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], Hairong Lv[1], Lei Zhai[3,*], 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

[3] Department of Critical Rehabilitation，Beijing Bo'ai Hospital, China Rehabilitation Research Center, Beijing 100068, China

[*] Corresponding author: drzhailei@163.com (Lei Zhai), ruijiang@tsinghua.edu.cn (Rui Jiang)

Abstract

Spatially resolved sequencing technologies have revolutionized our understanding of biological regulatory processes within the microenvironment by accessing the states of genomic regions, genes and proteins as well as spatial coordinates of cells. However, discrepancies between different modalities and samples hinder the analysis of spatial omics data, necessitating the development of advanced computational methods. In this article, we propose PRESENT, an effective and scalable contrastive learning framework, for the cross-modality representation and multi-sample integration of spatial multi-omics, epigenomics and transcriptomics data. Through comprehensive experiments on spatial datasets, PRESENT demonstrates superior performance across various species, tissues, and technologies. Specifically, PRESENT effectively integrates spatial dependency and omics information simultaneously, facilitating the detection of spatially functional domains and the exploration of biological regulatory mechanisms. Furthermore, PRESENT can be extended for the integrative analysis of tissue samples across different dissected regions or developmental stages, promoting the identification of hierarchical structures from systematic and spatiotemporal perspectives.

Introduction

Single-cell sequencing technologies have enabled the characterization of states and activities within individual cells, promoting molecular cell biology and disease research in areas such as cell atlas production, tumor immunology, and cancer genetics[1]. However, an indispensable step of single-cell sequencing techniques involves the dissociation of cells from the tissue of interest, resulting in the loss of spatial information[2]. Complex biological regulatory processes and cellular activities within the microenvironment are driven by intricate cell-cell interactions mediated by ligand-receptor partners, which heavily depend on the spatial distance between cells[3-5]. Recently, rapid advancements in spatially resolved sequencing technologies have enabled the simultaneous profiling of both spatial and omics data, such as the spatially resolved transcriptomics (SRT)[6, 7], spatial epigenomics assay for transposase-accessible chromatin sequencing (ATAC-seq)[8, 9] and spatial multi-omics such as spatial CITE-seq[10], SPOTS[11] and MISAR-seq[12]. These technologies have accumulated substantial spatially resolved omics data in repositories[9, 12] and databases[13, 14], thereby granting us a new perspective to decipher biological regulatory mechanisms in the context of microenvironment.

However, spatial coordinate data and omics data exhibit significant modality discrepancies, posing substantial challenges for the analysis of spatial omics data. Specifically, spatial coordinate data anchors the relative positions of cells within tissues, while omics data depicts the states of chromatin regions, genes and proteins involved in the gene regulatory processes within cells. Additionally, different layers of omics data further present distinct patterns and distributions[15-18]. For example, compared with RNA-seq, ATAC-seq datasets suffer from larger dimensionality, more severe sparsity and higher dropout rates, while proteomics datasets, sequenced by antibody-derived tags (ADT), demonstrate opposite characteristics, requiring targeted modeling for each omics layer. Cross-modality representation, a critical step in spatial omics data analysis, addresses these challenges by integrating multi-modal data from individual spatial

50 datasets into a low-dimensional embedding matrix. This approach generates efficient and informative
51 features for various downstream applications, including the detection of spatial domains, identification of
52 domain-specific markers, and elucidation of domain-related functions, thereby facilitating the
53 understanding of biological regulatory mechanisms within the microenvironment[19-21].

54 Due to technical limitations, such as size restrictions of the captured area during data acquisition,
55 individual spatial sequencing samples typically focus on a specific dissected region or developmental
56 stage[22, 23]. Consequently, discrepancies in dissection and placement prevent the direct comparison of
57 spatial coordinates between different samples[23]. Additionally, the intricate coupling between batch effects
58 and biological variations hinders the applications of single-sample representation methods to the joint
59 embedding of multiple samples, impeding the exploration of tissues under different conditions[23-25].
60 Therefore, the integrative analysis of multiple samples, known as multi-sample integration, is another
61 critical focus in the research of spatially resolved omics data. While facing greater challenges compared
62 to representation of individual samples, multi-sample integration offers a more comprehensive and
63 systematic perspective for deciphering the hierarchical structures of cellular organizations[22, 23, 26, 27]. For
64 example, literatures have demonstrated that integrated analysis of samples under different stages reveals
65 developmental dynamics in the early mouse embryo data[27, 28]. Moreover, recent studies have confirmed
66 that integrating disease and normal samples facilitates the identification of disease-specific domains and
67 promotes the investigation of disease onset and progression mechanisms[27, 29].

68 Many computational methods have been proposed to address the integrative analysis of SRT data[30],
69 including single-sample representation algorithms like SpaGCN[31], STAGATE[32] and PAST[19], as well as
70 multi-sample integration methods like STAligner[27], Stitch3D[33] and GraphST[22]. Despite the achievements,
71 the development of algorithms for spatial ATAC-seq[8, 9] and spatial multi-omics data[11, 12, 34] remains in
72 preliminary stage. Recently, SpatialGlue[35] was introduced as the first attempt for the single-sample

representation of spatial multi-omics data using an attentional feature fusion mechanism. Tian et al. proposed a dependency-aware generative variational autoencoder framework for the single-sample representation of SRT, spatial ATAC-seq and spatial CITE-seq data[36]. While these studies have addressed the cross-modality representation of individual spatial epigenomics and multi-omics samples, they have overlooked the integrated analysis of multiple samples across different conditions or time points, impeding the understanding of tissue structures from a spatiotemporal perspective. To address these challenges, there is an urgent need for advanced computational algorithms capable of integrating and analyzing cross-modality and multi-sample spatially resolved omics data.

To this end, we propose PRESENT, an advanced and scalable contrastive learning framework for the cross-modality representation and multi-sample integration of s**P**atially **RE**solved omics data, including spatial multi-omic**S, E**pigenomics, tra**N**scrip**T**omics (Fig. 1). Through comprehensive experiments on datasets generated by various spatial sequencing technologies, including spatial multi-omics[11, 12, 34], epigenomics[8, 9], and transcriptomics[6, 7], we demonstrated that PRESENT can simultaneously capture spatially coherent patterns and complementary multi-omics information, obtaining interpretable representations for various downstream analyses, including spatial visualization and the detection of spatial domains. Based on the Bayesian neural networks, PRESENT also offers the potential to incorporate reference data generated by single-cell or spatial technologies, addressing issues including low sequencing depth and signal-to-noise ratio associated with spatial data. Furthermore, PRESENT can be extended for the integrative analysis of multiple samples across different dissection regions and developmental stages, effectively eliminating batch effects while maintaining shared and sample-specific biological variations, paving the way to explore the hierarchical tissue structures and sophisticated cellular activities from systematic and spatiotemporal perspectives.

Results

Overview of PRESENT

Using a spatial multi-omics dataset that includes spatial coordinates, epigenomic profiles, transcriptomic expressions, and proteomic profiles, we illustrate the cross-modality representation architecture of PRESENT in Fig. 1a. Key challenges in the cross-modality representation of spatial multi-omics data include recognizing patterns within individual modalities and integrating multi-omics biological signals. Built on a multi-view autoencoder, PRESENT recognizes spatially biological variations contained in each omics layer via an omics-specific encoder consisting of graph attention neural networks[37] (GAT) and a Bayesian neural network[38] (BNN). The GAT module aims to incorporates spatial dependencies and omics information, while the BNN module, known for its effectiveness in non-linear regression problems[38], brings the potential to leverage reference data from existing sequencing data[19]. To further capture the distinct patterns of various omics layers, PRESENT employs omics-specific decoders which model the counts of transcriptomic and epigenomic data via zero-inflation negative binomial (ZINB) distribution[39] and zero-inflation Poisson (ZIP) distribution[40] (Methods), respectively. Moreover, to facilitate the integration of multi-omics biological signals, PRESENT fuses omics-specific embeddings from the encoders using a cross-omics alignment module which implements a multi-modal contrastive learning loss function, namely the inter-omics alignment (IOA) loss[41] (Methods).

Based on the above cross-modality representation framework, PRESENT can be further extended to the multi-sample integration of spatial multi-omics data, as illustrated in Fig. 1b. Critical challenges in multi-sample integration include addressing the spatial coordinate discrepancies between samples to capture cross-sample relationships[23] and distinguishing complexly coupled biological variations from technical batch effects[22, 26, 27, 33]. PRESENT adopts a two-stage training workflow to address these issues. In the first stage, PRESENT is pretrained based on the cross-sample graph and omics data acquired from

all the input samples. Specifically, the intra-sample edges of the cross-sample graph are established using the spatial coordinates, while the inter-sample connections are built using omics information (Methods). In the second stage, to align spots from multiple samples and eliminate technical batch effects from the coupled biological variations, we introduce a contrastive learning strategy[42] and a batch-adversarial learning strategy[43]. Additionally, we incorporate an intra-batch preserving loss during the training process to maintain biological signals within each sample while integrating multiple samples. Moreover, the second stage employs a cyclic iterative optimization process to refine the cross-sample graph using the joint embeddings generated by the newly trained model at each epoch, enabling a progressively improved characterization of cross-sample relationships (Methods).

On the other hand, PRESENT can be seamlessly adapted from spatial multi-omics data to spatial single-omics data, such as SRT and spatial epigenomics data, by simplifying the multi-view autoencoder framework into a single-view autoencoder (Methods). Through experiments on massive spatial datasets generated by spatial multi-omics, epigenomics and transcriptomics technologies, we have demonstrated that the joint embeddings obtained by PRESENT facilitate various downstream applications, including spatial visualization, domain detection, the integrative analysis of horizontal slices from different dissection regions, and vertical integration of samples from different developmental stages (Fig. 1c), paving the way to explore the cellular activities from spatiotemporal perspectives.

PRESENT facilitates accurate spatial domain identification in spatial RNA-ADT data

Spatial domain identification, which aims to decipher tissue regions with distinct anatomical structures, cell-type compositions, and cell-cell interactions[44, 45], is a fundamental step in revealing gene regulatory processes and tissue functions in spatial omics data[31, 32]. We first collected a spatial RNA-ADT human lymph node dataset generated by 10x Genomics Visium technology[35]. This dataset simultaneously

142 accessed the spatial locations, transcriptomic and proteomic expression of spots and was manually
143 annotated by the clinician group according to the histological image[35] (Fig. 2a). We benchmarked
144 PRESENT with SpatialGlue[35] and spaMultiVAE[36], two representation learning methods recently designed
145 for spatial multi-omics data, and several methods developed for single-cell RNA-ADT paired sequencing
146 data, including totalVI[46], CiteFuse[47] and Seurat-v4[48]. Based on the latent embeddings generated by these
147 methods, we implemented the Leiden algorithm[49] to acquire the same number of clusters as the ground
148 truth domains (Methods). Four commonly used metrics[50, 51], including adjusted rand index (ARI),
149 adjusted mutual information (AMI), normalized mutual information (NMI) and homogeneity (Homo),
150 were utilized to quantitatively evaluate the performance of spatial domain detection. As shown in Fig. 2b,
151 PRESENT achieved the best performance on all the four metrics for spatial domain identification. As a
152 qualitative evaluation and intuitive illustration of detected domains, we also demonstrated the uniform
153 manifold approximation and projection (UMAP) manifold structure[52] and spatial visualization results
154 (Fig. 2d, e, f). Specifically, CiteFuse and Seurat-v4 obtained unsatisfactory visualization results, where
155 domains such as medulla cords and sinuses scattered in the UMAP plot. spaMultiVAE detected stripe-like
156 spatial domains, resulting in biased spatial pattern identification in the lymph node tissue. Although
157 SpatialGlue and totalVI achieved relatively better results, they failed to recognize capsule from
158 pericapsular adipose tissue. In contrast, PRESENT successfully distinguished the pericapsular adipose
159 tissue (cluster 8), capsule (cluster 1), follicle (cluster 3) and cortex (cluster2 and cluster 10). Additionally,
160 we applied PRESENT to only a single omics layer of the human lymph node dataset, thus leveraging
161 expression information from only RNA data or ADT data, denoted as RNA-only and ADT-only,
162 respectively. As shown in Fig. 2c, PRESENT utilizing multiple omics layers (RNA & ADT) achieved
163 improved performance compared to the model that only leveraged RNA or ADT data, highlighting the
164 capabilities of PRESENT in integrating cross-modality spatial multi-omics information.

165 As the largest secondary lymphoid organ in the body, the spleen plays a significant role in
166 hematopoiesis and the clearance of erythrocytes, contributing significantly to immune functions[53]. We
167 further applied PRESENT to a SPOTS mouse spleen dataset[11], which measured spatial transcriptomic and
168 proteomic profiles simultaneously, to decipher the functionality of mouse spleen in the context of cellular
169 organizations. Through cross-modality representation and spatial clustering, PRESENT identified four
170 main spatial domains (Fig. 2g), denoted as the B cell-enriched domain, the T cell-enriched domain and
171 two macrophage-enriched subdomains, i.e., Mac1-enriched and Mac2-enriched domains, according to our
172 differentially expressed analysis (Fig. 2h, i). The differentially expressed analysis compared the protein
173 expression level of each domain with those of the rest domains based on the Wilcoxon rank-sum test[54]
174 (Fig. 2h), identifying protein markers distinguishing the B cell-enriched domain (CD3, CD4, CD8), the T
175 cell-enriched domain (CD19, B220, lgD) and the macrophage-enriched domain (F4-80, CD163, CD68),
176 consistent with the original study[11]. Specifically, the Mac1-enriched and Mac2-enriched domains had
177 nearly identical protein markers (Fig. 2h) as the macrophage markers identified in the original study[11],
178 making it hard to distinguish the two subdomains purely by ADT data. Therefore, we further conducted
179 differentially expressed analyses on the transcriptomic data between domains identified based on
180 PRESENT (Fig. 2i). Results demonstrated that Mac1-enriched domain could be marked by genes
181 including *Hbb-bt, Hbb-a1, Rsad2, Hbb-bs, Car2* and *Slc4a1*, while Mac-2 enriched domain could be
182 recognized with *Igkc, Jchain, Igha, Ighm, Gm2a* and *Cox8a.* The RNA expression levels between the
183 Mac1 and Mac2-enriched domains were also directly compared and numerous significantly differentially
184 expressed genes (DEGs) were found (Fig. 2j). These results showed that, despite the similarity of the two
185 macrophage-enriched subdomains in ADT expression levels, the distinction of transcriptomic patterns
186 between them could still be identified by PRESENT, emphasizing the ability of PRESENT in leveraging
187 complementary proteomic and transcriptomic information for effective domain identification.

188 To further decipher the functional distinction between the two macrophage-enriched subdomains, we
189 conducted Gene Ontology (GO) enrichment analysis based on the identified DEGs (Methods).
190 Specifically, for the Mac1-enriched domain, we enriched biological processes including erythrocyte
191 development, erythrocyte differentiation, erythrocyte homeostasis, myeloid cell development, myeloid
192 cell differentiation and myeloid cell homeostasis with false discovery rates (FDRs) of 1.29e-8, 5.48e-6,
193 6.70e-6, 3.59e-7, 8.71e-5 and 2.06e-5, respectively (Fig. 2k and Supplementary Table 1). Macrophages
194 are myeloid lineage cells which derive from bone marrow[55], reasonably explaining the enrichment of
195 development, differentiation and homeostasis of myeloid cell in the Mac1-enriched domain. In addition,
196 previous studies have stressed the importance of macrophages in the entire life of erythrocytes by
197 providing necessary factors for the survival, proliferation and development of erythrocytes[56, 57], agreeing
198 with the development, differentiation and homeostasis of erythrocyte found in the Mac1-enriched domain.
199 In comparison, the Mac2-enriched domain found biological processes including phagocytosis &
200 engulfment, plasma membrane invagination, membrane invagination, phagocytosis, humoral immune
201 response and lymphocyte mediated immunity with FDRs of 7.11e-5, 7.11e-5, 7.11e-5, 4.34e-4, 5.58e-4
202 and 5.94e-4, respectively (Fig. 2k and Supplementary Table 2). Phagocytosis and other related processes,
203 called the process of 'cell eating', are commonly considered as classical functions of macrophages[58, 59],
204 validating the enrichment of macrophages in the Mac2-enriched domain. On the other hand, compared
205 with the Mac1-enriched domain, the Mac2-enriched domain was spatially close to the B cell-enriched and
206 T cell-enriched domains (Fig. 2g), thereby explaining the enrichment of lymphocyte mediated immunity
207 process[60]. In conclusion, PRESENT effectively integrates multi-omics information and spatially coherent
208 patterns in spatial RNA-ADT data, thereby facilitating accurate spatial domain identification and
209 systematic exploration of biological implications in the identified domains.
210

211 **PRESENT integrates spatially aware multi-omics biological signals in spatial RNA-ATAC data**

212 Next, we collected MISAR-seq mouse brain datasets[12], which simultaneously measured the epigenomics
213 and transcriptomics, i.e., paired spatial RNA-ATAC profiles, of four mouse brain samples from four
214 developmental stages, involving E11.0, E13.5, E15.5 and E18.5. Compared with spatial-RNA-ADT co-
215 profiling data, spatial RNA-ATAC datasets suffer from higher dimension and sparsity. We firstly
216 compared the performance of PRESENT in spatial domain identification with the spatial multi-omics
217 method SpatialGlue[35] and six methods for the representation of single-cell RNA-ATAC data, including
218 MOFA+[61], scAI[62], scMDC[63], scMVP[21], Seurat-v4[48] and SnapATAC2[64]. Here spaMultiVAE[36], totalVI[46]
219 and CiteFuse[47] were not used as they are specially designed for single-cell or spatial RNA-ADT data. As
220 shown in Fig. 3a, PRESENT achieved overall the best spatial clustering performance, illustrating the its
221 superiority in extracting spatial multi-omics information into effective cross-modality embeddings for
222 accurate spatial domain identification. We also used PRESENT to obtain latent representations based on
223 only ATAC data or RNA data, denoted as ATAC-only and RNA-only respectively, which led to decreased
224 performance, highlighting the importance of integrating multi-omics information in deciphering
225 biological variations in the context of cellular organizations (Fig. 3b).

226 Taking the E18.5 mouse brain sample with the most complex structures as an example (Fig. 3c), we
227 intuitively visualized the latent embeddings learned by PRESENT using heatmap (Fig. 3d). The spots
228 from different spatial domains could be successfully distinguished via the combination of several latent
229 features or even a single feature in the latent feature space, including large domains such as cartilage 1-4,
230 mantle zone of dorsal pallium (DPallm), ventricular zone of dorsal pallium (DPallv), diencephalon &
231 hindbrain and subpallium 1, and small domains like midbrain and thalamus. We further demonstrated the
232 UMAP visualization of latent representations obtained by different methods on the E18.5 mouse brain
233 sample (Fig. 3e, f). Most baseline methods, including MOFA+, scAI, scMDC, scMVP and Seurat-v4,

234 could only roughly separate large domains such as DPallm, DPallv, diencephalon & hindbrain and
235 subpallium 1. SnapATAC2 could additionally recognize the cartilage 1 but still failed to distinguish
236 others. In comparison, spatially aware methods, including SpatialGlue and PRESENT, not only identified
237 large spatial domains such as diencephalon & hindbrain, subpallium 1, DPallm, DPallv, cartilage 1-4, but
238 also accurately distinguished the two smaller spatial domains, i.e., midbrain and thalamus. Furthermore,
239 domains identified by these two methods demonstrated enhanced spatial smoothness, highlighting the
240 importance of integrating spatial information (Fig. 3g). The UMAP and spatial visualization of other three
241 samples of the MISAR-seq mouse brain data[12] achieved similar results (Supplementary Figs. 1-4).
242 Additionally, we have conducted comprehensive experiments to investigate the impact of biological
243 signals from different omics on model performance to provide deeper insights into the model's ability to
244 extract multi-omics biological signals. Specifically, PRESENT utilizing only the transcriptomic data of
245 E11.0 brain sample, denoted as RNA-only, successfully captured the difference between the muscle and
246 the cartilage 2 domain, while the model based solely on epigenomic data, depicted as ATAC-only, failed
247 to do so (red circles in Fig. 3h). On the other hand, compared with the failure of PRESENT using only
248 transcriptomic data (RNA-only) in the E15.5 brain sample, the model based on the epigenomic data
249 (ATAC-only) contributed more to the differentiation between cartilage 2, cartilage 3 and mesenchyme
250 domains (blue circles in Fig. 3i). Therefore, relying solely on a single omics layer might lead to a biased
251 interpretation of biological signals in spatial domain detection. When taking both omics layers as input
252 (RNA & ATAC), PRESENT successfully captured variations among those domains in both E11.0 and
253 E15.5 mouse brain samples (Fig. 3h, i), highlighting the capabilities of PRESENT in leveraging spatially
254 aware transcriptomic & epigenomic multi-omics information for cross-modality representation and the
255 accurate identification of spatial domains from a comprehensive perspective.
256

257 **PRESENT eliminates batch effects while preserving biological signals in spatial multi-omics data**

258 The MISAR-seq mouse brain dataset simultaneously measured the transcriptomic and epigenomic
259 profiles of developmental brain samples from four stages, including E11.0, E13.5, E15.5 and E18.5,
260 providing a multi-omics landscape of the developmental mouse brain in the context of cellular
261 organizations[12]. However, varying experimental conditions and sequencing batches often introduce
262 potential technical biases, resulting in significant discrepancies in omics data across samples. These
263 discrepancies, commonly referred to as batch effects in single-cell studies, pose challenges to the
264 integrative analysis of multiple samples from different conditions and developmental stages[24, 26]. For
265 illustration, we applied joint principal component analysis (PCA) and the UMAP algorithm to raw RNA
266 and ATAC data of the four MISAR-seq mouse brain samples, revealing pronounced batch effects (Fig.
267 4a). Moreover, during the development of mouse brain, several functional structures might be specific to
268 certain developmental stages (e.g., primary brain), while others may persist throughout the entire
269 developmental process (e.g., midbrain and muscle), as shown in Fig. 4b. Therefore, the key challenges in
270 integrating the MISAR-seq mouse brain samples lied not only in the incorporation of spatial multi-omics
271 information and the correction of batch effects across samples, but also in the conservation of shared and
272 sample-specific biological variations.

273 Due to the lack of multi-sample integration methods specially designed for spatial multi-omics data,
274 we compared the integration performance of PRESENT with six methods developed for single-cell data
275 (MNN[65], Scanorama[25], Harmony[24], SCALEX[66], scBasset-BC[67] and SnapATAC2[64]) and three recently
276 proposed methods designed for SRT data (STAligner[27], Stitch3D[33] and GraphST[22]). Specifically, MNN,
277 Scanorama, Harmony and SCALEX could be applied to batch integration of either RNA-seq or ATAC-
278 seq data, denoted as RNA-* and ATAC-*, respectively, where * represents these methods. scBasset-BC
279 was featured as a sequence-based method specially designed for single-cell ATAC-seq data, while

280 STAligner, Stitch3D and GraphST were intended for integrative analysis of SRT data. SnapATAC2 has
281 provided guidance for incorporating multi-omics (i.e., transcriptomics & epigenomics) information for
282 integrated analysis, serving as a multi-sample integration workflow for single-cell multi-omics.

283 We first applied the UMAP algorithm to the joint embeddings generated by PRESENT and other
284 baseline methods for visualization, with the spots colored by batch indices (Fig. 4c and Supplementary
285 Fig. 5a) and ground truth domain labels (Fig. 4d and Supplementary Fig. 5b). PRESENT successfully
286 integrated spots from different developmental stages (red boxes in Fig. 4c) while maintaining the
287 biologically distinct patterns of DPallm, DPallv, subpallium 1 & 2, thalamus and muscle (red boxes in
288 Fig. 4d). Specifically, spots of muscle domain persisted throughout the entire life of mouse brain
289 development, while those of subpallium 1&2, DPallm, DPallv and thalamus only existed in specific
290 stages. The success of PRESENT in distinguishing these domains illustrated its capabilities in eliminating
291 batch effects while capturing shared and stage-specific biological variance. Additionally, we conducted
292 joint spatial clustering across the four stages based on the joint embeddings obtained by different
293 methods, as shown in Supplementary Fig. 6. Except for the scBasset-BC and SnapATAC2, other baseline
294 methods delivered limited clustering results. In contrast, PRESENT successfully detected not only large
295 domains like DPallm (cluster 7), but also small domains including thalamus (cluster 16) and muscle
296 (cluster 11). It is worth noting that when employing individual analysis of the sample from the E18.5
297 stage (Supplementary Fig. 4), PRESENT was unable to identify these small domains, underscoring the
298 advantages of integrative analysis across multiple samples.

299 Furthermore, to intuitively evaluate the integration and biological conservation capabilities of
300 PRESENT and other baseline methods, we systematically benchmarked them using 14 metrics, evenly
301 divided into two categories: batch effect removal and biological variance conservation, following a recent
302 benchmarking framework for single-cell omics data integration[68] (Methods). As shown in Fig. 4e,

303 scBasset-BC and SnapATAC2 provided satisfactory performance through incorporating sequence and
304 multi-omics information, respectively. PRESENT, which simultaneously integrated spatial dependency
305 and multi-omics information of all the samples, achieved the best performance in both batch effect
306 removal and biological variance conservation among all the 14 methods, illustrating its superiority in
307 multi-sample integration for spatial multi-omics data.
308

309 **PRESENT facilitates effective cross-modal representation in spatial single-omics data**

310 PRESENT is highly extensive and can be seamlessly adapted from spatial multi-omics data to spatial
311 single-omics data, such as spatial epigenomics and transcriptomics data, by simplifying the multi-view
312 autoencoder into a single-view model (Methods). We first accessed the performance of PRESENT on
313 spatial epigenomics data by collecting the spatial ATAC mouse embryo datasets which consist of three
314 developmental stages[9], including E15.5, E13.5 and E12.5. The datasets include six samples, with each
315 developmental stage containing two replicated samples, denoted as *-S1 and *-S2. Initially, we evaluated
316 the data quality of the six samples using logarithmic total fragment counts, denoted as log10(nFrags), and
317 transcription starting site (TSS) enrichment scores. The former metric represents the cellular sequencing
318 depth of samples[69], while the latter reflects the signal-to-noise ratio of datasets[70]. As shown in Fig. 5a, b,
319 the samples of E15.5 stage, i.e., E15.5-S1 and E15.5-S2, achieved the highest sequencing depth and
320 signal-to-noise ratio, while the E13.5-S1 sample and E12.5-S2 sample suffered from the lowest
321 sequencing depth and signal-to-noise ratio, respectively.

322 For spatial domain identification, we benchmarked PRESENT and other six baseline methods,
323 including three methods for single-cell ATAC-seq data (SCALE[71], cisTopic[72] and SnapATAC2[64]) and
324 three methods for SRT data (SpaGCN[31], STAGATE[32] and PAST[19]). As shown in Fig. 5c, PRESENT
325 achieved the best performance, with improvements of more than 8.60%, 26.13%, 25.91%, and 22.94%

326 compared to the second-best baseline method across the clustering metrics, respectively. Using the E15.5-
327 S1 sample as an example, we further demonstrated the UMAP and spatial visualization of spots colored
328 by ground truth domains (Fig. 5d, e) and the spatial clusters identified based on different methods (Fig.
329 5f). Baseline methods like cisTopic and STAGATE could only identify the forebrain domain. Although
330 SCALE, SnapATAC2, SpaGCN and PAST roughly recognized the subdomains of central nervous system,
331 i.e., forebrain, hindbrain and midbrain, the low-dimensional representations acquired by them are
332 decentralized and scattered, impeding their performance for spatial clustering (Fig. 5c, f). In contrast,
333 PRESENT effectively distinguished domains including the forebrain, hindbrain, midbrain, facial & bone,
334 muscle & heart, and meningeal & peripheral nervous system (PNS) 2. Visualization of the remaining five
335 samples obtained similar results (Supplementary Figs. 7-13), demonstrating the capabilities of PRESENT
336 in cross-modal representation of spatial epigenomics data.

337 Despite the superiority of PRESENT over existing baseline methods, low sequencing depth and
338 signal-to-noise ratio stills constrained the effectiveness of the model, as illustrated in Fig. 5a, b and
339 Supplementary Fig. 13. A previous study has shown that integrating external reference data through BNN
340 module can effectively address issues related to high noise levels in SRT data[19]. Inspired by this, we
341 collected a single-nucleus ATAC-seq mouse embryo dataset involving cells from three developmental
342 stages, including E12, E13 and E15, which was sequenced in conjunction with the spatial ATAC mouse
343 embryo data[9]. We utilized the single-nucleus ATAC-seq dataset (denoted as PRESENT-SN) and the other
344 five spatial ATAC samples (denoted as PRESENT-SP) as the reference data for PRESENT when
345 analyzing each spatial ATAC mouse embryo sample (Methods). As shown in Fig. 5g and Supplementary
346 Fig. 14, the introduction of prior information from reference data facilitated enhanced performance of
347 PRESENT, especially in samples with lower sequencing depth and signal-to-noise ratio, such as the
348 E13.5-S1 and E12.5-S2 mouse embryo samples. This finding was further validated by the UMAP

visualization of the two low-quality samples (Fig. 5h and Supplementary Figs. 9 and 12). Specifically, PRESENT-SN and PRESENT-SP successfully identified spatial domains including liver, limb, forebrain, hindbrain and midbrain, which other baseline methods and PRESENT failed to distinguish. For further validation, we increased the noise level of datasets by converting lower fragment counts to zero with growing dropout rates for each sample, according to a study that models the dropout events in single-cell data[73]. As shown in Fig. 5i, PRESENT-SN and PRESENT-SP gained greater advantages as the noise level increased, highlighting the improved benefits of integrating prior information from reference data with the growing dropout rates in spatial ATAC-seq datasets.

We also carried out extensive experiments on SRT axolotl brain datasets generated by Stereo-seq, which include six samples from different developmental stages, including stage 44, stage 54, stage 57, juvenile, adult and metamorphosis[6]. The performance of PRESENT was benchmarked with several recently developed SRT methods, involving SpaGCN[31], STAGATE[32], PAST[19] and SpatialPCA[74]. As shown in Supplementary Figs. 15-21, experimental results demonstrated the consistent superiority of PRESENT in cross-modal representation, spatial visualization and domain identification of SRT data. In addition, utilizing an SRT regenerated axolotl brain data at 20 days post-injury as reference data[6], PRESENT-SP again gained promoted performance through leveraging prior information contained in reference data (Supplementary Fig. 15). In summary, PRESENT not only promotes effective cross-modal representation in both spatial ATAC-seq and SRT data, but also empowers the integration of reference data to address the low sequencing depth and signal-to-noise ratio of spatial datasets.

PRESENT enables vertical and horizontal integration of spatial single-omics samples

Integrated analysis of multiple samples across different developmental stages (vertically serial tissue samples) and dissected areas (horizontally adjacent tissue samples), commonly referred to as vertical and

372 horizontal integration, provides a comprehensive approach for deciphering the hierarchical structures of
373 cellular organizations[22]. We first evaluated the performance of PRESENT in the integrative analysis of
374 vertically serial tissue samples, using the mouse embryo samples from three developmental stages (E15.5,
375 E13.5 and E12.5) generated by spatial ATAC technique[9]. Due to the lack of methods specially designed
376 for the multi-sample integration of spatial ATAC data, we compared PRESENT with six methods
377 developed for single-cell data (MNN[65], Scanorama[25], Harmony[24], SCALEX[66], scBasset-BC[67] and
378 SnapATAC2[64]) and two recently proposed SRT methods (STAligner[27] and GraphST[22]). We employed the
379 benchmarking framework consisting of 14 metrics to simultaneously accessed the batch effect removal
380 and biological variance conservation capabilities of these integration methods (Methods). As shown in
381 Fig. 6a, PRESENT achieved the best overall score for multi-sample vertical integration, ranking first in
382 biological variance conservation and second in batch effects removal among all the integration methods.
383 The UMAP visualization of spot embeddings across different developmental stages also demonstrated
384 that, compared with other methods, PRESENT gained better performance in aggregating spots from the
385 same domain but in different batches (Supplementary Fig. 22). We further performed spatial clustering
386 based on the joint embeddings produced by different methods. As shown in Fig. 6b, c, Scanorama and
387 STAligner could hardly distinguish different spatial domains across different developmental stages.
388 MNN, Harmony, SCALEX and SnapATAC2 provided improved results by roughly recognizing forebrain
389 and liver in E12.5 and E13.5 stages, but they still failed in other domains. In comparison, PRESENT not
390 only identified forebrain (cluster 1) and liver (cluster 5) in all the three stages, but also gained satisfactory
391 results in domains including hindbrain (cluster 2) and midbrain (cluster 8).

392 Furthermore, we collected two horizontal adjacent SRT mouse brain samples generated by the 10x
393 Genomics Visium platform, which involves a sagittal anterior dissected section and a sagittal posterior
394 dissected section. We applied two SRT integration methods, namely STAligner[27] and GraphST[22], and

395 PRESENT to perform integrated analysis of these horizontally adjacent brain samples and utilized the
396 anatomic annotation of the sagittal region in postnatal day 56 (P56) mouse brain provided by Allen
397 Reference Atlas[75] to intuitively compare the performance of these methods (Fig. 6d). As shown in Fig. 6e-
398 g, compared with STAligner and GraphST, analytical results of PRESENT achieved improvements in
399 shared and sample-specific domain identification. For example, PRESENT captured the classical
400 hierarchical structures, i.e., from outer layer 1 to inner layer 6, of dorsal pallium in both sagittal anterior
401 and posterior sections, as illustrated in the red rectangle of Fig. 6d. In addition, PRESENT successfully
402 distinguished the typical cortex Ammonis (CA, cluster 18) and dentate gyrus (DG, cluster 25) structures
403 simultaneously in the hippocampus (black frames in Fig. 6d), which have been proved to be associated
404 with mood, anxiety and personality disorders caused by various factors[76]. We further conducted
405 differentially expressed analysis for domain CA and domain DG, acquiring marker genes including *C1ql2*
406 and *Hpca*, respectively (Supplementary Figs. 23 and 24). According to previous studies, *C1ql2* serves as a
407 marker gene for dentate gyrus granule cells (DGCs) enriched in the DG structure[77], while *Hpca* is the
408 encoding gene for a high-affinity calcium-binding protein restricted to the central nervous system,
409 abundant in pyramidal cells of the hippocampus CA region[78].

410 Moreover, PRESENT also exhibited great potential in capturing sample-specific functional domains
411 and uncovering biological implications in identified spatial domains. PRESENT successfully revealed the
412 sophisticated substructures of the cerebellar vermis, which only exists in the sagittal posterior section of
413 mouse brain and plays important roles in coordinating movements of the central body such as trunk, head,
414 and proximal limbs[79]. Specifically, PRESENT successfully detected the molecular layer (cluster 3),
415 Purkinje cell layer (cluster 22) and internal granular layer (cluster 17) of cerebellar vermis cortex and the
416 white matter (cluster 20) of cerebellar vermis. The differentially expressed genes in the corresponding
417 domains have also been validated in previous studies, such as *S100b*[80], *Car8*[81], *Atp1a1*[82] and *Mbp*[83]

418 (Supplementary Figs. 25-28). For example, *Car8* and *Atp1a1*, the mutation of which cause the motor
419 coordination defect and dominant Charcot-Marie-Tooth type 2, were proved to be highly expressed in the
420 granular layers and Purkinje cells in recent studies[81, 82], respectively. In summary, these experimental
421 results demonstrate the superior performance of PRESENT in the integrative analysis of both vertically
422 serial and horizontally adjacent spatial samples generated by various spatial omics technologies, thereby
423 enabling the discovery of biological insights within identified spatial domains across multiple samples.
424

425 **Discussion**

426 Recent breakthroughs in spatial omics sequencing technologies have enabled the simultaneous acquisition
427 of spatial coordinate information and the regulatory status of the biomolecules within cells, such as
428 chromatin regions, genes, and proteins, accumulating massive spatial omics data in repositories and
429 databases[13, 14]. In this article, we proposed PRESENT, an effective and scalable contrastive learning
430 framework, to address the cross-modality representation and multi-sample integration of spatial omics
431 data, including spatial multi-omics, epigenomics and transcriptomics data. To capture the patterns of
432 various omics layers, we designed omics-specific encoders and decoders for different omics data, where
433 the transcriptomic and epigenomic decoders model the fragment counts via the zero-inflation negative
434 binomial (ZINB) distribution[39] and zero-inflation Poisson (ZIP) distribution[40], respectively (Methods). In
435 addition, a cross-omics alignment module built on a contrastive learning inter-omics alignment loss
436 function[41] and graph attention networks[37] (GAT) were also introduced to capture the spatial dependency
437 and complementary multi-omics information (Methods). We further extended the framework to the multi-
438 sample integration of spatial omics data, addressing the challenges including the inconsistency of spatial
439 coordinates between samples as well as the intricate coupling between batch effects and biological
440 signals. Specifically, we adopted a contrastive learning-based inter-batch alignment strategy[42] and a

441 batch-adversarial learning strategy[43] to eliminate batch effects as well as an intra-batch preserving loss to
442 preserve biological signals simultaneously. To validate of the model design, we have conducted a
443 comprehensive ablation study, as shown in Supplementary Figs. 29 and 30.

444 Through extensive experiments on different species involving human, mouse and axolotl, we
445 highlighted the superior performance of PRESENT in cross-modality representation and multi-sample
446 integration of spatial omics data generated by various spatial sequencing technologies, encompassing
447 spatial multi-omics techniques (spatial RNA-Protein[11, 12] and spatial RNA-ATAC[12]) and spatial single-
448 omics technologies (spatial ATAC[9] and spatial transcriptomics[6]). Specifically, PRESENT effectively
449 integrated the spatial dependency and complex multi-omics information contained in spatial multi-omics
450 or spatial single-omics techniques (Figs. 2, 3 and 5). Additionally, PRESENT also showed great potential
451 for leveraging various reference data based on the BNN module to address issues related to low
452 sequencing depth and signal-to-noise ratio in spatial omics data (Fig. 5). More importantly, PRESENT
453 can be further extended to the integrative analysis of multiple spatially resolved samples across distinct
454 dissection areas and developmental stages, thereby promoting the detection of hierarchical structures and
455 biological functions within cellular organizations from spatiotemporal views (Figs. 4 and 6).

456 Nevertheless, several avenues remain for further improvement of PRESENT. First, Due to technical
457 and cost limitations, the number of unpaired spatial samples exceeds that of paired spatial samples[84],
458 emphasizing the importance of unpaired integration. While PRESENT can anchor shared spots to analyze
459 paired multi-omics sequencing data or leverage shared genomic features to integrate multiple spatial
460 samples, it still struggles with the integration of unpaired samples with distinct feature space[84], such as
461 the integration of two samples generated by spatial ATAC-seq and SRT techniques, respectively. Second,
462 explorative analyses based on single-cell data and spatial data are limited by lack of spatial information
463 and low spatial resolution, respectively. As an integrative framework for spatially resolved omics data,

PRESENT could be further extended to integrate single-cell and spatial data, computationally addressing these technological limitations. Third, beyond integrating samples into the latent space, the alignment of spatial coordinates in physical space also plays an important role in joint analysis of multiple spatial samples[23, 26]. Finally, with the development of spatially resolved sequencing technology, imaged-based spatial techniques, such as 10x Genomics Xenium, produce high-resolution histology image of dissected samples, providing additional morphological features for enhanced analysis of spatial samples.

Methods

This study aims to investigate spatially functional domains as well as biological regulatory processes within microenvironments through incorporating cross-modality information, including spatial coordinate data and omics data, which involves 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). To achieve this goal, we have established a self-supervised contrastive learning framework, namely PRESENT, for the cross-modality representation and multi-sample integration of spatial multi-omics, epigenomics and transcriptomics data. Furthermore, we have also systematically evaluated the performance of PRESENT on massive real-world spatial datasets across various sequencing technologies, species and tissues.

Cross-modality representation of spatial multi-omics data

First, we detail the modeling approach of PRESENT for cross-modality representation of spatial multi-omics data, using a complex scenario as an example, which involves the comprehensive integration of spatial coordinate data, epigenomic data (ATAC data), transcriptomic data (RNA data), and proteomic data (ADT data). As shown in Fig. 1a, PRESENT is based on a multi-view autoencoder framework and comprises omics-specific encoders, a cross-omics alignment module, and omics-specific decoders, aiming to incorporate the cross-modality information into a low-dimensional feature matrix.

Multi-modality input of PRESENT. The input of PRESENT is a spatial coordinate matrix $\boldsymbol{M} \in R^{n\times 2}$ and three 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 genomic regions, g the number of highly variable genes and p the number of input proteins. Before feeding the spatial and omics data into the omics-specific encoders, we transform the spatial coordinate matrix $\boldsymbol{M}$ into an

undirected neighborhood graph $\boldsymbol{G}$ through adding undirected edges 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 ATAC data[85-88], library size normalization followed by log-transformation to RNA data[19, 31, 32] and the centered log ratio (CLR) transformation to ADT data [17, 89, 90], 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[37] (GAT), have showcased successful application in integrating spatial and transcriptomic information for spatial transcriptomics[19, 22, 31, 32]. On the other hand, research has proved that introducing weight uncertainty into models through Bayesian neural networks[38] (BNN) improves generalization in nonlinear regression problems[38] and offers an extension for leveraging prior information from existing sequencing data[19]. 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[37], depicted as:

$$\boldsymbol{h_i^2} = \sigma(\sum_{j \in \mathcal{N}_i} \alpha_{ij} \boldsymbol{W_2} \boldsymbol{h_j^1})$$

$$\alpha_{ij} = \frac{\sigma_{LeakyReLU}\left(\boldsymbol{a}^T\left(\boldsymbol{W_2}\boldsymbol{h_i^1} \oplus \boldsymbol{W_2}\boldsymbol{h_j^1}\right)\right)}{\sum_{k \in \mathcal{N}_i} \sigma_{LeakyReLU}\left(\boldsymbol{a}^T\left(\boldsymbol{W_2}\boldsymbol{h_i^1} \oplus \boldsymbol{W_2}\boldsymbol{h_k^1}\right)\right)}$$

where α_{ij} is the attention score between spot i and j, d_2 denotes the dimension of the second hidden layer, $\boldsymbol{W_2} \in R^{d_1 \times d_2}$ 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 a previous work, BNN module provides an opportunity to leverage biological prior information from existing sequencing data to improve the analytical performance in spatial transcriptomic data[19]. 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 the original settings[19], 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[91].

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

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

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[15-18]. Inspired by a recent study that achieved multi-modal entity alignment through contrastive learning[41], 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

562 function of the MLP module, d is the latent dimension of PRESENT, and d_{ATAC}, d_{RNA} and d_{ADT} refers

563 to the output dimensions of omics-specific encoders for ATAC, RNA and ADT data, respectively. The

564 MLP module is composed of three FCNs, the first two of which are followed by a layer normalization and

565 an ELU activation function.

566 However, only the MLP module cannot perfectly incorporate omics-specific knowledge from

567 different omics layers into the final cross-omics embeddings[41], which is also illustrated in our ablation

568 study shown in Supplementary Fig. 29. Therefore, we design an inter-omics alignment (IOA) loss to

569 minimize the multi-omics directional KLD over the output distributions between the joint embeddings

570 and other omics-specific embeddings:

571 $$\mathcal{L}_{IOA} = \frac{1}{M}\Big[\frac{1}{n}\sum_{i=1}^{n} KLD\left(q_{ed}^{(i)}(\boldsymbol{e}), q_{ed}^{(i)}(\boldsymbol{h}^{ATAC})\right) + \frac{1}{n}\sum_{i=1}^{n} KLD\left(q_{ed}^{(i)}(\boldsymbol{e}), q_{ed}^{(i)}(\boldsymbol{h}^{RNA})\right)$$

572 $$+\frac{1}{n}\sum_{i=1}^{n} KLD\left(q_{ed}^{(i)}(\boldsymbol{e}), q_{ed}^{(i)}(\boldsymbol{h}^{ADT})\right)\Big]$$

573 where M represents the number of omics layers, $\boldsymbol{e}, \boldsymbol{h}^{ATAC}, \boldsymbol{h}^{RNA}$ and $\boldsymbol{h}^{ADT}$ denotes the joint, ATAC,

574 RNA and ADT embeddings respectively, and $q_{ed}^{(i)}(*)$ represents the embedding distribution of spot i

575 over other spots, defined as (taking $\boldsymbol{e}$ as an example):

576 $$q_{ed}^{(i)}(\boldsymbol{e}) = \left(\frac{\delta_{ed}(\boldsymbol{e_i}, \boldsymbol{e_1})}{\sum_{j=1}^{n}\delta_{ed}(\boldsymbol{e_i}, \boldsymbol{e_j})}, \frac{\delta_{ed}(\boldsymbol{e_i}, \boldsymbol{e_2})}{\sum_{j=1}^{n}\delta_{ed}(\boldsymbol{e_i}, \boldsymbol{e_j})}, \dots, \frac{\delta_{ed}(\boldsymbol{e_i}, \boldsymbol{e_n})}{\sum_{j=1}^{n}\delta_{ed}(\boldsymbol{e_i}, \boldsymbol{e_j})}\right)$$

577 $$\delta_{ed}(\boldsymbol{u}, \boldsymbol{v}) = exp\left(\frac{\boldsymbol{u}^T\boldsymbol{v}}{\tau}\right)$$

578 where τ is a temperature parameter and set to 0.1 according to the original study[41].

579 **Omics-specific decoder.** The patterns and distributions of sequencing data vary significantly across

580 different omics layers, revealing complementary biologically regulatory information[15-18]. For example,

581 ADT sequencing data, which focuses on dozens of surface proteins, provides intuitive and reliable

582 quantification of cellular activities with low dimensionality, few dropout events and low sparsity[17]. In

583 contrast, RNA and ATAC data, which measure the expression levels of genes and accessibility scores of

584 genomic regions respectively, suffer from high dropout rates, dimensionality, and sparsity[92]. Therefore,

we designed omics-specific decoder for each modality to address the discrepancies between omics layers.

Inspired by a recently published study which modeled the fragment counts of single-cell ATAC-seq data with Poisson distribution[40], 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[9, 12]. 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 $f_{MLP}^{ATAC}(*)$, 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[39], 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:

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

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

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

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

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

Loss function for cross-modality representation of spatial multi-omics data. Based on the above descriptions, the overall loss function for cross-modality representation of spatial multi-omics data 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, denoted as:

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

where α_1, α_2, α_3 and α_4 are coefficients of the corresponding loss term. The default values of α_1, α_2, α_3 and α_4 are all set to 1.

Multi-sample integration of spatial multi-omics data

Multi-sample integration offers a comprehensive perspective for deciphering the structures of cellular organizations[22, 23, 26, 27]. For example, integrated analysis of samples under different stages can reveal developmental dynamics[27, 28], while integrating disease and normal samples facilitates the identification of disease-specific domains and promotes the investigation of disease mechanisms[27, 29]. Therefore, we have extended the PRESENT framework to multi-sample integration of spatial multi-omics data and further incorporated mechanisms to address the discrepancies of spatial coordinates as well as batch effects between samples. Overall, the multi-sample integration workflow based on the PRESENT framework involves two stages. The first stage serves as a warmup to pretrain the PRESENT model, while the second stage leverages a cyclic iterative optimization progress to eliminate batch effects while preserving biological variations across samples. Analogous to cross-modality representation of spatial multi-omics data, we use a scenario of integrating two spatial multi-omics sequencing samples as an

652 example to illustrate the workflow of multi-sample integration with PRESENT, where ATAC, RNA, ADT,
653 and spatial coordinate data are simultaneously provided for each sample (Fig. 1b).

654 **Feature sets unification.** The first step in multi-sample integration is to align the feature sets of different
655 sequencing samples, ensuring that all samples share a common feature space. For ATAC data, following
656 the merging objects vignette provided by the Signac package[91], a unified set of peak regions is first
657 obtained by merging all intersecting peaks from all the samples. The fragment counts of each spot within
658 the unified peak set are then quantified. For RNA and ADT data, whose feature sets are predefined genes
659 or proteins, we directly leverage the intersection of gene or protein sets of all the samples as the unified
660 gene or protein set. After the gene intersection step, highly variable genes will be selected based on the
661 unified gene set. For convenience, we denoted the ATAC, RNA, ADT and spatial data of the two samples
662 as $\boldsymbol{X}_1^{\boldsymbol{ATAC}} \in R^{n_1 \times r}$, $\boldsymbol{X}_2^{\boldsymbol{ATAC}} \in R^{n_2 \times r}$, $\boldsymbol{X}_1^{\boldsymbol{RNA}} \in R^{n_1 \times g}$, $\boldsymbol{X}_2^{\boldsymbol{RNA}} \in R^{n_2 \times g}$, $\boldsymbol{X}_1^{\boldsymbol{ADT}} \in R^{n_1 \times p}$, $\boldsymbol{X}_2^{\boldsymbol{ADT}} \in R^{n_2 \times p}$,
663 $\boldsymbol{M_1} \in R^{n_1 \times 2}$ and $\boldsymbol{M_2} \in R^{n_2 \times 2}$, respectively. Here, r represents the number of unified peak regions, g
664 the number of highly variable genes, p the number of unified proteins, and n_1 and n_2 the number of
665 spots of the two samples. The unified ATAC, RNA, ADT data are then concatenated and preprocessed
666 using the methods described in the cross-modality representation section, obtaining normalized data, i.e.,
667 $\widetilde{\boldsymbol{X}}^{\boldsymbol{ATAC}} \in R^{n \times r}$, $\widetilde{\boldsymbol{X}}^{\boldsymbol{RNA}} \in R^{n \times g}$, $\widetilde{\boldsymbol{X}}^{\boldsymbol{ADT}} \in R^{n \times p}$ and $\boldsymbol{M} \in R^{n \times 2}$, where $n = n_1 + n_2$ represents the number
668 of spots from the two samples.

669 **Cross-sample graph construction.** Due to inconsistent dissection and placement of sequencing samples,
670 the spatial coordinates between these samples cannot be directly compared with each other, resulting in
671 the failure of cross-modality representation methods in capturing relationships across samples[23, 26]. To
672 address this gap, we proposed a strategy to construct a cross-sample graph. Firstly, we find intra-sample
673 spatial neighbors for each spot based on the spatial coordinate matrix, building undirected edges between
674 all spots and their k nearest neighbors in Euclidean space from the same sample. Secondly, we find inter-

sample neighbors based on omics features, including ATAC, RNA and ADT profiles, to establish 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 PRESENT framework based on the cross-sample graph and the concatenated omics data of different samples, including $\widetilde{\boldsymbol{X}}^{\boldsymbol{ATAC}} \in R^{n\times r}$, $\widetilde{\boldsymbol{X}}^{\boldsymbol{RNA}} \in R^{n\times g}$ and $\widetilde{\boldsymbol{X}}^{\boldsymbol{ADT}} \in R^{n\times p}$. To integrate batch information into the PRESENT model, we concatenate the one-hot encoded batch labels with the output of the cross-omics alignment module, using the combined matrix as the input of the omics-specific decoders. The objective function and training process of PRESENT model during the first pretraining stage are identical to those used for cross-modality representation of individual samples, as detailed in the previous section.

Second-stage integration. The technical noise between different samples, i.e., batch effects, have not been fully addressed in the first stage. Therefore, we incorporate a contrastive learning strategy and a batch-adversarial learning strategy in the second stage to eliminate the batch effects across samples while preserving the biological signals simultaneously, as detailed in the following subsections.

Inter-batch alignment. We designed an inter-batch alignment (IBA) loss based on contrastive learning[42], aiming to align spots with similar omics features across different samples. Specifically, we first find mutual nearest neighbors from other samples as the positive set of for spot i, denoted as N_i^{pos}, referring to the inter-sample neighbors that are connected to spot i in the cross-sample graph. Second, we randomly select $1/4$ proportion of spots from the same sample as the negative set of spot i, depicted as N_i^{neg}. Finally, the IBA loss of spot i is:

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

where $\boldsymbol{e_i} \in \boldsymbol{R}^d$ denotes the latent representation of spot i, d is the output dimension of PRESENT and δ_{sim} denotes the cosine similarity. The inter-batch alignment loss of all the spots is:

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

Batch-adversarial learning. Inspired by a domain-adversarial learning research[93] and its's success in multi-batch integration of single-cell ATAC-seq data[43], we implement the batch-adversarial learning strategy to remove batch effect across spatial samples. 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 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 contrastive learning-based inter-batch alignment strategy and the batch-adversarial learning strategy aim for batch correction across samples. However, a key challenge of multi-sample integration lies in eliminating batch effects while simultaneously preserving the biological signals within each individual sample. We therefore proposed an intra-batch preserving (IBP) loss to maintain the biological pattern of latent representations within each sample during training process, 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 sample 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 within

sample k, defined as (taking the embeddings of current iteration $\boldsymbol{e}^{(t)}$ as an example):

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

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

where $s_k^{(j)}$ denotes the j-th spot in sample k, τ is a temperature parameter and set to 0.1 by default[41]. As shown in the above formula, the IBP loss is designed to preserve the overall manifold structure of spots within each sample during training process.

Training process of the second stage. The training process of the second stage is carried out in a cyclic iterative manner. First, we obtain the joint embeddings of multiple samples using the trained PRESENT model. Next, we construct a cross-sample graph according to the strategy described above, 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 model for one epoch using the newly constructed cross-sample graph. This cyclic iterative optimization process, aiming to iteratively refine the cross-sample graph, is repeated until the model converges. When updating the PRESENT model, we first 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 (i.e., the omics-specific encoders and the cross-omics alignment module) and omics-specific decoders fixed. Next, we fix the parameters of batch-discriminator and update the parameters of generator and omics-specific decoders based on an overall loss function, including the cross-modality representation loss, the inter-batch alignment loss, the intra-batch preserving loss and the negative cross-entropy loss for the batch-adversarial learning strategy:

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

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

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

Applications of PRESENT to spatial single-omics data

PRESENT is highly versatile and can be seamlessly adapted from spatial multi-omics data to spatial single-omics data, such as SRT and spatial epigenomics data, by simplifying the multi-view autoencoder framework into a single-view autoencoder. Specifically, for cross-modality representation or multi-sample integration of spatial single-omics data, only the respective omics-specific encoder and decoder are required. For example, only the ATAC omics-specific encoder and decoder are required when modeling spatial single-omics ATAC-seq data, while only the RNA omics-specific encoder and decoder are needed when modeling spatial transcriptomic data. Additionally, both the cross-omics alignment module and the associated inter-omics alignment loss are excluded from the PRESENT model in these cases, as spatial single-omics data consist of only a single omics layer.

Implementation details

With the development of spatial omics sequencing technologies, the number of spots in sequenced tissue has been quickly increasing from thousands to tens and even hundreds of thousands[9, 10, 13, 14, 94]. PRESENT is built upon graph neural networks, which face challenges such as memory explosion and neighborhood expansion when the number of nodes in graph grows large[95, 96]. To address these problems, we adopted necessary sparse adjacency matrix implementation based on the PyG framework when deploying our model. Additionally, we implemented the ClusterGCN algorithm to divide the entire graph into multiple small subgraphs[95], which enables our model to be trained on GPUs with limited memory and extend to larger spatial omics data.

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

Data collection and description

10x Visium RNA-Protein human lymph node data. The human lymph node dataset, generated by 10x Genomics Visium transcriptomics and proteomics co-profiling technology, was collected from a spatial multi-omics study[35]. This dataset was manually annotated by the clinician group according to the histological image[35], containing the spatial coordinates and the expression level of 18,085 genes and 31 proteins on 3,484 spots.

SPOTS mouse spleen data. The mouse spleen data[11] was generated by a high-throughput spatial transcriptomics and proteomics co-profiling technology, namely Spatial PrOtein and Transcriptome Sequencing (SPOTS). This dataset provides the spatial coordinates and the expression level of 32,285 genes and 21 proteins on 2,568 spots.

MISAR-seq mouse brain data. The MISAR-seq mouse brain data[12] provides the paired sequencing data of spatial gene expression and chromatin accessibility during mouse brain development. We collected

786 four manually annotated samples from original study[12], including E11.0 (1,258 spots), E13.5 (1,777
787 spots), E15.5 (1,949 spots) and E18.5 (2,129 spots). The manual annotations were produced with
788 reference to Kaufman's Atlas of Mouse Development and Allen Brain Atlas[12]. We called peak for each
789 sample based on the MACS2 v2.2.9.1 package, obtaining 242,024, 271,126, 265,014 and 294,734 peak
790 regions for downstream analyses, respectively.

791 **Single-nucleus ATAC-seq and spatial ATAC mouse embryo data.** The single-nucleus ATAC-seq
792 mouse embryo data[9] measured the chromatin accessibility of 1,879 cells from three developmental stages,
793 including E12, E13 and E15. The spatial ATAC mouse embryo data[9] is composed of six samples,
794 including E12.5-S1 (3,274 spots), E12.5-S2 (3,903 spots), E13.5-S1 (3,160 spots), E13.5-S2 (2,970
795 spots), E15.5-S1 (2,246 spots) and E15.5-S2 (2,234 spots). Associated domain annotations were also
796 provided within these samples, which have been validated through extensive experiments in the original
797 study[9]. The single-cell ATAC-seq datasets and spatial ATAC datasets were merged and 248,047 peak
798 regions were called based on MACS2 v2.2.6 package.

799 **Stereo-seq axolotl brain data.** We collected the Stereo-seq axolotl brain developmental datasets from a
800 recent study which explored the regeneration and development of axolotl telencephalon using spatial
801 transcriptomic atlas[6]. The datasets consist of six sections from developmental stage 44 (1,477 spots),
802 stage 54 (2,929 spots), stage 57 (4,410 spots), juvenile (11,698 spots), adult (8,243 spots), and
803 metamorphosis (7,441 spots), measuring the expression level of 12,704 genes. These samples were
804 manually annotated with reference to eHistology Kaufman Annotations and Allen Brain Atlas[12].

805 **10x Visium mouse brain data.** The spatial transcriptomic postnatal day 56 (P56) mouse brain data,
806 including a sagittal anterior section and a sagittal posterior section, were collected from the 10x Genomics
807 Visium platform. The mouse sagittal anterior section contains the expression level of 32,285 genes on
808 2,823 spots, while the posterior section provides the expression level of 32,285 genes on 3,289 spots.

809

810 **Experiment setup and baseline methods**

811 **Cross-modality representation of 10x Visium RNA-Protein human lymph node data.** We excluded
812 genes or proteins with zero expression across all spots. Baseline methods includes totalVI[46], CiteFuse[47],
813 Seurat-v4[48], SpatialGlue[35] and spaMultiVAE[36], all implemented following their respective tutorials.

814 **Cross-modality representation of MISAR-seq mouse brain data.** We filtered out peaks accessible in
815 less than 3% spots and genes with zero expression in all spots for quality control. The filtered count
816 matrices of ATAC data and RNA data along with the spatial coordinate matrix were then fed into the
817 workflow of PRESENT and other baseline methods, involving SpatialGlue[35], MOFA+[61], scAI[62],
818 scMDC[63], scMVP[21], Seurat-v4[48] and SnapATAC2[64]. Note that totalVI[46], CiteFuse[47] and spaMultiVAE[36]
819 were not used as they are specially designed for single-cell or spatial RNA-ADT data. All the baseline
820 methods were implemented following the respective tutorials and default settings. Additionally, following
821 the workflow of processing single-cell[97] and spatial[19, 32, 74] RNA-seq data, PRESENT selected 3,000
822 highly variable genes as the transcriptomic input for further analysis.

823 **Multi-sample integration of MISAR-seq mouse brain data.** We merged the peak sets of different
824 mouse brain samples according to the merging objects vignette provided by the Signac package[91], which
825 first obtained a unified peak set by merging all intersecting peaks from all the samples and then quantified
826 the fragment count of each spot on the newly obtained peak set. Consistent with the experimental settings
827 in cross-modality representation, peaks accessible in less than 3% spots in the concatenated matrix were
828 filtered out and 3,000 highly variable genes of transcriptomic data were selected for PRESENT. Due to
829 the lack of methods specially designed for the integration of spatial multi-omics data, we compared the
830 batch effect removal and biological variance conservation performance of PRESENT with six methods
831 developed for single-cell data (MNN[65], Scanorama[25], Harmony[24], SCALEX[66], scBasset-BC[67] and

832 SnapATAC2[64]) and three recently proposed SRT multi-sample integration methods (STAligner[27],
833 Stitch3D[33] and GraphST[22]). Specifically, MNN, Scanorama, Harmony, SCALEX could be applied to both
834 transcriptomic and epigenomic data, denoted as RNA-* or ATAC-*, respectively, where * denotes
835 respective methods. MNN, Scanorama and Harmony performed batch correction based on embeddings
836 extracted from RNA or ATAC data, which were accomplished based on the workflow of SCANPY[97] and
837 EpiScanpy[98], respectively. All the methods were implemented using default settings.

838 **Cross-modality representation of spatial ATAC mouse embryo data.** The spatial ATAC mouse embryo
839 data firstly went through a quality control process, where peaks accessible in less than 0.5% spots were
840 filtered out. Using the filtered ATAC count matrices as input, we then benchmarked the performance of
841 PRESENT with four methods designed for single-cell ATAC-seq data, including SCALE[71], cisTopic[72] and
842 SnapATAC2[64], and three methods designed for SRT data, i.e., SpaGCN[31], STAGATE[32] and PAST[19]. All
843 the baseline methods were implemented using their default settings, except for STAGATE and PAST. In
844 original tutorials, the two SRT methods adopted a feature selection process which is specially designed
845 for transcriptomic data and only selected 3,000 features, resulting significantly lower performance
846 (Supplementary Fig. 31). Since other methods utilized all the input peaks for dimension reduction, we
847 ignored the feature selection step of these two methods for fairness and better performance.

848 **Cross-modality representation of Stereo-seq axolotl brain data.** The genes with zero expression in all
849 spots were filtered out. The filtered count matrices and the spatial locations were fed into PRESENT and
850 other baseline methods, including SpaGCN[31], STAGATE[32], PAST[19] and SpatialPCA[74]. All baseline
851 methods were implemented with default settings in their tutorials. PRESENT again employed a gene
852 selection procedure, obtaining 3,000 highly variable genes as input.

853 **Multi-sample integration of spatial ATAC mouse embryo data.** The three spatial ATAC mouse embryo
854 samples from three developmental stages have the unified peak sets and were directly concatenated.

855 Following the experimental settings in cross-modality representation of a single spatial sample, peaks
856 accessible in less than 0.5% spots were filtered out. We benchmarked PRESENT with six integration
857 methods developed for single-cell data (MNN[65], Scanorama[25], Harmony[24], SCALEX[66], scBasset-BC[67]
858 and SnapATAC2[64]) and two recently proposed SRT methods (STAligner[27] and GraphST[22]). Here
859 Stitch3D[33] was not used as it encountered error. MNN, Scanorama and Harmony performed batch
860 correction based on the latent embeddings obtained by EpiScanpy[98]. All the methods were implemented
861 according to their default settings except for STAligner. Since other methods utilized all the input peaks,
862 we ignored the feature selection step of STAligner to ensure better performance and fairness.
863 **Multi-sample integration of 10x Visium mouse brain data.** The intersected gene set of the two mouse
864 brain samples, including a sagittal anterior section and a sagittal posterior section, were utilized as the
865 unified gene set. Next, gene count matrices of the two samples were concatenated and genes with zero
866 expression in all spots were filtered out. Similarly, 3,000 highly variable genes were selected as the input
867 for PRESENT. PRESENT was compared with two recently developed SRT methods, i.e., STAligner[27] and
868 GraphST[22], which were conducted according to their default settings. Specifically, we utilized PASTE[23] to
869 accomplish the initial coordinate alignment step for GraphST according to the original study[22].
870
871 **Evaluation metrics**
872 **Spatial clustering metrics.** We implemented four commonly used metrics[50, 51] to quantitatively evaluate
873 the performance of spatial clustering, including adjusted rand index (ARI), adjusted mutual information
874 (AMI), normalized mutual information (NMI) and homogeneity (Homo).
875 ARI, ranging from 0 to 1, denotes the consistent adjustment of the expectations of rand index:
876 $$\mathrm{ARI} = \frac{RI(P,T) - E[RI(P,T)]}{\max[RI(P,T)] - E[RI(P,T)]}$$
877 where $RI(\cdot)$ denotes the function of rand index, T refers to the ground truth labels, and P represents

878 the clustering labels. Rand index is the probability of the agreement between the clustering result and the

879 ground truth labels on a randomly selected spot set.

880 Mutual Information (MI) accesses the similarity between the clustering labels and ground truth

881 labels, from which both AMI and NMI originate. AMI ranges from -1 to 1, while NMI ranges from 0 to 1.

882 AMI is calculated as:

883 $$\mathrm{AMI} = \frac{MI(P,T) - E[MI(P,T)]}{avg[H(P),H(T)] - E[MI(P,T)]}$$

884 where $MI(\cdot)$ refers to the mutual information, $E(\cdot)$ represents the expectation function and $H(\cdot)$

885 denotes the function of information entropy. NMI can be calculated as:

886 $$\mathrm{NMI} = \frac{MI(P,T)}{\sqrt{H(P)H(T)}}$$

887 Homogeneity is another metric to measure the clustering performance. Specifically, if all the clusters

888 involve only data points that are members of a single class in the ground truth labels, the clustering results

889 could be interpreted as satisfying homogeneity, calculated as:

890 $$\mathrm{Homo} = 1 - \frac{H(T|P)}{H(T)}$$

891 **Integration metrics.** According to a recent benchmarking framework for single-cell omics data

892 integration, two important perspectives need to be considered when evaluating the performance of multi-

893 sample integration, namely the batch effect removal and the biological variance conservation[68].

894 To quantitatively evaluate the performance of batch effect removal, we introduced several metrics

895 used in the benchmarking framework[68], including graph connectivity(GC), k-nearest-neighbor batch-

896 effect test (kBET) and batch average silhouette width (bASW). In addition, we also implemented a cross-

897 validation strategy to evaluate the cross-batch annotation performance based on latent representations

898 obtained by different integration methods, which is also an important assessment of batch correction[99].

899 Specifically, using the spots from a specific sample as the test set and those from other samples along

900 with the domain labels as the training set, we transferred the domain labels of the training set into the test

901 set based on a k-NN classifier and the joint embeddings of all the samples. k-NN classifier was used as
902 its performance mainly depended on the quality of joint embeddings obtained by different methods. The
903 performance of label transfer was evaluated with four commonly used metrics implemented in the scikit-
904 learn package, i.e., accuracy (Acc), kappa, macro F1 score (mF1) and the weighted F1 score (wF1). After
905 all the samples were used as the test set in a cross-validation manner, we calculated the average score for
906 each metric to evaluate the performance of cross-batch annotation.

907 The quantitative evaluation for biological variance conservation involves the assessment of
908 representative quality and the clustering performance. The representative quality was accessed via the
909 mean average precision (MAP)[100] and two metrics included in the benchmarking framework[68], namely
910 the spot average silhouette width (sASW) and spot local inverse Simpson's index (sLISI) , while the
911 evaluation of clustering performance was based on the four metrics, i.e., ATI, AMI, NMI and Homo.

912 We calculated category scores of the batch effect removal and the biological variance conservation
913 through averaging the metrics mentioned in the two aspects, respectively. We also calculated an overall
914 score using a 40/60 weighted mean of the two category scores following previous studies[68, 100].

915

916 **Downstream analyses**

917 **Visualization.** Based on the latent representations obtained by various methods, we utilized the uniform
918 manifold approximation and projection (UMAP)[52] algorithm to further reduce the dimension of latent
919 representation matrices into two-dimensional matrices for visualization. In addition, the two-dimensional
920 UMAP and spatial coordinate matrices were visualized using the functions *scanpy.pl.umap()* and
921 *scanpy.pl.spatial()* wrapped in the SCANPY package[97].

922 **Spatial clustering.** We implemented a binary search strategy to tune the resolution of Leiden algorithm[49],
923 a graph-based clustering method, to match the number of spatial clusters as closely as possible with the

number of ground truth domains. Specifically, we set a minimum resolution res_{min} and a maximum resolution res_{max}, and obtain clustering results using a resolution equal to $\frac{res_{min}+res_{max}}{2}$. If the number of clusters exceeds the number of ground truth domains, res_{max} is updated to $\frac{res_{min}+res_{max}}{2}$. Otherwise, res_{min} is updated to $\frac{res_{min}+res_{max}}{2}$. The iterative process continues until the number of clusters matches the number of ground truth domains or the current iteration reaches the maximum number of iterations.

Differentially expressed gene and protein analysis. We implemented differential analysis based on the Wilcoxon rank-sum test[54] wrapped in *scanpy.tl.rank_genes_groups()* function of the SCANPY package[97].

Gene ontology analysis. Given a specific gene set, we conducted gene ontology (GO) analysis using the clusterProfiler package[101], which automated the process of biological-term classification and the enrichment analysis of gene clusters.

Data availability

The single-nucleus ATAC-seq and spatial ATAC mouse embryo datasets[9] are available at Gene Expression Omnibus (GEO) using the accession code GSE214991. The Stereo-seq axolotl brain datasets[6] as well as the reference dataset, i.e., axolotl brain data at 20 days post-injury, can be obtained through their database https://db.cngb.org/stomics/artista/. The 10x Visium mouse brain datasets, including a sagittal anterior section and a sagittal posterior section, are accessible at the 10x Genomics websites https://www.10xgenomics.com/datasets/mouse-brain-serial-section-2-sagittal-anterior-1-standard and https://www.10xgenomics.com/datasets/mouse-brain-serial-section-2-sagittal-posterior-1-standard, respectively. The SPOTS mouse spleen data[11] is accessible at GEO with accession code GSE198353. The 10x Visium RNA & protein human lymph node dataset[35] can be downloaded from the Zenodo database https://zenodo.org/records/10362607. The MISAR-seq mouse brain data[12] is available from the National Genomics Data Center via accession number OEP003285.

Code availability

The code as well as the detailed tutorials for implementing PRESENT can be found at the Github website https://github.com/lizhen18THU/PRESENT.

Acknowledgements

This work was partly supported by the National Key Research and Development Program of China (grant nos. 2022YFF1202400 and 2023YFF1204802), the National Natural Science Foundation of China (grant nos. 62273194), and Beijing Natural Science Foundation (L242026).

Author contributions

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

Competing interests

The authors declare no competing interests.

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. Adema, K. et al. Lost in space: what single-cell RNA sequencing cannot tell you. *Trends Plant Sci.* (2024).
3. Armingol, E. et al. Deciphering cell-cell interactions and communication from gene expression. *Nat. Rev. Genet.* **22**, 71-88 (2021).
4. Cang, Z.X. et al. Screening cell-cell communication in spatial transcriptomics via collective optimal transport. *Nat. Methods* **20**, 218-228 (2023).
5. Shao, X. et al. Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with SpaTalk. *Nat. Commun.* **13**, 4429 (2022).
6. Wei, X. et al. Single-cell Stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration. *Science* **377**, eabp9444 (2022).
7. Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nat. Methods* **19**, 534-546 (2022).
8. Deng, Y.X. et al. Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature* **609**, 375-383 (2022).
9. Llorens-Bobadilla, E. et al. Solid-phase capture and profiling of open chromatin by spatial ATAC. *Nat. Biotechnol.* **41**, 1085-1088 (2023).
10. 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).
11. Ben-Chetrit, N. et al. Integration of whole transcriptome spatial profiling with protein markers. *Nat. Biotechnol.* **41**, 788-193 (2023).
12. Jiang, F.Q. et al. Simultaneous profiling of spatial gene expression and chromatin accessibility during mouse brain development. *Nat. Methods* **20**, 1048-1057 (2023).
13. Yuan, Z.Y. et al. SODB facilitates comprehensive exploration of spatial omics data. *Nat. Methods* **20**, 387-399 (2023).
14. Xu, Z. et al. STOmicsDB: a comprehensive database for spatial transcriptomics data sharing, analysis and visualization. *Nucleic Acids Res.* **52**, D1053-D1061 (2024).
15. Berger, S.L. The complex language of chromatin regulation during transcription. *Nature* **447**, 407-412 (2007).
16. MacNeil, L.T. & Walhout, A.J.M. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Res.* **21**, 645-657 (2011).
17. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865-868 (2017).
18. Peterson, V.M. et al. Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* **35**, 936-939 (2017).
19. Li, Z. et al. Latent feature extraction with a prior-based self-attention framework for spatial transcriptomics. *Genome Res.* **33**, 1757-1773 (2023).
20. Shang, L. & Xiang, Z. Spatially aware dimension reduction for spatial transcriptomics. *Nat. Commun.* **13**, 7203 (2022).
21. 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).
22. Long, Y. et al. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nat. Commun.* **14**, 1155 (2023).
23. Zeira, R. et al. Alignment and integration of spatial transcriptomics data. *Nat. Methods* **19**, 567-575 (2022).
24. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289-1296 (2019).

1006 25. Hie, B., Bryson, B. & Berger, B. Efficient integration of heterogeneous single-cell transcriptomes using
1007 Scanorama. *Nat. Biotechnol.* **37**, 685-691 (2019).
1008 26. Guo, T.T. et al. SPIRAL: integrating and aligning spatially resolved transcriptomics data across different
1009 experiments, conditions, and technologies. *Genome Biol.* **24**, 241 (2023).
1010 27. Zhou, X., Dong, K.N. & Zhang, S.H. Integrating spatial transcriptomics data across different conditions,
1011 technologies and developmental stages. *Nat. Comput. Sci.* **3**, 894-906 (2023).
1012 28. Chen, A. et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned
1013 arrays. *Cell* **185**, 1777-+ (2022).
1014 29. Chen, S. et al. Spatially resolved transcriptomics reveals genes associated with the vulnerability of middle
1015 temporal gyrus in Alzheimer's disease. *Acta Neuropathol. Commun.* **10**, 188 (2022).
1016 30. Yuan, Z. et al. Benchmarking spatial clustering methods with spatially resolved transcriptomics data. *Nat.*
1017 *Methods* **21**, 1-11 (2024).
1018 31. Hu, J. et al. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains
1019 and spatially variable genes by graph convolutional network. *Nat. Methods* **18**, 1342-1351 (2021).
1020 32. Dong, K.N. & Zhang, S.H. Deciphering spatial domains from spatially resolved transcriptomics with an
1021 adaptive graph attention auto-encoder. *Nat. Commun.* **13**, 1739 (2022).
1022 33. Wang, G. et al. Construction of a 3D whole organism spatial atlas by joint modelling of multiple slices with
1023 deep neural networks. *Nat. Mach. Intell.* **5**, 1200-1213 (2023).
1024 34. Vandereyken, K. et al. Methods and applications for single-cell and spatial multi-omics. *Nat. Rev. Genet.* **24**,
1025 494-515 (2023).
1026 35. Long, Y. et al. Deciphering spatial domains from spatial multi-omics with SpatialGlue. *Nat. Methods* (2024).
1027 36. Tian, T. et al. Dependency-aware deep generative models for multitasking analysis of spatial omics data. *Nat.*
1028 *Methods* (2024).
1029 37. Veličković, P. et al. Graph attention networks. Preprint at *arXiv* https://arxiv.org/abs/1710.10903 (2017).
1030 38. Blundell, C. et al. Weight uncertainty in neural network. In *32nd International Conference on Machine*
1031 *Learning* (2015).
1032 39. Eraslan, G. et al. Single-cell RNA-seq denoising using a deep count autoencoder. *Nat. Commun.* **10**, 390
1033 (2019).
1034 40. Martens, L.D. et al. Modeling fragment counts improves single-cell ATAC-seq analysis. *Nat. Methods* **21**,
1035 28-31 (2023).
1036 41. Lin, Z. et al. Multi-modal Contrastive Representation Learning for Entity Alignment. In *29th International*
1037 *Conference on Computational Linguistics* (2022).
1038 42. Chen, T. et al. A simple framework for contrastive learning of visual representations. In *37th International*
1039 *Conference on Machine Learning* (2020).
1040 43. Kopp, W., Akalin, A. & Ohler, U. Simultaneous dimensionality reduction and integration for single-cell
1041 ATAC-seq data using deep learning. *Nat. Mach. Intell.* **4**, 162-168 (2022).
1042 44. Swanton, C. Intratumor heterogeneity: evolution through space and time. *Cancer Res.* **72**, 4875-4882 (2012).
1043 45. Janiszewska, M. The microcosmos of intratumor heterogeneity: the space-time of cancer evolution.
1044 *Oncogene* **39**, 2031-2039 (2020).
1045 46. Gayoso, A. et al. Joint probabilistic modeling of single-cell multi-omic data with totalVI. *Nat. Methods* **18**,
1046 272-282 (2021).
1047 47. Kim, H.J. et al. CiteFuse enables multi-modal analysis of CITE-seq data. *Bioinformatics* **36**, 4137-4143
1048 (2020).
1049 48. Hao, Y.H. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 (2021).
1050 49. Traag, V.A., Waltman, L. & Van Eck, N.J. From Louvain to Leiden: guaranteeing well-connected
1051 communities. *Sci. Rep.* **9**, 5233 (2019).

50. Chen, S. et al. VPAC: Variational projection for accurate clustering of single-cell transcriptomic data. *BMC Bioinformatics* **20**, 139-151 (2019).
51. Jiang, R. et al. SINFONIA: scalable identification of spatially variable genes for deciphering spatial domains. *Cells* **12**, 604 (2023).
52. McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. Preprint at *arXiv* https://arxiv.org/abs/1802.03426 (2018).
53. Lewis, S.M., Williams, A. & Eisenbarth, S.C. Structure and function of the immune system in the spleen. *Sci. Immunol.* **4**, eaau6085 (2019).
54. Soneson, C. & Robinson, M.D. Bias, robustness and scalability in single-cell differential expression analysis. *Nat. Methods* **15**, 255-261 (2018).
55. Schmid, M.C. & Varner, J.A. Myeloid cells in tumor inflammation. *Vasc. Cell* **4**, 1-7 (2012).
56. Borges, M.D. & Sesti-Costa, R. Macrophages: key players in erythrocyte turnover. *Hematol. Transfus. Cell Ther.* **44**, 574-581 (2022).
57. Seyfried, A.N., Maloney, J.M. & MacNamara, K.C. Macrophages orchestrate hematopoietic programs and regulate HSC function during inflammatory stress. *Front. Immunol.* **11**, 548463 (2020).
58. Cui, K. et al. Endocytic adaptors in cardiovascular disease. *Front. Cell Dev. Biol.* **8**, 624159 (2020).
59. Baranov, M.V. et al. Modulation of immune responses by particle size and shape. *Front. Immunol.* **11**, 607945 (2021).
60. Kurup, S.P., Butler, N.S. & Harty, J.T. T cell-mediated immunity to malaria. *Nat. Rev. Immunol.* **19**, 457-471 (2019).
61. Argelaguet, R. et al. MOFA+ : a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* **21**, 1-17 (2020).
62. Jin, S.Q., Zhang, L.H. & Nie, Q. scAI: an unsupervised approach for the integrative analysis of parallel single-cell transcriptomic and epigenomic profiles. *Genome Biol.* **21**, 1-19 (2020).
63. Lin, X. et al. Clustering of single-cell multi-omics data with a multimodal deep learning method. *Nat. Commun.* **13**, 7705 (2022).
64. Zhang, K. et al. A fast, scalable and versatile tool for analysis of single-cell omics data. *Nat. Methods* **21**, 217-227 (2024).
65. Haghverdi, L. et al. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421-427 (2018).
66. Xiong, L. et al. Online single-cell data integration through projecting heterogeneous datasets into a common cell-embedding space. *Nat. Commun.* **13**, 6118 (2022).
67. Yuan, H. & Kelley, D.R. scBasset: sequence-based modeling of single-cell ATAC-seq using convolutional neural networks. *Nat. Methods* **19**, 1088-1096 (2022).
68. Luecken, M.D. et al. Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* **19**, 41-50 (2022).
69. Heumos, L. et al. Best practices for single-cell analysis across modalities. *Nat. Rev. Genet.* **24**, 550-572 (2023).
70. Zhang, H. et al. Extensive evaluation of ATAC-seq protocols for native or formaldehyde-fixed nuclei. *BMC Genomics* **23**, 214 (2022).
71. Xiong, L. et al. SCALE method for single-cell ATAC-seq analysis via latent feature extraction. *Nat. Commun.* **10**, 4576 (2019).
72. González-Blas, C.B. et al. cisTopic: cis-regulatory topic modeling on single-cell ATAC-seq data. *Nat. Methods* **16**, 397-400 (2019).
73. Pierson, E. & Yau, C. ZIFA: Dimensionality reduction for zero-inflated single-cell gene expression analysis. *Genome Biol.* **16**, 1-10 (2015).

74. Shang, L. & Zhou, X. Spatially aware dimension reduction for spatial transcriptomics. *Nat. Commun.* **13**, 7203 (2022).
75. Sunkin, S.M. et al. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res.* **41**, D996-D1008 (2012).
76. Teicher, M.H., Anderson, C.M. & Polcari, A. Childhood maltreatment is associated with reduced volume in the hippocampal subfields CA3, dentate gyrus, and subiculum. *Proc. Natl. Acad. Sci.* **109**, E563-E572 (2012).
77. Matsuda, K. et al. Transsynaptic modulation of kainate receptor functions by C1q-like proteins. *Neuron* **90**, 752-767 (2016).
78. Park, S.-Y. et al. Hippocalcin promotes neuronal differentiation and inhibits astrocytic differentiation in neural stem cells. *Stem Cell Rep.* **8**, 95-111 (2017).
79. Coffman, K.A., Dum, R.P. & Strick, P.L. Cerebellar vermis is a target of projections from the motor areas in the cerebral cortex. *Proc. Natl. Acad. Sci.* **108**, 16068-16073 (2011).
80. Hachem, S. et al. Expression of S100B during embryonic development of the mouse cerebellum. *BMC Dev. Biol.* **7**, 1-15 (2007).
81. Hirasawa, M. et al. Carbonic anhydrase related protein 8 mutation results in aberrant synaptic morphology and excitatory synaptic function in the cerebellum. *Mol. Cell. Neurosci.* **35**, 161-170 (2007).
82. Lassuthova, P. et al. Mutations in ATP1A1 cause dominant Charcot-Marie-Tooth type 2. *Am. J. Hum. Genet.* **102**, 505-514 (2018).
83. Pacey, L.K. et al. Delayed myelination in a mouse model of fragile X syndrome. *Hum. Mol. Genet.* **22**, 3920-3930 (2013).
84. Argelaguet, R. et al. Computational principles and challenges in single-cell data integration. *Nat. Biotechnol.* **39**, 1202-1215 (2021).
85. Chen, S.Q. et al. RA3 is a reference-guided approach for epigenetic characterization of single cells. *Nat. Commun.* **12**, 13 (2021).
86. Chen, X. et al. Cell type annotation of single-cell chromatin accessibility data via supervised Bayesian embedding. *Nat. Mach. Intell.* **4**, 116-126 (2022).
87. Cusanovich, D.A. et al. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910-914 (2015).
88. Cusanovich, D.A. et al. A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell* **174**, 1309-1324 (2018).
89. Mulè, M.P., Martins, A.J. & Tsang, J.S. Normalizing and denoising protein expression data from droplet-based single cell profiling. *Nat. Commun.* **13**, 2099 (2022).
90. Stoeckius, M. et al. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 1-12 (2018).
91. Stuart, T. et al. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333-1341 (2021).
92. Qiu, P. Embracing the dropouts in single-cell RNA-seq analysis. *Nat. Commun.* **11**, 1169 (2020).
93. Ganin, Y. et al. Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* **17**, 1-35 (2016).
94. Zhang, D. et al. Spatial epigenome-transcriptome co-profiling of mammalian tissues. *Nature* **616**, 113-122 (2023).
95. Chiang, W.-L. et al. Cluster-GCN: An efficient algorithm for training deep and large graph convolutional networks. In *25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2019).
96. Bai, J., Ren, Y. & Zhang, J. Ripple walk training: A subgraph-based training framework for large and deep graph neural network. In *2021 International Joint Conference on Neural Networks (IJCNN)* (2021).
97. Wolf, F.A., Angerer, P. & Theis, F.J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 1-5 (2018).
98. Danese, A. et al. EpiScanpy: integrated single-cell epigenomic analysis. *Nat. Commun.* **12**, 5228 (2021).

99. Song, Y. et al. Benchmarking strategies for cross-species integration of single-cell RNA sequencing data. *Nat. Commun.* **14**, 6495 (2023).
100. Cao, Z.-J. & Gao, G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nat. Biotechnol.* **40**, 1458-1466 (2022).
101. Yu, G. et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS: J. Integrative Biol.* **16**, 284-287 (2012).

Figures

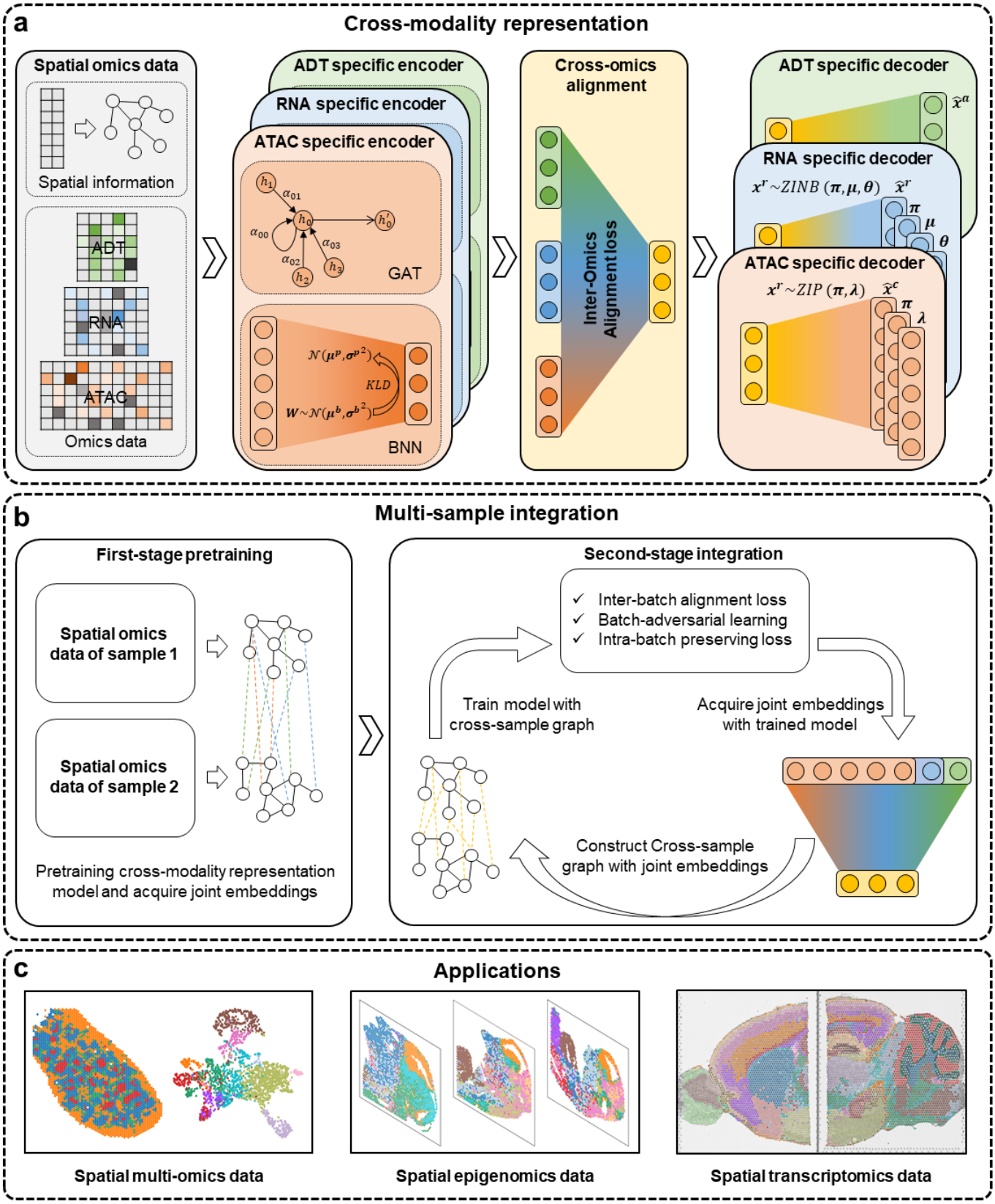

Fig. 1 | The framework of PRESENT. a, Cross-modality representation of spatial multi-omics data. PRESENT is capable to integrate spatial coordinate data and multi-omics data, including proteomic data sequenced by antibody-derived tags (ADT), transcriptomic data by RNA sequencing (RNA), and epigenomic data generated by the assay for transposase-accessible chromatin with sequencing (ATAC). 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 a Bayesian neural network (BNN) module. All the omics-specific feature matrices extracted by encoders are fused by a cross-omics alignment module, which integrates multi-omics information via the 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 of spatial omics data. PRESENT adopts a two-stage workflow for the integration of multiple spatial samples. In the first pretraining stage, PRESENT is pretrained using omics data and the cross-sample graph obtained from all the samples. 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 features across samples. In the second integration stage, PRESENT implements a cyclic iterative strategy to refine the cross-sample graph using the joint embeddings obtained by the trained model. PRESENT further introduces a contrastive learning-based 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 spatial omics data. PRESENT can be easily adapted from spatial multi-omics to spatial single-omics data by simplifying the multi-view autoencoder into a single-view autoencoder. The joint embeddings obtained by PRESENT can be applied to various downstream analyses, including spatial visualization, domain identification and the integrative analysis of samples across conditions and developmental stages.

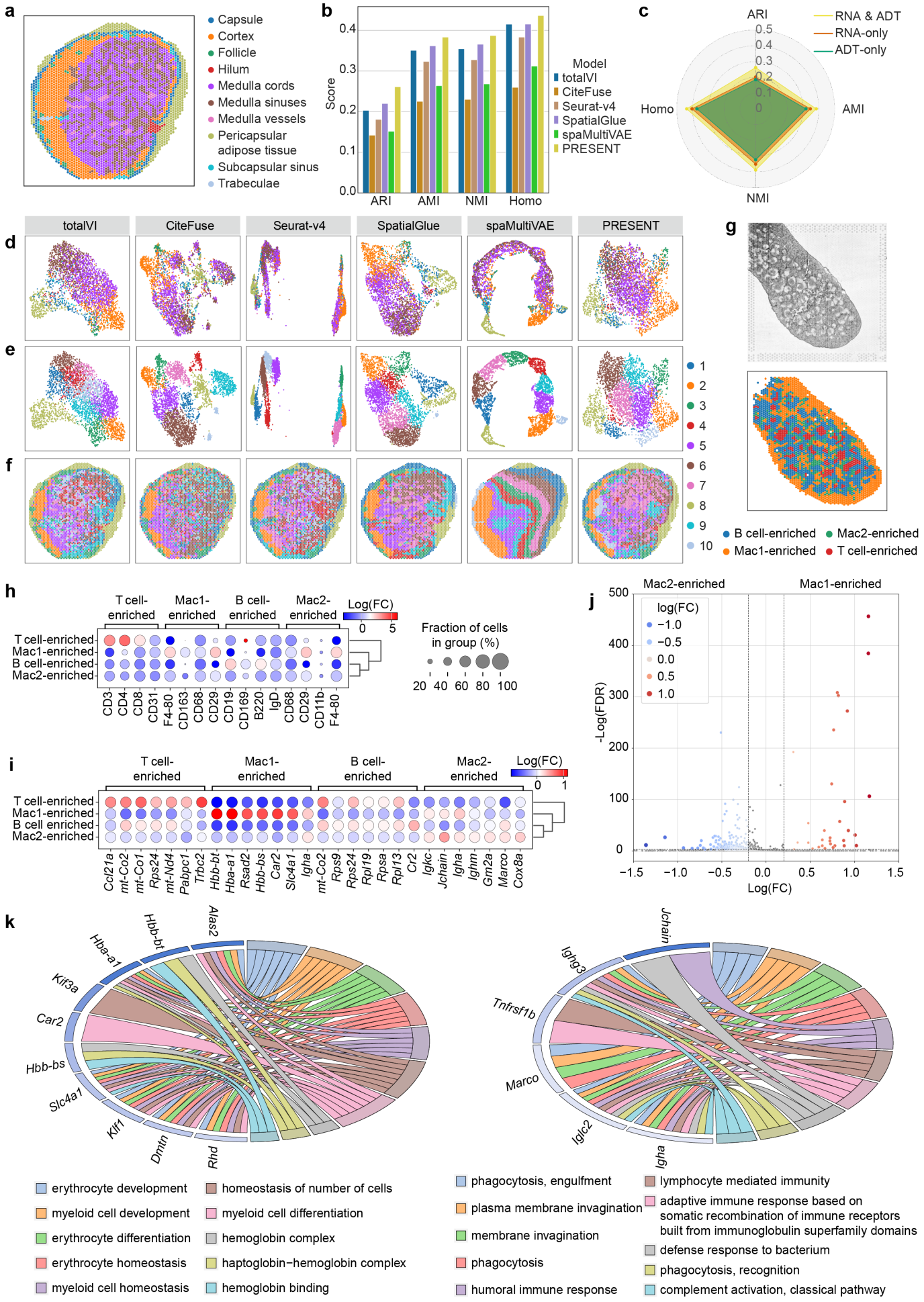
a
Capsule
Cortex
Follicle
Hilum
Medulla cords
Medulla sinuses
Medulla vessels
Pericapsular adipose tissue
Subcapsular sinus
Trabeculae
b
Score
ARI
AMI
NMI
Homo
Model
totalVI
CiteFuse
Seurat-v4
SpatialGlue
spaMultiVAE
PRESENT
c
RNA & ADT
RNA-only
ADT-only
d
e
f
g
B cell-enriched
Mac2-enriched
Mac1-enriched
T cell-enriched
h
Log(FC)
Fraction of cells in group (%)
i
j
Mac2-enriched
Mac1-enriched
log(FC)
-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. 2 | PRESENT facilitates accurate spatial domain identification in 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 embeddings obtained by different methods colored by ground truth domain labels and **e,** 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.

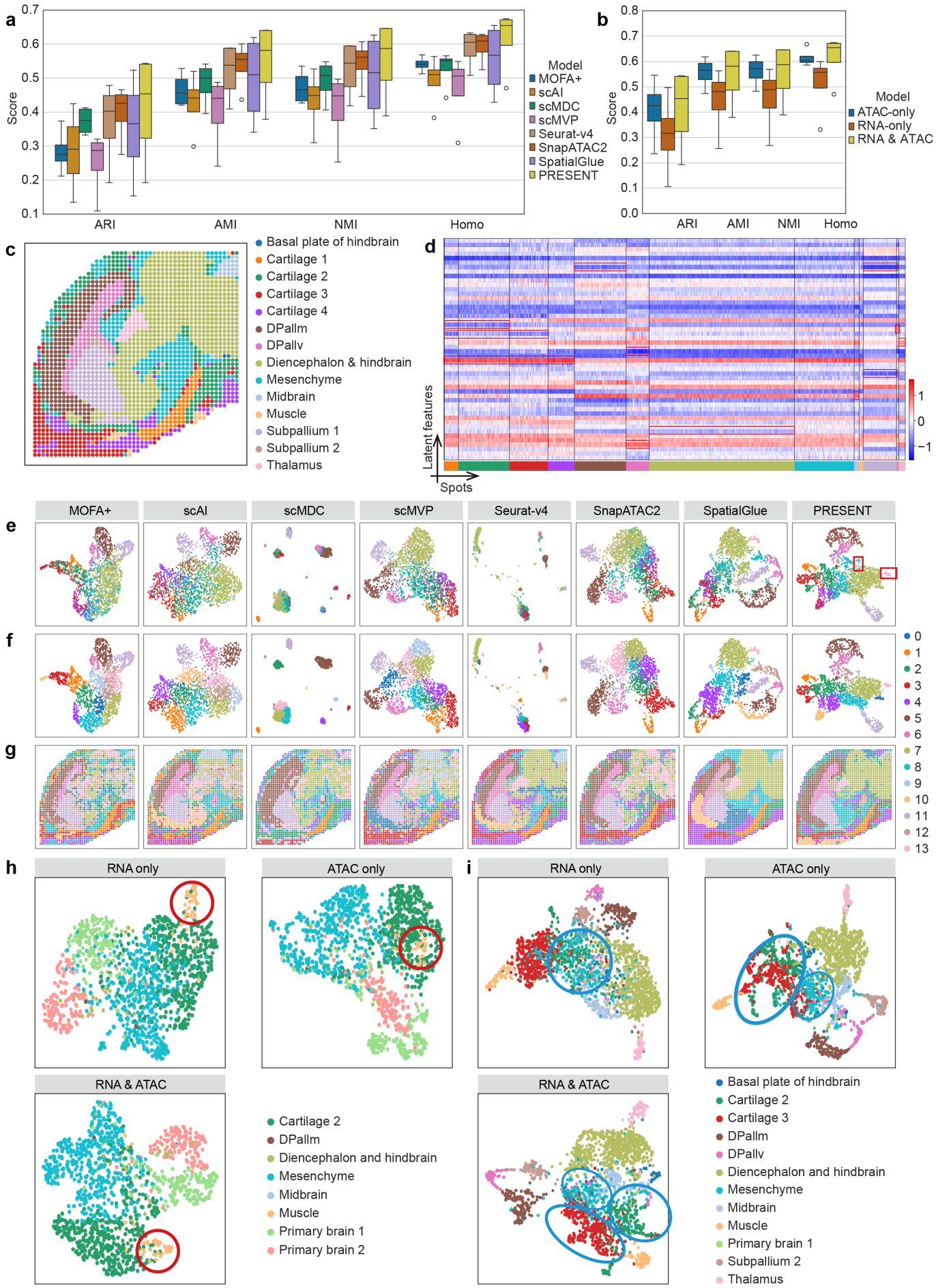

Fig. 3 | PRESENT integrates spatially aware multi-omics biological signals in spatial RNA-ATAC data. a, The quantitative comparison between PRESENT and other baseline methods and **b,** the comparison between PRESENT

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.

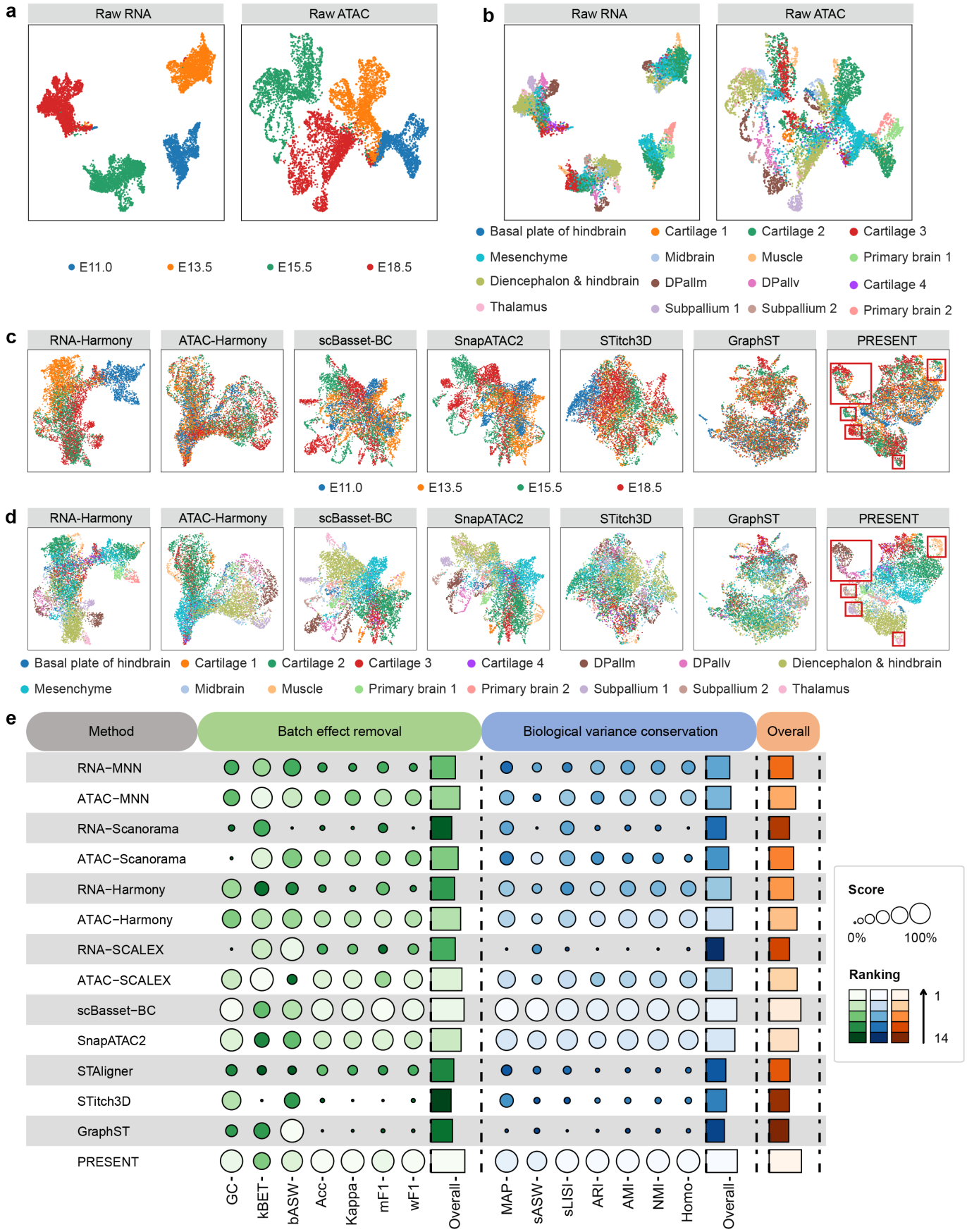

Fig. 4 | PRESENT eliminates batch effects while preserving biological signals in spatial multi-omics data. The UMAP visualization of spots across the four MISAR-seq mouse brain samples based on raw RNA or ATAC data colored by **a,** sample indices and **b,** ground truth domain labels. Note that prior to applying the UMAP algorithm to raw RNA or ATAC data, we first reduce the dimensionality to 50 using PCA transformation. **c,** The UMAP visualization of spots across the four MISAR-seq mouse brain samples based on the joint embeddings produced by different integration methods colored

by **c,** sample indices and **d,** ground truth domain labels. **e,** 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.

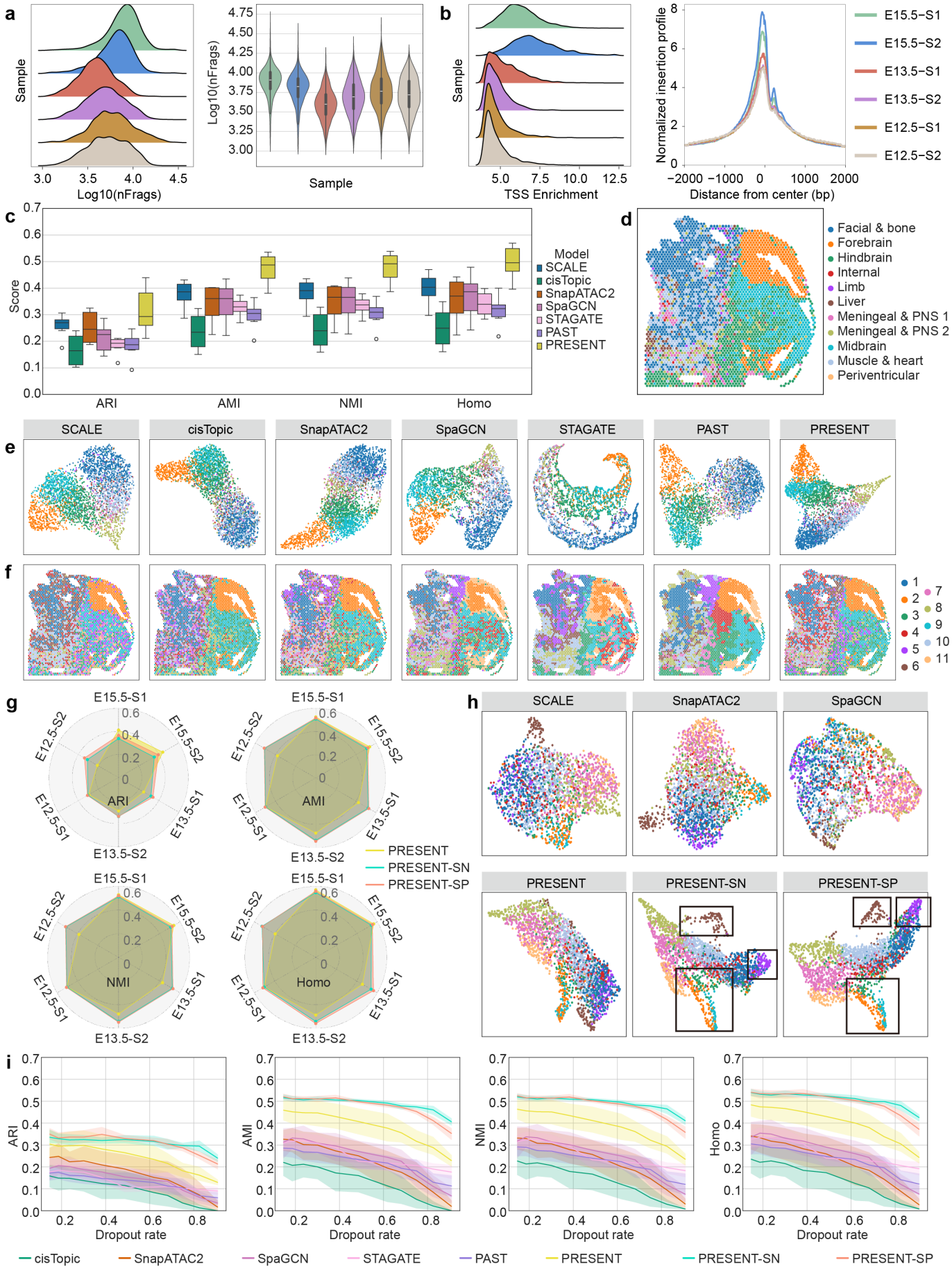

Fig. 5 | PRESENT facilitates effective cross-modal representation in spatial single-omics data. a, The ridge plot and violin plot of log10 total fragment counts, namely log10(nFrags), across the six spatial ATAC mouse embryo samples. **b,** The ridge plot of TSS enrichment scores for spots and the line 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 embeddings obtained by different methods colored by ground truth domain labels and **f,** the spatial visualization of clusters identified using the embeddings 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 levels of dropout rate. Here SCALE encountered error when the dropout rate was larger than 0.2.

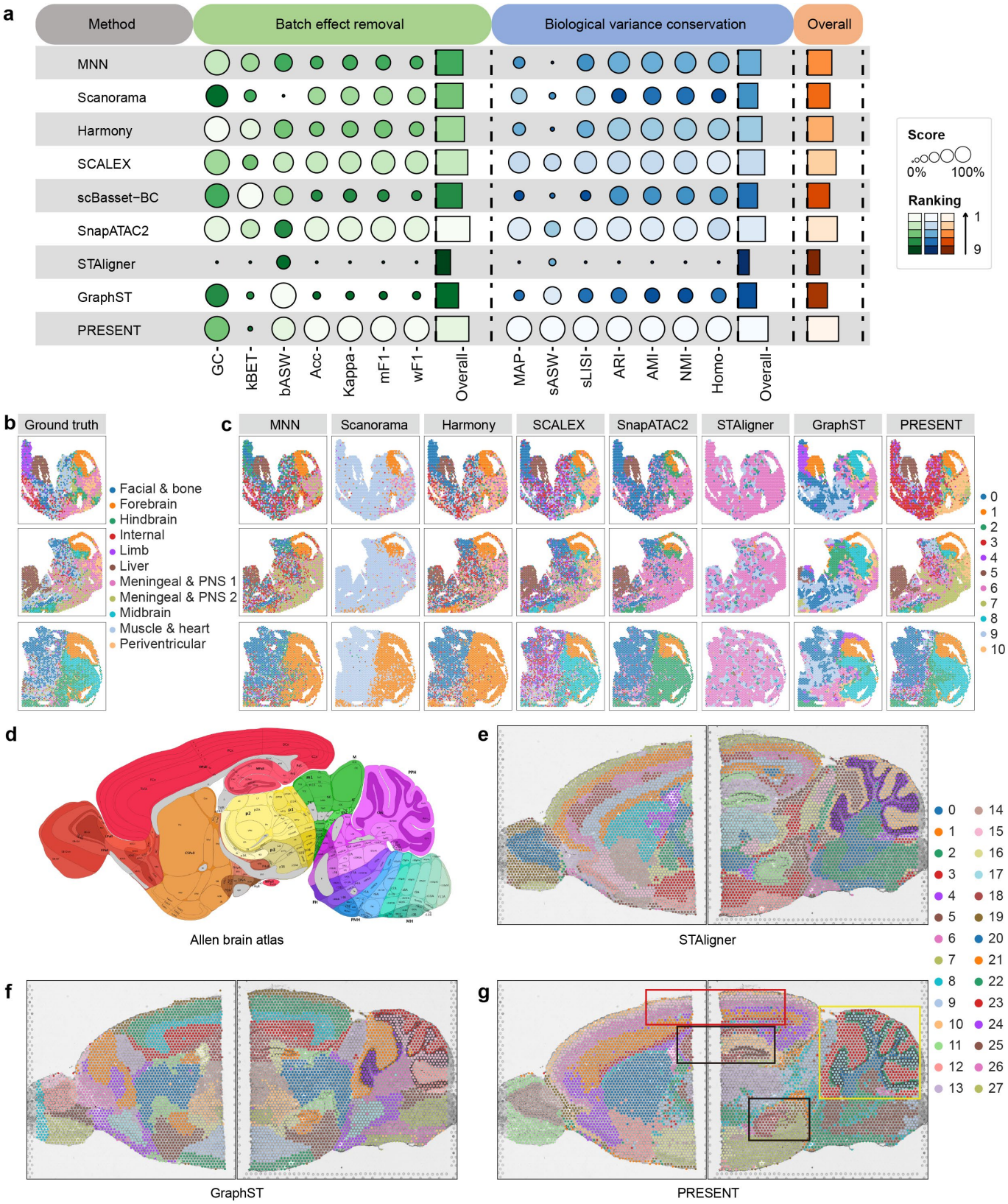

Fig. 6 | PRESENT enables vertical and horizontal integration of spatial single-omics samples. a, The quantitative evaluation of different integration methods on the three spatial ATAC mouse embryo samples using 14 metrics divided into two categories, namely batch effect removal and biological variance 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). **b,** The spatial visualization of spots across the three spatial ATAC mouse embryo samples colored by ground truth spatial domains and **c,** the spatial clusters identified based on different integration methods. The first, second and third row of **b** and **c** denotes the samples from E12.5, E13.5 and E15.5 stages, respectively. **d,** 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 **e,** STAligner, **f,** GraphST and **g,** PRESENT on the two horizontal mouse brain sagittal samples generated by the 10x Genomics Visium platform.