




# Modeling and analyzing single-cell multimodal data with deep parametric inference

Huan Hu, Zhen Feng, Hai Lin, Junjie Zhao, Yaru Zhang, Fei Xu, Lingling Chen, Feng Chen, Yunlong Ma , Jianzhong Su , Qi Zhao  and Jianwei Shuai 

Corresponding authors. Jianwei Shuai, Department of Physics, and Fujian Provincial Key Laboratory for Soft Functional Materials Research, Xiamen University, Xiamen 361005, China. E-mail: [jianweishuai@xmu.edu.cn](mailto:jianweishuai@xmu.edu.cn); Qi Zhao, School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, China. E-mail: [zhaoqi@lnu.edu.cn](mailto:zhaoqi@lnu.edu.cn)

## Abstract

The proliferation of single-cell multimodal sequencing technologies has enabled us to understand cellular heterogeneity with multiple views, providing novel and actionable biological insights into the disease-driving mechanisms. Here, we propose a comprehensive end-to-end single-cell multimodal analysis framework named Deep Parametric Inference (DPI). DPI transforms single-cell multimodal data into a multimodal parameter space by inferring individual modal parameters. Analysis of cord blood mononuclear cells (CBMC) reveals that the multimodal parameter space can characterize the heterogeneity of cells more comprehensively than individual modalities. Furthermore, comparisons with the state-of-the-art methods on multiple datasets show that DPI has superior performance. Additionally, DPI can reference and query cell types without batch effects. As a result, DPI can successfully analyze the progression of COVID-19 disease in peripheral blood mononuclear cells (PBMC). Notably, we further propose a cell state vector field and analyze the transformation pattern of bone marrow cells (BMC) states. In conclusion, DPI is a powerful single-cell multimodal analysis framework that can provide new biological insights into biomedical researchers. The python packages, datasets and user-friendly manuals of DPI are freely available at <https://github.com/studentiz/dpi>.

**Keywords:** multi-omics, data integration, deep learning, single-cell, COVID-19

## Introduction

The advances in the quantitative, high-throughput measurement of single-cell molecular composition are solving elusive biological and medical problems [1–4]. Increasingly, single-cell multimodal sequencing techniques are available to further improve our understanding of cellular function by profiling multiple distinct omics views. For example, it is now possible to simultaneously measure transcriptome and chromatin accessibility, spatial

location of cells in tissues, DNA methylation, and nucleosome occupancy. These single-cell multimodal sequencing technologies can not only reduce the differences between omics experiments, but also reveal the heterogeneous cell functions more comprehensively.

Since the outbreak of COVID-19, more and more researchers have applied single-cell multimodal techniques to analyze disease progression [5, 6]. It is well known that disease progression

---

**Huan Hu** is a PhD student at Department of Physics, Xiamen University. His research interests include bioinformatics, single cell data analysis, deep learning and complex network algorithm.

**Zhen Feng**, PhD, is a lecturer at the First Affiliated Hospital of Wenzhou Medical University. His research interests include bioinformatics, computer-aided diagnosis and quantum machine learning.

**Feng Cheng** is a graduate student at Department of Physics, Xiamen University. His research interests include single cell data analysis and deep learning.

**Hai Lin**, PhD, is an assistant research scientist in Wenzhou Institute, University of Chinese Academy of Sciences. His research interests include complex system, machine learning and single-cell data analysis.

**Junjie Zhao** is a graduate student at Cyberspace Institute of Advanced Technology, Guangzhou University. His research interests include single cell data analysis and deep learning.

**Yaru Zhang** is a PhD student at the School of Biomedical Engineering, Wenzhou Medical University. Her research interests include bioinformatics and integrative genomics on cancer and other complex diseases.

**Fei Xu** is a PhD student at Department of Physics, Xiamen University. His research interests include mathematics and complex network algorithm.

**Lingling Chen** is a graduate student at Department of Physics, Xiamen University. Her research interests include bioinformatics, single cell data analysis.

**Yunlong Ma**, PhD, is an associate professor at the School of Biomedical Engineering, Wenzhou Medical University. His research interests include bioinformatics and integrative genomics on cancer and other complex diseases.

**Jianzhong Su**, PhD, is a professor at Wenzhou Institute, University of Chinese Academy of Sciences. His research interests in bioinformatics analysis of single-cell sequencing data.

**Qi Zhao**, PhD, is a professor at School of Computer Science and Software Engineering, University of Science and Technology Liaoning. His research interests include bioinformatics, complex network and machine learning.

**Jianwei Shuai**, PhD, is a professor at department of Physics, Xiamen University. He is the deputy dean of National Institute for Data Science in Health and Medicine, and State Key Laboratory of Cellular Stress Biology, Innovation Center for Cell Signaling Network. He is also the director of Wenzhou Institute, University of Chinese Academy of Sciences and Oujian Laboratory (Zhejiang Lab for Regenerative Medicine, Vision and Brain Health). His research interests include biophysics, deep learning and bioinformatics.

**Received:** September 6, 2022. **Revised:** December 11, 2022. **Accepted:** January 2, 2023

© The Author(s) 2023. Published by Oxford University Press. All rights reserved. For Permissions, please email: [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

is closely related to cell surface proteins of immune cells. In recent years, a class of extended single-cell sequencing has been developed to simultaneously measure the transcriptome information and linked cell surface protein abundance [7–9]. CITE-seq and REAP-seq are representatives of this technique, and have similar experimental protocols [7, 8]. They utilize oligonucleotide-conjugated antibodies to simultaneously quantify RNA and surface protein abundance in single cells by sequencing antibody-derived tags. As the extension of CITE-seq, ASAP-seq [10, 11], DOGMA-seq [10] and ECCITE-seq [12] enable simultaneous determination of chromatin accessibility and CRISPR screening [13].

The rapid development of single-cell multimodal sequencing experimental technologies, led by CITE-seq, poses a challenge to the computational framework. Earlier CITE-Seq data methods focused on the analysis of only one modality, with the other modality being overlaid in the context. This analytical approach is biased towards one modality and is inefficient in utilizing other modalities.

In recent years, several methods have been proposed to combine multiple modalities. They can be divided into two types: machine learning-based and deep learning-based models. The former include SeuratV3 (2019) [14], MOFA+ (2020) [15], BREM-SC (2020) [16], Schema (2021) [17], SeuratV4 (2021) [18], CITEMO (2022) [19], etc. The latter include TotalVI (2021) [20], Multigrade (2022) [21], UMINT (2022) [22], GLUE (2022) [23], etc. The goal of both types of models is to build a low-dimensional embedding that represents single-cell multimodal data. Many studies have shown that low-dimensional embeddings can be competent for a series of downstream analyses such as cell clustering, visualization and quasi-sequential analysis.

However, the existing models mainly focus on the features of low-dimensional representations and ignore the biological properties of data. As a fact, the distribution is one of the most important properties of data. For example, previous studies have shown that RNA data measured by CITE-seq follow a negative binomial (NB) distribution [24], while protein data follow a Poisson distribution [20, 24]. Differences in the distributions of RNA and protein data also reflect differences in their biological functions. Considering that the low-dimensional embedding represents the biological properties of single cells, it is also necessary to model the distribution of low-dimensional embedding. A hard part of modeling a distribution is to obtain the parameters of distributions.

In this study, we design a generative model that can automatically infer the parameters of a distribution, which is called deep parameter inference (DPI). It is a single-cell multimodal integration framework that simultaneously models and integrates each modality into a multimodal parameter space. We use the CBMC dataset to investigate the performance of DPI and find that the multimodal parameter space represents a more comprehensive cellular heterogeneity than RNA and protein embeddings. Comparisons with the state-of-the-art methods also show that DPI has superior and efficient performance. Then, we apply DPI to the analysis of COVID-19 disease progression data and find that the multimodal parameter space can ignore batch effects of samples and serve as a reference for cell type annotation. Furthermore, the multimodal parameter space can perform cell state vector field analysis to reveal the effects of changes in genes and proteins on cell states. In conclusion, DPI is a powerful single-cell multimodal analysis framework, which not only reveals cellular heterogeneity more comprehensively, but also facilitates biomedical disease research.

## Material and methods

### Datasets and data preprocessing

Six publicly available datasets are introduced in this work, including CBMC, COVID-19 PBMC, BMC, PBMC5k, PBMC10k and MALT10k (Supplementary Material S1 available online at <http://bib.oxfordjournals.org/>). In the DPI pipeline, the preprocessed RNA and protein data are represented by  $X_{\text{scaled\_RNA}}$  and  $X_{\text{scaled\_protein}}$ , respectively, (Supplementary Material S1 available online at <http://bib.oxfordjournals.org/>). All datasets can be accessed on our project website (<https://github.com/studentiz/dpi/tree/main/data>), where 2D coordinates for UMAP visualization are also included. We recommend using DPI to analyze and visualize these datasets.

### The DPI model

DPI is a comprehensive single-cell multimodal analysis framework that includes a range of functions from data preprocessing to downstream analysis. The current version of DPI is specifically designed for CITE-Seq and REAP-Seq data. Considering that RNA and protein have different biological properties, DPI preprocesses RNA and protein data separately (Figure 1A). The DPI model has three sub-models: RNA parameter inference model, protein parameter inference model and multimodal parameter inference model.

The RNA data are fed into the RNA parameter inference model, and transformed into the RNA latent space (Figure 1B). The latent space is a normally distributed space, which is the embedding space of cell features. DPI introduces variational autoencoder to construct the RNA and protein latent space, respectively, based on the inferred mean and variance.

The parameters of the RNA and protein parameter spaces are mixed with multimodal parameter inference model to generate a multimodal parameter space. Similar to RNA and protein latent spaces, the multimodal parameter space is also a normally distributed space. The multimodal parameter space represents the comprehensive cellular heterogeneity that covers the features of the RNA and protein latent spaces.

DPI performs downstream analysis tasks such as cell clustering, visualization, reference and query of cell types, and cell state vector fields based on cell embeddings in multimodal parameter space (Figure 1C). The RNA and protein latent spaces are used to reconstruct the distributions of RNA and protein data (Figure 1C). Among them, RNA data are assumed to be NB distribution and protein data are assumed to be Poisson distribution. Their distribution parameters are also inferred by the neural network. The assumption of data distribution needs to be modified if DPI is used for other forms of multimodal data.

### RNA parameter inference model

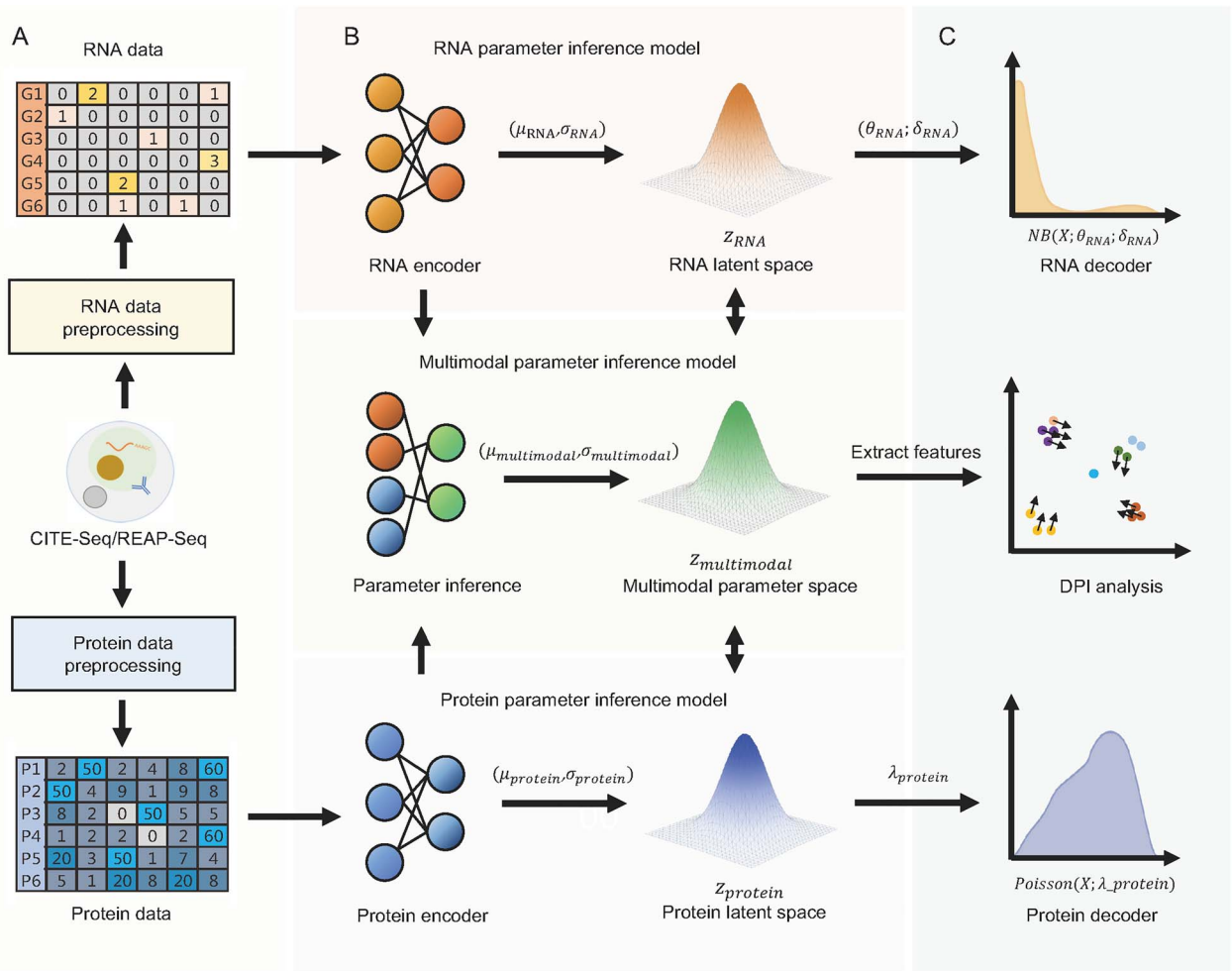
The RNA parameter inference model is specially designed for RNA data features, which has two objectives: (i) to construct a canonical data space to recapitulate RNA heterogeneity and (ii) to infer the parameters of the RNA data distribution to reconstruct noise-free RNA data.

To achieve objective (i), we design an asymmetric autoencoder, given as follows:

$$X_{\text{RNA}} = \text{Input}(X_{\text{scaled\_RNA}}) \quad (1)$$

$$X_{\text{encoded\_RNA}} = \text{Encoder}(X_{\text{RNA}}) \quad (2)$$

$$\mu_{\text{RNA}} = \text{Linear}(X_{\text{encoded\_RNA}}) \quad (3)$$



**Figure 1.** Overview of DPI in three steps. **(A)** First, DPI preprocesses RNA and protein data separately. **(B)** Second, RNA and protein data are encoded by RNA and protein parameter inference models, respectively, to generate RNA and protein latent spaces. A multimodal parameter space is generated from the parametric features provided by the RNA and protein parameter inference models. **(C)** Third, the features of the multimodal parameter space are used for downstream analysis of DPI. The RNA and protein latent spaces reconstruct the data distribution of RNA and protein.

$$\sigma_{RNA} = \text{Linear}(X_{\text{encoded\_RNA}}) \quad (4)$$

$$z_{RNA} = \mu_{RNA} + N(0, I) e^{\frac{\sigma_{RNA}}{2}} \quad (5)$$

$$X'_{RNA} = \text{Linear}(z_{RNA}) \quad (6)$$

$X_{\text{scaled\_RNA}}$  is the data source for the RNA parameter inference model, which is renamed as  $X_{RNA}$  (Eq. (1)). The dimensionality of  $X_{RNA}$  is consistent with the feature dimensionality of RNA data. It is encoded as  $X_{\text{encoded\_RNA}}$  by a classical encoder (a multi-layer neural network structure with gradually decreasing dimension) (Eq. (2)) [25–27]. By default,  $X_{\text{encoded\_RNA}}$  consists of three-layer neural network, with 1024, 256 and 128 neurons, respectively. According to the principle of classical variational autoencoder (VAE),  $\mu_{RNA}$  and  $\sigma_{RNA}$  can be generated from  $X_{\text{encoded\_RNA}}$  by a simple linear transformation [28], which are the mean and variance parameters for constructing the RNA latent space, respectively (Eqs (3) and (4)). By default, both  $\mu_{RNA}$  and  $\sigma_{RNA}$  are 128-dimensional neural network layers.

Next, a normal distribution space  $z_{RNA}$  can be constructed according to  $\mu_{RNA}$ ,  $\sigma_{RNA}$  and standard normal distribution  $N(0, I)$  (Eq. (5)) [28].  $z_{RNA}$  is the latent space of the RNA parameter inference model, which is considered to have the effective information

of  $X_{RNA}$ . By default,  $z_{RNA}$  is a 128-dimensional neural network layer. The classic VAE restores  $z_{RNA}$  to  $X_{RNA}$  through the decoder, which is usually symmetrical with encoder [28]. In our study, the decoder designed here is not symmetrical with the encoder, and it is just a linear neural network layer. This simplest decoding structure puts forward higher requirements for  $z_{RNA}$ , which can further ensure the effectiveness of  $z_{RNA}$  to capture  $X_{RNA}$  information. The output of the decoder is denoted as  $X'_{RNA}$  (Eq. (6)). The dimensionality of  $X'_{RNA}$  is consistent with that of  $X_{RNA}$ .

According to the principles of VAE,  $X'_{RNA}$  should be as close as possible to  $X_{RNA}$ . However, the previous studies have shown that  $X_{RNA}$  data may suffer from ‘Dropout’ due to technical defects, that is,  $X_{RNA}$  has noise [29–31]. In this case, the proximity of  $X'_{RNA}$  to  $X_{RNA}$  can cause noise in the latent space of the RNA. To solve this problem, we devise a scheme to make  $X_{RNA}$  close to its distribution. The distribution of data is noise-free. A distribution can be generated when the parameters of the distribution are determined. Therefore, objective (ii) is to infer the distribution parameters of  $X_{RNA}$  to obtain a noise-free distribution, which is given as follows:

$$\theta_{RNA} = \text{Linear}(z_{RNA}) \quad (7)$$

$$\delta_{RNA} = \text{Linear}(z_{RNA}) \quad (8)$$

$$\text{NB}(X; \theta; \delta) = \frac{\Gamma(X + \theta)}{\Gamma(\theta)} \left( \frac{\theta}{\theta + \delta} \right)^\theta \left( \frac{\delta}{\theta + \delta} \right)^X \quad (9)$$

$$P(X_{\text{RNA}}) = \text{NB}(X_{\text{RNA}}; \theta_{\text{RNA}}, \delta_{\text{RNA}}) \quad (10)$$

$$P(X'_{\text{RNA}}) = \text{NB}(X'_{\text{RNA}}; \theta_{\text{RNA}}, \delta_{\text{RNA}}) \quad (11)$$

Previous work has shown that RNA data follows an NB distribution. We infer the parameters  $\theta_{\text{RNA}}$  and  $\delta_{\text{RNA}}$  of the NB distribution from  $z_{\text{RNA}}$  using two linear neural networks, respectively (Eqs (7) and (8)). The dimensions of  $\theta_{\text{RNA}}$  and  $\delta_{\text{RNA}}$  output are the same as the dimensions of  $X_{\text{RNA}}$ . The probability density function of NB distribution can be generated according to  $\theta_{\text{RNA}}$  and  $\delta_{\text{RNA}}$  (Eq. (9)). The probability density values of  $X_{\text{RNA}}$  and  $X'_{\text{RNA}}$  in this NB distribution are  $P(X_{\text{RNA}})$  and  $P(X'_{\text{RNA}})$ , respectively, (Eqs (10) and (11)).

The loss function to train the VAE model for RNA is given by

$$\text{COS}(A, B) = \frac{A \cdot B}{\|A\| \|B\|} \quad (12)$$

$$\text{REC\_Loss}_{\text{RNA}} = \text{COS}(X_{\text{RNA}}, X'_{\text{RNA}}) + \text{COS}(P(X_{\text{RNA}}), P(X'_{\text{RNA}})) \quad (13)$$

$$\text{KL\_Loss}_{\text{RNA}} = \sum [\exp(\mu_{\text{RNA}}) - (1 + \sigma_{\text{RNA}}) + \mu_{\text{RNA}}^2] \quad (14)$$

$$\text{RNA}_{\text{Loss}} = \text{REC\_Loss}_{\text{RNA}} + \text{KL\_Loss}_{\text{RNA}} \quad (15)$$

Cosine similarity (COS) is introduced as a measure of proximity (Eq. (12)). The proximity of  $X_{\text{RNA}}$  and  $X'_{\text{RNA}}$  is denoted as  $\text{COS}(X_{\text{RNA}}, X'_{\text{RNA}})$  (Eq. (13)). The proximity of  $P(X_{\text{RNA}})$  and  $P(X'_{\text{RNA}})$  is denoted as  $\text{COS}(P(X_{\text{RNA}}), P(X'_{\text{RNA}}))$  (Eq. (13)). The sum of  $\text{COS}(X_{\text{RNA}}, X'_{\text{RNA}})$  and  $\text{COS}(P(X_{\text{RNA}}), P(X'_{\text{RNA}}))$  is used as the loss function of reconstructed RNA data, denote as  $\text{REC\_Loss}_{\text{RNA}}$  (Eq. (13)). Besides, according to the principles of VAE, the RNA latent space  $z_{\text{RNA}}$  is constrained to be as close as possible to the standard normal distribution, and its difference from the normal distribution is defined as  $\text{KL\_Loss}_{\text{RNA}}$  (Eq. (14)). The loss of reconstructed RNA data and the loss of RNA latent space together constitute the total loss of the RNA parameter inference model, which is defined as  $\text{RNA}_{\text{Loss}}$  (Eq. (15)).

## Protein parameter inference model

The protein parameter inference model is similar in structure and function to the RNA parameter inference model. The protein parameter inference model is specially designed for protein data features, which has two objectives: (i) to construct a canonical data space to recapitulate protein heterogeneity and (ii) to infer the parameters of the protein data distribution to reconstruct noise-free protein data.

To achieve objective (i), we also design an asymmetric autoencoder for the protein parameter inference model, which is like the RNA parameter inference model, given by

$$X_{\text{protein}} = \text{Input}(X_{\text{scaled\_protein}}) \quad (16)$$

$$X_{\text{encoded\_protein}} = \text{Encoder}(X_{\text{protein}}) \quad (17)$$

$$\mu_{\text{protein}} = \text{Linear}(X_{\text{encoded\_protein}}) \quad (18)$$

$$\sigma_{\text{protein}} = \text{Linear}(X_{\text{encoded\_protein}}) \quad (19)$$

$$z_{\text{protein}} = \mu_{\text{protein}} + N(0, I) e^{\frac{\sigma_{\text{protein}}}{2}} \quad (20)$$

$$X'_{\text{protein}} = \text{Linear}(z_{\text{protein}}) \quad (21)$$

We take  $X_{\text{scaled\_protein}}$  as the input to the protein parameter inference model, and it is renamed as  $X_{\text{protein}}$  (Eq. (16)). The dimensionality of  $X_{\text{protein}}$  is consistent with the feature dimensionality of protein data. It is encoded as  $X_{\text{encoded\_protein}}$  by a classical encoder (a multi-layer neural network structure with gradually decreasing dimension) (Eq. (17)). By default,  $X_{\text{encoded\_protein}}$  consists of three-layer neural network, with 1024, 256 and 128 neurons, respectively. According to the principle of classical VAE,  $\mu_{\text{protein}}$  and  $\sigma_{\text{protein}}$  can be generated from  $X_{\text{encoded\_protein}}$  by a simple linear transformation, which are the mean and variance parameters for constructing the protein latent space, respectively (Eqs (18) and (19)). Next, a normal distribution space  $z_{\text{protein}}$  can be constructed according to  $\mu_{\text{protein}}$ ,  $\sigma_{\text{protein}}$  and  $N(0, I)$  (Eq. (20)).  $z_{\text{protein}}$  is the latent space of the protein parameter inference model, which is considered to have the effective information of proteins. By default,  $\mu_{\text{protein}}$ ,  $\sigma_{\text{protein}}$  and  $z_{\text{protein}}$  are 128-dimensional neural network layers. The output of the decoder is denoted as  $X'_{\text{protein}}$  (Eq. (21)). The dimensionality of  $X'_{\text{protein}}$  is consistent with that of  $X_{\text{protein}}$ .

The objective (ii) is to infer the distribution parameters of  $X_{\text{protein}}$  to obtain a noise-free protein data distribution, which is given as follows:

$$\text{Poisson}(X; \lambda) = \frac{e^{-\lambda} \lambda^X}{X!} \quad (22)$$

$$\lambda_{\text{protein}} = \text{Linear}(z_{\text{protein}}) \quad (23)$$

$$P(X_{\text{protein}}) = \text{Poisson}(X_{\text{protein}}; \lambda_{\text{protein}}) \quad (24)$$

$$P(X'_{\text{protein}}) = \text{Poisson}(X'_{\text{protein}}; \lambda_{\text{protein}}) \quad (25)$$

Previous studies have shown that protein data obey Poisson distribution (Eq. (22)) [7, 20]. We infer the parameter  $\lambda_{\text{protein}}$  of the Poisson distribution from  $z_{\text{protein}}$  by a linear neural network layer transformation (Eq. (23)). The dimensionality of outputs from  $\lambda_{\text{protein}}$  is the same as the dimensionality of  $X_{\text{protein}}$ . The projections of  $X_{\text{protein}}$  and  $X'_{\text{protein}}$  on the Poisson distribution probability density function are denoted by  $P(X_{\text{protein}})$  and  $P(X'_{\text{protein}})$ , respectively (Eqs (24) and (25)).

The loss function to train the VAE model for protein is given by

$$\begin{aligned} \text{REC\_Loss}_{\text{protein}} &= \text{COS}(X_{\text{protein}}, X'_{\text{protein}}) \\ &+ \text{COS}(P(X_{\text{protein}}), P(X'_{\text{protein}})) \end{aligned} \quad (26)$$

$$\text{KL\_Loss}_{\text{protein}} = \sum [\exp(\mu_{\text{protein}}) - (1 + \sigma_{\text{protein}}) + \mu_{\text{protein}}^2] \quad (27)$$

$$\text{Protein}_{\text{Loss}} = \text{REC\_Loss}_{\text{protein}} + \text{KL\_Loss}_{\text{protein}} \quad (28)$$

$\text{REC\_Loss}_{\text{protein}}$  is the loss of reconstructed protein data. It also uses cosine similarity as the loss function.  $\text{COS}(X_{\text{protein}}, X'_{\text{protein}})$  represents the difference between the input protein data and the output data of the model.  $\text{COS}(P(X_{\text{protein}}), P(X'_{\text{protein}}))$  indicates the difference between input and output data in Poisson distribution with the same parameters (Eq. (26)). According to the principles

of VAE, the protein latent space  $z_{\text{protein}}$  is constrained to be as close as possible to the standard normal distribution, and its difference from the normal distribution is defined as  $\text{KL\_Loss}_{\text{protein}}$  (Eq. (27)). The loss of reconstructed protein data and the loss of protein latent space together constitute the total loss of the protein parameter inference model, which is defined as  $\text{Protein\_Loss}$  (Eq. (28)).

### Multimodal parameter inference model

Multimodal parametric model also has two objectives: (i) to construct a multimodal parameter space from RNA and protein modalities, and (ii) to ensure that the multimodal parameter space can reconstruct RNA and protein modalities.

The autoencoder for multimodal parameter inference space is described as

$$\mu_{\text{multimodal}} = \text{Encoder}([\mu_{\text{RNA}}, \mu_{\text{protein}}]) \quad (29)$$

$$\sigma_{\text{multimodal}} = \text{Encoder}([\sigma_{\text{RNA}}, \sigma_{\text{protein}}]) \quad (30)$$

$$z_{\text{multimodal}} = \mu_{\text{multimodal}} + N(0, I) e^{\sigma_{\text{multimodal}}} \quad (31)$$

$$z'_{\text{RNA}} = \text{Linear}(z_{\text{multimodal}}) \quad (32)$$

$$z'_{\text{protein}} = \text{Linear}(z_{\text{multimodal}}) \quad (33)$$

To achieve objective (i), a trick used in the model is that we do not directly fuse RNA and protein potential space, but instead, we fuse their parameters. On the one hand,  $\mu_{\text{RNA}}$  and  $\mu_{\text{protein}}$  are mixed by an encoder to generate  $\mu_{\text{multimodal}}$  (Eq. (29)). Specifically,  $\mu_{\text{RNA}}$  and  $\mu_{\text{protein}}$  are concatenated and then fed to  $\mu_{\text{multimodal}}$ , which is a 128-dimensional neural network layer. On the other hand,  $\sigma_{\text{RNA}}$  and  $\sigma_{\text{protein}}$  are mixed by another encoder to generate  $\sigma_{\text{multimodal}}$  (Eq. (30)). Specifically,  $\sigma_{\text{RNA}}$  and  $\sigma_{\text{protein}}$  are concatenated and then fed to  $\sigma_{\text{multimodal}}$ , which is a 128-dimensional neural network layer. According to the principle of VAE,  $\mu_{\text{multimodal}}$ ,  $\sigma_{\text{multimodal}}$  and  $N(0, I)$  can construct a normal distribution space  $z_{\text{multimodal}}$ , which is the multimodal parameter space (Eq. (31)). By default,  $z_{\text{multimodal}}$  is a 128-dimensional neural network layer.  $z_{\text{multimodal}}$  can decode back to the latent space of RNAs and proteins, which are defined as  $z'_{\text{RNA}}$  and  $z'_{\text{protein}}$ , respectively (Eqs (32) and (33)). The dimensionality of  $z'_{\text{RNA}}$  is the same as that of  $z_{\text{RNA}}$ . The dimensionality of  $z'_{\text{protein}}$  is the same as that of  $z_{\text{protein}}$ .

Objective (ii) is to make  $z'_{\text{RNA}}$  and  $z'_{\text{protein}}$  closer to  $z_{\text{RNA}}$  and  $z_{\text{protein}}$ . Although the multimodal parameters are derived from independent modalities, this does not guarantee that the multimodal parameter space covers the independent model features. To ensure that the multimodal parameter space can reconstruct the RNA and protein latent space, we propose objective (ii).

Thus, the loss function to train the VAE model for multimodal data is given by

$$\text{REC\_Loss}_{\text{multimodal}} = \text{MSE}(z_{\text{RNA}}, z'_{\text{RNA}}) + \text{MSE}(z_{\text{protein}}, z'_{\text{protein}}) \quad (34)$$

$$\text{KL\_Loss}_{\text{multimodal}} = \sum [\exp(\mu_{\text{multimodal}}) - (1 + \sigma_{\text{multimodal}}) + \mu_{\text{multimodal}}^2] \quad (35)$$

$$\text{Multimodal\_Loss} = \text{REC\_Loss}_{\text{multimodal}} + \text{KL\_Loss}_{\text{multimodal}} \quad (36)$$

$$\text{DPI\_Loss} = \text{RNA\_Loss} + \text{Protein\_Loss} + \text{Multimodal\_Loss} \quad (37)$$

The reconstruction of RNA and protein latent space is calculated by mean square error (MSE), which jointly constitute the multimodal reconstruction loss  $\text{REC\_Loss}_{\text{multimodal}}$  (Eq. (34)). When  $\text{REC\_Loss}_{\text{multimodal}}$  achieves a minimum value, it can be considered that the multimodal parameter space has successfully reconstructed the RNA and protein latent space. According to the principles of VAE, the multimodal parameter space  $z_{\text{multimodal}}$  is constrained to be as close as possible to the standard normal distribution, and its difference from the normal distribution is defined as  $\text{KL\_Loss}_{\text{multimodal}}$  (Eq. (35)). The loss of reconstructed multimodal data and the loss of multimodal parameter space together constitute the total loss of the multimodal parameter inference model, which is defined as  $\text{Multimodal\_Loss}$  (Eq. (36)).

As a result, the total loss of the DPI model  $\text{DPI\_Loss}$  consists of three parts:  $\text{RNA\_Loss}$ ,  $\text{Protein\_Loss}$  and  $\text{Multimodal\_Loss}$ . The optimization goal of the DPI model is to minimize  $\text{DPI\_Loss}$  (Eq. (37)). When it achieves the minimum value, we believe that the DPI model captures the cellular heterogeneity at three views of RNA, protein and multimodal, respectively, and reconstructs the distribution of RNA and protein data.

### Reference and query

The multimodal parameter space generated by DPI can not only profile the heterogeneity of cells, but also serve as a reference for cell types.

The multimodal parameter space of samples with the known cell subtypes can be used to annotate cell subtypes in other samples. DPI encodes each cell as an embedding in the multimodal parameter space. The distance of the embedding in the multimodal parameter space can characterize the similarity of cells. The multimodal parameter space with cell types is a reference for locating cell types. Other unannotated cells can query the cell types against this reference. Since the multimodal parameter space only encodes information about the heterogeneity of cells, it ignores batches of samples. Annotating samples using the multimodal parameter space with cell types requires the following steps:

- (i) To train DPI models using annotated single-cell multimodal data and obtain multimodal parameter spaces as well as cell embeddings, in which the embeddings are called annotated cell embeddings
- (ii) To feed unannotated single-cell multimodal data into the trained DPI model and obtain the cell embeddings in the multimodal parameter space, in which the embeddings are referred to as the unannotated cell embeddings
- (iii) To calculate the cosine distance of each unannotated cell to all annotated cells on the embedding, in which the cell type of the unannotated cells is derived from the cell type of the annotated cell with the maximum cosine distance.

Considering that the annotated and unannotated cell embeddings are located in the same multimodal parameter space, they can be visualized in the same UMAP model. Embedding the unannotated cells into a UMAP model of annotated cells allows further visualization of cell types and batches.

### Cell state vector field

The cell embedding of the multimodal parameter space profiles the cell state determined by the expression of genes and proteins. Since the multimodal parameter space is a continuous space, it covers the expression of all possible genes and proteins in the

sample. Altering the expression of genes and proteins will change the embedding of cells in the multimodal parameter space. We propose a cell state vector field to describe the effects of gene and protein changes on cell states based on a multimodal parameter space. Executing the cell state vector field from the multimodal parameter space requires the following steps:

- (i) To train a DPI model using single-cell multimodal data and output a multimodal parameter space as well as the cell embeddings, in which the cell embeddings are called the original cell embeddings
- (ii) To train a UMAP model using the multimodal parameter space
- (iii) To change gene/protein expression in single-cell multimodal data. For gene/protein up-regulation, the gene/protein expression data are expanded, by default, by a factor of 2. For down-regulation of a gene/protein, the gene/protein expression data are scaled down by a default factor of  $-2$ . Typically, we do not recommend to modify the expression values of multiple genes and proteins at the same time, which will make the experimental results difficult to interpret
- (iv) To input the modified single-cell multimodal data into the trained DPI and output the cell embeddings, in which the cell embeddings are referred to as the regulated cell embeddings
- (v) To build the cell state vector in the multimodal parameter space. The starting point of the vector is the original cell embedding. The direction of the vector is from the original cell–cell embedding to the regulated cell embedding. The magnitude of the vector is the difference between the original cell embedding and the regulated cell embedding
- (vi) To visualize all the cell state vectors in UMAP, which constitute the cell state vector field, the visualization code of the cell state vector field refers to the UMAP function of scVelo [32].

## Results

### Combining individual modalities to infer multimodal parameter space

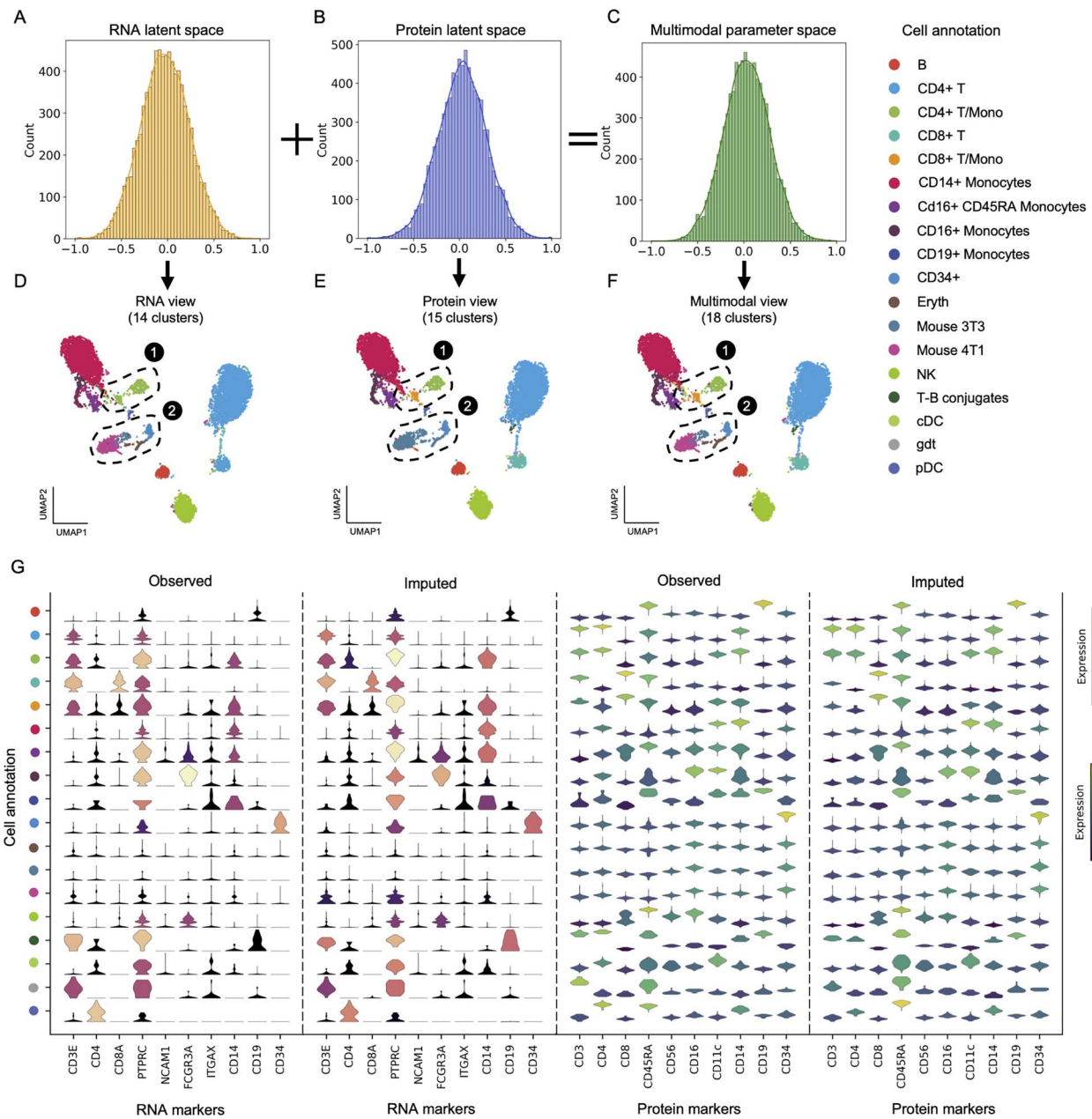
We take the CBMC sample as an example to demonstrate the performance of DPI for inferring multimodal heterogeneity from individual modalities. The CBMC samples sequenced by Stoeckius *et al.*, measure more than 20 000 types of genes and 10 immune-related types in 8617 cells simultaneously [7]. CBMC is fed into the framework of DPI. After training, we extract the embeddings of the RNA latent space (Figure 2A), the protein latent space (Figure 2B) and the multimodal parameter space (Figure 2C). These embeddings represent RNA, protein and multimodal features, respectively. We find that these features are normally distributed, and the latent space reconstructed from the multimodal parameter space is also normally distributed (Supplementary Figure S1 available online at <http://bib.oxfordjournals.org/>). These results imply that DPI successfully transforms RNA and protein features into normal distributions and generates a recoverable multimodal parameter space.

The RNA and protein latent spaces represent features of RNA and protein modalities, respectively. The multimodal parameter space represents the multimodal features. We perform cell clustering and visualization with RNA, protein and multimodal features, respectively (Supplementary Figure S2A–C available online at <http://bib.oxfordjournals.org/>). To directly compare RNA,

protein and multimodal differences, we visualize the cell clusters in all views with multimodal UMAPs (Figure 2D–F). Among them, 14 cell subtypes are identified in the RNA view, 15 cell subtypes are discriminated in the protein view, and 18 cell subtypes are found in the multimodal view (Figure 2D–F). While all three views successfully capture the cellular heterogeneity of CBMC sample, they find different cellular subtypes.

For example, CD8+ T/Mono is found in the protein view, but not in the RNA view (Circle 1 in Figure 2D and E) [33], probably due to the specific high expression of CD8 protein in CD8+ T/Mono (Supplementary Figure S2B and C available online at <http://bib.oxfordjournals.org/>). Another example involving mouse cells is that several mouse cells spike into the CBMC dataset to detect the sensitivity of antibodies to the protein [7]. The RNA view successfully identifies subtypes of these mouse cells (Circle 2 in Figure 2D and Supplementary Figure S2A and C available online at <http://bib.oxfordjournals.org/>). However, the protein view is unable to identify the subtypes of these cells due to the lack of protein marker (Circle 2 in Figure 2E). These results show the possible limitations of a single modal view, which may be overcome by the multimodal views. All the cell subtypes annotated by the RNA and protein views, including CD8+ T/Mono and mouse cell subtypes are successfully discovered by the multimodal view (Supplementary Figure S2C available online at <http://bib.oxfordjournals.org/>). These results confirm that a multimodal view can capture more comprehensive cellular heterogeneity than single modality views.

Considering that the quality of RNA and protein latent spaces directly affects the performance of multimodal parameter space, we output the imputed RNA and protein data from RNA and protein latent spaces. We find that the imputed data are similar to the observed data for both RNA and protein (Figure 2G and Supplementary Figure S2A–C available online at <http://bib.oxfordjournals.org/>). This demonstrates that the RNA and protein latent spaces successfully compress the real multi-omics data. This guarantees the reliability of the multimodal parameter space feature source. Furthermore, we validate the robustness of DPI in low-quality multi-omics data (Supplementary Figure S3A–C available online at <http://bib.oxfordjournals.org/>). On one hand, we simulate Dropout to generate low-quality RNA data (Supplementary Figure S3B available online at <http://bib.oxfordjournals.org/>) [29]. Dropout is not random, and previous studies have shown that it usually occurs on low-expressed RNAs [30, 31]. To simulate Dropout more realistically, we define RNAs with expression below the quartile as low-expressing RNAs. Considering the different RNAs expressed inside each cell, we count the low-expressing RNAs for each cell separately. Furthermore, each low-expressing RNA is set to zero with 90% probability. The RNA data of all cells are involved in the simulated Dropout. These Dropout RNAs and normal proteins are fed to train DPI and apply UMAP for visualization (Supplementary Figure S3B available online at <http://bib.oxfordjournals.org/>). On the other hand, we simulated contamination to produce low-quality protein data [7] (Supplementary Figure S3C available online at <http://bib.oxfordjournals.org/>). There are two types of protein contamination. In one case, the tag of the protein binds the protein non-specifically [7]. The other is that the tag of the protein does not bind to the protein tightly enough [7]. These two types of contamination are almost random. To simulate protein contamination, all protein data are randomly increased/decreased in abundance by 0–20%. Contaminated protein and normal RNA data are fed into DPI to train the model and visualized



**Figure 2.** Application of DPI to analyze CBMC sample. DPI maps CBMC data into (A) RNA latent space, (B) protein latent space and (C) multimodal parameter space. (D) The RNA and (E) the protein latent spaces reveal the cellular heterogeneity from its own view. (F) The multimodal parameter space reveals the integrated cellular heterogeneity. (G) Violin plots of the observed and the imputed multimodal data.

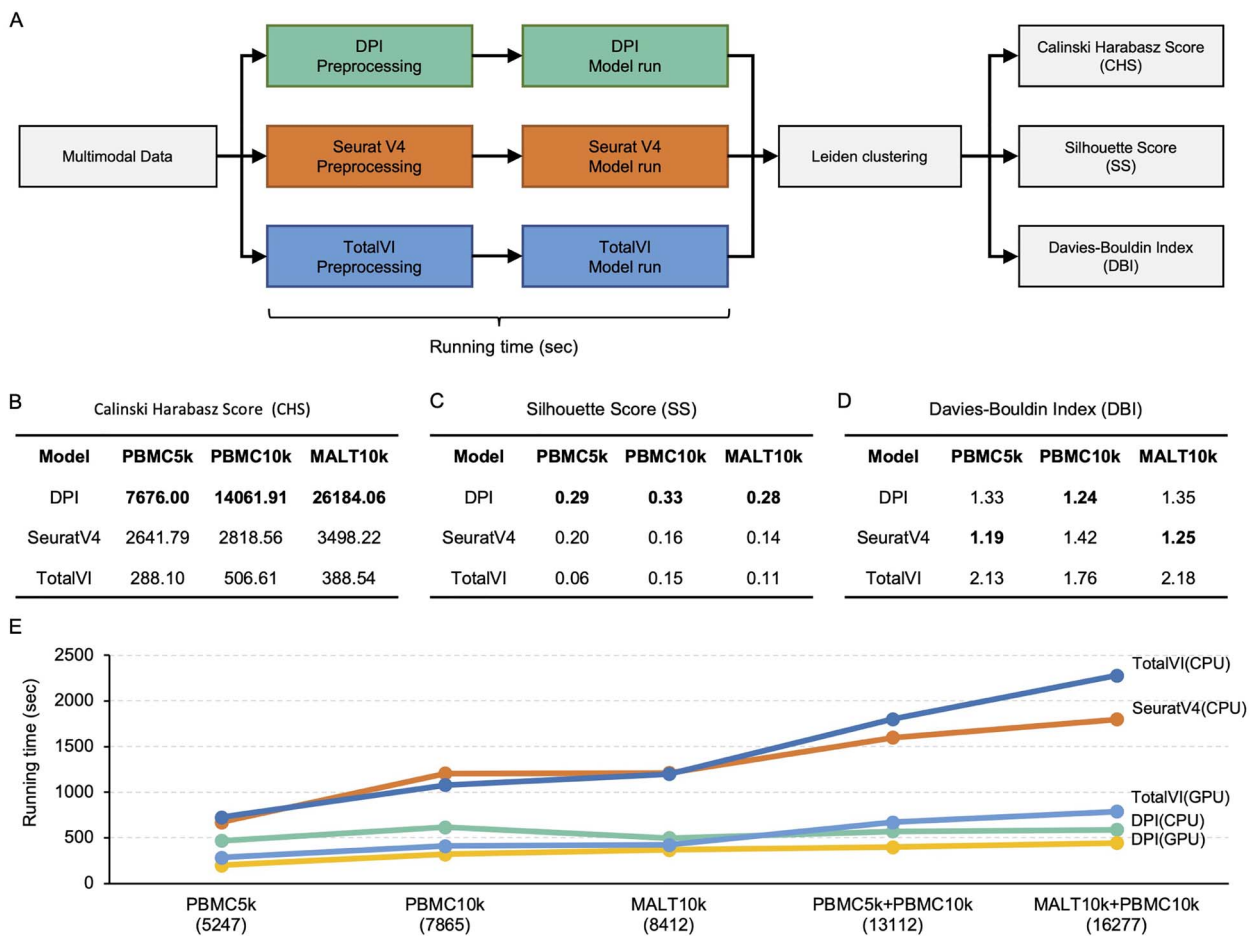
using UMAP (Supplementary Figure S3C available online at <http://bib.oxfordjournals.org/>). We find that both the low-quality RNA data and the low-quality protein data are similar to the conventional data integration results (Supplementary Figure S3A–C available online at <http://bib.oxfordjournals.org/>). This means that DPI has a degree of robustness in integrating low-quality data. In summary, DPI characterizes cellular heterogeneity by inferring a reliable multimodal parameter space from each reliable single modality.

### Comparison and evaluation with state-of-the-art models

We compare the performance of DPI with two types of state-of-the-art methods, namely the machine learning-based methods and neural network-based methods. We choose SeuratV4 as a

representative of machine learning methods, which is one of the state-of-the-art methods for integrating CITE-seq data. SeuratV4 proposes a weighted nearest neighbor analysis to learn the information content of each modality in each cell and defines cell subtypes based on a weighted combination of two modalities. We apply the default configuration of SeuratV4; i.e. RNA and protein data are reduced to 30 and 18 dimensions by PCA, respectively. These dimensions are used by SeuratV4 to find multimodal neighbors. The representative of neural network integration methods is TotalVI, which integrates multimodal data using variational inference and autoencoders. We apply the default configuration of TotalVI, where the ‘latent\_distribution’ parameter is set to ‘normal’. SeuratV4, TotalVI and DPI follow the same pipeline to ensure a fair comparison (Figure 3A).

Typically, the previous annotations of cellular data are derived from RNA data. Considering that multimodal data can identify



**Figure 3.** Comparison of DPI with state-of-the-art models. (A) We design a process to fairly evaluate the performance of DPI and state-of-the-art methods. To objectively evaluate the performance of these models, we introduce three unsupervised evaluation metrics: CHS (B), SS (C) and DBI (D). In addition, we test the performance of models in GPU and CPU environments (E).

more cell subtypes, cell annotations of RNA views are not appropriate to assess the clustering performance of multimodal views. Three unlabeled clustering metrics, Calinski Harabasz score (CHS) [34], Silhouette score (SS) [35] and Davies Bouldin Index (DBI) [36], are introduced to evaluate the performance of these three models (Supplementary Material S2 available online at <http://bib.oxfordjournals.org/>).

CHS evaluates the between-cluster variance and within-cluster variance of multimodal data to calculate scores. A higher CHS means a better performance. The results show that DPI outperforms SeuratV4 and TotalVI under the CHS metric (Figure 3B). This shows that DPI is a good description of the differences between cell clusters and the homogeneity within cell clusters.

SS is another commonly used unsupervised clustering evaluation metric. It ranges from  $-1$  to  $1$ , in which  $-1$  means the worst clustering and  $1$  means the perfect clustering. SS around zero indicates the overlapping clusters. The performance of DPI under the SS indicator is also better than those of SeuratV4 and TotalVI (Figure 3C). This implies that DPI-generated cell clusters are dense and well separated.

Furthermore, we introduce DBI to evaluate the performance of unsupervised cell clustering. DBI stands for the average 'similarity' between clusters, where similarity is a measure of how close the clusters are to the size of the clusters themselves. Note that zero is the lowest possible score for DBI, the values closer to zero indicates a better partitioning. DPI and SeuratV4 achieve lower DBI compared with TotalVI, indicating that DPI and

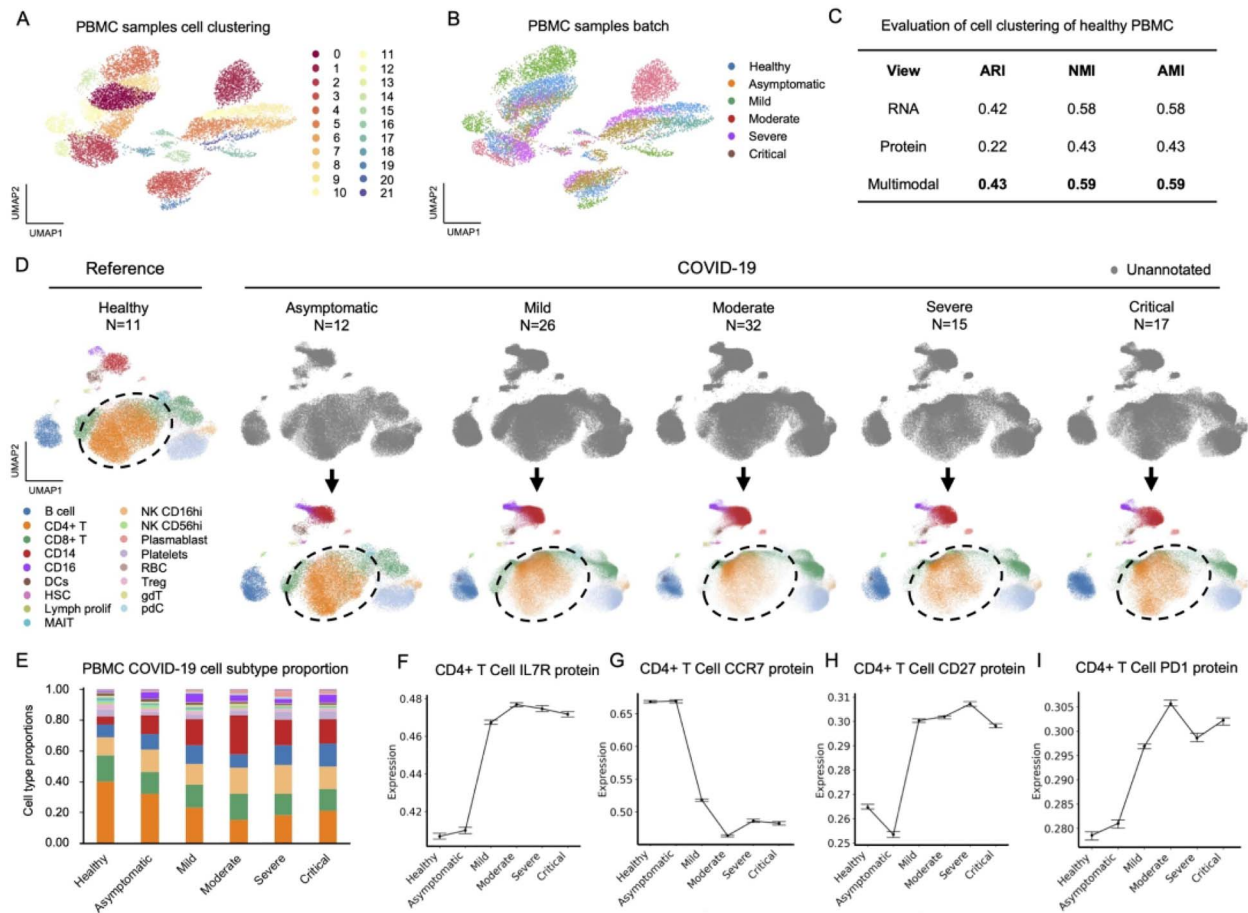
SeuratV4 can better profile the relationship between cell clusters than TotalVI (Figure 3D).

These results show that DPI has superior performance compared with the above-mentioned methods. Besides, DPI enables efficient analysis of large-scale single-cell multimodal data (Figure 3E). It is worth noting that DPI also supports GPU acceleration, in this mode DPI has higher efficiency. In summary, DPI is a powerful single-cell multimodal analysis framework that can accurately and efficiently analyze single-cell multimodal data.

### Reference and query cell subtypes based on the multimodal parameter space

DPI can not only characterize cell heterogeneity, but also annotate cell types regardless of batch effects. Batch effects are a major challenge for single-cell sequencing analysis. In this case, more than 600 000 PBMC with COVID-19 are introduced. COVID-19 PBMC samples are classified as healthy, asymptomatic, mild, moderate, severe and critical ones. To exclude the effect of sample size, we randomly sample 2000 cells from each disease progression. There are total 12 000 PBMC with six disease progressions. These cells are fed into DPI for clustering (Figure 4A). We find batch effects for these samples (Figure 4B). To illustrate the ability of DPI to eliminate batch effects and annotate cell types, we conduct the analysis on the full COVID-19 dataset (not the sampled dataset).





**Figure 4.** Application of DPI to reference and query PBMC for COVID-19 disease progression. **(A)** Multiple COVID-19 disease progression PBMC samples are clustered and **(B)** visualized by batch. **(C)** We evaluate the clustering results from three views of RNA, protein and multimodality in healthy PBMC, respectively. **(D)** Visualization of the progression of COVID-19 disease with healthy PBMC as a reference. **(E)** The proportion of PBMC subtypes during COVID-19 disease progression. Changes in CCR7 **(F)**, IL7R **(G)**, CD27 **(H)** and PD-1 **(I)** expression levels of CD4+ T cells in PBMC with the progression of COVID-19 disease.

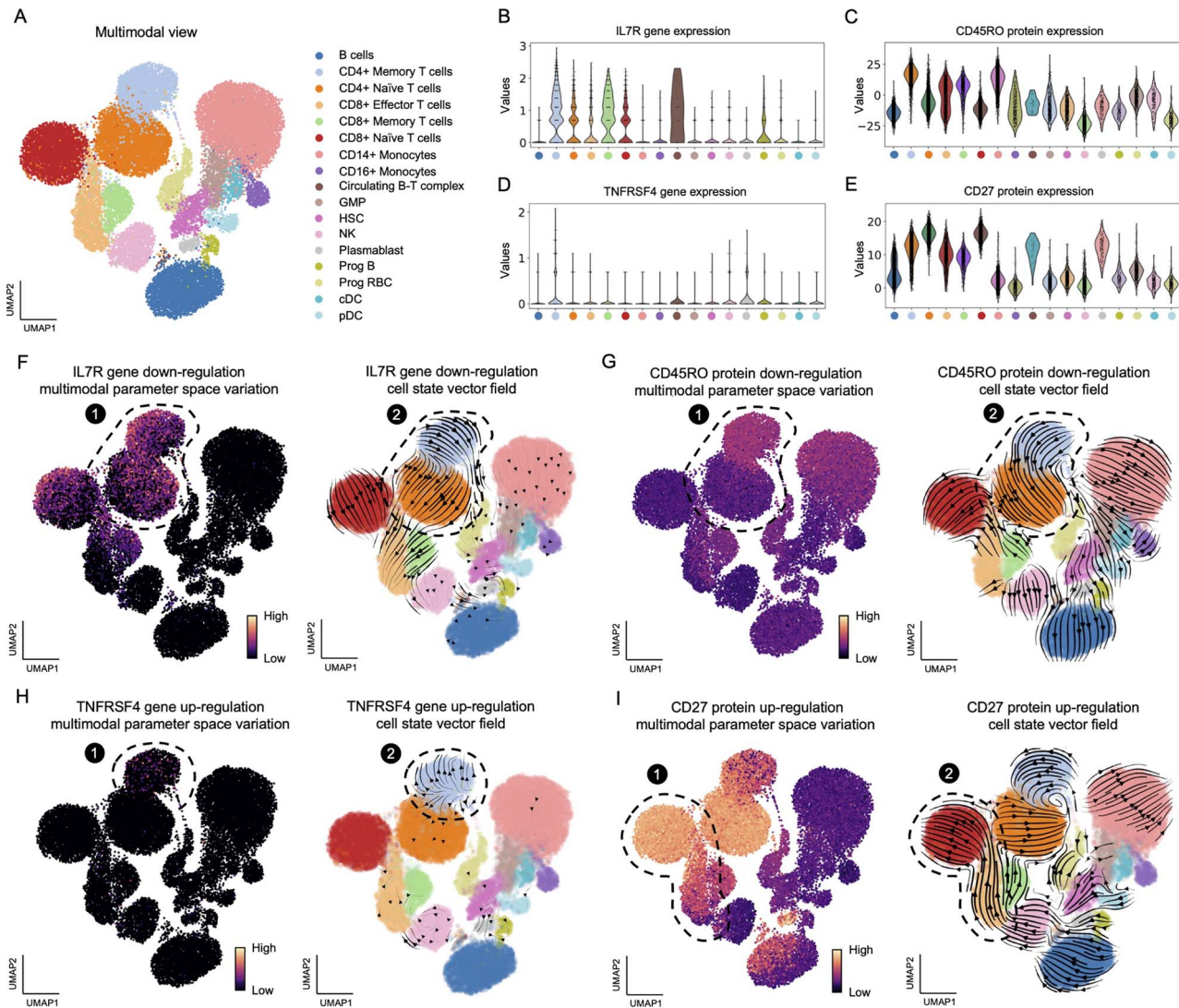
DPI annotates cell types based on reference and query strategies. Healthy PBMC are used as the reference to annotate mild, asymptomatic, severe and critically COVID-19 PBMC. Based on the previous studies, we annotate healthy PBMC in RNA, protein, and multimodal views as detailed as possible (Supplementary Figure S4A–C available online at <http://bib.oxfordjournals.org/>).

The adjusted Rand Index (ARI) [37], Normalized Mutual Information (NMI) [38] and Adjusted Mutual Information (AMI) [38] are introduced to evaluate the performance of these three views (Supplementary Material S3 available online at <http://bib.oxfordjournals.org/>). We find that the multimodal view performs best (Figure 4C).

We further merge cellular subtypes of the healthy PBMC (Supplementary Figure S4D available online at <http://bib.oxfordjournals.org/>). We separately query and visualize the PBMC subtypes of these COVID-19 patients, using the healthy PBMC multimodal parameter space as a reference (Figure 5D and Supplementary Figure S5B–F available online at <http://bib.oxfordjournals.org/>). By UMAP visualization, we find that the coordinates of cells with the same cell subtype are very close (Figure 4D). For example, CD4+ T cells in all patient states are located in the middle of the UMAP visualization (Circle in Figure 4D). The UMAP visualization reflects the location of the data in the embedding space.

Although the healthy and COVID-19 PBMC have the same cell composition, the proportions of their cellular subtypes are different (Figure 4E). It can be found that with the progression of COVID-19 disease, the proportion of CD4+ T cells gradually decreases (Figure 4E). It is well known that T cells are essential lymphoid immune cells. Numerous studies have shown that T cells play many key roles in fighting against COVID-19 [39–42]. Studying the reduction in the proportion of CD4+ T cells can help reveal the progression of COVID-19 disease. As a fact, the proportion of cells is directly related to the function of the cell, which is determined by the expression of genes and proteins. We find that the expression level of IL7R protein increases in CD4+ T cells as the COVID-19 disease progresses (Figure 4F). IL7R is associated with the growth and proliferation of lymphocytes. An increase in IL7R implies a rapid CD4+ T growth [43].

We also find a decrease in CCR7 protein (Figure 4G) and an increase in CD27 protein (Figure 4H) in CD4+ T cells as the disease progresses. It is known that naïve CD4+ T cells express higher levels of CCR7, and the previous studies have shown that CD27 is increased upon T cell activation. The decreased CCR7 and increased CD27 suggest that CD4+ T cells are rapidly activated as the disease progresses [44–47]. In addition, we notice that the expression level of PD1 protein increases as the disease progresses (Figure 4I). PD1 is a marker of T cell exhaustion [48]. With the development of COVID-19, T cells are rapidly exhausted [49].



**Figure 5.** Predicting the effect of gene and protein regulation on BMC cell state. (A) BMC samples are annotated with 17 cell subtypes. Violin plot visualization of IL7R gene (B), CD45RO protein (C), TNFRSF4 gene (D) and CD27 protein (E) expression. The cell state changes are induced by IL7R gene down-regulation (F), CD45RO protein down-regulation (G), TNFRSF4 gene up-regulation (H) and CD27 protein up-regulation (I).

These results suggest that CD4+ T cells rapidly grow and become activated after infection with COVID-19. However, these activated CD4+ T cells are rapidly exhausted. Rapid depletion of CD4+ T cells is manifested by a decrease in the proportion of CD4+ T cells. A similar pattern is also found in CD8+ T cells (Supplementary Figure S6 available online at <http://bib.oxfordjournals.org/>). Previous studies have also shown that as COVID-19 disease progresses, T lymphocytes are activated and exhausted [48, 49]. The above results demonstrate that DPI can help researchers to analyze disease progression across multiple samples in an intuitive manner. In addition, DPI not only automates the annotation of cell subtypes, but also preserves the biological properties of the sample. These results are attributed to the fact that the multimodal parameter space constructed with the reference dataset is batch-independent.

### Cell state vector fields based on multimodal parameter space

The multimodal parameter space generated by DPI can also be used to predict the effects of changes in genes and proteins on cell states. It can be profiled on 'cell state vector fields'. In

this case, we introduce BMC samples to illustrate the function of the cell state vector field. The BMC samples include 30 672 human bone marrow cells that are simultaneously measured for transcriptome and 25 types of proteins. We find 17 cell subtypes in the BMC (Figure 5A and Supplementary Figure S7 available online at <http://bib.oxfordjournals.org/>). We provide examples of the IL7R gene (Figure 4B), CD45RO protein (Figure 4C), TNFRSF4 gene (Figure 4D) and CD27 protein (Figure 4E) to illustrate the biomedical function of cell state vector field.

We simulate IL7R gene down-regulation with the cell state vector field of DPI (Figure 4F). As seen from the variations in the multimodal parameter space (Figure 5F), CD4+ T cells are affected by the down-regulation of the IL7R gene (Circle 1 in Figure 5F). It is found by the cell state vector field that CD4+ Memory T cells point to CD4+ Naive T cells (Circle 2 in Figure 5F). This represents a shift in the state of CD4+ Memory T cells to that of CD4+ Naive T cells. The CD4+ Naive T cells are resting immune cells, which do not have the immune function of CD4+ Memory T cells. This means that the down-regulation of the IL7R gene results in the conversion of CD4+ T cells to a state similar to that of CD4+ Naive T cells.

Previous studies have shown that the deletion of the IL7R gene leads to severe combined immunodeficiency (SCID) [50–52]. SCID is a fatal immune disease. Since the immune cells lack immunological function, the immune system of SCID patients has little effect on defending against bacteria, viruses and fungi [53]. Our modelling result successfully predicts the transformation of CD4+ T cells to a state without immunological function owing to IL7R down-regulation.

In addition, we simulate the changes in immune cells after IL7R up-regulation (Supplementary Figure S8A available online at <http://bib.oxfordjournals.org/>). We find that CD4+ naïve T cells transform into a state like CD4+ memory cells. It is known that IL7R up-regulation implies lymphocyte activation [54]. Our model successfully predicts the activation of CD4+ naïve T cells to CD4+ memory T cells due to IL7R up-regulation.

Furthermore, DPI can simulate the effect of protein changes on the cell state. Similar to the deletion of the IL7R gene, the deletion of CD45 family proteins also causes SCID [55]. We simulate the down-regulation of CD45RO protein (Circles 1 and 2 in Figure 5G) and find similar results to IL7R gene down-regulation. We also simulate the up-regulation of CD45RO protein (Supplementary Figure S8B available online at <http://bib.oxfordjournals.org/>), which also shows similar results as the up-regulation of the IL7R gene. These results suggest that both gene and protein regulation can be accurately predicted by DPI.

Another case is immunodeficiency disease (IDD), a group of diseases caused by a deficiency in immune function due to an underdeveloped or compromised immune system. Previous studies have shown that the deletion of TNFRSF4 expression causes IDD. We simulate the change in cell state after the down-regulation of the TNFRSF4 gene (Circle 1 in Figure 5H). As shown in Figure 5H, the arrows of the CD4+ Memory T cell state diverge outward (Circle 2 in Figure 5H), which implies that CD4+ Memory T cells cannot be maintained. Previous studies have shown that, when TNFRSF4 gene expression is down-regulated, Naïve T cells have difficulty activating into Memory T cells [56–58]. Our model successfully predicates the state change of CD4+ T memory cells. On the contrary, the up-regulation of TNFRSF4 shows a trend opposite to that in Figure 5H (Circle 2 in Figure 5H), that is, CD4+ T memory cells could develop normally (Supplementary Figure S8C available online at <http://bib.oxfordjournals.org/>) [58].

The above results have demonstrated the biomedical analysis capabilities of the cell state vector field. In fact, the utility of the cell state vector field arises from the expression of genes and proteins. Here, we use the CD27 protein as an example to explain the functional origin of the cell state vector field. It is well known that CD27 is a member of the TNF receptor superfamily and is constitutively expressed on Naïve T cells, Memory B cells, NK cells and HSC (Figure 4E) [59]. CD27 is a transmembrane phosphoglycoprotein expresses on CD4+ and CD8+ T cells. CD27 expression increases upon T cell activation and is shed from the cell surface to form the soluble CD27 upon activation. Therefore, the expression level of CD27 in CD8+ Naïve T cells is higher than that in CD8+ Effector T cells (Figure 5E and Supplementary Figure S9A available online at <http://bib.oxfordjournals.org/>).

We model the up-regulation of CD27 protein for each cell. Since each cell in the cell state vector field is localized according to gene and protein expression, the vector of CD8+ Effector T cells is oriented in the same direction as CD8+ Naïve T cells with higher CD27 expression (circle 2 in Figure 5I). Conversely, the down-regulation of CD27 protein in each cell will lead to the cell state vector pointing from CD8+ Naïve T cells to CD8+

Effector T cells (Supplementary Figure S9B available online at <http://bib.oxfordjournals.org/>).

These results show that the cell state vector field profiles the changes in cell state according to the expression of genes and proteins. The ability of the cell state vector field to profile the cell state changes stems from the continuity of the multimodal parameter space. Limited by experimental conditions, it is unlikely that one sample will cover the possible expression of all genes and proteins. In short, the cellular state represented by the sample is discrete. The multimodal parameter space generated by DPI is continuous, covering all possible values of genes and proteins in a sample with a specific distribution. Therefore, the cell state vector field based on the multimodal parameter space can characterize the change of cell state. In conclusion, the cell state vector fields provide novel and actionable biological insights into the mechanistic drivers behind disease.

## Discussion

In this paper, we propose a novel unsupervised generative model named DPI to analyze single-cell multimodal data. The functions of DPI include multimodal integration, cell clustering and visualization, batch-free reference and automatic cell type annotation, cell state vector field analysis, etc. These capabilities allow DPI to systematically analyze single-cell multimodal data from physiological and disease states. The success of DPI is attributed to innovative parameter inference methods and multimodal parameter spaces. The results of the CBMC analysis show that the multimodal parameter space inferred from the parameters of the RNA and protein latent spaces can better characterize the heterogeneity of cells.

Besides, the distribution of RNA and proteins restored by parameter inference is consistent with the distribution of the original data. It is worth noting that DPI can not only integrate biological data, but also be applied to the integration of other types of multimodal data. DPI can inspire other deep learning research and has a wide range of applications. Comparisons with the state-of-the-art methods illustrate the superior performance of the parameter inference method.

As deep learning model, TotalVI is a great single-cell multi-omics integration model. Both TotalVI and DPI transform multimodal embedding into a standard normal distribution. However, they have different modeling ideas. TotalVI, including SeuratV4, can be simplified to extract features for each modality and then integrate these features. While DPI does not directly integrate the data of each mode, but rather the parameters of each modal distribution. Parameters are not equivalent to features. Parameters describe the data as a whole, while features focus on the individuals in the data. The parameter fusion strategy enables the model to learn data from the whole. Since the distribution of latent layers is predetermined, the difficulty of the model to learn parameters is lower than that of extracting features. Furthermore, RNA, protein and multimodal spaces are designed to have standard normal distributions, avoiding model bias to either side. These modeling ideas enable DPI to model single-cell multimodal data fast and accurately.

The application of COVID-19 indicates that the multimodal parameter space is robust and can serve as a reference for cell types. A multimodal parameter space consisting of annotated samples is a ‘map’ that can be used to query cell types. Similar cell types are approached in the multimodal parameter space.

The multimodal parameter space is a continuous generative space that covers all possible cell states in the sample. Single-cell

multimodal data are transformed into a multimodal parameter space by DPI, which is equivalent to setting 'coordinates' for each possible cell state. Analysis of the BMC dataset reveals that the changes in RNA and protein signatures lead to the changes in the 'coordinates' of cells in a multimodal parameter space. We present a cell state vector field to visualize this change. The cell state vector field relies on continuous cell states generated from a multimodal parameter space. However, the current version of the cell state vector field cannot account for the changes in cell signaling networks. We plan to introduce kinetic methods [60–62] in future release to model cell signaling networks as a function of sampling time.

In conclusion, DPI is a powerful single-cell multimodal analysis framework that not only integrates and analyzes multimodal data, but also reveals new biological insights into biomedical researchers.

### Key Points

- Our integrated multimodal data are not only immune to noise but also aligned with each cell modality.
- We present an efficient framework, which also achieves better clustering performance.
- Our approach can refer and query cell types without batch effects, which is successfully applied to reveal the progression of the COVID-19 disease across multiple samples.
- Our proposed cell state vector fields can visualize the effects of genes and proteins on cell state transitions and reveal the mechanism of disease occurrence and development.
- The present multimodal integration method, fusing the parameters of the data rather than the data itself, may inspire other multimodal integration tasks.

## Supplementary Data

Supplementary data are available at Briefings in Bioinformatics Online.

## Acknowledgements

The authors would like to thank the members in our team, especially, Ruiqi Liu and Chunlin Zhao for valuable discussions. This work was also supported by National Institute for Data Science in Health and Medicine, and State Key Laboratory of Cellular Stress Biology, Innovation Center for Cell Signaling Network.

## Funding

The National Science and Technology Major Project of the Ministry of Science and Technology of China (grant no. 2021ZD0201900), the National Natural Science Foundation of China (grant nos 12090052, 11874310), Foundation of Education Department of Liaoning Province (grant no. LJKZ0280), and the Fujian Province Foundation (grant no. 2020Y4001).

## Data availability

This study uses six publicly available datasets, which are available at <https://github.com/studentiz/dpi/tree/main/data>. The Python

source code and user-friendly documentation for DPI are freely available on GitHub at <https://github.com/studentiz/dpi>.

## References

1. Perez RK, Gordon MG, Subramaniam M, et al. Single-cell RNA-seq reveals cell type-specific molecular and genetic associations to lupus. *Science* 2022;**376**:eabf1970.
2. Marsh SE, Walker AJ, Kamath T, et al. Dissection of artifactual and confounding glial signatures by single-cell sequencing of mouse and human brain. *Nat Neurosci* 2022;**25**:306–16.
3. Liu R, Hu H, McNeil M, et al. Dormant Nfatc1 reporter-marked basal stem/progenitor cells contribute to mammary lobuloalveoli formation. *iScience* 2022;**25**:103982.
4. Peng L, Wang F, Wang Z, et al. Cell-cell communication inference and analysis in the tumour microenvironments from single-cell transcriptomics: data resources and computational strategies. *Brief Bioinform* 2022;**23**:bbac234.
5. Tian Y, Carpp LN, Miller HE, et al. Single-cell immunology of SARS-CoV-2 infection. *Nat Biotechnol* 2022;**40**:30–41.
6. Shen L, Liu F, Huang L, et al. VDA-RWLRLS: an anti-SARS-CoV-2 drug prioritizing framework combining an unbalanced bi-random walk and Laplacian regularized least squares. *Comput Biol Med* 2022;**140**:105119.
7. Stoeckius M, Hafemeister C, Stephenson W, et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* 2017;**14**:865–8.
8. Peterson VM, Zhang KX, Kumar N, et al. Multiplexed quantification of proteins and transcripts in single cells. *Nat Biotechnol* 2017;**35**:936–9.
9. Todorovic V. Single-cell RNA-seq—now with protein. *Nat Methods* 2017;**14**:1028–9.
10. Mimitou EP, Lareau CA, Chen KY, et al. Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nat Biotechnol* 2021;**39**:1246–58.
11. Lareau CA, Ludwig LS, Muus C, et al. Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nat Biotechnol* 2021;**39**:451–61.
12. Mimitou EP, Cheng A, Montalbano A, et al. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat Methods* 2019;**16**:409–12.
13. Tang L. Arsenal of single-cell multi-omics methods expanded. *Nat Methods* 2021;**18**:858–8.
14. Stuart T, Butler A, Hoffman P, et al. Comprehensive integration of single-cell data. *Cell* 2019;**177**:1888–1902.e1821.
15. Argelaguet R, Arnol D, Bredikhin D, et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol* 2020;**21**:111.
16. Wang X, Sun Z, Zhang Y, et al. BREM-SC: a bayesian random effects mixture model for joint clustering single cell multi-omics data. *Nucleic Acids Res* 2020;**48**:5814–24.
17. Singh R, Hie BL, Narayan A, et al. Schema: metric learning enables interpretable synthesis of heterogeneous single-cell modalities. *Genome Biol* 2021;**22**:131.
18. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell* 2021;**184**:3573–3587.e3529.
19. Hu H, Liu R, Zhao C, et al. CITEMO(XMBD): a flexible single-cell multimodal omics analysis framework to reveal the heterogeneity of immune cells. *RNA Biol* 2022;**19**:290–304.
20. Gayoso A, Steier Z, Lopez R, et al. Joint probabilistic modeling of single-cell multi-omic data with TotalVI. *Nat Methods* 2021;**18**:272–82.

21. Lotfollahi M, Litinetskaya A, Theis FJ. Multigrade: single-cell multi-omic data integration. *BioRxiv* 2022; 2022.2003.2016.484643.
22. Maitra C, Seal DB, Das V, et al. UMINT: unsupervised neural network for single cell multi-omics integration. *BioRxiv* 2022; 2022.2004.2021.489041.
23. Cao Z-J, Gao G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nat Biotechnol* 2022;**40**:1458–66.
24. Eraslan G, Simon LM, Mircea M, et al. Single-cell RNA-seq denoising using a deep count autoencoder. *Nat Commun* 2019;**10**:390.
25. Wang W, Huang Y, Wang Y, et al. Generalized autoencoder: a neural network framework for dimensionality reduction. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014, 496–503.
26. Devroye L. 1986 Sample-based non-uniform random variate generation. In *Proceedings of the 18th conference on Winter simulation (WSC '86)*. Association for Computing Machinery, New York, NY, USA, 260–65
27. Sun F, Sun J, Zhao Q. A deep learning method for predicting metabolite-disease associations via graph neural network. *Brief Bioinform* 2022;**23**:bbac266.
28. Kingma DP, Welling M. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 2013.
29. Pierson E, Yau C. ZIFA: dimensionality reduction for zero-inflated single-cell gene expression analysis. *Genome Biol* 2015;**16**:241.
30. Qiu P. Embracing the dropouts in single-cell RNA-seq analysis. *Nat Commun* 2020;**11**:1169.
31. Xu J, Cui L, Zhuang J, et al. Evaluating the performance of dropout imputation and clustering methods for single-cell RNA sequencing data. *Comput Biol Med* 2022;**11**:1169.
32. Bergen V, Lange M, Peidli S, et al. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* 2020;**38**:1408–14.
33. Burel JG, Pomaznoy M, Arlehamn CSL, et al. Circulating T cell-monocyte complexes are markers of immune perturbations. *Elife* 2019;**8**:e46045.
34. Caliński T, Harabasz J. A dendrite method for cluster analysis. *Commun Stat* 1974;**3**:1–27.
35. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 1987;**20**: 53–65.
36. Davies DL, Bouldin DW. A cluster separation measure. *IEEE Trans Pattern Anal Mach Intell* 1979;**PAMI-1**:224–7.
37. Steinley D. Properties of the Hubert-Arabie adjusted Rand index. *Psychol Methods* 2004;**9**:386–96.
38. Vinh NX, Epps J, Bailey J. 2010 Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance. *J. Mach. Learn. Res.* **11**:2837–54.
39. Nelson RW, Chen Y, Venezia OL, et al. SARS-CoV-2 epitope-specific CD4+ memory T cell responses across COVID-19 disease severity and antibody durability. *Sci Immunol* 2022;**7**: eabl9464.
40. Ssemaganda A, Nguyen HM, Nuhu F, et al. Expansion of cytotoxic tissue-resident CD8+ T cells and CCR6+ CD161+ CD4+ T cells in the nasal mucosa following mRNA COVID-19 vaccination. *Nat Commun* 2022;**13**:1–9.
41. Popescu I, Snyder ME, Iasella CJ, et al. CD4+ T cell dysfunction in severe COVID-19 disease is TNF $\alpha$ /TNFR1-dependent. *Am J Respir Crit Care Med* 2022;**205**:1403–18.
42. Ma Y, Huang Y, Zhao S, et al. Integrative genomics analysis reveals a 21q22.11 locus contributing risk to COVID-19. *Hum Mol Genet* 2021;**30**:1247–58.
43. Borgoni S, Kudryashova KS, Burka K, et al. Targeting immune dysfunction in aging. *Ageing Res Rev* 2021;**70**:101410.
44. Meckiff BJ, Ramírez-Suástegui C, Fajardo V, et al. Imbalance of regulatory and cytotoxic SARS-CoV-2-reactive CD4+ T cells in COVID-19. *Cell* 2020;**183**:1340–1353. e1316.
45. Fritsch RD, Shen X, Sims GP, et al. Stepwise differentiation of CD4 memory T cells defined by expression of CCR7 and CD27. *J Immunol* 2005;**175**:6489–97.
46. Bacher P, Rosati E, Esser D, et al. Low-avidity CD4+ T cell responses to SARS-CoV-2 in unexposed individuals and humans with severe COVID-19. *Immunity* 2020;**53**:1258–1271. e1255.
47. Liu X, Sun W, Ma W, et al. Smoking related environmental microbes affecting the pulmonary microbiome in Chinese population. *Sci Total Environ* 2022;**829**:154652.
48. Modabber Z, Shahbazi M, Akbari R, et al. TIM-3 as a potential exhaustion marker in CD4+ T cells of COVID-19 patients. *Immun Inflamm Dis* 2021;**9**:1707–15.
49. Zheng H-Y, Zhang M, Yang C-X, et al. Elevated exhaustion levels and reduced functional diversity of T cells in peripheral blood may predict severe progression in COVID-19 patients. *Cell Mol Immunol* 2020;**17**:541–3.
50. Zago CA, Jacob CMA, de Albuquerque Diniz EM, et al. Autoimmune manifestations in SCID due to IL7R mutations: Omenn syndrome and cytopenias. *Hum Immunol* 2014;**75**: 662–6.
51. Meyer A, Parmar PJ, Shahrara S. Significance of IL-7 and IL-7R in RA and autoimmunity. *Autoimmun Rev* 2022;**21**:103120.
52. Oliveira ML, Veloso A, Garcia EG, et al. Mutant IL7R collaborates with MYC to induce T-cell acute lymphoblastic leukemia. *Leukemia* 2022;**36**:1533–40.
53. Currier R, Puck JM. SCID newborn screening: what we've learned. *J Allergy Clin Immunol* 2021;**147**:417–26.
54. Soskic B, Cano-Gamez E, Smyth DJ, et al. Immune disease risk variants regulate gene expression dynamics during CD4+ T cell activation. *Nat Genet* 2022;1–10.
55. Al Barashdi MA, Ali A, McMullin MF, et al. Protein tyrosine phosphatase receptor type C (PTPRC or CD45). *J Clin Pathol* 2021;**74**: 548–52.
56. Webb GJ, Hirschfield GM, Lane PJ. OX40, OX40L and autoimmunity: a comprehensive review. *Clin Rev Allergy Immunol* 2016;**50**: 312–32.
57. Mousavi SF, Soroosh P, Takahashi T, et al. OX40 costimulatory signals potentiate the memory commitment of effector CD8+ T cells. *J Immunol* 2008;**181**:5990–6001.
58. Soroosh P, Ine S, Sugamura K, et al. OX40-OX40 ligand interaction through T cell-T cell contact contributes to CD4 T cell longevity. *J Immunol* 2006;**176**:5975–87.
59. So T, Ishii N. The TNF-TNFR family of co-signal molecules. *Adv Exp Med Biol* 2019;**1189**:53–84.
60. Li X, Zhang P, Yin Z, et al. Caspase-1 and Gasdermin D afford the optimal targets with distinct switching strategies in NLRP1b Inflammasome-induced cell death. *Research (Wash D C)* 2022;**2022**:9838341.
61. Xu F, Yin Z, Zhu L, et al. Oscillations governed by the incoherent dynamics in necroptotic signaling. *Front Phys* 2021;**9**: 726638.
62. Li X, Zhong CQ, Wu R, et al. RIP1-dependent linear and nonlinear recruitments of caspase-8 and RIP3 respectively to necrosome specify distinct cell death outcomes. *Protein Cell* 2021;**12**:858–76.