nature communications

Article

https://doi.org/10.1038/s41467-026-72388-7

Charting spatial ligand-target activity using Renoir

Received: 13 November 2024

Accepted: 8 April 2026

Published online: 05 May 2026

Check for updates

Narein Rao[1], Tanush Kumar[2,16], Dina Kazemi[3,16], Shaozhi Hou[3,4], Atefeh Khakpoor[3,4], Merrin Mary Eapen[3,4], Rhea Pai[5,6], Liang Qiao[7], Archita Mishra[8,9,10], Florent Ginhoux[11,12], Jerry Kok Yen Chan[13], Jacob George[7], Ankur Sharma[3,4,5,6,13] ✉ & Hamim Zafar[1,14,15] ✉

The advancement of single-cell RNA sequencing and spatial transcriptomics has enabled the inference of cellular interactions in a tissue microenvironment. Despite advances in cell-cell interaction inference, methods capable of mapping the influence of ligands on downstream target genes across spatial niches harboring specific cell type composition, crucial for resolving niche-specific relationship between ligands and their downstream targets are still lacking. Here, we present Renoir for charting the ligand-target activities across a spatial topology, delineating spatial communication niches harboring specific ligand-target activities and spatially mapping pathway-level activity of genesets. Across spatial datasets with varying resolution (spot to single-cell) ranging from development to disease, Renoir infers cellular niches with distinct ligand-target interactions, spatially maps pathway activities, and identifies context-specific cell-cell interactions, including hepatocyte-macrophage interactions in fetal liver and interactions between onco-fetal and bipotent cells in hepatocellular carcinoma. Renoir uncovers biological insights and therapeutically-relevant cellular crosstalk from spatial transcriptomics data.

Coordination among the cells drives 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 and developmental trajectories at the resolution of individual cells[2–4] and have shown great potential in investigating CCC[5]. 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[6,7]. These methods, including CellChat[7], CellphoneDB[6], iCELLNET[8], NATMI[9], scTensor[10], and SingleCellSignalR[11] mostly rely on defining functions based on the

[1]Department of Computer Science and Engineering, Indian Institute of Technology Kanpur, Kanpur, India. [2]Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur, India. [3]Translational Genomics Program, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia. [4]School of Clinical Medicine, Faculty of Medicine and Health, University of New South Wales, Kensington, NSW, Australia. [5]Harry Perkins Institute of Medical Research, Nedlands, Perth, WA, Australia. [6]Curtin Medical School, Curtin University, Perth, WA, Australia. [7]Storr Liver Centre, The Westmead Institute for Medical Research and Westmead Hospital, University of Sydney, Sydney, NSW, Australia. [8]Telethon Kids Institute, University of Western Australia, Perth, WA, Australia. [9]School of Medical Sciences, The University of Sydney, Sydney, NSW, Australia. [10]Sydney Infectious Diseases Institute (Sydney ID), Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia. [11]Singapore Immunology Network (SIgN), Agency for Science, Technology and Research (ASTAR), Singapore, Singapore. [12]Gustave Roussy Cancer Campus, Villejuif, France. [13]KK Research Center, KK Women's and Children's Hospital, Singapore, Singapore. [14]Department of Biological Sciences and Bioengineering, Indian Institute of Technology Kanpur, Kanpur, India. [15]Mehta Family Center for Engineering in Medicine, Indian Institute of Technology Kanpur, Kanpur, India. [16]These authors contributed equally: Tanush Kumar, Dina Kazemi.
✉e-mail: ankur.sharma@garvan.org.au; hamim@iitk.ac.in

expression of ligand and receptor pair. Beyond ligand-receptor interactions, methods such as NicheNet[12], Cytotalk[13], SoptSC[14] and CCCExplorer[15] 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[10].

Recent breakthroughs in sequencing[16,17] and imaging-based[18] 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[19]). Moreover, CCC inference methods that specifically use spatial data have also been developed. Methods such as Giotto[20] and NCEM[21] use a spatial neighbor graph for inferring cell-cell interactions; CellPhoneDB v3[22] uses the spatial information to define microenvironments within which interactions are restricted; stLearn[23] and SpatialDM[24] rely on the co-expression of ligand and receptor genes in relation to the spatial location diversity; SVCA[25], MISTy[26] and COMMOT[27] 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 equivalent resolution given a specific spatial context, even after the adoption of a ligand-target database. Robust estimation of ligand-target activity requires consideration of the abundance of sender and receiver cell types at a spatial context and the expression of the cognate receptors by the receiver cell type[28]. Ligand-mediated modulation of target gene expression can only occur if a cognate receptor of the ligand is expressed by the recipient cell. Furthermore, the existing methods employing low-resolution spatial datasets (e.g., 10x Visium) cannot quantify cell type-specific interactions from the heterogeneous mixture of cell types within individual spots[1,28]. This represents a critical limitation, as different signaling pathways exhibit strong cell type-specificity[29]. For example, tumor necrosis factors (TNFs), multifunctional pro-inflammatory cytokines, are known to regulate expression of a large number of genes in a cell-type-specific manner by binding to and activating two distinct receptors: TNF receptor 1 (TNFR1) and TNFR2[30]. Finally, as existing computational methods were primarily developed for low-resolution spatial datasets, their application to emerging single-cell resolution platforms, such as NanoString CosMx and 10x Xenium, may yield suboptimal results. Deciphering the influence of ligands on target gene expressions in receiving cells in a tissue microenvironment with unique cell type composition remains a central challenge, underscoring the need for robust computational approaches capable of elucidating such context-specific signaling dynamics.

To address these challenges, here 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 single-cell resolution spatial transcriptomic data or matched low-resolution spatial transcriptomic data and scRNA-seq data from the same tissue to delineate the spatial activity of candidate ligand-target pairs curated from an existing database[12]. 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 in the spatial transcriptomics data by integrating 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 measures, and a cell type-specific mutual information between the ligand-target pairing as obtained from single-cell resolution spatial data or scRNA-seq 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 enable 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 de novo 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 demonstrate that Renoir outperforms the state-of-the-art methods in inferring spatial activity of ligand-target pairs and communication niches. To demonstrate Renoir's ability to generate spatial maps of ligand-target activities across diverse tissue contexts and spatial transcriptomics platforms, we apply it to datasets spanning development and disease, including adult mouse brain and human triple-negative breast cancer (TNBC) profiled with 10x Visium, as well as human fetal liver profiled with the higher-resolution 10x Visium HD technology, for which Renoir uncovers distinct spatial domains characterized by specific ligand-target activities. Renoir characterizes the interactions of regional astrocyte subtypes with other spatially co-localized cell types in the mouse brain, identifies several spatially restricted interactions between cancer, stromal, and immune cell populations in TNBC, and uncovers hepatocyte-macrophage interactions in developing fetal liver. To further demonstrate Renoir's ability to handle single-cell resolution spatial datasets, we apply it on NanoString CosMx and 10x Xenium datasets from hepatocellular carcinoma, for which Renoir identifies interactions between onco-fetal cells and bipotent stem like cells mediated by interleukins, which we validate using cell line experiments.

Results

Charting spatial ligand-target interactions using Renoir

In order to build a comprehensive computation framework to investigate spatial cell-cell interactions, we developed Renoir. As shown in Fig. 1a, 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[9] and NicheNet[12]. For a low-resolution spatial transcriptomic dataset (e.g., 10x Visium), using a cell type-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

https://doi.org/10.1038/s41467-026-72388-7

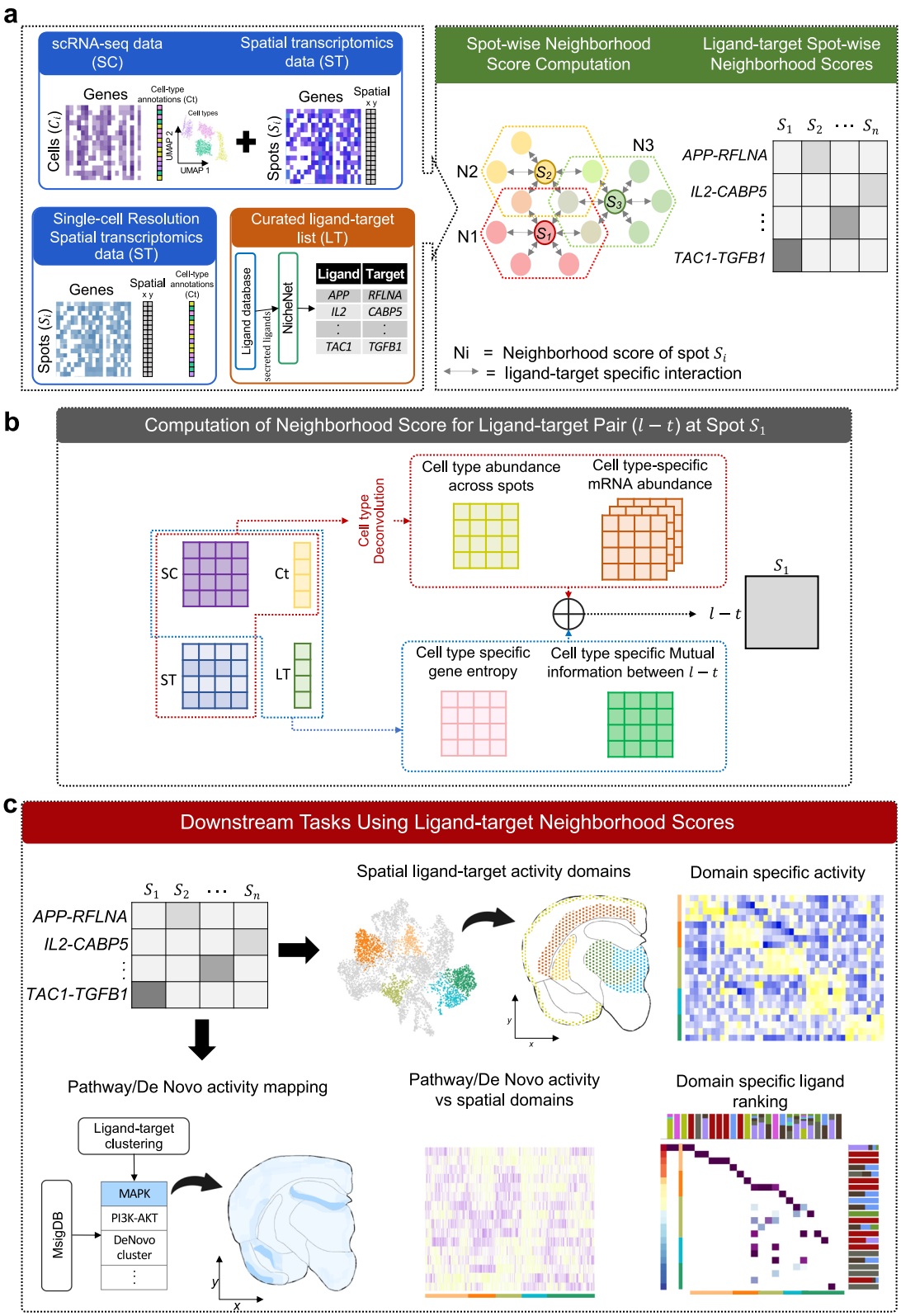
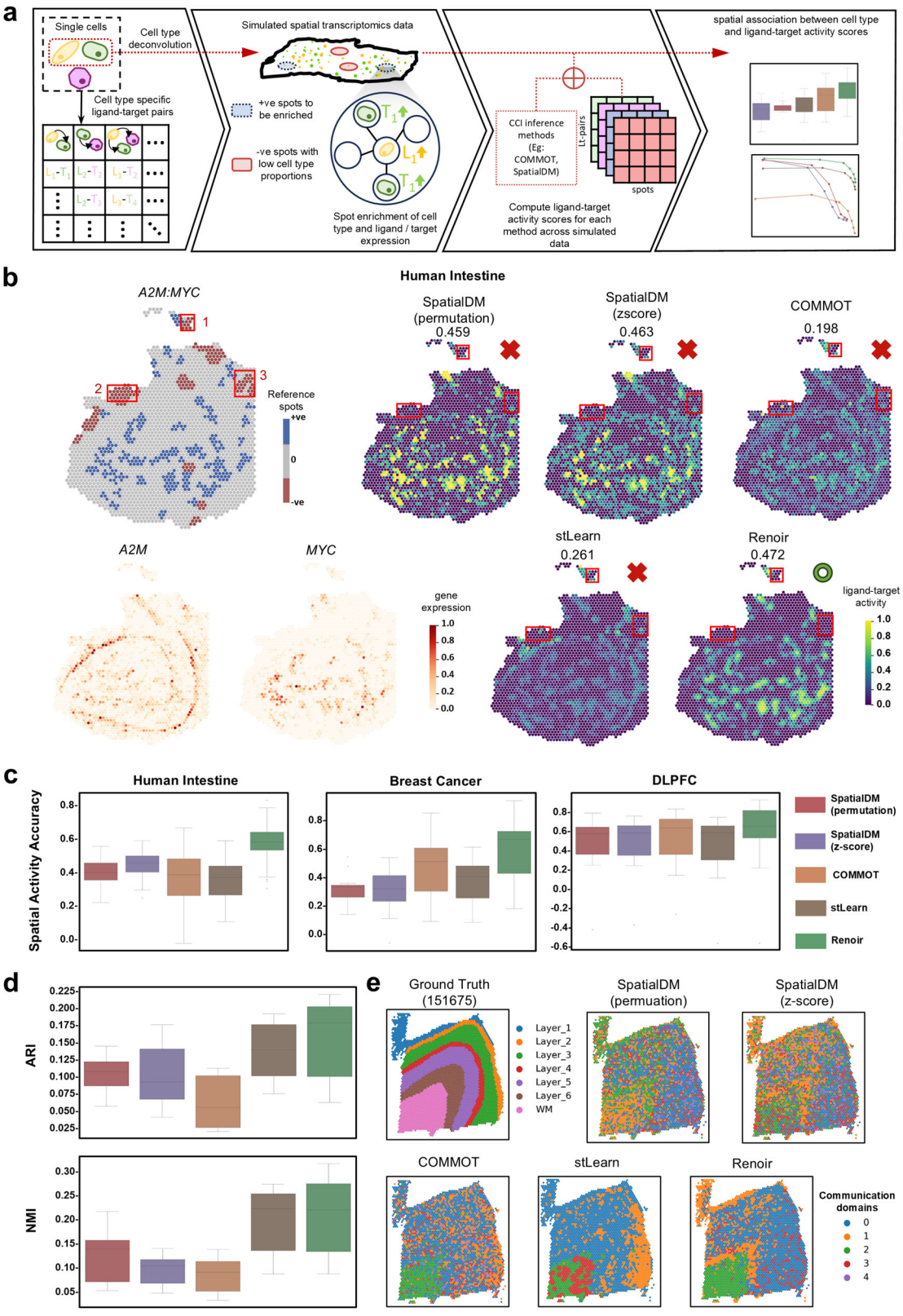


Fig. 1 | Overview of Renoir. a Renoir utilizes either spot-resolution spatial transcriptomic data and cell type-annotated scRNA-seq data from the same tissue or single-cell resolution spatial transcriptomic data to infer spatial neighborhood activity scores for a set of curated ligand-target pairs at each spot. **b** To estimate the neighborhood activity score for a ligand-target pair $l-t$ at a given spot S_1; 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.

https://doi.org/10.1038/s41467-026-72388-7



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 cell type 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 (Fig. 1b). Thus, Renoir generates a neighborhood activity score matrix (of dimensions number of ligand-target pairs × number of spots). Using the ligand-target neighborhood activity score matrix, Renoir performs several downstream analyses (Fig. 1c). First, Renoir infers spatial communication domains through unsupervised

Fig. 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. For each method, the numerical value above the spatial plot denotes the spatial activity accuracy of the method for the given ligand-target pair. **c** Comparison of Renoir, SpatialDM, COMMOT, and stLearn in terms of spatial activity accuracy across all ligand-target pairs for three tissue types: Human Intestine ($n = 43$), Breast Cancer ($n = 15$), and DLPFC ($n = 26$). The box denotes the interquartile range (IQR, the range between the 25th and 75th percentile) with the median value, whiskers indicate the maximum and minimum value within 1.5 times the IQR. **d** ARI and NMI comparison of the spatial communication domains computed across DLPFC samples ($n = 12$) based on the ligand-target activity scores inferred by Renoir, SpatialDM, COMMOT, and stLearn. The box denotes the interquartile range (IQR, the range between the 25th and 75th percentile) with the median value, whiskers indicate the maximum and minimum value within 1.5 times the IQR. **e** Qualitative comparison of the spatial communication domains inferred by Renoir, SpatialDM, COMMOT, and stLearn for DLPFC sample 151675. Source data are provided as a Source Data file.

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.

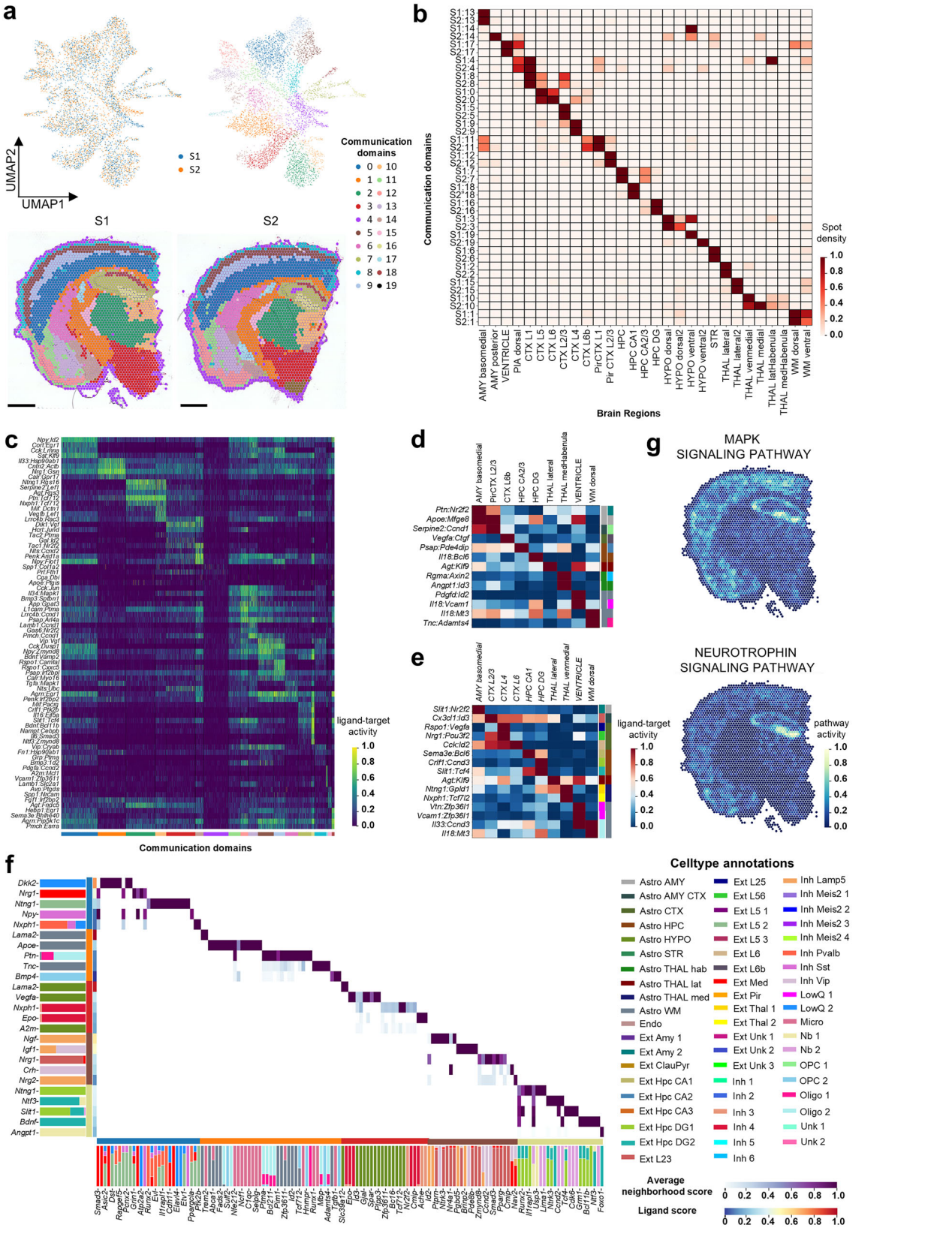
Renoir outperforms state-of-the-art methods in inferring ligand-target interactions

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, based on paired single-cell and spatial datasets, 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[31], triple negative breast cancer (TNBC)[32] and dorsolateral prefrontal cortex (DLPFC)[33] for simulation. The simulation workflow as described in Fig. 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[27], stLearn[34] and SpatialDM[24]. 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 (Supplementary Table 1). 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 quantified by a method with the true simulated activity of the ligand-target pair in the positive and negative reference spatial locations (see Methods 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.

To ensure that Renoir can capture the activities of the ligand-target pairs involved in canonical signaling pathways where the ligand and target genes can be expressed by multiple cell types, we

https://doi.org/10.1038/s41467-026-72388-7



performed another benchmarking based on validated pathways from CellChatDB[7] and used our enrichment-based simulation workflow to generate 111, 71, and 81 simulated datasets corresponding to the human intestine, TNBC, and DLPFC tissue types, respectively. For each dataset, we selected pathways for which at least one ligand, receptor, and target gene were expressed in the paired scRNA-seq data. For these simulated datasets, interactions were simulated spatially by enriching all the cell types associated with the ligand and target genes in the positive reference spots. A qualitative comparison of the methods for candidate pathways (Supplementary Fig. 2a-c) for the three tissue types, respectively, demonstrates how Renoir was more accurate in capturing the ligand-target activities in the positive spots

Fig. 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. Black scale bars, 1 mm. **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 five 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. Source data are provided as a Source Data file.

and the lack of activity in the negative spots as compared to the other methods, which often reported false-positive activities. Quantitatively, Renoir outperformed other CCC inference methods in terms of spatial activity accuracy (63.99, 94.72, and 83.98% improvement across the human intestine, TNBC, and DLPFC datasets) (Supplementary Fig. 3a-c) and average precision and recall across different threshold values (Supplementary Fig. 3d-f) across the tissue types.

We further simulated datasets (using pathways from CellChatDB) where, instead of enrichment of ligand and target cell types, spatial activity was simulated by directly enriching ligand and target expression values in the positive spots. The negative spots were determined on the basis of the lack of ligand and target gene expression. For these datasets also, Renoir outperformed the other methods by achieving higher average precision and recall across different threshold values (Supplementary Fig. 4a-c) across the tissue types.

Since the quality of ST data can be affected by low UMI count, we evaluated the robustness of Renoir to variations in UMI count by performing a downsampling experiment where the raw total UMI counts of the spots were downsampled to predefined quantiles of their original distribution. We simulated multiple downsampled datasets for varying degrees of data sparsity (90%, 75%, and 60%) across the three tissue types (TNBC, human intestine, and DLPFC). Renoir was able to maintain high spatial activity accuracy for the ligand-target pairs even for downsampled datasets with 60% of the original UMI count (Supplementary Fig. 5a). To further evaluate Renoir's tolerance to errors in cell type annotations, we performed an additional experiment with the TNBC dataset where we artificially shuffled single-cell annotations at varying degrees of perturbation, and the shuffled single-cell annotations were used for computing Renoir's ligand-target activity scores. Renoir was able to maintain stable performance across varying levels of cell label perturbation (10–30%), indicating its robustness against moderate level of noise in cell type annotation (Supplementary Fig. 5b).

Finally, we evaluated the accuracy of the spatial communication 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[33], in which the authors manually annotated six cortical layers and white matter for each sample based on cytoarchitecture and selected marker genes, and also showed that specific signaling pathways exhibit layer-specific activity in DLPFC. Considering the manual annotations as ground truth, we evaluated the domains inferred using Renoir's ligand-target scores with respect to how well they align with known biological structures 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.

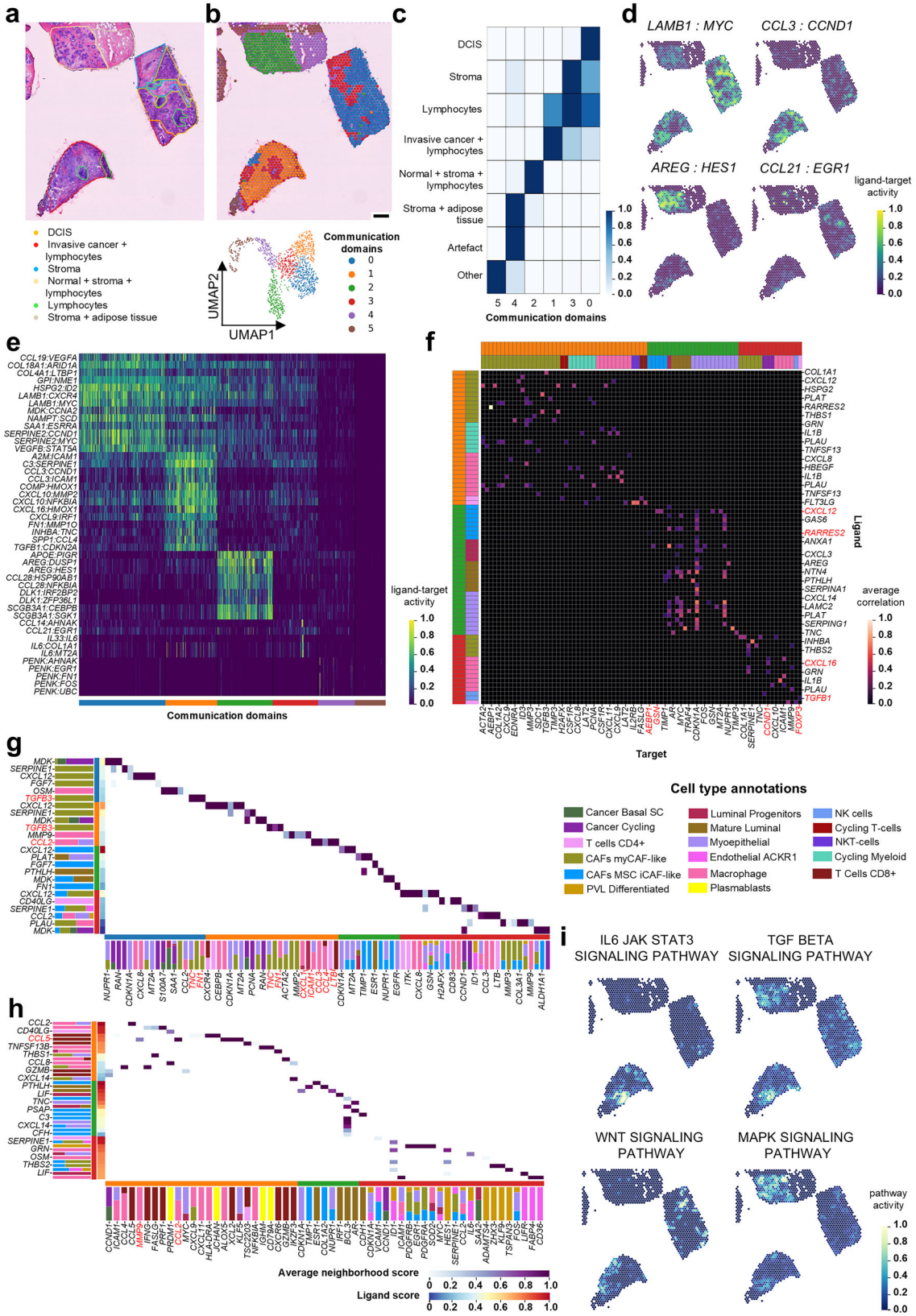
Renoir identifies distinct communication domains in mouse brain

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[35] (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 20 spatial communication domains which had distinct latent space embeddings (Fig. 3a) and were similarly distributed across the sections (Fig. 3a). 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. 3b).

The communication domains were found to harbor diverse cell types that were spatially localized and abundance pattern concordant across the replicates (Supplementary Fig. 6a-b). The domains further harbored differential activity of specific ligand-target pairs (Fig. 3c, Supplementary Fig. 7b-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[36,37], and domain 3, specific to the hypothalamus, exhibited *Dlk1* activity known to be expressed in hypothalamic arginine vasopressin and oxytocin neurons[38]. 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[39,40], 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[41,42]. Supplementary Table 2 lists the domain-specific ligand-target interactions with literature support for the associated brain region.

Kleshchevnikov et al.[35] 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 Fig. 7d 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 Fig. 7e 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*

https://doi.org/10.1038/s41467-026-72388-7



expressed by diencephalic astrocytes (Thalamus) was 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. 7f) that are

Fig. 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[32], $n = 1322$ 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. Black scale bar, 500 µm. **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** Unique ligand-target pairs within each domain are represented. 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. Source data are provided as a Source Data file.

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[43] displayed high activity within domain 16 associated with the hippocampus. *Nrg1*, known to have high impact on cortical functions[44] exhibited high activity within domain 5. Supplementary Table 3 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. 7g and Supplementary Fig. 8). Consistent with prior works[45,46], 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[32]. 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. 4b), were primarily associated with morphologically distinct pathological regions (Fig. 4c). 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. 9a). Since two (clusters 4 and 5) of the six domains had low overall UMI count (Supplementary Fig. 9b), 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. 9a); cancer-associated fibroblasts (CAFs) and different T cell populations (Supplementary Fig. 9a); and demonstrated differential activity of *MYC* and *SAA1* (*LAMB1:MYC*, *SERPINE2:MYC*, *SAA1:ESRRA*) known to be overexpressed in basal-subtype of breast cancer[47–49]; and MDK (*MDK:CCNA2*), a known cancer progression marker[50]. Domain 0 also harbored high activity of *VEGF* (*VEGFA*, *VEGFB*) (known to be highly up-regulated in breast cancer[51–53]), 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. 9a). Consequently, interactions associated with tumor progression and invasion (e.g., *CXCL10:MMP2*, *CCL3:ICAM1*[54], *FN1:MMP10*[55]), as well as suppression (e.g., *CXCL9:IRF1*[56]) 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. 9a). Interestingly, this domain harbored high activity of the ligand *CCL28* which is known to promote breast cancer and metastasis[57], target *HES1* which plays a crucial role in malignant cell proliferation[58] and *AREG* which is expressed by myoepithelial cells and loss of which can inhibit tumor proliferation, growth and invasiveness[59] (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. 10a) 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*[60] 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[30,61], (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)[62] and it was found to target *ICAM1* (a molecular target for TNBC)[54] 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[63]) which activated *MMP9* and *CCL2* in macrophages.

Since this dataset contained cell type labels at three different hierarchical levels (lv1, lv2 and lv3), we leveraged this to evaluate Renoir's robustness to varying levels of cell type annotation granularity. We applied Renoir using each level of annotation to compute ligand-target activity scores. Renoir maintained consistent performance across all levels of annotation. The inferred spatial communication domains across different levels of annotation were highly similar (Supplementary Fig. 11a) and demonstrated similar distributions of pathological regions (Supplementary Fig. 11b), indicating Renoir's robustness to the granularity of cell type annotations. Furthermore, we compared the differential ligand-target activities across the spatial communication domains inferred using different levels of cell type annotations. The differential ligand-target activities for each communication domain were highly correlated across different levels of cell type annotations (Supplementary Fig. 11c).

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

https://doi.org/10.1038/s41467-026-72388-7

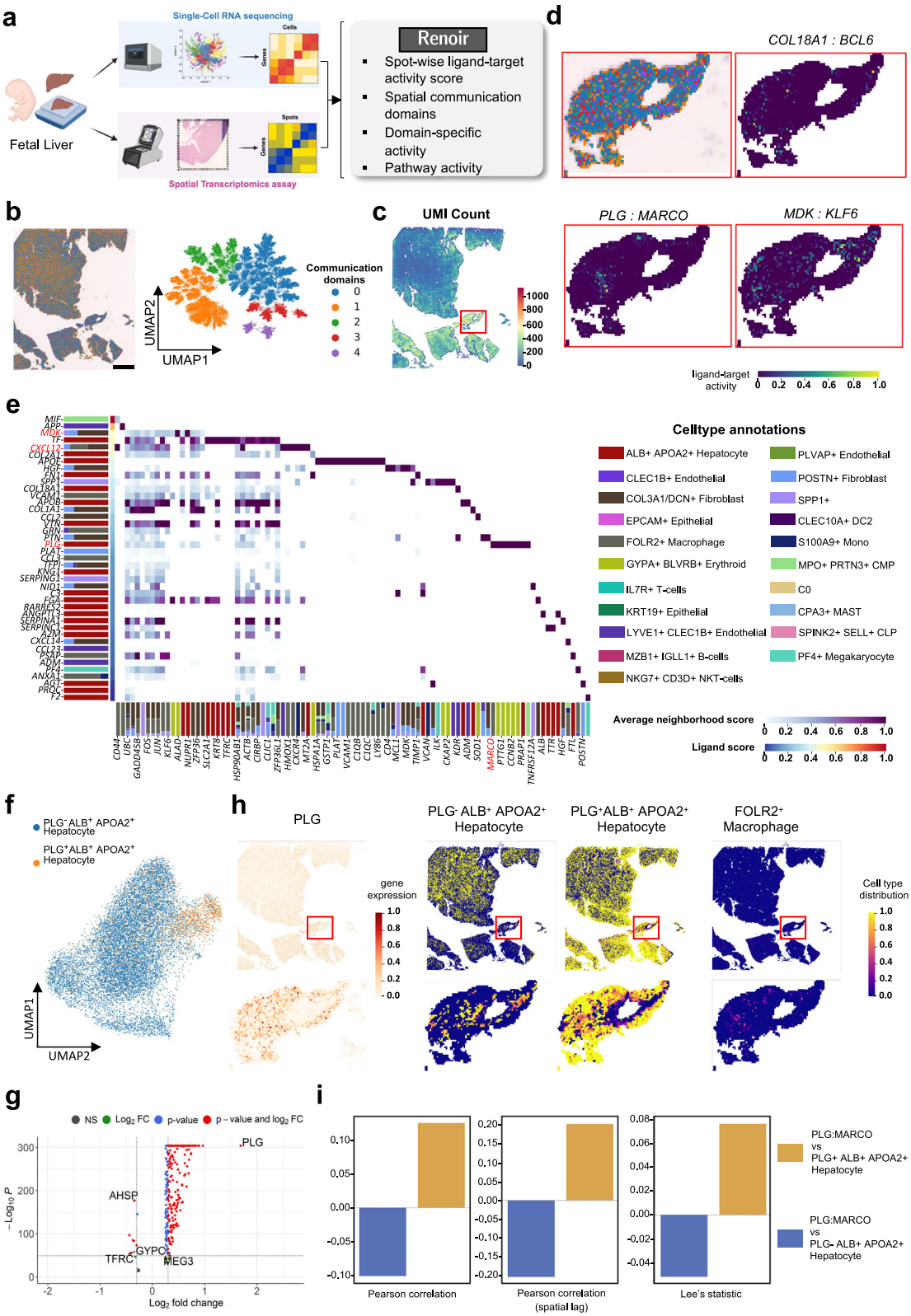


Fig. 5 | Renoir identifies hepatocyte-macrophage interactions in developing fetal liver. a Preparation of FFPE fetal liver tissue. Created in BioRender. Zafar, H. (2026) (https://BioRender.com/gmj8e7g) **b** Spatial communication domains inferred by Renoir based on ligand-target neighborhood scores obtained from the 10x Genomics Visium HD dataset (sample D1). Black scale bar, 1 mm. **c** Distribution of UMI count across the tissue section. **d** Spatial map of communication domains and neighborhood activity scores for differentially active ligand-target pairs (*COL18A1:BCL6*, *PLG:MARCO*, and *MDK:KLF6*). **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** Volcano plot depicting differentially expressed genes between PLG^+ and PLG^- Hepatocytes calculated using non-parametric Wilcoxon rank sum test ($-\log_{10}$P threshold = 50 and $\log_2$foldchange threshold = 0.3). **h** Spatial feature plot of *PLG* expression and cell type abundances of Hepatocytes (PLG^+ and PLG^-) and $FOLR2^+$ Macrophages overlaid onto the spots in ST data. **i** Spatial similarity measures between *PLG:MARCO* neighborhood activity scores and cell type abundances of PLG^+ and PLG^- Hepatocytes across neighborhoods containing $FOLR2^+$ Macrophages (n = 39651 spots). Source data are provided as a Source Data file.

abundance and entailed distinct ligand-target activities (Supplementary Figs. 12-15). Amongst the six subdomains in domain 0 (Supplementary Fig. 12), in subdomain 4 *TGFB1* and *FBN1* activated several EMT-associated genes (*FBN1*, *MMP2*, *COL5A1*, *PDGFRB*, etc.)[64] and cancer progression and regulation markers (*DUSP1*, *JUNB*, and *MYC*)[65,66] 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. 13), 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[67–70]. Similarly, subdomains of other domains showcased interesting activity particular to the cell type abundant across the subdomains (Supplementary Figs. 14–15).

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. 10b). 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[71]. 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 Genomics Visium HD spatial transcriptomic dataset (see Methods for details) from a fetal liver sample (D1) for which single-cell RNA sequencing data was available[72] (Fig. 5a) and applied Renoir on data binned at 16-μm resolution. Renoir identified 16577 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. 16a). Domain 0 exhibited ligand-target activity predominantly in regions with high UMI counts. In contrast, domain 1 showed sparse activity in low-UMI regions (Fig. 5c), whereas domains 2, 3, and 4 displayed moderate ligand-target activity with specificity toward distinct subsets of target genes. While hepatocytes constituted the majority of cells in all domains (Supplementary Fig. 16b), macrophages and fibroblasts were consistently observed across domains. Activities pertinent to fetal liver were found to be widespread (Fig. 5d and Supplementary Fig. 16a). For example, *COL18A1* associated with type XVIII collagen (a prominent component of the Extracellular Matrix (ECM) in the liver[73]) was found to be highly expressed across the tissue and activate multiple targets including hematopoietic transcription factors *BCL2L1* and *BCL6*[74]. Higher *CDKN1A* activity was observed, in line with the proliferative nature of hepatocytes in developing fetal liver tissue.

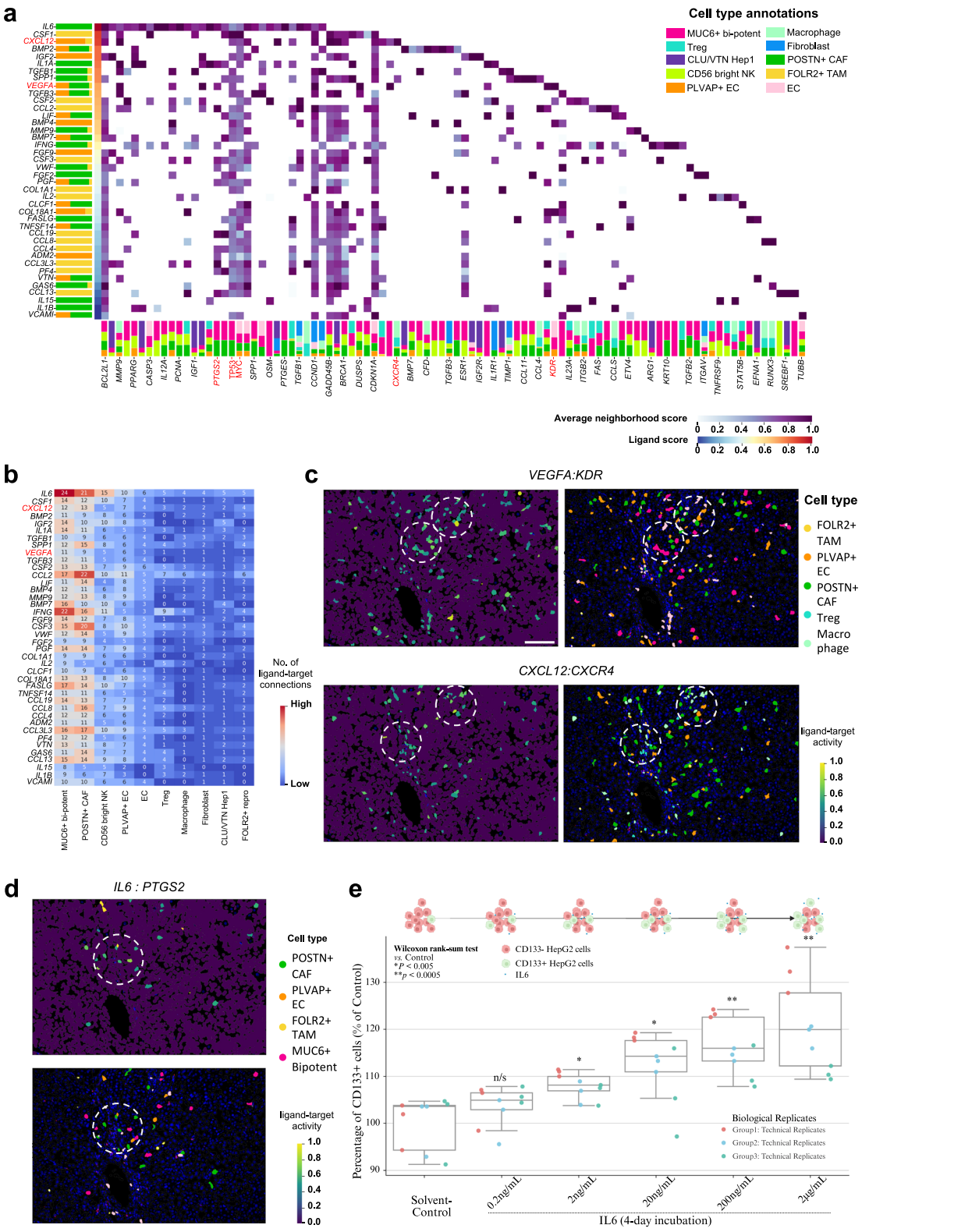
We further utilized Renoir to perform ligand rankings for each communication domain. Specifically, we observed prominent activity of the ligands *MDK*, an important growth factor for fetal liver hematopoietic stem cells and multipotent progenitor expansion[75] and *CXCL12* which helps in retaining hematopoietic stem cells in the fetal liver[76] (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[77] and fetal-like reprogramming of liver macrophages plays an important role in immunosuppressive phenotype in liver cancer[72]. As shown in Fig. 5e, we observed that *MARCO*$^+$ macrophages interact with *PLG*$^+$/*APOB*$^+$ subset of hepatocytes. The *PLG* (plasminogen) has been known to be important for hepatic regeneration[78,79] and *PLG* deficiency can lead to impaired remodelling in mouse models after acute injury. Next, we subdivided hepatocytes into two groups: *ALB*$^+$*APOA*$^+$*PLG*$^+$ and *ALB*$^+$*APOA*$^+$*PLG*$^-$ (Figs. 5f-g). As shown in Figs. 5g-h and Supplementary Fig. 16c, these two subsets are transcriptionally and spatially distinct from each other indicating heterogeneity of the fetal liver hepatocyte population. We observed that the majority of *ALB*$^+$*APOA*$^+$ hepatocytes (in the matched single-cell dataset) lack *PLG* expression, whereas a distinct subset exhibits high *PLG* expression (Fig. 5f). In the Visium HD dataset, the spots harboring *ALB*$^+$*APOA*$^+$*PLG*$^+$ hepatocytes also contained *FOLR2*$^+$ macrophages (Fig. 5h). *FOLR2*$^+$ macrophages are embryonic macrophages present in the fetal liver microenvironment[72]. We performed spatial correlation analyses using three different metrics (Pearson correlation, Pearson correlation with spatial lag adjustment, and Lee's L statistic) between the cell type abundance of *ALB*$^+$*APOA*$^+$*PLG*$^+$ hepatocytes/*PLG*$^-$ hepatocytes and the *PLG:MARCO* activity scores inferred by Renoir (see Methods for details). *PLG:MARCO* activity showed stronger correlation with *ALB*$^+$*APOA*$^+$*PLG*$^+$ hepatocytes than with *PLG*$^-$ hepatocytes in all neighborhoods containing *FOLR2*$^+$ macrophages across all three spatial correlation measures (Fig. 5i). To further validate this interaction, we generated and analyzed an additional 10x Visium HD fetal liver dataset (sample A1). Consistent with our previous findings, for the additional fetal liver dataset also, similar communication domains were inferred, domain-specific ligand-target activities were observed (Supplementary Fig. 17), and *PLG:MARCO* was inferred by Renoir to be an important ligand-target interaction with better spatial correlation with *PLG*$^+$ hepatocytes as compared to the *PLG*$^-$ hepatocytes (Supplementary Fig. 18a-c). Taken together, these results provide important insights into the macrophage-hepatocyte niche in the developing human liver.

Renoir identifies onco-fetal interactions in hepatocellular carcinoma

To evaluate Renoir's ability in handling single-cell resolution ST datasets, we applied it on a Nanostring CosMx dataset generated from a hepatocellular carcinoma (HCC) patient[80]. The dataset consisted of 70166 cells distributed across 46 cell types and 959 genes.

We recently[72,80] discovered the presence of onco-fetal tumor microenvironment (TME) in hepatocellular carcinoma, where the *FOLR2*$^+$ TAMs, *PLVAP*$^+$ ECs and *POSTN*$^+$ CAFs were identified as onco-fetal cell types that shared transcriptional programs with their fetal liver counterparts. We evaluated Renoir's performance in identifying the onco-fetal interactions from the single molecule imaging (CosMx) dataset. To perform an unbiased analysis, we used Renoir to quantify the activity of the ligands expressed by the onco-fetal cell types with the other major cell types present in the neighborhoods of the onco-fetal cells (Supplementary Fig. 19a) using Renoir's ligand-ranking formulation (Fig. 6a). Renoir was able to characterize the activity of several known core onco-fetal ligands (e.g., *VEGFA*, *TGFB3*) as well as ligands shared by two onco-fetal cell types (e.g., *BMP7*, *CXCL12*, *TGFB1*, *IFNG*) (Supplementary Fig. 19b)[80]. 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

https://doi.org/10.1038/s41467-026-72388-7



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). To further support our findings from the Nanostring CosMx dataset, we employed 10x Xenium (10x Genomics) technology, developed a new panel with known genes for onco-fetal biology and applied this on two HCC samples (Methods) followed by application of Renoir for quantifying the activity of the onco-fetal ligands. The Xenium datasets also annotated three different onco-fetal cell types: OF Macrophage, OF Endothelial and OF CAF. From the Xenium datasets also, Renoir identified the activity of different onco-fetal ligands (e.g., *VEGFA*, *TGFB3*, *CXCL12*, *TGFB1*, *IFNG*) (Supplementary Figs. 20a, 21a), many of which were shared with the CosMx datasets (Supplementary Fig. 19e). Particularly, the known onco-fetal

Fig. 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 for the CosMx hepatocellular carcinoma (HCC) dataset. Stacked color bars represent the cell types that express the ligand (left) and target (bottom). **b** 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 **c** (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). White scale bar, 125 µm. **d** (top) Spatial map of cell type specific neighborhood activity scores across cell types associated with the ligand and target for *IL6:PTGS2* (fov 10); (bottom) Spatial distribution of cell types associated with the ligand and target for *IL6:PTGS2* (fov 10) **e** IL6 induced bipotent stem cell-like properties in HepG2 cells. Treatment with IL6 induced CD133$^+$ population in a dose-dependent manner. The percentage of CD133$^+$ HepG2 cells was measured by flow cytometry after the HepG2 cells were treated with IL6 protein for increasing concentration for 96 h. All data presented are from three independent experiments containing three technical replicates ($n = 9$). p-values (*$p < 0.005$ vs control groups; **$p < 0.0005$ vs control groups) are calculated using two-sided Wilcoxon rank-sum test. p-values: 0.2ng = 0.051938, 2ng = 0.001987, 20 ng = 0.002666, 200 ng = 0.000409, 2 µg = 0.000409. The box denotes the interquartile range (IQR, the range between the 25th and 75th percentile) with the median value, whiskers indicate the maximum and minimum value within 1.5 times the IQR. Individual datapoints are overlaid on the boxplots. Created in BioRender. Hou, S. (2026) (https://BioRender.com/7pacpt3). Source data are provided as a Source Data file.

interactions *CXCL12*:*CXCR4* and *VEGFA*:*KDR* were detected by Renoir in multiple spatial neighborhoods where these onco-fetal cells colocalized (Supplementary Figs. 20b-c, 21b).

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

Renoir not only validated known onco-fetal interactions but also identified ligand-target interactions between onco-fetal niche and putative stem-like *MUC6*$^+$ cells in liver cancer microenvironments. From the CosMx data, 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. 19c). 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[81] 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 (Supplementary Fig. 19d) to be enriched. Since prostaglandin signaling is known to promote HCC progression[82], we further investigated 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 onco-fetal cells indicating the potential role of onco-fetal cells in the activation of *PTGS2* in the *MUC6*$^+$ bipotent cells (Fig. 6d). To experimentally validate the interaction between the onco-fetal niche and *MUC6*$^+$ bipotent cells, we focused on the activity of *IL6* as it was predicted by Renoir to be a key ligand from the onco-fetal niche which interacts with *MUC6*$^+$ bipotent cells. This bipotent population is known to express *CD133*[83,84] and has been shown to be involved in tumor initiation[85]. We treated HepG2 cells with increasing concentrations of IL6 (ligand predicted by Renoir). As shown in Fig. 6e and Supplementary Fig. 22a-c, addition of IL6 induced CD133$^+$ population in a dose-dependent manner indicating that the onco-fetal niche can impact bipotent cells as predicted by Renoir. For both Xenium datasets also, Renoir identified the bipotent cells as having the highest number of interactions with onco-fetal cells (Supplementary Figs. 20d, 21c), even though they were not the most abundant cell type within the onco-fetal neighborhood (Supplementary Figs. 20e, 21d). Furthermore for both Xenium samples, pathway analysis of the target genes in bipotent cells activated by onco-fetal ligands identified enrichment of similar pathways as observed in the CosMx dataset (interleukin signaling pathways, prostaglandin signaling) (Supplementary Figs. 20f, 21e). Importantly, these observations indicate that application of Renoir could lead to the identification of biological insights in the tumor microenvironment.

Discussion

Cell-cell interactions play a crucial role in various biological processes, including tissue development, homeostasis, and disease progression. 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. However, estimation of the effects of extracellular ligand signals on downstream target gene expression within spatially organized tissue microenvironments remains a major challenge. Existing computational approaches have primarily focused on inferring ligand-receptor interactions, often neglecting the downstream transcriptional consequences of these interactions, particularly at a spatial context with unique composition of interacting cell types.

In this regard, our computational method, Renoir, offers a powerful tool for investigating cellular interactions within their native spatial context, from development to disease. Renoir fundamentally advances spatial CCC analysis by enabling 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 based on validated pathways from real datasets across diverse tissue types demonstrated its superiority over existing CCC inference methods in quantifying the ligand-target activities across spatial locations. Specifically, by accounting for the expression of cognate receptors in the receiving cell type harboring target gene, 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. The robustness of Renoir to noisy cell type annotation noise and data sparsity further highlights its utility across a range of experimental conditions and spatial transcriptomics platforms.

Our application of Renoir to 10x Visium and Visium HD 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 and technologies with varying spatial resolution. 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 regional astrocyte subtypes with other co-localized excitatory and inhibitory neurons highlighting the complex architecture of neuro-glial communication. For the triple negative breast cancer, Renoir uncovered spatial communication domains and subdomains 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. In developing fetal liver, Renoir uncovered interaction between hepatocytes and macrophages mediated via the ligand *PLG*. The application of Renoir on single-molecule imaging (Nanostring CosMx and 10x Xenium) hepatocellular carcinoma datasets further shows Renoir's utility in handling single-cell resolution large-scale spatial datasets. In hepatocellular carcinoma, Renoir recovered canonical onco-fetal interactions involving endothelial cells, macrophages and CAFs while also revealing potential reprogramming of bipotent hepatocyte-like cells mediated by onco-fetal niche-derived ligands. Importantly, experimental validation of IL6-mediated expansion of $CD133^+$ HepG2 cells further underscored the functional relevance of Renoir's predictions.

Collectively, these findings demonstrate that Renoir offers a comprehensive and generalizable framework for understanding the dynamic relationships between different spatially co-localized cell types and their roles in various biological processes. Renoir's powerful 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. Renoir's modeling framework also offers a promising foundation for several important extensions. Future developments could include more elaborate modeling of receptor abundance and longitudinal dynamics of ligand-receptor-target network into the inference framework for modeling dose-dependent signaling. Graph neural networks can be utilized for modeling continuous interaction strengths based on ligand abundance, receptor expression, spatial proximity and binding affinity, and such an approach will enable a more quantitative representation of signaling dynamics and better reflect graded communication responses across heterogeneous cellular contexts. Furthermore, Renoir can be extended to jointly analyze datasets across multiple conditions to allow for systematic comparisons of communication landscapes across different stages of development and disease progression. Finally, future work can integrate CRISPR-based or pharmacological perturbation experiments for further validation of predicted signaling interactions. As spatial transcriptomic technologies continue to mature, approaches such as Renoir will be critical for translating these rich datasets into mechanistic insights, offering new avenues for understanding complex biological systems and developing novel therapeutic strategies targeting cellular interactions.

Methods

Ethical statement

All human tissue samples used in this study were obtained in accordance with relevant ethical regulations. Fetal liver tissues were collected under approval from the Centralised Institutional Review Board of the Singapore Health Services (SingHealth) and the National Healthcare Group Research Ethics Committees, Singapore. Written informed consent for the use of fetal tissues was obtained from all participants, in accordance with internationally recognized guidelines including the principles outlined in the Polkinghorne Report 1989. Formalin-fixed paraffin-embedded (FFPE) hepatocellular carcinoma (HCC) tumor resection blocks were collected for two patients from the Westmead Institute for Medical Research under approval from the University of New South Wales (UNSW) Human Research Ethics Committee (Ethics ID: iRECS6671), with written informed consent obtained from all participants.

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 is 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. In case the user does not have access to a well-annotated single-cell sequencing data, Renoir's software package provides a preprocessing pipeline for identifying major cell clusters from single-cell data using which spot deconvolution can be performed. This preprocessing uses either scVI (for single-batch data)[86] or scDREAMER (for multi-batch data)[3] for cellular embedding inference followed by Leiden clustering using Scanpy[87].

Acquiring curated ligand-target pairs. We first curate a set of secreted ligands from NATMI's connectomeDB2020[9] 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[12]. 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, $\mathcal{S}_{lt}$ 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,

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

where $\mathcal{N}_s$ 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,

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

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 s_1 and s_2 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

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

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

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

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

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), $MI(lc, tc)$ denotes the mutual information between the expression of l and the expression of t across cell type c. $H(lc)$ denotes the Shannon entropy of the expression of ligand l across cell type c and $H(tc)$ denotes the Shannon entropy of the expression of target t across cell type c. $MI(lc_1, tc_1)$, $MI(lc_2, tc_2)$, $H(lc_1)$, $H(lc_2)$, $H(tc_1)$ 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[35] that takes as input the ST data X^{st} and reference cell type signatures ($\mathcal{G}^C : C \times G$) 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:

$$U_{S\times C\times G} = \{u_{scg}\} = W * \mathcal{G}^C * M \quad (5)$$

where * represents matrix multiplication without sum-reduction. For high-resolution datasets with larger number of spots (e.g., 10x Visium HD), if cell2location fails to perform deconvolution, we use RCTD[88] to acquire $W_{S\times C}$. $U_{S\times C\times G}$ is then computed as $W^T_{S\times C} * X^{st}$. 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,

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

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

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

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

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

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, $\mathcal{A}_{l,t}$ of ligand-target pairs $(l,t) \in \mathcal{S}_{lt}$ and based on these ligand-target activity score vectors, spots are clustered using the scanpy library[87]. 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 ≥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[9] 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[12]. 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[89,90], 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 ref. 91. 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[35]. 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. The simulation follows one of two possible scenarios - ligand and target are marker genes of cell types, ligand and target genes are expressed by multiple cell types

Simulation of data using ligand and target marker genes:

- 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.
- 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.
- 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.

Simulation of data using ligand and target genes expressed by multiple cell types

- Using the pathway database of CellChatDB V2[7], a reference database of validated interactions, we first acquire a set of pathways such that atleast one ligand-receptor pair for the pathway is expressed in the paired scRNA-seq data. We then get a list of all possible ligand-receptor pairs across all selected pathways that are present in the scRNA-seq data. Using the existing list of ligand-target pairs and ligand-receptor pairs, we associate every ligand, receptor and target to a cell type if it is expressed in at least 10% of cells for the given cell type ensuring that the ligand and target genes are associated with multiple cell types. Ligand-target pairs are considered if the ligand and target are associated to at least one cell type each and there exists at least one ligand-receptor pair such that the ligand and receptor are also expressed in the cell types associated to the ligand and target respectively. This generates a set of ligand-target pairs LT_{pw} such that both the ligand and target genes are members of validated pathways and expressed by more than one cell type in the paired scRNA-seq data.
- For a specific ligand-target pair $lt \in LT_{pw}$, the ligand is expressed in a subset of cell types CT_l, $ct_l \in CT_l$ and target celltype in a subset of cell types CT_t, $ct_t \in CT_t$. In the ST dataset, we generate two types of reference spots, positive and negative spots respectively. Positive reference spots harbor the interaction between the ligand and target genes. We simulated two sets of datasets by following two different strategies for positive reference spots.
  - Scenario 1: Positive reference spots denote the spots within the tissue where ligand-target interactions for lt are simulated through the enrichment of subset of CT_l ($s(CT_l)$) and subset of target cell types ($s(CT_t)$), 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. Each of the target cell types $s(CT_t)$ enriched is ensured to have at least one receptor associated with ligand l expressed in the corresponding target cell type.
  - Scenario 2: Positive reference spots reflect enrichment of ligand-target expression, only the ligand and target expression values at the positive reference spots are increased equivalent to that of the expression in n_e single cells. The cell type abundances are not enriched.

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 two types of spots. 1) the spots within the tissue that do not harbor any activity for lt as indicated by the low proportion of ligand cell types CT_l and target cell types CT_t. Specifically, negative reference spots are selected as a set of spots for which both ligand cell type and target cell type 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. 2) Spots enriched with co-localization of $s(CT_l)$ and $s'(CT_t)$ (say $\widehat{s}$) where $s'(CT_t)$ refers to those target cell types where no known receptor of ligand l is expressed. For Scenario 2 of positive spot enrichment, negative reference spots are determined as spots (and their neighbors) devoid of $s(CT_t)$ and reflect the absence of ligand and target expression.

Downsampling and cell annotation perturbation experiments. We performed downsampling experiments with the simulated datasets. The datasets for the downsampling experiment were generated such that the total UMI count of the downsampled dataset equals x% of the total UMI count of the original dataset. This was performed using Scanpy's 'downsample_counts_function', with the 'total_count' parameter set to x% of the total count of the original dataset.

We further performed perturbation of single-cell annotation using the breast cancer dataset used in simulation. Shuffled single-cell annotations were generated at varying degrees of perturbation over the simulated Breast Cancer dataset. Here, for a given perturbation level x%, x% of cells labeled as a cell type 'ct' were reassigned to randomly selected alternative cell types. The newly annotated single-cell data (with x% cell labels perturbed) is then used as input to Renoir for the computation of ligand-target activities.

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,

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

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

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 lower 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

Formalin-fixed paraffin-embedded (FFPE) fetal liver tissues were obtained under ethical approval from the Centralised Institutional Research Board of the Singapore Health Services (SingHealth) and National Health Care Group Research Ethics Committees, with written informed consent obtained from all participants (mothers) in accordance with internationally recognized guidelines. To determine the quality of RNA in FFPE tissue, RNA was extracted from 10 µm 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, FFPE human fetal liver tissue blocks were sectioned at 5 µm thickness using a microtome and the sections were mounted onto positively charged glass slides according to 10x Genomics Visium HD sample preparation guidelines and stored at 4 °C with desiccant until further processing. Tissue sections were deparaffinized, stained with hematoxylin and eosin (H&E), imaged at 20 × magnification using a Zeiss Axioscan 7 prior to probe hybridization, and subjected to probe hybridization using the Visium Human Transcriptome Probe Set v2.1.0. Following hybridization and templated ligation, probe-derived products were transferred onto 10x Genomics Visium HD Spatial Gene Expression Slides (slide serial: H1-GVJMQ3Q) using the CytAssist instrument according to the manufacturer's protocol. Subsequent library preparation steps were performed following the Visium HD workflow. Library construction included Spatial transcriptomic libraries which were generated using the 10x Genomics Visium HD Spatial Gene Expression platform (SPATIAL-HD-v1 chemistry). Imaging data were used for tissue detection and nucleus segmentation during downstream analysis. Gene expression profiling was performed using the Visium Human Transcriptome Probe Set v2.1.0 (GRCh38-2024-A reference). Probes were hybridized to target transcripts in situ and ligated using a templated ligation approach. Following probe ligation, captured molecules were released and subjected to library construction including amplification and indexing according to the Visium HD workflow. Libraries were sequenced on an Illumina NovaSeq 6000 using paired-end 150 bp sequencing. During downstream processing, read 2 was trimmed to 50 bp in Space Ranger Count v4.0.1. Raw sequencing data were processed using Space Ranger Count v4.0.1 with probe filtering enabled, intronic reads excluded, UMI registration applied, and nucleus segmentation performed during image processing. The FFPE fetal liver tissue samples used in this study are stored at KK Research Center, KK Women's and Children's Hospital, Singapore.

Integration of mouse brain datasets

The two mouse brain samples were integrated using the stLearn workflow[23]. 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[92], 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

In order to assess the relationship between *PLG:MARCO* activity scores and the abundance of relevant cell types- PLG^+ and PLG^- hepatocytes and $FOLR2^+$ Macrophages, we first selected the spots that harbored the $FOLR2^+$ Macrophages and these spots along with their spatial neighborhoods were retained for downstream comparison. This filtering ensures that the analysis focuses on regions where the cell types underlying the *PLG:MARCO* interaction are present, thereby improving the biological interpretability of the scores. For these selected spots, we examined Renoir's inferred *PLG:MARCO* interaction scores and compared them across subsets of PLG^+ vs PLG^- hepatocytes, in conjunction with $FOLR2^+$ Macrophages, which are known to express the *MARCO* target. The correspondence between interaction scores and hepatocyte abundance was then quantified using the following metrics:

- *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[93]. Pearson correlation over spatial lag considers the sum of expression or abundance values over a spot and its corresponding neighbors before applying Pearson correlation. Formally, Pearson correlation in the context of expression or abundance values can be defined as,

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

X and Y are the spatial lags defined by,

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

where $N(x_i)$ represents the neighborhood of x_i. $X = x_i$, $i \in S$; $S_x = \{s_j, (X_{s_j} > 0) | (Y_{s_j} > 0)\}$ The spatial lags are calculated with respect to gene expressions from the spatial transcriptomics data or the spatial map of cell type abundances (x or y). While considering neighborhood scores, the scores are considered as they are since the neighborhood score at each spot takes into account the neighbors of the spot.

- *Lee's L statistic*: We utilised Lee's L statistic bivariate spatial association measure to incorporate spatial autocorrelation alongside spatial patterning between the two variables of interest[93]. 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,

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

Preparation of xenium in situ hepatocellular carcinoma datasets

HCC samples from two male patients (IDs: M293 and M316), aged 45-65 years, were used for 10x Xenium In Situ spatial transcriptomics experiments. The tumour tissues were preserved as Formalin-Fixed Paraffin-Embedded (FFPE) blocks, which were collected from the Westmead Institute for Medical Research (University of New South Wales (UNSW) Human Research Ethics and Compliances, Ethics ID: iRECS6671), with written informed consent from all participants. Sex of the participants was determined based on self reporting by the participants at the time of sample collection. The tissue samples were sectioned at 5 μm thickness and placed within the Sample Area (10.45 × 22.45 mm) of a Xenium slide according to the CG000578 Xenium In Situ for FFPE—Tissue Preparation Guide (10x Genomics). Deparaffinization of the tissue sections was performed, followed by rehydration of the tissues and decrosslinking, according to the manufacturer's protocol (CG000580 Xenium In Situ for FFPE—Deparaffinization & Decrosslinking, 10x Genomics). This was followed by overnight probe hybridization at 50 °C using a 319-gene panel (Xenium Standalone Custom Probes) according to the user guide—CG000749 Xenium In Situ Gene Expression with Cell Segmentation Staining (10x Genomics). Following post-hybridization wash, ligation and rolling circle amplification were performed. Next, cell segmentation staining was carried out overnight using the Xenium Multi-Tissue Stain Mix. Finally, autofluorescence quenching and nuclei staining were performed. The processed Xenium slide was loaded onto the Xenium Analyzer (10x Genomics) along with the required buffers and decoding modules, according to the CG000584 Xenium Analyzer user guide. The FFPE blocks of the samples used for Xenium In Situ experiments are currently stored at the Westmead Institute for Medical Research.

Preprocessing of xenium in situ hepatocellular carcinoma datasets

Standard preprocessing steps were applied to raw Xenium outputs using Scanpy[87] package in Python. This included removing low quality cells, defined as those within the bottom and top 2% of total transcript counts, to exclude outliers. Subsequently, a single-cell RNA reference dataset, consisting of 36 hepatocellular carcinoma (HCC) patients[80], was used for annotating cell types in the Xenium data. RCTD[88], a reference-based computational tool for cell type decomposition in spatial transcriptomic data from single cell reference datasets, was employed for annotation. Following the reference dataset's structure, RCTD annotations were performed at two levels: level 1 comprising main cell types with lower granularity, and level 2, where myeloids, fibroblasts, and endothelial cells were further annotated to identify oncofetal subtypes. Annotation results from RCTD were integrated into the h5ad file generated during the preprocessing steps. In order to determine the most appropriate radius for Renoir, and to ensure consistency with CosMx analysis, a neighbourhood size was selected to include an average of 10 to 20 spots. This corresponded to a neighbourhood radius of 30 microns. Next, Renoir was run on the processed Xenium outputs obtained from the above steps. To visualize cell segmentation masks as part of Renoir's output, the spatialdata and spatialdata-plot Python packages[94] were used to overlay segmentation information onto the spatial plots generated by Renoir.

IL6 incubation and analysis of bipotent cell marker CD133 expression

HepG2 cells line (ATCC, HB-8065) was cultured in DMEM (gibco, REF: 11995065) with 10% FBS (Garvan Institute of Medical Research) and 1% Penicillin Streptomycin (gibco, REF: 15070063). IL6 (gibcon, REF: PHC0063) stock was 1mg/ml dissolved in 100mM acetic acid (Sigma, REF: A6283-100ML). Day1, 3×10^5 HepG2 cells were seeded into 6-well plate; Day 2, cells were treated with designed groups as solvent control and IL6-experiment groups with multiple concentrations (0.2ng/ml, 2ng/ml, 20ng/ml, 200ng/

ml and 2 μg/ml) and same medium with/without IL6 was changed on day 3. Day 5, CD133 (BD Biosciences, REF: 567907) expression was measured by flow cytometry (BECKMAN COULTER, CYTOFLEX-11901398), meanwhile Propidium Iodide (Sigma, REF: 537059-50MG) was utilized to isolate live/dead cells. Flow cytometry data was analyzed using FlowJo software.

Sex and gender considerations

Sex and gender were not incorporated into the study design, as the primary objective of this study was to develop and validate a general-purpose computational framework, Renoir for the analysis of spatial transcriptomics datasets. Accordingly, the method is designed to be broadly applicable across datasets and does not include sex- or gender-specific modeling assumptions.

Statistics and reproducibility

No statistical method was used to predetermine sample size. All simulations and analyses were performed using fixed random seeds to ensure reproducibility. No data were excluded from the analyses. The analyses were conducted using automated computational pipelines; therefore, randomization and blinding were not applicable. Relevant simulations were repeated with independent random seeds to confirm robustness, and consistent results were obtained. Statistical tests used for comparing groups are detailed in the corresponding figure legends and Methods subsections.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

10x Genomics Visium HD datasets (both raw and processed) for the fetal liver samples generated in this study have been deposited in the GEO database under the accession code GSE319617. The 10x Genomics Xenium datasets for hepatocellular carcinoma generated for this manuscript are available at https://figshare.com/s/2c3c30dd4e1b51f875ae. The original publicly available datasets that were 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) (samples CID44971 and CID4290); Triple Negative Breast Cancer raw scRNA-seq data: downloaded from the Gene Expression Omnibus under accession number GSE176078; Human Intestinal development spatially resolved transcriptomics data: downloaded from the Gene Expression Omnibus under accession number GSE158328 (sample A2); Human Intestinal development raw scRNA-seq data: downloaded from the Gene Expression Omnibus under accession number GSE15870; Human Dorsolateral Prefrontal Cortex spatially resolved transcriptomic data and raw scRNA-seq data: downloaded from the spatialLIBD project (https://research.libd.org/spatialLIBD/). The scRNA-seq data for fetal liver was obtained from the Gene Expression Omnibus under accession number GSE156337. The Nanostring CosMx data for hepatocellular carcinoma has been obtained from https://doi.org/10.6084/m9.figshare.23972991. Source data are provided with this paper.

Code availability

Renoir has been implemented in Python and is freely available at https://github.com/Zafar-Lab/Renoir under the MIT license and is also available on Zenodo[95]. All analyses and results presented in the manuscript are available at https://github.com/Zafar-Lab/Renoir-reproducibility.

References

1. Armingol, E., Officer, A., Harismendy, O. & Lewis, N. E. Deciphering cell–cell interactions and communication from gene expression. *Nat. Rev. Genet.* **22**, 71–88 (2021).
2. Svensson, V., Vento-Tormo, R. & Teichmann, S. A. Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* **13**, 599–604 (2018).
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. *Nat. Commun.* **14**, 7781 (2023).
4. Pandey, K. & Zafar, H. Inference of cell state transitions and cell fate plasticity from single-cell with MARGARET. *Nucleic Acids Res.* **50**, e86–e86 (2022).
5. Camp, J. G. et al. Multilineage communication regulates human liver bud development from pluripotency. *Nature* **546**, 533–538 (2017).
6. 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. *Nat. Protoc.* **15**, 1484–1506 (2020).
7. Jin, S. et al. Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
8. Noel, F. et al. Dissection of intercellular communication using the transcriptome-based framework ICELLNET. *Nat. Commun.* **12**, 1089 (2021).
9. Hou, R., Denisenko, E., Ong, H. T., Ramilowski, J. A. & Forrest, A. R. R. Predicting cell-to-cell communication networks using NATMI. *Nat. Commun.* **11**, 5011 (2020).
10. Almet, A. A., Cang, Z., Jin, S. & Nie, Q. The landscape of cell-cell communication through single-cell transcriptomics. *Curr. Opin. Syst. Biol.* **26**, 12–23 (2021).
11. Cabello-Aguilar, S. et al. SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.* **48**, e55–e55 (2020).
12. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020).
13. Dimitrov, D. et al. Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. *Nat. Commun.* **13**, 3224 (2022).
14. Wang, S., Karikomi, M., MacLean, A. L. & Nie, Q. Cell lineage and communication network inference via optimization for single-cell transcriptomics. *Nucleic Acids Res.* **47**, e66–e66 (2019).
15. Choi, H. et al. Transcriptome analysis of individual stromal cell populations identifies stroma-tumor crosstalk in mouse lung cancer model. *Cell Rep.* **10**, 1187–1201 (2015).
16. Rodriques, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
17. Chen, A. et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792.e21 (2022).
18. Eng, C.-H. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235–239 (2019).
19. Cang, Z. & Nie, Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat. Commun.* **11**, 2084 (2020).
20. Dries, R. et al. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol.* **22**, 78 (2021).
21. Fischer, D. S., Schaar, A. C. & Theis, F. J. Modeling intercellular communication in tissues using spatial graphs of cells. *Nat. Biotechnol.* **41**, 332–336 (2023).
22. Garcia-Alonso, L. et al. Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. *Nat. Genet.* **53**, 1698–1711 (2021).

23. 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).
24. Li, Z., Wang, T., Liu, P. & Huang, Y. SpatialDM for rapid identification of spatially co-expressed ligand–receptor and revealing cell–cell communication patterns. *Nat. Commun.* **14**, 3995 (2023).
25. 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 Rep.* **29**, 202–211.e6 (2019).
26. 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 Biol.* **23**, 97 (2022).
27. Cang, Z. et al. Screening cell–cell communication in spatial transcriptomics via collective optimal transport. *Nat. Methods* **20**, 218–228 (2023).
28. Armingol, E., Baghdassarian, H. M. & Lewis, N. E. The diversification of methods for studying cell–cell interactions and communication. *Nat. Rev. Genet.* **25**, 381–400 (2024).
29. Miller-Jensen, K., Janes, K. A., Brugge, J. S. & Lauffenburger, D. A. Common effector processing mediates cell-specific responses to stimuli. *Nature* **448**, 604–608 (2007).
30. Oskarsson, T. et al. Breast cancer cells produce tenascin C as a metastatic niche component to colonize the lungs. *Nat. Med.* **17**, 867–874 (2011).
31. Fawkner-Corbett, D. et al. Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell* **184**, 810–826.e23 (2021).
32. Wu, S. Z. et al. A single-cell and spatially resolved atlas of human breast cancers. *Nat. Genet.* **53**, 1334–1347 (2021).
33. Maynard, K. R. et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat. Neurosci.* **24**, 425–436 (2021).
34. Pham, D. et al. Robust mapping of spatiotemporal trajectories and cell–cell interactions in healthy and diseased tissues. *Nat. Commun.* **14**, 7739 (2023).
35. Kleshchevnikov, V. et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).
36. 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. *Front. Immunol.* **10** (2019).
37. Fairlie-Clarke, K. et al. Expression and function of IL-33/ST2 axis in the central nervous system under normal and diseased conditions. *Front. Immunol.* **9** (2018).
38. 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).
39. 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).
40. Miller, S. J. et al. Molecularly defined cortical astroglia subpopulation modulates neurons via secretion of Norrin. *Nat. Neurosci.* **22**, 741–752 (2019).
41. Koeller, H. B., Ross, M. E. & Glickstein, S. B. Cyclin D1 in excitatory neurons of the adult brain enhances kainate-induced neurotoxicity. *Neurobiol. Dis.* **31**, 230–241 (2008).
42. 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).
43. 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. *Mol. Psychiatry* **12**, 656–670 (2007).
44. Agarwal, A. et al. Dysregulated expression of neuregulin-1 by cortical pyramidal neurons disrupts synaptic plasticity. *Cell Rep.* **8**, 1130–1145 (2014).
45. Di Benedetto, B. et al. Differential mRNA distribution of components of the ERK/MAPK signalling cascade in the adult mouse brain. *J. Comp. Neurol.* **500**, 542–556 (2007).
46. Vilar, M. & Mira, H. Regulation of neurogenesis by neurotrophins during adulthood: expected and unexpected roles. *Front. Neurosci.* **10** (2016).
47. Sørlie, T. et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA* **98**, 10869–10874 (2001).
48. Sotiriou, C. et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc. Natl. Acad. Sci. USA* **100**, 10393–10398 (2003).
49. Ignacio, R. M. C. et al. Serum amyloid A predisposes inflammatory tumor microenvironment in triple negative breast cancer. *Oncotarget* **10**, 511–526 (2019).
50. 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).
51. Ribatti, D. et al. Angiogenesis and antiangiogenesis in triple-negative breast cancer. *Transl. Oncol.* **9**, 453–457 (2016).
52. 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. *Ann. Surg. Oncol.* **5**, 271–278 (1998).
53. 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. *J. Pathol.* **192**, 150–158 (2000).
54. Guo, P. et al. ICAM-1 as a molecular target for triple negative breast cancer. *Proc. Natl. Acad. Sci. USA* **111**, 14710–14715 (2014).
55. McGowan, P. & Duffy, M. Matrix metalloproteinase expression and outcome in patients with breast cancer: analysis of a published database. *Ann. Oncol.* **19**, 1566–1572 (2008).
56. House, I. G. et al. Macrophage-derived CXCL9 and CXCL10 are required for antitumor immune responses following immune checkpoint blockade. *Clin. Cancer Res.* **26**, 487–504 (2020).
57. Yang, L., Xiao, L. Y. et al. CCL28 promotes breast cancer growth and metastasis through MAPK-mediated cellular anti-apoptosis and pro-metastasis. *Oncol. Rep.* **38**, 1393–1401 (2017).
58. 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. *J. Cancer* **9**, 757–766 (2018).
59. Mao, S. P. H. et al. Loss of amphiregulin reduces myoepithelial cell coverage of mammary ducts and alters breast tumor growth. *Breast Cancer Res.* **20**, 131 (2018).
60. Moustakas, A. & Heldin, C.-H. Mechanisms of TGFβ-induced epithelial-mesenchymal transition. *J. Clin. Med.* **5** (2016).
61. Wang, Y., Xu, H., Zhu, B., Qiu, Z. & Lin, Z. Systematic identification of the key candidate genes in breast cancer stroma. *Cell. Mol. Biol. Lett.* **23**, 44 (2018).
62. Kitamura, T. et al. CCL2-induced chemokine cascade promotes breast cancer metastasis by enhancing retention of metastasis-associated macrophages. *J. Exp. Med.* **212**, 1043–1059 (2015).
63. 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 Res.* **23**, 394–408 (2013).
64. Lehmann, B. D. et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Investig.* **121**, 2750–2767 (2011).
65. Tuglu, M. et al. 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).

66. 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. *Nat. Commun.* **7**, 11321 (2016).
67. Wittamer, V. et al. Specific recruitment of antigen-presenting cells by chemerin, a novel processed ligand from human inflammatory fluids. *J. Exp. Med.* **198**, 977–985 (2003).
68. Wittamer, V. et al. Neutrophil-mediated maturation of chemerin: a link between innate and adaptive immunity. *J. Immunol.* **175**, 487–493 (2005).
69. 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. *J. Immunol.* **174**, 244–251 (2005).
70. Vermi, W. et al. Role of ChemR23 in directing the migration of myeloid and plasmacytoid dendritic cells to lymphoid organs and inflamed skin. *J. Exp. Med.* **201**, 509–515 (2005).
71. Johnson, D. E., O'Keefe, R. A. & Grandis, J. R. Targeting the IL-6/JAK/STAT3 signalling axis in cancer. *Nat. Rev. Clin. Oncol.* **15**, 234–248 (2018).
72. Sharma, A. et al. Onco-fetal Reprogramming of Endothelial Cells Drives Immunosuppressive Macrophages in Hepatocellular Carcinoma. *Cell* **183**, 377–394.e21 (2020).
73. Duncan, M. B. et al. Type XVIII collagen is essential for survival during acute liver injury in mice. *Dis. Models Mechan.* **6**, 942–951 (2013).
74. 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).
75. Gao, S. et al. Identification of HSC/MPP expansion units in fetal liver by single-cell spatiotemporal transcriptomics. *Cell Res.* **32**, 38–53 (2022).
76. Lewis, K., Yoshimoto, M. & Takebe, T. Fetal liver hematopoiesis: from development to delivery. *Stem Cell Res. Ther.* **12**, 139 (2021).
77. MacParland, S. A. et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun.* **9**, 4383 (2018).
78. Bezerra, J. A. et al. Plasminogen deficiency leads to impaired remodeling after a toxic injury to the liver. *Proc. Natl. Acad. Sci. USA* **96**, 15143–15148 (1999).
79. 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).
80. Li, Z. et al. Presence of onco-fetal neighborhoods in hepatocellular carcinoma is associated with relapse and response to immunotherapy. *Nat. Cancer* **5**, 167–186 (2024).
81. 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 Res.* **35**, W193–W200 (2007).
82. Wang, Q., Morris, R. J., Bode, A. M. & Zhang, T. Prostaglandin pathways: opportunities for cancer prevention and therapy. *Cancer Res.* **82**, 949–965 (2022).
83. Rountree, C. B. et al. A CD133-expressing murine liver oval cell population with bilineage potential. *Stem Cells* **25**, 2419–2429 (2007).
84. Lu, W.-Y. et al. Hepatic progenitor cells of biliary origin with liver repopulation capacity. *Nat. Cell Biol.* **17**, 971–983 (2015).
85. Ma, S. et al. miR-130b promotes CD133+ liver tumor-initiating cell growth and self-renewal via tumor protein 53-induced nuclear protein 1. *Cell Stem Cell* **7**, 694–707 (2010).
86. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).
87. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
88. Cable, D. M. et al. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.* **40**, 517–526 (2022).
89. Subramanian, A. et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
90. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
91. Liu, Z., Sun, D. & Wang, C. Evaluation of cell-cell interaction methods by integrating single-cell RNA sequencing data with spatial information. *Genome Biol.* **23**, 218 (2022).
92. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
93. Lee, S.-I. Developing a bivariate spatial association measure: An integration of Pearson's r and Moran's I. *J. Geograph. Syst.* **3**, 369–385 (2001).
94. Marconato, L. et al. SpatialData: an open and universal data framework for spatial omics. *Nat. Methods* **22**, 58–62 (2025).
95. Rao, N. & Zafar, H. Zafar-lab/renoir: v1.0.0-beta (2025).

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

H.Z. and A.S. designed the study. H.Z. and N.R. developed the model and algorithm. N.R. implemented the software and performed all computational experiments and benchmarking. T.K. implemented the simulation experiment and performed benchmarking. D.K. performed the computational analyses of 10x Xenium datasets, and AK and MME performed wet-lab experiments for generating the 10x Xenium datasets. S.H. performed the IL6 incubation experiment with HepG2 cells. L.Q. and J.G. collected the clinical hepatocellular carcinoma samples for 10x Xenium experiments. R.P., A.M., F.G., and J.C. generated the fetal liver datasets. H.Z., A.S., and N.R. wrote the manuscript, and all authors approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-72388-7.

Correspondence and requests for materials should be addressed to Ankur Sharma or Hamim Zafar.

Peer review information *Nature Communications* thanks Youqiong Ye, Ming Liu, and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.