



## QuadST identifies cell-cell interaction-changed genes in spatially resolved transcriptomics data

Xiaoyu Song, Yuqing Shang, Michelle Ehrlich, et al.

*Genome Res.* published online June 23, 2025

Access the most recent version at doi:[10.1101/gr.279859.124](https://doi.org/10.1101/gr.279859.124)

---

|                                 |                                                                                                                                                                                                                                                                                                    |
|---------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>P&lt;P</b>                   | Published online June 23, 2025 in advance of the print journal.                                                                                                                                                                                                                                    |
| <b>Accepted Manuscript</b>      | Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.                                                                                                                                               |
| <b>Open Access</b>              | Freely available online through the <i>Genome Research</i> Open Access option.                                                                                                                                                                                                                     |
| <b>Creative Commons License</b> | This manuscript is Open Access. This article, published in <i>Genome Research</i> , is available under a Creative Commons License (Attribution 4.0 International license), as described at <a href="http://creativecommons.org/licenses/by/4.0/">http://creativecommons.org/licenses/by/4.0/</a> . |
| <b>Email Alerting Service</b>   | Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or <a href="#">click here</a> .                                                                                                                                          |

---

**Comprehensive** immune receptor profiling.  
Discover the **DriverMap™ AIR Assay** difference.



---

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

---

Published by Cold Spring Harbor Laboratory Press

# QuadST identifies cell–cell interaction-changed genes in spatially resolved transcriptomics data

Xiaoyu Song <sup>1,\*</sup>, Yuqing Shang<sup>1</sup>, Michelle E. Ehrlich <sup>2</sup>, Panos Roussos <sup>3</sup>, Guo-Cheng Yuan <sup>4</sup>, Pei Wang <sup>5</sup>

[1] Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore

[2] Departments of Neurology, Pediatrics, and Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[3] Center for Disease Neurogenetics, Department of Psychiatry, Department of Genetics and Genomics Sciences, Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, NY, USA; Mental Illness Research Education, and Clinical Center (VISN 2 South), James J. Peters VA Medical Center, Bronx, NY, USA

[4] Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[5] Department of Genetics and Genomic Sciences, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

\*Correspondence: [song.xiaoyu@duke-nus.edu.sg](mailto:song.xiaoyu@duke-nus.edu.sg)

17

## Abstract

18 Recent advances in spatially resolved transcriptomics (SRT) have provided valuable  
19 avenues for identifying cell–cell interactions and their critical roles in diseases. We  
20 introduce QuadST, a novel statistical method for the robust and powerful identification  
21 of cell–cell interactions and their impacted genes in single-cell SRT. QuadST models  
22 interactions at different cell–cell distance quantile levels and innovatively contrasts  
23 signals to identify interaction-changed genes, which exhibit stronger signals at shorter  
24 distances. Unlike other methods, QuadST does not require the specification of  
25 interacting cell pairs. It is also robust against unmeasured confounding factors and  
26 measurement errors of the data. Simulation studies demonstrate that QuadST  
27 effectively controls the type I error, even in misspecified settings, and significantly  
28 improves power over existing methods. Applications of QuadST to real datasets have  
29 successfully revealed biologically significant interaction-changed genes across various  
30 cell types.

31 **Key words:** cell–cell interaction, spatially resolved transcriptomics, interaction-changed  
32 gene

33 **Running title:** QuadST identifies cell interaction in spatial data

34

## Introduction

36 Cells rarely function in isolation; instead, they interact with neighboring cells to carry out  
37 diverse biological functions. These cell–cell interactions are integral to numerous

biological processes (Baszkin and Norde 1999), such as development and immune response. Malfunctions in the cell–cell interactions underlie many complex human diseases, such as cancer (Maffuid and Cao 2023) and neurodegenerative diseases (Szepesi et al. 2018). Spatially resolved transcriptomics (SRT) provides an invaluable avenue for understanding the cell–cell interactions, by allowing researchers to study the gene expression profiles of cells within the native tissue context (Burgess 2019).

Several statistical methods (Arnol et al. 2019; Cang and Nie 2020; Li et al. 2023; Pham et al. 2023; Ma et al. 2024; Dries et al. 2019; Fischer et al. 2023) have been developed for SRT data to identify cell–cell interactions and interaction-changed genes (ICGs), which are the genes whose expression levels are influenced by the spatial proximity of other cells. Among these methods, SVCA (Arnol et al. 2019), SpaOTsc (Cang and Nie 2020), SpatialDM (Li et al. 2023), and stLearn (Pham et al. 2023) analyze interactions by summarizing all cells collectively, regardless of their specific cell types, providing limited insights into cell-type specific interactions. Additionally, SVCA only estimates the proportion of total variations explained by cell–cell interactions without identifying specific genes. SpaOTsc, SpatialDM, and stLearn focus on ligand-receptor pairs, overlooking other interaction mechanisms (e.g., gap junctions) and failing to capture genes involved in the upstream and downstream pathways of the cell–cell interactions.

Meanwhile, RECCIPE (Ma et al. 2024), Giotto (Dries et al. 2021), and NCEM (Fischer et al. 2023) aim at screening the gene expression profiles to identify those involved in cell-type-specific cell–cell interactions. RECCIPE is designed for low resolution multi-cell SRT data, and Giotto and NCEM are for high resolution single-cell SRT data. Unfortunately, Giotto and NCEM require categorizing cells into neighboring

interacting pairs versus distant non-interacting pairs, a process prone to errors due to the limited knowledge on the cell-type-specific interacting distances. Additionally, current SRT technologies struggle to quantify cell distances accurately, as they cannot accurately capture cell boundaries. These methods often rely on the biased cell center-to-center distance for analysis but lack features to minimize the impact of measurement errors on their identifications. Finally, factors influencing cell–cell interactions in their local microenvironment, such as extracellular elements from a third cell type, are complex and largely unexplored. These factors can confound the analysis, causing inflated false discoveries and reduced power, while existing methods cannot address the unmeasured confounders.

We propose QuadST, a highly robust and powerful approach for identifying cell–cell interactions and ICGs on single-cell SRT data. For each cell-type pair, QuadST directly models the measured cell–cell distance as a continuous outcome on gene expression, avoiding the need for cell pair categorization and safeguarding against measurement errors. By innovatively contrasting the associations at different distance quantile levels, QuadST captures enhanced signals from closer cells, providing a strategy for powerful and robust identification of ICGs, even in the presence of inaccurately measured cell–cell distances, unmeasured confounding factors, and gene-gene correlations.

## Results

[Insert Figure 1 here]

Figure 1: **Schematic overview of QuadST.** Taking the spatial and transcriptomic data of single-cell SRT as input, QuadST creates an integrated matrix for each anchor–neighbor cell-type pair,

models the association between expression and a weighted anchor–neighbor cell–cell distance at different distance quantile levels, and contrasts signal levels at different quantile levels to identify cell–cell interactions and their involved genes with well-controlled FDR.

**QuadST framework.** QuadST has three key steps (Fig. 1 and Methods). First, it constructs an integrated matrix for each “anchor” and “neighbor” cell-type pair of analysis. The “anchor” and “neighbor” cells can come either from the same cell type or two different cell types, making a total  $K^2$  cell type pairs for  $K$  cell types of interest. The integrated matrix includes the expression profiles and cell-level covariates for all “anchor” cells, along with their distances to their  $k$  nearest cells in the “neighbor” cell type. Second, QuadST models the association between gene expression and a weighted average of  $k$  anchor–neighbor cell–cell distances, in parallel for all genes, at a grid of evenly distributed quantile levels of the distance, using quantile regression. By modeling distance as the outcome in quantile regression, we can observe its local associations with gene expression levels, facilitating the comparison how interactions differ when two cells are close versus far away from each other. Third, QuadST contrasts the signal levels at different distance quantile levels to identify ICGs with false discovery rate (FDR) control. This contrast calculates the ratio of the number of genes with small p-values at symmetric upper versus lower quantile levels, such as at above 90<sup>th</sup> versus below 10<sup>th</sup> quantile levels, which provides an upper bound for empirical FDR (eFDR), even when p-values are ill-behaved, such as arising from regressions with unmeasured confounders and gene-gene correlations.

[insert Figure 2 here]

Figure 2: Evaluation of FDR and power for QuadST across different parameters and datasets, and comparison of its performance with alternative methods in various settings. A. FDR and power of QuadST using different number of quantile levels. B. FDR and power of QuadST using different number of  $k$  nearest neighbors (KNN) at  $k = 1, \dots, 5$ . C. FDR and power of QuadST in data with different cell numbers per cell type. D. Comparison of different methods in an ideal case where measurement error and confounding factor do not exist and interacting cell pairs needed for Giotto and NCEM are perfectly known. E. Comparison of different methods where interacting cell pairs needed for Giotto and NCEM cannot be perfectly identified. F. Comparison of different methods under the existence of measurement error for cell–cell distance. G. Comparison of different methods under the existence of unobtained confounding factor.

**Simulation studies.** We performed extensive simulations to understand the performance of QuadST under varying parameters (e.g. quantile levels,  $k$ -nearest neighbors) and datasets (e.g. sample sizes) (Method). First, we evaluated the impact of numbers of quantile levels used in QuadST on its performance (Fig. 2A). We observed that type I error rates were consistently well-controlled, even if extreme quantile levels were used, such as using 999 quantile levels at 0.001, 0.002, ..., 0.999 for a total of 1000 anchor cells. The study powers initially increased with the number of quantile levels but plateaued after approximately 10 quantile levels. This is mostly due to highly correlated results at additional quantile levels that contribute little to the ICG identification. These results suggest that the specification of quantile levels in QuadST need not consider error control, but rather balance between study power and computation cost at a given sample size of the data. Next, we evaluated the impact of  $k$ -

128 nearest neighbors used to construct anchor–neighbor matrix in QuadST on its  
129 performance, and observed consistently well-controlled type I error rate and slight  
130 power decrease as the number of nearest neighbors increased from 1 to 5 (Fig. 2B).  
131 Finally, to understand the performance of QuadST on small samples, we ranged sample  
132 size from 100 to 1000 and observed well-controlled type I error and steadily increased  
133 study power with sample size (Fig. 2C).

134 Additionally, we performed extensive simulations to compare the performance of  
135 QuadST with alternative approaches. First, we evaluated all methods in an ideal case,  
136 where cell–cell distance was perfectly measured, no unmeasured confounding factors  
137 existed in the data, and interacting vs. noninteracting cell pairs could be perfectly known  
138 for use in Giotto and NCEM. All methods had well-controlled type I error, and QuadST  
139 and NCEM reached similar highest study power at 60.5% and 61.9%, followed by Giotto  
140 at 51.7% and Linear Regression (LR) at 15% (Fig. 2D). Secondly, we evaluated  
141 methods when interacting vs. noninteracting cell pairs could not be perfectly identified.  
142 Giotto and NCEM were impacted with powers reduced to 28.1% and 36.5%, while  
143 QuadST not requiring this information maintained the high power at 60.5% (Fig. 2E).  
144 Thirdly, we evaluated methods under the existence measurement errors, which biased  
145 the observed cell–cell distance. While all methods had well controlled type I error rate,  
146 QuadST had much higher study power at 50.9% than other methods at 37.2% for  
147 NCEM, 28.4% for Giotto, and 14.4% for LR (Fig. 2F). Finally, we evaluated these  
148 methods under existence of unmeasured confounding factors. QuadST was the only  
149 method that can control type I error rate, albeit conservatively at 2% (Fig. 2G), while all  
150 others have inflations to 16% in NCEM, 17.2% in Giotto, and 88.3% in LR. Although

QuadST was overconservative, it ensured the validity of its significant identifications. Conversely, all compared methods exhibited largely inflated type I error rates, jeopardizing their validities.

[Insert Figure 3 here]

Figure 3: A publicly available SeqFISH+ dataset from mouse cortex. A. Spatial map by annotated cell types in the SeqFISH+ data. B. UMAP of the highly variable genes in the data. C. Partial R-squared explained by cell type and FOV for UMAP Dimension 1 and 2.

**Cell–cell interaction studies in mouse cortex.** We first applied QuadST to study the cell–cell interaction in a mouse cortex dataset profiled by the seqFISH+ technology (Eng et al. 2019) (Methods). The data included expression levels of 10,000 genes from 511 cells of six cell types (Fig. 3A), imaged from 5 field of views (FOVs), from a 23-day old male mouse (C57BL/6J) cortex. The uniform manifold approximation and projection (UMAP) plot showed that cells clustered not only by cell types but also by FOVs (Fig.3B), which was confirmed by Partial  $R^2$  analysis (Fig.3C), instructing us to adjust FOVs as covariates in the analysis.

[Insert Figure 4 here]

Figure 4: Application of QuadST to a SeqFISH+ dataset from mouse cortex. A. Number of significant ICGs by anchor–neighbor cell-type pairs. B. Functional categories of significant ICGs. C. GO cellular component enrichment of ICGs in interactions among excitatory neurons. D. Directional association scores of ICGs and all other genes among all anchor–neighbor cell pairs.

E. Boxplot of directional association scores of genes in and out of the glutamatergic synapse gene set. F. Distance-expression profiles of the top two ICGs.

We ran QuadST for a total of 36 anchor–neighbor cell-type pairs from six cell types on 1,235–2,500 highly variable genes and identified 626 ICGs across 19 out of 36 pairs at 10% FDR (Fig. 4A; Table S1). Out of these significant ICGs, only 5.1% were ligands and receptors and 4.4% were transcription factors and cofactors (Fig. 4B), indicating that ICGs could be genes with different functions directly or indirectly involved in cell–cell interactions. The most frequent interactions were observed among excitatory neurons, with a total of 326 significant ICGs. These ICGs were primarily involved in synapse-related functions, particularly glutamatergic synapses, as revealed by the Gene Ontology (GO) cellular component enrichment analysis (Fig. 4C). A vast majority of these ICGs (97%) displayed positively associations with cell–cell distance (Fig. 4D), further suggesting that they play crucial roles in establishing connections with distant excitatory neurons via axons and dendrites through synapses. The top two ICGs were *Cplx1* and *Nsmf* (Fig. 3E–F). *Cplx1* regulates synaptic vesicle fusion (Chang et al. 2015), and *Nsmf* facilitates the transmission of signals from both synaptic and extrasynaptic NMDA receptors to the nucleus (Gomes et al. 2023). The discoveries of these critical genes and their association patterns demonstrated the effectiveness of our method in revealing intricate details of cell–cell interactions.

We next applied QuadST to a different mouse cortex dataset (Allen et al. 2023) profiled by the MERFISH technology (Methods). This dataset included 374 pre-selected genes in 13,745 cells of eight cell types from the cortical layers II–V of a 4-week-old female mouse (C57BL/6J) (Fig. 5A), and QuadST was applied to 280 highly variable genes of

the cells. Cortical layers contributed slightly to variations of gene expression as demonstrated by UMAP (Fig. 5B) and partial  $R^2$  (Fig. 5C) and was included as covariates in QuadST analysis. We ran QuadST for a total 64 anchor–neighbor cell-type pairs for the eight cell types and identified 132 significant ICGs for 36 out of 64 cell-type pairs (Fig. 5D; Table S2). Consistent with the SeqFISH+ results, interactions among excitatory neurons were most frequently identified (Fig. 5E), with all the 19 significant ICGs showing upregulated expression with cell–cell distance (Fig. 5E).

[Insert Figure 5 here]

Figure 5: Application of QuadST to a MERFISH dataset from mouse cortex. A. Spatial map by annotated cell types in the MERFISH data. B. UMAP by cell type and cortical layer. C. Partial R-squared by cell type and cortical layer. D. Number of significant ICGs by anchor–neighbor cell-type pairs. E. Directional association scores of ICGs and all other genes among all anchor–neighbor cell pairs.

In addition to detect cell–cell interactions and their ICGs, QuadST also provides an estimation of interaction distance specific to each cell-type pair (Fig. S1). In these two datasets, the distances varied between 10 and 70  $\mu m$  for different cell types. NCEM reported distances on the same scale but did not demonstrate variations across cell types due to their analyses of all cells collectively.

Finally, to understand how QuadST results differ from those produced by other methods, we also applied NCEM, Giotto, and LR to the SeqFISH+ (Supplementary Table 3) and MERFISH (Supplementary Table 4) data. A challenging step for both NCEM and Giotto is to group cells into neighboring pairs. NCEM suggests a cutoff

around 69um on cell–cell physical distance for grouping. Using this cutoff, NCEM identified 2450 out of 2500 highly-variable genes as ICGs with interactions among excitatory neurons in SeqFISH+, suggesting a potential high proportion of false discoveries. We further performed sensitivity analysis by perturbing cell–cell distance cutoffs (59um, 69um, 79um), and found that it was highly sensitive to the cell pair groups. For instance, NCEM detected 155 ICGs for microglia – astrocyte interactions in SeqFISH+ at 59um but no ICGs at 69um. While distance-based cutoff is not robust, Giotto offers an alternative network-based algorithm for defining neighborhoods. However, this approach may lack sufficient resolution. For instance, in SeqFISH+ data, it classified all excitatory neurons into each other’s neighborhoods, failing to detect any ICGs. Finally, when applying the naïve LR to the real data sets, we identified 15 ICGs out of approximately 2,500 highly variable genes across all 36 cell-type pairs in SeqFISH+, and 212 ICGs out of 280 genes across 64 cell type pairs in MERFISH, yielding a MERFISH-to-SeqFISH+ signal ratio of 70.98. In contrast, QuadST identified 626 ICGs in SeqFISH+ and 127 in MERFISH, resulting in a ratio of 1.02. Given that MERFISH includes a large number of canonical cell markers who function distinctly from ICGs, we do not expect a high MERFISH-to-SeqFISH+ signal ratio. The observed ratio in the naïve LR likely reflects its inability to account for unmeasured confounders, measurement errors, or gene-gene correlations.

## Discussion

We introduce QuadST for the robust and powerful identification of cell–cell interactions and their changed genes on single-cell SRT data. QuadST effectively controls the FDRs, even under the existence of inaccurate measurements of cell–cell distances,

unmeasured confounding factors, and gene-gene correlations. Simulation studies demonstrate that QuadST controls the type I error across various settings, including the under the existence of unmeasured confounding factors. It also provides higher power than existing methods, especially when interacting cell pairs cannot be perfectly specified or cell–cell distance cannot be accurately measured. Applying QuadST to two datasets obtained from the mouse cortex via SeqFISH+ and MERFISH technologies, we uncover interacting cell types, genes involved in their interactions, and the corresponding distances of these interactions. We showed that our identified ICGs among excitatory neurons were enriched in synapse functions, demonstrating the effectiveness of our method in revealing intricate details of cell–cell interactions.

The attractive features of QuadST arise from three-tiered modeling design: treating cell–cell distance as outcome, modeling its quantile levels, and contrasting signals across different quantile levels to identify ICGs. First, treating cell–cell distance as outcome reduces the influence of measurement errors on ICG identification. Current technologies often introduce significant errors when quantifying cell–cell distance. Traditional methods typically grouping cells into neighboring versus non-neighboring pairs based on their distances, which are largely vulnerable to these measurement errors. By treating distance as an outcome rather than a grouping variable, QuadST allows these errors to be taken in the regression error and thus limits its impact (Chen et al. 2011). Notably, QuadST employs the test statistics against the null hypothesis that gene expression and cell–cell distance are independent, which can be achieved regardless of treating distance as predictor or outcome. Second, modeling association between gene expression and cell–cell distance at different quantile levels enables QuadST to capture

local signals for ICG identification. Existing methods often model two binary states, "on" and "off", ignoring finer variations, such as stronger effects when cells are closer within the "on" state. Regressions that associate genes with the mean distance fail to capture abundant interactions that occur primarily at very short distances. By employing quantile regression, QuadST surveys the local cell–cell interaction patterns across the entire range of cell–cell distances, capturing the nuanced dynamics of cellular interactions. Lastly, contrasting signals across different quantile levels guarantees robust ICG identifications. Cellular interactions occur within complex microenvironments influenced by numerous factors, such as additional cells and extracellular elements, that may confound the analysis and cannot be fully adjusted. These factors can distort p-value distributions, leading to small p-value under the null. Additionally, genes in cells are often correlated, causing correlated p-values as well. Claiming significance directly using these ill-behaved p-values may result in erroneous conclusions. By comparing signals across quantile levels, QuadST effectively mitigates the impact of confounding factors and gene-gene correlations, yielding reliable and accurate identifications of ICGs even complex data sets, as statistically proved in Method.

Successful application and interpretation of QuadST involves careful attention to several key aspects. First, QuadST performs cell-type-pair-specific analysis, assuming cell types have been pre-identified. Inaccuracies in identifying cell types can reduce QuadST's study power, so careful preparation work, such as data preprocessing, clustering, and annotation, is needed for high-quality analysis. Second, QuadST surveys signals at a grid of user-specified cell–cell distance quantile levels. While it maintains well-controlled FDR even at extreme quantile levels, using too many quantile

levels can reduce computational efficiency but have limited power gain. Conversely, too few quantile levels may result in missed regions of interest. To make a balance, we recommend using at least five cells between two neighboring quantile levels and a maximum of 49 quantile levels. Thirdly, QuadST is designed to borrow information across genes for FDR-controlled identification. Technologies such as CODEX, which spatially profiles fewer than 100 proteins, might not be directly applicable by QuadST. Future studies could enhance QuadST's capabilities by integrating transcriptomics-based signals as priors. Finally, the QuadST identified ICGs are directional and not symmetric, and their interpretations cannot be reversed. For example, a ICG identified for the excitatory neuron (anchor) – astrocyte (neighbor) pair indicates this gene expressed in excitatory neurons is more confidently associated with the distance to astrocytes when the cells close together. This does not provide information of this gene's expression in astrocytes.

## Methods

### *QuadST.*

**Overview.** QuadST is applicable to each pre-defined anchor–neighbor cell-type pair and consists of three key steps: (1) creating the anchor–neighbor integrated matrix, (2) modeling the association between anchor cells' gene expression and cell–cell distance at different distance quantile levels, and (3) identifying ICGs by contrasting the signal levels between nearby and distant quantile levels (Fig. 1).

**Anchor–neighbor integrated matrix.** Suppose a SRT data contains expression levels of  $G$  genes in  $N$  cells. These cells can be categorized into  $K$  cell types based on their

expression (and sometimes spatial) profiles. For each cell type pair  $(l, m)$ , where  $l, m \in (1, \dots, K)$ , we are interested in identifying the existence of their cell–cell interaction and ICGs. Note, when  $l \neq m$ , we study the impact of neighboring cells in  $m$  cell type on gene expression of cells in  $l$  cell type, and when  $l = m$ , we study the impact of neighboring cells in  $m$  cell type on gene expression of cells in its own cell type.

The first step is to define the anchor  $l$  - neighbor  $m$  integrated matrix. We let  $X_{ig}^l$  be the expression level of gene  $g \in (1, \dots, G)$  in each cell  $i \in (1, \dots, N_l)$  in the anchor cell type  $l$ , where  $N_l$  is the number of cells for cell type  $l$ ,  $\mathbf{Z}_i^l$  be the vector of cell-level covariates, and  $Y_i^{lm}$  be the weighted averaged distance from cell  $i$  to its  $k$  nearest cells in cell type  $m$ , respectively. Parameter  $k$  denotes number of nearest neighbor cells under consideration. Let  $Y_{i,t}^{lm}$  be the distance from the anchor cell  $i$  to its  $t$ th nearest neighbor, then  $Y_i^{lm} = k / \sum_{t=1}^k (1/Y_{i,t}^{lm})$ . When  $k = 1$ ,  $Y_i^{lm}$  simplifies to  $Y_{i,1}^{lm}$ , the distance from cell  $i$  to its nearest neighbor cell. Then  $(Y_i^{lm}, X_{i1}^l, \dots, X_{iG}^l, \mathbf{Z}_i^{lT})_{i=(1, \dots, N_l)}$  is the anchor  $l$  -neighbor  $m$  integrated matrix. For a total of  $K$  cell types of interest, such matrices can be constructed for  $K^2$  anchor–neighbor cell-type pairs. As QuadST is designed to analyze on these pairs separately, we will omit the index  $l$  and  $m$ , using  $Y_i$ ,  $X_{ig}$ , and  $\mathbf{Z}_i$  in the following section for simplicity.

**Distance-quantile based model.** Assuming that nearby cells are more likely to interact with each other than distant cells, with stronger interactions likely perturbing more genes, we exam the association between gene expression and cell–cell distance at different quantile levels of the distance. Unlike models focusing on mean distance, our model allows for distinct associations in nearby versus distant cells. Specifically, we

consider a grid of evenly spaced quantile levels  $\tau = (\tau_1, \tau_2, \dots, \tau_J)$ , such as (0.02, 0.04, ..., 0.98) for 49 quantile levels, across the entire cell–cell distance distribution. At each quantile level  $\tau_j \in (0,1)$ , where  $j \in (1,2, \dots, J)$ , we model the association with gene  $g$  as follows:

$$Q_{Y_i}(\tau_j | X_{ig}, \mathbf{Z}_i) = X_{ig} \beta_{g,\tau_j} + \mathbf{Z}_i^T \boldsymbol{\alpha}_{g,\tau_j}. \quad (1)$$

Here,  $\beta_{g,\tau}$  captures the effects of gene  $g$  on the  $\tau_j$ th quantile of cell–cell distance and similarly  $\boldsymbol{\alpha}_{g,\tau}$  captures the effects of cell-level covariates. The p-value under the null hypothesis  $\beta_{g,\tau_j} = 0$  can be obtained using quantile rank-score test (Song et al. 2017).

It is worth noting that single-cell transcriptomics data often contains excessive zeros due to both biological and technical factors (Jiang et al. 2022). We address this issue by recognizing that the excessive zeros complement the continuous counts in providing expression abundance information, and thus extend our model under zero inflation to include a nonzero indicator as an extra predictor:

$$Q_{Y_i}(\tau_j | X_{ig}, \mathbf{Z}_i) = X_{ig} \beta_{g,\tau_j}^{(1)} + I(X_{ig} \neq 0) \beta_{g,\tau_j}^{(2)} + \mathbf{Z}_i^T \boldsymbol{\alpha}_{g,\tau_j}, \quad (2)$$

where  $I(\cdot)$  is an indicator function indicating whether gene expression is nonzero. This structure allows zero counts to have a discrete effect on the outcome from continuous values. Both  $\beta_{g,\tau_j}^{(1)}$  and  $\beta_{g,\tau_j}^{(2)}$  capture the effects of gene  $g$  on  $\tau_j$ th quantile of cell–cell distance and their consistent directions matter in interpreting the associations. Therefore, we propose to jointly test their effects while taking into consideration of the directions. Specifically, we first project the gene expression predictors  $A = (X, I(X \neq 0))$

to the column space of  $Z$  to obtain orthogonal signals of covariates, such that  $A^* = A - Z(Z^T Z)^{-1} Z^T A$ , where  $X = \{X_{ig}\}$  and  $Z = \{Z_i\}$  are the matrix forms of the predictors. Then, the quantile rank score function can be defined as  $S_n = n^{-1/2} A^{*T} \phi_\tau(Y - Z^T \hat{\alpha})$ , where  $Y = \{Y_i\}$  is the matrix form of  $Y_i$ ,  $\phi_\tau(u) = \tau - I(u < 0)$  is an asymmetric sign function for quantile regression, and  $\hat{\alpha}$  is the estimated coefficient under the null. Under the null hypothesis,  $S_n$  is a  $p \times 1$  matrix following a normal distribution with mean zero and variance  $\Sigma_n = n^{-1} \tau(1 - \tau) A^{*T} A^*$ . We standardize  $S_n$  and take an average across the  $p$  predictors, such that  $\widetilde{S}_n = n^{-1} 1^T \Sigma_n^{-1/2} S_n$ . Then,  $\widetilde{S}_n$  is our test statistics that follows  $N(0,1)$  under the null.

**Identification of ICGs.** Designed to exam associations at a grid of evenly distributed quantile levels across the entire cell–cell distance distribution, QuadST does not require knowledge of a specific distance where cell–cell interaction occurs. Specifically, from a total of  $G$  genes examed at a total of  $J$  distance quantile levels, we obtain a  $G \times J$  matrix for p-values. Its each element  $p_{g,\tau_j}$  captures the association at the quantile level  $\tau_j$  for gene  $g$  from Equation (1) or (2). For the  $j$  upper quantile levels ( $J, J - 1, \dots, J - j + 1$ ), we define  $A_j$  as the  $G \times j$  subset of the p-value matrix covering all  $G$  genes; similarly, for the  $j$  lower quantile levels ( $1, 2, \dots, j$ ), we define  $B_j$  as the  $G \times j$  subset of the p-value matrix. Then, each element of  $A_j$  and  $B_j$ , denoted as  $p_{g,h}^{A_j}$  and  $p_{g,h}^{B_j}$ , where  $h \in (1, \dots, j)$  and  $g \in (1, \dots, G)$ , capture aggregated signals from the same number of quantile levels in two tails. Then, we contrast  $p_{g,h}^{A_j}$  and  $p_{g,h}^{B_j}$  to identify ICGs who are more likely to be associated with cell–cell distance in nearby cells versus distant cells. Specifically, we

368 calculate  $T_j(c)$  as the ratio of the number of genes with p-values below a small positive  
 369 cutoff  $c$  for the upper versus lower tails:

$$T_j(c) = \frac{\sum_{g=1}^G \sum_{h=1}^j (p_{g,h}^{A_j} < c)}{\sum_{g=1}^G \sum_{h=1}^j (p_{g,h}^{B_j} < c)}. \quad (3)$$

370  $T_j(c)$  close to 1 indicates a similar number of genes with small p-values for both nearby  
 371 and distant cells, suggesting a lack evidence for cell–cell interaction. Conversely, a  
 372 small ratio suggests more genes with small p-values in nearby cells than distant cells,  
 373 indicating the presence of cell–cell interactions. Therefore, an evaluation  $T_j(c)$  can be  
 374 used to infer the presence of cell–cell interaction and its ICGs.

375 It is critical to note that under two simple assumptions,  $T_j(c)$  provides an upper bound of  
 376 the eFDR (Tusher et al. 2001; Efron and Tibshirani 2002). The eFDR is a data-driven,  
 377 unbiased estimator of FDR that uses observed test statistics under the null hypothesis  
 378 and the actual data to estimate the proportion of null hypotheses for FDR. Occasionally  
 379 eFDR exceeds 1, which can be interpreted as FDR=1. The two simple assumptions we  
 380 need are:

381 1. The  $p_{g,h}^{A_j}$  and  $p_{g,h}^{B_j}$  are symmetric under the null, such that  $\#(null\ g: p_{g,h}^{A_j} < c) =$   
 382  $\#(null\ g: p_{g,h}^{B_j} < c)$ .

383 2. The  $p_{g,h}^{A_j}$  is smaller under the alternative than under the null, resulting in  
 384  $\#(null\ g: p_{g,h}^{A_j} < c) \leq \#(g: p_{g,h}^{A_j} < c)$ .

385 In this context, the eFDR can be expressed as:

$$eFDR_j(c) = \frac{\#(\text{null } g: p_{g,h}^{B_j} < c)}{\#(g: p_{g,h}^{B_j} < c)} = \frac{\#(\text{null } g: p_{g,h}^{A_j} < c)}{\#(g: p_{g,h}^{B_j} < c)} \leq \frac{\#(g: p_{g,h}^{A_j} < c)}{\#(g: p_{g,h}^{B_j} < c)} = T_j(c) \quad (4)$$

386 This feature of  $T_j(c)$  is highly attractive, as it bounds the FDR without assuming well-

387 behaved distributions of  $p_{g,h}^{A_j}$  and  $p_{g,h}^{B_j}$ . For example, under the existence of unmeasured

388 confounders, such as with unknown effects from extracellular factors or unknown biases

389 from technical handling, the  $p_{g,h}^{A_j}$  and  $p_{g,h}^{B_j}$  can be biased to small values under the null.

390 Similarly, under gene-gene correlation,  $p_{g,h}^{A_j}$  and  $p_{g,h}^{B_j}$  can have distorted distributions. In

391 both cases, by contrasting  $p_{g,h}^{A_j}$  and  $p_{g,h}^{B_j}$  instead of comparing each of them to a

392 theoretical distribution like uniform(0,1),  $T_j(c)$  ensures the FDR control. Therefore,

393 QuadST provides computationally efficient and highly robustness FDR control under

394 complex p-value behaviors.

395 To identify ICGs at a desired nominal eFDR level  $\alpha$  (say 10%), we search for the p-

396 value threshold  $c$  as

$$C_j = \max \left\{ c: \frac{\sum_{g=1}^G \sum_{h=1}^j (p_{g,h}^{A_j} < c) + d}{\sum_{g=1}^G \sum_{h=1}^j (p_{g,h}^{B_j} < c) + d} \leq \alpha \right\}, \quad (5)$$

397 where  $d > 0$  is a small constant to make eFDR control conservative. Finding the

398 appropriate threshold  $C_j$  indicates that cell-cell interaction exists at or below the desired

399 nominal FDR level  $\alpha$  at the quantile level  $\tau_j$ . In other words,  $C_j$  cannot be found, we

claim that there is no cell–cell interaction in this cell type pair. Under the existence of  $C_j$ , the significant genes  $\tilde{G}_j$  are the genes with  $\text{eFDR} < \alpha$ .

ICGs are the significant genes at quantile level  $I$ , i.e.,  $ICG = \tilde{G}_I$ , where  $I$  maximizes the number of identified genes, i.e.,  $I = (j: \max(\#\tilde{G}_j))$ . Meanwhile, the cell–cell interaction distance is the distance corresponding to quantile level  $I$ .

### *Simulation studies.*

**Simulation Settings.** We have performed comprehensive simulations to evaluate the performance of QuadST, including evaluating the impact of quantile levels, number of nearest neighbors ( $k$ ), and number of cells (sample size). We also compared the results with three alternative methods, Giotto and NCEM and LR, that identify genes involved in cell-type-specific cell–cell interactions. Here, we first describe a basic simulation set up that serves as the skeleton of all simulations, as well as the general implementation of QuadST and comparison methods, and then explain the alternations in each individual simulation study.

*Basic simulation setup (skeleton).* We randomly generated 5000 cells uniformly distributed in a unit square. These cells were equally divided into five cell types, with one cell type randomly chosen as the anchor cell type and another cell type randomly selected as the neighbor cell type. In each anchor cell, we simulated 2000 genes whose expression levels were block-wise correlated, with a block size of 100. Within each block, gene expression levels followed a multivariate normal distribution with an autocorrelation coefficient 0.7 and homogenous variance. We also assumed that 10% of

the anchor cells, which were most proximal to cells in the neighbor cell type, were involved in cell–cell interactions, which perturbed the expression levels of 20% genes (i.e., 4 blocks out of 20 blocks) in these cells. For genes not involved in cell–cell interactions, their expression levels had a mean 0 and variance 1. For the four blocks of genes involved in cell–cell interactions, we simulated four different distance-expression association patterns as follows:

*Block 1: Constant Mean Shift.* This pattern represented an "on/off" pattern, where cell–cell interaction was turned "off" above the 10% cell distance threshold and "on" below the threshold. Under the interaction, the effect was a constant mean shift from 0 to  $\mu_\delta$ , where  $\mu_\delta \sim N(0.4, 0.2)$ .

*Block 2: Dosage Mean Shift.* This pattern represented a "dosage" pattern, where cell–cell interaction increased as cells become closer after being turned "on" at the distance threshold. The mean shift was generated as a piecewise linear function of distance ( $Y$ ), such that  $\mu_\delta(Y) = a(Y - d_c)I(Y < d_c)$ , where  $d_c$  was the distance cutoff for 10% of cells and  $a \sim N(-150, 75)$  was the slope for dosage effect.

*Block 3: Constant Mean Shift with Variance Increase.* Cell–cell interaction not only impacted mean in the "on/off" pattern same as Block 1, but also impact the variance of gene expression from 1 to  $var_\delta = 2$ .

*Block 4: Dosage Mean Shift with Variance Increase.* Cell–cell interaction not only impacted mean in the "dosage" pattern same as Block 2, but also the variance of gene expression from 1 to  $var_\delta = 2$ .

*Implementation of QuadST and comparison methods.* We compared QuadST with the three other methods, Giotto and NCEM and LR in simulation studies. Unless revised in the individual settings, we used one nearest neighbor and 49 quantile levels in QuadST analysis. Both Giotto and NCEM analysis required the determination of interacting versus noninteracting cell pairs, such as by using cell–cell distance or network. These interacting cell pairs can hardly be well estimated in real data, but in simulations we used the simulated known ground truth to demonstrate their best performances. Finally, we developed a LR approach to demonstrate the importance of modeling and contrasting different quantile levels, as designed in QuadST, rather than focusing on mean distance. This method modeled the mean cell–cell distance on gene expression level using linear regression, as such  $E[Y_i] = X_{ig}\beta_g + \mathbf{Z}_i^T\alpha_g$ . ICGs in all methods were all claimed significant at 10% FDR. Averaged results from 10 times Mante Carlo simulations were reported.

*Individual Setting 1* (relevant to Fig. 2A). To evaluate the impact of number of quantile levels used in QuadST on its performance, we preformed simulations using the basic simulation setup and evaluated QuadST that uses six different sets of evenly distributed quantile levels, including three quantile levels at 0.25, 0.5, 0.75, four quantile levels at 0.2, 0.4, 0.6, 0.8, nine quantile levels at 0.1, 0.2, ..., 0.9, 19 quantile levels at 0.05, 0.1, ..., 0.95, 49 quantile levels at 0.02, 0.04, ..., 0.98, 99 quantile levels at 0.01, 0.02, ... , 0.99, and 999 quantile levels at 0.001, 0.002, ..., 0.999. In our simulated data with 1000 cells, specifying 999 quantile levels provides assessment of QuadST under very extreme tails.

*Individual Setting 2* (relevant to Fig. 2B). To evaluate the impact of using different number of  $k$  in  $k$ -nearest neighbors in QuadST on its performance, we performed simulations also using the basic simulation setup and evaluated QuadST using five  $k$  values at 1, 2, ..., 5.

*Individual Setting 3* (relevant to Fig. 2C). To evaluate the impact of small sample size the performance of QuadST, instead of simulating 1000 cells per cell type as in the basic simulation setup, we also simulated 100, 300, 500, and 700 cells per cell type.

*Individual Setting 4* (relevant to Fig. 2D). We evaluate the FDR and power of QuadST and other methods in the ideal case using the basic simulation setup.

*Individual Setting 5* (relevant to Fig. 2E). As both Giotto and NCEM required the determination of interacting versus noninteracting cell pairs, we compared QuadST and LR to these two methods using a distance threshold that was 1.4 times of the simulated value, approximately doubling the interacting cell pairs from the simulated values. This demonstrated the performance of these methods when interacting cell pairs could not be estimated with accuracy.

*Individual Setting 6* (relevant to Fig. 2F). To evaluate the robustness of these methods, under the biasedly quantified cell–cell distance ( $Y$ ), we specified that observed distance  $\tilde{Y} = Y + \text{uniform}(0, 0.005)$ , and evaluated their performance under observed  $\tilde{Y}$ .

*Individual Setting 7* (relevant to Fig. 2G). To understand the impact of unmeasured confounding factors on analysis, we introduced region as a factor impacting cell locations and expression levels. We split the unit square into four equal-sized regions

( $R_1 - R_4$ ), altered the locations of neighboring cells to appear in  $R_1 - R_3$ , and increased by 0.2 the mean expression level for all genes in anchor cells in  $R_4$ . In this case, both cell composition and gene expression level varied by region, making it a confounding factor for the association between cell–cell distance and expression. We assumed region information was not accessible to us, and compared QuadST with other methods under the unmeasured confounder.

## Data analysis.

**seqFISH+ dataset.** High-throughput single-cell SRT data obtained from the cortex of a 23-day-old male mouse (C57BL/6J) using seqFISH+ were downloaded for analysis (Eng et al. 2019). It included spatial map and expression levels for 10,000 genes in 523 cells, imaged from 5 field of views (FOVs). The cells were categorized by the original study into nine major cell types, including excitatory neuron (n=325), interneuron (n=42), astrocyte (n=54), endothelial (n=45), oligodendrocyte (n=29), microglia (n=16), neural stem cells (n=6), neuroblast (n=5), and ependymal (n=1).

**MERFISH dataset.** We also analyzed the single-cell SRT data from mouse frontal cortex, profiled using MERFISH (Allen et al. 2023). The analysis focused on a 4-week-old female mouse (C57BL/6J) similar in age to seqFISH+ data. The data consists of gene expression levels for 374 genes in 13,745 cells collected from cortical layers II-V. These cells have been categorized into eight major cell types: excitatory neuron (n=7129), inhibitory neuron (n=1161), astrocyte (n=1333), microglia (n=787), oligodendrocyte (n=1153), oligodendrocyte-precursor (n=462), endothelial (n=1360),

and pericyte (n=360)<sup>12</sup>. The spatial coordinates of cells were provided in CELL x GENE repository (shown in the key resource table of the reference(Allen et al. 2023)).

**Dataset preprocessing.** To preprocess seqFISH+ data, we stitched cell's coordinates obtained from 5 FOVs to provide the global cell spatial coordinates. We required a minimal of 10 cells for QuadST analysis and thus focused our analyses six cell types. Then, we normalized gene counts across all remaining cells using scran(Lun et al. 2016), which adjusted for cell-specific bias due to cell-to-cell difference in library size and capture efficiency. Following the practice in the source (Eng et al. 2019), our analysis focused on highly-expressed genes whose expression levels belong to top 25% quantile in each cell type. Then the cell specific-bias adjusted counts were transformed to normal distributions by (1) calculating the empirical cumulative distribution function, (2) using inverse cumulative distribution function to transform it into standard normal distribution, and (3) shifting the minimum to 0. Similarly, in the preprocess of MERFISH data, we normalized gene counts using scran to adjust for cell-specific bias, kept genes whose expression levels belong to top 75% quantile in each cell type, and transferred the counts into normal distributions.

**UMAP and covariate effect analysis.** In each SRT dataset, we performed UMAP (Uniform Manifold Approximation and Projection) analysis on gene expression data using R package *umap* (R Core Team). Then, we calculated the partial  $R^2$  of each covariate on UMAP coordinates using a generalized linear model using R package *rsq*, where each UMAP coordinate was considered as a response and covariates of interest were considered as predictors.

**QuadST analysis in SeqFISH+ dataset.** We ran QuadST on the preprocessed SeqFISH+ dataset. For a total of six cell types annotated by the source, we created a total of 36 (6 times 6) anchor–neighbor integrated matrices for all pairs of the same and different cell types. Depending on the cell type, a total of 1,235-2,500 highly variable genes were considered for analysis, and fields of view were included as covariates. Given that excessive zeros were observed in these genes, we used the distance-quantile based model given by Eq. (2) to test distance-expression association at the evenly spaced distance quantile levels. For cell types with a relatively large sample size, such as excitatory neuron (n=325), we chose a maximum number of 49 quantile levels at 0.02, 0.04, ..., 0.98 for analysis. For cell types that had few cells, we chose the number of quantile levels to ensure a minimal of 5 cells between two adjacent quantile levels. For example, for microglia (n=16), we used only three quantile levels at 0.25, 0.5, 0.75. We contrasted the resulting p-values at different quantile levels as described in QuadST to identify ICGs for each cell-type pair at 10% FDR. To determine the association direction and understand the association confidence, we extracted the combined Z-score and derived the directional association score as the  $\text{sign}(Z\text{-score}) \cdot (-\log_{10} p\text{-value})$  at the interaction quantile level (l).

**QuadST analysis in MERFISH dataset.** Similar to SeqFISH+ analysis, we applied QuadST to the preprocessed MERFISH dataset for a total of eight cell types annotated by the source. We analyzed the gene expression levels of 280 highly variable genes for 64 cell type pairs, adjusting cortical layers as covariates. The distance-quantile-based model for zero-inflated genes were also used in this data to test the association. Given

the large sample size for all cell types, 49 quantile levels at 0.02, 0.04, ..., 0.98 were used for analysis. ICGs were identified at 10% FDR.

**ICG function annotation using existing databases.** To examine known biological roles of ICGs identified in SeqFISH+ dataset, we looked up in existing databases on ligands and receptors(Shao et al. 2021) as well as transcription factors and cofactors (Zhou et al. 2017). We obtained a set of mouse ligands and receptors from CellTalkDB (Shao et al. 2021) (mouse\_lr\_pair.txt) on <http://tcm.zju.edu.cn/celltalkdb/>. We obtained the mouse transcription factors and cofactors from a supplementary table (41467\_2017\_BFncomms15089\_MOESM3449\_ESM.xlsx) of the reference(Zhou et al. 2017).

**GO cellular component enrichment analysis.** To test the GO cellular component terms enriched for ICGs among excitatory neurons, we performed Wilcoxon Rank Sum test comparing p-values from genes overlapping versus nonoverlapping with a given GO cellular component gene set. The reported p-values were adjusted for multiple testing (i.e., FDR 10%). The GO cellular component gene sets for mouse (m5.go.cc.v2023.1.Mm.symbols.gmt) were downloaded from GSEA website <https://www.gsea-msigdb.org/gsea/msigdb/mouse/collections.jsp>.

**Analyses with comparison methods in SeqFISH+ and MERFISH datasets.** To understand how QuadST produces different results from the comparison methods, we also applied NCEM, Giotto, and LR to analyze the SeqFISH+ and MERFISH data. To implement Giotto, we first build a spatial network based on the Delaunay triangulation as suggested by the method, and identified cells with a common edge in the network as

neighbors and cells that are not indirectly connected as non-neighbors. Then, we compare expression levels of the highly variable genes separately for each cell-type pair between neighboring with non-neighboring cell pairs. To control for the covariate effects from the FOVs in SeqFISH+ and cortex layers in MERFISH within the association, we revised the Giotto pipeline from using t-test to regression. To implement NCEM, which suggested an averaged 69um cell—cell interacting distance on multiple platforms including MERFISH, we applied three distance cutoffs (59um, 69um, and 79um), controlling for the same covariates as above. Notably, as NCEM performs analysis using all cells, instead of by cell type pairs, we analyzed the 2500 highly variable genes in SeqFISH+ identified across all cells, rather than by cell types as done in QuadST and Giotto, as well as the same 280 genes in MERFISH. In LR, we regressed cell—cell distance on expression controlling FOVs in SeqFISH+ and cortex layers in MERFISH as covariates. In all analyses, we controlled for 10% FDR for significance calling.

**Software availability.** QuadST software is freely available in the Supplementary Material 1. QuadST, together with this source codes for reproducing all the analyses is available at <https://github.com/songxiaoyu/QuadST>.

## Acknowledgements

This work was supported by US NIH R03AG075567, US NIH RF1MH133703, Singapore Ministry of Education FY2024-MOET1-0004, and Singapore Ministry of Health Duke-NUS Signature Research Programme. Conceptualization: XS, PW, G-CY; Methodology: XS, PW, G-CY; Simulation: XS, YS, G-CY; Real data analysis: XS, YS,

ME, PR, PW, G-CY; Writing—original draft: XS; Writing- review and editing: XS, YS, ME,  
PR, PW, G-CY; Software: XS.

## Competing interest statement

The authors declare no competing interests.

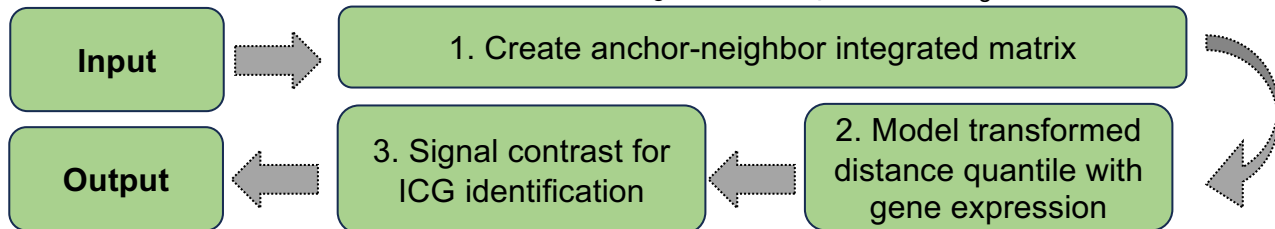
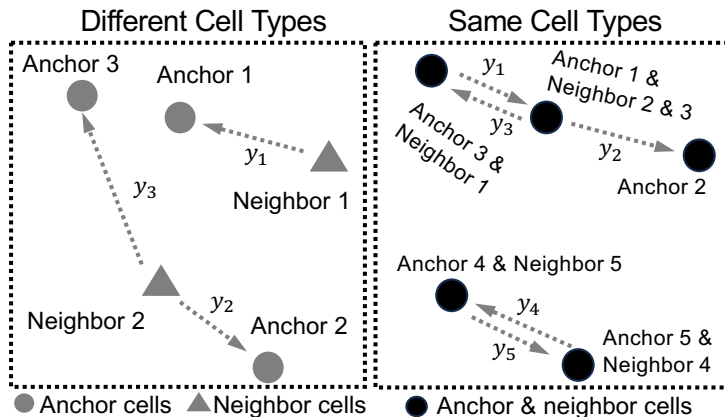
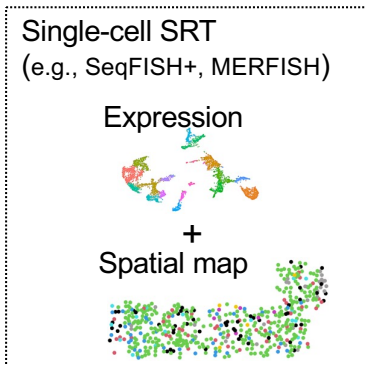
## References

- Allen WE, Blosser TR, Sullivan ZA, Dulac C, Zhuang X. 2023. Molecular and spatial signatures of mouse brain aging at single-cell resolution. *Cell* **186**: 194-208.e18.
- 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.
- Baszkin A, Norde W. 1999. *Physical chemistry of biological interfaces*. CRC Press.
- Burgess DJ. 2019. Spatial transcriptomics coming of age. *Nat Rev Genet* **20**: 317.
- Cang Z, Nie Q. 2020. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat Commun* **11**: 2084.
- Chang S, Reim K, Pedersen M, Neher E, Brose N, Taschenberger H. 2015. Complexin stabilizes newly primed synaptic vesicles and prevents their premature fusion at the mouse calyx of held synapse. *J Neurosci Off J Soc Neurosci* **35**: 8272–8290.
- Chen X, Hong H, Nekipelov D. 2011. Nonlinear models of measurement errors. *J Econ Lit* **49**: 901–37.
- Dries R, Zhu Q, Dong R, Eng C-HL, Li H, Liu K, Fu Y, Zhao T, Sarkar A, Bao F. 2021. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol* **22**: 1–31.
- Dries R, Zhu Q, Dong R, Linus Eng C-H, Li H, Liu K, Fu Y, Zhao T, Sarkar A, Bao F, et al. 2019. *Giotto, a toolbox for integrative analysis and visualization of spatial expression data*. Bioinformatics <http://biorxiv.org/lookup/doi/10.1101/701680> (Accessed February 2, 2021).
- Efron B, Tibshirani R. 2002. Empirical bayes methods and false discovery rates for microarrays. *Genet Epidemiol* **23**: 70–86.

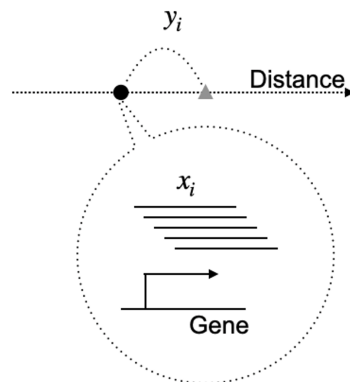
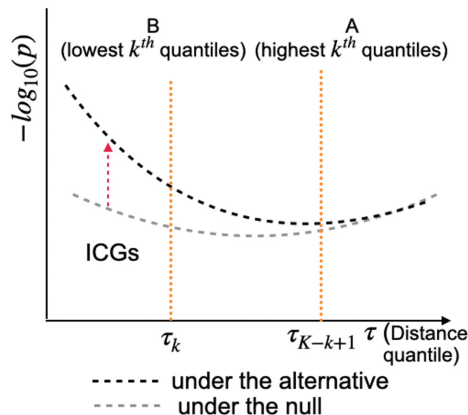
- 624 Eng C-HL, Lawson M, Zhu Q, Dries R, Kouloua N, Takei Y, Yun J, Cronin C, Karp C,  
625 Yuan G-C, et al. 2019. Transcriptome-scale super-resolved imaging in tissues by  
626 RNA seqFISH. *Nature* **568**: 235–239.
- 627 Fischer DS, Schaar AC, Theis FJ. 2023. Modeling intercellular communication in tissues  
628 using spatial graphs of cells. *Nat Biotechnol* **41**: 332–336.
- 629 Gomes GM, Bär J, Karpova A, Kreutz MR. 2023. A Jacob/nsmf gene knockout does not  
630 protect against acute hypoxia- and NMDA-induced excitotoxic cell death. *Mol*  
631 *Brain* **16**: 23.
- 632 Jiang R, Sun T, Song D, Li JJ. 2022. Statistics or biology: the zero-inflation controversy  
633 about scRNA-seq data. *Genome Biol* **23**: 31.
- 634 Li Z, Wang T, Liu P, Huang Y. 2023. SpatialDM for rapid identification of spatially co-  
635 expressed ligand-receptor and revealing cell-cell communication patterns. *Nat*  
636 *Commun* **14**: 3995.
- 637 Lun ATL, McCarthy DJ, Marioni JC. 2016. A step-by-step workflow for low-level analysis  
638 of single-cell RNA-seq data with Bioconductor. *F1000Research* **5**: 2122.
- 639 Ma W, Song X, Yuan G-C, Wang P. 2024. RECCIPE: A new framework assessing  
640 localized cell-cell interaction on gene expression in multicellular ST data. *Front*  
641 *Genet* **15**: 1322886.
- 642 Maffuid K, Cao Y. 2023. Decoding the Complexity of Immune-Cancer Cell Interactions:  
643 Empowering the Future of Cancer Immunotherapy. *Cancers* **15**: 4188.
- 644 Pham D, Tan X, Balderson B, Xu J, Grice LF, Yoon S, Willis EF, Tran M, Lam PY,  
645 Raghubar A, et al. 2023. Robust mapping of spatiotemporal trajectories and cell-  
646 cell interactions in healthy and diseased tissues. *Nat Commun* **14**: 7739.
- 647 R Core Team. R: A Language and Environment for Statistical Computing.  
648 <https://www.R-project.org>.
- 649 Shao X, Liao J, Li C, Lu X, Cheng J, Fan X. 2021. CellTalkDB: a manually curated  
650 database of ligand-receptor interactions in humans and mice. *Brief Bioinform* **22**:  
651 bbaa269.
- 652 Song X, Li G, Zhou Z, Wang X, Ionita-Laza I, Wei Y. 2017. QRank: a novel quantile  
653 regression tool for eQTL discovery. *Bioinforma Oxf Engl* **33**: 2123–2130.
- 654 Szepesi Z, Manouchehrian O, Bachiller S, Deierborg T. 2018. Bidirectional Microglia-  
655 Neuron Communication in Health and Disease. *Front Cell Neurosci* **12**: 323.
- 656 Tusher VG, Tibshirani R, Chu G. 2001. Significance analysis of microarrays applied to  
657 the ionizing radiation response. *Proc Natl Acad Sci U S A* **98**: 5116–5121.

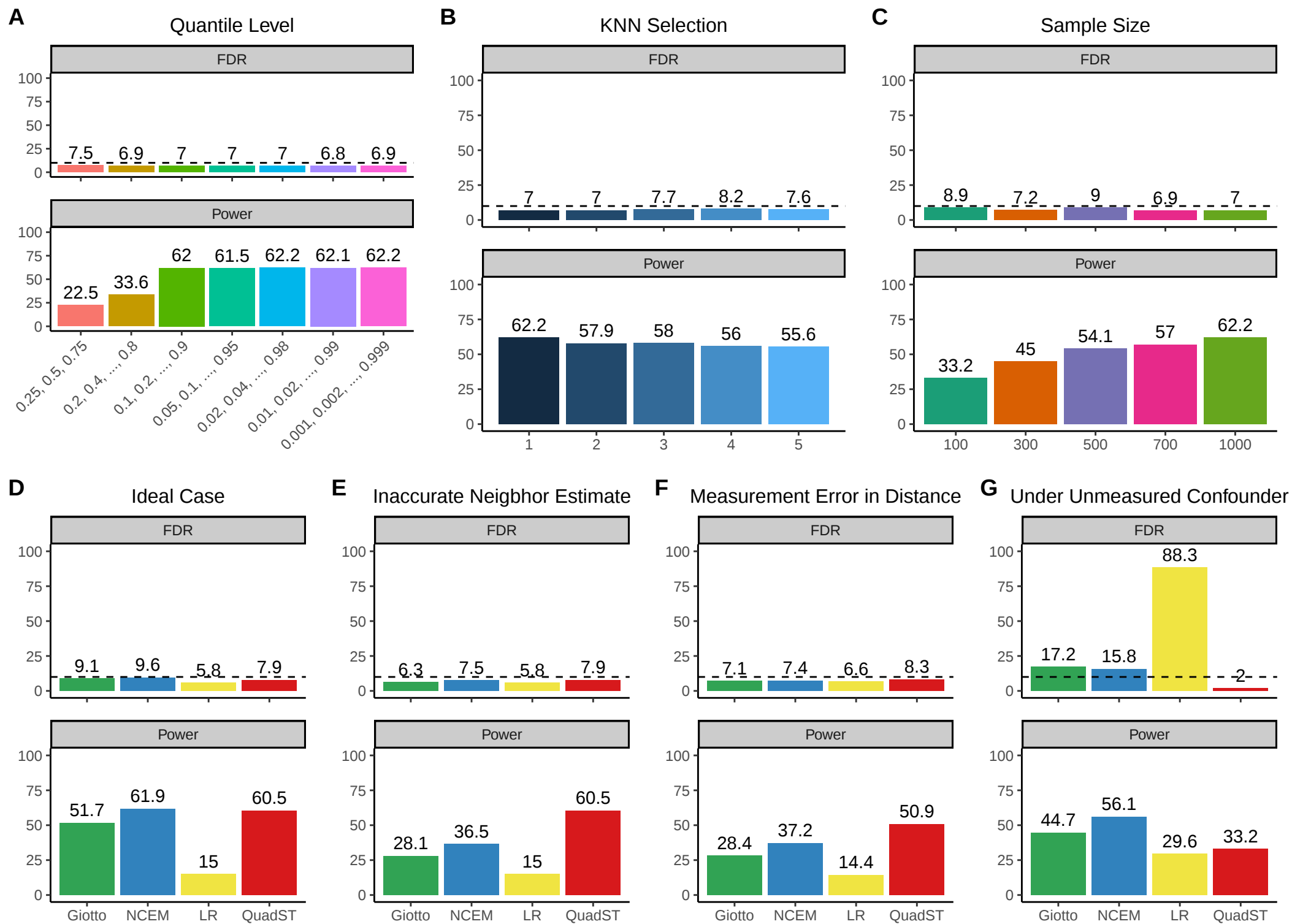
658 Zhou Q, Liu M, Xia X, Gong T, Feng J, Liu W, Liu Y, Zhen B, Wang Y, Ding C, et al.  
659 2017. A mouse tissue transcription factor atlas. *Nat Commun* **8**: 15089.

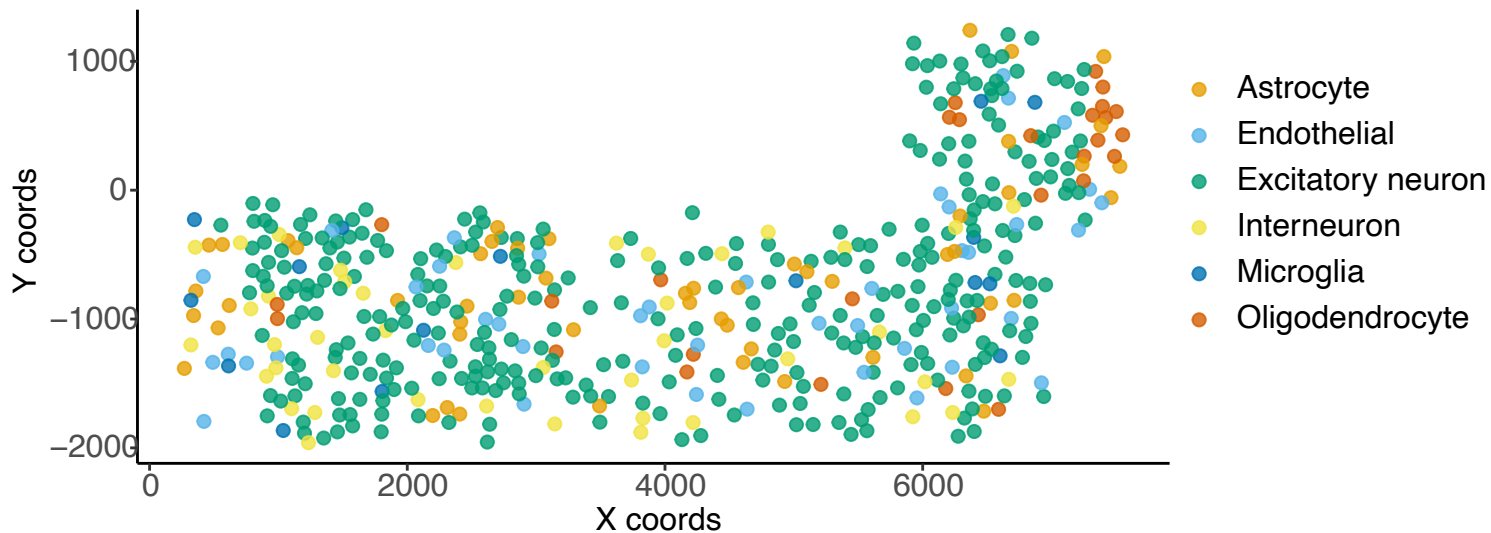
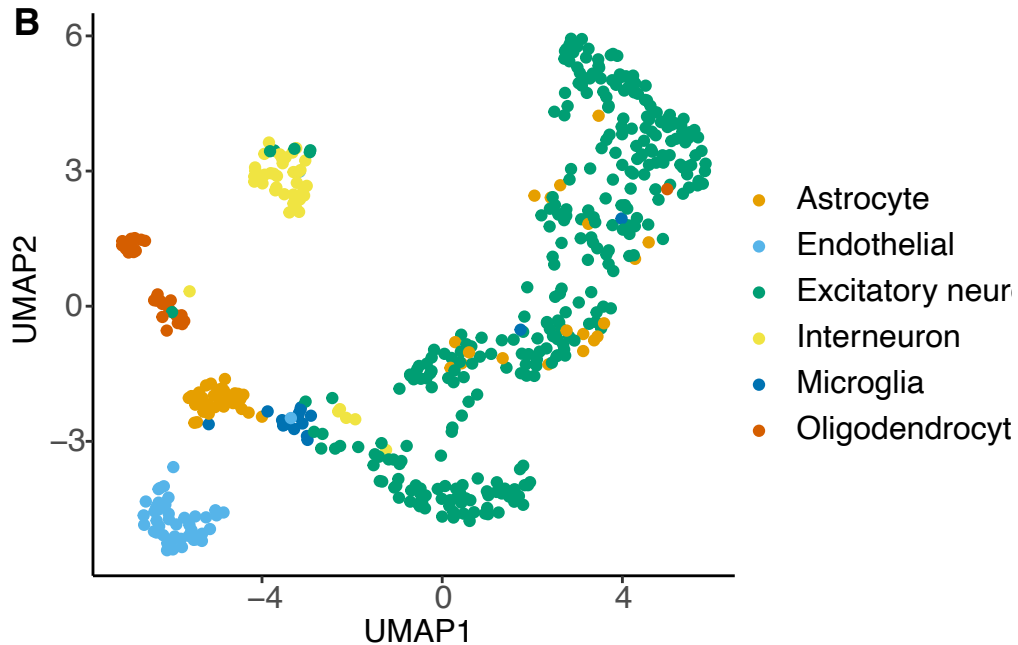
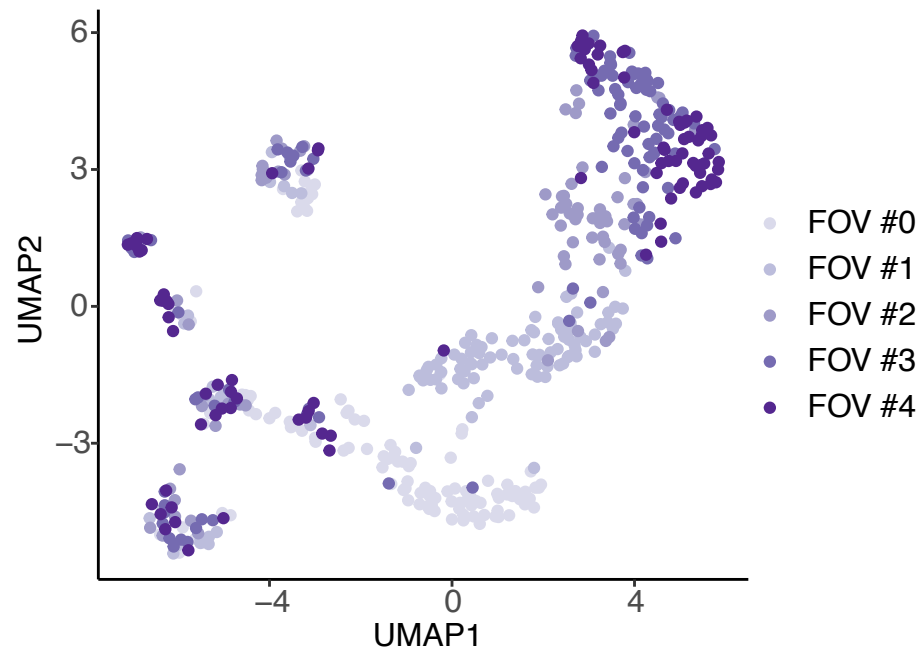
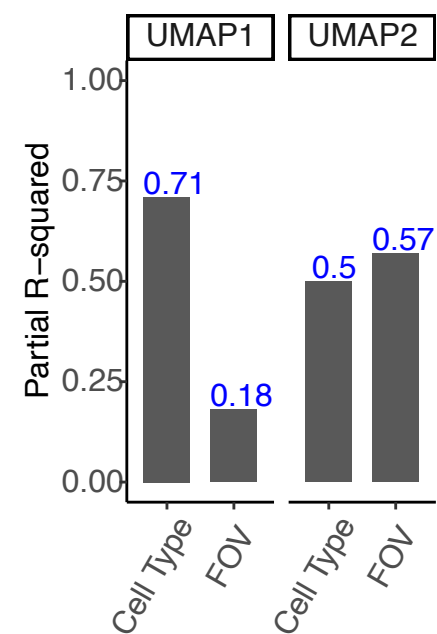
660



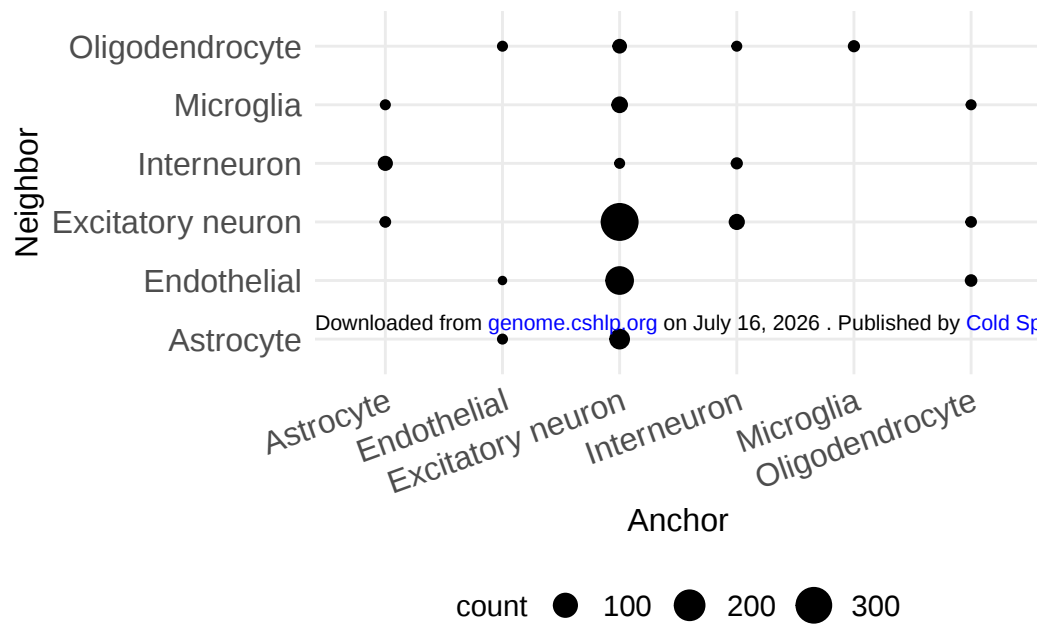
- a. Interacting cell types.
- b. ICGs with association directions.
- c. Cell-cell interaction distance.



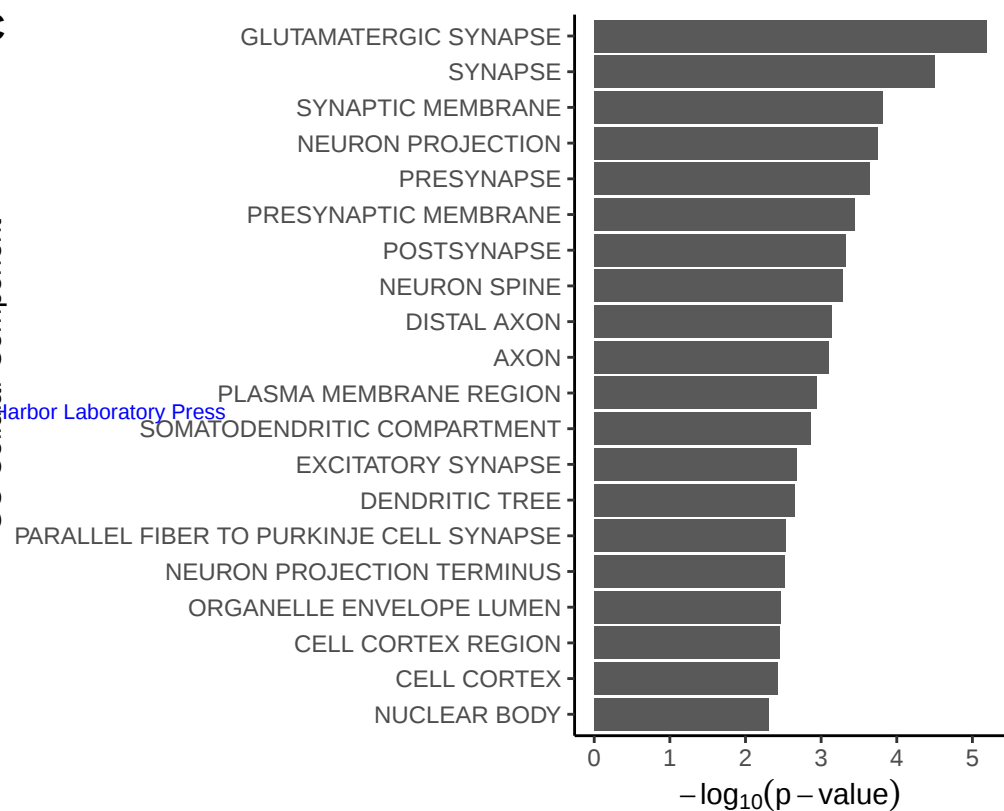


**A****B****C**

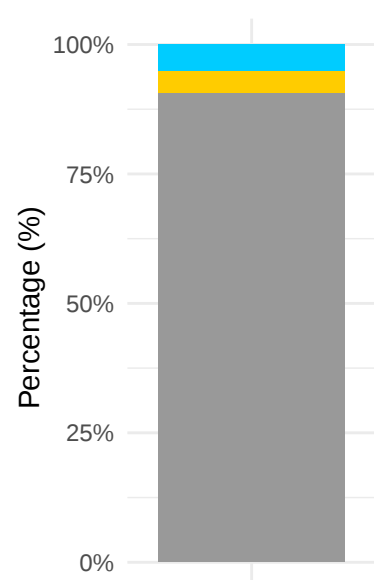
A



C



B

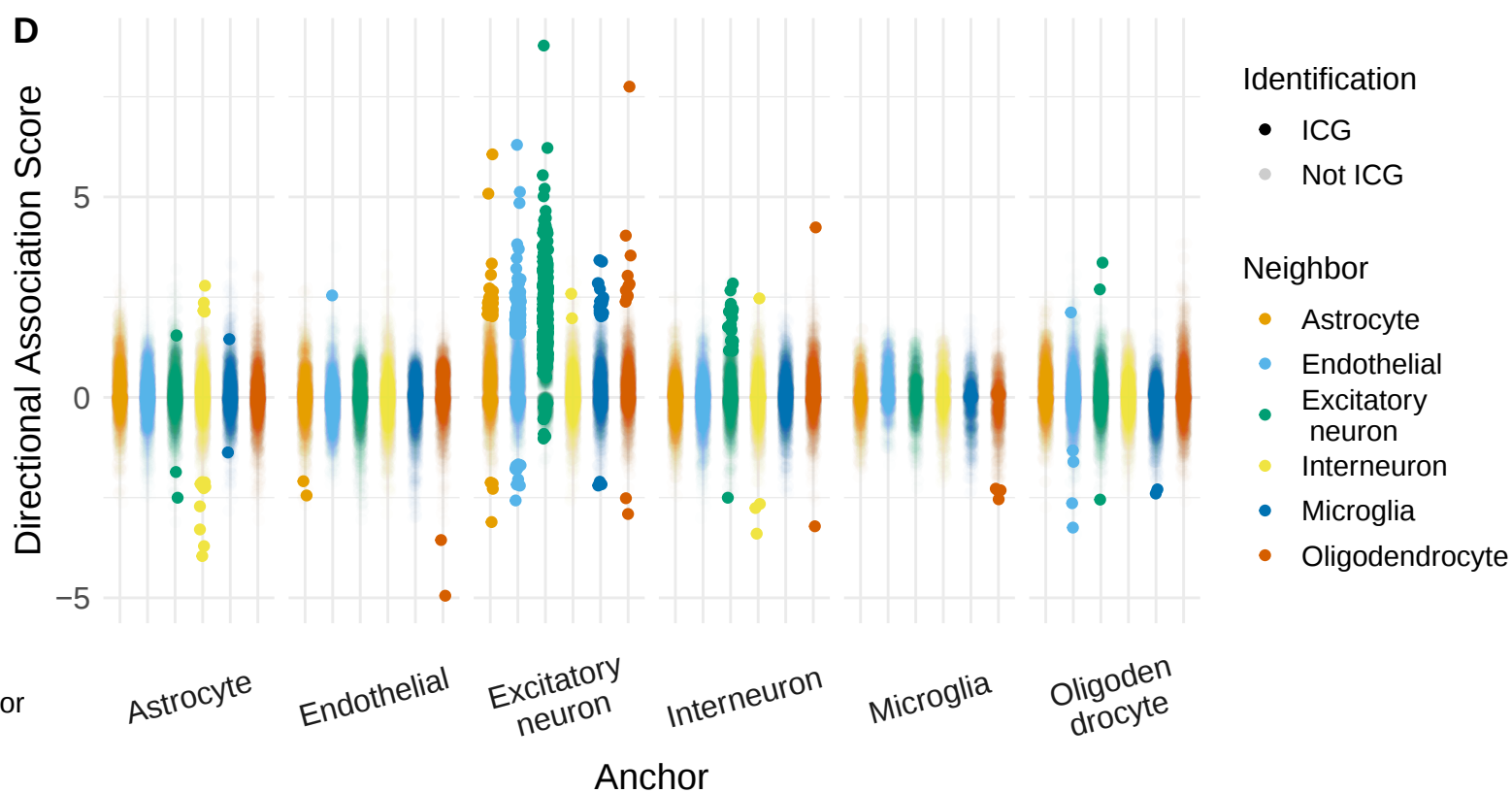


Ligand & Receptor

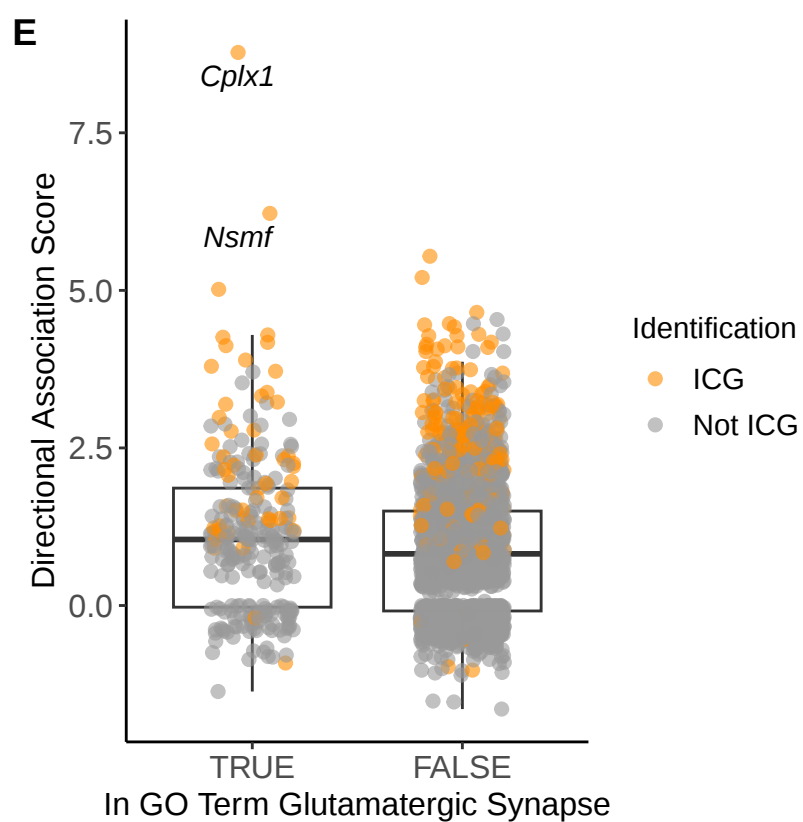
Transcription factor & cofactor

Others

D



E



F

