scPathCF: Intervention-Based Path Decomposition for Perturbation-Aware Protein Prediction

Anonymous authors
Paper under double-blind review

Abstract

A central goal in functional genomics and drug discovery is to predict how protein levels change after a genetic perturbation, especially for surface proteins that serve as immune checkpoints and therapeutic targets. A protein change can follow a corresponding RNA change or arise from regulation that is not explained by the measured RNA. Existing methods typically predict protein directly from RNA or predict the perturbed transcriptional state, without separating these two sources of protein change. We present scPathCF, a perturbation-aware predictor that decomposes protein abundance into a control baseline, an RNA-mediated increment, and a perturbation-dependent residual increment. The mediated head combines cell-level RNA representations with a training-RNA prototype indexed by perturbation identity. The residual head reads the perturbation embedding and an optional offline screening prior, but not factual RNA or cell context. Hard branch switches assemble a control, a mediator-blocked, and a factual state; their differences define the two increments as executable interventions. Across three public CRISPR–CITE-seq screens, the decomposition retains competitive predictive accuracy with a strict-zero prior on Papalexi, a full-data screen on Frangieh, and curated annotations on CaRPool, and the branch interventions reproduce contrasts consistent with known post-transcriptional and RNA-coupled regulation. The state differences provide an explicit account of how the fitted predictor allocates each response between its branches. Code is available at `https://osf.io/xhzes/?view_only=d03a9e76234b425ead29d60f35586c85`.

1 Introduction

Accurate prediction of how a cell responds to a genetic perturbation is a central challenge in functional genomics and drug discovery. The clinically relevant readout is often the protein rather than the transcript, since immune checkpoints, signalling receptors, and therapeutic targets are surface proteins. Multimodal CRISPR screens measure perturbation identity, transcript abundance, and surface protein levels in the same cell (Stoeckius et al., 2017; Papalexi et al., 2021; Frangieh et al., 2021), making protein responses directly observable and exposing structure that transcript-only screens cannot capture. A protein response can follow two routes: some effects are coupled to a corresponding RNA change, whereas others primarily reflect protein stability, trafficking, or degradation. CMTM6 regulation of PD-L1 illustrates the second route, because disrupting CMTM6 alters surface PD-L1 with little change in CD274 transcript abundance (Burr et al., 2017; Mezzadra et al., 2017). The route taken by a perturbation determines whether RNA can serve as a proxy for the protein response.

A model that sees the post-perturbation transcript and perturbation identity together can reach an accurate protein value through either route, because standard objectives reward accuracy without specifying which part of the response the RNA supports. Existing models capture parts of this structure: perturbation predictors extrapolate transcriptional responses through latent arithmetic, optimal transport, disentanglement, or gene-relation representations (Lotfollahi et al., 2019; Bunne et al., 2023; Lotfollahi et al., 2023; Piran et al., 2024; Baek et al., 2025; Roohani et al., 2024; Chen & Zou, 2024b); RNA-to-protein models optimize translation or imputation accuracy (Zhou et al., 2020; Lakkis et al., 2022; Liu et al., 2026); joint multi-omic models learn shared latent spaces (Gayoso et al., 2021; Lan et al., 2023); and MultiPert predicts perturbed RNA and protein together (Zhao et al.,

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

2026). These approaches do not provide a hard intervention that blocks the RNA-mediated route while retaining perturbation identity and cellular context. An accurate prediction therefore does not reveal which part of the response the RNA supports.

We address this gap with scPathCF, an explicitly decomposed protein predictor. The model writes a factual prediction as the sum of a control baseline, an RNA-mediated increment, and a residual increment. The mediated head combines cell-level RNA representations with a training-RNA prototype indexed by perturbation identity. The residual head reads an additive gene-level perturbation embedding and an optional offline screening prior, but not factual RNA or cell context; matched-control context enters through the baseline. The resulting increments describe the contributions of these two input paths. Hard branch switches assemble three states, $\widehat{\mathbf{Y}}_{0M0}$ for the control state, $\widehat{\mathbf{Y}}_{1M0}$ for the mediator-blocked state, and $\widehat{\mathbf{Y}}_{1M1}$ for the factual state. Their differences define the mediated and residual increments by executable branch interventions.

Our contributions are threefold.

First, we propose an intervention-based computational framework, scPathCF, that separates a protein response into an RNA-mediated increment and a perturbation-dependent residual increment.

Second, we implement the framework with a frozen translation encoder, trainable low-rank adapters, separate RNA and residual heads, and hard branch switches.

Third, we evaluate scPathCF on three public CRISPR–CITE-seq screens covering a canonical residual case, a large patient-derived screen, and a multi-gene combination screen. The evaluations test both prediction accuracy and the resulting path assignments.

2 Related Work

Single-cell perturbation prediction. scGen introduced latent-space arithmetic for transferring perturbation effects (Lotfollahi et al., 2019), and neural optimal transport learned distribution-level maps between control and perturbed cells (Bunne et al., 2023). CPA, biolord, and Cradle-VAE improved compositionality or counterfactual disentanglement (Lotfollahi et al., 2023; Piran et al., 2024; Baek et al., 2025). GEARS uses gene relations for unseen single- and multi-gene perturbations (Roohani et al., 2024), while GenePert couples gene identity representations with a regression model (Chen & Zou, 2024b;a). These methods return a transcriptional response or perturbation-level effect. scPathCF instead returns protein increments that can be measured by blocking an information channel.

Transcript-to-protein and multi-omic modeling. cTP-net and sciPENN infer surface proteins from transcriptomes (Zhou et al., 2020; Lakkis et al., 2022), and scTranslator scales this mapping through pretraining (Liu et al., 2026). totalVI jointly models RNA and protein counts, and scMOG generates paired multi-omic profiles (Gayoso et al., 2021; Lan et al., 2023). MultiPert targets multimodal perturbation response with adversarial alignment and dual attention (Zhao et al., 2026). A single mapping or shared latent space leaves the two routes entangled, so a correct protein value does not identify which route produced it. scPathCF exposes states in which either route is removed, making the increments directly readable.

Counterfactual and causal single-cell models. StateXDiff models multimodal perturbation responses with conditional diffusion (Shi et al., 2026), while CausCell learns disentangled representations for controllable counterfactual generation (Gao et al., 2025). CINEMA-OT estimates perturbation effects using causal identification and optimal transport (Dong et al., 2023). T2Pdecoder infers protein representations from transcriptomic data for protein-centric analyses (Wang et al., 2026). These methods address distinct prediction and inference tasks. scPathCF focuses on differences between protein predictions obtained by switching its RNA and residual branches. The interpretation of these differences depends on the information supplied to each branch (Section 5.1).

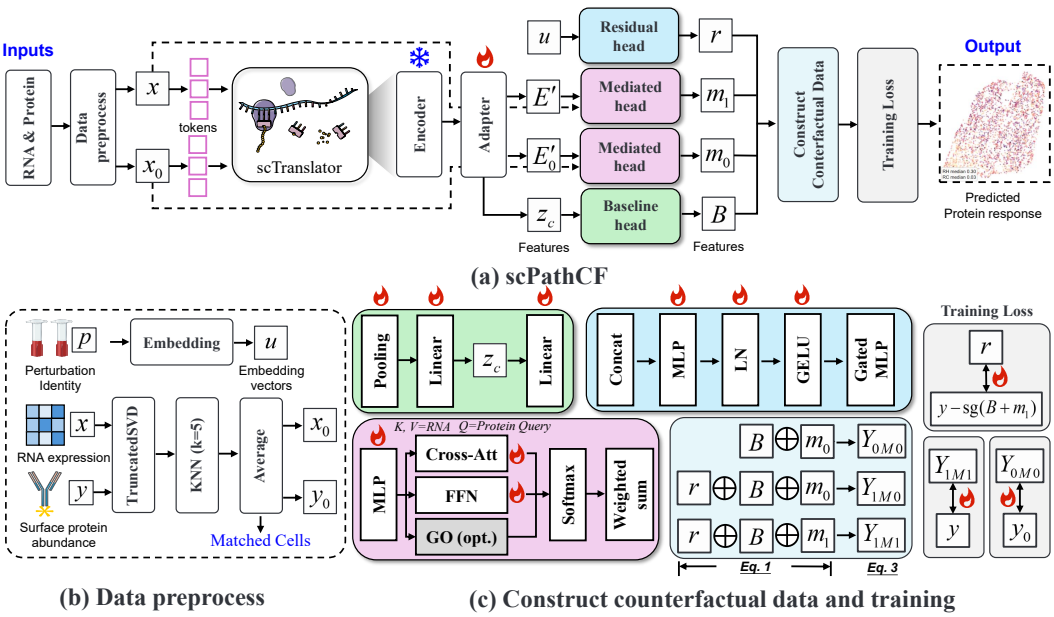

Figure 1: Overview of SCPATHCF. **(a)** End-to-end prediction. A frozen scTranslator backbone with trainable low-rank adapters encodes factual and matched control RNA. The control context z_C, derived solely from control encodings, provides the shared baseline B. The mediated head H_{med} uses RNA readouts, while the residual head H_{res} takes only the perturbation embedding $\mathbf{u}$ and screening prior $\mathbf{q}$. The mediated path also uses an identity-indexed training-RNA prototype (equation 3); this smoother is omitted from the schematic. Snowflakes and flames denote frozen and trainable modules, respectively. **(b)** Data preprocessing. Preprocessing produces RNA features and replicate-matched control references x_0 and y_0. The perturbation identity p is encoded as $\mathbf{u}$. Surface protein is the prediction target; cell-level protein measurements enter the training targets, while screen-derived annotations may enter through $\mathbf{q}$. **(c)** Counterfactual construction and training. The mediated head combines complementary RNA readouts through a softmax gate; the residual head uses a gated MLP. Hard branch interventions define the control, mediator-blocked and factual states in equation 4, whose differences yield the reported increments. Training combines factual reconstruction with path-consistency and mismatch-contrast terms.

3 METHOD

3.1 NOTATION AND PROBLEM SETTING

For a perturbed cell, let $\mathbf{x} \in \mathbb{R}^G$ and $\mathbf{y} \in \mathbb{R}^P$ denote normalized RNA and log-transformed ADT measurements. We match the cell to $k = 5$ nearest non-targeting controls after truncated SVD and average their RNA and ADT profiles, producing $(\mathbf{x}_0, \mathbf{y}_0)$. The perturbation embedding $\mathbf{u}$ is the mean of the member-gene embeddings, so it is defined for unseen combinations. An optional screening prior $\mathbf{q}$ stores perturbation-level annotations from an offline screen. We distinguish four provenance modes: strict zero, a screen estimated from training cells only, an explicitly supplied external screen, and a legacy oracle screen computed from the full dataset. The oracle mode is target-derived. It appears in the Papalexi sensitivity analysis and in the Frangieh legacy configuration reported in the main comparison. The learnable quantities are the low-rank adapters, the three prediction heads, the perturbation embedding table, and the gate weights; the observed inputs and any supplied prior are fixed. The goal is to predict $\mathbf{y}$ for a held-out guide and expose how the prediction is allocated between the two paths.

Figure 1 summarizes the model. SCPATHCF predicts

$$\widehat{\mathbf{Y}} = B(\mathbf{z}_C) + H_{\text{med}}(\mathbf{M}) + H_{\text{res}}(\mathbf{u}, \mathbf{q}), \tag{1}$$

where B is the control baseline, H_{med} is the mediated head, and H_{res} is the residual head. This decomposition is predictive and operational: the residual increment is the contribution of the perturbation-

dependent branch. Input exclusion alone does not identify the protein variation that RNA cannot explain. A biochemical interpretation would additionally require the matching, exclusion, additivity, and screening-prior assumptions discussed in Section 5.1.

3.2 Frozen Encoder and Control Baseline

We pass $\mathbf{x}$ and $\mathbf{x}_0$ through the official scTranslator encoder (Liu et al., 2026) and keep its weights frozen. A trainable low-rank adapter transforms the factual and control token representations into $\mathbf{E}'$ and $\mathbf{E}'_0$, providing the mediated head with a parameter-efficient interface to the pretrained features. We do not use the pretrained decoder for the main prediction. Pooling $\mathbf{E}'_0$ yields the control context $\mathbf{z}_C$, and a linear head maps it to the control baseline $B(\mathbf{z}_C)$. The baseline is therefore shared across all counterfactual states.

3.3 Mediated and Residual Paths

The mediated path combines complementary RNA readouts: cross-attention over encoder tokens, dense RNA heads, a stem-fusion head, and an optional pathway head. Their softmax-weighted output is

$$\mathbf{h}(\mathbf{x}) = \sum_k \alpha_k h_k(\mathbf{x}, \mathbf{E}'), \qquad \boldsymbol{\alpha} = \mathrm{softmax}(\mathbf{w}). \tag{2}$$

This readout is shared by factual, control, and prototype RNA. For perturbation p, a training-RNA prototype $\bar{\mathbf{x}}_p$ supplies a neighbourhood estimate, and a learned gate g_p gives

$$\mathbf{m}_1 = (1 - g_p)\mathbf{h}(\mathbf{x}) + g_p \mathbf{h}(\bar{\mathbf{x}}_p), \qquad \mathbf{m}_0 = \mathbf{h}(\mathbf{x}_0). \tag{3}$$

The gate reads the factual and prototype readouts and their absolute difference; screen-enabled runs also supply the RH flag. Each prototype averages RNA from the training split. For unseen identities, the gene-composition configuration averages available member-gene prototypes; otherwise the fallback is the global training-RNA mean. Thus the raw readouts use RNA, while the complete mediated path also uses identity through prototype selection. In equation 1, $\mathbf{M}$ includes these inputs. The dense heads are pretrained in Phase A, then frozen on Papalexi and Frangieh and kept trainable on CaRPool.

The residual path receives the perturbation embedding and, when available, the screening prior. It drops the control context because that information is already carried by the baseline $B(\mathbf{z}_C)$. An MLP–LayerNorm–GELU trunk followed by a gated MLP maps this input to an ADT increment $\mathbf{r}$. The residual path never receives $\mathbf{x}$ or $\mathbf{E}'$, and $\mathbf{r} = \mathbf{0}$ for controls. Its output defines the perturbation-dependent residual increment in the state assembly.

3.4 Hard Counterfactual Assembly and Training

The same discrete assembly is used for training and inference,

$$\widehat{\mathbf{Y}}_{0M0} = B + \mathbf{m}_0, \qquad \widehat{\mathbf{Y}}_{1M0} = B + \mathbf{m}_0 + \mathbf{r}, \qquad \widehat{\mathbf{Y}}_{1M1} = B + \mathbf{m}_1 + \mathbf{r}, \tag{4}$$

so the reported path contributions are $\widehat{\boldsymbol{\tau}}_{\mathrm{med}} = \widehat{\mathbf{Y}}_{1M1} - \widehat{\mathbf{Y}}_{1M0} = \mathbf{m}_1 - \mathbf{m}_0$ and $\widehat{\boldsymbol{\tau}}_{\mathrm{res}} = \widehat{\mathbf{Y}}_{1M0} - \widehat{\mathbf{Y}}_{0M0} = \mathbf{r}$. Each quantity is produced by an executable branch intervention rather than a soft attribution score.

Phase B jointly trains the adapters, baseline, unfrozen mediated readouts, residual path, embeddings, and gates. Each update evaluates the weighted sum

$$\mathcal{L} = w_f^{(e)} \mathcal{L}_{\mathrm{fact}} + w_p^{(e)} \lambda_{\mathrm{path}} \mathcal{L}_{\mathrm{path}} + w_m^{(e)} \lambda_{\mathrm{path}} \mathcal{L}_{\mathrm{mm}} + \mathcal{R}, \tag{5}$$

where factual reconstruction fits $\mathbf{y}$, the path loss fits the matched control and compares factual and blocked errors, and mismatch contrast substitutes another cell's RNA while keeping identity, prototype, and control inputs fixed. The regularizer $\mathcal{R}$ includes residual-gate sparsity and a stop-gradient leftover fit. Additive state identities hold by construction; their numerical consistency terms do not impose an additional learned constraint. Screen-enabled runs also use RH/RC weights, RH oversampling, and a residual-share prior. Strict $q = 0$ removes these screen-dependent terms while retaining the unstratified objective. The base loss weights are $0.55/0.30/0.15$ for Papalexi and CaRPool and $0.68/0.20/0.12$ for Frangieh, with a two-epoch Phase-B weight schedule. Appendix D.2 defines the terms and schedule.

4 EXPERIMENTS

Table notation. Bold blue marks the best displayed mean within each comparison; ties at the reported precision are marked together. The marking does not imply statistical significance.

4.1 EXPERIMENTAL SETUP

Datasets. We evaluate three public CRISPR–CITE-seq datasets along two axes relevant to path decomposition: platform and perturbation composition. The screens use different perturbation and readout chemistries; Papalexi and Frangieh are dominated by single-gene targets, whereas CaRPool also contains multi-gene combinations. Papalexi ECCITE-seq provides a canonical residual case through CMTM6 and PD-L1 and uses CRISPR knockout with four measured ADTs (Papalexi et al., 2021). Frangieh Perturb-CITE-seq is a larger patient-derived screen with 24 measured ADTs, allowing the split to be evaluated in a different biological system (Frangieh et al., 2021). CaRPool-seq uses Cas13 targeting in HEK293FT cells and includes two-gene and three-gene combinations, enabling an unseen-combination test (Wessels et al., 2023). Together, the datasets contain 154,681 cells, 7–241 perturbations per screen, and both single and combinatorial targets.

Table 1: Dataset statistics under the unseen-guide evaluation. Frangieh contains 24 measured proteins, of which 14 map to the backbone. RH denotes residual-high and RC denotes RNA-coupled; both are evaluation strata derived from target-level measurements for Papalexi and Frangieh and from explicit curated annotations for CaRPool. They are not independent mechanism labels.

Dataset	Cells	Genes	ADTs	Pert.	Test	RH / RC
Papalexi	20,729	18,649	4	25	3,871	1,749 / 458
Frangieh	126,966	22,097	24 (14)	241	21,591	2,779 / 2,109
CaRPool	6,986	20,631	5	7	1,293	518 / 775

Papalexi uses non-targeting (NT) controls and contains four ADTs (CD86, PDL1, PDL2, and CD366). Frangieh uses the dataset's control label and evaluates the 14 proteins supported by the backbone. CaRPool uses NT controls and contains CD46, CD55, CD71, CD29, and CD56. RNA is normalized and log-transformed; the backbone mapping uses 6,000 genes, while the Papalexi strict-zero dense input uses 2,000 variable genes. ADTs are log-transformed. The main split holds out guides, so no guide appears in both train and test. The mixed test strata contain 1,664, 16,703, and 0 cells for Papalexi, Frangieh, and CaRPool, respectively. Guide labels are grouped before splitting, so cells from one guide remain in one partition. RNA feature selection, SVD, and replicate-matched control construction precede the split in the supplied implementation. Complete per-dataset statistics appear in Appendix A.

Metrics. We report mean squared error and Pearson correlation after flattening cells and proteins, both over all test cells and over the residual-high (RH) evaluation subset. The split is fixed at $s_{\text{split}} = 0$ and train seeds are $s_{\text{train}} \in \{0, 1, 2\}$; values are mean $\pm$ sample standard deviation. Appendix F reports post-hoc paired training-seed MSE intervals and tests for Table 2; these quantify optimization variation on the reported fixed splits. Appendix D gives the metric definitions.

Baselines. We compare against GEARS (Roohani et al., 2024), n-gram ridge, scTranslator-FT (Liu et al., 2026), cTP-net (Zhou et al., 2020), sciPENN (Lakkis et al., 2022), totalVI (Gayoso et al., 2021), scMOG (Lan et al., 2023), MultiPert (Zhao et al., 2026), ridge regression (Hoerl & Kennard, 1970), and a perturbation-mean baseline. The n-gram ridge comparator uses hashed character-trigram features of gene symbols and ridge regression; it does not use the GenePT embeddings of GenePert (Chen & Zou, 2024b;a). The unseen-guide baselines use the same fixed split and protein dimensions, and complete comparisons appear in Appendix B.

Hardware and configuration. The legacy runs used a single NVIDIA vGPU-48GB partition with 224 CPU cores and 754 GB of host memory, under Python 3.12.2, PyTorch 2.5.1 with CUDA 12.4, and the official scTranslator checkpoint. The Papalexi strict-zero run uses 3,000 training cells from the 15,166-cell training partition. The reported configurations use batch size 1 and train for 10 epochs at 3.0×10^{-4} on Papalexi, 16 epochs at 2.5×10^{-4} on Frangieh, and 12 epochs at 3.5×10^{-4} on

Table 2: Unseen-guide protein prediction. Lower MSE and higher Pearson are better. Entries are mean ± standard deviation over three training seeds on the fixed split. Papalexi uses strict $q = 0$, Frangieh uses a target-derived full-data screen, and CaRPool uses curated annotations. The Frangieh rows therefore compare methods with different access to target-derived prior information. N-gram ridge uses hashed gene-symbol features. Bold marks the best observed mean, not statistical significance; paired intervals and tests appear in Table 15.

Dataset	Method	Overall MSE ↓	Overall r ↑	RH MSE ↓	RH r ↑
Papalexi	scPATHCF ($q = 0$)	**0.3158±.0055**	**0.7989±.0038**	**0.3334±.0104**	**0.7886±.0072**
	cTP-net	0.3332±.0091	0.7935±.0030	0.3489±.0110	0.7839±.0048
	N-gram ridge	0.4792±.0001	0.6696±.0001	0.5030±.0005	0.6530±.0004
	GEARS	0.5190±.0163	0.6495±.0107	0.5521±.0184	0.6282±.0137
	MultiPert	0.6327±.0074	0.5349±.0016	0.6212±.0038	0.5399±.0038
Frangieh	scPATHCF (legacy q)	**0.2960±.0006**	**0.9501±.0001**	**0.3115±.0003**	**0.9473±.0000**
	cTP-net	0.3005±.0019	0.9495±.0005	0.3139±.0011	0.9471±.0003
	N-gram ridge	0.3148±.0001	0.9469±.0000	0.3212±.0001	0.9456±.0000
	GEARS	0.3302±.0049	0.9444±.0006	0.3450±.0039	0.9415±.0005
	MultiPert	0.4508±.0003	0.9230±.0001	0.4582±.0021	0.9213±.0004
CaRPool	scPATHCF (curated q)	**0.3292±.0186**	**0.8395±.0090**	**0.3384±.0122**	**0.8552±.0054**
	N-gram ridge	0.3377±.0003	0.8363±.0001	0.3542±.0004	0.8530±.0001
	GEARS	0.3418±.0110	0.8337±.0009	0.3620±.0168	0.8479±.0015
	MultiPert	0.3773±.0168	0.8147±.0035	0.3956±.0171	0.8312±.0028
	cTP-net	0.8898±.0325	0.5585±.0125	0.8412±.0817	0.6211±.0399

CaRPool. The dense mediated head receives a separate pretraining stage; it is then frozen on Papalexi and Frangieh and remains trainable on CaRPool.

4.2 Unseen-Guide Prediction

Table 2 reports the cross-method comparison and identifies the prior used by each scPATHCF row. On Papalexi, the strict $q = 0$ model reaches overall MSE 0.3158 and RH Pearson 0.7886, compared with overall MSE 0.3332 and RH Pearson 0.7839 for cTP-net. On Frangieh, the legacy screen-prior configuration reaches MSE 0.2960, compared with 0.3005 for cTP-net and 0.3148 for N-gram ridge. On CaRPool, the curated-prior configuration reaches MSE 0.3292 and Pearson 0.8395, compared with 0.3377 and 0.8363 for N-gram ridge. The Frangieh screen uses full-data target measurements, so its comparison does not establish performance without target-derived prior information. The paired 95% MSE-difference intervals versus the closest listed baselines include zero on all three datasets (Appendix F); these small mean differences do not establish superiority or noninferiority.

cTP-net, sciPENN, and ridge read factual RNA at test time. GEARS and n-gram ridge use control RNA, control ADTs, and perturbation identity in their CITE adapters. scPATHCF combines factual RNA with a separate perturbation-dependent residual path. The RH columns are evaluation strata defined from target measurements. Seed counts and metric definitions are given in the protocol and appendix.

4.3 Screen-Prior Provenance and Sensitivity

The legacy prior was computed from perturbation-level RNA and protein changes in the full Papalexi dataset, so it is a target-derived provenance condition. We reran the same three train seeds with four provenance modes. Strict $q = 0$ removes the prior table, RH/RC oversampling, stratum-specific structural weighting, and the residual-share prior. Train-only estimates the screen from training cells, external uses a separately supplied annotation file, and oracle reproduces the legacy full-data screen.

Relative to matched oracle seeds, strict $q = 0$ changes overall MSE by -0.00062 and Pearson by $+0.00148$; train-only changes them by $+0.00283$ and -0.00101; and external changes them by approximately -0.00100 and $+0.00134$. Each difference is smaller than the across-seed standard deviation, so overall accuracy is similar across provenance modes. The train-only screen selects two

Table 3: Papalexi screening-prior audit under the same unseen-guide split and three train seeds. RH remains an evaluation-only stratum in the strict $q = 0$ run.

Prior mode	Overall MSE	Overall r	RH MSE	RH r
Strict $q = 0$	0.3158±.0055	**0.7989±.0038**	0.3334±.0104	0.7886±.0072
Train-only	0.3193±.0034	0.7964±.0024	0.3356±.0047	0.7866±.0040
External	**0.3155±.0036**	0.7988±.0028	**0.3288±.0060**	**0.7907±.0043**
Legacy oracle	0.3165±.0072	0.7974±.0049	0.3306±.0130	0.7894±.0083

RH and one RC target, compared with eight RH and four RC targets under the oracle screen. The screen composition therefore depends on the data used to construct it.

4.4 Biological Consistency Analysis

Aggregate accuracy does not test the semantics of a decomposition, so we examine a biological contrast as a consistency analysis. On Papalexi, CMTM6 changes PD-L1 protein while CD274 RNA stays within 0.02 of control, whereas IFNGR, JAK, and STAT perturbations move RNA and protein together. scPathCF assigns the CMTM6 response mainly to $\mathbf{r}$ and the IFN response mainly to $\mathbf{m}_1 - \mathbf{m}_0$; the hard branch removals show the same contrast. These assignments are compatible with the known post-transcriptional role of CMTM6. They remain model-relative because the paths are defined by the architecture's information restrictions, and the static measurements support consistency with the decomposition rather than a temporal biochemical mechanism.

Across three Frangieh seeds, the median residual share $\rho_r = \|\mathbf{r}\|/(\|\mathbf{m}_1 - \mathbf{m}_0\| + \|\mathbf{r}\|)$ is 0.250 ± 0.034 for RH perturbations and 0.078 ± 0.021 for RNA-coupled perturbations. The offline screen correlates with the learned residual share at Spearman $\rho = 0.85$ ($P = 3.7 \times 10^{-3}$, $n = 9$) on Papalexi and $\rho = 0.58$ ($P = 0.060$, $n = 11$) on Frangieh. Because screen annotations also guide training, these associations measure agreement with that supervision. Table 3 evaluates prediction accuracy across prior modes; it does not test whether the same branch assignments persist without the screen.

A frozen-checkpoint RNA-noise diagnostic leaves $\mathbf{r}$ unchanged while changing the mediated output (Appendix G). It tests the residual input restriction at inference; it does not test how noisy training data redistribute effects between the branches.

4.5 Ablations

The ablations use the archived screen-enabled reference configurations (Figure 2); the Papalexi reference therefore differs from the strict-zero run in Table 2. Removing each path reveals dataset-dependent effects. Removing the mediated head changes overall MSE by 0.1719 ± 0.0093 on Papalexi and 0.1153 ± 0.0926 on CaRPool. Removing the residual head changes overall MSE by 0.0175 ± 0.0007 on Frangieh and 0.1016 ± 0.0459 on CaRPool. Table 10 reports paired changes for all six removal variants on the overall and RH readouts. Removing the screening prior changes Frangieh MSE by 0.0195 ± 0.0020, while the changes on Papalexi and CaRPool fall within seed variation.

The loss weights balance factual, path, and mismatch terms within each training update. We vary their mix while keeping the sum at one (Figure 3; Appendix Table 11). Papalexi and Frangieh show non-monotonic error variation across the sampled weights, whereas both CaRPool error readouts decrease as the factual weight increases. This single-seed, reduced-budget sweep is a sensitivity diagnostic; the main comparisons retain their reported configurations.

We also hold out entire perturbation identities. On Papalexi, scPathCF has lower overall MSE than the listed comparators. For the held-out CaRPool three-gene combination, gene-mean embeddings improve on learned-ID embeddings, while GEARS and n-gram ridge obtain lower MSE. Appendix C reports these comparisons and identifies the reused Frangieh baseline entries.

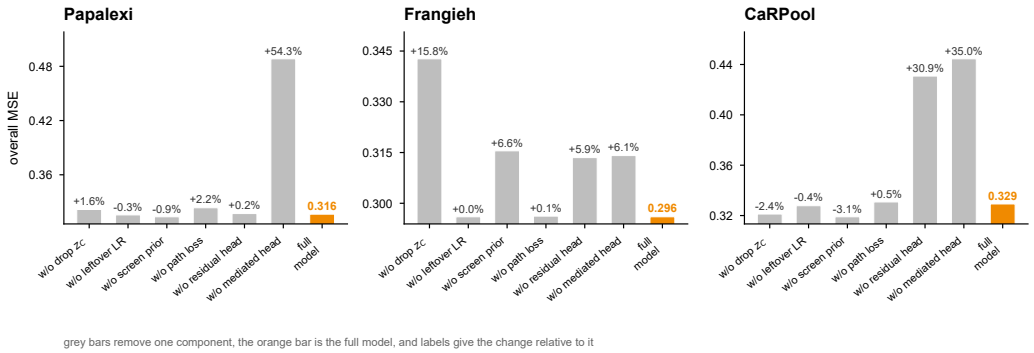

Figure 2: Leave-one-out ablation of the archived screen-enabled configurations, one panel per dataset; Papalexi uses a different reference from the strict-zero main table. Grey bars remove one component and the orange bar is the full model; labels give the change relative to the full model. The vertical axis shows absolute overall MSE on a range that starts above zero, so smaller effects remain visible. On Papalexi the mediated head has the largest effect, on CaRPool both paths contribute, and on Frangieh the two regularizers have small changes.

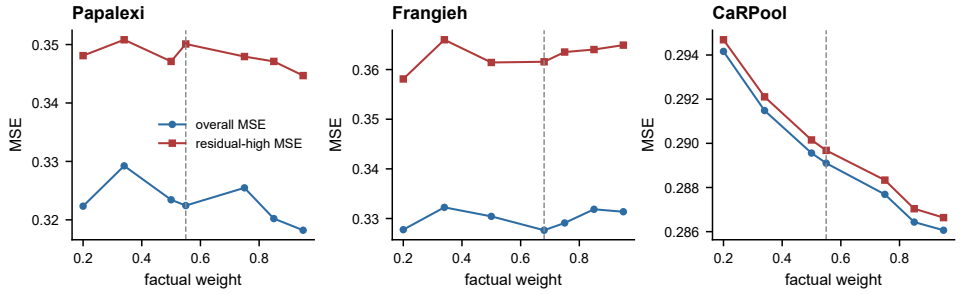

Figure 3: Loss-weight sweep on the three datasets, one panel each. Within a panel all three weights vary while their sum remains one; the dashed vertical line marks the configuration used in the main tables, and the two curves show overall error and residual-high error on the same axes. Each point is a single training run at a reduced uniform budget, so the curves describe the trade-off rather than a seed-averaged ranking.

5 Conclusion

We introduced scPathCF, an executable three-state decomposition for perturbation-aware protein prediction. Hard branch switches define RNA-branch and residual increments relative to a shared control state. The experiments report unseen-guide prediction under three prior settings: strict $q = 0$ on Papalexi, a full-data screen on Frangieh, and curated annotations on CaRPool. Papalexi has similar mean prediction accuracy across the four prior modes. The branch assignments summarize the fitted model; identifying biological mechanisms requires evidence beyond these state differences.

5.1 Limitations

The decomposition depends on matching, prototype construction, and additive state assembly. The mediated path includes identity-indexed RNA prototypes, so its increment is not an identity-free mediator effect. The legacy RH labels and oracle prior are derived from the same target measurements, so agreement with them is a consistency analysis. CaRPool uses curated strata because the original feature naming does not provide a reliable automatic cognate-RNA screen for every measured protein; mechanism-specific results therefore depend on the annotation. Some extended baselines have one reported seed, and the complete held-out multi-gene analysis is provided in Appendix C.

AI USE STATEMENT

We used generative AI tools (DeepSeek Harness, Claude, and OpenAI Codex) for model and evaluation-code implementation, development of the experimental methodology, literature retrieval and summarisation, manuscript organisation and editing, interpretation of results, figure creation and revision, and translation into English. The analyses use public single-cell datasets; generative AI was not used to generate the experimental data. The authors are responsible for the manuscript, its claims, and the accompanying code and figures.

REFERENCES

Seungheun Baek, Soyon Park, Yan Ting Chok, et al. Cradle-VAE: Enhancing single-cell gene perturbation modeling with counterfactual reasoning-based artifact disentanglement. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(15):15445–15452, 2025. doi: 10.1609/aaai.v39i15.33695.

Charlotte Bunne, Stefan G Stark, Gabriele Gut, et al. Learning single-cell perturbation responses using neural optimal transport. *Nature Methods*, 20(11):1759–1768, 2023. doi: 10.1038/s41592-023-01969-x.

Marian L Burr, Christina E Sparbier, Yih-Chih Chan, et al. CMTM6 maintains the expression of PD-L1 and regulates anti-tumour immunity. *Nature*, 549(7670):101–105, 2017. doi: 10.1038/nature23643.

Yiqun Chen and James Zou. GenePT: A simple but effective foundation model for genes and cells built from ChatGPT. *bioRxiv*, pp. 2023.10.16.562533, 2024a. doi: 10.1101/2023.10.16.562533. URL https://www.biorxiv.org/content/10.1101/2023.10.16.562533v2. Version 2.

Yiqun Chen and James Zou. GenePert: Leveraging GenePT embeddings for gene perturbation prediction. *bioRxiv*, pp. 2024.10.27.620513, 2024b. doi: 10.1101/2024.10.27.620513.

Mingze Dong, Bao Wang, Jessica Wei, et al. Causal identification of single-cell experimental perturbation effects with CINEMA-OT. *Nature Methods*, 20(11):1769–1779, 2023. doi: 10.1038/s41592-023-02040-5.

Chris J Frangieh, Johannes C Melms, Pratiksha I Thakore, et al. Multimodal pooled Perturb-CITE-seq screens in patient models define mechanisms of cancer immune evasion. *Nature Genetics*, 53(3):332–341, 2021. doi: 10.1038/s41588-021-00779-1.

Yicheng Gao, Kejing Dong, Caihua Shan, et al. Causal disentanglement for single-cell representations and controllable counterfactual generation. *Nature Communications*, 16(1):6775, 2025. doi: 10.1038/s41467-025-62008-1.

Adam Gayoso, Zoë Steier, Romain Lopez, et al. Joint probabilistic modeling of single-cell multi-omic data with totalVI. *Nature Methods*, 18(3):272–282, 2021. doi: 10.1038/s41592-020-01050-x.

Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970. doi: 10.1080/00401706.1970.10488634.

Justin Lakkis, Amelia Schroeder, Kenong Su, et al. A multi-use deep learning method for CITE-seq and single-cell RNA-seq data integration with cell surface protein prediction and imputation. *Nature Machine Intelligence*, 4(11):940–952, 2022. doi: 10.1038/s42256-022-00545-w.

Meng Lan, Shixiong Zhang, and Lin Gao. Efficient generation of paired single-cell multiomics profiles by deep learning. *Advanced Science*, 10(21):2301169, 2023. doi: 10.1002/advs.202301169.

Linjing Liu, Wei Li, Fang Wang, et al. A pre-trained large generative model for translating single-cell transcriptomes to proteomes. *Nature Biomedical Engineering*, 10(8):1672–1691, 2026. doi: 10.1038/s41551-025-01528-z.

Mohammad Lotfollahi, F Alexander Wolf, and Fabian J Theis. scGen predicts single-cell perturbation responses. *Nature Methods*, 16(8):715–721, 2019. doi: 10.1038/s41592-019-0494-8.

Mohammad Lotfollahi, Anna Klimovskaia Susmelj, Carlo De Donno, et al. Predicting cellular responses to complex perturbations in high-throughput screens. *Molecular Systems Biology*, 19(6): e11517, 2023. doi: 10.15252/msb.202211517.

Riccardo Mezzadra, Chong Sun, Lucas T Jae, et al. Identification of CMTM6 and CMTM4 as PD-L1 protein regulators. *Nature*, 549(7670):106–110, 2017. doi: 10.1038/nature23669.

Efthymia Papalexi, Eleni P Mimitou, Andrew W Butler, et al. Characterizing the molecular regulation of inhibitory immune checkpoints with multimodal single-cell screens. *Nature Genetics*, 53(3): 322–331, 2021. doi: 10.1038/s41588-021-00778-2.

Zoe Piran, Niv Cohen, Yedid Hoshen, et al. Disentanglement of single-cell data with biolord. *Nature Biotechnology*, 42(11):1678–1683, 2024. doi: 10.1038/s41587-023-02079-x.

Yusuf Roohani, Kexin Huang, and Jure Leskovec. Predicting transcriptional outcomes of novel multigene perturbations with GEARS. *Nature Biotechnology*, 42(6):927–935, 2024. doi: 10.1038/s41587-023-01905-6.

Peiting Shi, Ningfeng Que, Xianzhe Huang, et al. StateXDiff: Cell state-contextualized multimodal diffusion for single-cell perturbation prediction. *arXiv preprint arXiv:2605.16104*, 2026. doi: 10.48550/arXiv.2605.16104. URL `https://arxiv.org/abs/2605.16104`.

Marlon Stoeckius, Christoph Hafemeister, William Stephenson, et al. Simultaneous epitope and transcriptome measurement in single cells. *Nature Methods*, 14(9):865–868, 2017. doi: 10.1038/nmeth.4380.

Hui Wang, Jihong Tang, Yimeng Qiao, et al. T2Pdecoder enables protein-centric analyses from transcriptomic data. *Nature Communications*, 17:7458, 2026. doi: 10.1038/s41467-026-74209-3.

Hans-Hermann Wessels, Alejandro Méndez-Mancilla, Yuhan Hao, et al. Efficient combinatorial targeting of RNA transcripts in single cells with Cas13 RNA Perturb-seq. *Nature Methods*, 20(1): 86–94, 2023. doi: 10.1038/s41592-022-01705-x.

Mengyuan Zhao, Xinyue Tang, Jiawei Li, et al. MultiPert: An adversarial alignment and dual attention framework for single-cell multi-omics perturbation prediction. *PLOS Computational Biology*, 22(3):e1014054, 2026. doi: 10.1371/journal.pcbi.1014054.

Zilu Zhou, Chengzhong Ye, Jingshu Wang, and Nancy R Zhang. Surface protein imputation from single cell transcriptomes by deep neural networks. *Nature Communications*, 11(1):651, 2020. doi: 10.1038/s41467-020-14391-0.

A Dataset Details

Papalexi uses non-targeting (NT) controls and contains four ADTs (CD86, PDL1, PDL2, and CD366). Frangieh uses the dataset's control label and evaluates the 14 proteins supported by the backbone. CaRPool uses NT controls and contains CD46, CD55, CD71, CD29, and CD56; its seven perturbations include multi-gene combinations. For the unseen-guide evaluation, cells sharing a guide remain in one partition. The implementation selects RNA features, fits SVD, and matches controls before making this split, so held-out RNA contributes to preprocessing.

A.1 Implementation Details

The backbone mapping uses 6,000 RNA genes. The Papalexi strict-zero checkpoint additionally takes a 2,000-gene dense RNA input; its manifest records 3,000 training cells, 1,692 validation cells, and 3,871 test cells. Control matching uses five nearest controls within each replicate in a 64-dimensional truncated-SVD space fitted before splitting. The scTranslator encoder and decoder are frozen, and only the listed adapters and task heads are optimized. The shared task dimensions are a 256-dimensional control context, a 128-dimensional perturbation embedding, a 128-dimensional latent width, a 64-dimensional latent state, a 256-dimensional fusion width, an eight-dimensional screening-prior vector, and a 768-dimensional residual hidden layer. The mediated module uses

Table 4: Dataset-specific training settings. The loss weights multiply the factual, path, and mismatch terms in each update. RH oversampling applies to screen-enabled runs; the Papalexi strict-zero run uses an effective factor of 1.0.

Dataset	LoRA r	Epochs	LR	Loss weights	RH oversample
Papalexi	16	10	3.0×10^{-4}	.55/.30/.15	2.0
Frangieh	32	16	2.5×10^{-4}	.68/.20/.12	4.5
CaRPool	16	12	3.5×10^{-4}	.55/.30/.15	1.25

Table 5: Training and prediction for scPATHCF. Prototype selection, hard state assembly, and the weighted objective are shared by the equations and the implementation.

Algorithm 1: path-decomposed protein prediction

Inputs: factual RNA $\mathbf{x}$, identity p, prior $\mathbf{q}$, matched control RNA $\mathbf{x}_0$; protein targets $\mathbf{y}, \mathbf{y}_0$ during training.
Precompute: training-RNA prototypes $\bar{\mathbf{x}}_p$ and the unseen-identity fallback of equation 3.
Predict$(\mathbf{x}, \mathbf{x}_0, p, \mathbf{q})$:
  Encode RNA with the frozen backbone and trainable adapters.
  $\mathbf{z}_C \leftarrow \mathrm{pool}(\mathbf{E}_0')$; $B \leftarrow B(\mathbf{z}_C)$.
  Compute shared RNA readouts $\mathbf{h}(\mathbf{x})$, $\mathbf{h}(\mathbf{x}_0)$, and $\mathbf{h}(\bar{\mathbf{x}}_p)$.
  $\mathbf{m}_0 \leftarrow \mathbf{h}(\mathbf{x}_0)$; $\mathbf{m}_1 \leftarrow (1 - g_p)\mathbf{h}(\mathbf{x}) + g_p \mathbf{h}(\bar{\mathbf{x}}_p)$.
  $\mathbf{u} \leftarrow \mathrm{mean}_{g \in p}\, \mathrm{embed}(g)$; $\mathbf{r}_{\mathrm{raw}} \leftarrow H_{\mathrm{res}}(\mathbf{u}, \mathbf{q})$.
  $\mathbf{r} \leftarrow \mathbf{0}$ for controls, and $\mathbf{r}_{\mathrm{raw}}$ otherwise.
  $\widehat{\mathbf{Y}}_{0M0} \leftarrow B + \mathbf{m}_0$; $\widehat{\mathbf{Y}}_{1M0} \leftarrow B + \mathbf{m}_0 + \mathbf{r}$; $\widehat{\mathbf{Y}}_{1M1} \leftarrow B + \mathbf{m}_1 + \mathbf{r}$.
  return the three states and increments $\mathbf{m}_1 - \mathbf{m}_0$, $\mathbf{r}$.
Train:
  Phase A: pretrain the dense RNA heads on protein reconstruction.
  Phase B: freeze those heads except on CaRPool; keep the backbone frozen.
  For each epoch e and sampled training cell:
    Compute all three states with **Predict**.
    Substitute another cell's RNA, retaining the original identity, prototype, and controls, to obtain the mismatch prediction.
    Evaluate $\mathcal{L}_{\mathrm{fact}}$, $\mathcal{L}_{\mathrm{path}}$, $\mathcal{L}_{\mathrm{mm}}$, and $\mathcal{R}$ (Appendix D.2).
    Update trainable parameters using their weighted sum in equation 5.
  return the selected adapters and heads.

two layers and eight attention heads, and the control matching described above uses five nearest neighbours. Dataset-specific optimization settings are listed below.

The corresponding path-loss weights are 1.0, 0.55, and 0.9 for Papalexi, Frangieh, and CaRPool respectively, and the residual-prior weights in screen-enabled runs are 0.8, 0.2, and 0.35. The LoRA dropout is 0.05, the residual path drops $\mathbf{z}_C$ in all reported runs, and the CaRPool configuration keeps the dense mediated head trainable after warm-up. The implementation exposes `zero`, `train_only`, `external`, and `oracle` q modes; `zero` also disables RH/RC oversampling, stratum-specific structural weighting, and the residual-share prior. All random seeds, split manifests, preprocessing choices, and remaining loss coefficients are included in the anonymous configuration package.

A.2 Training Procedure

Algorithm 1 (Table 5) restates the two-stage procedure from the Method section. Every reported increment is produced by the discrete state assembly in equation 4, not by a soft attribution score.

B Complete Unseen-Guide Comparisons

Tables 6–8 report every baseline in the source evaluation. scTranslator-FT, sciPENN, totalVI, scMOG, and PertMean were available only for $s_{\mathrm{train}} = 0$ in the extended comparison, so their displayed standard deviation is zero; Frangieh has no totalVI result.

Table 6: Papalexi complete comparison. The scPathCF row uses strict $q = 0$; RH is an evaluation-only stratum. Values are mean ± sample standard deviation over three train seeds when available. Rows marked † have one reported run; other rows summarize three seeds. A displayed ±0.0000 may reflect rounding and does not by itself identify the run count.

Method	Overall MSE	Overall r	RH MSE	RH r
scPathCF ($q = 0$)	**0.3158±.0055**	**0.7989±.0038**	**0.3334±.0104**	**0.7886±.0072**
GEARS	0.5190±.0163	0.6495±.0107	0.5521±.0184	0.6282±.0137
N-gram ridge	0.4792±.0001	0.6696±.0001	0.5030±.0005	0.6530±.0004
scTranslator-FT†	0.6027±.0000	0.5557±.0000	0.6196±.0000	0.5389±.0000
cTP-net	0.3332±.0091	0.7935±.0030	0.3489±.0110	0.7839±.0048
sciPENN†	17.2086±.0000	0.6013±.0000	16.7941±.0000	0.6150±.0000
totalVI†	3.5421±.0000	0.6581±.0000	3.1831±.0000	0.6256±.0000
scMOG†	18.4375±.0000	0.4912±.0000	17.8560±.0000	0.4686±.0000
MultiPert	0.6327±.0074	0.5349±.0016	0.6212±.0038	0.5399±.0038
Ridge	0.3925±.0000	0.7519±.0000	0.4103±.0000	0.7430±.0000
PertMean†	0.6119±.0000	0.5442±.0000	0.5560±.0000	0.5024±.0000

Table 7: Frangieh complete comparison. Values are mean ± sample standard deviation over three train seeds when available. Rows marked † have one reported run; other rows summarize three seeds. A displayed ±0.0000 may reflect rounding and does not by itself identify the run count.

Method	Overall MSE	Overall r	RH MSE	RH r
scPathCF	**0.2960±.0006**	**0.9501±.0001**	**0.3115±.0003**	**0.9473±.0000**
GEARS	0.3302±.0049	0.9444±.0006	0.3450±.0039	0.9415±.0005
N-gram ridge	0.3148±.0001	0.9469±.0000	0.3212±.0001	0.9456±.0000
scTranslator-FT†	0.3968±.0000	0.9325±.0000	0.4099±.0000	0.9300±.0000
cTP-net	0.3005±.0019	0.9495±.0005	0.3139±.0011	0.9471±.0003
sciPENN†	5.2008±.0000	0.2712±.0000	5.2139±.0000	0.2621±.0000
scMOG†	3.9348±.0000	0.9117±.0000	4.0073±.0000	0.9144±.0000
MultiPert	0.4508±.0003	0.9230±.0001	0.4582±.0021	0.9213±.0004
Ridge	0.3044±.0000	0.9487±.0000	0.3170±.0000	0.9463±.0000
PertMean†	0.4490±.0000	0.9233±.0000	0.4357±.0000	0.9255±.0000

Table 8: CaRPool complete comparison. Values are mean ± sample standard deviation over three train seeds when available. Rows marked † have one reported run; other rows summarize three seeds. A displayed ±0.0000 may reflect rounding and does not by itself identify the run count.

Method	Overall MSE	Overall r	RH MSE	RH r
scPathCF	**0.3292±.0186**	**0.8395±.0090**	**0.3384±.0122**	**0.8552±.0054**
GEARS	0.3418±.0110	0.8337±.0009	0.3620±.0168	0.8479±.0015
N-gram ridge	0.3377±.0003	0.8363±.0001	0.3542±.0004	0.8530±.0001
scTranslator-FT†	0.7479±.0000	0.5644±.0000	0.7399±.0000	0.6413±.0000
cTP-net	0.8898±.0325	0.5585±.0125	0.8412±.0817	0.6211±.0399
sciPENN†	24.8267±.0000	0.2967±.0000	24.9977±.0000	0.3073±.0000
totalVI†	7.8016±.0000	−0.1309±.0000	7.7928±.0000	−0.1033±.0000
scMOG†	28.2678±.0000	0.2748±.0000	27.9393±.0000	0.3353±.0000
MultiPert	0.3773±.0168	0.8147±.0035	0.3956±.0171	0.8312±.0028
Ridge	1.0446±.0000	0.5155±.0000	1.0317±.0000	0.5579±.0000
PertMean†	0.3604±.0000	0.8182±.0000	0.3811±.0000	0.8329±.0000

C UNSEEN PERTURBATION IDENTITIES

CaRPool holds out the three-gene combination `CD46_CD55_CD71`. Replacing a learned combination ID with the mean of member-gene embeddings changes SCPATHCF MSE from 0.9645 ± 0.2734 to 0.6904 ± 0.0926 and Pearson from 0.6762 ± 0.0835 to 0.7566 ± 0.0264. GEARS and N-gram ridge reach MSE 0.4846 ± 0.0134 and 0.5661 ± 0.0001, respectively, with Pearson 0.8244 ± 0.0045 and 0.8175 ± 0.0001. Gene-mean composition improves SCPATHCF on this combination, while the listed GEARS and n-gram ridge comparators have lower error.

D PAIRED LEAVE-ONE-OUT ABLATIONS

Table 10 reports the Frangieh ablation in full. Each row is a paired change relative to the screen-enabled full model on the same split and seed, so a positive entry in an MSE column denotes a larger error after component removal.

The ordering in Table 10 characterizes the contribution of the two paths and the screening prior, while removing either path consistency or the residual-share penalty produces smaller changes in this setting. Table 12 reports the same removals on all three datasets, and the pattern depends on how each screen is built. On Papalexi, removing the mediated head changes overall MSE by 0.1719 while removing the residual head changes it by 0.0007, showing a larger predictive contribution from the RNA branch in this configuration. On CaRPool both paths contribute, with overall MSE changes of 0.1153 and 0.1016; on the RH subset, the residual-head change is 0.0840 compared with -0.0068 for the mediated head. The relative contribution of each branch therefore varies by dataset and evaluation stratum. Frangieh shows smaller changes for both heads and for the screening prior, with low variance across seeds.

D.1 METRIC DEFINITIONS

For n cells and P protein dimensions, mean squared error is

$$\mathrm{MSE} = \frac{1}{nP} \sum_{i=1}^{n} \sum_{j=1}^{P} (\widehat{y}_{ij} - y_{ij})^2 , \tag{6}$$

and Pearson correlation is computed after flattening all cell–protein pairs. If k indexes the resulting vector, then

$$r = \frac{\sum_k (\widehat{y}_k - \bar{\widehat{y}})(y_k - \bar{y})}{\sqrt{\sum_k (\widehat{y}_k - \bar{\widehat{y}})^2} \sqrt{\sum_k (y_k - \bar{y})^2}}. \tag{7}$$

The same definitions are applied to the full test set and to the fixed residual-high subset.

D.2 TRAINING OBJECTIVE IN FULL

Let $d_F = \mathrm{MSE}(\widehat{\mathbf{Y}}_{1M1}, \mathbf{y})$, $d_B = \mathrm{MSE}(\widehat{\mathbf{Y}}_{1M0}, \mathbf{y})$, and d_N be the mismatch prediction's MSE against the same target. With $[a]_+ = \max(a, 0)$, the gradient-carrying terms are

$$\mathcal{L}_{\mathrm{fact}} = a_f d_F, \tag{8}$$

$$\mathcal{L}_{\mathrm{path}} = a_c \left[\mathcal{L}_{\mathrm{anchor}} + \mathrm{MSE}(\widehat{\mathbf{Y}}_{0M0}, \mathbf{y}_0) + \lambda_s \mathcal{L}_{\mathrm{struct}} \right], \tag{9}$$

$$\mathcal{L}_{\mathrm{mm}} = a_m [d_F - d_N + \mu]_+. \tag{10}$$

Here a_f, a_c, a_m are the configured factual and cell-stratum weights. The anchor is the mean squared *raw* residual on control cells, before the hard control mask; it is zero when no controls are present. The implementation also evaluates the three additive-identity errors implied by equation 4. Those errors vanish algebraically and serve as numerical consistency checks.

The structural term compares prediction errors, rather than ordering protein coordinates:

$$\mathcal{L}_{\mathrm{struct}} = \begin{cases} \beta_{\mathrm{RH}} d_B + [d_F - d_B + 0.02]_+, & \mathrm{RH}, \\ [d_F - d_B + 0.05]_+, & \mathrm{RC}, \\ 0.25\{d_B + [d_F - d_B + 0.02]_+\}, & \text{unstratified}. \end{cases} \tag{11}$$

Table 9: Unseen-identity results. CaRPool holds out `CD46_CD55_CD71`, so its overall and RH sets coincide. Values are mean ± sample standard deviation over three train seeds. The archived Frangieh baseline entries are copied from the unseen-guide evaluation and are retained as reference values, not independently rerun identity-holdout results. The sciPENN-style rows use a local RNA encoder–decoder, distinct from the sciPENN rows in the complete unseen-guide tables. On Papalexi, holding out perturbation identities leaves SCPATHCF ahead of cTP-net, with a smaller margin than under the unseen-guide split.

Dataset	Method	Overall MSE	Overall r	RH MSE	RH r
Papalexi	SCPATHCF	**0.3184±.0048**	**0.7845±.0021**	**0.3950±.0165**	0.7498±.0070
	cTP-net	0.3305±.0035	0.7763±.0088	0.4034±.0156	0.7434±.0152
	sciPENN-style	0.3318±.0184	0.7838±.0013	0.3994±.0507	**0.7594±.0085**
	Ridge	0.3847±.0000	0.7377±.0000	0.4566±.0000	0.7072±.0000
	GEARS	0.4950±.0033	0.6275±.0026	0.5646±.0202	0.5859±.0034
	N-gram ridge	0.4960±.0004	0.6263±.0004	0.5685±.0009	0.5903±.0008
Frangieh	SCPATHCF	**0.2959±.0007**	**0.9502±.0001**	**0.3110±.0002**	**0.9474±.0000**
	cTP-net	0.3005±.0019	0.9495±.0005	0.3139±.0011	0.9471±.0003
	N-gram ridge	0.3148±.0001	0.9469±.0000	0.3212±.0001	0.9456±.0000
	GEARS	0.3302±.0049	0.9444±.0006	0.3450±.0039	0.9415±.0005
CaRPool	SCPATHCF (gene-mean)	0.6904±.0926	0.7566±.0264	0.6904±.0926	0.7566±.0264
	SCPATHCF (learned ID)	0.9645±.2734	0.6762±.0835	0.9645±.2734	0.6762±.0835
	GEARS	**0.4846±.0134**	**0.8244±.0045**	**0.4846±.0134**	**0.8244±.0045**
	N-gram ridge	0.5661±.0001	0.8175±.0001	0.5661±.0001	0.8175±.0001
	cTP-net	0.8898±.1421	0.6874±.0155	0.8898±.1421	0.6874±.0155
	sciPENN-style	0.8883±.0463	0.6958±.0134	0.8883±.0463	0.6958±.0134

Table 10: Frangieh leave-one-out ablation, reported as the paired change against the full model. Values are mean ± sample standard deviation over the available train seeds, and the sign convention is ablation minus full, so a positive MSE entry and a negative correlation entry both indicate a larger error or lower correlation after removal. The full model reaches 0.2960 overall MSE and 0.3115 residual-high MSE on this dataset.

Removed component	ΔMSE ↓	Δr ↑	RH ΔMSE ↓	RH Δr ↑
w/o drop $\mathbf{z}_C$	+.0467 ± .0559	−.0079 ± .0094	+.0277 ± .0458	−.0047 ± .0078
w/o screening prior	+.0195 ± .0020	−.0033 ± .0003	+.0174 ± .0030	−.0029 ± .0005
w/o mediated head	+.0181 ± .0018	−.0030 ± .0003	+.0169 ± .0013	−.0029 ± .0002
w/o residual head	+.0175 ± .0007	−.0029 ± .0001	+.0140 ± .0008	−.0024 ± .0001
w/o path loss	+.0002 ± .0001	**−.0000 ± .0000**	−.0004 ± .0005	**+.0001 ± .0001**
w/o leftover LR	**+.0000 ± .0001**	**−.0000 ± .0000**	**−.0005 ± .0005**	**+.0001 ± .0001**

The configured RH mode sets β_{RH} to 0, 0.25, or 1. Strict-zero runs use the unstratified case. The mismatch replaces factual dense and token RNA with another sampled cell's RNA; that cell need not have a different perturbation. Identity, prior, prototype, and matched controls remain fixed. The margin μ is configuration-specific (0.1 for the Papalexi strict-zero run).

The regularizer $\mathcal{R}$ contains residual-gate sparsity and the leftover fit

$$\mathcal{L}_{\mathrm{left}} = \mathrm{MSE}\Big(\mathbf{r}, \mathbf{y} - \mathrm{stopgrad}(\widehat{\mathbf{Y}}_{1M1} - \mathbf{r})\Big)\,. \tag{12}$$

The latter trains the residual against the error left by the other paths; it is not an orthogonality penalty. In screen-enabled configurations, $\mathcal{R}$ can also include an RH gate floor, an RH correlation term, and a residual-share prior with configuration-specific lower and upper targets. The sparsity multiplier is 0.35 on RH cells and 1 otherwise. Any configured vector-quantization loss is added here. Strict-zero runs clear the RH/RC labels and disable the screen-dependent regularizers and oversampling.

All three losses are evaluated in each update. For the first two Phase-B epochs, the configured boost b changes the base weights to

$$(\widetilde{w}_f, \widetilde{w}_p, \widetilde{w}_m) = \big(w_f/b, \min(0.7, bw_p), \min(0.35, bw_m)\big), \qquad w_j^{(e)} = \widetilde{w}_j / \sum_k \widetilde{w}_k. \tag{13}$$

Subsequent epochs use the base weights. This loss schedule is separate from training-cell sampling and from Phase-A dense-head pretraining. The Papalexi strict-zero configuration uses $b = 1.25$, $a_f = 2.2$, $\lambda_{\mathrm{path}} = 1$, $\lambda_s = 0.4$, sparsity coefficient 0.06, and leftover coefficient 1.

Table 11: Loss-weight sweep on the three datasets. Each block moves the factual, path-intervention and mismatch weights across the unit interval with their sum held fixed, and the row marked with ⋆ is the reported configuration for that dataset. The sweep uses a uniform reduced budget for every row so that the rows are directly comparable, while the main tables report the full-budget runs.

Dataset	Weight	Value	Overall MSE ↓	Overall r ↑	RH MSE ↓
Papalexi	loss weights	0.95/0.04/0.01	**0.3182**	**0.7954**	**0.3447**
	loss weights	0.85/0.11/0.04	0.3202	0.7946	0.3471
	loss weights	0.75/0.18/0.07	0.3255	0.7915	0.3479
	loss weights	0.55/0.30/0.15⋆	0.3224	0.7937	0.3501
	loss weights	0.50/0.33/0.17	0.3234	0.7926	0.3471
	loss weights	0.34/0.44/0.22	0.3292	0.7889	0.3508
	loss weights	0.20/0.53/0.27	0.3223	0.7947	0.3481
Frangieh	loss weights	0.95/0.04/0.01	0.3314	0.9445	0.3649
	loss weights	0.85/0.11/0.04	0.3319	0.9443	0.3640
	loss weights	0.75/0.18/0.07	0.3291	0.9448	0.3635
	loss weights	0.68/0.20/0.12⋆	**0.3276**	**0.9449**	0.3616
	loss weights	0.50/0.33/0.17	0.3304	0.9447	0.3614
	loss weights	0.34/0.44/0.22	0.3322	0.9444	0.3660
	loss weights	0.20/0.53/0.27	0.3278	**0.9449**	**0.3581**
CaRPool	loss weights	0.95/0.04/0.01	**0.2861**	**0.8592**	**0.2866**
	loss weights	0.85/0.11/0.04	0.2864	0.8591	0.2870
	loss weights	0.75/0.18/0.07	0.2877	0.8586	0.2883
	loss weights	0.55/0.30/0.15⋆	0.2891	0.8578	0.2897
	loss weights	0.50/0.33/0.17	0.2896	0.8576	0.2902
	loss weights	0.34/0.44/0.22	0.2915	0.8566	0.2921
	loss weights	0.20/0.53/0.27	0.2942	0.8552	0.2947

Table 12: Paired change in each metric relative to the archived screen-enabled full model on the same split and seed, for all three datasets. Values are ablation minus full, so a positive MSE entry indicates a larger error after removal. Only the two MSE columns are shown here; the corresponding correlation deltas for Frangieh appear in Table 10. Values are mean ± sample standard deviation over the available train seeds. The archived screen-enabled reference reaches 0.3164 overall MSE on Papalexi, 0.2960 on Frangieh, and 0.3292 on CaRPool.

	Papalexi		Frangieh		CaRPool	
Variant	Overall MSE	RH MSE	Overall MSE	RH MSE	Overall MSE	RH MSE
w/o drop $\mathbf{z}_C$	+.0052±.0090	+.0013±.0094	+.0467±.0559	+.0277±.0458	−.0080±.0069	−.0095±.0059
w/o leftover LR	−.0008±.0040	−.0008±.0091	**+.0000±.0001**	**−.0005±.0005**	−.0014±.0108	+.0009±.0068
w/o screening prior	**−.0029±.0043**	**−.0035±.0059**	+.0195±.0020	+.0174±.0030	**−.0102±.0087**	**−.0111±.0059**
w/o path loss	+.0071±.0075	+.0009±.0071	+.0002±.0001	−.0004±.0005	+.0016±.0103	+.0054±.0096
w/o residual head	+.0007±.0038	+.0020±.0059	+.0175±.0007	+.0140±.0008	+.1016±.0459	+.0840±.0515
w/o mediated head	+.1719±.0093	+.1666±.0121	+.0181±.0018	+.0169±.0013	+.1153±.0926	−.0068±.0128

D.3 Screening Prior Construction

When enabled, the perturbation-level vector supplied to the residual head is built by the training script from an offline screen and never from cell-level factual RNA at inference. Branch exclusion and statistical independence are distinct: dependence on the evaluation target is determined by the screen's provenance. With the configured width of eight, the vector has a fixed layout: The legacy oracle screen is computed from the same measurements as the prediction target: for each perturbation and measured protein it records the cognate-transcript and protein changes, then labels a perturbation residual-high when the protein moves while the transcript is stable. It is therefore a target-derived analysis variable rather than an independent annotation. The implementation records the construction mode and source path for every run. In `train_only` mode the screen is estimated only from training cells; `external` requires an explicit supplied file; `oracle` reproduces the legacy full-data screen; and `zero` sets all channels to zero and disables every screen-dependent loss and sampler. Table 3 shows similar overall Papalexi accuracy across the prior modes. CaRPool uses explicitly curated strata because its original feature identifiers do not yield a reliable automatic cognate-RNA screen for every measured protein.

Table 13: Channels of the eight-dimensional screening-prior vector.

Channel	Content
0	1 if the perturbation is residual-high in the offline screen, else 0
1	1 if the perturbation is RNA-coupled, else 0
2	the same residual-high flag repeated, as an open-gate bias
3	the same RNA-coupled flag repeated, as a mediated-preference bias
4	the screen residual score, clipped to $[-3, 3]$ when scores are available
5–7	unused padding

Table 14: Alternative metrics from the archived Papalexi prediction set. These values summarize that set separately from the strict-zero results in Table 2.

Method	Flattened r	Macro per-protein r	Differential r
sciPENN	0.8004	**0.6993**	**0.7038**
cTP-net	**0.8007**	0.6979	0.7017
scPathCF	0.7996	0.6941	0.7000
Ridge (RNA)	0.7519	0.6231	0.6288
N-gram ridge	0.6697	0.4433	0.4694
GEARS	0.6606	0.4292	0.4547
Control	0.6353	0.3844	0.4046
scTranslator-FT	0.5187	0.1085	0.0819
MultiPert	0.5226	0.0601	0.0747

D.4 Additional Metric Views

The flattened correlation in equation 7 pools every cell and protein into one vector, so it is dominated by the constitutive abundance differences between antibody targets rather than by the perturbation effect. Two views that remove that component were computed from the same archived predictions on Papalexi. The macro per-protein correlation averages the correlation computed separately for each protein across test cells, and the differential-effect correlation is taken on $\Delta y = y - \bar{y}_0$, with $\bar{y}_0$ the per-protein control mean, which subtracts the static hierarchy before correlating. The three translation baselines remain within 0.001 of each other on the flattened correlation and within 0.006 on both alternative views. These diagnostics show that metric choice changes the ordering among translation models with similar aggregate values. The larger separation among the remaining methods is consistent with the flattened number being dominated by the static abundance hierarchy rather than by perturbation-specific signal.

E ADDITIONAL FIGURES

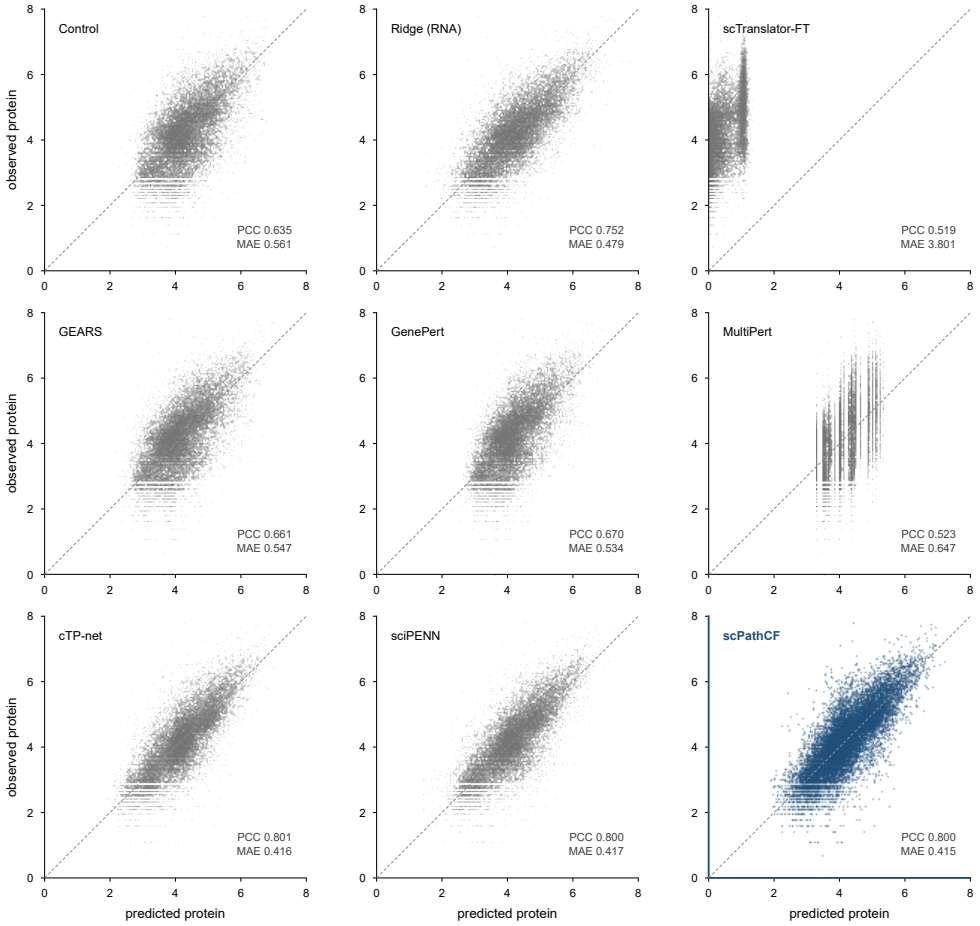

Figure 4: Observed and predicted protein abundance in the archived Papalexi perturbed-cell diagnostic, flattened across cells and proteins. The dashed line is the identity; each panel reports MAE and Pearson correlation. The panel labelled GenePert uses the n-gram ridge comparator. These aggregate metrics summarize abundance prediction and do not measure the accuracy of the branch assignments.

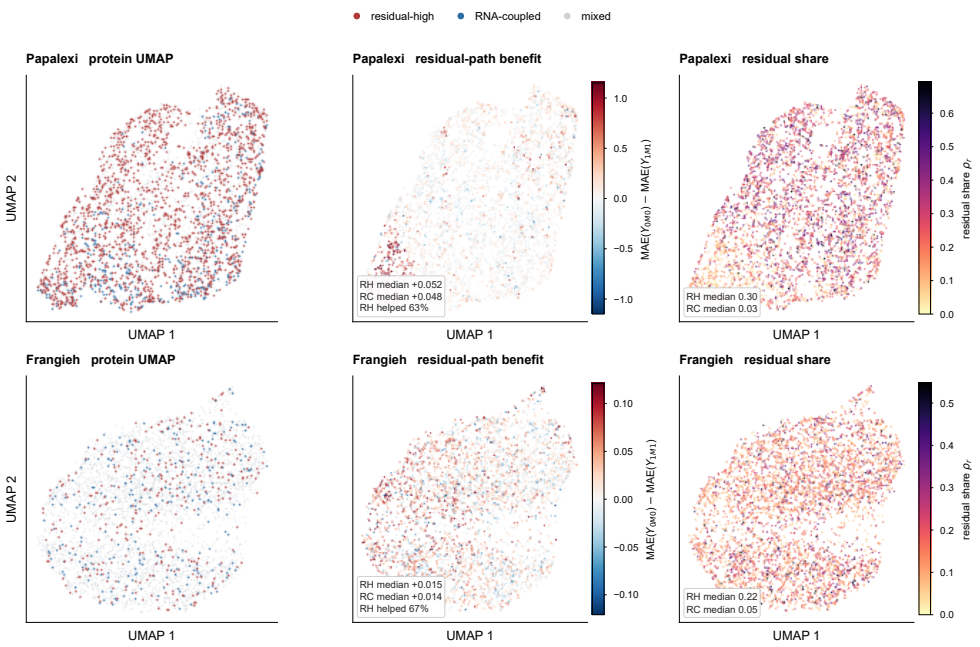

Figure 5: Protein UMAP of test cells. The left panel shows the residual-high and RNA-coupled strata; the middle panel shows $\mathrm{MAE}(\hat{\mathbf{Y}}_{0M0}) - \mathrm{MAE}(\hat{\mathbf{Y}}_{1M1})$; and the right panel shows the residual share. The middle contrast measures the combined prediction change from both branches, because it replaces control RNA with factual RNA and enables the residual branch. The strata are derived from target measurements.

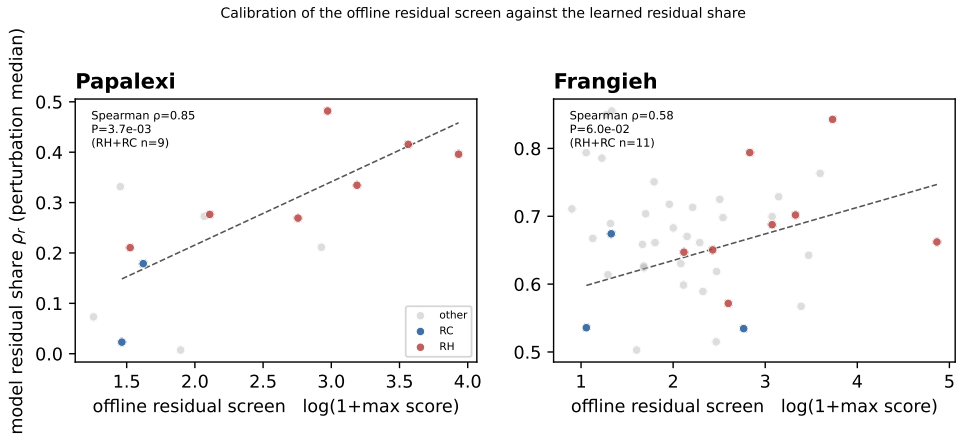

Figure 6: Offline screen score and fitted residual share. Each point is a perturbation; the horizontal axis is the log screen score and the vertical axis is the median residual share over the plotted cells. Screen annotations also enter the screen-enabled model and its training constraints, so this plot measures agreement with the screen.

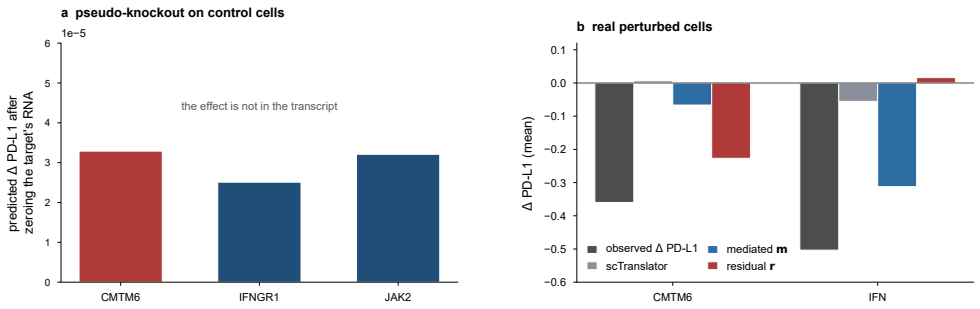

Figure 7: RNA-input pseudo-knockout and predictions on measured perturbations. Left: changing selected RNA inputs in non-targeting cells produces near-zero predicted PD-L1 changes for CMTM6, IFNGR1, and JAK2. This input intervention does not distinguish the three perturbations. Right: observed PD-L1 changes in perturbed cells, scTranslator predictions, and the fitted components of scPathCF; the fitted CMTM6 response lies mainly in the residual branch.

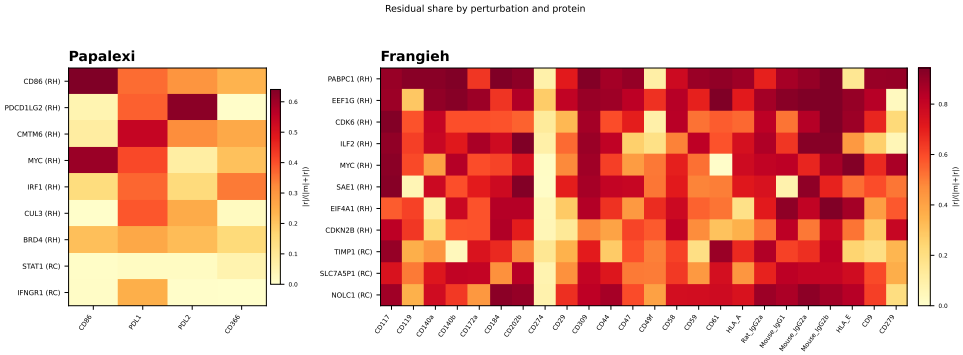

Figure 8: Archived residual-share profiles across the plotted ADT panels. Rows are representative residual-high and RNA-coupled perturbations; columns are measured ADTs. The Frangieh panel contains 24 ADTs, whereas the main prediction tables evaluate the 14 backbone-supported proteins. The panel-level display therefore covers a different set of outputs from the main comparison.

F Paired Uncertainty for the Main Comparison

We added a post-hoc analysis of the three saved training-seed results for each row of Table 2. For each baseline, we pair runs by seed label $s \in \{0, 1, 2\}$ on the reported fixed split and define $d_s = \mathrm{MSE}_{\mathrm{scPathCF},s} - \mathrm{MSE}_{\mathrm{baseline},s}$. Negative values favour scPathCF. We report $\bar{d} \pm t_{0.975,2} s_d/\sqrt{3}$ and a two-sided paired t-test, with Holm adjustment over the 12 overall-MSE comparisons. The intervals are individual 95% intervals, not simultaneous intervals. They describe training-seed variation, conditional on the split and each method's reported input, prior, and training budget; they are not cell- or perturbation-level confidence intervals.

The closest listed comparators are cTP-net on Papalexi and Frangieh and N-gram ridge on CaRPool. Their intervals all cross zero. Three seed pairs give little information about the distribution of training variation; the exact sign-flip test cannot attain a two-sided probability below 0.25. We therefore treat the t-based results as conditional uncertainty summaries, rather than evidence of a general performance ordering. No noninferiority margin was specified, and these tests do not establish equivalence or noninferiority. The Frangieh prior-access asymmetry in Table 2 also applies here.

The Papalexi values come from the strict-zero checkpoints and their saved comparison records, whose test-index hashes agree across seeds. Baseline values and the Frangieh/CaRPool rows come from the archived main-comparison records. Pairing uses the saved seed labels and matching reported test-cell counts; per-cell baseline predictions were not available for an independent paired perturbation analysis.

Table 15: Post-hoc paired training-seed differences for Table 2. $n = 3$ pairs per row. The t-based analysis assumes normally distributed seed differences. The last column gives the unadjusted exact two-sided seed sign-flip probability as a small-sample sensitivity check.

Dataset	Baseline	ΔMSE	95% CI	p_{Holm}	p_{flip}
Papalexi	cTP-net	-0.0174	$[-0.0533, 0.0185]$	0.518	0.25
Papalexi	N-gram ridge	-0.1634	$[-0.1770, -0.1498]$	0.003	0.25
Papalexi	GEARS	-0.2031	$[-0.2381, -0.1682]$	0.013	0.25
Papalexi	MultiPert	-0.3169	$[-0.3237, -0.3101]$	< 0.001	0.25
Frangieh	cTP-net	-0.0045	$[-0.0106, 0.0015]$	0.340	0.25
Frangieh	N-gram ridge	-0.0188	$[-0.0203, -0.0173]$	0.003	0.25
Frangieh	GEARS	-0.0342	$[-0.0461, -0.0223]$	0.039	0.25
Frangieh	MultiPert	-0.1548	$[-0.1570, -0.1526]$	< 0.001	0.25
CaRPool	cTP-net	-0.5606	$[-0.6866, -0.4345]$	0.019	0.25
CaRPool	N-gram ridge	-0.0085	$[-0.0546, 0.0377]$	0.618	0.75
CaRPool	GEARS	-0.0126	$[-0.0527, 0.0275]$	0.618	0.50
CaRPool	MultiPert	-0.0480	$[-0.0868, -0.0093]$	0.167	0.25

G RNA-Noise Test of Residual Routing

This post-hoc diagnostic asks whether factual RNA can change the residual output at fixed weights. We load the three Papalexi strict-zero checkpoints and run the full model in evaluation mode. Before measuring outputs, we sample two test cells per available perturbation with seed 20260926, giving 28 cells from 14 perturbations. The reconstructed test-index hash matches the training manifest. Perturbation identity, priors, matched control RNA, prototype RNA, gene identities, masks, and all weights remain fixed.

For noise level $\sigma \in \{0, 0.25, 0.5, 1\}$, we add Gaussian noise to factual dense RNA, scaled by each gene's training-set standard deviation, and to packed RNA values, scaled by the within-cell standard deviation of unmasked values. Dense RNA is clipped below at zero and packed values to $[0, 1]$; padding is unchanged. We reuse the same noise direction across levels and checkpoints. A change in the residual would falsify the implemented inference-time exclusion of these RNA inputs. The mediated response is a positive control for an effective input intervention.

Table 16: Frozen-checkpoint RNA-noise diagnostic. Residual change is the maximum absolute change over all tested cells, proteins, and checkpoints. Mediated change is the cell/protein mean absolute change from the uncorrupted output, reported as mean ± standard deviation across three checkpoints.

Noise σ	Max. $\lvert\Delta\mathbf{r}\rvert$	Mean $\lvert\Delta\mathbf{m}_1\rvert$
0	0	0.0000 ± 0.0000
0.25	0	0.0839 ± 0.0110
0.5	0	0.1656 ± 0.0122
1	0	0.3258 ± 0.0032

The residual is exactly invariant in these forward passes, while the mediated response grows with the input corruption. This result checks residual routing in the saved implementation. It does not establish a biological mechanism or test whether retraining on noisier RNA would increase the learned residual contribution. The identity-indexed prototype remains fixed, so the diagnostic also does not establish identity exclusion from the mediated path.