Inferring a spatial code of cell-cell interactions across a whole animal body

Erick Armingol[1,2], Abbas Ghaddar[3], Chintan J. Joshi[2], Hratch Baghdassarian[1,2], Isaac Shamie[1,2], Jason Chan[4], Hsuan-Lin Her[1], Eyleen J. O'Rourke[3,5*], Nathan E. Lewis[2,6*]

[1] Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, CA 92093, USA
[2] Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093, USA
[3] Department of Biology, University of Virginia, Charlottesville, VA 22903, USA
[4] Poway High School, Poway, CA 92064, USA
[5] Department of Cell Biology, School of Medicine of University of Virginia, Charlottesville, VA 22903, USA
[6] Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093, USA
*Correspondence: Nathan E. Lewis, nlewisres@ucsd.edu; Eyleen J. O'Rourke, ejo8b@virginia.edu

Abstract

Cell-cell interactions shape cellular function and ultimately organismal phenotype. However, the code embedded in the molecular interactions driving and sustaining the spatial organization of cells remains to be elucidated. Here we present a computational framework to infer the spatial code underlying cell-cell interactions from the transcriptomes of the cell types across the whole body of a multicellular organism. As core of this framework, we introduce our tool *cell2cell*, which uses the coexpression of ligand-receptor pairs to compute the potential for intercellular interactions, and we test it across the *Caenorhabditis elegans*' body. Leveraging a 3D atlas of *C. elegans*' cells, we also implement a genetic algorithm to identify the ligand-receptor pairs most informative of the spatial organization of cells. Validating the spatial code extracted with this strategy, the resulting intercellular distances are negatively correlated with the inferred cell-cell interactions. Furthermore, for selected cell-cell and ligand-receptor pairs, we experimentally confirm the communicatory behavior inferred with *cell2cell* and the genetic algorithm. Thus, our framework helps identify a code that predicts the spatial organization of cells across a whole-animal body.

Keywords

Caenorhabditis elegans / cell-cell interactions / cell-cell communication / ligand-receptor interactions / single-cell RNA-sequencing

Introduction

Cell-cell interactions (CCIs) are fundamental to all facets of multicellular life. They shape cellular differentiation and the functions of tissues and organs, which ultimately influence organismal physiology and behavior. CCIs often take the form of secreted or surface proteins produced by a sender cell (ligands) interacting with their cognate surface proteins in a receiver cell (receptors).The nature of CCIs is constrained by the distance between interacting cells (*1–3*), and, in turn, CCIs follow spatial patterns of interaction (*4*). These patterns are important since they allow CCIs to define cell location and community spatial structure (*3*, *5*). For instance, some molecules mediating CCIs form gradients that serve as a spatial cue for other cells to migrate (*6*, *7*). In addition, co-occurrence of ligands and receptors are strongly defined by their spatial neighborhoods (*8*). Thus, it is reasonable to speculate that there is a spatial code embedded in ligand-receptor (LR) interactions across the body of multicellular organisms; a code that defines the distribution of cells in tissues and organs and enables cells to sense their spatial proximity.

CCIs can be inferred from the gene expression levels of ligands and receptors (*9*). Although spatial information is lost during tissue dissociation in conventional bulk and single-cell RNA-sequencing technologies (scRNA-seq) (*10*), inferring CCIs from transcriptomics can help elucidate how multicellular functions are coordinated by both the molecules mediating CCIs and their spatial context. Indeed, previous studies have proven that gene expression levels still encode spatial information that can be recovered by adding information such as protein-protein interactions and/or microscopy data (*10–13*). For example, RNA-Magnet inferred cellular contacts in the bone marrow by considering the coexpression of adhesion molecules present on cell surfaces (*12*), while ProximID used gene expression coupled with microscopy of cells to construct a spatial map of cell-cell contacts in bone marrow (*11*). Thus, we propose that CCIs inferred from transcriptomics could be extended to assess whether one can find, in RNA, a spatial code of intercellular messages that defines spatial organization and cellular functions across the whole body of a multicellular organism.

Caenorhabditis elegans is an excellent model for studying CCIs in a spatial context across a whole body (*14*). This animal has fewer than 1,000 somatic cells stereotypically arranged across the body, whose locations have been described in a 3D atlas (*15*). Despite the small number of cells, the intercellular organization in *C. elegans* shows complexity comparable to higher-order organisms. Taking advantage of these features, here we use scRNA-seq data from *C. elegans* to compute CCIs and assess which ligand-receptor pairs could govern an intercellular spatial code across the body. For this purpose, single-cell transcriptome data (*16*) were integrated with a 3D-atlas of cells of *C. elegans* (*15*), while we built the most comprehensive list of ligand-receptor interactions in *C. elegans* for CCI analyses. Next, we compared our CCI predictions to literature and found them consistent with previous studies independently reporting relevant roles of the identified LR interactions as encoders of spatial information. Additionally, we experimentally tested uncharacterized CCIs, and validated *in situ* that adjacent cells co-express the LR pairs computationally inferred to contribute to the spatial code. Thus, together, we demonstrate that single-cell RNAseq data can be used to define a genotype-spatial phenotype link for the whole body in a multicellular organism.

Results

Computing cell-cell interactions

To study cell-cell interactions, a first step can be revealing active intercellular communication pathways from the coexpression of the corresponding LR pairs in any particular pair of cells. Communication scores can be assigned to each LR pair based on the RNA expression levels of their encoding genes in a given pair of sender and receiver cells (*17–22*). Communication scores are then aggregated into an overall CCI score for each pair of cells, often represented by the number of active (expressed) LR pairs, wherein the higher the number of active LR pairs, the stronger the cell-cell interaction is assumed to be (*9*, *18*). However, this method disregards that a high CCI score could result just by chance when one of the interacting cells promiscuously expresses many different ligands and/or receptors. In contrast, we propose a novel CCI score that is based on the idea that high CCI scores should represent a high but also specific complementarity in the production of ligands and receptors between the interacting cells (Fig. 1).

Among other biological mechanisms that justify this idea, cells can sense the number of receptors that are occupied by signals from surrounding cells (*3*, *23*). Higher occupancy can indicate greater proximity of communicating cells (*24*). Thus, to represent a cell-cell potential of interaction that may respond to or drive intercellular proximity, we propose a Bray-Curtis-like score computed from the mRNA expression of ligands and receptors in pairs of interacting cells (Fig. 1A). In contrast to the number of active LR pairs, our Bray-Curtis-like score weighs the number of LR pairs that a pair of cells use to communicate by the aggregate total of ligands and receptors expressed by each cell in the pair (see Materials and Methods, and Fig. 1B). The main assumption of our CCI score is that more proximal cells co-express more complementary ligands and receptors between the pair of cells. In other words, for any given pair of cells, cells are defined as closer when a greater fraction of the ligands produced by one cell interacts with cognate receptors on the other cell and vice versa, as this increases their potential of interaction.

To facilitate the implementation of our computational framework to predict a spatial code of CCIs and perform other general CCI analyses that do not rely on spatial information, we developed *cell2cell*. This open source tool infers intercellular interactions and communication using any gene expression matrix and list of LR pairs as inputs (https://github.com/earmingol/cell2cell). Although here we chose our Bray-Curtis-like score for representing a spatial-dependent complementarity of interaction, *cell2cell* is flexible in terms of the scoring strategies applied to decipher CCIs. Thus, depending on the purpose of a study, other CCI scores can be used (e.g., the number of active LR pairs to represent the strength of the interaction).

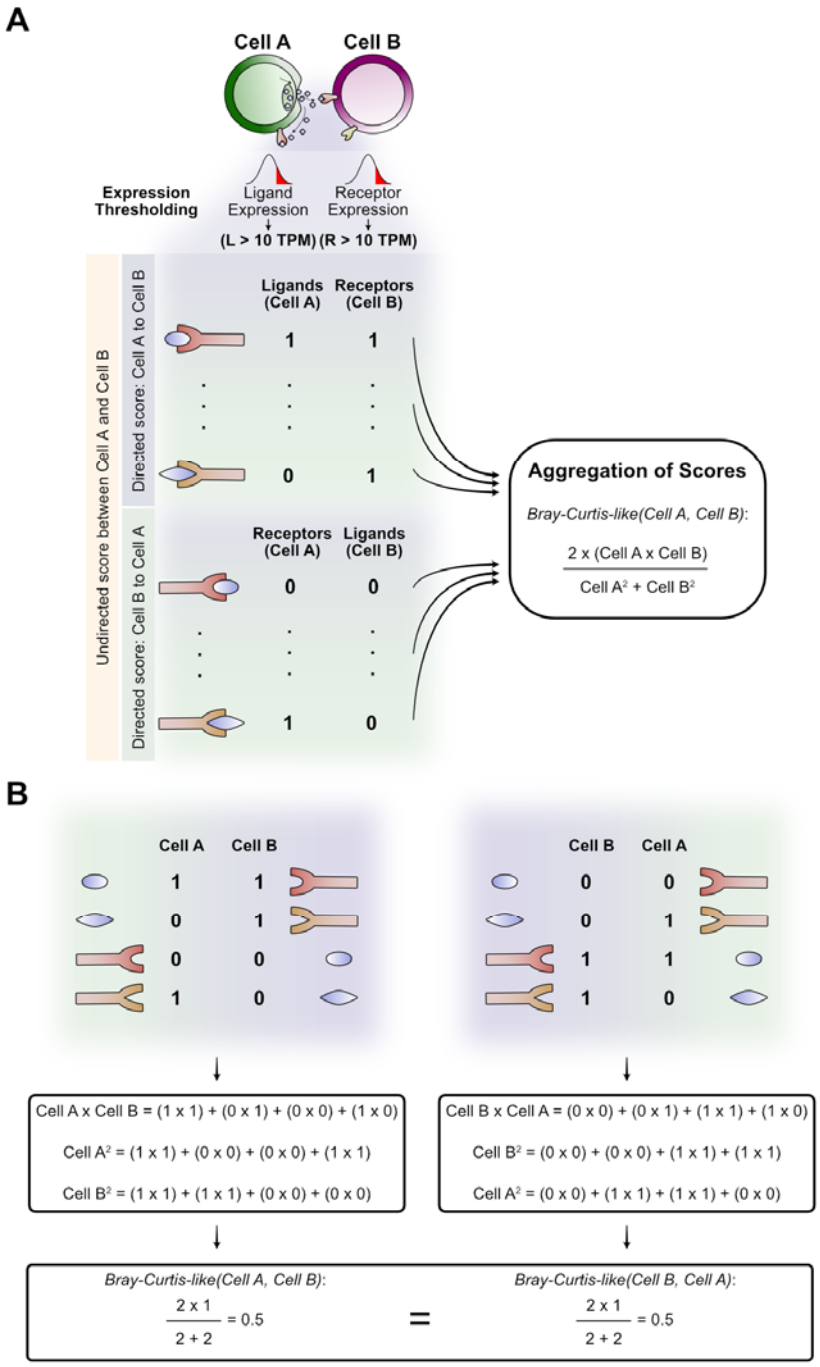


Fig. 1. Calculation of the modified Bray-Curtis CCI score. (**A**) To represent the overall interaction potential between cell A and cell B, our CCI score is computed from two vectors representing the ligands and receptors independently expressed in each cell. If only the ligands from one cell and the cognate receptors on the other are considered ("Cell A to Cell B" half or "Cell B to Cell A" half, independently), the score would be a directed score for representing the interaction (one cell is the sender and the other is the receiver). However, our score is undirected by considering both ligands and receptors of each cell to build the vector (both halves simultaneously, indicated with the yellow rectangle on the left). Thus, the vector of each cell is built with both directed halves of molecule production (e.g., top half possess ligands of cell A while the bottom half considers its receptors, generating a unique vector with both the ligands and the receptors of cell A). (**B**) Toy examples for computing our score for the interaction of Cell A and Cell B. Here, both possible directions of interaction are represented to show that they result in the same (undirected) score.

Cell-type specific properties are captured by computed cell-cell interactions

To assess whether our Bray-Curtis-like score captures the relationship between the expression of ligand-receptor pairs in interacting cells and their intercellular distances, we used *C. elegans* since, among other relevant characteristics, it has a stereotypical distribution of cells that has been extensively studied through microscopy and reported for 357 of its individual cells in a 3D atlas (*15*). However, to compute the complementarity of interaction between *C. elegans* cells, an extensive list of functional LR interactions is needed. While much is known about *C. elegans*, knowledge of its LR interactions remains dispersed across literature or contained in protein-protein interaction (PPI) networks that include other categories of proteins. Thus, we generated a list of 245 ligand-receptor interactions in *C. elegans* (table S1). Next, we used this list to determine the presence or absence of mRNAs encoding ligands and receptors in each cell identified in the single-cell transcriptome of *C. elegans* (*16*). Briefly, this dataset takes a matrix of gene expression data with the aggregated values from all individual cells with the same annotation. Of the 27 cell types identified across the body of *C. elegans*, we considered only the 22 cell types that we were able to assign a spatial location in a previously published 3D atlas (*15*). After integrating this aggregated single-cell transcriptomic data with the list of LR pairs, we inferred the active (expressed) LR pairs in all pairs of cell types by using a binary communication score (table S2). Next, we aggregated the respective communication scores for each cell pair with our Bray-Curtis-like metric, generating the first predicted network of CCIs in *C. elegans* that measures the complementarity of interacting cell types given their active LR pairs (Fig. 2A).

After determining the potential for interaction between every pair of cell types from the single-cell transcriptome of *C. elegans*, we grouped the different cell types based on their interactions with other cells through an agglomerative hierarchical clustering (Fig. 2A). This analysis generated clusters that seem to represent known roles of the defined cell types in their tissues. For instance, neurons have the largest potential for interactions with other cell types, especially with themselves and muscle cells. This suggests that these cell types use a higher fraction of all possible communication pathways, which is consistent with the high molecule interchange that occurs at the neuronal synapses and the neuromuscular junctions (*25*). Also, seemingly in line with basement membranes surrounding germline cells and physically constraining their ability to communicate with other cell types (*26*, *27*), germline cells have the lowest CCI potential with other cell types. Thus, the results suggest that our method may be properly capturing the nature of the interactions between vastly different cell pairs.

We further observed that pairs of interacting cells tend to be grouped by the sender cells (i.e., those expressing the ligands), but not by the receiver cells (i.e., those expressing the receptors) (Fig. 2B). Remarkably, our result is consistent with previous findings that ligands are produced in a cell type-specific manner by human cells, but receptors are promiscuously produced (*28*). While the study used a network-based clustering of ligands and receiver cell connections, we used UMAP (*29*, *30*) to visually summarize the Jaccard similarity (*31*) between pairs of interacting cell types, indicating this similar result from two different approaches could be biologically meaningful. Correspondingly, the coexpression of ligands and their cognate receptors follows a more similar behavior in cell pairs where the sender cells are of the same type, while the receiver cell types can be disregarded (fig. S1).

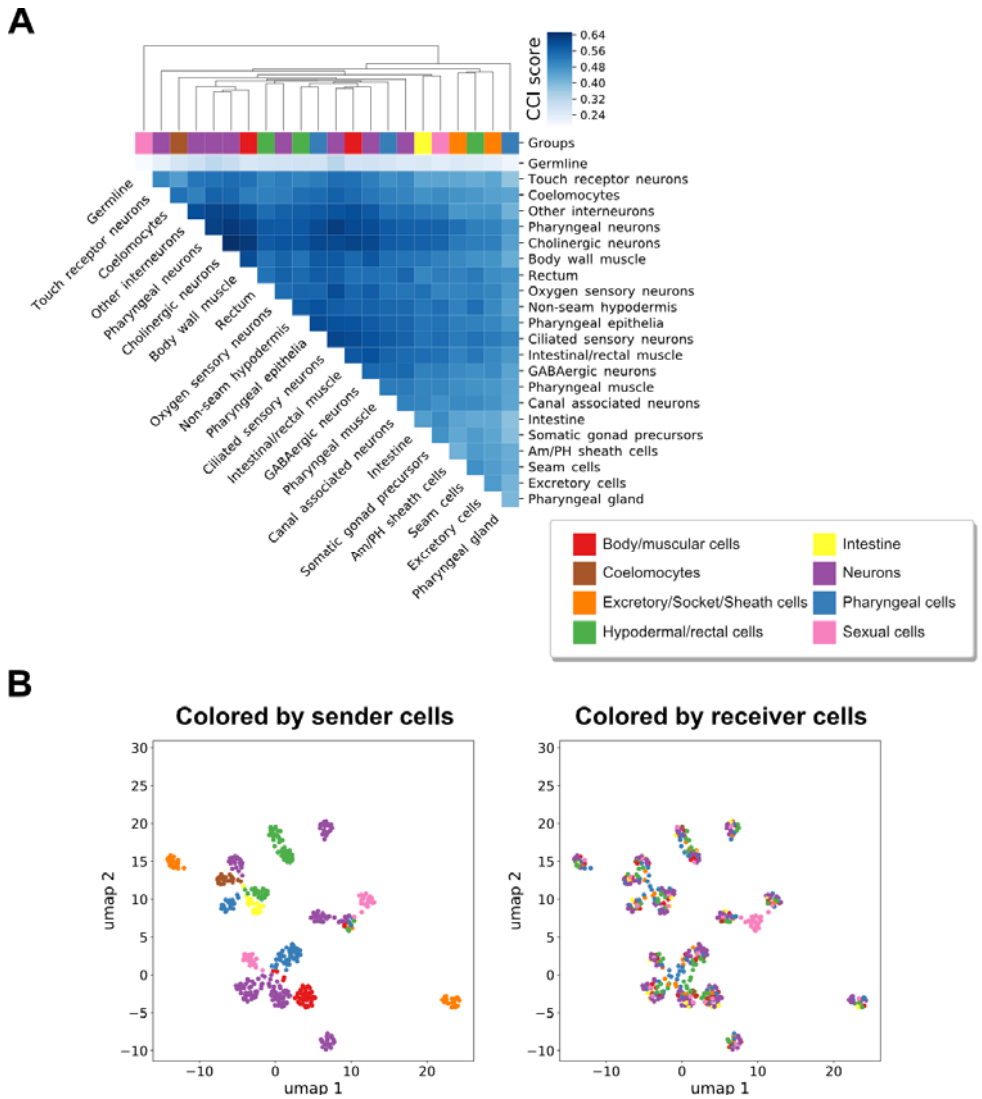


Fig. 2. Cell-cell interactions and communication in *C. elegans*. (**A**) Heatmap of CCI scores obtained for each pair of cell types using the curated list of LR pairs. An agglomerative hierarchical clustering was performed on a dissimilarity-like metric by taking the complement (1-score) of CCI scores, disregarding autocrine interactions. Cell types are colored by their lineages as indicated in the legend. Lineages and colors were assigned previously (*16*). (**B**) UMAP visualization of CCIs. Dots represent pairs of interacting cells and they were projected based on their Jaccard distances, which were computed from the LR pairs expressed in the directed interactions between cells (one cell is producing the ligands and the other the receptors). Dots are colored by either the sender cell (left) or the receiver cell (right), depending on their lineages as indicated in the legend of (A). A readable version of the data used for this projection is available in table S2, where names of LR pairs and their communication scores are specified for each cell pair.

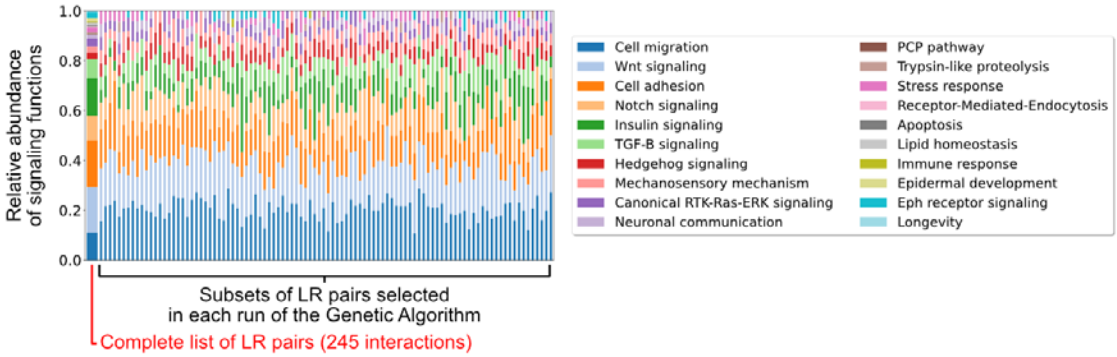
Signaling pathways involved in spatial patterning underlie the anticorrelation between cell distance and interaction potential

Under the assumption that larger physical distances would decrease the potential of cells to interact, we evaluated the relationship between our undirected CCI score and the Euclidean distance between cell pairs (Supplementary Notes). As expected, the correlation coefficient was negative (Spearman = -0.21; P-value = 0.0016). However, although negative, the anticorrelation is weak. Therefore, we hypothesized that there is a subset of

key LR pairs encoding spatial organization. To identify this LR subset, we used a genetic algorithm (GA), that is based on natural evolution, to select a subset of our initial list of LR pairs that maximizes the anticorrelation between the CCI scores and the Euclidean distances (Supplementary Notes). Using this approach, we found 100 candidate subsets from independent runs that led to an average Spearman coefficient of -0.67 ± 0.01. The genetic algorithm-optimized subsets of LR pairs (hereinafter referred to as 'initial GA-LR pairs') may therefore constitute good predictors of biological functions driving or sustaining intercellular proximity.

Using our literature-based functional annotations (see column "LR Function" in table S1), we next investigated the specific biological roles of the initial GA-LR pairs. Specifically, we computed the relative abundance of each functional annotation within each initial GA-LR pair (Fig. 3A). Considering the relative abundances in our complete list containing the 245 LR pairs as the expected values, we performed a one-sided Wilcoxon signed-rank test to evaluate whether the relative abundance of each signaling function either increased or decreased across all GA runs (Fig. 3B). Remarkably, LR pairs involved in cell migration, Hedgehog signaling, mechanosensory mechanisms, and canonical RTK-Ras-ERK signaling increased their relative abundance in the resulting subsets from the GA runs. Thus, the GA prioritizes LR pairs associated with processes such as cell patterning, morphogenesis, and tissue maintenance (*32*).

A



B

Signaling function	Change	Adj. P-value	Signaling function	Change	Adj. P-value
Cell migration	↑	3.983×10^{-17}	PCP pathway	↓	3.048×10^{-23}
Wnt signaling	–	7.221×10^{-01}	Trypsin-like proteolysis	↓	2.478×10^{-10}
Cell adhesion	↓	2.711×10^{-10}	Stress response	–	2.684×10^{-02}
Notch signaling	–	9.341×10^{-01}	Receptor-Mediated-Endocytosis	↓	4.755×10^{-15}
Insulin signaling	↓	8.158×10^{-13}	Apoptosis	↓	3.048×10^{-23}
TGF-B signaling	–	4.759×10^{-02}	Lipid homeostasis	↓	3.048×10^{-23}
Hedgehog signaling	↑	5.439×10^{-11}	Immune response	↓	5.507×10^{-18}
Mechanosensory mechanism	↑	8.074×10^{-17}	Epidermal development	↓	3.048×10^{-23}
Canonical RTK-Ras-ERK signaling	↑	3.724×10^{-05}	Eph receptor signaling	↓	4.472×10^{-13}
Neuronal communication	–	4.431×10^{-01}	Longevity	↓	3.048×10^{-23}

Fig. 3. Relative abundances of signaling functions across initial GA-LR pairs. (**A**) Composition plot given the signaling functions that LR pairs are associated with. Relative abundances are shown for the complete list of LR pairs (containing 245 interactions) and the subsets obtained in each of the 100 runs of the genetic algorithm (GA). Here, relative abundance is the number of LR pairs involved in a given pathway with respect to the total number of LR pairs in the list. Signaling functions are colored according to the legend. (**B**) Summary of changes for each of the signaling functions given their relative abundance across

the runs of the GA with respect to the relative abundances in the complete list of LR pairs. A one-tailed Wilcoxon's test was performed to evaluate their increase or decrease, as indicated with the arrows. An adjusted P-value is reported (FDR < 1%). For those cases without a significant change (gray rectangles), the smallest adjusted P-value is reported among the increase or decrease.

Considering that the GA is a non-deterministic approach, different optimal solutions can be obtained from independent runs. Thus, we next looked for a consensus set of LR pairs among all optimal solutions generated by our GA (Supplementary Notes). This resulted in a list of 37 LR pairs (table S3), hereinafter referred to as GA-LR pairs, yielding a Spearman coefficient of -0.63 (P-value = 2.629×10^{-27}) between the CCI scores and Euclidean distances; a correlation that is not a result of randomly selecting LR pairs (P-value = 0.0002, permutation tests detailed in Supplementary Notes). While the CCI scores computed from the complete list of LR pairs led to functional interactions of cell types (Fig. 2A), the GA-LR pairs seem to group cell types by more specific associations that may be attributable to their spatial localization (Fig. 4A). For example, the complete list grouped neurons and muscles together, while the GA-LR pairs increased the specificity of this association by grouping both excitatory and inhibitory neurons (cholinergic and GABAergic neurons, respectively) directly with all muscles. Furthermore, these GA-LR pairs group all cells composing the pharynx (pharyngeal gland, epithelia, muscle and neurons) together. Another interesting observation was the high CCI score between oxygen sensing neurons and intestinal cells, consistent with the extensive communication between these cells to link oxygen availability with nutrient status (*33–35*). Thus, the LR interactions prioritized by the GA capture cellular properties that define not only intercellular proximity, but more importantly, cell-community structure of tissues and organs. Moreover, some participating ligands and receptors are expressed in few cell types while others are found in most cell types (Fig. 4B), suggesting that our algorithm captures both communication between specific pairs of cells and more promiscuous interactions. Thus, we hypothesized these PPIs represent a spatial code, that when used in different combinations can encode different spatial proximities.

The GA-LR pairs define a spatial code of intercellular interactions along the body

To start assessing the hypothesis that the GA-LR pairs represent a spatial code, we performed an enrichment analysis of CCI mediators along the body of *C. elegans*. We first divided the *C. elegans* body in 3 sections, encompassing different cell types (Fig. 5A). We then computed all pairwise CCIs within each section and counted the number of times that each LR pair was used. With this number, we performed a Fisher's exact test on each bin for a given LR interaction. We observed enrichment or depletion of specific LR pairs in different parts of the body (Fig. 5B). Interestingly, we observed LR pairs enriched only in one section and depleted in the others and vice versa (Table 1), following a pattern mostly congruent with existing experimental data (table S4). For instance, *col-99* shows prominent expression in the head, especially during L1-L2 larvae stages of development (*36*), while LIN-44 is secreted by hypodermal cells exclusively in the tail during larval development (*37*, *38*), both cases coinciding with the results in Table 1. Although a few spurious results emerged, they are mainly associated with the limitations of the current scRNAseq methods and their analysis tools (see Discussion section). Therefore, the *col-99* and *lin-44* examples support the notion that our strategy captures the spatial distribution of gene expression and therefore of CCIs across the *C. elegans* body.

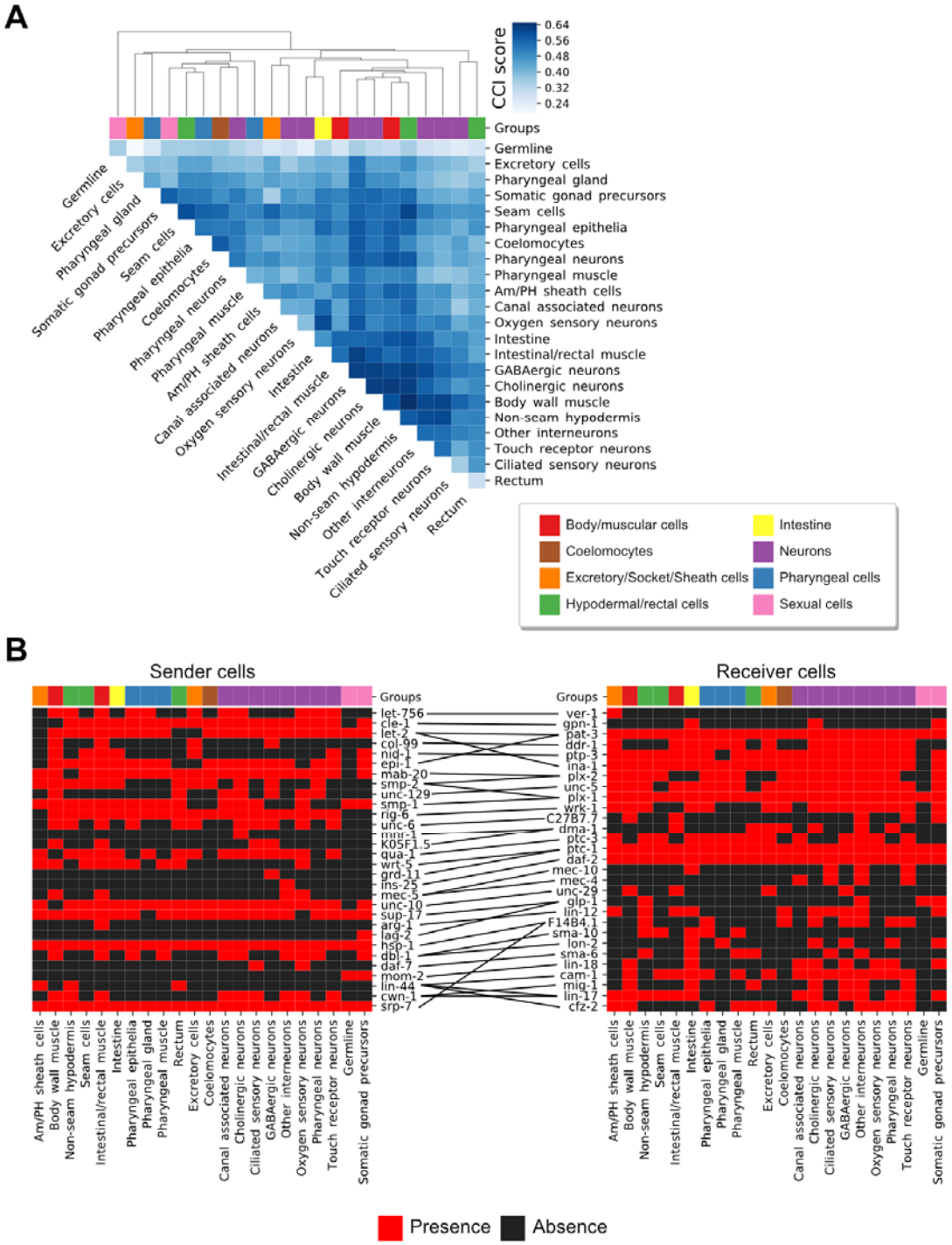


Fig. 4. CCI analyses based on LR pairs associated with intercellular distances. (**A**) Heatmap of CCI scores obtained for each pair of cells using the consensus GA-LR pairs. An agglomerative hierarchical clustering was performed on a dissimilarity-like metric by taking the complement of CCI scores (1-score), excluding autocrine interactions. Cell types are colored by their lineage groups as indicated. (**B**) Heatmaps representing the presence or absence of ligands (left) and receptors (right) after expression thresholding (>10 TPM) in sender and receiver cells, respectively. Lines at the center connect ligands with their cognate receptors according to the GA-selected interactions. Cell types are colored as in (A).

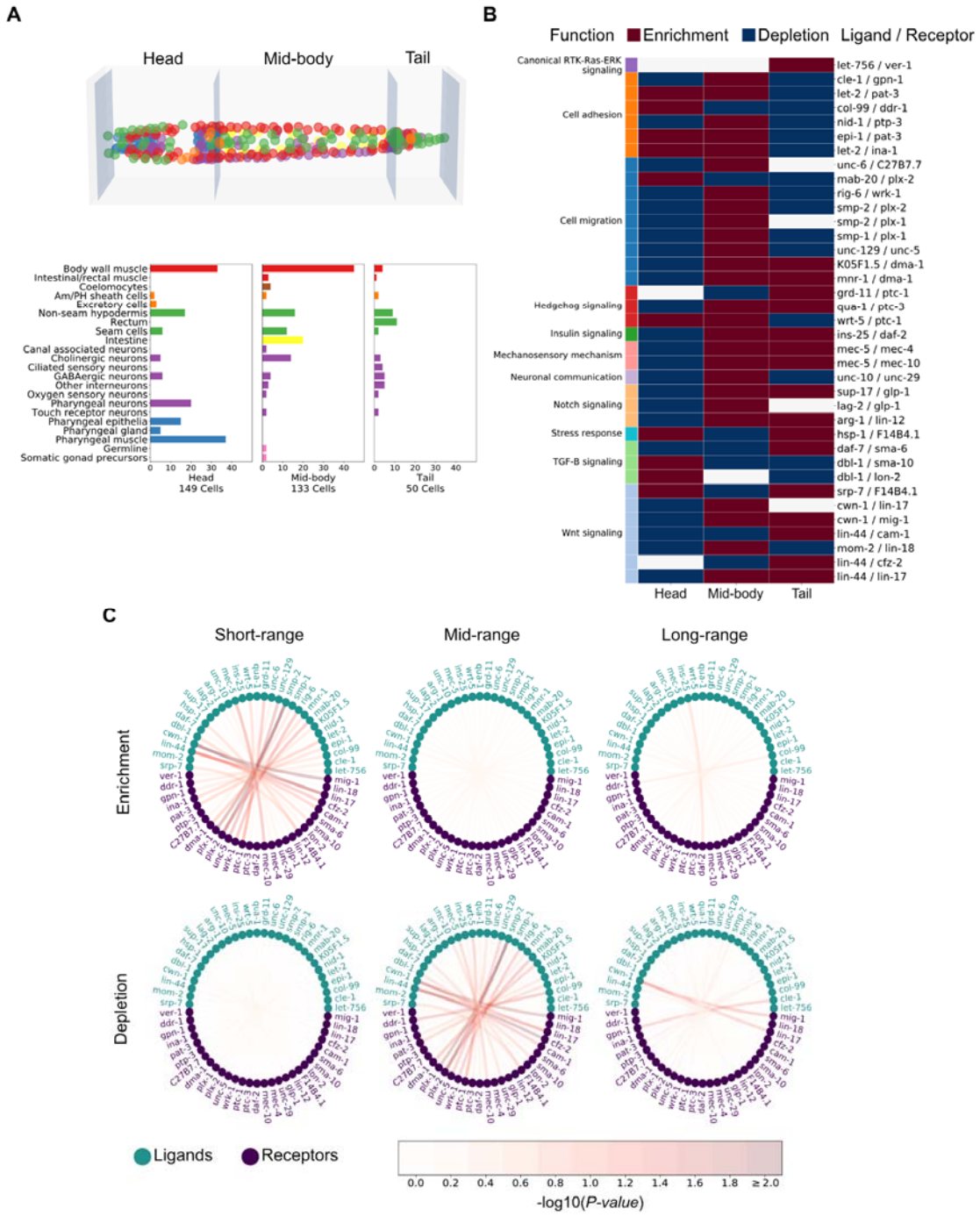


Fig. 5. Spatial enrichment and depletion of communication pathways. (**A**) To study the anteroposterior use of communication pathways, the body of *C. elegans* was divided into three sections along the anteroposterior axis (top) and cell-type composition of each section (bottom) given a previously published 3D atlas. The mid-body section is defined by the presence of the intestine cells, and the head and tail are the anterior and posterior sections to it, respectively. Cells in the 3D atlas (top) are colored according to the cell types as delineated in barplots (y-axis, bottom). (**B**) Enrichment/depletion (FDR < 1%) of ligand-receptor pairs (y-axis) in each of the three sections (x-axis), calculated from their usage across all pairs of cells of each section. Communication pathways are also colored by their annotated functions (left column). (**C**) Circos plots for representing the importance of cell-cell communication occurring at different ranges of distance. A Fisher exact test was performed to find enriched/depleted LR pairs among all pairs of cells for a given proximity. The ranges of distances were defined as explained in fig. N1C. Nodes represent ligands or receptors and edges connect those ligands and receptors that interact in the GA-LR pairs (table S3). The color of the nodes represent whether they are ligands or receptors and the color of the edges indicate the negative value of the logarithmic transformation on the Benjamini-Hochberg adjusted P-values, according to the colored bar at the bottom. Interactions that resulted significantly enriched or depleted (FDR < 1%) are equivalent to the color assigned to a value of 2.0 or bigger.

To better understand the importance of the GA-LR pairs in identifying spatially-constrained CCIs, we searched for LR pairs enriched or depleted across all cell pair interactions in any of the different distance-ranges of communication. We found five LR pairs that were either enriched or depleted in at least one of the three distance ranges given the corresponding pairs of cell types (FDR < 1%) (Fig. 5C). Three of these LR pairs are associated with Wnt signaling (*lin-44*/*cfz-2*, *cwn-1*/*lin-17* and *cwn-1*/*mig-1*) and the other two with cell migration (*smp-2*/*plx-1* and *smp-2*/*plx-2*). Members of the Wnt signaling act as a source of positional information for cells (*3*). For example, in *C. elegans*, *cwn-1* and *lin-44* follow a gradient along its body, enabling cell migration (*39–42*). While semaphorins (encoded by *smp-1*, *smp-2* and *mab-20*) and their receptors (plexins, encoded by *plx-1* and *plx-2*) can control cell-cell contact formation (*43*), and their mutants show cell positioning defects, especially along the anterior/posterior axis of *C. elegans* (*44*, *45*), affecting axon guidance, cell migration (*46*), epidermal and vulval morphogenesis (*47*, *48*). Thus, the GA-LR pairs may influence local or longer-range interactions and help encode intercellular proximity.

Our hypothesis that key LR pairs encode spatial CCI information also implies the assumption that cell-type localization is crucial for organismal phenotypes and functions. Thus, we performed a phenotype enrichment analysis for *C. elegans* (*49*) using the GA-LR genes (sampled genes) and the complete LR pair list (background genes). The 'organ system phenotype' was the only enriched term, with odds ratio of 4.13 (Fisher's exact test; adj. P-value = 0.0029). According to WormBase (*50*), this term represents a generalization for phenotypes affecting the morphology of organs, consistent with the clustering of cell types by their tissue lineage groups when considering genes associated with this phenotype (fig. S2). Thus, our GA-LR pairs seem to encode a relationship between CCIs and organ organization across the body, which in higher organisms could have an impact leading to diseases when perturbed.

GA-LR pairs are proximally expressed in *C. elegans*

So far, our computational framework seems to be able to identify LR interactions driving the spatial organization of cell-cell interactions. Therefore, based on the precedents presented here, especially the strong anticorrelation between our CCI score and the intercellular distance, we expected that the ligand and the receptor in each GA-LR pair are expressed in proximal cells. To test whether the LR pairs selected by our algorithm are actually co-expressed in proximal cell pairs, we searched the literature for established interactions in addition to experimentally testing new CCIs. We found several LR pairs with known expression patterns in *C. elegans* that coincide with the predictions of our algorithm (reported elsewhere; and summarized in table S4). Furthermore, we used single-molecule Fluorescent *In Situ* Hybridization (smFISH) to test whether previously uncharacterized LR pairs are co-expressed in adjacent cells as predicted by our model. Specifically, we focused on the uncharacterized interactions between *arg-1*/*lin-12*, *let-756*/*ver-1* and *lin-44*/*lin-17* (see Materials and Methods for selection criteria).

Confirming the predictions of our algorithm, we found the ligand and receptor genes expressed in spatially proximal cells. In particular, our *in situ* results confirmed our computational prediction that *arg-1* and *lin-12* are proximally expressed in the intestinal/rectal muscle and the non-seam hypodermal cells (Fig. 6A). We also confirmed that *let-756* is expressed in the non-seam hypodermal cells of the head, proximally to *ver-1* in the amphid sheath cells (Fig. 6B). Finally, *lin-44* in the non-seam hypodermal cells of the tail is proximally expressed to *lin-17* in the seam cells of the tail (Fig. 6C).

Additionally, we performed 3D projections of the smFISH images (movies S1-3), which more clearly show the extent of the spatial adjacency of the cells expressing the cognate LR pairs. We also noticed that while *arg-1*/*lin-12* and *let-756*/*ver-1* were expressed in cells that were juxtaposed, *lin-44*/*lin-17* were expressed in proximal but not necessarily juxtaposed cells. Interestingly, the ligands of the pairs expressed in juxtaposed cells are known to encode membrane bound proteins, while LIN-44 is a secreted ligand (*51*) to surrounding areas. Therefore, the experimental observations are consistent with our inferences for the respective pairs of cells, which presented CCI scores among the highest values, being of 0.61, 0.58 and 0.64, respectively (Fig. 4A). Thus, the results not only further support the notion that higher Bray-Curtis-like scores represent a higher potential of cells to be spatially proximal, but they also show that applying our computational framework is a hypothesis generator of unknown biology.

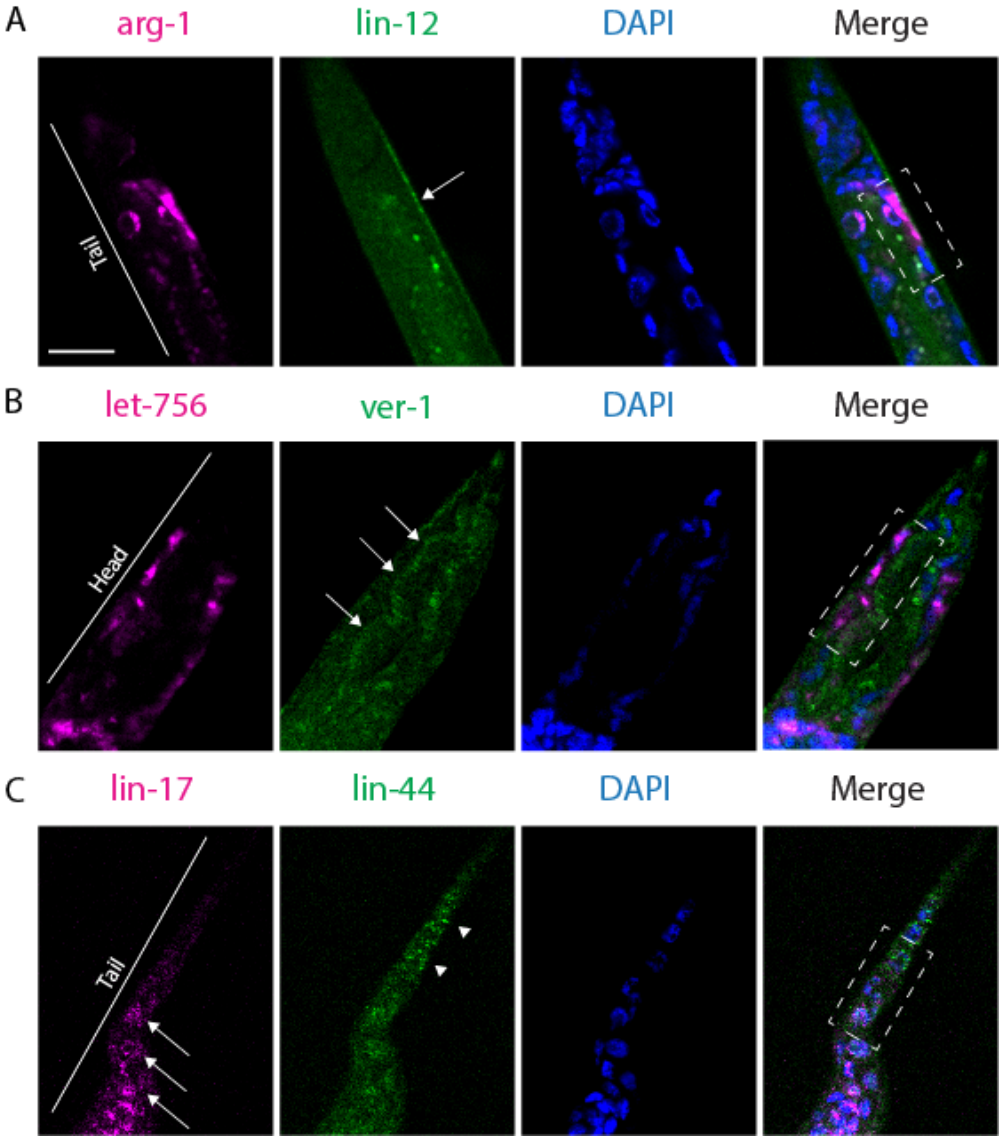


Fig. 6. Validation of the spatial expression of specific GA-LR pairs. Single-molecule Fluorescent In Situ Hybridization of genes encoding three GA-LR pairs in *C. elegans* L2 larvae. (**A**) Intestinal/rectal cells expressing *arg-1* (magenta) and non-seam hypodermal cells (arrow) expressing *lin-12* (green) are adjacent (see rectangle in the merge channel and movie S1). (**B**) Non-seam hypodermal cells expressing *let-756* (magenta) and amphid sheath cells (arrows) expressing *ver-1* (green). Amphid sheath cells are surrounded by hypodermal cells (see rectangle in the merge channel and movie S2). (**C**) Seam cells in the tail (arrows) expressing *lin-17* (magenta) and non-seam hypodermal cells in the tail (arrowheads) expressing *lin-44*

(green). The two genes are expressed in proximal cells (see ends of rectangle in the merge channel and movie S3). In all cases a DAPI staining was performed to distinguish cell nuclei. Scale bar = 10µm.

Discussion

Here we present a computational strategy to quantify the potential of cells to interact and communicate, which we named *cell2cell*. Using scRNA-seq data and a list of validated and predicted LR pairs, this tool uses the gene expression level of the ligands and the receptors in all cells in the dataset to infer how they communicate. Furthermore, we implemented in *cell2cell* a new scoring function to compute an overall potential of cells to interact, the Bray-Curtis-like CCI score. By applying this approach to infer interactions in *C. elegans* we identified pairs of cells that showed a potential to interact as measured by our Bray-Curtis-like CCI score, wherein there is a negative correlation with intercellular distance (negative Spearman coefficient). Thus, our computational framework is useful for associating phenotypes with CCIs in a space-dependent fashion.

By using a genetic algorithm to search for a combination of fewer LR pairs (as measured with our Bray-Curtis-like score), we enhanced the negative correlation and found a consensus subset of 37 LR pairs that decreased the Spearman coefficient from -0.21 to -0.63. Importantly, this further indicates that specific LR pairs can encode spatial organization of cells in *C. elegans*, supporting the notion that the GA-LR pairs impact organismal-level phenotypes. Furthermore, we compared our Bray-Curtis-like score to the number of active LR pairs (or LR count score), which is the most employed overall CCI score. In this analysis, we observed that our score resulted in significantly better correlation than the LR count score across all separate runs of the GA (fig. S3A). However, the consensus LR lists in both cases resulted to be biologically comparable (fig. S3B), strengthening the evidence that a spatial code across the whole body is dependent on specific LR interactions. Moreover, the consensus list found with our CCI score includes most of the ligands and receptors contained in the consensus list found with the LR count score, while also resulting in a larger number of LR that could be useful for new hypothesis generation tested through experimental validation. For example, in our case, the *arg-1*/*lin-12* pair was exclusive of the GA-LR list obtained with our Bray-Curtis-like score, and we experimentally demonstrated that this pair is co-expressed by spatially proximal cells (Fig. 6A).

To apply *cell2cell* in *C. elegans*, we needed a database of known LR interactions in this organism, which had not yet been reported. Hence, we collected LR interactions of *C. elegans* from the literature and databases of PPIs to build the most comprehensive database of LR interactions in *C. elegans* (table S1). We anticipate this will be a valuable resource and hypothesis generator for the study of CCIs in *C. elegans*, either at a spatial or functional level. In this regard, our CCI analysis based on this database identified a core set of LR pairs associated with spatial patterning in *C. elegans*. For example, we found that interacting cells were grouped based on cell type-specific production of ligands (Fig. 2B), which is consistent with the principles underlying a communication network reported for human haematopoietic cells (*28*). Our results are also consistent with previous experimental studies of *C. elegans* (table S4). For instance, the GA-driven selection of LR pairs prioritized mediators with a role in cell migration, Hedgehog signaling, mechanosensory mechanisms and canonical RTK-Ras-ERK signaling (Fig. 3B). These GA-LR pairs also included LR interactions that are crucial for the larval development of *C. elegans*, especially of processes driven by Notch and TGF-β signaling, as well as cellular positioning, and organ morphogenesis, which are particularly active at the cognate

stages of the datasets we used (*52–55*). Thus, the GA-selected LR pairs are enriched in processes that contribute to defining the spatial properties of tissues and organs. Furthermore, the GA-LR pairs more likely act in short distance interactions (Fig. 5C) and in specific body regions (Fig. 5B). Therefore, the genetic algorithm prioritized a core list of LR pairs whose active/inactive combination seems to define a cellular spatial code across the *C. elegans* body.

Importantly, the LR pairs selected by the GA can affect different phenotypes that are related among them in *C. elegans*, suggesting that new biology can be inferred when including less-studied LR pairs. For instance, the GA-LR pairs enriched in short distance interactions (Fig. 5C) include: 1) The LR pair composed of *smp-2*/*plx-1*, which mediates epidermal morphogenesis, as demonstrated by the defects in epidermal functions exhibited by *C. elegans* lacking *smp-2* (*44*); and 2) *cwn-1*/*mig-1*, which mediates cell positioning, as demonstrated by the abnormal migration of hermaphrodite specific motor neurons in the mutants (*41*, *56*). Additionally, by using smFISH we experimentally showed that mediators used by the cell pairs with high CCI scores (Fig. 4A), such as *arg-1*/*lin-12*, *let-756/ver-1* and *lin-44/lin-12*, are expressed in spatially adjacent or proximal cells (Fig. 6). While previous studies reported that hypodermal cells form a gradient of LIN-44 in the tail (*57*), and that LIN-44 can affect seam cell polarity through LIN-17 (*58*), the spatial proximity necessary for this LR pair to mediate a CCI had not been shown before. Although smFISH is not a direct proof of CCIs, mRNA co-expression serves as a good proxy for experimentally supporting those CCIs (*9*, *59*). *lin-44/lin-17* is proximally co-expressed in tail hypodermal and seam cells (Fig. 6C and movie s3). Thus, the previous reports and the smFISH results show congruence with the predictions of our algorithm. This not only increases the confidence in our approach and results, but it also exposes the potential of our computational framework to uncoverLR interactions that were not previously studied in specific cell types.

Overall, our strategy captures mechanisms underlying the spatial and functional organization of cells in a manner that is consistent with prior and new experimental evidence (table S4 and Fig. 6). Nevertheless, our approach has some limitations. Conventional scRNA-seq does not preserve spatial information, so labeling cells in a 3D atlas by using cell types as annotated in a transcriptomic dataset might be a confounder. For example, *C. elegans* possesses sub-types of non-seam hypodermal cells, and their gene expression varies depending on their antero/posterior location. However, the scRNA-seq data set employed here pooled all non-seam hypodermal cell subtypes as one cell type, artificially generating a generic hypodermal seam cell with a uniform gene expression profile across the body. An illustrative case where this impacted our predictions is the expression of *lin-44*, which is exclusively expressed in hypodermal cells of the tail (Fig. 6C) (*39*, *42*), but our method inferred that *lin-44* was also important in the mid-body (Fig. 5B, pair *lin-44*/*lin-17*). Similarly, *daf-7* is expressed only in sensory neurons in the head (*60*); however, our results show an enrichment in the tail (Table 1). This discrepancy is likely due to pooling the transcriptome of the two types of sensory neurons that express *daf-7*. Similarly, *ver-1* is expressed by amphid and phasmid sheath cells, which are located in the head and tail, respectively; however, these cells are annotated as the same cell type in the transcriptome: amphid/phasmid (Am/PH) sheath. Thus, the labeling of both groups of cells as Am/PH sheath cells could explain an enrichment of the *let-756/ver-1* interaction only in the tail (Fig. 5B) even though it is an important communication also happening in the head (Fig. 6). Therefore, relying only on conventional scRNA-seq enables us to infer the LR interactions that a pair of cells can theoretically use but may not actually use. These limitations of scRNA-seq may also explain the strong but imperfect correlation

obtained between CCI scores and intercellular distances, which may be evaluated by spatial transcriptomics in future studies of CCIs (*61*, *62*).

In summary, our computational framework combines the use of *cell2cell* and a GA to find a combination of LR pairs mediating overall CCIs that best correlates with the intercellular distances. As shown in this work, when considering spatial information, our approach is capable of recovering spatial properties lost in the traditional transcriptomics methods, either bulk or single cell, which is important since these technologies are easier to access than the technologies preserving spatial properties (e.g., spatial transcriptomics). Also, as long as a pertinent objective function can be defined for the GA, our strategy can be used to identify LR pairs associated with phenotypes of interest. Thus, our strategy provides a framework for unraveling the molecular and spatial features of cell-cell interactions and communication across a whole animal body, and potentially their phenotypic consequences. Finally, our approach can be extended to study the role of CCIs in physiological and diseased states in higher organisms.

Table 1. Ligand-receptor interactions enriched or depleted in one body section and depleted or enriched in the rest.

Interactions enriched in a body section and depleted in the rest			**Interactions depleted in a body section and enriched in the rest**		
Ligand	**Receptor**	**Section**	**Ligand**	**Receptor**	**Section**
col-99	*ddr-1*	Head	*K05F1.5*	*dma-1*	Head
mab-20	*plx-2*	Head	*mnr-1*	*dma-1*	Head
dbl-1	*sma-10*	Head	*qua-1*	*ptc-3*	Head
cle-1	*gpn-1*	Mid-Body	*ins-25*	*daf-2*	Head
nid-1	*ptp-3*	Mid-Body	*mec-5*	*mec-4*	Head
rig-6	*wrk-1*	Mid-Body	*mec-5*	*mec-10*	Head
smp-2	*plx-2*	Mid-Body	*sup-17*	*glp-1*	Head
smp-1	*plx-1*	Mid-Body	*arg-1*	*lin-12*	Head
unc-129	*unc-5*	Mid-Body	*cwn-1*	*mig-1*	Head
unc-10	*unc-29*	Mid-Body	*lin-44*	*lin-17*	Head
mom-2	*lin-18*	Mid-Body	*hsp-1*	*F14B4.1*	Mid-Body
daf-7	*sma-6*	Tail*	*srp-7*	*F14B4.1*	Mid-Body
lin-44	*cam-1*	Tail	*let-2*	*pat-3*	Tail
			epi-1	*pat-3*	Tail
			let-2	*ina-1*	Tail
			wrt-5	*ptc-1*	Tail

* See the discussion section for details about this prediction.

Materials and Methods

Single-cell RNA-seq data

A previously published transcriptome of 27 cell types of *C. elegans* in the larval L2 stage was used (*16*). The cell types in this dataset belong to different kinds of neurons, sexual cells, muscles and organs such as the pharynx and intestine. We used the published preprocessed gene expression matrix for cell-types provided previously (*16*), wherein the values are transcripts per million (TPM).

Intercellular distances of cell types

A 3D digital atlas of cells in *C. elegans* in the larval L1 stage, encompassing the location of 357 nuclei, was used for spatial analyses of the respective cell types (*15*). Each of the nuclei in this atlas was assigned a label according to the cell types present in the transcriptomics dataset, which resulted in a total of 322 nuclei with a label and therefore a transcriptome. To compute the Euclidean distance between a pair of cell types, all nuclei of each cell type were used to compute the distance between all element pairs (one in each cell type). Then, the minimal distance among all pairs is used as the distance between the two cell types (fig. N1A in Supplementary Notes). In this step, it is important to consider that this map is for the L1 stage, while the transcriptome is for the L2 stage. However, we should not expect major differences in the reference location of cells between both stages.

Generating a list of ligand-receptor interaction pairs

To build the list of ligand-receptor pairs of *C. elegans*, a previously published database of 2,422 human pairs (*18*) was used as reference for looking for respective orthologs in *C. elegans*. The search for orthologs was done using OrthoDB (*63*), OrthoList (*64*) and gProfiler (*65*). Then, a network of protein-protein interactions for *C. elegans* was obtained from RSPGM (*66*) and high-confidence interactions in STRING-db (confidence score > 700 and supported at least by one experimental evidence) (*67*). Ligand-receptor pairs were selected if a protein of each interaction was in the list of ortholog ligands and the other was in the list of ortholog receptors. Additionally, ligands and receptors mentioned in the literature were also considered (table S5). Finally, a manual curation as well as a functional annotation according to previous studies were performed, leading to our final list of 245 annotated ligand-receptor interactions, encompassing 127 ligands and 66 receptors (table S1).

Communication and CCI scores

To detect active communication pathways and to compute CCI scores between cell pairs, first it was necessary to infer the presence or absence of each ligand and receptor. To do so, we used an expression threshold over 10 TPM as previously described (*18*). Thus, those ligands and receptors that passed this filter were considered as expressed (a binary value of one was assigned). Then, a communication score of one was assigned to each ligand-receptor pair with both partners expressed; otherwise a communication score of 0 was assigned. To compute the CCI scores, a vector for each cell in a pair of cells was generated using their communication scores as indicated in Fig. 1. These vectors containing the scores were aggregated into a Bray-Curtis-like score to represent the potential of interaction. This potential aims to measure how complementary are the signals that interacting cells produce. To do so, our Bray-Curtis-like score considers the number of active LR pairs that a pair of cells has while also incorporating the potential that each cell has to communicate independently (Fig. 1). In other words, this score normalizes the

number of active LR pairs used by a pair of cells by the total number of ligands and receptors that each cell expresses independently. Unlike other CCI scores that represent a directed relationship of cells by considering, for instance, only the number of ligands produced by one cell and the receptors of another, our CCI score is also undirected. To make our score undirected, it includes all ligands and receptors in cell A, and all cognate receptors and ligands, respectively, in cell B (Fig. 1). Thus, pairs of cells interacting through all their ligands and receptors are represented by a value of 1 while those using none of them are assigned a value of 0.

Genetic algorithm for selecting ligand-receptor pairs that maximize correlation between physical distances and CCI scores

An optimal correlation between intercellular distances and CCI scores was sought through a genetic algorithm (GA). This algorithm used as an objective function the absolute value of the Spearman correlation, computed after passing a list of ligand-receptor pairs to compute the CCI scores. In this case, only non-autocrine interactions were used (elements of the diagonal of the matrix with CCI scores were set to 0). The absolute value was considered because it could result either in a positive or negative correlation. A positive correlation would indicate that the ligand-receptor pairs used as inputs are preferably used by cells that are not close, while a negative value would indicate the opposite. The GA generated random subsets of the curated list of ligand-receptor pairs and used them as inputs to evaluate the objective function (as indicated in fig. N2A in Supplementary Notes). The maximization process was run 100 times, generating 100 different lists that resulted in an optimal correlation. As shown in Supplementary Notes (fig. N2C-D), a selection of the consensus ligand-receptor pairs was done according to their co-occurrence across the 100 runs of the GA and presence in most of the runs.

Defining short-, mid- and long-range communication

The physical distances between all pairs of cell types in *C. elegans'* body were classified into different ranges of distances used for CCIs (short-, mid- or long-range distance) by using a Gaussian mixed model. This model was implemented using the scikit-learn library for Python *(68)* and a number of components equal to 3.

Statistical analyses

For each function annotated in the list of ligand-receptor pairs (table S1), a one-sample Wilcoxon signed rank test was used to evaluate whether the relative abundance increased or decreased with respect to the distribution generated with the GA runs. In this case, two one-tail analyses were performed for each function, one to evaluate an increase of the relative abundance and the other to assess a decrease. Finally, the smallest P-value was considered and the respective change was assigned if the adjusted P-value passed the threshold.

A permutation analysis was done on the list of consensus ligand-receptor pairs obtained from the GA. To do so, three scenarios were considered: (1) a column-wise permutation (one column is for the ligands and the other for the receptors); (2) a label permutation (run independently on the ligands and the receptors); and (3) a random subsampling from the original list, generating multiple subsets with similar size to the consensus list. In each of these scenarios, the list of ligand-receptor interactions was permuted 9,999 times.

All enrichment analyses in this work corresponded to a Fisher exact test. In all cases a P-value was obtained for assessing the enrichment and another for the depletion. The

analysis of enriched ligand-receptor pairs along the body of *C. elegans* (head, mid-body and tail) was performed by considering all pairs of cells in each section and evaluating the number of those interactions that use each of the ligand-receptor pairs. The total number of pairs corresponded to the sum of cell pairs in all sections of the body. Similarly, the enrichment analysis performed for the different ranges of distance (short-, mid- and long-ranges) was done by considering all cell pairs in each range and the total number of pairs was the sum of the pairs in each range. To evaluate the enrichment of phenotypes (obtained through the tissue enrichment tool for *C. elegans (49)*), all genes in the GA-selected list were used as background. Then, the genes associated with the respective phenotype tested were used to assess the enrichment.

When necessary, P-values were adjusted using Benjamini-Hochberg's procedure. In those cases, a significance threshold was set as FDR < 1% (or adj. P-value < 0.01).

Selection of ligand-receptor pairs to analyze *in animal*

The ligand-receptor pairs selected for experimental validation of their gene expression had to met the following criteria: 1) the gene pairs have not been shown to interact in the cell types of interest, 2) they are expressed in only a few specific cell types (non-ubiquitous gene expression, based on Fig. 4B), 3) just one of the gene pairs is highly expressed in one interacting cell and the other gene does so in the other interacting cell (to discard autocrine communication and evaluate interactions between LR pairs in different cell types) and 4) they are not highly expressed in cell types that are hard to differentiate under the microscope (e.g., GABAergic neurons are hard to distinguish from cholinergic neurons).

***C. elegans* strains and husbandry**

C. elegans PD4443 (ccIs4443[arg-1::GFP + dpy-20(+)]), KS411 (lin-17(n671) I; unc-119(e2498) III; him-5(e1490) V; mhIs9[lin-17::GFP]), BC12925 (dpy-5(e907) I; sIs10312[rCesC05D11.4::GFP + pCeh361]), LX929 (vsIs48 [unc-17::GFP]), BC12890 (dpy-5(e907) I; sIs11337[rCesY37A1B.5::GFP + pCeh361]) and PS3729 (unc-119(ed4) III; syIs78[ajm-1::GFP + unc-119(+)]) strains were obtained from the Caenorhabditis Genome Center (CGC). For maintenance, the worms were typically grown at 20°C on NGM plates seeded with *E. coli* strain OP50.

Single molecule fluorescent *in-situ* hybridization

Single molecule fluorescent *in-situ* hybridization (smFISH) of L2 stage *C. elegans* was performed as previously described with some modifications (*69*). Briefly, gravid worms of the strains of interest were bleached and the eggs rocked at 20□ for 18 hours to synchronize the population. The L1 worms were then counted and around 5000 worms were seeded on NGM plates containing OP50 *E. coli* strain. Once the worms reached the L2 stage, they were harvested and then incubated in a fixation solution (3.7% formaldehyde in 1x PBS) for 45 minutes. The worms were then washed in 1x PBS and left in 70% ethanol overnight. The next day, the worms were incubated in wash buffer (10% formamide in 2x SSC) for 5 minutes before being incubated overnight at 30□ in the hybridization solution containing the appropriate custom-made Stellaris FISH probes (Biosearch Technologies, United Kingdom). The samples were then washed twice in wash buffer for 30 minutes at 30□ before being incubated in DAPI solution for nuclear counterstaining (10ng/mL in water) for 30 minutes at 30□. Finally, the stained worms were resuspended in 100µL 2x SSC and mounted on agar pads for fluorescent imaging on a Leica confocal microscope (Leica, Germany).

PD4443 worms expressing *arg-1*::GFP were incubated in probes targeting *lin-12* (CAL Fluor Red 590 dye) and GFP (Quasar 670 Dye), KS411 worms expressing *lin-17*::GFP were incubated in probes targeting *lin-44* (CAL Fluor Red 590 dye) and GFP (Quasar 670 Dye), and BC12925 worms expressing *let-756*::GFP were incubated in probes targeting *ver-1* (CAL Fluor Red 590 dye) and GFP (Quasar 670 Dye). Fluorescent imaging of GFP in PD4443 (*arg-1*), KS411 (*lin-17*) and BC12925 (*let-756*) was performed to ensure the expression patterns observed with smFISH were comparable (fig. S4). Additionally, imaging of semo-1::GFP (BC12890) and ajm-1::GFP (PS3729) (fig. S5), which have previously been used to define the location of the hypodermal cells (*70*), was performed to ensure correct annotation of the probe signal observed in smFISH. All images obtained from these conditions were analyzed and processed on Fiji (*71*).

Data and code availability

The single-cell RNA-seq dataset was obtained from the Supplemental Table S4 in (*16*). The 3D digital atlas of *C. elegans* was obtained from the Supplemental Data 1 in (*15*). An annotated version of this atlas with the cell types in the scRNAseq dataset is available in table S6 of this work. The manual curated list containing 245 ligand-receptor interactions is present in the table S1 of this work, while the consensus list from the GA-selection, which contains 37 interactions, is available in the table S3.

All analyses performed in this work, their respective codes (implemented in Python and Jupyter Notebooks), all data, and instructions to use them are available in a public repository (https://github.com/LewisLabUCSD/Celegans-cell2cell). Reproducible runs of our analyses can be performed in a public Code Ocean capsule (https://doi.org/10.24433/CO.4688840.v1). Our open-source suite, *cell2cell*, is for inferring cell-cell interactions from bulk or single-cell RNA-seq data, using or not spatial information, and is available in a GitHub repository (https://github.com/earmingol/cell2cell).

References

1. B. Belardi, S. Son, J. H. Felce, M. L. Dustin, D. A. Fletcher, Cell–cell interfaces as specialized compartments directing cell function. *Nat. Rev. Mol. Cell Biol.* **21**, 750–764 (2020).

2. K. Francis, B. O. Palsson, Effective intercellular communication distances are determined by the relative time constants for cyto/chemokine secretion and diffusion. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 12258–12262 (1997).

3. A. D. Lander, How cells know where they are. *Science*. **339**, 923–927 (2013).

4. Y. Dang, D. A. J. Grundel, H. Youk, Cellular Dialogues: Cell-Cell Communication through Diffusible Molecules Yields Dynamic Spatial Patterns. *Cell Syst*. **10**, 82–98.e7 (2020).

5. J. E. Purvis, G. Lahav, Encoding and decoding cellular information through signaling dynamics. *Cell*. **152**, 945–956 (2013).

6. D. Wu, F. Lin, Modeling cell gradient sensing and migration in competing chemoattractant fields. *PLoS One*. **6**, e18805 (2011).

7. A. M. Pani, B. Goldstein, Direct visualization of a native Wnt in vivo reveals that a long-range Wnt gradient forms by extracellular dispersal. *Elife*. **7** (2018), doi:10.7554/eLife.38325.

8. G. Palla, D. S. Fischer, A. Regev, F. J. Theis, Spatial components of molecular tissue biology. *Nat. Biotechnol.* (2022), doi:10.1038/s41587-021-01182-1.

9. E. Armingol, A. Officer, O. Harismendy, N. E. Lewis, Deciphering cell–cell interactions and communication from gene expression. *Nat. Rev. Genet.*, 1–18 (2020).

10. X. Ren, G. Zhong, Q. Zhang, L. Zhang, Y. Sun, Z. Zhang, Reconstruction of cell spatial organization from single-cell RNA sequencing data based on ligand-receptor mediated self-assembly. *Cell Res.* (2020), doi:10.1038/s41422-020-0353-2.

11. J.-C. Boisset, J. Vivié, D. Grün, M. J. Muraro, A. Lyubimova, A. van Oudenaarden, Mapping the physical network of cellular interactions. *Nat. Methods*. **15**, 547–553 (2018).

12. C. Baccin, J. Al-Sabah, L. Velten, P. M. Helbling, F. Grünschläger, P. Hernández-Malmierca, C. Nombela-Arrieta, L. M. Steinmetz, A. Trumpp, S. Haas, Combined single-cell and spatial transcriptomics reveal the molecular, cellular and spatial bone marrow niche organization. *Nat. Cell Biol.* **22**, 38–48 (2020).

13. Z. Cang, Q. Nie, Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat. Commun.* **11**, 2084 (2020).

14. T. Kaletta, M. O. Hengartner, Finding function in novel targets: C. elegans as a model organism. *Nat. Rev. Drug Discov.* **5**, 387–398 (2006).

15. F. Long, H. Peng, X. Liu, S. K. Kim, E. Myers, A 3D digital atlas of C. elegans and its application to single-cell analyses. *Nat. Methods*. **6**, 667–672 (2009).

16. J. Cao, J. S. Packer, V. Ramani, D. A. Cusanovich, C. Huynh, R. Daza, X. Qiu, C. Lee, S. N. Furlan, F. J. Steemers, A. Adey, R. H. Waterston, C. Trapnell, J. Shendure, Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*. **357**, 661–667 (2017).

17. T. G. Graeber, D. Eisenberg, Bioinformatic identification of potential autocrine signaling loops in cancers from gene expression profiles. *Nat. Genet.* **29**, 295–300 (2001).

18. J. A. Ramilowski, T. Goldberg, J. Harshbarger, E. Kloppmann, M. Lizio, V. P. Satagopam, M. Itoh, H. Kawaji, P. Carninci, B. Rost, A. R. R. Forrest, A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat. Commun.* **6**, 7866 (2015).

19. M. P. Kumar, J. Du, G. Lagoudas, Y. Jiao, A. Sawyer, D. C. Drummond, D. A. Lauffenburger, A. Raue, Analysis of Single-Cell RNA-Seq Identifies Cell-Cell Communication Associated with Tumor Characteristics. *Cell Rep.* **25**, 1458–1468.e4 (2018).

20. R. Vento-Tormo, M. Efremova, R. A. Botting, M. Y. Turco, M. Vento-Tormo, K. B. Meyer, J.-E. Park, E. Stephenson, K. Polański, A. Goncalves, L. Gardner, S. Holmqvist, J. Henriksson, A. Zou, A. M. Sharkey, B. Millar, B. Innes, L. Wood, A. Wilbrey-Clark, R. P. Payne, M. A. Ivarsson, S. Lisgo, A. Filby, D. H. Rowitch, J. N. Bulmer, G. J. Wright, M. J. T. Stubbington, M. Haniffa, A. Moffett, S. A. Teichmann, Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature*. **563**, 347–353 (2018).

21. H. Choi, J. Sheng, D. Gao, F. Li, A. Durrans, S. Ryu, S. B. Lee, N. Narula, S. Rafii, O. Elemento, N. K. Altorki, S. T. C. Wong, V. Mittal, Transcriptome analysis of individual stromal cell populations identifies stroma-tumor crosstalk in mouse lung cancer model. *Cell Rep.* **10**, 1187–1201 (2015).

22. S. Cabello-Aguilar, M. Alame, F. Kon-Sun-Tack, C. Fau, M. Lacroix, J. Colinge, SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.* **48**, e55 (2020).

23. M. Nickerson, Receptor occupancy and tissue response. *Nature*. **178**, 697–698 (1956).

24. P. Zhong, J. F. Cara, H. S. Tager, Importance of receptor occupancy, concentration differences, and ligand exchange in the insulin-like growth factor I receptor system. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 11451–11455 (1993).

25. M. P. Daniels, Intercellular communication that mediates formation of the neuromuscular junction. *Mol. Neurobiol.* **14**, 143–170 (1997).

26. E. J. A. Hubbard, Caenorhabditis elegans germ line: a model for stem cell biology. *Dev. Dyn.* **236**, 3343–3357 (2007).

27. N. Pazdernik, T. Schedl, Introduction to germ cell development in Caenorhabditis elegans. *Adv. Exp. Med. Biol.* **757**, 1–16 (2013).

28. W. Qiao, W. Wang, E. Laurenti, A. L. Turinsky, S. J. Wodak, G. D. Bader, J. E. Dick, P. W. Zandstra, Intercellular network structure and regulatory motifs in the human hematopoietic system. *Mol. Syst. Biol.* **10**, 741 (2014).

29. L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [stat.ML]* (2018), (available at http://arxiv.org/abs/1802.03426).

30. E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I. W. H. Kwok, L. G. Ng, F. Ginhoux, E. W. Newell, Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* (2018), doi:10.1038/nbt.4314.

31. P. Jaccard, The distribution of the flora in the alpine zone.1. *New Phytol.* **11**, 37–50 (1912).

32. P. W. Ingham, Y. Nakano, C. Seger, Mechanisms and functions of Hedgehog signalling across the metazoa. *Nat. Rev. Genet.* **12**, 393–406 (2011).

33. T. Noble, J. Stieglitz, S. Srinivasan, An integrated serotonin and octopamine neuronal circuit directs the release of an endocrine signal to control C. elegans body fat. *Cell Metab.* **18**, 672–684 (2013).

34. E. Witham, C. Comunian, H. Ratanpal, S. Skora, M. Zimmer, S. Srinivasan, C. elegans Body Cavity Neurons Are Homeostatic Sensors that Integrate Fluctuations in Oxygen Availability and Internal Nutrient Reserves. *Cell Rep.* **14**, 1641–1654 (2016).

35. R. Hussey, N. K. Littlejohn, E. Witham, E. Vanstrum, J. Mesgarzadeh, H. Ratanpal, S. Srinivasan, Oxygen-sensing neurons reciprocally regulate peripheral lipid metabolism via neuropeptide signaling in Caenorhabditis elegans. *PLoS Genet.* **14**, e1007305 (2018).

36. H. Tu, P. Huhtala, H.-M. Lee, J. C. Adams, T. Pihlajaniemi, Membrane-associated collagens with interrupted triple-helices (MACITs): evolution from a bilaterian common ancestor and functional conservation in C. elegans. *BMC Evolutionary Biology*. **15** (2015), doi:10.1186/s12862-015-0554-3.

37. J. E. Gleason, E. A. Szyleyko, D. M. Eisenmann, Multiple redundant Wnt signaling components function in two processes during C. elegans vulval development. *Dev. Biol.* **298**, 442–457 (2006).

38. M. P. Klassen, K. Shen, Wnt signaling positions neuromuscular connectivity by inhibiting synapse formation in C. elegans. *Cell*. **130**, 704–716 (2007).

39. M. A. Herman, L. L. Vassilieva, H. R. Horvitz, J. E. Shaw, R. K. Herman, The C. elegans gene lin-44, which controls the polarity of certain asymmetric cell divisions, encodes a Wnt protein and acts cell nonautonomously. *Cell*. **83**, 101–110 (1995).

40. J. Whangbo, C. Kenyon, A Wnt signaling system that specifies two patterns of cell migration in C. elegans. *Mol. Cell*. **4**, 851–858 (1999).

41. C.-L. Pan, J. E. Howell, S. G. Clark, M. Hilliard, S. Cordes, C. I. Bargmann, G. Garriga, Multiple Wnts and frizzled receptors regulate anteriorly directed cell and growth cone migrations in Caenorhabditis elegans. *Dev. Cell*. **10**, 367–377 (2006).

42. M. Harterink, D. H. Kim, T. C. Middelkoop, T. D. Doan, A. van Oudenaarden, H. C. Korswagen, Neuroblast migration along the anteroposterior axis of C. elegans is controlled by opposing gradients of Wnts and a secreted Frizzled-related protein. *Development*. **138**, 2915–2924 (2011).

43. M. A. Miller, I. D. Chin-Sang, Eph receptor signaling in C. elegans. *WormBook*, 1–17 (2012).

44. V. E. Ginzburg, P. J. Roy, J. G. Culotti, Semaphorin 1a and semaphorin 1b are required for correct epidermal cell positioning and adhesion during morphogenesis in C. elegans. *Development*. **129**, 2065–2078 (2002).

45. F. Nakao, M. L. Hudson, M. Suzuki, Z. Peckler, R. Kurokawa, Z. Liu, K. Gengyo-Ando, A. Nukazuka, T. Fujii, F. Suto, Y. Shibata, G. Shioi, H. Fujisawa, S. Mitani, A. D. Chisholm, S. Takagi, The PLEXIN PLX-2 and the ephrin EFN-4 have distinct roles in MAB-20/Semaphorin 2A signaling in Caenorhabditis elegans morphogenesis. *Genetics*. **176**, 1591–1607 (2007).

46. R. Ikegami, H. Zheng, S.-H. Ong, J. Culotti, Integration of semaphorin-2A/MAB-20, ephrin-4, and UNC-129 TGF-beta signaling pathways regulates sorting of distinct sensory rays in C. elegans. *Dev. Cell*. **6**, 383–395 (2004).

47. Z. Liu, T. Fujii, A. Nukazuka, R. Kurokawa, M. Suzuki, H. Fujisawa, S. Takagi, C. elegans PlexinA PLX-1 mediates a cell contact-dependent stop signal in vulval precursor cells. *Dev. Biol.* **282**, 138–151 (2005).

48. X. Wang, W. Zhang, T. Cheever, V. Schwarz, K. Opperman, H. Hutter, D. Koepp, L. Chen, The C. elegans L1CAM homologue LAD-2 functions as a coreceptor in MAB-20/Sema2–mediated axon guidance. *J. Cell Biol.* **180**, 233–246 (2008).

49. D. Angeles-Albores, R. Y. N. Lee, J. Chan, P. W. Sternberg, Two new functions in the WormBase enrichment suite. *microPublication Biology* (2018) (available at https://www.micropublication.org/media/2018/03/microPublication.biology-10.17912-W25Q2N.pdf).

50. T. W. Harris, I. Antoshechkin, T. Bieri, D. Blasiar, J. Chan, W. J. Chen, N. De La Cruz, P. Davis, M. Duesbury, R. Fang, J. Fernandes, M. Han, R. Kishore, R. Lee, H.-M. Müller, C. Nakamura, P. Ozersky, A. Petcherski, A. Rangarajan, A. Rogers, G. Schindelman, E. M. Schwarz, M. A. Tuli, K. Van Auken, D. Wang, X. Wang, G. Williams, K. Yook, R. Durbin, L. D. Stein, J. Spieth, P. W. Sternberg, WormBase: a comprehensive resource for nematode research. *Nucleic Acids Res.* **38**, D463–7 (2010).

51. UniProt Consortium, UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).

52. G. Ou, R. D. Vale, Molecular signatures of cell migration in C. elegans Q neuroblasts. *J. Cell Biol.* **185**, 77–85 (2009).

53. W. Keil, L. M. Kutscher, S. Shaham, E. D. Siggia, Long-Term High-Resolution Imaging of Developing C. elegans Larvae with Microfluidics. *Dev. Cell*. **40**, 202–214 (2017).

54. S. Tuck, The control of cell growth and body size in Caenorhabditis elegans. *Exp. Cell Res.* **321**, 71–76 (2014).

55. V. Lažetić, D. S. Fay, Molting in C. elegans. *Worm*. **6**, e1330246 (2017).

56. A. Y. Zinovyeva, Y. Yamamoto, H. Sawa, W. C. Forrester, Complex network of Wnt signaling regulates neuronal migrations during Caenorhabditis elegans development. *Genetics*. **179**, 1357–1371 (2008).

57. M. Lu, K. Mizumoto, Gradient-independent Wnt signaling instructs asymmetric neurite pruning in C. elegans. *bioRxiv* (2019), p. 715912.

58. Y. Yamamoto, H. Takeshita, H. Sawa, Multiple Wnts redundantly control polarity orientation in Caenorhabditis elegans epithelial stem cells. *PLoS Genet.* **7**, e1002308 (2011).

59. X. Shao, X. Lu, J. Liao, H. Chen, X. Fan, New avenues for systematically inferring cell-cell communication: through single-cell transcriptomics data. *Protein Cell* (2020), doi:10.1007/s13238-020-00727-5.

60. T. L. Gumienny, C. Savage-Dunn, TGF-β signaling in C. elegans. *WormBook*, 1–34 (2013).

61. J. Liao, X. Lu, X. Shao, L. Zhu, X. Fan, Uncovering an Organ's Molecular Architecture at Single-Cell Resolution by Spatially Resolved Transcriptomics. *Trends Biotechnol.* (2020), doi:10.1016/j.tibtech.2020.05.006.

62. L. Larsson, J. Frisén, J. Lundeberg, Spatially resolved transcriptomics adds a new dimension to genomics. *Nat. Methods*. **18**, 15–18 (2021).

63. R. M. Waterhouse, F. Tegenfeldt, J. Li, E. M. Zdobnov, E. V. Kriventseva, OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. *Nucleic Acids Res.* **41**, D358–65 (2013).

64. W. Kim, R. S. Underwood, I. Greenwald, D. D. Shaye, OrthoList 2: A New Comparative Genomic Analysis of Human and Caenorhabditis elegans Genes. *Genetics*. **210**, 445–461 (2018).

65. J. Reimand, M. Kull, H. Peterson, J. Hansen, J. Vilo, g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* **35**, W193–W200 (2007).

66. X.-T. Huang, Y. Zhu, L. L. H. Chan, Z. Zhao, H. Yan, An integrative C. elegans protein–protein interaction network with reliability assessment based on a probabilistic graphical model. *Mol. Biosyst.* **12**, 85–92 (2016).

67. D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, L. J. Jensen, C. von Mering, STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).

68. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, Others, Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*. **12**, 2825–2830 (2011).

69. N. Ji, A. van Oudenaarden, Single molecule fluorescent in situ hybridization (smFISH) of C. elegans worms and embryos. *WormBook* (2012), doi:10.1895/wormbook.1.153.1.

70. Z. F. Altun, D. H. Hall, WormAtlas hermaphrodite handbook - epithelial system - hypodermis. *WormAtlas* (2002), doi:10.3908/wormatlas.1.13.

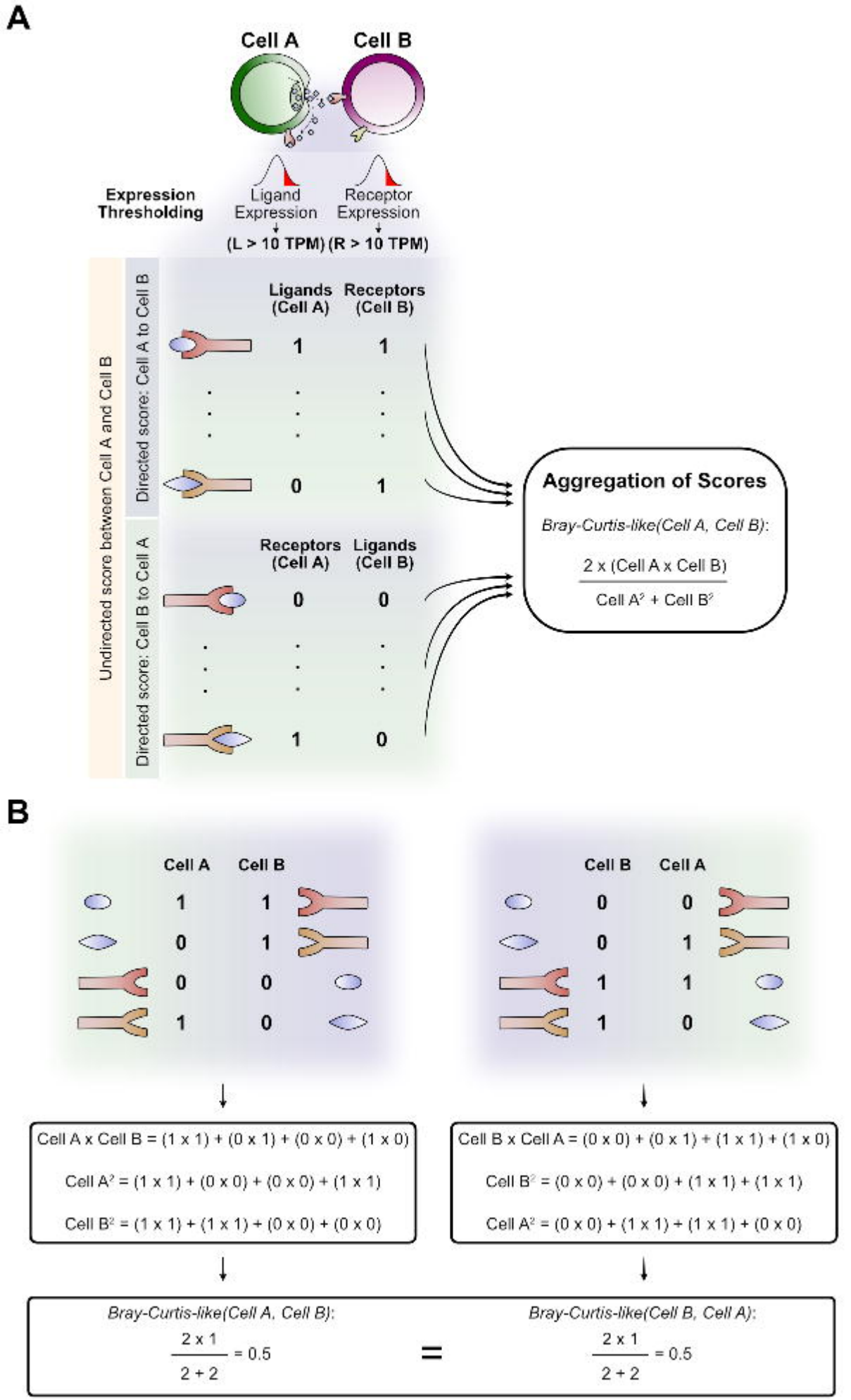
71. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis. *Nat. Methods*. **9**, 676–682 (2012).

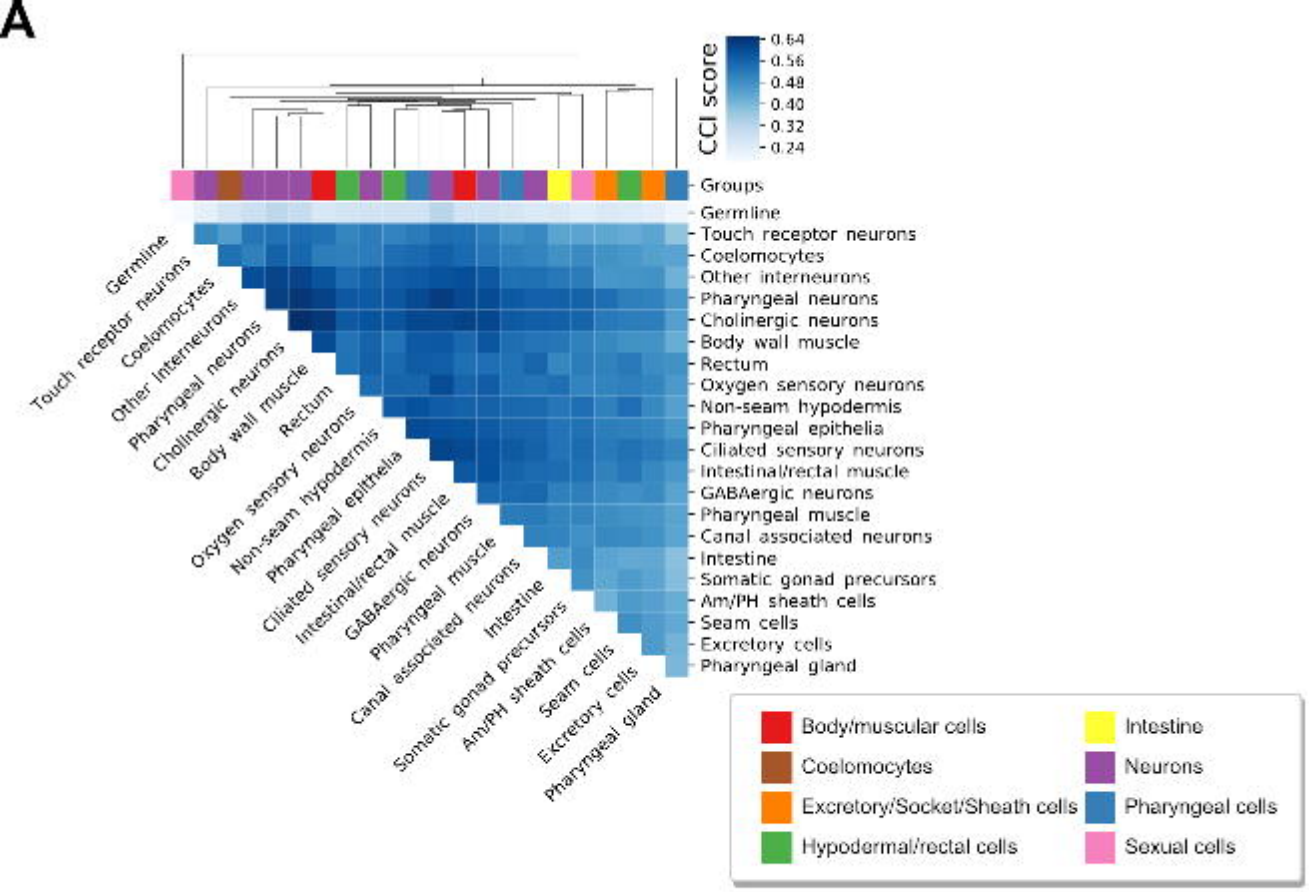
Acknowledgments

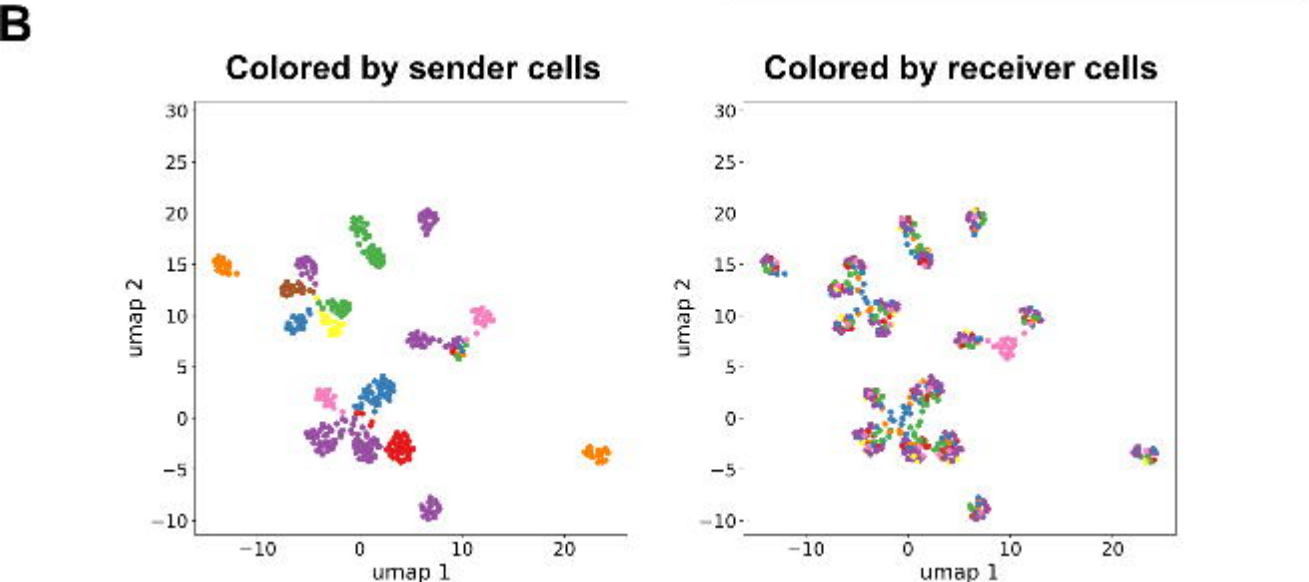
EA is supported by the Chilean Agencia Nacional de Investigación y Desarrollo (ANID) through its scholarship program DOCTORADO BECAS CHILE/2018 - 72190270 and by the Fulbright Commission. This work was further supported by NIGMS grant R35 GM119850 to NEL, a Lilly Innovation Fellows Award to CJ, Jefferson Foundation Award to AG, J Yang Foundation Fellowship to HH, PEW Charitable Trust Award to EJO, and generous funding from the W. M. Keck Foundation. We thank Ariel Pani for helpful comments.

Author contributions: EA, CJ, EJO, and NEL conceived the work. EA and HMB annotated the cells in the previously published 3D digital atlas of *C. elegans* and analyzed their physical locations. EA, CJ, EJO and NEL analyzed the single-cell RNA-seq data. EA and IS generated the list of ligand-receptor pairs of *C. elegans*. EA developed *cell2cell*, performed the CCI analyses and created the corresponding GitHub repositories for both the tool and the analyses. JC helped with the visualization of ligand-receptor interactions. HH implemented enrichment analyses in *cell2cell*. EA, AG and EJO compiled *C. elegans* data and compared results of this work with previous findings in literature. AG and EJO planned *in situ* validations and AG performed the smFISH experiments. EA wrote the paper and all authors carefully reviewed and edited the paper.

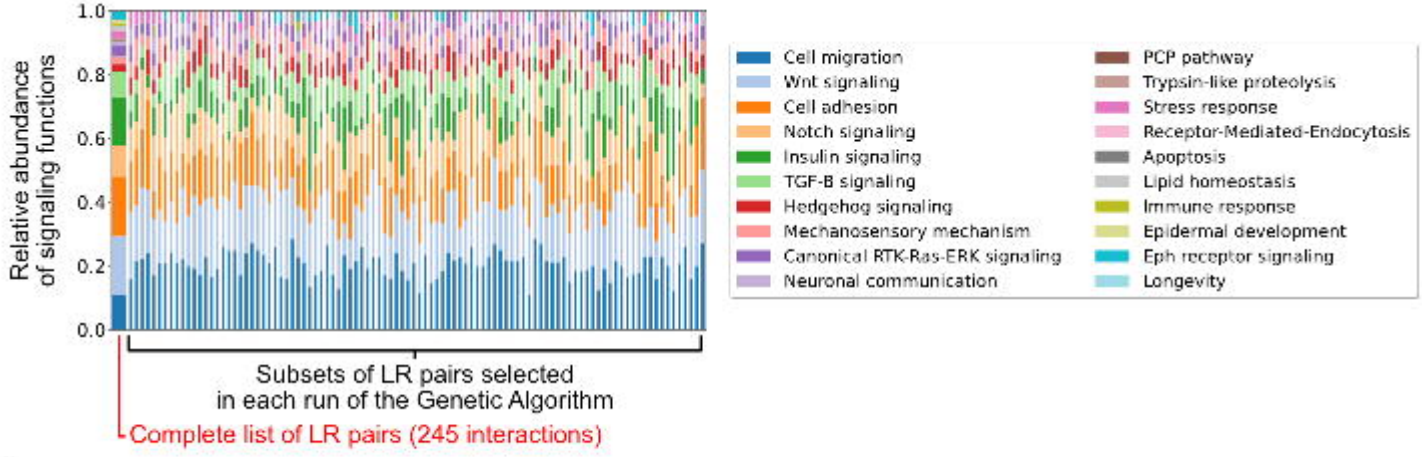
Competing interests: Authors declare that they have no competing interests.


A
Cell A
Cell B
Expression Thresholding
Ligand Expression
Receptor Expression
(L > 10 TPM)
(R > 10 TPM)
Undirected score between Cell A and Cell B
Directed score: Cell A to Cell B
Ligands (Cell A)
Receptors (Cell B)
1
1
0
1
Directed score: Cell B to Cell A
Receptors (Cell A)
Ligands (Cell B)
0
0
1
0
Aggregation of Scores
Bray-Curtis-like(Cell A, Cell B):
2 x (Cell A x Cell B)
Cell A² + Cell B²
B
Cell A
Cell B
1
1
0
1
0
0
1
0
Cell B
Cell A
0
0
0
1
1
1
1
0
Cell A x Cell B = (1 x 1) + (0 x 1) + (0 x 0) + (1 x 0)
Cell A² = (1 x 1) + (0 x 0) + (0 x 0) + (1 x 1)
Cell B² = (1 x 1) + (1 x 1) + (0 x 0) + (0 x 0)
Cell B x Cell A = (0 x 0) + (0 x 1) + (1 x 1) + (1 x 0)
Cell B² = (0 x 0) + (0 x 0) + (1 x 1) + (1 x 1)
Cell A² = (0 x 0) + (1 x 1) + (1 x 1) + (0 x 0)
Bray-Curtis-like(Cell A, Cell B):
2 x 1
2 + 2
= 0.5
=
Bray-Curtis-like(Cell B, Cell A):
2 x 1
2 + 2
= 0.5


A
CCI score
0.64
0.56
0.48
0.40
0.32
0.24
Groups
Germline
Touch receptor neurons
Coelomocytes
Other interneurons
Pharyngeal neurons
Cholinergic neurons
Body wall muscle
Rectum
Oxygen sensory neurons
Non-seam hypodermis
Pharyngeal epithelia
Ciliated sensory neurons
Intestinal/rectal muscle
GABAergic neurons
Pharyngeal muscle
Canal associated neurons
Intestine
Somatic gonad precursors
Am/PH sheath cells
Seam cells
Excretory cells
Pharyngeal gland
Body/muscular cells
Coelomocytes
Excretory/Socket/Sheath cells
Hypodermal/rectal cells
Intestine
Neurons
Pharyngeal cells
Sexual cells

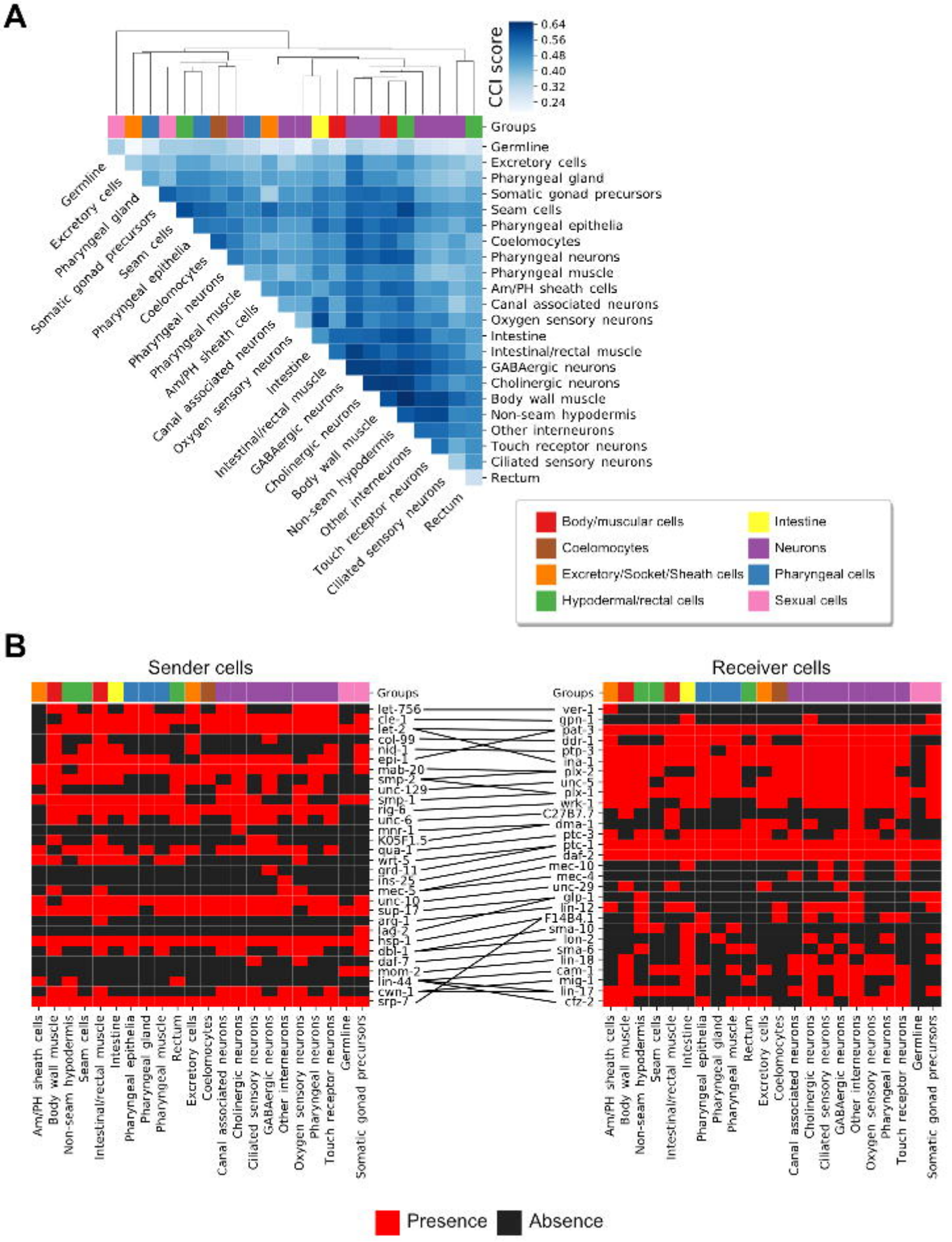

B
Colored by sender cells
Colored by receiver cells
umap 1
umap 2

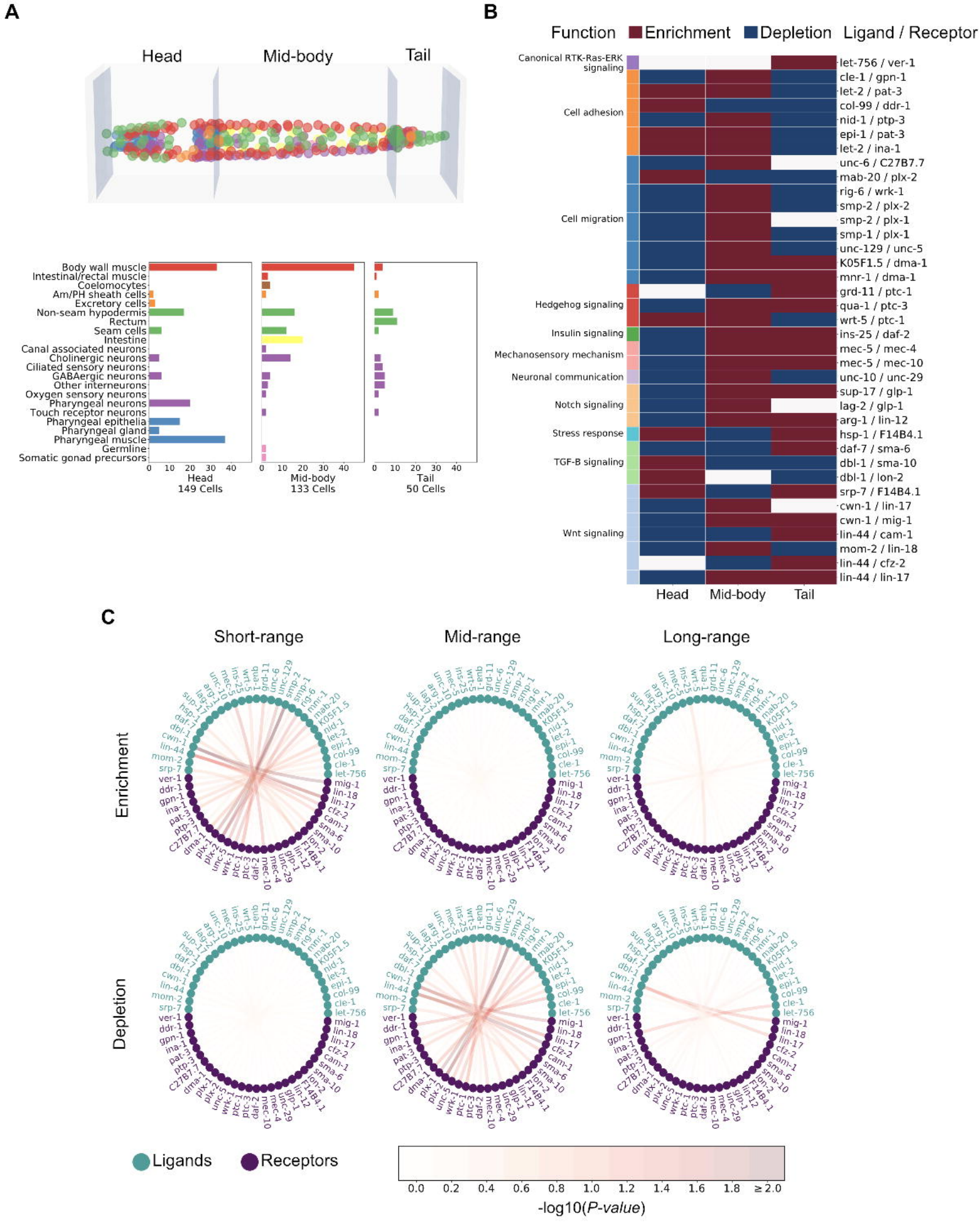
A

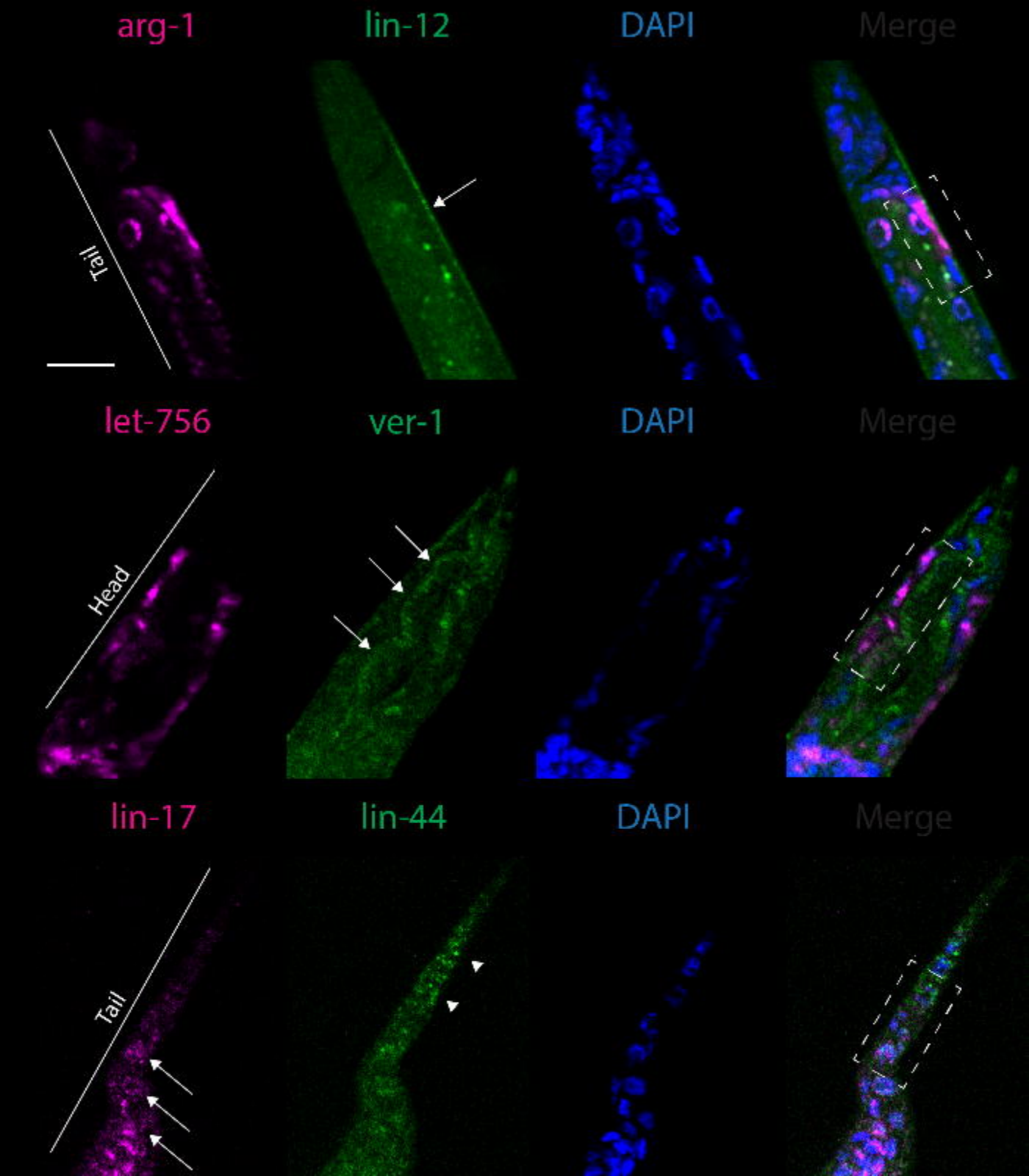


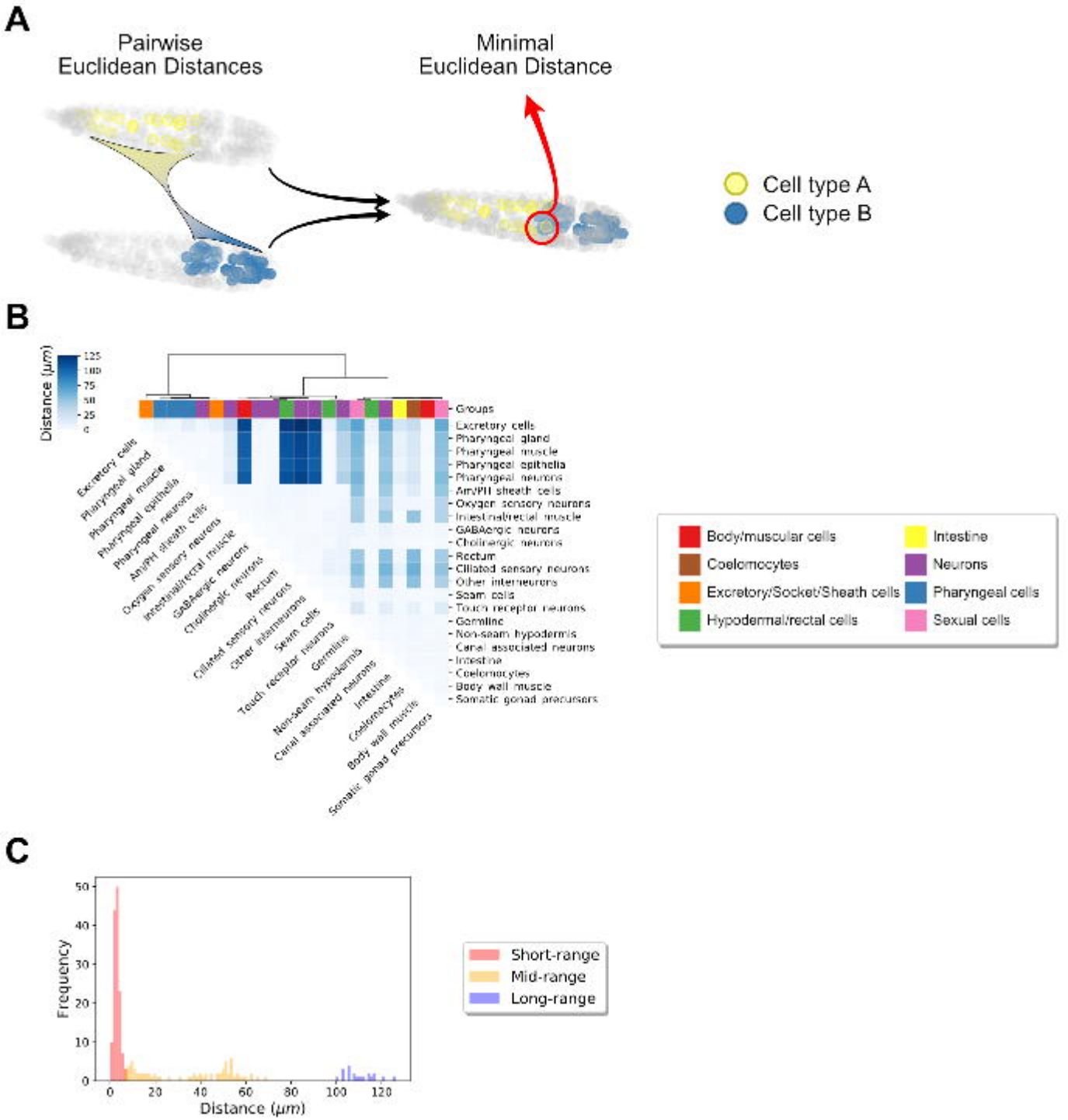
B

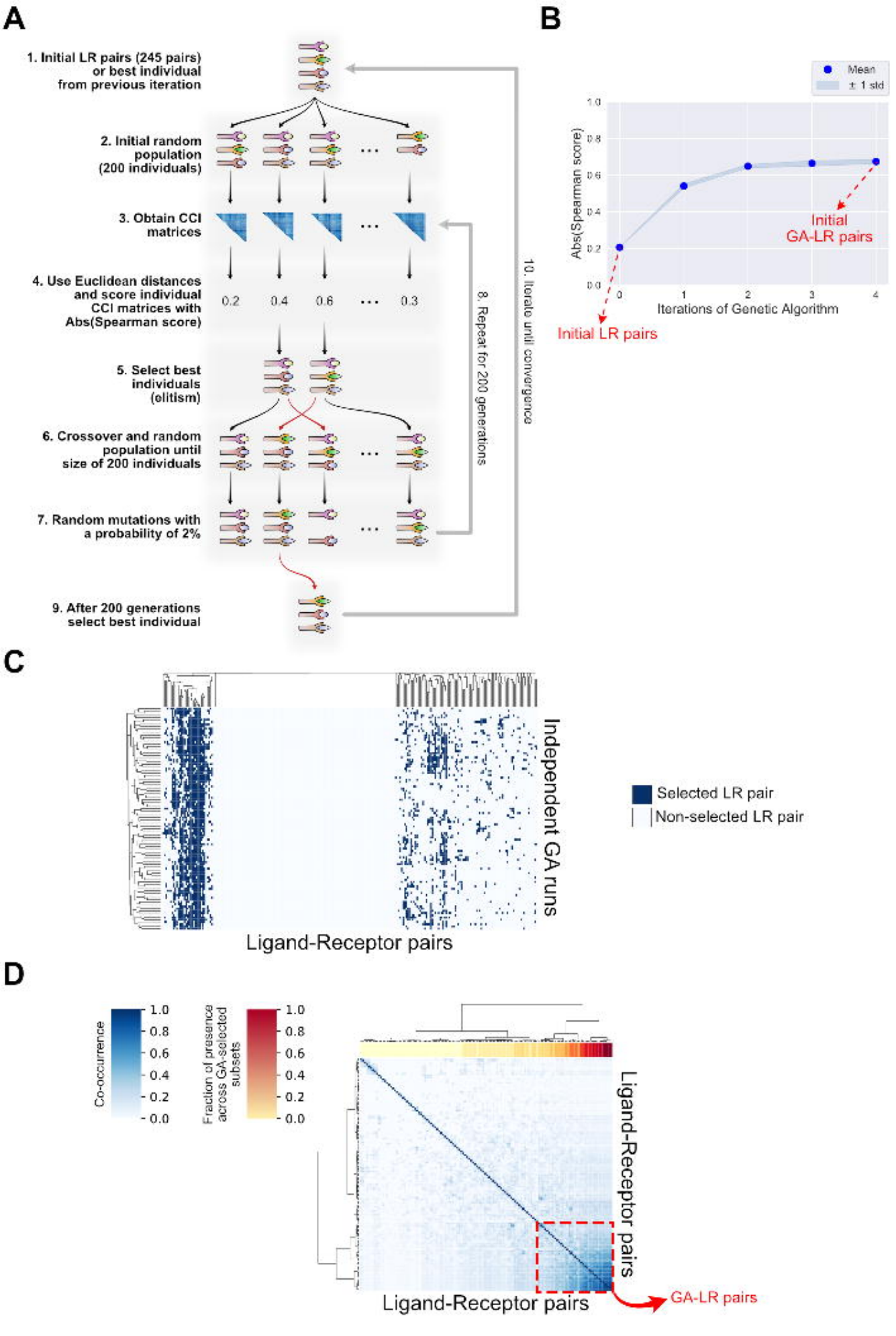
Signaling function	Change	Adj. P-value	Signaling function	Change	Adj. P-value
Cell migration	↑	3.983×10^{-17}	PCP pathway	↓	3.048×10^{-23}
Wnt signaling	—	7.221×10^{-21}	Trypsin-like proteolysis	↓	2.478×10^{-10}
Cell adhesion	↓	2.711×10^{-13}	Stress response	—	2.684×10^{-02}
Notch signaling	—	9.341×10^{-21}	Receptor-Mediated-Endocytosis	↓	4.755×10^{-15}
Insulin signaling	↓	8.158×10^{-13}	Apoptosis	↓	3.048×10^{-23}
TGF-B signaling	—	4.759×10^{-02}	Lipid homeostasis	↓	3.048×10^{-23}
Hedgehog signaling	↑	5.439×10^{-11}	Immune response	↓	5.507×10^{-16}
Mechanosensory mechanism	↑	8.074×10^{-17}	Epidermal development	↓	3.048×10^{-23}
Canonical RTK-Ras-ERK signaling	↑	3.724×10^{-25}	Eph receptor signaling	↓	4.472×10^{-13}
Neuronal communication	—	4.431×10^{-21}	Longevity	↓	3.048×10^{-21}

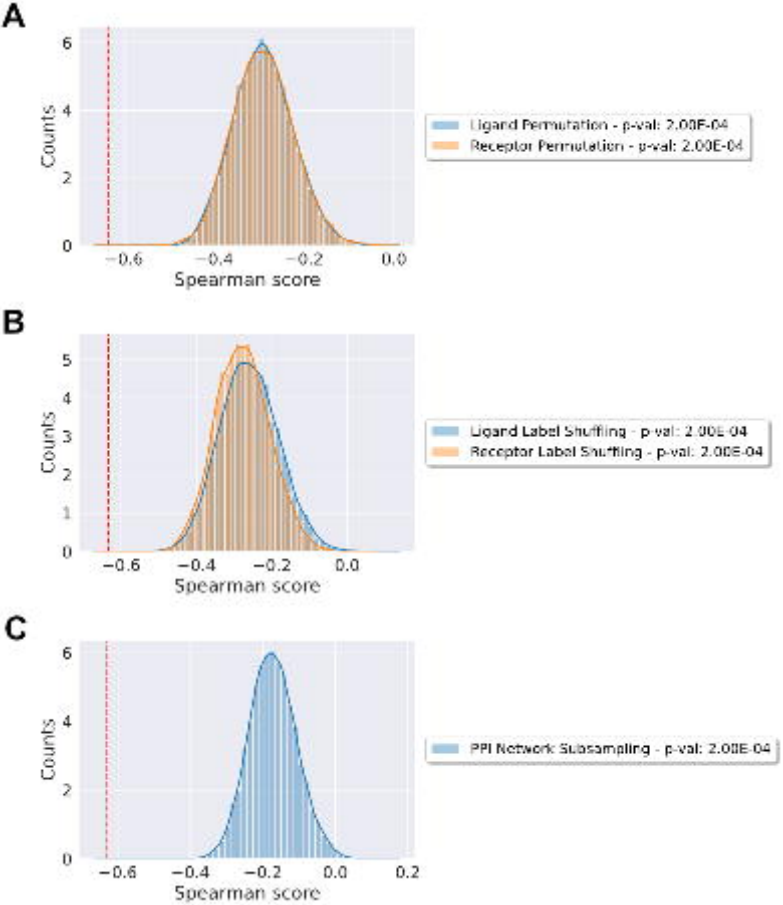

A
CCI score
0.64
0.56
0.48
0.40
0.32
0.24
Groups
Germline
Excretory cells
Pharyngeal gland
Somatic gonad precursors
Seam cells
Pharyngeal epithelia
Coelomocytes
Pharyngeal neurons
Pharyngeal muscle
Am/PH sheath cells
Canal associated neurons
Oxygen sensory neurons
Intestine
Intestinal/rectal muscle
GABAergic neurons
Cholinergic neurons
Body wall muscle
Non-seam hypodermis
Other interneurons
Touch receptor neurons
Ciliated sensory neurons
Rectum
Body/muscular cells
Coelomocytes
Excretory/Socket/Sheath cells
Hypodermal/rectal cells
Intestine
Neurons
Pharyngeal cells
Sexual cells
B
Sender cells
Receiver cells
Groups
let-756
cle-1
let-2
col-99
nid-1
epi-1
mab-20
smp-2
unc-129
smp-1
rig-6
unc-6
mnr-1
K05F1.5
qua-1
wrt-5
grd-11
ins-25
mec-5
unc-10
sup-17
arg-1
lad-2
hsp-1
dbl-1
daf-7
mom-2
lin-44
cwn-1
srp-7
ver-1
gpn-1
pat-3
ddr-1
ptp-3
ina-1
plx-2
unc-5
plx-1
wrk-1
C27B7.7
dma-1
ptc-3
ptc-1
daf-2
mec-10
mec-4
unc-29
glp-1
lin-12
F14B4.1
sma-10
lon-2
sma-6
lin-18
cam-1
mig-1
lin-17
cfz-2
Am/PH sheath cells
Body wall muscle
Non-seam hypodermis
Seam cells
Intestinal/rectal muscle
Intestine
Pharyngeal epithelia
Pharyngeal gland
Pharyngeal muscle
Rectum
Excretory cells
Coelomocytes
Canal associated neurons
Cholinergic neurons
Ciliated sensory neurons
GABAergic neurons
Other interneurons
Oxygen sensory neurons
Pharyngeal neurons
Touch receptor neurons
Germline
Somatic gonad precursors
Presence
Absence

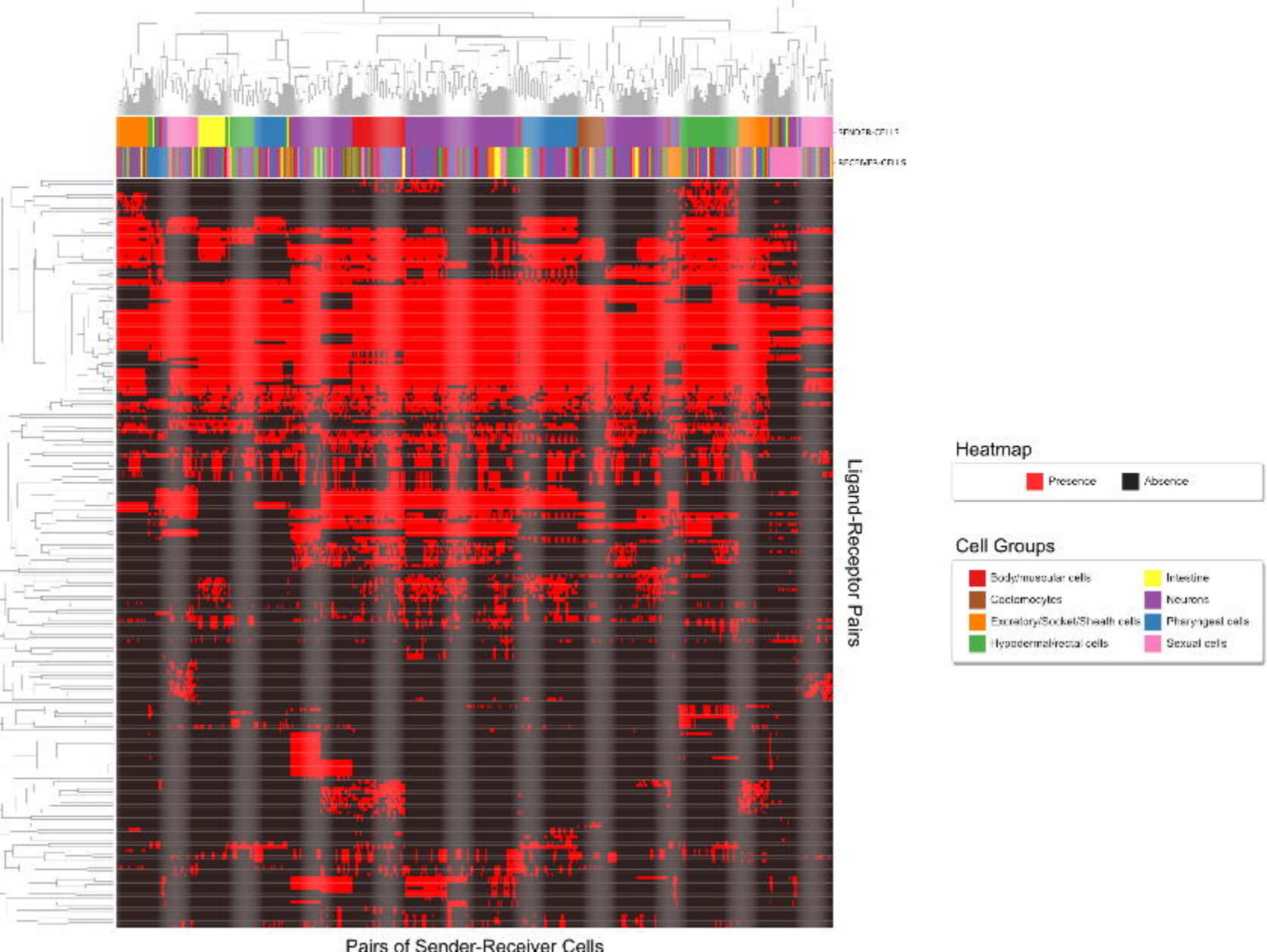

A
Head
Mid-body
Tail
Body wall muscle
Intestinal/rectal muscle
Coelomocytes
Am/PH sheath cells
Excretory cells
Non-seam hypodermis
Rectum
Seam cells
Intestine
Canal associated neurons
Cholinergic neurons
Ciliated sensory neurons
GABAergic neurons
Other interneurons
Oxygen sensory neurons
Pharyngeal neurons
Touch receptor neurons
Pharyngeal epithelia
Pharyngeal gland
Pharyngeal muscle
Germline
Somatic gonad precursors
0 10 20 30 40
Head
149 Cells
Mid-body
133 Cells
Tail
50 Cells
B
Function
Enrichment
Depletion
Ligand / Receptor
Canonical RTK-Ras-ERK signaling
let-756 / ver-1
Cell adhesion
cle-1 / gpn-1
let-2 / pat-3
col-99 / ddr-1
nid-1 / ptp-3
epi-1 / pat-3
let-2 / ina-1
Cell migration
unc-6 / C27B7.7
mab-20 / plx-2
rig-6 / wrk-1
smp-2 / plx-2
smp-2 / plx-1
smp-1 / plx-1
unc-129 / unc-5
K05F1.5 / dma-1
mnr-1 / dma-1
Hedgehog signaling
grd-11 / ptc-1
qua-1 / ptc-3
wrt-5 / ptc-1
Insulin signaling
ins-25 / daf-2
Mechanosensory mechanism
mec-5 / mec-4
Neuronal communication
mec-5 / mec-10
unc-10 / unc-29
Notch signaling
sup-17 / glp-1
lag-2 / glp-1
Stress response
arg-1 / lin-12
hsp-1 / F14B4.1
TGF-B signaling
daf-7 / sma-6
dbl-1 / sma-10
dbl-1 / lon-2
Wnt signaling
srp-7 / F14B4.1
cwn-1 / lin-17
cwn-1 / mig-1
lin-44 / cam-1
mom-2 / lin-18
lin-44 / cfz-2
lin-44 / lin-17
Head
Mid-body
Tail
C
Short-range
Mid-range
Long-range
Enrichment
Depletion
Ligands
Receptors
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 ≥2.0
-log10(P-value)

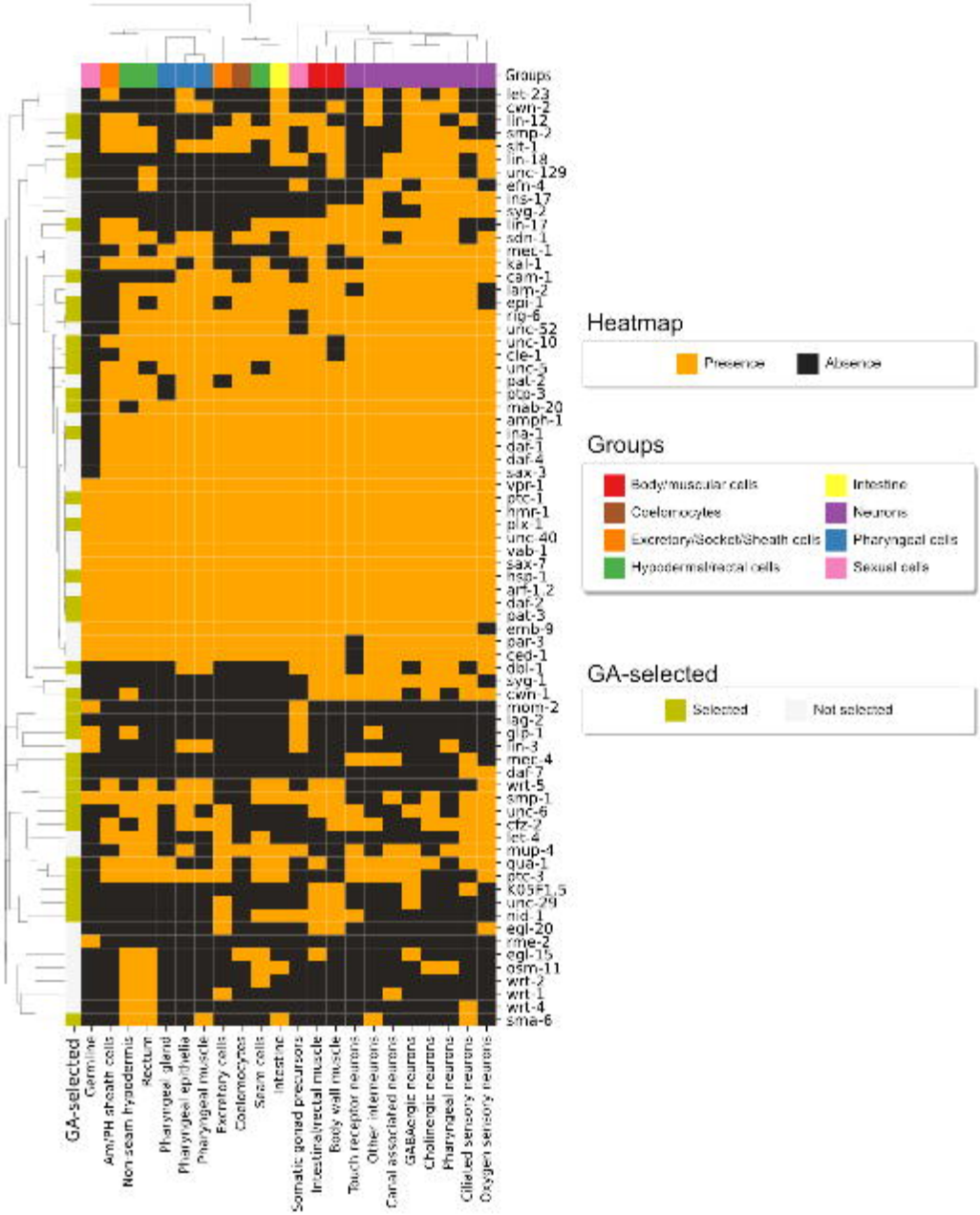

arg-1
lin-12
DAPI
Merge
Tail
let-756
ver-1
DAPI
Merge
Head
lin-17
lin-44
DAPI
Merge
Tail


A
Pairwise
Euclidean Distances
Minimal
Euclidean Distance
Cell type A
Cell type B
B
Distance (μm)
125
100
75
50
25
0
Groups
Excretory cells
Pharyngeal gland
Pharyngeal muscle
Pharyngeal epithelia
Pharyngeal neurons
Am/PH sheath cells
Oxygen sensory neurons
Intestinal/rectal muscle
GABAergic neurons
Cholinergic neurons
Rectum
Ciliated sensory neurons
Other interneurons
Seam cells
Touch receptor neurons
Germline
Non-seam hypodermis
Canal associated neurons
Intestine
Coelomocytes
Body wall muscle
Somatic gonad precursors
Body/muscular cells
Coelomocytes
Excretory/Socket/Sheath cells
Hypodermal/rectal cells
Intestine
Neurons
Pharyngeal cells
Sexual cells
C
Frequency
0
10
20
30
40
50
0
20
40
60
80
100
120
Distance (μm)
Short-range
Mid-range
Long-range

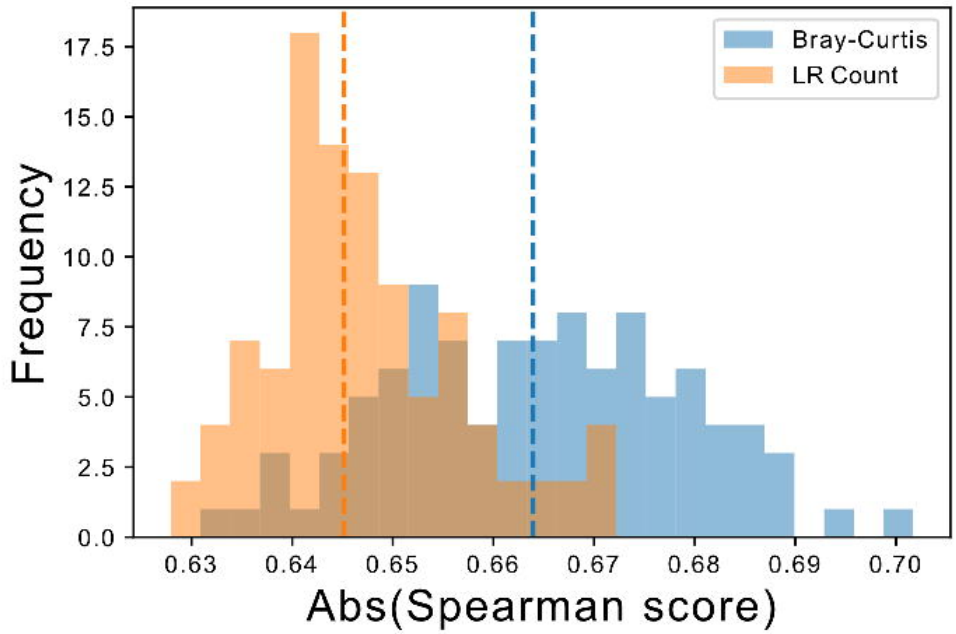

A
1. Initial LR pairs (245 pairs) or best individual from previous iteration
2. Initial random population (200 individuals)
3. Obtain CCI matrices
4. Use Euclidean distances and score individual CCI matrices with Abs(Spearman score)
0.2
0.4
0.6
0.3
5. Select best individuals (elitism)
6. Crossover and random population until size of 200 individuals
7. Random mutations with a probability of 2%
8. Repeat for 200 generations
9. After 200 generations select best individual
10. Iterate until convergence
B
Mean
± 1 std
Abs(Spearman score)
Iterations of Genetic Algorithm
Initial GA-LR pairs
Initial LR pairs
C
Independent GA runs
Ligand-Receptor pairs
Selected LR pair
Non-selected LR pair
D
Co-occurrence
Fraction of presence across GA-selected subsets
Ligand-Receptor pairs
Ligand-Receptor pairs
GA-LR pairs


A
Counts
Spearman score
Ligand Permutation - p-val: 2.00E-04
Receptor Permutation - p-val: 2.00E-04
B
Counts
Spearman score
Ligand Label Shuffling - p-val: 2.00E-04
Receptor Label Shuffling - p-val: 2.00E-04
C
Counts
Spearman score
PPI Network Subsampling - p-val: 2.00E-04


Ligand-Receptor Pairs
Pairs of Sender-Receiver Cells
Heatmap
Presence
Absence
Cell Groups
Body/muscular cells
Coelomocytes
Excretory/Socket/Sheath cells
Hypodermal/rectal cells
Intestine
Neurons
Pharyngeal cells
Sexual cells


Groups
let-23
cwn-2
lin-12
smp-2
slt-1
lin-18
unc-129
efn-4
ins-17
syg-2
lin-17
sdn-1
met-1
kal-1
cam-1
lam-2
epi-1
rig-6
unc-52
unc-10
cle-1
unc-5
pat-2
pbo-3
mab-20
amph-1
ina-1
daf-1
daf-4
sax-3
vpr-1
ptc-1
hmr-1
plx-1
unc-40
vab-1
sax-7
nsg-1
art-1.2
daf-2
pat-3
emb-9
par-3
ced-1
dbl-1
syg-1
cwn-1
mom-2
lag-2
glp-1
lin-3
mor-4
daf-7
wrt-5
smp-1
unc-6
cfz-2
let-4
mup-4
qua-1
ptc-3
K05F1.5
unc-29
nid-1
egl-20
rme-2
egl-15
osm-11
wrt-2
wrt-1
wrt-4
sma-6
GA-selected
Germline
Am/PH sheath cells
Non-seam hypodermis
Rectum
Pharyngeal gland
Pharyngeal epithelia
Pharyngeal muscle
Excretory cells
Coelomocytes
Seam cells
Intestine
Somatic gonad precursors
Intestine/rectal muscle
Body wall muscle
Touch receptor neurons
Other interneurons
Canal associated neurons
GABAergic neurons
Cholinergic neurons
Pharyngeal neurons
Ciliated sensory neurons
Oxygen sensory neurons
Heatmap
Presence
Absence
Groups
Body/muscular cells
Intestine
Coelomocytes
Neurons
Excretory/Socket/Sheath cells
Pharyngeal cells
Hypodermal/rectal cells
Sexual cells
GA-selected
Selected
Not selected

A



Mann-Whitney U Test

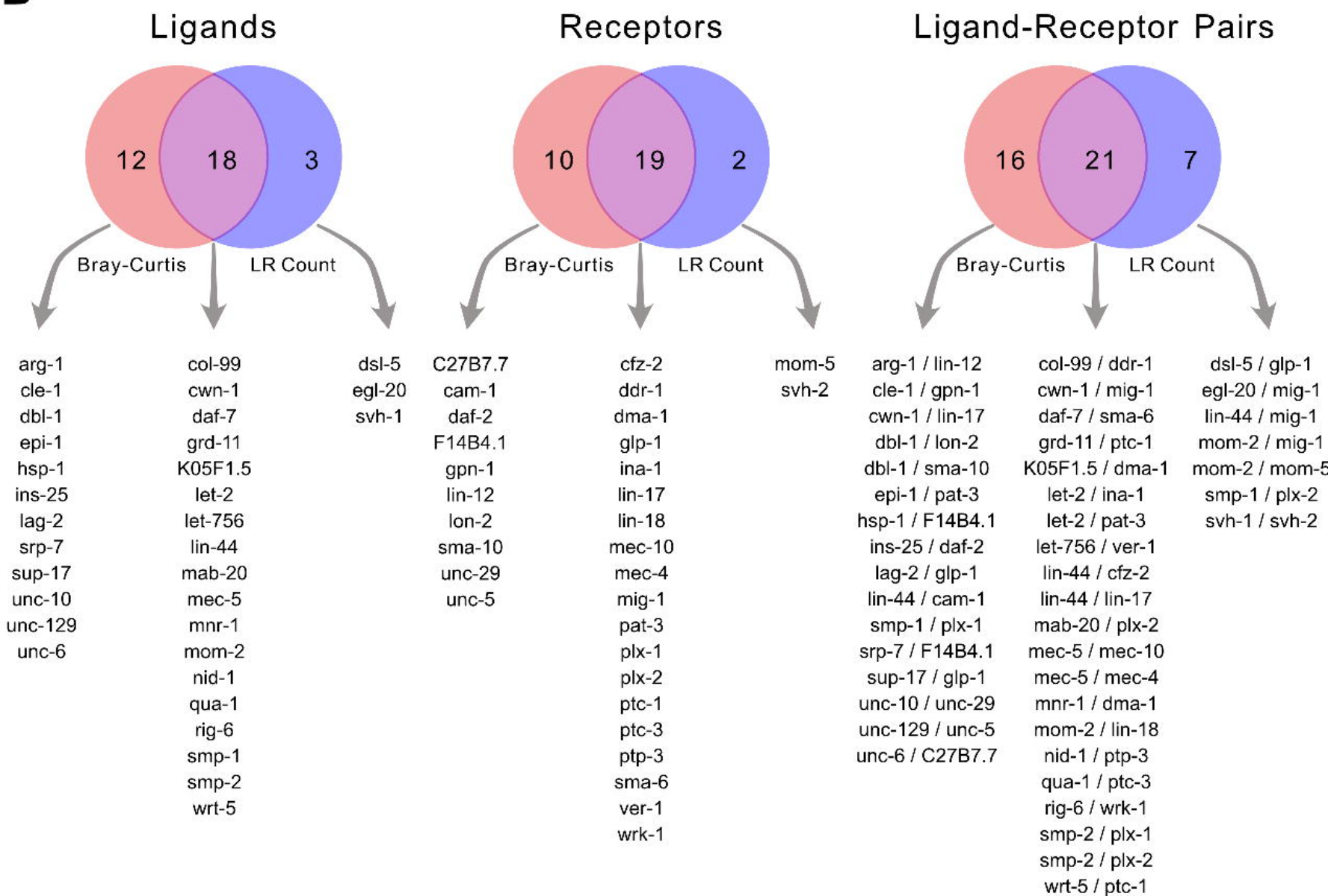
Test-statistic : 8339

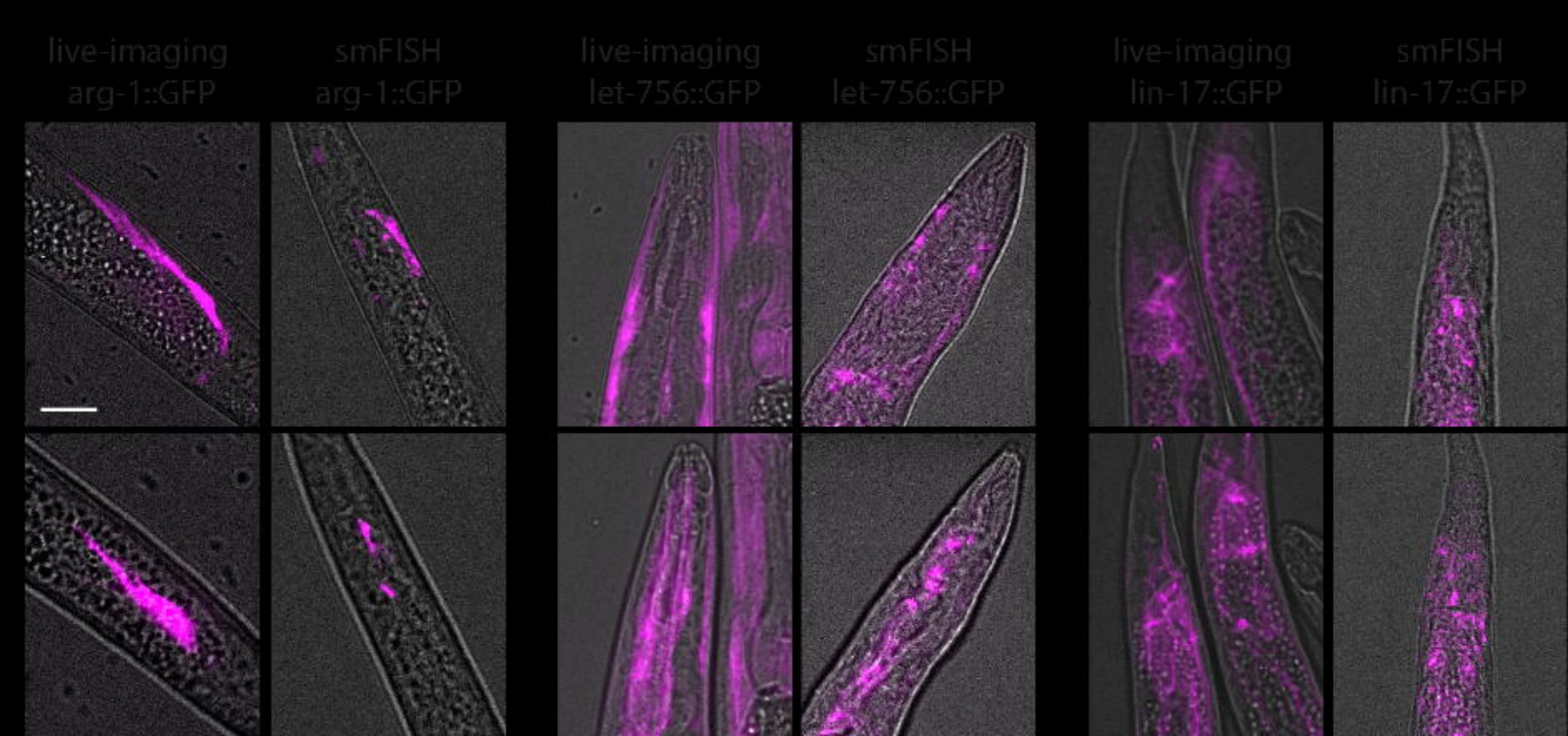
P-value : 3.43×10^{-16}

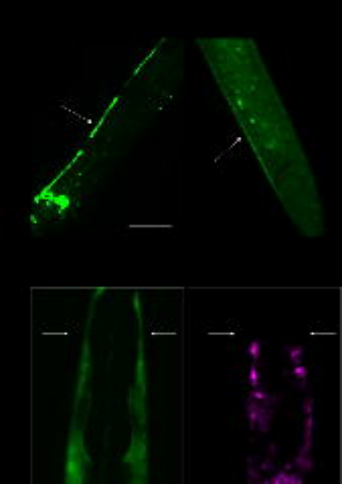
Bray-Curtis median : 0.664

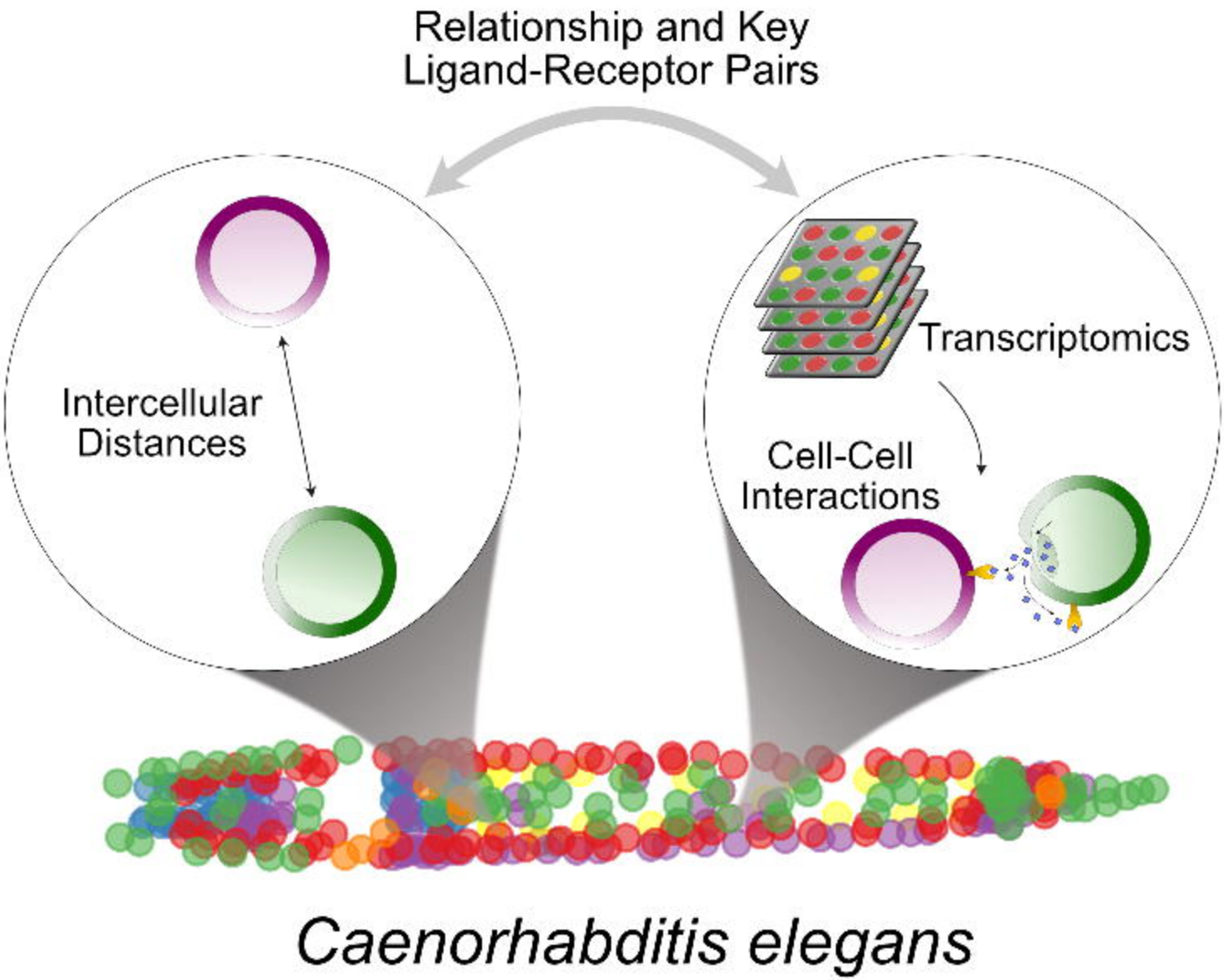
LR Count median : 0.645

B




live-imaging
arg-1::GFP
smFISH
arg-1::GFP
live-imaging
let-756::GFP
smFISH
let-756::GFP
live-imaging
lin-17::GFP
smFISH
lin-17::GFP




Relationship and Key
Ligand-Receptor Pairs
Intercellular
Distances
Transcriptomics
Cell-Cell
Interactions
Caenorhabditis elegans