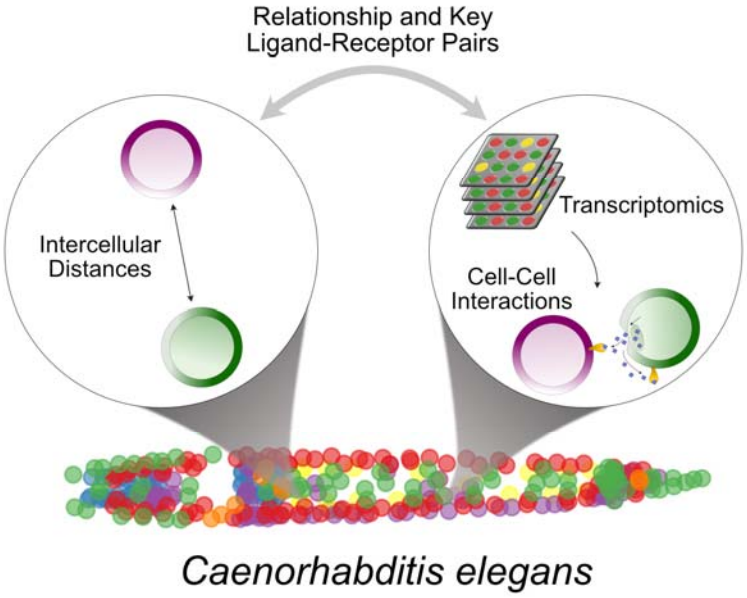
bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

Graphical abstract



bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Summary

Cell-cell interactions shape cellular function and ultimately organismal phenotype. However, the spatial code embedded in the molecular interactions driving and sustaining spatial organization remains to be elucidated. Here we present a computational framework to infer the spatial code underlying cell-cell interactions in a whole animal. Using the transcriptomes of the cell types composing *Caenorhabditis elegans*' body, we compute the potential for intercellular interactions from the coexpression of ligand-receptor pairs. Leveraging a 3D atlas of *C. elegans*' cells and a genetic algorithm we identify the ligand-receptor pairs most informative of the spatial organization of cells. The resulting intercellular distances are negatively correlated with the potential for cell-cell interaction, validating this strategy. Further, for selected ligand-receptor pairs, we experimentally confirm the algorithm-generated cell-cell interactions. Thus, our computational framework helps identify a code associated with spatial organization and cellular functions across a whole-animal body, showing that single-cell molecular measurements provide spatial information that may help elucidate organismal phenotypes and disease.

Highlights

- A cell-cell interaction network in the whole body of *C. elegans* is presented.
- Intercellular distance and interactions are negatively correlated.
- A combination of ligand-receptor pairs carries a spatial code of cell-cell interactions.
- Spatial expression of specific ligand-receptor pairs is validated *in vivo*.

Keywords

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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. In physiological, stress, and pathological conditions, CCIs relay signals between cells and trigger downstream signaling events that culminate in altered gene expression. Through these and other functional interactions, CCIs form spatial patterns of interaction (Dang et al., 2020) and mediate collective cellular behaviors (Sgro et al., 2015; Sun et al., 2012).

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). Ligands can mediate CCIs across a range of distances and can encode positional information for cells within tissues, which is critical for cellular function and decision-making, and therefore organismal phenotypes (Lander, 2013). For instance, some ligands form gradients that serve as a cue for cell migration (Pani and Goldstein, 2018; Wu and Lin, 2011). Thus, studying CCIs can help elucidate how multicellular functions are coordinated by both the molecules mediating CCIs and their spatial context.

CCIs can be inferred from transcriptomic data (Armingol et al., 2020). Computational analysis of CCIs usually involves examining the co-expression ligands and their cognate receptors. RNA-based analyses have informed CCIs and their mediators in small communities of cells, such as embryos (Cang and Nie, 2020; Xue et al., 2019) and tissues (Camp et al., 2017; Joost et al., 2016, 2018; Pavličev et al., 2017; Popescu et al., 2019; Sheikh et al., 2019; Vento-Tormo et al., 2018; Wang et al., 2019a, 2019b; Yuzwa et al., 2016). Moreover, CCI analyses have enabled the study of all cell types in the whole body of a multicellular organism in post-embryonic stages (Hou et al., 2020), yet without considering spatial properties held by the cell locations.

CCIs can define cell location and community spatial structure, enabling coordination of functions between interacting cells (Lander, 2013; Purvis and Lahav, 2013). As such, molecules mediating CCIs encode and pass a spatial code between cells. Thus, the study of CCIs can help decode spatial organization and the functions embedded in this code. Although spatial information is lost during tissue dissociation in conventional single-cell RNA-sequencing technologies (scRNA-seq) (Ren et al., 2020), previous studies have proven that gene expression levels still encode spatial information that can be recovered by adding extra information such as protein-protein interactions and/or microscopy data (Baccin et al., 2020; Boisset et al., 2018; Cang and Nie, 2020; Ren et al., 2020). For example, RNA-Magnet inferred cellular contacts in the bone marrow by considering the coexpression of adhesion molecules present on cell surfaces (Baccin et al., 2020), while ProximID used gene expression coupled with microscopy of cells to construct a spatial map of cell-cell contacts in bone marrow (Boisset et al., 2018). However, it remains unclear if one can find, in RNA, a spatial code of messages transmitted between cells that defines spatial organization and cellular functions across a whole animal body.

Caenorhabditis elegans is an excellent model for studying CCIs in a spatial context across a whole body (Kaletta and Hengartner, 2006). This animal has fewer than 1,000 somatic cells stereotypically arranged across the body. Despite the small number of cells, the intercellular organization shows complexity comparable to higher-order organisms. Here we use *C. elegans'* scRNA-seq data to compute CCIs and assess, across the *C. elegans* body, which ligand-receptor pairs govern a spatial code passed between cell types in different locations and ranges of distance. For this purpose, a previously published single-cell transcriptome (Cao et al., 2017) was integrated with a 3D-atlas of cells of *C. elegans* (Long et al., 2009), while a list of ligand-

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

receptor interactions was built, which to date is the most comprehensive one for CCI analyses of *C. elegans*. Next, we compared our predictions to literature and found them to be consistent with several publications independently reporting relevant roles of the identified LR interactions as encoders of spatial information. Additionally, we experimentally tested a few uncharacterized CCIs, and validated that adjacent cells co-express the LR pairs computationally inferred to contribute to the spatial code. Thus, together, we demonstrate that as long as single-cell RNAseq data is available and there is a sufficient number of characterized ligand-receptor pairs, a genotype-spatial phenotype link can be found at the whole body level in a multicellular organism.

Results

Computing cell-cell interactions

To study cell-cell interactions, a first step can be revealing active communication pathways from the coexpression of the corresponding LR pairs. Communication scores can be assigned to these interactions based on the RNA expression levels of the genes encoding the secreted and receiver proteins in a pair of cells (Choi et al., 2015; Graeber and Eisenberg, 2001; Kumar et al., 2018; Ramilowski et al., 2015; Vento-Tormo et al., 2018). Communication scores are then aggregated into an overall CCI score, which estimates the overall interaction state between two cells. Although focusing on overall CCIs is useful for quantifying the extent of interaction between pairs of cells, the interpretation of overall CCI scores depends on the assumptions made (Armingol et al., 2020). While previous studies have assumed that CCI scores such as the number of active LR pairs represents the strength of cell-cell interactions (Ramilowski et al., 2015), and the odds ratios of LR pairs in differentially expressed genes can reveal cell-type specific priming of cells prior to interacting (Krausgruber et al., 2020), here we propose a novel CCI score that informs interaction properties associated with spatial information. Briefly, our score is based on the idea that proximal cells coordinate their gene expressions such that their total production of ligands and receptors is more complementary than with distant cells (more details below).

Intercellular communication allows cells to coordinate their gene expression (Sgro et al., 2015) and to form spatial patterns of molecule exchange (Dang et al., 2020). These events also allow cells to sense their spatial proximity, which is essential for both the formation and the homeostasis of tissues and organs (Lander, 2013). For example, one mechanism includes sensing the occupancy of receptors by signals from surrounding cells (Lander, 2013; Nickerson, 1956); higher occupancy can indicate greater proximity of communicating cells (Zhong et al., 1993). Thus, to represent a cell-cell potential of interaction that may respond to or drive intercellular proximity, we propose a Bray-Curtis-like score (Figure 1). This CCI score is computed from the mRNA expression of ligands and receptors to represent the molecular complementarity of a pair of interacting cells. In contrast to previous CCI scores, we weigh the number of LR pairs that a pair of cells use to communicate by the aggregate total of ligands and receptors that each cell in the pair produces (see Methods and Figure 1B). The main assumption of our CCI score is that the smaller the intercellular distance, the more complementary is the production of the ligands and receptors in a 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 application of the CCI analyses, we developed *cell2cell*, an open source tool to infer intercellular interactions and communication with any gene expression matrix and list of LR pairs as inputs (https://github.com/earmingol/cell2cell). Although we chose our Bray-Curtis-like

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

score for representing a spatial-dependent complementarity of interaction, *cell2cell* is flexible in terms of the scoring strategies applied to decipher CCIs, so depending on the purpose of study other CCI scores can be used (e.g. the number of active LR pairs to represent the strength of the interaction).

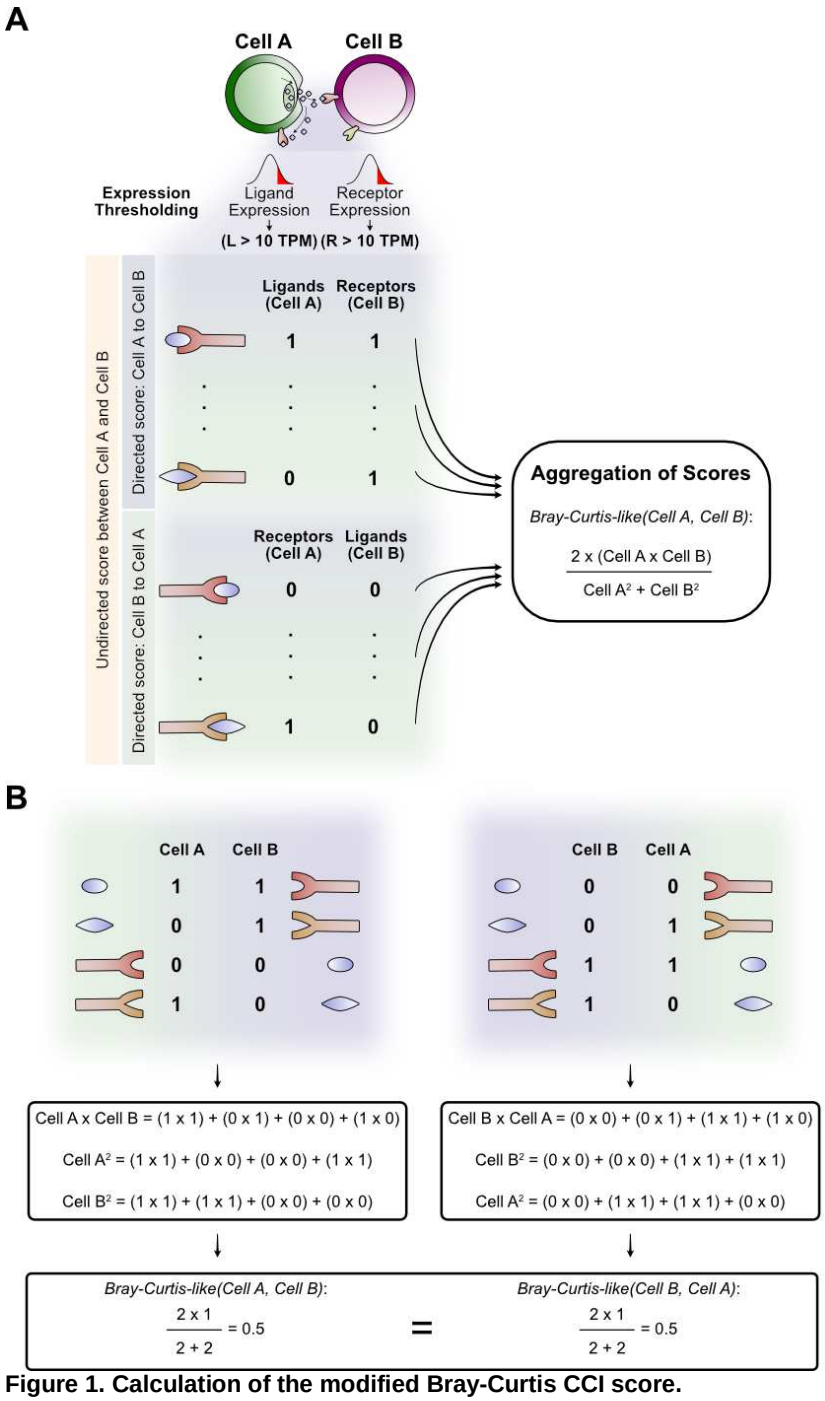


Figure 1. Calculation of the modified Bray-Curtis CCI score.
(A) To represent the overall interaction potential between cell A and cell B, our CCI score is computed from two vectors representing the ligands and receptors independently expressed in each cell. If only the ligands from one cell and the 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 compute CCIs in *C. elegans*, a list containing the interactions between its ligands and receptors is needed. Although much is known about this organism, 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 that consists of 245 ligand-receptor interactions in *C. elegans* (Supplemental Table 1). Next, we used this list to determine the presence or absence of ligands and receptors in each cell identified in the single-cell transcriptome of *C. elegans* (Cao et al., 2017), and ultimately inferred the active LR pairs in all pairs of cells through a binary communication score (Supplemental Table 2). Thus, 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 cells given their active LR pairs (Figure 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 (Figure 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 (Daniels, 1997). Similarly, germline cells have the lowest CCI potential with other cell types. In this regard, the limited intercellular communication of germline cells with other cell types has been previously reported (Hubbard, 2007; Pazdernik and Schedl, 2013), as the basement membranes surrounding them impose a physical constraint that may affect the expression coordination of LR pairs. Thus, these results suggest that our method is exposing interactions between cell pairs based on signals transmitted between them.

The Jaccard similarity between pairs of interacting cells was also analyzed given the LR pairs used. By using UMAP (Becht et al., 2018; McInnes et al., 2018) to visualize the similarity that cell types have (Figure 2B), we observe 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). This result is consistent with previous findings that ligands are produced in a cell type-specific manner by human cells, but receptors are promiscuously produced (Qiao et al., 2014), which suggests that this phenomenon may be conserved across multicellular organisms.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

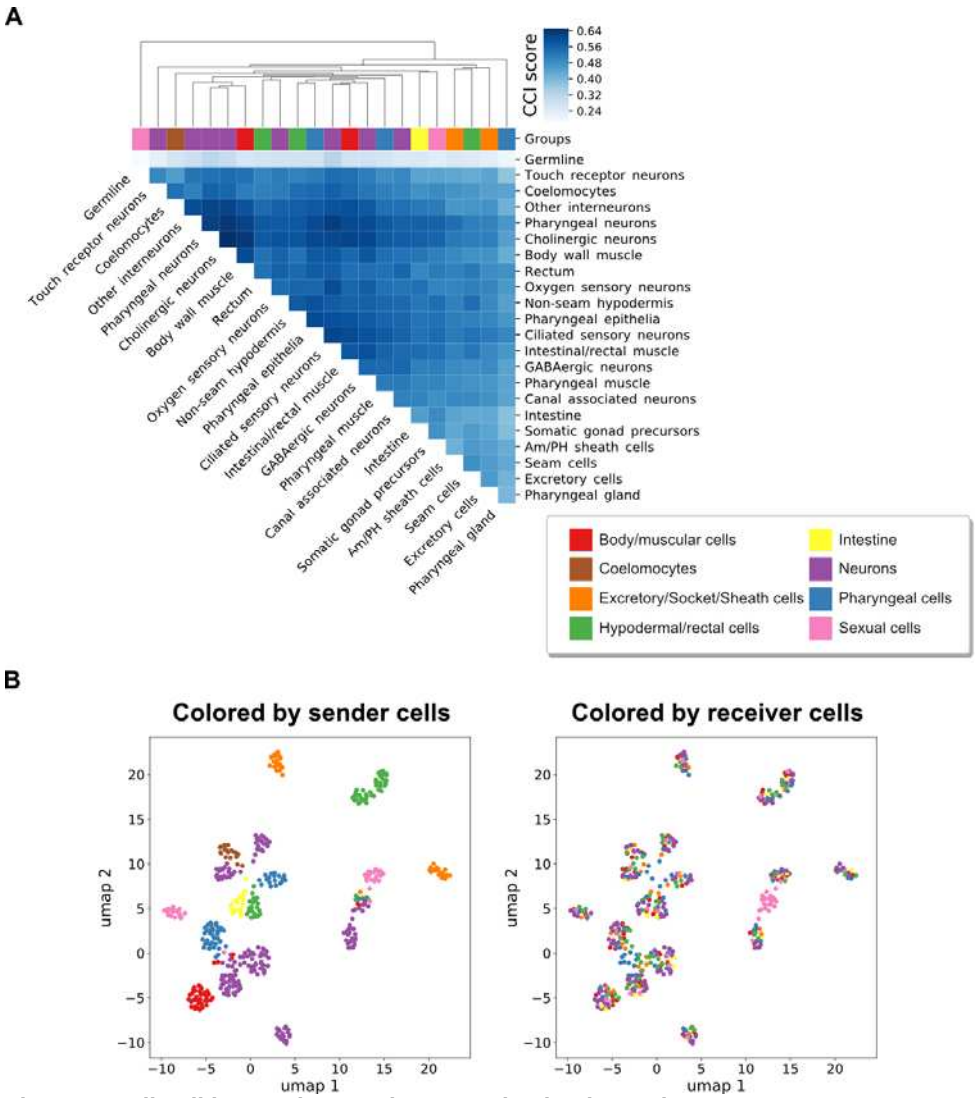


Figure 2. Cell-cell interactions and communication in *C. elegans*.
(A) Heatmap of CCI scores obtained for each pair of cell types using the curated list of LR pairs. An agglomerative hierarchical clustering was performed on a dissimilarity-like metric by taking the complement (1-score) of CCI scores, disregarding autocrine interactions. Cell types are colored by their lineages as indicated in the legend. Lineages and colors were assigned previously (Cao et al., 2017). (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 Supplemental Table 2, 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 should decrease the potential of cells to interact, we evaluated the relationship of our undirected CCI score with the Euclidean distance between cell pairs (Supplemental Notes). As expected, the correlation coefficient was negative (Spearman = -0.21; P-value = 0.0016). However, due to its low value, we also hypothesized that there is a subset of key LR pairs encoding spatial organization. To identify the LR subset maximizing this correlation, we used a genetic algorithm (Supplemental Notes). Using this approach, we found 100 candidate subsets from independent runs that led to an average

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

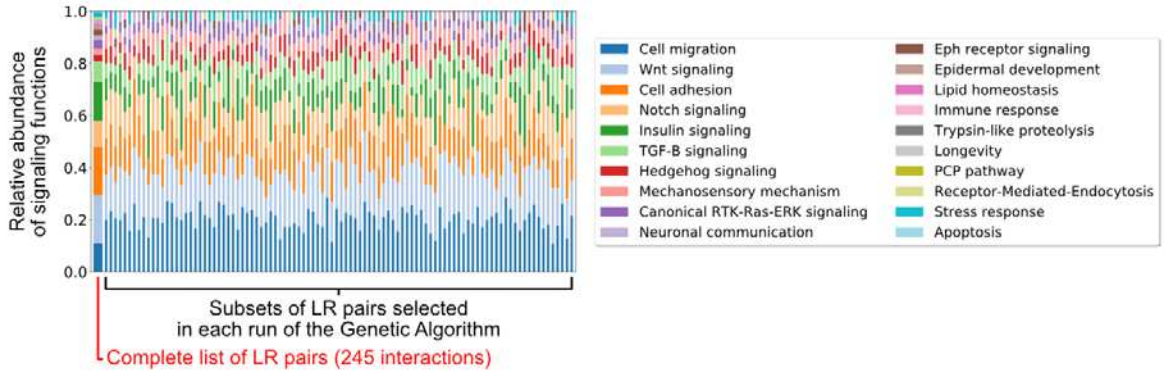
Spearman coefficient of -0.67 ± 0.01. The genetic algorithm-optimized subsets of LR pairs (hereinafter referred to as 'initial GA-LR pairs') may therefore constitute good predictors of biological functions driving or sustaining intercellular proximity.

Using our literature-based functional annotations (see column "LR Function" in Supplemental Table 1), we next investigated the specific biological roles of the initial GA-LR pairs. Specifically, for each initial GA-LR pair list, we computed the relative abundance of each signaling pathway (Figure 3A). Considering the relative abundance of these pathways in our complete list containing the 245 LR pairs, we used the distribution of abundances from the GA runs and performed a one-sided Wilcoxon signed-rank test to evaluate whether the fractions of each function either increased or decreased with respect to the fraction in all LR pairs (Figure 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 seems to prioritize LR pairs associated with processes such as cell patterning, morphogenesis, and tissue maintenance (Ingham et al., 2011).

We next looked for a consensus core set of LR pairs representing the optimal solutions generated by the GA (Supplemental Notes). We found a list containing 37 LR pairs (Supplemental Table 3), 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; correlation that is not a result of randomly selecting LR pairs (P-value < 0.0001, Supplemental Notes). While the CCI scores computed from the complete list of LR pairs led to functional interactions of cell types (Figure 2A), the GA-LR pairs seem to group cell types by more specific associations that may be attributable to their spatial localization (Figure 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, this list grouped 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 (Hussey et al., 2018; Noble et al., 2013; Witham et al., 2016). Thus, the PPIs prioritized by the GA capture cellular properties that define intercellular proximity, especially defining functional roles of tissues and organs.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A



B

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

Figure 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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

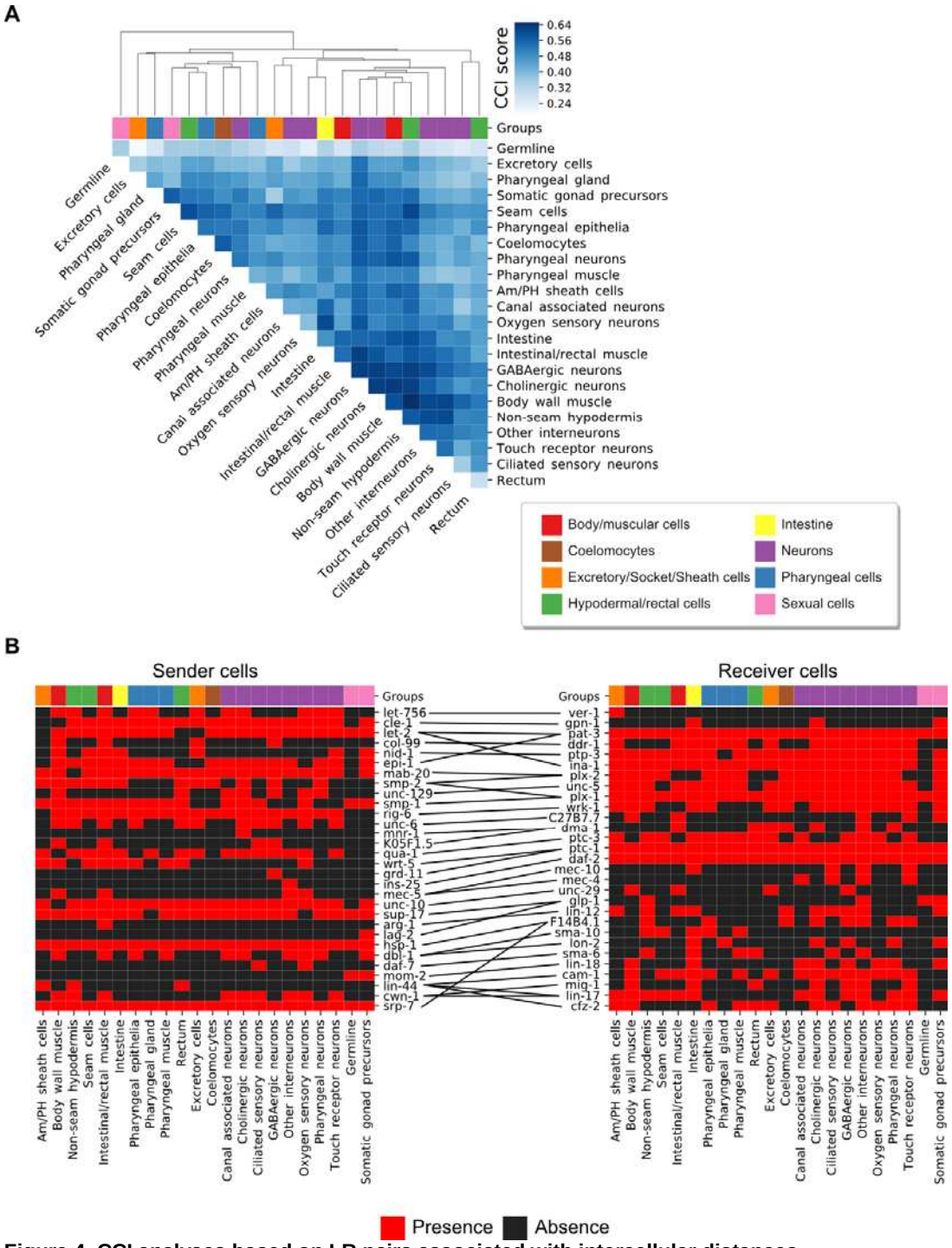


Figure 4. CCI analyses based on LR pairs associated with intercellular distances.
(A) Heatmap of CCI scores obtained for each pair of cells using the consensus GA-LR pairs. An agglomerative hierarchical clustering was performed on a dissimilarity-like metric by taking the complement of CCI scores (1-score), 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).

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

To explore the spatial distribution of the GA-LR pairs along the body of *C. elegans* (Figure 5), we performed an enrichment analysis of CCI mediators along its body. We first divided the *C. elegans* body in 3 sections, encompassing different cell types (Figure 5A). Then, we 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 (Figure 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. For instance, *col-99* shows prominent expression in the head, especially during L1-L2 larvae stages of development (Tu et al., 2015), while LIN-44 is secreted by hypodermal cells exclusively in the tail during larval development (Gleason et al., 2006; Klassen and Shen, 2007), 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 demonstrate our strategy captures true biological spatial behaviors of gene expression and therefore of CCIs.

To better understand the importance of the GA-LR pairs in defining 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 (Figure 5C). 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%). 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 (Lander, 2013). For example, in *C. elegans*, *cwn-1* and *lin-44* follow a gradient along its body, enabling cell migration (Harterink et al., 2011; Herman et al., 1995; Pan et al., 2006; Whangbo and Kenyon, 1999). 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 (Miller and Chin-Sang, 2012), their mutants show cell positioning defects, especially along the anterior/posterior axis of *C. elegans* (Ginzburg et al., 2002; Nakao et al., 2007), affecting axon guidance, cell migration (Ikegami et al., 2004), epidermal and vulval morphogenesis (Liu et al., 2005; Wang et al., 2008). 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 relies on the assumption that cell-type localization in tissues may have evolved towards optimizing CCIs and functions. Therefore, we postulate that these LR pairs may be associated with crucial traits or phenotypes that can be attributed to such facilitation. To test this, we performed a phenotype enrichment analysis for *C. elegans* (Angeles-Albores et al., 2018) 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 3.84 (Fisher's exact test; adj. P-value = 0.0074). According to WormBase (Harris et al., 2010), 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 (Figure S1). Thus, we show a relationship between CCIs and tissue organization. Moreover, this evidence may be extrapolated to the orthologs of these genes, wherein perturbation of their expression might affect organismal functions and result in disease in higher organisms.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

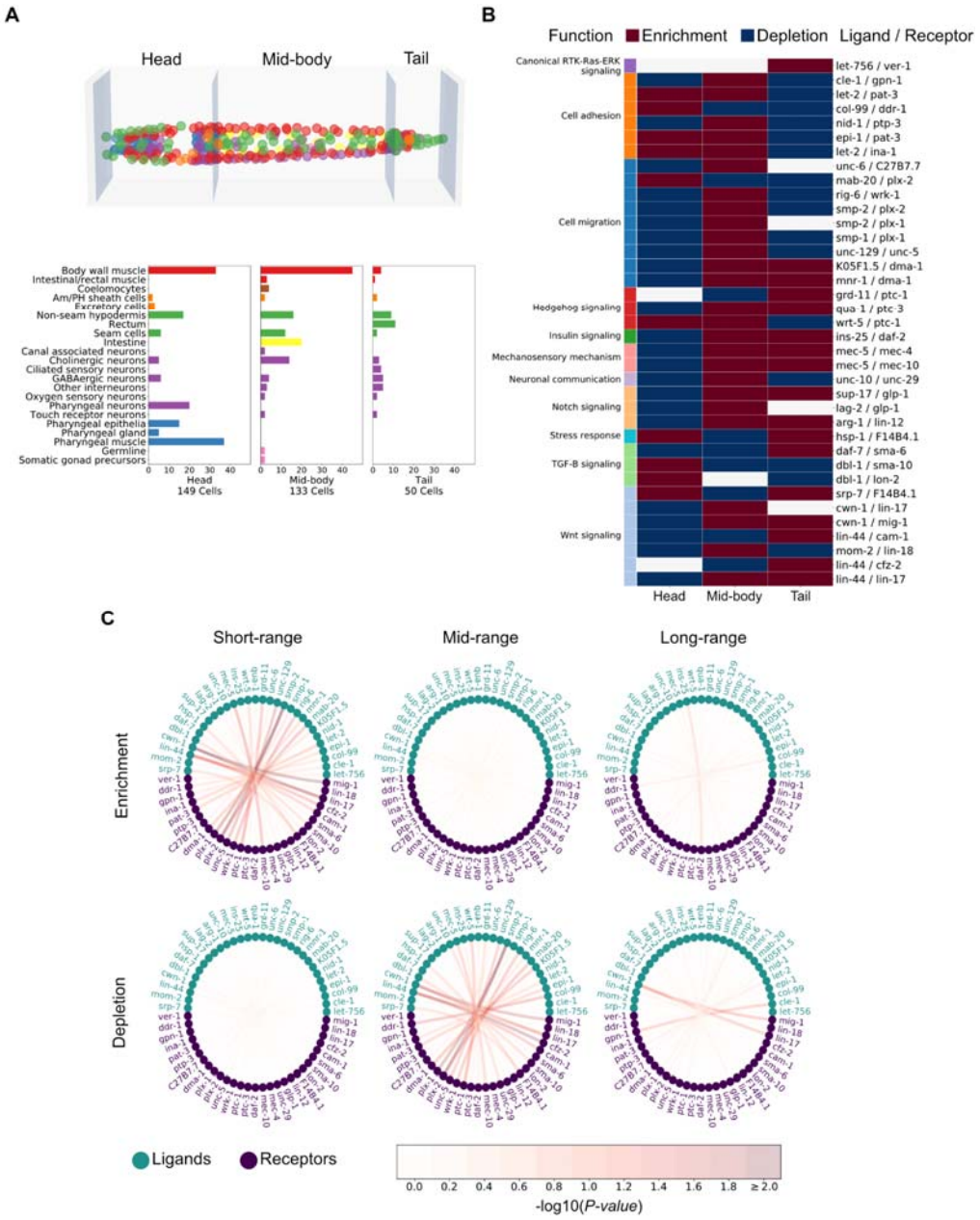


Figure 5. Spatial enrichment and depletion of communication pathways.
(A) To study the anteroposterior use of communication pathways, the body of *C. elegans* was divided into three sections along the anteroposterior axis (top) and cell-type composition of each section (bottom) given a previously published 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) according to the legend in Figure 3A. (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 Figure S1C. Nodes represent ligands or receptors and edges connect those ligands and receptors that interact in the GA-LR pairs (Supplemental Table 3). The color of the nodes represents 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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

As demonstrated here, our computational framework is expected to identify key LR interactions driving the spatial organization of cell-cell interactions. To test whether the LR pairs selected by our algorithm are actually co-expressed in the cell pairs participating in the predicted CCIs, 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; Supplemental Table 4). We further used single-molecule Fluorescent *In Situ* Hybridization (smFISH) and tested if the encoding genes of the 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 Methods for selection criteria).

Confirming the predictions of our algorithm, we found the ligand and receptor genes expressed in spatially proximal cells, suggesting that the signaling through these pathways is occurring as expected in the specific cell types. In particular, our *in situ* results confirmed our computational prediction that *arg-1* in the intestinal/rectal muscle interacts with *lin-12* in non-seam hypodermal cells (Figure 6A). They also confirmed that *let-756* in the non-seam hypodermal cells of the head interacts with *ver-1* in the amphid sheath cells (Figure 6B), and that *lin-44* in the non-seam hypodermal cells of the tail interacts with *lin-17* in the seam cells of the tail (Figure 6C). Consistently, the pairs of cells wherein these LR pairs were analyzed presented CCI scores of 0.61, 0.58 and 0.64, respectively. Remarkably, these values are among the highest CCI scores (Figure 4A), which further supports the notion that higher values of our Bray-Curtis-like score represent a higher potential of cells to be spatially proximal. Additionally, we performed 3D projections of the smFISH images (Supplemental Videos 1-3), 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 (UniProt Consortium, 2021).

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

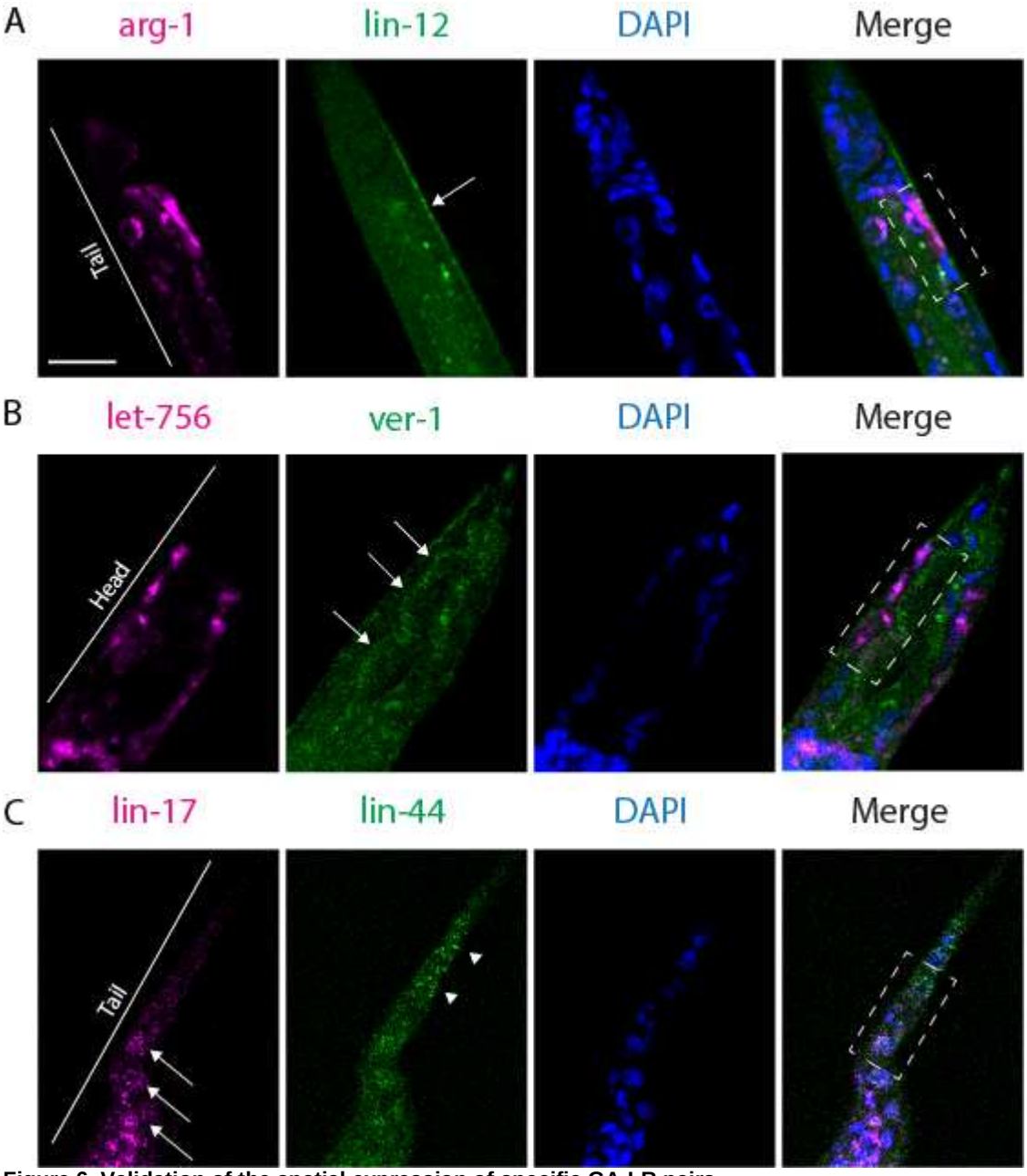


Figure 6. Validation of the spatial expression of specific GA-LR pairs.
Single-molecule Fluorescent In Situ Hybridization of genes encoding three GA-LR pairs in *C. elegans* L2 larvae. (A) Intestinal/rectal cells expressing *arg-1* (magenta) and non-seam hypodermal cells (arrow) expressing *lin-12* (green) are adjacent (see rectangle in the merge channel and Supplemental Video 1). (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 Supplemental Video 2). (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 Supplemental Video 3). In all cases a DAPI staining was performed to distinguish cell nuclei. Scale bar = 10µm.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Discussion

Here we present a computational strategy for quantifying the potential of cells to interact and communicate. This approach quantifies ligand and receptor co-expression between cells by integrating scRNA-seq data and extra information such as intercellular distances. We applied this approach in *C. elegans* to find the molecular interactions associated with spatial organization of CCIs. In particular, we found a negative Spearman coefficient between intercellular distance and CCIs computed with our Bray-Curtis-like score. By optimizing this correlation with a genetic algorithm, we found ligand-receptor pairs that are key for encoding properties associated with impact on organismal-level phenotypes. Thus, we propose that the optimized LR pairs transmit phenotypically critical information from cell to cell in a space-dependent fashion.

In this study, we collected ligand-receptor interactions of *C. elegans* from literature and PPI databases and built an essential resource for CCI analyses of *C. elegans* (Supplemental Table 1). Using this list, our CCI analysis identified a core set of LR pairs associated with spatial patterning. For example, we found that interacting cells were grouped based on cell type-specific production of ligands (Figure 2B), which is consistent with the principles underlying a communication network reported for human haematopoietic cells (Qiao et al., 2014). Our results are also consistent with experimental studies of *C. elegans* (Supplemental Table 4). 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 (Figure 3B), and mediators that are crucial for the larval development of *C. elegans*, especially in processes associated with Notch and TGF-β signaling, cellular positioning, and organ morphogenesis, which are particularly active at the cognate stages of the datasets we used (Keil et al., 2017; Lažetić and Fay, 2017; Ou and Vale, 2009; Tuck, 2014). Remarkably, these main roles of the GA-LR pairs contribute to defining the spatial properties of tissues and organs. Thus, this evidence may explain why the genetic algorithm prioritized a core list of LR pairs whose active/inactive combination seems to define a spatial code across the *C. elegans* body.

Our analysis successfully recapitulated the spatial organization of the cells in *C. elegans* and associated ligand-receptor interactions. The roles of the LR pairs selected by the genetic algorithm include determination and maintenance of cell-to-cell distances. For instance, the GA-LR pairs enriched in short distance interactions (Figure 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* (Ginzburg et al., 2002); and 2) *cwn-1*/*mig-1*, which mediates cell positioning, as demonstrated by the abnormal migration of hermaphrodite specific motor neurons in the mutants (Pan et al., 2006; Zinovyeva et al., 2008). Additionally, by using smFISH we experimentally showed that mediators used by cell pairs with high CCI scores (Figure 4A), such as *arg-1*/*lin-12*, *let-756/ver-1* and *lin-44/lin-12,* are expressed in spatially adjacent or proximal cells (Figure 6). While previous studies reported that hypodermal cells form a gradient of LIN-44 in the tail (Lu and Mizumoto, 2019), and that LIN-44 can affect seam cell polarity through LIN-17 (Yamamoto et al., 2011), a CCI between hypodermal and seam cells mediated by this LR pair had not been shown before. Although smFISH is not a direct proof of protein interaction, it is a good proxy for experimentally supporting CCIs (Armingol et al., 2020; Shao et al., 2020), which in our case helped to support that *lin-44/lin-17* participates in the proximal interaction between hypodermal and seam cells in the tail (Figure 6C and Supplemental Video 3). Thus, the previous reports and the smFISH results presented here demonstrate congruence between the predictions of our algorithm and the actual distance between the GA selected LR pairs in whole *C. elegans*. This not only increases the confidence in our approach and results, but also supports potential roles of LR interactions that were not previously studied in specific cell types.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Overall, our strategy captures mechanisms underlying the spatial and functional organization of cells in a manner that is consistent with experimental evidence. Nevertheless, our approach has some limitations. Conventional scRNA-seq does not preserve spatial information, so labelling 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, 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 (Figure 6C) (Harterink et al., 2011; Herman et al., 1995), but our method inferred that *lin-44* is also important in the mid-body (Figure 5B, pair *lin-44*/*lin-17*). Similarly, *daf-7* is expressed only in sensory neurons in the head (Gumienny and Savage-Dunn, 2013); 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*, which leads to spurious results wherein all sensory neurons have a similar transcriptome across the body. 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 (Figure 5B) even though it is an important communication also happening in the head (Figure 6). Finally, physical constraints, such as physical barriers between cells, are not considered in our analyses, which could lead to false positives. Examples include the Notch pathway interactions between germline cells and cells that are not the distal tip cell or sheath cells since a basal lamina physically blocks interactions with other cell types (Pazdernik and Schedl, 2013). 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, and may also explain the strong but imperfect correlation obtained between CCI scores and intercellular distances. Thus, spatial transcriptomics will be important in further study of CCIs (Larsson et al., 2021; Liao et al., 2020), since they can distinguish specific cell subtypes given their locations and physical constraints.

In summary, the strategy presented here provides a framework to associate CCIs with phenotypes and detect ligand-receptor interactions associated with cell patterning. In contrast to most previously proposed overall CCI scores (Armingol et al., 2020), ours is undirected. Thus, it can be successfully integrated with any undirected information between two cells, such as intercellular distance. When considering, for example, 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 technologies preserving spatial properties (e.g. spatial transcriptomics). In this regard, another approach that computed an undirected affinity matrix for representing CCIs successfully inferred the spatial assembly of tissues (Ren et al., 2020), supporting the value of undirected scores for recapitulating spatial properties of CCIs from traditional transcriptomes. Nevertheless, potential pitfalls could emerge given the nature of CCI analyses (Supplemental Notes). Finally, defining a pertinent objective function for the genetic algorithm can enable the use of our method for diverse studies identifying LR pairs associated with a phenotype 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, which could be extended to study normal and diseased states in higher organisms.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Methods

Single-cell RNA-seq data

A previously published single-cell RNA-seq dataset containing 27 cell types of *C. elegans* in the larval L2 stage was used as a transcriptome (Cao et al., 2017). 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 (Cao et al., 2017), 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 (Long et al., 2009). 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 (Figure N1A in Supplemental 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 (Ramilowski et al., 2015) was used as reference for looking for respective orthologs in *C. elegans*. The search for orthologs was done using OrthoDB (Waterhouse et al., 2013), OrthoList (Kim et al., 2018) and gProfiler (Reimand et al., 2007). Then, a network of protein-protein interactions for *C. elegans* was obtained from RSPGM (Huang et al., 2016) and high-confidence interactions in STRING-db (confidence score > 700 and supported at least by one experimental evidence) (Szklarczyk et al., 2019). 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 (Supplemental Table 5). 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 (Supplemental Table 1).

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 (Ramilowski et al., 2015). 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 Figure 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

while also incorporating the potential that each cell has to communicate independently (Figure 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 (Figure 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 Figure N2A in Supplemental Notes). The maximization process was run 100 times, generating 100 different lists that resulted in an optimal correlation. As shown in Supplemental Notes (Figures 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 (Pedregosa et al., 2011) and a number of components equal to 3.

Statistical analyses

For each function annotated in the list of ligand-receptor pairs (Supplemental Table 1), 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 10,000 times.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 (Angeles-Albores et al., 2018)*), 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 Figure 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 (Ji and van Oudenaarden, 2012). 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 (Figure S2). Additionally, imaging of semo-1::GFP (BC12890) and ajm-1::GFP (PS3729) (Figure S3), which have previously been used to define the location of the hypodermal cells (Altun and Hall, 2002), 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 (Schindelin et al., 2012).

Data and code availability

The single-cell RNA-seq dataset was obtained from the Supplemental Table S4 in (Cao et al., 2017). The 3D digital atlas of *C. elegans* was obtained from the Supplemental Data 1 in (Long et al., 2009). An annotated version of this atlas with the cell types in the scRNAseq dataset is available in the Supplemental Table 6 of this work. The manual curated list containing 245 ligand-receptor interactions is present in the Supplemental Table 1 of this work, while the consensus list from the GA-selection, which contains 37 interactions, is available in the Supplemental Table 3.

All analyses performed in this work, their respective codes (implemented in Python and Jupyter Notebooks) and instructions to use them are available in a public repository (https://github.com/LewisLabUCSD/Celegans-cell2cell). This repository also includes all the input datasets employed in the analyses. While *cell2cell* is an open source suite 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).

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

Altun, Z.F., and Hall, D.H. (2002). WormAtlas hermaphrodite handbook - epithelial system - hypodermis. WormAtlas.

Angeles-Albores, D., Lee, R.Y.N., Chan, J., and Sternberg, P.W. (2018). Two new functions in the WormBase enrichment suite. microPublication Biology.

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

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

Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. Nat. Biotechnol.

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

Camp, J.G., Sekine, K., Gerber, T., Loeffler-Wirth, H., Binder, H., Gac, M., Kanton, S., Kageyama, J., Damm, G., Seehofer, D., et al. (2017). Multilineage communication regulates human liver bud development from pluripotency. Nature *546*, 533–538.

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

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

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

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

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

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

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

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

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

Harris, T.W., Antoshechkin, I., Bieri, T., Blasiar, D., Chan, J., Chen, W.J., De La Cruz, N., Davis, P., Duesbury, M., Fang, R., et al. (2010). WormBase: a comprehensive resource for nematode research. Nucleic Acids Res. *38*, D463–D467.

Harterink, M., Kim, D.H., Middelkoop, T.C., Doan, T.D., van Oudenaarden, A., and Korswagen, H.C. (2011). 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.

Herman, M.A., Vassilieva, L.L., Horvitz, H.R., Shaw, J.E., and Herman, R.K. (1995). 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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Hou, R., Denisenko, E., Ong, H.T., Ramilowski, J.A., and Forrest, A.R.R. (2020). Predicting cell-to-cell communication networks using NATMI. Nat. Commun. *11*, 1–11.

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

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

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

Ikegami, R., Zheng, H., Ong, S.-H., and Culotti, J. (2004). 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.

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

Ji, N., and van Oudenaarden, A. (2012). Single molecule fluorescent in situ hybridization (smFISH) of C. elegans worms and embryos. WormBook.

Joost, S., Zeisel, A., Jacob, T., Sun, X., La Manno, G., Lönnerberg, P., Linnarsson, S., and Kasper, M. (2016). Single-Cell Transcriptomics Reveals that Differentiation and Spatial Signatures Shape Epidermal and Hair Follicle Heterogeneity. Cell Syst *3*, 221–237.e9.

Joost, S., Jacob, T., Sun, X., Annusver, K., La Manno, G., Sur, I., and Kasper, M. (2018). Single-Cell Transcriptomics of Traced Epidermal and Hair Follicle Stem Cells Reveals Rapid Adaptations during Wound Healing. Cell Rep. *25*, 585–597.e7.

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

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

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

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

Krausgruber, T., Fortelny, N., Fife-Gernedl, V., Senekowitsch, M., Schuster, L.C., Lercher, A., Nemc, A., Schmidl, C., Rendeiro, A.F., Bergthaler, A., et al. (2020). Structural cells are key regulators of organ-specific immune responses. Nature.

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

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

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

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

Liao, J., Lu, X., Shao, X., Zhu, L., and Fan, X. (2020). Uncovering an Organ's Molecular Architecture at Single-Cell Resolution by Spatially Resolved Transcriptomics. Trends Biotechnol.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Lu, M., and Mizumoto, K. (2019). Gradient-independent Wnt signaling instructs asymmetric neurite pruning in C. elegans.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.

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

Nakao, F., Hudson, M.L., Suzuki, M., Peckler, Z., Kurokawa, R., Liu, Z., Gengyo-Ando, K., Nukazuka, A., Fujii, T., Suto, F., et al. (2007). 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.

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

Noble, T., Stieglitz, J., and Srinivasan, S. (2013). 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.

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

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

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

Pavličev, M., Wagner, G.P., Chavan, A.R., Owens, K., Maziarz, J., Dunn-Fletcher, C., Kallapur, S.G., Muglia, L., and Jones, H. (2017). Single-cell transcriptomics of the human placenta: inferring the cell communication network of the maternal-fetal interface. Genome Res. *27*, 349–361.

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

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

Popescu, D.-M., Botting, R.A., Stephenson, E., Green, K., Webb, S., Jardine, L., Calderbank, E.F., Polanski, K., Goh, I., Efremova, M., et al. (2019). Decoding human fetal liver haematopoiesis. Nature *574*, 365–371.

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

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

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

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

Ren, X., Zhong, G., Zhang, Q., Zhang, L., Sun, Y., and Zhang, Z. (2020). Reconstruction of cell spatial organization from single-cell RNA sequencing data based on ligand-receptor mediated self-assembly. Cell Res.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Sgro, A.E., Schwab, D.J., Noorbakhsh, J., Mestler, T., Mehta, P., and Gregor, T. (2015). From intracellular signaling to population oscillations: bridging size- and time-scales in collective behavior. Mol. Syst. Biol. *11*, 779.

Shao, X., Lu, X., Liao, J., Chen, H., and Fan, X. (2020). New avenues for systematically inferring cell-cell

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

communication: through single-cell transcriptomics data. Protein Cell.

Sheikh, B.N., Bondareva, O., Guhathakurta, S., Tsang, T.H., Sikora, K., Aizarani, N., Sagar, Holz, H., Grün, D., Hein, L., et al. (2019). Systematic Identification of Cell-Cell Communication Networks in the Developing Brain. iScience *21*, 273–287.

Sun, B., Lembong, J., Normand, V., Rogers, M., and Stone, H.A. (2012). Spatial-temporal dynamics of collective chemosensing. Proc. Natl. Acad. Sci. U. S. A. *109*, 7753–7758.

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

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

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

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

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

Wang, S., Karikomi, M., MacLean, A.L., and Nie, Q. (2019a). Cell lineage and communication network inference via optimization for single-cell transcriptomics. Nucleic Acids Res. *47*, e66.

Wang, S., Drummond, M.L., Guerrero-Juarez, C.F., Tarapore, E., MacLean, A.L., Stabell, A.R., Wu, S.C., Gutierrez, G., That, B.T., Benavente, C.A., et al. (2019b). Single cell transcriptomics of human epidermis reveals basal stem cell transition states.

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

Waterhouse, R.M., Tegenfeldt, F., Li, J., Zdobnov, E.M., and Kriventseva, E.V. (2013). OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. Nucleic Acids Res. *41*, D358–D365.

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

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

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

Xue, Y., Liu, D., Cui, G., Ding, Y., Ai, D., Gao, S., Zhang, Y., Suo, S., Wang, X., Lv, P., et al. (2019). A 3D Atlas of Hematopoietic Stem and Progenitor Cell Expansion by Multi-dimensional RNA-Seq Analysis. Cell Rep. *27*, 1567–1578.e5.

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

Yuzwa, S.A., Yang, G., Borrett, M.J., Clarke, G., Cancino, G.I., Zahr, S.K., Zandstra, P.W., Kaplan, D.R., and Miller, F.D. (2016). Proneurogenic Ligands Defined by Modeling Developing Cortex Growth Factor Communication Networks. Neuron *91*, 988–1004.

Zhong, P., Cara, J.F., and Tager, H.S. (1993). 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.

Zinovyeva, A.Y., Yamamoto, Y., Sawa, H., and Forrester, W.C. (2008). Complex network of Wnt signaling regulates

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.22.392217; this version posted July 23, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

neuronal migrations during Caenorhabditis elegans development. Genetics *179*, 1357–1371.

Acknowledgements

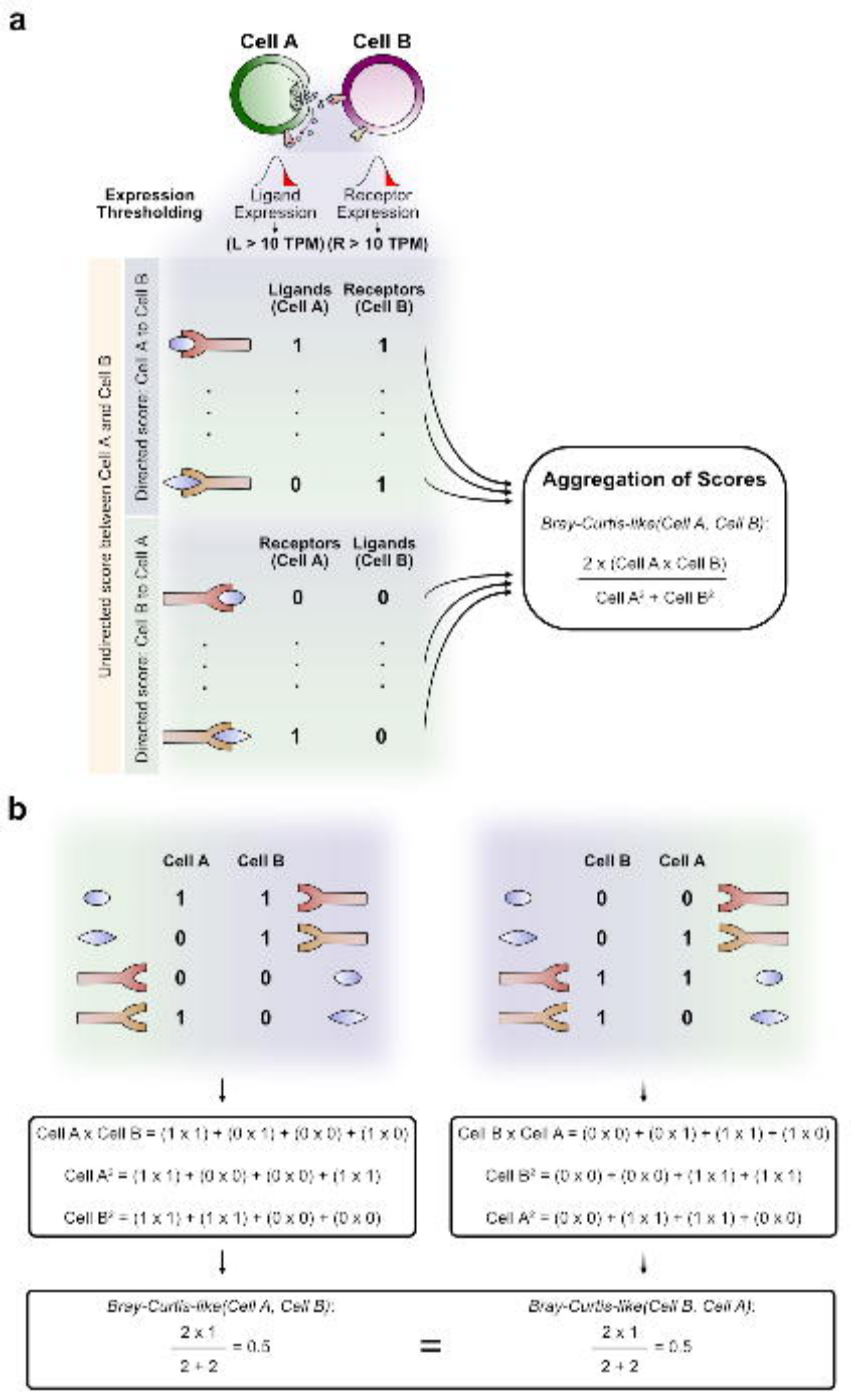
EA is supported by the Chilean Agencia Nacional de Investigación y Desarrollo (ANID) through its scholarship program DOCTORADO BECAS CHILE/2018 - 72190270 and by the Fulbright Commission. This work was further supported by NIGMS grant R35 GM119850 to NEL, a Lilly Innovation Fellows Award to CJ, Jefferson 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.

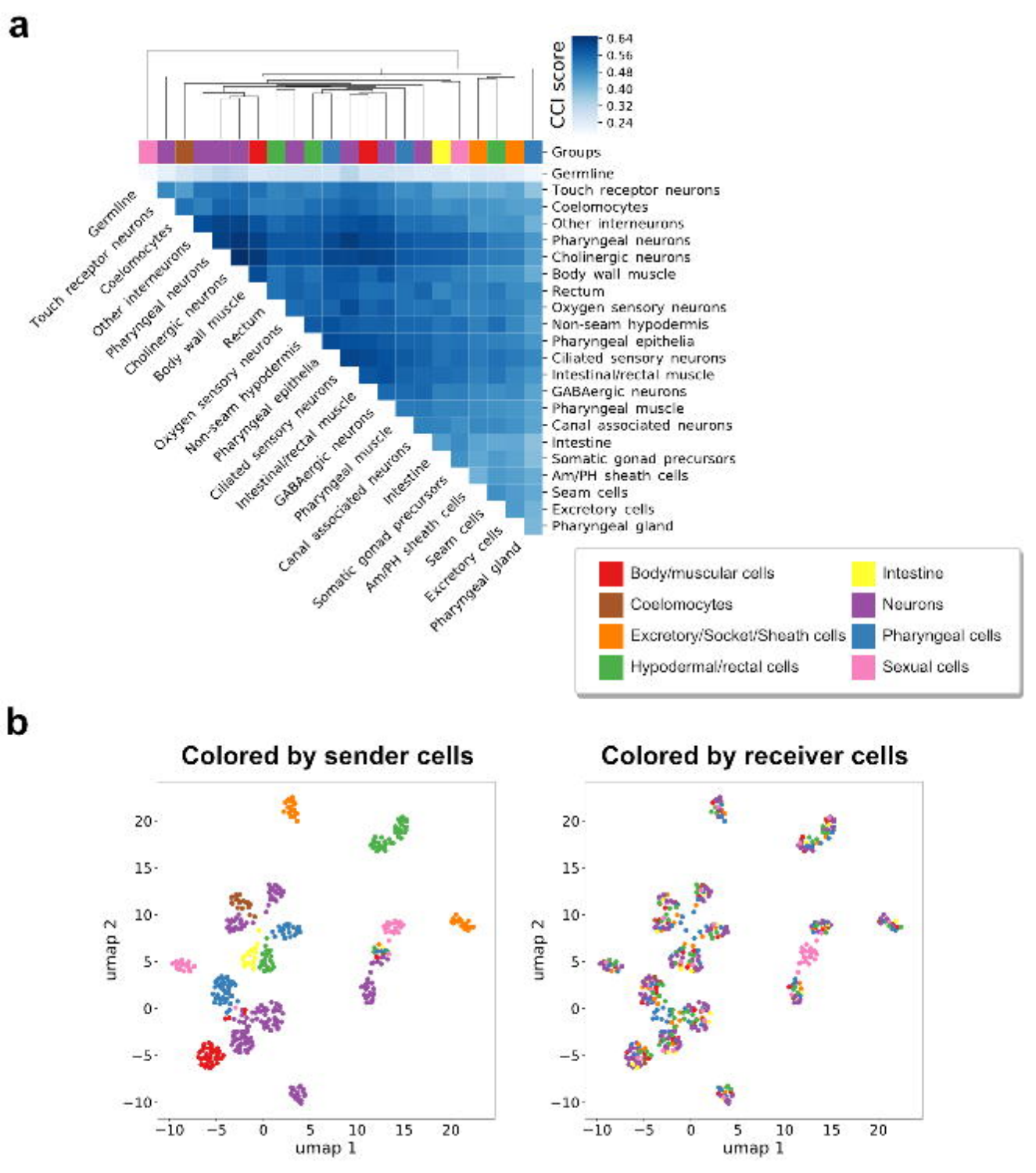
Author contributions

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 vivo* validations and AG performed the smFISH experiments. EA wrote the paper and all authors carefully reviewed and edited the paper.

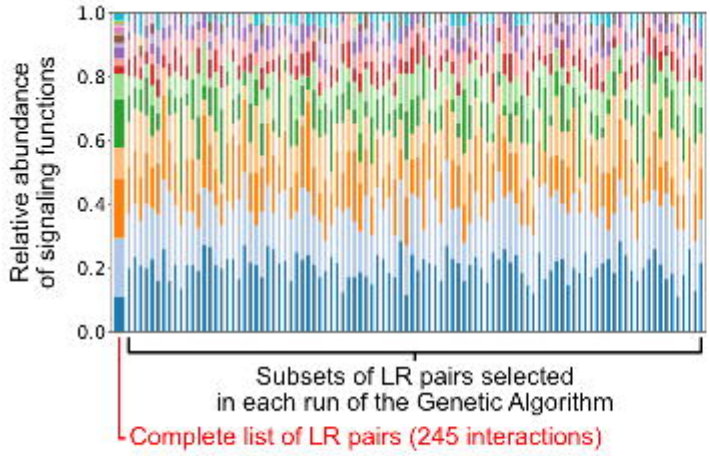
Competing interests

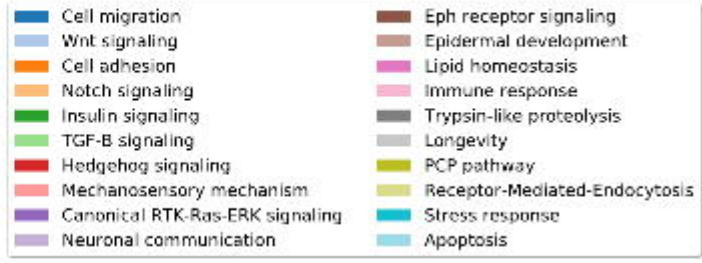
The authors declare no competing interests.


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

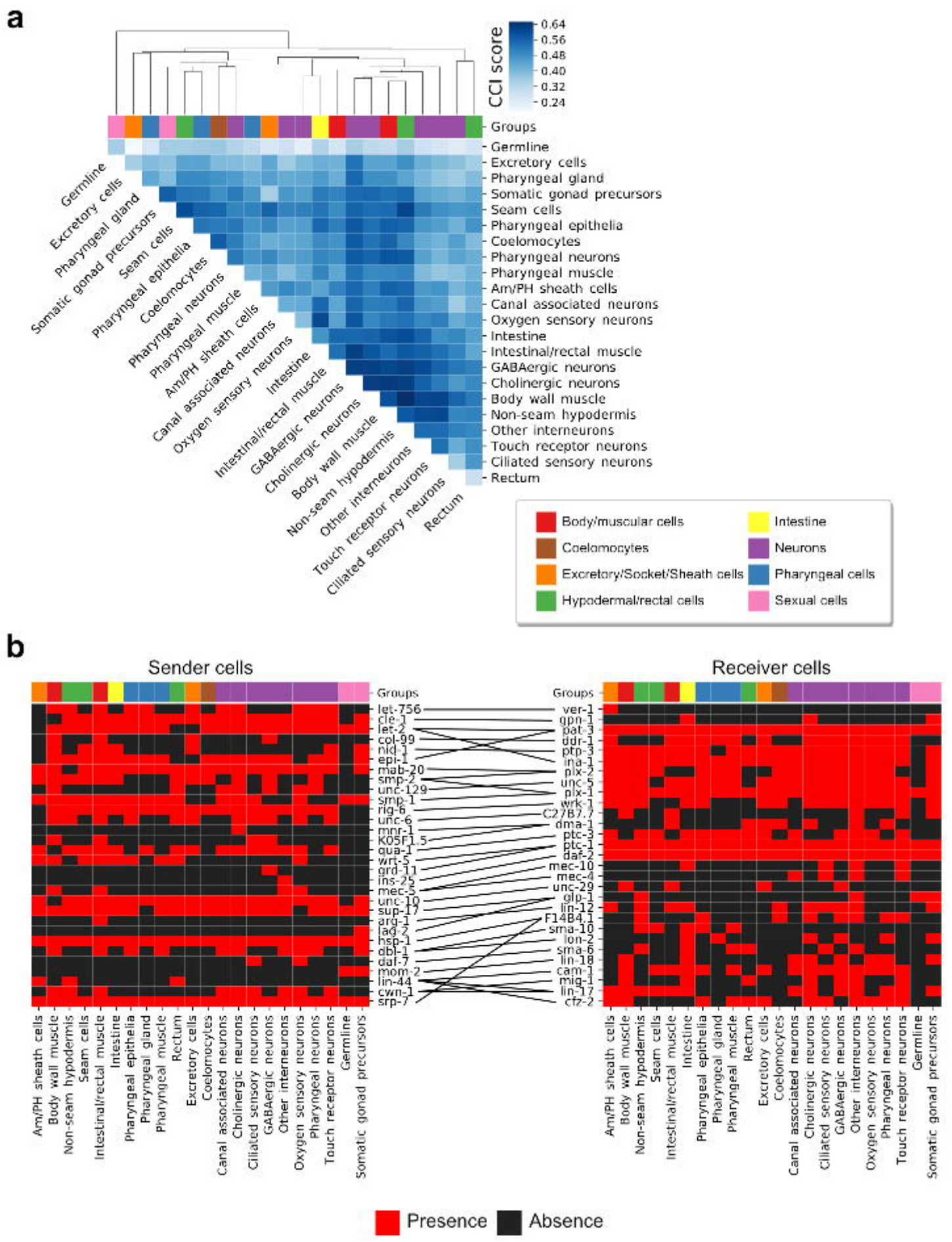


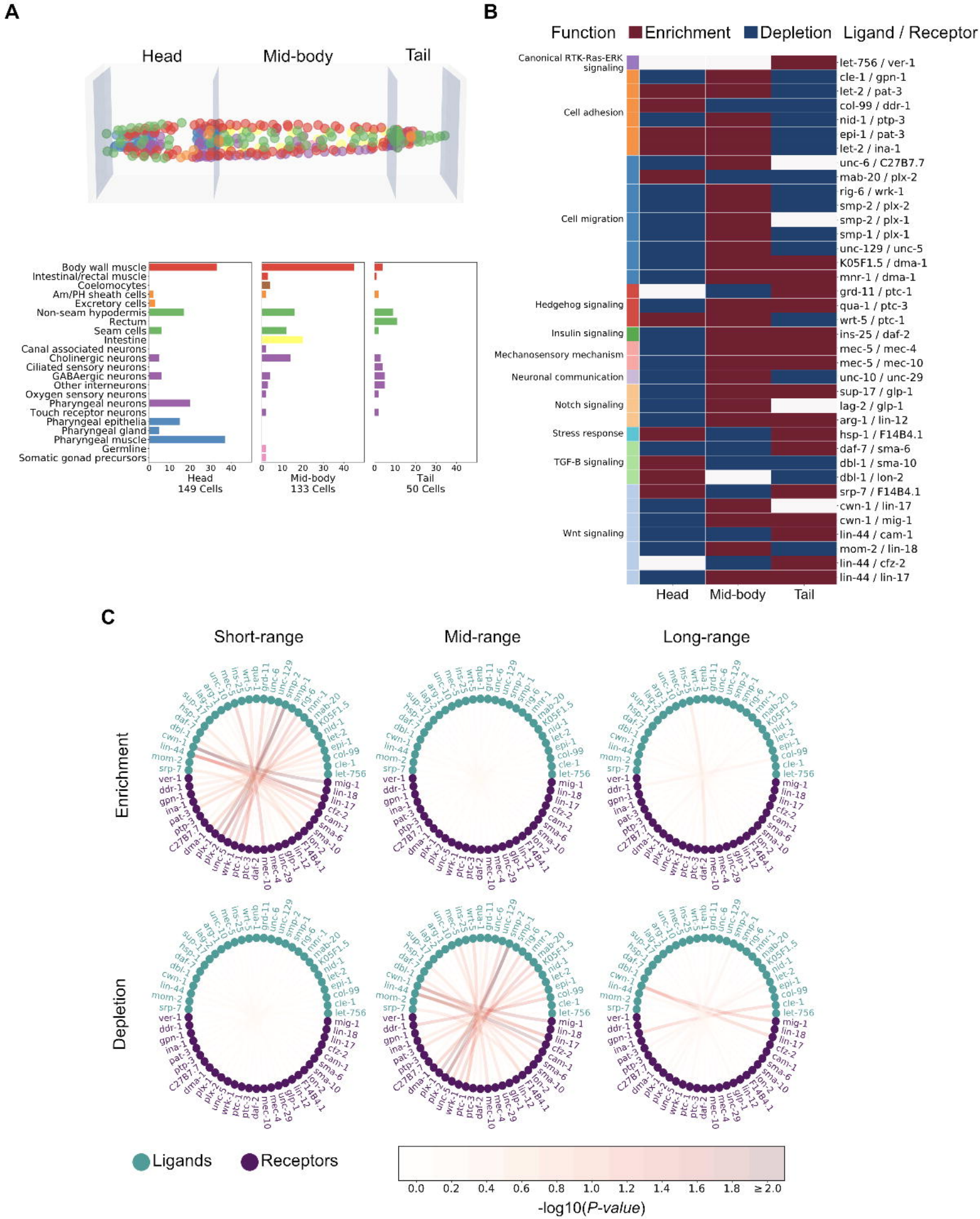


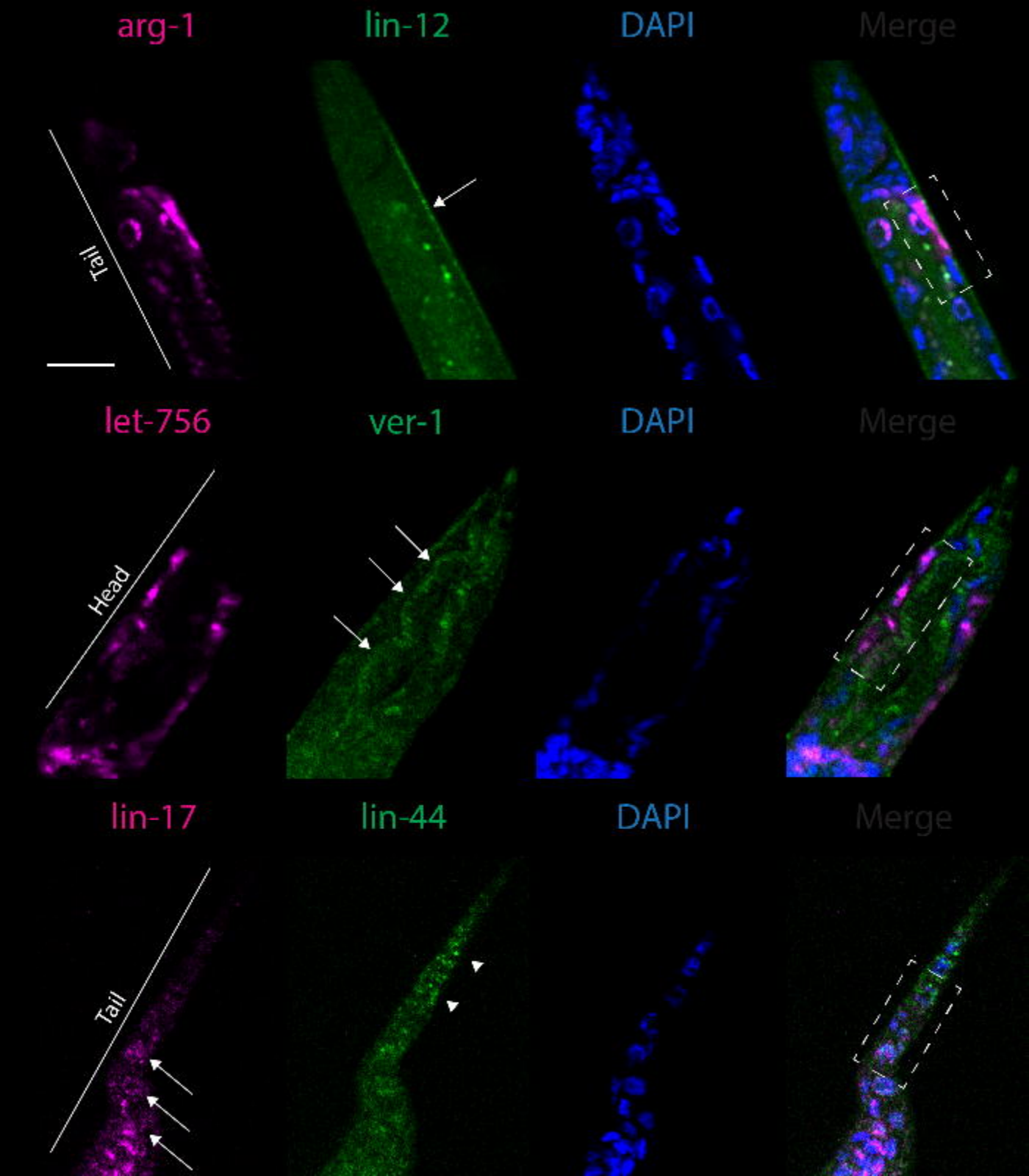
b

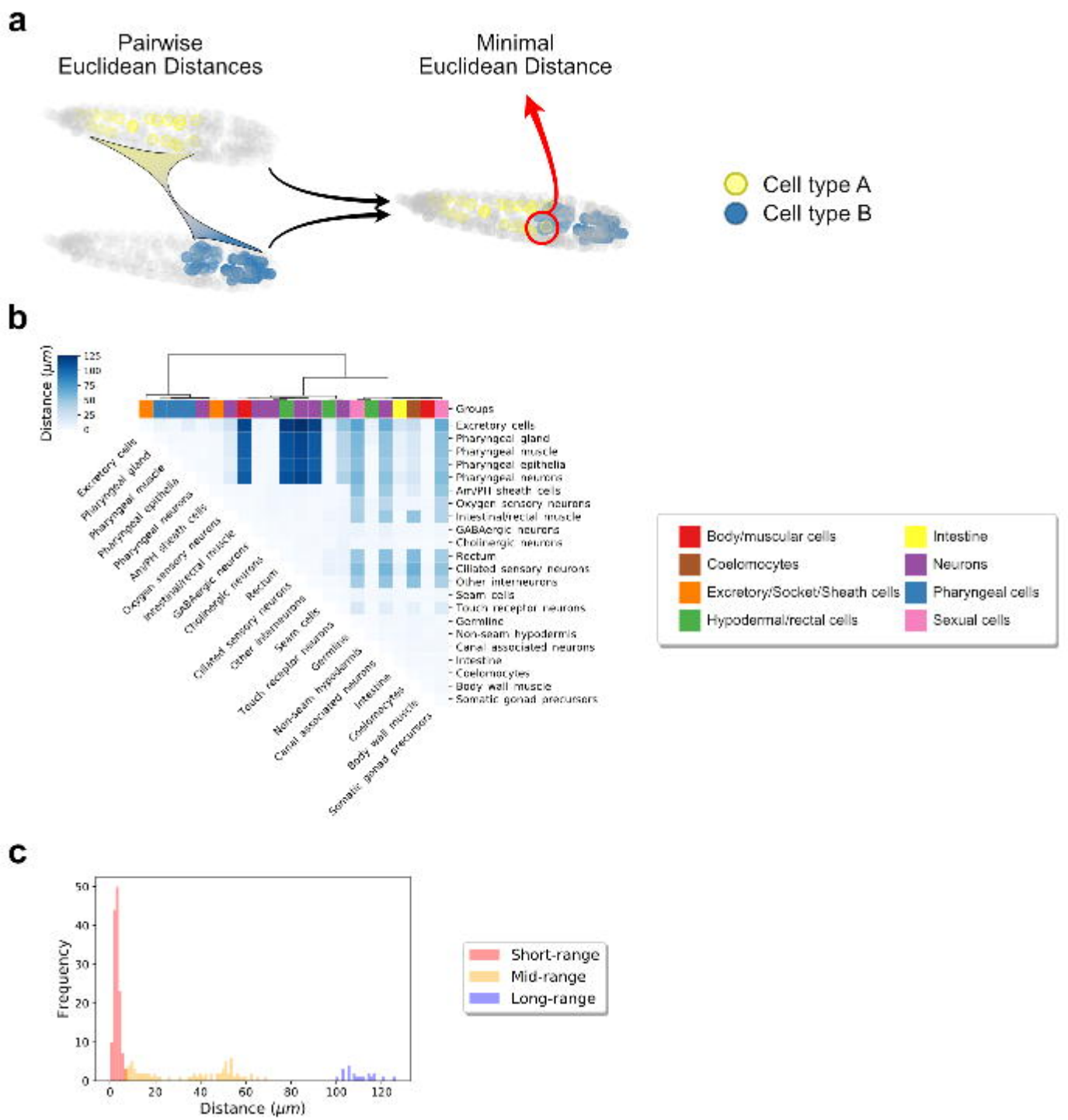
Signaling function	Change	Adj. P-value
Cell migration	↑	3.983×10^{-17}
Wnt signaling	–	7.221×10^{-01}
Cell adhesion	↓	2.711×10^{-13}
Notch signaling	–	9.341×10^{-01}
Insulin signaling	↓	8.158×10^{-13}
TGF-B signaling	–	4.759×10^{-02}
Hedgehog signaling	↑	5.439×10^{-11}
Mechanosensory mechanism	↑	8.074×10^{-17}
Canonical RTK-Ras-ERK signaling	↑	3.724×10^{-25}
Neuronal communication	–	4.431×10^{-01}

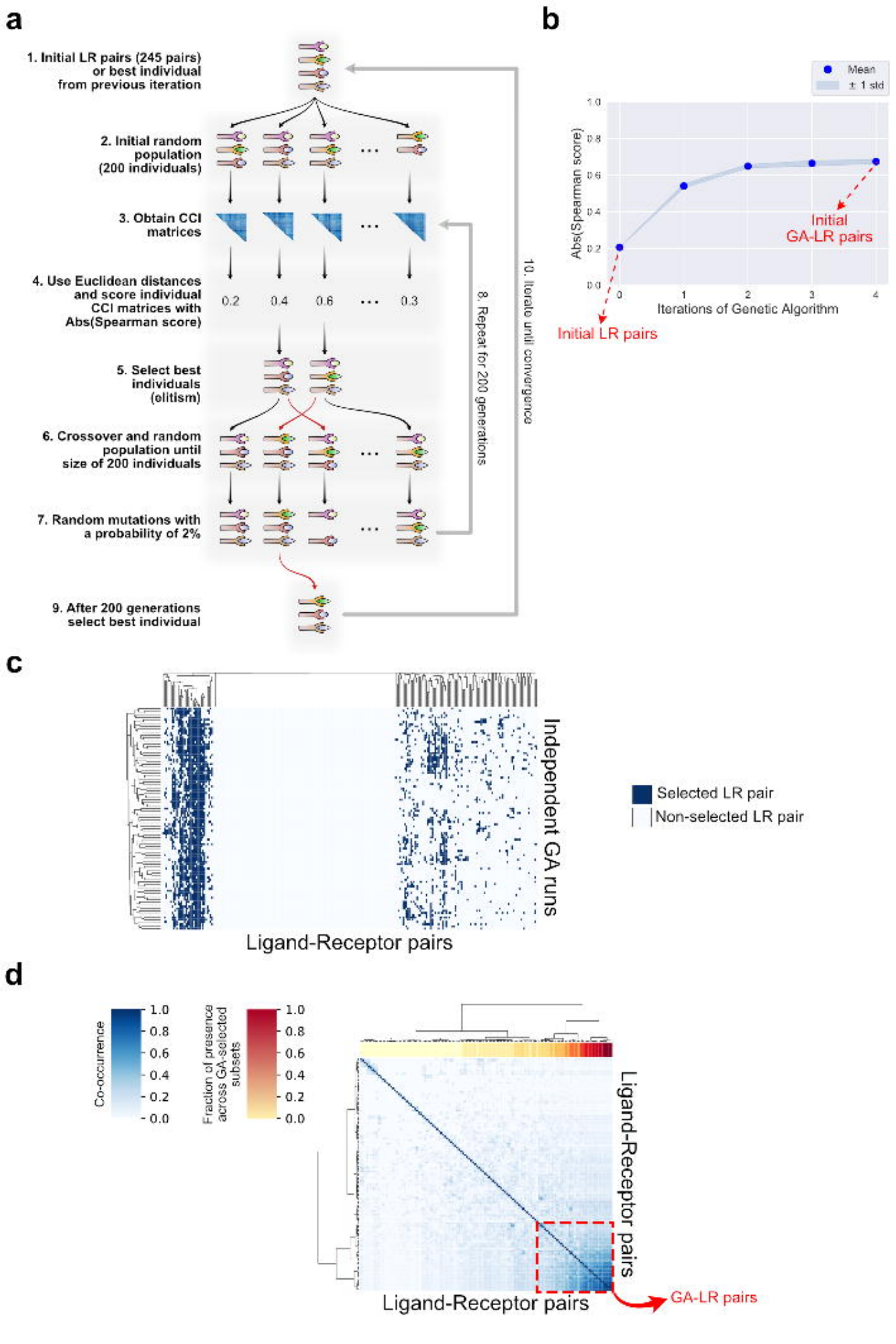
Signaling function	Change	Adj. P-value
Eph receptor signaling	↓	4.472×10^{-13}
Epidermal development	↓	3.048×10^{-23}
Lipid homeostasis	↓	3.048×10^{-23}
Immune response	↓	5.507×10^{-10}
Trypsin-like proteolysis	↓	2.478×10^{-10}
Longevity	↓	3.048×10^{-23}
PCP pathway	↓	3.048×10^{-23}
Receptor-Mediated-Endocytosis	↓	4.755×10^{-15}
Stress response	–	2.684×10^{-02}
Apoptosis	↓	3.048×10^{-23}

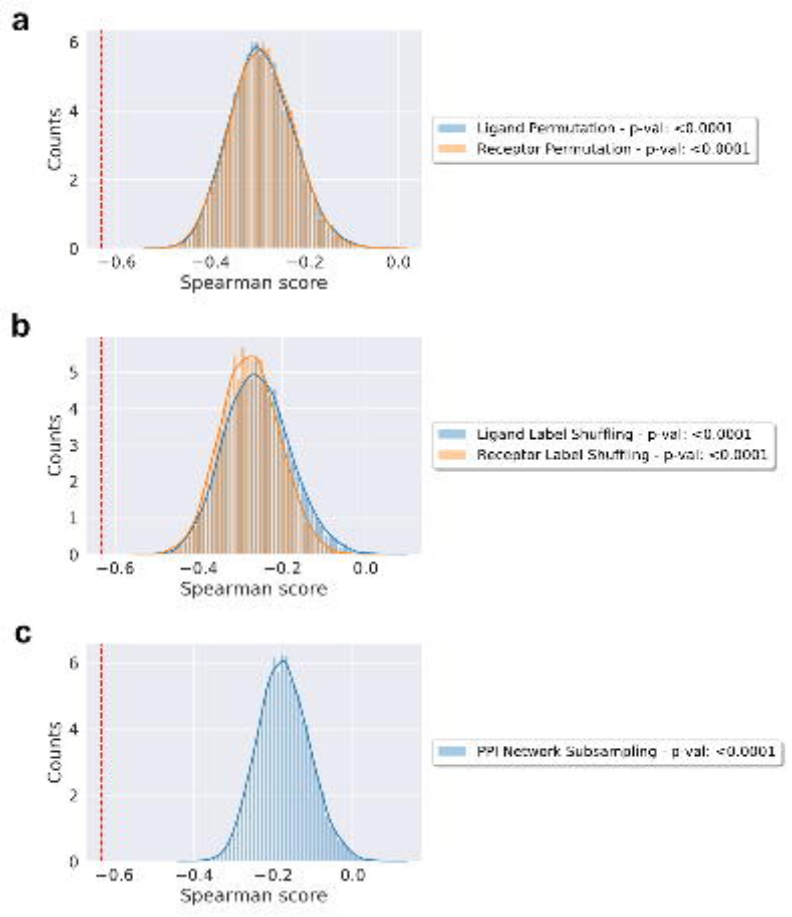

a
CCI score
0.64
0.56
0.48
0.40
0.32
0.24
Groups
Germline
Excretory cells
Pharyngeal gland
Somatic gonad precursors
Seam cells
Pharyngeal epithelia
Coelomocytes
Pharyngeal neurons
Pharyngeal muscle
Am/PH sheath cells
Canal associated neurons
Oxygen sensory neurons
Intestine
Intestinal/rectal muscle
GABAergic 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
lag-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

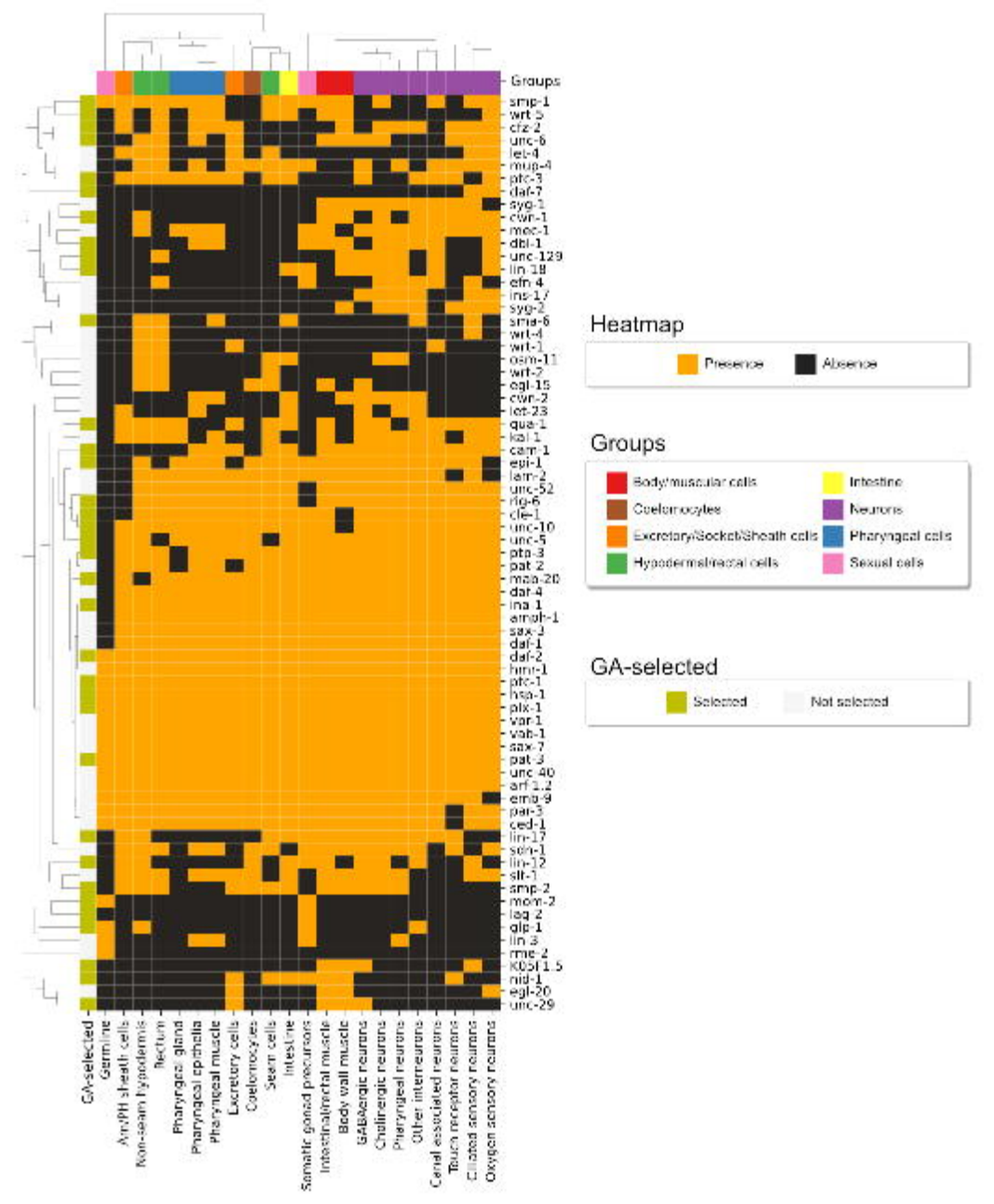

A
Head
Mid-body
Tail
Body wall muscle
Intestinal/rectal muscle
Coelomocytes
Am/PH sheath cells
Excretory cells
Non-seam hypodermis
Rectum
Seam cells
Intestine
Canal associated neurons
Cholinergic neurons
Ciliated sensory neurons
GABAergic neurons
Other interneurons
Oxygen sensory neurons
Pharyngeal neurons
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
Cell adhesion
Cell migration
Hedgehog signaling
Insulin signaling
Mechanosensory mechanism
Neuronal communication
Notch signaling
Stress response
TGF-B signaling
Wnt signaling
let-756 / ver-1
cle-1 / gpn-1
let-2 / pat-3
col-99 / ddr-1
nid-1 / ptp-3
epi-1 / pat-3
let-2 / ina-1
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
grd-11 / ptc-1
qua-1 / ptc-3
wrt-5 / ptc-1
ins-25 / daf-2
mec-5 / mec-4
mec-5 / mec-10
unc-10 / unc-29
sup-17 / glp-1
lag-2 / glp-1
arg-1 / lin-12
hsp-1 / F14B4.1
daf-7 / sma-6
dbl-1 / sma-10
dbl-1 / lon-2
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)

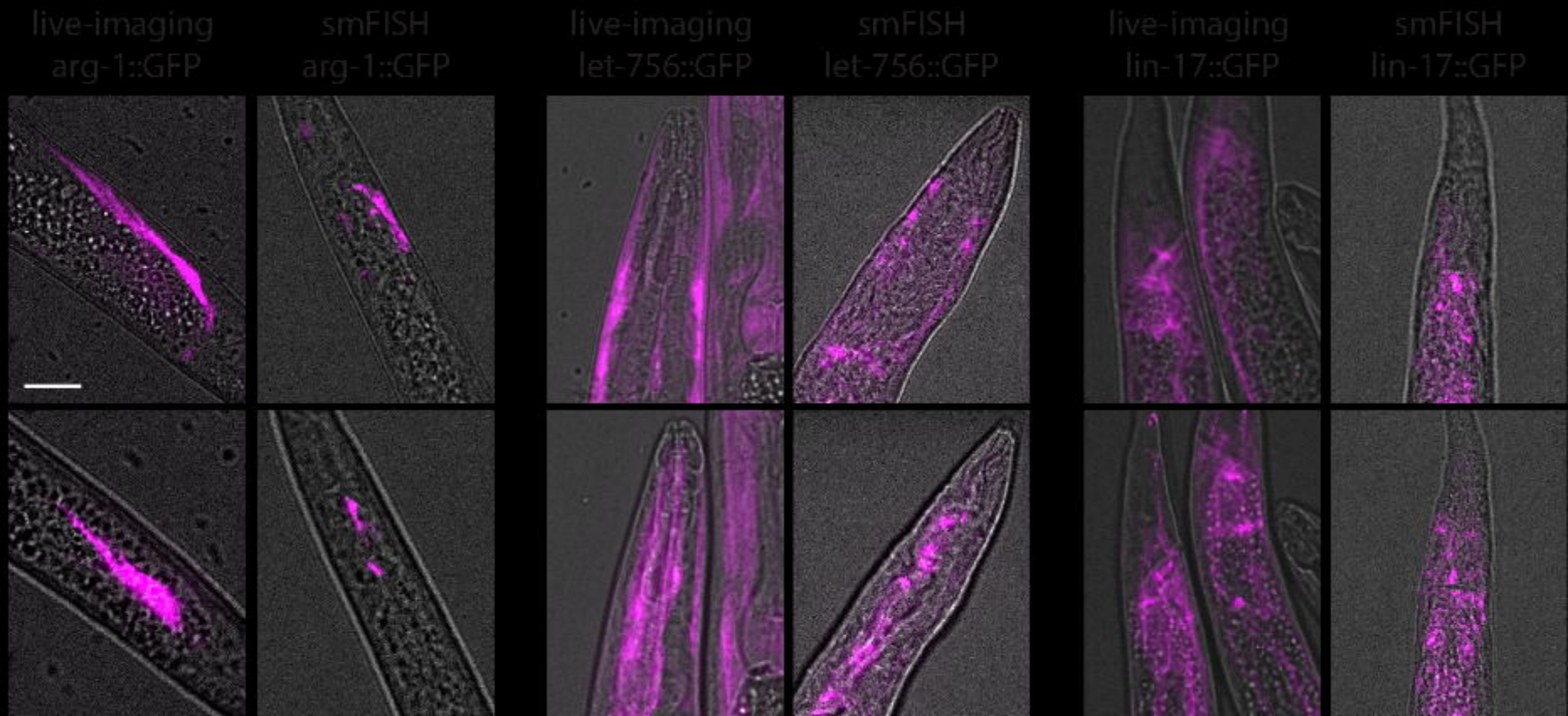
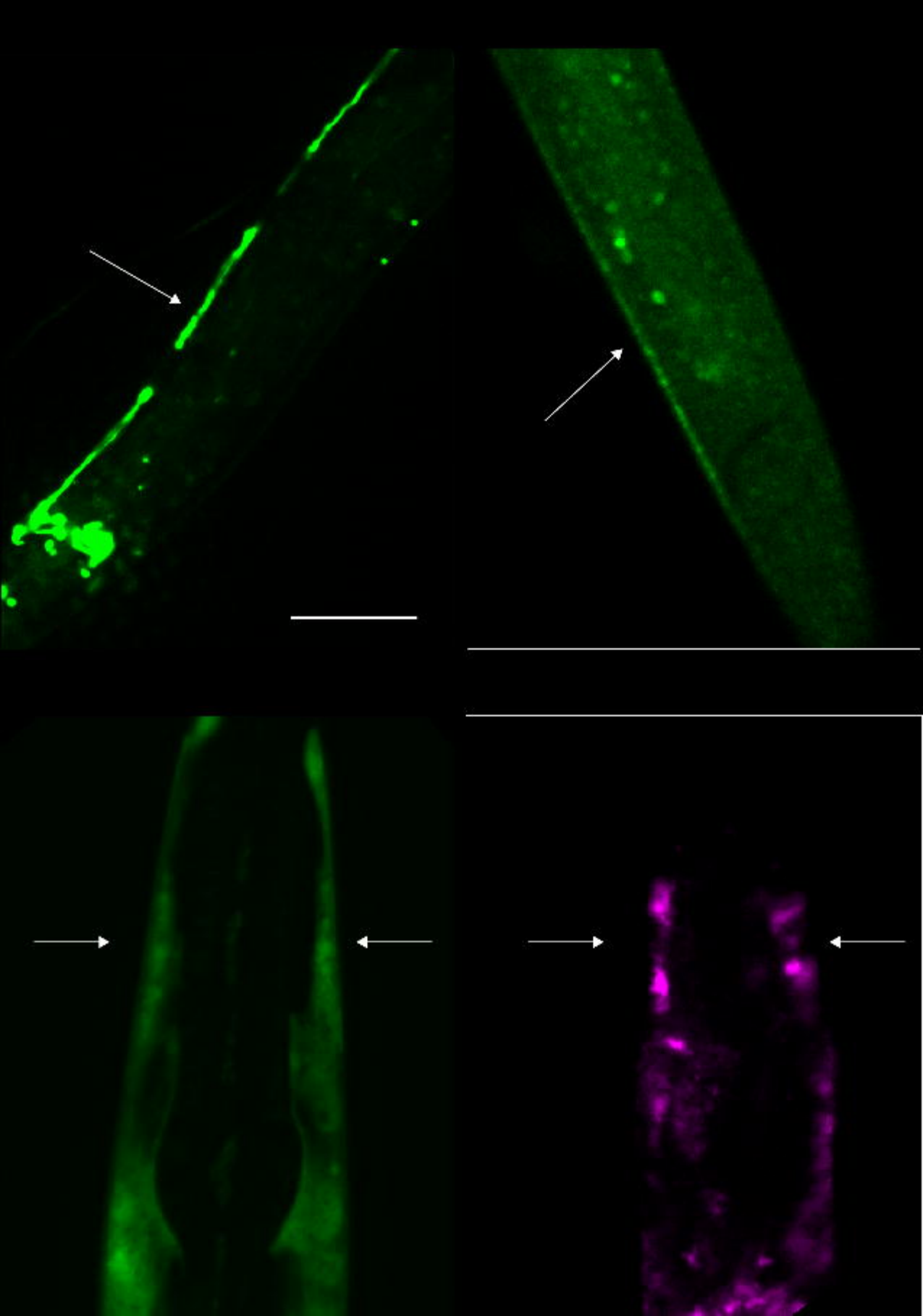

arg-1
lin-12
DAPI
Merge
Tail
let-756
ver-1
DAPI
Merge
Head
lin-17
lin-44
DAPI
Merge
Tail


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
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: <0.0001
Receptor Permutation - p-val: <0.0001
b
Counts
Spearman score
Ligand Label Shuffling - p-val: <0.0001
Receptor Label Shuffling - p-val: <0.0001
c
Counts
Spearman score
PPI Network Subsampling - p val: <0.0001


Groups
smp-1
wrt 5
cfz-2
unc-6
let-4
mup-4
ptc-3
daf-7
syg-1
cwn-1
mec-1
dbl-1
unc-129
lin 18
efn 4
ins-1/
syg-2
sma-6
wrt-4
wrt-1
osm-11
wrt-2
egl-15
cwn-2
let-23
qua-1
kal-1
cam-1
epi-1
lam-2
unc-52
rig-6
cle-1
unc-10
unc-5
pta-3
pat 2
mab-20
dar-4
ina 1
amph-1
sax-2
daf-1
daf-2
hmr-1
ptc-1
hsp-1
plx-1
var-1
vab-1
sax-7
pat-3
unc 40
arf 1.2
emb-9
par-3
ced-1
lin-17
sem-1
lin-12
slt-1
smp-2
mom-2
laq 2
glp-1
lin 3
rme-2
KO31 1.5
nid-1
egl-20
unc-29
GA-selected
Germline
Arn/PH sheath cells
Non-seam hypodermis
Rectum
Pharyngeal gland
Pharyngeal epithelia
Pharyngeal muscle
Excretory cells
Coelomocytes
Seam cells
Intestine
Somatic gonad precursors
Intestinal/rectal muscle
Body wall muscle
GABAergic neurons
Cholinergic neurons
Pharyngeal neurons
Other interneurons
Canal associated neurons
Touch receptor neurons
Ciliated sensory neurons
Oxygen sensory neurons
Heatmap
Presence
Absence
Groups
Body/muscular cells
Coelomocytes
Excretory/Socket/Sheath cells
Hypodermis/rectal cells
Intestine
Neurons
Pharyngeal cells
Sexual cells
GA-selected
Selected
Not selected


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