



## Harnessing agent-based frameworks in CellAgentChat to unravel cell–cell interactions from single-cell and spatial transcriptomics

Vishvak Raghavan, Yumin Zheng, Yue Li, et al.

*Genome Res.* 2025 35: 1646-1663 originally published online May 2, 2025

Access the most recent version at doi:[10.1101/gr.279771.124](https://doi.org/10.1101/gr.279771.124)

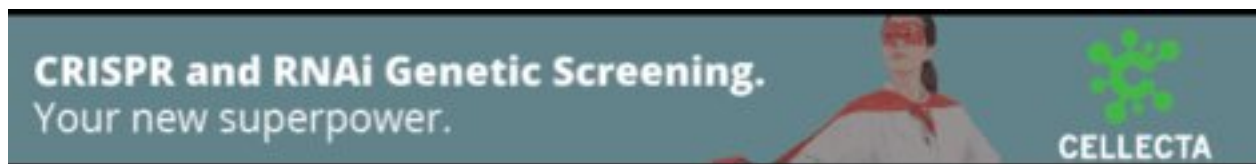
---

**References** This article cites 85 articles, 4 of which can be accessed free at:  
<http://genome.cshlp.org/content/35/7/1646.full.html#ref-list-1>

**Open Access** Freely available online through the *Genome Research* Open Access option.

**Creative Commons License** This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

## Method

# Harnessing agent-based frameworks in CellAgentChat to unravel cell–cell interactions from single-cell and spatial transcriptomics

Vishvak Raghavan,<sup>1,2,3</sup> Yumin Zheng,<sup>2</sup> Yue Li,<sup>1,3</sup> and Jun Ding<sup>1,2,3</sup>

<sup>1</sup>*School of Computer Science, McGill University, Montreal, Quebec H3A 2A7, Canada;* <sup>2</sup>*Meakins-Christie Laboratories, Translational Research in Respiratory Diseases Program, Research Institute of the McGill University Health Centre, Montreal, Quebec H4A 3J1, Canada;* <sup>3</sup>*Mila - Quebec Artificial Intelligence Institute, Montreal, Quebec H2S 3H1, Canada*

Understanding cell–cell interactions (CCIs) is essential yet challenging owing to the inherent intricacy and diversity of cellular dynamics. Existing approaches often analyze global patterns of CCIs using statistical frameworks, missing the nuances of individual cell behavior owing to their focus on aggregate data. This makes them insensitive in complex environments where the detailed dynamics of cell interactions matter. We introduce CellAgentChat, an agent-based model (ABM) designed to decipher CCIs from single-cell RNA sequencing and spatial transcriptomics data. This approach models biological systems as collections of autonomous agents governed by biologically inspired principles and rules. Validated across eight diverse single-cell data sets, CellAgentChat demonstrates its effectiveness in detecting intricate signaling events across different cell populations. Moreover, CellAgentChat offers the ability to generate animated visualizations of single-cell interactions and provides flexibility in modifying agent behavior rules, facilitating thorough exploration of both close and distant cellular communications. Furthermore, CellAgentChat leverages ABM features to enable intuitive *in silico* perturbations via agent rule modifications, facilitating the development of novel intervention strategies. This ABM method unlocks an in-depth understanding of cellular signaling interactions across various biological contexts, thereby enhancing *in silico* studies for cellular communication–based therapies.

[Supplemental material is available for this article.]

Cells, as the fundamental units of most life forms, are dynamic and continuously adapt their states in response to changes in their microenvironment (Hu and Polyak 2008; Shao et al. 2020). In multicellular organisms, cellular activities rely on the coordination of cell–cell interactions (CCIs) across different cell types and tissues (Rouault and Hakim 2012; Bonnans et al. 2014; Zhou et al. 2018; Armingol et al. 2021). CCIs are critical in numerous biological processes, including cell differentiation (Vento-Tormo et al. 2018), disease pathogenesis (Stubington et al. 2017), and immune responses (Giladi et al. 2020). By studying these interactions, we can uncover mechanisms behind biological processes, informing research on disorders and aiding in the discovery of diagnostic and therapeutic strategies (Giladi et al. 2020).

Inferring CCIs in biological processes is challenging owing to cellular heterogeneity in most biological systems (Yuan et al. 2017). Despite advancements in single-cell sequencing technologies like single-cell RNA sequencing (scRNA-seq) and spatial transcriptomics measurements (Eng et al. 2019; Rodrigues et al. 2019; Xia et al. 2019; Marx 2021; Chen et al. 2022; Janesick et al. 2023), existing methods for CCI inference face significant drawbacks. Tools, such as CellPhoneDB (Efremova et al. 2020), CellChat (Jin et al. 2021), Connectome (Raredon et al. 2022), SingleCellSignalR (Caballo-Aguilar et al. 2020), and NATMI (Hou et al. 2020), rely on mean expression values from single-cell clusters, focusing on cell population levels and losing single-cell resolution. Scriabin (Wilk et al. 2024) analyzes LR interactions at the

single-cell level but primarily relies on unimodal scRNA-seq data, limiting CCI inference ability owing to lack of spatial information. Because CCIs in tissue environments are heavily influenced by spatial structures and can occur at both long and short ranges (Asp et al. 2020), integrating spatial information is essential for studying endocrine and paracrine signaling. To address this, recent methods such as COMMOT (Cang et al. 2023), NICHES (Raredon et al. 2023), and CellChat v2 (Jin et al. 2025) have been developed for inferring CCIs from single-cell spatial measurements. However, no method effectively models both long-range and short-range interactions. Finally, CCIs play crucial roles in the pathogenesis of many diseases (Stubington et al. 2017), and the analysis of these interactions holds the potential for developing effective therapeutics. Yet, existing methods like COMMOT and scSeqComm (Baruzzo et al. 2022) do not support *in silico* perturbation to utilize these impacts for exploring potential treatments, such as receptor antagonists.

Agent-based models (ABMs) simulate the actions and interactions of autonomous agents to understand complex system behaviors (Bonabeau 2002). Agents in ABMs represent entities with specific characteristics, behaviors, and interaction rules. ABMs have several advantages for modeling biological systems. First, ABMs capture emergent phenomena arising from collective agent interactions, revealing insights into cellular community function (Bonabeau 2002; Yu and Bagheri 2020). Second, they model heterogeneous cells as agents with distinct statuses, capturing variations in cell types, states, and behaviors (Bonabeau 2002; Yu and

**Corresponding authors:** [yueli@cs.mcgill.ca](mailto:yueli@cs.mcgill.ca), [jun.ding@mcgill.ca](mailto:jun.ding@mcgill.ca)

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.279771.124>. Freely available online through the *Genome Research* Open Access option.

© 2025 Raghavan et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Bagheri 2020). Lastly, ABMs facilitate the straightforward and effective adjustment of agent rules, enabling perturbations in agent behavior.

To bridge the aforementioned gaps, we introduce CellAgentChat, a comprehensive ABM framework for inferring and visualizing CCIs from scRNA-seq and spatial transcriptomics data. Leveraging the advantages of ABMs, CellAgentChat offers versatile functionalities for CCI inference and visualization. By adopting an agent-based perspective, CellAgentChat integrates spatial and biological data, providing a more precise and comprehensive understanding of cellular interaction dynamics at both individual cell and system levels. Moreover, CellAgentChat enables *in silico* perturbation analysis of the downstream effects of cellular interactions on gene expression, facilitating the identification of potential therapeutic receptor targets. Notably, the adaptive framework of CellAgentChat allows for diverse functionalities by adjusting the behavior rules associated with cell agents. This feature enables the identification of long-range and short-range LR interactions. Finally, CellAgentChat can analyze temporal dynamics, shedding light on how CCIs influence cellular state and behavior over short time frames. Our evaluation of CellAgentChat, using several publicly available scRNA-seq and spatial transcriptomics data sets, reveal superior CCI inference compared to other benchmarked methods. Together, we introduce a new framework based on ABMs for future advancements in the study of cellular interactions implicated in vital biological processes.

## Results

### Overview of CellAgentChat

CellAgentChat constitutes a comprehensive ABM framework integrating gene expression data and existing knowledge of signaling LR interactions to compute the probabilities of CCI. Utilizing the principles of ABM, we characterize each cell agent through various cellular attributes (states), including cell identities (e.g., cell type), gene expression profiles, LR universe, and spatial coordinates (Fig. 1A). CCIs are measured by the quantity of ligands emitted by sender cells and subsequently absorbed by receiver cells. This process hinges upon three interrelated components: ligand diffusion rate ( $\gamma^l$ ), receptor receiving rate ( $\alpha^r$ ), and receptor conversion rate ( $\beta^r$ ) (Fig. 1B–D).

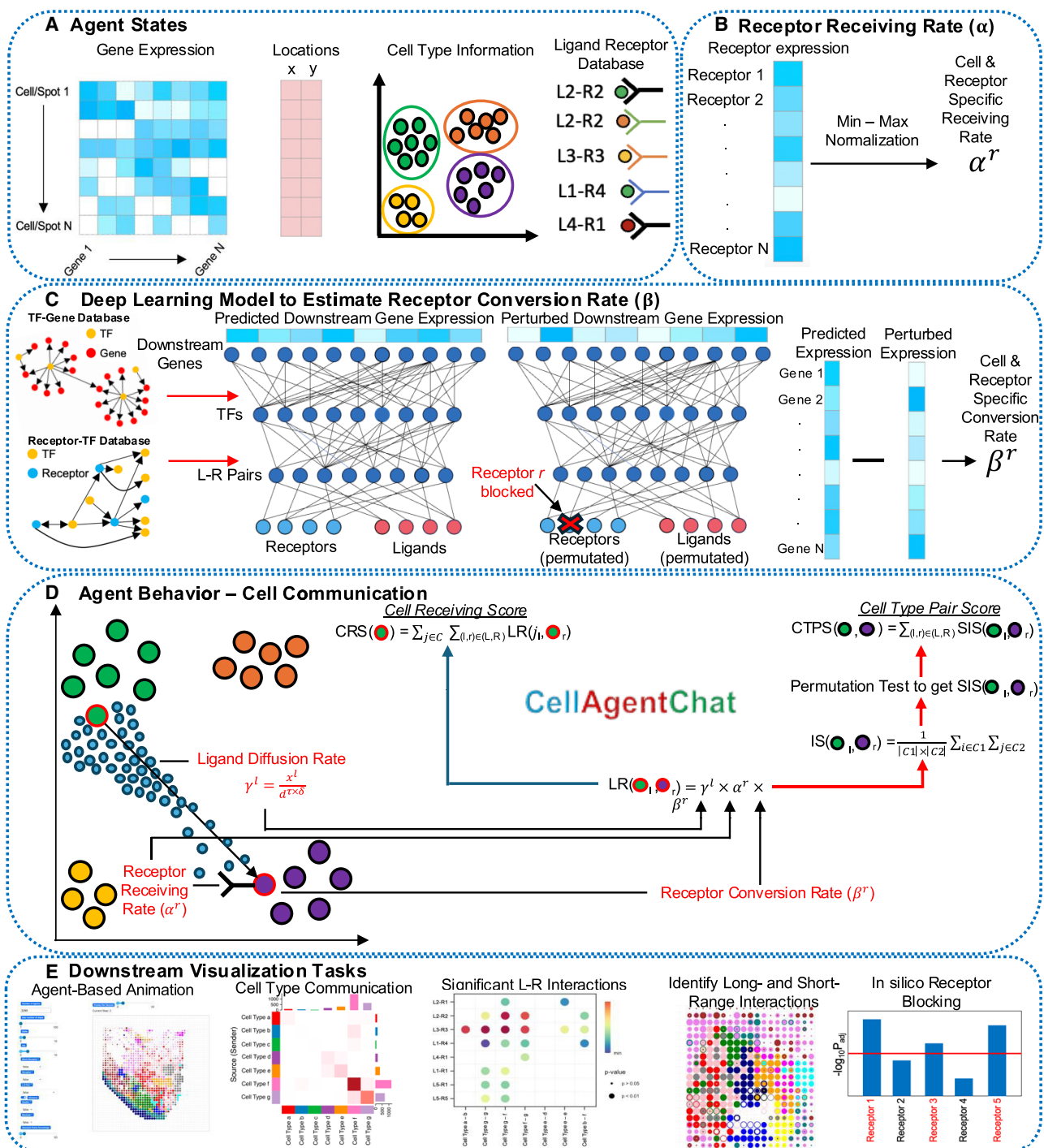
The initial step involves the estimation of the receptor receiving rate ( $\alpha^r$ ), which signifies the fraction of secreted ligands received by the receptor on the receiver cell (Fig. 1B). We then leverage a neural network to evaluate the receptor conversion rate ( $\beta^r$ ), which indicates how strongly an interaction influences downstream gene expression (Fig. 1C). Considering the possibility that a similar interaction strength between different LR pairs can yield diverse impacts on downstream gene expression dynamics (Armingol et al. 2021), this step is essential. Next, we compute the ligand diffusion rate ( $\gamma^l$ ), which quantifies the ligands secreted by the sender cell that reach the target receiver cell (Fig. 1D; Eq. 1 in Methods). This rate is a function of both ligand expression,  $x^l$ , intercellular distance,  $d$ , and two parameters:  $\tau$  and  $\delta$ .  $\tau$  represents the degrees of spatial freedom concerning the single-cell data, typically set to two for spatial transcriptomics data that is derived from two-dimensional slices.  $\delta$  signifies the decay rate of ligand diffusion, which prioritizes interactions over long or short distances. CellAgentChat quantifies interactions at multiple levels: between individual cells using LR scores, per-cell receiving strength via the cell receiving score, and between cell type populations.

Together, CellAgentChat offers a suite of downstream functionalities (Fig. 1E), several of which are absent in existing methods.

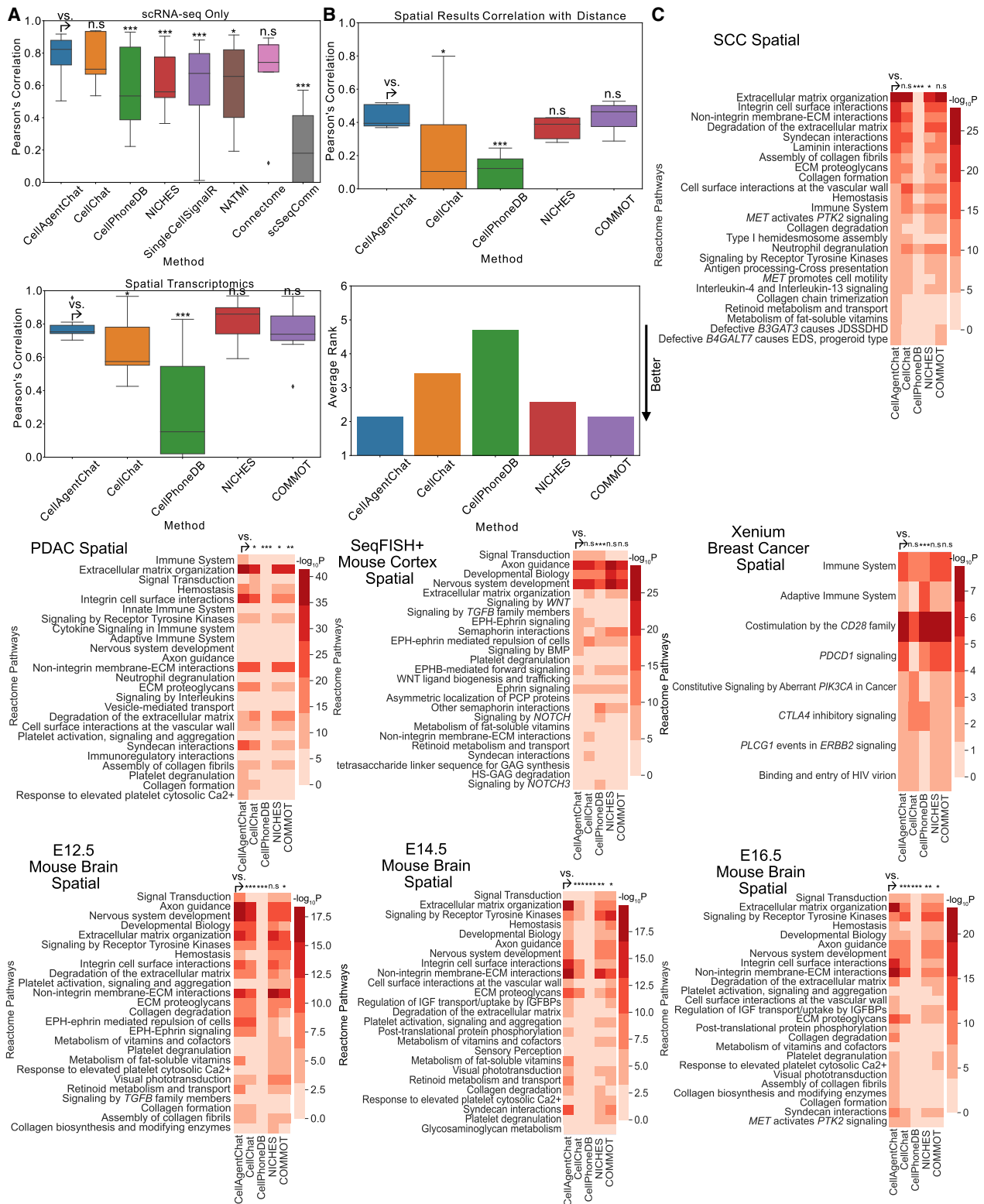
### Benchmarking CellAgentChat’s CCI inference across diverse application scenarios

We undertook a comprehensive benchmarking of CellAgentChat against nine prominent CCI inference methods with the help of the LIANA+ benchmark (Dimitrov et al. 2024): CellPhoneDB (v5) (Efremova et al. 2020), CellChat (v2) (Jin et al. 2021, 2025), NICHES (Raredon et al. 2023), COMMOT (Cang et al. 2023), Scriabin (Wilk et al. 2024), Connectome (Raredon et al. 2022), SingleCell-SignalR (Cabello-Aguilar et al. 2020), NATMI (Hou et al. 2020), and scSeqComm (Baruzzo et al. 2022). We evaluated the performance of all these methods across five different spatial transcriptomics data sets at both single-cell (or subcellular) and spot-level resolution: Stereo-seq mouse developing hippocampus at three developmental time points (E12.5, E14.5, E16.5) (Chen et al. 2022), 10x Genomics Visium human squamous cell carcinoma (Ji et al. 2020), 10x Genomics Visium pancreatic ductal adenocarcinoma (Moncada et al. 2020), SeqFISH+ mouse somatosensory cortex (Eng et al. 2019), and Xenium human breast cancer (Janesick et al. 2023). To deliver a comprehensive benchmarking, we compared all methods on these data sets from several key perspectives.

First, we compared the agreement of the cell–cell communication networks between each benchmarked method. We began by benchmarking across methods using only scRNA-seq data to validate the robustness of our CCI calculation framework before integrating spatial data. We assessed the cell–cell communication networks generated by CellAgentChat and compared them with the outcomes of other methods. Because of the absence of a definitive ground-truth, we adopted a consensus approach. This involved normalizing the interaction matrices computed by each method and averaging them to create an ensemble interaction matrix, serving as our new consensus reference. Subsequently, we computed the Pearson’s correlation between the ensemble interaction matrix and the inferred communication patterns of each method for comparison. Across all data sets, the results from CellAgentChat demonstrated superior performance, achieving a higher median Pearson’s correlation compared to all methods (Fig. 2A, top). Notably, CellAgentChat showed significantly better results than CellPhoneDB, NICHES, SingleCellSignalR, NATMI, and scSeqComm, further underscoring its consistency and accuracy in inferring communication networks (Fig. 2A, top). On examining individual data sets, CellAgentChat consistently outperformed CellPhoneDB, scSeqComm, and Scriabin in all cases (Wilcoxon one-sided test;  $P < 0.05$ ) (Supplemental Figs. 1–3). Additionally, CellAgentChat and CellChat consistently showed high performance, but CellAgentChat exhibited less variability across all evaluated data sets (Fig. 2A; Supplemental Figs. 1–3). Upon incorporating spatially resolved transcriptomics data, CellAgentChat continued to perform with a high degree of consistency among the state-of-the-art methods. Specifically, CellAgentChat maintained superior performance compared to CellPhoneDB and also outperformed CellChat (Wilcoxon one-sided test;  $P < 0.05$ ) (Fig. 2A, bottom; Supplemental Figs. 1, 4, 5). NICHES had a slightly higher median score (Wilcoxon one-sided test;  $P = 0.76$ ), and COMMOT showed a slightly lower median (Wilcoxon one-sided test;  $P = 0.53$ ), but neither difference reached statistical significance. Thus, CellAgentChat performed comparably to these methods, consistently exhibiting lower variability (9.9% smaller



**Figure 1.** Overview of CellAgentChat. (A) CellAgentChat requires the following agent states for each cell: gene expression from scRNA-seq, cell type/cluster information, and LR database. Spatial coordinates (from spatial transcriptomics data) are optional. (B) The receptor receiving rate is the probability that an interaction is received by a receptor, calculated by non-parametric min-max normalization of receptor expression. (C) Our deep learning model inputs the gene expression of ligands and receptors, predicting the expression of all genes. The model incorporates prior knowledge of TF interactions. Receptor blocking is simulated by permuting receptor features and scaling down expression. The receptor conversion  $\beta^r$  of the receptor is estimated from the total change of the predicted expression after the perturbation. (D) LR interaction scores depend on ligand diffusion rate ( $\gamma^l$ ), receptor receiving rate ( $\alpha^r$ ), receptor conversion rate ( $\beta^r$ ), and optionally spatial distance (d) between cells. Hyperparameters  $\tau$  (spatial freedom, default = 2) and  $\delta$  (ligand decay rate, default = 1) control interaction range. Using the LR score, we can calculate the interaction score between two clusters/cell types and the cell receiving score (CRS) for a cell. After a permutation test, significant interactions are identified and used to calculate the cell type pair score (CTPS) between two clusters. (E) CellAgentChat enables agent-based animation to visualize cell communication, identify significant LR pairs, and simulate receptor blocking to find the most perturbed downstream genes for potential therapeutic drug discovery.



**Figure 2.** Benchmarking of CellAgentChat with existing state-of-the-art methods. (A) Box plot comparing the Pearson's correlation between inferred cell communication by individual methods and the ensemble of all methods, without (*top*) and with (*bottom*) spatial information across all data sets. *P*-values calculated using Wilcoxon one-sided test: (\*)  $P < 0.05$ ; (\*\*)  $P < 0.01$ ; (\*\*\*)  $P < 0.001$ . (B) Box plot comparing the Pearson's correlation between the inferred cell communication for each method and the inverse cell-distance matrix (*top*), across data sets. *P*-values calculated using Wilcoxon one-sided test: (\*)  $P < 0.05$ ; (\*\*)  $P < 0.01$ ; (\*\*\*)  $P < 0.001$ . *Bottom* plot in panel shows the average ranking (lower is better) of each method's correlation across data sets. (C) The Reactome pathway enrichment analysis, performed on ligands and receptors sourced from the top 100 LR pairs inferred by each method across various data sets using spatial transcriptomics data. *P*-values calculated using Wilcoxon one-sided test: (\*)  $P < 0.05$ ; (\*\*)  $P < 0.01$ ; (\*\*\*)  $P < 0.001$ .

interquartile range compared to NICHES and 10.9% compared to COMMOT) and reinforcing its position among the top-performing methods.

Secondly, we examined the consistency of CellAgentChat's predictions with spatial information, focusing on the spatial proximity of stronger interactions. Specifically, we measured the Pearson's correlation between the probability of interaction and the inverse distance between cells, grounding this analysis on the established understanding that cells are more likely to interact when they are in closer proximity (Armingol et al. 2021). We excluded Connectome, SingleCellSignalR, NATMI, scSeqComm, and Scriabin for this comparison as they lack spatial data support. Indeed, cells predicted by CellAgentChat as highly interacting were closer in proximity compared to predictions from other methods (Fig. 2B). We demonstrated that our spatially informed predictions outperformed those of CellPhoneDB and CellChat (Wilcoxon one-sided test;  $P < 0.05$ ) (Fig. 2B), showing superior performance across every data set compared to CellPhoneDB and in all but one data set compared to CellChat (Supplemental Fig. 6). Furthermore, we show comparable performance to NICHES and COMMOT across data sets (Fig. 2B; Supplemental Fig. 6). Overall, CellAgentChat and COMMOT had the best average rank (lower is better) across all spatial data sets (Fig. 2B, bottom).

### Benchmarking CellAgentChat's CCI predictions via pathway enrichment analysis and ligand–receptor prediction agreement

After validating the accuracy of CellAgentChat in predicting CCIs through consensus and spatial proximity metrics, we conducted Reactome pathway enrichment analysis to further confirm the biological relevance of the inferred LR pairs. This step serves as an additional layer of validation. We conducted computational validations 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 conducted for the results using only scRNA-seq data underscored CellAgentChat's strength over other methods. CellAgentChat consistently outperformed Scriabin across all data sets (Supplemental Fig. 7). Compared to CellPhoneDB and SingleCellSignalR, CellAgentChat achieved superior performance in six of the seven data sets and comparable results in the remaining one. Additionally, it outperformed scSeqComm and NATMI in the majority of data sets, while showing comparable performance in the rest (Supplemental Fig. 7). Upon integrating spatial data, pathway enrichment analysis continued to highlight CellAgentChat's superiority over CellPhoneDB and CellChat (Fig. 2C). CellAgentChat continues to perform better than CellPhoneDB in all data sets (Fig. 2C; Supplemental Fig. 7B). CellAgentChat also demonstrated superior performance to CellChat, NICHES, and COMMOT in the majority of data sets (Fig. 2C; Supplemental Fig. 7B). Overall, pathway enrichment analysis supported the CCI inference precision of CellAgentChat over other benchmarked methods.

Finally, we calculated the overlap in LR pairs among the top 100 identified LR pairs across the nine methods. We observed that the overlap between CellAgentChat and the other methods, particularly CellChat, NICHES, and COMMOT, was high (Supplemental Figs. 8, 9). This further confirms CellAgentChat's effectiveness, as its LR signaling predictions align with state-of-the-art methods.

Building on the findings from the preceding benchmarking sections, which underscore CellAgentChat's superior, or at least comparable, performance compared to other methods, it is crucial to recognize that beyond these achievements, CellAgentChat also introduces distinct functionalities not available in other tools (Table 1). Unique to CellAgentChat is its integrated ABM animation platform, which facilitates the visualization of individual-level cellular interactions and supports dynamic simulations. Additionally, through easy and flexible adjustment of agent behavior, CellAgentChat empowers the use of *in silico* receptor blocking to explore potential interventions effectively and serves as a valuable tool for identifying both long- and short-range interactions. These functionalities will be discussed in more detail in the following sections, showcasing how they contribute to the broader utility and potential of CellAgentChat.

### CellAgentChat identifies signaling pathways using spatial transcriptomics data at single-cell and multi-cell/spot resolutions

To demonstrate CellAgentChat's accuracy in predicting CCIs, we applied it to single-cell spatial transcriptomics data from the Stereo-seq platform during mouse hippocampus development (Chen et al. 2022). We focused on identifying key signaling pathways involved in the developmental lineage between cells at embryonic day E12.5, E14.5, and E16.5 (Fig. 3A).

CellAgentChat identified 1531 significant interactions involving 579 LR pairs at E12.5 (Fig. 3B,C). Microglial (Micro), fibroblasts (Fibro), erythrocytes (Ery), endothelial cells (Endo), and glioblasts (GlioB) had the highest number of interactions (Fig. 3B), aligning with their known role in neurogenesis (Tong and Vidyadaran 2016; Li et al. 2022; Vogenstahl et al. 2022). Key LR pairs identified included the *ApoE* signaling pathway (Fig. 3C), consistent with previous findings on the regulatory role of *ApoE* in neurogenesis and regulating neurodegenerative diseases (Tensaouti et al. 2018; Yin et al. 2023). Additionally, we identified significant collagen signaling (*Col1a1*, *Col1a2*, and *Col8a1*), promoting hippocampal neurogenesis in mice (Kakoi et al. 2012; Swonger et al. 2016). To systematically assess the reliability of the LR interactions predicted by CellAgentChat, we employed Reactome pathway enrichment analysis. Our analysis revealed significant enrichment of pathways related to neurogenesis, including axon guidance, nervous system development, developmental biology, and extracellular matrix (ECM) organization (Cope and Gould 2019; Fromme and Zigrino 2022), with CellAgentChat outperforming existing methods (Fig. 2C). Further analyses at E14.5 and E16.5 yielded consistent results (Supplemental Results; Supplemental Figs. 10, 11).

CellAgentChat seamlessly integrates multiple spatial sections into a single analysis, as demonstrated with three combined slices at E16.5 (Supplemental Fig. 10B). This reveals consistent communication patterns across all slices (Supplemental Fig. 12A–C), demonstrating the tool's robustness and efficacy in capturing coherent signaling networks (Supplemental Fig. 12D). To further showcase broad applicability, we also applied CellAgentChat without spatial information, obtaining consistent results (Supplemental Fig. 11).

To showcase the broad applicability of CellAgentChat, we applied it to spot-level Visium spatial transcriptomics data from squamous cell carcinomas (SCC) (Ji et al. 2020), with a total of 13 cell populations, including normal keratinocytes (NKer), tumor keratinocyte (TKer), and myeloid dendritic cell (MDC) (Fig. 3D). Treating each spot as a cell, CellAgentChat uncovered key LR pairs driving SCC progression, notably *PKM-CD44*, *MMP1-CD44*, and *HLA-B-*

**Table 1.** Summary of properties and functionalities for all benchmarked CCI inference methods

Methods	Properties				Functions			
	CCI score <sup>a</sup>	Spatial transcriptomics data	Spatial slices integration	Multi-subunit	Intracellular signaling	In silico receptor blocking	Long-/short-range interactions	Dynamic animation
CellAgentChat	Individual cell	✓	✓	X	✓	✓	✓	✓
CellChat	Cell population	✓	✓	✓	X	X	X	X
CellPhoneDB	Cell population	✓	X	✓	X	X	X	X
NICHES	Individual cell	✓	X	X	X	X	X	X
COMMOT	Individual cell	✓	X	✓	✓	X	X	X
Scriabin	Individual cell	X	X	X	✓	X	X	X
Connectome	Cell population	X	X	✓	X	X	X	X
SingleCellSignalR	Cell population	X	X	✓	X	X	X	X
NATMI	Cell population	X	X	✓	X	X	X	X
scSeqComm	Cell population	X	X	X	X	X	X	X

Note: Features of CellAgentChat features include integration of multiple spatial slices, single-cell resolution, in silico receptor blocking, identification of short- and long-range interactions, and dynamic animation of CCIs over time.

<sup>a</sup>Individual cell CCI score can be aggregated to obtain cell population CCI scores.

CANX, consistent with prior studies (Fig. 3E; Pontén et al. 2008; Atasoy et al. 2009; Ludwig et al. 2019; Schaafsma et al. 2021; Boschert et al. 2022; Zhang et al. 2022). Reactome pathway analysis revealed significant enrichment of pathways relevant to SCC, including ECM organization (Fromme and Zigrino 2022; Parker et al. 2022), immune system (Ahrends and Borst 2018), *MET* signaling (Szturz et al. 2017; Centuori and Bauman 2022), and laminin interactions (Fig. 2C; Ramovs et al. 2017; Meireles Da Costa et al. 2021). CellAgentChat outperformed other methods in identifying these pathways, highlighting its superior ability to predict biologically relevant interactions in cancer (Fig. 2C).

To further demonstrate the broad applicability of our model, we also applied CellAgentChat to diverse single-cell spatial data sets across different platforms, including Xenium breast cancer, SeqFISH+ mouse somatosensory cortex, and human pancreatic ductal adenocarcinoma (PDAC) (Supplemental Figs. 13, 14), thereby confirming its compatibility and functionality across diverse spatial technologies.

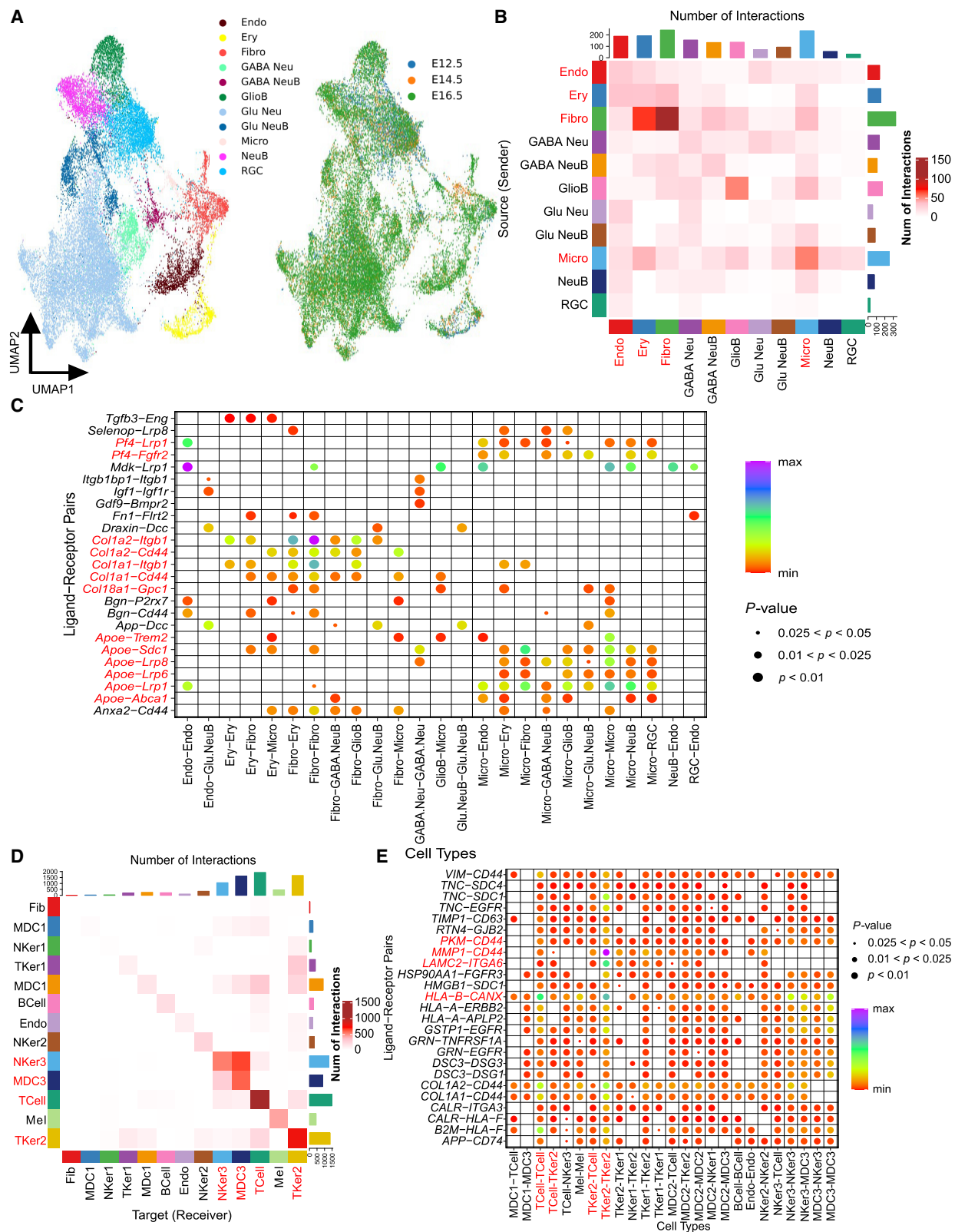
### CellAgentChat enables in silico receptor block perturbation to identify potential therapeutics

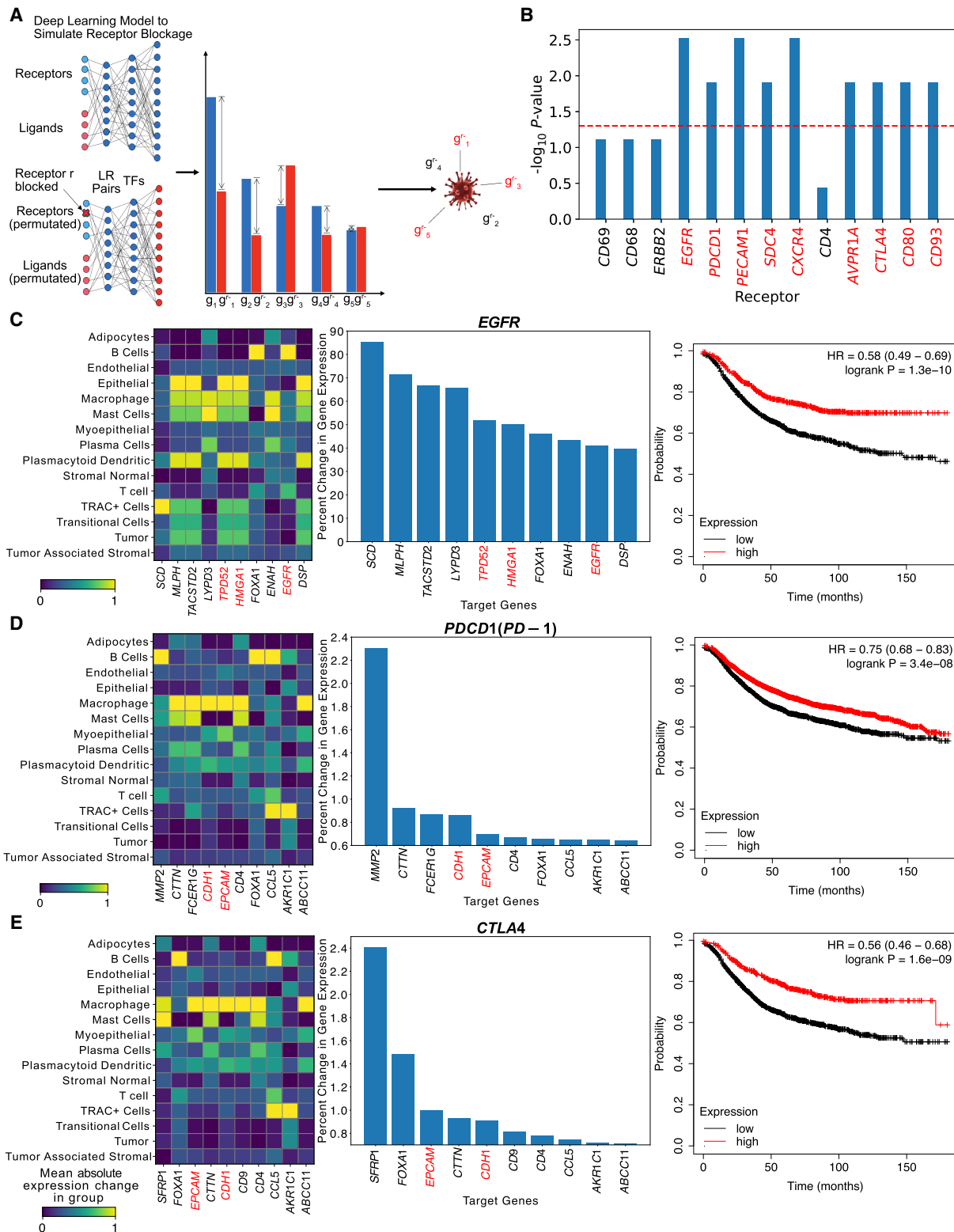
CellAgentChat facilitates in silico receptor blocking to simulate the interruption of specific signaling pathways. By computationally blocking receptors, researchers can observe and predict the impact on downstream genes, which may be associated with disease processes. This approach allows for the identification of potential therapeutic receptor targets without the need for physical experiments. Such in silico perturbations offer a methodical approach to therapeutic discovery, providing valuable insights into

the molecular mechanisms of diseases and accelerating the identification of potential treatments.

We harnessed CellAgentChat to build upon previous research on breast cancer exploring the Xenium human breast cancer data set (Janesick et al. 2023), aiming to identify significant receptors related to the disease and cultivate innovative therapeutic approaches. The first step of our process entailed using CellAgentChat to pinpoint 13 key receptors involved in the top 25 LR interactions for the cellular communications in the data set (Supplemental Figs. 13D, 15A). We then simulated the blocking of these receptors individually, akin to the actions of receptor inhibitor drugs. We deployed a feedforward neural network that was designed to calculate each receptor's conversion rate and was trained to predict the expression of all genes using the ligand and receptor expression as inputs. The network connections were constrained to the known LR interactions, LR transcription factor (TF), and TF genes along the forward pass, respectively (Fig. 4A). Using the trained network, we predicted the impact of receptor blocking on downstream gene expression and identified the most perturbed downstream genes for each in silico blocked receptor. We subsequently assessed these perturbed genes using DisGeNET (Piñero et al. 2015) enrichment analysis to evaluate the potential therapeutic implications of blocking specific receptors (Fig. 4A). Our analysis with CellAgentChat indicated that blocking nine of the 13 identified top interacting receptors in the breast cancer data set—specifically *EGFR*, *PDCD1*, *PECAM1*, *SDC4*, *CXCR4*, *AVPR1A*, *CTLA4*, *CD80*, and *CD93* (hereafter referred to as candidate receptors)—significantly disrupts genes associated with breast cancer (Fig. 4B).

In an ablation study to highlight the benefits of integrating biological priors, we tested a fully connected feedforward network





**Figure 4.** CellAgentChat enables in silico receptor blocking to identify potential receptor antagonists for therapeutics. (A) Schema of the in silico perturbation analysis using our neural network to block receptors and identify the most perturbed downstream genes. Blue bars show unperturbed predicted expression; red bars show perturbed predictions. (B) Breast cancer gene set enrichment ( $-\log_{10}$  adjusted  $P$ -values; binomial test, FDR corrected) for significantly perturbed genes after blocking 13 selected receptors. Receptors highlighted in red are significantly enriched. (C–E) Matrix plots showing mean expression changes per cell type for *EGFR*, *PDCD1 (PD-1)*, and *CTLA4*, respectively (left). Top 10 downstream genes with highest expression change after blocking *EGFR*, *PDCD1 (PD-1)*, and *CTLA4*, respectively; red highlights indicate breast cancer-associated genes (middle). Survival analysis comparing high versus low receptor expression groups for *EGFR*, *PDCD1 (PD-1)*, and *CTLA4*, respectively (right).

without the LR TF and TF target prior constraints. The removal of these biological constraints resulted in a noticeable decrease in specificity: four receptors previously identified as nonsignificant were now marked as significant (Supplemental Fig. 15B).

CellAgentChat's analysis reveals that specific cell types, including tumor cells, stromal cells, endothelial cells, and macrophages, are highly engaged in interactions with the candidate receptors (Supplemental Fig. 16). By simulating the blocking of these receptors, CellAgentChat predicts a significant impact on these cellular interactions, demonstrating the tool's potential to identify effective therapeutic strategies.

Although studies have associated all of the candidate receptors examined in this study with breast cancer, the candidate receptors we identified—*EGFR*, *PDCD1* (also known as *PD-1*), and *CTLA4*—appear to play a particularly crucial role in breast cancer (Fig. 4C–E; Masuda et al. 2012; Ralser et al. 2021; Ghahremani Dehbokri et al. 2023). Furthermore, therapeutic treatments targeting the three receptors in breast cancer have been previously documented. Both lapatinib and pembrolizumab are FDA-approved drugs designed to specifically target and disrupt pathways involving *EGFR* and *PDCD1*, respectively. Ipilimumab and tremelimumab are approved by the FDA to suppress the role of *CTLA4* in melanoma (Ghahremani Dehbokri et al. 2023). Although ipilimumab and tremelimumab are not FDA-approved for breast cancer treatment, studies have shown their positive impact in this context (Ghahremani Dehbokri et al. 2023). This corroborates the potential application of CellAgentChat to identify receptor-mediated therapeutics. Moreover, blocking the *EGFR*, *PDCD1*, and *CTLA4* receptors in silico perturbs the genes associated with breast cancer (from DisGeNET) (Fig. 4C–E). Our analysis revealed that blocking these receptors perturbs several disease genes, including *TPD52*, *CDH1*, and *EPCAM*, among others, which have been previously linked to breast cancer (Soysal et al. 2013; Corso et al. 2018; Ren et al. 2021). Numerous perturbed genes notably influenced immune cell types like mast cells and macrophages (Fig. 4C–E, left), which play a crucial role in the tumor microenvironment and immune response against cancer cells. All other breast cancer disease target genes perturbed are highlighted in red in the figure (Fig. 4C–E; Supplemental Figs. 15C–E, 17A–C).

To validate the accuracy of the candidate drug targets captured by CellAgentChat, we conducted a survival analysis of breast cancer patients over time, comparing populations with high and low expression levels of each candidate therapeutic receptor. The analysis revealed statistically significant differences in overall survival probabilities over 150 months for patients with varying levels of expression of *EGFR*, *PDCD1*, and *CTLA4* (Kaplan-Meier log-rank test;  $P = 1.3 \times 10^{-10}$ ,  $P = 3.4 \times 10^{-8}$ ,  $P = 1.6 \times 10^{-9}$ , respectively) (Fig. 4C–E). Additionally, survival analysis of all other candidate receptors exhibited statistically significant differences in survival probabilities between the two patient groups (Supplemental Figs. 15C–E, 17A–C). These findings indicate that the identified candidate receptors would serve as suitable therapeutic targets for treating breast cancer, highlighting the powerful functionality of CellAgentChat's in silico receptor blocking to predict and validate potential therapeutic interventions.

To demonstrate the applicability of CellAgentChat's in silico perturbation on other spatial transcriptomics technology, we also demonstrate the in silico perturbation capabilities of our model on a spot-level resolution PDAC data set (Moncada et al. 2020), where we identified *NOTCH3*, *PLAUR*, *ITGA11*, and *CD74* as potential therapeutic targets (see Supplemental Results; Supplemental Figs. 18, 19).

To validate the effectiveness of CellAgentChat in predicting the effects of receptor perturbation on disease target genes, we compared our approach with scGen (Lotfollahi et al. 2019). Using the breast cancer data set, we found that scGen did not identify any receptors whose perturbation significantly disrupted genes associated with breast cancer (Supplemental Fig. 20A). Additionally, receptor perturbation using CellAgentChat showed a higher specificity in affecting breast cancer-related genes compared to scGen, especially for well-known breast cancer receptors *EGFR* and *ERBB2* (Supplemental Fig. 20B). This validation was also conducted for the PDAC data set (Supplemental Results; Supplemental Fig. 20C–E).

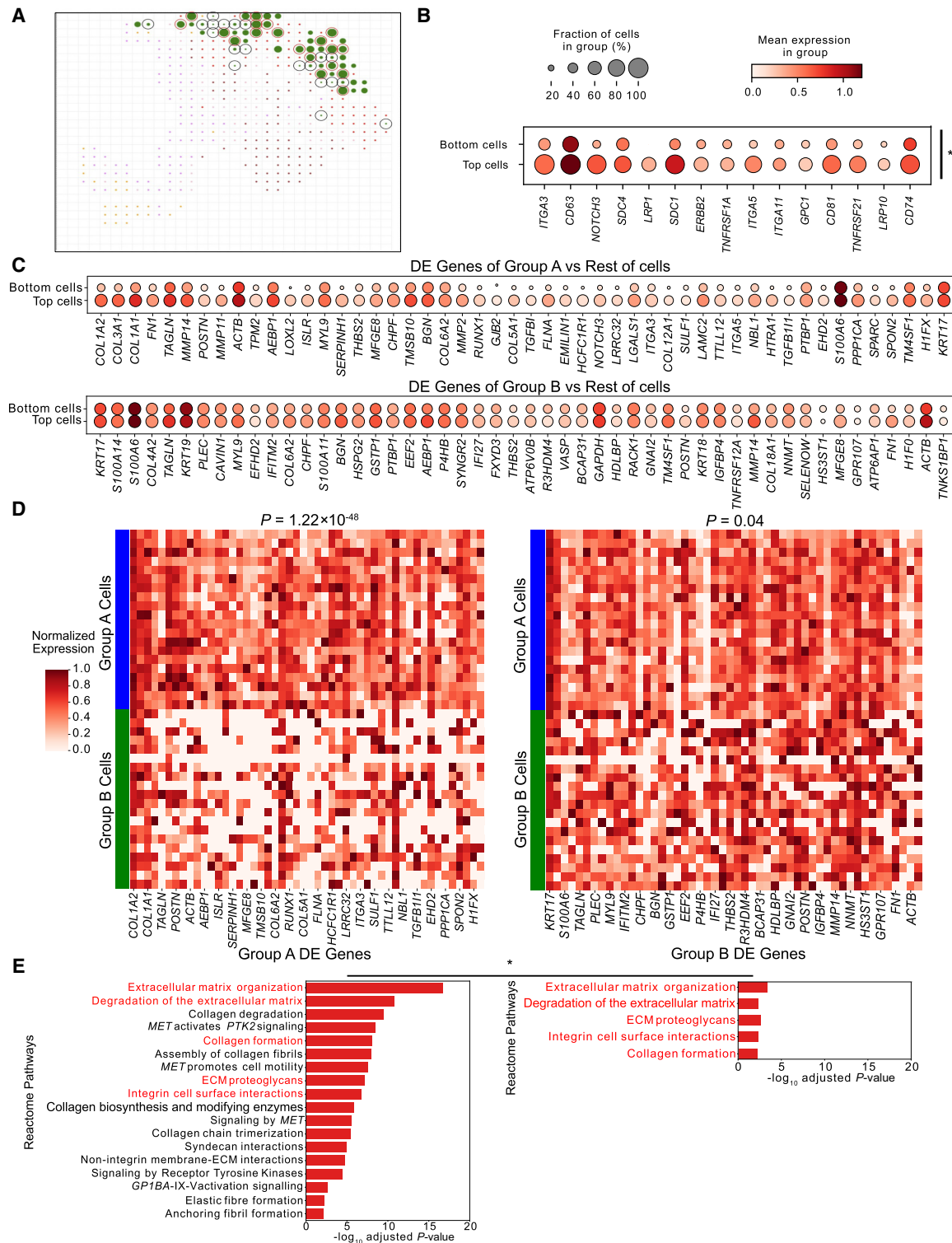
Analyzing perturbation effects in both diseased and normal tissues provides critical insights into underlying communication networks. CellAgentChat's ABM capabilities revealed three receptors—*LRP1*, *CAV1*, and *CD93*—that significantly disrupt PDAC-associated genes when blocked in tumor regions, while having no effect in normal tissue (see Supplemental Results; Supplemental Figs. 21–25).

### CellAgentChat facilitates agent-based visualization of interactions associated with individual cells

The ABM approach used in CellAgentChat allows us to examine individual cell interactions rather than solely focusing on cell populations. We applied CellAgentChat to a spot-level human PDAC data set (Moncada et al. 2020), with six distinct cell populations: ductal (Duct), acinar (Acin1, Acin2, and Acin3), stromal (Strom) cells, and two classes of cancer (Can1 and Can2) (Supplemental Fig. 13B,E,F). To demonstrate CellAgentChat's capability in revealing insights into cell interactions, we identified the 20 most receiving spots (group A, red circles) based on the cell receiving score (Methods) and the least receiving spots (group B, black circles) within Can2 cells (green cells) (Fig. 5A).

The spots with higher CRS (more interactive) generally reside in the outer layer of the tissue (invasive front), consistent with existing studies (Friedl and Alexander 2011). These two cell groups of distinct interacting patterns 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. 5B). 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. 5B).

Next, we systematically explored the differentially expressed (DE) genes of these two groups by comparing the spots within the group with the rest of the spots in the data set. The differential genes associated with group A show significantly higher expression in group A cells compared with group B (Wilcoxon one-sided test;  $P = 1.22 \times 10^{-48}$ ) (Fig. 5C,D). In contrast, the expression of group B DE genes is relatively similar across groups (Fig. 5C,D). To identify the biological functions associated with these groups, we conducted Reactome pathway enrichment on the top 50 DE genes for both groups (Fig. 5E). Further, group A was associated with several pathways related to the extracellular matrix and collagen, known to be crucial in PDAC tumor progression (Perez et al. 2021). Altogether, CellAgentChat identifies complex interactions at the individual cell level, revealing diverse roles in disease progression, highlighting the tool's potential to uncover new insights in cellular communication.



**Figure 5.** CellAgentChat facilitates inference of cellular interactions associated with individual cells in the PDAC data set. (A) CellAgentChat's animation platform enables the detection of communication strength variations in individual cancer cells. The platform highlights the cell receiving score (Methods) of Can2 cells (colored green). The top 20 cells with the highest CRS (group A) are marked by red circles and the bottom 20 cells with the lowest CRS (group B) are marked by black circles. (B) Expression analysis of receptors in the top 25 LR interactions shows significantly higher expression in group A compared to group B (Wilcoxon test one-sided;  $P < 0.05$ ). (C) Mean expression of differentially expressed (DE) genes upregulated for group A spots relative to rest of the spots (top) and upregulated for group B cells relative to rest of the spots (bottom). (D) Heat maps display the DE genes for group A (left) and group B (right). The DE genes of group A are mainly expressed by group A and not by group B (Wilcoxon test one-sided;  $P < 0.001$ ). Both groups express the DE genes of group B but are more significantly expressed by group B spots (Wilcoxon test one-sided;  $P < 0.05$ ). (E) The results of the Reactome pathway analysis using group A DE genes (left) and group B DE genes (right). Red highlights indicate shared pathways.

## CellAgentChat can be tuned to identify long- and short-range interactions

Our understanding of the specific distances at which LR pairs operate is evolving but remains incomplete. These interactions can occur closely between cells or span longer distances, influenced by factors like specific ligands/receptors, tissue environment, and physiological conditions. CellAgentChat's ABM captures detailed cell behaviors, including the diffusion of signaling ligands, enabling modeling of both short- and long-range interactions at the single-cell level by adjusting the delta ( $\delta$ ) factor to control the ligand diffusion rates (Methods).

Using CellAgentChat, we investigated the spatial distance tendencies of cellular communication within the mouse somatosensory cortex leveraging data from SeqFISH+ technology (Supplemental Fig. 14; Eng et al. 2019). To this end, we utilized the CellChat database, which categorizes LR pairs into “secreted signaling” for long-range interactions and “cell–cell contact” for short-range interactions based on known distance ranges.

To identify short-range interactions, we used a  $\delta=10$  to constrain interactions to shorter distances and used  $\delta=0.1$  to identify long-range interactions. To validate our results, we compared the top interacting LR pairs with CellChat's annotations. Under our short-range mode, the majority of our top 10, 15, 20, and 25 LR pairs were annotated as “cell–cell contact” (short-range) (Fig. 6A). Specifically, 80% of the top 15 LR pairs identified in the short-range mode fell into the “cell–cell contact” category (Fig. 6A). In the long-range mode, most of the top LR pairs identified were classified as “secreted signaling” (long-range), with 75% of the top 15 LR pairs falling into this category (Fig. 6A). Recently, Liu et al. (2022) have developed a method to identify long-range and short-range interactions using optimal transport (OT). Compared to their approach, CellAgentChat demonstrated superior accuracy in capturing short-range interactions and similar accuracy in identifying long-range interactions (Fig. 6B).

In our short-range mode, we detected LR pairs like *Sema4a-Plxna4*, known for its role in axon guidance in the mouse cortex, over short distances (Neufeld et al. 2016). We further employed our animation platform to visualize the CRS of individual cells for the *Sema4a-Plxna4* LR pair in the short-range mode originating from cluster 9 cells (turquoise cells) (Fig. 6C). The results highlighted higher CRS among cells close in proximity to the cluster 9 cells represented by larger sizes (Fig. 6C, within the red box), compared to cells situated further away (Fig. 6C, within the blue box). This observation underscores the capabilities of CellAgentChat in identifying and modeling short-range interactions.

We conducted pathway enrichment comparisons of the LR interaction pairs predicted by CellAgentChat for both long- and short-range interactions (Fig. 6D). The genes associated with short-range interactions showed enrichment in pathways related to cell adhesion and cell–cell junctions. In contrast, genes associated with long-range interactions were enriched in numerous signaling pathways known for their wide regulatory range (Fig. 6D). This analysis underscores CellAgentChat's ability to distinguish between different types of cellular interactions and its broader utility in understanding how these interactions influence genomic processes.

To further validate the accuracy of CellAgentChat's predictions, we compared the molecular weights of ligands identified in the long-range mode ( $\delta=0.1$ ) versus the short-range mode ( $\delta=10$ ). Our analysis revealed statistically significant differences in the molecular weights of ligands across all top pair ranges

(Supplemental Fig. 26). Specifically, ligands predicted in the long-range mode exhibited significantly lower molecular weights compared to those in the short-range mode. This finding aligns with expectations, as heavier ligands are less likely to diffuse over long distances and are thus more prevalent in short-range interaction (Li et al. 2009). In contrast, lighter ligands, which exhibit higher diffusivity, are more likely to mediate long-range interactions and are enriched in the long-range mode (Li et al. 2009). These results support the biological plausibility of our model's predictions and underscore the potential for integrating ligand molecular weight information into real-world data sets.

Ultimately, we can leverage our understanding of the identified short-range and long-range LR pairs to assign the ligand-specific decay rates ( $\delta$ ), enabling precise control over ligand diffusion ( $\gamma'_{i,j}$ ) and accurate modeling of both interaction types simultaneously.

## CellAgentChat empowers dynamic simulations using agent-based modeling

CellAgentChat provides dynamic simulation capabilities to analyze how CCIs impact the cell states over short periods. At each time point (time step), CCIs are computed, followed by the prediction of gene expression levels for each cell using our trained neural network (Methods). Given uncertainties and various influencing factors affecting gene expression, directly substituting existing values with predicted ones is impractical. Instead, we opt for assessing the directional change of each gene—whether they increase or decrease—which proves to be more robust and easier to predict (Supplemental Fig. 27A; Ding et al. 2018). For each cell at every time step, genes undergoing significant changes are identified, and their expressions are adjusted in the simulation for subsequent steps (Methods; Supplemental Fig. 27A). This iterative process captures the dynamic modulation of gene expression by CCIs, providing insights into how cellular interactions influence genomic responses.

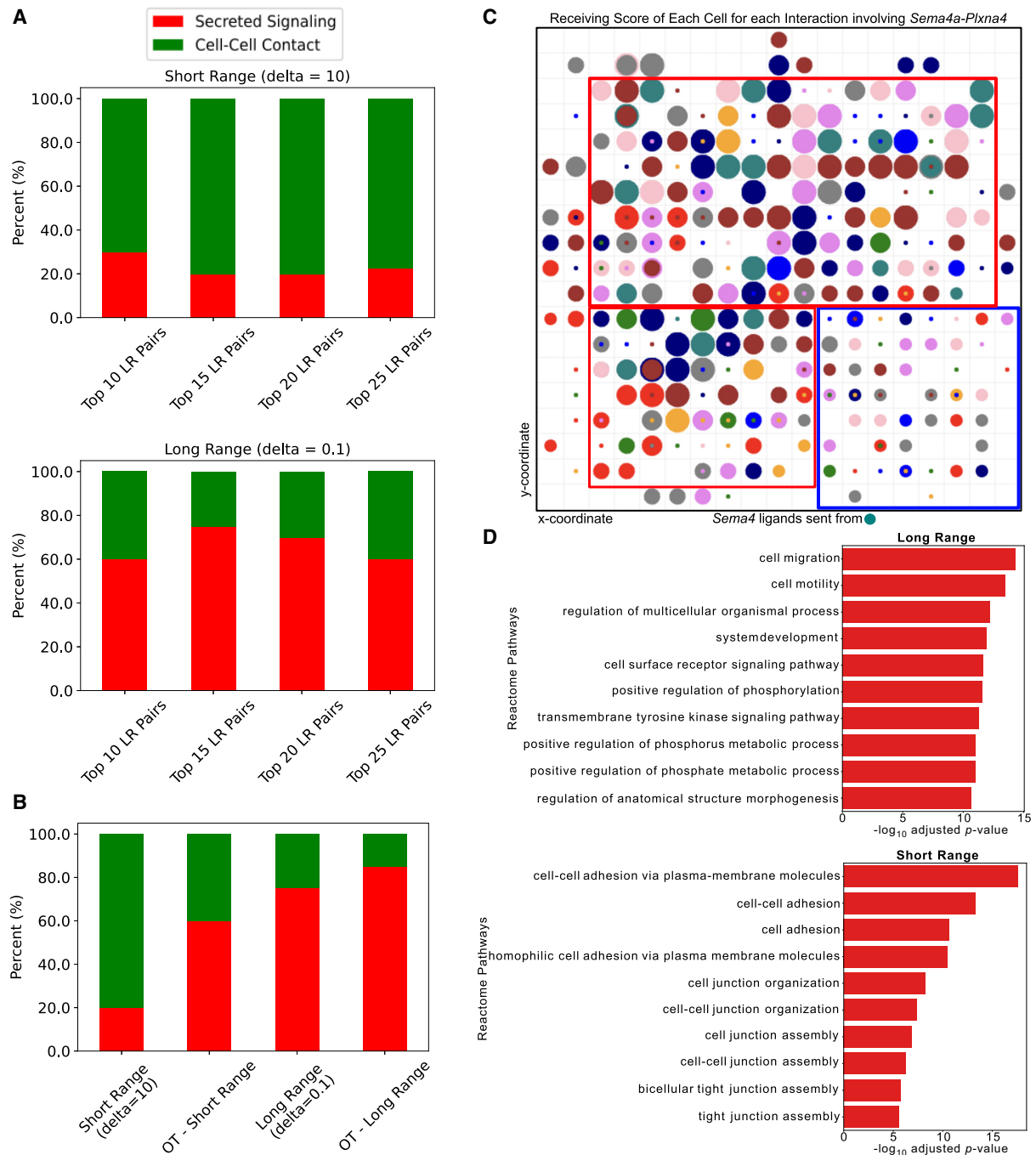
To validate the accuracy of our dynamic simulations, we determined the correlation between the pseudotime trajectory of each gene and its trajectory as established by the dynamic simulation of CellAgentChat across five time steps. Across all genes, we observed a correlation between the pseudotime trajectory and the ABM dynamic trajectory (Supplemental Fig. 27B). We found strong agreement between the pseudotime trajectory and the dynamic ABM trajectory for genes like *ITGA6* (receptor) and *MMP1* (ligand), which showed overall increasing trends over time ( $r^2=0.93$  and  $r^2=0.97$ , respectively) (Supplemental Fig. 27C,D). Similarly, both trajectories indicated a decreasing expression trend for *CALR* (ligand) over time ( $r^2=0.90$ ) (Supplemental Fig. 27E). Moreover, CellAgentChat accurately identified the initial decrease of expression of *CD44* (receptor) followed by a rapid increase ( $r^2=0.89$ ) (Supplemental Fig. 27F). These findings highlight the effectiveness of CellAgentChat in capturing dynamic cellular changes resulting from CCIs within a time frame comparable to the pseudotime represented by the cells. More validation into our dynamic simulations can be found in the Supplemental Results and Supplemental Figure 27.

## Discussion

We introduced CellAgentChat, an agent-based model capable of elucidating cell–cell interactions from single-cell and spatial transcriptomics data. By modeling individual cell agents with simple

behavior rules, CellAgentChat outperforms existing methods in reconstructing CCI networks, integrating spatial context, and identifying meaningful ligand–receptor (LR) interactions (Fig. 2).

CellAgentChat, uncovered key LR pairs in neural development and cancer, such as *ApoE* and collagen signaling in mouse neurogenesis (Fig. 3), *LAMC2-ITGA6* in squamous cell carcinoma



**Figure 6.** CellAgentChat identifies long-range and short-range interactions in the adult mouse cortex data set. (A) Bar plot illustrates the percentage of interactions found by CellAgentChat annotated by CellChat’s database as either “cell–cell contact” (short-range) or “secreted signaling” (long-range) with a delta ( $\delta$ ) value set to 10 (top) and 0.1 (bottom). (B) Bar plot illustrates the percentage of interactions found by CellAgentChat annotated by CellChat’s database as either “cell–cell contact” or “secreted signaling” in comparison to an optimal transport (OT)-based approach (Liu et al. 2022). (C) CellAgentChat’s animation platform depicts the cell receiving score for a “cell–cell contact” *Sema4a-Plxn4* ligand–receptor pair using the short-range mode ( $\delta = 10$ ). The color of the cells (circles) represents the cell cluster. This animation depicts the release of *Sema4a* (ligand) from cells in cluster 9 (represented by turquoise cells). In the animation, cell size corresponds to the cell receiving score. Cells closer to cluster 9 cells exhibit higher CRS (within red square) than cells that are further away (within blue square), which are smaller in size. (D) Reactome pathway analysis of the receptors involved in the top 25 LR interactions in the long-range mode (top) and short-range mode (bottom).

(Fig. 3), and therapeutic receptor targets *EGFR*, *PDCD1*, and *CTLA4* in breast cancer (Fig. 4). CellAgentChat also distinguishes long- and short-range interactions, identifying *Sema4a-Plxna4* as a short-range axon-guidance signal in the mouse cortex (Fig. 6; Neufeld et al. 2016). These results underscore CellAgentChat's versatility and robustness in advancing our understanding of complex biological systems and its potential impact on therapeutic strategies in biomedical research.

In contrast to the existing methods for predicting CCI, CellAgentChat offers various distinctive features. Firstly, it offers a dynamic visualization of cellular interactions related to individual cell agents. This animation tool allows researchers to explore the intricate heterogeneity within cell populations, revealing insights at both the cluster and individual cell levels that were previously inaccessible (Fig. 5). Furthermore, the animation feature enables researchers to visualize and explore the dynamic impact of CCIs over short time frames. As cell agents interact, their states—such as gene expression profiles—continuously update in response to CCIs. This dynamic visualization provides insights into how cellular behaviors and interactions evolve over time, offering a powerful tool to study emergent phenomena and identify key patterns or critical transitions driven by CCIs in a visually intuitive manner. 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. The perturbation pipeline is robust, specifically targeting disease-related genes rather than random genes, ensuring relevance to the biological context. By investigating these genes, CellAgentChat can propose systematic strategies for potential therapeutic interventions, making it a valuable tool for exploring targeted treatments. Thirdly, CellAgentChat can be tuned to detect both long- and short-range interactions, capturing cellular communication across spatial scales. This flexibility allows researchers to study diverse biological processes, such as endocrine signaling at a distance or juxtacrine signaling through direct contact, providing insights into tissue development, immune responses, and disease progression.

In future work, we will explore model-level optimizations tailored specifically for spatial transcriptomics data. These improvements may enhance the model's ability to capture spatial relationships, leading to better biological interpretability in our downstream functionalities. We also plan to extend CellAgentChat by incorporating cell movement and cell death, enhancing the model's realism. Although technically challenging, another promising extension is representing ligands, receptors, and cofactor molecules as individual agents, enabling more comprehensive simulations of complex cellular interactions. This approach would offer deeper insights into cell-molecule dynamics and improve the efficiency of *in silico* perturbations like receptor blocking.

Because of low gene coverage in single-cell resolution spatial datasets (e.g., SeqFISH+, Xenium), spot-level technologies remain popular. CellAgentChat allows agents to represent individual cells or spots for flexibility. Although spot-level representation may overlook some CCIs within spots, it simplifies complex analyses. For higher resolution, tools like SpatialScope (Wan et al. 2023) or scSemiProfiler (Wang et al. 2024) can deconvolve spots into single cells, enabling CellAgentChat to infer interactions, perform *in silico* perturbations, analyze long/short-range interactions, and explore key LR pairs (see GitHub repository for details; Software availability).

In this study, we avoided imputation or denoising techniques, focusing instead on developing CellAgentChat to infer

CCIs directly from raw normalized spatial transcriptomics data. However, imputation methods such as MAGIC (Van Dijk et al. 2018), can enhance gene expression quality. Interested researchers can preprocess data accordingly before applying CellAgentChat, highlighting our framework's flexibility and ease of use (see GitHub repository for guidance; Software availability).

CellAgentChat enhances the study of cellular interactions by providing a versatile ABM framework. It facilitates dynamic visualization, prioritization of interaction ranges and *in silico* receptor blocking, aiding in the identification of therapeutic targets. By integrating spatial information and accommodating various data resolutions, CellAgentChat offers a valuable tool for biomedical research by enhancing the understanding of disease mechanisms and developing therapeutic strategies through the lens of CCIs.

## Methods

### Ligand–receptor database

In this study, we used the ligand–receptor interactions from CellTalkDB (Shao et al. 2021), a curated database of almost 3400 human and over 2000 mouse literature-supported interactions, as the default LR interaction resource for our CellAgentChat method. Besides the default CellTalkDB LR interactions, CellAgentChat also allows for customized LR interactions that users obtain from other sources or manually curate. For our analysis of long-range and short-range interactions, we instead utilized the CellChat database, which classifies LR pairs into “secreted signaling” for long-range interactions and “cell–cell contact” for short-range interactions, based on established distance ranges.

### Scoring cell–cell interactions with the agent-based model

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

#### Cell agent state description

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 record 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. (3) *Cell Agent Batch*: To utilize multiple samples (biological replicates or technical samples) concurrently, we designate the batch of each cell to confine interactions solely between cells of the same batch. (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 dynamic simulations and animation of CCIs” in Methods). By doing so, they can remove all interactions associated with ligands or receptors (or specific LR interactions) from the universe of all cells. This 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 include the gene expression of the cell agent, the spatial position of the cell agent (optional), and the LR universe. To be more specific, the cellular interaction strengths (scores) between cell agents are quantified based on the set of defined rules as discussed below.

1. Ligand diffusion rate ( $\gamma_{i,j}^l$ ): The diffusion rate  $\gamma_{i,j}^l$  estimates the rate of the diffusion of a ligand  $l$  in the microenvironment from sender cell  $i$  to receiver cell  $j$ . 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.

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 hyperparameter,  $\tau$ , is a dimensionality parameter representing the degrees of spatial freedom for spatial data.  $\tau$  is typically set to two for spatial transcriptomics data derived from two-dimensional slices. However, with the advancement of technology and the increasing prevalence of three-dimensional slices (Rao et al. 2021), we can adjust the  $\tau$  parameter accordingly ( $\tau=3$ ). When spatial information is unavailable (e.g., in scRNA-seq data),  $\tau$  is set to 0 so that the distance between all cells in the data will degenerate to 1 ( $d^{\tau=0}=1$ ).

In the absence of precise information about the rates of ligand travel or the operational distance range of LR pairs, we employ the delta ( $\delta$ ) parameter as an approximation for the distance range of LR pairs and model the decay rate of ligands. A user-adjustable slider  $\delta$  (between 0 and 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. Keeping  $\delta$  at the default value of 1 provides no bias toward long-range or short-range interactions. We modeled short-range interactions with a  $\delta=10$  and long-range interactions with  $\delta=0.1$ . By investigating various  $\delta$  values and the corresponding agent communication behaviors defined by these values, we can also determine a specific  $\delta$  value for each LR pair. This approach avoids the assumption that all ligands share the same diffusion rate. The distance between sender cell  $i$  and receiver cell  $j$  is calculated as the Euclidean distance  $d(i, j)$ . The expression of ligand  $l$  in sender cell  $i$  is denoted as  $x_i^l$ . Then, the diffusion rate  $\gamma_{i,j}^l$  can be calculated as below:

$$\gamma_{i,j}^l = \frac{x_i^l}{d(i, j)^{\tau \times \delta}} \quad (1)$$

To address the variability in distance scales across different spatial techniques and data sets, we normalize spatial distances within each data set to a uniform range of 0–100. This normalization ensures that communication strengths are comparable across data sets while preserving the relative spatial relationships. Furthermore, this normalization is crucial for visualizing cell interactions within our ABM framework (see “Agent-based dynamic simulations and animation of CCI” in Methods), as maintaining consistent spatial scales enhances interpretability.

2. Receptor receiving rate ( $\alpha_j^r$ ): The receiving rate  $\alpha_j^r$  associated with a receptor  $r$  represents the probability that an interaction is received by the receptor in cell  $j$ , 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 that the receptor

cannot “receive” any ligands that reach it, and 1 denotes that reached ligands will all be “received” by the receptor. Here, we employed the non-parametric min-max normalization to calculate the receptor receiving rate for each receptor. Min-max normalization has been used in other methods for cell interaction inference (Tran et al. 2022; Peng et al. 2023, 2024; Feng et al. 2024), lending support to our choice. For each cell, the minimum expression value in the cell is subtracted from the expression value of each receptor in that cell and then divided by the difference between the maximum and minimum expression values in that cell (min-max normalization), enabling the identification of highly expressed receptors within each cell, thereby establishing a cell-specific metric.

3. Receptor transcriptome conversion rate ( $\beta^r$ ): CCI play a crucial role in the regulation of gene expression, impacting the transcriptome and ultimately influencing cellular function and fate. 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 (here, we term it as the receptor conversion rate  $\beta^r$ ). The neural network takes as input two vectors,  $x_j^R$ , representing expression of all receptors in cell  $j$ , and  $\frac{1}{N} \sum_{i=1}^N x_i^L$ , representing the average expression of all ligands across the set of all  $N$  cells. These two vectors are then fed into the neural network ( $\mathcal{F}$ ) to predict the expression  $\hat{x}_j$  of all genes in cell  $j$  (regression task). The neural network structure comprises the following components: two input layers, two hidden layers, and one output layer. The size of the two input layers corresponds to the number of ligands and receptors, respectively. The first hidden layer is dedicated to representing specific LR pairs. Each node in this layer receives two connections, one corresponding to a ligand and the other to a receptor, thus encapsulating the unique LR pair interactions. The second hidden layer represents transcription factors involved in downstream signaling pathways. We leverage scSeqComm (Baruzzo et al. 2022), a database that incorporates insights from KEGG (Kanehisa and Goto 2000) and REACTOME (Jassal et al. 2020) pathways to delineate the association between a receptor and its downstream transcription factors, to inform the connections between this layer and the previous one. The output layer is responsible for outputting downstream gene expression predictions. To establish connections between known TFs and their downstream gene targets, we utilize prior knowledge from databases such as TRRUST v2 (Han et al. 2018), HTRIDb (Bovolenta et al. 2012), and RegNetwork. If specific information regarding the target genes of certain TFs is unavailable, we utilize dense connections. We use a stochastic gradient descent optimizer to train the network  $\mathcal{F}$  that minimizes the mean squared error loss  $L$  (between the true observed expression for the set of all cells  $X$  and the predicted expression):

$$L = \frac{1}{N} \sum_{j=1}^N \left( x_j - \mathcal{F} \left( x_j^R, \frac{1}{N} \sum_{i=1}^N x_i^L \right) \right)^2 \quad (2)$$

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 conduct a perturbation-based feature calculation (Mi et al. 2021), by permuting feature values for each receptor, independent of the output. Next, we adjust the expression of the permuted receptor to reflect a ~50% reduction, scaling it by the median expression of all receptors. This represents the perturbed input receptor vector ( $x_j^{R,r}$ ) simulating a ~50%

repression of receptor  $r$  in expression. Although this 50% repression serves as the default setting, we also offer the option to simulate a 100% repression, equivalent to a complete knockout, catering to users who prefer a full knockout scenario. Then, we quantify the importance of the receptor  $r$  ( $I_r$ ) based on the loss increase caused by the perturbation:

$$I_r = L_r - L$$

$$= \frac{1}{N} \sum_{j=1}^N \left( x_j - \mathcal{F} \left( x_j^{Rr}, \frac{1}{N} \sum_{i=1}^N x_i^L \right) \right)^2 - \frac{1}{N} \sum_{j=1}^N \left( x_j - \mathcal{F} \left( x_j^R, \frac{1}{N} \sum_{i=1}^N x_i^L \right) \right)^2 \quad (3)$$

Employing a similar approach as the receiving rate, we normalize the receptor importance into range (0,1) via min-max normalization. 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 (Arnol et al. 2019). We confirmed that this approach does not significantly alter the resulting gene expression of the cells of the training grid (Supplemental Figs. 28, 29).

We included various validations of our deep learning module to estimate the conversion rate as well its capabilities to perform in silico perturbation (see Supplemental Results; Supplemental Figs. 30–33). We also include ablation studies that show the positive impact of using biologically informed neural network using a gene network and the conversion rate as a whole, in line with other studies (Supplemental Fig. 34; Lin et al. 2017). Please refer to the Supplemental Results for further details.

- 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 ( $\gamma^l, \alpha^r, \beta^r$ ):

$$LR(i, j, r) = \gamma_{i,j}^l \times \alpha_j^r \times \beta^r = \frac{x_i^l}{d(i, j)^{\tau \times \delta}} \times \alpha_j^r \times \beta^r \quad (4)$$

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

$$CRS(i) = \sum_{j \in C} \sum_{(l,r) \in (L,R)} LR(j, i, r) \quad (5)$$

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

$$CPS(i, j) = \sum_{(l,r) \in (L,R)} LR(i, j, r) \quad (6)$$

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  $C2$ :

$$IS(C1_l, C2_r) = \frac{1}{|C1| \times |C2|} \sum_{i \in C1} \sum_{j \in C2} LR(i, j, r)$$

$$= \frac{1}{|C1| \times |C2|} \sum_{i \in C1} \sum_{j \in C2} \frac{x_i^l}{d(i, j)^{\tau \times \delta}} \times \alpha_j^r \times \beta^r \quad (7)$$

We also introduce a method to optionally integrate the utilization of pseudotime ordering of cells into the calculation of CCI scores. Please refer to the Supplemental Methods for further details.

Unlike typical ABMs, CellAgentChat integrates various statistical methods to estimate parameters governing agent behaviors. However, it stands out as a robust ABM framework rather than a workflow. Whereas workflows consist of discrete, sequential steps operating independently or semi-independently, frameworks provide an integrated structure where components function cohesively as a unified system. CellAgentChat adheres to core ABM principles (Bonabeau 2002; Masad and Kazil 2015; Yu and Bagheri 2020), modeling cells as agents with unique states (e.g., gene expression profiles and spatial coordinates) that dynamically evolve through interactions with their environment and neighboring cells. The statistical and neural network components in CellAgentChat are not standalone modules, as in a workflow, but integral elements that enhance the framework's ability to simulate realistic interaction rules from data.

### Finding significant ligand–receptor pairs between cell populations

After computing interaction scores (IS) for LR pairs between cell clusters, we assess their statistical significance using a random permutation test. Specifically, we randomly permute cell group labels and recalculate LR interaction scores, repeating this until each LR pair accumulates approximately 10,000 random scores. We then estimate the parameters (alpha and scale) of a gamma distribution from these scores. The significance ( $P$ -value) of each observed IS is computed based on its position in the right tail of this distribution, followed by FDR (Benjamini–Hochberg) correction. Interaction scores with adjusted  $P$ -values below 0.05 are deemed significant (Significant Interaction Scores, SIS). Using the SIS, we can calculate the cell type pair scores between two clusters,  $C1$  (sender) and  $C2$  (receiver):

$$CTPS(C1, C2) = \sum_{(l,r) \in (L,R)} SIS(C1_l, C2_r) \quad (8)$$

Because of the computational expense of calculating the distances between all cells in the data set, the permutation test can be quite costly 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 this method significantly more efficiently (Supplemental Fig. 35B,C).

### 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 perturbation analysis.

We permuted feature values for each receptor, independent of the output and simulated receptor blocking and scaled the expression of the permuted receptor by the median expression of all receptors. This manipulation of the input receptor vector,  $x_j^{Rr}$ , allowed us to gauge the ensuing influence on the gene expression downstream. After each receptor's perturbation, we singled out the most impacted genes, on average, across all cells. Specifically, we looked at the top 50 most impacted genes 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 (Piñero et al. 2015) database as the potential validation. We verified whether the perturbed genes are associated with the specific diseases that piqued our interest—breast cancer/PDAC (i.e., diseases associated with the data set). To quantify the strength of the association between each receptor and the disease, we applied a binomial test to compute the statistical significance of the overlap between the perturbed genes and the genes associated with each disease (breast cancer, PDAC). We further refined our results by adjusting the *P*-values over all disease gene sets, utilizing the FDR-Benjamini–Hochberg method, ensuring the robustness of our findings. Additionally, we confirmed the effectiveness of these receptor-inhibitor therapeutics using Kaplan–Meier survival plots (Györfy 2021).

To validate CellAgentChat's *in silico* receptor blocking functionality, we compared its performance with scGen (Lotfollahi et al. 2019). We trained scGen to reconstruct cell gene expression profiles and reduced each receptor's expression by 50% during inference to replicate CellAgentChat's strategy. The model predicted the resulting downstream gene expression changes, and we analyzed these changes to identify the most affected genes. We then cross-referenced these genes with DisGeNET to evaluate their relevance to the disease of interest, using the same evaluation pipeline as CellAgentChat.

### Agent-based dynamic simulations and animation of CCIs

CellAgentChat offers dynamic simulation capabilities for analyzing the impact of CCIs on cell states over short durations. Users can either execute the model once, similar to existing methods, inferring CCIs at a single moment, or run it across multiple short time steps to simulate temporal dynamics. At each step, CCIs are computed, and a trained neural network predicts gene expression levels for the next step. Although predicted gene expression values cannot directly replace actual data owing to uncertainties, we trust their directional changes—whether genes increase or decrease (Ding et al. 2018). Genes with statistically significant changes are then adjusted slightly ( $\pm 0.01$ ) within the ABM, simulating minor expression variations. This iterative process reflects the dynamic modulation of gene expression by CCIs.

Uniquely, CellAgentChat infers interactions between individual cells, not just cell types. Its real-time animation platform visualizes cell response scores, where size indicates CRS strength—larger cells represent stronger responses (Supplemental Fig. 35A). Users can customize hyperparameters, such as the number of steps (default = 1),  $\delta$  (default = 1), and  $\tau$  (default = 2), to perform effective *in silico* permutations. See Supplemental Methods for more information to effectively set up hyperparameters to facilitate simulations.

In addition, CellAgentChat offers hyperparameters for controlling the LR universe. Users can block a specific LR pair or a receptor to observe downstream gene expression changes. The animation platform also allows for visualizing specific cell types, with modifications only applied to those selected. Furthermore, users have the option to visualize specific receptors or LR pairs.

Further details on standard data set preprocessing and methods related to benchmarking and functional evaluation of inferred ligand–receptor pairs can be found in the Supplemental Methods.

### Data sets

CellTalkDB is available at GitHub (<https://github.com/ZJUfanLab/CellTalkDB>). The SCC and PDAC data sets analyzed in this study

are available from the NCBI Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) repository under the following accession numbers: GSE144240 (GSM4565826) and GSE111672 (GSM3036911), respectively. The data for the SeqFISH+ mouse somatosensory cortex data set can be found at GitHub (<https://github.com/CaiGroup/seqFISH-PLUS>). The Xenium human breast cancer data set (sample 2) can be found at <https://www.10xgenomics.com/products/xenium-in-situ/preview-dataset-human-breast>. The Stereo-seq developmental mouse hippocampus data set can be found at STOmicsDB (Xu et al. 2024) <https://db.cngb.org/stomics/mosta/>. We used the h5ad file titled “Dorsal\_midbrain\_cell\_bin.h5ad”. The PDAC data set with tumor and healthy regions can be found at Zenodo (<https://zenodo.org/records/10712047>).

### Software availability

The CellAgentChat software package and source code are available at GitHub (<https://github.com/mcgillinglab/CellAgentChat>) and as Supplemental Code.

### Competing interest statement

The authors declare no competing interests.

### Acknowledgments

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 and 295299 to J.D.); the Natural Sciences and Engineering Research Council of Canada (NSERC) (RGPIN2022-04399 to J.D.; RGPIN-2016-05174 to Y.L.); and the Meakins-Christie Chair in Respiratory Research (to J.D.). This publication is part of the Human Cell Atlas (HCA) publication bundle (HCA-10).

*Author contributions:* J.D. conceived the study. J.D. and Y.L. supervised the study. V.R. built the framework. V.R. conducted the experiments, with assistance from Y.Z. on one experiment. V.R., J.D., and Y.L. wrote and revised the manuscript. All authors read and approved the final manuscript.

### References

- Ahrends T, Borst J. 2018. The opposing roles of CD4<sup>+</sup> T cells in anti-tumour immunity. *Immunology* **154**: 582–592. doi:10.1111/imm.12941
- Armingol E, Officer A, Harismendy O, Lewis NE. 2021. Deciphering cell–cell interactions and communication from gene expression. *Nat Rev Genet* **22**: 71–88. doi:10.1038/s41576-020-00292-x
- Arnol D, Schapiro D, Bodenmiller B, Saez-Rodriguez J, Stegle O. 2019. Modeling cell–cell interactions from spatial molecular data with spatial variance component analysis. *Cell Rep* **29**: 202–211.e6. doi:10.1016/j.celrep.2019.08.077
- Asp M, Bergenstråhle J, Lundeberg J. 2020. Spatially resolved transcriptomes—next generation tools for tissue exploration. *Bioessays* **42**: 1900221. doi:10.1002/bies.201900221
- Atasoy M, Anadol-Braise R, Pirim I, Dogan H, Ikbilal M. 2009. HLA antigen profile differences in patients with SCC (squamous cell carcinoma) *in situ*/actinic keratosis and invasive SCC: is there a genetic susceptibility for invasive SCC development? *Eurasian J Med* **41**: 162.
- Baruzzo G, Cesaro G, Di Camillo B. 2022. Identify, quantify and characterize cellular communication from single-cell RNA sequencing data with *scSeqComm*. *Bioinformatics* **38**: 1920–1929. doi:10.1093/bioinformatics/btac036
- Bonabeau E. 2002. Agent-based modeling: methods and techniques for simulating human systems. *Proc Natl Acad Sci* **99**: 7280–7287. doi:10.1073/pnas.082080899
- Bonnans C, Chou J, Werb Z. 2014. Remodelling the extracellular matrix in development and disease. *Nat Rev Mol Cell Biol* **15**: 786–801. doi:10.1038/nrm3904

- Boschert V, Teusch J, Müller-Richter UD, Brands RC, Hartmann S. 2022. PKM2 modulation in head and neck squamous cell carcinoma. *Int J Mol Sci* **23**: 775. doi:10.3390/ijms23020775
- Bovolenta L, Acencio M, Lemke N. 2012. HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genomics* **13**: 405. doi:10.1186/1471-2164-13-405
- Cabello-Aguilar S, Alame M, Kon-Sun-Tack F, Fau C, Lacroix M, Colinge J. 2020. Singlecellsignalr: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res* **48**: e55–e55. doi:10.1093/nar/gkaa183
- Cang Z, Zhao Y, Almet AA, Stabel A, Ramos R, Plikus MV, Atwood SX, Nie Q. 2023. Screening cell–cell communication in spatial transcriptomics via collective optimal transport. *Nat Methods* **20**: 218–228. doi:10.1038/s41592-022-01728-4
- Centuori SM, Bauman JE. 2022. c-Met signaling as a therapeutic target in head and neck cancer. *Cancer J* **28**: 346–353. doi:10.1097/PP0.0000000000000619
- Chen A, Liao S, Cheng M, Ma K, Wu L, Lai Y, Qiu X, Yang J, Xu J, Hao S, et al. 2022. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**: 1777–1792.e21. doi:10.1016/j.cell.2022.04.003
- Cope EC, Gould E. 2019. Adult neurogenesis, glia, and the extracellular matrix. *Cell Stem Cell* **24**: 690–705. doi:10.1016/j.stem.2019.03.023
- Corso G, Veronesi P, Sacchini V, Galimberti V. 2018. Prognosis and outcome in CDH1-mutant lobular breast cancer. *Eur J Cancer Prev* **27**: 237–238. doi:10.1097/CEJ.0000000000000405
- Dimitrov D, Schäfer PSL, Farr E, Rodriguez-Mier P, Lobentanzer S, Badia-i-Mompel P, Dugourd A, Tanevski J, Ramirez Flores RO, Saez-Rodriguez J. 2024. LIANA+ provides an all-in-one framework for cell–cell communication inference. *Nat Cell Biol* **26**: 1613–1622. doi:10.1038/s41556-024-01469-w
- Ding J, Hagood JS, Ambalavanan N, Kaminski N, Bar-Joseph Z. 2018. iDREM: interactive visualization of dynamic regulatory networks. *PLoS Comput Biol* **14**: e1006019. doi:10.1371/journal.pcbi.1006019
- Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. 2020. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat Protoc* **15**: 1484–1506. doi:10.1038/s41596-020-0292-x
- Eng C-HL, Lawson M, Zhu Q, Dries R, Kouloua N, Takei Y, Yun J, Cronin C, Karp C, Yuan G-C, et al. 2019. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**: 235–239. doi:10.1038/s41586-019-1049-y
- Feng J, Goedegebuure SP, Zeng A, Bi Y, Wang T, Payne P, Ding L, DeNardo D, Hawkins W, Fields RC, et al. 2024. Sc2menetdrug: a computational tool to uncover inter-cell signaling targets and identify relevant drugs based on single cell RNA-seq data. *PLoS Comput Biol* **20**: e1011785. doi:10.1371/journal.pcbi.1011785
- Friedl P, Alexander S. 2011. Cancer invasion and the microenvironment: plasticity and reciprocity. *Cell* **147**: 992–1009. doi:10.1016/j.cell.2011.11.016
- Fromme JE, Zigrino P. 2022. The role of extracellular matrix remodeling in skin tumor progression and therapeutic resistance. *Front Mol Biosci* **9**: 864302. doi:10.3389/fmolb.2022.864302
- Ghahremani Dehbokri S, Alizadeh N, Isazadeh A, Baghbanzadeh A, Abbaspour-Ravassani S, Hajiasgarzadeh K, Baradaran B. 2023. CTLA-4: as an immunosuppressive immune checkpoint in breast cancer. *Curr Mol Med* **23**: 521–526. doi:10.2174/1566524022666220610094716
- Giladi A, Cohen M, Medaglia C, Baran Y, Li B, Zada M, Bost P, Blecher-Gonen R, Salame T-M, Mayer JU, et al. 2020. Dissecting cellular crosstalk by sequencing physically interacting cells. *Nat Biotechnol* **38**: 629–637. doi:10.1038/s41587-020-0442-2
- Györfy B. 2021. Survival analysis across the entire transcriptome identifies biomarkers with the highest prognostic power in breast cancer. *Comput Struct Biotechnol J* **19**: 4101–4109. doi:10.1016/j.csbj.2021.07.014
- Han H, Cho J-W, Lee S, Yun A, Kim H, Bae D, Yang S, Kim CY, Lee M, Kim E, et al. 2018. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res* **46**: D380–D386. doi:10.1093/nar/gkx1013
- Hou R, Denisenko E, Ong HT, Ramiłowski JA, Forrest AR. 2020. Predicting cell-to-cell communication networks using NATMI. *Nat Commun* **11**: 5011. doi:10.1038/s41467-020-18873-z
- Hu M, Polyak K. 2008. Microenvironmental regulation of cancer development. *Curr Opin Genet Dev* **18**: 27–34. doi:10.1016/j.gde.2007.12.006
- Janesick A, Shelansky R, Gottscho AD, Wagner F, Williams SR, Rouault M, Beliakoff G, Morrison CA, Oliveira MF, Sicherman JT, et al. 2023. High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nat Commun* **14**: 8353. doi:10.1038/s41467-023-43458-x
- Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, Sidiroopoulos K, Cook J, Gillespie M, Haw R, et al. 2020. The reactome pathway knowledgebase. *Nucleic Acids Res* **48**: D498–D503. doi:10.1093/nar/gkz1031
- Ji AL, Rubin AJ, Thrane K, Jiang S, Reynolds DL, Meyers RM, Guo MG, George BM, Mollbrink A, Bergensträhle J. 2020. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell* **182**: 497–514.e22. doi:10.1016/j.cell.2020.05.039
- Jin S, Guerrero-Juarez CF, Zhang L, Chang J, Ramos R, Kuan C-H, Myung P, Plikus MV, Nie Q. 2021. Inference and analysis of cell–cell communication using CellChat. *Nat Commun* **12**: 1088. doi:10.1038/s41467-021-21246-9
- Jin S, Plikus MV, Nie Q. 2025. Cellchat for systematic analysis of cell–cell communication from single-cell transcriptomics. *Nat Protoc* **20**: 180–219. doi:10.1038/s41596-024-01045-4
- Kakoi C, Udo H, Matsukawa T, Ohnuki K. 2012. Collagen peptides enhance hippocampal neurogenesis and reduce anxiety-related behavior in mice. *Biomed Res* **33**: 273–279. doi:10.2220/biomedres.33.273
- Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **28**: 27–30. doi:10.1093/nar/28.1.27
- Li W, You L, Schaffler MB, Wang L. 2009. The dependency of solute diffusion on molecular weight and shape in intact bone. *Bone* **45**: 1017–1023. doi:10.1016/j.bone.2009.07.076
- Li W, Mandeville ET, Durán-Laforet V, Fukuda N, Yu Z, Zheng Y, Held A, Park J-H, Nakano T, Tanaka M, et al. 2022. Endothelial cells regulate astrocyte to neural progenitor cell *trans*-differentiation in a mouse model of stroke. *Nat Commun* **13**: 7812. doi:10.1038/s41467-022-35498-6
- Lin C, Jain S, Kim H, Bar-Joseph Z. 2017. Using neural networks for reducing the dimensions of single-cell RNA-seq data. *Nucleic Acids Res* **45**: e156. doi:10.1093/nar/gkx681
- Liu Z, Sun D, Wang C. 2022. Evaluation of cell–cell interaction methods by integrating single-cell RNA sequencing data with spatial information. *Genome Biol* **23**: 218. doi:10.1186/s13059-022-02783-y
- Lotfollahi M, Wolf FA, Theis FJ. 2019. ScGen predicts single-cell perturbation responses. *Nat Methods* **16**: 715–721. doi:10.1038/s41592-019-0494-8
- Ludwig N, Szczepanski MJ, Glusko A, Szafarowski T, Azambuja JH, Dolg L, Gellrich N-C, Kampmann A, Whiteside TL, Zimmerer RM. 2019. CD44 (+) tumor cells promote early angiogenesis in head and neck squamous cell carcinoma. *Cancer Lett* **467**: 85–95. doi:10.1016/j.canlet.2019.10.010
- Marx V. 2021. Method of the year: spatially resolved transcriptomics. *Nat Methods* **18**: 9–14. doi:10.1038/s41592-020-01033-y
- Masad D, Kazil J. 2015. MESA: an agent-based modeling framework. In *Proceedings of the 14th PYTHON in Science Conference*, Vol. 2015, pp. 53–60. Citeseer. Austin, TX.
- Masuda H, Zhang D, Bartholomeusz C, Doihara H, Hortobagyi GN, Ueno NT. 2012. Role of epidermal growth factor receptor in breast cancer. *Breast Cancer Res Treat* **136**: 331–345. doi:10.1007/s10549-012-2289-9
- Meireles Da Costa N, Mendes FA, Pontes B, Nasciutti LE, Ribeiro Pinto LF, Palumbo Júnior A. 2021. Potential therapeutic significance of laminin in head and neck squamous carcinomas. *Cancers (Basel)* **13**: 1890. doi:10.3390/cancers13081890
- Mi X, Zou B, Zou F, Hu J. 2021. Permutation-based identification of important biomarkers for complex diseases via machine learning models. *Nat Commun* **12**: 3008. doi:10.1038/s41467-021-22756-2
- Moncada R, Barkley D, Wagner F, Chiodin M, Devlin JC, Baron M, Hajdu CH, Simeone DM, Yanai I. 2020. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat Biotechnol* **38**: 333–342. doi:10.1038/s41587-019-0392-8
- Neufeld G, Mumbat Y, Smolkin T, Toledano S, Nir-Zvi I, Ziv K, Kessler O. 2016. The role of the semaphorins in cancer. *Cell Adh Migr* **10**: 652–674. doi:10.1080/19336918.2016.1197478
- Parker AL, Bowman E, Zingone A, Ryan BM, Cooper WA, Kohonen-Corish M, Harris CC, Cox TR. 2022. Extracellular matrix profiles determine risk and prognosis of the squamous cell carcinoma subtype of non-small cell lung carcinoma. *Genome Med* **14**: 126. doi:10.1186/s13073-022-01127-6
- Peng L, Yuan R, Han C, Han G, Tan J, Wang Z, Chen M, Chen X. 2023. CellEnBoost: a boosting-based ligand–receptor interaction identification model for cell-to-cell communication inference. *IEEE Trans Nanobioscience* **22**: 705–715. doi:10.1109/TNB.2023.3278685
- Peng L, Xiong W, Han C, Li Z, Chen X. 2024. CellDialog: a computational framework for ligand–receptor-mediated cell–cell communication analysis. *IEEE J Biomed Health Inform* **28**: 580–591. doi:10.1109/JBHI.2023.3333828
- Perez VM, Kearney JF, Yeh JJ. 2021. The PDAC extracellular matrix: a review of the ECM protein composition, tumor cell interaction, and therapeutic strategies. *Front Oncol* **11**: 751311. doi:10.3389/fonc.2021.751311
- Piñero J, Queralt-Rosinach N, Bravo A, Deu-Pons J, Bauer-Mehren A, Baron M, Sanz F, Furlong LI. 2015. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database* **2015**: bav028. doi:10.1093/database/bav028

- Pontén F, Jirstrom K, Uhlen M. 2008. The Human Protein Atlas—a tool for pathology. *J Pathol* **216**: 387–393. doi:10.1002/path.2440
- Ralsler DJ, Klumper N, Gevensleben H, Zarbl R, Kaiser C, Landsberg J, Hölzel M, Strieth S, Faridi A, Abramian A, et al. 2021. Molecular and immune correlates of PDCD1 (PD-1), PD-L1 (CD274), and PD-L2 (PDCD1LG2) DNA methylation in triple negative breast cancer. *J Immunother* **44**: 319–324. doi:10.1097/CJI.0000000000000384
- Ramovs V, Te Molder L, Sonnenberg A. 2017. The opposing roles of laminin-binding integrins in cancer. *Matrix Biol* **57–58**: 213–243. doi:10.1016/j.matbio.2016.08.007
- Rao A, Barkley D, França GS, Yanai I. 2021. Exploring tissue architecture using spatial transcriptomics. *Nature* **596**: 211–220. doi:10.1038/s41586-021-03634-9
- Raredon MSB, Yang J, Garritano J, Wang M, Kushnir D, Schupp JC, Adams TS, Greaney AM, Leiby KL, Kaminski N, et al. 2022. Computation and visualization of cell–cell signaling topologies in single-cell systems data using *Connectome*. *Sci Rep* **12**: 4187. doi:10.1038/s41598-022-07959-x
- Raredon MSB, Yang J, Kothapalli N, Lewis W, Kaminski N, Niklason LE, Kluger Y. 2023. Comprehensive visualization of cell–cell interactions in single-cell and spatial transcriptomics with NICHES. *Bioinformatics* **39**: btac775. doi:10.1093/bioinformatics/btac775
- Ren J, Chen Y, Kong W, Li Y, Lu F. 2021. Tumor protein D52 promotes breast cancer proliferation and migration via the long non-coding RNA NEAT1/microRNA-218-5p axis. *Ann Transl Med* **9**: 1008. doi:10.21037/atm-21-2668
- Rodrigues SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, Welch J, Chen LM, Chen F, Macosko EZ. 2019. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**: 1463–1467. doi:10.1126/science.aaw1219
- Rouault H, Hakim V. 2012. Different cell fates from cell–cell interactions: core architectures of two-cell bistable networks. *Biophys J* **102**: 417–426. doi:10.1016/j.bpj.2011.11.4022
- Schaafsma E, Fugle CM, Wang X, Cheng C. 2021. Pan-cancer association of HLA gene expression with cancer prognosis and immunotherapy efficacy. *Br J Cancer* **125**: 422–432. doi:10.1038/s41416-021-01400-2
- Shao X, Lu X, Liao J, Chen H, Fan X. 2020. New avenues for systematically inferring cell–cell communication: through single-cell transcriptomics data. *Protein Cell* **11**: 866–880. doi:10.1007/s13238-020-00727-5
- Shao X, Liao J, Li C, Lu X, Cheng J, Fan X. 2021. CellTalkDB: a manually curated database of ligand–receptor interactions in humans and mice. *Brief Bioinformatics* **22**: bbaa269. doi:10.1093/bib/bbaa269
- Soysal SD, Muenst S, Barbie T, Fleming T, Gao F, Spizzo G, Oertli D, Viehl CT, Obermann EC, Gillanders WE. 2013. EpCAM expression varies significantly and is differentially associated with prognosis in the luminal B HER2+, basal-like, and HER2 intrinsic subtypes of breast cancer. *Br J Cancer* **108**: 1480–1487. doi:10.1038/bjc.2013.80
- Stubbington MJ, Rozenblatt-Rosen O, Regev A, Teichmann SA. 2017. Single-cell transcriptomics to explore the immune system in health and disease. *Science* **358**: 58–63. doi:10.1126/science.aan6828
- Swonger JM, Liu JS, Ivey MJ, Tallquist MD. 2016. Genetic tools for identifying and manipulating fibroblasts in the mouse. *Differentiation* **92**: 66–83. doi:10.1016/j.diff.2016.05.009
- Szturz P, Raymond E, Abitbol C, Albert S, De Gramont A, Faivre S. 2017. Understanding c-MET signalling in squamous cell carcinoma of the head & neck. *Crit Rev Oncol Hematol* **111**: 39–51. doi:10.1016/j.critrevonc.2017.01.004
- Tensaouti Y, Stephanz EP, Yu T-S, Kernie SG. 2018. Apoe regulates the development of adult newborn hippocampal neurons. *eNeuro* **5**: ENEURO.0155-18.2018. doi:10.1523/ENEURO.0155-18.2018
- Tong CK, Vidyadaran S. 2016. Role of microglia in embryonic neurogenesis. *Exp Biol Med* **241**: 1669–1675. doi:10.1177/1535370216664430
- Tran M, Yoon S, Teoh M, Andersen S, Lam P, Purdue B, Raghuram A, Hanson S, Devitt K, Jones K, et al. 2022. A robust experimental and computational analysis framework at multiple resolutions, modalities and coverages. *Front Immunol* **13**: 911873. doi:10.3389/fimmu.2022.911873
- Van Dijk D, Sharma R, Nainys J, Yin K, Kathail P, Carr AJ, Burdzyak C, Moon KR, Chaffer CL, Pattabiraman D. 2018. Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**: 716–729.e27. doi:10.1016/j.cell.2018.05.061
- Vento-Tormo R, Efremova M, Botting RA, Turco MY, Vento-Tormo M, Meyer KB, Park J-E, Stephenson E, Polański K, Goncalves A, et al. 2018. Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature* **563**: 347–353. doi:10.1038/s41586-018-0698-6
- Vogenstahl J, Parrilla M, Acker-Palmer A, Segarra M. 2022. Vascular regulation of developmental neurogenesis. *Front Cell Dev Biol* **10**: 890852. doi:10.3389/fcell.2022.890852
- Wan X, Xiao J, Tam SST, Cai M, Sugimura R, Wang Y, Wan X, Lin Z, Wu AR, Yang C. 2023. Integrating spatial and single-cell transcriptomics data using deep generative models with SpatialScope. *Nat Commun* **14**: 7848. doi:10.1038/s41467-023-43629-w
- Wang J, Fonseca GJ, Ding J. 2024. scSemiProfiler: advancing large-scale single-cell studies through semi-profiling with deep generative models and active learning. *Nat Commun* **15**: 5989. doi:10.1038/s41467-024-50150-1
- Wilk AJ, Shalek AK, Holmes S, Blish CA. 2024. Comparative analysis of cell–cell communication at single-cell resolution. *Nat Biotechnol* **42**: 470–483. doi:10.1038/s41587-023-01782-z
- Xia C, Babcock HP, Moffitt JR, Zhuang X. 2019. Multiplexed detection of RNA using MERFISH and branched DNA amplification. *Sci Rep* **9**: 7721. doi:10.1038/s41598-019-43943-8
- Xu Z, Wang W, Yang T, Li L, Ma X, Chen J, Wang J, Huang Y, Gould J, Lu H, et al. 2024. STOmicsDB: a comprehensive database for spatial transcriptomics data sharing, analysis and visualization. *Nucleic Acids Res* **52**: D1053–D1061. doi:10.1093/nar/gkad933
- Yin Z, Rosenzweig N, Kleemann KL, Zhang X, Brandão W, Margeta MA, Schroeder C, Sivanathan KN, Silveira S, Gauthier C, et al. 2023. APOE4 impairs the microglial response in Alzheimer's disease by inducing TGFβ-mediated checkpoints. *Nat Immunol* **24**: 1839–1853. doi:10.1038/s41590-023-01627-6
- Yu JS, Bagheri N. 2020. Agent-based models predict emergent behavior of heterogeneous cell populations in dynamic microenvironments. *Front Bioeng Biotechnol* **8**: 249. doi:10.3389/fbioe.2020.00249
- Yuan G-C, Cai L, Elowitz M, Enver T, Fan G, Guo G, Irizarry R, Kharchenko P, Kim J, Orkin S. 2017. Challenges and emerging directions in single-cell analysis. *Genome Biol* **18**: 84. doi:10.1186/s13059-017-1218-y
- Zhang G, Li T, Tan G, Song Y, Liu Q, Wang K, Ai J, Zhou Z, Li W. 2022. Identity of MMP1 and its effects on tumor progression in head and neck squamous cell carcinoma. *Cancer Med* **11**: 2516–2530. doi:10.1002/cam4.4623
- Zhou X, Franklin RA, Adler M, Jacox JB, Bailis W, Shyer JA, Flavell RA, Mayo A, Alon U, Medzhitov R. 2018. Circuit design features of a stable two-cell system. *Cell* **172**: 744–757.e17. doi:10.1016/j.cell.2018.01.015

Received July 9, 2024; accepted in revised form April 23, 2025.