HYPOTHESIS-EVIDENCE FEEDBACK REASONING FOR CELLULAR PERTURBATION PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Predicting cellular responses to unseen perturbations requires generalizing from sparse experimental knowledge across interventions and cellular contexts. Which experimental evidence is most useful can depend not only on the perturbation query but also on the current response prediction, motivating feedback between prediction and evidence selection. We therefore introduce Hypothesis-Evidence Feedback Reasoning (**HEFR**), a unified PROPOSE–REASON–REALIZE framework that turns response prediction into an adaptive process of hypothesis initialization, hypothesis-evidence interaction, and cell-population realization. Starting from an initial latent hypothesis of the perturbation response, HEFR iteratively uses the current hypothesis to retrieve evidence that grounds its revision into the next hypothesis. This loop terminates once further reasoning is no longer expected to yield substantive refinement. The final hypothesis is decoded into an aggregate response and realized as a heterogeneous population of perturbed cells consistent with that response. Across chemical and genetic perturbation benchmarks, HEFR substantially outperforms existing methods in key metrics of aggregate accuracy and distributional fidelity. Theoretical and biological interpretability analyses further demonstrate the validity of HEFR. Ultimately, HEFR shifts virtual-cell modeling from passive pattern mapping toward adaptive, feedback-driven scientific reasoning. Code will be released publicly upon publication.

1 INTRODUCTION

A central goal of virtual-cell modeling is to predict cellular responses to unseen genetic and chemical interventions across diverse contexts. Such predictions require synthesizing partial clues distributed across prior studies, including responses to related perturbations, cross-context effects, and biological networks. The primary challenge is not merely encoding this prior knowledge, but determining which evidence is relevant to a given query. While existing models powerfully represent these priors (Lotfollahi et al., 2023; Hetzel et al., 2022; Roohani et al., 2024; Adduri et al., 2026), they usually follow a prescribed inference path that treats prior knowledge as static, one-shot context rather than dynamic evidence whose relevance evolves alongside the response hypothesis (Fig. 1a).

Crucially, prior knowledge becomes informative evidence primarily when evaluated against an active hypothesis. An initial hypothesis reveals what remains unexplained, causing the relevance of prior observations to shift as the hypothesis evolves. A principled model should therefore use its evolving predictions to guide *what knowledge to consult next* and *when further refinement is unnecessary*, establishing a tight feedback loop between hypothesis revision and evidence selection. However, grounding revisions solely in retrieved evidence risks mere interpolation. Effective generalization requires balancing experimental grounding with bounded innovation to explore response directions beyond observed contexts. Finally, because cellular responses are inherently stochastic, the refined aggregate hypothesis must be *realized* across single cells, capturing population heterogeneity rather than collapsing to an average profile.

To address these challenges, we introduce Hypothesis-Evidence Feedback Reasoning (HEFR), establishing adaptive hypothesis-evidence interaction as the central organizing principle of cellular perturbation prediction. HEFR follows a unified PROPOSE–REASON–REALIZE process: PROPOSE forms an initial latent response hypothesis from the perturbation and cell context. During REASON, a hypothesis-driven retriever utilizes the current hypothesis to extract relevant evidence from prior

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.

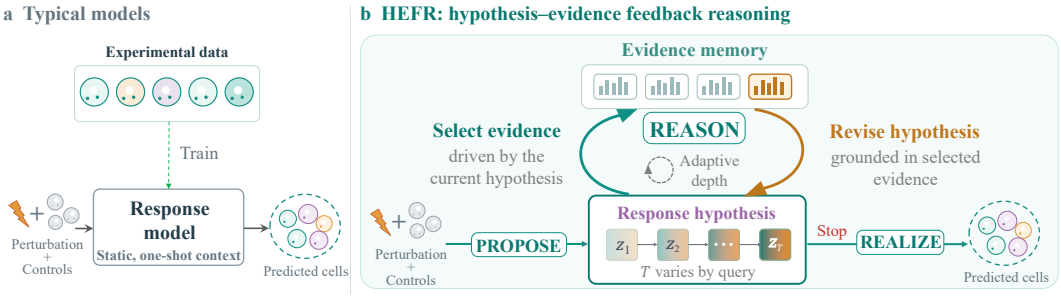

Figure 1: **Comparison of typical perturbation models and HEFR.** (a) Typical models rely on static, one-shot context throughout inference. (b) HEFR couples evidence-grounded latent hypothesis revision with evidence selection: each revised hypothesis guides the next evidence retrieval, forming a feedback loop that terminates adaptively before cell-population realization.

knowledge at each step. Next, an evidence-grounded revision module updates the response hypothesis by balancing retrieved evidence with bounded innovation. Subsequently, a value-guided critic accepts the proposed update only if its expected gain justifies further reasoning; otherwise, the process terminates at the present reasoning step. The reasoning path is specific to the biological query: both the evidence sequence and the reasoning depth emerge dynamically from the evolving prediction itself. Finally, REALIZE conditions individual control cells on the refined hypothesis, producing a heterogeneous perturbed population while preserving the predicted overall response (Fig. 1b).

Across chemical and genetic perturbation benchmarks, HEFR substantially outperforms other methods in key metrics of aggregate accuracy and distributional fidelity. Theoretical and biological interpretability analyses provide further support for its adaptive, feedback-loop reasoning mechanism.

We summarize our key contributions as follows:

- We introduce HEFR, a unified PROPOSE–REASON–REALIZE framework that reformulates cellular perturbation prediction as an adaptive hypothesis-evidence feedback reasoning process rather than a prescribed feedforward mapping.
- We design a *hypothesis-driven retriever* that queries prior knowledge to select evidence relevant to the current hypothesis, an *evidence-grounded revision module* that refines the response hypothesis by balancing experimental grounding with bounded innovation, and a *value-guided critic* that enables adaptive reasoning termination.
- We demonstrate strong performance across chemical and genetic perturbation benchmarks at both aggregate and distributional levels, supported by theoretical insights and interpretable biological analyses.

2 RELATED WORK

Single-cell perturbation tasks. Single-cell perturbation modeling aims to predict transcriptomic responses to unmeasured genetic and chemical interventions across diverse cellular contexts (Norman et al., 2019; Srivatsan et al., 2020). To track progress, community benchmarks evaluate generalization to unseen perturbations, novel combinations, and out-of-distribution cell types (Szałata et al., 2024; Wu et al., 2026; Wei et al., 2026). However, recent analyses reveal that model performance remains highly sensitive to evaluation protocols and systematic variation (Viñas Torné et al., 2025; Schäfer et al., 2026), with complex non-linear architectures often failing to consistently outperform simple linear or additive baselines (Ahlmann-Eltze et al., 2025).

Cellular perturbation modeling. Existing methods typically combine latent response modeling, biological priors, and population generation. scGen models perturbations through latent shifts (Lotfollahi et al., 2019), while CPA (Lotfollahi et al., 2023), chemCPA (Hetzel et al., 2022), biolord (Piran et al., 2024), and SAMS-VAE (B. & K., 2023) disentangle or compose perturbation and cellular factors. GEARS (Roohani et al., 2024) incorporates gene-interaction structure; Geneformer (Theodoris et al., 2023), scGPT (Cui et al., 2024), and scFoundation (Hao et al., 2024a) learn representations from cell atlases, while Tahoe-x1 (Gandhi et al., 2025) and State (Adduri et al., 2026)

exploit large perturbation collections. For population prediction, CellOT (Bunne et al., 2023) learns transport maps, CellFlow (Klein et al., 2025) and scDFM (Yu et al., 2026) use conditional flow matching, and PerturbNet (Yu et al., 2025) and PerturbDiff (Yuan et al., 2026) generate response distributions for unseen interventions. While these approaches substantially advance response representation and generative modeling, they operate under static conditioning without feedback between response prediction and evidence selection.

Retrieval and iterative inference. Recent methods incorporate experimental observations directly into prediction: STACK (Dong et al., 2026) utilizes context cells as in-context examples, while PT-RAG (Di Francesco et al., 2026) aggregates embeddings of retrieved perturbations via a learned cell-type-aware selector. Although these methods demonstrate the value of non-parametric experimental evidence, their retrieval remains static and anchored solely to the initial query. Conversely, iterative computation in perturbation modeling, such as latent denoising in X-Cell (Wang et al., 2026) or sequential conditioning in D^2R^2 (Fan et al., 2026), serves the generative sampling process rather than evidence-grounded hypothesis refinement. While adaptive computation and iterative retrieval are widely studied in broader machine learning for multi-step reasoning (Trivedi et al., 2023; Asai et al., 2024; Hao et al., 2024b; Geiping et al., 2025; Graves, 2016; Banino et al., 2021), HEFR is distinct in allowing the evolving response hypothesis itself to dynamically steer evidence acquisition and dictate reasoning depth.

3 METHODOLOGY

Overview. A query $q = (p, c, \mathcal{X}^{\mathrm{ctrl}})$ specifies a genetic or chemical perturbation p, a target cellular context c, and N control cells measured over G genes, $\mathcal{X}^{\mathrm{ctrl}} = \{\mathbf{x}_i^{\mathrm{ctrl}}\}_{i=1}^{N} \subset \mathbb{R}_+^G$. We predict $\widehat{\mathcal{X}}^{\mathrm{pert}} = \{\widehat{\mathbf{x}}_i^{\mathrm{pert}}\}_{i=1}^{N}$. Training observes an independent population of $\widetilde{N}$ perturbed cells, $\mathcal{X}^{\mathrm{pert}} = \{\mathbf{x}_\ell^{\mathrm{pert}}\}_{\ell=1}^{\widetilde{N}} \subset \mathbb{R}_+^G$; cell pairing and equal population sizes are unnecessary. *For concise notation, we omit query subscripts q on single-instance variables*. The perturbed mean $\overline{\mathbf{x}}^{\mathrm{pert}}$ and control mean $\overline{\mathbf{x}}^{\mathrm{ctrl}}$ define the aggregate logFC target

$$\mathbf{y} = \log_2\left(\frac{\overline{\mathbf{x}}^{\mathrm{pert}} + \delta}{\overline{\mathbf{x}}^{\mathrm{ctrl}} + \delta}\right) \in \mathbb{R}^G, \tag{1}$$

where $\delta > 0$ is a pseudocount and operations are element-wise. HEFR follows the PROPOSE–REASON–REALIZE computation (Fig. 2):

$$\boxed{\begin{aligned} \mathbf{z}_1 &= \mathrm{Propose}(q), \\ \big((\mathbf{z}_k)_{k=1}^{T},\, T\big) &= \mathrm{Reason}(\mathbf{z}_1, \mathcal{M}), \\ \widehat{\mathcal{X}}^{\mathrm{pert}} &= \mathrm{Realize}_\psi(\mathcal{X}^{\mathrm{ctrl}}, \mathbf{z}_T). \end{aligned}} \tag{2}$$

The d-dimensional hypothesis $\mathbf{z}_k$ evolves through the ordered trajectory $(\mathbf{z}_k)_{k=1}^{T}$, using evidence memory $\mathcal{M}$ until adaptive stopping at T. REALIZE combines the terminal hypothesis with control cells to generate the final predictions.

3.1 PROPOSE: AN INITIAL LATENT HYPOTHESIS

The PROPOSE stage forms an initial response hypothesis from the perturbation and the basal state of the target cells, Fig. 2. We represent the perturbation p by a descriptor $\mathbf{h}_p \in \mathbb{R}^r$: a Morgan fingerprint derived from compound SMILES or an ESM embedding of the protein encoded by a perturbed gene. A fixed basal embedding $\mathbf{b}$ summarizes the gene-wise means and standard deviations of the matched control cells. Together with cellular context c, these inputs yield the initial aggregate response $\mathbf{a} = f_0(\mathbf{h}_p, c, \mathbf{b}) \in \mathbb{R}^G$. We fit the response prediction module f_0 by kernel ridge regression, combining perturbation similarity, context agreement, and basal similarity for initial prediction.

To define a latent space for hypothesis revision, we calculate the empirical response mean $\overline{\mathbf{y}}_{\mathrm{tr}} \in \mathbb{R}^G$ and leading PCA axes $\mathbf{U} \in \mathbb{R}^{G\times d}$ ($\mathbf{U}^\top\mathbf{U} = \mathbf{I}_d$) from the training perturbation responses. The latent projection module maps the initial prediction to an initial latent hypothesis:

$$\mathbf{z}_1 = \mathbf{U}^\top\big(\mathbf{a} - \overline{\mathbf{y}}_{\mathrm{tr}}\big) \in \mathbb{R}^d. \tag{3}$$

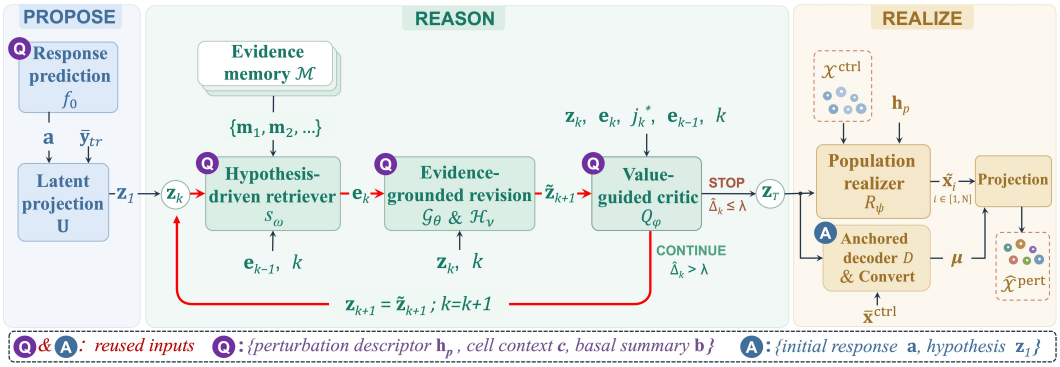

Figure 2: **The PROPOSE–REASON–REALIZE framework of HEFR.** PROPOSE forms an initial latent hypothesis. REASON combines hypothesis-driven retrieval, evidence-grounded revision, and a value-guided critic that accepts revisions or stops reasoning. REALIZE generates a heterogeneous perturbed-cell population whose mean expression yields the final logFC response.

Reasoning revises this hypothesis in latent space, while an anchored decoder D maps it to a feasible aggregate logFC response. For the query's control mean, we define the element-wise logFC lower bound $\mathbf{L} = \log_2\left(\frac{\delta}{\bar{\mathbf{x}}^{\text{ctrl}}+\delta}\right)$. We define the decoder as

$$D(\mathbf{z}) = \max\{\mathbf{a} + \mathbf{U}(\mathbf{z} - \mathbf{z}_1), \mathbf{L}\} \in \mathbb{R}^G, \tag{4}$$

where the maximum is element-wise. This lower bound ensures that the inverse logFC transform yields a nonnegative expression mean. Thus, stopping at $\mathbf{z}_1$ recovers $\max\{\mathbf{a}, \mathbf{L}\}$, the feasible response associated with the initial prediction, and recovers $\mathbf{a}$ itself whenever $\mathbf{a} \ge \mathbf{L}$ element-wise.

3.2 REASON: THE HYPOTHESIS–EVIDENCE–VALUE LOOP

Memory construction. Let $\mathcal{I}$ index the perturbation experiment records in the training set, each comprising perturbation p_j, context c_j, response $\mathbf{y}_j$, and control embedding $\mathbf{b}_j$. They define the evidence memory $\mathcal{M} = \{(p_j, c_j, \mathbf{y}_j, \mathbf{b}_j) : j \in \mathcal{I}\}$. Responses are stored in the shared latent coordinates and accessed in the target context c as

$$\mathbf{m}_j = \mathcal{T}_{c_j \to c}\left[\mathbf{U}^\top(\mathbf{y}_j - \bar{\mathbf{y}}_{\text{tr}})\right], \qquad j \in \mathcal{I}. \tag{5}$$

Here $\mathcal{T}_{c_j \to c}$ is an affine map learned by ridge regression on paired latent responses to perturbations shared between the source and target training contexts; it reduces to the identity when $c_j = c$. For an unseen context, a control-based RBF router assigns weights over observed contexts and combines their transported memories and context statistics. Transport details are provided in Appendix A.1. Every record remains eligible; transport changes its representation, not memory membership.

Hypothesis-driven retriever. At step k, a scorer s_ω estimates the utility of each memory record $\mathbf{m}_j$ for refining the current latent hypothesis $\mathbf{z}_k$. The scorer applies a tree ensemble to predefined features capturing hypothesis–evidence geometry, perturbation similarity, context affinity, basal-state mismatch, normalized depth, and similarity to the preceding evidence. We select the record with the highest predicted utility as the retrieved evidence:

$$\mathbf{e}_k = \mathbf{m}_{j_k^\star}, \quad j_k^\star = \arg\max_{j \in \mathcal{I}} \widehat{u}_{jk}, \quad \widehat{u}_{jk} = s_\omega(\mathbf{h}_p, c, \mathbf{b}, \mathbf{z}_k, \mathbf{m}_j, \mathbf{e}_{k-1}, k), \tag{6}$$

where $\mathbf{e}_0 = \mathbf{0}$. We train s_ω to regress an interpolation-based utility proxy: the normalized reduction in latent error obtained by moving from $\mathbf{z}_k$ toward $\mathbf{m}_j$. With $\mathbf{t} = \mathbf{U}^\top(\mathbf{y} - \bar{\mathbf{y}}_{\text{tr}})$ and $\ell_{\text{lat}}(\mathbf{z}) = \|\mathbf{t} - \mathbf{z}\|_2^2$, the target is

$$u_{jk} = \frac{\ell_{\text{lat}}(\mathbf{z}_k) - \ell_{\text{lat}}(\mathbf{z}_k + \beta(\mathbf{m}_j - \mathbf{z}_k))}{\ell_{\text{lat}}(\mathbf{z}_k) + \varepsilon_u}, \tag{7}$$

where β controls interpolation and $\varepsilon_u > 0$ ensures numerical stability. Scorer training uses group-disjoint episodes, with each query response excluded from the training set of f_0 and evidence memory. Ground-truth responses are used only to construct training targets; inference selects evidence using $\widehat{u}_{jk}$. Because scores depend jointly on the current hypothesis and candidate evidence, candidate utilities are reassessed after each hypothesis revision.

Evidence-grounded revision. Conditioned on the retrieved evidence $\mathbf{e}_k$, a conditional MLP $\mathcal{G}_\theta$ predicts a correction to $\mathbf{z}_k$, while an MLP $\mathcal{H}_\nu$ predicts the residual left by that correction:

$$\begin{aligned} \mathbf{g}_k &= \mathcal{G}_\theta(\mathbf{h}_p, c, \mathbf{b}, \mathbf{z}_k, \mathbf{e}_k, k), \\ \mathbf{v}_k &= \mathcal{H}_\nu(\mathbf{h}_p, c, \mathbf{b}, \mathbf{z}_k, \mathbf{e}_k, k). \end{aligned} \tag{8}$$

The evidence-relative direction $\mathbf{e}_k - \mathbf{z}_k$ and the learned correction $\mathbf{g}_k$ define the local span $\mathcal{L}_k = \mathrm{span}\{\mathbf{e}_k - \mathbf{z}_k, \mathbf{g}_k\}$. We retain only the component of $\mathbf{v}_k$ orthogonal to this span, bound its magnitude, and update the hypothesis:

$$\widetilde{\mathbf{z}}_{k+1} = \mathbf{z}_k + \eta\big(\mathbf{g}_k + \gamma \mathbf{i}_k\big), \quad \mathbf{i}_k = \mathrm{clip}_{\kappa(\|\mathbf{g}_k\|_2 + \varepsilon_{\mathrm{inn}})}(\mathbf{v}_k^\perp), \quad \mathbf{v}_k^\perp = \mathbf{v}_k - \mathrm{proj}_{\mathcal{L}_k}(\mathbf{v}_k). \tag{9}$$

Here $\mathrm{proj}_{\mathcal{L}_k}$ denotes orthogonal projection, and clip_R rescales vectors whose norm exceeds R to norm R. The parameters $\kappa \geq 0$, $\eta > 0$, and $\gamma \geq 0$ control the innovation bound, step size, and innovation contribution, respectively; $\varepsilon_{\mathrm{inn}} > 0$ is a stabilizer. Thus, evidence conditions the learned revision without restricting it to interpolation toward $\mathbf{e}_k$, while $\mathbf{i}_k$ provides bounded innovation orthogonal to the local span. The candidate hypothesis $\widetilde{\mathbf{z}}_{k+1}$ is then fed into the value-guided critic.

Value-guided critic. The learnable critic Q_ϕ predicts whether the candidate hypothesis improves the aggregate logFC response. Its offline supervision is the realized cosine gain computed using the decoder D for both the current and candidate hypotheses:

$$\Delta_k = \mathcal{C}\big(D(\widetilde{\mathbf{z}}_{k+1}), \mathbf{y}\big) - \mathcal{C}\big(D(\mathbf{z}_k), \mathbf{y}\big), \tag{10}$$

where $\mathcal{C}$ denotes cosine similarity. To predict the expected revision gain, the critic conditions on features summarizing the query, current and candidate hypotheses, step depth, and evidence context:

$$\widehat{\Delta}_k = Q_\phi\big(\mathbf{h}_p, c, \mathbf{b}, \mathbf{z}_k, \mathbf{e}_k, \widetilde{\mathbf{z}}_{k+1}, k; j_k^\star, \mathbf{e}_{k-1}\big). \tag{11}$$

We fit the critic by squared-error regression on calibration trajectories. For $k < K_{\max}$, threshold $\lambda \geq 0$ determines STOP or CONTINUE:

$$\begin{cases} \widehat{\Delta}_k \leq \lambda & \Longrightarrow \quad T = k, \qquad \mathbf{z}_T = \mathbf{z}_k, \\ \widehat{\Delta}_k > \lambda & \Longrightarrow \quad \mathbf{z}_{k+1} = \widetilde{\mathbf{z}}_{k+1}. \end{cases} \tag{12}$$

Reasoning ends at the first stop decision or upon reaching $K_{\max}$ retained states, with at most $K_{\max} - 1$ candidate attempts. The terminal latent hypothesis $\mathbf{z}_T$ then conditions population realization.

3.3 REALIZE: A TERMINAL-CONDITIONED CELL POPULATION

An encoder–decoder $\mathcal{R}_\psi$ modulates each control cell with $\mathbf{z}_T$ and the perturbation descriptor $\mathbf{h}_p$, and decodes the modulated embeddings into nonnegative preliminary cell profiles:

$$\widetilde{\mathbf{x}}_i = \mathcal{R}_\psi\big(\mathbf{x}_i^{\mathrm{ctrl}}, \mathbf{z}_T, \mathbf{h}_p\big) \in \mathbb{R}_+^G, \qquad i = 1, \ldots, N. \tag{13}$$

To anchor these profiles to the aggregate response, we convert the terminal prediction $\widehat{\mathbf{y}} = D(\mathbf{z}_T)$ into the target expression mean. Since $\widehat{\mathbf{y}} \geq \mathbf{L}$ element-wise, its inverse logFC is nonnegative:

$$\boldsymbol{\mu} = \big(\overline{\mathbf{x}}^{\mathrm{ctrl}} + \delta\big) \odot 2^{\widehat{\mathbf{y}}} - \delta \in \mathbb{R}_+^G, \tag{14}$$

where $\odot$ is element-wise multiplication. A gene-wise projection then rescales the preliminary profiles to match $\boldsymbol{\mu}$ exactly:

$$\widehat{x}_{i,g}^{\mathrm{pert}} = \mu_g \frac{\big(\widetilde{x}_{i,g} + \varepsilon\big)}{N^{-1} \sum_{\ell=1}^{N} \widetilde{x}_{\ell,g} + \varepsilon}, \qquad i = 1, \ldots, N, \quad g = 1, \ldots, G. \tag{15}$$

Here $\varepsilon > 0$ is a stabilizer. Thus, $\frac{1}{N}\sum_{i=1}^{N} \widehat{\mathbf{x}}_i^{\mathrm{pert}} = \boldsymbol{\mu}$, coupling cell profiles to the predicted logFC.

3.4 MODEL LEARNING

We first fit the initial response prediction module f_0, then train the scorer s_ω and revision modules $\mathcal{G}_\theta$ and $\mathcal{H}_\nu$. With these modules fixed, we fit the critic Q_ϕ. Finally, with PROPOSE and REASON frozen, we train the population realizer $\mathcal{R}_\psi$ to match predicted and observed perturbed-cell distributions.

4 THEORETICAL ANALYSIS

We characterize the attainable value of adapting evidence and reasoning depth, holding the memory and revision operators fixed.

4.1 ADAPTIVE EVIDENCE RETRIEVAL

Let $\ell_{\cos}(\mathbf{z}) = 1 - \mathcal{C}(D(\mathbf{z}), \mathbf{y})$ and let $\mathbf{d}_k(\mathbf{e})$ denote the revision direction in Eq. 9. Define its conditional value as $\overline{\Delta}_k(\mathbf{e}) = \mathbb{E}[\ell_{\cos}(\mathbf{z}_k) - \ell_{\cos}(\mathbf{z}_k + \eta \mathbf{d}_k(\mathbf{e})) \mid \mathcal{F}_k]$, where $\mathcal{F}_k$ contains the available target-free information. For a fixed query and finite memory, compare selectors on a common distribution of reachable hypotheses and depths. A fixed selector retains one record; an adaptive selector may reassess its choice.

Proposition 1 (strict value of evidence adaptation). Their optimal expected one-step gains differ by

$$\Gamma_{\mathrm{ev}} = \mathbb{E}\left[\max_{j \in \mathcal{I}} \overline{\Delta}_k(\mathbf{m}_j)\right] - \max_{j \in \mathcal{I}} \mathbb{E}\left[\overline{\Delta}_k(\mathbf{m}_j)\right] \geq 0. \tag{16}$$

The inequality is strict if no single record is optimal almost surely, identifying an advantage when hypotheses benefit from complementary evidence. The learned retriever uses a latent-interpolation surrogate, which need not preserve this actual-revision gain ranking. Appendix B.1.2 evaluates this alignment using realized revision gains.

4.2 ADAPTIVE STOPPING

Fix the evidence and revision policies and construct a potential trajectory by accepting every candidate up to $K_{\max}$. Define $U(t) = \mathcal{C}(D(\mathbf{z}_t), \mathbf{y}) - \lambda(t-1)$, $1 \leq t \leq K_{\max}$. The penalty regularizes accepted revisions rather than measuring total computation. Let $b_k = \overline{\Delta}_k(\mathbf{e}_k) - \lambda$. Assume that once b_k becomes nonpositive, it remains nonpositive at all later steps almost surely. This is a sufficient condition, not a property enforced by the architecture.

Proposition 2 (optimal depth and adaptation gain). Under this condition, the oracle rule

$$\tau^\star = \min(\{k < K_{\max} : b_k \leq 0\} \cup \{K_{\max}\}) \tag{17}$$

maximizes expected utility over stopping rules adapted to $\mathcal{F}_k$. Moreover, with $(x)_+ = \max(x, 0)$,

$$\mathbb{E}[U(\tau^\star)] - \max_{1 \leq K \leq K_{\max}} \mathbb{E}[U(K)] \quad \geq \sum_{k=1}^{K_{\max}-1} \min\{\mathbb{E}[(b_k)_+], \mathbb{E}[(-b_k)_+]\}\,. \tag{18}$$

Expectations range over queries and model randomness. The advantage is strict whenever b_k takes both signs with positive probability at some depth. In practice, the learned critic estimates the conditional expected gain used by this oracle stopping rule. Appendices B.2.2 and B.2.3 assess critic calibration, missed improvements after stopping, and performance against fixed depths with measured computational cost. Proofs appear in Appendices B.1 and B.2.

5 EXPERIMENTS

We evaluate HEFR on various generalization regimes covering chemical and genetic perturbations, complemented by component ablations and analyses of model mechanisms and biological interpretability. Additional experiments across broader settings are provided in the Appendix.

5.1 EXPERIMENTAL SETUP

Benchmarks. For fair comparison, we adopt four generalization settings built on the official splits from Wu et al. (2026) and Wei et al. (2026). Here, we focus on three main settings spanning chemical and genetic perturbations: (1) *unseen perturbations* (holding out drugs in sciPlex3-K562 and genes in Replogle-K562-essential); (2) *unseen combinations* (holding out drug pairs in sciPlex3-comb

Table 1: Performance comparison on unseen perturbations.

		Aggregate Metrics				Distributional Metrics		
	Method	Cos-FC ↑	R-Cos ↓	RMSE ↓	R-RMSE ↓	MMD-G ↓	MMD-P ↓	DEG-R ↑
Drug	Identity floor	–	0.4761	0.0346	0.5087	0.2218	0.2107	0.0124
	chemCPA (Hetzel et al., 2022)	0.6565	0.3890	0.0288	0.3907	9.1520	4.4947	0.1079
	cycleCDR (Huang & Liu, 2024)	0.0276	0.5025	0.2319	0.4979	13.5579	6.7455	0.0302
	CellFlow (Klein et al., 2025)	0.0904	0.4559	0.1338	0.4696	4.8238	3.0490	0.0024
	PerturbNet (Yu et al., 2025)	0.1770	0.4401	0.2476	0.4774	8.7488	7.2971	0.0086
	State (Adduri et al., 2026)	0.6119	0.3791	0.0310	0.4003	0.2417	0.2275	**0.1224**
	HEFR (ours)	**0.6986**	**0.3685**	**0.0231**	**0.3561**	**0.1046**	**0.1095**	0.1019
Gene	Identity floor	–	0.4923	0.0593	0.4977	0.4478	0.4885	0.0074
	GEARS (Roohani et al., 2024)	0.1488	0.4056	0.0633	0.3950	0.4407	0.4788	0.0929
	scGPT (Cui et al., 2024)	0.2724	**0.1661**	0.0572	0.2728	9.0227	3.1956	0.1384
	GenePert (Chen & Zou, 2024)	0.3084	0.2405	0.0547	**0.2456**	–	–	–
	PerturbNet (Yu et al., 2025)	0.0992	0.3104	0.1057	0.4021	5.8664	2.0667	0.0722
	CellFlow (Klein et al., 2025)	0.0967	0.3375	0.0988	0.4057	4.3776	2.7579	0.0783
	State (Adduri et al., 2026)	0.1302	0.4146	0.0665	0.4062	0.6082	0.5827	0.0663
	PT-RAG (Di et al., 2026)	0.1311	0.4108	0.0701	0.3988	0.6111	0.5832	0.0712
	scDFM (Yu et al., 2026)	0.1631	0.4923	0.1016	0.5818	3.2351	2.0880	0.0916
	HEFR (ours)	**0.3192**	0.1903	**0.0521**	0.3026	**0.3347**	**0.3981**	**0.1402**

Identity floor is a no-effect baseline that returns the context-matched control population unchanged.

and gene pairs in Norman19); and (3) *unseen cellular contexts* (holding out entire cell lines A549, K562, and MCF7 in sciPlex3). For datasets with multiple seeds or splits, we report pooled metrics. Detailed per-split breakdowns and results under additional settings are provided in Appendix E.

Evaluation metrics. Following PerturBench (Wu et al., 2026), we evaluate models across aggregate and distributional levels. **Aggregate metrics** comprise logFC cosine (Cos-FC, response direction), mean-expression RMSE (magnitude error), and rank metrics (R-Cos, R-RMSE) for specificity. **Distributional metrics** for single-cell outputs assess cell-population fidelity via MMD in gene (MMD-G) and PCA (MMD-P) spaces, alongside top-50 DEG recall (DEG-R) for responsive gene recovery. Dashes mark metrics that are undefined for a predictor. Details of the evaluator, subsampling, and baseline exclusions are provided in Appendix C.

5.2 Generalization experiments

Unseen perturbations. HEFR leads 6 of 7 metrics on unseen drugs and 5 on unseen genes (Table 1), achieving Cos-FC scores of 0.6986 and 0.3192, respectively. Its leading MMD-G and MMD-

Table 2: Performance comparison on unseen perturbation combinations.

		Aggregate Metrics				Distributional Metrics		
	Method	Cos-FC ↑	R-Cos ↓	RMSE ↓	R-RMSE ↓	MMD-G ↓	MMD-P ↓	DEG-R ↑
Drug	Identity floor	–	0.5429	0.1313	0.5330	1.4082	2.9251	0.0115
	LatentAdditive (Wu et al., 2026)	0.4065	0.4161	0.1043	0.3747	9.7801	5.7517	0.0003
	Decoder-only (Wu et al., 2026)	0.4470	0.3936	**0.1033**	0.3676	–	–	–
	chemCPA (Hetzel et al., 2022)	0.2848	0.4067	0.1062	0.3490	10.0849	6.0635	0.1061
	CPA (Lotfollahi et al., 2023)	0.4773	0.3028	0.2201	0.3160	8.4611	6.0191	**0.1661**
	SAMS-VAE (B. & K., 2023)	0.4695	0.4103	0.1080	0.3607	7.6666	4.4397	0.0000
	biolord (Piran et al., 2024)	0.3985	0.5419	0.1244	0.4223	3.5989	3.0817	0.0003
	CellFlow (Klein et al., 2025)	0.3964	0.3854	0.1119	0.3464	1.3397	2.4111	0.1034
	HEFR (ours)	**0.5181**	**0.3001**	0.1071	**0.3008**	**1.1095**	**2.2406**	0.1128
Gene	Identity floor	–	0.5696	0.0959	0.4836	1.0395	1.7624	0.0052
	LatentAdditive (Wu et al., 2026)	0.6461	0.0812	0.0520	0.0478	6.6223	3.1624	0.0000
	Decoder-only (Wu et al., 2026)	0.7320	0.0213	0.0480	0.0324	–	–	–
	CPA (Lotfollahi et al., 2023)	0.4150	0.0507	0.0863	0.0609	4.2536	2.0021	0.0222
	SAMS-VAE (B. & K., 2023)	0.4846	0.0150	0.0847	0.0203	4.3273	1.9979	0.0000
	scGPT (Cui et al., 2024)	0.6062	0.0628	0.0598	0.0357	5.9125	3.2318	0.0009
	biolord (Piran et al., 2024)	0.0128	0.5053	0.0971	0.5039	3.0665	1.9880	0.0000
	GEARS (Roohani et al., 2024)	0.3974	0.1877	0.0725	0.1229	0.6386	0.8748	0.0381
	GenePert (Chen & Zou, 2024)	0.7449	0.0058	0.0584	0.0425	–	–	–
	PerturbNet (Yu et al., 2025)	0.1961	0.3019	0.1089	0.3193	4.3279	2.4942	0.0239
	CellFlow (Klein et al., 2025)	0.4762	0.0700	0.1113	0.1469	1.4180	2.1629	0.1952
	State (Adduri et al., 2026)	0.4035	0.5367	0.0820	0.5179	0.8206	1.3544	0.2304
	scDFM (Yu et al., 2026)	0.5001	0.1016	0.0705	0.0612	0.7861	0.8622	0.2716
	HEFR (ours)	**0.8153**	**0.0029**	**0.0392**	**0.0106**	**0.3712**	**0.2438**	**0.2941**

Table 3: Performance comparison on unseen cellular contexts.

	Aggregate Metrics				Distributional Metrics		
Method	Cos-FC ↑	R-Cos ↓	RMSE ↓	R-RMSE ↓	MMD-G ↓	MMD-P ↓	DEG-R ↑
Identity floor	–	0.5324	0.0670	0.4028	0.5600	0.8438	0.0348
LatentAdditive (Wu et al., 2026)	0.2568	0.2963	0.1900	0.3333	8.5530	7.0798	0.0000
scGen (Lotfollahi et al., 2019)	0.2235	0.2269	0.0932	0.2824	5.0301	2.7849	0.0015
chemCPA (Hetzel et al., 2022)	0.1480	0.4398	0.3130	0.5000	15.7958	12.2045	0.0059
CellOT (Bunne et al., 2023)	0.2888	0.2870	0.1793	0.3565	9.5958	7.6072	0.0052
cycleCDR (Huang & Liu, 2024)	0.1125	0.3843	0.2520	0.3611	11.0657	8.3605	0.1267
biolord (Piran et al., 2024)	0.1521	0.1806	0.1242	0.3519	2.7150	2.1529	0.0000
CellFlow (Klein et al., 2025)	0.1265	0.3380	0.3472	0.4676	7.5224	5.9827	0.0348
State (Adduri et al., 2026)	0.2926	0.3056	**0.0622**	0.1667	0.5013	0.6053	**0.2022**
HEFR (ours)	**0.3836**	**0.0815**	0.0702	**0.1141**	**0.4216**	**0.5008**	0.1861

P scores on both tasks show that accurate aggregate response prediction is accompanied by improved population fidelity, supporting generalization at both resolutions.

Unseen perturbation combinations. HEFR ranks first across all 7 metrics on genetic combinations and 5 on drug pairs (Table 2). On genetic combinations, it improves upon the LatentAdditive in both response direction and population fidelity. On drug pairs, HEFR leads in Cos-FC (0.5181) and distributional alignment (MMD-G: 1.11, MMD-P: 2.24); while CPA achieves better DEG recall, its population fidelity severely collapses (MMD-G > 8). These results highlight its ability to recover both aggregate effects and heterogeneous response distributions for unseen combinations.

Unseen cellular contexts. HEFR leads 5 of 7 metrics under context shifts (Table 3), with the strongest response direction (Cos-FC: 0.3836), perturbation specificity (R-Cos: 0.0815; R-RMSE: 0.1141), and population fidelity. Although State achieves lower RMSE and higher DEG recall, HEFR better distinguishes perturbation-specific responses while recovering the target population, demonstrating complementary strengths beyond mean-expression fit.

5.3 Ablation studies

Using the unseen-perturbation setting on sciPlex3, we assess the contributions of Propose, Reason, and Realize, as well as the individual mechanisms within Reason. Detailed ablation definitions and additional controls are provided in Appendix D.1.

Distinct roles of the three stages. Relative to *Full* HEFR, replacing the learned proposal with the mean training response lowers Cos-FC from 0.6986 to 0.6581, while bypassing Reason and returning $\mathbf{z}_1$ lowers it to 0.6218 (Table 4). These stage-level ablations support the contributions of Propose and Reason as a whole. Replacing the realized population with rescaled control cells preserves the terminal aggregate response but increases MMD-G from 0.1046 to 0.2903, demonstrating the contribution of Realize to population structure.

Evidence guides a learned revision. As shown in Table 4, *Fixed-state* lowers Cos-FC from 0.6986 to 0.6583, supporting evidence reassessment as the hypothesis evolves. The $\gamma = 0$ variant outperforms *Interpol.* (0.6708 versus 0.6601); both disable innovation, so this comparison isolates the benefit of learned $\mathbf{g}_k$ over direct interpolation toward $\mathbf{e}_k$. *Full* further improves on $\gamma = 0$ (0.6986 versus 0.6708), showing the additional contribution of bounded innovation.

Table 4: Contributions of the three stages and the revision mechanisms.

	Full	Propose	Reason				Realize
Metric	HEFR	Mean	$T = 1$	Fixed-state	Interpol.	$\gamma = 0$	Control
Cos-FC ↑	**0.6986**	0.6581	0.6218	0.6583	0.6601	0.6708	0.6986
MMD-G ↓	**0.1046**	0.1591	0.2042	0.2276	0.1825	0.1676	0.2903

Mean: mean training response as anchor; $T = 1$: bypass Reason; *Fixed-state*: score at $\mathbf{z}_1$, update $\mathbf{z}_k$; *Interpol.*: $\mathbf{g}_k = \mathbf{e}_k - \mathbf{z}_k$, $\gamma = 0$; $\gamma = 0$: learned $\mathbf{g}_k$ without innovation; *Control*: control cells rescaled to *Full*'s terminal mean.

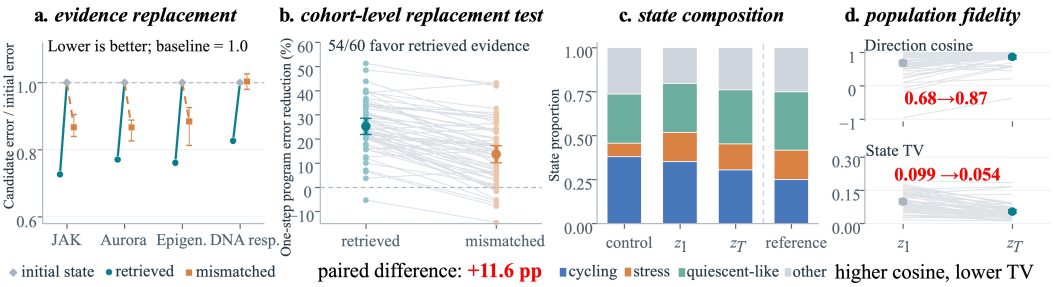

Figure 3: **Biological interpretation analysis.** **(a)** Relative program error for four mechanism-family examples under retrieved and mismatched evidence; error bars show the replacement range. **(b)** Paired one-step error reduction across 60 queries. **(c)** Control, initial, terminal, and reference state compositions for CYC116–MCF7. **(d)** Initial-to-terminal changes in direction cosine and state-proportion total variation (TV). Gray lines denote queries; large points and error bars show means and 95% drug-cluster bootstrap CIs. Higher cosine and lower TV indicate better agreement.

Multi-step feedback. As shown in Table 5, *Fixed-depth* reasoning with current-state retrieval improves Cos-FC from 0.6746 to 0.6807 over *One revision* and outperforms initial-state retrieval at the same depth (0.6526), separating the benefits of repeated reasoning and hypothesis feedback. *Adaptive stopping* further raises Cos-FC to 0.6986 while reducing mean depth from 4.00 to 3.20.

5.4 Mechanistic and Biological Interpretation

Evidence guides biologically coherent refinement. We compare the originally retrieved evidence with mechanism-mismatched replacements at the same initial hypothesis, holding the revision module and step size fixed. Fig. 3a shows one representative query from each of four prespecified mechanism families. Within each family, the displayed query is closest to the median paired difference. In all examples, retrieved evidence yields lower normalized candidate program error than the mismatched replacement average. Across the cohort, the mean error reduction is 25.3% with retrieved evidence and 13.7% with replacements, a paired difference of 11.6 (95% CI: 9.6–13.5) percentage points (Fig. 3b).

Table 5: Separating revision, hypothesis feedback, and adaptive stopping.

Reasoning policy	Retrieval state	Cos-FC	Mean T
No revision	–	0.6218	1.00
One revision	Initial $\mathbf{z}_1$	0.6746	2.00
Fixed depth	Initial $\mathbf{z}_1$	0.6526	4.00
Fixed depth	Current $\mathbf{z}_k$	0.6807	4.00
Adaptive	Initial $\mathbf{z}_1$	0.6583	3.65
Adaptive	Current $\mathbf{z}_k$	0.6986	3.20

Terminal hypotheses capture shifts in population composition. For the illustrative CYC116–MCF7 query, the terminal population more closely matches the reference state composition, reducing total-variation distance from 0.102 to 0.055 (Fig. 3c). Across the cohort, mean direction Cos-FC increases from 0.68 to 0.87, while mean state-proportion total variation decreases from 0.099 to 0.054, with improvements in 56/60 queries (Fig. 3d). These complementary readouts evaluate both the direction of population change and the accuracy of the resulting composition.

6 Conclusion

In this paper, we introduced HEFR, an adaptive hypothesis-evidence feedback reasoning framework for cellular perturbation prediction. Under a Propose–Reason–Realize paradigm, the current prediction guides a *hypothesis-driven retriever* to gather empirical evidence, which in turn informs an *evidence-grounded revision* module to refine latent dynamics, while a *value-guided critic* governs adaptive stopping. Theoretical and biological interpretation analyses further confirm the validity of this feedback loop, while benchmark evaluations across unseen perturbations, combinations, and contexts validate robust aggregate accuracy and population fidelity. Overall, HEFR shifts virtual-cell modeling from static feedforward generation toward hypothesis-evidence feedback reasoning.

AI USE STATEMENT

In this work, we used generative AI tools for no tasks with required disclosure. We have not used generative AI tools for developing theoretical models or conceptual frameworks, formulating mathematical claims, providing critical ingredients for proving mathematical claims, assisting in the writing of proofs, proposing or refining hypotheses, designing or providing feedback on research methodology or experiments, implementing methods, or interpreting results, and generating synthetic datasets, assisting with translation, cleaning and reformatting datasets, or supporting qualitative, and thematic data analysis are not applicable to this work. Additionally, we used generative AI tools for language polishing, grammatical editing, and formatting LaTeX code. We have reviewed all AI-assisted work. Specifically, AI assistance was strictly limited to language refinement and LaTeX code formatting; all underlying research hypotheses, mathematical claims, model architectures, code implementations, experimental designs, and result interpretations were developed entirely by the human authors and verified for correctness. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

In this work, we adhere strictly to established research integrity and ethical standards. All datasets utilized in our empirical evaluations, including the sciPlex3 single-cell perturbation benchmark, are derived entirely from previously published, publicly available, and open-access scientific repositories. No new human or animal experiments were performed by the authors, and no personally identifiable information or private medical records were accessed, raising no privacy, consent, or Institutional Review Board (IRB) compliance concerns. Furthermore, the proposed framework, HEFR, is designed exclusively for basic scientific research in computational biology and virtual-cell perturbation modeling. Its predictive outputs consist of abstract transcriptomic response distributions intended to accelerate biological hypothesis generation, posing no dual-use biosecurity threats, harmful biological synthesis risks, or direct negative societal impacts.

REPRODUCIBILITY STATEMENT

We are committed to ensuring the full reproducibility of all theoretical formulations and empirical results presented in this work. The core architecture of HEFR, including the PROPOSE–REASON–REALIZE paradigm and its three constituent modules, is formally described in Section 3. All mathematical assumptions, theoretical analyses, and complete proofs are provided in Appendix B. For empirical evaluations, all experiments rely on publicly available single-cell perturbation datasets (e.g., sciPlex3, Norman); the data preprocessing and benchmark split protocols are fully detailed in Section 5 and Appendix C.1. Furthermore, training details, hyperparameter configurations, and hardware environments are documented in Appendix A.3. To facilitate exact reproduction of our results, an anonymized source code repository containing model implementations, training scripts, and evaluation pipelines will be released publicly upon publication.

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.

Constantin Ahlmann-Eltze, Wolfgang Huber, and Simon Anders. Deep-learning-based gene perturbation effect prediction does not yet outperform simple linear baselines. *Nature Methods*, 22(8): 1657–1661, 2025.

Akari Asai, Zeqiu Wu, Yizhong Wang, Avirup Sil, and Hannaneh Hajishirzi. Self-RAG: Learning to retrieve, generate, and critique through self-reflection. In *International Conference on Learning Representations*, 2024.

Andrea Banino, Jan Balaguer, and Charles Blundell. PonderNet: Learning to ponder. *arXiv preprint arXiv:2107.05407*, 2021.

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.

Charlotte Bunne, Stefan G Stark, Gabriele Gut, Jacobo Sarabia Del Castillo, Mitch Levesque, Kjong-Van Lehmann, Lucas Pelkmans, Andreas Krause, and Gunnar Rätsch. Learning single-cell perturbation responses using neural optimal transport. *Nature Methods*, 20(11):1759–1768, 2023.

Yiqun Chen and James Zou. GenePert: Leveraging GenePT embeddings for gene perturbation prediction. *bioRxiv*, 2024.

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.

Andrea Giuseppe Di Francesco, Andrea Rubbi, and Pietro Liò. Retrieval-augmented generation for predicting cellular responses to gene perturbation. *arXiv preprint arXiv:2603.07233*, 2026.

Mingze Dong, Abhinav Adduri, Dhruv Gautam, Christopher Carpenter, Rohan Shah, Chiara Ricci-Tam, Yuval Kluger, Dave P Burke, and Yusuf H Roohani. STACK: In-context learning of single-cell biology. *bioRxiv*, 2026.

Ninghan Fan, Qi Liu, Xunuo Zhu, Yukai Sun, Luyuan Chen, Xuheng Zhou, Yuetian Du, Ming Kong, Xiaojun Zhu, Jie Liu, et al. D^2R^2: Discrete diffusion with regulation reinforcement for single-cell perturbation prediction. *arXiv preprint arXiv:2608.15288*, 2026.

Chris J. Frangieh, Johannes C. Melms, Pratiksha I. Thakore, Kathryn R. Geiger-Schuller, Patricia Ho, Adrienne M. Luoma, Brian Cleary, Livnat Jerby-Arnon, Shruti Malu, Michael S. Cuoco, Maryann Zhao, Casey R. Ager, Meri Rogava, Lila Hovey, Asaf Rotem, Chantale Bernatchez, Kai W. Wucherpfennig, Bruce E. Johnson, Orit Rozenblatt-Rosen, Dirk Schadendorf, Aviv Regev, and Benjamin Izar. Multimodal pooled Perturb-CITE-seq screens in patient models define mechanisms of cancer immune evasion. *Nature Genetics*, 53(3):332–341, 2021. doi: 10.1038/s41588-021-00779-1.

Shreshth Gandhi et al. Tahoe-x1: Scaling perturbation-trained single-cell foundation models to 3 billion parameters. *bioRxiv*, 2025.

Jonas Geiping, Sean McLeish, Neel Jain, John Kirchenbauer, Siddharth Singh, Brian R Bartoldson, Bhavya Kailkhura, Abhinav Bhatele, and Tom Goldstein. Scaling up test-time compute with latent reasoning: A recurrent depth approach. *arXiv preprint arXiv:2502.05171*, 2025.

Alex Graves. Adaptive computation time for recurrent neural networks. *arXiv preprint arXiv:1603.08983*, 2016.

Minsheng Hao, Jing Gong, Xin Zeng, Chiming Liu, Yucheng Guo, Xingyi Cheng, Taifeng Wang, Jianzhu Ma, Xuegong Zhang, and Le Song. Large-scale foundation model on single-cell transcriptomics. *Nature Methods*, 21(8):1481–1491, 2024a.

Shibo Hao, Sainbayar Sukhbaatar, DiJia Su, Xian Li, Zhiting Hu, Jason Weston, and Yuandong Tian. Training large language models to reason in a continuous latent space. *arXiv preprint arXiv:2412.06769*, 2024b.

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.

Wei Huang and Hui Liu. Predicting single-cell cellular responses to perturbations using cycle consistency learning. *Bioinformatics*, 40(Supplement_1):i462–i470, 2024.

Dominik Klein, Jonas Simon Fleck, Daniil Bobrovskiy, et al. CellFlow enables generative single-cell phenotype modeling with flow matching. *bioRxiv*, 2025.

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, Leon Hetzel, Yuge Ji, Ignacio L Ibarra, Sanjay R Srivatsan, Mohsen Naghipourfar, Riza M Daza, Beth Martin, et al. Predicting cellular responses to complex perturbations in high-throughput screens. *Molecular Systems Biology*, 19(6):MSB202211517, 2023.

Thomas M Norman, Max A Horlbeck, Joseph M Replogle, Alex Y Ge, Albert Xu, Marco Jost, Luke A Gilbert, and Jonathan S Weissman. Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. *Science*, 365(6455):786–793, 2019.

Zoe Piran, Niv Cohen, Yedid Hoshen, and Mor Nitzan. Disentanglement of single-cell data with biolord. *Nature Biotechnology*, 42(11):1678–1683, 2024.

Joseph M Replogle, Reuben A Saunders, Angela N Pogson, Jeffrey A Hussmann, Alexander Lenail, Alina Guna, Lauren Mascibroda, Eric J Wagner, Karen Adelman, Gila Lithwick-Yanai, et al. Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell*, 185(14):2559–2575, 2022.

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

Philipp Sven Lars Schäfer, Kendall Reid, Zach Boldyga, Ekin Deniz Aksu, Hugo Hakem, and Julio Saez-Rodriguez. Towards principled evaluation of single-cell perturbation prediction models. *bioRxiv*, 2026.

Sanjay R Srivatsan, José L McFaline-Figueroa, Vijay Ramani, Lauren Saunders, Junyue Cao, Jonathan Packer, Hannah A Pliner, Dana L Jackson, Riza M Daza, Lena Christiansen, et al. Massively multiplex chemical transcriptomics at single-cell resolution. *Science*, 367(6473):45–51, 2020.

Artur Szałata, Andrew Benz, Robrecht Cannoodt, Mauricio Cortes, Jason Fong, Sunil Kuppasani, Richard Lieberman, Tianyu Liu, Javier A Mas-Rosario, Rico Meinl, et al. A benchmark for prediction of transcriptomic responses to chemical perturbations across cell types. *Advances in Neural Information Processing Systems*, 37:20566–20616, 2024.

Christina V Theodoris, Ling Xiao, Anant Chopra, Mark D Chaffin, Zeina R Al Sayed, Matthew C Hill, Helene Mantineo, Elizabeth M Brydon, Zexian Zeng, X Shirley Liu, et al. Transfer learning enables predictions in network biology. *Nature*, 618(7965):616–624, 2023.

Harsh Trivedi, Niranjan Balasubramanian, Tushar Khot, and Ashish Sabharwal. Interleaving retrieval with chain-of-thought reasoning for knowledge-intensive multi-step questions. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, 2023.

Ramon Viñas Torné, Maciej Wiatrak, Zoe Piran, Shuyang Fan, Liangze Jiang, Sarah A Teichmann, Mor Nitzan, and Maria Brbić. Systema: a framework for evaluating genetic perturbation response prediction beyond systematic variation. *Nature Biotechnology*, pp. 1050–1059, 2025.

Chloe Wang, Mehran Karimzadeh, Neal G Ravindra, Lexi R Bounds, Nader Alerasool, Ann C Huang, Shihao Ma, Daniel R Gulbranson, Haotian Cui, Yongju Lee, et al. X-Cell: Scaling causal perturbation prediction across diverse cellular contexts via diffusion language models. *bioRxiv*, 2026.

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.

Yan Wu, Esther Wershof, Sebastian Schmon, Marcel Nassar, Błażej Osiński, Ridvan Eksi, Zichao Yan, Rory Stark, Kun Zhang, and Thore Graepel. PerturBench: Benchmarking machine learning models for cellular perturbation analysis. *Advances in Neural Information Processing Systems*, 38, 2026.

Chenglei Yu, Chuanrui Wang, Bangyan Liao, and Tailin Wu. scDFM: Distributional flow matching model for robust single-cell perturbation prediction. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=QSGanMEcUV.

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):960–982, 2025.

Xinyu Yuan, Xixian Liu, Ya Shi Zhang, Zuobai Zhang, Hongyu Guo, and Jian Tang. PerturbDiff: Functional diffusion for single-cell perturbation modeling. *arXiv preprint arXiv:2602.19685*, 2026.

A Implementation and Training Details

Notation. We follow the main text and omit the query subscript q on single-query quantities. Each hypothesis, response, control reference, and terminal depth is understood to belong to the query under consideration. Indices identifying evidence records, reasoning steps, or individual cells are retained. Across-query averages and expectations are specified explicitly where needed. HEFR contains approximately 14.5 million fitted scalar parameters, including neural parameters, regression coefficients, and tree thresholds and leaf values.

A.1 Evidence memory and context transport

The evidence memory contains all eligible training records $(p_j, c_j, \mathbf{y}_j, \mathbf{b}_j)$ in the current split. Every record is scored after transport into the query context; perturbation descriptors are scorer features, not filters for a preliminary shortlist. For a supported source–target pair, the affine map $\mathcal{T}_{c_j \to c}(\mathbf{u}) = \mathbf{A}\mathbf{u} + \mathbf{t}$ is fitted by

$$\min_{\mathbf{A},\mathbf{t}} \sum_{(\mathbf{u},\mathbf{w}) \in \mathcal{P}_{c_j,c}} \|\mathbf{A}\mathbf{u} + \mathbf{t} - \mathbf{w}\|_2^2 + \alpha_{\mathrm{tr}} \|\mathbf{A}\|_F^2, \tag{19}$$

where $\mathcal{P}_{c_j,c}$ pairs latent responses to perturbations observed in both training contexts, with codes $\mathbf{U}^\top(\mathbf{y} - \overline{\mathbf{y}}_{\mathrm{tr}})$. Repeated target records with the same (p, c) are averaged before pairing. These are pairs of aggregate responses, not individual cells. Within-context transport is the identity. For an unseen context, a control-based RBF router weights observed contexts and combines their transported memory members and context statistics. No held-out treated response is used to fit the transport or construct the memory.

A.2 Module inputs and staged learning

The ExtraTrees scorer s_ω uses state–evidence cosine, norms and distances, evidence-history alignment, descriptor similarity, context agreement, normalized depth, and the basal features $\mathbf{b}$, $\mathbf{b} - \mathbf{b}_j$, and their distance. History features use a fixed missing-history encoding at the first step. The conditional MLP $\mathcal{G}_\theta$ and MLP $\mathcal{H}_\nu$ receive query descriptors, context, $\mathbf{b}$, $\mathbf{z}_k$, $\mathbf{e}_k$, their displacement, and norm/cosine/depth summaries. Parameters are shared across reasoning steps.

Training follows Section 3.4: fit the kernel-ridge response prediction module f_0 and training-response PCA, then the scorer and revision modules, followed by the critic with these modules frozen. Scorer targets use the latent interpolation utility in Eq. 7; group-disjoint episodes exclude each training query response from its f_0 fit and memory. The critic regresses realized decoded cosine gains on calibration trajectories. Finally, the realizer is trained to match perturbed-cell distributions with the Propose and Reason frozen. Structural ablations refit affected modules and recalibrate the critic when trajectories change; policy diagnostics instead retain identical frozen modules.

Mean consistency and export checks. The decoder D in Eq. 4 already applies the feasibility floor. Under the same control reference, the complete mean-projected population therefore recovers $D(\mathbf{z}_T)$ exactly, up to numerical precision. Only the pre-floor linear readout can differ from this population-derived logFC; at $T = 1$, the decoded response is $\max\{\mathbf{a}, \mathbf{L}\}$. Cell subsampling can

introduce a separate discrepancy. Export checks verify gene order, query IDs, control references, finite nonnegative expression, and full-population mean agreement with $\boldsymbol{\mu}$.

A.3 TRAINING SPECIFICATION

Data flow and intermediate states. Let $\mathcal{D}_{\mathrm{tr}}$ and $\mathcal{D}_{\mathrm{cal}}$ denote the released training and validation partitions. Response PCA and feature normalization use $\mathcal{D}_{\mathrm{tr}}$ only. Four group-disjoint folds within $\mathcal{D}_{\mathrm{tr}}$ generate training episodes: each episode fits its own f_0, memory, and transport maps on the support folds and initializes queries from the excluded fold. The shared response basis remains fitted on the outer training partition.

The initial scorer is fitted using the interpolation targets in Eq. 7 at these initial states. Training trajectories then alternate full-memory retrieval with $\mathbf{z}_{k+1} = \mathbf{z}_k + \beta(\mathbf{e}_k - \mathbf{z}_k)$, recording candidate utilities and revision examples at every step. The scorer is refitted on the accumulated examples, and the revision heads are fitted as below. These interpolation trajectories construct supervision; deployment uses the learned update in Eq. 9.

Revision and critic objectives. Write $\boldsymbol{\xi}$ for the revision inputs specified above, $\mathbf{t} = \mathbf{U}^\top(\mathbf{y} - \overline{\mathbf{y}}_{\mathrm{tr}})$, and $\mathbf{r} = \mathbf{t} - \mathbf{z}$. With expectations denoting empirical averages over recorded training transitions, the objectives are

$$\mathcal{L}_G = \mathbb{E}\|\mathcal{G}_\theta(\boldsymbol{\xi}) - \mathbf{r}\|_2^2, \tag{20}$$

$$\mathcal{L}_H = \mathbb{E}\,\|\mathcal{H}_\nu(\boldsymbol{\xi}) - [\mathbf{r} - \mathcal{G}_\theta(\boldsymbol{\xi})]\|_2^2\,, \tag{21}$$

$$\mathcal{L}_Q = \mathbb{E}_{\mathcal{D}_{\mathrm{cal}}}(\widehat{\Delta}_k - \Delta_k)^2. \tag{22}$$

The conditional revision MLP is frozen before training the residual MLP on its remaining prediction errors. These heads predict unscaled corrections: η damps their combined update, while projection and clipping constrain the innovation only when constructing candidates.

With retrieval and revision frozen, calibration trajectories accept every candidate up to $K_{\max}$ to obtain the decoded gain targets in Eq. 10. Grouped four-fold out-of-fold critic predictions select λ by mean terminal cosine, breaking ties toward shorter trajectories; the final critic is then fitted on all calibration transitions. Thus calibration responses train the critic and select settings, but never enter the evidence memory.

Population objective and optimization. Let X and Y be independently sampled predicted and observed perturbed populations for one query, with X evaluated after the mean-preserving projection. Let P be a frozen d_P-dimensional cell-PCA transform fitted only on training cells, and let $\mathcal{E}$ denote the biased empirical energy discrepancy in Eq. 35. The population objective is

$$\begin{aligned}\mathcal{L}_R = \frac{\mathcal{E}(X,Y)}{\sqrt{G}} + \frac{\mathcal{E}(PX,PY)}{\sqrt{d_P}} \\ + \frac{0.05}{d_P}\,\|\log(1+\mathrm{Var}(PX)) - \log(1+\mathrm{Var}(PY))\|_2^2\,,\end{aligned} \tag{23}$$

where variances use population normalization. Gradients pass through the realizer and mean projection; the terminal hypothesis and upstream modules remain fixed.

For training queries, terminal conditions are regenerated with fold-specific f_0 fits and memories excluding their query groups. The shared PCA, fitted revision heads, and critic remain frozen: this excludes direct target retrieval but is not full cross-fitting of every model parameter. Calibration queries use the final training memory. The same calibration partition selects the population checkpoint; it is not an independent evaluation set. Test responses are used only for final evaluation.

Descriptors, doses, and unseen contexts. For a perturbation with m components, let $\mathbf{v}_a$ be the component's Morgan fingerprint or fixed ESM embedding. For chemical components define $s_a = \log(1 + d_a/(1\,\mathrm{nM}))$; for genetic interventions without dose metadata set $s_a = 0$. A permutation-invariant descriptor is

$$\mathbf{h}_p = \left[\frac{1}{m}\sum_a \mathbf{v}_a;\ \frac{1}{m}\sum_a s_a\mathbf{v}_a;\ \frac{1}{m}\sum_a s_a;\ m\right]. \tag{24}$$

Chemical fingerprints use radius 2 and 2048 bits. For chemical tasks without dose metadata, the dose blocks are omitted consistently from training and inference. Descriptor scaling and dimensionality reduction are fitted on training descriptors only.

For an unseen context with basal embedding $\mathbf{b}$, let $\overline{\mathbf{b}}_c$ be the mean training basal embedding of observed context c. Define

$$w_c(\mathbf{b}) = \frac{\exp[-\|\mathbf{b} - \overline{\mathbf{b}}_c\|_2^2/(2\sigma_b^2)]}{\sum_{c'} \exp[-\|\mathbf{b} - \overline{\mathbf{b}}_{c'}\|_2^2/(2\sigma_b^2)]}, \qquad \mathbf{m}_j = \sum_c w_c(\mathbf{b})\, \mathcal{T}_{c_j \to c} \left[\mathbf{U}^\top (\mathbf{y}_j - \overline{\mathbf{y}}_{\mathrm{tr}})\right]. \quad (25)$$

The bandwidth σ_b^2 is the median positive squared distance between training context centroids, with fallback 1. The same weights mix known-context response predictions, context statistics, and context encodings. Transport fitting uses shared training perturbations; an unsupported source–target map falls back to the identity.

Hardware and compute environment. All neural modules are trained and evaluated on a single NVIDIA A800 GPU, with CPU execution utilized for PCA, kernel/ridge regression, and ExtraTrees fitting. Latency comparisons are conducted on the same A800 GPU, maintaining consistent batch sizes and cell counts across all inference policies.

Model architectures and fixed hyperparameters. The response and descriptor dimensions are capped at 64 and 16, respectively, subject to available training rank. The conditional revision MLP and innovation MLP each use two 256-unit ReLU hidden layers and a linear d-dimensional output layer. The retriever/scorer and critic use 96 and 160 ExtraTrees estimators, with minimum leaf sizes of 16 and 4, respectively. The realizer architecture employs a 128-dimensional cell representation, 256-unit conditioning layers, 512-unit encoder/decoder hidden layers (with descriptor conditioning enabled), and a 32-dimensional PCA space ($d_P = 32$) fitted on training data for its distributional loss. Hyperparameters governing revision dynamics and context bounds are fixed to $\beta = 0.25$, $\eta = 0.5$, $\gamma = 0.1$, $\kappa = 0.35$, and $K_{\max} = 4$.

Training specifications and sampling. The conditional revision MLP, residual MLP, and population realizer are optimized using AdamW with a learning rate of 2×10^{-4}, weight decay of 10^{-4}, and gradient norm clipping capped at 5. The conditional revision MLP and innovation MLP are each trained for 100 epochs with a batch size of 256 transitions. The realizer is trained for ten epochs using a batch size of four queries, with 32 independently sampled control and treated cells per query; checkpoint selection evaluates 64 cells per query population. Cell sampling is performed without replacement when sufficient cells are available, and with replacement otherwise.

Hyperparameter search and model selection. Only three parameters are grid-searched, while all others remain fixed: (1) f_0 kernel ridge regularization $\alpha_0 \in \{0.01, 0.1, 1, 10\}$, selected to maximize calibration cosine similarity; (2) context sharing ratio $\rho \in \{0.25, 0.75\}$ (fixed to $\rho = 1$ for single-context scenarios); and (3) stopping threshold $\lambda \in \{0, 0.001, 0.002, 0.0025, 0.005, 0.01, 0.02, 0.04, 0.08, 1\}$, selected using grouped out-of-fold calibration predictions. Model selection follows the procedure described above; test set responses are strictly excluded from all training, search, and validation steps.

Training procedure.

1. Fit training-only representations; select f_0 settings and construct support-excluded episodes.
2. Fit the initial scorer, collect interpolation trajectories, refit the scorer, and fit $\mathcal{G}_\theta$ followed by $\mathcal{H}_\nu$.
3. Freeze these modules, generate calibration trajectories, fit the critic, and select its stopping threshold.
4. Generate terminal population conditions, freeze all upstream modules, and optimize $\mathcal{R}_\psi$ with $\mathcal{L}_R$.
5. Freeze the selected configuration and evaluate test queries.

B THEORETICAL ANALYSIS AND EMPIRICAL VALIDATION

We pair each theoretical result with diagnostics of its implementation. The experiments use frozen model weights and memory on the sciPlex3 unseen-perturbation setting used for the main ablations. Offline trajectories continue to $K_{\max}$ irrespective of the recorded stop decision, enabling evaluation of rejected and later candidates.

B.1 EVIDENCE ADAPTATION

B.1.1 PROOF OF PROPOSITION 1

Fix the query, fitted model, finite eligible memory, and a common distribution of reachable decision contexts, including hypothesis, depth, and history. Let $\mathcal{F}_k$ be the target-free information and $a_j = \overline{\Delta}_k(\mathbf{m}_j)$. Expectations integrate over these contexts and the unknown response; bounded cosine gains are integrable under a common convention for zero-norm inputs. For any $\mathcal{F}_k$-measurable selector J, the tower property gives

$$\mathbb{E}[\ell_{\cos}(\mathbf{z}_k) - \ell_{\cos}(\mathbf{z}_k + \eta \mathbf{d}_k(\mathbf{m}_J))] = \mathbb{E}[a_J]. \tag{26}$$

With deterministic tie breaking, $J^\star \in \arg\max_j a_j$ is measurable and attains $\mathbb{E}[\max_j a_j]$. The best fixed record attains $\max_j \mathbb{E}[a_j]$, and randomizing a fixed choice cannot improve it. Hence

$$\Gamma_{\mathrm{ev}} = \min_{j \in \mathcal{I}} \mathbb{E}[\max_{l \in \mathcal{I}} a_l - a_j] \geq 0. \tag{27}$$

Because the memory is finite and each integrand is nonnegative, the gap is zero exactly when some record is optimal almost surely.

For a quantitative sufficient condition, suppose distinct records j_A, j_B are uniquely optimal on positive-probability events A, B, with margins at least $\delta_A, \delta_B > 0$ over every other record. A fixed choice must lose on at least one event, yielding

$$\Gamma_{\mathrm{ev}} \geq \min\{\Pr(A)\delta_A, \Pr(B)\delta_B\} > 0. \tag{28}$$

Comparison scope. A fixed-state scorer can still vary its selection with depth and history; it is not the fixed-record comparator of Proposition 1. For any target-free selector J^0, the matched-context comparison is

$$\mathbb{E}[\Delta_k(\mathbf{m}_{J^\star}) - \Delta_k(\mathbf{m}_{J^0})] = \mathbb{E}[\max_j a_j - a_{J^0}] \geq 0, \tag{29}$$

where $\Delta_k(\mathbf{e})$ is the realized decoded cosine gain from revising with $\mathbf{e}$. Strictness requires J^0 to be suboptimal with positive probability. These are one-step comparisons at matched states, not guarantees for divergent trajectories or learned surrogate rankings. The following diagnostic tests the latter empirically.

B.1.2 RETRIEVAL SCORES AND ACTUAL REVISION GAINS

At identical states, we apply the actual revision operator to alternative evidence records and measure decoded cosine gains. Uniform selection averages eligible records; the fixed-record control retains the initially retrieved evidence. The hindsight gap compares the best realized candidate gain with the selected gain at the same state.

Current-state selection improves actual revision gains over uniform and fixed-record selection (Table 6). The positive score–gain correlation supports surrogate alignment, while the remaining hindsight gap indicates imperfect selection. Hindsight uses test responses and is not deployable; it differs from the conditional expected-gain oracle in Proposition 1.

B.2 ADAPTIVE STOPPING

B.2.1 PROOF OF PROPOSITION 2

Freeze retrieval and revision, and generate the potential trajectory by accepting every candidate up to $K_{\max}$. Let $\mathcal{F}_k$ be the increasing target-free information at decision k, including the selected

Table 6: Alignment between evidence selection and actual revision gains. Entries are means [95% CIs], in units of 10^{-3} cosine gain. The hindsight reference uses test responses and is not a deployable selector.

Evidence selector	Realized gain ↑	Hindsight gap ↓
Uniform selection	-1.8 [-3.6, 0.1]	14.4 [11.8, 17.0]
Fixed initial record	1.1 [-0.5, 2.8]	11.5 [9.1, 14.0]
Learned current-state selection	4.3 [2.0, 6.6]	8.3 [6.4, 10.4]
Hindsight best record	12.6 [9.8, 15.3]	0

Within-state score–gain Spearman correlation: 0.43 [0.31, 0.54].

evidence and candidate. For each query, write $U(t) = \mathcal{C}(D(\mathbf{z}_t), \mathbf{y}) - \lambda(t-1)$, as in the main text. Expectations in this proof range over queries and any model randomness; suppressing the query index does not fix the query across these expectations. For $H = K_{\max} - 1$, define

$$\Delta_k = \mathcal{C}(D(\mathbf{z}_{k+1}), \mathbf{y}) - \mathcal{C}(D(\mathbf{z}_k), \mathbf{y}), \qquad b_k = \mathbb{E}[\Delta_k \mid \mathcal{F}_k] - \lambda. \tag{30}$$

For any adapted stopping time $\tau \le K_{\max}$, telescoping and $\{\tau > k\} \in \mathcal{F}_k$ imply

$$\begin{aligned}
\mathbb{E}[U(\tau)] &= \mathbb{E}[U(1)] + \sum_{k=1}^{H} \mathbb{E}[\mathbf{1}_{\{\tau>k\}}(\Delta_k - \lambda)] \\
&= \mathbb{E}[U(1)] + \sum_{k=1}^{H} \mathbb{E}[\mathbf{1}_{\{\tau>k\}} b_k] \\
&\le \mathbb{E}[U(1)] + \sum_{k=1}^{H} \mathbb{E}[(b_k)_+].
\end{aligned} \tag{31}$$

Assume the stopping region is persistent:

$$\{b_k \le 0\} \subseteq \{b_{k+1} \le 0\} \quad \text{almost surely}, \qquad 1 \le k < H. \tag{32}$$

Positive net gains then form a prefix. The first-nonpositive rule $\tau^\star$ satisfies $\mathbf{1}_{\{\tau^\star>k\}} b_k = (b_k)_+$ almost surely and attains

$$\sup_\tau \mathbb{E}[U(\tau)] = \mathbb{E}[U(\tau^\star)] = \mathbb{E}[U(1)] + \sum_{k=1}^{H} \mathbb{E}[(b_k)_+]. \tag{33}$$

For fixed terminal depth K, the expected gain is $\sum_{k<K} \mathbb{E}[b_k]$. Therefore

$$\begin{aligned}
&\mathbb{E}[U(\tau^\star)] - \max_{1\le K\le K_{\max}} \mathbb{E}[U(K)] \\
&= \sum_{k=1}^{H} \mathbb{E}[(b_k)_+] - \max_{0\le L\le H} \sum_{k=1}^{L} \mathbb{E}[b_k] \\
&\ge \sum_{k=1}^{H} (\mathbb{E}[(b_k)_+] - (\mathbb{E}[b_k])_+) \\
&= \sum_{k=1}^{H} \min\{\mathbb{E}[(b_k)_+], \mathbb{E}[(-b_k)_+]\}.
\end{aligned} \tag{34}$$

The inequality bounds any prefix sum by the sum of positive expected increments; the last identity uses $b_k = (b_k)_+ - (-b_k)_+$. If both signs occur with positive probability at some depth, the gap is strictly positive.

Assumptions and interpretation. Persistence requires only that nonpositive conditional gains remain nonpositive; the stronger condition $b_{k+1} \le b_k$ is sufficient and makes the gap bound an equality. Persistence is not enforced by the architecture. Without it, net increments $(-0.02, 0.12)$ give a simple counterexample: stopping immediately misses a total gain of 0.10. The utility penalizes accepted revisions, not total runtime or rejected candidates; at $\lambda = 0$ it reduces to cosine. The next diagnostics assess the learned critic and practical trade-offs, rather than proving the persistence assumption.

Table 7: Critic and stopping diagnostics. Intervals are 95% perturbation-cluster CIs. Missed-utility statistics are conditional on stopping before $K_{\max}$.

Diagnostic	Value [95% CI]
Gain RMSE	0.0076 [0.0061, 0.0092]
Calibration gap	0.0014 [0.0008, 0.0022]
Calibration gap near λ	0.0019 [0.0010, 0.0030]
Rejected candidates with $\Delta_k > \lambda$ (%)	15.0 [7.0, 25.0]
Accepted candidates with $\Delta_k \leq \lambda$ (%)	7.5 [3.5, 13.0]
Early-stopped queries with positive missed utility (%)	20.0 [10.0, 33.0]
Mean missed utility among early-stopped queries	0.0012 [0.0005, 0.0021]

B.2.2 Critic calibration and stopping diagnostics

Calibration uses bins fixed on validation data. The calibration gap is the weighted absolute difference between predicted and realized bin means, evaluated globally and near the threshold: $|\widehat{\Delta}_k - \lambda| \leq 0.005$, with $\lambda = 0.002$. For each query, missed utility is $\max_{T \leq t \leq K_{\max}} \{U(t) - U(T)\}$ along the continued trajectory, where T is the deployed stopping depth. Including $t = T$ makes this quantity zero when no later state improves utility. Statistics for missed utility are conditional on stopping before $K_{\max}$.

Table 7 shows small calibration error with occasional incorrect decisions. Positive missed utility occurs for some early-stopped queries, but its small mean limits the observed cost of premature stopping.

B.2.3 Fixed-depth comparison and computational cost

All policies share frozen modules with $K_{\max} = 4$ and $\lambda = 0.002$. Fixed-depth policies accept every candidate and bypass the critic; $K = 3$ is selected on validation data. Latency includes all three stages under identical hardware, batching, and cell counts. The reported mean utility averages $U(T)$ across evaluated queries; for a fixed-depth policy, $T = K$ for every query.

Table 8: Depth and cost comparison. Cos-FC and mean depth at fixed $K = 1$, $K = 2$, $K = 4$, and adaptive stopping match the main-text ablations. Utility is computed with $\lambda = 0.002$. Relative latency includes all three stages and is normalized to fixed $K = 4$.

Policy	Cos-FC ↑	Mean $U(T)$ ↑	Mean T	Candidates	Rel. latency
Fixed $K = 1$	0.6218	0.6218	1.00	0.00	0.40
Fixed $K = 2$	0.6746	0.6726	2.00	1.00	0.60
Fixed $K = 3$ (validation-selected)	0.6867	0.6827	3.00	2.00	0.80
Fixed $K = 4$	0.6807	0.6747	4.00	3.00	1.00
Adaptive	**0.6986**	**0.6942**	3.20	2.70	0.97

The observed proportions at $T = 1, 2, 3, 4$ are $10\%, 10\%, 30\%, 50\%$. Early stopping evaluates and rejects one last candidate, whereas reaching $T = 4$ requires no further candidate. Thus the respective candidate counts are $1, 2, 3, 3$, yielding mean $T = 3.20$ and 2.70 candidate evaluations.

Adaptive stopping exceeds the validation-selected fixed depth by 0.0119 in Cos-FC and 0.0115 in utility (Table 8). Relative to $K = 4$, it reduces retained depth by 20%, candidate evaluations by 10%, and latency by 3%; these distinguish trajectory length from actual computational savings.

C Experimental Protocols

C.1 Datasets, splits, and available information

We use the released partitions of PerturBench and the parallel benchmark (Wu et al., 2026; Wei et al., 2026). The settings are unseen perturbations, unseen combinations, unseen cellular contexts,

and unseen perturbation–context pairs. Table 9 summarizes all seven grids. Split seeds identify data partitions, not repeated training runs.

Table 9: Evaluation grid. Cell counts are train/validation/test counts from released manifests. Pooled counts enumerate held-out unit–split evaluations; parentheses give the number of distinct held-out entities (drugs, genes, combinations, single-drug conditions, or cell lines), which is the resampling unit for the intervals described in Appendix C.2, or the split structure for single-split grids.

Setting	Cells	Held-out units per split	Pooled
Unseen perturbations (drug)	130,692 / 14,539 / 4,782 (seed 1); 108,538 / 11,991 / 29,484 (seed 2); 107,273 / 11,852 / 30,888 (seed 3)	9, 37, 38 drugs	84 (75 distinct drugs)
Unseen perturbations (gene)	188,588 / 21,145 / 53,512 (seed 1); 190,500 / 21,060 / 51,685 (seed 2); 193,605 / 19,390 / 50,250 (seed 3)	324 genes	972 (786 distinct genes)
Unseen combinations (drug)	8,304 / 8,055 / 47,019 (seed 4); 14,110 / 6,888 / 42,380 (seed 5); 18,772 / 8,563 / 36,043 (seed 6)	22, 20, 17 conditions (20, 18, 14 combinations)	59 (24 distinct combinations)
Unseen combinations (gene)	62,913 / 13,055 / 15,200	46 two-gene combinations	46 (one split)
Unseen cellular contexts	20,925 / – / 5,121 (A549); 17,736 / – / 8,310 (K562); 13,431 / – / 12,615 (MCF7)	one line per fold; 9 drugs per fold	27 (3 folds)
Unseen perturbation–context pairs (drug)	120,222 / 27,044 / 30,947	86 drug–line pairs	86 (one split)
Unseen perturbation–context pairs (gene)	171,478 / 23,113 / 23,740	87 gene–condition pairs	87 (one split)

Unseen perturbations. The sciPlex3-K562 drug benchmark has 5,839 genes (Srivatsan et al., 2020; Wei et al., 2026); released `train`, `test`, and `ood` labels map to training, validation, and test, respectively. Seeds 1–3 hold out 9, 37, and 38 drugs from training catalogues of 178, 150, and 149 drugs. Replogle-K562-essential has 5,822 genes (Replogle et al., 2022; Wei et al., 2026); the GEARS simulation protocol holds out 324 targets per split, with 1,163 training perturbations (Roohani et al., 2024). This essential-gene screen emphasizes strong-effect perturbations. Held-out identities are excluded from training, calibration, and memory for each split.

Unseen combinations. The sciPlex3-comb A549 release has 5,000 genes, 25 two-drug combinations, six single-drug conditions, and controls (Srivatsan et al., 2020; Wei et al., 2026). Seeds 4–6 contain 6, 8, and 11 training conditions. The complete released test partitions contribute 59 evaluations: 52 combinations and 7 single-drug conditions. The per-split combination/single-drug counts are 20/2, 18/2, and 14/3. Among held-out combinations, neither/exactly one component is profiled alone in training for 15/5, 11/7, and 3/11 cases; none has both. Single-agent measurements at 10 nM are not dose-matched references for combinations at 1000 nM, and held-out single-agent profiles are never added to training. Norman19 contains 91,168 K562 CRISPRa cells, 5,575 genes, and 237 conditions: control, 105 single-gene perturbations, and 131 combinations (Norman et al., 2019). The seed-42 PerturBench split assigns 145/45/46 perturbations to training/validation/test. Both component singletons of each test combination occur in training.

Unseen cellular contexts. The multi-dose sciPlex3 release covers nine drugs in A549, K562, and MCF7 (Wei et al., 2026). Each fold trains on two lines and holds out the third, producing 27 drug–line evaluations. The held-out line is absent from training, while its untreated controls are supplied in the test partition (2,041 A549, 3,528 K562, and 5,351 MCF7 cells). These controls, matched on `condition1`, provide the query reference without exposing treated responses. This multi-dose release differs from the highest-dose transfer slice used for unseen perturbation–context pairs.

Unseen perturbation–context pairs. The sciPlex3/Srivatsan20 highest-dose slice contains 178,213 cells, 8,630 genes, and 189 conditions including control across three lines (Srivatsan et al., 2020). Each evaluated drug remains observed in another training context, but the target treated pair is withheld. Frangieh21 has 9,140 genes in unperturbed culture, IFNγ, and co-culture contexts (Frangieh et al., 2021). The evaluated units are 87 held-out gene–condition pairs in co-culture; their perturbations remain observed in other training contexts. Unperturbed culture names a cellular context, not the absence of genetic perturbation.

Controls and preprocessing. All methods use the same context-matched untreated reference, fixed before cell subsampling. Unseen perturbation–context pairs and unseen combinations (gene) use controls from the evaluation split; unseen perturbations and unseen combinations (drug) use training controls of the single evaluation context; the unseen cellular contexts setting uses the held-out line's test-side controls as above. Gene order, expression layer, normalization, and pseudocount $\delta = 0.1$ are fixed across methods and recorded with prediction exports.

C.2 Metrics and statistical aggregation

Aggregate accuracy and specificity. Cos-FC is cosine similarity between predicted and observed aggregate logFC; RMSE measures error between mean-expression profiles. R-Cos and R-RMSE rank the matching response against other queries within the same context/dose block, using respectively the cross-query cosine and mean-expression error matrices. Lower ranks indicate greater specificity; chance is approximately 0.5, depending on the finite-sample rank convention. Comparisons share the evaluator's tie, normalization, and zero-norm rules. Population-derived logFC is distinguished from the pre-floor readout as described in Appendix A.2.

Population discrepancies and DEG recovery. For cell sets $X = \{\mathbf{x}_i\}_{i=1}^n$ and $Y = \{\mathbf{y}_j\}_{j=1}^m$, the benchmark's MMD-G and MMD-P columns use the empirical energy discrepancy

$$\mathcal{E}(X,Y) = \frac{2}{nm}\sum_{i,j}\|\mathbf{x}_i - \mathbf{y}_j\|_2 - \frac{1}{n^2}\sum_{i,i'}\|\mathbf{x}_i - \mathbf{x}_{i'}\|_2 - \frac{1}{m^2}\sum_{j,j'}\|\mathbf{y}_j - \mathbf{y}_{j'}\|_2, \tag{35}$$

in gene-expression and cell-PCA space, respectively. We retain the benchmark column names; these are not Gaussian-kernel MMD values. Evaluation PCA is fitted to observed evaluation cells with 100 components and is unavailable to the predictor. Predicted and reference populations are each capped at 300 cells per unit; controls are capped at 1000 cells per covariate, unlike the uncapped released evaluator. The formulation requires neither cell pairing nor equal sample sizes.

For each query, let $\widehat{\mathcal{D}}_{50}$ and $\mathcal{D}_{50}$ denote the top-50 DEG sets identified from its predicted and observed populations, respectively. DEG-R is $|\widehat{\mathcal{D}}_{50} \cap \mathcal{D}_{50}|/50$. Predicted and observed populations are tested against the same permitted controls using the released evaluator's gene universe, test statistic, and tie rules. Aggregate-only predictions are ineligible for population discrepancies and DEG recovery; broadcasting a mean does not constitute a heterogeneous population.

Pooling and uncertainty. For pooled Cos-FC, we combine the test predictions and corresponding observed responses across splits and recompute cosine similarity on the combined data. Split-level scores are computed separately within each split. Thus, the pooled score is computed from the underlying prediction and reference data, rather than by averaging the reported split-level scores. Repeated appearances of a drug or gene across splits are dependent; paired bootstrap resampling retains all evaluations of the same entity. Unseen perturbations (drug) and unseen perturbations (gene) cover 75 distinct drugs and 786 distinct genes. For unseen combinations (drug), resampling groups repeated combination identities or single-drug conditions, reflecting its mixed test set. For unseen cellular contexts, resampling groups evaluations by held-out line; its three folds are reported separately, and any interval is conditional on these lines rather than a precise estimate across possible contexts. Training, split, and cell-sampling seeds are distinguished. Paired comparisons use per-split intersections of eligible queries and the same preprocessing and sampling schedule. Distances on different expression spaces are not averaged across datasets.

Reporting conventions. Missing predictions and reasons are recorded per unit. GEARS evaluates 971/972 unseen perturbations (gene) units, 43/46 unseen combinations (gene) units, and 64/87

Table 10: Reported comparator coverage in the main and supplementary result tables. Representation and inference-access exceptions are described below.

Setting	Comparators
Unseen perturbations (drug)	chemCPA, State, CellFlow, PerturbNet, cycleCDR, Identity floor
Unseen perturbations (gene)	GEARS, scGPT, PerturbNet, GenePert, State, PT-RAG, CellFlow, scDFM, Identity floor
Unseen combinations (drug)	LatentAdditive, CPA, SAMS-VAE, Decoder-only, biolord, chemCPA, CellFlow, Identity floor
Unseen combinations (gene)	LatentAdditive, CPA, SAMS-VAE, Decoder-only, biolord, GEARS, scGPT, State, scDFM, CellFlow, GenePert, PerturbNet, Identity floor
Unseen cellular contexts	LatentAdditive, biolord, chemCPA, State, CellFlow, scGen, CellOT, cycleCDR, Identity floor
Unseen perturbation–context pairs (drug)	LatentAdditive, CPA, SAMS-VAE, Decoder-only, biolord, chemCPA, State, scDFM, CellFlow, scGen, PerturbNet, cycleCDR, CellOT, Identity floor
Unseen perturbation–context pairs (gene)	LatentAdditive, CPA, SAMS-VAE, Decoder-only, biolord, GEARS, State, PT-RAG, scDFM, CellFlow, scGen, GenePert, scGPT, CellOT, Identity floor

unseen perturbation–context pairs (gene) units; GenePert evaluates 971/972 unseen perturbations (gene) units; scGPT evaluates 83/87 unseen perturbation–context pairs (gene) units. GEARS and GenePert omit different unseen perturbations (gene), so paired comparisons use their intersection within each split. A dash marks a metric that is undefined for a predictor: aggregate-only predictors emit no cell population, and the identity floor's Cos-FC is undefined because its predicted logFC is identically zero.

C.3 Comparator coverage and protocol exceptions

Table 10 summarizes reported comparator coverage, including supplementary results. Models are evaluated on the shared gene space, controls, pseudocount, and splits. For pre-trained large models, we load their official pre-trained weights. The following exceptions distinguish representation limitations from differences in training access.

Perturbation and modality support. Catalogue-based PerturBench endpoint models encode perturbations as multi-hot vectors over the training atom vocabulary, so they are excluded from unseen perturbations, where a held-out identity contributes no trained atom; on unseen combinations (drug), 29 of the 52 held-out combinations have neither component in that vocabulary and 23 have one, leaving the LatentAdditive, SAMS-VAE, Decoder-only, and CPA rows there wholly or partly degenerate. GEARS, scGPT, GenePert, and PT-RAG apply only to genetic perturbations; PT-RAG's retrieval index additionally holds single-gene entries only, so it appears on the two genetic grids whose queries are single genes and not on unseen combinations (gene), and unseen cellular contexts, a chemical setting of nine compounds per held-out line, excludes it too. scDFM keys drugs by identity with no structure-derived path and is omitted from both drug-holdout settings; its genetic conditioning extends to unseen targets on all three splits.

Context support and degenerate encodings. GEARS and scGPT lack a cell-line conditioning channel; CPA, SAMS-VAE, and Decoder-only lack an embedding for an unseen line, excluding them from unseen cellular contexts. Biolord encodes perturbations and contexts over the training catalogue, so its unseen combinations (drug) and unseen cellular contexts rows reflect a missing-entry fallback rather than a learned representation of the held-out entity; State on unseen combinations (gene) likewise treats a held-out combination as an unseen one-hot label. chemCPA covers unseen drugs through structure-based encodings, but its predicted cells carry near-zero within-unit variance and its context channel carries no trained parameters for a held-out line; its population-level columns and its unseen cellular contexts row are therefore uninformative even where its aggregate

scores are competitive. These results should be interpreted with their representation limitations and the identity floor reported alongside them.

Untreated controls and missing targets. scGen and CellOT obtain target context through untreated cells rather than a context embedding. Their unseen cellular contexts use of test-side controls follows the same permitted-control rule as HEFR and does not expose treated responses. For unseen perturbation–context pairs (gene), GEARS removes held-out genes from its training loader across contexts, so its effective task is gene-level rather than pair-level holdout. Missing GO-graph entries or vocabulary targets produce the reduced unit counts reported above; they are not removed from other methods' evaluation sets.

D Additional Ablations and Biological Interpretation

D.1 Component controls

All structural variants use the main sciPlex3 ablation split and training and validation budgets. Affected modules are refitted, including critic recalibration after trajectory changes. In the main-text controls, *Mean* replaces the query-specific anchor by the training response mean; $T = 1$ bypasses revision; *Fixed-state* scores at $\mathbf{z}_1$ while updating $\mathbf{z}_k$, allowing depth/history dependence. *Interpol.* sets $\mathbf{g}_k = \mathbf{e}_k - \mathbf{z}_k$ and $\gamma = 0$; $\gamma = 0$ alone retains the learned grounded correction. *Control* replaces preliminary realized profiles by control cells while preserving the terminal mean. One revision stops at $\mathbf{z}_2$. Policy comparisons in Table 5 and Appendix B.2.3 use frozen modules.

D.1.1 Evidence access and innovation constraints

No evidence removes evidence vectors, source metadata, and history features while retaining query, state, depth, and revision capacity. *No orthogonal projection* retains the innovation norm bound; *No norm clipping* retains orthogonal projection.

Table 11: Additional revision ablations on sciPlex3. The Full-model baseline is aligned with the main-text component ablation.

Variant	Cos-FC ↑	MMD-G ↓
Full HEFR	**0.6986**	**0.1046**
No evidence	0.6705	0.1936
No orthogonal projection	0.6876	0.1394
No norm clipping	0.6824	0.1567

Table 11 separates the contribution of record-specific evidence from query-conditioned parameters and shows complementary benefits of the two innovation constraints. Together with the main interpolation and $\gamma = 0$ controls, it supports learned evidence-conditioned correction with bounded orthogonal innovation.

D.1.2 Terminal conditioning and mean consistency

We freeze the Propose and Reason and refit only the realizer. *No terminal-state input* removes $\mathbf{z}_T$ from $\mathcal{R}_\psi$ while retaining control cells, perturbation features, and projection to the same $\boldsymbol{\mu}$. *No mean projection* outputs the nonnegative preliminary profiles directly. Mean deviation is the relative ℓ_1 discrepancy between the predicted population mean and $\boldsymbol{\mu}$.

Table 12: Realization ablations with frozen terminal hypotheses on sciPlex3. Full-model performance aligns with the main component ablation.

Realization	Cos-FC ↑	MMD-G ↓	MMD-P ↓	Mean deviation ↓
Full HEFR	**0.6986**	**0.1046**	**0.1095**	$< 10^{-6}$
No terminal-state input	**0.6986**	0.1492	0.1635	$< 10^{-6}$
No mean projection	0.6864	0.1276	0.1358	0.0364

Without terminal input, projection preserves aggregate Cos-FC but population discrepancies worsen (Table 12), isolating a role for terminal conditioning beyond the prescribed mean. Removing projection introduces a 3.64% mean deviation and worsens both aggregate and population metrics. These controls complement the mean-preserving *Control* replacement in the main text.

D.2 Biological interpretation protocol

Cohort and evidence intervention. The sciPlex3 screen profiles 188 compounds in three cell lines at four doses after 24 h, with biological duplicates (Srivatsan et al., 2020). The diagnostic cohort fixes $10\,\mu$M and 24 h and samples 20 compounds uniformly without replacement, yielding 60 queries across A549, K562, and MCF7. Eligible compounds have external mechanism/program annotations, adequate treated and vehicle-control cells, and at least five eligible replacement records per query. They are excluded from training, calibration, and memory in every line. This dedicated cohort is separate from the K562 benchmark split.

At $\mathbf{z}_1$, selected evidence is compared with five memory records with non-overlapping broad mechanism annotations, matched on source line and dose and drawn from a distance- and norm-matched pool in the transported latent space. Calipers and the sampling seed are fixed before inspecting outcomes; insufficient matches are reported. All branches share the initial state, revision parameters, step size, and random inputs. Candidates are evaluated before the critic decision, and replacement outcomes are averaged within query.

Program and population readouts. For each query, let $E(\mathbf{z})$ denote its prespecified response-program error at hypothesis $\mathbf{z}$. Program error reduction is $[E(\mathbf{z}_1) - E(\widetilde{\mathbf{z}}_2)]/E(\mathbf{z}_1)$, with program definitions and scales fixed from external annotations and training/reference data. A prespecified near-zero threshold and absolute-gain sensitivity analysis address unstable denominators. Drug-cluster bootstrap resampling retains all three lines and both branches; replacement draws and biological replicates are not independent drug observations. For each of four prespecified families, the example in Fig. 3a is nearest its family's median paired difference; all 60 queries enter the cohort analysis.

A state map fitted on training/reference cells is frozen separately for each line and applied to predicted and observed populations. States need not denote the same cell types across lines. Control and treated cells are sampled independently, retaining replicate and plate information, with up to 256 QC-passing cells per population. Aggregate reference means are fixed before subsampling. Initial-state realizations are offline diagnostics; ordinary inference realizes only the terminal hypothesis.

E Supplementary Benchmark Results

Tables 13 and 14 report unseen perturbation–context pairs transfer results. Table 15 resolves multi-split grids; Table 16 resolves held-out cell lines. Metric definitions and eligible-unit exceptions are given in Appendix C.

Table 13: Unseen perturbation–context pairs (drug): 86 held-out drug–line pairs. Metrics follow Appendix C.2; aggregate-only outputs receive dashes for population metrics.

	Aggregate Metrics				Distributional Metrics		
Method	Cos-FC ↑	R-Cos ↓	RMSE ↓	R-RMSE ↓	MMD-G ↓	MMD-P ↓	DEG-R ↑
Identity floor	–	0.4626	0.0226	0.5639	0.2095	0.2483	0.0093
LatentAdditive (Wu et al., 2026)	0.4887	0.1351	**0.0173**	0.1480	4.1581	1.6561	0.0000
SAMS-VAE (B. & K., 2023)	0.4666	0.2502	0.0197	0.2476	2.5045	0.5698	0.0002
Decoder-only (Wu et al., 2026)	0.3534	0.1794	0.0181	0.1463	–	–	–
biolord (Piran et al., 2024)	0.3443	0.2360	0.0221	0.2136	0.8367	0.2434	0.0000
CPA (Lotfollahi et al., 2023)	0.3095	0.3636	0.0209	0.3463	4.2178	1.7716	0.0002
State (Adduri et al., 2026)	0.1295	0.4347	0.0560	0.4735	3.2738	1.3546	0.0344
scDFM (Yu et al., 2026)	0.0856	0.4064	0.0854	0.4480	1.5772	2.5764	0.0584
CellFlow (Klein et al., 2025)	0.1040	0.4029	0.0574	0.3947	1.1611	1.9830	0.0067
chemCPA (Hetzel et al., 2022)	0.1242	0.4887	0.0819	0.4535	6.1349	4.8738	0.0105
scGen (Lotfollahi et al., 2019)	0.2983	0.3421	0.0246	0.3156	1.8336	0.3913	0.0007
PerturbNet (Yu et al., 2025)	0.1356	0.4204	0.0381	0.4071	2.4181	0.8070	0.0000
cycleCDR (Huang & Liu, 2024)	0.1195	0.4186	0.1240	0.4932	8.1315	6.0167	**0.1433**
CellOT (Bunne et al., 2023)	0.1358	0.3916	0.0804	0.4010	5.6591	4.6169	0.0000
HEFR (ours)	**0.5149**	**0.1331**	0.0221	**0.1176**	**0.1610**	**0.1736**	0.1123

Table 14: Unseen perturbation–context pairs (gene): 87 held-out gene–condition pairs in co-culture. GEARS evaluates 64 units and scGPT 83; all other rows use the full query set. Metric definitions follow Appendix C.2.

	Aggregate Metrics				Distributional Metrics		
Method	Cos-FC ↑	R-Cos ↓	RMSE ↓	R-RMSE ↓	MMD-G ↓	MMD-P ↓	DEG-R ↑
Identity floor	–	0.4973	0.0264	0.4615	0.2917	0.1923	0.0103
LatentAdditive (Wu et al., 2026)	0.1626	0.3957	**0.0246**	0.3776	6.5306	3.1276	0.0000
CPA (Lotfollahi et al., 2023)	0.1249	0.2915	0.0257	0.2793	2.3777	0.3302	0.0044
Decoder-only (Wu et al., 2026)	0.0754	0.3756	0.0253	0.3742	–	–	–
SAMS-VAE (B. & K., 2023)	0.1143	0.4833	0.0273	0.4818	2.5431	0.4019	0.0000
biolord (Piran et al., 2024)	0.0590	0.3990	0.0310	0.4022	1.5546	0.2482	0.0000
GEARS (Roohani et al., 2024)	0.0282	0.4673	0.0369	0.4891	0.3842	0.2728	0.0059
State (Adduri et al., 2026)	0.1100	0.3455	0.0292	0.3887	0.4122	0.3892	0.0568
scDFM (Yu et al., 2026)	−0.0067	0.4158	0.1084	0.4691	2.1701	2.3345	0.0389
CellFlow (Klein et al., 2025)	−0.0255	0.3806	0.0464	0.4288	0.4601	0.4568	0.0103
scGen (Lotfollahi et al., 2019)	0.1545	**0.2021**	0.0266	0.2374	1.9657	0.2331	0.0000
PT-RAG (Di et al., 2026)	0.1129	0.3488	0.0317	0.4348	0.4480	0.3006	0.0701
GenePert (Chen & Zou, 2024)	0.1028	0.2082	0.0270	0.2463	–	–	–
CellOT (Bunne et al., 2023)	0.0407	0.2780	0.0885	0.4256	6.7319	3.8540	0.0000
scGPT (Cui et al., 2024)	0.0731	0.2086	0.0338	0.2440	5.4142	2.5811	0.0000
HEFR (ours)	**0.1930**	0.2032	0.0258	**0.1982**	**0.2692**	**0.1491**	**0.0826**

Table 15: Per-split and pooled Cos-FC. Pooled Cos-FC is recomputed from the combined test predictions and corresponding observed responses across the three splits, not averaged from the split-level scores. Seeds A/B/C are 1/2/3 for unseen perturbations and 4/5/6 for unseen combinations (drug). Split sizes are listed in Table 9; methods with ineligible units are scored on the subsets given in Appendix C.2. Repeated entities across splits are dependent.

Setting	Method	Seed A ↑	Seed B ↑	Seed C ↑	Pooled ↑
Unseen perturbations (drug)	State	0.5302	0.6062	0.6369	0.6119
	chemCPA	0.4732	0.6650	0.6916	0.6565
	CellFlow	0.0035	0.0909	0.1105	0.0904
	cycleCDR	0.0865	0.0199	0.0211	0.0276
	PerturbNet	0.1207	0.1833	0.1841	0.1770
	HEFR (ours)	**0.6141**	**0.6998**	**0.7184**	**0.6986**
Unseen perturbations (gene)	State	0.0929	0.1579	0.1399	0.1302
	CellFlow	0.1471	0.0527	0.0904	0.0967
	PerturbNet	0.0944	0.1054	0.0977	0.0992
	scGPT	0.2689	0.2764	0.2720	0.2724
	PT-RAG	0.0938	0.1599	0.1385	0.1311
	GEARS	0.1405	0.1554	0.1505	0.1488
	GenePert	0.3027	0.3035	0.3190	0.3084
	scDFM	0.1576	0.1622	0.1689	0.1631
	HEFR (ours)	**0.3121**	**0.3190**	**0.3268**	**0.3192**
Unseen combinations (drug)	chemCPA	−0.0202	0.2938	0.6688	0.2848
	CPA	0.2998	0.5900	0.5744	0.4773
	biolord	0.0851	0.5344	0.6442	0.3985
	CellFlow	0.0139	0.5589	0.7001	0.3964
	HEFR (ours)	**0.3215**	**0.6213**	**0.7208**	**0.5181**

Table 16: Unseen cellular contexts by held-out cell line, with 9 drugs per fold. Pooled results cover 27 evaluations over three distinct lines; folds are not repeated training seeds. Pooled values are reported in Table 3. Pooled metrics follow the aggregation procedure in Appendix C.2.

		Aggregate Metrics				**Distributional Metrics**		
Fold	Method	Cos-FC ↑	R-Cos ↓	RMSE ↓	R-RMSE ↓	MMD-G ↓	MMD-P ↓	DEG-R ↑
A549	State	0.3115	0.2639	**0.0686**	0.1111	0.5293	0.6576	**0.2356**
	chemCPA	0.1272	0.3750	0.3910	0.5000	19.6554	10.8937	0.0044
	biolord	0.1285	0.1250	0.1326	0.4444	2.4956	2.0508	0.0000
	CellFlow	0.0709	0.4167	0.2612	0.4167	4.3142	4.4487	0.0267
	CellOT	0.2888	0.1944	0.1866	0.3333	9.9532	7.7103	0.0089
	cycleCDR	0.0964	0.4167	0.2629	0.3194	11.4424	8.9055	0.1089
	scGen	0.2364	0.1944	0.0880	0.2083	4.9689	2.6176	0.0000
	HEFR (ours)	**0.3958**	**0.0694**	0.0754	**0.0903**	**0.4398**	**0.5321**	0.2184
K562	State	0.2418	0.2639	**0.0597**	0.1250	0.4705	0.4941	**0.1156**
	chemCPA	0.1383	0.4444	0.3035	0.5000	15.4478	14.3853	0.0000
	biolord	0.1923	0.1528	0.1082	0.2639	3.1160	1.9279	0.0000
	CellFlow	0.1286	0.4306	0.6550	0.5000	16.7009	12.4708	0.0511
	CellOT	0.2638	0.3194	0.1906	0.3611	10.5871	8.3234	0.0022
	cycleCDR	0.1380	0.2500	0.3063	0.4306	13.4553	8.6587	0.0133
	scGen	0.2183	0.2361	0.1172	0.3472	6.2215	3.2906	0.0022
	HEFR (ours)	**0.3417**	**0.0833**	0.0668	**0.1042**	**0.3947**	**0.4376**	0.1103
MCF7	State	0.3246	0.3889	**0.0583**	0.2639	0.5042	0.6643	0.2556
	chemCPA	0.1786	0.5000	0.2444	0.5000	12.2841	11.3346	0.0133
	biolord	0.1354	0.2639	0.1317	0.3472	2.5333	2.4798	0.0000
	CellFlow	0.1799	0.1667	0.1253	0.4861	1.5520	1.0287	0.0267
	CellOT	0.3137	0.3472	0.1608	0.3750	8.2471	6.7878	0.0044
	cycleCDR	0.1032	0.4861	0.1870	0.3333	8.2995	7.5174	**0.2578**
	scGen	0.2157	0.2500	0.0744	0.2917	3.9000	2.4466	0.0022
	HEFR (ours)	**0.4133**	**0.0918**	0.0684	**0.1478**	**0.4303**	**0.5327**	0.2296