scGANT: Cell-Conditioned Biological Guidance for Single-Cell Foundation Models

Anonymous authors
Paper under double-blind review

Abstract

Single-cell foundation models learn transferable cellular representations from large-scale transcriptomic atlases (*e.g.* scRNA-seq). Most of these Transformer-based foundation models represent each cell as a sequence of gene tokens and rely on data-driven attention to infer gene dependencies from raw data. However, this principle is fundamentally mismatched to single-cell biology, where *gene dependencies are biologically structured and cell-state specific*. To model such cell-specific gene dependencies, we introduce **scGANT**, a novel single-cell foundation model framework with **cell-conditioned biological guidance**. scGANT includes two key innovations: (i) *Cell-conditioned prior induction* is designed to adapt multi-channel biological priors (including transcription-factor regulation, pathway co-membership, co-expression, and chromosomal proximity) to the current cellular context. (ii) The resulting cell-conditioned prior signals are then incorporated into the *biologically guided attention*, which determines which gene pairs are allowed to communicate and how strongly their interactions are weighted. Across five benchmark datasets spanning *cell-level and gene-level* tasks, scGANT improves macro-F1 on cell type annotation by up to 12.2%, achieves the SOTA frozen-encoder batch-integration overall scIB score, and the best performance on all four differential-expression metrics over 22 held-out perturbation predictions. In an additional case study, scGANT identifies a 12-gene Leiden cluster containing five genes with no direct prior edges among them. Published studies suggest that these genes are involved in related processes. This illustrates scGANT's hypothesis generation beyond direct prior-edge copying. Collectively, these results support biological guidance of gene-token communication as a promising design principle for single-cell foundation models. Code is available at: `https://anonymous.4open.science/r/scGANT-1BFB`.

1 Introduction

AI for Virtual Cell (AIVC) aims to develop computational models that can represent, simulate, and predict cellular behavior across diverse biological contexts. By learning from large-scale single-cell data, AIVC has the potential to support cell-state characterization, perturbation-response prediction, drug discovery, and disease mechanism analysis Bunne et al. (2024); Qian et al. (2025). Inspired by language foundation models, recent single-cell foundation models such as scGPT Cui et al. (2024), scBERT Yang et al. (2022), scFoundation Hao et al. (2024), Geneformer Theodoris et al. (2023), and scPRINT Kalfon et al. (2025) tokenize each cell as a sequence of genes and pretrain transformer-based models on transcriptomic atlases Zhang et al. (2025); Szałata et al. (2024); Liu et al. (2024), providing general-purpose cellular and gene representations that transfer across datasets, cell types, and downstream tasks.

Despite this progress, transformer-based single-cell foundation models inherit a core assumption from language modeling: token representations emerge from expression data-driven self-attention. However, this assumption is fundamentally mismatched with cellular biology, where **genes are not exchangeable linguistic tokens** but act as components of regulatory programs, pathways, and functional modules Feng et al. (2025); Bravo González-Blas et al. (2023). Recent efforts try to incorporate biological gene-to-gene priors into the model, yet typically as input-level token features through gene rankings or positional codes Theodoris et al. (2023), graph-derived embeddings Fang et al. (2025), or functional annotations Kalfon et al. (2025), which augment what each gene token

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

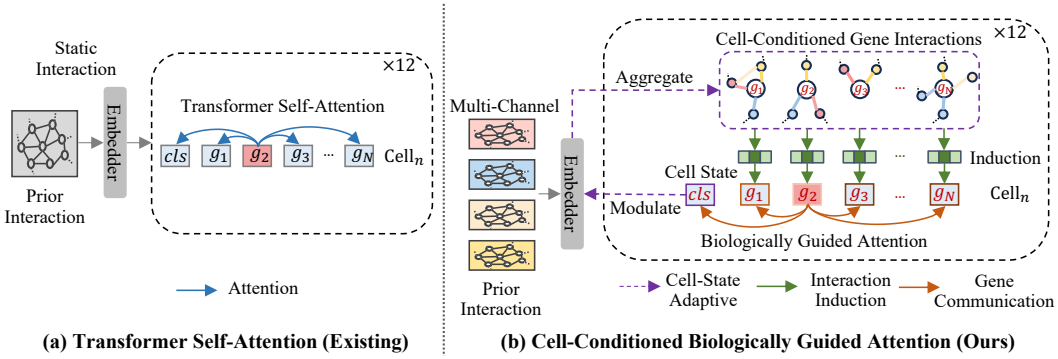


Figure 1: Design motivation of scGANT. (a) Existing models use priors as static input embeddings while gene communication remains governed by self-attention. (b) scGANT converts multi-channel priors into cell-conditioned prior signals that guide dynamic gene communication.

represents but not how gene tokens *interact*. Therefore, we argue that the next step for single-cell foundation models is: ***biological priors should guide how gene tokens exchange information, rather than merely augment gene-token features.***

This raises a deeper challenge that gene regulatory networks are biologically universal, but cells are not: each cell type, disease state, or perturbation condition, activates only a cell-specific subset of prior gene networks Zhang et al. (2023). The same gene pair interaction may be important in one cellular state, but less relevant in another. This makes a fixed prior graph insufficient: although it provides valuable general biological knowledge, it cannot define the effective gene-to-gene communication pattern for all cells Bravo González-Blas et al. (2023); Geiler-Samerotte et al. (2020); Sheinin et al. (2025). Collectively, ***we need to adapt shared biological priors to the current cellular state and use them as cell-conditioned biological guidance for gene-to-gene communication.***

To address these challenges, we introduce ***scGANT***, a novel single-cell foundation model framework with cell-conditioned biological guidance for cellular and gene representation modeling. scGANT models gene-to-gene communication, defined as *attention-mediated information exchange among gene tokens*, through two integrated components. First, *Cell-Conditioned Prior Induction* adapts shared biological prior channels to the current cellular representation, producing cell-conditioned prior signals for scoring candidate gene relationships. Second, *Biologically Guided Attention* combines these signals to route and weight information exchange among genes rather than relying on a fixed biological graph. In this way, scGANT introduces cell-conditioned biological bias directly into gene-token communication.

With this unified architecture, scGANT transfers the same pretrained model across cell-level and gene-level analyses. Across five benchmark datasets spanning three downstream tasks, scGANT improves macro-F1 on Multiple Sclerosis *cell type annotation* by 12.2% over the strongest reported baseline, attains the highest five-metric scIB aggregate on both frozen-embedding *batch integration* benchmarks, and improves differential-expression-focused *perturbation prediction* on held-out targets. Additionally, a gene-level case study examines a scGANT-discovered Leiden cluster whose literature-supported candidate members are not connected to one another by direct edges in the supplied prior graph, illustrating the model's use for biological hypothesis generation. These results motivate *cell-conditioned biological guidance* as a promising design principle for gene-token communication.

2 Related Work

Single-cell foundation models. Recent single-cell foundation models learn transferable cellular representations from large-scale transcriptomic atlases by representing each cell as a sequence of gene tokens and pretraining transformer-based models with self-supervised objectives. Representative models include scBERT Yang et al. (2022) and scGPT Cui et al. (2024), which learn contextualized gene-token representations through masked expression modeling, Geneformer Theodoris et al. (2023), which ranks expressed genes as tokens for network biology prediction, and larger-scale models such as scFoundation Hao et al. (2024), cellFM Zeng et al. (2025), and scPRINT Kalfon et al.

(2025). However, their gene-token communication is still mainly governed by data-driven self-attention, where gene pairs interact under a largely structure-agnostic pattern, but gene interactions are structured and biologically constrained.

Biological priors in single-cell modeling. Another line of work incorporates biological priors such as pathways, gene regulatory networks or functional gene sets. Pathway-guided models such as PASNet Hao et al. (2018) and TOSICA Chen et al. (2023) use curated pathway memberships, while regulation-aware and hierarchy-aware models such as DCell Ma et al. (2018), KPNN Fortelny & Bock (2020), GEARS Roohani et al. (2024), and GRNFormer Qiu et al. (2025) leverage biological networks for cell-state modeling and perturbation prediction. More recent graph-aware foundation models, including GraphCompass Fang et al. (2025), scPRINT Kalfon et al. (2025), scNET Sheinin et al. (2025), and DeepMAPS Ma et al. (2023), further combine foundation learning with gene graph information. These methods show the value of biological structure, but often use fixed graphs that do not adapt to cell state.

3 Design Principles

As shown in Figure 1, the design of scGANT follows two principles: 1) biologically constrained gene-level communication; and 2) cell-conditioned biological guidance.

Principle 1: Biologically constrained gene-level communication. Biological priors should be applied directly to the gene-level communication process in attention rather than only encoded to augment gene-token features. Specifically, prior gene-to-gene interaction graphs, such as gene regulatory networks, are converted into attention guidance that encourages biologically relevant genes to exchange information and down-weights less relevant gene pairs.

Principle 2: Cell-conditioned biological guidance. A shared biological prior graph provides a useful communication scaffold, but it cannot fully capture cell-specific gene interactions. Since the relevance of different biological knowledge sources can vary across cell types, disease states, and perturbation conditions, shared priors should be adapted to the current cell state before guiding gene communication. scGANT therefore uses cellular context to modulate the relative contributions of the four prior channels, producing cell-conditioned prior signals for attention.

4 Methodology: The Proposed scGANT Framework

4.1 Biological Prior Construction

We construct four channel-specific support matrices for TF regulation, pathway co-membership, co-expression, and chromosomal proximity. The data sources, edge definitions, and preprocessing procedures for four prior channels are detailed in Appendix B. For $r \in \{1,2,3,4\}$, $P^{(r)}_{uv} = s^{(r)}_{uv} / \max s^{(r)} \in [0,1]$ is the normalized support for routing from query gene u to gene v. The latter three channels are symmetric, and the TF channel retains its TF→target direction for candidate lookup. For a token pair (g_i, g_j), we collect

$$p_{ij} = \big[P^{(1)}_{g_i g_j},\, P^{(2)}_{g_i g_j},\, P^{(3)}_{g_i g_j},\, P^{(4)}_{g_i g_j}\big]. \tag{1}$$

Here, $p_{ij,r}$ denotes the support from prior channel r.

4.2 scGANT Model

scGANT combines cell-state modeling with biologically guided gene communication. The entering `[CLS]` representation modulates the strengths of four prior channels, and the resulting prior signals are combined with representation-dependent information to select and weight gene-to-gene communication routes. The routed updates are then integrated through gated fusion.

4.2.1 Embedder

Each cell is represented as $s = [g_{\text{cls}}, g_1, \ldots, g_M]$, with a learnable `[CLS]` token and M gene tokens Cui et al. (2024). Its base embeddings are

$$b_i = E_g(g_i) + E_x(x_i) + E_f(f_i), \qquad i = 0, \ldots, M, \tag{2}$$

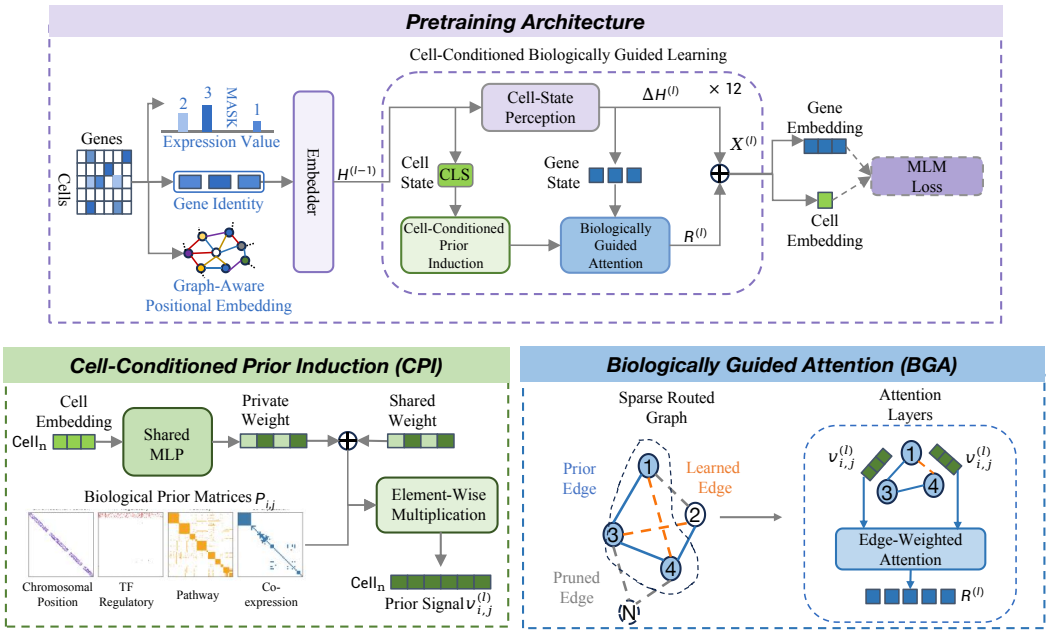


Figure 2: **Overview of scGANT.** The embedder combines gene identity, expression value, masking role, and graph-aware structural information into [CLS] and gene tokens. Cell-State Perception applies pre-normalized dense attention to produce $\Delta H^{(\ell)}$, while the entering [CLS] state conditions four biological priors through Cell-Conditioned Prior Induction (CPI), yielding the cell-conditioned prior signal. Biologically Guided Attention (BGA) combines this signal with attention mechanism to construct a sparse routed graph and produce the updated embedding $R^{(\ell)}$. The residual and dense updates are summed directly, while $R^{(\ell)}$ is scaled by a learned gate before the feed-forward sublayer. After 12 layers, the final cell and gene embeddings are trained with masked-expression reconstruction.

where E_g, E_x, and E_f embed gene identity, expression x_i, and mask role f_i. Masked genes retain their identities but receive zero expression input, and the [CLS] input is also zero. A dedicated initialization router operates directly on the base embeddings $B = [b_0, \ldots, b_M]$ to construct the transition matrix used by standard RWSE Rampášek et al. (2022) for positional embedding. The projected RWSE descriptors are then added to B to obtain $H^{(0)}$, with $K_{\mathrm{rw}} = 8$.

4.2.2 Cell-State Perception

Let $H^{(\ell-1)} = [h_0^{(\ell-1)}, \ldots, h_M^{(\ell-1)}]$ enter layer ℓ. Cell-State Perception is the dense pre-normalized multi-head attention (MHA) update:

$$\Delta H^{(\ell)} = \mathrm{MHA}^{(\ell)}\left(\mathrm{LN}_1^{(\ell)}\left(H^{(\ell-1)}\right)\right). \tag{3}$$

Write its input as $\bar{H}^{(\ell)} = \mathrm{LN}_1^{(\ell)}(H^{(\ell-1)})$. In parallel, prior induction reads the entering [CLS] state $c^{(\ell)} = h_0^{(\ell-1)}$, and routed attention reads the gene rows of $\bar{H}^{(\ell)}$.

4.2.3 Cell-Conditioned Prior Induction

Cell-conditioned prior induction uses the entering [CLS] state to modulate the prior channels. Each layer has a baseline $\bar{\gamma}^{(\ell)} \in \mathbb{R}^4$, while a shared $\mathrm{MLP}_\theta : \mathbb{R}^d \to \mathbb{R}^4$ supplies a cell-conditioned channel offset:

$$\gamma^{(\ell)} = \mathrm{softplus}\left(\bar{\gamma}^{(\ell)} + \mathrm{MLP}_\theta(c^{(\ell)})\right), \tag{4}$$

The softplus yields positive channel strengths $\gamma^{(\ell)}$. CPI performs cell-dependent and channel-wise modulation of the multi-channel prior. Within a given cell and layer, the same four-dimensional $\gamma^{(\ell)}$ is shared across all gene pairs. Pair specificity comes from the static supports $P^{(r)}_{g_i g_j}$. The

initialization-routing pass uses $c^{(0)} = b_0$, so $\gamma^{(0)}$ is shared across cells. These channel strengths define the cell-conditioned prior signal

$$v_{ij}^{(\ell)} = \sum_{r=1}^{4} \gamma_r^{(\ell)} \log\Big(1 + \frac{p_{ij,r}}{\varepsilon}\Big), \qquad \varepsilon = 10^{-6}, \tag{5}$$

where a missing edge in channel r ($p_{ij,r} = 0$) contributes zero.

4.2.4 Biologically Guided Attention

Biologically Guided Attention first selects communicating gene pairs (Sparse routing) and then weights their messages (Guided attention).

Sparse routing. The candidate list combines the B_{prior} strongest prior-supported partners, ranked by $\sum_r p_{ij,r}$, with B_{learn} representation-learned partners as complementary candidates. The latter are the highest-scoring non-self positions under a no-gradient scaled dot-product similarity computed in a d_{proj}-dimensional subspace using the corresponding dense attention Q/K projections. This similarity is used only for candidate proposal. The two lists are concatenated, after which each candidate entry is rescored using the representation term and cell-conditioned prior signal below:

$$\begin{aligned} e_{ij}^{(\ell)} &= \frac{(W_r^{(\ell)} \bar{h}_i^{(\ell)})^\top (W_r^{(\ell)} \bar{h}_j^{(\ell)})}{\sqrt{d_{\text{proj}}}} + v_{ij}^{(\ell)}, \\ \mathcal{N}_i^{(\ell)} &= (\{i\} \setminus \mathcal{M}_{\text{gen}}) \cup \underset{j \in \mathcal{C}_i^{(\ell)} \setminus (\mathcal{M}_{\text{gen}} \cup \{i\})}{\overset{K_r}{\text{ArgTopK}}} \; \text{sg}\Big(e_{ij}^{(\ell)}\Big). \end{aligned} \tag{6}$$

Here, $W_r^{(\ell)}$ projects to d_{proj} dimensions, and $\mathcal{M}_{\text{gen}}$ contains masked reconstruction targets. ArgTopK^{K_r} returns the K_r highest-scoring indices, and sg blocks gradients only through this selection.

Guided attention. For $\mathcal{S}_i^{(\ell)} = \mathcal{N}_i^{(\ell)} \setminus \{i\}$, the selected-edge router weights remain differentiable:

$$w_{ii}^{(\ell)} = 1, \qquad w_{ij}^{(\ell)} = \underset{j \in \mathcal{S}_i^{(\ell)}}{\text{softmax}}\Big(e_{ij}^{(\ell)}\Big), \quad j \in \mathcal{S}_i^{(\ell)}. \tag{7}$$

Each attention head, of dimension d_h, uses $\log w_{ij}^{(\ell)}$ as a routing bias:

$$a_{ij}^{(\ell,h)} = \underset{j \in \mathcal{N}_i^{(\ell)}}{\text{softmax}}\left(\frac{(W_Q^{(\ell,h)} \bar{h}_i^{(\ell)})^\top (W_K^{(\ell,h)} \bar{h}_j^{(\ell)})}{\sqrt{d_h}} + \log w_{ij}^{(\ell)} \right), \tag{8}$$

Values are aggregated as $r_i^{(\ell)} = W_O^{(\ell)} \,\text{Concat}_h\big(\sum_{j \in \mathcal{N}_i^{(\ell)}} a_{ij}^{(\ell,h)} W_V^{(\ell,h)} \bar{h}_j^{(\ell)}\big)$.

Fusion. Let $R^{(\ell)} = [0, r_1^{(\ell)}, \ldots, r_M^{(\ell)}]$, with zero `[CLS]` refinement. A scalar gate fuses the routed, dense, and residual paths:

$$X^{(\ell)} = H^{(\ell-1)} + \Delta H^{(\ell)} + \sigma(\alpha^{(\ell)}) R^{(\ell)}. \tag{9}$$

Here, σ is the sigmoid. A standard pre-normalized FFN residual maps $X^{(\ell)}$ to $H^{(\ell)}$.

4.3 Training Objective and Downstream Use

After $L = 12$ layers, $\widehat{H} = \text{LN}_{\text{out}}(H^{(L)})$. Pretraining minimizes masked expression reconstruction:

$$\mathcal{L}_{\text{rec}} = \frac{1}{|\mathcal{M}_{\text{gen}}|} \sum_{i \in \mathcal{M}_{\text{gen}}} \Big(\text{Dec}(\widehat{h}_i) - x_i^\star\Big)^2, \tag{10}$$

where $x_i^\star$ is the target expression.

For downstream tasks, $\widehat{h}_0$ is used as the cell representation and $\{\widehat{h}_i\}_{i=1}^M$ as gene representations.

5 EXPERIMENTS

5.1 EXPERIMENTAL SETUP

Datasets. We use six public datasets: five benchmarks covering three downstream tasks (Table 1). Cell type annotation on Multiple Sclerosis, with a healthy-donor reference and a patient query, and on hPancreas, with a cross-study split. Frozen-encoder batch integration on the held-out COVID-19 subset and the in-corpus Perirhinal Cortex dataset. Perturbation prediction on Adamson under the GEARS Roohani et al. (2024) *unseen_single* split, where no test target is seen in training. We additionally use Immune Human for a gene-program case study based on frozen contextualized gene embeddings.

Table 1: Datasets and evaluation settings.

Evaluation	Dataset	Tissue	Cell	Gene	Class	Batch	Split / protocol
Cell type annotation	Multiple Sclerosis Schirmer et al. (2019)	Brain	21,312	3,000	18	–	7,059 / 785 / 13,468
	hPancreas Chen et al. (2023)	Pancreas	14,818	3,000	14	–	9,540 / 1,060 / 4,218
Batch integration	COVID-19 Lotfollahi et al. (2022)	Lung / blood / BM	6,883	18,474	28	10	held-out dataset
	Perirhinal Cortex Siletti et al. (2023)	Brain	17,535	59,357	10	2	in-corpus dataset
Perturbation prediction	Adamson Adamson et al. (2016)	K562	68,603	5,060	86	–	57 / 7 / 22 perturbations
Gene-program case study	Immune Human Luecken et al. (2022)	Immune	33,506	12,303	16	10	frozen encoder

BM: bone marrow.

Baselines. For cell type annotation we compare with TOSICA Chen et al. (2023), scBERT Yang et al. (2022), Geneformer Theodoris et al. (2023), scFoundation Hao et al. (2024), scGPT Cui et al. (2024), and TABULA Ding et al. (2026). For batch integration we compare with the dedicated integration methods scVI Lopez et al. (2018) and Seurat Stuart et al. (2019), and with Geneformer, scPRINT Kalfon et al. (2025), and scGPT. For perturbation prediction we compare with GEARS Roohani et al. (2024), PRESAGE Littman et al. (2025), Geneformer, scPRINT, and scGPT.

Implementation. scGANT is pretrained on 37.5M human cells from CellxGene Census Program et al. (2025). Full-corpus pretraining runs for six epochs on four NVIDIA A800 GPUs and takes 180 hours. Cell type annotation attaches a lightweight three-layer classifier head and fine-tunes the backbone on the reference set. Batch integration is evaluated using frozen cell encoders. Perturbation prediction is fine-tuned under the GEARS protocol with a learnable perturbation token and a decoder head. The gene-program case study uses frozen contextualized gene embeddings from scGANT without fine-tuning.

5.2 MAIN RESULTS

5.2.1 CELL-LEVEL EVALUATION

Cell type annotation. Cell type annotation evaluates whether the pretrained encoder learns transferable cellular representations across datasets. We fine-tune scGANT and compare it with representative single-cell foundation models and task-specific baselines on Multiple Sclerosis and hPancreas. As shown in Table 2, scGANT achieves the best score on all four metrics on Multiple Sclerosis, improving macro-F1 over the strongest baseline scFoundation from 0.690 to 0.774, a 12.2% relative gain. On hPancreas, scGANT obtains the best macro-precision, macro-recall, and macro-F1 compared with baselines. Overall, these results indicate that cell-conditioned and biologically guided modeling improves transferable cell representations. Additional visualizations are provided in Appendix E.3.

Batch integration. Batch integration evaluates whether frozen cell representations preserve biological structure while reducing unwanted batch effects. We evaluate the foundation models on the held-out COVID-19 subset of 6,883 cells from 10 batches whose source studies are absent from scGANT pretraining, and on the in-corpus Perirhinal Cortex cohort. On held-out COVID-19, scGANT obtains the best Overall score, 0.681 versus 0.677 for scGPT. Its Avg Bio ties scGPT at 0.549, while its Avg Batch is highest at 0.878. On Perirhinal Cortex, scGANT improves Overall from 0.920 for Seurat to 0.935, driven by the highest Avg Bio score of 0.915 while maintaining a competitive Avg Batch score of 0.964. Appendix Figure 10 visualizes the batch structure of the four frozen foundation-model embeddings on the held-out COVID-19 subset.

Table 2: Cell type annotation. We compare scGANT with representative task-specific and foundation-model baselines. The best is **bolded** and the second is underlined.

Method	Venue	*Multiple Sclerosis (brain, 18 cell types)* Acc↑	Prec↑	Rec↑	F1↑	*hPancreas (pancreas, 14 cell types)* Acc↑	Prec↑	Rec↑	F1↑
TOSICA	NatComm'23	0.758	0.664	0.585	0.578	0.960	0.661	0.681	0.656
scBERT	NMI'22	0.785	0.604	0.624	0.599	0.964	0.699	0.689	0.685
Geneformer	Nature'23	0.612	0.579	0.578	0.554	0.940	0.653	0.649	0.643
scFoundation	NatMeth'24	0.788	0.700	0.695	0.690	0.882	0.604	0.582	0.582
scGPT	NatMeth'24	0.856	0.684	0.703	0.675	0.954	0.793	0.693	0.710
TABULA	NeurIPS'25	—	—	—	—	**0.981**	0.781	0.780	0.771
scGANT (Ours)	—	**0.870**	**0.818**	**0.777**	**0.774**	0.966	**0.847**	**0.787**	**0.784**

Table 3: Batch integration. COVID-19 is restricted to the 10 batches whose source studies are absent from the pretraining corpus (held-out). All Perirhinal Cortex cells are in the pretraining corpus (in-corpus). The best is **bolded** and the second is underlined.

Method	Venue	Bio conservation NMI↑	ARI↑	ASW_c↑	Avg Bio↑	Batch correction ASW_b↑	GC↑	Avg Batch↑	Aggregate Overall↑
COVID-19, held-out (6,883 cells, 10 batches, 28 cell types; 0% pretraining overlap)									
scVI	NatMeth'18	0.648	0.430	0.483	0.520	0.852	0.894	0.873	0.662
Seurat	Cell'19	0.593	0.315	0.486	0.465	0.835	0.845	0.840	0.615
Geneformer	Nature'23	0.642	0.434	0.532	0.536	0.836	0.841	0.838	0.657
scGPT	NatMeth'24	0.662	0.462	0.523	**0.549**	0.844	0.895	0.869	0.677
scPRINT	NatComm'25	0.591	0.406	0.514	0.504	0.898	0.801	0.850	0.642
scGANT (Ours)	—	0.672	0.460	0.515	**0.549**	0.887	0.870	**0.878**	**0.681**
Perirhinal Cortex, in-corpus (17,535 cells, 2 batches, 10 cell types; 100% pretraining overlap)									
scVI	NatMeth'18	0.943	0.952	0.530	0.808	0.921	0.992	0.957	0.868
Seurat	Cell'19	0.926	0.920	0.824	0.890	0.938	0.992	0.965	0.920
Geneformer	Nature'23	0.930	0.914	0.803	0.882	0.951	0.967	0.959	0.913
scGPT	NatMeth'24	0.926	0.911	0.817	0.885	0.931	0.967	0.949	0.910
scPRINT	NatComm'25	0.960	0.976	0.611	0.849	0.966	0.965	**0.966**	0.896
scGANT (Ours)	—	0.972	0.982	0.791	**0.915**	0.962	0.966	0.964	**0.935**

5.2.2 Gene-Level Evaluation

Perturbation prediction. Perturbation prediction tests whether representations capture gene-response structure under unseen interventions. We evaluate on the Adamson dataset using the GEARS `simulation` split in the *unseen_single* setting. All methods are trained and scored under the same split and evaluation code. As shown in Figure 3, scGANT performs best on the differential expression (DE)-focused correlation metrics and also attains the lowest DE-focused MSE, including against PRESAGE, whose priors enter only at the input level. The representative response profiles in Appendix Figure 11 show that scGANT closely tracks ground-truth expression changes, suggesting that it recovers functional response patterns rather than merely smoothing expression profiles.

Gene-program case study. Leiden clustering of frozen scGANT gene embeddings identifies a 12-gene cluster with an automatically assigned label P122 (Figure 4). We selected P122 post hoc as an illustrative case because it combines external annotation support with high prior disconnection. Seven members carry the DisGeNET *Inflammation* annotation Piñero et al. (2020), while the other five (*BCL2A1*, *G0S2*, *KLF10*, *NINJ1*, and *STX11*) have no direct prior edges among themselves, and only 11 of all 66 within-cluster pairs are directly connected (Figure 4a).

Independent studies link these candidates to NF-κB-mediated survival (BCL2A1) Zong et al. (1999), monocyte inflammation (G0S2) Okabe et al. (2022), macrophage responses to infection (KLF10) Madrid-Paulino et al. (2022), inflammatory membrane rupture (NINJ1) Kayagaki et al. (2021), and familial hemophagocytic lymphohistiocytosis type 4 (STX11) zur Stadt et al. (2005). P122 therefore illustrates hypothesis generation beyond direct prior edges. Protocols are provided in Appendix D.3.2.

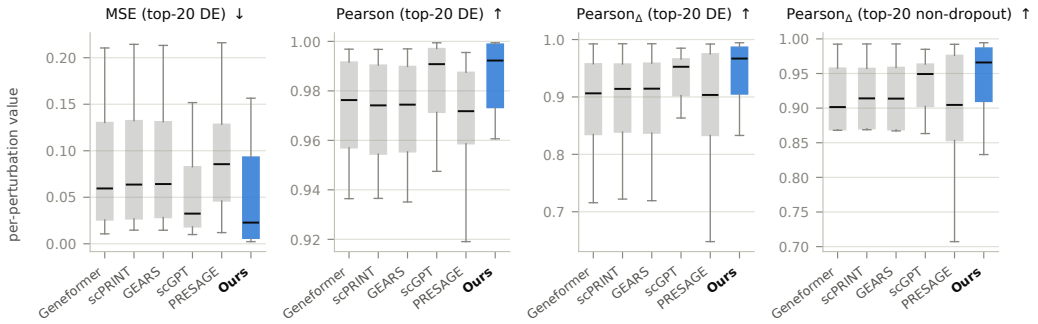


Figure 3: **Adamson results across 22 held-out perturbations on the top-20 DE genes.** Metrics include MSE, Pearson correlation, Pearson$_\Delta$, and Pearson$_\Delta$ on non-dropout genes.

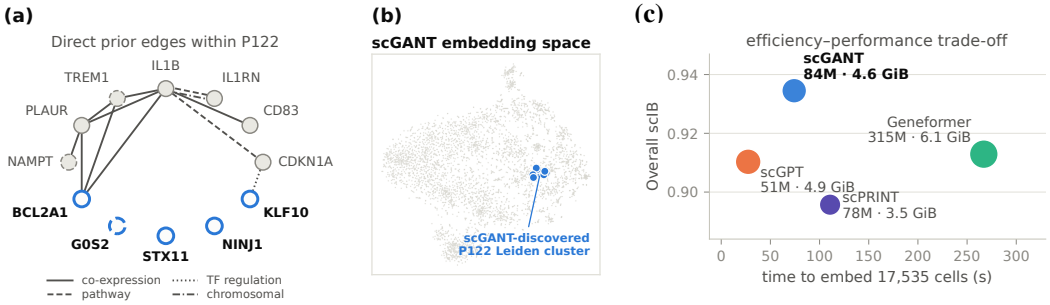


Figure 4: **Case study and efficiency analysis. (a)** Direct prior edges among the 12 genes in the scGANT-discovered Leiden cluster P122. Blue outlines mark the five genes without the *Inflammation* annotation. They have no mutual direct edges. Dashed outlines mark genes absent from panel (b). **(b)** t-SNE of 3,887 scGANT gene embeddings. Blue points mark the nine visualized P122 genes and lines join their 29 prior-disconnected pairs. **(c)** Overall scIB versus the time required to embed 17,535 Perirhinal Cortex cells. Marker area denotes peak memory, and labels report parameter count and peak memory.

5.3 Ablation Study

Ablation Study on Model Components. Table 4 compares five configurations pretrained on 500K cells. For the SPI- and CPI-only configurations (BGA $\times$), the induced prior signal is incorporated into dense self-attention as a head-shared additive attention bias. Starting from the CSP-only baseline, static prior induction (SPI) improves performance across all three tasks, while replacing SPI with cell-conditioned prior induction (CPI) yields further consistent gains, supporting the benefit of adapting prior contributions to the current cell state. The prior-free BGA variant also improves over CSP alone, with particularly strong gains on integration and perturbation prediction. Full model with CPI and BGA achieves the best performance on all three tasks, suggesting that cell-conditioned prior induction and biologically guided attention provide complementary benefits.

Ablation Study on Priors. Table 5 removes one prior channel at a time while leaving the remaining model unchanged. Every removal lowers performance on all three tasks, and removing TF regulation causes the largest degradation in annotation, held-out integration, and perturbation prediction. The remaining channels show smaller but consistent contributions, supporting complementary multi-prior fusion rather than dependence on a single source.

Model Efficiency Analysis. At the fixed 512-token length, scGANT occupies the strongest overall position in the efficiency-performance plane (Figure 4c). Relative to Geneformer, it achieves a higher Overall scIB score while running 3.6$\times$ faster (74 s vs. 267 s), with 3.8$\times$ fewer parameters and 1.6 GiB less peak memory. scGPT remains the fastest and smallest model, but its Overall score is 0.025 lower and its peak memory is higher than that of scGANT. Together, these comparisons show that scGANT improves representation quality without sacrificing practical efficiency, yielding a favorable performance, runtime, and memory trade-off at $M = 512$.

Table 4: Ablation on model components.

Component				Downstream task		
CSP	SPI	CPI	BGA	Anno. Acc.↑	Integ. Overall↑	Pert. P_{Δ}^{DE} ↑
✓	×	×	×	0.8392	0.6219	0.7776
✓	✓	×	×	0.8440	0.6382	0.7828
✓	×	✓	×	0.8478	0.6429	0.7877
✓	×	×	✓	0.8427	0.6473	0.7901
✓	×	✓	✓	**0.8563**	**0.6548**	**0.7925**

Table 5: Ablation on prior channels.

Prior channel				Downstream task		
chrpos	TF	coexp	pathway	Anno. Acc.↑	Integ. Overall↑	Pert. P_{Δ}^{DE} ↑
×	✓	✓	✓	0.8407	0.6361	0.7808
✓	×	✓	✓	0.8383	0.6319	0.7608
✓	✓	×	✓	0.8394	0.6480	0.7769
✓	✓	✓	×	0.8385	0.6421	0.7783
✓	✓	✓	✓	**0.8563**	**0.6548**	**0.7925**

Notes. *CSP*: cell state perception; *SPI*: static prior induction; *CPI*: cell-conditioned prior induction; *BGA*: biologically guided attention. The best result is **bolded**.

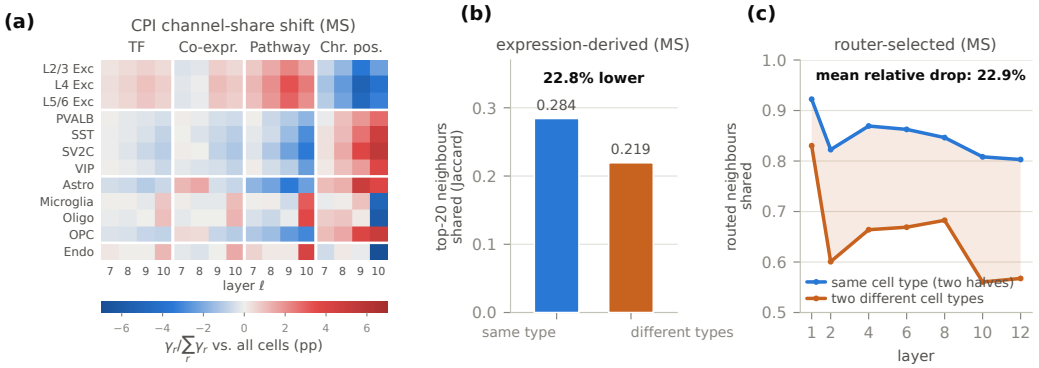


Figure 5: **Mechanism analyses of prior modulation and routed communication. (a)** Differences in normalized prior-channel shares relative to the population mean across cell types and layers. **(b-c)** On Multiple Sclerosis, Jaccard overlap of top-20 expression-derived and router-selected neighbours within versus across cell types. Router cross-type overlap is lower at every evaluated layer.

Visualization of Cell-Conditioned Prior Modulation. To examine whether CPI can learn context-dependent prior weighting, Figure 5(a) reports deviations in the normalized channel shares $\gamma_r / \sum_r \gamma_r$ from the population mean across cell types and layers. The cell-type-dependent and layer-dependent deviations across channels show that CPI learns context-dependent relative channel mixtures rather than a single global weighting.

Visualization of Routing Analysis. Expression-derived neighbourhood overlap falls from 0.284 within cell types to 0.219 across types (22.8%; Figure 5b), consistent with cell-type-dependent transcriptional structure. Router-selected neighbourhoods show the same separation at every evaluated layer, with a 22.9% average relative drop (Figure 5c). This persistent gap shows that routing remains cell-type dependent rather than collapsing to a shared neighbourhood.

6 Discussion

Conclusion. Our results support a biologically guided principle for single-cell foundation models that biological priors should guide gene communication, not merely augment initial gene features. scGANT combines cell-conditioned prior induction, which adapts the contributions of multiple prior channels, and biologically guided attention, which combines the resulting prior signals with attention mechanism to construct dynamic gene communication routes. The strong improvements across cell- and gene-level downstream tasks suggest that this benefit is architectural rather than task-specific.

Limitations. Our experiments focus on human scRNA-seq datasets. Future work will assess the transferability of the framework to other species and single-cell modalities. Cross-species transfer will require careful treatment of ortholog mapping and the reconstruction of species-specific biological priors. Extension to modalities such as scATAC-seq and spatial transcriptomics will further require modality-aware tokenization. Such validation would clarify whether cell-conditioned biological guidance remains effective across different gene vocabularies and settings.

AI USE STATEMENT

In this work, generative AI tools were used to assist with language polishing. No AI-generated output was treated as experimental evidence or used to determine the reported results. The authors reviewed, revised, and verified all AI-assisted material and take full responsibility for the final text, claims, and artifacts in this work.

ETHICS STATEMENT

This study uses only previously published, publicly available datasets and does not collect new human-subject data or personally identifiable information. This study therefore raises no ethical concerns.

REPRODUCIBILITY STATEMENT

The model formulation and training objectives are described in Section 4, and the datasets, evaluation protocols, and metrics are summarized in Section 5.1. Hyperparameters, training configurations, and detailed results are provided in the appendix. Key source code is available through the anonymous repository at `https://anonymous.4open.science/r/scGANT-1BFB`.

REFERENCES

Britt Adamson, Thomas M Norman, Marco Jost, Min Y Cho, James K Nuñez, Yuwen Chen, Jacqueline E Villalta, Luke A Gilbert, Max A Horlbeck, Marco Y Hein, et al. A multiplexed single-cell crispr screening platform enables systematic dissection of the unfolded protein response. *Cell*, 167(7):1867–1882, 2016.

Carmen Bravo González-Blas, Seppe De Winter, Gert Hulselmans, Nikolai Hecker, Irina Matetovici, Valerie Christiaens, Suresh Poovathingal, Jasper Wouters, Sara Aibar, and Stein Aerts. Scenic+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nature methods*, 20 (9):1355–1367, 2023.

Charlotte Bunne, Yusuf Roohani, Yanay Rosen, Ankit Gupta, Xikun Zhang, Marcel Roed, Theo Alexandrov, Mohammed AlQuraishi, Patricia Brennan, Daniel B Burkhardt, et al. How to build the virtual cell with artificial intelligence: Priorities and opportunities. *Cell*, 187(25):7045–7063, 2024.

Jiawei Chen, Hao Xu, Wanyu Tao, Zhaoxiong Chen, Yuxuan Zhao, and Jing-Dong J Han. Transformer for one stop interpretable cell type annotation. *Nature Communications*, 14(1):223, 2023.

Haotian Cui, Chloe Wang, Hassaan Maan, Kuan Pang, Fengning Luo, Nan Duan, and Bo Wang. scgpt: toward building a foundation model for single-cell multi-omics using generative ai. *Nature methods*, 21(8):1470–1480, 2024.

Jiayuan Ding, Jianhui Lin, Shiyu Jiang, Yixin Wang, Ziyang Miao, Zhaoyu Fang, Jiliang Tang, Min Li, and Xiaojie Qiu. Tabula: A tabular self-supervised foundation model for single-cell transcriptomics. *Advances in Neural Information Processing Systems*, 38:103789–103808, 2026.

Chen Fang, Wentao Cui, Zhilong Hu, Wenhao Liu, Shubai Chen, Shaole Chang, Qingqing Long, Cong Li, Yana Liu, Haiping Jiang, et al. Cell-graphcompass: modeling single cells with graph structure foundation model. *National Science Review*, 12(10):nwaf255, 2025.

Zhanying Feng, Xi Chen, Zhana Duren, Jingxue Xin, Hao Miao, Qiuyue Yuan, Yong Wang, and Wing Hung Wong. Modeling combinatorial regulation from single-cell multi-omics provides regulatory units underpinning cell type landscape using cregulon. *Genome Biology*, 26(1):220, 2025.

Nikolaus Fortelny and Christoph Bock. Knowledge-primed neural networks enable biologically interpretable deep learning on single-cell sequencing data. *Genome biology*, 21(1):190, 2020.

Luz Garcia-Alonso, Christian H Holland, Mahmoud M Ibrahim, Denes Turei, and Julio Saez-Rodriguez. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome research*, 29(8):1363–1375, 2019.

Kerry A Geiler-Samerotte, Shuang Li, Charalampos Lazaris, Austin Taylor, Naomi Ziv, Chelsea Ramjeawan, Annalise B Paaby, and Mark L Siegal. Extent and context dependence of pleiotropy revealed by high-throughput single-cell phenotyping. *PLoS biology*, 18(8):e3000836, 2020.

Jie Hao, Youngsoon Kim, Tae-Kyung Kim, and Mingon Kang. Pasnet: pathway-associated sparse deep neural network for prognosis prediction from high-throughput data. *BMC bioinformatics*, 19(1):510, 2018.

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

Peter W Harrison, M Ridwan Amode, Olanrewaju Austine-Orimoloye, Andrey G Azov, Matthieu Barba, If Barnes, Arne Becker, Ruth Bennett, Andrew Berry, Jyothish Bhai, et al. Ensembl 2024. *Nucleic acids research*, 52(D1):D891–D899, 2024.

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

Nobuhiko Kayagaki, Opher S Kornfeld, Bettina L Lee, Irma B Stowe, Karen O'Rourke, Qingling Li, Wendy Sandoval, Donghong Yan, Jing Kang, Min Xu, et al. Ninj1 mediates plasma membrane rupture during lytic cell death. *Nature*, 591(7848):131–136, 2021.

Russell Littman, Jacob Levine, Sepideh Maleki, Yongju Lee, Vladimir Ermakov, Lin Qiu, Alexander Wu, Kexin Huang, Romain Lopez, Gabriele Scalia, et al. Gene-embedding-based prediction and functional evaluation of perturbation expression responses with presage. *bioRxiv*, pp. 2025–06, 2025.

Tianyu Liu, Kexing Li, Yuge Wang, Hongyu Li, and Hongyu Zhao. Evaluating the utilities of foundation models in single-cell data analysis. *Advanced Science*, pp. e14490, 2024.

Romain Lopez, Jeffrey Regier, Michael B Cole, Michael I Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nature methods*, 15(12):1053–1058, 2018.

Mohammad Lotfollahi, Mohsen Naghipourfar, Malte D Luecken, Matin Khajavi, Maren Büttner, Marco Wagenstetter, Žiga Avsec, Adam Gayoso, Nir Yosef, Marta Interlandi, et al. Mapping single-cell data to reference atlases by transfer learning. *Nature biotechnology*, 40(1):121–130, 2022.

Malte D Luecken, Maren Büttner, Kridsadakorn Chaichoompu, Anna Danese, Marta Interlandi, Michaela F Müller, Daniel C Strobl, Luke Zappia, Martin Dugas, Maria Colomé-Tatché, et al. Benchmarking atlas-level data integration in single-cell genomics. *Nature methods*, 19(1):41–50, 2022.

Anjun Ma, Xiaoying Wang, Jingxian Li, Cankun Wang, Tong Xiao, Yuntao Liu, Hao Cheng, Juexin Wang, Yang Li, Yuzhou Chang, et al. Single-cell biological network inference using a heterogeneous graph transformer. *Nature Communications*, 14(1):964, 2023.

Jianzhu Ma, Michael Ku Yu, Samson Fong, Keiichiro Ono, Eric Sage, Barry Demchak, Roded Sharan, and Trey Ideker. Using deep learning to model the hierarchical structure and function of a cell. *Nature methods*, 15(4):290–298, 2018.

Edgardo Madrid-Paulino, Dulce Mata-Espinosa, Juan Carlos León-Contreras, Isela Serrano-Fujarte, Sol Díaz de León-Guerrero, Tomás Villaseñor, Lucero Ramon-Luing, José L Puente, Leslie Chavez-Galan, Rogelio Hernández-Pando, et al. Klf10 favors mycobacterium tuberculosis survival by impairing ifn-γ production and preventing macrophages reprograming to macropinocytosis. *Journal of Leukocyte Biology*, 112(3):475–490, 2022.

Marija Milacic, Deidre Beavers, Patrick Conley, Chuqiao Gong, Marc Gillespie, Johannes Griss, Robin Haw, Bijay Jassal, Lisa Matthews, Bruce May, et al. The reactome pathway knowledgebase 2024. *Nucleic acids research*, 52(D1):D672–D678, 2024.

Sophia Müller-Dott, Eirini Tsirvouli, Miguel Vazquez, Ricardo O Ramirez Flores, Pau Badia-i Mompel, Robin Fallegger, Dénes Türei, Astrid Lægreid, and Julio Saez-Rodriguez. Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities. *Nucleic acids research*, 51(20):10934–10949, 2023.

Mako Okabe, Shinya Takarada, Nariaki Miyao, Hideyuki Nakaoka, Keijiro Ibuki, Sayaka Ozawa, Kazuhiro Watanabe, Harue Tsuji, Ikuo Hashimoto, Kiyoshi Hatasaki, et al. G0s2 regulates innate immunity in kawasaki disease via lncrna hsd11b1-as1. *Pediatric research*, 92(2):378–387, 2022.

Janet Piñero, Juan Manuel Ramírez-Anguita, Josep Saüch-Pitarch, Francesco Ronzano, Emilio Centeno, Ferran Sanz, and Laura I Furlong. The disgenet knowledge platform for disease genomics: 2019 update. *Nucleic acids research*, 48(D1):D845–D855, 2020.

CZI Cell Science Program, Shibla Abdulla, Brian Aevermann, Pedro Assis, Seve Badajoz, Sidney M Bell, Emanuele Bezzi, Batuhan Cakir, Jim Chaffer, Signe Chambers, et al. Cz cellxgene discover: a single-cell data platform for scalable exploration, analysis and modeling of aggregated data. *Nucleic acids research*, 53(D1):D886–D900, 2025.

Liujia Qian, Zhen Dong, and Tiannan Guo. Grow ai virtual cells: three data pillars and closed-loop learning. *Cell Research*, 35(5):319–321, 2025.

Mufan Qiu, Xinyu Hu, Fengwei Zhan, Sukwon Yun, Jie Peng, Ruichen Zhang, Bhavya Kailkhura, Jiekun Yang, and Tianlong Chen. Grnformer: A biologically-guided framework for integrating gene regulatory networks into rna foundation models. In *Findings of the Association for Computational Linguistics: ACL 2025*, pp. 3805–3819, 2025.

Ladislav Rampášek, Michael Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, and Dominique Beaini. Recipe for a general, powerful, scalable graph transformer. *Advances in Neural Information Processing Systems*, 35:14501–14515, 2022.

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

Lucas Schirmer, Dmitry Velmeshev, Staffan Holmqvist, Max Kaufmann, Sebastian Werneburg, Diane Jung, Stephanie Vistnes, John H Stockley, Adam Young, Maike Steindel, et al. Neuronal vulnerability and multilineage diversity in multiple sclerosis. *Nature*, 573(7772):75–82, 2019.

Ron Sheinin, Roded Sharan, and Asaf Madi. scnet: learning context-specific gene and cell embeddings by integrating single-cell gene expression data with protein–protein interactions. *Nature Methods*, 22(4):708–716, 2025.

Kimberly Siletti, Rebecca Hodge, Alejandro Mossi Albiach, Ka Wai Lee, Song-Lin Ding, Lijuan Hu, Peter Lönnerberg, Trygve Bakken, Tamara Casper, Michael Clark, et al. Transcriptomic diversity of cell types across the adult human brain. *Science*, 382(6667):eadd7046, 2023.

Tim Stuart, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M Mauck, Yuhan Hao, Marlon Stoeckius, Peter Smibert, and Rahul Satija. Comprehensive integration of single-cell data. *cell*, 177(7):1888–1902, 2019.

Artur Szałata, Karin Hrovatin, Sören Becker, Alejandro Tejada-Lapuerta, Haotian Cui, Bo Wang, and Fabian J Theis. Transformers in single-cell omics: a review and new perspectives. *Nature methods*, 21(8):1430–1443, 2024.

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

Fan Yang, Wenchuan Wang, Fang Wang, Yuan Fang, Duyu Tang, Junzhou Huang, Hui Lu, and Jianhua Yao. scbert as a large-scale pretrained deep language model for cell type annotation of single-cell rna-seq data. *Nature machine intelligence*, 4(10):852–866, 2022.

Yuansong Zeng, Jiancong Xie, Ningyuan Shangguan, Zhuoyi Wei, Wenbing Li, Yun Su, Shuangyu Yang, Chengyang Zhang, Jinbo Zhang, Nan Fang, et al. Cellfm: a large-scale foundation model pre-trained on transcriptomics of 100 million human cells. *Nature Communications*, 16(1):4679, 2025.

Fan Zhang, Hao Chen, Zhihong Zhu, Ziheng Zhang, Zhenxi Lin, Ziyue Qiao, Yefeng Zheng, and Xian Wu. A survey on foundation language models for single-cell biology. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 528–549, 2025.

Shilu Zhang, Saptarshi Pyne, Stefan Pietrzak, Spencer Halberg, Sunnie Grace McCalla, Alireza Fotuhi Siahpirani, Rupa Sridharan, and Sushmita Roy. Inference of cell type-specific gene regulatory networks on cell lineages from single cell omic datasets. *Nature Communications*, 14(1):3064, 2023.

Wei-Xing Zong, Leonard C Edelstein, Cailin Chen, Judy Bash, and Céline Gélinas. The prosurvival bcl-2 homolog bfl-1/a1 is a direct transcriptional target of nf-κb that blocks tnfα-induced apoptosis. *Genes & development*, 13(4):382–387, 1999.

Udo zur Stadt, Susanne Schmidt, Brigitte Kasper, Karin Beutel, A Sarper Diler, Jan-Inge Henter, Hartmut Kabisch, Reinhard Schneppenheim, Peter Nurnberg, Gritta Janka, et al. Linkage of familial hemophagocytic lymphohistiocytosis (fhl) type-4 to chromosome 6q24 and identification of mutations in syntaxin 11. *Human molecular genetics*, 14(6):827–834, 2005.

A Pretraining Data Collection and Processing

Data Collection Our pretraining corpus is built from the human CellxGene Census collection Program et al. (2025), a large cross-study single-cell compendium with a unified gene vocabulary and harmonized metadata. The raw corpus used in this work contains 37,489,811 cells over the 60,694-gene Census vocabulary. We use this scale to expose the model to broad cross-tissue, cross-study, and cross-batch variation before downstream supervision, which is essential for learning transferable foundation-model representations.

For scalable training, we store the corpus as 192 per-tissue Census partitions and stream them during full-corpus pretraining. To avoid ordering effects from file partitions or tissue grouping, we randomly shuffle cells across the pretraining stream before constructing mini-batches. The same shuffled loading pipeline is reused for smaller controlled subsets used in the ablation studies, ensuring that differences are not caused by data-format, partition order, or sampling artifacts. Throughout the paper, the full-corpus model refers to the model pretrained on this shuffled CellxGene stream with scGANT tokenizer and objective.

Quality Control and Data Preprocessing All cells are converted into sparse gene-token sequences. No task-specific labels or metadata are used during pretraining, and the model sees only expression-derived token sequences. For sequence construction, cells with more non-zero genes than the maximum token budget are sampled to a fixed length without replacement, with a default maximum sequence length of 512. As shown in Figure 6, the pretrained representation already organizes broad tissue structure before task-specific fine-tuning, suggesting that scGANT learns transferable biological geometry from unsupervised pretraining.

B Biological Prior Knowledge Collection and Preprocessing

Multi-Channel Prior Graph scGANT uses $n = 4$ parallel prior channels over the shared gene vocabulary: TF regulation (r=1), pathway co-membership (r=2), co-expression (r=3), and chromosomal proximity (r=4). Because no single resource is complete across all cellular contexts, the channels are built independently. Table 6 specifies each channel.

Table 6: Construction of the four biological prior channels of scGANT.

Prior channel	Resource	Edge criterion	Affinity construction	Normalization $P^{(r)}_{uv} = s^{(r)}_{uv} / \max s^{(r)}$
TF regulation	DoRothEA (levels A+B) Garcia-Alonso et al. (2019) + CollecTRI Müller-Dott et al. (2023)	Directed TF→target edges	Binary adjacency	identity ($\max s^{(r)} = 1$)
Pathway	Reactome Milacic et al. (2024)	Shared pathway membership (≥ 5 genes)	Size-weighted co-membership	rescaling
Co-expression	Pretraining CellxGene Census Program et al. (2025)	Pearson $\lvert r\rvert > 0.36$ on $\log 1p$	Absolute Pearson correlation affinity	rescaling
Chromosomal proximity	Ensembl Harrison et al. (2024)	Same chromosome, $d \leq 1$ Mb	Inverse-distance affinity	identity ($\max s^{(r)} = 1$)

Statistical Overview of the Multi-Channel Prior As shown in Figure 7, the four channels differ substantially in edge density, gene coverage, and degree distribution, and their pairwise overlap is limited, indicating that each contributes distinct evidence about gene organization. This heterogeneity motivates the cell-conditioned fusion module, which learns to re-weight prior sources according to cellular context.

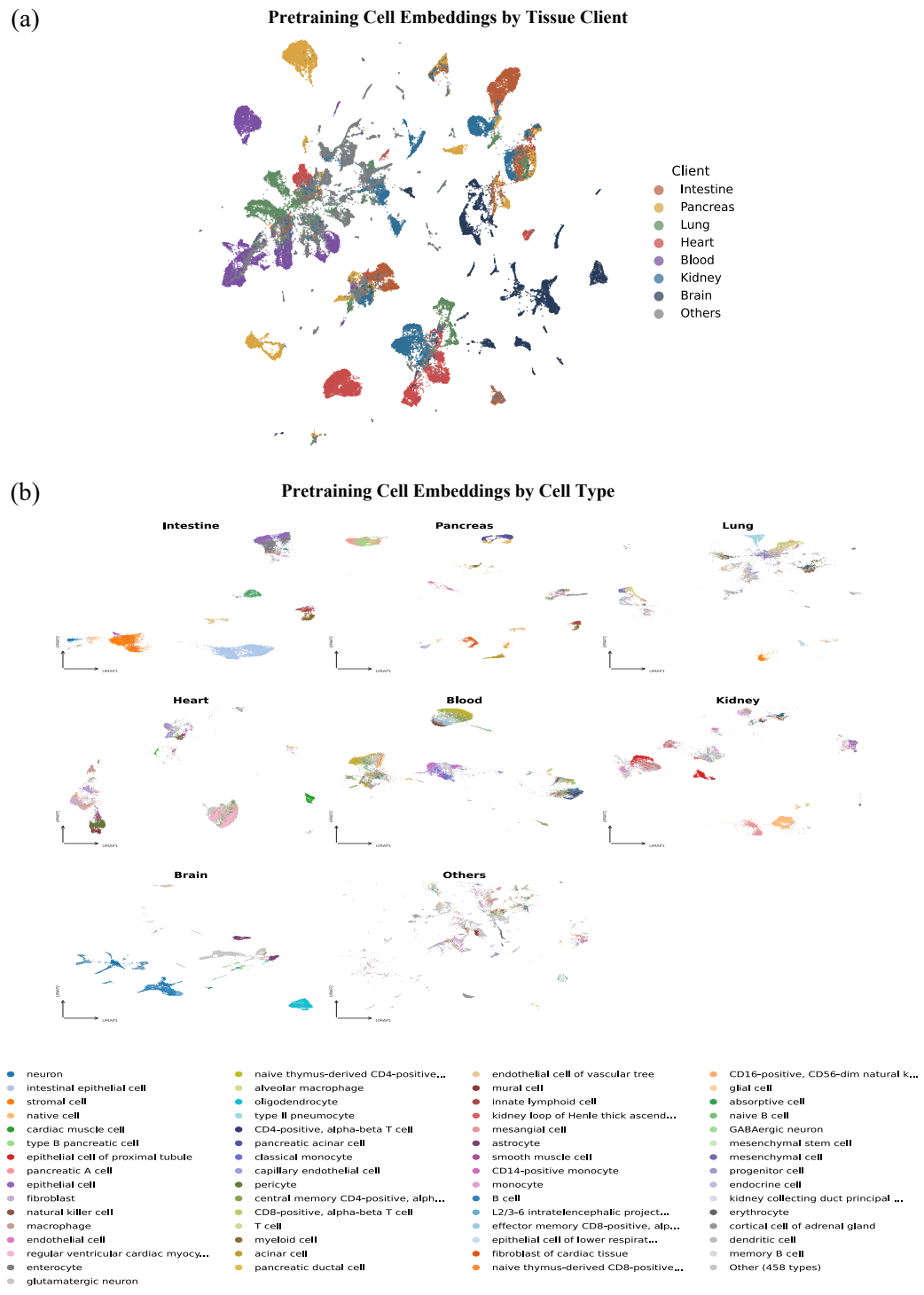


Figure 6: Pretrained cell embeddings across tissue clients. (a) UMAP visualization of sampled CellxGene pretraining cells colored by broad tissue client. (b) Cell-type-colored UMAPs within each tissue group.

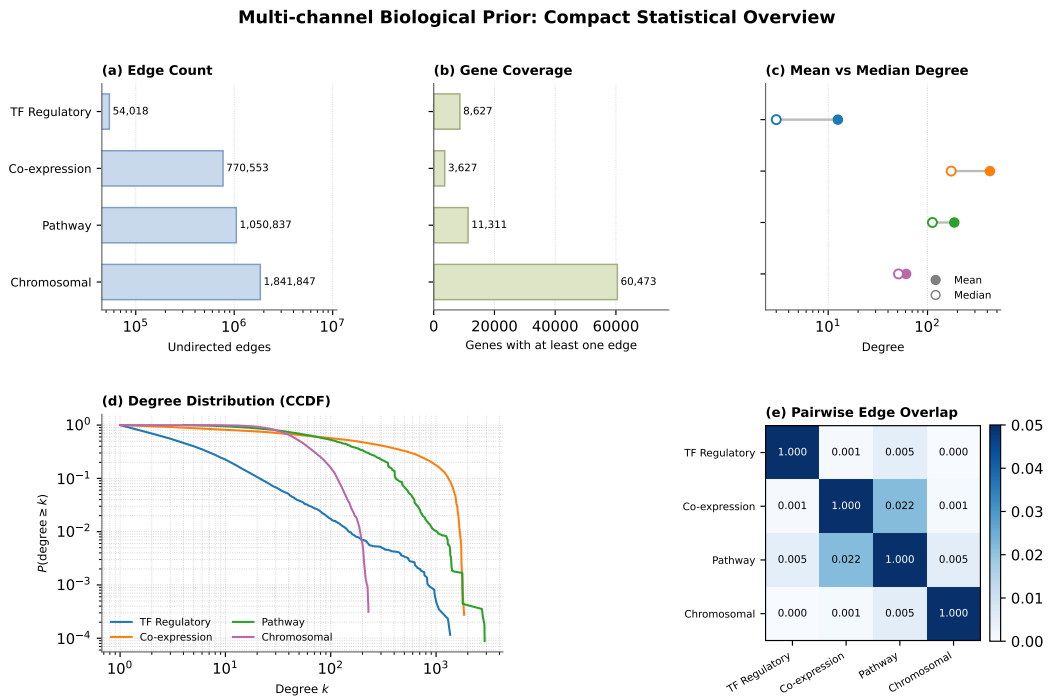


Figure 7: Statistics of the four biological prior channels used by scGANT. We summarize edge count, gene coverage, mean and median degree, degree distribution, and pairwise edge overlap for TF-regulatory, co-expression, pathway, and chromosomal-position channels. The channels differ substantially in density and coverage and have limited overlap, motivating cell-conditioned multi-prior fusion rather than reliance on a single fixed graph.

C Pretraining Implementation Details

Main Model Configuration Table 7 lists the architecture of the scGANT configuration used in the full-corpus experiments.

Table 7: scGANT architecture in the full-corpus experiments.

	Component	Value
Attention layers	Gene vocabulary	60,694
	Hidden size d	512
	Layers	12
	Attention heads	8
	Feed-forward size	2,048
	Dropout	0.1
Biologically guided refinement	Prior candidates B_{prior}	32
	Learned candidates B_{learned}	64
	Routed neighborhood K_r	32
	Routing projection d_{proj}	32
	Dynamic prior conditioning	enabled
Gate / encoding	RWSE	enabled, $K_{\text{rw}} = 8$
	Fusion gate init α_{init}	-1.0
Decoder	Expression head (per gene token)	$512 \to 512 \to 512 \to 1$ MLP
	Prediction target	scalar $\log 1p$ expression

Pretraining Hyperparameter Details All reported scGANT variants use the same masked expression reconstruction objective, tokenizer, maximum sequence length, and optimizer family. Table 8 summarizes the full-corpus recipe. Masking is applied per cell: a uniformly sampled 15% of

that cell's gene tokens retains its gene identity but receives a zero expression input and a masked-role flag. Its expression is reconstructed from the resulting sequence.

Table 8: Full-corpus pretraining recipe.

	Hyperparameter	Value
Data / objective	Maximum sequence length	512
	Mask ratio	0.15 (uniform)
	Prior candidate size	32
	Learned candidate size	64
Optimization	Optimizer	AdamW
	Learning rate	1×10^{-4}
	Weight decay	0.01
	Warmup ratio	0.05 (linear)
	LR schedule	cosine, $\eta_{\min} = 10^{-6}$
Run	Epochs	6
	GPUs	$4\times$ A800
	Batch size per GPU	32
	Precision	mixed precision

Computational complexity The routed aggregation operates on at most $K_r = 32$ selected neighbors per query and costs $O(MK_r)$ when feature dimensions are treated as fixed. The complete architecture is not asymptotically sparse: every layer retains dense MHA with quadratic sequence-length dependence, and learned-candidate construction at each routing update computes a query-key similarity matrix in a $d_{\text{proj}} = 32$ subspace, costing $O(M^2 d_{\text{proj}})$. The prior-candidate lookup costs $O(MB_{\text{prior}})$. Learned-candidate construction reuses the dense-MHA query and key projection parameters, runs without gradients, and is evaluated once every two layers. These choices reduce its constant and memory overhead but do not remove its quadratic dependence on M.

Pretraining Compute Budget The full-corpus run is executed on four NVIDIA A800 GPUs in distributed data-parallel mode, with four-channel priors. It takes 180 hours of wall-clock time, roughly 30 hours per epoch. Checkpoints are written periodically and at every epoch boundary.

D Evaluation Datasets and Protocols

D.1 Overview

We evaluate three downstream benchmark tasks: the cell-level tasks of cell type annotation and batch integration, and the gene-level task of perturbation prediction. We additionally conduct a gene-program case study using frozen contextualized gene embeddings.

D.2 Cell-Level Benchmarks

D.2.1 Cell Type Annotation

Datasets Multiple Sclerosis Schirmer et al. (2019) is a disease-state transfer: nine healthy control donors form the reference set and twelve multiple sclerosis patients the query set, so the two share neither individuals nor disease state. hPancreas Chen et al. (2023) uses the fixed train and test split distributed with the benchmark. In both cases the reference set is split $90\%/10\%$ into training and validation, stratified by cell type, and the entire query set is held out for testing. We report accuracy, macro-precision, macro-recall, and macro-F1, emphasizing macro metrics because both datasets contain minority cell populations.

Protocols A lightweight three-layer classifier head is attached to the final cell embedding and trained with cross-entropy over cell-type labels. The head is optimized with Adam at learning rate 10^{-4}, gradient clipping at 1.0, batch size 32, and a StepLR schedule with multiplicative decay 0.9 per epoch, for 10 epochs, selecting the checkpoint by validation accuracy.

D.2.2 BATCH INTEGRATION

Datasets COVID-19 Lotfollahi et al. (2022) is restricted to the held-out subset of 6,883 cells from 10 batches and 28 cell types whose source studies are absent from the scGANT pretraining corpus. Perirhinal Cortex Siletti et al. (2023) uses all 17,535 cells from its two batches and is in-corpus. Within each evaluation cohort, all batches are embedded jointly and scored with scIB Luecken et al. (2022), where NMI, ARI, and $\mathrm{ASW}_{\mathrm{cell}}$ measure biological conservation and $\mathrm{ASW}_{\mathrm{batch}}$ and graph connectivity measure batch mixing. Bio and Batch are the means of the two groups, and Overall $= 0.6\,\mathrm{Bio} + 0.4\,\mathrm{Batch}$.

Protocols For the foundation models, the pretrained encoder is frozen and the cell embedding is read from each model's cell-level readout, L_2-normalized, and scored with the same scIB implementation. Seurat and scVI are instead fitted separately on each evaluation dataset using batch labels, and their integrated representations are scored with the same scIB implementation.

D.3 GENE-LEVEL EVALUATION

D.3.1 PERTURBATION PREDICTION

Datasets We use the Adamson Adamson et al. (2016) benchmark with the GEARS Roohani et al. (2024) `simulation` split in the *unseen_single* setting: 57 training, 7 validation, and 22 test perturbations, whose target genes are never perturbed during training. We report MSE, Pearson correlation, Pearson correlation on differential expression, and the top-20 DE-focused metrics, on both the full reconstructed profile and the differential response relative to control.

Protocols We follow the GEARS setting and fine-tune the full-corpus checkpoint with a learnable perturbation token and a three-layer decoder head under plain MSE training on the perturbed profile. Fine-tuning runs for 10 epochs at learning rate 10^{-4} with batch size 16, keeping the checkpoint with the best validation loss.

D.3.2 GENE-PROGRAM CASE STUDY

Datasets Frozen contextualized gene embeddings are extracted by scGANT from Immune Human Luecken et al. (2022). To preserve the original discovery analysis exactly, Leiden clustering uses its fixed 1,918-gene universe: the top 2,000 highly variable genes retained in Immune Human and available to the three encoders used when that universe was constructed. The separate visualization in Figure 4b retains the existing 3,887-gene display universe and therefore contains nine of the twelve P122 members.

Protocols Frozen scGANT gene embeddings are L_2-normalized to construct a 15-nearest-neighbor graph, followed by Leiden clustering (resolution 40, two iterations, seed 0. Minimum cluster size five). P122 is the automatic label for community 122 and has no external biological meaning. External annotations and literature are consulted only after clustering. We select P122 post hoc because it has the largest number of significant external annotation terms (359), while 55 of its 66 pairs (83.3%) lack a direct edge in the four-channel prior (treated as undirected; Appendix B). Seven genes carry the DisGeNET *Inflammation* annotation, while the other five have no mutual direct edges ($0/10$).

E ADDITIONAL EXPERIMENTS

E.1 GATE INITIALIZATION SENSITIVITY

Table 9 shows a task-dependent trade-off: $\alpha = -1$ performs best on annotation and perturbation prediction, whereas the balanced initialization $\alpha = 0$ performs best on held-out integration. The conservative $\alpha = -3$ setting is not optimal for any task, indicating that no single initialization dominates all downstream settings.

Table 9: Gate initialization sensitivity.

α	Gate $\sigma(\alpha)$	Annotation Accuracy	Integration Overall	Perturbation P_{Δ}^{DE}
-3	0.047	0.8404	0.6486	0.7818
-1	0.269	**0.8563**	0.6548	**0.7925**
0	0.500	0.8402	**0.6692**	0.7848

E.2 Preliminary Results

The cell-conditioned design rests on a premise that can be tested without any model: that no single gene interaction graph is correct for every cell. Figure 8 tests it with expression values and cell-type labels alone. Panel (a) shows individual gene pairs that are nearly uncoupled in one cell type and strongly correlated in another. Panel (b) shows that this is systematic across the four annotation and integration datasets: each gene's top-20 correlated partners are shared far less between real cell types (0.11 to 0.27) than between size-matched random splits of the same cells (0.59 to 0.90), and a graph pooled over all cells is no more representative of any single type (0.21 to 0.35). A static prior therefore cannot be correct for all cells at once.

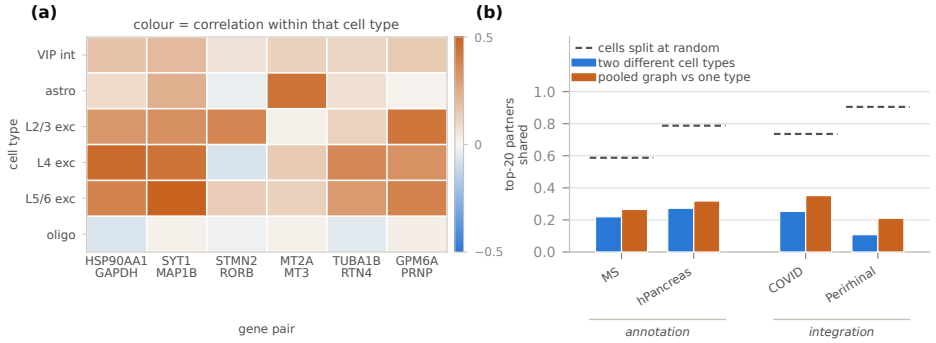


Figure 8: **Gene-gene co-expression structure is not shared across cell states.** (a) Correlation of individual gene pairs within each cell type of Multiple Sclerosis over a shared gene panel. The pairs shown are the most variable across cell types, not typical ones. (b) Agreement between gene-gene correlation structures estimated within individual cell types, measured as the overlap of each gene's top-20 correlated partners.

E.3 Task 1: Cell Type Annotation

E.3.1 Additional Results

Figure 9 visualizes the query-cell embeddings on both benchmarks. scGANT produces embeddings with clearer cell-type organization, and its cluster structure closely matches the annotated reference.

E.4 Task 2: Batch Integration

E.4.1 Additional Results

Figure 10 visualizes the batch structure of the four frozen foundation-model embeddings on the held-out COVID-19 subset.

E.5 Task 3: Perturbation Prediction

E.5.1 Additional Results

Figure 11 provides complementary qualitative evidence with representative response profiles on the top-20 DE genes for four held-out perturbations. For TIMM44, DAD1, and AMIGO3 knockdown, scGANT closely tracks the direction and magnitude of the ground-truth expression changes. We also show a failure case, BHLHE40 knockdown ($r = -0.80$), where the predicted response is largely

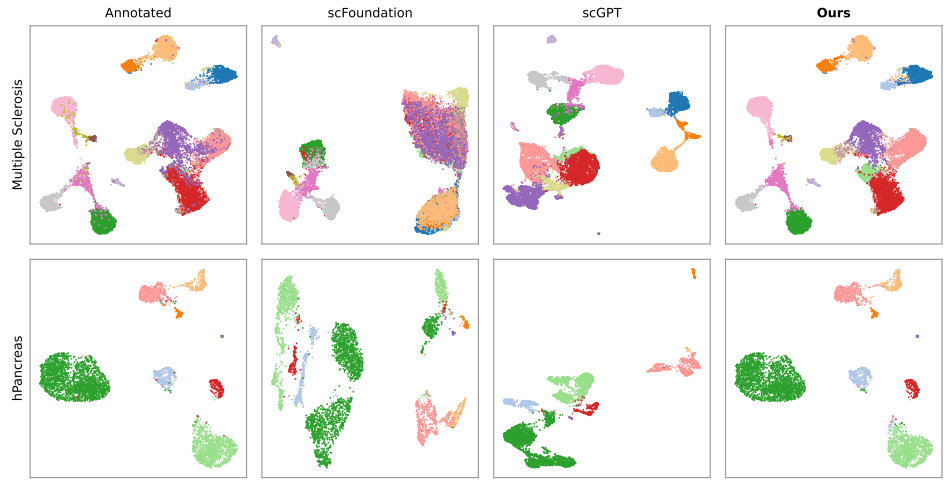


Figure 9: UMAPs of query-cell embeddings on Multiple Sclerosis (top) and hPancreas (bottom) for scFoundation, scGPT, and scGANT, colored by cell type. The first column shows the annotated reference labels.

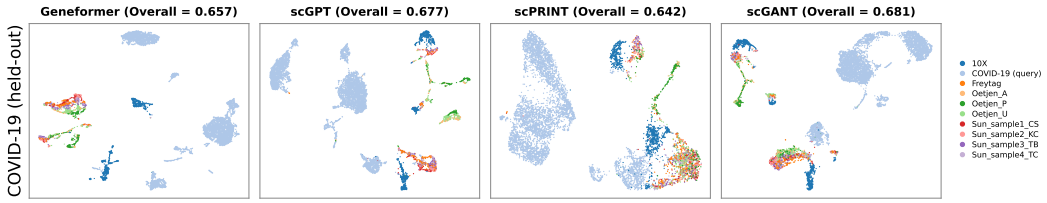


Figure 10: UMAPs of scGANT and baselines on the held-out COVID-19 subset colored by batch.

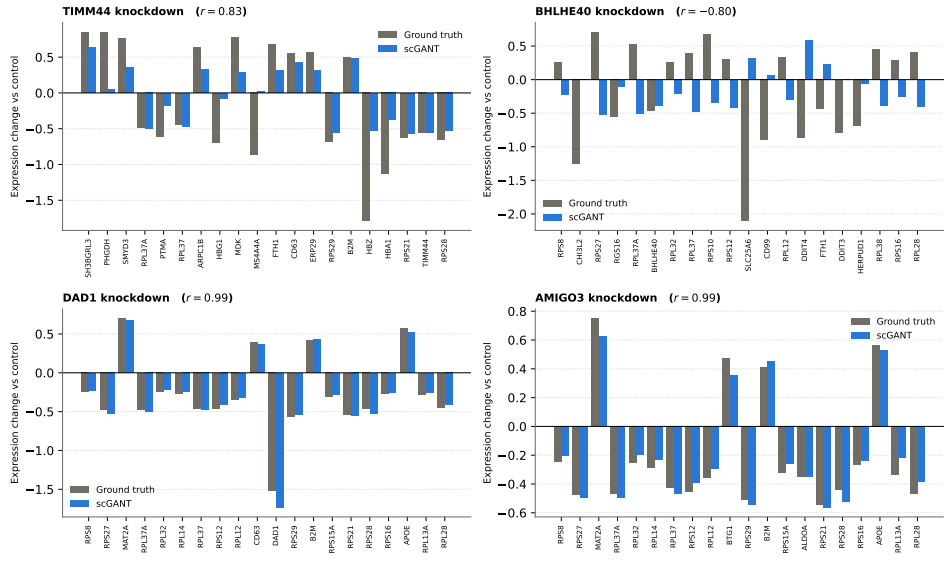


Figure 11: Representative perturbation response profiles comparing predicted (scGANT) and ground-truth expression changes relative to control on the top-20 DE genes for four held-out Adamson perturbations; r is the Pearson correlation between the two profiles.

inverted, illustrating that the remaining errors concentrate on a few perturbations rather than being spread uniformly.