OXFORD

Briefings in Bioinformatics, 2025, **26(1)**, bbae711
https://doi.org/10.1093/bib/bbae711
Problem Solving Protocol

Integrating scRNA-seq and scATAC-seq with inter-type attention heterogeneous graph neural networks

Lingsheng Cai [1], Xiuli Ma [1,*], Jianzhu Ma [2,3,*]

[1]State Key Laboratory of General Artificial Intelligence, School of Intelligence Science and Technology, Peking University, 100871 Beijing, China
[2]Department of Electronic Engineering, Tsinghua University, 100084 Beijing, China
[3]Institute for AI Industry Research, Tsinghua University, 100084 Beijing, China

*Corresponding authors. Xiuli Ma, E-mail: xlma@pku.edu.cn; Jianzhu Ma, E-mail: majianzhu@tsinghua.edu.cn

Abstract

Single-cell multi-omics techniques, which enable the simultaneous measurement of multiple modalities such as RNA gene expression and Assay for Transposase-Accessible Chromatin (ATAC) within individual cells, have become a powerful tool for deciphering the intricate complexity of cellular systems. Most current methods rely on motif databases to establish cross-modality relationships between genes from RNA-seq data and peaks from ATAC-seq data. However, these approaches are constrained by incomplete database coverage, particularly for novel or poorly characterized relationships. To address these limitations, we introduce **s**ingle-**c**ell **M**ulti-omics Integration (scMI), a heterogeneous graph embedding method that encodes both cells and modality features from single-cell RNA-seq and ATAC-seq data into a shared latent space by learning cross-modality relationships. By modeling cells and modality features as distinct node types, we design an inter-type attention mechanism to effectively capture long-range cross-modality interactions between genes and peaks. Benchmark results demonstrate that embeddings learned by scMI preserve more biological information and achieve comparable or superior performance in downstream tasks including modality prediction, cell clustering, and gene regulatory network inference compared to methods that rely on databases. Furthermore, scMI significantly improves the alignment and integration of unmatched multi-omics data, enabling more accurate embedding and improved outcomes in downstream tasks.

Keywords: single-cell multi-omics integration; heterogeneous graphs; deep learning; modality prediction

Introduction

Single-cell sequencing techniques, such as single-cell RNA sequencing (scRNA-seq) and single-cell Assay for Transposase-Accessible Chromatin sequencing (scATAC-seq), have revolutionized the exploration of cellular heterogeneity. These techniques allow scientists to analyze genomics, transcriptomics, epigenomics, and other molecular characteristics of cells in unprecedented detail, revealing crucial insights in diverse fields such as cancer biology, immuno-oncology, and neuroscience [1]. Traditional single-cell studies typically concentrate on a single modality, such as scRNA-seq or scATAC-seq data, which can lead to an incomplete view of cellular information. Those approaches may overlook crucial interactions and information, limiting the holistic understanding of cellular systems [2].

Advancements in sequencing technology have enabled single-cell multi-omics techniques to simultaneously measure multiple modalities within individual cells [3]. These techniques facilitate a more comprehensive analysis of cellular processes by capturing diverse modalities from the same cell. In particular, the integration of data from scRNA-seq and scATAC-seq technologies has become a critical topic of research [4]. This integration necessitates sophisticated representation learning strategies to effectively interpret the vast and complex multi-omics datasets. The representations enable simultaneous profiling of transcriptomic and epigenomic features within individual cells, providing a comprehensive and integrated perspective on cell phenotypes and regulatory mechanisms. For instance, integrated representations of scRNA-seq and scATAC-seq data can be used to identify different cell types, gaining deeper insights into cellular heterogeneity. Additionally, gene regulatory networks (GRNs) inferred from integrated representations help understand the regulatory mechanisms driving gene expression and cellular functions [5].

Several approaches have been developed to integrate scRNA-seq and scATAC-seq data. These methods typically involve aligning or correlating different modalities at the single-cell level to create a unified representation that captures the full spectrum of cell phenotypes. Typical solutions can be classified into three categories: correlation-based methods, such as canonical correlation analysis [6] and matrix factorization [7], which identify shared components across modalities for integration; joint dimensionality reduction methods, like multi-omics factor analysis (MOFA) [8], which reduce the dimensionality of multiple omics datasets while preserving inter-modality relationships; and neighbor-based approaches, such as Seurat v4's weighted nearest neighbor algorithm [9], which identifies mutual nearest neighbors across modalities to facilitate cohesive analysis. While these approaches show promise, they often rely on predefined relationships between modalities, limiting their ability to uncover complex

Received: September 11, 2024. **Revised:** December 7, 2024. **Accepted:** January 2, 2025

interactions. This highlights the need for more robust and flexible integration techniques.

Recently, heterogeneous graphs have emerged as a powerful tool in multimodal learning. In the context of single-cell multi-omics analysis, researchers encode cells and modality features from different omics, such as genes and chromatin accessibility peaks, as distinct types of nodes and model their relationships within a heterogeneous graph. By capturing the complex relationships and interactions between different types of nodes, heterogeneous graphs facilitate the discovery of intricate biological insights that are otherwise challenging to uncover using traditional methods. Heterogeneous graph attention network models show their effectiveness in node classification tasks, particularly in biological contexts [10]. Furthermore, DeepMAPS employs a heterogeneous graph transformer to analyze cells and genes from single-cell multi-omics data, aiming to infer cell-type-specific gene regulatory networks [11]. However, DeepMAPS integrates multi-omics data primarily based on prior knowledge instead of utilizing heterogeneous graphs to capture the complex cross-modality relationships. Some researchers have begun incorporating modality features, such as genes and peaks, into heterogeneous graphs using prior knowledge. For example, SIMBA links genes and peaks through motifs and utilizes heterogeneous graph learning to generate general embeddings for downstream analysis [12]. Unlike traditional methods, heterogeneous graphs provide greater flexibility by avoiding the need for direct projection of ATAC-seq features onto RNA-seq genes. Instead, they construct heterogeneous graphs based on cross-modality relationships from databases, allowing the integration of multi-omics information within a heterogeneous graph framework. However, the graph construction in these methods heavily relies on prior knowledge, specifically by identifying active transcription factor (TF) binding sites using known motifs.

While these approaches have merit, their effectiveness is compromised when motif information in existing databases is incomplete. Additionally, the inherent cellular heterogeneity at the single-cell level can lead to distinct motif expression patterns across different cell subpopulations, complicating motif enrichment analysis and potentially resulting in an incomplete understanding of TFs activity. In those cases, the nodes representing genes and peaks in the heterogeneous graph are not directly connected, forcing information to be propagated solely through intermediary cell nodes. This limitation hampers the ability of these methods to effectively capture long-range dependencies and uncover hidden cross-modality relationships. As a result, the integration of data from scRNA-seq and scATAC-seq remains challenging, with current methods often failing to discover novel cross-modality interactions beyond those already encoded in the motif databases. Consequently, there is a need for more advanced approaches that can autonomously detect and model these complex relationships to improve the integration of multi-omics data.

In this paper, to tackle the above challenges and limitations, we propose an inter-type relationship deep learning method on heterogeneous graphs for **s**ingle-**c**ell **M**ulti-omics **I**ntegration (scMI). scMI models cells, genes and peaks from multi-omics datasets within a unified heterogeneous graph. By leveraging a frequency-based multiple-restart Random Walk (RW) strategy, scMI ensures the consistent capture of both structural patterns and data features inherent in the heterogeneous graph. Furthermore, we design an inter-type attention mechanism to capture long-range topological structures across different node types within heterogeneous graphs, effectively enhancing the discovery of accurate cross-modality relationships between peaks and genes. Compared to existing heterogeneous graph methods, scMI excels by autonomously learning and optimizing cross-modality relationships without relying on extensive prior knowledge, making it more adaptable to complex, novel datasets. In our framework, we combine scMI with existing multi-omics integration methods, particularly for unmatched multi-omics datasets, and leverage the inter-type relationship deep learning method to improve cell alignment in these datasets, thus enhancing the accuracy and robustness of multi-omics data integration. To validate scMI, we conduct extensive experiments on real-world datasets involving modality prediction, cell clustering, and GRN inference. Results demonstrate that scMI achieves comparable or superior performance compared to previous methods.

Materials and methods

Overview

Figure 1(a) depicts the overview of our proposed scMI framework. scMI starts by preprocessing scRNA-seq and scATAC-seq data, filtering out genes and peaks that are not expressed in a substantial portion of the dataset. The preprocessed data are then used to construct a heterogeneous graph, modeling the intricate relationships between cells and modality features. To handle the data's complexity, a frequency-based subgraph sampling method generates smaller, more manageable subgraphs for analysis. scMI then jointly learns low-dimensional embeddings for cells and features, employing an inter-type attention mechanism to highlight key relationships. Finally, scMI incorporates a collaborative learning strategy that simultaneously optimizes the embeddings with downstream tasks, such as cell clustering and GRN inference, for further biological analysis.

Datasets

Peripheral blood mononuclear cells

Peripheral blood mononuclear cells (PBMCs) consist of a heterogeneous mix of immune cells, including lymphocytes (such as T cells, B cells, and NK cells), monocytes, and dendritic cells, each playing a critical role in the immune system. These cells are isolated from peripheral blood and are widely used in research. Three PBMC datasets, PBMC 3k, PBMC 7k, and PBMC 10k datasets, derived from healthy donors, are selected for single-cell multi-omics integration.

The PBMC 3k (https://www.10xgenomics.com/datasets/pbmc-from-a-healthy-donor-granulocytes-removed-through-cell-sorting-3-k-1-standard-2-0-0) and 10k (https://www.10xgenomics.com/datasets/pbmc-from-a-healthy-donor-granulocytes-removed-through-cell-sorting-10-k-1-standard-2-0-0) datasets are available at 10X Genomics official website. The two datasets are processed using the 10x Genomics Chromium platform, and annotated with the Cell Ranger pipeline. The PBMC 7k datasets can be obtained from NCBI GEO under accession GSE156478. All datasets provide valuable insights into immune cell diversity and serve as essential resources for advancing the field of single-cell multi-omics analysis.

Bone marrow mononuclear cells

Bone marrow mononuclear cells (BMMCs) are a diverse population of cells found in the bone marrow, consisting primarily of lymphocytes, monocytes, and hematopoietic stem and progenitor cells. These cells are crucial for studying hematopoiesis, immune function, and various hematological disorders. BMMCs serve as an important model for understanding the complex processes of blood cell formation and the regulation of immune responses.

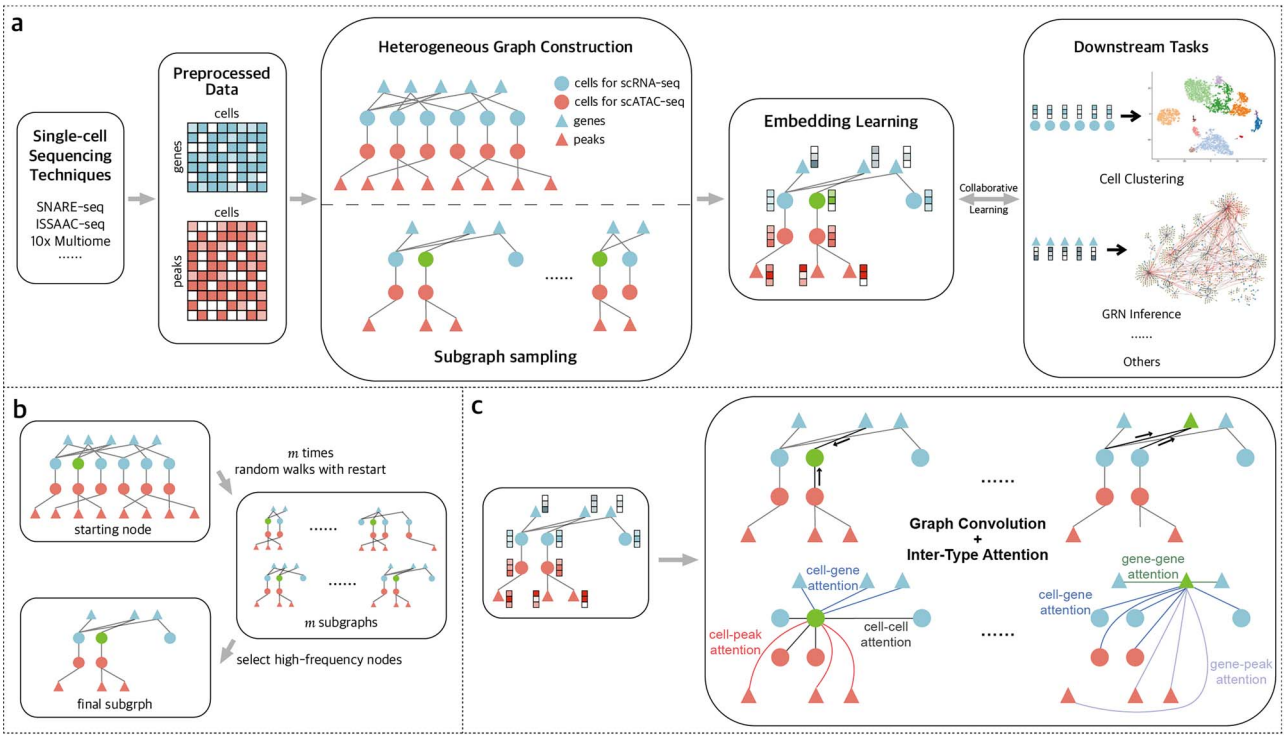


Figure 1. The architecture of scMI. (a) The overview of scMI. (b) A frequency-based RW algorithm with restart to sample subgraphs. The algorithm samples m subgraphs starting from the same node, and the final subgraph is obtained by filtering based on frequency. (c) Representation learning with inter-type attention heterogeneous graph neural networks. Graph convolutions preserve the topological structure information of the subgraph, while the inter-type attention mechanism aims to capture the implicit cross-modality relationships within the multi-omics data.

The BMMC dataset, sourced from the NeurIPS 2021 competition and available from NCBI GEO under accession GSE194122, served as a benchmark for evaluating model performance in modality prediction and data integration tasks [13].

The GSE194122 dataset includes data from 4 sites and 10 donors. In our study, to ensure the generalization ability of the model, we selected three datasets—s1d2, s2d1, and s3d6—for our experiments, where s stands for sequencing site and d stands for donor number.

Human brain tissue

The flash-frozen Human healthy Brain Tissue dataset was obtained from 10X Genomics official website(https://www.10xgenomics.com/datasets/frozen-human-healthy-brain-tissue-3-k-1-standard-2-0-0). Nuclei were isolated from a 2605-mg section of the brain and prepared following established protocols for single-cell multiome ATAC and gene expression sequencing. Paired ATAC and gene expression libraries were generated and sequenced to provide key insights into chromatin accessibility and gene expression in human cerebellar cells and is a valuable resource for single-cell multi-omics analysis.

Mouse embryonic and adult brain cortex

The Mouse Embryonic Brain Cortex (MEBC) dataset, available under the accession number GSE126074, was generated by Chen *et al*. using SNARE-seq, a pioneering single-cell sequencing technique that simultaneously captures scRNA-seq and scATAC-seq data from the same cell [14]. We select joint profiles of 5081 cells from neonatal mouse brain cortices for analysis.

Additionally, we include a dataset of cells from the adult Mouse Brain Cortex, which can be accessed in BioStudies under the ID E-MTAB-11264. This dataset provides further insights into the cellular landscape of the mature mouse brain cortex.

Mouse skin tissue and retina cells

The Mouse Skin Tissue (MST) dataset, available under the accession number GSE140203, was obtained by Ma *et al*., offering valuable insights into the regulatory mechanisms that govern skin cell differentiation and function [15].

This dataset provides a detailed view of the transcriptional programs and epigenetic states within mouse skin tissue by integrating scRNA-seq and scATAC-seq data. The combined analysis of RNA-seq and chromatin accessibility data facilitates a deeper understanding of how chromatin potentially influences gene expression during skin development and maintenance.

The adult Mouse Retina Cells dataset, available under the accession number GSE201402, was generated using the 10x Genomics Multiome ATAC+RNA kit to capture both scRNA-seq and scATAC-seq data from an adult mouse retina sample. This dataset offers a comprehensive view of the gene expression and chromatin accessibility profiles, enabling the study of various neuronal and non-neuronal cell types in the retina.

The detailed statistical information for all datasets above is provided in Supplementary Table 1.

Building single-cell multi-omics heterogeneous graphs using scRNA-seq and scATAC-seq data

The construction of a single-cell multi-omics heterogeneous graph network involves integrating four distinct types of nodes: cells from scRNA-seq data, cells from scATAC-seq data, genes, and chromatin accessibility peaks, as shown in Fig. 1(a). In the graph, edges are established based on specific biological relationships. First, corresponding cells from scRNA-seq and scATAC-seq data are connected by edges, representing the same underlying biological cell across different modalities. Additionally, edges are formed between scRNA-seq cells and genes that are

highly expressed in those cells, reflecting the transcriptional activity captured in the scRNA-seq data. To identify highly expressed genes, we first select highly variable genes within the dataset. For each highly variable gene, we determine its expression threshold by calculating the 95th percentile of its expression values across all cells. A gene is considered highly expressed in a cell if its expression in that cell exceeds this threshold, thereby establishing an edge between the cell and the corresponding gene to represent transcriptional activity. Similarly, scATAC-seq cells are connected by edges to chromatin regions (peaks) that are accessible within those cells, indicating potential regulatory regions. This graph structure allows for the simultaneous representation and integration of multiple omics layers, enabling the exploration of complex cellular relationships and regulatory mechanisms within a unified framework.

For multi-omics datasets lacking cell matching information, however, the edges between cells are unknown. To facilitate the integration of unmatched datasets, we employ established algorithms, Seurat v4, GLUE [16], and bindSC [17], to perform initial pairing among cells. All three methods ultimately map cells from different modalities into a shared latent space. In this space, biologically similar cells originating from different modalities are positioned near each other. To establish connections in our heterogeneous graph, we identify the nearest scATAC-seq cell for each scRNA-seq cell in this shared space, based on the cosine similarity of their embeddings. An edge is then established between these paired cells in the heterogeneous graph, facilitating the integration of multi-omics data in the absence of direct cell-to-cell correspondence.

Subgraph sampling via Frequency-based Random Walk with Restart

In heterogeneous graph representation learning, sampling subgraphs plays a crucial role due to the complexity and size of real-world graphs. Traditional RW methods randomly select nodes and their neighbors, reducing computational complexity while preserving the structural features of the original graph. However, in biological heterogeneous graphs, the nodes are more massive and the graphs are sparser. Traditional subgraph sampling methods exhibit significant uncertainty, which can introduce noise and instability to subsequent learning processes. In addition, disparities in the quantities of different edge types lead to limitations like uneven coverage across node types, biases towards more prevalent edge types, and inefficiencies. To address this, as illustrated in Fig. 1(b), we introduce a Frequency-based Random Walks with Restart (FRWR) algorithm to sample subgraphs. Starting from a node, the algorithm performs multiple restart RWs based on edge weights to generate m distinct subgraphs, promoting path diversification. For each type of node, the k most frequently sampled nodes, along with their associated edges, form the final subgraph, which is subsequently used for representation learning. This algorithm reliably captures the most salient aspects of the graph's topology and feature distribution in a stable manner.

Inter-type attention heterogeneous graph neural networks

To extract biological information, it is essential to extract the topological and data features of nodes within subgraphs. Heterogeneous graph convolutional networks (HGCNs) extend GCNs to operate on heterogeneous graphs that contain multiple types of nodes and edges. These networks leverage the graph's complex structure to aggregate feature information from different types of nodes and their interactions, thereby learning rich and comprehensive node representations. The fundamental operation of an HGCN layer can be expressed as follows:

$$\tilde{A}_r = A_r + I, \tag{1}$$

$$\tilde{D}_r = \mathrm{diag}\left(\sum_j \tilde{A}_{r,ij}\right), \tag{2}$$

$$H^{(l+1)} = \sigma\left(\sum_{r\in\mathcal{R}} \tilde{D}_r^{-\frac{1}{2}} \tilde{A}_r \tilde{D}_r^{-\frac{1}{2}} H^{(l)} W_r^{(l)}\right), \tag{3}$$

where $\tilde{A}_r$ is the adjacency matrix for edge type r with added self-loops (where A_r is the original adjacency matrix for edge type r and I is the identity matrix), $\tilde{D}_r$ is the degree matrix corresponding to $\tilde{A}_r$, and $W_r^{(l)}$ is the trainable weight matrix for edge type r at the lth layer and $\mathcal{R}$ is a set of edge types in the heterogeneous graph.

In current multi-omics integration methods, such as SIMBA, traditional heterogeneous graph learning techniques are employed to integrate information from genes and peaks connected through prior knowledge including motif databases. To uncover potential cross-modality relationships that may be missing from the database, scMI needs to identify and focus on inter-type nodes that are not connected by edges but exhibit co-occurrence or similarity in data expression. To capture these long-range inter-type relationships within subgraphs, we propose an inter-type attention mechanism. Subgraphs based on frequency sampling reflect the relationships between nodes, where any two nodes within a subgraph exhibit high correlation. Based on this, we extend the original subgraph into a fully connected graph by connecting all pairs of nodes in the subgraph, and then compute the attention weights between nodes. For each node v_i, attention coefficients are computed between node v_i and its neighboring nodes v_j, considering of the type of edges and nodes:

$$e_{ij} = \sigma\left(h_i W_{r_{ij}} h_j\right), \tag{4}$$

where h_i and h_j are the feature vectors of nodes v_i and v_j, respectively. Notably, $W_{r_{ij}}$ is a learnable inter-type attention matrix according to the edge type between nodes v_i and v_j. The inter-type attention focuses on the relationship between different types of nodes, especially the attention parameter matrix between peaks and genes, which is the key to multimodal integration. Furthermore, these coefficients are normalized across all nodes using the softmax function:

$$\alpha_{r,ij} = \frac{\exp(e_{ij})}{\sum_{e_{ik}\in \mathbf{r}_{ij}} \exp(e_{ik})}, \tag{5}$$

where e_{ik} denotes to all edges that have the same type as e_{ij} incident to node v_i. Node embeddings are obtained by aggregating the features of neighboring nodes, weighted by the normalized attention coefficients:

$$H^{(l+1)} = \sigma\left(\sum_{r\in\mathcal{R}} \mathcal{A}_r H^{(l)} W_r^{(l)}\right), \tag{6}$$

where $\mathcal{A}_r$ is the matrix of $\alpha_{r,ij}$, and $W_r^{(l)}$ is the trainable weight matrix for edge type r at the lth layer.

In order to jointly consider the subgraph's topological structure and long-range inter-type interaction information, we combine

equations (3) and (6) to obtain the final aggregation:

$$H^{(l+1)} = \sigma\left[\sum_{r\in\mathcal{R}}\left(\tilde{D}_r^{-\frac{1}{2}}\tilde{A}_r\tilde{D}_r^{-\frac{1}{2}} + \mathcal{A}_r\right)H^{(l)}W_r^{(l)}\right], \quad (7)$$

This approach effectively captures the complex interplay between different node types and their relationships within the heterogeneous graph. By capturing rich contextual information, The aggregated embeddings form a fundamental basis for downstream analysis.

Since most scRNA-seq and scATAC-seq data lack labels for supervised learning, scMI employs a contrastive learning approach to refine node embeddings by ensuring that similar nodes within a subgraph are closer in the embedding space, while dissimilar nodes, from outside the subgraph, are farther apart. For each node, positive samples are drawn from other nodes within the same subgraph, leveraging the inherent similarities in their properties or relationships. Negative samples are randomly chosen nodes from other subgraphs.

The training employs the Noise Contrastive Estimation (NCE) loss to differentiate between these positive and negative samples. The NCE loss for a node v and the graph G are given by

$$\mathcal{L}_v = -\sum_{v^+\in V_i}\log\frac{\exp(\cos(h_v, h_{v^+})/\tau)}{\exp(\cos(h_v, h_{v^+})/\tau) + \sum_{v^-}\exp(\cos(h_v, h_{v^-})/\tau)}, \quad (8)$$

$$\mathcal{L}_G = \sum_{v\in V}\mathcal{L}_v, \quad (9)$$

where h_{v^+} is the embedding of the positive sample, h_{v^-} are embeddings of negative samples, $\cos(.)$ denotes the cosine similarity, and τ is a temperature parameter.

Contrastive learning enables embeddings to capture the multi-omics information inherent in the heterogeneous graph. To enhance the embeddings' ability to represent data patterns and improve generalization, scMI incorporates modality reconstruction as a supervisory signal during training. Node features and associated edges for 50% of the cells from the scATAC-seq data in training sets are masked in a manner that varies across epochs, ensuring that different nodes are masked each time and promoting robustness in learning. This strategy enhances the model's ability to generalize by training it to infer missing information and adapt to variability. The reconstruction of masked modality feature values is assessed using the Root Mean Square Error (RMSE). The loss is formulated as follows:

$$\mathcal{L}_{\text{omics}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{y}_i - \hat{\mathbf{y}}_i\|_2^2}, \quad (10)$$

where $\mathbf{y}_i$ represents the modality feature values of the masked cells, $\hat{\mathbf{y}}_i$ is the predicted values from the embedding, and n is the total number of masked cells. By minimizing the RMSE loss, scMI refines the embeddings to accurately capture interconnections among features across modalities.

Downstream tasks and evaluation

Modality prediction

Modality prediction is a key task for assessing whether multi-omics data has been successfully integrated. The objective is to predict one modality of multi-omics data from another, such as inferring scATAC-seq data from scRNA-seq data, or vice versa. This process is similar to modality reconstruction but is carried out on the test set to ensure that the data being predicted is completely new to the model. This task is essential for understanding the relationships between different omics and for imputing missing data in multi-omics datasets.

The effectiveness of models performing modality prediction is evaluated using multiple metrics, including RMSE, Pearson correlation coefficient (PCC), and for RNA-seq to ATAC-seq predictions, the Area Under the Receiver Operating Characteristic Curve (AUROC) is additionally calculated. A lower RMSE indicates a smaller discrepancy between the predicted and actual data, reflecting the model's accuracy in inferring missing modalities. Higher PCC values demonstrate stronger linear correlations between predicted and true values, while higher AUROC scores assess the model's classification performance in the restoration of chromatin accessibility.

Cell clustering

Clustering is crucial in single-cell analysis as it allows for the identification of distinct cell populations and states. We employ the k-means algorithm with the embeddings derived from our heterogeneous graph model. The algorithm aims to partition cells into k clusters, with each cell assigned to the nearest mean, which acts as a cluster prototype. The optimization employs the within-cluster sum of squares as the loss function:

$$\mathcal{L}_{\text{task}} = \sum_{i=1}^{N}\sum_{j=1}^{k} z_{ij}\cdot||h_i - \mu_j||^2, \quad (11)$$

where N is the number of cells, k is the number of clusters, h_i is the embedding of the ith cell, μ_j is the centroid of the jth cluster, and z_{ij} is a binary indicator variable that equals 1 if the ith cell is assigned to the jth cluster and 0 otherwise.

To comprehensively evaluate the clustering quality, we use metrics such as the Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), and Average Silhouette Width (ASW). ARI measures the similarity between the clustering labels assigned by the algorithm and true labels, while NMI quantifies the mutual dependence between the predicted and true cluster assignments. ASW assesses how well each cell is clustered by calculating the average silhouette value, where higher values indicate that cells are appropriately clustered with their closest counterparts.

GRN inference

GRN inference is a essential downstream task that aims to unravel the complex network of interactions that control gene expression within a cell. This process involves identifying the regulatory relationships between TFs and their target genes. Transcription factor genes encode proteins that bind to specific DNA sequences and modulate the transcription of other genes. By accurately determining which genes are regulated by specific TFs, researchers can reconstruct the biological networks that dictate cellular functions and behaviors.

For this task, we select representative TFs and their corresponding target genes for each cell type. GRN inference can be framed as a link prediction task. We apply a pairwise scoring function that calculates the likelihood of a regulatory link between a TF gene and a target gene. This scoring function takes the embeddings of the TF gene and the potential target gene as input, evaluating their compatibility for forming a regulatory link. The

loss function for link prediction is formulated as a binary cross-entropy loss:

$$\mathcal{L}_{\text{task}} = -\sum_{i}\sum_{j} y_{g_{\text{TF}_i},g_j} \log(\hat{y}_{g_{\text{TF}_i},g_j}) + (1 - y_{g_{\text{TF}_i},g_j}) \log(1 - \hat{y}_{g_{\text{TF}_i},g_j}), \tag{12}$$

where $y_{g_{\text{TF}_i},g_j}$ represents the ground truth label indicating the presence or absence of a regulatory link between the TFs gene g_{TF_i} and the target gene g_j, while $\hat{y}_{g_{\text{TF}_i},g_j}$ is the predicted probability of such a link as determined by the model.

In the context of GRN inference, we assess the performance of model using AUROC for prediction of regulons associated with specific TFs. By focusing on regulons, which are groups of genes regulated by the same TF, we can precisely measure the model's ability to capture the underlying regulatory relationships governed by these TFs. Higher AUROC values demonstrate that the model accurately identifies the regulatory targets of the TFs, reflecting its effectiveness in reconstructing the gene regulatory network. The ground-truth of GRNs is obtained from three public functional databases, including Reactome [18], DoRothEA [19], and TRRUST v2 [20]. These three databases provide extensive information on gene regulation and biological pathways and have been used for validating GRN enrichment analysis by several single-cell multi-omics GRN inference methods, such as DeepMAPS. Additionally, these databases provide specific annotations for TFs and their target genes for both human and mouse data. To ensure species consistency, we carefully separated the regulatory data into human and mouse categories when selecting the ground-truth, aligning the datasets used for GRN inference with species-specific regulatory information.

Among three databases, we gathered 1,186 TFs and 43,178 associated regulons from human data, and 1,072 TFs with 38,665 associated regulons from mouse data. For each dataset, we selected TFs and their associated regulons based on biological matching and filtered them using highly variable genes to serve as the ground-truth for that dataset. This selection process further refines the choice of ground-truth, ensuring that the regulons are representative within the dataset.

Results and discussion

Overall performance evaluation on matched datasets

The results of modality prediction on datasets of four types are shown in Fig. 2(a–c). LS. Lab and Cajal are the winning methods for RNA-seq to ATAC-seq and ATAC-seq to RNA-seq predictions in the NeurIPS 2021 competition, respectively [21]. scJoint [22] and scMoGNN [23] are deep learning-based prediction methods. In the ATAC-seq to RNA-seq prediction, scMI achieves the lowest average RMSE and the highest PCC among all compared methods. Specifically, scMI demonstrates the best RMSE performance in seven out of eight datasets (Supplementary Table 4) and achieves the highest PCC in five datasets (Supplementary Table 5). For the RNA-seq to ATAC-seq task, scMI still achieves the best RMSE performance and ranks second in terms of PCC. Additionally, it obtains the highest average AUROC in five out of eight datasets (Supplementary Table 8), with the overall average being the highest among the compared methods.

For cell clustering, we compare scMI with Seurat v1 [24], MOFA+ [25], DeepMAPS , MultiVI [26], SIMBA, and scEMC [27]. As shown in Fig. 2(d–e), scMI outperforms other methods in ARI, NMI, and ASW. In particular, scMI achieves the highest ARI and NMI in five out of the seven datasets (Supplementary Tables 9–10), and it achieves the highest ASW score across six datasets(Supplementary Table 11). Compared to the single-omic method Seurat v1, scMI significantly improves across all seven datasets, with an average increase of 73.5% ARI. By integrating data from multiple omics, the model captures the complex biological relationships between cells more effectively, leading to more accurate and distinct clustering. Compared to baseline methods, scMI better captures and preserves biological differences between cell types, resulting in clearer and more biologically meaningful clusters.

The UMAP visualization of the embeddings learned by SIMBA and scMI is shown in Fig. 3, colored by clusters. Compared to SIMBA, scMI exhibits more distinct clustering boundaries, with less overlap among the embeddings of different clusters. This result is attributed to scMI's ability to simultaneously capture long-range inter-type relationships and local graph structure, thereby enhancing the distinction between different cell types and avoiding over-smoothing. On the other hand, SIMBA focuses solely on representation learning for biological networks without differentiating downstream tasks. In contrast, our dynamic collaborative learning approach optimizes the loss function to simultaneously refine cell embeddings and minimize the loss for downstream clustering tasks, enabling fine-tuned embeddings tailored for cell clustering and resulting in superior performance.

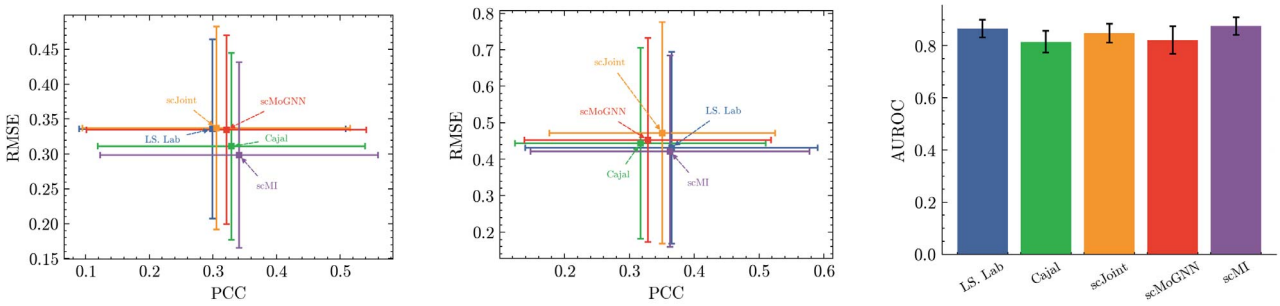
To validate the GRN inference capability of scMI, we select GRNBoost2 [28], scMTNI [29], DeepMAPS, and SIMBA as baseline methods. Considering real-world scenarios in single-cell multi-omics, we calculate the AUROC for inference of regulons (Fig. 2f). GRNBoost2 infers GRNs solely using scRNA-seq data, while scMTNI, DeepMAPS, and SIMBA integrate multi-omics information to comprehensively infer regulatory relationships. scMI exhibits superior performance across all datasets (Supplementary Table 12), excelling in identifying regulons associated with specific TFs.

The enhancement of GRNs inference through multi-omics integration lies in supplementing missing gene information from single-omics data with scATAC-seq data, uncovering more regulatory relationships. Our inter-type attention mechanism specifically targets gene-peak relationships, enabling the discovery of relationships that prior-knowledge-based models may have overlooked.

It is observed that scMI does not surpass existing methods in modality prediction and cell clustering on certain datasets. Unlike baseline methods, scMI does not rely on cross-modality relationships derived from motif databases; instead, it aims to discover these relationships directly from the data. When dealing with smaller datasets, the inherent information may be insufficient to fully capture the cross-modality relationships specific to certain cell types, resulting in performance that falls short of existing methods. Nevertheless, scMI maintains robustness across various tasks and achieves the best average performance across multiple metrics in different task scenarios.

Performance of omics alignment on unmatched datasets

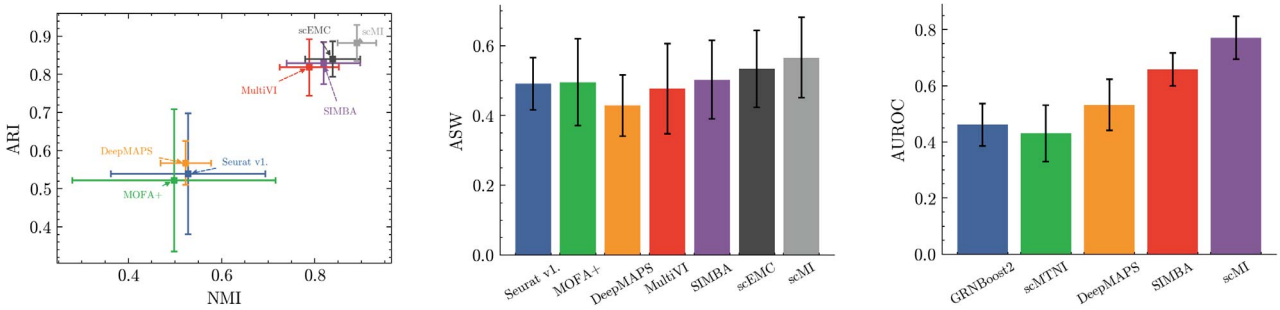
The integration of unmatched multi-omics data, such as scRNA-seq and scATAC-seq, poses considerable challenges due to the inherent difficulty of establishing direct correspondence between cells across different omics. This lack of direct alignment complicates efforts to accurately combine the datasets, as traditional integration methods often rely on matched or paired data. For unmatched multi-omics datasets, scMI leverages existing



(a) Modality Prediction: scATAC-seq to scRNA-seq

(b) Modality Prediction: scRNA-seq to scATAC-seq

(c) Modality Prediction: scRNA-seq to scATAC-seq



(d) Cell Clustering

(e) Cell Clustering

(f) GRN Inference

Figure 2. Overall performance evaluation on matched datasets. (a) Average RMSE and PCC values of scMI and baselines predicting scRNA-seq data from scATAC-seq data across eight matched datasets. The x and y axes are average PCC and RMSE values, respectively. Error bars indicate the SD of eight datasets. (b) Same as (a), but the algorithms predict scATAC-seq data from scRNA-seq data. (c) Bar plots illustrate the AUROC values of scMI and baselines predicting scATAC-seq data from scRNA-seq data across eight matched datasets. (d) Same as (a), but the results are evaluated by the average ARI and NMI values of cell clustering across seven datasets. (e) Bar plots illustrate the ASW values of cell clustering across seven datasets. (f) Same as (e), but the results are evaluated by the average AUROC values of GRN Inference. All the source data can be found in Supplementary Tables 4–12.

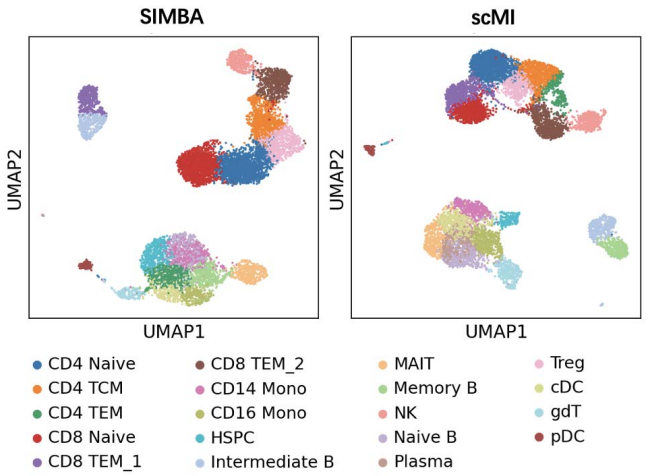


Figure 3. The UMAP visualization of clustering results for SIMBA (ARI=0.836) and scMI(ARI=0.867) on the PBMC 10k dataset, with each color representing a cluster. Even without relying on motif databases, scMI's clustering results still exhibit clearly boundaries.

alignment methods for initial pairing and further integrates the data with the inter-type attention mechanism. Our results demonstrate that scMI not only enhances the alignment between these datasets but also significantly improves the overall integration process, leading to more accurate and biologically meaningful insights from the combined multi-omics data.

We assess the alignment quality by visualizing the embeddings of cells from PMBC scRNA-seq and scATAC-seq data using UMAP. The embeddings of both node types are combined, and the degree of mixing is analyzed. As shown in Fig. 4, we visualize the embeddings obtained through Seurat v4, GLUE, and bindSC methods, as well as the concatenated embeddings (Concat) from these three methods. The plots reveal that the enhanced embeddings achieved by scMI based on the integration of these three methods exhibit a higher degree of mixing between cells from the two omics sources, indicating a more seamless integration facilitated by scMI.

To quantitatively analyze the alignment and integration performance of the model on unmatched datasets, we define a metric match_k to evaluate the effectiveness of cell pairing between scRNA-seq and scATAC-seq data in the latent space. For each scRNA-seq cell (c_i^{RNA}), we identify the k nearest scATAC-seq cells (c_j^{ATAC}) based on their embeddings in the latent space. A pairing is considered successful if the true corresponding c_i^{ATAC} is within the k nearest neighbors of c_i^{RNA}. The match_k is defined as follows:

$$\text{match}_k = \frac{\left|\left\{c_i^{\text{RNA}} \mid c_i^{\text{ATAC}} \in \text{KNN}_k(c_i^{\text{RNA}})\right\}\right|}{\left|\{c_i^{\text{RNA}}\}\right|} \tag{13}$$

where $\text{KNN}_k(c_i^{\text{RNA}})$ represents the set of k nearest neighbors of c_i^{RNA} in the latent space. This metric provides a quantitative measure of the quality of the alignment of the scRNA-seq and scATAC-seq data, with higher match_k values indicating better alignment and integration between the two modalities. As shown in Fig. 4, scMI achieves a match_k value of 0.2218, significantly outperforming the baseline methods. This substantial improvement

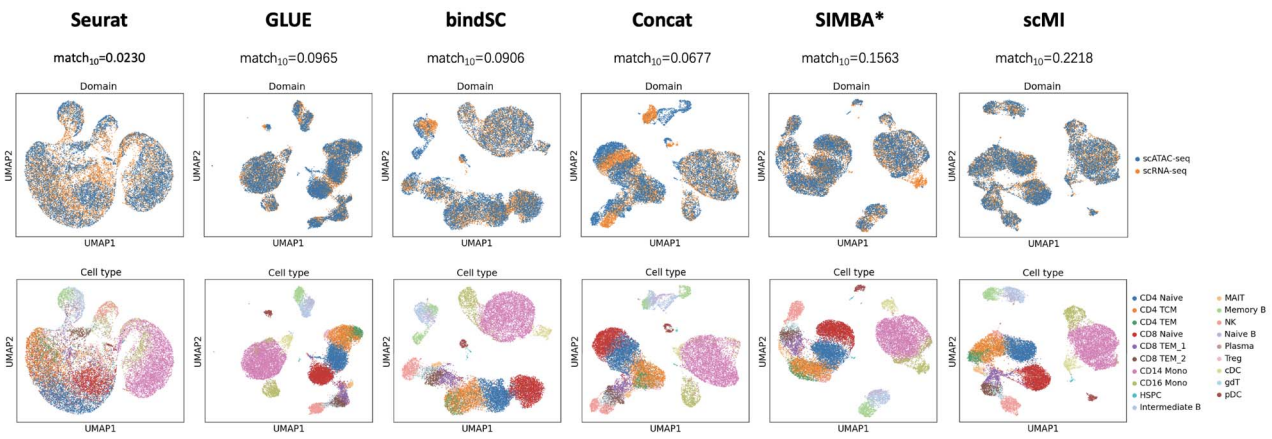


Figure 4. The UMAP of cell embeddings in unmatched PBMC 10k datasets. The six columns represent embeddings generated by Seurat v4, GLUE, bindSC, their concatenation (Concat), SIMBA, and the proposed scMI. (SIMBA* refers to the multi-omics integration analysis module in SIMBA.) Cells are colored by data source (scRNA-seq or scATAC-seq) in the first row, and by type in the second row. The metric $match_{10}$, used to assess alignment performance, indicates the ratio of cases where the corresponding scATAC-seq cell is among the 10 nearest neighbors of scRNA-seq cell in the latent space. The corresponding t-SNE visualization is shown in Supplementary Fig. 1.

indicates that scMI is more effective in accurately aligning scRNA-seq and scATAC-seq cells in the latent space, leading to more successful pairings between the two modalities.

To validate the biological functionality of embeddings obtained through unmatched multi-omics methods and scMI, we further visualize the embeddings from the perspective of cell types. As shown in Fig. 4, the embeddings enhanced by scMI produce more distinct boundaries between different cell types. This improvement highlights the ability of scMI to maintain and enhance the biological distinctions among various cell populations.

To further evaluate the alignment and integration performance of our model in cell subtype heterogeneity, we perform UMAP visualization on three types of B cells, as shown in Fig. 5. Naive B cells are activated to become transitional B cells, with a subset eventually differentiating into memory B cells. The alignment effectiveness of the four baseline methods has certain limitations, resulting in embeddings that do not adequately distinguish between the three types of B cells after integration. Due to these limitations, the boundaries between the different types of B cells become blurred, making it difficult to achieve accurate clustering and downstream analyzes. In contrast, the embeddings enhanced by scMI exhibit significantly improved alignment, which allows for clearer separation of the three B-cell types. The scMI-enhanced embeddings reveal distinct boundaries between the different cell populations, demonstrating the method's ability to preserve and highlight biological differences in the data.

Additionally, to evaluate the omics alignment and degree of mixing of the scRNA-seq and scATAC-seq data on unmatched datasets, we introduce the integrated Local Inverse Simpson's Index (iLISI) [30] as an evaluation metric. iLISI assesses the batch mixing quality by measuring the diversity of batch labels within the neighborhood of each cell in the integrated space. For this analysis, we treat the scRNA-seq and scATAC-seq data as two distinct batches and compute the iLISI score for each cell type. The final iLISI score is obtained by averaging across all cell types, providing a comprehensive metric that indicates how well the integration method aligns and mixes cells from different omics sources. Higher iLISI scores suggest better integration, reflecting effective omics alignment. We compare scMI with existing methods for unmatched datasets, including the baeline methods Seurat v4, GLUE, bindSC as well as Harmony [30], scJoint, SIMBA and scBridge [31]. As shown in Table 1, scMI consistently achieves competitive iLISI scores, demonstrating superior integration capability across most datasets. It also shows the highest average iLISI score of 0.810, indicating that it maintains strong alignment performance across various cell types and datasets. scMI's overall results highlight its robust and effective integration of scRNA-seq and scATAC-seq data, particularly in cases where datasets are unmatched.

Evaluation of downstream task performance on unmatched datasets

Enhanced alignment improves the accuracy of data integration and embedding learning, leading to more precise representations of the underlying biological relationships. This, in turn, enables the model to perform better in downstream tasks, such as cell clustering and gene regulatory network inference, by providing a more coherent and biologically relevant foundation for these analyses.

To validate the effectiveness of scMI on unmatched datasets, we compare scMI with three methods designed for unmatched datasets. The results are shown in Table 2, with the last row representing the results for the fully matched dataset as a reference. The results demonstrate that scMI not only surpasses all existing state-of-the-art methods for unmatched datasets but also approaches the performance levels achieved with fully matched datasets. This indicates scMI's superior capability in integrating multi-omics data and improving the accuracy and reliability of downstream analyses, even in the absence of direct cell-to-cell correspondence between different omics.

Furthermore, the GRN inference task also shows excellent performance on unmatched datasets in Table 3, reinforcing scMI's effectiveness in deciphering complex regulatory interactions and providing robust insights into cellular mechanisms.

While scMI demonstrates superior performance on both matched and unmatched datasets, computational efficiency remains a critical consideration, especially for large-scale datasets. To further evaluate the efficiency of scMI, we compare its computation time for cell clustering with the baseline methods mentioned above, providing insights into the computational scalability of scMI. We record the computational time of scMI and the baselines across three datasets: PBMC 3k (2,711 cells), MST (5,692 cells), and PBMC 10k (11,898 cells). All experiments were conducted under the same hardware configuration. As

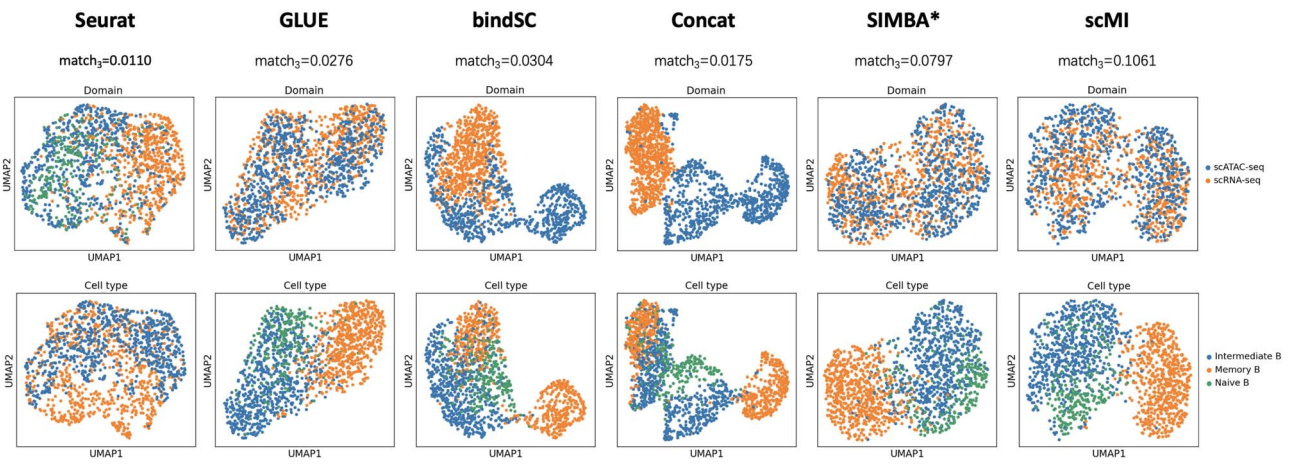


Figure 5. The UMAP of cell embeddings among three B cell subtypes in unmatched PBMC 10k datasets. The interpretation of each image is consistent with that of Fig. 4. Similarly, $match_3$ indicates the ratio of cases where the corresponding scATAC-seq cell is among the three nearest neighbors of scRNA-seq cell in the latent space. The corresponding t-SNE visualization is shown in Supplementary Fig. 2.

Table 1. Performance comparison on unmatched datasets of iLISI for omics alignment. SIMBA* refers to the multi-omics integration analysis module in SIMBA, which takes unmatched multi-omics data as input. The best-performing results for unmatched datasets are highlighted in bold

	iLISI↑							
	PBMC 3k	**PBMC 10k**	**BMMC 1**	**BMMC 2**	**BMMC 3**	**MEBC**	**MST**	**Average**
Seurat v4	0.593	0.517	0.676	0.579	0.753	0.454	0.414	0.570
GLUE	0.827	0.687	0.671	0.637	0.740	0.697	0.715	0.710
bindSC	0.721	0.715	0.739	0.583	**0.764**	0.686	0.624	0.690
Harmony	0.853	0.756	0.759	0.510	0.694	0.649	0.746	0.710
scJoint	0.828	0.749	0.612	0.557	0.666	0.584	0.699	0.671
SIMBA*	0.899	0.775	0.661	0.705	0.709	**0.750**	0.743	0.749
scBridge	0.859	**0.835**	0.714	0.721	0.734	0.653	0.821	0.762
scMI(unmatched)	**0.890**	0.815	**0.856**	**0.757**	0.757	0.738	**0.857**	**0.810**

Table 2. Performance comparison on unmatched datasets of cell clustering. SIMBA* refers to the multi-omics integration analysis module in SIMBA, which takes unmatched multi-omics data as input. The best-performing results for unmatched datasets are highlighted in bold

	ARI↑							
	PBMC 3k	**PBMC 10k**	**BMMC 1**	**BMMC2**	**BMMC 3**	**MEBC**	**MST**	**Average**
Seurat v4	0.883	0.629	0.652	0.360	0.526	0.382	0.364	0.542
GLUE	0.913	0.731	0.657	0.450	0.553	0.564	0.578	0.635
bindSC	0.905	0.752	0.717	0.586	0.538	0.541	0.603	0.663
Harmony	0.892	0.693	0.725	0.491	0.533	0.472	0.594	0.629
scJoint	0.868	0.750	0.656	0.537	0.491	0.549	0.583	0.633
SIMBA*	0.909	0.822	0.675	0.694	0.627	0.598	0.733	0.723
scBridge	0.927	0.819	0.798	0.773	0.528	0.570	0.769	0.741
scMI(unmatched)	**0.953**	**0.854**	**0.860**	**0.823**	**0.813**	**0.742**	**0.849**	**0.842**
scMI(matched)	0.968	0.867	0.924	0.855	0.840	0.841	0.881	0.882

shown in Fig. 6, traditional R-based methods (e.g. Seurat and MOFA+) consistently exhibit lower computational complexity due to their reliance on pre-computed databases or comparatively simple model architectures. However, these methods trade computational efficiency for limited performance and flexibility when dealing with complex or unmatched datasets.

Among deep learning methods, scMI demonstrates comparable computation time to other graph-based methods, including SIMBA and DeepMAPS. On the largest dataset, scMI achieves a shorter running time than SIMBA and DeepMAPS. Notably, as the number of cells increases, scMI becomes more efficient compared to SIMBA and DeepMAPS. This improvement can be attributed to the FRWR algorithm and the inter-type attention mechanism, which optimize cross-modality relationship learning and enable scMI to handle larger datasets efficiently.

Ablation study

Frequency-based subgraph sampling

Since the learning of embeddings in scMI relies on subgraph sampling, we evaluate the efficacy of the FRWR algorithm on the

Table 3. AUROC on unmatched datasets of GRN inference comparing regulons from specific TFs. SIMBA* refers to the multi-omics integration analysis module in SIMBA, which takes unmatched multi-omics data as input. The best-performing results for unmatched datasets are highlighted in bold

	AUROC↑							
	PBMC 3k	**PBMC 10k**	**BMMC 1**	**BMMC 2**	**BMMC 3**	**MEBC**	**MST**	**Average**
GLUE	0.290	0.379	0.356	0.551	0.594	0.405	0.392	0.424
bindSC	0.419	0.376	0.390	0.423	0.503	0.342	0.449	0.415
SIMBA*	0.527	0.620	0.494	0.408	0.653	0.362	0.470	0.505
scMI(unmatched)	**0.769**	**0.737**	**0.654**	**0.628**	**0.838**	**0.667**	**0.724**	**0.717**
scMI(matched)	0.796	0.798	0.719	0.695	0.887	0.678	0.823	0.771

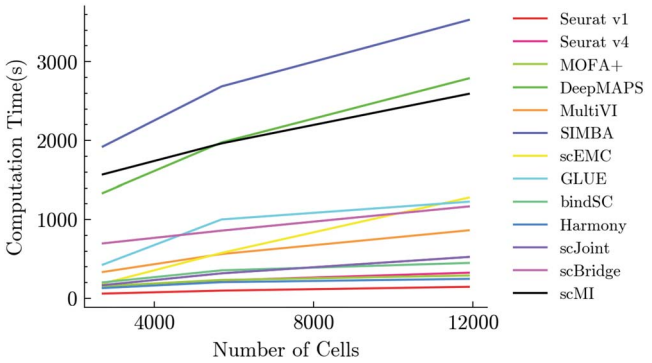


Figure 6. Computational time comparison between scMI and baseline methods for cell clustering across three datasets (PBMC 3k, MST, and PBMC 10k). The source data and hardware configuration can be found in Supplementary Table 13.

Table 4. Ablation study on subgraph sampling algorithms. The best-performing results are highlighted in bold

			ARI↑	
	m	***k***	**PBMC 3k**	**PBMC 10k**
RW	–	–	0.722	0.683
RWR	–	–	0.721	0.754
FRWR	10	5	0.931	0.815
	30	5	**0.968**	0.840
	20	5	0.963	**0.867**
	20	2	0.942	0.855
	20	8	0.950	0.852

PBMC datasets, as shown in Table 4. We compare our method with standard RW and Random Walk with Restart (RWR) approaches. Our results demonstrate that FRWR achieves better performance and stability, ensuring that the sampled subgraphs retain biologically meaningful relationships.

The performance of FRWR depends on several key hyperparameters, including the number of subgraphs m and parameter k regulating the size of subgraphs. To ensure optimal sampling and representation learning, we systematically optimize these hyperparameters through extensive experiments. We conduct grid search to optimize m and k within predefined ranges: $m \in \{10, 20, 30\}$ and $k \in \{5, 8, 10\}$. The results are shown in Table 4. The number of subgraph samples m significantly influences the model's performance by affecting the diversity and stability of captured subgraph structures. As m increases, the generated subgraphs better represent the graph's overall structure, leading to enhanced performance in GRN inference. Specifically, we observe that setting $m = 20$ strikes a balance between stability and computational efficiency. While further increasing m (e.g. $m = 30$) continues to improve subgraph stability, it results in diminishing diversity and increased computational costs. When m exceeds 30, the diversity of subgraphs sampled from the same starting node decreases significantly during training, which can lead to overfitting.

The parameter k, representing the number of nodes of each type in a subgraph, plays a crucial role in balancing structural detail and information aggregation. If k is too small, the subgraph lacks sufficient context to capture complex graph structures, limiting the learning process. Conversely, as k continues to increase, additional noise can be introduced into the subgraph. For example, when learning the embedding of an scRNA-seq cell, the inclusion of multiple unrelated scATAC-seq cell nodes can lead to blurred clustering boundaries, adversely affecting the model's performance in clustering tasks. This noise complicates the representation and can impair the model's ability to accurately distinguish cell types.

In our study, $m = 20, k = 5$ provides the best trade-off, enabling scMI to maintain effective representation learning without over-smoothing or excessive noise.

To enhance the performance of scMI, we conduct grid search for the other hyperparameters of scMI across different downstream tasks. The tuning experiments for model hyperparameters are presented in Supplementary Figs 3–5, and the hyperparameters for all downstream tasks are summarized in Supplementary Table 2.

Inter-type attention mechanism

Inter-type attention mechanism automatically captures cross-modality relationships, enhancing scMI's capability to integrate multi-omics data. In this ablation study, we evaluate the role of the inter-type attention mechanism within scMI by conducting experiments with three different setups: using only HGCN, using only the Inter-type Attention mechanism (I.A.), and using both HGCN and inter-type attention together. These experiments are carried out on the PBMC 3k and PBMC 7k datasets to assess their performance on modality prediction tasks. As shown in Table 5, the combination of HGCN and the inter-type attention mechanism achieved the best performance across all three experimental setups.

HGCN primarily focuses on aggregating local neighborhood information and preserving topological structure. This means that while it can capture structural and connectivity patterns in the graph, it lacks the capability to highlight interactions between distinct node types, such as those between genes and peaks. As a result, the model's performance is constrained by its inability to fully represent the complex, long-range dependencies that are crucial for cross-modality learning.

Table 5. Ablation study on inter-type attention mechanism

	RMSE↓			
	ATAC-seq to RNA-seq		RNA-seq to ATAC-seq	
	PBMC 3k	PBMC 7k	PBMC 3k	PBMC 7k
HGCN	0.1970	0.5317	0.3276	1.0310
I.A.	0.1776	0.2824	0.2689	0.8375
HGCN+I.A.	0.1214	0.2065	0.2238	0.7796

Table 6. Ablation study on multi-omics integration. The best-performing results are highlighted in bold

	AUROC↑		
	BMMC 1	BMMC 2	BMMC 3
GRNBoost2	0.440	0.592	0.524
scMI(embedding)	0.691	0.601	0.848
scMI	**0.719**	**0.695**	**0.887**

The inter-type attention mechanism addresses some of these issues by allowing the model to focus on significant cross-modality relationships between different node types, helping to capture these long-range dependencies. By learning cross-modality relationships and combining with information from one modality, the model can more accurately predict another modality. Notably, the improvement brought by the inter-type attention mechanism is more pronounced on larger-scale datasets, indicating that larger datasets enable the model to learn more comprehensive cross-modality relationships.

Multi-omics integration in GRN inference

To quantify the enhancement of GRN inference by multi-omics information and our model, we employ the single-omic GRN inference method GRNBoost2 for comparison. Three setups include inferring GRN using scRNA-seq data with GRNBoost2, using embeddings generated by scMI as input for GRNBoost2, and inferring GRN with scMI. F-scores for the top K edges with the highest confidence from each setup are presented in Table 6. Comparing the first two experimental setups, using integrated embeddings as input for GRNBoost2 provides more information than scRNA-seq alone, resulting in higher accuracy in GRN generation. The outcomes of the latter two experiments show that our framework and collaborative learning approach, based on omics integration, better adapt to GRN inference.

Conclusion

In this paper, we introduce scMI, a novel deep learning method based on heterogeneous graphs for single-cell multi-omics integration. This approach effectively integrates scRNA-seq and scATAC-seq data, addressing the limitations of traditional methods that rely heavily on motif databases. By employing a frequency-based multiple Random Walk with Restart strategy for subgraph sampling, scMI effectively balances the representation of different node types while maintaining biological interpretability. A key innovation of scMI is the introduction of an inter-type attention mechanism, which captures long-range cross-modality relationships between different node types, such as genes and chromatin accessibility peaks. This mechanism enhances the model's ability to uncover complex biological interactions that may be missed by other approaches.

To evaluate the effectiveness of our model, we conduct extensive experiments to compare scMI with baselines. The results on various datasets consistently demonstrate the superiority and robustness of scMI in tasks including cell clustering and GRN inference. Notably, scMI enhances the alignment and integration of unmatched multi-omics data, leading to more accurate and biologically meaningful insights. The method's ability to capture long-range cross-modality relationships without the need for motif databases further underscores its robustness and adaptability.

Key Points

- We propose scMI, a novel heterogeneous graph embedding approach to integrate scRNA-seq and scATAC-seq data. Our framework models cells and modality features from multi-omics datasets onto heterogeneous graphs for representation learning, eliminating the reliance on motif databases.
- We introduce an inter-type attention mechanism to capture long-range cross-modality relationships on heterogeneous graphs, enhancing the biological interpretability of gene-peak interactions.
- For unmatched datasets, scMI builds upon existing alignment methods and employs the inter-type attention mechanism to adjust misaligned pairings, enhancing alignment accuracy.
- Extensive evaluations are conducted to validate the effectiveness of scMI. Visualization elucidates the biological interpretability and insights gained from our study.

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Conflict of interest: None declared.

References

1. Ma A, McDermaid A, Jennifer X. *et al*. Integrative methods and practical challenges for single-cell multi-omics. *Trends Biotechnol* 2020;**38**:1007–22. https://doi.org/10.1016/j.tibtech.2020.02.013.
2. Kashima Y, Sakamoto Y, Kaneko K. *et al*. Single-cell sequencing techniques from individual to multiomics analyses. *Exp Mol Med* 2020;**52**:1419–27. https://doi.org/10.1038/s12276-020-00499-2.
3. Lee J, Hyeon DY, Hwang D. Single-cell multiomics: technologies and data analysis methods. *Exp Mol Med* 2020;**52**:1428–42. https://doi.org/10.1038/s12276-020-0420-2.
4. Sorrenti V, Gabbia D, Fortinguerra S. *et al*. Chapter 4 - single-cell omics: cellular functions. In: Barh D, Azevedo V (eds). *Single-Cell Omics*, Academic Press, Cambridge, Massachusetts, USA, 2019, 45–59. https://doi.org/10.1016/B978-0-12-814919-5.00004-X.
5. Badia-i-Mompel P, Wessels L, Müller-Dott S. *et al*. Gene regulatory network inference in the era of single-cell multi-omics. *Nat Rev Genet* 2023;**24**:739–54. https://doi.org/10.1038/s41576-023-00618-5.
6. Butler A, Hoffman P, Smibert P. *et al*. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018;**36**:411–20. https://doi.org/10.1038/nbt.4096.

7. Liu J, Gao C, Sodicoff J. *et al*. Jointly defining cell types from multiple single-cell datasets using liger. *Nat Protoc* 2020;**15**:3632–62. https://doi.org/10.1038/s41596-020-0391-8.
8. Argelaguet R, Velten B, Arnol D. *et al*. Multi-omics factor analysis–a framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol* 2018;**14**:e8124. https://doi.org/10.15252/msb.20178124.
9. Hao Y, Hao S, Andersen-Nissen E. *et al*. Integrated analysis of multimodal single-cell data. *Cell* 2021;**184**:3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.
10. Wang X, Ji H, Shi C. *et al*. Heterogeneous graph attention network. In *The World Wide Web Conference*, WWW '19, pp. 2022–32, New York, NY, USA: Association for Computing Machinery, 2019. https://doi.org/10.1145/3308558.3313562.
11. Ma A, Wang X, Li J. *et al*. Single-cell biological network inference using a heterogeneous graph transformer. *Nat Commun* 2023;**14**:964. https://doi.org/10.1038/s41467-023-36559-0.
12. Chen H, Ryu J, Vinyard ME. *et al*. SIMBA: single-cell embedding along with features. *Nat Methods* 2024;**21**:1003–13. https://doi.org/10.1038/s41592-023-01899-8.
13. Luecken M, Burkhardt D, Cannoodt R. *et al*. A sandbox for prediction and integration of DNA, RNA, and proteins in single cells. In: J Vanschoren, S Yeung (eds). *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, Curran Associates Inc., NY, USA, 2021, **1**.
14. Chen S, Lake BB, Zhang K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat Biotechnol* 2019;**37**:1452–7. https://doi.org/10.1038/s41587-019-0290-0.
15. Ma S, Zhang B, LaFave LM. *et al*. Chromatin potential identified by shared single-cell profiling of rna and chromatin. *Cell* 2020;**183**:1103–1116.e20. https://doi.org/10.1016/j.cell.2020.09.056.
16. Cao Z-J, Gao G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nat Biotechnol* 2022;**40**:1458–66. https://doi.org/10.1038/s41587-022-01284-4.
17. Dou J, Liang S, Mohanty V. *et al*. Bi-order multimodal integration of single-cell data. *Genome Biol* 2022;**23**:112. https://doi.org/10.1186/s13059-022-02679-x.
18. Joshi-Tope G, Gillespie M, Vastrik I. *et al*. Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res* 2005;**33**:D428–32.
19. Garcia-Alonso L, Holland CH, Ibrahim MM. *et al*. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res* 2019;**29**:1363–75. https://doi.org/10.1101/gr.240663.118.
20. Han H, Cho J-W, Lee S. *et al*. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res* 2018;**46**:D380–6. https://doi.org/10.1093/nar/gkx1013.
21. Lance C, Luecken MD, Burkhardt DB. *et al*. Multimodal single cell data integration challenge: results and lessons learned. In: *NeurIPS 2021 Competitions and Demonstrations Track*, PMLR, Cambridge, Massachusetts, United States, 2022, pp. 162–76.
22. Lin Y, Tung-Yu W, Wan S. *et al*. scJoint integrates atlas-scale single-cell RNA-seq and ATAC-seq data with transfer learning. *Nat Biotechnol* 2022;**40**:703–10. https://doi.org/10.1038/s41587-021-01161-6.
23. Wen H, Ding J, Jin W. *et al*. Graph neural networks for multimodal single-cell data integration. In: *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, Association for Computing Machinery, New York, NY, United States, 2022, pp. 4153–63.
24. Satija R, Farrell JA, Gennert D. *et al*. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 2015;**33**:495–502. https://doi.org/10.1038/nbt.3192.
25. Argelaguet R, Arnol D, Bredikhin D. *et al*. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol* 2020;**21**:1–17.
26. Ashuach T, Gabitto MI, Koodli RV. *et al*. MultiVI: deep generative model for the integration of multimodal data. *Nat Methods* 2023;**20**:1222–31. https://doi.org/10.1038/s41592-023-01909-9.
27. Dayu H, Liang K, Dong Z. *et al*. Effective multi-modal clustering method via skip aggregation network for parallel scRNA-seq and scATAC-seq data. *Brief Bioinform* 2024;**25**:bbae102. https://doi.org/10.1093/bib/bbae102.
28. Moerman T, Santos SA, González-Blas CB. *et al*. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics* 2019;**35**:2159–61. https://doi.org/10.1093/bioinformatics/bty916.
29. Zhang S, Pyne S, Pietrzak S. *et al*. Inference of cell type-specific gene regulatory networks on cell lineages from single cell omic datasets. *Nat Commun* 2023;**14**:3064. https://doi.org/10.1038/s41467-023-38637-9.
30. Korsunsky I, Millard N, Fan J. *et al*. Fast, sensitive and accurate integration of single-cell data with harmony. *Nat Methods* 2019;**16**:1289–96. https://doi.org/10.1038/s41592-019-0619-0.
31. Li Y, Zhang D, Yang M. *et al*. scBridge embraces cell heterogeneity in single-cell RNA-seq and ATAC-seq data integration. *Nat Commun* 2023;**14**:6045. https://doi.org/10.1038/s41467-023-41795-5.

Briefings in Bioinformatics, 2025, **26(1)**, bbae711
https://doi.org/10.1093/bib/bbae711
Problem Solving Protocol