


# DECIPHER for learning disentangled cellular embeddings in large-scale heterogeneous spatial omics data

Received: 12 March 2025

Chen-Rui Xia <sup>1,2,3</sup>, Zhi-Jie Cao <sup>1,2,3</sup> ✉ & Ge Gao <sup>1,2</sup> ✉

Accepted: 1 August 2025

Published online: 27 August 2025

 Check for updates

The functional role of a cell, shaped by the sophisticated interplay between its molecular identity and spatial context, is often obscured in current spatial modeling. In efforts to model large-scale heterogeneous spatial data in silico effectively and efficiently, we introduce DECIPHER, which disentangles cells' intra-cellular and extra-cellular representation through a novel cross-scale contrast learning strategy. In addition to superior performance over state-of-the-arts, systematic benchmarks and various real-world case studies showed that the disentangled embeddings produced by DECIPHER enable delineating cell-environment interaction across multiple scales. Of note, DECIPHER is highly scalable, capable of handling spatial atlases with millions of cells which is largely infeasible for existing methods.

In multicellular organisms, cells must interact to organize into three-dimensional tissue structures. The spatial context of an individual cell is as critical as its intrinsic properties in determining its physiological function and potential pathological alterations<sup>1,2</sup>. While spatial omics technologies enable systematic characterization of both these aspects, their proper in silico representation remains a serious challenge<sup>3–5</sup>.

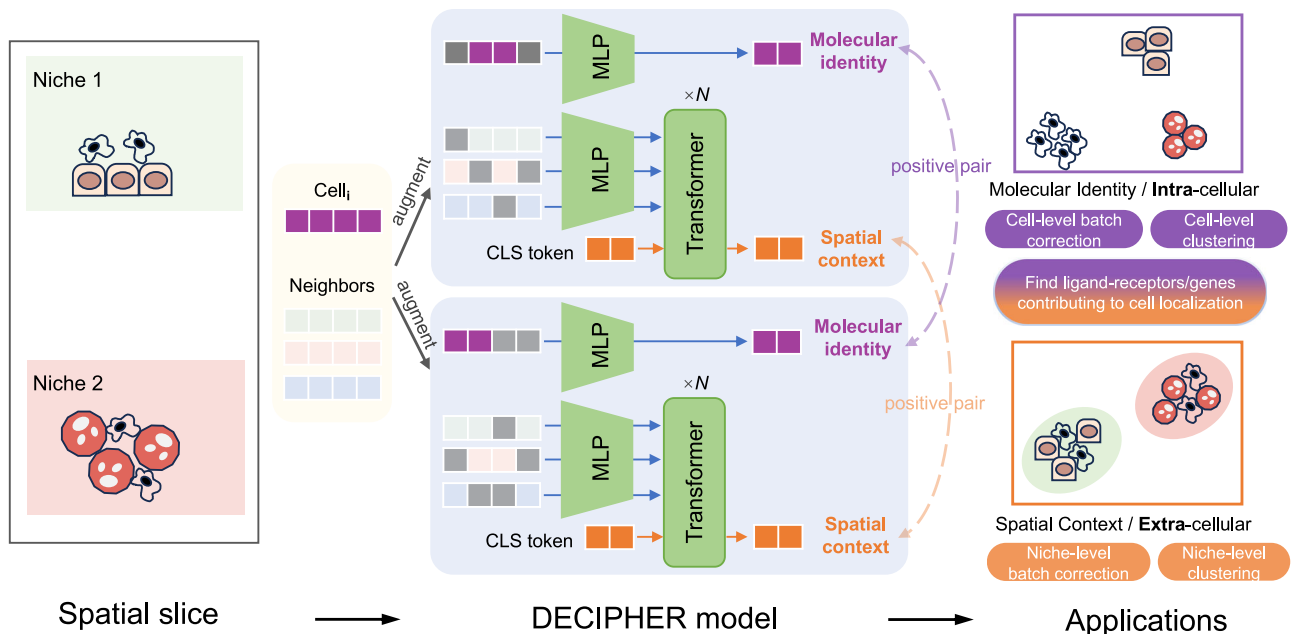
Embedding, or the procedure of representing original high-dimensional spatial data as low-dimensional vectors, has been widely adopted to enrich biologically relevant signals while simultaneously removing technical artifacts<sup>6,7</sup>. Inspired by its great success in single-cell omics<sup>6</sup>, current methods often aim to generate a holistic embedding for spatial data by encoding both the cellular transcriptome and spatial context into a single low-dimensional space<sup>3–5</sup> as the backbone of downstream analyses. For example, STAGATE<sup>4</sup>, Banksy<sup>5</sup>, BASS<sup>8</sup>, and scNiche<sup>9</sup> are designed for spatial domain clustering to identify the spatial niches of cells; GraphST<sup>10</sup> and STADIA<sup>11</sup> additionally support correcting batch effects from different experiments and biological conditions; SLAT can align multiple spatial slices. While this holistic embedding approach eases model design and implementation, it effectively introduces intrinsic entanglement between the molecular identity and spatial context, obfuscating direct dissection of their interplay. This approach is also

hard to connect with cell-cell communication, where embedding-free methods such as NicheNet<sup>12</sup> and CellChat<sup>13</sup> are commonly used to find active ligand-receptor across given cell types. Besides, scalability remains a significant limitation, e.g., graph neural network (GNN)-based methods (STAGATE, GraphST, SLAT, scNiche) often need to load entire spatial graphs into memory (or GPU memory), which is impractical for datasets containing millions of cells.

Here, we present DECIPHER, a context-aware deep learning model designed for large, heterogeneous spatial datasets. DECIPHER employs a novel cross-scale contrastive learning framework to learn disentangled representations as separate intra- and extra-cellular embeddings (Fig. 1). Notably, benefiting from a transformer-based architecture<sup>14</sup>, DECIPHER's scalability is significantly improved compared to previous methods, which is crucial in the context of increasing throughput of spatial omics. In addition to its outstanding performance showed by systematic benchmarks, DECIPHER's disentangled embeddings enable quantitative characterization of intra- and extra-cellular interactions which is crucial for various real-world applications, such as calling cell interaction molecules for B cell localization in Germinal Center (GC) and identifying key genes for immune cells infiltration in tumor micro-environments (TME). DECIPHER is publicly available at <https://github.com/gao-lab/DECIPHER>.

<sup>1</sup>State Key Laboratory of Gene Function and Modulation Research, School of Life Sciences, Biomedical Pioneering Innovative Center (BIOPIIC) and Beijing Advanced Innovation Center for Genomics (ICG), Center for Bioinformatics (CBI), Peking University, Beijing, China. <sup>2</sup>Changping Laboratory, Beijing, China.

<sup>3</sup>These authors contributed equally: Chen-Rui Xia, Zhi-Jie Cao. ✉e-mail: [caozj@mail.cbi.pku.edu.cn](mailto:caozj@mail.cbi.pku.edu.cn); [gaog@mail.cbi.pku.edu.cn](mailto:gaog@mail.cbi.pku.edu.cn)



**Fig. 1 | Overview of DECIPHER.** Architecture of DECIPHER. DECIPHER models the disentangled intracellular molecular identity and extracellular spatial context of cells, respectively. Augmented views of  $Cell_i$  and its spatial neighbors are obtained by random dropout (Methods). Then, DECIPHER uses an omics encoder (MLP) to obtain the molecular identity embedding of  $Cell_i$ , and uses a spatial encoder (transformer-based) to obtain the spatial context embedding based on  $Cell_i$ 's

spatial neighbors. The model is optimized by contrastive learning to get positive pairs closer and pushed apart negative pairs (Supplementary Fig. 1). The right panel shows downstream applications based on cell identity or/and context embeddings, such as batch correction, clustering in both cell type and niche levels, as well as finding ligand-receptors/genes contributing to cell localization (see Fig. 3a, b for details).

## Results

### DECIPHER for learning disentangled intra- and extra-cellular embeddings

Unlike current spatial omics embedding methods seeking to find a holistic embedding<sup>3–5</sup>, DECIPHER employs two separate but interconnected components to preserve both shared and specific information in gene expression and spatial context. In particular, the “Omics Encoder” is a multi-layer perceptron (MLP) that learns an intracellular, molecular identity-oriented embedding from the expression profile. The “Spatial Encoder” is based on the transformer architecture, which treats each neighboring cell as a token, and projects the cell's spatial information into an independent latent space (Fig. 1, Methods). Both components are optimized simultaneously through a dedicated cross-scale contrastive learning procedure: augmented views from the same molecular profile or spatial neighbors are considered positive samples for molecular or spatial embedding, respectively (Supplementary Fig. 1, Methods). Additionally, the contrastive learning strategy is also used to remove batch effects (Methods). After training, the disentangled embeddings can be applied to downstream tasks (Fig. 1). Specifically, the molecular identify embedding is suitable for omics-oriented tasks like cell typing, while the spatial context embedding can be applied to spatial-oriented tasks, such as spatial domain identification.

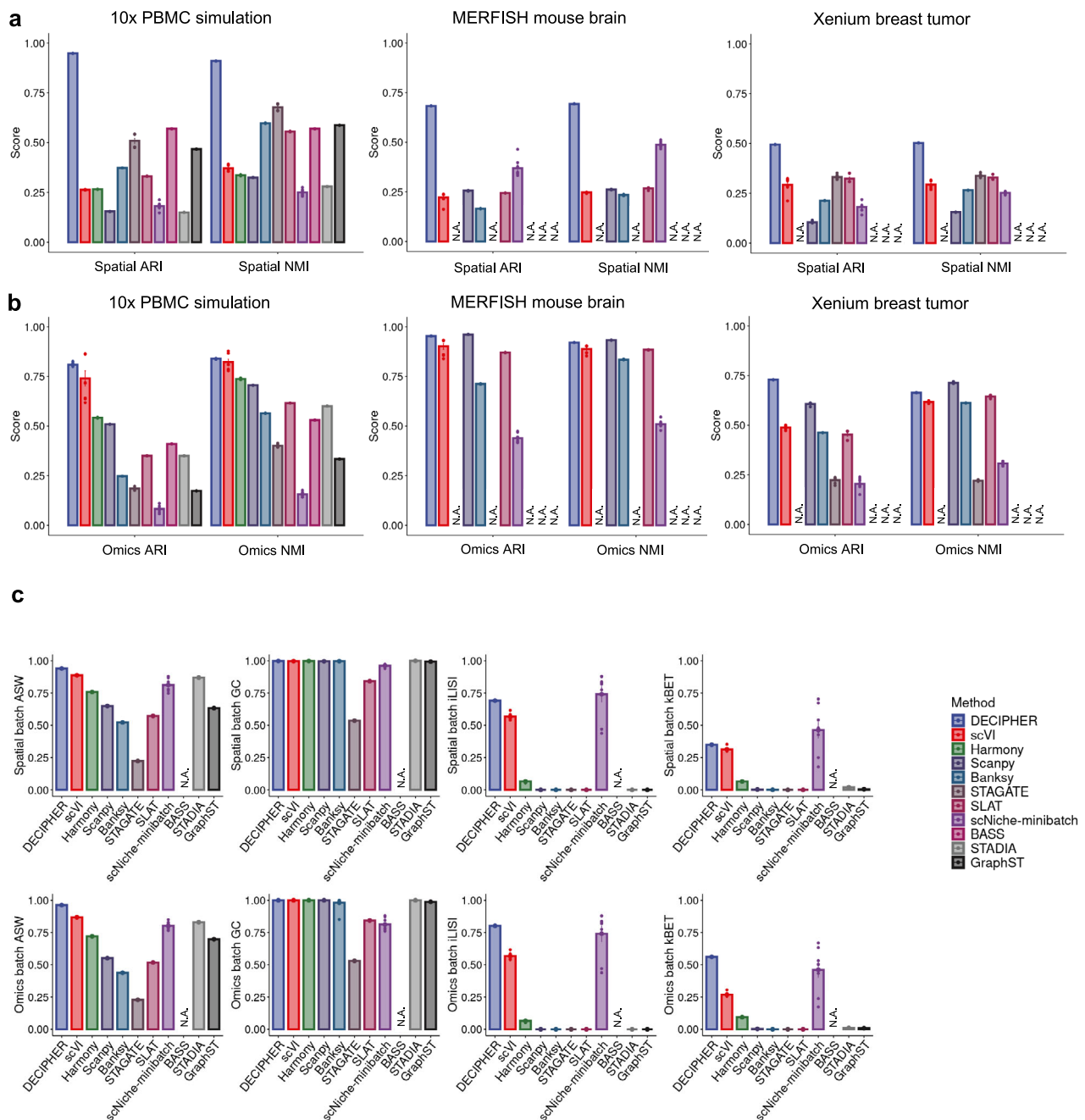
### Benchmarks show DECIPHER's superior performance

To evaluate the quality of the resulting embeddings, we first compared them with several state-of-the-art methods (Banksy<sup>5</sup>, STAGATE<sup>4</sup>, SLAT<sup>3</sup>, scNiche<sup>9</sup>, GraphST<sup>10</sup>, BASS<sup>8</sup> for spatial data modeling; scVI<sup>6</sup>, Scanpy<sup>15</sup> for single cell modeling and integration, STADIA<sup>11</sup> and Harmony<sup>16</sup> for spatial and single cell data batch correction). The selected methods adopt diverse strategies to incorporate spatial information: STAGATE, SLAT, GraphST and scNiche are based on GNN; Banksy uses an azimuthal Gabor filter to weight neighbor expression; BASS utilizes a Bayesian hierarchical model. Only scNiche requires cell type

annotation as input, while all other methods operate in a fully unsupervised manner. Notably, a key distinction of DECIPHER is its simultaneous optimization of disentangled embeddings, in contrast to other approaches that generate a single holistic embedding. We applied these methods to two well-curated single cell resolution spatial datasets (Supplementary Data 1): the Xenium breast cancer dataset<sup>17</sup> (single slice, 164,079 cells with 313 genes), and the MERFISH brain dataset<sup>18</sup> (31 slices, 378,918 cells with 374 genes). We also used two 10x PBMC datasets to simulate a synthetic spatial dataset (2 slices, 8405 cells with 31,915 genes) with unambiguous ground truths for cell types, spatial domains and batch effects (Methods). Harmony and STADIA are designed for single cell or spatial data batch correction. They require at least two batches and not applicable to Xenium and MERFISH datasets.

For spatial-oriented benchmarks, we quantify the consistency of embeddings with ground-truth spatial region annotations via normalized mutual information (NMI) and adjusted rand index (ARI). DECIPHER's spatial context embedding outperformed others across all three datasets (Fig. 2a and Supplementary Fig. 2a), fully demonstrating the advantages of its disentangled design. STAGATE is a representative GNN-based model and ranked second in both the simulation and Xenium dataset. In contrast, another GNN-based model, scNiche-minibatch, performed well only on the MERFISH dataset. Interestingly, the simplest GNN model, SLAT, which contains no learnable parameter, achieved overall best performance among GNN-based models (Fig. 2a and Supplementary Fig. 2a). These results suggested that no single GNN-based model consistently outperforms others across all datasets.

We found that lack of scalability prevents the application of several methods on real datasets, even with abundant computational resource (A100-80GB GPU and 64 CPU cores equipped with 256GB RAM). For example, although GraphST and BASS are top-ranked in 10x PBMC simulation dataset (BASS is 2nd in spatial ARI, GraphST is 3rd in spatial NMI), they both failed in larger Xenium and MERFISH datasets: GraphST failed due to memory overflow (capping at 256 GB) while



**Fig. 2 | Benchmarking of DECIPHER and state-of-the-art methods.** **a** Spatial-oriented benchmark results, in which NMI and ARI were reported against the original spatial region annotations. From left to right are results on the 10x mimic dataset, the MERFISH brain dataset and the Xenium breast tumor dataset. The UMAP visualization of spatial regions are shown in Supplementary Fig. 3. **b** Omics-oriented benchmark results, in which NMI and ARI are reported against the original cell type annotations. From left to right are the results on the 10x mimic dataset, the MERFISH brain dataset and the Xenium breast tumor dataset. STAGATE failed on MERFISH brain dataset because of GPU memory overflow (capping at 80 GB). GraphST failed on MERFISH brain and Xenium breast dataset because of memory overflow (capping at 256 GB). BASS failed on MERFISH brain and Xenium breast dataset because of overtime (capping at 72 h). Harmony and STADIA are not

applicable to MERFISH brain and Xenium breast datasets which do not contain batch effects. The UMAP visualization cell types are shown in Supplementary Fig. 4. **c** Benchmark results for batch correction on 10x simulation dataset. batch ASW, batch GC, batch iLISI, and batch kBET were calculated against spatial region annotations (first row) and cell type annotations (second row), respectively. The UMAP visualization of batches were shown in Supplementary Fig. 6.  $n = 8$  repeats with different random seeds. The error bars indicate the means  $\pm$  s.d. Raw data was provided as Supplementary Data 6. These methods can be mainly divided into three groups: spatial modeling (DECIPHER, Banksy, STAGATE, SLAT, scNiche, BASS), single cell modeling/integration (scVI, Scanpy, Harmony), and spatial batch correction (STADIA, GraphST). Source data are provided as a Source Data file.

BASS did not finish within a reasonable time limit (72 h), consistent with an independent benchmark work<sup>19</sup>. STAGATE failed in MERFISH datasets due to GPU memory overflow. The vanilla scNiche failed in Xenium and MERFISH datasets due to GPU memory overflow, we had to run a subsampling-based version by adopting the “subgraph-based

batch training strategy” suggested by the original authors (denoted as ‘scNiche-minibatch’, *Methods*) for all benchmarks.

In the simulation dataset, cell types and spatial domains are designed to be independent (*Method*, Supplementary Data 2). Under the setting, spatial-aware methods outperformed non-spatial methods

overall (Fig. 2a, left panel). However, in the Xenium breast tumor dataset where cell types and spatial domains are interrelated (Supplementary Data 2), a non-spatial method (e.g. scVI) performed comparably to spatially-aware methods. Notably, DECIPHER consistently achieved the best performance (Fig. 2a, right panel).

Existing spatial omics embedding models generally exhibit subpar performance in omics-oriented metrics (Fig. 2b, Supplementary Fig. 2a), especially two GNN-based models STAGATE and scNiche-minibatch, likely due to resolution degradation from the holistic embedding strategy (Supplementary Figs. 3–4). DECIPHER avoids this issue by its disentangled design (Fig. 1). We also computed spatial-oriented metrics using the molecular identity embeddings (and vice versa) as a negative control, and found significant performance decline (Supplementary Fig. 5). This suggests each embedding captures distinct and complementary aspects of spatial omics data as designed.

We found that DECIPHER showed superior batch correction performance (Fig. 2c, Supplementary Fig. 2b and Supplementary Fig. 6, *Methods*). The subsampling-based scNiche (scNiche-minibatch) ranked second-best, coming at the cost of declined performance for both cell typing and spatial region annotation, which is consistent with the intrinsic information loss during graph subsampling. GraphST and STADIA outperformed non-batch correction spatial methods (such as SLAT, STAGATE and Banksy), but still worse than DECIPHER and single cell batch correction methods scVI, and Harmony (Fig. 2c).

### The disentangled embeddings enable delineating cell-environment interaction across multiple scale

Cell-environment interactions often orchestrate cell localization mediated by ligand-receptors (LR) pairs, and further contribute to organismal development and homeostasis<sup>20</sup>. The disentangled embeddings produced by DECIPHER enable direct modeling of such sophisticated interactions which is rather challenging with canonical holistic embedding-based approaches. We focus on identifying key LR pairs or genes that contribute to the spatial localization of cells. The key idea is to learn a mask from a cell's molecular identity embedding to select specific LR/genes that can reconstruct the edge in kNN graph based on spatial context embedding (Fig. 3a, b, *Methods*). In other words, the selected LR/genes are associated with the spatial environment of the cell.

Firstly, we used a lymph node Xenium 5k dataset to investigate key cell-cell communication (CCC) molecules contributing to the localization of maturing B cells within GCs (Fig. 3c, Supplementary Fig. 7a, b, *Methods*). It is well established that the LR pairs CXCL12\_CXCR4 and CXCL13\_CXCR5 play critical roles in B cell maturation and localization<sup>21–24</sup>: CXCL12\_CXCR4 is essential for localizing follicular B cells within the dark zone, whereas CXCL13\_CXCR5 guides the migration of dark zone B cells towards the light zone (Fig. 3d, e and Supplementary Fig. 7c). DECIPHER successfully identified CXCL12\_CXCR4 and CXCL13\_CXCR5 as the top two most important LR pairs (Fig. 3f). In comparison, neither CXCL12\_CXCR4 nor CXCL13\_CXCR5 was ranked at the top by NicheNet<sup>12</sup> (Fig. 2g). CellChatV2<sup>13</sup> (Fig. 2h) and simple DEGs (differentially expressed genes) via canonical holistic embedding-based methods consistently missed CXCL13\_CXCR5 (Supplementary Fig. 7d, e). Considering that CellChat, NicheNet and DEG baselines focus more on genes with larger differences between groups by design, we speculated that the small difference caused by low expression level of CXCL13\_CXCR5 led to it being overlooked (Supplementary Fig. 7f, g). In contrast, by identifying LR pairs most correlated with cells' spatial embeddings, DECIPHER is insensitive to gene expression levels. Meanwhile, we also applied DECIPHER to a well-explored human skin dataset, and found that it accurately identified the important LR pairs GAS6-TYRO3 and PROS1-TYRO3, consistent with the results of NicheNet (Supplementary Fig. 8).

For spatial technologies with limited gene detectability, ligand and receptor pairs could hardly be detected simultaneously (Fig. 3i).

DECIPHER's high-fidelity embeddings empower calling key localization-related genes in such case. As demonstrated in a breast cancer Xenium dataset covering only 313 genes, lipocalin-type prostaglandin D2 synthase (*PTGDS*) emerged as the most relevant gene for the localization of both T and B cells (Fig. 3j). T and B cells expressing high levels of *PTGDS* were either enriched at the tumor interface or found infiltrating the tumor (Fig. 3k). A previous breast cancer study using IMC imaging independently identified *PTGDS* as a crucial marker for infiltrating lymphocytes (CD4+/CD8+/CD19+)<sup>25</sup>, confirming our findings. Among the remaining top-ranked genes, *CXCL12*<sup>26</sup>, *CD86*<sup>27</sup>, *SFPR4*<sup>28</sup> and *CXCR4*<sup>29</sup> have also been linked to lymphocyte activation and recruitment (Supplementary Fig. 9a, b). On the contrary, differential expression analysis based on canonical holistic embeddings (SLAT, Banksy) failed to identify *PTGDS* and others (Supplementary Fig. 9c–f). Additionally, CCC methods (NicheNet, CellChatv2) failed due to the limited number of LR pairs detected in this dataset (Fig. 3i).

### Profiling of atlas-level spatial data via DECIPHER

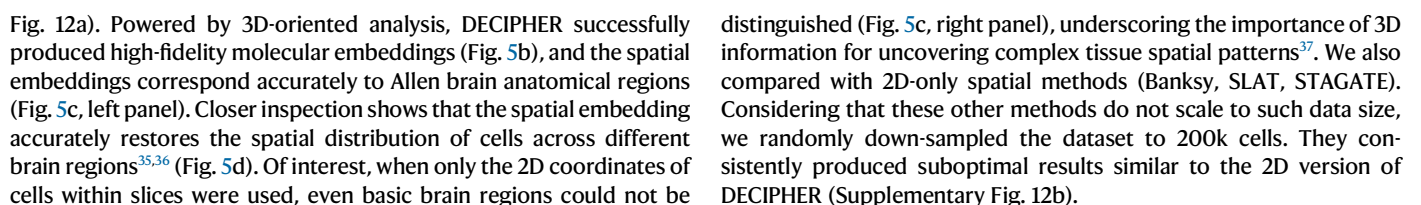
Technological advances have enabled the generation of heterogeneous spatial omics atlas exceeding millions of cells<sup>30</sup>, raising significant challenge to the scalability of computational methods. DECIPHER's advantages in both performance and scalability facilitate global analysis of such complex data. For instance, the human pancreatic spatial atlas consists of 8.7 million cells spanning eight different cancer types<sup>31</sup> (Fig. 4a), with significant batch effects across slices. DECIPHER successfully corrected batch effects, distinguished cell types in the molecular identity embeddings, and produced high-quality spatial context embeddings (Fig. 4b, Supplementary Fig. 10a). Unlike regular tissues, spatial patterns in cancer lack distinct boundaries formed by specific cell types. Instead, they typically exhibit a continuous variation in cell type density (Supplementary Fig. 10b) reflecting the dynamic nature of tumor growth, invasion, and lymphocyte infiltration. DECIPHER captured such continuous cell type composition variation globally and identified three classes of niches (Fig. 4c): the tumor core (dominated by tumor cells), lymph-low infiltrated tumor, and lymph-high infiltrated tumor (where T cells and B cells are significantly colocalized with the formation of tertiary lymphoid structures, Supplementary Fig. 10c, d).

We further identified genes most associated with T-cell localization and identified *CCR7* and *CCL5* as the top two genes (Fig. 4d). T cells with high *CCR7* expression tend to reside in lymph-high infiltrated regions, consistent with its key function in naïve T-cell activation<sup>32</sup> (Fig. 4e). In contrast, T cells with high *CCL5* expression are more likely to infiltrate into the tumor core (Fig. 4e), and also exhibit high expression of immunosuppressive checkpoint genes including PD-1, LAG-3, and CTLA-4, which aligns with the “terminally exhausted T cells” (Tex terminal, characterized by PD-1+, LAG3+, CTLA4+, GZMA+, GLNY+, GZMB+ and IFNG+) as defined in a non-spatial pan-cancer T cell study<sup>33</sup> (Supplementary Fig. 10e). These findings are consistent with *CCL5*'s well-known role in promoting T-cell infiltration and exhaustion<sup>34</sup> in the TME. Our findings underscore a strong association between T cell state and their spatial localization in the TME.

Notably, even using a single GPU, the total running time of DECIPHER is under 4 h, and can be further accelerated through multi-GPU parallelization (Fig. 4f). In contrast, state-of-the-art spatial modeling methods not only failed to process the entire datasets but also produced suboptimal results on down-sampled data (Fig. 4g and Supplementary Fig. 11).

3D-oriented global analysis is essential for accurately modeling tissue architecture. DECIPHER natively supports 3D spatial data. Specifically, we used the mouse brain 3D atlas, which comprises 151 consecutive slices of a single brain with 3.5 million cells, depicting the complex central nervous system<sup>35</sup>. These slices were registered to 3D CCF coordinates by the original authors (Fig. 5a, Supplementary





**Fig. 3 | DECIPHER enables identifying localization-related LR pairs/genes.**

**a** Detailed illustration of the LR/gene selection model. This model uses a MLP with a Gumbel-sigmoid layer to generate a learnable LR/gene mask from the molecular identity embedding. Then the product of selected LR/genes of  $Cell_i$  and  $Cell_j$  were used to reconstruct the edge of  $k$ -NN graph from spatial embedding: if  $Cell_i$  and  $Cell_j$  were connected by an edge, the product is expected to be larger, otherwise, to be smaller (*Methods*).  $\odot$  means Hadamard product. **b** Illustration of LR activity calculation, which is the production of LR expression<sup>48</sup> (*Methods*) and an extra LR mask: if cell  $i$  expresses the receptor gene of one LR pair and cell  $i$ 's neighbors within 100  $\mu\text{m}$  express the corresponding ligand gene, the corresponding value in the mask is set to "1" (green); otherwise, it is set to "0" (red). **c** Spatial visualization of cell types on lymph node slices (left panel) and the marker of each cell type (right panel). **d** Schematic diagram illustrating the well-studied process of B cell maturation in germinal center, with the regulation of specific ligand-receptors<sup>23</sup>.

**e** Visualization of a typical GC; the dark zone and light zone were annotated by an anatomical expert. The experiment was repeated three times with similar results. The left panel shows the original H&E-stained image of the GC, and the right panel is colored by cell types. **f** Bar plot showing the top 10 LR pairs most important for B-cell localization calculated by the LR selection model. **g** NicheNet result on the lymph node dataset. **h** CellChatV2 result on the lymph node dataset. **i** Barplot showing the number detected and undetected LR pairs in Xenium (313 genes in total), Xenium 5k (4624 genes in total) and 10x v3 (31,915 genes in total) dataset, LR pairs information came from CellChatDB. **j** Bar plot showing the top 5 localization-related genes of T cells (left panel) and B cells (right panel) identified by the gene selection model in the breast cancer Xenium dataset. **k** Spatial expression of *PTGDS* in T cells (left panel) and B cells (right panel). Pink circles indicate the DICS tumor region annotated by the original author. Source data are provided as a Source Data file.

## Discussion

Intra- and extra-cellular factors in spatial biology are interconnected but should not be conflated. To model spatial data properly, we propose DECIPHER for learning high-fidelity, disentangled embeddings of molecular identity and spatial context, through a novel cross-scale contrastive learning framework. We demonstrated superior performance of DECIPHER in various benchmarks and real-world case studies. More importantly, DECIPHER's unique ability to quantify omics factors highly correlated with spatial locations enhances the understanding of intracellular and extracellular regulation.

Given the rapidly increasing throughput of new spatial technologies<sup>38,39</sup>, scalability becomes a critical challenge for existing method. For example, BASS, GraphST performed well on the 10x simulation dataset (Fig. 2), but failed to scale to larger datasets such as Xenium and MERFISH datasets, consistent with previous benchmark studies<sup>19</sup>. By employing a transformer-based spatial encoder, DECIPHER scales to datasets of arbitrary sizes. Additionally, DECIPHER only uses shallow networks and low embedding dimensions, as we found more layers and larger dimension brings marginal improvement. This lightweight architecture appears sufficient to capture the key features present in current spatial omics data. Additionally, DECIPHER supports multi-GPU parallelization for further acceleration.

DECIPHER's contrastive self-supervised learning strategy combined with the transformer architecture is well suited for large-scale heterogeneous spatial pretraining<sup>40</sup>. While current spatial data's limited quantity and quality (in terms of resolution, gene coverage and drop-out ratio) makes it more difficult to benefit from the "scaling law"<sup>41</sup>, we anticipate that with continued technological advances, high-quality spatial data covering entire transcriptome, entire organs or organisms will eventually emerge. Once such data become available, DECIPHER could serve as the backbone of pretrained spatial foundation model, providing a reliable starting point for addressing various biological questions involving cellular functions within native contexts.

We also quantitatively compared the divergence between the two embeddings by calculating the coefficient of determination ( $R^2$ ) (*Methods*). Interestingly, the  $R^2$  values appeared to be cell-type-specific (Supplementary Data 3). For instance, in the mouse brain 3D atlas dataset (Supplementary Fig. 13), higher  $R^2$  values were observed in certain neuron subtypes (e.g., IT-ET Glut, HY GABA, HY Glut), suggesting their conserved molecular-spatial profiles (Fig. 5d). In contrast, lower  $R^2$  values were found for OPC-oligodendrocyte, immune, and vascular cell types which are more broadly distributed across the brain.

We'd note that the current study has some limitations. Model-wise, DECIPHER only use random dropout as data augmentation, more diverse data augment approaches may improve the performance<sup>42</sup>. In terms of benchmarking, our evaluation only contains two technologies, Xenium and MERFISH. Although DECIPHER is broadly applicable to various single-cell resolution spatial technologies other than

Xenium and MERFISH, e.g., CosMx and CODEX, we could not incorporate them into the current benchmark partly due to the lack of high-quality annotation of cell types and spatial regions. Thus, more diverse and gold-standard spatial datasets are still required for the systematic evaluation of methods in the community.

In summary, DECIPHER is a high-performance, high-scalable tool which provides a new analytical perspective for exploring spatial omics data. To promote its application by the research community, the DECIPHER package, along with detailed tutorials, are available at <https://github.com/gao-lab/DECIPHER>.

## Methods

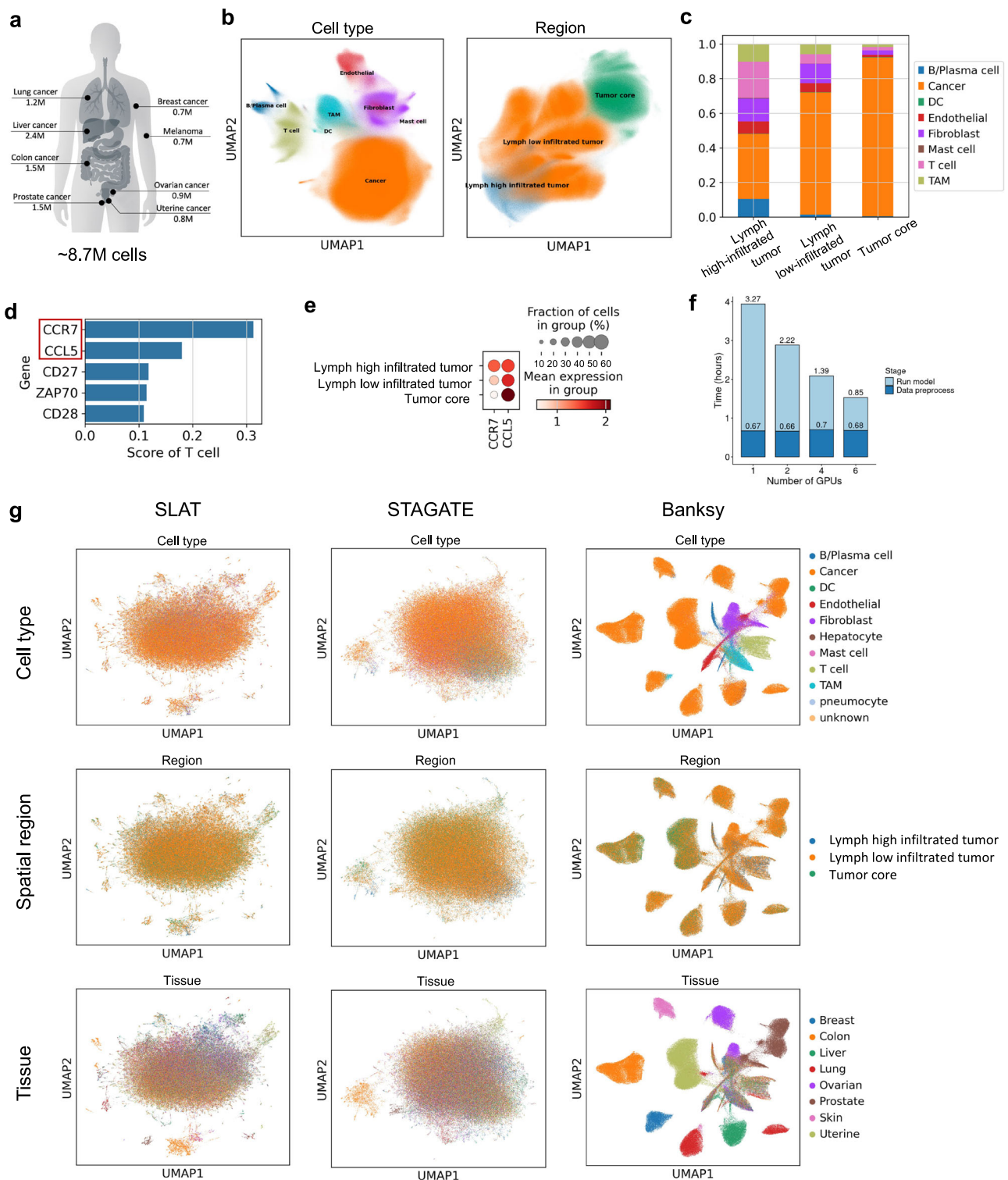
### Framework

DECIPHER requires one or multiple single cell spatial omics datasets consisting of the raw expression matrix and spatial coordinates of cells (supporting 2D or 3D coordinates). At the omics level, the raw expression matrix undergoes standard processing<sup>15</sup>, including log-based normalization, top 2000 highly variable gene selection (via 'seurat\_v3' mode of the scanpy.pp.highly\_variable\_genes function), and scaling. At the spatial level, for each cell  $i$ ,  $k$ -nearest neighbors in the spatial coordinate space are considered the spatial neighbors of cell  $i$ , denoted as  $\mathcal{N}_n$  (with  $k$  defaulting to 20).

The DECIPHER model is composed of an "omics encoder" and a "spatial encoder" (Fig. 1), which are used to learn intra-cellular molecular identity and extra-cellular spatial context embeddings, respectively. Permutation invariance is an important feature of spatial neighborhoods, due to the spatial neighbors of a cell not having a meaningful order. Self-attention mechanism suits spatial data modeling since it is naturally permutation invariant, i.e., each token can access any token through attention operation<sup>43</sup>. Thus, the "spatial encoder" adopts a transformer architecture<sup>14,44</sup> without the position embedding procedure (by default, 3 layers with 1 attention head). It takes the omics features of all spatial neighbors of  $cell_i$  (excluding  $cell_i$  itself) and a trainable token (CLS token) as input tokens. The "omics encoder" is a multilayer perceptron (MLP) that projects high-dimensional single-cell expression data into the molecular identity latent space (by default, 32 dimensions), denoted as  $z_i^{\text{omics}}$ . We do not use a Transformer-based architecture here because its high computational cost. To reduce computational complexity, the input tokens of spatial encoder are not high-dimensional expressions but rather latent representations learned by the omics encoder. The output embedding of the CLS token is ultimately used as the spatial context embedding  $z_i^{\text{spatial}}$  of  $cell_i$  (by default, 32 dimensions).

### Training process

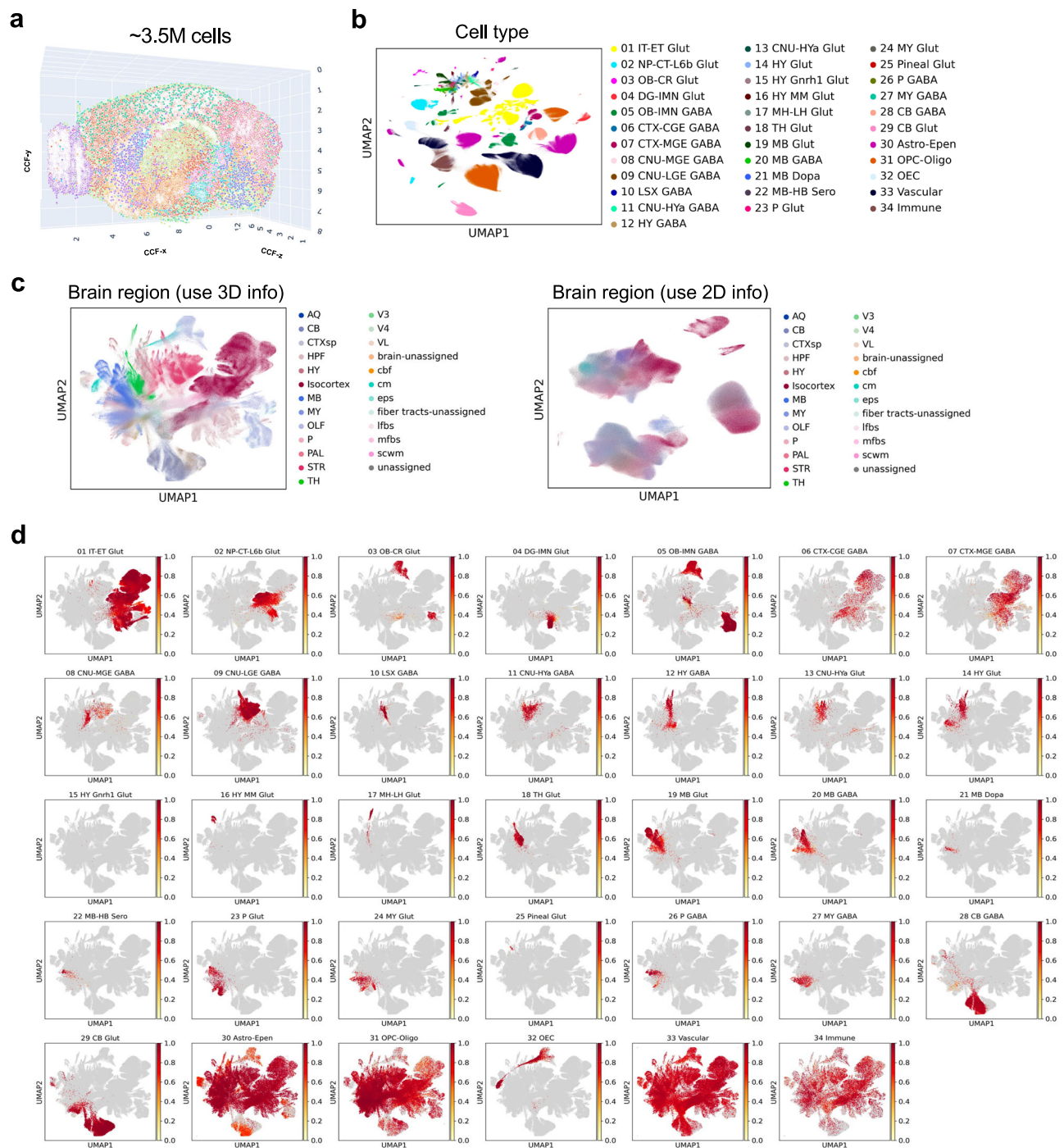
The DECIPHER model is optimized through novel cross-scale contrastive learning we proposed (Supplementary Fig. 1). Specifically, two contrastive losses from omics scale and spatial scale are used to optimized the whole model simultaneously:



**Fig. 4 | Profiling human pan-cancer spatial atlas.** **a** Diagram illustrating the cancer type and cell numbers of the human pan-cancer spatial atlas. **b** UMAP visualization of DECIPHER molecular identity embedding (left panel) and spatial context embedding (right panel) of the atlas. **c** Cell type composition in different spatial regions defined in (b). **d** Bar plot showing the top 5 localization-related genes of T cells in the human pan-cancer spatial atlas. **e** Expression of *CCR7* and *CCL5* in T cells located in different spatial regions defined in (b). **f** Time taken of DECIPHER

with different number of GPUs for the atlas. **g** Results of state-of-the-art spatial modeling methods on down-sampled human pan-cancer spatial atlas. UAMP visualizations of SLAT, STAGATE and Banksy embeddings on a randomly down-sampled human pan-cancer spatial atlas with 200,000 cells, colored by cell type (first row), spatial region (second row) and tissue source (third row), respectively. Source data are provided as a Source Data file.





**Fig. 5 | Profiling mouse 3D brain atlas. a** Diagram illustrating the mouse brain atlas in 3D CCF coordinates<sup>54</sup>. **b** UMAP visualization of DECIPHER molecular identity embeddings on the mouse brain atlas. **c** UMAP visualization of DECIPHER spatial context embeddings on the mouse brain atlas when using 3D coordinates (left panel) or only 2D coordinates (right panel). **d** Distribution density of every cell type on UMAP result based on DECIPHER spatial context embedding. Astro astrocyte, AQ aqueduct of Sylvius, CB cerebellum, CGE caudal ganglionic eminence, CNU cerebral nuclei, CR Cajal-Retzius, CT corticothalamic, CTX cerebral cortex, CTXsp cortical subplate, DG dentate gyrus, EA extended amygdala, Epen ependymal, EPI epithalamus, ET extratelencephalic, GC granule cell, HB hindbrain, HPF hippocampal formation, HY hypothalamus, HYA anterior hypothalamic, IMN immature

neurons, IT intratelencephalic, L6b layer 6b, LGE lateral ganglionic eminence, LH lateral habenula, LSX lateral septal complex, MB midbrain, MGE medial ganglionic eminence, MH medial habenula, MM medial mammillary nucleus, MY medulla, NP near-projecting, OB olfactory bulb, OEC olfactory ensheathing cells, OLF olfactory areas, Oligo oligodendrocytes, OPC oligodendrocyte precursor cells, P pons, PAL pallidum, STR striatum, TH thalamus, V3 third ventricle, V4 fourth ventricle, VL ventral lateral nucleus, CM centromedian nucleus, EPS entorhinal-perirhinal-solitary complex, IFBS internal fiber bundles, MFBS medial forebrain bundles, SCWM sub-cortical white matter. Neurotransmitter types: DOPA dopaminergic, GABA GABAergic, Glut glutamatergic, Nora noradrenergic, Sero serotonergic<sup>36</sup>.



For each training batch, we randomly selected  $n$  cells according to the batch size (defaulting to 256). Once a cell is sampled as the central cell, all its spatial neighbors are also loaded. Then we augmented all sampled cells in both omics and spatial terms. In omics terms, we simulated dropout events to obtain augmented views of single-cell omics data. Specifically, for each cell, we randomly mask genes twice on the basis of probability  $p$  (defaulting to 0.6) to obtain two different “dropout views” of the same cell. NT-Xent loss<sup>45</sup> (Formula 1) is used to move the two dropout views of the same cell close to each other, while views of different cells are pushed apart<sup>45</sup>.

$$\ell_{omics} = -\log \frac{\exp\left(\frac{\text{sim}(z_m, z_n)}{\tau}\right)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq m]} \exp\left(\frac{\text{sim}(z_m, z_k)}{\tau}\right)} \quad (1)$$

Here,  $z_m$  and  $z_n$  are two views of the same cell,  $z_m$  and  $z_k$  are two views of different cells,  $N$  is the batch size, and  $\tau$  is the temperature coefficient. Data augmentation occurs only during training. The complete omics feature of each cell is used directly as the input in the inference stage. Additionally, to handle batch effects across multiple slices, we include mutual nearest neighbor (MNN) pairs across different batches as extra positive pairs in addition to the positive pairs obtained through the above data augmentation strategy.

In spatial terms, since each neighbor of cell  $i$  has two “dropout views”, we naturally obtain two views of the spatial context embedding of cell  $i$  via the spatial encoder (Supplementary Fig. 1). We used another NT-Xent loss (same as Formula 1) term to encourage positive spatial context embedding pairs to be close to each other and to push negative pairs apart.

Finally, the total loss of the model is the weighted average of the contrastive losses at the omics and spatial scale:

$$\ell = \alpha * \ell_{omics} + (1 - \alpha) \ell_{spatial} \quad (2)$$

where  $\alpha$  is 0.5 by default. According to our evaluation, the default setting of  $\alpha = 0.5$  achieves the best balance in gene expression and spatial context (Supplementary Data 4). It is expected because the omics encoder and spatial encoder were optimized synergistically through  $\ell_{omics}$  and  $\ell_{spatial}$ . The imbalanced weight of the two losses may desynchronize these two components.

DECIPHER is built on the PyTorch framework and supports multi-GPU parallelization. By default, DECIPHER first warms up the omics encoder for 3 epochs with a learning rate of 0.01 and then trains the omics encoder and spatial encoder synergistically for 5 epochs or 10,000 steps with a decayed learning rate of 0.00001.

We tested DECIPHER with different hyperparameters in the 10x PBMC simulation dataset, including (1) dropout ratio, (2) the latent space dimension, (3) the layer of the transformer, (4) the number of attention heads, (5) the neighbor size  $k$ , and (6) the loss weight  $\alpha$ ; the results are reported in Supplementary Data 4.

### Localization-related genes/LR pairs selection model

For the gene selection model, we employed a parameterized MLP to quantify the importance of each gene to the of cell localization. The input and output dimension of MLP were the same as molecular embedding and HVG-filtered expression matrix, respectively. It also contained a hidden layer with dimension of 256. The output was then differentially binarized as a mask vector through Gumbel-sigmoid sampling<sup>46</sup>. Then mask vector was subsequently applied to the gene expression of cell to obtain the selected genes only:

$$\text{mask} = \text{Gumbel}\left(\text{MLP}\left(z_i^{\text{spatial}}\right)\right) \text{selected}_i = \text{expression} \odot \text{mask} \quad (3)$$

Thus,  $\text{selected}_i$  contains only the selected gene set  $S$  of cell  $i$  (Fig. 3a).  $\odot$  means the Hadamard product.

The training objective is to optimize the selected gene set  $S$  to identify genes most associated with cell localization (spatial context embedding). We framed this problem as the reconstruction of the spatial context embedding similarity graph and constructed a  $k$ -nearest neighbors graph  $G_{\text{spatial}}$  on the basis of spatial context embedding. During training, we used the selected gene set  $S$  to reconstruct the edges of  $G_{\text{spatial}}$  (Fig. 3a) following the graph auto-encoder loss<sup>47</sup>: if  $\text{Cell}_i$  and  $\text{Cell}_j$  were connected in the kNN graph, the product of  $\text{selected}_i$  and  $\text{selected}_j$  should be larger, and vice versa. The loss function is the negative log-likelihood of edge reconstruction:

$$\ell_g = -\sum_{i=1}^N \sum_{j=1}^N \left( y_{ij} \log \hat{y}_{\theta, ij} + (1 - y_{ij}) \log (1 - \hat{y}_{\theta, ij}) \right) \quad (4)$$

In the above equation,  $y_{ij}$  indicates whether there is an edge between cell  $i$  and cell  $j$  in  $G_{\text{spatial}}$ .  $\hat{y}_{\theta, ij}$  is the predicted value. Additionally, to filter out irrelevant genes, we added an extra  $L_1$  regularization to the size of  $S$ . The overall loss is written as:

$$\ell = \ell_g + \lambda |S|_1 \quad (5)$$

$\lambda$  is set to 1 by default. During training, the model applies back propagation to learn the gene mask, gradually filtering out genes that do not contribute to predicting the spatial context embedding. By default, the model is trained for 300 epochs with a learning rate of 0.01. After training is complete, the resulting  $S_i$  represents the key genes for the spatial positioning of single cell  $i$ . As for a group interested cells (e.g., cell types, cells in specific niches, or any custom groups), we determined the localization-related score of gene  $g$  for these interested cells as the frequency it is selected:

$$\text{score}_g = \frac{X_g}{N} \quad (6)$$

where  $N$  is number of cells,  $X_g$  is the number of cells select the gene  $g$ .

The ligand-receptor selection model uses the same framework, but instead of genes, the basic units are LR pairs (Fig. 3b). The activity of each LR pair in cell  $i$  ( $e_i^{\text{lr}}$ ) is defined as the product of expression between the receptor genes  $e^{\text{receptor}}$  in cell  $i$  and the sum of the ligand gene expression  $e^{\text{ligand}}$ <sup>48</sup> within the 100  $\mu\text{m}$  radius (a typical cell-cell interaction distance<sup>49</sup>), denoted as  $\mathcal{N}_i$ :

$$e_i^{\text{lr}} = e_i^{\text{receptor}} * \sum_{j \in \mathcal{N}_i} e_j^{\text{ligand}} \quad (7)$$

### Coefficient of determination between embeddings

To quantitatively compare the divergence between the molecular identity and spatial context embedding spaces, we used the coefficient of determination ( $R^2$ ). Inspired by a previous work<sup>50</sup>, we split cells into training (80%) and testing (20%) subsets. An MLP (1 hidden layer with dimension of 256) was trained to predict the spatial context embedding from the molecular identity embedding. The MLP was trained for 100 epoches using mean squared error loss, with a learning rate of  $1e-3$ .  $R^2$  values were then calculated on the held-out testing set as following:

$$R^2 = 1 - \frac{\sum_i (y_i - y_{\text{pred}})^2}{\sum_i (y_i - y_{\text{mean}})^2} \quad (8)$$

where  $y_{\text{pred}}$  is the predicted value and  $y_{\text{mean}}$  is the mean of actual value.

### Benchmark

**Public datasets.** The MERFISH brain dataset<sup>18</sup> and Xenium breast cancer dataset<sup>17</sup> are publicly available. The MERFISH brain dataset

contains 378,918 cells and 374 genes, and the Xenium breast tumor dataset contains 164,079 cells and 313 genes. We used the cell type and spatial domain annotations from the original publications.

**Simulation datasets.** We simulated spatial slices in which the ground truths of the cell types and spatial patterns were clear. Specifically, we selected one  $10 \times 5'$  (named dataset 1) and one  $10 \times 3'$  (named dataset 2) based PBMC dataset from the official 10x datasets from the same donor. These two datasets contain technical batch effects (Supplementary Fig. 14a). For each dataset, we annotated cell types via the same criteria, retaining only three distinct cell types: T cells (5278 cells, marked by CD3D), B cells (1068 cells, marked by CD79A), and monocytes (2059 cells, marked by CD14 and CD16) (Supplementary Fig. 14b). The dataset 1 and 2 contained 4512 cells and 3893 cells (8405 in total), and 31,915 shared genes.

Then we used dataset 1 and dataset 2 to simulate the slice 1 and slice 2 following same procedure. In detail, three identical spatial patterns were simulated in each slice (Supplementary Fig. 14c):

1. Pattern 1: Monocytes and T cells randomly mixed (x-axis boundary is [0, 0.5], y-axis boundary is [0.5, 1]).
2. Pattern 2: Only contains T cells (x-axis boundary is [0.5, 1], y-axis boundary is [0, 1]).
3. Pattern 3: B cells and T cells randomly mixed (x-axis boundary is [0, 0.5], y-axis boundary is [0, 0.5]).

Thus, the two datasets are homogeneous in terms of cell types and spatial patterns. Numbers of each cell type in each pattern are kept in Supplementary Data 2.

## Methods

The benchmark methods Scanpy, Harmony, scVI, Banksy, STAGATE, and SLAT were executed via the Python packages “Scanpy” (v1.10.1), “harmonypy” (v0.0.6), “scvi-tools” (v1.1.2), “Banksy\_py” (last commit 10770c2), “STAGATE\_pyG” (latest commit 8b9c8ef), “scNiche” (v1.1.0), “GraphST” (v1.0.0), and “scSLAT” (v0.2.1), respectively, in Python (v3.11). Methods STADIA and BASS were executed via the R packages “STADIA” (v1.0.1) and “BASS” (latest commit 7121c89), respectively, in R (v4.4.1). Each method was allocated 8 CPU cores (Intel Xeon Platinum 8358), 256 GB of memory, and one GPU (A100-80G) and 48 h for running. For all benchmark methods, including DECIPHER, we used the default parameters and original data preprocessing steps. For scNiche, we adopted the “subgraph-based batch training strategy” proposed by the original authors, denoted as “scNiche-minibatch”, because the default training strategy failed to scale to the Xenium and MERFISH datasets. Additionally, in the simulation dataset where the default scNiche strategy did run successfully, it exhibited embedding collapse, with more than 80% cells/spots allocated the same latent embedding (Supplementary Fig. 15). A possible cause is the insufficient mutual information between different views (e.g., gene expression and spatial neighborhoods) in scNiche’s multi-view consensus strategy in the simulated dataset, where cell types within each “region” were randomly distributed.

## Metrics

For identity and context embedding, we used two common metrics: the adjusted rand index (ARI) and normalized mutual information (NMI). For batch effect removal, we used batch ASW, batch GC, batch iLSI, and batch kBET. We also recorded the run time and memory usage of each method (Supplementary Data 5). The ‘overall score’ was the average of each metric included in the current benchmark panel<sup>51</sup>. All metrics were calculated following the standard procedures in scIB<sup>51</sup>:

### NMI

NMI compares the overlap of two clusterings. The overlap was scaled using the mean of the entropy terms for predicted and true cluster

labels. NMI scores of 0 or 1 correspond to uncorrelated clustering or a perfect match, respectively.

### ARI

The Rand index compares the overlap of two clusterings, considering both agreements in cluster assignments and agreements in cluster separations. The adjustment of the Rand index corrects for random agreements. An ARI of 0 or 1 corresponds to random labeling or a perfect match, respectively.

### Batch ASW

For the batch mixing, we consider the absolute silhouette width on batch labels per cell. The silhouette width measures the relationship between the within-cluster distances of a cell and the between-cluster distances of that cell to the closest cluster. Batch ASW of 1 represents ideal batch mixing and a value of 0 indicates strongly separated batches.

### Batch GC

The batch graph connectivity (GC) metric assesses whether the kNN graph of the integrated data directly connects all cells with the same label. The resultant score has a range of [0;1], where 1 indicates that all cells with the same labels are connected in the integrated kNN graph and the lowest possible score indicates a graph where no cell is connected.

### Batch kBET

The kBET algorithm determines whether the label composition of a  $k$ -nearest neighborhood of a cell is similar to the expected (global) label composition. The test is repeated for random subsets of cells, and the results are summarized as a rejection rate over all tested neighborhoods.

### Batch LISI

LISI scores are computed from neighborhood lists per node from integrated kNN graphs. Specifically, the inverse Simpson’s index is used to determine the number of cells that can be drawn from a neighbor list before one batch is observed twice. Thus, LISI scores range from 1 to  $N$ , where  $N$  is the total number of batches in the dataset.

## Calculation procedure of metrics

Spatial and omics clustering metrics (NMI and ARI): for DECIPHER, we clustered the molecular and spatial embeddings and get the molecular cluster and spatial cluster. Then we calculated ARI, NMI between molecular cluster against ground truth cell type (provided by original authors), as well as spatial cluster against ground truth spatial domain (provided by original authors). For rest holistic embedding methods (STAGATE, Banksy, SLAT, scVI, Harmony, Scanpy, STADIA, scNiche and GraphST), we clustered the embedding and get the holistic cluster. Then we calculated ARI, NMI between the holistic cluster against ground truth cell type (provided by original authors) and spatial domain (provided by original authors). The output of BASS was the cell cluster and spatial cluster labels, thus we directly calculated metrics (ARI, NMI) between cell cluster against ground truth cell type (provided by original authors), as well as spatial cluster against ground truth spatial domain (provided by original authors).

Batch correction metrics (batch ASW, batch GC, batch iLSI, batch kBET): we calculated these metrics following similar manner in calculating spatial and omics clustering metrics. Notably, BASS only provide the cluster labels rather than embedding, thus the batch correction metrics were not applicable.

## Case studies

**Identification of key CCC molecules contributing to cellular localization in the Xenium 5k data.** The Xenium 5k lymph node dataset includes 708,983 cells and 4624 elaborately selected genes<sup>52</sup>, covering

important cell-cell communication and signaling pathways. The corresponding cell segmentation and H&E-stained images were provided by 10x Genomics. First, we annotated the major cell types in the Xenium 5k dataset according to the following classical markers (Fig. 3c): T cells (*CD3E*), dendritic cells (DCs, *ITGAX*), B cells (*CD79A*), endothelial cells (*PLVAP*), macrophages (*MMP9*), plasma cells (*MZB1*), follicular dendritic cells (fDCs, *CR2*), plasmacytoid dendritic cells (pDCs, *IRF7*), and vascular smooth muscle cells (VSMCs, *NOTCH3*). A group of cells with significantly lower counts were annotated as “low\_quality” (Supplementary Fig. 7a). We also annotated the B cells into three spatial-related groups (Follicular B cell, Dark zone B cell, and Light zone B cell) based on (Fig. 3e). Ligand-receptor data is sourced from the latest version of CellChatDB<sup>48</sup> (<http://www.cellchat.org/cellchatdb/>), which contains 3124 LR pairs, among which 1078 pairs were detected in this dataset.

DECIPHER was run with the default parameters (Supplementary Fig. 7b). We then ran LR selection model for all B cells and found CXCL13\_CXCR5 and CXCL12\_CXCR4 were two most import LR pairs (Fig. 3f). We further checked the paired H&E image to annotated the light zone and dark zone in a typical GC by a pathologist, then visualized CXCL13\_CXCR5 activity and CXCL12\_CXCR4 activity, and cell types in the H&E image (Supplementary Fig. 7c). We also compared the averaged ratio of CXCL13\_CXCR5 to CXCL12\_CXCR4 in B cells which located in different regions: follicular B cell, dark zone B cell and light zone B cell (Supplementary Fig. 7c).

We compared CCC methods and holistic spatial embedding methods. As for CCC methods, we ran NicheNet<sup>12</sup> and CellChatV2<sup>13</sup> following their standard protocols with default parameters. Neither CXCL12\_CXCR4 nor CXCL13\_CXCR5 was among the top-ranking LR pairs by NicheNet (Fig. 3g). CellChatV2 missed the CXCL13\_CXCR5 signal (Fig. 3h). As for holistic spatial embedding methods, we clustered their embeddings (Banksy, SLAT) to get spatial domains and then identified differentially expressed LR pairs between the GC and non-GC regions using the “scanpy.pp.rank\_genes\_groups()” function (activity of each LR pair is calculated via Formula 7). They both missed the CXCL13\_CXCR5 signal (Supplementary Fig. 7d, e). STAGATE failed in this dataset (capping at 1024 GB system memory and 80 GB GPU memory).

**Identification of key localization-related genes in the Xenium breast tumor dataset.** The Xenium breast tumor dataset contains 164,079 cells and 313 genes and was annotated by the original authors. We ran DECIPHER on this dataset with default parameters. Based on DECIPHER embedding, the gene selection model was run separately for T cells and B cells. For each cell type, genes were ranked according to their localization-related scores and the top 5 ranked genes are presented in Fig. 3j.

We compared CCC methods and embedding methods. Notably, we were unable to run CCC methods (CellChatV2, and NicheNet) on this dataset because only 313 genes were captured and did not contain enough paired ligand-receptor genes for analysis (Fig. 3i). For embedding methods, we clustered the embeddings from Banksy and SLAT to obtain spatial domains (Supplementary Fig. 9c, d) and then identified DEGs between tumor domains and non-tumor domains using the “scanpy.tl.rank\_genes\_groups()” function (Supplementary Fig. 9e, f).

**Human pan-cancer spatial atlas profiling.** The human pan-cancer spatial atlas includes 16 slices from 8 cancer types (2 colon cancer, 2 liver cancer, 2 melanoma, 4 ovarian cancer, 2 prostate cancer, 2 lung cancer, 1 breast cancer, and 1 uterine cancer), containing a total of 8,696,580 cells and 500 genes, analyzed using the MERSCOPE platform. We manually annotated the major cell types in the dataset via following classical markers: T cell (*CD3D*), B cell (*CD79A*), TAM (*CIQC*), DC (*CD207*), MAST (*KIT*), and cancer cell (*VCAM1*, *MKI67*).

Owing to the significant batch effects between slices, we used the batch effect removal option in DECIPHER (as mentioned above). The

runtime included both data processing and algorithm execution, which was performed on a server equipped with dual Intel Platinum 8358 CPUs, 8 A100-80G GPUs and 1024 GB system memory. Notably, Other spatial modeling methods (SLAT, STAGATE, Banksy) failed to process the entire dataset on the same server. Thus, we randomly down-sampled the dataset to 200,000 cells (the maximum size for STAGATE on our server) and subsequently ran each model with its default parameters (Fig. 4g).

**Mouse brain 3D atlas profiling.** The mouse brain atlas consists of ~3.5 M quality-controlled cells with 1122 detected genes, derived from 151 consecutive MERFISH slices of a single adult mouse brain<sup>35</sup>, including detailed annotations of cell types and spatial domains. These slices are also registered to 3D CCF coordinates by authors (Supplementary Fig. 12a).

We trained DECIPHER using 3D CCF coordinates or only 2D coordinates inside slices (Fig. 5b, c). Other spatial modeling methods (SLAT, STAGATE, Banksy) failed to process the entire dataset (capping at 1024 GB system memory, 80 GB GPU memory). Thus, we randomly down-sampled the dataset to 200,000 cells (the maximum size for STAGATE on our server) and subsequently ran each model with its default parameters (Supplementary Fig. 12b).

## Reporting summary

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

## Data availability

All datasets used in this study have already been published and were obtained from public data repositories. Supplementary Data 1 provides the associated publications and download URLs. Source data are provided with this paper.

## Code availability

DECIPHER package, along with code and detailed tutorials for reproducibility, is available at <https://github.com/gao-lab/DECIPHER> under MIT license, and in Zenodo (<https://zenodo.org/records/15860235>)<sup>53</sup>.

## References

- Lewis, S. M. et al. Spatial omics and multiplexed imaging to explore cancer biology. *Nat. Methods* **18**, 997–1012 (2021).
- Wang, X. Q. et al. Spatial predictors of immunotherapy response in triple-negative breast cancer. *Nature* **621**, 868–876 (2023).
- Xia, C.-R., Cao, Z.-J., Tu, X.-M. & Gao, G. Spatial-linked alignment tool (SLAT) for aligning heterogenous slices. *Nat. Commun.* **14**, 1–12 (2023).
- Dong, K. & Zhang, S. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nat. Commun.* **13**, 1739 (2022).
- Singhal, V. et al. BANKSY unifies cell typing and tissue domain segmentation for scalable spatial omics data analysis. *Nat. Genet.* **56**, 431–441 (2024).
- Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).
- Cui, H. et al. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nat. Methods* **21**, 1470–1480 (2024).
- Li, Z. & Zhou, X. BASS: multi-scale and multi-sample analysis enables accurate cell type clustering and spatial domain detection in spatial transcriptomic studies. *Genome Biol.* **23**, 168 (2022).
- Qian, J. et al. Identification and characterization of cell niches in tissue from spatial omics data at single-cell resolution. *Nat. Commun.* **16**, 1693 (2025).



10. Long, Y. et al. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nat. Commun.* **14**, 1155 (2023).
11. Li, Y. & Zhang, S. Statistical batch-aware embedded integration, dimension reduction, and alignment for spatial transcriptomics. *Bioinformatics* **40**, btae611 (2024).
12. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling inter-cellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020).
13. Jin, S., Plikus, M. V. & Nie, Q. CellChat for systematic analysis of cell–cell communication from single-cell transcriptomics. *Nat. Protoc.* 1–40 <https://doi.org/10.1038/s41596-024-01045-4> (2024).
14. Vaswani, A. et al. Attention is all you need. *Adv. Neural Inform. Process. Syst.* **30**, 5998–6008 (2017).
15. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
16. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
17. Janesick, A. et al. High resolution mapping of the tumor micro-environment using integrated single-cell, spatial and in situ analysis. *Nat. Commun.* **14**, 8353 (2023).
18. Allen, W. E., Blosser, T. R., Sullivan, Z. A., Dulac, C. & Zhuang, X. Molecular and spatial signatures of mouse brain aging at single-cell resolution. *Cell* **186**, 194–208.e18 (2023).
19. Hu, Y. et al. Benchmarking clustering, alignment, and integration methods for spatial transcriptomics. *Genome Biol.* **25**, 212 (2024).
20. Armingol, E., Officer, A., Harismendy, O. & Lewis, N. E. Deciphering cell–cell interactions and communication from gene expression. *Nat. Rev. Genet.* **22**, 71–88 (2021).
21. Allen, C. D. C. et al. Germinal center dark and light zone organization is mediated by CXCR4 and CXCR5. *Nat. Immunol.* **5**, 943–952 (2004).
22. Young, C. & Brink, R. The unique biology of germinal center B cells. *Immunity* **54**, 1652–1664 (2021).
23. Robert, P. A., Rastogi, A., Binder, S. C. & Meyer-Hermann, M. How to simulate a germinal center. in *Germinal Centers: Methods and Protocols* (ed. Calado, D. P.) 303–334. [https://doi.org/10.1007/978-1-4939-7095-7\\_22](https://doi.org/10.1007/978-1-4939-7095-7_22) (Springer, 2017).
24. De Silva, N. S. & Klein, U. Dynamics of B cells in germinal centres. *Nat. Rev. Immunol.* **15**, 137–148 (2015).
25. Jiang, J. et al. Tumour-infiltrating immune cell-based subtyping and signature gene analysis in breast cancer based on gene expression profiles. *J. Cancer* **11**, 1568–1583 (2020).
26. Janssens, R., Struyf, S. & Proost, P. The unique structural and functional features of CXCL12. *Cell Mol. Immunol.* **15**, 299–311 (2018).
27. Zhang, Q. et al. CD86 is associated with immune infiltration and immunotherapy signatures in AML and promotes its progression. *J. Oncol.* **2023**, 9988405 (2023).
28. Yang, M.-W. et al. SFRP4 is a prognostic marker and correlated with Treg cell infiltration in pancreatic ductal adenocarcinoma. *Am. J. Cancer Res.* **9**, 363–377 (2019).
29. Kohli, K., Pillarisetty, V. G. & Kim, T. S. Key chemokines direct migration of immune cells in solid tumors. *Cancer Gene Ther.* **29**, 10–21 (2022).
30. Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nat. Methods* **19**, 534–546 (2022).
31. MERSCOPE FFPE Solution. <https://info.vizgen.com/merscope-ffpe-solution>.
32. Bromley, S. K., Thomas, S. Y. & Luster, A. D. Chemokine receptor CCR7 guides T cell exit from peripheral tissues and entry into afferent lymphatics. *Nat. Immunol.* **6**, 895–901 (2005).
33. Zheng, L. et al. Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science* <https://doi.org/10.1126/science.abe6474> (2021).
34. Dangaj, D. et al. Cooperation between constitutive and inducible chemokines enables T cell engraftment and immune attack in solid tumors. *Cancer Cell* **35**, 885–900.e10 (2019).
35. Zhang, M. et al. Molecularly defined and spatially resolved cell atlas of the whole mouse brain. *Nature* **624**, 343–354 (2023).
36. Yao, Z. et al. A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *Nature* **624**, 317–332 (2023).
37. Tissue histology in 3D. *Nat. Methods* **21**, 1133–1133 (2024).
38. Chen, A. et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792.e21 (2022).
39. Oliveira, M. F. et al. Characterization of immune cell populations in the tumor microenvironment of colorectal cancer using high definition spatial profiling. 2024.06.04.597233 Preprint at <https://doi.org/10.1101/2024.06.04.597233> (2024).
40. Radford, A. et al. Learning transferable visual models from natural language supervision. In *Proc. 38th Int. Conference on Machine Learning* **139**, 8748–8763 (2021).
41. Kaplan, J. et al. Scaling laws for neural language models. Preprint at <https://doi.org/10.48550/arXiv.2001.08361> (2020).
42. Shorten, C. & Khoshgoftaar, T. M. A survey on image data augmentation for deep learning. *J. Big Data* **6**, 60 (2019).
43. Lee, J. et al. Set transformer: a framework for attention-based permutation-invariant neural networks. In *Proc. 36th International Conference on Machine Learning* 3744–3753 (PMLR, 2019).
44. Dosovitskiy, A. et al. An image is worth 16×16 words: transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR, 2021)*; <https://openreview.net/forum?id=YicbFdNTTy>.
45. Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. A simple framework for contrastive learning of visual representations. In *Proc. 37th International Conference on Machine Learning* 1597–1607 (PMLR, 2020).
46. Kool, W., van Hoof, H. & Welling, M. Stochastic beams and where to find them: the gumbel-top-k trick for sampling sequences without replacement. In *International Conference on Machine Learning*, 3499–3508 (PMLR, 2019).
47. Kipf, T. N. & Welling, M. Variational graph auto-encoders. In *Neural Information Processing Systems Workshop on Bayesian Deep Learning* (Curran Associates, Inc., 2016).
48. Jin, S. et al. Inference and analysis of cell–cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
49. Sarkar, H., Chitra, U., Gold, J. & Raphael, B. J. A count-based model for delineating cell–cell interactions in spatial transcriptomics data. *Bioinformatics* **40**, i481–i489 (2024).
50. Fischer, D. S., Schaar, A. C. & Theis, F. J. Modeling intercellular communication in tissues using spatial graphs of cells. *Nat. Biotechnol.* 1–5 <https://doi.org/10.1038/s41587-022-01467-z> (2022).
51. Luecken, M. D. et al. Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* **19**, 41–50 (2022).
52. Datasets. *10x Genomics* <https://www.10xgenomics.com/datasets>.
53. Chen-Rui, X. et al. DECIPHER for learning disentangled cellular embeddings in large-scale heterogeneous spatial omics data. DECIPHER: v0.1.2. Zenodo <https://doi.org/10.5281/zenodo.15860235> (2025).
54. The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell* **181**, 936–953.e20 (2020).

## Acknowledgements

We thank Mr. Z. Qi for his assistance with H&E image annotation, as well as Drs. Yuan Lin, Z. Zhang, F. Tang, and X.S. Xie at Peking University for their helpful discussions and comments during the study. This work was supported by funds from the National Key Research and Development Program of China (2023YFF1204703, 2022ZD0115004), as well as the State Key Laboratory of Gene Function and Modulation Research, the

Beijing Advanced Innovation Center for Genomics (ICG) at Peking University, the Changping Laboratory, and the Shaw Foundation Hong Kong Limited. The research of C.R.X. is supported in part by the National Natural Science Foundation of China (grant no. 323B2017). The research of Z.J.C. is supported in part by the China Postdoctoral Science Foundation (grant no. 2023T160009 to Z.J.C.).

### Author contributions

G.G. and Z.J.C. conceived the study and supervised the research. C.R.X. and Z.J.C. designed and implemented the computational framework and conducted benchmarks and case studies with guidance from G.G. C.R.X., Z.J.C., and G.G. wrote the manuscript.

### 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/s41467-025-63140-8>.

**Correspondence** and requests for materials should be addressed to Zhi-Jie Cao or Ge Gao.

**Peer review information** *Nature Communications* thanks Wei Wei and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025