TRUST-VC: MULTI-VIEW CLUSTERING FOR TRANSCRIPTOMICS VIA TRUST-AWARE VIRTUAL-PRIOR RESIDUAL FUSION

Anonymous authors
Paper under double-blind review

ABSTRACT

Multi-view clustering (MVC) of paired multi-omics data must reconcile complementary but unequally reliable views, and the problem becomes harder when external representations from pretrained foundation models are introduced. We present Trust-VC, a reliability-aware MVC framework that treats a foundation-model embedding as a virtual-cell prior rather than an equally trusted observed view. RNA is directly measured in all evaluated datasets. We use a fixed RNA-derived representation as the reference for fusion, and express contributions from the measured auxiliary modality and the virtual prior as gated residual corrections relative to that reference. For each cell, virtual-prior trust combines cross-modal consistency, a learned latent-dispersion proxy, and a self-supervised assignment-confidence gain proxy. A reference-preservation objective penalizes excessive net displacement from the fixed RNA reference, while partial-to-complete distillation aligns the two residual-enhanced partial views with the complete view without updating the RNA reference. Extensive experiments on eight paired single-cell multi-omics datasets demonstrate that Trust-VC not only achieves the best performance in 28 of 32 dataset–metric comparisons, with up to a 10.8 percentage-point ACC improvement, but also regulates transferred corrections in fixed-checkpoint prior-permutation diagnostics.

1 INTRODUCTION

Multimodal single-cell assays jointly profile complementary molecular layers, such as RNA with surface proteins in CITE-seq or RNA with chromatin accessibility in multiome data Stoeckius et al. (2017); Ma et al. (2020). These measurements provide richer views of cellular state, but their statistical properties differ substantially: RNA and ATAC are sparse, ADT is panel-limited, and cross-omics relationships can be weak, nonlinear, or cell-state dependent Hu et al. (2024); Fu et al. (2025); Wang et al. (2025). Consequently, treating heterogeneous modalities as equally reliable, whether through direct concatenation or aggressive alignment, can propagate modality-specific noise and conflicting signals into the shared representation. This motivates specifying both a reference representation for fusion and how complementary sources are allowed to modify it.

Existing integration and clustering methods address multimodal heterogeneity through generative modeling, structural learning, contrastive objectives, and uncertainty-aware optimization Ashuach et al. (2023); Yang et al. (2024); Yun et al. (2024); Dong et al. (2025); Ding et al. (2026); Hu et al. (2025); Li et al. (2026). A common limitation, however, is that reliability is mainly modeled among views observed in the target dataset. This becomes more consequential when an additional representation is transferred from a pretrained model.

Single-cell foundation models such as Geneformer, scGPT, scFoundation, CellPLM, and UCE provide transferable cellular representations learned from large transcriptomic corpora Theodoris et al. (2023); Cui et al. (2024); Hao et al. (2024); Wen et al. (2024); Rosen et al. (2026). Yet large-scale pretraining does not guarantee reliable transfer to a specific target dataset: zero-shot evaluations and recent multi-omics benchmarks report that foundation-model representations do not consistently outperform conventional or task-specific alternatives Kedzierska et al. (2025); Wang et al. (2026). More generally, source–target representation mismatch can induce negative transfer even when pretrained features are informative on average Xu et al. (2026). Thus, the central question is not only

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.

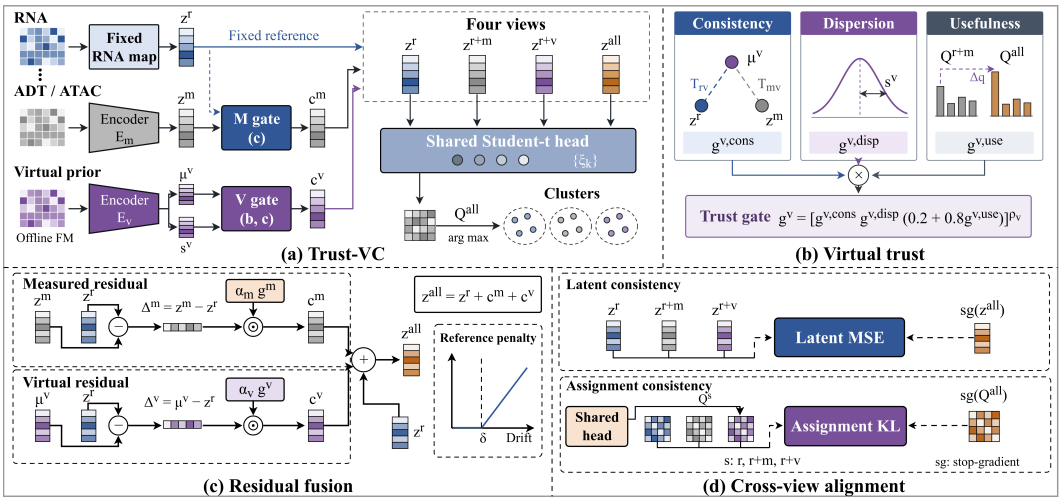

Figure 1: Overview of Trust-VC. A fixed RNA reference is augmented by gated residual corrections from the measured auxiliary modality and a trust-regulated virtual prior. Reference preservation penalizes excessive net displacement, while distillation aligns the two partial fused views with the complete view for clustering.

whether external biological knowledge is available, but when it should be trusted and how strongly it should influence each cell.

We propose Trust-VC, a trust-aware multi-view clustering framework built around a fixed RNA reference. A representation constructed from the observed RNA measurements defines the reference coordinates throughout optimization; the measured auxiliary modality and the foundation-model virtual prior contribute learned, gated residual corrections relative to it. RNA is available in all evaluated datasets, which motivates this common reference choice without assuming that RNA is noise-free or optimal for every cell type. A cell-adaptive gate controls the measured auxiliary correction, and a composite trust mechanism controls the virtual correction using cross-modal agreement, latent dispersion, and assignment-confidence gain. A reference-preservation penalty acts on their combined displacement from RNA. Partial-to-complete distillation transfers complete-view geometry and assignments to the two residual-enhanced partial views while leaving the RNA reference fixed.

Our contributions are threefold:

- We formulate multi-view fusion around a fixed RNA reference, with the measured auxiliary modality and a foundation-model virtual-cell prior providing distinct residual corrections. This separates the experimentally observed reference from complementary measured information and transferred knowledge.
- We regulate virtual corrections using cell-specific trust, penalize the combined residual displacement from the RNA reference, and coordinate the two residual-enhanced partial views through distillation from the complete view.
- Extensive experiments across eight paired datasets, five foundation models, and ten baselines demonstrate consistent gains, while ablations assess component contributions and prior-permutation diagnostics examine adaptive regulation of transferred corrections.

A more detailed discussion of related work is provided in Appendix A.

2 Method

2.1 Problem Formulation and Model Overview

Given N paired cells indexed by $i \in \{1, \ldots, N\}$, we consider three information sources: an RNA view $\mathbf{x}_i^r$, an auxiliary experimentally observed modality $\mathbf{x}_i^m$, and a precomputed representation from

a foundation model $\mathbf{x}_i^v$. Trust-VC aims to learn a unified representation $\mathbf{z}_i^{\text{all}} \in \mathbb{R}^d$ for unsupervised clustering into K groups.

The three sources have distinct roles in Trust-VC. RNA defines a fixed reference representation, while the auxiliary modality and virtual prior provide cell-specific residual corrections relative to that reference:

$$\mathbf{z}_i^{\text{all}} = \mathbf{z}_i^r + \mathbf{c}_i^m + \mathbf{c}_i^v, \qquad \mathbf{z}_i^r = B_r(\mathbf{x}_i^r). \tag{1}$$

where $\mathbf{c}_i^m$ and $\mathbf{c}_i^v$ denote the residual contributions from the auxiliary measured modality and the virtual prior, respectively. $B_r(\cdot)$ denotes the RNA preprocessing map, comprising library-size normalization, log transformation, highly variable gene selection, standardization, and PCA as specified in Appendix B. It is constructed before clustering optimization and held fixed, so $\mathbf{z}_i^r$ is not updated by gradient descent. The learned corrections determine the displacement of the fused representation from this reference. Choosing RNA as the reference is an architectural choice, not an assumption that its measurements are noise-free or that its neighborhoods are necessarily correct.

2.2 RNA-Referenced Adaptive Residual Fusion

The experimentally observed auxiliary modality supplies complementary information whose contribution is learned relative to the fixed RNA reference. We encode it as $\mathbf{z}_i^m = E_m(\mathbf{x}_i^m)$ and define its residual in the reference coordinates as:

$$\mathbf{\Delta}_i^m = \mathbf{z}_i^m - \mathbf{z}_i^r. \tag{2}$$

To modulate this residual on a per-cell basis, we estimate its reliability from both latent-space agreement and learnable cross-view interaction. For numerical stability, we use the normalized representation

$$\bar{\mathbf{z}} = \frac{\mathbf{z}}{\max(\|\mathbf{z}\|_2, \epsilon)}, \tag{3}$$

where $\epsilon > 0$ is a small constant. The agreement-based score is defined as

$$g_{i,\text{dist}}^m = \exp\left(-\frac{\|\bar{\mathbf{z}}_i^r - \bar{\mathbf{z}}_i^m\|_2^2}{\tau_m}\right), \tag{4}$$

where $\tau_m > 0$ controls the sensitivity to cross-view discrepancy. We then combine this explicit agreement score with the learnable interaction operator introduced in Eq. (11):

$$g_i^m = \lambda_b g_{i,\text{dist}}^m + (1 - \lambda_b) G_{\theta_m}(\mathbf{z}_i^r, \mathbf{z}_i^m), \tag{5}$$

where $\lambda_b \in [0, 1]$ balances geometric agreement and learned cross-view compatibility. The resulting residual correction is

$$\mathbf{c}_i^m = \alpha_m g_i^m \mathbf{\Delta}_i^m, \qquad \mathbf{z}_i^{r+m} = \mathbf{z}_i^r + \mathbf{c}_i^m, \tag{6}$$

with $\alpha_m \geq 0$ controlling the overall contribution of the auxiliary modality. Throughout this work, g_i^m is implemented as a scalar gate shared across latent dimensions.

Residual fusion refines the fused representation while leaving the RNA reference itself fixed. Unconstrained corrections may nevertheless produce excessive displacement from that reference. We therefore regularize the final fused representation using a margin-based reference-preservation objective:

$$\mathcal{L}_{\text{RP}} = \frac{1}{N} \sum_{i=1}^{N} \left[\|\mathbf{z}_i^{\text{all}} - \mathbf{z}_i^r\|_2 - \delta\right]_+, \tag{7}$$

where $[x]_+ = \max(0, x)$ and $\delta \geq 0$ specifies the tolerated displacement from RNA. Since

$$\mathbf{z}_i^{\text{all}} - \mathbf{z}_i^r = \mathbf{c}_i^m + \mathbf{c}_i^v, \tag{8}$$

$\mathcal{L}_{\text{RP}}$ penalizes the aggregate residual displacement rather than each branch independently. It encourages the fused representation to remain near the fixed RNA reference while allowing complementary corrections within the margin. This is a soft penalty: it neither guarantees a hard distance bound nor directly preserves every RNA neighborhood, and opposing residuals can partially cancel.

2.3 TRUST-AWARE VIRTUAL-PRIOR FUSION

Unlike the experimentally observed auxiliary modality, the virtual-cell representation is transferred from an external pretraining distribution and may not be equally reliable for all target cells. We therefore treat it as a selective prior rather than an additional observed view. Specifically, its residual contribution is modulated by three complementary signals capturing (i) cross-view consistency, (ii) latent dispersion, and (iii) clustering-oriented utility.

Cross-View Consistency. We first assess whether the virtual representation is compatible with the experimentally observed views. Given two latent representations $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$, we combine explicit geometric agreement with a learnable interaction score:

$$g_{\mathrm{dist}}(\mathbf{a}, \mathbf{b}) = \exp\left(-\frac{\|\bar{\mathbf{a}} - \bar{\mathbf{b}}\|_2^2}{\tau_m}\right), \tag{9}$$

where $\bar{\mathbf{a}}$ and $\bar{\mathbf{b}}$ denote their normalized representations. The learnable component is defined as

$$G_\theta(\mathbf{a}, \mathbf{b}) = \mathrm{sigmoid}(\mathbf{W}_2\,\mathrm{ReLU}(\mathbf{W}_1[\mathbf{a}; \mathbf{b}; |\mathbf{a} - \mathbf{b}|; \mathbf{a} \odot \mathbf{b}] + \mathbf{b}_1) + \mathbf{b}_2), \tag{10}$$

yielding the pairwise trust operator

$$\mathcal{T}_\theta(\mathbf{a}, \mathbf{b}) = \lambda_b g_{\mathrm{dist}}(\mathbf{a}, \mathbf{b}) + (1 - \lambda_b) G_\theta(\mathbf{a}, \mathbf{b}). \tag{11}$$

Using a dedicated parameter set θ_v, the virtual consistency score is computed against both observed views:

$$g_i^{v,\mathrm{cons}} = \frac{1}{2}\left[\mathcal{T}_{\theta_v}(\mathbf{z}_i^r, \boldsymbol{\mu}_i^v) + \mathcal{T}_{\theta_v}(\mathbf{z}_i^m, \boldsymbol{\mu}_i^v)\right]. \tag{12}$$

The measured-modality gate in Section 2.2 uses an independent parameter set θ_m.

Latent-Dispersion Proxy. Cross-view agreement alone does not capture the reliability of the transferred representation itself. We therefore parameterize the virtual encoder as a diagonal Gaussian,

$$q_\phi(\mathbf{z}_i^v \mid \mathbf{x}_i^v) = \mathcal{N}\big(\boldsymbol{\mu}_i^v, \mathrm{diag}\big[(\mathbf{s}_i^v)^2\big]\big), \tag{13}$$

where

$$\mathbf{s}_i^v = \mathrm{softplus}(\boldsymbol{\eta}_i^v - b_s) + \epsilon_s. \tag{14}$$

We use the posterior mean $\boldsymbol{\mu}_i^v$ for subsequent fusion and define the virtual residual relative to the same fixed RNA reference as

$$\boldsymbol{\Delta}_i^v = \boldsymbol{\mu}_i^v - \mathbf{z}_i^r. \tag{15}$$

The average posterior scale provides a compact cell-wise dispersion statistic,

$$d_i^v = \frac{1}{d}\sum_{j=1}^{d} s_{ij}^v, \qquad g_i^{v,\mathrm{disp}} = \frac{1}{1 + \tau_d d_i^v}, \tag{16}$$

such that diffuse latent representations receive smaller virtual corrections. Because no explicit uncertainty-calibration objective is imposed, d_i^v is interpreted only as a learned dispersion proxy rather than calibrated predictive uncertainty. The Gaussian latent space is regularized by a standard KL term, detailed in Appendix B.

Assignment-Confidence Gain. A virtual representation can be consistent and concentrated while still providing little benefit to the current clustering state. We therefore introduce a task-aware usefulness predictor. For each cell, we construct

$$\boldsymbol{\chi}_i = [\mathbf{z}_i^r; \boldsymbol{\Delta}_i^m; \boldsymbol{\Delta}_i^v; |\boldsymbol{\Delta}_i^v|; |\boldsymbol{\Delta}_i^v - \boldsymbol{\Delta}_i^m|] \in \mathbb{R}^{5d}, \tag{17}$$

and estimate

$$\widetilde{g}_i^{v,\mathrm{use}} = G_{\mathrm{use}}(\boldsymbol{\chi}_i). \tag{18}$$

Its self-supervised target is derived from the change in maximum assignment confidence induced by virtual-prior injection:

$$\widetilde{h}_i^v = \mathrm{sigmoid}\left(\frac{\mathrm{sg}\big(\max_k q_{ik}^{\mathrm{all}} - \max_k q_{ik}^{r+m}\big)}{\tau_{\mathrm{use}}}\right), \tag{19}$$

where $\mathrm{sg}(\cdot)$ denotes stop-gradient and q_{ik}^s is defined in Eq. (28). To maintain a nonzero virtual pathway, both quantities are lower bounded by ϵ_u:

$$g_i^{v,\mathrm{use}} = \epsilon_u + (1-\epsilon_u)\widetilde{g}_i^{v,\mathrm{use}}, \qquad \widehat{h}_i^v = \epsilon_u + (1-\epsilon_u)\widetilde{h}_i^v. \tag{20}$$

The usefulness predictor is optimized by

$$\mathcal{L}_{v,\mathrm{use}} = \frac{1}{N}\sum_{i=1}^{N}\left(g_i^{v,\mathrm{use}} - \widehat{h}_i^v\right)^2. \tag{21}$$

Importantly, $\widehat{h}_i^v$ measures confidence improvement rather than assignment correctness.

Composite Virtual Trust. The three signals capture complementary aspects of virtual-prior reliability: consistency measures cross-view compatibility, dispersion characterizes latent concentration, and usefulness reflects task-specific contribution. We combine them multiplicatively so that strong virtual injection requires joint support from all three criteria:

$$g_i^v = \left[g_i^{v,\mathrm{cons}} g_i^{v,\mathrm{disp}} \left(0.2 + 0.8 g_i^{v,\mathrm{use}}\right)\right]^{\rho_v}, \tag{22}$$

where $\rho_v \geq 1$ controls the sharpness of trust modulation. The affine transformation of $g_i^{v,\mathrm{use}}$ prevents this task-dependent signal from independently shutting down the virtual branch.

The resulting virtual correction is

$$\mathbf{c}_i^v = \alpha_v g_i^v \boldsymbol{\Delta}_i^v, \tag{23}$$

and the complete fused representation becomes

$$\mathbf{z}_i^{\mathrm{all}} = \mathbf{z}_i^r + \mathbf{c}_i^m + \mathbf{c}_i^v, \tag{24}$$

where α_v controls the global strength of the transferred prior.

Virtual-Prior Regularization. To avoid systematic reliance on the transferred prior, we penalize excessive gate activation:

$$\mathcal{L}_{v,\mathrm{sparse}} = \frac{1}{N}\sum_{i=1}^{N} g_i^v. \tag{25}$$

The virtual-prior objective is

$$\mathcal{L}_{\mathrm{VP}} = \lambda_{v,\mathrm{sparse}}\mathcal{L}_{v,\mathrm{sparse}} + \lambda_{v,\mathrm{use}}\mathcal{L}_{v,\mathrm{use}} + \mathcal{R}_v, \tag{26}$$

where $\mathcal{R}_v$ contains the auxiliary regularizers associated with the Gaussian latent space and virtual residual. Their explicit definitions and implementation settings are deferred to Appendix B.

2.4 Partial-to-Complete Cross-View Distillation

The gated residual design naturally induces two partial fused representations and a complete representation:

$$\mathbf{z}_i^{r+m} = \mathbf{z}_i^r + \mathbf{c}_i^m, \qquad \mathbf{z}_i^{r+v} = \mathbf{z}_i^r + \mathbf{c}_i^v, \qquad \mathbf{z}_i^{\mathrm{all}} = \mathbf{z}_i^r + \mathbf{c}_i^m + \mathbf{c}_i^v. \tag{27}$$

All three fused representations retain the same fixed RNA reference and differ in which residual corrections are included. They share a clustering head with the RNA-only representation. Distillation, however, applies only to $r+m$ and $r+v$; computing assignments for the RNA reference does not make it a trainable residual branch or an active distillation student.

Complete-View Clustering. For any representation $\mathbf{z}_i^s$, a shared Student-t assignment head with trainable cluster centers $\{\boldsymbol{\xi}_k\}_{k=1}^K$ produces

$$q_{ik}^s = \frac{\left(1 + \|\mathbf{z}_i^s - \boldsymbol{\xi}_k\|_2^2/\nu\right)^{-(\nu+1)/2}}{\sum_{k'}\left(1 + \|\mathbf{z}_i^s - \boldsymbol{\xi}_{k'}\|_2^2/\nu\right)^{-(\nu+1)/2}}, \tag{28}$$

where $\nu > 0$ denotes the degree-of-freedom parameter.

Following the self-training principle of DEC Xie et al. (2016), we construct a sharpened target distribution from the complete view:

$$p_{ik}^{\mathrm{all}} = \frac{(q_{ik}^{\mathrm{all}})^2 / f_k}{\sum_{k'} (q_{ik'}^{\mathrm{all}})^2 / f_{k'}}, \qquad f_k = \sum_i q_{ik}^{\mathrm{all}}. \tag{29}$$

The complete representation is optimized by

$$\mathcal{L}_{\mathrm{clu}} = \mathrm{KL}\big(\mathrm{sg}(\mathbf{P}^{\mathrm{all}}) \| \mathbf{Q}^{\mathrm{all}}\big), \tag{30}$$

where $\mathrm{sg}(\cdot)$ denotes stop-gradient. In practice, the target distribution is estimated from the current mini-batch for scalable joint optimization of the residual encoders, gates, and clustering head. The RNA reference remains fixed, while the cluster centers are trainable after K-means initialization. Numerical stabilization details are provided in Appendix B.

Partial-to-Complete Distillation. Optimizing only the complete view does not ensure that representations formed from individual residual sources preserve the same latent structure and clustering semantics. We therefore distill the complete representation into the two partial fused views at both the representation and assignment levels:

$$\mathcal{L}_{\mathrm{lat}} = \sum_{s \in \{r+m, r+v\}} \lambda_{\mathrm{lat}}^s \frac{1}{N} \sum_{i=1}^{N} \left\| \mathbf{z}_i^s - \mathrm{sg}(\mathbf{z}_i^{\mathrm{all}}) \right\|_2^2, \tag{31}$$

and

$$\mathcal{L}_q = \sum_{s \in \{r+m, r+v\}} \lambda_q^s \mathrm{KL}\big(\mathrm{sg}(\mathbf{Q}^{\mathrm{all}}) \| \mathbf{Q}^s\big). \tag{32}$$

The complete view is detached in both objectives and therefore acts as a one-way teacher rather than participating in symmetric cross-view contraction. $\mathcal{L}_{\mathrm{lat}}$ transfers complete-view geometry, whereas $\mathcal{L}_q$ transfers its clustering semantics. In our implementation, distillation is applied only to the two residual-enhanced partial views; the RNA-only consistency weights are set to zero. The RNA-only assignments may still change as the shared cluster centers are updated, even though $\mathbf{z}_i^r$ itself remains fixed.

Notably, these objectives encourage each partial view to approximate the information encoded by the complete representation rather than explicitly shrinking individual residual corrections. Reference drift and excessive virtual-prior contribution are controlled separately by the corresponding regularization terms introduced above.

Multi-View Clustering Objective. The resulting clustering objective is

$$\mathcal{L}_{\mathrm{MVC}} = \mathcal{L}_{\mathrm{clu}} + \mathcal{L}_{\mathrm{lat}} + \mathcal{L}_q. \tag{33}$$

This formulation couples complete-view self-training with partial-view distillation, encouraging the complementary residual pathways to remain individually informative while sharing a consistent clustering structure.

2.5 Overall Objective

The complete training objective is

$$\boxed{\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{MVC}} + \sigma\, \mathcal{L}_{\mathrm{RP}} + \beta\, \mathcal{L}_{v,\mathrm{sparse}} + \gamma\, \mathcal{L}_{v,\mathrm{use}} + \mathcal{R}_v} \tag{34}$$

where σ, β, and γ control RNA-reference preservation, virtual-prior sparsity, and usefulness supervision, respectively. These three coefficients are varied in the sensitivity analysis, while the remaining hyperparameters are fixed as specified in Appendix B.

The cluster centers are initialized with K-means and subsequently optimized jointly with the model. At inference, the final cluster label of each cell is determined by the maximum-probability assignment of the complete fused view.

3 EXPERIMENTS

3.1 EXPERIMENTAL SETUP

We evaluate eight paired single-cell multi-omics datasets against ten baselines using ACC, NMI, ARI, and F1 over repeated runs. The transductive protocol sets K to the annotated cell-type count. Validation labels are used only to select the virtual-prior source; apart from this selection and K, labels are excluded from representation learning, clustering, early stopping, and checkpoint selection. Held-out test labels are reserved for final evaluation. Dataset statistics, baselines, and protocol details appear in Appendix C.

3.2 VIRTUAL-PRIOR SOURCE SELECTION

We compare scGPT Cui et al. (2024), Geneformer Theodoris et al. (2023), scFoundation Hao et al. (2024), UCE Rosen et al. (2026), and CellPLM Wen et al. (2024) under the same fusion and clustering protocol, selecting the source separately for each dataset using validation performance. Figure 2 shows that UCE leads in 15 of 24 dataset–metric comparisons and scFoundation in the remaining nine, including all metrics on BMNC-3K and BMNC-5K. Neither model is uniformly optimal, motivating dataset-specific selection and reliability-aware use of the transferred prior.

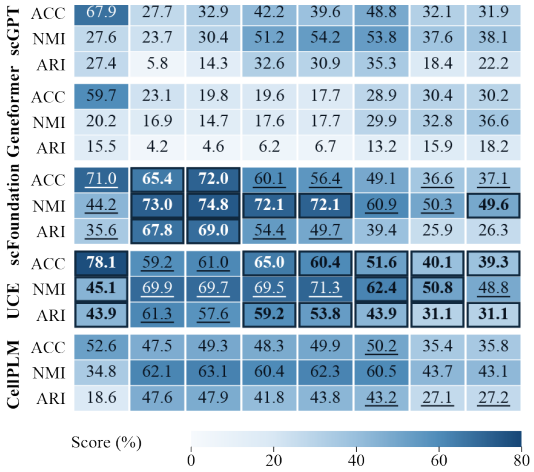

Figure 2: Foundation-model comparison across eight datasets using ACC, NMI, and ARI (%).

3.3 OVERALL MULTI-VIEW CLUSTERING PERFORMANCE

Trust-VC ranks first by mean score in 28 of 32 comparisons and second in the remaining four (Table 1). It achieves the best ACC, NMI, and F1 on all eight datasets and a top-two ARI throughout. The largest ACC gain is 10.8 percentage points on GSE128639 (69.9% versus 59.1%); it also leads in 11 of 12 comparisons across the BMNC series. Prior-mismatch diagnostics are evaluated separately below.

3.4 ABLATION STUDY

We remove cross-view consistency ($\mathcal{L}_{\text{lat}}$ and $\mathcal{L}_q$), RNA-reference preservation ($\sigma\mathcal{L}_{\text{RP}}$), trust-aware virtual learning, or the virtual prior. Removing reference preservation retains the fixed RNA reference and both residual pathways. The full model achieves the highest mean ACC, NMI, and ARI in all 24 dataset–metric comparisons (Fig. 3; Appendix D), supporting these components. These ablations do not isolate each constituent trust signal; the trained trust-aware ablation also differs from the inference-time GateOff diagnostic below.

3.5 ADAPTIVE VIRTUAL CORRECTIONS UNDER PRIOR MISMATCH

We permute cell–prior correspondences while fixing trained parameters, the RNA reference, and measured auxiliary inputs. GateOff sets the effective virtual gate to one in the same model. On GSE128639, from 0% to 100% permutation, mean virtual correction norms change from 3.556 to 3.581 under Trust-VC versus 3.888 to 4.902 under GateOff; net RNA displacement changes from 7.502 to 6.986 versus 7.767 to 7.846 (Fig. 4). Absolute ARI reaches 0.569 versus 0.553, exceeding both clean-prior values (0.523 and 0.524). Thus, the diagnostic supports adaptive correction under mismatch, without assuming permutation always harms clustering. Appendix E details the separate diagnostic checkpoints, PBMC5K results, and paired analyses.

Table 1: Clustering performance comparison on eight datasets (%, mean ± std). The best and second-best results are highlighted with light red and light blue backgrounds, respectively.

Dataset	Metric	scMVAE	scMoE	scMHNN	ACCMVC	DCG	STCMC-UR	AEMVC	scMCs	MOH	scMUSCLE	Trust-VC
ASSP	ACC	69.7 ± 0.2	59.0 ± 4.5	50.8 ± 1.2	65.1 ± 2.8	66.9 ± 3.4	66.6 ± 1.2	32.8 ± 3.3	51.1 ± 3.9	40.2 ± 1.8	40.5 ± 11.5	**69.8 ± 0.5**
	NMI	66.1 ± 0.6	54.9 ± 3.1	54.7 ± 1.1	65.3 ± 0.7	63.5 ± 1.5	64.2 ± 0.4	23.6 ± 5.2	58.4 ± 2.0	39.1 ± 1.1	35.6 ± 12.0	**66.7 ± 0.4**
	ARI	**52.7 ± 0.1**	41.9 ± 7.4	31.5 ± 2.6	47.0 ± 2.6	48.6 ± 1.5	50.1 ± 1.1	12.7 ± 4.5	39.0 ± 3.6	26.6 ± 0.2	17.9 ± 10.2	51.1 ± 1.3
	F1	65.7 ± 1.8	49.5 ± 2.6	49.2 ± 1.8	61.5 ± 2.5	60.1 ± 3.2	58.5 ± 0.6	32.3 ± 3.9	46.3 ± 3.4	35.4 ± 0.1	33.8 ± 10.9	**66.9 ± 2.1**
BMNC-3K	ACC	48.9 ± 2.8	48.3 ± 7.1	37.1 ± 1.1	52.2 ± 0.6	56.5 ± 2.1	53.2 ± 3.0	45.0 ± 1.8	40.2 ± 2.2	17.5 ± 7.3	41.0 ± 3.5	**62.9 ± 0.6**
	NMI	73.2 ± 1.1	61.4 ± 4.0	64.0 ± 1.3	67.4 ± 1.7	71.9 ± 0.3	70.4 ± 4.3	72.2 ± 0.1	57.8 ± 0.5	29.0 ± 18.0	60.3 ± 0.5	**74.8 ± 1.4**
	ARI	44.9 ± 3.1	37.8 ± 8.4	32.9 ± 1.7	50.3 ± 1.7	54.3 ± 4.4	52.9 ± 0.6	41.9 ± 1.1	34.4 ± 5.2	11.3 ± 8.5	42.0 ± 7.7	**55.7 ± 1.8**
	F1	43.5 ± 4.0	32.6 ± 4.4	28.3 ± 0.6	37.0 ± 1.5	33.0 ± 5.9	39.6 ± 2.2	41.3 ± 4.0	31.7 ± 1.3	16.6 ± 8.2	28.0 ± 3.4	**50.1 ± 1.1**
BMNC-5K	ACC	51.3 ± 2.7	43.7 ± 4.9	40.4 ± 1.2	48.4 ± 4.4	59.1 ± 2.8	56.0 ± 4.7	46.2 ± 0.4	44.4 ± 1.0	17.5 ± 1.6	36.8 ± 1.7	**66.3 ± 2.6**
	NMI	74.3 ± 0.5	62.8 ± 1.5	65.1 ± 1.3	71.7 ± 0.8	73.3 ± 1.1	72.6 ± 4.4	70.1 ± 0.7	60.3 ± 2.1	34.2 ± 6.3	51.2 ± 3.0	**76.3 ± 1.6**
	ARI	48.5 ± 2.8	36.5 ± 5.2	35.8 ± 1.5	50.9 ± 4.7	60.1 ± 0.9	55.7 ± 0.7	41.0 ± 1.4	38.6 ± 0.9	12.3 ± 3.4	23.7 ± 1.3	**60.2 ± 2.1**
	F1	46.0 ± 1.8	29.3 ± 1.6	35.0 ± 1.6	36.5 ± 3.3	33.6 ± 2.0	41.2 ± 6.5	42.4 ± 1.8	37.9 ± 1.4	17.4 ± 3.2	23.0 ± 1.7	**52.8 ± 1.5**
BMNC-10K	ACC	51.9 ± 1.4	54.8 ± 3.3	43.6 ± 2.0	50.0 ± 1.4	61.9 ± 2.4	58.2 ± 1.8	42.8 ± 1.2	43.5 ± 4.1	18.7 ± 2.2	40.1 ± 3.0	**65.8 ± 1.3**
	NMI	74.8 ± 0.2	65.5 ± 0.6	65.8 ± 1.0	74.5 ± 0.6	74.5 ± 2.8	76.4 ± 1.5	67.9 ± 1.2	59.5 ± 1.5	36.9 ± 3.7	51.6 ± 2.9	**76.8 ± 0.8**
	ARI	46.5 ± 0.9	49.1 ± 2.1	41.3 ± 2.1	44.0 ± 0.6	**62.7 ± 5.3**	54.8 ± 1.9	38.7 ± 1.0	34.9 ± 4.0	13.5 ± 3.7	26.9 ± 4.7	62.3 ± 2.8
	F1	48.1 ± 1.2	35.1 ± 0.6	38.9 ± 1.5	41.6 ± 1.1	31.2 ± 6.2	47.5 ± 3.4	39.1 ± 1.7	36.6 ± 3.3	18.7 ± 3.5	23.4 ± 4.3	**48.6 ± 1.6**
GSE128639	ACC	51.5 ± 0.6	51.5 ± 2.1	41.3 ± 2.4	50.1 ± 1.2	59.1 ± 2.6	56.8 ± 1.8	35.7 ± 2.6	15.0 ± 1.2	22.3 ± 14.3	35.9 ± 7.8	**69.9 ± 3.2**
	NMI	74.6 ± 0.8	66.3 ± 0.4	66.9 ± 0.0	74.7 ± 0.5	75.0 ± 0.8	75.9 ± 0.7	53.3 ± 2.7	25.7 ± 2.4	31.0 ± 25.9	42.8 ± 7.8	**79.5 ± 2.1**
	ARI	46.5 ± 0.9	48.7 ± 5.0	41.6 ± 2.8	44.2 ± 0.5	61.1 ± 3.2	52.1 ± 2.7	27.3 ± 2.2	9.1 ± 1.4	14.6 ± 12.9	20.2 ± 12.4	**65.8 ± 4.1**
	F1	46.6 ± 0.2	34.4 ± 2.5	34.6 ± 3.1	43.0 ± 0.5	34.0 ± 5.7	45.9 ± 2.5	31.6 ± 3.6	11.2 ± 0.8	20.0 ± 12.4	18.5 ± 3.8	**51.5 ± 1.6**
PBMC5K	ACC	51.8 ± 2.1	41.5 ± 6.0	40.4 ± 3.1	49.2 ± 2.0	56.7 ± 3.0	54.0 ± 4.7	39.8 ± 3.0	49.6 ± 0.9	44.0 ± 2.8	39.0 ± 5.5	**60.7 ± 0.0**
	NMI	61.5 ± 0.6	42.7 ± 3.8	51.9 ± 1.6	59.0 ± 0.6	60.5 ± 1.2	59.9 ± 1.2	51.2 ± 0.1	57.6 ± 0.4	45.2 ± 2.0	43.8 ± 4.4	**63.3 ± 0.1**
	ARI	38.3 ± 0.7	28.5 ± 7.5	27.8 ± 3.9	37.2 ± 2.7	**47.9 ± 6.6**	40.7 ± 5.0	24.8 ± 0.9	37.9 ± 0.5	31.6 ± 5.5	23.5 ± 5.6	46.9 ± 0.0
	F1	42.6 ± 2.7	25.4 ± 2.2	34.6 ± 3.4	34.9 ± 0.3	34.6 ± 1.2	39.5 ± 2.2	38.2 ± 2.5	37.1 ± 0.5	40.9 ± 5.7	24.4 ± 3.6	**49.5 ± 0.2**
SLN111	ACC	31.3 ± 1.4	32.8 ± 2.0	28.8 ± 0.6	28.4 ± 1.5	37.4 ± 2.0	30.3 ± 1.5	14.8 ± 0.6	23.3 ± 0.7	12.5 ± 0.9	26.3 ± 3.5	**41.0 ± 0.4**
	NMI	55.3 ± 1.1	47.1 ± 1.3	49.5 ± 0.1	51.6 ± 1.2	51.3 ± 1.6	52.7 ± 1.0	18.6 ± 1.8	41.5 ± 0.2	15.1 ± 2.9	34.1 ± 4.6	**56.0 ± 0.6**
	ARI	21.3 ± 1.0	24.0 ± 2.7	22.0 ± 0.3	18.7 ± 0.7	**30.2 ± 1.0**	22.6 ± 1.1	6.3 ± 0.5	14.2 ± 0.1	5.0 ± 0.8	15.3 ± 4.0	30.1 ± 0.6
	F1	33.7 ± 1.9	19.4 ± 1.5	24.5 ± 2.0	21.0 ± 0.4	18.2 ± 3.1	23.5 ± 2.0	9.0 ± 0.3	21.6 ± 0.0	10.9 ± 0.6	18.5 ± 1.5	**34.5 ± 2.0**
SLN206	ACC	30.3 ± 2.2	32.3 ± 0.6	27.5 ± 1.3	28.3 ± 0.6	37.0 ± 0.9	30.4 ± 2.3	14.2 ± 0.9	21.5 ± 0.4	15.2 ± 5.5	29.3 ± 5.4	**39.7 ± 0.2**
	NMI	53.8 ± 0.7	46.8 ± 0.8	48.2 ± 0.8	51.6 ± 0.3	48.5 ± 4.2	51.5 ± 2.0	17.4 ± 1.2	39.8 ± 0.4	23.9 ± 14.8	35.4 ± 4.8	**55.7 ± 0.7**
	ARI	20.4 ± 1.0	24.4 ± 1.2	20.6 ± 1.2	20.6 ± 0.3	29.8 ± 0.7	23.0 ± 1.9	5.3 ± 0.4	13.3 ± 0.4	7.1 ± 5.3	15.2 ± 7.4	**30.3 ± 0.9**
	F1	32.4 ± 1.0	19.5 ± 0.8	23.2 ± 1.5	19.5 ± 0.7	19.7 ± 3.4	22.2 ± 3.3	8.1 ± 0.7	19.2 ± 0.7	13.4 ± 5.2	21.7 ± 2.9	**33.0 ± 1.9**

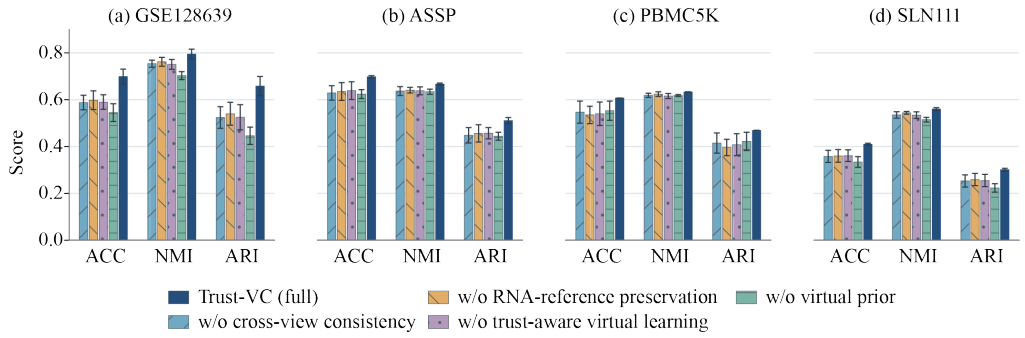

Figure 3: Ablations on (a) GSE128639, (b) ASSP, (c) PBMC5K, and (d) SLN111. Bars and error bars denote means and standard deviations. ACC, NMI, and ARI use their original scale (higher is better).

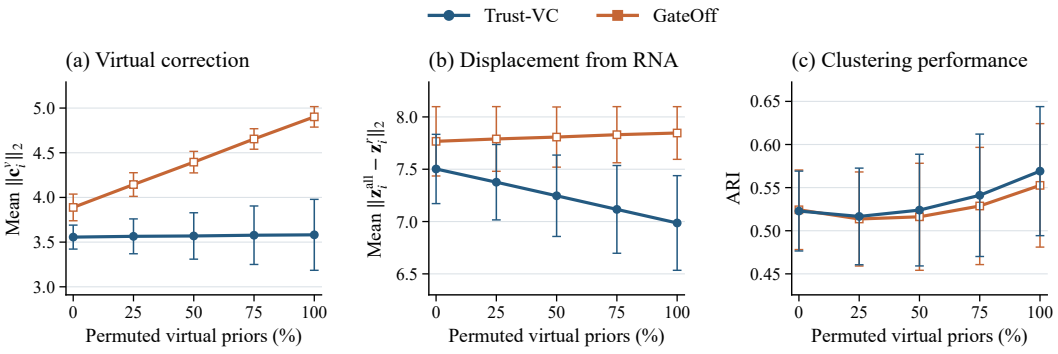

Figure 4: Adaptive virtual corrections around a fixed RNA reference on GSE128639. (a) Mean virtual correction norm. (b) Mean net displacement from RNA. (c) Absolute ARI. GateOff sets the effective virtual gate to one in the same trained model. Points and error bars show means and standard deviations across five training seeds, using the same permutation at each level.

3.6 Parameter Sensitivity Analysis

Figure 5 shows sensitivity on ASSP to reference preservation (σ), virtual-gate sparsity (β), and confidence-proxy fitting (γ). Each column fixes γ while varying (σ, β); all other coefficients remain fixed within each sweep. Within the tested grids, σ generally produces more visible score changes, while β and γ have smoother effects. Results for the other seven datasets appear in Appendix F; Appendix G provides representation visualizations and biological characterization.

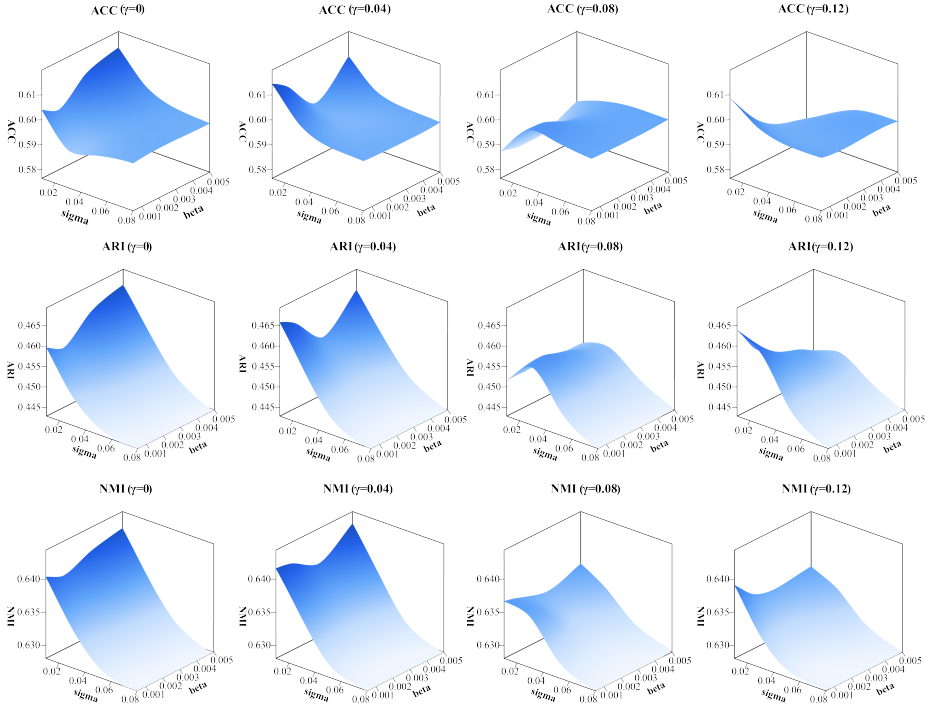

Figure 5: Hyperparameter sensitivity on ASSP. Rows: ACC, ARI, and NMI; columns: fixed γ. Each surface varies (σ, β), with other coefficients fixed within the sweep. Scores use their original scale.

4 Conclusion

Trust-VC incorporates transferred priors through reliability-aware residual corrections around a fixed RNA reference. Across eight paired datasets, it leads in 28 of 32 mean-score comparisons. Ablations support the main components, and fixed-checkpoint diagnostics show regulated virtual corrections and smaller RNA displacement than GateOff, with dataset-dependent clustering effects. These results support selective transfer while leaving the individual trust signals and broader perturbation regimes for further study.

Reproducibility Statement

Dataset statistics, baseline descriptions, optimization details, full ablation results, extended virtual-prior mismatch analyses, cross-dataset sensitivity results, and additional qualitative analyses are provided in the appendix. Code and preprocessing scripts will be provided through an anonymous repository or supplementary archive in the double-blind submission package.

AI Use Statement

During preparation of this manuscript, the authors used ChatGPT for language editing, organization, and LaTeX assistance. The authors reviewed and verified the scientific content and take full responsibility for the submitted manuscript.

REFERENCES

Ricard Argelaguet, Damien Arnol, Danila Bredikhin, Yonatan Deloro, Britta Velten, John C Marioni, and Oliver Stegle. Mofa+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome biology*, 21(1):111, 2020.

Tal Ashuach, Mariano I Gabitto, Rohan V Koodli, Giuseppe-Antonio Saldi, Michael I Jordan, and Nir Yosef. Multivi: deep generative model for the integration of multimodal data. *Nature methods*, 20(8):1222–1231, 2023.

Zhi-Jie Cao and Ge Gao. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nature biotechnology*, 40(10):1458–1466, 2022.

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.

Yun Ding, Yangzhen Jiang, Jing Wang, Dayu Tan, Yansen Su, and Chunhou Zheng. Clustering single-cell multi-omics data via multi-subspace contrastive learning with structural smoothness. *Briefings in Bioinformatics*, 27(1), 2026.

Yao Dong, Chen Chen, Jin Shi, Yushan Hu, Xiaowen Cao, Yongfeng Dong, and Xuekui Zhang. Moh: a novel multilayer multi-omics heterogeneous graph for single-cell clustering. *IEEE Journal of Biomedical and Health Informatics*, 30(1):760–772, 2025.

Shaliu Fu, Shuguang Wang, Duanmiao Si, Gaoyang Li, Yawei Gao, and Qi Liu. Benchmarking single-cell multi-modal data integrations. *Nature Methods*, 22(11):2437–2448, 2025.

Adam Gayoso, Zoë Steier, Romain Lopez, Jeffrey Regier, Kristopher L Nazor, Aaron Streets, and Nir Yosef. Joint probabilistic modeling of single-cell multi-omic data with totalvi. *Nature methods*, 18(3):272–282, 2021.

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.

Shizhe Hu, Binyan Tian, Weibo Liu, and Yangdong Ye. Self-supervised trusted contrastive multi-view clustering with uncertainty refined. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 17305–17313, 2025.

Yinlei Hu, Siyuan Wan, Yuanhanyu Luo, Yuanzhe Li, Tong Wu, Wentao Deng, Chen Jiang, Shan Jiang, Yueping Zhang, Nianping Liu, et al. Benchmarking algorithms for single-cell multi-omics prediction and integration. *Nature Methods*, 21(11):2182–2194, 2024.

Kasia Z Kedzierska, Lorin Crawford, Ava P Amini, and Alex X Lu. Zero-shot evaluation reveals limitations of single-cell foundation models. *Genome Biology*, 26(1):101, 2025.

Wenyuan Kong, Zhibin Gu, and Bing Li. Dual-branch representations with dynamic gated fusion and triple-granularity alignment for deep multi-view clustering. In *International Conference on Learning Representations*, 2026.

Jiawei Li, Taoyuan Ye, Yilang Xiao, Mengyuan Zhao, Limin Jiang, Shizhan Chen, Fei Guo, and Jijun Tang. Udec-mo: an uncertainty-guided deep embedded clustering framework for bulk and single-cell multi-omics data. *Briefings in Bioinformatics*, 27(4):bbag435, 2026.

Pengyuan Li, Man Liu, Dongxia Chang, Yiming Wang, Zisen Kong, and Yao Zhao. Aemvc: Mitigate imbalanced embedding space in multi-view clustering. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 6461–6470, 2025.

Wei Li, Bin Xiang, Fan Yang, Yu Rong, Yanbin Yin, Jianhua Yao, and Han Zhang. scmhnn: a novel hypergraph neural network for integrative analysis of single-cell epigenomic, transcriptomic and proteomic data. *Briefings in bioinformatics*, 24(6):bbad391, 2023.

Sai Ma, Bing Zhang, Lindsay M LaFave, Andrew S Earl, Zachary Chiang, Yan Hu, Jiarui Ding, Alison Brack, Vinay K Kartha, Tristan Tay, et al. Chromatin potential identified by shared single-cell profiling of rna and chromatin. *Cell*, 183(4):1103–1116, 2020.

Shuangmei Peng, Zhe Chen, Tianyang Xu, and Xiao-Jun Wu. Relationship alignment for view-aware multi-view clustering. In *International Conference on Learning Representations*, 2026.

Liangrui Ren, Jun Wang, Zhao Li, Qingzhong Li, and Guoxian Yu. scmcs: a framework for single-cell multi-omics data integration and multiple clusterings. *Bioinformatics*, 39(4):btad133, 2023.

Yanay Rosen, Yusuf Roohani, Ayush Agrawal, Leon Samotorčan, Tabula Sapiens Consortium, Stephen R Quake, and Jure Leskovec. Universal cell embedding provides a foundation model for cell biology. *Nature*, pp. 1–9, 2026.

Marlon Stoeckius, Christoph Hafemeister, William Stephenson, Brian Houck-Loomis, Pratip K Chattopadhyay, Harold Swerdlow, Rahul Satija, and Peter Smibert. Simultaneous epitope and transcriptome measurement in single cells. *Nature methods*, 14(9):865–868, 2017.

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.

Gefei Wang, Jia Zhao, Yingxin Lin, Tianyu Liu, Yize Zhao, and Hongyu Zhao. scmodal: a general deep learning framework for comprehensive single-cell multi-omics data alignment with feature links. *Nature Communications*, 16(1):4994, 2025.

Yixuan Wang, Yimin Fan, Xuesong Wang, Tingyang Yu, Yongshuo Zong, Xinyuan Liu, Gaoyang Zhong, Meitong Liu, Qing Li, Kin Hei Lee, et al. Scmbench: benchmarking domain-specific and foundation models for single-cell multi-omics data integration. *Nature Communications*, 2026.

Hongzhi Wen, Wenzhuo Tang, Xinnan Dai, Jiayuan Ding, Wei Jin, Yuying Xie, and Jiliang Tang. Cellplm: Pre-training of cell language model beyond single cells. In *International Conference on Learning Representations*, volume 2024, pp. 5649–5673, 2024.

Yichen Xu, Ryumei Nakada, Linjun Zhang, and Lexin Li. Residual feature integration is sufficient to prevent negative transfer. In *International Conference on Learning Representations*, volume 2026, pp. 77738–77763, 2026.

Weiqing Yan, Yuanyang Zhang, Chang Tang, Wujie Zhou, and Weisi Lin. Anchor-sharing and cluster-wise contrastive network for multiview representation learning. *IEEE Transactions on Neural Networks and Learning Systems*, 36(2):3797–3807, 2024.

Xiuhui Yang, Koren K Mann, Hao Wu, and Jun Ding. sccross: a deep generative model for unifying single-cell multi-omics with seamless integration, cross-modal generation, and in silico exploration. *Genome Biology*, 25(1):198, 2024.

Sukwon Yun, Jie Peng, Namkyeong Lee, Yanyong Zhang, Chanyoung Park, Zunpeng Liu, and Tianlong Chen. scmoe: single-cell multi-modal multi-task learning via sparse mixture-of-experts. *bioRxiv*, pp. 2024–11, 2024.

Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48, pp. 478–487, 2016.

Yuanyang Zhang, Yijie Lin, Weiqing Yan, Li Yao, Xinhang Wan, Guangyuan Li, Chao Zhang, Guanzhou Ke, and Jie Xu. Incomplete multi-view clustering via diffusion contrastive generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 22650–22658, 2025.

Chunman Zuo and Luonan Chen. Deep-joint-learning analysis model of single cell transcriptome and open chromatin accessibility data. *Briefings in Bioinformatics*, 22(4):bbaa287, 2021.

A Related Work

A.1 Multi-View Clustering for Transcriptomics

Multi-view clustering aims to integrate complementary information from heterogeneous views while preserving their shared structure. Existing methods mainly rely on representation alignment, contrastive learning, generative modeling, or graph-based constraints. Representative approaches such as ACCMVC Yan et al. (2024), AEMVC Li et al. (2025), and DCG Zhang et al. (2025) improve clustering by aligning view-specific representations or reconstructing incomplete information. Reliability-aware methods further account for heterogeneous view quality. For example, STCMC-UR Hu et al. (2025) incorporates evidential uncertainty into contrastive clustering, while recent methods employ adaptive gated fusion or view-aware alignment to reduce the influence of redundant or conflicting information Kong et al. (2026); Peng et al. (2026).

In single-cell transcriptomics, multi-view learning has been widely used to integrate heterogeneous omics measurements. Methods such as scMVAE Zuo & Chen (2021), scMHNN Li et al. (2023), scMoE Yun et al. (2024), and scMCs Ren et al. (2023) model shared and complementary information across modalities using variational learning, hypergraphs, mixture-of-experts, or multi-view representation learning. Recent approaches including MOH Dong et al. (2025) and scMUSCLE Ding et al. (2026) further exploit graph structure and contrastive objectives for clustering. However, these methods mainly consider views directly observed in the target dataset, whereas externally transferred representations may introduce an additional source of mismatch that requires different reliability control. Trust-VC assigns distinct roles to these sources: RNA supplies a fixed reference, and auxiliary measurements and transferred representations supply gated corrections relative to it.

A.2 Foundation-Model Priors for Transcriptomic Representation Learning

Single-cell foundation models provide transferable representations learned from large-scale transcriptomic corpora. Representative models include Geneformer Theodoris et al. (2023), scGPT Cui et al. (2024), scFoundation Hao et al. (2024), CellPLM Wen et al. (2024), and UCE Rosen et al. (2026). These models encode broad biological knowledge and can provide complementary information for downstream transcriptomic analysis beyond the measurements available in a specific target dataset.

Nevertheless, pretrained representations are not uniformly transferable across datasets. Recent evaluations show that single-cell foundation models do not consistently outperform conventional or task-specific representations under zero-shot or multi-omics settings Kedzierska et al. (2025); Wang et al. (2026). Differences in gene coverage, sequencing protocols, cellular composition, and pretraining distributions may lead to source–target mismatch. Therefore, directly treating a foundation-model embedding as an equally reliable additional view may be suboptimal. Trust-VC instead regards it as a *virtual-cell prior* and adaptively controls its residual contribution to the fused representation relative to a fixed RNA reference. The prior changes the fused representation; it does not update the reference itself.

B Additional Optimization Details

The auxiliary virtual-prior regularization term in Eq. (26) is

$$\mathcal{R}_v = \lambda^v_{\mathrm{KL}}\mathcal{L}^v_{\mathrm{KL}} + \lambda_{v,\mathrm{contrib}}\mathcal{L}_{v,\mathrm{contrib}} + \lambda_{\mathrm{use,reg}}\mathcal{L}_{\mathrm{use,reg}}, \tag{35}$$

where

$$\mathcal{L}^v_{\mathrm{KL}} = \frac{1}{2N}\sum_{i,j}\left[(\mu^v_{ij})^2 + (s^v_{ij})^2 - \log\big((s^v_{ij})^2\big) - 1\right], \tag{36}$$

$$\mathcal{L}_{v,\mathrm{contrib}} = \frac{1}{N}\sum_i \|\mathbf{c}^v_i\|_2, \tag{37}$$

$$\mathcal{L}_{\mathrm{use,reg}} = \frac{1}{N}\sum_i (g^{v,\mathrm{use}}_i - 0.5)^2. \tag{38}$$

The first term is a Gaussian latent KL regularizer between $q_\phi(\mathbf{z}^v_i \mid \mathbf{x}^v_i)$ and the standard Gaussian prior $\mathcal{N}(\mathbf{0}, \mathbf{I})$; it is not used as a variational information-bottleneck objective in the strict sense because stochastic samples do not enter the downstream clustering losses in the reported configuration.

Table 2: Key Trust-VC coefficients and main-experiment defaults. Only σ, β, and γ are varied in the sensitivity sweeps. The separate checkpoint settings for prior-permutation diagnostics are given in Table 5.

Symbol	Domain	Default	Role / strategy
ϵ	> 0	10^{-8}	Numerical stabilizer for latent normalization; fixed.
ϵ_{DEC}	> 0	10^{-8}	Numerical stabilizer for mini-batch DEC frequencies; fixed.
τ_m	> 0	10.0	Distance temperature in pairwise trust; fixed.
λ_b	$[0, 1]$	0.4	Blend ratio between distance and learned trust; fixed.
α_m	≥ 0	0.48	Scale of the measured-modality residual; fixed.
δ	≥ 0	1.0	Margin of RNA-reference preservation; fixed.
τ_d	> 0	1.0	Temperature of latent-dispersion attenuation; fixed.
b_s	≥ 0	5.0	Offset in the softplus posterior-scale parameterization; fixed.
ϵ_s	> 0	10^{-6}	Minimum posterior standard deviation; fixed.
ϵ_u	$[0, 1)$	0.05	Floor applied to both confidence-gain prediction and target; fixed.
τ_{use}	> 0	0.1	Temperature of the confidence-gain target; fixed.
ρ_v	≥ 1	1.05	Power shaping of the composite virtual gate; fixed.
α_v	≥ 0	0.20	Scale of the virtual residual; linearly warmed up over the first 5 epochs.
ν	> 0	1.0	Student-t degree of freedom in the DEC head; fixed.
$\lambda_{\mathrm{lat}}^{r+m}$	≥ 0	0.01	RNA+measured latent distillation weight; fixed.
$\lambda_{\mathrm{lat}}^{r+v}$	≥ 0	0.015	RNA+virtual latent distillation weight; fixed.
λ_q^{r+m}	≥ 0	0.01	RNA+measured assignment distillation weight; fixed.
λ_q^{r+v}	≥ 0	0.015	RNA+virtual assignment distillation weight; fixed.
$\sigma = \lambda_{\mathrm{back}}$	≥ 0	0.03	Reference-preservation weight; varied in sensitivity analysis.
$\beta = \lambda_{v,\mathrm{sparse}}$	≥ 0	0.0015	Virtual-gate sparsity weight; varied in sensitivity analysis.
$\gamma = \lambda_{v,\mathrm{use}}$	≥ 0	0.08	Confidence-gain proxy supervision weight; varied in sensitivity analysis.
λ_{KL}^v	≥ 0	4×10^{-5}	Gaussian latent KL weight; linearly warmed up over the first 10 epochs.
$\lambda_{v,\mathrm{contrib}}$	≥ 0	3×10^{-4}	Virtual residual-magnitude penalty; fixed.
$\lambda_{\mathrm{use,reg}}$	≥ 0	10^{-4}	Confidence-score regularization weight; fixed.

The other two terms constrain virtual residual magnitude and discourage saturation of the predicted confidence-gain score toward either boundary.

Implementation details. The RNA reference is constructed once by library-size normalization to 10^4, $\log(1 + x)$ transformation, variance-based selection of 2,000 highly variable genes, standardization, and PCA to 32 dimensions. The resulting RNA representations are held fixed throughout clustering optimization; no trainable RNA encoder is included in the reported configuration. The measured auxiliary encoder first applies a linear projection to 256 dimensions, LayerNorm, ReLU, and dropout ($p = 0.1$), followed by two 256-dimensional residual blocks and a final linear projection to $d = 32$. Each residual block contains two Linear–LayerNorm–ReLU–Dropout stages with an identity skip connection. The virtual encoder uses the same 256-dimensional input projection, LayerNorm, ReLU, and dropout, but no residual blocks in the reported configuration; separate linear heads map the hidden representation to $\boldsymbol{\mu}_i^v$ and $\boldsymbol{\eta}_i^v$. The latter head uses Xavier-uniform weight initialization and zero bias before the softplus-offset transformation in Section 2.3. The pairwise trust MLP is $4d \to 128 \to 1$ with ReLU and sigmoid output, while the confidence-gain MLP is $5d \to 64 \to 1$ with the same activations. All reported experiments use scalar gates. The reported implementation also supports vector-valued gates, but that option is not used for the results in this paper. The main-experiment configuration uses Adam with learning rate 10^{-3}, weight decay 10^{-5}, batch size 64, and at most 100 epochs. The prior-permutation diagnostic uses the separately documented checkpoints in Table 5.

The reported configuration uses $\boldsymbol{\mu}_i^v$ for both virtual residual fusion and virtual trust evaluation. The posterior scale $\mathbf{s}_i^v$ remains trainable through the Gaussian KL term and the dispersion-dependent

Table 3: Summary of the datasets used in this study. Size is the number of cells, and Clusters is the number of annotated cell types used to set K.

Dataset	Size	Clusters	Dimensions	Platform
ASSP	8689	12	4768/216	Illumina NovaSeq 6000
BMNC-3K	3000	27	17009/25	Illumina HiSeq 2500
BMNC-5K	5000	27	17009/25	Illumina HiSeq 2500
BMNC-10K	10000	27	17009/25	Illumina HiSeq 2500
GSE128639	30672	27	17009/25	Illumina NovaSeq 6000
PBMC5K	5247	14	22538/32	10x Genomics
SLN111	16828	35	13553/112	Illumina NovaSeq 6000
SLN206	15820	35	13553/209	Illumina NovaSeq 6000

gate. For epoch index $t \in \{1, 2, \ldots\}$, the two linear warm-up schedules are

$$\alpha_v^{(t)} = \alpha_v \min\left(\frac{t}{E_v}, 1\right), \qquad E_v = 5, \tag{39}$$

and

$$(\lambda_{\mathrm{KL}}^v)^{(t)} = \lambda_{\mathrm{KL}}^v \min\left(\frac{t}{E_{\mathrm{KL}}}, 1\right), \qquad E_{\mathrm{KL}} = 10. \tag{40}$$

Thus the virtual residual is introduced gradually, while the Gaussian latent regularizer reaches its nominal weight over ten epochs. The confidence-gain floor is applied with the same affine transform to both the prediction and target in Eq. (21); hard thresholding of the final virtual gate is disabled.

The RNA-only latent and assignment consistency weights are both zero, so the fixed RNA reference is not a distillation student. The active weights for the two residual-enhanced partial views are those listed in Table 2. Before the first gradient-based training epoch, cluster centers are initialized by K-means ($n_{\mathrm{init}} = 20$, random state 0) on the initial complete-view representations, before the training epochs indexed by $t \geq 1$. Thereafter, the cluster centers are optimized jointly with the model. At every mini-batch update, the mini-batch DEC target distribution is recomputed from the current $\mathbf{Q}^{\mathrm{all}}$ and detached before the KL loss is evaluated. Best-checkpoint selection is based on an unsupervised training loss after the first 20 epochs, and final labels are obtained from the maximum-probability assignments of $\mathbf{Q}^{\mathrm{all}}$.

C Experimental Details

The eight datasets span peripheral blood, bone marrow, spleen, and lymph-node tissues, with 3,000–30,672 cells and 12–35 annotated cell types. We compare scMVAE, scMoE, scMHNN, ACCMVC, DCG, STCMC-UR, AEMVC, scMCs, MOH, and scMUSCLE Zuo & Chen (2021); Yun et al. (2024); Li et al. (2023); Yan et al. (2024); Zhang et al. (2025); Hu et al. (2025); Li et al. (2025); Ren et al. (2023); Dong et al. (2025); Ding et al. (2026). Results use repeated runs of transductive clustering with K fixed to the annotated cell-type count. Validation labels serve only for virtual-prior source selection, and held-out test labels only for final evaluation.

We evaluated Trust-VC against ten representative baselines, comprising six single-cell multi-omics methods and four general multi-view clustering methods.

- **scMVAE** Zuo & Chen (2021) integrates paired scRNA-seq and scATAC-seq through VAE-based joint learning and Gaussian-mixture clustering.
- **scMoE** Yun et al. (2024) combines cross-attention with sparse mixture-of-experts routing for multimodal and multitask representation learning.
- **scMHNN** Li et al. (2023) models higher-order cross-omics relationships using multimodal hypergraphs and self-supervised contrastive learning.
- **ACCMVC** Yan et al. (2024) learns view-common representations through shared-anchor aggregation and cluster-wise contrastive alignment.
- **DCG** Zhang et al. (2025) reconstructs missing views through diffusion generation and aligns cross-view representations using contrastive learning.
- **STCMC-UR** Hu et al. (2025) incorporates evidential uncertainty into contrastive clustering to reduce the influence of unreliable pseudo-labels and sample pairs.

- **AEMVC** Li et al. (2025) alleviates directional imbalance in the embedding space through feature activation and inter-view mutual supervision.
- **scMCs** Ren et al. (2023) jointly captures omics-common and omics-specific information and derives multiple low-redundancy clustering structures.
- **MOH** Dong et al. (2025) propagates information within and across omics layers using a multilayer heterogeneous graph.
- **scMUSCLE** Ding et al. (2026) combines structural enhancement, multi-subspace contrastive learning, and adaptive graph smoothing for multimodal clustering.

D Full Ablation Results

To complement the representative ablation results in the main text, we report the complete evaluation across all eight benchmark datasets: ASSP, BMNC-3K, BMNC-5K, BMNC-10K, GSE128639, PBMC5K, SLN111, and SLN206. We use the same four variants defined in Section 3.4: removing cross-view consistency, the RNA-reference preservation penalty, trust-aware virtual learning, or the virtual prior. The cross-view-consistency ablation retains the complete-view clustering objective. All four variants retain the RNA reference; removing the preservation penalty does not remove the reference itself.

The complete Trust-VC model achieves the highest mean score in all 24 dataset–metric comparisons. The magnitude of the reduction caused by each ablation varies across datasets. For example, removing the virtual prior on GSE128639 reduces mean ACC from 0.699 to 0.545 and mean ARI from 0.658 to 0.446, corresponding to 15.4 and 21.2 percentage points, respectively. These comparisons support the contributions of the four components under the reported configurations; they do not isolate the individual trust signals within the composite gate. The full-model entries agree with Table 1 after conversion from percentages.

E Additional Analysis of Virtual-Prior Permutation

E.1 Diagnostic Protocol and Checkpoint Configuration

We analyze previously saved, frozen-checkpoint evaluations on GSE128639 and PBMC5K. Both datasets use UCE as the virtual-prior source and five training seeds: 3407, 2027, 1997, 42, and 12345. This diagnostic uses its own clean-trained checkpoints and the full exported benchmark cell sets (30,672 and 5,247 cells, respectively). Its absolute scores are therefore reported separately from the main comparison in Table 1. The checkpoint configurations are summarized in Table 5; these settings override the main-experiment defaults for this analysis.

At each nominal level $r \in \{0, 0.25, 0.50, 0.75, 1.00\}$, virtual-prior rows are permuted within a selected subset of cells. The subsets are nested and generated with permutation seed 20260815. GSE128639 uses a cyclic derangement of the selected indices, while PBMC5K uses a random derangement. These permutations have no fixed indices within the selected subset and preserve the empirical distribution of virtual-prior rows. The same corrupted prior is used for all five training seeds and both inference variants at a given level. Consequently, the reported variability is across training seeds conditional on the saved permutations, not across independent permutation replicates. Rounding to whole cells causes negligible deviations from the nominal fractions on PBMC5K.

For each checkpoint, Trust-VC retains its learned virtual gate, while GateOff forces the effective gate to one. Model parameters and cluster centers remain unchanged, and evaluation uses the virtual posterior mean. This is an inference intervention, not the trained trust-aware ablation in Section 3.4. We verified that the saved RNA representations and measured-modality corrections are identical across permutation levels and gate settings, and that the saved outputs satisfy $\mathbf{z}_i^{\mathrm{all}} = \mathbf{z}_i^r + \mathbf{c}_i^m + \mathbf{c}_i^v$. The additional diagnostics below are calculated from those saved outputs; no retraining or new model selection is performed.

We distinguish two representation diagnostics:

$$C_v(r) = \frac{1}{N}\sum_i \|\mathbf{c}_i^v(r)\|_2, \qquad D_{\mathrm{RNA}}(r) = \frac{1}{N}\sum_i \|\mathbf{z}_i^{\mathrm{all}}(r) - \mathbf{z}_i^r\|_2. \tag{41}$$

Table 4: Ablation study of Trust-VC on eight datasets (mean ± standard deviation, original metric scale). Bold marks the highest mean for each dataset and metric. Multiply values by 100 to compare with Table 1.

(a) Setting	ASSP			BMNC-3K		
	ACC	NMI	ARI	ACC	NMI	ARI
w/o Cross-view Consistency	0.629 ± 0.031	0.637 ± 0.019	0.448 ± 0.033	0.553 ± 0.024	0.696 ± 0.022	0.463 ± 0.048
w/o RNA Reference Preservation	0.635 ± 0.038	0.641 ± 0.012	0.456 ± 0.037	0.566 ± 0.018	0.733 ± 0.015	0.507 ± 0.027
w/o Trust-aware Virtual Learning	0.639 ± 0.038	0.638 ± 0.017	0.457 ± 0.024	0.573 ± 0.017	0.727 ± 0.012	0.512 ± 0.018
w/o Virtual Prior	0.624 ± 0.019	0.634 ± 0.011	0.444 ± 0.017	0.512 ± 0.048	0.662 ± 0.012	0.403 ± 0.038
Trust-VC	**0.698** ± 0.005	**0.667** ± 0.004	**0.511** ± 0.013	**0.629** ± 0.006	**0.748** ± 0.014	**0.557** ± 0.018

(b) Setting	BMNC-5K			BMNC-10K		
	ACC	NMI	ARI	ACC	NMI	ARI
w/o Cross-view Consistency	0.546 ± 0.039	0.695 ± 0.033	0.455 ± 0.066	0.551 ± 0.037	0.716 ± 0.028	0.474 ± 0.055
w/o RNA Reference Preservation	0.585 ± 0.030	0.744 ± 0.022	0.530 ± 0.038	0.541 ± 0.041	0.731 ± 0.022	0.483 ± 0.057
w/o Trust-aware Virtual Learning	0.575 ± 0.025	0.725 ± 0.021	0.505 ± 0.041	0.566 ± 0.028	0.733 ± 0.014	0.505 ± 0.039
w/o Virtual Prior	0.512 ± 0.023	0.659 ± 0.010	0.404 ± 0.026	0.548 ± 0.024	0.695 ± 0.016	0.467 ± 0.031
Trust-VC	**0.663** ± 0.026	**0.763** ± 0.016	**0.602** ± 0.021	**0.658** ± 0.013	**0.768** ± 0.008	**0.623** ± 0.028

(c) Setting	GSE128639			PBMC5K		
	ACC	NMI	ARI	ACC	NMI	ARI
w/o Cross-view Consistency	0.588 ± 0.031	0.754 ± 0.015	0.524 ± 0.046	0.547 ± 0.047	0.619 ± 0.009	0.415 ± 0.043
w/o RNA Reference Preservation	0.598 ± 0.040	0.762 ± 0.019	0.540 ± 0.049	0.535 ± 0.037	0.624 ± 0.010	0.396 ± 0.035
w/o Trust-aware Virtual Learning	0.590 ± 0.031	0.751 ± 0.021	0.525 ± 0.054	0.540 ± 0.050	0.616 ± 0.011	0.408 ± 0.047
w/o Virtual Prior	0.545 ± 0.038	0.703 ± 0.017	0.446 ± 0.037	0.553 ± 0.041	0.618 ± 0.004	0.422 ± 0.039
Trust-VC	**0.699** ± 0.032	**0.795** ± 0.021	**0.658** ± 0.041	**0.607** ± 0.000	**0.633** ± 0.001	**0.469** ± 0.000

(d) Setting	SLN111			SLN206		
	ACC	NMI	ARI	ACC	NMI	ARI
w/o Cross-view Consistency	0.358 ± 0.026	0.535 ± 0.014	0.253 ± 0.026	0.369 ± 0.010	0.529 ± 0.014	0.282 ± 0.010
w/o RNA Reference Preservation	0.360 ± 0.027	0.544 ± 0.006	0.259 ± 0.026	0.356 ± 0.013	0.538 ± 0.009	0.262 ± 0.024
w/o Trust-aware Virtual Learning	0.361 ± 0.025	0.534 ± 0.014	0.255 ± 0.026	0.380 ± 0.008	0.540 ± 0.007	0.296 ± 0.006
w/o Virtual Prior	0.334 ± 0.023	0.515 ± 0.010	0.223 ± 0.018	0.359 ± 0.015	0.514 ± 0.012	0.271 ± 0.018
Trust-VC	**0.410** ± 0.004	**0.560** ± 0.006	**0.301** ± 0.006	**0.397** ± 0.002	**0.557** ± 0.007	**0.303** ± 0.009

Table 5: Configuration of the clean-trained checkpoints used for the virtual-prior permutation diagnostic. Epoch lists follow the seed order 3407, 2027, 1997, 42, and 12345.

Setting	GSE128639	PBMC5K
Virtual-prior source	UCE	UCE
Learning rate	10^{-4}	10^{-4}
Configured training epochs	200	200
Batch size	64	64
Measured residual scale α_m	0.44	0.48
Virtual residual scale α_v	0.20	0.10
(σ, β, γ)	$(0.03, 0.0015, 0.08)$	$(0.03, 0.0015, 0.08)$
Reference margin δ	1.0	1.0
Selected checkpoint epochs	128, 180, 135, 130, 155	28, 93, 21, 22, 24

The second equals the mean norm of the combined residual $\mathbf{c}_i^m + \mathbf{c}_i^v(r)$, not the sum of the two branch norms. Norms are computed per cell and averaged within each seed before aggregation across seeds. We report means and sample standard deviations over five seeds. ACC uses Hungarian matching; NMI and ARI use the cluster assignments directly; F1 is macro-averaged after Hungarian matching. All scores in these diagnostic figures use the original metric scale.

E.2 Residual Regulation on PBMC5K

Figure 6 repeats the main diagnostic on PBMC5K. From 0% to 100% permutation, the mean virtual correction norm decreases from 1.809 to 1.672 under Trust-VC, while increasing from 2.641 to 2.752 under GateOff. Mean net displacement from RNA changes from 6.558 to 6.259 and from 7.320 to 7.186, respectively. Thus, the net displacement decreases for both variants on this dataset, and remains smaller under adaptive gating. At 100% permutation, mean absolute ARI is 0.415 for Trust-VC and 0.412 for GateOff; the clustering effect is smaller than on GSE128639. The reference-preservation objective is a soft penalty, so D_{RNA} is not required to fall below δ. These distances describe the learned representation geometry and do not directly establish preservation of biological neighborhoods or isolate the effect of $\mathcal{L}_{\mathrm{RP}}$.

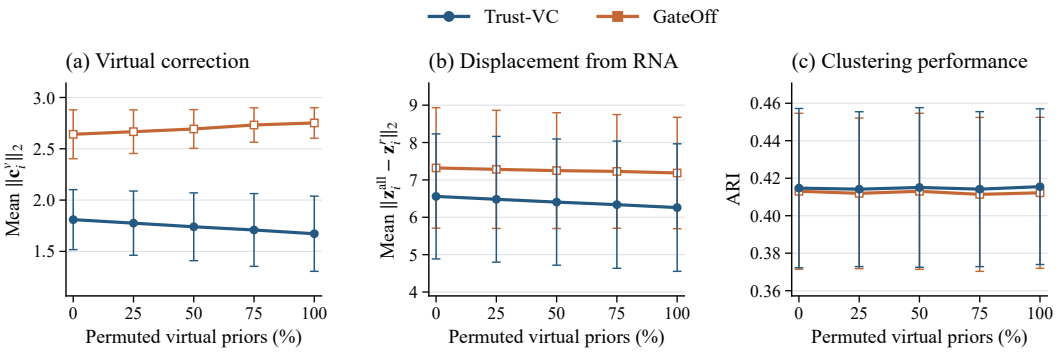

Figure 6: Virtual-prior permutation diagnostic on PBMC5K. (a) Mean virtual correction norm. (b) Mean net displacement from the fixed RNA reference. (c) Absolute ARI. Points and error bars are means and standard deviations across five training seeds.

E.3 Trust Responses for the Same Cells

Mean virtual trust decreases from 0.933 to 0.767 on GSE128639 and from 0.698 to 0.616 on PBMC5K as permutation increases from 0% to 100% (Fig. 7a,b). To distinguish a response to reassigned priors from differences in cell composition, we also compare each cell's gate with its own clean-prior gate at 50% permutation. For the permuted subset, the seed-averaged gate change is -0.166 ± 0.059 on GSE128639 and -0.082 ± 0.058 on PBMC5K. The gates of unpermuted cells are unchanged in the saved outputs. Each replicate in this analysis is a seed-level average over the same cells before and after permutation, rather than an independent-cell replicate. This supports sensitivity to reassigned priors without establishing calibrated reliability or separate effects of the three constituent trust signals.

E.4 Absolute Clustering Scores

Figure 8 shows dataset- and metric-dependent responses: GSE128639 ACC and ARI increase at high permutation levels, whereas NMI and F1 differ; PBMC5K scores change little. Permutation need not worsen clustering and may exchange priors within an annotated cell type. At 100% permutation, 89.3% of GSE128639 cells and 78.0% of PBMC5K cells receive a prior from a different type. These post-evaluation fractions use annotations; the perturbations themselves do not. Smaller RNA displacement alone does not establish better clustering or imply a noise-free RNA reference.

E.5 Paired Changes Relative to Clean Baselines

To complement the absolute scores, we retain the paired difference in changes from each variant's own clean-prior baseline:

$$A_M(r) = \Delta M_{\text{Trust-VC}}(r) - \Delta M_{\text{GateOff}}(r), \tag{42}$$

where

$$\Delta M(r) = M(r) - M(0). \tag{43}$$

We aggregate within-seed differences. Positive $A_M(r)$ indicates a more favorable change, not the absolute score gap at r. On GSE128639, all 16 nonzero-permutation–metric combinations are

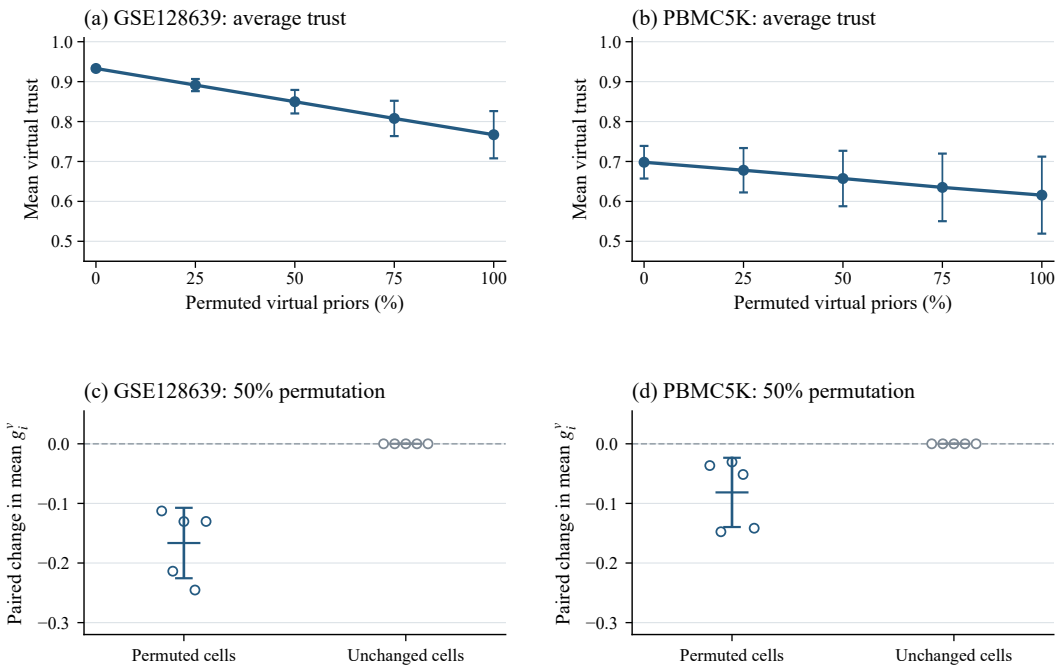

Figure 7: Virtual-trust responses under fixed-checkpoint permutation. (a,b) Mean trust on GSE128639 and PBMC5K, respectively. (c,d) Paired changes in mean trust at 50% permutation, evaluated on the same cells relative to their clean-prior gates, separately for permuted and unchanged subsets. Each open circle is one training seed; horizontal marks and error bars denote the mean and standard deviation across five seeds. The upper panels use the same mean and standard-deviation summary.

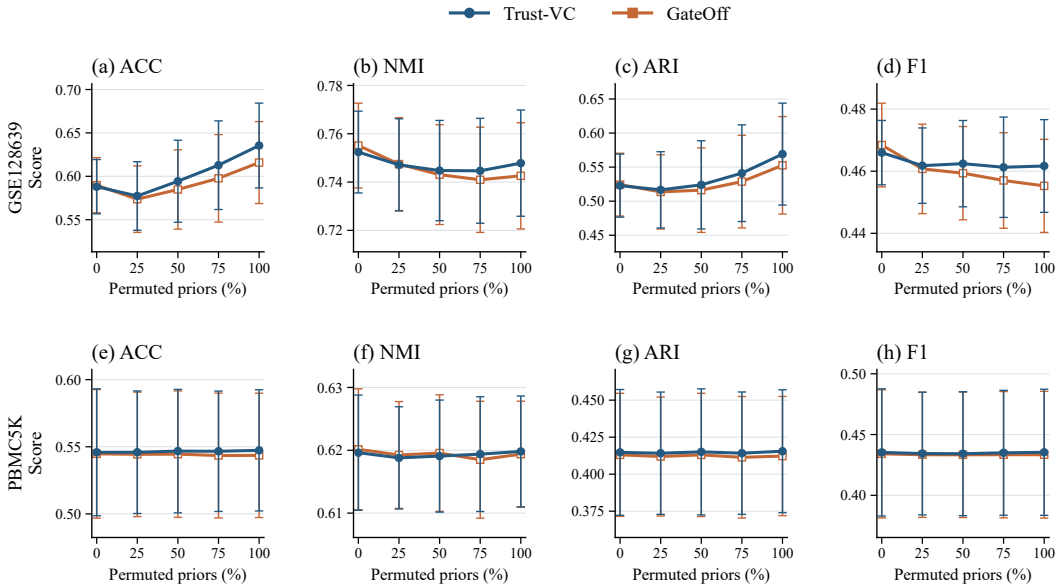

Figure 8: Absolute clustering scores under virtual-prior permutation. Panels (a–d) show ACC, NMI, ARI, and macro-F1 on GSE128639; panels (e–h) show the corresponding PBMC5K scores. Points and error bars show means and standard deviations across five training seeds. These scores use the diagnostic checkpoints and cell sets specified in Table 5.

positive in all five seeds. At 100% permutation, mean ACC, NMI, ARI, and F1 differences are $(21.31, 7.95, 17.96, 8.85) \times 10^{-3}$; on PBMC5K they are $(2.59, 0.98, 1.58, 0.83) \times 10^{-3}$, positive in 5/5, 3/5, 4/5, and 4/5 seeds. The absolute ARI gaps, 0.01650 and 0.00325, differ from the paired changes, 0.01796 and 0.00158. The heatmap reports descriptive effects and seed-direction counts without significance claims.

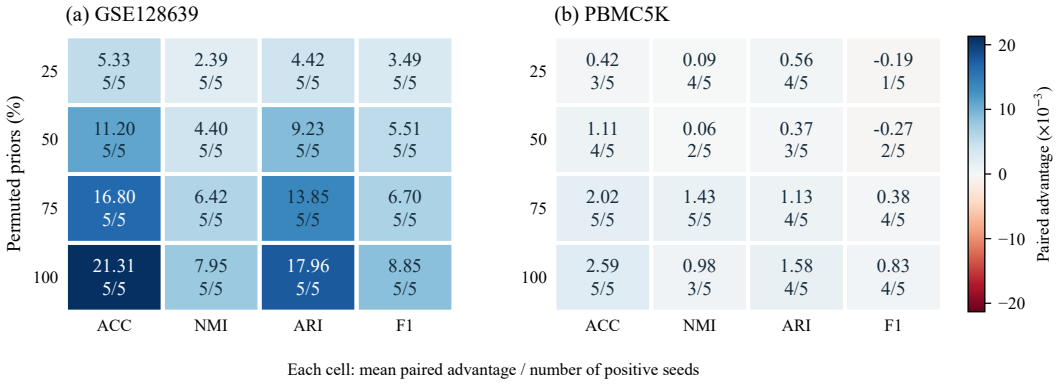

Figure 9: Paired differences in changes from clean-prior baselines on (a) GSE128639 and (b) PBMC5K. Each cell shows the mean $A_M(r) \times 10^3$ above the number of seeds with $A_M(r) > 0$. Both panels share a symmetric color scale centered at zero. One displayed unit corresponds to 0.1 percentage points. No significance markers are used; the counts summarize five training seeds under the same saved permutation at each level.

E.6 Scope of the Diagnostic

Adaptive gating regulates virtual corrections and yields smaller mean RNA displacement than Gate-Off, with dataset-dependent clustering effects. This does not establish an optimal RNA reference, uniformly harmful perturbations, or the necessity of each trust signal. Independent permutation replicates and separate trained component ablations are needed to examine those questions.

F Additional Parameter Sensitivity Analysis

We extend the ASSP analysis in Fig. 5 to the remaining seven datasets. We vary RNA-reference preservation ($\sigma = \lambda_{\text{back}}$), virtual-prior sparsity ($\beta = \lambda_{v,\text{sparse}}$), and assignment-confidence proxy fitting ($\gamma = \lambda_{v,\text{use}}$). Each column fixes γ and varies (σ, β); all other hyperparameters remain fixed within each sweep. Rows show ACC, ARI, and NMI on their original scale. The displayed grids characterize local sensitivity and do not establish insensitivity to arbitrary parameter choices.

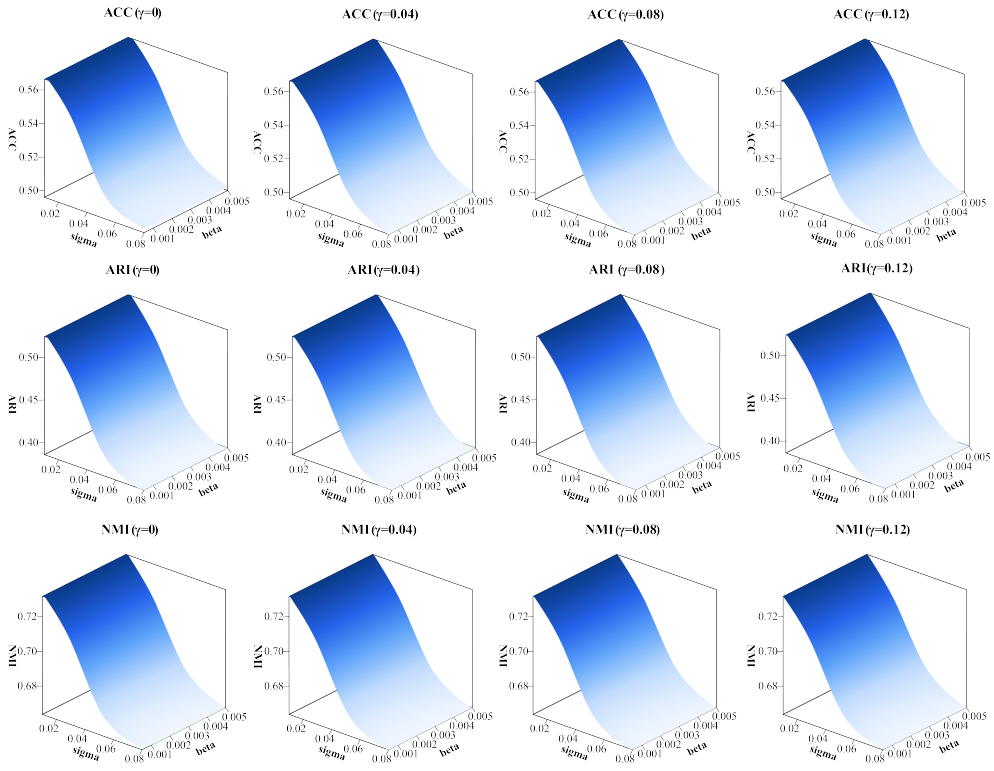

Figure 10: Sensitivity on BMNC-3K. Rows: ACC, ARI, NMI; columns: fixed γ. Surfaces vary (σ, β); other coefficients are fixed within each sweep.

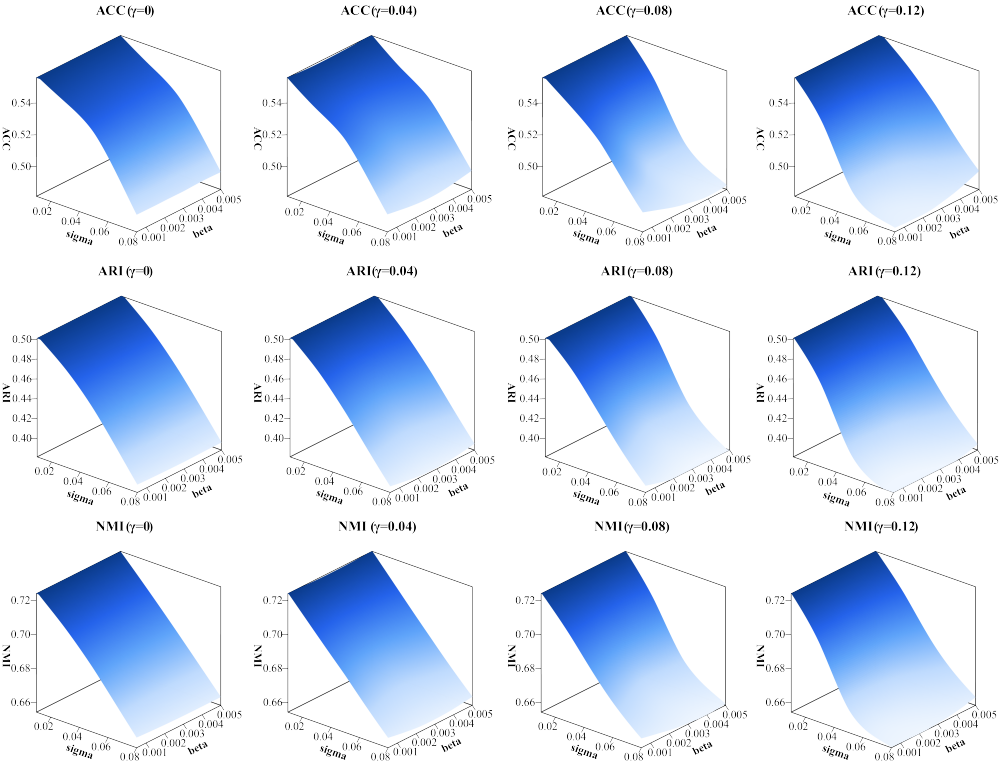

Figure 11: Sensitivity on BMNC-5K. Rows: ACC, ARI, NMI; columns: fixed γ. Surfaces vary (σ, β); other coefficients are fixed within each sweep.

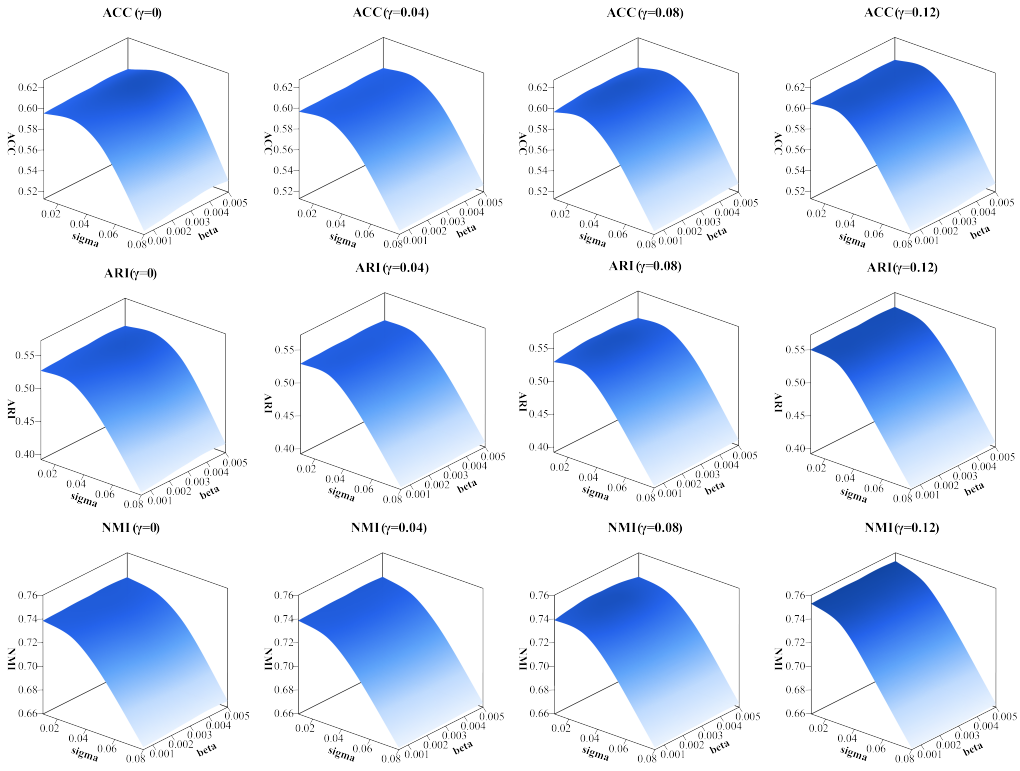

Figure 12: Sensitivity on BMNC-10K. Rows: ACC, ARI, NMI; columns: fixed γ. Surfaces vary (σ, β); other coefficients are fixed within each sweep.

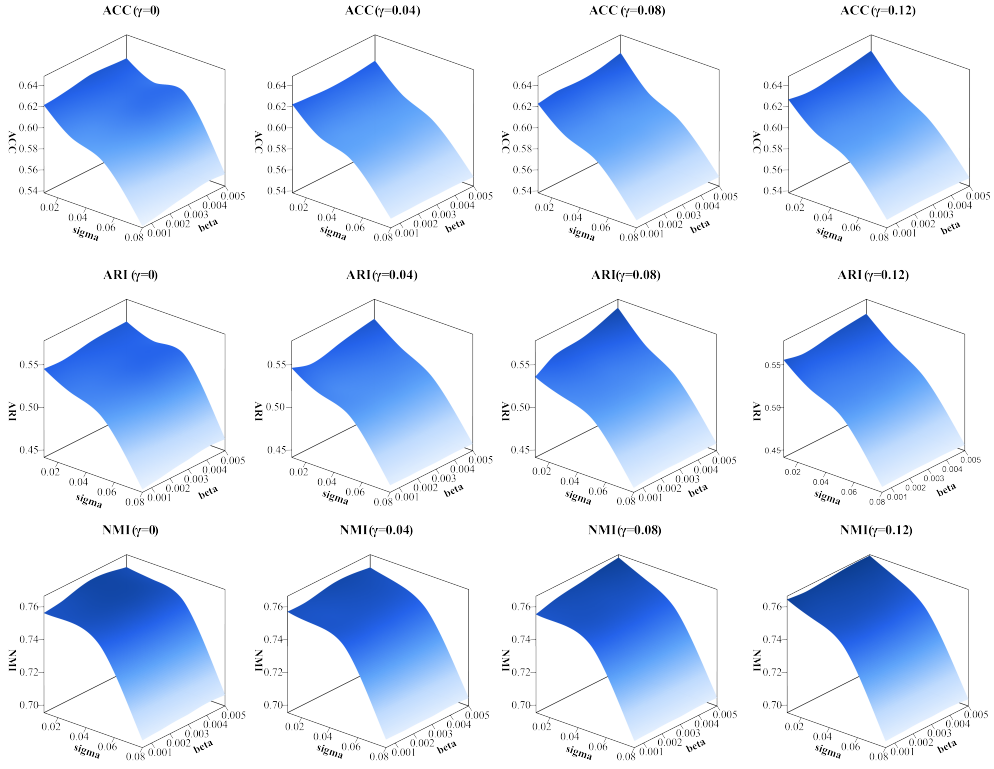

Figure 13: Sensitivity on GSE128639. Rows: ACC, ARI, NMI; columns: fixed γ. Surfaces vary (σ, β); other coefficients are fixed within each sweep.

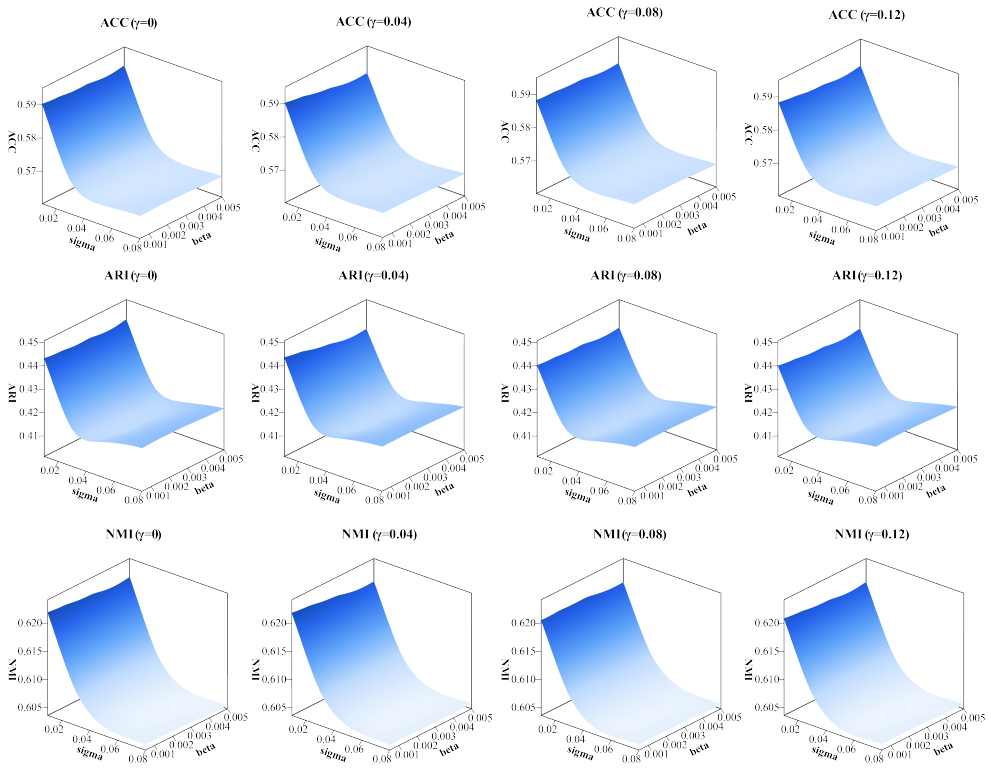

Figure 14: Sensitivity on PBMC5K. Rows: ACC, ARI, NMI; columns: fixed γ. Surfaces vary (σ, β); other coefficients are fixed within each sweep.

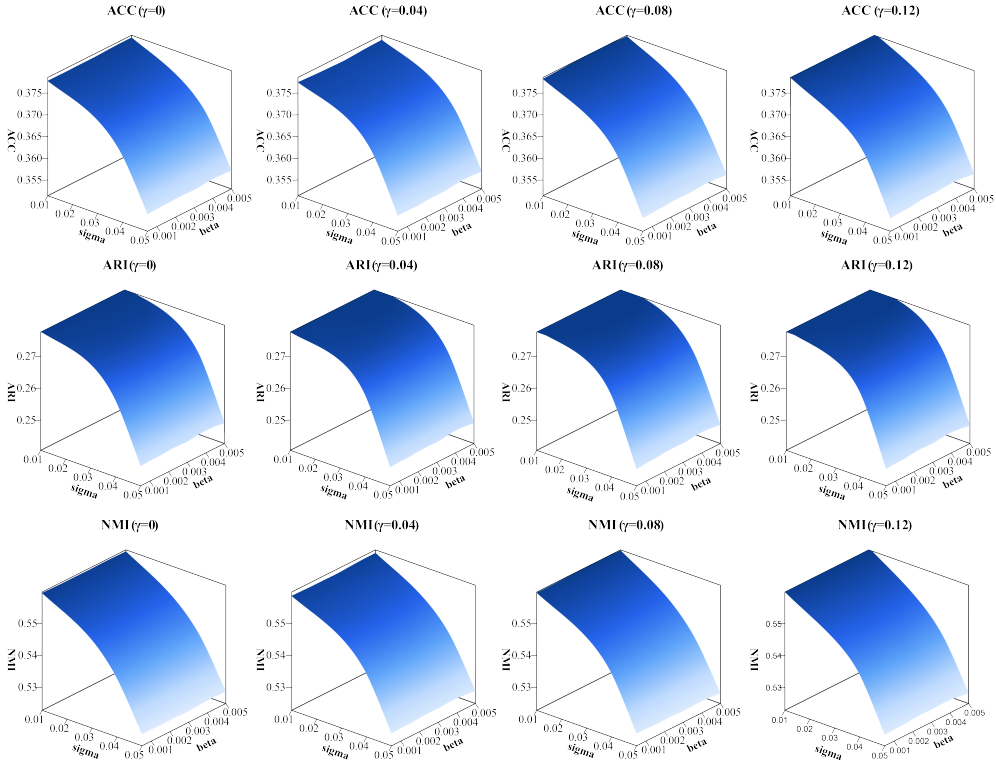

Figure 15: Sensitivity on SLN111. Rows: ACC, ARI, NMI; columns: fixed γ. Surfaces vary (σ, β); other coefficients are fixed within each sweep.

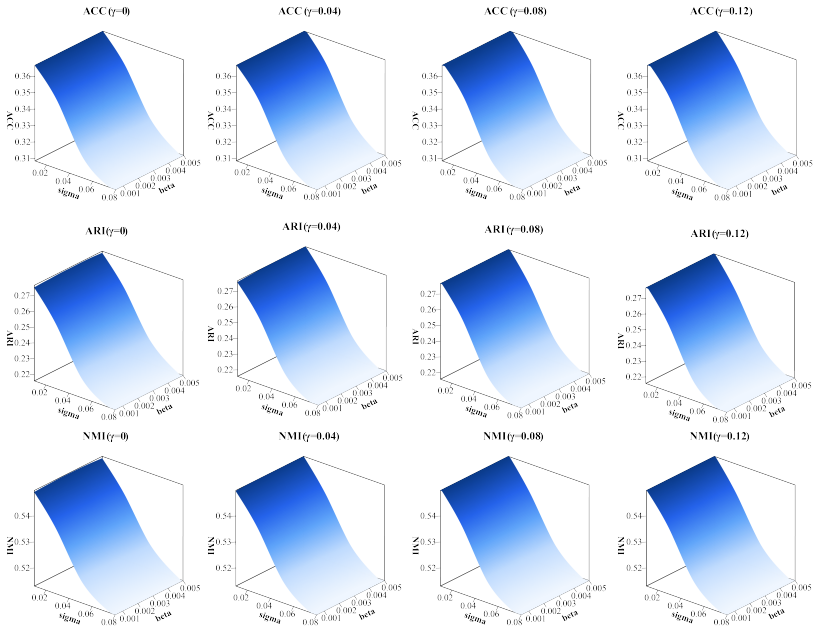

Figure 16: Sensitivity on SLN206. Rows: ACC, ARI, NMI; columns: fixed γ. Surfaces vary (σ, β); other coefficients are fixed within each sweep.

G Additional Representation and Biological Analyses

G.1 Integrated-Representation Visualization

We compare the learned representations on GSE128639. The raw input and most competing methods retain substantial mixing among related cell populations or produce fragmented structures. In the displayed projection, Trust-VC yields more distinct groups consistent with the reference annotations, including progenitor and dendritic populations. These two-dimensional visualizations provide qualitative context for the quantitative metrics in Table 1.

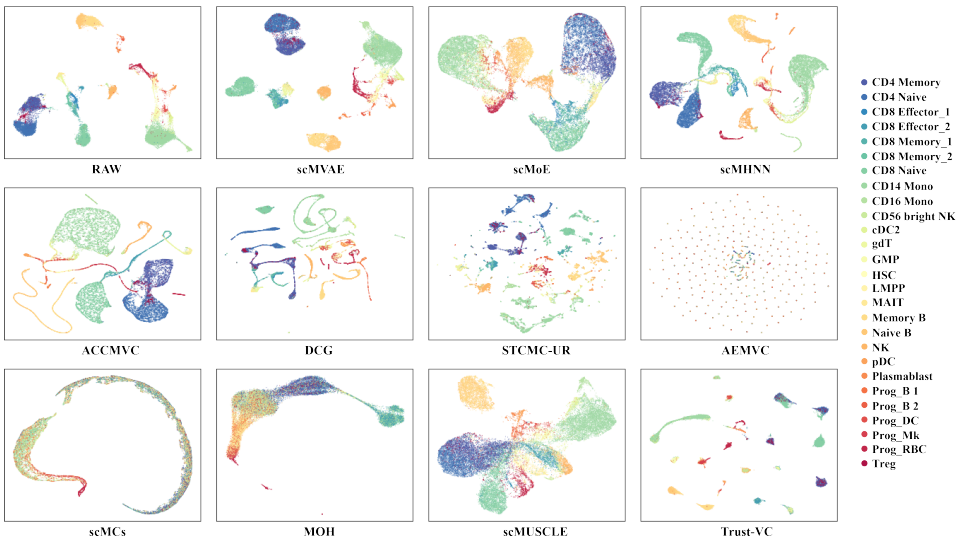

Figure 17: UMAP visualization of the raw input and integrated representations learned by Trust-VC and competing methods on GSE128639, colored by reference cell type.

G.2 Effect of the Virtual Prior

Figure 18 compares RNA, ADT, UCE, RNA+ADT without the virtual prior, and Trust-VC. On PBMC5K, measured-only and full representations retain recognizable immune populations; on GSE128639, Trust-VC shows more compact, distinct groups. These projections complement the ablations without establishing neighborhood preservation in the original latent space.

G.3 Biological Coherence of the Learned Clusters

On PBMC5K, cluster-specific RNA markers recover major immune lineages, and the inferred identities are supported by concordant ADT surface-protein profiles. Gene Ontology enrichment identifies lineage-associated immune processes, while ligand–receptor analysis suggests a predominantly myeloid-centered pattern of putative communication. These descriptive analyses support the biological coherence of the learned clusters; the inferred communication links are not direct measurements of cell–cell signaling.

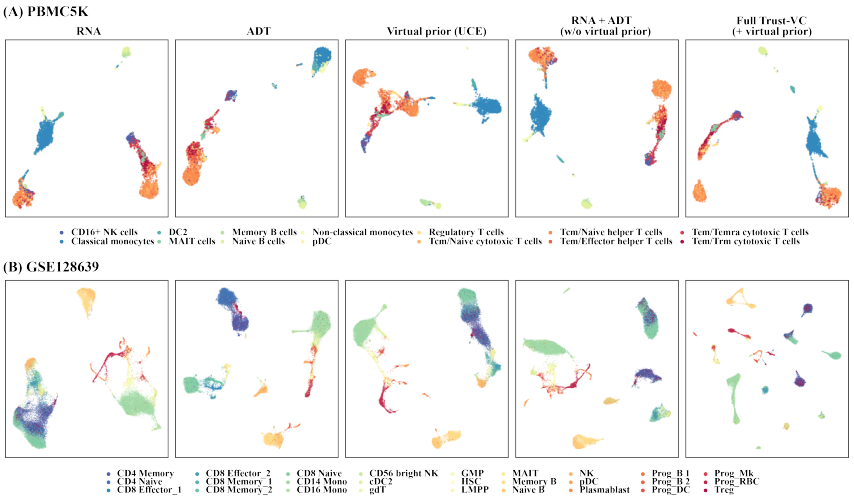

Figure 18: Effect of the virtual-cell prior on PBMC5K and GSE128639. Columns show RNA, ADT, the UCE virtual prior, RNA+ADT without the virtual prior, and full Trust-VC, colored by reference cell type.

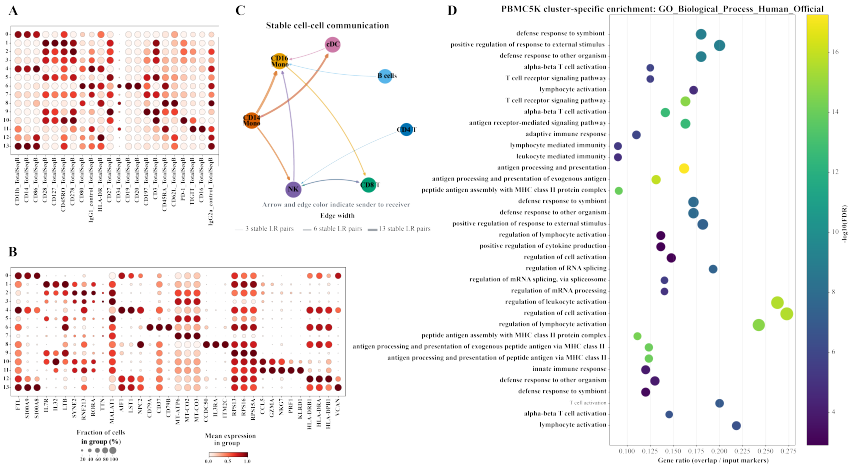

Figure 19: Biological characterization of Trust-VC clusters on PBMC5K. (a) Cluster-specific RNA markers. (b) ADT marker profiles. (c) Inferred putative ligand–receptor communication network. (d) GO Biological Process enrichment of cluster-specific RNA markers.