scGEMOC, A Graph Embedded Contrastive Learning Single-cell Multiomics Clustering Model

1st Bingjun Li
Department of Computer Science and Engineering
University of Connecticut
Storrs, USA
bingjun.li@uconn.edu

2nd Sheida Nabavi
Department of Computer Science and Engineering
University of Connecticut
Storrs, USA
sheida.nabavi@uconn.edu

2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 979-8-3503-3748-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/BIBM58861.2023.10385267

***Abstract*—Recent advancements in single-cell multiomics sequencing create new research opportunities but also pose challenges, particularly in cell clustering. One major challenge is feature fusion. Early fusion models are robust but ignore the unique distributions of omics and cannot handle various omic dimensions. Most current clustering methods use late fusion, employing independent encoders for each omic. However, the extracted omic features belong to different latent spaces, leading to difficulties in aligning omics. Additionally, current cell clustering methods do not incorporate prior biological knowledge, such as interactions within and across omics, which has been shown plays a key role in defining cell types.**

To address these shortcomings, we propose a novel, scalable, end-to-end clustering method, called single-cell graph embedding multiomics cluster (scGEMOC). scGEMOC utilizes prior biological knowledge to represent inter- and intra-omics connections as a heterogeneous graph. It applies graph embedding to aggregate omics interaction data as a pseudo omic and employs contrastive learning for effectively aligning omics in the latent space. We evaluated scGEMOC on three public datasets against five state-of-the-art baseline models. scGEMOC achieves superior clustering performance compared to the baseline models on all datasets. An ablation study confirms the significant contribution of each component and identifies the most impactful one.

***Index Terms*—single-cell, multiomics, cell clustering, contrastive learning, graph embedding, gene regulatory network**

I. Introduction

The rapid development of single-cell sequencing technology has enabled researchers to simultaneously profile multiple omics of biological information, including gene expression, chromatin accessibility, and surface protein [1]–[3]. This multiomics data provides a more comprehensive perspective of cellular processes, such as embryonic development and disease progression [4]. However, integrative analysis of the generated multiomics data presents new challenges, especially in the field of cell clustering. Cell clustering is a crucial step in single-cell analysis. It facilitates the construction of cell subpopulations for any downstream analysis, reveals the trajectory of cell development among samples, and uncovers common biomarkers associated with a specific cell type. Clustering on single-cell multiomics sequencing data introduces new challenges.

Fusion of different omics is a major challenge in multiomics data analysis. There are two types of fusion model, early fusion [5], [6] and late fusion, but most models use late fusion because this approach can better accommodate different omic dimensions and unique distributions of different omics. Current late-fusion methods for single-cell clustering can be categorized into three types based on their feature extractors [7]. The first involves using matrix factorization to decompose the multiomics data, such as MOFA+ (multi-omics factor analysis v2) [8]. The second contains variations of autoencoder-style neural-network-based models, which emerged with the recent advancement of machine learning. For example scMVAE (single-cell multimodal variational autoencoder) uses variational autoencoder as the backbone [9], MoClust uses an autoencoder-like model to integrate different omics [10], and GLUE utilizes autoencoders to encode omics data and knowledge-based guidance graph [11]. The third features models use graph to represent relationship among cells such as WNN (weighted nearest neighbour) and sigDGCNb [12], [13]. However, one disadvantage of these methods is that they don't align the extracted features in the latent space.

One solution is to use adversarial training by using a discriminator to differentiate among latent features from different omics during the training stage [11], [14]. Different omic features can be considered in the same latent space if a discriminator cannot distinguish them. However, the use of discriminators has been found to possess several major drawbacks: I) it is difficult to prioritize different omic features if the features are forced to be indistinguishable from each other; II) since not all clusters can be distinctly separated in every omic, the presence of a incomplete clustering structure within each omic prevents the clear cluster separation after the fusion; III) adversarial training only aligns feature distributions, which could cause a cluster in one omic aligns with a different cluster in another omic and hampers the clustering results [15].

Contrastive learning has recently shown promising results for self-supervised clustering in computer vision [15], [16]. The principle of contrastive learning is to assign positive and negative pairs to different features, and maximizing the similarities of positive pairs while minimizing the similarities of negative ones. Contrastive learning preserves structures and distributions within each omic during omic alignment, which makes it a better measure than adversarial training. However, the application of contrastive learning in single-cell multiomics clustering has not been studied thoroughly [10]. We aim to utilize contrastive learning for more effective omic alignment.

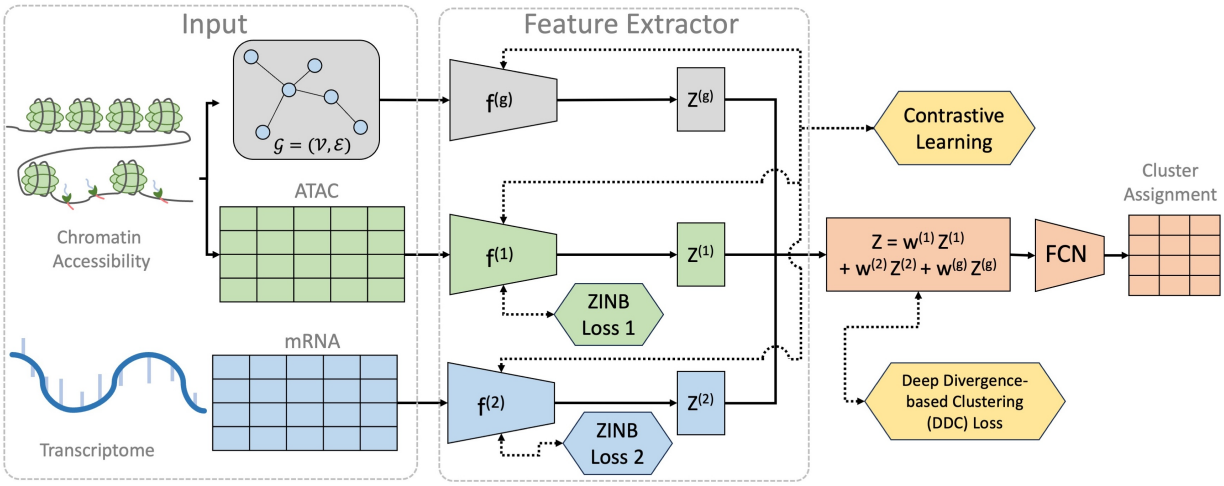


Fig. 1: The overall structure of the model. Trapezoids labeled with $f^{(o)}$ represent feature extraction functions, rectangles represent features, and hexagons represent loss terms. Dashed lines with double arrowheads represent terms regulated by the loss function. $f^{(g)}$ is the variational graph encoder, with its detailed inner structure depicted in Figure 2.

An important insight from the bulk-cell multiomics analysis is that the inter- and intra-omics connections plays a crucial role in defining cell types and utilizing such information improves cell type classification [5], [17]–[19]. To our best knowledge, there is no clustering method that incorporates such biological knowledge [11]. Overall, our goals for single-cell multiomics clustering are: to utilize the prior knowledge, to use a late-fusion model for modeling omic distributions, and to utilize contrastive learning for better omic alignment.

We propose a novel single-cell Graph Embedded Mulitomics clustering model, scGEMOC. Our contributions are:

- The proposed model, scGEMOC, is the first single-cell multiomics clustering model to incorporate prior biological knowledge in the form of omics interaction network for better clustering performance;
- scGEMOC utilizes contrastive learning that preserves omic-specific geometric structure and omic in-variance while avoiding the drawbacks of adversarial training;
- scGEMOC offers a scalable model structure so that it can easily accommodate additional omics with different dimensions as future sequencing technology emerges.

II. Method

As shown in Figure 1, the model uses a distinct encoder for each omic to extract features, along with a variational graph encoder to extract features from the graph representing inter- and intra-omics interactions, which we refer to as the mulitomics network (MON). The features on regular omics are regulated by a zero-inflated negative binomial (ZINB) loss and all features are aligned by a contrastive learning module [20]. A fused feature is derived from a weighted sum with learnable omic weights. Deep divergence-based clustering (DDC) loss is used to regulated the cluster assignment for each cell.

A. Omics and Graph Construction

Assuming the data consists of N samples with O omics data, denoted as $\boldsymbol{X} = [\boldsymbol{X}^{(1)}, ..., \boldsymbol{X}^{(O)}]$, the o-th omic is denoted as $\boldsymbol{X}^{(o)} \in \mathbb{R}^{N \times M^{(o)}}$ for all $o \in \{1, ..., O\}$, where $M^{(o)}$ is the the feature dimension of the o-th omic.

Utilizing prior biological information, we constructed a heterogeneous multi-layer graphs, representing both inter-omics and intra-omic connections, referred to as MON. Prior study in bulk-cell multiomics has shown both connections provides crucial information on cell types [17]. Let's denote the graph as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{X})$, where $\mathcal{V}$ represents a set of vertices indicating omic entities, such as genes or ATAC peaks with their attributes $\boldsymbol{X}$, and $|\mathcal{V}| = \sum_{o=1}^{O} |X^{(o)}|$. Here, $\mathcal{E}$ denotes the set of inter-omic and intra-omic edges. The adjacency matrix $\boldsymbol{A}$ is constructed as follows:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & ... & \boldsymbol{A}_{1O} \\ \vdots & \ddots & \vdots \\ \boldsymbol{A}_{O1} & ... & \boldsymbol{A}_{OO} \end{bmatrix}, \tag{1}$$

where $\boldsymbol{A}_{ii} \in \mathbb{R}^{M^{(i)} \times M^{(i)}}, i \in 1, .., O$ is the intra-omic connections for i-th omic, and $\boldsymbol{A}_{ij} = \boldsymbol{A}_{ji}^{T} \in \mathbb{R}^{M^{(i)} \times M^{(j)}}, i \neq j \in 1, ..., O$ is the inter-omic connections between i-th and j-th omics. When there is no applicable connections for $\boldsymbol{A}_{ii}$ or $\boldsymbol{A}_{ij}$, $\boldsymbol{A}_{ii} = \boldsymbol{I}_{M^i}$ and $\boldsymbol{A}_{ij} = \boldsymbol{0}_{M^{(i)} \times M^{(j)}}$ respectively. All nodes in $\mathcal{G}$ are self-connected.

In our experiments, we used multiomics datasets that include transcriptomics and epigenomics, specifically mRNA and ATAC. The intra-omic edges for mRNA are based on gene-gene interactions sourced from BioGrid [21], while intra-omic edges for ATAC are not applicable. The inter-omic edges between mRNA and ATAC peak are determined based on their proximity. If an ATAC peak falls within the region of a gene or its 1000 bases upstream, we consider them connected.

For mRNA data, we filtered the scRNA-seq data by top variance after log transformation and normalization. For ATAC data, we aggregated scATAC-seq count in the region of a gene or its 1kb-upstream as the promoter activity for that gene.

B. Omic Encoder with ZINB Loss

The latent feature of each regular omic, $Z^{(o)}, o = 1, ..., O$, is extracted by the function, $f^{(o)}(\cdot)$. $f^{(o)}$ consists of a three-layer shallow fully-connected network (FCN). To better model the sparse distribution of transcriptomics (scRNA-seq) and chromatin accessibility (scATAC-seq), a ZINB-based encoder is utilized for each omic [20]. There are three parameters for ZINB distribution, mean (μ), dispersion (θ), and the weight of the point mass (π). Its distribution can be expressed as follows.

$$\text{ZINB}(x; \pi, \mu, \theta) = \pi\delta_0(x) + (1 - \pi)\text{NB}(x; \mu, \theta), \quad (2)$$

where $\text{NB}(x; \mu, \theta) = \frac{\Gamma(x+\theta)}{\Gamma(\theta)} \left(\frac{\theta}{\theta+\mu}\right)^{\theta} \left(\frac{\mu}{\theta+\mu}\right)^{x}$ and $\delta_0(\cdot)$ is a Dirac function. Each parameter is estimated by an individual network. The loss function of the ZINB encoder is:

$$\mathcal{L}_{\text{ZINB}} = -\sum_{o=1}^{O} \gamma_o \log(\text{ZINB}(X^{(o)}|\mu^{(o)}, \pi^{(o)}, \theta^{(o)})), \quad (3)$$

where γ_o is a positive weight for the o-th omic.

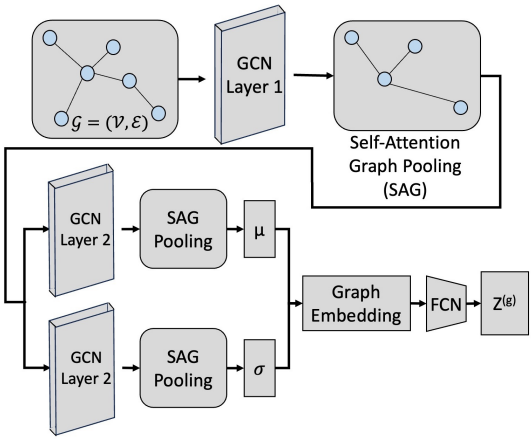


Fig. 2: The structure of graph encoder is plotted. Two GCN layers are used to obtain μ and σ. The graph embedding is obtained in a variational-encoder style. Then the embedding is passed through a transformer layer to obtain $Z^{(g)}$.

C. Graph Encoder

The structure of the graph encoder is shown in Figure 2. We build the graph encoder based on the variational graph autoencoder [22]. The MON pseudo omic feature is obtained by two graph convolution network (GCN) layers combined with corresponding pooling layers defined as follows:

$$\boldsymbol{\mu} = \text{GCN}_{\mu}(\boldsymbol{X}, \boldsymbol{A}), \quad \log \boldsymbol{\sigma} = \text{GCN}_{\sigma}(\boldsymbol{X}, \boldsymbol{A}), \quad (4)$$

where $\text{GCN}(\boldsymbol{X}, \boldsymbol{A}) = \tilde{\boldsymbol{A}}\text{ReLU}(\tilde{\boldsymbol{A}}\boldsymbol{X}\boldsymbol{W}_{\text{GCN}_0})\boldsymbol{W}_{\text{GCN}_1}$, GCN_{μ} and GCN_{σ} shares the same weight $\boldsymbol{W}_{\text{GCN}_0}$, and $\tilde{\boldsymbol{A}} = \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$. The graph embedding is defined as $\boldsymbol{\mu}+\boldsymbol{\sigma}\times\epsilon$, $\epsilon \in (0, 1)$. The graph embedding is then passed through a transform layer comprising a two-layer FCN to obtain the final MON pseudo omic feature , which is denoted as $Z^{(g)}$. The fused feature is obtained by summing all features by weights.

$$Z = \sum_{o=1}^{O} w_o Z^{(o)} + w_g Z^{(g)}, \quad (5)$$

where the weights, w_o and w_g are learnable scalar parameters.

Since all samples share the same graph structure, using only the GCN layer yields indistinct features across clusters. After each GCN layer, we introduce a self-attention graph (SAG) pooling layer. The SAG pooling layer is used to preserve only the most crucial nodes for the sample, thereby enhancing the distinguishability of the graph embedding across all cells [23]. The SAG pooling layer first computes an attention score, $U \in \mathbb{R}^{|\mathcal{V}|}$ for every node defined as the following.

$$U = \sigma(\tilde{\boldsymbol{A}}\boldsymbol{X}\Theta_{att}), \quad (6)$$

where $\Theta_{att} \in \mathbb{R}^{F\times 1}$ is the only parameter of the SAG pooling layer, and F is the node output feature size from the prior GCN layer. Let's denote $q \in (0, 1]$ as the pooling ratio parameter. Nodes are then selected based on U score.

$$\text{idx} = \text{top-rank}(U, \lceil q|\mathcal{V}| \rceil), Z^{(g)}_{mask} = Z^{(g)}_{\text{idx}} \quad (7)$$

where top-rank$(\cdot)$ returns the indices of the top $\lceil q|\mathcal{V}| \rceil$ nodes.

D. Contrastive Learning

Initially, we define the similarity score $s_{ij}^{(u,v)}$ between two latent features $z_i^{(u)}$ (the u-th omic feature of i-th cell) and $z_j^{(v)}$ in Equation (8).This equation defines the cosine similarity, an optimal metric for measuring similarities between feature vectors in the latent feature space.

$$s_{ij}^{(u,v)} = \left(z_i^{(u)}\right)^T z_j^{(v)} / \left[\|z_i^{(u)}\| \cdot \|z_j^{(v)}\|\right] \quad (8)$$

The contrastive loss is based on NT-Xent loss [16] shown as follows.

$$\mathcal{L}_{\text{Contrastive}} = \frac{1}{nO(O-1)} \sum_{i=1}^{N} \sum_{u=1}^{O} \sum_{v=1}^{O} \mathbb{1}_{u\neq v} l_i^{(u,v)}, \quad (9)$$

where $\mathbb{1}$ is an indicator function and

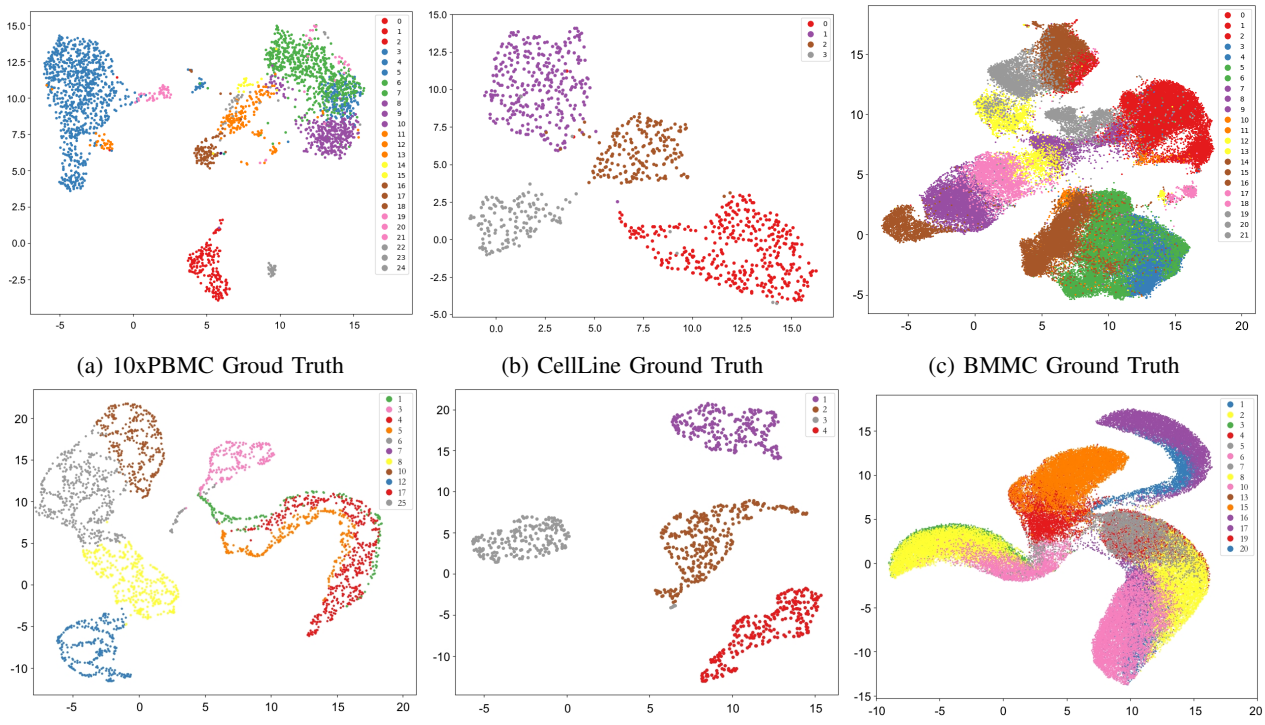
$$l_i^{(u,v)} = -\log \frac{\exp(s_{ii}^{u,v}/\tau)}{\sum_{s'\in \text{Neg}(z_i^u, z_i^v)} \exp(s'/\tau)}, \quad (10)$$

where τ is a hyperparameter set to 0.1 and $\text{Neg}(z_i^u, z_i^v)$ is the set of similarity scores for all the negative pairs of sample i.

An intuitive approach involves assigning positive pairs to the omic features of the same object, and negative pairs to the omic features of different objects. However, this strategy forces the omic features of cells from the same clusters far away too, which deteriorates the downstream clustering performance [15]. In the proposed model, we assign positive pairs to omic features of samples that originate from the same cluster, while assigning negative pairs to omic features of samples that come from different clusters. We construct the negative similarity score set as follows.

$$\begin{aligned} \mathcal{N}_i = \{s_{ij}^{(u,v)} : j \neq i \text{ and } j \in \{1, ..., N\}, \\ u, v \in \{1, ..., O\}, \arg\max \alpha_i \neq \arg\max \alpha_j\}, \end{aligned} \quad (11)$$

where $\alpha_i \in \mathbb{R}^K$ is the clustering assignment vector for i-th cell and K is the number of clusters. And $\text{Neg}_{(z_i^u, z_i^v)}$ is obtained by sampling a constant number of similarity scores from $\mathcal{N}_i$.



(a) 10xPBMC Groud Truth (b) CellLine Ground Truth (c) BMMC Ground Truth

(d) 10xPBMC Prediction (e) CellLine Prediction (f) BMMC Prediction

Fig. 3: UMAP visualizations for three datasets colored by the ground truth cell types is plotted in the upper row and UMAP visualizations colored by the predicted cell types is plotted in the lower row.

E. Clustering

As mentioned in section II-C, we obtained fused features via aggregating the individual omic features by scalar weights. Then, these fused features are passed through an FCN to obtain h_i and another FCN to obtain the soft cluster assignment vector α_i with a softmax activation function. We use the DDC loss to regulate the clustering process, which shows promising results in single modality clustering [24].

The DDC loss comprises three major components: cluster separability and compactness, cluster assignment separability, and closeness of cluster assignment vectors to simplex corners. The first term $\mathcal{L}_1$ is a generalization form derived from Cauchy-Shwarz divergence to achieve separation between clusters in the latent feature space.

$$\mathcal{L}_1 = \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \frac{\binom{k}{2}^{-1} g_{ij}}{\sqrt{g_{ii} g_{ij}}}, \tag{12}$$

where K is the number of clusters and $g_{ij} = \sum_{a=1}^{N} \sum_{b=1}^{N} \alpha_{ai} \tilde{h}_{ab} \alpha_{bj}$, In this context, $\tilde{h}_{ij} = \exp(-\|h_i - h_j\|^2/(2\sigma^2))$, h_i is the i-th omic latent feature, σ is a hyperparameter, and α_{ai} is the i-th element on the cluster assignment vector for sample a.

We use the second term $\mathcal{L}_2$ to separate cluster assignment vectors α_i by encouraging them to be orthogonal.

$$\mathcal{L}_2 = \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \alpha_i^T \alpha_j \tag{13}$$

The third term $\mathcal{L}_3$ is to enforce cluster assignment vectors to a standard simplex.

$$\mathcal{L}_3 = \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \frac{\binom{K}{2}^{-1} u_{ij}}{\sqrt{u_{ii} u_{jj}}}, \tag{14}$$

where $u_{ij} = \sum_{a=1}^{N} \sum_{b=1}^{N} m_{ai} \tilde{h}_{ab} m_{bj}$, in which $m_{ij} = \exp(-\|\alpha_i - e_j\|^2)$, and e_j is the j-th corner of the standard simplex in $\mathbb{R}^K$.

The overall loss function is $\mathcal{L} = \mathcal{L}_{\text{ZINB}} + \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3 + \delta \min(w_1, .., w_O, w_g) \mathcal{L}_{\text{Contrastive}}$, where δ is a hyperparameter for the strength of contrastive learning loss, and $w_1, ..., w_O, w_g$ are regular omics and MON pseudo omic's fusion weights.

III. Experiment

We tested the proposed model against five baseline models that cover all three categories of feature extractors: MoClust, GLUE, MOFA+, EarlyFusion+Leiden, and sigDGCNb.

- **MoClust:** an autoencoder-based clustering model with contrastive learning for alignment but no MON [10].
- **GLUE:** only integrative model for single-cell multiomics data that incorporates similar knowledge graph with regular omic alignment through adversarial training. It is not clustering-focused so KMeans is used for clustering [11].
- **MOFA+:** a widely adopted statistical clustering model that utilizes matrix factorization [8].
- **EarlyFusion+Leiden:** a conventional approach for early-fusion clustering. Both omics are filtered by variance.
- **sigDGCNb:** a network-based clustering method for single omic data that constructs a cell-cell similarity graph [13].

TABLE I: Detail Description of Three Datasets

Datasets	mRNA		ATAC		# of Samples	# of Clusters	Average Number of Samples per Clusters
	Input Dimension	Sparsity	Input Dimension	Sparsity (After Agg.)			
10xPBMC	36,601	94.8%	98,318	84.7%	2,711	25	108.4 ± 194.6
CellLine	18,666	91.0%	136,771	60.4%	1,047	4	261.8 ± 108.7
BMMC	13,431	97.5%	116,490	97.9%	69,249	22	3147.7 ± 3203.9

TABLE II: Clustering Results of the Proposed Model and Baseline Models on Three Datasets

Model	10xPBMC			CellLine			BMMC		
	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
scGEMOC	**0.552**	**0.620**	**0.416**	**0.893**	0.719	**0.771**	**0.574**	**0.635**	**0.466**
MoClust	0.501	0.613	0.339	0.721	0.634	0.556	0.557	0.603	0.420
GLUE+KMeans	0.426	0.510	0.373	0.571	0.462	0.397	0.336	0.319	0.275
MOFA+	0.461	0.473	0.394	0.708	0.691	0.742	0.481	0.568	0.410
EarlyFusion+Leiden	0.351	0.508	0.229	0.336	0.447	0.226	0.444	0.597	0.321
sigDGCNb	0.504	0.457	0.363	0.848	**0.760**	0.764	0.469	0.542	0.412

We implemented scGEMOC in Python. The source code is available at https://github.com/NabaviLab/scgemoc.

A. Datasets and Evaluation Criteria

We conducted experiments on three public single-cell multiomics mRNA and ATAC datasets, 10xPBMC, CellLine, and BMMC dataset [25]–[27]. The details of the datasets are presented in Table I. The cell types for the CellLine and BMMC datasets are annotated by experts, and the cell types for the 10xPBMC dataset are annotated by WNN [12]. Both the mRNA omic and the aggregated ATAC omic show clear sparsity. The UMAP visualizations for cell type groud truth of all datasets are shown in Figure 3a, 3b and 3c. Table I reveals evident cluster imbalances within the 10xPBMC and BMMC datasets, which may cause prediction issues for smaller clusters and potentially lead to a suboptimal overall performance. We selected three criteria to evaluate the models' performance: clustering accuracy (ACC), normalized mutual information (NMI), and adjusted rand index (ARI).

B. Results and Discussions

Table II illustrates that the proposed model outperforms baselines in all metrics across datasets, except for NMI in the CellLine dataset. This superior performance attests to the benefits of incorporating the MON information and the effectiveness of contrastive learning in omic alignment.

When compared to the other contrastive learning method, MoClust, scGEMOC outperforms MoClust in every criterion across datasets, which shows the additional information from the MON pseudo omic helps scGEMOC to better cluster cells. Among five baseline models, MoClust, as the only one with contrastive learning, achieves third best performance at the 10xPBMC and CellLine datasets, and second best performance at the BMMC dataset. The proposed model, scGEMOC and baseline model, MoClust outperform others, showing that contrastive learning enhances clustering in single-cell multiomics data. Moreover, their superior performance over MOFA+ demonstrates that neural-network-based models are preferable for clustering on single-cell multiomics data.

The proposed model, scGEMOC displays evident improvement in performance over sigDGCNb. This underscores the benefit of utilizing multiomics sequencing data for cell type clustering. However, sigDGCNb achieves second best performance at two datasets, which indicates that the relationship among cells also contribute significantly to the results.

GLUE and EarlyFusion+Leiden are the only two non end-to-end clustering models among baselines. EarlyFusion+Leiden achieves the worst performance twice out of all three datasets, which shows early fusion and conventional models are not suitable for high-dimensional complex single-cell multiomics data. Despite GLUE's inclusion of the additional knowledge graph information, it yields the second-to-worst performance on all datasets, which highlights the superiority of dedicated end-to-end clustering methods.

As mentioned earlier, there is an imbalance problem in the 10xPBMC and BMMC datasets. The UMAP visualizations, with cell types predicted by scGEMOC, are shown in Figures 3d, 3e and 3f. These figures demonstrate scGEMOC captures 11 clusters in the 10xPMBC dataset and 15 in the BMMC dataset, capturing both large and some small clusters despite the imbalance. However, scGEMOC fails to capture the extremely small clusters. Enhancing the model's robustness to such imbalances and its ability to identify these tiny clusters is one of our future research goals.

C. Ablation Study

To evaluate the impact of scGEMOC's key components, we conducted an ablation study. The full model was compared to seven variations, each omitting different combinations of the three key elements: MON pseudo omic, contrastive learning loss, and ZINB loss. The results are shown in Table III.

Among three cases with one component off, the one without ZINB loss exhibits the poorest performance. On the other hand, the one-off combination without the MON pseudo omic shows the best performance across all datasets. This finding suggests that ZINB module has the biggest performance impact while only omitting one key component. All three cases shows performance deterioration compared to the full model.

Among three cases with two components off, they all perform relatively similar across all datasets, except for the combination without MON pseudo omic and contrastive learning on the CellLine dataset. We believe the reasons are two fold. First, the omission of MON pseudo omic compensates the omission of contrastive learning module. Since the DDC loss regulates the learned feature to be compact, the result does

TABLE III: Testing Effectiveness of the Three Components in the Proposed Model on 10xPBMC Dataset

Components			10xPBMC			CellLine			BMMC		
Graph	Contrastive	ZINB	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
✓	✓	✓	0.552	0.620	0.416	0.893	0.719	0.771	0.574	0.635	0.466
×	✓	✓	0.501	0.613	0.339	0.721	0.580	0.481	0.557	0.603	0.420
✓	×	✓	0.438	0.588	0.365	0.697	0.524	0.522	0.498	0.514	0.398
✓	✓	×	0.408	0.600	0.324	0.662	0.535	0.520	0.476	0.491	0.401
×	×	✓	0.383	0.555	0.299	0.798	0.628	0.588	0.417	0.440	0.354
×	✓	×	0.391	0.587	0.309	0.654	0.469	0.452	0.419	0.451	0.372
✓	×	×	0.388	0.596	0.284	0.622	0.498	0.458	0.413	0.447	0.359
×	×	×	0.363	0.562	0.268	0.619	0.452	0.387	0.398	0.411	0.405

not show clear deterioration. Second, the CellLine dataset is a relatively simple one, where simpler model can easily achieve good result. Except for this combination on CellLine data, all the other two-off combinations show a clear decrease in performance compared to those with one off. The combination with all three components off shows the worst performance among all variations. Thus, we can conclude that each components contributes differently to the model performance and ZINB loss has the greatest single-component impact.

IV. Conclusion

In this study, we introduced scGEMOC, a novel and scalable single-cell multiomics clustering method that combines the strength of multiomics interaction information and contrastive learning. To the best of our knowledge, scGEMOC is the first single-cell multiomics clustering model that utilizes MON. scGEMOC considers MON as pseudo omic, and extracts MON features through variational graph encoder and other omics features through individual encoders. It employs both contrastive learning and clustering loss to align and regulate the latent feature space for effective fusion. Additionally, scGEMOC utilizes ZINB loss to accommodate the sparsity characteristics of single-cell multiomics data.

We evaluated scGEMOC against five baseline models across three datasets. scGEMOC consistently outperforms the baseline models in terms of clustering accuracy, ARI, and NMI on all datasets. Our ablation study reveals that the three important components of the proposed model (MON, contrastive loss, and ZINB) enhance performance. Omitting ZINB loss notably impacts the model performance, while excluding the contrastive learning module could compensate the lack of the MON information on simpler datasets in terms of performance.

While our experiments focused on two-omic datasets, scGEMOC's scalable design can extend to additional omics [28]. scGEMOC has paved a path to integrate inter- and intra-omics connections into multiomics fusion. With the advancement of spatial multiomics sequencing, we aim to extend our work to include spatial information and explore cell-cell relationship as our future work.

References

[1] S. Ma and et al., "Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin," *Cell*, vol. 183, 2020.

[2] H. Chen and et al., "Assessment of computational methods for the analysis of single-cell ATAC-seq data," *Genome Biology*, vol. 20, 2019.

[3] M. Stoeckius and et al., "Simultaneous epitope and transcriptome measurement in single cells," *Nature methods*, vol. 14, 2017.

[4] I. C. Macaulay, C. P. Ponting, and T. Voet, "Single-cell multiomics: multiple measurements from single cells," *Trends in genetics*, 2017.

[5] B. Li, T. Wang, and S. Nabavi, "Cancer molecular subtype classification by graph convolutional networks on multi-omics data," in *Proceedings of the 12th ACM BCB*, 2021, pp. 1–9.

[6] X. Lin and et al., "Clustering of single-cell multi-omics data with a multimodal deep learning method," *Nature Communications*, 2022.

[7] S. Stanojevic, Y. Li, A. Ristivojevic, and L. X. Garmire, "Computational methods for single-cell multi-omics integration and alignment," *Genomics, Proteomics & Bioinformatics*, 2022.

[8] R. Argelaguet, D. Arnol, D. Bredikhin, Y. Deloro, B. Velten, J. C. Marioni, and O. Stegle, "Mofa+: a statistical framework for comprehensive integration of multi-modal single-cell data," *Genome biology*, 2020.

[9] C. Zuo and L. Chen, "Deep-joint-learning analysis model of single cell transcriptome and open chromatin accessibility data," *Briefings in Bioinformatics*, vol. 22, no. 4, p. bbaa287, 2021.

[10] M. Yuan, L. Chen, and M. Deng, "Clustering single-cell multi-omics data with MoClust," *Bioinformatics*, vol. 39, no. 1, Jan. 2023.

[11] Z.-J. Cao and G. Gao, "Multi-omics single-cell data integration and regulatory inference with graph-linked embedding," *Nature Biotechnology*, May 2022.

[12] Y. Hao and et al., "Integrated analysis of multimodal single-cell data," *Cell*, vol. 184, no. 13, pp. 3573–3587.e29, Jun. 2021.

[13] T. Wang, B. Li, and S. Nabavi, "Single-cell rna sequencing data clustering using graph convolutional networks," in *IEEE BIBM*, 2021.

[14] Q. Liu, S. Chen, R. Jiang, and W. H. Wong, "Simultaneous deep generative modeling and clustering of single cell genomic data," 2020.

[15] D. J. Trosten and et al., "Reconsidering Representation Alignment for Multi-view Clustering," in *IEEE/CVF CVPR*, 2021, pp. 1255–1265.

[16] T. Chen and et al., "A simple framework for contrastive learning of visual representations," in *ICML*. PMLR, 2020, pp. 1597–1607.

[17] B. Li and S. Nabavi, "A multimodal graph neural network framework for cancer molecular subtype classification," *arXiv preprint arXiv:2302.12838*, 2023.

[18] C. Gupta and et al., "Single-cell network biology characterizes cell type gene regulation for drug repurposing and phenotype prediction in alzheimer's disease," *PLoS Computational Biology*, vol. 18, 2022.

[19] S. Zhang and et al., "Inference of cell type-specific gene regulatory networks on cell lineages from single cell omic datasets," *Nature Communications*, vol. 14, 2023.

[20] G. Eraslan and et al., "Single-cell rna-seq denoising using a deep count autoencoder," *Nature communications*, vol. 10, 2019.

[21] R. Oughtred and et al., "The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions," *Protein Sci*, vol. 30, pp. 187–200, 2021.

[22] T. N. Kipf and M. Welling, "Variational graph auto-encoders," *arXiv preprint arXiv:1611.07308*, 2016.

[23] J. Lee, I. Lee, and J. Kang, "Self-Attention Graph Pooling." arXiv, Jun. 2019, arXiv:1904.08082 [cs, stat].

[24] M. Kampffmeyer and et al., "Deep Divergence-Based Approach to Clustering," *Neural Networks*, vol. 113, pp. 91–101, 2019.

[25] X. Genomics, "Pbmc from a healthy donor, single cell multiome atac gene expression demonstration data by cell ranger arc 2.0.0."

[26] S. Chen and et al., "High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell," *Nat. Biotechnol*, 2019.

[27] M. D. Luecken and et al., "A sandbox for prediction and integration of dna, rna, and proteins in single cells," in *NeurIPS 2021*.

[28] K. Vandereyken and et al., "Methods and applications for single-cell and spatial multi-omics," 2023.