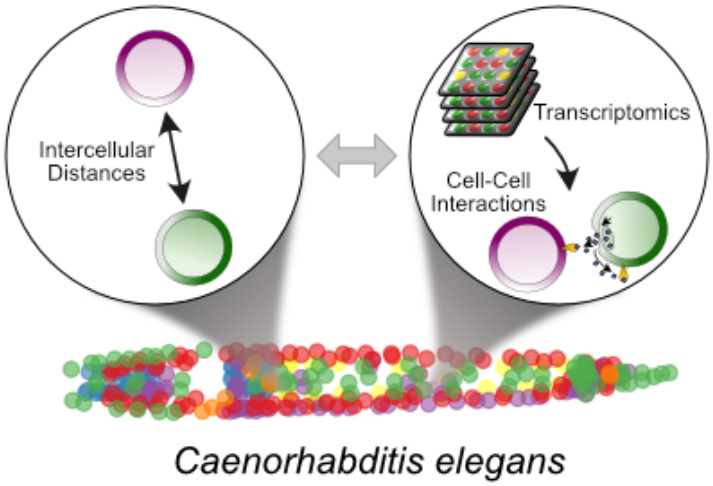
Inferring the spatial code of cell-cell interactions and communication across a whole animal body

Erick Armingol[1,2,3], Chintan J. Joshi[2], Hratch Baghdassarian[1,2], Isaac Shamie[1,2,3], Abbas Ghaddar[4], Jason Chan[5], Hsuan-Lin Her[1], Eyleen J. O'Rourke[4,6], Nathan E. Lewis[2,3,7*]

[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] Novo Nordisk Foundation Center for Biosustainability at the University of California, San Diego, La Jolla, CA 92093, USA
[4] Department of Biology, University of Virginia, Charlottesville, VA 22903, USA
[5] Poway High School, Poway, CA 92064, USA
[6] Department of Cell Biology, School of Medicine of University of Virginia, Charlottesville, VA 22903, USA
[7] Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093, USA
*Correspondence: nlewisres@ucsd.edu

Graphical abstract



Abstract

Cell-cell interactions are crucial for multicellular organisms as they shape cellular function and ultimately organismal phenotype. However, the spatial code embedded in the molecular interactions that drive and sustain spatial organization, and in the organization that in turns drives intercellular interactions across a living animal remains to be elucidated. Here we use the expression of ligand-receptor pairs obtained from a whole-body single-cell transcriptome of *Caenorhabditis elegans* larvae to compute the potential for intercellular interactions through a Bray-Curtis-like metric. Leveraging a 3D atlas of *C. elegans*' cells, we implement a genetic algorithm to select the ligand-receptor pairs most informative of the spatial organization of cells. Validating the strategy, the selected ligand-receptor pairs are involved in known cell-migration and morphogenesis processes and we confirm a negative correlation between cell-cell distances and interactions. Thus, our computational framework helps identify cell-cell interactions and their relationship with intercellular distances, and decipher molecular bases encoding spatial information in a whole animal. Furthermore, it can also be used to elucidate associations with any other intercellular phenotype and applied to other multicellular organisms.

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

Introduction

Cell-cell interactions (CCIs) are fundamental to all facets of multicellular life. They shape cellular differentiation and the functions of tissues and organs, which ultimately influence organismal physiology and behavior. Cell-cell communication (CCC) is a subtype of CCIs and involves a cell sending a signal to another cell, usually triggering downstream signaling events that culminate in altered gene expression. Thus, CCIs allow cells to coordinate their gene expression[1], form spatial patterns of interaction[2], and perform collective behaviors[3]. Signals passed between cells are often ligands which are received by receptors in other cells. Ligands can mediate CCC across a range of distances and they can encode positional information for cells within tissues, which is critical for cellular decision-making[4]. For instance, some ligands form a gradient that serves as a cue for cells to migrate[5,6]. Thus, studying CCIs elucidates how cells coordinate different functions depending on both the molecules mediating CCC and their spatial context.

CCIs and CCC can be inferred from transcriptomic data[7]. Computational analysis of CCIs usually consists of examining the coexpression of secreted proteins by a sender cell (e.g. ligands) and their cognate surface proteins in a receiver cell (e.g. receptors). To reveal active communication pathways from the coexpression of the corresponding ligand-receptor (LR) pairs, communication scores can be assigned to these interactions based on the RNA expression levels of genes encoding the secreted and receiver proteins[8–12]. RNA-based analyses have informed CCIs and their mediators in small communities of cells, such as embryos[13,14] and tissues[11,15–23]. Moreover, CCI analyses have enabled the study of all cell types in the whole body of a multicellular organism in post-embryonic stages[24].

CCIs allow cells to know their location in their communities, enabling a coordination of functions[4,25]. Molecules mediating these interactions are used by cells to encode and pass this spatial code. Thus, the study of CCIs can help decode spatial organization and function. However, studying CCIs in a spatial context cannot be directly done from conventional single-cell RNA-sequencing technologies (scRNA-seq) since spatial information is lost during tissue dissociation[26]. Nevertheless, previous studies have proved 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[13,26–28]. For example, RNA-Magnet inferred cellular contacts in the bone marrow by considering the coexpression of adhesion molecules present on cell surfaces[28], while ProximID used gene expression coupled with microscopy of cells to construct a spatial map of cell-cell contacts in bone marrow[27]. Thus, as previously done for understanding a few multicellular niches[13,29–31], one can study CCIs in a spatial context by adding appropriate information to the RNA-based CCI analysis. Hence, integrating CCI analyses with spatial information represents a great opportunity to deepen our understanding of intercellular communication in a spatial context. Moreover, this task holds potential to identify the signals that cells use to encode spatial organization.

Caenorhabditis elegans is an excellent model for studying CCIs in a spatial context[32]. Its cellular organization shows complexity comparable to higher-order organisms. It also has well defined tissues, and organs in post-embryonic stages. *C. elegans* can also be easily cultured and all individuals possess the same number and location of cells. Previous studies about its CCIs have focused on subsets of cells or very-early developmental stages. For example, cell contacts were mapped by microscopy during embryogenesis[33,34], which helped identify cell pairs using a specific Notch signaling pathway[34]. Thus, *C. elegans* represents a well controlled experimental system that is sophisticated enough for performing CCI analyses. Importantly, a microscopy-based 3D atlas of cells[35] and a single-cell transcriptome[36] exist for *C. elegans* larvae; a stage at which most of the *C. elegans'* organs have already formed and most cells have migrated to their

final adult position[37]. Therefore, *C. elegans* can readily be exploited to decode the signals that define the spatial organization of cells in a whole living animal. To this end, here we compute CCIs from scRNA-seq data and assess how intercellular distance is associated with the potential of cells to interact. Also, we inspected which signals govern the CCC that occur in different locations and ranges of distance across the *C. elegans* body. For the latter task, a list of ligand-receptor interactions was built, which to date is the most comprehensive one for CCI analyses of *C. elegans*. Our computational framework detects molecules that link CCIs and intercellular distances. Importantly, this approach can be extended to any other organism that has the necessary input data to study the molecular bases underlying intercellular distances and many other traits driving and sustaining cellular organization in multicellular organisms.

Results

Computing cell-cell interactions

Intercellular communication allows cells to coordinate their gene expression[1] and to form spatial patterns of molecule exchange[2]. These events also allow cells to sense their spatial proximity, which is essential for both the formation and the homeostasis of tissues and organs[4]. For example, one mechanism includes sensing the occupancy of receptors by signals from surrounding cells[4,38]; higher occupancy can indicate greater proximity of communicating cells[39]. 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 cell-cell interaction (CCI) score is computed from the expression of ligands and receptors to represent the molecular complementarity of a pair of interacting cells. Specifically, it weighs the number of LR pairs that both cells use to communicate by the aggregate total of complementary ligands and receptors 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.

For computing CCIs in *C. elegans*, besides the gene expression levels of its cells, 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 scattered across literature or contained in protein-protein interaction (PPI) networks that include other categories of proteins. Thus, we generated a list for *C. elegans* that consists of 245 ligand-receptor interactions (Supplementary Table 1), which was built from interactions described in literature and high-confidence published PPIs (see Methods). 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*[36], and ultimately the active LR pairs in all pairs of cells. To determine presence and absence of proteins, we used expression thresholding[9,24], the most common strategy for analyzing CCIs and CCC due to its binary nature and easy interpretation[7], and used the derived ligand and receptor scores as input of our CCI score to represent the overall potential of its cell types to interact.

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). Thus, we used our Bray-Curtis-like score to generate the first predicted network of CCIs in *C. elegans* that measures the complementarity of interacting cells given their active LR pairs (Figure 2a). Although we chose this score for representing a spatial-dependent complementarity of interaction, *cell2cell* is flexible in terms of

the scoring strategies applied to decipher CCIs and CCC, 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).

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

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, we found germline cells to have the lowest CCI potential with other cell types. This is consistent with the physical constraint that germline cells have as they are surrounded by basal membranes that limit their intercellular communication with other cell types[40,41], and may uncouple the coordination of their gene expressions. For example, germline cell fate into mitosis or meiosis is almost exclusive on CCIs with distal tip cells, especially through mediators of Notch signaling such as *glp-1*[42]. In contrast, neurons have the largest potential for interactions with other cell types, suggesting that these cell types use a higher fraction of all possible communication pathways. This occurs especially in interactions between neurons and muscle cells (Figure 2a), which is consistent with the high molecule interchange that occurs at the neuromuscular junctions[43], suggesting that our method is exposing complementary signals actually transmitted between cells.

Similarity between pairs of interacting cells was also analyzed given the LR pairs they use. By using UMAP[44,45] to visualize the similarity they 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[46], which suggests that this phenomenon may be conserved across multicellular organisms.

Key signaling pathways link distance between cells with their potential of interaction

Given that our CCI score is undirected, it can also be compared with spatial properties such as the distance between cells, which represent a unique state that does not change with the order that cells in a pair are considered. Under the hypothesis that larger distances should decrease the potential of cells to interact, we expected our CCI scores to be negatively correlated with the Euclidean distances between cells. Thus, we used the distances between cells, calculated by taking the Euclidean distance between cells from a 3D atlas of *C. elegans* (Supplementary Figure 1a), as our reference data to assess our methodology and assumptions in calculating our CCI score (see ***Computing cell-cell interactions***)

We annotated each cell in the 3D atlas with a corresponding cell type in the scRNA-seq dataset (Supplementary Table 2), and computed the minimal Euclidean distances between each pair of cell types (Supplementary Figure 1b). In this case, the minimal distance was considered because this case would represent the maximal potential that cells have to interact. Thus, we next calculated the Spearman correlation between the CCI score matrix and the Euclidean distance matrix. This originally resulted in a correlation coefficient of -0.21 (P-value = 0.0016). Although the correlation was negative as expected, it was a low value. This may be due to noise introduced by comprehensively incorporating all LR pairs into the computation, which may include several LR pairs not necessarily encoding spatial information.

Given that the CCI scores computed by using all LR pairs in our list led to a low correlation with the distances between cells, we hypothesized that there is a subset of key LR pairs that follow a spatial pattern of co-expression and would allow cells to specifically sense their spatial relationship with other cells or drive the spacing. In this scenario, a CCI score that is a function of only these LR pairs would better represent the potential of two cells to functionally interact and it is expected to better correlate with cell proximity. Under this hypothesis, we looked for key LR pairs of *C. elegans* that better capture the potential of cells to interact given their physical locations. To do so, we ran a genetic algorithm (GA) to maximize the correlation between the CCI score matrix and the Euclidean distance matrix by randomly generating different size subsets of the LR pairs in our complete list (Supplementary Figure 2a-b). This algorithm was run 100 times, obtaining in each case a different optimal list of LR pairs due to the stochastic nature of this algorithm (Supplementary Figure 2c). Nevertheless, across all solutions, an average Spearman coefficient of -0.67 ± 0.01 was obtained (shown as an absolute value in Supplementary Figure 2b) and the maximal correlation resulted in a value of -0.70 (P-value = 1.435×10^{-35}). Thus, the resulting optimized subsets of LR pairs (hereinafter referred to as initial GA-LR pairs) may constitute good predictors of biological functions driving or sustaining intercellular proximity, and they support the hypothesis that a subset of the LR pairs would drive the distance-dependent potential of interaction between two cells.

To identify the specific biological roles that the initial GA-LR pairs may have, we used our functional annotations about the signaling processes they are involved in (see column "LR Function" in Supplementary Table 1). Specifically, for each initial GA-LR pair list, we computed the relative abundance of each signaling pathway (i.e. the number of LR pairs involved in a given pathway with respect to the total number of LR pairs in the list) (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 Canonical RTK-Ras-ERK signaling, cell migration, Hedgehog signaling and mechanosensory mechanisms increased their relative abundance in the resulting subsets from the GA runs. Thus, the GA seems to be prioritizing LR pairs associated with processes such as cell patterning, morphogenesis and tissue maintenance[47].

The established or predicted roles of the enriched initial GA-LR pairs are congruent with a role in establishing and/or sustaining the proximity of cells. This notion is further supported by the congruence between our predictions and targeted studies demonstrating the essential role in spatial organization of the initial GA-LR pairs detected in most of the GA runs. For instance: 1) The LR pair composed of *smp-2*/*plx-1* mediates epidermal morphogenesis, as demonstrated by the defects in epidermal functions exhibited by *C. elegans* lacking *smp-2*[48]; 2) *cwn-1*/*mig-1* mediate cell positioning, as demonstrated by the abnormal migration of hermaphrodite specific motor neurons in the mutants[49,50]; 3) *K05F1.5*/*dma-1*, as *K05F1.5* was described as a novel gene (named *lect-2*) that, similar to *mnr-1* and *dma-1*, is key for dendrite guidance of sensory neurons innervating the muscle-skin interface[51,52]; and 4) the *let-756*/*ver-1* interaction, an homolog of the mammalian FGF-VEGF pathway, that is essential for *C. elegans* development, especially in the L2 stage where *let-756* has its peak of expression[53], and for the positioning of ray 1 in the tail[54].

Consensus GA-selected LR pairs contribute to the formation of spatial patterns of communication along the *C. elegans* body

Next, we looked for a core set of LR pairs that were representative of the optimal solutions generated by the GA. We first clustered LR pairs by their co-occurrence across the GA runs and

then selected the cluster with members that were simultaneously present in a high fraction of the optimal subsets (Supplementary Figure 2c-d). This resulted in a consensus list of 37 LR pairs (Supplementary Table 3), hereinafter referred to as GA-LR pairs, whose combined appearance seemed to encode proximity across cell-cell interactions.

This consensus list yields a Spearman coefficient of -0.63 (P-value = 2.629×10^{-27}) between the CCI score matrix and the Euclidean distance matrix. To test whether the correlation stems from the specific LR pairs in the consensus list, we did a series of permutation analyses (Supplementary Figure 3). We evaluated if the correlation computed with the consensus list was greater than the value stemming from the null distribution generated with random interactions generated from the ligands and receptors in the GA-LR interactions, either by randomly permuting the ligands and the receptors (Supplementary Figure 3a) or by shuffling their labels to keep the topology of the interactions (Supplementary Figure 3b). We also subsampled the complete list of LR pairs (Supplementary Table 1) to obtain random subsets of similar size to the list of GA-LR pairs (Supplementary Figure 3c). In each scenario, the randomized lists yielded a smaller negative Spearman correlation than the consensus list (P-value < 0.0001, see Supplementary Figure 3).

Interestingly, when compared to the complete list of LR pairs, the GA-LR pairs led to more heterogeneity in the cell-cell interaction potential (Figures 2a and 4a). The heterogeneity seems to stem from the proximity that cells have since cells are grouped by functional interactions within their tissues (Figure 4a). For example, the cells composing the pharynx (pharyngeal gland, epithelia, muscle and neurons) group together, which may reflect a pharynx-specific pattern of interactions between these cell types that may not translate to the cell types composing other organs or locating in other regions of the worm. Similarly, we observed seemingly functional associations between neurons and muscles. In particular, the interactions between muscle cells and GABAergic and cholinergic neurons presented a high CCI score, which may represent a protein-based priming for the known exchange of GABA and acetylcholine in neuromuscular junctions. We also found that the most complementary interaction of amphid/phasmid sheath cells was with seam cells. This is consistent with the role of seam cells during larval development, where they form adherens junctions with sheath cells and hypodermal cells and function as sockets for the phasmid sensilla[55]. Another interesting observation was the high cell-cell interaction score between oxygen sensing neurons and intestinal cells, which is consistent with the extensive communication between these cells to link oxygen availability with nutrient status[56–58]. Thus, the protein-protein interactions prioritized by the GA seem to capture cellular properties that define physical proximity, especially defining functional roles of tissues and organs.

To explore the spatial distribution of consensus GA-LR pairs along the body of *C. elegans* (Figure 5), we performed an enrichment analysis of CCC 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[59], while LIN-44 is secreted by hypodermal cells exclusively in the tail during larval development[60,61], both cases coinciding with the results in Table 1. By contrast, *daf-7* is known to be expressed only by sensory neurons in the head[62]; however, our results suggest an enrichment in the tail (Table 1). This is likely due to the

transcriptome including two types of sensory neurons that indeed express *daf-7* (Figure 4b) suggesting that clustering of single cells cannot distinguish all subtypes of sensory neurons. Thus, mapping cell types in the 3D atlas resulted in all sensory neurons having a similar transcriptome across the body (Figure 5a), explaining this spurious result. Nevertheless, although the *daf-7* example points to limitations of the current scRNAseq methods and their analysis tools, the *col-99* and *lin-44* examples demonstrate that when cellular identification is sufficiently detailed, our strategy captures true biological spatial behaviors of gene expression and therefore of CCC.

To gain more understanding on the importance that the GA-LR pairs may have 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 6). 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*). As previously reported, 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[63]; while their mutants show cell positioning defects, especially along the anterior/posterior axis of *C. elegans*[48,64]. They are also key for axon guidance and cell migration[65] and necessary for epidermal[66] and vulval morphogenesis[67]. Members of the Wnt signaling are biomolecules known to act as a source of positional information for cells[4]. In *C. elegans*, *cwn-1* and *lin-44*, for example, follow a gradient along its body, enabling cell migration[49,68–70]. Thus, the GA-LR pairs may influence local or longer-range interactions and help encode intercellular proximity.

GA-selected mediators are enriched in morphology and cell migration phenotypes

As many LR interactions have not been spatially or functionally characterized in *C. elegans*, we anticipate that several pairs in our GA-selected list may either be “yet to be discovered functional interactions” or false positives. To minimize false positives, we incorporated phenotypic data. The rationale here being that phenotypic associations between the genes likely indicate coordinated function, which in this case may imply CCIs. Thus, we next examined the phenotypic associations between genes composing the GA-LR pair list. Given that, as defined before, these LR interactions are important for cell patterning, morphogenesis and tissue maintenance, we focused on phenotypes with annotations such as “morphology phenotype” and “cell migration”. We tested for enrichment of GA selected genes with a Fisher’s exact test, using the lists of genes whose mutants have been demonstrated to affect that phenotype. Using our complete list of LR pairs to define the background list of genes and those in the consensus list as the sample list, we observed that morphology and cell migration phenotypes have odds ratios of 4.83 (P-value = 0.027) and 3.07 (P-value = 0.0019), respectively, indicating an enrichment and therefore, supporting their role as drivers of spatial organization.

Among the genes associated with either morphology or cell migration phenotypes in our complete list of LR pairs, 51% of them were selected by our genetic algorithm (Figure 7). Remarkably, many of the GA-selected genes associated with morphology phenotypes are involved in the Wnt signaling pathway (*lin-17*, *lin-18*, *lin-44* and *mom-2*) and the rest with cell adhesion (*epi-1*) and insulin signaling (*daf-2*). On the other hand, GA-selected genes associated with cell migration include Wnt signaling (*cam-1*, *cfz-2*, *cwn-1*, *lin-17*, *lin-44*, *mig-1* and *mom-2*), cell migration pathways (*mab-20*, *unc-5*, *unc-6*, *unc-129* and *rig-6*), cell adhesion (*ddr-1*, *epi-1*, *ina-1*, *let-2*, *nid-1* and *pat-3*), Notch signaling (*lag-2* and *lin-12*) and TGF-β signaling (*dbl-1*). Noteworthily, previous studies have shown that these genes and their interactions are key in spatial allocation of cells. Examples include: 1) *pat-3*, which is involved in post-embryonic

organogenesis and tissue function[71]; 2) the interaction between the discoidin domain receptor ddr-1 and the collagen col-99, which plays a role in axonal guidance and asymmetry establishment of the ventral nerve cord[72]; 3) *lin-12*, *mab-20* and *unc-6* and their respective receptors, which are involved in intestine morphogenesis[73]; and 5) the interaction between *nid-1* and *ptp-3*, which participates in cell migration and axon guidance[74]. Cell adhesion molecules such as collagen and proteins from the immunoglobulin superfamily also play a role in intercellular contact and communication, encompassing genes such as *let-2*, *cle-1*, *gpn-1*, *rig-6*, *wrk-1*, *unc-5* and *ina-1*[75–79]. Interestingly, when considering the expression of genes associated with the phenotypes reported here, cell types seem to be clustered by spatial proximity of their lineage groups (Figure 7), suggesting that these genes may be markers of spatial properties. Thus, the congruence between spatial proximity and biological function strongly supports the notion that the strategy presented here provides insights into the spatial code behind CCIs and CCC.

Discussion

Here we developed a computational strategy for inferring complementarity between cells given their ligand and receptor expression in scRNA-seq data. With this approach, we identified spatial properties in *C. elegans* associated with the potential of cells to interact and communicate. Particularly, we found a negative Spearman coefficient between intercellular distance and CCIs computed with our Bray-Curtis-like score; a correlation that was stronger when inferring CCIs from ligand-receptor pairs selected with a genetic algorithm. Thus, these LR pairs resulted to be informative of spatial properties and may direct how cells transmit this kind of information to other cells.

In this study, we also collected ligand-receptor interactions of *C. elegans* that were available in literature and PPI databases, building an essential resource for CCI analyses of *C. elegans* (Supplementary Table 1). Using this list, our CCI analysis led to results consistent with previous findings. For example, we found that interacting cells were grouped given cell type-specific production of ligands (Figure 2b), which was previously shown in a work that analyzed a communication network of human haematopoietic cells[46]. Our results are also consistent with experimental studies of *C. elegans*. For instance, the GA-driven selection of LR pairs significantly prioritized interactions participating in cell migration, Hedgehog signaling, mechanosensory mechanisms and canonical RTK-Ras-ERK signaling. Remarkably, these pathways and few other mediators, also involved in cell migration (e.g. members of Notch and TGF-β signaling), are crucial for the larval development of *C. elegans*[80–83], coinciding with the cognate stages of the datasets we used.

Our analysis successfully recapitulated known biology regarding spatial organization of the cells in *C. elegans* and associated ligand-receptor interactions. Many of the ligands and receptors included in the GA-LR pairs (Supplementary Table 3 and Figure 4b) have been reported to contribute to cellular positioning, migration and/or organ morphogenesis, which may explain why the genetic algorithm selected them as encoders of spatial information. Netrin UNC-6 is an example ligand that serves as a cue for enabling DTC migration given its spatial pattern of expression[84]. Moreover, TGFβ-related signaling pathways regulate body size and therefore location of cells, sometimes involving proteins selected by the GA, such as DAF-7, DBL-1, SMA-6, SMA-10, LON-2, SRP-7, F14B4.1 and UNC-129[85–87]. Similarly, mediators associated with Wnt signaling are required for axon guidance[88] and mutants of some of their encoding genes, such as *cam-1* (ROR homolog that sequesters Wnts), *cwn-1*, *lin-44*, *cfz-2*, *lin-17*, *lin-18* and *mig-1*, have previously been linked to defects of cell migration, positioning and patterning[50,89–91]. Thus, the spatial distributions and functions determined in previous studies

provide support to our predictions, which suggests that analyzing CCIs and CCC through RNA-based approaches can decipher important spatial and functional properties.

Considering the inputs of our analyses, false positives may arise from the preprocessing of datasets. For instance, selecting a threshold value to consider ligands and receptors as expressed can affect the number of false positives and negatives[9]. In this regard, different values could be explored to infer the presence of biologically active protein, as previously addressed[92,93]. Moreover, other approaches such as using expression products to compute the usage of a ligand-receptor pair may also help[10], but further adaptations should be done to use it with our Bray-Curtis like score. Our CCI score is also dependent on the input list of LR interactions, so including other LR pairs that were not considered in our complete list (Supplementary Table 1) could improve the predictions. More comprehensive lists might generate a better correlation or contain more important LR pairs for the GA selection than the ones selected here (Supplementary Table 3). Furthermore, lists of ligand-receptor pairs considering the formation of multimeric complexes can improve the reliability of the results[11,94], so considering structural information of proteins may also improve the predictions. However, in contrast to mammals such as mice and humans, *C. elegans* has considerably less information for building comprehensive lists of ligand-receptor interactions containing multimeric complexes.

Although our strategy captured underlying mechanisms that are consistent with experimental evidence in literature, our approach has limitations that can be related, for instance, to the nature of the dataset. Conventional scRNA-seq technologies do not preserve spatial information, so labelling cells in a 3D atlas by using the cell types in the transcriptome might be a confounder. For example, *C. elegans* possesses sub-types of non-seam hypodermal cells, and their gene expressions vary depending on the antero/posterior location of them. However, in the scRNA-seq data set employed here there was only one type of non-seam hypodermal cells to represent all subtypes, so they virtually share the same gene expression. An illustrative case where this represents an issue is the expression of *lin-44*, which is expressed by hypodermis cells located in the tail[68,70], but not in other sections as our results showed for *lin-44*/*lin-17* (Figure 5b). Besides that, physical constraints, such as physical barriers between cells, are not considered in our analyses, which could lead to false positives. Examples of this class of false positives are 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 types of cells[41]. Therefore, relying only on the gene expression 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, emerging spatial transcriptomics methods are expected to be an important advance in the study of CCIs[95], since they can distinguish specific cell subtypes given their locations and physical constraints.

The strategy presented in this work provides a framework to associate CCIs with phenotypes and detect ligand-receptor interactions that are crucial for those phenotypes. In contrast to previously proposed overall CCI scores[7], ours is undirected. A benefit of it is that it can be successfully integrated with any information also representing an undirected state between two cells, as it is the intercellular distance. When considering, for example, spatial information, this holds the potential of recovering spatial properties that are lost in the traditional transcriptomics methods, either bulk or single cell. Thus, it also holds the potential to be used for other correlative or predictive purposes about important LR pairs associated with a phenotype of interest. Importantly, substantial support to our predictions is observed in the literature, which suggests that our strategy captures underlying mechanisms and functions of mediators associated with the spatial allocation of cells. Thus, although future experiments evaluating

spatial allocation of signals will help refine our predictions, the strategy presented here lays the foundation to unveil the code of cell-cell interactions and communication that defines the spatial distribution of cells across a whole animal body.

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 transcriptome[36]. 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[36], where 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[35]. 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 (Supplementary Figure 1a). 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[9] was used as reference for looking for respective orthologs in *C. elegans*. The search for orthologs was done using OrthoDB[96], OrthoList[97] and gProfiler[98]. Then, a network of protein-protein interactions for *C. elegans* was obtained from RSPGM[99] and high-confidence interactions in STRING-db (confidence score > 700 and supported at least by one experimental evidence)[100]. 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 (Supplementary Table 4). 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 (Supplementary Table 1).

Communication and CCI scores

To detect active communication pathways and to compute CCI scores between cell pairs, first it was necessary to detect the presence or absence of each ligand and receptor. To do so, we used a threshold over 10 TPM as previously described[9]. Thus, those ligands and receptors that passed this filter were considered as expressed (a binary value of 1 was assigned). Then, a communication score of 1 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 as indicated in Figure 1. Using the respective vectors for both interacting cells, a Bray-Curtis-like score was calculated to represent the potential of interaction. This potential aims to measure how complementary are the signals that interacting cells produce. To do so, our Bray-Curtis-like score considers the number of

active LR pairs that a pair of cells has while also incorporating the potential that each cell has to communicate independently (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 Supplementary Figure 2a). The maximization process was run 100 times, generating 100 different lists that resulted in an optimal correlation. As shown in Supplementary Figures 2c-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 different ranges of distance used for CCC were defined by using a Gaussian mixed model on the distributions of distance between all pairs of cell types. This model was implemented using the scikit-learn library for Python[101] and a number of components equal to 3.

Statistical analyses

For each function annotated in the list of ligand-receptor pairs (Supplementary 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.

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 evaluated the number of those interactions that the corresponding ligand-receptor pair was used. 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 from the phenotype ontology association available in WormBase[102]), 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).

Data availability

- Single-cell RNA-seq dataset**:** Supplementary Table S4 in https://doi.org/10.1126/science.aam8940 and GitHub repository for the analyses of this work.
- 3D digital atlas of *C. elegans*: Supplementary Data 1 in https://doi.org/10.1038/nmeth.1366 and GitHub repository for the analyses of this work.
- Lists of ligand-receptor pairs of *C. elegans*: The manual curated list containing 245 interactions is present in the Supplementary Table 1, while the consensus list from the GA-selection, which contains 37 interactions, is available in the Supplementary Table 3. Both lists are also available in the GitHub repository for the analyses of this work.

Code availability

All analyses performed in this work, their respective codes (implemented in Python and Jupyter Notebooks) and instruction to use them are available in a public repository (https://github.com/LewisLabUCSD/Celegans-cell2cell). Similarly, *cell2cell* is available as an open source tool in a GitHub repository (https://github.com/earmingol/cell2cell).

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 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 score (undirected score).

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 according to REF[36]. **b**, UMAP visualization of CCC. 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 Supplementary Table 5, where names of LR pairs and their communication scores are specified for each cell pair.

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). 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 respective 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 test.

Figure 4. CCI and CCC analyses based on LR pairs driving high correlation 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), disregarding 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**).

Figure 5. Anteroposterior use of communication pathways.

a, Division of the body of *C. elegans* into three main sections along the anteroposterior axis (top) and cell-type composition of each section (bottom) given a previously published 3D atlas. In this case, 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.

Figure 6. Importance of ligand-receptor interactions for communication at different distances.

Circos plots for representing the Fisher exact test on cell-cell communication occurring at different ranges of distance. The ranges of distances were defined as explained in Supplementary Figure 1c. Nodes represent ligands or receptors and edges connect those ligands and receptors that interact in the GA-LR pairs (Supplementary Table 3). The color of the nodes represent whether they are ligands or receptors and the color of the edges indicate the negative value of the logarithmic transformation on the adjusted P-values (Benjamini-Hochberg's method), 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.

Figure 7. Expression of cell migration- and morphogenesis-associated genes in the complete list of LR and GA-LR pairs.

The presence or absence of proteins encoded by genes associated with cell migration and/or morphology phenotypes (y-axis) is indicated for each cell type (x-axis) according to *C. elegans* phenotype ontology. The threshold for presence is a gene expression value greater than 10 TPM; otherwise is labeled as absence. Only genes that are present in our complete list of LR pairs are shown, and members also in the GA-LR list are denoted with ochre cells (y-axis). Color key for groups of cell types and phenotype annotations are depicted to the right. Agglomerative hierarchical clustering was performed using a Jaccard similarity for both genes and cell types, independently.

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 main text for discussion on this prediction.

Supplementary Figure 1. Euclidean distances between cells.

a, Schematic representation of computing the minimal Euclidean distance for a pair of cell types in *C. elegans*. The distance for all pairs of cells between those belonging to cell type A and those belonging to cell type B is computed; then the minimal one is selected. **b**, Heatmap of resulting Euclidean distances among all pairs of cells. Depicted color key defines the eight major cell groups that were previously defined by Cao et al.[36]. **c**, Distribution of intercellular distances given the diagonal matrix in (**b**). Short-, mid- and long-range distances were defined using a Gaussian mixed model built with three components.

Supplementary Figure 2. Genetic algorithm-based selection of ligand-receptor pairs leading to high CCI score-distance correlation.

a, Workflow for running the genetic algorithm (GA) and selecting a subset of ligand-receptor pairs leading to an optimal correlation between CCI scores and intercellular distances. Each run consists of ten steps described in the schematic representation. This framework was run 100 times independently, leading to 100 different subsets. **b**, Absolute value of the correlation score obtained at each iteration of the GA. 100 independent runs are plotted, showing the mean and standard deviation across them. **c**, Heatmap indicating when each of the 245 initial LR pairs was selected by one of the 100 GA runs. **d**, Heatmap of the co-occurrence of two LR pairs. Only LR pairs that were selected in at least one GA run are shown here. The co-occurrence is defined as the number of runs that two pairs were selected together with respect to the total number when at least one pair was selected. The dashed red square represents the cluster that was finally selected as the consensus subset of LR pairs because of the high values of co-occurrence and high fraction of presence that its interactions had across the 100 GA runs.

Supplementary Figure 3. Permutation-based null distributions of the correlation obtained using the GA-LR pairs.

a, Null distributions resulting from column-wise random permutation of the interactors, either ligands or receptors, separately. **b**, Null distributions resulting from random permutation of interactor labels, either of ligands or receptors, separately. **c**, Null distribution resulting from random subsampling of the initial LR pairs (245 interactions) to generate subsets equivalent to the GA-LR pairs (37 interactions). Each analysis was run independently 10,000 times and for each run a correlation coefficient was computed, generating a null distribution in each case. In (**a**-**c**) the dashed lines represent the correlation score obtained from the consensus GA-LR pairs.

References

1. Sgro, A. E. *et al.* From intracellular signaling to population oscillations: bridging size- and time-scales in collective behavior. *Mol. Syst. Biol.* **11**, 779 (2015).
2. Dang, Y., Grundel, D. A. J. & Youk, H. Cellular Dialogues: Cell-Cell Communication through Diffusible Molecules Yields Dynamic Spatial Patterns. *Cell Syst* **10**, 82–98.e7 (2020).
3. Sun, B., Lembong, J., Normand, V., Rogers, M. & Stone, H. A. Spatial-temporal dynamics of collective chemosensing. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 7753–7758 (2012).
4. Lander, A. D. How cells know where they are. *Science* **339**, 923–927 (2013).
5. Wu, D. & Lin, F. Modeling cell gradient sensing and migration in competing chemoattractant fields. *PLoS One* **6**, e18805 (2011).
6. Pani, A. M. & Goldstein, B. Direct visualization of a native Wnt in vivo reveals that a long-range Wnt gradient forms by extracellular dispersal. *Elife* **7**, (2018).
7. Armingol, E., Officer, A., Harismendy, O. & Lewis, N. E. Deciphering cell–cell interactions and communication from gene expression. *Nat. Rev. Genet.* 1–18 (2020).
8. Graeber, T. G. & Eisenberg, D. Bioinformatic identification of potential autocrine signaling loops in cancers from gene expression profiles. *Nat. Genet.* **29**, 295–300 (2001).
9. Ramilowski, J. A. *et al.* A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat. Commun.* **6**, 7866 (2015).
10. Kumar, M. P. *et al.* Analysis of Single-Cell RNA-Seq Identifies Cell-Cell Communication Associated with Tumor Characteristics. *Cell Rep.* **25**, 1458–1468.e4 (2018).
11. Vento-Tormo, R. *et al.* Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* **563**, 347–353 (2018).
12. Choi, H. *et al.* Transcriptome analysis of individual stromal cell populations identifies stroma-tumor crosstalk in mouse lung cancer model. *Cell Rep.* **10**, 1187–1201 (2015).
13. Cang, Z. & Nie, Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat. Commun.* **11**, 2084 (2020).
14. Xue, Y. *et al.* A 3D Atlas of Hematopoietic Stem and Progenitor Cell Expansion by Multi-dimensional RNA-Seq Analysis. *Cell Rep.* **27**, 1567–1578.e5 (2019).
15. Yuzwa, S. A. *et al.* Proneurogenic Ligands Defined by Modeling Developing Cortex Growth Factor Communication Networks. *Neuron* **91**, 988–1004 (2016).
16. Sheikh, B. N. *et al.* Systematic Identification of Cell-Cell Communication Networks in the Developing Brain. *iScience* **21**, 273–287 (2019).
17. Popescu, D.-M. *et al.* Decoding human fetal liver haematopoiesis. *Nature* **574**, 365–371 (2019).
18. Wang, S., Karikomi, M., MacLean, A. L. & Nie, Q. Cell lineage and communication network inference via optimization for single-cell transcriptomics. *Nucleic Acids Res.* **47**, e66 (2019).
19. Joost, S. *et al.* Single-Cell Transcriptomics Reveals that Differentiation and Spatial Signatures Shape Epidermal and Hair Follicle Heterogeneity. *Cell Syst* **3**, 221–237.e9 (2016).
20. Wang, S. *et al.* Single cell transcriptomics of human epidermis reveals basal stem cell transition states. *bioRxiv* 784579 (2019) doi:10.1101/784579.
21. Joost, S. *et al.* Single-Cell Transcriptomics of Traced Epidermal and Hair Follicle Stem Cells Reveals Rapid Adaptations during Wound Healing. *Cell Rep.* **25**, 585–597.e7 (2018).

22. Camp, J. G. *et al.* Multilineage communication regulates human liver bud development from pluripotency. *Nature* **546**, 533–538 (2017).
23. Pavličev, M. *et al.* Single-cell transcriptomics of the human placenta: inferring the cell communication network of the maternal-fetal interface. *Genome Res.* **27**, 349–361 (2017).
24. Hou, R., Denisenko, E., Ong, H. T., Ramilowski, J. A. & Forrest, A. R. R. Predicting cell-to-cell communication networks using NATMI. *Nat. Commun.* **11**, 1–11 (2020).
25. Purvis, J. E. & Lahav, G. Encoding and decoding cellular information through signaling dynamics. *Cell* **152**, 945–956 (2013).
26. Ren, X. *et al.* Reconstruction of cell spatial organization from single-cell RNA sequencing data based on ligand-receptor mediated self-assembly. *Cell Res.* (2020) doi:10.1038/s41422-020-0353-2.
27. Boisset, J.-C. *et al.* Mapping the physical network of cellular interactions. *Nat. Methods* **15**, 547–553 (2018).
28. Baccin, C. *et al.* Combined single-cell and spatial transcriptomics reveal the molecular, cellular and spatial bone marrow niche organization. *Nat. Cell Biol.* **22**, 38–48 (2020).
29. Cohen, M. *et al.* Lung Single-Cell Signaling Interaction Map Reveals Basophil Role in Macrophage Imprinting. *Cell* **175**, 1031–1044.e18 (2018).
30. Li, L. *et al.* Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and Gonadal Niche Interactions. *Cell Stem Cell* **20**, 891–892 (2017).
31. Solé-Boldo, L. *et al.* Single-cell transcriptomes of the human skin reveal age-related loss of fibroblast priming. *Commun Biol* **3**, 188 (2020).
32. Kaletta, T. & Hengartner, M. O. Finding function in novel targets: C. elegans as a model organism. *Nat. Rev. Drug Discov.* **5**, 387–398 (2006).
33. Hench, J., Henriksson, J., Lüppert, M. & Bürglin, T. R. Spatio-temporal reference model of Caenorhabditis elegans embryogenesis with cell contact maps. *Dev. Biol.* **333**, 1–13 (2009).
34. Chen, L. *et al.* Establishment of Signaling Interactions with Cellular Resolution for Every Cell Cycle of Embryogenesis. *Genetics* **209**, 37–49 (2018).
35. Long, F., Peng, H., Liu, X., Kim, S. K. & Myers, E. A 3D digital atlas of C. elegans and its application to single-cell analyses. *Nat. Methods* **6**, 667–672 (2009).
36. Cao, J. *et al.* Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).
37. Girard, L. R. *et al.* WormBook: the online review of Caenorhabditis elegans biology. *Nucleic Acids Res.* **35**, D472–5 (2007).
38. Nickerson, M. Receptor occupancy and tissue response. *Nature* **178**, 697–698 (1956).
39. Zhong, P., Cara, J. F. & Tager, H. S. Importance of receptor occupancy, concentration differences, and ligand exchange in the insulin-like growth factor I receptor system. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 11451–11455 (1993).
40. Hubbard, E. J. A. Caenorhabditis elegans germ line: a model for stem cell biology. *Dev. Dyn.* **236**, 3343–3357 (2007).
41. Pazdernik, N. & Schedl, T. Introduction to germ cell development in Caenorhabditis elegans. *Adv. Exp. Med. Biol.* **757**, 1–16 (2013).
42. Kimble, J. & Crittenden, S. L. Germline proliferation and its control. in *WormBook* (ed. The C. elegans Research Community) (2005).

43. Daniels, M. P. Intercellular communication that mediates formation of the neuromuscular junction. *Mol. Neurobiol.* **14**, 143–170 (1997).
44. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [stat.ML]* (2018).
45. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* (2018) doi:10.1038/nbt.4314.
46. Qiao, W. *et al.* Intercellular network structure and regulatory motifs in the human hematopoietic system. *Mol. Syst. Biol.* **10**, 741 (2014).
47. Ingham, P. W., Nakano, Y. & Seger, C. Mechanisms and functions of Hedgehog signalling across the metazoa. *Nat. Rev. Genet.* **12**, 393–406 (2011).
48. Ginzburg, V. E., Roy, P. J. & Culotti, J. G. Semaphorin 1a and semaphorin 1b are required for correct epidermal cell positioning and adhesion during morphogenesis in C. elegans. *Development* **129**, 2065–2078 (2002).
49. Pan, C.-L. *et al.* Multiple Wnts and frizzled receptors regulate anteriorly directed cell and growth cone migrations in Caenorhabditis elegans. *Dev. Cell* **10**, 367–377 (2006).
50. Zinovyeva, A. Y., Yamamoto, Y., Sawa, H. & Forrester, W. C. Complex network of Wnt signaling regulates neuronal migrations during Caenorhabditis elegans development. *Genetics* **179**, 1357–1371 (2008).
51. Zou, W. *et al.* A multi-protein receptor-ligand complex underlies combinatorial dendrite guidance choices in C. elegans. *Elife* **5**, (2016).
52. Liu, O. W. & Shen, K. The transmembrane LRR protein DMA-1 promotes dendrite branching and growth in C. elegans. *Nat. Neurosci.* **15**, 57–63 (2011).
53. Roubin, R. *et al.* let-756, a C. elegans fgf essential for worm development. *Oncogene* **18**, 6741–6747 (1999).
54. Dalpe, G., Tarsitano, M., Persico, M. G., Zheng, H. & Culotti, J. C. elegans PVF-1 inhibits permissive UNC-40 signalling through CED-10 GTPase to position the male ray 1 sensillum. *Development* **140**, 4020–4030 (2013).
55. White, J. The Anatomy. Chapter 4. in *The nematode C. elegans* (ed. Wood, W. B.) 81–122 (Cold Spring Harbor Laboratory Press, 1988).
56. Noble, T., Stieglitz, J. & Srinivasan, S. An integrated serotonin and octopamine neuronal circuit directs the release of an endocrine signal to control C. elegans body fat. *Cell Metab.* **18**, 672–684 (2013).
57. Witham, E. *et al.* C. elegans Body Cavity Neurons Are Homeostatic Sensors that Integrate Fluctuations in Oxygen Availability and Internal Nutrient Reserves. *Cell Rep.* **14**, 1641–1654 (2016).
58. Hussey, R. *et al.* Oxygen-sensing neurons reciprocally regulate peripheral lipid metabolism via neuropeptide signaling in Caenorhabditis elegans. *PLoS Genet.* **14**, e1007305 (2018).
59. Tu, H., Huhtala, P., Lee, H.-M., Adams, J. C. & Pihlajaniemi, T. Membrane-associated collagens with interrupted triple-helices (MACITs): evolution from a bilaterian common ancestor and functional conservation in C. elegans. *BMC Evolutionary Biology* **15**, (2015).
60. Gleason, J. E., Szyleyko, E. A. & Eisenmann, D. M. Multiple redundant Wnt signaling components function in two processes during C. elegans vulval development. *Dev. Biol.* **298**, 442–457 (2006).
61. Klassen, M. P. & Shen, K. Wnt signaling positions neuromuscular connectivity by inhibiting

synapse formation in C. elegans. *Cell* **130**, 704–716 (2007).
62. Gumienny, T. L. & Savage-Dunn, C. TGF-β signaling in C. elegans. *WormBook* 1–34 (2013).
63. Miller, M. A. & Chin-Sang, I. D. Eph receptor signaling in C. elegans. *WormBook* 1–17 (2012).
64. Nakao, F. *et al.* The PLEXIN PLX-2 and the ephrin EFN-4 have distinct roles in MAB-20/Semaphorin 2A signaling in Caenorhabditis elegans morphogenesis. *Genetics* **176**, 1591–1607 (2007).
65. Ikegami, R., Zheng, H., Ong, S.-H. & Culotti, J. Integration of semaphorin-2A/MAB-20, ephrin-4, and UNC-129 TGF-beta signaling pathways regulates sorting of distinct sensory rays in C. elegans. *Dev. Cell* **6**, 383–395 (2004).
66. Wang, X. *et al.* The C. elegans L1CAM homologue LAD-2 functions as a coreceptor in MAB-20/Sema2–mediated axon guidance. *J. Cell Biol.* **180**, 233–246 (2008).
67. Liu, Z. *et al.* C. elegans PlexinA PLX-1 mediates a cell contact-dependent stop signal in vulval precursor cells. *Dev. Biol.* **282**, 138–151 (2005).
68. Herman, M. A., Vassilieva, L. L., Horvitz, H. R., Shaw, J. E. & Herman, R. K. The C. elegans gene lin-44, which controls the polarity of certain asymmetric cell divisions, encodes a Wnt protein and acts cell nonautonomously. *Cell* **83**, 101–110 (1995).
69. Whangbo, J. & Kenyon, C. A Wnt signaling system that specifies two patterns of cell migration in C. elegans. *Mol. Cell* **4**, 851–858 (1999).
70. Harterink, M. *et al.* Neuroblast migration along the anteroposterior axis of C. elegans is controlled by opposing gradients of Wnts and a secreted Frizzled-related protein. *Development* **138**, 2915–2924 (2011).
71. Lee, M., Cram, E. J., Shen, B. & Schwarzbauer, J. E. Roles for beta(pat-3) integrins in development and function of Caenorhabditis elegans muscles and gonads. *J. Biol. Chem.* **276**, 36404–36410 (2001).
72. Taylor, J., Unsoeld, T. & Hutter, H. The transmembrane collagen COL-99 guides longitudinally extending axons in C. elegans. *Mol. Cell. Neurosci.* **89**, 9–19 (2018).
73. Asan, A., Raiders, S. A. & Priess, J. R. Correction: Morphogenesis of the C. elegans Intestine Involves Axon Guidance Genes. *PLoS Genet.* **12**, e1006077 (2016).
74. Ackley, B. D. *et al.* The two isoforms of the Caenorhabditis elegans leukocyte-common antigen related receptor tyrosine phosphatase PTP-3 function independently in axon guidance and synapse formation. *Journal of Neuroscience* **25**, 7517–7528 (2005).
75. Ackley, B. D. *et al.* The NC1/endostatin domain of Caenorhabditis elegans type XVIII collagen affects cell migration and axon guidance. *J. Cell Biol.* **152**, 1219–1232 (2001).
76. Schwarz, V., Pan, J., Voltmer-Irsch, S. & Hutter, H. IgCAMs redundantly control axon navigation in Caenorhabditis elegans. *Neural Development* **4**, 13 (2009).
77. Leung-Hagesteijn, C. *et al.* UNC-5, a transmembrane protein with immunoglobulin and thrombospondin type 1 domains, guides cell and pioneer axon migrations in C. elegans. *Cell* **71**, 289–299 (1992).
78. Forrester, W. C. & Garriga, G. Genes necessary for C. elegans cell and growth cone migrations. *Development* **124**, 1831–1843 (1997).
79. Kubota, Y., Ohkura, K., Tamai, K. K., Nagata, K. & Nishiwaki, K. MIG-17/ADAMTS controls cell migration by recruiting nidogen to the basement membrane in C. elegans. *Proc. Natl.*

Acad. Sci. U. S. A. **105**, 20804–20809 (2008).
80. Ou, G. & Vale, R. D. Molecular signatures of cell migration in C. elegans Q neuroblasts. *J. Cell Biol.* **185**, 77–85 (2009).
81. Keil, W., Kutscher, L. M., Shaham, S. & Siggia, E. D. Long-Term High-Resolution Imaging of Developing C. elegans Larvae with Microfluidics. *Dev. Cell* **40**, 202–214 (2017).
82. Tuck, S. The control of cell growth and body size in Caenorhabditis elegans. *Exp. Cell Res.* **321**, 71–76 (2014).
83. Lažetić, V. & Fay, D. S. Molting in C. elegans. *Worm* **6**, e1330246 (2017).
84. Gritti, N., Kienle, S., Filina, O. & van Zon, J. S. Long-term time-lapse microscopy of C. elegans post-embryonic development. *Nat. Commun.* **7**, 12500 (2016).
85. Wang, J., Tokarz, R. & Savage-Dunn, C. The expression of TGFβ signal transducers in the hypodermis regulates body size in C. elegans. *Development* **129**, 4989–4998 (2002).
86. Colavita, A., Krishna, S., Zheng, H., Padgett, R. W. & Culotti, J. G. Pioneer axon guidance by UNC-129, a C. elegans TGF-beta. *Science* **281**, 706–709 (1998).
87. McGehee, A. M., Moss, B. J. & Juo, P. The DAF-7/TGF-β signaling pathway regulates abundance of the Caenorhabditis elegans glutamate receptor GLR-1. *Mol. Cell. Neurosci.* **67**, 66–74 (2015).
88. Hilliard, M. A. & Bargmann, C. I. Wnt signals and frizzled activity orient anterior-posterior axon outgrowth in C. elegans. *Dev. Cell* **10**, 379–390 (2006).
89. Inoue, T. *et al.* C. elegans LIN-18 is a Ryk ortholog and functions in parallel to LIN-17/Frizzled in Wnt signaling. *Cell* **118**, 795–806 (2004).
90. Zinovyeva, A. Y. & Forrester, W. C. The C. elegans Frizzled CFZ-2 is required for cell migration and interacts with multiple Wnt signaling pathways. *Dev. Biol.* **285**, 447–461 (2005).
91. Green, J. L., Inoue, T. & Sternberg, P. W. The C. elegans ROR receptor tyrosine kinase, CAM-1, non-autonomously inhibits the Wnt pathway. *Development* **134**, 4053–4062 (2007).
92. Richelle, A., Joshi, C. & Lewis, N. E. Assessing key decisions for transcriptomic data integration in biochemical networks. *PLoS Comput. Biol.* **15**, e1007185 (2019).
93. Joshi, C. J. *et al.* StanDep: Capturing transcriptomic variability improves context-specific metabolic models. *PLoS Comput. Biol.* **16**, e1007764 (2020).
94. Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *bioRxiv 2020.07.21.214387; doi: https://doi.org/10.1101/2020.07.21.214387* (2020) doi:10.1101/2020.07.21.214387.
95. Liao, J., Lu, X., Shao, X., Zhu, L. & Fan, X. Uncovering an Organ's Molecular Architecture at Single-Cell Resolution by Spatially Resolved Transcriptomics. *Trends Biotechnol.* (2020) doi:10.1016/j.tibtech.2020.05.006.
96. Waterhouse, R. M., Tegenfeldt, F., Li, J., Zdobnov, E. M. & Kriventseva, E. V. OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. *Nucleic Acids Res.* **41**, D358–65 (2013).
97. Kim, W., Underwood, R. S., Greenwald, I. & Shaye, D. D. OrthoList 2: A New Comparative Genomic Analysis of Human and Caenorhabditis elegans Genes. *Genetics* **210**, 445–461 (2018).
98. Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* **35**,

W193–W200 (2007).
99. Huang, X.-T., Zhu, Y., Chan, L. L. H., Zhao, Z. & Yan, H. An integrative C. elegans protein–protein interaction network with reliability assessment based on a probabilistic graphical model. *Mol. Biosyst.* **12**, 85–92 (2016).
100. Szklarczyk, D. *et al.* STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
101. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* **12**, 2825–2830 (2011).
102. Harris, T. W. *et al.* WormBase: a comprehensive resource for nematode research. *Nucleic Acids Res.* **38**, D463–7 (2010).

Acknowledgements

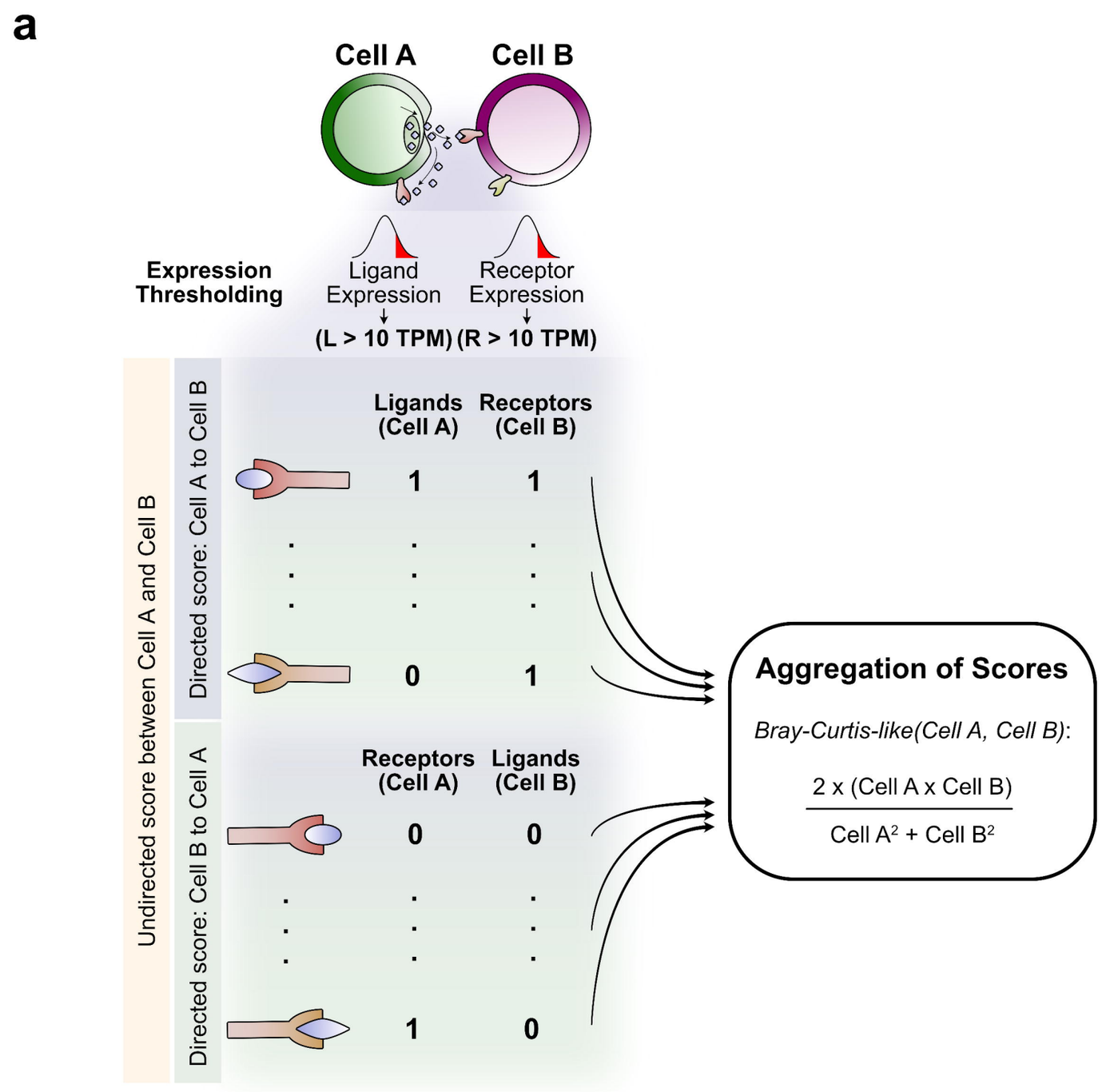
EA is supported by the Chilean Agencia Nacional de Investigación y Desarrollo (ANID) through its scholarship program DOCTORADO BECAS CHILE/2018 - 72190270 and by the Fulbright Commission. This work was further supported by NIGMS grant R35 GM119850 to NEL, a Lilly Innovation Fellows Award to CJ, Jefferson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

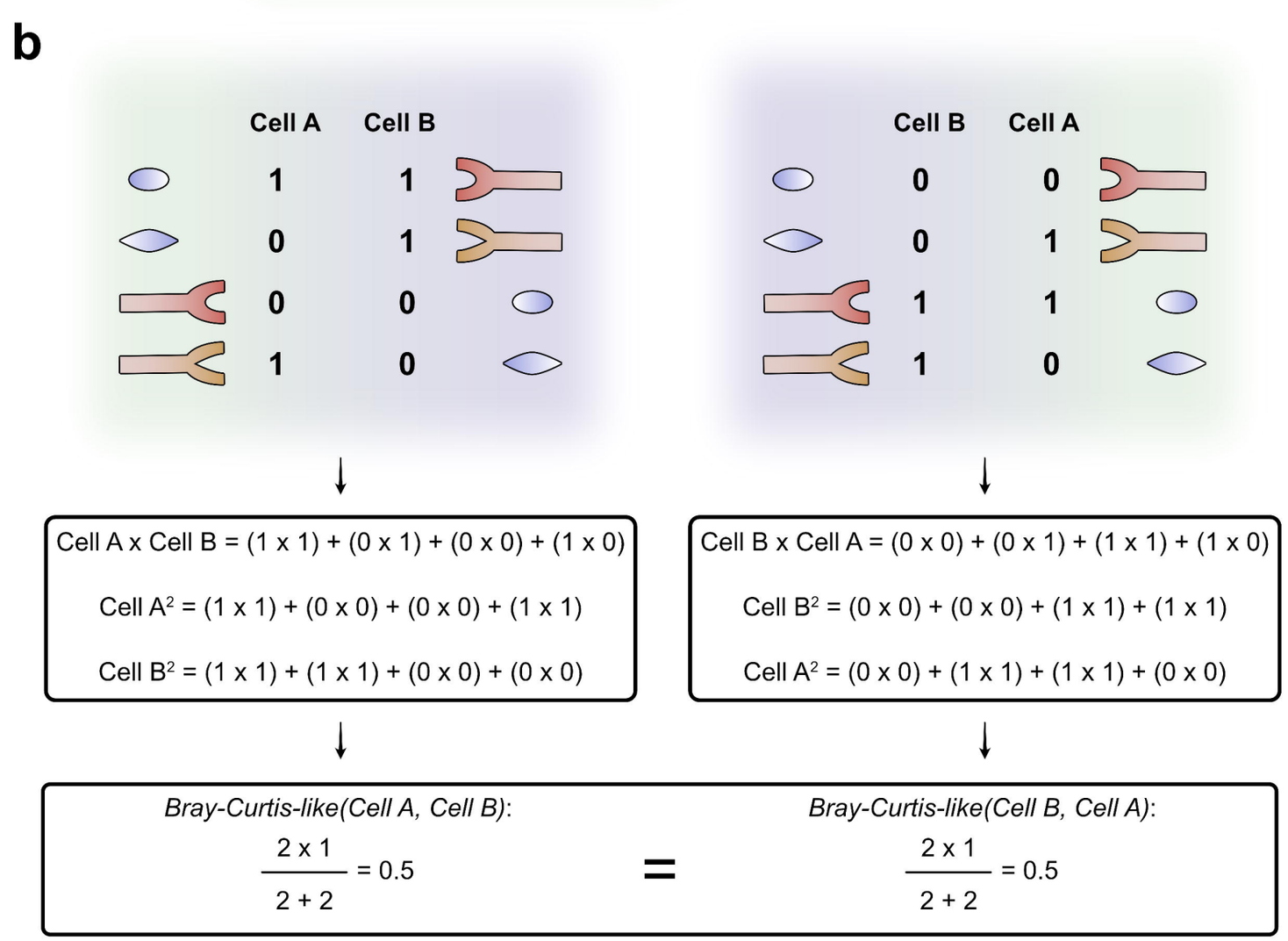
Author contributions

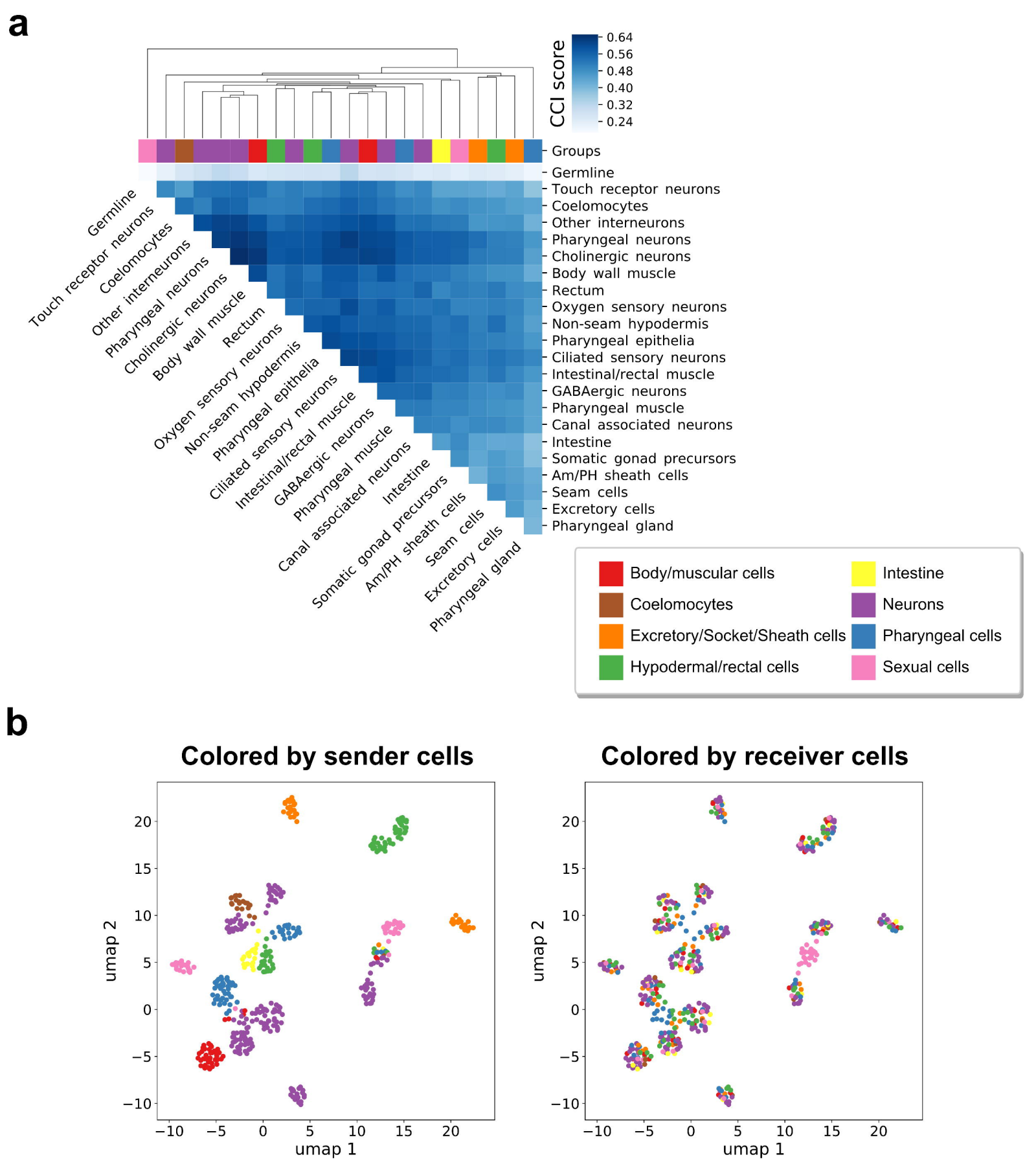
EA, CJ, EJO, and NEL conceived the work. EA and HMB annotated the cells in the previously published 3D digital atlas of *C. elegans* and analyzed their physical locations. EA, CJ, EJO and NEL analyzed the single-cell RNA-seq data. EA and IS generated the list of ligand-receptor pairs of *C. elegans*. 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. EA wrote the paper and all authors carefully reviewed and edited the paper.

Competing interests

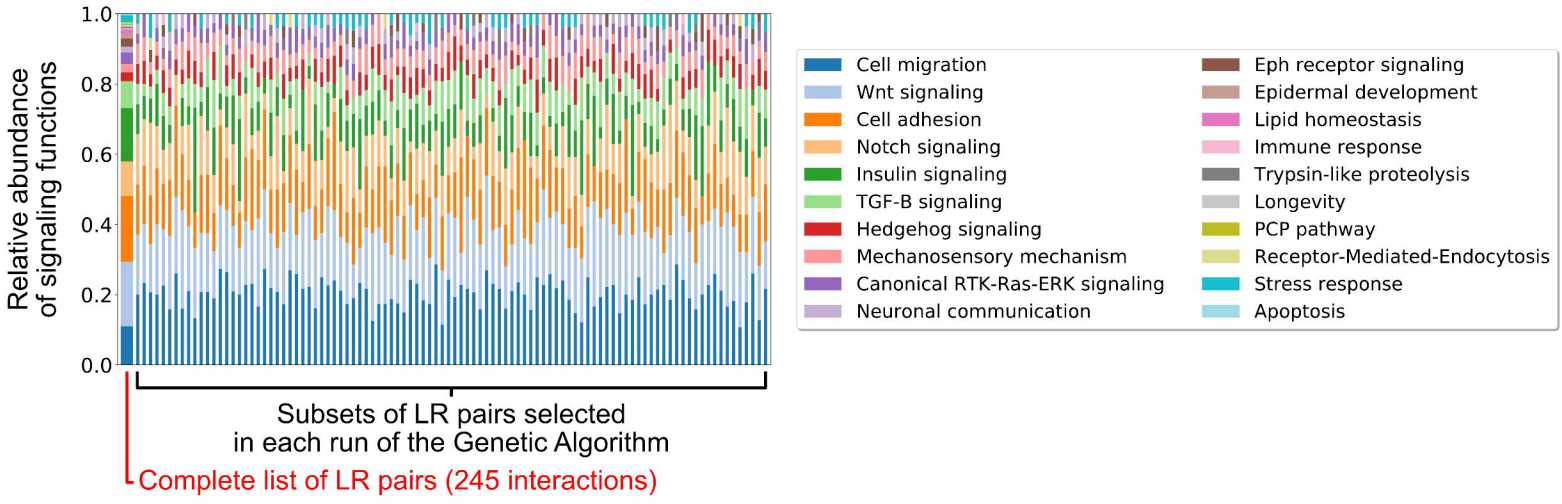
The authors declare no competing interests.


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

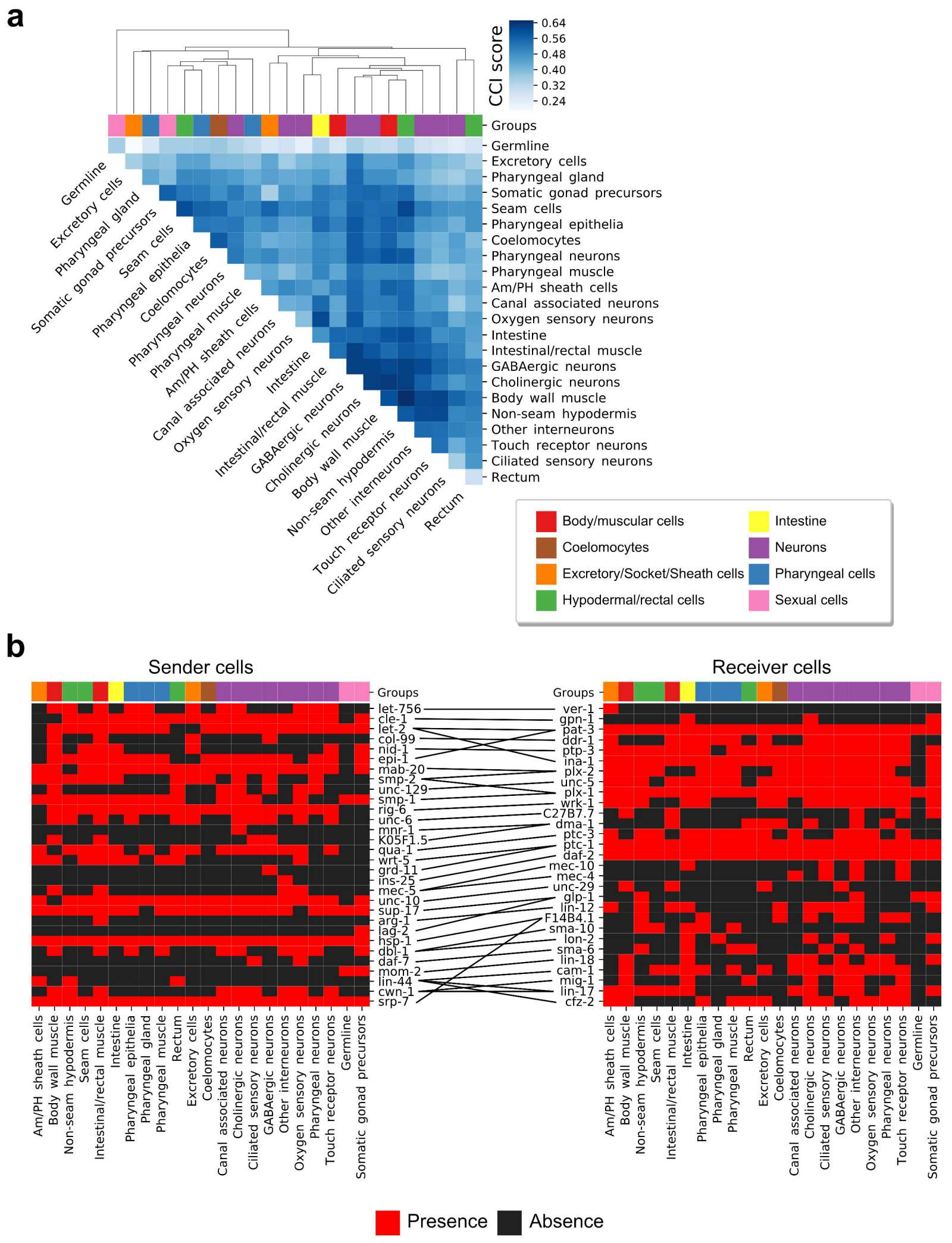

a
CCI score
0.64
0.56
0.48
0.40
0.32
0.24
Groups
Germline
Touch receptor neurons
Coelomocytes
Other interneurons
Pharyngeal neurons
Cholinergic neurons
Body wall muscle
Rectum
Oxygen sensory neurons
Non-seam hypodermis
Pharyngeal epithelia
Ciliated sensory neurons
Intestinal/rectal muscle
GABAergic neurons
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

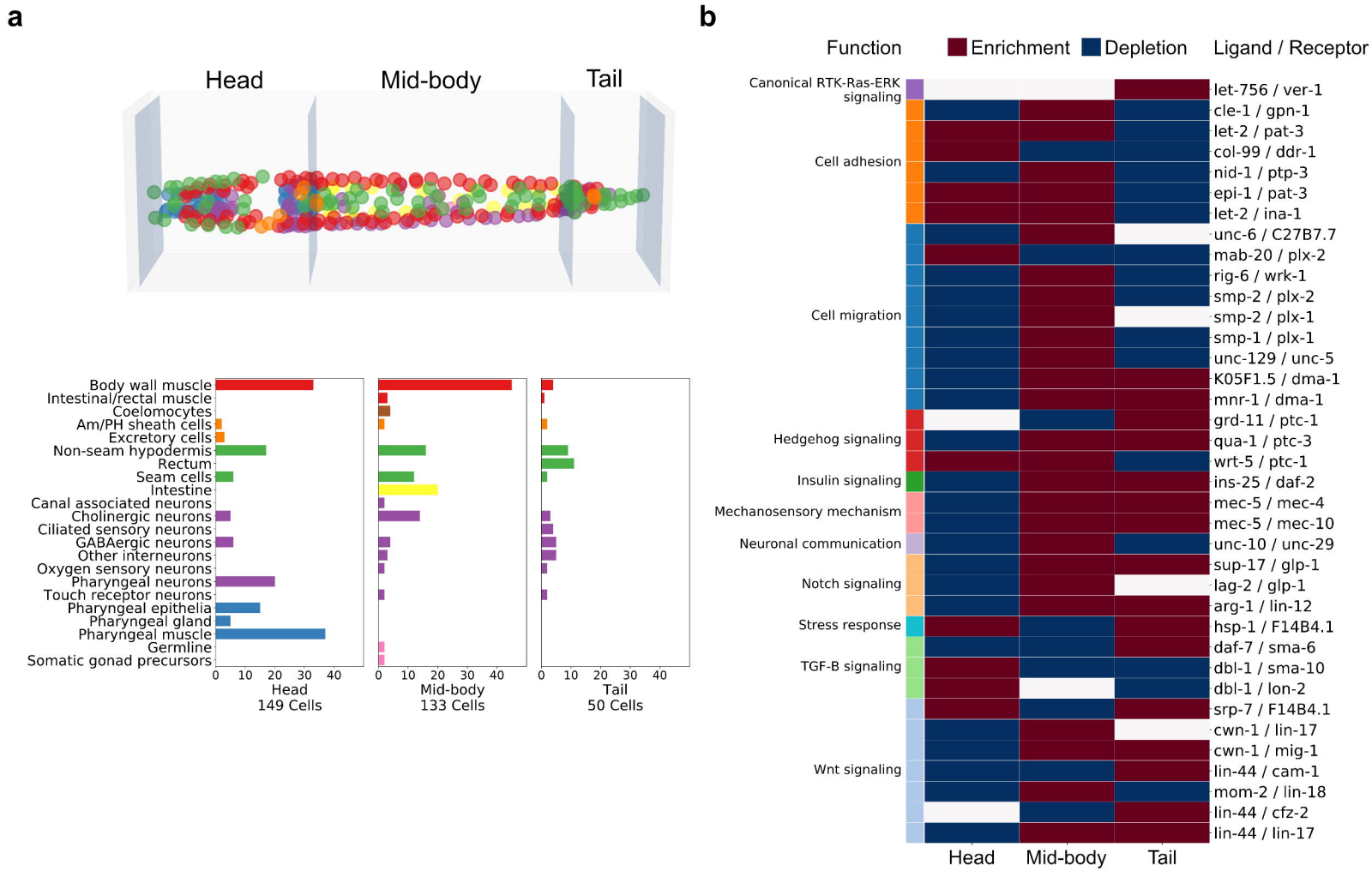
a

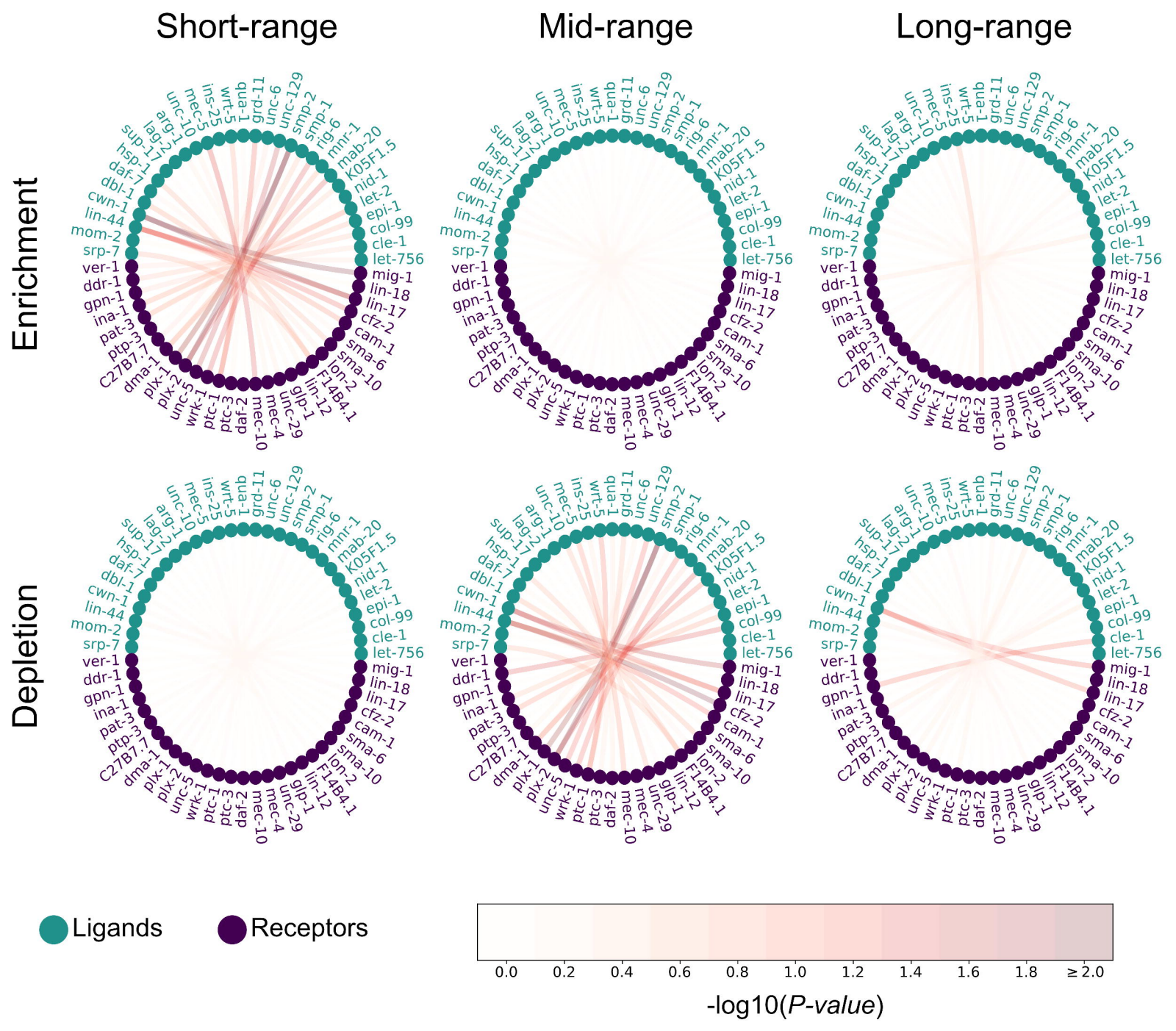


b

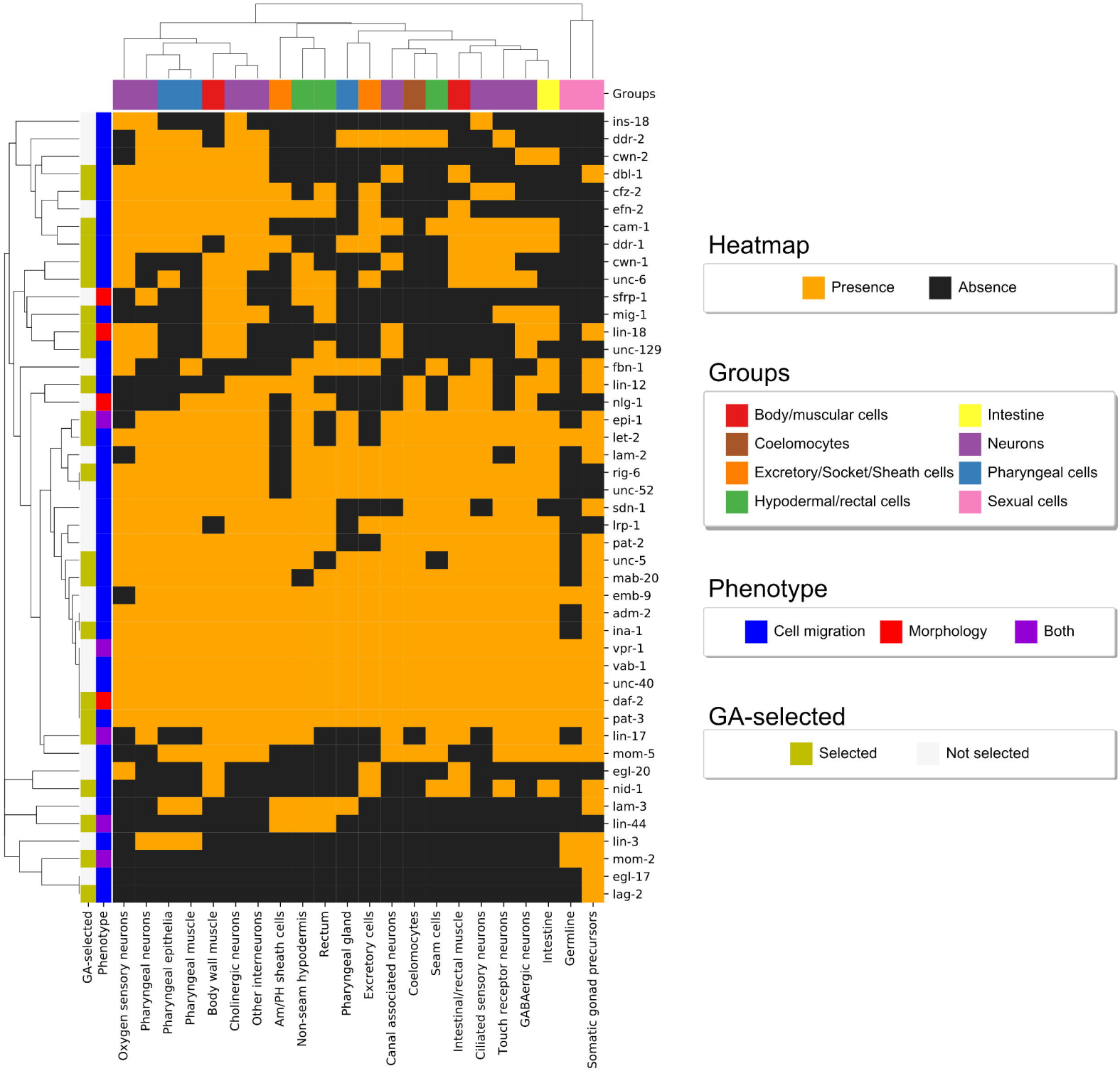
Signaling function	Change	Adj. P-value	Signaling function	Change	Adj. P-value
Cell migration	↑	3.983×10^{-17}	Eph receptor signaling	↓	4.472×10^{-13}
Wnt signaling	–	7.221×10^{-01}	Epidermal development	↓	3.048×10^{-23}
Cell adhesion	↓	2.711×10^{-10}	Lipid homeostasis	↓	3.048×10^{-23}
Notch signaling	–	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}

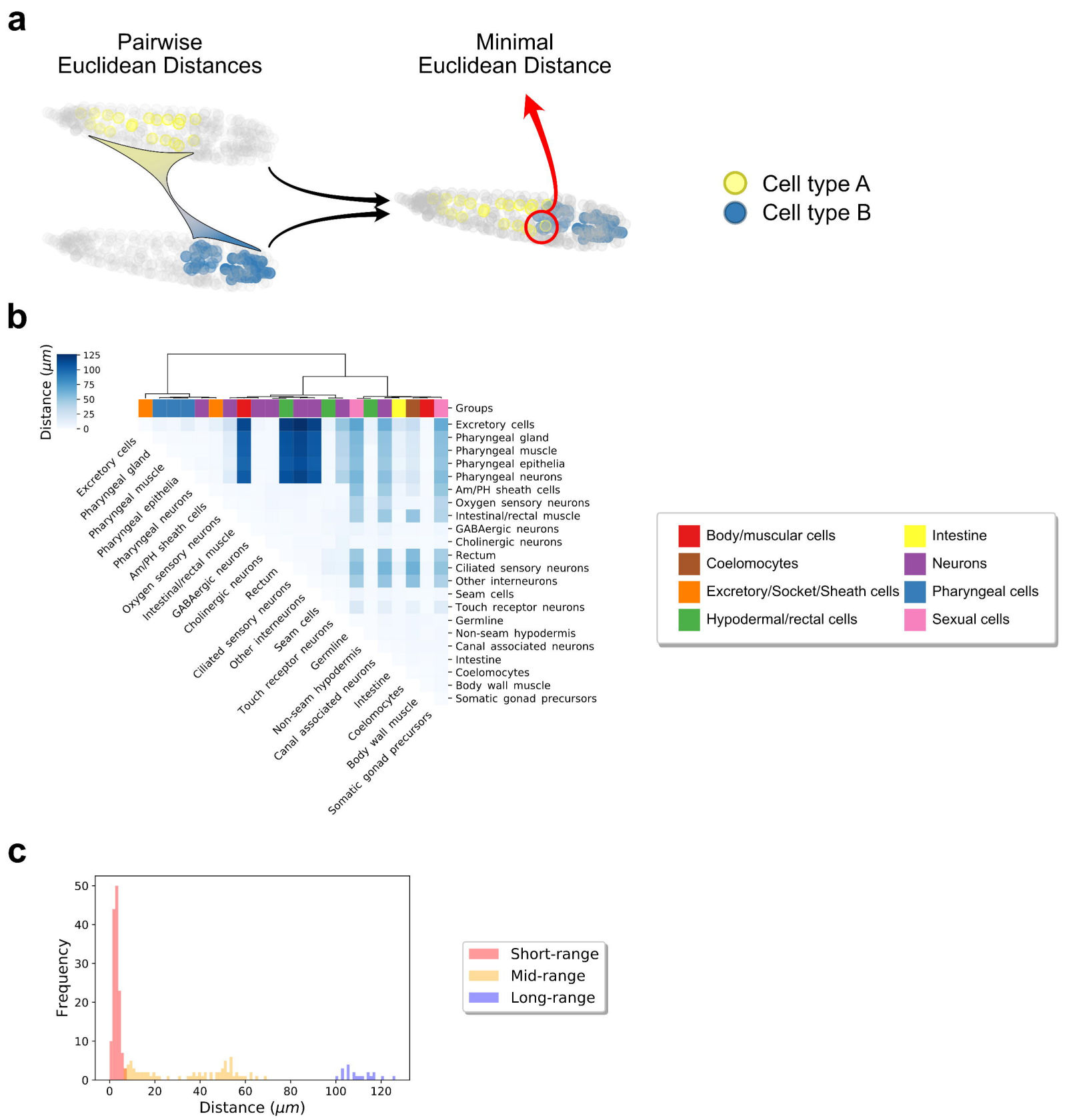

a
CCI score
0.64
0.56
0.48
0.40
0.32
0.24
Groups
Germline
Excretory cells
Pharyngeal gland
Somatic gonad precursors
Seam cells
Pharyngeal epithelia
Coelomocytes
Pharyngeal neurons
Pharyngeal muscle
Am/PH sheath cells
Canal associated neurons
Oxygen sensory neurons
Intestine
Intestinal/rectal muscle
GABAergic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

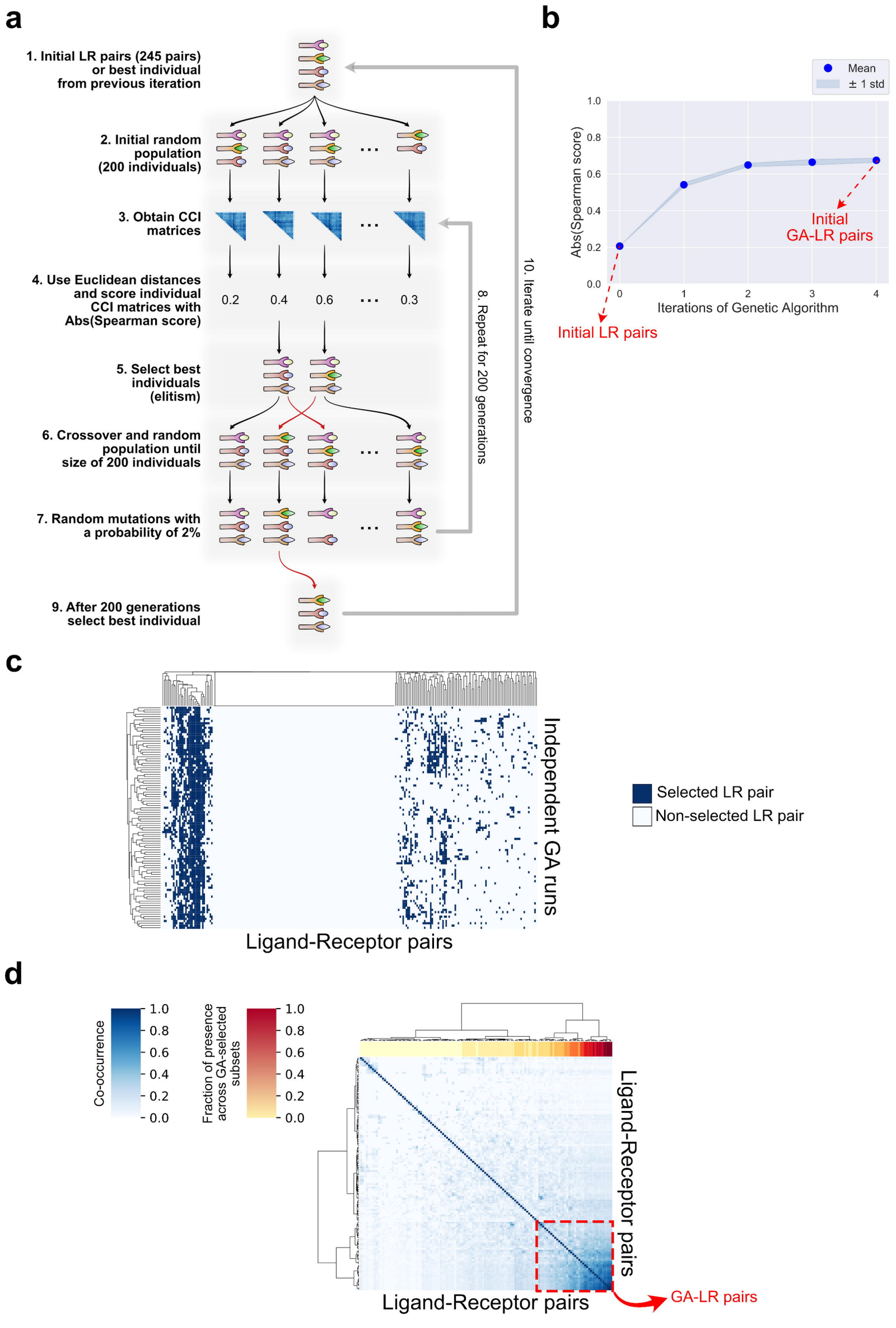

a
Head
Mid-body
Tail
Body wall muscle
Intestinal/rectal muscle
Coelomocytes
Am/PH sheath cells
Excretory cells
Non-seam hypodermis
Rectum
Seam cells
Intestine
Canal associated neurons
Cholinergic neurons
Ciliated sensory neurons
GABAergic neurons
Other interneurons
Oxygen sensory neurons
Pharyngeal neurons
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

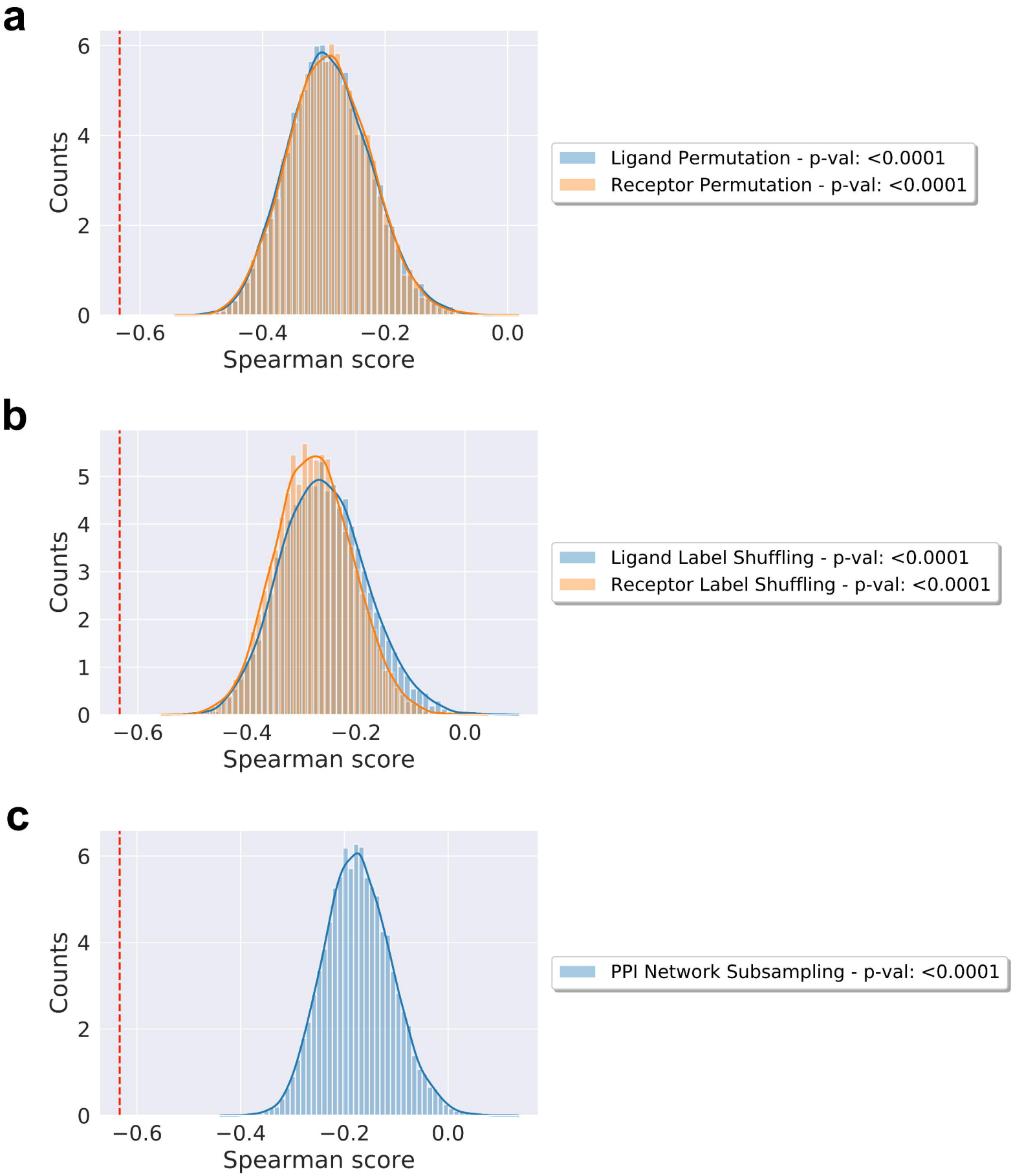


Short-range
Mid-range
Long-range
Enrichment
Depletion
Ligands
Receptors
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
≥ 2.0
-log10(P-value)


Groups
ins-18
ddr-2
cwn-2
dbl-1
cfz-2
efn-2
cam-1
ddr-1
cwn-1
unc-6
sfrp-1
mig-1
lin-18
unc-129
fbn-1
lin-12
nlg-1
epi-1
let-2
lam-2
rig-6
unc-52
sdn-1
lrp-1
pat-2
unc-5
mab-20
emb-9
adm-2
ina-1
vpr-1
vab-1
unc-40
daf-2
pat-3
lin-17
mom-5
egl-20
nid-1
lam-3
lin-44
lin-3
mom-2
egl-17
lag-2
GA-selected
Phenotype
Oxygen sensory neurons
Pharyngeal neurons
Pharyngeal epithelia
Pharyngeal muscle
Body wall muscle
Cholinergic neurons
Other interneurons
Am/PH sheath cells
Non-seam hypodermis
Rectum
Pharyngeal gland
Excretory cells
Canal associated neurons
Coelomocytes
Seam cells
Intestinal/rectal muscle
Ciliated sensory neurons
Touch receptor neurons
GABAergic neurons
Intestine
Germline
Somatic gonad precursors
Heatmap
Presence
Absence
Groups
Body/muscular cells
Coelomocytes
Excretory/Socket/Sheath cells
Hypodermal/rectal cells
Intestine
Neurons
Pharyngeal cells
Sexual cells
Phenotype
Cell migration
Morphology
Both
GA-selected
Selected
Not selected


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
50
40
30
20
10
0
0
20
40
60
80
100
120
Distance (μm)
Short-range
Mid-range
Long-range


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


a
Counts
Spearman score
Ligand Permutation - p-val: <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