OXFORD

Briefings in Bioinformatics, 2026, **27(3)**, bbaf169

https://doi.org/10.1093/bib/bbaf169

Problem Solving Protocol

scMVAF: a multi-view adaptive fusion clustering approach for single-cell RNA-sequencing data

Jinfeng Wang[1], Qixiong Long[1,*], Deyu Tang[1], Jin Deng[1], Yong Liang[2]

[1]College of Mathematics and Informatics, South China Agricultural University, No. 483 Wushan Road, 510642 Guangzhou, China
[2]Peng Cheng Laboratory, No. 6001 Shahe West Road, 518055 Shenzhen, China

*Corresponding author. College of Mathematics and Informatics, South China Agricultural University, No. 483 Wushan Road, 510642 Guangzhou, China. E-mail: 2272734687@qq.com

Abstract

Single-cell RNA-sequencing (scRNA-seq) can excavate cellular heterogeneity and distinguish different types of cells. Clustering cells into subpopulations is essential in analyzing scRNA-seq data as it can help subsequent downstream analysis. However, scRNA-seq data are high-dimensional, sparse, and contain erroneous zero counts, which poses a great challenge for clustering. Although various methods have emerged in recent years, they cannot fully grasp the information of cells by characterizing scRNA-seq data from a single perspective, resulting in poor learned embedding representation and poor clustering performance. In this paper, we propose a multi-view clustering framework scMVAF for scRNA-seq data, which can learn more discriminative embedding representations by integrating feature information from multiple cell views. First, to comprehensively capture the data information, we generate multiple diverse views by down-sampling features, and then scMVAF learns a strong embedding representation for each cell view using an autoencoder based on a denoising zero-inflated negative binomial model. Next, to explore the correlation between cells in different views, a multi-view fusion module is introduced to fuse the embeddings from different views into a unified feature space. Concurrently, the fused embeddings are clustered to generate pseudo labels to improve the embedding process, and finally updating the embedding features and pseudo-labels in turn to obtain better clustering performance. Experiments are implemented on 16 real datasets and verify that scMVAF is superior to the other eight advanced technologies. Our code script can be obtained at https://github.com/LQXLE/scMVAF/

Keywords: scRNA-seq data; multi-view clustering; autoencoder; fusion

Introduction

Live creatures are composed mostly of cells, which are also responsible for ensuring that live organisms develop and operate properly [1]. Traditional RNA sequencing (Bulk RNA-seq) sequences all cells in a sample simultaneously at the tissue level. However, it fails to reveal cellular heterogeneity. In recent years, the development of single-cell sequencing (scRNA-seq) has enabled us to analyze tens of thousands of cells at single-cell resolution. It is an important tool for studying cellular heterogeneity [2] and plays an important role in studying complex diseases [3], exploring biological development processes [4], and assisting drug development [5]. In the early days, single-cell sequencing technology mainly focused on low-throughput sequencing methods. With the maturity of next-generation sequencing technology, the emergence of high-throughput sequencing platforms has greatly increased data output [6]. Due to the limited transcript coverage of each cell during the sequencing process, the expression values of a large number of genes were zero. Additionally, polymerase chain reaction amplification bias and sequencing errors can also interfere with the detection of true biological signals [7]. Compared with bulk RNA-seq data, scRNA-seq data are characterized by high noise [8] and sparseness. These characteristics impact downstream analysis and make cell grouping extremely difficult. Therefore, it becomes extremely important to develop a clustering model that can cope with these challenges.

In recent years, clustering has received widespread attention in the analysis of single-cell RNA sequencing (scRNA-seq) data. As a key analytical tool, clustering aims to divide cells into different subpopulations based on their gene expression characteristics, thereby revealing the diversity of cell types and identifying new cell populations. This is of great significance for understanding complex biological systems and disease mechanisms. In order to cluster scRNA-seq data, some classic clustering methods have been proposed. For instance, K-means [9] clustering is a widely used unsupervised machine learning algorithm. CIDR [10] is a typical hierarchical clustering method that incorporates an implicit zeroing method into distance calculation to provide stable cell-to-cell distance estimates. SIMLR [11] is a clustering method that combines spectral clustering with multiple kernel similarity measures. Seurat [12] is a popular community detection-based method that constructs a k-nearest neighbor graph by selecting genes with large variations, and finally uses a community

Received: September 12, 2024. **Revised:** March 13, 2025. **Accepted:** March 23, 2025

detection algorithm to complete clustering. While these methods offer distinct benefits, they also come with some drawbacks. For example, several methods are not suitable for small sample data, which are usually sparser and noisier in single-cell sequencing, and these methods cannot handle the noise in the data well, resulting in poor performance in feature selection and clustering. Furthermore, some of these methods have high complexity and cannot be scaled to large datasets.

There are analysis-relevant factors in scRNA-seq data. A key technical challenge is the occurrence of missing values in the data. In a gene expression matrix, zero values may be due to technical reasons that the gene is not expressed or detected, a situation often referred to as false zero [13]. This implies that gene expression values are unobserved due to factors such as low capture efficiency or inadequate sequencing depth [14]. Deep learning models have attracted widespread attention because they can reduce the impact of technical factors in single-cell data. scGMAI [15] is a Gaussian mixture model that first reconstructs the data using an autoencoder, then reduces the dimensionality, and finally performs clustering using a Gaussian mixture model. DCA [16] applied two different noise models combined with an autoencoder to infer distribution parameters to better characterize scRNA-seq data. scDeepCluster [17] simultaneously learns feature representation and clustering through generative explicit modeling by projecting the original data space into a zero-inflated negative binomial (ZINB) distribution and combining DCA to impute missing values. Contrast-sc [18] extracts features by comparing the similarities between cells. scDCCA [19] not only learns the similarity of samples but also takes into account the characteristics of the data itself by using dual contrastive learning to obtain the pairwise similarity of cells.

Clustering methods based on deep learning perform well in capturing complex data patterns, but a single model has limitations in dealing with data diversity. To overcome this limitation, recent studies have adopted ensemble learning methods to integrate the results of multiple deep learning models to comprehensively mine data. scDFC [20] effectively integrates the attribute and structure information of scRNA-seq data by combining the attribute feature clustering module and the structure-attention feature clustering module. SC3 [21] is an ensemble-based clustering method that first applies PCA and Laplace transform to reduce the dimensionality of genetic data, calculates pairwise distances between cells, and constructs a consensus matrix, which is then used as input for hierarchical clustering. SAFE [22] is another ensemble-based method that integrates the results of four clustering algorithms: SC3, CIDR, Seurat, and t-SNE [23] + K-means, and applies three hypergraph-based segmentation algorithms to fuse the clustering solutions of each method to generate the final clustering results. scMVFI [24] is a multi-view clustering method that integrates multiple perspectives to reveal potential biological patterns that are difficult to capture with a single perspective, helping to discover new cell types. However, it ignores the redundant information between perspectives and the differences in the contributions of perspectives to clustering, which affects clustering accuracy.

Given the constraints of current approaches, this paper proposes a multi-view clustering framework called scMVAF. The proposed method first generates multiple views by downsampling the original data several times and then learns the embedding representations of each view using a ZINB-based denoising autoencoder. This scheme can not only reduce the problem of data sparsity, but also better fill in the missing information and reduce the impact of single-view data noise through comprehensive analysis of multiple views. Next, we design a multi-view feature fusion module which can capture the differences between different views and dynamically assign weights to each view. Finally, the fusion view is used to generate a unified target distribution to guide the optimization of the autoencoder, thereby mining an embedded representation with higher discriminative information to identify cell types. Our model is multiple compared on 16 real datasets, and the experimental results show that the proposed model scMVAF outperforms the other 8 baseline models and shows excellent performance in cell clustering.

Materials and methods

Datasets and data preprocessing

Sixteen real single-cell datasets were collected from public platforms for experiments and compared with other baseline methods. All datasets have a credible real label, and the label information of the cell type is annotated by its author. The details of these datasets are shown in Table 1.

For the original dataset, our preprocessing is as follows. First, datasets with sample sizes >50 were log-transformed (base 2), and then cells in which >95% of the cells had a gene expression value of 0 were removed. We then normalized the data based on library size by calculating the size factor. Specifically, the library size of a specific cell i is denoted by s_i, the library size of all cells is denoted by s, and the size factor for cell i is expressed as the ratio s_i/s_m, where s_m represents the median of s.

Finally, the python package SCANPY [25] was used to screen out 2000 genes with large variations from the original data. By selecting genes with large variations between cells, the impact of technical noise can be effectively reduced, making it possible to better distinguish different cell types because these genes show significant expression differences between different cell types, thereby improving the accuracy and stability of clustering.

The framework of scMVAF

Figure 1 shows the framework diagram of scMVAF. It consists of two key modules. Specifically, the preprocessed expression matrix is first sampled at a fixed sampling rate of 0.9 to generate multiple views. Then multiple views are input into multiple denoising ZINB distributed autoencoders, respectively, and projected into the latent space to obtain multiple independent single-view latent features. Considering that multiple views are generated in different scenarios and there may be diversity between them, simple fusion may ignore the cell consensus information learned from multiple views. Therefore, this paper proposes an adaptive weighted fusion module that can dynamically perceive the importance of each view and perform feature fusion of multiple views, which helps to combine multi-view information for comprehensive prediction of cell types. Finally, the fused views are used to obtain a unified target distribution and each view encoder obtains a soft cluster distribution. The private soft distribution of each view and the unified target distribution of the fused view optimize each other, thereby guiding each view encoder to obtain better discriminative information.

Generate cell views

Due to technical limitations in the scRNA-seq process, there are missing values in the data, also known as dropout events. This blurs the relationship between genes and greatly affects downstream analysis. Recently, ensemble learning has been used for dropout imputation in scRNA-seq [32, 33], which can achieve

Table 1. Dataset information including cell types, source, and reference

Datasets	Cells	Genes	Cell types	Cell source	Reference
Yan	90	20 214	6	Human	[26]
Goolam	124	41 480	5	Mouse	[27]
Deng	268	22 431	6	Mouse	[28]
Camp	777	19 020	7	Mouse	[29]
QS_Limb_Muscle	1090	23 341	6	Mouse	[30]
QS_Trachea	1350	23 341	4	Mouse	[30]
Qx_Bladder	2500	23 341	4	Mouse	[30]
Qx_Limb_Muscle	3909	23 271	6	Mouse	[30]
QS_Heart	4365	23 341	8	Mouse	[30]
QxSpleen	9552	23 341	5	Mouse	[30]
PBMC-Zheng4k	4340	33 694	8	Human	[31]
PBMC-Ding	7111	20 428	10	Human	[31]
PBMC-A	11 432	14 504	8	Human	[31]
PBMC-C	11 989	14 222	8	Human	[31]
PBMC-B	12 261	14 473	8	Human	[31]
AD-brain	13 214	10 850	8	Human	[31]

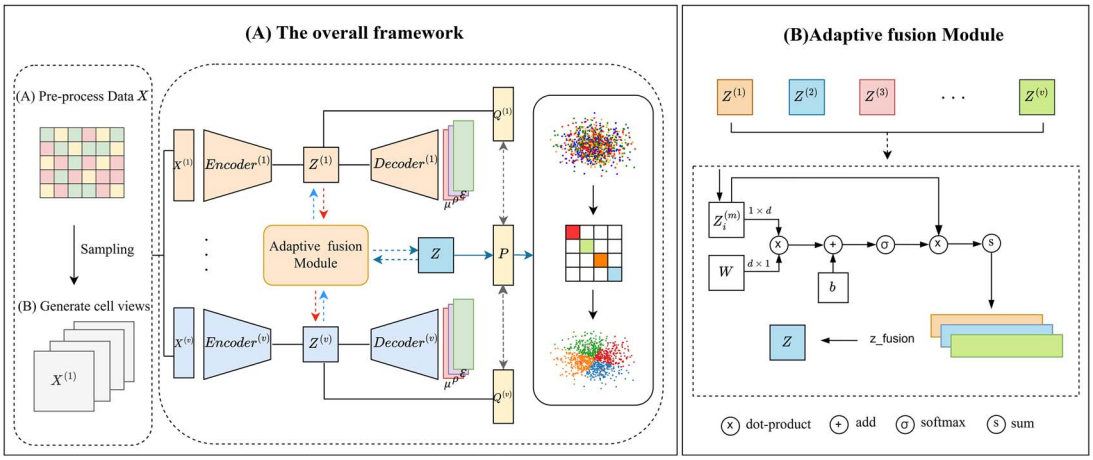

Figure 1. Illustration of scMVAF framework. (A) The overall framework. The preprocessed X is first generated by multiple times down-sampling to generate V views, each of which is independently fed into the ZINB-based model for denoising projection to obtain the embedding features of a single view, followed by adaptive weighted fusion. A uniform target distribution P is obtained by the clustering algorithm on the fused features and $Q^{(v)}$ is the soft clustering distribution for each view. Finally, the uniform objective distribution P is used to guide the view of private soft distribution $Q^{(v)}$ for clustering optimization. (B) Adaptive weighted fusion Module.

better prediction performance by integrating the classification results of multiple weak classifiers [34]. For ensemble clustering, the first task is to generate multiple base cluster partitions. A common approach is to repeatedly sample a certain number of gene features from the preprocessed data [35]. In this study, in order to eliminate possible redundant features and generate a robust view, the original data were repeatedly sampled multiple times using a fixed sampling rate to generate multiple subsets. Multiple random subsampled submatrices can reveal different patterns and variabilities within the data. Training with several distinct views helps mitigate the noise effects of any single view, thereby enhancing the model's generalization ability. Additionally, random subsampling helps prevent overfitting associated with using a single view (the complete data matrix), particularly when dealing with sparse and noisy scRNA-seq data. Specifically, these subsets have the same number of cells as the original data but only contain a subset of genes. The preprocessed expression matrix is represented as X, and scMVAF randomly samples features of X to generate multiple cell views $\{X^{(1)}, X^{(2)}, \ldots, X^{(v)}\}$. Each cell view is obtained by sampling the original expression matrix, and the generated multiple cell views are input to the ZINB-based denoising multi-view autoencoder to learn multiple independent feature embeddings.

Denoising multi-view autoencoder

Autoencoder is a popular neural network model used for feature learning and data compression in unsupervised learning. It consists of two parts: an encoder and a decoder. The basic idea is to compress the input data into low-dimensional representations in the latent space through the encoder and reconstruct these representations into outputs as close to the original input as possible through the decoder. However, scRNA-seq data are highly sparse and contain numerous zero values and noise due to dropout events, potentially leading to features learned by the autoencoder that lack robustness and generalizability. To mitigate the impact of noise and dropout, this paper adopts a ZINB-based denoising autoencoder to map data from different cell views into multiple embedding spaces and perform feature fusion. Denoising autoencoders can recover undamaged data from noisy input data, which helps to alleviate the impact of noise [36]. Compared with simple reconstruction methods, the ZINB distribution model can

provide more accurate estimates of gene expression levels by considering the zero-inflated nature of the data, thereby improving the performance of downstream analysis. Specifically the vth cell view data are $X^v \in R^{(n\times m)}$, each cell view data are first corrupted by random Gaussian noise e to obtain X_v^{corrupt}:

$$X_v^{\text{corrupt}} = X^v + e, \tag{1}$$

the noise e can be passed to any layer of the encoder.

We define W as the learning parameter of the encoder, and the encoder function is defined as:

$$z^{(v)} = f_W(X_v^{\text{corrupt}}), \tag{2}$$

where $f_w(\cdot)$ denotes the encoder function parameterized by W, z(v) represents the potential representation obtained after being processed by the encoder function $f_w(\cdot)$.

Subsequently, the decoder reconstructs the hidden mapping, W' is the learning parameter of the decoder, and the decoder function is $g_{W'}(\cdot)$:

$$X_v' = g_{W'}(z^{(v)}), \tag{3}$$

where X_v' is the result obtained by decoder.

The encoder and decoder consist of fully connected neural networks using rectified linear units as their activation function, and the minimization loss function L during the learning process of the denoising autoencoder is expressed as:

$$L(X^v, g_{W'}(f_w(X_v^{\text{corrupt}}))). \tag{4}$$

Denoising Autoencoder Based on ZINB Distribution consists of an encoder and three decoders. The encoder compresses high-dimensional gene expression data into a low-dimensional embedding space. The three decoders are used to predict the mean (μ), discrete value (θ), and zero-inflated probability (ε), respectively. The ZINB denoising model uses μ as the interpolated gene expression value. For zero values in the data, the model will determine the source of the zero value based on the ε value. If it is introduced by technical noise, the μ value will be used for interpolation. The θ value captures the overdispersion of scRNA-seq data by adjusting the variance of the negative binomial distribution, and combines it with the zero-inflation part to deal with excess zero values, so that the ZINB model can more accurately reflect the true characteristics of single-cell gene expression data. It is expressed as follows:

$$\text{NB}(X^{(v)}|\mu,\theta) = \frac{\Gamma(X^{(v)}+\theta)}{X^{(v)}!\,\Gamma(\theta)}\left(\frac{\theta}{\theta+\mu}\right)^{\theta}\left(\frac{\mu}{\theta+\mu}\right)^{X^{(v)}}, \tag{5}$$

$$\text{ZINB}(X^{(v)}|\varepsilon,\mu,\theta) = \varepsilon\delta(X^{(v)}) + (1-\varepsilon)\ \text{NB}(X^{(v)}|\mu,\theta). \tag{6}$$

During the reconstruction process, the ZINB model estimates the parameters μ, θ, and ε. The estimation process can be defined as follows:

$$D = g_{W'}(f_w(X_v^{\text{corrupt}})), \tag{7}$$

$$M = \text{diag}(s_i^{(v)}) \times \exp(W_\mu D), \tag{8}$$

$$\Theta = \exp(W_\theta D), \tag{9}$$

$$E = \text{sigmoid}(W_\varepsilon D), \tag{10}$$

where D stands for the last layer of the decoder, three independent fully connected layers attached after D. M, Θ, E the probability estimation matrices for the mean, discrete, and lost probabilities, respectively. W_μ, W_θ, and W_ε are learnable parameter matrices of μ, θ, and ε, respectively. The loss of the ZINB-based autoencoder uses a negative log-likelihood, and the loss function L_{zinb} for pretraining on a single-cell view and the minimized loss function L_{pre} during learning are defined as follows:

$$L_{\text{zinb}} = -\log(\text{ZINB}(X^{(v)}|\varepsilon,\mu,\theta)), \tag{11}$$

$$L_{\text{pre}} = \min(L_{\text{zinb}}). \tag{12}$$

Adaptive weighted fusion module

In multi-view learning, each view contain different information, and there may be correlations and differences between these information. Equally weighted view fusion ignores the differences between views, and the fusion result is more sensitive to noise and outliers, which reduces the robustness and stability of the model. The attention mechanism has achieved remarkable results in capturing the differences between views, which can dynamically perceive the differences between views and assign weights to them. Therefore, this paper designs a module that automatically perceives the weights of each view. It assigns different weights to the embedded representation of each view based on the improved attention mechanism, and makes full use of the information of different unit views for fusion. Specifically, for the N views in a cell sample, the embedding representation $\{z_1^v, z_2^v, \ldots, z_N^v\}$, $v \in \{1, 2, \ldots, m\}$. The global information is captured by computing the correlation between different views. The core is to calculate the weight matrix ω_m to perceive the importance of different views. Calculate the importance score h_m for each cell view, which can be defined as:

$$h_m = \tanh(\omega_m z_i^m + b_m), \tag{13}$$

where b_m represents the bias term.

Then, the importance score of each view is normalized and converted into view weight. These weights reflect the relative importance of each view in the current clustering task. It can be defined as:

$$a_i^m = \frac{\exp(h_m)}{\sum_{m'}\exp(h_{m'})}, \tag{14}$$

where a_i^m represents the weight coefficient of the m th view in the ith sample.

Finally, the comprehensive feature of each sample is calculated by weighted summing the features of multiple views. Specifically, for each sample, its features under different views are weighted and summed according to the calculated weights to obtain the final feature vector of the sample. For N samples, the feature vector of each sample is a weighted combination of its multiple view features. The fusion z of a specific sample can be defined as:

$$z = z^{\text{fusion}} = \sum_{v=1}^{V} a_i^v z_i^v, \tag{15}$$

where $z_i^v, i \in \{1, 2, \ldots, N\}$ is the embedding representation learned for the vth view in the ith sample.

Multi-view self-supervised learning module

In multi-view learning, the complementarity of views can improve performance [37], and the additional information provided by different views helps to distinguish cell subpopulations more accurately. Since the information learned in these views is related, each view may contain information that other views cannot capture. The complementary information provided by different views can effectively reduce the noise and bias in the data. However, existing multi-view clustering algorithms for scRNA-seq data do not fully consider the complementarity between views. To address these limitations, this paper introduces a multi-view self-supervised learning module to promote information complementarity between views. It enhances information fusion and consistency of feature representation between views by narrowing the difference between the distribution of individual views and the fused view. Specifically, the centroid μ_j of z is first initialized by the K-means method, and then the soft distribution is obtained using the Student t distribution to generate the target distribution. The soft distribution r_{ij} and the target distribution p_{ij} can be expressed as follows:

$$r_{ij} = \frac{(1 + \|z - \mu_j\|^2)^{-1}}{\sum_j (1 + \|z - \mu_j\|^2)^{-1}}, \tag{16}$$

$$p_{ij} = \frac{(r_{ij}^2 / \sum_i r_{ij})}{\sum_j (r_{ij}^2 / \sum_i r_{ij})}. \tag{17}$$

Each cell view obtains a soft label $Q^v = \{q_1^v, q_2^v, \ldots, q_N^v\}$ through K-means. Define q_{ij}^v as the probability that the ith cell in the vth cell view belongs to the jth cluster, which can be expressed as:

$$q_{ij}^v = \frac{(1 + \|z_i^v - \mu_j^v\|^2)^{-1}}{\sum_j (1 + \|z_i^v - \mu_j^v\|^2)^{-1}}, \tag{18}$$

where μ_j^v denotes the jth cluster center of mass of the vth cell view and z_i^v represents the ith sample in the vth view.

Finally, the relative entropy between the target distribution p obtained from the fused representation and the private distribution Q^v learned for each cell view is used to guide the encoder training, which can be expressed as:

$$L_s = \mathrm{KL}(P\|Q) = \sum_{v=1}^{m} \sum_{i=1}^{n} \sum_{j=1}^{k} p_{ij} \log \frac{p_{ij}}{q_{ij}^v}. \tag{19}$$

In order to better learn robust embedding features, the model jointly optimizes the autoencoder reconstruction loss and Kullback–Leibler (KL) loss to learn more discriminative features to better cluster cells. The overall loss function L_{total} that minimizes the training process is defined as follows:

$$L_{\text{total}} = \min(L_{\text{zinb}} + \beta L_s), \tag{20}$$

where β is the correlation coefficient of KL loss.

Results

Parameter setting

scMVAF is run in pytorch based on Python 3.9. The encoding layer size of each view encoder is (256, 64), the bottleneck layer is 32, and the decoding layer size is (64, 256). The view encoder uses a ZINB-based denoising autoencoder, which is pretrained for 90 epochs using the optimizer AdamW [38] with a pretraining learning rate of 0.001. Experimental verification shows that when the number of view encoders reaches 9 and the sampling rate is 0.9, the fusion effect is the best. scMVAF down-sampled the preprocessed cell count matrix using a fixed sampling rate of 0.9 to generate nine views. During the training phase, the optimizer Adadelta [39] was used, and the learning rate was adjusted to 0.001. The model was trained for 150 epochs. The batch size is 256 in both the pretraining and training stages. All experiments are performed on NVIDIA GeForce GTX 4060Ti (16G).

Clustering performance

This study uses ACC (clustering accuracy) [40], ARI (adjusted random index) [41], and NMI (normalized mutual information) [42] as evaluation indicators. The definitions of these three evaluation metrics are provided in Section A of the Supplementary Material.

To assess the performance of scMVAF, we tested it on 16 real-world scRNA-seq datasets and compared its results with those of eight leading methods, where K-means and Seurat are traditional clustering methods, and scDeepCluster, scGMAI, scDFC, and scDCCA are deep clustering methods, SC3, and SAFE are integrated methods. To ensure fairness, scMVAF and all baseline methods were run five times using default parameters, and the mean and standard deviation of the results were calculated. After multiple experimental verifications, our model shows excellent performance in all three evaluation indicators, as shown in Tables 2–4. Some values are missing in some positions in the table because SAFE and scDFC have too high time complexity or cannot run on some datasets due to the biological complexity of the dataset. In contrast, scMVAF shows excellent performance in almost all tasks. Supplementary Tables S1–S3 list the standard deviations of the ACC, ARI, and NMI scores for all methods across the five runs.

scMVAF consistently maintains the top two results in the comparison with all other baseline methods. In terms of ACC score, scMVAF improves ACC by 21.1% on the QxSpleen dataset compared with the second-best algorithm. In terms of ARI score, our method improves the second-best result by 11%, 25.9%, and 26.9% on the PBMC-Zheng4k, QS_Limb_Muscle, and QS_Heart datasets, respectively. In terms of NMI score, except for Camp, which ranks second, all other datasets rank first in NMI score. For the Camp dataset, which ranks second, scMVAF performs similarly to the other eight methods. The experimental results show that our model has stronger generalization ability than other baseline methods on most datasets. This may be due to the significant advantages of multi-view clustering in enhancing cell heterogeneity identification and improving clustering accuracy. It can better handle the uncertainty in the data and effectively integrate information from different perspectives, thereby reducing the noise and bias of a single perspective. In addition, data from different perspectives can provide different levels of biological information. Multi-view clustering can reveal complex biological relationships that cannot be discovered from a single view.

Multi-view learning promotes cell clustering

Most methods cluster data from a single view, which may not be comprehensive enough to capture valuable information. In contrast, multi-view learning addresses the limitations of single-view learning and enhances the feature expressiveness. In scMVAF, one of the views can be optimized by the guidance of the other views, and each view learns a robust representation of cell clustering. To verify the contribution of scMVAF in multi-view

Table 2. ACC scores for scMVAF and baseline methods across 16 datasets, with the highest values highlighted in bold

Datasets	K-means	scDFC	scGMAI	scDeepCluster	Seurat	scDCCA	SC3	SAFE	Ours
Yan	0.762	0.800	0.773	0.687	0.689	0.769	0.756	0.644	**0.858**
Goolam	0.700	0.758	0.653	0.658	0.710	0.815	0.758	0.629	**0.965**
Deng	0.538	0.705	0.701	0.554	0.649	0.745	0.629	0.761	**0.938**
Camp	0.773	0.804	0.608	0.610	0.660	0.693	0.786	0.726	**0.852**
QS_Limb_Muscle	0.751	0.740	0.532	0.735	0.667	0.621	0.738	0.635	**0.990**
QS_Trachea	0.781	**0.956**	0.464	0.814	0.343	0.625	0.642	0.582	0.955
Qx_Bladder	0.606	0.834	0.625	0.773	0.312	0.639	0.771	0.665	**0.864**
Qx_Limb_Muscle	0.735	0.819	0.628	0.759	0.553	0.762	0.935	0.794	**0.994**
QS_Heart	0.681	0.747	0.383	0.718	0.474	0.672	0.724	0.554	**0.926**
QxSpleen	0.723	–	0.530	0.666	0.412	0.755	0.744	0.145	**0.966**
PBMC-Zheng4k	0.696	0.615	0.418	0.743	0.650	0.739	0.762	0.856	**0.879**
PBMC-Ding	0.553	0.458	0.277	0.430	0.457	0.317	0.361	0.256	**0.690**
PBMC-A	0.669	–	0.491	0.776	0.633	**0.794**	0.521	–	0.785
PBMC-C	0.649	–	0.492	0.699	0.588	0.734	0.499	–	**0.793**
PBMC-B	0.656	–	0.488	0.671	0.574	0.737	0.417	–	**0.759**
AD-brain	0.536	–	0.328	0.567	0.485	0.641	0.53	–	**0.682**

Table 3. ARI scores for scMVAF and baseline methods across 16 datasets, with the highest values highlighted in bold

Datasets	K-means	scDFC	scGMAI	scDeepCluster	Seurat	scDCCA	SC3	SAFE	Ours
Yan	0.734	0.752	0.790	0.613	0.690	0.714	0.708	0.602	**0.874**
Goolam	0.614	0.687	0.593	0.561	0.580	0.747	0.687	0.390	**0.991**
Deng	0.448	0.632	0.620	0.470	0.500	0.743	0.580	0.835	**0.922**
Camp	0.691	0.715	0.461	0.553	0.620	0.640	0.679	0.680	**0.722**
QS_Limb_Muscle	0.700	0.656	0.281	0.654	0.551	0.519	0.669	0.536	**0.976**
QS_Trachea	0.592	0.881	0.149	0.683	0.221	0.426	0.525	0.510	**0.928**
Qx_Bladder	0.515	0.768	0.458	0.758	0.231	0.523	0.757	0.575	**0.816**
Qx_Limb_Muscle	0.698	0.804	0.507	0.746	0.479	0.742	0.926	0.740	**0.989**
QS_Heart	0.609	0.660	0.190	0.617	0.369	0.518	0.668	0.445	**0.937**
QxSpleen	0.496	–	0.192	0.431	0.231	0.614	0.317	0.051	**0.946**
PBMC-Zheng4k	0.610	0.445	0.100	0.634	0.581	0.626	0.643	0.729	**0.839**
PBMC-Ding	0.367	0.302	−0.008	0.262	0.308	0.039	0.111	0.178	**0.532**
PBMC-A	0.543	–	0.261	0.609	0.496	0.672	0.146	–	**0.742**
PBMC-C	0.569	–	0.287	0.556	0.491	0.626	0.118	–	**0.785**
PBMC-B	0.640	–	0.281	0.568	0.471	0.603	0.084	–	**0.783**
AD-brain	0.357	–	0.194	0.392	0.344	**0.560**	0.107	–	0.550

Table 4. NMI scores for scMVAF and baseline methods across 16 datasets, with the highest values highlighted in bold

Datasets	K-means	scDFC	scGMAI	scDeepCluster	Seurat	scDCCA	SC3	SAFE	Ours
Yan	0.827	0.815	0.840	0.740	0.663	0.824	0.831	0.628	**0.881**
Goolam	0.775	0.857	0.749	0.697	0.740	0.761	0.776	0.523	**0.968**
Deng	0.689	0.803	0.784	0.696	0.662	0.787	0.743	0.691	**0.879**
Camp	0.800	**0.842**	0.633	0.742	0.726	0.765	0.802	0.758	0.819
QS_Limb_Muscle	0.766	0.816	0.399	0.806	0.668	0.668	0.760	0.655	**0.958**
QS_Trachea	0.641	0.857	0.297	0.725	0.399	0.457	0.535	0.415	**0.872**
Qx_Bladder	0.634	0.790	0.490	0.806	0.370	0.618	0.704	0.505	**0.814**
Qx_Limb_Muscle	0.787	0.852	0.507	0.846	0.624	0.775	0.909	0.753	**0.976**
QS_Heart	0.766	0.798	0.351	0.746	0.556	0.643	0.712	0.551	**0.892**
QxSpleen	0.604	–	0.230	0.544	0.352	0.568	0.346	0.198	**0.870**
PBMC-Zheng4k	0.715	0.603	0.108	0.744	0.614	0.689	0.660	0.770	**0.826**
PBMC-Ding	0.527	0.464	0.020	0.383	0.466	0.065	0.232	0.366	**0.626**
PBMC-A	0.665	–	0.361	0.676	0.629	0.696	0.320	–	**0.726**
PBMC-C	0.629	–	0.362	0.671	0.571	0.653	0.281	–	**0.732**
PBMC-B	0.706	–	0.364	0.675	0.551	0.644	0.230	–	**0.752**
AD-brain	0.559	–	0.343	0.589	0.477	0.622	0.244	–	**0.649**

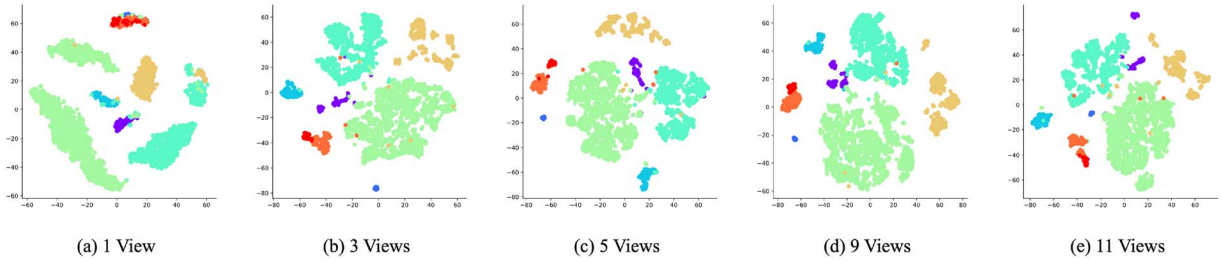

Figure 2. Comparing embedded representations for single-view learning and multi-view learning on the QS_Heart dataset.

cell clustering, we conducted a comparative experiment of single-view and multi-view learning on the QS_Heart dataset. In the experiment, we selected single-view, three-view, five-view, nine-view, and eleven-view settings, and visualized the corresponding embedding representations. The colors in the figures correspond to distinct cell types. As shown in Fig. 2, in the embedded representation obtained using a single view, the boundaries between different clusters are blurred, the overlap between clusters is high, and it is difficult to clearly distinguish different types of cell populations. One possible reason is that single-view learning only uses reconstruction loss to learn representations, which may fail to fully capture the complex correlations and heterogeneity between cells. The multi-view method learns rich features from multiple views and then integrates the complementary information of multiple views. In contrast, the model can effectively reduce the impact of noise and finally separate cells belonging to different clusters to form clear boundaries. Therefore, in the visualization using multi-view clustering, the boundaries between different clusters are clearer and the data points within the cluster are more densely distributed.

Additional comparison with scDeepCluster in real datasets

To further verify the effectiveness of the scMVAF method, we compared scMVAF with the scDeepCluster method combined with Dropout Layer. Dropout Layer, as a regularization technique, helps reduce the overfitting of the model by randomly dropping neurons during training. In this experiment, we selected eight real datasets from different scales to evaluate the performance of the two methods, as shown in Fig. 3 and Supplementary Table S4. The experimental results show that scMVAF outperforms the method combining scDeepCluster with Dropout Layer in multiple evaluation indicators, further verifying the effectiveness of our method. By building multiple sub-models and keeping the model structure clear, scMVAF outperforms the method combining scDeepCluster and Dropout Layer in performance, proving that scMVAF is a superior clustering method.

The impact of hyperparameters

The number of views is a crucial parameter in our scMVAF model, as it determines the number of distinct views for each cell and significantly impacts clustering performance [43]. In this experiment, our method is performed with different numbers of views to analyze its impact on performance. To ensure the reliability of the scMVAF results, we set 5 random seeds 1234, 3333, 2222, 142, 42, and performed 5 independent runs. Then, we calculated the median ARI score under different views as a metric to evaluate its performance. Figure 4 and Supplementary Table S5 shows the ARI results of our method using different numbers of views on the datasets Goolam, QS_Heart, Qx_Bladder, PBMC-Ding, and PBMC-Zheng4k. It can be seen that in most cases, when the

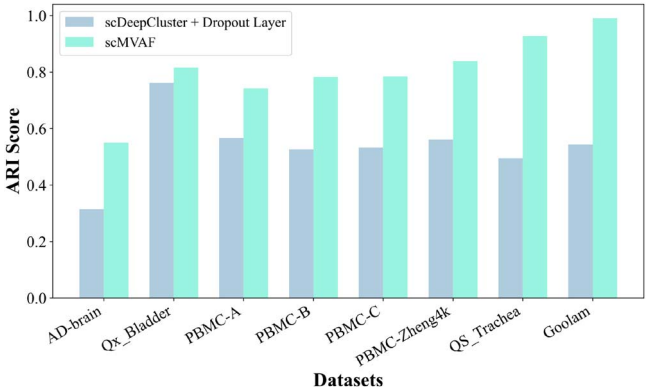

Figure 3. Comparison of ARI scores between scMVAF and the method combining scDeepCluster and Dropout Layer on datasets of different sizes.

number of views is set to 9, the model achieves better clustering performance. Taking the PBMC-Ding dataset as an example, when the number of views is <9, the increase in the number of views improves the performance. The possible reason is that as the number of views increases, cells can be characterized from more different angles and better results can be obtained in the identification of cell types. When the number of views is >9, the clustering performance decreases. Too many cell views may lead to redundant views, which will harm the performance of the model. Therefore, the number of views is set to 9 as our default parameter. Figure 5, Supplementary Tables S6 and S7 show the running time and memory usage for five datasets with different numbers of views.

The down-sampling rate controls the size of each generated unit view. To explore the impact of the down-sampling rate on model performance, our method is run with different down-sampling rates in this experiment. Figure 6 and Supplementary Table S8 present the ARI results. It can be seen that when the down-sampling rate is 0.9, our model takes satisfactory performance. In most cases, when the down-sampling rate is <0.9, the clustering performance of the model improves as the down-sampling rate increases. The possible reason is that fewer features retained lead to loss of information. As the number of features increases, the clustering performance of the model improves. When down-sampling is not performed, the model performance decreases. The possible reason is that there is redundant information in the complete data. Therefore, the down-sampling rate is set to 0.9 as the default parameter.

In the scMVAF model, we design a hyperparameter β to balance the reconstruction loss and the KL loss. β was adjusted on different datasets to explore its impact on model performance. We set the value of β from 0.2 to 1 with a step size of 0.2. When β = 0.2, the model relies more on the reconstruction loss, and when β = 1,

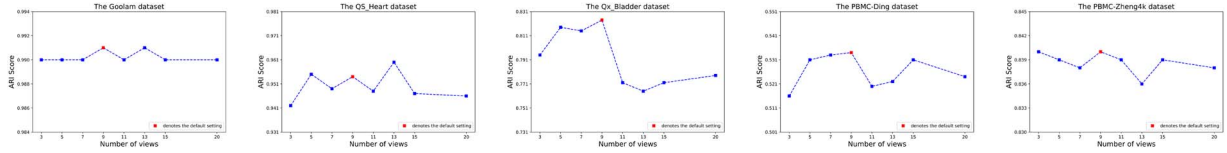

Figure 4. The effect of the number of views in scMVAF.

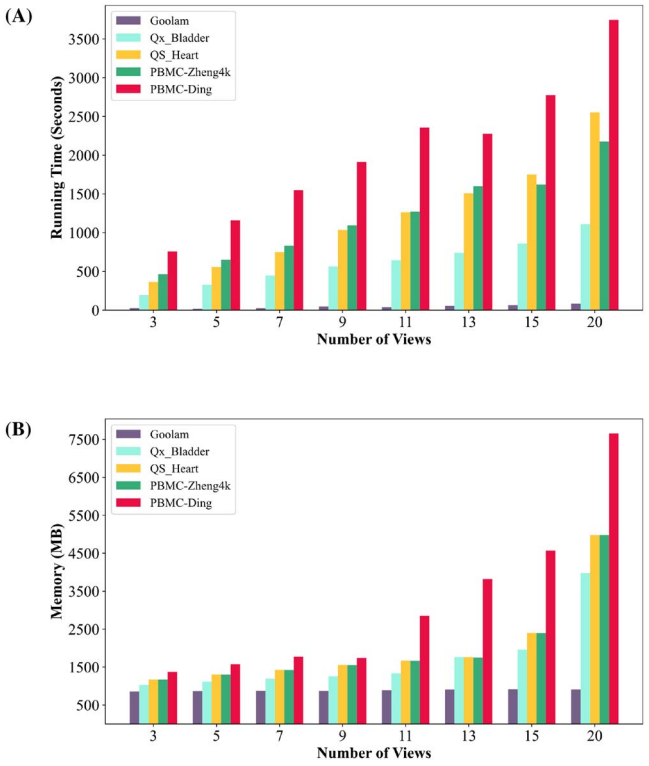

Figure 5. Running time and memory consumption of scMVAF in different views. (A) Running time; (B) memory consumption.

the reconstruction loss and KL loss are equally important. As shown in Fig. 7 and Supplementary Table S9, there are significant differences in the sensitivity of different datasets to β. Specifically, on the PBMC-Ding dataset, as the β value increases, the model performance gradually improves, while in the other four datasets, the model performance remains stable under different β values. Overall, when β is close to 1, the model performs relatively well on the five datasets. This shows that balancing the reconstruction loss and the KL divergence loss to make full use of the information of different cell views is crucial to improving model performance.

Impact of dropout rate on methods

We selected three deep clustering-based baseline methods and compared the impact of dropout events on performance on four datasets of different sizes. To this end, we set the dropout rate to 15%–30% and kept other parameters at default values to simulate a "dropout" event. Each method was run five times, and the clustering performance of each method under different dropout rates was evaluated by calculating the average value of ARI. As shown in Fig. 8 and Supplementary Table S10, the average ARI of scMVAF under different dropout rates remains stable. In particular, in the Goolam dataset, scMVAF always maintains a high stability, while the performance of other methods drops significantly with the increase of dropout rate, and only a few results drop <0.5. In the PBMC-Ding dataset, the ARI scores for all methods are relatively close and change little under different dropout rates. In the Qx_Bladder dataset, we observed that the performance of scDFC improved with the increase of dropout rate, while in the PBMC-Zheng4k dataset, the performance of scDFC and scDCCA also improved with the increase of dropout rate. This may be because the dropout simulation events removed some noise information that was not relevant to the task, which helped to improve the performance of these methods in clustering tasks. Overall, scMVAF always outperformed the other three methods at different dropout rates and maintained high stability, which shows that this method can effectively deal with the common sparsity and random loss problems in gene expression data. Even in the case of data loss, scMVAF can still maintain relatively good clustering performance.

Comparison of embeddings learned by different methods

In order to intuitively verify whether scMVAF can obtain better embeddings compared with other deep learning-based methods, this paper visualizes the embeddings learned by three deep clustering-based methods on the Qx_Bladder and QS_Trachea datasets. The embeddings learned by all methods are visualized by t-SNE [23]. As shown in Fig. 9, scMVAF can separate cells well compared with other methods. In Fig. 9A, scDCCA [19] and scDeepCluster [17] tend to mix different types of cells together. The possible reason is that the multidimensional feature learning of cells is incomplete, which affects the accuracy of clustering. In Fig. 9B, the cell distribution of scDCCA [19] shows a high degree of overlap, which indicates that it has certain limitations in distinguishing different cell populations. In contrast, scMVAF can separate cell populations more clearly and show clearer cluster boundaries. In contrast, our method shows better performance. This may be due to the adoption of a multi-view learning strategy. The fusion module integrates data from multiple views to achieve multi-view analysis of samples. Each view provides partial information of the sample, and the fusion module effectively captures the complementary relationship between views by learning a shared embedding space, thereby significantly improving the clustering performance.

Marker gene identification

Marker genes can be used to distinguish different types of cells and identify the functional state of cells. To further test the ability of

scMVAF in identifying specific genes, for the embryonic dataset Yan containing 90 cells, we used the non-parametric Wilcoxon rank sum test to identify the top 3 differentially expressed genes in each predicted cluster and used them as marker genes for the cluster. In addition, to ensure the accuracy of the statistical test results, we averaged the repeated ranks.

Figure 10 shows the expression levels of the top 3 marker genes in each cluster identified from the Yan dataset. As shown in Fig. 10, the expression of the selected specific genes in each cell population is significantly different, indicating that scMVAF can

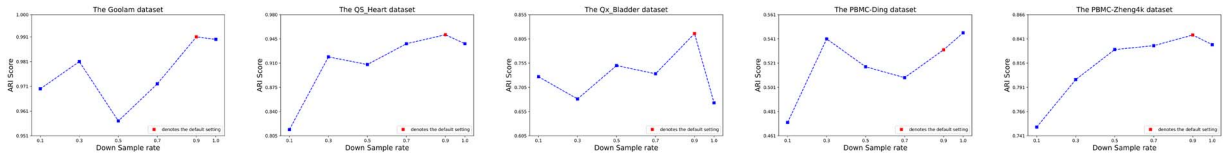

Figure 6. The effect of down-sampling rate in scMVAF.

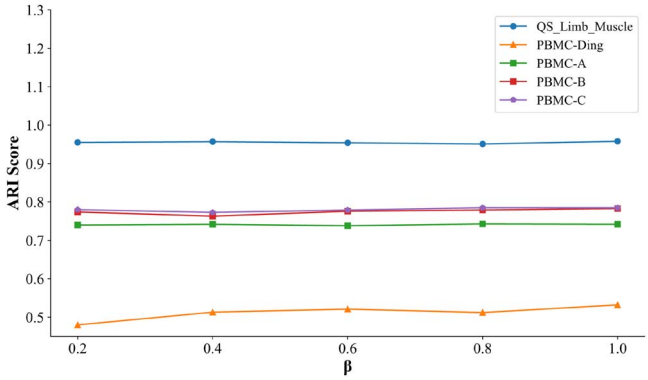

Figure 7. ARI scores of scMVAF with different β values.

better distinguish cell types. For example, in cluster 2, the top two marker genes identified are DUX4L2 and DUX4. DUX4L2 is a gene related to embryonic development. DUX4 activates a series of genes related to the 2-cell embryonic stage. Studies have shown that DUX4L2 and DUX4 are marker genes for 2-cell stage cells [44]. In addition, although some expressed genes do not match the expressed genes in existing studies. As shown in Fig. 10, they are highly expressed in specific clusters, which may help discover new marker genes.

Ablation experiments

scMVAF mainly consists of an adaptive weighted fusion module and a self-supervised learning module. In order to verify the contribution of these two modules to the model performance, this paper proposes two scMVAF variants. The two variants are scMVAF-FK (the adaptive fusion module and the self-supervised learning module are deleted) and scMVAF-K (only the adaptive fusion module is retained). As shown in Fig. 11 and Supplementary Table S11, this paper selects eight real datasets of different sizes for ablation experiments.

As shown in Fig. 11, compared with its two variants, scMVAF performs the best and scMVAF-FK performs the worst. This result shows that it is crucial to obtain consistency between cell views, which helps to more accurately capture the relationship between cells from different angles. In addition, individual cell views show differences, and the adaptive weighted fusion module can effectively capture these differences, resulting in better discriminative features. In the PBMC-Ding and PBMC-Zheng4k datasets, the clustering performance of scMVAF-K is significantly better than that of scMVAF-FK, which further verifies the effectiveness of the weighted fusion module. Overall, scMVAF achieves the best performance in all cases, which shows that our network is efficient and robust and can handle datasets of various sizes well.

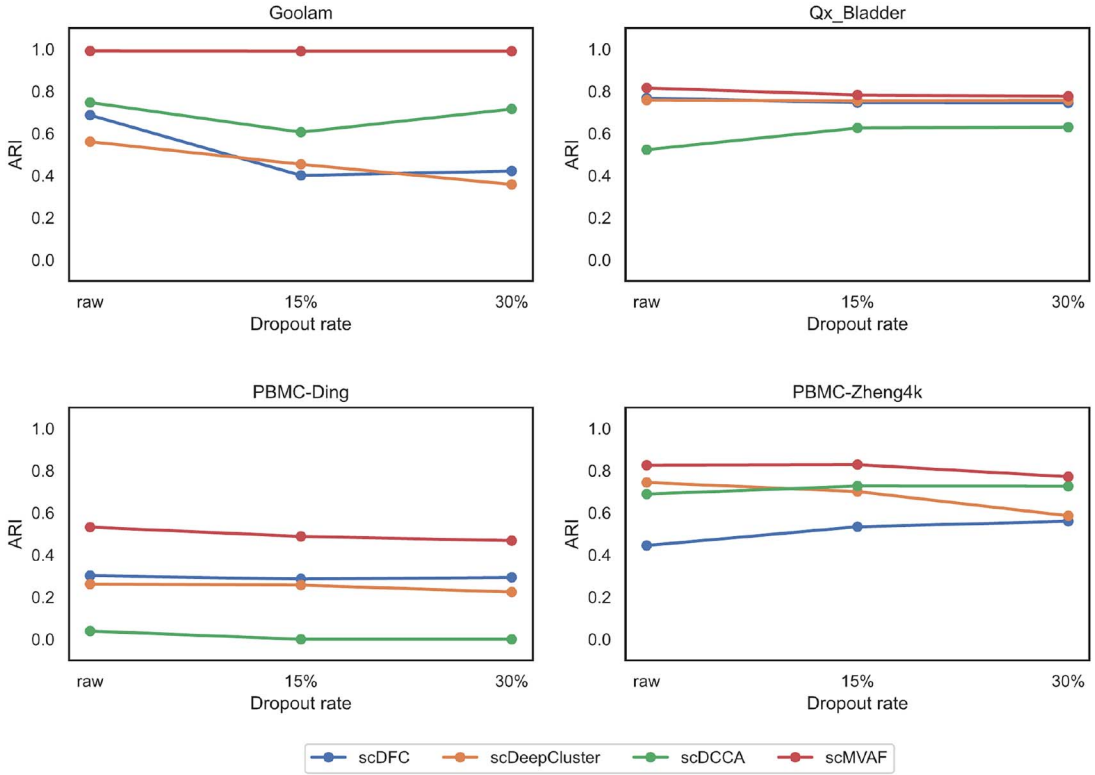

Figure 8. Comparison of methods across different dropout rates

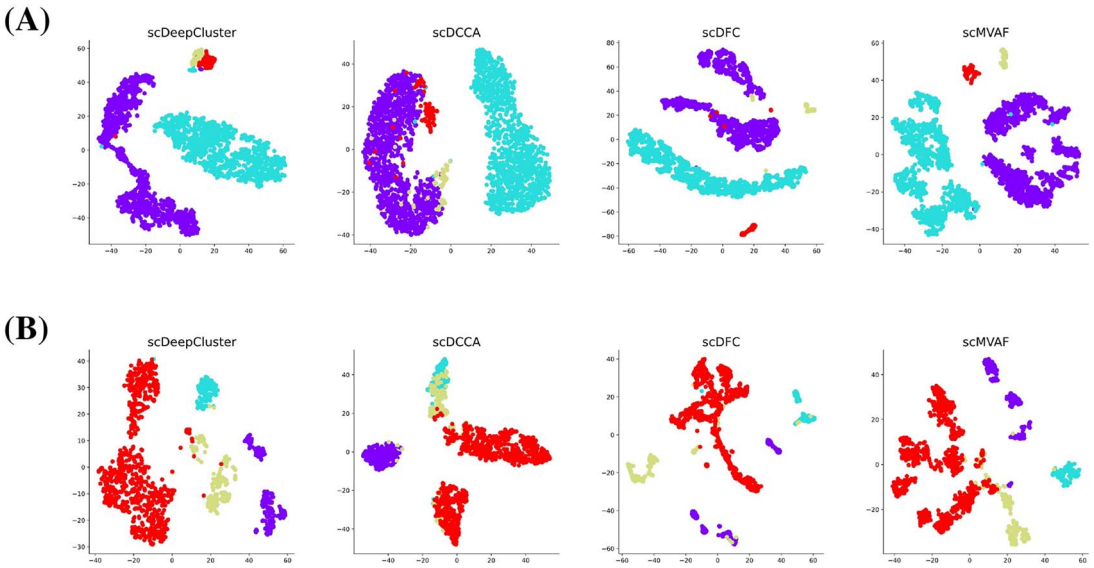

Figure 9. Visual analysis of the three deep learning methods applied on the two datasets using t-SNE.(A) Qx_Bladder dataset; (B) Trachea dataset.

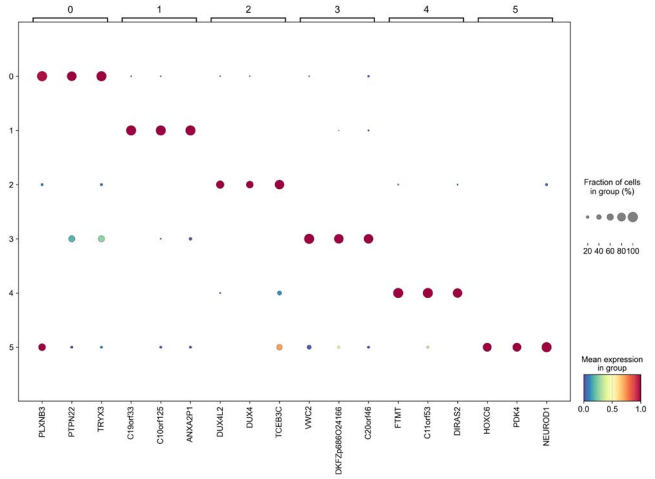

Figure 10. Expression dot plot for the Yan dataset (The color of the dot represents the average gene expression level within each cluster, while the dot size reflects the proportion of cells expressing the gene in each group)

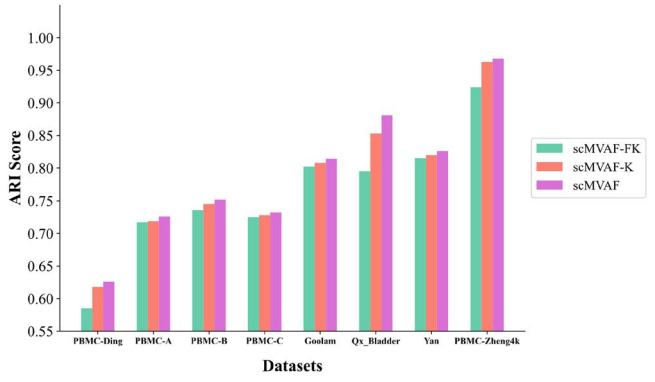

Figure 11. ARI scores of scMVAF-FK, scMVAF-K, and scMVAF. The results show that scMVAF achieves the first place in all eight datasets, and scMVAF always outperforms scMVAF-K, which indicates that multi-view self-supervised learning can improve clustering performance.

Conclusion

Single-cell sequencing can reveal cellular heterogeneity from a molecular perspective. Over the past decade, scRNA-seq datasets have grown exponentially, leading to the development of various cell clustering approaches to identify cell types. However, these methods consider cells from only a single perspective, failing to fully capture the differences and complex relationships between cells from multiple perspectives. This limitation leads to poor results in cell subtype clustering. In this paper, a novel multi-view clustering framework for single-cell data is proposed to enhance feature diversity by fusing features from multiple views to facilitate the identification of cell subtypes. The adaptive fusion of different views enhances the representation learning of data, and the complementarity between views helps to more comprehensively capture the diversity and complexity of data, thereby improving the robustness of clustering. Experimental results show that our method outperforms other baseline methods on 16 real datasets.

In the future, feature selection methods are planned to be added to further improve the method. Since interactions between genes may enhance feature representation, gene–gene pairwise interactions can be used to promote learning of low-dimensional data, thereby improving the robustness of the network. In addition, the complexity of the model can be further reduced by studying lightweight models.

Key Points

- Observing cells from different perspectives is crucial to understanding their characteristics. The integration of multi-perspective data can more comprehensively capture the differences and complex relationships between cells.
- We fuse cell features learned from different perspectives to help improve the accuracy of cell clustering. This feature fusion can more accurately reveal the intrinsic structure of cell populations.
- We integrated the multi-view adaptive fusion module and the self-supervision module into scMVAF to enhance the ability to distinguish cell diversity.

- Experimental results show that our proposed model performs well on 16 datasets, outperforming all other baseline methods.

Acknowledgements

The authors acknowledge Qiong Huang for supervision and funding support. The authors thank all editors and reviewers for the suggestions and comments.

Supplementary material

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

Conflicts of interest

None declared.

Funding

The work was supported by the Special Fund for Marine Economic Development (six marine industries) of Guangdong Province, China (GDNRC[2024]18) and the Fund of Southern Marine Science and Engineering Guangdong Laboratory (Zhanjiang)(ZJW-2023-04), Guangdong Provincial Education Department (2023ZDZX4002, 2021ZDZX1078), and Guangzhou Key Laboratory of Smart Agriculture (2019020-10081).

Data availability

QS_Limb_Muscle, QS_Trachea, Qx_Bladder, Qx_Limb_Muscle, QS_Heart, and QxSpleen can be downloaded from the Tabula Muris database, while other datasets can be downloaded for free at https://hemberg-lab.github.io/scRNA.seq.datasets/ and https://github.com/LQXLE/scMVAF/.

References

1. Artis D, Spits H. The biology of innate lymphoid cells. *Nature* 2015; **517**:293–301. https://doi.org/10.1038/nature14189
2. Grosselin K, Durand A, Marsolier J *et al*. High-throughput single-cell chip-seq identifies heterogeneity of chromatin states in breast cancer. *Nat Genet* 2019;**51**:1060–6. https://doi.org/10.1038/s41588-019-0424-9
3. Ellsworth DL, Blackburn HL, Shriver CD *et al*. Single-cell sequencing and tumorigenesis: improved understanding of tumor evolution and metastasis. *Clin Transl Med* 2017;**6**:15.
4. Marioni JC, Arendt D. How single-cell genomics is changing evolutionary and developmental biology. *Annu Rev Cell Dev Biol* 2017;**33**:537–53. https://doi.org/10.1146/annurev-cellbio-100616-060818
5. Claudio Vinegoni J, Dubach M, Thurber GM *et al*. Advances in measuring single-cell pharmacology in vivo. *Drug Discov Today* 2015;**20**:1087–92. https://doi.org/10.1016/j.drudis.2015.05.011
6. Parkhomchuk D, Amstislavskiy V, Soldatov A *et al*. Use of high throughput sequencing to observe genome dynamics at a single cell level. *Proc Natl Acad Sci USA* 2009;**106**:20830–5. https://doi.org/10.1073/pnas.0906681106
7. Tian HC, Benitez JJ, Craighead HG *et al*. Single cell on-chip whole genome amplification via micropillar arrays for reduced amplification bias. *PLoS One* 2018;**13**:e0191520. https://doi.org/10.1371/journal.pone.0191520
8. Kharchenko PV, Silberstein L, Scadden DT. Bayesian approach to single-cell differential expression analysis. *Nat Methods* 2014;**11**:740–2. https://doi.org/10.1038/nmeth.2967
9. Chen L, Wang W, Zhai Y *et al*. Deep soft k-means clustering with self-training for single-cell RNA sequence data. *NAR Genom Bioinform* 2020;**2**. https://doi.org/10.1093/nargab/lqaa039
10. Lin P, Troup M, Ho JWK. CIDR: ultrafast and accurate clustering through imputation for single-cell RNA-seq data. *Genome Biol* 2017;**18**:1–11.
11. Wang B, Ramazzotti D, De Sano L *et al*. SIMLR: a tool for large-scale genomic analyses by multi-kernel learning. *Proteomics* 2018;**18**:1700232. https://doi.org/10.1002/pmic.201700232
12. Satija R, Farrell JA, Gennert D *et al*. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 2015;**33**:495–502. https://doi.org/10.1038/nbt.3192
13. Grün D, Kester L, Van Oudenaarden A. Validation of noise models for single-cell transcriptomics. *Nat Methods* 2014;**11**:637–40. https://doi.org/10.1038/nmeth.2930
14. Kiselev VY, Andrews TS, Hemberg M. Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat Rev Genet* 2019;**20**:273–82. https://doi.org/10.1038/s41576-018-0088-9
15. Bin Y, Chen C, Qi R *et al*. scGMAI: a Gaussian mixture model for clustering single-cell RNA-seq data based on deep autoencoder. *Brief Bioinform* 2021;**22**:bbaa316.
16. Eraslan G, Simon LM, Mircea M *et al*. Single-cell RNA-seq denoising using a deep count autoencoder. *Nat Commun* 2019;**10**:390. https://doi.org/10.1038/s41467-018-07931-2
17. Tian T, Wan J, Song Q *et al*. Clustering single-cell RNA-seq data with a model-based deep learning approach. *Nat Mach Intell* 2019;**1**:191–8. https://doi.org/10.1038/s42256-019-0037-0
18. Ciortan M, Defrance M. Contrastive self-supervised clustering of scRNA-seq data. *BMC bioinformatics* 2021;**22**:280. https://doi.org/10.1186/s12859-021-04210-8
19. Wang J, Xia J, Wang H *et al*. scDCCA: deep contrastive clustering for single-cell RNA-seq data based on auto-encoder network. *Brief Bioinform* 2023;**24**:bbac625. https://doi.org/10.1093/bib/bbac625
20. Dayu H, Liang K, Zhou S *et al*. scDFC: a deep fusion clustering method for single-cell RNA-seq data. *Brief Bioinform* 2023;**24**:bbad216.
21. Kiselev VY, Kirschner K, Schaub MT *et al*. Sc3: consensus clustering of single-cell RNA-seq data. *Nat Methods* 2017;**14**:483–6. https://doi.org/10.1038/nmeth.4236
22. Yang Y, Huh R, Culpepper HW *et al*. Safe-clustering: single-cell aggregated (from ensemble) clustering for single-cell RNA-seq data. *Bioinformatics* 2019;**35**:1269–77. https://doi.org/10.1093/bioinformatics/bty793
23. Van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008;**9**:2579–2605.
24. Li F, Liu Y, Liu J *et al*. A framework for scRNA-seq data clustering based on multi-view feature integration. *Biomedical Signal Processing and Control* 2024;**89**:105785. https://doi.org/10.1016/j.bspc.2023.105785
25. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* 2018;**19**:1–5.
26. Yan L, Yang M, Guo H *et al*. Single-cell RNA-seq profiling of human preimplantation embryos and embryonic stem cells. *Nat Struct Mol Biol* 2013;**20**:1131–9. https://doi.org/10.1038/nsmb.2660

27. Goolam M, Scialdone A, Graham SJL *et al*. Heterogeneity in Oct4 and Sox2 targets biases cell fate in 4-cell mouse embryos. *Cell* 2016;**165**:61–74. https://doi.org/10.1016/j.cell.2016.01.047
28. Deng Q, Ramsköld D, Reinius B *et al*. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* 2014;**343**:193–6. https://doi.org/10.1126/science.1245316
29. Gray Camp J, Sekine K, Gerber T *et al*. Multilineage communication regulates human liver bud development from pluripotency. *Nature* 2017;**546**:533–8. https://doi.org/10.1038/nature22796
30. Iram T. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* 2018;**562**:367–72. https://doi.org/10.1038/s41586-018-0590-4
31. Zhang W, Ruochen Y, Zeqi X *et al*. scCompressSA: dual-channel self-attention based deep autoencoder model for single-cell clustering by compressing gene–gene interactions. *BMC Genomics* 2024;**25**:423. https://doi.org/10.1186/s12864-024-10286-2
32. Gan L, Vinci G, Genevera AI. Correlation imputation in single cell RNA-seq using auxiliary information and ensemble learning. In: *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, Virtual Event, USA*, pp. 1–6, 2020.
33. Zhang X-F, Ou-Yang L, Yang S *et al*. EnImpute: imputing dropout events in single-cell RNA-sequencing data via ensemble learning. *Bioinformatics* 2019;**35**:4827–9. https://doi.org/10.1093/bioinformatics/btz435
34. Yin HC, Tao JX, Peng Y *et al*. MSPJ: discovering potential biomarkers in small gene expression datasets via ensemble learning. *Comput Struct Biotechnol J* 2022;**20**:3783–95. https://doi.org/10.1016/j.csbj.2022.07.022
35. Zuguang G, Schlesner M, Hübschmann D. cola: an R/Bioconductor package for consensus partitioning through a general framework. *Nucleic Acids Res* 2021;**49**:e15-5.
36. Vincent P, Larochelle H, Bengio Y *et al*. Extracting and composing robust features with denoising autoencoders. In: *Proceedings of the 25th international conference on Machine learning*, Helsinki, Finland, pp. 1096–1103, 2008.
37. Xinlei X, Wang Z, Ren S *et al*. Local–global geometric information and view complementarity introduced multiview metric learning. *IEEE Trans Neural Netw Learn Syst* 2025;**36**:5428–5441.
38. Loshchilov I, Hutter F. Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101. 2017.
39. Zeiler MD. Adadelta: An adaptive learning rate method. arXiv preprint arXiv:1212.5701. 2012.
40. Xie J, Girshick R, Farhadi A. Unsupervised deep embedding for clustering analysis. In: Balcan MF, Weinberger KQ (eds.), *International Conference on Machine Learning*, pp. 478–487. PMLR, Brooklyn, NY, USA, 2016.
41. Hubert L, Arabie P. Comparing partitions journal of classification 2 193–218. *Google Scholar* 1985;**2**:193–218. https://doi.org/10.1007/BF01908075
42. Strehl A, Ghosh J. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *J Mach Learn Res* 2002;**3**:583–617.
43. Kumar A, Yadav J. A review of feature set partitioning methods for multi-view ensemble learning. *Inform Fusion* 2023;**100**:101959. https://doi.org/10.1016/j.inffus.2023.101959
44. Alberto DI, Evarist P, Andrea C *et al*. DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat Genet* 2017;**49**:941–5. https://doi.org/10.1038/ng.3858

Briefings in Bioinformatics, 2026, **27(3)**, bbaf169
https://doi.org/10.1093/bib/bbaf169
Problem Solving Protocol