Cell lineage predicts neural connectivity beyond cell type

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

[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

The probability that a given neuron connects to another neuron is known to depend on their cell-types (e.g. sensory-interneuron). However, the developmental lineage could, in principle, allow each neuron to be unique, in a way making each neuron its own cell-type. Here we test this hypothesis by building predictors of connectivity that depend on cell-type and lineage on top of physical distance, using data from the *Caenorhabditis elegans* nematode. We find that the projection patterns of neurons between developmentally related cells (siblings) are more similar than that of cells with the same cell type, even after correcting for spatial distance. In the context of neural connectivity, the concept of cell type needs to be broadened to include lineage.

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

Introduction

Cell types, a way of grouping cells into similar phenotypes or genotypes, are central to modern biology. These cell types are often based upon gene expression (1), cell physiology (2), or other phenotypic properties (3, 4). Massive efforts, like the Allen Institute for Brain Science Cell Types Database, seek to harmonize these efforts across the breadth of the brain (4–7). But such cell type definitions are most meaningful only if the cell type really does capture a complete characterization of a cell's activity and phenotype. How much variance does cell type explain, and how much is missing?

Cell types — among other things — guide how the body is made during development. In the brain, cell types guide the wiring of the brain through many mechanisms (1, 6, 8, 9), including the selective steering of neural growth cones (10). But cell type alone does not explain the complex but highly consistent connectivity of the mature brain (4, 11). We propose that cellular lineage — the family tree of mitotic cell splits — plays a role in steering connectivity patterns. Techniques like serial-section electron microscopy connectomics now yield comprehensive, synapse-resolution connectivity datasets (12–14). 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 (15).

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 (16–18). We find that cellular lineage does indeed correlate more closely with synaptic connectivity than does cell type, and we confirm that this remains true after correcting for confounds like spatial (3D) distance. This discovery suggests that current cell types 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 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, and then we compared with results predicted by the *C. elegans* cell lineage tree (**Fig. 2**).

We discovered that more closely related neurons tended to have a greater overlap in their synaptic partners. This effect was most pronounced 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).

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).

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." (8, 19, 20) 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, refuting the core claim of Peters' Rule (**Fig. 4**). When we regress

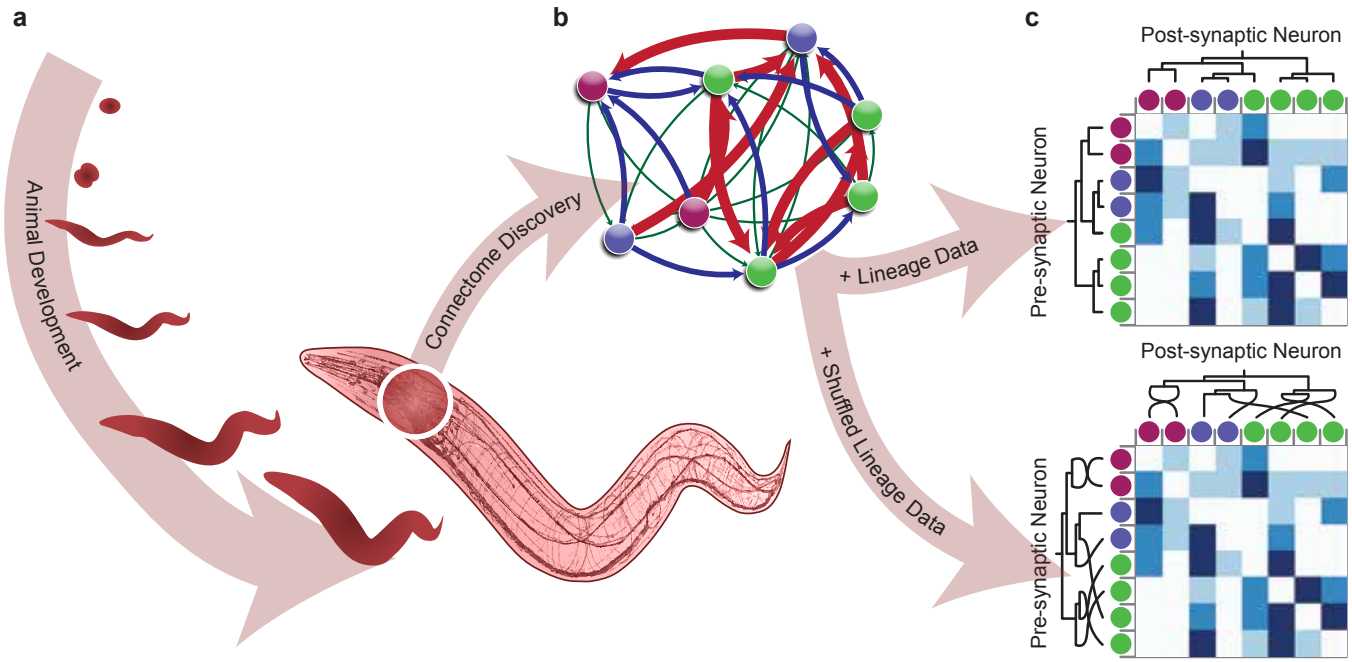

Fig. 1. Closely related cells have more synaptic partners in common than predicted by cell type alone. a. As an animal develops and matures, progenitor cells split and divide into daughter cells. These cells eventually assume specific cell roles in the mature organism. As neural cells develop, they also form synaptic connections with local and distributed partners. **b.** 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.** By comparing this graph — 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.

physical distance between neurons out, the observed effects remain. All analyses below 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.

A null model for lineage trees eliminates the effect of lineage on neural connectivity. In order to contextualize these results and ensure that they were not simply some emergent property of densely connected networks, we produced a simple lineage-tree null model by randomly shuffling neuron 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 that our null model with shuffled cell-identity mapping eliminates the effects seen in these analyses.

Discussion

Here we have shown that cellular lineage can explain more of the variance of the connectome than our current definition of cell types alone. Cell lineage information was 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 underlying biological mechanisms (i.e., we propose that it is unlikely that target cells are producing local chemical signals which source cells search for.)

There are several confounds to consider, not least that there is a great deal of inter-individual variance in connectivity, even between isogenic individuals (16). In order to control for this, we inspected 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 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 (5, 15), 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 net-

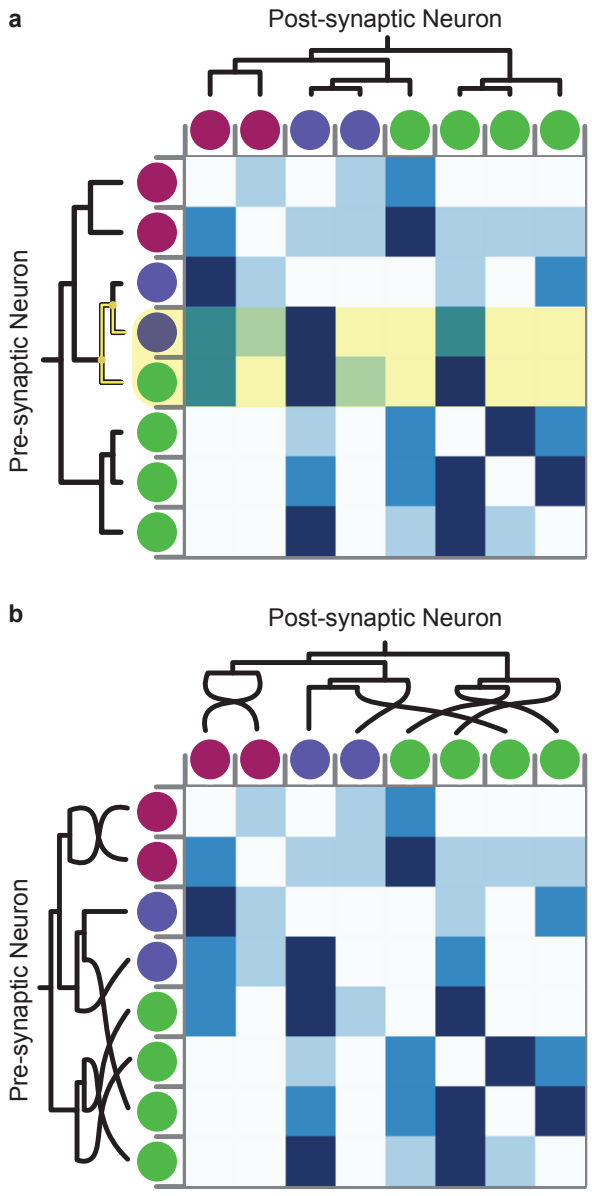

Fig. 2. A model of a connectome adjacency matrix with coregistered lineage information. In this model nervous system, there are eight neurons, with three cell types (red, blue, and green). The adjacency matrix of this nervous system is presented with blue squares where synapses are present: The shaded region at $(1, 2)$ indicates that the cell with ID 1 has a synapse (directed edge) on the cell with ID 2. On the left and top edges of the matrix, we overlay the tree of lineage information. The root of the tree is the zygote cell, 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.

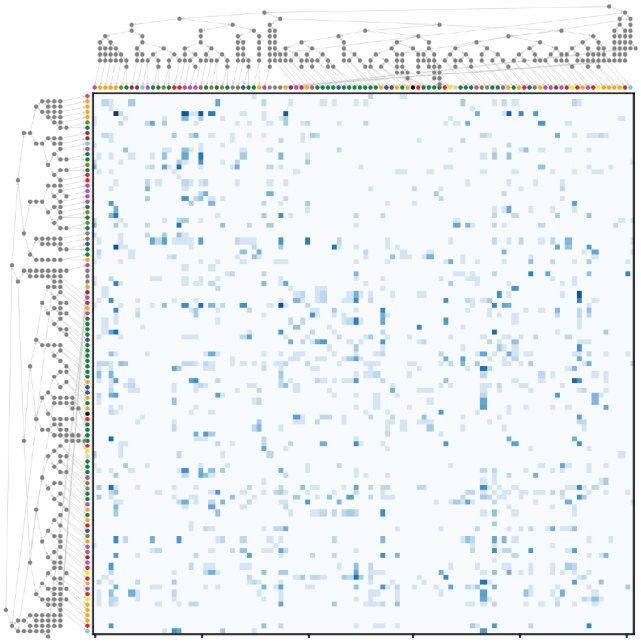

Fig. 3. The *C. elegans* connectome adjacency matrix with coregistered lineage information. Cell types are indicated with leaf-vertex coloring. Lineages are only shown for cells with annotated chemical synapses. Matrix ordering is based upon a topological sort of the cell lineage.

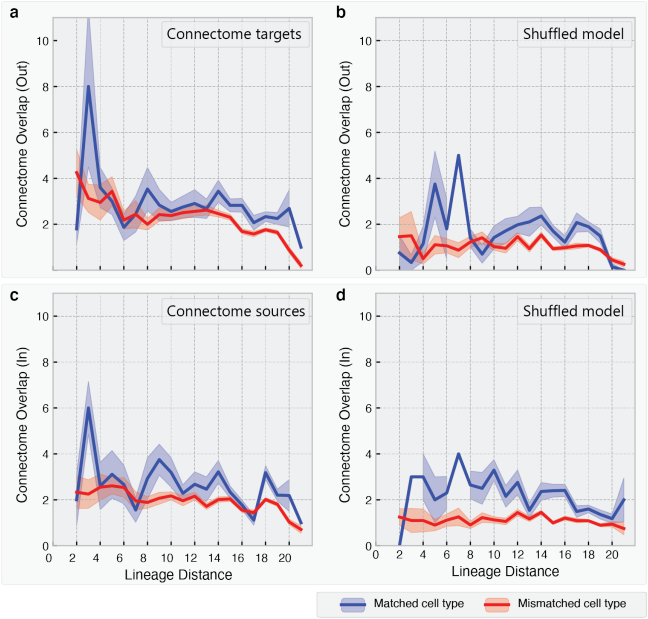

Fig. 4. The relationship between lineage distance and neighbor set overlap. Neurons that are more closely related to each other tend to have more downstream-synapse partners in common. **a & b: Connectome similarity, measured by the size of the set-intersect of synaptic targets. a. The *C. elegans* connectome.** 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 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.** The effect size is lessened when predicting synapse sources rather than targets.

work structure than can be efficiently encoded in the limited information storage of the genome (21).

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*).

Methods

C. elegans nematode connectomes were downloaded from NemaNode.org (16, 18) and BossDB.org (22) 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. (17, 23, 24)

Cell identities were coregistered with cell lineage and location information from WormWeb.org and WormBase.org, indexing by cell name. (17, 25) Cell locations were recorded as the centroid of the soma of the 3D reference reconstruction from OpenWorm.org (code provided in *Supplemental Materials*). (25, 26) 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. (27, 28) 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 of the electron microscopy connectome, and Nikhil Bhatla for the collection of the *C. elegans* lineage tree data.

Bibliography

1. Mathias Uhlén, Linn Fagerberg, Björn M Hallström, Cecilia Lindskog, Per Oksvold, Adil Mardinoglu, Åsa Sivertsson, Caroline Kampf, Evelina Sjöstedt, Anna Asplund, et al. Tissue-based map of the human proteome. *Science*, 347(6220), 2015.
2. Joshua P Neunuebel and James J Knierim. Spatial firing correlates of physiologically distinct cell types of the rat dentate gyrus. *Journal of Neuroscience*, 32(11):3848–3858, 2012.
3. Stewart HC Hendry and David J Calkins. Neuronal chemistry and functional organization in the primate visual system. *Trends in neurosciences*, 21(8):344–349, 1998.
4. Hongkui Zeng and Joshua R Sanes. Neuronal cell-type classification: challenges, opportunities and the path forward. *Nature Reviews Neuroscience*, 18(9):530–546, 2017.
5. Carol L Thompson, Lydia Ng, Vilas Menon, Salvador Martinez, Chang-Kyu Lee, Katie Glattfelder, Susan M Sunkin, Alex Henry, Christopher Lau, Chinh Dang, et al. A high-resolution spatiotemporal atlas of gene expression of the developing mouse brain. *Neuron*, 83(2): 309–323, 2014.
6. Csaba Erö, Marc-Oliver Gewaltig, Daniel Keller, and Henry Markram. A cell atlas for the mouse brain. *Frontiers in neuroinformatics*, 12:84, 2018.
7. Bosiljka Tasic, Zizhen Yao, Lucas T Graybuck, Kimberly A Smith, Thuc Nghi Nguyen, Darren Bertagnolli, Jeff Goldy, Emma Garren, Michael N Economo, Sarada Viswanathan, et al. Shared and distinct transcriptomic cell types across neocortical areas. *Nature*, 563(7729): 72–78, 2018.
8. Christopher L Rees, Keivan Moradi, and Giorgio A Ascoli. Weighing the evidence in peters' rule: does neuronal morphology predict connectivity? *Trends in neurosciences*, 40(2): 63–71, 2017.
9. Arthur W Toga and Paul M Thompson. Connectopathy in ageing and dementia. *Brain*, 137 (12):3104–3106, 2014.
10. Till Marquardt, Ryuichi Shirasaki, Sourav Ghosh, Shane E. Andrews, Nigel Carter, Tony Hunter, and Samuel L. Pfaff. 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. ISSN 0092-8674. doi: 10.1016/j.cell.2005.01.020.
11. Dániel L Barabási and Dániel Czégel. Constructing graphs from genetic encodings. *Scientific Reports*, 11(1):1–13, 2021.
12. Matthew F Glasser, Stephen M Smith, Daniel S Marcus, Jesper LR Andersson, Edward J Auerbach, Timothy EJ Behrens, Timothy S Coalson, Michael P Harms, Mark Jenkinson, Steen Moeller, et al. The human connectome project's neuroimaging approach. *Nature neuroscience*, 19(9):1175–1187, 2016.
13. Susie Y Huang, Thomas Witzel, Boris Keil, Alina Scholz, Mathias Davids, Peter Dietz, Elmar Rummert, Rebecca Ramb, John E Kirsch, Anastasia Yendiki, 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.
14. Alexander Shapson-Coe, Michał Januszewski, Daniel R Berger, Art Pope, Yuelong Wu, Tim Blakely, Richard L Schalek, Peter Li, Shuohong Wang, Jeremy Maitin-Shepard, et al. A connectomic study of a petascale fragment of human cerebral cortex. *bioRxiv*, 2021.
15. Stan Kerstjens, Gabriela Michel, and Rodney J. Douglas. 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.
16. Daniel Witvliet, Ben Mulcahy, James K Mitchell, Yaron Meirovitch, Daniel R Berger, Yuelong Wu, Yufang Liu, Wan Xian Koh, Rajeev Parvathala, Douglas Holmyard, et al. Connectomes across development reveal principles of brain maturation. *BioRxiv*, pages 2020–04, 2021.
17. Nikhil Bhatla. C. elegans Interactive Cell Lineage, 2011.
18. Steven J Cook, Travis A Jarrell, Christopher A Brittin, Yi Wang, Adam E Bloniarz, Maksim A Yakovlev, Ken CQ Nguyen, Leo T-H Tang, Emily A Bayer, Janet S Duerr, et al. Whole-animal connectomes of both caenorhabditis elegans sexes. *Nature*, 571(7763):63–71, 2019.
19. Alan Peters and Martin L Feldman. The projection of the lateral geniculate nucleus to area 17 of the rat cerebral cortex. *Journal of neurocytology*, 5(1):63–84, 1976.
20. Valentino Braitenberg and Almut Schüz. *Cortex: statistics and geometry of neuronal connectivity*. Springer Science & Business Media, 2013.
21. Alexei Koulakov, Sergey Shuvaev, and Anthony Zador. Encoding innate ability through a genomic bottleneck. Mar 2021. doi: 10.1101/2021.03.16.435261.
22. Robert Hider, Dean M Kleissas, Derek Pryor, Timothy Gion, Luis Rodriguez, Jordan Matelsky, William Gray-Roncal, and Brock Wester. The block object storage service (bossdb): A cloud-native approach for petascale neuroscience discovery. *bioRxiv*, page 217745, 2019.
23. Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.
24. Jordan K. Matelsky, Elizabeth P. Reilly, Erik C. Johnson, Jennifer Stiso, Danielle S. Bassett, Brock A. Wester, and William Gray-Roncal. DotMotif: an open-source tool for connectome subgraph isomorphism search and graph queries. *Scientific Reports*, 11(1), Jun 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-91025-5.
25. Todd W Harris, Igor Antoshechkin, Tamberlyn Bieri, Darin Blasiar, Juancarlos Chan, Wen J Chen, Norie De La Cruz, Paul Davis, Margaret Duesbury, Ruihua Fang, et al. Wormbase: a comprehensive resource for nematode research. *Nucleic acids research*, 38(suppl_1): D463–D467, 2010.
26. Balázs Szigeti, Padraig Gleeson, Michael Vella, Sergey Khayrulin, Andrey Palyanov, Jim Hokanson, Michael Currie, Matteo Cantarelli, Giovanni Idili, and Stephen Larson. Openworm: an open-science approach to modeling caenorhabditis elegans. *Frontiers in computational neuroscience*, 8:137, 2014.
27. Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.
28. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.