ATLAS: Adaptive Temporal Learning for Single-Cell Multi-Omics Alignment and Dynamics

Ye Zhang [1], **Zijie Fang** [2], **Zhixiang Lin** [1,3,4] ✉

[1] Department of Statistics and Data Science, The Chinese University of Hong Kong
[2] The Hong Kong University of Science and Technology, Hong Kong
[3] Shenzhen Loop Area Institute, Shenzhen, China
[4] CUHK Shenzhen Research Institute, Shenzhen, China

yezhang@cuhk.edu.hk, zfangay@connect.ust.hk, zhixianglin@cuhk.edu.hk

Abstract

Single-cell multi-omics technologies provide a powerful basis for characterizing cell-state transitions and dynamic regulatory processes. However, existing methods for multi-omics dynamic modeling still face two major limitations. From the data perspective, many methods require high-quality paired multi-omics measurements; from the modeling perspective, dynamic inference often relies on predefined regulatory structures or kinetic assumptions. These limitations restrict their applicability to partially paired, unpaired, and broader cross-omics settings. To address these challenges, we propose ATLAS, a unified framework for single-cell multi-omics alignment and dynamic modeling that jointly learns cross-omics consistency and temporal dynamics from partially paired or unpaired data. ATLAS adaptively models temporal lag effects between omics layers to characterize asynchronous cross-omics regulation, and further introduces a reliability-guided temporal distillation strategy to improve model-based temporal ordering. We systematically evaluate ATLAS on five datasets across four tasks, showing strong performance in multi-omics alignment, cross-omics prediction, trajectory inference, and future-state prediction. Our code is available at https://github.com/zhangye-zoe/ATLAS.

1 Introduction

Understanding how cells evolve over time is fundamental to reconstructing developmental trajectories and predicting future states (Traxler et al., 2023; Vinyard et al., 2025; Estévez et al., 2026). However, cellular transitions are governed by coordinated regulatory processes across multiple molecular layers rather than any single modality (Li et al., 2023; Deng et al., 2026). For example, chromatin accessibility reflects regulatory potential, while RNA expression captures downstream transcriptional activity, providing complementary yet temporally distinct perspectives. Therefore, jointly modeling these modalities is essential for understanding the temporal regulation underlying cell-state transitions.

Single-cell multi-omics technologies provide a powerful approach for studying cross-omics regulation by jointly profiling complementary molecular layers (Bravo González-Blas et al., 2023; Li et al., 2025b). However, obtaining fully paired measurements across modalities remains challenging in practice (Baysoy et al., 2023). Instead, many datasets contain only a subset of cells with multi-modal measurements, while other cells are observed in a single modality, leading to partially paired or fully unpaired settings (Sun et al., 2025; Tang et al., 2026). The absence of cell-level correspondence makes it difficult to infer cross-omics regulatory dynamics and limits existing approaches that rely on matched multi-omics observations (Li et al., 2023, 2025a; Baião et al., 2025).

✉ Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

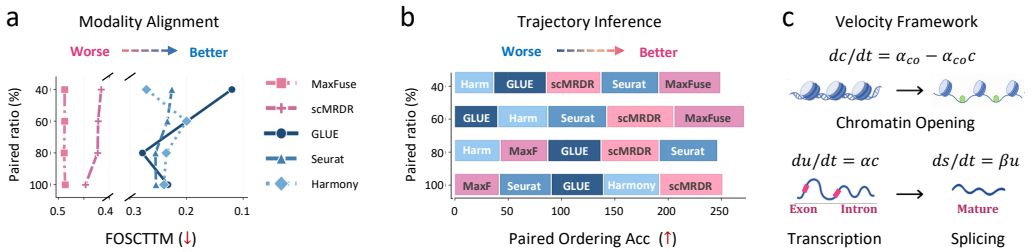

Figure 1: **Challenges in multi-omics dynamic modeling. a.** Cross-modal alignment performance across paired-data ratios, evaluated by FOSCTTM. **b.** Trajectory inference performance of methods, evaluated by paired ordering accuracy. The inconsistent rankings between a and b show that strong modality alignment does not necessarily preserve dynamic structure. **c.** Velocity-based models couple chromatin accessibility, transcription, and splicing through predefined kinetic assumptions.

To address incomplete cell-level correspondence, existing multi-omics methods typically align different modalities in a shared representation space using factor analysis (Lock et al., 2013), canonical correlation analysis (Singh et al., 2019), graph-based integration (Cao and Gao, 2022), or deep generative models (Sun et al., 2025), followed by downstream trajectory inference or fate prediction (Trapnell et al., 2014; Qiu et al., 2017). However, alignment and dynamic modeling are usually optimized separately, so strong cross-modal correspondence does not necessarily preserve the temporal structure required for dynamic analysis (Li et al., 2026). As shown in Fig. 1a,b, GLUE (Cao and Gao, 2022) and Harmony (Korsunsky et al., 2019) achieve stronger modality alignment but less accurate trajectory ordering than MaxFuse (Chen et al., 2024). This discrepancy suggests that cross-modal alignment and dynamic fidelity are complementary objectives that should be jointly optimized.

A complementary line of work models cellular dynamics through RNA velocity (Argelaguet et al., 2021; Gao et al., 2022; Li et al., 2023; Gayoso et al., 2024). Methods such as MultiVelo (Li et al., 2023) and MultiVeloVAE (Li et al., 2025a) use ODE-based kinetic models to couple chromatin accessibility, unspliced RNA, and spliced RNA, enabling the inference of cell-state transition directions and temporal progression. As shown in Fig. 1c, these models assume a predefined regulatory sequence from chromatin opening to transcription and splicing. However, their reliance on fixed kinetic assumptions limits their extension to other omics combinations with different regulatory relationships, while their dependence on sufficiently paired measurements restricts their applicability to incomplete multi-omics data. These limitations motivate a more flexible framework that jointly models cross-omics alignment and temporal dynamics while learning regulatory delays from data.

In this work, we propose **ATLAS**, an **A**daptive **T**emporal **L**earning for Single-Cell Multi-Omics **A**lignment and Dynamic**S**. The method learns shared representations that improve cross-modal consistency while preserving temporal structure for downstream dynamic analysis. To infer cellular progression without temporal annotations, the framework introduces a reliability-guided temporal distillation strategy that integrates unsupervised signals to learn latent time. Building on this temporal organization, it incorporates an adaptive lag module to capture context-dependent regulatory delays without imposing predefined kinetic forms. It supports fully paired, partially paired, and unpaired multi-omics data, bridging cross-modal alignment with dynamic regulatory modeling. The method achieves the best FOSCTTM across six PBMC and BMMC settings and the highest future-state AUROC on three HSPC benchmarks, while remaining competitive in cross-omics prediction and trajectory inference. Our main contributions are summarized as follows:

- We propose ATLAS, a unified framework that jointly models multi-omics alignment and cellular dynamics, enabling dynamic analysis under fully paired, partially paired, and unpaired settings.
- We develop a reliability-guided temporal distillation strategy that integrates complementary unsupervised temporal signals through reliability-aware fusion to learn robust latent cellular time, with theoretical analysis supporting the stability of the resulting temporal teacher.
- We introduce an adaptive cross-omics lag module that models temporal dependencies as learnable, context-dependent delays, enabling cell-specific characterization of asynchronous regulation across omics layers without relying on predefined kinetic forms.
- We conduct comprehensive experiments across four complementary tasks and five datasets, systematically evaluating cross-modal alignment, cross-omics prediction, trajectory inference, and future-state prediction under diverse pairing ratios and modality settings.

2 Related Work

Multi-omics Integration and Alignment. A substantial body of research has explored the integration of heterogeneous multi-omics modalities, particularly in partially paired or unpaired settings (Ma et al., 2022; Miao et al., 2025). Existing methods can be broadly categorized into three groups. The first integrates modalities through nearest-neighbor fusion or latent factor modeling, as exemplified by Seurat WNN (Hao et al., 2021) and MOFA+ (Argelaguet et al., 2020). The second learns shared latent representations using probabilistic or deep generative models, including totalVI (Gayoso et al., 2021), MultiVI (Ashuach et al., 2023), Cobolt (Gong et al., 2021), and Multigrate (Lotfollahi et al., 2022). The third aligns modalities through graph learning, matrix factorization, or bridge-based strategies, including GLUE (Cao and Gao, 2022), scJoint (Lin et al., 2022), uniPort (Cao et al., 2022), UINMF (Kriebel and Welch, 2022), StabMap (Ghazanfar et al., 2024), and scMoMaT (Zhang et al., 2023). Despite their effectiveness in static multi-omics integration, these methods generally lack explicit mechanisms for modeling dynamic cross-omics regulatory relationships.

Velocity-based Modeling of Cellular Dynamics. Velocity-based models infer cellular dynamics from single-cell snapshots. scVelo (Bergen et al., 2020) extends RNA velocity by explicitly modeling unspliced and spliced RNA dynamics, enabling the analysis of transient cellular states beyond steady-state assumptions. Subsequent methods further refine velocity estimation through unified temporal modeling in UniTVelo (Gao et al., 2022), cell-specific kinetics in DeepVelo (Cui et al., 2024), stochastic dynamics in SDEvelo (Liao et al., 2024), relay kinetics (Li et al., 2024), and trajectory-aware estimation in TIVelo (Ge et al., 2025). More recently, MultiVelo (Li et al., 2023) and MultiVeloVAE (Li et al., 2025a) extend velocity modeling to multi-omics settings by jointly modeling chromatin accessibility and RNA dynamics. Despite these advances, existing methods often rely on biological priors and predefined kinetic assumptions, while multi-omics extensions typically require paired measurements, limiting their applicability to partially paired or unpaired settings.

Our method bridges these two lines of research by jointly learning cross-omics alignment and time-lagged regulatory dynamics directly from heterogeneous multi-omics data, without relying on predefined biological priors or restrictive kinetic assumptions.

3 Methods

As illustrated in Fig. 2, our proposed framework comprises four components: variational representation learning, multi-level cross-modal alignment, reliability-guided temporal distillation, and adaptive cross-omics lag-time modeling. These components learn unified multi-omics representations, infer temporal dynamics, and capture asynchronous regulatory relationships.

3.1 Variational Representation Learning

Let $\mathbf{x}^{(m)}$ denote the input data from modality m. As illustrated in Fig. 2a, a modality-specific encoder maps the input into two complementary latent representations, consisting of a shared cell-state representation $\mathbf{z}_s^{(m)}$ and a private modality-specific representation $\mathbf{z}_p^{(m)}$, as follows.

$$q_m(\mathbf{z}_s^{(m)}, \mathbf{z}_p^{(m)}|\mathbf{x}^{(m)}) = q_m(\mathbf{z}_s^{(m)}|\mathbf{x}^{(m)})q_m(\mathbf{z}_p^{(m)}|\mathbf{x}^{(m)}). \tag{1}$$

Then, the shared representations from observed modalities are integrated through a mask-aware attention mechanism, where $\alpha^{(m)}$ denotes a learned modality-specific attention weight.

$$\mathbf{h} = \textstyle\sum_m \alpha^{(m)} \mathbf{z}_s^{(m)}, \quad \sum_m \alpha^{(m)} = 1. \tag{2}$$

To preserve modality-specific information, we reconstruct each observed modality from the fused representation and its corresponding private component: $\hat{\mathbf{x}}^{(m)} = G_m(\mathbf{h}, \mathbf{z}_p^{(m)})$, where G_m denotes the modality-specific decoder. The reconstruction loss is computed only for observed modalities:

$$\mathcal{L}_{\text{rec}} = \textstyle\sum_m \delta^{(m)} \ell_{\text{rec}}(\mathbf{x}^{(m)}, \hat{\mathbf{x}}^{(m)}), \tag{3}$$

where $\delta^{(m)} \in \{0, 1\}$ indicate whether the modality is observed. For paired observations, we apply a contrastive loss $\mathcal{L}_{\text{con}}$ to align shared representations across modalities. The overall objective is:

$$\mathcal{L}_{\text{VAE}} = \lambda_{\text{rec}}\mathcal{L}_{\text{rec}} + \lambda_{\text{con}}\mathcal{L}_{\text{con}} + \lambda_{\text{KL}}\mathcal{L}_{\text{KL}}. \tag{4}$$

where $\mathcal{L}_{\text{KL}}$ regularizes the latent variables, and λ_{rec}, λ_{con}, and λ_{KL} balance the three objectives.

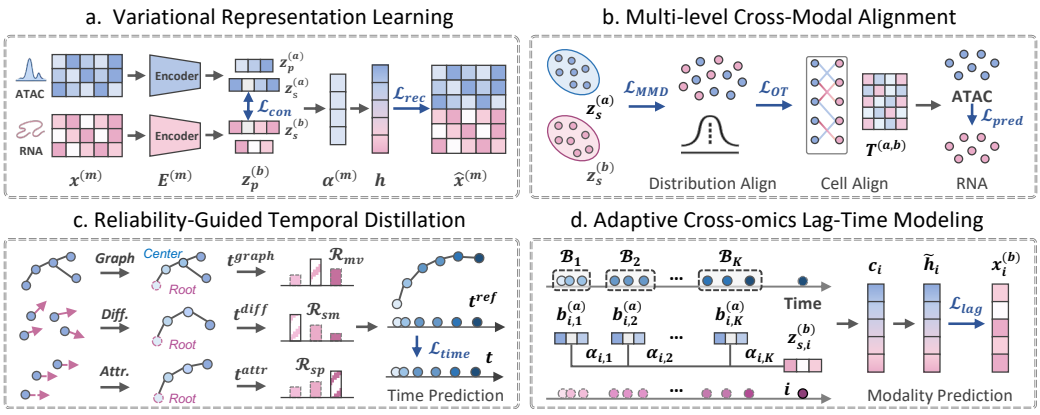

Figure 2: **Overview of the ATLAS framework. a.** Variational learning disentangles shared and modality-specific representations. **b.** Cross-modal alignment combines distribution- and cell-level matching. **c.** Reliability-guided distillation integrates temporal estimates for latent-time prediction. **d.** Time lag modeling captures asynchronous cross-omics dependencies through lag-aware prediction.

3.2 Multi-Level Cross-Modal Alignment

To reduce modality-specific discrepancies while preserving biological structure, we perform modality alignment at the distribution and cell levels, as shown in Fig. 2b. We first minimize the maximum mean discrepancy (MMD) between their shared latent distributions:

$$\mathcal{L}_{\mathrm{MMD}}^{(a,b)} = \mathrm{MMD}\big(\{\mathbf{z}_{s,i}^{(a)}\}_{i\in\mathcal{I}_a}, \{\mathbf{z}_{s,j}^{(b)}\}_{j\in\mathcal{I}_b}\big), \tag{5}$$

where $\mathcal{I}_a$ and $\mathcal{I}_b$ denote the sets of cells observed in modalities a and b, respectively. To encourage fine-grained cross-modal correspondence, we further introduce an optimal transport (OT) loss:

$$\mathcal{L}_{\mathrm{OT}}^{(a,b)} = \sum_{i,j} T_{ij}^{(a,b)} \left\| \mathbf{z}_{s,i}^{(a)} - \mathbf{z}_{s,j}^{(b)} \right\|_2^2, \tag{6}$$

where $T^{(a,b)}$ denotes the transport plan between modalities a and b in the shared latent space.

Following scMRDR (Sun et al., 2025), we further introduce a biological-structure conservation loss $\mathcal{L}_{\mathrm{bio}}$ that encourages the shared representation to preserve the pairwise geometry of the complete modality-specific latent representation. The overall alignment objective is defined as

$$\mathcal{L}_{\mathrm{align}} = \lambda_{\mathrm{MMD}}\mathcal{L}_{\mathrm{MMD}} + \lambda_{\mathrm{OT}}\mathcal{L}_{\mathrm{OT}} + \lambda_{\mathrm{bio}}\mathcal{L}_{\mathrm{bio}} + \lambda_{\mathrm{pred}}\mathcal{L}_{\mathrm{pred}}, \tag{7}$$

where λ_{MMD}, λ_{OT}, and λ_{bio} control the relative contributions of distribution-level, cell-level alignment, and biological-structure conservation, respectively.

3.3 Reliability-Guided Temporal Distillation

To support downstream dynamic modeling, we learn a continuous latent time representation for each cell, as shown in Fig. 2c. Since reliable temporal annotations are generally unavailable, we first construct multiple unsupervised temporal candidates $\mathcal{T} = \{\mathbf{t}^{\mathrm{graph}}, \mathbf{t}^{\mathrm{diff}}, \mathbf{t}^{\mathrm{attr}}\}$, which capture complementary graph-based, diffusion-based, and attribution-aware temporal signals, respectively.

To assess the reliability of cellular time estimates, we define a composite risk score $\mathcal{R}_k$ for each method $k \in \{\mathrm{graph}, \mathrm{diff}, \mathrm{attr}\}$, formulated as follows.

$$\mathcal{R}_k = \lambda_{\mathrm{sm}}\mathcal{R}_{\mathrm{sm}}\left(\mathbf{t}^{(k)}\right) + \lambda_{\mathrm{mv}}\mathcal{R}_{\mathrm{mv}}\left(\mathbf{t}^{(k)}\right) + \lambda_{\mathrm{sp}}\mathcal{R}_{\mathrm{sp}}\left(\mathbf{t}^{(k)}\right), \tag{8}$$

where $\mathbf{t}^{(k)}$ denotes the cellular time estimates produced by method k, with their relative contributions controlled by the corresponding weights λ_{sm}, λ_{mv}, and λ_{sp}. We convert the composite risk scores into normalized reliability weights, assigning higher weights to methods with lower risks:

$$w_k = \mathrm{softmax}_k\left(-\mathcal{R}_k/\tau_t\right). \tag{9}$$

where τ_t controls the sharpness of the weight distribution and each estimator's relative contribution.

To mitigate scale differences across methods, we aggregate their predictions in rank space with weights reflecting estimator reliability to obtain the final reference pseudotime:

$$\mathbf{t}^{\mathrm{ref}} = \textstyle\sum_k w_k \,\mathrm{Rank}\left(\mathbf{t}^{(k)}\right). \tag{10}$$

Building on the reliability-weighted reference pseudotime, we learn cell-specific latent times from the fused representations as $\mathbf{t} = \phi(\mathbf{h}) \in [0,1]^N$, where ϕ denotes a learnable time prediction head. To transfer the consensus temporal information into the model, we optimize the predicted latent time through the following distillation objective:

$$\mathcal{L}_{\mathrm{time}} = \left\|\mathbf{t} - \mathbf{t}^{\mathrm{ref}}\right\|_2^2 + \lambda_{\mathrm{reg}}\mathcal{L}_{\mathrm{reg}}, \tag{11}$$

where $\mathcal{L}_{\mathrm{reg}}$ encourages consistent temporal ordering and local smoothness while preventing degenerate time predictions, with λ_{reg} controlling its contribution.

3.4 Adaptive Cross-Omics Lag Modeling

Single-cell multi-omics signals often evolve asynchronously, with regulatory changes in one omics layer preceding responses in another. To capture such delayed dependencies, we introduces an adaptive lag modeling module for each directed cross-omics edge, as shown in Fig. 2d.

For each target cell i, ATLAS first constructs a source-modality lag bank $\mathbf{B}_i^{(a)}$ by aggregating information from cells with earlier inferred latent times:

$$\mathbf{B}_i^{(a)} = \left\{\mathbf{b}_{i,1}^{(a)}, \ldots, \mathbf{b}_{i,K}^{(a)}\right\}, \qquad \mathbf{b}_{i,k}^{(a)} = \mathrm{Pool}\left(\left\{\mathbf{z}_{s,j}^{(a)} \mid t_j < t_i,\ t_i - t_j \in \mathcal{B}_k\right\}\right), \tag{12}$$

where $\mathcal{B}_k$ denotes the k-th predefined temporal lag interval. Each bank element summarizes the source-modality representations of earlier cells within a specific temporal window.

ATLAS then applies an attention mechanism over the lag bank to obtain a lag-aware context:

$$\mathbf{c}_i = \sum_{k=1}^{K} \alpha_{i,k}\mathbf{b}_{i,k}^{(a)}, \qquad \alpha_{i,k} = \mathrm{softmax}_k\left(\psi\left(\mathbf{z}_{s,i}^{(b)}, \mathbf{b}_{i,k}^{(a)}\right)\right), \tag{13}$$

where ψ is a learnable scoring function. The attention weights determine the contributions of historical source states to the current target state. The lag-aware context is then combined with the current fused representation to obtain a lag-enhanced representation: $\tilde{\mathbf{h}}_i = \mathbf{U}\left(\mathbf{h}_i, \mathbf{c}_i\right)$. The lag module is optimized by predicting the target modality from the lag-enhanced representation:

$$\mathcal{L}_{\mathrm{lag}} = \mathcal{L}_p + \lambda_u\mathcal{L}_u + \lambda_{\mathrm{smooth}}\mathcal{L}_{\mathrm{smooth}}, \tag{14}$$

where $\mathcal{L}_p$ and $\mathcal{L}_u$ denote the paired and unpaired alignment losses, respectively. These terms encourage the lag-enhanced representation to capture the target-modality observations, while $\mathcal{L}_{\mathrm{smooth}}$ regularizes lag assignments across neighboring cells.

The learned attention weights $\alpha_{i,k}$ provide cell-specific and edge-specific estimates of cross-omics temporal delays, enabling interpretable characterization of asynchronous regulatory effects.

3.5 Training Objective and Progressive Optimization

The overall training objective integrates representation learning, cross-modal alignment, temporal distillation, and adaptive lag modeling:

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{VAE}} + \mathcal{L}_{\mathrm{align}} + \lambda_{\mathrm{time}}\mathcal{L}_{\mathrm{time}} + \lambda_{\mathrm{lag}}\mathcal{L}_{\mathrm{lag}}. \tag{15}$$

ATLAS is progressively optimized in three stages. First, the representation-learning module is trained to obtain stable latent representations. Next, the alignment loss is introduced to improve cross-modal consistency. Finally, temporal distillation and lag modeling are incorporated to capture cross-omics dependencies. This progressive training strategy improves optimization stability compared with joint optimization from the outset. Further implementation details are provided in Appendix A.

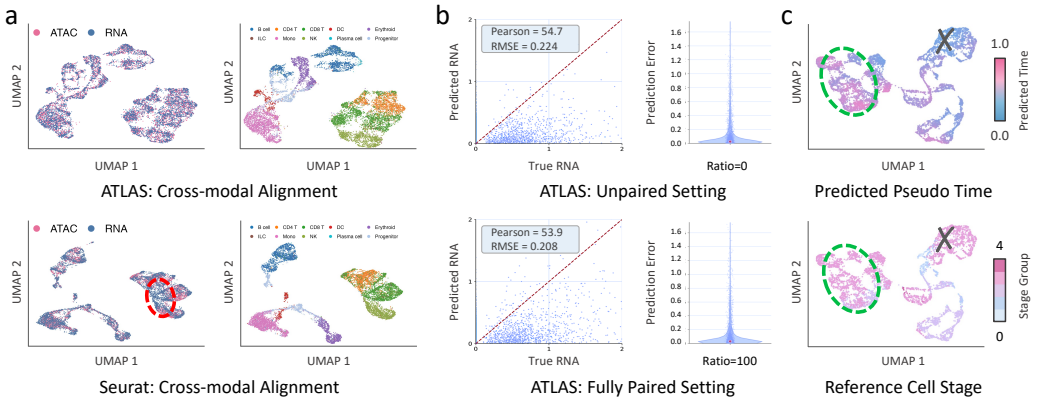

Figure 3: **Visualizations of multi-omics alignment, cross-omics prediction, and trajectory inference on BMMC dataset. a.** UMAP of ATLAS and Seurat cross-modal alignment under fully paired conditions. The red circle highlights residual modality misalignment in Seurat. **b.** Predicted versus ground-truth RNA expression and prediction error distributions under unpaired and fully paired settings. **c.** UMAPs colored by predicted pseudotime and reference cell stages. The green circles highlight consistency between inferred temporal progression and annotated stage ordering.

4 Experimental Setup

4.1 Task Settings and Evaluation Metrics

To comprehensively evaluate the proposed model in both Static Integration and Dynamic Modeling, we consider four tasks from two complementary perspectives. Specifically, static integration includes **T1:** *Cross-modal Alignment* and **T2:** *Cross-omics Prediction*, assessing cell-level modality correspondence and missing-omics recovery. Dynamic modeling includes **T3:** *Trajectory Inference* and **T4:** *Future State Prediction*, evaluating cellular state transitions and future cell-state prediction.

We further adopt task-specific metrics for quantitative evaluation. In detail, **T1** uses FOSCTTM and Paired Embedding Distance (PED); **T2** uses Pearson correlation and RMSE; **T3** uses Pairwise Ordering Accuracy (POA) and Pairwise Ordering AUC (POAUC); and **T4** uses F1 score and AUROC. Additional details of these metrics are provided in the Appendix B.

4.2 Datasets

We evaluate our method across the four tasks on five representative single-cell multi-omics datasets, as summarized in Table 1. **PBMC** (Zheng et al., 2017) is a 10x Genomics multiome dataset with paired RNA and ATAC measurements from peripheral blood mononuclear cells of healthy donor. **BMMC** (Luecken et al., 2021) is a large-scale benchmark dataset collected from 12 healthy donors for the NeurIPS 2021 multimodal single-cell integration challenge, comprising experimental subsets with RNA, ATAC, and protein measurements. These two datasets are used for the first three tasks to assess cross-omics alignment and trajectory inference performance.

Table 1: Dataset suitability across tasks.

Dataset	T1	T2	T3	T4
PBMC	✓	✓	✓	
BMMC	✓	✓	✓	
HSPC-2M			✓	✓
HSPC-MAC			✓	✓
HSPC-RNA			✓	✓

For dynamic evaluation, we follow MultiVeloVAE (Li et al., 2025a) and use three HSPC-centered fusion datasets released as post-processed AnnData objects. **HSPC-2M** combines two paired 10x Multiome HSPC samples for multi-sample hematopoietic differentiation. **HSPC-MAC** integrates HSPC and macrophage-induced samples to capture directed myeloid differentiation. **HSPC-RNA** extends HSPC-2M with a public RNA-only bone marrow dataset, providing a partially overlapping modality setting. Together, these datasets evaluate cross-omics temporal modeling, trajectory inference, and future-state prediction. Details are provided in the Appendix C.

4.3 Experimental Details

The baseline methods considered in this study rely on different software ecosystems and therefore require different execution environments. Specifically, Python/PyTorch-based methods are run on a

Table 2: **Performance comparison with static multi-omics integration methods under different pairing ratios.** All metrics are reported as percentages (%) except PED and RMSE. The 0%, 40%, and 100% settings correspond to unpaired, partially paired, and fully paired data, respectively.

Ratio	Model	T1: Cross-modal Alignment				T2: Cross-omics Prediction				T3: Trajectory Inference				Type
		PBMC		BMMC		PBMC		BMMC		PBMC		BMMC		
		FOS ↓	PED ↓	FOS ↓	PED ↓	Pear ↑	RMSE ↓	Pear ↑	RMSE ↓	POA ↑	PAUC ↑	POA ↑	PAUC ↑	
0%	GLUE†	29.1	12.6	17.5	**8.04**	62.2	0.236	45.4	0.243	61.8	84.9	56.2	83.1	△
	SeuratV5†	25.9	33.7	18.9	38.9	**68.6**	0.446	**57.1**	0.546	49.6	80.9	59.3	87.0	△
	scMRDR†	47.9	83.9	45.4	49.4	52.1	0.620	43.9	0.649	62.3	**88.6**	70.8	86.5	△
	MaxFuse†	48.7	185.7	48.2	100.2	30.3	1.744	31.0	2.047	46.3	82.3	57.5	85.8	○
	Harmony†	25.2	40.8	32.3	50.0	64.1	0.489	50.4	0.267	43.4	69.3	62.5	87.3	○
	ATLAS‡	**17.3**	**11.9**	**15.4**	8.91	65.7	**0.231**	54.7	**0.224**	**65.5**	85.0	**72.1**	**87.9**	–
40%	GLUE†	20.0	**10.2**	16.8	8.01	62.9	0.230	46.2	0.227	49.3	81.9	61.7	82.9	△
	SeuratV5†	23.4	32.5	17.8	38.6	**69.2**	0.442	57.3	0.533	54.2	84.1	63.0	**92.3**	△
	scMRDR†	45.2	84.1	45.0	52.8	52.8	0.768	43.9	0.751	58.2	87.2	76.2	84.8	△
	MaxFuse†	48.8	208.1	48.3	103.8	30.6	1.918	31.0	2.191	53.1	83.1	57.8	90.2	○
	Harmony†	20.0	42.7	28.6	47.7	63.1	0.488	51.7	0.244	37.3	78.8	**78.2**	91.3	○
	ATLAS‡	**4.56**	11.2	**5.69**	**5.50**	64.2	0.363	53.2	**0.209**	**62.7**	**87.7**	77.5	88.4	–
100%	GLUE†	12.7	**8.69**	15.2	7.76	64.3	**0.226**	46.7	0.227	49.1	79.6	63.6	87.7	△
	SeuratV5†	21.8	31.6	17.3	38.4	68.9	0.444	57.4	0.535	48.6	82.8	60.4	91.1	△
	scMRDR†	44.9	67.7	45.3	50.1	51.6	0.653	43.6	0.667	42.6	79.8	69.5	86.7	△
	MaxFuse†	48.7	202.8	48.2	96.4	30.9	1.795	31.0	1.911	**59.4**	**86.3**	61.4	88.6	○
	Harmony†	24.6	39.8	31.1	47.6	63.1	0.492	48.8	0.252	52.7	85.0	**72.5**	**92.5**	○
	ATLAS‡	**4.24**	10.2	**4.35**	**5.09**	64.6	0.361	53.9	**0.208**	56.3	83.7	70.3	86.9	–

† denotes trajectory inferred by Monocle3 (Cao et al., 2019) on learned embeddings; ‡ denotes trajectory inferred directly by ATLAS. FOS: FOSCTTM; Pear: Pearson correlation coefficient. △ denotes distribution-matching methods; ○ denotes cell-centric methods.

server with an NVIDIA RTX 3090 GPU, whereas R-based methods are executed on an Apple silicon device with 24 GB memory. Unless otherwise specified, ATLAS is trained for 180 epochs using a three-stage schedule of 80, 60, and 40 epochs, with a batch size of 256, a learning rate of 0.001, weight decay of 1×10^{-5}, and gradient clipping at a maximum norm of 5.0. The shared and private latent dimensions are set to 32 and 16, respectively, with a dropout rate of 0.1. The lag-time bank uses 4 lag bins with a 5-epoch warm-up. The reproduction codes are provided in the Appendix D.

5 Experimental Results

5.1 Comparison with Multi-omics Integration Methods

To evaluate the performance of ATLAS in multi-omics alignment and downstream dynamic analysis, we compare it with five representative multi-omics integration methods, including GLUE (Cao and Gao, 2022), SeuratV5 (Hao et al., 2024), scMRDR (Sun et al., 2025), MaxFuse (Chen et al., 2024), and Harmony (Korsunsky et al., 2019). To assess robustness to partial pairing, we evaluate all methods under different paired-data ratios by randomly masking paired observations. Throughout the tables, the best results are shown in **bold**, and the second-best results are underlined.

Cross-modal Alignment. We first assess cross-modal alignment using FOSCTTM and PED, as shown in Table 2. ATLAS achieves the best FOSCTTM scores on both PBMC and BMMC across all pairing settings, including the fully unpaired setting. On BMMC, for example, ATLAS obtains FOSCTTM scores of 15.4, 5.69, and 4.35 under 0%, 40%, and 100% pairing ratios, respectively. ATLAS also achieves the lowest PED in the partially paired and fully paired settings, and remains competitive under the unpaired setting. These results indicate that ATLAS learns accurate cell-level cross-modal correspondence even when paired measurements are unavailable or limited. Consistently, the UMAP visualizations in Fig. 3a show that ATLAS effectively integrates RNA and ATAC embeddings while preserving distinct cell-type structures. In contrast, Seurat exhibits residual modality-specific discrepancies in certain cell populations, as highlighted by the red dashed circle.

Cross-omics Prediction. We next evaluate cross-omics prediction by reconstructing RNA expression from ATAC profiles. As shown in Table 2, ATLAS achieves competitive prediction performance, yielding the lowest RMSE on BMMC across all pairing settings and on PBMC under fully unpaired training. These results suggest that the learned representations preserve cross-modal predictive infor-

Table 3: **Performance comparison with dynamic modeling methods.** Five representative methods are evaluated on trajectory inference and future state prediction across three HSPC datasets.

Model	T3: Trajectory Inference						T4: Future State Prediction						Omics
	HSPC-2M		HSPC-RNA		HSPC-MAC		HSPC-2M		HSPC-RNA		HSPC-MAC		
	POA ↑	POAUC ↑	POA ↑	POAUC ↑	POA ↑	POAUC ↑	F1 ↑	AUROC ↑	F1 ↑	AUROC ↑	F1 ↑	AUROC ↑	
MultiVeloVAE	0.717	0.738	0.682	0.699	**0.611**	0.613	0.589	0.670	0.590	0.652	0.600	0.691	◇
MultiVelo	0.803	0.833	0.383	0.372	0.406	0.405	0.623	0.679	0.579	0.637	0.658	0.726	◇
SDEvelo	0.215	0.188	0.216	0.177	0.480	0.429	0.575	0.654	0.605	0.663	**0.812**	0.781	▷
VeloVAE	0.783	0.812	0.670	0.692	0.448	0.418	0.589	0.667	0.610	0.649	0.725	0.721	▷
scVelo	0.792	0.821	0.665	0.687	0.440	0.425	0.580	0.663	**0.613**	0.653	0.683	0.711	▷
ATLAS	**0.859**	**0.885**	**0.701**	**0.717**	0.610	**0.639**	**0.637**	**0.851**	0.602	**0.770**	0.769	**0.792**	◇

◇ denotes methods based on RNA+ATAC multi-omics data, whereas ▷ denotes methods based on RNA-only data.

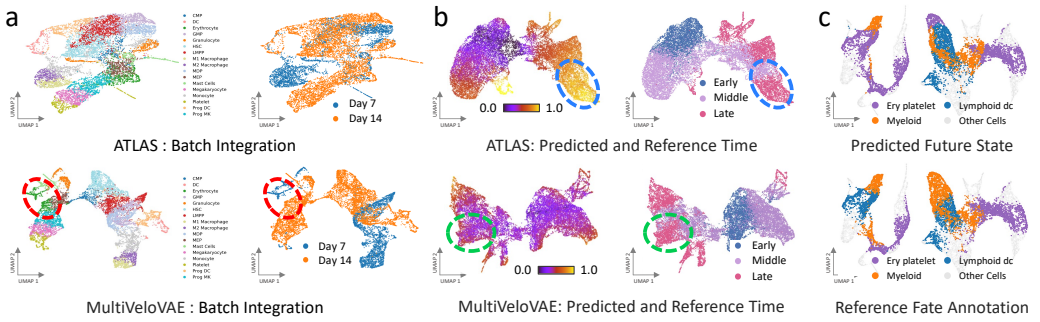

Figure 4: **Dynamic modeling comparison between ATLAS and MultiVeloVAE. a.** UMAP visualizations of cell-type distributions and batch integration. The red dashed circles highlight residual batch-specific separation in MultiVeloVAE. **b.** Predicted latent time versus reference developmental stages. The blue dashed circles highlight consistent temporal progression in ATLAS, whereas green circles indicate localized discrepancies in MultiVeloVAE. **c.** Comparison of predicted future cell states with reference fate annotations, showing their correspondence across different cell populations.

mation even without paired observations. Although Seurat achieves higher Pearson correlations in several settings, ATLAS consistently maintains lower reconstruction errors on BMMC. As illustrated in Fig. 3b, ATLAS predictions exhibit a positive association with ground-truth RNA expression under both unpaired and fully paired settings. The corresponding error distributions are concentrated near zero, with comparable Pearson correlations and slightly lower RMSE under full pairing.

Trajectory Inference. We further assess whether the learned representations support dynamic analysis. Baseline trajectories are inferred by applying Monocle3 (Cao et al., 2019) to the learned embeddings, whereas ATLAS directly estimates latent cellular time. As summarized in Table 2, ATLAS achieves strong trajectory performance across pairing settings, including the best POA on both PBMC and BMMC in the unpaired setting and the highest BMMC PAUC at 0%, demonstrating robust temporal ordering even without paired training observations. Consistently, the UMAP visualizations in Fig. 3c show that the inferred pseudotime largely follows the reference developmental structure, with the green-circled region highlighting a clear agreement between predicted temporal progression and annotated stage ordering. Some local inconsistencies remain, which is expected because the stage annotations are relatively coarse and the distribution of annotated cell groups is highly imbalanced.

5.2 Comparison with Dynamic Modeling Methods

We further compare ATLAS with representative dynamic modeling methods, including multi-omics velocity models MultiVeloVAE (Li et al., 2025a) and MultiVelo (Li et al., 2023), as well as RNA-only velocity models SDEvelo (Liao et al., 2024), VeloVAE (Gu et al., 2022), and scVelo (Weiler et al., 2024). We evaluate all methods on three HSPC-centered datasets, covering multi-sample integration, RNA-only modality transfer, and macrophage induction settings.

Trajectory Inference. We first evaluate whether methods recover the temporal ordering of cell-state transitions. As shown in Table 3, ATLAS achieves the best trajectory inference performance on HSPC-2M and HSPC-RNA, demonstrating robust temporal ordering across multi-sample and partially

Table 4: **Quantitative validation and robustness of adaptive lag modeling.** I. Temporal-reference agreement, II. sensitivity to lag-bin resolution, and III. branch-aware lag-bank construction.

I. Temporal-reference agreement				**II. Lag-resolution sensitivity**				**III. Branch-aware lag bank**			
Method	ρ	**NMAE**	**Agr.**	**Bins**	**Wt. corr.**	**Lag MAE**	**AUROC**	**Lag bank**	**F1**	**AUROC**	**Mix.**
Random	0.03	0.241	50.2	2	0.84 ± 0.04	0.121	$\mathbf{0.851} \pm 0.021$	**Time only**	0.621	0.842	0.187
Global lag	0.21	0.183	61.4	**4**	$\mathbf{0.89 \pm 0.03}$	**0.091**	$\mathbf{0.851} \pm 0.016$	Time + kNN	0.635	0.851	0.112
Win+kNN	0.58	0.118	80.6	6	0.86 ± 0.03	0.097	0.845 ± 0.014	Time + graph	0.641	**0.862**	**0.069**
ATLAS	**0.64**	**0.096**	**84.1**	8	0.80 ± 0.04	0.108	0.843 ± 0.019	Graph-aware	**0.646**	0.858	0.074

overlapping modality settings. The UMAP visualizations in Fig. 4a show that ATLAS effectively integrates batches while preserving distinct cell-type structures, whereas MultiVeloVAE exhibits residual batch-specific separation, as highlighted by the red dashed circles. Furthermore, Fig. 4b shows that ATLAS infers latent time consistent with reference developmental stages. Specifically, cells annotated as Late exhibit inferred latent time approaching 1.0 in ATLAS, as highlighted by the blue dashed circle, whereas MultiVeloVAE assigns values around 0.5 to corresponding cells, as indicated by the green dashed circle. This contrast demonstrates stronger agreement between the inferred temporal progression and reference stage ordering in ATLAS.

Future State Prediction. We next evaluate future-state prediction, which assesses whether inferred dynamics can predict cell fates. As shown in Table 3, ATLAS achieves the best AUROC across all three datasets. Although RNA-only methods such as scVelo, VeloVAE, and SDEvelo obtain strong F1 scores in some settings, their trajectory inference performance is less stable, suggesting that accurate fate classification does not necessarily imply reliable temporal ordering. The UMAP visualizations in Fig. 4c show that the predicted future states broadly agree with reference fate annotations. Specifically, ATLAS largely recovers the spatial distributions of Early platelet, Myeloid, and Lymphoid DC populations, with some localized discrepancies in regions containing mixed cell populations. These results demonstrate ATLAS supports both reliable temporal ordering and accurate future-state prediction, highlighting its ability to capture biologically meaningful cellular dynamics.

5.3 Lag-Time Interpretability Analysis

To evaluate the interpretability and robustness of adaptive lag modeling, three experiments examine temporal-reference agreement, lag-resolution sensitivity, and branch-aware lag-bank construction. These analyses investigate whether ATLAS captures meaningful temporal associations, maintains stable lag estimates, and preserves developmental trajectory structure.

Temporal-Reference Agreement. Independent temporal references are derived from unspliced–spliced peak timing and cross-correlation in HSPC-2M, and ATAC–RNA temporal ordering in HSPC-MAC. As shown in Table 4, Spearman correlation measures agreement in lag rankings, normalized mean absolute error (NMAE) quantifies prediction errors, and directional agreement (Agr.) evaluates temporal ordering consistency. ATLAS achieves the highest correlation of 0.64, lowest NMAE of 0.096, and strongest directional agreement of 84.1%, supporting the temporal consistency of learned lags across both datasets.

Lag-Resolution Sensitivity. Different numbers of lag bins are evaluated to assess the robustness of temporal discretization. Weight correlation measures the consistency of learned lag weights, while lag mean absolute error quantifies deviations from reference delays. Four bins achieve the highest weight correlation of 0.89 and lowest lag MAE of 0.091, while maintaining an area under the AUROC of 0.851. Higher resolutions provide no consistent improvement, suggesting that moderate lag discretization better balances temporal accuracy and predictive performance.

Branch-Aware Lag-Bank Construction. Different lag-bank strategies are compared to assess whether incorporating trajectory structure improves lag modeling. F1 measures the balance between precision and recall, AUROC evaluates predictive discrimination, and cross-branch mixing measures the inclusion of historical states from different developmental branches. Graph-aware achieves the highest F1 of 0.646, while Time + graph obtains the highest AUROC of 0.862 and lowest branch mixing of 0.069. These results highlight the benefits of graph constraints for reducing cross-branch interference. Additional comparisons are provided in Appendix E.

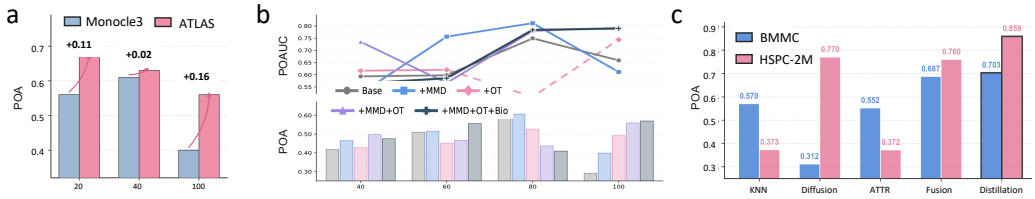

Figure 5: **Ablation study of model designs. a.** Comparison between Monocle3 trajectory inference on ATLAS embeddings and ATLAS direct latent-time inference. **b.** Effects of different alignment settings across masking ratios. **c.** Quantitative comparison of latent-time inference strategies.

5.4 Ablation Study

We conduct ablation studies to evaluate key ATLAS components. As shown in Fig. 5a, direct latent-time inference outperforms Monocle3 applied to ATLAS embeddings, improving POA by 0.11, 0.02, and 0.16 across three settings. These gains demonstrate the benefit of explicit temporal learning beyond embedding-based trajectory inference. In Fig. 5b, MMD and OT contribute differently to temporal performance across masking ratios, while biological-structure conservation improves robustness under severe masking, highlighting the importance of balancing cross-modal alignment and structural preservation. Fig. 5c shows that reliability-guided distillation achieves the highest POA on both BMMC and HSPC-2M, reaching 0.703 and 0.859, respectively, demonstrating the benefit of integrating complementary temporal signals. Finally, Table 5 shows that adaptive lag modeling improves future-state prediction on HSPC-2M, increasing F1 from 0.609 to 0.637. These results confirm that modeling asynchronous cross-omics dependencies provides information for predicting cellular transitions. Further ablation experiments appear in Appendix F.

Table 5: Lag ablation study.

Model	F1	AUROC
w/o lag	0.609	0.817
with lag	**0.637**	**0.851**

6 Discussion

Conclusion. We present ATLAS, a unified framework for single-cell multi-omics alignment and dynamic modeling. By combining reliability-guided latent-time learning with adaptive cross-omics lag modeling, ATLAS bridges static integration and dynamic analysis. Results show that strong cross-modal alignment does not guarantee accurate temporal inference, while adaptive lag modeling captures asynchronous dependencies associated with cell-state transitions. These findings highlight the importance of jointly modeling cross-modal correspondence and temporal dynamics.

Limitations. First, fully unpaired multi-omics alignment remains challenging without cell-level correspondence, and ATLAS remains sensitive to latent-time accuracy, potentially introducing training instability. Second, latent-time estimates vary across datasets. Integrating multiple estimators improves reliability but increases complexity and computational overhead, motivating simpler end-to-end temporal learning. Finally, learned lags reflect temporal associations rather than causal regulatory effects. Establishing causality remains difficult because regulatory events can occur within seconds to minutes, whereas longitudinal multi-omics measurements are often collected days or weeks apart.

Outlook. We are developing an end-to-end framework with biologically informed developmental-stage constraints to improve latent-time accuracy and stability while reducing reliance on multiple estimators. Interdisciplinary collaborations will provide temporally resolved multi-omics and perturbation data to validate inferred regulatory delays. Future work will extend asynchronous cross-omics modeling to disease research to uncover disrupted regulatory programs.

Acknowledgments

This work was supported in part by research grants awarded through the 1+1+1 CUHK-CUHK(SZ)-GDSTC Joint Collaboration Fund (Grant No. 2025A0505000057 to Z.L.), the Research Grants Council of the Hong Kong Special Administrative Region, China (GRF 14300923 to Z.L.), and the Shenzhen Loop Area Institute (Grant No. FPF10120250014).

The authors sincerely thank reviewers for their insightful comments and suggestions. To maintain narrative coherence, selected rebuttal experiments were incorporated into main text, while additional analyses are presented in Appendix G due to space constraints. Their thoughtful feedback substantially improved this work and is particularly valuable in an era of increasingly AI-assisted research.

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.

Ricard Argelaguet, Anna SE Cuomo, Oliver Stegle, and John C Marioni. Computational principles and challenges in single-cell data integration. *Nature biotechnology*, 39(10):1202–1215, 2021.

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.

Ana R Baião, Zhaoxiang Cai, Rebecca C Poulos, Phillip J Robinson, Roger R Reddel, Qing Zhong, Susana Vinga, and Emanuel Gonçalves. A technical review of multi-omics data integration methods: from classical statistical to deep generative approaches. *Briefings in bioinformatics*, 26 (4):bbaf355, 2025.

Alev Baysoy, Zhiliang Bai, Rahul Satija, and Rong Fan. The technological landscape and applications of single-cell multi-omics. *Nature Reviews Molecular Cell Biology*, 24(10):695–713, 2023.

Volker Bergen, Marius Lange, Stefan Peidli, F Alexander Wolf, and Fabian J Theis. Generalizing rna velocity to transient cell states through dynamical modeling. *Nature biotechnology*, 38(12): 1408–1414, 2020.

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

Junyue Cao, Malte Spielmann, Xiaojie Qiu, Xingfan Huang, Daniel M Ibrahim, Andrew J Hill, Fan Zhang, Stefan Mundlos, Lena Christiansen, Frank J Steemers, et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, 566(7745):496–502, 2019.

Kai Cao, Qiyu Gong, Yiguang Hong, and Lin Wan. uniport: a unified computational framework for single-cell data integration with optimal transport. *bioRxiv*, pages 2022–02, 2022.

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.

Shuxiao Chen, Bokai Zhu, Sijia Huang, John W Hickey, Kevin Z Lin, Michael Snyder, William J Greenleaf, Garry P Nolan, Nancy R Zhang, and Zongming Ma. Integration of spatial and single-cell data across modalities with weakly linked features. *Nature Biotechnology*, 42(7):1096–1106, 2024.

Haotian Cui, Hassaan Maan, Maria C Vladoiu, Jiao Zhang, Michael D Taylor, and Bo Wang. Deepvelo: deep learning extends rna velocity to multi-lineage systems with cell-specific kinetics. *Genome biology*, 25(1):27, 2024.

Chao Deng, Xinyi Ma, Hui Lu, Hongyu Zhao, Jingsi Ming, and Tao Wang. Bridging unpaired single-cell multimodal data for integrative analyses with supermap. *Proceedings of the National Academy of Sciences*, 123(6):e2505182123, 2026.

Silas Ruhrberg Estévez, Nicolas Huynh, Tennison Liu, Roderik M. Kortlever, Gerard I. Evan, David L. Bentley, and Mihaela Van Der Schaar. Cellbridge: Learning cellular trajectories via interaction-aware alignment. In Tong Zhang, Miroslav Dudik, Martin Jaggi, Alekh Agarwal, Sharon Li, Dale Schuurmans, Jerry Zhu, Felix Berkenkamp, Hanze Dong, and Alberto Bietti, editors, *Proceedings of the 43rd International Conference on Machine Learning*, volume 306 of *Proceedings of Machine Learning Research*, pages 28345–28381. PMLR, 06–11 Jul 2026.

Mingze Gao, Chen Qiao, and Yuanhua Huang. Unitvelo: temporally unified rna velocity reinforces single-cell trajectory inference. *Nature communications*, 13(1):6586, 2022.

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.

Adam Gayoso, Philipp Weiler, Mohammad Lotfollahi, Dominik Klein, Justin Hong, Aaron Streets, Fabian J Theis, and Nir Yosef. Deep generative modeling of transcriptional dynamics for rna velocity analysis in single cells. *Nature methods*, 21(1):50–59, 2024.

Muyang Ge, Jishuai Miao, Ji Qi, Xiaocheng Zhou, and Zhixiang Lin. Tivelo: Rna velocity estimation leveraging cluster-level trajectory inference. *Nature Communications*, 16(1):6258, 2025.

Shila Ghazanfar, Carolina Guibentif, and John C Marioni. Stabilized mosaic single-cell data integration using unshared features. *Nature biotechnology*, 42(2):284–292, 2024.

Boying Gong, Yun Zhou, and Elizabeth Purdom. Cobolt: integrative analysis of multimodal single-cell sequencing data. *Genome biology*, 22(1):351, 2021.

Yichen Gu, David T Blaauw, and Joshua Welch. Variational mixtures of odes for inferring cellular gene expression dynamics. In *International Conference on Machine Learning*, pages 7887–7901. PMLR, 2022.

Yuhan Hao, Stephanie Hao, Erica Andersen-Nissen, William M Mauck, Shiwei Zheng, Andrew Butler, Maddie J Lee, Aaron J Wilk, Charlotte Darby, Michael Zager, et al. Integrated analysis of multimodal single-cell data. *Cell*, 184(13):3573–3587, 2021.

Yuhan Hao, Tim Stuart, Madeline H Kowalski, Saket Choudhary, Paul Hoffman, Austin Hartman, Avi Srivastava, Gesmira Molla, Shaista Madad, Carlos Fernandez-Granda, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature biotechnology*, 42(2):293–304, 2024.

Ilya Korsunsky, Nghia Millard, Jean Fan, Kamil Slowikowski, Fan Zhang, Kevin Wei, Yuriy Baglaenko, Michael Brenner, Po-ru Loh, and Soumya Raychaudhuri. Fast, sensitive and accurate integration of single-cell data with harmony. *Nature methods*, 16(12):1289–1296, 2019.

April R Kriebel and Joshua D Welch. Uinmf performs mosaic integration of single-cell multi-omic datasets using nonnegative matrix factorization. *Nature communications*, 13(1):780, 2022.

Chen Li, Maria C Virgilio, Kathleen L Collins, and Joshua D Welch. Multi-omic single-cell velocity models epigenome–transcriptome interactions and improves cell fate prediction. *Nature biotechnology*, 41(3):387–398, 2023.

Chen Li, Yichen Gu, Maria C Virgilio, Kun H Lee, Kathleen L Collins, and Joshua D Welch. Inferring differential dynamics from multi-lineage, multi-omic, and multi-sample single-cell data with multivelovae. *Nature Communications*, 2025a.

Hechen Li, Ziqi Zhang, Michael Squires, Xi Chen, and Xiuwei Zhang. scmultisim: simulation of single-cell multi-omics and spatial data guided by gene regulatory networks and cell–cell interactions. *Nature methods*, 22(5):982–993, 2025b.

Shengyu Li, Pengzhi Zhang, Weiqing Chen, Lingqun Ye, Kristopher W Brannan, Nhat-Tu Le, Junichi Abe, John P Cooke, and Guangyu Wang. A relay velocity model infers cell-dependent rna velocity. *Nature biotechnology*, 42(1):99–108, 2024.

Shuang Li, Malte Lücken, John C Marioni, Sarah A Teichmann, and Peng He. Toward informed batch correction for single-cell transcriptome integration. *Nature Computational Science*, pages 1–11, 2026.

Xu Liao, Lican Kang, Yihao Peng, Xiaoran Chai, Peng Xie, Chengqi Lin, Hongkai Ji, Yuling Jiao, and Jin Liu. Multivariate stochastic modeling for transcriptional dynamics with cell-specific latent time using sdevelo. *Nature communications*, 15(1):10849, 2024.

Yingxin Lin, Tung-Yu Wu, Sheng Wan, Jean YH Yang, Wing H Wong, and YX Rachel Wang. scjoint integrates atlas-scale single-cell rna-seq and atac-seq data with transfer learning. *Nature biotechnology*, 40(5):703–710, 2022.

Eric F Lock, Katherine A Hoadley, James Stephen Marron, and Andrew B Nobel. Joint and individual variation explained (jive) for integrated analysis of multiple data types. *The annals of applied statistics*, 7(1):523, 2013.

Mohammad Lotfollahi, Anastasia Litinetskaya, and Fabian J Theis. Multigrate: single-cell multi-omic data integration. *BioRxiv*, 2022:484643, 2022.

Malte D Luecken, Daniel Bernard 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 *Thirty-fifth conference on neural information processing systems datasets and benchmarks track (Round 2)*, 2021.

Yuanyuan Ma, Zexuan Sun, Pengcheng Zeng, Wenyu Zhang, and Zhixiang Lin. Jsnmf enables effective and accurate integrative analysis of single-cell multiomics data. *Briefings in bioinformatics*, 23(3):bbac105, 2022.

Jishuai Miao, Jinzhao Li, Jingxue Xin, Jiajuan Tu, Muyang Ge, Ji Qi, Xiaocheng Zhou, Ying Zhu, Can Yang, and Zhixiang Lin. Multigate: integrative analysis and regulatory inference in spatial multi-omics data via graph representation learning. *Nature Communications*, 16(1):9403, 2025.

Xiaojie Qiu, Qi Mao, Ying Tang, Li Wang, Raghav Chawla, Hannah A Pliner, and Cole Trapnell. Reversed graph embedding resolves complex single-cell trajectories. *Nature methods*, 14(10): 979–982, 2017.

Amrit Singh, Casey P Shannon, Benoît Gautier, Florian Rohart, Michaël Vacher, Scott J Tebbutt, and Kim-Anh Lê Cao. Diablo: an integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics*, 35(17):3055–3062, 2019.

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 *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Zhenchao Tang, Guanxing Chen, Shouzhi Chen, Haohuai He, Jiehui Huang, Tiejun Dong, Jun Zhou, Lu Zhao, Linlin You, and Calvin Yu-Chian Chen. Modal-next: Towards unified heterogeneous cellular data integration. *Information Fusion*, 125:103479, 2026.

Cole Trapnell, Davide Cacchiarelli, Jonna Grimsby, Prapti Pokharel, Shuqiang Li, Michael Morse, Niall J Lennon, Kenneth J Livak, Tarjei S Mikkelsen, and John L Rinn. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology*, 32(4):381–386, 2014.

Larissa Traxler, Raffaella Lucciola, Joseph R Herdy, Jeffrey R Jones, Jerome Mertens, and Fred H Gage. Neural cell state shifts and fate loss in ageing and age-related diseases. *Nature Reviews Neurology*, 19(7):434–443, 2023.

Michael E Vinyard, Anders W Rasmussen, Ruitong Li, Allon M Klein, Gad Getz, and Luca Pinello. Learning cell dynamics with neural differential equations. *Nature Machine Intelligence*, pages 1–16, 2025.

Philipp Weiler, Marius Lange, Michal Klein, Dana Pe'er, and Fabian Theis. Cellrank 2: unified fate mapping in multiview single-cell data. *Nature methods*, 21(7):1196–1205, 2024.

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 (1):384, 2023.

Grace XY Zheng, Jessica M Terry, Phillip Belgrader, Paul Ryvkin, Zachary W Bent, Ryan Wilson, Solongo B Ziraldo, Tobias D Wheeler, Geoff P McDermott, Junjie Zhu, et al. Massively parallel digital transcriptional profiling of single cells. *Nature communications*, 8(1):14049, 2017.

Appendix

A Model Details **15**

A.1 Progressive Training . . . 15

A.2 Loss Weights Settings . . . 16

A.3 Transition Modeling . . . 17

B Evaluation Metrics **17**

B.1 Cross-Modal Alignment . . . 17

B.2 Cross-Omics Prediction . . . 17

B.3 Trajectory Inference . . . 18

B.4 Future-State Prediction . . . 18

C Datasets **18**

C.1 Dataset Description . . . 18

C.2 Data Preprocessing . . . 19

D Code Availability **22**

D.1 ATLAS . . . 22

D.2 Comparison Methods . . . 22

E Additional Experiments **22**

E.1 Other Pairing Ratio . . . 22

E.2 RNA–Protein Integration . . . 23

E.3 More Visualizations . . . 23

F Additional Ablation experiments **24**

G Rebuttal Experiments **24**

G.1 Robustness Analysis . . . 24

G.2 Integration Quality . . . 25

G.3 Sensitivity Analysis . . . 25

H Theoretical Analysis **25**

H.1 Temporal Teacher Reliability . . . 26

H.2 Optimization Analysis . . . 27

H.3 Reliability-Weighted Error Reduction . . . 28

A Model Details

A.1 Progressive Training

ATLAS is optimized in three phases. Let $\mathcal{L}_{\mathrm{rec}}$, $\mathcal{L}_{\mathrm{con}}$, and $\mathcal{L}_{\mathrm{KL}}$ denote reconstruction, paired contrastive, and KL losses. The first two phases optimize

$$\mathcal{L}_1 = \lambda_{\mathrm{rec}}\mathcal{L}_{\mathrm{rec}} + \lambda_{\mathrm{con}}\mathcal{L}_{\mathrm{con}} + \lambda_{\mathrm{KL}}\mathcal{L}_{\mathrm{KL}}, \tag{S1}$$

$$\mathcal{L}_2 = \mathcal{L}_1 + \lambda_{\mathrm{MMD}}\mathcal{L}_{\mathrm{MMD}} + \lambda_{\mathrm{OT}}\mathcal{L}_{\mathrm{OT}} + \lambda_{\mathrm{bio}}\mathcal{L}_{\mathrm{bio}} + \lambda_{\mathrm{pred}}\mathcal{L}_{\mathrm{pred}}. \tag{S2}$$

Phase I learns stable variational representations, while Phase II adds cross-modal alignment and auxiliary target prediction. In Phase III, the model further learns latent time and lag-conditioned target prediction:

$$\mathcal{L}_3 = \mathcal{L}_2 + \lambda_{\mathrm{time}}\mathcal{L}_{\mathrm{time}} + \lambda_{\mathrm{lag}}\mathcal{L}_{\mathrm{lag}}, \tag{S3}$$

where $\mathcal{L}_{\mathrm{time}}$ includes teacher distillation and time-head regularization, and $\mathcal{L}_{\mathrm{lag}}$ includes paired/unpaired target prediction and attention smoothness, with the internal weights given in Table S2. Algorithm 1 summarizes the progressive optimization of ATLAS. The model is trained in three phases: representation learning, cross-modal alignment, and temporal dynamics. During the final phase, the temporal teacher and lag banks are periodically refreshed using training data, while model parameters are optimized through mini-batches.

Algorithm 1 Progressive Training of ATLAS

Require: Training set $\mathcal{D}_{\mathrm{tr}}$, lag edges $\mathcal{E}_{\mathrm{lag}}$, epochs (E_1, E_2, E_3), refresh interval R
Ensure: Optimized parameters θ^*

Initialize parameters θ and AdamW optimizer.
/* Phase I: Representation Learning */
for $e = 1$ **to** E_1 **do**
  for mini-batch $\mathcal{B} \subset \mathcal{D}_{\mathrm{tr}}$ **do**
    Update θ by minimizing $\mathcal{L}_1(\mathcal{B})$.
  end for
end for
/* Phase II: Cross-Modal Alignment */
for $e = 1$ **to** E_2 **do**
  for mini-batch $\mathcal{B} \subset \mathcal{D}_{\mathrm{tr}}$ **do**
    Update θ by minimizing $\mathcal{L}_2(\mathcal{B})$.
  end for
end for
/* Phase III: Temporal and Lag-Aware Learning */
for $e = 1$ **to** E_3 **do**
  if $e = 1$ **or** $(e-1) \bmod R = 0$ **then**
    $\mathbf{H}, \{\mathbf{Z}_s^{(m)}\} \leftarrow \mathrm{Encode}_\theta(\mathcal{D}_{\mathrm{tr}})$
    $\{\mathbf{t}^{(k)}\}_{k\in\{g,d,a\}} \leftarrow \mathrm{TemporalCandidates}(\mathbf{H})$
    $w_k \leftarrow \mathrm{softmax}_k(-\mathcal{R}_k/\tau_t)$
    $\mathbf{t}^{\mathrm{teach}} \leftarrow \mathrm{Normalize}\left(\sum_k w_k \,\mathrm{Rank}(\mathbf{t}^{(k)})\right)$
    $\{\mathbf{B}^{(a)}\} \leftarrow \mathrm{LagBank}(\{\mathbf{Z}_s^{(a)}\}, \mathbf{t}^{\mathrm{teach}})$
  end if
  for mini-batch $\mathcal{B} \subset \mathcal{D}_{\mathrm{tr}}$ **do**
    $(\mathbf{t}_{\mathcal{B}}, \boldsymbol{\alpha}_{\mathcal{B}}) \leftarrow F_\theta(\mathcal{B}, \mathbf{B}_{\mathcal{B}})$
    Update θ by minimizing $\mathcal{L}_3^{(e)}(\mathcal{B})$.
  end for
end for
Refresh temporal teacher and lag banks.
return θ^*

A.2 Loss Weights Settings

Table S1 summarizes the loss coefficients. In addition to the representation, alignment, temporal, and lag losses described in the main text, the implementation uses an auxiliary cross-omics prediction objective, $\mathcal{L}_{\text{pred}}$, starting in Phase II. This objective uses observed paired targets and transport-based supervision for unpaired observations.

Table S1: Loss coefficients used in ATLAS training.

Coefficient	Value	Objective
Representation learning		
λ_{rec}	1.00	Observed-modality reconstruction
λ_{con}	1.00	Paired-cell contrastive alignment
λ_{KL}	0.001	Variational regularization
Cross-modal alignment		
λ_{MMD}	0.05	Distribution matching
λ_{OT}	0.20	Optimal transport alignment
λ_{bio}	0.50	Latent-geometry preservation
Prediction and dynamics		
λ_{pred}	1.00	Auxiliary cross-omics prediction
λ_{time}	1.00	Temporal distillation
λ_{lag}	0.30	Adaptive lag modeling

The temporal teacher combines graph-based, diffusion-like, and attribution-proxy estimates through reliability-weighted rank aggregation. The source-specific lag bank uses four equal-width intervals in the normalized teacher-time range. Table S2 lists the temporal and lag settings used in the released training configuration.

Table S2: Temporal-teacher and lag-modeling settings.

Parameter	Value	Purpose
Temporal teacher		
λ_{sm}	1.00	Neighborhood smoothness risk
λ_{mv}	1.00	Cross-modality agreement risk
λ_{sp}	1.00	Temporal spread risk
τ_t	0.85	Reliability-weight temperature
k_{NN}	15	Neighbors for teacher construction
Temporal and lag optimization		
λ_{reg}	0.10	Time-head ordering and spread penalty
λ_{smooth}	0.10	Lag-attention smoothness
$\lambda_{\text{u,pred}}$	1.00	Unpaired prediction supervision
$\lambda_{\text{u,lag}}$	1.00	Unpaired lag supervision
K	4	Number of lag intervals
R	5	Epochs between teacher/bank refreshes
E_{warm}	5	Lag-loss warm-up epochs

The lag-associated transport losses use a temporal window of 0.35, up to 32 local candidates, an entropic regularization of 0.05, and 30 Sinkhorn iterations. The temporal teacher is recomputed from deterministic training embeddings every five epochs during Phase III. Lag banks are constructed from the teacher time, rather than directly from the current time-head predictions; consequently, the discrete bank construction is not differentiated through the time head.

A.3 Transition Modeling

For downstream future-state analysis, a directed transition matrix can be defined over a neighborhood graph using the learned representations and latent times. For $j \in \mathcal{N}(i)$, consider the transition score

$$s_{ij} = -\frac{\|\mathbf{h}_i - \mathbf{h}_j\|_2^2}{\sigma_h^2} + \beta_t(t_j - t_i) + \beta_{\text{lag}} r_{ij}^{\text{lag}}, \tag{S4}$$

where the terms respectively encode neighborhood similarity, temporal direction, and optional lag consistency. Row normalization gives

$$P_{ij} = \frac{\exp(s_{ij})}{\sum_{k \in \mathcal{N}(i)} \exp(s_{ik})}, \qquad j \in \mathcal{N}(i), \tag{S5}$$

with $P_{ij} = 0$ outside the neighborhood. For a set of labeled reference terminal cells $\mathcal{C}_c$ of class c, an L-step transition score can be aggregated as

$$\hat{p}_{i,c} = \frac{\sum_{j \in \mathcal{C}_c} (\mathbf{P}^L)_{ij}}{\sum_{c'} \sum_{j \in \mathcal{C}_{c'}} (\mathbf{P}^L)_{ij}}. \tag{S6}$$

Reference terminal groups must be defined independently of the query labels. This downstream formulation is separate from ATLAS parameter optimization and is not part of the currently released T1–T3 training pipeline.

B Evaluation Metrics

B.1 Cross-Modal Alignment

Cross-modal alignment (T1) evaluates cell-level correspondence between observed and predicted RNA profiles. Given N cells, we jointly project the observed and predicted RNA profiles into a shared PCA space with up to 30 components. Let $\mathbf{u}_i$ and $\mathbf{v}_i$ denote their respective embeddings for cell i. We evaluate alignment using the Fraction of Samples Closer Than the True Match (FOSCTTM) and Paired Embedding Distance (PED):

$$\begin{aligned} \text{FOSCTTM} &= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbb{I}\left[d(\mathbf{u}_i, \mathbf{v}_j) < d(\mathbf{u}_i, \mathbf{v}_i)\right], \\ \text{PED} &= \frac{1}{N} \sum_{i=1}^{N} \|\mathbf{u}_i - \mathbf{v}_i\|_2, \end{aligned} \tag{S7}$$

where $d(\cdot, \cdot)$ denotes Euclidean distance. Lower FOSCTTM and PED indicate better alignment. Top-k matching accuracy ($k = 1, 5, 10$) is additionally reported.

B.2 Cross-Omics Prediction

Cross-omics prediction (T2) evaluates ATAC-to-RNA reconstruction using held-out cells. The model predicts RNA expression from ATAC observations, while ground-truth RNA profiles are used only for evaluation. Predictions are inverse-standardized using training-set statistics. We report cell-wise Pearson correlation and Root Mean Squared Error (RMSE):

$$\begin{aligned} \text{Pear}_{\text{cell}} &= \frac{1}{|\mathcal{V}|} \sum_{i \in \mathcal{V}} \text{corr}\left(\hat{\mathbf{x}}_i^{\text{RNA}}, \mathbf{x}_i^{\text{RNA}}\right), \\ \text{RMSE} &= \sqrt{\frac{1}{NG} \sum_{i=1}^{N} \sum_{g=1}^{G} \left(\hat{x}_{i,g}^{\text{RNA}} - x_{i,g}^{\text{RNA}}\right)^2}, \end{aligned} \tag{S8}$$

where $\mathcal{V}$ contains cells with valid correlations and G denotes the number of evaluated genes. Higher Pearson correlation and lower RMSE indicate better prediction. Gene-wise Pearson correlation is also reported as a supplementary metric.

B.3 Trajectory Inference

Trajectory inference (T3) measures agreement between predicted latent time and reference developmental ordering. Cell-type annotations are mapped to ordinal stages $s_i = \psi(y_i)$. Since pseudotime direction is ambiguous, we select the orientation of predicted time that maximizes POA against the reference stages during evaluation, denoted by $\tilde{t}_i$. Pairwise Ordering Accuracy (POA) is defined as

$$\mathrm{POA} = \frac{\sum_{\substack{i<j \\ s_i \neq s_j}} |s_i - s_j| \, \mathbb{I}\left[(\tilde{t}_i - \tilde{t}_j)(s_i - s_j) > 0\right]}{\sum_{\substack{i<j \\ s_i \neq s_j}} |s_i - s_j|}. \tag{S9}$$

Pairwise Ordering AUC (POAUC) evaluates the discrimination between earlier and later stages:

$$\begin{aligned} \mathrm{AUC}_{a,b} &= \mathrm{AUC}\left(\mathbb{I}[s_i = b], \tilde{t}_i\right), \\ \mathrm{POAUC} &= \frac{\sum_{a<b} |a-b| \, \mathrm{AUC}_{a,b}}{\sum_{a<b} |a-b|}, \end{aligned} \tag{S10}$$

where a and b denote distinct reference stages, and the sums include only valid cells and represented stage pairs. Higher POA and POAUC indicate more accurate temporal ordering.

B.4 Future-State Prediction

Future-state prediction (T4) evaluates the prediction of future cellular states using reference labels y_i and predicted class probabilities $\hat{p}_{i,c}$. We report F1 score and Area Under the Receiver Operating Characteristic Curve (AUROC). For binary classification, predicted labels are obtained using a threshold of 0.5, and the metrics are defined as

$$\begin{aligned} \mathrm{F1} &= \frac{2\,\mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}, \\ \mathrm{AUROC} &= \mathrm{AUC}\left(\mathbb{I}[y_i = 1], \hat{p}_{i,1}\right). \end{aligned} \tag{S11}$$

For multi-class prediction, we use macro-averaged F1 and one-vs-rest AUROC:

$$\begin{aligned} \mathrm{F1}_{\mathrm{macro}} &= \frac{1}{C} \sum_{c=1}^{C} \mathrm{F1}_c, \\ \mathrm{AUROC}_{\mathrm{macro}} &= \frac{1}{C} \sum_{c=1}^{C} \mathrm{AUC}\left(\mathbb{I}[y_i = c], \hat{p}_{i,c}\right). \end{aligned} \tag{S12}$$

Higher F1 and AUROC indicate better future-state prediction performance.

C Datasets

C.1 Dataset Description

We evaluate ATLAS on five single-cell multi-omics datasets covering immune cell heterogeneity and hematopoietic differentiation. PBMC and BMMC are used for cross-modal alignment (T1), cross-omics prediction (T2), and trajectory inference (T3), while three HSPC-centered datasets derived from the MultiVeloVAE benchmark are used for trajectory inference (T3) and future-state prediction (T4). Together, these datasets cover paired, partially paired, and multi-sample settings with diverse cellular dynamics.

PBMC Multiome (10x Genomics). The PBMC dataset contains paired RNA and ATAC measurements from approximately 10,000 human peripheral blood mononuclear cells. It captures heterogeneous immune cell populations, providing a benchmark for evaluating cell-level alignment, cross-omics prediction, and temporal ordering. The filtered feature-barcode matrix is available at: https://cf.10xgenomics.com/samples/cell-arc/1.0.0/pbmc_granulocyte_sorted_10k/pbmc_granulocyte_sorted_10k_filtered_feature_bc_matrix.h5.

BMMC (NeurIPS 2021 Multimodal Benchmark). The BMMC dataset (Luecken et al., 2021) comprises bone marrow mononuclear cells collected across multiple donors and experimental batches. It contains two complementary subsets: paired RNA–ATAC measurements obtained by 10x Multiome and paired RNA–protein measurements obtained by CITE-seq. These heterogeneous modality combinations provide a benchmark for evaluating cross-omics integration and prediction under different pairing ratios. The dataset is available at: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE194122.

Dual-HSPC Multiome (HSPC-2M). HSPC-2M integrates two paired 10x Multiome samples of human hematopoietic stem and progenitor cells (HSPCs). The post-processed dataset, denoted by `3423-MV-2-8489-MV-1`, contains chromatin accessibility and spliced/unspliced RNA measurements. It captures multilineage hematopoietic differentiation, including erythroid, myeloid, and megakaryocytic trajectories, enabling evaluation of temporal ordering and future-state prediction across samples.

HSPC–Macrophage Multiome (HSPC-MAC). HSPC-MAC combines an HSPC sample with a macrophage-induced sample, corresponding to `8489-MV-1-9060-MV-3`. The two samples represent different stages of an induced differentiation process, including progenitor, monocyte, and macrophage states. This dataset provides a directed differentiation setting for evaluating temporal dynamics and future-state prediction across distinct cell states.

HSPC Multiome–scRNA Integration (HSPC-RNA). HSPC-RNA extends HSPC-2M with an additional RNA-only bone marrow sample, corresponding to `3423-MV-2-8489-MV-1-GSM5460410`. The Multiome samples provide paired chromatin accessibility and RNA measurements, whereas the additional sample contains RNA measurements only. This partially paired setting evaluates whether ATLAS can recover consistent cellular dynamics across samples with incomplete modality coverage.

The three HSPC datasets use post-processed AnnData objects released with MultiVeloVAE, available at: https://doi.org/10.6084/m9.figshare.30280333.

C.2 Data Preprocessing

PBMC Cell-type Annotation The downloaded PBMC multiome dataset does not provide cell-type annotations, making it unsuitable for trajectory inference evaluation directly. To obtain reference cell-type labels, we followed the annotation strategy used for BMMC and generated PBMC annotations using Azimuth with the official PBMC reference from the Azimuth web portal. Specifically, we loaded the raw 10x PBMC multiome count matrix, constructed a Seurat object, mapped the cells to the PBMC reference using `RunAzimuth`, and used the predicted level-2 cell-type labels for downstream evaluation as following.

```r
library(Seurat)
library(SeuratData)
library(Azimuth)

# 1. Load the raw 10x PBMC multiome count matrix.
counts <- Read10X_h5(
  "xx_Project/data/PBMC/pbmc_granulocyte_sorted_10k_filtered_feature_bc_matrix.h5"
)

```

```r
# 2. Create a Seurat object using the RNA assay.
pbmc <- CreateSeuratObject(
  counts = counts,
  assay = "RNA",
  project = "PBMC"
)

# 3. Map PBMC cells to the Azimuth PBMC reference.
pbmc <- RunAzimuth(
  pbmc,
  reference = "pbmcref"
)

# 4. Inspect the predicted level-2 cell-type annotations.
print(table(pbmc$predicted.celltype.l2))

# 5. Save all Azimuth-predicted annotation columns.
pred_cols <- grep("^predicted", colnames(pbmc@meta.data), value = TRUE)

anno <- data.frame(
  cell_id = colnames(pbmc),
  pbmc@meta.data[, pred_cols, drop = FALSE],
  row.names = NULL
)

write.csv(
  anno,
  file = "pbmc_azimuth_annotations.csv",
  row.names = FALSE,
  quote = FALSE
)

# 6. Save the full mapped Seurat object.
saveRDS(
  pbmc,
  file = "pbmc_azimuth_mapped.rds"
)

# 7. Save a UMAP visualization colored by predicted level-2 cell type.
pdf("pbmc_azimuth_celltype_l2.pdf", width = 10, height = 8)
print(
  DimPlot(
    pbmc,
    group.by = "predicted.celltype.l2",
    label = TRUE
  ) + NoLegend()
)
dev.off()
```

Listing 1: PBMC cell-type annotation using Azimuth.

Cell Stage Order Generation To evaluate the T3 trajectory-inference task, we converted discrete cell-type annotations into coarse ordinal developmental stages. Since the available annotations are categorical rather than continuous pseudotime labels, we defined a stage-order map based on hematopoietic differentiation progression. Smaller values indicate earlier progenitor or stem-like states, whereas larger values indicate more differentiated or mature cell states. These ordinal labels were used only as reference ordering for T3 evaluation.

```r
# PBMC stage-order map for T3 trajectory inference.
# Smaller values indicate earlier developmental stages.
```

```

pbmc_stage_order_map = {
  "HSPC": 0,

  # Early lineage-committed states
  "B naive": 1,
  "CD4 Naive": 1,
  "CD8 Naive": 1,
  "Eryth": 1,

  # Intermediate states
  "B intermediate": 2,
  "CD4 TCM": 2,
  "CD8 TCM": 2,
  "CD14 Mono": 2,
  "CD16 Mono": 2,
  "cDC1": 2,
  "cDC2": 2,
  "pDC": 2,
  "CD4 Proliferating": 2,
  "NK Proliferating": 2,

  # Mature immune states
  "B memory": 3,
  "Plasmablast": 3,
  "CD4 TEM": 3,
  "CD8 TEM": 3,
  "Treg": 3,
  "MAIT": 3,
  "NK": 3,
  "dnT": 3,
  "gdT": 3
}
```

Listing 2: PBMC stage-order map used for T3 trajectory-inference evaluation.

```
# BMMC stage-order map for T3 trajectory inference.
# Smaller values indicate earlier hematopoietic stages.

bmmc_stage_order_map = {
  "HSC": 0,

  # Early progenitors
  "Lymph prog": 1,
  "G/M prog": 1,
  "MK/E prog": 1,
  "ID2-hi myeloid prog": 1,

  # Intermediate lineage-committed states
  "CD4+ T naive": 2,
  "CD8+ T naive": 2,
  "Naive CD20+ B": 2,
  "Transitional B": 2,
  "B1 B": 2,
  "Proerythroblast": 2,

  # Mature immune and erythroid states
  "NK": 3,
  "CD8+ T": 3,
  "CD4+ T activated": 3,
  "CD14+ Mono": 3,
```

```
    "CD16+ Mono": 3,
    "pDC": 3,
    "cDC2": 3,
    "ILC": 3,
    "Plasma cell": 3,
    "Erythroblast": 3,

    # Late erythroid state
    "Normoblast": 4
}
```

Listing 3: BMMC stage-order map used for T3 trajectory-inference evaluation.

D Code Availability

D.1 ATLAS

The official ATLAS repository provides implementations of variational representation learning, multi-level cross-modal alignment, reliability-guided temporal distillation, and adaptive lag modeling, together with data preprocessing and evaluation pipelines.

Repository: https://github.com/zhangye-zoe/ATLAS

D.2 Comparison Methods

We provide separate reproduction repositories for the static multi-omics integration and velocity-based dynamic modeling baselines considered in our experiments:

- **GLUE:** Graph-linked embedding for multi-omics integration. https://github.com/zhangye-zoe/GLUE_RE
- **Seurat V5:** Anchor-based integration and cross-modal transfer. https://github.com/zhangye-zoe/Seurat_RE
- **scMRDR:** Multi-resolution representation learning for cross-modal integration. https://github.com/zhangye-zoe/scMRDR_RE
- **MaxFuse:** Cross-modal cell matching through shared features. https://github.com/zhangye-zoe/MaxFuse_RE
- **Harmony:** Batch correction and integration in latent space. https://github.com/zhangye-zoe/Harmony_RE
- **Velocity-based methods:** RNA and multi-omics velocity benchmarks. https://github.com/zhangye-zoe/VelocityBench

E Additional Experiments

E.1 Other Pairing Ratio

To further evaluate model robustness under limited paired measurements, we provide additional comparisons under the 20% paired-data ratio in Table S3. This setting represents a highly incomplete multi-omics scenario where only a small fraction of cells have paired observations, while the remaining cells are observed in a single modality.

Across PBMC and BMMC, ATLAS achieves the best FOSCTTM scores, indicating more accurate cross-modal alignment under limited pairing. On BMMC, ATLAS also obtains the lowest PED, suggesting improved cell-level correspondence in the shared latent space. For cross-omics prediction, ATLAS achieves the lowest RMSE on BMMC and competitive RMSE on PBMC, although Seurat

obtains higher Pearson correlation in this setting. For trajectory inference, ATLAS achieves the highest POA on both PBMC and BMMC, showing that its direct latent-time inference remains effective when paired measurements are scarce. These results further support that ATLAS learns representations that are not only well aligned, but also informative for downstream dynamic modeling.

Table S3: **Performance comparison with static multi-omics integration methods under different paired-data ratios.** All metrics are reported as percentages (%) except PED and RMSE.

Ratio	Model	T1: Cross-modal Alignment				T2: Cross-omics Prediction				T3: Trajectory Inference				Type
		PBMC		BMMC		PBMC		BMMC		PBMC		BMMC		
		FOS ↓	PED ↓	FOS ↓	PED ↓	Pear ↑	RMSE ↓	Pear ↑	RMSE ↓	POA ↑	PAUC ↑	POA ↑	PAUC ↑	
20%	SeuratV5†	25.5	33.2	17.9	38.6	**65.1**	0.444	57.8	0.533	45.7	81.5	60.6	89.6	△
	scMRDR†	47.8	126.9	45.6	53.4	51.5	1.226	43.5	0.746	64.3	**88.4**	72.8	87.1	△
	MaxFuse†	48.9	206.0	48.2	98.9	30.0	1.812	31.1	2.023	56.9	86.0	59.5	88.0	○
	Harmony†	23.6	38.9	30.3	48.2	64.5	0.485	50.6	0.247	41.4	75.3	63.6	**90.4**	○
	ATLAS‡	**6.50**	12.4	**6.69**	**5.81**	63.6	0.366	52.5	**0.211**	**66.5**	86.8	**74.6**	88.2	–

† denotes trajectory inferred by Monocle3 (Cao et al., 2019) on learned embeddings; ‡ denotes trajectory inferred directly by ATLAS. FOS: FOSCTTM; Pear: Pearson correlation coefficient. △ denotes distribution-matching methods; ○ denotes cell-centric methods.

E.2 RNA–Protein Integration

To evaluate whether ATLAS generalizes beyond RNA–ATAC integration, we further test it on RNA–Protein static integration under different single-modality ratios. As shown in Table S4, ATLAS maintains stable cross-omics prediction performance across masking ratios, with cell-wise Pearson remaining around 73% and RMSE around 0.27. Cross-modal matching becomes more challenging as the single-modality ratio increases, but Top-10 matching improves under higher masking ratios, suggesting that ATLAS can still recover meaningful RNA–Protein correspondence.

E.3 More Visualizations

To further evaluate trajectory inference and cross-modal alignment, we provide additional UMAP visualizations of ATLAS. As shown in Fig. S1, ATLAS exhibits clearer temporal progression and stronger consistency with reference developmental stages than MultiVeloVAE, demonstrating its ability to capture biologically meaningful cellular dynamics. Furthermore, Fig. S2 illustrates cross-modal alignment under 20%, 40%, 60%, and 80% pairing ratios. ATLAS maintains substantial overlap between RNA and ATAC embeddings while preserving distinct cell-type structures across different levels of paired supervision, highlighting the robustness of learned representations.

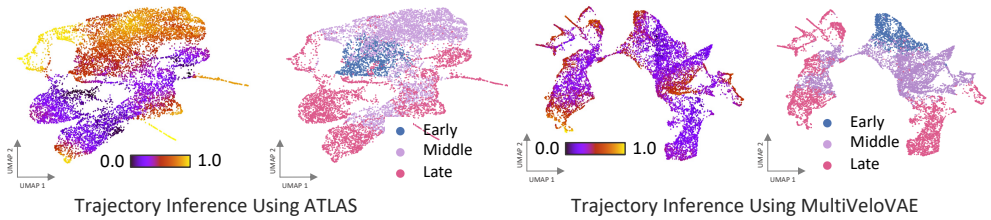

Figure S1: **Trajectory inference comparison between ATLAS and MultiVeloVAE.** UMAP visualizations of predicted latent time and reference developmental stages. ATLAS shows stronger consistency between inferred temporal progression and reference stage ordering, while MultiVeloVAE exhibits greater discrepancies across cell populations.

Table S4: **RNA–Protein static integration results.** All metrics are reported as percentages (%) except PED and RMSE. Best and second-best results are highlighted in bold and underline, respectively.

Paired Ratio	T1: Alignment		T2: Prediction	
	FOS↓	PED↓	Cell Pear.↑	RMSE↓
0%	**3.77**	**2.113**	**73.61**	**0.270**
20%	<u>3.84</u>	<u>2.118</u>	73.50	**0.270**
40%	<u>3.84</u>	2.119	73.46	0.271
60%	3.93	2.119	73.40	**0.270**
80%	4.01	2.125	<u>73.58</u>	0.271
100%	4.05	2.138	73.30	0.273

F Additional Ablation experiments

To examine whether the trajectory improvement of ATLAS comes from the learned representation or from the dedicated latent-time inference module, we performed a fairness analysis by comparing different temporal inference strategies under controlled settings. Specifically, we first applied the same Monocle3 trajectory inference procedure to embeddings learned by different methods. We then fixed the ATLAS embedding and compared different temporal inference strategies.

As shown in Table S5, ATLAS embedding substantially improves Monocle3-based trajectory inference over GLUE and Seurat embeddings, increasing POA from 43.9 and 54.2 to 61.0, respectively. This indicates that ATLAS learns representations that are intrinsically more informative for temporal ordering, rather than merely relying on a dedicated time head. Moreover, when using the same ATLAS embedding, DPT, the reliability-guided teacher, and the ATLAS time head all achieve strong trajectory performance. In particular, the ATLAS time head obtains the highest POAUC of 87.7, suggesting that reliability-guided temporal distillation improves the stability and separability of the inferred latent time. These results demonstrate that the trajectory gains of ATLAS arise from both dynamically informative representation learning and robust latent-time modeling.

Table S5: **Fairness analysis for trajectory inference.** We compare Monocle3 across different embeddings and temporal inference strategies on the same ATLAS embedding, demonstrating the benefits of dynamic representation learning and reliability-guided latent-time modeling.

Controlled Setting	Representation / Time Inference	POA ↑	POAUC ↑
Same temporal inference on different embeddings	GLUE embedding + Monocle3	43.9	81.9
	Seurat embedding + Monocle3	54.2	84.1
	ATLAS embedding + Monocle3	**61.0**	84.0
Different temporal inference on ATLAS embedding	ATLAS embedding + Monocle3	61.0	84.0
	ATLAS embedding + DPT	62.9	85.6
	ATLAS embedding + Reliability teacher	**63.1**	87.3
	ATLAS embedding + ATLAS time head	62.7	**87.7**

G Rebuttal Experiments

G.1 Robustness Analysis

To investigate whether the results depend on a particular initialization or modality-masking configuration, we repeated the PBMC experiments under three pairing ratios using three independent runs. Both network initialization and modality-masking splits were regenerated for each run. Table S6 reports the mean and standard deviation. Across the reported runs, ATLAS achieves better mean FOSCTTM, RMSE, POA, and POAUC than GLUE in all three pairing settings. These results support

Table S6: **Robustness to initialization and masking on PBMC**. FOSCTTM, POA, and POAUC are reported as percentages; RMSE is reported in its original scale.

Paired	Method	FOS ↓	RMSE ↓	POA ↑	POAUC ↑
0%	GLUE	30.6 ± 2.4	0.259 ± 0.031	61.7 ± 2.2	84.8 ± 1.5
	ATLAS	$\mathbf{18.6 \pm 1.8}$	$\mathbf{0.237 \pm 0.027}$	$\mathbf{64.8 \pm 2.1}$	$\mathbf{85.1 \pm 1.5}$
40%	GLUE	19.2 ± 2.6	0.235 ± 0.023	50.1 ± 2.3	82.4 ± 1.3
	ATLAS	$\mathbf{5.10 \pm 0.98}$	$\mathbf{0.219 \pm 0.019}$	$\mathbf{61.9 \pm 2.1}$	$\mathbf{86.9 \pm 1.2}$
100%	GLUE	13.4 ± 1.1	0.229 ± 0.015	48.7 ± 1.8	80.1 ± 1.2
	ATLAS	$\mathbf{4.75 \pm 0.60}$	$\mathbf{0.187 \pm 0.012}$	$\mathbf{56.0 \pm 1.9}$	$\mathbf{83.1 \pm 1.1}$

the reproducibility of the main findings across the tested configurations, although the variability remains non-negligible, particularly in the fully unpaired setting.

At 40% pairing, the POA/POAUC standard deviations were 2.1/1.2 when only initialization varied, 2.3/1.4 when only the mask varied, and 2.4/1.6 when both varied. On HSPC-MAC, three-run F1 and AUROC were 0.804 ± 0.019 and 0.791 ± 0.015 for ATLAS, compared with 0.794 ± 0.021 and 0.773 ± 0.017 for SDEvelo. These analyses address sensitivity to the tested initialization and simulated masking choices, not all possible cross-technology shifts.

G.2 Integration Quality

Beyond the FOSCTTM and paired embedding distance metrics used in the main evaluation, we examine cross-modal mixing and the conservation of biological cell-type structure using an expanded set of integration metrics. For mixing, we report integration local inverse Simpson's index (iLISI), batch average silhouette width (batch ASW), k-nearest-neighbor batch effect test (kBET), and graph connectivity. For biological conservation, we report cell-type ASW, normalized mutual information (NMI), adjusted Rand index (ARI), and isolated-label score. The reported scores are oriented so that higher values indicate better performance. We also report aggregate biological-conservation, mixing, and overall scores.

Table S7 summarizes BMMC integration performance across pairing ratios. ATLAS achieves the highest iLISI at all tested ratios and strong mixing scores, but its biological conservation metrics, particularly NMI and ARI, are often below those of Seurat and scMRDR. At full pairing, scMRDR achieves an overall score of 0.664 versus 0.646 for ATLAS, despite ATLAS having a higher mixing aggregate (0.855 versus 0.675). The results illustrate that optimizing modality mixing and preserving fine-grained cell-type organization are distinct goals.

G.3 Sensitivity Analysis

We further investigate sensitivity to principal loss coefficients, lag-bank resolution, and training schedule. For the loss-weight study, each configuration is evaluated over three random seeds, and the central value of each tested triplet is the default. Table S8 reports that BMMC POAUC and HSPC-2M dynamic metrics remain relatively stable across the tested two-fold variations of the loss coefficients. While the results do not establish robustness to arbitrary settings, they suggest that performance is not restricted to a single sharply tuned configuration.

For lag-bank sizes $K \in \{2, 4, 6, 8\}$, normalized lag MAE values were 0.121, 0.091, 0.097, and 0.108, respectively, with the lowest error at $K = 4$. Progressive training improved BMMC POAUC from 86.8 to 88.4 and HSPC-2M AUROC from 84.0 to 85.1 relative to joint optimization. These additional checks support the selected lag-bank resolution and staged optimization within the evaluated settings.

H Theoretical Analysis

Unsupervised latent-time learning is inherently ambiguous because the true temporal order is not directly observed. A single pseudotime estimator may be unreliable due to root-selection errors,

Table S7: **Comprehensive BMMC integration evaluation across pairing ratios.** BC denotes biological-conservation aggregate and Mix denotes mixing aggregate.

Pair	Model	iLISI	Batch ASW	kBET	Graph conn.	Cell ASW	NMI	ARI	Isolated	BC	Mix	Overall
100%	Seurat	0.000	0.678	0.372	0.420	0.550	0.763	0.686	0.549	0.662	0.372	0.546
	scMRDR	0.687	0.909	0.319	0.849	0.556	0.743	0.647	0.558	0.657	0.675	**0.664**
	ATLAS	**0.903**	**0.971**	**0.807**	0.759	0.500	0.536	0.360	0.526	0.507	**0.855**	0.646
80%	Seurat	0.000	0.674	0.397	0.426	0.551	0.756	0.671	0.550	0.677	0.377	0.557
	scMRDR	0.665	0.892	0.338	0.836	0.560	0.749	0.631	0.557	0.648	0.665	**0.655**
	ATLAS	**0.898**	**0.964**	**0.818**	0.753	0.496	0.548	0.429	0.522	0.520	**0.852**	0.653
60%	Seurat	0.000	0.676	0.407	0.441	0.549	0.754	0.659	0.551	0.652	0.383	0.544
	scMRDR	0.539	0.903	0.306	0.847	0.551	0.731	0.614	0.551	0.634	0.632	0.633
	ATLAS	**0.881**	**0.955**	**0.759**	0.745	0.494	0.539	0.416	0.522	0.509	**0.826**	**0.636**
40%	Seurat	0.001	0.675	0.438	0.418	0.552	0.755	0.652	0.552	0.671	0.386	0.557
	scMRDR	0.578	0.909	0.306	0.845	0.546	0.732	0.657	0.550	0.641	0.644	**0.642**
	ATLAS	**0.828**	**0.939**	**0.685**	0.714	0.500	0.566	0.495	0.526	0.540	**0.781**	0.637
20%	Seurat	0.001	0.666	0.448	0.432	0.553	0.764	0.668	0.551	0.682	0.387	0.564
	scMRDR	0.569	0.895	0.338	0.839	0.552	0.720	0.581	0.550	0.638	0.640	**0.639**
	ATLAS	**0.792**	**0.920**	**0.604**	0.726	0.497	0.546	0.436	0.527	0.523	**0.747**	0.612
0%	Seurat	0.002	0.667	0.549	0.423	0.553	0.759	0.663	0.549	0.631	0.410	0.543
	scMRDR	0.500	**0.902**	0.338	**0.833**	0.548	0.737	0.638	**0.550**	0.618	0.643	**0.628**
	ATLAS	**0.757**	0.734	0.459	0.645	0.470	0.564	0.433	0.450	0.479	**0.649**	0.547

Table S8: **Loss-weight sensitivity.** Results are reported as mean $\pm$ standard deviation over three seeds. The central configuration is the default.

Parameter	Tested Values	Evaluation	Results
λ_{MMD}	0.025 / 0.05 / 0.10	BMMC POAUC	87.6 ± 1.0 / 88.4 ± 0.8 / 88.0 ± 0.9
λ_{OT}	0.10 / 0.20 / 0.40	BMMC POAUC	87.3 ± 1.1 / 88.4 ± 0.8 / 87.9 ± 1.0
λ_{time}	0.50 / 1.00 / 2.00	HSPC-2M POA	83.1 ± 2.4 / 85.9 ± 1.9 / 84.7 ± 2.2
λ_{lag}	0.15 / 0.30 / 0.60	HSPC-2M AUROC	84.3 ± 1.2 / 85.1 ± 1.0 / 84.6 ± 1.3

graph noise, branch mixing, or modality-specific artifacts. A naive average of multiple temporal candidates can also be harmful when unreliable candidates are given the same weight as reliable ones. To address this issue, ATLAS constructs a reliability-guided temporal teacher by adaptively weighting multiple unsupervised temporal candidates according to label-free reliability risks. We provide a theoretical interpretation of this teacher construction below.

H.1 Temporal Teacher Reliability

Given the aligned multi-omics representation

$$\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_N]^\top \in \mathbb{R}^{N \times d},$$

ATLAS constructs a set of unsupervised temporal candidates

$$\mathcal{T} = \left\{ \mathbf{t}^{(1)}, \ldots, \mathbf{t}^{(K)} \right\}, \qquad \mathbf{t}^{(k)} \in [0,1]^N.$$

In the main model, these candidates correspond to neighborhood-based, diffusion-based, and attribution-aware temporal signals. Since different candidates may have different scales, we aggregate them in rank space:

$$\mathbf{r}^{(k)} = \mathrm{Rank}\left(\mathbf{t}^{(k)} \right), \qquad \mathbf{r}^{(k)} \in [0,1]^N.$$

This transformation makes the teacher invariant to monotonic rescaling of each temporal candidate and focuses on the induced cell ordering.

For each candidate $\mathbf{t}^{(k)}$, ATLAS computes a label-free reliability risk:

$$\mathcal{R}\Big(\mathbf{t}^{(k)}\Big) = \lambda_{\mathrm{sm}}\mathcal{R}_{\mathrm{sm}}\Big(\mathbf{t}^{(k)}\Big) + \lambda_{\mathrm{mv}}\mathcal{R}_{\mathrm{mv}}\Big(\mathbf{t}^{(k)}\Big) + \lambda_{\mathrm{sp}}\mathcal{R}_{\mathrm{sp}}\Big(\mathbf{t}^{(k)}\Big),$$

where $\mathcal{R}_{\mathrm{sm}}$, $\mathcal{R}_{\mathrm{mv}}$, and $\mathcal{R}_{\mathrm{sp}}$ measure local graph smoothness, cross-modality agreement, and temporal spread, respectively. These terms do not require cell-type labels, stage annotations, or terminal-state labels.

The reliability weights are obtained by a temperature-scaled softmax over the negative risks:

$$w_k = \frac{\exp\left(-\mathcal{R}(\mathbf{t}^{(k)})/\tau_t\right)}{\sum_{j=1}^{K}\exp\left(-\mathcal{R}(\mathbf{t}^{(j)})/\tau_t\right)}, \qquad \sum_{k=1}^{K} w_k = 1.$$

A smaller temperature τ_t makes the teacher closer to selecting the lowest-risk candidate, whereas a larger τ_t produces a smoother ensemble.

The final temporal teacher is defined as

$$\mathbf{t}^{\mathrm{teach}} = \mathrm{Normalize}\left(\sum_{k=1}^{K} w_k \mathbf{r}^{(k)}\right).$$

The TimeHead ϕ is then trained to predict this teacher:

$$\mathbf{t}^{\mathrm{head}} = \phi(\mathbf{H}), \qquad \mathcal{L}_{\mathrm{time}} = \left\|\mathbf{t}^{\mathrm{head}} - \mathbf{t}^{\mathrm{teach}}\right\|_2^2 + \lambda_{\mathrm{reg}}\mathcal{L}_{\text{time-reg}}.$$

The regularization term encourages smooth, ordered, and non-collapsed temporal predictions, as described in the main text.

H.2 Optimization Analysis

The reliability-weighted teacher can be interpreted as the solution to an entropy-regularized risk minimization problem. Let $\mathbf{w} = (w_1, \ldots, w_K)$ denote the weights assigned to the K temporal candidates. We consider the following optimization problem over the probability simplex:

$$\min_{\mathbf{w}\in\Delta_K} \sum_{k=1}^{K} w_k \mathcal{R}\Big(\mathbf{t}^{(k)}\Big) + \tau_t \sum_{k=1}^{K} w_k \log w_k,$$

where

$$\Delta_K = \left\{\mathbf{w} : w_k \geq 0,\ \sum_{k=1}^{K} w_k = 1\right\}.$$

The first term favors temporal candidates with lower reliability risk, while the entropy term prevents degenerate selection when multiple candidates have similar risks.

Proposition 1. The minimizer of the entropy-regularized reliability objective is

$$w_k^\star = \frac{\exp\left(-\mathcal{R}(\mathbf{t}^{(k)})/\tau_t\right)}{\sum_{j=1}^{K}\exp\left(-\mathcal{R}(\mathbf{t}^{(j)})/\tau_t\right)}.$$

Proof. The Lagrangian is

$$\mathcal{J}(\mathbf{w}, \alpha) = \sum_{k=1}^{K} w_k \mathcal{R}\Big(\mathbf{t}^{(k)}\Big) + \tau_t \sum_{k=1}^{K} w_k \log w_k + \alpha\left(\sum_{k=1}^{K} w_k - 1\right).$$

Taking the derivative with respect to w_k and setting it to zero gives

$$\mathcal{R}\Big(\mathbf{t}^{(k)}\Big) + \tau_t(\log w_k + 1) + \alpha = 0.$$

Therefore,

$$w_k = \exp\left(-\frac{\mathcal{R}(\mathbf{t}^{(k)}) + \alpha + \tau_t}{\tau_t}\right) \propto \exp\left(-\frac{\mathcal{R}(\mathbf{t}^{(k)})}{\tau_t}\right).$$

Normalizing over all candidates yields the stated softmax form. □

This result shows that the teacher weights used in ATLAS are not heuristic: they are the optimal solution of an entropy-regularized reliability minimization problem.

H.3 Reliability-Weighted Error Reduction

We next provide a simple analysis showing why reliability-adaptive weighting can reduce negative transfer from unreliable temporal candidates.

Assumption. Let $\mathbf{t}^\star \in [0,1]^N$ denote an unknown ideal latent time. Each rank-transformed temporal candidate satisfies

$$\mathbf{r}^{(k)} = \mathbf{t}^\star + \boldsymbol{\epsilon}^{(k)},$$

where $\boldsymbol{\epsilon}^{(k)}$ is candidate-specific noise with

$$\mathbb{E}\left[\boldsymbol{\epsilon}^{(k)}\right] = 0, \qquad \mathbb{E}\left[\left\|\boldsymbol{\epsilon}^{(k)}\right\|_2^2\right] \le \sigma_k^2.$$

We assume that candidate errors are uncorrelated across k, and that the reliability risk is positively associated with temporal noise, i.e., noisier candidates tend to have larger expected risk.

Proposition 2. Let

$$\mathbf{t}^{\text{teach}} = \sum_{k=1}^{K} w_k \mathbf{r}^{(k)}$$

be the reliability-weighted teacher. Then

$$\mathbb{E}\left[\left\|\mathbf{t}^{\text{teach}} - \mathbf{t}^\star\right\|_2^2\right] \le \sum_{k=1}^{K} w_k^2 \sigma_k^2.$$

In comparison, the uniform teacher

$$\mathbf{t}^{\text{uni}} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{r}^{(k)}$$

satisfies

$$\mathbb{E}\left[\left\|\mathbf{t}^{\text{uni}} - \mathbf{t}^\star\right\|_2^2\right] \le \frac{1}{K^2}\sum_{k=1}^{K} \sigma_k^2.$$

When some temporal candidates have substantially larger noise, assigning them smaller weights reduces their contribution to the teacher error.

Proof. Using $\mathbf{r}^{(k)} = \mathbf{t}^\star + \boldsymbol{\epsilon}^{(k)}$, we have

$$\mathbf{t}^{\text{teach}} - \mathbf{t}^\star = \sum_{k=1}^{K} w_k \boldsymbol{\epsilon}^{(k)}.$$

Taking the squared norm and expectation gives

$$\mathbb{E}\left[\left\|\mathbf{t}^{\text{teach}} - \mathbf{t}^\star\right\|_2^2\right] = \mathbb{E}\left[\left\|\sum_{k=1}^{K} w_k \boldsymbol{\epsilon}^{(k)}\right\|_2^2\right].$$

Expanding the squared norm yields

$$\sum_{k=1}^{K} w_k^2 \mathbb{E}\left[\left\|\boldsymbol{\epsilon}^{(k)}\right\|_2^2\right] + 2\sum_{k<l} w_k w_l \mathbb{E}\left[\left\langle \boldsymbol{\epsilon}^{(k)}, \boldsymbol{\epsilon}^{(l)} \right\rangle\right].$$

Since candidate errors are assumed to be uncorrelated, the cross terms vanish. Therefore,

$$\mathbb{E}\left[\left\|\mathbf{t}^{\text{teach}} - \mathbf{t}^{\star}\right\|_2^2\right] = \sum_{k=1}^{K} w_k^2 \mathbb{E}\left[\left\|\boldsymbol{\epsilon}^{(k)}\right\|_2^2\right] \leq \sum_{k=1}^{K} w_k^2 \sigma_k^2.$$

The uniform-teacher bound follows by setting $w_k = 1/K$. □

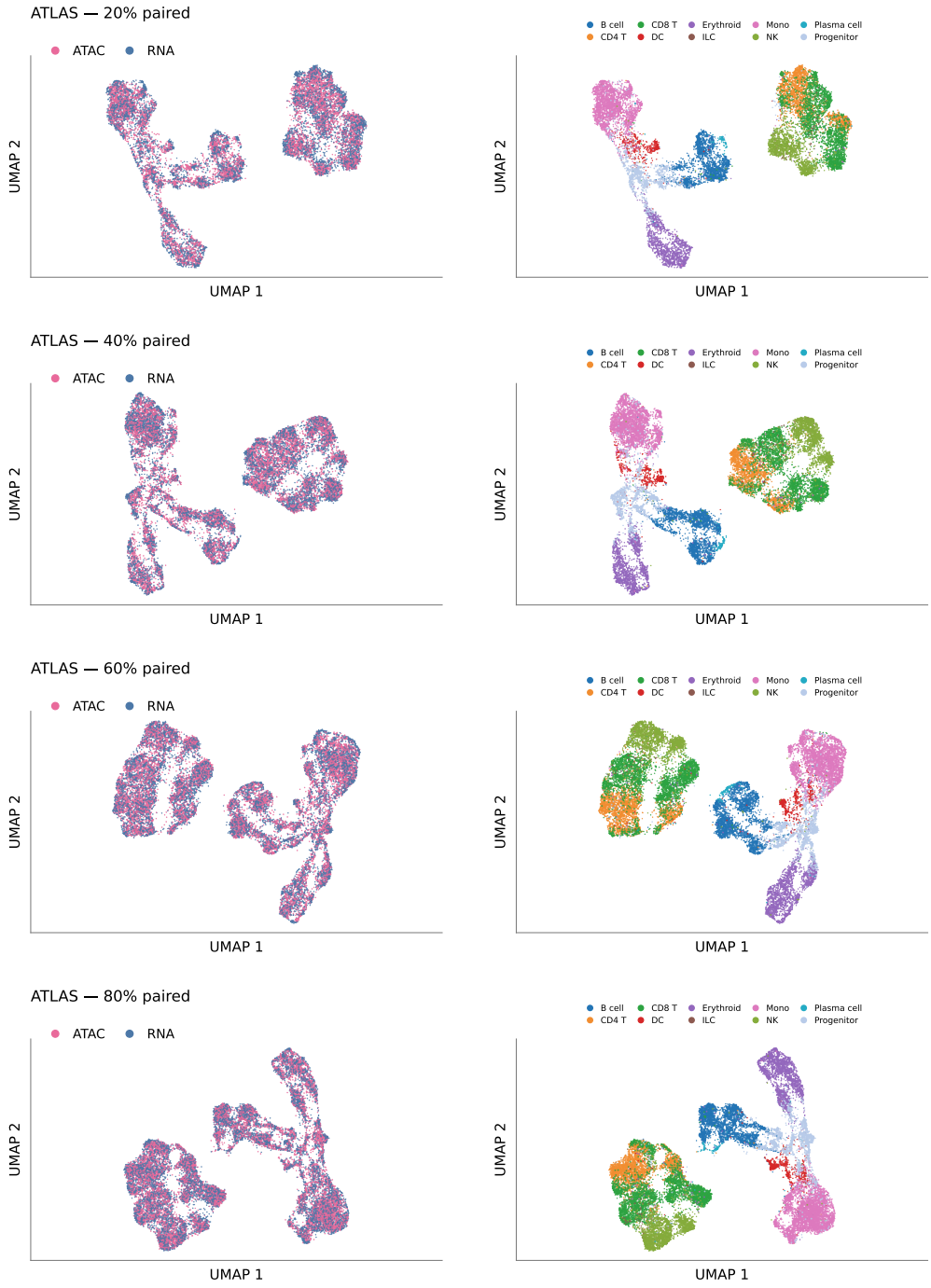

Figure S2: **Cross-modal alignment under different pairing ratios.** UMAP visualizations of ATLAS embeddings at 20%, 40%, 60%, and 80% pairing ratios. The left column shows RNA and ATAC distributions, while the right column shows the corresponding cell-type annotations. ATLAS maintains substantial modality mixing while preserving distinct biological structures across different levels of paired supervision.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: In this paper, we propose ATLAS, a unified framework for learning cell-level cross-omics alignment in dynamic regulatory processes through adaptive time-lag modeling. The limitations of existing multi-omics alignment methods and the key contributions of ATLAS are clearly discussed in both the abstract and Section 1.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The limitations of the proposed ATLAS framework, together with directions for future work, are discussed in main text.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: The theoretical results, together with the full set of assumptions and a proof sketch, are provided in Appendix H.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification:We present the proposed ATLAS model in Section 3, with additional details provided in Appendix A. The experimental setups are described in Section 4.3. The code for ATLAS is available through the anonymous link provided in the abstract, and the code for the comparison methods is described in Appendix D.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: This paper provides open-access code through the link in the abstract, along with sufficient instructions to faithfully reproduce the main experimental results. Dataset download links and data preparation instructions are provided in Appendix C.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The experimental settings are detailed in Section 4.3. An anonymous GitHub link is also provided to facilitate reproducibility of the experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: Error bars are not reported due to the high computational overhead required to perform repeated runs.

   Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We described the computation resources used in experiments in Section 4.3.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have read the NeurIPS Code of Ethics and have made every effort to abide by it throughout this research.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

Justification: We summarize both the positive and negative societal impacts of the proposed ATLAS. We also discuss possible mitigation strategies for the negative impacts.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: This work does not introduce models that pose foreseeable risks of misuse or dual use.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: We cite the comparison code, together with the corresponding URLs, in Section 5.1 and Appendix D. The datasets used for comparison are cited, and their URLs are provided in Appendix C. The licenses of the datasets and code used in this work are also reported in Appendix D and Appendix C, except for Velocity-Bench, which contains mixed licenses, including BSD-3-Clause, BSD-4-Clause, MIT, EUPL-1.2, Unlicense, GPL-2.0-only, GPL-3.0-only, GPL-3.0-or-later, and LGPL-3.0-or-later.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: New assets introduced in this paper are well documented in the anonymous GitHub repository linked in the abstract.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing or research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The paper does not involve crowdsourcing or research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method developed in this paper does not involve LLMs as important, original, or non-standard components.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.