

scMAG : Integrating Single-Cell Multi-Omics Data via Multi-Stage Deep Fusion with Manifold-Aware Gating

Junhao Zou^{1,2}, Jiaqi Liu^{2,3}, Shuangquan Li¹, Liu Liu⁴, Chunquan Li^{2,3,4,5*}, Huifang Tang^{3,5*}

1 School of Automation, Harbin University of Science and Technology, Harbin, Heilongjiang, 150080, China.

2 The First Affiliated Hospital, Cardiovascular Lab of Big Data and Imaging Artificial Intelligence, Hengyang Medical School, University of South China, Hengyang, Hunan, 421001, China.

3 Hunan Provincial Key Laboratory of Multi-omics and Artificial Intelligence of Cardiovascular Diseases, University of South China, Hengyang, Hunan, 421001, China.

4 School of Computer, University of South China, Hengyang, Hunan, 421001, China.-

5 The First Affiliated Hospital, Institute of Cardiovascular Disease, Hengyang Medical School, University of South China, Hengyang, Hunan, 421001, China

† The authors wish to be known that the first two authors should be regarded as joint first authors.

*Corresponding author. E-mail: lcqbio@163.com, 2018010003@usc.edu.cn

Abstract

Background: Single-cell omics technologies now enable simultaneous profiling of multiple genomic modalities within individual cells. Integration of these multi-omics data necessitates computational frameworks that establish cross-modal associations while preserving biological fidelity. A central challenge lies in balancing two competing objectives: alignment of heterogeneous omics layers and retention of modality-specific distributions. Overly aggressive alignment risks semantic distortion, whereas strict distribution preservation may cause modal segregation in latent representations. This fundamental trade-off underscores the need for advanced strategies to harmonize integrative accuracy with biological authenticity.

Results: This paper innovatively proposes a multi-stage feature deep fusion formalization method with omics multi-core manifold preservation and guided gating optimization strategy and designs the scMAG algorithm. This is a framework aimed at improving the clustering accuracy and data visualization of single-cell multi-omics, while achieving adaptive alignment of multi-omics latent spaces and optimizing the distribution of omics data in the environmental space, effectively suppressing biological noise and measurement errors. To comprehensively evaluate the scMAG algorithm, we compared the paired datasets scRNA-seq with scATAT-seq and ADT as a benchmark. We consistently observed that scMAG outperformed other algorithms in terms of clustering effect and data visualization clarity. Further multi-task experimental analysis indicates that this enhancement stems from the ability of scMAG to improve the distribution of potential features in the data space and adaptively balance the shared signals and modality-specific signals across multi-omics.

Conclusion: scMAG not only significantly improved clustering performance, but also showed excellent performance in feature dimensionality reduction, batch effect removal, multimodal data integration, cell trajectory inference, etc., and also had good biological interpretability. This method provides new theoretical support and practical reference for in-depth analysis of multimodal single-cell data.

Keywords: Multi-stage fusion, Manifold learning, Guided gating mechanism, Distribution Alignment

1 Background

The rapid development of single-cell technology has significantly advanced cell research, making multi-dimensional analyses such as single-cell transcriptome (scRNA-seq), epigenome (scATAC-seq), and proteome possible. With continuous breakthroughs in these technologies and the accumulation of data resources, there is an increasing need for computational methods capable of integrating information from different modalities to conduct joint analysis of single-cell multi-omics data, thereby providing a more comprehensive understanding of cell states and functions [1, 2].

The rapid development of single-cell multi-omics technologies has provided unprecedented opportunities for analyzing cell heterogeneity. However, the integration of multi-omics data faces three core challenges. Firstly, the complexity of biological associations significantly affects the integration outcome. The strength of feature correlations between different omics modalities varies fundamentally [3, 4]. Chromatin accessibility (scATAC-seq) and gene expression (scRNA-seq) have clear regulatory relationships and can be integrated with high accuracy through linear or nonlinear models; while protein abundance and mRNA levels are affected by post-transcriptional modifications and show weak correlations, resulting in a significant decline in performance of traditional methods, such as Canonical Correlation Analysis (CCA) when integrating CITE-seq (RNA and protein) data [5]. Subsequently, the limitations of omics technologies exacerbate data noise. scRNA-seq can capture the entire transcriptome, but antibody-based protein detection can only cover tens to hundreds of targets, and the imbalance in throughput weakens the available signal intensity [6]. More seriously, the inherent noise of single-cell data leads to false zero values for up to 90% of genes due to technical omissions, resulting in data following a zero-inflated negative binomial distribution (ZINB), and the bottleneck of proteomics detection further amplifies the noise [7, 8]. Thereafter, the differences in feature dimensions of different omics are huge and their distributions are heterogeneous. Transcriptional data is a count-negative binomial distribution, epigenetic data is a binary distribution, and protein data may be a continuous Gaussian distribution. Forced alignment may lead to information loss [9].

The cornerstone framework for current data integration is to use the MLP calculation method to achieve end-to-end dimensionality reduction and denoising. However, when dealing with single-cell multi-omics data, there are significant limitations in aspects such as topological preservation, weak association learning, and computational efficiency. By jointly modeling the multi-omics distributions using variational autoencoders (VAEs), it is possible to map different omics data from the same cell to a shared embedding space [10, 11]. Utilizing this shared embedding space, one can effectively obtain a joint low-dimensional representation of multi-omics. This framework, as the current mainstream method for single-cell data processing, has evolved to develop methods such as scMM and totalVI for joint dimensionality reduction, cell clustering, and batch correction, and has demonstrated advantages in uncertainty quantification and feature extraction [12, 13]. However, these VAEs rely on the assumption of simplified real posterior distributions and are difficult to capture the multi-modal characteristics of the real cell state, and are prone to overfitting on low-throughput protein data. The current research on single-cell multi-omics data integration under the VAE framework has further developed the scDiffusion-X framework based on the fusion of VAE and diffusion processes, achieving the generation and translation of multi-omics data through a dual cross-attention mechanism [14]. The uniPort framework and

moscot framework based on the combination of VAE and optimal transport theory achieve efficient mapping of multi-omics data [15, 16]. The scMODAL framework based on feature-guided geometric regularization uses known associated features to construct anchors and align the low-dimensional representation distributions of multi-omics data through GAN alignment [17]. The scCross deep generative model realizes the closed loop of cross-modal generation from the unified latent space of multi-omics to any modality [18].

The VAE converts high-dimensional data into low-dimensional latent space representations, which can reduce noise but will lose the local topological structure of the environmental space. The latent variables being forced to follow a Gaussian distribution will further lead to the inability of the latent space to retain the complex structure of the original data [19]. To address the distribution deviation problem of multi-omics data, a mechanism for optimizing the environmental space distribution based on manifold fitting can be introduced [20, 21]. This directly corrects the data distribution in the high-dimensional space and retains the biological topological structure. The aim of this method is to reconstruct a smooth manifold within the original space of the measurement data, capturing the low-dimensional structure of the data in a way that minimizes information loss and effectively eliminates noise. The latest innovative flow fitting provides a solution to the limitations of existing methods [22].

This article innovatively improves the existing single-cell multi omics integration framework. In terms of model structure, independent optimization and distribution alignment are performed on each set of learning data before inputting the Gated Variational Autoencoder. This not only preserves the manifold structure of each modality itself, but also effectively alleviates potential spatial conflicts caused by distribution differences during data fusion. Subsequently, through the combination of segmented fusion and guided gating mechanisms, it was achieved to improve the information transmission efficiency of the bottleneck layer while preserving the manifold nature of the data, providing theoretical and structural support for the accuracy and clustering performance of potential representations. The main contributions are reflected in three aspects.

1. Manifold regularized representation: In the original measurement data space, a multi view manifold representation was constructed and similarity measurement was performed through multi kernel learning to capture the low dimensional manifold structure of the data in a way that minimizes information loss and effectively eliminates noise, while preserving the expression distribution patterns and feature differences under each view as much as possible. By utilizing flexible neighborhood definitions and kernel function combinations, the spatial distribution of data is significantly enhanced, which facilitates the use of multi-layer perceptron (MLP) to fit distribution parameters and comprehensively model zero mean negative binomial distribution (ZINB).

2. Latent feature gating mechanism: Combining the heuristic optimization strategy of forgetting gate and EM (Expectation Maximization), and guiding the optimization process through smooth clustering, not only can effective information be retained in each clustering process, but also the accumulation of useless features can be avoided. On the one hand, it can optimize the information transmission efficiency of the VAE bottleneck layer, and on the other hand, force the model to rely only on cell embeddings to reconstruct gene expression, thereby improving clustering performance and reconstruction quality through bottleneck learning tasks, achieving more obvious intra class aggregation, and ensuring clearer separation between classes.

3. Multi-stage feature deep fusion: Under the framework of multi task collaborative optimization, the fusion

of feature spaces is dynamically regulated at different stages through independent fusion strategies, achieving effective alignment and complementarity of data. Combined with hierarchical guidance of information flow and fusion, it not only enhances the adaptability of the model to complex data structures, but also further strengthens the discriminative and expressive ability of features in the latent space.

The article is structured as follows, and in Section 2 we give a detailed description of the scMAG method and the specific analysis steps. In Section 3, we present the results and the experimental proof for the model. In Section 4, we perform a multi-task downstream analysis and meanwhile also giving the relevant biological validation. Finally, in sections 5 and 6, we provided relevant discussions and conclusions.

2 Methods

2.1 The scMAG workflow

scMAG is an end-to-end single cell omics clustering analysis method, and its overall process is shown in Figure 1A. After the single-cell omics raw data is transformed to improve the signal-to-noise ratio in the data and ensure a more accurate and reliable data representation, scMAG model performs multiple kernel learning for sparse representation on a manifold, low dimensional manifold space and efficient clustering results are obtained, which are then used for multi downstream task analysis and the corresponding results and analysis are given. The specific model architecture is shown in Figure 1B, and scMAG is a latent feature gating cluster-guided variational autoencoder. Firstly, the three transformations of the original data are matched with kernel function mapping, and then manifold fitting is performed to obtain the fitting data corresponding to the three transformations. Finally, the optimal fitting data is selected through multi kernel learning, as shown in Figure 1C. The random noise is added to the two encoders of the model, forcing the model to learn feature representations that are insensitive to input perturbations, thereby enhancing the model's robustness, generalization ability, and learning a more representative latent space. The encoder adopts a partially shared coding mode to achieve the alignment of data feature space, which can also better remove batch effects. We designed an EM heuristic optimization strategy based on forgetting gate to fuse the feature space to improve the clustering effect, named Selective Entropy Optimization (SER), as shown in Figure 1D. Staged fusion allows for the initial focus on omics data reconstruction, and then the gradual introduction of KL divergence and clustering objectives, thereby avoiding optimization conflicts and obtaining a more stable model.

2.2 Manifold regularized representation

We used a total of 9 paired omics for validation, including 7 omics containing RNA and ADT counts and 2 omics containing RNA and ATAC counts, as well as their cell type labels. We performed three rounds of data transformation on the original data and matched it with a kernel function mapping. Then, we obtained a popular regularized omics representation through popular fitting, and finally obtained the optimal fitting data through multi kernel learning [22] as the input data for the model.

2.2.1 Data conversion and kernel function mapping

Using three types of data transformation to extract potential biological structural differences from multiple perspectives, providing multiple possibilities for data structure, avoiding loss of key information, and laying the

foundation for constructing the most authentic manifold structure. These three conversions are: logarithmic conversion (Log), Z-score normalization (Z-score), and concentrated logarithmic ratio conversion (CLR).

Log transformation is a commonly used method in single-cell data preprocessing. It is mainly used to narrow the expression span, reduce the dominance of highly expressed genes, and stabilize the variance of the data, making the data closer to the normal distribution, which is conducive to subsequent dimension reduction, clustering and other analysis. Each expression vector x_i is transformed as follows:

$$y_{1i} = \log 2(x_i + 1) \quad (1)$$

Z-score normalization is used to convert gene expression data into a distribution with a mean of 0 and a standard deviation of 1, so as to eliminate the influence of different gene expression levels and make each gene comparable in dimension reduction and clustering. Each expression vector x_i is transformed as follows:

$$y_{2i} = \frac{x_i - \mu}{\sigma} \quad (2)$$

The central logarithmic ratio transformation is mainly used for preprocessing proportional data, and this technique is quite common in CITE seq protein data (ADT). It can effectively process the composition data, avoid the problem of spurious correlation caused by the constraint of the sum of proportional values, and make the data more suitable for dimensionality reduction, clustering, and other operations. The conversion method for each expression vector x_i is as follows:

$$y_{3i} = \log\left(\frac{x_i}{g(x)}\right), g(x) = \left(\prod_{j=1}^D x_j\right)^{\frac{1}{D}} \quad (3)$$

where x_i is the i th component in the sample, $g(x)$ is the geometric mean of all components of the sample, and D is the number of features.

The converted data are $y \in \{y_1, y_2, y_3\}$. For a transformed data $y_i \in \{y_{1i}, \dots, y_{mi}\}$, the manifold hypothesis assumes that the y_i can be viewed as a noisy representation of the underlying low-dimensional manifold M of dimension d ($d \ll D$), with D is the number of features. However, real single-cell data is nonlinear and separable, and its manifold structure may be highly nonlinear, and manifold fitting is a linear operation. Therefore, we introduce nonlinear similarity structures through kernel functions, implicitly mapping the raw data to a linearly separable high-dimensional kernel space [23,24]. In this space, nonlinear structures become "linearly separable" or easier to fit, thereby enhancing the modeling ability and expression effect of manifold fitting on potential low dimensional manifold structures [25]. Due to the different statistical distributions and structural characteristics of different expression transformations, kernel functions that match their structural characteristics are selected for high-dimensional spatial mapping and structural modeling.

In order to fully explore the potential structural features under different data transformation perspectives, we designed four combination strategies of kernel functions and transformation methods. Firstly, the data after logarithmic transformation (Log) preserves the nonlinear trend in the expression pattern while compressing extreme values. Therefore, we use Gaussian kernels to capture local nonlinear relationships and polynomial kernels to supplement the explanatory power of nonlinear combinations [24]. Secondly, the standardized (Z-score) processed data has the property of zero mean and one variance, and a polynomial kernel is selected to

enhance its feature interaction ability [26]. Finally, for the centralized logarithmic ratio transform (CLR) commonly used in omics proportional data, due to its ability to significantly improve the comparability of closed data and reveal internal correlation structures, it is suitable to use Gaussian kernels to characterize and measure local structures[27].

We use Gaussian kernel (RBF) and Polynomial kernels for nonlinear mapping and obtain four data K_1, K_2, K_3, K_4 .

$$K(y_{ai}, y_{aj}) = \exp\left(-\frac{\|y_{ai} - y_{aj}\|^2}{2\sigma^2}\right), K(y_{ai}, y_{aj}) = (y_{ai}^T y_{aj} + c)^d, a = 1, 2, 3 \quad (4)$$

2.2.2 Sparse representation on manifolds

The fitting process includes two parts, direction estimation and projection estimation.

Direction estimation. Determining the projection direction is to determine the direction of point movement, so that the data is closer to the true manifold structure. For any point $K \in \{K_1, K_2, K_3, K_4\}$, its projection direction can be expressed as $d_K = F(K) - K$, where $F(K)$ is local weighted average of y . $F(K)$ is defined by the following formula [28]:

$$F(K) = \sum_i \alpha_i(K) K^{(i)} \quad (5)$$

where defined as:

$$\alpha_i(K) = \frac{\tilde{\alpha}_i(K)}{\sum_{i \in I_y} \tilde{\alpha}_i(K)}, \tilde{\alpha}_i(K) = \begin{cases} \left(1 - \frac{\|K - K^{(i)}\|_2^2}{r_0^2}\right)^t, & \|K - K^{(i)}\|_2 \leq r_0 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

and $t > 2$ (t is a fixed integer, usually set to 3) ensures smoothness.

Computing the shared nearest neighborhood. For each $K^{(i)}$, we denote $N_p(i)$ as p (default by 15) nearest neighborhood of $K^{(i)}$, as determined by a given metric. $N_p(i)$ contains the key sample points where $K^{(i)}$ should pay attention to. Hence, the shared nearest neighborhood (SNN) of $K^{(i)}$ and $K^{(j)}$ is defined as:

$$SNN(i, j) = |N_p(i) \cap N_p(j)| \quad (7)$$

The final projection direction $F(K^{(i)})$ is defined as:

$$\mathbb{R}(i) = \arg \max_{S \subset K, |S|=p} \sum_{K^{(j)} \in \mathbb{R}(i)} SNN(i, j), F(K^{(i)}) = \frac{1}{|\mathbb{R}(i)|} \sum_{K^{(j)} \in \mathbb{R}(i)} K^{(j)} \quad (8)$$

Projection estimation. Considering that single-cell data has high-dimensional features and irregular noise, this method simplifies projection estimation $G(K^{(i)})$ by identifying the maximum density point on the line connecting $K^{(j)}$ and $F(K^{(i)})$, and replacing the original $G(K^{(i)})$ with that point. The calculation process is as follows:

$$G(K^{(i)}) = \arg \max_{K_t} \rho(K_t), K_t = K^{(i)} + t(F(K^{(i)}) - K^{(i)}), \rho(K^{(i)}) = \frac{1}{\sum_{K^{(j)} \in \mathbb{R}(i)} \|K^{(i)} - K^{(j)}\|_2^2} \quad (9)$$

The fitting data is represented as :

$$Y = \left\{ G(K^{(i)}) \right\}_{i=1}^N \quad (10)$$

After manifold fitting, four corresponding fitting data were obtained, which are Y_1, Y_2, Y_3, Y_4 .

2.2.3 Multi kernel weighted fusion learning (MKL)

The weighted data Y is the input data of the model, and its weighted expression is:

$$Y = \sum_{i=1}^4 \alpha_i Y_i \quad (11)$$

where the optimization objectives are:

$$\min_{\alpha} \sum_{i=1}^m \alpha_i \cdot \|Y - Y_i\|_F^2 + \lambda \|\alpha\|^2 \quad (12)$$

and $\alpha \in \mathbb{R}^m, \alpha_i \geq 0, \sum \alpha_i = 1$.

2.3 Model training

scMAG reduces dimensionality while preserving some inherent connections in the data through the first stage of fusion, then imports its connections during the second stage of fusion, and finally maximizes the information flow of the bottleneck layer through two independent manifold spaces.

Firstly, we will add random noise to the preprocessed RNA and ADT/ATAC data, e , which are expressed as:

$$X_1 = X_R + k_R \times N_R, \quad X_2 = X_A + k_A \times N_A \quad (13)$$

Where $X_R \in \mathbb{R}^{N \times D_R}$ and $X_A \in \mathbb{R}^{N \times D_A}$ are the pretreated RNA and ADT/ATAC data, and N is the number of cells, D_R and D_A are the characteristic dimensions of RNA and ADT/ATAC data. N_R and N_A are the random noises that are added to both (the mean is 0 and the variance is 1). k_R and k_A are the coefficients of the random noise, set to 2 and 1.5, respectively.

We adopt a variational autoencoder as the overall architecture of the model [29] to achieve nonlinear mapping and feature dimensionality reduction. The model extracts features through two partially shared encoders E_1 and E_2 [11], through feature splicing and Bayesian inference, two feature spaces Z_1 and Z_2 are obtained, which are expressed as:

$$X_1^c = E_1(X_1), \quad X_2^c = E_2(X_2) \quad (14)$$

$$Z_1 = X_1^c \oplus X_2^c, \quad Z_2 = \mu_1 + \sigma_1 \times \varepsilon \quad (15)$$

In the reconstruction process, the feature space needs to fully preserve the detailed information of the original data to accurately restore its potential biological structure. Therefore, we adopt a weighted fusion strategy of two feature spaces to construct a reconstructed manifold space Z_r with more structural expression ability, thereby improving the reconstruction quality and biological consistency.

$$Z_r = Z_2 + \eta Z_1 \quad (16)$$

where X_1^c and X_2^c are the output of the two parts that share the encoder. μ_1 and σ_1 are the mean and standard deviation of Z_1 , and ε is the noise sampled from the standard normal distribution. η is a weight factor of Z_1 and is set to 0.3.

In the clustering process, we use the SER optimization strategy to obtain the clustering manifold space

$Z_c = \alpha \times Z_1$ for clustering analysis. The forget gate dynamically adjusts the information flow, enabling the model to have the ability to "actively select memory and forget". In feature extraction, it can improve the model's response to key signals while suppressing noise and redundant information. However, in the SER strategy, we utilized the results of soft clustering to gradually maximize the clustering expectation by completing the feature information in the iterative process. This approach can preserve effective information during each clustering process, avoid accumulating useless features, and help construct more robust and biologically interpretable potential spaces.

$$\alpha = \varphi(W_\alpha \cdot [Z_c, Q] + b) \quad (17)$$

where α is the coefficient matrix of Z_1 , φ is the sigmoid activation function, and W_α and b are the learnable parameters. Q is the soft allocation weight matrix, which is calculated as follows:

$$Q = [q_{ij}]_{m \times n}, q_{ij} \geq 0, \sum_{j=1}^n q_{ij} = 1 \quad (18)$$

where the calculation of q_{ij} can be found in the clustering loss.

Finally, the reconstructed manifold space Z_r passes through two completely independent decoders D_1 and D_2 , and three independent fully connected layers are added after the output of each decoder to obtain the three parameters M , θ and Π of the ZINB model [30], which are defined as: \hat{X}_2

$$X_1' = D_1(Z_r), \quad X_2' = D_2(Z_r) \quad (19)$$

$$M_R = \hat{X}_1 = \text{diag}(s_i^r) \times \exp(w_\mu^r X_1'), \quad \theta_R = \exp(w_\theta^r X_1'), \quad \Pi_R = \text{sigmoid}(w_\pi^r X_1') \quad (20)$$

$$M_A = \hat{X}_2 = \text{diag}(s_i^a) \times \exp(w_\mu^a X_2'), \quad \theta_A = \exp(w_\theta^a \hat{X}_2), \quad \Pi_A = \text{sigmoid}(w_\pi^a X_2') \quad (21)$$

where M_R , θ_R and Π_R are the mean, dispersion and exit probability of RNA data in ZINB loss, respectively, and M_A , θ_A and Π_A are the average, dispersion and exit probability of ADT/ATAC data in ZINB loss, respectively, M_R and M_A are also used as reconstruction data of RNA and ADT/ATAC. w_μ^r , w_θ^r , w_π^r , w_μ^a , w_θ^a and w_π^a is the learnable parameter for the six fully ligated layers, and s_i^r and s_i^a are the factor sizes for the RNA and ADT/ATAC data.

2.4 Parameter configuration

The model performed 400 iterations in pre-training and 100 iterations in training, and also set an iterative convergence threshold to avoid meaningless multiple trainings. Due to the large difference in the feature dimension between ADT and ATAC, and the difference in dimensionality between different datasets, we will set different encoding and decoding structures according to different datasets. For the RNA and ADT datasets, the two partially shared encoders are set to $\{256, 64, 32, 16\}$ and $\{64, 32, 16\}$, respectively, and the two decoders are set to $\{32, 64, 256\}$ and $\{32, 64\}$, respectively. For the RNA and ATAC datasets, the encoders shared by both parts are $\{256, 64, 32, 16\}$, and both decoders are $\{32, 64, 256\}$.

2.5 Loss function

ZINB loss. We use the zero-bloated negative binomial (ZINB) distribution for both RNA and ADT/ATAC data

as the reconstructed loss function of the model. It combines the modeling ability of negative binomial distribution for over discretization and an additional zero inflation mechanism, specifically targeting single-cell data with zero inflation and over discretization, to more realistically simulate the statistical characteristics of raw data, thereby improving the accuracy of VAE reconstruction and enhancing the expression ability of hidden space [30]. First, define the representation of the negative binomial (NB) distribution as:

$$NB(X_R | \mu_R, \theta_R) = \frac{\Gamma(X_R + \theta_R)}{X_R \Gamma(\theta_R)} \left(\frac{\theta_R}{\theta_R + \mu_R} \right)^{\theta_R} \left(\frac{\mu_R}{\theta_R + \mu_R} \right)^{X_R} \quad (22)$$

$$NB(X_A | \mu_A, \theta_A) = \frac{\Gamma(X_A + \theta_A)}{X_A \Gamma(\theta_A)} \left(\frac{\theta_A}{\theta_A + \mu_A} \right)^{\theta_A} \left(\frac{\mu_A}{\theta_A + \mu_A} \right)^{X_A} \quad (23)$$

where μ_R , θ_R , μ_A and θ_A are the mean and dispersion of RNA and ADT/ATAC data in the NB distribution, respectively.

The ZINB distribution can be parameterized as an NB distribution with an additional coefficient of exit probability, which represents the weight of the mass of the zero-probability point, which can be expressed as:

$$ZINB(X_R | \pi_R, \mu_R, \theta_R) = \pi_R \delta_0(X_R) + (1 - \pi_R) NB(X_R | \mu_R, \theta_R) \quad (24)$$

$$ZINB(X_A | \pi_A, \mu_A, \theta_A) = \pi_A \delta_0(X_A) + (1 - \pi_A) NB(X_A | \mu_A, \theta_A) \quad (25)$$

where π_R and π_A are exit probabilities for RNA and ADT/ATAC data, respectively.

Eventually, the two ZINB loss functions of this model are the refactoring loss functions, which are defined as:

$$L_R = -\log(ZINB(X_R | \pi_R, \mu_R, \theta_R)) \quad (26)$$

$$L_A = -\log(ZINB(X_A | \pi_A, \mu_A, \theta_A)) \quad (27)$$

Kullback-Leibler(KL) Divergence Loss. In the model, we retained the standard KL divergence term to ensure regularization in the total loss function, constraining the potential distribution output by the encoder to be closer to the prior distribution, thereby improving the continuity and interpretability of the latent space, providing structured support for subsequent generation and clustering tasks, and enhancing the model's generalization ability [31]. The KL loss function is expressed as:

$$L_{KL} = D_{KL}(q(z_2 | x) \| p(z_2)) = \frac{1}{2} (1 + \log(\sigma_1^2) - \mu_1^2 - \sigma_1^2) \quad (28)$$

where μ_1 and σ_1 are the mean and standard deviation of feature space Z_1 .

Clustering loss. We use a weighted soft-allocated K-means clustering method [32] to cluster clustered manifold space Z_c , which can generate smooth cluster assignment weights instead of hard assignments to a single cluster center. This can improve the robustness of the model to noise and clustering performance. More importantly, it provides clustering guidance for EM heuristic optimization strategies, optimizes the clustering manifold space, enhances representation ability, and preserves internal information. In this process, the scaled squared Euclidean distance between the eigenvector z_i of Z_c and each of the centricities v_j can be expressed as:

$$d_{ij} = \|z_i - v_j\|^2 \quad (29)$$

The mean of each row is then subtracted from that row to calculate its normalized distance matrix:

$$td_{ij} = d_{ij} - \frac{1}{K} \sum_{j=1}^K d_{ij} \quad (30)$$

The normalized soft-allocation weight q_{ij} is calculated using exponential operation and double normalization, and from this weight the soft-assigned weights matrix Q mentioned above is formed.

$$q_{ij} = \frac{\exp(-td_{ij})}{\sum_{j=1}^K \exp(-td_{ij})}, \quad q_{ij} = \frac{q_{ij}^2}{\sum_{j=1}^K q_{ij}^2} \quad (31)$$

Finally, the weighted soft classification K-means clustering loss is calculated using the weighted squared Euclidean distance of d_{ij} , as follows:

$$L_{\text{wscK-means}} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^K d_{ij} q_{ij} \quad (32)$$

Total loss. By calculating the sum of two reconstructed losses for RNA and ADT/ATAC data, KL loss and clustering loss, the final overall loss is expressed as:

$$L_{\text{total}} = L_R + L_A + \alpha L_{\text{KL}} + \beta L_{\text{wscK-means}} \quad (33)$$

where the weight α of KL loss and the weight of cluster loss β are set at 0.01 and 0.001, respectively, and the clustering loss ($\beta = 0$) is not used in the pre-training stage, and the KL loss is not used in the first half of the pre-training ($\alpha = 0$).

2.6 Evaluate metrics

Adjusted Rand Index (ARI). ARI is a measure of the consistency between clustering results and true labels [33], ranging from -1 to 1, with closer to 1 indicating better clustering. We use the ARI value calculated by the `adjusted_rand_score` function in the sklearn library, which is defined as:

$$ARI = \frac{(RI - E[RI])}{\max(RI) - E[RI]} \quad (34)$$

where RI is the Rand Index, which indicates the proportion of sample pairs whose clustering results are consistent with the real label. $E[RI]$ is the expected value of the Rand index in the case of random clustering, and $\max(RI)$ is the maximum possible Rand index.

Normalized Mutual Information (NMI). NMI is normalized by dividing the mutual information by the average information entropy (or maximum information entropy) [34], so that the range is [0,1]. NMI is expressed as:

$$NMI(U, V) = \frac{2MI(U, V)}{H(U) + H(V)} \quad (35)$$

Where $H(U)$ and $H(V)$ are denoted as the entropy of the predicted label U and the true label V , respectively, representing their respective uncertainties. $MI(U, V)$ is mutual information.

Adjusted Mutual Information (AMI). AMI is the result of adjusting for Mutual Information (MI), which takes into account the impact of randomness on mutual information [35]. The value range of AMI is [-1,1], and the specific formula is expressed as:

$$AMI(U, V) = \frac{MI(U, V) - E[MI(U, V)]}{\max(H(U), H(V)) - E[MI(U, V)]} \quad (36)$$

Where is the mathematical expectation of mutual information, $E[MI(U,V)]$ which is used to remove the influence of random factors on mutual information.

3 Results

3.1 Clustering performance

We evaluated the clustering performance of scMAG on 9 datasets using ARI and NMI clustering indicators, and for RNA and ADT datasets, there were 5 single-batch datasets (SLN111-D1, SLN111-D2, SLN206-D1, SLN206-2, PBMC) and 2 two-batch datasets (SLN111, SLN206), and 5 single-cell multiomics models (BREM-SC, TotalVI, CiteFuse, Seurat, Sputter). For both RNA and ATAC datasets, they were single-batch datasets and were compared to three single-cell multiomics models (Seurat, Cobolt, scMM). In Figure 2 and 3, we use boxplots, histograms, and dot plots to demonstrate the performance of scMAG on clustering. Specifically, scMAG has significant advantages in both stability and clustering performance, particularly in terms of ARI metrics.

As shown in the box plot in Figure 2a, the height of the scMAG box is the highest in both indicators, indicating that there is a high level of accuracy on different datasets. Moreover, the box is also the narrowest, indicating that the clustering results on different datasets have good stability. Combined with Figure 2b, we can see that in the BREM-SC, CiteFuse and Seurat models, there are abnormal clustering results on a certain dataset, and the stability of all the comparison model methods on the ARI index is not good, indicating that their models only pay attention to the global effect and the inter-class clustering effect is stable, but do not pay attention to the local effect, that is, the performance improvement of intra class clustering was not significant. However, scMAG not only greatly improves the ARI index, but also has a certain stability, which truly improves the ability of intra-class clustering. Although the clustering performance improvement on the NMI index is small, it is basically the best result, except for a little deficiency on the SLN111-D2 dataset.

In Figure 3, we give the model comparison dot plots of RNA and ADT datasets and RNA and ATAC datasets, respectively, and we can intuitively see the results of simultaneous comparison of ARI and NMI indicators, and we can see that the clustering performance of scMAG is still outstanding.

3.2 Ablation experiments

Single cell data has high-dimensional characteristics, but its expression is often distributed on low-dimensional manifolds. Therefore, the introduction of manifold fitting not only helps to retain nonlinear structural information, but also enhances the expression of key biological signals, which is conducive to feature extraction and subsequent modeling. In the model design, we believe that there are some limitations in relying on Bayesian inference for data reconstruction, which makes it difficult to completely retain the original cell information and its internal structure, resulting in the loss of potential details. Therefore, this paper does not directly adopt the variational fusion strategy, but first carries out feature fusion, and then introduces a multi-stage feature space fusion mechanism to integrate multi-source information and build a more biologically meaningful manifold space. To maximize the information flow of the bottleneck layer, we constructed independent manifold spaces for decoding reconstruction and dimension reduction clustering respectively: the former adopts a weighted space

fusion strategy to ensure the integrity and accuracy of the reconstruction process [36]; The latter introduces the EM heuristic optimization strategy based on forgetting gate (SER) to enhance the clustering effect and reveal the potential heterogeneity and connections between cells [37].

In order to verify the effectiveness of popular fitting methods, two-stage fusion methods, and their corresponding fusion strategies, we conducted ablation experiments to demonstrate the improvement of model performance by these methods. We have made modular modifications to the scMAG model, including removing the preprocessed manifold fitting, removing the two-stage structure, removing the SER optimization strategy, and removing all designs.

The results of the ablation experiment are shown in Figure 4, which only shows the ablation results of some of the datasets (the ablation experiments for the remaining datasets can be found in Supplementary Material Figure 1). We can clearly observe that for each dataset, only the original model has the highest metric on different indicators, which precisely indicates that our inference is correct. By constructing a data manifold structure and combining two-stage fusion with independent fusion strategies to further preserve detailed information and global structure, the representation ability of the manifold space is enhanced. In the end, we also observed a significant performance improvement in the ARI metric, which not only demonstrates the effectiveness of the scMAG model for intra class clustering of the dataset in this paper, but also demonstrates its effectiveness in improving the model methodology.

3.3 Cluster annotation

As shown in Figure 5, the clustering results of the scMAG on the SLN206 dataset are shown. Through UMAP dimensionality reduction visualization and cell type annotation, it can be seen that the model can not only effectively classify cells, but also maintain the distribution relationship between cell clusters to a certain extent. This indicates that the scMAG model can balance the clustering accuracy and biological correlation well (the UMAP clustering results of the remaining datasets are shown in Supplementary Material Figure 2).

4 Downstream analysis

4.1 De-batch effect

We assessed the debatching effect capacity of scMAG, as shown in Figure 6. As can be seen from the comparison chart, the scMAG significantly reduces the impact of batch effect while achieving high-quality clustering. This indicates that the model is highly robust and adaptable in terms of cross-batch integration, and the clustered manifold space constructed by it has high expressive ability (the clustering, cell type annotation, and debate results of the scMAG on the SLN111 dataset are presented in Supplementary Material Figure 3).

4.2 Marker gene identification

In Figure 7, by comparing the performance of the reconstructed data with the original data, we can observe that the reconstructed data can more significantly express the marker genes of different cell clusters, while the original data performs poorly in characterizing the marker genes. To a large extent, this phenomenon indicates that the data reconstruction process improves the quality of the data and allows for a more accurate

representation of the biological differences between cell clusters. This improvement may be due to the fact that the refactoring method effectively reduces technical noise and enhances the biological signal, thus providing more reliable underlying data support for downstream analysis [38]. This result further validates the importance of data reconstruction in single-cell transcriptome analysis, especially in the potential applications of cell type identification and functional elucidation (the labeled protein map of the SLN206 dataset is shown in Supplementary Material Figure 4).

To verify the biological effectiveness of the model, we analyzed the marker genes and explained them in conjunction with existing biological research, demonstrating that the model can effectively identify gene differences between different cells at the biological level. The results showed that *Bach2* gene was highly expressed in both mature B cells and transition B cells, which was usually related to its important role in B cell differentiation and function regulation. In Mature B cells, it further regulates antibody production and humoral immune responses, allowing cells to respond appropriately when confronted with antigens [39, 40]. *SLC15A3* acts as a proton-coupled oligopeptide transporter and is involved in the regulation of signaling pathways in immune cells, particularly in inflammatory responses. Studies have shown that *SLC15A3* function in endolysosomes and is associated with the conduction of inflammatory signals [41]. And neutrophils show *SLC15A3* high expression levels at homeostasis. *CD24a* is typically expressed higher in transitional B cells, a critical transition period for the differentiation of immature B cells into mature B cells. Studies have shown that *CD24a* expression is higher in the early stages of B cell maturation [42], then gradually decreases as it differentiates into mature antibody-producing cells.

4.3 Genetic difference analysis

The volcano map simultaneously displays the expression changes of each gene, allowing for rapid identification of important differentially expressed genes. Figure 8 (a, b) shows the analysis of multiple genetic differences between mature B cells and CD4 T cells, CD8 T cells and transitional B cells. And there are many proven findings [43-45], such as *Setbp1*, *Pik3ap1*, *Slc15a3*, and *Klhl14* being highly expressed in MatureB cells, and *Cd40lg* and *Cd28* being highly expressed in CD4T cells. *Prkcq*, *Sla2*, *Cd7*, and *Nkg7* are highly expressed in CD8 T cells, while *Vpreb3*, *Eaf2*, and *Iglc1* are highly expressed in transitional B cells.

4.4 Protein differences analysis

The protein heat map is shown in Figure 8c, where we found that different B cell types were clustered together and shared multiple significantly enriched proteins. *ADT_CD44_A0073* is enriched in neutrophil granules, and this protein plays an important role in neutrophil chemotaxis and aggregation. *ADT_IgD_A0571* exhibited the highest enrichment fraction in both Transitional B cells and Mature B cells, reflecting its key function during B cell maturation and differentiation. *ADT_CD5_A0111* is significantly enriched in Tregs, ICOS-high Tregs, *Ilft3*-high CD4 T cells, CD4 T cells, and its role in T cell signaling and immune regulation has been extensively studied and confirmed, and the expression level of CD5 in Tregs is related to its regulatory properties [46], especially in highly autoreactive Tregs.

These analysis results further support the protein expression characteristics in each cell cluster and provide an important basis for understanding the function of immune cells. We also found that some subtype cell clusters

of the same type had identical protein expression between them (the corresponding protein heatmap is shown in Figure 5 of the material), which made the ADT data not useful in post-fusion clustering and could interfere with model clustering. This is also the reason why the model does not have a significant improvement effect on NMI indicators, and cannot better distinguish different cell subtypes, resulting in an insignificant overall clustering effect.

4.5 Trajectory analysis

Figure 8d shows the trajectory of the cell clusters, showing the connections between cells including regulatory T cells, natural killer T cells, marginal band (MZ) B cells, and GD T cells. We have found many proven connections for a variety of cell clusters, which also shows that the latent space not only represents the differences between cell clusters, but also retains the intrinsic connections between clusters. Specifically, between NKT cells29 and B-CD8 T3 cells, CD8⁺ NKT cells express both the T cell markers TCR β and CD3, as well as the NK cell receptors CD49b and NKG2D. However, there is also a link between cD122⁺ CD8 T cells7, ICOS-high Tregs cells12, and CD8 T 9 cells, and naturally occurring CD8(+) CD122(+) T cells are also Tregs cells with the ability to suppress T cell responses and suppress autoimmune and allogeneic immunity [47]. The link between Ifit3-high B14 cells and Tregs33 cells is more likely to show that Ifit3-high B cells can regulate the PI3K/AKT signaling pathway by binding to the JAK-PI3K signaling pathway, which in turn affects the expression of IFN γ and the stability of FOXP3 in Treg cells.

4.6 Enrichment analysis

The differentially expressed gene sequences for some of the cell clusters are shown in Figure 9a (the differentially expressed gene sequences of all cell clusters are shown in Supplementary Material Figure 6). For these differentially expressed genes, we performed a gene ontology (GO) enrichment analysis and demonstrated the enriched biological pathways in Figure 9b.

We conducted Gene Ontology (GO) enrichment analysis on these differentially expressed genes and displayed the enriched functional entries in Figure 8b. The significantly enriched GO terms are mainly distributed in the GO: MF and GO: BP classifications. In the GO: MF (molecular function) category, typical enriched functions include protein binding (GO: 0005515), MHC protein complex binding (GO: 0042288), and immune receptor activity (GO: 0041075), indicating the presence of an active protein interaction network and immune recognition activity in the sample. In addition, the significant enrichment of enzyme regulation related functions (such as GO: 0030234, GO: 0061135) and ribosome structural functions (GO: 0003735) suggests that related genes may be involved in protein translation and metabolic regulation. In the GO: BP (biological processes) category, enriched entries include ribosome subunit biosynthesis (GO: 0042254), protein folding (GO: 0006457), cellular stress response (GO: 0033554), and oxidative phosphorylation (GO: 0006119), reflecting the active involvement of the sample in protein synthesis, energy metabolism, and stress response processes. In addition, immune related processes such as T cell activation (GO: 0041110) were significantly enriched, further emphasizing their potential role in immune regulation. In terms of GO: CC (cellular components), enriched entries include ribosomes (GO: 0005840), cytoplasm (GO: 0005737), and cell surface (GO: 0009986), indicating

that the above functional activities are mainly localized in the cytoplasmic and membrane structural regions, supporting their spatial characteristics in translation, signal transduction, and cellular communication.

In summary, the GO enrichment results reveal that differentially expressed genes in this study are widely involved in core biological processes such as immune regulation, protein synthesis and folding, energy metabolism, and stress resistance, providing a theoretical basis for further exploring their potential functional mechanisms and biological significance [48].

5 Discussion

This article proposes a method for integrating single-cell multi omics data through multi-stage deep fusion and manifold aware gating, aiming to achieve feature dimensionality reduction, batch effect removal, data integration, trajectory inference, and improve data quality while still improving clustering performance. Through experimental research on a large number of single-cell multi omics datasets, it has been proven that scMAG has indeed achieved significant improvements in clustering accuracy through multi-stage deep fusion, and is also superior to similar methods.

This model achieves sparse representation of raw data through multi kernel manifold fitting, minimizing information loss and effectively eliminating noise, thanks to its ability to capture complex nonlinear structures and changing patterns in the data. This method does not rely on specific data distribution assumptions, but maintains the local properties of the data by modeling within local regions, thereby more accurately reflecting the true structural relationships of the data and preserving its specific modal distribution. In addition, to improve the model's generalization ability, random noise was added to the model to complement preprocessing and eliminate noise.

Although variational autoencoders (VAEs) have nonlinear dimensionality reduction and denoising capabilities, they often overlook the preservation of local topological structures when learning latent representations, resulting in the loss of some original information during the mapping process. To overcome this limitation, further preserve the modal distribution and manifold structure, and achieve data heterogeneity alignment, we propose an improved method based on a two-stage fusion strategy. It is a progressive feature integration mechanism, whose core lies in dividing the fusion process of multimodal data into two stages of clear structure and complementary functions, in order to gradually improve the representation ability and task performance of potential space. The first stage focuses on the construction of basic features, preserving the local adjacency relationships and nonlinear structures of the original data to the greatest extent possible through aligned modal concatenation and preservation of manifold structure, providing high-quality input for subsequent information extraction. In the second stage, a differentiated information fusion strategy is introduced based on this, combined with optimization objectives guided by clustering, to further strengthen the interaction and semantic separation ability between features. A progressive relationship is formed between the two stages, with the first stage ensuring structural integrity and the second stage strengthening task expression, thereby achieving efficient, stable, and interpretable latent spatial learning.

On this basis, we combine the forget gate mechanism with the expectation maximization (EM) heuristic optimization strategy to jointly improve the representation quality of latent space in clustering tasks. Specifically,

as a dynamic regulation mechanism, the forget gate can finely screen and retain the transmitted information based on context, thereby enhancing the VAE bottleneck layer's perception and expression ability of key features, effectively suppressing the propagation of redundant and noisy information. At the same time, the EM heuristic strategy introduces a clustering oriented optimization process in the latent space, which iteratively updates between "hidden variable estimation of clustering labels" and "model parameter maximization" to make the model more dependent on cell embedding expression to reconstruct the gene expression matrix. This strategy not only enhances the task guidance role of the bottleneck layer, but also achieves the extraction of biologically meaningful clustering structures from potential distributions.

Through the synergistic effect of these two mechanisms, the model can achieve clearer classification when learning latent space, thereby improving the compactness and consistency of similar cells in the embedding space, while maintaining clearer boundaries between cells of different categories, significantly improving clustering performance and reconstruction quality. Meanwhile, this strategy also helps to achieve better feature decoupling and structural preservation, providing a more stable and interpretable representation space for downstream analysis tasks. In addition, this mechanism has demonstrated good adaptability and generalization ability in the integration of heterogeneous data from multiple omics. It not only helps alleviate the distribution bias between modalities, but also improves the efficiency of cross modal information fusion while ensuring the authenticity of biological structures, providing solid support for potential biometric mining in high-dimensional complex data.

Although the model has shown good performance in feature extraction, structure preservation, and multimodal integration in latent space, there are still some limitations. Firstly, this strategy may face the problem of reduced generalization ability when dealing with highly sparse or extremely imbalanced multi omics data, especially in cases where certain modal signals are extremely weak or missing, and the expression of the latent space may deviate. Secondly, the model introduces additional regularization terms and gating structures when constructing collaborative mechanisms, which may make the training process more complex, parameter selection sensitive, and require relatively high computational resources. In addition, in the context of high-dimensional low sample, the trade-off between feature decoupling and distribution preservation still needs to be further optimized to prevent model overfitting or misleading modeling of potential structures. Finally, due to the incomplete interpretability of the integrated potential space, its biological significance still needs to be further explored through practical annotation or experimental verification, which limits its direct promotion in certain clinical translational applications.

6 Conclusions

This article proposes a method for integrating single-cell multi omics data through multi-level deep fusion and manifold aware gating (scMAG), providing an effective research tool for heterogeneous omics integration and clustering analysis. This method uses manifold regularization to preserve the data distribution and manifold structure, effectively enhancing the spatial distribution of the data. This is beneficial for multi-layer perceptrons to fit distribution parameters and fully model ZINB. By combining a two-stage deep fusion framework and a selection entropy optimization strategy, combined with hierarchical guidance and fusion of information flow, the

bottleneck layer information flow is maximized, which not only enhances the adaptability of the model to complex data structures, but also further enhances the discriminative and expressive ability of features in the latent space. scMAG maximizes the balance between data distribution and alignment, not only improving the quality of reconstructed data while enhancing clustering performance, but also meeting the various requirements of single-cell clustering models for spatial alignment, de batching effects, and preserving connections between data. Ultimately, the multi task applicability of scMAG was demonstrated through multiple downstream tasks, and important biological significance was discovered, providing more possibilities for biomedical analysis.

List of abbreviations

VAE: Variational Autoencoder
ZINB: Zero-Inflated Negative Binomial
SER: Selective Entropy Optimization
ARI: Adjusted Rand Index
NMI: Normalized Mutual Information
AMI: Adjusted Mutual Information
KL: Kullback – Leibler Divergence

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and materials:

7 sets of RNA and ADT datasets were used in this paper, among which the PBMC dataset was available on the 10X website (<https://www.10xgenomics.com/datasets>), and the cell type labels is available from GitHub of Specter (<https://github.com/canzarlab/Specter>). There were also cells from the spleen and lymph nodes of two wild-type mice (biological replicates), which were stained with 111 antibodies or 208 antibodies for each mouse over two days, yielding 4 sets of datasets (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>, GSE150599), and the two sets of 111 antibody staining data were recombined into a new dataset SLN111, and the two groups of 208 antibody staining data were recombined into a new dataset SLN206. Its cell type labels is available on TotalVI (https://github.com/Yosef-Lab/totalVI_reproducibility) on GitHub. 2 sets of RNA and ATAC datasets were also used. PBMCs of human peripheral blood mononuclear cells were expressed as pbmc3k and pbmc10k, respectively, which can be obtained from the 10X genomics website (<https://www.10xgenomics.com/datasets>), and the cell labels are transferred by Signac (v1.4.0) from the annotated datasets.

Competing interests: The authors declare that they have no competing interests.

Funding: This research received no external funding.

Author Contributions: This article was independently conceived and written by the author J.H. Zou, with academic guidance provided by mentor J.Q. Liu through regular discussions.

Acknowledgements: Not applicable.

References

1. Golomb S M, Guldner I H, Zhao A, et al. Multi-modal single-cell analysis reveals brain immune landscape plasticity during aging and gut microbiota dysbiosis. *Cell reports*, 2020, 33(9).
2. Zhang Z. Integrative Analysis of Single Cell Multi-Omics Data. Princeton University, 2022.
3. Miao Z, Humphreys B D, McMahon A P, et al. Multi-omics integration in the age of million single-cell data. *Nature Reviews Nephrology*, 2021, 17(11): 710-724.
4. Vandereyken K, Sifrim A, Thienpont B, et al. Methods and applications for single-cell and spatial multi-omics. *Nature Reviews Genetics*, 2023, 24(8): 494-515.
5. Li J, Zhang Y, Yang C, et al. Discrepant mRNA and protein expression in immune cells. *Current genomics*, 2020, 21(8): 560-563.
6. Stoekius M, Hafemeister C, Stephenson W, et al. Simultaneous epitope and transcriptome measurement in single cells. *Nature methods*, 2017, 14(9): 865-868.
7. Jiang R, Sun T, Song D, et al. Statistics or biology: the zero-inflation controversy about scRNA-seq data. *Genome biology*, 2022, 23(1): 31.
8. Mulè M P, Martins A J, Tsang J S. Normalizing and denoising protein expression data from droplet-based single cell profiling. *Nature communications*, 2022, 13(1): 2099.
9. Stanojevic S, Li Y, Ristivojevic A, et al. Computational methods for single-cell multi-omics integration and alignment. *Genomics, Proteomics & Bioinformatics*, 2022, 20(5): 836-849.
10. Wang L, Zhang H, Yi B, et al. FactVAE: a factorized variational autoencoder for single-cell multi-omics data integration analysis. *Briefings in Bioinformatics*, 2025, 26(2): bbaf157.

11. He Z, Hu S, Chen Y, et al. Mosaic integration and knowledge transfer of single-cell multimodal data with MIDAS. *Nature biotechnology*, 2024, 42(10): 1594-1605.
12. Minoura K, Abe K, Nam H, et al. A mixture-of-experts deep generative model for integrated analysis of single-cell multiomics data. *Cell reports methods*, 2021, 1(5).
13. Gayoso A, Steier Z, Lopez R, et al. Joint probabilistic modeling of single-cell multi-omic data with totalVI. *Nature methods*, 2021, 18(3): 272-282.
14. Luo E, Liu Q, Hao M, et al. Multi-modal Diffusion Model with Dual-Cross-Attention for Multi-Omics Data Generation and Translation. *bioRxiv*, 2025: 2025.02. 27.640020.
15. Cao K, Gong Q, Hong Y, et al. uniPort: a unified computational framework for single-cell data integration with optimal transport. *bioRxiv*, 2022: 2022.02. 14.480323.
16. Tang L. Optimal transport for single-cell genomics. *Nature Methods*, 2025, 22(3): 452-452.
17. Wang G, Zhao J, Lin Y, et al. scMODAL: a general deep learning framework for comprehensive single-cell multi-omics data alignment with feature links. *Nature Communications*, 2025, 16(1): 1-13.
18. Yang X, Mann K K, Wu H, et al. scCross: a deep generative model for unifying single-cell multi-omics with seamless integration, cross-modal generation, and in silico exploration. *Genome Biology*, 2024, 25(1): 198.
19. Falorsi L, De Haan P, Davidson T R, et al. Explorations in homeomorphic variational auto-encoding. *arXiv preprint arXiv:1807.04689*, 2018.
20. Yao Z, Li B, Lu Y, et al. Single-cell analysis via manifold fitting: A framework for RNA clustering and beyond. *Proceedings of the National Academy of Sciences*, 2024, 121(37): e2400002121.
21. Jain M S, Polanski K, Conde C D, et al. MultiMAP: dimensionality reduction and integration of multimodal data. *Genome biology*, 2021, 22: 1-26.
22. Yao Z, Li B, Lu Y, et al. Single-cell analysis via manifold fitting: A framework for RNA clustering and beyond. *Proceedings of the National Academy of Sciences*, 2024, 121(37): e2400002121.
23. Weston J, Watkins C. Multi-class support vector machines. Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London, May, 1998.
24. Schölkopf B, Smola A J. Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press, 2002.
25. Tenenbaum J B, Silva V, Langford J C. A global geometric framework for nonlinear dimensionality reduction. *science*, 2000, 290(5500): 2319-2323.
26. Andrew A M. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods by Nello Cristianini and John Shawe-Taylor, Cambridge University Press, Cambridge, 2000, xiii+ 189 pp., ISBN 0-521-78019-5 (Hbk,£ 27.50). *Robotica*, 2000, 18(6): 687-689.
27. Aitchison J. The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1982, 44(2): 139-160.
28. Yao Z, Su J, Li B, et al. Manifold Fitting. *arXiv preprint arXiv:2304.07680*, 2023.
29. Kingma D P, Welling M. Auto-encoding variational bayes.(2013-12-20).
30. Eraslan G, Simon L M, Mircea M, et al. Single-cell RNA-seq denoising using a deep count autoencoder. *Nature communications*, 2019, 10(1): 390.
31. Pereyra G, Tucker G, Chorowski J, et al. Regularizing neural networks by penalizing confident output distributions. *arXiv preprint arXiv:1701.06548*, 2017.
32. Zhu B, Bedeer E, Nguyen H H, et al. Improved soft-k-means clustering algorithm for balancing energy consumption in wireless sensor networks. *IEEE Internet of Things Journal*, 2020, 8(6): 4868-4881.
33. Steinley D, Brusco M J, Hubert L. The variance of the adjusted Rand index. *Psychological methods*, 2016, 21(2): 261.
34. Mahmoudi A, Jemielniak D. Proof of biased behavior of Normalized Mutual Information. *Scientific Reports*, 2024, 14(1): 9021.
35. Schütze H, Manning C D, Raghavan P. Introduction to information retrieval. Cambridge: Cambridge University Press, 2008.
36. Alemi A A, Fischer I, Dillon J V, et al. Deep variational information bottleneck. *arXiv preprint arXiv:1612.00410*, 2016.
37. Jiang Z, Zheng Y, Tan H, et al. Variational deep embedding: An unsupervised and generative approach to clustering. *arXiv preprint arXiv:1611.05148*, 2016.
38. Golub T R, Slonim D K, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *science*, 1999, 286(5439): 531-537.
39. Muto A, Ochiai K, Kimura Y, et al. Bach2 represses plasma cell gene regulatory network in B cells to promote antibody class switch. *The EMBO journal*, 2010, 29(23): 4048-4061.
40. Hu Q, Xu T, Zhang W, et al. Bach2 regulates B cell survival to maintain germinal centers and promote B cell memory. *Biochemical and biophysical research communications*, 2022, 618: 86-92.
41. Yu S, Yang J, Zhang R, et al. SLC15A3 is transcriptionally regulated by HIF1 α and p65 to worsen neuroinflammation in experimental ischemic stroke. *Molecular Neurobiology*, 2024: 1-16.
42. Mensah F F K, Armstrong C W, Reddy V, et al. CD24 expression and B cell maturation shows a novel link with energy metabolism: potential implications for patients with myalgic encephalomyelitis/chronic fatigue syndrome. *Frontiers in immunology*, 2018, 9: 2421.
43. Troutman T D. B-Cell Adapter for Phosphoinositide 3-Kinase Is a Signaling Adapter in the Toll-Like Receptor/Interleukin-1 Receptor Superfamily. 2014.
44. Ham H, Hirdler J B, Bihnam D T, et al. Lysosomal NKG7 restrains mTORC1 activity to promote CD8+ T cell durability and tumor control. *Nature Communications*, 2025, 16(1): 1628.
45. Lelliott E J, Ramsbottom K M, Dowling M R, et al. NKG7 enhances CD8+ T cell synapse efficiency to limit inflammation. *Frontiers in immunology*, 2022, 13: 931630.
46. Tung J W, Kunnavatana S S, Herzenberg L A, et al. The regulation of CD5 expression in murine T cells. *BMC molecular biology*, 2001, 2: 1-13.

47. Liu J, Chen D, Nie G D, et al. CD8+ CD122+ T-cells: a newly emerging regulator with central memory cell phenotypes. *Frontiers in immunology*, 2015, 6: 494.
48. Ashburner M, Ball C A, Blake J A, et al. Gene ontology: tool for the unification of biology. *Nature genetics*, 2000, 25(1): 25-29.

Figure legends

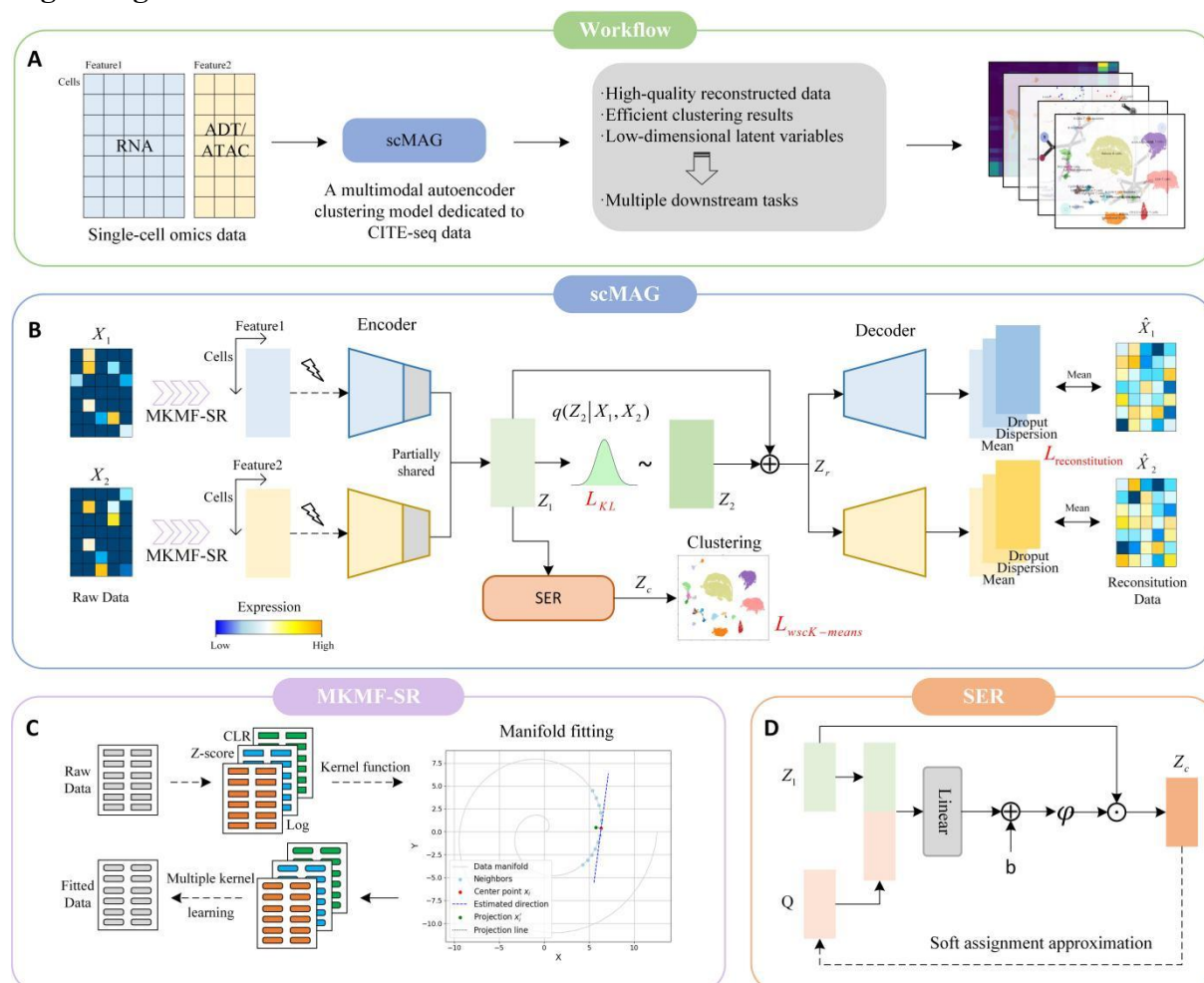
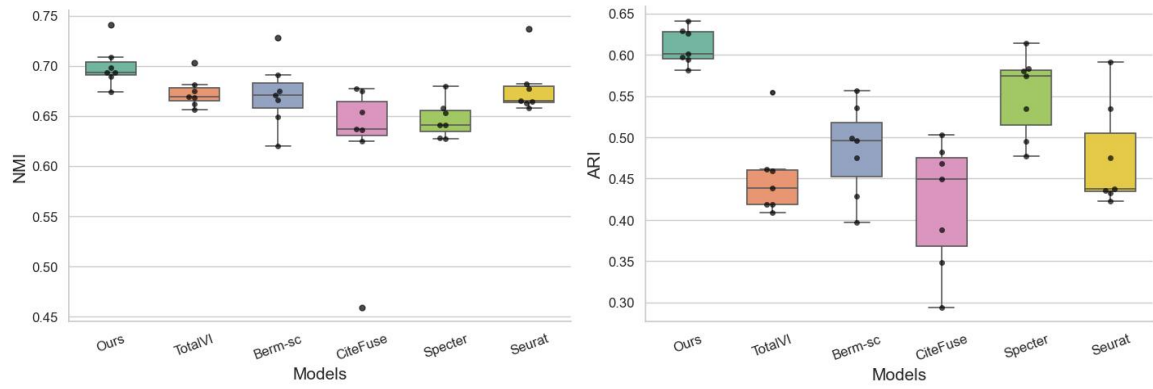


Fig.1 The scMAG framework for the analysis of Single-cell omics data. (A) First, the single-cell multimodal data are input into the scMAG model, and then the reconstructed data, clustering results and low dimensional manifold space are obtained. Finally, the multi downstream task analysis is carried out. (B) Schematic of the scMAG model: RNA counts, protein counts and ATAC peaks for each cell n are jointly transformed by an encoder neural network into the parameters of the posterior distributions. The scMAG model effectively integrates single-cell multi-omics data through manifold fitting technology(C) and Selective Entropy Optimization (SER) strategy (D), achieves high-quality data reconstruction and clustering.

A



B

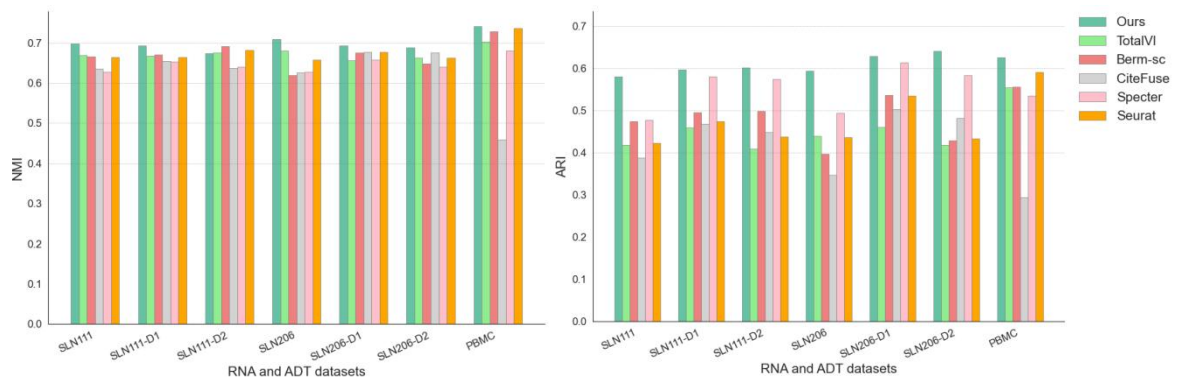


Fig.2 scMAG was compared with the other 5 methods, boxplot (A), histogram (B). This reflects that scMAG has good stability and excellent clustering ability, and the ARI index has been greatly improved, indicating that the intra-class clustering ability has been further improved.

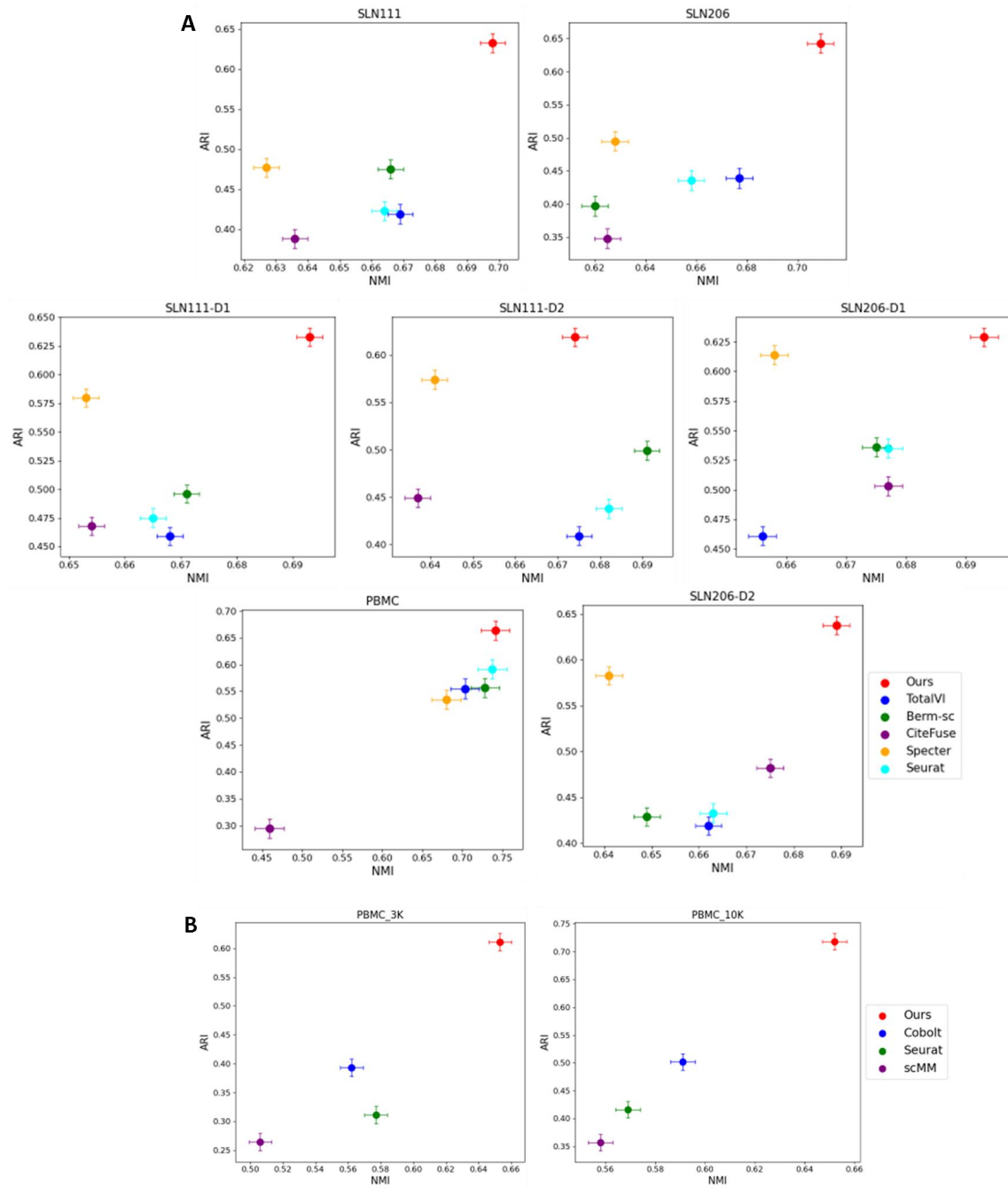


Fig.3 Model comparison dot plot on RNA and ADT datasets (A), model comparison dot plot on RNA and ATAC datasets(B).

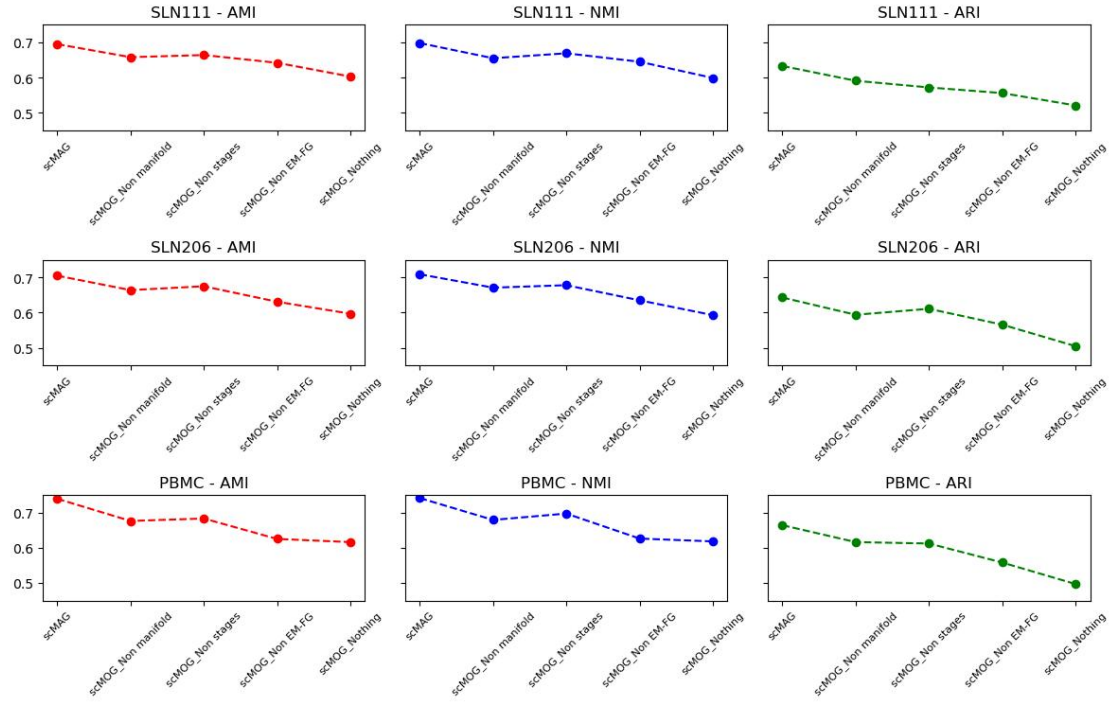


Fig.4 Ablation experiments of 5 modes on the SLN206, SLN111, and PBMC datasets.

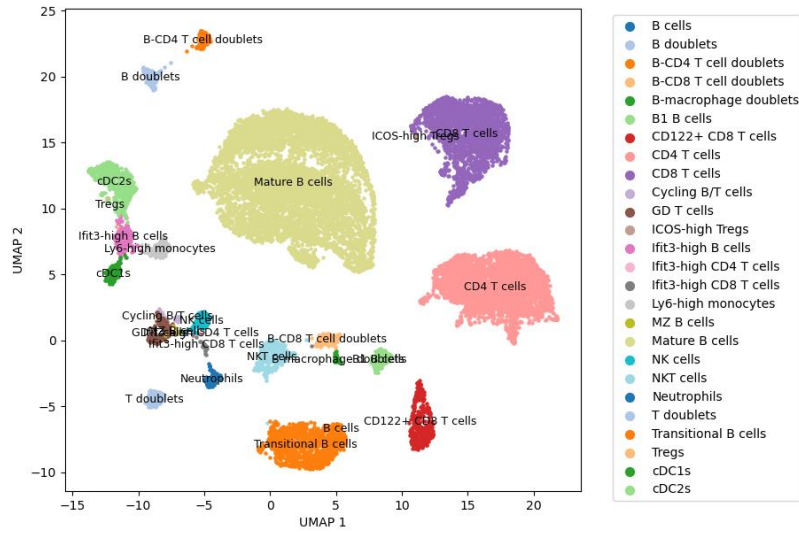


Fig.5 The clustering and labeling on the SLN206 dataset are clear, and there is a certain relationship between the clusters.

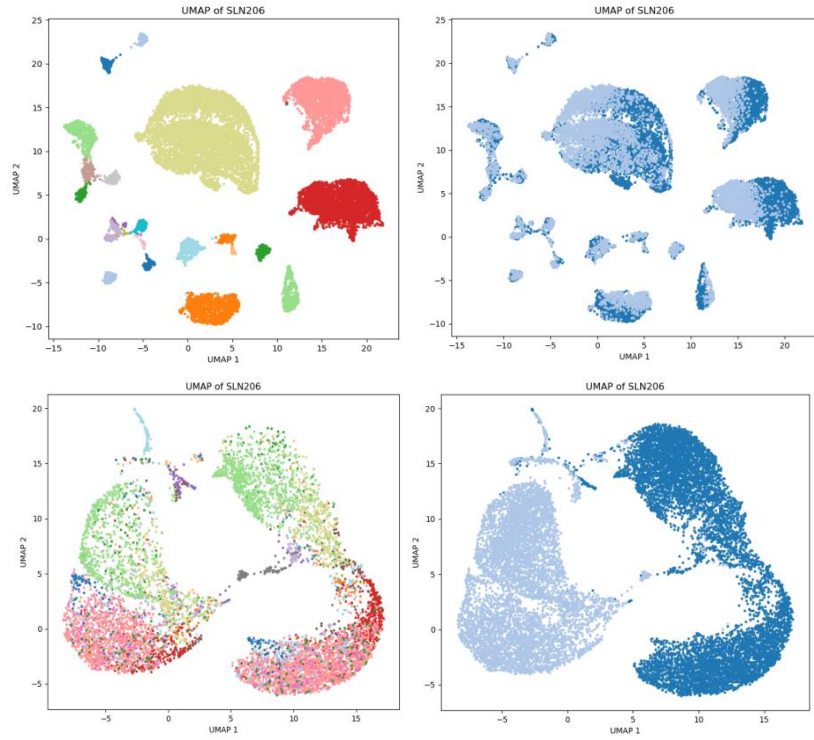


Fig.6 Debatching task on the SLN206 dataset. The first row is the clustered manifold space of the scMAG model, the second row is the raw data, the first column is the clustering results, and the second column is the batch results.

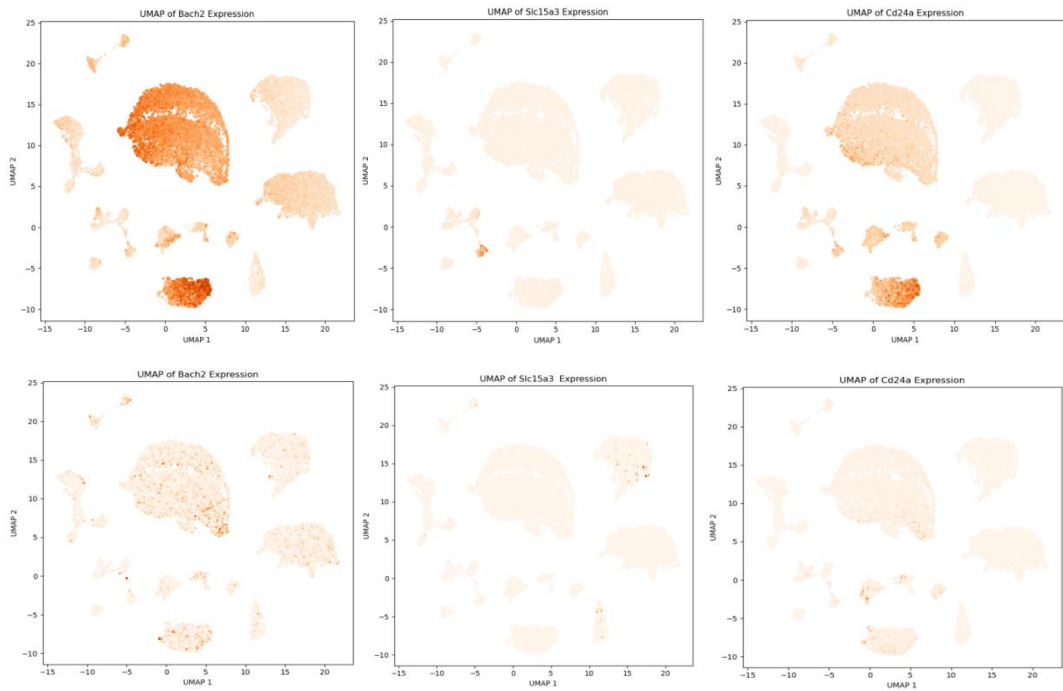


Fig.7 Comparison of marker genes on the SLN206 dataset. The first row reconstructs the data, and the second row is the raw data, with each column being a different marker gene.

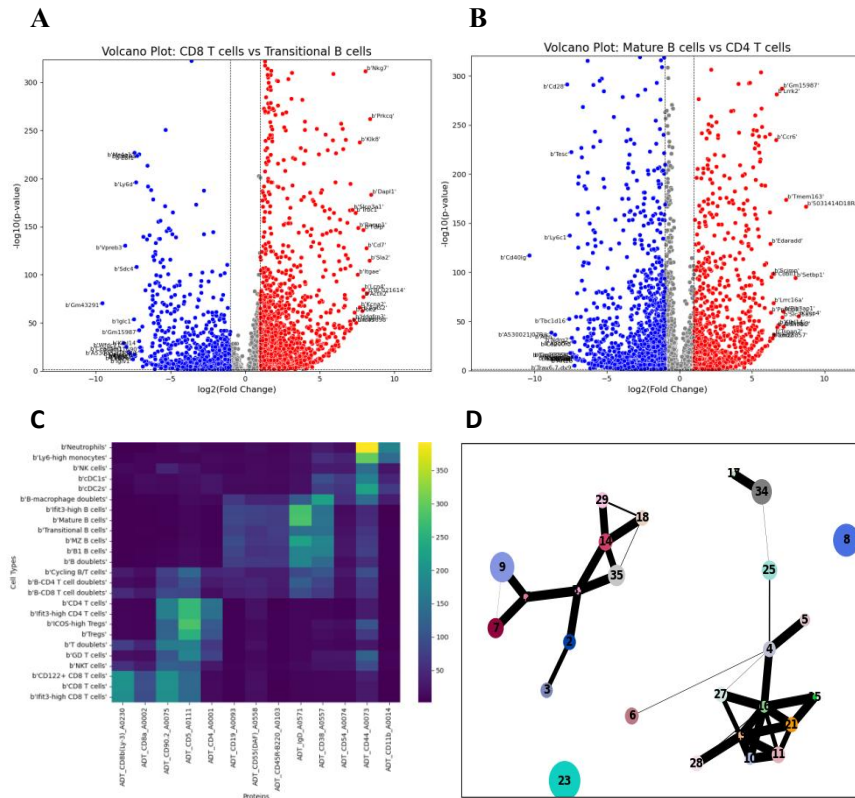
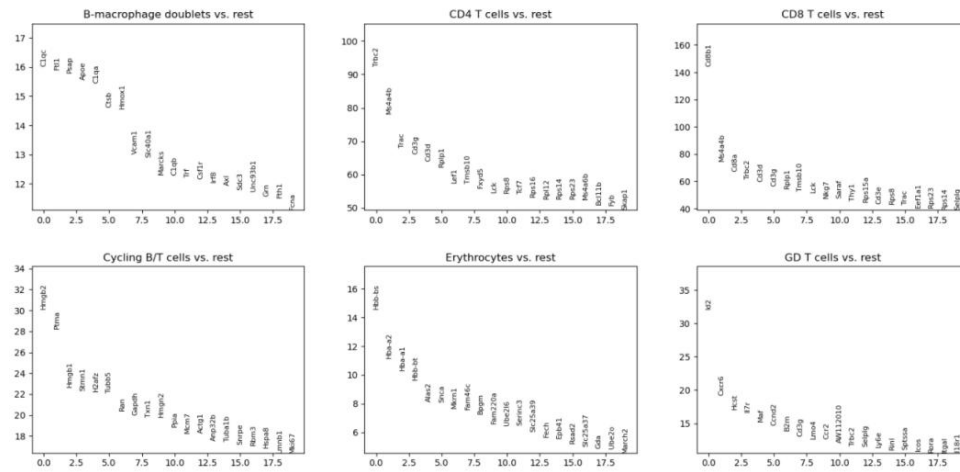
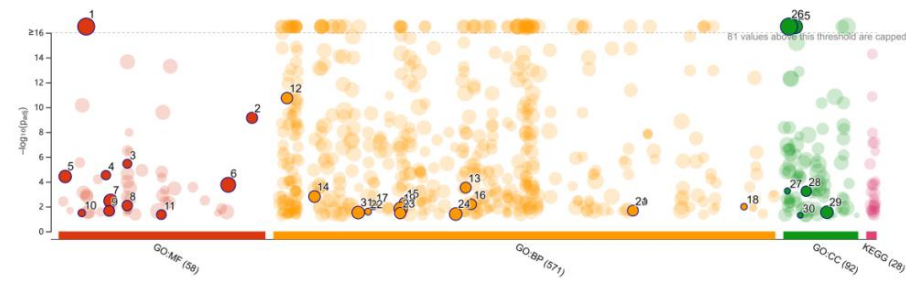


Fig.8 The volcano plot (A, B) shows the gene differences between the different clusters, the protein heat map (C) shows the protein expression differences between the different clusters, and the cell cluster trajectory map (D) shows the potential connections between the different clusters.

A



B



ID	Source	Term ID	Term Name	Padj (query, 1)
1	GO:MF	GO:0005515	protein binding	3.712×10^{-35}
2	GO:MF	GO:0140375	immune receptor activity	7.045×10^{-18}
3	GO:MF	GO:0019865	immunoglobulin binding	3.521×10^{-5}
4	GO:MF	GO:0015453	oxidoreduction-driven active transmembrane tran...	2.900×10^{-5}
5	GO:MF	GO:0003735	structural constituent of ribosome	3.697×10^{-5}
6	GO:MF	GO:0098772	molecular function regulator activity	1.752×10^{-4}
7	GO:MF	GO:0016491	oxidoreductase activity	3.466×10^{-4}
8	GO:MF	GO:0019843	rRNA binding	5.277×10^{-4}
9	GO:MF	GO:0016209	antioxidant activity	2.206×10^{-4}
10	GO:MF	GO:0005031	tumor necrosis factor receptor activity	3.245×10^{-4}
11	GO:MF	GO:0042056	chemoattractant activity	4.432×10^{-4}
12	GO:BP	GO:0002181	cytoplasmic translation	1.815×10^{-11}
13	GO:BP	GO:0042274	ribosomal small subunit biogenesis	2.924×10^{-4}
14	GO:BP	GO:0006457	protein folding	1.537×10^{-4}
15	GO:BP	GO:0030574	collagen catabolic process	4.476×10^{-4}
16	GO:BP	GO:0042773	ATP synthesis coupled electron transport	6.513×10^{-4}
17	GO:BP	GO:0018119	peptidyl-cysteine S-nitrosylation	6.399×10^{-4}
18	GO:BP	GO:1905686	positive regulation of plasma membrane repair	1.010×10^{-4}
19	GO:BP	GO:0030036	actin cytoskeleton organization	1.463×10^{-4}
20	GO:BP	GO:0098883	synapse pruning	1.789×10^{-4}
21	GO:BP	GO:0098869	cellular oxidant detoxification	2.087×10^{-4}
22	GO:BP	GO:0015990	electron transport coupled proton transport	2.504×10^{-4}
23	GO:BP	GO:0022900	electron transport chain	3.322×10^{-4}
24	GO:BP	GO:0036293	response to decreased oxygen levels	5.963×10^{-4}
25	GO:CC	GO:0009986	cell surface	1.987×10^{-36}
26	GO:CC	GO:0005737	cytoplasm	3.179×10^{-36}
27	GO:CC	GO:0005602	complement component C1 complex	5.438×10^{-4}
28	GO:CC	GO:0032040	small-subunit processome	5.742×10^{-4}
29	GO:CC	GO:0062023	collagen-containing extracellular matrix	2.847×10^{-4}
30	GO:CC	GO:0030142	COPI-coated Golgi to ER transport vesicle	4.993×10^{-4}
31	GO:BP	GO:0010959	regulation of metal ion transport	2.000×10^{-4}

Fig.9 Gene sequences of different clusters (A), GO enrichment analysis (B).