CLAMP: Steady-State ODE Inference of Gene Regulatory Networks from Single-Cell Perturbations

Anonymous Authors[1]

Abstract

Recovering mechanistic gene regulatory networks from single-cell interventional data is a fundamental challenge of computational biology. While ordinary differential equations (ODEs) offer an interpretable framework, existing single-cell ODE methods primarily target temporal trajectories rather than the steady-state transcriptomic snapshots provided by certain high-throughput CRISPR Perturb-seq protocols. To bridge this gap, we introduce CLAMP, a differentiable framework that reframes sustained genetic perturbation as a coordinate constraint on a Hill-kinetics regulatory network. By clamping perturbed genes to their observed expression and computing implicit gradients through a fixed-point solver, CLAMP jointly fits a single shared regulatory adjacency across all single-gene conditions. This formulation naturally resolves k-gene perturbations as k simultaneous clamps, enabling compositional prediction without combinatorial training data. CLAMP is also encoder-agnostic, comparing predicted and observed states through any frozen and differentiable encoder. Evaluated on the Norman 2019 dataset, CLAMP accurately predicts non-additive combinatorial effects and structurally aligns with independent ChIP-seq and lineage signatures. Deploying CLAMP as a mechanistic probe further reveals that current single-cell foundation models struggle to preserve the geometric directions where regulatory signal lives, providing a structural explanation for their performance on perturbation prediction tasks.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Submitted to the 2026 Workshop on Generative and Agentic AI for Biology (ICML 2026). Do not distribute.

1. Introduction

Mechanistic understanding of biological processes is at the heart of our ability to explore and expand on current biological knowledge. Among these, gene regulation can naturally be represented as a network: a structured representation of which genes influence which, and through what dynamics. Recovering this structure directly from measurement data, rather than curating it from the literature, has been a long-standing goal of computational biology. CRISPR Perturb-seq has reshaped this problem by providing high-throughput interventional data at single-cell resolution (Dixit et al., 2016; Norman et al., 2019), where thousands of cells are assayed under hundreds of targeted genetic perturbations, each revealing how the system responds when a specific regulator is forced to a specific value. A regulatory model recovered from this data would yield interpretable structure and kinetic parameters and would predict untested perturbations through forward simulation of the fit dynamics. This is a stronger requirement than predictive accuracy on perturbed expression, which can be achieved without recovering any underlying mechanistic system.

1.1. Related Work

Four lines of prior work approach mechanistic recovery from single-cell data, each with limitations that motivate our framework.

Regression-based perturbation prediction. A large body of recent work models perturbation response as a learned mapping from a perturbation identifier p, control expression x_{ctrl}, and covariates c (e.g. cell type, modality, or dose) to the perturbed profile, $f_\theta : (p, x_{\mathrm{ctrl}}, c) \mapsto \hat{x}_p$. GEARS (Roohani et al., 2024) embeds this mapping in a graph neural network over a Gene Ontology graph. CPA (Lotfollahi et al., 2023) learns disentangled latent representations of perturbation, cell type, and covariate effects, predicting unseen combinations through latent-space arithmetic. AttentionPert (Bai et al., 2024) attends over perturbation embeddings and gene features, while fine-tuned single-cell foundation models (Cui et al., 2024; Theodoris et al., 2023) apply the same framing through pretrained representations. These methods produce predicted expression profiles, and their compositional

generalization relies on seeing doubles in training or on latent-space arithmetic over disentangled representations, rather than on simulation through a recovered structure.

Dynamical-system methods for single-cell data. A second family fits ordinary differential equations to single-cell expression, serving as our closest conceptual starting point. However, these methods are primarily designed for temporal trajectories rather than endpoint snapshots. Cell-MNN (von Bassewitz et al., 2026) fits a locally linearized ODE in PCA space, constraining it to linear latent dynamics. FLeCS (Bertin et al., 2025) retains nonlinear kinetics but requires pseudotime estimation to construct training trajectories and fits parameters per-condition rather than assembling a shared regulatory model. PerturbODE (Lin et al., 2025) uses a neural drift function where regulatory structure remains implicit in MLP weights rather than interpretable kinetic parameters. PHOENIX (Hossain et al., 2024) fits Hill kinetics but targets bulk temporal forecasting rather than interventional structural recovery. Finally, continuous Lyapunov models like Bicycle (Rohbeck et al., 2024) leverage steady-state data but rely on linear stochastic differential equations. Together, these methods establish the tractability of single-cell ODEs but highlight a critical gap: none jointly recovers a single, nonlinear regulatory model across all perturbation conditions directly from the snapshots provided by Perturb-seq (see Appendix A for more details).

Static-prior propagation methods. A third family treats the regulatory network as a static object assembled from auxiliary data and simulates perturbation effects by propagating linearized responses through the fixed network. CellOracle (Kamimoto et al., 2023) constructs the network from scATAC-seq accessibility and TF motif presence, then iteratively updates expression by linear propagation through the fixed graph. scTenifoldKnk (Osorio et al., 2022) constructs gene-gene networks from control expression via manifold alignment and computes virtual knockouts by comparing the original network to one with the perturbed gene's influence removed. While mechanism-aware, the structure is taken as input rather than learned from perturbations, and predictions reflect local linearizations rather than new dynamical steady states.

Foundation-model evaluation for perturbation prediction. Single-cell foundation models have been evaluated systematically through regression-based benchmarks (Wenteler et al., 2025; Ahlmann-Eltze et al., 2025), which establish that foundation models trained on observational expression data underperform additive and mean baselines on perturbation prediction tasks. These results report what foundation models fail at, but do not explain why: regression accuracy alone cannot distinguish failures of a flexible readout from failures of the underlying representation. What is missing is an evaluation regime that tests representations against the structural requirements of mechanistic recovery directly.

1.2. Contributions

We introduce CLAMP, a framework that shares a single regulatory object across all perturbation conditions through joint optimization, accommodates arbitrary nonlinear kinetic forms without restriction to linear latent dynamics, and operates natively on the snapshots that Perturb-seq actually provides. The framework is encoder-agnostic by construction: because the loss is computed on steady states rather than along trajectories, any differentiable encoder can serve as the lens through which predicted and observed states are compared. This recovers the PCA setting as a special case and admits scGPT, scPRINT, and other frozen foundation models without modification to the fitting procedure. Our contributions are threefold:

1. **A reframing of perturbation prediction as a learned ODE evaluated at a steady state solution.** We clamp the perturbed gene's steady-state value at its observed perturbation level and solve for the induced state of the remaining network.
2. **A differentiable steady-state framework for joint multi-condition ODE recovery.** We fit a single shared regulatory adjacency across all single-gene perturbation conditions by matching predicted to observed per-condition steady states. Gradients through the fixed-point solver are computed by implicit differentiation of the steady-state equations rather than by unrolling iterations through autograd, decoupling backward pass cost from solver depth and making joint fitting tractable at Perturb-seq gene-panel scales.
3. **An encoder-agnostic interface, which enables probing of mechanistic recovery.** Applying any differentiable encoder as a projection of both predicted and observed steady states enables substitution of Identity, PCA, and frozen foundation-model representations without changing the fitting procedure. Recovery quality corresponding to a given encoder serves as a mechanistic probe of whether the representation preserves causal regulatory structure, rather than of what a flexible readout can extract from it.

2. Methods

In the following section we outline our CLAMP framework. Section 2.1 formulates a CRISPR perturbation as a steady state constraint on the regulatory ODE. Section 2.2 states the joint multi-condition recovery objective, with a frozen encoder serving as a differentiable metric on observations. Section 2.3 develops the implicit-differentiation backward pass through the steady-state solver. Section 2.4 specifies

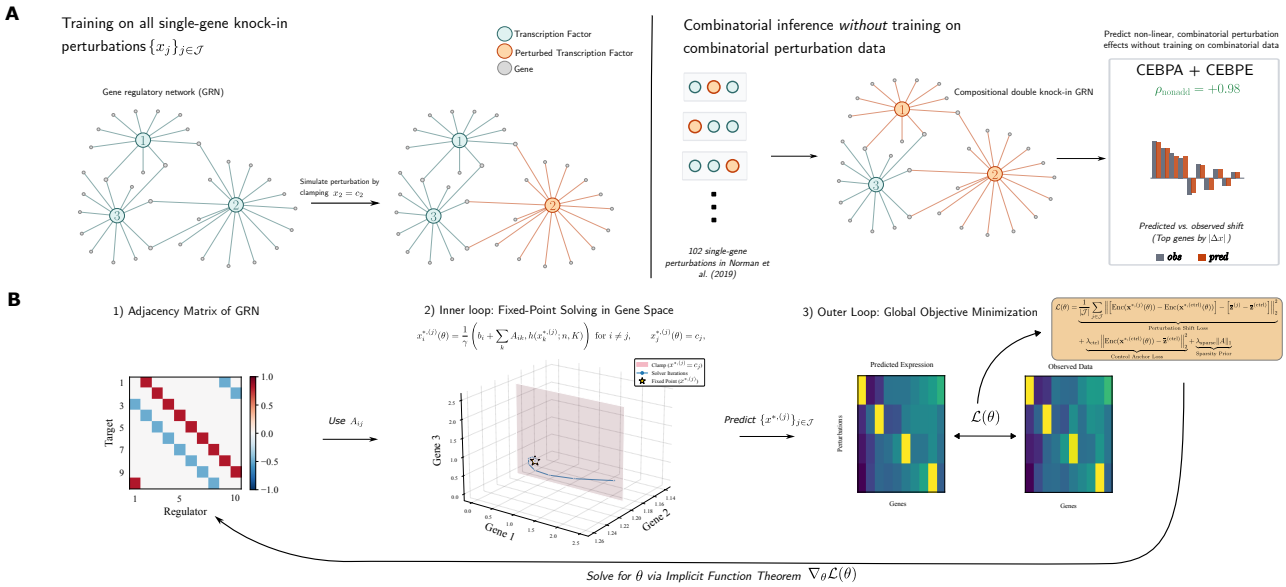

Figure 1. **Overview of the CLAMP pipeline. (A) Conceptual schema.** The framework is fit on single-gene CRISPR perturbations $\{\mathbf{x}_j\}_{j\in\mathcal{J}}$, with each perturbation simulated as a coordinate clamp $x_j = c_j$ on a shared regulatory network (left). The fitted model predicts combinatorial responses to held-out double-gene perturbations by imposing two simultaneous clamps, illustrated here for the redundant pair CEBPA+CEBPE ($\rho_{\text{nonadd}} = +0.98$, with the top genes by $|\Delta x|$ shown as a waterfall of observed and predicted shifts, right). **(B) Optimization loop.** The recovered adjacency A (1) parameterizes a Hill-kinetics ODE whose constrained steady states are obtained by damped fixed-point iteration in gene space (2), with the clamp plane $x_j^{*,(j)} = c_j$ reimposed at every step. Per-condition predicted steady states are then compared to the observed data through the encoder lens via the joint loss $\mathcal{L}(\theta)$ of Eq. 4, comprising perturbation-shift, control-anchor, and sparsity terms (3). Parameter gradients are obtained by implicit differentiation of the fixed-point equation, closing the loop without unrolling solver iterations through autograd. We provide a full walkthrough of the pipeline in Appendix B.

the encoders and adjacency parameterizations we evaluate. Section 2.5 defines our evaluation protocol. An overview schema of the full pipeline is provided in Figure 1.

2.1. Steady-state characterization of perturbation response

We model gene regulation as a deterministic ODE in gene space. For G genes with expression vector $\mathbf{x} \in \mathbb{R}^G$, the dynamics take the form

$$\frac{dx_i}{dt} = b_i + \sum_{j=1}^{G} A_{ij}\, h(x_j;\, n,\, K) - \gamma\, x_i, \tag{1}$$

where A_{ij} encodes the signed regulatory effect of gene j on gene i, b_i is the basal transcription rate, γ is a shared strictly positive mRNA decay rate, and $h(x;\, n,\, K) = x^n/(K^n + x^n)$ is the Hill function with cooperativity n and half-maximal constant K (Appendix D gives the biophysical derivation and discusses shared kinetic parameters). We collect the parameters as $\theta = (A,\, \mathbf{b},\, \gamma,\, n,\, K)$.

Steady-state recovery under coordinate constraints Two features of CRISPR Perturb-seq experiments define the use of Eq. 1 for network recovery. First, CRISPR perturbations represent a sustained force rather than a transient pulse, driving target genes toward stationary expression levels. Second, the equilibration of a regulatory network is rate-limited by the turnover of its functional proteins, which in human cell lines exhibit median half-lives of roughly 38-46 hours compared to just 9 hours for mRNA transcripts (Zecha et al., 2018; Schwanhäusser et al., 2011). In benchmark datasets such as the K562 cell profiles in (Norman et al., 2019), cells are cultured for 5-6 days post-infection before single-cell sequencing, allowing sufficient time for the slow-moving protein cascades to propagate and for the faster mRNA variables to equilibrate into a stable cellular attractor.

We therefore model the observed perturbed state not as a snapshot along a trajectory, but as a constrained steady state, a fixed point of the dynamics where the perturbed coordinate is clamped at its observed value c_j while the remaining genes settle into a self-consistent equilibrium. This assumption transforms the dynamical recovery into an algebraic root-finding problem defined by the network's topology. Given the observed clamped value c_j, the steady state $x^{*,(j)}$ satisfies:

$$x_i^{*,(j)}(\theta) = \frac{1}{\gamma}\left(b_i + \sum_k A_{ik}\, h(x_k^{*,(j)}; n, K)\right) \text{ for } i \neq j, \qquad x_j^{*,(j)}(\theta) = c_j. \tag{2}$$

where the first expression enforces the equilibrium condition $dx_i/dt = 0$ for all unclamped genes, and the second

represents the coordinate clamp. We resolve this system numerically using damped fixed-point iteration, reimposing the clamp c_j at each step to ensure the final state is consistent with the intervention. This approach generalizes naturally to multi-gene perturbations by applying multiple simultaneous coordinate constraints, which we exploit for compositional prediction in Section 2.5.

2.2. Joint multi-condition recovery via a differentiable encoder

Perturb-seq measures single-cell gene expression under a battery of single-gene CRISPR perturbations. Let $\mathcal{J}$ index the perturbed genes, with $\mathcal{X}^{(j)} \subset \mathbb{R}^G$ denoting cells receiving perturbation j and $\mathcal{X}^{(\text{ctrl})}$ the unperturbed controls. We assume one regulatory model governs the cell's response to every intervention, and therefore fit a single θ jointly across all conditions, rather than per-condition as in prior nonlinear-ODE work (Bertin et al., 2025). CLAMP operates through a frozen encoder $\text{Enc} : \mathbb{R}^G \to \mathbb{R}^d$ that serves as a metric on observations: predicted and observed states are compared in its representation space rather than directly in gene space. The encoder is frozen but must be differentiable in its input, since $\mathbf{x}^{*,(j)}(\theta)$ depends on θ, requiring backpropagation through the encoder. Critically, θ is recovered in gene space, keeping A interpretable across encoder choices: Identity ($\text{Enc}(\mathbf{x}) = \mathbf{x}$) recovers the unencoded baseline, and other differentiable encoders substitute without changing the fitting procedure.

Shift-based steady-state loss. We fit θ by matching predicted steady states to observations through three loss components. First, we match the perturbation-induced *shift* between encoded representations rather than the encoded states themselves,

$$\mathcal{L}_{\text{pert}}(\theta) = \frac{1}{|\mathcal{J}|} \sum_{j \in \mathcal{J}} \Big\| \big[\text{Enc}(\mathbf{x}^{*,(j)}(\theta)) - \text{Enc}(\mathbf{x}^{*,(\text{ctrl})}(\theta))\big] - \big[\bar{\mathbf{z}}^{(j)} - \bar{\mathbf{z}}^{(\text{ctrl})}\big] \Big\|_2^2. \tag{3}$$

where $\bar{\mathbf{z}}^{(j)}$ is the mean encoding of cells under perturbation j and $\bar{\mathbf{z}}^{(\text{ctrl})}$ is the control mean. The shift formulation removes the baseline component of $\text{Enc}(\mathbf{x}^{*,(\text{ctrl})})$ that is shared across conditions and would otherwise dilute the regulatory signal. Second, we anchor the predicted wild-type steady state to the observed control mean via $\mathcal{L}_{\text{ctrl}}(\theta) = \|\text{Enc}(\mathbf{x}^{*,(\text{ctrl})}(\theta)) - \bar{\mathbf{z}}^{(\text{ctrl})}\|_2^2$, fixing a reference point that the shift formulation alone leaves undetermined. Third, we add a weak L_1 sparsity penalty $\|A\|_1$ encoding the prior that biological regulatory networks are sparse. The full objective is

$$\mathcal{L}(\theta) = \mathcal{L}_{\text{pert}}(\theta) + \lambda_{\text{ctrl}}\, \mathcal{L}_{\text{ctrl}}(\theta) + \lambda_{\text{sparse}}\, \|A\|_1, \tag{4}$$

with hyperparameter values reported in Appendix E.

2.3. Implicit differentiation through the fixed-point solver

Backpropagating through the unrolled fixed-point iteration is impractical at the scale of this work: the activation trace must be retained for every perturbation condition until the backward pass, with both memory and time scaling linearly in iteration count. Following the approach established in Bai et al. (2019) and Blondel et al. (2022), we instead compute gradients analytically via the implicit function theorem. The forward iteration is run without an autograd graph and returns only the converged $\mathbf{x}^*$. Writing $g(\mathbf{x}, \theta) = (\mathbf{b} + A\, h(\mathbf{x}; n, K))/\gamma$, the converged state satisfies $F(\mathbf{x}^*, \theta) = \mathbf{x}^* - g(\mathbf{x}^*, \theta) = \mathbf{0}$. For a scalar loss $\mathcal{L}$ with upstream gradient $\mathbf{v} = \partial\mathcal{L}/\partial\mathbf{x}^*$, the parameter gradient is

$$\frac{\partial \mathcal{L}}{\partial \theta} = -\mathbf{w}^\top \frac{\partial F}{\partial \theta}, \quad J^\top \mathbf{w} = \mathbf{v}, \quad \text{with} \quad J = I - \tfrac{1}{\gamma} A\, \text{diag}(h'(\mathbf{x}^*)). \tag{5}$$

We solve the linear system $J^\top \mathbf{w} = \mathbf{v}$ once per condition and then evaluate the analytic vector-Jacobian product (VJP), which reads: $-\mathbf{w}^\top \partial F/\partial\theta$ for each parameter group. The Hill derivatives have closed forms, so all parameter gradients are evaluated exactly at the fixed point. This linear system is well-posed provided J is non-singular, a condition typically satisfied when the forward iteration converges to a stable attractor. Full closed-form expressions for $\partial\mathcal{L}/\partial A$, $\partial\mathcal{L}/\partial\mathbf{b}$, $\partial\mathcal{L}/\partial\gamma$, $\partial\mathcal{L}/\partial n$, and $\partial\mathcal{L}/\partial K$ are reported alongside a discussion of numerical stability in Appendix F. To reflect that the expression of perturbed genes is fixed by the CRISPR intervention rather than the learned dynamics, we enforce coordinate constraints on the gradient side by zeroing the entry of $\mathbf{v}$ for clamped genes and replacing the corresponding row of J with $\mathbf{e}_j^\top$. This zeroes the gradient through clamped coordinates, since their values are set by the experimental perturbation rather than predicted by θ, so backpropagation should not contribute to parameter updates from them. Backward cost then reduces to a single $G \times G$ solve per condition, independent of forward solver depth, and we wrap the entire fixed-point solve in a `torch.autograd.Function` so that it composes seamlessly with gradients through the encoder lens.

2.4. Encoders and adjacency parameterization

The framework requires only that $\text{Enc} : \mathbb{R}^G \to \mathbb{R}^d$ be differentiable in its input. We evaluate four encoders. Identity is the unencoded baseline. PCA, fitted on control cells, gives a linear compression baseline. scGPT (Cui et al., 2024) and scPRINT (Kalfon et al., 2025) are pretrained transformer-based single-cell foundation models, used as frozen wrappers that map an expression vector to a cell-level embedding. Per-encoder vocabulary mapping, input format, and depth-token details are reported in Appendix G.

We restrict the recovered adjacency $A \in \mathbb{R}^{G \times G_{\text{TF}}}$ to fit edges only for TFs with interventional evidence, where

rows index the G measured genes and columns index the $G_{\mathrm{TF}} = |\mathcal{J}| = 102$ perturbed TFs, so A_{ij} encodes the effect of perturbed TF j on gene i. This reduces the parameter count from $G^2 \approx 6.5 \times 10^6$ to $G \cdot G_{\mathrm{TF}} \approx 2.6 \times 10^5$, giving G scalar measurements per TF column and a roughly one-to-one parameter-to-observation regime. We use a dense parameterization with the TF mask reapplied after each optimizer step to prevent AdamW second-moment statistics from drifting masked entries off zero.

2.5. Evaluation: held-out prediction, structural validation, and encoder transmission

We evaluate along three axes: whether the recovered ODE generalizes to held-out perturbations and captures interactions beyond single-gene superposition; whether the recovered adjacency A aligns with independent biological references; and whether, independently of any fitted ODE, each encoder transmits the geometry of regulatory signal prior to fitting.

Held-out prediction. We evaluate on Norman 2019 (Norman et al., 2019), to our knowledge the only published Perturb-seq dataset with double-gene perturbations at the scale required for per-pair compositional evaluation. The framework is fit on single-gene perturbations alone, with a subset held out for single-gene validation (SKI val) and all double-gene perturbations held out for compositional evaluation (DKI val). A k-gene perturbation imposes k simultaneous coordinate constraints on the same Hill ODE, with unclamped genes determined by Eq. 2. The same learned θ resolves unseen combinatorial perturbations through the same solver, requiring no additional training data or architectural extension.

For each held-out pair (j, k) we compute the Pearson correlation between predicted and observed shifts relative to control,

$$\rho_{j,k} = \mathrm{Pearson}\Big(\mathbf{x}^{*,(j,k)}(\theta) - \mathbf{x}^{*,(\mathrm{ctrl})}(\theta), \; \bar{\mathbf{x}}^{(j,k)} - \bar{\mathbf{x}}^{(\mathrm{ctrl})}\Big). \tag{6}$$

with the perturbed coordinates excluded, and the same correlation restricted to the non-additive residual $\rho_{j,k}^{\mathrm{nonadd}}$, obtained by replacing each shift in Eq. 6 with its difference from the additive prediction $\hat{\mathbf{s}}_{\mathrm{add}}^{(j,k)} = \bar{\mathbf{x}}^{(j)} + \bar{\mathbf{x}}^{(k)} - 2\bar{\mathbf{x}}^{(\mathrm{ctrl})}$. The two metrics probe different aspects of recovery. Aggregate $\rho_{j,k}$ measures overall fit to the predicted shift, the bulk of which on Perturb-seq is additive: $\hat{\mathbf{s}}_{\mathrm{add}}$ alone outperforms fine-tuned single-cell foundation models on this metric (Ahlmann-Eltze et al., 2025). The non-additive correlation $\rho_{j,k}^{\mathrm{nonadd}}$ isolates the interaction component (synergy, buffering, suppression), which is zero for any additive predictor and positive only when the recovered model captures structure beyond single-gene superposition. We additionally report Top-1 and Top-5 retrieval among held-out pairs by cosine similarity, which is sensitive to direction in the high-dimensional shift space and complements per-pair Pearson, alongside the Frobenius norm and effective rank of the recovered A as descriptive diagnostics of the fitted structure.

Structural validation. We assess the recovered adjacency against three independent reference types: curated hematopoietic lineage signatures, ChEA 2022 ChIP-seq target sets (Griffon et al., 2015), and ENCODE TF ChIP-seq 2015 target sets (Consortium et al., 2012), with hypergeometric enrichment of the top-25 recovered targets per TF (Appendix I). As these references are unused during fitting, agreement between the recovered A and any of them is independent evidence of structural recovery rather than a consistency check.

Encoder transmission of regulatory geometry. We probe each encoder by injecting a controlled signal along orthonormal directions of gene space and measuring its response. For each single-gene CRISPR condition j, we define the observed shift $\Delta x_j = \mathbb{E}[x \mid \mathrm{SKI}_j] - \bar{x}_{\mathrm{ctrl}}$ in log1p HVG space and stack these across j into a regulatory-signal matrix S. We obtain two orthonormal bases of gene space by SVD: α from centered control expression, and γ from S. The two probe complementary geometries. α tests whether regulatory signal lies along directions of natural variation in unperturbed cells, while γ tests whether the encoder preserves the directions along which observed perturbations actually moved cells. For each basis vector v_k, the regulatory weight $w_k = \sum_j (\Delta x_j \cdot v_k)^2$ then measures how much signal energy the perturbation battery places along v_k. The encoder transmission

$$t_k = \frac{1}{n_a} \sum_{i=1}^{n_a} \frac{\|\mathrm{Enc}(x_i + c\, v_k) - \mathrm{Enc}(x_i)\|_2}{c} \tag{7}$$

is the embedding response when v_k is injected onto each of n_a control cells, with c set to the median observed shift norm so the encoder is probed in the regime where real perturbations operate. Identity gives $t_k = 1$ for all k by construction, and any linear encoder M gives $t_k = \|M v_k\|_2$, so PCA transmits 1 inside its retained subspace and 0 outside. The preserved regulatory fraction is the share of regulatory signal energy that survives encoding:

$$\rho_{\mathrm{enc}} = \frac{\sum_k w_k t_k^2}{\sum_k w_k} \tag{8}$$

3. Experimental Results

We evaluate the framework along the three axes set out in §2.5. The first tests predictive accuracy on Norman 2019 Perturb-seq, asking whether single-gene perturbations fit during training compose to predict held-out double-gene responses and capture the non-additive interactions that an additive baseline cannot express (§3.1). The second assesses structural fidelity by comparing the recovered adjacency against independent ChIP-seq and lineage signature references (§3.2). The third probes each encoder directly, measuring how they transmit or attenuate the geometric directions where regulatory signal lives (§3.3).

Table 1. **Prediction, retrieval, and learned structure on Norman 2019.** SKI val, DKI val, retrieval, and learned-structure columns are mean ± SD across 5 seeds; ρ^{nonadd} is reported as per-pair bootstrap mean and 95% CI over $n = 129$ held-out double-gene pairs. Additive baseline included as reference: ρ^{nonadd} is zero by construction for additive predictors and isolates the interaction component.

	Prediction			Retrieval (%)		Learned structure	
Condition	SKI val	DKI val	DKI val ρ^{nonadd}	Top-1	Top-5	$\|A\|_F$	Eff. rank
Identity	0.706 ± 0.011	0.782 ± 0.008	$0.492\ [0.459, 0.526]$	77.3 ± 1.3	85.7 ± 2.6	2.12 ± 0.06	69.4 ± 2.3
PCA	0.540 ± 0.069	0.582 ± 0.067	$0.312\ [0.150, 0.474]$	78.6 ± 1.9	83.7 ± 1.1	4.35 ± 0.97	77.2 ± 1.3
scGPT	-0.001 ± 0.006	-0.011 ± 0.019	$0.373\ [0.088, 0.658]$	7.6 ± 1.7	17.8 ± 3.0	0.23 ± 0.13	98.4 ± 1.5
scPRINT	0.090 ± 0.022	0.171 ± 0.032	$0.283\ [0.133, 0.433]$	41.0 ± 15.2	62.0 ± 12.3	3.56 ± 0.49	46.4 ± 1.7
Additive	–	0.801 ± 0.125	–	–	–	–	–

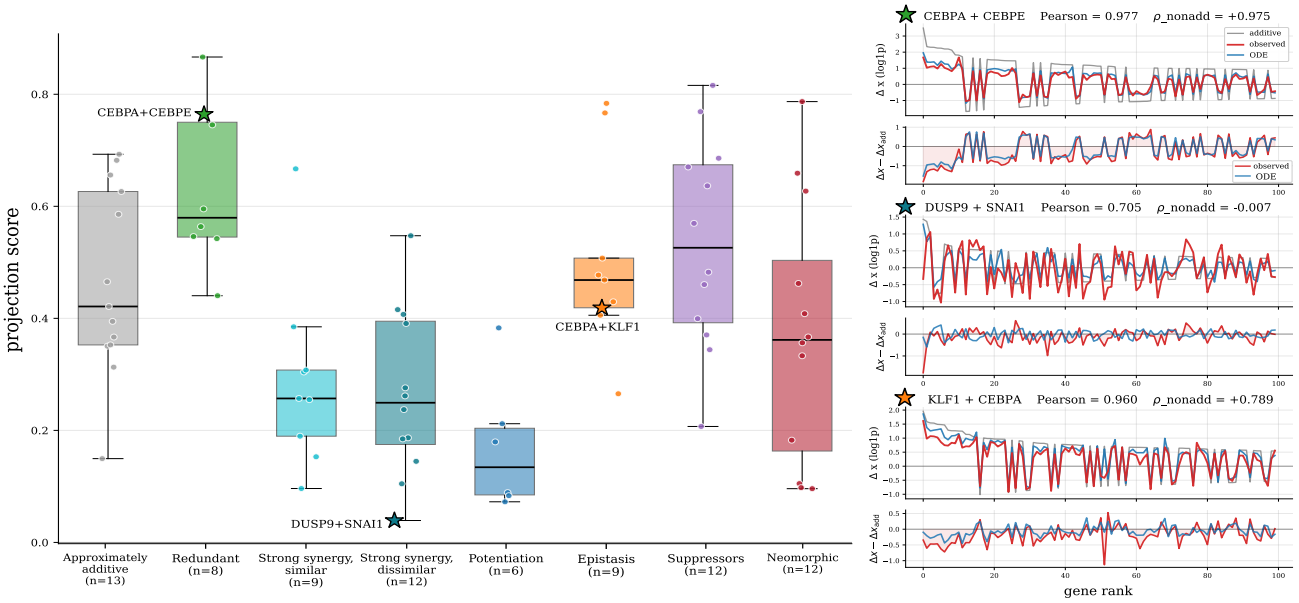

Figure 2. **Non-additive recovery on held-out double-KI pairs, stratified by Norman 2019 gene categories. Left:** Projection of the predicted residual onto the observed residual direction, normalized by $\|e_{\text{obs}}\|$. 1 = predicted residual matches observed in magnitude and direction; 0 = no component along observed; negative = anti-aligned. Boxes show median and IQR, circles are individual held-out pairs. Stars mark the three pairs shown on the right. **Right:** per-pair waterfall plots. Top sub-panel shows Δx from control across the top 100 genes ranked by $|\Delta x_{\text{add}}|$, comparing observed (red), ODE prediction (blue), and additive baseline (grey). Bottom sub-panel shows the residual $\Delta x - \Delta x_{\text{add}}$ for observed (red) and ODE (blue). Per-panel y-axis ranges differ.

3.1. Held-out prediction

Table 1 reports prediction, retrieval, and learned structure for the four encoder conditions and the additive baseline. Identity recovers a Hill-kinetics adjacency from single-KI data alone that predicts held-out double-KI shifts at 0.782 ± 0.008 aggregate Pearson, captures their non-additive component at $\rho^{\text{nonadd}} = 0.492$ with 95% bootstrap CI $[0.459, 0.526]$ over $n = 129$ pairs, reaches 0.706 ± 0.011 on single-KI prediction, retrieves the matching observed shift as the predicted nearest neighbour for 77.3% of held-out pairs, and recovers an adjacency of Frobenius norm $\|A\|_F = 2.12$ at effective rank 69.4. Top-1 retrieval well above chance ($\approx 1\%$) shows that predictions discriminate between perturbation conditions, and the moderate Frobenius norm at sub-maximal effective rank (bounded above by the 102 TF columns) indicates that the recovered adjacency concentrates signal on a structured subset of regulatory directions rather than spreading it diffusely.

The additive baseline is statistically indistinguishable from Identity within error bars (0.801 ± 0.125 vs 0.782 ± 0.008) but has zero non-additive component by construction: aggregate Pearson is dominated by the additive component of double-gene responses, so ρ^{nonadd} is the metric that isolates the interaction structure distinguishing mechanistic recovery from arithmetic composition. PCA tracks Identity at reduced magnitude with a denser higher-rank adjacency ($\|A\|_F = 4.35$, rank 77.2). scGPT collapses to near-zero correlation on every prediction column (Top-1 retrieval 7.6%) with $\|A\|_F = 0.23$ and near-maximal effective rank 98.4, a signature consistent with an ODE finding no learnable signal under the encoder rather than absent perturbation signal at the input (§3.3). scPRINT places

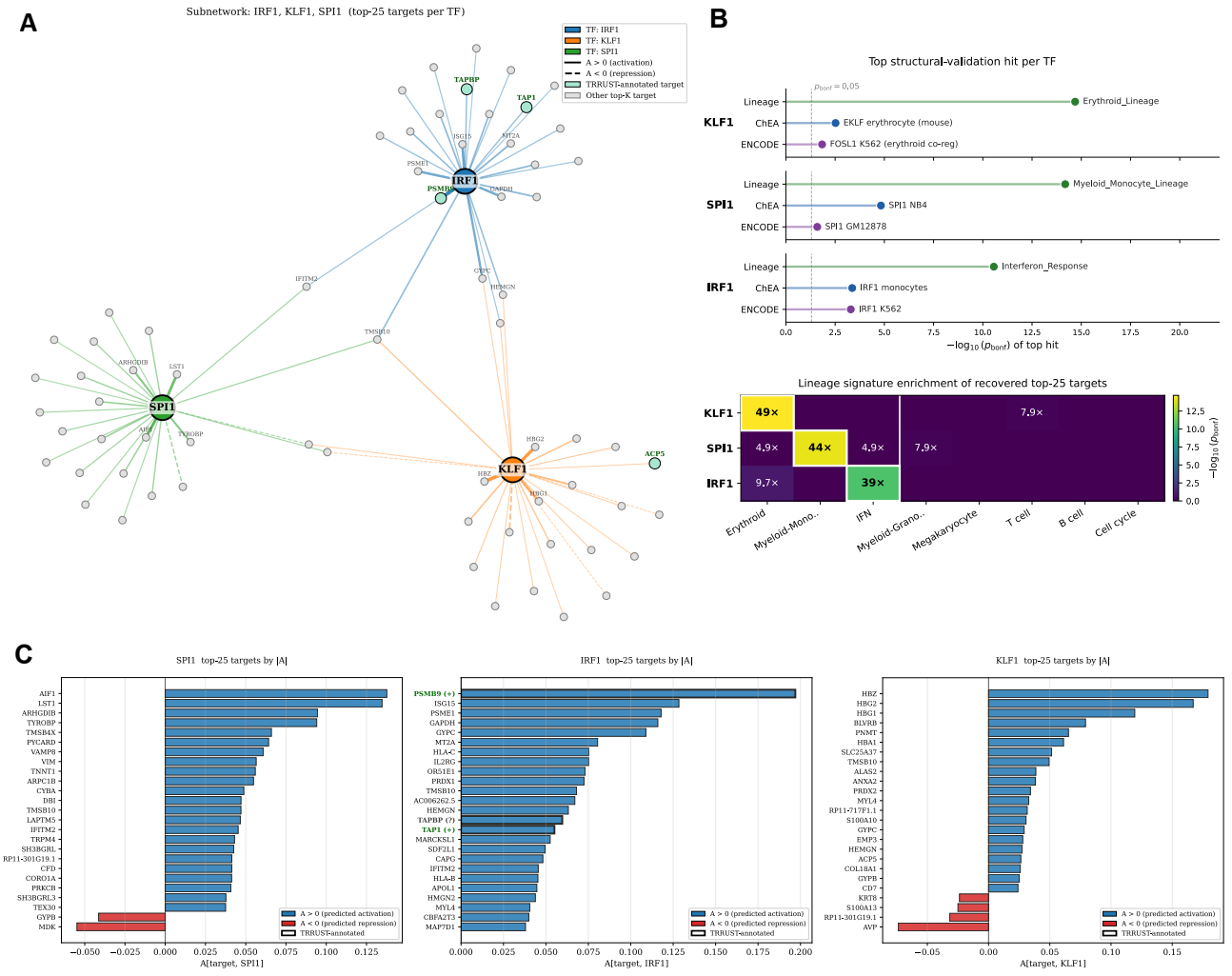

Figure 3. **Structural validation of the recovered adjacency. (A)** Combined subnetwork of the top-25 recovered targets per master TF (*IRF1*, *KLF1*, *SPI1*). Edges colored by source; solid = activation ($A > 0$), dashed = repression. TRRUST-annotated targets shown with colored fill. **(B)** Top hit per TF across three reference types (lineage signatures, ChEA 2022, ENCODE 2015), reported as $-\log_{10}(p_{\text{bonf}})$ of the strongest matched-TF or matched-lineage gene set. Dashed line: $p_{\text{bonf}} = 0.05$. **(C)** Fold enrichment of each TF's top-25 against eight curated hematopoietic lineage signatures, colored by $-\log_{10}(p_{\text{bonf}})$. Expected hits boxed.

between scGPT and PCA on prediction (DKI val 0.171) at lower rank ($\|A\|_F = 3.56$, rank 46.4), suggesting a low-dimensional regulatory program.

Stratifying held-out pairs by Norman 2019 functional categories (Figure 2, left) shows a structured pattern. Recovery is strongest on Redundant, Suppressor, and Epistasis pairs (median projection $\sim$ 0.5–0.6), intermediate on Approximately additive and Neomorphic pairs ($\sim$ 0.4), and weakest on Strong synergy and Potentiation pairs ($\sim$ 0.15–0.25) where the joint phenotype is qualitatively new and not implicit in either TF's single-gene response. The ODE recovers non-additive structure cleanly on redundant interactions (CEBPA+CEBPE, $r = 0.98$, $\rho^{\text{nonadd}} = +0.98$) and on epistatic masking (KLF1+CEBPA, $r = 0.96$, $\rho^{\text{nonadd}} = +0.79$), and weakens on dissimilar-phenotype synergy (DUSP9+SNAI1, $r = 0.71$, $\rho^{\text{nonadd}} = -0.01$) where high aggregate Pearson hides limited residual recovery.

3.2. Structural validation

Predictive accuracy on held-out perturbations does not identify which adjacency from the consistency family has been recovered. We assess structural recovery against three independent reference types for the Identity encoder on three master hematopoietic regulators chosen to span orthogonal axes of the K562 bipotential state. KLF1 and SPI1 serve as the definitive anchors for its two mutually exclusive developmental fates (erythroid and myeloid, respectively (Miller & Bieker, 1993; Scott et al., 1994)) while IRF1 provides a biologically independent axis of external immune response (Fujita et al., 1989). Top-25 recovered

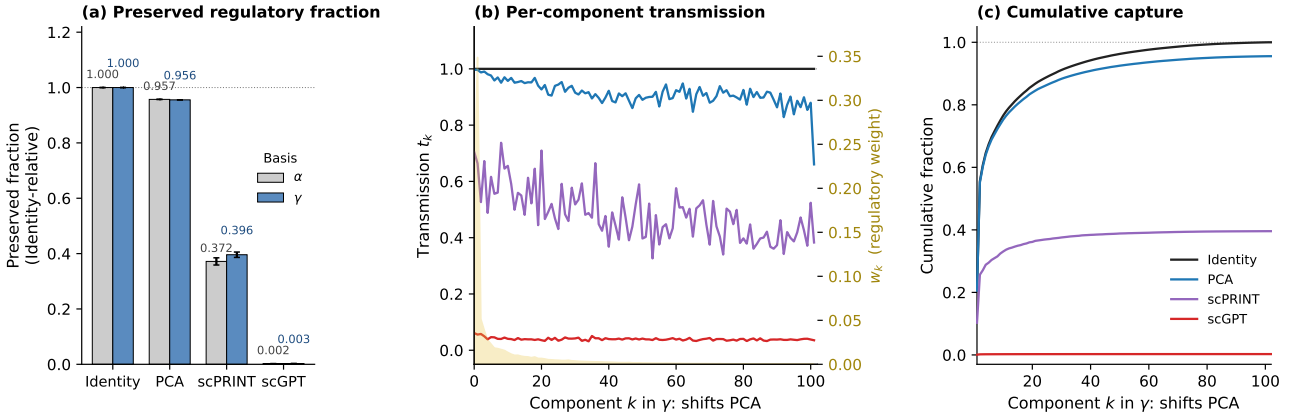

Figure 4. **Encoder transmission of regulatory signal on Norman Perturb-seq.** Each encoder is evaluated by injecting orthonormal basis directions onto control cells and measuring the resulting embedding shift across two bases of gene space: α, the PCA of control expression, and γ, the PCA of observed perturbation shifts. **(a)** Preserved regulatory fraction, the share of perturbation-shift energy that survives encoding, normalized so Identity equals 1. **(b)** Per-component transmission t_k in γ, the fraction of basis direction v_k (the k-th principal axis of the perturbation shifts) that the encoder passes through. Yellow shading marks the regulatory weight w_k, the share of regulatory signal energy that lives along v_k. **(c)** Cumulative captured fraction as components are accumulated, with the no-encoder reference (Identity) tracing where the regulatory signal lives. Mean ± SD across 5 seeds.

targets per TF are visualized as a combined subnetwork (Figure 3A) and ranked individually (Figure 3C).

Each TF's top-25 enriches strongly against its matched lineage signature: KLF1 against erythroid (48.6×, $p_{\text{bonf}} = 2.1\text{e}{-15}$), SPI1 against myeloid monocyte (44.4×, $p_{\text{bonf}} = 6.6\text{e}{-15}$), and IRF1 against the combined interferon response signature (38.9×, $p_{\text{bonf}} = 2.8\text{e}{-11}$). Figure 3B shows the diagonal of expected matches across all eight lineage signatures: matched lineages reach $p_{\text{bonf}} < 10^{-10}$ for all three TFs while every wrong-lineage and negative-control signature (T cell, B cell, cell cycle) remains above $p_{\text{bonf}} = 0.1$. Independently curated ChIP-seq references confirm these matches, with each TF's top-25 enriching against same-TF target sets in ChEA 2022 and ENCODE 2015 (top hits in Figure 3B, full statistics in Appendix I).

Beyond matched-TF hits, each TF's top-25 also enriches against ChIP target sets of its known co-regulators in the matched cellular context (Appendix I). KLF1's recovered targets enrich against GATA1 and GATA2 K562 ChIP (canonical erythroid co-regulators), IRF1's against STAT2 K562 ChIP (the IRF1–STAT2 axis in type I interferon signaling), and SPI1's against TAL1, RUNX1, and LYL1 ChIP in mouse hematopoietic progenitors, three of seven members of the *heptad* TF complex that co-regulates hematopoietic stem and progenitor programs alongside SPI1 (Wilson et al., 2010; Subramanian et al., 2023). The framework fits each TF independently and is given no information about co-regulatory relationships, so these recovered structures emerge as a consequence of the single shared adjacency formulation: fitting one A across all conditions allows information from each perturbation to constrain entries beyond the directly perturbed gene.

3.3. Encoder transmission

The compositional gap between Identity and FMs could reflect either an ODE that cannot extract regulatory information from FM embeddings, or embeddings that destroyed it before the ODE saw it. We distinguish these by measuring encoder transmission directly, without an ODE in the loop. Identity preserves 1.00 of regulatory signal energy by construction, PCA 0.96, scPRINT 0.37–0.40 depending on basis (an intermediate regime between PCA and scGPT), and scGPT 0.003 (Figure 4a). The two bases yield closely matching values for every encoder, so the conclusion does not depend on whether preservation is measured along directions of natural variation in unperturbed cells (α) or along directions established by the perturbations themselves (γ). The per-component curves in γ make the failure mode explicit (Figure 4b). Regulatory weight concentrates in the first few components, with over half the signal energy in the top two and ninety percent in the top thirty. PCA's transmission tracks Identity throughout, dropping only at the very end where its retained subspace runs out. scGPT's transmission sits uniformly near zero across all components, including those carrying the most signal: the encoder is not selectively suppressing low-weight directions, it is attenuating every direction the regulatory signal could possibly use. scPRINT's transmission stays cleanly between PCA and scGPT across all components, consistent with the headline preservation falling in the same gap. The cumulative captured fraction (Figure 4c) reinforces these patterns visually.

These measurements bound how much of the FM compo-

sitional failure is attributable to the encoder rather than to ODE fitting. Whatever a structured ODE could in principle recover from the scGPT embedding, it can recover only what survives encoding, and on this measurement very little does. The ordering across encoders is consistent with the compositional Pearson ordering on held-out double-KIs: preservation and compositional accuracy track each other across three encoders that span two orders of magnitude on both metrics. We read this as evidence that compositional ODE recovery is rate-limited by the regulatory geometry the encoder transmits, with the caveat that the present analysis does not exclude the possibility that a more flexible fitter could extract more from a low-transmission representation than our Hill-kinetics ODE does. Yet the 100-fold gap in preserved regulatory fraction between scPRINT (0.40) and scGPT (0.003) shows the bottleneck is not an inevitability of FM training: expression-FM training varies substantially in how much regulatory geometry it preserves, with scGPT actively attenuating regulatory directions while scPRINT retains a substantial fraction.

4. Conclusion

We presented CLAMP, a steady-state ODE recovery framework that reframes sustained CRISPR perturbation as a coordinate constraint on a regulatory network and fits a single shared adjacency jointly across all single-gene conditions. Implicit differentiation through the fixed-point solver makes this tractable at Perturb-seq scales, and the encoder-agnostic interface admits any frozen differentiable representation as a metric on observations. On Norman 2019, the framework predicts held-out double-KI shifts at 0.78 aggregate Pearson and captures their non-additive component at $\rho^{\text{nonadd}} = 0.49$ (95% bootstrap CI $[0.46, 0.53]$), recovers an adjacency that aligns with curated lineage signatures and ChIP-seq references at $p_{\text{bonf}} < 10^{-10}$, and respects known co-regulatory groupings without being given them. Deployed as a probe, the framework reveals that single-cell foundation models vary substantially in how much regulatory geometry they preserve, with scGPT actively attenuating regulatory directions while scPRINT retains a substantial fraction. The compositional accuracy ordering across encoders matches their preservation ordering on a measurement that involves no ODE at all, suggesting that mechanistic recovery is rate-limited by representational geometry. The most promising direction for future work therefore concerns FM development itself: representation-learning objectives that explicitly preserve high-regulatory-weight directions could close most of the gap between FM and Identity recovery, with the encoder transmission probe providing a fitter-independent metric for regulatory-geometry preservation that serves as the evaluation criterion. We present a deeper analysis of the scope, assumptions and limitations of this work in Appendix C.

References

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

Bai, D., Ellington, C. N., Mo, S., Song, L., and Xing, E. P. Attentionpert: accurately modeling multiplexed genetic perturbations with multi-scale effects. *Bioinformatics*, 40 (Supplement_1):i453–i461, 2024.

Bai, S., Kolter, J. Z., and Koltun, V. Deep equilibrium models. *Advances in Neural Information Processing Systems*, 32, 2019.

Bertin, P., Viviano, J. D., Tejada-Lapuerta, A., Wang, W., Bauer, S., Theis, F. J., and Bengio, Y. A scalable gene network model of regulatory dynamics in single cells. *arXiv preprint arXiv:2503.20027*, 2025.

Blondel, M., Berthet, Q., Cuturi, M., Frostig, R., Hoyer, S., Llinares-López, F., Pedregosa, F., and Vert, J.-P. Efficient and modular implicit differentiation. *Advances in Neural Information Processing Systems*, 35:5230–5242, 2022.

Consortium, E. P. et al. An integrated encyclopedia of dna elements in the human genome. *Nature*, 489(7414):57, 2012.

Cui, H., Wang, C., Maan, H., Pang, K., Luo, F., Duan, N., and Wang, B. scgpt: toward building a foundation model for single-cell multi-omics using generative ai. *Nature methods*, 21(8):1470–1480, 2024.

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C. P., Jerby-Arnon, L., Marjanovic, N. D., Dionne, D., Burks, T., Raychowdhury, R., et al. Perturb-seq: dissecting molecular circuits with scalable single-cell rna profiling of pooled genetic screens. *cell*, 167(7):1853–1866, 2016.

Fujita, T., Kimura, Y., Miyamoto, M., Barsoumian, E. L., and Taniguchi, T. Induction of endogenous ifn-α and ifn-β genes by a regulatory transcription factor, irf-1. *Nature*, 337(6204):270–272, 1989.

Griffon, A., Barbier, Q., Dalino, J., van Helden, J., Spicuglia, S., and Ballester, B. Integrative analysis of public chip-seq experiments reveals a complex multi-cell regulatory landscape. *Nucleic acids research*, 43(4):e27–e27, 2015.

Han, H., Cho, J.-W., Lee, S., Yun, A., Kim, H., Bae, D., Yang, S., Kim, C. Y., Lee, M., Kim, E., et al. Trrust v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic acids research*, 46(D1):D380–D386, 2018.

Hay, S. B., Ferchen, K., Chetal, K., Grimes, H. L., and Salomonis, N. The human cell atlas bone marrow single-cell interactive web portal. *Experimental Hematology*, 68:51–61, 2018.

Hossain, I., Fanfani, V., Fischer, J., Quackenbush, J., and Burkholz, R. Biologically informed neuralodes for genome-wide regulatory dynamics. *Genome Biology*, 25(1):127, 2024.

Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., et al. The reactome pathway knowledgebase. *Nucleic acids research*, 48(D1):D498–D503, 2020.

Kalfon, J., Samaran, J., Peyré, G., and Cantini, L. scprint: pre-training on 50 million cells allows robust gene network predictions. *Nature Communications*, 16(1):3607, 2025.

Kamimoto, K., Stringa, B., Hoffmann, C. M., Jindal, K., Solnica-Krezel, L., and Morris, S. A. Dissecting cell identity via network inference and in silico gene perturbation. *Nature*, 614(7949):742–751, 2023.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. The molecular signatures database hallmark gene set collection. *Cell systems*, 1(6): 417–425, 2015.

Lin, Z., Chang, S., Zweig, A., Kang, M., Azizi, E., and Knowles, D. A. Interpretable neural odes for gene regulatory network discovery under perturbations. *arXiv preprint arXiv:2501.02409*, 2025.

Loshchilov, I. and Hutter, F. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Lotfollahi, M., Klimovskaia Susmelj, A., De Donno, C., Hetzel, L., Ji, Y., Ibarra, I. L., Srivatsan, S. R., Naghipourfar, M., Daza, R. M., Martin, B., et al. Predicting cellular responses to complex perturbations in high-throughput screens. *Molecular systems biology*, 19(6): MSB202211517, 2023.

Miller, I. J. and Bieker, J. J. A novel, erythroid cell-specific murine transcription factor that binds to the caccc element and is related to the krüppel family of nuclear proteins. *Molecular and cellular biology*, 13(5):2776–2786, 1993.

Norman, T. M., Horlbeck, M. A., Replogle, J. M., Ge, A. Y., Xu, A., Jost, M., Gilbert, L. A., and Weissman, J. S. Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. *Science*, 365(6455):786–793, 2019.

Novershtern, N., Subramanian, A., Lawton, L. N., Mak, R. H., Haining, W. N., McConkey, M. E., Habib, N., Yosef, N., Chang, C. Y., Shay, T., et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell*, 144(2):296–309, 2011.

Osorio, D., Zhong, Y., Li, G., Xu, Q., Yang, Y., Tian, Y., Chapkin, R. S., Huang, J. Z., and Cai, J. J. sctenifoldknk: An efficient virtual knockout tool for gene function predictions via single-cell gene regulatory network perturbation. *Patterns*, 3(3), 2022.

Rohbeck, M., Clarke, B., Mikulik, K., Pettet, A., Stegle, O., and Ueltzhöffer, K. Bicycle: Intervention-based causal discovery with cycles. In *Causal Learning and Reasoning*, pp. 209–242. PMLR, 2024.

Roohani, Y., Huang, K., and Leskovec, J. Predicting transcriptional outcomes of novel multigene perturbations with gears. *Nature Biotechnology*, 42(6):927–935, 2024.

Schoenfelder, S. and Fraser, P. Long-range enhancer–promoter contacts in gene expression control. *Nature Reviews Genetics*, 20(8):437–455, 2019.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. Global quantification of mammalian gene expression control. *Nature*, 473(7347):337–342, 2011.

Scott, E. W., Simon, M. C., Anastasi, J., and Singh, H. Requirement of transcription factor pu. 1 in the development of multiple hematopoietic lineages. *Science*, 265(5178): 1573–1577, 1994.

Subramanian, S., Thoms, J. A., Huang, Y., Cornejo-Páramo, P., Koch, F. C., Jacquelin, S., Shen, S., Song, E., Joshi, S., Brownlee, C., et al. Genome-wide transcription factor–binding maps reveal cell-specific changes in the regulatory architecture of human hspcs. *Blood*, 142(17):1448–1462, 2023.

Tenen, D. G. Disruption of differentiation in human cancer: Aml shows the way. *Nature Reviews Cancer*, 3(2):89–101, 2003.

Theodoris, C. V., Xiao, L., Chopra, A., Chaffin, M. D., Al Sayed, Z. R., Hill, M. C., Mantineo, H., Brydon, E. M., Zeng, Z., Liu, X. S., et al. Transfer learning enables predictions in network biology. *Nature*, 618(7965):616–624, 2023.

von Bassewitz, J.-P., Pervez, A., Fumero, M., Robinson, M., Karaletsos, T., and Locatello, F. Learning explicit single-cell dynamics using ode representations. *International Conference on Learning Representation*, 2026.

Wenteler, A., Occhetta, M., Branson, N., Curean, V., Huebner, M., Dee, W., Connell, W., Chung, S. P., Hawkins-Hooker, A., Ektefaie, Y., Córdova, C. M. V.,

and Gallagher-Syed, A. Perteval-scFM: Benchmarking single-cell foundation models for perturbation effect prediction. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=t04D9bkKUq.

Wilson, N. K., Foster, S. D., Wang, X., Knezevic, K., Schütte, J., Kaimakis, P., Chilarska, P. M., Kinston, S., Ouwehand, W. H., Dzierzak, E., et al. Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell stem cell*, 7(4):532–544, 2010.

Zecha, J., Meng, C., Zolg, D. P., Samaras, P., Wilhelm, M., and Kuster, B. Peptide level turnover measurements enable the study of proteoform dynamics. *Molecular & Cellular Proteomics*, 17(5):974–992, 2018.

Supplementary Material

This supplementary material provides methodological details deferred from the main text.

- Appendix A (supporting Section 1.1) gives an overview of various pre-existing methods that use dynamical systems approaches to model gene regulatory networks and how our CLAMP fits into this.
- Appendix B (supporting Section 2) gives an overview of the CLAMP framework and how it is optimized.
- Appendix C contains a detailed description of the scope, assumptions and limitations of CLAMP.
- Appendix D (supporting Section 2.1) derives the Hill function from cooperative TF binding and discusses our use of shared kinetic parameters.
- Appendix E (supporting Section 2.2) reports the optimizer, loss-weight, solver, and training-schedule values used across all experiments.
- Appendix F (supporting Section 2.3) gives the implicit-gradient derivation, closed-form Jacobian and Hill derivatives, per-parameter gradient expressions, and numerical-stability and verification details.
- Appendix G (supporting Section 2.4) documents the per-encoder wrappers, including vocabulary mapping, input format, depth-token handling, and the diagnostic checks each wrapper passes before being used in experiments.
- Appendix H details the dense adjacency parameterization, including the transcription-factor restriction, the reapplication of the TF mask under AdamW, and the initialization scheme.
- Appendix I reports the full structural validation results against curated hematopoietic lineage signatures, ChEA 2022 ChIP-seq target sets, ENCODE TF ChIP-seq 2015 target sets, MSigDB Hallmark, Reactome 2022, and TRRUST literature-curated targets, alongside a discussion of the recovered local regulatory neighborhood and methodological notes on multiple-testing correction and background gene-set choice.
- Appendix J documents the curated hematopoietic lineage signature library used as the primary structural validation reference, including per-signature gene lists and curation sources, and specifies the disease-context and generic catch-all denylist used to filter MSigDB Hallmark and Reactome 2022 prior to enrichment testing.

A. Overview of dynamical-system approaches for gene regulatory network modeling

Table 2. Comparison of dynamical systems methods for gene regulatory network inference

Method	Dynamical System Type	Temporal Scope	Latent Space Constraint	Interpretability Mechanism	Unseen Perturbations
Cell-MNN	Locally Linearized ODE	Trajectory / Interpolation	PCA Subspace	Linear Transition Matrix	×
FLECS	Message-Passing ODE	Pseudotime / Trajectory	None (Gene Space)	Sparse Prior Network Weights	×
PerturbODE	Neural Drift ODE	Trajectory / Steady-State	None (Gene Space)	Post-hoc Jacobian Extraction	×
PHOENIX	Hill-Langmuir Neural ODE	Bulk Temporal Microarray	None (Gene Space)	Imposed Functional Forms	×
Bicycle	Linear SDE (Ornstein-Uhlenbeck)	Steady-State Snapshot	Unconstrained Latent Space	Explicit Causal Matrix via Lyapunov	✓
CLAMP	Hill-Kinetics Fixed-Point ODE	Steady-State Snapshot	Encoder-Agnostic	Explicit Kinetic Parameters	✓

B. Full Pipeline

Figure 1 summarizes the full CLAMP pipeline and the relationship between its three moving parts: a shared regulatory adjacency, an inner solver that maps that adjacency to a constrained steady state for each perturbation condition, and an outer optimizer that updates the adjacency to match observations. The schema is intended as a roadmap for the formal treatment in Section 2, and we walk through it here in the order in which information flows through the framework at fitting time and again at inference time.

Panel A states the modeling assumption that drives the rest of the framework. A CRISPR perturbation of gene j is treated as a coordinate constraint $x_j = c_j$ imposed on the shared regulatory ODE of Eq. 1, with the unperturbed coordinates relaxing to the self-consistent fixed point of Eq. 2. The same shared adjacency A governs every condition in $\mathcal{J}$, so a k-gene perturbation is resolved by imposing k simultaneous clamps on the same equations. Panel A right illustrates this compositional step on the held-out pair CEBPA+CEBPE, where the predicted and observed shifts agree closely on both the additive and the non-additive components of the joint response. Crucially, no information about double-gene perturbations enters the fit: the recovered θ comes from single-gene conditions alone, and the double-gene response is read out through a forward solve of the same dynamics under a second clamp.

Panel B traces the optimization loop that produces θ. Step (1) is the parameter object: the adjacency $A \in \mathbb{R}^{G \times G_{\mathrm{TF}}}$, restricted to columns corresponding to the $G_{\mathrm{TF}} = 102$ perturbed transcription factors as discussed in Section 2.4. Step (2) is the inner loop. For each condition j, the constrained steady state $\mathbf{x}^{*,(j)}(\theta)$ is computed by damped fixed-point iteration in gene space, with the clamp $x_j^{*,(j)} = c_j$ reimposed at every iteration so that the converged state lies exactly on the clamp plane shown in the inset. The forward iteration is run without an autograd graph and returns only the converged $\mathbf{x}^*$, decoupling memory cost from solver depth.

Step (3) is the outer loop. The predicted steady states $\{\mathbf{x}^{*,(j)}(\theta)\}_{j \in \mathcal{J}}$ are compared to the observed condition means through the frozen encoder via the three-component loss of Eq. 4, with the perturbation-shift term carrying the regulatory signal, the control-anchor term fixing the reference point that the shift formulation alone leaves undetermined, and the L_1 sparsity penalty encoding the prior that biological regulatory networks are sparse. Gradients with respect to θ are obtained by implicit differentiation of the fixed-point equation $F(\mathbf{x}^*, \theta) = \mathbf{0}$, as developed in Section 2.3, reducing the backward pass to a single $G \times G$ linear solve per condition. These gradients flow back to step (1), closing the loop. The encoder appears only at the loss layer and is held fixed throughout, so substituting Identity, PCA, or a frozen foundation model changes which projection of the steady state the loss compares but leaves the rest of the pipeline untouched.

At inference time only the inner loop is exercised. A new perturbation, possibly involving multiple target genes, is encoded as a set of simultaneous clamps on the same fitted θ, and the predicted response is read off from the converged steady state. This is the mechanism behind the held-out compositional results of Section 3.1: a k-gene perturbation is resolved by the same solver that produced the single-gene training states, with no architectural extension and no additional training data.

C. Scope, assumptions and limitations

The framework rests on four assumptions that delimit where it applies and where it does not.

1. **Steady-state response.** The steady-state assumption (Section 2.1) requires that the cellular response to perturbation has equilibrated by the time of measurement. This is well-supported for the 5–6 day post-infection sampling window of Norman 2019 given the protein turnover times of human cell lines, but would not hold for short-timescale perturbations such as drug treatments sampled within a few hours, where the observed state lies along a trajectory rather than at a fixed point.

2. **Single deterministic attractor per condition.** The framework treats each condition as a single deterministic attractor and fits to the condition mean. It does not model cell-to-cell variability within a condition, multistability, or the case where a single perturbation drives the population towards two or more distinct fates.

3. **Adjacency support restricted to perturbed TFs.** The recovered adjacency is restricted to the columns of perturbed transcription factors, so A_{ij} is identified only for $j \in \mathcal{J}$. This restriction is both a methodological choice and a reflection of what the data supports: only the perturbed TFs carry direct interventional evidence, and fitting columns for unperturbed genes would require the model to extrapolate regulatory effects from observational variation alone. The restriction also keeps the parameter count tractable, reducing the adjacency from $G^2 \approx 6.5 \times 10^6$ to $G \cdot G_{\mathrm{TF}} \approx 2.6 \times 10^5$

and yielding the roughly one-to-one parameter-to-observation regime discussed in Section 2.4. Edges between genes that were never directly perturbed are not estimated, and the framework is silent on regulators outside the intervention battery.

4. **Hill kinetics as the instantiation choice.** The kinetic form used in this work is fixed to Hill activation with a shared cooperativity (n), decay rate (γ) and half-maximal constant (K) across all regulators (Appendix D). The framework itself does not require this: any differentiable kinetic form admitting a fixed point (mass action, Michaelis–Menten, or learned neural drift functions, among others) can be substituted without changing the fitting procedure, and Hill activation is the choice made in this work for biophysical interpretability and parameter parsimony. The Hill choice is a deliberate compression of biological reality: real transcription factors vary in their cooperativity and binding affinities, and individual transcripts exhibit heterogeneous decay rates. The signed scalar A_{ij} and shared kinetic scalars collapse these into a single effective coupling. This compression is what makes joint multi-condition fitting tractable at the parameter counts of Section 2.4, and the structural and predictive results of Section 3.1 and Section 3.2 indicate it is empirically adequate at this scale, but we do not claim that the recovered A disentangles direct from indirect regulation or recovers the underlying biophysical kinetics of any individual edge.

The fourth assumption is the source of the identifiability concern noted in Section 3.2, and the structural validation results presented there are the response: independent ChIP-seq and lineage signatures provide external evidence that the recovered A aligns with the true regulatory structure rather than merely a steady-state-equivalent surrogate.

D. Hill kinetics: biophysical derivation and parameter choices

This appendix derives the Hill function from cooperative TF-promoter binding equilibria, justifies our use of shared scalar parameters in the synthetic benchmark, and notes the framework's generality over kinetic forms beyond Hill.

D.1. Derivation of the Hill function

Consider a promoter with n binding sites for a transcription factor (TF) at intracellular concentration x. Under the cooperative-binding assumption, the promoter is transcriptionally active only when all n sites are simultaneously occupied by the TF. Writing the cooperative binding reaction as $\mathrm{P} + n\,\mathrm{TF} \rightleftharpoons \mathrm{P}_{\text{bound}}$, the equilibrium dissociation constant is

$$K_d = \frac{[\mathrm{P}]\,x^n}{[\mathrm{P}_{\text{bound}}]}. \tag{9}$$

The fraction of time the promoter spends in the active state is therefore

$$\frac{[\mathrm{P}_{\text{bound}}]}{[\mathrm{P}] + [\mathrm{P}_{\text{bound}}]} = \frac{x^n/K_d}{1 + x^n/K_d} = \frac{x^n}{K^n + x^n} = h(x;\, n,\, K), \tag{10}$$

where $K = K_d^{1/n}$ is the concentration at which the promoter is half-occupied. The Hill function is sigmoidal in x, with steepness controlled by n: at $n = 1$ it reduces to Michaelis-Menten saturation, and as $n \to \infty$ it approaches a hard threshold at $x = K$. We use $n = 2$ as a default, corresponding to dimerized cooperative binding, the most common regime in eukaryotic transcriptional regulation.

D.2. Per-regulator versus multi-regulator nonlinearity

When a target gene i is regulated by multiple transcription factors $\{x_j\}_{j \in \text{regs}(i)}$, two natural functional forms are available. The *per-regulator* form, used in this work,

$$\frac{dx_i}{dt} \supset \sum_j A_{ij}\, h(x_j;\, n,\, K), \tag{11}$$

applies the Hill nonlinearity to each regulator independently before summing. The *multi-regulator* form,

$$\frac{dx_i}{dt} \supset h\left(\sum_j A_{ij}\, x_j;\, n,\, K\right), \tag{12}$$

applies a single Hill activation to a linear combination of regulators. Per-regulator nonlinearity is consistent with how cooperativity is modeled in deterministic systems biology when each TF binds independently at distinct sites, and it makes the individual edge coefficients A_{ij} interpretable as marginal regulatory contributions. The multi-regulator form, used by FLeCS (Bertin et al., 2025), is a coarser approximation that compresses regulator interactions into a sigmoidal threshold on a weighted sum and obscures the contribution of any single TF to the response. Both forms are admissible in our framework without changing the recovery procedure. We adopt the per-regulator form for interpretability.

D.3. Generality over kinetic forms

The recovery framework treats the production term in Equation 1 as a black-box differentiable function. Hill kinetics is one instantiation, chosen for biological interpretability. Other admissible parameterizations include the following.

- **Michaelis-Menten** kinetics, the $n = 1$ special case of Hill, suited to non-cooperative regulation.
- **Mass action** kinetics, $\frac{dx_i}{dt} \supset \sum_j A_{ij} x_j$, the linear special case in which each regulator contributes proportionally to its concentration.
- **Generalized Hill** with regulator-specific scalars $h(x_j;\, n_j,\, K_j)$, relaxing the shared-scalar simplification at the cost of additional parameters.
- **Neural drift**, $\frac{dx_i}{dt} \supset f_\phi(\mathbf{x})$ with f_ϕ a differentiable neural network, abandoning interpretability of individual edges in exchange for representational flexibility (cf. PerturbODE, (Lin et al., 2025)).

The implicit-differentiation backward of Section 2.3 requires only that $\partial F/\partial \mathbf{x}$ and $\partial F/\partial \theta$ be computable, either in closed form (Hill, Michaelis-Menten, mass action) or via autograd through the production term (neural drift). The shift-based loss of Equation 3 and the joint multi-condition objective are independent of the kinetic form. Choosing among parameterizations is therefore a modeling decision rather than a methodological constraint.

E. Training hyperparameters

This appendix lists the hyperparameter values used in our experiments. Values that vary by encoder or by gene-panel size are noted. The exact values used in the camera-ready experiments are stored in `configs/perturbseq.yaml` and `configs/synthetic_data.yaml`; what follows reflects the settings used throughout development and held stable across reported runs.

E.1. Optimizer

We use AdamW (Loshchilov & Hutter, 2017) with two parameter groups. The adjacency parameters (the active entries of A) receive weight decay; the basal rates $\mathbf{b}$ and the log-parameterized scalars $\log \gamma$, $\log n$, $\log K$ receive zero weight decay. This avoids drifting the kinetic scalars away from biologically reasonable values and avoids drifting $\mathbf{b}$ away from its initialization at the control mean.

Parameter	Value
Learning rate	1×10^{-3}
β_1, β_2	$0.9, 0.999$
Weight decay (adjacency parameters)	1×10^{-4}
Weight decay (other parameters)	0
Gradient clipping (max norm)	5.0

E.2. Loss weights

The total objective $\mathcal{L} = \mathcal{L}_{\text{pert}} + \lambda_{\text{ctrl}} \mathcal{L}_{\text{ctrl}} + \lambda_{\text{sparse}} \|A\|_1$ uses weights given below. The L_1 sparsity weight is held constant from epoch zero; we do not use a warmup schedule.

Weight	Value	Notes
λ_{ctrl}	1.0	Control anchor
λ_{sparse}	1×10^{-5}	Weak L_1 on A, applied from epoch 0

E.3. Fixed-point solver settings

The forward solver uses damped fixed-point iteration with the constraints reimposed at every step (Section 2.1). The backward pass uses the implicit gradient with Tikhonov regularization on the linear solve (Appendix F).

Parameter	Value	Notes
Damping factor α	0.5	In $x_{k+1} = \alpha\, g(x_k) + (1-\alpha)\, x_k$
Max iterations (cold start)	200	First 20 epochs
Max iterations (warm start)	50	Subsequent epochs
Convergence tolerance	1×10^{-5}	On $\|x_{k+1} - x_k\|_\infty$
Clamp bounds	$[10^{-6}, 50.0]$	Numerical stability of $h(x)$
Backward Tikhonov ϵ	1×10^{-8}	See Equation 30

The forward solve for each perturbation condition is warm-started from its converged state at the previous epoch. The wild-type solve uses a separate cache.

E.4. Training procedure

We split cells within each perturbation condition into train and validation subsets, with the same fraction across all conditions. Validation perturbation loss is the early-stopping criterion and the best-checkpoint metric.

Setting	Value	Notes
Train/val split	80%/20%	Stratified within each condition
Random seed	42	Used consistently for all splits
Maximum epochs	1000	
Early stopping patience	100 epochs	On validation $\mathcal{L}_{\text{pert}}$
Best-checkpoint metric	validation $\mathcal{L}_{\text{pert}}$	

For dense parameterizations under TF restriction, the TF mask is reapplied after each optimizer step to prevent AdamW's second-moment statistics from drifting masked entries off zero.

E.5. Encoder output dimensions

The encoder output dimension d is inherited from the encoder's pretraining configuration. Per-encoder values used in this work are:

Encoder	Output dimension d
Identity	G (gene-panel size)
PCA	512
scGPT (Cui et al., 2024)	512
scPRINT (Kalfon et al., 2025)	512

Encoder-internal settings (vocabulary mapping, raw-count handling, depth tokens, soft-binning temperature) are reported in Appendix G.

E.6. Compute requirements

All experiments were run on a single NVIDIA H200 NVL (141 GB HBM3e), 16 CPU cores, and approximately 320 GB of system memory. No multi-GPU training was used. Foundation model forward passes, fixed-point ODE rollouts, and adjacency optimization all fit within a single device. Encoder forward and backward passes dominate wall-clock time for the transformer encoders, because the loss requires evaluating the frozen differentiable encoder on both the predicted and target states at every gradient step.

Encoder batch sizes are set per-encoder by VRAM constraints. Identity, PCA, and scGPT run at the global batch size of 64 cells. scPRINT requires per-cell batching with batch size 1 and the internal context length restricted to `ncell`$=100$ in order to fit a forward and backward pass through the full encoder stack on the H200, due to its larger working memory footprint at the soft-binning and gene-vocabulary embedding stages. Live-encoder embeddings for control cells are cached on first use and reused across seeds within a condition.

Encoder	Cell batch size	Notes
Identity	64	No encoder forward
PCA	64	Linear projection, fixed weights
scGPT	64	Differentiable continuous-value encoder
scPRINT	1	`ncell`$=100$ to fit H200 VRAM

The headline real-data run on Norman 2019 (102 single-perturbation conditions, 129 double-perturbation conditions, 2,553 genes after Highly Variable Gene (HVG) filtering, 11,855 control cells) covered five random seeds across four encoder conditions plus the additive baseline, for a total of twenty independent ODE fits. The job completed in 7 hours 32 minutes of wall-clock time. Per-condition wall-clock aggregated across the five seeds is reported below.

Condition	Total wall-clock (5 seeds)	Mean per seed
Identity	1.11 h	13.4 min
PCA	0.50 h	6.1 min
scGPT	4.08 h	49.0 min
scPRINT	1.81 h	21.7 min

Adjacency optimization uses a dense $G \times G$ parameterization with a TF column mask that holds $260{,}406$ active entries (the 102 perturbed-TF columns) and zeros the remaining $G \times (G - G_{\mathrm{TF}})$ entries at every step (Appendix H.1). Memory cost is dominated by the encoder activations rather than by A itself.

We estimate total project compute, including the synthetic benchmark, encoder ablations, and structural validation runs, at roughly 60 to 80 H200 GPU-hours. Foundation model embedding caches and ODE checkpoints are retained on cluster scratch between runs to avoid redundant tokenization and encoder evaluation when only downstream analysis code changes.

F. Implicit-differentiation gradients: derivation and closed forms

This appendix derives the implicit-gradient form used in Section 2.3, gives closed-form expressions for the Jacobian J and the per-parameter gradients of the Hill ODE, and discusses numerical stability of the linear solve.

F.1. Implicit function theorem at the fixed point

The forward solver returns the converged steady state $\mathbf{x}^*(\theta)$ defined implicitly by $F(\mathbf{x}^*, \theta) = \mathbf{x}^* - g(\mathbf{x}^*, \theta) = \mathbf{0}$, where $g(\mathbf{x}, \theta) = (\mathbf{b} + A\, h(\mathbf{x}; n, K))/\gamma$ is the right-hand side of Equation 2. Differentiating the implicit equation with respect to θ yields

$$\left.\frac{\partial F}{\partial \mathbf{x}}\right|_{\mathbf{x}^*} \frac{\partial \mathbf{x}^*}{\partial \theta} + \left.\frac{\partial F}{\partial \theta}\right|_{\mathbf{x}^*} = \mathbf{0}, \tag{13}$$

so that

$$\frac{\partial \mathbf{x}^*}{\partial \theta} = -J^{-1} \left.\frac{\partial F}{\partial \theta}\right|_{\mathbf{x}^*}, \qquad J = \left.\frac{\partial F}{\partial \mathbf{x}}\right|_{\mathbf{x}^*}. \tag{14}$$

For a scalar loss $\mathcal{L}$ that depends on θ only through $\mathbf{x}^*$, with upstream gradient $\mathbf{v} = \partial \mathcal{L}/\partial \mathbf{x}^*$, the parameter gradient is

$$\frac{\partial \mathcal{L}}{\partial \theta} = \mathbf{v}^\top \frac{\partial \mathbf{x}^*}{\partial \theta} = -\mathbf{v}^\top J^{-1} \frac{\partial F}{\partial \theta} = -\mathbf{w}^\top \frac{\partial F}{\partial \theta}, \tag{15}$$

where $\mathbf{w} = J^{-\top}\mathbf{v}$ is obtained by solving the linear system

$$J^\top \mathbf{w} = \mathbf{v}. \tag{16}$$

This is a vector-Jacobian product. One $G \times G$ linear solve provides the backward signal for all parameter groups simultaneously, after which the per-parameter gradient is read off from the analytic form of $\partial F/\partial \theta$.

F.2. Closed-form Jacobian

For the Hill production term used in this work, the components of F are

$$F_i(\mathbf{x}, \theta) = x_i - \frac{1}{\gamma}\left(b_i + \sum_k A_{ik}\, h(x_k;\, n,\, K)\right). \tag{17}$$

Differentiating with respect to x_l and using $\partial h(x_k)/\partial x_l = \delta_{kl}\, h'(x_l)$,

$$J_{il} = \frac{\partial F_i}{\partial x_l} = \delta_{il} - \frac{A_{il}\, h'(x_l)}{\gamma}, \tag{18}$$

or equivalently in matrix form

$$J = I - \frac{1}{\gamma} A \,\mathrm{diag}\big(h'(\mathbf{x}^*)\big). \tag{19}$$

The construction of J requires only the Hill derivative h' at $\mathbf{x}^*$ and the current adjacency A, both available without additional computation at backward time.

F.3. Closed-form Hill derivatives

The Hill function $h(x;\, n,\, K) = x^n/(K^n + x^n)$ admits closed-form derivatives with respect to its argument and its parameters. Direct differentiation yields

$$h'(x) = \frac{n\, K^n\, x^{n-1}}{(K^n + x^n)^2} = \frac{n}{x}\, h(x)\big(1 - h(x)\big), \tag{20}$$

$$\frac{\partial h}{\partial n} = \frac{K^n\, x^n\, \ln(x/K)}{(K^n + x^n)^2} = h(x)\big(1 - h(x)\big)\, \ln(x/K), \tag{21}$$

$$\frac{\partial h}{\partial K} = -\frac{n\, K^{n-1}\, x^n}{(K^n + x^n)^2} = -\frac{n}{K}\, h(x)\big(1 - h(x)\big). \tag{22}$$

The right-hand-side identities are convenient at runtime: the saturation factor $h(x)(1 - h(x))$ is computed once and reused across all three derivative quantities.

F.4. Per-parameter gradients

Each parameter gradient follows from the VJP formula $\partial\mathcal{L}/\partial\theta = -\mathbf{w}^\top \partial F/\partial\theta$ with $\mathbf{w}$ obtained from Equation 16. The relevant partial derivatives of F are

$$\frac{\partial F_i}{\partial A_{kl}} = -\frac{\delta_{ik}\, h(x_l)}{\gamma}, \qquad \frac{\partial F_i}{\partial b_k} = -\frac{\delta_{ik}}{\gamma}, \tag{23}$$

$$\left.\frac{\partial F_i}{\partial \gamma}\right|_{\mathbf{x}^*} = \frac{x_i^*}{\gamma}, \qquad \frac{\partial F_i}{\partial \alpha} = -\frac{1}{\gamma}\sum_k A_{ik}\,\frac{\partial h(x_k)}{\partial \alpha} \quad \text{for} \quad \alpha \in \{n, K\}, \tag{24}$$

where the γ identity uses the fixed-point relation $b_i + \sum_k A_{ik}\, h(x_k^*) = \gamma\, x_i^*$ to simplify. Substituting into the VJP formula and simplifying yields the closed-form parameter gradients

$$\frac{\partial\mathcal{L}}{\partial A_{ij}} = \frac{w_i\, h(x_j^*)}{\gamma}, \tag{25}$$

$$\frac{\partial\mathcal{L}}{\partial b_i} = \frac{w_i}{\gamma}, \tag{26}$$

$$\frac{\partial\mathcal{L}}{\partial \gamma} = -\frac{1}{\gamma}\sum_i w_i\, x_i^*, \tag{27}$$

$$\frac{\partial\mathcal{L}}{\partial n} = \frac{1}{\gamma}\,\mathbf{w}^\top A\left.\frac{\partial \mathbf{h}}{\partial n}\right|_{\mathbf{x}^*}, \tag{28}$$

$$\frac{\partial\mathcal{L}}{\partial K} = \frac{1}{\gamma}\,\mathbf{w}^\top A\left.\frac{\partial \mathbf{h}}{\partial K}\right|_{\mathbf{x}^*}, \tag{29}$$

where $\partial\mathbf{h}/\partial n$ and $\partial\mathbf{h}/\partial K$ are vectors whose k-th entries are given by Equations 21 and 22 evaluated at x_k^*.

Clamped genes constraints on gradients. Clamped genes have $x_j^{*,(j)} = c_j$ independent of θ, so their contributions to all parameter gradients should vanish. As described in Section 2.3, this is enforced on the gradient side by zeroing the corresponding entry of $\mathbf{v}$ and replacing the j-th row of J with $\mathbf{e}_j^\top$ before solving Equation 16, which forces $w_j = 0$. With $w_j = 0$, the contributions of clamped indices to Equations 25-29 vanish automatically, and no further masking of the analytic VJPs is required.

F.5. Numerical stability of the linear solve

The system in Equation 16 is well-posed when J is non-singular. By Equation 19, $J = I - (A\,\mathrm{diag}(h'))/\gamma$, and J approaches singularity precisely when the spectral radius of $(A\,\mathrm{diag}(h'))/\gamma$ approaches one, which is the same condition under which the forward fixed-point iteration becomes slow or marginally stable. In our experiments J remains well-conditioned throughout training, but to guard against numerical instability we solve a Tikhonov-regularized system

$$(J^\top + \epsilon I)\,\mathbf{w} = \mathbf{v}, \qquad \epsilon = 10^{-8}, \tag{30}$$

which adds a perturbation that vanishes in the well-conditioned limit and bounds the gradient norm when J is near-singular. The introduced bias is negligible at this ϵ scale relative to the floating-point precision of the inputs. Persistent ill-conditioning indicates that the forward solver is operating near the stability boundary, in which case the appropriate response is to revisit the solver damping factor or iteration depth rather than to increase ϵ.

G. Encoder wrapper details

This appendix documents the per-encoder implementation details for the differentiable wrappers used in our experiments. All wrappers conform to the abstract interface Enc : $\mathbb{R}^G \to \mathbb{R}^d$ described in Section 2.2, return embeddings differentiable in their input expression vector, and pass a common suite of diagnostic checks documented at the end of this appendix.

G.1. Identity and PCA

The Identity encoder returns its input unchanged, with $d = G$. Differentiability is trivial. The shift-based loss in this case reduces to gene-space matching of perturbation shifts.

The PCA encoder is fit on control cells from the data being analyzed and projects onto the top d principal components, $\mathrm{Enc}(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})\, V$, where $\boldsymbol{\mu}$ is the control mean and $V \in \mathbb{R}^{G \times d}$ collects the top right singular vectors of the centered control matrix. We use $d = 512$ throughout, which captures the bulk of expression variance while leaving a sufficient compression to make PCA a meaningful baseline against the foundation-model encoders.

G.2. scGPT

We use scGPT as a frozen wrapper that maps an expression vector to a cell-level embedding via the published transformer architecture and pretrained weights. Two design points are important for differentiability and correctness.

Continuous value encoder. scGPT supports both continuous and categorical encodings of expression values. Categorical encoding bins values into discrete tokens and is non-differentiable in the input. We require checkpoints trained with the continuous variant, in which expression values pass through a learned MLP before being added to gene-token embeddings. The wrapper asserts this at initialization and refuses checkpoints that use the categorical path.

Gene vocabulary mapping. scGPT operates on a fixed gene vocabulary tied to the pretrained checkpoint. We map each input gene name in the dataset to its corresponding vocabulary index via case-aware matching (exact match preferred, then uppercase, lowercase, capitalized variants). Genes that fail to map are excluded from the input sequence rather than mapped to a placeholder token.

Cell embedding extraction. The wrapper returns the cell embedding from the model's CLS token.

G.3. scPRINT

We use scPRINT as a frozen wrapper. Its interface differs from scGPT in two important ways.

Raw counts and depth token. scPRINT applies its own internal $\log_2(1 + \text{count})$ transformation and conditions on a depth token derived from the per-cell library size. The wrapper therefore expects raw integer counts, not log-normalized expression, and we pass the full-transcriptome library size (sum over all genes prior to highly-variable-gene filtering) as the depth token. Failing to provide the full-transcriptome library size, for example by passing the sum over the highly-variable subset, produces a depth token that systematically underestimates the cell's true sequencing depth and degrades the embedding quality.

Gene identifiers. scPRINT's vocabulary indexes genes by Ensembl ID. The wrapper reads the dataset's Ensembl IDs directly when available and falls back to symbol-to-Ensembl lookup via mygene only for genes without an Ensembl annotation. Direct vocabulary lookup is faster and avoids ambiguity introduced by symbol aliases.

Numerical safety. scPRINT's internal $\log_2(1 + x)$ transformation produces `NaN` on negative inputs. Because Equation 3 fits a steady-state predictor whose intermediate states can transiently violate non-negativity during optimization, we clamp predicted expression to $x \geq 0$ before passing to scPRINT. For real-data raw-count inputs the clamp is a no-op.

G.4. Embedding caching

For encoders that perform substantial compute on each forward pass (the foundation models), we cache encoded observations to disk and reuse them across runs. The cache is keyed by a fingerprint hash that includes the encoder identity and wrapper

version, the dataset and gene-panel hash, the input format flag (raw counts versus log-normalized), and the wrapper's vocabulary mapping. Any change that alters the encoded output invalidates the cache, including code changes to the wrapper itself.

G.5. Diagnostic validation

Each wrapper is validated against a common suite of checks: gradient flow (autograd produces non-zero gradients with respect to a perturbed input), perturbation sensitivity (encoded shifts have non-trivial magnitude under realistic input perturbations), discriminability (encoded shifts under different perturbations are distinguishable in cosine similarity), and reference-result reproducibility (encoded outputs match published embeddings on a small held-out set, where applicable). Wrappers that fail any check are not used in experiments.

H. Adjacency parameterization details

This appendix gives details for the dense parameterization of the regulatory adjacency matrix A introduced in Section 2.4, including the transcription-factor restriction and the initialization scheme.

H.1. Transcription-factor restriction

We restrict the recovered adjacency $A \in \mathbb{R}^{G \times G_{\mathrm{TF}}}$ so that only the perturbed transcription factors are admissible as regulators, fitting edges only for TFs for which we have interventional evidence. This reflects the identifiability constraint that a column of A can only be directly constrained by perturbing the corresponding regulator: TFs whose perturbations are not observed contribute no direct interventional signal and would have their columns determined only through indirect downstream effects.

Because we use a dense parameterization for A, this restriction is implemented as a strict multiplicative mask on the parameter matrix. Entries corresponding to non-perturbed columns are set to exactly zero at initialization. Crucially, this mask is reapplied and the non-perturbed entries are zeroed again after every optimizer step. Reapplication is necessary because we optimize the network using AdamW (Loshchilov & Hutter, 2017): even if the gradient for a masked entry is zero, AdamW's second-moment running statistics for that parameter can become non-zero by virtue of past momentum during the run. Without continuous reapplication of the mask, these optimizer statistics would cause the masked, non-perturbed entries to slowly drift away from zero over time.

H.2. Initialization scheme

The active, learnable entries of the dense adjacency matrix A (corresponding to the G_{TF} columns) are initialized by sampling independently from a zero-mean Gaussian distribution, $A_{ij} \sim \mathcal{N}(0, \sigma^2)$. The standard deviation σ is kept small to calibrate the expected magnitude of the initial regulatory inputs. This ensures that the initial interaction weights provide a stable, well-conditioned starting point for the forward fixed-point solver, preventing the Hill activation functions from being driven into hard saturation prematurely during the earliest epochs of training. All non-TF columns remain strictly zero throughout initialization and training.

I. Structural validation details

This appendix reports the full structural validation results summarized in Figure 3 of the main text. The recovered adjacency for the Identity encoder is tested against three independent reference types: curated hematopoietic lineage signatures, ChEA 2022 ChIP-seq target sets, and ENCODE TF ChIP-seq 2015 target sets. Pathway enrichment against MSigDB Hallmark (Liberzon et al., 2015) and Reactome 2022 (Jassal et al., 2020) is also reported as a fourth tier for completeness.

For each TF, we take the top-25 recovered targets ranked by $|A[:, \mathrm{TF}]|$ (the same lists shown in Figure 3 panel C), compute hypergeometric enrichment against every gene set in each library after restricting to the gene panel, and apply Bonferroni correction over the size of each library. Background size is the panel target count ($n = 2553$ measured genes). Self-loops are excluded from the top-25 ranking.

I.1. Lineage signature enrichment

The lineage signature library (Appendix J) consists of eight curated hematopoietic gene sets: three matched-lineage signatures (Erythroid, Myeloid Monocyte, Interferon Response) and five negative or alternative controls (Myeloid Granulocyte, Megakaryocyte, T cell, B cell, Cell cycle).

Table 3. Fold enrichment and Bonferroni-corrected hypergeometric p-values of top-25 recovered targets against each lineage signature, per TF. Bold entries highlight the matched-lineage hit per TF. $m = 8$ for Bonferroni.

TF	Lineage signature	Overlap	Fold	p_{bonf}
KLF1	**Erythroid Lineage**	**10/21**	**48.6×**	**2.1e-15**
KLF1	Lymphoid T cell Lineage	1/13	7.9×	7.2e-1
KLF1	Myeloid Granulocyte Lineage	0/13	0.0×	1.0
KLF1	Myeloid Monocyte Lineage	0/23	0.0×	1.0
KLF1	Interferon Response	0/21	0.0×	1.0
KLF1	Megakaryocyte Lineage	0/–	0.0×	1.0
KLF1	Lymphoid B cell Lineage	0/–	0.0×	1.0
KLF1	Cell Cycle	0/–	0.0×	1.0
SPI1	**Myeloid Monocyte Lineage**	**10/23**	**44.4×**	**6.6e-15**
SPI1	Myeloid Granulocyte Lineage	1/13	7.9×	7.2e-1
SPI1	Erythroid Lineage	1/21	4.9×	1.0
SPI1	Interferon Response	1/21	4.9×	1.0
SPI1	Lymphoid T cell Lineage	0/13	0.0×	1.0
SPI1	Megakaryocyte Lineage	0/–	0.0×	1.0
SPI1	Lymphoid B cell Lineage	0/–	0.0×	1.0
SPI1	Cell Cycle	0/–	0.0×	1.0
IRF1	**Interferon Response**	**8/21**	**38.9×**	**2.8e-11**
IRF1	Erythroid Lineage	2/21	9.7×	1.0e-1
IRF1	Myeloid Granulocyte Lineage	0/13	0.0×	1.0
IRF1	Myeloid Monocyte Lineage	0/23	0.0×	1.0
IRF1	Lymphoid T cell Lineage	0/13	0.0×	1.0
IRF1	Megakaryocyte Lineage	0/–	0.0×	1.0
IRF1	Lymphoid B cell Lineage	0/–	0.0×	1.0
IRF1	Cell Cycle	0/–	0.0×	1.0

Each TF's matched-lineage hit reaches $p_{\mathrm{bonf}} < 10^{-10}$ with fold enrichment between $38.9\times$ and $48.6\times$. Every wrong-lineage and negative-control signature has $p_{\mathrm{bonf}} > 0.1$. The closest off-diagonal hit is IRF1 against Erythroid Lineage at $9.7\times$ ($p_{\mathrm{bonf}} = 0.10$), consistent with the K562 hematopoietic context where interferon-responsive genes overlap with erythroid co-regulated genes (notably *HEMGN* and *GYPC*, which appear in both the IRF1 top-25 and the broader erythroid program).

I.2. ChEA 2022 ChIP-Seq target enrichment

ChEA 2022 contains 1546 gene sets, each representing the targets of one TF in one ChIP-Seq experiment with cell-line and reference annotated. Bonferroni correction uses $m = 685$ (size after restricting sets to the gene panel and applying minimum size threshold of 5; full library spans 1546 entries before panel restriction).

Table 4. Top 5 ChEA hits per TF for top-25 recovered targets. Cell line and species are part of the gene set name. Same-TF entries are bold.

TF	ChEA hit (top 5)	Overlap	Fold	p_{bonf}
KLF1	**EKLF 21900194 ChIP-Seq erythrocyte (mouse)**	**8/107**	**7.6×**	**3.0e-3**
KLF1	GATA2 19941826 ChIP-Seq K562 (human)	9/238	3.9×	1.7e-1
KLF1	GATA1 19941826 ChIP-Seq K562 (human)	6/110	5.6×	3.5e-1
KLF1	**KLF1 20508144 ChIP-Seq fetal liver erythroid (mouse)**	**6/112**	**5.5×**	**3.8e-1**
KLF1	CLOCK 20551151 ChIP-Seq 293T (human)	5/73	7.0×	3.9e-1
SPI1	**SPI1 23547873 ChIP-Seq NB4 (human)**	**16/377**	**4.3×**	**1.5e-5**
SPI1	TAL1 20887958 ChIP-Seq HPC-7 (mouse)	9/191	4.8×	3.0e-2
SPI1	RUNX1 20887958 ChIP-Seq HPC-7 (mouse)	7/123	5.8×	8.1e-2
SPI1	KLF6 26769127 ChIP-Seq PDAC cell line (human)	8/174	4.7×	1.1e-1
SPI1	LYL1 20887958 ChIP-Seq HPC-7 (mouse)	6/89	6.9×	1.1e-1
IRF1	IRF8 22096565 ChIP-ChIP GC-B (human)	7/32	22.3×	6.8e-6
IRF1	**IRF1 21803131 ChIP-Seq monocytes (human)**	**6/35**	**17.5×**	**4.3e-4**
IRF1	MAF 26560356 ChIP-Seq Th1 (human)	11/231	4.9×	2.6e-3
IRF1	SA1 27219007 ChIP-Seq B cells (human)	9/200	4.6×	4.3e-2
IRF1	UTX 26944678 ChIP-Seq Jurkat (human)	9/219	4.2×	8.9e-2

Each TF's top-25 recovered targets enrich significantly ($p_{\mathrm{bonf}} < 10^{-3}$) against same-TF ChIP target sets in ChEA. The IRF1 top hit is IRF8, the closest IRF-family paralog that shares many target genes with IRF1; the IRF1 same-TF entry is the second-strongest. The KLF1 top hit uses the gene synonym EKLF (Erythroid Krüppel-Like Factor), which is the same gene as KLF1.

The next-strongest hits in each TF's table reflect the local regulatory neighborhood. KLF1's recovered targets enrich against GATA2 and GATA1 ChIP in K562, the canonical erythroid co-regulators that share targets with KLF1 in erythroid commitment. SPI1's top targets enrich against TAL1, RUNX1, and LYL1 ChIP in mouse hematopoietic progenitors, which together with SPI1 form the heptad transcription factor complex characterized in early myeloid specification (Wilson et al., 2010). IRF1 enriches against MAF in Th1 cells, consistent with the IRF1-MAF interplay in interferon-responsive immune programs.

I.3. ENCODE TF ChIP-Seq 2015 target enrichment

ENCODE TF ChIP-Seq 2015 contains 816 gene sets, each representing one TF binding event in one cell line under the ENCODE consortium. K562 is the most-represented human cell line in this library, making it directly relevant to the Norman 2019 dataset.

Table 5. Top 5 ENCODE hits per TF for top-25 recovered targets. Same-TF entries are bold.

TF	ENCODE hit (top 5)	Overlap	Fold	p_{bonf}
KLF1	FOSL1 K562 hg19	7/92	7.8×	1.5e-2
KLF1	GATA1 erythroblast hg19	9/177	5.2×	1.9e-2
KLF1	TAL1 G1E-ER4 mm9	6/80	7.7×	7.0e-2
KLF1	TAL1 erythroblast mm9	6/81	7.6×	7.5e-2
KLF1	ZNF274 K562 hg19	6/88	7.0×	1.2e-1
SPI1	TFAP2C HeLa-S3 hg19	7/145	4.9×	2.7e-1
SPI1	**SPI1 GM12878 hg19**	**8/215**	**3.8×**	**2.55e-2**
SPI1	EP300 CH12.LX mm9	8/258	3.2×	1.0
SPI1	**SPI1 HL-60 hg19**	**6/148**	**4.1×**	**1.0**
SPI1	ZNF384 ES-E14 mm9	6/150	4.1×	1.0
IRF1	RCOR1 CH12.LX mm9	10/143	7.1×	2.9e-4
IRF1	MYC CH12.LX mm9	9/112	8.2×	4.0e-4
IRF1	**IRF1 K562 hg19**	**13/291**	**4.6×**	**5.0e-4**
IRF1	POLR2A K562 hg19	14/420	3.4×	5.5e-3
IRF1	STAT2 K562 hg19	5/36	14.2×	1.5e-2

KLF1 and IRF1 enrich strongly against ENCODE entries with K562 specificity. KLF1's top ENCODE hits reflect erythroid co-regulation: FOSL1 in K562, GATA1 in erythroblast, and TAL1 in two erythroid contexts (G1E-ER4 is a GATA1-rescuable erythroid line, erythroblast is primary cell). IRF1's IRF1 K562 entry recovers 13 of 291 ENCODE-defined IRF1 targets in the matched cell line at $p_{\text{bonf}} = 5.0 \times 10^{-4}$. The IRF1 STAT2 K562 hit ($14.2\times$ fold) reflects the canonical IRF1-STAT2 axis in type I interferon signaling.

SPI1 ENCODE enrichment is weaker and does not pass Bonferroni correction at the 0.05 level. SPI1 same-TF entries (GM12878, HL-60) do appear in the top 5 but the available ENCODE SPI1 ChIP experiments are in cell lines (lymphoblast and APL respectively) less directly matched to K562 than the ChEA SPI1 NB4 entry (Table 4). The SPI1 lineage signature enrichment (Table 3, $44.4\times$, $p_{\text{bonf}} = 6.6 \times 10^{-15}$) is the strongest evidence of SPI1 recovery.

I.4. MSigDB Hallmark and Reactome enrichment

For completeness, we report top hits against the broader pathway databases used as initial validation references. Bonferroni correction over $m = 49$ (Hallmark, after panel restriction) and $m = 575$ (Reactome 2022, after panel restriction and exclusion of disease categories per Appendix J).

Table 6. Top hit per TF in MSigDB Hallmark and Reactome 2022.

TF	Library	Top hit	Overlap	Fold	p_{bonf}
KLF1	Hallmark	Heme Metabolism	9/42	21.9×	3.0e-9
KLF1	Reactome	Metabolism of Porphyrins	2/5	40.8×	5.2e-1
SPI1	Hallmark	(none significant at top-25, $m = 49$)	–	–	n.s.
SPI1	Reactome	Signaling by Rho GTPases	5/66	7.7×	2.0e-1
IRF1	Hallmark	Interferon Gamma Response	8/60	13.6×	2.2e-6
IRF1	Reactome	ER-Phagosome Pathway	5/23	22.2×	1.0e-3

KLF1's Hallmark Heme Metabolism hit recovers 9 of 42 panel-resident heme metabolism genes in the top-25, which is consistent with the lineage signature result (10/21 erythroid hit, Table 3). IRF1's Hallmark Interferon Gamma Response enrichment is the expected primary signal. SPI1 does not reach Bonferroni significance in either Hallmark or Reactome at top-25, because neither library contains a coarse-grained myeloid lineage category that captures the recovered target program. This finding motivated the curated lineage signature library used as the primary structural validation in Figure 3 of the main text.

I.5. TRRUST literature-curated reference

TRRUST v2 contains 9396 human TF-target pairs curated from literature (Han et al., 2018). After restriction to the gene panel and the perturbed TF set, TRRUST annotates 16 IRF1 targets, 23 SPI1 targets, and 1 KLF1 target. Recovered top-25 overlap with TRRUST: IRF1 3/25 (*PSMB9*, *TAPBP*, *TAP1*, with *PSMB9* and *TAP1* marked as activations and *TAPBP* as unsigned), SPI1 0/25, KLF1 0/25.

The IRF1 3/25 TRRUST overlap is statistically significant against the random baseline of 16/2553 expected overlap ($p_{\text{hypergeom}} < 10^{-3}$). The SPI1 and KLF1 TRRUST results are non-significant, but the recovered targets are strongly biologically interpretable in the lineage signature, ChEA, and ENCODE references (Tables 3–5). This pattern reflects TRRUST's known sparsity for K562-context lineage regulation rather than failure of the recovered adjacency: KLF1 has only one TRRUST-annotated target in the panel, and SPI1's TRRUST annotations weight toward immune contexts not well-captured in K562 erythromyeloid biology.

We retain TRRUST as a tertiary validation tier because it provides a positive control for IRF1 (a well-curated TF in TRRUST) and because the TRRUST annotation shown in Figure 3 panel A (TRRUST-annotated targets highlighted with colored fill) helps reviewers familiar with literature-curated GRNs orient to the recovered subnetworks.

I.6. Co-regulator recovery and the local regulatory neighborhood

A consistent pattern across Tables 4 and 5 is that each TF's recovered top-25 enriches not only against its own ChIP targets but against the ChIP targets of its known co-regulators in the matched cellular context.

For KLF1, the co-regulator hits are GATA1, GATA2, FOSL1, and TAL1, all canonical erythroid TFs that share regulatory

targets with KLF1 during erythroid maturation. GATA1 in particular is the master erythroid TF upstream of KLF1, and the recovered KLF1 adjacency captures targets that GATA1 binds in K562 chromatin.

For SPI1, the co-regulator hits in ChEA mouse hematopoietic progenitors are TAL1, RUNX1, and LYL1, three of the seven members of the heptad transcription factor complex (FLI1, ERG, GATA2, RUNX1, LYL1, LMO2, TAL1) that co-regulates myeloid commitment. The SPI1 top-25 recovers shared heptad targets even though our model fits SPI1 individually.

For IRF1, the co-regulator hit is STAT2 in K562 (Table 5, $14.2\times$ fold), reflecting the IRF1-STAT2 axis in canonical interferon signaling, plus IRF8 in B cells (Table 4, $22.3\times$ fold) reflecting the IRF1-IRF8 paralog target overlap.

This local-regulatory-neighborhood pattern is consistent with the single shared regulatory adjacency formulation (Section 2.2 of the main text): the model fits one A across all conditions, and the implicit constraint that targets co-regulated by a related TF should retain similar A entries propagates information beyond the directly perturbed TF. We do not claim the framework explicitly recovers co-regulator structure, but the structural validation tables show that the recovered targets respect known co-regulatory groupings.

I.7. Methodological notes

Multiple-testing correction. Bonferroni correction is applied within each library independently. We do not jointly correct across libraries because they are independent reference types being reported separately, not a single multiple-comparison test. Within-library correction uses the panel-restricted library size (after minimum-size and disease/generic filtering for Hallmark and Reactome; no filtering applied to ChEA, ENCODE, or lineage signatures, where pathway names do not match the disease/generic substring lists).

Background gene set. All hypergeometric tests use the gene panel ($n = 2553$ measured genes) as the background, not the full transcriptome. This is the conditional null appropriate for enrichment tests where pathway membership outside the panel cannot affect ranking.

Self-loops. Self-loops (the perturbed gene's own row in its column of A) are excluded from the top-25 ranking before enrichment, because the perturbed gene is clamped to its observed value and does not represent a learned regulatory edge.

J. Lineage signature library and pathway library filtering

This appendix documents the curated hematopoietic lineage signature library used as the primary structural validation reference in Section 3.2, Figure 3 and Table 3 and the denylist used to filter disease-context and generic catch-all pathways from MSigDB Hallmark and Reactome 2022 before enrichment testing. Both are referenced from Appendix I.

J.1. Lineage signature library

The lineage signature library consists of eight curated hematopoietic gene sets, three of which are matched-lineage signatures expected to enrich for the master transcription factors evaluated in Section 3.2 and five of which are negative or alternative controls. The sets are intentionally compact (20–47 genes each) to focus on canonical lineage markers rather than broader expression programs, since the goal is structural validation against the most strongly lineage-defining genes rather than against the full transcriptional response of each lineage.

The eight signatures and their biological role in the validation are as follows.

ERYTHROID_LINEAGE (44 genes). Primary expected hit for KLF1. Includes the human globin cluster (HBA1, HBA2, HBB, HBG1, HBG2, HBD, HBE1, HBM, HBQ1, HBZ), heme biosynthesis enzymes (ALAS2, FECH, HMBS, ALAD, UROD, UROS, CPOX, PPOX), iron metabolism and heme catabolism genes (SLC25A37, TFRC, FTH1, FTL, BLVRB, HMOX1), erythroid transcription factors (KLF1, GATA1, NFE2, TAL1, LDB1, LMO2, EPOR), red blood cell membrane and cytoskeletal proteins (GYPA, GYPB, GYPC, EPB42, EPB41, SPTA1, SPTB, ANK1, SLC4A1, RHCE, RHD), and erythroid maturation markers (AHSP, HEMGN).

MYELOID_GRANULOCYTE_LINEAGE (34 genes). Alternative myeloid signature for SPI1. Includes primary granule proteases and antimicrobials (ELANE, MPO, PRTN3, CTSG, AZU1, BPI), secondary granule contents (LCN2, LTF, CAMP, MMP8, MMP9), neutrophil-enriched calcium-binding proteins (S100A8, S100A9, S100A12),

defensins (DEFA1, DEFA3, DEFA4), granulocyte receptors and surface markers (CSF3R, FCGR3B, CEACAM3, CEACAM8, FPR1, FPR2, ITGAM), the NADPH oxidase complex (CYBA, CYBB, NCF1, NCF2, NCF4), granulocytic transcription factors (CEBPA, CEBPE, GFI1), and other markers (ALOX5, PGLYRP1).

MYELOID_MONOCYTE_LINEAGE (39 genes). Primary expected hit for SPI1. Includes canonical monocyte and macrophage markers (CD14, CD68, CD163, CD86, MRC1, MARCO, MERTK), Fc receptors (FCGR1A, FCGR2A, FCGR2B, FCGR3A), macrophage and dendritic cell transcription factors (SPI1, IRF8, MAFB, KLF4), lysozyme and lysosomal enzymes (LYZ, CTSS, CTSB, CTSD, LAPTM5), the CSF1 axis (CSF1R, CSF2RA), monocyte-enriched cytoskeletal proteins (AIF1, LST1, TYROBP, CORO1A, ARPC1B, ARHGDIB), complement components (CFD, C1QA, C1QB, C1QC), inflammation effectors (TLR2, TLR4, NLRP3, FCN1), and other myeloid markers (CD33, VAMP8, PYCARD).

MEGAKARYOCYTE_LINEAGE (22 genes). Discriminative control. Shares GATA1 and NFE2 with the erythroid lineage, so a recovered KLF1 adjacency that reflects erythroid commitment specifically (rather than the generic erythroid–megakaryocyte progenitor program) should hit ERYTHROID_LINEAGE more strongly than MEGAKARYOCYTE_LINEAGE. Includes platelet integrins and glycoproteins (ITGA2B, ITGB3, GP1BA, GP1BB, GP9, GP5, GP6), alpha granule contents (PF4, PPBP, VWF, FGA, FGB), the thrombopoietin receptor (MPL), megakaryocytic transcription factors (FLI1, RUNX1, MEIS1), P-selectin and adhesion (SELP, TLN1), shared erythroid–megakaryocyte transcription factors (GATA1, NFE2), and other markers (TUBB1, PROS1).

LYMPHOID_T_CELL_LINEAGE (33 genes). Negative control. K562 lacks T cell lineage commitment and no enrichment is expected for any of KLF1, SPI1, or IRF1. Includes the T cell receptor complex (CD3D, CD3E, CD3G, CD247, TRAC, TRBC1, TRBC2), coreceptors and accessory molecules (CD2, CD4, CD7, CD8A, CD8B, CD27, CD28), T cell transcription factors (TCF7, LEF1, TBX21, GATA3, FOXP3, RUNX3, BCL11B), effector molecules (GZMA, GZMB, GZMK, PRF1, IFNG), interleukin receptors (IL7R, IL2RA, IL2RB), and T-specific signaling components (ZAP70, LCK, ITK, THEMIS).

LYMPHOID_B_CELL_LINEAGE (28 genes). Negative control. Includes the B cell receptor complex (CD19, MS4A1, CD22, CD79A, CD79B), immunoglobulin chains (IGHM, IGHD, IGLL1, IGKC), B cell transcription factors (PAX5, EBF1, POU2AF1, TCF3, BCL6), plasma cell terminal differentiation transcription factors (PRDM1, XBP1, IRF4), surface and signaling molecules (CD72, BANK1, BLNK, MZB1, TNFRSF13B, TNFRSF13C), and other B cell markers (VPREB1, VPREB3, FCRL1, FCRL2, FCRL5).

INTERFERON_RESPONSE (47 genes). Primary expected hit for IRF1. Combines type I and type II interferon-stimulated genes into a single signature to avoid splitting power across two highly overlapping sets. Includes canonical interferon-stimulated genes (ISG15, ISG20, IFIT1, IFIT2, IFIT3, IFIT5), the OAS family (OAS1, OAS2, OAS3, OASL), MX dynamins (MX1, MX2), antiviral IFITM proteins (IFITM1, IFITM2, IFITM3), other ISGs (RSAD2, IFI6, IFI27, IFI44, IFI44L, USP18, BST2, HERC5, HERC6, EIF2AK2), the GBP family (GBP1, GBP2, GBP4, GBP5), JAK–STAT signaling components (STAT1, STAT2, IRF1, IRF7, IRF9), antigen presentation machinery (HLA-A, HLA-B, HLA-C, TAP1, TAP2, TAPBP, PSMB8, PSMB9, PSME1, PSME2), and type II effectors (CIITA, IDO1, WARS1).

CELL_CYCLE (38 genes). Negative control. No enrichment is expected for KLF1, SPI1, or IRF1 if the recovered adjacency captures lineage-specific regulation rather than generic proliferation. Includes cyclins (CCNA1, CCNA2, CCNB1, CCNB2, CCND1, CCND2, CCND3, CCNE1, CCNE2), cyclin-dependent kinases (CDK1, CDK2, CDK4, CDK6, CDK7), CDK inhibitors (CDKN1A, CDKN1B, CDKN2A, CDKN2B), proliferation markers (MKI67, PCNA, TOP2A), replication licensing factors (MCM2–MCM7), spindle and mitotic regulators (AURKA, AURKB, PLK1, BUB1, BUB1B, MAD2L1), and E2F transcription factors with RB1 and TFDP1.

Sources used to curate the gene lists are in (Hay et al., 2018), (Novershtern et al., 2011), the MSigDB C8 cell type signatures collection, (Schoenfelder & Fraser, 2019) for erythroid biology, and (Tenen, 2003) for myeloid transcription factor circuitry. The total number of unique genes across all eight signatures is 283. The library is provided as a tab-separated file in Enrichr format (one set per line: `<set_name>\t\t<gene1>\t<gene2>\t...`) so that it is consumed directly by the same enrichment infrastructure used for MSigDB Hallmark and Reactome 2022. Bonferroni multiple-testing correction over $m = 8$ provides modest correction with substantial statistical power for the matched-lineage hits expected to dominate.

J.2. Pathway library filtering for Hallmark and Reactome

For the MSigDB Hallmark 2020 and Reactome 2022 enrichment runs reported in Appendix I (Table 6), we apply a denylist filter before testing to exclude two categories of pathways that are statistically real annotations but biologically inappropriate for evaluating a regulatory network recovered from K562 perturbation data.

The first category is disease and pathogen response pathways. These pathways describe the cellular response to infection or pathology rather than physiological regulation, and pathways such as "Tuberculosis Infection Response" or "Influenza Infection" are statistically enrichable for any general inflammatory or interferon program but do not describe normal regulatory biology. We exclude pathways whose names contain any of the following substrings (case-insensitive): `infection`, `infectious`, `disease`, `diseases`, `tuberculosis`, `mtb`, `mycobacterium`, `hiv`, `influenza`, `sars`, `coronavirus`, `covid`, `hepatitis`, `leishmania`, `malaria`, `plasmodium`, `carcinoma`, `neoplasm`, `leukem`, `tumor`, `tumour`, `cancer`, `syndrome`, `disorder`.

The second category is generic catch-all pathways at the top of the Reactome hierarchy. These pathways are broad superset annotations such as "Generic Transcription Pathway" or "Signal Transduction" that aggregate large numbers of genes across many specific functions. They are consistently enriched in any sufficiently broad gene list and provide no biological discrimination. We exclude pathways whose names match the following exact strings: `Generic Transcription Pathway`, `Developmental Biology`, `Signal Transduction`, `Metabolism`, `Metabolism of proteins`, `Metabolism of RNA`, `Gene expression (Transcription)`, `Gene expression`, `RNA Polymerase II Transcription`, `Cellular responses to stimuli`, `Cellular responses to stress`, `Cellular responses to external stimuli`, `Hemostasis`, `Programmed Cell Death`, `Cell Cycle`, `Cell Cycle, Mitotic`, `Mitotic Cell Cycle`, `Vesicle-mediated transport`, `Membrane Trafficking`, `Post-translational protein modification`.

The exact-name match is applied after stripping the trailing Reactome stable identifier (e.g., `R-HSA-NNNNNN`) that Enrichr appends to Reactome pathway names. We do not pattern-match on the substring `Metabolism`, because specific subpathways such as "Heme Metabolism" or "Metabolism of Lipids" are legitimate and biologically informative annotations that should be tested. The exact-string filter targets only the top-level catch-all entries.

The filter is not applied to the lineage signature library (Appendix J.1), to ChEA 2022, or to ENCODE TF ChIP-Seq 2015. The lineage signatures are curated specifically for this evaluation and contain no disease or generic-catch-all entries by construction. The ChEA and ENCODE libraries name gene sets by transcription factor and cell line rather than by pathway concept, so neither the disease substring list nor the generic-catch-all exact match list captures any entries in those libraries.