Linking the Caenorhabditis elegans nervous system neural connections and developmental history

Ronan Perry

1 Introduction

The connectome of an organism is a map of its neural connections and provides a way to quantify aspects such as the distribution of connections and topological organization. A mapping of the human connectome could provide deep understandings of brain damage and recovery and yet a full map is not feasible in the near future [1]. The *Caenorhabditis elegans* nematode, however, has had its entire connectome well mapped out and extensively reported and thus is a model organism for connectomal studies [2][3]. Analysis of the *C. elegans*connectome has been used to predict locomotion [4] and study behavior on the basis of inferred computational circuits [5]. Recent work has also illuminated relations between the lineages of the *C. elegans*cells and their properties in the connectome, such as connection strength [6]. Methods of cell lineage discovery in more complex organisms have shown promise [7] and so uncovering relations between the neuronal cell lineages and connectome structure could shed light on the connectome properties of more complex organisms.

However, analysis of cell lineages and the connectome relies on having preprocessed and conveniently stored data. The *C. elegans*lineage data is neither of these. Rather, it contains information about both the male and hermaphrodite sexes and is not matched to any given connectome, among other complications. Using an extensive mapping of the hermaphroditic chemical connectome [2] as well as the full cell lineage data [8], a partial mapping is provided along with the corresponding connectome and lineage tree. Such data provides the opportunity for others to readily access the information and begin their own investigations.

2 Results

The connectome used here is extensive, taking into account non-muscle end organs which are usually ignored as well as 95 body wall muscle cells [2]. Of the 453 connectome cells, 347 or 76.6%, had their lineages' identified. Section 3 gives detail on these unmapped cells, most of which are the body wall muscle cells. A set of csv files have been generated to store this mapping and to make analysis of the relationship between the *C. elegans*chemical connectome and developmental history convenient. These csv files and their respective headers are described as follows. The data are visualized in Figure 1.

1. *herm_cell_to_lineage_cell.csv*: The list of cells in the connectome and the corresponding cells in the lineage list, if determined.
2. *herm_chem_A_lineage.csv*: The synaptic weight adjacency matrix of the mapped cells, a subset of the full adjacency matrix.
3. *herm_chem_A_lineage_cells.csv*: The list of cells forming the synaptic weight adjacency matrix. The ith cell corresponds to the ith row and column in the matrix.
4. *herm_lineage_instance.csv*: The list of observed cells and their developmental paths from the zygote to the connectome cells. This assumes an arbitrary realization of the stochastic splits, see Section 4.1 for details.
5. *herm_lineage_instance_edgelist.csv*: The edge list of the full developmental tree of the connectome cells, assuming an arbitrary realization of the stochastic splits. The edge list is a list of vertex

pairs, parent to child, with a developmental path serving as a unique identifier for each vertex. Approximately half of the vertices, all of the leaves in the tree, are the connectome cells. Some of the remaining vertices are key cells observed during the development, such as the zygote. However, most of the remaining vertices are unobserved stages in the developmental history whose developmental paths are inferred from the necessary divisions needed to form the binary tree.

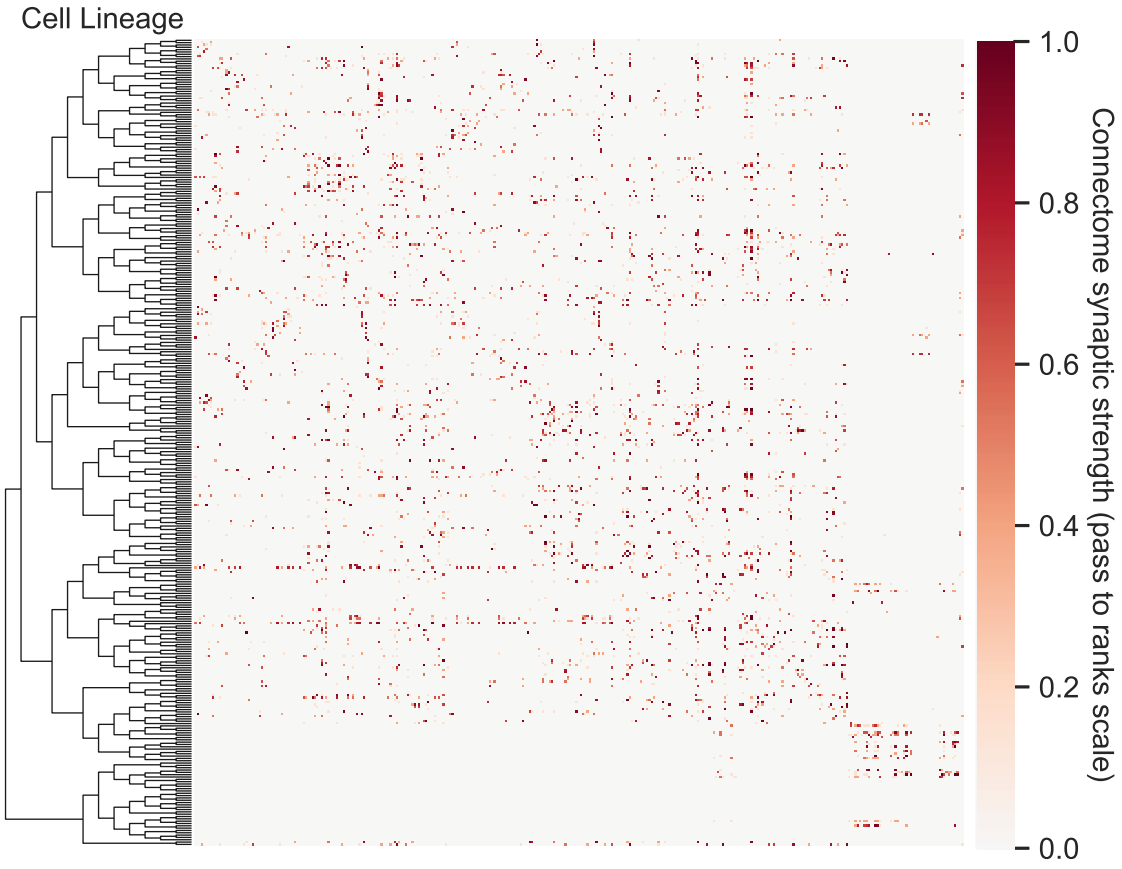

Figure 1: The connectome matched with and sorted according to the developmental history of its cells. Weights are scaled by the pass to ranks method in the open-source software package GraSPy.

3 Discussion

A partial mapping of these cells has been made and preprocessed into easily parsable formats. It is hoped that this will allow for the continuation of future work in this area. However, not all cells have been mapped. These cells fall into one of six categories, listed below.

- *syncytial pharyngeal muscle cells*: These cells are numerically distinguished as *pm1D-pm8D* in the connectome but are differentiated further in the cell lineage list into left and right divisions such as *pm1DR* and *pm1DL*. These are syncytial cells resulting from the fusion of multiple, in this case two, uninuclear cells.
- *body wall muscles*: These 95 end organ cells are labeled but not in a way that allows for them to be uniquely distinguished.

- *excretory glands*: These face the same left and right problem as the syncytial pharyngeal muscle cells.
- *basement membrane*: This is not a cell but is targeted by some of the pharyngeal cells.
- *intestinal cells*: The connectome has a single intestinal, *int*, cell listed and yet the lineage has many different intestinal cells. They, and only they, come from the E progenitor cell.
- *hypodermis*: The connectome has a single hypodermal, *hyp*, cell listed. Yet, the hypodermis is very complex, there being one large syncytium, *hyp7*, as well as many smaller syncytia which are all listed in the lineage list.

Most of these cells are end organs and so it is debatable how much their inclusion is necessary in understanding the structure of the connectome. With future work, however, these discrepancies can be dealt with and a better understanding of the relation between the cell lineages and connectome can be had.

4 Methods

4.1 Raw Data Description

The *C. elegans*hermaphrodite connectome data are provided as an adjacency matrix by the work of Cook et. al. [2]. Edges between nodes in the connectome graph correspond to synaptic connections in the animal, where the edge weight is proportional to a measure of synaptic strength [2]. The chemical synapse graph is directed and is used as directed connections better reveal sensor-to-neuron and neuron-to-endorgan connections. This chemical synapse connectome of the hermaphrodite consists of 453 nodes and 4879 edges, representing a sparse 2.4% of the total number of possible connections.

The cell lineage of *C. elegans*was downloaded from WormAtlas 1.0 [8], an adaption of Sulston [3]. The lineage cell list consists of 1358 observed cells from various stages of the development and includes cells from both the male and hermaphrodite. The cell lineage describes the divisions and cell fates leading from the single cell zygote to the adult animal. Not only is the *C. elegans*cell lineage fully known, the somatic cell lineage is mostly invariant. That is, most cell divisions are deterministic and produce specific cells such that the exact patterns of differentiation can be traced from the zygote to specific adult cells [9].

The lineage of each cell is coded by a string of characters. The uppercase characters at the start denote the blast cell ancestor. The following lower case characters are separated from the blast cell label either by a space, for embryonic cells, or by a period, for post-embryonic cell. Each lower case letter marks a division and the relative position of the cell following that division to its sister cell [9]. A cell and its sister cell take on one of the following pairs of positions: left or right, indicated by *l* and *r*, anterior or posterior, indicated by *a* and *p*, or lateral or ventral, indicated by *l* and *v*. For example, a cell with the lineage *AB apl* is a cell resulting from the successive anterior, posterior, and lateral divisions following the *AB* blast cell.

Some lineage mappings are not deterministic, however, as in cases where two lineages stochastically map to a pair of cells. A pair of cell names separated by a forward slash denotes two possible cells which may take on a specific linage. The slash-separated cell names are postfixed with a capitalized letter corresponding to the position of the cell in the worm. For instance, in the cell list there exist *P9/10L* and its left/right homolog *P9/10R*. During development, these cells arrange themselves about the midline of the developing gonad and then split left and right by pair. Thus, *P9* may end up on either the left or the right side of the nematode and so will any of the cells resulting from it [9].

4.2 Processing

A mapping from each cell in the connectome cell list to a cell in the lineage list was attempted. Most names held a one-to-one correspondence or differed on just a basis of notation. Cells whose lineages are stochastically determined were kept in a lineage-ambiguous format. Some cells in the connectome could not be mapped to a single cell in the lineage list and so were not included, see Section 3 for details on these cells. The new chemical connectome adjacency matrix was simply the 347 vertices which could be mapped to a cell in the lineage list.

The lineage list was processed to split the branching path from the blast cell. Additionally, cells which held identical names in the lineage list but were mapped, on the basis of their lineage, were then renamed with the distinct names from the connectome cell list. In order to create an edge list representation of the full hierarchical lineage tree, arbitrary realizations of the stochastic lineages were selected. This represents a single possible developmental trajectory. Each mapped cell from the connectome corresponds to a leaf in the developmental tree and although the root and some of the interior nodes are known, most interior cells are not observed. However, using the stored lineages of the leaves and propagating back up the tree to the root, all such interior nodes with their respective lineages were generated. Propagating from the leaves ensured that only cells relevant to the hermaphrodite's chemical connectome were included. A given cell is the parent of any cell whose lineage is the string formed by appending a lowercase letter to the parent's lineage. Using the full node list and this edge relation, an edge list was constructed for the tree defining the developmental trajectory from the initial zygote to the mapped cells in the hermaphrodite's connectome.

This processing is easily run given an updated connectome to lineage mapping, generating an updated edge list and lineage tree.

Data Availability The data are contained in csv files located in the private Github repository *eleGraSPy* on the Neurodata Design account at `https://github.com/neurodata/eleGraSPy/tree/rflperry/lineages`.

Code Availability The code is located in Python jupyter notebooks in the same private repository as the data.

Bibliography

[1] O. Sporns, G. Tononi, and R. Kötter, "The human connectome: A structural description of the human brain," *PLOS Computational Biology*, vol. 1, no. 4, 09 2005. 1

[2] S. J. Cook *et al.*, "Whole-animal connectomes of both c. elegans sexes," *Nature*, 2019. 1, 3

[3] J. Sulston, "The nematode *Caenorhabditis elegans*," vol. 17, no. Ch. 5, pp. 123–155, 1988. 1, 3

[4] B. L. J. C. Beckman, "Neuronal network of c. elegans: from anatomy to behavior," 2007. 1

[5] T. A. Jarrell *et al.*, "The connectome of a decision-making neural network," *Science*, vol. 337, no. 6093, pp. 437–444, 2012. 1

[6] A. Pathak, N. Chatterjee, and S. Sinha, "Developmental trajectory of the Caenorhabditis elegans nervous system governs its structural organization," *arXiv e-prints*, p. arXiv:1903.11972, Mar 2019. 1

[7] E. Shapiro, “On the journey from nematode to human, scientists dive by the zebrafish cell lineage tree,” *Genome Biology*, vol. 19, no. 1, p. 63, May 2018. 1

[8] Z. Altun *et al.*, “Wormatlas,” *http://www.wormatlas.org*, (ed.s) 2002-2019. 1, 3

[9] J. Sulston and H. Horvitz, “Post-embryonic cell lineages of the nematode, caenorhabditis elegans,” *Developmental Biology*, vol. 56, no. 1, pp. 110 – 156, 1977. 3