nature computational science

Article

https://doi.org/10.1038/s43588-025-00891-w

Integrative deep learning of spatial multi-omics with SWITCH

Received: 10 February 2025

Accepted: 9 September 2025

Published online: 29 October 2025

Check for updates

Zhongzhan Li[1], Sanqing Qu[2], Haixin Liang[1], Ruohui Tang[1], Xudong Zhang[2], Fan Lu[2], Jiani Yang[1], Ziling Gan[1], Shaorong Gao[1] ✉, Yanping Zhang[1] ✉ & Guang Chen[2] ✉

Advancements in spatial omics permit spatially resolved measurements across several biological modalities. The high cost of acquiring co-profiled multimodal data limits the analysis. This underscores the necessity for computational methods to integrate unpaired spatial multi-omics data and perform cross-modal predictions on single-modality data. The integration of spatial omics is challenging due to typically low signal-to-noise ratios. Here we introduce SWITCH (Spatially Weighted Multi-omics Integration and Cross-modal Translation with Cycle-mapping Harmonization), a deep generative model for spatial multi-omics integration. SWITCH presents a cycle-mapping mechanism that produces dependable cross-modal translations without requiring additional paired data. These cross-modal translations function as pseudo-pairs to provide supplementary signals. Systematic evaluations demonstrate that SWITCH outperforms existing methods in terms of integration accuracy and achieves more precise spatial domain delineation, resolving brain cortical structures at higher resolution. The reliability of cross-modal translations was validated, facilitating various downstream analyses such as differential analysis, trajectory inference and gene regulatory network inference.

The progress of single-cell multi-omics technology has transformed our capacity to investigate gene regulation processes across several omics layers. However, tissue dissociation processes result in the loss of spatial context, which is crucial for comprehending cellular processes. Spatial transcriptomics (ST)[1–4] was developed to overcome this constraint and has evolved to encompass spatial multi-omics, including spatial epigenomics[5–7] and proteomics[8,9]. Despite advancements enabling spatially resolved measurements of multiple omics within a single tissue section[10–12], the higher costs relative to single-omics methods, along with limitations in resolution and cellular throughput, lead to the current spatial omics data generally being restricted to single modality. Consequently, there is an urgent necessity for computational methods to efficiently integrate unpaired spatial multi-omics datasets[13].

Numerous methods have been proposed for the integration of unpaired single-cell multi-omics data, including Seurat V3[14], LIGER[15], bindSC[16], GLUE[17], SCALEX[18], SIMBA[19], MaxFuse[20], scConfluence[21] and Monae[22]. However, none of these methods use spatial information, which is crucial for accurate detection of spatial domains. Furthermore, as these methods are primarily designed for single-cell omics, they fail to fully account for the inherently low signal-to-noise ratio characteristic of spatial omics, which is particularly evident in modalities beyond transcriptomics[20,23]. This much noise may result in erroneous or failed integration and limit the precise delineation of spatial domains.

In addition, the limitations of co-profiling spatial multi-omics technologies highlight the necessity for predictive methods that can infer comprehensive multi-omics information from single-modal data. Previous methods, such as JAMIE[24], MultiVI[25] and scButterfly[26],

[1]Clinical and Translational Research Center of Shanghai First Maternity and Infant Hospital, Frontier Science Center for Stem Cell Research, School of Life Sciences and Technology, Tongji University, Shanghai, China. [2]School of Computer Science and Technology, Tongji University, Shanghai, China. ✉e-mail: gaoshaorong@tongji.edu.cn; yanpingzhang@tongji.edu.cn; guangchen@tongji.edu.cn

https://doi.org/10.1038/s43588-025-00891-w

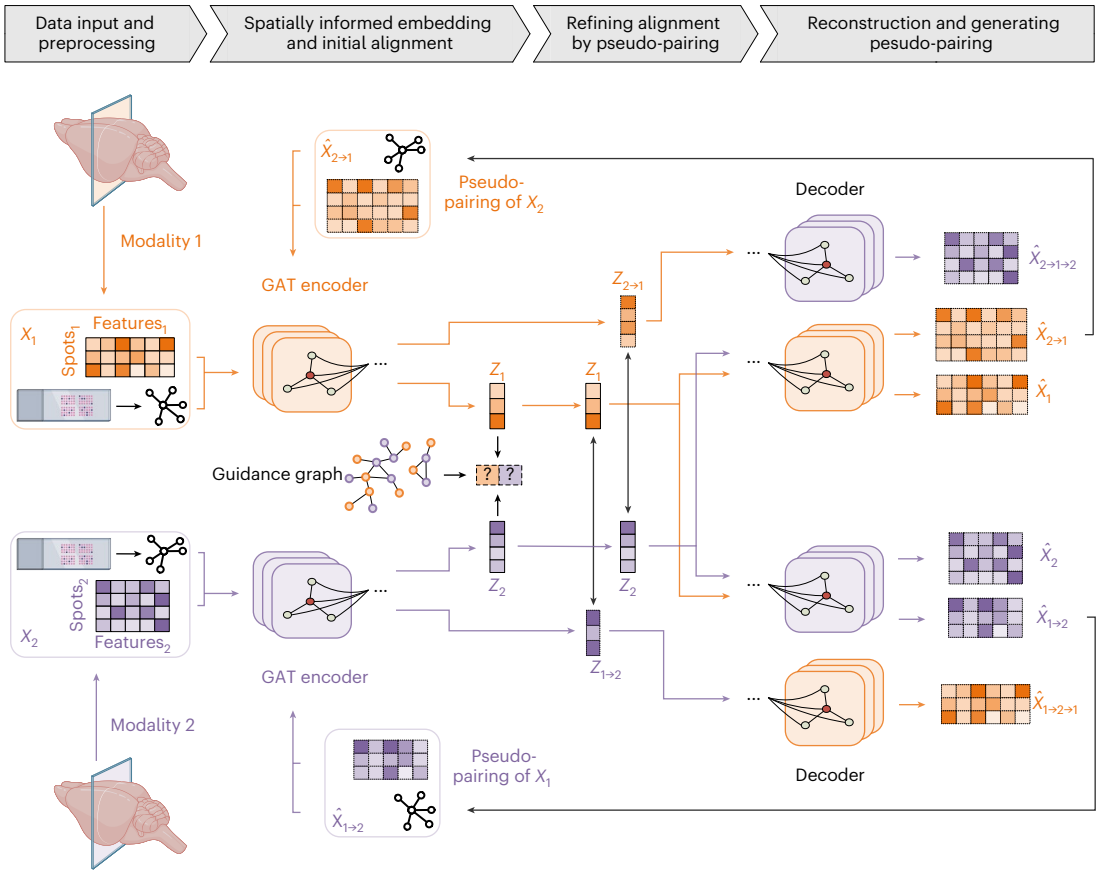


Fig. 1 | Architecture of SWITCH framework. SWITCH constructs spatial proximity graphs for each modality and utilizes modality-specific GAT encoders to generate low-dimensional embeddings of cell states. Probabilistic decoders reconstruct these embeddings into their respective feature spaces. The alignment procedure comprises two iterative phases: (1) adversarial alignment guided by a knowledge-based feature graph and (2) refinement through the minimization of differences between original embeddings and pseudo-pairings generated by cross-modal translation. A cycle-mapping mechanism validates cross-modal translations by re-encoding and reverting them to their original modality, ensuring reliable cross-modal translation without paired data. Figure created with BioRender.com.

utilize additional cross-modal pairing or cell-type information to embed different modalities into a unified space, enabling transformation across modalities in a supervised manner. However, pairing or cell-type information is often inaccessible, and it remains unclear if these transformations can effectively generalize from paired data to single-modality data[27].

To address these challenges, we propose SWITCH (Spatially Weighted Multi-omics Integration and Cross-modal Translation with Cycle-mapping Harmonization), a deep generative model designed to integrate unpaired spatial multi-omics data and perform cross-modal prediction as a unified task. SWITCH employs graph attention networks (GATs) to learn low-dimensional embeddings for each modality, followed by a two-stage alignment strategy. To effectively handle the low signal-to-noise ratio in spatial omics data, the model iteratively generates pseudo-pairs to introduce additional supervision signals, enabling high-quality integration and precise delineation of spatial domains. A cycle-mapping mechanism is incorporated to ensure consistency and reliability of the cross-modal translations throughout the optimization process. Using multiple datasets, we demonstrate that SWITCH outperforms existing methods in integrating spatial omics data. Comprehensive evaluations validate the efficacy of SWITCH's unsupervised cross-modal translations, facilitating diverse downstream analysis.

Results

The SWITCH framework

SWITCH takes feature matrices and spatial coordinates from different modalities as input (Fig. 1). For each modality, it constructs a spatial proximity graph based on the coordinates and employs a GAT encoder to learn low-dimensional embeddings. Thereafter, modality-specific probabilistic decoders map these embeddings back to the feature space. The alignment process proceeds in two stages. Initially, the alignment is attained by adversarial learning guided by a knowledge-based feature graph[17] ('guidance graph'), which models regulatory relationships among omics layers to ensure biologically meaningful alignment. Subsequently, the model generates pseudo-pairings by cross-decoding the aligned embeddings. By minimizing the discrepancy between the pseudo-pairings and original embeddings, the model gains additional supervisory signals to refine the alignment. These pseudo-pairings are then updated on the basis of the refined alignment and used in subsequent iterations. Through this iterative refinement, SWITCH enables precise integration of spatial omics and reliable cross-modal translation without requiring paired data. To ensure the accuracy of pseudo-pairings during training, SWITCH incorporates a cycle-mapping mechanism that re-encodes the pseudo-pairings and projects them back into the original modality to enforce consistency. See Methods for details.

Benchmarking SWITCH for spatial multi-omics integration

We first evaluated the integration performance of SWITCH using the E13 mouse embryo spatial ATAC–RNA-seq dataset[11], which profiles gene expression and chromatin accessibility from the same tissue section. We artificially unpaired the data and compared SWITCH with nine state-of-the-art methods: Seurat V3, GLUE, Monae, SCALEX, MaxFuse, scConfluence, SIMBA, bindSC and LIGER. The original study provided

anatomical annotations on the basis of hematoxylin and eosin staining, which served as the ground truth (Fig. 2a).

Effective integration of spatial multi-omics data requires aligning spots across modalities while accurately identifying spatial domains that are either shared or modality specific. SWITCH exhibits superior performance in both aspects, achieving unified alignment across omics layers and precisely delineating consistent domains in both modalities. It was the only approach that identified the dorsal root ganglion and spinal structures in both modalities (Fig. 2b and Supplementary Fig. 1). In contrast, competing methods revealed limitations: Seurat and SCALEX could only integrate a limited number of spots, while GLUE and Monae struggle to detect key structures in the ATAC modality. Other methods merely aligned and delineated the forebrain and midbrain structures, and their outputs were accompanied by substantial noise (Supplementary Fig. 1). For quantitative evaluation, integration performance was assessed from two perspectives: omics mixing and biological variation conservation, with each aspect measured using three distinct metrics. Across all replicates, SWITCH outperformed the other methods in both aspects, achieving an overall score 59% higher than that of the second-best method (Fig. 2c,d and Supplementary Fig. 5a,b; see Supplementary Section 1 for metric details). Moreover, SWITCH attained the best performance on the unsupervised Moran's *I* score and intraclass correlation coefficient (ICC), which respectively measure the spatial autocorrelation of clusters and the homogeneity within clusters (Supplementary Fig. 5a).

To provide a finer-scale assessment of integration performance, we evaluated spot-level alignment accuracy using the FOSCTTM (fraction of samples closer than true match)[17,28] and FOSKNN (fraction of samples whose true matches are among their *k*-nearest neighbors)[20]. Ideally, true paired spots across modalities should be in close proximity within the co-embedding space, as they represent the same underlying biological state. Thus, lower FOSCTTM and higher FOSKNN values indicate more accurate alignment. In all replicates, SWITCH outperformed the other methods on both metrics, showing a 46% improvement in FOSCTTM and a 137% improvement in FOSKNN compared with the second-best method (Fig. 2e).

We next benchmarked SWITCH on a P22 mouse brain dataset from spatial CUT&TAG–RNA-seq, profiling RNA and H3K27ac (histone H3 lysine 27 acetylation) within the same section[11]. In the absence of annotations, we used the Allen Brain Atlas[29] reference to annotate major anatomical regions such as the cortex (ctx), genu of the corpus callosum (ccg) and lateral ventricle (vl; Fig. 2f). Compared with the first embryo dataset, this dataset featured higher spatial resolution but lower signal-to-noise ratio[11]. Under these conditions, SWITCH demonstrated more pronounced advantages by accurately delineating major anatomical regions in both modalities (Fig. 2g). Notably, SWITCH precisely captured the layered structure of the cortex, including cortical layers 1 (15-ctx), 2/3 (8-ctx), 4 (13-ctx), 5 (5-ctx) and 6 (10-ctx) (Fig. 2g). This level of partitioning accuracy surpasses that of all competing methods, including the original study[11]. Additionally, the structures of these layers are well reflected in the uniform manifold approximation and projection (UMAP) embedding, indicating adequate preservation of biological variation (Supplementary Fig. 2b). All other methods introduced substantial noise and identified only a few structures, such as the ccg (Fig. 2g and Supplementary Fig. 2a). These results are further evident in Moran's *I* and ICC scores (Fig. 2h,i). Regarding spot-level alignment accuracy, SWITCH achieved the best performance again, with an 81% improvement in FOSCTTM and a 249% improvement in FOSKNN compared with the second-best method (Fig. 2j).

We further extended the analysis to a similar P22 mouse brain dataset with co-profiled RNA and H3K4me3 (trimethylation of lysine 4 on histone H3), and a P21 brain dataset from spatial RNA–ATAC-seq. In both datasets, SWITCH achieved superior performance in terms of visualization and quantitative metrics (Supplementary Figs. 3, 4 and 6a,b). We also compared SWITCH with the paired integration methods, including SpatialGlue[30] and COSMOS[31], and observed that SWITCH achieved comparable domain identification performance (Supplementary Fig. 7). Moreover, SWITCH showed robustness to hyperparameter variation, with default settings yielding near-optimal performance across all benchmarks (Supplementary Fig. 8).

Cross-modal imputation with uncertainty estimation

As a generative model, SWITCH can impute missing modalities without requiring paired data. We evaluated its cross-modal imputation performance using the artificially unpaired embryo spatial ATAC–RNA-seq dataset, predicting chromatin accessibility from gene expression and vice versa. The imputed gene expression showed a Pearson correlation of 0.41 with the observed values, while the imputed chromatin accessibility had a correlation of 0.22, with an area under the receiver operating characteristic curve of 0.28 for binarized chromatin accessibility (Fig. 3a,b and Supplementary Fig. 9a). Inspired by previous studies[23,25], we hypothesize that this relatively low correlation may stem from the low sensitivity and high noise in spatial omics data. Following procedures from a recent study[25], we smoothed the original data to mitigate these issues. As anticipated, the imputation results exhibited high correlations with the smoothed data, with a Pearson correlation of 0.87 for gene expression and 0.71 for chromatin accessibility (Fig. 3c).

In addition, SWITCH provides uncertainty estimates for its imputed results, enabling users to assess their reliability for downstream analyses. Since uncertainty is substantially influenced by expression and accessibility levels[25], we focused on the uncertainty estimates of binary accessibility. We first calculated the imputation uncertainty across domain clusters and found that regions such as the eye and midbrain exhibited higher-than-average uncertainty, whereas connective tissue and muscle showed lower uncertainty (Fig. 3d and Supplementary Fig. 9b). This is understandable, as the eye and midbrain typically exhibit higher chromatin accessibility levels during development. A strong correlation was observed between the model-estimated uncertainty and the prediction errors, suggesting that uncertainty could serve as a benchmark for assessing prediction quality (Fig. 3e). Notably, we identified a small subset of imputed results (approximately 0.2%) with low uncertainty but high imputation errors (Fig. 3e, data points in purple box). This indicates that the model was highly confident in these predictions, yet the predicted values were nearly the opposite of the observations (Fig. 3f, left panel). To investigate these discrepancies, we examined both the observed and smoothed accessibility values and found that the smoothed values aligned with SWITCH's imputed results, indicating that the cases with low uncertainty yet high errors in SWITCH's predictions may correspond to false negatives or false positives in the original data (Fig. 3f, right panel).

To evaluate the efficacy of imputed data in downstream analyses, we performed differential analyses using SWITCH's imputations. The log fold change values of imputed and observed data exhibited strong correlations, with Spearman correlations of 0.64 for gene expression and 0.51 for chromatin accessibility (Fig. 3g). When restricted to high-confidence data points (q-value < 0.01), these correlations further strengthened, reaching 0.87 for gene expression and 0.59 for chromatin accessibility (Supplementary Fig. 9c). As a specific example, we examined the differential analysis results for the eye domain. The five most significantly differentially expressed genes were all related to eye development (Fig. 3h, left panel). Although these genes were detectable in the observed expression data, their signals were marred by substantial noise (for example, for *Vax2*; Fig. 3i, left panel). In contrast, SWITCH's imputed expression precisely recovered these signals, showing enhanced enrichment in the eye domain (Fig. 3i, right panel). Similarly, the five most significantly differentially accessible peaks were associated with eye development, with two of these peaks undetectable in the observed data (Fig. 3h, right panel). The peak at chr17:85607669-85608247, corresponding to the transcription start

https://doi.org/10.1038/s43588-025-00891-w

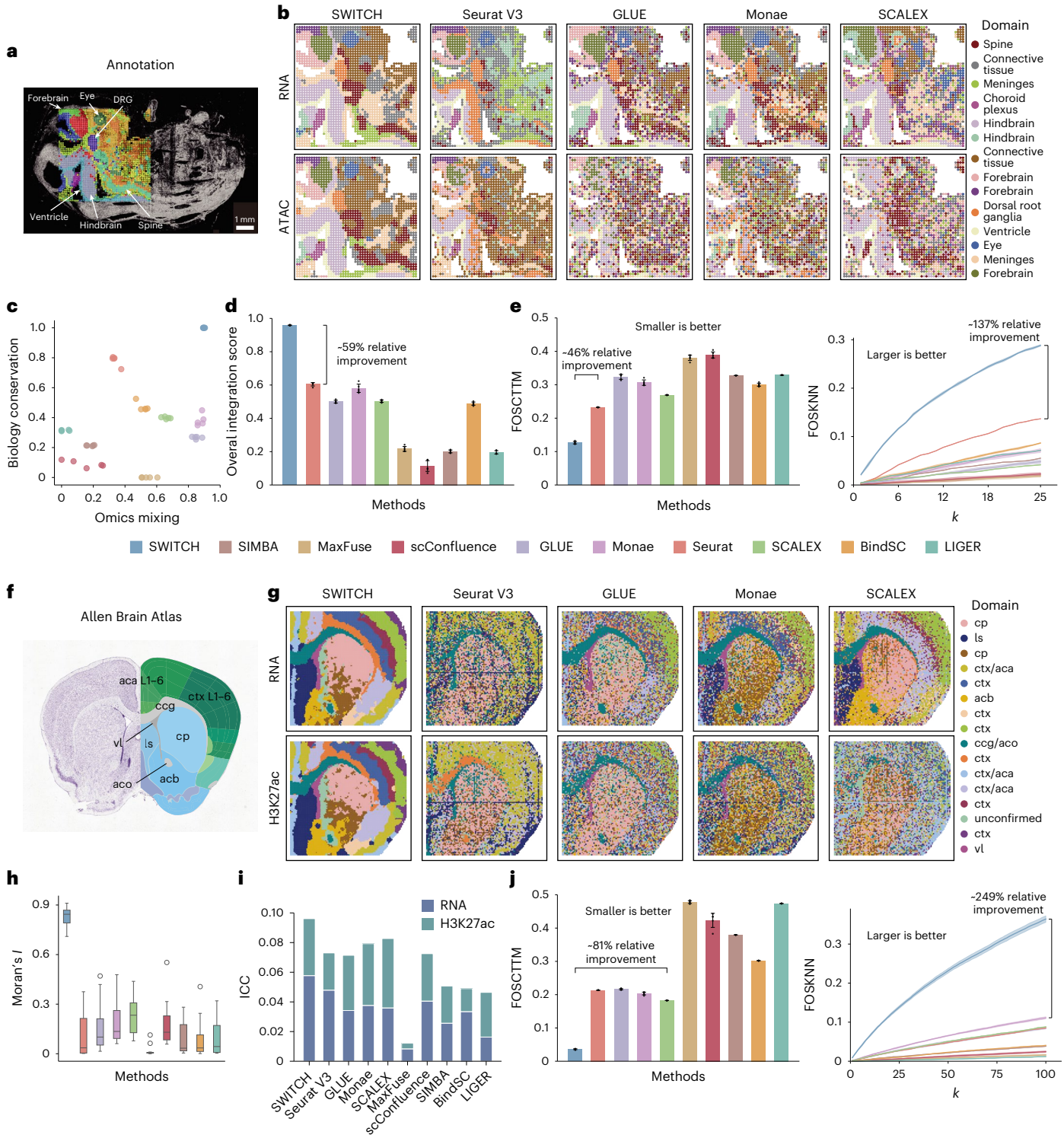


Fig. 2 | Benchmarking the integration performance of SWITCH. a, Anatomical annotation of E13 mouse embryo. DRG, dorsal root ganglion. **b**, Spatial plots of the mouse embryo data showing domains identified by different methods. The annotated labels correspond to SWITCH results, and the clustering colors and structures of other methods may not necessarily match. **c**, Omics mixing scores versus biological conservation scores for different methods over five repetitions. **d**, Overall integration scores for different methods, calculated as a weighted combination of biological conservation (60%) and omics mixing (40%) in **c** ($n = 5$ repeats with different random seeds). **e**, FOSCTTM and FOSKNN scores for each method ($n = 5$ repeats with different random seeds). **f**, Major anatomical structures of the mouse brain coronal section, annotated using the Allen Brain Atlas[29]. **g**, Spatial plots of the mouse brain data showing domains identified by different methods. The annotated labels correspond to SWITCH results, and the clustering colors and structures of other methods may not necessarily match. **h**, Box plots of Moran's *I* scores for each method ($n = 16$ clusters). In the box plots, the central line represents the median, the lower and upper hinges indicate the first and third quartiles, respectively, and the whiskers extend to 1.5 times the interquartile range. Data points beyond the whiskers are plotted individually. **i**, ICC for each method. **j**, FOSCTTM and FOSKNN scores for each method ($n = 5$ repeats with different random seeds). The error bars or shaded areas indicate mean ± s.d. Panels adapted from: **a**, ref. 11 under a Creative Commons license CC BY 4.0; **f**, ref. 29 under a Creative Commons license CC BY 4.0.

https://doi.org/10.1038/s43588-025-00891-w

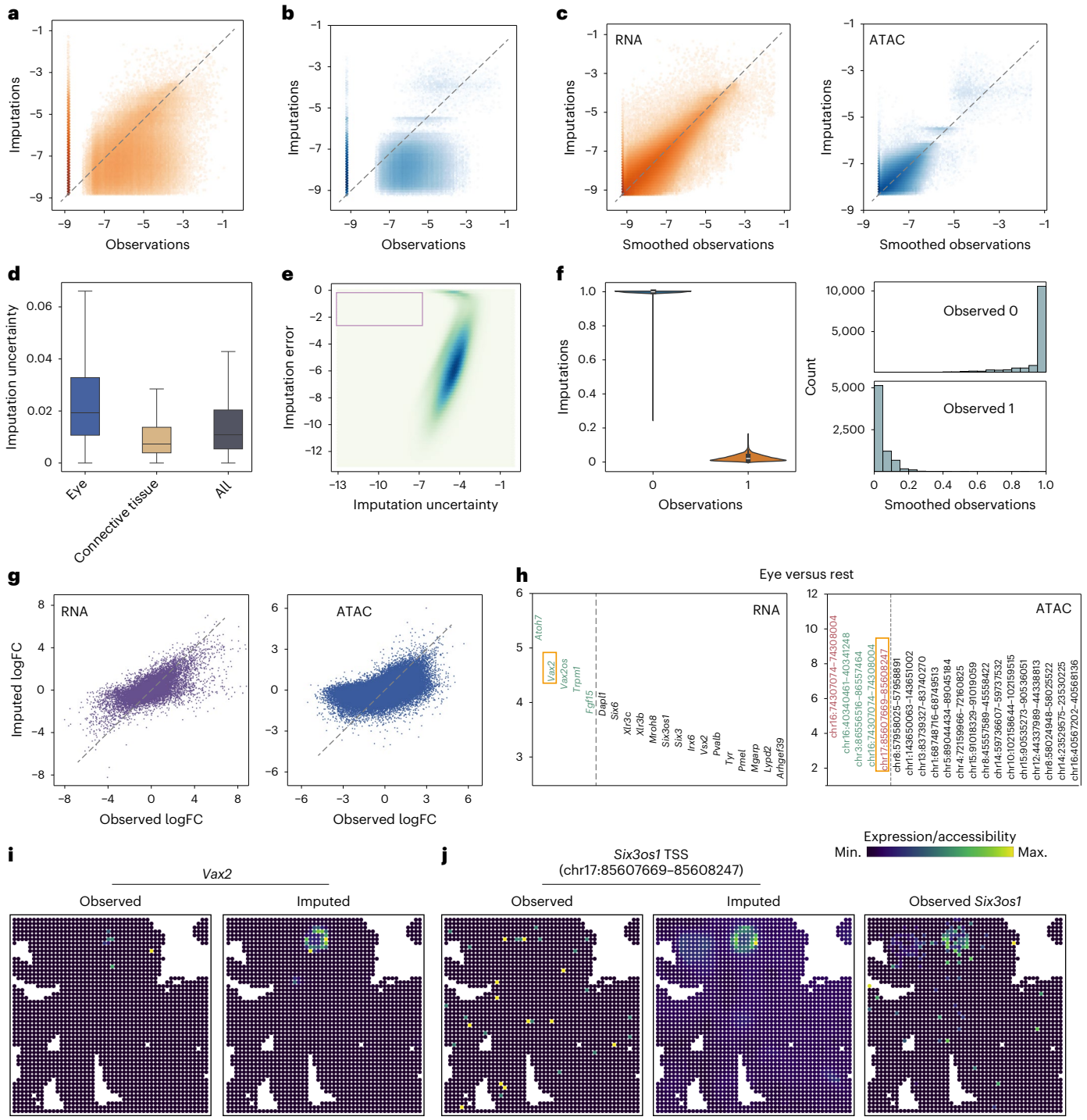


Fig. 3 | SWITCH achieves reliable cross-modal imputation and uncertainty estimation. a, SWITCH-imputed RNA versus observed RNA, with all values log-scaled. **b**, SWITCH-imputed ATAC versus observed ATAC, with all values log-scaled. **c**, SWITCH-imputed values versus smoothed RNA and ATAC observations, with all values log-scaled. **d**, Box plots of imputation uncertainty across domain clusters (n = 10,659,438). In the box plots, the central line represents the median, the lower and upper hinges indicate the first and third quartiles, respectively, and the whiskers extend to 1.5 times the interquartile range. **e**, The imputation error versus imputation uncertainty, with all values log-scaled. The purple box highlights data points with low uncertainty but high error. **f**, Comparison of binary imputed and observed accessibility for data points marked in the purple box in **e** (n = 21,102), and smoothed binary accessibility for points where observations are 1 (top right) and 0 (bottom right). In the box plots, the central line represents the median, the lower and upper hinges indicate the first and third quartiles, respectively, and the whiskers extend to 1.5 times the interquartile range. **g**, Comparison of log fold change (logFC) between imputed and observed values for expression (left) and accessibility (right). **h**, Differential analysis results for the eye domain based on imputed expression (left) and imputed accessibility (right). Features detected in observed data are shown in green; undetected ones in red. **i**, Observed and SWITCH-imputed expression values for *Vax2*. **j**, Observed and SWITCH-imputed accessibility values for *Six3os1* TSS, alongside observed expression of *Six3os1*.

site (TSS) of *Six3os1* (a regulator of the eye development gene *Six3*), was almost entirely noise in the observed data but exhibited clear specificity in the model-imputed accessibility data, aligning well with *Six3os1* expression (Fig. 3j). We further benchmarked SWITCH against four state-of-the-art cross-modal prediction methods: JAMIE[24], MultiVI[25], scButterfly[26] and Monae[22]. SWITCH consistently outperformed all competing methods across four translation tasks on two datasets (Supplementary Fig. 10).

To assess the impact of cross-modal cell-type imbalance on imputation performance, we conducted two additional perturbation experiments. In the first, gene expression data from the eye domain were removed; in the second, chromatin accessibility data were removed. Despite the absence of eye expression, SWITCH's imputations remained well correlated with the observed values, with a Pearson correlation of 0.39 with raw and 0.84 with smoothed data. Further analysis revealed that imputation performance remained stable for spots outside the eye region, but the imputed values within the eye region were systematically lower than those from the complete dataset, indicating that the model adopts a more conservative strategy when predicting without direct evidence (Supplementary Fig. 11a). A similar trend was observed in the second experiment, where the imputed accessibility data showed Pearson correlations of 0.21 and 0.69 with the raw and smoothed data, respectively (Supplementary Fig. 11b). Extending the analysis to other domain-removal scenarios revealed consistent trends (Supplementary Fig. 12). Moreover, SWITCH maintained robust integration performance under these unbalanced conditions, with only a slight decrease in overall accuracy (Supplementary Fig. 13).

Downstream analyses enabled by cross-modal imputation

Previous experiments were conducted on artificially unpaired datasets, which may introduce certain biases. To address this, we applied SWITCH to two co-sequenced microfluidic indexing-based spatial ATAC and RNA sequencing[12] mouse embryo datasets, using gene expression data from one section and chromatin accessibility data from the other. The original study provided morphology-based annotations, which served as ground truth (Fig. 4a).

Visually, SWITCH recovered spatial domains closely matching the ground-truth annotations (Fig. 4b). UMAP plots demonstrated effective mixing of omics layers, with clear delineation of domain clusters (Fig. 4c). As individual spot labels were unavailable, we used ICC and Moran's *I* to quantitatively assess the performance of domain identification. SWITCH outperformed other methods on both metrics, indicating better cluster homogeneity and less noise in the identified domains (Fig. 4d and Supplementary Fig. 14a). To validate the imputation accuracy, we conducted differential expression analysis on the imputed expression data. The Gene Ontology analysis of the top 30 differentially expressed genes for each domain cluster aligned well with their biological functions (Supplementary Fig. 15a). Similarly, we performed differential accessibility analysis using the imputed accessibility data, followed by motif enrichment analysis of the domain-specific peaks (Supplementary Fig. 15b). The enriched motifs and associated transcription factors (TFs) exhibited cell-type specificity and were consistent with known regulatory programs.

Next, we concentrated on the cortical development process in mouse embryos, using Monocle3[32] to infer the developmental trajectories of the cortical regions (domains 4 and 6; Fig. 4e, left panel). By mapping the pseudotime of spots onto their spatial location, we observed consistent spatiotemporal changes from the ventricular zone to the cortical plate in both modalities, aligning with current research findings (Fig. 4e). To verify whether the imputed expression captured changes along the pseudotime, we identified a series of genes that varied over pseudotime using the gene expression in RNA modality, and observed that these genes displayed similar trends in the ATAC-imputed expression (Pearson correlation 0.98) (Fig. 4f). For instance, the marker *Pax6* for radial glial progenitor cells is highly expressed in early cortical development and declines as development progresses. The marker *Eomes* for intermediate progenitor cells has elevated expression throughout the mid-developmental stage. The post-mitotic neuron marker *Tbr1* displays increasing expression during development, reaching its zenith in the later stages (Fig. 4g). This expression pattern is consistent with the cell lineage transition from radial glial cells to intermediate progenitor cells and then to post-mitotic neurons over cortical development[12,33].

By imputing missing expression and accessibility data, SWITCH enables systematic identification of peak–gene links. Visualization of highly correlated connections (Pearson correlation >0.45) demonstrated their consistency and domain specificity (Fig. 4h). Further integrating these links with peak–motif connections enables the construction of gene regulatory networks (GRNs). As an illustrative example, we extracted the *Neurod1*-centered regulatory subnetwork in the DPallm region (Fig. 4i). Within this network, we identified several well-established regulatory relationships, such as the regulation of *Bhlhe22* by *Neurod1*[30], and the regulation of *Cux2* and *Satb2* by *Neurod2*[34–36]. Notably, our network reveals a cascading regulatory pathway in which *Neurod1* regulates *Neurod2* via *Bhlhe22*. This mechanism is indirectly supported by previous studies[30] showing that *Bhlhe22* is rapidly induced within one day of *Neurod1* activation, whereas *Neurod2* expression increases three days later.

Despite SWITCH's success in integrating truly unpaired data, it may still face challenges when integrating modalities with lower similarity. To quantitatively assess alignment quality under such conditions, we developed a metric termed the integration consistency score, which evaluates alignment accuracy on the basis of biological prior knowledge. This metric was first validated on four paired datasets, where scores substantially declined when spots were randomly mismatched to simulate misalignment (Supplementary Fig. 16a). On the mouse embryo datasets at different developmental stages, the scores declined as the developmental stages of the two modalities became increasingly distant (Supplementary Fig. 16c–e).

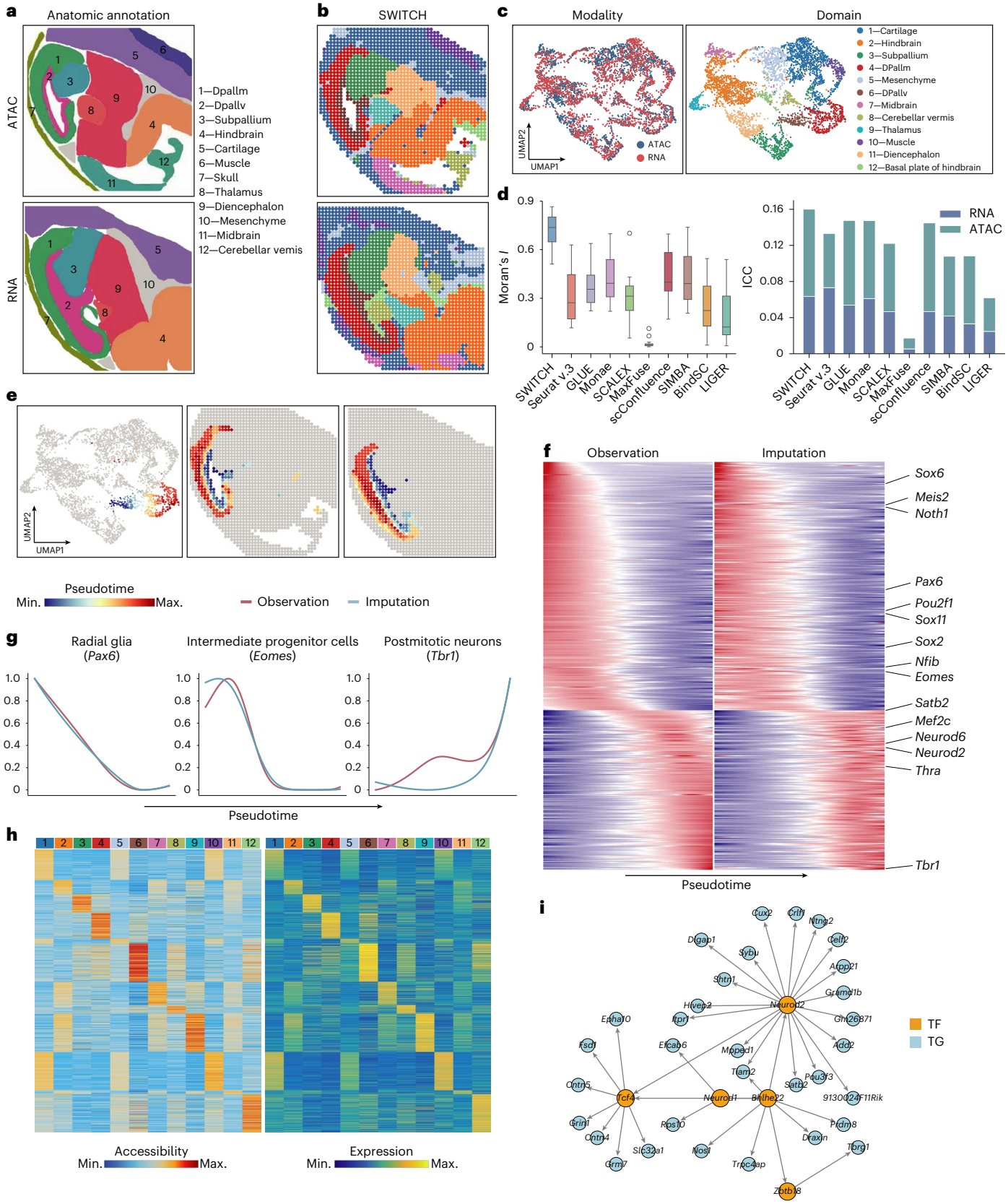
Integration of ST and single-nucleus ATAC-seq data

As an emerging technology, spatial omics still trails behind single-cell data in resolution, cell throughput and cost-effectiveness. A practical alternative is to combine these two approaches to leverage their complementary strengths. Previous studies have primarily focused on integrating spatial and single-cell transcriptomics data, whereas SWITCH enables integration across different omics layers. Moreover, by supporting cross-modal translation, SWITCH extends unimodal spatial data into multi-omics, allowing for more comprehensive downstream analyses.

Fig. 4 | Cross-modal imputation in unpaired mouse embryo data enables diverse downstream analyses. a, Anatomic annotation of major structural regions in the mouse embryo. DPallm, mantle zone of the dorsal pallium; DPallv, ventricular zone of the dorsal pallium. **b**, Spatial plots of the mouse embryo data showing domains identified by SWITCH. **c**, UMAP representation computed from the latent space generated by SWITCH, with spots colored by modality and domain. **d**, Moran's *I* score ($n = 12$ clusters) and ICC for each method. In the box plots, the central line represents the median, the lower and upper hinges indicate the first and third quartiles, respectively, and the whiskers denote 1.5 times the interquartile range. Data points beyond the whiskers are plotted individually. **e**, UMAP plot showing the trajectory of cortical development (left) and the trajectory mapped onto the spatial plots (middle and right). **f**, Heatmaps of gene expression changes over pseudotime, showing observed and imputed. **g**, Dynamic changes of genes *Pax6*, *Eomes* and *Tbr1* over pseudotime. **h**, Heatmaps depicting chromatin accessibility (left) and their linked genes (right), with rows clustered on the basis of accessibility. **i**, Subgraph of the GRN for Dpallm, showing the first- and second-order targets of *Neurod1*. TG, target gene. Panel **a** adapted with permission from ref. 12, Springer Nature Limited.

https://doi.org/10.1038/s43588-025-00891-w

In this section, we applied SWITCH to integrate ST data from an adult mouse coronal brain, obtained using Stereo-seq[4], with single-nucleus ATAC-seq data from similar tissue[37] (Fig. 5). The UMAP plot revealed effective mixing of the two modalities while maintaining clear separation of different cell types (Supplementary Fig. 18a). The original ATAC data provided the sampling regions for cells, allowing us to estimate the alignment accuracy by contrasting the sampled regions with their respective cell-type compositions. For instance, cells sampled from region B were primarily annotated as L2/3 and L5 cortical excitatory neurons (EX L2/3 and EX L5); cells from region G



https://doi.org/10.1038/s43588-025-00891-w

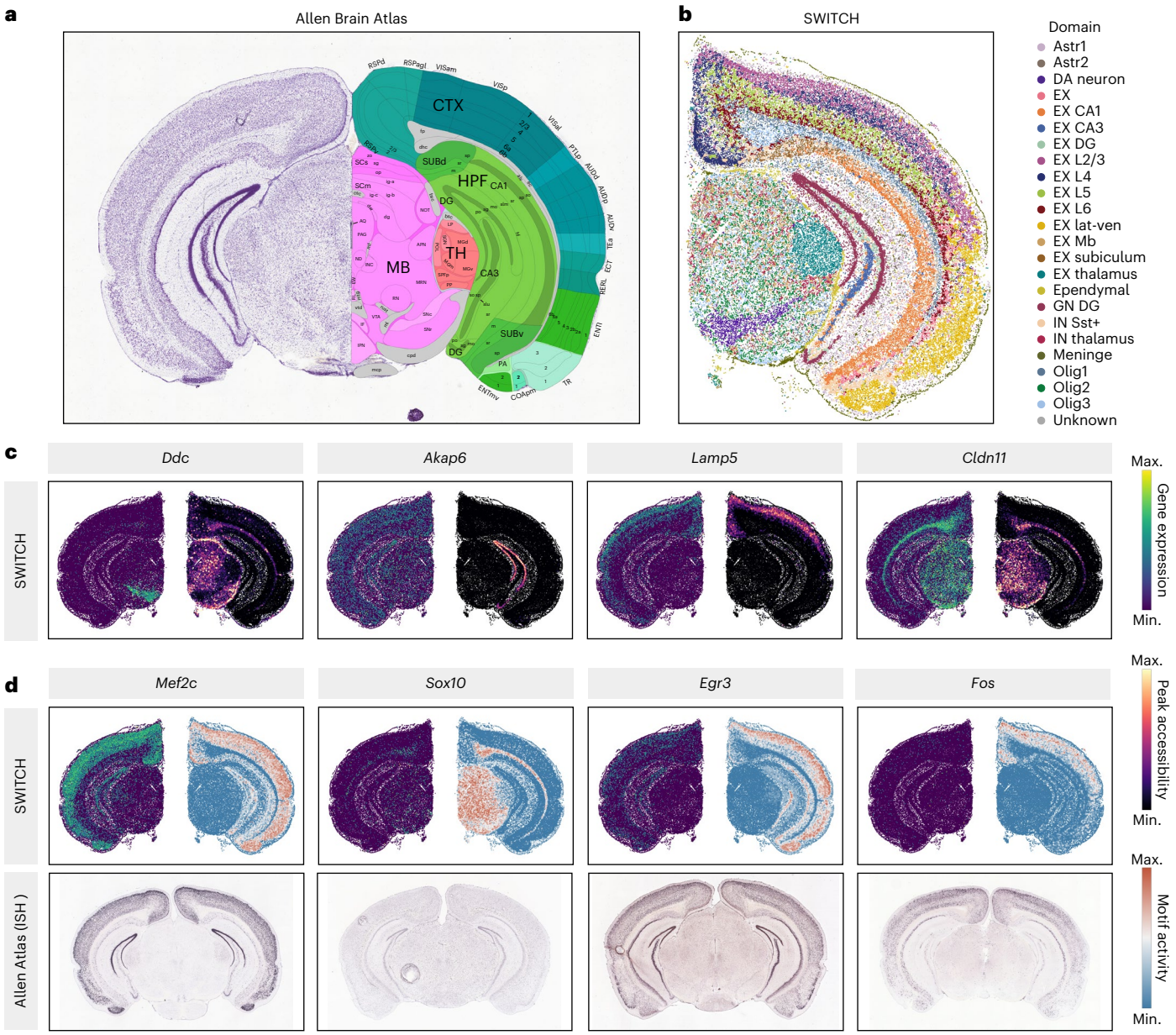


Fig. 5 | SWITCH integrates ST and single-cell ATAC data to infer spatial chromatin accessibility. a, Annotation of the mouse brain coronal section from the Allen Brain Atlas. **b**, Visualization of spatial domains identified by integrating ST and single-cell ATAC data using SWITCH. **c**, Observed gene expression in ST (left) and SWITCH-imputed accessibility at the corresponding TSS (right). **d**, Observed expression (top left) and activity scores (inferred from SWITCH-imputed accessibility) of selected TFs (top right), and the corresponding ISH images of the mouse brain coronal section from the Allen Brain Atlas (bottom). Panel **a** adapted from ref. 29 under a Creative Commons license CC BY 4.0.

were mainly annotated as excitatory neurons from the midbrain and thalamic regions (EX Mb and EX thalamic; Supplementary Fig. 18b). Further validation with the Allen Brain Atlas[29] confirmed that SWITCH accurately captured major anatomical structures, such as cortical layers L1–L6, fiber tracts and hippocampal regions CA1, CA3 and DG (Fig. 5a,b).

We next evaluated SWITCH's imputation performance, focusing on translating gene expression in the ST data into chromatin accessibility (Fig. 5c). Given the absence of ground truth, we cross-validated the imputed results using an external paired single-cell multi-omics dataset from mouse brain[38], in which RNA and ATAC were jointly profiled. Using the RNA modality as a shared anchor, we transferred single-cell ATAC signals to the Stereo-seq data via Seurat[14]. Comparison of the transferred and imputed values showed a moderate Pearson correlation of 0.45. Considering the inherent technical noise in single-cell ATAC data, we applied *k*-nearest neighbor smoothing before transfer, which improved the correlation to 0.65 (Supplementary Fig. 19).

We next examined the spatial patterns of imputed chromatin accessibility and compared them with the expression of proximal genes. In some cases, genes were enriched in specific regions, but the imputed accessibility did not show localization (the case for *Ddc*). In other cases, even when genes did not exhibit enrichment in specific regions, the imputed accessibility showed distinct spatial patterns (the case for *Akap6*). Only a few genes, such as *Lamp5* and *Cldn11*, displayed consistent spatial patterns across the two modalities. These observations reflect the inherent complexity and nonlinearity of gene–peak regulatory relationships.

We further inferred TF motif activity scores from the imputed accessibility profiles and observed strong correlations with gene expression levels for certain TFs such as *Mef2c*, *Sox10* and *Egr3*, as validated

by in situ hybridization (ISH) data from the Allen Brain Atlas[29] (Fig. 5d). Notably, the TF *Fos*, despite showing substantial noise in the expression data, had activity scores that closely matched the Allen ISH results.

Finally, we reconstructed GRNs by integrating observed gene expression with SWITCH-inferred chromatin accessibility. The resulting GRNs captured several known regulatory relationships. For instance, *Ascl2* is known to regulate *Car2*[39], *Mobp* is regulated by *Sox10*[40] and *Tcf4* regulates *Grik3*, *Csmd2* and *Pde10a*[41].

Discussion

Although SWITCH was primarily designed for unpaired spatial multi-omics data, it is capable of incorporating paired datasets. Such pairing information can aid modality alignment and refinement of pseudo-pair generation, further improving integration accuracy. Nevertheless, external paired datasets should be incorporated with caution, as the resulting batch effects or cell-type discrepancies may undermine the accuracy of predictions. Moreover, the conceptual framework of SWITCH may also be extended to other related tasks, such as perturbation effect prediction or the translation between histological images and gene expression.

SWITCH operates under the assumption that different modalities share a common semantic space. This assumption may be challenged in cases where the inter-modality correspondence is weak, such as between epigenomic and proteomic data. A promising strategy to address this issue is the introduction of an intermediate modality to bridge weakly linked data types—for example, transcriptomics could serve as a mediator to improve integration between epigenomic and proteomic profiles. Optimizing SWITCH for these challenging scenarios is a potential direction for future research and would enhance its applicability across diverse biological contexts. In addition, as spatial omics technologies continue to advance and generate increasingly high-throughput data, further optimization of SWITCH for large-scale datasets will be important for its practical utility.

Methods

The framework of SWITCH

Construction of the spatial neighbor graph. Assuming that spatially adjacent spots have similar cell states, we transform the spatial information into an undirected neighbor graph $G = (V, E)$, where V represents the set of all spots and E represents the set of edges connecting these spots.

The adjacency matrix A encodes the neighbor relationships between spots: if the Euclidean distance between two spots i and j is less than a predefined threshold r, then $A_{i,j} = 1$; otherwise, $A_{i,j} = 0$. r can be flexibly adjusted on the basis of data characteristics or experimental requirements, ensuring that each spot has an appropriate number of neighbors to accurately reflect spatial structure features. To preserve the graph's completeness and capture features of local regions, we introduce self-loops by setting $A_{i,i} = 1$ for i in the adjacency matrix.

GAT-based modality-specific encoder. Let $X_1 \in \mathbb{R}^{N_1 \times F_1}$ and $X_2 \in \mathbb{R}^{N_2 \times F_2}$ depict the two modalities to be integrated, where N_1 and N_2 represent the number of spots in the tissue, and F_1 and F_2 denote the number of features for each modality, respectively. For instance, in the integration of ST and epigenomics, X_1 corresponds to a gene-by-spot matrix, while X_2 corresponds to a region-by-spot matrix.

Each modality $m \in 1, 2$ is equipped with an independent encoder f_z^m, which encodes the input features $x_{m,i}$ of spot i into the parameters of a d-dimensional multivariate normal (MVN) distribution $z_{m,i} \in \mathbb{R}^d$, specifically the mean $\mu_{m,i} \in \mathbb{R}^d$ and variance $\sigma^2_{m,i} \in \mathbb{R}^d$. This design allows each modality to capture its unique characteristics. To represent both expression patterns and the neighborhood microenvironment, we employ GATs as the primary encoder. To mitigate the oversmoothing effect common in graph-based methods, the final layer of the encoder is implemented as a linear layer. Specifically, the embedding of spot i produced by the lth ($l \in 1, 2, \ldots, L-1$) layer of the encoder for modality m is computed as

$$h_{m,i}^{(l)} = \text{Leaky ReLU}\left(\sum_{j \in \mathcal{N}i} \alpha_{m,ij}^{(l)} \times \left(W_m^{(l)} h_{m,j}^{(l-1)}\right)\right) \tag{1}$$

where $\mathcal{N}i$ denotes the set of neighboring spots for spot i, $W_m^{(l)}$ is the weight matrix for the lth layer of the encoder for modality m, Leaky ReLU is the activation function and $\alpha_{m,ij}^{(l)}$ represents the attention coefficient between spot i and its neighbor j in the lth layer of the encoder for modality m, computed as follows:

$$\alpha_{m,ij}^{(l)} = \frac{\exp\left(e_{m,ij}^{(l)}\right)}{\sum_{k \in \mathcal{N}_i} \exp\left(e_{m,ik}^{(l)}\right)} \tag{2}$$

$$e_{m,ij}^{(l)} = \text{Leaky ReLU}\left(\mathbf{a}_m^{(l)\top}\left[W_m^{(l)} h_{m,i}^{(l-1)} \parallel W_m^{(l)} h_{m,j}^{(l-1)}\right]\right) \tag{3}$$

where $\mathbf{a}_m^{(l)}$ is a trainable weight vector and $\parallel$ represents vector concatenation.

In the Lth encoder layer, the features processed through the $L-1$ GAT layers are mapped to the distribution parameters of $z_{m,i}$ via two linear layers as follows:

$$\mu_{m,i} = h_{m,i}^{(L-1)} W_{\mu_m} + \mathbf{b}_{\mu_m} \tag{4}$$

$$\log \sigma^2_{m,i} = h_{m,i}^{(L-1)} W_{\sigma_m} + \mathbf{b}_{\sigma_m} \tag{5}$$

where W_{μ_m} and W_{σ_m} denote trainable weight matrices and b_{μ_m} and b_{σ_m} are bias vectors.

Directly sampling $z_{m,i}$ from the MVN distribution complicates backpropagation due to the non-differentiability of the sampling process. To address this, we apply the reparameterization trick, where $z_{m,i}$ is expressed as a deterministic function of the learned parameters $\mu_{m,i}$ and $\sigma^2_{m,i}$, combined with a random noise term ϵ drawn from a standard normal distribution $\mathcal{N}(0, I)$. This is formulated as

$$z_{m,i} = \mu_{m,i} + \sigma_{m,i} \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I) \tag{6}$$

where $\odot$ denotes Hadamard multiplication. In summary, the encoder f_z^m maps the input features $x_{m,i}$ to a probability latent representation described by the following distribution:

$$q(z_{m,i} | x_{m,i}; \psi_m) = \text{MVN}\left(z_{m,i}; \mu_{m,i}, \sigma^2_{m,i} I\right) \tag{7}$$

where ψ_m encompasses all trainable parameters of the encoder for modality m, including weights and biases across all layers.

Adversarial alignment via prior guidance graph. To align the spot embeddings z_m generated by the modality-specific encoders, we introduce a discriminator D with an m-dimensional softmax output:

$$\hat{y} = \text{softmax}(D(z_{m,i}; \phi)) \tag{8}$$

where ϕ denotes the learnable parameters of the discriminator, $\hat{y} \in \mathbb{R}^m$ and $\hat{y}_m$ represents the probability of spot i being predicted as belonging to the mth modality. The discriminator's objective is to classify the modality of spot i on the basis of its embedding $z_{m,i}$. This adversarial mechanism encourages the latent representations produced by the encoders to converge, ensuring consistency across modalities.

However, inherent differences between modalities may lead to semantically inconsistent latent variables, potentially causing

misalignment. To address this, we leverage a guidance graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, inspired by GLUE, to impose constraints on the alignment process on the basis of prior biological regulatory relationships. In this graph, $\mathcal{V}$ denotes the set of nodes, each corresponding to a modality-specific feature such as a gene (from RNA-seq) or an open chromatin peak (from ATAC-seq). The set of edges $\mathcal{E}$ encodes known or hypothesized regulatory relationships, typically linking features across different omics layers, with each edge from node i to node j further annotated by a weight w_{ij} and a sign s_{ij}. Here, $w_{ij} \in (0, 1]$ represents the strength of the interaction, and $s_{ij} \in \{-1, 1\}$ indicates the direction of regulation, with $s_{ij} = 1$ denoting positive regulation and $s_{ij} = -1$ denoting negative regulation.

In all analyses presented in this study, we constructed regulatory edges in the guidance graph solely on the basis of the genomic proximity between peaks and genes. Specifically, a directed edge was established between a peak and a gene if the peak overlapped with either the gene body or the proximal promoter region, defined as the 2-kb region upstream of the TSS. For each such edge, we assigned $w_{ij} = 1.0$ and $s_{ij} = 1$, indicating a presumed positive regulatory relationship between the peak and the gene. Reverse edges and self-loops were also included for all nodes to facilitate message passing and maintain numerical stability.

Similar to GLUE, we treat the guidance graph as an observed variable and employ a dedicated graph convolutional network (GCN) encoder to represent each node j in graph $\mathcal{G}$ as a d-dimensional MVN random variable $v_j \in \mathbb{R}^d$. The encoding process is defined as

$$q(v_j|\mathcal{G}; \psi_{\mathcal{G}}) = \mathrm{MVN}\left(v_j; \mu_j, \sigma_j^2 I\right) \tag{9}$$

$$\mu_j = f_{\mathrm{GCN}_{\mu_j}}(\mathcal{G}) \tag{10}$$

$$\log \sigma_j^2 = f_{\mathrm{GCN}_{\sigma_j^2}}(\mathcal{G}) \tag{11}$$

where $\psi_{\mathcal{G}}$ represents the set of parameters of the encoder for the guidance graph $\mathcal{G}$. Next, the guidance graph decoder g_v maps the node embeddings V back to the original graph:

$$\hat{\mathcal{G}} \sim \prod_{i,j \in V} \mathrm{Bernoulli}\left(p(e_{ij})\right) \tag{12}$$

$$p(e_{ij}) = g_v(v_i, v_j; \varphi_{\mathcal{G}}) \tag{13}$$

where $\varphi_{\mathcal{G}}$ represents the set of parameters of the decoder for the guidance graph $\mathcal{G}$, $p(e_{ij})$ denotes the probability of an edge existing between nodes i and j, and $\hat{\mathcal{G}}$ is the reconstructed graph.

Modality-specific decoder. The decoder g_z^m for modality m is designed to generate the full-dimensional features corresponding to modality m from the spot embeddings z, which can originate from any modality, and the feature embeddings V_m. Instead of directly reconstructing the original feature matrix, we map the latent space to the parameters of a specific distribution to account for noise in the original data. For example, we employ a negative binomial (NB) distribution for the RNA-seq and ATAC-seq data. The generation process for the full-dimensional features of modality m from a spot i belonging to any modality is expressed as

$$\hat{x}_{m,i} \sim \prod_{j \in V_m} \mathrm{NB}\left(\rho_{m,j}, \theta_{m,j}\right) \tag{14}$$

$$\rho_{m,j} = \mathrm{softmax}_j\left(\alpha_m \odot V_m^{\top} z_i + \beta_m\right) \times l_i \tag{15}$$

where α_m, β_m and $\theta_{m,j}$ are learnable parameters of the decoder for modality m. $\rho_{m,j}$ and $\theta_{m,j}$ represent the mean and dispersion of the NB distribution, respectively. α_m is a scaling factor, β_m is a bias factor and $\mathrm{softmax}_j$ represents performing softmax on the jth dimension of the data. l_i is the total count for spot i, which represents a scaling factor that adjusts the feature for spot i. In the case where spot i does not belong to modality m, l_i is a scalar value greater than 0 that scales the values in a meaningful way to maintain consistency with the modality's feature distribution.

In summary, through the decoder g_z^m, we map the latent representation z_i to a probabilistic distribution over the features of modality m:

$$\hat{x}_{m,i} = g_z^m(z_i, V_m; \varphi_m) \tag{16}$$

where φ_m represents the set of parameters of the decoder for modality m.

Model training of SWITCH

Reconstruction loss. The reconstruction loss measures the difference between the original data and the reconstructed data generated by the model. It quantifies the model's ability to regenerate input features from the latent representations. Minimizing this loss ensures that the learned latent representations preserve the characteristics of the original features. For omics data, the loss is defined as the negative log-likelihood of the observed counts given the predicted parameters:

$$\mathcal{L}_{\mathrm{data}}^{\mathrm{recon}} = \frac{1}{2} \sum_{m \in \{1,2\}} \mathcal{L}_{x_m}^{\mathrm{recon}} \tag{17}$$

$$\mathcal{L}_{x_m}^{\mathrm{recon}} = -\mathbb{E}_{q(z_m|x_m;\psi_m)}\left[\log p\left(x_m | z_m, v_m; \varphi_m\right)\right] \tag{18}$$

where $\log p(x_m \mid z_m, v_m; \varphi_m)$ is the data likelihood, conditioned on the latent variables z_m and the feature embeddings v_m, with φ_m being the parameters of the decoder for modality m.

For the guiding graph data, the reconstruction loss is defined as

$$\mathcal{L}_{\mathcal{G}}^{\mathrm{recon}} = -\mathbb{E}_{q(V|\mathcal{G};\psi_{\mathcal{G}})}\left[\log p\left(\mathcal{G} | V; \varphi_{\mathcal{G}}\right)\right] \tag{19}$$

where $\log p(\mathcal{G}|V; \varphi_{\mathcal{G}})$ is the graph likelihood, conditioned on the node embeddings V.

Adversarial loss. The adversarial loss is employed to train the discriminator to accurately classify the modality of the point embeddings. Specifically, the discriminator learns to differentiate between the different modalities present in the data. On the other hand, the encoder is trained to produce modality-invariant representations by minimizing the ability of the discriminator to distinguish between these modalities. This adversarial framework encourages the encoder to extract features that are independent of any particular modality, thereby promoting generalization across different input types. Formally, the adversarial loss is defined as

$$\mathcal{L}^{\mathrm{adv}} = \frac{1}{2} \sum_{m \in \{1,2\}} \mathcal{L}_m^{\mathrm{adv}} \mathcal{L}_m^{\mathrm{adv}} = -\mathbb{E}_{q(z_m|x_m;\psi_m)}\left[\sum_{m \in \{1,2\}} y_m \log \hat{y}_m\right]. \tag{20}$$

Cycle-mapping loss. In SWITCH, each modality m is associated with a modality-specific decoder g_z^m, which takes modality-specific features V_m and spot embeddings z (originating from any modality) as inputs and generates the modality-specific features for that spot. When the input embeddings z originate from modality m, this process is referred to as reconstruction. Conversely, when the embeddings arise from a different modality n, it is termed cross-modal translation. While reconstruction loss ensures that the embeddings can map back to their original modality, it does not guarantee accurate mapping to another

modality. Previous methods address this by relying on external paired data, which limits their generalizability. To overcome this, we introduce a cycle-mapping mechanism that enables spot embeddings z to translate between modalities using dedicated decoders, ensuring cross-modal consistency without requiring paired data.

For the translation from modality m to n, the cycle-mapping process is defined as follows:

$$z_{m,i} = f_z^m(x_{m,i}) \tag{21}$$

$$\hat{x}_{m\to n,i} = g_z^n(z_{m,i}, V_n) \tag{22}$$

$$\hat{z}_{m\to n,i} = f_z^n(\hat{x}_{m\to n,i}) \tag{23}$$

$$\hat{x}_{m\to n\to m,i} = g_z^m(\hat{z}_{m\to n,i}, V_m) \tag{24}$$

where $z_{m,i}$ denotes the latent representation of spot i in modality m and $\hat{x}_{m\to n,i}$ is the cross-modal translation of $z_{m,i}$ to modality n. $\hat{z}_{m\to n,i}$ re-encodes latent representation in modality n, and $\hat{x}_{m\to n\to m,i}$ represents the cycle-mapped reconstruction in modality m. The cycle-mapping loss quantifies the difference between the reconstructed features $\hat{x}_{m\to n\to m,i}$ and the original input $x_{m,i}$. For translation from modality m to n, the loss is defined as

$$\mathcal{L}_{m\to n}^{\text{cycle}} = -\mathbb{E}_{q(\hat{z}_{m\to n}|\hat{x}_{m\to n};\psi_m)}\left[\log p(x_m|\hat{z}_{m\to n}, V_m;\varphi_m)\right]. \tag{25}$$

A similar formulation applies for the reverse translation $n \to m$. Thus, the overall cycle-mapping loss is given by

$$\begin{aligned}\mathcal{L}^{\text{cycle}} = &-\mathbb{E}_{q(\hat{z}_{m\to n}|\hat{x}_{m\to n};\psi_n)}\left[\log p(x_m|\hat{z}_{m\to n}, V_m;\varphi_m)\right]\\ &-\mathbb{E}_{q(\hat{z}_{n\to m}|\hat{x}_{n\to m};\psi_m)}\left[\log p(x_n|\hat{z}_{n\to m}, V_n;\varphi_n)\right].\end{aligned} \tag{26}$$

The effectiveness of the cycle-mapping loss can be understood from two perspectives:

- Interpretability within the target modality. It ensures that cross-modal imputation results are interpretable within the target modality by requiring the imputed data to be re-encoded into latent representations by the target modality's encoder. For example, in the translation $m \to n$, the imputed features $\hat{x}_{m\to n,i}$ are meaningful if their re-encoding $\hat{z}_{m\to n,i} = f_z^n(\hat{x}_{m\to n,i})$ aligns with the latent structure of modality n.
- Faithfulness to the biological state. It ensures that imputed features reflect the true biological state of the spots. This is achieved by requiring that cycle-mapped reconstructions $\hat{x}_{m\to n\to m,i}$ closely resemble the original features $x_{m,i}$. This consistency demonstrates that the cross-modal translation retains biological information and structural integrity.

In summary, the cycle-mapping loss plays a pivotal role in maintaining the biological and structural integrity of cross-modal imputations. By ensuring that imputed results are interpretable within the target modality and faithful to the original biological state, it allows SWITCH to achieve accurate and meaningful cross-modal integration without relying on paired data.

Embedding-alignment loss. The embedding-alignment loss ensures consistency between the latent representations generated during cross-modal translation and the original embeddings. As described earlier, SWITCH facilitates cross-modal translation, enabling spot representations to be transformed across different modalities. This process generates pseudo-paired data for the spots, which assists the encoder in aligning representations across modalities.

Specifically, for a spot i in modality m, its latent variable $z_{m,i}$ should align with the latent variable $\hat{z}_{m\to n,i}$ obtained from the cross-modal translation result $\hat{x}_{m\to n,i}$. Since these latent variables represent the same underlying cell state, they should be close to each other in the latent space. We employ the cosine similarity to quantify this alignment:

$$\text{dist}(z_{m,i}, \hat{z}_{m\to n,i}) = \frac{z_{m,i}\cdot\hat{z}_{m\to n,i}}{\| z_{m,i}\| \| \hat{z}_{m\to n,i}\|}. \tag{27}$$

The embedding-alignment loss is then defined as

$$\mathcal{L}^{\text{align}} = e^{-\text{dist}(z_{m,i},\hat{z}_{m\to n,i})} + e^{-\text{dist}(z_{n,i},\hat{z}_{n\to m,i})}. \tag{28}$$

Kullback–Leibler-regularization loss. To ensure that the latent variables z are well structured and aligned with a prior distribution $p(z)$, we further employ the Kullback–Leibler (KL) divergence as a regularization term. The prior distribution $p(z)$ is typically chosen to be a standard normal distribution $\mathcal{N}(0, I)$. This regularization promotes smoothness and continuity within the latent space, enabling the model to learn meaningful latent representations and exhibit robust generative capabilities.

$$\begin{aligned}\mathcal{L}^{\text{KL}} = &\,\text{KL}(q(z_m|x_m;\psi_m)\parallel p(z))\\ &+\text{KL}(q(z_n|x_n;\psi_n)\parallel p(z))\\ &+\text{KL}(q(\hat{z}_{m\to n}|\hat{x}_{m\to n};\psi_n)\parallel p(z))\\ &+\text{KL}(q(\hat{z}_{n\to m}|\hat{x}_{n\to m};\psi_m)\parallel p(z)).\end{aligned} \tag{29}$$

Overall optimization objective. The total loss function used for training SWITCH is a weighted sum of individual loss components, including reconstruction losses, adversarial loss, cycle-mapping loss, embedding-alignment loss and KL-regularization loss. This composite objective ensures that the model optimally balances data reconstruction, modality alignment and latent-space regularization. The total loss function is defined as

$$\begin{aligned}\mathcal{L}_{\text{total}} = &\,\lambda_{\text{data}}^{\text{recon}}\mathcal{L}_{\text{data}}^{\text{recon}} + \lambda_{\mathcal{G}}^{\text{recon}}\mathcal{L}_{\mathcal{G}}^{\text{recon}} + \lambda^{\text{adv}}\mathcal{L}^{\text{adv}}\\ &+\lambda^{\text{cycle}}\mathcal{L}^{\text{cycle}} + \lambda^{\text{align}}\mathcal{L}^{\text{align}} + \lambda^{\text{KL}}\mathcal{L}^{\text{KL}}\end{aligned} \tag{30}$$

where $\lambda_{\text{data}}^{\text{recon}}$, $\lambda_{\mathcal{G}}^{\text{recon}}$, λ^{adv}, λ^{cycle}, $\lambda^{\text{align and}}$ λ^{KL} are hyperparameters that control the relative contributions of each loss term total objective.

Implementation details. For all datasets, the model is trained using a learning rate of 0.0002. The maximum number of training epochs is automatically determined on the basis of the dataset size and learning rate, ensuring efficient convergence. A default set of hyperparameters, optimized for general use across a wide range of scenarios, is provided in Supplementary Table 1.

Benchmarking integration methods

We compared SWITCH with nine other methods, including Seurat V3[14] (v.3.2.3), LIGER[15] (v.0.2.0), bindSC[16] (v.1.0.0), GLUE[17] (v.0.2.0), SCALEX[18] (v.1.0.2), MaxFuse[20] (v.09302022V), SIMBA[19] (v.1.2), scConfluence[21] (v.0.1.0) and Monae[22]. For methods that require a gene activity matrix as input, such as Seurat V3, LIGER and bindSC, gene activity scores were computed from the peak matrix using the ArchR R package[42] (v.1.0.2). All competing methods were applied with their respective recommended default hyperparameter settings and preprocessing steps, and the same number of highly variable genes and peaks was selected to ensure a fair comparison under consistent conditions. The benchmarking code is publicly available at https://github.com/zzli123/SWITCH/.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Source data are available with this paper. All datasets used in this study are publicly available. Detailed information about the datasets, as well as the accessible code and links, can be found in Supplementary Table 2. The processed datasets are freely available at https://doi.org/10.5281/zenodo.15602076 (ref. 43).

Code availability

The source code of SWITCH, along with Jupyter notebooks for reproducing the results in this study, is available at https://github.com/zzli123/SWITCH/ and https://doi.org/10.5281/zenodo.16522594 (ref. 44).

References

1. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
2. Rodriques, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
3. Liu, Y. et al. High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell* **183**, 1665–1681 (2020).
4. Chen, A. et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792 (2022).
5. Deng, Y. et al. Spatial-CUT&tag: spatially resolved chromatin modification profiling at the cellular level. *Science* **375**, 681–686 (2022).
6. Deng, Y. et al. Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature* **609**, 375–383 (2022).
7. Lu, T., Ang, C. E. & Zhuang, X. Spatially resolved epigenomic profiling of single cells in complex tissues. *Cell* **185**, 4448–4464 (2022).
8. Goltsev, Y. et al. Deep profiling of mouse splenic architecture with codex multiplexed imaging. *Cell* **174**, 968–981 (2018).
9. He, S. et al. High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nat. Biotechnol.* **40**, 1794–1806 (2022).
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. Zhang, D. et al. Spatial epigenome–transcriptome co-profiling of mammalian tissues. *Nature* **616**, 113–122 (2023).
12. Jiang, F. et al. Simultaneous profiling of spatial gene expression and chromatin accessibility during mouse brain development. *Nat. Methods* **20**, 1048–1057 (2023).
13. Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell and spatial multi-omics. *Nat. Rev. Genet.* **24**, 494–515 (2023).
14. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
15. Welch, J. D. et al. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* **177**, 1873–1887 (2019).
16. Dou, J. et al. Bi-order multimodal integration of single-cell data. *Genome Biol.* **23**, 112 (2022).
17. Cao, Z. J. & Gao, G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nat. Biotechnol.* **40**, 1458–1466 (2022).
18. Xiong, L. et al. Online single-cell data integration through projecting heterogeneous datasets into a common cell-embedding space. *Nat. Commun.* **13**, 6118 (2022).
19. Chen, H., Ryu, J., Vinyard, M. E., Lerer, A. & Pinello, L. SIMBA: single-cell embedding along with features. *Nat. Methods* **21**, 1003–1013 (2024).
20. Chen, S. et al. Integration of spatial and single-cell data across modalities with weakly linked features. *Nat. Biotechnol.* **42**, 1096–1106 (2024).
21. Samaran, J., Peyré, G. & Cantini, L. scConfluence: single-cell diagonal integration with regularized inverse optimal transport on weakly connected features. *Nat. Commun.* **15**, 7762 (2024).
22. Tang, Z. et al. Modal-nexus auto-encoder for multi-modality cellular data integration and imputation. *Nat. Commun.* **15**, 9021 (2024).
23. You, Y. et al. Systematic comparison of sequencing-based spatial transcriptomic methods. *Nat. Methods* **21**, 1743–1754 (2024).
24. Cohen Kalafut, N., Huang, X. & Wang, D. Joint variational autoencoders for multimodal imputation and embedding. *Nat. Mach. Intell.* **5**, 631–642 (2023).
25. Ashuach, T. et al. MultiVI: deep generative model for the integration of multimodal data. *Nat. Methods* **20**, 1222–1231 (2023).
26. Cao, Y. et al. scButterfly: a versatile single-cell cross-modality translation method via dual-aligned variational autoencoders. *Nat. Commun.* **15**, 2973 (2024).
27. Wu, K. E., Yost, K. E., Chang, H. Y. & Zou, J. Babel enables cross-modality translation between multiomic profiles at single-cell resolution. *Proc. Natl Acad. Sci. USA* **118**, e2023070118 (2021).
28. Liu, J., Huang, Y., Singh, R., Vert, J. P. & Noble, W. S. Jointly embedding multiple single-cell omics measurements. *Algorithms Bioinform.* **143**, 10 (2019).
29. 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).
30. Ma, N. X., Puls, B. & Chen, G. Transcriptomic analyses of NeuroD1-mediated astrocyte-to-neuron conversion. *Dev. Neurobiol.* **82**, 375–391 (2022).
31. Zhou, Y. et al. Cooperative integration of spatially resolved multi-omics data with COSMOS. *Nature Commun.* **16**, 27 (2025).
32. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
33. Englund, C. et al. Pax6, Tbr2, and Tbr1 are expressed sequentially by radial glia, intermediate progenitor cells, and postmitotic neurons in developing neocortex. *J. Neurosci.* **25**, 247–251 (2005).
34. Bayam, E. et al. Genome-wide target analysis of NEUROD2 provides new insights into regulation of cortical projection neuron migration and differentiation. *BMC Genom.* **16**, 681 (2015).
35. Bormuth, I. et al. Neuronal basic helix–loop–helix proteins Neurod2/6 regulate cortical commissure formation before midline interactions. *J. Neurosci.* **33**, 641–651 (2013).
36. Hahn, M. A. et al. Reprogramming of DNA methylation at NEURO2-bound sequences during cortical neuron differentiation. *Sci. Adv.* **5**, eaax0080 (2019).
37. Zu, S. et al. Single-cell analysis of chromatin accessibility in the adult mouse brain. *Nature* **624**, 378–389 (2023).
38. Chai, H et al. Tri-omic single-cell mapping of the 3D epigenome and transcriptome in whole mouse brains throughout the lifespan. *Nat. Methods* **22**, 994–1007 (2025).
39. Bogutz, A. B. et al. Transcription factor ASCL2 is required for development of the glycogen trophoblast cell lineage. *PLoS Genet.* **14**, e1007587 (2018).
40. Kunke, M. et al. SOX10-mediated regulation of enteric glial phenotype in vitro and its relevance for neuroinflammatory disorders. *J. Mol. Neurosci.* **75**, 26 (2025).
41. Forrest, M. P. et al. The psychiatric risk gene transcription factor 4 (TCF4) regulates neurodevelopmental pathways associated with schizophrenia, autism, and intellectual disability. *Schizophr. Bull.* **44**, 1100–1110 (2018).
42. Granja, J. M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).

43. Li, Z. Benchmark datasets for SWITCH. *Zenodo* https://doi.org/10.5281/zenodo.15602076 (2025).
44. Zhongzhan, L. SWITCH: a deep generative model for spatial multi-omics integration and cross-modal prediction. *Zenodo* https://doi.org/10.5281/zenodo.16522594 (2025).

Acknowledgements

This work was supported by the National Key Research and Development Program of China (2022YFA1103102), the National Natural Science Foundation of China (32170589, 32370616, 32488101, 32330030), Xiaomi Young Talents Program and Shanghai Eastern Youth Talent Program.

Author contributions

G.C., Y.Z. and S.G. conceived and supervised the study. Z.L. and S.Q. designed the method with help of X.Z. and F.L. Z.L. conducted the experiment with the help of J.Y., R.T., H.L. and Z.G. Z.L. designed and created all main figures. Z.L., Y.Z., S.Q. and G.C. wrote the paper. All authors read and approved the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43588-025-00891-w.

Correspondence and requests for materials should be addressed to Shaorong Gao, Yanping Zhang or Guang Chen.

Peer review information *Nature Computational Science* thanks Christopher A. Jackson, Michelle Y. Y. Lee and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Ananya Rastogi, in collaboration with the *Nature Computational Science* team. Peer reviewer reports are available.

nature portfolio

Corresponding author(s): Shaorong Gao, Yanping Zhang, Guang Chen

Last updated by author(s): Jul 28, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was employed for data collection.
Data analysis	Data analysis was performed using our custom-developed SWITCH software (v1.0.0), available at [https://github.com/zzli123/SWITCH/]. Benchmark analyses were conducted using the following software packages: Seurat (v5.1.0), LIGER (v0.2.0), BindSC (v1.0.0), GLUE (v0.2.0), SCALEX (v1.0.2), MaxFuse (v09302022V), SIMBA (v1.2.0), scConfluence (v0.1.0) and Monae (vN.A.). For data preprocessing and downstream analyses, we utilized the following software: anndata (v0.9.2), ArchR (v1.0.2), MACS2 (v2.2.9.1), matplotlib (v3.7.5), numpy (v1.22.4), pandas (v1.5.3), scanpy (v1.9.8), scib (v1.1.5), scikit-learn (v1.3.2), scipy (v1.10.1), seaborn (v0.13.2), snapatac2 (v2.4.0), sparse (v0.14.0), squidpy (v1.2.3), torch (v2.1.1) and torch-geometric (v2.4.0).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Source data for Figs. 2–5 is available with this paper. All dataset used in this study are already published and publicly available. They can be downloaded from the following sources:

1. Mouse embryo, Spatial-ATAC-RNA-seq
ATAC: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM6204624
RNA: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM6204637
2. Mouse brain, Spatial CUT&TAG-RNA-seq (RNA+H3K27ac)
H3K27ac: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM6204621
RNA: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM6204635
3. Mouse brain, Spatial CUT&TAG-RNA-seq (RNA+H3K4me3)
H3K4me3: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM6704980
RNA: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM6753046
4. Mouse brain, Spatial-ATAC-RNA-seq
ATAC: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM6204623
RNA: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM6204636
5. Mouse embryo, MISAR-seq
ATAC: https://www.biosino.org/node/download/node/data/public/OED00772634
RNA: https://www.biosino.org/node/download/node/data/public/OED00772595
6. Mouse brain, Stereo-seq and snATAC-seq
ATAC: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE246791
RNA: https://ftp.cngb.org/pub/SciRAID/stomics/STDS0000058/stomics/Mouse_brain_cell_bin.h5ad
7. Mouse brain, ChAIR data
https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA024774

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N.A.
Reporting on race, ethnicity, or other socially relevant groupings	N.A.
Population characteristics	N.A.
Recruitment	N.A.
Ethics oversight	N.A.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Seven publicly available datasets were used in this study.
Data exclusions	Lowly expressed spots and features were removed based on commonly used and pre-established criteria in the field.

Replication	In this study, all experiments were repeated multiple times to ensure stable evaluation. For the assessment of different methods, we used n = 5 independent runs with different random seeds. For the calculation of Moran's I scores, the sample size n was determined based on the number of clusters identified in the dataset. No formal sample size calculation was performed; the chosen sample sizes are consistent with prior studies and were deemed sufficient to ensure reliable and reproducible results.
Randomization	N.A. Randomization is not relevant to this study since no new experimental data were collected.
Blinding	N.A. Our experiments did not involve human participants or subjective assessments.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*