Harnessing Agent-Based Modeling in CellAgentChat to Unravel Cell-Cell Interactions from Single-Cell Data

Vishvak Raghavan[1,2] and Jun Ding[1,2*]
[1]School of Computer Science, McGill University, Montreal, Canada
[2]Meakins-Christie Laboratories, Department of Medicine, McGill University, Montreal, Canada

[*]All correspondence should be addressed to jun.ding@mcgill.ca

Abstract

Deciphering cell-cell interactions is crucial yet challenging, given the inherent intricacy and heterogeneity of cellular dynamics within biological systems. Existing methodologies, primarily those grounded in statistical or machine-learning approaches, often demand intricate mathematical representations that limit their efficacy in biological settings. Herein, we introduce CellAgentChat, a pioneering application of agent-based modeling (ABM) devised for unraveling cell-cell interactions from single-cell RNA sequencing and spatial transcriptomics data. Our approach encapsulates biological systems as collections of autonomous agents, subject to clear, decipherable rules, and adeptly sidesteps conventional restrictions. Tested against three diverse single-cell datasets, CellAgentChat proved its robustness in delineating intricate signaling events spanning varied cell populations. A unique attribute of our model lies in its ability to yield animated visualizations of single-cell interactions and the flexibility to modify agent behavior rules, thereby enabling comprehensive exploration of both proximal and remote cellular communications. Furthermore, by leveraging the inherent ABM characteristics of CellAgentChat, we can perform intuitive in silico perturbations of cellular interactions through agent rule adjustments, thus creating opportunities for novel intervention strategies. This cutting-edge ABM approach constitutes a significant stride towards a holistic understanding of cellular signaling interactions across a broad spectrum of biological contexts.

Keywords:

Cell-cell interaction, Agent-based model, Single-cell, Spatial transcriptomics data, Cellular dynamics, In silico perturbation.

INTRODUCTION

As the fundamental units of most life forms, cells are dynamic and continuously adapt their states in response to changes in their microenvironment[1,2]. In multicellular organisms, cellular activities rely on the coordination of cell-cell interactions (CCIs) across different cell types and tissues[3-6]. CCIs are critical in numerous biological processes, including cell differentiation[7], disease pathogenesis[8], and immune responses[9]. Signaling molecules, such as ligands and receptors, facilitate cellular interactions and influence cell behavior, leading to dynamic changes in cellular functions[3]. By studying these interactions, we can gain insights into the mechanisms underlying biological processes, such as organ development and tumor growth, ultimately informing the study of various disorders and diseases and aiding in the identification of potential diagnostic and therapeutic strategies[9].

However, inferring CCIs in biological processes is challenging due to cellular heterogeneity in most biological systems[10]. The advent of single-cell sequencing, particularly single-cell RNA-sequencing (scRNA-seq) and single-cell spatial measurements like 10x genomics Visium[11], Slide-Seq[12], and Stereo-seq[13], offers unparalleled opportunities to investigate CCIs in diverse biological processes. Numerous tools have been developed in recent years for CCI inference by integrating known ligand-receptor pairs with gene expression data from scRNA-seq data, including state-of-the-art methods such as CellPhoneDB[14], CellChat[15], and SingleCellSignalR[16]. Despite advancements, earlier methods and versions primarily rely on unimodal scRNA-seq data, which restricts their capacity to infer CCIs due to the absence of spatial distance information between cells[14-16]. As CCIs in tissue environments are heavily influenced by spatial structures and can occur at both long and short ranges[17], integrating spatial information is essential for studying paracrine and endocrine signaling. Thus, several recent methods, such as Giotto[18] and CellChat version 1.6.0[15], have been developed for inferring cellular interactions from single-cell spatial measurements[19].

Existing methods still face significant drawbacks, whether or not they integrate spatial information. Most methods primarily focus on cell population levels, inferring cell-cell communication between cell populations (clusters) while ignoring within-cluster diversity. This oversight hampers the accurate discovery of CCIs in various biological processes. For example, cancer cells often interact with each other in the tumor microenvironment via cell adhesion molecules to control cell survival and lineage[20]. Further, incorporating pseudotime ordering of cells within cell clusters can improve CCI inference[21]. Besides, most existing approaches lack a view of cellular interactions for individual cells within the same cluster. Thus, existing methods often neglected the heterogeneous cellular interaction within the same cell population. Additionally, most existing methods rely on statistical or machine learning approaches that necessitate mathematical descriptions of complex biological systems, often requiring numerous assumptions, which inevitably limits the CCI inference capabilities of these approaches[14-16,22].

Due to their unique characteristics, agent-based models (ABMs) are well-suited for exploring complex CCIs and their impact on spatial and temporal cellular dynamics in biological systems. First, ABMs capture emergent phenomena arising from collective agent interactions, revealing insights into cellular community function[23,24]. Second, they represent heterogeneous cells as agents with different statuses, enabling accurate modeling of diverse cell populations and capturing variations in cell types, states, and behaviors critical for understanding biological processes[23,24]. Third, ABMs can explicitly integrate temporal and spatial information, which is crucial for studying complex biological processes like tissue development, immune responses, and tumor growth[23,24]. Fourth, rule-based interactions in ABMs allow direct incorporation of domain-specific knowledge, enhancing the biological relevance and interpretability of simulated CCIs[23,24]. Lastly, ABMs enable easy and effective in silico perturbations, allowing predictions of changes in cellular interactions based on perturbed cell agent behavior rules, such as blocking a receptor.

Although ABMs hold great potential, to our knowledge, researchers have not yet developed them for inferring cellular interactions from single-cell data. To address this gap, we developed CellAgentChat, an innovative ABM for inferring and visualizing CCIs from scRNA-seq and

spatial transcriptomics data. Due to the inherent advantages of ABMs, CellAgentChat offers versatile functionalities for CCI inference and visualization.

CellAgentChat presents a unique agent-based perspective on cellular interactions, seamlessly integrating temporal, spatial, and biological data, offering a more precise and comprehensive understanding of cellular interaction dynamics. The approach enables in silico perturbations and in-depth analysis of the effects of cellular interactions on downstream gene expression, which could facilitate the identification of potential therapeutic receptor targets. Our evaluation of CellAgentChat, using several publicly available single-cell sequencing datasets, revealed superior CCI inference accuracy compared to other benchmarked methods. Notably, the flexible CellAgentChat framework permits diverse functionalities by adjusting the behavior rules associated with cell agents, a feature not attainable with traditional statistical or machine learning methods. This approach sets the stage for future advancements in the study of cellular interactions with significant implications for our understanding of complex biological processes.

RESULTS

Overview of CellAgentChat

CellAgentChat constitutes a comprehensive framework integrating gene expression data and existing knowledge of signaling ligand-receptor (LR) interactions to compute the probabilities of cell-cell communication. Figure 1 schematically outlines the CellAgentChat methodology. Utilizing the principles of agent-based modeling (ABM), we describe each cell agent through various attributes, including cell identification (e.g., cell type or clusters), gene expression profiles, LR universe, spatial coordinates (optional), and pseudotime (optional) (Fig. 1a). We quantify cellular interactions between sender and receiver cells based on the volume of ligands secreted by the sender cells and subsequently received by the receiver cells. This process hinges upon three interrelated components: ligand diffusion rate (γ_l), receptor receiving rate (α_r), and receptor conversion rate (β_r) (Fig. 1b).

The initial step in our approach involves the computation of the ligand diffusion rate (γ_l), which quantifies the ligands secreted by the sender cell that reach the target receiver cell. This rate is a function of both ligand expression and intercellular distance. In the subsequent step, CellAgentChat estimates the receptor receiving rate (α_r), which signifies the fraction of secreted ligands received by the receptor on the receiver cell. We then leverage a neural network to evaluate the receptor conversion rate (β_r), which measures the potency of a cellular interaction by studying its effect on downstream gene expression dynamics. Considering the possibility that a similar interaction strength between LR pairs can yield diverse impacts on downstream gene expression dynamics[3], it becomes crucial to assess the conversion rate for each receptor. To the best of our knowledge, we are the first to estimate the downstream gene expression from CCIs and integrate this information into the interaction inference. We quantify the interaction between a specific LR pair from sender and receiver cell agents using LR scores, which depend on the above rates (Fig. 1b). The LR score for two cluster pairs is the Interaction Score (IS) (Fig. 1b). We consider an IS for an LR pair significant if its IS score significantly surpasses the background score derived from a permutation test (Fig. 1c). The cumulative interaction between two cell clusters is computed as

the total number of significant LR pairs. The ABM framework further allows us to adjust the agent behavior rules, facilitating the exploration of long and short-range cellular interactions and the effects of in silico receptor blocking (Fig. 1d). This approach also supports the agent-based animation and visualization of cellular interactions associated with individual cells within the biological system, enhancing the interpretability and visualization of the modeling outcomes (Fig. 1d).

CellAgentChat identifies cellular communications in Human Squamous Cell Carcinoma

An application of CellAgentChat to a dataset of Squamous Cell Carcinomas (SCC) demonstrates its capabilities and performance. The dataset utilizes single-cell spatial transcriptomics data to accurately define the cellular transcriptomics and cellular composition of SCC in humans[25]. Although many single-cell spatial transcriptomics data remain at the spot level (often comprised of dozens of cells), here we did not distinguish cells and spots and thus term them all as cell agents (and cells) for easier description. We preprocessed the cells using the Scanpy[26] pipeline to filter for highly variable genes and cluster the cells into cell populations (i.e., clusters representing different cell types or sub-types). The dataset comprises 3,295 cells from a total of thirteen distinct cell populations. Among them are three normal keratinocytes (NKer), two tumor keratinocytes (TKer), and three myeloid dendritic cells (MDC) (Fig. 2a, Supplementary Fig. 1a). Additionally, the dataset includes several other cell types, such as endothelial (Endo), fibroblast (Fib), T-cells, B-cells, and melanocytes (Mel).

CellAgentChat will calculate the interaction score based on the spatial coordinates of cells and the distance between individual cells when spatial information is provided (Fig. 1). When considering the spatial coordinates of cells, CellAgentChat identified 9163 significant interactions involving 1571 LR pairs (Fig. 2b, d). Only a few cell types were involved in these interactions. NKer3, MDC3, T cells, and TKer2 participated in the most interactions. At the same time, Mel appeared to be farther away from other cell types, resulting in fewer interactions (Fig. 2b). Additionally, CellAgentChat found that the LR pair involved in the most interactions was GRN-TNFRSF1B. (Fig. 2d, Supplementary Fig. 2d). Both factors play a crucial role in the progression and growth of SCC tumors[27,28].

By incorporating both spatial data and pseudotime ordering of cells in the SCC dataset, CellAgentChat was able to identify 1477 significant interactions involving 810 LR pairs. (Fig. 2c, e). The top three interaction pairs involve self-interactions among T cells, Mel, and TKer2 (Fig. 2c). The most significant LR pair is LAMC2-ITGB4 (Fig. 2e), both of which are highly expressed in SCC and have essential roles in tumor progression and migration[29-31]. This finding is consistent with the high expression of these genes in TKer2 cells (Supplementary Fig. 2e), suggesting an early stage of cancer migration.

Since spatial information is not always provided, we utilized CellAgentChat with scRNA-seq data (see Supplemental Results). The analysis highlighted specific cell types, such as MDC subpopulation three and TKer subpopulation two, as prominent receiving and transmitting cells, respectively, within the communication network (Supplementary Fig. 1b, c). The LR pair HLA-B-CANX stood out as the leading LR interaction, showcasing robust expression of HLA-B across all

cell types and notably elevated CANX expression in TKer2 (Supplementary Figs. 1c, 2b). This compelling association has been demonstrated to be linked with SCC[32-34]. We also addressed the challenge of overlooking cell heterogeneity within the same population by incorporating pseudotime ordering (Supplementary Fig. 1e, f). The MMP1-CD44 pair was involved in numerous interactions, with MMP1 highly expressed in tumor keratinocyte subpopulations and CD44 overexpressed in cancer cells (Supplementary Figs. 1f, 2c)[35,36]. These findings align with previous studies linking MMP1 expression to squamous cell carcinoma, highlighting the significance of CD44 in cancer cells[37,38].

To systematically assess the reliability of the LR interaction pairs predicted by CellAgentChat, we employed reactome pathway enrichment analysis. We compared the enrichment P-values of these pathways' terms using the significant LR pairs with high scores (see methods). Many pathways associated with the LRs predicted by CellAgentChat are relevant to the disease of interest. The most significant reactome pathway in all scenarios was extracellular matrix (ECM) organization (Fig. 2f, 2g, Supplementary Fig. 1d, 1g). Additional pathways involving ECM interactions had high adjusted P-values, indicating that these interactions play a role in SCC. Studies have shown that the ECM regulates cell behavior and tumor progression in SCC[39]. Additionally, changes in ECM organization occur early in disease progression[40], indicating that this patient may have an early stage of SCC. The hemostasis pathway was highly observed in all scenarios (Fig. 2f, 2g, Supplementary Fig. 1d, 1g) and is associated with cancer, including SCC[41]. Further, the immune system was present in two of the four scenarios (Fig. 2f, Supplementary Fig. 1d). Cancer patients often experience an elevated immune response, with T-cells playing a critical role in recognizing and responding to tumor antigens in all types of cancers, including SCC[42]. We observed cytokine signaling (Fig. 2f, Supplementary Fig. 1d), which plays a role in the immune response during cancer development and progression, including the suppression of tumor cells[43]. MET signaling pathway regulates a range of downstream processes, including proliferation and invasiveness of tumor cells, and researchers have associated it with poor prognosis in SCC[44,45]. ERBB2 signaling is highly expressed in cancer and is associated with poor prognosis in patients[46,47]. These pathways were observed in pathway enrichment analysis along with kinase activation and tyrosine activity pathways, which CellAgentChat also predicted (Fig. 2f, 2g Supplementary Fig. 1d, 1g). Previous studies on SCC cell communication also identified these pathways[48].

Amongst the pathways shared between pseudotime and non pseudotime modes (marked with a blue star), the $-\log_{10}$ adjusted P-values are more significant when using pseudotime and spatial information than with just spatial information (Fig. 2f, g). Indeed, the results with pseudotime are also superior without using spatial information (Supplementary Fig. 1d, g). Integrating pseudotime and spatial data provides a more comprehensive understanding of cellular behavior and its relationship to tissue biology. By combining information about the trajectory of cells (pseudotime) and their spatial organization within a tissue, researchers can identify significant interactions that consider both spatial distance and temporal progression. This integrated approach holds greater functional relevance to the key reactome pathways associated with tissue biology.

CellAgentChat enables the agent-based visualization of cellular interactions associated with individual cells

We employed an ABM approach through CellAgentChat, allowing us to examine individual cell interactions rather than solely focusing on cell populations. This approach sets CellAgentChat apart from other available methodologies. We applied the method to a human pancreatic ductal adenocarcinoma (PDAC) dataset[49], which profiles single-cell gene expression and spatial positions. After preprocessing the PDAC dataset, we discerned six distinct cell populations: ductal (Duct), three subpopulations of acinar cells (Acin1, Acin2, and Acin3), stromal (Strom) cells, and two classes of cancer cells (Can1 and Can2) (Supplementary Fig. 3a). To effectively exhibit the insights that CellAgentChat can provide into individual cell interactions, we distinguished between the twenty most interacting cells (group A, red circles) and the least interacting cells (group B, black circles) within the second cancer cell class (green cells) (Fig. 3a).

The cells with higher levels of interactions generally reside in the outer layer of the tissue (invasive front), consistent with the discovery from many existing studies[50]. These two cell groups within the same cluster also demonstrated significant transcriptomic differences. Group A showed higher mean expression levels of receptors involved in the top 25 LR pairs than Group B (Fig 3b, Supplementary Fig. 3g). For example, CD63 was highly expressed in both groups, with group A having the highest expression levels (Fig. 3b). Additionally, receptors such as LRP1, SDC1, ITGA3, NOTCH3, and CD74 displayed notable differences in mean expression between the two groups, with group A again exhibiting higher expression (Fig. 3b). Next, we systematically explored the differentially expressed (DE) genes of these two groups. To identify the DE genes for each group (A and B), we compare the cells within the group with the rest of the cells in the dataset. The differential genes associated with group A show significantly higher expression in group A compared with group B (P-value=1.22×10^{-48}). In contrast, the expression of group B DE genes is relatively similar across groups (Fig. 3c, d).

We showed the top 50 DE genes for both groups by comparing the expression of cells in groups A and B (Fig. 3d). While both groups showed high expression levels of S100A6 and ACTB, group B cells exhibited elevated expression of KRT19 (Fig. 3d). Conversely, COL1A1, responsible for type I collagen production, was particularly prominent in group A cells (Fig. 3d). Notably, COL1A1, a gene found enriched in PDAC tumor cells, is associated with tumor growth and metastasis in cells[51]. To identify the biological functions associated with these cell groups, we conducted reactome pathway enrichment on the top 50 DE genes for both groups (Fig. 3e). Intriguingly, group A was associated with several pathways related to the extracellular matrix (ECM) and collagen, known to be crucial in PDAC tumor progression[52]. On the other hand, group B cells showed lesser involvement in ECM-related pathways, aligning with their lower level of interaction compared to group A (Fig. 3e).

Our findings underscore the unique behaviors exhibited by individual cells within a subclass, such as cancer cells. These behaviors, effectively captured through CellAgentChat, suggest that group A cells, demonstrating stronger cellular interactions, might be in a more advanced cancerous stage than group B cells, as indicated by their increased ECM and collagen-associated functions. It reinforces the idea that studying individual cell behaviors can unveil critical insights into disease

mechanisms, which might otherwise go unnoticed when studying cell types in aggregate. As such, tools like CellAgentChat can offer a more nuanced understanding of complex cellular interactions that underpin disease processes, allowing for more precise interrogation of gene expression and signaling activity variations, even within seemingly homogeneous cell populations.

CellAgentChat enables effective in silico receptor blocking for analysis of perturbed downstream genes

The unique feature of ABMs is their inherent adaptability, which facilitates straightforward manipulation of cell agent behavior rules. This adaptability facilitates the effective execution of in silico receptor blocking perturbations. Such perturbations are often challenging to implement in traditional models that require an accurate mathematical representation of complex biological systems. In this context, we put the broad applicability of our tool, CellAgentChat, on display, illustrating how simple agent behavior adjustments can unlock powerful in silico permutations.

We harnessed CellAgentChat to build upon previous research exploring the PDAC dataset, aiming to identify significant receptors related to the disease and cultivate innovative therapeutic approaches. For our in silico permutations, we chose to exclusively utilize scRNA-seq data, abstaining from spatial transcriptomic data and pseudotime cell ordering, in line with scRNA-seq's standing as the method of choice in most contexts (Supplementary Fig. 3c, d). The first step of our process entailed using CellAgentChat to pinpoint key receptors involved in LR interactions, specifically those contributing to the most notable cross-talks in our data—between two cancer subpopulations (Supplementary Fig. 3d, e). We then simulated the blocking of these receptors individually, akin to the actions of receptor inhibitor drugs. To accomplish this, we set the receptor's receiving rate to zero, thereby altering the communication behavior of the agents.

Following this, we deployed the previously trained deep learning model designed to calculate each receptor's conversion rate (Fig. 4a). Employing this novel approach, we predicted the potential impact of receptor blocking on downstream gene expression. Specifically, we identified the most perturbed downstream genes for each in silico blocked receptor—defined as the top 50 target genes most impacted by each receptor's blocking (Fig. 4a). We then proceeded to examine these target genes through diseaseGeNet[53] enrichment analysis to assess potential therapeutic benefits of receptor blocking (see methods) (Fig. 4a). From the cellular interaction analysis, we identified three receptors, CD81, LRP1, and CD44 (henceforth termed candidate receptors), which exhibited the highest -log10 P-values (FDR corrected, <0.05) for PDAC, all demonstrating statistical significance (Fig. 4b). This suggests that in silico blocking of the candidate receptors significantly shifts the expression of PDAC-associated disease genes reported in diseaseGeNet. Additional analysis revealed that, when blocked, LRP1, CD81, and CD151 yielded statistically significant P-values (<0.05; with LRP1 registering a P-value <0.01) for a disease similar to PDAC, Pancreatic Carcinoma (PC) (Fig. 4b).

The receiving score of each cell, as identified from our animation platform, depicts that Acin3, Strom, Duct, Can1, and Can2 cells exhibit the highest levels of involvement in interactions with the candidate receptors (Fig. 4c). We have found significant interactions between these populations, which supports this conclusion (Supplementary Fig. 3c, d). While ductal cells

traditionally serve as the primary origin of PDAC, emerging evidence suggests that acinar cells may also contribute to the development of PDAC[54]. This intriguing possibility could explain the extensive involvement of both Duct and Acin3 cells in interactions with all three receptors (Fig. 4c). Furthermore, studies have demonstrated that stromal cells significantly influence extracellular matrix (ECM) formation and tumor progression in PDAC[55]. Through an agent-based view, we gain robust confirmation of the extensive involvement of Strom, Duct, Can1, and Can2 cells in the context of PDAC. Notably, the agent-based approach reveals even more pronounced interactions with Acin3 cells, further enriching our understanding of the intricate dynamics at play in this disease.

Although studies have associated all of the receptors examined in this study with PDAC, the candidate receptors we identified—CD81, LRP1, and CD44—appear to play a particularly crucial role in PDAC development[56-58]. This finding validates the efficacy of the in silico receptor blocking technique facilitated by CellAgentChat. Blocking the CD81, LRP1, and CD44 receptors in silico perturbs the disease genes associated with PDAC (from diseaseGeNET) (Fig. 4d, e). In particular, when LRP1 was blocked, ST6GAL1 emerged as the second most affected gene, previously linked with PDAC (Fig. 4e)[59]. Similarly, HRAS, another gene associated with PDAC[60], was the fifth most affected gene when CD44 was blocked (Fig. 4e). All other PDAC disease target genes perturbed by blocking the candidate receptors are depicted in Figure. 4e, highlighted in red.

CellAgentChat can be tuned to target long and short-range interactions

Leveraging CellAgentChat, we embarked on an investigation of cellular communication within a human prostate cancer dataset. CellAgentChat discerned 41,972 LR interactions spanning 15 cell types (Supplementary Fig. 4a, b, d, e) while focusing solely on pseudotime. Upon incorporating spatial transcriptomics data, we identified 2,787 LR interactions (Supplementary Fig. 4c, f, g).

CellAgentChat extends its analytical capabilities beyond intercellular interactions, enabling the discernment of interactions occurring at varying distance ranges. By modeling these distance ranges, we can identify long- and short-range cellular communications, which can help us further understand the tumor microenvironment. The ABM employed by CellAgentChat provides nuanced representations of cell agent behaviors, including the diffusion of signaling ligands. As a result, this facilitates single-cell level modeling of both long- and short-range interactions. By manipulating the delta factor (δ), which influences the degree of cell-to-cell distance, CellAgentChat can model interactions at varying ranges (refer to methods). Modeling interactions at various distance ranges may provide a better understanding of the stages of tumor progression. For example, short-range interactions are more prevalent during the early stages of tumor development[61]. As the tumor matures, longer-range intercellular communication comes into play and impacts the overall growth of the tumor[61]. For our study, we simulated long-range interactions using a δ value of 0.5 (Fig. 5a, b) and short-range interactions with a δ value of 5 (Fig. 5c, d). A larger δ implies faster decay of ligand diffusion with increasing distance.

Our findings suggest that CellAgentChat identifies a higher number of interactions attributed to an increased diffusion rate in modeling long-range interactions. The resulting communication network appears similar to one observed in the absence of spatial information (Fig. 5a,

Supplementary Fig. 4d). Conversely, when modeling short-range interactions, the network resembles what we would observe in the "normal" range, albeit with slightly stronger self-interactions (Fig. 5c, Supplementary Fig. 4f). This is due to proximal cells within the same cluster exhibiting stronger interactions. Following this, we explored and compared the top 25 LR interaction pairs for short-range and long-range distances (Fig. 5b, d). The long-range mode (δ=0.5) revealed eight LR pairs not within the top 100 LR pairs in the short-range mode (δ=5). Notably, these included interactions involving VEGFA, such as VEGFA-ITGAV and VEGFA-GPC1, identified using the long-range mode (Fig. 5b). Given that VEGFA originates from endocrine glands and targets its receptors through endocrine signaling[62], this is an intriguing observation. Additionally, two poorly ranked interactions in short-range mode, CALR-LRP1 and CALR-ITGAV, ranked among the top 25 LR pairs in the long-range mode (Fig. 5b). In response to different stress stimuli, including inflammation and hypoxia, CALR is actively released into the extracellular space to reach its targets[63]. These interactions play significant roles in prostate cancer signaling pathways[64]. In our short-range mode, we detected LR pairs involving HSPG2, such as HSPG2-PTPRS and HSPG2-FGFR1 (Fig. 5d). Previous research has linked these pairs to cell adhesion processes occurring over short distances[65]. We further employed our animation platform to visualize the receiving strength of individual cells for the HSPG2-FGFR1 LR pair in the short-range mode originating from cluster 4 cells (green cells) (Fig. 5e). The results highlighted stronger ligand reception amongst cells close in proximity to the cluster 4 cells represented by larger sizes (within the black circle), compared to cells situated further away (red, green, and blue circles). This observation reinforces that modeling short-range interactions boosts self-interaction among cells within the same cluster.

Finally, we conducted pathway enrichment comparisons of the LR interaction pairs predicted by CellAgentChat for both long- and short-range interactions (Fig. 5f). Remarkably, we found cytokine signaling to be the most enriched pathway in the short-range interaction analysis, while it was not in the long-range interaction analysis (Fig. 5f). This aligns with our understanding of cytokine signaling in the immune system, which primarily functions through autocrine and short-range paracrine interactions[66]. Moreover, pathways like ER phagosome[67], antigen presentation, and processing[68], known to occur at short distances, emerged in the short-range mode analysis (Fig. 5f). In summary, our analysis demonstrates that CellAgentChat proficiently discerns between and accurately represents both short-range and long-range interactions. Furthermore, the tool proved adept at identifying the associated functional properties of these interactions. The biologically relevant and accurate LR interaction pairs identified by CellAgentChat validate the tool's utility, bolstering our understanding of cellular communication dynamics within the context of human prostate cancer.

Benchmarking of CellAgentChat with other cell–cell communication inference tools

We undertook a comparative analysis of CellAgentChat against two prominent CCI inference methods based on scRNA-seq data: CellPhoneDB (version 4) and CellChat (version 1.6.0). CellPhoneDB and CellChat predicate their inference on aggregated gene expression within cell clusters, while CellAgentChat harnesses an ABM approach for a more encompassing view of individual cells and their interactions. Although each employs distinct methodologies, all three

methods can integrate spatial information for CCI inference. CellChat leverages spatial coordinates from Visium 10x data, while CellPhoneDB necessitates an understanding of the microenvironment of cells, restricting interaction inference to cells within shared microenvironments. Among the three, only CellAgentChat can accommodate pseudotime trajectory calculations. More importantly, only CellAgentChat provides an agent-based animation for the cellular interactions for individual cells and empowers the use of in silico receptor blocking for exploring potential effective interventions.

We examined the cell-cell communication networks generated by CellAgentChat with pseudotime but sans spatial data, comparing them with the outcomes from CellChat and CellPhoneDB. Across all three datasets, CellAgentChat's results demonstrated consistency with the communication networks inferred by CellPhoneDB and CellChat, with all significant P-values < 0.01 (Fig. 6a-c). Each method forecasted NKer3, MDC3, T-Cells, and TKer2 as the most dominant cell types in the communication network within the SCC dataset (Fig. 6a). Furthermore, in the human prostate cancer dataset, clusters 0, 1, and 8 displayed minimal cell-cell communication across all three methods (Fig. 6b). Interestingly, both CellAgentChat and CellChat highlighted Acin3 as dominant receivers within the PDAC dataset (Fig. 6c). Conversely, CellPhoneDB anticipated a more equitable interaction distribution between Can2, Acin3, and Strom cells (Fig. 6c). Furthermore, CellPhoneDB overlooked self-interaction within cells in the same cluster, a factor captured by both CellAgentChat and CellChat.

We conducted computational validations, as previously described, comparing the enrichment P-values for relevant pathways using ligands and receptors from the top 100 significant LR pairs from each method. The pathway enrichment analysis underscored CellAgentChat's superiority over other methods. Specifically, in the SCC dataset, CellAgentChat outperformed CellChat and CellPhoneDB (Fig. 6d, left). In the human prostate cancer dataset and the PDAC dataset, CellAgentChat performed comparably to CellChat while still outshining CellPhoneDB (Fig. 6d, middle and right).

Additionally, we conducted a comparative analysis that included spatial data to evaluate the performance of CellAgentChat, CellChat, and CellPhoneDB. CellAgentChat and CellChat exhibited consistent communication networks with significant P-values for the SCC and human prostate cancer datasets (Supplementary Fig. 6a, b). However, CellChat could not utilize spatial inference for the PDAC dataset. CellPhoneDB generated slightly different networks for SCC and PDAC datasets and a significantly different network for the human prostate cancer dataset (Supplementary Fig. 6a-c). Pathway enrichment analysis supported the superior performance of CellAgentChat over CellPhoneDB and comparable performance with CellChat when using spatial data (Supplementary Fig. 6d).

DISCUSSION

We introduced CellAgentChat, an innovative agent-based model (ABM) capable of elucidating cell-cell interactions (CCI) from single-cell data, both with and without spatial information. Differing from established methodologies that depend on exact mathematical or statistical characterizations of intricate biological systems, CellAgentChat employs individual cell agents

guided by simple behavior rules to investigate the arising complexity of cellular interactions within the system. Demonstrated across three diverse single-cell datasets, CellAgentChat has proven its efficacy in determining CCIs, identifying key ligand-receptor pairs known to play crucial roles in the studied diseases. As an illustration, our findings reveal a significant LAMC2-ITGB4 interaction among tumor keratinocytes (TKer cells) in squamous cell carcinoma (Fig. 2e). This interaction has previously been linked to critical roles in tumor progression and migration, suggesting its involvement in the early stages of cancer development[29-31]. Further, we validated LRP1, CD44, and CD81 as key receptors in pancreatic ductal adenocarcinoma (Fig. 4). We have also shown that integrating pseudotime with spatial data provides a more comprehensive understanding of cellular behavior and its relationship to tissue biology.

Uniquely characterized as an ABM for predicting cellular interactions, CellAgentChat boasts distinct advantages over traditional methods. Firstly, it offers a dynamic visualization of cellular interactions related to individual cell agents. Through CellAgentChat's cellular interaction animation, one can explore cellular heterogeneity within a population, often uncovering insights at the cluster level that are impossible to obtain. Thus, CellAgentChat provides both a cluster- and cell-specific view of cellular communication. Secondly, CellAgentChat enables efficient in silico receptor blocking, facilitated by modulating the agent rules (for instance, receptor receiving rate), allowing for a systematic assessment of a receptor blocking's impact on downstream gene expression dynamics. CellAgentChat can propose systematic strategies for potential therapeutic interventions by investigating disease-related genes. Finally, researchers can tailor CellAgentChat to prioritize long- or short-range interactions controlled by the diffusion rate employed in quantifying cellular interactions. Using this approach, we identified HSPG2-FGFR1 as a short-range interaction involved in cell adhesion signaling pathways in human prostate cancer and a long-range endocrine signaling interaction CALR-LRP1 in response to cancer-related inflammation (Fig. 5).

With the increasing availability of single-cell data, especially single-cell spatial transcriptomics, novel technologies and datasets promise unprecedented opportunities for understanding cellular interactions in complex biological systems. CellAgentChat, an ABM not reliant on precise descriptions of these systems, is well-suited for a wide range of application scenarios. Its ability to comprehensively examine cellular interactions from multiple perspectives, including individual cells and long- or short-range interactions, provides unmatched potential for deeper insights into cellular dynamics. Moreover, in silico receptor blocking, achieved by perturbing agent rules, allows for identifying effective biological system interventions. As an instance, CellAgentChat can highlight potential receptor inhibitors for therapeutic purposes through its in silico receptor blocking capability.

ABMs are highly suitable for investigating various aspects of single-cell genomics and multi-omics research. By representing cells as individual agents with distinct statuses, ABMs enable precise modeling of heterogeneous cell populations, effectively capturing the diversity in cell types, states, and behaviors essential for comprehending biological phenomena[23,24]. Expanding CellAgentChat's capabilities to include cell movement and even cell death would be natural extensions, enhancing its capacity to simulate realistic scenarios. Although technically

challenging, further improvements by incorporating ligands, receptors, and other cofactor molecules as agents in addition to cells would allow for a more comprehensive and detailed simulation of biological systems' complex interactions and dynamics. By considering these additional components as agents, the model could capture the intricate interplay between cells and their surrounding molecular environment, providing a more nuanced understanding of the underlying mechanisms at play. It would also improve the efficiency of in silico perturbations such as receptor blocking.

In conclusion, CellAgentChat offers a robust framework for inferring and deepening our understanding of CCIs from single-cell data. CellAgentChat, therefore, emerges as a potent tool in the arsenal of researchers dedicated to studying intricate cellular interactions and their potential therapeutic implications.

METHODS

Ligand-Receptor database

In this study, we used the ligand-receptor (LR) interactions from CellTalkDB[69], a curated database of almost 3400 human and over 2000 mouse literature-supported interactions, as the default LR interaction resource for our CellAgentChat method. The interactions from the curated database specify the universe of all LR interactions that can occur. Besides the default CellTalkDB LR interactions, CellAgentChat also allows for customized LR interactions that users obtain from other sources or manually curated.

Single-cell sequencing data preprocessing

We preprocessed the single-cell RNA-sequencing (scRNA-seq) data using the Scanpy pipeline[26]. We began by filtering out cells with more than 20% mitochondrial reads, which are indicative of poor-quality cells[70,71]. We also removed genes not expressed by a minimum of 3 cells. Finally, cell counts are normalized against the library size and transformed into log space (with the *log1p* function). The spatial coordinates of each cell are incorporated into the "anndata" object (in h5ad format) with the gene expression.

Scoring cell-cell interactions with the agent-based model

To build the agent-based model (ABM), we need to define the state and behavior of each agent. Agents here represent the cells in the biosystem. The agent states describe the state and feature associated with each cell, while the agent rules define how each agent behaves.

<u>Cell Agent State Description</u>

As we are primarily interested in studying the interactions between cells (or clusters), we describe each cell agent with characteristics and features related to cellular interactions. The remainder of this section provides the detailed cell agent state description. (1) *Cell Agent Expression*: The expression value of all genes, including ligands and receptors, is stored for each cell. The gene expression profile presents the quantitative description of the cell agent state, which is essential to infer the interaction behaviors. (2) *Cell Agent Spatial Position*: We use spatial transcriptomics data to describe the position of each cell agent. All the cells are placed on the grid at their spatial

coordinate (if provided). Otherwise, CellAgentChat randomly distributes all cells on the grid, and it sets the degree of distance (τ) as 0, ignoring the cell distances in the subsequent CCI inference steps. (3) *Cell Agent Pseudotime*: Even the cells within the same cluster exhibit differences, in many continuous biological processes, such as cell differentiation or disease progression. Pseudotime provides ordering of the cells within the cluster (and the entire dataset). The interactions of cells with varying pseudotime may differ substantially, even if they belong to the same cluster[21]. Therefore, here we include the pseudotime as one cell agent state description. In order to calculate the pseudotime trajectory of each cell, we use Slingshot[72], a cell lineage and pseudotime inference tool. Using the pseudotime values generated from Slingshot, we group each cluster's cells into different pseudotime bins ("stages"). The user sets the number of bins (default = 10). CellAgentChat stores the pseudotime bin that each cell resides in and uses it to calculate the interaction score for the single-cell data with spatial information. (4) *Cell Agent Ligand-Receptor Universe*: Every cell agent has the same LR universe using interactions curated from CellTalkDB or user-customized LR interactions. Users can modify the LR Universe by blocking a specific receptor or LR interaction (see Agent-Based Animation of CCIs). By doing so, they can remove all interactions associated with ligands or receptors (or specific LR interactions) from the universe of all cells. This unique characteristic of ABMs enables effective and flexible ligand/receptor perturbations.

Cell Agent Communication Behavior Rules

In this study, the cellular communications between agents (cells) depend on the states of the cells, which includes the gene expression of the cell agent, the spatial position of the cell agent (optional), the pseudotime of the cell agent (optional), and the LR universe. To be more specific, the cellular interaction strengths (scores) between cell agents will be quantified based on the set of defined rules as discussed below.

(1) Ligand Diffusion Rate (γ_l)

The diffusion rate γ_l estimates the rate of the diffusion of a ligand l in the microenvironment. The diffusion rate depends on a few factors. First, the diffusion rate is proportional to the number of ligands a sender cell secretes into the environment. We can estimate this from the number of RNA molecules proportional to its gene expression. Using the ligands database, we multiplied each ligand gene's expression value from a given scRNA-seq dataset by a scale factor (SF). Here we provide a default empirical SF value of 1000. The value of the scaling factor SF does not influence the relative abundance of the ligands since the same factor scales all the genes. Instead, the SF is just employed to scale the real value gene expression (often the case after preprocessing and normalization).

Second, the diffusion rate is inversely proportional to the distance (of the receiver agent) from the sender cell agent. We calculate the Euclidean distance between two cells based on the spatial coordinates of all the cell agents. The degree of freedom for the distance is a hyperparameter τ, with a default value of two, as the cells are from a two-dimensional slice. When spatial information is unavailable (e.g., in single-cell RNA-seq data), τ is set to 0 so that the distance between all cells in the data will degenerate to 1. A user-settable slider δ (between *0-10*, default = *1*) is available to

control the decay rate of the diffusion along with the increase of distance, which allows for inferring long-range or short-range interactions (see below). The distance between two cell agents (i, j) is calculated as the Euclidean Distance $d(i, j)$. The diffusion rate γ_l can be calculated as below:

$$\gamma_l = \frac{x_l}{d(i,j)^{\tau \times \delta}} \times SF$$

(2) Receptor receiving rate (α_r)

The receiving rate α_r associated with a receptor r represents the probability that an interaction is received by the receptor, and it is proportional to the expression of the receptor. The receptor receiving rate will be a value in the range (0,1) where 0 represents the receptor cannot "receive" all ligands that reach it, while 1 denotes that reached ligands will all be "received" by the receptor. Here, we employed a Gaussian distribution to model the distribution of all gene expressions $x_r \sim N(\mu, \sigma)$. With the expression distribution, we then can convert the receptor expression x_r into range (0,1) via the cumulative density frequency. Specifically, the receiving rate α_r for the receptor r is defined as the following:

$$\alpha_r = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x_r - \mu}{\sigma\sqrt{2}}\right)\right]$$

(3) Receptor transcriptome conversion rate (β_r)

Existing cellular interaction inference methods only quantify the level of interactions between cell types. They do not describe how the CCIs affect the downstream gene expression dynamics. However, CCIs play a crucial role in regulating gene expression, impacting the transcriptome, and ultimately influencing cellular function and fate. In different application scenarios, LR pairs may have distinct impacts on the downstream gene expression and affect the cell phenotype. In order to account for the effect of cellular interaction on downstream gene expression, we employed a neural network regressor to estimate the downstream gene expression impact associated with each receptor (conversion rate β_r). The neural network takes as input a vector vr_i representing the expression of all receptors in cell i. This receptor expression vector is then fed into the neural network ($\mathcal{F}$) to predict the expression x_i of all genes in cell i (regression task). The neural network structure comprises the following components: one input layer, one hidden layer, and one output layer. The input layer is of a size equivalent to the number of receptors. The output layer is the same size as the number of all genes. The hidden layer is of size one thousand, representing transcription factors involved in downstream signaling pathways. We use a stochastic gradient descent optimizer to search for a network $\mathcal{F}$ that minimizes the mean squared error loss L (between the true observed expression X and predicted expression $\widehat{X} = \mathcal{F}(vr_i)$:

$$L = \frac{1}{|X|} \sum_{x_i \in X} \left(x_i - \mathcal{F}(vr_i)\right)^2$$

With the trained neural network regressor, we then perform a sensitivity analysis to score each input receptor based on its impact on the downstream gene expression. Specifically, we simulate

blocking a receptor r by setting its value to zero (perturbed input receptor input vector ${vr_i}^{r-}$). Then, we quantify the importance of the receptor r (I_r) based on the loss increase caused by the perturbation:

$$I_r = L^{r-} - L$$
$$= \frac{1}{|X|}\sum_{x_i \in X}\left(x_i - \mathcal{F}(vr_i^{r-})\right)^2 - \frac{1}{|X|}\sum_{x_i \in X}\left(x_i - \mathcal{F}(vr_i^{r})\right)^2$$

Employing a similar approach as the receiving rate, we normalize the receptor importance into range (0,1) via a Gaussian ($I_r \sim N(\mu_I, \sigma_I)$) cumulative density frequency strategy $\beta_r = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{I_r - \mu_I}{\sigma_I\sqrt{2}}\right)\right]$. Here, we set a minimum value for the conversion rate to be 0.4 based on previous studies that measure the effect of cellular interaction on gene expression[73].

(4) Cell-cell interaction score between cells

The interaction score between a ligand-receptor pair (l, r) in a sender-receiver pair (i, j) is calculated based on the rates calculated above ($\gamma_l, \alpha_r, \beta_r$):

$$LR(i_l, j_r) = \gamma_l \times \alpha_r \times \beta_r = \frac{x_l}{d(i,j)^{\tau \times \delta}} \times SF \times \alpha_r \times \beta_r$$

The receiving score for a given cell i is calculated as the summation over all ligand-receptor pairs (l,r) across all sending cells:

$$CRS(i) = \sum_{j \in C}\sum_{(l,r) \in (L,R)} LR(j_l, i_r)$$

The combined interaction score between a sender-receiver cell pair (i, j) is calculated as the summation over all ligand-receptor pairs (l,r):

$$CPS(i,j) = \sum_{(l,r) \in (L,R)} LR(i_l, j_r)$$

The overall interaction score $IS(C1,\ C2)$ between two clusters C1 (sender) and C2 (receiver) for a ligand-receptor pair (l,r) is quantified as the average interaction score of all possible sender-receptor cell agent pairs between the sender C1 and receiver C2:

$$IS(C1_l,\ C2_r) = \frac{1}{|C1| \times |C2|}\sum_{i \in C1}\sum_{j \in C2} LR(i_l, j_r)$$
$$= \frac{1}{|C1| \times |C2|}\sum_{i \in C1}\sum_{j \in C2} \frac{x_l}{d(i,j)^{\tau \times \delta}} \times SF \times \alpha_r \times \beta_r$$

(5) Integration of cell agent pseudotime in calculating cell-cell interaction score

The CCI score discussed above ignores the pseudotime time differences between all cell agents in the same cluster. However, the integration of the cell pseudotime improves the CCI inference[21]. The main difference in these score calculations lies in the fact that, instead of computing scores between every cell in two clusters, we now exclusively calculate scores between cells that reside within the same pseudotime bin (b):

$$CRS_{PT}(i) = \begin{cases} \sum_{j \in C} \sum_{(l,r) \in (L,R)} LR(i_l, j_r), & if\ b_i = b_j \\ 0 & , else \end{cases},$$

$$CPS_{PT}(i,j) = \begin{cases} \sum_{(l,r) \in (L,R)} LR(i_l, j_r), & if\ b_i = b_j \\ 0 & , else \end{cases},$$

We take the average score of all non-zero bins to produce a final pseudotime-integrated interaction score (IS_{PT}) for a given ligand-receptor (l,r) pair between two clusters:

$$IS_{PT}(C1_l,\ C2_r) = \begin{cases} \frac{1}{|C1| \times |C2|} \sum_{i \in C1} \sum_{j \in C2} LR(i_l, j_r), & if\ b_i = b_j \\ 0 & , else \end{cases},$$

Find significant ligand-receptor pairs between cell populations

Once we get the interaction score for an LR pair between two cell clusters, we will examine whether such an interaction score is statistically significant (stronger compared to the random background). Specifically, we employed a random permutation test to calculate the statistical significance P-value associated with each LR interaction score. Significant interactions are identified between two cell groups using a permutation test. It randomly permutes the group labels of cells and then recalculates the interaction score for each LR pair. The LR score is stored after each permutation regardless of the cell type pairing. We perform enough permutations so that each LR pair, has on average 10,000 score values. We fit the random score distribution with a gamma distribution for each LR pair. By fitting the observed interaction scores (for a specific LR pair) to a gamma distribution, we get the estimated parameters (alpha and scale) of the gamma distribution to represent the random interaction score distribution of the LR pair. Using the background distribution, we can compute a P-value for a given interaction score (for the specific LR pair) by evaluating its position in the distribution's right tail. The P-value is then corrected using FDR (Benjamini/Hochberg) correction. Interaction Scores (IS) with a P-value above the threshold (0.05) are considered Significant Interaction Scores (SIS). Using the SIS, we can calculate the cell type pair scores (CTPS) between two clusters, C1 (sender) and C2 (receiver):

$$CTPS(C1, C2) = \sum_{(l,r) \in (L,R)} SIS(C1_l, C2_r)$$

Due to the computational expense of calculating the distances between all cells in the dataset, the permutation test can be quite costly, especially when considering spatial information. To address this, we have devised a method to match the background gamma distribution of non-spatial calculations with that of spatial ones. This is achieved by dividing the scale parameter by the average distance between all cells. Additionally, we halve both the alpha and the modified scale values. Although this does not equate to perfect scaling, we present very similar results using the scaling method significantly more efficiently (Supplementary Fig. 8b, c).

Long and Short range CCI inference

Through the power of ABM and CellAgentChat's ability to consider spatial transcriptomics data when modeling individual cell interactions, we can capture long- and short-range interactions. We

use the δ parameter to affect the degree of distance and model interactions at different ranges. Short-range interactions can be modeled using a δ greater than 1 (1-10). Long-range interaction is modeled using a δ value less than 1 (0-1). Keeping δ at the default value of 1 provides no bias toward long-range or short-range interactions.

In silico receptor perturbation

After identifying the top interactions between cell types, we conducted in silico perturbation using our trained deep learning model $\mathcal{F}$. To illustrate the process of in silico receptor blocking, we centered our attention on the receptors involved in the top 25 LR interactions, spanning all cell type pairs. We then subjected these receptors to a sensitivity analysis for perturbation. We simulated the blocking of each receptor r by setting its gene expression value to zero. This manipulation of the input receptor vector, vr_i^{r-}, allowed us to gauge the ensuing influence on the gene expression downstream. After each receptor's perturbation, we singled out the top 50 most impacted genes (target genes). This approach enabled us to assess the effects of receptor blocking at the transcriptome level. To further demonstrate the potency of our model in executing in silico perturbation, we utilized the DisGeNET[53] database. The objective was to link the perturbed target genes with the specific diseases that pique our interest - PDAC and PC (i.e., the disease associated with the dataset). To quantify the strength of the correlation between each receptor and the disease, we applied a binomial test. We used this test to compute the statistical significance of the overlap between the perturbed target genes and the genes associated with each disease. We further refined our results by adjusting the P-values, utilizing the FDR-Bonferroni Correction method, ensuring the robustness of our findings.

Agent Based Animation of CCIs

CellAgentChat offers a unique feature by allowing the inference of cellular interactions between individual cells rather than just cell types. Its real-time animation platform displays the receiving strength of each cell, with size representing the strength; larger cells indicate stronger receiving scores (Supplementary Fig. 8a, Supplementary Movie 1). The platform also includes customizable hyperparameters, such as the maximum number of steps (default of 1), δ for controlling long-range and short-range interactions (default of 1), the degree of freedom τ (default of 2), and Gaussian noise applied to the ligand diffusion rate (default of 5%) at each step, enabling effective in silico permutations (Supplementary Movie 1).

In addition, CellAgentChat offers hyperparameters for controlling the LR universe (Supplementary Movie 1). Users can block a specific receptor to observe downstream gene expression changes, which can aid in identifying potential therapeutic targets. The animation platform also allows for visualizing specific cell types/clusters, with modifications only applied to those selected. Furthermore, users have the option to visualize specific receptor or LR pairs. This visualization tool enables researchers to understand cell communication patterns at a granular level.

Functional Evaluation of Ligand-Receptor pairs inferred by CellAgentChat

We conducted pathway enrichment analysis to validate the interactions inferred by CellAgentChat functionally. We utilized both ligands and receptors from the top 25 highest scoring LR pairs among all cell type pairs as the input gene set. To perform pathway enrichment analysis, we employed gProfiler[74] and selected all terms from the reactome pathway with a minimum adjusted P-value below 1.00×10^{-2}. This approach was consistent across all datasets, regardless of spatial or pseudotime combinations. In cases where we compared the functional reactome pathways between two groups (i.e., pseudotime vs non pseudotime), we used a Wilcoxon test (one-sided).

Benchmarking

We compared the performance of CellAgentChat with two other state-of-the-art methods, CellChat and CellPhoneDB. These methods also enable the inference of CCIs using spatial and non-spatial data. However, unlike CellAgentChat, they do not employ an ABM approach. We used the LR universe derived from CellTalkDB for CellAgentChat, and the existing databases for CellChat and CellPhoneDB, respectively.

To compare the accuracies of the methods, we determined the number of shared LR interactions from the top 100 pairs identified by each method (Supplementary Figs. 5a-b, 6a-b). We then used the cumulative density frequency of the hypergeometric distribution to calculate a P-value. The population size was the number of LR pairs in CellTalkDB, and the number of successes in the population was the number of LR pairs unique to the method compared to CellAgentChat (i.e., CellPhoneDB or CellChat). The sample size was the number of LR pairs unique to CellAgentChat, and the number of successes in the sample was the overlapping LR interactions between the two methods.

Additionally, we measured the Spearman's rank correlation (two-sided) to evaluate the communication pattern across cell types of each of the three methods. As described above, we also conducted pathway enrichment analysis using the genes from the top 100 LR pairs to evaluate the functional validation of the LR interactions inferred by all three methods. We took the pathways found using pathway enrichment analysis of the inferred interactions found by CellAgentChat and compared its -log10 adjusted P-value with CellChat and CellPhoneDB for the same pathways. Finally, we used the Wilcoxon test (one-sided) to differentiate the functional validity of the LR interactions across the three methods.

Data availability

CellTalkDB is available at (https://github.com/ZJUFanLab/CellTalkDB). Two of the three datasets analyzed in this study are available from the Gene Expression Omnibus (GEO) repository under the following accession numbers: GSE144240 (GSM4565826) and GSE111672 (GSM3036911). The third dataset is from StomicDB (https://db.cngb.org/stomics/), a public resource that stores scRNA-seq and spatial transcriptomic datasets[75], under the name: *Human Prostate Cancer, Acinar Cell Carcinoma (FFPE)*.

Code availability

The CellAgentChat software package and source code are available at our GitHub repository (https://github.com/mcgilldinglab/CellAgentChat). All preprocessed .h5ad files used in this study are also available in the same GitHub repository.

Acknowledgements

This work is supported by grants from the Canadian Institutes of Health Research (CIHR) [PJT-180505 to J.D]; the Fonds de recherche du Québec - Santé (FRQS) [295298 to J.D., 295299 to J.D.]; the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN2022-04399 to J.D.]; and the Meakins-Christie Chair in Respiratory Research [to J.D.]. This publication is part of the Human Cell Atlas – www.humancellatlas.org/publications/.

FIGURES

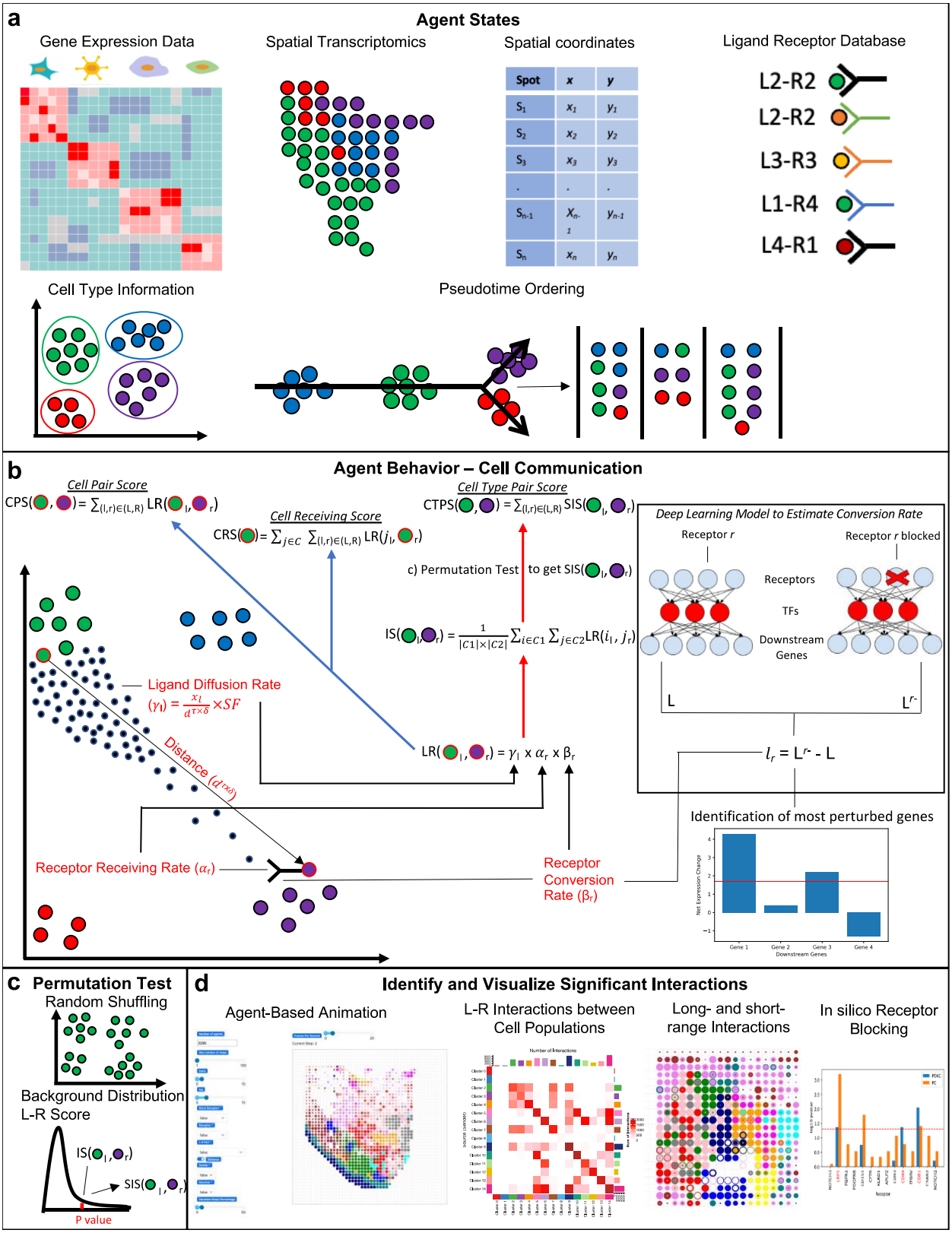
a
Agent States
Gene Expression Data
Spatial Transcriptomics
Spatial coordinates
Spot x y
S1 x1 y1
S2 x2 y2
S3 x3 y3
Sn-1 Xn-1 yn-1
Sn xn yn
Ligand Receptor Database
L2-R2
L2-R2
L3-R3
L1-R4
L4-R1
Cell Type Information
Pseudotime Ordering
b
Agent Behavior – Cell Communication
Cell Pair Score
CPS = ∑(l,r)∈(L,R) LR(l, r)
Cell Receiving Score
CRS = ∑j∈C ∑(l,r)∈(L,R) LR(j, r)
Cell Type Pair Score
CTPS = ∑(l,r)∈(L,R) SIS(l, r)
c) Permutation Test to get SIS(l, r)
IS(l, r) = 1/(|C1|×|C2|) ∑i∈C1 ∑j∈C2 LR(il, jr)
LR(l, r) = γl x αr x βr
Ligand Diffusion Rate
(γl) = xl/dτ×δ ×SF
Distance (dτ×δ)
Receptor Receiving Rate (αr)
Receptor Conversion Rate (βr)
Deep Learning Model to Estimate Conversion Rate
Receptor r
Receptor r blocked
Receptors
TFs
Downstream Genes
L
Lr-
lr = Lr- - L
Identification of most perturbed genes
Net Expression Change
Gene 1
Gene 2
Gene 3
Gene 4
Downstream Genes
c Permutation Test
Random Shuffling
Background Distribution
L-R Score
IS(l, r)
SIS(l, r)
P value
d
Identify and Visualize Significant Interactions
Agent-Based Animation
L-R Interactions between Cell Populations
Number of Interactions
Long- and short-range Interactions
In silico Receptor Blocking
PSAC
PC
Receptor

Figure 1: Overview of CellAgentChat. a CellAgentChat Requires the following agent states. The states of each agent (cell) are its gene expression derived from scRNA-seq data, its cell type/cluster information, and the ligand-receptor database (by default uses CellTalkDB). Optionally, users can provide spatial coordinates of each agent (from spatial transcriptomics data) and the pseudotime ordering of each cell. **b** Cell communication behaviour of each agent. Calculating the ligand-receptor score between two cells for a specific ligand-receptor pair requires the ligand diffusion rate γ, the receptor receiving rate α_r, and the receptor conversion rate β_r. If spatial data is provided, the ligand diffusion rate γ incorporates the Euclidean distance d between. The τ hyperparameter controls the degree of freedom for the distance (default τ=2). The δ hyperparameter (default δ=1) controls whether to model long-range ($0 > \delta < 1$) or short-range ($1 > \delta < 10$). A deep learning model (right box) calculates the receptor conversion rate β_r. The inputs of our deep learning model are the receptor's gene expression, and the outputs are the gene expression of all genes. The hidden layers represent transcription factors. The neural network is fully connected and estimates the weights during training. We calculate l_r, the absolute value of the mean squared error loss L with L^{r-} , derived from blocking receptor r in the deep learning model. l_r can then be used to estimate β_r. Using the LR score, we can calculate the interaction score between two clusters for a cell type, the cell receiving score for a cell, and the cell pair score for two cells. After a permutation test, significant interactions are identified and used to calculate the cell type pair score between two clusters. **c** Permutation test. We shuffle the cell type labels of all cells and generate a gamma distribution of the scores for each ligand-receptor pair. The P-value is the right tail of the gamma distribution. **d** Our animation platform can identify and visualize significant interactions between individual cells. We also provide visualization plots to analyze interactions between cell types and identify ligand-receptor interactions. Further, we support the ability to identify long-range and short-range interactions. The deep learning model can also be used for in silico-permutations to simulate blocking receptors to identify the most perturbed downstream genes to discover therapeutic drugs.

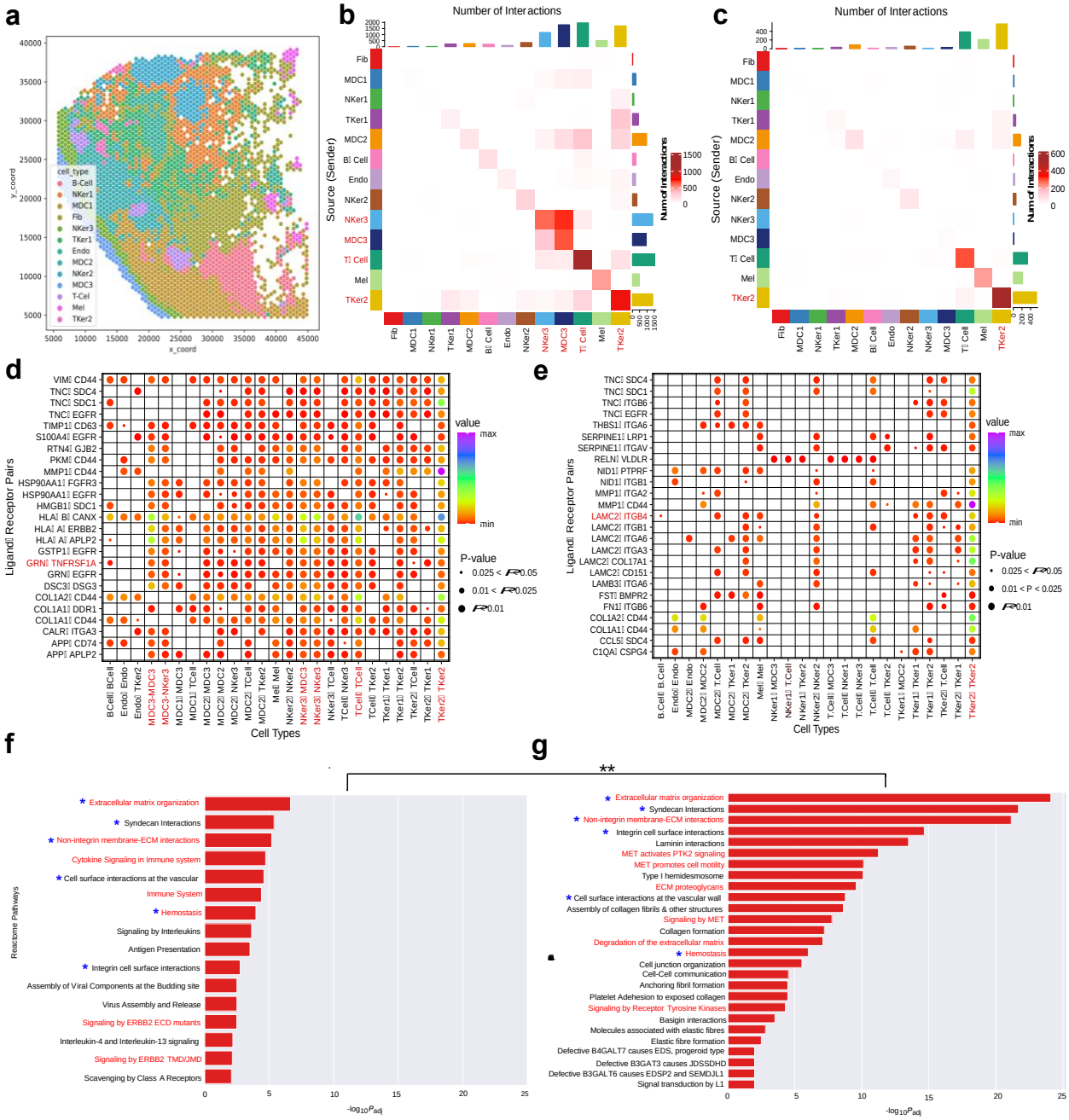

Figure 2: Analysis of Cell-Cell Communications in Squamous Cell Carcinoma (SCC) using CellAgentChat with Spatial Data. a Scatter plot of the coordinates of each cell. **(b, c)** Heatmaps display the communication network between each cell type inferred by CellAgentChat, without and with considering the pseudotime ordering of cells, respectively. **(d, e)** Dot plots illustrate the top 25 ligand-receptor pairs involved in the communication between cell types as inferred by CellAgentChat, without and with considering the pseudotime ordering of cells, respectively. **(f, g)** Reactome pathways analysis using the ligand-receptor interactions inferred by CellAgentChat, without and with considering the pseudotime ordering of cells, respectively. Pseudotime improves the inference of cellular communication with spatial data (Wilcoxon test one-sided; $P = 5.62\text{x}10^{-4}$). SCC is associated with the terms highlighted in red. Blue stars indicate terms shared between plots.

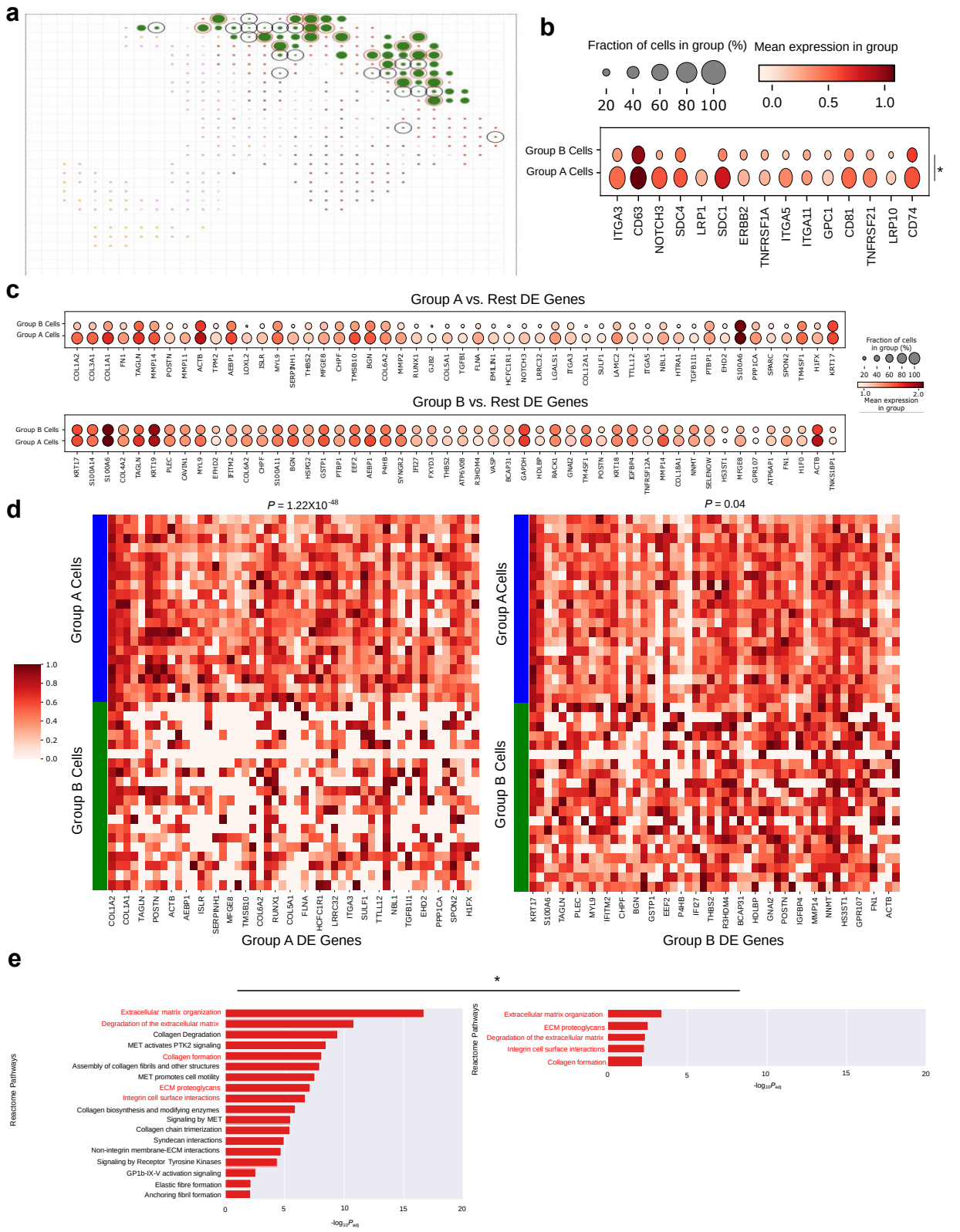

Figure 3: Application of CellAgentChat to detect variations in intercellular communication among individual cancer cells. a CellAgentChat's animation platform enables the detection of communication strength variations in individual cancer cells. The platform highlights the receiving

capacity of cancer subpopulation one cells (colored green). It identifies the top 20 receiving cells (group A) marked by red circles and the bottom 20 receiving cells (group B) marked by black circles. **b** Differential expression analysis of receptors involved in the top 25 ligand-receptor interactions reveals that Group A, consisting of the top 20 receiving cancer cells identified by CellAgentChat's animation platform, shows significantly higher expression of key receptors compared to Group B, consisting of the bottom 20 receiving cancer cells (Wilcoxon test one-sided; $P < 0.05$). Moreover, we detect the expression of these receptors in more cells within group A than group B. **c** Mean expression of differentially expressed (DE) genes for group A cells (top) and group B cells (bottom). **d** Heatmaps display the differentially expressed (DE) genes for group A (left) and group B (right). The DE genes of group A cells are mainly expressed by group A cells and not by group B cells (Wilcoxon test one-sided; $P < 0.001$). Both groups express the DE genes of group B cells but are more significantly expressed by group A cells (Wilcoxon test one-sided; $P < 0.05$). **e** The results of the reactome pathway analysis using group A DE genes (on the left) and group B DE genes (on the right). Pathways affected by PDAC are more associated with group A DE genes (Wilcoxon test one-sided; $P = 0.0313$). Red highlights indicate shared pathways.

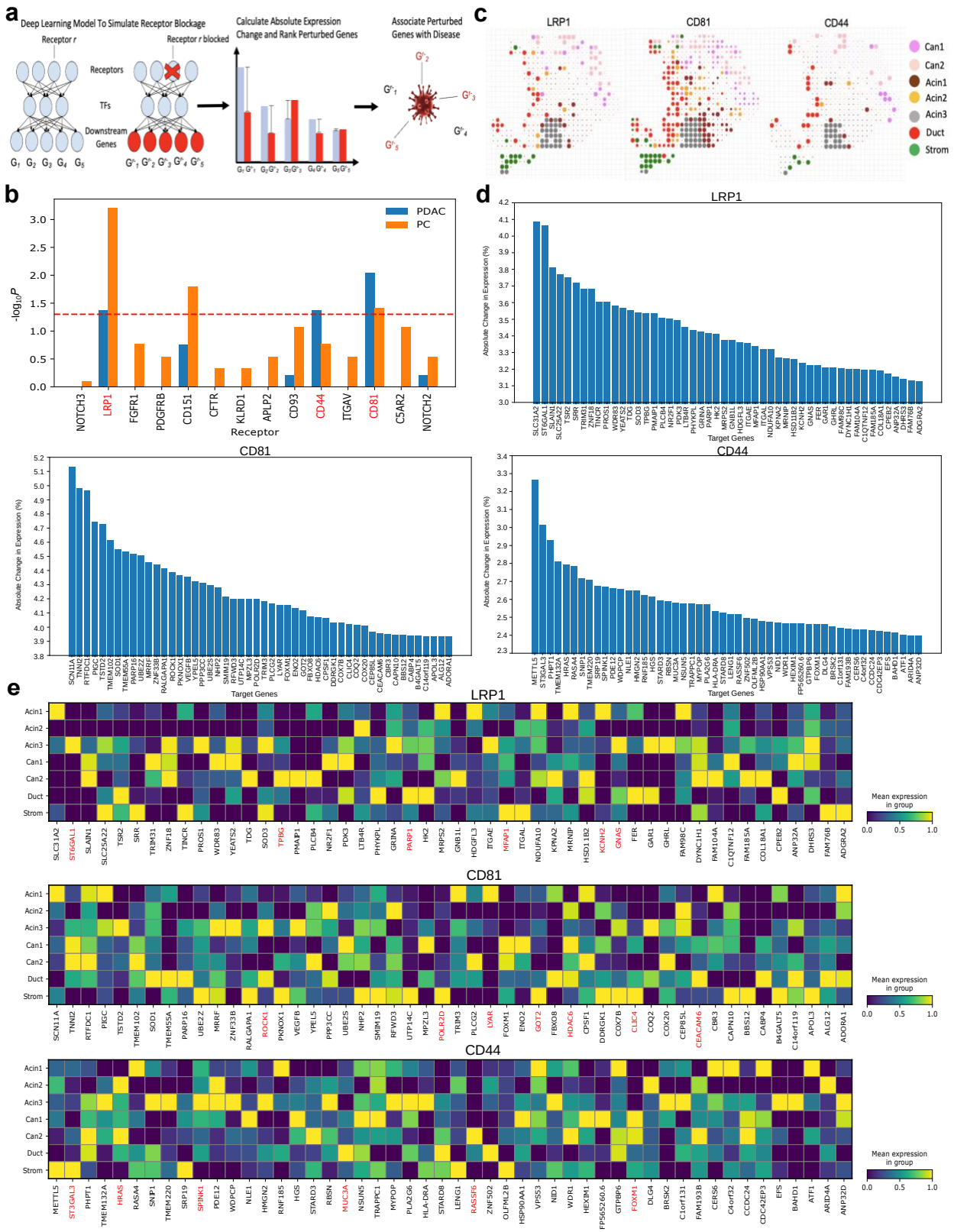

Figure 4: Identification of key downstream genes related to PDAC after in silico blockage of top receptors. a Schema of the in silico perturbation analysis using our neural network to block receptors and identify the most perturbed downstream genes linked with PDAC. **b** Significance of

receptor association with pancreatic ductal adenocarcinoma (PDAC) and its similarity to a related disease (PC), as represented by $-\log_{10}$ adjusted P-values (FDR corrected-Bonferroni Correction). The analysis includes receptors from the top 25 ligand-receptor interactions inferred by CellAgentChat without spatial data. LRP1, CD44, and CD81, highlighted in red, exhibit a significant association with PDAC. **c** The receiving strength of each cell for all interactions involving the receptors LRP1, CD81, and CD44, determined using CellAgentChat's animation platform. The size of the cell in the animation indicates the receiving strength. **d** Identification of the top 50 downstream genes (target genes) with the highest percent change in absolute expression upon blocking LRP1, CD81, and CD44, respectively. **e** The matrix plot displays the mean expression of each target gene for each cell type. A red highlight indicates genes associated with PDAC.

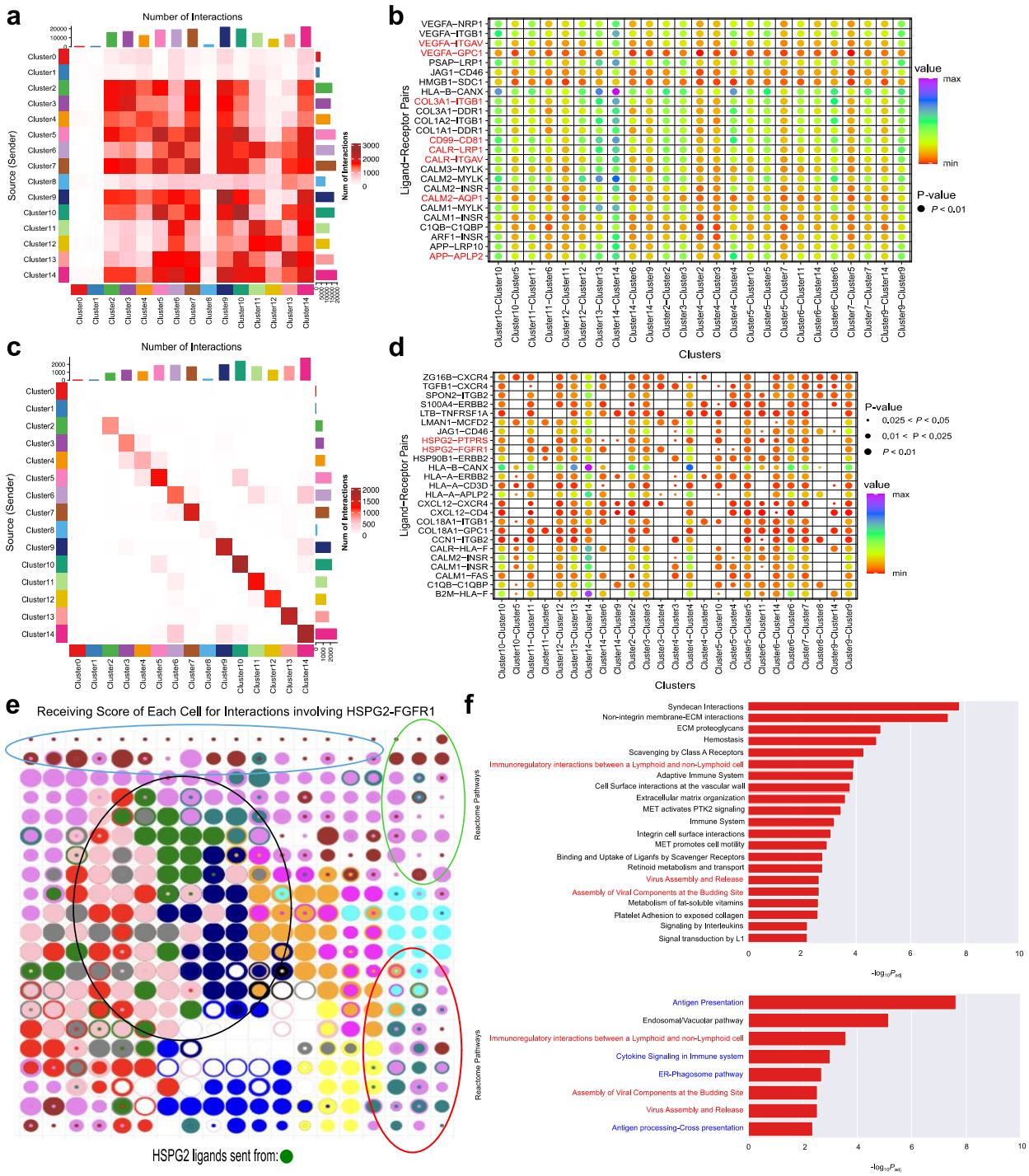

Figure 5: Application of CellAgentChat to infer Long-range and Short-range interactions in Human Prostate Cancer Cells. a Heatmap displays the communication network between each cell type inferred by CellAgentChat's long-range mode (δ = 0.5). **b** Dot plot illustrates the top 25 ligand-receptor pairs involved in the communication between cell types inferred by CellAgentChat's long-range mode. **c** Heatmap displays the communication network between each cell type inferred by CellAgentChat's short-range mode (δ = 5). **d** Dot plot illustrates the top 25 ligand-receptor pairs involved in the communication between cell types inferred by CellAgentChat's short-range mode. **e** CellAgentChat's animation platform depicts the receiving

strength of cells for the HSPG2-FGFR1 ligand-receptor pair using the short-range mode. This animation sends HSPG2 (ligand) from cells in cluster 4 (represented by green cells). The size of the cell in the animation indicates the receiving strength. Cells closer to cluster 4 cells display higher receiving strengths (indicated by a black circle) than cells that are further away (indicated by red, blue, and green circles). The results demonstrate that HSPG2-FGFR1 occurs primarily at short ranges. **f** Reactome pathway analysis of the receptors involved in the top 25 ligand-receptor interactions in the long-range mode (top) and short-range mode (bottom). Red highlights indicate shared terms between long- and short-range analysis. Highlighted in blue are pathways associated with human prostate cancer that occur at short ranges.

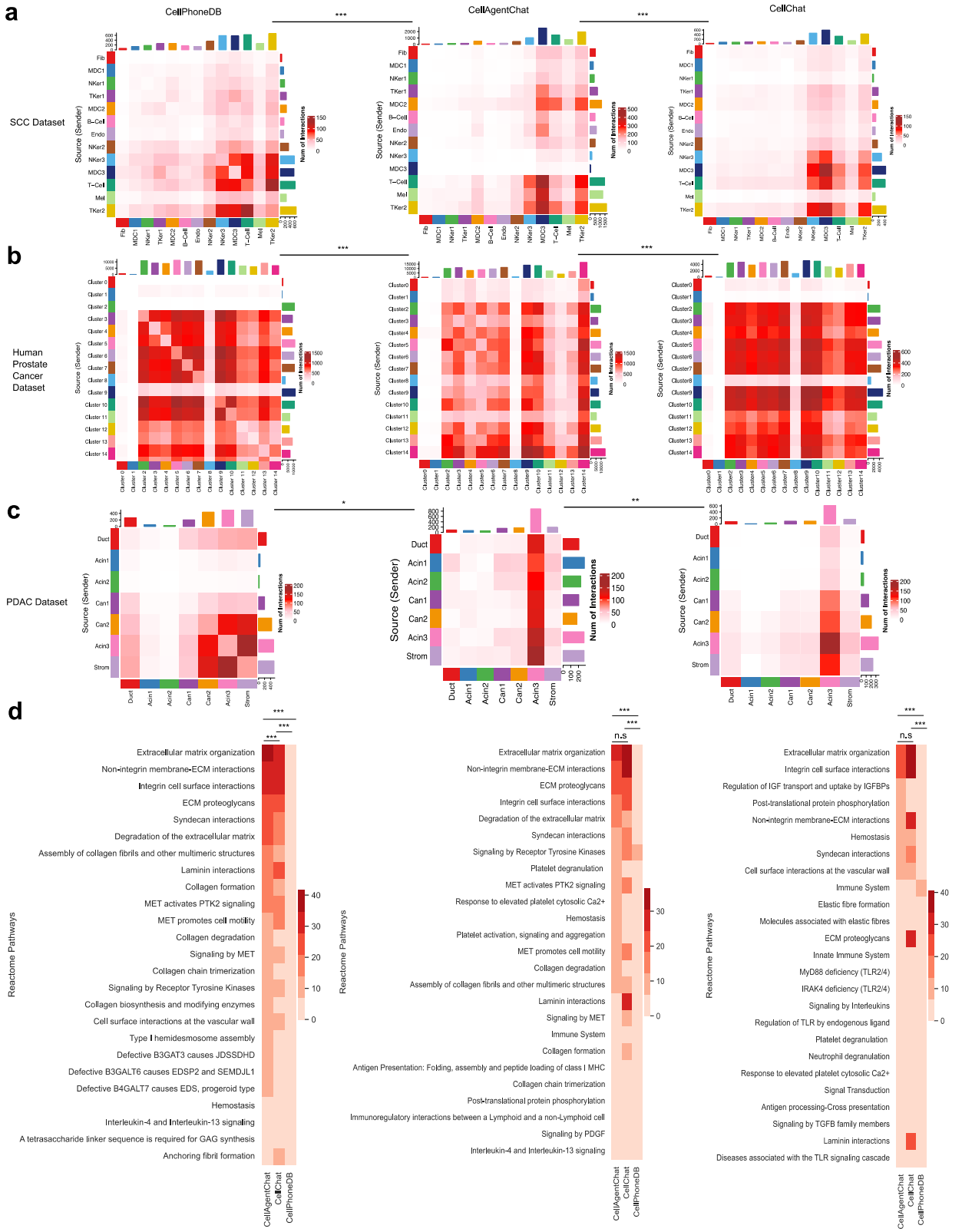

Figure 6: Comparison of CellAgentChat with CellPhoneDB and CellChat while considering pseudotime but without spatial data. a Heatmaps display the communication network between each cell type inferred by CellAgentChat, CellPhoneDB, and CellChat for the SCC dataset.

CellAgentChat shows similar communication networks to CellPhoneDB and CellChat (Spearman rank correlation two-sided; $P = 1.04\text{x}10^{-11}$, $P = 1.34\text{x}10^{-5}$). **b** Heatmaps display the communication network between each cell type inferred by CellAgentChat, CellPhoneDB, and CellChat for the human prostate cancer dataset. CellAgentChat shows similar communication networks to CellPhoneDB and CellChat (Spearman rank correlation two-sided; $P = 8.61\text{x}10^{-38}$, $P = 5.55\text{x}10^{-41}$). **c** Heatmaps display the communication network between each cell type inferred by CellAgentChat, CellPhoneDB, and CellChat for the PDAC dataset. CellAgentChat shows similar communication networks to CellPhoneDB and CellChat (Spearman rank correlation two-sided; $P = 0.0114$, $P = 0.00950$). **d** Reactome pathway analysis of receptors from the top 25 ligand-receptor pairs inferred by each method. CellAgentChat performs significantly better in terms of functional relevance of inferred interactions than CellPhoneDB and CellChat for the SCC dataset (left) (Wilcoxon test one-sided; $P = 4.81\text{x}10^{-4}$, $P = 2.99\text{x}10^{-8}$). CellAgentChat performs comparably in terms of the functional relevance of inferred interactions with CellChat but significantly better than CellPhoneDB for the human prostate cancer dataset (middle) (Wilcoxon test one-sided; $P = 0.489$ (n.s), $P = 2.98\text{x}10^{-8}$) and for the PDAC dataset (right) (Wilcoxon test one-sided; $P = 0.386$ (n.s), $P = 1.28\text{x}10^{-6}$).

REFERENCES

1 Hu, M. & Polyak, K. Microenvironmental regulation of cancer development. *Current opinion in genetics & development* **18**, 27-34 (2008).

2 Shao, X., Lu, X., Liao, J., Chen, H. & Fan, X. New avenues for systematically inferring cell-cell communication: through single-cell transcriptomics data. *Protein & cell* **11**, 866-880 (2020).

3 Armingol, E., Officer, A., Harismendy, O. & Lewis, N. E. Deciphering cell–cell interactions and communication from gene expression. *Nature Reviews Genetics* **22**, 71-88 (2021).

4 Zhou, X. *et al.* Circuit design features of a stable two-cell system. *Cell* **172**, 744-757. e717 (2018).

5 Rouault, H. & Hakim, V. Different cell fates from cell-cell interactions: core architectures of two-cell bistable networks. *Biophysical Journal* **102**, 417-426 (2012).

6 Bonnans, C., Chou, J. & Werb, Z. Remodelling the extracellular matrix in development and disease. *Nature reviews Molecular cell biology* **15**, 786-801 (2014).

7 Vento-Tormo, R. *et al.* Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature* **563**, 347-353 (2018).

8 Stubbington, M. J., Rozenblatt-Rosen, O., Regev, A. & Teichmann, S. A. Single-cell transcriptomics to explore the immune system in health and disease. *Science* **358**, 58-63 (2017).

9 Giladi, A. *et al.* Dissecting cellular crosstalk by sequencing physically interacting cells. *Nature Biotechnology* **38**, 629-637 (2020).

10 Yuan, G.-C. *et al.* Challenges and emerging directions in single-cell analysis. *Genome biology* **18**, 1-8 (2017).

11 Marx, V. Method of the Year: spatially resolved transcriptomics. *Nature methods* **18**, 9-14 (2021).

12 Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463-1467 (2019).

13 Chen, A. *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777-1792. e1721 (2022).

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

15 Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nature communications* **12**, 1088 (2021).

16 Cabello-Aguilar, S. *et al.* SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Research* **48**, e55-e55 (2020).

17 Asp, M., Bergenstråhle, J. & Lundeberg, J. Spatially resolved transcriptomes—next generation tools for tissue exploration. *BioEssays* **42**, 1900221 (2020).

18 Dries, R. *et al.* Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome biology* **22**, 1-31 (2021).

19 Williams, C. G., Lee, H. J., Asatsuma, T., Vento-Tormo, R. & Haque, A. An introduction to spatial transcriptomics for biomedical research. *Genome Medicine* **14**, 1-18 (2022).

20 Brady-Kalnay, S. M. Vol. 6 344-345 (Taylor & Francis, 2012).

21 Li, D. *et al.* TraSig: inferring cell-cell interactions from pseudotime ordering of scRNA-Seq data. *Genome biology* **23**, 73 (2022).
22 Xu, C. & Jackson, S. A. Vol. 20 1-4 (Springer, 2019).
23 Yu, J. S. & Bagheri, N. Agent-based models predict emergent behavior of heterogeneous cell populations in dynamic microenvironments. *Frontiers in Bioengineering and Biotechnology* **8**, 249 (2020).
24 Bonabeau, E. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the national academy of sciences* **99**, 7280-7287 (2002).
25 Ji, A. L. *et al.* Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell* **182**, 497-514. e422 (2020).
26 Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome biology* **19**, 1-5 (2018).
27 Dong, D. *et al.* A GRN Autocrine-Dependent FAM135B/AKT/mTOR Feedforward Loop Promotes Esophageal Squamous Cell Carcinoma ProgressionFunction and Mechanism of FAM135B in ESCC. *Cancer Research* **81**, 910-922 (2021).
28 Han, Y., Ding, Z., Chen, B., Liu, Y. & Liu, Y. A Novel Inflammatory Response–Related Gene Signature Improves High-Risk Survival Prediction in Patients With Head and Neck Squamous Cell Carcinoma. *Frontiers in Genetics* **13** (2022).
29 Liang, Y. *et al.* LncRNA CASC9 promotes esophageal squamous cell carcinoma metastasis through upregulating LAMC2 expression by interacting with the CREB-binding protein. *Cell Death & Differentiation* **25**, 1980-1995 (2018).
30 Moon, Y. *et al.* LAMC2 enhances the metastatic potential of lung adenocarcinoma. *Cell Death & Differentiation* **22**, 1341-1352 (2015).
31 Nagata, M. *et al.* ITGA3 and ITGB4 expression biomarkers estimate the risks of locoregional and hematogenous dissemination of oral squamous cell carcinoma. *BMC cancer* **13**, 1-9 (2013).
32 Schaafsma, E., Fugle, C. M., Wang, X. & Cheng, C. Pan-cancer association of HLA gene expression with cancer prognosis and immunotherapy efficacy. *British Journal of Cancer* **125**, 422-432 (2021).
33 Atasoy, M., Anadolu-Braise, R., Pirim, I., Dogan, H. & Ikbal, M. HLA antigen profile differences in patients with SCC (squamous cell carcinoma) in-situ/actinic keratosis and invasive SCC: Is there a genetic succeptibility for invasive SCC development? *The Eurasian Journal of Medicine* **41**, 162 (2009).
34 Pontén, F., Jirström, K. & Uhlen, M. The Human Protein Atlas—a tool for pathology. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland* **216**, 387-393 (2008).
35 Zhao, S. & Yu, M. Identification of MMP1 as a potential prognostic biomarker and correlating with immune infiltrates in cervical squamous cell carcinoma. *DNA and cell biology* **39**, 255-272 (2020).
36 Zhang, G. *et al.* Identity of MMP1 and its effects on tumor progression in head and neck squamous cell carcinoma. *Cancer medicine* **11**, 2516-2530 (2022).
37 Ludwig, N. *et al.* CD44 (+) tumor cells promote early angiogenesis in head and neck squamous cell carcinoma. *Cancer Letters* **467**, 85-95 (2019).
38 Yuan, S.-S. F. *et al.* CD44 Mediates Oral Squamous Cell Carcinoma-Promoting Activity of MRE11 via AKT Signaling. *Journal of Personalized Medicine* **12**, 841 (2022).

39 Parker, A. L. *et al.* Extracellular matrix profiles determine risk and prognosis of the squamous cell carcinoma subtype of non-small cell lung carcinoma. *Genome Medicine* **14**, 1-29 (2022).
40 Fromme, J. E. & Zigrino, P. The role of extracellular matrix remodeling in skin tumor progression and therapeutic resistance. *Frontiers in Molecular Biosciences*, 370 (2022).
41 Lottin, M. *et al.* Molecular landscape of the coagulome of oral squamous cell carcinoma. *Cancers* **14**, 460 (2022).
42 Ahrends, T. & Borst, J. The opposing roles of CD 4+ T cells in anti-tumour immunity. *Immunology* **154**, 582-592 (2018).
43 Sharma, V. The interplay of cytokine signaling and non-coding RNAs in head and neck squamous cell carcinoma pathobiology. *Molecular Biology Reports* **49**, 10825-10847 (2022).
44 Centuori, S. M. & Bauman, J. E. c-Met signaling as a therapeutic target in head and neck cancer. *The Cancer Journal* **28**, 346-353 (2022).
45 Szturz, P. *et al.* Understanding c-MET signalling in squamous cell carcinoma of the head & neck. *Critical reviews in oncology/hematology* **111**, 39-51 (2017).
46 Birkeland, A. C. *et al.* Identification of targetable ERBB2 aberrations in head and neck squamous cell carcinoma. *JAMA Otolaryngology–Head & Neck Surgery* **142**, 559-567 (2016).
47 Dahlhoff, M., Muzumdar, S., Schäfer, M. & Schneider, M. R. ERBB2 is essential for the growth of chemically induced skin tumors in mice. *Journal of Investigative Dermatology* **137**, 921-930 (2017).
48 Liu, Z., Sun, D. & Wang, C. Evaluation of cell-cell interaction methods by integrating single-cell RNA sequencing data with spatial information. *Genome Biology* **23**, 1-38 (2022).
49 Moncada, R. *et al.* Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nature biotechnology* **38**, 333-342 (2020).
50 Friedl, P. & Alexander, S. Cancer invasion and the microenvironment: plasticity and reciprocity. *Cell* **147**, 992-1009 (2011).
51 Chen, P., Cescon, M. & Bonaldo, P. Collagen VI in cancer and its biological mechanisms. *Trends in molecular medicine* **19**, 410-417 (2013).
52 Perez, V. M., Kearney, J. F. & Yeh, J. J. The PDAC extracellular matrix: a review of the ECM protein composition, tumor cell interaction, and therapeutic strategies. *Frontiers in Oncology* **11**, 751311 (2021).
53 Piñero, J. *et al.* DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database* **2015** (2015).
54 Xu, Y., Liu, J., Nipper, M. & Wang, P. Ductal vs. acinar? Recent insights into identifying cell lineage of pancreatic ductal adenocarcinoma. *Annals of pancreatic cancer* **2** (2019).
55 Mun, J.-Y., Leem, S.-H., Lee, J. H. & Kim, H. S. Dual relationship between stromal cells and immune cells in the tumor microenvironment. *Frontiers in Immunology* **13**, 864739 (2022).
56 Gheysarzadeh, A., Ansari, A., Emami, M. H., Razavi, A. E. & Mofid, M. R. Over-expression of low-density lipoprotein receptor-related Protein-1 is associated with poor prognosis and invasion in pancreatic ductal adenocarcinoma. *Pancreatology* **19**, 429-435 (2019).

57 Ruivo, C. F. *et al.* Extracellular vesicles from pancreatic cancer stem cells lead an intratumor communication network (EVNet) to fuel tumour progression. *Gut* **71**, 2043-2068 (2022).
58 Zhao, S. *et al.* CD44 Expression Level and Isoform Contributes to Pancreatic Cancer Cell Plasticity, Invasiveness, and Response to TherapyCD44 and Gemcitabine Resistance. *Clinical Cancer Research* **22**, 5592-5604 (2016).
59 Chakraborty, A. *et al.* ST6Gal-I sialyltransferase promotes chemoresistance in pancreatic ductal adenocarcinoma by abrogating gemcitabine-mediated DNA damage. *Journal of Biological Chemistry* **293**, 984-994 (2018).
60 Brody, J. R. *et al.* Adenosquamous carcinoma of the pancreas harbors KRAS2, DPC4 and TP53 molecular alterations similar to pancreatic ductal adenocarcinoma. *Modern Pathology* **22**, 651-659 (2009).
61 West, J. & Newton, P. K. Cellular interactions constrain tumor growth. *Proceedings of the National Academy of Sciences* **116**, 1918-1923 (2019).
62 Corlan, A. S., Cîmpean, A. M., Jitariu, A.-A., Melnic, E. & Raica, M. Endocrine gland-derived vascular endothelial growth factor/prokineticin-1 in cancer development and tumor angiogenesis. *International Journal of Endocrinology* **2017** (2017).
63 Kielbik, M., Szulc-Kielbik, I. & Klink, M. Calreticulin—multifunctional chaperone in immunogenic cell death: potential significance as a prognostic biomarker in ovarian cancer patients. *Cells* **10**, 130 (2021).
64 Li, Y. *et al.* Bioinformatics analysis for the role of CALR in human cancers. *Plos one* **16**, e0261254 (2021).
65 Murdoch, A. D., Dodge, G., Cohen, I., Tuan, R. & Iozzo, R. Primary structure of the human heparan sulfate proteoglycan from basement membrane (HSPG2/perlecan). A chimeric molecule with multiple domains homologous to the low density lipoprotein receptor, laminin, neural cell adhesion molecules, and epidermal growth factor. *Journal of Biological Chemistry* **267**, 8544-8557 (1992).
66 Leonard, W. J. & Lin, J.-X. Cytokine receptor signaling pathways. *Journal of Allergy and Clinical Immunology* **105**, 877-888 (2000).
67 Nunes-Hasler, P. & Demaurex, N. The ER phagosome connection in the era of membrane contact sites. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1864**, 1513-1524 (2017).
68 Vyas, J. M., Van der Veen, A. G. & Ploegh, H. L. The known unknowns of antigen processing and presentation. *Nature Reviews Immunology* **8**, 607-618 (2008).
69 Shao, X. *et al.* CellTalkDB: a manually curated database of ligand–receptor interactions in humans and mice. *Briefings in bioinformatics* **22**, bbaa269 (2021).
70 Islam, S. *et al.* Quantitative single-cell RNA-seq with unique molecular identifiers. *Nature methods* **11**, 163-166 (2014).
71 Ilicic, T. *et al.* Classification of low quality cells from single-cell RNA-seq data. *Genome biology* **17**, 1-15 (2016).
72 Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC genomics* **19**, 1-16 (2018).
73 Arnol, D., Schapiro, D., Bodenmiller, B., Saez-Rodriguez, J. & Stegle, O. Modeling cell-cell interactions from spatial molecular data with spatial variance component analysis. *Cell reports* **29**, 202-211. e206 (2019).

74 Raudvere, U. *et al.* g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic acids research* **47**, W191-W198 (2019).
75 Xu, Z. *et al.* STOmicsDB: a database of Spatial Transcriptomic data. *bioRxiv*, 2022.2003.2011.481421 (2022).