Cross-modality representation and multi-sample integration of spatially resolved omics data

Zhen Li[1], Xuejian Cui[1], Xiaoyang Chen[1], Zijing Gao[1], Yuyao Liu[1], Yan Pan[1], Shengquan Chen[2], and Rui Jiang[1,*]

[1] Ministry of Education Key Laboratory of Bioinformatics, Bioinformatics Division at the Beijing National Research Center for Information Science and Technology, Center for Synthetic and Systems Biology, Department of Automation, Tsinghua University, Beijing 100084, China

[2] School of Mathematical Sciences and LPMC, Nankai University, Tianjin 300071, China

[*] Corresponding author: ruijiang@tsinghua.edu.cn

Abstract

Spatially resolved sequencing technologies have revolutionized the characterization of biological regulatory processes within microenvironment by simultaneously accessing the states of genomic regions, genes and proteins, along with the spatial coordinates of cells, necessitating advanced computational methods for the cross-modality and multi-sample integrated analysis of spatial omics datasets. To address this gap, we propose PRESENT, an effective and scalable contrastive learning framework, for the cross-modality representation of spatially resolved omics data. Through comprehensive experiments on massive spatially resolved datasets, PRESENT achieves superior performance across various species, tissues, and sequencing technologies, including spatial epigenomics, transcriptomics, and multi-omics. Specifically, PRESENT empowers the incorporation of spatial dependency and complementary omics information simultaneously, facilitating the detection of spatial domains and uncovering biological regulatory mechanisms within microenvironment. Furthermore, PRESENT can be extended to the integrative analysis of horizontal and vertical samples across different dissected regions or developmental stages, thereby promoting the identification of hierarchical structures from a spatiotemporal perspective.

Introduction

Single-cell sequencing technologies have enabled the characterization of states and activities within individual cells from specific or even multi-omics perspectives, promoting molecular cell biology and disease research in areas such as cell atlas production, tumor immunology, and cancer genetics[1, 2]. However, an indispensable step of single-cell sequencing techniques involves the dissociation of cells from the tissue of interest, resulting in the loss of positional information[3]. Complex biological regulatory processes and cellular activities within tissue microenvironment are driven by sophisticated cell-cell interactions mediated by ligand-receptor partners, underscoring the importance of spatial distance between cells[4-6]. Recently, spatially resolved omics technologies have achieved simultaneous profiling of spatial locations and various omics layers, encompassing spatially resolved transcriptomics (SRT) [7, 8], spatial assay for transposase-accessible chromatin sequencing (ATAC-seq)[9, 10] and spatial multi-omics[11] (such as spatial CITE-seq[12], SPOTS[13] and MISAR-seq[14]), and accumulated substantial spatial omics sequencing data in repositories and databases[15, 16], thereby granting us a new perspective to decipher gene regulatory mechanisms in the context of tissue microenvironment.

However, spatial coordinate data and omics data exhibit significant modality differences, posing substantial challenges for the analysis of spatial omics data. Specifically, spatial coordinate data anchors the relative positions of cells within the tissue microenvironment, while omics data depicts the states of chromatin regions, genes and proteins involved in the gene regulatory processes within cells. Intended for this problem, cross-modality representation integrates the multi-modal data from spatially resolved datasets, including spatial coordinates, epigenomics, transcriptomics and proteomics, into a low-dimensional matrix. As a crucial step for spatial omics data analysis, cross-modality representation provides efficient and integrative features for various downstream applications and exploratory analyses[17-19], particularly the identification of spatial domains with distinct anatomical structures, cell-

type compositions, omics heterogeneity, and cell-cell interactions[20, 21], thereby facilitating the comprehension of tissue functions achieved through the coordination of diverse spatial domains.

Due to technical limitations such as size restrictions of captured area during data acquisition, every sample generated by spatial sequencing technologies typically focuses on a specific dissected region under a certain condition, while cross-modality representation methods could only be applied to a single sample because of spatial coordinate discrepancies between samples[22], impeding the exploration of spatially resolved samples across multiple dissected areas or developmental stages[23]. Therefore, integrative analysis of multiple samples across spatiotemporal scales, denoted as the multi-sample integration, is also critical for spatial omics data analysis by providing a comprehensive and systematic perspective for deciphering hierarchical and complex structures of cellular organizations[22-25]. Moreover, recent studies have highlighted that integrating disease and normal samples facilitates the identification of disease-specific domains and promotes the comprehensive exploration of disease onset and progression mechanisms[25, 26], further emphasizing the significance of incorporating multiple spatially resolved samples across various conditions.

Several computational methods have been proposed to address the exploration of SRT data[27, 28], encompassing cross-modality representation methods, e.g., SpaGCN[29], STAGATE[30] and PAST[17], along with some multi-sample integration methods such as STAligner[25], Stitch3D[31] and GraphST[23]. Nevertheless, compared with the rapid advancement in SRT data, the development of analytical methods for new spatial sequencing data, namely the spatial ATAC-seq[9, 10] and spatial multi-omics[11, 13, 14] data, is still stuck in the preliminary stage. Recently, SpatialGlue[32] was proposed as the first attempt for the cross-modality representation of spatial multi-omics data by utilizing an attentional feature fusion mechanism. Tian et al. proposed a dependency-aware generative variational autoencoder framework for the cross-modality representation of SRT, spatial ATAC-seq and spatial CITE-seq data[33]. However, there are still no

72 methods specially designed for the multi-sample integration of spatial ATAC-seq and spatial multi-omics
73 data, forcing researchers to resort to integrative algorithms intended for single-cell or SRT data, which
74 overlook the critical spatial information[3] and the significant gaps between different omics layers[34-37],
75 respectively. Additionally, most current algorithms are only capable of accomplishing either cross-modal
76 representation or multi-sample integration for specific spatial omics data, without addressing the common
77 key issues underlying these two tasks. Therefore, for the widespread application of spatial omics data,
78 there is an urgent need for a unified, advanced, and scalable framework for the cross-modality and multi-
79 sample integrated analysis of various spatial omics data.

80 Both cross-modality representation and multi-sample integration of spatially resolved omics data face
81 two common key issues. Firstly, different omics layers of spatial data, encompassing transcriptomics,
82 epigenomics and proteomics, exhibit distinct patterns and distributions[34-37]. For example, compared with
83 RNA-seq data, ATAC-seq datasets suffer from larger dimensionality, more severe sparsity and higher
84 dropout rates, while those of proteomics data sequenced by antibody-derived tags (ADT) are opposite,
85 requiring targeted modeling of each omics layer. Secondly, while spatial multi-omics technologies
86 provide a systematic landscape of gene regulatory processes within cellular organizations[14] using cross-
87 modal spatial coordinate and omics data, they also require analytical algorithms with enhanced
88 capabilities to simultaneously fuse complementary multi-omics signals and spatial information.
89 Furthermore, there are additional challenges for the integrated analysis of multiple spatial samples. On the
90 one hand, due to discrepancies in the dissection and placement, the spatial coordinate data between
91 different samples cannot be directly compared[22], necessitating effective mechanisms to model the cross-
92 sample relationships. On the other hand, a major challenge in multi-sample integration lies in the intricate
93 coupling between technical noise and biological variations[38, 39], requiring methods to eliminate batch
94 effects while preserving important biological signals.

To this end, we propose PRESENT, an effective and scalable contrastive learning framework developed for the cross-modality representation of s**P**atially **RES**olved omics data, encompassing **E**pigenomics, tra**N**scriptomics, and pro**T**eomics. Through comprehensive experiments on massive real-world datasets generated by diverse spatial sequencing technologies, including but not limited to spatial transcriptomics[7, 8], spatial epigenomics[9, 10] and spatial multi-omics[11, 13, 14], we demonstrated the superior performance of PRESENT in cross-modality representation. Specifically, PRESENT can simultaneously capture spatial dependency and complementary multi-omics information, obtaining interpretable cross-modality representations for various downstream analyses, particularly the spatial domain identification. PRESENT also offers the potential to incorporate various reference data to address issues related to the low sequencing depth and signal-to-noise ratio in spatial omics data. Furthermore, PRESENT can also be extended to the integrative analysis of multiple samples across different dissection regions and developmental stages, effectively eliminating redundant batch effects while maintaining shared and sample-specific biological variations as well as detecting hierarchical functional structures from a spatiotemporal perspective. In summary, as a scalable and integrative tool in the field of spatial omics data analysis, PRESENT is anticipated to provide great potential for systematic and comprehensive exploration of gene regulatory processes and cellular activities in the context of tissue microenvironment.

Results

Overview of PRESENT

Taking a spatial coordinate matrix and an arbitrary combination of epigenomic, transcriptomic, or proteomic count matrices as input (Methods, Fig. 1a), PRESENT is built on a multi-view autoencoder and extracts spatially coherent biological variations contained in each omics layer via an omics-specific encoder consisting of a graph attention neural network[40] (GAT) and a Bayesian neural network[41] (BNN).

118 The GAT module aims to incorporates spatial dependency and corresponding omics information, while
119 the BNN module, proved powerful in non-linear regression problems in original studies[41], brings the
120 potential to leverage reference data from existing sequencing data, according to our previous study[17].
121 Afterwards, the omics-specific embeddings obtained by encoders are fused by a cross-omics alignment
122 module which implements a multi-modal contrastive learning loss function, i.e., inter-omics alignment
123 (IOA) loss[42] (Methods). Additionally, to capture the distinct patterns of various omics layers, PRESENT
124 includes omics-specific decoders which model the counts of transcriptomic and epigenomic data via a
125 zero-inflation negative binomial (ZINB) distribution[43] and a zero-inflation Poisson (ZIP) distribution[44]
126 (Methods), respectively. Due to significantly lower dropout rates, dimensionality, and sparsity, ADT
127 expression matrix is reconstructed using a multi-layer perceptron (MLP) module.

128 Compared to cross-modality representation on a single spatial sample, integrative analysis of multiple
129 samples faces the common challenges such as the characterization of different omics layers and the cross-
130 modality integration of spatial and omics information. There, we extend the cross-modality representation
131 framework PRESENT intended for a single sample to the multi-sample integration scenario, thereby
132 obtaining PRESENT-BC (Methods, Fig. 1b). The two additional critical issues of multi-sample
133 integration are overcoming the spatial coordinate discrepancies between samples to capture cross-sample
134 relationships[22], as well as distinguishing between complexly coupled biological variations and technical
135 batch effects[23-25, 31]. PRESENT-BC adopts a two-stage strategy to address these issues. In the first stage,
136 PRESENT-BC has almost the same architecture as PRESENT and is pretrained using a cross-sample
137 graph and omics data from multiple samples. The intra-sample edges of the cross-sample graph are built
138 using the spatial coordinates, while the inter-sample connections are established using omics information
139 (Methods). Additionally, the input of decoders of PRESENT-BC concatenates the one-hot encoding of
140 batch indices of spots, facilitating the extraction of batch effect-free representations. In the second stage,

141 we introduce an inter-batch alignment loss[45] and a batch-adversarial learning strategy[46] to align spots
142 from multiple samples and eliminate redundant batch effects. We further propose an intra-batch
143 preserving loss to maintain biological signals within each batch while integrating multiple samples.
144 Additionally, we implement a cyclic iterative optimization process to refine the cross-sample graph based
145 on the joint representations newly obtained from the trained model in each epoch, as the graph not only
146 enables the characterization of cross-sample relationships, but also empowers the construction of positive
147 neighbor sets for spots in the inter-batch alignment loss (Methods).
148 The low-dimensional cross-modality representations obtained by PRESENT facilitate various
149 downstream applications (Fig. 1c), including visualization and spatial domain detection, and further
150 reveal biological implications in the identified domains. When extended to multi-sample integration,
151 PRESENT-BC not only promotes the integrated of horizontal slices from different dissection regions, but
152 also enables vertical integration of samples from different developmental stages (Fig. 1c), paving the way
153 to explore the hierarchical cellular activities from a spatiotemporal perspective.
154
155 **PRESENT facilitates accurate spatial domain identification in spatial ATAC-seq and SRT data**
156 We first accessed the performance of PRESENT on the spatial ATAC mouse embryo datasets which
157 consist of three developmental stages[10], including E15.5, E13.5 and E12.5. The datasets include six
158 samples, with each developmental stage containing two replicated samples, denoted as *-S1 and *-S2. We
159 evaluated the data quality of the six samples using logarithmic total fragment counts, denoted as
160 Log10(nFrags), and transcription starting site (TSS) enrichment scores. The former metric represents the
161 cellular sequencing depth of samples[47], while the latter reflected the signal-to-noise ratio of datasets[48]. As
162 shown in Fig. 2a, b, the samples of E15.5 stage, i.e., E15.5-S1 and E15.5-S2, achieved the highest
163 sequencing depth and signal-to-noise ratio, while the E13.5-S1 sample and E12.5-S2 sample suffered

164 from the lowest sequencing depth and signal-to-noise ratio, respectively.

165 Spatial domain identification, which aims to decipher regions with distinct anatomical structures,
166 cell-type compositions, and cell-cell interactions[20, 21], is a fundamental step in revealing gene regulatory
167 processes and tissue functions in spatial omics data[29, 30]. To evaluate the performance of spatial domain
168 identification based on latent representations, we applied PRESENT and other six baseline methods,
169 including three methods (SCALE[49], cisTopic[50] and SnapATAC2[51]) for single-cell ATAC-seq data and
170 three methods (SpaGCN[29], STAGATE[30] and PAST[17]) for SRT data, to the six mouse embryo samples.
171 Based on the latent representations from these methods, we implemented the Leiden algorithm[52] to
172 acquire the same number of clusters as the ground truth domains (Methods). Four commonly used
173 metrics[53, 54], including adjusted rand index (ARI), adjusted mutual information (AMI), normalized mutual
174 information (NMI) and homogeneity (Homo), were utilized to quantitatively evaluate the performance of
175 spatial domain detection. As shown in Fig. 2c, PRESENT achieved the best performance, with
176 improvements of more than 8.60%, 26.13%, 25.91%, and 22.94% compared to the second-best baseline
177 method across the four metrics, respectively. As a qualitative evaluation and intuitive illustration of the
178 spatial domain identification, we also demonstrated the uniform manifold approximation and projection
179 (UMAP) visualization[55] and spatial visualization of E15.5-S1 sample, colored by ground truth domains
180 (Fig. 2d, e) and the spatial clusters identified based on different methods (Fig. 2f). As shown in Fig. 2e,
181 baseline methods like cisTopic and STAGATE could only identify the forebrain domain from others.
182 Although SCALE, SnapATAC2, SpaGCN and PAST roughly recognized the subdomains of central
183 nervous system, i.e., forebrain, hindbrain and midbrain, the low-dimensional representations acquired by
184 them are decentralized and scattered, impeding their performance for spatial clustering (Fig. 2c, f). In
185 contrast, PRESENT effectively distinguished domains including the forebrain, hindbrain, midbrain, facial
186 & bone, muscle & heart, and meningeal & peripheral nervous system (PNS) 2. Similar analytical results

187 were also observed on the other five samples (Supplementary Figs. 1-6), exhibiting the potential of
188 PRESENT in cross-modality representation and downstream visualization and domain identification.
189 Despite the superiority of PRESENT across the six mouse embryo samples, low sequencing depth
190 and signal-to-noise ratio stills significantly constrained the effectiveness of the model, illustrated in Fig.
191 2a, b and Supplementary Fig. 7. Our previous study has systematically demonstrated that integrating
192 external reference data can effectively address such issues in SRT data[17]. Inspired by this, we collected a
193 single-nucleus ATAC-seq mouse embryo dataset involving cells from three developmental stages,
194 including E12, E13 and E15, which was generated in conjunction with the spatial ATAC mouse embryo
195 data[10]. We utilized the single-nucleus ATAC-seq dataset (denoted as PRESENT-SN) and the other five
196 spatial ATAC samples (denoted as PRESENT-SP) as the reference data for PRESENT when analyzing
197 each mouse embryo sample (Methods). As shown in Fig. 2g and Supplementary Fig. 8, the introduction
198 of prior information contained in reference data enabled enhanced performance of PRESENT, especially
199 in samples with lower sequencing depth and signal-to-noise ratio, such as the E13.5-S1 and E12.5-S2
200 samples. This finding was further validated by the UMAP visualization of the two low-quality samples
201 (Fig. 2h and Supplementary Figs. 3 and 6). Specifically, PRESENT-SN & SP successfully identified
202 spatial domains including liver, limb, forebrain, hindbrain and midbrain, which other baseline methods or
203 even PRESENT failed to distinguish. For systematic evaluation, we further increased the nose level of
204 datasets by converting lower fragment counts to zero with growing dropout rates, according to a study
205 that models the dropout events in single-cell data[56]. As shown in Fig. 2i, PRESENT-SN & SP gained
206 greater advantages as the noise level increased, highlighting the improved benefits of integrating prior
207 information from reference data with the growing dropout rates in spatial ATAC-seq datasets.
208 Furthermore, we also carried out comprehensive experiments on SRT axolotl brain datasets generated
209 by Stereo-seq technique, consisting of six samples from six developmental stages, including stage 44,

210 stage 54, stage 57, juvenile, adult and metamorphosis[7]. The performance of PRESENT was benchmarked
211 with several recently developed SRT methods, involving SpaGCN[29], STAGATE[30], PAST[17] and
212 SpatialPCA[18]. As shown in Supplementary Figs. 9-15, experimental results demonstrated the consistent
213 superiority of PRESENT in representation, visualization and spatial domain identification. In addition,
214 utilizing an SRT regenerated axolotl brain data at 20 days post-injury as reference data[7], PRESENT-SP
215 again gained promoted performance through leveraging prior information contained in reference data
216 (Supplementary Fig. 9). In summary, PRESENT not only enables effective cross-modality representation
217 and spatial domain identification in both spatial ATAC-seq and SRT data, but also empowers the
218 integration of reference data from various sources to address the low sequencing depth and signal-to-
219 noise ratio of spatially resolved datasets.
220

221 **PRESENT leverages complementary omics information from paired spatial RNA-ADT data**

222 Next, we collected a spatial RNA-ADT human lymph node dataset generated by 10x Genomics Visium
223 technology, an in-house data from SpatialGlue[32]. This dataset simultaneously accessed the spatial
224 locations, transcriptomic and proteomic expression of spots and was manually annotated by the clinician
225 group according to the histological image[32] (Fig. 3a). We benchmarked PRESENT with SpatialGlue[32], a
226 representation learning method specially designed for spatial multi-omics data, and several methods
227 developed for single-cell RNA-ADT paired sequencing data, including totalVI[57], CiteFuse[58] and Seurat-
228 v4[59]. As shown in Fig. 3b, PRESENT again achieved the best performance on all the four metrics for
229 spatial domain identification. The UMAP and spatial visualization results were consistent with the
230 quantitative results (Fig. 3d, e, f). Specifically, CiteFuse and Seurat-v4 obtained unsatisfactory
231 visualization results, where domains such as medulla cords and sinuses scattered. Although SpatialGlue
232 and totalVI achieved relatively better results, they failed to recognize capsule from pericapsular adipose

233 tissue. In contrast, PRESENT successfully distinguished the pericapsular adipose tissue (cluster 8),
234 capsule (cluster 1), follicle (cluster 3) and cortex (cluster2 and cluster 10). We also directly applied
235 PRESENT to a single omics layer of the human lymph node dataset, thus leveraging expression
236 information from only a single omics layer, denoted as RNA-only and ADT-only. As shown in Fig. 3c,
237 PRESENT, which utilized multiple omics layers (RNA & ADT), achieved improved performance
238 compared to the model that only leveraged RNA or ADT data, highlighting the capabilities of PRESENT
239 in integrating complementary omics information.

240 As the largest secondary lymphoid organ in the body, the spleen plays a significant role in
241 hematopoiesis and the clearance of erythrocytes, contributing significantly to immune functions[60]. We
242 further applied PRESENT to a SPOTS mouse spleen dataset[13], which measured spatial transcriptomic and
243 proteomic profiles simultaneously, to decipher the functionality of spleen in the context of cellular
244 organizations. Through cross-modality representation and spatial clustering, PRESENT identified four
245 main spatial domains (Fig. 3g), denoted as the B cell-enriched domain, the T cell-enriched domain and
246 two macrophage-enriched subdomains, i.e., Mac1-enriched and Mac2-enriched domains, according to our
247 differentially expressed analysis (Fig. 3h, i). The differentially expressed analysis compared the protein
248 expression level of each domain with those of the rest domains based on the Wilcoxon rank-sum test[61]
249 (Fig. 3h), identifying protein markers distinguishing the B cell-enriched domain (CD3, CD4, CD8), the T
250 cell-enriched domain (CD19, B220, lgD) and the macrophage-enriched domain (F4-80, CD163, CD68),
251 consistent with the original study[13]. Specifically, the Mac1-enriched and Mac2-enriched domains had
252 nearly identical protein markers (Fig. 3h), which corresponded to the macrophage markers identified in
253 the original study[13], thereby making it hard to distinguish the two subdomains purely by ADT data.
254 Therefore, we further conducted differentially expressed analyses on the transcriptomic data between
255 different domains (Fig. 3i). Analytical results demonstrated that Mac1-enriched domain could be marked

256 by genes including *Hbb-bt, Hbb-a1, Rsad2, Hbb-bs, Car2* and *Slc4a1*, while Mac-2 enriched domain
257 could be recognized with *Igkc, Jchain, Igha, Ighm, Gm2a* and *Cox8a.* The RNA expression levels
258 between the Mac1 and Mac2-enriched domains were also directly compared and numerous significantly
259 differentially expressed genes (DEGs) were found (Fig. 3j). In conclusion, despite the similarity of the
260 two macrophage-enriched subdomains in ADT expression levels, the distinction of transcriptomic
261 patterns between them could still be identified by PRESENT, emphasizing the ability of PRESENT in
262 leveraging complementary proteomic and transcriptomic information for effective domain identification.
263 To further decipher the functional distinction between the two macrophage-enriched subdomains, we
264 conducted Gene Ontology (GO) enrichment analysis based on the identified DEGs (Methods).
265 Specifically, for the Mac1-enriched domain, we enriched biological processes including erythrocyte
266 development, erythrocyte differentiation, erythrocyte homeostasis, myeloid cell development, myeloid
267 cell differentiation and myeloid cell homeostasis with false discovery rates (FDRs) of 1.29e-8, 5.48e-6,
268 6.70e-6, 3.59e-7, 8.71e-5 and 2.06e-5, respectively (Fig. 3k and Supplementary Table 1). Macrophages
269 are myeloid lineage cells which derive from bone marrow[62], reasonably explaining the enrichment of
270 development, differentiation and homeostasis of myeloid cell in the Mac1-enriched domain. In addition,
271 previous studies have stressed the importance of macrophages in the entire life of erythrocytes by
272 providing necessary factors for the survival, proliferation and development of erythrocytes[63, 64], agreeing
273 with the development, differentiation and homeostasis of erythrocyte found in the Mac1-enriched domain.
274 In comparison, the Mac2-enriched domain found biological processes including phagocytosis &
275 engulfment, plasma membrane invagination, membrane invagination, phagocytosis, humoral immune
276 response and lymphocyte mediated immunity with FDRs of 7.11e-5, 7.11e-5, 7.11e-5, 4.34e-4, 5.58e-4
277 and 5.94e-4, respectively (Fig. 3k and Supplementary Table 2). Phagocytosis and other related processes,
278 also called the process of 'cell eating', are commonly considered as classical functions of macrophages[65,

279 [66], validating the enrichment of macrophages in the Mac2-enriched domain. On the other hand, compared
280 with the Mac1-enriched domain, the Mac2-enriched domain was spatially close to the B cell-enriched and
281 T cell-enriched domains (Fig. 3g), thereby explaining the enrichment of lymphocyte mediated immunity
282 process[67]. In summary, PRESENT effectively integrates complementary omics information in spatial
283 RNA-ADT data, thereby facilitating accurate spatial domain identification and systematic exploration of
284 biological processes in the identified domains.
285
286 **PRESENT extracts cross-modality representations from parallel spatial RNA-ATAC datasets**
287 We further collected the MISAR-seq mouse brain datasets[14], which simultaneously measured the
288 epigenomics and transcriptomics, i.e., paired spatial RNA-ATAC profiles, of four mouse brain samples
289 from four developmental stages, involving E11.0, E13.5, E15.5 and E18.5. Compared with spatial-RNA-
290 ADT data, spatial RNA-ATAC datasets suffer from higher dimension and sparsity. We firstly compared
291 the performance of PRESENT in spatial domain identification with the spatial multi-omics method
292 SpatialGlue[32] and six methods designed for the representation of single-cell multi-omics data, including
293 MOFA+[68], scAI[69], scMDC[70], scMVP[19], Seurat-v4[59] and SnapATAC2[51]. As shown in Fig. 4a, PRESENT
294 achieved overall the best spatial clustering performance, illustrating the superiority of PRESENT in
295 extracting spatial multi-omics information into effective cross-modality embeddings for accurate spatial
296 domain identification. We also used PRESENT to obtain latent representations based on only ATAC data
297 or RNA data, denoted as ATAC-only and RNA-only respectively, which resulted in decreased
298 performance, highlighting the importance of integrating complementary multi-omics information in
299 deciphering biological variations in the context of cellular organizations (Fig. 4b).
300 Taking the E18.5 mouse brain sample with the most complex spatial structures as an example (Fig.
301 4c), we intuitively visualized the latent embeddings learned by PRESENT using heatmap (Fig. 4d). The

302 spots from different spatial domains could be successfully distinguished via the combination of several
303 latent features or even a single feature in the latent representation space, including large domains such as
304 cartilage 1-4, mantle zone of dorsal pallium (DPallm), ventricular zone of dorsal pallium (DPallv),
305 diencephalon & hindbrain and subpallium 1, and small domains like midbrain and thalamus. We further
306 demonstrated the UMAP visualization of latent representations obtained by different methods on the
307 E18.5 mouse brain sample (Fig. 4e). Baseline methods, including MOFA+, scAI, scMDC, scMVP and
308 Seurat-v4, could only roughly separate large domains such as DPallm, DPallv, diencephalon & hindbrain
309 and subpallium 1. SnapATAC2 could additionally recognize the cartilage 1 but still failed to distinguish
310 others. In comparison, SpatialGlue and PRESENT not only identified large spatial domains such as
311 diencephalon & hindbrain, subpallium 1, DPallm, DPallv, cartilage 1-4, but also accurately distinguished
312 two rare spatial domains, i.e., midbrain and thalamus. In addition, the visualization of spatial clusters
313 identified based on different methods consistently demonstrated the excellent domain identification
314 performance of PRESENT, shown in Fig. 4f, g. For example, PRESENT also successfully deciphered
315 most large domains, including DPallm (cluster 5), DPallv (cluster 6), subpallium 1 (cluster 11), cartilage
316 1 (cluster 1) and diencephalon & hindbrain (cluster 7), and the two smaller domains, namely midbrain
317 (cluster 9) and thalamus (cluster 13).

318 The UMAP and spatial visualization of other three samples of the MISAR-seq mouse brain dataset[14]
319 achieved similar results (Supplementary Figs. 16-19) and further highlighted the capabilities of
320 PRESENT in integrating complementary transcriptomic and epigenomic omics information for effective
321 cross-modality representation. Specifically, PRESENT utilizing only the transcriptomic data of E11.0
322 brain sample, denoted as RNA-only, successfully captured the difference between the muscle and the
323 cartilage 2 domain, while the model based solely on epigenomic data, depicted as ATAC-only, failed to do
324 so (red circles in Fig. 4h). On the other hand, compared with the failure of PRESENT using only

325 transcriptomic data (RNA-only) in the E15.5 brain sample, the model based on the epigenomic data
326 (ATAC-only) contributed more to the differentiation between cartilage 2, cartilage 3 and mesenchyme
327 domains (blue circles in Fig. 4i). Therefore, purely utilizing a single omics layer resulted in biased
328 interpretation of biological signals in spatial domains. In comparison, when taking both omics layers as
329 input (RNA & ATAC), PRESENT successfully captured variations among those domains in both E11.0
330 and E15.5 mouse brain samples (Fig. 4h, i), highlighting the capabilities of PRESENT in leveraging
331 spatial and complementary transcriptomic & epigenomic information for cross-modality representation
332 and further facilitating the identification of spatial domains from a systematic perspective.
333

334 **PRESENT-BC enables effective integration of horizontal and vertical spatially resolved samples**

335 Compared with cross-modality representation on a single spatial sample, integrated analysis of multiple
336 samples enables the identification of hierarchical structures of cellular organizations across different
337 dissected areas (horizontally adjacent tissue samples) or developmental stages (vertically serial tissue
338 samples)[23]. Therefore, building upon the cross-modality representation framework PRESENT, we further
339 extended PRESENT-BC for the multi-sample integration of horizontal and vertical samples generated by
340 spatially resolved technologies (Methods, Fig. 1b).

341 We firstly collected two horizontal SRT mouse brain samples generated by the 10x Genomics Visium
342 platform, which involves a sagittal anterior section and a sagittal posterior section. We applied two SRT
343 integration methods, namely STAligner[25] and GraphST[23], and PRESENT-BC to carry out integrated
344 analysis of these two adjacent brain samples and utilized the anatomic annotation of the sagittal region in
345 postnatal day 56 (P56) mouse brain provided by Allen Reference Atlas[71] to intuitively compare the
346 performance of these methods (Fig. 5a). As shown in Fig. 5b-d, analytical results illustrated significant
347 improvements of PRESENT-BC in shared and sample-specific domain identification compared with

348 STAligner and GraphST. For example, PRESENT-BC captured the classical hierarchical structures, i.e.,
349 from outer layer 1 to inner layer 6, of dorsal pallium in both sagittal anterior and posterior sections, as
350 illustrated in the red rectangle of Fig. 5d. In addition, PRESENT-BC successfully distinguished the
351 typical cortex Ammonis (CA, cluster 18) and dentate gyrus (DG, cluster 25) structures simultaneously in
352 the hippocampus (black frames in Fig. 5d), which have been proved to be associated with mood, anxiety
353 and personality disorders caused by various factors[72]. We further conducted differentially expressed
354 analysis for domain CA and domain DG, acquiring marker genes including *C1ql2* and *Hpca*, respectively
355 (Supplementary Figs. 20 and 21). According to previous studies, *C1ql2* serves as a marker gene for
356 dentate gyrus granule cells (DGCs) enriched in the DG structure[73], while *Hpca* is the encoding gene for a
357 high-affinity calcium-binding protein restricted to the central nervous system, abundant in pyramidal cells
358 of the hippocampus CA region[74].

359 Furthermore, PRESENT-BC also exhibited great potential in capturing sample-specific functional
360 domains. For instance, PRESENT-BC successfully revealed the sophisticated substructures of the
361 cerebellar vermis, which only exists in the sagittal posterior section of mouse brain and plays an
362 important role in coordinating movements of the central body such as trunk, head, and proximal limbs.
363 Specifically, PRESENT-BC successfully detected the molecular layer (cluster 3), Purkinje cell layer
364 (cluster 22) and internal granular layer (cluster 17) of cerebellar vermis cortex and the white matter
365 (cluster 20) of cerebellar vermis. The differentially expressed genes in the corresponding domains
366 (Supplementary Figs. 22-25), including *S100b*[75], *Car8*[76], *Atp1a1*[77] and *Mbp*[78], have also been reported in
367 previous studies, further illustrating the effectiveness of PRESENT-BC in uncovering biological
368 implications in identified spatial domains. For example, *Car8* and *Atp1a1*, the mutation of which cause
369 the motor coordination defect and dominant Charcot-Marie-Tooth type 2, were proved to be highly
370 expressed in the granular layers and Purkinje cells in recent studies[76, 77], respectively.

371 On the other hand, we also quantitatively evaluated the performance of PRESENT-BC in the
372 integrative analysis of vertically serial tissue samples involving mouse embryo slices from three
373 developmental stages (E15.5, E13.5 and E12.5) generated by spatial ATAC technique[10]. Due to the lack of
374 methods specially designed for spatial ATAC data, we compared PRESENT-BC with six integration
375 methods developed for single-cell data (MNN[79], Scanorama[39], Harmony[38], SCALEX[80], scBasset-BC[81]
376 and SnapATAC2[51]) and two recently proposed SRT methods (STAligner[25] and GraphST[23]). According to
377 a recent benchmarking framework for single-cell omics data integration[82], we simultaneously accessed
378 the batch effect removal and biological variance conservation capabilities of these methods, with seven
379 metrics for each aspect (Methods). As shown in Fig. 5e, PRESENT-BC achieved the best overall score for
380 multi-sample integration, ranking first in preserving biological variance and second in removing batch
381 effects among all the integration methods. The UMAP visualization of spot embeddings across different
382 developmental stages also demonstrated that compared with other methods, PRESENT-BC gained better
383 performance in aggregating spots from the same domain but in different batches (Supplementary Fig. 26).
384 Furthermore, we performed joint spatial clustering based on the joint embeddings obtained by different
385 methods. As shown in Fig. 5f, g, Scanorama and STAligner could hardly distinguish different spatial
386 domains across different developmental stages. MNN, Harmony, SCALEX and SnapATAC2 provided
387 improved results by roughly recognizing forebrain and liver in E12.5 and E13.5 stages, but they still
388 failed in other domains. In comparison, PRESENT-BC not only identified forebrain (cluster 1) and liver
389 (cluster 5) in all the three stages, but also gained superior results in domains including hindbrain (cluster
390 2) and midbrain (cluster 8). In summary, these experimental results have highlighted the superior
391 performance of PRESENT-BC in integrative analysis of both horizontally adjacent and vertically serial
392 spatially resolved samples generated by various spatial omics technologies.
393

394 **PRESENT-BC captures biological signals while eliminating batch effects in spatial RNA-ATAC data**

395 The MISAR-seq mouse brain dataset simultaneously measured the transcriptomic and epigenomic
396 profiles of developmental brain samples from four stages, including E11.0, E13.5, E15.5 and E18.5,
397 providing a multi-omics landscape of the developmental mouse brain in the context of cellular
398 organizations[14]. During the development of mouse brain, some functional structures may be specific to
399 certain developmental stages (e.g., primary brain), while others may persist throughout the entire
400 developmental process (e.g., midbrain and muscle). Therefore, the key challenges of integrating the
401 MISAR-seq mouse brain samples included not only the incorporation of multi-omics information across
402 multiple samples, but also the conservation of shared and stage-specific biological variations.

403 Due to the lack of multi-sample integration methods specially designed for spatial multi-omics data,
404 we compared the performance of PRESENT-BC with six methods developed for single-cell data (MNN[79],
405 Scanorama[39], Harmony[38], SCALEX[80], scBasset-BC[81] and SnapATAC2[51]) and three recently proposed
406 SRT methods (STAligner[25], Stitch3D[31] and GraphST[23]). MNN, Scanorama, Harmony and SCALEX could
407 be applied to batch integration of both RNA-seq and ATAC-seq data, denoted as RNA-* and ATAC-*,
408 respectively, where * represents the different methods. scBasset-BC was featured as a sequence-based
409 method specially designed for single-cell ATAC-seq data, while STAligner, Stitch3D and GraphST were
410 intended for integrative analysis of SRT data. SnapATAC2 has provided guidance for incorporating both
411 transcriptomic and epigenomic information for latent representation and integrated analysis. These
412 methods were systematically benchmarked using 14 metrics divided into two aspects, i.e., batch effect
413 removal and biological variance conservation. As shown in Fig. 6a, scBasset-BC and SnapATAC2
414 provided satisfactory performance through incorporating sequence and multi-omics information,
415 respectively. PRESENT-BC, which simultaneously integrated spatial dependency and multi-omics
416 information of all the samples, achieved the best performance in both batch effect removal and biological

417 variance conservation among all the 14 methods, thereby illustrating the superiority in multi-sample
418 integration for spatial RNA-ATAC data.

419 The UMAP visualization of joint embeddings colored by batch indices (Fig. 6b and Supplementary
420 Fig. 27a) and ground truth domain labels (Fig. 6c and Supplementary Fig. 27b) also validated the
421 quantitative results. Specifically, PRESENT-BC successfully fused spots from different developmental
422 stages (red boxes in Fig. 6b) while maintaining the biologically distinct patterns of DPallm, DPallv,
423 subpallium 1 & 2, thalamus and muscle (red boxes in Fig. 6c). Specifically, spots of muscle domain
424 persisted throughout the entire life of mouse brain development, while those of subpallium 1&2, DPallm,
425 DPallv and thalamus only existed in specific stages. The success of PRESENT-BC in distinguishing these
426 domains illustrated its capabilities in capturing shared and stage-specific biological variance while
427 eliminating batch effects. Furthermore, we conducted joint spatial clustering across the four stages based
428 on the integrated representations obtained by different methods, as shown in Fig. 6d and Supplementary
429 Fig. 28. Except for the scBasset-BC and SnapATAC2, other baseline methods provide unreliable
430 clustering results. In contrast, PRESENT-BC successfully detected not only large domains like DPallm
431 (cluster 7), but also small domains including thalamus (cluster 16) and muscle (cluster 11), which
432 PRESENT failed to identify through individual analysis on a single sample from E18.5 stage
433 (Supplementary Fig. 19), highlighting the advantages of integrative analysis of multiple samples. In
434 summary, these results emphasize that PRESENT-BC enables the conservation of biological signals while
435 removing batch effects in the integrated analysis of spatial RNA-ATAC samples.

436

437 **Discussion**

438 Recent breakthroughs in spatial omics sequencing technologies have enabled the simultaneous acquisition
439 of spatial coordinate information and the regulatory status of the biomolecules within cells, such as

440 chromatin regions, genes, and proteins, accumulating massive spatial omics data in repositories and
441 databases[15, 16]. In this article, we proposed PRESENT, an effective and scalable contrastive learning
442 framework, to address the cross-modality representation of spatially resolved omics data, including
443 spatial epigenomic, transcriptomic and multi-omics data. To model the distinct patterns and distributions
444 of different omics layers, we designed omics-specific encoders and decoders for different omics data,
445 where the transcriptomic and epigenomic decoders model the counts via a zero-inflation negative
446 binomial (ZINB) distribution[43] and a zero-inflation Poisson (ZIP) distribution[44], respectively (Methods).
447 In addition, a cross-omics alignment module built on a contrastive learning loss function[42] and graph
448 attention networks[40] (GAT) and were also introduced to capture the spatial dependency and
449 complementary multi-omics information (Methods). We further extended the cross-modality
450 representation framework to the multi-sample integration of spatial omics data, addressing the key
451 challenges including the inconsistency of spatial coordinates between samples as well as the conservation
452 of biological variance while eliminating batch effects. Specifically, we adopted an inter-batch alignment
453 loss[45] and a batch-adversarial learning strategy[46] to eliminate redundant batch effects as well as an intra-
454 batch preserving loss to maintain biological signals simultaneously. As part of the validation of the model
455 design, we conducted a comprehensive ablation study to illustrate the effectiveness of these modules, as
456 shown in Supplementary Figs. 29 and 30.

457 Throughout comprehensive experiments on different species involving human, mouse and axolotl,
458 we highlight the superior performance of PRESENT in cross-modality representation of spatial omics
459 data generated by various spatial sequencing technologies, encompassing spatial ATAC-seq[10], spatial
460 transcriptomics[7], as well as the spatial RNA-Protein[11, 12] and RNA-ATAC[14] co-profiling techniques.
461 Specifically, PRESENT effectively integrates the spatial dependency and diverse omics information
462 contained in spatial data generated by spatial ATAC-seq[10] or spatial transcriptomics[7] techniques (Fig. 2a-

463 f). It also shows great potential for leveraging various reference data to address issues related to low
464 sequencing depth and signal-to-noise ratio in spatial omics data (Fig. 2g-i). Additionally, PRESENT
465 exhibits powerful capabilities in incorporating spatial and complementary multi-omics information,
466 thereby facilitating explorative analysis in spatial multi-omics data (Fig. 3, 4). More importantly, the
467 extended multi-sample integration version, namely PRESENT-BC, facilitates the integrative analysis of
468 multiple spatially resolved samples across distinct dissection areas (Fig. 5a-d) and developmental stages
469 (Fig. 5e-g, 6), thereby promoting the detection of hierarchical structures and biological functions within
470 cellular organizations from a spatiotemporal perspective.

471 Nevertheless, several avenues remain for further improvement of PRESENT. First, although
472 PRESENT can anchor on spots to analyze a single sample with paired omics sequencing data and
473 PRESENT-BC is capable of anchoring on shared genomic features to integrate multiple samples, the
474 framework still struggles with the integration of unpaired samples with distinct features[83], such as the
475 integration of two samples generated by spatial ATAC-seq and SRT techniques, respectively. Due to
476 technological and cost limitations, the number of unpaired spatial samples significantly outweighs the
477 number of paired spatial samples[83], highlighting the importance of unpaired integration. Second, it is also
478 inspiring to extend the application scenarios of PRESENT to the integration of single-cell and spatial
479 omics data, which lack crucial spatial information and single-cell resolution, respectively. Third, in
480 addition to the integrative analysis of multiple samples in the latent space, the alignment of spatial
481 coordinates in physical space also plays an important role in joint analysis of multiple samples. Moreover,
482 advanced spatial sequencing technologies like 10x Genomics Visium additionally provide high-resolution
483 histology image of dissected samples, providing additional morphological features which could be further
484 leveraged by PRESENT for enhanced analysis of spatially resolved samples.
485

Methods

Cross-modality representation with PRESENT

PRESENT is constructed upon a multi-view autoencoder framework for representation learning of spots sequenced by spatial omics technologies, adeptly incorporating cross-modality information, encompassing spatial coordinate data, epigenomic data generated by the assay for transposase-accessible chromatin with sequencing (ATAC), transcriptomic data measured by RNA sequencing (RNA), and proteomic data sequenced by antibody-derived tags (ADT), into a low-dimensional feature matrix. Note that while the spatial coordinate matrix is a necessary input, the three omics layers are not all mandatory. In other words, PRESENT can take an arbitrary combination of count matrices from the three omics layers, including ATAC, RNA and ADT data, as input. In the following sections, taking all the three omics layers and spatial coordinates as input, we elaborate the architecture of the PRESENT, which consists of omics-specific encoders, a cross-omics alignment module, and omics-specific decoders (Fig. 1a).

Multi-modality input of PRESENT. The input of PRESENT is a spatial coordinate matrix $\boldsymbol{M} \in R^{n\times 2}$ and three optional fragment count matrices, namely the ATAC data $\boldsymbol{X^{ATAC}} \in R^{n\times r}$, RNA data $\boldsymbol{X^{RNA}} \in R^{n\times g}$ and ADT data $\boldsymbol{X^{ADT}} \in R^{n\times p}$. Here n denotes the number of spots, r the number of input regions, g the number of highly variable genes and p the number of input proteins. Before feeding the spatial and omics input into the omics-specific encoders, we transform the spatial coordinate matrix $\boldsymbol{M}$ into an undirected neighborhood graph $\boldsymbol{G}$ through adding edge connections between all spots and their k nearest neighbors, where k is set to 6 by default. Afterwards, we apply term frequency-inverse document frequency (TF-IDF) normalization to the ATAC data[84-87], library size normalization followed by log-transformation to RNA data[17, 29, 30] and the centered log ratio (CLR) transformation to ADT data [36, 88, 89], obtaining normalized omics data $\widetilde{\boldsymbol{X}}^{\boldsymbol{ATAC}} \in R^{n\times r}$, $\widetilde{\boldsymbol{X}}^{\boldsymbol{RNA}} \in R^{n\times g}$ and $\widetilde{\boldsymbol{X}}^{\boldsymbol{ADT}} \in R^{n\times p}$, respectively.

Omics-specific encoder. Graph neural networks, such as graph attention networks[40] (GAT), have showcased successful application in integrating spatial and transcriptomic information for spatial transcriptomics[17, 23, 29, 30]. On the other hand, research has proved that introducing weight uncertainty into models through Bayesian neural networks[41] (BNN) improves generalization in nonlinear regression problems[41] and offers an extension for leveraging prior information from existing sequencing data[17]. Inspired by these studies, we design an omics-specific encoder, consisting of a GAT module and a BNN module, to capture spatially organized and biologically regulatory signals embedded in each omics layer. In the following two subsections, taking the ATAC omics-specific encoder as an example, we delineate the detailed design of GAT and BNN module.

Graph attention network. Given the input undirected neighborhood graph $\boldsymbol{G}$ and the normalized ATAC matrix $\widetilde{\boldsymbol{X}}^{\boldsymbol{ATAC}} \in R^{n\times r}$, we denote the normalized accessibility vector of spot i in $\widetilde{\boldsymbol{X}}^{\boldsymbol{ATAC}}$ as $\widetilde{\boldsymbol{x}}_{\boldsymbol{i}}^{\boldsymbol{ATAC}} \in R^{r}$ and the spatial neighborhood of spot i as $\mathcal{N}_{\mathrm{i}}$. Initially, we acquire the embedding from the first hidden layer of spot i, i.e., $\boldsymbol{h}_{\boldsymbol{i}}^{\boldsymbol{1}} \in R^{d_1}$ using a fully connected network (FCN):

$$\boldsymbol{h}_{\boldsymbol{i}}^{\boldsymbol{1}} = \sigma(\boldsymbol{W}_{\boldsymbol{1}}\widetilde{\boldsymbol{x}}_{\boldsymbol{i}}^{\boldsymbol{ATAC}} + \boldsymbol{b}_{\boldsymbol{1}})$$

where d_1 represents dimension of the first hidden layer, σ indicates the combination of layer normalization and exponential linear unit (ELU) activation function, and $\boldsymbol{W}_{\boldsymbol{1}} \in R^{d_1\times r}$ and $\boldsymbol{b}_{\boldsymbol{1}} \in R^{d_1}$ indicate the weight and bias parameters of the FCN, respectively. Afterwards, The second hidden-layer embedding of spot i, namely $\boldsymbol{h}_{\boldsymbol{i}}^{\boldsymbol{2}} \in R^{d_2}$, is obtained based on a graph attention layer[40], depicted as:

$$\boldsymbol{h}_{\boldsymbol{i}}^{\boldsymbol{2}} = \sigma(\sum_{j\in\mathcal{N}_i} \alpha_{ij}\boldsymbol{W}_{\boldsymbol{2}}\boldsymbol{h}_{\boldsymbol{j}}^{\boldsymbol{1}})$$

$$\alpha_{ij} = \frac{\sigma_{LeakyReLU}\left(\boldsymbol{a}^T\left(\boldsymbol{W}_{\boldsymbol{2}}\boldsymbol{h}_{\boldsymbol{i}}^{\boldsymbol{1}} \oplus \boldsymbol{W}_{\boldsymbol{2}}\boldsymbol{h}_{\boldsymbol{j}}^{\boldsymbol{1}}\right)\right)}{\sum_{k\in\mathcal{N}_i}\sigma_{LeakyReLU}\left(\boldsymbol{a}^T\left(\boldsymbol{W}_{\boldsymbol{2}}\boldsymbol{h}_{\boldsymbol{i}}^{\boldsymbol{1}} \oplus \boldsymbol{W}_{\boldsymbol{2}}\boldsymbol{h}_{\boldsymbol{k}}^{\boldsymbol{1}}\right)\right)}$$

where α_{ij} is the attention score between spot i and j, d_2 denotes the dimension of the second hidden layer, $\boldsymbol{W}_{\boldsymbol{2}} \in R^{d_1\times d_2}$ and $\boldsymbol{a} \in R^{2d_2}$ are learnable parameters of the graph attention layer, $\sigma_{LeakyReLU}$

denotes the LeakyReLU activation function with a negative slope of 0.2, and $\oplus$ refers to the concatenation function. Finally, $\boldsymbol{h_i^2}$ undergoes transformation by another graph attention layer to obtain $\boldsymbol{h_i^{gat}} \in R^{d_{gat}}$ as the output, where d_{gat} denotes the output dimension of GAT module.

Bayesian neural network. According to our previous work, BNN module provides an opportunity to leverage biological prior information from various sequencing data to improve the analytical performance in spatial transcriptomic data[17]. Inspired by this, PRESENT includes the BNN module, which assumes that the parameters of weight matrix, denoted as $\boldsymbol{W} \in R^{d_{bnn} \times r}$, follow Gaussian distributions. Specifically, each element of $\boldsymbol{W}$ is governed by a Gaussian distribution parameterized by the corresponding elements in the mean matrix $\boldsymbol{\mu^b}$ and variance matrix $\boldsymbol{\sigma^{b^2}}$, depicted as:

$$\boldsymbol{W_{ij}} \sim \mathcal{N}\left(\boldsymbol{\mu_{ij}^b}, \boldsymbol{\sigma_{ij}^{b^2}}\right)$$

In addition, $\boldsymbol{W_{ij}}$ is also governed by a Gaussian prior parameterized by mean $\boldsymbol{\mu_{ij}^p}$ and variance $\boldsymbol{\sigma_{ij}^{p^2}}$. Inheriting our previous work[17], we calculate the PCA projection weight matrix of normalized ATAC matrix $\boldsymbol{\widetilde{X}^{ATAC}}$ as the mean matrix $\boldsymbol{\mu^p}$ and utilize the standard deviation of all the elements in mean matrix $\boldsymbol{\mu^p}$ to construct the standard deviation matrix $\boldsymbol{\sigma^p}$. The distance between the prior and the distributions of the parameters in BNN module is restricted with Kullback-Leibler divergence (KLD):

$$\mathcal{L}_{BNN} = KLD(\mathcal{N}\left(\boldsymbol{\mu^b}, \boldsymbol{\sigma^{b^2}}\right), \mathcal{N}\left(\boldsymbol{\mu^p}, \boldsymbol{\sigma^{p^2}}\right))$$

It is notable that the parameters of prior Gaussian distributions, i.e., the mean matrix $\boldsymbol{\mu^p}$ and the variance matrix $\boldsymbol{\sigma_{ij}^{p^2}}$, can also be constructed with reference ATAC data, enabling PRESENT to incorporate prior information from existing ATAC-seq data generated by various technologies, including single-cell ATAC-seq or spatial ATAC-seq. The unification of peak sets between reference ATAC-seq data and target ATAC-seq data is conducted according to the merging objects vignette provided by the Signac package[90].

Therefore, the output embeddings of spot i from BNN module, i.e., $\boldsymbol{h_i^{bnn}} \in R^{d_{bnn}}$, is denoted as:

$$\boldsymbol{h_i^{bnn}} = \boldsymbol{W}\boldsymbol{\widetilde{x}_i^{ATAC}} = (\boldsymbol{\mu^b} + \boldsymbol{\eta} \odot \boldsymbol{\sigma^b})\boldsymbol{\widetilde{x}_i^{ATAC}}$$

where d_{bnn} denotes the output dimension of BNN module, $\boldsymbol{\eta}$ is the standard Gaussian noise matrix and $\odot$ denotes element-wise product.

Finally, the output embedding of spot i from the ATAC omics-specific encoder, denoted as $\boldsymbol{h_i^{ATAC}} \in R^{d_{ATAC}}$, is the concatenation of output embeddings from the GAT and BNN module:

$$\boldsymbol{h_i^{ATAC}} = \boldsymbol{h_i^{gat}} \oplus \boldsymbol{h_i^{bnn}}$$

where $d_{ATAC} = d_{gat} + d_{bnn}$ is the output dimension of omics-specific encoder for ATAC data.

Cross-omics alignment module. Based on the omics-specific encoders, the embeddings of different omics layers are separately extracted. However, accurately deciphering sophisticated gene regulation process requires integrating the status of biological macromolecules in cells from different omics perspectives[34-37]. Inspired by a recent study that achieved multi-modal entity alignment through contrastive learning[42], we designed a cross-omics alignment module to model the complex regulatory interactions between different omics layers.

Given the omics-specific embeddings of spot i obtained by omics-specific encoders, including $\boldsymbol{h_i^{ATAC}} \in R^{d_{ATAC}}$ for ATAC, $\boldsymbol{h_i^{RNA}} \in R^{d_{RNA}}$ for RNA and $\boldsymbol{h_i^{ADT}} \in R^{d_{ADT}}$ for ADT, we utilized a multi-layer perception (MLP) module to fuse the embeddings from different omics layers to acquire a cross-omics representation of spot i, denoted as:

$$\boldsymbol{e_i} = f_{MLP}(\boldsymbol{h_i^{ATAC}} \oplus \boldsymbol{h_i^{RNA}} \oplus \boldsymbol{h_i^{ADT}})$$

where $\boldsymbol{e_i} \in R^d$ represents the final cross-omics representation of spot i, f_{MLP} denotes the transformation function of the MLP module, d is the latent dimension of PRESENT, and d_{ATAC}, d_{RNA} and d_{ADT} refers to the output dimensions of omics-specific encoders for ATAC, RNA and ADT data, respectively. The MLP module is composed of three FCNs, the first two of which are followed by a layer normalization and an ELU activation function. However, only the MLP fusion module cannot perfectly incorporate complementary omics-specific knowledge from different omics layers into the final cross-omics

embeddings [42], which is also illustrated in our ablation study shown in Supplementary Fig. 29. Therefore, we design an inter-omics alignment (IOA) loss to minimize the multi-omics directional KLD over the output distributions between the joint embeddings and other omics-specific embeddings:

$$\mathcal{L}_{IOA} = \frac{1}{M}\left[\frac{1}{n}\sum_{i=1}^{n} KLD\left(q_{ed}^{(i)}(\boldsymbol{e}), q_{ed}^{(i)}(\boldsymbol{h}^{ATAC})\right) + \frac{1}{n}\sum_{i=1}^{n} KLD\left(q_{ed}^{(i)}(\boldsymbol{e}), q_{ed}^{(i)}(\boldsymbol{h}^{RNA})\right) + \frac{1}{n}\sum_{i=1}^{n} KLD\left(q_{ed}^{(i)}(\boldsymbol{e}), q_{ed}^{(i)}(\boldsymbol{h}^{ADT})\right)\right]$$

where M represents the number of omics layers, $\boldsymbol{e}, \boldsymbol{h}^{ATAC}, \boldsymbol{h}^{RNA}$ and $\boldsymbol{h}^{ADT}$ denotes the joint, ATAC, RNA and ADT embeddings respectively, and $q_{ed}^{(i)}(*)$ represents the embedding distribution of spot i over other spots, defined as (taking $\boldsymbol{e}$ as an example):

$$q_{ed}^{(i)}(\boldsymbol{e}) = \left(\frac{\delta_{ed}(\boldsymbol{e_i}, \boldsymbol{e_1})}{\sum_{j=1}^{n}\delta_{ed}(\boldsymbol{e_i}, \boldsymbol{e_j})}, \frac{\delta_{ed}(\boldsymbol{e_i}, \boldsymbol{e_2})}{\sum_{j=1}^{n}\delta_{ed}(\boldsymbol{e_i}, \boldsymbol{e_j})}, \dots, \frac{\delta_{ed}(\boldsymbol{e_i}, \boldsymbol{e_n})}{\sum_{j=1}^{n}\delta_{ed}(\boldsymbol{e_i}, \boldsymbol{e_j})}\right)$$

$$\delta_{ed}(\boldsymbol{u}, \boldsymbol{v}) = exp\left(\frac{\boldsymbol{u}^T\boldsymbol{v}}{\tau}\right)$$

where τ is a temperature parameter and set to 0.1 according to the original study[42].

Omics-specific decoder. The patterns and distributions of sequencing data vary significantly across different omics layers, revealing complementary biologically regulatory information[34-37]. For example, ADT sequencing data, which focuses on dozens of surface proteins, provides intuitive and reliable quantification of cellular activities with low dimensionality, few dropout events and low sparsity[36]. In contrast, RNA and ATAC data, which measure the expression levels of genes and accessibility scores of genomic regions respectively, suffer from high dropout rates, dimensionality, and sparsity[91]. Therefore, we designed omics-specific decoder for each modality to address this challenge.

Inspired by a recently published study which modeled the fragment counts of single-cell ATAC-seq data with Poisson distribution[44], we fit the fragment counts of spatial ATAC-seq data $\boldsymbol{X}^{ATAC} \in R^{n\times r}$ with the zero-inflation Poisson (ZIP) distribution to accommodate the lower capture rate of spatial technologies[10, 14]. Taking the fragment count vector of spot i, i.e., $\boldsymbol{x_i}^{ATAC}$, as an example, the omics-specific decoder of ATAC data as well as the probability model is described as:

$$\boldsymbol{\pi_i} = \sigma_{sigmoid}\left(f_{MLP}^{ATAC,\pi}(\boldsymbol{e_i})\right) \quad \boldsymbol{\rho_i} = f_{MLP}^{ATAC,\rho}(\boldsymbol{e_i}) \quad \hat{\boldsymbol{x}}_{\boldsymbol{i}}^{c} = f_{MLP}^{ATAC,x}(\boldsymbol{e_i})$$

$$\boldsymbol{\omega_i} = softmax(\boldsymbol{\rho_i} + \boldsymbol{r^c})$$

$$\boldsymbol{\lambda_i} = l_i \cdot \boldsymbol{\omega_i}$$

where $\sigma_{sigmoid}$ represents the sigmoid activation function, $\boldsymbol{r^c}$ is designed to capture region specific bias such as peak length, $l_i = \sum_{j=1}^{r} x_{ij}^{c}$ refers to the total fragment counts of spot i, $f_{MLP}^{ATAC,\pi}(*)$, $f_{MLP}^{ATAC,\rho}(*)$ and $f_{MLP}^{ATAC,x}(*)$ are three output layers of $f_{MLP}^{ATAC}(*)$, which denotes the MLP module of ATAC omics-specific decoder. Specifically, $f_{MLP}^{ATAC,\pi}(*)$ and $f_{MLP}^{ATAC,\rho}(*)$ output parameters of the ZIP probability model, while $f_{MLP}^{ATAC,x}(*)$ outputs reconstructed normalized profiles of ATAC data. Then for each region $j = 1,2,\dots,r$, the fragment count of i-th spot x_{ij}^{ATAC} satisfies:

$$x_{ij}^{ATAC} \sim ZIP\left(x_{ij}^{ATAC}; \pi_{ij}, \lambda_{ij}\right)$$

$$ZIP\left(x_{ij}^{ATAC}; \pi_{ij}, \lambda_{ij}\right) = \pi_{ij}\delta_0\left(x_{ij}^{ATAC}\right) + (1-\pi_{ij})Poisson(x_{ij}^{ATAC}; \lambda_{ij})$$

$$\delta_0\left(x_{ij}^{ATAC}\right) = \begin{cases} 1 \; if \; x_{ij}^{ATAC} \; equals \; 0 \\ 0 \; otherwise \end{cases} \quad Poisson\left(x_{ij}^{ATAC}; \lambda_{ij}\right) = \frac{\lambda_{ij}^{x_{ij}^{ATAC}} e^{-\lambda_{ij}}}{x_{ij}^{ATAC}!}$$

Therefore, the negative log likelihood (NLL) loss function of the ZIP probability model is:

$$\mathcal{L}_{NLL}^{ATAC} = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r}\left[-log\, ZIP\left(x_{ij}^{ATAC}; \pi_{ij}, \lambda_{ij}\right) + \gamma\pi_{ij}^{2}\right]$$

where the $\gamma\pi_{ij}^{2}$ refers to a regularization term and γ is set to 0.5 by default. The reconstruction loss of ATAC omics layer is measured with mean squared error (MSE):

$$\mathcal{L}_{MSE}^{\mathrm{ATAC}} = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r}\left(\hat{x}_{ij}^{ATAC} - \tilde{x}_{ij}^{ATAC}\right)^2$$

Similarly, we also model the counts of RNA data $\boldsymbol{X^{RNA}} \in R^{n\times g}$ using the zero-inflation negative binomial (ZINB) distribution according to a single-cell RNA-seq analytical research[43], denoted as:

$$\boldsymbol{\pi_i} = \sigma_{sigmoid}\left(f_{MLP}^{RNA,\pi}(\boldsymbol{e_i})\right) \quad \boldsymbol{\mu_i} = \sigma_{softplus}\left(f_{MLP}^{RNA,\mu}(\boldsymbol{e_i})\right)$$

$$\boldsymbol{\theta_i} = \sigma_{softplus}\left(f_{MLP}^{RNA,\theta}(\boldsymbol{e_i})\right) \quad \hat{\boldsymbol{x}}_{\boldsymbol{i}}^{r} = f_{MLP}^{RNA,x}(\boldsymbol{e_i})$$

where $\sigma_{softplus}$ refers to the softplus activation function, $f_{MLP}^{RNA,\pi}(*)$, $f_{MLP}^{RNA,\mu}(*)$, $f_{MLP}^{RNA,\theta}(*)$ and

$f_{MLP}^{RNA,x}(*)$ are four output layers of f_{MLP}^{RNA}, the MLP module of the RNA omics-specific decoder. Then the count of spot i on gene j satisfies the following model:

$$x_{ij}^{RNA} \sim ZINB\left(x_{ij}^{RNA}; \pi_{ij}, \mu_{ij}, \theta_{ij}\right)$$

$$ZINB\left(x_{ij}^{RNA}; \pi_{ij}, \mu_{ij}, \theta_{ij}\right) = \pi_{ij}\delta_0\left(x_{ij}^{RNA}\right) + (1-\pi_{ij})NB(x_{ij}^{RNA}; \mu_{ij}, \theta_{ij})$$

$$NB\left(x_{ij}^{RNA}; \mu_{ij}, \theta_{ij}\right) = \frac{\Gamma\left(x_{ij}^{RNA} + \theta_{ij}\right)}{\Gamma\left(\theta_{ij}\right)}\left(\frac{\theta_{ij}}{\theta_{ij} + \mu_{ij}}\right)^{\theta_{ij}}\left(\frac{\mu_{ij}}{\theta_{ij} + \mu_{ij}}\right)^{x_{ij}^{RNA}}$$

The NLL loss function of the ZINB probability model for RNA data can be denoted as:

$$\mathcal{L}_{NLL}^{RNA} = \frac{1}{ng}\sum_{i=1}^{n}\sum_{j=1}^{g}\left[-\log ZINB\left(x_{ij}^{RNA}; \pi_{ij}, \mu_{ij}, \theta_{ij}\right) + \gamma\pi_{ij}^2\right]$$

where the regularization parameter γ is also set to 0.5 by default. In addition, the reconstruction loss of RNA omics layer is depicted as:

$$\mathcal{L}_{MSE}^{\mathrm{RNA}} = \frac{1}{ng}\sum_{i=1}^{n}\sum_{j=1}^{g}\left(\hat{x}_{ij}^{RNA} - \tilde{x}_{ij}^{RNA}\right)^2$$

Compared with ATAC and RNA data, lower dropout rates, dimensionality and sparsity make the analysis of ADT data less challenging. Therefore, the ADT omics-specific decoder is built on the MLP module with one output layer to reconstruct the normalized expression level of proteins, denoted as:

$$\hat{\boldsymbol{x}}_{\boldsymbol{i}}^{ADT} = f_{MLP}^{ADT}(\boldsymbol{e}_{\boldsymbol{i}})$$

$$\mathcal{L}_{MSE}^{\mathrm{ADT}} = \frac{1}{np}\sum_{i=1}^{n}\sum_{j=1}^{p}\left(\hat{x}_{ij}^{ADT} - \tilde{x}_{ij}^{ADT}\right)^2$$

Therefore, the overall NLL loss and reconstruction loss of PRESENT are summarized as:

$$\mathcal{L}_{NLL} = \mathcal{L}_{NLL}^{\mathrm{ATAC}} + \mathcal{L}_{NLL}^{\mathrm{RNA}}$$

$$\mathcal{L}_{MSE} = \mathcal{L}_{MSE}^{\mathrm{ATAC}} + \mathcal{L}_{MSE}^{\mathrm{RNA}} + \mathcal{L}_{MSE}^{\mathrm{ADT}}$$

Overall loss functions of PRESENT. Based on the above descriptions, the overall loss function of PRESENT consists of four parts, including the loss of BNN module, the inter-omics alignment loss, the negative log likelihood loss of ZIP and ZINB probability model, and the mean squared error loss for reconstruction of normalized omics profiles, which could be denoted as:

645 $$\mathcal{L}_{repre} = \alpha_1 \mathcal{L}_{BNN} + \alpha_2 \mathcal{L}_{IOA} + \alpha_3 \mathcal{L}_{NLL} + \alpha_4 \mathcal{L}_{MSE}$$

646 where α_1, α_2, α_3 and α_4 are coefficients of the corresponding loss term, and the default values of them

647 are all set to 1.

648

649 **Multi-sample integration with PRESENT-BC**

650 We further extend the PRESENT framework to PRESENT-BC, which not only retains the ability of

651 PRESENT in incorporating multi-modal information, including various omics layers along with the

652 necessary spatial coordinate data, but also effectively eliminates batch effects while capturing shared

653 biological variations across samples. The architecture of PRESENT-BC also includes omics-specific

654 encoders, a cross-omics alignment module, and omics-specific decoders. However, compared with

655 PRESENT, the input to the omics-specific decoders in PRESENT-BC is the concatenation of the output

656 from the cross-omics alignment module and the one-hot encoding of batch labels. Additionally,

657 PRESENT-BC is enhanced for multi-slice integration based on a two-stage workflow. The first stage

658 serves as a warmup to pretrain the model, while the second stage leverage a cyclic iterative optimization

659 strategy to eliminate batch effects while preserving biological variations across samples. Using a scenario

660 of fusing two spatial multi-omics sequencing samples as an example, where ATAC, RNA, ADT, and

661 spatial coordinate data are simultaneously sequenced for each spot, we illustrate the workflow of multi-

662 sample integration with PRESENT-BC in the subsequent sections (Fig. 1b).

663 **Feature sets unification.** Before multi-sample integration, the first step is to align the feature sets of

664 different sequencing samples, which serves as the anchor for multi-sample integration. For ATAC data,

665 following the merging objects vignette provided by the Signac package[90], a unified set of peak regions is

666 first obtained by merging all intersecting peaks from all the samples. The fragment counts of each spot

667 within the unified peak set are then quantified for each sample. For RNA and ADT data, whose feature

sets are predefined genes or proteins, we directly leverage the intersection of gene or protein sets of all the samples as the unified gene or protein set. After the gene intersection step, highly variable genes will be selected based on the unified gene set. For convenience, we denoted the ATAC, RNA, ADT and spatial data of the two samples as $\boldsymbol{X}_1^{ATAC} \in R^{n_1 \times r}$, $\boldsymbol{X}_2^{ATAC} \in R^{n_2 \times r}$, $\boldsymbol{X}_1^{RNA} \in R^{n_1 \times g}$, $\boldsymbol{X}_2^{RNA} \in R^{n_2 \times g}$, $\boldsymbol{X}_1^{ADT} \in R^{n_1 \times p}$, $\boldsymbol{X}_2^{ADT} \in R^{n_2 \times p}$, $\boldsymbol{M_1} \in R^{n_1 \times 2}$ and $\boldsymbol{M_2} \in R^{n_2 \times 2}$, respectively. Here, r represents the number of unified regions, g the number of highly variable genes, p the number of input proteins, and n_1 and n_2 the number of spots of the two samples. The unified ATAC, RNA, ADT data are then concatenated and preprocessed using the omics-specific methods described in the previous section, obtaining normalized data, i.e., $\widetilde{\boldsymbol{X}}^{ATAC} \in R^{n \times r}$, $\widetilde{\boldsymbol{X}}^{RNA} \in R^{n \times g}$, $\widetilde{\boldsymbol{X}}^{ADT} \in R^{n \times p}$ and $\boldsymbol{M} \in R^{n \times 2}$, where $n = n_1 + n_2$ represents the number of spots from the two samples.

Cross-sample graph construction. Due to inconsistent dissection and placement of sequencing tissues, the spatial coordinates between different samples can hardly be directly compared with each other, resulting in the failure of representation learning models in capturing the relationship between different samples[22, 24]. To address this gap, we proposed a strategy to construct a cross-sample graph. Firstly, we find intra-sample spatial neighbors for each spot based on the spatial coordinate matrix, building undirected edges between all spots and their k nearest neighbors in Euclidean space from the same sample. Secondly, we find inter-sample neighbors based on specified feature matrices, such as ATAC, RNA and ADT profiles, or even latent embeddings, by establishing inter-sample undirected edges between all spots and their k most similar neighbors in cosine space from other samples. Note that the parameter k is set to 6 by default. Finally, we acquire a cross-sample graph which consists of both intra-sample edges and inter-sample edges.

First-stage pretraining. In the first stage, we pretrain the parameter of PRESENT-BC based on a cross-sample graph and multi-omics sequencing data of different samples, including $\widetilde{\boldsymbol{X}}^{ATAC} \in R^{n \times r}$, $\widetilde{\boldsymbol{X}}^{RNA} \in$

691 $R^{n\times g}$ and $\widetilde{\boldsymbol{X}}^{\boldsymbol{ADT}} \in R^{n\times p}$. Specifically, the intra-sample edges in the cross-sample graph are built based on

692 the Euclidean distance between spots using their spatial coordinates, while the inter-sample edges are

693 establish based on the cosine similarity between different spots using their ATAC, RNA and ADT profiles.

694 The objective function and training process of PRESENT-BC in the first pretraining stage are identical to

695 those of PRESENT, as described in the previous section.

696 **Second-stage integration.** The first pretraining stage enables PRESENT-BC to capture biological

697 variations within each sample, while the technical variations between different samples, i.e., batch effects,

698 have not been fully addressed. In the second stage, we design effective mechanisms, including an inter-

699 batch alignment loss, a batch-adversarial learning strategy and an intra-batch preserving loss, to eliminate

700 the batch effects across samples while simultaneously preserving the shared and sample-specific

701 biological signals, which are described in detail in the subsequent subsections.

702 ***Inter-batch alignment.*** We designed a inter-batch alignment (IBA) loss to align spots from different

703 samples based on contrastive learning[45]. Firstly, we find mutual nearest neighbors from other samples as

704 the positive set of for spot i, denoted as N_i^{pos}, referring to the inter-sample neighbors that are connected

705 to spot i in the cross-sample graph. Secondly, we randomly select $1/4$ proportion of spots from the

706 same sample as the negative set of spot i, depicted as N_i^{neg}. Finally, the IBA loss of spot i is:

707 $$\mathcal{L}_{IBA}^{(i)} = -log\frac{\sum_{j\in N_i^{pos}} exp(\delta_{sim}(\boldsymbol{e_i}, \boldsymbol{e_j}))}{\sum_{j\in N_i^{pos}} exp(\delta_{sim}(\boldsymbol{e_i}, \boldsymbol{e_j})) + \sum_{j\in N_i^{neg}} exp(\delta_{sim}(\boldsymbol{e_i}, \boldsymbol{e_j}))}$$

708 where $\boldsymbol{e_i} \in \boldsymbol{R}^d$ denotes the latent representation of spot i, d is the output dimension of PRESENT-BC

709 and δ_{sim} denotes the cosine similarity. Therefore, the inter-batch alignment loss of all the spots is:

710 $$\mathcal{L}_{IBA} = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}_{IBA}^{(i)}$$

711 ***Batch-adversarial learning.*** Inspired by a domain-adversarial learning research[92] and its's success in

712 multi-batch integration of single-cell ATAC-seq data[46], we implement a batch-adversarial learning

strategy to remove batch effect contained in joint representations. Specifically, we introduce a batch-discriminator, aiming to predict the batch index of each spot based on its latent embedding, while the omics-specific encoders and the cross-omics alignment module of PRESENT-BC are treated as the generator, aiming to produce latent embeddings without batch effect, thereby confusing the batch-discriminator in an adversarial manner. The loss function of the batch-adversarial learning is:

$$\mathcal{L}_{ADV} = \max_{G} \min_{D} CE(P, T)$$

where G refers to the generator, D denotes the batch-discriminator, CE represents the cross-entropy loss, P denotes the predicted batch indices and T refers to the ground truth batch labels.

Intra-batch preserving. The first pretraining stage empowers PRESENT-BC to capture the biological variations within each sample, while the inter-batch alignment loss and the batch-adversarial learning strategy in the second integration stage aims for batch correction across samples. However, the key challenge of multi-sample integration is how to eliminate batch effects while simultaneously preserving biological signals. We therefore proposed an intra-batch preserving loss to iteratively maintain the pattern of latent representations within each batch during batch correction, denoted as:

$$\mathcal{L}_{IBP} = \frac{1}{S}\sum_{k=1}^{S} \frac{1}{|s_k|} \sum_{i \in s_k} KLD\left(q_{s_k}^{(i)}\left(\boldsymbol{e}^{(t)}\right), q_{s_k}^{(i)}\left(\boldsymbol{e}^{(t-1)}\right)\right)$$

where S represents the number of samples, s_k denotes the spots in sample k, $|s_k|$ denotes the number of spots in s_k, $\boldsymbol{e}^{(t)}$ and $\boldsymbol{e}^{(t-1)}$ refer to the joint embeddings of all spots in current epoch and the last epoch, respectively. $q_{s_k}^{(i)}(*)$ represents the embedding distribution of spot i over other spots in sample k, defined as (taking the embeddings of current iteration $\boldsymbol{e}^{(t)}$ as an example):

$$q_{s_k}^{(i)}\left(\boldsymbol{e}^{(t)}\right) = \left(\frac{\delta_{ed}\left(\boldsymbol{e}_{\boldsymbol{i}}^{(t)}, \boldsymbol{e}_{s_k^{(1)}}^{(t)}\right)}{\sum_{j=1}^{|s_k|} \delta_{ed}\left(\boldsymbol{e}_{\boldsymbol{i}}^{(t)}, \boldsymbol{e}_{s_k^{(j)}}^{(t)}\right)}, \frac{\delta_{ed}\left(\boldsymbol{e}_{\boldsymbol{i}}^{(t)}, \boldsymbol{e}_{s_k^{(2)}}^{(t)}\right)}{\sum_{j=1}^{|s_k|} \delta_{ed}\left(\boldsymbol{e}_{\boldsymbol{i}}^{(t)}, \boldsymbol{e}_{s_k^{(j)}}^{(t)}\right)}, \ldots, \frac{\delta_{ed}\left(\boldsymbol{e}_{\boldsymbol{i}}^{(t)}, \boldsymbol{e}_{s_k^{(|s_k|)}}^{(t)}\right)}{\sum_{j=1}^{|s_k|} \delta_{ed}\left(\boldsymbol{e}_{\boldsymbol{i}}^{(t)}, \boldsymbol{e}_{s_k^{(j)}}^{(t)}\right)}\right)$$

$$\delta_{ed}(\boldsymbol{u}, \boldsymbol{v}) = exp\left(\frac{\boldsymbol{u}^T \boldsymbol{v}}{\tau}\right)$$

where $s_k^{(j)}$ denotes the j-th spot in sample k, τ is a temperature parameter and set to 0.1 by default[42].

Loss functions of PRESENT-BC and the parameter updating process. Batch-adversarial learning strategy separate the parameter updating process of each epoch into two steps. In the first step, we update the parameters of the batch-discriminator using the cross-entropy loss $\mathcal{L}_{ADV}^{disc} = \beta_3 CE(P, G)$, with parameters of the generator and omics-specific decoders fixed. The generator denotes the omics-specific encoders and cross-omics alignment module of PRESENT-BC. In the second step, we fix the parameters of batch-discriminator and update the parameters of generator and omics-specific decoders based on an overall loss function, consisting of the cross-omics representation loss, the inter-batch alignment loss, the intra-batch preserving loss and a negative cross-entropy loss for the batch-adversarial learning process:

$$\mathcal{L}_{integ} = \mathcal{L}_{repre} + \beta_1 \mathcal{L}_{IBA} + \beta_2 \mathcal{L}_{IBP} + \beta_3 \mathcal{L}_{ADV}^{gen}$$

$$\mathcal{L}_{ADV}^{gen} = -CE(P, G)$$

where β_1, β_2 and β_3 are coefficients of the corresponding loss function, and set to 1 by default.

Cyclic iterative optimization. The training process of the second integration stage is carried out in a cyclic iterative manner. First, we obtain the joint embeddings of multiple samples using the trained PRESENT-BC. Next, we construct a cross-sample graph according to the strategy described in the previous section, where intra-sample connections are established based on the spatial coordinates of spots and the inter-sample edges are based on the cosine similarity between the newly acquired joint embeddings of spots from different samples. Finally, we train the PRESENT-BC model for one epoch using the newly constructed cross-sample graph. The cyclic process, aiming to iteratively refine the cross-sample graph, is repeated until the model converges. The cross-sample graph is not only significant for integration of cross-sample relationships based on the GAT module in omics-specific encoders, but also important for the inter-batch alignment loss, which utilizes the inter-sample edges in cross-sample graph to construct positive neighbor sets for spots.

757 **Implementation details**

758 With the development of spatial omics sequencing technologies, the number of spots in sequenced tissue
759 has been quickly increasing from thousands to tens and even hundreds of thousands[10, 12, 15, 16, 93].
760 PRESENT and its extension PRESENT-BC are built upon graph neural networks, which face challenges
761 such as memory explosion and neighborhood expansion when the number of nodes in graph grows
762 large[94, 95]. To address these problems, we adopted necessary sparse adjacency matrix implementation
763 based on the PyG framework when deploying our model. Additionally, we implemented the ClusterGCN
764 algorithm to divide the entire graph into multiple small subgraphs[94], which enables our model to be
765 trained on GPUs with limited memory and extend to larger spatial omics data.

766 The input dimensions of ATAC and RNA data are generally much higher than the ADT data, resulting
767 in different hidden dimension settings for them. Specifically, the hidden and output dimensions of omics-
768 specific encoders for ATAC and RNA data were (1024, 512) and 50, respectively, while those of omics-
769 specific encoder for ADT data were $\left({2d_{adt}^{in}}/{3} + {d_{adt}^{out}}/{3}, {d_{adt}^{in}}/{3} + {2d_{adt}^{out}}/{3}\right)$ and d_{adt}^{out}. Here d_{adt}^{in}
770 denotes the input dimension of ADT data and $d_{adt}^{out} = \min(50, d_{adt}^{in})$ refers to the output dimension of
771 omics-specific encoder for ADT data. The final output dimensions of both PRESENT and PRESENT-BC
772 were set to 50 by default. Additionally, the hidden dimensions of omics-specific decoders were symmetric
773 to the corresponding omics-specific encoders, i.e., $({2d_{adt}^{in}}/{3} + {d_{adt}^{out}}/{3}, {d_{adt}^{in}}/{3} + {2d_{adt}^{out}}/{3})$ for ADT data
774 and (512, 1024) for ATAC and RNA data. We trained the model for at most 100 epochs using the Adam
775 optimizer, updating the parameters with an initial learning rate of 1e-3 and a weight decay of 1e-4.

776

777 **Data collection and description**

778 **Single-nucleus ATAC-seq and spatial ATAC mouse embryo data.** The single-nucleus ATAC-seq
779 mouse embryo data[10] measured the chromatin accessibility of 1,879 cells from E12, E13 and E15

780 developmental stages. The spatial ATAC mouse embryo data[10] is composed of six samples, including
781 E12.5-S1 (3,274 spots), E12.5-S2 (3,903 spots), E13.5-S1 (3,160 spots), E13.5-S2 (2,970 spots), E15.5-
782 S1 (2,246 spots) and E15.5-S2 (2,234 spots). The original study[10] merged the single-cell ATAC-seq
783 datasets and spatial ATAC datasets and called 248,047 peak regions for the merged dataset based on
784 MACS2 v2.2.6 package.

785 **Stereo-seq axolotl brain data.** We collected the Stereo-seq axolotl brain developmental datasets from a
786 recent study which explored the regeneration and development of axolotl telencephalon using spatial
787 transcriptomic atlas[7]. The datasets consist of six sections from developmental stage 44 (1,477 spots),
788 stage 54 (2,929 spots), stage 57 (4,410 spots), juvenile (11,698 spots), adult (8,243 spots), and
789 metamorphosis (7,441 spots), measuring the expression level of 12,704 genes.

790 **10x Visium mouse brain data.** The spatial transcriptomic postnatal day 56 (P56) mouse brain data,
791 including a sagittal anterior section and a sagittal posterior section, are collected from the 10x Genomics
792 Visium platform. The mouse sagittal anterior section contains the expression level of 32,285 genes on
793 2,823 spots, while the posterior section provides the expression level of 32,285 genes on 3,289 spots.

794 **SPOTS mouse spleen data.** The mouse spleen data[13] is generated by a high-throughput spatial
795 transcriptomics and proteomics co-profiling technology, i.e., Spatial PrOtein and Transcriptome
796 Sequencing (SPOTS). This dataset provides the spatial coordinates and the expression level of 32,285
797 genes and 21 proteins on 2,568 spots.

798 **10x Visium RNA-Protein human lymph node data.** The human lymph node dataset, generated by 10x
799 Genomics Visium transcriptomics and proteomics co-profiling technology, is collected from a spatial
800 multi-omics representation learning study[32]. The in-house dataset contains the spatial coordinates and the
801 expression level of 18,085 genes and 31 proteins on 3,484 spots.

802 **MISAR-seq mouse brain data.** The MISAR-seq mouse brain data[14] provides the paired sequencing data

803 of spatial gene expression and chromatin accessibility during mouse brain development. We collected
804 four manually annotated samples from original study[14], including E11.0 (1,258 spots), E13.5 (1,777
805 spots), E15.5 (1,949 spots) and E18.5 (2,129 spots). We called peak for each sample based on the MACS2
806 v2.2.9.1 package, obtaining 242,024, 271,126, 265,014 and 294,734 peak regions, respectively.
807
808 **Experiment setup and baseline methods**
809 **Cross-modality representation of spatial ATAC mouse embryo data.** The spatial ATAC mouse embryo
810 data firstly went through a quality control process, where peaks accessible in less than 0.5% spots were
811 filtered out. Using the filtered ATAC count matrices as input, we then benchmarked the performance of
812 PRESENT with four methods designed for single-cell ATAC-seq data, including SCALE[49], cisTopic[50] and
813 SnapATAC2[51], and three methods designed for SRT data, i.e., SpaGCN[29], STAGATE[30] and PAST[17]. All
814 the baseline methods were implemented according to default settings except for STAGATE and PAST. In
815 original tutorials, the two SRT methods adopted a feature selection process which is specially designed
816 for transcriptomic data and only selected 3,000 features, resulting significantly lower performance
817 (Supplementary Fig. 31). Since other methods utilized all the input peaks for dimension reduction, we
818 ignored the feature selection step of these two methods for fairness and better performance.
819 **Cross-modality representation of Stereo-seq axolotl brain data.** The genes with zero expression in all
820 spots were filtered out. The filtered count matrices and the spatial locations are fed into PRESENT and
821 other baseline methods, including SpaGCN[29], STAGATE[30], PAST[17] and SpatialPCA[18]. All baseline
822 methods are implemented with default settings in their tutorials. Following the workflow of processing
823 single-cell[96] and spatial[17, 18, 30] RNA-seq data, PRESENT also employed a gene selection procedure,
824 obtaining 3,000 highly variable genes as input.
825

826 **Cross-modality representation of 10x Visium RNA-Protein human lymph node data.** We filtered out
827 genes and proteins with zero expression in all spots. Baseline methods includes SpatialGlue[32], totalVI[57],
828 CiteFuse[58] and Seurat-v4[59]. All baseline methods were conducted according to their tutorials.
829 **Cross-modality representation of MISAR-seq mouse brain data.** We filtered out peaks accessible in
830 less than 3% spots and genes with zero expression in all spots for quality control. The filtered count
831 matrices of ATAC data and RNA data along with the spatial coordinate matrix were then fed into the
832 workflow of PRESENT and other baseline methods, involving SpatialGlue[32], MOFA+[68], scAI[69],
833 scMDC[70], scMVP[19], Seurat-v4[59] and SnapATAC2[51]. All these baseline methods are implemented
834 according to their default settings. PRESENT again selected 3,000 highly variable genes as the
835 transcriptomic input for further analysis.
836 **Multi-sample integration of 10x Visium mouse brain data.** The intersected gene set of the two mouse
837 brain samples, including a sagittal anterior section and a sagittal posterior section, were utilized as the
838 unified gene set. Afterwards, gene count matrices of the two samples were concatenated and genes with
839 zero expression in all spots were filtered out. Next, 3,000 highly variable genes were selected as the input
840 for PRESENT-BC. PRESENT-BC was compared with two recently developed SRT methods, i.e.,
841 STAligner[25] and GraphST[23], which were conducted according to their default settings. Specifically, we
842 utilized PASTE[22] to accomplish the initial coordinate alignment step for GraphST according to the
843 recommendation of the original study[23].
844 **Multi-sample integration of spatial ATAC mouse embryo data.** The three spatial ATAC mouse embryo
845 samples from three developmental stages have the same peak sets and were directly concatenated.
846 Following the experimental settings in cross-modality representation of a single spatial sample, peaks
847 accessible in less than 0.5% spots were filtered out. Using the filtered concatenated matrix, concatenated
848 spatial matrix and batch indices as input, we benchmarked PRESENT-BC with six integration methods

849 developed for single-cell data (MNN[79], Scanorama[39], Harmony[38], SCALEX[80], scBasset-BC[81] and
850 SnapATAC2[51]) and two recently proposed SRT methods (STAligner[25] and GraphST[23]). MNN, Scanorama
851 and Harmony performed batch correction based on the latent embeddings obtained by a widely used
852 epigenomic toolkit, EpiScanpy[97]. All the methods are implemented according to their default settings
853 except for STAligner. Due to similar reasons in the cross-modality representation for a single sample, we
854 ignored the feature selection step of STAligner to ensure better performance and fairness.
855 **Multi-sample integration of MISAR-seq mouse brain data.** We merged the peak sets of different
856 mouse brain samples according to the merging objects vignette provided by the Signac package[90], which
857 first obtained a unified peak set by merging all intersecting peaks from all the samples and then quantified
858 the fragment count of each spot on the newly obtained peak set. Consistent with the experimental settings
859 in cross-modality representation, peaks accessible in less than 3% spots in the concatenated matrix were
860 filtered out and 3,000 highly variable genes of transcriptomic data were selected for PRESENT-BC. Due
861 to the lack of methods specially designed for the integration of spatial multi-omics data, we compared the
862 batch effect removal and biological variance conservation performance of PRESENT-BC with six
863 methods developed for single-cell data (MNN[79], Scanorama[39], Harmony[38], SCALEX[80], scBasset-BC[81]
864 and SnapATAC2[51]) and three recently proposed SRT multi-sample integration methods (STAligner[25],
865 Stitch3D[31] and GraphST[23]). Specifically, MNN, Scanorama, Harmony, SCALEX could be applied to both
866 transcriptomic and epigenomic data, denoted as RNA-* or ATAC-*, respectively, where * denotes the
867 different methods. MNN, Scanorama and Harmony performed batch correction based on low-dimensional
868 embeddings extracted from RNA or ATAC data, which were accomplished based on the workflow of
869 SCANPY[96] and EpiScanpy[97], respectively. All the methods were implemented using default settings.
870
871

Evaluation metrics

Spatial clustering metrics. We implemented four commonly used metrics[53, 54] to quantitatively evaluate the performance of spatial clustering, including adjusted rand index (ARI), adjusted mutual information (AMI), normalized mutual information (NMI) and homogeneity (Homo).

ARI, ranging from 0 to 1, denotes the consistent adjustment of the expectations of rand index:

$$\mathrm{ARI} = \frac{RI(P,T) - E[RI(P,T)]}{\max[RI(P,T)] - E[RI(P,T)]}$$

where $RI(\cdot)$ denotes the function of rand index, T refers to the ground truth labels, and P represents the clustering labels. Rand index is the probability of the agreement between the clustering result and the ground truth labels on a randomly selected spot set.

Mutual Information (MI) accesses the similarity between the clustering labels and ground truth labels, from which both AMI and NMI originate. AMI ranges from -1 to 1, while NMI ranges from 0 to 1. AMI is calculated as:

$$\mathrm{AMI} = \frac{MI(P,T) - E[MI(P,T)]}{avg[H(P), H(T)] - E[MI(P,T)]}$$

where $MI(\cdot)$ refers to the mutual information, $E(\cdot)$ represents the expectation function and $H(\cdot)$ denotes the function of information entropy. NMI can be calculated as:

$$\mathrm{NMI} = \frac{MI(P,T)}{\sqrt{H(P)H(T)}}$$

Homogeneity is another metric to measure the clustering performance. Specifically, if all the clusters involve only data points that are members of a single class in the ground truth labels, the clustering results could be interpreted as satisfying homogeneity, calculated as:

$$\mathrm{Homo} = 1 - \frac{H(T|P)}{H(T)}$$

Integration metrics. According to a recent benchmarking framework for single-cell omics data integration, two important perspectives need to be considered when evaluating the performance of multi-sample integration, namely the batch effect removal and the biological variance conservation[82].

895 To quantitatively evaluate the performance of batch effect removal, we introduced several metrics
896 used in the benchmarking framework[82], including graph connectivity(GC), k-nearest-neighbor batch-
897 effect test (kBET) and batch average silhouette width (bASW). In addition, we also implemented a cross-
898 validation strategy to evaluate the cross-batch annotation performance based on latent representations
899 obtained by different integration methods, which is also an important assessment of batch correction[98].
900 Specifically, using the spots from a specific sample as the test set and those from other samples along
901 with the domain labels as the training set, we transferred the domain labels of the training set into the test
902 set based on a k-NN classifier and the joint embeddings of all the samples. k-NN classifier was used as
903 its performance mainly depended on the quality of joint embeddings obtained by different methods. The
904 performance of label transfer was evaluated with four commonly used metrics implemented in the scikit-
905 learn package, i.e., accuracy (Acc), kappa, macro F1 score (mF1) and the weighted F1 score (wF1). After
906 all the samples were used as the test set in a cross-validation manner, we calculated the average score for
907 each metric to evaluate the performance of cross-batch annotation.
908 The quantitative evaluation for biological variance conservation involves the accessment of
909 representative quality and the clustering performance. The representative quality was accessed via the
910 mean average precision (MAP)[99] and two metrics included in the benchmarking framework[82], namely the
911 spot average silhouette width (sASW) and spot local inverse Simpson's index (sLISI) , while the
912 evaluation of clustering performance was based on the four metrics, i.e., ATI, AMI, NMI and Homo.
913 We calculated category scores of the batch effect removal and the biological variance conservation
914 through averaging the metrics mentioned in the two aspects, respectively. We also calculated an overall
915 score using a 40/60 weighted mean of the two category scores[82, 99].
916

917 **Downstream analyses**

918 **Visualization.** Based on the latent representations obtained by various methods, we utilized the uniform
919 manifold approximation and projection (UMAP)[55] algorithm to further reduce the dimension of latent
920 representation matrices into two-dimensional matrices for visualization. In addition, the two-dimensional
921 UMAP and spatial coordinate matrices were visualized using the functions *scanpy.pl.umap()* and
922 *scanpy.pl.spatial()* wrapped in the SCANPY package[96].

923 **Spatial clustering.** We implemented a binary search strategy to tune the resolution of Leiden algorithm[52],
924 a graph-based clustering method, to match the number of spatial clusters as closely as possible with the
925 number of ground truth domains. Specifically, we set a minimum resolution res_{min} and a maximum
926 resolution res_{max}, and obtain clustering results using a resolution equal to $\frac{res_{min}+res_{max}}{2}$. If the number
927 of clusters exceeds the number of ground truth domains, res_{max} is updated to $\frac{res_{min}+res_{max}}{2}$. Otherwise,
928 res_{min} is updated to $\frac{res_{min}+res_{max}}{2}$. The iterative process continues until the number of clusters matches
929 the number of ground truth domains or the current iteration reaches the maximum number of iterations.

930 **Differentially expressed gene and protein analysis.** We implemented differential analysis based on the
931 Wilcoxon rank-sum test[61] wrapped in *scanpy.tl.rank_genes_groups()* function of the SCANPY package[96].

932 **Gene ontology analysis.** Given a specific gene set, we conducted gene ontology (GO) analysis using the
933 clusterProfiler package[100], which automated the process of biological-term classification and the
934 enrichment analysis of gene clusters.

935

936 ## Data availability

937 The single- nucleus ATAC-seq and spatial ATAC mouse embryo datasets[10] are available at Gene
938 Expression Omnibus (GEO) using the accession code GSE214991. The Stereo-seq axolotl brain datasets[7]
939 as well as the reference dataset, i.e., axolotl brain data at 20 days post-injury, can be obtained through

940 their database https://db.cngb.org/stomics/artista/. The 10x Visium mouse brain datasets, including a
941 sagittal anterior section and a sagittal posterior section, are accessible at the 10x Genomics websites
942 https://www.10xgenomics.com/datasets/mouse-brain-serial-section-2-sagittal-anterior-1-standard and
943 https://www.10xgenomics.com/datasets/mouse-brain-serial-section-2-sagittal-posterior-1-standard,
944 respectively. The SPOTS mouse spleen data[13] is accessible at GEO with accession code GSE198353. The
945 10x Visium RNA & protein human lymph node dataset[32] can be downloaded from the Zenodo database
946 https://zenodo.org/records/10362607. The MISAR-seq mouse brain data[14] is available from the National
947 Genomics Data Center via accession number OEP003285.

948 **Code availability**

949 The code for implementing PRESENT and PRESENT-BC can be found at the Github website
950 https://github.com/lizhen18THU/PRESENT.

951 **Acknowledgements**

952 This work was supported by the National Key Research and Development Program of China (grant nos.
953 2023YFF1204802 and 2021YFF1200902) and the National Natural Science Foundation of China (grant
954 nos. 62273194).

955 **Author contributions**

956 R.J. conceived the study and supervised the project. Z.L. designed, implemented and validated the
957 PRESENT framework. X.J.C., X.Y.C., Z.G., Y. L., Y. P. and S.C. helped with analyzing the results. Z.L.
958 and R.J. wrote the manuscript, with input from all the authors.

959 **Competing interests**

960 The authors declare no competing interests.

961

References

1. Baysoy, A. et al. The technological landscape and applications of single-cell multi-omics. *Nat. Rev. Mol. Cell Biol.* **24**, 695-713 (2023).
2. Kashima, Y. et al. Single-cell sequencing techniques from individual to multiomics analyses. *Exp. Mol. Med.* **52**, 1419-1427 (2020).
3. Adema, K. et al. Lost in space: what single-cell RNA sequencing cannot tell you. *Trends Plant Sci.* (2024).
4. Armingol, E. et al. Deciphering cell-cell interactions and communication from gene expression. *Nat. Rev. Genet.* **22**, 71-88 (2021).
5. Cang, Z.X. et al. Screening cell-cell communication in spatial transcriptomics via collective optimal transport. *Nat. Methods* **20**, 218-228 (2023).
6. Shao, X. et al. Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with SpaTalk. *Nat. Commun.* **13**, 4429 (2022).
7. Wei, X. et al. Single-cell Stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration. *Science* **377**, eabp9444 (2022).
8. Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nat. Methods* **19**, 534-546 (2022).
9. Deng, Y.X. et al. Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature* **609**, 375-383 (2022).
10. Llorens-Bobadilla, E. et al. Solid-phase capture and profiling of open chromatin by spatial ATAC. *Nat. Biotechnol.* **41**, 1085-1088 (2023).
11. Vandereyken, K. et al. Methods and applications for single-cell and spatial multi-omics. *Nat. Rev. Genet.* **24**, 494-515 (2023).
12. Liu, Y. et al. High-plex protein and whole transcriptome co-mapping at cellular resolution with spatial CITE-seq. *Nat. Biotechnol.* **41**, 1405-1409 (2023).
13. Ben-Chetrit, N. et al. Integration of whole transcriptome spatial profiling with protein markers. *Nat. Biotechnol.* **41**, 788-193 (2023).
14. Jiang, F.Q. et al. Simultaneous profiling of spatial gene expression and chromatin accessibility during mouse brain development. *Nat. Methods* **20**, 1048-1057 (2023).
15. Yuan, Z.Y. et al. SODB facilitates comprehensive exploration of spatial omics data. *Nat. Methods* **20**, 387-399 (2023).
16. Xu, Z. et al. STOmicsDB: a comprehensive database for spatial transcriptomics data sharing, analysis and visualization. *Nucleic Acids Res.* **52**, D1053-D1061 (2024).
17. Li, Z. et al. Latent feature extraction with a prior-based self-attention framework for spatial transcriptomics. *Genome Res.* **33**, 1757-1773 (2023).
18. Shang, L. & Zhou, X. Spatially aware dimension reduction for spatial transcriptomics. *Nat. Commun.* **13**, 7203 (2022).
19. Li, G.Y. et al. A deep generative model for multi-view profiling of single-cell RNA-seq and ATAC-seq data. *Genome Biol.* **23**, 20 (2022).
20. Swanton, C. Intratumor heterogeneity: evolution through space and time. *Cancer Res.* **72**, 4875-4882 (2012).
21. Janiszewska, M. The microcosmos of intratumor heterogeneity: the space-time of cancer evolution. *Oncogene* **39**, 2031-2039 (2020).
22. Zeira, R. et al. Alignment and integration of spatial transcriptomics data. *Nat. Methods* **19**, 567-575 (2022).
23. Long, Y. et al. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nat. Commun.* **14**, 1155 (2023).
24. Guo, T.T. et al. SPIRAL: integrating and aligning spatially resolved transcriptomics data across different experiments, conditions, and technologies. *Genome Biol.* **24**, 241 (2023).

1007 25. Zhou, X., Dong, K.N. & Zhang, S.H. Integrating spatial transcriptomics data across different conditions,
1008 technologies and developmental stages. *Nat. Comput. Sci.* **3**, 894-906 (2023).
1009 26. Chen, S. et al. Spatially resolved transcriptomics reveals genes associated with the vulnerability of middle
1010 temporal gyrus in Alzheimer's disease. *Acta Neuropathol. Commun.* **10**, 188 (2022).
1011 27. Cheng, A., Hu, G. & Li, W.V. Benchmarking cell-type clustering methods for spatially resolved
1012 transcriptomics data. *Brief. Bioinform.* **24**, bbac475 (2023).
1013 28. Yuan, Z. et al. Benchmarking spatial clustering methods with spatially resolved transcriptomics data. *Nat.*
1014 *Methods* **21**, 1-11 (2024).
1015 29. Hu, J. et al. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains
1016 and spatially variable genes by graph convolutional network. *Nat. Methods* **18**, 1342-1351 (2021).
1017 30. Dong, K.N. & Zhang, S.H. Deciphering spatial domains from spatially resolved transcriptomics with an
1018 adaptive graph attention auto-encoder. *Nat. Commun.* **13**, 1739 (2022).
1019 31. Wang, G. et al. Construction of a 3D whole organism spatial atlas by joint modelling of multiple slices with
1020 deep neural networks. *Nat. Mach. Intell.* **5**, 1200-1213 (2023).
1021 32. Long, Y. et al. Integrated analysis of spatial multi-omics with SpatialGlue. Preprint at *bioRxiv*
1022 https://doi.org/10.1101/2023.04.26.538404 (2023).
1023 33. Tian, T. et al. Dependency-aware deep generative models for multitasking analysis of spatial omics data. *Nat.*
1024 *Methods* (2024).
1025 34. Berger, S.L. The complex language of chromatin regulation during transcription. *Nature* **447**, 407-412 (2007).
1026 35. MacNeil, L.T. & Walhout, A.J.M. Gene regulatory networks and the role of robustness and stochasticity in
1027 the control of gene expression. *Genome Res.* **21**, 645-657 (2011).
1028 36. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**,
1029 865-868 (2017).
1030 37. Peterson, V.M. et al. Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.*
1031 **35**, 936-939 (2017).
1032 38. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods*
1033 **16**, 1289-1296 (2019).
1034 39. Hie, B., Bryson, B. & Berger, B. Efficient integration of heterogeneous single-cell transcriptomes using
1035 Scanorama. *Nat. Biotechnol.* **37**, 685-691 (2019).
1036 40. Veličković, P. et al. Graph attention networks. Preprint at *arXiv* https://doi.org/10.48550/arXiv.1710.10903
1037 (2017).
1038 41. Blundell, C. et al. Weight uncertainty in neural network. In *International conference on machine learning*
1039 (2015).
1040 42. Lin, Z. et al. Multi-modal Contrastive Representation Learning for Entity Alignment. In *Proceedings of the*
1041 *29th International Conference on Computational Linguistics* (2022).
1042 43. Eraslan, G. et al. Single-cell RNA-seq denoising using a deep count autoencoder. *Nat. Commun.* **10**, 390
1043 (2019).
1044 44. Martens, L.D. et al. Modeling fragment counts improves single-cell ATAC-seq analysis. *Nat. Methods* **21**,
1045 28-31 (2023).
1046 45. Chen, T. et al. A simple framework for contrastive learning of visual representations. In *International*
1047 *conference on machine learning* (2020).
1048 46. Kopp, W., Akalin, A. & Ohler, U. Simultaneous dimensionality reduction and integration for single-cell
1049 ATAC-seq data using deep learning. *Nat. Mach. Intell.* **4**, 162-168 (2022).
1050 47. Heumos, L. et al. Best practices for single-cell analysis across modalities. *Nat. Rev. Genet.* **24**, 550-572
1051 (2023).
1052 48. Zhang, H. et al. Extensive evaluation of ATAC-seq protocols for native or formaldehyde-fixed nuclei. *BMC*

1053 *Genomics* **23**, 214 (2022).
1054 49. Xiong, L. et al. SCALE method for single-cell ATAC-seq analysis via latent feature extraction. *Nat. Commun.*
1055 **10**, 4576 (2019).
1056 50. González-Blas, C.B. et al. cisTopic: cis-regulatory topic modeling on single-cell ATAC-seq data. *Nat.*
1057 *Methods* **16**, 397-400 (2019).
1058 51. Zhang, K. et al. A fast, scalable and versatile tool for analysis of single-cell omics data. *Nat. Methods* **21**,
1059 217-227 (2024).
1060 52. Traag, V.A., Waltman, L. & Van Eck, N.J. From Louvain to Leiden: guaranteeing well-connected
1061 communities. *Sci. Rep.* **9**, 5233 (2019).
1062 53. Chen, S. et al. VPAC: Variational projection for accurate clustering of single-cell transcriptomic data. *BMC*
1063 *Bioinformatics* **20**, 139-151 (2019).
1064 54. Jiang, R. et al. SINFONIA: scalable identification of spatially variable genes for deciphering spatial domains.
1065 *Cells* **12**, 604 (2023).
1066 55. McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for dimension
1067 reduction. Preprint at *arXiv* https://doi.org/10.48550/arXiv.1802.03426 (2018).
1068 56. Pierson, E. & Yau, C. ZIFA: Dimensionality reduction for zero-inflated single-cell gene expression analysis.
1069 *Genome Biol.* **16**, 1-10 (2015).
1070 57. Gayoso, A. et al. Joint probabilistic modeling of single-cell multi-omic data with totalVI. *Nat. Methods* **18**,
1071 272-282 (2021).
1072 58. Kim, H.J. et al. CiteFuse enables multi-modal analysis of CITE-seq data. *Bioinformatics* **36**, 4137-4143
1073 (2020).
1074 59. Hao, Y.H. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 (2021).
1075 60. Lewis, S.M., Williams, A. & Eisenbarth, S.C. Structure and function of the immune system in the spleen. *Sci.*
1076 *Immunol.* **4**, eaau6085 (2019).
1077 61. Soneson, C. & Robinson, M.D. Bias, robustness and scalability in single-cell differential expression analysis.
1078 *Nat. Methods* **15**, 255-261 (2018).
1079 62. Schmid, M.C. & Varner, J.A. Myeloid cells in tumor inflammation. *Vasc. Cell* **4**, 1-7 (2012).
1080 63. Borges, M.D. & Sesti-Costa, R. Macrophages: key players in erythrocyte turnover. *Hematol. Transfus. Cell*
1081 *Ther.* **44**, 574-581 (2022).
1082 64. Seyfried, A.N., Maloney, J.M. & MacNamara, K.C. Macrophages orchestrate hematopoietic programs and
1083 regulate HSC function during inflammatory stress. *Front. Immunol.* **11**, 548463 (2020).
1084 65. Cui, K. et al. Endocytic adaptors in cardiovascular disease. *Front. Cell Dev. Biol.* **8**, 624159 (2020).
1085 66. Baranov, M.V. et al. Modulation of immune responses by particle size and shape. *Front. Immunol.* **11**, 607945
1086 (2021).
1087 67. Kurup, S.P., Butler, N.S. & Harty, J.T. T cell-mediated immunity to malaria. *Nat. Rev. Immunol.* **19**, 457-471
1088 (2019).
1089 68. Argelaguet, R. et al. MOFA+ : a statistical framework for comprehensive integration of multi-modal single-
1090 cell data. *Genome Biol.* **21**, 1-17 (2020).
1091 69. Jin, S.Q., Zhang, L.H. & Nie, Q. scAI: an unsupervised approach for the integrative analysis of parallel
1092 single-cell transcriptomic and epigenomic profiles. *Genome Biol.* **21**, 1-19 (2020).
1093 70. Lin, X. et al. Clustering of single-cell multi-omics data with a multimodal deep learning method. *Nat.*
1094 *Commun.* **13**, 7705 (2022).
1095 71. Sunkin, S.M. et al. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous
1096 system. *Nucleic Acids Res.* **41**, D996-D1008 (2012).
1097 72. Teicher, M.H., Anderson, C.M. & Polcari, A. Childhood maltreatment is associated with reduced volume in
1098 the hippocampal subfields CA3, dentate gyrus, and subiculum. *Proc. Natl. Acad. Sci.* **109**, E563-E572 (2012).

1099 73. Matsuda, K. et al. Transsynaptic modulation of kainate receptor functions by C1q-like proteins. *Neuron* **90**,
1100 752-767 (2016).
1101 74. Park, S.-Y. et al. Hippocalcin promotes neuronal differentiation and inhibits astrocytic differentiation in
1102 neural stem cells. *Stem Cell Rep.* **8**, 95-111 (2017).
1103 75. Hachem, S. et al. Expression of S100B during embryonic development of the mouse cerebellum. *BMC Dev.*
1104 *Biol.* **7**, 1-15 (2007).
1105 76. Hirasawa, M. et al. Carbonic anhydrase related protein 8 mutation results in aberrant synaptic morphology
1106 and excitatory synaptic function in the cerebellum. *Mol. Cell. Neurosci.* **35**, 161-170 (2007).
1107 77. Lassuthova, P. et al. Mutations in ATP1A1 cause dominant Charcot-Marie-Tooth type 2. *Am. J. Hum. Genet.*
1108 **102**, 505-514 (2018).
1109 78. Pacey, L.K. et al. Delayed myelination in a mouse model of fragile X syndrome. *Hum. Mol. Genet.* **22**, 3920-
1110 3930 (2013).
1111 79. Haghverdi, L. et al. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual
1112 nearest neighbors. *Nat. Biotechnol.* **36**, 421-427 (2018).
1113 80. Xiong, L. et al. Online single-cell data integration through projecting heterogeneous datasets into a common
1114 cell-embedding space. *Nat. Commun.* **13**, 6118 (2022).
1115 81. Yuan, H. & Kelley, D.R. scBasset: sequence-based modeling of single-cell ATAC-seq using convolutional
1116 neural networks. *Nat. Methods* **19**, 1088-1096 (2022).
1117 82. Luecken, M.D. et al. Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* **19**,
1118 41-50 (2022).
1119 83. Argelaguet, R. et al. Computational principles and challenges in single-cell data integration. *Nat. Biotechnol.*
1120 **39**, 1202-1215 (2021).
1121 84. Chen, S.Q. et al. RA3 is a reference-guided approach for epigenetic characterization of single cells. *Nat.*
1122 *Commun.* **12**, 13 (2021).
1123 85. Chen, X. et al. Cell type annotation of single-cell chromatin accessibility data via supervised Bayesian
1124 embedding. *Nat. Mach. Intell.* **4**, 116-126 (2022).
1125 86. Cusanovich, D.A. et al. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular
1126 indexing. *Science* **348**, 910-914 (2015).
1127 87. Cusanovich, D.A. et al. A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell* **174**, 1309-
1128 1324 (2018).
1129 88. Mulè, M.P., Martins, A.J. & Tsang, J.S. Normalizing and denoising protein expression data from droplet-
1130 based single cell profiling. *Nat. Commun.* **13**, 2099 (2022).
1131 89. Stoeckius, M. et al. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for
1132 single cell genomics. *Genome Biol.* **19**, 1-12 (2018).
1133 90. Stuart, T. et al. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333-1341 (2021).
1134 91. Qiu, P. Embracing the dropouts in single-cell RNA-seq analysis. *Nat. Commun.* **11**, 1169 (2020).
1135 92. Ganin, Y. et al. Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* **17**, 1-35 (2016).
1136 93. Zhang, D. et al. Spatial epigenome-transcriptome co-profiling of mammalian tissues. *Nature* **616**, 113-122
1137 (2023).
1138 94. Chiang, W.-L. et al. Cluster-GCN: An efficient algorithm for training deep and large graph convolutional
1139 networks. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery &*
1140 *data mining* (2019).
1141 95. Bai, J., Ren, Y. & Zhang, J. Ripple walk training: A subgraph-based training framework for large and deep
1142 graph neural network. In *2021 International Joint Conference on Neural Networks (IJCNN)* (2021).
1143 96. Wolf, F.A., Angerer, P. & Theis, F.J. SCANPY: large-scale single-cell gene expression data analysis. *Genome*
1144 *Biol.* **19**, 1-5 (2018).

1145 97. Danese, A. et al. EpiScanpy: integrated single-cell epigenomic analysis. *Nat. Commun.* **12**, 5228 (2021).
1146 98. Song, Y. et al. Benchmarking strategies for cross-species integration of single-cell RNA sequencing data. *Nat.*
1147 *Commun.* **14**, 6495 (2023).
1148 99. Cao, Z.-J. & Gao, G. Multi-omics single-cell data integration and regulatory inference with graph-linked
1149 embedding. *Nat. Biotechnol.* **40**, 1458-1466 (2022).
1150 100. Yu, G. et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS: J.*
1151 *Integrative Biol.* **16**, 284-287 (2012).
1152
1153

Figures

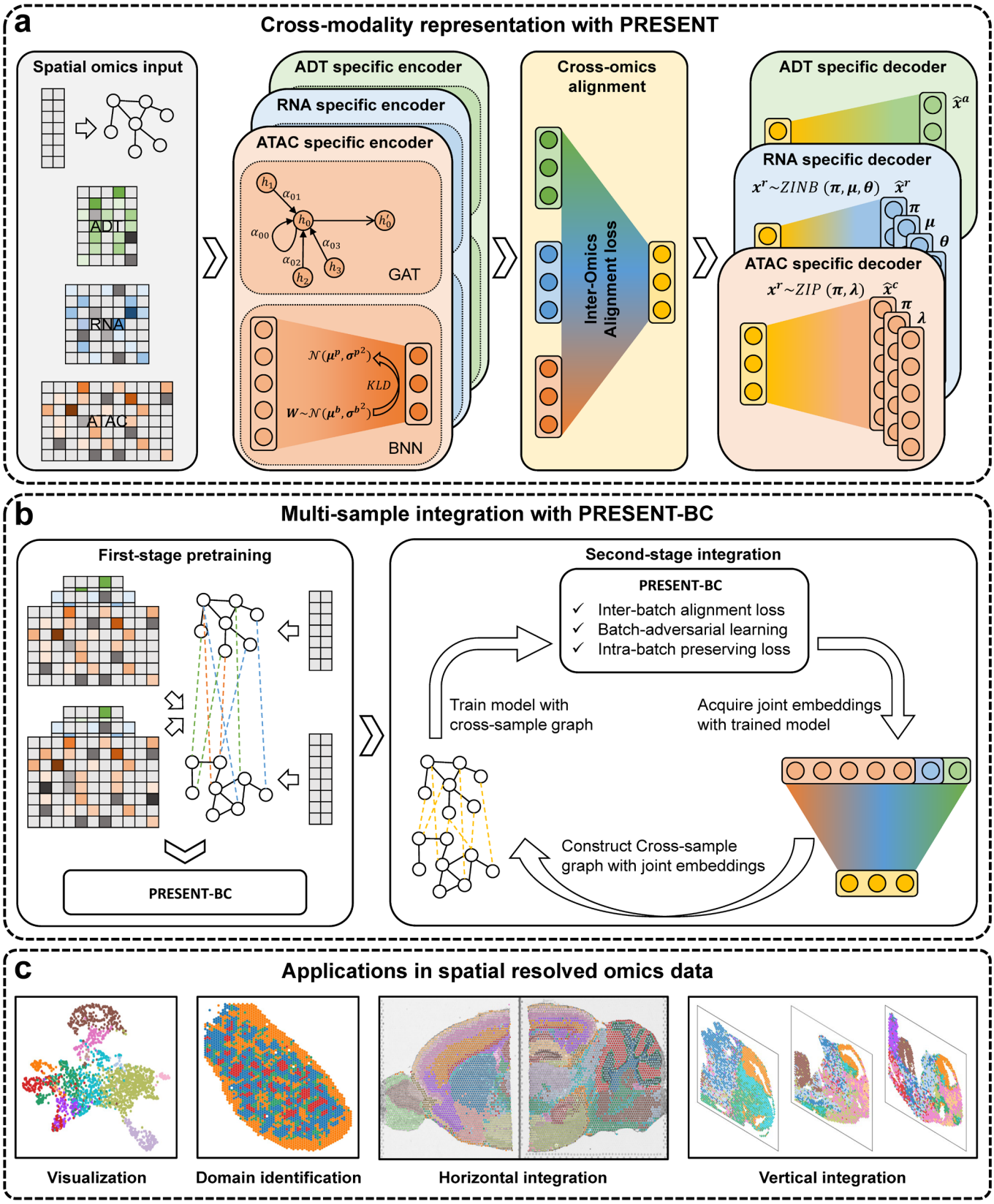

Fig. 1 | The framework of PRESENT. a, Cross-modality representation with PRESENT. PRESENT is capable to integrate spatial coordinate data and an arbitrary combination of different omics data, including proteomic data sequenced by antibody-derived tags (ADT), transcriptomic data by RNA sequencing (RNA), and epigenomic data generated by the assay for transposase-accessible chromatin with sequencing (ATAC). Here we illustrate the cross-modality representation of spatial information and all the three omics layers as an example. Firstly, an omics-specific encoder is used to extract

specific biological variations contained in each omics layer. Each omics-specific encoder is composed of a graph attention neural network (GAT) and Bayesian neural network (BNN) module. All the omics-specific feature matrices obtained by encoders are fused by a cross-omics alignment module consisting of an MLP module, which integrates multi-omics information via a contrastive learning inter-omics alignment loss function. The output of cross-omics alignment module refers to the final representation matrix of PRESENT. The distinct patterns of different omics layers are addressed via omics-specific decoders, which take the output of cross-omics alignment module as input and model the count matrices of RNA and ATAC data using the zero-inflation negative binomial (ZINB) and the zero-inflation Poisson (ZIP) distribution, respectively. **b,** Multi-sample integration with PRESENT-BC. PRESENT-BC adopts a two-stage workflow. In the first pretraining stage, PRESENT-BC has similar architecture as PRESENT and is pretrained using omics data and a cross-sample graph. The intra-sample edges of the cross-sample graph are built using spatial coordinates, while inter-sample connections are established based on the similarity of omics data. In the second integration stage, PRESENT-BC implements a cyclic iterative strategy to refine the cross-sample graph using the joint embeddings obtained by the trained model. PRESENT-BC introduces an inter-batch alignment loss, a batch-adversarial learning strategy and an intra-batch preserving loss in the second stage to eliminate batch effects while preserving biological signals. **c,** Applications of PRESENT in cross-modality representation and PRESENT-BC in multi-sample integration of spatially resolved omics data. The representation obtained by PRESENT can be applied to various downstream analysis, including but not limited to visualization and spatial domain identification. On the other hand, PRESENT-BC enables effective integration of both horizontal and vertical spatial samples generated by various technologies, facilitating integrated analysis of spatial datasets across different dissection regions and developmental stages.

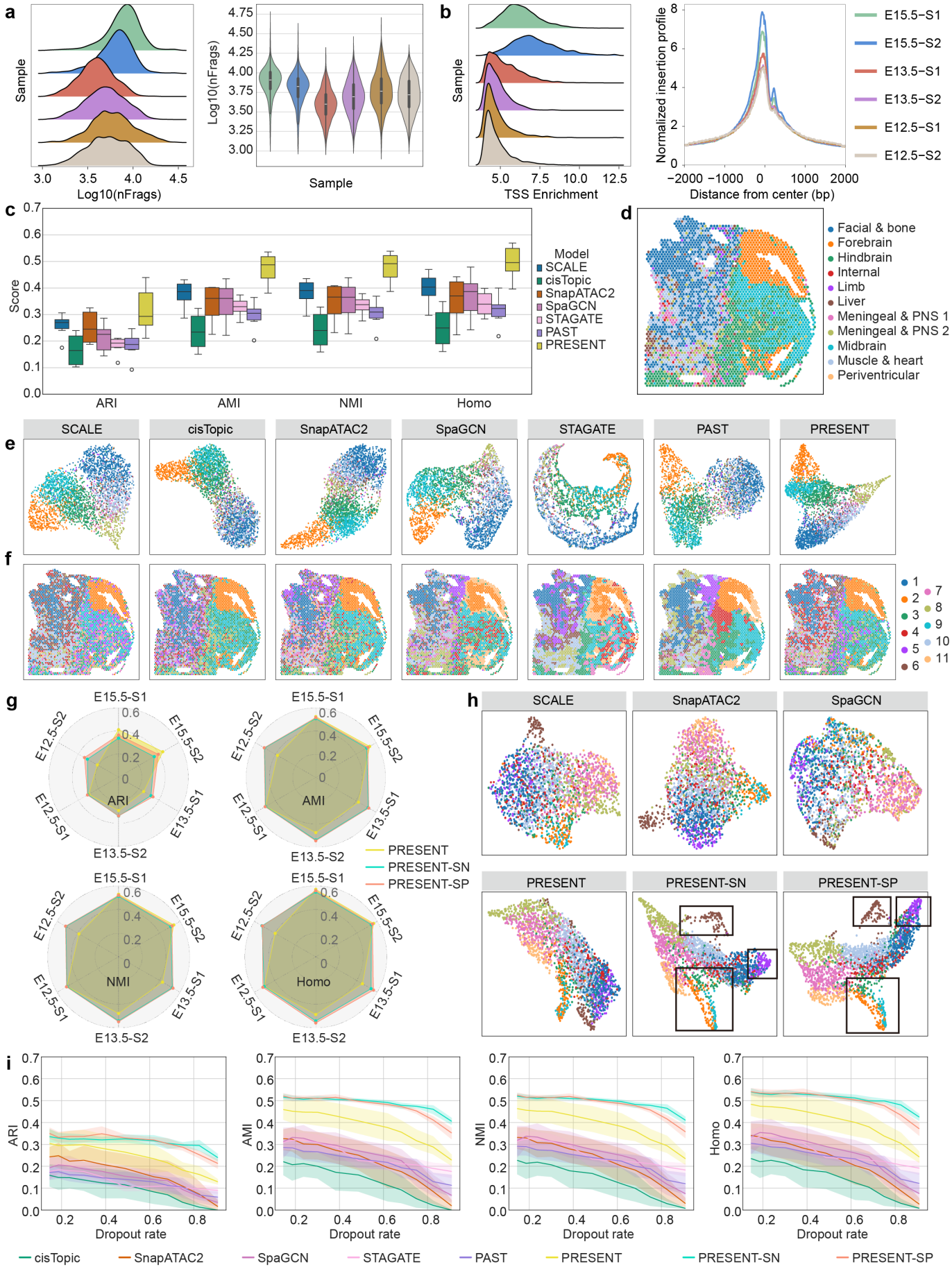

Fig. 2 | PRESENT facilitates accurate spatial domain identification in spatial ATAC-seq and SRT data. a, The ridge plot and violin plot of log10 total fragment counts, namely log10(nFrags), across the six spatial ATAC mouse embryo samples. **b,** The ridge plot of TSS enrichment scores for spots and the line plot of the enrichment profiles of fragments around the TSS across the six mouse embryo samples. **c,** The quantitative evaluation of spatial clustering performance of

different methods using four metrics, including ARI, AMI, NMI and Homo. The center line, box limits and whiskers in the box and violin plots are the median, upper and lower quartiles, and 1.5× interquartile range, respectively. **d,** The spatial visualization colored by ground truth domain labels, **e,** the UMAP visualization of the representations obtained by different methods colored by ground truth domain labels and **f,** the spatial visualization of clusters identified using the latent representations obtained by different methods on the E15.5-S1 mouse embryo sample. **g,** The detailed comparison of performance of PRESENT and PRESENT-SN & SP on the six mouse embryo samples based on radar plots. **h,** The UMAP visualization of the latent representations obtained by different methods colored by ground truth domain labels on E12.5-S2 mouse embryo sample. **i,** The spatial clustering performance of baseline methods, PRESENT and PRESENT-SN & SP on the six mouse embryo samples under different level of dropout rate. Note that SCALE encountered error when the dropout rate was larger than 0.2.

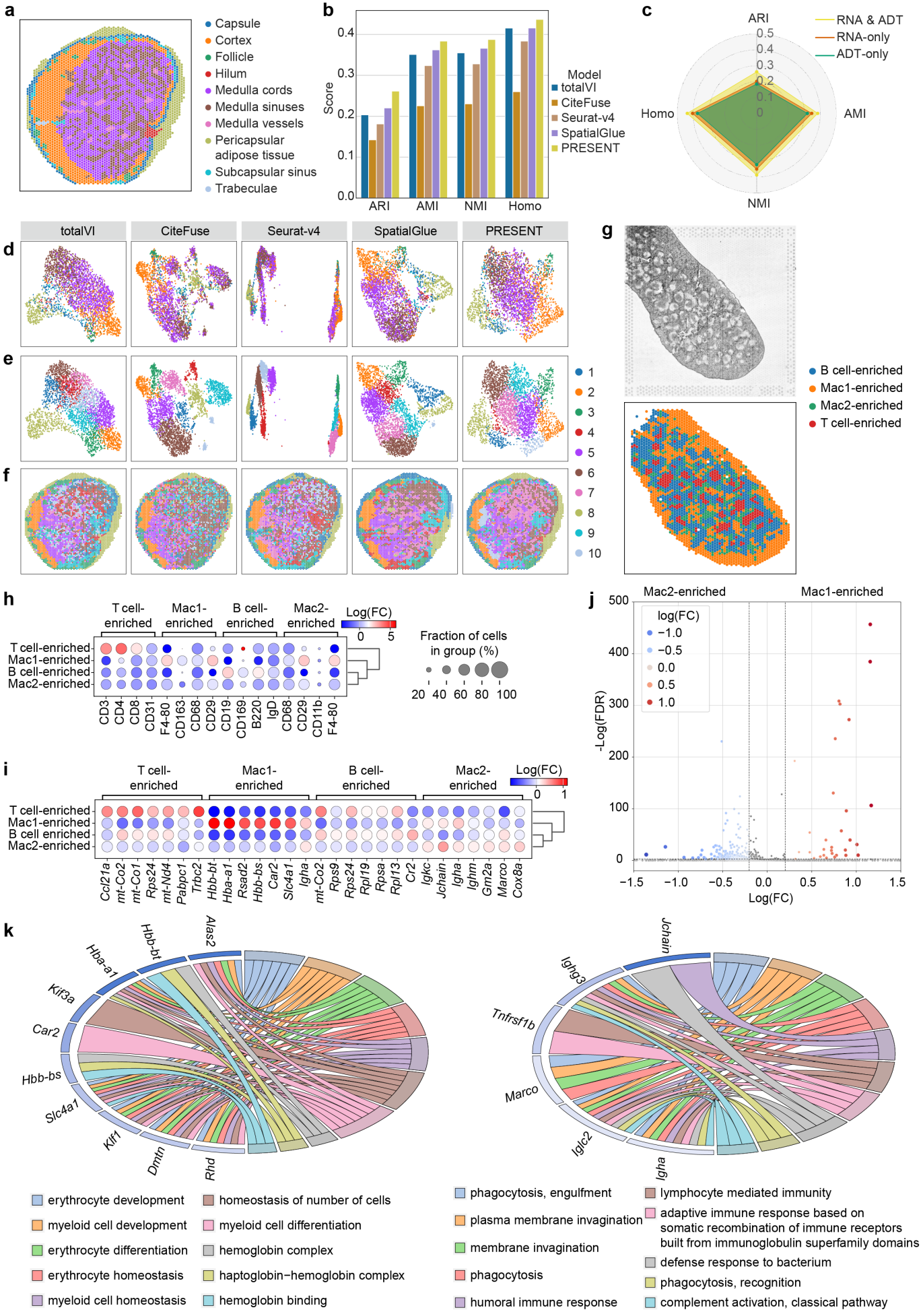
a
Capsule
Cortex
Follicle
Hilum
Medulla cords
Medulla sinuses
Medulla vessels
Pericapsular adipose tissue
Subcapsular sinus
Trabeculae
b
Score
ARI
AMI
NMI
Homo
Model
totalVI
CiteFuse
Seurat-v4
SpatialGlue
PRESENT
c
RNA & ADT
RNA-only
ADT-only
d
e
f
g
B cell-enriched
Mac1-enriched
Mac2-enriched
T cell-enriched
h
Log(FC)
Fraction of cells in group (%)
i
j
Mac2-enriched
Mac1-enriched
-Log(FDR)
Log(FC)
k
erythrocyte development
myeloid cell development
erythrocyte differentiation
erythrocyte homeostasis
myeloid cell homeostasis
homeostasis of number of cells
myeloid cell differentiation
hemoglobin complex
haptoglobin−hemoglobin complex
hemoglobin binding
phagocytosis, engulfment
plasma membrane invagination
membrane invagination
phagocytosis
humoral immune response
lymphocyte mediated immunity
adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains
defense response to bacterium
phagocytosis, recognition
complement activation, classical pathway

Fig. 3 | PRESENT leverages complementary omics information from paired spatial RNA-ADT data. a, The spatial visualization of the 10x Genomics Visium RNA-Protein human lymph node sample colored by ground truth domain labels. **b,** The quantitative comparison of spatial domain identification performance between PRESENT and other baseline methods based on bar plot and, **c,** the comparison between PRESENT utilizing both RNA and ADT data (RNA & ADT) and PRESENT using only RNA (RNA-only) or ADT data (ADT-only) based on radar plot on the human lymph node dataset. **d,** The UMAP visualization of latent representations obtained by different methods colored by ground truth domain labels and **e,** spatial cluster labels identified based on different methods on the human lymph node dataset. **f,** The spatial visualization of clusters identified based on different methods on the human lymph node dataset. **g,** The histology image and spatial visualization of spatial clusters identified based on PRESENT on the SPOTS mouse spleen dataset. **h,** The dot plot demonstrating the difference of ADT expression levels and **i,** RNA expression levels between each domain and the rest domains. **j,** The volcano plot illustrating the difference of RNA expression levels between Mac1 and Mac2-enriched domains. The vertical dashed line represents the threshold for $\log(\mathrm{FC}) = \pm 0.2$, while the horizontal dashed line denotes the threshold for $-\log(\mathrm{FDR} = 0.05)$. **k,** The chordal plots of the Mac1-enriched domain (left) and the Mac2-enriched domain (right) demonstrating the linkage of differential expressed genes (DEGs) in each domain and the corresponding enriched biological processes. In each chordal plot, the left semicircle represents DEGs while the right semicircle denotes the enriched biological processes.

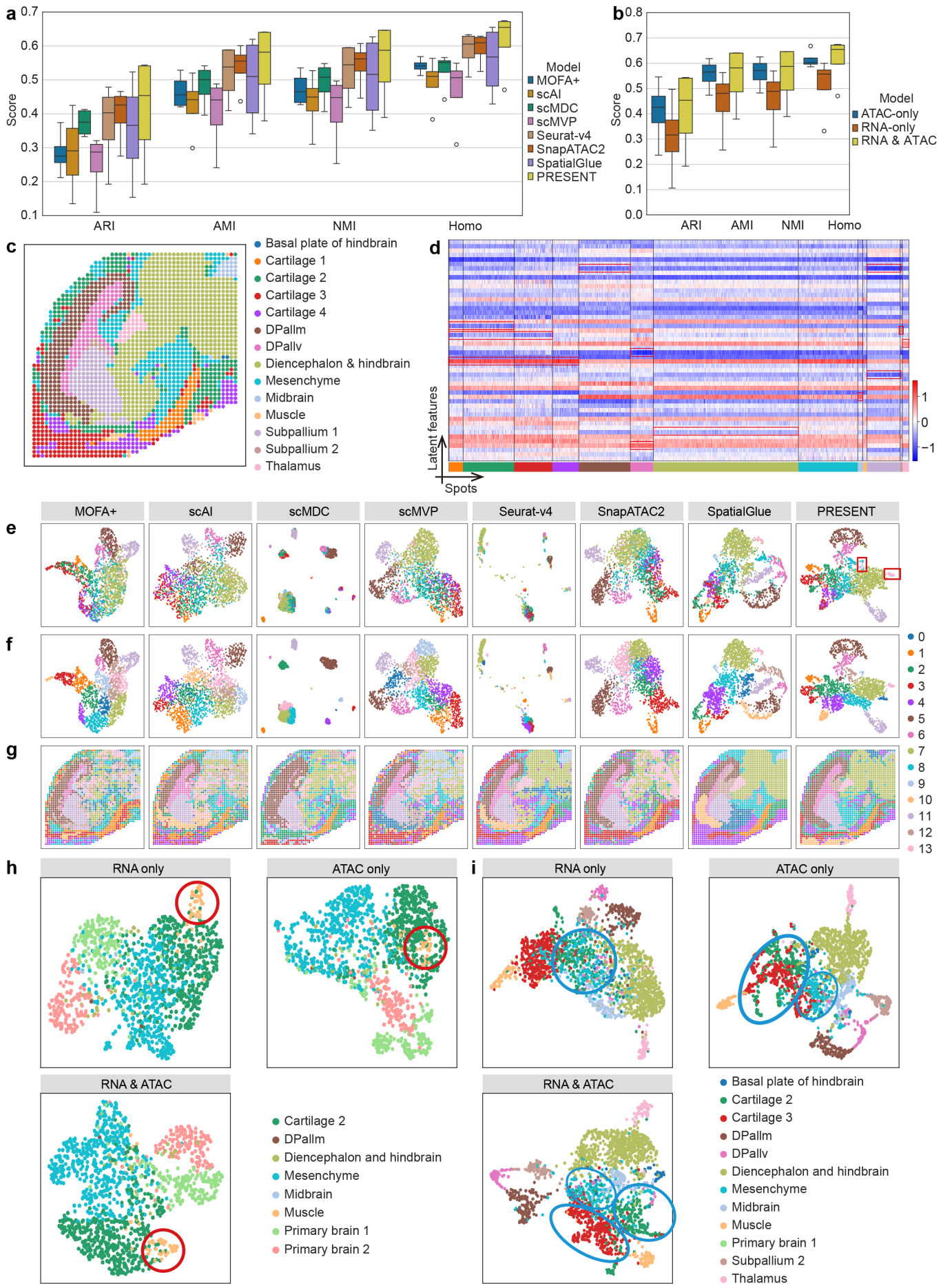

Fig. 4 | PRESENT extracts cross-modality representations from parallel spatial RNA-ATAC datasets. a, The quantitative comparison between PRESENT and other baseline methods and **b,** the comparison between PRESENT utilizing both RNA and ATAC data (RNA & ATAC) and PRESENT using only RNA (RNA-only) or ATAC (ATAC-only)

data based on box plot on the MISAR-seq mouse brain dataset. The center line, box limits and whiskers in the box plots are the median, upper and lower quartiles, and 1.5× interquartile range, respectively. **c,** The spatial visualization of spots from the E18.5 mouse brain sample colored by ground truth domain labels. **d,** The heatmap visualization of latent representations obtained by PRESENT. **e,** The UMAP visualization of latent representations obtained by different methods colored by ground truth domain labels. **f,** The UMAP visualization of latent representations obtained by different methods and **g,** the spatial visualization colored by spatial clusters identified based on different methods on the E18.5 mouse brain sample. The UMAP visualization of latent embeddings obtained by PRESENT using only RNA data, only ATAC data and both RNA & ATAC data, colored by ground truth domain labels on the **h,** E11.0 and **i,** E15.5 mouse brain samples.

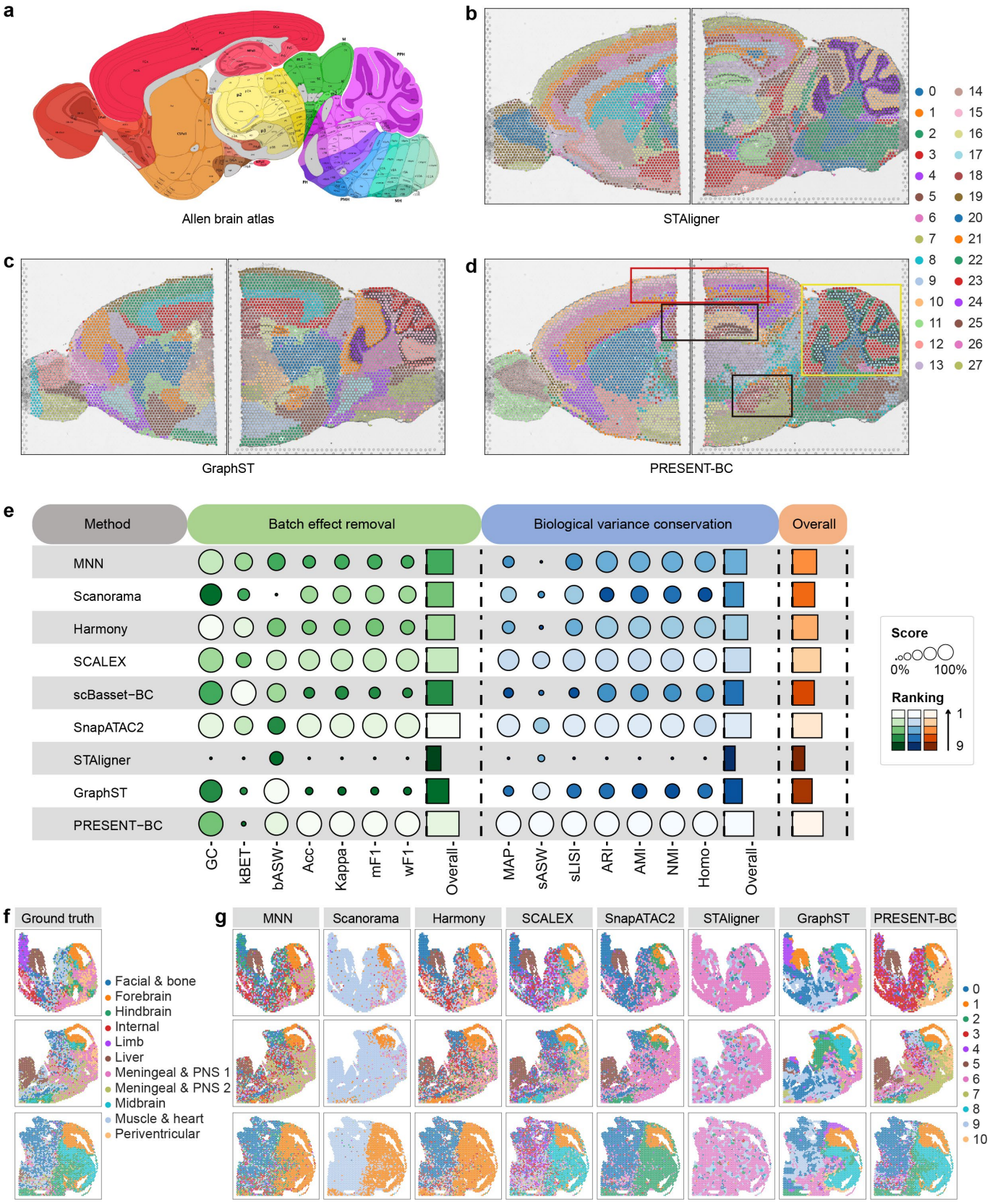

Fig. 5 | PRESENT-BC enables effective integration of horizontal and vertical spatially resolved samples. a, The anatomic annotation of the sagittal region in P56 mouse brain provided by Allen Reference Atlas. The joint spatial clustering results based on the latent embeddings obtained by **b,** STAligner, **c,** GraphST and **d,** PRESENT-BC on the two horizontal mouse brain sagittal samples generated by the 10x Genomics Visium platform. **e,** The quantitative evaluation of different integration methods on the three spatial ATAC mouse embryo samples using 14 metrics divided into two categories, namely batch effect removal and biological variance conservation. The category scores of these two aspects were calculated by averaging the metrics within each category. An overall score for each integration method was computed

using a 40/60 weighted mean of the category scores for batch effect removal and biological variance conservation (Methods). **f,** The spatial visualization of spots across the three spatial ATAC mouse embryo samples colored by ground truth spatial domains and **g,** the spatial clusters identified based on different integration methods. The first, second and third row of **f** and **g** denotes the samples from E12.5, E13.5 and E15.5 stages, respectively.

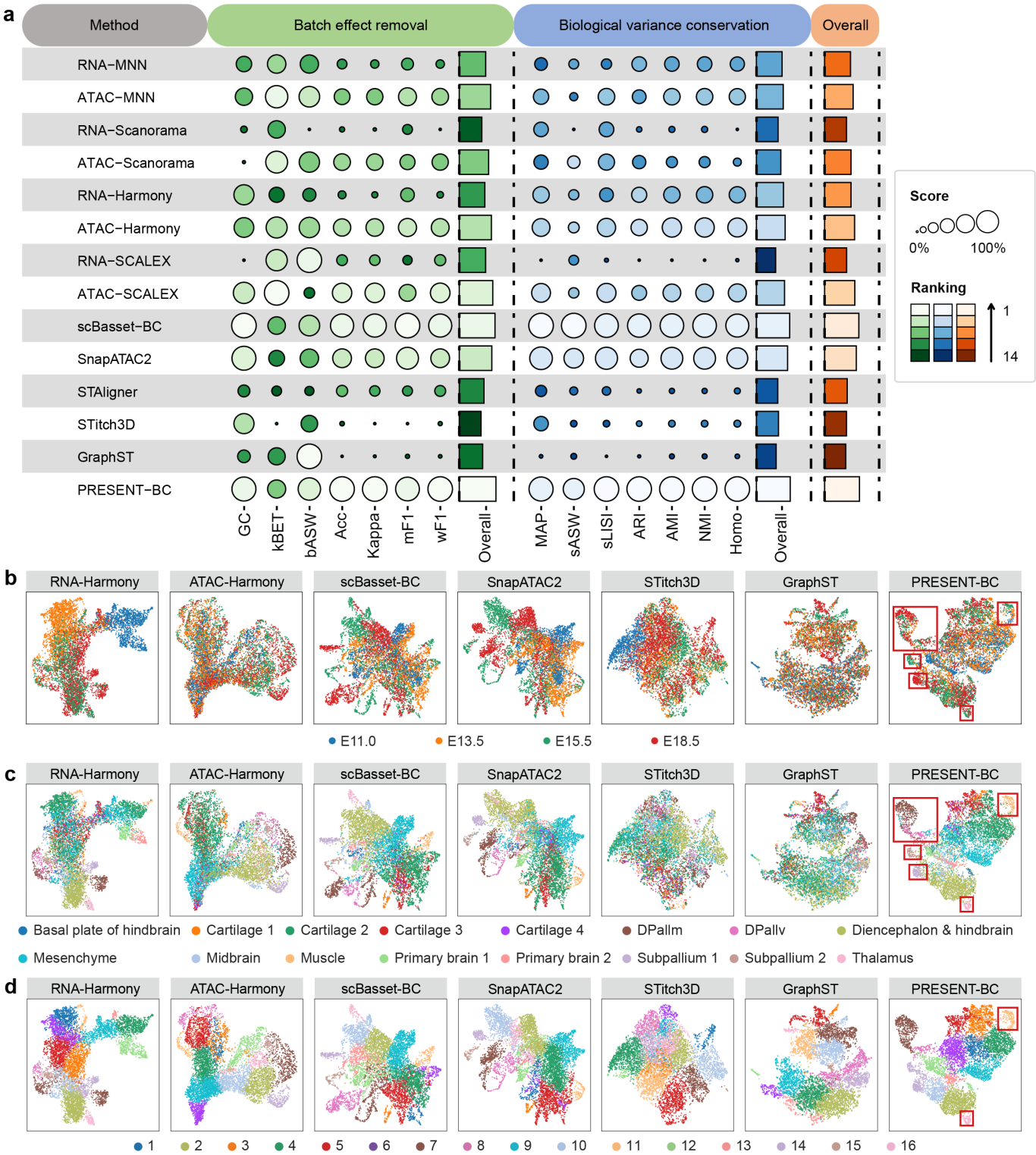

Fig. 6 | PRESENT-BC captures biological signals while eliminating batch effects in spatial RNA-ATAC data. a, The quantitative evaluation of different integration methods on the four MISAR-seq mouse brain samples using 14 metrics divided into two categories, namely batch effect removal and biological variance conservation. The category scores of these two aspects were calculated by averaging the metrics within each category. An overall score for each integration method was computed using a 40/60 weighted mean of the category scores for batch effect removal and biological variance conservation. **b,** The UMAP visualization of spots across the four MISAR-seq mouse brain samples colored by sample indices, **c,** ground truth domain labels and **d,** spatial clustering labels identified based on different integration methods.