# Cell lineage predicts neural connectivity beyond cell type

**Jordan K. Matelsky**[a,b], **Brock Wester**[a], **and Konrad P. Kording**[b,c,d]

[a]Johns Hopkins Applied Physics Laboratory, Laurel, Maryland, United States; [b]Department of Bioengineering, University of Pennsylvania, Philadelphia, Pennsylvania, United States; [c]Department of Neuroscience, University of Pennsylvania, Philadelphia, Pennsylvania, United States; [d]CIFAR, LMB program

**As an organism develops, a zygote becomes the body through repeated cell division. This process produces the mitotic family tree, and cells specialize into their ultimate phenotype through interaction with other cells but also through the history of divisions. Biologists often cluster cells in the body into "cell types," using e.g. their morphologies and molecular makeup. While morphomolecular cell type carries information about phenotype and function, there may be unknown, missing information available in this family tree. Emerging methods are making these lineage trees progressively observable. Here, using the complete mitotic family tree and connectome of the nematode *C. elegans* we ask about the role of cell-types and their family tree. We can thus evaluate how well we can predict synaptic connection with only cell-type information versus with the family tree lineage. We show that neglecting lineage can produce misleading insights into the mechanisms underlying neural wiring: underlying lineage can confound the effect of cell type. These results suggest that the concept of cell-type needs to be re-thought in the context of this emerging knowledge about cell lineage.**

**connectomics | development | *C. elegans* | cell lineage | cell types**
**Correspondence:** ***matelsky@upenn.edu***

## Introduction

Cell types, a way of grouping cells with similar phenotype, are central to modern biology and are a major topic of ongoing research. Cell types are often based on gene expression Uhlén et al. (2015), cell physiology Neunuebel and Knierim (2012), or other phenotypic properties Hendry and Calkins (1998); Zeng and Sanes (2017). Massive efforts, like the Allen Institute for Brain Science Cell Types Database, seek to harmonize these efforts across the breadth of the brain Thompson et al. (2014); Erö et al. (2018); Tasic et al. (2018); Zeng and Sanes (2017). But such cell type definitions are most meaningful if the cell type really does capture a unique or complete characterization of a cell's activity and phenotype. How much phenotypic variance does cell type explain, and how much is missing?

Cell types — among other things — are thought to guide how the body is made during development. In the brain, cell types guide the wiring of the brain through many mechanisms Uhlén et al. (2015); Rees et al. (2017); Erö et al. (2018); Toga and Thompson (2014), including the selective steering of neural growth cones Marquardt et al. (2005). But cell type alone does not explain the complex but highly consistent connectivity of the mature brain Barabási and Czégel (2021); Zeng and Sanes (2017). We propose that cellular lineage — the family tree of mitotic cell splits — plays a major role in steering connectivity patterns. Techniques like electron microscopy connectomics now yield comprehensive, synapse-resolution connectivity datasets Glasser et al. (2016); Huang et al. (2021); Shapson-Coe et al. (2021); Witvliet et al. (2021). In combination with these connectomes, we can now ask how much of the connectivity of a brain is affected by cell types, and how much by developmental lineage Kerstjens et al. (2022).

Is it important to know how lineage and cell type relate to cell phenotypes distinctly? Suppose only cell lineage mattered when determining a cell phenotype. As cell types often occur in a compact part of the lineage tree, it would appear that cell type "explains" much of the variance that is actually attributable to lineage. And the same would be true of the opposite condition. As such, a disentangling of cell type from mitotic lineage is essential in order for the concept of a cell type to be a meaningful and rigorous part of causal biological theories.

Here we hypothesize that cellular lineage can explain synaptic connections beyond that which can be explained by cell type alone. In this work, we test this hypothesis in the nematode *Caenorhabditis elegans*, the only organism to our knowledge for which exhaustive chemical synapse connectome data and exhaustive cellular lineage data are known Witvliet et al. (2021); Bhatla (2011); Cook et al. (2019). We find that cellular lineage does indeed correlate more closely with synaptic connectivity than cell type, and we confirm that this remains true after correcting for confounds like spatial (3D) distance. This discovery suggests that current cell types in *C. elegans* may be too broad of a categorization to accurately reflect cell-level properties in the nervous system and beyond.

## Results

Because cells close in the developmental lineage tree tend to have similar cellular fates, we hypothesized that

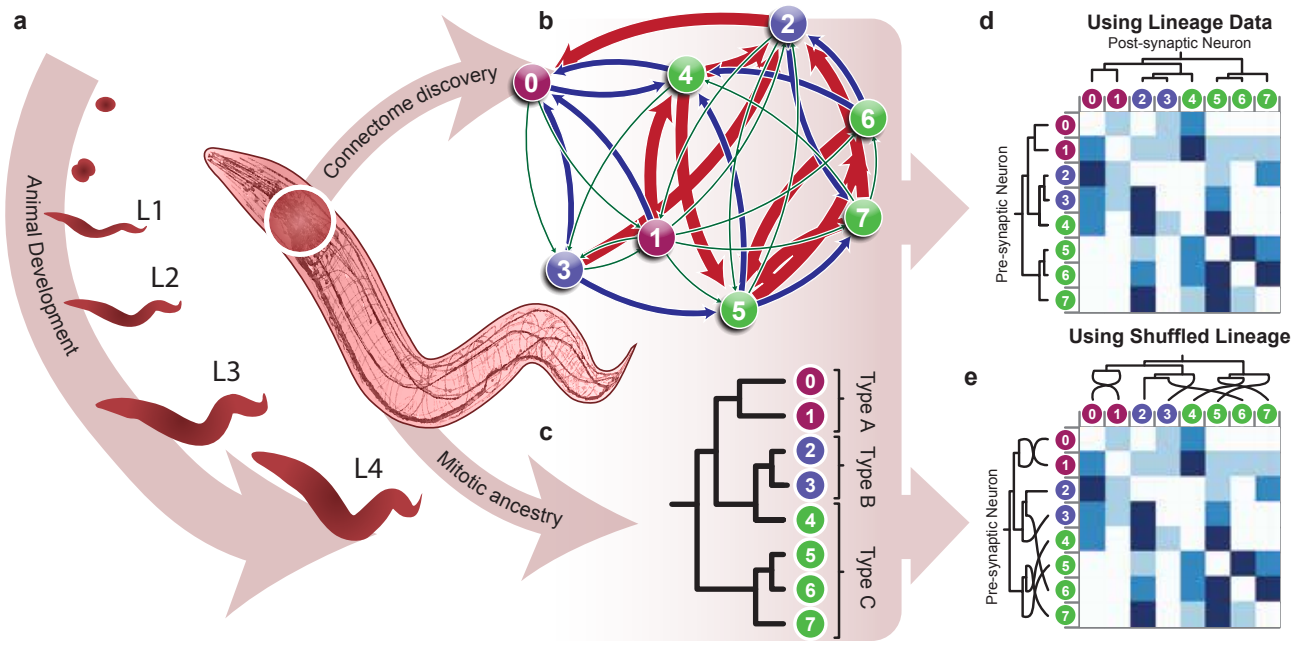


**Figure 1. Closely related cells have more similar synapsing patterns than predicted by cell type alone. a. Animal development.** As an animal develops and matures, progenitor cells split and divide into daughter cells. These cells eventually assume specific cell roles in the mature organism. As neural cells develop, they also form synaptic connections with local and distributed partners. **b. Connectome discovery.** Through modern electron microscopy connectomics efforts, we can now reconstruct a wiring diagram of these synaptic connections, in the form of a graph, where vertices represent neurons, and directed edges represent the synapses that communicate information from one neuron to the next. We can overlay on this network additional metadata such as human-described cell-types (here, represented by vertex color). **c. Mitotic ancestry.** By tracking highly stereotyped organism development, each neuron can be placed in a tree of mitotic ancestry, with the tree's root at the zygote cell. **d. Comparing lineage data with connectivity.** By comparing the graph of chemical synapses — in the form of an adjacency matrix — to the neural cells' lineages, we determine that closely-related cells in the cell lineage tend to have more similar wiring than distantly-related cells. **e. A null model of lineage.** When we shuffle the lineage tree cell identities, we see that this phenomenon is eliminated, as expected. Cell and animal morphology graphics are based upon OpenWorm resources Szigeti et al. (2014).

more closely related neurons would have greater overlap in their synaptic partners. To test this, we first compared the power of cell type in predicting the synaptic adjacency matrix (**Fig. 3**), and then we compared with results predicted by the *C. elegans* cell lineage tree (**Fig. 1**).

We discovered that more closely related neurons tended to have a greater overlap in their synaptic partners. This effect was more significant in downstream partners (i.e., the number of synaptic targets two neurons have in common, as a function of those neurons' distance in the lineage family tree), but was also true of presynaptic partners (i.e., the number of predecessors that two neurons have in common). This significance was most pronounced in the mature adult, and least pronounced in early (i.e., L1 – L2) development.

### Cell lineage explains more variance in synaptic targets than cell type

We explored the relationship between cell type and connectivity patterns, and the relationship between cell lineage and connectivity patterns. We discovered that the closeness of two neurons in the developmental lineage of an organism can explain more variance in connectivity than the cell type. The effect was more pronounced in cells of mismatched cell-type (**Fig. 4 a & b**). No neuron pairs in our dataset had a lineage distance of 1 (which would represent a synapse between mother and daughter cells, which never occurs in the dataset) or of 0 (representing a self-loop, or autapse) (**Table 1**).

### The effect of cellular lineage on connectivity remains after accounting for physical distance

Previous work has suggested that neurons that are physically nearer tend to synapse on each other with greater frequency. This phenomenon is commonly called "Peters' Rule" Peters and Feldman (1976); Braitenberg and Schüz (2013); Rees et al. (2017); Jonas and Kording (2015). We find that the path-length distance between two neurons in the lineage tree of an organism is a stronger predictor of connectivity than the physical distance between the neurons, providing preliminary evidence of a reduced relationship/correlation between proximity and connectivity which runs counter to Peters' Rule (**Fig. 4**). When we regress physical distance between neurons out, the observed effects remain. All subsequent analyses are reported after accounting for physical distance.

### The effect of cell lineage on synaptic sources is less pronounced than on synaptic targets

We repeated our analysis, comparing synaptic sources instead of synaptic targets. Though the effect was less pronounced (**Fig. 4 c & d**), we discovered the same effect: More closely related neurons in the lineage family tree shared more similar synaptic targets, and more distantly related neurons tended to have more diverse synaptic sources.

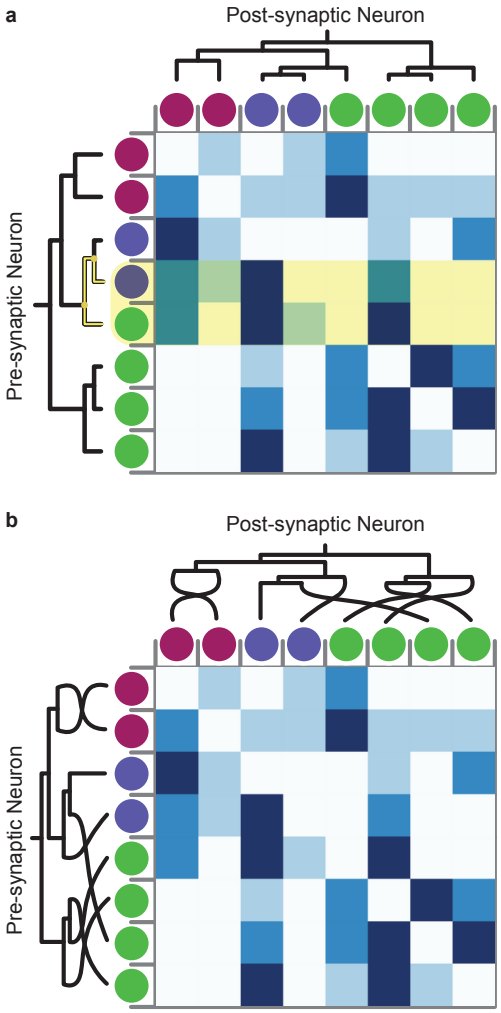


**Figure 2. A model of a connectome adjacency matrix with coregistered lineage information; and a shuffled version in which the structure of the lineage tree and the structure of the adjacency graph remain the same, but the mapping between the two is shuffled.** In this model nervous system, there are eight neurons, with three cell types (red, blue, and green). The adjacency matrix of this nervous system is presented with blue squares where synapses are present: The blue-shaded region at $(1, 2)$ indicates that the cell with ID $1$ has a synapse (directed edge) on the cell with ID $2$. Darkness indicates synaptic count. On the left and top edges of the matrix, we overlay the tree of lineage information. The root of the tree is the zygote, and each bifurcation represents a single instance of mitosis. (Non-neural cells are not shown.) **a. Two cells of different cell type, with similar connectivity in the adjacency matrix.** We highlight two cells (yellow rows) which, despite being different cell-types, have similar connectivity patterns due to their close relation in the cell lineage family tree. These cells have a lineage distance of 2 according to our metric, because the shortest path (yellow line) between the two nodes passes through two bifurcations (yellow nodes). We discover that more closely-related cells have similar connectivity properties, whether or not they are of the same cell type. **b. A null model, where cell identities have been shuffled.** In this perturbed model, the effect that cell lineage has on downstream similarity vanishes entirely.

### Shuffling the lineage tree eliminates the effect of lineage on neural connectivity

In order to ensure that our results were not simply some emergent property of densely connected networks, we produced a simple lineage-tree null model by randomly shuffling neuron lineage identities prior to associating with neurons in the connectome network. This model avoids changing the underling true structure of the lineage tree and the connectome network. We confirm

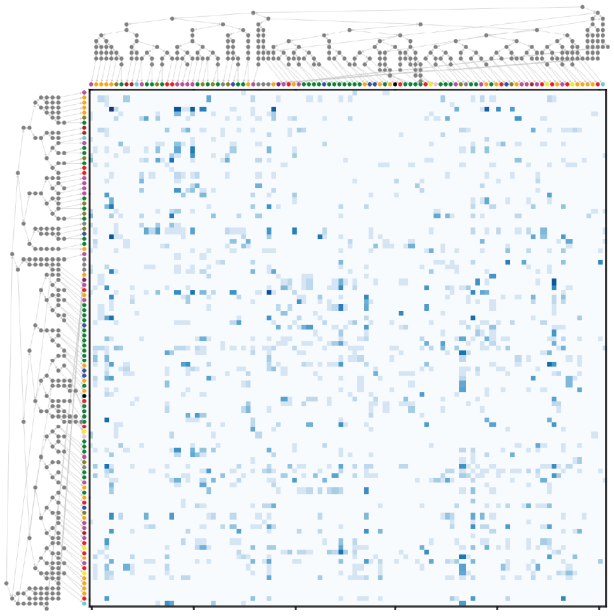

**Figure 3. The *C. elegans* connectome adjacency matrix with coregistered lineage information.** Cell types are indicated with leaf-vertex coloring Witvliet et al. (2021). All other vertices in the lineage tree represent progenitor cells — cells that eventually derive cells that express chemical synapses represented in this connectome dataset. Lineage tree information is computed for the entire dataset, but is only rendered in this figure for cells with annotated chemical synapses Bhatla (2011); Witvliet et al. (2021). Matrix ordering is based on a topological sort of the cell lineage, and the rendering here represents the connectome after pre-processing and folding as described in *Methods*.

that our null model with shuffled cell-identity mapping eliminates the effects seen in these analyses (**Table 1**), even though the structure of the tree and adjacency matrix remained the same. This confirms that the effect we see is not due purely to the topology of these two graphs.

## Discussion

The ability to predict the phenotypic propeties of a cell — such as synaptic connectivity — without observing it directly is a long-standing goal of biology research. Traditionally, human-defined cell types have served as a valuable proxy-measure. Here we have shown that cellular lineage can explain more of the variance of the connectome than our current definition of cell types alone. We discover that cell lineage is much more predictive of a cell's synaptic *targets* than its synaptic *sources*. We therefore conclude that (1) two closely related cells are more likely to synapse on the same downstream targets than two distantly related cells; and (2) neurons do not incorporate the lineage information of their own targets when selecting which cells to synapse on. This helps to narrow down the possible biological mechanisms that underlie the formation of the connectome.

There are several confounds to consider, not least that there is a great deal of inter-individual variance in connectivity, even between isogenic individuals Witvliet et al. (2021). In order to control for this, we inspected

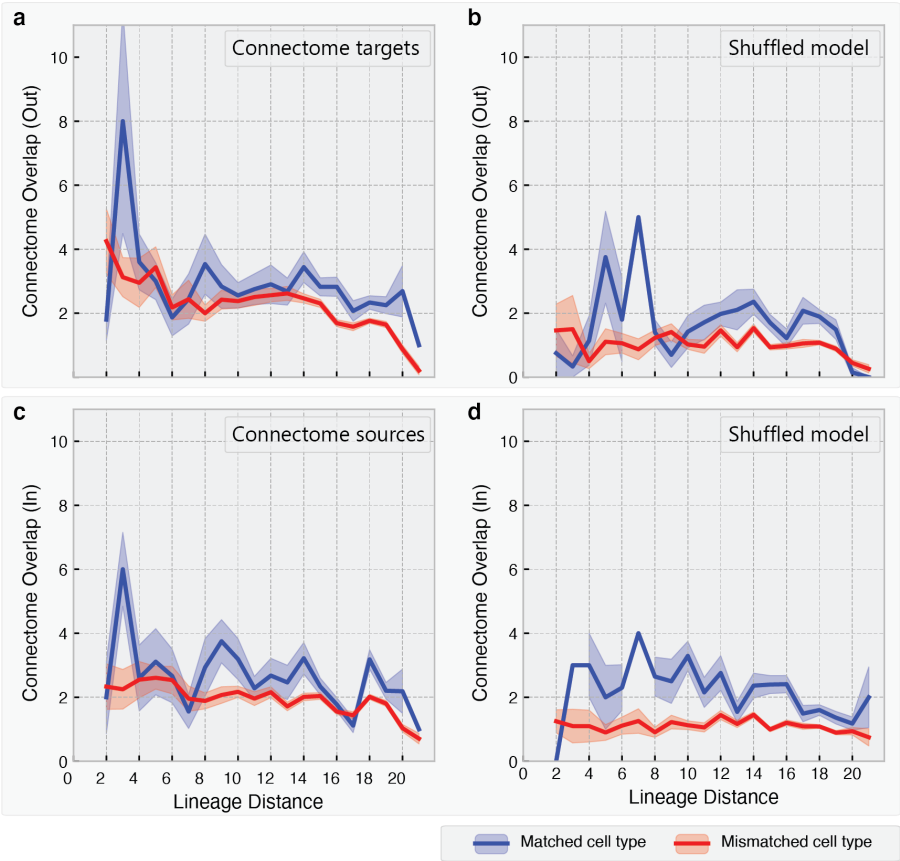


**Figure 4. The relationship between lineage distance and neighbor set overlap.** Neurons that are more closely related to each other tend to have more downstream-synapse partners in common. **a & b: Connectome similarity, measured by the size of the set-intersect of synaptic targets. a. The *C. elegans connectome.*** This effect is more pronounced between neurons of mismatched cell-type (red) than in neuron pairs of the same cell type (blue). Shaded areas indicate standard error of the mean (SEM). Note that there are no values at $x = 0$ and $x = 1$, as these indicate ($x = 0$) self-loops (omitted from our study) and ($x = 1$) synaptic connections to cell-lineage parent or daughter cells, which do not occur in this neural population. **b. A shuffled null model.** A connectome, when node identities are randomly shuffled, does not exhibit this same relationship. **c & d: Connectome similarity, measured by the size of the set-intersect of synaptic *sources*; and respective null model.** The effect size is lessened when predicting synapse sources rather than targets, but is still statistically significant. The null model (**Fig. 2**) likewise eliminates the effect of cell lineage on connectivity source selection.

the connectomes of eight *C. elegans* individuals at different developmental stages. Our findings were consistent across all adult individuals, though the effect was less pronounced (albeit still statistically significant) in adolescent individuals. We also found that sensorimotor cells tended to also be highly predictive of other sensorimotor neurons' targets. We believe this is due to the small neural population of this type, and the relatively wide repertoire of motor outputs per neuron. Despite these controls, the fact remains that the nematode connectome is quite sparse, and is highly constrained by the physical morphology of the *C. elegans* body plan. It is therefore possible that this effect may be attenuated in a larger or more densely-connected connectome.

For now, *C. elegans* remains the only model organism for which cellular lineage and chemical synapse-resolution connectomes are available. Parallel efforts have demonstrated a similar effect in aggregate in rodent Thompson et al. (2014); Kerstjens et al. (2022), and we believe there is still much more to be learned by achieving cell-resolution data. Beyond simply predicting the connectome, cellular lineage may explain organization of other complex anatomical structures, or carry more information in its network structure than can be efficiently encoded in the limited information storage of the genome Koulakov et al. (2021).

One simple explanation may be that there are simply more "true" cell types than the field has found so far. In future work, this same analysis might be repurposed to establish the minimal number of cell types to explain the connectome (or some other variable of interest) to a desired level of accuracy. We have released our tools as open-source resources to the community in order to encourage this type of analysis, or application of this same technique to other connectomes of interest (*Supplemental Materials*).

For the community of scientists studying generative models of connectomes, our paper adds an important extra aspect: good models should use knowledge of cell lineage, an approach that is progressively becoming possible. Our approaches can be combined with other generative models, *e.g.*, those that deal with distance Jonas and Kording (2015) and bring them closer to the rich developmental program implemented in brain development.

More generally, there is an ongoing conflict in modeling biological phenomena, where models are exceptionally simple for rather complex phenomena. For ex-

ample, connectomes are often modeled using graph theoretic random models. However, all these models ignore the rich biological knowledge that we have about the way biology actually organizes itself. As such, our paper is a motivation towards bringing models closer to reality without sacrificing data-driven rigor.

Our results show that cell type is confounded by underlying shared lineage data, at least in *C. elegans*. Due to the correlation of cell type and lineage, it may seem as if cell type explains more variance in projection data than it really does. This is often a problem in multimodal data, and such confounding factors are known across various domains ranging from biology to economics Pearl (2009); Angrist and Pischke (2014). Our results suggest that the success of cell type-based biology needs additional tests in order to validate the various domains to which cell type is often applied.

Here we have shown that lineage matters for neuron connectivity in *C. elegans*, beyond the effect of cell type. Previous research has shown that in many organisms, lineage matters for physiology Li et al. (2012); Cadwell et al. (2020); He et al. (2020). As such, lineage may be a relevant factor for many if not most biological phenomena. Quantifying lineage is starting to be more feasible with the advent of novel approaches to lineage estimation Masuyama et al. (2019). We thus suggest additional experiments across all domains where cell types are used, in order to disentangle their effects from that of mitotic cellular lineage.

## Methods

*C. elegans* nematode connectomes were downloaded from NemaNode.org Witvliet et al. (2021); Cook et al. (2019) and BossDB.org Hider et al. (2019) as directed GraphML files and *node-link* formatted JSON. Cell lineage information was downloaded from WormWeb.org as a JSON tree, and organized using the *NetworkX* and *Grand-Graph* Python graph analysis packages. Bhatla (2011); Hagberg et al. (2008); Matelsky et al. (2021)

Cell identities were coregistered with cell lineage and location information from WormWeb.org and WormBase.org, indexing by cell name. Bhatla (2011); Harris et al. (2010) Cell locations were recorded as the centroid of the soma of the 3D reference reconstruction from OpenWorm.org (code provided in *Supplemental Materials*). Harris et al. (2010); Szigeti et al. (2014) We computed the distances and similarity of pairwise sets of neurons by calculating the Euclidean (3D) distance between cell origins, and the lineage distance as the (undirected) shortest path between two nodes. Finally, we computed the similarity of two neurons in the connectome according to: (1) the size of the intersection of their in-common downstream synaptic targets; and (2) the size of the intersection of their in-common upstream synaptic partners.

Many neurons in the nematode nervous system have corresponding left- and right-side variants. Due to the way in which the lineage tree was constructed, these cells appear to be "distantly" related despite having nearly identical cell identities. Prior to analysis, we standardized the connectomes by combining these left- and right-side symmetric neurons. This "folding" process was performed by collapsing pairs of corresponding neurons into one, and using the union of edges of the resulting node. The list of symmetric neurons was manually reviewed for errors.

We performed an ordinary-least-squares regression on cell types (binary feature), 3D distance, and lineage distance (continuous features) in order to compute the correlation for each feature. Controlling for 3D distance, we compared the pairwise lineage distance — the minimum number of edges required to walk from one vertex to the other — with the pairwise connectome similarity — the number of upstream or downstream synaptic partners in common between two vertices (**Fig. 4**). Regressions were performed using the *statsmodels* (v0.13.1) and *scikit-learn* (v1.0.2) Python packages. Seabold and Perktold (2010); Pedregosa et al. (2011) All analyses are available as Jupyter Notebooks (see *Supplemental Materials*).

## Supplemental Materials

Data access tools, preprocessing tools, and notebooks to reproduce our analyses are available online at github.com/KordingLab/LineageConnectomics.

**ACKNOWLEDGEMENTS**

The authors thank Daniel Witvliet and all creators and annotators of the electron microscopy connectome, and Nikhil Bhatla for the collection of the *C. elegans* lineage tree data.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 (Adult) | 8 (Adult) |
|---|---|---|---|---|---|---|---|---|
| Matched type=i | $9.01 \times 10^{-6}$ | $4.86 \times 10^{-13}$ | $3.90 \times 10^{-6}$ | $1.70 \times 10^{-5}$ | $1.31 \times 10^{-10}$ | $3.51 \times 10^{-3}$ | $1.11 \times 10^{-2}$ | $6.94 \times 10^{-7}$ |
| Matched type=in | $8.06 \times 10^{-1}$ | $8.95 \times 10^{-1}$ | $8.53 \times 10^{-1}$ | $5.29 \times 10^{-1}$ | $6.30 \times 10^{-1}$ | $9.17 \times 10^{-1}$ | $4.70 \times 10^{-2}$ | $5.06 \times 10^{-2}$ |
| Matched type=sm | $1.51 \times 10^{-22}$ | $3.99 \times 10^{-9}$ | $1.91 \times 10^{-13}$ | $5.99 \times 10^{-17}$ | $3.22 \times 10^{-11}$ | $6.02 \times 10^{-10}$ | $3.40 \times 10^{-7}$ | $2.36 \times 10^{-6}$ |
| Matched type=sn | $5.90 \times 10^{-1}$ | $5.76 \times 10^{-1}$ | $8.99 \times 10^{-1}$ | $5.78 \times 10^{-1}$ | $7.35 \times 10^{-3}$ | $1.31 \times 10^{-2}$ | $8.38 \times 10^{-4}$ | $7.43 \times 10^{-5}$ |
| Matched type=s | $8.31 \times 10^{-1}$ | $5.44 \times 10^{-2}$ | $6.49 \times 10^{-1}$ | $7.85 \times 10^{-1}$ | $6.41 \times 10^{-1}$ | $3.50 \times 10^{-2}$ | $1.17 \times 10^{-1}$ | $1.18 \times 10^{-1}$ |
| Matched type=imn | $6.87 \times 10^{-2}$ | $5.59 \times 10^{-1}$ | $8.48 \times 10^{-1}$ | $9.26 \times 10^{-1}$ | $8.37 \times 10^{-1}$ | $4.67 \times 10^{-1}$ | $7.57 \times 10^{-1}$ | $4.67 \times 10^{-1}$ |
| Matched type=m | $3.08 \times 10^{-5}$ | $8.53 \times 10^{-6}$ | $9.00 \times 10^{-6}$ | $5.24 \times 10^{-3}$ | $6.02 \times 10^{-3}$ | $6.04 \times 10^{-10}$ | $3.92 \times 10^{-4}$ | $1.75 \times 10^{-4}$ |
| Matched type=n | $3.80 \times 10^{-1}$ | $6.61 \times 10^{-1}$ | $3.07 \times 10^{-1}$ | $8.61 \times 10^{-1}$ | $9.04 \times 10^{-1}$ | $8.99 \times 10^{-1}$ | $7.82 \times 10^{-1}$ | $5.68 \times 10^{-1}$ |
| Matched type=si | $6.08 \times 10^{-1}$ | $8.44 \times 10^{-1}$ | $5.22 \times 10^{-1}$ | $5.30 \times 10^{-1}$ | $1.40 \times 10^{-1}$ | $8.54 \times 10^{-1}$ | $6.78 \times 10^{-1}$ | $9.80 \times 10^{-1}$ |
| Matched type=sin | $4.76 \times 10^{-1}$ | $6.69 \times 10^{-1}$ | $6.54 \times 10^{-1}$ | $7.51 \times 10^{-1}$ | $9.45 \times 10^{-1}$ | $8.99 \times 10^{-1}$ | $8.41 \times 10^{-1}$ | $8.23 \times 10^{-1}$ |
| Matched type=mn | $6.22 \times 10^{-1}$ | $7.00 \times 10^{-2}$ | $5.63 \times 10^{-2}$ | $9.54 \times 10^{-1}$ | $6.26 \times 10^{-1}$ | $4.90 \times 10^{-1}$ | $2.26 \times 10^{-1}$ | $7.83 \times 10^{-1}$ |
| Matched type=im | | | | | | | $1.37 \times 10^{-1}$ | $9.44 \times 10^{-1}$ |
| Matched type=u | $2.93 \times 10^{-1}$ | $2.71 \times 10^{-1}$ | $2.11 \times 10^{-1}$ | $2.02 \times 10^{-1}$ | $1.21 \times 10^{-1}$ | $1.22 \times 10^{-1}$ | $5.55 \times 10^{-2}$ | $8.43 \times 10^{-2}$ |
| Geom. Distance | $6.28 \times 10^{-1}$ | $6.55 \times 10^{-1}$ | $3.98 \times 10^{-1}$ | $2.13 \times 10^{-1}$ | $3.41 \times 10^{-1}$ | $3.34 \times 10^{-2}$ | $2.77 \times 10^{-3}$ | $3.72 \times 10^{-4}$ |
| Lineage Distance | $1.32 \times 10^{-9}$ | $8.63 \times 10^{-6}$ | $4.18 \times 10^{-16}$ | $3.51 \times 10^{-11}$ | $4.10 \times 10^{-20}$ | $2.53 \times 10^{-11}$ | $1.95 \times 10^{-19}$ | $1.50 \times 10^{-21}$ |

**Table 1. Results of one-hot encoding and distance features.** Here we show the $p$-values of the features measured in our analysis. We show both the significance of one-hot encoded cell-type variables Witvliet et al. (2021) as well as variables for physical 3D distance and the path-length distance between two cells in the cell lineage "family tree" (**Fig. 3**). Entries are excluded from this table when there were insufficient data to perform the computation.

# Bibliography


Angrist, J. D. and Pischke, J.-S. *Mastering metrics: The path from cause to effect.* Princeton university press, 2014.

Barabási, D. L. and Czégel, D. Constructing graphs from genetic encodings. *Scientific Reports*, 11(1):1–13, 2021.

Bhatla, N. C. elegans Interactive Cell Lineage, 2011. URL http://wormweb.org/celllineage.

Braitenberg, V. and Schüz, A. *Cortex: statistics and geometry of neuronal connectivity*. Springer Science & Business Media, 2013.

Cadwell, C. R., Scala, F., Fahey, P. G., Kobak, D., Mulherkar, S., Sinz, F. H., Papadopoulos, S., Tan, Z. H., Johnsson, P., Hartmanis, L., et al. Cell type composition and circuit organization of clonally related excitatory neurons in the juvenile mouse neocortex. *Elife*, 9: e52951, 2020.

Cook, S. J., Jarrell, T. A., Brittin, C. A., Wang, Y., Bloniarz, A. E., Yakovlev, M. A., Nguyen, K. C., Tang, L. T.-H., Bayer, E. A., Duerr, J. S., et al. Whole-animal connectomes of both caenorhabditis elegans sexes. *Nature*, 571(7763):63–71, 2019.

Erö, C., Gewaltig, M.-O., Keller, D., and Markram, H. A cell atlas for the mouse brain. *Frontiers in neuroinformatics*, 12:84, 2018.

Glasser, M. F., Smith, S. M., Marcus, D. S., Andersson, J. L., Auerbach, E. J., Behrens, T. E., Coalson, T. S., Harms, M. P., Jenkinson, M., Moeller, S., et al. The human connectome project's neuroimaging approach. *Nature neuroscience*, 19(9):1175–1187, 2016.

Hagberg, A., Swart, P., and S Chult, D. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

Harris, T. W., Antoshechkin, I., Bieri, T., Blasiar, D., Chan, J., Chen, W. J., De La Cruz, N., Davis, P., Duesbury, M., Fang, R., et al. Wormbase: a comprehensive resource for nematode research. *Nucleic acids research*, 38(suppl_1):D463–D467, 2010.

He, S., Tian, Y., Feng, S., Wu, Y., Shen, X., Chen, K., He, Y., Sun, Q., Li, X., Xu, J., et al. In vivo single-cell lineage tracing in zebrafish using high-resolution infrared laser-mediated gene induction microscopy. *Elife*, 9:e52024, 2020.

Hendry, S. H. and Calkins, D. J. Neuronal chemistry and functional organization in the primate visual system. *Trends in neurosciences*, 21(8):344–349, 1998.

Hider, R., Kleissas, D. M., Pryor, D., Gion, T., Rodriguez, L., Matelsky, J., Gray-Roncal, W., and Wester, B. The block object storage service (bossdb): A cloud-native approach for petascale neuroscience discovery. *bioRxiv*, page 217745, 2019.

Huang, S. Y., Witzel, T., Keil, B., Scholz, A., Davids, M., Dietz, P., Rummert, E., Ramb, R., Kirsch, J. E., Yendiki, A., et al. Connectome 2.0: Developing the next-generation ultra-high gradient strength human mri scanner for bridging studies of the micro-, meso-and macro-connectome. *Neuroimage*, 243:118530, 2021.

Jonas, E. and Kording, K. Automatic discovery of cell types and microcircuitry from neural connectomics. *Elife*, 4:e04250, 2015.

Kerstjens, S., Michel, G., and Douglas, R. J. Constructive connectomics: how neuronal axons get from here to there using gene-expression maps derived from their family trees. Feb 2022. doi: 10.1101/2022.02.26.482112.

Koulakov, A., Shuvaev, S., and Zador, A. Encoding innate ability through a genomic bottleneck. Mar 2021. doi: 10.1101/2021.03.16.435261.

Li, Y., Lu, H., Cheng, P.-l., Ge, S., Xu, H., Shi, S.-H., and Dan, Y. Clonally related visual cortical neurons show similar stimulus feature selectivity. *Nature*, 486(7401):118–121, 2012.

Marquardt, T., Shirasaki, R., Ghosh, S., Andrews, S. E., Carter, N., Hunter, T., and Pfaff, S. L. Coexpressed epha receptors and ephrin-a ligands mediate opposing actions on growth cone navigation from distinct membrane domains. *Cell*, 121(1):127–139, Apr 2005. doi: 10.1016/j.cell.2005.01.020.

Masuyama, N., Mori, H., and Yachie, N. DNA barcodes evolve for high-resolution cell lineage tracing. *Current Opinion in Chemical Biology*, 52:63–71, 2019.

Matelsky, J. K., Reilly, E. P., Johnson, E. C., Stiso, J., Bassett, D. S., Wester, B. A., and Gray-Roncal, W. DotMotif: an open-source tool for connectome subgraph isomorphism search and graph queries. *Scientific Reports*, 11(1), Jun 2021. doi: 10.1038/s41598-021-91025-5.

Neunuebel, J. P. and Knierim, J. J. Spatial firing correlates of physiologically distinct cell types of the rat dentate gyrus. *Journal of Neuroscience*, 32(11):3848–3858, 2012.

Pearl, J. *Causality*. Cambridge university press, 2009.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Peters, A. and Feldman, M. L. The projection of the lateral geniculate nucleus to area 17 of the rat cerebral cortex. *Journal of neurocytology*, 5(1):63–84, 1976.

Rees, C. L., Moradi, K., and Ascoli, G. A. Weighing the evidence in peters' rule: does neuronal morphology predict connectivity? *Trends in neurosciences*, 40(2):63–71, 2017.

Seabold, S. and Perktold, J. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

Shapson-Coe, A., Januszewski, M., Berger, D. R., Pope, A., Wu, Y., Blakely, T., Schalek, R. L., Li, P., Wang, S., Maitin-Shepard, J., et al. A connectomic study of a petascale fragment of human cerebral cortex. *bioRxiv*, 2021.

Szigeti, B., Gleeson, P., Vella, M., Khayrulin, S., Palyanov, A., Hokanson, J., Currie, M., Cantarelli, M., Idili, G., and Larson, S. Openworm: an open-science approach to modeling caenorhabditis elegans. *Frontiers in computational neuroscience*, 8:137, 2014.

Tasic, B., Yao, Z., Graybuck, L. T., Smith, K. A., Nguyen, T. N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M. N., Viswanathan, S., et al. Shared and distinct transcriptomic cell types across neocortical areas. *Nature*, 563(7729):72–78, 2018.

Thompson, C. L., Ng, L., Menon, V., Martinez, S., Lee, C.-K., Glattfelder, K., Sunkin, S. M., Henry, A., Lau, C., Dang, C., et al. A high-resolution spatiotemporal atlas of gene expression of the developing mouse brain. *Neuron*, 83(2):309–323, 2014.

Toga, A. W. and Thompson, P. M. Connectopathy in ageing and dementia. *Brain*, 137(12): 3104–3106, 2014.

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. Tissue-based map of the human proteome. *Science*, 347(6220), 2015.

Witvliet, D., Mulcahy, B., Mitchell, J. K., Meirovitch, Y., Berger, D. R., Wu, Y., Liu, Y., Koh, W. X., Parvathala, R., Holmyard, D., et al. Connectomes across development reveal principles of brain maturation. *BioRxiv*, pages 2020–04, 2021.

Zeng, H. and Sanes, J. R. Neuronal cell-type classification: challenges, opportunities and the path forward. *Nature Reviews Neuroscience*, 18(9):530–546, 2017.