± 0.0623
DeFo	**0.0017 ± 0.0009**	0.5627 ± 0.0800	**0.8021 ± 0.0290**	**0.0010 ± 0.0007**	**0.4667 ± 0.0574**	**0.9465 ± 0.0110**

4.2 Controlled Validation of Decomposed Dynamics on Synthetic Data

Before evaluating real-world forecasting performance, we first examine the behavior of the proposed decomposition in a controlled setting with known ground-truth dynamics. We construct a two-dimensional synthetic held-out perturbation task in which each cell-level observation follows

$$x_i^c(t) = x_{\text{base}}(t) + x_{\text{resp}}^c(t) + \epsilon_{i,t}^c, \tag{15}$$

where the base follows shared nonlinear dynamics, the response varies across time and perturbations, and $\epsilon_{i,t}^c$ denotes independently sampled observation noise. The corresponding noise-free mean trajectory has velocity

$$u^c(t) = u_{\text{base}}(t) + u_{\text{resp}}^c(t). \tag{16}$$

To examine the effect of decomposition under this controlled setting, we compare DeFo with a non-decomposed variant (w/o Deco) under the same held-out forecasting protocol. As shown in Figure 3, the shared base trajectory defines the reference temporal progression, while the response component provides a correction that moves the trajectory toward the true perturbation-specific evolution. DeFo achieves lower trajectory RMSE on both held-out perturbations, with reductions of 45.6% and 52.6% relative to w/o Deco. These results support the utility of explicit decomposition when the underlying dynamics admit a shared-plus-response structure. Detailed generation settings and additional analyses are provided in Appendix B.1.

4.3 Held-Out Perturbation Forecasting on Real Single-Cell Datasets

We next evaluate DeFo under the held-out perturbation setting on HSPC-Mk and RENGE. As shown in Table 1, DeFo achieves the best mean performance in five of the six dataset–metric combinations and ranks second in the remaining one. In particular, DeFo obtains the lowest MSE and highest DE Direction on both datasets, indicating more accurate reconstruction of future population states and perturbation-specific response directions.

The results suggest that explicitly modeling transferable response dynamics is beneficial for held-out perturbation forecasting. By decomposing dynamics into a shared base progression and a trans-

Table 2: Ablation study. Results are reported as mean ± standard deviation.

Variant	HSPC-Mk			RENGE		
	MSE ↓	Pearson Δ ↑	DE Dir. ↑	MSE ↓	Pearson Δ ↑	DE Dir. ↑
w/o Deco	0.0053 ± 0.0015	0.2939 ± 0.1276	0.6540 ± 0.0534	**0.0010 ± 0.0006**	0.4552 ± 0.0795	0.9385 ± 0.0343
w/o Temp	0.0020 ± 0.0009	0.5291 ± 0.1007	0.7865 ± 0.0359	0.0011 ± 0.0007	0.4367 ± 0.0645	0.9424 ± 0.0249
w/o Comp	0.0039 ± 0.0008	0.3529 ± 0.1230	0.7101 ± 0.0360	0.0011 ± 0.0007	0.4472 ± 0.0585	0.9358 ± 0.0320
DeFo	**0.0017 ± 0.0009**	**0.5627 ± 0.0800**	**0.8021 ± 0.0290**	**0.0010 ± 0.0007**	**0.4667 ± 0.0574**	**0.9465 ± 0.0110**

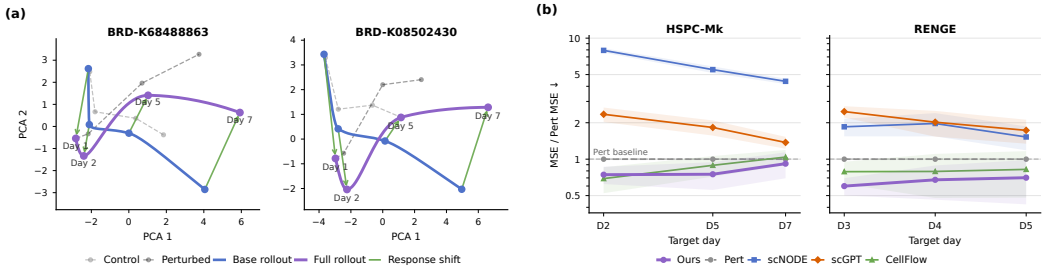


Figure 4: Analysis of learned dynamics. (a) Decomposition geometry in PCA space for representative perturbations. (b) Time-resolved forecasting performance measured by normalized MSE across target time points.

ferable response component, DeFo can infer future perturbation trajectories from limited early-response evidence.

4.4 Ablation Study

To investigate the contribution of individual design components, we perform ablation studies on three key modules of DeFo: w/o Deco removes the explicit base–response decomposition, w/o Temp replaces time-varying response dynamics with a static representation, and w/o Comp replaces soft response composition with the single most similar known perturbation.

As shown in Table 2, removing the decomposition substantially degrades all three metrics on HSPC-Mk. On RENGE, DeFo attains the highest mean Pearson-Δ and DE Direction while tying with w/o Deco for the lowest MSE, suggesting that the main benefit of decomposition in this setting is at the response level. Removing temporal response modeling or soft composition also lowers the mean response-level performance on both datasets. Overall, these trends are consistent with complementary roles of decomposition, temporal response modeling, and soft composition in held-out perturbation forecasting.

4.5 Analysis of Learned Dynamics

Beyond aggregate forecasting performance, we further examine the learned decomposition and its behavior across prediction horizons.

Decomposition geometry. We first visualize representative HSPC-Mk perturbations by jointly projecting observed control and perturbed trajectories, the base rollout, and the final DeFo predictions into a two-dimensional PCA space (Figure 4a). In these examples, the base rollout follows the major temporal progression of the matched control trajectory, while the response component provides time-dependent corrections toward the perturbed trajectory. This qualitative analysis illustrates how DeFo represents perturbation-specific responses relative to a shared temporal backbone.

Time-resolved forecasting performance. We next examine how forecasting error evolves over increasing prediction horizons (Figure 4b). MSE is normalized by the empirical Pert baseline, such that values below 1 indicate lower error than Pert. DeFo maintains favorable relative performance

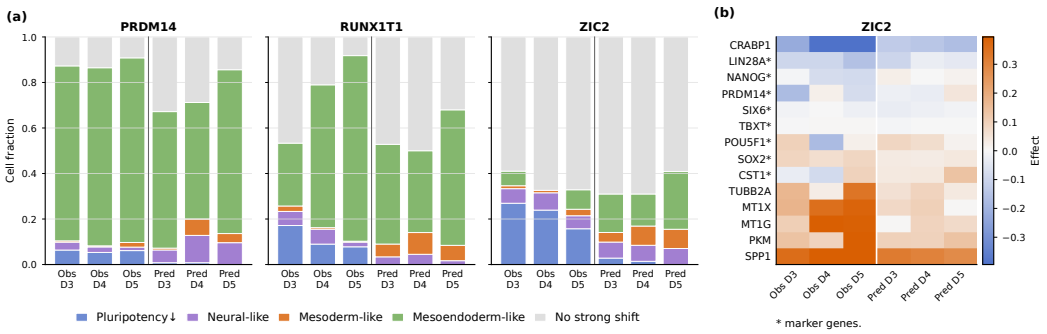


Figure 5: Biological case study of forecasted responses. (a) Observed and predicted temporal response programs for PRDM14, RUNX1T1, and ZIC2. (b) Observed and predicted gene-level perturbation effects for ZIC2 across time.

across the evaluated horizons on both HSPC-Mk and RENGE. Complete results are provided in Appendix B.2.

4.6 Biological Plausibility of Forecasted Dynamics

To assess whether DeFo preserves interpretable biological response patterns beyond aggregate expression agreement, we examine representative perturbations from RENGE at the population-composition and gene-expression levels.

We first evaluate perturbation-induced changes in cell-state composition. Following previously characterized response programs, cells are assigned to five control-referenced state-shift categories: pluripotency loss, neural-like, mesoderm-like, mesoendoderm-like, and no strong shift. We compare the predicted and observed population compositions over time for representative perturbations, including PRDM14, RUNX1T1, and ZIC2 (Figure 5a). The predicted trajectories capture the major temporal trends of these state-shift patterns, including the emergence of mesoendoderm-like responses and changes in pluripotency-associated marker programs Ishikawa et al. (2023).

We further examine ZIC2 as a representative held-out perturbation. DeFo predicts an increasing mesoendoderm-like response over time, which is biologically plausible given previous evidence linking ZIC2 to pluripotency-state regulation and early lineage specification Hossain et al. (2024).

At the gene level, we analyze control-referenced expression changes for representative marker genes (Figure 5b). For ZIC2, DeFo recovers the direction of several lineage- and pluripotency-associated marker changes, including the day-5 increase of CST1. Together, these examples suggest that the forecasts retain biologically interpretable temporal response patterns beyond aggregate expression recovery.

5 Conclusion

We presented DeFo, a control-referenced framework for forecasting temporal responses to held-out perturbations from limited early observations. DeFo decomposes perturbed dynamics into shared control-associated progression and perturbation-specific response dynamics, and introduces Temporal Response Memory Attention to retrieve and compose trajectory-valued response patterns from known perturbations.

Across chemical and genetic perturbation datasets, DeFo achieved competitive recovery of population-level expression, control-referenced perturbation effects, and differential-expression directions. Further analyses demonstrated that the learned decomposition captures interpretable response corrections and preserves biologically meaningful temporal programs. Overall, DeFo provides a principled framework for transferring temporal perturbation dynamics beyond observed interventions, with its generalization capability benefiting from diverse and informative response trajectories.

AI USE STATEMENT

In this work, generative AI tools were used to assist with manuscript writing and revision, including language polishing, sentence rephrasing, literature retrieval and summarization, and limited discussion of research methodology and result interpretation. Generative AI tools did not independently conduct experiments or generate experimental data or results. All AI-assisted content was reviewed, verified where appropriate, and finalized by the authors.

ETHICS STATEMENT

This work follows the ICLR Code of Ethics. We use publicly available single-cell transcriptomic datasets and do not conduct any new experiments involving human participants or animals. All datasets are used in accordance with their original release conditions and associated usage policies. Our work is intended for methodological research in computational biology and does not provide clinical or diagnostic recommendations.

REPRODUCIBILITY STATEMENT

We provide detailed descriptions of the datasets, preprocessing procedures, condition-level data splits, evaluation metrics, baseline adaptations, and implementation settings in Appendix A. Model hyperparameters, optimization settings, and computational resources are also reported to facilitate reproduction of the experiments. The datasets used in this work are publicly available from their original sources. Code and files required to reproduce the reported experiments are included in the supplementary material.

REFERENCES

Abhinav K Adduri, Dhruv Gautam, Beatrice Bevilacqua, Mohsen Naghipourfar, Alishba Imran, Rohan Shah, Noam Teyssier, Rishi Verma, Christopher Carpenter, Basak Eraslan, et al. Predicting cellular responses to perturbation across diverse contexts with state. *Cell*, 2026.

Michael Bereket and Theofanis Karaletsos. Modelling cellular perturbations with the sparse additive mechanism shift variational autoencoder. *Advances in Neural Information Processing Systems*, 36:1–12, 2023.

Changxi Chi, Yufei Huang, Jun Xia, Jiangbin Zheng, Yunfan Liu, Zelin Zang, and Stan Z Li. Departures: Distributional transport for single-cell perturbation prediction with neural schrödinger bridges. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 20544–20552, 2026.

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.

Benjamin DeMeo, Charlotte Nesbitt, Samuel A Miller, Daniel B Burkhardt, Inna Lipchina, Doris Fu, Peter Holderrieth, David Kim, Sergey Kolchenko, Artur Szalata, et al. Active learning framework leveraging transcriptomics identifies modulators of disease phenotypes. *Science*, 390(6776): eadi8577, 2025.

Wenbo Guo, Zeyu Chen, and Jin Gu. Computational modeling of single-cell dynamics data. *Briefings in Bioinformatics*, 26(3):bbaf305, 2025.

Siyu He, Yuefei Zhu, Daniel Naveed Tavakol, Haotian Ye, Yeh-Hsing Lao, Zixian Zhu, Cong Xu, Sharadha Chauhan, Guy Garty, Raju Tomer, et al. Squidiff: predicting cellular development and responses to perturbations using a diffusion model. *bioRxiv*, pp. 2024–11, 2024.

Leon Hetzel, Simon Boehm, Niki Kilbertus, Stephan Günnemann, Fabian Theis, et al. Predicting cellular responses to novel drug perturbations at a single-cell resolution. *Advances in Neural Information Processing Systems*, 35:26711–26722, 2022.

Ishtiaque Hossain, Pierre Priam, Sofia C Reynoso, Sahil Sahni, Xiao X Zhang, Laurence Côté, Joelle Doumat, Candus Chik, Tianxin Fu, Julie A Lessard, et al. Zic2 and zic3 promote swi/snf recruitment to safeguard progression towards human primed pluripotency. *Nature Communications*, 15 (1):8539, 2024.

Masato Ishikawa, Seiichi Sugino, Yoshie Masuda, Yusuke Tarumoto, Yusuke Seto, Nobuko Taniyama, Fumi Wagai, Yuhei Yamauchi, Yasuhiro Kojima, Hisanori Kiryu, et al. Renge infers gene regulatory networks using time-series single-cell rna-seq data with crispr perturbations. *Communications Biology*, 6(1):1290, 2023.

Qi Jiang and Lin Wan. A physics-informed neural sde network for learning cellular dynamics from time-series scrna-seq data. *Bioinformatics*, 40(Supplement_2):ii120–ii127, 2024.

Wenkang Jiang, Yuhang Liu, Erdun Gao, Ehsan Abbasnejad, Lina Yao, and Javen Qinfeng Shi. Learning latent dynamical causal processes for single-cell perturbation prediction. *arXiv preprint arXiv:2605.25581*, 2026.

Dominik Klein, Jonas Simon Fleck, Daniil Bobrovskiy, Lea Zimmermann, Sören Becker, Alessandro Palma, Leander Dony, Alejandro Tejada-Lapuerta, Guillaume Huguet, Hsiu-Chuan Lin, et al. Cellflow enables generative single-cell phenotype modeling with flow matching. *bioRxiv*, pp. 2025–04, 2025.

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

Mohammad Lotfollahi, F Alexander Wolf, and Fabian J Theis. scgen predicts single-cell perturbation responses. *Nature methods*, 16(8):715–721, 2019.

Mohammad Lotfollahi, Anna Klimovskaia Susmelj, Carlo De Donno, Yuge Ji, Ignacio L Ibarra, F Alexander Wolf, Nafissa Yakubova, Fabian J Theis, and David Lopez-Paz. Compositional perturbation autoencoder for single-cell response modeling. *BioRxiv*, 2021.

Erpai Luo, Haoxiang Gao, Haiyang Bian, Yaru Li, Chen Li, Minsheng Hao, Mo Chen, Yuli She, Lei Wei, Kai Liu, et al. Autoregressive forecasting of future single-cell state transitions. *bioRxiv*, pp. 2026–02, 2026.

Martin Rohbeck, Edward De Brouwer, Charlotte Bunne, Jan-Christian Huetter, Anne Biton, Kelvin Chen, Aviv Regev, and Romain Lopez. Modeling complex system dynamics with flow matching across time and conditions. In *International Conference on Learning Representations*, volume 2025, pp. 12164–12188, 2025.

Yusuf Roohani, Kexin Huang, and Jure Leskovec. Gears: Predicting transcriptional outcomes of novel multi-gene perturbations. *BioRxiv*, pp. 2022–07, 2022.

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

Chen Su and Amin Emad. Cellpace: A temporal diffusion-forcing framework for simulation, interpolation and forecasting of single-cell dynamics. *bioRxiv*, pp. 2026–02, 2026.

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

Zhiting Wei, Yiheng Wang, Yicheng Gao, Shuguang Wang, Ping Li, Duanmiao Si, Yuli Gao, Siqi Wu, Danlu Li, Kejing Dong, et al. Benchmarking algorithms for generalizable single-cell perturbation response prediction. *Nature Methods*, 23(2):451–464, 2026.

Xing Xu, Qianxi Wen, Tianchen Lan, Liuqing Zeng, Yonghao Zeng, Shiyan Lin, Minghao Qiu, Xing Na, and Chaoyong Yang. Time-resolved single-cell transcriptomic sequencing. *Chemical Science*, 15(46):19225–19246, 2024.

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

Chenglei Yu, Chuanrui Wang, Bangyan Liao, and Tailin Wu. scdfm: Distributional flow matching model for robust single-cell perturbation prediction. *arXiv preprint arXiv:2602.07103*, 2026.

Hengshi Yu, Weizhou Qian, Yuxuan Song, and Joshua D Welch. Perturbnet predicts single-cell responses to unseen chemical and genetic perturbations. *Molecular Systems Biology*, 21(8):960, 2025.

Xinyu Yuan, Xixian Liu, Ya Shi Zhang, Zuobai Zhang, Hongyu Guo, and Jian Tang. Perturbdiff: Functional diffusion for single-cell perturbation modeling. In *The 2026 Workshop on Generative and Agentic AI for Biology*, 2026.

Jiaqi Zhang, Erica Larschan, Jeremy Bigness, and Ritambhara Singh. scnode: generative model for temporal single cell transcriptomic data prediction. *Bioinformatics*, 40(Supplement_2):ii146–ii154, 2024a.

Ziqi Zhang, Xinye Zhao, Mehak Bindra, Peng Qiu, and Xiuwei Zhang. scdisinfact: disentangled learning for integration and prediction of multi-batch multi-condition single-cell rna-sequencing data. *Nature Communications*, 15(1):912, 2024b.

APPENDICES

A EXPERIMENTAL DETAILS

A.1 DATASETS AND PREPROCESSING

HSPC-Mk. HSPC-Mk contains 21 chemical perturbations and a control condition measured at Days 0, 1, 2, 5, and 7. We preprocess the expression matrix and select 2,000 highly variable genes (HVGs) for downstream analysis. Day 0 contains only control cells and is excluded from trajectory training and evaluation; Day 1 is therefore used as the anchor time, with Days 2, 5, and 7 as forecasting targets.

RENGE. RENGE contains 23 genetic perturbations measured at Days 2, 3, 4, and 5. We concatenate the 10x Genomics expression matrices across time using their intersecting gene set. For each cell, we assign the guide with the largest guide-capture UMI count and exclude cells with no detected guide. We remove cells expressing fewer than 200 genes and genes detected in fewer than 10 cells, normalize library size to 10^4 counts per cell, and apply $\log(1+x)$. We select 2,000 HVGs using the Seurat procedure and retain perturbation target genes when absent from the selected HVGs.

A.2 DATA SPLITS AND EVALUATION PROTOCOL

HSPC-Mk and RENGE use five condition-level splits generated with seeds $0, \ldots, 4$, holding out four of 21 and five of 23 perturbations, respectively. The same predefined splits are used for all methods.

For each held-out perturbation, only its earliest observed population is available at inference, and all later perturbed populations are used exclusively as forecasting targets. An optimal transport coupling between the held-out and control populations at the anchor time is propagated along the observed control trajectory to construct temporally consistent control references.

A.3 BASELINE ADAPTATION

For each held-out perturbation c, let N_c denote the number of cells observed at the anchor time. Each generative method produces N_c cells at every future evaluation time.

Ctrl samples N_c cells from the observed control population at the target time. *Pert* forms a condition-balanced empirical mixture from known perturbations at the same target time, allocating samples as evenly as possible across conditions and sampling without replacement before reuse. Neither baseline accesses future cells from the held-out perturbation.

For condition-aware baselines, we derive an expression-based condition representation from the observed anchor population:

$$\mathbf{e}_c = \overline{\mathbf{z}}_{c,t_0} - \overline{\mathbf{z}}_{\mathrm{ctrl},t_0}, \qquad \mathbf{z} = \mathrm{PCA}_{50}(\mathbf{x}).$$

scGPT, CellFlow, scDFM, and Departures use this representation in place of, or in addition to, their original condition input while otherwise retaining their released model backbones and generation procedures. scNODE is trained without held-out perturbations using its native temporal forecasting formulation and initialized from the observed anchor population. Held-out future populations are excluded from all training and adaptation procedures.

MMFM is not included as a separate baseline because our w/o Deco variant provides the closest controlled MMFM-style counterpart under the same held-out protocol: it retains the multi-time flow-matching backbone while replacing the proposed decomposed dynamics with a single condition-aware vector field.

A.4 EVALUATION METRICS

We evaluate population-level expression recovery, perturbation-effect recovery, and differential-expression direction recovery. Metric implementations follow the `cell-eval` library.[1] For each

[1] `https://github.com/ArcInstitute/cell-eval`

split, metrics are first averaged across held-out perturbations within each forecast target time point, and then averaged across target time points. The main results table reports the mean ± standard deviation of these split-level scores.

MSE. For perturbation c at target time t, let $\bar{\mathbf{x}}_{c,t}^{\mathrm{pred}}$ and $\bar{\mathbf{x}}_{c,t}^{\mathrm{obs}}$ denote the predicted and observed pseudobulk expression profiles. MSE is defined as

$$\mathrm{MSE}_{c,t} = \frac{1}{G} \left\| \bar{\mathbf{x}}_{c,t}^{\mathrm{pred}} - \bar{\mathbf{x}}_{c,t}^{\mathrm{obs}} \right\|_2^2,$$

where G is the number of evaluated genes. Lower values indicate better global expression recovery.

Pearson-Δ. Using the observed control pseudobulk profile $\bar{\mathbf{x}}_{\mathrm{ctrl},t}^{\mathrm{obs}}$ at the same time point, we define

$$\widehat{\boldsymbol{\Delta}}_{c,t} = \bar{\mathbf{x}}_{c,t}^{\mathrm{pred}} - \bar{\mathbf{x}}_{\mathrm{ctrl},t}^{\mathrm{obs}}, \boldsymbol{\Delta}_{c,t}^{\mathrm{obs}} = \bar{\mathbf{x}}_{c,t}^{\mathrm{obs}} - \bar{\mathbf{x}}_{\mathrm{ctrl},t}^{\mathrm{obs}}.$$

Pearson-Δ is

$$\text{Pearson-}\Delta_{c,t} = \mathrm{corr}\left(\widehat{\boldsymbol{\Delta}}_{c,t}, \boldsymbol{\Delta}_{c,t}^{\mathrm{obs}} \right),$$

with higher values indicating better perturbation-effect recovery.

DE Direction Match. At each target time, differential expression is computed separately for the observed and predicted perturbation populations against the same observed control population using the Wilcoxon rank-sum test. The evaluation set is defined from the observed data as

$$\mathcal{G}_{c,t}^{\mathrm{obs}} = \left\{ g : \mathrm{FDR}_{g,c,t}^{\mathrm{obs}} < 0.05 \right\}.$$

DE Direction Match is

$$\mathrm{DDM}_{c,t} = \frac{1}{|\mathcal{G}_{c,t}^{\mathrm{obs}}|} \sum_{g \in \mathcal{G}_{c,t}^{\mathrm{obs}}} \mathbb{I}\Big[\mathrm{sign}(\widehat{\mathrm{LFC}}_{g,c,t}) = \mathrm{sign}(\mathrm{LFC}_{g,c,t}^{\mathrm{obs}}) \Big].$$

Higher values indicate better agreement in differential-expression direction.

A.5 Implementation Details

For each condition-holdout split, a 50-dimensional PCA transform is fitted using only control and known-perturbation cells and then fixed for training, validation, and inference. We sample 300 cells from each known condition–time block, using replacement when necessary, and split them 80/20 into training and validation sets.

DeFo uses 64-dimensional state and time embeddings, SELU activations, Xavier-uniform initialization, and a response-attention temperature of 0.1. We optimize using Adam with initial learning rate 10^{-3} and cosine decay to 10^{-5}. Training runs for at most 1,000 epochs with batch size 360, gradient clipping at norm 1, and early stopping evaluated every five epochs with patience 40 and a minimum improvement threshold of 10^{-2}.

For the multi-time flow-matching backbone, we use linear conditional paths with variance 0.1 and set $\lambda_{\mathrm{base}} = 1.0$ and $\lambda_{\mathrm{resp}} = 0.5$. The full velocity loss is weighted by $1 + 1.5t^2$ for normalized $t \in [0, 1]$. Forecast trajectories are integrated using fourth-order Runge–Kutta. Model training uses random seed 42 within each predefined split.

CellFlow is run in its official JAX environment; DeFo and the remaining baselines are implemented in PyTorch. Experiments are run on a single NVIDIA RTX 3090 GPU.

B Additional Results

B.1 Synthetic Data Generation

We constructed a controlled two-dimensional perturbation system to assess whether an explicit base–response decomposition improves held-out trajectory recovery when the data-generating process has the same additive structure. For each condition c, individual synthetic cells were sampled as

$$\mathbf{x}_{c,i}(t) = \mathbf{x}_{\mathrm{base}}(t) + \mathbf{x}_{\mathrm{resp}}^{c}(t) + \boldsymbol{\epsilon}_{i,t}^{c}, \tag{17}$$

where $x_{\text{base}}(t)$ denotes the shared temporal dynamics, $x^c_{\text{resp}}(t)$ denotes the condition-specific response, and the observation noise is independently sampled for each condition–time–cell observation:

$$\boldsymbol{\epsilon}^c_{i,t} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \boldsymbol{\Sigma} = \begin{bmatrix} 0.035^2 & 1.5 \times 10^{-4} \\ 1.5 \times 10^{-4} & (1.15 \times 0.035)^2 \end{bmatrix}. \tag{18}$$

Equivalently, the condition-specific mean trajectory is

$$\boldsymbol{\mu}_c(t) = \mathbf{x}_{\text{base}}(t) + \mathbf{x}^c_{\text{resp}}(t). \tag{19}$$

The corresponding analytic velocity satisfies

$$\mathbf{u}_c(t) = \frac{\mathrm{d}\boldsymbol{\mu}_c(t)}{\mathrm{d}t} = \mathbf{u}_{\text{base}}(t) + \mathbf{u}^c_{\text{resp}}(t). \tag{20}$$

These analytic derivatives define the controlled ground-truth dynamics, but are not provided to the models during training.

The shared base trajectory was defined as

$$\mathbf{x}_{\text{base}}(t) = \begin{bmatrix} 2.35t \\ 0.42\sin(2\pi t) + 0.10\sin(4\pi t + 0.25) \end{bmatrix}. \tag{21}$$

For the control condition, $\mathbf{x}^{\text{ctrl}}_{\text{resp}}(t) = \mathbf{0}$. For perturbation c, the condition-specific response was

$$\mathbf{x}^c_{\text{resp}}(t) = A_c g_c(t)\mathbf{d}_c + 0.12A_c \sin(2\pi t + \phi_c)(0.35 + 0.65t)\mathbf{o}_c. \tag{22}$$

where

$$g_c(t) = 0.22 + 0.78\,\text{sigmoid}\left(\frac{t - \tau_c}{0.095}\right). \tag{23}$$

Here, A_c is the response amplitude, $\mathbf{d}_c$ is the primary response direction, $\mathbf{o}_c$ is its orthogonal direction, τ_c controls the response onset, and ϕ_c controls the transient phase. The nonzero offset of $g_c(t)$ produces a weak response at $t = 0$, so that each held-out perturbation has an informative initial state for inference.

We considered ten perturbations and one control condition. The perturbations Pert_{05} and Pert_{08} were held out, whereas the remaining eight perturbations were used as known conditions. Thus, the synthetic task approximately matches the condition-level 80/20 holdout setting used for HSPC-Mk and RENGE. States were sampled at $t \in \{0, 0.15, 0.30, 0.50, 0.70, 0.90, 1.00\}$, with 72 cells independently sampled for each condition–time pair.

For held-out perturbations, only the initial-time observations were used to construct the early-response query; observations at later time points were excluded from training and reserved for evaluation.

Because the state space is two-dimensional, no PCA projection was applied. DeFo and w/o Deco were trained with the same data split, initialization seed, training budget, and inference protocol. The latter removes the base–response decomposition and models the dynamics with a single condition-aware vector field.

For each held-out perturbation, trajectory error was evaluated using the mean predicted and observed states at the six future time points $\mathcal{T}_{\text{fut}} = \{0.15, 0.30, 0.50, 0.70, 0.90, 1.00\}$:

$$\text{RMSE}_c = \sqrt{\frac{1}{2|\mathcal{T}_{\text{fut}}|} \sum_{t \in \mathcal{T}_{\text{fut}}} \|\widehat{\boldsymbol{\mu}}_c(t) - \overline{\mathbf{x}}_c(t)\|_2^2}, \tag{24}$$

where $\widehat{\boldsymbol{\mu}}_c(t)$ is the mean generated state and $\overline{\mathbf{x}}_c(t)$ is the empirical mean of the observed synthetic cells. This analysis tests a controlled inductive-bias claim. It does not establish that temporal and perturbation effects are uniquely identifiable in real biological data.

B.2 Additional Analysis of Learned Dynamics

Effect-space validation. To complement the PCA visualization in Figure 4a, we further evaluate the learned decomposition directly in the control-referenced effect space. At each time point, observed

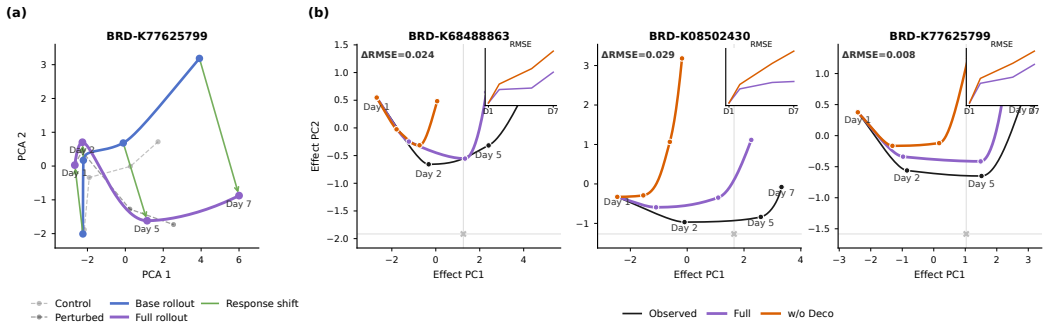


Figure 6: Additional analysis of decomposed dynamics. (a) Decomposition geometry for a representative perturbation. (b) Control-referenced effect-space trajectories of DeFo and w/o decomposition for three representative perturbations, with high-dimensional effect RMSE shown in the insets.

Table 3: Unnormalized time-resolved forecasting performance measured by MSE.

Method	HSPC-Mk			RENGE		
	Day 2	Day 5	Day 7	Day 3	Day 4	Day 5
Ctrl	0.0025 ± 0.0010	0.0023 ± 0.0010	0.0050 ± 0.0016	0.0012 ± 0.0006	0.0014 ± 0.0012	0.0019 ± 0.0018
Pert	0.0011 ± 0.0005	0.0019 ± 0.0009	0.0033 ± 0.0012	0.0011 ± 0.0006	0.0014 ± 0.0011	0.0020 ± 0.0016
scNODE	0.0086 ± 0.0008	0.0102 ± 0.0014	0.0144 ± 0.0017	0.0020 ± 0.0007	0.0028 ± 0.0014	0.0030 ± 0.0015
scGPT	0.0025 ± 0.0008	0.0034 ± 0.0011	0.0045 ± 0.0011	0.0027 ± 0.0006	0.0029 ± 0.0016	0.0034 ± 0.0017
CellFlow	**0.0007 ± 0.0004**	0.0017 ± 0.0007	0.0034 ± 0.0011	0.0009 ± 0.0004	0.0011 ± 0.0010	0.0016 ± 0.0015
scDFM	0.0947 ± 0.0649	0.0813 ± 0.0650	0.0813 ± 0.0659	0.0182 ± 0.0117	0.0188 ± 0.0133	0.0187 ± 0.0130
Departures	0.0171 ± 0.0064	0.0140 ± 0.0052	0.0131 ± 0.0046	0.0035 ± 0.0011	0.0033 ± 0.0007	0.0040 ± 0.0007
DeFo	0.0008 ± 0.0003	**0.0014 ± 0.0008**	**0.0030 ± 0.0016**	**0.0006 ± 0.0002**	**0.0010 ± 0.0007**	**0.0014 ± 0.0012**

and predicted perturbation effects are computed relative to the matched control reference and jointly projected into a shared PCA space.

As shown in Figure 6b, DeFo more closely follows the observed effect trajectories than the non-decomposed variant for the three representative perturbations. The corresponding high-dimensional effect RMSE is also lower for DeFo, with reductions of 0.024, 0.029, and 0.008, respectively. These examples provide additional real-data evidence that the decomposed formulation improves recovery of perturbation-specific effects.

Time-resolved forecasting performance. Table 3 reports the unnormalized MSE values corresponding to the normalized comparison in Figure 4b. DeFo achieves the lowest mean MSE at five of the six target time points and the second-lowest MSE at HSPC-Mk Day 2. Although absolute prediction error increases at later HSPC-Mk horizons, DeFo remains among the strongest-performing methods across the evaluated target times. These results complement the normalized analysis by showing the underlying absolute forecasting errors.

B.3 Additional Biological Response-Program Analysis

We further evaluate four predefined biological response programs across all held-out RENGE perturbations included in the evaluation splits: pluripotency loss, neural-like, mesoderm-like, and mesoendoderm-like responses. For each perturbation and time point, cells are scored using the corresponding program signatures, and the resulting scores are aggregated at the condition–time level.

For each perturbation–program pair, we compare the predicted and observed response trajectories across the evaluated time points. The validation score summarizes temporal agreement, Day 5 response-direction consistency, relative response amplitude, and Day 5 response magnitude. Fig-

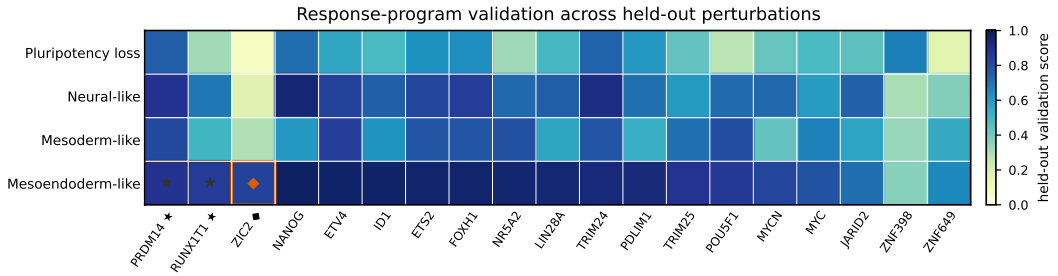


Figure 7: Response-program validation across held-out RENGE perturbations. The heatmap summarizes agreement between predicted and observed temporal responses for four predefined biological programs.

ure 7 reports these scores across all held-out perturbations, providing a systematic view of response-program recovery beyond the representative examples shown in Figure 5.

PRDM14 and RUNX1T1 show strong agreement in literature-supported response programs and are therefore included as representative validation examples in the main text. ZIC2 exhibits a comparatively strong and temporally consistent mesoendoderm-like response, with its agreement concentrated in this program relative to the other response programs. We therefore include ZIC2 as an additional representative case for examining the corresponding temporal and gene-level response patterns.