scPLAD: Multi-source biological prior-guided latent diffusion for single-cell population generation under genetic perturbations

Zhigang Zhang[1,2,3], Xiaobo Zhu[1,2,3], Yu-An Huang[4], Lu You[1,2,3]
Pengyu Chen[1,2,3], Xun Deng[1,2,3], Lun Hu[1,2,3], Thomas Herget[5]
Feng Tan[6], Chenke Xu[7], Zhu-Hong You[4], Jianying Hu[7]
Huili Hu[8,9], Pengwei Hu[1,2,3,*]

[1]Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Urumqi 830011, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China
[3]Xinjiang Laboratory of Minority Speech and Language Information Processing, Urumqi 830011, China
[4]School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China
[5]Technical University of Darmstadt, Darmstadt, Germany
[6]AI Hub, Merck KGaA, Darmstadt, Germany
[7]Department of Environmental Science and Engineering, Fudan University, Shanghai 200433, China
[8]Key Laboratory of Experimental Teratology, Ministry of Education, Department of Systems Biomedicine, School of Basic Medical Sciences, Shandong University, Jinan 250012, China
[9]Qilu Hospital, Cheeloo College of Medicine, Shandong University, Jinan 250012, China

[*]Correspondence and requests for materials should be addressed to Pengwei Hu (hpw@ms.xjb.ac.cn).

Motivation

Single-cell perturbation screens reveal diverse cellular responses, but measuring every perturbation in every cell type is difficult. Predicting unobserved responses requires models that use biological knowledge and transfer information across cellular contexts while preserving variation among cells. We developed scPLAD to generate perturbation response populations by combining biological priors with target cell context.

Summary

Single-cell genetic perturbation screens reveal average transcriptional changes and heterogeneous responses among cells exposed to the same perturbation. Predicting these population responses remains challenging for perturbations unseen during training. Prediction is also difficult in target contexts without measured perturbation responses. Here we present scPLAD, a generative framework integrating multi-source biological priors for perturbed genes with the target cellular context to model distributions of single-cell responses. In K562 cells, scPLAD generated responses to unseen perturbations while recovering perturbation directions and population-level distributional structure. Across cell lines, scPLAD generated heterogeneous expression distributions characteristic of the target context without training on perturbed target cells. These results support the use of biological priors and target context information to predict responses to unseen perturbations and transfer responses across cell lines. By separately encoding perturbation semantics and cellular context and explicitly modeling their interaction, scPLAD provides a strategy for incorporating prior knowledge into virtual cell models.

Introduction

Large-scale single-cell perturbation screens are reshaping experimental studies of gene function and cellular state responses. Perturb-seq and CROP-seq combine pooled CRISPR perturbations with single-cell RNA sequencing, enabling perturbation identities to be linked to transcriptomic responses across large numbers of cells.[1–4] Related approaches, including sci-Plex and cytokine stimulation assays such as IFN-β treatment, have extended this strategy to small molecules and extracellular signals.[5,6] These data provide systematic resources for investigating gene function, regulatory relationships and cellular state transitions, and have supported studies of disease mechanisms, drug response, cell state conversion and fate control.[3,5,7–11] Nevertheless, even for genetic interventions alone, the number of possible perturbations and combinations grows rapidly with the number of candidate genes, making exhaustive experimental characterization impractical.[12] A central computational objective is therefore to learn perturbation response relationships from observed conditions and predict or generate single-cell expression distributions for new perturbations or cellular contexts.[12,13]

Existing methods address this problem from complementary perspectives. scGen learns perturbation vectors in an autoencoder latent space to predict responses in unobserved cellular contexts,[13] whereas the compositional perturbation autoencoder represents perturbations, doses and covariates as composable factors.[14] CellOT treats unpaired control and perturbed cells as distributions and learns a neural optimal-transport map between them.[15] For genetic perturbations, GEARS incorporates prior gene relationships to predict multigene and unseen combinatorial perturbations,[12] and TxPert integrates multiple knowledge graphs to predict transcriptomic effects of unseen perturbations.[16] Semantic gene representations have also been used for extrapolation: Scouter predicts unseen genetic perturbations using gene embeddings derived from language models,[17] whereas scLAMBDA incorporates foundation model gene embeddings

into a generative framework for single-gene and multigene perturbations.[18] STATE models perturbation effects across sets of cells and supports prediction across cellular contexts,[19] while PIE predicts population-level perturbation effects using external biological knowledge, baseline expression and response evidence from observed contexts.[20] Other generative approaches, including scDiffusion, Squidiff, CellFlow and scLDM, model cell distributions after perturbation using diffusion, flow matching or latent diffusion.[21–24] Together, these studies have expanded perturbation modeling from mean expression prediction towards conditional generation of cellular populations.

Distributional generation is more demanding than prediction of a single mean expression vector. A perturbation changes a population of cells, within which individual cells can exhibit different response magnitudes, state shifts and technical noise patterns.[13,15,25] Recent benchmarks further show that means or medians from the training set, predictors that assume no perturbation effect, additive models and linear baselines can remain competitive with more complex deep models.[26,27] Evaluating a generated perturbation response therefore requires complementary measurements of pseudobulk agreement, perturbation direction, conditional heterogeneity and distributional structure. Diffusion, flow matching and latent generative models provide a natural basis for this objective because they represent cell populations through conditional sampling.[21,23,24,28–30]

Generalization to unseen perturbations and cellular contexts also depends on how perturbations are represented. Discrete perturbation identifiers support memorization of observed conditions, whereas biological representations can connect genes through shared functions, pathways and regulatory contexts and can aid interpretation of generated responses. Related work on chemical perturbation transcriptomics has shown that biologically structured knowledge graph embeddings can support inference when compounds or targets are unseen.[31] We therefore represent each perturbed gene with a vector assembled from multiple public biological resources. The representation combines functional annotations, pathway context, protein sequence, interaction networks and regulatory information. It represents genes in a common feature space for perturbation prediction.

scPLAD is a biologically conditioned latent generative framework for single-cell perturbation responses. It maps expression profiles to a reconstructable latent space and generates perturbed cell populations conditioned on the biological priors of the perturbed gene and the control context of the target cells. We evaluate scPLAD on unseen perturbations in K562 and on transfer across cell lines with K562 reserved as the target context. Across both tasks, the evaluation covers mean expression, perturbation direction, response gene recovery and the distributional structure of the population.

Results

Cell line transfer generation requires priors and context

scPLAD formulates single-cell perturbation prediction as conditional population generation. Given a target cellular context and a perturbed gene, the model generates a set of single-cell expression profiles after perturbation that approximates the conditional distribution of the measured perturbed population. Training data are collected from several source cell lines, for which the sets of measured perturbations are only partially overlapping. Functional annotations, pathways, protein interactions and regulatory relationships provide shared biological semantics across these perturbations. At test time, the target cell line contributes control cells that define its basal context, and the model transfers response patterns learned in source cell lines to this target background (Fig. 1a).

This setting evaluates two related forms of extrapolation. Perturbations observed in source cell lines have measured responses in one or more source contexts but not in the target cell line, testing transfer of a learned perturbation effect across contexts. Perturbations unobserved in source cell lines lack measured responses in every source context, jointly testing context transfer and semantic extrapolation. Generation for both groups combines transferable biological priors with the control context of the target cell line.

scPLAD separates representation, priors and generation

scPLAD comprises three connected modules. First, the pathway-aware PatchAE is an autoencoder that uses pathway-organized gene patches as local input units. It maps each expression profile to latent tokens that can be reconstructed by a frozen decoder (Fig. 1b), providing a reconstructable latent space for the downstream generative model.

Second, a perturbed gene is encoded using public biological priors from GO, Reactome, ESM3, PPI, gene regulatory networks, OmniPath and protein complexes (Fig. 1c). This representation provides a shared semantic interface for perturbations observed and unobserved in source cell lines, allowing the model to relate perturbations through functional and network context.

Third, the generative model predicts a latent displacement relative to a control anchor (Fig. 1d). An AdaLN-conditioned denoiser models the displacement distribution and integrates perturbation priors, control context information and diffusion time throughout the network. During generation, the recovered displacement is added to the clean control anchor and decoded into a perturbed single-cell expression profile. A separately trained mean response predictor conditioned on the priors subsequently refines the mean for each condition by applying the same gene-specific shift to every generated cell within that condition while preserving cell-to-cell deviations. This modular design permits separate evaluation of PatchAE reconstruction and the contributions of distinct biological priors.

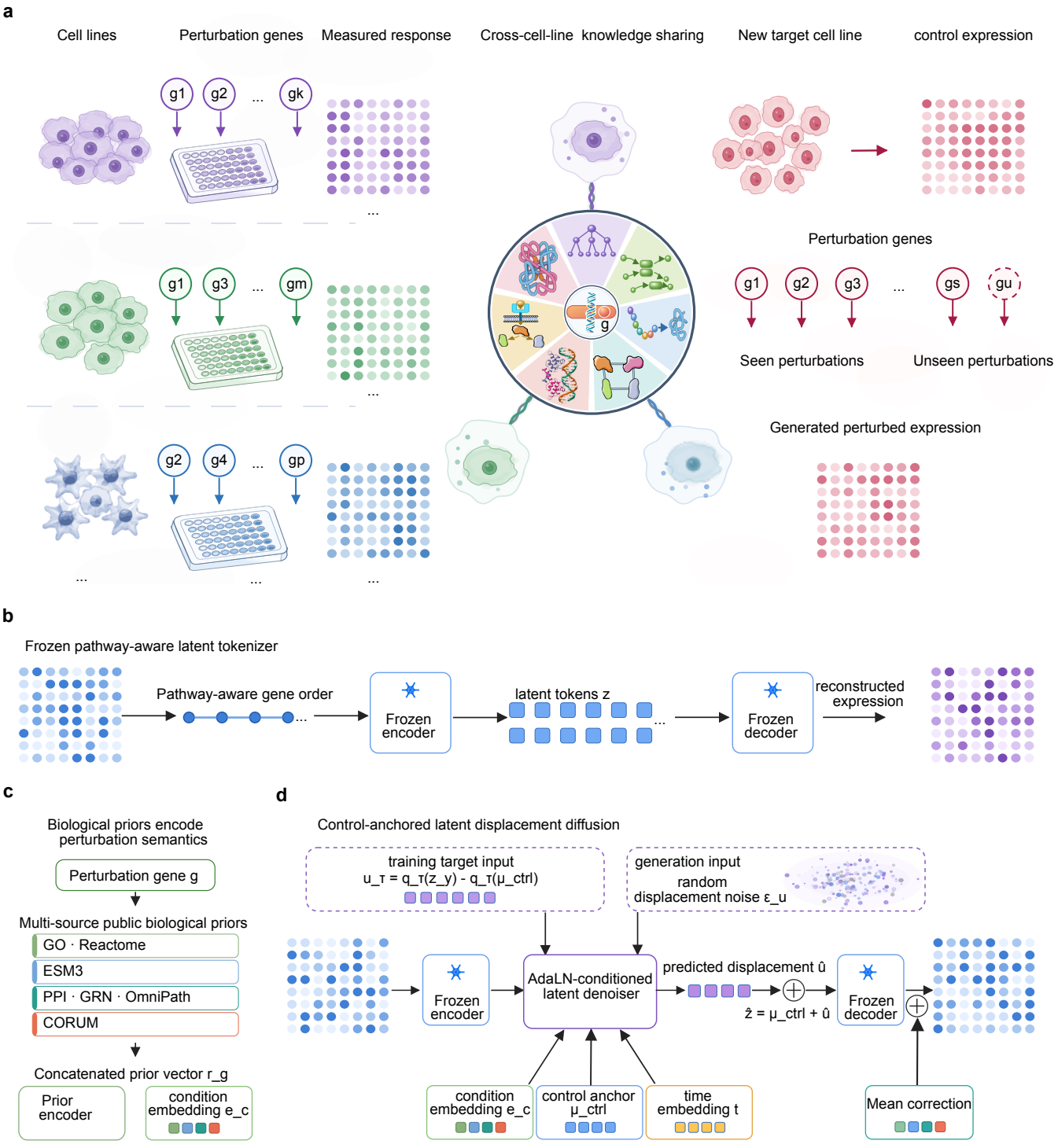


Figure 1: Task definition and architecture of scPLAD. **a,** Cell line transfer perturbation response generation. The model learns transferable perturbation semantics from genes and measured responses across source cell lines, and generates responses to perturbation in a target cell line for perturbations either observed or unobserved in the source cell lines. **b,** Frozen pathway-aware latent tokenizer. Expression profiles are organized by a pathway-aware gene order, encoded into latent tokens and reconstructed by a frozen decoder, providing a reconstructable latent space for displacement diffusion. **c,** Biological prior encoding of perturbations. Each perturbed gene is represented using public priors from multiple sources, including GO, Reactome, ESM3, PPI, gene regulatory networks, Omni-Path and CORUM protein complexes. **d,** Control-anchored latent displacement diffusion. An AdaLN-conditioned latent denoiser generates a perturbation displacement from biological priors and control context information. The recovered displacement is added to the clean control anchor and decoded into a perturbed expression profile.

Multilevel evaluation captures perturbation specificity and population structure

We evaluated generated populations at the levels of mean responses and population structure. Mean PCC, mean Spearman and error measures quantified agreement in average expression. Delta correlations assessed agreement in perturbation effects relative to controls. Top-100 DE overlap and AUPRC assessed recovery of response genes.

We next evaluated population structure at the gene and local cell-state levels. Expression range coverage (ERC) measured how much of the observed expression interval was covered by the generated interval. Interval IoU assessed agreement in interval location and width, while the width ratio distinguished contraction from expansion. Correlation structure agreement (CSA) measured preservation of correlations among response genes. Realistic cell fraction (RCF) measured the fraction of generated cells falling within observed cell neighborhoods. Observed cell coverage (OCC) measured the fraction of observed neighborhoods reached by generated cells. Their harmonic mean defined the fidelity–coverage score (FCS), jointly summarizing local fidelity and coverage.

Finally, we assessed broader distributional differences and recovery of subpopulations. GGE[32] Wasserstein, MMD^2 and Energy distances compared observed and generated distributions along components of a shared PCA representation. Complete total variation distance (Complete TVD) quantified differences in subpopulation proportions, including the fraction of unassigned generated cells. Cluster detection rate measured the fraction of observed subpopulations recovered among generated cells. OCC within clusters measured coverage of observed cell states within each subpopulation. These metrics distinguish recovery of subpopulation proportions from coverage within subpopulations.

PatchAE provides a reconstructable latent space

Before training displacement diffusion, we pretrained a separate PatchAE for each task to accommodate the different gene spaces. Both encoders and decoders were then frozen during diffusion training.

In the cell line transfer task, PatchAE reconstructed source validation cells with a PCC of 0.9963. A separate reconstruction check on K562 control cells included in autoencoder training yielded a PCC of 0.9957 (Supplementary Table S4). The processed expression distributions contained many zero values and a long tail of nonzero values (Fig. 2c). Latent distributions differed between the two training settings (Fig. 2d). The K562 model, trained with weak Gaussian moment regularization, showed latent values closer to zero mean and unit scale. The cell line transfer model, trained without this regularization, showed a broader latent distribution.

We next examined local structure using a shared reference UMAP containing K562 controls and held-out HSPA9, NCBP2 and SLC39A9 perturbation cells. PatchAE reconstructions retained the major regions occupied by observed cells and their relative positions in this embedding (Fig. 2e). Reconstructions from a global VAE with 128 latent dimensions showed more

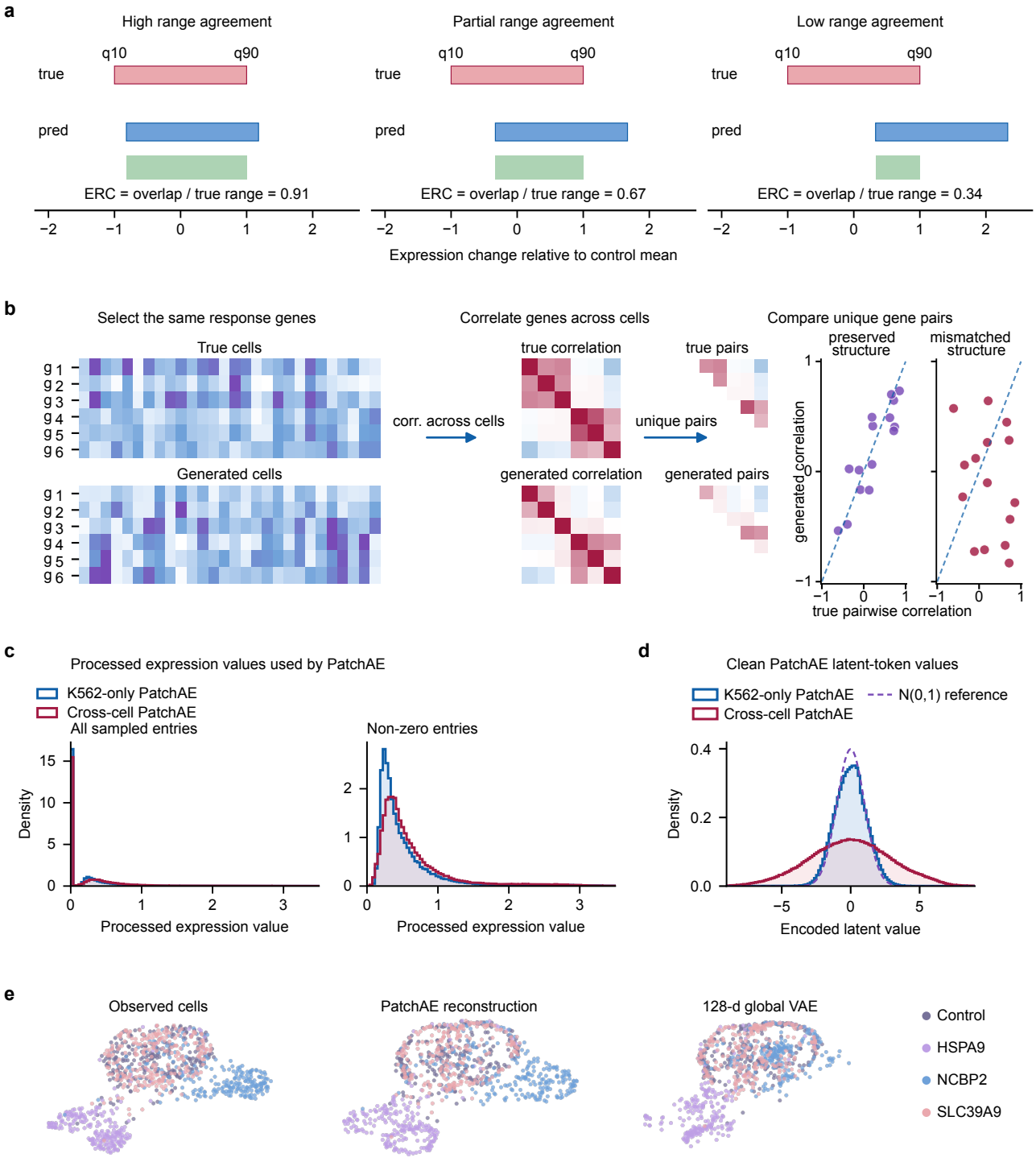


Figure 2: Population-level metrics and diagnostics of PatchAE representations. **a,** ERC is the intersection between generated and true 10th to 90th percentile expression intervals divided by the length of the true interval. **b,** CSA compares pairwise gene correlation matrices computed over top response genes. **c,** Distributions of processed expression values in held-out validation cells for the two PatchAE settings, shown for all sampled entries and for nonzero entries. **d,** Distributions of clean encoded latent token values for the two PatchAE models; the dashed curve denotes an $N(0, 1)$ reference. **e,** A shared-reference UMAP comparing observed cells with PatchAE and 128-dimensional global-VAE reconstructions. PCA50 and UMAP were fitted only to observed cells, and both reconstructions were projected through the same transformations.

apparent displacement and compression. These visual comparisons supported the preservation of local cell state organization in PatchAE reconstructions.

scPLAD recovers unseen K562 perturbation responses

To assess whether scPLAD can predict responses to unseen perturbations in a known cellular context, we evaluated the model on 272 unseen K562 perturbations using three independent random seeds.

Among the evaluated prediction models, scPLAD achieved the highest ERC across all genes and the highest interval IoU across all genes and top-100 response genes. It also achieved the lowest GGE distributional distances (Fig. 3a–g; Supplementary Fig. S2d,f). Its fidelity–coverage score was 0.925, compared with 0.813 for TxPert, 0.756 for CellFlow and 0.751 for Scouter (Fig. 3h; Supplementary Fig. S2g). TxPert achieved higher delta PCC and Top-100 DE overlap, whereas scPLAD better recovered expression intervals and population structure (Supplementary Fig. S2c–g). Scouter achieved slightly higher CSA Pearson than scPLAD, with scores of 0.6504 and 0.6401, respectively (Supplementary Fig. S2e). The unconditional global cloud baseline yielded lower GGE distances but lower perturbation-specific and local support scores (Fig. 3a–h). Together, these comparisons show that mean response, global distribution and local population structure provide complementary views of prediction performance.

We next asked whether scPLAD also recovered the composition and internal structure of individual subpopulations. We evaluated the 255 test conditions containing multiple observed subclusters. scPLAD achieved the lowest mean Complete TVD of 0.172 and the highest mean OCC within clusters of 0.896 (Fig. 3i; Supplementary Fig. S2h). Observed and generated subpopulation proportions across these conditions are shown in Fig. 3j. These results support recovery of both subpopulation proportions and cell states within subpopulations.

Biological priors and gene ordering contribute to distinct response properties

Having established that scPLAD recovers responses to unseen K562 perturbations, we next asked which aspects of its design contribute to this performance. We compared how different combinations of biological priors and gene orders affect prediction performance (Fig. 3k). Adding molecular interaction and regulatory priors to GO+Reactome increased delta PCC from 0.5551 to 0.5630 and Top-100 DE overlap from 0.4577 to 0.4633. These gains support the contributions of molecular interaction and regulatory priors to perturbation effect agreement and response gene recovery (Supplementary Fig. S3c).

ESM3 sequence representations primarily improved expression interval coverage and agreement. Adding ESM3 increased ERC across all genes from 0.8157 to 0.8235 and interval IoU from 0.8231 to 0.8278 (Supplementary Fig. S3b). Changes in perturbation direction and response gene recovery were comparatively modest.

We then compared pathway-aware and original gene orders using the complete prior representation. Mean-response metrics were similar, whereas pathway-aware ordering increased ERC from 0.8069 to 0.8255 and interval IoU from 0.8163 to 0.8307 across all genes. CSA Pearson also increased from 0.6293 to 0.6401, indicating better preservation of correlations among response genes (Supplementary Fig. S3a). Together, these analyses indicate that biological priors and

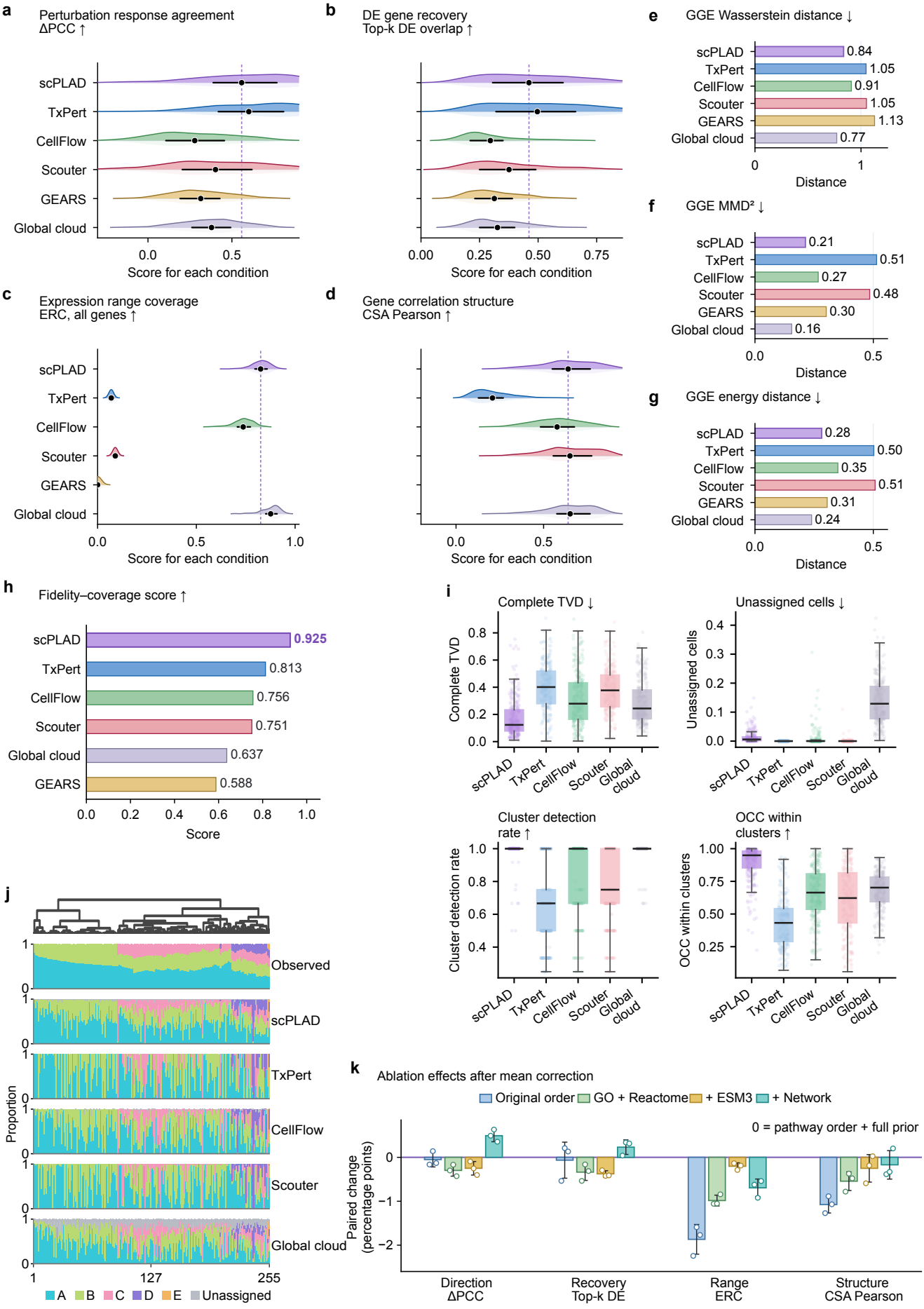

a
Perturbation response agreement
ΔPCC ↑
scPLAD
TxPert
CellFlow
Scouter
GEARS
Global cloud
0.0
0.5
Score for each condition
b
DE gene recovery
Top-k DE overlap ↑
0.00
0.25
0.50
0.75
Score for each condition
c
Expression range coverage
ERC, all genes ↑
0.0
0.5
1.0
Score for each condition
d
Gene correlation structure
CSA Pearson ↑
0.0
0.5
Score for each condition
e
GGE Wasserstein distance ↓
0.84
1.05
0.91
1.05
1.13
0.77
0
1
Distance
f
GGE MMD² ↓
0.21
0.51
0.27
0.48
0.30
0.16
0.0
0.5
Distance
g
GGE energy distance ↓
0.28
0.50
0.35
0.51
0.31
0.24
0.0
0.5
Distance
h
Fidelity–coverage score ↑
0.925
0.813
0.756
0.751
0.637
0.588
0.0
0.2
0.4
0.6
0.8
1.0
Score
i
Complete TVD ↓
Complete TVD
Unassigned cells ↓
Unassigned cells
Cluster detection rate ↑
Cluster detection rate
OCC within clusters ↑
OCC within clusters
scPLAD
TxPert
CellFlow
Scouter
Global cloud
j
Observed
scPLAD
TxPert
CellFlow
Scouter
Global cloud
Proportion
1
127
255
A
B
C
D
E
Unassigned
k
Ablation effects after mean correction
Original order
GO + Reactome
+ ESM3
+ Network
0 = pathway order + full prior
Paired change (percentage points)
0
−1
−2
Direction
ΔPCC
Recovery
Top-k DE
Range
ERC
Structure
CSA Pearson

Figure 3: Benchmarking and ablation analyses for unseen K562 perturbations. **a–d,** Condition-level distributions of perturbation response agreement, DE gene recovery, expression range coverage and gene correlation structure. Black points denote arithmetic means, black bars denote interquartile ranges and purple dashed lines mark the mean of scPLAD. **e–g,** GGE Wasserstein, MMD^2 and Energy distances in PCA50 space; lower values indicate closer distributions. Global cloud denotes an unconditional baseline sampled from the training perturbation-cell pool independently of test-perturbation identity. **h,** Fidelity–coverage score (FCS), the harmonic mean of RCF and OCC; higher values indicate a better joint balance of local fidelity and observed cell coverage. **i,** Recovery of data-driven subpopulations across the 255 test conditions containing more than one observed subcluster. Boxes show the interquartile range, center lines show medians, whiskers extend to 1.5 times the interquartile range and points denote conditions. Complete TVD includes the unassigned fraction; arrows indicate the preferred direction. **j,** Observed and generated subpopulation proportions for the same 255 conditions, hierarchically ordered by observed composition. Subclusters A–E are ranked by observed abundance separately within each condition; unassigned denotes generated cells outside the q99 observed-subcluster support threshold. **k,** Paired changes for alternative biological prior combinations and the original gene order relative to full scPLAD after applying the same fixed mean correction to every setting. Bars show the mean paired change across three random seeds, error bars show the standard deviation and open circles denote individual seeds. The zero line corresponds to pathway-aware gene order with the complete prior representation.

gene ordering contribute to distinct aspects of generated perturbation responses.

scPLAD transfers responses across cell lines

We next tested whether scPLAD could transfer perturbation responses learned in other cell lines to K562. Training used perturbed cells from RPE1, HepG2 and Jurkat, with K562 represented only by control cells. We evaluated 1,086 K562 perturbations using three independent random seeds.

Across all test perturbations, scPLAD achieved a delta PCC of 0.4687 ± 0.0031, a Top 100 DE overlap of 0.3825 ± 0.0026 and an AUPRC of 0.3478 ± 0.0029. Its delta PCC and delta Spearman scores were approximately 0.02 lower than those of TxPert, whereas its Top 100 DE overlap was higher. PIE achieved the highest AUPRC, but its delta PCC and Top 100 DE overlap were lower than those of scPLAD (Fig. 4a; Supplementary Fig. S4a,b). Across these four metrics, scPLAD and TxPert shared the best average rank of 1.75 among the four evaluated methods.

We then assessed whether generated populations recovered subpopulation composition and cell states within subpopulations (Fig. 4b,c). Perturbations were grouped by whether their responses had been observed in at least one source cell line. Of the 1,086 test perturbations, 1,062 contained multiple observed subclusters, comprising 568 source-observed and 494 source-unobserved perturbations. We compared scPLAD with TxPert in both groups and with STATE in the source-observed group, for which STATE provided direct population predictions.

For perturbations observed in at least one source cell line, scPLAD achieved 22.9% lower Complete TVD and 46.9% higher OCC within clusters than STATE. For perturbations absent from all source cell lines, scPLAD achieved 49.5% lower Complete TVD and 104.2% higher OCC within clusters than TxPert (Fig. 4b,c).

To examine the association between source coverage and prediction quality, we grouped perturbations by the number of source cell lines with measured responses. Coverage ranged from zero, indicating absence from all source cell lines, to three, indicating observations in RPE1, HepG2 and Jurkat. From coverage zero to three, delta PCC increased from $0.4278 \pm$

0.0033 to 0.5135 ± 0.0037, and Top-100 DE overlap increased from 0.3664 ± 0.0030 to 0.4002 ± 0.0036. FCS was higher for scPLAD than TxPert in every coverage group and higher than STATE across coverage levels one to three (Fig. 4d). TxPert retained higher perturbation effect correlations, whereas scPLAD achieved comparable or higher response gene recovery at broader source coverage.

We further examined whether predictions for the 506 source-unobserved perturbations captured responses beyond a common source effect. We constructed a baseline that applied the average perturbation effect across source conditions to the K562 control mean. Across three seeds, scPLAD achieved a delta PCC of 0.4278 ± 0.0033 and an AUPRC of 0.3283 ± 0.0025, compared with 0.3069 and 0.2201 for this baseline. These results support recovery of perturbation-specific effects and response gene rankings beyond the shared source response (Supplementary Fig. S5a–d).

Finally, we tested how target context contributed through its interaction with the perturbation prior, using generated populations before mean correction. Model weights, perturbation priors, the noisy K562 diffusion anchor and the clean K562 reconstruction anchor were held fixed. Only the projected anchor paired with the perturbation prior was replaced by the corresponding RPE1, HepG2 or Jurkat representation. The K562 interaction context yielded the highest three-seed means for delta PCC, Top-100 DE overlap, ERC_{100} and CSA_{100} Pearson (Fig. 4e; Supplementary Fig. S5e–h). These results indicate that target cell context influences predictions through its interaction with the perturbation prior.

Factors shaping scPLAD population predictions

We tested whether sampling empirical residuals could reproduce the population properties captured by scPLAD. The baseline combined a predicted condition mean with complete residual vectors sampled from observed training perturbations (Fig. 5a,b). In K562, scPLAD achieved higher delta PCC, Top 100 DE overlap, AUPRC and FCS. Its FCS was 0.9250, compared with 0.7125 for the residual baseline. The residual baseline achieved slightly higher interval IoU_{100} and CSA Pearson scores. In the cell line transfer task, scPLAD achieved higher scores on all six displayed metrics, including an FCS of 0.6388 compared with 0.3423.

To compare population structure at a common condition mean, we recentered both populations to the same frozen mean prediction. Under this fixed mean setting, scPLAD retained higher FCS than the residual baseline in both tasks, with scores of 0.9170 in K562 and 0.5593 across cell lines. The residual baseline retained slightly higher IoU_{100} in both tasks and slightly higher CSA Pearson in K562. These comparisons showed that scPLAD improved local fidelity and coverage even when the condition means were matched. Additional metrics are shown in Supplementary Fig. S6.

We next tested whether the generated populations matched the responses to the requested perturbations. Reassigning each predicted population to an incorrect perturbation reduced all six displayed scores in both tasks (Fig. 5c). In K562, delta PCC decreased from 0.5581 for

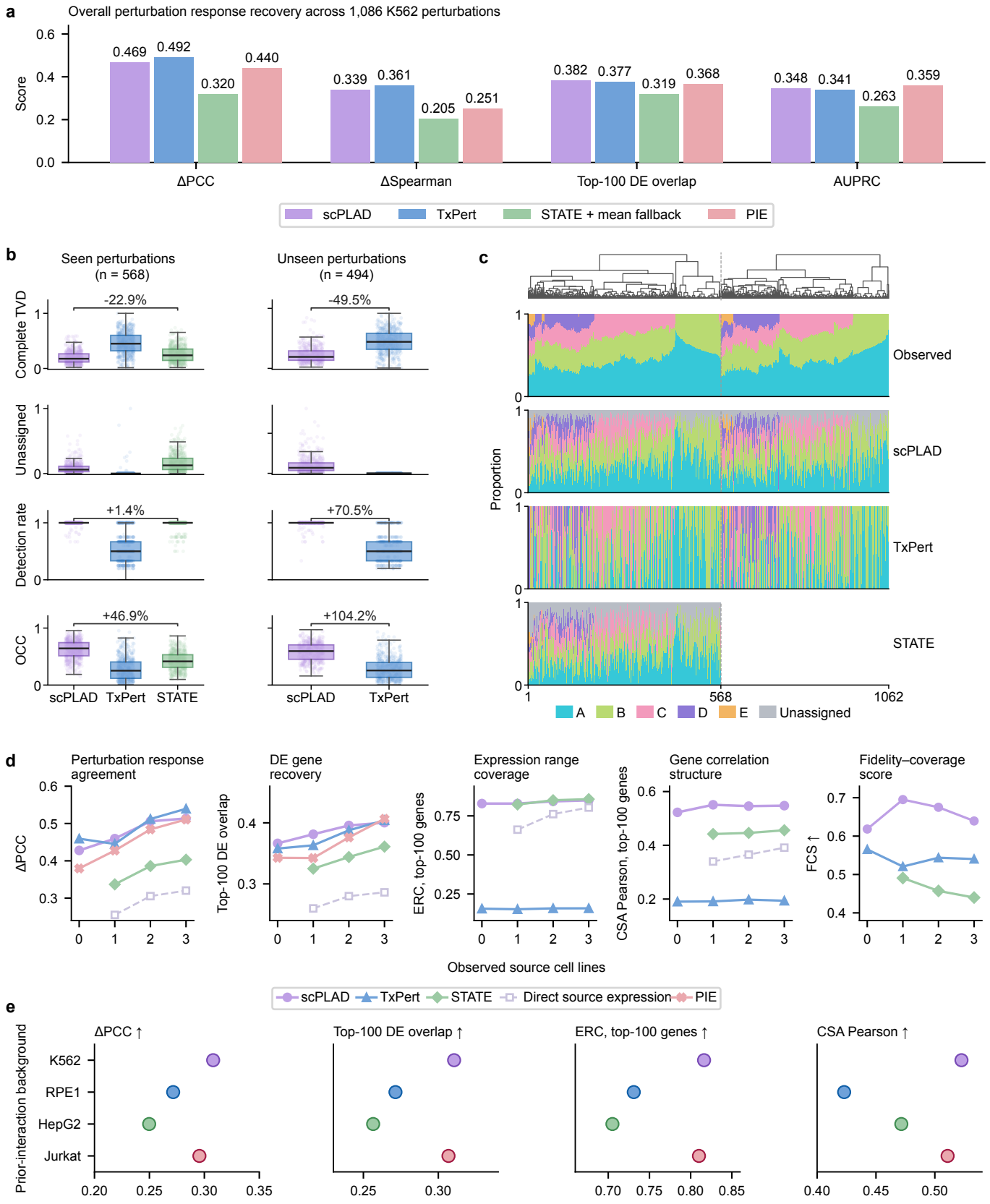


Figure 4: Generation of K562 perturbation responses across cellular contexts. **a,** Overall mean response metrics. STATE directly predicts 580 conditions; the remaining 506 use a global mean-effect fallback. **b,** q99 subpopulation recovery for seen ($n = 568$; source coverage 1–3; scPLAD, TxPert and STATE) and unseen perturbations ($n = 494$; source coverage 0; scPLAD and TxPert). Brackets are shown only when scPLAD leads the strongest comparator for that metric. Labels report descriptive relative differences in condition-weighted means versus STATE for the displayed seen comparisons and TxPert for unseen comparisons; negative values denote reductions in Complete

TVD. **c,** Observed and generated proportions across all 1,062 multi-cluster conditions, ordered within the seen and unseen groups. STATE is blank in the unseen group; cells outside all observed-cluster q99 supports are unassigned. **d,** Delta PCC, Top-100 DE overlap, ERC_{100}, CSA_{100} Pearson and fidelity–coverage score by source cell line coverage. The dashed direct-source baseline covers levels 1–3; STATE population metrics use its direct predictions. **e,** Coverage-0 metrics under alternative prior interaction contexts. Points are three-seed means from the uncorrected generated populations. PIE contributes condition level mean results in a and d.

the correct mapping to a permutation mean of 0.2639 ± 0.0152. FCS decreased from 0.9254 to 0.7790 ± 0.0071. In the cell line transfer task, delta PCC decreased from 0.4687 to 0.2521 ± 0.0034, and FCS decreased from 0.6390 to 0.5168 ± 0.0018. Across all displayed metrics, the correct mapping scored higher than each of the 20 permutations. Generated populations therefore matched the observed responses more closely under the correct perturbation labels.

Finally, we evaluated sensitivity to the number of K562 control cells available at inference, with all model weights fixed. We used five matched subsamples at each size of 64 and 128 cells for scPLAD, TxPert, CellFlow and Scouter (Fig. 5d). With 64 and 128 controls, scPLAD achieved mean FCS values of 0.9219 and 0.9244, respectively, compared with 0.9269 using all 10,691 controls. Delta PCC was 0.4429 with 64 controls, 0.4860 with 128 controls and 0.5580 with the full set. Across both subsampled settings, scPLAD achieved the highest mean FCS and delta PCC among the four evaluated models. FCS remained high with fewer control cells, whereas delta PCC benefited from a larger control reference. Additional metrics are shown in Supplementary Fig. S7.

scPLAD recovers HSPA9 responses and gene correlation structure

We examined HSPA9 as a case study of mean response and population structure recovery in the cell line transfer task. HSPA9 perturbation responses were observed in the source cell lines, and the K562 test set contained 374 perturbed cells. We evaluated mean responses across 159 direct downstream targets identified in the K562 network inferred by Brown *et al.* using inspre.[33] The network was inferred from the same Replogle K562 dataset used in our benchmark. For these targets, we compared predicted and observed mean expression changes relative to K562 controls (Fig. 6a). For scPLAD and PIE, predictions for individual genes were averaged across three seeds before calculating correlations. Pearson and Spearman correlations were 0.89 and 0.82 for PIE, 0.87 and 0.80 for scPLAD, and 0.75 and 0.72 for STATE. TxPert showed a negative Pearson correlation on this target set. PIE achieved the highest correlations, followed by scPLAD, indicating that both methods recovered the mean response across these targets.

We next evaluated population structure using the 100 genes with the largest absolute observed mean responses to HSPA9 perturbation. CSA_{100} Pearson was calculated across all 4,950 unique gene pairs, with the 36 genes with the largest absolute observed mean responses displayed for visualization (Fig. 6b). Mean CSA_{100} Pearson across the three scPLAD seeds was 0.77, compared with 0.02 for STATE and 0.20 for TxPert. The scPLAD heat map also showed major correlation patterns similar to those observed among the displayed genes.

We then compared expression intervals for the same 36 displayed genes using one representative scPLAD run (Fig. 6c). ERC_{36} was 0.82 for both scPLAD and STATE, compared with

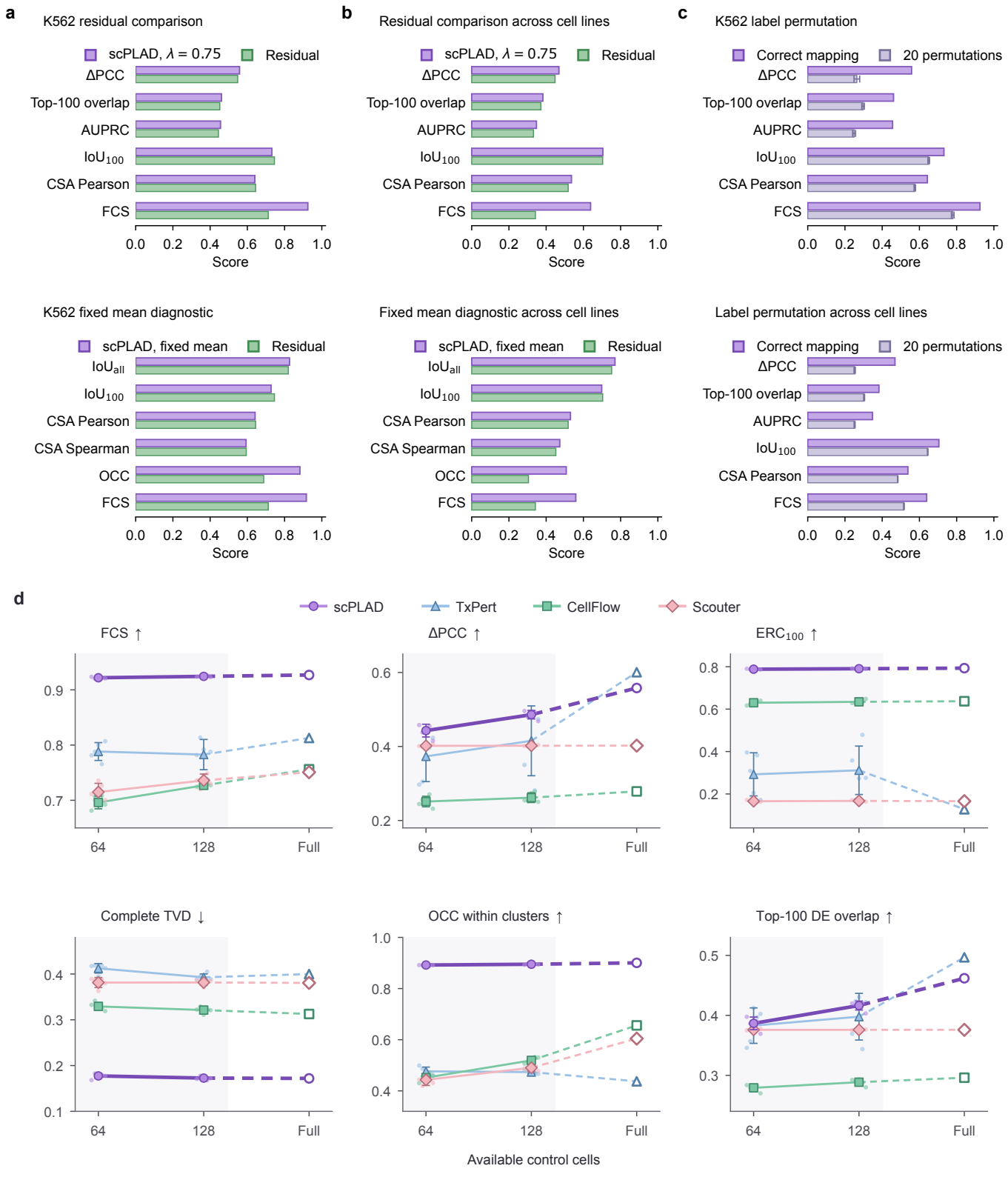


Figure 5: Empirical residual, fixed mean, perturbation label and control availability analyses. **a,b,** Comparisons with a nonparametric training perturbation residual baseline and fixed mean diagnostics in K562 (a) and across cell lines (b). The residual baseline adds complete zero mean residual vectors from observed training perturbations to the same frozen mean response prediction. For the fixed mean diagnostics, both methods were recentered to the same frozen mean response prediction before population structure evaluation. **c,** Perturbation label permutation controls in K562 and across cell lines. Each mapping changes perturbation labels while keeping the complete

predicted populations unchanged. Grey-purple bars show means across 20 one-to-one derangements, with error bars denoting the standard deviation across mappings; permutations across cell lines were performed within source coverage strata. **d,** Sensitivity to the number of available K562 control cells for scPLAD, TxPert, CellFlow and Scouter. Filled symbols show means across five matched control cell draws of size 64 or 128, small points show individual draws and error bars show their sample standard deviation. Open symbols show the single reference evaluation using all 10,691 control cells. Solid lines connect the two subsampled settings and dashed lines connect the 128 cell setting to the full reference as visual guides. FCS, delta PCC, ERC_{100} and Top-100 DE overlap use 272 unseen perturbations; Complete TVD and OCC within clusters use the 255 conditions with multiple observed clusters. Higher values indicate better agreement except for Complete TVD, for which lower is better. Exact numerical values are provided as source data.

0.14 for TxPert. In this case, scPLAD and STATE achieved similar expression range coverage for the displayed genes. However, scPLAD more closely reproduced the observed correlation structure among the top 100 response genes.

Discussion

scPLAD generated populations that recovered mean perturbation responses, local cell states and subpopulation composition across the evaluated tasks. It supported prediction of unseen perturbations within K562 and transfer from source cell lines to a target represented by control cells. These results support joint examination of average perturbation effects and variation among responding cells.

Model comparisons showed that performance depended on the response properties being evaluated. TxPert achieved higher perturbation effect correlations, whereas scPLAD achieved higher fidelity and coverage scores in both tasks. The HSPA9 case illustrated a further distinction between expression range coverage and gene correlation structure. scPLAD and STATE covered similar fractions of the observed intervals for the displayed genes, while scPLAD more closely reproduced correlations among response genes. These findings support evaluating mean responses and population structure through complementary metrics.

The empirical residual comparisons clarified the value of learned population generation. Sampling complete residual vectors from training perturbations remained competitive on several interval and correlation measures. scPLAD achieved higher fidelity and coverage scores in both tasks, including when condition means were matched. This result supports a contribution from the generated cellular deviations to local population recovery. Reassigning complete predictions to incorrect perturbation labels also reduced response and population agreement, showing that the generated populations better matched their requested perturbations.

The ablations linked biological information and gene organization to different aspects of prediction quality. Molecular interaction and regulatory priors improved perturbation direction and response gene recovery, while ESM3 representations improved expression interval coverage and agreement. Pathway guided gene ordering improved interval and correlation measures, with relatively little change in mean response metrics. These patterns suggest evaluating prior selection and gene ordering against both mean responses and population structure.

Transfer performance was associated with the number of source cell lines in which a perturbation had been observed. For perturbations absent from all source cell lines, scPLAD also

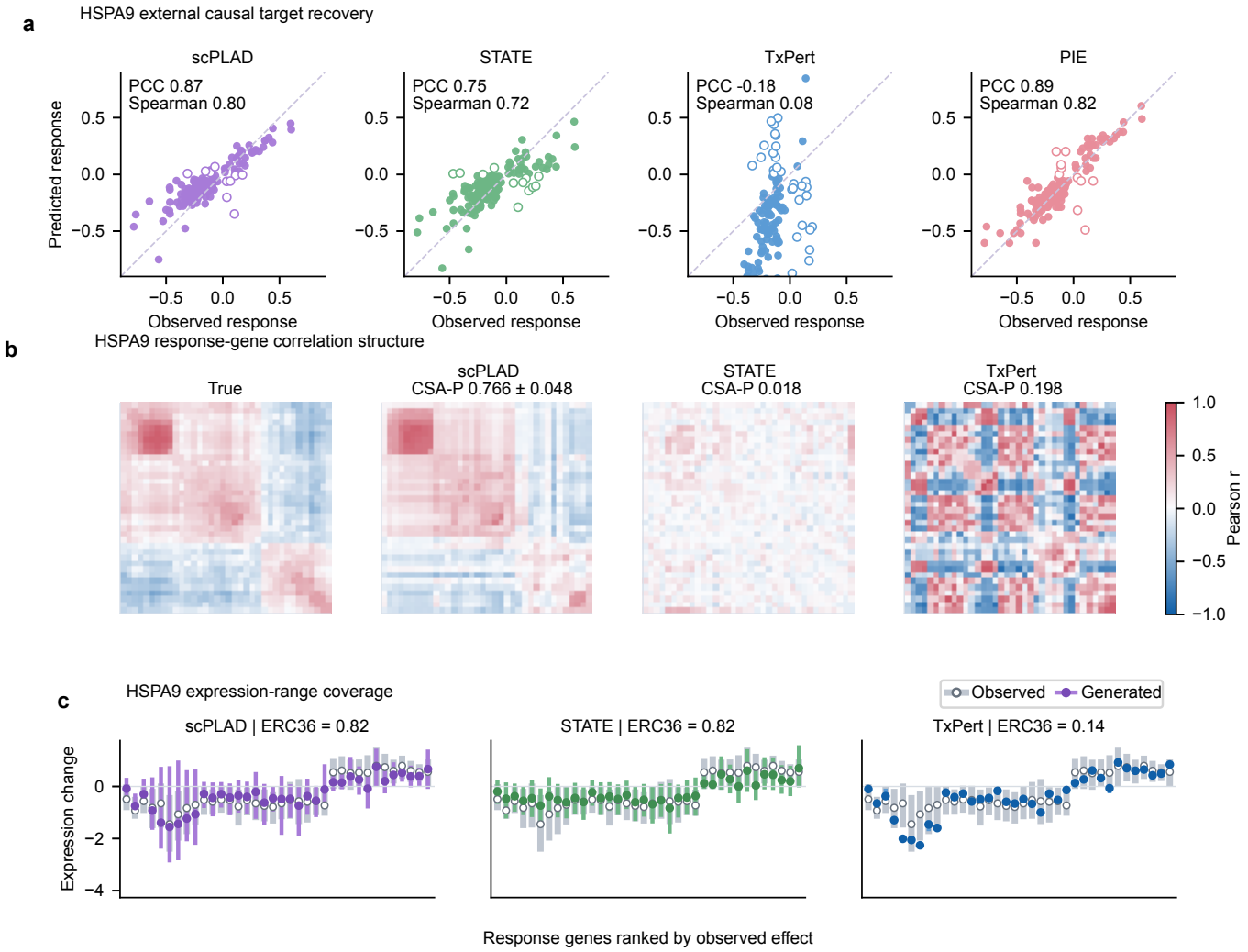


Figure 6: Response-gene and population structure for HSPA9 perturbation. **a,** True and predicted mean responses of 159 direct downstream targets from an external causal network. Targets were obtained from the K562 directed network inferred by Brown *et al.* using inspre;[33] direct HSPA9 edges with $G_{\text{hat}} \neq 0$ and $p_{\text{FDR}} \leq 0.05$ were retained and mapped to the 3,352-gene expression space. The scPLAD and PIE gene-wise predictions are averaged across three random seeds before calculating the displayed correlations. Filled points indicate concordant true and predicted directions, open points indicate discordant directions and the dashed line is the identity reference. PIE is included in panel a. **b,** Within-population correlation structure of the true top-100 HSPA9 response genes. CSA_{100} Pearson was computed using all 4,950 unique pairs; heat maps display 36 genes for visualization. The scPLAD matrix is averaged across three seeds and self-correlations are suppressed for display. **c,** True and generated 10th to 90th percentile expression intervals and means for the same 36 response genes relative to K562 controls. Grey intervals and open points denote true cells; coloured intervals and filled points denote generated cells. ERC_{36} is the average coverage of the true interval across these genes. The scPLAD panel shows one representative model run for this single-condition range analysis.

improved response prediction over a common source effect baseline. Replacing the target context in the prior interaction reduced the evaluated mean scores, supporting its role in conditioning predictions. Control availability affected prediction dimensions differently. With trained model weights fixed, fidelity and coverage scores remained high under control subsampling, whereas perturbation effect correlations benefited from a larger control reference. This distinction can

inform the choice of control reference size for different prediction objectives.

scPLAD combines biological priors with target cell context to generate populations for unseen perturbations and transfer responses across cell lines. We also developed scPLAD Response Lab, a browser interface for generating and exploring these predictions. This interface may help researchers explore virtual perturbation responses and formulate hypotheses for experimental testing.

STAR Methods

Key resources table

Reagent or resource	Source	Identifier
Deposited data		
K562 and RPE1 Perturb-seq data	Replogle et al.[3]	DOI: 10.1016/j.cell.2022.05.013
HepG2 and Jurkat perturbation data	Nadig et al.[34]	DOI: 10.1038/s41588-025-02169-3
Processed TxPert benchmark data	Wenkel et al.[16]	`https://github.com/valence-labs/TxPert`
Processed task data, priors and checkpoints	This study	`https://huggingface.co/zhangzhigang/scPLAD`
Software and algorithms		
scPLAD code and workflows	This study	`https://github.com/zzgzanyr/scPLAD`
scPLAD Response Lab	This study	`https://scplad.net`
TxPert	Wenkel et al.[16]	DOI: 10.1038/s41587-026-03113-4
CellFlow	Klein et al.[23]	DOI: 10.1101/2025.04.11.648220
Scouter	Zhu and Li[17]	DOI: 10.1038/s43588-025-00912-8
GEARS	Roohani et al.[12]	DOI: 10.1038/s41587-023-01905-6
STATE	Adduri et al.[19]	DOI: 10.1016/j.cell.2026.07.052
PIE	Verma et al.[20]	DOI: 10.64898/2026.10.02.756297
Biological databases and protein representations		
Gene Ontology	Gene Ontology Consortium[35]	`https://geneontology.org`
Reactome	Gillespie et al.[36]	`https://reactome.org`
UniProt	UniProt Consortium[37]	`https://www.uniprot.org`
ESM3 sequence representations	Hayes et al.[38]	ESM3-sm-open-v1
STRING	Szklarczyk et al.[39]	`https://string-db.org`
DoRothEA and CollecTRI	Garcia-Alonso et al.; Müller-Dott et al.[40,41]	DOI: 10.1101/gr.240663.118; 10.1093/nar/gkad841
OmniPath	Türei et al.[42]	`https://omnipathdb.org`
CORUM	Giurgiu et al.[43]	`https://mips.helmholtz-muenchen.de/corum/`

Resource availability

Lead contact. Further information and requests for resources should be directed to the lead contact, Pengwei Hu (hpw@ms.xjb.ac.cn).

Materials availability. This study does not report newly generated physical materials.

Data and code availability. The processed datasets, biological priors, gene order files and selected model checkpoints are available in the scPLAD Hugging Face repository (`https://huggingface.co/zhangzhigang/scPLAD`). The K562 and cell line transfer task datasets are provided in separate directories. The original public data sources and the additional processing performed in this study are described in the Methods and Supplementary Information.

The source code and reproducibility workflows are available in the scPLAD GitHub repository (`https://github.com/zzgzanyr/scPLAD`). An interactive browser-based implementation is available through scPLAD Response Lab (`https://scplad.net`). Inference runs locally in the user's browser, and generated expression profiles are not uploaded to an inference server.

Experimental model and study details

Data-use boundaries. Data partitions were fixed before model fitting. In the K562 task, cells from unseen perturbations did not enter PatchAE, displacement diffusion or mean response predictor training, pathway aware gene order construction, or fitted training statistics. In the cell line transfer task, perturbed K562 cells were excluded from the corresponding model fitting stages, whereas perturbed RPE1, HepG2 and Jurkat cells supplied the source data. Control cells from the target context were permitted where the method requires basal context: for control anchors, correlations in control expression used in gene ordering and, in the cell line transfer setting, PatchAE pretraining. Biological prior features were constructed from public annotations, protein sequences and molecular networks without using reserved expression means, differential expression results or generated outputs. Observed test perturbation cells were accessed after prediction as evaluation references, including in the explicitly described clustering and support calibration procedures; they were not used to fit the predictive models or select their checkpoints.

The K562 experiment uses the TxPert K562 pathway5000 module benchmark. Training, validation and test cells originate from the same K562 data universe but are separated by perturbation condition. Comparisons include TxPert, CellFlow, Scouter, GEARS and a non-parametric global cloud baseline. All methods are evaluated on the same 272 unseen conditions, expression space of 5,000 genes and K562 control reference. The global cloud baseline samples 2,000 cells per test condition from the pooled training perturbation cells independently of the test perturbation identity.

The cell line transfer experiment uses the TxPert K562 pathway3352 benchmark. Perturbed cells from RPE1, HepG2 and Jurkat form the source training data, totaling 254,170 cells, and K562 controls define the anchor for the target cell line. TxPert is evaluated on all 1,086 K562

test perturbations. STATE directly predicts the 580 perturbations observed in at least one source cell line. For the mean response comparison across all conditions, the remaining 506 perturbations are completed using a global mean perturbation effect estimated from available STATE predictions; ERC, interval IoU, width ratio and CSA use the 580 direct predictions. The direct source expression baseline pools observed cells for the same perturbation across source cell lines in their original expression backgrounds.

Control anchors are means of encoded control cells. The K562, RPE1, HepG2 and Jurkat control sets contain 10,691, 11,485, 4,976 and 12,013 cells, respectively. These anchors represent cellular context in the diffusion model.

Method details

Problem formulation and model overview

Let $x \in \mathbb{R}^G$ denote a single-cell expression vector, where G is the number of genes retained in a given experimental setting. We denote the perturbed gene by g, the target cell line or cellular context by c, and the set of control cells from that context by $\mathcal{C}_c$. The objective is to generate a set of perturbed single-cell expression profiles $\{\hat{y}_i\}_{i=1}^{M}$ that approximates the conditional population distribution $p_{\text{pert}}(y \mid g, c)$. The generated population should recover the average expression response, response genes, cell-to-cell heterogeneity and the geometry of the perturbed cell population.

As a motivating view, a perturbation response can be regarded as a conditional residual process acting on a basal cell state:

$$y \approx x + R_{g,c}(x), \qquad x \sim p_{\text{ctrl}}(x \mid c), \qquad y \sim p_{\text{pert}}(y \mid g, c).$$

Single-cell perturbation experiments provide unpaired control and perturbed populations. Accordingly, scPLAD formulates the task as a distribution-level conditional mapping in which control cells from the target context define a basal-state anchor and the model learns the distribution of perturbed-cell displacements relative to that anchor in latent space.

The model separates expression reparameterization from conditional latent displacement generation. This design follows the motivation of latent diffusion, adopts the use of latent diffusion for single-cell expression modeling in scLDM, and is informed by the principle in Unified Latents that the latent representation should serve both the downstream generative prior and the decoder.[24,29,44]

The PatchAE encoder E_η maps an expression vector to latent tokens,

$$z = E_\eta(x),$$

and its decoder D_η reconstructs the expression profile,

$$\tilde{x} = D_\eta(z).$$

After pretraining, both modules are frozen. For cellular context c, the clean latent control anchor is

$$a_c = \frac{1}{|\mathcal{C}_c|} \sum_{x_i \in \mathcal{C}_c} E_\eta(x_i).$$

For a perturbed cell y_j, with latent representation $z_{y_j} = E_\eta(y_j)$, we define the control-anchored displacement

$$u_j = z_{y_j} - a_c.$$

Perturbation generation is thus modeled as control-anchored latent displacement diffusion. Given a multi-source biological prior vector r_g and control anchor a_c, the model samples a displacement and reconstructs a complete perturbed latent state:

$$\hat{u} \sim p_\psi(u \mid r_g, a_c), \qquad \hat{z} = a_c + \hat{u},$$

$$\hat{y} = D_\eta(\hat{z}).$$

Equivalently, the generated expression distribution is the decoder-induced push-forward of the conditional displacement distribution:

$$p_\theta(y \mid g, c, \mathcal{C}_c) = \int \delta(y - D_\eta(a_c + u))\, p_\psi(u \mid r_g, a_c)\, \mathrm{d}u.$$

Here, $\delta(\cdot)$ denotes the point mass induced by the deterministic decoder.

Inspired by the large-patch tokenization strategy of JiT,[45] the input gene vector is divided into $N = \lceil G/P \rceil$ patches, and each patch is encoded as one latent token, yielding a structured latent tensor of shape $N \times d$. Each token represents a local gene patch and incorporates cross-patch context through Transformer blocks.[46]

Pathway-aware gene patches

Single-cell expression profiles comprise gene sets without an intrinsic sequence order.[24] In a patch-based Transformer or autoencoder, the imposed gene order determines which genes share a local token. We therefore construct a fixed pathway-aware order that places genes with related pathway annotations or control-expression patterns near one another.

Let $\mathcal{G}$ be the retained gene set, with $|\mathcal{G}| = G$, and let P be the patch size. For genes i and j, we combine non-negative control cell Pearson correlation C_{ij} with normalized Reactome co-membership R_{ij}:[36]

$$S_{ij} = 0.7C_{ij} + 0.3R_{ij}.$$

Gene order construction uses public pathway annotations and permitted control cells as its inputs. In the K562 experiment, gene patches of fixed size are built greedily from S. In the cell line transfer experiment, average linkage hierarchical clustering and optimal leaf ordering are applied to the distance $1 - S$, after which the resulting global order is divided into patches. Genes without Reactome annotations remain eligible through correlations in control expression.

Full filtering criteria and pseudocode are provided in the Supplementary Methods.

For the resulting order $\pi^\star$, the expression vector is rewritten as

$$x_{\pi^\star} = (x_{\pi^\star(1)}, \ldots, x_{\pi^\star(G)}).$$

The i-th gene patch and expression patch are

$$\mathcal{P}_i = \{\pi^\star((i-1)P+1), \ldots, \pi^\star(iP)\},$$

$$x_i^{\text{patch}} = \left(x_{\pi^\star((i-1)P+1)}, \ldots, x_{\pi^\star(iP)}\right) \in \mathbb{R}^P.$$

We denote this ordering and partitioning operation by $T_{\pi^\star}(x) = \{x_i^{\text{patch}}\}_{i=1}^N$. Both experiments use $P = 20$. The 5,000-gene K562 setting yields 250 patches, whereas the 3,352-gene cell line transfer setting is padded to 168 patches. The order supplies a functional-module inductive bias for patch construction.

Noisy Gaussian PatchAE

PatchAE is pretrained independently of perturbation prediction to provide a reconstructable latent space and a decoder that is robust to modest sampling deviations. The term "Gaussian" refers to two optional training components: Gaussian noise added before latent decoding, and a weak Gaussian moment regularizer on the minibatch latent distribution. The moment regularizer is used in the K562 experiment with weight 0.001 and omitted in the cell line transfer experiment.

Encoder and decoder. The ordered expression profile is reshaped into an $[N, P]$ patch sequence. A shared local encoder maps each patch to a 32-dimensional vector,

$$h_i = f_{\text{patch}}(x_{\mathcal{P}_i}), \qquad h_i \in \mathbb{R}^{32}.$$

Positional embeddings are added, and two Transformer encoder blocks exchange information across patches:[46]

$$z = \text{Transformer}(h + e_{\text{pos}}), \qquad z \in \mathbb{R}^{N \times 32}.$$

The decoder maps every token back to its expression patch and concatenates the patches into a full expression profile. The reconstruction objective is

$$\mathcal{L}_{\text{rec}} = \|x - \hat{x}\|_2^2.$$

For a minibatch latent tensor $z \in \mathbb{R}^{B \times N \times 32}$, the optional Gaussian moment regularizer is computed over all flattened latent entries:

$$\mathcal{L}_{\text{gauss}} = \mu(z^{\text{flat}})^2 + \Big(\sigma(z^{\text{flat}}) - 1\Big)^2.$$

The complete autoencoder objective is

$$\mathcal{L}_{\mathrm{AE}} = \mathcal{L}_{\mathrm{rec}} + \lambda_{\mathrm{gauss}}\mathcal{L}_{\mathrm{gauss}}.$$

For K562 training, $\lambda_{\mathrm{gauss}} = 0.001$; for cell line transfer training, $\lambda_{\mathrm{gauss}} = 0$. This deterministic moment regularizer weakly constrains the first two minibatch moments and introduces no latent sampling.

Noisy latent decoding. Latents produced by the displacement model need not lie exactly on the encoder manifold. To expose the decoder to nearby deviations, we add noise to the latent tokens during autoencoder training, following the local-robustness motivation of denoising autoencoders and the objective of reducing mismatch between encoded and prior-generated latents:[44,47]

$$\tilde{z} = z + \sigma_{\mathrm{AE}}\epsilon, \qquad \epsilon \sim \mathcal{N}(0, I), \qquad \sigma_{\mathrm{AE}} = 0.08.$$

The decoder learns to reconstruct x from $\tilde{z}$, thereby covering clean encoder latents and their local neighborhoods.

Global 128-dimensional VAE reconstruction control

Figure 2e includes a conventional global VAE with a 128-dimensional latent vector and multilayer perceptron encoder and decoder.[48] Its input is the processed 3,352-gene expression vector used in the cell line transfer experiment. The encoder comprises input LayerNorm, a $3352 \to 64$ linear layer and GELU, followed by separate linear heads for a 128-dimensional posterior mean and log-variance. The decoder is a $128 \to 64 \to 3352$ MLP with GELU. The objective is

$$\mathcal{L}_{\mathrm{VAE}} = \mathcal{L}_{\mathrm{MSE}} + \beta\mathcal{L}_{\mathrm{KL}}, \qquad \beta = 10^{-3}.$$

The VAE was trained on the same 293,335-cell autoencoder training split as the cell line transfer PatchAE, excluding validation and test cells. Training used 100 epochs, batch size 1,024, AdamW with learning rate 10^{-3}, weight decay 10^{-4}, gradient-norm clipping at 5.0 and random seed 20260721. Reconstruction used the posterior mean without stochastic sampling.

The shared-reference UMAP used K562 control cells and held-out K562 HSPA9, NCBP2 and SLC39A9 perturbation cells, with 250 cells sampled per condition. The K562 controls were included in the cell line transfer autoencoder training split, whereas the three perturbed conditions were excluded. PCA50 and UMAP were fitted only to observed expression profiles, and reconstructions from both autoencoders were projected through the same fitted transformations. UMAP used 30 neighbors, `min_dist`=0.35, Euclidean distance and random seed 20260721.

Multi-source biological prior representation

The perturbation condition is represented by public biological information about the perturbed gene. We first constructed a gene-level prior table for TxPert perturbation genes and reused

the same feature schema in the K562 and cell line transfer experiments. The K562-adapted table was produced by reindexing the cell line transfer prior table to the K562 perturbation set, preserving identical feature definitions. The raw table contains 2,342 numeric features. DepMap/Chronos fields are retained only for optional analyses and are excluded from the main model using `exclude_feature_prefixes=depmap_`, leaving a 2,336-dimensional input.

GO256. Gene Ontology annotations were converted to gene-level profiles after gene-alias matching.[35] Among 2,164 cell line transfer perturbation genes, 2,156 had usable GO annotations and 55 were recovered through aliases. Genes without usable annotations received a zero vector and a missingness indicator. GO256 contains 246 selected term indicators and 10 summary features, including direct and propagated term counts, BP/MF/CC branch counts and `go_has_annotation`. The retained dimensions therefore remain linked to readable ontology terms or annotation summaries.

Reactome256. Reactome fine pathways were mapped to the Reactome event hierarchy and assigned to second-level modules.[36,49] For each perturbation gene, the block records membership in second-level modules and six pathway summaries: the number of fine pathways, the number of second-level modules, pathways per module, mean and maximum hierarchy depth, and an annotation mask. The cell line transfer table contains 2,825 mapped pathways and 216 second-level modules; module features and summaries are padded with zeros to 256 dimensions.

Sequence and molecular-network priors. ESM3-PCA256 was derived from reviewed UniProt protein sequences. Representations from ESM3-sm-open-v1 were averaged across tokens to 1,536 dimensions and reduced to 256 dimensions by PCA over the perturbation gene universe; an availability indicator was retained.[37,38] The STRING block selects the top protein interaction neighbors and forms a 256-dimensional neighbor GO profile by averaging with edge weights, together with summaries of neighbor count, edge weight and availability.[39] CollecTRI and DoRothEA regulatory networks provide signed GO profiles of downstream targets and upstream transcription factors, counts of activating, repressing and unknown edges, and masks for transcription factors and targets.[40,41] OmniPath features analogously summarize upstream, downstream, activating and inhibitory signaling relationships.[42] CORUM features record complex membership, the mean GO profile of members in the same complex, counts of complexes and other members, and a membership mask.[43]

The prior blocks retain their source-specific interpretation; the subsequent learned projection combines them in a shared hidden representation. Their data provenance and use boundaries are summarized under Experimental model and study details.

Control-anchored latent displacement diffusion

Let the frozen encoder and decoder be E and D. A perturbed cell y from context c is encoded as $z_y = E(y)$, and the context-specific clean control anchor is

$$a_c = \frac{1}{|\mathcal{C}_c|} \sum_{x \in \mathcal{C}_c} E(x).$$

The clean displacement is defined as $u = z_y - a_c$. The model uses two diffusion indices with distinct roles. The fixed level τ constructs the displacement representation and its noisy-anchor condition, whereas t is the randomly sampled denoising step of the outer diffusion process. Inspired by DRDD's residual modeling in a fixed-noise domain,[50] at $\tau = 500$ the perturbed latent and control anchor receive the same Gaussian noise instance:

$$q_\tau(z_y; \epsilon) = \sqrt{\bar{\alpha}_\tau} z_y + \sqrt{1 - \bar{\alpha}_\tau}\epsilon,$$

$$q_\tau(a_c; \epsilon) = \sqrt{\bar{\alpha}_\tau} a_c + \sqrt{1 - \bar{\alpha}_\tau}\epsilon.$$

Their difference is

$$u_\tau = q_\tau(z_y; \epsilon) - q_\tau(a_c; \epsilon).$$

Because the noise is shared, the stochastic term cancels exactly:

$$u_\tau = \sqrt{\bar{\alpha}_\tau}(z_y - a_c).$$

Thus, u_τ is a deterministically scaled clean displacement, while τ also determines the noise level of the anchor supplied to the denoiser as conditioning. A global scalar transformation $s(\cdot)$ then standardizes u_τ, and standard outer diffusion training is applied to the standardized displacement. In the 5,000-gene K562 experiment, the training-set mean and standard deviation used by the scaler were 0.0055 and 0.9699; a separate scaler was fitted to the cell line transfer training set.

At every minibatch, an outer diffusion time t is sampled and the standardized target is noised as

$$v_t = \sqrt{\bar{\alpha}_t}\, s(u_\tau) + \sqrt{1 - \bar{\alpha}_t}\epsilon', \qquad \epsilon' \sim \mathcal{N}(0, I).$$

The denoiser uses an x_0-prediction objective:

$$\mathcal{L}_{\text{diff}} = \| f_\psi(v_t, t, r_g, a_{c,\tau}) - s(u_\tau) \|_2^2,$$

where $a_{c,\tau} = q_\tau(a_c; \epsilon)$ is used only as noisy-anchor conditioning during diffusion. Final reconstruction always adds the recovered clean displacement to the clean anchor a_c.

Both experiments use a linear schedule with 1,000 diffusion steps and DDIM sampling with 50 steps.[28,51] The denoiser has hidden dimension 256, six Transformer blocks, eight attention

heads and zero dropout. At generation, the clean displacement is recovered by

$$\hat{u} = \frac{s^{-1}(\widehat{s(u_\tau)})}{\sqrt{\bar{\alpha}_\tau}},$$

and the expression sample is

$$\hat{y} = D(a_c + \hat{u}).$$

Prior-anchor interaction conditioning

Prior-table construction and denoiser conditioning are separate stages. GO, Reactome, ESM3, STRING and the other sources first form distinct feature blocks, several of which are 256-dimensional. After excluding DepMap, these blocks are concatenated into the 2,336-dimensional vector r_g. A shared condition encoder maps the concatenated vector to a global hidden token of dimension $H = 256$:

$$c_g = f_{\text{prior}}(r_g) \in \mathbb{R}^H.$$

This token is broadcast across patches and interacts with the projected control-anchor token. For patch i,

$$\tilde{a}_{c,i} = W_a a_{c,i},$$

$$m_i = f_{\text{int}}\left([c_g, \tilde{a}_{c,i}, c_g \odot \tilde{a}_{c,i}]\right).$$

The interaction token m_i represents the perturbation prior in the local target-control background. It is combined with timestep and context embeddings:

$$q_i = e_t + e_{\text{ctx}} + m_i.$$

In the main configuration, the context-feature mode is zero and cellular background enters principally through the anchor a_c.

Each Transformer block maps q_i to adaptive normalization parameters. In abbreviated form,

$$\text{AdaLN}(h, c) = \gamma(c) \cdot \text{LN}(h) + \beta(c).$$

More specifically,

$$(\beta_i^{\text{attn}}, \gamma_i^{\text{attn}}, g_i^{\text{attn}}, \beta_i^{\text{mlp}}, \gamma_i^{\text{mlp}}, g_i^{\text{mlp}}) = f_{\text{AdaLN}}(q_i).$$

The attention branch is

$$\tilde{h}_i^{\text{attn}} = (1 + \gamma_i^{\text{attn}}) \odot \text{LN}(h_i) + \beta_i^{\text{attn}},$$

$$h' = h + g^{\text{attn}} \odot \text{SA}(\tilde{h}^{\text{attn}}),$$

and the MLP branch is

$$\tilde{h}_i^{\text{mlp}} = (1 + \gamma_i^{\text{mlp}}) \odot \text{LN}(h_i') + \beta_i^{\text{mlp}},$$

$$h^{\text{next}} = h' + g^{\text{mlp}} \odot \text{MLP}(\tilde{h}^{\text{mlp}}).$$

This conditioning follows FiLM and the principles of adaptive layer normalization used in diffusion Transformers.[52,53] The perturbation prior can therefore alter normalization scale, shift and residual gating throughout the denoiser.

Prior-conditioned mean response prediction and population correction

The stochastic generator models the distribution of cellular responses, and a separate predictor refines the mean perturbation effect for each condition in both experimental settings. The two tasks used independently trained predictors and independently fitted preprocessing statistics in their corresponding expression spaces: 5,000 genes for the K562 task and 3,352 genes for the cell line transfer task. Each predictor receives a 256-dimensional GO block, a 256-dimensional Reactome block and the matched control mean. GO, Reactome and control inputs are processed by separate 32-dimensional linear layers with SiLU activation. Their three outputs are concatenated, projected through a 128-dimensional SiLU layer with dropout 0.1 and mapped to a perturbation effect vector in the corresponding gene space.

The training target is the pseudobulk mean of each perturbation condition minus its matched cell-line control mean. The K562 predictor was trained and validated on the K562 training and validation perturbation splits. The cell line transfer predictor was trained and validated on separate source perturbation splits from RPE1, HepG2 and Jurkat, paired with the corresponding source cell line control means. Within each task, prior features were standardized feature-wise, the control mean was standardized by a global scalar mean and standard deviation, and the effect target was divided by a global scalar standard deviation, with all quantities fitted from that task's training split. Three predictors with independent random seeds were optimized using AdamW. The loss combines mean-squared error on the normalized effect vector with a row-wise correlation term,

$$\mathcal{L}_{\text{mean}} = \text{MSE}(\hat{\delta}_c, \delta_c) + 0.05 \left[1 - \text{PCC}(\hat{\delta}_c, \delta_c)\right].$$

The checkpoint with the highest validation delta PCC was retained for each seed, and the three predicted effect vectors were averaged. Additional optimization settings are provided in the Supplementary Methods.

Let μ_c^{mean} denote the condition mean obtained by adding the ensemble-predicted effect to the target context control mean, and let $\bar{\hat{x}}_c$ be the mean of the population generated by the diffusion model. The same correction was applied to every generated cell in condition c,

$$\hat{x}'_{c,i} = \hat{x}_{c,i} + 0.75 \left(\mu_c^{\text{mean}} - \bar{\hat{x}}_c\right),$$

where the coefficient 0.75 achieved the highest mean delta PCC in a within-context calibration using source cell lines. We used this coefficient for both tasks and held it fixed across diffusion seeds, conditions and ablation variants. Details of the calibration are provided in the Supplementary Methods. This operation moves the generated mean towards the independently predicted response while retaining the diffusion model's cell-to-cell deviations around that mean. Because

it is a constant gene-wise translation within each condition, it leaves within-condition variance, covariance and correlation matrices unchanged.

Quantification and statistical analysis

For perturbation condition c, let the true and generated cell sets be $X_c = \{x_i\}_{i=1}^{n_c}$ and $\hat{X}_c = \{\hat{x}_i\}_{i=1}^{m_c}$, both in $\mathbb{R}^G$. Their pseudobulk means are

$$\bar{x}_c = \frac{1}{n_c}\sum_{i=1}^{n_c} x_i, \qquad \bar{\hat{x}}_c = \frac{1}{m_c}\sum_{i=1}^{m_c} \hat{x}_i.$$

Unless otherwise specified, metrics are computed separately for each perturbation condition and averaged with equal weight across conditions. Exceptions, including the median summaries for width ratio, are stated in the corresponding metric definitions.

Mean-expression and perturbation-direction metrics. Mean PCC and mean Spearman compare $\bar{x}_c$ and $\bar{\hat{x}}_c$ across genes. Pseudobulk errors are

$$\mathrm{MSE}_c = \frac{1}{G}\sum_{g=1}^{G}(\bar{\hat{x}}_{c,g} - \bar{x}_{c,g})^2,$$

$$\mathrm{MAE}_c = \frac{1}{G}\sum_{g=1}^{G}|\bar{\hat{x}}_{c,g} - \bar{x}_{c,g}|.$$

With target context control mean $\bar{x}_{\mathrm{ctrl}}$, the true and generated perturbation effects are

$$\Delta x_c = \bar{x}_c - \bar{x}_{\mathrm{ctrl}}, \qquad \Delta\hat{x}_c = \bar{\hat{x}}_c - \bar{x}_{\mathrm{ctrl}}.$$

Delta PCC and delta Spearman compare these effect vectors across genes.

Response-gene recovery. Let $T_c^{(k)}$ and $\hat{T}_c^{(k)}$ be the top-k genes ranked by $|\Delta x_{c,g}|$ and $|\Delta\hat{x}_{c,g}|$, respectively. Top-k DE overlap is

$$\mathrm{Overlap}_c^{(k)} = \frac{|T_c^{(k)} \cap \hat{T}_c^{(k)}|}{k}.$$

For each perturbation, the 100 genes with the largest absolute observed effects were treated as positives, and the absolute predicted effects were used as ranking scores. AUPRC was computed as average precision using `average_precision_score` from scikit-learn.

GGE-PCA50 distributional metrics. For the fixed size K562 comparisons and ablations, true and generated cells are embedded in a common PCA50 space following the GGE evaluation framework.[32] Wasserstein-1, maximum mean discrepancy squared (MMD^2) and Energy distance are computed separately for each principal component and then averaged. For component ℓ,

unbiased squared MMD with an RBF kernel is

$$\mathrm{MMD}^2_{c,\ell} = \frac{1}{n_c(n_c-1)}\sum_{i\neq j} k(z_{i\ell}, z_{j\ell}) + \frac{1}{m_c(m_c-1)}\sum_{i\neq j} k(\hat{z}_{i\ell}, \hat{z}_{j\ell}) - \frac{2}{n_c m_c}\sum_{i=1}^{n_c}\sum_{j=1}^{m_c} k(z_{i\ell}, \hat{z}_{j\ell}),$$

where the bandwidth is selected by the median heuristic. Negative MMD^2 estimates were set to zero before component averaging. Energy distance is

$$\mathcal{E}_{c,\ell} = 2\mathbb{E}|Z_\ell - \hat{Z}_\ell| - \mathbb{E}|Z_\ell - Z'_\ell| - \mathbb{E}|\hat{Z}_\ell - \hat{Z}'_\ell|.$$

Lower distances indicate closer empirical cell distributions.

Fidelity–coverage score. Local support was evaluated using a nearest-neighbor manifold construction adapted from generative precision and coverage metrics.[54,55] A fixed PCA50 representation was fitted on 10,000 randomly sampled K562 training cells in the 5,000-gene benchmark space. For each condition and repeat, equal numbers of observed and generated cells were sampled without replacement, capped at 100 cells per group. Let z_i denote an observed cell in PCA space and let r_i be its Euclidean distance to its fifth-nearest other observed cell. For generated cells $\hat{z}_j$, the realistic cell fraction (RCF) and observed cell coverage (OCC) are

$$\mathrm{RCF}_c = \frac{1}{m_c}\sum_{j=1}^{m_c} \mathbb{I}\left[\exists\, i \in \{1,\ldots,n_c\} : \|\hat{z}_j - z_i\|_2 \leq r_i\right],$$

$$\mathrm{OCC}_c = \frac{1}{n_c}\sum_{i=1}^{n_c} \mathbb{I}\left[\min_j \|\hat{z}_j - z_i\|_2 \leq r_i\right].$$

RCF measures the fraction of generated cells that fall within at least one observed cell support neighborhood, whereas OCC measures the fraction of observed cell neighborhoods reached by at least one generated cell. We define their harmonic mean as the Fidelity–coverage score,

$$\mathrm{FCS}_c = \frac{2\,\mathrm{RCF}_c\,\mathrm{OCC}_c}{\mathrm{RCF}_c + \mathrm{OCC}_c},$$

with $\mathrm{FCS}_c = 0$ when both components are zero. Sampling was repeated ten times. FCS was calculated within each repeat before averaging repeats within a condition; values for individual conditions were then averaged with equal weight, and the three scPLAD seeds were averaged within condition. The same neighborhood order, sampling cap and aggregation rule were used for all models. RCF and OCC are reported separately, and FCS summarizes their joint performance in Fig. 3h. For the cell line transfer analysis, the same procedure was applied in a fixed PCA50 representation of the shared 3,352-gene space and summarized within source coverage strata (Fig. 4d). Prediction matrices were reordered to the benchmark gene order before PCA projection.

Subpopulation composition recovery. For each K562 test condition, observed cells were projected into the same fixed PCA50 representation and clustered independently using a Euclidean 10-nearest-neighbor graph and Leiden clustering at resolution 0.40. The graph was undirected, Leiden optimization used two iterations and seed 20261023, and clusters were ranked by observed abundance within each condition (A for the largest cluster, B for the second largest and so on). This yielded 17 single-cluster and 255 multicluster conditions. Generated cells were assigned with a distance-weighted 10-nearest-neighbor classifier fitted to all observed cells from the corresponding condition. For support rejection, the mean distance to the five nearest observed cells in the predicted cluster was divided by that cluster's within-reference distance scale. Thresholds were calibrated by fivefold cross-validation on observed cells; predictions exceeding the 99th percentile of the corresponding observed support-score distribution were labeled unassigned. Clusters represented by fewer than 20 calibration cells used the condition-pooled threshold.

For observed cluster proportions $p_{c,r}$, accepted generated proportions $\hat{p}_{c,r}$, and unassigned fraction u_c, complete total variation distance (Complete TVD) was

$$\mathrm{TVD}_c = \frac{1}{2}\left(\sum_r |\hat{p}_{c,r} - p_{c,r}| + u_c\right).$$

A cluster was considered detected when it contained at least 1% of all generated cells. OCC within clusters applied the same fifth-neighbor support definition as above after assignment, using equal observed and generated samples capped at 100 cells, ten repeats and macro-averaging across the observed clusters in a condition. Summary comparisons were restricted to the 255 multicluster conditions and weighted conditions equally; scPLAD values were first averaged across its three seeds. These clusters serve as data-driven within-condition partitions for composition-recovery analysis.

For the cell line transfer analysis, a separate PCA50 transform was fitted to 10,000 cells sampled from the shared 3,352-gene training partition and then frozen. All observed K562 test cells were projected with this transform, and each of the 1,086 perturbation conditions was clustered independently using the same Euclidean 10-nearest-neighbor Leiden procedure at resolution 0.40. This yielded 24 single-cluster and 1,062 multicluster conditions. The distance-weighted 10-nearest-neighbor assignment, normalized five-neighbor support score and fivefold observed cell calibration described above were applied without model-specific changes, using the q99 support threshold. All models used the same observed clusters, PCA transform and thresholds.

Seen perturbations were the 580 conditions with direct STATE predictions, equivalently source coverage 1–3; excluding 12 single-cluster conditions left 568 multicluster conditions. Unseen perturbations were the 506 source coverage-0 conditions; excluding 12 single-cluster conditions left 494 multicluster conditions. The unseen population analysis compares scPLAD with TxPert because STATE provides no population output for coverage-0 perturbations. Each

model contributed 2,000 generated cells per available condition. scPLAD metrics were averaged across its three seeds within each condition before conditions were weighted equally. Relative-change labels in Fig. 4b were calculated as $100(M_{\text{scPLAD}} - M_{\text{comp}})/|M_{\text{comp}}|$, using the strongest comparator mean for each metric. The brackets summarize descriptive relative mean differences.

Expression range coverage and interval agreement. For condition c and gene g, define the true and generated central expression intervals

$$R_{c,g}^{\text{true}} = [q_{0.1}(x_{\cdot,g}^{c}), q_{0.9}(x_{\cdot,g}^{c})], \qquad R_{c,g}^{\text{pred}} = [q_{0.1}(\hat{x}_{\cdot,g}^{c}), q_{0.9}(\hat{x}_{\cdot,g}^{c})].$$

Expression range coverage (ERC) is

$$\text{ERC}_{c,g} = \frac{|R_{c,g}^{\text{true}} \cap R_{c,g}^{\text{pred}}|}{|R_{c,g}^{\text{true}}| + \varepsilon}, \qquad \text{ERC}_{c} = \frac{1}{G}\sum_{g=1}^{G} \text{ERC}_{c,g},$$

where ε is a constant used for numerical stability. ERC ranges from 0 to 1 and measures the fraction of the true central interval for each gene covered by generated cells.

To evaluate whether the generated interval also matches the location and width of the observed interval, we additionally compute interval intersection-over-union (IoU). Let

$$I_{c,g} = |R_{c,g}^{\text{true}} \cap R_{c,g}^{\text{pred}}|, \quad w_{c,g}^{\text{true}} = |R_{c,g}^{\text{true}}|, \quad w_{c,g}^{\text{pred}} = |R_{c,g}^{\text{pred}}|.$$

For genes with $w_{c,g}^{\text{true}} > 10^{-8}$,

$$\text{IoU}_{c,g} = \frac{I_{c,g}}{w_{c,g}^{\text{true}} + w_{c,g}^{\text{pred}} - I_{c,g}}, \qquad \rho_{c,g} = \frac{w_{c,g}^{\text{pred}}}{w_{c,g}^{\text{true}}}.$$

IoU penalizes intervals that are too broad, too narrow or displaced. The width ratio ρ distinguishes broadening ($\rho > 1$) from narrowing ($\rho < 1$). We report all gene and true top-100 response gene versions from the same candidate sets used for ERC. ERC is averaged over the complete candidate set, with zero-width observed intervals contributing zero under the definition above. IoU and width ratio are instead summarized over the evaluable subset with $w_{c,g}^{\text{true}} > 10^{-8}$, and the evaluable fraction is recorded. Condition-level IoU is the arithmetic mean over this subset. Width-ratio summaries use the median within each condition and the median across conditions.

Correlation structure agreement. For each condition, let S_c be the 100 genes with the largest true perturbation effects. We compute

$$C_c^{\text{true}} = \text{corr}(X_c[:, S_c]), \qquad C_c^{\text{pred}} = \text{corr}(\hat{X}_c[:, S_c]).$$

Using the vectorized off-diagonal upper triangles,

$$\text{CSA}_c = \text{corr}\left(\text{vec}_{\text{upper}}(C_c^{\text{true}}), \text{vec}_{\text{upper}}(C_c^{\text{pred}})\right).$$

Pearson and Spearman versions are reported. CSA quantifies preservation of within-condition pairwise gene correlation structure among response genes.

Empirical residual and perturbation label controls. The training perturbation residual baseline combined the frozen mean response prediction with complete residual vectors sampled from observed training perturbations. Each training cell was centered by the mean of its own perturbation condition; training conditions were sampled with equal probability, followed by a cell within the selected condition. The sampled residual matrix was recentered to exact zero mean and added to the predicted condition mean, producing 2,000 cells per test condition without fitting another population generator. The K562 task used K562 training perturbations, whereas the cell line transfer task used training perturbations from RPE1, HepG2 and Jurkat. The complete method comparison used the reported scPLAD populations after fixed $\lambda = 0.75$ mean correction; implementation details and the fixed mean diagnostic are provided in the Supplementary Methods.

For the perturbation label control, we reassigned each complete final predicted population to an incorrect perturbation label while keeping its expression values unchanged. In K562, one-to-one derangements spanned all 272 unseen perturbations. In the cell line transfer task, derangements were performed separately within source coverage strata 0–3. Each mapping contained no fixed points, and the same 20 mappings were applied to all three diffusion seeds. Evaluation references, observed cells and response gene sets followed the target label. We report the score averaged across three seeds for the correct mapping and the mean and standard deviation across 20 derangements after first averaging each mapping across seeds.

Control-cell availability analysis. We evaluated inference sensitivity to the number of available K562 control cells for scPLAD, TxPert, CellFlow and Scouter while keeping all trained model weights fixed. From the full pool of 10,691 controls, we drew five matched subsets of 64 cells and five matched subsets of 128 cells without replacement; the same control identifiers were supplied to all compatible model pipelines in each draw. Each method used its native control representation. scPLAD used the sampled cells to construct its target-control anchor and applied the fixed $\lambda = 0.75$ mean correction. TxPert used batch-matched control means, with selected-pool means for batches absent from a subset. CellFlow and Scouter used individual controls resampled from the selected pool. Each setting generated 2,000 cells for each of the 272 held-out perturbations. Complete TVD and OCC within clusters were evaluated on the 255 conditions containing multiple observed clusters. Reported subset values are equal-weight condition means averaged over the five control draws; error bars denote the sample standard deviation across draws. The full-control result is a single reference evaluation and therefore has no sampling error bar.

HSPA9 external target evaluation. The HSPA9 analysis uses a directed K562 causal network inferred by Brown *et al.* with inverse sparse regression.[33] Direct outgoing HSPA9 edges

with $G_{\text{hat}} \neq 0$ and $p_{\text{FDR}} \leq 0.05$ were retained, mapped from Ensembl identifiers to gene symbols and intersected with the 3,352-gene expression space, yielding 159 unique downstream targets. For each model, their true and predicted pseudobulk perturbation effects were compared using Pearson correlation, Spearman correlation and directional agreement. For scPLAD, HSPA9 effects for individual genes from the three runs with fixed $\lambda = 0.75$ were averaged before calculating the correlations reported in Fig. 6a. PIE effects were inferred independently for HSPA9 from each of the three trained checkpoints, aligned to the common gene order and averaged across seeds before the same correlation calculation. The network was applied as a post hoc evaluation reference after model training and condition encoding. Because the network and benchmark both draw on Replogle K562 Perturb-seq data, this analysis measures concordance with an externally inferred network rather than independent experimental validation.

The HSPA9 expression range panel displays a fixed set of 36 genes from the true top response set. Expression values are centered by the K562 test control mean. The scPLAD intervals and means were computed from 2,000 cells generated by one representative model run; STATE and TxPert were each evaluated from their corresponding single generated population.

Paired statistics for prior ablations. For the ESM3 and network prior contrasts, condition level metrics were first averaged across the three random seeds. Each augmented-prior model and the GO+Reactome baseline were then paired across the same 272 held-out perturbations. Two-sided Wilcoxon signed-rank tests assessed paired differences, with zero differences excluded, and 95% confidence intervals for mean paired changes were obtained from 10,000 condition level bootstrap samples. Benjamini–Hochberg correction was applied separately within each contrast across five metrics: delta PCC, delta Spearman, Top-100 DE overlap, all-gene ERC and all-gene interval IoU. Effects with $q < 0.05$ are displayed in Supplementary Fig. S3c; all raw p values and adjusted q values are provided in the supplementary source data.

Use of generative AI. During the preparation of this work, the authors used OpenAI ChatGPT and Codex to improve the language and readability of the manuscript and to assist with code development. The core ideas of the study were developed through discussions among the authors. The authors reviewed and edited the resulting text and code and take full responsibility for the content of the manuscript and the work reported.

References

1. 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. (2016). Perturb-Seq: Dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. Cell *167*, 1853–1866.e17. https://doi.org/10.1016/j.cell.2016.11.038.

2. Datlinger, P., Rendeiro, A.F., Schmidl, C., Krausgruber, T., Traxler, P., Klughammer, J., Schuster, L.C., Kuchler, A., Alpar, D., and Bock, C. (2017). Pooled CRISPR screening with

single-cell transcriptome readout. Nat. Methods *14*, 297–301. https://doi.org/10.1038/nmeth.4177.

3. Replogle, J.M., Saunders, R.A., Pogson, A.N., Hussmann, J.A., Lenail, A., Guna, A., Mascibroda, L., Wagner, E.J., Adelman, K., Lithwick-Yanai, G., et al. (2022). Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. Cell *185*, 2559–2575.e28. https://doi.org/10.1016/j.cell.2022.05.013.

4. Feng, C., Peets, E.M., Zhou, Y., Crepaldi, L., Usluer, S., Dunham, A., Braunger, J.M., Su, J., Strauss, M.E., Muraro, D., et al. (2026). A genome-scale single-cell CRISPRi map of trans gene regulation across human pluripotent stem cell lines. Cell Genom. *6*, 101076. https://doi.org/10.1016/j.xgen.2025.101076.

5. Srivatsan, S.R., McFaline-Figueroa, J.L., Ramani, V., Saunders, L., Cao, J., Packer, J., Pliner, H.A., Jackson, D.L., Daza, R.M., Christiansen, L., et al. (2020). Massively multiplex chemical transcriptomics at single-cell resolution. Science *367*, 45–51. https://doi.org/10.1126/science.aax6234.

6. Kang, H.M., Subramaniam, M., Targ, S., Nguyen, M., Maliskova, L., McCarthy, E., Wan, E., Wong, S., Byrnes, L., Lanata, C.M., et al. (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. Nat. Biotechnol. *36*, 89–94. https://doi.org/10.1038/nbt.4042.

7. Frangieh, C.J., Melms, J.C., Thakore, P.I., Geiger-Schuller, K.R., Ho, P., Luoma, A.M., Cleary, B., Jerby-Arnon, L., Malu, S., Cuoco, M.S., et al. (2021). Multimodal pooled Perturb-CITE-seq screens in patient models define mechanisms of cancer immune evasion. Nat. Genet. *53*, 332–341. https://doi.org/10.1038/s41588-021-00779-1.

8. McFarland, J.M., Paolella, B.R., Warren, A., Geiger-Schuller, K., Shibue, T., Rothberg, M., Kuksenko, O., Colgan, W.N., Jones, A., Chambers, E., et al. (2020). Multiplexed single-cell transcriptional response profiling to define cancer vulnerabilities and therapeutic mechanism of action. Nat. Commun. *11*, 4296. https://doi.org/10.1038/s41467-020-17440-w.

9. McFaline-Figueroa, J.L., Hill, A.J., Qiu, X., Jackson, D., Shendure, J., and Trapnell, C. (2019). A pooled single-cell genetic screen identifies regulatory checkpoints in the continuum of the epithelial-to-mesenchymal transition. Nat. Genet. *51*, 1389–1398. https://doi.org/10.1038/s41588-019-0489-5.

10. Duan, J., Li, B., Bhakta, M., Xie, S., Zhou, P., Munshi, N.V., and Hon, G.C. (2019). Rational reprogramming of cellular states by combinatorial perturbation. Cell Rep. *27*, 3486–3499.e6. https://doi.org/10.1016/j.celrep.2019.05.079.

11. Zhou, P., Shi, H., Huang, H., Sun, X., Yuan, S., Chapman, N.M., Connelly, J.P., Lim, S.A., Saravia, J., KC, A., et al. (2023). Single-cell CRISPR screens in vivo map T cell fate regulomes in cancer. Nature *624*, 154–163. https://doi.org/10.1038/s41586-023-06733-x.

12. Roohani, Y., Huang, K., and Leskovec, J. (2024). Predicting transcriptional outcomes of novel multigene perturbations with GEARS. Nat. Biotechnol. *42*, 927–935. https://doi.org/10.1038/s41587-023-01905-6.

13. Lotfollahi, M., Wolf, F.A., and Theis, F.J. (2019). scGen predicts single-cell perturbation responses. Nat. Methods *16*, 715–721. https://doi.org/10.1038/s41592-019-0494-8.

14. 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. (2023). Predicting cellular responses to complex perturbations in high-throughput screens. Mol. Syst. Biol. *19*, e11517. https://doi.org/10.15252/msb.202211517.

15. Bunne, C., Stark, S.G., Gut, G., Sarabia del Castillo, J., Levesque, M., Lehmann, K.V., Pelkmans, L., Krause, A., and Rätsch, G. (2023). Learning single-cell perturbation responses using neural optimal transport. Nat. Methods *20*, 1759–1768. https://doi.org/10.1038/s41592-023-01969-x.

16. Wenkel, F., Tu, W., Masschelein, C., Shirzad, H., Hodgson, L., Bendidi, I., Eastwood, C., Whitfield, S.T., Russell, C., El Mesbahi, Y., et al. (2026). TxPert: using multiple knowledge graphs for prediction of transcriptomic perturbation effects. Nat. Biotechnol. Advance online publication. https://doi.org/10.1038/s41587-026-03113-4.

17. Zhu, O., and Li, J. (2026). Scouter predicts transcriptional responses to genetic perturbations with large language model embeddings. Nat. Comput. Sci. *6*, 21–28. https://doi.org/10.1038/s43588-025-00912-8.

18. Wang, G., Liu, T., Zhao, J., Cheng, Y., and Zhao, H. (2024). Modeling and predicting single-cell multi-gene perturbation responses with scLAMBDA. bioRxiv Preprint. https://doi.org/10.1101/2024.12.04.626878.

19. Adduri, A.K., Gautam, D., Bevilacqua, B., Naghipourfar, M., Imran, A., Shah, R., Teyssier, N., Verma, R., Carpenter, C., Eraslan, B., et al. (2026). Predicting cellular responses to perturbation across diverse contexts with State. Cell *189*, 5914–5931.e20. https://doi.org/10.1016/j.cell.2026.07.052.

20. Verma, R., Adduri, A., Bevilacqua, B., Eraslan, B., Burke, D.P., Goodarzi, H., and Roohani, Y.H. (2026). PIE: Generalizing perturbation effects across unseen perturbations, contexts and datasets. bioRxiv Preprint. https://doi.org/10.64898/2026.10.02.756297.

21. Luo, E., Hao, M., Wei, L., and Zhang, X. (2024). scDiffusion: conditional generation of high-quality single-cell data using diffusion model. Bioinformatics *40*, btae518. https://doi.org/10.1093/bioinformatics/btae518.

22. He, S., Zhu, Y., Tavakol, D.N., Ye, H., Lao, Y.H., Zhu, Z., Xu, C., Chauhan, S., Garty, G., Tomer, R., et al. (2026). Squidiff: predicting cellular development and responses to

perturbations using a diffusion model. Nat. Methods *23*, 65–77. https://doi.org/10.1038/s41592-025-02877-y.

23. Klein, D., Fleck, J.S., Bobrovskiy, D., Zimmermann, L., Becker, S., Palma, A., Dony, L., Tejada-Lapuerta, A., Huguet, G., Lin, H.C., et al. (2025). CellFlow enables generative single-cell phenotype modeling with flow matching. bioRxiv Preprint. https://doi.org/10.1101/2025.04.11.648220.

24. Palla, G., Babu, S., Dibaeinia, P., Pearce, J.D., Li, D., Khan, A.A., Karaletsos, T., and Tomczak, J.M. (2026). Scalable single-cell gene expression generation with latent diffusion models. In Proceedings of the 43rd International Conference on Machine Learning, Proceedings of Machine Learning Research, *306*, 95607–95648, PMLR. URL https://proceedings.mlr.press/v306/palla26a.html.

25. Lopez, R., Regier, J., Cole, M.B., Jordan, M.I., and Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. Nat. Methods *15*, 1053–1058. https://doi.org/10.1038/s41592-018-0229-2.

26. Kernfeld, E., Yang, Y., Weinstock, J.S., Battle, A., and Cahan, P. (2025). A comparison of computational methods for expression forecasting. Genome Biol. *26*, 388. https://doi.org/10.1186/s13059-025-03840-y.

27. Ahlmann-Eltze, C., Huber, W., and Anders, S. (2025). Deep-learning-based gene perturbation effect prediction does not yet outperform simple linear baselines. Nat. Methods *22*, 1657–1661. https://doi.org/10.1038/s41592-025-02772-6.

28. Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. In Advances in Neural Information Processing Systems, *33*, 6840–6851.

29. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 10684–10695.

30. Lipman, Y., Chen, R.T.Q., Ben-Hamu, H., Nickel, M., and Le, M. (2023). Flow matching for generative modeling. In International Conference on Learning Representations. URL https://openreview.net/forum?id=PqvMRDCJT9t.

31. Ni, S., Kong, X., Zhang, Y., Chen, Z., Wang, Z., Fu, Z., Huo, R., Tong, X., Qu, N., Wu, X., et al. (2024). Identifying compound-protein interactions with knowledge graph embedding of perturbation transcriptomics. Cell Genom. *4*, 100655. https://doi.org/10.1016/j.xgen.2024.100655.

32. Rubbi, A., Di Francesco, A.G., Lotfollahi, M., and Liò, P. (2026). A standardized framework for evaluating gene expression generative models. arXiv Preprint, arXiv:2603.11244; presented at the GenAI in Genomics Workshop at ICLR 2026. https://doi.org/10.48550/arXiv.2603.11244.

33. Brown, B.C., Tokolyi, A., Morris, J.A., Lappalainen, T., and Knowles, D.A. (2025). Large-scale causal discovery using interventional data sheds light on gene network structure in K562 cells. Nat. Commun. *16*, 9628. https://doi.org/10.1038/s41467-025-64353-7.

34. Nadig, A., Replogle, J.M., Pogson, A.N., Murthy, M., McCarroll, S.A., Weissman, J.S., Robinson, E.B., and O'Connor, L.J. (2025). Transcriptome-wide analysis of differential expression in perturbation atlases. Nat. Genet. *57*, 1228–1237. https://doi.org/10.1038/s41588-025-02169-3.

35. The Gene Ontology Consortium (2021). The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. *49*, D325–D334. https://doi.org/10.1093/nar/gkaa1113.

36. Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., Senff-Ribeiro, A., Griss, J., Sevilla, C., Matthews, L., Gong, C., et al. (2022). The Reactome pathway knowledgebase 2022. Nucleic Acids Res. *50*, D687–D692. https://doi.org/10.1093/nar/gkab1028.

37. The UniProt Consortium (2023). UniProt: the universal protein knowledgebase in 2023. Nucleic Acids Res. *51*, D523–D531. https://doi.org/10.1093/nar/gkac1052.

38. Hayes, T., Rao, R., Akin, H., Sofroniew, N.J., Oktay, D., Lin, Z., Verkuil, R., Tran, V.Q., Deaton, J., Wiggert, M., et al. (2025). Simulating 500 million years of evolution with a language model. Science *387*, 850–858. https://doi.org/10.1126/science.ads0018.

39. Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A.L., Fang, T., Doncheva, N.T., Pyysalo, S., et al. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. *51*, D638–D646. https://doi.org/10.1093/nar/gkac1000.

40. Garcia-Alonso, L., Holland, C.H., Ibrahim, M.M., Turei, D., and Saez-Rodriguez, J. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. *29*, 1363–1375. https://doi.org/10.1101/gr.240663.118.

41. Müller-Dott, S., Tsirvouli, E., Vazquez, M., Ramirez Flores, R.O., Badia-i-Mompel, P., Fallegger, R., Türei, D., Lægreid, A., and Saez-Rodriguez, J. (2023). Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities. Nucleic Acids Res. *51*, 10934–10949. https://doi.org/10.1093/nar/gkad841.

42. Türei, D., Valdeolivas, A., Gul, L., Palacio-Escat, N., Klein, M., Ivanova, O., Ölbei, M., Gábor, A., Theis, F., Módos, D., et al. (2021). Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. Mol. Syst. Biol. *17*, e9923. https://doi.org/10.15252/msb.20209923.

43. Giurgiu, M., Reinhard, J., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Ruepp, A. (2019). CORUM: the comprehensive resource of mammalian protein complexes–2019. Nucleic Acids Res. *47*, D559–D563. https://doi.org/10.1093/nar/gky973.

44. Heek, J., Hoogeboom, E., Mensink, T., and Salimans, T. (2026). Unified latents (UL): How to train your latents. arXiv Preprint, arXiv:2602.17270. https://doi.org/10.48550/arXiv.2602.17270.

45. Li, T., and He, K. (2026). Back to basics: Let denoising generative models denoise. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 36115–36125. URL https://openaccess.thecvf.com/content/CVPR2026/html/Li_Back_to_Basics_Let_Denoising_Generative_Models_Denoise_CVPR_2026_paper.html.

46. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems, *30*.

47. Vincent, P., Larochelle, H., Bengio, Y., and Manzagol, P.A. (2008). Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th International Conference on Machine Learning, 1096–1103. https://doi.org/10.1145/1390156.1390294.

48. Kingma, D.P., and Welling, M. (2014). Auto-encoding variational Bayes. In International Conference on Learning Representations. URL https://arxiv.org/abs/1312.6114.

49. Good, B.M., Van Auken, K., Hill, D.P., Mi, H., Carbon, S., Balhoff, J.P., Albou, L.P., Thomas, P.D., Mungall, C.J., Blake, J.A., et al. (2021). Reactome and the Gene Ontology: digital convergence of data resources. Bioinformatics *37*, 3343–3348. https://doi.org/10.1093/bioinformatics/btab325.

50. Lin, Z., Hou, J., Xia, H., Xie, X., Wang, F., Zhou, Y., Wang, W., Liu, J., and Qu, L. (2026). Decoupled residual denoising diffusion models for unified and data efficient image-to-image translation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 35967–35977. URL https://openaccess.thecvf.com/content/CVPR2026/html/Lin_Decoupled_Residual_Denoising_Diffusion_Models_for_Unified_and_Data_Efficient_CVPR_2026_paper.html.

51. Song, J., Meng, C., and Ermon, S. (2021). Denoising diffusion implicit models. In International Conference on Learning Representations. URL https://openreview.net/forum?id=St1giarCHLP.

52. Perez, E., Strub, F., de Vries, H., Dumoulin, V., and Courville, A. (2018). FiLM: Visual reasoning with a general conditioning layer. In Proceedings of the AAAI Conference on Artificial Intelligence, *32*, 3942–3951. https://doi.org/10.1609/aaai.v32i1.11671.

53. Peebles, W., and Xie, S. (2023). Scalable diffusion models with transformers. In Proceedings of the IEEE/CVF International Conference on Computer Vision, 4195–4205.

54. Kynkäänniemi, T., Karras, T., Laine, S., Lehtinen, J., and Aila, T. (2019). Improved precision and recall metric for assessing generative models. In Advances in Neural Information Processing Systems, *32*.

55. Naeem, M.F., Oh, S.J., Uh, Y., Choi, Y., and Yoo, J. (2020). Reliable fidelity and diversity metrics for generative models. In Proceedings of the 37th International Conference on Machine Learning, Proceedings of Machine Learning Research, *119*, 7176–7185, PMLR. URL `https://proceedings.mlr.press/v119/naeem20a.html`.

Acknowledgements

This work was supported by the National Key R&D Program of China under grant no. 2025YFC3409003.

Author contributions

Z.Z. and P.H. conceived and designed the study. Z.Z. developed the methodology and software, curated the data, performed the experiments and analyses, validated the results, prepared the visualizations and drafted the manuscript. X.Z., X.D., L.Y. and P.C. contributed to figure design and visualization. Y.-A.H., L.H., T.H., F.T., C.X., Z.-H.Y., J.H. and H.H. contributed to scientific discussions and interpretation of the results. P.H. supervised the study, administered the project, acquired funding and revised the manuscript. All authors reviewed and approved the final manuscript.

Declaration of interests

The authors declare no competing interests.