Published online 15 September 2022

NAR Genomics and Bioinformatics, 2022, Vol. 4, No. 3 **1**
https://doi.org/10.1093/nargab/lqac069

SPROUT: spectral sparsification helps restore the spatial structure at single-cell resolution

Jingwan Wang [1,2,†], Shiying Li [1,2,†], Lingxi Chen [1,2] and Shuai Cheng Li [1,2,*]

[1]Department of Computer Science, City University of Hong Kong, 83 Tat Chee Ave, Kowloon Tong, Hong Kong, China and [2]City University of Hong Kong Shenzhen Research Institute, Shenzhen, 518057 Guangdong, China

Received March 13, 2022; Revised July 11, 2022; Editorial Decision August 22, 2022; Accepted August 26, 2022

ABSTRACT

Single-cell RNA sequencing thoroughly quantifies the individual cell transcriptomes but renounces the spatial structure. Conversely, recently emerged spatial transcriptomics technologies capture the cellular spatial structure but skimp cell or gene resolutions. Ligand-receptor interactions reveal the potential of cell proximity since they are spatially constrained. Cell–cell affinity values estimated by ligand–receptor interaction can partially represent the structure of cells but falsely include the pseudo affinities between distant or indirectly interacting cells. Here, we develop a software package, SPROUT, to reconstruct the single-cell resolution spatial structure from the transcriptomics data through diminished pseudo ligand–receptor affinities. For spatial data, SPROUT first curates the representative single-cell profiles for each spatial spot from a candidate library, then reduces the pseudo affinities in the intercellular affinity matrix by partial correlation, spectral graph sparsification, and spatial coordinates refinement. SPROUT embeds the estimated interactions into a low-dimensional space with the cross-entropy objective to restore the intercellular structures, which facilitates the discovery of dominant ligand–receptor pairs between neighboring cells at single-cell resolution. SPROUT reconstructed structures achieved shape Pearson correlations ranging from 0.91 to 0.97 on the mouse hippocampus and human organ tumor microenvironment datasets. Furthermore, SPROUT can solely *de novo* reconstruct the structures at single-cell resolution, *i.e.*, reaching the cell-type proximity correlations of 0.68 and 0.89 between reconstructed and immunohistochemistry-informed spatial structures on a human developing heart dataset and a tumor microenvironment dataset, respectively.

INTRODUCTION

Revealing the spatial context and molecular abundance of cells and tissue is critical for understanding the composition and functions of complex tissues. Single-cell RNA sequencing (scRNA-seq) technologies quantify the single-cell transcriptome by a high sequencing depth with whole-transcriptome coverage (1). The thorough scope of single-cell transcriptome enables investigations on cell heterogeneities, subpopulations and interactions (2,3). However, the isolation procedure renounces the spatial context of these cells.

Spatial transcriptomics (ST) technologies have been developed to acquire spatial context and expression profiles simultaneously. High-plex RNA imaging technologies (4–6) only localize dozens to hundreds of genes, and spatial barcoding technologies such as 10X Visium, Slide-Seq and HDST (7–9) yield a greater magnitude. However, they have achieved unsatisfied abundances or inadequate cell resolution, which restricts the potential of ST data for downstream analyses. Resolution-enhancement methods, such as the BayesSpace method powered by imposing a Bayesian model on the spatial neighborhoods of spots, approach but do not yet arrive at the single-cell resolution (10).

Except for wet-lab approaches, researchers also proposed computational methods to restore the spatial structure from the scRNA-seq data. NovoSpaRc (11) assigns cells to tissue locations by probability. Its premise only considers the similarity in gene expression as the neighboring factor, neglecting the heterogeneity of, for instance, the transition areas (12) or immune cell infiltration regions (13). CSOmap reconstructs the intercellular proximity based on the contact-required ligand–receptor (LR) interactions (14,15). Specifically, CSOmap estimates the affinity of two cells by the mRNA expression summation of the interacting LR pairs, forming a k-nearest neighbor affinity graph simulating cell-cell interactions. However, two distant cells could have a 'pseudo' high affinity due to the share of strong similarity with the other two cells which interact. Without any spatial reference, the estimated affinity contains *pseudo* affinities from distance cells.

*To whom correspondence should be addressed. Tel: +852 34429412; Fax: +852 34420503; Email: shuaicli@cityu.edu.hk
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Researchers also started to integrate the ST data with the scRNA-seq data. Early attempts for integration focus on reconstructing cellular spatial structure based on spatial references such as immunohistochemistry (IHC) or fluorescence *in situ* hybridization (FISH) (16,17). Spatial barcoding presents a new aspect for integrating scRNA-seq and spatial data, leading to two primary integration approaches: deconvolution and mapping (18). One objective of deconvolution methods is to infer the proportion of cell types from each ST capture location or spot in the ST data. Provided with a labeled scRNA-seq dataset, non-negative least squares and dampened weighted least squares linear regression can deconvolute the captured spot mixtures (19,20). Alternatively, deconvolution can be accomplished by fitting a model of negative binomial distribution or Poisson distribution to the scRNA-seq expression with the empirical data of ST spot as a prior. Subsequently, maximized posterior yields an estimation of the cell-type distribution (21–23). Moreover, several studies on the tumor microenvironment (TME) map subgroups of single-cell to specific subregions in ST data by the enrichment score (24,25). These mappings improve the resolution on the subpopulation level but require prior clustering and annotation on both data types, which is inaccurate when mapping tissue regions comprised of mixed cell types. SpaOTsc (26) maps cells by minimizing the gene expression dissimilarity between single-cell data and ST with the optimal transport distance, neglecting the heterogeneity in the spot, and taking all provided single-cell without selection.

Here, we present a software package, SPROUT, that recapitulates the single-cell resolution cell structure of the spatial transcriptome from a sparsified affinity graph where the pseudo affinities are reduced by partial correlation (27), spectral sparsification (28), and spatial coordinates refinement. Instead of solely delivering cell-type acknowledgment, SPROUT locates single-cell expression profiles in spots from a candidate library, hence enabling the exploration of the spatial intercellular communication mechanisms at single-cell resolution.

MATERIALS AND METHODS

Constructing the single-cell aggregates to reproduce ST expression profiles

We propose a preprocessing module to integrate ST data with scRNA-seq data. The module takes two parameters, the cell number ℓ_s and cell-type proportion $p_{s,t}$, $t \in T$ for a ST spot s, T denotes the set of cell types. The parameter ℓ_s denotes the average number of cells in a spot. The number of cells captured in a spot varies according to sequencing methods and tissue density; our module allows users to specify it. Many accurate and efficient deconvolution methods have been developed to infer the cell-type proportion of each spot, such as stereoscope, RCTD, SPOTlight, and cell2location (19,21,23,29). Therefore, SPROUT accepts the deconvolute result of these packages to approximate the cell-type proportion $p_{s,t}$.

Let $k_{s,t}$ denote the cell number of type t at ST spot s, then $k_{s,t} \approx \ell_s \times p_{s,t} = f_{s,t}$. Note that $f_{s,t}$ can be fractional. Here, we round on $f_{s,t}$ randomly (30) to acquire the integer number of $k_{s,t}$ while stabilizing the expectation of ℓ_s. Randomized rounding round up or down with the probability proportional to the value of the variable in a fractional form. Denoting the decimal part of $f_{s,t}$ as $\{f_{s,t}\} \in [0, 1)$, $f_{s,t}$ randomly rounds up or down to $k_{s,t}$ according to the probability $P(k_{s,t} = \lceil f_{s,t} \rceil) = \{f_{s,t}\}$. The preprocessing module chooses cell set $\mathbf{M_s}$ from a predefined single-cell candidate library to reproduce the single-cell resolution for each spot s. The summed expression profile of all chosen cells in $\mathbf{M_s}$ termed the *aggregated expressions* $\mathbf{E}(\mathbf{M_s})$, which calculated by $\sum_{i=1}^{|\mathbf{M_s}|} \mathbf{E}(c_i)$. It curates the single-cell aggregates set $\mathbf{M_s}$ by maximizing the Pearson correlation between $\mathbf{E}(\mathbf{M_s})$ and the expression $\mathbf{E}(s)$ of spot s in the set of all spots S; that is, by the following objective function.

$$\text{Maximize} \quad \sum_{s \in S, \mathbf{M_s} \subset \mathcal{L}} \rho(\mathbf{E}(M_s), \mathbf{E}(s))$$

$$\text{Subject to} \quad k_{s,t} = \{c \in \mathbf{M_s} | t(c) = t\}, \ \forall t \in T$$

where $\mathcal{L} \in \mathbb{R}^{m \times n}$ is the expression matrix of the single-cell candidate library composed of m cells and n genes, $t(c)$ represents the type of cell c. The number of chosen cells from each type in $\mathbf{M_s}$ should be the same as the value of $k_{s,t}$.

The module adopts a heuristic method of two steps, initialization and *swapping* to optimize the objective function. The initialization selects top $k_{s,t}$ cells of type t for spot s according to the Pearson correlation coefficients between the spot and the cell from the single-cell candidate library.

If a better objective value is obtained, the swapping step swaps a cell in aggregates with a cell from the candidate library. We provide two optional cell sampling modes, one mode gathers cells from the same type, and another mode gathers cells from the whole candidate library. The process is repeated until convergence, or a predefined maximum number of iterations is achieved. The swapping process can be time-consuming, and we adopted a local sensitive hash (LSH) strategy to accelerate the swapping step (31), the runtime of an exemplary dataset can be found in (Supplementary information, Supplementary Table S4). During the swapping procedure, the module removes one cell from the aggregate $\mathbf{M_s}$ at spot s randomly, denoting the aggregate after the removal as $\mathbf{M_s}'$. The module chooses a new cell $\mathbf{m}$ in each iteration to further increases the $\rho(\mathbf{E}(\mathbf{M_s}' \cup \{\mathbf{m}\}), \mathbf{E}(s))$. It can be chosen by querying a cell in LSH that has the highest correlation with $\mathbf{E}(s) - \mathbf{E}(\mathbf{M_s}')$. The module performs feature selection (32) on the single-cell candidates before the aggregate construction to reduce the noise introduced by sequencing and low variable genes by choosing the top 3,000 highly variable genes and 80% highly variable LR genes to maintain the capability to infer the intercellular affinity.

Measuring the intercellular affinity by ligand–receptor interactions in single-cell profiles

We gather the selected single-cell expression profile of all spots to form a new matrix denotes by $\mathbf{T} \in \mathbb{R}^{r \times n}$ consisting of r cells and n genes. We collect our human ligand–receptor (LR) pairs from the curated database under the FANTOM5 project (33), and the mouse LR pairs from the CellTalkDB (34). With collected LR pairs, we extract the expressions of LR genes from $\mathbf{T}$ and form the ligand and receptor expres-

sion matrices as $\mathbf{T}_L$ and $\mathbf{T}_R \in \mathbb{R}^{r \times n_{lr}}$ where n_{lr} denotes the number of LR pairs. Specifically, for the ith LR pair, the ith columns of $\mathbf{T}_L$ and $\mathbf{T}_R$ are the corresponding ligand and receptor expressions subset from $\mathbf{T}$, respectively. The multiplication of the two expression matrices yields the affinity between each pair of cells suggested by the co-expression of each LR pair.

As a cell can simultaneously express both ligand and receptor genes, we have two symmetric terms $\mathbf{A}_1 = \mathbf{T}_L\mathbf{T}_R^T$ and $\mathbf{A}_2 = \mathbf{T}_R\mathbf{T}_L^T$ representing two possible LR orders in each cell pair. We formulate the *initial affinity matrix* $\mathbf{W}$ as $\mathbf{A}_1 + \mathbf{A}_2 = \mathbf{T}_L\mathbf{T}_R^T + \mathbf{T}_R\mathbf{T}_L^T$ of size $r \times r$.

Reducing the pseudo affinities to refine the affinity matrix by sparsification

The initial affinity matrix includes pseudo affinities between distant or indirectly interacting cells. Here we present three different approaches for diminishing the pseudo affinities, that is, partial correlation, spectral graph sparsification, and spatial coordinates refinement for ST coupled datasets.

We first adopt partial correlation to reduce the pseudo affinities for initial affinities of high variance (27). Two distant cells could have a high affinity value due to the share of strong interaction with another cell. Partial correlation is widely used in network analyses such as protein contact predictions and brain region interactions as it measures the latent variables representing direct causation and thus removes indirect relationships among entities (35–37). While a covariance matrix represents the relations between any two entities, the inverse of a covariance matrix, also known as the precision matrix, approximates the partial correlations among entities (38). For the block expression matrix $\mathbf{T}_{LR} = \begin{pmatrix}\mathbf{T}_L \\ \mathbf{T}_R\end{pmatrix}$, we denote its covariance matrix as $\mathbf{K}$, that is, $\mathbf{K} = \mathbf{T}_{LR}\mathbf{T}_{LR}^T$. In particular, we have the block form of $\mathbf{K} = \begin{pmatrix}\mathbf{S}_L & \mathbf{A}_1 \\ \mathbf{A}_2 & \mathbf{S}_R\end{pmatrix}$, where $\mathbf{A}_1 = \mathbf{T}_L\mathbf{T}_R^T$, $\mathbf{A}_2 = \mathbf{T}_R\mathbf{T}_L^T$, and $\mathbf{S}_L = \mathbf{T}_L\mathbf{T}_L^T$ and $\mathbf{S}_R = \mathbf{T}_R\mathbf{T}_R^T$ represent the ligand and receptor gene expression similarity between any two cells. We could distinguish direct and indirect LR interactions among cells and keep the direct ones by using the precision matrix of $\mathbf{K}$, that is, $\mathbf{K}^{-1} = \begin{pmatrix}\mathbf{K}_{11}^{-1} & \mathbf{K}_{12}^{-1} \\ \mathbf{K}_{21}^{-1} & \mathbf{K}_{22}^{-1}\end{pmatrix}$. Therefore, we have $\mathbf{W} = \mathbf{K}_{11}^{-1} + \mathbf{K}_{12}^{-1} + \mathbf{K}_{21}^{-1} + \mathbf{K}_{22}^{-1}$ representing direct LR interactions.

Sparsification of the affinity matrix aims to keep the association between adjacent cells while reducing affinities with distant ones. Sparsification based on k-nearest neighbor consider local neighborhood relationships by values. There is no proof that the spectral features of the graph are preserved (39). While spectral graph sparsification maintains high spectral similarity while reducing unnecessary edges between distant cells by finding a sparse approximation of the original graph (28). Therefore, we apply the Spielman-Srivastava spectral graph sparsification algorithm (40) to remove pseudo affinities. We build the affinity graph $\mathbf{G}$ by regarding the cells as vertices and the cell-cell affinity as the edge weight. When the context is clear, we also refer to $\mathbf{W}$ as the adjacency matrix for $\mathbf{G}$ for notation simplicity. We further denote the Laplacian matrix of $\mathbf{G}$ as $\mathbf{L}$. In the Spielman-Srivastava algorithm, the effective resistance, that is, the distance between two vertices connected by an edge is proportional to the reciprocal of its edge weight. In the sparsification step, edges are sampled by the probabilities proportional to their effective resistances. The algorithm preserves the spectrum of the graph Laplacian, that is, the eigenspaces spanned by eigenvalues, and their relations by requiring high similarity between the two Laplacian matrices, while some previous works only maintain the span of the dominant eigenvectors (41,42). We define the effective resistance between two cells u and v as

$$\mathrm{Reff}(u, v) = (\delta_u - \delta_v)^T\mathbf{L}^{-1}(\delta_u - \delta_v) \tag{1}$$

where $\delta_u \in \{0, 1\}^r$ is the indicator vector of vertex u. Following the definition, the sparse graph preserves the crucial edges of the original graph. We sample the edge (u, v) by the probability $p_{u,v} = min\{1, C \cdot (\log r) W_{u,v} \cdot \mathrm{Reff}(u, v)/\epsilon^2\}$, where C is some constant and ϵ is the approximation parameter. We further adjust the weight of the sampled edge (u, v) as $W_{u,v}/p_{u,v}$. We determine the value of the term C/ϵ^2 by the user-defined proportion of preserved edges $q = 2\sum_{u,v} p_{u,v}/r(r-1)$. Since the expected number of chosen edges can be bounded by

$$\sum_{u,v} p_{u,v} = \sum_{u,v} min\{1, C \cdot (\log r) W_{u,v} \times \mathrm{Reff}(u, v)/\epsilon^2\} \leq \frac{Cr \log r}{\epsilon^2} \tag{2}$$

where $\frac{C}{\epsilon^2} \geq \frac{\sum_{u,v} p_{u,v}}{r \log r} = \frac{q(r-1)}{2 \log r}$, thus by adjusting the parameter q we can control the percentage of preserved edges.

Moreover, we utilize the spot coordinates in the coupled spatial data as one sparsification approach. If two cells belong to nonadjacent spots, the affinity between them is considered to be pseudo affinities.

Reconstructing the quasi-structure with fuzzy set cross-entropy embedding

The embedding of a cell-cell affinity graph to a low-dimensional space consists of two stages: (a) forming a topological representation $\mathbb{W}$ of sparsified the cell-cell affinity $\mathbf{W}$; and (b) finding an embedding $\mathbb{E}$ in the low-dimensional space of the topological representation to minimize the discrepancy between the embedding and the representation. A reliable topological representation of $\mathbf{W}$ should maintain the affinity relations while restricting the number of neighbors for each cell. Here, we maintain the top k_n affinities in $\mathbf{W}$ for each cell while setting other values to be zeros. Subsequently, we perform min-max normalization on the remaining affinities to obtain the membership strength in the range of [0, 1], denoting the matrix as $\mathbb{W}$. The fuzzy simplicial set expands the classical binary definition of membership by allowing continuous membership strength in the range of [0, 1] (43), and the union of the fuzzy simplicial sets (44) yields the fuzzy topological representation. Hence, $\mathbb{W}$ is the fuzzy topological representation of $\mathbf{W}$.

Subsequently, we apply strategies from UMAP (45) to minimize the fuzzy set cross-entropy between the embed-

ding $\mathbb{E}$ and the topological representation $\mathbb{W}$, that is,

$$CE(\mathbb{E}, \mathbb{W}) = \mathbf{P}_{\mathbb{E}} \log \frac{\mathbf{P}_{\mathbb{E}}}{\mathbf{Q}_{\mathbb{W}}} + (1 - \mathbf{P}_{\mathbb{E}}) \log \frac{1 - \mathbf{P}_{\mathbb{E}}}{1 - \mathbf{Q}_{\mathbb{W}}} \quad (3)$$

where $\mathbf{P}_{\mathbb{E}}$ and $\mathbf{Q}_{\mathbb{W}}$ represent the min-max normalized adjacency matrices of $\mathbb{E}$ and $\mathbb{W}$, respectively. We use a spectral layout, that is, the Laplacian matrix of $\mathbb{W}$ to as the initial Cartesian coordinates of $\mathbb{E}$ (46). By regarding edges as attractive forces and vertices as repulsive forces, we alternatively apply the attractive and repulsive forces until $CE(\mathbb{E}, \mathbb{W})$ converges to a local minimum.

Evaluating the reconstruction performance of SPROUT

A major metric for assessing the quality of the reconstructed spatial structure is its reproduction of the spatial characteristics of the tissue. Given a spatial structure of cells, we construct a fixed-volume neighbor graph, where the radius is the median distance between any cell to its third-nearest neighbor. According to the fixed-volume neighbor graph, we quantify the spatial characteristics as the number of neighboring pairs between any two cell types, indicating whether the two are enriched or depleted near each other. Therefore, we evaluate the cell type enrichment or depletion discrepancy by the Kullback–Leibler (KL) divergence (47) of the neighboring pair numbers for any two cell types between a given spatial structure and the embedding structure. To further evaluate the statistical significance of observed possible enrichment or depletion, we compare the number of neighboring pairs with 1000 random permutations of the cell type labels. We test the enrichment hypothesis, that is, the observed number of neighboring pairs is larger than the random expectation by P-values from both the right-tailed and left-tailed tests. We further adjust the P-values following the Benjamini-Hochberg procedure (48) and obtain the q-values with a cutoff of 0.05 for significance.

Revealing the dominating LR pairs contributing to intercellular affinity

Given a pair of cell i and j in the expression matrix $\mathbf{T}$, the expression profile is $\mathbf{T}_i$ and $\mathbf{T}_j$, the contribution from the k-th LR pair to the total cell-cell interacting affinity can be formulated as:

$$b_k^{ij} = \frac{\mathbf{T}^i_{L_k} {\mathbf{T}^j_{R_k}}^T + \mathbf{T}^i_{R_k} {\mathbf{T}^j_{L_k}}^T}{\mathbf{T}^i_L {\mathbf{T}^j_R}^T + \mathbf{T}^i_R {\mathbf{T}^j_L}^T} \quad (4)$$

where $\mathbf{T}^i_{L_k}$ denotes the ligand gene expression of kth LR pair in cell i, $\mathbf{T}^j_{R_k}$ denotes the receptor gene expression of kth LR pair in cell j. The contribution of each LR pair between two cell types t_1 and t_2 is calculated by:

$$b_k^{t_1,t_2} = \frac{1}{N} \sum_{i \in t_1, j \in t_2} b_k^{ij} \quad (5)$$

where N is the number of neighboring cell pairs between t_1 and t_2.

Assessment of the state of the art resolution enhancement methods and simulation datasets construction

Comparison of state of the art resolution enhancement methods. We perform SPROUT, BayesSpace (10), CellTrek (49) and SpaOTsc (26) on a mouse cortex dataset (50,51). We run BayesSpace spatialEnhance function following the default parameters in the BayesSpace package vignette. Given the qplot result of BayesSpace, we set the estimated cluster number as eight. We select sets of marker genes for each of 15 cell types and predict their expression at subspot resolution using enhanceFeatures of BayesSpace. To assign the cell type to each enhanced subspot, we assign the cell type by the highest scaled enhanced cell-type expression in each subpot. For CellTrek, we follow the same protocol in its Quick Tour. For SpaOTsc, we follow the parameters in its short tutorial. We acquire the cost matrix by mapping the datasets to a common low dimensional space by its recommending software—scanorama (52) and calculated their Euclidean distance in the low dimensional space.

Simulated ST datasets. Our simulated ST datasets were generated by the combined expression profiles of cells drawn from the mouse hippocampus scRNA-seq data (53). We utilize cells from mice age 28 postnatal days (P28) as both the scRNA-seq reference data to build the ST simulation data and the paired candidate library. Meanwhile, we use the mice aged 60 postnatal days (P60) as the unpaired candidate library. We randomly sampled cells from the scRNA-seq reference and summed the expression of each gene in the selected cells to generate the expression profile of the ST simulation data. To underline the adaptability of the preprocessing module regarding various spot internal densities, we simulate ST data with 1–10, 30 and 40 cells per spot. Moreover, to assess preprocessing module performance in dealing with different cell mixing ratios within the spot, the ST simulation was generated by the combined expression profiles of cells with different astrocyte ratios ($per = 1, 0.8, 0.5, 0.2, 0$) of astrocytes and neurons, since astrocytes and neurons are the main constituents of the brain and their interactions are crucial to the central nervous system (54). Each preliminary simulated ST data consist of 40 spots with the same cell per spot and astrocyte ratio. We concatenated the preliminary simulated ST datasets with the same cell number and different astrocyte ratios together by a sequence of decreasing astrocyte ratios, forming simulated ST data with 200 spots to simulate the transition of two cell types. We generated five repeats for each simulation with a various sampling of cells from the scRNA-seq reference data. We performed SPROUT on the simulated ST data with paired and unpaired candidate libraries.

Correlation measures the degree of similarity between two datasets while mean squared error (MSE) quantifies their difference in absolute value, however, correlation prone to reduce the noise (55). Therefore, to quantitatively evaluate the cell selection performance of preprocessing module, we calculated the Pearson correlation and the MSE between the summed expression of the selected cell of each spot with its corresponding ST simulations.

Furthermore, to assess the performance of SPROUT, we computed the correlation of pairwise distance between the SPROUT reconstruction and the simulated spatial structure.

RESULTS

Overview of SPROUT algorithm: reconstructing spatial organization at single-cell resolution from the spatial transcriptome

SPROUT provides a preprocessing module for ST datasets that select and aggregate single-cell profiles representing the expression profile of each spot. For a spot of the spatial data, the module derives the quantities of each cell type by deconvoluted cell type proportions produced by deconvolution software and a prespecified parameter ℓ_s representing the average number of cells in a spot (Figure 1A). The comparison of SPROUT processing human lymph node and mouse brain dataset (29) with the deconvoluted cell type proportions produced by stereoscope, RCTD, SPOTlight, and cell2location displays a small difference (Supplementary information, Supplementary Figure S1). The module then aggregates a set of single cells agreeing with the derived quantities and maximizing the correlation between the aggregated cell expression profile and the ST spot. Note that if the paired single-cell data are unavailable, we can use a labeled single-cell candidate library of similar tissue to create aggregations and achieves similar performance (Figure 1 B, Supplementary information, Supplementary Figure S3). Cells interact with proximal cells, and in this work, we use the term *affinity* as the measurement for the interaction strengths between interacting cells. We can build a cellular spatial configuration, termed *quasi-structure*, from the affinity values. We first assume that the cell-cell affinity can be estimated by the concentration of LR complexes which can be approximated by their mRNA abundance. Furthermore, we assume that cells compete for space because of the limitation of biological constraints. SPROUT has no prior knowledge of cell proximity when forming the initial affinity matrix. It calculates the affinity value between any two cells. Therefore, the approximated affinities based on the first assumption contain pseudo affinities between distant or indirect interacting cells. Following the above assumptions, SPROUT reconstructs the quasi-structure from scRNA-seq data with four steps: (a) establishing the initial affinity matrix by the LR expression profiles, which falsely includes the pseudo affinities between distant or indirect interacting cells; (b) constructing an affinity graph regards cells as vertices and the initial affinity matrix as the adjacency matrix; (c) reducing the underlying pseudo affinities in the initial affinity graph by partial correlation, spectral graph sparsification and spatial coordinates refinement and (d) embedding the sparsified affinity graph into a low-dimensional space as the quasi-structure in the single-cell resolution.

SPROUT approximates the cell-cell affinity by the mRNA abundance of interacting LR pairs (Figure 1C). For initial affinities of high variances, SPROUT replaces the initial cell-cell affinity matrix with the precision matrix to reduce the indirect correlations for subsequent procedures (Figure 1D). SPROUT reduces the pseudo affinities from the initial affinity matrix by imposing spectral graph sparsification and spatial coordinates refinement on the affinity matrix (Figure 1D). SPROUT adopts a local fuzzy set (LFS) embedding method to embed the processed affinity matrix in a low-dimensional space. The LFS step first builds a fuzzy topological representation from the processed affinity matrix, limiting the number of neighbors required by the second assumption (Figure 1E, top panel). Subsequently, the LFS step optimizes the representation in the low-dimensional space by minimizing the fuzzy set cross-entropy between the two representations (Figure 1E, bottom panel). SPROUT can take the curated single-cell aggregates, yielding the reconstructed quasi-structure for downstream analyses (Figure 1F). The embedding result, that is, the reconstructed quasi-structure by SPROUT, facilitates further evaluation of discovering dominant ligand–receptor pairs between neighboring cells at single-cell resolution (Figure 1G). Furthermore, with proper sparsification, SPROUT is capable of *de novo* reconstruction from the single-cell transcriptome. In the head and neck cancer (HNC) scRNA-seq dataset, SPROUT recapitulates the quasi-structure features which are commonly observed in the partial epithelial to mesenchymal transition (p-EMT) process: (a) p-EMT cells located at the interface between malignant cells and cancer-associated fibroblasts (CAF) cells; (b) CAF-1 cells presenting at closer proximity to the p-EMT cells compared to CAF-2 cells; (c) malignant cells showing minimum interactions with immune cells due to immune evasion (Figure 1H).

Assessing the performance of SPROUT in processing ST datasets

Benchmarking SPROUT on real and simulation datasets. We apply SPROUT to public mouse cortex scRNA-seq (50) and ST datasets from 10× (51). The scRNA-seq data contains 4785 cells from 15 types. The ST dataset contains 1075 spots. We then compared SPROUT on a mouse cortex dataset with three additional resolution enhancement methods: (i) BayesSpace (10), which imposes a Bayesian model on the spatial neighborhoods of spots to enhance the resolution of spots; (ii) CellTrek (49), which achieves single-cell spatial mapping by the embedding of spatial and single-cell transcriptomics profiles and metric learning and (iii) spots (26), which maps cells by optimally transporting. SPROUT and CellTrek reconstructed the original spatial pattern of the mouse cortex, while BayesSpace enhanced the resolution *in situ* and SpaOTsc captures only a part of spatial structure (Figure 2A). Compared with CellTrek, the reconstruction of SPROUT spread evenly and has a higher density, which has high consistency, $r = 0.99$, with its coupling ST spot structure, regarding the pairwise distance. The cell-type proximity summarized by cell locations is vital for downstream analyses. Thus, the recapitulation of such information should also be a metric for evaluating the reconstructed quasi-structure. Specifically, we use KL-divergence to assess the difference in the cell-type proximity between the original and reconstructed structure. SPROUT achieves the lowest KL-divergences among the four approaches. Moreover, we calculate the *K*-distance between neuron cells of all layers to the L2/3 IT cells. The gradually

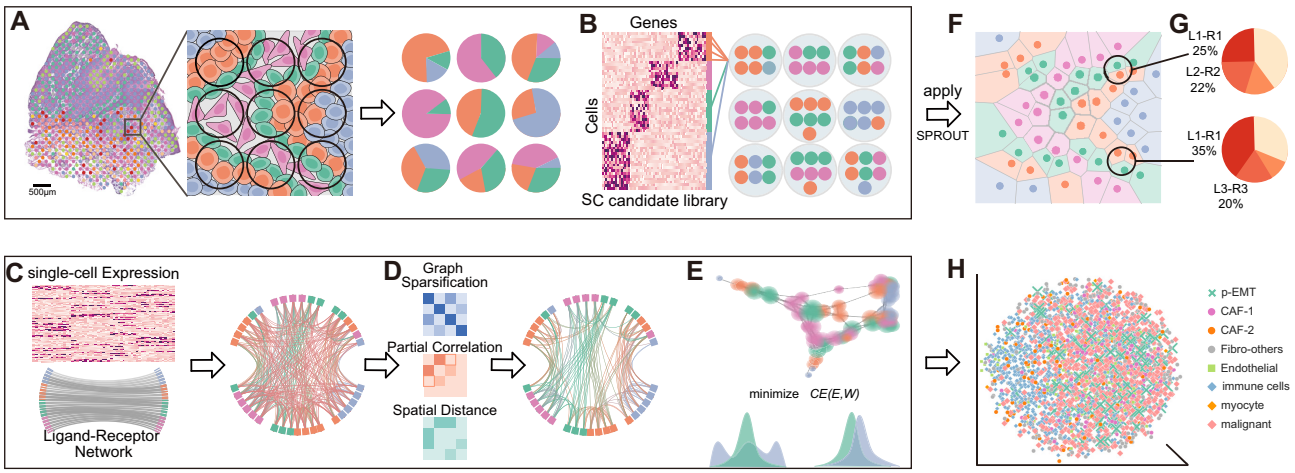

Figure 1. Schematics of SPROUT. (A–C) Workflow of the preprocessing module. (**A**) The preprocessing module of SPROUT adopts existing deconvolution software to decompose cell-type mixtures of ST profiles. (**B**) The preprocessing module selects a designated amount of cells from the single-cell candidate library, equal to the estimated cell number per cell type in each spot. (C–E) Workflow of the SPROUT. (**C**) SPROUT derives the initial cell-cell affinity graph from the single-cell profiles by the LR interactions. (**D**) SPROUT applies partial correlation, spectral graph sparsification, and spatial coordinates refinement on the cell-cell affinity graph to reduce pseudo affinities. (**E**) SPROUT utilizes LFS embedding to embed interactions into a low-dimensional space. (**F**) The 2D embedding of the selected single cells reconstructed by SPROUT. (**G**) The determination of dominant ligand–receptor pairs between neighboring cells at single-cell resolution. (**H**) The 3D embedding of the HNC data reconstructed by SPROUT.

rising K-distance results of SPROUT and CellTrek showed a well-captured spatial structure of the mouse cortex layers, while BayesSpace fails due to the overspread of L5 PT cells (Figure 2B). The cell-cell connection between layers in SPROUT's reconstruction shows a clear connection between each anatomically adjacent layer while CellTrek fails to fully capture (Figure 2C). The feature genes of the main layers (*Calb*, *Fam19al*, *Crym*, *Cplx3*) also express in the corresponding regions in the quasi-structure (Figure 2D).

We next evaluate the performance of SPROUT and its preprocessing module on simulated datasets. The spot internal density varies according to the tissue density and the spot diameter, thus the spot internal density varies (24,51,56). We generated the simulated datasets to reflect the validity and robustness of SPROUT under two simulation conditions: spot internal density and cell mixing ratio within the spot.

We perform SPROUT and its preprocessing module with both paired and unpaired single-cell candidate libraries. We quantitatively evaluated the performance of SPROUT by the expression correlation between the summed expression of selected cells and its corresponding simulated spatial expression, and the shape correlation of pairwise distance between SPROUT reconstruction and simulated spatial structure. Under different simulation conditions, the summed expression profiles of each single-cell aggregate achieve an average Pearson correlation coefficient $r = 0.94$ with their corresponding ST profiles for the paired library, and $r = 0.90$ for the unpaired library (Supplementary information, Supplementary Figure S2A). Meanwhile, the difference in average MSE of genes is higher in the unpaired datasets (Supplementary information, Supplementary Figure S2A). The expression correlation increase with the growth of the actual cell number per spot, since the fault tolerance of choosing a similar cell increased. Moreover, to assess the stability of choosing different parameters ℓ_s, we perform the preprocessing module with parameters ℓ_s of 5, 10, 20 and 40 on the simulated ST dataset generated with ten cells per spot. The expression correlation between the simulated ST datasets and sc aggregates under various choices of ℓ_s has a small difference (Supplementary information, Supplementary Figure S2B).

Subsequently, SPROUT reconstructs the quasi-structure from the selected single-cell aggregates. The quasi-structure has a concordant inter-spot organization as the cells originating from the same spot remain in the same compartment (Supplementary information, Supplementary Figure S2C). The quasi-structure of each simulation reached a high average shape correlation of $r = 0.97$ with the simulated spot organization (Supplementary information, Supplementary Figure S2D). Furthermore, we compare the shape correlation between various parameters, the average correlations decrease with the increase in cell number per spot (Supplementary information, Supplementary Figure S2D). The shape correlation of the paired data dropped from 0.97 to 0.95 in 40 cells per spot, which may caused by the overcrowding in the spot. Furthermore, the expression pattern of *Apoe* and *Ppp3ca*, the feature gene of astrocytes and neurons, is concordant in the spatial structure and the reconstructed quasi-structure, achieving expression correlation at 0.96 and 0.93 (Supplementary information, Supplementary Figure S2E). The high-quality single-cell aggregates and the quasi-structure demonstrate the accuracy and robustness of the preprocessing module of SPROUT.

SPROUT reconstructed a high-quality quasi-structure for the mouse hippocampus dataset. We apply SPROUT to reconstruct the single-cell resolution quasi-structure for the mouse hippocampus dataset. The spatial data provided by stereoscope (21) contains 609 spots, and the single-cell candidate library from the *mousebrain.org* contains 8449 cells,

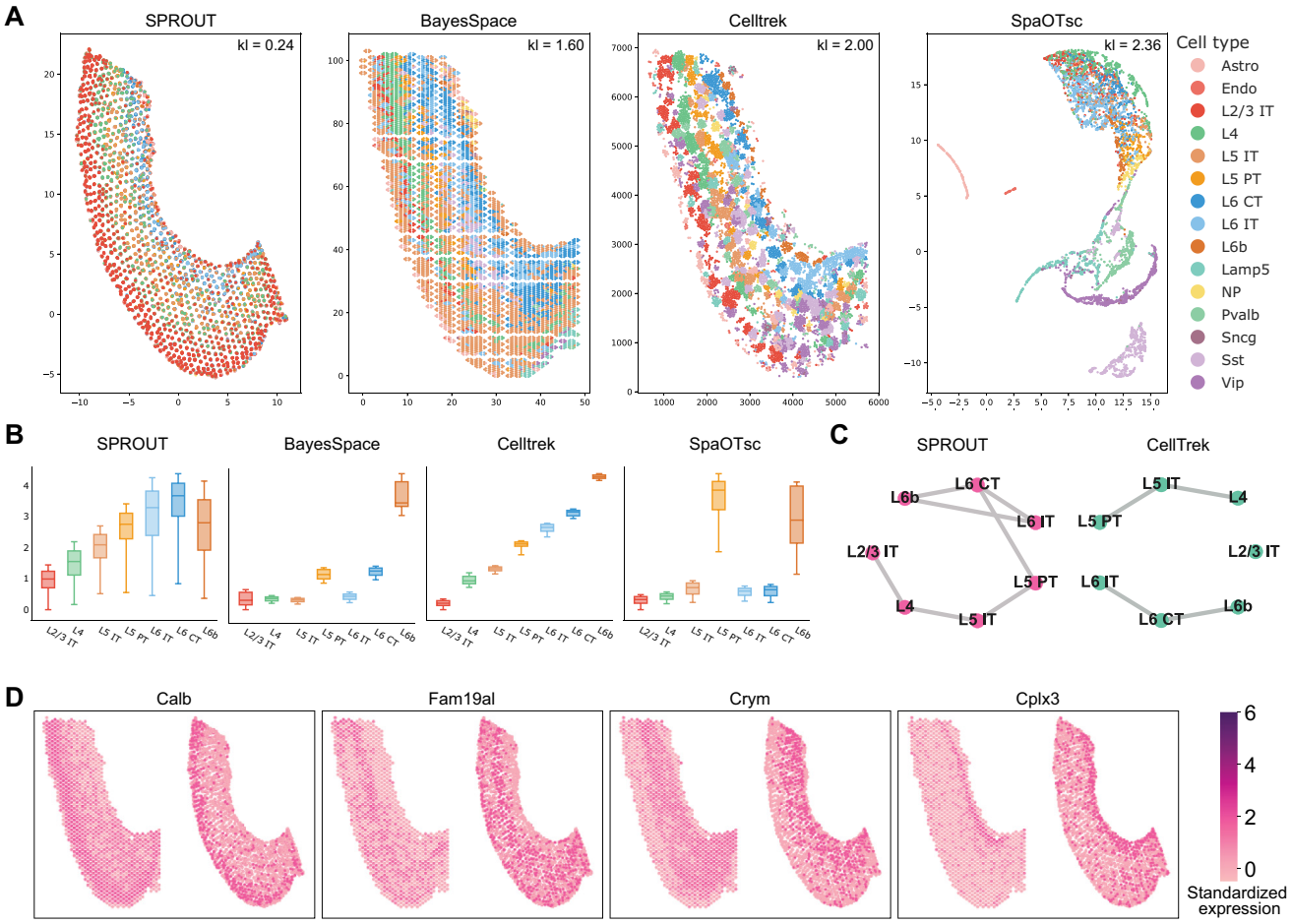

Figure 2. Benchmarking SPROUT on reconstructing the spatial organization in a mouse cortex tissue. (**A**) Comparison of SPROUT, BayesSpace, CellTrek, and SpaOTsc results for reconstruction in a mouse cortex tissue. (**B**) *K*-distance of neuron cells to L2/3 IT cells based on the result of each method. (**C**) Comparison of SPROUT (left) and CellTrek (right) results of normalized significant colocalization graph using permutation tests. (**D**) The standardized gene expression of layer feature genes in ST (left) and reconstructed quasi-structure of SPROUT (right).

re-clustered and annotated by stereoscope, covering 56 subtypes across seven major groups.

We use the deconvolution software package to infer $p_{s,t}$. Based on the deconvolution result of *stereoscope* (21), the preprocessing module of SPROUT then selects 6071 cells that constitute 609 single-cell aggregates ($\ell_s = 10$) from the single-cell candidate library to represent the expression profile of ST spots. Then SPROUT reconstructs the quasi-structure from the selected single-cell aggregates (Figure 3A). The reconstructed quasi-structure of SPROUT achieves a 0.97 Pearson correlation with its coupling ST spots in the pairwise distance (Supplementary information, Supplementary Table S1).

Furthermore, we calculate the expression correlation of each gene between ST and single-cell aggregates. *Cnp*, *Plp1* and *Ppp3ca*, achieve high expression correlations, $r = 0.71$, $r = 0.70$, $r = 0.65$, between ST spots and single-cell aggregates (Figure 3B, Supplementary information, Supplementary Table S3). Meanwhile, the aggregated expression profile of each single-cell aggregate achieves a median Pearson correlation coefficient $r = 0.66$ with their corresponding ST profiles (Supplementary information, Supplementary Figure S3).

The quasi-structure achieves a low KL-divergence, 0.067, in the cell-type proximity (Figure 3C, Supplementary information, Supplementary Table S2). Moreover, we assess the effectiveness of each step in SPROUT by comparing the KL-divergence with different combinations of embedding and sparsification methods (Figure 3 C). Comparing the LFS embedding that SPROUT utilizes with constrained t-SNE used by CSOmap, the lower median KL-divergence and higher shape correlation in the combination of LFS embedding with a sparsification method is demonstrated (Figure 3C, Supplementary information, Supplementary Figure S4). For sparsification methods, spectral graph sparsification partially reduces the pseudo affinities in the cell-cell affinity matrix, hence achieving a smaller median KL-divergence compared to the hard-filtering method of keeping the top fifty high-affinity edges for each node. The additional distance metric provided by spatial information effectively reduces more pseudo affinities in the cell-cell affinity graph, leading to a smaller median KL-divergence. The

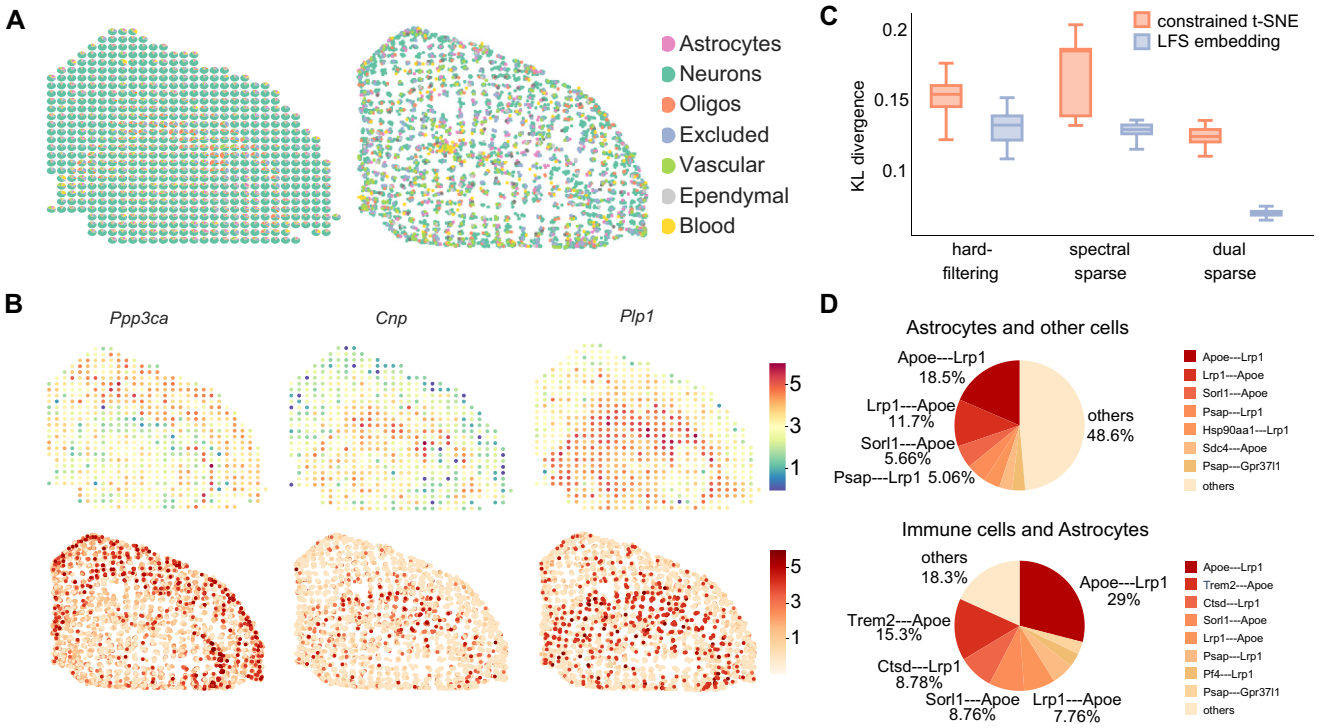

Figure 3. The reconstructed quasi-structure of mouse hippocampus. (**A**) The 2D visualization of the ST spots (left) and the reconstructed quasi-structure of the mouse hippocampus (right), colored by cell types. (**B**) The standardized gene expression of exemplary genes in ST (top) and reconstructed quasi-structure (bottom). (**C**) The cell-type proximity KL-divergence for the combinations of two different embedding methods and three sparsification methods. (**D**) The pie charts of the LR pair contributions to the interactions of astrocytes with all other cells (top) and with only immune cells (bottom).

smallest KL-divergence, 0.067, is acquired in the combination of LFS embedding and dual sparsification, which suggests the validity of each step in SPROUT.

The well-captured neighboring information in the reconstructed quasi-structure enables identifying the driver LR pairs mediating interactions between cell types. We only consider the cell-cell communication between neighboring cells, which increases the credibility of the evaluation of the dominant force in cell-cell communication. In the reconstructed quasi-structure, we observe that the interactions between lipoprotein receptor-related protein 1 (*Lrp1*) and apolipoprotein E (*apoE*) is the leading interactions among neurons, vascular cells, and astrocytes (Figure 3D, Supplementary information, Supplementary Figure S5). LRP1 mediates the metabolism of Amyloid-beta ($A\beta$), whose accumulation is a vital pathogenic element of Alzheimer's disease. Yet apoE can block the LRP1-mediated pathway in astrocytes, hindering the clearance of $A\beta$ (57). Hence, certain immunotherapy targeting apoE has been applied to APP/PS1 mice to meliorate the accumulation of $A\beta$ (58). The reveal of the fundamental interaction between Lrp1 and apoE in our quasi-structure consolidates the validity of SPROUT and, therefore, its capability of providing valuable biological insights.

SPROUT uncovers the dominating LR pairs mediating intercellular interactions in human diseases. High-quality reconstructions of SPROUT help reveal the underlying molecular mechanisms of human diseases. To evaluate the performance of SPROUT processing heterogeneous tissues, we apply SPROUT on two spatial datasets paired with scRNA-seq data from pancreatic ductal adenocarcinoma (PDAC) and human squamous cell carcinoma (SCC) patients. The spatial data of the PDAC dataset contains 428 spots, and the scRNA-seq data contains 1926 cells annotated by 17 cell types from Moncada *et al.*'s work (24). The spatial data of the SCC dataset contains 666 spots, and the matching scRNA-seq data contains 2689 cells across 14 cell types (25). The ST and scRNA-seq data are processed from the same malignant tissue.

Based on the deconvolution result of *stereoscope* (21), the preprocessing module of SPROUT curates 4289 and 6625 cells with replacement ($\ell_s = 10$), constructing single-cell aggregates to represent the expression profile of each spot in the spatial data for PDAC and SCC dataset. The feature gene of the main regions identified in Moncada *et al.*'s work, *CRISP3*, *PRSS1*, *TM4SF1*, also express in the corresponding regions in the quasi-structure (Figure 4 A). Moreover, we calculated the expression correlation of each gene between ST and single-cell aggregates (Supplementary information, Supplementary Table S3). Several cell-type marker genes annotated in Andrew *et al.*'s work, for example, *CALML5*, *SPRR1B*, *KRT2*, achieve high expression correlations, $r = 0.79$, $r = 0.65$, $r = 0.61$, between ST spots and single-cell aggregates (Figure 4B).

Subsequently, SPROUT rebuilds the quasi-structure from the curated single-cell aggregates. Given the high variance in the affinity values of the PDAC dataset, SPROUT reconstructs the quasi-structure of the curated single-cell aggregates with the precision matrix form of affinity ma-

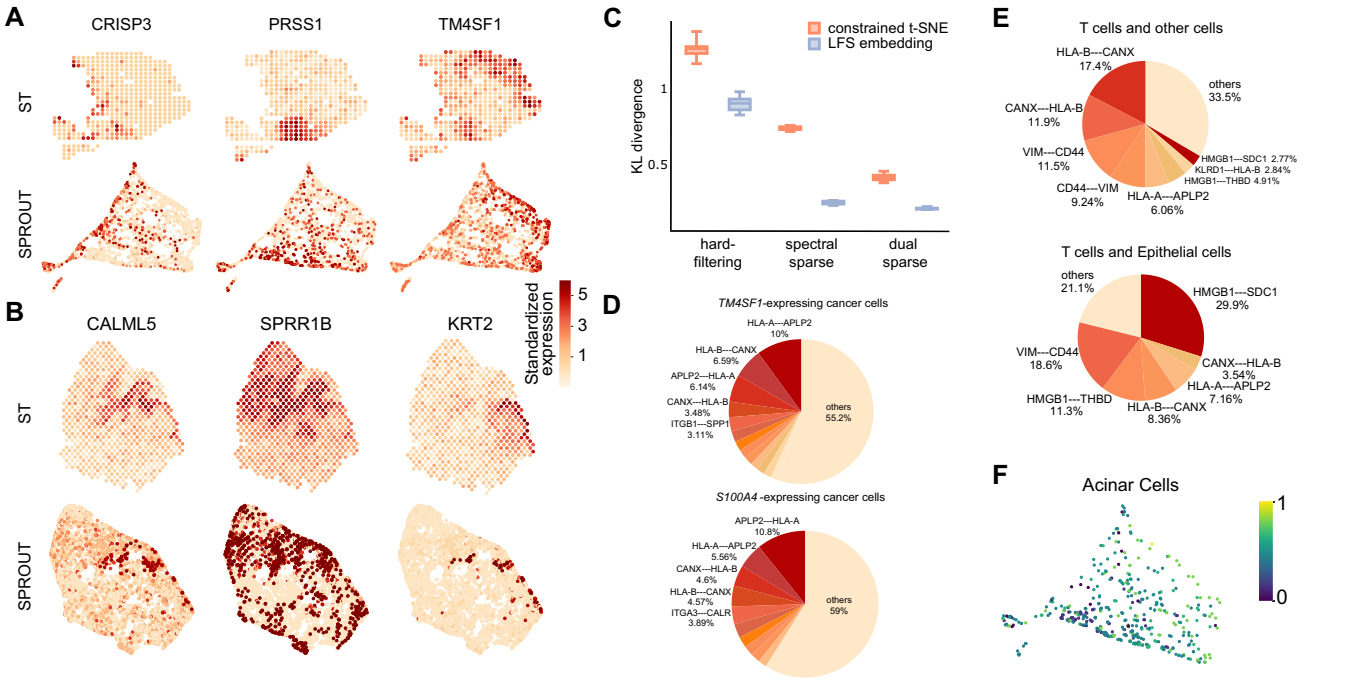

Figure 4. Performance of SPROUT in rebuilding quasi-structure from PDAC dataset. (**A**) The standardized expression of region-specific genes in ST (top) and single-cell aggregates (bottom) in PDAC dataset. (**B**) The standardized expression of region-specific genes in ST (top) and single-cell aggregates (bottom) in SCC dataset. (**C**) The cell-type proximity KL-divergence for the combination of three sparsification methods and two different embedding methods. (**D**) The pie chart of LR pair contributions in *TM4SF1*- and *S100A4*-expressing cells. (**E**) The pie charts of the LR pair contributions to the interaction of T cell and other cells (top), T cell and epithelial cells in particular (bottom). (**F**) Trajectory analysis for acinar cells in PDAC dataset.

trix. The reconstructed quasi-structures achieve high similarity, $r = 0.93$ and $r = 0.91$, of the pairwise distance with its coupling spatial data for PDAC and SCC datasets (Supplementary information, Supplementary Figure S6, Supplementary information, Supplementary Table S1). Furthermore, the quasi-structure achieves low KL-divergence of 0.13 and 0.42 in the cell-type proximity between the original and the quasi-structure for the PDAC and SCC dataset (Figure 4C, Supplementary information, Supplementary Table S2). When comparing across different combinations of embedding and sparsification methods, SPROUT also achieves the smallest median KL-divergence while combining dual sparsification and LFS embedding, which emphasizes the stability of SPROUT on cancer datasets. The preprocessing procedure of SPROUT used 20 minutes to process the PDAC dataset, the average runtime of one embedding is 9.59 s (Supplementary information, Supplementary Table S4).

Tumor heterogeneity has been an obstacle to cancer therapy since mutant clones escape and thrive from the targeted therapy. Our spatially informed single-cell transcriptome can characterize the driver interactions between distinct subpopulations. Leveraging the high-quality quasi-structure SPROUT reconstructed for the PDAC dataset, we observe that the interaction between HLA-A and APLP2 contributes around 16% to the overall interaction potential in both *TM4SF1*- and *S100A4*-expressing cancer cells (Figure 4D). APLP2 can cause a reduction in the expression of the total cell surface major histocompatibility complex (MHC) class I (59), which is a crucial molecule for cancer cell recognition and elimination. The high interaction between HLA-A and APLP2 observed in the quasi-structure indicates a potential immune escape mechanism adopted by both *TM4SF1*- and *S100A4*-expressing cancer cells. Expect for the mutual LR interactions, we also found distinct dominating LR pairs in these two cancer types (Figure 4D). The LR pair ITGB1-SPP1 is a major contributing factor to the interaction for *TM4SF1*-expressing cancer cells. SPP1 has been proved to abet immune escape in lung adenocarcinoma through its mediation on macrophage polarization (60). Experiments have also revealed how ITGB1-SPP1 interaction incites the cancer progression in ovarian cancer (61). Our finding suggests that the interaction between ITGB1 and SPP1 potentially triggers the immune escape of *TM4SF1*-expressing cancer cells. However, in *S100A4*-expressing cancer cells, the interaction between ITGA3 and CALR is more prevalent (Figure 4D, right). ITGA3 has been identified as a biomarker for diagnosing and prognostic predicting pancreatic cancer (62). The LR pair ITGA3-CALR has also been predicted as a poor-prognostic LR pair by other datasets from the same tissue in the recent work of Suzuki *et al.* (63). These discoveries demonstrate that researchers can characterize tumor heterogeneity with the high-quality quasi-structure by revealing the driver interactions between distinct subpopulations.

Subsequently, by evaluating the LR contribution to the cell-cell affinity in the SCC dataset, we identify the LR pair HLA-B-CANX as a driving force behind the interaction of T cells, constituting about 29% of the T cell affinities. Our finding is supported by a report regarding an

impaired CD8+ T cell-mediated immune response due to the disturbance in HLA-B-CANX interaction in colorectal cancer (64). Moreover, we investigate the dominating LR pairs facilitating the crosstalk between T and epithelial cells. We identify that the interaction between HMGB1 and SDC1 contributes around 30% to the affinity between T and epithelial cells. HMGB1 and SDC1 have been reported to associate with drug resistance in glioma (65). Furthermore, the increase in HMGB1 promotes tissue invasion and metastasis of cancer (66), and SDC1 influences the migration of mouse keratinocytes (67). Our finding connects HMGB1 with SDC1, indicating that the reported promotion of metastasis may result from the interaction between HMGB1 and SDC1. These discoveries demonstrate that the high-quality quasi-structure reconstructed by SPROUT facilitates disclosing the decisive LR interaction underneath the cell-cell communications. We also discover the spatial-temporal distribution of acinar cells (Figure 4F), which indicates the developmental trajectory of the acinar cells.

Evaluating the effectiveness of SPROUT on *de novo* reconstruction of single-cell datasets

We have demonstrated that the quasi-structure can be reconstructed from cell–cell affinity with proper sparsification. Therefore, we further evaluate the validity of SPROUT in reconstructing the spatial organization of scRNA-seq data without prior spatial structure.

SPROUT outperforms CSOmap on the hepatocellular carcinoma (HCC) dataset. We apply SPROUT on the HCC dataset consisting of 1329 cells from Ren *et al.*'s work, for which the reconstruction of CSOmap obtains a Spearman correlation of $r = 0.69$ in the cell-type proximity with the IHC image of the same tumor sample. Given the large variance in the initial affinity values of the HCC dataset, SPROUT rebuilds the quasi-structure with the precision matrix form of the affinity matrix. Compared with CSOmap, the reconstructed quasi-structure of SPROUT is visually less compact (Figure 5A) and achieves higher cell-type proximity, that is, a Spearman correlation of $r = 0.89$, with its IHC image reference (Figure 5B).

Subsequently, we evaluate the performance of combinations in embedding and sparsification methods regarding the cell-type proximity similarity (Figure 5B, Supplementary information, Supplementary Table S5). A higher correlation is observed in the combination of LFS embedding and any sparsification method when comparing LFS embedding with constrained t-SNE. Moreover, spectral graph sparsification reduces the pseudo affinities, achieving a higher correlation than the hard-filtering method. The comparison between different combinations reveals the collaborative contribution of LFS embedding and spectral graph sparsification for reconstruction.

The high-quality reconstructed structure enables investigations on intercellular regulatory mechanisms. The interaction between regulatory T cells (Tregs) and CD8+ T cells suggests an ongoing suppression of the immune response (68), during which Treg cells induce the p38 and ERK1/2 signaling pathways ineffective T cells, which initiate DNA damage, resulting in cell senescence (69). Consistent with the previous study, we observe an increase in the mRNA expression of ERK1 in the Treg-CD8+ T cell interacting area, indicating the potential of SPROUT in discovering the immune response signals hidden in the scRNA-seq data.

Furthermore, the well-captured cell-type proximity in the quasi-structure enables the analysis of the dominating LR pairs contributing to the cell-cell affinity. We analyze the main LR pairs between any two cell types. Specifically, we identified the difference in the dominating LR pair between Tregs and CD8+ T cells as well as between Treg and exhausted T cells, which indicates a distinct regulation mechanism of Treg in these two types of cells. *CCL5* is one of the signature genes identified in exhausted T cells (70). The contribution of CXCR3-CCL5 increases in the interaction between Treg and exhausted T cells compared with CD8+ T cells. Indicating that the Tregs originated expression of CXCR3 may trigger the exhaustion.

The discovery demonstrates that the high-quality quasi-structure reconstructed *de novo* by SPROUT promotes the reveal of the LR interaction underneath the cell-cell regulatory mechanism.

SPROUT recapitulates the signal transmission process in the developing human heart. We apply SPROUT on a human developing heart dataset consisting of 3717 cells from the 6.5 post-conception weeks (PCW) heart (71). We apply SPROUT to reconstruct the quasi-structure of the heart dataset. The 3D quasi-structure of the developing human heart demonstrates a compact structure (Figure 6 A, left). The atrial cardiomyocytes are spatially segregated from ventricular cardiomyocytes (Figure 6 A, middle), which is consistent with the separation of the atrium and the ventricle in anatomy (Figure 6 A, right). Moreover, we evaluate the cell-type proximity similarity between the quasi-structure and the *in situ* sequencing data. The quasi-structure achieves a high normalized Spearman correlation of $r = 0.68$ in the cell-type proximity.

When comparing the different combinations of embedding and sparsification methods, Figure 6B demonstrates that the reconstructed quasi-structure rebuilt by the combination of spectral sparsification and LFS embedding achieves the highest resemblance in cell-type proximity (Supplementary information, Supplementary Table S5). The cell-type proximity SPROUT recapitulated includes fibroblasts and cardiac cells (Figure 6C), enabling fibroblasts to modify gene and protein expression, and ultimately cardiac function (72). Ang II activates the paracrine secretion of TGF-β1 (*TGFB1*, transforming growth factor-β) and endothelin-1 (*EDN1*) in fibroblasts, leading to the cardiac myocyte hypertrophy (Figure 6D) (73). Angiotensinogen (*AGT*) is a precursor for angiotensin I, which will be eventually converted to Ang II for further activities (74). Therefore, we inspect the proximity of *AGT* high-express cell and *TGFB1*, *EDN1* high-express cell through the neighboring cell pair numbers between these cells in the quasi-structure (Figure 6D). We consider a pair of cells are neighboring if the distance is less than the median distance between any cell to its third-nearest neighbor. The proximity between cells that express critical signaling genes provides conditions for signaling through paracrine, consistent with the experimentally validated signaling pathway. This con-

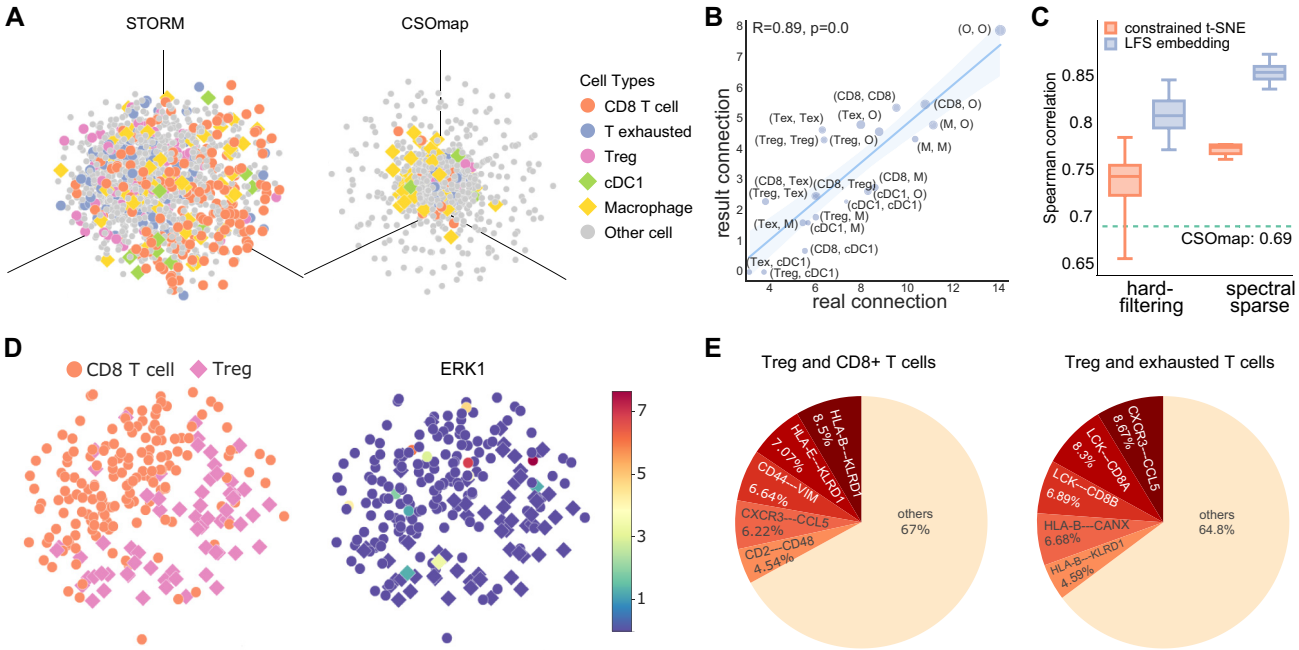

Figure 5. Application of SPROUT in restoring the quasi-structure of HCC. (**A**) The 3D embedding of the reconstructed quasi-structure of SPROUT (left) and the prediction of CSOmap (right) on the HCC scRNA-seq data. (**B**) Spearman correlation between IHC image-based cell connections (X-axis) and SPROUT reconstruction (Y-axis). CD8: CD8+ T cells; Tex: exhausted T cell; Treg: Foxp3+ regulatory T cells; M: macrophages; cDC1: CLEC9A+ dendritic cells; O: other cells. (**C**) Comparison of cell-type proximity (Spearman) between different embedding and sparsification methods. The green dotted line represents the best Spearman correlation of the CSOmap prediction. (**D**) CD8+ T cell and Treg cells in the quasi-structure of SPROUT colored by cell types (left) and standardized expressions (right). (**E**) The pie charts of dominating LR pairs in the interaction of regulatory T cells with CD8+ T cells and exhausted T cells, respectively.

sistency indicates the effectiveness of the quasi-structure rebuild by SPROUT to reveal the local signal transmission process in the tissue.

DISCUSSION

The combination of the spatial context and expression profile of each cell enables our understanding of the intercellular regulation mechanism of tissue homeostasis and pathogenesis. The scRNA-seq discards the spatial context, and ST technologies skimp the cell resolution. Therefore, current technologies are inadequate to produce the spatial structure of tissues with single-cell resolution. In this work, we presented SPROUT to reconstruct the single-cell resolution spatial structure from the spatial and/or single-cell transcriptome. SPROUT rebuilds the quasi-structure of cells by embedding the sparsified affinity graph to a low-dimensional space. The reconstruction accuracy of SPROUT has been demonstrated in the mouse hippocampus, human heart, and tumor microenvironment of different organs in expression similarity, shape similarity, and cell-type proximity.

Although SPROUT relies on a comprehensive and valid LR pair database, extensive tests across different organisms and diseases demonstrate a consistent performance of SPROUT. The recapitulation of literature-supported major LR interactions in TMEs and immune responses also shows the effectiveness of the default LR datasets in providing valid biological observations. However, SPROUT can delineate a broader range of interactions with a higher accuracy if a more extensive LR pair network is expected with future developments. Moreover, our method is extendable to weighted LR pairs. For example, CellPhoneDB weights LR pairs by considering subunits in a ligand–receptor complex that allows a multiple-to-multiple interaction of ligands and receptors (75). In addition, the preprocessing module benefits from a comprehensive single-cell candidate library. It is therefore subjected to the influence of sequencing depth of ST data, the imbalanced sizes, inconsistent cell-type constitution, and batch effects between ST and scRNA-seq data, and the accuracy of the estimated cell numbers per spot. Nevertheless, our evaluations consistently show that SPROUT produces high-correlation quasi-structures across various paired and unpaired datasets with different library sizes. In particular, we recommend using paired datasets for disease studies to ensure an accurate reconstruction against high heterogeneity among samples. In contrast, unpaired datasets have little influence on normal tissues which have smaller divergence in mRNA expression across different samples. SPROUT also supports the ST and single-cell data with their batch effects removed by other computational approaches. Previous deconvolution methods (19–23) failed to achieve a single-cell resolution, integrative methods either fall short in dealing with heterogeneous tissue (11,24,25) or omit single-cell datasets without spatial reference (26), and LR-based reconstruction (14) neglected the pseudo affinities of distant or indirect interacting cells. Unlike previous methods, SPROUT

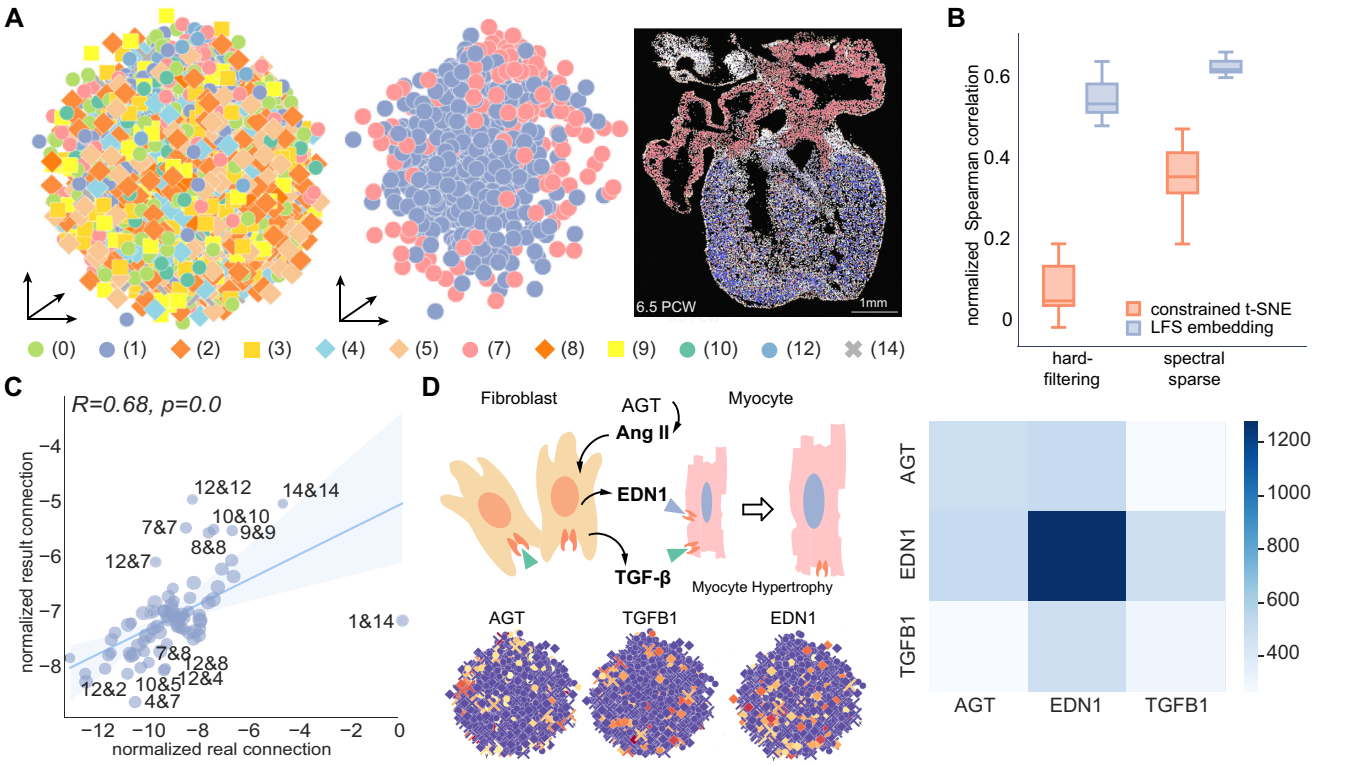

Figure 6. SPROUT recapitulates the quasi-structure of the developing human heart. (**A**) 3D visualization of the reconstructed quasi-structure of the developing human heart (left). Ventricular and atrial cardiomyocytes are separately displayed (middle). The tissue section of 6.5 PCW (scale bar: 1 mm), where the ventricular and atrial cardiomyocytes are manually labeled (right). Cell type label is the same as the original data: (0): Capillary endothelium; (1): Ventricular cardiomyocytes; (2): Fibroblast-like (related to cardiac skeleton connective tissue); (3): Epicardium-derived cells; (4): Fibroblast-like (smaller vascular development); (5): Smooth muscle cells / fibroblast-like; (7): Atrial cardiomyocytes; (8): Fibroblast-like (larger vascular development); (9): Epicardial cells; (10): Endohelium / pericytes / adventia; (12): Myoz2-enriched Cardiomyocytes; (14): Cardiac neural crest cells & Schwann progenitor cells. (**B**) The normalized Spearman correlation of cell-type proximity in the result of hard-filtering and sparsified graphs embedded by constrained t-SNE (orange) and LFS embedding (blue). (**C**) The normalized Spearman correlation between cell-type connections based on spots in the ST section (X-axis) and the quasi-structure reconstructed by SPROUT (Y-axis), with biases introduced by uneven cell counts among different cell types reduced after normalization. (**D**) Mechanism illustration and evaluation of the regulation network between fibroblast and cardiomyocyte. Top-left: schematic diagram of molecular mediation between fibroblast and cardiomyocyte. Bottom-left: standardized expression of above intermediate genes. Right: heatmap of the numbers of neighboring pairs of cells expressing different marker genes.

utilizes the single-cell transcriptome, spatial transcriptome, and LR interactions to reconstruct a quasi-structure of cells in single-cell resolution by a curated affinity graph.

A limitation of the preprocessing module is that the actual number of cells in each spot varies according to spots and tissues. For instance, tissue like the lung, which contains many alveoli, leaves plenty of cavities in the tissue section (76). Therefore hard to estimate the cell number in each spot accurately. For future development, we intend to include an algorithm for accurate quantification of cell numbers per spot by the high-resolution histological image of the tissue section.

SPROUT reconstructs the spatial structure in single-cell resolution, utilizing the spatial context of each cell. The quasi-structure facilitates the acquisition of the dominating LR pairs in each cell pair, leading to the discovery of subpopulations based on dominating LR since cell talk subdivides cell functions. With a precise reconstruction, SPROUT reveals the co-occurrence of different types of cells and divergent colonization of subpopulations, which cannot be detected solely by scRNA-seq or ST technologies. Besides, SPROUT can acquire the dominating LR pairs in each cell pair, leading to the discovery of novel subpopulations based on dominating LR since cell talk subdivides cell functions. These abilities shed light on the studies on tumor heterogeneity and immune therapy. For instance, identifying the disparity of immune microenvironment around different cancer subpopulations could guide medication and metastatic evaluation. Furthermore, the quantification of intercellular interactions between the cancer cell and immune cell can predicate the prognosis of patients with clinical information.

DATA AVAILABILITY

The ST and scRNA-seq data we use have been previously published (14,21,24,25,50,53,71) and are available online at cell2location tutorial, satijalab vignette, mousebrain.org, and https://github.com/almaan/stereoscope. The PDAC, HNC, and SCC datasets are deposited at the Gene Expression Omnibus under GSE111672, GSE103322, and

GSE144240. The count matrix of the developing human heart is available at https://www.spatialresearch.org with the erythrocytes and immune cells removed, and the labels we use in this work remain consistent with the original publication. The HCC dataset CSOmap used is deposited at EGA with accession number EGAS00001003449.

The software implementation of SPROUT is available at https://github.com/deepomicslab/SPROUT.

SUPPLEMENTARY DATA

Supplementary Data are available at NARGAB Online.

ACKNOWLEDGEMENTS

We would like to express sincere gratitude to Dr Wenji Ma for the suggestions on data collection. We appreciate Ms Wenqian Zhang for the manual cell-type annotation of the heart tissue section.

Author contributions: SCL conceived and supervised the project. J.W. developed the software. J.W. and SYL designed and performed the analysis and validation experiment. SCL, J.W., SYL, and L.C performed manuscript writing, review, and editing.

FUNDING

SIRG [CityU SIRG 7020005]. The work was supported by The Science Technology and Innovation Committee of Shenzhen Municipality (Project No. JCYJ20200109143216036).

Conflict of interest statement. None declared.

REFERENCES

1. Wu,A.R., Neff,N.F., Kalisky,T., Dalerba,P., Treutlein,B., Rothenberg,M.E., Mburu,F.M., Mantalas,G.L., Sim,S., Clarke,M.F. *et al.* (2014) Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods*, **11**, 41–46.
2. Kolodziejczyk,A.A., Kim,J.K., Svensson,V., Marioni,J.C. and Teichmann,S.A. (2015) The technology and biology of single-cell RNA sequencing. *Mol. Cell*, **58**, 610–620.
3. Hwang,B., Lee,J.H. and Bang,D. (2018) Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.*, **50**, 1–14.
4. Ke,R., Mignardi,M., Pacureanu,A., Svedlund,J., Botling,J., Wählby,C. and Nilsson,M. (2013) In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods*, **10**, 857–860.
5. Lubeck,E., Coskun,A.F., Zhiyentayev,T., Ahmad,M. and Cai,L. (2014) Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods*, **11**, 360–361.
6. Eng,C.-H.L., Lawson,M., Zhu,Q., Dries,R., Koulena,N., Takei,Y., Yun,J., Cronin,C., Karp,C., Yuan,G.-C. *et al.* (2019) Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature*, **568**, 235–239.
7. Ståhl,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.
8. Rodriques,S.G., Stickels,R.R., Goeva,A., Martin,C.A., Murray,E., Vanderburg,C.R., Welch,J., Chen,L.M., Chen,F. and Macosko,E.Z. (2019) Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science*, **363**, 1463–1467.
9. Stickels,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.
10. Zhao,E., Stone,M.R., Ren,X., Guenthoer,J., Smythe,K.S., Pulliam,T., Williams,S.R., Uytingco,C.R., Taylor,S.E., Nghiem,P. *et al.* (2021) Spatial transcriptomics at subspot resolution with BayesSpace. *Nat. Biotechnol.*, **39**, 1375–1384.
11. Moriel,N., Senel,E., Friedman,N., Rajewsky,N., Karaiskos,N. and Nitzan,M. (2021) NovoSpaRc: flexible spatial reconstruction of single-cell gene expression with optimal transport. *Nat. Protoc.*, **16**, 4177–4200.
12. Thrane,K., Eriksson,H., Maaskola,J., Hansson,J. and Lundeberg,J. (2018) Spatially resolved transcriptomics enables dissection of genetic heterogeneity in stage III cutaneous malignant melanoma. *Cancer Res.*, **78**, 5970–5979.
13. Vestweber,D. (2015) How leukocytes cross the vascular endothelium. *Nat. Rev. Immunol.*, **15**, 692–704.
14. Ren,X., Zhong,G., Zhang,Q., Zhang,L., Sun,Y. and Zhang,Z. (2020) Reconstruction of cell spatial organization from single-cell RNA sequencing data based on ligand–receptor mediated self-assembly. *Cell Res.*, **30**, 763–778.
15. Wells,A. and Wiley,H.S. (2018) A systems perspective of heterocellular signaling. *Essays Biochem.*, **62**, 607–617.
16. Achim,K., Pettit,J.-B., Saraiva,L.R., Gavriouchkina,D., Larsson,T., Arendt,D. and Marioni,J.C. (2015) High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.*, **33**, 503–509.
17. Satija,R., Farrell,J.A., Gennert,D., Schier,A.F.t. and Regev,A. (2015) Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.*, **33**, 495–502.
18. 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.
19. 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.
20. Dong,R. and Yuan,G.-C. (2021) SpatialDWLS: accurate deconvolution of spatial transcriptomic data. *Genome Biol.*, **22**, 145.
21. Andersson,A., Bergenstråhle,J., Asp,M., Bergenstråhle,L., Jurek,A., Navarro,J.F. and Lundeberg,J. (2020) Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Commun. Biol.*, **3**, 565.
22. Kleshchevnikov,V., Shmatko,A., Dann,E., Aivazidis,A., King,H.W., Li,T., Lomakin,A., Kedlian,V., Jain,M.S., Park,J.S. *et al.* (2020) Comprehensive mapping of tissue cell architecture via integrated single cell and spatial transcriptomics. bioRxiv doi: https://doi.org/10.1101/2020.11.15.378125, 17 November 2020, preprint: not peer reviewed.
23. Cable,D.M., Murray,E., Zou,L.S., Goeva,A., Macosko,E.Z., Chen,F. and Irizarry,R.A. (2021) Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.*, **40**, 517–526.
24. 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.
25. Ji,A.L., Rubin,A.J., Thrane,K., Jiang,S., Reynolds,D.L., Meyers,R.M., Guo,M.G., George,B.M., Mollbrink,A., Bergenstråhle,J. *et al.* (2020) Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell*, **182**, 497–514.
26. Cang,Z. and Nie,Q. (2020) Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat. Commun.*, **11**, 2084.
27. de la Fuente,A., Bing,N., Hoeschele,I. and Mendes,P. (2004) Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics*, **20**, 3565–3574.
28. Spielman,D.A. and Teng,S.-H. (2011) Spectral Sparsification of Graphs. *SIAM J. Comput.*, **40**, 981–45
29. 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.

30. Raghavan,P. and Tompson,C.D. (1987) Randomized rounding: a technique for provably good algorithms and algorithmic proofs. *Combinatorica*, **7**, 365–374.
31. Datar,M., Immorlica,N., Indyk,P. and Mirrokni,V.S. (2004) Locality-sensitive hashing scheme based on p-stable distributions. In: *Proceedings of the twentieth annual symposium on Computational geometry*. pp. 253–262.
32. Stuart,T., Butler,A., Hoffman,P., Hafemeister,C., Papalexi,E., Mauck,W.M., Hao,Y., Stoeckius,M., Smibert,P. and Satija,R. (2019) Comprehensive Integration of Single-Cell Data. *Cell*, **177**, 1888–1902.
33. Ramilowski,J.A., Goldberg,T., Harshbarger,J., Kloppmann,E., Lizio,M., Satagopam,V.P., Itoh,M., Kawaji,H., Carninci,P., Rost,B. *et al.* (2015) A draft network of ligand–receptor-mediated multicellular signalling in human. *Nat. Commun.*, **6**, 7866.
34. Shao,X., Liao,J., Li,C., Lu,X., Cheng,J. and Fan,X. (2020) CellTalkDB: a manually curated database of ligand–receptor interactions in humans and mice. *Brief. Bioinfor.*, **22**, bbaa269.
35. Fan,J., Liao,Y. and Liu,H. (2016) An overview of the estimation of large covariance and precision matrices. *Econometrics J.*, **19**, C1–C32.
36. Li,Y., Hu,J., Zhang,C., Yu,D.-J. and Zhang,Y. (2019) ResPRE: high-accuracy protein contact prediction by coupling precision matrix with deep residual neural networks. *Bioinformatics*, **35**, 4647–4655.
37. Das,A., Sampson,A.L., Lainscsek,C., Muller,L., Lin,W., Doyle,J.C., Cash,S.S., Halgren,E. and Sejnowski,T.J. (2017) Interpretation of the precision matrix and its application in estimating sparse brain connectivity during sleep spindles from human electrocorticography recordings. *Neural Comput.*, **29**, 603–642.
38. DeGroot,M.H. (2005) In: *Optimal Statistical Decisions*. Vol. **82**, John Wiley & Sons.
39. Chakeri,A., Farhidzadeh,H. and Hall,L.O. (2016) Spectral sparsification in spectral clustering. In: *2016 23rd international conference on pattern recognition (icpr)*. IEEE, pp. 2301–2306.
40. Spielman,D.A. and Srivastava,N. (2011) Graph sparsification by effective resistances. *SIAM J. Comput.*, **40**, 1913–1926.
41. Frieze,A., Kannan,R. and Vempala,S. (2004) Fast Monte-Carlo algorithms for finding low-rank approximations. *J. ACM*, **51**, 1025–1041.
42. Batson,J.D., Spielman,D.A. and Srivastava,N. (2009) Twice-Ramanujan Sparsifiers. In: *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*. Association for Computing Machinery STOC '09, NY, p. 255–262.
43. Chang,S.S. and Zadeh,L.A. (1996) On fuzzy mapping and control. In: *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A. Zadeh*. World Scientific, pp. 180–184.
44. Gathigi,S.M., Gichuki,M.N., Otieno,P.A. and Were,H.S. (2013) Normality and Its Variants on Fuzzy Isotone Spaces. *Adv. Pure Math.*, **3**, 639–642.
45. McInnes,L., Healy,J. and Melville,J. (2018) Umap: Uniform manifold approximation and projection for dimension reduction. arXiv doi: https://arxiv.org/abs/1802.03426, 18 September 2020, preprint: not peer reviewed.
46. Koren,Y. (2003) On spectral graph drawing. In: *International Computing and Combinatorics Conference*. Springer, Berlin, Heidelberg, pp. 496–508.
47. Kullback,S. and Leibler,R.A. (1951) On information and sufficiency. *Annals Math. Stat.*, **22**, 79–86.
48. Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.: Series B (Methodological)*, **57**, 289–300.
49. 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.
50. Tasic,B., Menon,V., Nguyen,T.N., Kim,T.K., Jarsky,T., Yao,Z., Levi,B., Gray,L.T., Sorensen,S.A., Dolbeare,T. *et al.* (2016) Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.*, **19**, 335–346.
51. Ståhl,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.
52. Hie,B., Bryson,B. and Berger,B. (2019) Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat. Biotechnol.*, **37**, 685–691.
53. Zeisel,A., Hochgerner,H., Lönnerberg,P., Johnsson,A., Memic,F., Van Der Zwan,J., Häring,M., Braun,E., Borm,L.E., La Manno,G. *et al.* (2018) Molecular architecture of the mouse nervous system. *Cell*, **174**, 999–1014.
54. Benarroch,E.E. (2005) Neuron-astrocyte interactions: partnership for normal function and disease in the central nervous system. In: *Mayo Clinic Proceedings*. Vol. **80**, Elsevier, pp. 1326–1338.
55. Benesty,J., Chen,J. and Huang,Y. (2008) On the importance of the Pearson correlation coefficient in noise reduction. *IEEE Trans. Audio, Speech Lang. Process.*, **16**, 757–765.
56. Vickovic,S., Eraslan,G., Salmén,F., Klughammer,J., Stenbeck,L., Schapiro,D., Äijö,T., Bonneau,R., Bergenstråhle,L., Navarro,J.F. *et al.* (2019) High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods*, **16**, 987–990.
57. Liu,C.-C., Hu,J., Zhao,N., Wang,J., Wang,N., Cirrito,J.R., Kanekiyo,T., Holtzman,D.M. and Bu,G. (2017) Astrocytic LRP1 mediates brain Aβ clearance and impacts amyloid deposition. *J. Neurosci.*, **37**, 4023–4031.
58. Kim,J., Eltorai,A.E., Jiang,H., Liao,F., Verghese,P.B., Kim,J., Stewart,F.R., Basak,J.M. and Holtzman,D.M. (2012) Anti-apoE immunotherapy inhibits amyloid accumulation in a transgenic mouse model of Aβ amyloidosis. *J. Exp. Med.*, **209**, 2149–2156.
59. Tuli,A., Sharma,M., Wang,X., Simone,L.C., Capek,H.L., Cate,S., Hildebrand,W.H., Naslavsky,N., Caplan,S. and Solheim,J.C. (2009) Amyloid precursor-like protein 2 association with HLA class I molecules. *Cancer Immun. Immunother.*, **58**, 1419.
60. Zhang,Y., Du,W., Chen,Z. and Xiang,C. (2017) Upregulation of PD-L1 by SPP1 mediates macrophage polarization and facilitates immune escape in lung adenocarcinoma. *Exp. Cell Res.*, **359**, 449–457.
61. Zeng,B., Zhou,M., Wu,H. and Xiong,Z. (2018) SPP1 promotes ovarian cancer progression via Integrin β1/FAK/AKT signaling pathway. *OncoTargets Ther.*, **11**, 1333.
62. Jiao,Y., Li,Y., Liu,S., Chen,Q. and Liu,Y. (2019) ITGA3 serves as a diagnostic and prognostic biomarker for pancreatic cancer. *OncoTargets Ther.*, **12**, 4141.
63. Suzuki,S.R., Kuno,A. and Ozaki,H. (2021) Cell-to-cell interaction analysis of prognostic ligand–receptor pairs in human pancreatic ductal adenocarcinoma. *Biochem. Biophys. Rep.*, **28**, 101126.
64. Zheng,J., Yang,T., Gao,S., Cheng,M., Shao,Y., Xi,Y., Guo,L., Zhang,D., Gao,W., Zhang,G. *et al.* (2021) miR-148a-3p silences the CANX/MHC-I pathway and impairs CD8+ T cell-mediated immune attack in colorectal cancer. *FASEB J.*, **35**, e21776.
65. Yuan,D., Tao,Y., Chen,G. and Shi,T. (2019) Systematic expression analysis of ligand–receptor pairs reveals important cell-to-cell interactions inside glioma. *Cell Commun. Signal.*, **17**, 48.
66. Seidu,R.A., Wu,M., Su,Z. and Xu,H. (2017) Paradoxical role of high mobility group box 1 in glioma: a suppressor or a promoter?*Oncol. Rev.*, **11**, 325.
67. Stepp,M.A., Liu,Y., Pal-Ghosh,S., Jurjus,R.A., Tadvalkar,G., Sekaran,A., LoSicco,K., Jiang,L., Larsen,M., Li,L. *et al.* (2007) Reduced migration, altered matrix and enhanced TGFβ1 signaling are signatures of mouse keratinocytes lacking Sdc1. *J. Cell Sci.*, **120**, 2851–2863.
68. Chen,M.-L., Pittet,M.J., Gorelik,L., Flavell,R.A., Weissleder,R., Von Boehmer,H. and Khazaie,K. (2005) Regulatory T cells suppress tumor-specific CD8 T cell cytotoxicity through TGF-β signals in vivo. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 419–424.
69. Liu,X., Mo,W., Ye,J., Li,L., Zhang,Y., Hsueh,E.C., Hoft,D.F. and Peng,G. (2018) Regulatory T cells trigger effector T cell DNA damage and senescence caused by metabolic competition. *Nat. Commun.*, **9**, 249.
70. Hsu,C.-L., Ou,D.-L., Bai,L.-Y., Chen,C.-W., Lin,L., Huang,S.-F., Cheng,A.-L., Jeng,Y.-M. and Hsu,C. (2021) Exploring markers of exhausted CD8 T cells to predict response to immune checkpoint inhibitor therapy for hepatocellular carcinoma. *Liver Cancer*, **10**, 346–359.
71. Asp,M., Giacomello,S., Larsson,L., Wu,C., Fürth,D., Qian,X., Wärdell,E., Custodio,J., Reimegård,J., Salmén,F. *et al.* (2019) A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart. *Cell*, **179**, 1647–1660.
72. Howard,C.M. and Baudino,T.A. (2014) Dynamic cell–cell and cell–ECM interactions in the heart. *J. Mol. Cell. Cardiol.*, **70**, 19–26.
73. Gray,M.O., Long,C.S., Kalinyak,J.E., Li,H.-T. and Karliner,J.S. (1998) Angiotensin II stimulates cardiac myocyte hypertrophy via

paracrine release of TGF-β1 and endothelin-1 from fibroblasts. *Cardiovasc. Res.*, **40**, 352–363.
74. Jeunemaitre,X., Soubrier,F., Kotelevtsev,Y.V., Lifton,R.P., Williams,C.S., Charru,A., Hunt,S.C., Hopkins,P.N., Williams,R.R., Lalouel,J.-M. *et al.* (1992) Molecular basis of human hypertension: role of angiotensinogen. *Cell*, **71**, 169–180.
75. Efremova,M., Vento-Tormo,M., Teichmann,S.A. and Vento-Tormo,R. (2020) CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat. Protoc.*, **15**, 1484–1506.
76. Brunnström,H., Johansson,L., Jirström,K., Jönsson,M., Jönsson,P. and Planck,M. (2013) Immunohistochemistry in the differential diagnostics of primary lung cancer: an investigation within the Southern Swedish Lung Cancer Study. *Am. J. Clin. Pathol.*, **140**, 37–46.