# Charting spatial ligand-target activity using Renoir

Narein Rao[1], Rhea Pai[2,3], Archita Mishra[4,5], Florent Ginhoux[6,7], Jerry Chan[5], Ankur Sharma[2,3,5,8,*], and Hamim Zafar[1,9,10,*]

[1]Department of Computer Science and Engineering, Indian Institute of Technology Kanpur, India
[2]Harry Perkins Institute of Medical Research, Nedlands, Perth, Western Australia
[3]Curtin Medical School, Curtin University, Perth, Western Australia
[4]Telethon Kids Institute, University of Western Australia, Perth, Australia
[5]KK Research Center, KK Women's and Children's Hospital, Singapore
[6]Singapore Immunology Network (SIgN), Agency for Science, Technology and Research (ASTAR), Singapore
[7]Gustave Roussy Cancer Campus, Villejuif, France
[8]Institute of Molecular and Cellular Biology (IMCB), Agency for Science, Technology and Research (ASTAR)
[9]Department of Biological Sciences and Bioengineering, Indian Institute of Technology Kanpur, India
[10]Mehta Family Center for Engineering in Medicine, Indian Institute of Technology Kanpur, India
[*]*Corresponding Author*
[1,9,10]hamim@iitk.ac.in (H.Z.), [2,3,5,8]ankur.sharma@perkins.org.au (A.S.)

## Abstract

The advancement of single-cell RNA sequencing (scRNA-seq) and spatial transcriptomics has made it possible to infer interactions amongst heterogeneous cells and their surrounding cellular environments. Existing methods assist in the analysis of ligand-receptor interactions by either adding spatial information to the currently available scRNA-seq data or utilizing spot-level or high-resolution spatial transcriptomics data. However, till date, there is a lack of methods capable of mapping ligand-target interactions across a spatial topology with specific cell type composition, with the potential to shed further light on the niche-specific relationship between ligands and their downstream targets. Here we present Renoir for charting the ligand-target activities across a spatial topology and delineating spatial communication niches harboring specific ligand-target activities and cell type composition. Renoir can also spatially map pathway-level aggregate activity of ligand-target gene sets and identify domain-specific activities between ligands and targets. We applied Renoir to three spatial datasets ranging from development to disease to demonstrate its effectiveness in inferring cellular niches with distinct ligand-target interactions, spatially mapping hallmark pathway activities, ranking ligand activity across spatial niches, and visualizing overall cell type-specific, ligand-target interactions in spatial niches.

## keywords

Cell-cell communication, ligand-target activity, spatial transcriptomics, pathway

## Introduction

High-resolution single-cell RNA sequencing (scRNA-seq) data as well as spatial transcriptomics that retain information on spatial localization are paramount to studying cell-cell communication. Methods for inferring cell-cell communication have been developed that rely solely on scRNA-seq and utilise pre-existing ligand-receptor databases such as CellChat (1) and CellphoneDB (2), manually curated databases such as iCELLNET (3), iTALK (4), NATMI (5), scTensor (6), and SingleCellSignalR (7), or are independent of such databases (e.g., COMUNET (8) and PyMINEr (9)). Although these methods have provided robust approaches to infer potential ligand-receptor interactions, they lack the spatial context of cell type arrangement in the tissue , which is necessary when inferring cellular interactions. More importantly, cellular communications are often spatially constrained as signalling pathways are activated when ligands diffuse from sender cells to neighbouring receiver cells in close proximity (6). To address this challenge, additional approaches have been developed which either spatially map previously collected scRNA-seq data such as SpaOTsc (10), or make use of highly informative spatial transcriptomic data such as CellPhoneDBv3 (11), Cell2Cell (12), Giotto (13), MISTy (14), stLEARN (15), SVCA (16), NICHES (17) and COMMOT (18). However, some of these methods do not take into account the fine-grained spatial distribution of cell types, or prior information regarding plausible ligand-receptor interactions from scRNA-seq data or examine downstream target gene response caused by ligand–receptor binding. On the other hand, methods such as Cytotalk (19), scMLnet (20), NicheNet (21), SoptSC (22) and CCCExplorer (23) can infer the interactions of ligands with possible downstream targets by making use of bulk data or scRNA-seq data, however, the informative spatial context is still lacking. Very recently, a handful of methods such as COMMOT and SpaOTsc, attempted to investigate the correlation between ligand-receptor interactions and target gene expression, but these approaches were unable to clearly exhibit probable ligand-target relationships over a spatial context. As a result, there is a critical requirement for reliable computational approaches that can offer a unique perspective on the probable spatial influence that ligands can have on known targets in a particular spatial context with unique cell type composition.

In this paper, we introduce Renoir (ligand-taRgEt iNteractions acrOss spatIal topogRaphy), an end-to-end computational framework for spatially charting the ligand-target activities and inferring

spatial communication niches that harbor specific ligand-target activities and cell type composition. Renoir requires spatial transcriptomics data as well as scRNA-seq data from the same tissue in order to delineate the spatial influence between possible ligand-target pairs curated from an existing database (21). Therefore, Renoir allows one to determine which ligands lead to the activation of which target genes in a given spatial context. For each curated ligand-target pair, Renoir quantifies a neighborhood activity score for each spatial spot of the spatial transcriptomics data by taking into account the cell type abundances, cell type-specific mRNA abundance values, cell type-specific gene entropy measure and a cell type-specific mutual information between the ligand-target pairing as obtained from scRNAseq data. The neighborhood activity score computed by Renoir denotes the strength of the ligand-target activity within the context of a spot's neighbourhood.

In addition to mapping potential ligand-target relationships throughout a spatial topology, the neighbourhood scores computed by Renoir further enables a wide variety of downstream analyses. Using the neighbourhood scores, Renoir can infer communication niches, which constitute tissue microenvironments harboring specific ligand-target activities pertaining to major co-localized cell types The neighbourhood scores also allow for the spatial mapping of the total activity of gene sets corresponding to well-established signaling pathways as well as inferred denovo through unsupervised clustering. Renoir can also identify the major ligands active in a communication niche and rank them based on their cumulative activity across the target genes expressed in the niche. As a proof-of-principle, we applied Renoir to three distinct spatial transcriptomics datasets ranging from development to disease, including adult mouse brain, human fetal liver, and human triple negative breast cancer (TNBC), for each of which Renoir uncovered distinct spatial domains characterized by specific ligand-target activities. Renoir characterized the interactions of regional astrocyte subtypes with other spatially co-localized cell types in the mouse brain, uncovered hepatocyte-macrophage interactions in developing fetal liver, and identified several spatially restricted interactions between cancer, stromal and immune cell populations in TNBC.

# Results

## Charting spatial ligand-target interactions using Renoir

In order to build a comprehensive computation framework to investigate spatial cell-cell interaction we developed Renior. As shown in Figure 1, Renoir is an end-to-end computational framework for exploring the spatial map of ligand-target activities by integrating spatial transcriptomics and scRNA-seq datasets from the same tissue. Renoir first curates a set of ligand-target pairs using NATMI's ConnectomeDB (5) and NicheNet (21). Using a well-annotated scRNA-seq data and the spatial transcriptomic dataset, Renoir quantifies a neighborhood activity score for each curated ligand-target pair for each spot in the spatial transcriptomic data. The quantification of the neighborhood activity score for a specific ligand-target pair for a particular spot is performed using the proportions and cell type-specific mRNA abundances present in each spot within the defined neighborhood of the spot (see Methods for details). The activity between a ligand-target pair for the given spot and another spot in the neighborhood is scored based on the cell type–specific mRNA abundances of the two genes weighted by the cell type proportions as well as the mutual information between the two genes across the cell types present within the two spots. Thus, Renoir generates a novel neighborhood activity score matrix (of dimensions number of ligand-target pairs $\times$ number of spots). Using the ligand-target neighborhood activity score matrix, Renoir performs several downstream analyses. First of all, Renoir infers spatial communication domains through unsupervised clustering of spots based on ligand-target neighborhood activity scores and determines communication domain-specific ligand-target activities. Moreover, Renoir can infer cell type-specific ligand-target activities within a communication domain. Renoir further provides an algorithm for ranking the ligands within a communication domain based on their cumulative activity across target genes expressed by the major cell types in the domain. Renoir also provides a mapping of pathway-level activity across spots and communication domains. Finally, Renoir encompasses a visualization module for the visualisation of ligand-target activity across spots as well as a Sankey plot that provides an overview of user-specified ligand-target activities across cell types and communication domains.

## Renoir identifies distinct communication domains in Mouse brain

To demonstrate Renoir's ability to uncover region-specific ligand-target interactions, we first applied Renoir to 10X Visium data from two Mouse brain sections (samples S1 and S2) and its corresponding single nuclei data (24) (see data availability). Renoir identified 9409 (S1) and 9405 (S2) ligand-target pairs that were expressed in this dataset and quantified their spatial neighborhood activity scores for both samples. Based on the ligand-target neighborhood activity scores, Renoir clustered the ST spots across the two tissue sections and identified 19 spatial communication domains which had distinct latent space embeddings (Figure 2a) and were similarly distributed across the sections (Figure 2b). We further quantified the distribution of the spatial communication domains across different anatomical brain regions and found that the domains across both samples primarily corresponded to the same anatomical regions and were highly concordant across the two samples in terms of their distribution in anatomical brain regions (Figure 2c).

The communication domains were found to harbor diverse cell types that were spatially localized and abundance pattern concordant across the replicates (Supplementary Figure 1a-b). The domains further harbored differential activity of specific ligand-target pairs (Figure 2d, Supplementary Figure 2b-c). For instance, domain 1 specific to white matter exhibited *Il33* and *Csf1* activity which are known to be required for microglia development and maintenance in white matter (25; 26) and domain 3 specific to the hypothalamus exhibited *Dlk1* activity known to be expressed in hypothalamic arginine vasopressin and oxytocin neurons (27). Domains 2 and 13 located in the thalamus region were enriched in *Tcf7l2* and *Lef1* activity which are known to be essential for neurogenesis in the developing mouse neocortex (28; 29). We also noticed the existence of ligand or target-specific activity pertaining to distinct anatomical regions, wherein the ligand or target was known to be abundant within the brain regions corresponding to the domain. For example, domain 11 exhibited high activity of *Rspo1*, which is known to be highly expressed within the cortex (30; 31), domain 10 exhibited high activity of hippocampal marker *Crlf1*, and domain 9 exhibited high activity of *Ccnd1* and *Nr2f2*, which are known to be highly expressed within the Amygdala region (32; 33). Supplementary Table 1 lists the domain-specific ligand-target interactions with literature support for the associated brain region. Furthermore, we also observed section-specific interactions that were specific to S1 or S2. S2 specifically showed *Tgfa* activity within domain 7 which was located

mainly in the striatum region. This is in line with previous findings of TGF-$\alpha$ synthesis across regions which included the tuber olfactorium, a part of the ventral striatum (34).

Kleshchevnikov et al. (24) identified and molecularly characterized 10 regional astrocyte subtypes using this spatial and matched single-cell data. However, interactions of these regional astrocyte subtypes with other spatially co-localized cell types were not explored. We applied Renoir to characterize the interaction of these regional astrocyte subtypes with other cell types. Figure 2d and Supplementary Figure 2d illustrate the interactions where ligands expressed by regional astrocyte subtypes activate other target genes in other cell types across samples S1 and S2 respectively. Figure 2e and Supplementary Figure 2e demonstrate the interactions where ligands expressed by other cell types activate specific marker genes in astrocyte subtypes. Ligands expressed by the telencephalic astrocytes in Cortex, and Amygdala were found to activate genes in excitatory neurons in Layer 6b (*Vegfa:Ctgf*), Amygdala (*Apoe:Ociad2*), and Layer 6 (*Ptn:Per1*) respectively. On the other hand, *Agt* expressed by diencephalic astrocytes (Hypothalamus) were found to activate genes in inhibitory neurons (*Agt:Psmc6*). Lateral Thalamus astrocytes that were abundant in the ventrolateral thalamic nuclei, expressed *Serpine2* which activated different genes in different excitatory neurons in the thalamus. White matter astrocytes interacted with different oligodendrocytes via *Nrg1* which activated the genes *Slc2a1* and *Smad7*. Similarly, different genes in telencephalic astrocytes in Amygdala, Cortex, and Hippocampus were found to be activated by ligands expressed by excitatory neurons in Layer 5 (*Rspo1:Vegfa*), Amygdala (*Bmp3:Sptbn1*), and Hippocampus DG2 (*Il16:Id3*, *Tgb2:Slc1a2*) respectively. Excitatory neurons in the thalamus were found to activate genes in diencephalic astrocytes (Thalamus Habenula (*Igf1:Agt*) and Thalamus Medial (*Ntng1:Gpld1*, *Nxph1:Tcf7l2*)).

For each communication domain, based on the agreement of neighborhood activity and prior regulatory potential scores, we ranked the prominent ligands that are expressed by the major cell types in the domain (see Methods for details). As shown in Figure 2f and Supplementary Figure 2f, the top-ranked ligands for some of the major domains with literature support for sections S1 and S2 respectively. This further highlighted the region-specific enrichment of certain ligands. For instance, *Tgfb1* which is known to be well expressed within the central amygdaloid nucleus (35)(36) showcases high activity within domain 9 which is primarily located in the amygdala. *Nrg1*, known to be expressed within the components of the basal ganglia (37) exhibited high activity within

domain 7 (associated with regions in and around the striatum). Supplementary Table 2 lists such major ligands with their corresponding literature support.

Using Renoir, we spatially mapped major pathway activities based on the activity scores of the ligand-target pairs associated with the pathways (see Methods). We found the pathway activity to be concentrated in brain regions that are typically associated with the said pathway (Figure 2g, Supplementary Figure 2g and Supplementary Figure 3). For example, the cortex and hippocampus were highly enriched in MAPK signalling and NEUROTROPHIN signalling pathways, respectively. This is consistent with prior works which demonstrated high concentrations of MAPK components in the cortex and hippocampus of the adult mouse brain (38), and NEUROTROPHIN signaling's role in the regulation of neurogenesis (39).

### Renoir identifies hepatocyte-macrophage interactions in developing fetal liver

Cell-cell communication shapes tissue functioning and organogenesis during development. In order to evaluate the versatility of Renoir in uncovering cell-cell interactions during development, we applied Renoir to data from fetal liver development. We generated a 10X Visium spatial transcriptomic dataset (see Methods for details) from a fetal liver sample for which single-cell RNA sequencing data was available (40) (Figure 3a). Renoir identified 17630 ligand-target pairs active in the tissue and quantified their spatial neighborhood activity scores. Based on the ligand-target neighborhood activity scores, Renoir identified five spatial communication domains (Figure 3b), which differed based on ligand-target activity and distribution of UMI count across the tissue section (Figure 3b-c and Supplementary Figure 4a). Domain 0 demonstrated the majority of ligand-target activity, followed by domain 1. On the other hand, domains 2 and 3 showed sparse activity (UMI counts were low) (Figure 3c) while domain 4 was specific to the target *CDKN1A* (Supplementary Figure 4a). This is in line with the UMI count distribution across spots, where the UMI count was most abundant in domains 0 and 4, followed by domains 1, 2 and 3. Cell types were found to be evenly distributed across domains 0, 1 and 4 (Supplementary Figure 4b). Activities pertinent to fetal liver were found to be widespread (Figure 3d). For example, *COL18A1* associated with type XVIII collagen (a prominent component of the Extracellular Matrix (ECM) in the liver (41)) was found to be highly expressed across the tissue and activate multiple targets including hematopoietic transcription factors *BCL2L1* and *BCL6* (42). Higher *CDKN1A* activity was observed, in line with

the proliferative nature of hepatocytes in developing fetal liver tissue.

Further analyses involved acquiring previously described ligand rankings and prominent cell type-specific ligand-target interactions for each communication domain. Specifically, we observed prominent activity of the ligands *MDK*, an important growth factor for fetal liver hematopoietic stem cells and multipotent progenitor expansion (43) and *CXCL12* which helps in retaining hematopoietic stem cells in the fetal liver (44) (Figure 3e). We further observed the occurrence of ligand-target pairs *CXCL12-CXCR4* and *HGF-COL1A2* as prominent interactions (Supplementary Figure 5a). While *CXCL12* is known to interact with *CXCR4* to control the regeneration of liver amongst other organs (45), *CXCR4* is known to be expressed at the highest density by the lymphoid cells (46). Moreover, the deletion of one of *CXCL12* or *CXCR4* is known to cause defective B-cell lymphopoiesis (45). Importantly, we also observed $MARCO^+$ tissue-resident macrophage population (Kupffer cells) interacting with $APOB^+$ hepatocytes. *MARCO* (MAcrophage Receptor with COllagenous structure) is predominantly expressed in non-inflammatory Kupffer cells (47) and fetal-like reprogramming of liver macrophages plays an important role in immunosuppressive phenotype in liver cancer (40). As shown in Figure 3e and Supplementary Figure 5a, we observed that $MARCO^+$ macrophages interact with $PLG^+/APOB^+$ subset of hepatocytes. The *PLG* (plasminogen) has been known to be important for hepatic regeneration (48; 49) and *PLG* deficiency can lead to impaired remodelling in mouse models after acute injury. Next, we subdivided hepatocytes into two groups: $ALB^+APOA^+PLG^+$ and $ALB^+APOA^+PLG^-$ (Figures 4f-g). As shown in Figures 4g-h, these two subsets are transcriptionally and spatially distinct from each other indicating heterogeneity of the fetal liver hepatocyte population. We observed that the majority of the $ALB^+APOA^+$ hepatocyte population does not express *PLG* while there is a specific cluster with a high expression of *PLG* (Figure 3f). In the Visium data, the spots expressing $ALB^+APOA^+PLG^+$ hepatocytes also contained $FOLR2^+$ macrophages (Figure 3h). $FOLR2^+$ macrophages are embryonic macrophages present in the fetal liver microenvironment (40). Spatial correlation analyses further showed higher similarity of $FOLR2^+$ macrophages and $ALB^+APOA^+PLG^+$ hepatocytes as compared to $PLG^-$ hepatocytes in terms of cell type distribution as well as neighborhood activity score (Figure 3i).

Major pathway activities were also spatially mapped onto the Fetal Liver sample (Supplementary Figure 5c). Although domain specific activities were undetected, relevant activity notably,

KRAS signaling (plays an important role in Fetal Liver hematopoiesis (50)) and canonical JAK STAT signaling (recognised to be integral in initiating a signalling cascade required for embryonic development, haemopoietic development and differentiation (51)) were distinctly observed. Taken together, these results provide important insights into the macrophage-hepatocyte niche in the developing human liver.

## Renoir uncovers intratumor spatial communication sub-domains within triple negative breast cancer

To demonstrate Renoir's ability to characterize complex heterogeneous tissues in terms of ligand-target interactions, we analyzed single-cell and Visium datasets (Figure 4a) from a Triple Negative Breast Cancer (TNBC) patient (52). From the set of curated ligand-target pairs, Renoir identified 18654 ligand-target pairs that were expressed in this dataset and quantified their spatial neighborhood activity scores. Based on the neighborhood activity scores of ligand-target pairs, Renoir clustered the spatial spots into six communication domains which were also distinct in low-dimensional embeddings (Figure 4b). While each of the spatial communication domains was primarily associated with morphologically distinct pathological regions, some communication domains were distributed across multiple pathological regions (Figure 4c). The domains were further differentiated based on the ligand-target activities and the composition and abundance of the cell types they harbored (Figure 4d-e, Supplementary Figure 6a-b). Since two (clusters 4 and 5) of the six domains had low overall UMI count (Supplementary Figure 6c), we focused on the first four communication domains in our further analyses. In accordance with their respective cell type compositions, interesting differential ligand-target activities were observed across the six domains. For instance, domain 0 which was primarily associated with the histological region annotated as ductal carcinoma in situ (DCIS) and secondarily contained spots pathologically annotated as stroma and lymphocytes, harbored a high abundance of cancer cycling and cancer basal SC cell types (Figure 4c, Supplementary Figure 6a). Moreover, it also harbored cancer-associated fibroblasts (CAFs) and different T cell populations (Supplementary Figure 6a). Domain 0 demonstrated differential activity of *MYC* (e.g., *LAMB1-MYC*, *SERPINE2-MYC*), *MDK* (*MDK-CCNA2*) and *SAA1* (*SAA1-ESRRA*) which were all marker genes for the cancer cell types in this dataset (Figure 4d-e). While *MYC* and *SAA1* are known to be overexpressed in basal-subtype of breast cancer (53; 54; 55),

*MDK* is a known cancer progression marker (56). Furthermore, domain 0 also harbored high activity of *VEGF* (*VEGFA*, *VEGFB*) (known to be highly up-regulated in breast cancer (57; 58; 59)), and other stromal ligands (*COL18A1*, *COL4A1*) (Figure 4e). Domain 1 was associated with the histological regions lymphocytes and invasive cancer and majorly harbored cancer-associated myofibroblasts (myCAF), macrophages, and T cells (cycling and CD8+) along with the two cancer cell types (Figure 4c, Supplementary Figure 6a). Consequently, interactions associated with tumor progression and invasion (e.g., *CXCL10-MMP2*, *CCL3-ICAM1* (60), *FN1-MMP10* (61)), as well as suppression (e.g., *CXCL9-IRF1* (62)) among these cell types were enriched in domain 1 (Figure 4e). Domain 2 was associated with the histological region normal, stroma and lymphocytes and harbored four major cell types including inflammatory-like CAFs (iCAF-like) and myoepithelial cells (Figure 4c, Supplementary Figure 6a). Interestingly, this domain harbored high activity of the ligand *CCL28* which is known to promote breast cancer and metastasis (63), target *HES1* which plays a crucial role in malignant cell proliferation (64) and *AREG* which is expressed by myoepithelial cells and loss of which can inhibit tumor proliferation, growth and invasiveness (65) (Figure 4e). Domains 3 and 4, despite low UMI count, showcased distinct ligand-target activities. For example, domain 3 displayed distinct interleukin activity (e.g., *IL6*, *IL33*, *IL1RN*) which are known to play important roles in metastasis (66),(67),(68) and *PENK* was found to be differentially active in domain 4 (Figure 4e).

We further investigated the most prominent cell type-specific ligand-target interactions (Figure 4f, Supplementary Figure 7a) within each spatial communication domain (see Methods for details). By emphasizing domains with the highest ligand-target activities and UMI count (0, 1, 2, and 3), we uncovered cell type-specific interactions such as *RARRES2-AEBP1*, *CXCL12-GSN* (ligand and target are markers of iCAFs, domain 2). We also noticed cross-communication between myCAFs and CD4+ T cells (*CXCL12-ITK*) in domain 1. Tumor metastasis and progression-promoting interactions between macrophages and cancer cycling cells (*CXCL16-CCND1*) (69; 70) were observed in domain 3. Additionally in domain 3, we observed interaction between *TGFB1* and *FOXP3* which are markers of NK cells and CD4+ T cells respectively indicating *TGFB1*-mediated suppression of immune response (71; 72).

For each communication domain, we further ranked the major ligands (expressed by the major cell types in the domain) according to their activities (see Methods for details). A key EMT

marker, *TGFB3* (73) was one of the top-ranked ligands in both domains 0 and 1 (which harbored a high abundance of cancer cells) and its top targets included *FN1* (in myCAFs and macrophages) and *TNC* (in myoepithelial cells and myCAFs) which are associated with breast cancer metastasis (74), (75) (Figure 4g). In domain 1 which contained high abundance of macrophages, another top ligand was *CCL2*, which is known to promote metastasis via retention of metastasis-associated macrophages (MAMs) (76) and it was found to target *ICAM1* (a molecular target for TNBC) (60) in myoepithelial cells and macrophages, other chemokines *CCL3* and *CCL4* as well as lymphocyte inhibitory molecule *LTB* in CD4+ T cells (Figure 4g). Within the top ligand lists of each communication domain, we also found ligands that were specifically active in a single communication domain (Figure 4h). For example, in domain 1, CD8+ T cells secreted the ligand *CCL5* (known to promote TNBC progression through the regulation of immunosuppressive cells (77)) which further activated *MMP9* and *CCL2* in macrophages.

Subsequently, utilising cell type abundance values of spatial locations, we sub-clustered each communication domain (domains 0, 1, 2, and 3) into subdomains that were unique in terms of cell type abundance and entailed distinct ligand-target activities (Supplementary Figures 8-11). Amongst the five subdomains in domain 0 (Supplementary Figure 8), subdomain 2 was rich in myCAFs, myoepithelial cells, and endothelial cells and it exhibited *TGFB1* and *FBN1* activity, where *TGFB1* activated several EMT-associated genes (*FBN1*, *MMP2*, *COL5A1*, *PDGFRB*, etc.) (78) and *FBN1*, as produced by myCAFs, (79; 80), was found to activate cancer progression and regulation markers (*DUSP1*, *JUNB*, and *MYC*) (81; 82). Contrastingly, subdomain 3 which harbored different immune cells (CD4+ and CD8+ T cells, DCs, NKT cells) also was abundant in iCAFs, and subsequently showcased extensive chemokine activity (*CCL5*) which targeted different genes in different immune cells and iCAFs (e.g., *EGRN*). Domain 1 was segregated into 3 distinct subdomains (Supplementary Figure 9), of which subdomain 0 abundant with immune cells showcased high *RARRES2* activity (secreted by myCAFs) and subsequently a lower abundance of cancer cells compared to other subdomains. This is coherent with previous findings which show that *RARRES2* serves as an immune-dependent tumor suppressor by acting as a chemo-attractant to recruit anti-cancer immune cells (83; 84; 85; 86). The subdomain 1, which harbored a high abundance of macrophages, myoepithelial cells, myCAFs, and cycling myeloid cells displayed *FBN1* and *IL1B*-mediated activation of *MMP1*, which is known to promote brain metastasis in TNBC (87).

Additionally, subdomain 2 abundant with cancer cell types showcased high activity of *TGFA*, a ligand associated with tumor invasion (88) and *DKK1-HES1*, which are known to promote tumor metastasis, proliferation and invasion (89; 64). High *DLK1* (a prominent marker for iCAFs (90)) activity was observed in subdomain 2 of domain 2 (Supplementary Figure 10) as it was abundant with iCAFs. Interestingly, the only subdomain (subdomain 3) of domain 2 that harbored cancer basal cell type showed *NTN4*- (expressed by myoepithelial and mature luminal cells) and *SEMA3C*- (expressed by mature luminal cells) mediated activity of *FOXA1* indicating potential roles of these co-localized cell types in EMT-progression (91) (Supplementary Figure 10). Domain 3 was segregated into five subdomains (Supplementary Figure 11) of which subdomain 0 abundant with immune cells showed high *CCL19* and *CCL21* activity which are known to attract lymphocytes (92) and subdomain 1 abundant with cancer cells, myCAFs, and myoepithelial cells showcased activity of *SERPING1* (metastasis-predictor extracellular matrix gene (93)), *TGFA*, *COL4A1*, and *SPP1* respectively.

Using Renoir, we further spatially mapped the major pathway activities based on the activity scores of the ligand-target pairs associated with the pathways (see Methods). While some pathways showed activity across multiple communication domains, some other pathways were more active in specific communication domains (Supplementary Figure 7b). For example, the WNT signaling pathway was found to be active in subdomain 1 of domain 0, and subdomain 2 of domain 1 which harbored a high abundance of cancer cell types (Figure 4i). IL-6/JAK/STAT3 signaling was highly active in subdomain 1 of domain 1 (Figure 4i) that harbored a high abundance of myCAFs, macrophages and myoepithelial cells along with cancer cells but lacked immune cells other than cycling T-cells indicating a highly immunosuppressive tumour microenvironment (94). TGF beta signaling was active in domains 0 and 1, and MAPK signaling had the highest activity in domain 2 (Figure 4i). Taken together, Renoir was able to provide comprehensive insights into the spatial heterogeneity of ligand-target activity and its implication in tumor biology.

## Discussion

Cell-cell interactions play a crucial role in various biological processes, including tissue development, homeostasis, and disease. These interactions are essential for the proper functioning of multicellular

organisms, as they facilitate communication and coordination among different cell types. Single-cell RNA-seq approach has allowed us to comprehend the heterogeneity of cell types in a complex tissue. However, traditional single-cell RNA-seq data lacks spatial context, making it challenging to uncover the spatial organization and intercellular relationships that underpin complex biological systems. Spatial transcriptomics is a burgeoning field that addresses this limitation by providing spatially resolved gene expression data, enabling researchers to explore cells in a spatial context and bridging the gap between molecular and morphological observations.

In this regard, our newly developed computational method, Renoir, offers a powerful tool for investigating cellular interactions within their native spatial context, from development to disease. Renoir enables the spatial mapping of ligand-target activities through the quantification of a neighborhood activity score, infers spatial communication domains by grouping spots based on the similarity of ligand-target activities and uncovers ligand-target interactions restricted to each communication domain. Renoir further enables the identification of the major ligands active in each domain and the spatial mapping of pathway activities based on known or denovo genesets.

We applied Renoir to three distinct spatial transcriptomics datasets ranging from development to disease, including adult mouse brain, human fetal liver, and human triple negative breast cancer (TNBC). For each dataset, Renoir identified distinct spatial communication domains which harbored distinct ligand-target activities pertaining to the major cell types abundant in the domain and these were also consistent with known literature. In the adult mouse brain, Renoir characterized the interactions of different regional astrocyte subtypes with other co-localized excitatory and inhibitory neurons. For the triple negative breast cancer, Renoir uncovered spatial communication domains and sub-domains harboring distinct chemokine interactions between cancer, stromal and immune cells which were associated with tumor proliferation and metastasis. Renoir further identified signaling pathways that were active in cancer (WNT signaling) or immune/stromal cell type (IL-6/JAK/STAT3 signaling) dominated domains of the TNBC tissue. For developing fetal liver, Renoir uncovered an interaction between hepatocytes and macrophages mediated via the ligand PLG.

By integrating single-cell RNA-seq and spatial transcriptomics data, Renoir enables a more comprehensive understanding of the dynamic relationships between different spatially co-localized cell types and their roles in various biological processes. This approach allows for the identification

of key players in cellular communication and provides insights into the mechanisms governing tissue organization, functionality, and response to disease. While in our analyses, we primarily used 10X Visium datasets, Renoir can be widely applied to datasets generated from any other spatial transcriptomic technologies. The several visualization modules provided by Renoir further enable the user to explore the spatial map of ligand-target and pathway activities of their interest. Ultimately, Renoir has the potential to advance our understanding of complex biological systems and contribute to the development of novel therapeutic strategies targeting cellular interactions.

# Methods

## Overview of Renoir

Using scRNA-seq data, its corresponding cell type annotations, spatial transcriptomics data, and a set of curated ligand-target pairs, Renoir calculates spatially enriched neighbourhood scores for the curated set of ligand-target pairs for a given sample. The neighbourhood scores can subsequently be leveraged to track ligand-target activity across the geographic topology, as well as identify spatially enriched communication regions, cell type-specific interactions, and pathway activity.

## Required input data

**Single-cell RNA-seq and Spatial transcriptomic data:** As input, Renoir requires (1) a spatial transcriptomics dataset $X^{st} \in \mathbb{R}^{S \times G}$ with gene expression for $S$ spots and $G$ genes, (2) spatial coordinates for each spot, and (3) a corresponding cell type annotated single-cell gene expression dataset $X^{sc} \in \mathbb{R}^{N_c \times G}$ for $N_c$ cells and $G$ genes, where each cell mapped to one of $C$ cell types.

**Acquiring curated ligand-target pairs:** We first curate a set of secreted ligands from NATMI's connectomeDB2020 (5) database. For each ligand, we obtain a list of putative targets sorted according to the descending order of ligand–target regulatory potential scores as calculated by NicheNet's prior knowledge model (21). The number of targets per ligand is user-defined. For our analyses, we have used the top 100 targets for each ligand. The ligand-target pair is retained for further analysis if both genes are among the highly variable genes in the spatial transcriptomic

dataset. Thus, we obtain a set of ligand-target pairs, $\mathscr{S}_{lt}$ specific to our spatial transcriptomic dataset.

### Calculation of neighborhood activity score for a ligand-target pair for a spot

**Constructing neighborhoods:** We utilize the spatial transcriptomic data to construct a weighted, directed graph such that each spot within the dataset represents a node in the graph and there exists a directed edge between a node (spot) and every node within its neighborhood. A neighborhood for a spot is defined with respect to the distribution of spots within the spatial dataset considered. For data wherein the spots are equidistant from their immediate neighbors (e.g., 10X Visium and Spatial Transcriptomics), the neighborhood for a spot $s$ is defined to be a set of spots inclusive of $s$ such that every spot other than $s$ is at a "one-hop" distance away from $s$.

**Formulating neighborhood activity scores:** On constructing the neighborhood graph, we quantify the potential interaction between a ligand $l$ and target $t$ within the neighborhood of a spot $s$ using a neighborhood activity score defined by,

$$\mathscr{A}_{l,t}(s) = ES_{ls,ts} + \Sigma_{j \in \mathscr{N}_s \setminus s}(ES_{lj,ts} + ES_{ls,tj}) \tag{1}$$

where $\mathscr{N}_s$ denotes the neighborhood of $s$ and $ES_{ls_1,ts_2}$ denotes the edge score (edge weight) of an edge directed from a spot $s_1$ expressing ligand $l$ to a neighboring spot $s_2$ expressing target $t$ and is defined by,

$$ES_{ls_1,ts_2} = norm(ECS_{ls_1,ts_2}) \times norm(ISM_{ls_1,ts_2})$$

where $norm(ECS_{ls_1,ts_2})$ and $norm(ISM_{ls_1,ts_2})$ are the min-max normalized expression correspondence score and inherent similarity measure for the ligand-target pair $l - t$ for the spots $s_1$ and $s_2$ respectively. The edge score quantifies the possible interaction between a ligand $l$ at spot $s_1$ and target $t$ at spot $s_2$ as the product of the inferred similarity between the expressions of $l$ and $t$ across all the cell types present at $s1$ and $s2$ respectively ($ECS_{ls_1,ts_2}$) and the inherent similarity (acquired from the single cell data) between the expressions of $l$ and $t$ amongst cell types present

at $s_1$ and $s_2$ respectively.

**Calculation of Expression Correspondence Score and Inherent Similarity Measure:** To calculate the Expression Correspondence Score and Inherent Similarity Measure between two neighboring spots $s_1$ and $s_2$ where a ligand $l$ is expressed at spot $s_1$ and target $t$ is expressed at spot $s_2$, we utilize the following formulations

$$ECS_{ls_1,ts_2} = \frac{\Sigma_{c_1 \in C_1} p_{c_1} PEM_{l,c_1,s_1} + \Sigma_{c_2 \in C_2} p_{c_2} PEM_{t,c_2,s_2}}{2} \tag{2}$$

$$ISM_{ls_1,ts_2} = \frac{\Sigma_{c_1 \in C_1} - p_{c_1} log_{10}(\frac{MI(lc_1,tc_1)}{min(H_{lc_1},H_{tc_1})}) + \Sigma_{c_2 \in C_2} - p_{c_2} log_{10}(\frac{MI(lc_2,tc_2)}{min(H_{lc_2},H_{tc_2})})}{2} \tag{3}$$

where, $C_1$ and $C_2$ denote the set of cell types present at $s_1$ and $s_2$ respectively, $p_{c_1}$ and $p_{c_2}$ denote the cell type proportions for cell types $c_1 \in C_1$ and $c_2 \in C_2$ respectively. $ECS_{ls_1,ts_2}$ is defined as the average of cell type proportion-weighted PEM () of ligand $l$ and target $t$ in spots $s_1$ and $s_2$ respectively. The PEM value for cell type specificity of gene $g$ in cell type $c$ for spot $s$ is given by

$$PEM_{g,c,s} = log_{10}(\frac{u_{scg}}{e_{scg}}), \tag{4}$$

where, $u_{scg}$ and $e_{scg}$ denote the inferred and expected expression of gene $g$ specific to cell type $c$ at spot $s$ respectively. In eq. (3), $MI(lc_1, tc_2)$ denotes the mutual information between the expression of $l$ across cell type $c_1$ and the expression of $t$ across cell type $c_2$. $H(lc_1)$ denotes the Shannon entropy of the expression of ligand $l$ across cell type $c_1$ and $H(tc_2)$ denotes the Shannon entropy of the expression of target $t$ across cell type $c_2$. $MI(lc_1, tc_2)$, $H(lc_1)$, and $H(tc_2)$ have been estimated from the scRNA-seq data using the Miller-Madow correction ().

**Quantifying inferred and expected cell type-specific gene expression:** In order to quantify the cell type-specific gene expressions at each spot, we estimate the abundance of all cell types in each spot ($W_{S \times C}$) and gene-specific scaling parameters ($M_{G \times 1}$) using Cell2location (24) that takes as input the ST data $X^{st}$ and reference cell type signatures ($\mathscr{G}^C : C \times G$) estimated from the single-cell data $X^{sc}$. We then compute the absolute number of mRNA molecules contributed by each cell type $c$ to each spot $s$ to quantify the gene expression of a specific gene $g$ at spot $s$ for cell

type $c$ as given by:

$$U_{S\times C\times G} = \{u_{scg}\} = W * \mathscr{G}^C * M \tag{5}$$

where $*$ represents matrix multiplication without sum-reduction. The expected expression of a specific gene $g$ at spot $s$ for cell type $c$ can then be calculated as:

$$E_{S\times C\times G} = \{e_{scg}\} = \Sigma_{n=1}^{C} u_{sng} \times \frac{\Sigma_{g'=1}^{G} u_{scg'}}{\Sigma_{n=1}^{C}\Sigma_{g'=1}^{G} u_{sng'}} \tag{6}$$

## Inference of spatial communication domains based on ligand-target neighborhood scores

In order to identify distinct spatial communication domains, which we define as a group of spots harboring similar ligand-target interactions, we performed clustering of the spots based on the ligand-target neighborhood activity scores. After computing the ligand-target neighborhood scores for all curated ligand-target pairs, each spot $s$ can be represented as a vector of ligand-target activity scores, $\mathscr{A}_{l,t}$ of ligand-target pairs $(l,t) \in \mathscr{S}_{lt}$ and based on these ligand-target activity score vectors, spots are clustered using the scanpy library (95). First, the top 2000 highly variable ligand-target pairs are identified followed by principal component analysis. Leiden clustering is performed on the k-nearest neighbor graph obtained based on the principal components.

**Inferring cell type-specific sub-communication domains:** Additionally, we further sub-clustered the spots present within each of the spatial communication domains based on the cell type abundance of the spots. For sub-clustering also, Leiden clustering was adopted and implemented using scanpy.

## Inference of communication domain-specific ligand-target activities

**Differential ligand-target activity across communication domains:** The Wilcoxon-rank sum test was utilized to estimate the ranked, differentially active ligand-target pairs for each communication domain using scanpy's 'rank_genes_groups()' function.

**Ranking of ligands based on activity within communication domains:** Renoir further allows the ranking of ligand activity within a communication domain. In order to perform ranking

of ligand activity, we first determine the major cell types (cell types that together constitute $\geq 90\%$ of cell type abundance) within a communication domain based on the average proportion of the cell types across all the spots within that domain. For each major cell type in a domain, we identify the top marker genes ($n = 100$) from the scRNA-seq dataset. We acquire a list of ligand-receptor pairs from NATMIs connectomeDB (5) and identify the major ligands that are also present within the marker gene lists of the major cell types within the domain. A cell type is said to express a receptor if the receptor gene is expressed in more than 5% of cells of that cell type (computed based on scRNA-seq data). For each ligand, we identify a set of marker targets across the major cell types within the domain. A target is selected if it falls within the marker gene list of the cell type and the cell type expresses a receptor for that ligand. For each ligand, we compute the average neighborhood activity score for each of its marker targets (averaged across the spots expressing the associated receptor cell type). Each ligand is then scored based on the Pearson correlation between the average neighborhood scores of the ligand-target pairs of the cell type marker targets and the regulatory potential scores for each ligand-target pair acquired from NicheNet's prior model (21). Using the ligand score, Renoir ranks the ligands, where the ranking represents the amount of agreement between the ligand-target activity in our dataset and the prior knowledge of the pairs. While in our analyses, we used ligand-receptor and ligand-target pairs from specific sources, Renoir can also work with user-defined ligand-receptor and ligand-target pairs.

**Inferring major cell type-specific ligand-target activity within a communication domain:** To identify cell type-specific ligand-target activity within each communication domain, we first considered the ligand-target pairs for which both the ligand and target fell within the top 100 markers for the major cell types of the communication domain. For each domain where the ligand and target were markers of the major cell types, a simple average of the Pearson correlations between the neighborhood activity scores of the ligand-target pair and the cell type proportions of the cell types expressing the ligand and target respectively were calculated across the spots pertaining to the domain being considered. The correlation score measures how prominent a cell type-specific ligand-target activity is within a communication domain.

## Inference of spatial mapping of pathway activity

**Construction of genesets of known pathways or Denovo clusters:** Renoir further allows the spatial mapping of a pathway activity by utilizing existing or denovo gene sets. First, for a particular pathway represented by a gene set, we determine the ligand-target pairs active in the dataset by intersecting the set of all possible gene pairs within a gene set with the set of curated ligand-target pairs. In this work, we restricted our usage to the GSEA Molecular Signatures Database (96; 97), specifically HALLMARK and CANONICAL PATHWAYS (KEGG, WIKIPATHWAYS) gene sets. Renoir can also utilize any predefined gene set provided by the user, specific to the user's requirements. Apart from utilizing gene sets for existing pathways, Renoir can also identify de novo clusters of ligand-target pairs based on their similarity of the neighborhood activity scores across the spots. For performing the clustering of ligand-target pairs, Renoir employs hierarchical clustering strategies, namely HDBSCAN and dynamic hclust. HDBSCAN was implemented using the hdbscan python package with the min_pts parameter set to the lowest possible value and dynamic hclust was implemented using the dynamicTreeCut package.

**Spatial mapping of existing pathways and Denovo clusters:** To generate a spatial map of previously known pathways or Denovo clusters, we computed the first principal component after performing dimensionality reduction over the neighborhood activity scores of the subset of ligand-target pairs corresponding to a pathway/Denovo cluster using the 'TruncatedSVD' function of sklearn. This generates a pathway activity score for each spot in the spatial transcriptomics data.

## Preparation of fetal liver data

FFPE fetal liver tissue was obtained under ethical approval from the Centralised Institutional Research Board of the Singapore Health Services, Singapore SingHealth and National Health Care Group Research Ethics Committees. To determine the quality of RNA in FFPE tissue, RNA was extracted from $10\mu m$ sections using a Qiagen FFPE RNA extraction kit (#73504) and following the manufacturer's protocol. RNA quality was determined as "good" if the percentage of RNA fragments >200 nucleotides (DV200%) was at least 50% of all fragments. Next, a $5\mu m$ tissue section was placed on spatially barcode oligonucleotides, within the capture area (6.5 x $6.5mm$) of the Visium slide (4 capture areas/slide) and processed according to the manufacturer's protocol.

Briefly, Visium slides with affixed FFPE tissue sections were deparaffinised in fresh xylene and rehydrated in reducing ethanol (in water) concentration gradient. The slides were stained with Haematoxylin and Alcoholic Eosin and imaged at 10 times optical magnification using a Nikon Ni-E Microscope. RNA was then de-crosslinked and tissue sections were hybridised with Visium Human Transcriptome Probe Set, overnight at 50°C. The following day, the slides were washed with supplied post-hybridisation wash buffer and hybridized probe pairs were ligated using supplied ligase buffer at 37°C. After ligation, the slides were washed with room temperature and pre-heated post-ligation wash buffer. The ligated probes were then released by digesting hybridized RNA with supplied RNase enzyme at 37°C. The tissue sections were then permeabilised at 37°C using the supplied permeabilising enzyme. Permeabilization enabled ligated probes to be captured by spatially barcoded oligonucleotides immobilised on the Visium slide. Captured hybridised probes were then extended using probe extension mix at 45°C and extended probes were eluted accordingly. To determine appropriate amplification cycle numbers for library preparation in the next step, $1\mu l$ of eluted probes were quantitively amplified (qPCR) in accordance with the manufacturer's instructions. Quantification cycle (Cq) value at 25% peak relative fluorescence unit (RFU) was used to set cycle numbers during library preparation. The library was then quantified using TapeStation (Agilent High Sensitivity D5000) to determine the average fragment size and concentration before sequencing.

### Integration of mouse brain datasets

The two mouse brain samples were integrated using the stlearn workflow (15). Neighborhood scores generated for each mouse brain sample were concatenated and min-max normalized across ligand-target pairs for every spot. The principal components calculated over the spatial transcriptomic data of each sample via scanpy were used as input to harmony algorithm (98), which projected the spots onto a shared embedding space and thus presenting a new set of adjusted principal components for the concatenated dataset. Leiden clustering was performed via scanpy over the concatenated dataset with adjusted principal components at a resolution of 1.3 to obtain a set of spatial communication domains shared across the two brain samples.

### Mapping of regional astrocyte subtype activity in mouse brain

In order to find localized astrocyte subtype activity in appropriate regions of the mouse brain dataset, we first identified the top 100 marker genes for every cell type (including astrocyte subtypes) using the available mouse brain snRNA-seq dataset. We then identified the top 100 differentially active ligand-target pairs for each region of the Mouse Brain dataset. Two sets of pairs were then generated, (i) the first set consisted of ligand-target pairs where the ligand was a marker gene of a specific astrocyte subtype, the ligand-target pair displayed differential activity in a region associated with said astrocyte subtype and the target was a marker of a cell type associated with said region (ii) the second set consisted of ligand-target pairs where the target was a marker gene of a specific astrocyte subtype, the ligand-target pair was differentially active in a region associated with said astrocyte subtype and the ligand was a marker of another cell type associated with said region. The above steps were carried out using the scanpy library 'rank_genes_group' function.

### Correlation measures used for analyzing *PLG:MARCO* interaction in fetal liver

The following correlation measures were utilized to measure the similarity between *PLG:MARCO* neighborhood scores with $PLG^+$ and $PLG^-$ hepatocytes and $FOLR2^+$ Macrophages for the Fetal Liver dataset,

- *Pearson correlation*: We utilized Pearson correlation to find the correlation amongst specific ligand/target expression values (either derived from scRNA-seq data or spatial transcriptomic data) and cell type abundances. We also used Pearson correlation to acquire an initial estimate of the correlation amongst the neighborhood scores of a specific ligand-target pair with the abundances of the cell types expressing the ligand and the target. The overall correlation measure can be summarised as,

$$r_{l,t,ct_l,ct_t} = \frac{P_{lt}(ct_l) + P_{lt}(ct_t)}{2}$$

  where,

$$P_{lt}(ct_x) = \frac{\Sigma_{i \in S_x}(NS_{lt}(i) - \mu_{NS_{lt}})(w_{i,ct_x} - \mu_{w_{ct_x}})}{\sqrt{\Sigma_{i \in S_x}(NS_{lt}(i) - \mu_{NS_{lt}})^2 \Sigma_{i \in S_x}(w_{i,ct_x} - \mu_{w_{ct_x}})^2}}$$

$l$ and $t$ represent a specific ligand and target respectively

$ct_l$ and $ct_t$ represent the cell types expressing the ligand and the target respectively (as a marker)

$NS_{lt}(s)$ represents the neighborhood scores of ligand-target pair $lt$ at spot $s$

$NS_{lt}$ represents a vector of neighborhood scores of ligand-target pair $lt$ across all $S_x$ spots

$w_{s,y}$ represents the cell type abundance of the cell type the ligand($ct_l$) or target($ct_t$) is a marker of at spot $s$

$w_{ct_x}$ is a vector of cell type abundances of the cell type the ligand($ct_l$) or target($ct_t$) is a marker of, across all $S_x$ spots

$\mu_{NS_{lt}}$ is the average neighborhood score for a specific ligand-target pair $lt$

$\mu_{w_{ct_x}}$ is the average cell type abundance for a specific cell type $ct_x$

$S_x = \{s_j, (NS_{lt}(s_j) > 0)|(w_{s_j,ct_x} > 0)\}$ where $x$ is either a ligand $l$ or target $t$.

- *Pearson correlation over spatial lag*: We furher used Pearson correlation over spatial lag as a measure of spatial correlation (99). Pearson correlation over spatial lag considers the sum of expression or abundance values over a spot and its corresponding neighbors before applying Pearson correlation. Formally, Pearson correlation in the context of expression or abundance values can be defined as,

$$r_{xy} = \frac{\Sigma_{i \in S}(X_i - \mu_X)(Y_i - \mu_Y)}{\sqrt{\Sigma_{i \in S}(X_i - \mu_X)^2 \Sigma_{i \in S}(Y_i - \mu_Y)^2}}$$

$X$ and $Y$ are the spatial lags defined by,

$$X_i = \Sigma_{n \in N(x_i)} x_n$$

where $N(x_i)$ represents the neighborhood of $x_i$. $X = x_i, i \in S$; $S_x = \{s_j, (X_{s_j} > 0)|(Y_{s_j} > 0)\}$

The spatial lags are calculated with respect to gene expressions from the spatial transcriptomics data or the spatial map of cell type abundances ($x$ or $y$). While considering neighborhood scores, the scores are considered as they are since the neighborhood score at each spot takes into account the neighbors of the spot.

- *Lee's L statistic*: We utilised Lee's L statistic bivariate spatial association measure to incorporate spatial autocorrelation alongside spatial patterning between the two variables of interest (99). In our case, we adopted the measure to view the spatial co-patterning of neighborhood scores, gene expressions from the spatial transcriptomics data or the spatial map of cell type abundances with one another. This is defined as,

$$L_{X,Y} = \sqrt{L_{X,X} L_{Y,Y}} r_{xy}$$

where,

$$L_{X,X} = \frac{\Sigma_{i \in S}(X_i - \mu_x)^2}{\Sigma_{i \in S}(x_i - \mu_x)^2}$$

## Data availability

The original public data that was utilised in this study are provided below: snRNA-seq from adjacent sections in the mouse brain: downloaded from ArrayExpress under accession number E-MTAB-11115; Visium from adjacent sections in the mouse brain: downloaded from ArrayExpress under accession number E-MTAB-11114 (sample S1 = ST8059048 and sample S2 = ST8059052); Triple Negative Breast Cancer spatially resolved transcriptomics data: downloaded from the Zenodo data repository (https://doi.org/10.5281/zenodo.4739739); Triple Negative Breast Cancer raw scRNA-seq data: Downloaded from the Gene Expression Omnibus under accession number GSE176078. The scRNA-seq data for fetal liver was obtained from the Gene Expression Omnibus under accession number GSE156337. 10X Visium data for the fetal liver generated for this manuscript is available here.

### Code availability

Renoir has been implemented in Python and is freely available at https://github.com/Zafar-Lab/Renoir under the MIT license.

## Acknowledgements

This work was funded by the DBT/Wellcome Trust India Alliance Early Career Fellowship [grant IA/E/21/1/506298], IIT Kanpur initiation grant [IITK /CS /2019236], and Science and Engineering Research Board (SERB), Government of India Startup Research Grant [SRG/2020/001333] to H.Z. A.S. Laboratory is supported by funding from NHMRC Ideas grant [2021/GNT2010795], MRFF EMCR grant [2022/MRF2016215], Perkins-Curtin start-up fellowship and Cancer Research Trust (CRT) Programme Grant to Liver Cancer Collaborative.

## Author Contributions

HZ and AS designed the study. HZ and NR developed the model and algorithm. NR implemented the software and performed all experiments. RP, AM, FG and JC generated the fetal liver dataset. HZ, AS, NR, and RP wrote the manuscript and all authors approved the manuscript.

## Competing Interests

The authors declare no competing interests.

## References

[1] Jin, S. *et al.* Inference and analysis of cell-cell communication using cellchat. *Nature Communications* **12**, 1088 (2021). URL `https://doi.org/10.1038/s41467-021-21246-9`.

[2] Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. Cellphonedb: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nature Protocols* **15**, 1484–1506 (2020). URL `https://doi.org/10.1038/s41596-020-0292-x`.

[3] Noël, F. *et al.* Dissection of intercellular communication using the transcriptome-based framework icellnet. *Nature Communications* **12**, 1089 (2021). URL `https://doi.org/10.1038/s41467-021-21244-x`.

[4] Wang, Y. *et al.* italk: an r package to characterize and illustrate intercellular communication (2019). URL `https://doi.org/10.1101/507871`.

[5] Hou, R., Denisenko, E., Ong, H. T., Ramilowski, J. A. & Forrest, A. R. R. Predicting cell-to-cell communication networks using natmi. *Nature Communications* **11**, 5011 (2020). URL https://doi.org/10.1038/s41467-020-18873-z.

[6] Almet, A. A., Cang, Z., Jin, S. & Nie, Q. The landscape of cell–cell communication through single-cell transcriptomics. *Current Opinion in Systems Biology* **26**, 12–23 (2021). URL https://www.sciencedirect.com/science/article/pii/S2452310021000081.

[7] Cabello-Aguilar, S. *et al.* SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Research* **48**, e55–e55 (2020). URL https://doi.org/10.1093/nar/gkaa183. https://academic.oup.com/nar/article-pdf/48/10/e55/33326200/gkaa183.pdf.

[8] Solovey, M. & Scialdone, A. COMUNET: a tool to explore and visualize intercellular communication. *Bioinformatics* **36**, 4296–4300 (2020). URL https://doi.org/10.1093/bioinformatics/btaa482. https://academic.oup.com/bioinformatics/article-pdf/36/15/4296/33796656/btaa482.pdf.

[9] Tyler, S. R. *et al.* Pyminer finds gene and autocrine-paracrine networks from human islet scrna-seq. *Cell Reports* **26**, 1951–1964.e8 (2019). URL https://www.sciencedirect.com/science/article/pii/S2211124719300920.

[10] Cang, Z. & Nie, Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nature Communications* **11**, 2084 (2020). URL https://doi.org/10.1038/s41467-020-15968-5.

[11] Garcia-Alonso, L. *et al.* Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. *Nature Genetics* **53**, 1698–1711 (2021). URL https://doi.org/10.1038/s41588-021-00972-2.

[12] Armingol, E. *et al.* Context-aware deconvolution of cell–cell communication with tensor-cell2cell. *Nature Communications* **13**, 3665 (2022). URL https://doi.org/10.1038/s41467-022-31369-2.

[13] Dries, R. *et al.* Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biology* **22**, 78 (2021). URL https://doi.org/10.1186/s13059-021-02286-2.

[14] Tanevski, J., Flores, R. O. R., Gabor, A., Schapiro, D. & Saez-Rodriguez, J. Explainable multiview framework for dissecting spatial relationships from highly multiplexed data. *Genome Biology* **23**, 97 (2022). URL https://doi.org/10.1186/s13059-022-02663-5.

[15] Pham, D. *et al.* stlearn: integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. *bioRxiv* (2020). URL https://www.biorxiv.org/content/early/2020/05/31/2020.05.31.125658. https://www.biorxiv.org/content/early/2020/05/31/2020.05.31.125658.full.pdf.

[16] Arnol, D., Schapiro, D., Bodenmiller, B., Saez-Rodriguez, J. & Stegle, O. Modeling cell-cell interactions from spatial molecular data with spatial variance component analysis. *Cell Reports* **29**, 202–211.e6 (2019). URL https://www.sciencedirect.com/science/article/pii/S2211124719311325.

[17] Raredon, M. S. B. *et al.* Comprehensive visualization of cell–cell interactions in single-cell and spatial transcriptomics with NICHES. *Bioinformatics* **39** (2022). URL https://doi.org/10.1093/bioinformatics/btac775. Btac775, https://academic.oup.com/bioinformatics/article-pdf/39/1/btac775/48520748/btac775_supplementary_data.pdf.

[18] Cang, Z. *et al.* Screening cell–cell communication in spatial transcriptomics via collective optimal transport. *Nature Methods* **20**, 218–228 (2023). URL https://doi.org/10.1038/s41592-022-01728-4.

[19] Dimitrov, D. *et al.* Comparison of methods and resources for cell-cell communication inference from single-cell rna-seq data. *Nature Communications* **13**, 3224 (2022). URL https://doi.org/10.1038/s41467-022-30755-0.

[20] Cheng, J., Zhang, J., Wu, Z. & Sun, X. Inferring microenvironmental regulation of gene expression from single-cell RNA sequencing data using scMLnet with an application to COVID-19. *Briefings in Bioinformatics* **22**, 988–1005 (2020). URL https://doi.org/10.1093/bib/bbaa327. https://academic.oup.com/bib/article-pdf/22/2/988/36653976/supplementary_information_bbaa327.pdf.

[21] Browaeys, R., Saelens, W. & Saeys, Y. Nichenet: modeling intercellular communication by linking ligands to target genes. *Nature Methods* **17**, 159–162 (2020). URL https://doi.org/10.1038/s41592-019-0667-5.

[22] Wang, S., Karikomi, M., MacLean, A. L. & Nie, Q. Cell lineage and communication network inference via optimization for single-cell transcriptomics. *Nucleic Acids Research* **47**, e66–e66 (2019). URL https://doi.org/10.1093/nar/gkz204. https://academic.oup.com/nar/article-pdf/47/11/e66/28837990/gkz204_supplemental_file.pdf.

[23] Choi, H. *et al.* Transcriptome analysis of individual stromal cell populations identifies stroma-tumor crosstalk in mouse lung cancer model. *Cell Reports* **10**, 1187–1201 (2015). URL https://www.sciencedirect.com/science/article/pii/S2211124715000650.

[24] Kleshchevnikov, V. *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics. *Nature Biotechnology* **40**, 661–671 (2022). URL https://doi.org/10.1038/s41587-021-01139-4.

[25] Easley-Neal, C., Foreman, O., Sharma, N., Zarrin, A. A. & Weimer, R. M. Csf1r ligands il-34 and csf1 are differentially required for microglia development and maintenance in white and gray matter brain regions. *Frontiers in Immunology* **10** (2019). URL https://www.frontiersin.org/articles/10.3389/fimmu.2019.02199.

[26] Fairlie-Clarke, K. *et al.* Expression and function of il-33/st2 axis in the central nervous system under normal and diseased conditions. *Frontiers in Immunology* **9** (2018). URL https://www.frontiersin.org/articles/10.3389/fimmu.2018.02596.

[27] Villanueva, C., Jacquier, S. & de Roux, N. Dlk1 is a somato-dendritic protein expressed in hypothalamic arginine-vasopressin and oxytocin neurons. *PLOS ONE* **7**, 1–9 (2012). URL https://doi.org/10.1371/journal.pone.0036134.

[28] Chodelkova, O., Masek, J., Korinek, V., Kozmik, Z. & Machon, O. Tcf7l2 is essential for neurogenesis in the developing mouse neocortex. *Neural Development* **13**, 8 (2018). URL https://doi.org/10.1186/s13064-018-0107-8.

[29] Nagalski, A. *et al.* Postnatal isoform switch and protein localization of lef1 and tcf7l2 transcription factors in cortical, thalamic, and mesencephalic regions of the adult mouse brain. *Brain Structure and Function* **218**, 1531–1549 (2013). URL https://doi.org/10.1007/s00429-012-0474-6.

[30] Li, J.-Y. *et al.* LGR4 and Its Ligands, R-Spondin 1 and R-Spondin 3, Regulate Food Intake in the Hypothalamus of Male Rats. *Endocrinology* **155**, 429–440 (2014). URL https://doi.org/10.1210/en.2013-1550. https://academic.oup.com/endo/article-pdf/155/2/429/10824578/endo0429.pdf.

[31] Miller, S. J. *et al.* Molecularly defined cortical astroglia subpopulation modulates neurons via secretion of norrin. *Nature Neuroscience* **22**, 741–752 (2019). URL https://doi.org/10.1038/s41593-019-0366-7.

[32] Koeller, H. B., Ross, M. E. & Glickstein, S. B. Cyclin d1 in excitatory neurons of the adult brain enhances kainate-induced neurotoxicity. *Neurobiology of Disease* **31**, 230–241 (2008). URL https://www.sciencedirect.com/science/article/pii/S0969996108000867.

[33] Tang, K., Rubenstein, J. L. R., Tsai, S. Y. & Tsai, M.-J. COUP-TFII controls amygdala patterning by regulating neuropilin expression. *Development* **139**, 1630–1639 (2012). URL https://doi.org/10.1242/dev.075564. https://journals.biologists.com/dev/article-pdf/139/9/1630/1175064/1630.pdf.

[34] Wilcox, J. & Derynck, R. Localization of cells synthesizing transforming growth factor-alpha mrna in the mouse brain. *Journal of Neuroscience* **8**, 1901–1904 (1988). URL https://www.jneurosci.org/content/8/6/1901. https://www.jneurosci.org/content/8/6/1901.full.pdf.

[35] Vincze, C. *et al.* Distribution of mrnas encoding transforming growth factors-$\beta$1, -2, and -3 in the intact rat brain and after experimentally induced focal ischemia. *Journal of Comparative Neurology* **518**, 3752–3770 (2010). URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cne.22422. https://onlinelibrary.wiley.com/doi/pdf/10.1002/cne.22422.

[36] Dobolyi, A., Vincze, C., Pál, G. & Lovas, G. The neuroprotective functions of transforming growth factor beta proteins.

*International Journal of Molecular Sciences* **13**, 8219–8258 (2012). URL https://www.mdpi.com/1422-0067/13/7/8219.

[37] Mei, L. & Xiong, W.-C. Neuregulin 1 in neural development, synaptic plasticity and schizophrenia. *Nature Reviews Neuroscience* **9**, 437–452 (2008). URL https://doi.org/10.1038/nrn2392.

[38] Di Benedetto, B. *et al.* Differential mrna distribution of components of the erk/mapk signalling cascade in the adult mouse brain. *Journal of Comparative Neurology* **500**, 542–556 (2007). URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cne.21186. https://onlinelibrary.wiley.com/doi/pdf/10.1002/cne.21186.

[39] Vilar, M. & Mira, H. Regulation of neurogenesis by neurotrophins during adulthood: Expected and unexpected roles. *Frontiers in Neuroscience* **10** (2016). URL https://www.frontiersin.org/articles/10.3389/fnins.2016.00026.

[40] Sharma, A. *et al.* Onco-fetal reprogramming of endothelial cells drives immunosuppressive macrophages in hepatocellular carcinoma. *Cell* **183**, 377–394.e21 (2020). URL https://www.sciencedirect.com/science/article/pii/S0092867420310825.

[41] Duncan, M. B. *et al.* Type XVIII collagen is essential for survival during acute liver injury in mice. *Disease Models & Mechanisms* **6**, 942–951 (2013). URL https://doi.org/10.1242/dmm.011577. https://journals.biologists.com/dmm/article-pdf/6/4/942/1865579/942.pdf.

[42] Guo, Y. *et al.* Relationships between hematopoiesis and hepatogenesis in the midtrimester fetal liver characterized by dynamic transcriptomic and proteomic profiles. *PLOS ONE* **4**, 1–10 (2009). URL https://doi.org/10.1371/journal.pone.0007641.

[43] Gao, S. *et al.* Identification of hsc/mpp expansion units in fetal liver by single-cell spatiotemporal transcriptomics. *Cell Research* **32**, 38–53 (2022). URL https://doi.org/10.1038/s41422-021-00540-7.

[44] Lewis, K., Yoshimoto, M. & Takebe, T. Fetal liver hematopoiesis: from development to delivery. *Stem Cell Research & Therapy* **12**, 139 (2021). URL https://doi.org/10.1186/s13287-021-02189-w.

[45] Bianchi, M. E. & Mezzapelle, R. The chemokine receptor cxcr4 in cell proliferation and tissue regeneration. *Frontiers in Immunology* **11** (2020). URL https://www.frontiersin.org/articles/10.3389/fimmu.2020.02109.

[46] Aiuti, A. *et al.* Expression of cxcr4, the receptor for stromal cell-derived factor-1 on fetal and adult human lymphohematopoietic progenitors. *European Journal of Immunology* **29**, 1823–1831 (1999). URL https://onlinelibrary.wiley.com/doi/abs/10.1002/. https://onlinelibrary.wiley.com/doi/pdf/10.1002/.

[47] MacParland, S. A. *et al.* Single cell rna sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nature Communications* **9**, 4383 (2018). URL https://doi.org/10.1038/s41467-018-06318-7.

[48] Bezerra, J. A. *et al.* Plasminogen deficiency leads to impaired remodeling after a toxic injury to the liver. *Proceedings of the National Academy of Sciences* **96**, 15143–15148 (1999). URL https://www.pnas.org/doi/abs/10.1073/pnas.96.26.15143. https://www.pnas.org/doi/pdf/10.1073/pnas.96.26.15143.

[49] Okada, K. *et al.* Binding of plasminogen to hepatocytes isolated from injured mice liver and nonparenchymal cell-dependent proliferation of hepatocytes. *Blood Coagulation & Fibrinolysis* **19** (2008). URL https://journals.lww.com/bloodcoagulation/Fulltext/2008/09000/Binding_of_plasminogen_to_hepatocytes_isolated.6.aspx.

[50] Damnernsawad, A. *et al.* Kras is Required for Adult Hematopoiesis. *Stem Cells* **34**, 1859–1871 (2016). URL https://doi.org/10.1002/stem.2355. https://academic.oup.com/stmcls/article-pdf/34/7/1859/42026704/stmcls_34_7_1859.pdf.

[51] Kiu, H. & Nicholson, S. E. Biology and significance of the jak/stat signalling pathways. *Growth Factors* **30**, 88–106 (2012). URL https://doi.org/10.3109/08977194.2012.660936. PMID: 22339650, https://doi.org/10.3109/08977194.2012.660936.

[52] Wu, S. Z. *et al.* A single-cell and spatially resolved atlas of human breast cancers. *Nature Genetics* **53**, 1334–1347 (2021). URL https://doi.org/10.1038/s41588-021-00911-1.

[53] Sørlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.

*Proceedings of the National Academy of Sciences* **98**, 10869–10874 (2001). URL https://www.pnas.org/doi/abs/10.1073/pnas.191367098. https://www.pnas.org/doi/pdf/10.1073/pnas.191367098.

[54] Sotiriou, C. *et al.* Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proceedings of the National Academy of Sciences* **100**, 10393–10398 (2003). URL https://www.pnas.org/doi/abs/10.1073/pnas.1732912100. https://www.pnas.org/doi/pdf/10.1073/pnas.1732912100.

[55] Ignacio, R. M. C. *et al.* Serum amyloid a predisposes inflammatory tumor microenvironment in triple negative breast cancer. *Oncotarget* **10**, 511–526 (2019). URL https://www.oncotarget.com/article/26566/.

[56] Filippou, P. S., Karagiannis, G. S. & Constantinidou, A. Midkine (mdk) growth factor: a key player in cancer progression and a promising therapeutic target. *Oncogene* **39**, 2040–2054 (2020). URL https://doi.org/10.1038/s41388-019-1124-8.

[57] Ribatti, D. *et al.* Angiogenesis and antiangiogenesis in triple-negative breast cancer. *Translational Oncology* **9**, 453–457 (2016). URL https://www.sciencedirect.com/science/article/pii/S1936523316300717.

[58] Harmey, J. H., Dimitriadis, E., Kay, E., Redmond, H. P. & Bouchier-Hayes, D. Regulation of macrophage production of vascular endothelial growth factor (vegf) by hypoxia and transforming growth factor $\beta$-1. *Annals of Surgical Oncology* **5**, 271–278 (1998). URL https://doi.org/10.1007/BF02303785.

[59] Lewis, J., Landers, R., Underwood, J., Harris, A. & Lewis, C. Expression of vascular endothelial growth factor by macrophages is up-regulated in poorly vascularized areas of breast carcinomas. *The Journal of Pathology* **192**, 150–158 (2000).

[60] Guo, P. *et al.* Icam-1 as a molecular target for triple negative breast cancer. *Proceedings of the National Academy of Sciences* **111**, 14710–14715 (2014). URL https://www.pnas.org/doi/abs/10.1073/pnas.1408556111. https://www.pnas.org/doi/pdf/10.1073/pnas.1408556111.

[61] McGowan, P. & Duffy, M. Matrix metalloproteinase expression and outcome in patients with breast cancer: analysis of a published database. *Annals of Oncology* **19**, 1566–1572 (2008). URL https://www.sciencedirect.com/science/article/pii/S0923753419402160.

[62] House, I. G. *et al.* Macrophage-Derived CXCL9 and CXCL10 Are Required for Antitumor Immune Responses Following Immune Checkpoint Blockade. *Clinical Cancer Research* **26**, 487–504 (2020). URL https://doi.org/10.1158/1078-0432.CCR-19-1868. https://aacrjournals.org/clincancerres/article-pdf/26/2/487/2062480/487.pdf.

[63] Yang, L., Xiao, Liu, Y., Kai, Lin, J., Feng, Shi, M., Hui & Ou, L., Zhou. Ccl28 promotes breast cancer growth and metastasis through mapk-mediated cellular anti-apoptosis and pro-metastasis. *Oncol Rep* **38**, 1393–1401 (2017). URL https://doi.org/10.3892/or.2017.5798.

[64] Li, X., Cao, Y., Li, M. & Jin, F. Upregulation of hes1 promotes cell proliferation and invasion in breast cancer as a prognosis marker and therapy target via the akt pathway and emt process. *Journal of Cancer* **9**, 757–766 (2018).

[65] Mao, S. P. H. *et al.* Loss of amphiregulin reduces myoepithelial cell coverage of mammary ducts and alters breast tumor growth. *Breast Cancer Research* **20**, 131 (2018). URL https://doi.org/10.1186/s13058-018-1057-0.

[66] Ara, T. & DeClerck, Y. A. Interleukin-6 in bone metastasis and cancer progression. *European Journal of Cancer* **46**, 1223–1231 (2010). URL https://www.sciencedirect.com/science/article/pii/S095980491000153X. Stopping Cancer in its Tracks: Metastasis as a Therapeutic Target.

[67] Liu, J., Shen, J.-X., Hu, J.-L., Huang, W.-H. & Zhang, G.-J. Significance of interleukin-33 and its related cytokines in patients with breast cancers. *Frontiers in Immunology* **5** (2014). URL https://www.frontiersin.org/articles/10.3389/fimmu.2014.00141.

[68] Shiiba, M. *et al.* Interleukin-1 receptor antagonist (il1rn) is associated with suppression of early carcinogenic events in human oral malignancies. *Int J Oncol* **46**, 1978–1984 (2015). URL https://doi.org/10.3892/ijo.2015.2917.

[69] Xiao, G. *et al.* Cxcl16/cxcr6 chemokine signaling mediates breast cancer progression by perk1/2-dependent mechanisms.

*Oncotarget* **6**, 14165–14178 (2015). URL https://www.oncotarget.com/article/3690/.

[70] Mir, H. *et al.* Cxcr6-cxcl16 axis promotes breast cancer by inducing oncogenic signaling. *Cancers* **13** (2021). URL https://www.mdpi.com/2072-6694/13/14/3568.

[71] Li, Y.-Q. *et al.* Tumor secretion of ccl22 activates intratumoral treg infiltration and is independent prognostic predictor of breast cancer. *PLOS ONE* **8**, null (2013). URL https://doi.org/10.1371/journal.pone.0076379.

[72] Ahrends, T. & Borst, J. The opposing roles of cd4+ t cells in anti-tumour immunity. *Immunology* **154**, 582–592 (2018). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/imm.12941. https://onlinelibrary.wiley.com/doi/pdf/10.1111/imm.12941.

[73] Moustakas, A. & Heldin, C.-H. Mechanisms of tgf$\beta$-induced epithelial–mesenchymal transition. *Journal of Clinical Medicine* **5** (2016). URL https://www.mdpi.com/2077-0383/5/7/63.

[74] Wang, Y., Xu, H., Zhu, B., Qiu, Z. & Lin, Z. Systematic identification of the key candidate genes in breast cancer stroma. *Cellular & Molecular Biology Letters* **23**, 44 (2018). URL https://doi.org/10.1186/s11658-018-0110-4.

[75] Oskarsson, T. *et al.* Breast cancer cells produce tenascin c as a metastatic niche component to colonize the lungs. *Nature Medicine* **17**, 867–874 (2011). URL https://doi.org/10.1038/nm.2379.

[76] Kitamura, T. *et al.* CCL2-induced chemokine cascade promotes breast cancer metastasis by enhancing retention of metastasis-associated macrophages. *Journal of Experimental Medicine* **212**, 1043–1059 (2015). URL https://doi.org/10.1084/jem.20141836. https://rupress.org/jem/article-pdf/212/7/1043/1162569/jem_20141836.pdf.

[77] Zhang, Y. *et al.* A novel role of hematopoietic ccl5 in promoting triple-negative mammary tumor progression by regulating generation of myeloid-derived suppressor cells. *Cell Research* **23**, 394–408 (2013). URL https://doi.org/10.1038/cr.2012.178.

[78] Lehmann, B. D. *et al.* Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *The Journal of Clinical Investigation* **121**, 2750–2767 (2011). URL https://www.jci.org/articles/view/45014.

[79] Sabatier, L. *et al.* Heparin/heparan sulfate controls fibrillin-1, -2 and -3 self-interactions in microfibril assembly. *FEBS Letters* **588**, 2890–2897 (2014). URL https://www.sciencedirect.com/science/article/pii/S0014579314005481.

[80] Summers, K. M., Raza, S., Van Nimwegen, E., Freeman, T. C. & Hume, D. A. Co-expression of fbn1 with mesenchyme-specific genes in mouse cell lines: implications for phenotypic variability in marfan syndrome. *European journal of human genetics* **18**, 1209–1215 (2010).

[81] Tuglu, M., Matilda, Bostanabad, Y., Saber, Ozyon, G., Dalkiliç, B. & Gurdal, H. The role of dual-specificity phosphatase 1 and protein phosphatase 1 in $\beta$2-adrenergic receptor-mediated inhibition of extracellular signal regulated kinase 1/2 in triple negative breast cancer cell lines. *Mol Med Rep* **17**, 2033–2043 (2018). URL https://doi.org/10.3892/mmr.2017.8092.

[82] Lukey, M. J., Greene, K. S., Erickson, J. W., Wilson, K. F. & Cerione, R. A. The oncogenic transcription factor c-jun regulates glutaminase expression and sensitizes cells to glutaminase-targeted therapy. *Nature Communications* **7**, 11321 (2016). URL https://doi.org/10.1038/ncomms11321.

[83] Wittamer, V. *et al.* Specific recruitment of antigen-presenting cells by chemerin, a novel processed ligand from human inflammatory fluids. *The Journal of experimental medicine* **198**, 977–985 (2003).

[84] Wittamer, V. *et al.* Neutrophil-mediated maturation of chemerin: a link between innate and adaptive immunity. *The Journal of immunology* **175**, 487–493 (2005).

[85] Zabel, B. A., Silverio, A. M. & Butcher, E. C. Chemokine-like receptor 1 expression and chemerin-directed chemotaxis distinguish plasmacytoid from myeloid dendritic cells in human blood. *The Journal of Immunology* **174**, 244–251 (2005).

[86] Vermi, W. *et al.* Role of chemr23 in directing the migration of myeloid and plasmacytoid dendritic cells to lymphoid organs and inflamed skin. *The Journal of experimental medicine* **201**, 509–515 (2005).

[87] Liu, H. *et al.* The role of mmp-1 in breast cancer growth and metastasis to the brain in a xenograft model. *BMC Cancer* **12**, 583 (2012). URL https://doi.org/10.1186/1471-2407-12-583.

[88] Giricz, O., Calvo, V., Peterson, E. A., Abouzeid, C. M. & Kenny, P. A. Tace-dependent tgf$\alpha$ shedding drives triple-negative breast cancer cell invasion. *International Journal of Cancer* **133**, 2587–2595 (2013). URL https://onlinelibrary.wiley.com/doi/abs/10.1002/ijc.28295. https://onlinelibrary.wiley.com/doi/pdf/10.1002/ijc.28295.

[89] Zhu, G. *et al.* Expression and role of dickkopf-1 (dkk1) in tumors: from the cells to the patients. *Cancer Management and Research* **13**, 659 (2021).

[90] Kieffer, Y. *et al.* Single-Cell Analysis Reveals Fibroblast Clusters Linked to Immunotherapy Resistance in Cancer. *Cancer Discovery* **10**, 1330–1351 (2020). URL https://doi.org/10.1158/2159-8290.CD-19-1384. https://aacrjournals.org/cancerdiscovery/article-pdf/10/9/1330/1818529/1330.pdf.

[91] Anzai, E. *et al.* Foxa1 induces e-cadherin expression at the protein level via suppression of slug in epithelial breast cancer cells. *Biological and Pharmaceutical Bulletin* **40**, 1483–1489 (2017).

[92] Vicari, A. P. *et al.* Antitumor effects of the mouse chemokine 6ckine/slc through angiostatic and immunological mechanisms. *The Journal of Immunology* **165**, 1992–2000 (2000).

[93] Crawford, N. P. S. *et al.* ¡i¿bromodomain 4¡/i¿ activation predicts breast cancer survival. *Proceedings of the National Academy of Sciences* **105**, 6380–6385 (2008). URL https://www.pnas.org/doi/abs/10.1073/pnas.0710331105. https://www.pnas.org/doi/pdf/10.1073/pnas.0710331105.

[94] Johnson, D. E., O'Keefe, R. A. & Grandis, J. R. Targeting the il-6/jak/stat3 signalling axis in cancer. *Nature Reviews Clinical Oncology* **15**, 234–248 (2018). URL https://doi.org/10.1038/nrclinonc.2018.8.

[95] Wolf, F. A., Angerer, P. & Theis, F. J. Scanpy: large-scale single-cell gene expression data analysis. *Genome Biology* **19**, 15 (2018). URL https://doi.org/10.1186/s13059-017-1382-0.

[96] Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545–15550 (2005). URL https://www.pnas.org/doi/abs/10.1073/pnas.0506580102. https://www.pnas.org/doi/pdf/10.1073/pnas.0506580102.

[97] Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011). URL https://doi.org/10.1093/bioinformatics/btr260. https://academic.oup.com/bioinformatics/article-pdf/27/12/1739/48864724/bioinformatics_27_12_1739.pdf.

[98] Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with harmony. *Nature Methods* **16**, 1289–1296 (2019). URL https://doi.org/10.1038/s41592-019-0619-0.

[99] Lee, S.-I. Developing a bivariate spatial association measure: An integration of pearson's r and moran's i. *Journal of Geographical Systems* **3**, 369–385 (2001). URL https://doi.org/10.1007/s101090100064.

## Figure Legends

Figure 1: **Overview of Renoir.** (a) Renoir utilizes spatial transcriptomic data and its corresponding cell type-annotated scRNA-seq data to infer spatial neighborhood activity scores for a set of curated ligand-target pairs at each spot. (b) To estimate the neighborhood activity score for a ligand-target pair $l - t$ at a given spot $S1$; Renoir utilizes cell type abundance, cell type-specific mRNA abundance for each spot as inferred by a cell type deconvolution module, gene entropy and mutual information between l-t specific to cell types. (c) Renoir performs various downstream tasks based on the ligand-target neighborhood scores - inference of spatial communication domains and domain-specific ligand-target activities, spatial mapping of pathway activity and ranking of ligand activity.

Figure 2: **Renoir identifies distinct communication domains in mouse brain.** (a) Spatial communication domains with distinct latent embeddings inferred by Renoir based on ligand-target neighborhood scores for two mouse brain sections. (b) Distribution of spots in spatial communication domains across different anatomical regions of the mouse brain. The rows correspond to spatial communication domains (in both S1 and S2) and columns correspond to known brain regions. Each cell represents the proportion of spots of a communication domain in the corresponding anatomical brain region. (c) Differentially active ligand-target interactions across the communication domains for sample S1. (d) - (e) Characterization of ligand-target interactions between regional astrocyte subtypes and other co-localized cell types in sample S1. (d) Ligands expressed by regional astrocyte subtypes activating target genes in other cell types. (e) Target genes in regional astrocyte subtypes activated by ligands expressed by other co-localized cell types. (f) Communication domain-specific ranking of ligands based on cumulative activities over target genes expressed by major cell types in the domain for sample S1. Top three ligands for each of eight communication domains are represented. Stacked color bars represent the cell types that express the ligand (right) and target (top). (g) Spatial map of MAPK SIGNALING and NEUROTROPHIN SIGNALING pathway activity across the spots in sample S1.

Figure 3: **Renoir identifies hepatocyte-macrophage interactions in developing fetal liver**. (a) Preparation of FFPE fetal liver tissue (b) Spatial communication domains inferred by Renoir based on ligand-target neighborhood scores. (c) Distribution of UMI count across the tissue section. (d) Spatial map of neighborhood activity scores for the ligand-target pairs (e) Ranking of ligands based on their cumulative activities over target genes expressed by major cell types in domain 0. (f) UMAP plot of latent embedding for hepatocyte population in scRNA-seq data, cells are colored as either $PLG^+$ or $PLG^-$. (g) Cell type abundances of Hepatocytes ($PLG^+$ and $PLG^-$) and $FOLR2^+$ Macrophages overlaid onto the spots in ST data. (h) Volcano plot depicting differentially expressed genes between $PLG^+$ and $PLG^-$ Hepatocytes ($-log_{10}P$ threshold = 32 and $log_2 foldchange$ threshold = 0.5). (i) Spatial similarity measures between *PLG:MARCO* neighborhood activity scores and cell type abundances of $PLG^+$, $PLG^-$ Hepatocytes and $FOLR2^+$ Macrophages.

Figure 4: **Renoir uncovers intratumor spatial communication sub-domains in triple negative breast cancer tissue**. (a) The original histopathological annotations of H&E-stained triple negative breast cancer tissue (CID44971 (52), $n = 1322$ spots) showing six pathologically distinct regions. (b) Spatial communication domains inferred by Renoir based on ligand-target neighbor activity scores. The communication domains display distinct latent embeddings as can be seen from UMAP. (c) Distribution of spots in inferred communication domains across pathological annotations as in a. Each cell represents the proportion of spots of a communication domain in the corresponding pathological region. (d) Spatial map of neighborhood activity scores for the ligand-target pairs (e) Differentially active ligand-target pairs inferred by Renoir across all communication domains. (f) Cell type-specific ligand-target interactions for the major cell types in domains 1, 2, and 3. The inner color bar represents the cell types of the ligand (left) and target (top) and the outer color bar represents the communication domain. Each cell represents the average Pearson correlation between the ligand-target neighborhood scores and the abundances of the cell types expressing the ligand and target across the spots pertaining to the domain being considered. (g) - (h) Domain-specific ligand ranking based on their cumulative activities over target genes expressed by major cell types in the domain. (g) Top six ligands for communication domains 0, 1, 2, and 3 are represented. (h) represents the unique ligand-target pairs within each domain. Stacked color bars represent the cell types that express the ligand (right) and target (top). (i) Spatial map of four hallmark pathways - IL6 JAK STAT SIGNALING, TGF BETA SIGNALING, WNT SIGNALING and MAPK signalling pathway activity.

Figure 1

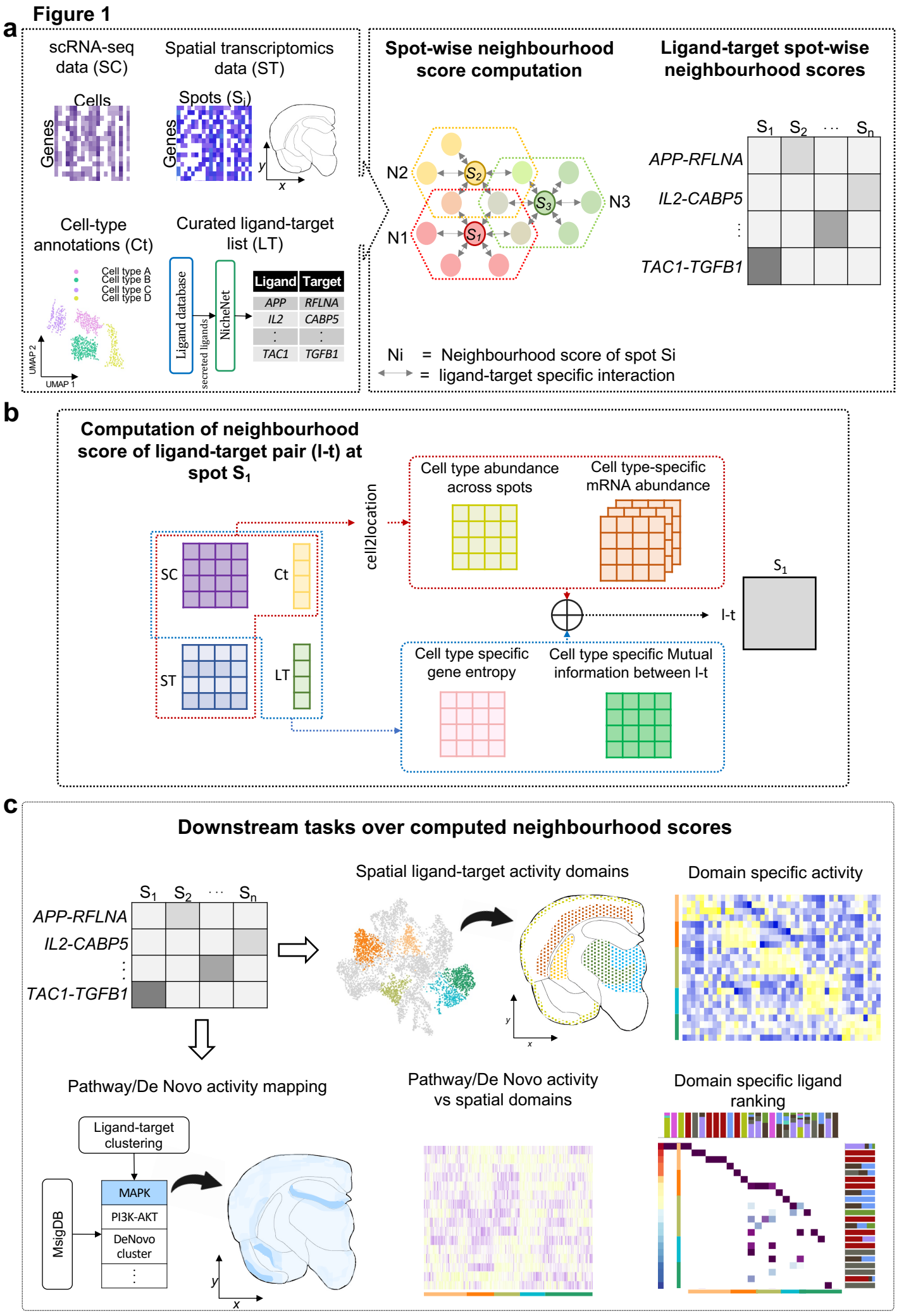

a
scRNA-seq data (SC)
Spatial transcriptomics data (ST)
Cells
Genes
Spots (Si)
Genes
y
x
Cell-type annotations (Ct)
Cell type A
Cell type B
Cell type C
Cell type D
UMAP 2
UMAP 1
Curated ligand-target list (LT)
Ligand database
secreted ligands
NicheNet
Ligand
Target
APP
RFLNA
IL2
CABP5
TAC1
TGFB1
Spot-wise neighbourhood score computation
N2
N1
N3
S2
S3
S1
Ligand-target spot-wise neighbourhood scores
S1
S2
Sn
APP-RFLNA
IL2-CABP5
TAC1-TGFB1
Ni = Neighbourhood score of spot Si
= ligand-target specific interaction
b
Computation of neighbourhood score of ligand-target pair (l-t) at spot S1
SC
Ct
ST
LT
cell2location
Cell type abundance across spots
Cell type-specific mRNA abundance
Cell type specific gene entropy
Cell type specific Mutual information between l-t
l-t
S1
c
Downstream tasks over computed neighbourhood scores
S1
S2
Sn
APP-RFLNA
IL2-CABP5
TAC1-TGFB1
Spatial ligand-target activity domains
Domain specific activity
Pathway/De Novo activity mapping
Ligand-target clustering
MsigDB
MAPK
PI3K-AKT
DeNovo cluster
Pathway/De Novo activity vs spatial domains
Domain specific ligand ranking

Figure 2

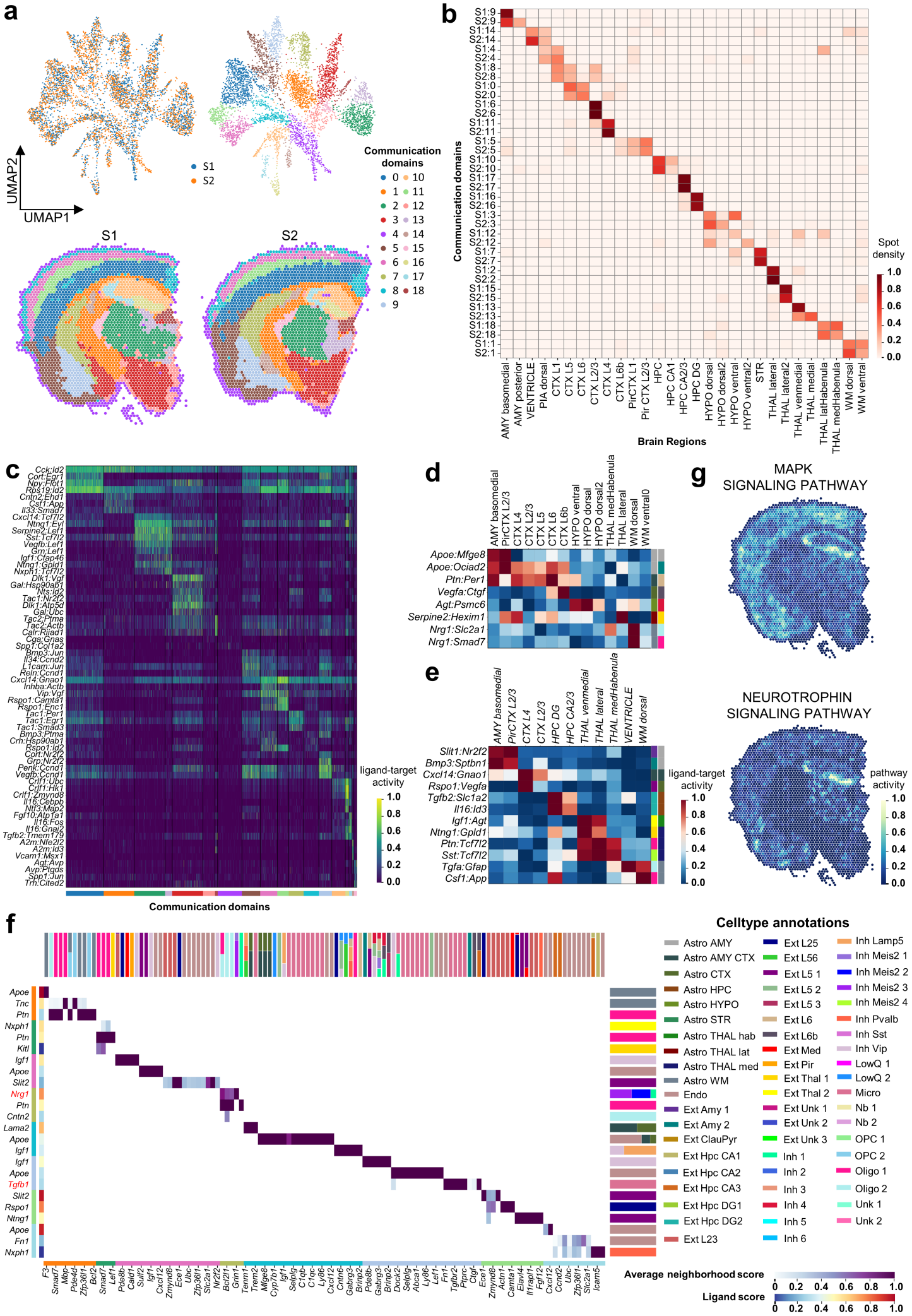

a
UMAP2
UMAP1
S1
S2
Communication domains
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
b
Communication domains
Brain Regions
Spot density
1.0 0.8 0.6 0.4 0.2 0.0
AMY basomedial
AMY posterior
VENTRICLE
PIA dorsal
CTX L1
CTX L5
CTX L6
CTX L2/3
CTX L4
CTX L6b
PirCTX L1
Pir CTX L2/3
HPC
HPC CA1
HPC CA2/3
HPC DG
HYPO dorsal
HYPO dorsal2
HYPO ventral
HYPO ventral2
STR
THAL lateral
THAL lateral2
THAL venmedial
THAL medial
THAL latHabenula
THAL medHabenula
WM dorsal
WM ventral
c
Communication domains
ligand-target activity
1.0 0.8 0.6 0.4 0.2 0.0
d
e
g
MAPK SIGNALING PATHWAY
NEUROTROPHIN SIGNALING PATHWAY
pathway activity
f
Celltype annotations
Astro AMY
Astro AMY CTX
Astro CTX
Astro HPC
Astro HYPO
Astro STR
Astro THAL hab
Astro THAL lat
Astro THAL med
Astro WM
Endo
Ext Amy 1
Ext Amy 2
Ext ClauPyr
Ext Hpc CA1
Ext Hpc CA2
Ext Hpc CA3
Ext Hpc DG1
Ext Hpc DG2
Ext L23
Ext L25
Ext L56
Ext L5 1
Ext L5 2
Ext L5 3
Ext L6
Ext L6b
Ext Med
Ext Pir
Ext Thal 1
Ext Thal 2
Ext Unk 1
Ext Unk 2
Ext Unk 3
Inh 1
Inh 2
Inh 3
Inh 4
Inh 5
Inh 6
Inh Lamp5
Inh Meis2 1
Inh Meis2 2
Inh Meis2 3
Inh Meis2 4
Inh Pvalb
Inh Sst
Inh Vip
LowQ 1
LowQ 2
Micro
Nb 1
Nb 2
OPC 1
OPC 2
Oligo 1
Oligo 2
Unk 1
Unk 2
Average neighborhood score
Ligand score
0 0.2 0.4 0.6 0.8 1.0

Figure 3

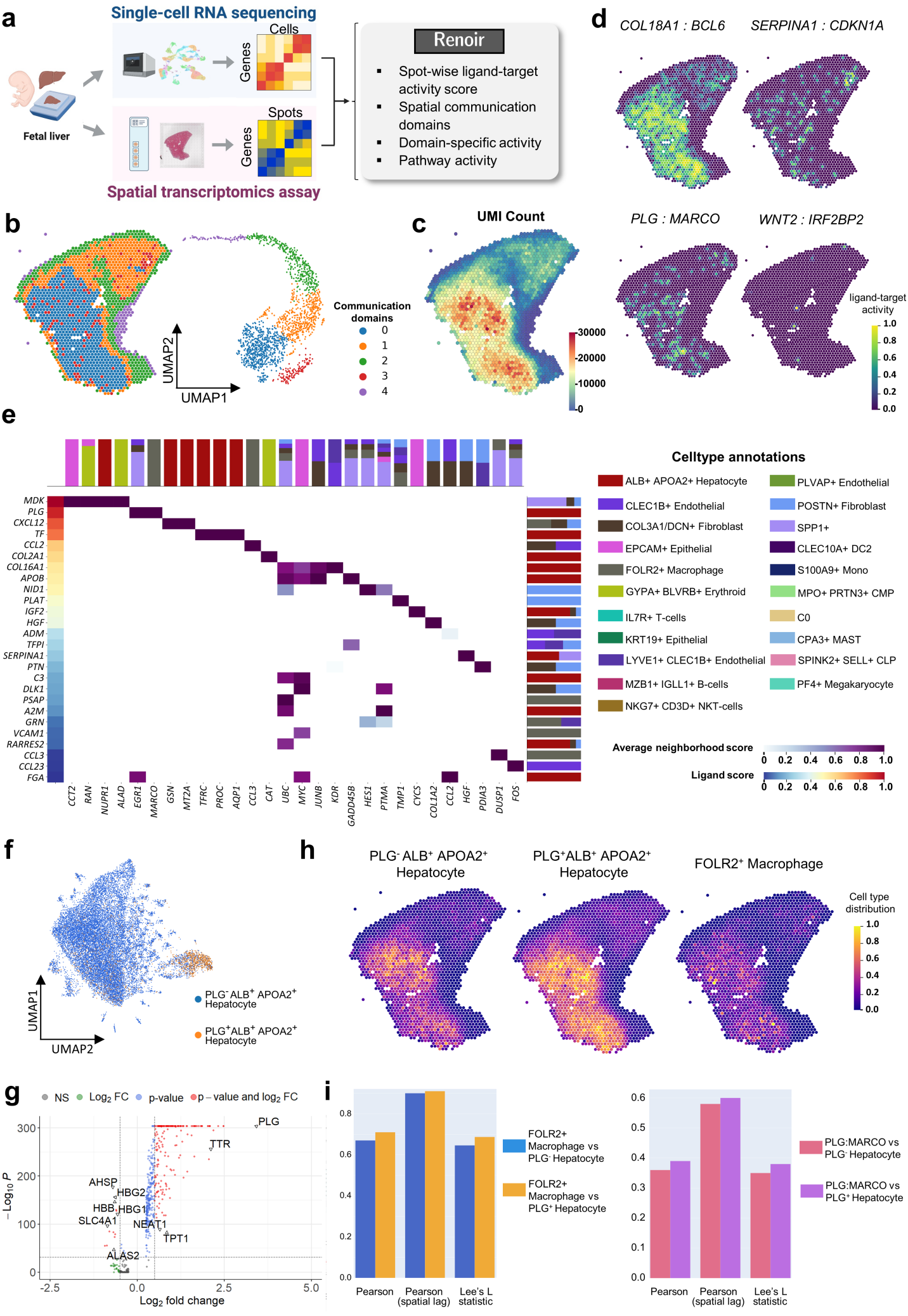
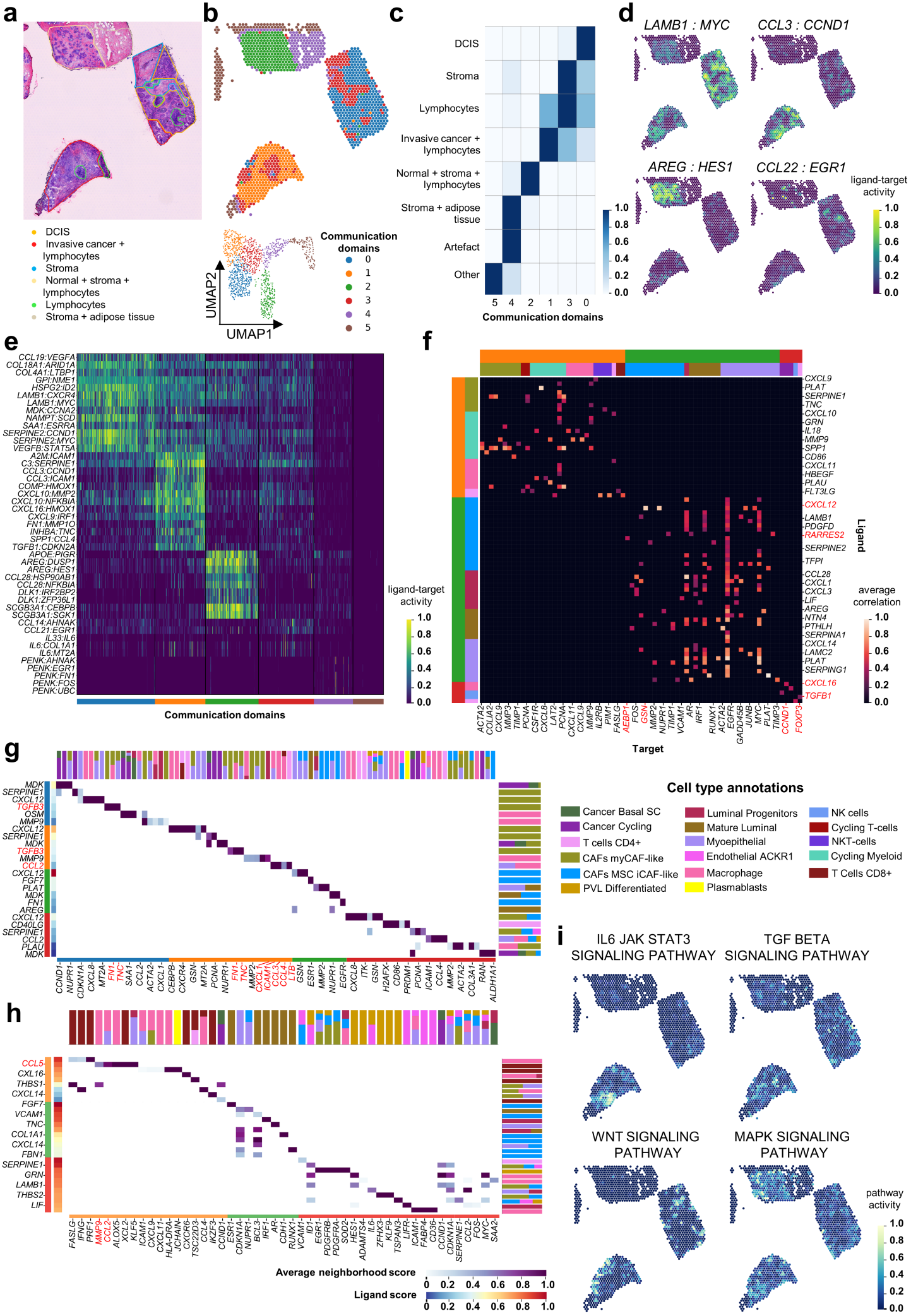

a
Single-cell RNA sequencing
Cells
Genes
Fetal liver
Spots
Genes
Spatial transcriptomics assay
Renoir
Spot-wise ligand-target activity score
Spatial communication domains
Domain-specific activity
Pathway activity
d
COL18A1 : BCL6
SERPINA1 : CDKN1A
b
UMAP2
UMAP1
Communication domains
0
1
2
3
4
c
UMI Count
30000
20000
10000
0
PLG : MARCO
WNT2 : IRF2BP2
ligand-target activity
1.0
0.8
0.6
0.4
0.2
0.0
e
Celltype annotations
ALB+ APOA2+ Hepatocyte
CLEC1B+ Endothelial
COL3A1/DCN+ Fibroblast
EPCAM+ Epithelial
FOLR2+ Macrophage
GYPA+ BLVRB+ Erythroid
IL7R+ T-cells
KRT19+ Epithelial
LYVE1+ CLEC1B+ Endothelial
MZB1+ IGLL1+ B-cells
NKG7+ CD3D+ NKT-cells
PLVAP+ Endothelial
POSTN+ Fibroblast
SPP1+
CLEC10A+ DC2
S100A9+ Mono
MPO+ PRTN3+ CMP
C0
CPA3+ MAST
SPINK2+ SELL+ CLP
PF4+ Megakaryocyte
Average neighborhood score
0 0.2 0.4 0.6 0.8 1.0
Ligand score
0 0.2 0.4 0.6 0.8 1.0
MDK
PLG
CXCL12
TF
CCL2
COL2A1
COL16A1
APOB
NID1
PLAT
IGF2
HGF
ADM
TFPI
SERPINA1
PTN
C3
DLK1
PSAP
A2M
GRN
VCAM1
RARRES2
CCL3
CCL23
FGA
CCT2
RAN
NUPR1
ALAD
EGR1
MARCO
GSN
MT2A
TFRC
PROC
AQP1
CCL3
CAT
UBC
MYC
JUNB
KDR
GADD45B
HES1
PTMA
TMP1
CYCS
COL1A2
CCL2
HGF
PDIA3
DUSP1
FOS
f
UMAP1
UMAP2
PLG- ALB+ APOA2+ Hepatocyte
PLG+ALB+ APOA2+ Hepatocyte
h
PLG- ALB+ APOA2+ Hepatocyte
PLG+ALB+ APOA2+ Hepatocyte
FOLR2+ Macrophage
Cell type distribution
1.0
0.8
0.6
0.4
0.2
0.0
g
NS
Log2 FC
p-value
p – value and log2 FC
PLG
TTR
AHSP
HBG2
HBB
HBG1
SLC4A1
NEAT1
TPT1
ALAS2
−Log10 P
300
200
100
0
-2.5
0.0
2.5
5.0
Log2 fold change
i
0.8
0.6
0.4
0.2
0.0
Pearson
Pearson (spatial lag)
Lee's L statistic
FOLR2+ Macrophage vs PLG- Hepatocyte
FOLR2+ Macrophage vs PLG+ Hepatocyte
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Pearson
Pearson (spatial lag)
Lee's L statistic
PLG:MARCO vs PLG- Hepatocyte
PLG:MARCO vs PLG+ Hepatocyte

Figure 4


a
DCIS
Invasive cancer + lymphocytes
Stroma
Normal + stroma + lymphocytes
Lymphocytes
Stroma + adipose tissue
b
UMAP2
UMAP1
Communication domains
0
1
2
3
4
5
c
DCIS
Stroma
Lymphocytes
Invasive cancer + lymphocytes
Normal + stroma + lymphocytes
Stroma + adipose tissue
Artefact
Other
5 4 2 1 3 0
Communication domains
d
LAMB1 : MYC
CCL3 : CCND1
AREG : HES1
CCL22 : EGR1
ligand-target activity
e
Communication domains
ligand-target activity
f
Ligand
Target
average correlation
g
Cell type annotations
Cancer Basal SC
Cancer Cycling
T cells CD4+
CAFs myCAF-like
CAFs MSC iCAF-like
PVL Differentiated
Luminal Progenitors
Mature Luminal
Myoepithelial
Endothelial ACKR1
Macrophage
Plasmablasts
NK cells
Cycling T-cells
NKT-cells
Cycling Myeloid
T Cells CD8+
h
Average neighborhood score
Ligand score
i
IL6 JAK STAT3 SIGNALING PATHWAY
TGF BETA SIGNALING PATHWAY
WNT SIGNALING PATHWAY
MAPK SIGNALING PATHWAY
pathway activity