# **Cell Systems**

# scTenifoldXct: A semi-supervised method for predicting cell-cell interactions and mapping cellular communication graphs

### **Graphical abstract**



### Authors

Yongjian Yang, Guanxun Li, Yan Zhong, ..., Cristhian Roman-Vicharra, Robert S. Chapkin, James J. Cai

### Correspondence

r-chapkin@tamu.edu (R.S.C.), jcai@tamu.edu (J.J.C.)

### In brief

scTenifoldXct is a semi-supervised computational tool for detecting LRmediated cell-cell interactions. It uncovers weak but biologically important interactions often overlooked by other methods. It can also be used to compare different samples to identify differential interactions, revealing functional implications associated with changes in cellular communication status.

### **Highlights**

- scTenifoldXct detects LR-mediated cell-cell interactions
- scTenifoldXct leverages manifold alignment of gene regression networks
- scTenifoldXct detects weak but biologically important interactions
- scTenifoldXct detects differential interactions between two different samples



# **Cell Systems**

### Methods



# scTenifoldXct: A semi-supervised method for predicting cell-cell interactions and mapping cellular communication graphs

Yongjian Yang,<sup>1</sup> Guanxun Li,<sup>2</sup> Yan Zhong,<sup>4</sup> Qian Xu,<sup>3</sup> Yu-Te Lin,<sup>5</sup> Cristhian Roman-Vicharra,<sup>3</sup> Robert S. Chapkin,<sup>6,\*</sup> and James J. Cai<sup>1,3,7,8,\*</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA

<sup>2</sup>Department of Statistics, Texas A&M University, College Station, TX 77843, USA

<sup>3</sup>Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843, USA

<sup>4</sup>Key Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE, School of Statistics, East China Normal University, 3663 North Zhongshan Road, Shanghai 200062, China

<sup>5</sup>Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei, Taiwan

<sup>6</sup>Department of Nutrition and the Program in Integrative Nutrition & Complex Diseases, Texas A&M University, College Station, TX 77843, USA <sup>7</sup>Interdisciplinary Program of Genetics, Texas A&M University, College Station, TX 77843, USA

<sup>8</sup>Lead contact

\*Correspondence: r-chapkin@tamu.edu (R.S.C.), jcai@tamu.edu (J.J.C.) https://doi.org/10.1016/j.cels.2023.01.004

#### SUMMARY

We present scTenifoldXct, a semi-supervised computational tool for detecting ligand-receptor (LR)-mediated cell-cell interactions and mapping cellular communication graphs. Our method is based on manifold alignment, using LR pairs as inter-data correspondences to embed ligand and receptor genes expressed in interacting cells into a unified latent space. Neural networks are employed to minimize the distance between corresponding genes while preserving the structure of gene regression networks. We apply scTenifoldXct to real datasets for testing and demonstrate that our method detects interactions with high consistency compared with other methods. More importantly, scTenifoldXct uncovers weak but biologically relevant interactions overlooked by other methods. We also demonstrate how scTenifoldXct can be used to compare different samples, such as healthy vs. diseased and wild type vs. knockout, to identify differential interactions, thereby revealing functional implications associated with changes in cellular communication status.

### **INTRODUCTION**

Single-cell technology has revolutionized biomedical research. For example, single-cell RNA sequencing (scRNA-seq) allows the transcriptomic information from tens of thousands of cells to be gathered in parallel in a robust and reproducible way.<sup>3–5</sup> The unprecedented resolution in scRNA-seq data has been exploited to reveal cellular diversity, such as various cell types and cellular states in tissue samples.<sup>6</sup> With scRNA-seq data, it is possible to study the cellular communication network through the mapping of cell-specific ligand-receptor (LR) connectivity in complex tissues.<sup>1,7</sup> The evolving scRNA-seq data space has sparked the development of numerous computational tools for mining cell-to-cell communication information.<sup>8–19</sup> Nevertheless, robust statistical confidence in detecting results has proved difficult to achieve.<sup>20,21</sup>

This paper presents scTenifoldXct—a computational tool that incorporates intra- and intercellular gene networks to detect cellcell interaction using scRNA-seq data. scTenifoldXct is semi-supervised and thus can be used with or without reference LR pairs. A redesigned crosstalk scoring metric is also introduced to estimate the interaction strength of each LR pair. The crosstalk scoring metric extends the commonly used metric, i.e., the product of ligand and receptor gene expressions, which is adopted by almost all existing methods. The built-in functionality involving single-cell gene network construction allows scTenifoldXct to couple cell-cell interactions with intracellular activities. scTenifoldXct is also able to perform differential interaction analysis in which cell-cell interaction patterns are compared between tissue samples. When performing comparative analysis, scTenifoldXct combines and analyzes data from two samples in an integrative manner rather than processing the two samples separately, making the analysis more powerful in detecting subtle differential interactions. To demonstrate these features of scTenifoldXct, we applied scTenifoldXct to scRNA-seq datasets under the single-sample and two-sample application settings.

### RESULTS

# The scTenifoldXct framework for single-sample application

The simple application for scTenifoldXct is single-sample based, i.e., an application using scRNA-seq data from one sample that contains different cell types. The goal is to detect LR pairs with

# CellPress

### Cell Systems Methods



#### Figure 1. Overview of the scTenifoldXct workflow

scTenifoldXct is designed to identify LR-mediated interactions using scRNA-seq data. The scTenifoldXct workflow involves three successive steps, namely (1) gene similarity matrix construction, (2) manifold alignment, and (3) significant LR pairs detection. PC regression is used in the construction of single-cell gene regression networks (scGRNs),  $W_A$  and  $W_B$ , to approximate gene regulatory networks for cell types A and B, respectively. The joint similarity matrix is composed of  $W_A$ ,  $W_B$ , and **S**, the crosstalk score matrix. Neural networks are used to learn latent representations of each gene pair. Two genes of an LR pair are more likely to interact when their low-dimensional latent representations are more similar.

gene products that interact between two given cell types. In such a single-sample application setting, scTenifoldXct requires two inputs: a gene-by-cell count matrix,  $X \in \mathbb{R}^{n \times m}$  and a cell type label vector,  $c \in \mathbb{R}^{m \times 1}$ , where *n* is the number of genes and *m* is the number of cells. Given any cell type *A*, let  $X_A$  be the count matrix where cells belong to type *A*. The analytical objectives include the following: (1) to examine LR pairs in cell-cell interaction databases and identify those pairs that show significant activities and contribute to the LR-mediated interactions in the given sample, and (2) to discover LR pairs that are absent in the database. The framework of scTenifoldXct involves three successive steps (Figure 1), as briefly described below. More technical details are given in STAR Methods.

# Step 1: Construction of single-cell gene networks and crosstalk score matrix

We started by creating two gene networks for the two cell types of interest, respectively (substep 1.1). Subsequently, we computed a gene-gene crosstalk score matrix (substep 1.2). Finally, we concatenated the two gene networks and the genegene crosstalk score matrix to create a joint similarity matrix (substep 1.3), subject to manifold alignment in step 2.

Substep 1.1: Constructing gene network for each of the cell types. We used principal component (PC) regression to

construct single-cell gene networks as previously described.<sup>22</sup> Briefly, for each gene, the expression of the gene was used as a response variable, and the expressions of all other genes were used as dependent variables. The constructed gene networks for cell types A and B were then saved as graphs with signed, weighted, and directional edge weights represented as adjacency matrices  $W_A$  and  $W_B$ . Each column of an adjacency matrix stores the PC regression coefficients of a gene, indicating the regulatory relationships between this gene and all other genes.  $W_A$  and  $W_B$  were normalized separately by dividing the maximum absolute value of all entries of the matrix. Note that PC regression infers gene-gene expression relationships to approximate gene regulatory networks without requiring any information on transcription factors (TFs) and their targets or knowledge of regulatory elements such as enhancers and promoters. To avoid confusion, the networks,  $W_A$  and  $W_B$ , are hereafter referred to as gene regression networks. Users can supply their own  $W_A$  and  $W_B$  at this step to replace these two gene regression networks.

Substep 1.2: Calculating crosstalk score between gene pairs from different cell types. We computed the crosstalk score,  $score(i_A, j_B)$ , between gene *i* in cell type *A* and gene *j* in cell type *B*. The score was designed to incorporate information

### Cell Systems Methods



from the mean and variance of expression of the two genes. The score is greater when two genes have a higher and more variable expression (see STAR Methods). The crosstalk scores between pairs of all genes form the matrix **S**.

Substep 1.3: Constructing gene similarity matrix. The joint similarity matrix W was composed of four blocks. The gene regression networks,  $W_A$  and  $W_B$ , were set as the diagonal blocker matrices and the crosstalk score matrix **S** as the off-diagonal blocker matrix to yield the similarity matrix W between genes from two cell types, A and B.

#### Step 2: Manifold alignment

The manifold alignment method was applied to the joint similarity matrix **W** to recover latent representations of gene expression of two cell types. We used neural networks to learn the unified low-dimensional latent representation.<sup>23</sup> The manifold alignment problem solved by neural networks has been shown to be computationally efficient while preserving nonlinearity properties,<sup>24</sup> and numerically more stable than methods based on eigen decomposition.<sup>25,26</sup>

### Step 3: Determination of significant LR pairs

Euclidean distance between projected genes on the aligned manifold subspace was used to determine significant LR pairs. The distances were computed between all gene pairs from cell type A to B. Based on the distance values, gene pairs were ranked. The closer two genes are, the more likely they are to interact. To select significant LR pairs, p values were calculated using a nonparametric test. A list of significant LR pairs was selected using a 0.05 p value cutoff.

# Single-sample scTenifoldXct analysis with inflammatory skin data

We validated scTenifoldXct outcomes using publicly available scRNA-seq datasets. The first dataset was generated with lesional and non-lesional skin samples in a study of atopic dermatitis.<sup>27</sup> This inflammatory skin dataset contains ten different subpopulations of cells, including three subpopulations of fibroblasts and two subpopulations of dendritic cells. Experimental evidence shows that lesional skin has enhanced CCL19-CCR7 interaction between inflammatory fibroblasts and dendritic cells. CCL19 is required for lymphocyte recirculation, homing, and migration to secondary lymphoid organs.<sup>28</sup> Lesional skin in atopic dermatitis shows enhanced chemokine signals with higher expression of CCL19 in inflammatory fibroblasts.<sup>29</sup>

We first focused on interactions between fibroblasts and dendritic cells and identified 30 significant LR pairs (Table S1), including CCL19-CCR7. These significant interactions shed light on the function of inflammatory fibroblasts interacting with lymphoid cells and regulating type 2 inflammation, an inflammatory pathway involving a subpopulation of CD4<sup>+</sup> T cells known as Th2 cells, which is consistent with the findings of the original study.<sup>27</sup> Gene ontology (GO) enrichment analysis further suggested that significant LR genes were enriched in dendritic cell chemotaxis and dendritic cell migration pathways. This result is in line with another previous study,<sup>30</sup> showing that the cutaneous immune response depends on dendritic cell migration from the skin to draining lymph nodes. VCAM1-ITGB2 and CCL2-CXCR4 were identified by scTenifoldXct as the top-ranked pairs. These genes do not have a higher expression (measured by either the average expression of a single gene among cells or the product of the average expressions of two genes) than many other LR pairs. But, both VCAM1-ITGB2 and CCL2-CXCR4 are biologically relevant, i.e., VCAM1-associated type 2 inflammation causes the over-expression of cytokines (CCL2 and CCL19) in fibroblasts,<sup>27</sup> and CCL2 is known to be induced by inflammatory stimuli such as tumor necrosis factor  $\alpha$ (TNF- $\alpha$ ).<sup>31</sup> scTenifoldXct also identified CCL2-TNF. Again, neither of the genes was highly expressed. These examples demonstrate the capability of scTenifoldXct in detecting lowly expressed LR pairs by considering intracellular regulatory activities.

We then focused on interactions between dendritic cells and T cells and identified 33 significant LR pairs (Table S2). These interactions centered around two ligand genes: CCL17 and CCL22, both closely associated with the pathophysiology of atopic dermatitis.<sup>32</sup> Serum levels of CCL17 and CCL22 are known to be correlated with the disease severity.<sup>33</sup> In the context of inflamed skin, dendritic cell-derived CCL17 and CCL22 primarily attract T cells that express the cutaneous homing receptor-they bind to C-C motif chemokine receptors (CCRs), which are preferentially expressed in T cells, resulting in inflammation.<sup>34</sup> Genes of significant LR pairs also included CCR6, CXCR4, CCL17, S100A8, and S100B. The abundant expression of CCR6 in T cells in skin lesions suggests its important role in early inflammatory T cell recruitment. CXCR4 is a critical receptor involved in both homeostatic and pathological leukocyte trafficking, attracting cells to inflammatory sites and contributing to the activation of integrins required for T cell activation.<sup>35</sup> CCL17 binds to the receptor CCR4, which is known to be expressed on activated/memory T cells. Recent data show the enhanced expression of CCL17 in the skin lesions and serum of AD patients, leading to dendritic cell migration from the skin to the skin-draining lymph nodes.<sup>36</sup> S100 proteins such as S100A8 and S100B are required for immunological homeostasis and inflammation and have been linked to various inflammatory skin diseases, including psoriasis and atopic dermatitis.37,38 S100A8 and S100B were other examples lowly expressed but were detected and ranked as the top LR pairs by scTenifoldXct.

Finally, we conflated the interactions across three cell types, i.e., fibroblast-dendritic cell-T cell. The combined results implied a cascade of intercellular signaling pathways where fibroblasts activate dendritic cell inflammatory responsiveness. The latter interacts with T cells to facilitate T cell trafficking, lymphoid tissue organization, and type 2 cell recruitment (Figure 2A). Gene regression networks allowed us to approximate and examine intracellular systems, locate LR genes in the intracellular networks, and trace the upstream TFs that regulate the expression of the LR genes. Figure 2B illustrates such an integrated intraand intercellular network centered around two significant interactions: CCL19-CCR7 and CXCL12-CXCR4, as predicted by scTenifoldXct. In dendritic cells, REL (proto-oncogene c-Rel, an NF-kB subunit) was strongly regulated by CCR7, which is consistent with the experimental results showing that CCR7 activates NF-κB.<sup>39</sup> NF-κB is also known to be co-activated with AP1 and regulates CCR7 expression.<sup>40</sup> JUN (the subunit of AP1) is a TF positively regulates the expression of CCR7. Figure 2C depicts another integrated network centered around ligands CCL17 and CCL22 and their receptors CCR6 and CXCR4 in dendritic and T cells. We found that AP1 subunits,

## CellPress

### Cell Systems Methods



**Figure 2. Cell-cell interactions between fibroblasts, dendritic cells, and T cells with their intracellular networks in inflamed skin** (A) An illustration of representative interactions between cell types in skin lesions.<sup>27</sup> (B) An integrated network across fibroblasts and dendritic cells with interactions: CCL19-CCR7, CXCL12-CXCR4, and CCL2-TNF (boldfaced in Table S1). Blue

(b) An integrated network across floroblasts and denoritic cells with interactions: CCL19-CCH7, CXCL12-CXCH4, and CCL2-TNF (boldfaced in Table S1). Blue and red edges indicate negative and positive regulatory relationships between genes, respectively. Genes like HES1 and JUNB are present in both cell types and thus appear twice in the figure.

(C) An integrated network across dendritic cells and T cells with interactions: CCL17-CCR6, CCL22-CCR6, and CCL17-CXCR4 (boldfaced in Table S2).

JUNB and JUN, are linked with these receptors, suggesting a role of JUN/AP-1 proteins in skin inflammation.<sup>41</sup> In conclusion, these constructed networks were consistent with our prior knowledge about gene regulatory relationships in the studied inflammatory skin system.

# Comparison between scTenifoldXct and other existing methods

We compared the prediction results of scTenifoldXct with those produced by five existing methods, namely CellChat,<sup>11</sup> Connectome,<sup>12</sup> iTALK,<sup>10</sup> NATMI,<sup>8</sup> and SingleCellSignalR.<sup>9</sup> The prediction of each method was generated from the same input data, i.e., fibroblasts and dendritic cells in the inflammatory skin dataset, and the same reference cell-cell interaction database, Omnipath,<sup>1</sup> was used. For each method, we implemented them at their own optimal threshold and retained the equal number of most significant LR pairs as reported (except that CellChat only produced nine significant pairs) to perform an overlap analysis using an upset plot (Figure S1). Five interactions (CCL2-CCR7, CCL19-CCR7, CCL19-CXCR4, CXCL12-CCR7, and CXCL14-CCR7) were detected by all six methods. Predictions made by scTenifoldXct tended to overlap more with consensus inferred by other methods. For example, scTenifoldXct detected all interactions that CellChat detected. scTenifoldXct detected six interactions missed by CellChat and Connectome, but the other three methods detected them. scTenifoldXct detected three additional interactions missed by CellChat but detected by all other methods. CytoTalk,<sup>19</sup> another method considering intracellular networks, detected four significant LR pairs (CCL19-CCR7, CCL26-CCR7, CCL2-VEGFA, and CCL26-CXCR4), and all were detected by scTenifoldXct (Table S1).

Additionally, scTenifoldXct detected S100A8-ITGB2 and CCL26-CCR6. All other methods failed to detect these two LR pairs because this observation involved genes that were lowly expressed (Figure S2), while the sensitivity of scTeni-

foldXct was augmented with the manifold alignment of intracellular networks. There is mounting evidence that shows that the two LR pairs are key factors involved in the pathogenesis of atopic dermatitis. S100A8 is an important molecule in the pathogenesis and progression of atopic dermatitis via altering cytokine and skin barrier protein expression levels.<sup>42</sup> S100A8 is the first S100 family member that has a potent chemokine-like activity to murine phagocytes in vitro and in vivo.43 S100A8 and S100A9 induce neutrophil adhesion to fibrinogen in vitro via upregulating Mac-1 (a heterodimer of CD11b and ITGB2), indicating S100A8-ITGB2 is an important pathogenic mechanism associated with the pathogenesis and the progression of atopic dermatitis. CCL26-CCR6 is another pair of molecules uniquely detected by scTenifoldXct. The lesional atopic dermatitis samples were characterized by the expansion of inflammatory dendritic cells and tissue-resident memory T cells.<sup>27</sup> CCR6, a  $\beta$ -chemokine receptor, mediates the migration of dendritic cells and several lymphocyte subsets to sites of epithelial inflammation.44,45 It has also been reported that CCR6 is required for IL-23-induced psoriasis-like inflammation in mice.<sup>46</sup> CCL26 is another molecule that may serve an important role in the pathogenesis of atopic dermatitis.47 The expression of CCL26 is known to be increased in lesional atopic dermatitis fibroblasts,<sup>27</sup> and the serum CCL26 levels in patients with atopic dermatitis tend to decrease after the treatment.<sup>47</sup> These analyses demonstrate that scTenifoldXct is capable of predicting both strong and relatively weak interactions.

To further assess the performance of scTenifoldXct quantitatively, we used 40 LR pairs that were identified by all the five existing methods as ground truth and plotted the receiver operating characteristic (ROC) curve and precision-recall curve. We found that scTenifoldXct achieved the greatest area under the ROC curve (AUROC, 0.89) and the highest average precision (AP, 0.87), respectively (Figure S3).

### **Cell Systems Methods**





**Two-sample** 

### The scTenifoldXct framework for two-sample application

The basic scTenifoldXct framework for single-sample applications can be extended to the comparison between two samples, detecting LR pairs that show a significant difference in interaction strength between samples such as healthy and diseased. Similar to the basic framework in which manifold alignment is achieved by solving a generalized eigenvalue problem with a joint matrix (Figure 3A), the extended framework applies the neural networks as in the single sample to solve the manifold alignment problem in the two-sample. The difference is that the coupled joint matrix is formed by including four gene regression networks (from two cell types of two samples) and two crosstalk score matrices (between two cell types from each sample) (Figure 3B).

### Step 1: Constructing joint similarity matrices for two samples separately

We computed a joint matrix for each sample separately using the procedure described in the single-sample application above.

#### Step 2: Constructing a coupled joint similarity matrix

We then placed the two joint matrices in the diagonal block of the coupled joint matrix V (Figure 3B). To make the low-dimensional representations of two samples numerically comparable, a constraint factor was included by setting  $\beta I$  to the off-diagonal block of the joint similarity matrix, where I is an identity matrix that reflects the binary correspondence between genes in different samples and  $\beta$  is a tuning hyperparameter.

### Step 3: Calculating the distance differences between two samples

To determine significant differentially interacted LR pairs, we calculated the Euclidean distance between every pairwise combination of gene *i* in cell type A and gene *j* in cell type B for two samples, respectively. The distance difference between each gene pair was then computed across samples.

### Step 4: Determining significantly differential LRmediated interactions

We considered gene pairs with a greater distance difference between two samples to be more significantly differentiated. With

#### Figure 3. Joint similarity matrices used in scTenifoldXct single- and two-sample analyses

CellPress

(A) Joint similarity matrix for manifold alignment in the single-sample analysis.  $W_A$  and  $W_B$  are the gene regression networks of cell types A and B, respectively. S is the crosstalk score matrix, and n is the number of genes.

(B) Coupled joint similarity matrix for manifold alignment in two-sample analysis.  $W_{1,A}$  and  $W_{1,B}$ are the gene regression networks of cell types A and B from sample 1,  $W_{2,A}$  and  $W_{2,B}$  from sample 2. S<sub>1</sub> and S<sub>2</sub> are the crosstalk matrices for samples 1 and 2, respectively.  $\beta I$  is an identity matrix with a tuning hyperparameter  $\beta$ .

this, we obtained a list of ranked gene pairs. We computed p values for the gene pairs using the chi-square test, adjusting the p values with a multiple testing

correction, and selected significant LR pairs using a 0.05 false discovery rate (FDR) cutoff against the OmniPath database.<sup>1</sup>

### Differential interaction analysis between normal and tumor samples

The additional applications of scTenifoldXct include multi-sample comparison. To demonstrate the performance of scTenifoldXct comparative analysis, we obtained scRNA-seq dataset from a hepatocellular carcinoma (HCC) study.<sup>48</sup> The dataset contains cells from the tumor and adjacent normal tissues, allowing for the comparison (Figure 4A). Using scTenifoldXct to compare tumor and normal samples, we identified 11 interactions that showed significantly different strengths of hepatocyte-endothelial cell interactions (Table 1). Among them, eight LR pairs interact more strongly in tumor samples. For example, PLA2G2A (secretory calciumdependent phospholipase A2), previously implicated in host antimicrobial defense, inflammatory response, and tissue regeneration,49-51 was found to interact more strongly with ITGA5 in the tumor. GO enrichment analysis with genes in these eight LR pairs enriched beta-1 integrin cell surface interactions (Table S3), highlighting the role of beta-1 integrin in the progression of HCC.<sup>52</sup> Figure 4B depicts an integrated network showing four out of the eight tumor-enhanced pairs of LR-mediated interactions: PLA2G2A-ITGA5, MDK-ITGA6, SPP1-ITGA5, and SPP1-ITGB5. Corresponding intracellular networks of different cell types were included. In this integrated network, TCF4 is linked with ITGA5 and ITGA6 in the endothelial cells, indicating a functional relationship between TCF4 and the two receptor genes. TCF4 is a key component of the Wnt signaling pathway, which has been linked to the proliferation of HCC cells. It could be an effective therapeutic target for blocking the growth of hepatocellular tumors.<sup>53</sup> Specifically, the mutation of β-catenin (leading to its nuclear and cytoplasmic accumulation) renders it capable of bypassing APC-targeted degradation and accumulating in the nucleus to form a complex with TCF4 aberrantly activating downstream transcriptional events.<sup>54</sup> Similarly, ERG is also linked with ITGA5 and ITGA6 in endothelial cells. A previous study has found more ERG-positive endothelial cells in HCC tissue than in adjacent normal tissue.<sup>55</sup> Fos proto-oncogene, AP-1 transcription

### CellPress



factor subunit (FOS) is linked with ITGA5 and ITGB5. FOS is an oncogene upregulated in HCC.<sup>56</sup> Figure 4C shows two identical cellular networks of endothelial cells connected with hepatocytes via LR-mediated interactions IGF2-FGFR3 and IGF2-PDGFRB. In the tumor sample, the reduced intracellular regulation between PDGFRB and ID1 and ID3 was evident. PDGFRB has been implicated in the development and metastasis of HCC, and ID1 and ID3 are associated with HCC dedifferentiation.<sup>57</sup>

# Differential interaction analysis between WT and gene KO samples

To test the generalizability of scTenifoldXct, especially when applying it to the comparison of two samples, we performed an additional two-sample comparison analysis. We obtained scRNA-seq data derived from lung alveolar type II (AT2) cells and basophils in a study utilizing wild-type (WT) and IL1RL1 (IL33R) knockout (KO) mice.<sup>58</sup> It is known that AT2 cells produce IL33 (ligand) that binds to IL1RL1 (receptor). In IL1RL1-deficient lungs, basophils lack the expression of a large number of lung basophil-specific genes. We predicted that scTenifoldXct-based comparison analysis between the WT and KO samples should be able to identify significant differential LR pairs. Indeed, IL33-IL1RL1 ranked at the top of all LR pairs that were predicted by scTenifoldXct to be significantly differentiated (Table S4). In addition to IL33-IL1RL1, scTenifoldXct also identified six other significant differentially interactive LR pairs, including CSF1-TNF, CSF2-TNF, EDN1-EDNRB, and EDN3-EDNRB. CSF1 and CSF2 are CSF family cytokines that have been shown to play crucial roles in shaping the lung microenvironment.59-61 The interaction strengths of CSF1-TNF and CSF2-TNF were all significantly diminished in the KO sample. This is consistent with the known functions of these genes and their interplay in lung development.58,62 Similarly, the interaction strengths of EDN1-EDNRB and EDN3-EDNRB were also diminished in the

#### Figure 4. Cell-cell interactions between hepatocytes (senders) and endothelial cells (receivers)

**Cell Systems** 

(A) An illustration of representative interactions in hepatocellular carcinoma (HCC).<sup>48</sup>

(B) An integrated network across hepatocytes and endothelial cells, connected by interactions PLA2G2A-ITGA5, MDK-ITGA6, SPP1-ITGA5, and SPP1-ITGB5 (boldfaced in Table 1). The top 30 weighted edges in each network of the cell type are shown.

(C) Integrated networks connected through IGF2-FGFR3 and IGF2-PDGFRB across hepatocytes and endothelial cells in the adjacent normal (left) and tumor (right) tissues.

KO sample, which is consistent with the observed downregulation of EDNRB in lung disease, where cells lose the ability to respond to endothelin stimulation.<sup>63</sup>

#### DISCUSSION

We have showcased the functionality and performance of scTenifoldXct. Using

real-data examples, we demonstrated single-sample and twosample applications of scTenifoldXct. A two-sample application involves comparative analysis between two samples for which none of the existing methods adopts the integrative strategy as does scTenifoldXct. Methodologically, scTenifoldXct leverages information on the topological placements of genes in gene regression networks to explore intracellular connections of LR pairs with other genes. Since the scTenifoldXct method takes into account the networks underlying genes of LR pairs, it allows for the reconstruction of a more detailed map of biological pathways linking the two cell types of interest. Incorporating intracellular networks enables a comprehensive inference of signaling pathways across cells. To the best of our knowledge, only a few existing tools, such as NATMI and scMLnet,<sup>8,18</sup> incorporate intracellular network or signaling pathway information in their analyses. The difference is that scTenifoldXct requires no prior knowledge of intracellular networks, whereas other methods depend on signaling pathways and networks in databases such as Kyoto Encyclopedia of Genes and Genomes (KEGG). scTenifoldXct computes gene regression networks from the input data and operates entirely in a data-driven manner. Furthermore, scTenifoldXct assigns scores to all combinations of gene pairs without requiring prior knowledge of known LR pairs. Thus, significant LR pairs detected by scTenifoldXct can be those undocumented in the databases.

scTenifoldXct adopts the redesigned crosstalk score to quantify the interaction strength between an LR pair. The product of ligand and receptor expression has been widely adopted as an indicator for this purpose, which is intuitively sound. When a ligand and its receptor interact, they are expected to be highly expressed, and the value of the product of their gene expression should be greater. The redesigned interaction score we designed for scTenifoldXct used the same principle, but we expanded its formula by incorporating gene expression

### Cell Systems Methods

CellPress

Table 1. Differential interactions between hepatocytes and endothelial cells						
LR pairs	LR distance on manifold (tumor)	LR distance on manifold (normal)	Difference in LR distances between tumor and normal	p value	Adj. p value	
PLA2G2A-ITGA5ª	0.0010	0.0203	-0.0193	3.6E-7	8.0E-4	
TFF3-ACKR3	0.0147	0.0330	-0.0183	1.5E-6	1.6E-3	
IGF2-PDGFRB	0.0290	0.0108	0.0182	1.8E-6	1.7E-3	
IGF2-FGFR3	0.0269	0.0105	0.0164	1.6E-5	6.1E-3	
MDK-PTPRB	0.0041	0.0193	-0.0152	6.6E-5	1.5E-2	
CXCL3-ACKR1	0.0078	0.0228	-0.0150	8.1E-5	1.8E-2	
MDK-ITGA6 <sup>a</sup>	0.0059	0.0206	-0.0147	1.1E-4	2.1E-2	
SPP1-ITGA5 <sup>a</sup>	0.0058	0.0203	-0.0145	1.3E-4	2.4E-2	
IGF2-ERBB2	0.0275	0.0135	0.0140	2.1E-4	3.3E-2	
SPP1-ITGB5 <sup>a</sup>	0.0173	0.0312	-0.0139	2.4E-4	3.6E-2	
SPP1-ITGA6	0.0068	0.0206	-0.0138	2.8E-4	3.9E-2	

LR pairs were ranked according to the difference in distances from the ligand gene to the receptor gene on their aligned manifold in the latent space between two cell types. The sign of difference indicates the direction of change in the interaction strength between the tumor and normal tissues. A positive difference indicates a greater LR distance or decreased interaction strength in the tumor. By contrast, a negative difference indicates a smaller LR distance or increased interaction strength in the tumor.

<sup>a</sup>LR pairs that appear in Figure 4B.

variability. The latter has been shown to be equally weighted as the average gene expression in reflecting cellular functions.<sup>64,65</sup> Rather than focusing just on the expression mean, scTenifoldXct considers gene expression variability across cells and calculates the interaction score, balancing the contributions of the mean and the variability. The effectiveness of this redesigned metric has been demonstrated in our results, showing that biologically significant LR pairs with low expression can be detected.

In terms of differential interaction analysis, although many tools allow users to compare two samples,<sup>8-12</sup> they initially infer interactions for each sample independently before comparison. The differences in obtained interaction scores are used to assess the differences between the two samples. By contrast, the scTenifoldXct comparative algorithm uses the integrated method to learn the correspondences rather than processing two samples independently. The level of differential interactions is measured by the distance differences in the joint latent space of aligned manifolds. Our manifold alignment algorithm automatically assumes that datasets share the same underlying structure across cell types. Such an assumption can be easily nullified by presenting cell type-specific cell states across heterogeneous singlecell datasets. Thus, it remains computationally challenging for manifold alignment algorithms to preserve both shared and dataset-specific cellular structures across samples.<sup>66</sup> Despite this concern, our common manifold-based integrative strategy has been shown to be highly effective. For example, retrospectively, we tried to apply the independent processing strategy to repeat the comparative analysis, but this strategy did not work with scTenifoldXct-i.e., no LR pairs were detected under the independent processing.

For further development, because temporally and spatially precise cell communication is the key to cellular differentiation, we consider the next version of scTenifoldXct could be directed toward incorporating time series and spatial transcriptomic information. For time series scRNA-seq data, existing analytical frameworks such as GraphFP and SoptSC<sup>67,68</sup> have demonstrated the feasibility of simultaneous inference of cell lineages

and cell-cell communications. Recent improvements in pseudotemporal ordering enable us to map the underlying regulatory networks over time.<sup>69,70</sup> Thus, the scTenifoldXct framework can be further updated by taking time series or pseudotime information and integrating dynamic inference modules, enabling the assessment of changes in interaction strength throughout the processes of cellular differentiation or organ development.

#### **STAR**\***METHODS**

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
- Data preprocessing
  - Network construction
  - Crosstalk score
  - Manifold alignment
  - Determining the statistical significance of interactions between cell types in a single sample
  - Comparative manifold alignment
  - Determining LR pair statistical significance of differential interactions between two samples
  - Validate the predicted interactions between LR pairs
  - Visualization of integrated networks
  - Systematic comparisons between scTenifoldXct and existing tools

#### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j. cels.2023.01.004.

### CellPress

#### ACKNOWLEDGMENTS

This research was funded by Texas A&M University 2019 X-Grants and DOD grant GW200026 for J.J.C.; NIH R35 CA197707, the Allen Endowed Chair in Nutrition & Chronic Disease Prevention, and with support from the Institute for Advancing Health through Agriculture for R.S.C.

#### **AUTHOR CONTRIBUTIONS**

J.J.C. conceived the study and designed the workflow. Y.Y. designed the workflow, performed data analysis, and implemented the software tool. Y.Z. and G.L. contributed to the workflow design. Q.X. contributed to the preparation of testing datasets. Y.-T.L. contributed to the software tool implementation. Y.Y., G.L., Q.X., and R.S.C. contributed to the writing of the manuscript. R.S.C. and J.J.C. supervised the study.

#### **DECLARATION OF INTERESTS**

Y.Y., G.L., Y.Z., and J.J.C. are listed as inventors on a patent application related to this work.

#### REFERENCES

- Türei, D., Valdeolivas, A., Gul, L., Palacio-Escat, N., Klein, M., Ivanova, O., Ölbei, M., Gábor, A., Theis, F., Módos, D., et al. (2021). Integrated intraand intercellular signaling knowledge for multicellular omics analysis. Mol. Syst. Biol. 17, e9923. https://doi.org/10.15252/msb.20209923.
- Ng, A.H.M., Khoshakhlagh, P., Rojo Arias, J.E., Pasquini, G., Wang, K., Swiersy, A., Shipman, S.L., Appleton, E., Kiaee, K., Kohman, R.E., et al. (2021). A comprehensive library of human transcription factors for cell fate engineering. Nat. Biotechnol. 39, 510–519. https://doi.org/10.1038/ s41587-020-0742-6.
- Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell *161*, 1202–1214. https://doi.org/10.1016/j.cell.2015. 05.002.
- Zilionis, R., Nainys, J., Veres, A., Savova, V., Zemmour, D., Klein, A.M., and Mazutis, L. (2017). Single-cell barcoding and sequencing using droplet microfluidics. Nat. Protoc. 12, 44–73. https://doi.org/10.1038/nprot. 2016.154.
- Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. 8, 14049. https://doi.org/10.1038/ncomms14049.
- Aldridge, S., and Teichmann, S.A. (2020). Single cell transcriptomics comes of age. Nat. Commun. *11*, 4307. https://doi.org/10.1038/s41467-020-18158-5.
- Armingol, E., Officer, A., Harismendy, O., and Lewis, N.E. (2021). Deciphering cell-cell interactions and communication from gene expression. Nat. Rev. Genet. 22, 71–88. https://doi.org/10.1038/s41576-020-00292-x.
- Hou, R., Denisenko, E., Ong, H.T., Ramilowski, J.A., and Forrest, A.R.R. (2020). Predicting cell-to-cell communication networks using NATMI. Nat. Commun. *11*, 5011. https://doi.org/10.1038/s41467-020-18873-z.
- Cabello-Aguilar, S., Alame, M., Kon-Sun-Tack, F., Fau, C., Lacroix, M., and Colinge, J. (2020). SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. Nucleic Acids Res. 48, e55. https://doi.org/10.1093/nar/gkaa183.
- Wang, Y., Wang, R., Zhang, S., Song, S., Jiang, C., Han, G., Wang, M., Ajani, J., Futreal, A., and Wang, L. (2019). iTALK: an R package to characterize and illustrate intercellular communication. Preprint at bioRxiv. https://doi.org/10.1101/507871.
- 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. https://doi. org/10.1038/s41467-021-21246-9.

- Raredon, M.S.B., Yang, J., Garritano, J., Wang, M., Kushnir, D., Schupp, J.C., Adams, T.S., Greaney, A.M., Leiby, K.L., Kaminski, N., et al. (2022). Computation and visualization of cell-cell signaling topologies in singlecell systems data using connectome. Sci. Rep. 12, 4187. https://doi.org/ 10.1038/s41598-022-07959-x.
- 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. https://doi.org/10.1038/s41596-020-0292-x.
- Ramilowski, J.A., Goldberg, T., Harshbarger, J., Kloppmann, E., Lizio, M., Satagopam, V.P., Itoh, M., Kawaji, H., Carninci, P., Rost, B., and Forrest, A.R. (2015). A draft network of ligand-receptor-mediated multicellular signalling in human. Nat. Commun. 6, 7866. https://doi.org/10.1038/ ncomms8866.
- Noël, F., Massenet-Regad, L., Carmi-Levy, I., Cappuccio, A., Grandclaudon, M., Trichot, C., Kieffer, Y., Mechta-Grigoriou, F., and Soumelis, V. (2021). Dissection of intercellular communication using the transcriptome-based framework ICELLNET. Nat. Commun. *12*, 1089. https://doi.org/10.1038/s41467-021-21244-x.
- Wang, Y. (2020). talklr uncovers ligand-receptor mediated intercellular crosstalk. Preprint at bioRxiv. https://doi.org/10.1101/2020.02.01.930602.
- Browaeys, R., Saelens, W., and Saeys, Y. (2020). NicheNet: modeling intercellular communication by linking ligands to target genes. Nat. Methods 17, 159–162. https://doi.org/10.1038/s41592-019-0667-5.
- Cheng, J., Zhang, J., Wu, Z., and Sun, X. (2021). Inferring microenvironmental regulation of gene expression from single-cell RNA sequencing data using scMLnet with an application to COVID-19. Brief. Bioinform. 22, 988–1005. https://doi.org/10.1093/bib/bbaa327.
- Hu, Y., Peng, T., Gao, L., and Tan, K. (2021). CytoTalk: de novo construction of signal transduction networks using single-cell transcriptomic data. Sci. Adv. 7, eabf1356. https://doi.org/10.1126/sciadv.abf1356.
- Almet, A.A., Cang, Z., Jin, S., and Nie, Q. (2021). The landscape of cell-cell communication through single-cell transcriptomics. Curr. Opin. Syst. Biol. 26, 12–23. https://doi.org/10.1016/j.coisb.2021.03.007.
- Dimitrov, D., Türei, D., Garrido-Rodriguez, M., Burmedi, P.L., Nagai, J.S., Boys, C., Ramirez Flores, R.O., Kim, H., Szalai, B., Costa, I.G., et al. (2022). Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. Nat. Commun. *13*, 3224. https://doi. org/10.1038/s41467-022-30755-0.
- Osorio, D., Zhong, Y., Li, G., Huang, J.Z., and Cai, J.J. (2020). scTenifoldNet: A machine learning workflow for constructing and comparing transcriptomewide gene regulatory networks from single-cell data. Patterns (N Y) 1, 100139. https://doi.org/10.1016/j.patter.2020.100139.
- Golovko, V., Kroshchanka, A., and Treadwell, D. (2016). The nature of unsupervised learning in deep neural networks: A new understanding and novel approach. Opt. Mem. Neural Netw. 25, 127–141. https://doi.org/ 10.3103/S1060992X16030073.
- Nguyen, N.D., Huang, J., and Wang, D. (2022). A deep manifold-regularized learning model for improving phenotype prediction from multi-modal data. Nat Comput. Sci. 2, 38–46. https://doi.org/10.1038/s43588-021-00185-x.
- Vu, H.T., Carey, C.J., and Mahadevan, S. (2012). Manifold warping: manifold alignment over time. AAAI Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI Press) vol 26, pp. 1155–1161.
- Wang, C., and Mahadevan, S. (2009). A general framework for manifold alignment. AAAI Fall Symposium: Manifold Learning and Its Applications (Association for the Advancement of Artificial Intelligence).
- 27. He, H., Suryawanshi, H., Morozov, P., Gay-Mimbrera, J., Del Duca, E., Kim, H.J., Kameyama, N., Estrada, Y., Der, E., Krueger, J.G., et al. (2020). Single-cell transcriptome analysis of human skin identifies novel fibroblast subpopulation and enrichment of immune subsets in atopic



dermatitis. J. Allergy Clin. Immunol. *145*, 1615–1628. https://doi.org/10. 1016/j.jaci.2020.01.042.

- Takamura, K., Fukuyama, S., Nagatake, T., Kim, D.Y., Kawamura, A., Kawauchi, H., and Kiyono, H. (2007). Regulatory role of lymphoid chemokine CCL19 and CCL21 in the control of allergic rhinitis. J. Immunol. *179*, 5897–5906. https://doi.org/10.4049/jimmunol.179.9.5897.
- Nedoszytko, B., Sokołowska-Wojdyło, M., Ruckemann-Dziurdzińska, K., Roszkiewicz, J., and Nowicki, R.J. (2014). Chemokines and cytokines network in the pathogenesis of the inflammatory skin diseases: atopic dermatitis, psoriasis and skin mastocytosis. Postepy Dermatol. Alergol. *31*, 84–91. https://doi.org/10.5114/pdia.2014.40920.
- Saalbach, A., Janik, T., Busch, M., Herbert, D., Anderegg, U., and Simon, J.C. (2015). Fibroblasts support migration of monocyte-derived dendritic cells by secretion of PGE2 and MMP-1. Exp. Dermatol. 24, 598–604. https://doi.org/10.1111/exd.12722.
- Gschwandtner, M., Derler, R., and Midwood, K.S. (2019). More than just attractive: how CCL2 influences myeloid cell behavior beyond chemotaxis. Front. Immunol. 10, 2759. https://doi.org/10.3389/fimmu.2019.02759.
- Mantovani, A., Gray, P.A., Van Damme, J., and Sozzani, S. (2000). Macrophage-derived chemokine (MDC). J. Leukoc. Biol. 68, 400–404.
- Saeki, H., and Tamaki, K. (2006). Thymus and activation regulated chemokine (TARC)/CCL17 and skin diseases. J. Dermatol. Sci. 43, 75–84. https:// doi.org/10.1016/j.jdermsci.2006.06.002.
- 34. Hirota, T., Saeki, H., Tomita, K., Tanaka, S., Ebe, K., Sakashita, M., Yamada, T., Fujieda, S., Miyatake, A., Doi, S., et al. (2011). Variants of C-C motif chemokine 22 (CCL22) are associated with susceptibility to atopic dermatitis: case-control studies. PLoS One 6, e26987. https://doi. org/10.1371/journal.pone.0026987.
- García-Cuesta, E.M., Santiago, C.A., Vallejo-Díaz, J., Juarranz, Y., Rodríguez-Frade, J.M., and Mellado, M. (2019). The role of the CXCL12/ CXCR4/ACKR3 axis in autoimmune diseases. Front. Endocrinol. (Lausanne) 10, 585. https://doi.org/10.3389/fendo.2019.00585.
- Stutte, S., Quast, T., Gerbitzki, N., Savinko, T., Novak, N., Reifenberger, J., Homey, B., Kolanus, W., Alenius, H., and Förster, I. (2010). Requirement of CCL17 for CCR7- and CXCR4-dependent migration of cutaneous dendritic cells. Proc. Natl. Acad. Sci. USA *107*, 8736–8741. https://doi.org/ 10.1073/pnas.0906126107.
- Xia, C., Braunstein, Z., Toomey, A.C., Zhong, J., and Rao, X. (2017). S100 proteins as an important regulator of macrophage inflammation. Front. Immunol. 8, 1908. https://doi.org/10.3389/fimmu.2017.01908.
- Broome, A.M., Ryan, D., and Eckert, R.L. (2003). S100 protein subcellular localization during epidermal differentiation and psoriasis. J. Histochem. Cytochem. 51, 675–685. https://doi.org/10.1177/002215540305100513.
- Liu, F.Y., Zhao, Z.J., Li, P., Ding, X., Guo, N., Yang, L.L., Zong, Z.H., and Sun, C.F. (2011). NF-kappaB participates in chemokine receptor 7-mediated cell survival in metastatic squamous cell carcinoma of the head and neck. Oncol. Rep. 25, 383–391. https://doi.org/10.3892/or. 2010.1090.
- Mburu, Y.K., Egloff, A.M., Walker, W.H., Wang, L., Seethala, R.R., van Waes, C., and Ferris, R.L. (2012). Chemokine receptor 7 (CCR7) gene expression is regulated by NF-kappaB and activator protein 1 (AP1) in metastatic squamous cell carcinoma of head and neck (SCCHN). J. Biol. Chem. 287, 3581–3590. https://doi.org/10.1074/jbc.M111.294876.
- Schonthaler, H.B., Guinea-Viniegra, J., and Wagner, E.F. (2011). Targeting inflammation by modulating the Jun/AP-1 pathway. Ann. Rheum. Dis. 70, i109–i112. https://doi.org/10.1136/ard.2010.140533.
- Kim, M.J., Im, M.A., Lee, J.S., Mun, J.Y., Kim, D.H., Gu, A., and Kim, I.S. (2019). Effect of S100A8 and S100A9 on expressions of cytokine and skin barrier protein in human keratinocytes. Mol. Med. Rep. 20, 2476– 2483. https://doi.org/10.3892/mmr.2019.10454.
- Wang, S., Song, R., Wang, Z., Jing, Z., Wang, S., and Ma, J. (2018). S100A8/A9 in inflammation. Front. Immunol. *9*, 1298. https://doi.org/10. 3389/fimmu.2018.01298.

- 44. Phadke, A.P., Akangire, G., Park, S.J., Lira, S.A., and Mehrad, B. (2007). The role of CC chemokine receptor 6 in host defense in a model of invasive pulmonary aspergillosis. Am. J. Respir. Crit. Care Med. 175, 1165–1172. https://doi.org/10.1164/rccm.200602-256OC.
- Paradis, T.J., Cole, S.H., Nelson, R.T., and Gladue, R.P. (2008). Essential role of CCR6 in directing activated T cells to the skin during contact hypersensitivity. J. Invest. Dermatol. *128*, 628–633. https://doi.org/10.1038/sj. jid.5701055.
- Hedrick, M.N., Lonsdorf, A.S., Shirakawa, A.K., Richard Lee, C.C., Liao, F., Singh, S.P., Zhang, H.H., Grinberg, A., Love, P.E., Hwang, S.T., and Farber, J.M. (2009). CCR6 is required for IL-23-induced psoriasis-like inflammation in mice. J. Clin. Invest. *119*, 2317–2329. https://doi.org/10. 1172/jci37378.
- 47. Kagami, S., Kakinuma, T., Saeki, H., Tsunemi, Y., Fujita, H., Nakamura, K., Takekoshi, T., Kishimoto, M., Mitsui, H., Torii, H., et al. (2003). Significant elevation of serum levels of eotaxin-3/CCL26, but not of eotaxin-2/CCL24, in patients with atopic dermatitis: serum eotaxin-3/CCL26 levels reflect the disease activity of atopic dermatitis. Clin. Exp. Immunol. *134*, 309–313. https://doi.org/10.1046/j.1365-2249.2003.02273.x.
- Sharma, A., Seow, J.J.W., Dutertre, C.A., Pai, R., Blériot, C., Mishra, A., Wong, R.M.M., Singh, G.S.N., Sudhagar, S., Khalilnezhad, S., et al. (2020). Onco-fetal reprogramming of endothelial cells drives immunosuppressive macrophages in hepatocellular carcinoma. Cell *183*, 377– 394.e21. https://doi.org/10.1016/j.cell.2020.08.040.
- 49. Kramer, R.M., Hession, C., Johansen, B., Hayes, G., McGray, P., Chow, E.P., Tizard, R., and Pepinsky, R.B. (1989). Structure and properties of a human non-pancreatic phospholipase A2. J. Biol. Chem. 264, 5768–5775.
- Suzuki, N., Ishizaki, J., Yokota, Y., Higashino, K., Ono, T., Ikeda, M., Fujii, N., Kawamoto, K., and Hanasaki, K. (2000). Structures, enzymatic properties, and expression of novel human and mouse secretory phospholipase A(2)s. J. Biol. Chem. 275, 5785–5793. https://doi.org/10.1074/jbc.275. 8.5785.
- Ishizaki, J., Suzuki, N., Higashino, K., Yokota, Y., Ono, T., Kawamoto, K., Fujii, N., Arita, H., and Hanasaki, K. (1999). Cloning and characterization of novel mouse and human secretory phospholipase A(2)s. J. Biol. Chem. 274, 24973–24979. https://doi.org/10.1074/jbc.274.35.24973.
- Tian, T., Li, C.L., Fu, X., Wang, S.H., Lu, J., Guo, H., Yao, Y., Nan, K.J., and Yang, Y.J. (2018). beta1 integrin-mediated multicellular resistance in hepatocellular carcinoma through activation of the FAK/Akt pathway. J. Int. Med. Res. 46, 1311–1325. https://doi.org/10.1177/0300060517740807.
- Zhao, D.H., Hong, J.J., Guo, S.Y., Yang, R.L., Yuan, J., Wen, C.Y., Zhou, K.Y., and Li, C.J. (2004). Aberrant expression and function of TCF4 in the proliferation of hepatocellular carcinoma cell line BEL-7402. Cell Res. 14, 74–80. https://doi.org/10.1038/sj.cr.7290205.
- Huang, H., Fujii, H., Sankila, A., Mahler-Araujo, B.M., Matsuda, M., Cathomas, G., and Ohgaki, H. (1999). Beta-catenin mutations are frequent in human hepatocellular carcinomas associated with hepatitis C virus infection. Am. J. Pathol. *155*, 1795–1801. https://doi.org/10.1016/s0002-9440(10)65496-x.
- Moh-Moh-Aung, A., Fujisawa, M., Ito, S., Katayama, H., Ohara, T., Ota, Y., Yoshimura, T., and Matsukawa, A. (2020). Decreased miR-200b-3p in cancer cells leads to angiogenesis in HCC by enhancing endothelial ERG expression. Sci. Rep. 10, 10418. https://doi.org/10.1038/s41598-020-67425-4.
- Arbuthnot, P., Kew, M., and Fitschen, W. (1991). c-fos and c-myc oncoprotein expression in human hepatocellular carcinomas. Anticancer Res. *11*, 921–924.
- 57. Damdinsuren, B., Nagano, H., Kondo, M., Yamamoto, H., Hiraoka, N., Yamamoto, T., Marubashi, S., Miyamoto, A., Umeshita, K., Dono, K., et al. (2005). Expression of ld proteins in human hepatocellular carcinoma: relevance to tumor dedifferentiation. Int. J. Oncol. 26, 319–327.
- Cohen, M., Giladi, A., Gorki, A.D., Solodkin, D.G., Zada, M., Hladik, A., Miklosi, A., Salame, T.M., Halpern, K.B., David, E., et al. (2018). Lung single-cell signaling interaction map reveals basophil role in macrophage



### CellPress

imprinting. Cell 175, 1031–1044.e18. https://doi.org/10.1016/j.cell.2018. 09.009.

- Shibata, Y., Berclaz, P.Y., Chroneos, Z.C., Yoshida, M., Whitsett, J.A., and Trapnell, B.C. (2001). GM-CSF regulates alveolar macrophage differentiation and innate immunity in the lung through PU.1. Immunity 15, 557–567. https://doi.org/10.1016/s1074-7613(01)00218-7.
- Guilliams, M., De Kleer, I., Henri, S., Post, S., Vanhoutte, L., De Prijck, S., Deswarte, K., Malissen, B., Hammad, H., and Lambrecht, B.N. (2013). Alveolar macrophages develop from fetal monocytes that differentiate into long-lived cells in the first week of life via GM-CSF. J. Exp. Med. 210, 1977–1992. https://doi.org/10.1084/jem.20131199.
- Ginhoux, F., and Jung, S. (2014). Monocytes and macrophages: developmental pathways and tissue homeostasis. Nat. Rev. Immunol. 14, 392–404. https://doi.org/10.1038/nri3671.
- Jones, C.V., Alikhan, M.A., O'Reilly, M., Sozo, F., Williams, T.M., Harding, R., Jenkin, G., and Ricardo, S.D. (2014). The effect of CSF-1 administration on lung maturation in a mouse model of neonatal hyperoxia exposure. Respir. Res. *15*, 110. https://doi.org/10.1186/s12931-014-0110-5.
- Wei, F., Ge, Y., Li, W., Wang, X., and Chen, B. (2020). Role of endothelin receptor type B (EDNRB) in lung adenocarcinoma. Thorac. Cancer 11, 1885–1890. https://doi.org/10.1111/1759-7714.13474.
- Dueck, H., Eberwine, J., and Kim, J. (2016). Variation is function: are single cell differences functionally important?: testing the hypothesis that single cell variation is required for aggregate function. BioEssays *38*, 172–180. https://doi.org/10.1002/bies.201500124.
- Osorio, D., Yu, X., Zhong, Y., Li, G., Yu, P., Serpedin, E., Huang, J.Z., and Cai, J.J. (2019). Single-cell expression variability implies cell function. Cells 9, E14. https://doi.org/10.3390/cells9010014.
- Cao, K., Hong, Y., and Wan, L. (2021). Manifold alignment for heterogeneous single-cell multi-omics data integration using Pamona. Bioinformatics 38, 211–219. https://doi.org/10.1093/bioinformatics/btab594.
- Jiang, Q., Zhang, S., and Wan, L. (2022). Dynamic inference of cell developmental complex energy landscape from time series single-cell transcriptomic data. PLoS Comput. Biol. 18, e1009821. https://doi.org/10. 1371/journal.pcbi.1009821.

 Wang, S., Karikomi, M., MacLean, A.L., and Nie, Q. (2019). Cell lineage and communication network inference via optimization for single-cell transcriptomics. Nucleic Acids Res. 47, e66. https://doi.org/10.1093/nar/ gkz204.

**Cell Systems** 

- Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods 14, 979–982. https://doi.org/10.1038/nmeth.4402.
- Xu, Q., Li, G., Osorio, D., Zhong, Y., Yang, Y., Lin, Y.T., Zhang, X., and Cai, J.J. (2022). scInTime: A computational method leveraging single-cell trajectory and gene regulatory networks to identify master regulators of cellular differentiation. Genes (Basel) *13*, 371. https://doi.org/10.3390/ genes13020371.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell *177*, 1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.
- Stiefel, E. (1935). Richtungsfelder und fernparallelismus in n-dimensionalen Mannigfaltigkeiten. Commentarii Math. Helv. 8, 305–353.
- Liaw, R., Liang, E., Nishihara, R., Moritz, P., Gonzalez, J.E., and Stoica, I. (2018). Tune: a research platform for distributed model selection and training. Preprint at arXiv. https://doi.org/10.48550/arXiv.1807.05118.
- Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128. https:// doi.org/10.1186/1471-2105-14-128.
- Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics 33, 2938–2940. https://doi.org/10.1093/bioinformatics/btx364.
- Carneiro, T., Medeiros Da Nobrega, R.V., Nepomuceno, T., Bian, G., De Albuquerque, V.H.C., and Filho, P.P.R. (2018). Performance analysis of Google Colaboratory as a tool for accelerating deep learning applications. IEEE Access 6, 61677–61685. https://doi.org/10.1109/ACCESS.2018. 2874767.

### Cell Systems Methods



### **STAR \* METHODS**

### **KEY RESOURCES TABLE**

REAGENT or RESOURCE	SOURCE	IDENTIFIER	
Deposited data			
Inflammatory skin data set	He et al. <sup>27</sup>	GEO: GSE147424	
Hepatocellular carcinoma data set	Sharma et al. <sup>48</sup>	GEO: GSE156337	
Lung microenvironment data set	Cohen et al. <sup>58</sup>	GEO: GSE119228	
Software and algorithms			
scTenifoldXct	This paper	https://doi.org/10.5281/ zenodo.7453377	
R	https://www.cran.r-project.org	v4.0.4	
Python	https://www.python.org/	v3.9	
OmniPath	Turei et al. <sup>1</sup>	v1.0.4	
Seurat (R package)	https://cran.r-project.org/web/ packages/Seurat	v4.0.2	
Ray Tune (Python package)	https://www.ray.io/ray-tune	v1.8.0	
Pytorch (Python package)	https://pytorch.org	v1.9.0	
igraph (Python package)	https://igraph.org	v0.9.6	
LIANA (R package)	Dimitrov et al. <sup>21</sup>	v0.0.5	
UpSetR (R package)	https://cran.r-project.org/web/ packages/UpSetR	v1.4.0	
CytoTalk (R package)	Hu et al. <sup>19</sup>	v0.99.0	

### **RESOURCE AVAILABILITY**

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, James J. Cai (jcai@tamu.edu).

### **Materials availability**

This study did not generate new materials.

### Data and code availability

- This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table. Known LR pairs (n = 8,047) were retrieved from OmniPath (v1.0.4),<sup>1</sup> a knowledge database of intracellular and intercellular signaling pathways. TF gene names (n = 1,564) were downloaded from the Human TFome database.<sup>2</sup>
- All original code has been deposited at GitHub (https://github.com/cailab-tamu/scTenifoldXct) and is publicly available as of the date of publication. This repository has been archived at Zenodo (https://doi.org/10.5281/zenodo.7453377).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

### **METHOD DETAILS**

#### **Data preprocessing**

The input of scTenifoldXct is a gene-by-cell count matrix with annotated cell types. In addition, sample information is required for the comparative analysis. For all scRNA-seq data sets, we performed log-normalization using the *NormalizeData* function in Seurat (v4.0.2).<sup>71</sup> Highly variable genes were selected using the *FindVariableFeatures* function in Seurat (selection.method = "vst"). For each gene, Seurat computed the standardized variance of its expression across cells, controlling for the mean expression.<sup>71</sup> For each data set, the top 3,000 highly variable genes were included in subsequent analyses. Cell annotations from original studies were retained and used.

### CelPress

### Cell Systems Methods

### **Network construction**

Given cell types *A* and *B*, we employed PC regression to construct the intracellular gene regression networks, denoting them as  $W_A$  and  $W_B$ . Specifically, suppose  $X \in \mathbb{R}^{n \times p}$  is the count matrix with *n* genes and *p* cells for cell type *A*. The gene expression level of the *i*<sup>th</sup> gene in all *p* cells is represented by the *i*<sup>th</sup> row of **X**, denoted by  $X_i \in \mathbb{R}^p$ . Denote  $X_{-i} \in \mathbb{R}^{(n-1) \times p}$  as the matrix by deleting  $X_i$  from **X**. To estimate the effects of other n - 1 genes on the *i*<sup>th</sup> gene, we constructed the PC regression model for  $X_i$ . First, we applied principal component analysis (PCA) to  $X_{-i}^{-}$  and selected the first *M* leading PCs to construct  $Z^i = (Z_1^i, \dots, Z_M^i) \in \mathbb{R}^{p \times M}$ , where  $Z_m^i \in \mathbb{R}^p$  is the  $m^{\text{th}}$  PC of  $X_{-i}^{-}$ ,  $m = 1, 2, \dots, M$  and  $M \ll \min(n, p)$ . Because PC regression only uses *M* PCs as covariates, it reduces the risk of overfitting and computation time. Denote  $V^i \in \mathbb{R}^{(n-1) \times M}$  as the PC loading matrix for the first *M* leading PCs, then  $Z^i = X_{-i}^T V^i$  where  $V^i$  satisfies  $(V^i)^T V^i = I_M$ . The next step was to generate the regression coefficients by regressing  $X_i$  on  $Z^i$  and solve the following optimization problem:

$$\widehat{\beta}^{i} = \arg \min_{\beta^{i} \in B^{M}} \|X_{i} - Z^{i}\beta^{i}\|_{2}^{2}$$

which can be easily solved by the ordinary least square (OLS) method. Then, the effects of the other n - 1 genes to the *i*<sup>th</sup> gene were obtained by  $\hat{\alpha}^i = V^i \hat{\beta}^i \in \mathbb{R}^{n-1}$ . Repeating this process for another n - 1 times, with one different gene as the response gene each time, we assembled  $\{\hat{\alpha}^i\}_{i=1}^n$  together and constructed an  $n \times n$  weighted adjacency matrix  $\boldsymbol{W}$  of the intracellular gene regression network. The *i*<sup>th</sup> row of  $\boldsymbol{W}$  is  $\hat{\alpha}^i$ , and all diagonal entries of  $\boldsymbol{W}$  are all 0. Eventually, the gene regression network  $W_A$  for cell type A and gene regression network  $W_B$  for cell type B were obtained.  $W_A$  and  $W_B$  could be symmetrized when necessary.

#### **Crosstalk score**

The product of the mean expression of known ligands and receptors has been widely used in most computational approaches to gauge cellular interactions. To account for the gene variance within cell groups, we additionally incorporated gene expression variances in the definition of the crosstalk score between gene *i* in cell *A* and gene *j* in cell *B*, as shown below:

$$\text{score}_{iA,iB} = \left[ (1 - \lambda) u_{iA}^2 + \lambda \sigma_{iA}^2 \right] \times \left[ (1 - \lambda) u_{iB}^2 + \lambda \sigma_{iB}^2 \right],$$

where  $u_{iA}$ ,  $\sigma_{iA}^2$  and  $u_{jB}$ ,  $\sigma_{iB}^2$  are the expression mean and variance of gene *i* in cell *A* and gene *j* in cell *B*, respectively. The hyperparameter  $\lambda$  scales the relative contribution of mean and variance. By default,  $\lambda = 1/2$ , and the crosstalk score could be treated as the product of half of the second moment between gene *i* in cell *A* and gene *j* in cell *B*.

#### **Manifold alignment**

Manifold alignment is a nonlinear feature projection method by which we embed the genes of sender and receiver cells into a unified space while simultaneously minimizing the distance between corresponding genes and preserving the original structure of gene regression networks of each cell type. Manifold alignment allows the low dimensional projections of genes to be comparable and preserves the information of gene regression networks. To summarize the strength of interactions for each pair of genes across the cell types, we defined the crosstalk score matrix **S** with  $S_{ij} = \mu \text{ score}_{iAjB}$ , where  $\mu$  is a scale factor indicating how much we want to preserve the weight of intercellular correspondences relative to the edge weight of gene regression networks. Following  $\sum_{i=1}^{N} (W_{B})_{ii} + \sum_{i=1}^{N} (W_{B})_{ii}$ 

Vu et al.,<sup>25</sup> we set  $\mu = \frac{\sum_{i,j}^{(W_A)_{ij}} + \sum_{i,j}^{(W_B)_{ij}}}{2\sum_{i,j} \text{score}_{iA_{jB}}}$ . In this way we ensured the correspondence and gene regression networks are in a compa-

rable scale so that the manifold alignment result is not biased toward either metric. We found that the above value setting of  $\mu$  generated a more robust and unbiased detecting result than other scaling settings, including  $\mu = 1$  (unscaled), 0.1 $\mu$ , and 10 $\mu$  (Figure S4B). The joint similarity matrix is then constructed as follows:

$$W = \begin{bmatrix} W_A & S \\ S^T & W_B \end{bmatrix}.$$

Note that **S** is asymmetric, while, when  $W_A$  and  $W_B$  are symmetric, **W** is symmetric.  $W_A$  and  $W_B$  may contain negative values when gene expressions are negatively correlated, and in such cases, the properties of the corresponding Laplacian are not well understood.<sup>26</sup> We resolved this problem by adding 1 to all entries of  $W_A$  and  $W_B$ , transforming the range of  $W_A$  and  $W_B$  from [-1, 1] to [0, 2]. As a result, the projected features of two genes with a positive correlation would be closer than those with a negative correlation. For convenience, we still used  $W_A$  and  $W_B$  to denote the transformed gene regression networks of the two cell types. The loss function for this manifold alignment is

$$\ell(\mathcal{H}) = \sum_{ij} ||\mathcal{H}_{i} - \mathcal{H}_{j}||^{2} W_{ij} = \sum_{ij} ||\mathcal{H}_{i}^{A} - \mathcal{H}_{j}^{A}||^{2} W_{ij}^{A} + \sum_{ij} ||\mathcal{H}_{i}^{B} - \mathcal{H}_{j}^{B}||^{2} W_{ij}^{B} + 2\sum_{ij} ||\mathcal{H}_{i}^{A} - \mathcal{H}_{j}^{B}||^{2} S_{ij},$$

### Cell Systems Methods

where  $\mathcal{H} = \begin{bmatrix} \mathcal{H}^{A} \\ \mathcal{H}^{B} \end{bmatrix} \in \mathbb{R}^{2n \times d}$  is the low-dimensional representation for cell type *A* and cell type *B*, and *d* (  $\ll$  *n*) is the dimension of the latent space. Moreover, we needed an additional constraint

 $\mathcal{H}^T D \mathcal{H} = I_d$ 

for this loss function to work properly, i.e., to avoid the trivial solution of mapping all instances to zero, where **D** is a diagonal matrix with  $D_{ii} = \sum_{i} W_{ij}$  and  $I_d$  is a  $d \times d$  identity matrix. Solutions for manifold alignment traditionally rely on the eigen decomposition, which is

computationally demanding. To speed up, neural networks were used to learn the unified low-dimensional latent representation  $\mathcal{H}$ .<sup>24</sup>

Let  $\mathcal{H} = \begin{bmatrix} \mathcal{H}^{A} \\ \mathcal{H}^{B} \end{bmatrix} = \begin{bmatrix} \mathcal{H}^{A}(\cdot; \theta^{A}) \\ \mathcal{H}^{B}(\cdot; \theta^{B}) \end{bmatrix}$ , where  $\theta^{A}$  and  $\theta^{B}$  are parameters for two neural networks. By minimizing the loss function  $\ell(\mathcal{H}; \theta_{A}, \theta^{B})$ 

 $\theta_B$ ), we obtained parameters  $\hat{\theta}_A$  and  $\hat{\theta}_B$ . To guarantee that our solution  $\hat{\mathcal{H}} = \mathcal{H}(\cdot; \hat{\theta}_A, \hat{\theta}_B)$  satisfies the constrain  $\hat{\mathcal{H}}^T D \hat{\mathcal{H}} = I_d$ , we followed the optimization method described in Nguyen et al.,<sup>24</sup> forcing the outputs of the optimization problem on the Stiefel manifolds.<sup>72</sup> This nonlinear method yielded the low-dimensional representation  $\hat{\mathcal{H}}$ , which reveals the information on both intra- and inter-cellular networks.

Neural networks have an input layer consisting of nodes of samples (cells) and two hidden layers (by default 32/16 hidden units) along with sigmoid nonlinearity, followed by a linear output layer (3 units). All the layers are fully connected. Hyperparameters such as the number of iterations and learning rate were optimized to ensure the functionality and reproducibility of the networks for single-and two-sample analyses. For each data set, we performed a random hyperparameters search of 100 trials by the module Ray Tune  $(v1.8.0)^{73}$  with the hidden neurons randomly selected from 64 and 32, the embedding dimension sampled from 2 to 10, and the learning rate sampled from a uniform distribution from  $10^{-5}$  to 0.1 in log space. Figure S4A shows some experimental runs including the optimal output for each data set. To implement, we initialized two neural networks with uniform weights following PyTorch (v1.9.0) recommended heuristic for linear layers and trained them from scratch. For the inflammatory skin data set of single-sample analysis, we trained the model for 1,000 iterations using the Adam optimizer with a learning rate of 0.01.

### Determining the statistical significance of interactions between cell types in a single sample

With  $\hat{\mathcal{H}}_A = \mathcal{H}^A(\bullet; \hat{\theta}_A)$  and  $\hat{\mathcal{H}}_B = \mathcal{H}^B(\bullet; \hat{\theta}_B) \in \mathbb{R}^{n \times d}$  being the representations of genes from cell types *A* and *B* in the low dimensional embedding, respectively, we calculated the Euclidean distance  $d_{ij}$  across cell types for every pairwise combination of gene *i* in cell *A* 

and gene *j* in cell *B* and denoted the square difference of projected representations between gene *i* and gene *j* as  $d_{ij}^2 = \|\hat{\mathcal{H}}_i^A - \hat{\mathcal{H}}_j^B\|^2$ . We implemented a nonparametric test to identify significant gene pairs among all combinations under the null hypothesis that there is no LR-mediated interaction between gene pairs. The null hypothesis distribution was obtained by collecting  $d_{ij}^2$  of all gene pairs, excluding LR pairs in the OmniPath database. Compared to gene pairs that appear in the database, those gene pairs absent in the database were considered much less likely to be gene pairs that confer LR-mediated interactions. Alternatively, the null distribution could be constructed by random shuffling of the data, which would make the test less conservative. Next, we calculated the quantiles for LR pairs under the null hypothesis distribution and set them as the original *p-values*. The threshold was set to 0.05 for all data sets. We excluded the Chi-square test in this case because the left tail targeting for close gene pairs with short distances is desired and is therefore incompatible with the Chi-square test.

#### **Comparative manifold alignment**

Comparative biology to study the function of gene sets looks for significant differential interactions between two samples, such as healthy (*H*) and diseased (*D*). One direct method is to walk through the above pipeline for two samples separately and obtain the difference  $d_{ij}^{H}$  and  $d_{ij}^{D}$  for different samples. Given that  $d_{ij}^{H}$  and  $d_{ij}^{D}$  were not comparable since the low-dimensional representations for two samples belonged to different latent spaces, we were asked to add constraints to make them numerically comparable so that the same genes from two samples would have similar low-dimensional representations. Following the general manifold alignment framework,<sup>26</sup> we constructed the coupled joint similarity matrix across two samples denoted by *V* as

$$V = \begin{bmatrix} W^H & \beta I \\ \beta I & W^D \end{bmatrix},$$

where  $W^H$  and  $W^D$  are the joint similarity matrices constructed by the above manifold alignment method for each sample, respectively. *I* is an  $2n \times 2n$  identical matrix, and  $\beta$  is a tuning hyperparameter, which is by default 0.9 times the mean value of the row sums of  $W^H$  and  $W^D$ . Intuitively, a smaller factor  $\beta$  would not enforce correspondence between identical genes from different samples, whereas a larger one would produce close distances between identical genes without much consideration of the given gene regression networks. We showed that within an optimal range of the scale factor in  $\beta$ , the aligned distances remained highly correlated (Figure S4C). The loss function is defined as

$$\ell(F) = \sum_{ij} ||F(i,) - F(j,)||^2 V(i,j).$$

### CelPress

### Cell Systems Methods

The solution is denoted as  $\hat{F} = \begin{bmatrix} \hat{F}^{H} \\ \hat{F}^{D} \end{bmatrix} \in \mathbb{R}^{4n \times d}$  where  $\hat{F}^{H}$  and  $\hat{F}^{D} \in \mathcal{R}^{2n \times d}$  are analogy to  $\hat{\mathcal{H}}$  in single-sample analysis, which are the

vertically stacked representations of genes from cell types *A* and *B* in each sample. In the two-sample analysis setting, four neural networks are initiated for solving the optimization problem described previously, the architecture of which are identical with ones used for the single-sample analysis. We performed a random search again for the learning rate. We trained the model for up to 3,000 iterations using the Adam optimizer with the same learning rate of 0.01 as in the single-sample analysis (Figure S4A).

#### Determining LR pair statistical significance of differential interactions between two samples

From  $\hat{F}^H$  and  $\hat{F}^D$ , we calculated the Euclidean distances  $d_{ij}^H$  and  $d_{ij}^D$  for every pairwise combination of gene *i* in cell *A* and gene *j* in cell *B* in samples *H* and *D*, respectively, similar to the single sample scenario. We then computed the squared distance difference of each gene pair across samples as  $\Delta d_{ij}^2 = \|d_{ij}^H - d_{ij}^D\|^2$ . Note that this test statistic would be higher for gene pairs that exhibit a significant difference between two samples. Thus, gene pairs with large  $\Delta d_{ij}^2$  are considered more differentiated. To determine the significance, we implemented the Chi-square test since our test statistic was demonstrated in the sum of square form. Here we set df = 1 to make a conservative selection of gene pairs with high precision. By using the right tail P(X > x) of the Chi-square distribution, we assigned *p*-values for each gene pair. Finally, we implemented FDR to generate the adjusted *p*-values and selected significant gene pairs with adjusted *p*-value < 0.05.

#### Validate the predicted interactions between LR pairs

After the Chi-square test, significant LR interactions were queried against the OmniPath database.<sup>1</sup> The predicted LR pairs in the database were retained for subsequent functional enrichment analyses performed by Enrichr.<sup>74</sup>

#### **Visualization of integrated networks**

Python package igraph (v0.9.6) was used to generate network plots. In each of the network plots, only direct connections between TFs and the enriched LR pair of interest were shown, and the LR interaction itself was highlighted in green. The edge thickness was adjusted to be proportional to the absolute value of the coefficient between gene pairs in the intracellular networks. Positive and negative coefficients were indicated in red and blue, respectively.

### Systematic comparisons between scTenifoldXct and existing tools

We compared scTenifoldXct with NATMI,<sup>8</sup> SingleCellSignalR,<sup>9</sup> Connectome,<sup>12</sup> iTALK,<sup>10</sup> and CellChat.<sup>11</sup> The test data was the inflammatory skin scRNA-seq data set. Two cell types were fibroblasts and dendritic cells. For all the methods, the reference LR database was OmniPath.<sup>1</sup> The comparison analysis was facilitated by using LIANA (v0.0.5).<sup>21</sup> To show the overlap between significant results, the upset plot was generated using UpSetR (v1.4.0).<sup>75</sup> Equal numbers of pairs ranked by each method's default scoring metric were retained. CytoTalk (v0.99.0)<sup>19</sup> was executed independently for reporting significant signaling pathways, and LR pairs from those pathways were subsequently used for the comparison. For plotting the ROC and precision-recall curves, a total of 160 LR pairs, for which both ligand and receptor genes were expressed in the two cell types, were included in the evaluation. Among them, 40 consensus pairs detected by all five methods except scTenifoldXct were used as positive pairs. The remaining 120 were used as negative pairs. The computation cost of scTenifoldXct was estimated on the public platform Google Colaboratory,<sup>76</sup> and the results are available in Table S5.