SpatialDM: Rapid identification of spatially co-expressed ligand-receptor reveals cell-cell communication patterns

Zhuoxuan Li[1], Tianjie Wang[2], Pengtao Liu[1,3*] and Yuanhua Huang[1,2,3*]

[1]School of Biomedical Sciences, University of Hong Kong, Hong Kong SAR, China.
[2]Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong SAR, China.
[3]Center for Translational Stem Cell Biology, Hong Kong Science and Technology Park, Hong Kong SAR, China.

*Corresponding author(s). E-mail(s): pliu88@hku.hk; yuanhua@hku.hk;

Abstract

Cell-cell communication is a key aspect of dissecting the complex cellular microenvironment. Existing single-cell and spatial transcriptomics-based methods primarily focus on identifying cell-type pairs for a specific interaction, while less attention has been paid to the prioritisation of interaction features or the identification of interaction spots in the spatial context. Here, we introduce SpatialDM, a statistical model and toolbox leveraging a bivariant Moran's statistic to detect spatially co-expressed ligand and receptor pairs, their local interacting spots (single-spot resolution), and communication patterns. By deriving an analytical null distribution, this method is scalable to millions of spots and shows accurate and robust performance in various simulations. On one melanoma and multiple intestinal datasets, SpatialDM reveals promising communication patterns and identifies differential interactions between conditions, hence enabling the discovery of context-specific cell cooperation and signalling.

1 Introduction

Cell-cell communication (CCC) plays essential roles in various biological processes and functional regulations [1, 2], for example, immune cooperation in a tumour microenvironment, organ development and stem cell niche maintenance, and wound healing. Protein interaction, as a media of CCC, has been widely studied in the past decades. Despite the relatively low throughput in proteomics technologies, a large number of ligand-receptor candidates still have been accumulated through broad experimental studies and compiled into databases, e.g., 1,396 pairs in CellPhoneDB [3], 1,940 pairs in CellChatDB [4], and 380 pairs in ICELLNET [5].

As a more accessible surrogate, the RNAs of ligand and receptor have been shown effective in the quantification of inter-cellular communications [1]. The advancement of single-cell transcriptomics technologies further enables LR interaction (LRI) and CCC in a cell state-specific manner, for example in the maternal-fetal interface [6] and intestinal stem cell niche [7]. Multiple computational methods have soon been developed to identify the interacting cell types and the mediating LR pairs [1, 8]. CellPhoneDB is a prominent example that considers multimeric proteins in manually curated LRIs and identifies communicating cell types by comparing the null with permuted cell type labels [3, 6]. Another widely used method, CellChat, extends the CCC analysis on multiple aspects, including a mass action model to quantify LR co-expression, expanded LRI candidates with more detailed annotations, and a set of useful plotting utilities [4]. Other methods, including NicheNet [9], PyMINEr [10], iTALK [11], ICELLNET, [5] and SingleCellSignalR [12], have also been introduced in the past two or three years with their unique features on LRI resources and/or testing methods [1]. A recent study further evaluated 16 LRI resources and 7 methods on their impact and consistency in CCC analysis from scRNA-seq data [13], while the direct assessment is generally challenging due to the lack of gold-standard data. Moreover, one major limitation of single-cell-based methods is the lack of spatial coordinates of cells. Therefore, it cannot guarantee physical proximity between the putative interacting cells and may lead to high false-positive rates [8].

In recent years, spatial transcriptomics (ST) technologies have also embraced a few major breakthroughs, on both sequencing and imaging-based platforms [14], therefore, ST is increasingly used to double-check the physical proximity of the LRI identified in single-cell data. Meanwhile, a few ST-based methods have been developed to identify CCC and LRIs directly from ST data [15, 16]. Giotto is a toolbox for multifaceted analyses of ST data, including detecting cell-type pairs that have increased interactions of proximal cells than those at random locations [17]. SVCA is a Gaussian process-like method that defines a universal cell-cell interaction covariance over spatially smoothed cell embeddings and consequently identifies genes with a high proportion of variance explained by this interaction term [18]. SpaOTsc leverages an optimal transport method to quantify the likelihood of interaction between any two cells, with spatial distance as one cost component [19]. SpaTalk is another

recently proposed toolbox to analyze spatial LRI and CCC by testing if a certain cell-type pair is enriched in those co-expressed spots [20]. Although these methods brought promise to directly analyze CCC in a spatial context, most of them focus on identifying interacting cell types for all LRIs instead of detecting the interacting LR pairs first, hence may over-interpret less informative LRIs. Additionally, these strategies may not be sensitive enough to identify regional CCC, as they aim to detect cell types with enriched interactions as a whole (Fig. 1A). Moreover, the conventional permutation test is not scalable and may slow down the computational analysis, particularly considering the fast advances in spatial resolution and cell numbers.

Here, to address these limitations, we introduce SpatialDM (Spatial Direct Messaging, or Spatial co-expressed ligand and receptor Detected by Moran's bivariant extension), a statistical model and toolbox that uses a bivariate Moran's statistic to identify the spatial co-expression (i.e., spatial association) between a pair of ligand and receptor. Critically, we introduced an analytical derivation of the null distribution, making it highly scalable to analyze millions of cells. This method also contains effective strategies to identify interacting local spots and the patterns shared by multiple LRIs or pathways. We evaluated the accuracy of SpatialDM and other counterparts with various simulations and demonstrated its effectiveness in detecting LRIs and differential interactions between conditions in melanoma and intestinal datasets.

2 Results

2.1 Overview of SpatialDM method

Identification of the communicating cells and the interacting LR pairs are the two major orthogonal tasks in dissecting CCC in scRNA-seq and ST data. Most existing methods mainly aim to address the former challenge (at cell-cluster or cell-type resolution) but omit the latter task of feature selection simply by relying on a curated database. However, we argue that identifying the dataset-specific interacting LR pairs is a crucial step for ensuring quality analysis and reliable interpretation of the putative CCC.

Therefore, the primary aim and the first step of our SpatialDM is to detect LR pairs that have significant spatial co-expression (i.e. ligand and receptor transcripts are expressed within a reasonable geographical distance) in ST data. The candidate LR pairs are generally from a comprehensively curated database, e.g. CellChatDB by default. Fig. 1A shows an example that the LR pair B has spatial co-expression and can be detected by SpatialDM, while pair A does not though its cluster-level enrichment may lead to false positives in existing approaches. Generally, this problem of spatial association between two variables can be formulated by a regression model, either via fixed effects, e.g., SDM and SDEM [21] or random effects, e.g., SVCA [18]. Here, we introduce a bi-variate Moran's R as a test statistic (Fig. 1A; Methods), which can well account for the spatial association, i.e., the spatial co-expression of ligand and receptor here. This method is an extension of the well-known Moran's I

in uni-variate auto-correlation analysis [22] to a bivariate setting initially by Wartenberg [23] and is still widely used in the broad field of spatial analysis [24, 25]. The computational convenience and effectiveness make it an appealing method for LRI in ST data (see evaluation below).

As a computational toolbox, SpatialDM has major functions for both global and local analyses (Fig. 1B). First, by leveraging this bivariate R, we introduce a hypothesis testing to reject the null that the ligand and receptor are spatially independent, hence allowing us to select the spatially co-expressed LR pairs. Second, we further adapted local Moran's I to their bivariate format to detect local hits for each significant LR pair (Methods). Based on the local interaction hits for each LR pair, SpatialDM allows grouping these significant LR pairs into a few distinct communication patterns, e.g., by the automatic expression histology model introduced in SpatialDE [26]. Third, to interpret the local communication patterns, it also provides an enrichment test and visualization of putative pathways for each local pattern. Last, as a unique feature, SpatialDM further supports detecting LR pairs that have differential interaction density between conditions or along with a continuous covariate, which is highly demanded for biological discovery in both developmental and disease contexts.

2.2 Accurate and efficient z-score test

In order to obtain the null distribution in this hypothesis testing problem, a generic method is permutation as used by most CCC methods, where the test statistic R will be calculated by random shuffling of binding partners for each pair, e.g., 1,000 times. On the other hand, when the number of spatial spots is large, the permutation test often becomes a computational bottleneck for the analysis. Therefore, we derived the first and second moments of the null distribution to analytically obtain a z-score and its according p-value for the observed R (see Supp. Note 1). Strikingly, the z-score-based p-value has high correlations with the permutation-based p-value in datasets with different sizes (Fig. 1C,D; Spearman's $R > 0.98$, local statistics correlation: Extended Data Fig. 1D,E). Given the computational convenience, SpatialDM (the permutation mode, 1 CPU) ranks as the fastest method among all permutation-based methods, finishing testing 1000 LR pairs within 1.5 min for a 10,000-spot dataset (all using 50 CPUs except SpaTalk). Importantly, the z-score-based strategy further introduces over 100x speedups, therefore is exclusively scalable to a million spots within 12 minutes (even with a single CPU). Therefore, this innovation of analytical null distribution can be highly valuable for the analysis of ST data with increasingly large sizes.

To examine the accuracy of SpatialDM in detecting spatially correlated ligand-receptor pairs, we first generated multiple sets of simulated ST data by adapting a recent method SVCA [18]. In short, SVCA is a principled Gaussian process model that decomposes the variance of a certain gene (a ligand here) into cell states, spatial proximity, spatially weighted receptor (i.e., the ligand-receptor spatial interaction), and residual noise (see Methods). Here, based on

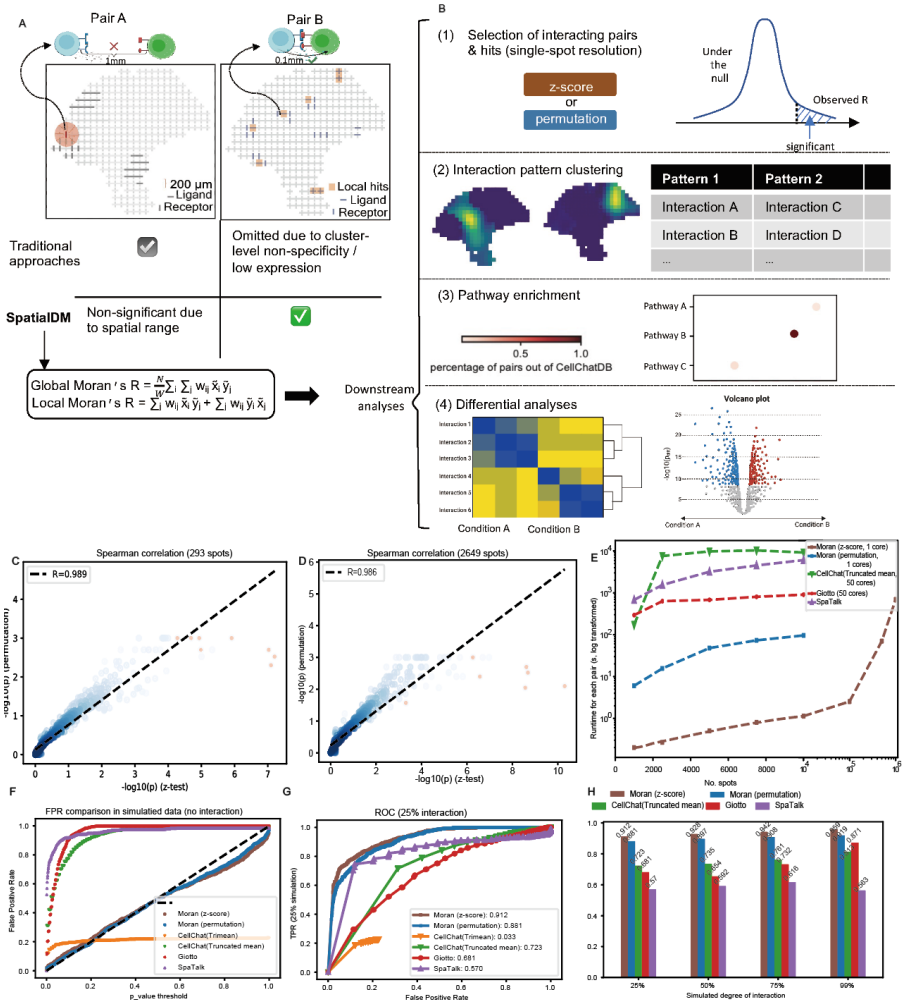

Fig. 1 (A) Illustration of SpatialDM method. Top: two examples of ligand-receptor pairs in ST where pair A cannot interact due to distant location while pair B can. Different from traditional approaches focusing on cell type enrichment, SpatialDM aims to detect spatially co-expressed LR pairs by a bivariate Moran's R. (B) The four major utility functions in the SpatialDM toolbox for CCC analysis, including (1) selecting LR pairs with global significance, local interaction hits and (2) patterns, (3) pathway enrichment analysis and (4) detecting differential interacting pairs between conditions. (C-D) Consistency between SpatialDM's permutation and z-score modes in a melanoma dataset (293 spots; C) and an intestine dataset (2,649 spots; D), respectively. Spearman correlation coefficients were specified on the top left. Fitted linear regression is in a black dash line. (E) Running time comparison of SpatialDM with CellChat, Giotto and SpaTalk, where only SpatialDM (z-score mode) is scalable to 1 million spots within 12 min. (F-H) Assessment by simulated datasets: (F) False Positive Rate (FPR) comparison when no interaction is simulated and only SpatialDM calibrates with the null distribution along with the diagonal line. (G) Receiver Operating Characteristic (ROC) under the 25% degree of interaction simulation scenario. Area Under ROC (AUROC) for each method is labelled in the legend. (H) Comparison of AUROC under 4 different degrees of interactions (25%, 50%, 75%, 99% interaction respectively).

a seed dataset with 293 spots and 1,180 LR pairs, we first generated a negative set with 0% variance explained by ligand-receptor spatial interaction. When applying SpatialDM to this negative data set (under the null), we found that the p-values of both permutation and z-score are well calibrated to a uniform distribution (Fig. 1F), despite the data being generated by a different model. In contrast, CellChat with 2 different parameter settings [4], Giotto [17], and SpaTalk [20] failed to control false positives, possibly due to the lack of effective modelling of spatial information.

To further evaluate the power of SpatialDM and its overall performance, we generated a positive set with 25% variance explained by the spatial correlation (Methods) and applied SpatialDM to the pool of positive and negative sets. With the default cutoff of p-value<0.05, SpatialDM achieves a power of 74.5% and controls a false positive rate of 8.2% with the z-score approach. By varying the p-value, it returns an AUROC of 0.912 (z-score mode; permutation AUROC=0.881), substantially outperforming the other existing methods (AUROC: 0.570 to 0.723; Fig. 1G). Similar results were also observed when generating positive samples with higher levels of variance explained by spatial interaction from 50%, 75% to 99%, where the AUROC increases accordingly up to 0.959 (Fig. 1H, Extended Data Fig. 1A-C). Note, as CellChat's Trimean mode has limited power (Fig. 1G and Extended Data Fig. 1A-C), we excluded it for further comparison and only keep CellChat's Truncated-mean mode.

2.3 Detecting spatial LRI in melanoma

Next, we applied our SpatialDM to the aforementioned seed data, a melanoma sample probed by ST platform (200 μm centre-to-centre distance), covering over 7 cell types from 293 spots [27]. Of note, given the small sample size, we employed SpatialDM's permutation approach. When applying to the 1,180 LR pairs from CellChatDB, SpatialDM detects 130 spatial co-expressed pairs (FDR<0.1; Fig. 2A and Supp. Table 1). In contrast, other methods generated 340 to 874 significant pairs, raising the possibility of false positives (Supp. Table 1). Indeed, all other methods suffer from high false positives when testing on a manually generated negative set by shuffling the ligand-receptor database to create a list of 663 non-documented ligand-receptor pairs (e.g., 285 pairs by Giotto as the best counterpart; Extended Data Fig. 2A, Supp Table 2). However, SpatialDM has good false-positive controls here (90 pairs; permutation p-value<0.05), which is consistent with the simulation (Extended Data Fig. 2A and Fig. 1F). A similar pattern is also observed on two expected irrelevant LR pairs (FGF2_FZD8 and PHF5A_EDEM3; Fig. 2A).

Interestingly, many known melanoma-related genes like VEGF, SPP1, and CSF1 have been included in the 130 LR pairs selected by SpatialDM. Further, we applied SpatialDM to identify local hits of interaction by local Moran's R ($p < 0.1$). Given the general low depth in spatial transcriptomics, the method proves sensitive enough by detecting pairs as sparse as 3 interaction spots, and also powerful by detecting as many as 200 spots. The 130 selected pairs were subjected to automatic expression histology from SpatialDE, which resulted in

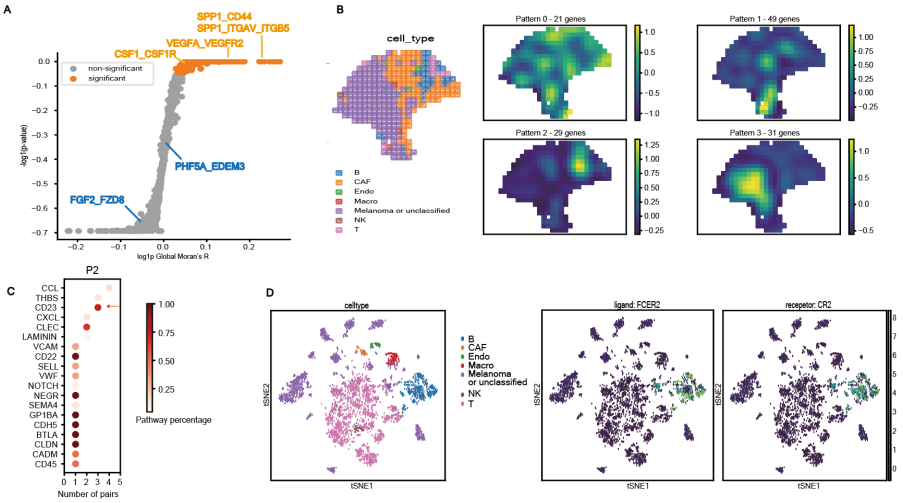

Fig. 2 SpatialDM detects spatially co-expressed LRs in melanoma data and identifies CCC patterns. (A) Scatterplot of Global Moran's R and permutation p-value. X-axis is log(Global Moran's R+1); y-axis is -log(p value+1). Significant pairs (FDR< 0.1) are highlighted in orange. Three well-known positive samples (in orange) and two negative samples (in blue) are highlighted. (B) Left: Spatial plot coloured by cell types that have the largest proportion estimated by RCTD. Right: Clustering of 130 significant pairs into 4 spatial patterns by SpatialDE. A representative plot is shown for each spatial pattern. Each plot is coloured by the posterior mean of genes in each spatial pattern. (C) Dot plots of enriched pathways in pattern 3 from panel B. Y-axis: the name of the enriched pathway; x-axis: the number of significant LR pairs in that pathway for each pattern; dot colour: the percentage of significant pairs for that pathway from CellChatDB. (D) tSNE plots of matched melanoma scRNA data, coloured by the original cell types and expression level of ligand FCER2 and receptor CR2, one pair of CD23 ligand-receptor interactions as highlighted in (C).

4 coarse patterns (Fig 2B, Supp Table 3). We observed that pattern 0 is pan-melanoma, patterns 1 maps to the cancer-associated fibroblast (CAF) region, patterns 2 correspond to the lymphoid region, and pattern 3 simulates the melanoma region, referenced to Thrane, et. al. and the predicted cell types from scRNA-seq by RCTD [28] (Fig 2B). Indeed, we found that the local interaction scores are good predictors of the cell types (Pearson's R=0.928; linear regression; Extended Data Fig. 2B).

We then identified pathways enriched in each pattern, and found that the melanoma region (i.e. pattern 3) shows signatures of angiogenesis and tumour progression (Extended Data Fig. 2C, Extended Data Fig. 3, Supp. Table 3). Immunity-related pathways (including CCL and CD23) were enriched in the lymphoid region (i.e. pattern 2, Extended Data Fig. 3), concordant with histology annotations provided by the authors and RCTD annotated results (Fig. 2B-C, Supp. Table 3). CD23, a less-discussed pathway in melanoma showed high relevance in pattern 2 (Fig 2C), which led us to examine the result in an annotated melanoma scRNA-seq dataset with greater sequencing

depth and resolution. CD23 (a.k.a, FCER2) could bind with CR2 or integrin complexes to trigger immunologic responses [29, 30]. Consistent with the identified region (pattern 2), it was mainly found in B cells in other reports. In another melanoma scRNA-seq data we examined [31], FCER2 and its receptors were also enriched in the B cells, which is 20-fold higher than any other cluster, validating the discoveries from spatial transcriptomic analyses (Fig 2D, Extended Data Fig. 2D). Interestingly, by examining the 3,500 genes that are up-regulated in the CD23 hot spots, we found that they are highly enriched in T cell and B cell activation pathways, supporting anti-tumour functions, instead of a pro-tumour role (Gene Ontology analysis with PANTHER; Extended Data Fig. 2E; Supp Table 4). Taken together, these identified LRI and their regional patterns may contribute to further signalling investigation and potential treatment targets.

2.4 Identifying consistent cell-cell communications in multiple intestine samples

Human intestines originate from all three germ layers, involving a variety of developmental cues at different post-conceptual weeks (PCW), and sophisticated self-renewing mechanisms of the crypt-villus structure throughout adult life. With time-stamped single-cell and spatial transcriptomic datasets from 12 post-conceptual weeks (12 PCW: 3 colon replicates from 2 donors, A3, A8 and A9, 2 TI replicates from one donor, A6, A7) or 19 PCW foetus sample (1 slice, A4) to adult samples (2 replicates from 1 donor, A1 and A2, with IBD or cancer), Corbett, et al. have identified several ligand-receptor interactions through customized analyses (100 μm spot-spot distance, Supp. Table 5) [7]. Briefly, Corbett, et al. screened through a database of over 2,000 LR pairs, giving each ligand and receptor specificity scores and expression scores across each of the 101 scRNA clusters; then, the putative list of LR interactions with high specificity and expression in a cluster-cluster combination was validated in spatial transcriptome regarding LR spatial colocalization. As a result, Corbett, et al. have identified CEACAM1_CEACAM5 toward the crypt top in adult samples, IL7_IL7R_IL2RG, CCL21_CCR7 and CCL19_CCR7 between Lymphoid Tissue Inducer (LTi) and S4, ANGPT2 in fetal vasculature, and many others [7]. Considering the large sample size, we leveraged the z-score approach in SpatialDM to re-analyze all samples in this dataset, and identified most of these reported interacting pairs (300 out of 301; Supp. Table 6, Extended Data Fig. 4A). More interestingly, 138 additional LR pairs are uniquely identified by SpatialDM, suggesting its potentially enhanced sensitivity in detecting sparsely expressed LR pairs.

Thanks to the multi-sample setting, we first used this dataset to assess the reproducibility of SpatialDM in both detecting spatially co-expressed LR pairs and their communicating regions. When comparing the global Moran's R, we observed high correlations between slices from the same sample versus low correlations among slices from different samples (Fig. 3A). Similarly, whole-interactome clustering revealed the dendrogram relationships that are close

to the sample kinship (e.g. A8 and A9 from one 12 PCW sample is close to another 12 PCW sample A3 but far from the adult samples A1 and A2; Extended Data Fig. 4B,C).

Next, we assessed whether local hits discovered by SpatialDM are consistent in technical or even biological replicates. The cell type weights of local selected spots are highly correlated between technical replicates (e.g. median Pearson's R=0.975 for A1 vs. A2 and R=0.871 for A8 vs. A9, Extended Data Fig. 4D-F), moderately correlated between biological replicates (e.g. A3 vs. A9), but poorly correlated in distinct samples (e.g. A3 vs. A7, Extended Data Fig. 4G). Given the sensitivity of SpatialDM, the consistency in local pattern detection is observed for both ubiquitously interacting pairs and sparse ones, from which we illustrate two concrete examples here. FN1_CD44 interacts more ubiquitously in adult and foetus colons (Extended Data Fig. 4D, Extended Data Fig. 5, Supp. Table 6), probably due to its versatile role during intestine development [32]. The interaction of PLG_F2RL1 is exclusively found in all fetal slices, and with consistent cell-type enrichment in enterocytes (Fig. 3B, Extended Data Fig. 5).

2.5 EGF pathway interactions are enriched in adult crypt top colonocytes

Seeing the consistency of SpatialDM between technical replicates, we then zoomed into sample A1 to reveal the interaction patterns in adult colons with IBD or cancer. Through similar procedures as in melanoma analysis, the 292 significant pairs (z-score FDR<0.1, hits in at least 10 spots) were classified into 4 patterns (Supp. Table 7). Pattern 0 is mostly enriched in immune cells, pattern 1 in crypt top colonocytes, and pattern 2-3 in myofibroblast (Fig. 3C-D, Extended Data Fig. 6A). Such cell-type enrichment patterns are consistent with pathway enrichment. For example, interactions under MHC-II and ICAM pathways show high relevance in pattern 0, which showed enrichment in immune cells, suggesting an inflammatory microenvironment in the adult colon (Extended Data Fig. 6B, Supp. Table 7). The EGF pathway comprises diverse ligands (including EGF, TGFA, AREG, EREG, and HBEGF) and receptors (including EGFR, ERBB2, ERBB3, and ERBB4), exerting distinct or redundant functions [33]. In the adult sample we analyzed, most EGF interactions were detected and enriched in pattern 1 (Fig. 3E-F, Extended Data Fig. 6C, Supp. Table 7).

As reported, the EGF signalling plays important roles primarily in intestinal epithelial cell proliferation and self-renewal, and has a complex interplay with other pathways [33]. Nászai, et al. have revealed that RAL GTPases, encoded by RALA and RALB, are necessary and sufficient to activate EGFR signalling and further MAPK signalling in the intestine [34]. Interestingly, we indeed found that the upstream RALA and RALB expression and downstream MAPK expression have great overlap with the local Moran selected spots (Fig. 3G, Extended Data Fig. 6D). It highlights the potential to detect interplay with upstream or downstream signalling of LRI captured by SpatialDM.

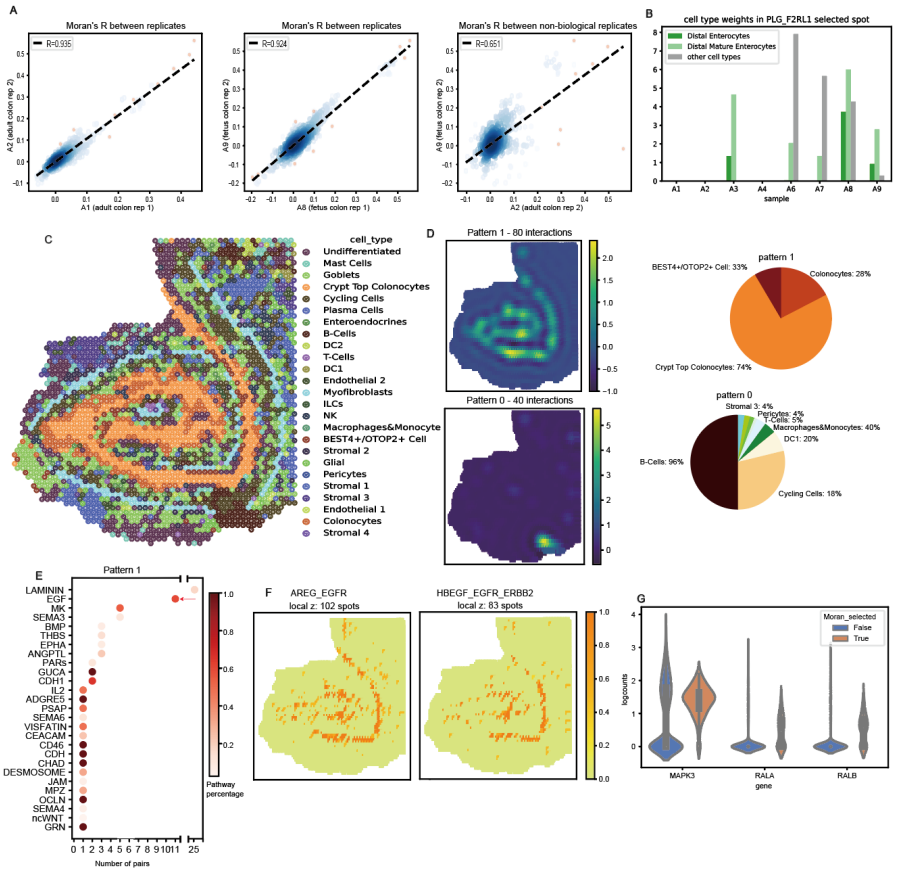

Fig. 3 Multiple intestinal samples for technical validation and CCC pattern discovery. (A) Cross-replicate correlations of Global Moran's R. Pearson coefficients were specified in the top left. (From left to right) Correlations between two adult replicates (A1 and A2), between two foetuses, replicates (A8 and A9), and between an adult and a foetus sample. (B) Additive cell type weights of PLG_F2RL1 interacting spots (local z-score p <0.1). PLG_F2RL1 was not identified in adult samples (A1, A2) and 19 PCW colons (A4). Other cell types than enterocytes have been grouped together. (C) Spatial plot of A1 sample coloured by cell type annotated from the original study. (D) Summary of pattern 0 and 1 in A1 from SpatialDE. Left: Spatial plots of the pattern. Right: Weighted sum of cell type composition for each pattern. For each cell type, the additive weights over total weights were specified in the bar charts. (E) Pathway enrichment result for pattern 1 in A1. (F) Selected local spots of two example pairs under EGF pathways. Color denotes $1-p$. The number of significant spots were specified ($p < 0.1$)(G) Comparison of MAPK3/RALA/RALB expression in AREG_EGFR local selected spots (z-score $p < 0.1$) vs. non-selected spots.

2.6 SpaitalDM identifies differential interactions between foetus and adults

Besides the sample-independent analysis, SpatialDM allows differential analysis of detailed interactive pairs between conditions or along with a continuous covariate, accounting for multiple replicates. Briefly, a (generalised) linear

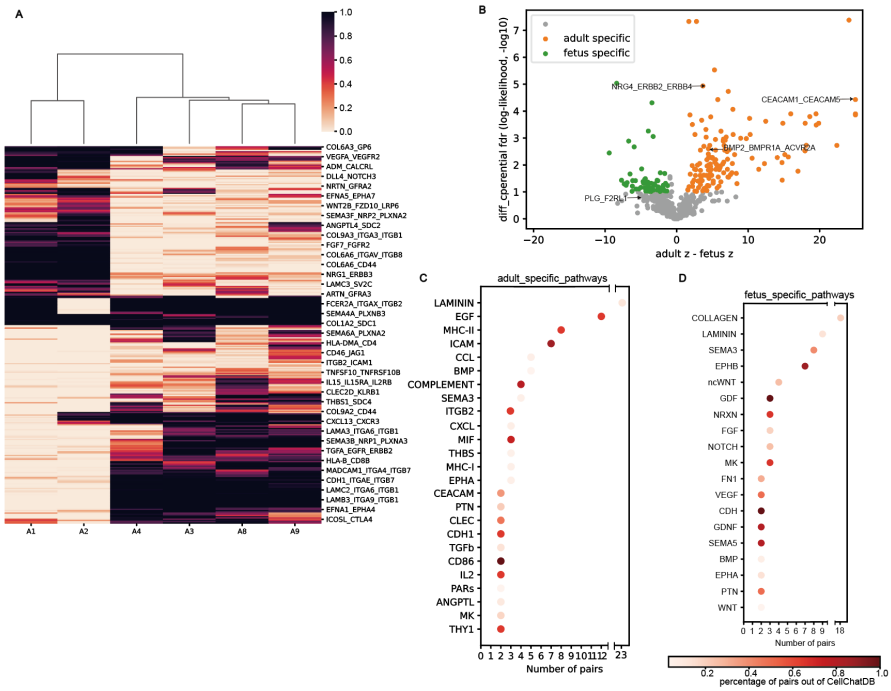

Fig. 4 SpatialDM identifies differential LR interaction between foetus and adult intestines. (A) Heatmap colored by global z-score p-values of adult-foetus differential pairs (likelihood ratio test FDR < 0.1), including 135 adult-specific (low p-values in A1 and A2), and 98 foetus-specific. (B) Volcano plot highlighting adult-specific and foetus-specific pairs, in orange and green, respectively (FDR < 0.1). z difference (x-axis) cutoffs were taken at 30% and 70% quantiles, respectively. (C-D) Pathway enrichment results for adult-specific and foetus-specific pairs. Only pathways with 2 or more selected pairs were visualized.

model is introduced to test if a certain covariate affects the interaction density (indicated by the z-score or permutation numbers; see Methods). Here, we showcase the differential analyses among adult vs. fetal colon samples based on the z-score inputs (Fig. 4A-B and Supp. Table 8), where 135 pairs of LR interactions are up-regulated in adult samples while 98 pairs in the foetus (FDR <0.1; likelihood ratio test, Fig. 4B).

By pathway enrichment analysis (Fig. 4C), we first noticed the adult-specific pairs enriched with chemokine and cytokine responses (e.g. ICAM, CCL and CXCL) as well as inflammatory and immune signatures (e.g. MHC-II, TGFB, COMPLEMENT, BMP and MIF), which is consistent with insights from previous comparative RNAseq analysis [35]. It was known that inflammation in the foetus can be associated with preterm parturition [36], Fetal Inflammatory Response Syndrome (FIRS) [37], impaired neurological outcomes [38], and other defects. In our analysis, some pathways like ICAM, COMPLEMENT and CXCL are generally exclusive in adults (Extended Data

Fig. 7), while other interactions like TGFB and CCL can be possibly established early in the foetus stage. For example, TGFB3_TGFBR1_TGFBR2 was identified across each time points (Supp Table 7, Extended Data Fig. 8). TGFBs are potent immunosuppressive cytokines, which drive the functional development of lymphocytes, therefore reinforcing the gut barrier. Such interactions may have critical roles during early intestine development at the foetus stage. We also observed that the foetus-enriched pathways are associated with neural processes (e.g. NRXN, GDNF, PTN), new blood vessel formation (e.g. SEMA, VEGF), and growth (e.g. GDF, MK; Fig. 4D, Extended Data Fig. 9). Such observations of early establishment prior to 12 PCW were consistent with Corbett, et al [7]. Overall, we provide evidence that the diseased adult intestine has a more pro-inflammatory environment, while the fetal intestine has more development-related signatures.

Beyond pathway-level comparison, SpatialDM allows differential analysis on a certain ligand-receptor pair (Supp. Table 8). While traditional pathway enrichment may only detect enriched BMP pathway in adults, SpatialDM further refines the adult-specific interactions to BMP2 and its receptors (BMPR1A/B and ACVR2A, Supp. Table 6,8). In fact, with the function of promoting apoptosis and inhibiting proliferation, BMP2 was previously revealed by RT-PCR and immunoblotting to be expressed by, and act on mature colon epithelial cells [39]. There have also been multiple reports of epithelium-immune orchestration in the adult intestine. We have identified NRG4_ERBB2 among various cell types in adult-specific interactions (FDR<0.0001, Fig. 4B, Extended Data Fig. 10), but not in foetus samples. Interestingly, NRG4 was found in human breast milk, and its oral supplementation can protect against inflammation in the intestine [40]. Our analysis consolidated that the anti-inflammation system may only be established after early conceptual weeks, likely at infant breastfeeding stages as reported.

In addition to adult-only pairs, our differential analysis allows the detection of LR pairs with a subtle but significant change in communication density between adult and foetus (Fig. 4B). CEACAM1_CEACAM5 is an example that was also demonstrated in adult samples by the authors. Although we have identified CEACAM1_CEACAM5 in A3 with a moderate signal (FDR = 0.003) in addition to two adult slices, CEACAM1_CEACAM5 was considered adult-specific in the differential analysis (Fig. 4B, differential $p<0.0001$, A1 R=0.433, A2 R=0.577). In fact, CEACAM1_CEACAM5 is only sparsely expressed in A3 (R=0.034), with few positive significant interaction spots (Extended Data Fig. 6E). Both molecules were recognized to be highly present in human colon epithelia and related to inflammation and tumorigenesis [41]. Defects in CEACAM signalling in intestinal epithelial cells are associated with Inflammatory Bowel Disease (IBD), and even Colitis-Associated Cancer (CAC) [41, 42]. As we revealed the interplay of various cell types including colonocytes and cycling cells in CEACAM1_CEACAM5 interaction in IBD or colorectal cancer patients, it might highlight targeting these cells to reverse the adverse conditions.

Overall, SpatialDM has not only validated a number of interactions discussed by the original report, but also uncovered multiple new insights into the inter-compartment orchestrations in the human intestine, especially by allowing differential analyses among multiple replicates. Therefore, SpatialDM enables the generation of new hypotheses for further experimental studies to discover more underlying mechanisms of intestinal disorders which are currently poorly understood.

3 Discussion

To tackle unaddressed questions in spatial transcriptome as what ligand-receptor interact and where they take place, we introduce SpatialDM, a statistical model in the form of bivariate Moran's method. This method uniquely aims to effectively detect the spatially co-expressed ligand-receptor at single-spot resolution as the primary task, ensuring the high-quality discovery of communication patterns. Critically, we also derived an analytical form of the null distribution, therefore SpatialDM does not need to rely on the time-consuming permutation test, and is scalable to millions of spots.

Following the significant LR pairs, SpatialDM further identifies the local communicating spots and their regional patterns, facilitating various downstream explorations. Notably, the concise framework also allows differential analyses under multi-sample settings with the likelihood-ratio test of global z-scores. This facilitates spatial-temporal analyses of cell-cell interactions in a time-series design or along with a pseudo-time trajectory. Such differential analyses are not only helpful in identifying disease mechanisms and potential treatment targets but also enable the detection of subtle changes during development on an interacting pair level instead of the pathway level.

Similar to most CCC methods, SpatialDM also takes a curated LR database as input. As SpatialDM is capable of detecting dataset-specific LR pairs, we generally recommend feeding a more comprehensive database, e.g., CellChatDB by default, while one can input a customised candidate list. Of note, all analyses of ST data here are only on the mRNA level, while other factors, e.g., alternative splicing, translation machinery, and post-translational modifications can further determine whether the interactions actually happen on the cell surface. While ST datasets have been examined in this paper given their prevalence, the same framework could, in principle, be directly applied to high-throughput spatial proteomic datasets to generate more direct interpretations, particularly considering the rapid development of spatial proteomics or multi-omics technologies, e.g., Deep Visual Proteomics (DVP) and DBiT-seq [43, 44].

Another open challenge is to identify the downstream targets of LR interactions, which can largely enhance the interpretation of the signalling pathway of a certain CCC. Though we showed one case that the literature-reported downstream targets are well supported here, a comprehensively curated database

with high quality will be largely appreciated to perform a systematic investigation; the scMLnet database might be an option [45]. Additionally, more sophisticated methods are desired in addressing this challenge.

Furthermore, there are also technical elements in the SpatialDM framework worth further exploration. For example, we only used the RBF kernel for defining the spatial similarity matrix, while other kernels may be applicable too, e.g., the Cauchy kernel or a mixture of multiple kernels. Also, given a small number of replicates, the detection of differential communicating LR pairs between conditions is generally challenging, hence a Bayesian treatment for jointly analysing all pairs may mitigate this issue to a certain degree.

To conclude, the method presented here resolved the selection of the spatially communicating LR pairs in ST data, allowing for effective CCC pattern discovery in a local region and identification of condition-specific communications. With the rapid development of spatial omics technologies, SpatialDM opens up an efficient and reliable way to dissect cell cooperation in a micro-environment.

Methods

Global Moran's R for spatial co-expression

In order to perform reliable cell-cell communication in ST data, SpatialDM aims to identify ligand-receptor with significant spatial co-expression, from a comprehensive candidate list. By default, we use LR lists from CellChatDB v.1.1.3 (mouse: 2,022 pairs, human: 1,940 pairs) as input [4], while users can use any customised list.

Here, for detecting the spatial co-expression, we extended the widely used Moran's I from a univariate to a bivariate setting. This is an extension which is closely related to the earlier use in geography proposed by Wartenberg [23]. In order to distinguish the spatial auto-correlation in a univariate setting, we call this bivariate statistic Moran's R, as follows

$$\text{Global Moran's } R = \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}}, \tag{1}$$

where x_i and y_j denotes log-transformed ligand and receptor expression at spot i and j, respectively. Spatial weight matrix computation is based on Radial Basis Function (RBF) kernel with an element-wise normalization,

$$w_{ij}^{(0)} = \exp\{-\frac{d_{ij}^2}{2l^2}\};\ w_{ij} = \frac{n}{W} w_{ij}^{(0)}, \tag{2}$$

where d_{ij} is the geographical distance between spot i and j (i.e., Euclidean distance on spatial coordinates), W is the sum of $w_{ij}^{(0)}$, and n is the number of spots. Optionally, if assuming single-cell resolution, the diagonal of the weight matrix can be made 0 to reduce the influence by auto-correlations when ligand

and receptor are encoded by the same gene: $w_{ii} = 0$ for any i. For the analysis in this work, all datasets are not of single-cell resolution, so the zero-diagonal is not used.

In addition to the scale factor l in the RBF kernel, alternative options through either cut-off (co) or the number of nearest neighbours ($n_neighbors$) can be customized to restrain the interaction within certain spots' diameter distance. In the melanoma data (200 μm center-to-center distance) analysis, we assigned $l = 1.2, co = 0.2$; In the intestine data (100 μm center-to-center distance) analysis, we assigned $l = 75, co = 0.2$ (according to larger coordinate scale). Such a setting is based on the assumption that CCC can occur in 100-200 μm.

For ligands or receptors composed of multiple subunits, we computed the algebraic means as inputs for SpatialDM, i.e.

$$x_i = \frac{\sum_{s=1}^{S_L} x_i^{(s)}}{S_L};\ y_j = \frac{\sum_{s=1}^{S_R} y_j^{(s)}}{S_R}, \tag{3}$$

where s is the sth subunit for ligand x_i (with S_L subunits) or receptor y_j (with S_R subunits). Users can also opt for geometric means for more stringent selection results.

Hypothesis testing with global Moran's R

In order to perform the hypothesis testing, the distribution of R statistic under the null (i.e., ligand and receptor are spatially independent). Two methods can be adopted to approximate the null distribution and calculate the p value: (1) Permutation method by shuffling w_{ij} for multiple times (e.g. 1000), and then calculate the p value as the proportion of the permutation R values that are as large as the observed value; (2) Analytical method by approximating the null distribution with a normal distribution by deriving its first and second moments (see Supp. Note 1), then a corresponding z-score can be calculated, as follows:

$$z = \frac{R - 0}{\sqrt{\texttt{Var}(R)}}, \tag{4}$$

where the final form of the variance can be written as:

$$\texttt{Var}(R) = \frac{n^2 \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} w_{ji} - 2n(\sum_{i=1}^{n}(\sum_{j=1}^{n} w_{ij} \sum_{j=1}^{n} w_{ji}) + (\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij})^2}{n^2(n-1)^2} \tag{5}$$

Then the p-value can be obtained by the survival function in a standard normal distribution from the z-score.

Significant interaction spots

Similar to the global R, we also introduce local R in a bivariate setting as a testing statistic to indicate the local interacting spots for each ligand-receptor pair. The local Moran's R_i for spot i is defined as follows,

$$\text{Local Moran's } R_i = (x_i - \bar{x}) \sum_{j=1}^{n} w_{ij}(y_j - \bar{y}) + (y_i - \bar{y}) \sum_{j=1}^{n} w_{ij}(x_j - \bar{x}), \quad (6)$$

where x and y denotes raw ligand and receptor expression, respectively. Similar to the Global counterpart, we applied both permutation and z-score approaches on Local Moran's R to identify significant interaction spots, where the variance for local R_i is derived as:

$$Var(R_i) = 2\frac{(n-1)^2}{n^2}\sigma_1^2\sigma_2^2 \sum_{j=1}^{n} w_{ij}^2 + 2\frac{(n-1)^2}{n^2}\sigma_1^2\sigma_2^2 w_{ii}^2 \quad (7)$$

where σ_1 and σ_2 are the standard deviations for ligand and receptor, respectively (see more details in Supp Note 1).

To avoid picking interacting spots with low sender signals and low receiver signals in the neighbourhood, which would result in a high positive Local Moran's R, we adapted to the quadrant method of Moran's I and refined the significant spots to be positive for both sender signals and receiver signals in the neighbourhood, i.e. Local $p_i = 1$ when $x_i = 0$ and $y_i = 0$

Simulation

The simulation approach was adapted from SVCA [18] and was based on Thrane's melanoma dataset with 293 spots [27]. In SVCA, the variance of each gene was decomposed using a multivariate normal model into the intrinsic factor which can be inferred from expression patterns of all other genes, the environmental factor which can be imputed from spatial adjacency, the noise factor, and most importantly, the interaction factor which is a linear combination of neighbour cell expression profiles. After fitting the model to real spatial data, SVCA rescales the interaction factor to simulate different degrees of interaction. Here, with the hypothesis that genes correlate more with binding partners instead of all other genes, we adapted SVCA by replacing the intrinsic factor modelled from all genes with corresponding receptor subunits for each ligand gene. Please refer to SVCA for detailed protocols [18]. SVCA settings were kept except the term X which was the expression profile across all spots of all genes except the molecule of interest (dimensions = number of molecules -1), and adapted as the expression profile across all spots only on the corresponding receptor genes (dimensions = number of receptor subunits). Briefly, the adapted SVCA model was fitted for each ligand gene in the ligand-receptor database using maximum likelihood. The cell-cell interaction covariance was then rescaled to simulate circumstances of no interaction (0%), 25% interaction, 50% interaction, 75% interaction, and 99% interaction. For negatively correlated pairs observed from all scenarios except 0% interaction, we reversed the signs for each simulated ligand expression value.

Comparison with other models

We applied SpatialDM (both approaches, non-single cell resolution, $l = 1.2$, cut-off=0.2), CellChat (v.1.1.3; default trimean setting and truncatedMean with trim = 0), Giotto (v.1.0.4; default setting), and SpaTalk (v.1.0; loss option changed to mse) to the positive-interaction simulations to compare the true positive rate (TPR), and to the no-interaction simulation to compare the false positive rate (FPR). As CellChat and Giotto results were presented on a cluster level, we kept the lowest p-value for each ligand-receptor pair across all cluster-cluster results. Receiver operating characteristic (ROC) was plotted for each method, and Area Under ROC (AUROC) were compared under each interaction scenario. We also compared the computation time for 1000 LR pairs of the aforementioned methods. Given the high computation efficiency of SpatialDM, 1 core was applied for the run time. We run all other methods using 50 cores except SpaTalk. The number of spots was varied from 1,000 to 10,000 (1 million for the z-score approach of SpatialDM).

We also applied different models in the melanoma dataset with the aforementioned settings. In addition, we shuffled the curated ligand-receptor to generate a 663-pair negative control list. In theory, interactions between these LR pairs were not documented before. We applied SpatialDM and the aforementioned methods with the same settings on the negative control list for FPR comparison.

Experimental datasets and processing settings

Two datasets of different sizes and from different sequencing platforms were used to showcase the framework, including 1) Thrane's melanoma dataset (sample 1 rep 2, 293 spots, ST [27]), and 2) All intestine samples probed by Visium from Corbett et al., containing 8 slices from 3 time points and 4 donors, respectively [7]. We mainly showcased the permutation approach in the melanoma dataset (Global: FDR < 0.1, Local: p-value <0.1), and the z-score approach in the intestine dataset (Global: FDR < 0.1, Local: p-value <0.1).

Cell type annotation

For the melanoma dataset, scRNA-seq and marker gene lists of each of the 7 cell types were obtained from [31]. Cell type composition in each spatial transcriptome spots were computed using RCTD v.2.0.1 [28] based on the spot mRNA expression and the marker gene list. For the intestine dataset, we directly used cell types annotations from the original study [7]. For selected pair(s), the decomposition weights were summed up across selected spots for each cell type.

Verification in scRNA-seq

Dimension reduction was performed in scRNA-seq using tSNE. Cell type annotations were performed by Tirosh, et al. [31]. FCER2 and CR2 expressions were visualized in tSNE plots.

Additional utility analyses in SpatialDM

Histology clustering of significant pair using SpatialDE

The input here is the binary matrix of either local permutation or z-score selected spots (0 for non-significant spots, 1 for selected spots). SpatialDE.aeh.spatial_patterns function was used to cluster all filtered pairs into 4 spatial patterns.

Pathway enrichment

For selected pairs, we counted the number of pairs belonging to each pathway as documented in CellChatDB v.1.1.3 which was visualized in the dot plot x-axis. We also computed the percentage of the pairs in relation to all pairs belonging to the respective pathway in the dataset.

Differential analyses

Colon samples (A1, A2 for adult, A3, A4, A8 and A9 for foetus) and their Global Moran z-scores (1377 pairs) were extracted for differential analyses. If either ligand or receptor was not detected in a sample, the z-score was forced to 0. For each pair, linear regression models were fitted to the 6 z-scores twice, with (full model) or without (reduced model) condition information. A likelihood ratio test was performed to calculate the p value for differential communication. Specifically, the difference between log-likelihood from the full vs. reduced models was then subjected to Chi-Squared test for the differential p-values [46].

Fine tune with auto-correlation weights

With a hypothesis that spatially significant pairs will have a certain degree of auto-correlation for the ligand or receptor, we integrated ligand/receptor Moran's R in simulated data.

Auto-correlation Moran's I_l (ligand) and I_r (receptor) are defined as:

$$\begin{aligned} I_l &= \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \\ I_r &= \frac{\sum_i \sum_j w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2}, \end{aligned} \tag{8}$$

where x denotes ligand expression, y denotes receptor expression. The fine-tuned $R = w_l * I_l + w_r * I_r + R_{lr}$. We used $w_l = 0.17$ and $w_r = 0$ in the simulation data (learned from a logistic regression on a separate dataset) and used $w_l = 0, w_r = 0$ for all experimental datasets.

Data availability

All datasets used here are previously published and publicly available (Raw mRNA counts log-transformed mRNA counts, and spatial coordinates of

the melanoma data were obtained from https://github.com/msto/spatial-datasets; Raw mRNA counts and spatial coordinates of the intestine data were obtained from https://simmonslab.shinyapps.io/FetalAtlasDataPortal/, GEO: GSE158328) For easier reuse, we also included them in the SpatialDM Python package as follows,
$spatialdm.datasets.melanoma()$
$spatialdm.datasets.intestine(sample = "A1")$.

Codes availability

SpatialDM is an open-source Python package freely available at https://github.com/StatBiomed/SpatialDM. Detailed documentation and the analysis notebooks used in this paper are also included in this repository.

Acknowledgement

We thank Rio Sugimura, Martin Cheung and Langqi Gong for biological insights on discussing melanoma analyses and Chen Qiao for technical discussion on ST data modelling. We also thank Shoufa Chen and Mingze Gao for helping the Python implementation and the package releasing process.

Author contributions

YH and PL conceived and supervised the study. ZL and YH designed the project. ZL implemented the SpatialDM package and performed all data analysis, with support from TW and YH. TW derived the analytical null distributions. ZL and YH wrote the manuscript with inputs from all authors.

References

[1] Armingol, E., Officer, A., Harismendy, O., Lewis, N.E.: Deciphering cell–cell interactions and communication from gene expression. Nature Reviews Genetics **22**(2), 71–88 (2021)

[2] Bloemendal, S., Kück, U.: Cell-to-cell communication in plants, animals, and fungi: a comparative review. Naturwissenschaften **100**(1), 3–19 (2013)

[3] Efremova, M., Vento-Tormo, M., Teichmann, S.A., Vento-Tormo, R.: Cellphonedb: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. Nature protocols **15**(4), 1484–1506 (2020)

[4] Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.-H., Myung, P., Plikus, M.V., Nie, Q.: Inference and analysis of cell-cell communication using cellchat. Nature communications **12**(1), 1–20 (2021)

[5] Noël, F., Massenet-Regad, L., Carmi-Levy, I., Cappuccio, A., Grandclaudon, M., Trichot, C., Kieffer, Y., Mechta-Grigoriou, F., Soumelis, V.: Dissection of intercellular communication using the transcriptome-based framework ICELLNET. Nature communications **12**(1), 1–16 (2021)

[6] Vento-Tormo, R., Efremova, M., Botting, R.A., Turco, M.Y., Vento-Tormo, M., Meyer, K.B., Park, J.-E., Stephenson, E., Polański, K., Goncalves, A., *et al.*: Single-cell reconstruction of the early maternal–fetal interface in humans. Nature **563**(7731), 347–353 (2018)

[7] Fawkner-Corbett, D., Antanaviciute, A., Parikh, K., Jagielowicz, M., Gerós, A.S., Gupta, T., Ashley, N., Khamis, D., Fowler, D., Morrissey, E., Cunningham, C., Johnson, P.R.V., Koohy, H., Simmons, A.: Spatiotemporal analysis of human intestinal development at single-cell resolution. Cell **184**(3), 810–82623 (2021)

[8] Almet, A.A., Cang, Z., Jin, S., Nie, Q.: The landscape of cell–cell communication through single-cell transcriptomics. Current opinion in systems biology **26**, 12–23 (2021)

[9] Browaeys, R., Saelens, W., Saeys, Y.: NicheNet: modeling intercellular communication by linking ligands to target genes. Nature methods **17**(2), 159–162 (2020)

[10] Tyler, S.R., Rotti, P.G., Sun, X., Yi, Y., Xie, W., Winter, M.C., Flamme-Wiese, M.J., Tucker, B.A., Mullins, R.F., Norris, A.W., *et al.*: PyMINEr finds gene and autocrine-paracrine networks from human islet scRNA-Seq. Cell reports **26**(7), 1951–1964 (2019)

[11] Wang, Y., Wang, R., Zhang, S., Song, S., Jiang, C., Han, G., Wang, M., Ajani, J., Futreal, A., Wang, L.: iTALK: an R package to characterize and illustrate intercellular communication. BioRxiv, 507871 (2019)

[12] Cabello-Aguilar, S., Alame, M., Kon-Sun-Tack, F., Fau, C., Lacroix, M., Colinge, J.: SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. Nucleic Acids Research **48**(10), 55–55 (2020)

[13] Dimitrov, D., Türei, D., Garrido-Rodriguez, M., Burmedi, P.L., Nagai, J.S., Boys, C., Ramirez Flores, R.O., Kim, H., Szalai, B., Costa, I.G., *et al.*: Comparison of methods and resources for cell-cell communication inference from single-cell rna-seq data. Nature Communications **13**(1), 1–13 (2022)

[14] Rao, A., Barkley, D., França, G.S., Yanai, I.: Exploring tissue architecture using spatial transcriptomics. Nature **596**(7871), 211–220 (2021)

[15] Walker, B.L., Cang, Z., Ren, H., Bourgain-Chang, E., Nie, Q.: Deciphering tissue structure and function using spatial transcriptomics. Communications biology **5**(1), 1–10 (2022)

[16] Longo, S.K., Guo, M.G., Ji, A.L., Khavari, P.A.: Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. Nature Reviews Genetics **22**(10), 627–644 (2021)

[17] Dries, R., Zhu, Q., Dong, R., Eng, C.-H.L., Li, H., Liu, K., Fu, Y., Zhao, T., Sarkar, A., Bao, F., *et al.*: Giotto: a toolbox for integrative analysis and visualization of spatial expression data. Genome biology **22**(1), 1–31 (2021)

[18] Arnol, D., Schapiro, D., Bodenmiller, B., Saez-Rodriguez, J., Stegle, O.: Modeling cell-cell interactions from spatial molecular data with spatial variance component analysis. Cell Reports **29**(1), 202–2116 (2019)

[19] Cang, Z., Nie, Q.: Inferring spatial and signaling relationships between cells from single cell transcriptomic data. Nature communications **11**(1), 1–13 (2020)

[20] Shao, X., Li, C., Yang, H., Lu, X., Liao, J., Qian, J., Wang, K., Cheng, J., Yang, P., Chen, H., et al.: Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with spatalk. bioRxiv (2022)

[21] Rüttenauer, T.: Spatial regression models: a systematic comparison of different model specifications using Monte Carlo experiments. Sociological Methods & Research **51**(2), 728–759 (2022)

[22] Moran, P.A.: The interpretation of statistical maps. Journal of the Royal Statistical Society. Series B (Methodological) **10**(2), 243–251 (1948)

[23] Wartenberg, D.: Multivariate spatial correlation: a method for exploratory geographical analysis. Geographical Analysis **17**(4), 263–283 (1985)

[24] Lee, S.-I.: Developing a bivariate spatial association measure: an integration of pearson's r and moran's i. Journal of geographical systems **3**(4), 369–385 (2001)

[25] Anselin, L.: A local indicator of multivariate spatial association: Extending geary's c. Geographical Analysis **51**(2), 133–150 (2019)

[26] Svensson, V., Teichmann, S.A., Stegle, O.: SpatialDE: identification of spatially variable genes. Nature methods **15**(5), 343–346 (2018)

[27] Thrane, K., Eriksson, H., Maaskola, J., Hansson, J., Lundeberg, J.:

Spatially Resolved Transcriptomics Enables Dissection of Genetic Heterogeneity in Stage III Cutaneous Malignant Melanoma. Cancer Research **78**(20), 5970–5979 (2018)

[28] Cable, D.M., Murray, E., Zou, L.S., Goeva, A., Macosko, E.Z., Chen, F., Irizarry, R.A.: Robust decomposition of cell type mixtures in spatial transcriptomics. Nature Biotechnology **40**(4), 517–526 (2022)

[29] Aubry, J.-P., Pochon, S., Gauchat, J.-F., Nueda-Marin, A., Holers, V.M., Graber, P., Siegfried, C., Bonnefoy, J.-Y.: Cd23 interacts with a new functional extracytoplasmic domain involving n-linked oligosaccharides on cd21. The Journal of Immunology **152**(12), 5806–5813 (1994)

[30] Khan, F., Chang, C.: Human autoantibodies in urticaria, angioedema, and other atopic diseases. In: Autoantibodies, pp. 93–101. Elsevier, ??? (2014)

[31] Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., *et al.*: Dissecting the multicellular ecosystem of metastatic melanoma by single-cell rna-seq. Science **352**(6282), 189–196 (2016)

[32] Walter, R.J., Sonnentag, S.J., Munoz-Sagredo, L., Merkel, M., Richert, L., Bunert, F., Heneka, Y.M., Loustau, T., Hodder, M., Ridgway, R.A., *et al.*: Wnt signaling is boosted during intestinal regeneration by a cd44-positive feedback loop. Cell death & disease **13**(2), 1–16 (2022)

[33] Abud, H.E., Chan, W.H., Jardé, T.: Source and impact of the egf family of ligands on intestinal stem cells. Frontiers in Cell and Developmental Biology **9**, 685665 (2021)

[34] Nászai, M., Bellec, K., Yu, Y., Román-Fernández, A., Sandilands, E., Johansson, J., Campbell, A.D., Norman, J.C., Sansom, O.J., Bryant, D.M., *et al.*: Ral gtpases mediate egfr-driven intestinal stem cell proliferation and tumourigenesis. Elife **10**, 63807 (2021)

[35] Senger, S., Ingano, L., Freire, R., Anselmo, A., Zhu, W., Sadreyev, R., Walker, W.A., Fasano, A.: Human fetal-derived enterospheres provide insights on intestinal development and a novel model to study necrotizing enterocolitis (nec). Cellular and Molecular Gastroenterology and Hepatology **5**(4), 549–568 (2018)

[36] Romero, R.J., Gomez, R., Ghezzi, F., Yoon, B.H., Mazor, M., Edwin, S.S., Berry, S.M.: A fetal systemic inflammatory response is followed by the spontaneous onset of preterm parturition. American journal of obstetrics and gynecology **179 1**, 186–93 (1998)

[37] Madsen-Bouterse, S.A., Romero, R., Tarca, A.L., Kusanovic, J.P., Espinoza, J., Kim, C.J., Kim, J.-S., Edwin, S.S., Gomez, R., Draghici, S.: Original article: The transcriptome of the fetal inflammatory response syndrome. American Journal of Reproductive Immunology **63**(1), 73–92 (2010)

[38] Mittendorf, R., Montag, A.G., MacMillan, W., Janeczek, S., Pryde, P.G., Besinger, R.E., Gianopoulos, J.G., Roizen, N.: Components of the systemic fetal inflammatory response syndrome as predictors of impaired neurologic outcomes in children. American journal of obstetrics and gynecology **188**(6), 1438–1446 (2003)

[39] Hardwick, J.C., Van Den Brink, G.R., Bleuming, S.A., Ballester, I., Van Den Brande, J.M., Keller, J.J., Offerhaus, G.J.A., Van Deventer, S.J., Peppelenbosch, M.P.: Bone morphogenetic protein 2 is expressed by, and acts upon, mature epithelial cells in the colon. Gastroenterology **126**(1), 111–121 (2004)

[40] McElroy, S.J., Castle, S.L., Bernard, J.K., Almohazey, D., Hunter, C.J., Bell, B.A., Al Alam, D., Wang, L., Ford, H.R., Frey, M.R.: The erbb4 ligand neuregulin-4 protects against experimental necrotizing enterocolitis. The American journal of pathology **184**(10), 2768–2778 (2014)

[41] Kelleher, M., Singh, R., O'Driscoll, C.M., Melgar, S.: Carcinoembryonic antigen (ceacam) family members and inflammatory bowel disease. Cytokine & growth factor reviews **47**, 21–31 (2019)

[42] Saiz-Gonzalo, G., Hanrahan, N., Rossini, V., Singh, R., Ahern, M., Kelleher, M., Hill, S., O'Sullivan, R., Fanning, A., Walsh, P.T., Hussey, S., Shanahan, F., Nally, K., O'Driscoll, C.M., Melgar, S.: Regulation of ceacam family members by ibd-associated triggers in intestinal epithelial cells, their correlation to inflammation and relevance to ibd pathogenesis. Frontiers in Immunology **12** (2021)

[43] Mund, A., Coscia, F., Kriston, A., Hollandi, R., Kovács, F., Brunner, A.-D., Migh, E., Schweizer, L., Santos, A., Bzorek, M., et al.: Deep visual proteomics defines single-cell identity and heterogeneity. Nature Biotechnology, 1–10 (2022)

[44] Liu, Y., Yang, M., Deng, Y., Su, G., Enninful, A., Guo, C.C., Tebaldi, T., Zhang, D., Kim, D., Bai, Z., *et al.*: High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. Cell **183**(6), 1665–1681 (2020)

[45] Cheng, J., Zhang, J., Wu, Z., Sun, X.: Inferring microenvironmental regulation of gene expression from single-cell rna sequencing data using scmlnet with an application to covid-19. Briefings in bioinformatics **22**(2),

988–1005 (2021)

[46] Seabold, S., Perktold, J.: statsmodels: Econometric and statistical modeling with python. In: 9th Python in Science Conference (2010)