Biological network inference from single-cell multi-omics data using heterogeneous graph transformer

Anjun Ma[1,2,*], Xiaoying Wang[3,*], Cankun Wang[1], Jingxian Li[3], Tong Xiao[2], Juexing Wang[4], Yuzhou Chang[1,2], Yang Li[1], Yutao Liu[3], Shaopeng Gu[1], Duolin Wang[4], Yuexu Jiang[4], Jinpu Li[4], Li Su[4], Zihai Li[2], Bingqiang Liu[3,$], Dong Xu[4,$], Qin Ma[1,2,$]

[1] Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, OH, 43210, USA,

[2] Pelotonia Institute for Immuno-Oncology, The Ohio State University Comprehensive Cancer Center, Columbus, OH, 43210, USA,

[3] School of Mathematics, Shandong University, Jinan, Shandong, 250100, China,

[4] Department of Electrical Engineering and Computer Science, and Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, MO 65211, USA.

* These authors contributed equally to the paper as first authors

$ To whom correspondence should be addressed

Abstract

We present DeepMAPS, a deep learning platform for cell-type-specific biological gene network inference from single-cell multi-omics (scMulti-omics). DeepMAPS includes both cells and genes in a heterogeneous graph to infer cell-cell, cell-gene, and gene-gene relations simultaneously. The graph attention neural network considers a cell and a gene with both local and global information, making DeepMAPS more robust to data noises. We benchmarked DeepMAPS on 18 datasets for cell clustering and network inference, and the results showed that our method outperforms various existing tools. We further applied DeepMAPS on a case study of lung tumor leukocyte CITE-seq data and observed superior performance in cell clustering, and predicted biologically meaningful cell-cell communication pathways based on the inferred gene networks. To improve the feasibility and ensure the reproducibility of analyzing scMulti-omics data, we deployed a webserver with multi-functions and various visualizations. Overall, we valued DeepMAPS as a novel platform of the state-of-the-art deep learning model in the single-cell study and can promote the use of scMulti-omics data in the community.

Keywords: Single cell multimodal omics, heterogeneous graph transformer, graph neural network, multi-head attention framework, biological network, single-cell sequencing data analysis webserver

Main

Single-cell sequencing, such as single-cell RNA sequencing (scRNA-seq) and single-cell ATAC sequencing (scATAC-seq), reshapes the way of investigating cellular heterogeneity and brings novel insights in neuroscience, cancer research, immuno-oncology, drug response, etc[1, 2]. Individual single-cell modality can only reflect the snapshot of genetic features and partially depict the peculiarity of cells, leading to characterization biases in complex biological systems[2, 3]. To fully capture the intricacy of complex molecular mechanism and cellular heterogeneity, single-cell multi-omics (scMulti-omics) generate and quantify multiple modalities in the single-cell level simultaneously, and such measurements advance various biological studies equipped with robust computational analysis methods[4].

The existing tools for integrative analyses of scMulti-omics data, e.g., Seurat[5], MOFA+[6], Harmony[7], and totalVI[8], achieve reliable prediction of cell types and states, recover dropouts in single-modality data analysis, remove batch effects, reveal relationships or alignment among various modalities, and toward a mechanistic understanding of cell-type-specific gene regulations. However, the cross-talk among cells and different molecular modalities (e.g., genes and proteins) are usually missing in an independent hypothesis. Hence, these models lack the ability to infer the underlying biological networks of diverse cell types and have limited power to elucidate the response of these complex networks to external stimuli in specific cell types.

Recently, graph neural network (GNN) shows its unique strength in learning low-dimensional representations of individual cells by propagating neighbor cells' features and constructing cell-cell relations in a global cell graph[9-12]. For example, our in-house tool scGNN is a novel GNN model that has demonstrated the superior performance of cell clustering and gene imputation based on the large-scale scRNA-seq data[13]. Furthermore, a heterogeneous graph is a multi-relational model which provides a natural representation framework for integrating scMulti-omics data and learning the underlying cell-type-specific biological networks. Moreover, the recent development in the attention mechanism for modeling and integrating heterogeneous relationships make deep learning models explainable and enable the inference of cell-type-specific biological networks[14, 15].

To this end, we develop the first-of-its-kind model **DeepMAPS** (**Deep** learning-based **M**ulti-omics **A**nalysis **P**latform for **S**ingle-cell data), which is a heterogeneous graph transformer framework for biological network inference from scMulti-omics data. DeepMAPS formulates high-level representations of relations among cells and genes in a heterogeneous graph, with cells and genes as the two disjoint node sets in this graph. Projecting the features of genes and cells into the same latent space is an effective way to harmonize the imbalance of different batches and lies a solid foundation of cell clustering (i.e., node clustering) and the prediction of cell-gene and gene-gene relations

81 in a specific cell cluster (i.e., link prediction). Most importantly, the attention mechanism in
82 this transformer model enhances the biological interpretability and enables the
83 identification of important gene modules in each cell cluster. Overall, DeepMAPS is an
84 end-to-end and hypotheses-free framework and provides the first deep learning tool to
85 infer the cell-type-specific biological networks from scMulti-omics data.
86
87 In this study, we evaluated DeepMAPS on 18 scMulti-omics datasets (including
88 transcriptomic, epigenomic, and proteomic data) and our method outperformed existing
89 tools in terms of cell clustering accuracy and biological network inference (i.e., gene
90 association network and gene regulatory network). Specifically, DeepMAPS shows its
91 superior power in characterizing cellular heterogeneity since it can pass messages of
92 neighbor cells and genes to employ the identification of highly important genes in each
93 cell cluster. To further validate the biological insight inferred from DeepMAPS, we
94 performed a case study on a lung tumor leukocyte CITE-seq data. By jointly analyzing
95 gene expression and protein abundance, DeepMAPS accurately identified and annotated
96 13 cell types based on curated markers, which cannot be fully elucidated by a single
97 modality. We also proved that the embedding features identified in DeepMAPS capture
98 the true signals and amplify them when the original signals are weak. Besides, we also
99 identified biologically meaningful cell-cell communication pathways between dendritic cells
100 and tissue resident memory CD4 T cells based on the gene network inferred in the two
101 clusters. We deployed DeepMAPS as a code-free web portal along with Dockers, to
102 ensure the reproducibility of scMulti-omics data analysis and lessen the programming
103 burden for biologists who lack sufficient computational skills or resources (**Fig. 1**).
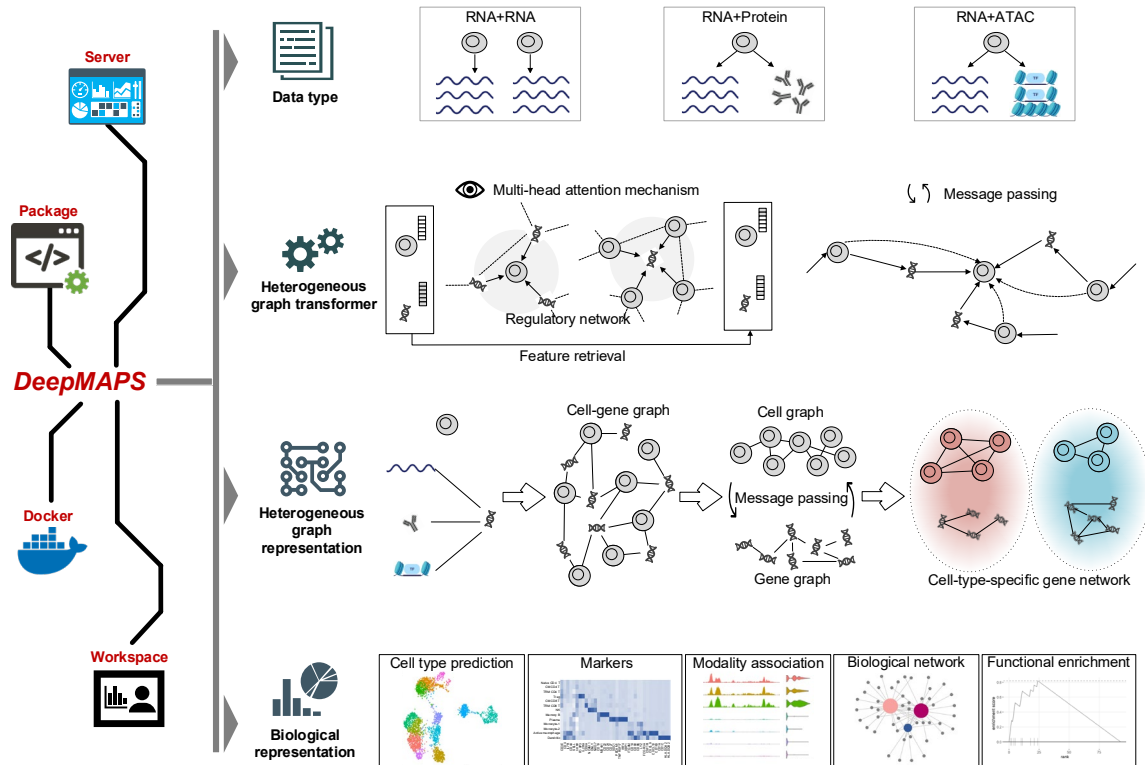
Fig. 1. DeepMAPS is a deep learning-based Multi-omics Analysis Platform for Single-cell data. It allows the joint analysis of multiple scRNA-seq, CITE-seq, and matched single-cell RNA and ATAC-seq data. The core method includes the representation of cell-gene relations via a heterogeneous graph and a transformer with a graph attention mechanism. DeepMAPS provides interactive and interpretable graphical representations to deliver cell clusters and various cell-type-specific biological networks depending on modality types. Eventually, DeepMAPS is delivered as a web portal to ensure robustness and reproducibility, along with a docker container. Workspace is committed to being provided for users for job saving and retrieval. DeepMAPS also supports diverse interpretations include but not limited to joint cell clustering, marker identification, modality associations, cell-type-specific biological network inference, and functional enrichment.

Results

Overview of DeepMAPS

There are five steps in DeepMAPS to fulfill the joint analysis of scMulti-omics data (**Fig. 2** and **Methods**). (*i*) Data is preprocessed by removing low-quality cells and modalities and then applied with different normalization methods according to specific data types (*ii*) An integrated cell-gene matrix will be generated by representing the combined activity of each gene in each cell. Different data integration methods are applied for different scMulti-omics data types. (*iii*) A heterogeneous graph transformer (**HGT**) model is built to jointly learn the low-dimensional embedding for cells and genes and generate an attention score to indicate the importance of a gene to a cell. (*iv*) Cell clustering and identification of functional gene modules in each cell cluster based on HGT-learned embeddings and

127 attention score. (*v*) Diverse biological networks, e.g., gene regulatory networks and gene
128 association networks, are inferred in each cell type.

130 To enable the simultaneous learning of joint representatives of cells and genes, we first
131 generate a cell-gene matrix integrating all heterogeneous information of the input scMulti-
132 omics data. A heterogeneous graph with cell nodes and gene nodes is then constructed,
133 where a cell-gene edge represents the integrated gene activity score in the matrix, and
134 the initial embedding of each node is learned from the gene-cell integrated matrix via a
135 two-layer GNN graph autoencoder. The entire heterogeneous graph is then sent to a graph
136 autoencoder to learn relations among cells and genes and update the embedding of each
137 node. Here, DeepMAPS adopts a heterogeneous multi-head attention mechanism to
138 models both overall topological information (global relationship) and neighbor message
139 passing (local relationship) on the heterogeneous graph.

141 In a HGT layer, each node (either a cell or a gene) will be considered as a target, and
142 DeepMAPS evaluates the importance of its neighbor nodes and the amount of information
143 that can be passed to the target based on the synergy of node embedding. As a result,
144 cells and genes with high positively correlated embeddings are more likely to exchange
145 messages with each other, thus tending to maximize the similarity and disagreement of
146 embeddings. To make the unsupervised training process feasible on a large
147 heterogeneous graph, DeepMAPS is first performed on 50 subgraphs subtracted from the
148 heterogeneous graph, covering a minimum coverage of 50% of all nodes, to train a shared
149 parameters between different nodes, which is later been used for testing the whole graph.
150 As an important training outcome, an attention score will be given to represent the
151 importance of a gene to a cell. A gene with high attention to a cell implies that the gene is
152 of relatively much importance for defining cell identity and characterizing cell heterogeneity.
153 This insight will lead to reliable gene association networks in each cell cluster as the final
154 output of DeepMAPS.

156 We applied the graph-based invariant-sampling method to make DeepMAPS feasible for
157 handling large-scale heterogeneous graphs[16]. The basic idea is to keep a separate node
158 budget for each node type (i.e., cell node and gene node) and to sample an equal number
159 of nodes per type with a sampling strategy to reduce variance. For an already sampled
160 node, we will add all its direct neighbors into the corresponding budget and add its
161 normalized degree to these neighbors, which will then be used to calculate the sampling
162 probability. Such normalization is equivalent to accumulating the random walk probability
163 of each sampled node to its neighborhood, avoiding the sampling being dominated by
164 high-degree nodes. The sampler constructs a number of small subgraphs from the given
165 giant graph, and these subgraphs can be fed in batches with multiple GPUs. These
166 distributed training results will be collected to build the whole graph with representations
167 on the head node.

168

169 The gene-gene and gene-cell relations predicted via the HGT model can only depict the
170 co-expression correlations as well as the gene activity in cells in the dataset rather than
171 reflect the gene module specificity in a cell cluster. Therefore, we build a Steiner Forest
172 Problem (SFP) model on a sparse heterogeneous graph to identify genes with higher
173 attention scores and similar embedding features uniquely in a cell cluster. We firstly
174 construct a sparse weighted heterogeneous graph based on the gene-gene and gene-cell
175 relations, and then build an SFP model based on this heterogeneous graph and cell
176 clusters. The gene-gene and gene-cell relations in the optimized solution of the SFP model
177 mirror the co-expression relations among genes and the specificity of genes to a cell type.
178 A gene network established from SFP contains genes that are highly associated based on
179 their gene activities and are of the most important in characterizing the identity of that cell
180 cluster, which is considered to be cell-type-active.

181
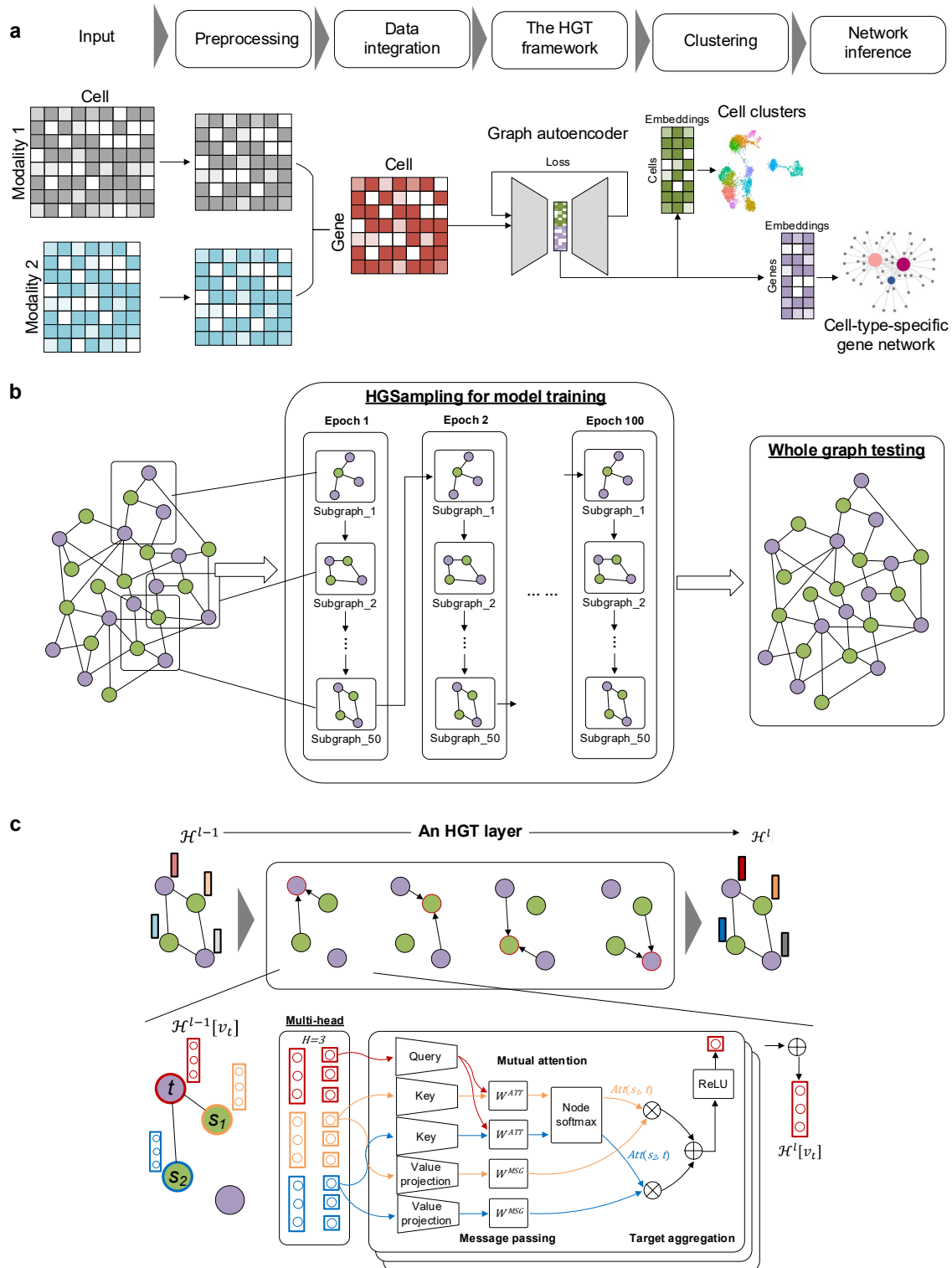
183   **Fig. 2**: DeepMAPS framework. (**a**) The five major steps and graphical illustration in DeepMAPS.
184   (**b**) Relations in the heterogeneous graph of cells and genes will be learned in an HGT-based graph
185   autoencoder. The hyperparameters are first trained in 50 subgraphs and 100 epochs and then
186   applied to the testing in the whole graph. (**c**) For each subgraph training and the whole graph testing,

187 multiple HGT layers are applied. A toy example with two cells (purple) and two genes (green) are
188 shown for one HGT layer. The embeddings of the target cell (red) and two neighbor genes (orange
189 and blue) are separated evenly into three heads. For each head, the HGT layer calculates the
190 attention score of neighbor genes to the target cell and updates the target's embedding of this head.
191 The entire target cell embedding is updated by concatenating all three heads to complete one HGT
192 layer.
193
194 ***DeepMAPS achieves superior performances in joint cell clustering and biological***
195 ***network inference from scMulti-omics data***
196 We collected 18 scMulti-omics datasets, including two multiple scRNA-seq data (Data 1-
197 2), eight CITE-seq data (Data 3-10), and eight matched scRNA-seq and scATAC-seq
198 (scRNA-ATAC-seq) data measured from the same cell (Data 11-18), to benchmark
199 DeepMAPS (**Supplementary Table S1**). Specifically, Data 1-2 and 17-18 have
200 benchmark annotations provided in the original manuscript. These data cover a number
201 of cells ranging from 549 to 30,672; an average read depth (consider scRNA-seq data
202 only) ranging from 2,933 to 645,526; a zero-expression rate (consider scRNA-seq data
203 only) from 71% to 97% (**Fig. 3a**).
204
205 To evaluate the performance of joint cell clustering in all three scMulti-omics data types,
206 we compare DeepMAPS with four benchmarking tools using default settings, including
207 Seurat, MOFA+, TotalVI, and Harmony (**Methods**), in terms of the Average Silhouette
208 Weight (ASW) (**Fig. 3a**) and Adjusted Rand Index (ARI) (**Fig. 3b**). DeepMAPS was trained
209 by each scMulti-omics data type and each dataset in an unsupervised way. The one set
210 of parameter was chosen as default for all datasets in the same data type based on the
211 grid optimization of hyper-parameter combinations (**Supplementary Table 3-5**). Results
212 clearly showed that, in all three scenarios of scMulti-omics data, DeepMAPS distinctly
213 outperformed the others in most cases. Note that, for Data 2 and 18, though DeepMAPS
214 did not achieve the best ASW than other tools, its performance regarding ARI comparison
215 to the benchmark label is the highest.
216
217 We showed UMAP results of cell clustering of Data 17, a cancer cell line data (n=549) with
218 benchmarked cell labels of scRNA-ATAC-seq data (**Fig. 3c**). By comparing with the
219 original cell line labels, we found that DeepMAPS is the only tool that accurately separate
220 each cell type with minimum mismatches (ARI=0.97), while for Seurat (ARI=0.88) and
221 MOFA+ (ARI=0.79), either PDX1 or PDX2 population was mistakenly divided into two
222 clusters and include more mismatches. For datasets without benchmark labels,
223 DeepMAPS also showed good UMAP visualization than other tools without having mixture
224 clusters or separated random cells (**Fig. 3d)**.
225
226 To evaluate the robustness of DeepMAPS, we performed a leave-out test on benchmark
227 datasets. For data with benchmark labels, we first filter cells by removing a cluster of cells

228   based on benchmark labels and then perform analysis using DeepMAPS and benchmark
229   tools; for other datasets, we removed cells based on clusters identified from each tool.
230   The results showed that DeepMAPS achieved superior performance compared to all other
231   benchmarking methods (**Fig. 3e**). Another test was performed on a series of data
232   simulated for different read depth rates (**Fig. 3f**). For each testing dataset, the clustering
233   results of DeepMAPS are consistent with high AWS or ARI, indicating that the message
234   passing and attention mechanism used in DeepMAPS helped maintain cell-cell relations
235   and tolerance to read depth.
236

237   We further evaluated the two kinds of biological networks the DeepMAPS can deliver. For
238   the gene association network (for all scMulti-omics data types), we considered using the
239   centrality scores and enriched pathways to compare DeepMAPS with IRIS3. IRIS3 is an
240   in-house tool for the identification of cell-type-specific regulon and gene regulatory network
241   construction from scRNA-seq data. It has superior performance than other public tools,
242   such as SCENIC. We also compared our results to the co-expression network constructed
243   from the whole gene list with significance cutoffs, which is a widely used way in single-cell
244   studies. Both the average closeness centrality and eigenvector centrality scores of
245   networks constructed in DeepMAPS showed significantly higher scores than the other two
246   methods using all 18 benchmark datasets (**Fig. 3g-h**). Moreover, for the gene regulatory
247   network constructed from scRNA-ATAC-seq data, we evaluated the number of significantly
248   enriched pathways in a TF-regulon and in a cell-type-specific regulon (**Fig. 3i-j**). The
249   results indicated that DeepMAPS is capable of constructing more compatible and
250   biologically reasonable gene networks in each cell type and outperformed the other
251   methods.
252

253
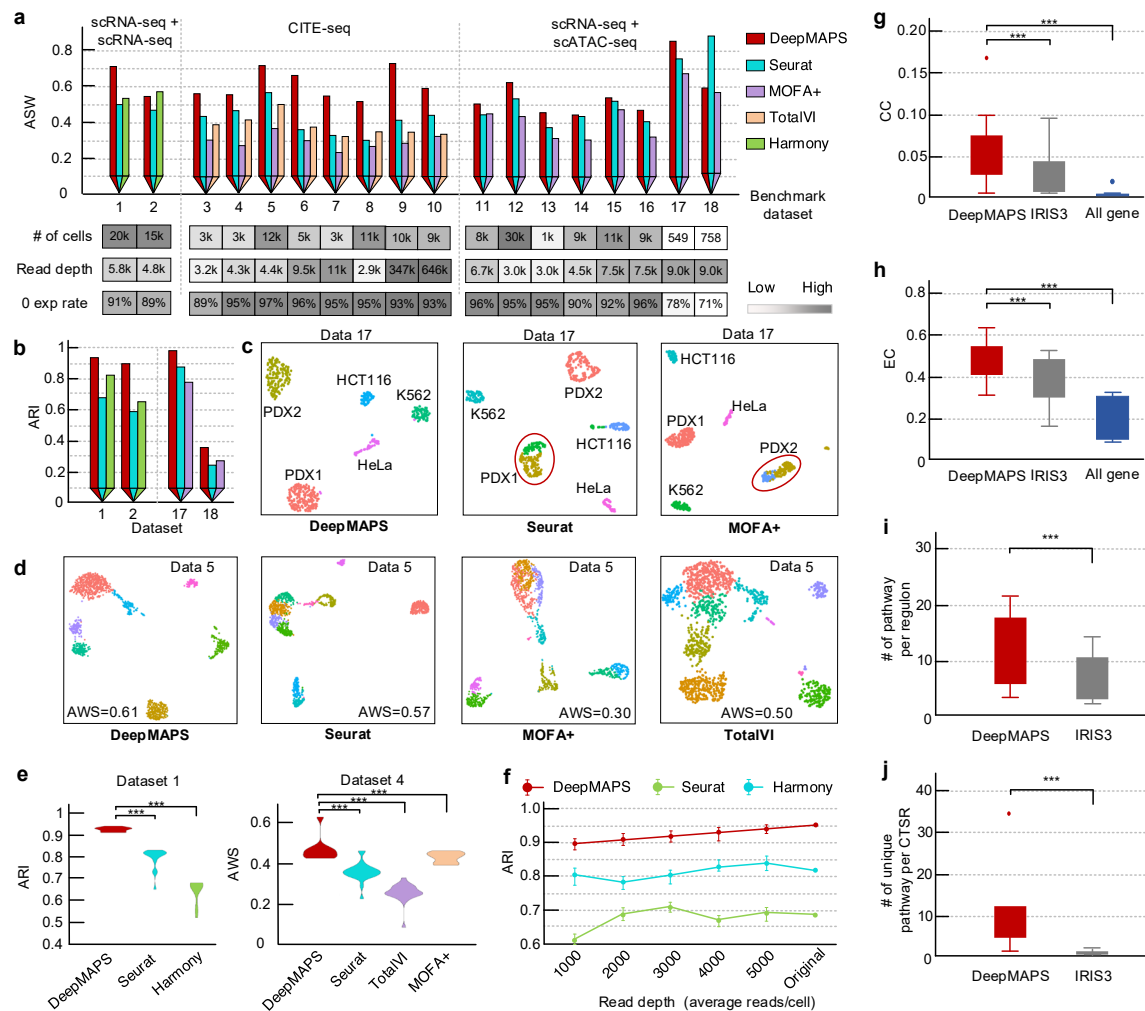
Fig. 3: Benchmarking of DeepMAPS in terms of cell clustering and biological network inference. (**a**) Benchmarking cell clustering results in all 18 datasets in terms of average silhouette weight without using benchmark cell labels. Due to the capability, different benchmarking tools were selected for the comparison. We also show heatmaps to indicate the number of total cells, average gene expression read depth per cell, and average RNA zero expression rate in each data. (**b**) Results comparison on four datasets with benchmarking cell labels in terms of adjusted rand index. (**c**) UMAP comparison of Data 17 (with benchmark labels) between DeepMAPS and other tools. Cluster labels were annotated based on cell correspondence to the original cell label. (**d**) UMAP comparison of Data 5 (without benchmark labels) between DeepMAPS and other tools. (**e**) Robustness test of DeepMAPS using cell cluster leave-out method for Data 1 (with benchmark label) and Data 4 (without benchmark label). Details can be found in the Method section. (**f**) Robustness test of DeepMAPS to different read depth on Data 1. (**g-h**) Evaluation and comparison of gene association network inference of DeepMAPS and other methods. Closeness centrality and betweenness centrality were used to indicate the compactness and connectivity of networks inferred from different methods. (**i-j**) Evaluation and comparison of gene regulatory network

270 (regulons) identified in DeepMAPS and IRIS3, based on the number of functional pathways
271 enriched in a regulon or cell-type-specific regulon.

272

273 ***DeepMAPS accurately identify cell types in PBMC and lung tumor immune CITE-***
274 ***seq data***

275 To demonstrate the joint representation of scMulti-omics in characterize cell identities, we
276 demonstrate a case study that applies DeepMAPS on a published PBMC and lung tumor
277 leukocytes CITE-seq data. The dataset includes RNA and protein measured on 3,485 cells.
278 We identified 13 cell clusters, including four CD4 T cell groups (Naïve, central memory
279 (CM), tissue resident memory (TRM), and regulatory (Treg)), two CD8 T cell groups (CM
280 and TRM), a natural killer group, a memory B cell group, a plasma cell group, two
281 monocyte groups, one tumor-associated macrophage group, and a dendritic cell (DC)
282 group, via DeepMAPS and annotated each cluster by visualizing expression levels of
283 curated maker genes and proteins (**Fig. 4a**). Compared to cell types identified using only
284 proteins or RNAs, we either isolated or accurately annotated cell populations that cannot
285 be characterized in the individual modality analysis. For example, the expression levels of
286 marker genes (CD4, IL7R, CD44, SELL, and CD69) in CD4 T cells are weak and
287 undistinguishable, thus only CM and naïve CD4 T cell groups can be recognized using
288 only scRNA-seq data (**Fig. 4b**), while the signal of marker proteins in CD4 T cells are much
289 stronger and differentiable. Another example lies in the recognition of tumor associated
290 macrophage in which the marker genes (CSF1R, CCR2, and MARCO) were identified to
291 be uniquely expressed in RNA, while no specific protein markers can be used to annotate
292 such cell group. Altogether, combining signals captured from both RNA and protein sides,
293 DeepMAPS successfully identified biologically reasonable and meaningful cell types in the
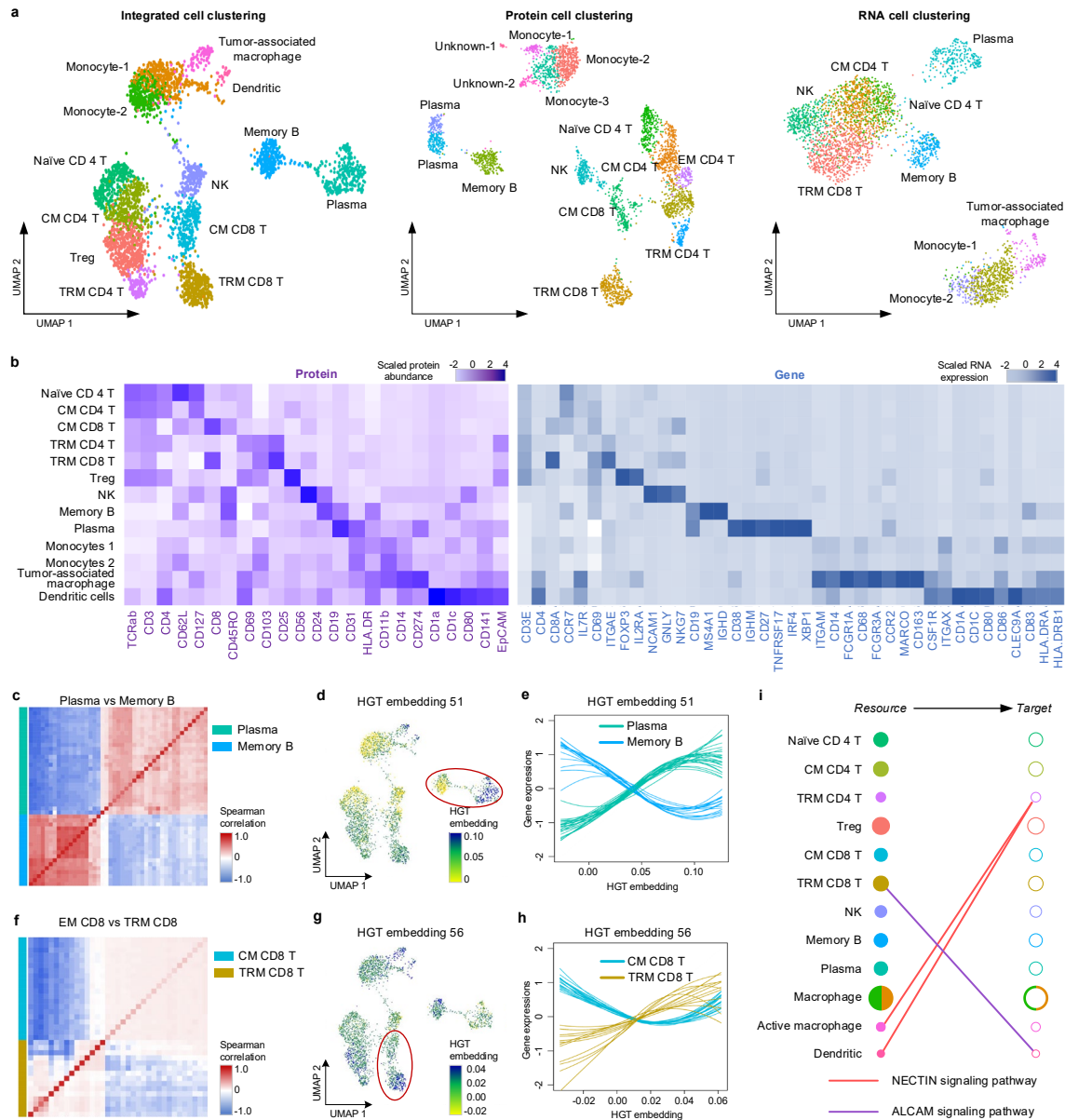294 CITE-seq data.

295

296 We then zoomed in to compare the modality correlation between two cell types. We used
297 the top differentially expressed genes and proteins between memory B cells and plasma
298 cells and performed hierarchical clustering of the correlation matrix. The result clearly
299 stratified these features into two anticorrelated modules: one associated with memory B
300 cells and the other with plasma cells (**Fig. 4c**). Furthermore, we found that the features in
301 the two modules significantly correlated with the axis of maturation captured by our HGT
302 embeddings. We observed that the 51th HGT embedding showed distinctive differences
303 between plasma cells and memory B cells (**Fig. 4d-e**). Similar findings were also observed
304 for the comparison of EM CD8 T cells and TRM CD8 T cells, which showed a much closer
305 relation when looking at expression correlations (**Fig. 4f**). Nevertheless, we can still find a
306 representative HGT embedding (56th) that maintains embedding signals to well separate
307 the two groups (**Fig. 4g-h**). These results point to a program of any two cell clusters
308 consisting of coordinated activation and repression of multiple genes and proteins, leading
309 to a gradual transition in cell state that can be captured by a specific dimension of the
310 DeepMAPS latent HGT space.

311

312   Based on the cell types and raw data of genes and proteins' expression, we inferred cell-

313   cell communication by using CellChat[17]. We constructed communication networks among

314   different cell types within multiple signaling pathways as **Fig. 4i.** We further applied

315   CellChat[17] to find any ligand-receptor interactions that have been validated between any

316   cell types we identified. We observed an ALCAM signaling pathway existing between TRM

317   CD4 T cells and DCs in the lung cancer tumor microenvironment (TME), of which DCs and

318   TRM CD4 T cells serve as the major source and ligand of CD166. Previous studies

319   showed that ALCAM on antigen-presenting DCs would interact with CD6 on T cell surface

320   and contribute to T-cell activation and proliferation[18-20]. For another example, we also

321   identified the involvement of NECTIN pathway during the interaction between the activated

322   macrophages (source) and TRM CD8 T cells (target). This is consistent with the previous

323   report that NECTIN (CD155) expressed on tumor-infiltrating macrophages could be

324   immunosuppressive when interacting with surface receptors on CD8+ T cells in the lung

325   cancer TME[21, 22].

326

Fig. 4: DeepMAPS identifies the heterogeneity in CITE-seq data of PBMC and lung tumor leukocytes. (**a**) UMAPs for DeepMAPS cell clustering results of integrating RNA and protein information, only using protein, and only using RNA. Cell clusters were annotated based on curated marker proteins and genes. (**b**) Heatmap of curated marker proteins and genes that determine the cell clustering and annotation. (**c**) Heatmap of correlation comparison of top differentially expressed genes and proteins in plasma cells and memory B cells. (**d**) UMAP is colored by the 51th embedding, indicating distinct embedding representations in plasma cells and memory B cells. (**e**) Expression of top differentially expressed genes and proteins in **c** as a function of the 51th embedding to observe the pattern relations between plasma cells and memory B cells. Each line represents a gene/protein, colored by cell types. For each gene, the line was drawn by a loess smoothing function based on the corresponding embedding and scaled gene expression in a cell. (**f-h**) Similar

339 visualization was given for the comparison of EM CD8 T cell s and TRM cd8 T cells in *c-e*. (**i**) Two
340 signaling pathways, NECTIN and ALCAM, were shown to indicate the predicted cell-cell
341 communications between two cell clusters.

342

343 ***DeepMAPS provides a multi-functional and user-friendly web portal for analyzing***
344 ***scMulti-omics data***
345 Researchers who lack sufficient computational skills prefer to use webservers or dockers
346 to lessen the programming burden of data analysis, and hence, a code-free and interactive
347 platform for single-cell sequencing data analysis are urgently needed in the public domain.
348 Considering the complexity of single-cell sequencing data, more and more webservers
349 and dockers have been developed in the past three years[23-35] (**Supplementary Table 6**).
350 However, most of these tools only provide minimal and basic functions such as cell
351 clustering and differential gene analysis, and do not support the joint analysis of scMulti-
352 omics data, especially lack sufficient support for biological network inference. To this end,
353 we deliver DeepMAPS as the-first-of-its kind web portal to support online and code-free
354 computational analysis for scMulti-omics data. The webserver supports the analysis of
355 multiple RNA-seq data, CITE-seq data, and scRNA-ATAC-seq data using DeepMAPS.
356 Other methods, e.g., Seurat, are also included as an alternative use for the users'
357 convenience. Three major steps, data preprocessing, cell clustering and annotation, and
358 network construction, are included in the server. The DeepMAPS server supports real-
359 time computing and interactive graph representations. Users may register for an account
360 to have their own workspace to store and share analytical results.

361

362 **Conclusion and discussion**
363 We highlighted DeepMAPS as the first deep learning framework that implements
364 heterogeneous graph representation learning and graph transformer in the study of
365 scMulti-omics data. By building a heterogeneous graph containing both cells and genes,
366 DeepMAPS identified the joint embedding of both sides simultaneously and enabled the
367 inference of cell-type-specific biological networks along with cell types in an intact
368 framework. The application of heterogeneous graph transformer takes the advantages
369 beyond graph nerual network that considering not only the message from neighbor cells
370 but also the attention of how such message should be passed to the target. In such a way,
371 the information training and learning process in a graph can be largely shortened to
372 consider cell impacts from a further distance.

373

374 While with the advantages and outperformed performances in analyzing scMulti-omics
375 data, there are still rooms to further improve the power of DeepMAPS. First of all, the
376 computational efficiency for super large datasets (e.g., more than 1 million cells) might be
377 a practical issue considering the complexity of the heterogeneous graph representation
378 (may contain billions of edges). Moreover, DeepMAPS is recommended to be run on
379 GPUs and supercomputers, which leads to a potential problem of reproducibility. Different

380 GPU models have different floating-point numbers that may influence the precision of loss
381 functions during the training process. That is to say, for different GPU models, DeepMAPS
382 may generate different cell clustering and network results, which is the main reason that
383 drives us to the development of webserver. Lastly, the current version of DeepMAPS is
384 based on a bipartite heterogeneous graph with genes and cells. Separate preprocessing
385 and integration steps are required to transfer different modalities all into genes and be
386 integrated into a unique cell-gene matrix. To fully achieve an end-to-end framework for
387 scMulti-omics analysis, the bipartite graph can be extended to a multipartite graph, where
388 each modality can be included as a node type. Such multipartite heterogeneous graph can
389 even include knowledgeable and biological information, such as known molecular
390 regulations, and more than two modalities all in one graph. However, by including more
391 node types, the computational burden will be increased geometrically, which requires a
392 dedicated discovery of model optimization in the future.
393
394 In summary, we evaluate our DeepMAPS as a pioneer study for the joint analysis of
395 scMulti-omics data and cell-type-specific biological network inference. It is likely to provide
396 new visions of deep learning deployment in single-cell biology. With the development and
397 maintainness of DeepMAPS webserver, our long-term goal is to create a deep learning-
398 based eco-community for AI-ready scMulti-omics data archiving, analyzing, visualizing,
399 and disseminating.
400
401

Methods

Data preprocessing and data integration

Multiple scRNA-seq data

DeepMaps takes the raw scRNA-seq gene expression profiles as input. Only genes that are expressed in more than 0.1% of total cells, and cells with a minimum of 0.1% genes expressed were kept. To integrate multiple scRNA-seq, we first reduce the dimension of multiple scRNA-seq in low-dimensional space by canonical correlation analysis (CCA) and search for mutual nearest neighbors (MNNs) in the shared low-dimensional space (integration method used in Seurat v3). It then calculated vector of each cell and for correcting gene expressions in different datasets. The output is an integrated matrix with combined cells from all datasets and shared genes with normalized and scaled expression values of x'_{ij} for gene *i* in cell *j*.

CITE-seq data

We also first removed low-quality genes and cells as described above. We applied log normalization on the RNA matrix and centered log-ratio (CLR) transformation on the protein matrix as below:

$$Lognormalize(x_{ij}) = \log\left(1 + \frac{x_{ij} \times 1e4}{\|\{x_i.\}\|}\right) \tag{1}$$

$$CLRnormalize(x_{ij}) = \log\left(1 + \frac{x_{ij}}{L_j}\right) \tag{2}$$

where x_{ij} represent for the expression of either a gene or protein in a cell. The top 2,000 highly variable genes in the RNA matrix were spliced with the protein matrix and performed a joint normalization as below:

$$x'_{ij} = exp\left(\frac{\sum_j \log(1 + x_{ij})}{\|K_i\|}\right), \tag{3}$$

where x_{ij} represents the expression for gene/protein *i* in cell *j*, and $\|K_j\|$ represents the sum of expressed values in of that gene/protein.

Matched scRNA-seq and scATAC-seq data (scRNA-ATAC-seq)

Data filtering and quality control were performed as usual. We first annotated peak regions in the scATAC-seq based on the method described in MAESTRO[36]. The regulatory potential R_{ij} of each $peak_j$ to each $gene_i$ are calculated independently by the exponential weight decay with the distance from the peak to the transcription start site (TSS):

$$R_{ij} = \begin{cases} 0, & d_{ij} > 150kb \text{ or } peak_i \text{ located in any nearby genes} \\ \frac{1}{l}, & peak_j \text{ located at the exons region of the } gene_i \\ 2^{-\frac{d_{ij}}{d_0}}, & else \end{cases} \tag{4}$$

where d_{ij} is the distance between the center of $peak_j$ and the TSS of $gene_i$. The default

435 d_0 is set to 10kb. If a $peak_j$ is located at the exon regions of the $gene_i$, the regulatory
436 potential R_{ij} will be normalized by the length of the exon region l. To ensure computational
437 efficiency, we set the R_{ij} as 0 if the $peak_j$ to TSS of $gene_i$ distance is over 150 kb. The
438 default parameters were applied. The gene regulatory activity A_{ik} of $gene_i$ in $cell_k$ is
439 defined as:

$$A_{ik} = \sum_i R_{ij} P_{jk}, \tag{5}$$

441 where P_{jk} is the binary value in the scATAC-seq count matrix of $peak_j$ in $cell_k$.

442 We assume that the activity of a gene to a cell is determined by both gene expression
443 activity and gene regulatory activity while with different contributions. Different than the
444 contribution weights determined directly based on the expression and chromatin
445 accessibility values in Seurat v4 (weighted nearest neighbor)[5], we hypothesized that the
446 contribution relation of gene expression and chromatin accessibility is dynamic to the
447 future state of the cell, which can be estimated by RNA velocity. The RNA velocity is
448 determined by the abundance of unspliced and spliced mRNA in a cell. The amount of
449 unspliced mRNA is determined by gene regulation and gene transcription rate (which is
450 considered to be the same among genes), and the amount of spliced mRNA is determined
451 by the difference between unsliced mRNA and degraded mRNA. We assume that for
452 genes with positive RNA velocities, there are higher potentials to drive genes to be
453 transcribed. Thus, its regulatory activity related to chromatin accessibility has a higher
454 influence than the gene expression in defining the overall gene activity in a cell of the
455 current snapshot; for genes with negative velocities, on the other hand, the transcription
456 rate tend to be decelerated, and regulatory activity has less influence to the cell than gene
457 expression activity. We define a gene activity score (GAS) of $gene_i$ in $cell_k$, which
458 integrates RNA and ATAC information, is defined as:

$$GAS_{ik} = \begin{cases} E_{ik} \cdot \sum_k E_{ik} + (1 + b^{v+}) A_{ik} \cdot \left((1 + b^{v+}) \sum_k A_{ik} \right), & for\ v_{ik} > 0 \\ E_{ik} \cdot \sum_k E_{ik} + (1 - b^{v-}) A_{ik} \cdot \left((1 + b^{v+}) \sum_k A_{ik} \right), & for\ v_{ik} < 0 \\ E_{ik} \cdot \sum_k E_{ik} + A_{ik} \cdot \sum_k A_{ik}, & for\ v_{ik} = 0 \end{cases} \tag{6}$$

460 where E_{ik} and A_{ik} represents the gene expression activity and gene regulatory activity,
461 respectively, of $gene_i$ in $cell_k$ which are normalized by rows. v_{ik} represents the RNA
462 velocity of $gene_i$ in $cell_k$, calculated by using CellRank[37]. The weighted of scATAC-seq is
463 defined as:

$$b^+ = \frac{\sqrt{\left(|V_{,k}^+| - rank(v_{ik}|V_{,k}^+) \right)^2 + \left(|V_{i,}^+| - rank(v_{ik}|V_{i,}^+) \right)^2}}{\sqrt{\left(|V_{,k}^+| - 1 \right)^2 + \left(|V_{i,}^+| - 1 \right)^2}} \tag{7}$$

$$b^- = \frac{\sqrt{\left(|V_{,k}^-| - rank(v_{ik}|V_{,k}^-)\right)^2 + \left(|V_{i,}^-| - rank(v_{ik}|V_{i,}^-)\right)^2}}{\sqrt{\left(|V_{,k}^-| - 1\right)^2 + \left(|V_{i,}^-| - 1\right)^2}} \tag{8}$$

where $V_{i,}^+$ is the positive velocity in $gene_i$, $V_{i,}^-$ is the negative velocity in $gene_i$. And $V_{,k}^+$ is the positive velocity in $cell_k$, $V_{,k}^-$ is the negative velocity in $cell_k$.

Construction of gene-cell heterogeneous graph

For any scMulti-omics data type, we now obtain a matrix that integrates information from both modalities, with columns as cells and rows as genes only. Values in the integrated matrix represent for either normalized gene expressions (for multiple scRNA-seq and CITE-seq) or GAS (for scRNA-ATAC-seq). Given an integrated matrix $X \in \mathbb{R}^{NXM}$ with N cells and M genes generated in the last step, we denote a heterogeneous graph $G = (V_C, V_G, E)$, where $V_C = \{c_1, c_2, \dots, c_N\} \in \mathbb{R}^N$ and $V_G = \{g_1, g_1, \dots g_M\} \in \mathbb{R}^M$ represents for all cells and genes in X, respectively, and $E = \{e_{ij}\}$ represents for the edge between c_i and g_j with the weight corresponding to gene expression of g_j in c_i. To learn an initial embedding of each node (c_i and g_j) in the matrix X, a feature autoencoder with two deep learning layers of dense networks in both encoder and decoder was used. The training objective of this feature autoencoder is to achieve a maximum similarity between X and reconstructed matrix X' by minimizing the mean squared error (MSE):

$$\sum (X - X_g)^2, and \sum (X - X_c)^2 \tag{9}$$

We denote the initial embedding for each c_i and g_j as $H_{c_i}^0$ and $H_{g_j}^0$.

Learning joint embeddings via a heterogeneous graph transformer

We propose an unsupervised HGT framework[14, 15] to learn joint embeddings of c_i and g_j. Given a heterogeneous bipartite graph $G = (V_C, V_G, E)$, DeepMAPS extracts all linked cell-gene node pairs which are denoted as (s, t), where t means the target node and s means the neighbor node of t. The training processes consist of the following four steps: 1) calculate multi-head attention; 2) pass heterogeneous message; 3) aggregate neighbors' information; 4) calculate loss function. We also denote the embedding of the $l\,th$ HGT layer as $\mathcal{H}^l$. To learn joint embeddings of c_i and g_j, we aggregate information from s to get a contextualized representation for t and simultaneously learn the representations of V_C and V_G. It is noteworthy that, to handle the heterogeneous relations in the graph, attention will be calculated via multiple heads, where each node type (cell node and gene node) has a unique head in attention and is linearly projected to a low dimensional space to maximally model the distribution differences.

1) Calculate multi-head attention.

We extract all linked node pairs, where a target node $v_t \in \{V_C, V_G\}$ is directly linked to its

501 source (neighbor) node $v_s = \{V_C, V_G\}$ through edge e. The contextualized representation of
502 v_t on the l th layer is denoted as $\mathcal{H}^l[v_t]$, which can be learned by its own presentation
503 $\mathcal{H}^{l-1}[v_t]$ and its neighbor $\mathcal{H}^{l-1}[v_s]$ from the $(l-1)^{th}$ layer, where $l \geq 1$. The overall model
504 is formulated as:

$$\mathcal{H}^l[v_t] \leftarrow \underset{\forall s,t \in V, \forall e \in E}{Aggregate}(\mathcal{H}^{l-1}[Attention(v_s, v_t, e) \cdot Message(v_s)]), \qquad (10)$$

506 where *Attention* estimates the importance of each neighbor; *Message* extracts the
507 information passed from the neighbors; *Aggregate* is the final step to aggregate the
508 neighborhood message by the attention weight.
509   The multi-head attention is proposed in the attention level to calculate h-head
510 attention for each edge $e = (v_s, v_t)$. Each target node t in the head $h \epsilon \{1, 2, ..., H\}$ was
511 mapped into a target-node vector $T^h(v_t)$ in each HGT layer via linear projection
512 $T^h_{\tau(t)}: R^d \to \times R^{\frac{d}{H}}$, where d is the dimension of initial node feature, H is the number of heads,
513 and $\frac{d}{H}$ is the feature dimension per head. Similarly, each neighbor node s in the head h
514 was mapped into a key vector $K^i(v_s)$ with a linear projection $K^h_{\tau(s)}: R^d \to \times R^{\frac{d}{H}}$. The
515 similarity between the queries and keys is measured (e.g., scaled dot product operator)
516 as attention.

$$T^h(v_t) = T^h_{\tau(t)}(\mathcal{H}^{(l-1)}[v_t]), \qquad (11)$$

$$K^h(v_s) = K^h_{\tau(s)}(\mathcal{H}^{(l-1)}[v_s]), \qquad (12)$$

519 Then we calculate the multi-head attention value for source node s to target node t by the
520 dot product. To maximize parameter sharing while still maintaining the specific
521 characteristics of different relations, we propose to parameterize weight matrices $W^{ATT}_{\phi(e)}$ of
522 the interaction operators. The i th head attention can be defined as:

$$ATThead^h(v_s, v_t, e) = \left(K^h(v_s)W^{ATT}_{\phi(e)}Q^h(v_t)^T\right) \cdot \frac{\mu\langle\tau(v_s), \tau(v_t)\rangle}{\sqrt{d}}, \qquad (13)$$

524 where T is the transposal function, μ is a prior tensor to denote the significance of each
525 edge e, serving as an adaptive scaling to the attention. The attention score in the h th
526 head in the l th layer is defined as:

$$Att(v_s, v_t, e) = \underset{\forall s \in V(t)}{Softmax}\left(\underset{H}{||} ATThead^h(v_s, v_t, e)\right). \qquad (14)$$

528
529 *2) Pass heterogeneous message.*
530 To alleviate the distribution differences of different types of nodes and edges, we
531 incorporate the types of edges into the message passing. The h th head message for each
532 edge (s, t) can be defined as:

533
$$M^h(v_s, v_t, e) = M^h_{\tau(s)}(\mathcal{H}^{l-1}[v_s])W^{MSG}_{\phi(e)} \tag{15}$$

534 where each source node s in the head i was mapped into a message vector by a linear

535 projection $M^h_{type(s)}: R^d \to \times\ R^{\frac{d}{H}}$, $W^{Msg}_e \in R^{\frac{d}{H} \times \frac{d}{H}}$ is a distinct edge-based matrix for each

536 edge, e is the edge type of the heterogeneous graph. After multi-head aggregate, the

537 degree of message passing can be defined as:

538
$$Msg(v_s, v_t, e) = \underset{H}{||} head^h(v_s, v_t) \tag{16}$$

539

540 *3) Aggregate neighbors' information.*

541 To obtain the represents of each node, we need to aggregate multi-head attention and

542 message. The attention vectors can be regarded as the weight for message representation.

543 The representation of target nodes $\mathcal{H}^l[v_t]$ can be updated as:

544
$$\mathcal{H}^l[v_t] = A_{Linear_{type(t)}}\left(\sigma\left(\underset{\forall s \in N(t)}{\bigoplus}\left(Att(v_s, v_t, e) \cdot Msg(v_s, v_t, e)\right)\right)\right) + \mathcal{H}^{l-1}[v_t] \tag{17}$$

545

546 *4) Calculate loss function.*

547 The original application of HGT was to solve node classification problems. Here, to train

548 cell and gene embeddings without supervised classification labels, we use a graph

549 autoencoder (GAE) framework. The whole HGT structure was deployed as an encoder in

550 the GAE, and we defined two embedding matrices E_c and E_g recording the trained

551 embeddings of cells and genes from the HGT encoder with both columns representing the

552 same embedding dimensions, and rows representing cells and genes, respectively. A

553 decoder was used to reconstruct the heterogeneous graph by the inner product of E_c and

554 E_g. The loss function of GAE is defined as:

555
$$loss = \sum_i [p(x_i)logp(x_i) - p(x_i)logq(x_i)], \tag{18}$$

556 where $p(x_i)$ and $q(x_i)$ represents the softmax operation for the i th row of E_c and E_g.

557

558 **HGT training on subgraphs**

559 To handle the efficiency and capability of applying HGT on such a giant heterogeneous

560 graph (tens of thousands of nodes and millions of edges), we performed model training

561 on subgraphs and multiple mini-batches based on the idea of HGSampling[14]. The core of

562 HGSampling is to sample heterogeneous subgraphs with similar proportions in different

563 type of nodes which can avoid sampling highly imbalanced subgraph in training process.

564 We sample n_batch number of subgraphs for each epoch training, and the one-hot of each

565 node will be put into the trained model to obtain all node embedding. Given a node t which

566 has been sampled, a dictionary $D[\tau]$ for each node type τ, we add all the first-neighbor

567 node of t into the corresponding $D[\tau]$ and add t's normalized degree to these neighbors to

568   calculate the sampling probability *p*.

569
$$p[\tau][s] = \frac{D[\tau][s]^2}{\left\|D[\tau]\right\|^2},\qquad(19)$$

570   where $p[\tau][s]$ is the sampling probability for each source node s of type τ, $D[\tau]$ is all node
571   for type τ with the normalized degree, $\|.\|$ is the 2-norm, $D[\tau][s]$ is the normalizer degree
572   for source node s of node type τ. Then we sampled all types of nodes according to the
573   probability in $D[\tau]$, and moved them out of $D[\tau]$. We repeated this sampling for 50 times
574   to obtain 50 subgraphs that maximize the coverage of the whole heterogeneous graph,
575   and each subgraph was trained with 100 epochs.
576

577   ***Cell clustering and cell-type-active gene association network prediction***
578   *Cell clustering*.
579   We applied the Louvain clustering method (igraph v1.2.7, R package) to predict cell
580   clusters on cell-embedding matrix E_c.
581

582   *Attention-based gene module detection.*
583   To infer the connection of genes and cell clusters, we extract the attention value of gene
584   g to cell c in head i through the step of multi-head attention calculation. We define the
585   importance I of g to c as:

586
$$I(g,c) = \sqrt{\sum_H ATThead^h(g,c)^2}\qquad(20)$$

587   We assign genes to each cell with a threshold of $mean_g\big(I(g,c)\big) + sd_g(I(g,c))$. The gene
588   g is considered to be one of the active genes in cell c, if the I is higher than the threshold.
589

590   *The Steiner Forest Problem (SFP) model*
591   We build an SFP model on a heterogeneous graph to extract the most critical gene-gene
592   and gene-cell relations contributing to the gene module specificity in a cell cluster. The
593   input of this model includes three parts:
594   1. Gene-gene relations defined by the embedding (resulted from GAE) Pearson's
595   correlation between genes (E),
596   2. Gene-cell relations are defined by the attention score of a gene to a cell (F),
597   3. A set of cell clusters, $\{V_i, i = 1, 2, \cdots, k\}$, predicted by the HGT model.
598

599   We define a weighted heterogeneous graph, $G = (U \cup V, E \cup F)$, in which nodes represent
600   genes (U) and cells $(V = \bigcup_i V_i)$, edges S represent both gene-gene (E) and gene-cell (F)
601   relations. We formulate this problem using a combinatorial optimization model defined as
602   below
603

$$\min_{S \subseteq E \cup F} \sum_{e \in S} w(e) \tag{21}$$

s.t.

$$r_S(u, v) = 1, \forall\, u, v \in V_i, i = 1, 2, \cdots, k. \tag{22}$$

where $r_S(u, v)$ is a binary indicator function to represent whether two nodes, u and v, are connected (1) or not (0) in the subgraph induced by S in G. We aim to identify the minimum weighted edge set, S, from the heterogeneous graph G, so that cells in the same cell type could be connected to each other via edges in S.

First, in view of the huge size of G, containing $|U \cup V|$ nodes and $|E \cup F|$ edges, we convert G into a sparser graph, G', by iteratively finding a global alignment between genes and cells based on the gene-cell edges, using the maximum matching theory[38]. In graph theory, a matching or independent edge set in an undirected graph is a set of edges without common nodes, and a maximum matching in a weighted graph is a matching M that yields the maximum sum of edge weights. To fulfill this task, we build a weighted bipartite graph, $G^B = (U \cup V, F)$, by only retaining the gene-cell edges, i.e., F. The objective is to identify an edge subset, M, to satisfy

$$\max_{M \subseteq F} \sum_{e \in M} w(e). \tag{23}$$

We calculate the M of G^B using the igraph R package[39]. Then, we remove the cell nodes incident by edges in M. Repeat the prediction of maximum matching and deletion of cell nodes incident to edges in previously identified maximum matchings, until there is no cell node in the remaining graph. We compute the union of all the matchings as F', and then construct $G' = (U \cup V, E \cup F')$. Finally, the weights of gene-cell edges, F', and gene-gene edges, E, are normalized by the following two functions, respectively.

$$\frac{\max(w(F')) - w(F')}{\max(w(F')) - \min(w(F'))} \tag{24}$$

$$\frac{\max(w(E)) - w(E)}{\max(w(E)) - \min(w(E))} \tag{25}$$

Second, we find the edge set, S, of the Steiner forest, T, as follows. To begin with, we utilize the igraph R package to calculate a minimum spanning forest (MSF)[38], T, of G'. A MSF means that each pair of nodes in the same connected component could be connected to each other. Herein, we only need the edges to connect cell nodes belonging to the same cell type. Therefore, we iteratively remove the gene nodes with degree one from T, until no gene node with degree one exists in T. Finally, we output the edge set of

639 T, i.e., S, as the solution to the SFP model.

640

641 For each connected component of the Steiner forest, the gene-gene edges denote the co-
642 expression relations among genes in the same module, while the set of gene-cell edges
643 represents the cell type specificity of this gene module, and this module is a cell-type-
644 active gene module.

645

646 ### Construct gene regulatory network from scRNA-ATAC-seq data

647 *Infer master TFs and GRNs in each cell type.*
648 To quantify the intensity of genes regulated by TFs, we design regulatory intensive (RI)
649 score, which can be decomposed into two components as: 1) the regulatory potential (R_{ij})
650 of peaks calculated in the preprocessing step, and 2) the binding affinity (BA) score of TFs
651 to the peaks. The TF binding profiles were obtained from JASPAR database. To reduce
652 false positives of binding site, we select significance binding sites with transformed *p*-
653 values for TF binding profile matches less than 1e-04. The BA score is the transformed
654 relative score which obtained from TF binding profiles. Then the RI score $RI_{L,i,k}$ of TF L
655 to the gene i in the cell k is defined as:

656
$$RI_{L,i,k} = \sum_p BA_{L,p} \cdot R_{ij} \qquad (26)$$

657 Master TFs are genes at the top of a gene regulation hierarchy, particularly in regulatory
658 pathways related to cell fate and differentiation. To infer cell type master TFs, we construct
659 cell-type-specific GRN with RI score as edges weight and calculate centrality which
660 reflects the importance of each node in the network to rank the TFs in each cell type. TFs
661 with high ranked are regarded as master TFs. Consider the RI score of TFs to genes,
662 eigenvector centrality which assigns relative scores to all nodes in the network based on
663 the concept that connections to high-scoring nodes contribute more to the score of the
664 node in question than equal connections to low-scoring nodes is applied to infer master
665 TFs. The eigenvector centrality of a node v in GRN can be defined as:

666
$$C_v = \alpha_{max}(v) \qquad (27)$$

667 Where α_{max} is the eigenvector corresponding to the largest eigenvalue of the weighted
668 adjacency matrix of a GRN.

669

670 *Identity differential regulon (CTSRs)*
671 To detect regulon associated with disease states, we identity CTSRs by logFC and
672 Wilcoxon rank-sum test. For a cell type active regulon, we define a regulon activity score
673 (RAS) as:

674
$$RAS(TF, CT) = \frac{\sum_i \sum_{k \in CT} GAS_{i,k} \cdot RI_{L,i,k}}{NG}, \qquad (28)$$

675 where L means TF, i means gene and k means cell. N is the cell number in a cell type, G
676 is the gene number in the regulon $L - CT$. Then we construct a RAS matrix with $|L - CT|$
677 rows, $|CT|$ columns. The significance of difference is calculated using the Wilcoxon rank-

678   sum test. If the BH-adjusted p-value is less than 0.05 between different cell clusters and
679   log fold change larger than 0.25, we consider the regulon is differentially active in this
680   cluster and defined as a CTSR.

681

682   ***Benchmarking quantification and statistics***

683   *Adjusted rand index (ARI)*

684   ARI is used to compute similarities by considering all pairs of the samples that are
685   assigned in clusters in the current and previous clustering adjusted by random permutation.
686   To calculate ARI, a contingency table is built to summarize the overlaps between the two
687   cell label lists with n elements (cells). Each entry denotes the number of objects in common
688   between the two label lists. The ARI score can be calculated as:

689
$$ARI = \frac{RI - E[RI]}{\max(RI) - E[RI]} \tag{29}$$

690   where RI is the unadjusted rand index which defined as:

691
$$RI = \frac{a + b}{C_n^2} \tag{30}$$

692

693   *Average Silhouette Weight (ASW)*

694   Different from ARI which requires known ground truth labels, silhouette refers to a method
695   of interpretation and validation of consistency within clusters of data. The silhouette value
696   is a measure of how similar an object is to its luster (cohesion) compared to other clusters
697   (separation). The silhouette ranges from −1 to +1, where a high value indicates that the
698   object is well matched to its cluster and poorly matched to neighboring clusters. The
699   silhouette score $s(i)$ can be calculated by:

700
$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} = \begin{cases} 1 - \dfrac{a(i)}{b(i)}, & if\ a(i) < b(i) \\ 0, & if\ a(i) = b(i) \\ \dfrac{b(i)}{a(i)} - 1, & if\ a(i) > b(i) \end{cases} \tag{31}$$

701   where $a(i)$ is the average distance between a cell i and all other cells in the same cluster,
702   and $b(i)$ be the average distance of i to all cell in the nearest cluster to which i does not
703   belong. We take the average silhouette of all cells in a cluster as the average silhouette
704   weight (ASW) to represent the whole cell cluster.

705

706   *Closeness centrality (CC)*

707   The closeness centrality (CC)[40] of a vertex u is defined by the inverse of the sum length
708   of the shortest paths to all the other vertices v in the undirected weighted graph. The
709   formulation is defined as:

710
$$CC(u) = \frac{1}{\sum_{i \neq j} d_w(u, v)} \tag{32}$$

711   Where $d_w(u, v)$ is the shortest weighted paths between u and v. If there is no path

712 between vertex u and v, the total number of vertices is used in the formula instead of the
713 path length. The CC is calculated by R package igraph with function igraph::betweenness.
714

715 *Eigenvector centrality (EC)*
716 Eigenvector centrality (EC)[41] scores correspond to the values of the first eigenvector of
717 the graph adjacency matrix. The EC score of vertex u is defined as:

718
$$EC(u) = x_u = c \sum_{v \in g} a_{uv} x_v \tag{33}$$

719 Where c is inverse of the eigenvalues of eigenvector $x = [x_1, x_2, \ldots, x_n]$, a_{uv} is the
720 weighted adjacent matrix of undirect graph g. The EC is calculated by R package igraph
721 with function igraph::evcent.
722

723 *Pathway enrichment test*
724 To evaluate the function of the regulatory network, we use pathway enrichment analysis[42]
725 to identify pathways that are significantly represented in each cell cluster active regulon,
726 and count the number of regulon-enriched pathways. The pathway enrichment analysis is
727 done by R package enrichR[43].
728

729 **Robustness evaluation**
730 *Cell cluster leave-out test*
731 For a benchmark dataset with a real cell type label, we removed all cells in one cell type
732 and ran DeepMAPS. We traverse all cell types (one at a time) to evaluate the robustness
733 with ARI. For data without benchmark labels, we removed cells in predicted cell clusters
734 from DeepMAPS and other benchmark tools, respectively.
735

736 *Read depth simulation test*
737 We performed a downsampling simulation for gene expressions to test the robustness of
738 DeepMAPS to read depth. Let matrix C be the $N \times M$ expression count matrix, where N
739 is the number of cells and M is the number of genes. Define the cell sequencing depths
740 $c_i = \sum_{j=1}^{M} C_{ij}$, i.e., the column sums of C. Thus, the average sequencing depth of the

741 experiment is $\bar{c} = \frac{\sum_{i=1}^{n} c_i}{N}$. Let $t < \bar{c}$ be our target downsampled sequencing depth and let

742 C^* be the $N \times M$ downsampled matrix. We perform the downsampling as follows:
743 For each spot $i = 1, \ldots, N$:

744 1) Define the total counts to be sampled in the cell i as $t_i = \frac{t \times c_i}{\bar{c}}$.

745 2) Construct the character vector of genes to be sampled as $G_i =$
746 $\{\underbrace{1, \ldots, 1}_{C_{i1}}, \underbrace{2, \ldots, 2}_{C_{i2}}, \ldots, \underbrace{M, \ldots, M}_{C_{iM}}\}$.

747 3) Sample t_i elements from G_i without replacement and define N_j as the number of
748 times gene j was sampled from G_i for $j = 1, \ldots, M$.

749   4) Let $C_{ij}^* = N_j$.

750 Using this method, the average downsampled sequencing depth is:

751
$$\overline{C}^* = \frac{\frac{t}{\bar{c}}c_1 + \frac{t}{\bar{c}}c_2 + \cdots + \frac{t}{\bar{c}}c_n}{N} = \frac{\frac{t}{\bar{c}}\sum_{i=1}^{N}c_i}{N} = \frac{t}{\bar{c}} \times \bar{c} = t \tag{34}$$

752 as desired. Note that, this method preserves the relative total counts of each cell, i.e.,
753 cells that had higher sequencing depths in the original matrix have proportionally higher
754 depths in the downsampled matrix.

755

756 ***Comparisons with existing tools***
757 In order to assess the performance of DeepMAPS alongside other proposed scMulti-omics
758 benchmark tools, we compare DeepMAPS with Seurat (v 3.2.3 and v 4.0.0,
759 https://github.com/satijalab/seurat), MOFA+ (v 1.0.0, https://github.com/bioFAM/MOFA2),
760 Harmony (v 0.1, https://github.com/immunogenomics/harmony), and TotalVI (v 0.10.0,
761 https://github.com/YosefLab/scvi-tools). Due to the integration capability, DeepMAPS was
762 compared with Seurat v 3.2.3 and Harmony on multiple scRNA-seq data, with Seurat
763 v4.0.0, MOFA+, and TotalVI on CITE-seq data, and with Seurat 4.0.0 and MOFA+ on
764 scRNA-ATAC-seq data. All benchmark tools used the default settings. We also evaluated
765 the performance of gene association network inference with IRIS3[11] and a normal gene
766 co-expression network inference method. Specifically, in IRIS3, cell-gene biclusters were
767 first identified based on the QUBIC2 algorithm. Cell-type-active gene modules were
768 identified in each cell cluster (using the same cell label predicted in DeepMAPS to ensure
769 the comparability) by performing a cell-wise hypergeometric enrichment. On the other
770 hand, all genes were selected to calculate a gene expression correlation score (Pearson's
771 correlation) between any pairs of two genes using cells in one cell cluster. Gene pair
772 expression correlations with a BH-adjusted p-value smaller than 0.05 were kept and used
773 to build the overall co-expression network in one cell cluster. Co-expressed sub gene
774 modules were inferred by performing Louvain clustering on the co-expression network.
775 For scRNA-ATAC-seq data, we compared regulon and cell-type-specific regulon inferred
776 from DeepMAPS with IRIS3 in terms of enriched biological pathways. Noted that, IRIS3
777 only supports regulon inference from scRNA-seq data based on de novo motif findings;
778 thus, here, we used the GAS matrix generated in DeepMAPS as an input of IRIS3.

779

780 ***DeepMAPS server construction***
781 DeepMAPS runs on an HPE XL675d RHEL system with 2 x 128-core AMD EPYC 7H12
782 CPU, 64GB RAM, and 2 x NVIDIA A100 40GB GPU. The backend is written in TypeScript
783 using the NestJs framework. Auth0 is used as an independent module to provide user
784 authentication and authorization services. Redis houses a queue of all pending analysis
785 jobs. There are two types of jobs in DeepMAPS: The stateful jobs are handled by the
786 Plumber R package to provide real-time interactive analysis; The stateless jobs, such as
787 CPU-bound bioinformatics pipelines and GPU training tasks that could take a very long

788   time, are constructed using Nextflow. All the running jobs are orchestrated using Nomad,
789   allowing each job to be assigned with proper cores and storage, as well as keeping the
790   jobs scalable based on the server load. The job results are deposited to a MySQL
791   database. The frontend is built with NUXT, Vuetify as the UI library, Apache ECharts, and
792   Cytoscape.js for data visualization. The frontend server and backend server are
793   communicated using REST API.
794

795   **Data availability**
796   All data used for benchmarking and case study are collected from the public domain and
797   can be retrieved using links or accession numbers provided in **Supplementary Tab. 1.**
798

799   **Code availability**
800   The source code of DeepMAPS Docker is freely available at (https://github.com/OSU-
801   BMBL/deepmaps). The DeepMAPS webserver is available at
802   https://bmblx.bmi.osumc.edu/.
803
804

References

805

1. Stuart, T. & Satija, R. Integrative single-cell analysis. *Nat Rev Genet* **20**, 257-272 (2019).

2. Ma, A., McDermaid, A., Xu, J., Chang, Y. & Ma, Q. Integrative Methods and Practical Challenges for Single-Cell Multi-omics. *Trends in Biotechnology* (2020).

3. Argelaguet, R., Cuomo, A.S.E., Stegle, O. & Marioni, J.C. Computational principles and challenges in single-cell data integration. *Nat Biotechnol* (2021).

4. S Teichmann, M.E. Method of the Year 2019: Single-cell multimodal omics. *Nat Methods* **17**, 1 (2020).

5. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e3529 (2021).

6. Argelaguet, R. et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biology* **21**, 111 (2020).

7. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods* **16**, 1289-1296 (2019).

8. Gayoso, A. et al. Joint probabilistic modeling of single-cell multi-omic data with totalVI. *Nature Methods* **18**, 272-282 (2021).

9. Li, Y. et al. Elucidation of Biological Networks across Complex Diseases Using Single-Cell Omics. *Trends Genet* (2020).

10. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083-1086 (2017).

11. Ma, A. et al. IRIS3: integrated cell-type-specific regulon inference server from single-cell RNA-Seq. *Nucleic Acids Res* **48**, W275-W286 (2020).

12. Han, P., Gopalakrishnan, C., Yu, H. & Wang, E. Gene Regulatory Network Rewiring in the Immune Cells Associated with Cancer. *Genes (Basel)* **8**, 308 (2017).

13. Wang, J. et al. scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. *Nature Communications* **12**, 1882 (2021).

14. Hu, Z., Dong, Y., Wang, K. & Sun, Y. in Proceedings of The Web Conference 2020 2704–2710 (Association for Computing Machinery, Taipei, Taiwan; 2020).

15. Wang, X. et al. in The World Wide Web Conference 2022–2032 (Association for Computing Machinery, San Francisco, CA, USA; 2019).

16. Zou, D. et al. in Advances in Neural Information Processing Systems 11247-11256 (2019).

17. Jin, S. et al. Inference and analysis of cell-cell communication using CellChat. *Nature Communications* **12**, 1088 (2021).

18. Ampudia, J. et al. CD6-ALCAM signaling regulates multiple effector/memory T cell functions. *The Journal of Immunology* **204**, 150.113-150.113 (2020).

19. Skonier, J.E. et al. Mutational analysis of the CD6 ligand binding domain. *Protein Engineering, Design and Selection* **10**, 943-947 (1997).

20. Gimferrer, I. et al. Relevance of CD6-mediated interactions in T cell activation and proliferation. *Journal of immunology (Baltimore, Md. : 1950)* **173**, 2262-2270

846   (2004).

847 21. Johnston, R.J., Lee, P.S., Strop, P. & Smyth, M.J. Cancer Immunotherapy and the
848   Nectin Family. *Annual Review of Cancer Biology* **5**, 203-219 (2021).

849 22. Li, X.-Y. et al. CD155 loss enhances tumor suppression via combined host and
850   tumor-intrinsic mechanisms. *J Clin Invest* **128**, 2613-2625 (2018).

851 23. Li, K. et al. cellxgene VIP unleashes full power of interactive visualization, plotting
852   and analysis of scRNA-seq data in the scale of millions of cells. *bioRxiv*,
853   2020.2008.2028.270652 (2020).

854 24. Pereira, W. et al. Asc-Seurat – Analytical single-cell Seurat-based web application.
855   *bioRxiv*, 2021.2003.2019.436196 (2021).

856 25. Gardeux, V., David, F.P.A., Shajkofci, A., Schwalie, P.C. & Deplancke, B. ASAP: a
857   web-based platform for the analysis and interactive visualization of single-cell
858   RNA-seq data. *Bioinformatics* **33**, 3123-3125 (2017).

859 26. Li, B. et al. Cumulus provides cloud-based data analysis for large-scale single-cell
860   and single-nucleus RNA-seq. *Nature Methods* **17**, 793-798 (2020).

861 27. Hillje, R., Pelicci, P.G. & Luzi, L. Cerebro: interactive visualization of scRNA-seq
862   data. *Bioinformatics* **36**, 2311-2313 (2019).

863 28. Prompsy, P. et al. Interactive analysis of single-cell epigenomic landscapes with
864   ChromSCape. *Nature Communications* **11**, 5702 (2020).

865 29. Bolisetty, M.T., Stitzel, M.L. & Robson, P. CellView: Interactive exploration of high
866   dimensional single cell RNA-seq data. *bioRxiv*, 123810 (2017).

867 30. Mohanraj, S. et al. CReSCENT: CanceR Single Cell ExpressioN Toolkit. *Nucleic
868   Acids Research* **48**, W372-W379 (2020).

869 31. Patel, M.V. iS-CellR: a user-friendly tool for analyzing and visualizing single-cell
870   RNA sequencing data. *Bioinformatics* **34**, 4305-4306 (2018).

871 32. Yousif, A., Drou, N., Rowe, J., Khalfan, M. & Gunsalus, K.C. NASQAR: a web-
872   based platform for high-throughput sequencing data analysis and visualization.
873   *BMC Bioinformatics* **21**, 267 (2020).

874 33. Zhu, Q. et al. PIVOT: platform for interactive analysis and visualization of
875   transcriptomics data. *BMC Bioinformatics* **19**, 6 (2018).

876 34. Innes, B. & Bader, G. scClustViz Single-cell RNAseq cluster assessment and
877   visualization [version 2; peer review: 2 approved]. *F1000Research* **7** (2019).

878 35. Granja, J.M. et al. ArchR is a scalable software package for integrative single-cell
879   chromatin accessibility analysis. *Nat Genet* **53**, 403-411 (2021).

880 36. Wang, C. et al. Integrative analyses of single-cell transcriptome and regulome
881   using MAESTRO. *Genome Biology* **21**, 198 (2020).

882 37. Lange, M. et al. CellRank for directed single-cell fate mapping. *bioRxiv*,
883   2020.2010.2019.345983 (2020).

884 38. Diestel, R., Schrijver, A. & Seymour, P.D. in MATHEMATISCHES
885   FORSCHUNGSINSTITUT OBERWOLFACH REPORT NO. 16/2007 (Citeseer,
886   2007).

39. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal, complex systems* **1695**, 1-9 (2006).

40. Li, G.S., Li, M., Wang, J.X., Li, Y.H. & Pan, Y. United Neighborhood Closeness Centrality and Orthology for Predicting Essential Proteins. *Ieee Acm T Comput Bi* **17**, 1451-1458 (2020).

41. Parisutham, N. & Rethnasamy, N. Eigenvector centrality based algorithm for finding a maximal common connected vertex induced molecular substructure of two chemical graphs. *J Mol Struct* **1244** (2021).

42. Reimand, J. et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc* **14**, 482-517 (2019).

43. Kuleshov, M.V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97 (2016).

Acknowledgments

Contributions

Q.M. and D.X. conceived the basic idea and designed the framework. X.W. and B.L. wrote the DeepMAPS code. C.W. built the backend and frontend server. S.G. designed dynamic figures on the server. Y.Liu carried out RNA velocity calculation. Y.Li designed the SFP model for gene module prediction. A.M, X.W., and J.L. carried out benchmark experiments. X.W., Y.C., and B.L. performed robustness tests. A.M., J.L., and T.X. carried out the case study. J.W., D.W., Y.J., J.L., and L.S. performed tool optimizations. A.M. lead the figure design and manuscript writing. All authors participated in interpretation and writing the manuscript.

Corresponding author

Correspondence to Prof. Qin Ma: Email: qin.ma@osumc.edu; Prof. Dong Xu: xudong@missouri.edu; Prof. Bingqiang Liu: bingqiang@sdu.edu.cn.

Ethics declarations
Competing interests
The authors declare no competing interests.