

# Pairpot: a database with real-time lasso-based analysis tailored for paired single-cell and spatial transcriptomics

Zhihan Ruan<sup>1,†</sup>, Fan Lin<sup>1,†</sup>, Zhenjie Zhang<sup>1,†</sup>, Jiayue Cao<sup>1</sup>, Wenting Xiang<sup>1</sup>, Xiaoyi Wei<sup>2</sup> and Jian Liu<sup>3,\*</sup>

<sup>1</sup>Centre for Bioinformatics and Intelligent Medicine, College of Computer Science, Nankai University, No.38 Tongyan Road, 300350 Tianjin, China

<sup>2</sup>Fifth Affiliated Hospital of Sun Yat-sen University, No.52 East Meihua Road, 519000 Zhuhai, China

<sup>3</sup>State Key Laboratory of Medical Chemical Biology, College of Computer Science, Nankai University, No.38 Tongyan Road, 300350 Tianjin, China

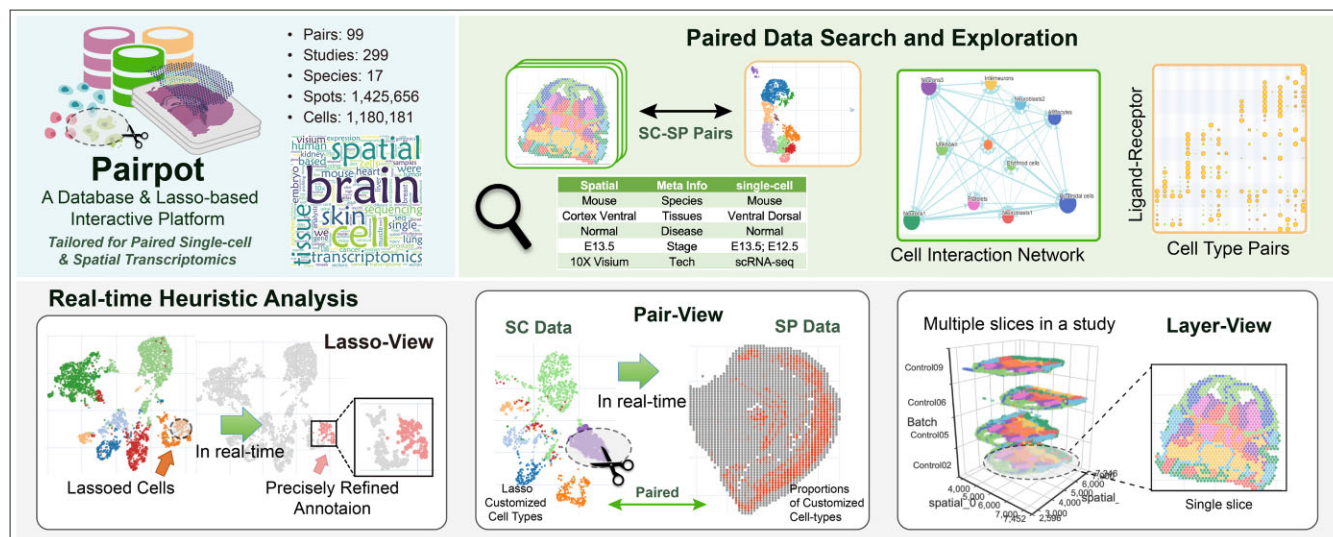
\*To whom correspondence should be addressed. Tel: +86 22 85358850; Fax: +86 22 85358850; Email: jianliu@nankai.edu.cn

†The first three authors should be regarded as Joint First Authors.

## Abstract

Paired single-cell and spatially resolved transcriptomics (SRT) data supplement each other, providing in-depth insights into biological processes and disease mechanisms. Previous SRT databases have limitations in curating sufficient single-cell and SRT pairs (SC-SP pairs) and providing real-time heuristic analysis, which hinder the effort to uncover potential biological insights. Here, we developed Pairpot (<http://pairpot.bioxai.cn>), a database tailored for paired single-cell and SRT data with real-time heuristic analysis. Pairpot curates 99 high-quality pairs including 1,425,656 spots from 299 datasets, and creates the association networks. It constructs the curated pairs by integrating multiple slices and establishing potential associations between single-cell and SRT data. On this basis, Pairpot adopts semi-supervised learning that enables real-time heuristic analysis for SC-SP pairs where Lasso-View refines the user-selected SRT domains within milliseconds, Pair-View infers cell proportions of spots based on user-selected cell types in real-time and Layer-View displays SRT slices using a 3D hierarchical layout. Experiments demonstrated Pairpot's efficiency in identifying heterogeneous domains and cell proportions.

## Graphical abstract



## Introduction

With the rapid development of single-cell technologies [e.g. single-cell RNA sequencing (scRNA-seq)], more and more cellular expression patterns are characterized, which allows the

construction of numerous cell atlases in the single-cell resolution (1). These massive atlases reveal significant cell-to-cell heterogeneity across diverse tissues and organ systems. The functions of tissues are closely related to the positional

Received: August 14, 2024. Revised: October 10, 2024. Editorial Decision: October 11, 2024. Accepted: October 15, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

contexts in microenvironments; however, cells usually suffer from dissociation during the whole tissue suspension (2). To effectively preserve the invaluable spatial contexts, spatially resolved transcriptomics (SRT) technologies are developed, making it promising to uncover the heterogeneous transcriptome landscapes (3).

SRT technologies fall into two categories: sequencing-based [i.e. 10× Visium (4), Slide-seq V2 (5) and Stereo-seq (6)] and imaging-based [i.e. MERFISH (7), STARmap (8) and NanoString COX (9)] strategies, which often require the presence of single-cell data to detect spatial contexts in single-cell levels (10). Sequencing-based technologies capture long-read transcripts *in situ* and then conduct next-generation sequencing in spot resolution (11). They require precisely annotated single-cell data as references to infer the cell proportions of each spot (12). Image-based technologies detect target sequences using fluorescent probes, which fetch hundreds of genes at most (10). They also require well-annotated single-cell data for cell-type label transfer (13). Integrating single-cell and spatial transcriptomics as pairs can facilitate discoveries in fields such as neuroscience (14) and tumor micro-environments (15). Specifically, these single-cell and spatial transcriptome data pairs (SC–SP pairs) should exhibit high similarity in terms of species, tissues, developmental stages and diseases (12).

To achieve SC–SP pairs, more and more studies performed single-cell and SRT sequencing simultaneously on samples of the same conditions and further conducted integrative analysis (16). Besides these studies, some SRT research used previously published single-cell data or atlases, which were generated under similar conditions with SRT data, to achieve SC–SP pairs for integrative analysis (17). In addition, many SRT researches are still not supplemented with paired single-cell data. Previous SRT databases mainly focus on collecting spatial transcriptomics data [e.g. SpatialDB (18), SOAR (19), SODB (20), STomicsDB (21) and Aquila (22)] but ignore curating paired single-cell data. Researchers have to frequently access corresponding single-cell data from different data resources to match SC–SP pairs, which is time-consuming. It is promising to construct a database that collects currently available single-cell and SRT data and creates valuable SC–SP pairs with cell-type annotations.

Recently, some databases have considered paired single-cell and spatial transcriptomics resources [e.g. SPASCR (23), ssREAD (24), SORC (25) and CROST (26)]. Existing databases are either limited to specific fields [e.g. cancer (25) and Alzheimer's disease (24)] or mainly focus on analysis within individual slices, ignoring potential associations across multiple slices of SRT data. Analysis within individual slices may pose challenges for capturing the global spatial distribution patterns that span across slices. These databases also ignore establishing potential associations between single-cell and SRT data, which may make it hard to discover cells and spots with similar biological processes and disease mechanisms. Further, these databases offer limited support for incorporating heuristic analysis according to user interaction patterns. They fail to dynamically personalize spatial segmentation or deconvolution according to user-selected spots (cells) in real-time, making it challenging for users to identify rare cell types and regional heterogeneity of SRT data. Suppose biologists would like to personalize the integrative analysis, such as modifying annotations in single-cell references according to their preferences. In that case, they would have to spend extra time and effort reor-

ganizing inputs, running pipelines and fine-tuning parameters repeatedly.

To deal with the challenges above, we developed Pairpot, a database with real-time lasso-based analysis tailored for paired single-cell and spatial transcriptomics. Pairpot currently curates 99 available single-cell and SRT data pairs from 299 studies, evaluates the quality of these SC–SP pairs, and establishes their association networks. Pairpot performs in-depth pre-analysis on the curated pairs, including clustering, cell-type annotation, marker detection, cell-proportion inference and cell/domain interaction. Pairpot provides Layer-View to display multiple slices in a 3D hierarchical layout. In addition, Pairpot constructs the neighbor graphs based on integrated data to establish connections among multiple slices. Pairpot assigns signatures of different cell types from PanglaoDB (27) using UCell (28) to evaluate the relative gene expression in single-cell and SRT data. On this basis, Pairpot proposes real-time lasso-based heuristic analysis based on semi-supervised machine learning. In particular, Lasso-View refines the customized domains lassoed by users within milliseconds using the neighbor graphs from integrated data. Moreover, Pair-View infers cell proportions of spots using the cell types defined by users based on UCell signatures. Pairpot also generates code for these heuristic analysis that allows users to integrate these analysis with their existing pipelines. We leveraged Lasso-View to uncover similar domains across multiple slices using lassoed spots, and utilized Pair-View to precisely infer proportions of user-defined cell sub-types. These case studies demonstrated Pairpot's efficiency in discovering heterogeneous domains and cell proportions. In summary, Pairpot provides valuable database resources and powerful heuristic analysis tools that empower researchers to streamline real-time single-cell and SRT integrative analysis.

## Materials and methods

### Data collection

Spatial transcriptomics and single-cell datasets were acquired and downloaded from databases such as the National Center for Biotechnology Information (29), European Bioinformatics Institute (30), China National Center for Bio-information (31) and 10× Genomics (<https://www.10xgenomics.com>).

Inspired by previous studies (12), we designed strategies for collecting paired single-cell datasets that hinge on the nature of the SRT studies. Initially, if the SRT studies themselves provide single-cell data, we would collect them directly. If the SRT studies lack their own single-cell data but use previous single-cell datasets for integration analysis, we would collect these corresponding single-cell datasets instead. In the case that the SRT studies did not mention any single-cell datasets, we would search for paired single-cell datasets according to the consistency of features such as species, tissues and diseases among the SRT and single-cell studies. If there exist multiple single-cell studies with the same features, Pairpot chooses the single-cell data that gives annotation files by default. When there exist multiple studies that provide annotation files, Pairpot chooses the study containing the widest variety of cell types by default. Lastly, when all the above strategies fail, we would resort to the cells of corresponding tissues from ovarian single-cell atlases.

We constructed association networks among different datasets using Neovis.js. For constructing the word clouds, we

used the term frequency-inverse document frequency strategy based on TfidfVectorizer from scikit-learn (version 1.0.2) to evaluate the word weight in the dataset summary and overall design.

## Data curation

### Quality control and preprocessing

Quality control was conducted on raw scRNA-seq and SRT data using Cell Ranger (version 7.2) and Space Ranger (version 3.0) (10× Genomics) to remove low-quality reads respectively. For BAM format files, we converted them to raw sequencing files using bamtofastq in CellRanger and then performed quality control. All datasets were manually converted to Anndata format after quality control. We used Scanpy (32) (version 1.9.1) as the standard downstream analysis pipeline. Genes expressing less than five cells were filtered out. Cells (spots) expressing >200 genes or showing a mitochondrial expression ratio above 5% of the total gene expression were also removed. Doublet cells in scRNA-seq data were identified and removed using Scrublet (33) (version 1.1.0). We performed log-normalization, selected 2000 highly variable genes using *sc.pp.highly\_variable\_genes* and performed principal component analysis (PCA) using *sc.pp.pca(adata, n\_comps=50, svd\_solver='arpack')*.

### Data integration and neighbor graph construction

To establish the associations among slices in SRT data, we integrated multiple samples presented in a dataset to correct batch effects using harmony (34) (version 0.0.9) based on 50 PCA components. Specifically for image-based data, we further integrated multiple image-based SRT samples with paired scRNA-seq data. For neighbor graph construction, assuming that the integrated dataset has  $n$  cells, we constructed the knn-based connectivity graph  $V(n \times n)$  using the function *sc.pp.neighbors*. In the connectivity graph, each node represents a cell, and each edge represents the distance between two cells. We used a Gaussian kernel to generate the similarity matrix  $W$ , where

$$W_{i,j} = -\exp(-\alpha V_{i,j}^2). \quad (1)$$

Here, the hyper-parameter  $\alpha$  is 0.5 by default. Next, we built the probability transition matrix  $P$ , where  $P_{ij}$  is the probability of label transition from cell  $i$  to cell  $j$ .

$$P_{i,j} = \frac{W_{i,j}}{\sum_{k=1}^n W_{i,k}}. \quad (2)$$

The dimensional reduction was performed for visualization using uniform manifold approximation and projection (UMAP) and t-distributed stochastic neighbor embedding based on the neighbor graphs. Afterwards, we used MENDER (35) to perform multi-scale cellular context representations of SRT data.

### Clustering and cell annotation

Pairpot segmented the spatial domains in multiple slices using MENDER (35) based on the multi-scale cellular context representations. For scRNA-seq data, Pairpot clusters the cells using the Leiden algorithm with the default resolution. Markers were detected using *sc.tl.rank\_genes\_groups(adata, groupby='leiden', method='wilcoxon')*.

Pairpot uses different strategies to replicate annotations for single-cell data in the following scenarios. In the scenario

where the original study provides a publicly available annotation file, Pairpot directly imports and verifies its integrity, maintaining a direct alignment with the source data. In the scenario where annotation files are not available, Pairpot refers to the markers and cell names provided by original studies and replicates annotations manually. In the replicating process, existing publications may provide detailed annotations for the clusters relevant to their research topics, while other clusters are loosely annotated (36). The annotations of the clusters irrelevant to the original research topics may also be valuable for other researchers. We annotate these clusters by matching their marker genes with existing canonical markers to improve annotation accuracy. In the scenario where studies lack cell-type names or markers, we search for the potential cell types according to the organs and species from existing cell marker databases [e.g. PanglaoDB (27) and CellMarker 2.0 (37)], and then perform manual annotation as same as the scenario above.

### UCell assignment

Evaluating signature scores in both single-cell and SRT data helps elucidate cell-type proportions and functions in cells and spots, and therefore establish potential associations between single-cell and SRT data. Pairpot uses UCell (28) to assign signature scores and evaluate cell types for the datasets lacking original annotations. Since UCell scores only depend on relative gene expression, they are also fundamental parameters for Pair-View to perform real-time cell proportion inference in SRT data.

For an SC-SP pair, we first found the potential cell types  $C = \{c_1, c_2, \dots, c_k\}$  and their corresponding gene signatures  $S = \{S_1, S_2, \dots, S_k\}$  from PanglaoDB (27) based on the species and tissues. Then we evaluated the UCell scores (28) of the cell-type gene signatures in  $S$ . Assuming that a dataset has  $m$  genes after pre-processing, the gene expression of a cell or spot  $x_i$  is  $G_i = \{g_{i1}, g_{i2}, \dots, g_{im}\}$ . Assuming a cell-type gene signature set has  $n$  genes  $S = \{s_1, s_2, \dots, s_n\}$ ,  $S \in G_i$  and the corresponding gene rank score is  $r_{i1}, r_{i2}, \dots, r_{in}$ , the UCell score of  $x_i$  is

$$UCell(x_i, S) = 1 - \frac{U}{U_{\max}}, \quad (3)$$

where

$$U = \sum_{j=1}^k r'_{ij} - R_{\min}$$

$$U_{\max} = R_{\max} - R_{\min}.$$

Here,  $U$  is the Mann-Whitney U statistic (bounded by 0 and  $U_{\max}$ ), and  $maxRank$  is a hyperparameter (default to 1500), which sets limitations to the rank of the top genes. Rank score  $R_{ij}$  larger than  $maxRank$  would be set to  $maxRank$ .  $R_{\max}$  and  $R_{\min}$  are the max and min rank sum scores, respectively.

$$r'_{ij} = \begin{cases} r_{ij}, & r_{ij} < maxRank \\ maxRank, & r_{ij} \geq maxRank \end{cases}$$

$$R_{\max} = n * maxRank$$

$$R_{\min} = \frac{n(n+1)}{2}$$



## Deconvolution

Deconvolution was performed using Cell2Location (12) (version 0.1.3), RCTD (38) (version 2.0.0), CARD (39) (version 1.0), SpaTalk (40) (version 1.0), Seurat (41) (version 5.1.0) and CellTrek (42) for integrative analysis to infer the cell proportions of each spot using pre-defined cell type annotation results.

## Cell communication

We employed CellphoneDB (43) (version 5.0), CellChat (44) (version 1.5.0), iTALK (<https://github.com/Coolgenome/iTALK>), COMMOT (45) (version 0.0.3) and SpaTalk (40) (version 1.0) for cell interaction inference leveraging the clustering and cell-type annotation results, which were used to analyze communication patterns and mechanisms between tissue regions and neighboring cell types.

## Real-time heuristic analysis

Real-time heuristic analysis refers to the rapid (usually within milliseconds) processing of analytical tasks leveraging user-provided prompts, enabling timely responses. In Pairpot, we developed two real-time heuristic analysis modules: Lasso-View, which refines user-selected spatial domains or cells; and Pair-View, which infers cell proportions of spots using the user-defined cell types.

Inspired by the label propagation algorithm (46), Lasso-View aims to propagate the prior labels according to the connectivity among cells. During data curation, Pairpot evaluates similarity using the Gaussian kernel function and generates the probability transition matrix ( $P$ ) for each dataset according to formula (2). Each element  $P_{ij}$  in  $P$  represents the probability of transferring from the cell (spot)  $i$  to  $j$ . During online analysis, based on annotations provided by Pairpot, users can generate customized annotations by selecting new cell types using lasso tools. Lasso-View generates the prior labels by masking 90% of the cells annotated by Pairpot and appends the cells selected by users. Assuming that the prior labels have  $m$  candidates  $\{l_1, l_2, \dots, l_m\}$ , Lasso-View generates the label probability matrix  $Y_{n \times m}$ , whose element  $Y_{ij}$  represents the probability of classifying cell (spot)  $i$  as label  $l_j$ ,  $0 \leq Y_{ij} \leq 1$ .  $Y_{n \times m}$  is decomposed into  $YL_{k \times m}$  containing labeled cells, and  $YU_{(n-k) \times m}$  containing unlabelled ones,  $n$  denotes the number of cells and  $k$  denotes the number of labeled ones. Lasso-View initializes the label probability matrix  $Y_0$  as follows:

$$Y_0 = \begin{bmatrix} YL_0 \\ YU_0 \end{bmatrix}, YL_{0ij} = \begin{cases} 1, & \text{if label}(i) = l_j \\ 0, & \text{if label}(i) \neq l_j \end{cases}, YU_{0ij} = \frac{1}{m}. \quad (4)$$

Here,  $1 \leq i \leq k$  in  $YL_0$ ,  $k+1 \leq i \leq n$  in  $YU_0$ ,  $1 \leq j \leq m$ . In  $YL_0$ , each of its elements  $YL_{0ij}$  is set to 1 if the label of cell  $i$  is  $l_j$ . In  $YU_0$ , all initial values are set to  $1/m$ , indicating that each cell (spot) has an equal probability of being classified as any label candidate. During the iteration, Lasso-View calculates a new label proportion matrix  $Y_t$  by multiplying the probability transition matrix  $P$  with  $Y_{t-1}$  and doing normalization subsequently. In epoch  $t$ ,

$$Y_t = [YL_t, YU_t]^T \quad (5)$$

$$YL_t = YL_0,$$

$$YU_t = \text{normalize}(YU'_t),$$

$$[YL'_t, YU'_t]^T = P \times Y_{t-1}.$$

Here, the normalize function denotes that each element in  $YU'_t$  is divided by the sum of all the elements in its row. After each iteration, Lasso-View evaluates the L1-norm of the subtraction between  $Y_t$  and  $Y_{t-1}$ . Based on our experimental evaluations (Supplementary Figure S1), we determined that the termination point most suitable for Lasso-View is at either 1000 epochs or upon reaching a convergence at

$$\frac{|Y_t - Y_{t-1}|}{n} < 10^{-5}.$$

After termination, for each cell (spot)  $i$ , Lasso-View assigns to it the label with the highest probability in  $Y_i$ ,

$$\text{label}(i) = \text{argmax}([Y_{i1}, Y_{i2}, \dots, Y_{im}]), 1 \leq j \leq m.$$

Indeed, a few of the cells selected by users may be dissimilar from the major selected cells. After the Lasso-View process, we rectify the label of each lassoeed cell with the most frequent label among its  $k$ -nearest cells. Ultimately, the result will be displayed in the right chart of Lasso-View.

Pair-View aims to infer cell proportions of spots using the user-selected cells from single-cell data in real time. Inspired by SPOTlight (47), Pair-View performs non-negative linear square regression (NNLS) using UCell scores from user-defined cell types and spots in SRT data. During online analysis, users select cells in the single-cell data, while Pair-View assigns the user-selected cell type  $u$  to these cells. Pair-View then generates the prior labels by appending the cell type  $u$  and other previously annotated cell types. Assuming that SRT data has  $n$  spots, single-cell data have  $m$ -cell types (including the user-selected cell type) and their pre-built UCell scores have  $k$  gene signature sets. Pair-View constructs a non-negative linear regression model to infer cell proportions in spatial transcriptomics using the user-selected cells in scRNA-seq data.

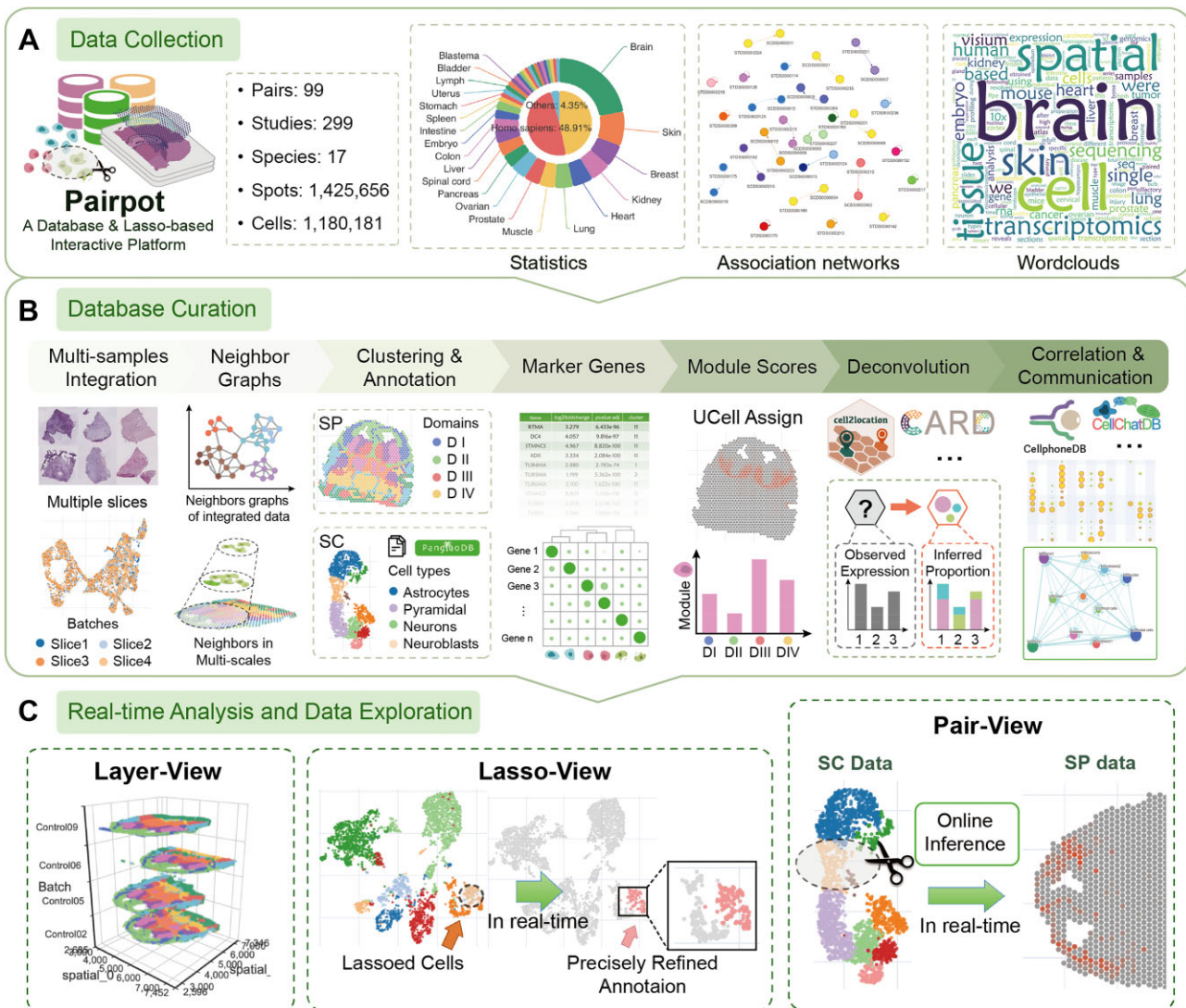
$$Y_{sp} \sim W * X_{sc} \quad (6)$$

Here,  $Y_{sp}$  is the  $n \times k$  UCell scores matrix of all spots in SRT data,  $X_{sc}$  is the  $m \times k$  mean UCell scores of all cell types in single-cell data. Each UCell score is defined by formula (3).  $W$  is the  $n \times m$  coefficient matrix, whose element  $W_{ij}$  denotes the weight of cell type  $j$  in spot  $i$ . The coefficient matrix  $W$  is inferred by `LinearRegression(positive=True)` in scikit-learn (version 1.3.2), Pair-View returns the weight of user-selected cell-type  $u$  in all spots as the cell proportions in SRT data.

$$\text{Prop}(u) = [w_{1u}, w_{2u}, \dots, w_{nu}] \quad (7)$$

## Database construction

The front-end frameworks were constructed with React.js (version 18.1.2) and ant-design (version 5.19.3). The real-time analysis modules were constructed with Echarts (version 5.4.3). The backend was built using Flask (version 3.1.0) and Python (version 3.10.12). Pairpot used SQLite3 (version 9.6) to store the metadata of publications and datasets. Neo4j (version 5.22.0) was used to store the relationships among datasets. Nginx (v1.22.0) was used as the reverse proxy server. Currently, Pairpot has been tested successfully in the following browsers: Google Chrome (v114.0), Safari (v17.5) and Firefox (v114.0).



**Figure 1.** Overview of Pairpot. **(A)** Data Collection of Pairpot. **(B)** Data Curation of Pairpot, including multi-sample integration, neighbor graphs construction, clustering and annotation, marker genes detection, module scores assignment, deconvolution and cell/domain communication evaluation. **(C)** Real-time analysis and data exploration of Pairpot.

## Database contents and usage

### Overview of Pairpot

Pairpot is a database tailored for paired single-cell and spatial transcriptomics data, providing data deployment, interactive data exploration and lasso-based real-time analysis (Figure 1). Pairpot collected 299 SRT datasets from 251 studies and generated 99 SC–SP pairs across 17 species and 25 spatial transcriptomics technologies, including 1,180,181 cells and 1,425,656 spots (Figure 1A). Pairpot provides association networks to establish connections among different single-cell and SRT studies, and wordclouds to excavate the focal points of the current available datasets in the research community. Pairpot also introduces a ‘Pair Score’ to evaluate the quality of each SC–SP pair.

Pairpot establishes the associations among SRT slices in a study by constructing neighbor graphs of multiple slices in multi-scales. The neighbor graphs enable users to heuristically explore and identify similar cells (spots) in other samples after selecting cells in one sample. Pairpot also provides cell-type annotations of single-cell data and their cell propor-

tions in SRT data. Moreover, Pairpot establishes the associations among single-cell and SRT data by evaluating their UCell scores, which enables users to infer the proportions of their customized cell types in spots of SRT data (Figure 1B). Each SRT dataset in Pairpot contains a gene expression matrix and the spatial coordinates for each spot. Each single-cell dataset contains cells with precise cell-type annotations, which are randomly downsampled to 3000 in order to accelerate the web-page loading while maintaining enough biological information. Both single-cell and SRT datasets are curated into a unified data format based on AnnData (48). On this basis, Pairpot provides real-time analysis including Layer-View, Lasso-View and Pair-View that supports for incorporating heuristic analysis according to user interaction patterns (Figure 1C).

Pairpot contains five modules: the data search and browse module, the Lasso-View module, the Layer-View module, the Pair-View module and the analysis results exploration module (Figure 2). Users can access each module using the navigation bar on the top of the Pairpot home page.

## Data search and browse

Pairpot enables efficient navigation through the extensive collection of datasets, which provides convenient access to meta-information, and literature resources (Figure 2A). Users can search for datasets by any words of interest, or apply filters by clicking preset keywords in species, tissues, technologies and diseases. Afterwards, they can take an overview of statistics and word clouds about the filtered results by browsing through an extensive list or association networks of all single-cell and SRT datasets. In the extensive list, each item contains a title, summary and a series of action buttons. By clicking ‘Visualization’, users can access the paired dataset, the static analysis results and the lasso-based real-time analysis module. By clicking ‘Descriptions’, users can access detailed meta-information and links to original datasets in each item. In the association networks, each node denotes a dataset and its color denotes the dataset’s species. Nodes with self-loops in the network represent datasets that contain SC–SP pairs. Paired single-cell and SRT datasets are linked by undirected edges. Users can click the nodes to explore detailed information and access the analysis modules. In brief, Pairpot provides visual insights into dataset relationships and characteristics.

## Lasso-view

Lasso-View is a heuristic analysis that discovers extra omitted cells (spots) similar to the lassoed ones in both single-cell and SRT datasets based on semi-supervised machine learning (Figure 2B). Users select cells (spots) of interest as a cluster (domain), then Lasso-View discovers unselected cells similar to those in the user-defined cluster and removes the cells mistakenly selected into the cluster (Figure 3). During data curation, the probability transition matrix of each dataset was generated by Pairpot in advance (Figure 3A). During online analysis, users can generate customized annotations by adding new cell types using lasso tools. Then, Lasso-View masks 90% of cell (spot) labels, appends the user-selected labels and initializes the initial label probability matrix (Figure 3B) and then enters iterations (Figure 3C). After the termination of iterations, Lasso-View highlights cells similar to the new cell type with rectification (Figure 3D). Combined with neighbor graphs constructed with integrated datasets, Lasso-View can further discover similar spatial domains across multiple slices based on user-selected domains in a slice. Users can click the lasso tools at TopRight and then select their cells of interest in the left chart. Lasso-View provides various select modes and tools for users to zoom, draw and erase their customized cells. Lasso-View also offers predefined annotations of spatial domains and cell annotations, which serve as a reference guide for users to select cells (spots) of their interests. After renaming the selected cells, users can click ‘Refine’ to generate refined annotations by calling the API of Lasso-View in the server. The API of Lasso-View is optimized by C++, ensuring efficient processing of single-cell and SRT datasets within a millisecond-level response time. The refined annotations can be confirmed and downloaded through their corresponding buttons, facilitating the sharing of refined results among collaborators.

## Layer-view

Inspired by 3D Landscape in De-spot (49), Pairpot develops Layer-View to dynamically display gene expressions and annotations of multiple slices in 3D hierarchical layouts (Figure 2C). Leveraging the integrated neighbor graphs, the spatial

domains are segmented across multiple slices using MENDER (35). Users can explore multiple slices of a study in the left chart, or focus on a single slice in the right chart. In the 3D layout, the x-axis and y-axis denote the spatial coordinates of the slices, while the z-axis represents different batches of these slices. Users can rotate the axis to switch perspectives, click ‘Inverse’ to hide all annotations, and highlight specific domains by clicking their legends.

## Pair-view

Pair-View is another heuristic analysis that aids users in quickly inferring cell proportions of spots using their customized cells from single-cell data in real-time (Figure 2D). The interface of Pair-View allows users to view SRT data (right) and paired single-cell data (left) simultaneously. Users can use lasso tools to select cells of their interests in the single-cell chart, similar to Lasso-View. They can also use options in ‘scConfigs’ to select customized cell types under different annotations and embeddings. After selecting customized cell types, users can click the ‘Deconv’ buttons to call the Pair-View API. The Pair-View API performs NNLS online based on the pre-analyzed UCell scores, which uses the Mann–Whitney U statistic to evaluate the related gene expression in single-cell and SRT data (see the ‘Materials and methods’ section). The inferred cell proportions would be subsequently displayed in the SRT chart. Users can further explore cell proportions in different batches through the options in ‘spConfigs’.

## Analysis results exploration

Pairpot also provides diverse pre-analysis results for user exploration (Figure 2E). Users can click a cell type in the left chart of Pair-View and then its proportions will be displayed in the right chart accordingly. Users can access spatial markers of domains through a marker heatmap, where domains are hierarchically clustered. When hovering over the scatters in the Marker Table, a tooltip will pop up with their cell types, marker names, average expressions and fractions. When clicking the ‘Rotation’ button, the marker table will exchange the x-axis and y-axis and provide a scalable slider to navigate through areas of interest. Users can browse interaction networks and ligand-receptor pairs for both single-cell and SRT data. For spatial domain interactions of SRT data, users can see the spatial locations of clusters by clicking nodes in the spatial domain interaction networks. For cell interactions of single-cell data, users can see the locations of cells in UMAP embeddings alongside. When hovering over a clustering node, the tooltip will pop up with main cell types, interaction counts and highly variable genes. In addition, users can select the provided legends to re-render the interaction networks. For ligand-receptor pair heatmaps, users can select the results generated by different interaction inference methods as preferences.

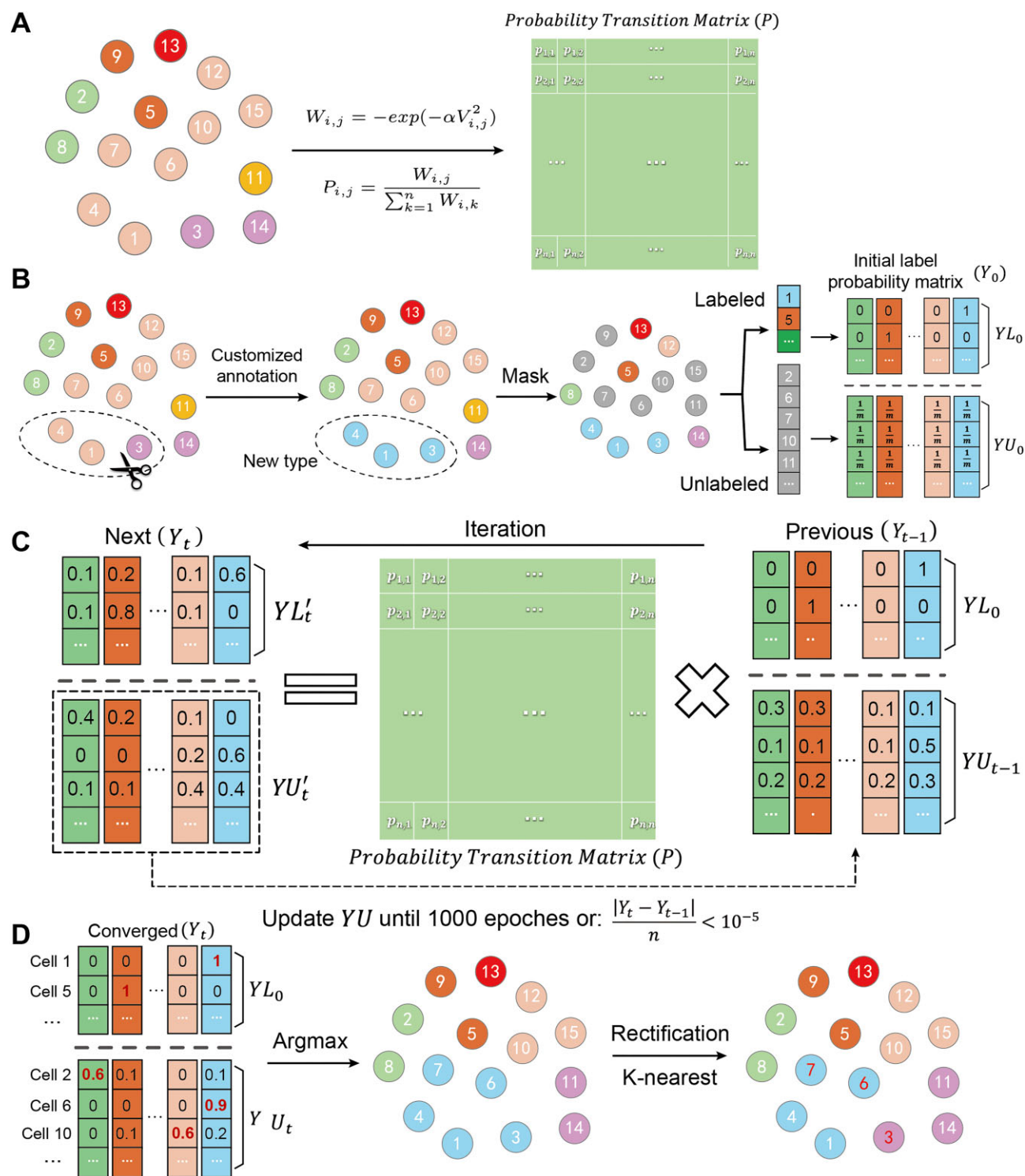
## Case study

We explored the embryonic mouse brain spatial transcriptomics data (STDS0000235), which contains four slices with 4597 spots and 17 different domains (17). Its corresponding single-cell data (SCDS0000001) are from dorsal and ventral E13.5 embryonic brains of a previous published dataset (50), which are further annotated to 11 cell types including astrocytes, neurons, neuroblasts, pyramidal cells, platelets and erythroid cells (Figure 4).



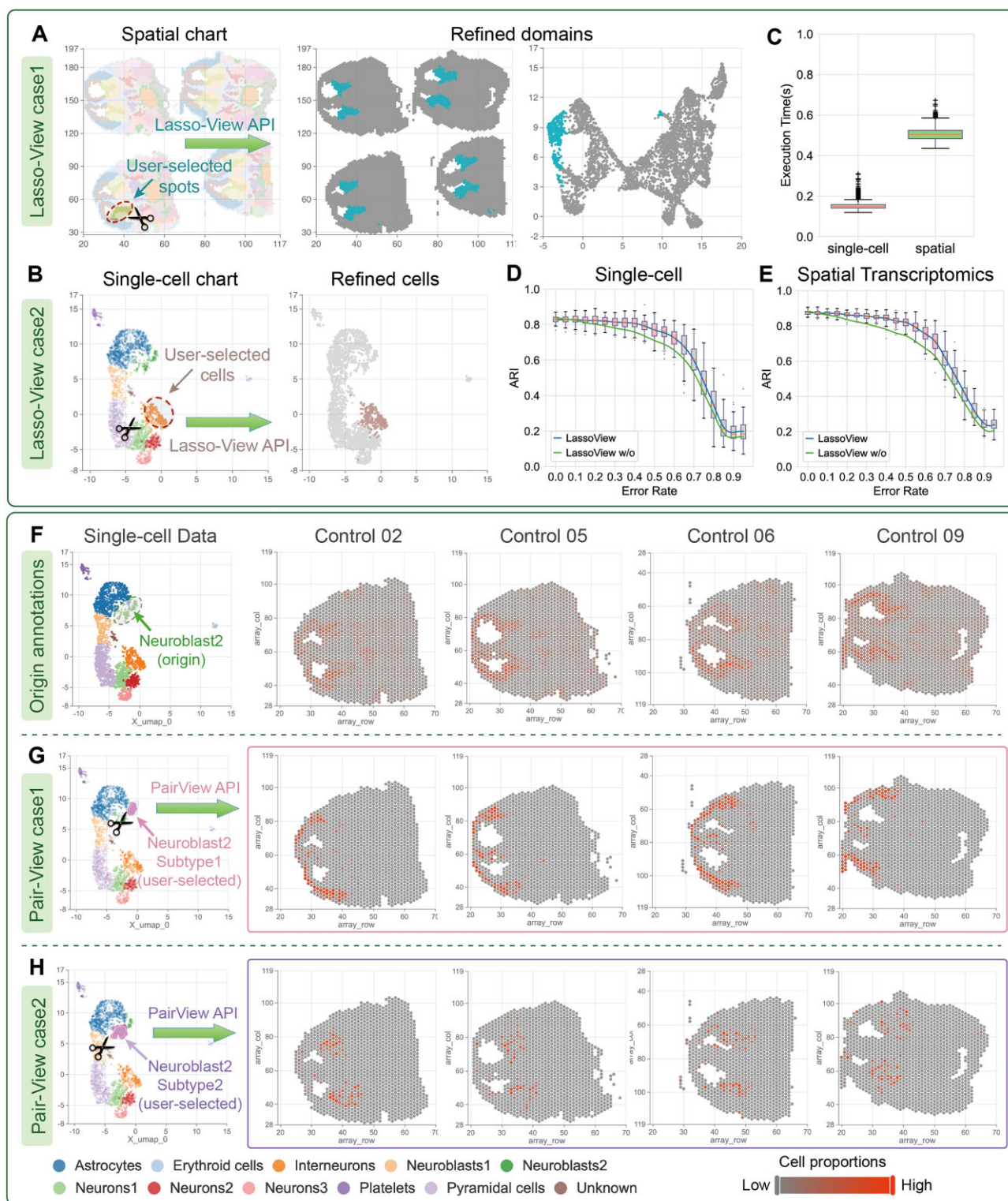


**Figure 2.** User interfaces and usages of modules in Pairpot database. **(A)** Data Search and Browse. **(B)** The Lasso-View module. **(C)** The Layer-View module. **(D)** The Pair-View module. **(E)** Analysis results exploration, including marker heatmaps, cell/domain interaction networks and ligand-receptor pair heatmaps.



**Figure 3.** Schema of Lasso-View. **(A)** Data preparation for Lasso-View. Pairpot evaluates the probability transition matrix  $P_{n \times n}$  for each dataset,  $n$  denotes the number of cells in the dataset. Nodes denote cells and colors denote their corresponding labels provided by Pairpot. **(B–D)** Real-time analysis in Lasso-View. **(B)** Initialization of label probability matrix. Users generate customized annotations using Lasso tools. Then Lasso-View masks 90% of cell (spot) labels, appends the user-selected labels and evaluates the initial label probability matrix. Each column in  $Y$  corresponds to a cell-type label. **(C)** Iterations of Lasso-View. **(D)** Labels assignment and rectification. The highlighted element in the converged matrix denotes the highest probability of each row. In the output of Lasso-View, Cells 6 and 7 are similar to the user-selected cells, and Cell 3 is rectified.





**Figure 4.** Case study of SC-SP pairs containing multiple mouse brain slices in Pairpot. **(A)** Lasso-View case 1. Left, the spatial chart where users select the spots of their interests in a slice. Middle, the refined domains of multiple slices in spatial embedding. Right, the refined domains of multiple slices in UMAP embedding. **(B)** Lasso-View case 2. Left, the single-cell chart where users select the cells of their interests. Right, the refined cells in UMAP embedding. **(C)** Execution time of Lasso-View API in current single-cell and SRT data [bars indicate the mean  $\pm$  standard error of the mean (SEM)]. **(D, E)** The adjusted Rand index (ARI) of Lasso-View tested in single-cell **(D)** and SRT **(E)** data with the condition of 90% masked annotations and 0–0.95 error rate in 10% known annotations (bars indicate the mean  $\pm$  SEM). The green line indicates mean ARI of Lasso-View without rectification. **(F)** Pair-View identifies the proportions of user-selected cells in multiple mouse brain slices. Left, Origin annotations of single-cell data. Right, cell proportions of neuroblast 2 in slices Control 02, Control 05, Control 06 and Control 09, pre-defined by Pairpot. **(G)** Pair-View case 1. Left, the single-cell chart where users select the neuroblast 2 subtype 1 evaluated by Pair-View. Right, cell proportions of neuroblast 2 subtype 1 evaluated by Pair-View. **(H)** Pair-View case 2. Left, the single-cell chart where users select the neuroblast 2 subtype 2. Right, cell proportions of neuroblast 2 subtype 2 evaluated by Pair-View.

We conducted case studies to evaluate the capability of Lasso-View to discover biological insights. In Lasso-View case 1, we selected spots in a lateral ganglionic eminence (LGE) domain of slice Control 02 (Figure 4A, left). After calling Lasso-View API, the refined domains were highlighted not only in slice Control 02 but also in LGE domains of other slices (Figure 4A, right). In addition, spots mistakenly selected in the cortex regions were removed. In Lasso-View case 2, we selected part of the interneuron cells in the single-cell chart, and the refined cells covered the interneurons after calling Lasso-View API (Figure 4B). The results demonstrated that Lasso-View can discover similar domains across different slices and remove unrelated spots.

To evaluate the accuracy and response speed of Lasso-View, we randomly select cells and call the Lasso-View API 10 000 times to simulate the user's behavior. As shown in Figure 4C, the average processing time of Lasso-View API is <200 milliseconds for single-cell data and about 500 milliseconds for SRT data. We masked 90% of labels, and randomly replaced the remained 10% of labels with errors as pseudo user-selected labels for Lasso-View. We used these pseudo-labels to predict the labels of masked cells and evaluate ARI for each cell type. The average ARI for single-cell data was over 0.8 in the non-error situation and remained above 0.7 even when the error rate was up to 0.6 (Figure 4D). For spatial transcriptome data, the average ARI reached 0.85 in the non-error situation and remained above 0.7 when the error rate was up to 0.65 (Figure 4E). Additionally, we tested the capability of rectification options. Lasso-View with rectification maintained significantly better ARI performance across different error rates (Figure 4D and E). The results above revealed that Lasso-View can accurately discover cells similar to user-selected ones and sensitively rectify errors in the user selection within milliseconds.

We also applied real cases to evaluate the capability of Pair-View in inferring cell proportions using user-selected cells from single-cell data. Neuroblast2 cells were originally detected in the dorsal and ventral embryonic brain (Figure 4F). To detect the proportions of neuroblast2 more precisely, we manually dichotomized neuroblast2 into subtype 1 and subtype 2 using lasso tools. In particular, subtype1 contains more cells from the dorsal batch, while subtype2 contains more cells from the ventral batch. After calling Pair-View API, we found that subtype 1 was correctly detected in the cortex of the dorsal embryonic brain, corresponding to the expression of *TBR1* (Figure 4G). In contrast, subtype 2 was detected in the LGE region of the ventral brain, associated with *GAD2* (Figure 4H). The locations of both subtype 1 and subtype 2 remain consistent across the four slices. The results demonstrated that Pair-View can perform precise cell proportion inference online and significantly improve the efficiency of exploring paired datasets.

## Conclusions and future developments

We presented Pairpot, a database with real-time lasso-based analysis tailored for single-cell and SRT pairs. The database aims to ease the process of searching SC-SP pairs and facilitate the applications of paired single-cell and SRT data within the research community. Pairpot collects available SC-SP pairs and constructs the network associations among them. Pairpot evaluates the quality of SC-SP pairs, integrates different slices, establishes potential associations between single-cell and SRT

data, and offers advanced real-time heuristic analysis modules to significantly promote understanding of biological processes and disease mechanisms. In particular, Lasso-View refines the user-selected cells by discovering unselected cells similar to those in the user-defined cluster and removing the mistakenly selected cells. Lasso-View is optimized by C++ and can respond within milliseconds. Pair-View can infer cell proportions of spots using the user-selected cells from single-cell data. Both Lasso-View and Pair-View achieve real-time in online scenarios, considerably reducing the extra work required by biologists in segmentation and deconvolution, thereby significantly improving the efficiency of personalized analysis. Pairpot is valuable to biologists not only because it offers easy-to-access paired datasets but also because its lasso-based modules provide a new perspective for developing efficient online computational platforms based on pre-analyzed information and user behavior. Pairpot is also useful for researchers to seek potential biological insights because it provides both overall observations for multiple slices in a 3D layout and precise real-time lasso-based analysis for cells or domains of user's interest.

Pairpot's intuitive interface and heuristic analysis modules empower users to tailor their workflows according to their specific needs. For researchers who prefer to download the data and conduct analyses offline, Pairpot generates code for online heuristic analysis and alternative bioinformatics tools that allow users to replicate the analysis results offline. If researchers would like to analyze their own paired single-cell and SRT data, they can integrate the interfaces provided by Pairpot into their pipelines, including preprocessing, data integration, neighbor graph construction and UCell scores evaluation. For researchers without programming capabilities, Pairpot provides valuable resources and powerful tools online to streamline the SRT integrative analysis and enhance the understanding of biological processes and disease mechanisms.

We will continue to update Pairpot by collecting newly published SRT data with annotations and matching them with paired single-cell data. In addition, we plan to conduct SC-SP pairs by matching an SRT dataset with multiple single-cell datasets simultaneously. Since single-cell and spatial transcriptomics are both fast-evolving fields, we plan to continuously provide results generated by newly developed alternative tools and develop more downstream analysis modules to facilitate biomedical research.

## Data availability

The Pairpot website is available online at <http://pairpot.bioxai.cn>. Its data resources are also freely available at <http://src.bioxai.cn>. The source code of Pairpot is available at GitHub: <https://github.com/lyotvincent/Pairpot> and Zenodo: <https://doi.org/10.5281/zenodo.13919102>.

## Supplementary data

Supplementary Data are available at NAR Online.

## Acknowledgements

**Author contributions:** J.L. conceived and designed the project. Z.R., F.L., Z.Z., J.C. and X.W. curated the database. Z.R., Z.Z. and J.L. developed the approach and conducted analysis pipelines. Z.R., F.L. and W.X. developed the web platforms.

Z.R. and J.L. wrote the manuscript with input from all authors. All authors read and approved the final manuscript.

## Funding

National Key Research and Development Program of China [2020YFA0908700 and 2020YFA0908702]; National Natural Science Foundation of China [62272246].

## Conflict of interest statement

None declared.

## References

- Lee, A.T., Chang, E.F., Paredes, M.F. and Nowakowski, T.J. (2024) Large-scale neurophysiology and single-cell profiling in human neuroscience. *Nature*, **630**, 587–595.
- van den Brink, S.C., Sage, F., Vértessy, B., Spanjaard, B., Peterson-Maduro, J., Baron, C.S., Robin, C. and van Oudenaarden, A. (2017) Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods*, **14**, 935–936.
- Marx, V. (2021) Method of the year: spatially resolved transcriptomics. *Nat. Methods*, **18**, 9–14.
- Stahl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., et al. (2016) Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, **353**, 78–82.
- Strickels, R.R., Murray, E., Kumar, P., Li, J., Marshall, J.L., Di Bella, D.J., Arlotta, P., Macosko, E.Z. and Chen, F. (2021) Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.*, **39**, 313–319.
- Wei, X., Fu, S., Li, H., Liu, Y., Wang, S., Feng, W., Yang, Y., Liu, X., Zeng, Y.-Y., Cheng, M., et al. (2022) Single-cell Stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration. *Science*, **377**, eabp9444.
- Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S. and Zhuang, X. (2015) Spatially resolved, highly multiplexed RNA profiling in single cells. *Science*, **348**, aaa6090.
- Wang, X., Allen, W.E., Wright, M.A., Sylwestrak, E.L., Samusik, N., Vesuna, S., Evans, K., Liu, C., Ramakrishnan, C., Liu, J., et al. (2018) Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science*, **361**, eaat5691.
- Merritt, C.R., Ong, G.T., Church, S.E., Barker, K., Danaher, P., Geiss, G., Hoang, M., Jung, J., Liang, Y., McKay-Fleisch, J., et al. (2020) Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat. Biotechnol.*, **38**, 586–599.
- Rao, A., Barkley, D., França, G.S. and Yanai, I. (2021) Exploring tissue architecture using spatial transcriptomics. *Nature*, **596**, 211–220.
- Longo, S.K., Guo, M.G., Ji, A.L. and Khavari, P.A. (2021) Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nat. Rev. Genet.*, **22**, 627–644.
- Kleshchevnikov, V., Shmatko, A., Dann, E., Aivazidis, A., King, H.W., Li, T., Elmentaite, R., Lomakin, A., Kedlian, V., Gayoso, A., et al. (2022) Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.*, **40**, 661–671.
- Allen, W.E., Blosser, T.R., Sullivan, Z.A., Dulac, C. and Zhuang, X. (2023) Molecular and spatial signatures of mouse brain aging at single-cell resolution. *Cell*, **186**, 194–208.
- Lein, E., Borm, L.E. and Linnarsson, S. (2017) The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science*, **358**, 64–69.
- Maniatis, S., Petrescu, J. and Phatnani, H. (2021) Spatially resolved transcriptomics and its applications in cancer. *Curr. Opin. Genet. Dev.*, **66**, 70–77.
- Moncada, R., Barkley, D., Wagner, F., Chiodin, M., Devlin, J.C., Baron, M., Hajdu, C.H., Simeone, D.M. and Yanai, I. (2020) Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol.*, **38**, 333–342.
- Di Marco, B., Vázquez-Marín, J., Monyer, H., Centanin, L. and Alfonso, J. (2023) Spatial transcriptomics map of the embryonic mouse brain – a tool to explore neurogenesis. *Biology Open*, **12**, bio060151.
- Fan, Z., Chen, R. and Chen, X. (2019) SpatialDB: a database for spatially resolved transcriptomes. *Nucleic Acids Res.*, **48**, D233–D237.
- Li, Y., Dennis, S., Hutch, M.R., Ding, Y., Zhou, Y., Li, Y., Pillai, M., Ghotbaldini, S., Garcia, M.A., Broad, M.S., et al. (2023) SOAR elucidates disease mechanisms and empowers drug discovery through spatial transcriptomics. *bioRxiv* doi: <https://doi.org/10.1101/2022.04.17.488596>, 15 December 2023, preprint: not peer reviewed.
- Yuan, Z., Pan, W., Zhao, X., Zhao, F., Xu, Z., Li, X., Zhao, Y., Zhang, M.Q. and Yao, J. (2023) SODB facilitates comprehensive exploration of spatial omics data. *Nat. Methods*, **20**, 387–399.
- Xu, Z., Wang, W., Yang, T., Li, L., Ma, X., Chen, J., Wang, J., Huang, Y., Gould, J., Lu, H., et al. (2023) STOmicsDB: a comprehensive database for spatial transcriptomics data sharing, analysis and visualization. *Nucleic Acids Res.*, **52**, D1053–D1061.
- Zheng, Y., Chen, Y., Ding, X., Wong, K.H. and Cheung, E. (2022) Aquila: a spatial omics database and analysis platform. *Nucleic Acids Res.*, **51**, D827–D834.
- Fan, Z., Luo, Y., Lu, H., Wang, T., Feng, Y., Zhao, W., Kim, P. and Zhou, X. (2022) SPASCR: spatial transcriptomics annotation at single-cell resolution. *Nucleic Acids Res.*, **51**, D1138–D1149.
- Wang, C., Acosta, D., McNutt, M., Bian, J., Ma, A., Fu, H. and Ma, Q. (2024) A single-cell and spatial RNA-seq database for Alzheimer's disease (ssREAD). *Nat. Commun.*, **15**, 4710.
- Zhou, W., Su, M., Jiang, T., Yang, Q., Sun, Q., Xu, K., Shi, J., Yang, C., Ding, N., Li, Y., et al. (2023) SORC: an integrated spatial omics resource in cancer. *Nucleic Acids Res.*, **52**, D1429–D1437.
- Wang, G., Wu, S., Xiong, Z., Qu, H., Fang, X. and Bao, Y. (2023) CROST: a comprehensive repository of spatial transcriptomics. *Nucleic Acids Res.*, **52**, D882–D890.
- Franzén, O., Gan, L.-M. and Björkegren, J. L.M. (2019) PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database*, **2019**, baz046.
- Andreatta, M. and Carmona, S.J. (2021) UCell: robust and scalable single-cell gene signature scoring. *Comput. Struct. Biotechnol. J.*, **19**, 3796–3798.
- Sayers, E., Beck, J., Bolton, E., Brister, J., Chan, J., Comeau, D., Connor, R., DiCuccio, M., Farrell, C., Feldgarden, M., et al. (2023) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **52**, D33–D43.
- Thakur, M., Buniello, A., Brooksbank, C., Gurwitz, K.T., Hall, M., Hartley, M., Hulcoop, D.G., Leach, A.R., Marques, D., Martin, M., et al. (2023) EMBL's European Bioinformatics Institute (EMBL-EBI) in 2023. *Nucleic Acids Res.*, **52**, D10–D17.
- Members, C.-N. Partners (2023) Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2024. *Nucleic Acids Res.*, **52**, D18–D32.
- Wolf, F.A., Angerer, P. and Theis, F.J. (2018) SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.*, **19**, 15.
- Wolock, S.L., Lopez, R. and Klein, A.M. (2019) Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Systems*, **8**, 281–291.
- Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.-R. and Raychaudhuri, S. (2019) Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods*, **16**, 1289–1296.
- Yuan, Z. (2024) MENDER: fast and scalable tissue structure identification in spatial omics data. *Nat. Commun.*, **15**, 207.



36. Lu,Z., Zhong,A., Liu,H., Zhang,M., Chen,X., Pan,X., Wang,M., Deng,X., Gao,L., Zhao,L., *et al.* (2022) Dissecting the genetic and microenvironmental factors of gastric tumorigenesis in mice. *Cell Rep.*, **41**, 111482.
37. Hu,C., Li,T., Xu,Y., Zhang,X., Li,F., Bai,J., Chen,J., Jiang,W., Yang,K., Ou,Q., *et al.* (2023) CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Res.*, **51**, D870–D876.
38. Cable,D.M., Murray,E., Zou,L.S., Goeva,A., Macosko,E.Z., Chen,F. and Irizarry,R.A. (2022) Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.*, **40**, 517–526.
39. Ma,Y. and Zhou,X. (2022) Spatially informed cell-type deconvolution for spatial transcriptomics. *Nat. Biotechnol.*, **40**, 1349–1359.
40. Shao,X., Li,C., Yang,H., Lu,X., Liao,J., Qian,J., Wang,K., Cheng,J., Yang,P., Chen,H., *et al.* (2022) Knowledge-graph-based cell–cell communication inference for spatially resolved transcriptomic data with SpaTalk. *Nat. Commun.*, **13**, 4429.
41. Butler,A., Hoffman,P., Smibert,P., Papalexi,E. and Satija,R. (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.*, **36**, 411–420.
42. Wei,R., He,S., Bai,S., Sei,E., Hu,M., Thompson,A., Chen,K., Krishnamurthy,S. and Navin,N.E. (2022) Spatial charting of single-cell transcriptomes in tissues. *Nat. Biotechnol.*, **40**, 1190–1199.
43. Garcia-Alonso,L., Lorenzi,V., Mazzeo,C.I., Alves-Lopes,J.P., Roberts,K., Sancho-Serra,C., Engelbert,C., Marečková,M., Gruhn,W.H., Botting,R.A., *et al.* (2022) Single-cell roadmap of human gonadal development. *Nature*, **607**, 540–547.
44. Jin,S., Guerrero-Juarez,C.F., Zhang,L., Chang,I., Ramos,R., Kuan,C.-H., Myung,P., Plikus,M.V. and Nie,Q. (2021) Inference and analysis of cell–cell communication using CellChat. *Nat. Commun.*, **12**, 1088.
45. Cang,Z., Zhao,Y., Almet,A.A., Stabell,A., Ramos,R., Plikus,M.V., Atwood,S.X. and Nie,Q. (2023) Screening cell–cell communication in spatial transcriptomics via collective optimal transport. *Nat. Methods.*, **20**, 218–228.
46. Raghavan,N., Albert,R. and Kumara,S. (2007) Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E Stat. Nonlin. Soft. Matter. Phys.*, **76**, 36106.
47. Elosua-Bayes,M., Nieto,P., Mereu,E., Gut,I. and Heyn,H. (2021) SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res.*, **49**, e50–e50.
48. Virshup,I., Bredikhin,D., Heumos,L., Palla,G., Sturm,G., Gayoso,A., Kats,I., Koutrouli,M., Scverse,Community, Berger,B., *et al.* (2023) The scverse project provides a computational ecosystem for single-cell omics data analysis. *Nat. Biotechnol.*, **41**, 604–606.
49. Ruan,Z., Zhou,W., Liu,H., Wei,J., Pan,Y., Yan,C., Wei,X., Xiang,W., Yan,C., Chen,S., *et al.* (2024) Precise detection of cell-type-specific domains in spatial transcriptomics. *Cell Rep. Methods*, **4**, 100841.
50. La Manno,G., Siletti,K., Furlan,A., Gyllborg,D., Vinsland,E., Mossi Albiach,A., Mattsson Langseth,C., Khven,I., Lederer,A.R., Dratva,L.M., *et al.* (2021) Molecular architecture of the developing mouse brain. *Nature*, **596**, 92–96.