SCHEDOMICS: ADAPTIVE MODALITY SCHEDULING FOR SINGLE-CELL MULTI-OMICS INTEGRATION ACROSS PAIRING TOPOLOGIES

Anonymous authors
Paper under double-blind review

ABSTRACT

Single-cell multi-omics integration combines complementary molecular assays to characterize cellular heterogeneity beyond any single modality. However, existing methods often rely on predefined alignment and fusion policies that overlook modality-dependent learning dynamics. Under incomplete cell-level pairing and without cell-type supervision, such policies may prematurely incorporate underoptimized or unstable representations, compromising integration. To address this challenge, we propose SchedOmics, a self-supervised framework for adaptive modality scheduling across pairing topologies. First, mask-aware representation learning encodes observed assays and provides comparable representations for state estimation. Second, state-aware modality scheduling estimates remaining difficulty, stable learning progress, and temporal volatility to allocate bounded additional optimization, prioritizing reducible learning demand while discounting instability. Third, topology-aware cross-modal interaction uses these states to weight directional guidance according to target learning demand and source stability. Together, these components coordinate modality learning to exploit complementary information while limiting interference from unstable representations, enabling effective integration across pairing topologies. Across six datasets spanning three pairing settings, SchedOmics achieves consistently strong performance in label discrimination, clustering, and manifold preservation, with clear advantages in cross-assay transfer under incomplete pairing.

1 INTRODUCTION

Cell identity and function reflect molecular layers that no single assay fully captures. Single-cell multi-omics combines RNA expression, chromatin accessibility (ATAC), and cell-surface protein abundance measured by antibody-derived tags (ADTs), helping resolve cell populations and transitional states beyond individual modalities. However, observations may be fully paired, partially paired, or unpaired. Sparse or absent cell-level correspondence entangles differences in assay feature spaces, distributions, and noise with sample composition and missingness, complicating recovery of shared cell states while preserving assay-specific information. Cell-type annotation at scale is labor-intensive and requires marker expertise, while suitable labeled references may be unavailable for new tissues or conditions (Pliner et al., 2019; Geuenich et al., 2024). These constraints call for self-supervised integration directly from observed assays across pairing settings.

Existing methods emphasize assay-specific distribution modeling, shared latent representations, and cross-modal relations. For example, totalVI (Gayoso et al., 2021) jointly models RNA and protein counts, while MultiVI (Ashuach et al., 2023) uses assay-specific likelihoods for RNA and ATAC. Cobolt (Gong et al., 2021) connects modalities and batches through shared latent variables, scMoMaT (Zhang et al., 2023) uses matrix factorization for mosaic integration, and MIDAS (He et al., 2024) and mmAAVI (Rong et al., 2024) use generative models for diverse assay combinations and observation patterns. scGLUE (Cao & Gao, 2022) links transcriptomic and epigenomic features through a regulatory graph, whereas ACE (Yan et al., 2025) uses staged contrastive learning. Despite these advances, predefined joint optimization policies seldom track each assay's evolution or adapt its participation to underlearning, stable progress, or persistent fluctuations. A well-designed shared space or alignment objective alone does not address this limitation, and incomplete pairing further

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.

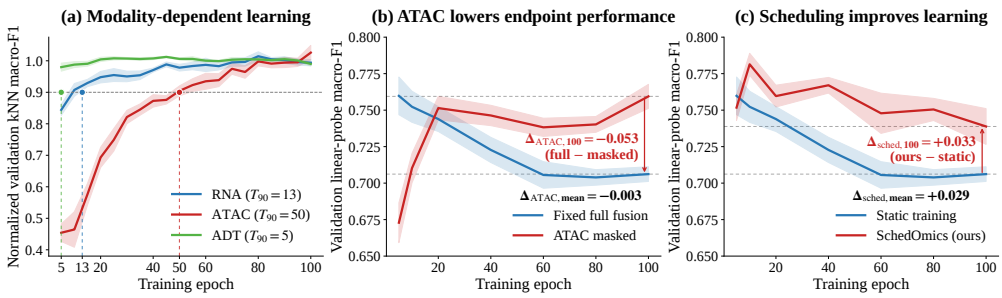

Figure 1: Controlled observations on DOGMA-seq. (a) Assay-specific validation kNN macro-F1 under static training, normalized to late-training performance. T_{90} denotes the fold-mean epoch reaching 90%. (b) Full fusion versus validation-time ATAC masking on identical frozen checkpoints ($\Delta_{\text{ATAC},100} = -0.053$). (c) The complete state-aware policy versus static training, with mean gains of 0.029 across folds and checkpoints and 0.033 at epoch 100. Panels (b) and (c) report validation linear-probe macro-F1. Lines and bands show five-fold means and standard errors.

obscures the distinction between learning-state variation and correspondence uncertainty. We therefore ask how a model can infer each assay's remaining learnable structure and stable progress from incomplete observations without cell-type supervision, and use these states to adapt within-assay optimization and cross-modal influence across pairing topologies.

To examine this issue, we conduct a controlled study on trimodal DOGMA-seq (Mimitou et al., 2021) using the same diagnostic backbone and evaluation settings. In Fig. 1(a), RNA, ATAC, and ADT reach 90% of their late-training validation performance after 13, 50, and 5 epochs, respectively, revealing unequal learning rates. In Fig. 1(b), retaining ATAC, the slowest assay to reach this threshold, lowers validation linear-probe macro-F1 by 0.053 at epoch 100 relative to masking it on the same frozen checkpoint. The changing fusion gap indicates stage-dependent modality contributions rather than a general disadvantage of ATAC. Fig. 1(c) shows that SchedOmics's complete state-aware policy improves the same readout by 0.033 over static training at epoch 100, supporting adaptive modality participation.

These observations motivate adapting *when and how assays participate, beyond designing a shared space or alignment objective.* This requires estimating learning states from unlabeled observed assays without fabricating missing measurements, distinguishing reducible demand from persistent instability, and translating these states into assay-specific updates. Cross-modal interaction must further separate optimization demand from source suitability and respect topology-supported relations, since an assay needing refinement may be an unsuitable source and learning states cannot justify unsupported interactions. We therefore seek a unified *state estimation, modality scheduling, and topology-constrained interaction* loop across pairing observations.

To this end, we propose SchedOmics, a self-supervised framework for adaptive modality scheduling in single-cell multi-omics integration across pairing topologies. First, mask-aware representation learning encodes observed assays and reconstructs them from a mask-aware joint representation, providing a common projected space for state estimation. Second, state-aware modality scheduling estimates remaining difficulty, stable learning progress, and temporal volatility from perturbed views over training, allocating bounded additional within-assay updates toward reducible learning demand while discounting instability. Third, topology-aware cross-modal interaction separates target demand from source suitability and restricts directional gradient influence to relations supported by cell-level co-observation or explicit feature priors. Our main contributions are as follows:

- We propose a unified self-supervised framework for single-cell multi-omics integration that adapts modality participation to evolving learning states across pairing topologies.
- We design a state-aware training method that separates optimization priority, source suitability, and cross-modal support to coordinate within-assay refinement, adaptive fusion, and directional interaction while discounting instability.
- We provide controlled evidence of modality-dependent learning dynamics and evaluate SchedOmics on six datasets, showing gains from the adaptive scheduling over static training, while transfer experiments demonstrate improved cross-assay organization under incomplete pairing.

2 Related Work

Likelihood-Based Generative Integration. Likelihood-based methods preserve assay-specific measurement characteristics while learning shared representations. totalVI (Gayoso et al., 2021) jointly models RNA and protein counts, while MultiVI (Ashuach et al., 2023) combines modality-specific likelihoods for RNA and ATAC. Cobolt (Gong et al., 2021) extends latent-variable integration to paired and single-modality profiles, and Multigrate (Lotfollahi et al., 2022) supports incomplete multi-view data via a generative framework. MIDAS (He et al., 2024) further combines a product of experts with self-supervised alignment and representation disentanglement. They provide flexible assay-aware modeling, but modality participation is generally governed by a predefined architecture and training objective.

Alignment, Factorization, and Mosaic Integration. Other methods focus on structural relations across assays. scGLUE (Cao & Gao, 2022) connects modality-specific autoencoders through a regulatory graph. MOFA+ (Argelaguet et al., 2020) learns sparse shared factors, while scMoMaT (Zhang et al., 2023) uses matrix tri-factorization for mosaic integration. mmAAVI (Rong et al., 2024) supports flexible assay combinations through hierarchical variational inference and attention, whereas ACE (Yan et al., 2025) combines contrastive alignment with modality completion. For unpaired data, scMRDR (Sun et al., 2025) learns regularized disentangled representations. Recent benchmarks (Liu et al., 2025; Wang et al., 2026) show that performance varies across datasets, modality combinations, and evaluation criteria, highlighting the importance of explicit input and pairing assumptions.

3 Method

3.1 Problem Definition

Single-cell multi-omics data vary in assay type and availability, so we distinguish observed measurements from supported cross-assay relations. Let $\mathcal{I} = \{1, \dots, N\}$ and $\mathcal{A} = \{1, \dots, M\}$ index cells and assays, respectively, with p_m features for assay m. The measurement $\mathbf{x}_i^m \in \mathbb{R}^{p_m}$ is observed iff $o_i^m = 1$ in the mask $\mathbf{O} \in \{0,1\}^{N \times M}$, and $\mathcal{A}_i = \{m \in \mathcal{A} \mid o_i^m = 1\}$ denotes the assays observed for cell i. We use *assay* and *modality* interchangeably. The mask $\mathbf{O}$ defines fully paired, connected-mosaic, or disjoint unpaired observations, but does not itself establish cross-assay support. For training cells $\mathcal{I}_{\text{train}} \subset \mathcal{I}$, we define

$$\mathcal{E}_{\text{sup}} = \mathcal{E}_{\text{obs}} \cup \mathcal{E}_{\text{prior}}, \qquad \mathcal{E}_{\text{obs}} = \{(m,n) \mid m \neq n, \textstyle\sum_{i \in \mathcal{I}_{\text{train}}} o_i^m o_i^n > 0\}, \tag{1}$$

where $\mathcal{E}_{\text{obs}}$ contains ordered assay pairs supported by cell-level co-observation and $\mathcal{E}_{\text{prior}}$ contains those supported by declared training-fold feature relations (Cao & Gao, 2022). We aim to learn $\mathbf{z}_i \in \mathbb{R}^{d_z}$ without treating missing assays as zeros or introducing relations outside $\mathcal{E}_{\text{sup}}$.

3.2 Overall Framework

Heterogeneous assays exhibit different learning dynamics, while incomplete pairing constrains cross-assay interactions. As shown in Fig. 2, SchedOmics addresses these challenges through three modules. First, mask-aware representation learning encodes and reconstructs only observed assays to learn a unified representation. Second, state-aware modality scheduling estimates assay learning states from perturbed views and allocates additional within-assay updates. Third, topology-aware cross-modal interaction uses these states to weight fusion and regulate directional guidance along relations supported by cell-level co-observation or explicit feature priors. Together, these modules coordinate assay-specific learning and supported cross-assay information sharing to exploit complementary signals while limiting interference from unstable representations.

3.3 Mask-Aware Representation Learning

An unobserved assay provides neither a valid input nor a reconstruction target, so fixed-modality pooling would treat absence as a value or fabricate a measurement. We therefore introduce mask-aware representation learning, which encodes and reconstructs only observed assays and fuses their projected representations over $\mathcal{A}_i$.

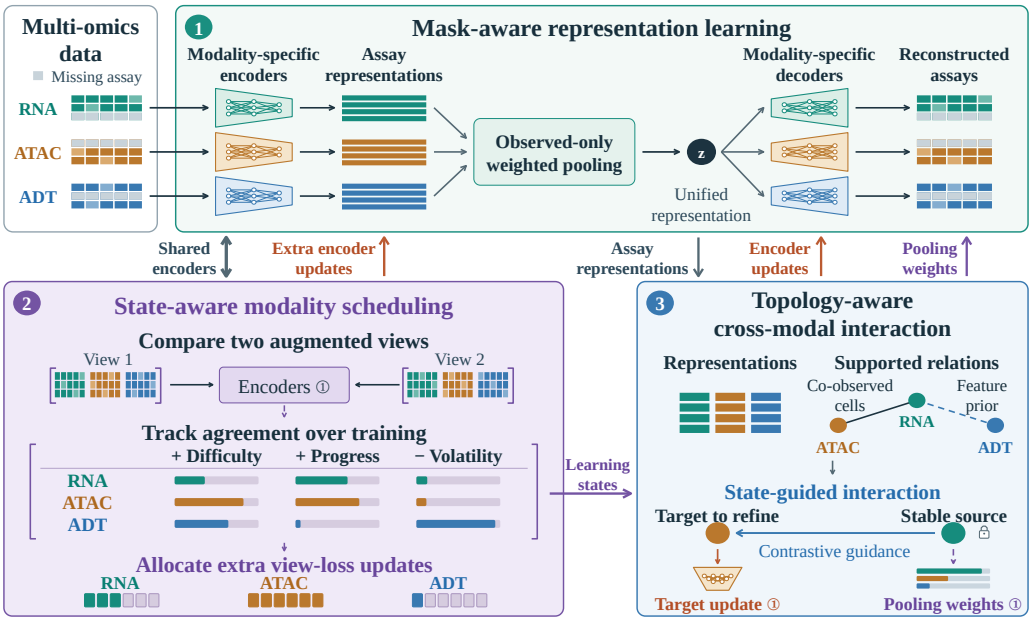

Figure 2: Overview of SchedOmics. Mask-aware representation learning processes observed assays and learns perturbation-consistent representations. State-aware scheduling estimates assay learning states and adaptively allocates additional updates. Topology-aware interaction uses supported cross-modal relations with state-dependent directional weighting to guide integration.

Encoding and fusing observed assays. Encoder E_m and projector P_m map each observed measurement to the shared dimension d_z, giving $\mathbf{s}_i^m = P_m(E_m(\mathbf{x}_i^m)) \in \mathbb{R}^{d_z}$. We use $\bar{\mathbf{s}}_i^m = \mathrm{norm}(\mathbf{s}_i^m)$, where $\mathrm{norm}(\mathbf{u}) = \mathbf{u}/\max(\|\mathbf{u}\|_2, \epsilon)$, where $\epsilon > 0$ is a numerical floor. The positive source-suitability score $w_m^{(t)}$, defined in Eq. (10), is fixed during epoch t and normalized over the assays available to each cell. The joint representation is formulated as follows:

$$\mathbf{h}_i = \mathrm{norm}\left(\sum_{m \in \mathcal{A}_i} \frac{w_m^{(t)}}{\max(\sum_{\ell \in \mathcal{A}_i} w_\ell^{(t)}, \epsilon)} \bar{\mathbf{s}}_i^m \right), \qquad \mathbf{z}_i = \mathrm{norm}(\mathbf{h}_i + R(\mathbf{h}_i)) \,. \tag{2}$$

Both $\mathbf{h}_i$ and $\mathbf{z}_i$ lie in $\mathbb{R}^{d_z}$, and the shared residual network R refines the pool. The positive coefficients sum to one over $\mathcal{A}_i$, so missing assays do not enter and a single observed assay receives coefficient one. Initialization gives every assay the same positive floor $w_m^{(1)} = \varepsilon_w$, so epoch-1 fusion is uniform; its small raw magnitude only keeps the unnormalized interaction conservative before a state is observed.

Reconstructing observed assays. A joint representation may discard assay-specific information, whereas a missing assay has no valid target. We therefore reconstruct only observed assays. For minibatch $\mathcal{B}$, let $\mathcal{B}_m = \{i \in \mathcal{B} \mid o_i^m = 1\}$ and $\mathcal{A}_\mathcal{B} = \{m \mid |\mathcal{B}_m| > 0\}$. Decoder D_m and its assay-appropriate negative log-likelihood ℓ_m define the reconstruction objective as follows:

$$\mathcal{L}_{\mathrm{rec}} = \frac{1}{|\mathcal{A}_\mathcal{B}|} \sum_{m \in \mathcal{A}_\mathcal{B}} \frac{1}{|\mathcal{B}_m|} \sum_{i \in \mathcal{B}_m} \ell_m(\mathbf{x}_i^m, D_m(\mathbf{z}_i)) \,. \tag{3}$$

The cell-level and assay-level averages prevent observation count from determining an assay's contribution. Likelihoods are given in Appendix A.1; the projected representations supply the comparable-space inputs used by the state estimator.

3.4 State-Aware Modality Scheduling

Fixed modality weights cannot capture different learning dynamics, while raw reconstruction losses are incomparable across assay likelihoods. We therefore introduce state-aware modality scheduling,

which estimates each assay's state from two perturbed views and allocates additional within-assay updates by its remaining learning demand. It uses only observed cells and requires neither cell-type labels nor comparable reconstruction-loss scales.

Estimating the learning state. We estimate each assay's learning state from representation consistency under controlled perturbations, providing a label-free signal in the common projected space. For each observed assay, two independent transformations $T_{m,1}$ and $T_{m,2}$ combine feature dropout with scale-adaptive Gaussian noise, yielding $\widetilde{\mathbf{x}}_{i,k}^m = T_{m,k}(\mathbf{x}_i^m)$ for $k \in \{1,2\}$. Both views share E_m and P_m. Let $\bar{\mathbf{S}}_k^m \in \mathbb{R}^{N_{m,\mathcal{B}} \times d_z}$ stack the normalized view embeddings of the $N_{m,\mathcal{B}} = |\mathcal{B}_m|$ observed cells. When $N_{m,\mathcal{B}} \geq 2$, ensuring at least one in-batch negative, we optimize

$$\mathcal{L}_{\text{view}}^m = \tfrac{1}{2}\left[\mathrm{NCE}_\tau(\bar{\mathbf{S}}_1^m, \bar{\mathbf{S}}_2^m) + \mathrm{NCE}_\tau(\bar{\mathbf{S}}_2^m, \bar{\mathbf{S}}_1^m)\right], \tag{4}$$

where $\tau > 0$ is the temperature and NCE_τ is defined in Appendix A.1. This nonnegative objective preserves cell identity across perturbations. Because its magnitude depends on the number and difficulty of in-batch negatives, it is not used as the state signal.

Instead, scheduling uses a detached linear CKA gap. Let $\mathbf{S}_k^{m,\mathrm{c}} \in \mathbb{R}^{N_{m,\mathcal{B}} \times d_z}$ denote the column-centered unnormalized view embeddings. We define

$$g_{m,\mathcal{B}} = 1 - \frac{\|(\mathbf{S}_1^{m,\mathrm{c}})^\top \mathbf{S}_2^{m,\mathrm{c}}\|_F^2}{\max(\|(\mathbf{S}_1^{m,\mathrm{c}})^\top \mathbf{S}_1^{m,\mathrm{c}}\|_F \|(\mathbf{S}_2^{m,\mathrm{c}})^\top \mathbf{S}_2^{m,\mathrm{c}}\|_F, \epsilon)}, \tag{5}$$

and clamp $g_{m,\mathcal{B}}$ to $[0,1]$. A larger gap indicates greater perturbation discrepancy and serves as a scale-independent proxy for remaining difficulty, not a direct measure of representation quality. Because the gap alone cannot distinguish reducible structure from irreducible assay noise, the scheduler also tracks learning progress and temporal volatility. Its next-epoch association is evaluated in Appendix C.3.

Tracking the state over time. Because minibatch gaps depend on sample composition, we aggregate detached gaps per epoch and track their level and temporal change. At the start of epoch t, the stored state $\mathbf{a}_m^{(t)} = (d_m^{(t)}, v_m^{(t)}, q_m^{(t)})$ contains remaining difficulty $d_m^{(t)} \in [0,1]$, temporal volatility $v_m^{(t)} \geq 0$, and stable learning progress $q_m^{(t)} \in [0,1]$. A flag $r_m^{(t)} \in \{0,1\}$ indicates whether assay m has a valid difficulty observation. We initialize

$$\mathbf{a}_m^{(1)} = (1,0,0), \qquad r_m^{(1)} = 0, \tag{6}$$

where $r_m^{(1)} = 0$ prevents the default difficulty from triggering adaptive refinement before observation.

Let $\mathfrak{B}_m^{(t)}$ denote the eligible minibatches for assay m in epoch t. Their mean gap $\widehat{d}_m^{(t)} \in [0,1]$ measures current difficulty. The first valid observation initializes the measured state directly. From the second onward, let $t_m^- < t$ be the most recent epoch with a valid observation and $[a]_+ = \max(a,0)$. We define

$$\widehat{d}_m^{(t)} = \frac{1}{|\mathfrak{B}_m^{(t)}|} \sum_{\mathcal{B} \in \mathfrak{B}_m^{(t)}} g_{m,\mathcal{B}}, \quad \delta_m^{(t)} = \frac{\widehat{d}_m^{(t_m^-)} - \widehat{d}_m^{(t)}}{\max(\widehat{d}_m^{(t_m^-)}, \varepsilon_d)}, \quad \widehat{q}_m^{(t)} = \frac{[\delta_m^{(t)}]_+}{\max(|\delta_m^{(t)}| + v_m^{(t)}, \epsilon)}. \tag{7}$$

Here $\delta_m^{(t)} > 0$ indicates decreasing difficulty, while $|\delta_m^{(t)}|$ measures change magnitude. The floor $\varepsilon_d > 0$ stabilizes the relative change when the previous gap is near zero. Thus $\widehat{q}_m^{(t)} \in [0,1]$ is positive only under improvement and is discounted by stored volatility.

After epoch t, the first valid observation initializes the state, while subsequent observations update it by EMA:

$$(\mathbf{a}_m^{(t+1)}, r_m^{(t+1)}) = \begin{cases} ((\widehat{d}_m^{(t)}, 0, 0), 1), & r_m^{(t)} = 0, \\ \left(\beta \mathbf{a}_m^{(t)} + (1-\beta)(\widehat{d}_m^{(t)}, |\delta_m^{(t)}|, \widehat{q}_m^{(t)}), 1\right), & r_m^{(t)} = 1, \end{cases} \tag{8}$$

where $\beta \in [0,1)$ controls smoothing and preserves the stated ranges of $d_m^{(t)}$, $v_m^{(t)}$, and $q_m^{(t)}$. If epoch 1 is eligible, $\mathbf{a}_m^{(2)} = (\widehat{d}_m^{(1)}, 0, 0)$, and the first temporal change is computed at the next valid observation and used only in the following epoch. If no eligible minibatch exists, $\mathbf{a}_m$ and r_m remain unchanged. Thus, all controls depend only on previously observed states.

How to schedule each assay. Because high perturbation discrepancy may reflect noise rather than learnable structure, we bound the allocation rate, strengthen it with stable progress, and discount temporal volatility:

$$\rho_m^{(t)} = r_m^{(t)} \min\left\{1, \frac{d_m^{(t)}(1+q_m^{(t)})}{1+v_m^{(t)}}\right\}. \tag{9}$$

Here $d_m^{(t)}$ prioritizes assays with greater remaining difficulty, $1+q_m^{(t)}$ strengthens refinement under stable progress, and $1+v_m^{(t)}$ discounts volatile states. Because progress is not a hard gate, remaining difficulty can still trigger refinement when recent progress is weak. The bounded rate $\rho_m^{(t)} \in [0,1]$ controls additional assay-specific view updates alongside the joint objective. Since $r_m^{(1)} = 0$, adaptive refinement starts only after a valid state is observed. Its deterministic realization is given in Appendix A.2, and sensitivity to the maximum additional rate is reported in Appendix C.3. This bounded monotone rule is a design choice rather than a uniquely derived estimator. The same state is next used to distinguish target demand from source suitability in cross-modal interaction.

3.5 Topology-Aware Cross-Modal Interaction

Learning demand alone cannot determine cross-modal influence. A difficult assay may be an unsuitable source, and no state can establish an unsupported relation. We therefore introduce topology-aware cross-modal interaction. The support graph first determines admissibility, after which the state determines source suitability, target demand, and directional strength. This differs from fixed symmetric alignment and prevents scheduling from creating unsupported edges.

State-guided paired interaction. For a supported co-observed direction $m \to n$, let $e_{mn}^{(t)} \in [0,1]$ denote the stored clipped cosine discrepancy. Using states fixed at the start of epoch t, we define

$$w_m^{(t)} = \max\left(\varepsilon_w, \frac{1-d_m^{(t)}}{1+v_m^{(t)}}\right), \qquad \omega_{mn}^{(t)} = w_m^{(t)} d_n^{(t)} (1-q_n^{(t)}) \exp(-e_{mn}^{(t)}). \tag{10}$$

Here $\varepsilon_w \in (0,1)$ is a small positive floor. The score $w_m^{(t)} \in [\varepsilon_w, 1]$ favors stable, low-difficulty assays and weights fusion in Eq. (2). The directional score $\omega_{mn}^{(t)} \in [0,1]$ combines target difficulty and paired consistency, with $1-q_n^{(t)}$ reducing assistance to targets already improving stably.

With at least two co-observed cells per minibatch, $\mathcal{L}_{m\to n}^{\text{pair}}$ denotes their compatibility-weighted contrastive retrieval loss. The same cell across assays provides the positive pair, while other co-observed cells provide negatives. Detaching the source and sample weights confines gradients from this loss to the target encoder and projector. We combine both directions as

$$\mathcal{L}_{mn}^{\text{obs},(t)} = \min(1, \omega_{mn}^{(t)} + \omega_{nm}^{(t)}) \frac{\omega_{mn}^{(t)} \mathcal{L}_{m\to n}^{\text{pair}} + \omega_{nm}^{(t)} \mathcal{L}_{n\to m}^{\text{pair}}}{\max(\omega_{mn}^{(t)} + \omega_{nm}^{(t)}, \epsilon)}. \tag{11}$$

The prefactor reduces total alignment strength when both scores are weak. Retrieval definitions, relation-state updates, and initialization are given in Appendix A.3.

Handling incomplete pairing. Paired and connected-mosaic data use only direct co-observation edges, never synthetic pairs from indirect paths. When co-observation and a feature prior coexist, only the direct loss is used. Disconnected assays require reciprocal genomic anchors constructed from an explicit training-fold feature prior. For these indirect relations, $\kappa_n^{(t)} = d_n^{(t)} q_n^{(t)} / (1+v_n^{(t)}) \in [0,1]$ measures stable reducible target demand. Unlike $1-q_n^{(t)}$ in paired assistance, $q_n^{(t)}$ gates state-dependent directional evidence, not the prior loss itself:

$$\mathcal{L}_{mn}^{\text{prior},(t)} = \frac{1}{M}\left(\pi_{mn}^{(t)} \mathcal{L}_{m\to n}^{\text{anchor}} + \pi_{nm}^{(t)} \mathcal{L}_{n\to m}^{\text{anchor}}\right). \tag{12}$$

Here $\mathcal{L}_{m\to n}^{\text{anchor}}$ is a confidence- and compatibility-weighted anchor-retrieval loss with its source and sample weights detached, active with at least two anchors per minibatch. The weights $\pi_{mn}^{(t)}$ combine source suitability, $\kappa_n^{(t)}$, and anchor confidence, with $\pi_{mn}^{(t)} + \pi_{nm}^{(t)} = 1$. A neutral residual preserves eligible prior interactions despite uncertain states. Appendix A.3 specifies anchor construction, bounded directional weighting, and initialization using unlabeled training cells only.

Table 1: Results on three dataset (mean±standard deviation; higher is better). Input lists assay modalities only; All denotes RNA+ATAC+ADT. **Bold**/underlined: best/second-best mean.

Data	Method	Input	Bal. Acc.	Macro-F1	kNN-F1	ARI	NMI	ASW	Graph
DOGMA-seq	ACE	All	0.758±0.025	0.737±0.011	0.731±0.021	0.109±0.006	0.272±0.007	-0.017±0.003	0.933±0.015
	Cobolt	All	0.799±0.047	0.790±0.025	0.719±0.021	0.127±0.018	0.298±0.019	0.091±0.015	0.948±0.024
	scGLUE	RNA+ATAC	0.810±0.058	0.749±0.037	0.652±0.025	0.127±0.029	0.265±0.038	0.041±0.034	0.943±0.042
	MIDAS	All	0.808±0.050	0.805±0.021	0.725±0.031	0.124±0.033	0.299±0.021	0.083±0.005	0.946±0.045
	mmAAVI	All	0.624±0.060	0.649±0.061	0.504±0.033	0.161±0.055	0.230±0.024	-0.174±0.245	0.727±0.039
	MOFA+	All	0.737±0.014	0.751±0.027	0.665±0.020	0.189±0.038	0.324±0.020	-0.031±0.104	0.912±0.019
	Multigrate	All	0.821±0.060	0.794±0.058	0.741±0.031	0.128±0.026	0.306±0.024	0.062±0.021	0.961±0.040
	MultiVI	RNA+ATAC	0.631±0.044	0.571±0.052	0.382±0.017	0.157±0.028	0.258±0.011	0.011±0.100	0.749±0.037
	scMoMaT	All	0.827±0.051	0.656±0.045	0.143±0.033	0.143±0.014	0.320±0.022	**0.092±0.053**	0.925±0.038
	scMRDR	All	0.729±0.086	0.709±0.057	0.545±0.040	0.087±0.018	0.242±0.031	0.016±0.025	0.876±0.040
	totalVI	RNA+ADT	0.826±0.060	0.764±0.026	0.692±0.034	0.146±0.020	0.321±0.032	0.077±0.014	0.932±0.040
	SchedOmics (ours)	All	**0.845±0.059**	**0.837±0.047**	**0.785±0.014**	**0.213±0.018**	**0.342±0.018**	0.091±0.079	**0.989±0.011**
BMMC Mosaic	ACE	All	0.418±0.028	0.418±0.028	0.375±0.063	0.221±0.049	0.593±0.046	-0.005±0.006	0.835±0.085
	scGLUE	RNA+ATAC	0.424±0.049	0.390±0.029	0.338±0.023	0.153±0.037	0.484±0.041	-0.082±0.028	0.788±0.022
	MIDAS	All	0.454±0.041	0.433±0.041	0.405±0.031	0.234±0.032	0.592±0.018	0.037±0.017	0.852±0.035
	mmAAVI	All	0.407±0.030	0.322±0.019	0.307±0.036	0.230±0.056	0.510±0.022	-0.220±0.093	0.664±0.060
	MOFA+	All	0.570±0.031	0.531±0.023	0.506±0.018	0.271±0.052	0.655±0.024	0.039±0.015	0.913±0.042
	Multigrate	All	0.531±0.048	0.482±0.013	0.439±0.015	0.263±0.040	0.606±0.033	-0.037±0.051	0.773±0.020
	scMoMaT	All	0.209±0.012	0.197±0.017	0.213±0.028	0.268±0.020	0.624±0.021	0.016±0.034	0.900±0.023
	scMRDR	All	0.539±0.046	0.497±0.021	0.402±0.023	0.247±0.044	0.609±0.030	-0.035±0.043	0.810±0.026
	totalVI	RNA+ADT	0.574±0.033	0.546±0.030	0.480±0.031	0.278±0.043	0.656±0.025	0.048±0.012	0.926±0.048
	SchedOmics (ours)	All	**0.576±0.041**	**0.548±0.023**	**0.513±0.025**	**0.282±0.053**	**0.668±0.013**	**0.052±0.020**	**0.945±0.010**
Muto-2021	Cobolt	RNA+ATAC	0.856±0.033	0.855±0.029	0.845±0.023	0.456±0.045	0.650±0.048	0.157±0.021	0.657±0.016
	scGLUE	RNA+ATAC	0.840±0.031	0.863±0.022	0.814±0.043	0.379±0.100	0.545±0.059	0.115±0.059	0.726±0.057
	MIDAS	RNA+ATAC	0.513±0.051	0.596±0.035	0.604±0.050	0.108±0.056	0.358±0.059	-0.145±0.088	0.556±0.055
	mmAAVI	RNA+ATAC	0.545±0.089	0.571±0.059	0.523±0.075	0.179±0.151	0.358±0.106	-0.176±0.066	0.561±0.068
	MOFA+	RNA+ATAC	0.763±0.028	0.875±0.030	0.873±0.031	0.393±0.149	0.589±0.049	0.174±0.024	0.647±0.035
	Multigrate	RNA+ATAC	0.087±0.016	0.040±0.010	0.062±0.007	0.003±0.008	0.032±0.019	-0.065±0.012	0.311±0.046
	MultiVI	RNA+ATAC	0.862±0.028	0.875±0.022	0.858±0.034	0.465±0.061	0.642±0.034	**0.196±0.036**	0.636±0.023
	scMoMaT	RNA+ATAC	0.483±0.065	0.541±0.055	0.064±0.009	0.150±0.054	0.438±0.052	-0.210±0.113	0.576±0.045
	scMRDR	RNA+ATAC	0.585±0.066	0.646±0.061	0.700±0.061	0.141±0.056	0.447±0.059	-0.114±0.104	0.623±0.045
	SchedOmics (ours)	RNA+ATAC	**0.884±0.035**	**0.895±0.025**	**0.888±0.026**	**0.476±0.090**	**0.882±0.034**	**0.196±0.037**	**0.921±0.024**

Objective and inference. Let $\mathcal{R}_{\mathcal{B}}^{\text{obs}}$ and $\mathcal{R}_{\mathcal{B}}^{\text{prior}}$ denote active unordered co-observed and prior-only assay pairs, respectively, with $K_{\mathcal{B}}$ their total number. The epoch-t base objective is

$$\mathcal{L}^{(t)} = \mathcal{L}_{\text{rec}} + \frac{1}{|\mathcal{A}_{\mathcal{B}}|} \sum_{m \in \mathcal{A}_{\mathcal{B}}} \mathcal{L}_{\text{view}}^{m} + \frac{\sum_{\{m,n\} \in \mathcal{R}_{\mathcal{B}}^{\text{obs}}} \mathcal{L}_{mn}^{\text{obs},(t)} + \sum_{\{m,n\} \in \mathcal{R}_{\mathcal{B}}^{\text{prior}}} \mathcal{L}_{mn}^{\text{prior},(t)}}{\max(1, K_{\mathcal{B}})}. \tag{13}$$

The averages stabilize scale as assay availability changes. Each minibatch receives one base update, with additional assay-specific view updates at rate $\rho_m^{(t)}$ as detailed in Appendix A.2. Thus, states control refinement, fusion, and directional strength, while $\mathcal{E}_{\text{sup}}$ determines admissibility. At inference, refinement is disabled and Equation (2) embeds observed assays using fixed terminal weights w_m. The label-free checkpoint rule is given in Appendix A.4.

4 Experiments

4.1 Experimental Setup

(1) Datasets: Six datasets span fully paired, connected-mosaic, and prior-linked unpaired settings: TEA-seq (Swanson et al., 2021), DOGMA-seq (Mimitou et al., 2021), PBMC 10k Multiome (10x Genomics, 2021), BMMC Mosaic (Luecken et al., 2021), Muto-2021 (Muto et al., 2021), and Yao-2021 (Yao et al., 2021). **(2) Baselines:** We compare eleven methods with native or explicitly adapted inputs: totalVI (Gayoso et al., 2021), MultiVI (Ashuach et al., 2023), Cobolt (Gong et al., 2021), Multigrate (Lotfollahi et al., 2022), scGLUE (Cao & Gao, 2022), MOFA+ (Argelaguet et al., 2020), scMoMaT (Zhang et al., 2023), MIDAS (He et al., 2024), mmAAVI (Rong et al., 2024), ACE (Yan et al., 2025), and scMRDR (Sun et al., 2025). Coverage is 11/11/10/9/9/9 in dataset order, determined by input compatibility rather than performance. Unsupported settings are marked N/A. **(3) Metrics:** Complementary metrics include linear-probe balanced accuracy and Macro-F1, 15-nearest-neighbor transfer Macro-F1, ARI, NMI, cell-type ASW, and graph connectivity (Luecken et al., 2022). **(4) Protocol:** Following prior work (Luecken et al., 2021; 2022; Liu et al., 2025), we use rotating outer test folds with a disjoint validation fold, keeping donor or batch groups intact when possible. Cell-type labels are used only for evaluation after freezing the embeddings. Training and implementation details are in Appendices B.1–B.4.

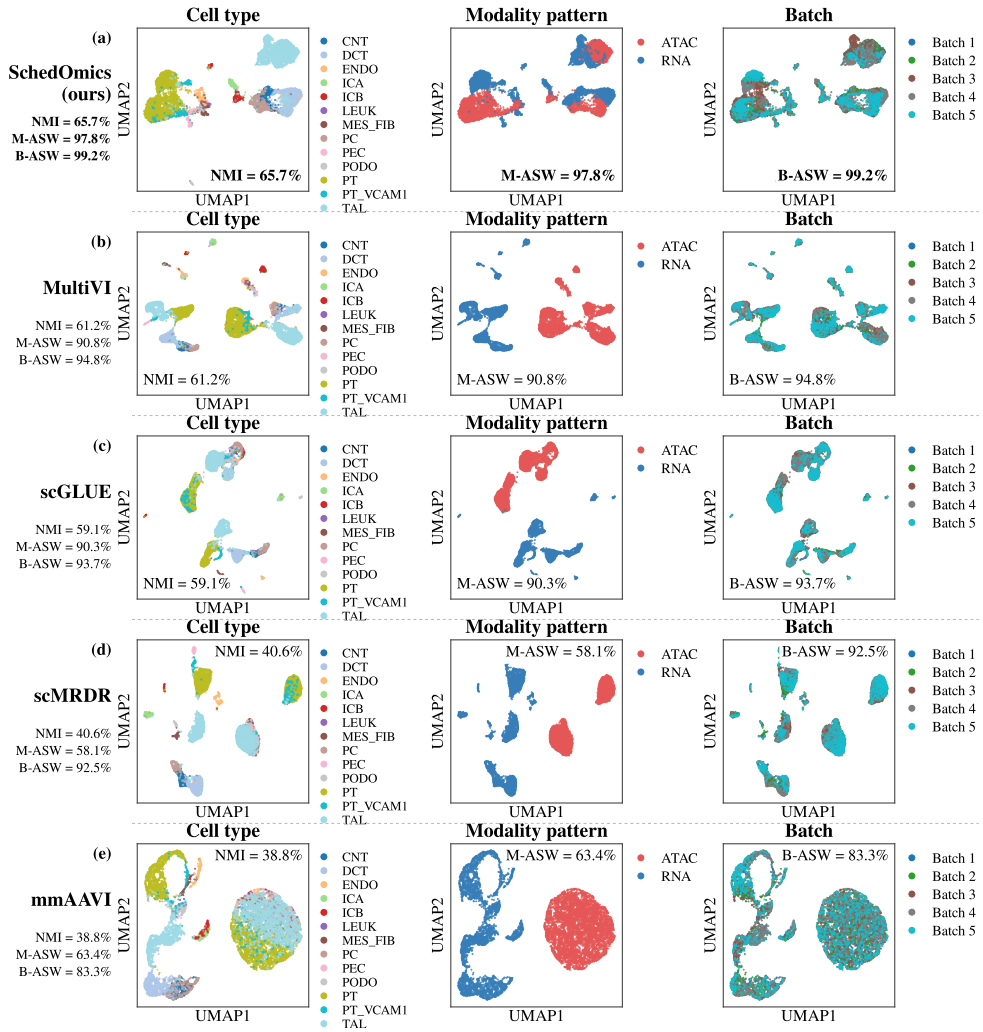

Figure 3: UMAP comparison on the unpaired Muto-2021. Columns are colored by cell type, modality pattern, and batch. Insets report pre-UMAP NMI, M-ASW, and B-ASW (higher is better).

4.2 Main Results Across Pairing Topologies

Table 1 reports results on DOGMA-seq (fully paired), BMMC Mosaic (connected mosaic), and Muto-2021 (prior-linked unpaired), with the remaining datasets in Appendix C.1. SchedOmics achieves the highest means on six of seven metrics on DOGMA-seq and Muto-2021 and all seven on BMMC Mosaic, also tying for the best ASW on Muto-2021. These results support consistently strong integration across pairing topologies, including incomplete pairing. Fig. 3 shows Muto-2021 embeddings, where RNA-only and ATAC-only cells lack cell-level correspondence. SchedOmics leads the displayed methods in NMI, M-ASW, and B-ASW, supporting better preservation of biological structure and reduced assay and batch separation. UMAP provides qualitative support, while conclusions rely primarily on quantitative outer-fold results. Additional visualizations appear in Appendix C.2.

4.3 Analysis of Learning-State Scheduling

Table 2 reports DOGMA-seq ablations of state-aware scheduling, fusion, and supported interaction. Scheduler-only improves five of seven metrics over Static, although additional updates prevent isolating the effect of allocation. Against compute-matched Uniform-extra, it achieves absolute gains of 1.9%, 0.9%, and 0.9% in mean ARI, NMI, and graph connectivity, respectively, supporting benefits to representation geometry. Adding fusion or supported interaction yields complementary gains, while Full achieves the highest means across all seven metrics. Together with the

Table 2: Compute-matched DOGMA-seq ablation over five outer folds (mean±standard deviation). **Bold** and underlined denote the best and second-best values.

Variant	Bal. Acc.	Macro-F1	*k*NN-F1	ARI	NMI	ASW	Graph
STATIC	0.771±0.072	0.722±0.013	0.684±0.023	0.158±0.039	0.289±0.042	0.039±0.089	0.898±0.034
UNIFORM-EXTRA	0.816±0.039	0.766±0.033	0.679±0.029	0.171±0.069	0.318±0.060	0.017±0.070	0.902±0.029
SCHEDULER-ONLY	0.797±0.050	0.768±0.021	0.680±0.026	0.190±0.055	0.327±0.055	0.013±0.055	0.911±0.037
SCHEDULE+FUSION	0.760±0.023	0.770±0.033	0.695±0.043	0.205±0.055	0.324±0.044	0.031±0.072	0.937±0.037
SCHEDULE+INTERACTION	0.762±0.042	0.774±0.044	0.687±0.015	0.169±0.036	0.311±0.032	0.039±0.089	0.920±0.047
FULL	**0.845**±0.059	**0.837**±0.047	**0.785**±0.014	**0.213**±0.018	**0.342**±0.018	**0.091**±0.079	**0.989**±0.011

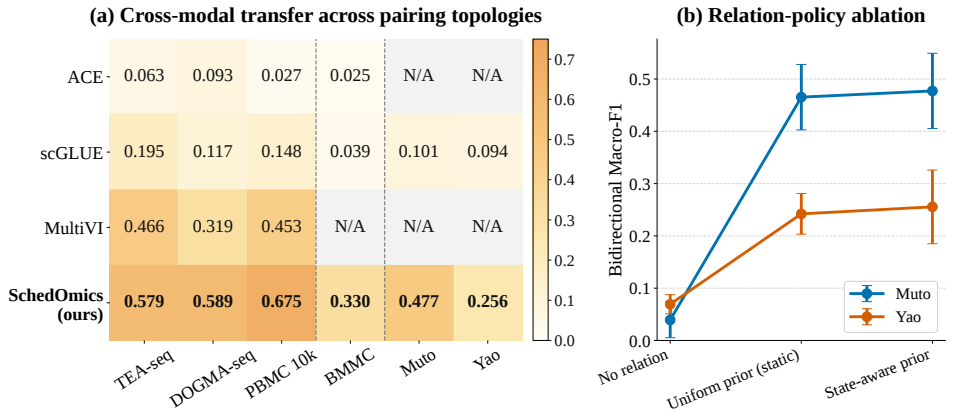

Figure 4: Cross-modal transfer across pairing topologies. (a) Mean bidirectional RNA–ATAC Macro-F1 across outer folds. (b) Matched relation-policy controls on prior-linked unpaired data (mean±standard deviation). N/A denotes unsupported topologies.

complete policy's favorable training trajectory in Fig. 1(c), these results support combining scheduling with both cross-modal mechanisms. Further results are in Appendices C.4 and C.3.

4.4 ANALYSIS OF TOPOLOGY-CONSTRAINED INTERACTION

We assess topology-constrained interaction through bidirectional RNA–ATAC label transfer in Fig. 4. A 15-nearest-neighbor classifier is fitted on source-assay outer-training cells and evaluated on target-assay outer-test cells. SchedOmics achieves the highest bidirectional mean among applicable methods on all six datasets, supporting strong cross-assay organization across pairing topologies. Absolute gains over the strongest applicable baseline range from 11.4% on TEA-seq to 37.6% on Muto-2021, including 29.1% on BMMC Mosaic. On prior-linked unpaired data, removing the genomic relation substantially reduces transfer, highlighting the importance of explicit support. State-aware directionality further improves transfer over static use of the same prior.

4.5 ADDITIONAL ROBUSTNESS AND SUPPLEMENTARY ANALYSES.

We further assess missing-assay robustness, resource use, and result stability in Appendix C.5. Under matched assay removal on BMMC Mosaic, SchedOmics achieves the highest means across all six metrics and four removal levels, although ACE shows slightly higher relative Macro-F1 retention at some levels. This supports stronger post-removal performance rather than uniformly lower relative degradation. Appendix C.7 reports parameter count, inference latency, and GPU memory, while Appendix C.6 examines molecular and biological preservation.

5 CONCLUSION

In this paper, we proposed SchedOmics that adapts modality participation to evolving learning states across pairing topologies. By separating optimization demand, source suitability, and interaction support, it coordinates within-assay refinement, fusion, and directional interaction. Experiments across six datasets demonstrate strong integration and cross-assay transfer performance. Future work will explore extensions to spatial and perturbational multi-omics.

REFERENCES

10x Genomics. 10k human pbmcs, multiome v1.0, chromium x. https://www.10xgenomics.com/datasets/10-k-human-pbm-cs-multiome-v-1-0-chromium-x-1-standard-2- 2021. Accessed 2026-09-11.

Ricard Argelaguet, Damien Arnol, Danila Bredikhin, Yonatan Deloro, Britta Velten, John C. Marioni, and Oliver Stegle. Mofa+: a statistical framework for comprehensive integration of multimodal single-cell data. *Genome Biology*, 21:111, 2020. doi: 10.1186/s13059-020-02015-1.

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*, 20:1222–1231, 2023. doi: 10.1038/s41592-023-01909-9.

Zhi-Jie Cao and Ge Gao. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nature Biotechnology*, 40:1458–1466, 2022. doi: 10.1038/s41587-022-01284-4.

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*, 18:272–282, 2021. doi: 10.1038/s41592-020-01050-x.

Michael J. Geuenich, Dae-won Gong, and Kieran R. Campbell. The impacts of active and self-supervised learning on efficient annotation of single-cell expression data. *Nature Communications*, 15:1014, 2024. doi: 10.1038/s41467-024-45198-y.

Boying Gong, Yun Zhou, and Elizabeth Purdom. Cobolt: integrative analysis of multimodal single-cell sequencing data. *Genome Biology*, 22:351, 2021. doi: 10.1186/s13059-021-02556-z.

Zhen He, Shuofeng Hu, Yaowen Chen, Sijing An, Jiahao Zhou, Runyan Liu, Junfeng Shi, Jing Wang, Guohua Dong, Jinhui Shi, Jiaxin Zhao, Le Ou-Yang, Yuan Zhu, Xiaochen Bo, and Xiaomin Ying. Mosaic integration and knowledge transfer of single-cell multimodal data with midas. *Nature Biotechnology*, 42:1594–1605, 2024. doi: 10.1038/s41587-023-02040-y.

Chunlei Liu, Sichang Ding, Hani Jieun Kim, Siqu Long, Di Xiao, Shila Ghazanfar, and Pengyi Yang. Multitask benchmarking of single-cell multimodal omics integration methods. *Nature Methods*, 22:2449–2460, 2025. doi: 10.1038/s41592-025-02856-3.

Mohammad Lotfollahi, Anastasia Litinetskaya, and Fabian J. Theis. Multigrate: single-cell multi-omic data integration. *bioRxiv*, 2022. doi: 10.1101/2022.03.16.484643.

Malte D. Luecken, Daniel B. Burkhardt, Robrecht Cannoodt, Christopher Lance, Aditi Agrawal, Hananeh Aliee, Ann T. Chen, Louise Deconinck, Angela M. Detweiler, Alejandro A. Granados, et al. A sandbox for prediction and integration of dna, rna, and proteins in single cells. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1, 2021.

Malte D. Luecken, Maren Büttner, Kridsadakorn Chaichoompu, Anna Danese, Marta Interlandi, Michaela F. Mueller, Daniel C. Strobl, Luke Zappia, Martin Dugas, Maria Colomé-Tatché, and Fabian J. Theis. Benchmarking atlas-level data integration in single-cell genomics. *Nature Methods*, 19:41–50, 2022. doi: 10.1038/s41592-021-01336-8.

Eleni P. Mimitou, Caleb A. Lareau, Kelvin Y. Chen, Andre L. Zorzetto-Fernandes, Yuhan Hao, Yusuke Takeshima, Wendy Luo, Tse-Shun Huang, Bertrand Z. Yeung, Efthymia Papalexi, Pratiksha I. Thakore, Tatsuya Kibayashi, James Badger Wing, Mayu Hata, Rahul Satija, Kristopher L. Nazor, Shimon Sakaguchi, Leif S. Ludwig, Vijay G. Sankaran, Aviv Regev, and Peter Smibert. Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nature Biotechnology*, 39(10):1246–1258, 2021. doi: 10.1038/s41587-021-00927-2.

Yoshiharu Muto, Parker C. Wilson, Nicolas Ledru, Haojia Wu, Henrik Dimke, Sushrut S. Waikar, and Benjamin D. Humphreys. Single cell transcriptional and chromatin accessibility profiling redefine cellular heterogeneity in the adult human kidney. *Nature Communications*, 12:2190, 2021. doi: 10.1038/s41467-021-22368-w.

Hannah A. Pliner, Jay Shendure, and Cole Trapnell. Supervised classification enables rapid annotation of cell atlases. *Nature Methods*, 16:983–986, 2019. doi: 10.1038/s41592-019-0535-3.

Zhiwei Rong, Jiali Song, Yipei Yu, Lan Mi, ManTang Qiu, Yuqin Song, and Yan Hou. Single-cell mosaic integration and cell state transfer with auto-scaling self-attention mechanism. *Briefings in Bioinformatics*, 25(6):bbae540, 2024. doi: 10.1093/bib/bbae540.

Jianle Sun, Chaoqi Liang, Ran Wei, Peng Zheng, Lei Bai, Wanli Ouyang, Hongliang Yan, and Peng Ye. scmrdr: A scalable and flexible framework for unpaired single-cell multi-omics data integration. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-5170. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/e34d214dfb5b80212ac1b565aa41af55-Abstract-Conference.html.

Elliott Swanson, Cara Lord, Julian Reading, Alexander T. Heubeck, Palak C. Genge, Zachary Thomson, Morgan D. A. Weiss, Xiao-jun Li, Adam K. Savage, Richard R. Green, Troy R. Torgerson, Thomas F. Bumol, Lucas T. Graybuck, and Peter J. Skene. Simultaneous trimodal single-cell measurement of transcripts, epitopes, and chromatin accessibility using tea-seq. *eLife*, 10:e63632, 2021. doi: 10.7554/eLife.63632.

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*, 17:5967, 2026. doi: 10.1038/s41467-026-72570-x.

Xuhua Yan, Jinmiao Chen, Ruiqing Zheng, and Min Li. Ace: A versatile contrastive learning framework for single-cell mosaic integration. *Genomics, Proteomics & Bioinformatics*, 23(4):qzaf062, 2025. doi: 10.1093/gpbjnl/qzaf062.

Zizhen Yao, Hanqing Liu, Fangming Xie, Stephan Fischer, Rachael S. Adkins, Andrew I. Aldridge, Seth A. Ament, et al. A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature*, 598:103–110, 2021. doi: 10.1038/s41586-021-03500-8.

Ziqi Zhang, Haoran Sun, Ragunathan Mariappan, Xi Chen, Xinyu Chen, Mika S. Jain, Mirjana Efremova, Sarah A. Teichmann, Vaibhav Rajan, and Xiuwei Zhang. scmomat jointly performs single cell mosaic integration and multi-modal bio-marker detection. *Nature Communications*, 14:384, 2023. doi: 10.1038/s41467-023-36066-2.