bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



# Global and cross-omics feature aggregation improves single-cell multi-omics integration and clustering

**Yingjie Guo**[1,2,*,†], **Chenyang Cheng**[1,2,*], **Zheqing Zhu**[1,2], **Jin Zhou**[1,2], **Haiping Lu**[3,4], **Yuhua Qian**[1,2,†] and **Zhen Liang**[5,†]

[1]Institute of Big Data Science and Industry, Shanxi University
[2]School of Computer and Information Technology, Shanxi University
[3]Centre for Machine Intelligence, University of Sheffield
[4]School of Computer Science, University of Sheffield
[5]School of Life Science, Shanxi University
[*]These authors contributed equally to this work.
[†]Corresponding author. zliang@sxu.edu.cn, jinchengqyh@126.com, yjguo@sxu.edu.cn

**Sequencing technology advancements enable the simultaneous measurement of multiple omics within single cells, thereby facilitating single-cell multi-omics analysis. Most existing deep multi-omics integration methods learn shared or omics-specific representations from multiple omics by omics-wise aggregation. However, these approaches often fail to account for the structural relationships among all cells. Here, we introduce scCODI, an innovative cross-omics deep learning model to integrate single-cell multi-omics data for cell clustering. This model processes paired omics datasets, combining single-cell transcriptomics with either epigenomics or proteomics data to facilitate a more comprehensive exploration of cellular heterogeneity. Specifically, the shared representation of two omics is derived through cross-omics feature aggregation, fully leveraging the complementarity of different omics within the same cell. Additionally, we align the omic-specific and shared representations of each cell using a global relationship-guided contrastive learning module, which enhances the similarity between the shared and omic-specific representations of the same cell. Extensive experiments demonstrate that scCODI outperforms previous methods in single-cell multi-omics data integration and cell clustering, as evidenced by its performance across nine paired multi-omics datasets. This makes it a promising tool for decoding complex biological mechanisms. The code is available at https://github.com/SXUI-BDSI/scCODI**

*Keywords: single cell, multi-omics, data integration, feature aggregation, clustering*

## 1. Introduction

Recent advancements in single-cell multi-omics sequencing technologies have facilitated the simultaneous and detailed profiling of multiple molecular modalities, such as RNA expression, protein abundance (ADT), and chromatin accessibility (ATAC) within individual cells [1]. SNARE-seq [2] and SMAGE-seq [3] integrate single-cell RNA data with chromatin accessibility information, and CITE-seq [4–6] links RNA data with protein measurements. These technologies reveal cellular heterogeneity across multiple molecular dimensions and allow more detailed characterization of cellular features to elucidate interactions between distinct omics layers[7]. They provide a holistic view of cellular interactions and regulatory networks at single-cell resolution [8]. Given the intrinsic complementarity of different omics [3], increasing the number of measured modalities further enhances our ability to decode complex biological mechanisms such as cellular function, tissue development, and disease pathogenesis [9]. However, with the rapid expansion of single-cell multi-omics datasets, there is an urgent need for advanced computational frameworks capable of harmonizing high-dimensional data from multiple sources. Such integrative tools are critical to enabling downstream analyses such

© 2025 Shanxi University. All rights reserved

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



as clustering and cell-type identification, and unlocking the full potential of increasingly powerful experimental platforms [10].

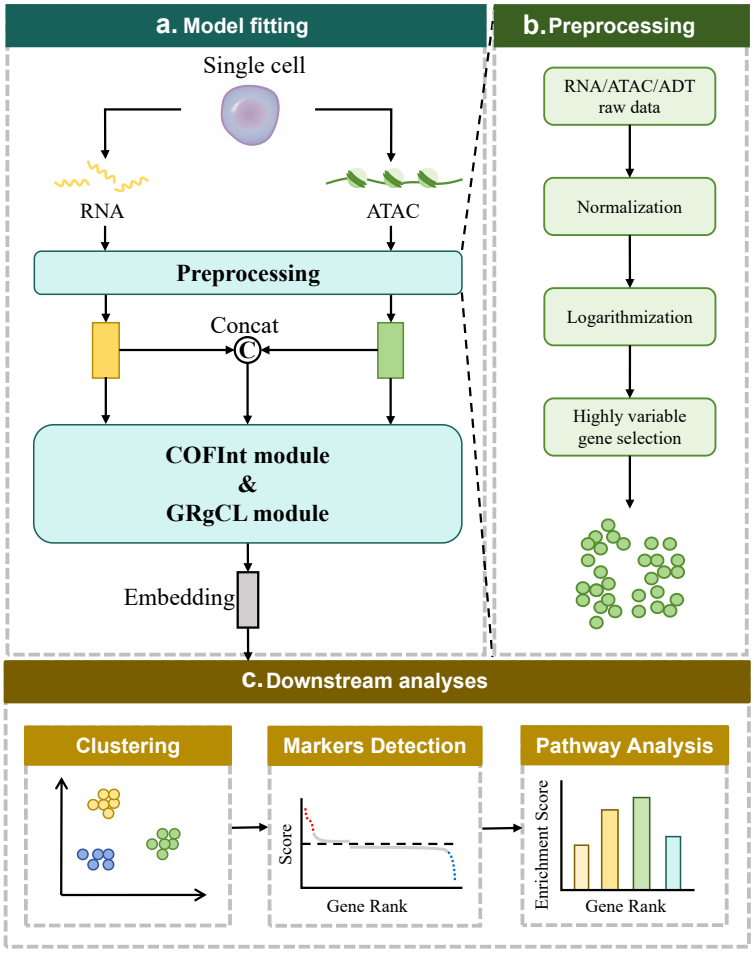



Figure 1. **Overview of the scCODI workflow.** The workflow begins with model fitting (a). Using single-cell multi-omics technology, RNA and ATAC (or ADT), can be obtained from the same cell. These raw data then undergo preprocessing (b), which includes normalization, logarithmic transformation, and the selection of highly variable genes to reduce noise and prepare the data for integration. By concatenating the preprocessed RNA and ATAC, an integrated input is obtained. This integrated input and preprocessed RNA and ATAC, are separately fed into the the Cross-Omics Feature Integration (COFInt) module for further data augmentation. Subsequently, the global relationship-guided Contrastive Learning (GRgCL) module employs contrastive learning to align shared and omic-specific representations. It selects negative samples using the global relationship matrix obtained from COFInt and treats the same cell's representations in both shared and omic-specific spaces as positive samples, thereby incorporating omic-specific information to generate the final embeddings. These embeddings are then used for cluster analysis and subsequent downstream analysis tasks (c) such as marker gene detection and pathway analysis for deeper biological insights.

Cluster analysis plays a pivotal role in single-cell research, and it serves as a foundational method for investigating cellular heterogeneity and biological processes at the level of cell types or subtypes. It provides critical insights that are often obscured in population-level analyses [11, 12]. Commonly used clustering methods for scRNA-seq data include TSCAN and SC3. TSCAN applies principal component analysis (PCA) to reduce the dimensionality of scRNA-seq data, followed by clustering in

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



low-dimensional space using a Gaussian mixture model (GMM) [13]. SC3 employs spectral clustering to generate clustering results from distance matrices based on Euclidean, Pearson, or Spearman metrics. These results are integrated into a shared feature matrix, which is analyzed subsequently through hierarchical clustering to determine the final clusters [14]. Although this is effective for single-omics data, these traditional clustering approaches are not optimized to exploit the complexity of multi-omics datasets, which limits their utility in integrative multi-omics analyses.

Deep learning methods have emerged as a dominant approach for single-cell data analysis because of their exceptional feature extraction capabilities [15]. Recently, various deep learning models have been developed to integrate bimodal data, which advances the integration of multi-omics datasets. For instance, the single-cell Multimodal Variational AutoEncoder (scMVAE) [16] has been utilized to integrate scRNA-seq and scATAC-seq data. This model combines probabilistic Gaussian mixture models with three distinct joint learning strategies to identify latent features that effectively represent multiomics data. However, embedding different omics modalities into a shared latent space can lead to the loss of omic-specific information, which potentially limits the model's capacity to capture unique omics characteristics.

In contrast to scMVAE, the Deep Cross-omics Cycle Attention (DCCA) model employs deep generative networks to process scRNA-seq and scATAC-seq data individually. It incorporates attention-transfer techniques to model interactions between omics layers while accounting for cellular heterogeneity [17]. scMCs are designed to tackle the individuality in data representation by capturing omic-specific features effectively, enabling better characterization of omic-specific variations. However, their strong emphasis on omic-specific features often limits the ability to extract shared characteristics across different omics layers [18]. The scMDC model integrates RNA and ATAC data or RNA and ADT data by directly concatenating the raw data and using an encoder to derive latent feature representations suitable for clustering [19]. scMDC employs a Zero-Inflated Negative Binomial (ZINB) model [20, 21] to represent single-cell RNA data more accurately, and it addresses dropout events that methods such as BREM-SC [22] and TotalVI [23] have limitations in to account for during the embedding or clustering process. However, scMDC also faces challenges to capture specific omics representations adequately.

This paper presents a novel single-cell multi-omics integration method, scCODI (single-cell Cross-Omics Data Integration), designed to integrate and analyze multimodal single-cell data. The scCODI workflow, illustrated in Figure 1, begins with the preprocessing of paired RNA and ATAC (or ADT) data from SNARE-seq, SMAGE-seq, and CITE-seq. After preprocessing, the data from two modality RNA and ATAC (or ADT) are concatenated to generate an integrated dataset. Both the integrated and single modal datasets are then fed into scCODI, where they undergo processing through two key modules: the global and cross-omics feature integration (COFInt) module and the global relationship-guided contrastive learning (GRgCL) module. The COFInt module performs global and cross-omics feature aggregation to obtain a shared representation of RNA and ATAC (or ADT) data, fully exploiting the complementarity between different omics within the same cell. The GRgCL module employs contrastive learning to align shared and omic-specific representations, ensuring the representations of the same cell across different omics become more similar. Additionally, it effectively mitigates the issue of negative pairs from the same cluster but belonging to different cells, which would otherwise result from low similarity scores. Together, these modules produce embeddings suitable for clustering. The resulting clusters are subsequently used for downstream tasks such as marker detection and pathway analysis. Extensive evaluations on SNARE-seq, SMAGE-seq, and CITE-seq data demonstrate that scCODI significantly outperforms previous methods in clustering multimodal single-cell data.

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



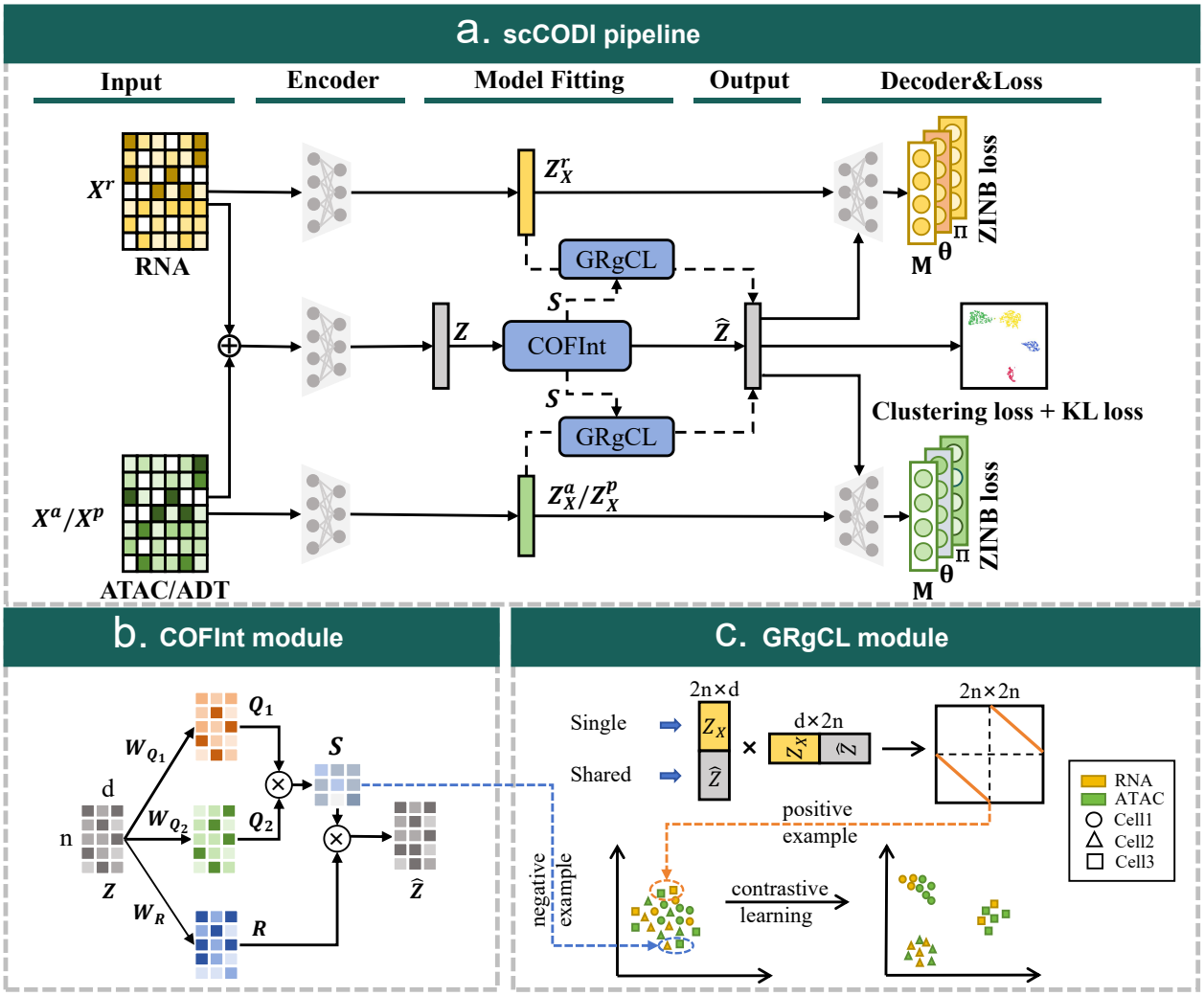



Figure 2. **The overall framework and key modules of scCODI**. (a) scCODI pipeline: The method integrates multimodal data from RNA and ATAC or RNA and ADT. The model includes an encoder to process concatenated raw data, along with two omic-specific encoders for each omic, and two decoders tailored to each omic. The encoded features are then passed through the Cross-omics Feature Integration (COFInt) module, which performs global and cross-omics feature aggregation, followed by the GRgCL (global relationship-guided Contrastive Learning) module for contrastive learning based on global relationships. The output consists of clustering representations and reconstructed data, optimized through clustering loss, KL divergence loss, and ZINB loss. (b) COFInt Module: The COFInt module utilizes matrices $W_{Q_1}$, $W_{Q_2}$, and $W_R$ for feature space transformation across omics, while **S** captures the global relationships among cells. (c) GRgCL Module: The GRgCL module applies contrastive learning to align shared and omic-specific representation. The blue dashed line indicates the selection of negative sample pairs, guided by the global relationship matrix **S**, while the orange dashed line denotes the identification of positive sample pairs, where the same cell exhibits similar representations in both the single modal representation and the shared representation.

## 2. Materials and methods

The network architecture of scCODI (Figure 2) includes three encoders: one for concatenated data and two for single modal data, along with two decoders tailored to each modality, it can be applied to clustering single-cell multi-omics RNA and ATAC (or ADT) datasets. The model integrates two

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



key components: COFInt and GRgCL. COFInt incorporates a transformer structure to capture global relationships across different feature spaces, producing a shared representation based on these relationships. The global relationship-guided contrastive learning module aligns the shared and omic-specific representations by selecting negative samples based on the global relationship matrix provided by COFInt, while treating the representations of the same cell in both shared and omic-specific spaces as positive samples.

### 2.1. Count data preprocessing

The raw data from SNARE-seq and SMAGE-seq are preprocessed and normalized using the Python package SCANPY [24]. The entire preprocessing process is illustrated in (Figure 1b). For each cell, the counts are normalized by a size factor that is denoted as $s_i$, with $s_i^a$ for ATAC data and $s_i^r$ for RNA data, which is calculated by dividing the library size of the cell by the median library size across all cells. This normalization ensures comparability by equalizing library sizes across cells. Subsequently, the counts are transformed logarithmically and standardized to achieve unit variance and zero mean. The processed count data for RNA and ATAC are utilized in the denoising multimodal autoencoder model, while the raw count matrix is retained for computing the zero-inflated negative binomial (ZINB) loss [15, 25]. CITE-seq data undergo a similar preprocessing and normalization procedure, with the size factor $s_i^p$ for ADT data calculated in the same manner.

### 2.2. Denoising multi-omics data encoder

In scCODI (Figure 2), RNA data is denoted as $\mathbf{X}^r \in \mathbb{R}^{N\times D^r}$, ATAC data as $\mathbf{X}^a \in \mathbb{R}^{N\times D^a}$, and ADT data as $\mathbf{X}^p \in \mathbb{R}^{N\times D^p}$. Autoencoders, a fundamental type of neural network, are highly effective for learning nonlinear representations of data [26]. Among their variants, denoising autoencoders are particularly well-suited for handling noisy datasets because they learn robust latent representations by reconstructing clean inputs from corrupted versions [27]. Given the inherently noisy nature of single-cell data, we employed denoising autoencoders for processing RNA, ADT, and ATAC data. The preprocessed RNA, ATAC, and ADT data are denoted as $\mathbf{X}^r$, $\mathbf{X}^a$, and $\mathbf{X}^p$, respectively. After introducing noise, these datasets are represented as $\mathbf{X}^r_{noise}$,$\mathbf{X}^a_{noise}$,$\mathbf{X}^p_{noise}$, which are formally defined as:

$$\mathbf{X}^r_{noise} = \mathbf{X}^r + \delta_r \cdot \eta_r, \tag{1}$$

$$\mathbf{X}^a_{noise} = \mathbf{X}^a + \delta_a \cdot \eta_a, \tag{2}$$

$$\mathbf{X}^p_{noise} = \mathbf{X}^p + \delta_p \cdot \eta_p, \tag{3}$$

For RNA, ATAC, and ADT data, we add artificial Gaussian noise $\eta_r$, $\eta_a$, and $\eta_p$, respectively. The noise weights $\delta_r$, $\delta_a$, and $\delta_p$ are set to 2.5 for all three data types.

Our autoencoder architecture includes a shared encoder for handling concatenated raw data, two omic-specific encoders for specific omics, and two decoders tailored to different omics datasets. Specifically, for the concatenated data, we denote the encoder for the concatenated RNA and ATAC data, and the encoder for the concatenated RNA and ADT data, as follows:

$$\mathbf{Z}^{ra} = E_w(\mathbf{X}^r_{noise} \oplus \mathbf{X}^a_{noise}), \mathbf{Z}^{rp} = E_w(\mathbf{X}^r_{noise} \oplus \mathbf{X}^a_{noise}), \tag{4}$$

Decoders are defined as:

$$(\mathbf{X}^{r'}, \mathbf{X}^{a'}) = D_{w'}(\mathbf{Z}_{ra}), (\mathbf{X}^{r'}, \mathbf{X}^{p'}) = D_{w'}(\mathbf{Z}_{rp}). \tag{5}$$

Regarding single omics data, we denote the encoder for the RNA, ATAC, and ADT data as:

$$\mathbf{Z}^r_X = E_{w_r}(\mathbf{X}^r_{noise}), \mathbf{Z}^a_X = E_{w_a}(\mathbf{X}^a_{noise}), \mathbf{Z}^p_X = E_{w_p}(\mathbf{X}^p_{noise}), \tag{6}$$

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



The decoder is defined as:

$$\mathbf{X}^{r'} = D_{w_{r'}}(\mathbf{Z}^{r}), \mathbf{X}^{a'} = D_{w_{a'}}(\mathbf{Z}^{a}), \mathbf{X}^{p'} = D_{w_{p'}}(\mathbf{Z}^{p}), \tag{7}$$

Here, $\mathbf{X}^{r'}$,$\mathbf{X}^{a'}$,$\mathbf{X}^{p'}$, represent the reconstructed data for RNA, ATAC, and ADT, respectively. $w$ and $w'$ correspond to the learnable weights of the encoder and decoder. The operator $\oplus$ denotes the concatenation of two matrices. For all the hidden layers within the encoders and decoders, the Exponential Linear Unit (ELU) [28] activation function was used. Additionally, dropout regularization and batch normalization were applied to the layer output to improve generalization and to stabilize training.

For all omics data, we used zero-inflated negative binomial (ZINB) as the reconstruction loss function [20]. The ZINB model is defined by three key parameters: the mean ($\mu_{\mathbf{X}}$) and dispersion ($\theta$) parameters of the negative binomial distribution, and a dropout probability coefficient ($\pi$) that accounts for zero-inflation due to dropout events. Formally, the model can be expressed as:

$$NB(\mathbf{X}; \mu_{\mathbf{X}}, \theta) = \frac{\Gamma(\mathbf{X}+\theta)}{\Gamma(\theta)\cdot \mathbf{X}!}\left(\frac{\theta}{\theta+\mu_X}\right)^{\theta}\left(\frac{\mu_X}{\theta+\mu_X}\right)^{X}, \tag{8}$$

$$ZINB(\mathbf{X}; \pi, \lambda, \theta) = \pi \cdot \delta_0(\mathbf{X}) + (1-\pi)\cdot \mathrm{NB}(\mathbf{X}; \mu_X, \theta). \tag{9}$$

To estimate parameters in the ZINB loss functions, we added three independent fully connected layers $\mathbf{M}$, $\mathbf{\Theta}$, and $\mathbf{\Pi}$ to the last hidden layer of each decoder. The layers are defined as:

$$\mathbf{M}^{\upsilon} = diag(s) \times \exp\left(w_{\mu}\mathbf{X}\right), \mathbf{\Theta}^{\upsilon} = \exp\left(w_{\theta}\mathbf{X}\right), \mathbf{\Pi}^{\upsilon} = \exp\left(w_{\pi}\mathbf{X}\right), \tag{10}$$

Here, let $\mathcal{D}$ represent the set of data types, where $\mathcal{D} = \{\mathrm{RNA}, \mathrm{ATAC}, \mathrm{ADT}\}$. For each data type $\upsilon \in \mathcal{D}$, we defined the following matrices: $\mathbf{M}^{\upsilon}$ as the estimated mean matrix, $\theta^{\upsilon}$ as the estimated dispersion matrix, and $\mathbf{\Pi}^{\upsilon}$ as the estimated dropout probability matrix. $w_{\mu}^{\upsilon}$, $w_{\theta}^{\upsilon}$, $w_{\pi}^{\upsilon}$ are the learnable weights. The size factors $s_i^{\upsilon}$ for each data type were determined in the preprocessing step.

Unlike traditional autoencoders that rely on mean squared error loss, the ZINB-based decoder network uses a loss function defined by the negative log-likelihood of the ZINB distribution:

$$\mathcal{L}_{\mathrm{ZINB}} = -\log(\mathrm{ZINB}(\mathbf{X} \mid \mathbf{\Pi}, \mathbf{M}, \mathbf{\Theta})). \tag{11}$$

Thus, the reconstruction loss for our autoencoder model using ZINB is the sum of the ZINB loss for specific omics and the ZINB loss for integrated omics data. For the integration of RNA and ATAC, the reconstruction loss is:

$$\mathcal{L}_{\mathrm{rec}} = \mathcal{L}_{\mathrm{ZINB}}^{\mathrm{r}} + \mathcal{L}_{\mathrm{ZINB}}^{\mathrm{a}} + \mathcal{L}_{\mathrm{ZINB}}^{\mathrm{ra}}, \tag{12}$$

where $\mathcal{L}_{\mathrm{ZINB}}^{\mathrm{r}}$ and $\mathcal{L}_{\mathrm{ZINB}}^{\mathrm{a}}$ are the ZINB losses for specific RNA and ATAC omics data, respectively, and $\mathcal{L}_{\mathrm{ZINB}}^{\mathrm{ra}}$ is the ZINB loss for the integrated RNA and ATAC data.

For the integration of RNA and ADT, the reconstruction loss is:

$$\mathcal{L}_{\mathrm{rec}} = \mathcal{L}_{\mathrm{ZINB}}^{\mathrm{r}} + \mathcal{L}_{\mathrm{ZINB}}^{\mathrm{p}} + \mathcal{L}_{\mathrm{ZINB}}^{\mathrm{rp}}, \tag{13}$$

where $\mathcal{L}_{\mathrm{ZINB}}^{\mathrm{r}}$ and $\mathcal{L}_{\mathrm{ZINB}}^{\mathrm{p}}$ are the ZINB losses for specific RNA and ADT omics data, respectively, and $\mathcal{L}_{\mathrm{ZINB}}^{\mathrm{rp}}$ is the ZINB loss for the integrated RNA and ADT data.

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



### 2.3. Global and Cross-omics Feature Integation

Although concatenating preprocessed data and directly inputting it into the autoencoder can facilitate learning shared features across different omics, this approach may overlook omic-specific patterns, the resulting data representations often contain omic-specific information that may be irrelevant or misleading. This issue can also impact the quality of clustering negatively [29]. Instead of simply using a weighted sum, we enhanced the shared representations of samples across different omics by focusing on those samples with strong inter-relationships within the same cluster.

To obtain a shared representation of multi-omics data, we first modeled the global relationships among samples and utilized these relationships to derive the shared feature representation (Figure 1b). Initially, the raw data was concatenated and input into the autoencoder, which resulted in the shared feature representation $\mathbf{Z}$, which is expressed as follows:

$$\mathbf{Z}^{ra} = [\mathbf{Z}^r, \mathbf{Z}^a], \mathbf{Z}^{rp} = [\mathbf{Z}^r, \mathbf{Z}^p], \tag{14}$$

where $\mathbf{Z}^r = [\mathbf{z}_1^r; \mathbf{z}_2^r; \ldots; \mathbf{z}_n^r] \in \mathbb{R}^{n \times d_r}$, $\mathbf{z}_i^r \in \mathbb{R}^{1 \times d_r}$, $\mathbf{Z} \in \mathbb{R}^{n \times d}$. Here, $n$ represents the number of samples, and $d_r$ represents the dimensionality of RNA data features in the shared feature representation derived from the autoencoder. The shared feature representations for ATAC and ADT data were obtained using the same method.

Inspired by the transformer attention mechanism [30], we projected $\mathbf{Z}$ into different feature spaces using $\mathbf{W}_R$ to facilitate cross-omics integration across all modalities:

$$\begin{bmatrix} \mathbf{R}_{1:} \\ \mathbf{R}_{2:} \\ \vdots \\ \mathbf{R}_{n:} \end{bmatrix} = \begin{bmatrix} \mathbf{z}_1^r & \mathbf{z}_1^a \\ \mathbf{z}_2^r & \mathbf{z}_2^a \\ \vdots & \vdots \\ \mathbf{z}_n^r & \mathbf{z}_n^a \end{bmatrix} \begin{bmatrix} \mathbf{W}_{R_{r:}} \\ \mathbf{W}_{R_{a:}}, \end{bmatrix} \tag{15}$$

where $\mathbf{R}_{j:} = \mathbf{z}_j^r \mathbf{W}_{R_{r:}} + \mathbf{z}_j^a \mathbf{W}_{R_{a:}}$; the joint representation mapping of RNA and ADT data can also be obtained using this method. Similarly, $\mathbf{Q}_1$ and $\mathbf{Q}_2$ were obtained using $\mathbf{W}_{Q1}$ and $\mathbf{W}_{Q2}$:

$$\begin{aligned} \mathbf{Q}_1 &= \mathbf{Z}\mathbf{W}_{Q_1}, \\ \mathbf{Q}_2 &= \mathbf{Z}\mathbf{W}_{Q_2}, \end{aligned} \tag{16}$$

where $\mathbf{Q}_1, \mathbf{Q}_2 \in \mathbb{R}^{n \times d}$. The parameters in this module are denoted by $\mathbf{W}_Z = \{\mathbf{W}_{Q1}, \mathbf{W}_{Q2}, \mathbf{W}_R\}$.

The global relationship among samples is represented as follows:

$$\mathbf{S} = \text{softmax}\left(\frac{\mathbf{Q}_1 \mathbf{Q}_2^T}{\sqrt{d}}\right), \tag{17}$$

where $\mathbf{S}$ is a matrix obtained by applying softmax normalization to the inner product of $\mathbf{Q}_1$ and $\mathbf{Q}_2$, which is scaled by $\sqrt{d}$.

This matrix $\mathbf{S}$ enhances the learned representation matrix $\mathbf{R}$ by incorporating global relationships among samples. Specifically, it allows each sample's representation to benefit from those of other samples with high correlations, thereby promoting similarity among data representations of samples within the same cluster. The enhanced shared feature representation $\hat{\mathbf{Z}}$ is computed as follows:

$$\hat{\mathbf{z}}_i = \sum_{j=1}^{n} \mathbf{S}_{ij} \mathbf{R}_{j:}, \hat{\mathbf{Z}} = [\hat{\mathbf{z}}_1; \hat{\mathbf{z}}_2; \ldots; \hat{\mathbf{z}}_n], \tag{18}$$

where $\mathbf{R}_{j:} \in \mathbb{R}^{1 \times d}$ represents the representation of the $j$-th sample in matrix $\mathbf{R}$, and $\mathbf{S}_{ij}$ denotes the relationship weight between the $i$-th and $j$-th samples. The resulting matrix $\hat{\mathbf{Z}} \in \mathbb{R}^{n \times d}$.

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



### 2.4. Global relationship-guided contrastive learning

We enhanced the learning of the common representation $\hat{\mathbf{Z}}$ by considering the global relationship matrix across all samples. Because samples within the same cluster should have similar shared feature representations across different omics data, their specific representations within each omics data should also be similar. Therefore, the shared feature representations from the same cluster and the omic-specific representations $\hat{\mathbf{Z}}_X$ should be mapped close to $\hat{\mathbf{Z}}$ [31, 32]. Based on this, we designed a global relationship-guided contrastive learning method [33], where the shared feature representation and the specific representation of the same sample under different omics were treated as positive pairs. For handling negative pairs, directly setting other samples as negative pairs may lead to inconsistent representations within the same cluster, which contradicts our clustering objective. To address this, we introduced a global relationship-guided multi-omics contrastive learning method (Figure 1c). Specifically, we computed the cosine distance between the shared feature representation $\hat{\mathbf{Z}}$ and the specific representation $\mathbf{Z}_X^v$ ($v \in \mathcal{D}$, $\mathcal{D} = \{\text{RNA}, \text{ATAC}, \text{ADT}\}$) to indicate their similarity:

$$\cos\left(\hat{\mathbf{Z}}_{i:}, \mathbf{Z}^v_{Xi:}\right) = \frac{\hat{\mathbf{Z}}^T_{i:}\mathbf{Z}^v_{Xi:}}{\left\|\hat{\mathbf{Z}}_{i:}\right\| \left\|\mathbf{Z}^v_{Xi:}\right\|}, \tag{19}$$

The loss function of the global relationship-guided contrastive learning is defined as:

$$\mathcal{L}_{\text{cl}} = -\frac{1}{2N}\sum_{i=1}^{N}\sum_{v=1}^{V}\log\frac{e^{\cos\left(\hat{Z}_{i:}, \mathbf{Z}^v_{Xi:}\right)/\tau}}{\sum_{j=1}^{N} e^{(1-\mathbf{S}_{ij})\cos\left(\hat{\mathbf{Z}}_{i:}, \mathbf{Z}^v_{Xj:}\right)/\tau} - e^{1/\tau}}, \tag{20}$$

In this equation, the relationship between $\cos\left(\hat{\mathbf{Z}}_{i:}, \mathbf{Z}^v_{Xj:}\right)$ and $\mathbf{S}_{ij}$ is determined by the value of $S_{ij}$ from Eq. (17) and the temperature parameter $\tau$. When $S_{ij}$ is small, $(1 - \mathbf{S}_{ij})$ is large, which means that the contribution of $\cos(\hat{\mathbf{Z}}_{i:}, \mathbf{Z}^v_{Xj:})$ to the loss function is large. Even if $\cos(\hat{\mathbf{Z}}_{i:}, \mathbf{Z}^v_{Xj:})$ is small, it still occupies a large proportion in the denominator, thereby affecting the overall loss value.

### 2.5. KL divergence in the latent layer

In cluster analysis, it is crucial to group similar data points into the same cluster. We incorporated a Kullback-Leibler (KL) divergence loss function [34], to strengthen the relationships between similar cells while preventing cluster centroids from becoming overly concentrated in the latent space. Similar to t-SNE [35], we utilized a $t$-distribution kernel function to represent the pairwise similarity between two cells $i$ and $i'$ in the latent space of our autoencoder:

$$q_{ii'} = \frac{(1 + |\mathbf{Z}_i - \mathbf{Z}_{i'}|^2)^{-1}}{\sum_{l \neq i}(1 + |\mathbf{Z}_i - \mathbf{Z}_l|^2)^{-1}}. \tag{21}$$

Here, $q_{ii} = 0$. During training, the target distribution $P$ adjusts affinities between cells by amplifying the similarities of highly similar cells and diminishing those with low similarities. $P$ is derived by the square of $Q$, followed by normalization:

$$p_{ii'} = \frac{q^2_{ii'}}{\sum_{l \neq i} q^2_{il}}. \tag{22}$$

With these two similarity distributions, $\mathbf{Q}$ and the corresponding target distribution $\mathbf{P}$, the KL divergence-based loss function is defined as follows:

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



$$\mathcal{L}_{\mathrm{kl}} = \mathrm{KL}(\mathbf{P}\|\mathbf{Q}) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}, \tag{23}$$

this loss function measures the difference in probability distributions between $P$ and $Q$. During training, $\mathbf{P}$ and $\mathbf{Q}$ were calculated per batch.

### 2.6. Deep K-means clustering objective function

We apply unsupervised clustering to the latent representations generated by the model [34]. The autoencoder learns a nonlinear mapping for each cell, which transforms the original concatenated data into a low-dimensional space $\hat{\mathbf{Z}}$ through cross-omics integration. The loss function $\mathcal{L}_c$ for K-means clustering is defined as:

$$\mathcal{L}_c = \sum_{i=1}^{N} \sum_{j=1}^{K} w_{ij} \tau f(\hat{\mathbf{Z}}_i, \mathbf{V}_j), \tag{24}$$

In this equation, $\mathbf{V}$ represents the $K$ clustering centroids, and $f$ computes the Euclidean distance between a cell in the latent space and a given centroid. The hyperparameter $\tau$ is assigned different values depending on the data type. To enhance the smoothness of the gradient descent optimization process, weights were determined using a Gaussian kernel function:

$$\widetilde{w}_{ij} = \frac{\exp(-f(\mathbf{Z}_i, \mathbf{V}_j))}{\sum_{k=1}^{K} \exp(-f(\mathbf{Z}_i, \mathbf{V}_k))}. \tag{25}$$

To accelerate convergence, the weights were inflated, and the final $w_{ij}$ was given by:

$$w_{ij} = \frac{\widetilde{w}_{ij}^{\alpha}}{\sum_{k=1}^{K} \widetilde{w}_{ij}^{\alpha}}, \tag{26}$$

where the hyperparameter $\alpha$ is assigned a value of 2.

In the proposed framework, the total loss in our network is defined as:

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_{\mathrm{rec}} + \lambda \mathcal{L}_{\mathrm{cl}} + \mu \mathcal{L}_{\mathrm{kl}} + \gamma \mathcal{L}_{\mathrm{c}} \\ &= \mathcal{L}_{\mathrm{rec}} \left(\{\mathbf{X}^r, \mathbf{X}^a, \mathbf{X}^{ra}\}; \{w, we\}\right) \\ &\quad + \lambda \mathcal{L}_{\mathrm{cl}} \left(\{\mathbf{Z}, \hat{\mathbf{Z}}, \mathbf{Z}_X^v\}_{v=1}^V; \{\mathbf{W}_Z, \mathbf{W}_{\hat{\mathbf{Z}}}, \mathbf{W}_{\mathbf{Z}_X^v}, we\}_{v=1}^V\right) \\ &\quad + \mu \mathcal{L}_{\mathrm{kl}} \left(\{\mathbf{X}^r, \mathbf{X}^a\}; \{w\}\right) \\ &\quad + \gamma \mathcal{L}_{\mathrm{c}} (\hat{\mathbf{Z}}; \mathbf{W}_{\hat{Z}}, \mathbf{U}), \end{aligned} \tag{27}$$

where $\mathcal{L}_{\mathrm{rec}}$ represents the reconstruction loss of three types of data, which include RNA, ATAC, and the joint reconstruction of the original RNA and ATAC data. The RNA and ADT data are processed similarly. $w$ and $w'$ are the parameters of the encoder and decoder. $\mathcal{L}_{\mathrm{cl}}$ represents the loss induced by contrastive learning, which captures the discrepancy between the relationship-enhanced shared feature representation and the omic-specific representations of specific omics data. $\mathcal{L}_c$ denotes the clustering loss, where the latent representations iteratively converge toward the centroids initialized by K-means, which are represented as $U$. In this framework, $\lambda$, $\mu$, and $\gamma$ are hyperparameters that regulate the weights of contrastive loss, KL divergence loss, and K-means clustering loss, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



# 3. Results

## 3.1. Experimental setup

**Datasets:** scCODI integrates different sets of single-cell omics data. In our experiments, we evaluated the model's ability to integrate RNA with chromatin data and RNA with protein data.

For the integration of RNA and ATAC, we used four preprocessed single-cell multi-omics datasets with paired profiles from a previous study [16]: (i) CellMix, which comprised 1,047 cells, obtained from GEO (GSE126074), where gene expression and chromatin accessibility were co-assayed simultaneously in each single cell using SNARE-seq; (ii) AdBrain, which consisted of 10,309 cells, also downloaded from GEO (GSE126074), where gene expression and chromatin accessibility were measured in single cells derived from the adult mouse cerebral cortex; and (iii) Human peripheral blood mononuclear cells (PBMCs) datasets, which consisted of approximately 3,000 cells and 10,000 cells.

For the integration of RNA and ADT, we collected five single-cell multi-omics datasets and preprocessed them: (i) Cord blood mononuclear (BMNC) cells with 1128 cells downloaded from GEO (GSE100866). the data were annotated using marker genes and ADTs [22] ; (ii) Bone marrow mononuclear cells with 30,672 cells were downloaded from GEO (GSE128639), and the cell type labels were downloaded from the "bmcite" dataset in the "SeuratData" package (v0.2.1); (iii) The PBMC dataset with 3762 cells that was available on the 10X website. We downloaded the preprocessed data and the cell type labels from the GitHub of Specter [36]; (iv) Mouse spleen lymph node with 16,828 cells was downloaded from GEO (SLN111, GSE150599); and (v) Mouse spleen lymph node with 15,820 cells was downloaded from GEO (SLN206, GSE150599). TotalVI provided cell-type labels on GitHub.

**Evaluation protocols:** For our multi-omics data integration clustering, we applied the K-means algorithm to the final low-dimensional co-embedding representation $\hat{\mathbf{Z}}$. Subsequently, we evaluated the clustering performance using Normalized Mutual Information (NMI) [37] and Adjusted Rand Index (ARI) [38]. Both NMI and ARI values ranged from 0 to 1, with higher values indicating better clustering performance.

**Comparing methods:**We implemented scCODI using the PyTorch deep learning framework and compared its performance with other single-cell multi-omics data integration methods.

For the integration of RNA and ATAC, we used the following methods: (i) **scMCs**, which employed a deep learning model for joint representation of multi-omics data, mapping them to a shared low-dimensional space and enhancing the relationships between different omics through contrastive learning to improve clustering performance [18]; (ii) **scMDC**, which used a multi-branch neural network to learn shared representations and applied clustering algorithms for identification of cell clusters [19]; (iii) **scMVAE**, which incorporated three strategies—scMVAE-PoE, scMVAE-NN, and scMVAE-Direct—for learning joint latent features to facilitate data fusion and clustering. Specifically, scMVAE-Direct concatenated the raw features from each omics dataset, scMVAE-NN integrated low-dimensional features extracted from different omics, and scMVAE-PoE utilized the product of experts framework to estimate a joint posterior distribution [16]; and (iv) **DCCA**, which projected different omics into their respective low-dimensional spaces and used a 'teacher-student' mechanism to fuse multi-omics data [17].

To integrate RNA and ADT, we compared with two multi-omics integration methods for RNA and ADT data integration. (i) **BREM-SC** [22] uses a Bayesian model to embed RNA and protein data into a shared low-dimensional space for joint analysis; and (ii) **TotalVI** [23] , which is a deep generative model designed specifically for the joint analysis of single-cell RNA and protein data. It utilizes

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



variational inference to model the complex relationships between transcriptomics and proteomics data. Additionally, we used the **scMDC** deep learning method for comparative experiments. scMDC demonstrates robust performance in integrating CITE-seq data, validating its effectiveness across different multi-omics modalities.

### 3.2. Evaluation on real RNA and ATAC integration data

Initially, we evaluated the clustering performance of scCODI on SNARE-seq and SMAGE-seq datasets, benchmarking it against 10 alternative methods. These approaches included models designed specifically for multimodal data clustering (DCCA, scMVAE, scMDC, and scMCs), two single-modal data clustering tools (SC3 and TScan), and two general clustering algorithms (IDEC and K-means). We evaluated these methods using four paired RNA and ATAC datasets; two datasets were derived from SNARE-seq and two were derived from SMAGE-seq (Table 1).

Using NMI and ARI metrics across different datasets, multimodal methods consistently outperformed single-modal approaches (Figure 3a). scCODI achieved superior performance across both metrics on all four datasets, surpassing other multimodal methods significantly, such as scMVAE-PoE, scMVAE-Direct, and scMVAE-NN (Figure 3a). This highlights the advantages of scCODI. Among these methods, scMVAE-Direct exhibited the weakest performance, likely due to the increased sparsity and complexity that arise from concatenating high-dimensional features. scMVAE-NN outperformed both scMVAE-Direct and scMVAE-PoE by leveraging a more compact feature space for representation. However, scMVAE failed to capture the unique characteristics of multi-omics data in cell clustering adequately. Conversely, scCODI balances the integration of shared features effectively, which is critical for consistent clustering, with the retention of specific features from single-omics views. Although DCCA uses separate neural networks to project multi-omics data into distinct representation spaces, it underperforms compared with scCODI due to its emphasis on specific omics features and insufficient consideration of shared features that are crucial for consistent clustering. In contrast, scCODI extracts both shared and specific features from different omics datasets simultaneously. Additionally, through contrastive learning, scCODI enhances feature similarity across different omics for the same samples, thereby capturing cellular heterogeneity more comprehensively to achieve superior clustering performance.

The positive values indicate that scCODI outperforms its counterparts (Figure 3b). The analysis reveals that the median differences in NMI and ARI were approximately 0.1, while the median

Table 1. Performance of clustering for different methods on RNA and ATAC integration datasets.The symbol '±' represents the standard deviation over five repeated experiments. The best results are highlighted in **bold font**, and the second-best results are indicated by underline.

| | GSE126074 | | Adbrain | | SMAGE10× 3k | | SMAGE10× 10k | |
|---|---|---|---|---|---|---|---|---|
| | NMI | ARI | NMI | ARI | NMI | ARI | NMI | ARI |
| SC3 | .942±.000 | .966±.000 | .418±.003 | .319±.001 | .551±.003 | .510±.002 | .501±.001 | .446±.002 |
| Kmeans | .726±.000 | .877±.000 | .383±.000 | .143±.000 | .464±.000 | .352±.000 | .501±.000 | .343±.000 |
| Tscan | .772±.001 | .804±.001 | .253±.003 | .189±.003 | .258±.002 | .120±.004 | .401±.001 | .376±.001 |
| IDEC | .518±.001 | .460±.001 | .149±.005 | .092±.003 | .300±.001 | .079±.001 | .253±.001 | .067±.001 |
| DCCA | .898±.000 | .940±.000 | .357±.002 | .406±.003 | .508±.000 | .583±.000 | .560±.000 | .600±.000 |
| scMCs | .869±.000 | .907±.000 | .510±.000 | .554±.001 | .278±.001 | .121±.000 | .393±.000 | .224±.001 |
| scMDC | .918±.000 | .952±.000 | .588±.001 | .434±.001 | .638±.001 | .516±.002 | .589±.000 | .613±.001 |
| scMVAE-POE | .899±.000 | .935±.000 | .457±.003 | .219±.005 | .503±.001 | .285±.001 | .531±.001 | .310±.001 |
| scMVAE-Direct | .892±.000 | .927±.000 | .302±.003 | .137±.001 | .506±.004 | .305±.001 | .540±.001 | .331±.001 |
| scMVAE-NN | .921±.000 | .952±.000 | .468±.003 | .222±.004 | .499±.000 | .265±.000 | .553±.001 | .311±.000 |
| scCODI | **.947±.000** | **.978±.000** | **.595±.005** | **.584±.006** | **.654±.003** | **.606±.006** | **.663±.001** | **.721±.001** |

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



differences in ARI was around 0.2, which underscores the advantages of scCODI. Furthermore, we ranked all methods based on their performance metrics for each dataset. scCODI consistently achieved the top rank in both metrics after averaging the rank of each method across the four datasets (Figure 3c). The second-best method, scMDC, attained an average rank of 2 for both NMI and ARI. Overall, these results from multiple real datasets demonstrated the stability and robustness of scCODI in clustering performance on SNARE-seq and SMAGE-seq datasets.

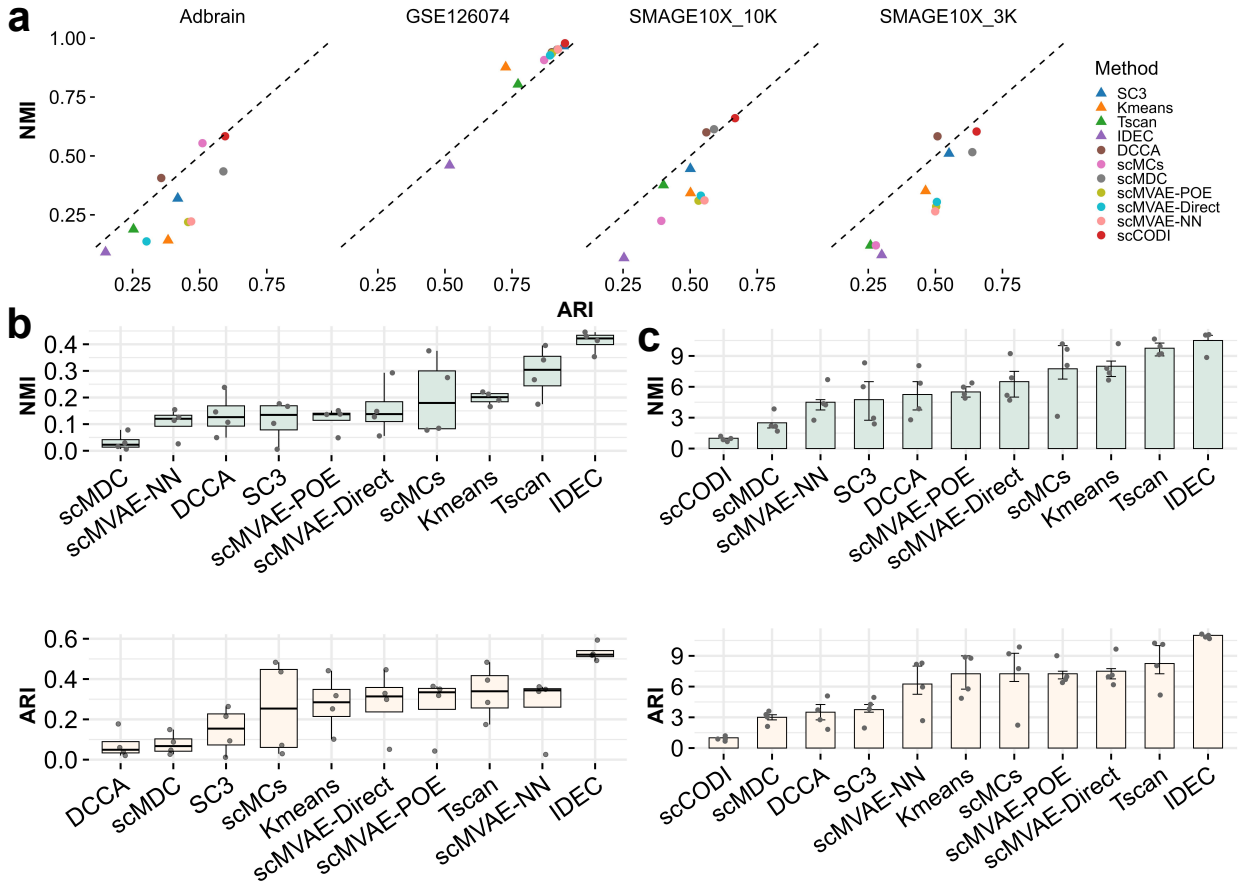



Figure 3. **Clustering performance of scCODI and the competing methods on different ATAC datasets.** The clustering performance of scCODI and competing methods across different ATAC datasets. All methods were evaluated across four datasets. Panels (a) of the figure visualizes clustering performance in a two-dimensional representation. Panel (b) uses box plots to illustrate the performance differences between scCODI and other methods. Each box plot displays key statistical measures, which include the minimum, first quartile (Q1), median, third quartile (Q3), and maximum values. The interquartile range (IQR), which is the difference between the third quartile (Q3) and the first quartile (Q1). The minimum and maximum are determined as the smallest and largest data points that are greater than or equal to Q1 - 1.5IQR and less than or equal to Q3 + 1.5IQR, respectively. Each dot represents a data point, which represents a performance difference in a dataset. Panels (c) summarize the average ranking of all methods, with dots indicating a method's rank for a specific dataset and error bars representing standard errors.

Furthermore, to demonstrate the quality of the final low-dimensional co-embedding representation $\hat{\mathbf{Z}}$, we used uniform manifold approximation and projection (UMAP) [39] to visualize the cell clustering points of scCODI and other baseline methods across various benchmark datasets.The clustering results of baseline methods on AdBrain, SMAGE10X 3K, and SMAGE10X 10K are presented in the

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Supplementary Material (Figures S3 to S5), which show the visualization and clustering results of different methods on both raw and imputed data. scCODI demonstrated more compact separation of samples within the same cluster and better separation of samples from different clusters (Figure 4). This validated the effectiveness of our proposed data augmentation and contrastive learning methods and explained why scCODI achieved superior clustering performance.

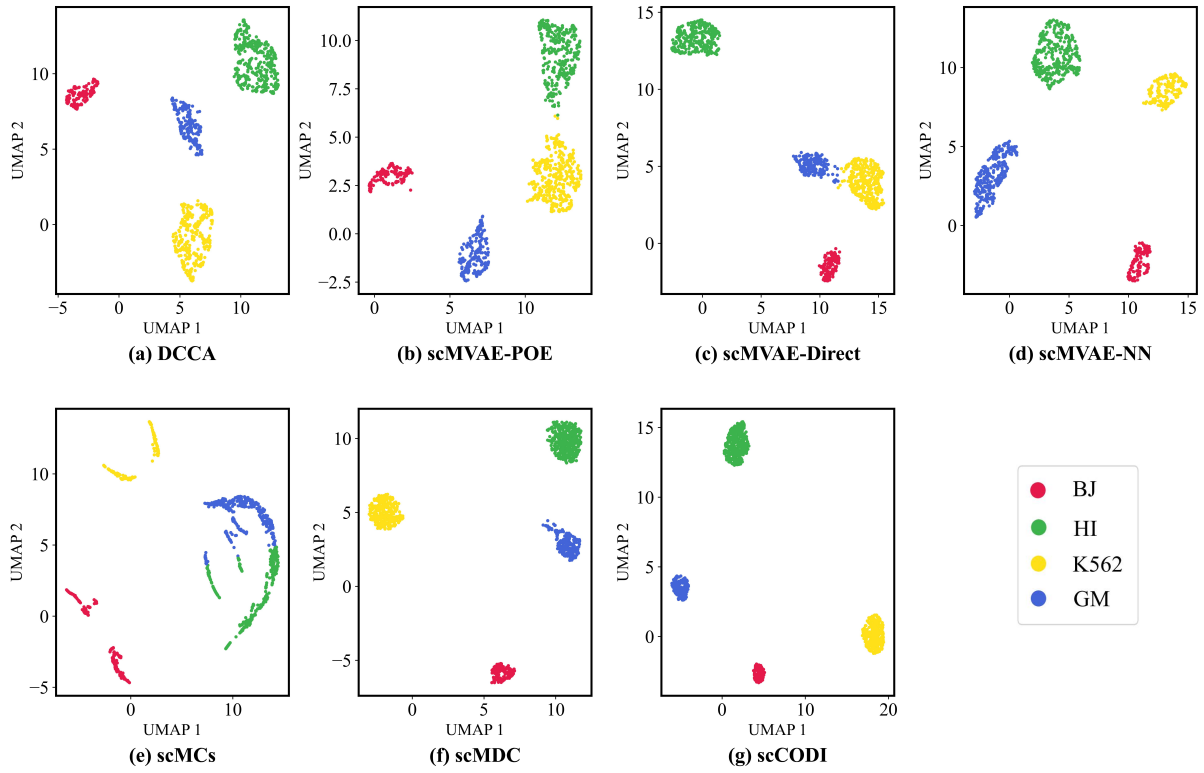



Figure 4. Cell clustering visualization of each method on GSE126074 dataset. (a) DCCA, (b) scMVAE-POE, (c) scMVAE-Direct, (d) scMVAE-NN, (e) scMCs, (f) scMDC, and (g) scCODI.

### 3.3. Evaluation on real RNA and ADT integration data

We evaluated the clustering performance of scCODI on the CITE-seq dataset and compared it with recent competitive methods, which included BREM-SC that was designed for clustering multimodal data, TotalVI, which addresses batch effects, and scMDC, which was developed for learning embeddings of single or multiple modal data. We tested these methods on five CITE-seq datasets. scCODI outperformed the competing methods across all five datasets (Figure S1), except for the PBMCspec-

Table 2. Performance of clustering for different methods on CITE-seq datasets. The best results are highlighted in **bold font**, and the second-best results are indicated by underline.

| | BMNC100866 | | BMNC128639 | | PBMC_spector | | spleen111 | | SLN206 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | NMI | ARI | NMI | ARI | NMI | ARI | NMI | ARI | NMI | ARI |
| BREM-SC | 0.8036 | 0.7424 | 0.6194 | 0.7285 | 0.5295 | 0.5111 | 0.3991 | 0.2538 | 0.4069 | 0.2765 |
| TotalVI | 0.7876 | 0.8279 | 0.7403 | 0.4645 | 0.6455 | 0.5016 | 0.6069 | 0.2551 | 0.6114 | 0.2529 |
| scMDC | 0.9679 | 0.9881 | 0.8567 | 0.8934 | 0.6377 | 0.5160 | 0.6593 | 0.5452 | 0.6068 | 0.3590 |
| scCODI | **0.9904** | **0.9950** | **0.8631** | **0.9075** | **0.7180** | **0.6149** | **0.6677** | **0.5934** | **0.6281** | **0.4249** |

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



tor dataset, where scMDC exhibited comparable performance. Although TotalVI was less effective than scMDC, it outperformed BREM-SC due to its ability to correct batch effects. The performance differences between scCODI and the competing methods are summarized (Figure S1a). A positive difference indicates superior performance of scCODI. Our analysis reveals that scCODI consistently outperformed all other methods across multiple datasets. Furthermore, we ranked all methods based on their performance metrics for each dataset. scCODI achieved the highest average rank across the five datasets for both metrics consistently (Figure S1c). scCODI ranked first in both metrics across all datasets. In contrast, the second-best method, scMDC, had an average rank of 2 for both NMI and ARI. Overall, our results across various real datasets confirmed that scCODI demonstrated stable and robust clustering performance on CITE-seq datasets (Table 2).

We used UMAP to visualize the cell clustering of scCODI and other baseline methods across various benchmark datasets after processing the CITE-seq data with our model. scCODI achieved clearer cluster boundaries, lower error rates, and a higher concentration of points within the same cluster (Figure S2). The clustering results on other CITE-seq integration data are presented in the Supplementary Material (Figures S6 to S9). These observations validated the superior clustering performance of scCODI and underscored the effectiveness of our innovative approach.

### 3.4. Evaluation of multimodal integrated data

As discussed in the introduction, different omics data provide unique and complementary information, which is essential for accurate cell clustering and identification of cell types. Consequently, integrating multi-omics data is expected to yield superior clustering performance compared with approaches that rely solely on single-source data. To evaluate the benefits of multimodal integration, we used two evaluation strategies. The first strategy involved comparing our model with methods focused on unimodal clustering. In earlier sections, we compared scCODI with existing single-cell clustering methods (SC3 and TSCAN) and two general clustering methods (IDEC and K-means), which are commonly used for scRNA-seq data. We extracted RNA gene expression matrices, chromatin accessibility matrices, and protein expression matrices from the four previously mentioned RNA and ATAC multimodal datasets and from five CITE-seq datasets. These matrices were then subjected to cluster analysis, and their performance was evaluated using NMI and ARI. The experimental results demonstrate the advantages of multimodal integration over unimodal methods (Figure 3a and Figure 3b).

Additionally, because scCODI's encoder is capable of performing cluster analysis on single-modal data, we got three scCODI variant models by keeping one omic dataset while setting the other omic data to all-zero values: a sub-model of scCODI with only RNA input and reconstruction loss (referred to as scCODI-RNA), and sub-models with only ADT or ATAC input and reconstruction loss (referred to as scCODI-ADT and scCODI-ATAC, respectively). scCODI outperformed the three variant models consistently across CITE-seq, SMAGE-seq, and SNARE-seq datasets; scCODI-RNA achieved the second-best performance (Figure 5a). This result aligns with our expectations because scRNA-seq data provide richer gene expression information and higher feature dimensions, which allowed the model to learn and to predict cell states more comprehensively.

In contrast, ADT data contained fewer features and more limited information, which resulted in lower performance when used alone compared with scRNA-seq data. On the other hand, scCODI-ATAC exhibited poorer performance across the four SNARE-seq and SMAGE-seq datasets (Figure 5b). Subsequently, we selected the higher-performing subset of the variant models across nine datasets for comparison with scCODI. There was a clear upward trend in performance with multimodal integration (Figure 5c).

scCODI is capable of integrating information from multimodal data to enhance clustering per-

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



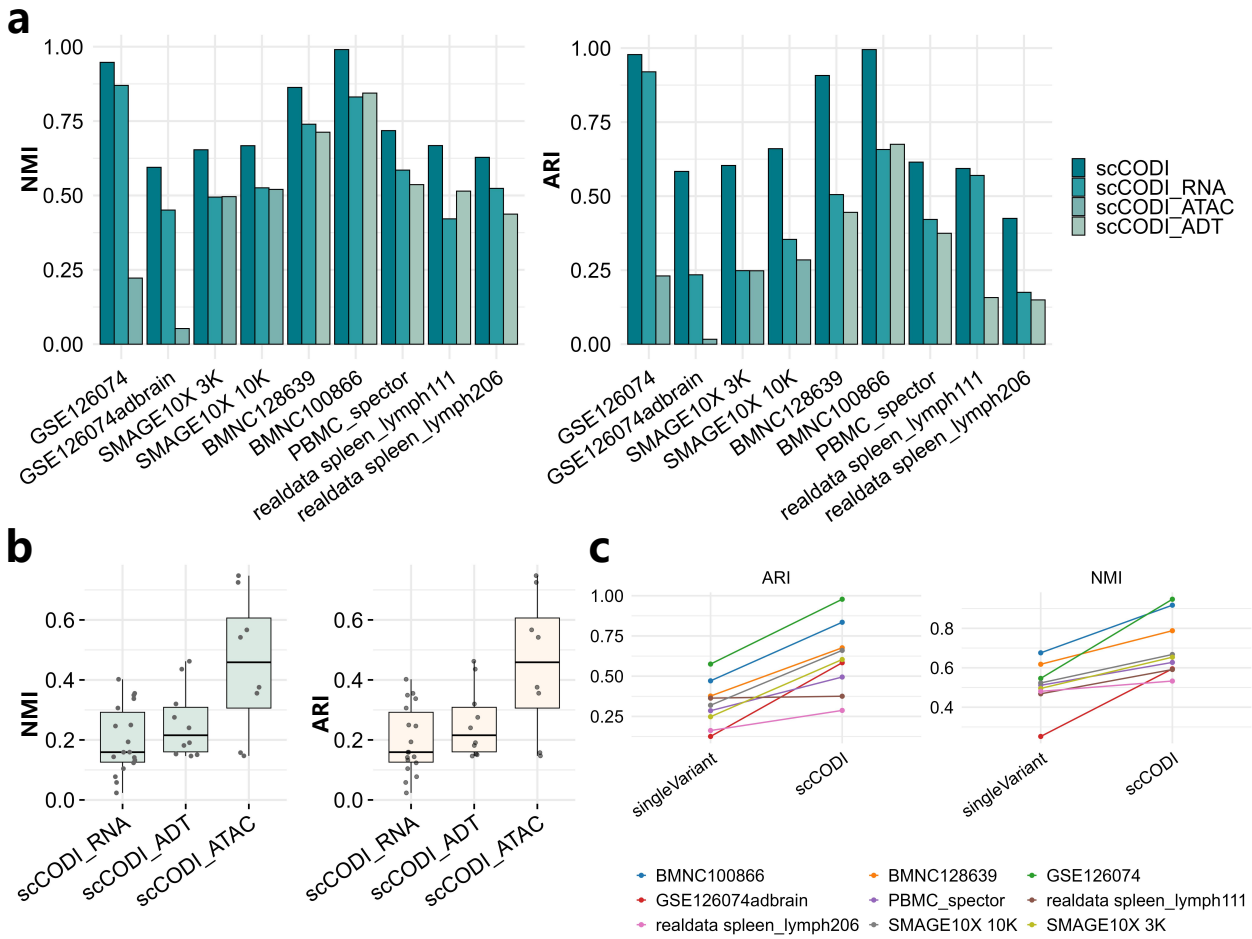



Figure 5. **Clustering performance of scCODI and its variant models on multimodal datasets.** scCODI, scCODI-RNA, and scCODI-ADT were evaluated across five CITE-seq datasets, and scCODI, scCODI-RNA, and scCODI-ATAC were evaluated across four SMAGE-seq datasets and SNARE-seq. (a) shows present clustering performance as a bar chart. (b) illustrates the performance differences between scCODI and other methods using box plots, separately for the CITE-seq datasets (left, n = 5) and the SMAGE-seq datasets (right, n = 4). Each box plot shows the minimum, first quartile (Q1), median, third quartile (Q3), and maximum values. The interquartile range (IQR), which is the difference between the third quartile (Q3) and the first quartile (Q1). The minimum and maximum are determined as the smallest and largest data points that are greater than or equal to Q1 - 1.5IQR and less than or equal to Q3 + 1.5IQR, respectively. The comparison between scCODI and its variants is shown in line plots in panel (c). Lines connect the paired performance for each dataset between the two methods.

formance. To this end, we constructed UMAP plots (Figures S10) for scCODI and the three variant models.

### 3.5. Downstream analysis

To validate the clustering results derived from the latent representations, we conducted two standard downstream analyses: differential expression (DE) analysis and gene set enrichment analysis (GSEA) [40]. DE analysis was performed using the Scanpy package, which enables pairwise comparisons between two clusters or between a target cluster and the remaining clusters. Based on the ranked

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



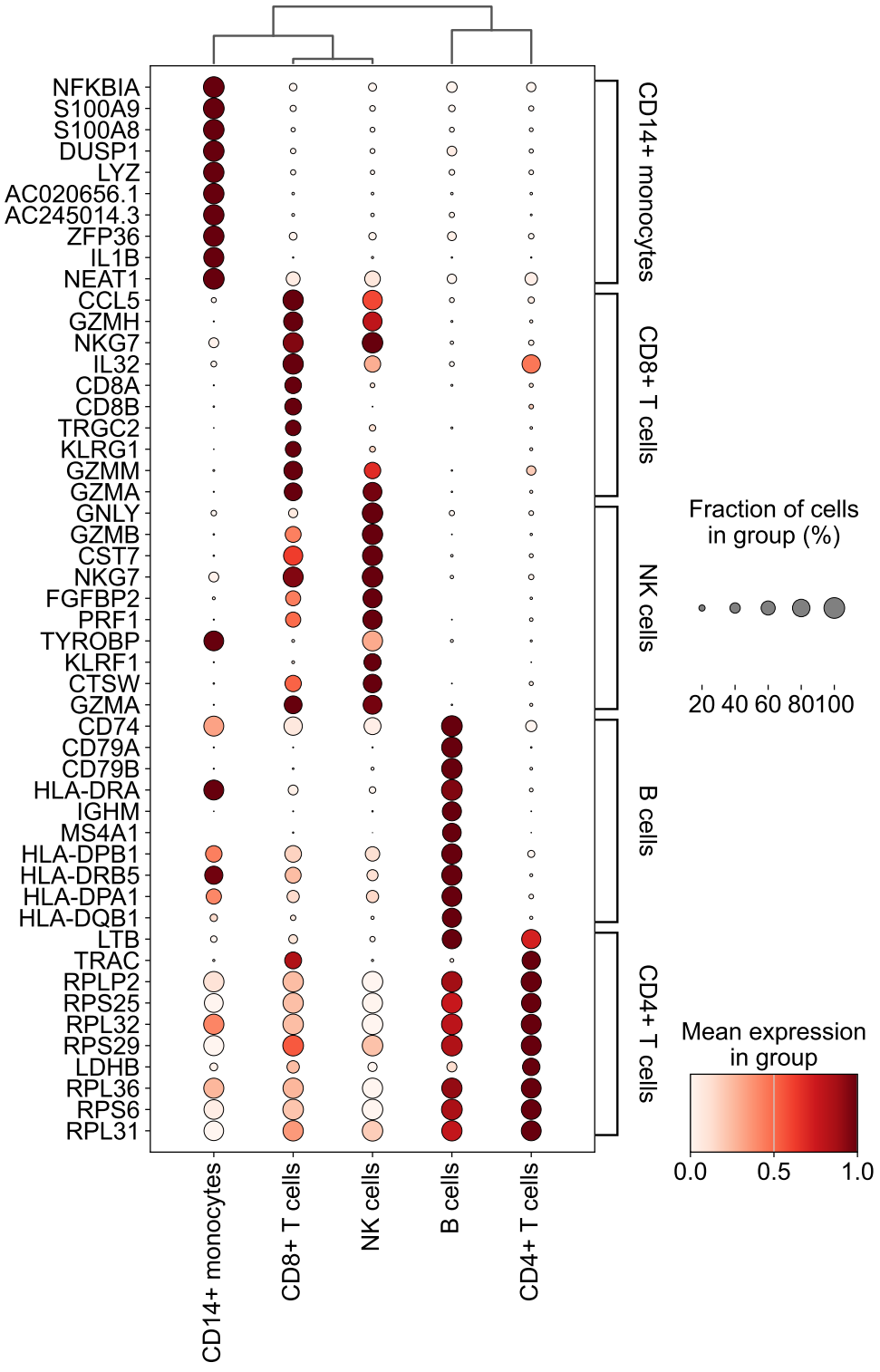



Figure 6. **Dot plot visualization of marker gene expression across cell clusters.** The dot plot illustrates the expression levels and the fraction of cells expressing specific marker genes across different cell types, including CD14+ monocytes, CD8+ T cells, NK cells, B cells, and CD4+ T cells. The size of the dots represents the percentage of cells in each group expressing a given gene, while the color intensity indicates the mean expression level of the gene within the group. Hierarchical clustering is applied to group genes and cell types based on their expression patterns.

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



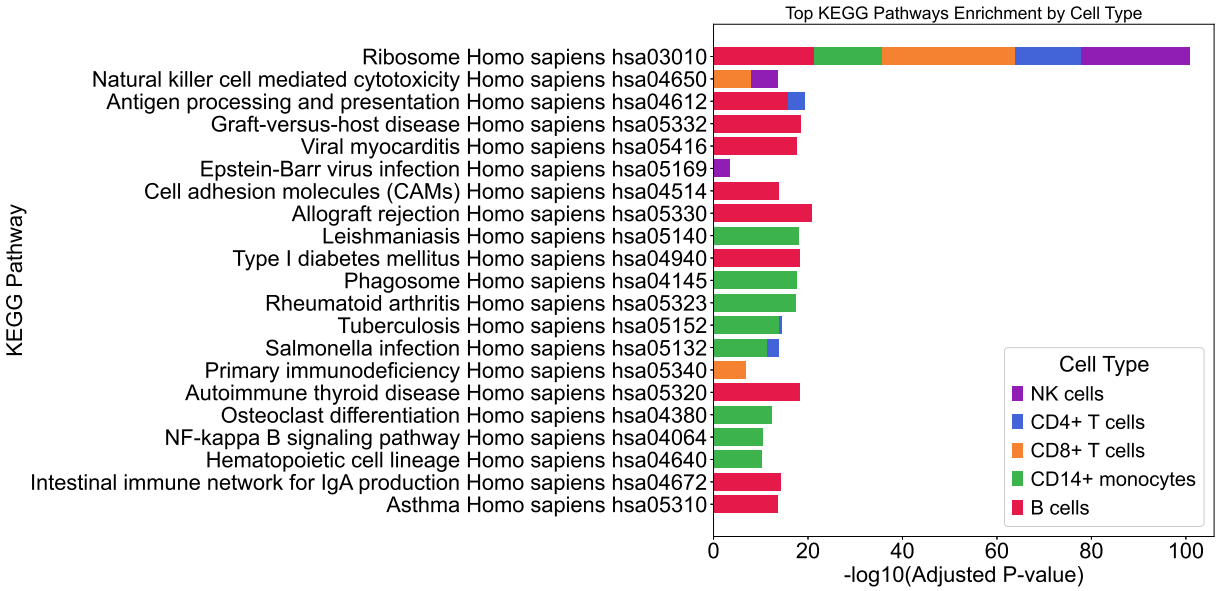



Figure 7. **KEGG Pathway Enrichment Analysis by Cell clusters**. This bar chart illustrates the KEGG pathway enrichment results for different cell types, including NK cells, CD4+ T cells, CD8+ T cells, CD14+ monocytes, and B cells. Each bar represents a KEGG pathway, with its length indicating the -log10 (adjusted P-value) as a measure of significance. Different colors correspond to different cell types. The pathway names are listed on the left, and longer bars indicate higher enrichment significance for the corresponding cell type in the respective pathway.

gene lists and their directions, GSEA was applied to identify enriched pathways within target clusters. As an example, we present the results from the BMNC100866 dataset (GSE100866), which focused on five clusters: B cells, CD14+ monocytes, CD4+ T cells, CD8+ T cells, and NK cells. DE and GSEA analyses were performed for each cluster compared to the others. The differentially expressed genes for these five cell clusters exhibited numerous well-known marker genes for each cell type (Figure 6). For example, LYZ was highly expressed in CD14+ monocytes [41]. MS4A1 is a marker gene for B cells, CD8A and CD8B are marker genes for CD8+ T cells [42], and NKG7 is highly expressed in NK cells [43].

The analysis of enriched pathways revealed the significant roles of different cell types in specific functional pathways (Figure 7). For instance, NK cells showed a marked enrichment in the "Natural killer cell-mediated cytotoxicity" pathway, highlighting their crucial role in immune defense [44, 45]. CD14+ monocytes exhibited significant enrichment in pathways such as "Antigen processing and presentation" and "NF-kappa B signaling pathway", which further highlight their central role in immune regulation and inflammatory response [46].CD8+ T cells were notably enriched in the "T cell receptor signaling pathway", which indicate the importance of this pathway in T cell function and antiviral defense [47, 48]. Similarly, CD4+ T cells were enriched in the "Antigen processing and presentation pathway", which demonstrate their critical role in immune response and cellular activation [49, 50]. B cells showed significant enrichment in pathways, such as the "Intestinal immune network for IgA production" and the "Complement system", which emphasizes their pivotal role in humoral immunity [51]. These findings further validate the specificity of functional activities across different immune cell types and support the accuracy and reliability of the clustering results obtained through our method.

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



### 3.6. Ablation study and parameter sensitivity analysis

**Convergence analysis.** To verify the convergence of the model, we analyzed the iterative changes in the objective values and evaluation metrics (Figure 8). The results demonstrated that while the objective value generally decreased over iterations, minor fluctuations were observed during the process before it eventually stabilized. Similarly, the NMI and ARI values showed an initial gradual increase during the iterative process, followed by stabilization and minor fluctuations within a narrow range. These observations confirmed the convergence behavior of scCODI.

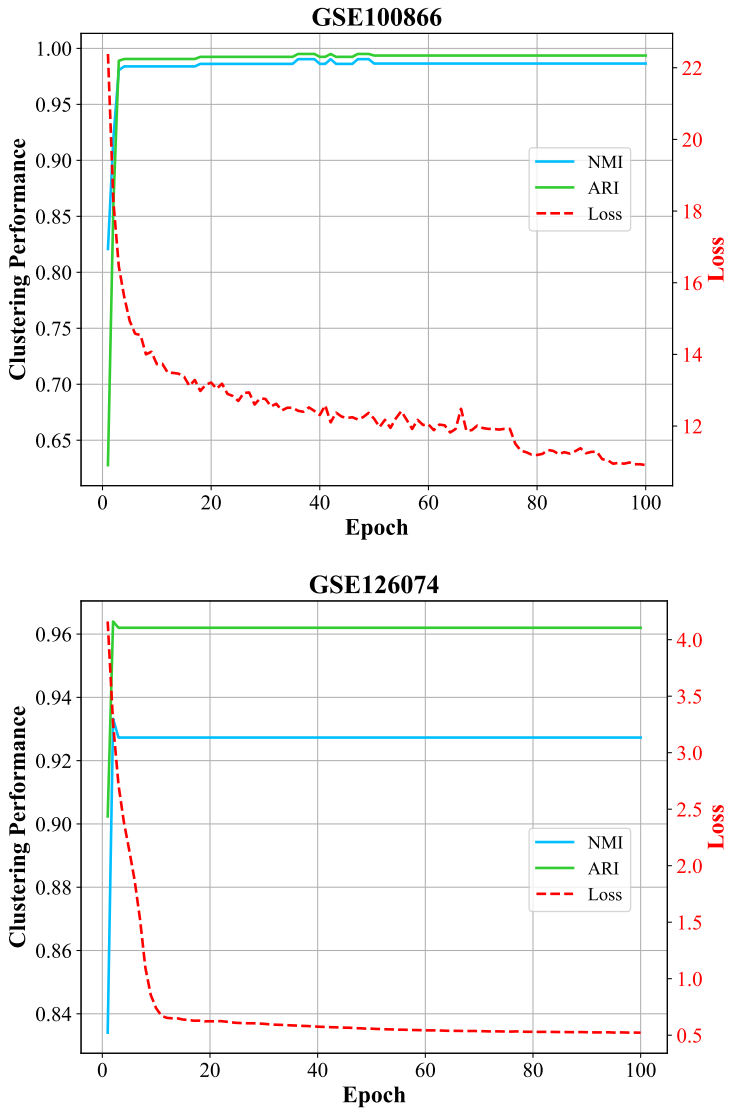



Figure 8. **Clustering Performance and Loss Curve**. This line plot illustrates the clustering performance and loss curve over training epochs for the BMNC100866 and GSE126074 datasets. The Y-axis on the left represents clustering performance metrics, including NMI and ARI, while the Y-axis on the right represents the loss value. The X-axis indicates the number of epochs. The blue and green lines represent the NMI and ARI scores, respectively, demonstrating convergence as training progresses. The red dashed line represents the loss, which decreases rapidly during the initial epochs and stabilizes as training continues.

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



**Parameter sensitivity analysis.** We evaluated the effect of hyperparameters on the clustering performance of scCODI, which included the trade-off coefficients $\lambda$, $\mu$, $\gamma$, and the temperature parameter $\tau$. NMI and ARI of scCODI when $\lambda$ varied from $10^{-2}$ to $10^{2}$, $\mu$ varied from 0 to 1, $\gamma$ varied from $10^{-2}$ to $10^{2}$, and $\tau$ varied from 0.2 to 0.8 . The clustering results of scCODI were insensitive to $\lambda$ and $\tau$ when $\lambda$ was in the range of 0.1 to 1, when $\mu$ was in the range of 0.001 to 0.005, when $\gamma$ is in the range of 0.01 to 0.1, and $\tau$ is in the range of 0.3 to 0.5 (Figure 9). Empirically, we set $\lambda$, $\mu$, $\gamma$, and $\tau$ to 1.0, 0.001, 0.1, and 0.5, respectively.

**Ablation study.** We performed an ablation study to assess the contribution of each component in the proposed model. For this purpose, we implemented two alternative models and compared their performance with that of the proposed method.

*Effectiveness of the COFInt module:* Because the output of the COFInt module was the input to the GRgCL module, we set the resulting global relationship matrix as an identity matrix. Specifically, we used the low-dimensional embedding ($z$) obtained by concatenating the raw data of the two omics through an autoencoder and analyzing it in conjunction with the identity matrix for subsequent analysis. The results of the proposed method are higher in terms of NMI than the No-COFInt model by 9.87%, 9.94%, and 4.73% on the three datasets, respectively (Table 3). This indicated that the concatenated representation $Z$ contained a large amount of view-specific information, which was not conducive to clustering. Although the No-COFInt model leveraged the complementarity of similar samples to reduce noise and redundancy across different views effectively, our COFInt module enhanced the consistency of data representation within the same cluster further.

*Effectiveness of the GRgCL loss:* The results of the proposed method were higher in terms of NMI than the No-CL model by 9.54%, 14.31%, and 0.84%, respectively (Table 3). Because the multimodal shared feature representation was improved by incorporating the global relationships across all samples. Contrastive learning further optimized this by maximizing the similarity between the shared feature representation and the view-specific representation of the same sample, while simultaneously minimizing the similarity between representations of different samples that exhibit low relationships; this improved clustering performance.

Table 3. Ablation study. The best results are highlighted in **bold font**.

| | noCOFInt | | noGRgCL | | scCODI | |
|---|---|---|---|---|---|---|
| | NMI | ARI | NMI | ARI | NMI | ARI |
| BMNC100866 | 0.9014 | 0.8089 | 0.9299 | 0.9422 | **0.9904** | **0.9950** |
| GSE126074 | 0.8617 | 0.8616 | 0.8288 | 0.8315 | **0.9474** | **0.9781** |
| PBMC_spetor | 0.6856 | 0.4810 | 0.7120 | 0.4667 | **0.7180** | **0.6149** |

## 4. Conclusion

In this paper, we proposed a novel cross-omics deep learning method called scCODI for cluster analysis of single-cell multi-omics data. Our method is capable of handling two types of omics data: single-cell transcriptomic data paired with epigenomic data, and single-cell transcriptomic data paired with proteomic data. We designed a cross-omics integration module that enhances the common representation by learning the global similarities between samples, which brings similar samples closer in space. Additionally, through a contrastive learning module, scCODI aligns the shared features and specific representations of omics data, which ensures that paired samples from different omics have similar specific representations. Our experimental results demonstrated that scCODI outperformed previous methods on diverse single-cell multi-omics datasets.

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



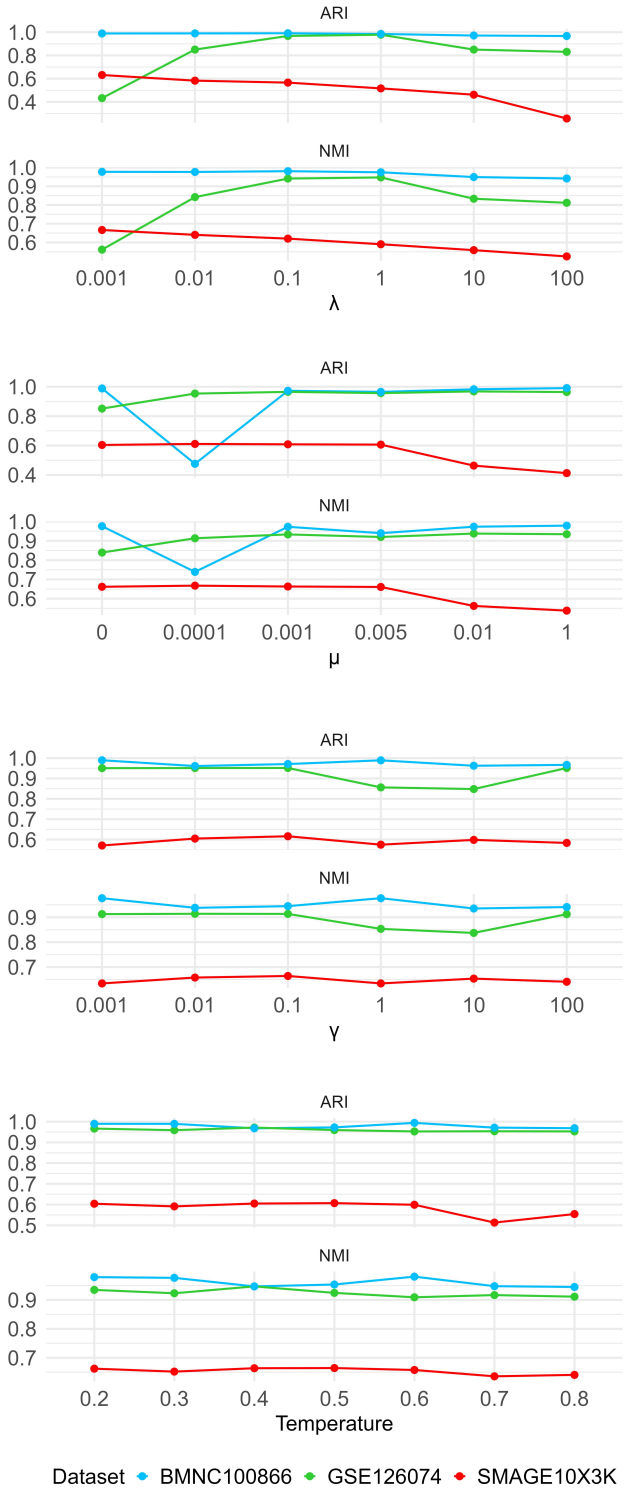



Figure 9. **Hyperparameter tuning of scCODI**. This figure demonstrates the impact of varying hyperparameters ($\lambda$, $\mu$, $\gamma$, and temperature) on clustering performance metrics, including ARI and NMI.

Overall, scCODI improves data clustering performance by combining a global and cross-view feature aggregation module with a multi-view contrastive learning module. Our method not only

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



fully leverages the complementarity of different omics data, but also enhances the consistency of data representation through contrastive learning, which results in more accurate clustering. These results indicate that scCODI has great potential and broad application prospects in handling complex single-cell multi-omics data.

## 5. Code and Data availability

The code is available at https://github.com/SXUI-BDSI/scCODI All data underlying this article are available in Gene Expression Omnibus, at https://www.ncbi.nlm.nih.gov/geo/

## 6. Author Contributions

**Conceptualization**: Yingjie Guo,Yuhua Qian.

**Formal analysis**: Yingjie Guo, Zhen Liang.

**Funding acquisition**: Yingjie Guo, Zhen Liang.

**Investigation**: Chenyang Cheng,Yingjie Guo, Zhen Liang

**Methodology**: Yingjie Guo, Chenyang Cheng, Zhen Liang.

**Project administration**: Yingjie Guo, Zhen Liang,Yuhua Qian.

**Resources**: Yuhua Qian.

**Software**: Chenyang Cheng, Zheqing Zhu, Jin Zhou.

**Supervision**: Yingjie Guo, Zhen Liang,Yuhua Qian.

**Visualization**: Chenyang Cheng, Zheqing Zhu, Jin Zhou.

**Writing – original draft**: Chenyang Cheng, Yingjie Guo.

**Writing – review & editing**: Chenyang Cheng, Haiping Lv, Yingjie Guo, Zhen Liang, Yuhua Qian.

## 7. Funding

This research was supported by the Science and Technology Innovation Young Talent Team of Shanxi Province to ZL ( NO. 202204051001019), the National Natural Science Foundation of China to ZL (NO. 32170410), the Natural Science Foundation of Shanxi Province for the Excellent Youth to ZL (NO. 202203021224002), Fundamental Research Program of Shanxi Province to YG (NO. 202403021221020), Shanxi Scholarship Council of China to YG (NO. 2023-006). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## References

[1] Alara Baysoy, Zhicheng Bai, Rahul Satija, and Rong Fan. The technological landscape and applications of single-cell multi-omics. *Nature Reviews Molecular Cell Biology*, 24:695–713, 2023.

[2] Song Chen, Brooklynn B. Lake, and Kun Zhang. High-throughput sequencing of the

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



transcriptome and chromatin accessibility in the same cell. *Nature Biotechnology*, 37(12):1452–1457, 2019.

[3] Shuai Ma et al. Chromatin potential identified by shared single-cell profiling of rna and chromatin. *Cell*, 183:1103–1116.e1120, 2020.

[4] Marlon Stoeckius, Shijie Zheng, Britta Houck-Loomis, et al. Cell hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biology*, 19(1):1–12, 2018.

[5] Marlon Stoeckius, Christoph Hafemeister, William Stephenson, et al. Simultaneous epitope and transcriptome measurement in single cells. *Nature Methods*, 14(9):865–868, 2017.

[6] Yuhan Hao, Stephanie Hao, Erica Andersen-Nissen, et al. Integrated analysis of multimodal single-cell data. *Cell*, 184(13):3573–3587, 2021.

[7] J. Gao, Y. Zheng, L. Li, M. Lu, X. Chen, Y. Wang, et al. Integrated transcriptomics and epigenomics reveal chamber-specific and species-specific characteristics of human and mouse hearts. *PLoS Biology*, 19(5):e3001229, 2021.

[8] Yufei Liu, Miaoqiang Yang, Yue Deng, et al. Integrating multi-omics data for gene regulatory network inference and application to breast cancer. *Scientific Reports*, 10(1):1–12, 2020.

[9] Yizheng Wang, Yixiao Zhai, Yijie Ding, and Quan Zou. Sbsm-pro: Support bio-sequence machine for proteins. *Science China Information Sciences*, 67(11):144–159, 2024.

[10] Tim Stuart, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M. Mauck, Yuhan Hao, Marlon Stoeckius, Peter Smibert, and Rahul Satija. Comprehensive integration of single-cell data. *Cell*, 177(7):1888–1902.e21, 2019.

[11] Andrew Butler, Paul Hoffman, Peter Smibert, et al. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology*, 36(5):411–420, 2018.

[12] Vladimir Y. Kiselev, Thomas S. Andrews, and Martin Hemberg. Challenges in unsupervised clustering of single-cell rna-seq data. *Nature Reviews Genetics*, 20(5):273–282, 2019.

[13] Zhe Ji and Hong Ji. Tscan: Pseudo-time reconstruction and evaluation in single-cell rna-seq analysis. *Nucleic Acids Research*, 44(13):e117, 2016.

[14] Vladimir Y. Kiselev, Kristina Kirschner, Marc T. Schaub, et al. Sc3: consensus clustering of single-cell rna-seq data. *Nature Methods*, 14(5):483–486, 2017.

[15] Gökcen Eraslan, Lukas M. Simon, Fabian J. Theis, et al. Single-cell rna-seq denoising using a deep count autoencoder. *Nature Communications*, 10(1):390, 2019.

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

[17] Cheng Li, Xinyu Li, Peng Jiang, et al. A deep learning approach to integrating single-cell rna-seq and atac-seq data. *Nature Communications*, 12(1):1–11, 2021.

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

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



[19] Xiaotao Lin, Tian Tian, Zongliang Wei, et al. Clustering of single-cell multi-omics data with a multimodal deep learning method. *Nature Communications*, 13(1):7705, 2022.

[20] Tian Tian, Jian Wan, Qian Song, and Zhi Wei. Clustering single-cell rna-seq data with a model-based deep learning approach. *Nature Machine Intelligence*, 1:191–198, 2019.

[21] Davide Risso, Franck Perraudeau, Svetlana Gribkova, Sandrine Dudoit, and Jean-Philippe Vert. A general and flexible method for signal extraction from single-cell rna-seq data. *Nature Communications*, 9:1–17, 2018.

[22] Xinzhu Wang et al. Brem-sc: a bayesian random effects mixture model for joint clustering single cell multi-omics data. *Nucleic Acids Research*, 48:5814–5824, 2020.

[23] Adam Gayoso et al. Joint probabilistic modeling of single-cell multiomic data with totalvi. *Nature Methods*, 18:272–282, 2021.

[24] Fabian A. Wolf, Philipp Angerer, and Fabian J. Theis. Scanpy: large-scale single-cell gene expression data analysis. *Genome Biology*, 19(1):15, 2018.

[25] Romain Lopez, Jeffrey Regier, Michael B. Cole, Michael I. Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nature Methods*, 15:1053–1058, 2018.

[26] Geoffrey E. Hinton and Ruslan R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[27] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th International Conference on Machine Learning*, pages 1096–1103, 2008.

[28] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, 2015.

[29] Lin Xie, Xia Yang, and Yifan Liu. Integrative clustering of multi-omics data using multi-view learning. *Bioinformatics*, 37(15):2073–2080, 2021.

[30] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

[31] Jun Chen, Zhen Xu, Qing Li, et al. A contrastive learning framework for integrating multi-omics data. *Nature Communications*, 12(1):3347, 2021.

[32] Jun Wu, Yang Zhang, Xia Wei, et al. Global relationship-guided contrastive learning for multi-omics integration. *IEEE Transactions on Biomedical Engineering*, 67(9):2456–2466, 2020.

[33] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 1597–1607. PMLR, 2020.

[34] Lijuan Chen, Wei Wang, Yilin Zhai, and Ming Deng. Deep soft k-means clustering with self-training for single-cell rna sequence data. *NAR Genomics and Bioinformatics*, 2(lqaa039), 2020.

[35] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

bioRxiv preprint doi: https://doi.org/10.1101/2025.04.10.648152; this version posted April 17, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



[36] Felix R. Ringeling and Stefan Canzar. Linear-time cluster ensembles of large-scale single-cell rna-seq and multimodal data. *Genome Research*, 31:677–688, 2021.

[37] Stephen Alexander and Ghosh Joydeep. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3:583–617, 2003.

[38] Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of Classification*, 2(1):193–218, 1985.

[39] Eike Becht, Leland McInnes, Jack Healy, et al. Dimensionality reduction for visualizing single-cell data using umap. *Nature Biotechnology*, 37(1):38–44, 2019.

[40] M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biology*, 15:550, 2014.

[41] Johannes C. M. Schlachetzki et al. A monocyte gene expression signature in the early clinical course of parkinson's disease. *Scientific Reports*, 8:1–13, 2018.

[42] D. Amsen et al. Cd8+ t cells: Markers and functionality. *Immunology Today*, 36(6):259–265, 2015.

[43] S. S. Ng, F. De Labastida Rivera, J. Yan, et al. The nk cell granule protein nkg7 regulates cytotoxic granule exocytosis and inflammation. *Nature Immunology*, 21:1205–1218, 2020.

[44] J. S. Orange. Natural killer cells and their role in immunity. *Nature Reviews Immunology*, 13(7):497–507, 2013.

[45] C. Gavin et al. The complement system is essential for the phagocytosis of mesenchymal stromal cells by monocytes. *Frontiers in Immunology*, 10:2249, 2019.

[46] J. Banchereau and R. M. Steinman. Antigen presentation by dendritic cells and monocytes. *Nature Reviews Immunology*, 1:1–10, 1998.

[47] G. Gaud, R. Lesourne, and P. E. Love. Regulatory mechanisms in t cell receptor signalling. *Nat Rev Immunol*, 18:485–497, 2018.

[48] K. Shah, A. Al-Haidari, J. Sun, et al. T cell receptor (tcr) signaling in health and disease. *Sig Transduct Target Ther*, 6:412, 2021.

[49] K. L. Hilligan and F. Ronchese. Antigen presentation by dendritic cells and their instruction of cd4+ t helper cell responses. *Cell Mol Immunol*, 17:587–599, 2020.

[50] C. Jolly, S. Welsch, S. Michor, and Q. J. Sattentau. The regulated secretory pathway in cd4+ t cells contributes to human immunodeficiency virus type-1 cell-to-cell spread at the virological synapse. *PLoS Pathogens*, 7(9):e1002226, 2011.

[51] C. Gutzeit, G. Magri, and A. Cerutti. Intestinal iga production and its role in host-microbe interaction. *Immunol Rev*, 260(1):76–85, 2014.