Graph-Regularized Optimal Transport for Single-Cell Data Integration

Zexuan Wang[1⋆], Qipeng Zhan[1⋆], and Li Shen[1⋆⋆]

University of Pennsylvania, Philadelphia PA 19104

Abstract. Recent advancements in single-cell omics technologies have enabled comprehensive analyses of cellular processes. However, due to the lack of co-assay sequencing technologies, single-cell omics datasets often comprise unpaired observations with varying feature dimensions. In this study, we present Graph-Regularized Optimal Transport for Single-Cell Data Integration (GROT), a computational method that aligns different multi-omics datasets without requiring correspondence information. GROT achieves global alignment through optimal transport while preserving local structure via graph regularization. We demonstrate its effectiveness on four simulated datasets and two real-world datasets, achieving superior performance compared to current state-of-the-art unsupervised alignment methods. The software is available at `https://github.com/PennShenLab/GROT`.

Keywords: Optimal Transport · Graph Laplacian · Single Cell· Multi Omics · Data Integration

1 Introduction

The advancement of single-cell technology offers a comprehensive understanding of cellular heterogeneity and the dynamic evolution of cell states. Various single-cell measurements reveal different aspects: scRNA-seq [1, 2] provides detailed gene expression profiles, while scATAC-seq [3] sheds light on chromatin accessibility in individual cells. Integrating these datasets is crucial as it allows for a more holistic view of cellular mechanisms, enabling the correlation of transcriptional activity with chromatin states to better understand gene regulation and cellular function.

Lots of computational methods have recently been developed to integrate data across multiple modalities [4, 5]. A critical challenge for these algorithms is their reliance on correspondence information to identify alignments between cells or features. In practice, such information is often only partially available, hindering the effectiveness of existing strategies [6, 7, 8, 9]. Consequently, research has focused on data integration under semi-supervised settings, where integration does not rely on correspondences between cells and features. The Generalized Unsupervised Manifold Alignment (GUMA) method by Cui et al. [10] aligns datasets by optimizing local geometric structures to establish a one-to-one correspondence. Building upon this, the UnionCom algorithm by Cao et al. [11] enables semi-supervised topological alignment, relaxing GUMA's strict one-to-one mapping requirement. Liu et al. [12] proposed an alternative manifold alignment strategy called MMD-MA, which employs the Maximum Mean Discrepancy (MMD) metric for alignment. Additionally, the SCOT method [13] utilizes Gromov-Wasserstein distances to align multi-omic single-cell data. However, even when integration is performed without correspondences, hyperparameters are often tuned

⋆ Equal contribution
⋆⋆ Corresponding author: li.shen@pennmedicine.upenn.edu

using cell label validation, rendering the process semi-supervised. Demetci et al, [13] demonstrated that most methods fail to adapt to fully unsupervised settings when no orthogonal alignment information is available.

This work shares a similar idea with the MMD-MA [12], where we focus on determining coefficients that define mappings between datasets within a reproducing kernel Hilbert space (RKHS) framework. We begin by embedding each dataset into a high-dimensional space using kernel functions that quantify cell similarities. We aim to learn mappings for each dataset that transform the data into a shared lower-dimensional space for direct comparison. Our proposed method emphasizes both global alignment and local geometry preservation. Global alignment is achieved through optimal transport to minimize discrepancies in the shared latent space. To preserve local geometry, we employ the graph Laplacian, ensuring that the intrinsic structures of the datasets are maintained during alignment. Unlike most existing methods, which are not fully unsupervised and often depend on label information for hyperparameter tuning—even when they do not require one-to-one correspondence or labels for integrating modalities—our method achieves better alignment results by relying solely on the loss function. Moreover, it requires tuning only two hyperparameters, reducing the tuning time when applied to other datasets.

We evaluate our model against SCOT, MMD-MA, and UnionCom across four simulated and two real-world datasets in unsupervised and semi-supervised settings. Our Graph-Regularized Optimal Transport (GROT) algorithm matches the performance of state-of-the-art methods. The schematic design of our approach is illustrated in **Figure 1**. The source code for GROT is publicly available at `https://github.com/PennShenLab/GROT`.

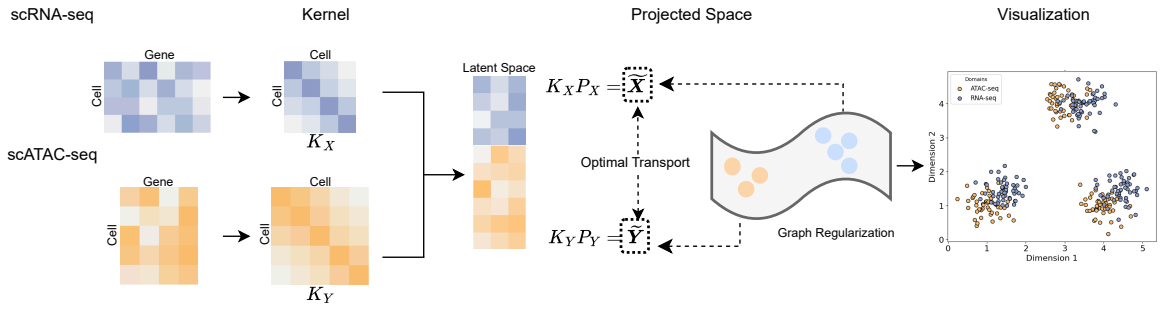

Fig. 1: Schematic Design of the Graph-Regularized Optimal Transport (GROT) Algorithm: The GROT algorithm begins by constructing a kernel matrix for each single-cell sequencing modality, such as scRNA-Seq and scATAC-Seq. It then learns a mapping matrix to project data from the RKHS space into a shared latent space. Global alignment within this latent space is facilitated through optimal transport. Additionally, graph regularization is applied to maintain the local structural integrity; this assumes that cells close in their original domains should exhibit similar structures in the shared space.

2 Method

2.1 Problem Formulation

We introduce a method to integrate single-cell datasets across different conditions or modalities. Let us consider two datasets, X and Y, with respective representations $X = \{x_1, \ldots, x_{n_x}\} \subset \mathscr{X}$ and $Y = \left\{y_1, \ldots, y_{n_y}\right\} \subset \mathscr{Y}$, where n_x and n_y denote the number cells in X and Y, respectively. We aim to uncover a shared manifold structure between X and Y without a priori correspondence between the datasets.

To achieve this, we first compute the intra-dataset kernels K_X and K_Y, which capture the internal structure of X and Y, respectively. As long as it is positive definite, each kernel corresponds to an implicit feature mapping $\phi_X : \mathscr{X} \to \mathscr{H}_X$ and $\phi_Y : \mathscr{Y} \to \mathscr{H}_Y$, where $\mathscr{H}_X$ and $\mathscr{K}_Y$ are the Reproducing Kernel Hilbert Spaces (RKHS) associated with K_X and K_Y. Subsequently, we seek mapping functions $f_X : \mathscr{X} \to \mathbb{R}^k$ and $f_Y : \mathscr{Y} \to \mathbb{R}^k$, where k is the dimensionality of the shared space. These functions are optimized so that the mapped representations $f_X(X)$ and $f_Y(Y)$ are well-aligned, thereby discovering the shared manifold structure.

2.2 Kernel Representation

To capture the intrinsic geometry of the datasets, we define the intra-dataset kernels K_X and K_Y using Gaussian kernel functions:

$$K_X(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_X^2}\right), \quad K_Y(y_i, y_j) = \exp\left(-\frac{\|y_i - y_j\|^2}{2\sigma_Y^2}\right) \tag{1}$$

where σ_X and σ_Y are bandwidth parameters specific to X and Y, respectively. These kernels define the feature maps $\phi_X : \mathscr{X} \to \mathscr{H}_X$ and $\phi_Y : \mathscr{Y} \to \mathscr{H}_Y$ into their respective Reproducing Kernel Hilbert Spaces (RKHS).

We adopt a data-driven approach to determine σ_X and σ_Y by setting each parameter to the mean of the pairwise Euclidean distances within the corresponding dataset. This heuristic adjusts the bandwidths to reflect the average spatial dispersion of the data points, thereby tuning the kernels to the specific scale of variability in each dataset.

2.3 Optimal Transport

For simplicity, we will use the notation: $\widetilde{X} = f_X(X) \in \mathbb{R}^{n_x \times k}, \widetilde{Y} = f_Y(Y) \in \mathbb{R}^{n_y \times k}$ to represent the mapped datasets. The Sinkhorn divergence between the projected representations $\widetilde{X}$ and $\widetilde{Y}$ is defined as:

$$\mathcal{L}_{\mathrm{OT}}(\widetilde{X}, \widetilde{Y}) = \mathrm{OT}_\varepsilon(\widetilde{X}, \widetilde{Y}) - \frac{1}{2}\left(\mathrm{OT}_\varepsilon(\widetilde{X}, \widetilde{X}) + \mathrm{OT}_\varepsilon(\widetilde{Y}, \widetilde{Y})\right) \tag{2}$$

where $\mathrm{OT}_\varepsilon(\cdot, \cdot)$ denotes the entropically regularized optimal transport cost between two distributions. Next, we define the entropic optimal transport cost between $\widetilde{X}$ and $\widetilde{Y}$. The cost is computed as:

$$\mathrm{OT}_\varepsilon(\widetilde{X}, \widetilde{Y}) = \min_{T \in \Pi(\mathrm{a},\ \mathrm{b})} \langle C, T \rangle + \varepsilon H(T) \tag{3}$$

where $\Pi(\mathbf{a}, \mathbf{b}) = \left\{ T \in \mathbb{R}_+^{n_x \times n_y} \mid T\mathbf{1}_{n_y} = \mathbf{a}, T^\top \mathbf{1}_{n_x} = \mathbf{b} \right\}$. The matrix $T \in \mathbb{R}_+^{n_x \times n_y}$ is the transport plan matrix, representing the amount of mass transported from $\widetilde{x}_i$ to $\tilde{y}_j$. The cost matrix $C \in \mathbb{R}^{n_x \times n_y}$ quantifies the pairwise distances between the projected samples. Each element is defined as: $C_{ij} = \|\widetilde{x}_i - \tilde{y}_j\|_2^2$. The parameter $\varepsilon > 0$ is the entropic regularization parameter that smooths the optimization problem and $H(T) = -\sum_{i=1}^{n_x} \sum_{j=1}^{n_y} T_{ij} (\log T_{ij} - 1)$ is the entropy of the transport plan T. The marginal distributions $\mathbf{a} \in \mathbb{R}^{n_x}$ and $\mathbf{b} \in \mathbb{R}^{n_y}$ are typically uniform distributions over the samples: $\mathbf{a} = \frac{1}{n_x}\mathbf{1}_{n_x}, \quad \mathbf{b} = \frac{1}{n_y}\mathbf{1}_{n_y}$, where $\mathbf{1}_{n_x}$ and $\mathbf{1}_{n_y}$ are vectors of ones with lengths n_x and n_y, respectively. We utilize optimal transport over maximum mean discrepancy due to its benefits, such as non-vanishing gradients and other theoretical advantages [14].

2.4 Graph Laplacian Regularization

To capture the local geometric structures of the datasets X and Y, we construct graph Laplacians based on the k-nearest neighbor relationships defined through the Gaussian kernels. These Laplacians serve as regularizers for the mapping functions, enforcing smoothness by ensuring that nearby data points in the RKHS spaces $\mathscr{H}_X$ and $\mathscr{H}_Y$ remain close in the shared latent space.

For each dataset, we begin by constructing a k-nearest neighbor graph in the RKHS. Specifically, for dataset X, we identify the set of k-nearest neighbors for each feature-mapped data point $\phi_X(x_i)$, denoted as $\mathcal{N}_k(\phi_X(x_i))$, based on the distance metric in $\mathscr{H}_X$. The adjacency matrix $W_X \in \mathbb{R}^{n_x \times n_x}$ is then defined with entries:

$$[W_X]_{ij} = \begin{cases} 1, & \text{if } \phi_X(x_j) \in \mathcal{N}_k(\phi_X(x_i)) \text{ or } \phi_X(x_i) \in \mathcal{N}_k(\phi_X(x_j)) \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

Similarly, for dataset Y, we construct the adjacency matrix $W_Y \in \mathbb{R}^{n_y \times n_y}$. Next, we compute the degree matrices D_X and D_Y, which are diagonal matrices where each diagonal entry represents the sum of the edge weights connected to a node:$[D_X]_{ii} = \sum_{j=1}^{n_x} [W_X]_{ij}$, $[D_Y]_{ii} = \sum_{j=1}^{n_y} [W_Y]_{ij}$. The graph Laplacians are then defined as the difference between the degree and adjacency matrices:

$$L_X = D_X - W_X, \quad L_Y = D_Y - W_Y \tag{5}$$

To regularize the projected representations $\widetilde{X}$ and $\widetilde{Y}$, we introduce smoothness terms based on the Laplacian quadratic form. Specifically, the smoothness term for $\widetilde{X}$ is given by:

$$\frac{1}{2} \sum_{i=1}^{n_x} \sum_{j=1}^{n_x} [W_X]_{ij} \|\widetilde{x}_i - \widetilde{x}_j\|^2 = \operatorname{Tr}\left(\widetilde{X}^\top L_X \widetilde{X}\right) \tag{6}$$

where $\widetilde{x}_i$ denotes the i-th row of $\widetilde{X}$. This expression encourages neighboring points in the original data space to have similar representations in the latent space, promoting smoothness in the mappings.

2.5 GROT Algorithm

To integrate the datasets X and Y into a shared latent space, we propose the Graph-Regularized Optimal Transport (GROT) algorithm. Our objective is to find the mapping $f_X : \mathscr{X} \to \mathbb{R}^k, f_Y : \mathscr{Y} \to \mathbb{R}^k$ that maps the data into a common k-dimensional space, effectively aligning their underlying manifold structures. The existence of such mapping is guaranteed by the representer theorem, which states that the optimal mappings can be expressed as finite linear combinations of the kernel functions:

$$[f_X]_j (x) = \sum_{i=1}^{n_x} \alpha_X^{ij} K_X (x_i, x) \tag{7}$$

where α_X^{ij} are the learned coefficients, K_X is the kernel function, and x_i are the samples from X. The coefficients α_X^{ij} are then organized into the matrix $P_X \in \mathbb{R}^{n_x \times k}$, and similarity for Y we have $P_Y \in \mathbb{R}^{n_y \times k}$. Thus, the final mapped representations are given by:

$$\widetilde{X} = K_X P_X, \quad \widetilde{Y} = K_Y P_Y \tag{8}$$

The optimization problem for the GROT algorithm is formulated as:

$$\min_{P_X, P_Y} \mathcal{L}_{\mathrm{OT}}(\widetilde{X}, \widetilde{Y}) + \lambda \left(\mathrm{Tr}(\widetilde{X}^\top L_X \widetilde{X}) + \mathrm{Tr}(\widetilde{Y}^\top L_Y \widetilde{Y}) \right) + \rho \left(\|P_X^\top K_X P_X - I_k\|_F^2 + \|P_Y^\top K_Y P_Y - I_k\|_F^2 \right) \tag{9}$$

The first term $\mathcal{L}_{\mathrm{OT}}(\widetilde{X}, \widetilde{Y})$ aligns the global distributions of the datasets in the latent space using sinkhorn divergence. Minimizing the Sinkhorn divergence between the latent space representations ensures that the overall structures of X and Y are closely matched after projection. This captures global structural similarities and facilitates the discovery of shared manifold features between the datasets.

The second term $\mathrm{Tr}\left(\widetilde{X}^\top L_X \widetilde{X}\right)$ and $\mathrm{Tr}\left(\widetilde{Y}^\top L_Y \widetilde{Y}\right)$ is used preserve the local geometric structures inherent in each dataset. These terms penalize the weighted differences between neighboring points in the latent space. Doing so encourages neighboring points in the $\mathscr{H}_{\mathscr{X}}, \mathscr{H}_{\mathscr{Y}}$ to remain close in the latent space.

To prevent degenerate solutions and ensure that the mappings retain meaningful structure, we impose orthogonality constraints on the mapping matrices through the terms $\left(\|P_X^\top K_X P_X - I_k\|_F^2\right)$ and $\left(\|P_Y^\top K_Y P_Y - I_k\|_F^2\right)$.

By jointly optimizing this objective function, the GROT algorithm effectively balances global alignment and local structure preservation while ensuring that the projections are meaningful and well-behaved.

3 Dataset

3.1 Simulated Datasets

We evaluated the GROT algorithm using four simulated datasets: three from Liu et al. [12] and one additional dataset from Demetci et al. [13]. The datasets by Liu et al. [12] comprise 300 samples each, projected non-linearly to dimensions of 1000 and 2000. Specifically, the first dataset presents a branch structure in a two-dimensional space, the second a Swiss roll

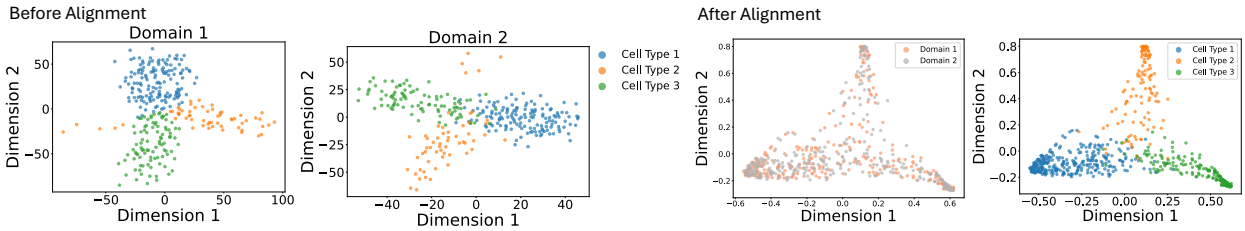

Fig. 2: Visualization of Simulation Dataset 1: From left to right, the first domain prior to alignment, the second domain prior to alignment, post-alignment visualization colored by domain, and post-alignment visualization colored by cell type.

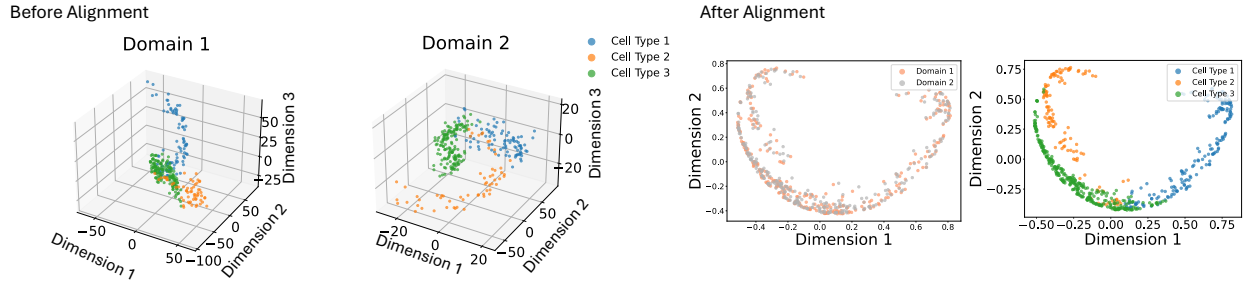

Fig. 3: Visualization of Simulation Dataset 2: From left to right, the first domain prior to alignment, the second domain prior to alignment, post-alignment visualization colored by domain, and post-alignment visualization colored by cell type.

in three-dimensional space, and the third a circular frustum in three dimensions. The dataset from Demetci et al. [13] includes 5000 cells, each characterized by gene measurements at two scales, 50 and 500. All datasets contain three distinct cell types. In all simulations, we Z-score normalize the features before running the alignment algorithms as in the study by Liu et al. [12] and Demetci et al. [13]. Visualization of each dataset before alignment are shown in Figure 2,3,4,5.

3.2 Real-World Datasets

We utilized two real-world datasets to test the model's performance. The first one comes from Cheow et al. [15], generated using the scGEM assay to evaluate our model's performance. These datasets simultaneously profile gene expression and DNA methylation in primary lung adenocarcinomas and human fibroblasts. The gene expression dataset comprises 177 samples with 34 features, while the DNA methylation dataset includes 177 samples with 27 features. The second dataset comes from Chen et al. [16]. It is generated using droplet-based single-nucleus chromatin accessibility and mRNA expression sequencing (SNARE-seq), which links chromatin accessibility with gene expression. The resulting data matrices for the SNARE-seq dataset include 1047 samples with 19 features for ATAC-seq and 1047 samples with 10 features for RNA-seq. Both datasets are unit normalized, consistent with the preprocessing steps described in the study by Demetci et al. [13]. Visualization of each dataset before alignment are shown in Figure 6,7.

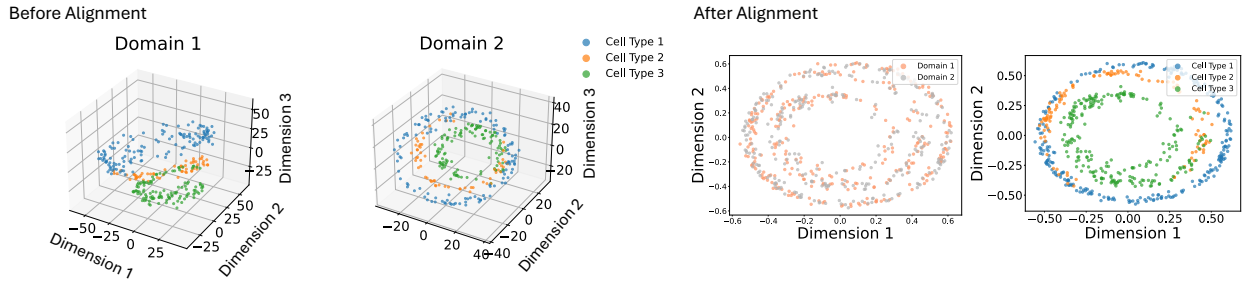

Fig. 4: Visualization of Simulation Dataset 3: From left to right, the first domain prior to alignment, the second domain prior to alignment, post-alignment visualization colored by domain, and post-alignment visualization colored by cell type.

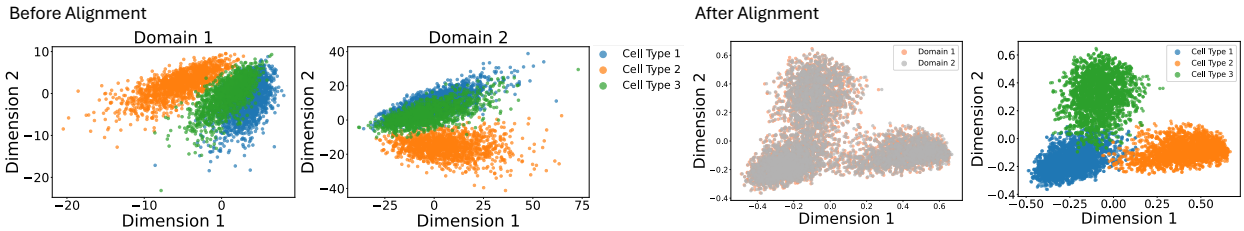

Fig. 5: Visualization of Simulation Dataset 4: From left to right, the first domain prior to alignment, the second domain prior to alignment, post-alignment visualization colored by domain, and post-alignment visualization colored by cell type.

4 Experiment Setup

4.1 Evaluation Metric

In our study, we benchmarked our alignment method, GROT, against three prominent single-cell alignment techniques: MMD-MA [12], UnionCom [11], and SCOT [13]. SCOT has a second version, SCOT-v2, specifically targeting datasets with disproportionate cell types. However, according to the literature, SCOT-v2 underperforms relative to SCOT-v1 in datasets lacking extreme cell type disproportionality [13]. Therefore, we opted to compare it against the original SCOT version. Additionally, we observed discrepancies in the reported results for SCOT between the sc-GEM and sc-SNARE datasets across the two versions. In our analysis, we utilized the most recent results for these datasets.

We evaluated alignment accuracy using the fraction of samples closer than the true match (FOSCTTM) metric proposed by Liu et al. [12]. This metric calculates the Euclidean distances from a selected sample in one domain to all samples in another, determining the proportion of samples closer to the selected sample than its true counterpart. Averaging these proportions across all samples yields an overall metric, with lower values indicating superior alignment. We also assessed the quality of label assignments using the label transfer accuracy (LTA) metric from Cao et al. [11]. LTA measures the effectiveness of label transfers between domains by training a k-NN classifier on one domain and using it to predict labels in the other, with higher values indicating more accurate label assignments.

4.2 Hyperparameter Tunning

Commonly in the literature, even when authors do not utilize a one-to-one correspondence between cells, hyperparameters are determined by metrics such as label accuracy or FOSCTTM. However, evaluating these metrics necessitates known true labels. Therefore, our experiments use two distinct hyperparameter tuning strategies across all methods. The first experiment employed a fully unsupervised approach, selecting hyperparameters based on the loss function. The FOSCTTM and LTA are reported. The second experiment employed a semi-supervised approach, using the hyperparameters with the lowest FOSCTTM scores. This approach can be beneficial for identifying domains where accurate cell type labels need to be assigned. The lowest FOSCTTM and the LTA with the same hyperparameter are reported using the semi-supervised approach.

In the unsupervised setting, hyperparameters were selected based on achieving the lowest total loss. In the semi-supervised setting for GROT, we fixed the latent dimension to 5 and set the number of nearest neighbors to 10, requiring tuning these two hyperparameters. We established a hyperparameter grid as follows: weights $\lambda_1 \in \{1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$ and $\rho \in \{10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}, 10^{-9}\}$. We divided the dataset into training and validation sets in a 2:1 ratio, allocating two-thirds of the samples for training and one-third for validation.

5 Results

5.1 Unsupervised Self-Tuning Result

In this experiment, we align four simulated datasets and two real-world datasets. Hyperparameters are tuned based on the unsupervised setting described in Section 4.2. We report the FOSCTTM and LTA metrics. Figures 2 through 7 visualize the latent space alignment, colored by domain or cell type, demonstrating that the domains are correctly aligned according to their corresponding cell types. The mean FOSCTTM values are summarized in Table 1, and the LTA values are presented in Table 2, where the first domain is used to train a classifier to predict labels in the second domain. GROT consistently achieves the lowest mean FOSCTTM and the highest LTA. It ranks second with a slight decrease in FOSCTTM on Simulation Dataset 3 and LTA on Simulation Dataset 1. GROT shows significant improvement in LTA on the two real-world datasets, highlighting its applicability for label prediction and assignment on previously unseen datasets. In contrast, MMD-MA failed to

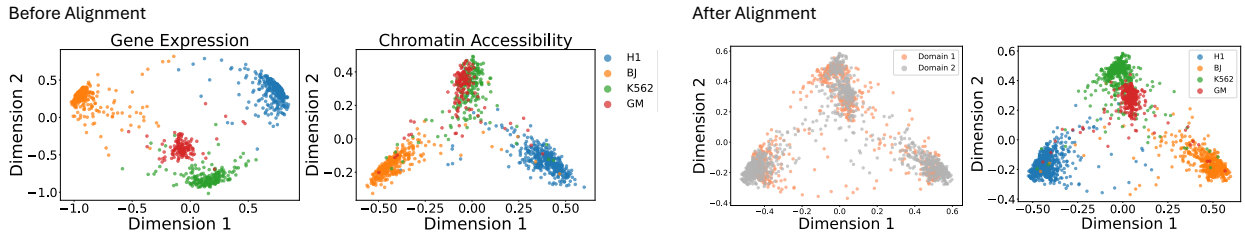

Fig. 6: Visualization of sc-SNARE dataset: From left to right, the first domain prior to alignment, the second domain prior to alignment, post-alignment visualization colored by domain, and post-alignment visualization colored by cell type.

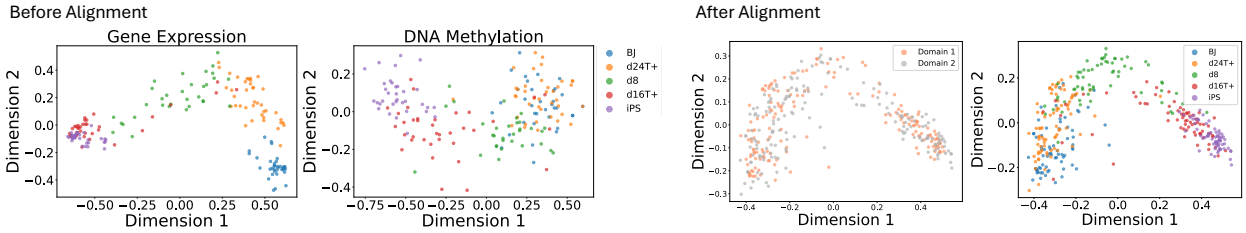

Fig. 7: Visualization of sc-GEM dataset: From left to right, the first domain prior to alignment, the second domain prior to alignment, post-alignment visualization colored by domain, and post-alignment visualization colored by cell type.

align Simulation Dataset 3 and sc-GEM, while UnionCom failed to align Simulation Dataset 3 and the two real-world datasets.

Table 1: Alignment performance by FOSCTTM under unsupervised setting (The lower the better).

	Simulation 1	Simulation 2	Simulation 3	Synthetic RNA-seq	scGEM	SNAREseq
SCOT	0.088	0.025	**0.009**	0.001	0.209	0.218
MMD-MA	0.125	0.012	0.739	0.384	0.437	0.473
UnionCom	0.091	0.028	0.684	0.028	0.691	0.510
GROT(Ours)	**0.076**	**0.008**	0.012	**1e-6**	**0.203**	**0.154**

Table 2: Alignment performance by label transfer accuracy ($k = 5$) under unsupervised setting (The higher the better).

	Simulation 1	Simulation 2	Simulation 3	Synthetic RNA-seq	scGEM	SNAREseq
SCOT	**0.977**	0.977	0.950	0.996	0.423	0.852
MMD-MA	0.897	0.957	0.700	0.506	0.237	0.412
UnionCom	0.947	0.947	0.133	0.948	0.107	0.288
GROT(Ours)	0.970	**0.983**	**0.956**	**0.997**	**0.632**	**0.975**

5.2 Semi-Supervised Tuning Result

In this experiment, we align four simulated datasets and two real-world datasets. The hyperparameters are tuned based on the unsupervised setting described in Section 4.2. The mean FOSCTTM values are summarized in Table 3, and the LTA values are reported in Table 4 when the first domain is used to train a classifier to predict labels in the second domain. GROT consistently achieves the lowest mean FOSCTTM and the highest LTA. GROT only ranks third with a slight decrease in LTA on Simulation Dataset 3. Other competitive methods show some performance increase, corresponding to hyperparameters chosen

in the semi-supervised version. Despite this, GROT demonstrates consistent results in both unsupervised and semi-supervised settings, showcasing its robustness.

Table 3: Alignment performance by FOSCTTM scores under semi-supervised setting (The lower the better).

	Simulation 1	Simulation 2	Simulation 3	Synthetic RNA-seq	scGEM	SNAREseq
SCOT	0.085	0.022	**0.009**	0.001	0.192	0.150
MMD-MA	0.124	0.023	0.012	0.112	0.201	0.150
UnionCom	0.083	0.016	0.152	0.038	0.209	0.265
GROT(Ours)	**0.071**	**0.006**	**0.009**	**1e-7**	**0.191**	**0.147**

Table 4: Alignment performance by label transfer accuracy ($k = 5$) under semi-supervised setting (The higher the better).

	Simulation 1	Simulation 2	Simulation 3	Synthetic RNA-seq	scGEM	SNAREseq
SCOT	0.937	0.977	**0.957**	**0.998**	0.576	0.982
MMD-MA	0.89	0.783	0.947	0.706	0.588	0.942
UnionCom	0.96	0.62	0.613	0.997	0.582	0.423
GROT(Ours)	**0.970**	**0.990**	0.943	**0.998**	**0.621**	**0.985**

6 Discussion

This study presents GROT, an algorithm that performs unsupervised single-cell multi-omics data integration. GROT employs optimal transport for global alignment across modalities and graph regularization to maintain local structure within each modality. Our approach shares similarities with MMD-MA in transforming data into reproducing kernel Hilbert spaces and finding mapping matrices from each space. Additionally, we introduce a local preservation term via graph regularization to better preserve local geometry. We evaluated GROT's performance in both unsupervised and semi-supervised settings, benchmarking it against current state-of-the-art methods and demonstrating superior results, indicating its robustness. In future work, we plan to explore the use of coupling matrices within the optimal transport framework. Additionally, we aim to extend our current method to enable the automatic alignment of more than two modalities.

Acknowledgements: This work is supported in part by NIH Grants R01 AG071470, U19 AG074879, U01 AG066833 and U01 AG068057.

References

[1] Lukas Heumos et al. "Best practices for single-cell analysis across modalities". In: *Nature Reviews Genetics* 24.8 (2023), pp. 550–572.

[2] Ashraful Haque et al. "A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications". In: *Genome medicine* 9 (2017), pp. 1–12.

[3] Fiorella C Grandi et al. "Chromatin accessibility profiling by ATAC-seq". In: *Nature protocols* 17.6 (2022), pp. 1518–1552.

[4] Bingjun Li and Sheida Nabavi. "scGEMOC, A Graph Embedded Contrastive Learning Single-cell Multiomics Clustering Model". In: *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE. 2023, pp. 2075–2080.

[5] Geert-Jan Huizing et al. "Paired single-cell multi-omics data integration with Mowgli". In: *Nature Communications* 14.1 (2023), p. 7711.

[6] Tim Stuart et al. "Comprehensive integration of single-cell data". In: *cell* 177.7 (2019), pp. 1888–1902.

[7] Nikolas Barkas et al. "Joint analysis of heterogeneous single-cell RNA-seq dataset collections". In: *Nature methods* 16.8 (2019), pp. 695–698.

[8] Keren Bahar Halpern et al. "Paired-cell sequencing enables spatial gene expression mapping of liver endothelial cells". In: *Nature biotechnology* 36.10 (2018), pp. 962–970.

[9] Matthew Amodio and Smita Krishnaswamy. "MAGAN: Aligning biological manifolds". In: *International conference on machine learning*. PMLR. 2018, pp. 215–223.

[10] Zhen Cui et al. "Generalized unsupervised manifold alignment". In: *Advances in Neural Information Processing Systems* 27 (2014).

[11] Kai Cao et al. "Unsupervised topological alignment for single-cell multi-omics integration". In: *Bioinformatics* 36.Supplement_1 (2020), pp. i48–i56.

[12] Jie Liu et al. "Jointly embedding multiple single-cell omics measurements". In: *Algorithms in bioinformatics:... International Workshop, WABI..., proceedings. WABI (Workshop)*. Vol. 143. NIH Public Access. 2019.

[13] Pinar Demetci et al. "SCOT: single-cell multi-omics alignment with optimal transport". In: *Journal of computational biology* 29.1 (2022), pp. 3–18.

[14] Jean Feydy et al. "Interpolating between optimal transport and mmd using sinkhorn divergences". In: *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR. 2019, pp. 2681–2690.

[15] Lih Feng Cheow et al. "Single-cell multimodal profiling reveals cellular epigenetic heterogeneity". In: *Nature methods* 13.10 (2016), pp. 833–836.

[16] Song Chen, Blue B Lake, and Kun Zhang. "High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell". In: *Nature biotechnology* 37.12 (2019), pp. 1452–1457.