Charting spatial ligand-target activity using Renoir

Narein Rao[1], Tanush Kumar[2], Rhea Pai[3,4], Archita Mishra[5,6], Florent Ginhoux[7,8], Jerry Chan[6], Ankur Sharma[3,4,6,9,10,*], and Hamim Zafar[1,11,12,*]

[1]Department of Computer Science and Engineering, Indian Institute of Technology Kanpur, India
[2]Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, India
[3]Harry Perkins Institute of Medical Research, Nedlands, Perth, Western Australia
[4]Curtin Medical School, Curtin University, Perth, Western Australia
[5]University of Sydney, Sydney, Australia
[6]KK Research Center, KK Women's and Children's Hospital, Singapore
[7]Singapore Immunology Network (SIgN), Agency for Science, Technology and Research (ASTAR), Singapore
[8]Gustave Roussy Cancer Campus, Villejuif, France
[9]Translational Genomics Program, Garvan Institute of Medical Research, Darlinghurst, Australia
[10]School of Clinical Medicine, Faculty of Medicine and Health, University of New South Wales, Kensington, Australia
[11]Department of Biological Sciences and Bioengineering, Indian Institute of Technology Kanpur, India
[12]Mehta Family Center for Engineering in Medicine, Indian Institute of Technology Kanpur, India
[*] *Corresponding Author*
[3,4,6,9,10]ankur.sharma@garvan.org.au (A.S.), [1,11,12]hamim@iitk.ac.in (H.Z.)

Abstract

The advancement of single-cell RNA sequencing (scRNA-seq) and spatial transcriptomics has enabled the inference of cellular interactions in a tissue microenvironment. Despite the development of cell-cell interaction inference methods, there is a lack of methods capable of mapping the influence of ligands on downstream target genes across a spatial topology with specific cell type composition, with the potential to shed light on 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. Renoir also spatially maps pathway-level activity of ligand-target genesets and identifies domain-specific ligand-target activities. Across spatial datasets with varying resolution (spot to single-cell) ranging from development to disease, Renoir inferred cellular niches with distinct ligand-target interactions, spatially mapped pathway activities, and identified context-specific novel cell-cell interactions. Renoir uncovers biological insights and therapeutically-relevant cellular crosstalk from spatial transcriptomics data.

keywords

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

Introduction

Coordination among the cells drive various functions in a multicellular tissue where cellular decision making is influenced by interactions with the environment composed of extrinsic stimuli and other cells. These interactions also known as cell–cell communication (CCC) play a crucial role in tissue development, biological processes and functions (1). Recently emerged single-cell RNA sequencing (scRNA-seq) technologies have made it possible to analyze tissue architecture at the resolution of individual cells (2; 3) and have shown great potential in investigating CCC (4). Computational methods have been developed for inferring cell-cell communication and mediating ligand-receptor (LR) pairs from scRNA-seq datasets by utilising ligand-receptor databases (5; 6). These methods including CellChat (6), CellphoneDB (5), iCELLNET (7), NATMI (8), scTensor (9), and SingleCellSignalR (10) mostly rely on defining functions based on the expression of ligand and receptor pair. Beyond ligand-receptor interactions, methods such as NicheNet (11), Cytotalk (12), SoptSC (13) and CCCExplorer (14) focused on inferring the effects of ligand–receptor interactions on possible downstream target gene expression in the receiving cells. While these methods have provided biological insights based on single-cell datasets, they often generate false-positives by failing to account for the spatial context of cell type arrangement in the tissue, as the spatial location information is not captured in scRNA-seq data. Retaining the spatial context is of paramount importance in CCC inference as cellular communications are often spatially constrained where signalling pathways are activated when ligands diffuse from sender cells to neighbouring receiver cells in close proximity (9).

Recent breakthroughs in sequencing (15; 16) and imaging-based (17) spatial transcriptomic technologies have enabled the measurement of gene expression profiles at varying cellular resolutions from single-cells to spots consisting of multiple or fractions of cells while preserving the spatial context information for the cells or spots. Such datasets have been utilized for spatially mapping cells from scRNA-seq data for inferring ligand-target interactions (e.g., SpaOTsc (18)). Moreover, CCC inference methods that specifically use spatial data have also been developed. Methods such as Giotto (19) and NCEM (20) use a spatial neighbor graph for inferring cell-cell interactions; CellPhoneDB v3 (21) uses the spatial information to define microenvironments within which interactions are restricted; stLearn (22) and SpatialDM (23) rely on the co-expression of ligand and

receptor genes in relation to the spatial location diversity; SVCA (24), MISTy (25) and COMMOT (26) use probabilistic model, machine learning model and optimal transport respectively for identifying spatially restricted ligand-receptor interactions. However, most of these methods focus on ligand-receptor interactions, they do not examine the effect of the cell-cell interaction on the downstream target gene expression. Moreover, most of these methods aim to uncover the interacting cell types globally and thus may be less sensitive in identifying local interactions. While a handful of methods including COMMOT, stLearn and SpatialDM can compute ligand-receptor interactions at the resolution of a spatial location, they fail to quantify the activity of ligand-target gene pairs at the same resolution given a specific spatial context with a fine-grained spatial distribution of cell types. Furthermore, the existing methods employing low-resolution spatial datasets (e.g., 10X Visium) cannot quantify cell type-specific interactions from the mixture of cell types present in a spot. Given that the existing computational methods were developed based on low-resolution spatial datasets, they may provide sub-optimal results for the recent single-cell resolution spatial technologies including NanoString CosMx and 10X Xenium. There is a critical requirement for reliable computational approaches that can elucidate the probable influence of the ligands on target gene expressions in the receiving cell in a particular spatial context with unique cell type composition.

To address these challenges, 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 either single-cell resolution spatial transcriptomic data or matched low-resolution spatial transcriptomic data and scRNA-seq data from the same tissue in order to delineate the spatial influence between possible ligand-target pairs curated from an existing database (11). 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 location of the spatial transcriptomics data by taking into account the cell type abundances, cell type-specific mRNA abundance values of the ligand and target genes, expression of suitable receptors for the ligand in the cell type harboring target gene, cell type-specific gene entropy measure and a cell type-specific mutual information between the ligand-target pairing as obtained from single-cell resolution spatial data or 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 neighborhood 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 neighborhood 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.

Using a wide range of synthetic datasets generated across a variety of tissue types (brain, cancer and intestine) as well as real datasets, we demonstrated that Renoir outperforms the state-of-the-art methods in inferring spatial activity of ligand-target pairs and communication niches. To demonstrate the utility of Renoir in uncovering spatial map of ligand-target activities across tissue types, we applied Renoir to three spatial transcriptomic datasets ranging from development to disease, including adult mouse brain, human triple negative breast cancer (TNBC), and human fetal liver, for 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, identified several spatially restricted interactions between cancer, stromal and immune cell populations in TNBC, and uncovered hepatocyte-macrophage interactions in developing fetal liver. To further demonstrate Renoir's ability to handle single-cell resolution spatial datasets, we applied it on a NanoString CosMx dataset from hepatocellular carcinoma for which Renoir identified known and novel onco-fetal interactions.

Results

Charting spatial ligand-target interactions using Renoir

In order to build a comprehensive computation framework to investigate spatial cell-cell interaction we developed Renoir. As shown in Figure 1, Renoir is an end-to-end computational framework for exploring the spatial map of ligand-target activities either by integrating spatial transcriptomics

and scRNA-seq datasets from the same tissue or by employing single-cell resolution spatial transcriptomics data. Renoir first curates a set of ligand-target pairs using NATMI's ConnectomeDB (8) and NicheNet (11). For a low-resolution spatial transcriptomic dataset (e.g., 10X Visium), using a well-annotated scRNA-seq data, Renoir quantifies a neighborhood activity score for each curated ligand-target pair for each spot in the spatial transcriptomic data. For a single-cell resolution spatial transcriptomic data, Renoir quantifies the neighborhood activity score for each curated ligand-target pair for each cell with a spatial context. 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/cell within the defined neighborhood of the spot/cell (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 outperforms state-of-the-art methods in inferring ligand-target interactions

We first evaluated Renoir's performance in inferring the spatial activity of ligand-target pairs using simulated datasets generated using a semi-synthetic enrichment-based simulation workflow (Fig. 2a,

see Methods for details). Specifically, we started with paired single-cell and spatial datasets based on which we simulated new spatial datasets where activity of a ligand-target pair was simulated in reference spatial locations (positive reference spots) through the enrichment of cell types associated with the ligand and target gene. Moreover, the simulated datasets also contained spatial locations which did not harbor any activity between the ligand and target (negative reference spots). To comprehensively cover the range of interactions across tissue types, we used three reference matched scRNA-seq and ST datasets from human intestine (27), triple negative breast cancer (TNBC) (28) and dorsolateral prefrontal cortex (DLPFC) (29) for the simulation. The simulation workflow as described in Figure 2a was applied to generate 43, 15 and 26 simulated datasets corresponding to the human intestine, TNBC and DLPFC datasets respectively, where, each simulated dataset enriched a specific ligand-target interaction. Since Renoir quantifies ligand-target activity at a specific spatial context and is the only method (to the best of our knowledge) to do so contrary to the other cell-cell interaction methods, which infer ligand-receptor interactions by employing the spatial data, as competitor methods, we selected state-of-the-art ligand-receptor interaction inference methods COMMOT (26), stLearn (30) and SpatialDM (23). These three methods are capable of quantifying ligand-receptor interactions at the resolution of spatial spots and we adopted them for quantifying ligand-target activity from our simulated datasets. The performance of each method was evaluated based on a spatial activity accuracy metric that computes the Pearson correlation of the spot-level ligand-target activity score measured by a method with the true simulated activity of the ligand-target pair in the positive and negative reference spatial locations (see Method for details). In addition, we also compared the methods in terms of precision and recall (see Methods) for the ligand-target pairs to evaluate their ability to distinguish the spots harboring ligand-target interactions from the spots devoid of such interactions.

In each simulated dataset across the tissue types, Renoir was able to capture the ligand-target interactions more precisely. To illustrate, we consider the ligand-target pair *A2M-MYC*, simulated from the human intestine dataset (Fig. 2b), where the positive (blue) and negative (red) reference spots included enrichment and abatement of said ligand-target pair interaction respectively as is also evident from the spatial expression plot of the two genes. Amongst the positive spots, we observed Renoir's ligand-target activity scores to truly reflect the enriched spots with high scores as compared to COMMOT and stLearn. While SpatialDM had high scores for positive spots,

it also had scores all across the tissue and failed to differentiate the positive and negative spots appropriately. Renoir was also able to correctly quantify the absence of ligand-target interactions in the negative reference spots. In comparison, all three competitor methods showed false-positive activity of the gene pair in multiple negative spots (red bounded regions labeled 1, 2 and 3) with near obsolete *A2M* and *MYC* expression. Renoir is capable of computing the activity of a ligand-target by considering the expression of a suitable receptor at the receiving cell. To showcase this, we considered the pair *IFNG-JCHAIN*, simulated using the TNBC dataset, where *IFNG* was associated with Natural Killer T Cells (NKT Cells) and $CD8^+$ T Cells; and *JCHAIN* was associated with B Cells and Dendritic Cells (DCs). However, B cells do not express a receptor for *IFNG* (Supplementary Fig. 1a) and thus the *IFNG* secreted from NKT Cells or $CD8^+$ T Cells cannot activate *JCHAIN* in B cells. Therefore, the positive reference spots harbored interactions with DCs as the target cell type whereas the negative spots contained the enrichment of B cells as the target cell type (see Methods section for more details). In this scenario, by accounting for the presence of suitable receptors while calculating ligand-target activity scores, Renoir correctly scored the negative spots with negligible scores; whereas other methods failed to consider such scenarios and therefore incorrectly gave high scores to the negative spots as depicted by the red bounded regions (Supplementary Fig. 1a). Similarly, for the DLPFC dataset (e.g., for the pair *NPY:CARTPT*) also, Renoir's spatial scores for the ligand-target activity were more precise as compared to the other methods (Supplementary Fig. 1b). SpatialDM and COMMOT had less distinction between the positive and non-positive spots and stLearn failed to capture the activity pattern due to the high range of *NPY* expression and relatively low range of *CARTPT* expression. Quantitatively, Renoir outperformed other CCC inference methods in inferring spatial ligand-target activity by achieving 45.8, 98.5, and 12.4% improvement in the median spatial activity accuracy across the human intestine, TNBC and DLPFC datasets (Fig. 2c) indicating Renoir's superiority in capturing the spatial variation of the ligand-target activity. Furthermore, in terms of average precision and recall over the ligand-target pairs, Renoir outperformed SpatialDM, COMMOT, and stLearn across the tissue types across different threshold values used for calculating precision and recall (Supplementary Fig. 1c). For varying threshold values, Renoir was able to consistently maintain high precision and recall values across the tissue types. In comparison, the precision and recall values for the other methods varied across datasets, with a method having high precision-

recall for one dataset but low precision-recall for another.

Finally, we evaluated the accuracy of the spatial communications domains inferred based on the ligand-target interaction scores inferred by a CCC inference method. For this, we utilized 10X Visium datasets from 12 dorsolateral prefrontal cortex (DLPFC) samples (29), in which the authors manually annotated six cortical layers and white matter for each sample based on cytoarchitecture and selected marker genes. Considering the manual annotations as ground truth, we evaluated the utility of the ligand-target scores inferred by Renoir in identifying the layers and compared its performance against that of COMMOT, SpatialDM and stLearn. Each method was evaluated based on adjusted Rand index (ARI) and normalized mutual information (NMI) scores calculated by comparing the communication domains inferred by the method against the ground truth annotations over all 12 DLPFC samples (see Methods for details). Renoir achieved the highest ARI and NMI scores considering all the samples and outperformed all the other methods (Fig. 2d) indicating that it is able to capture spatial variation and architecture better than the other CCC inference methods. Figure 2e shows how the communication domains inferred by Renoir captured the underlying spatial structure for the sample 151675 better than the other methods for which the inferred communication domains were much noisier.

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 (31) (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 (Fig. 3a) and were similarly distributed across the sections (Fig. 3b). 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 (Fig. 3c).

The communication domains were found to harbor diverse cell types that were spatially localized

and abundance pattern concordant across the replicates (Supplementary Fig. 2a-b). The domains further harbored differential activity of specific ligand-target pairs (Fig. 2d, Supplementary Fig. 3b-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 (32; 33) and domain 3 specific to the hypothalamus exhibited *Dlk1* activity known to be expressed in hypothalamic arginine vasopressin and oxytocin neurons (34). 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 9 exhibited high activity of *Rspo1*, which is known to be highly expressed within the cortex (35; 36), domain 7 exhibited high activity of hippocampal marker *Crlf1*, and domain 13 exhibited high activity of *Ccnd1* and *Nr2f2*, which are known to be highly expressed within the Amygdala region (37; 38). Supplementary Table 1 lists the domain-specific ligand-target interactions with literature support for the associated brain region.

Kleshchevnikov et al. (31) 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 3d and Supplementary Figure 3d 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 3e and Supplementary Figure 3e 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*) and Amygdala (*Ptn:Nr2f2*) respectively. On the other hand, *Agt* expressed by diencephalic astrocytes (Thalamus) were found to activate genes in inhibitory neurons (*Psap:Pde4dip*). White matter astrocytes interacted with oligodendrocytes via *Tnc* which activated gene *Adamts4*. Similarly, different genes in telencephalic astrocytes in Amygdala, Cortex, and Hippocampus were found to be activated by ligands expressed by excitatory neurons in Layer *4* (*Rspo1:Vegfa*), Amygdala (*Slit1:Nr2f2*), and Hippocampus DG (*Crlf1:Ccnd3*, *Slit1:Tcf4*) respectively. Excitatory neurons in the thalamus were found to activate genes in diencephalic astrocytes (*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 (Fig. 3f, Supplementary Fig. 3f) that are expressed by the major cell types in the domain (see Methods for details) which further highlighted the region-specific enrichment of certain ligands. For instance, *Bdnf* known to be associated with numerous functions within the hippocampus (39) displayed high activity within domain 16 associated with the hippocampus. *Nrg1*, known to have high impact on cortical functions (40) exhibited high activity within domain 5. 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) (Fig. 3g, Supplementary Fig. 3g and Supplementary Fig. 4). Consistent with prior works (41; 42), cortex and hippocampus were highly enriched in MAPK signalling and NEUROTROPHIN signalling pathways, respectively.

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 (Fig. 4a) from a triple negative breast cancer (TNBC) patient (28). Renoir quantified the spatial neighborhood activity scores for 18654 ligand-target pairs based on which, it clustered the spatial spots into six communication domains which in addition to being distinct in low-dimensional embeddings (Fig. 3b), were primarily associated with morphologically distinct pathological regions (Fig. 3c). The domains were further differentiated based on the ligand-target activities and the composition and abundance of the cell types they harbored (Fig. 4d-e, Supplementary Fig. 5a-b). Since two (clusters 4 and 5) of the six domains had low overall UMI count (Supplementary Fig. 5b), 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, histologically annotated as ductal carcinoma in situ (DCIS), harbored a high abundance of cancer cycling and cancer basal SC cell types (Fig. 4c, Supplementary Fig. 5a); cancer-associated fibroblasts (CAFs) and different T cell populations (Supplementary Fig. 5a); and demonstrated differential activity of MYC and SAA1 (*LAMB1-MYC*, *SERPINE2-MYC*,

SAA1-ESRRA) known to be overexpressed in basal-subtype of breast cancer (43; 44; 45); and MDK (*MDK-CCNA2*), a known cancer progression marker (46). Domain 0 also harbored high activity of *VEGF* (*VEGFA*, *VEGFB*) (known to be highly up-regulated in breast cancer (47; 48; 49)), and other stromal ligands (*COL18A1*, *COL4A1*) (Fig. 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 (Fig. 4c, Supplementary Fig. 5a). Consequently, interactions associated with tumor progression and invasion (e.g., *CXCL10-MMP2*, *CCL3-ICAM1* (50), *FN1-MMP10* (51)), as well as suppression (e.g., *CXCL9-IRF1* (52)) among these cell types were enriched in domain 1 (Fig. 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 (Fig. 4c, Supplementary Fig. 5a). Interestingly, this domain harbored high activity of the ligand *CCL28* which is known to promote breast cancer and metastasis (53), target *HES1* which plays a crucial role in malignant cell proliferation (54) and *AREG* which is expressed by myoepithelial cells and loss of which can inhibit tumor proliferation, growth and invasiveness (55) (Fig. 4e). Domains 3 and 4, despite low UMI count, showcased distinct ligand-target activities (e.g., *IL6*, *IL33*, *IL1RN* in domain 3 and *PENK* in domain 4 (Fig. 4e)).

We further investigated the most prominent cell type-specific ligand-target interactions (Fig. 4f, Supplementary Fig. 6a) 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), *CXCL16-CCND1* (domain 3). 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* (56) 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 (57; 58), (Fig. 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) (59) and it was found to target *ICAM1* (a molecular target for TNBC) (50) in myoepithelial cells and

macrophages, other chemokines *CCL3* and *CCL4* as well as lymphocyte inhibitory molecule *LTB* in *CD4*$^+$ T cells (Fig. 4g). Within each communication domain, we also found ligands that were specifically active in a single communication domain (Fig. 4h). For example, in domain 1, *CD8*$^+$ T cells secreted the ligand *CCL5* (known to promote TNBC progression through the regulation of immunosuppressive cells (60)) which 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 Figs. 7-10). Amongst the six subdomains in domain 0 (Supplementary Fig. 7), in subdomain 4 *TGFB1* and *FBN1* activated several EMT-associated genes (*FBN1*, *MMP2*, *COL5A1*, *PDGFRB*, etc.) (61) and cancer progression and regulation markers (*DUSP1*, *JUNB*, and *MYC*) (62; 63) respectively. In contrast, subdomain 1 which harbored different immune cells (*CD4*$^+$ and *CD8*$^+$ T cells, DCs, NKT cells) as well as iCAFs, showcased extensive chemokine activity (*CCL5*) which targeted different genes in different immune cells and iCAFs (e.g., *EGR1*). Domain 1 was segregated into 4 subdomains (Supplementary Fig. 8), of which subdomain 2 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 (64; 65; 66; 67). Similarly, subdomains of other domains showcased interesting activity particular to the cell type abundant across the subdomains (Supplementary Figs. 9-10).

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 Fig. 6b). 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 (Fig. 4i). IL-6/JAK/STAT3 signaling was highly active in subdomain 1 of domain 1 (Fig. 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 (68). Taken together, Renoir was able to provide comprehensive insights into the spatial heterogeneity of ligand-target activity

and its implication in tumor biology.

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 (69) (Fig. 5a). Renoir identified 17630 ligand-target pairs active in the tissue, quantified their spatial neighborhood activity scores and based on these scores, identified five spatial communication domains (Fig. 5b), which differed based on ligand-target activity and distribution of UMI count across the tissue section (Fig. 5b-c and Supplementary Fig. 4a). Domain 0 demonstrated the majority of ligand-target activity, followed by domain 1. On the other hand, domains 2 and 4 showed sparse activity (UMI counts were low) (Fig. 5c) while domain 3 was specific to the target *CDKN1A* (Supplementary Fig. 4a). This is in line with the UMI count distribution across spots, where the UMI count was most abundant in domains 0 and 3, followed by domains 1, 2 and 4. Cell types were found to be evenly distributed across domains 0, 1 and 3 (Supplementary Fig. 11b). Activities pertinent to fetal liver were found to be widespread (Fig. 5d). For example, *COL18A1* associated with type XVIII collagen (a prominent component of the Extracellular Matrix (ECM) in the liver (70)) was found to be highly expressed across the tissue and activate multiple targets including hematopoietic transcription factors *BCL2L1* and *BCL6* (71). 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 (72) and *CXCL12* which helps in retaining hematopoietic stem cells in the fetal liver (73) (Fig. 5e). Importantly, we 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 (74) and fetal-like reprogramming of liver macrophages plays an important role in im-

354 munosuppressive phenotype in liver cancer (69). As shown in Fig. 5e and Supplementary Fig. 12a,
355 we observed that $MARCO^+$ macrophages interact with $PLG^+/APOB^+$ subset of hepatocytes. The
356 PLG (plasminogen) has been known to be important for hepatic regeneration (75; 76) and PLG
357 deficiency can lead to impaired remodelling in mouse models after acute injury. Next, we subdi-
358 vided hepatocytes into two groups: $ALB^+APOA^+PLG^+$ and $ALB^+APOA^+PLG^-$ (Figs. 4f-g). As
359 shown in Figs. 5g-h, these two subsets are transcriptionally and spatially distinct from each other
360 indicating heterogeneity of the fetal liver hepatocyte population. We observed that the majority of
361 the ALB^+APOA^+ hepatocyte population does not express PLG while there is a specific cluster with
362 a high expression of PLG (Fig. 5f). In the Visium data, the spots expressing $ALB^+APOA^+PLG^+$
363 hepatocytes also contained $FOLR2^+$ macrophages (Fig. 5h). $FOLR2^+$ macrophages are embryonic
364 macrophages present in the fetal liver microenvironment (69). Spatial correlation analyses further
365 showed higher similarity of $FOLR2^+$ macrophages and $ALB^+APOA^+PLG^+$ hepatocytes as com-
366 pared to PLG^- hepatocytes in terms of cell type distribution as well as neighborhood activity score
367 (Fig. 5i). Taken together, these results provide important insights into the macrophage-hepatocyte
368 niche in the developing human liver.

369 Renoir identifies onco-fetal interactions in hepatocellular carcinoma

370 To evaluate Renoir's ability in handling single-cell resolution ST datasets, we applied it on a
371 Nanostring CosMx dataset generated from a hepatocellular carcinoma (HCC) patient (77). The
372 dataset consisted of 70166 cells distributed across 46 cell types and 959 genes.

373 We recently (69; 77) discovered the presence of onco-fetal tumor microenvironment (TME) in
374 hepatocellular carcinoma, where the $FOLR2^+$ TAMs, $PLVAP^+$ ECs and $POSTN^+$ CAFs were
375 identified as onco-fetal cell types that shared transcriptional programs with their fetal liver coun-
376 terparts. We evaluated Renoir's performance in identifying the onco-fetal interactions from the
377 single molecule imaging (CosMx) dataset. To perform an unbiased analysis, we used Renoir to
378 quantify the activity of the ligands expressed by the onco-fetal cell types with the other major cell
379 types present in the neighborhoods of the onco-fetal cells (Supplementary Fig. 13a) using Renoir's
380 ligand-ranking formulation (Fig. 6a). Renoir was able to characterize the activity of several known
381 core onco-fetal ligands (e.g., *VEGFA*, *TGFB3*) as well as ligands shared by two onco-fetal cell
382 types (e.g., *BMP7*, *CXCL12*, *TGFB1*, *IFNG*) (Supplementary Fig. 13b) (77). Specifically, Renoir

identified the activity of known onco-fetal interactions *CXCL12-CXCR4* and *VEGFA-KDR* (Fig. 6c). *CXCL12-CXCR4* signaling was particularly elevated in onco-fetal niche where *CXCL12* was expressed by *PLVAP*$^+$ ECs and *FOLR2*$^+$ TAMs and it activated *CXCR4* in *POSTN*$^+$ CAFs as well as macrophages and Tregs. Renoir correctly identified the co-localization of these cells and quantified the activity of the ligand-target pair in the neighborhood. Similarly, *VEGFA* was expressed by all three onco-fetal cell types and it interacted with *KDR* in co-localized ECs and *MUC6*$^+$ bipotent hepatocytes (Fig. 6c).

Renoir discovered onco-fetal niche mediated reprogramming of *MUC6*$^+$ bipotent cells

Renoir not only validated known onco-fetal interactions but also identified novel biological interactions between onco-fetal niche and putative stem-like *MUC6*$^+$ cells in liver cancer microenvironments. Renoir identified the *MUC6*$^+$ bipotent hepatocytes to have the highest number of interactions with the onco-fetal cell types (more specifically with *POSTN*$^+$ CAFs) (Fig. 6b). We delved into interactions specific to onco-fetal cells and *MUC6*$^+$ bipotent cells by quantifying the cell type-specific activity of the relevant ligand-target pairs in the onco-fetal neighborhoods (Supplementary Fig. 13c). We noticed the activation of *MYC* and *TP53* in *MUC6*$^+$ bipotent cells by multiple onco-fetal ligands including *BMP2, TGFB3*, and *CXCL12*. We utilized gprofiler (78) to infer the pathways in the *MUC6*$^+$ bipotent cells targeted by the onco-fetal ligands and identified various interleukin signaling pathways (e.g., IL4, IL13, IL10), cytokine signaling and prostaglandin signaling (Fig. 6d) to be enriched. Since prostaglandin signaling is known to promote HCC progression (79), we further looked into the activity of *IL6-PTGS2*. High scores for *IL6-PTGS2* was found in the tissue niche where *MUC6*$^+$ bipotent cells were co-localized with the oncofetal cells indicating the potential role of onco-fetal cells in the activation of PTGS2 in the *MUC6*$^+$ bipotent cells (Fig. 6e). Importantly, these observations indicate that application of Renoir could lead to the identification of novel biological insights in the tumor microenvironment.

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 approaches have 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. Renoir can work on datasets generated using spot-resolution spatial technologies as well as single-cell resolution spatial technologies.

Our comprehensive benchmarking of Renoir on a variety of datasets simulated from real datasets across tissue types demonstrate its superiority over existing CCC inference methods in quantifying the ligand-target activities across spatial locations. Specifically, by accounting for the receptor expression in the target cell type, Renoir is capable of reducing false positive interactions which are abundantly found by other methods. Using DLPFC spatial datasets with ground truth annotations of spatial domains, we further illustrated that the communication domains inferred by Renoir can better capture the underlying spatial structure of the tissue as compared to other methods.

Our application of Renoir to three distinct 10X Visium spatial transcriptomics datasets including adult mouse brain, human fetal liver, and human triple negative breast cancer (TNBC) demonstrate Renoir's versatility in characterizing ligand-target interactions across tissue types. 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*. The application of Renoir on a single-molecule imaging (Nanostring CosMx) hepatocellular carcinoma dataset further shows Renoir's utility in handling single-cell resolution large-scale spatial datasets. From the hepatocellular carcinoma dataset, Renoir identified several onco-fetal interactions involving endothelial cells, macrophages and CAFs and uncovered the role of these onco-fetal cell types in activating prostaglandin signaling in the bipotent hepatocytes.

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. 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 either paired cell type annotated scRNA-seq data and spatial transcriptomic (ST) data or single-cell resolution spatial transcriptomic 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 the following - (1) a spatial transcriptomics dataset $X^{st} \in \mathbb{R}^{S \times G}$ with gene expression for S spots and G genes. Depending on the technology, the spots can contain a mixture of cells (e.g., 10X Visium) or single-cell (e.g., Nanostring CosMX). In case of cell-resolution ST data, cell type annotations of the spots to one of C cell types is required. (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 (required for ST datasets where each spot contains mixture of cells), where each cell mapped to one of C cell types. Note that the single-cell gene expression dataset is optional when single-cell resolution spatial transciptomic dataset is available.

Acquiring curated ligand-target pairs: We first curate a set of secreted ligands from NATMI's connectomeDB2020 (8) 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 (11). 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 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. For data wherein the spots are not uniformly distributed across the sample (e.g., Slide-seq, StereoSeq, CosMx), the neighborhood for a spot s is defined to be a set of spots (inclusive of s) which are at most r unit distance away from s, where r is user defined.

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 $c_2 \in C_2$ and c_2 expresses at least one receptor of l respectively. c_2 is considered to express a receptor if the receptor gene is expressed in more than $thresh\%$ of cells of that cell type (computed based on scRNA-seq/single-cell ST data), where $thresh$ in our analysis was set to 5 following NicheNet. 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 single-cell (either scRNA-seq or single-cell ST) 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 (31) 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. For single-cell resolution datasets, the cell type annotation of the spots are directly used as cell type abundance information, where $W_{S\times C}$ represents the cell type annotation of each spot and $U_{S\times C\times G}$ represents the original single-cell resolution ST dataset. This gives us,

$$U_{S\times C\times G} = U_{S\times G} = \{u_{sg}\} = X^{st} \tag{6}$$

The expected expression of a specific gene g at spot s for cell type c can then be calculated as:

$$E_{S\times C\times G} = \{e_{scg}\} = \frac{\Sigma_{n\in C}\Sigma_{s'\in S}u_{s'ng}}{S} \times \frac{\Sigma_{g'\in G}u_{scg'}}{\Sigma_{n\in C}\Sigma_{g'\in G}u_{sng'}} \tag{7}$$

For single cell resolution datasets, since each cell is of a specific cell type the above formulation collapses to,

$$E_{S\times C\times G} = E_{C\times G} = \{e_{cg}\} = \frac{\Sigma_{s'\in S}u_{s'cg}}{S} \tag{8}$$

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 (80). 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 ($n = 500$) from the scRNA-seq dataset. We acquire a list of ligand-receptor pairs from NATMI's connectomeDB (8) and identify the major ligands that are also present within the marker gene lists of the major cell types within the domain. 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 (11). 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 200 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 (81; 82), 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.

Benchmarking of Renoir

Generation of simulated data

For benchmarking the performance of Renoir, we simulated spatial datasets based on a semi-synthetic approach inspired from the simulation approach in (83). In order to simulate a dataset with known spatial interaction between ligand and target, we start with a paired scRNA-seq and ST data (the reference dataset) and cell type deconvolution of the reference ST dataset is performed using cell2location (31). After that, we simulate the ST dataset using the following steps in order to maintain a controlled environment, where we can keep track of the different kinds of interactions enriched within the data.

1. Select a set of ligand-target pairs LT such that both the ligand and target are marker genes (top 100 markers) of at least one cell type in the paired scRNA-seq data. After that, the pairs in LT are segregated into two mutually exclusive subsets: LT_m and LT_s. LT_m consists of the ligand-target pairs where the target gene is a marker for multiple (two or more) cell types and LT_s consists of every other pair, where the target gene is a marker of exactly one cell type. In both LT_m and LT_s, the ligand could be a marker of one or more cell types. In our analysis, we were able to find a maximum of two cell types associated with each of the ligand or target genes.

2. For a specific ligand-target pair $lt \in LT_s$, where the target is a marker of a cell type ct_{t1}; in the ST dataset, we generate two types of reference spots, positive and negative spots respectively. Positive reference spots denote the spots within the tissue where ligand-target interactions for lt are simulated through the enrichment of the ligand cell type and target cell type ct_{t1}, where a spot is enriched with a cell type by adding $n_e = Uniform(2, 5)$ single cells of the said cell type into said spot. The UMI counts are then scaled back to the original UMI count to maintain the original UMI distribution of the sample being simulated. Negative references refer to the spots within the tissue that do not harbor any activity for lt as indicated by the low proportion of ligand cell type and target cell type ct_{t1}. Specifically, negative reference spots are selected as a set of spots for which both ligand cell type and target cell type ct_{t1} proportions are low and their neighbors also have low proportions (below acceptable

threshold) of ligand and target cell types. In the negative reference spots, interactions of lt can not occur given the low abundance of both the ligand and target cell types in these spots as well as in the neighboring spots.

3. For a specific ligand-target pair $lt \in LT_m$ where the target is a marker of two cell types ct_{t1} and ct_{t2}; in the corresponding ST dataset, we generate two types of reference spots, positive and negative. Without loss of generality, we assume that the cell type ct_{t1} does not express a known receptor of the ligand l but the cell type ct_{t2} expresses a known receptor of l (according to a ligand-receptor database, e.g., NATMI). Positive references are those spots within the tissue that have been enriched with the ligand cell type and target cell type ct_{t2}. Specifically, a set of spots are selected that are not devoid of the ligand cell type. This is to ensure that the enrichment follows the underlying distribution of cell types in the dataset. For each of the selected spots, the spot is enriched with the ligand cell type and the neighbors of the spot are enriched with the target cell type ct_{t2}. Negative references are those spots within the tissue that have been enriched with the ligand cell type and target cell type ct_{t1} (type 1 negative); or spots within the tissue that do not harbor any activity for lt as indicated by the low proportion of ligand cell type and target cell type ct_{t2} (type 2 negative). To simulate type 1 negative spots, a set of spots are selected from the set of all spots with low ct_{t1} and ct_{t2} proportions. We then enrich these spots with the ligand cell type and the target cell type ct_{t1}. Thus for the type 1 negative spots the ligand-target interactions pertinent to the ligand cell type and target cell type ct_{t1} could not occur since no known receptors are expressed by ct_{t1} despite the co-localization of the cell types. The simulation of type 2 negative spots follows the same strategy as in case of $lt \in LT_s$ considering ct_{t2} as the target cell type.

Scoring and evaluation strategy

To evaluate the performance of a CCC method in inferring the ligand-target activity at spot resolution, we employ the following scores.

Accuracy of spatial activity: The true activity of a ligand-target pair across the positive and negative reference spots is computed by employing the local Moran's bivariate test-statistic ($R_s^{ct_l ct_t}$) between the ligand and target cell types,

$$R_s^{ct_l ct_t} = \left(\frac{w_{s,ct_l} - \mu_{ct_l}}{\sigma_{w_{s,ct_l}}} \times SL_{s,ct_t} \right) + \left(\frac{w_{s,ct_t} - \mu_{ct_t}}{\sigma_{ct_t}} \times SL_{s,ct_l} \right)$$

$$SL_{s,x} = \frac{\sum_{i \in \mathscr{N}_s} x_i - \mu_x \cdot |\mathscr{N}_s|}{\sigma_x}$$

where $w_{s,x}$ represents the cell type abundance of cell type x at spot s, μ_x represents the mean cell type abundance of cell type x across all spots considered, σ_x represents the standard deviation of cell type abundance of cell type x across all spots considered and $SL_{s,x}$ represents the spatial lag of cell type x at spot s. This provides a measure of local interactions between the ligand and target cell types. Pearson's correlation (as discussed before) is then used to calculate the average correlation between the ligand-target activity scores inferred by each method and $R_s^{ct_l ct_t}$ which gives a measure of how well does the inferred activity scores align with the true interactions as quantified in terms of spatial cell type co-localization.

Precision and Recall: To calculate the precision and recall of a CCC inference method in inferring the spatial interactions of ligand and target, we divide the spots into two sets based on a predefined threshold set on the ligand-target scores computed by the CCC inference method. Given a specific method, the spots which have a score greater than or equal to the threshold are considered to be positive predicted spots (harboring the ligand-target interaction) and the spots with a score lesser than the threshold are considered to be negative predicted spots (no ligand-target interaction). We then consider the these positive and negative predicted spots and compare them against the ground truth positive and negative reference spots. A predicted spot is considered to be a true positive (TP) if the spot is a positive predicted spot and is also a positive reference spot, true negative (TN) if the spot is a negative predicted spot and is also a negative reference spot, false positive (FP) if the spot is a positive predicted spot but is a negative reference spot and a false negative (FN) if the spot is a negative predicted spot but a positive reference spot. For each method, we calculate the TP, TN, FP and FN values to calculate the precision and recall of a method for a given simulated dataset. Thus precision and recall are calculated for each simulated ligand-target pair and the final precision and recall for a dataset for a specific threshold is computed to be the average of the precision and recall values across all considered ligand-target pairs for the

same dataset. Instead of selecting a single threshold, the threshold is varied to determine the precision-recall curve for each method across various threshold values.

Comparison of communication domains with ground truth layers in DLPFC dataset In order to estimate how coherent the inferred communication domains are with known regions of the dataset, we first estimate the communication domains across all 12 DLPFC samples for each of the methods (Renoir, COMMOT, SpatialDM and stLearn). This is done so by applying leiden clustering (resolutions range from 0.1 to 1.6 in intervals of 0.1) over the ligand-target interaction scores generated by each of the methods across all samples. The NMI and ARI values are calculated between the cluster labels and ground truth annotations; for each of the methods across all samples over all resolutions. At each resolution, the averaged NMI and ARI values are calculated for each method across all samples. For each method, the cluster labels corresponding to the resolution at which the averaged NMI and ARI values is the best for said method in each sample, is now considered as the best communication domain inferred by a method for a specific sample. This results with a list of best NMI/ARI values for each method across each sample. The box plots are then generated over the final inferred NMI/ARI values.

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 (22). 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 (84), 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 (85). 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 incorpo-

rate spatial autocorrelation alongside spatial patterning between the two variables of interest (85). 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. The Nanostring CosMx data for hepatocellular carcinoma has been obtained from https://doi.org/10.6084/m9.figshare.23972991.

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], Science and Engineering Research Board (SERB), Government of India Startup Research Grant [SRG/2020/001333], and Har Gobind Khorana Innovative Young Biotechnologist Award Grant [BT/13/IYBA/2020/05] 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 and benchmarking. TK implemented the simulation experiment and performed benchmarking. RP, AM, FG and JC generated the fetal liver dataset. HZ, AS, and NR wrote the manuscript and all authors approved the manuscript.

Competing Interests

The authors declare no competing interests.

References

[1] Armingol, E., Officer, A., Harismendy, O. & Lewis, N. E. Deciphering cell–cell interactions and communication from gene expression. *Nature Reviews Genetics* **22**, 71–88 (2021). URL `https://doi.org/10.1038/s41576-020-00292-x`.

[2] Svensson, V., Vento-Tormo, R. & Teichmann, S. A. Exponential scaling of single-cell rna-seq in the past decade. *Nature Protocols* **13**, 599–604 (2018). URL `https://doi.org/10.1038/nprot.2017.149`.

[3] Shree, A., Pavan, M. K. & Zafar, H. scdreamer for atlas-level integration of single-cell datasets using deep generative model paired with adversarial classifier. *Nature Communications* **14**, 7781 (2023). URL `https://doi.org/10.1038/s41467-023-43590-8`.

[4] Camp, J. G. *et al.* Multilineage communication regulates human liver bud development from pluripotency. *Nature* **546**, 533–538 (2017). URL `https://doi.org/10.1038/nature22796`.

[5] 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`.

[6] 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.

[7] Noel, 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019). URL https://www.science.org/doi/abs/10.1126/science.aaw1219. https://www.science.org/doi/pdf/10.1126/science.aaw1219.

[16] Chen, A. *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using dna nanoball-patterned arrays. *Cell* **185**, 1777–1792.e21 (2022). URL https://www.sciencedirect.com/science/article/pii/S0092867422003993.

[17] Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by rna seqfish+. *Nature* **568**, 235–239 (2019). URL https://doi.org/10.1038/s41586-019-1049-y.

[18] 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.

[19] 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.

[20] Fischer, D. S., Schaar, A. C. & Theis, F. J. Modeling intercellular communication in tissues using spatial graphs of cells. *Nature Biotechnology* **41**, 332–336 (2023). URL https://doi.org/10.1038/s41587-022-01467-z.

[21] 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.

[22] 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.

[23] Li, Z., Wang, T., Liu, P. & Huang, Y. Spatialdm for rapid identification of spatially co-expressed ligand–receptor and

revealing cell–cell communication patterns. *Nature Communications* **14**, 3995 (2023). URL https://doi.org/10.1038/s41467-023-39608-w.

[24] 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.

[25] 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.

[26] 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.

[27] Fawkner-Corbett, D. *et al.* Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell* **184**, 810–826.e23 (2021). URL https://www.sciencedirect.com/science/article/pii/S009286742031686X.

[28] 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.

[29] Maynard, K. R. *et al.* Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature Neuroscience* **24**, 425–436 (2021). URL https://doi.org/10.1038/s41593-020-00787-0.

[30] Pham, D. *et al.* Robust mapping of spatiotemporal trajectories and cell–cell interactions in healthy and diseased tissues. *Nature Communications* **14**, 7739 (2023). URL https://doi.org/10.1038/s41467-023-43120-6.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] Heldt, S. A., Stanek, L., Chhatwal, J. P. & Ressler, K. J. Hippocampus-specific deletion of bdnf in adult mice impairs spatial memory and extinction of aversive memories. *Molecular Psychiatry* **12**, 656–670 (2007). URL https://doi.org/10.1038/sj.mp.4001957.

[40] Agarwal, A. *et al.* Dysregulated expression of neuregulin-1 by cortical pyramidal neurons disrupts synaptic plasticity. *Cell Reports* **8**, 1130–1145 (2014). URL https://www.sciencedirect.com/science/article/pii/S221112471400610X.

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] 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/.

[46] 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.

[47] 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.

[48] 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 β-1. *Annals of Surgical Oncology* **5**, 271–278 (1998). URL https://doi.org/10.1007/BF02303785.

[49] 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).

[50] 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.

[51] 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.

[52] 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.

[53] 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.

[54] 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).

[55] 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.

[56] Moustakas, A. & Heldin, C.-H. Mechanisms of tgfβ-induced epithelial–mesenchymal transition. *Journal of Clinical*

Medicine **5** (2016). URL https://www.mdpi.com/2077-0383/5/7/63.

[57] 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.

[58] 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.

[59] 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.

[60] 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.

[61] 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.

[62] Tuglu, M., Matilda, Bostanabad, Y., Saber, Ozyon, G., Dalkiliç, B. & Gurdal, H. The role of dual-specificity phosphatase 1 and protein phosphatase 1 in β2-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.

[63] 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.

[64] 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).

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

[66] 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).

[67] 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).

[68] 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.

[69] 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.

[70] 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.

[71] 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.

[72] 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.

[73] 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.

[74] 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.

[75] 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.

[76] 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.

[77] Li, Z. *et al.* Presence of onco-fetal neighborhoods in hepatocellular carcinoma is associated with relapse and response to immunotherapy. *Nature Cancer* **5**, 167–186 (2024). URL https://doi.org/10.1038/s43018-023-00672-2.

[78] Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Research* **35**, W193–W200 (2007). URL https://doi.org/10.1093/nar/gkm226. https://academic.oup.com/nar/article-pdf/35/suppl_2/W193/16746817/gkm226.pdf.

[79] Wang, Q., Morris, R. J., Bode, A. M. & Zhang, T. Prostaglandin Pathways: Opportunities for Cancer Prevention and Therapy. *Cancer Research* **82**, 949–965 (2022). URL https://doi.org/10.1158/0008-5472.CAN-21-2297. https://aacrjournals.org/cancerres/article-pdf/82/6/949/3187127/949.pdf.

[80] 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.

[81] 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.

[82] 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.

[83] Liu, Z., Sun, D. & Wang, C. Evaluation of cell-cell interaction methods by integrating single-cell rna sequencing data with spatial information. *Genome Biology* **23**, 218 (2022). URL https://doi.org/10.1186/s13059-022-02783-y.

[84] 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.

[85] 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 either spot-resolution spatial transcriptomic data and cell type-annotated scRNA-seq data from the same tissue or single-cell resolution spatial transcriptomic 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, expression of the receptor at the target cell types, 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.

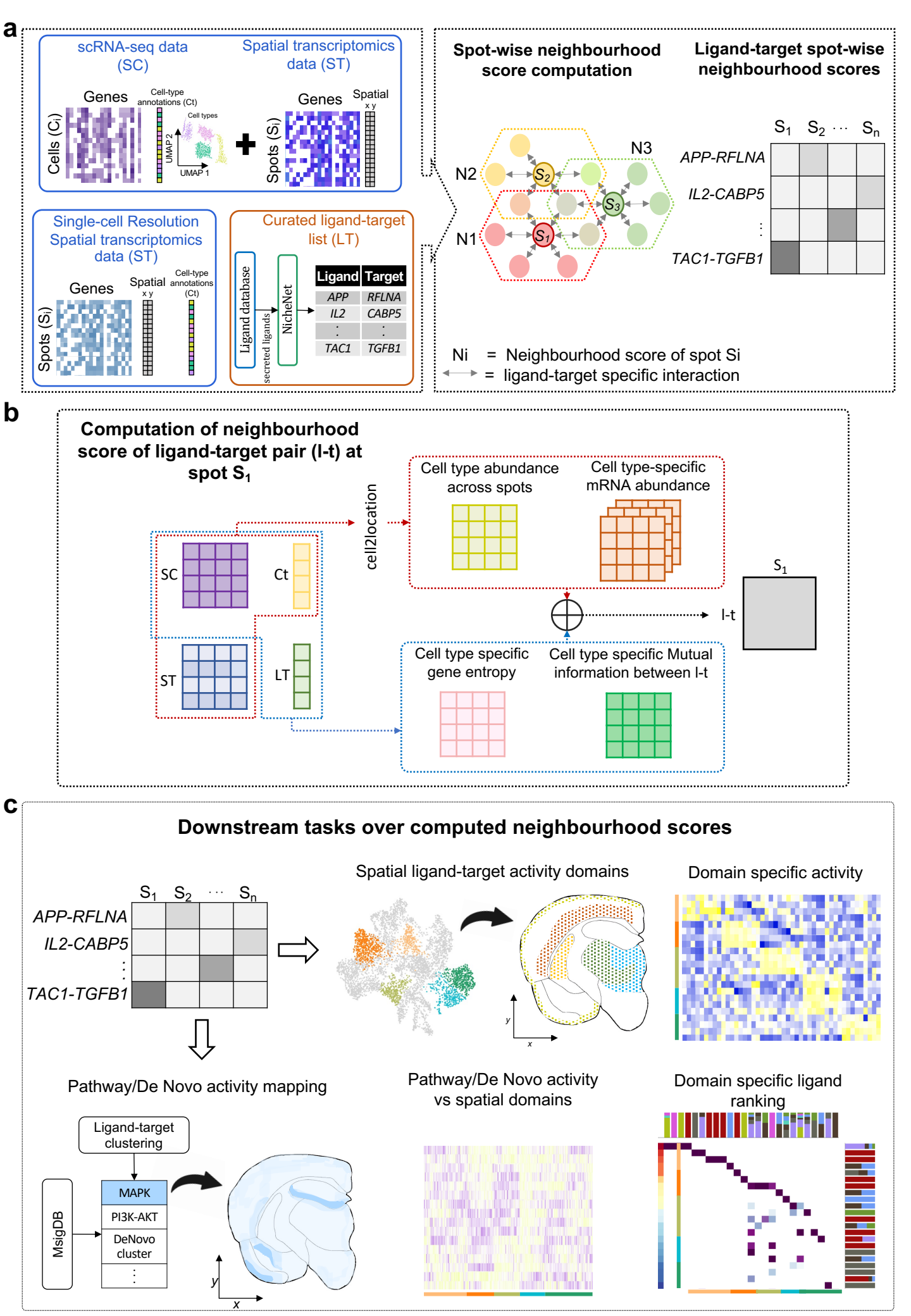
Figure 2: **Renoir outperforms state-of-the-art CCC inference methods in inferring ligand-target activities.** (a) Overview of the simulation strategy used for benchmarking. (b) Comparison of Renoir, SpatialDM, COMMOT, and stLearn for the simulated dataset for the ligand-target pair *A2M-MYC* simulated based on the human intestine dataset. Positive and negative reference spots are shown in the tissue. Regions 1, 2 and 3 contain negative spots for which SpatialDM, COMMOT, and stLearn falesly inferred activity of the pair but Renoir correctly inferred the absence of activity. (c) Comparison of Renoir, SpatialDM, COMMOT, and stLearn in terms of spatial activity accuracy over all the ligand-target pairs across the tissue types. (d) ARI and NMI comparison of the spatial communication domains computed based on the ligand-target activity scores inferred by Renoir, SpatialDM, COMMOT, and stLearn. (e) Qualitative comparison of the spatial communication domains inferred by Renoir, SpatialDM, COMMOT, and stLearn for DLPFC sample 151673.

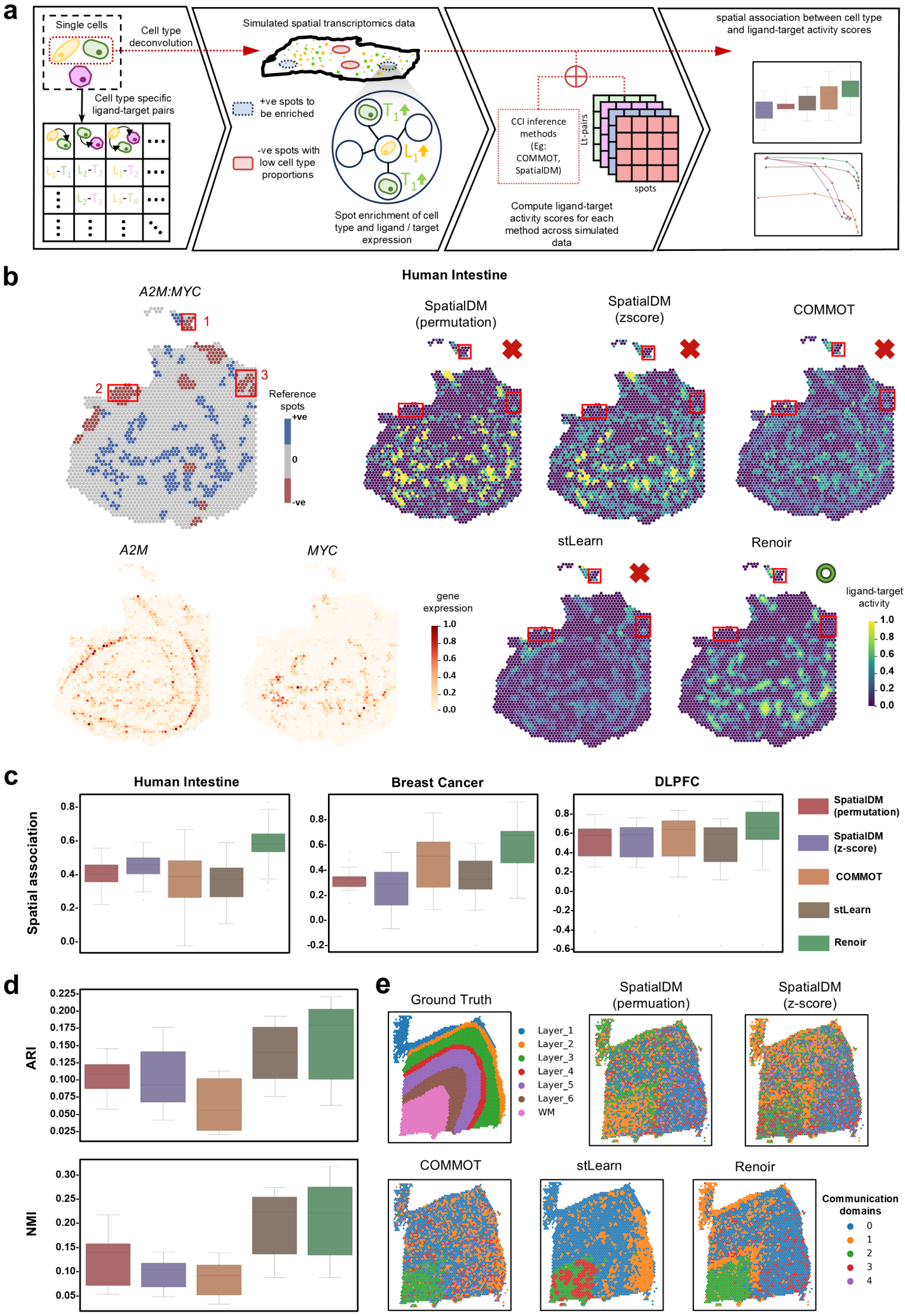
Figure 3: **Renoir identifies distinct communication domains in mouse brain.** (a) Spatial communication domains with distinct latent embeddings inferred by Renoir based on ligand-target neighborhood scores for two mouse brain sections. (b) Distribution of spots in spatial communication domains across 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.

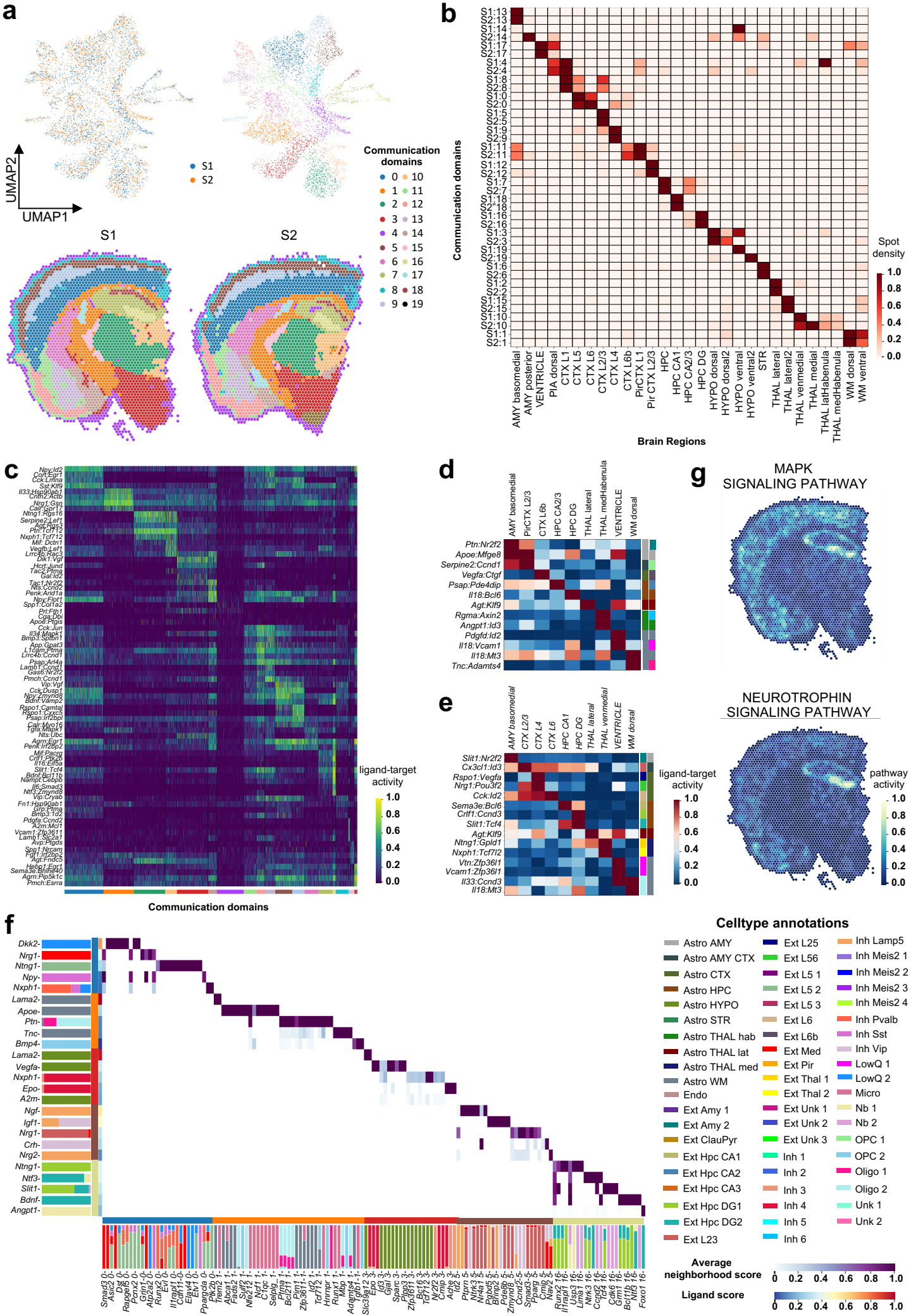
Figure 4: **Renoir uncovers intratumor spatial communication sub-domains in triple negative breast cancer tissue**. (a) The original histopathological annotations of H&E-stained triple negative breast cancer tissue (CID44971 (28), $n = 1322$ spots) showing six pathologically distinct regions. (b) Spatial 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.

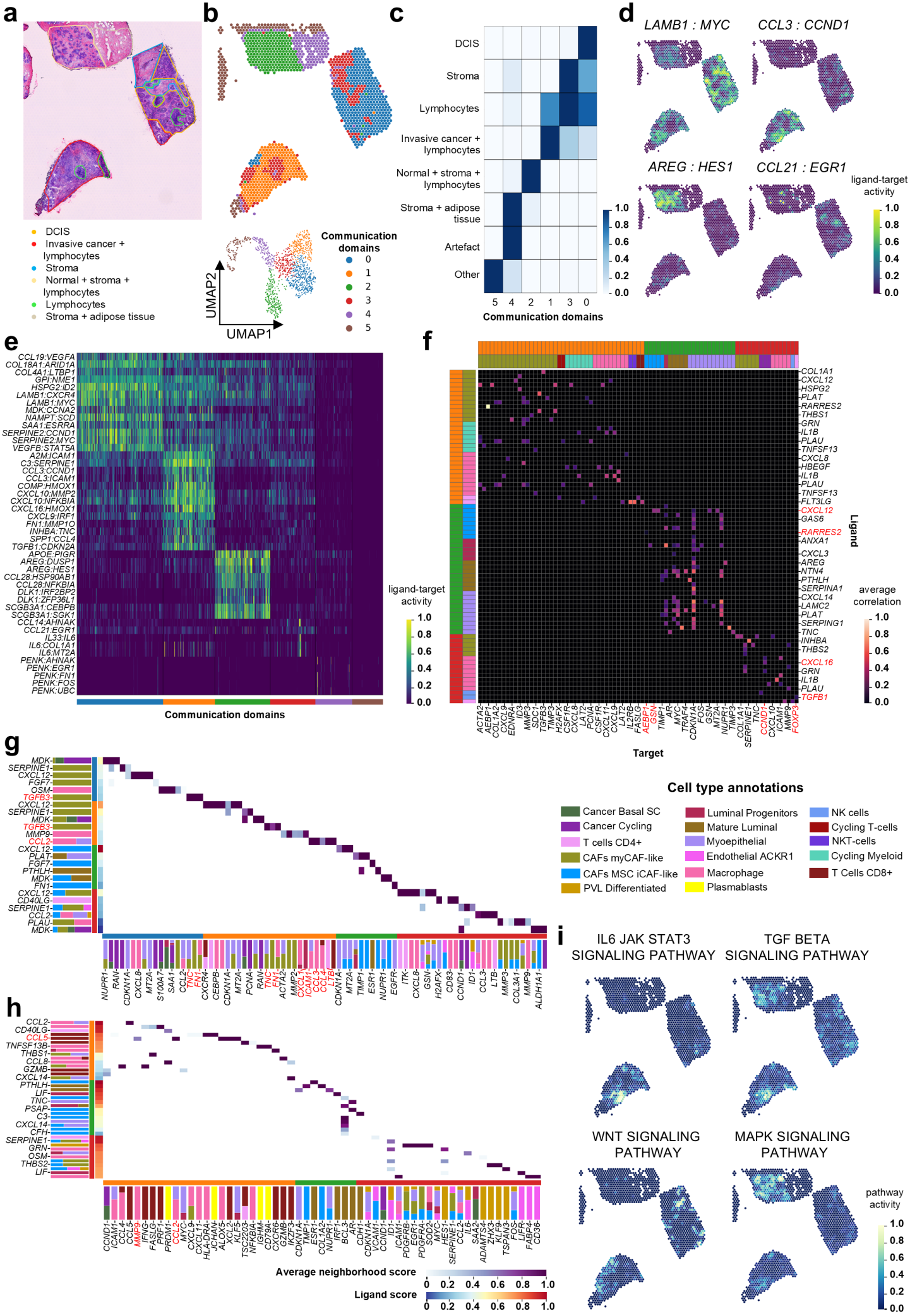
Figure 5: **Renoir identifies hepatocyte-macrophage interactions in developing fetal liver**. (a) Preparation of FFPE fetal liver tissue (b) Spatial communication domains inferred by Renoir based on ligand-target neighborhood scores. (c) Distribution of UMI count across the tissue section. (d) Spatial 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.

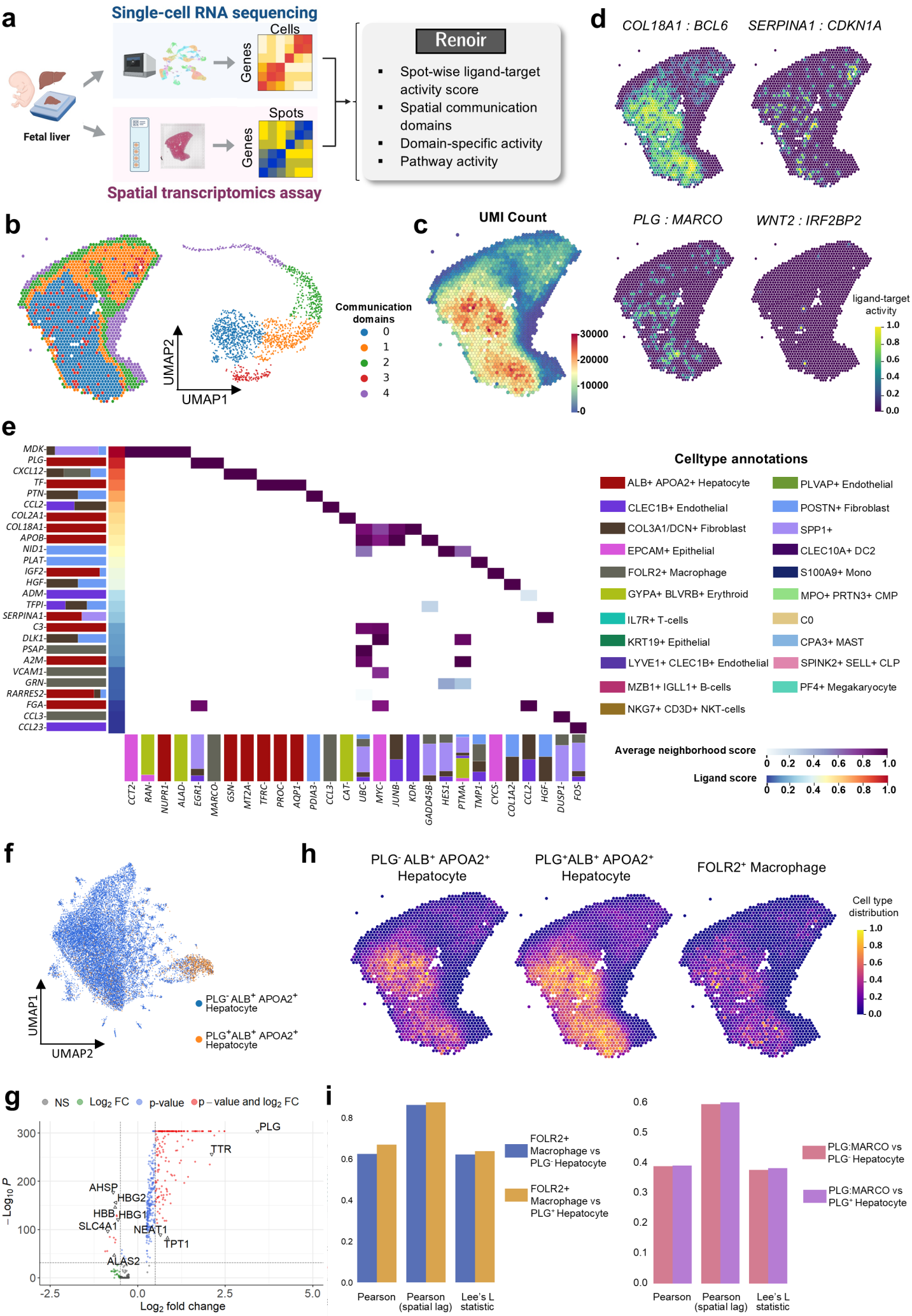
Figure 6: **Renoir identifies onco-fetal interactions in hepatocellular carcinoma**. (a) Ranking of activity of ligands expressed by onco-fetal cell types with other major cell types present in the neighborhoods of the onco-fetal cells. Stacked color bars represent the cell types that express the ligand (left) and target (bottom). (b) (left) Spatial map of cell type specific neighborhood activity scores across cell types associated with the ligand and target for $VEGFA : KDR$ and $CXCL12 : CXCR4$ respectively (fov 10); (right) Spatial distribution of cell types associated with the ligand and target for $VEGFA : KDR$ and $CXCL12 : CXCR4$ respectively (fov 10) (c) Distribution of ligand-target interactions across target cell types. The target cell types include major non-onco fetal cell types present in the neighborhoods of the onco-fetal cell types and onco-fetal cell types (d) Top pathways in the $MUC6^+$ bi-potent cells targeted by the oncofetal ligands (e) (From left to right) Spatial map of cell type specific neighborhood activity scores across cell types associated with $IL6 : PTGS2$; Spatial distribution of cell types associated with $IL6 : PTGS2$.

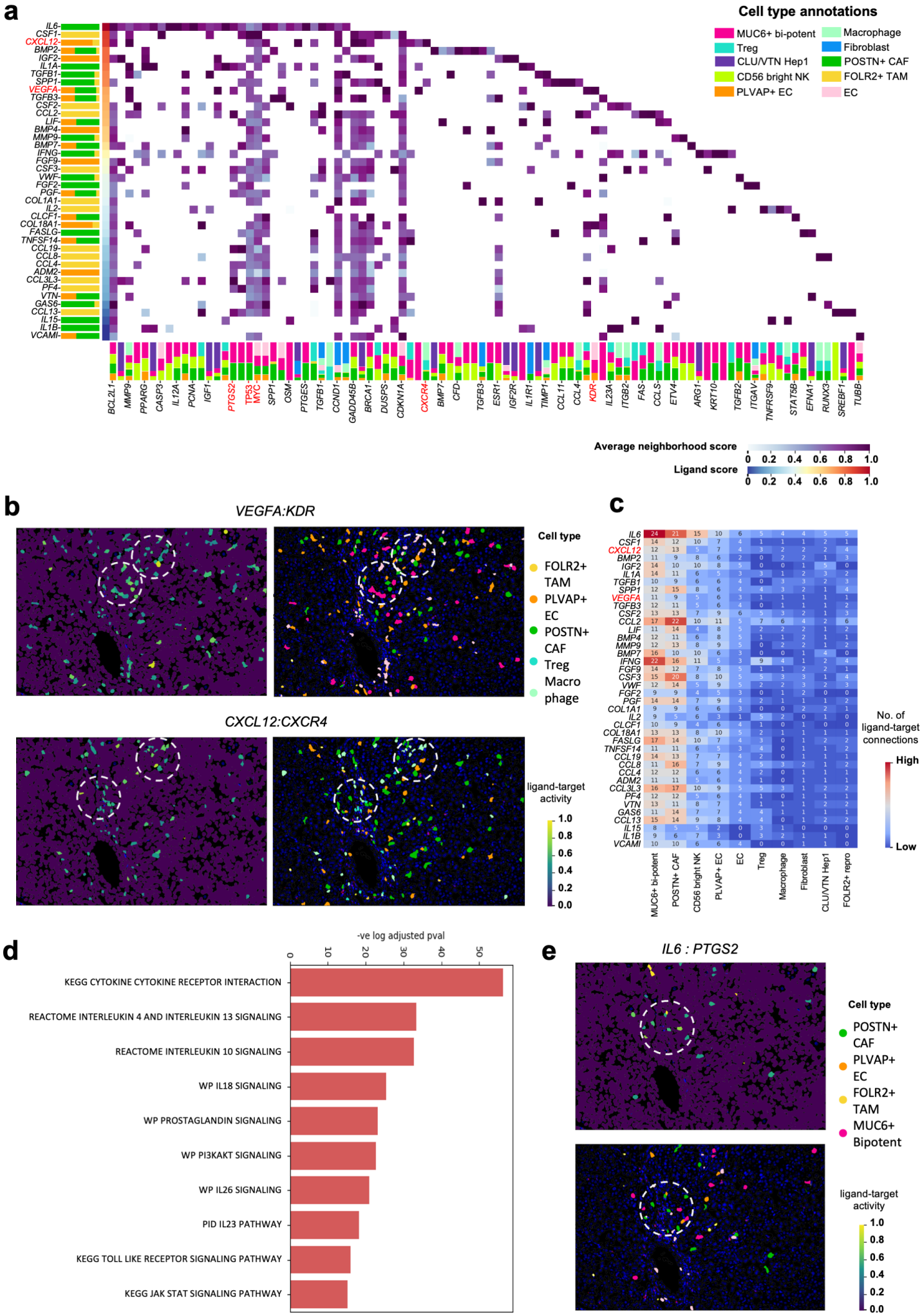

a
scRNA-seq data (SC)
Spatial transcriptomics data (ST)
Genes
Cell-type annotations (Ct)
Cell types
UMAP 2
UMAP 1
Cells (Ci)
Genes
Spatial
x y
Spots (Si)
Single-cell Resolution Spatial transcriptomics data (ST)
Genes
Spatial
x y
Cell-type annotations (Ct)
Spots (Si)
Curated ligand-target list (LT)
Ligand database
secreted ligands
NicheNet
Ligand
Target
APP
RFLNA
IL2
CABP5
TAC1
TGFB1
Spot-wise neighbourhood score computation
N1
N2
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
y
x


a
Single cells
Cell type deconvolution
Cell type specific ligand-target pairs
Simulated spatial transcriptomics data
+ve spots to be enriched
-ve spots with low cell type proportions
Spot enrichment of cell type and ligand / target expression
CCI inference methods (Eg: COMMOT, SpatialDM)
Lt-pairs
spots
Compute ligand-target activity scores for each method across simulated data
spatial association between cell type and ligand-target activity scores
b
Human Intestine
A2M:MYC
Reference spots
+ve
0
-ve
SpatialDM (permutation)
SpatialDM (zscore)
COMMOT
A2M
MYC
gene expression
stLearn
Renoir
ligand-target activity
c
Human Intestine
Breast Cancer
DLPFC
Spatial association
SpatialDM (permutation)
SpatialDM (z-score)
COMMOT
stLearn
Renoir
d
ARI
NMI
e
Ground Truth
Layer_1
Layer_2
Layer_3
Layer_4
Layer_5
Layer_6
WM
SpatialDM (permuation)
SpatialDM (z-score)
COMMOT
stLearn
Renoir
Communication domains
0
1
2
3
4


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 19
b
Communication domains
Brain Regions
Spot density
1.0 0.8 0.6 0.4 0.2 0.0
c
Communication domains
ligand-target activity
1.0 0.8 0.6 0.4 0.2 0.0
d
AMY basomedial
PirCTX L2/3
CTX L6b
HPC CA2/3
HPC DG
THAL lateral
THAL medHabenula
VENTRICLE
WM dorsal
Ptn:Nr2f2
Apoe:Mfge8
Serpine2:Ccnd1
Vegfa:Ctgf
Psap:Pde4dip
Il18:Bcl6
Agt:Klf9
Rgma:Axin2
Angpt1:Id3
Pdgfd:Id2
Il18:Vcam1
Il18:Mt3
Tnc:Adamts4
e
AMY basomedial
CTX L2/3
CTX L4
CTX L6
HPC CA1
HPC DG
THAL lateral
THAL venmedial
VENTRICLE
WM dorsal
Slit1:Nr2f2
Cx3cl1:Id3
Rspo1:Vegfa
Nrg1:Pou3f2
Cck:Id2
Sema3e:Bcl6
Crlf1:Ccnd3
Slit1:Tcf4
Agt:Klf9
Ntng1:Gpld1
Nxph1:Tcf7l2
Vtn:Zfp36l1
Vcam1:Zfp36l1
Il33:Ccnd3
Il18:Mt3
ligand-target activity
1.0 0.8 0.6 0.4 0.2 0.0
g
MAPK SIGNALING PATHWAY
NEUROTROPHIN SIGNALING PATHWAY
pathway activity
1.0 0.8 0.6 0.4 0.2 0.0
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
0 0.2 0.4 0.6 0.8 1.0
Ligand score
0 0.2 0.4 0.6 0.8 1.0


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
1.0
0.8
0.6
0.4
0.2
0.0
d
LAMB1 : MYC
CCL3 : CCND1
AREG : HES1
CCL21 : EGR1
ligand-target activity
1.0
0.8
0.6
0.4
0.2
0.0
e
CCL19:VEGFA
COL18A1:ARID1A
COL4A1:LTBP1
GPI:NME1
HSPG2:ID2
LAMB1:CXCR4
LAMB1:MYC
MDK:CCNA2
NAMPT:SCD
SAA1:ESRRA
SERPINE2:CCND1
SERPINE2:MYC
VEGFB:STAT5A
A2M:ICAM1
C3:SERPINE1
CCL3:CCND1
CCL3:ICAM1
COMP:HMOX1
CXCL10:MMP2
CXCL10:NFKBIA
CXCL16:HMOX1
CXCL9:IRF1
FN1:MMP10
INHBA:TNC
SPP1:CCL4
TGFB1:CDKN2A
APOE:PIGR
AREG:DUSP1
AREG:HES1
CCL28:HSP90AB1
CCL28:NFKBIA
DLK1:IRF2BP2
DLK1:ZFP36L1
SCGB3A1:CEBPB
SCGB3A1:SGK1
CCL14:AHNAK
CCL21:EGR1
IL33:IL6
IL6:COL1A1
IL6:MT2A
PENK:AHNAK
PENK:EGR1
PENK:FOS
PENK:UBC
Communication domains
f
ligand-target activity
1.0
0.8
0.6
0.4
0.2
0.0
average correlation
1.0
0.8
0.6
0.4
0.2
0.0
Ligand
COL1A1
CXCL12
HSPG2
PLAT
RARRES2
THBS1
GRN
IL1B
PLAU
TNFSF13
CXCL8
HBEGF
IL1B
PLAU
TNFSF13
FLT3LG
CXCL12
GAS6
RARRES2
ANXA1
CXCL3
AREG
NTN4
PTHLH
SERPINA1
CXCL14
LAMC2
PLAT
SERPING1
TNC
INHBA
THBS2
CXCL16
GRN
IL1B
PLAU
TGFB1
Target
ACTA2
AEBP1
COL1A2
CXCL9
EDNRA
IL13
MMP3
SDC1
TGFB3
TIMP3
H2AFX
CSF1R
CXCL8
LAT2
RPSA
CSF1R
CXCL11
CXCL9
LAT2
IL2RB
FASLG
AEBP1
GSN
TIMP1
AR
MYC
TRAF4
CDKN1A
FOS
GSN
MT2A
NUPR1
TIMP3
COL3A1
SERPINE1
TNC
CCND1
CXCL10
ICAM1
MMP9
FOXP3
g
MDK-
SERPINE1-
CXCL12-
FGF7-
OSM-
TGFB3-
CXCL12-
SERPINE1-
MDK-
TGFB3-
MMP9-
CCL2-
CXCL12-
PLAT-
FGF7-
PTHLH-
MDK-
FN1-
CXCL12-
CD40LG-
SERPINE1-
CCL2-
PLAU-
MDK-
NUPR1-
RAN-
CDKN1A-
CXCL8-
MT2A-
S100A7-
SAA1-
CCL2-
TNC-
FN1-
CXCR4-
CEBPB-
CDKN1A-
MT2A-
PCNA-
RAN-
TNC-
FN1-
ACTA2-
MMP2-
CXCL1-
ICAM1-
CCL4-
LTB-
CDKN1A-
MT2A-
TIMP1-
ESR1-
NUPR1-
EGFR-
ITK-
CXCL8-
GSN-
H2AFX-
CD83-
CCND1-
ID1-
CCL3-
LTB-
MMP3-
COL3A1-
MMP9-
ALDH1A1-
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
CCL2-
CD40LG-
CCL5-
TNFSF13B-
THBS1-
CCL8-
GZMB-
CXCL14-
PTHLH-
LIF-
TNC-
PSAP-
C3-
CXCL14-
CFH-
SERPINE1-
GRN-
OSM-
THBS2-
LIF-
CCND1-
ICAM1-
CCL4-
CCL5-
MMP9-
IFNG-
FASLG-
PRF1-
PRDM1-
CCL2-
MYC-
CXCL9-
CXCL11-
HLA-DRA-
JCHAN-
ALOX5-
XCL2-
XCL5-
TSC22D3-
NFKBIA-
IGHM-
CD79A-
CXCR6-
GZMB-
IKZF3-
CDKN1A-
TIMP1-
ESR1-
COL1A2-
NUPR1-
IRF1-
BCL3-
AR-
CCL4-
CDKN1A-
VCAM1-
CCND1-
ID1-
ICAM1-
PDGFRB-
EGFR-
PDGFRA-
SOD2-
MYC-
HES1-
SERPINE1-
CCL2-
IL6-
SAA2-
ADAMTS4-
ZHX3-
KLF9-
TSPAN3-
FOS-
LIFR-
FABP4-
CD36-
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
i
IL6 JAK STAT3 SIGNALING PATHWAY
TGF BETA SIGNALING PATHWAY
WNT SIGNALING PATHWAY
MAPK SIGNALING PATHWAY
pathway activity
1.0
0.8
0.6
0.4
0.2
0.0


a
Single-cell RNA sequencing
Cells
Genes
Spots
Genes
Fetal liver
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
MDK
PLG
CXCL12
TF
PTN
CCL2
COL2A1
COL18A1
APOB
NID1
PLAT
IGF2
HGF
ADM
TFPI
SERPINA1
C3
DLK1
PSAP
A2M
VCAM1
GRN
RARRES2
FGA
CCL3
CCL23
CCT2
RAN
NUPR1
ALAD
EGR1
MARCO
GSN
MT2A
TFRC
PROC
AQP1
PDIA3
CCL3
CAT
UBC
MYC
JUNB
KDR
GADD45B
HES1
PTMA
TMP1
CYCS
COL1A2
CCL2
HGF
DUSP1
FOS
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
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


a
Cell type annotations
MUC6+ bi-potent
Treg
CLU/VTN Hep1
CD56 bright NK
PLVAP+ EC
Macrophage
Fibroblast
POSTN+ CAF
FOLR2+ TAM
EC
Average neighborhood score
Ligand score
b
VEGFA:KDR
CXCL12:CXCR4
Cell type
FOLR2+ TAM
PLVAP+ EC
POSTN+ CAF
Treg
Macro phage
ligand-target activity
c
No. of ligand-target connections
High
Low
d
-ve log adjusted pval
KEGG CYTOKINE CYTOKINE RECEPTOR INTERACTION
REACTOME INTERLEUKIN 4 AND INTERLEUKIN 13 SIGNALING
REACTOME INTERLEUKIN 10 SIGNALING
WP IL18 SIGNALING
WP PROSTAGLANDIN SIGNALING
WP PI3KAKT SIGNALING
WP IL26 SIGNALING
PID IL23 PATHWAY
KEGG TOLL LIKE RECEPTOR SIGNALING PATHWAY
KEGG JAK STAT SIGNALING PATHWAY
e
IL6 : PTGS2
Cell type
POSTN+ CAF
PLVAP+ EC
FOLR2+ TAM
MUC6+ Bipotent
ligand-target activity