Neural Cartography: Methods for Mapping the Structure of Neural Networks

Leila Elabbady

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2025

Reading Committee:

John Tuthill, Co-Chair
Forrest Collman, Co-Chair
Osama Ahmed

Program Authorized to Offer Degree:
Neuroscience

University of Washington

Abstract

Neural Cartography: Methods for Mapping the Structure of Neural Networks

Leila Elabbady

Chair of the Supervisory Committee:

John Tuthill & Forrest Collman

Department of Neurobiology and Biophysics

Recent advancements in volumetric electron microscopy (vEM) allow neuroscientists to map the nervous system at unprecedented scale and resolution. This offers an opportunity to unravel structural organization, cellular morphology, and connectivity, providing new insights into circuit function. This dissertation presents new tools for scalable analytical solutions and circuit analyses to extract testable biological insights from vEM wiring diagrams.

In the first section, I utilize cell-body morphology and connectivity to train a hierarchical model for cell-type prediction in a cubic millimeter vEM dataset of mouse visual cortex. This method bypasses the need for complete cell reconstructions, is adaptable to multiple cell-typing schemes, and is computationally inexpensive. Further, this method produced predictions for nearly 100,000 cells with 91% accuracy compared to expert-labeled ground truth. These predictions,

publicly available with a new feature set, can be used for unsupervised search of rare cell-types and demonstrated the surprising sufficiency of the somatic region for cell identification.

In the second section, I reconstructed the wiring diagram of tactile sensory neurons in the fly ventral nerve cord to elucidate how spatial information is organized within somatosensory circuits, specifically regarding spatially targeted leg grooming. Using genetic labeling and vEM, I defined the foreleg somatotopic map. Downstream connectivity revealed 60 interneurons receiving substantial synaptic input exclusively from tactile neurons. These interneurons exhibit unique axonal projections, diverse dendritic morphologies, and distinct postsynaptic partners. Optogenetic experiments and kinematic analyses demonstrated that activating distinct interneurons initiates spatially guided grooming strategies consistent with our structurally derived receptive field predictions. From these results I propose a four-layer circuit where interneurons form distinct functional modules, each sampling a portion of the leg to elicit spatially targeted grooming.

The third section extends this spatial mapping analysis to chemosensory information from the fly leg, using a similar approach to examine how a distinct circuit structure might also maintain spatial information for a different sensory system within the same body segment.

Overall, this dissertation examines neural network maps broadly, quantifying structural variability in cell bodies and connectivity across large populations. It then zooms in on specific sensory circuits, exploring how their structural organization informs our understanding of neural computations.

Introduction

We think of the nervous system as the network of cells that allows us to respond to the world around us, facilitates thought and reflection, stores our memories, and regulates our personalities. Simply put, we think of our brains as the organs that make us who we are. The loss of a limb, removal of an appendix, or even a heart transplant are not considered a 'change of self'. However, an illness to the brain can cause one to act erratically, alter one's personality, or forget oneself entirely. Naturally one might wonder, how much of our character and individuality is coded within the structure of the brain?

For centuries, neuroscientists have been building maps of the brain to better understand this enigmatic network of cells. After a tragic railroad accident in 1848, Phineas Gage was left with severe trauma to his prefrontal cortex and an entirely new personality.[1,2] Today, scientists view him as one of the first recorded examples that linked specific brain areas to higher order function. It was not until the mid 20th century that Wilder Penfield published a functional map of the entire human brain.[3] Through a series of cortical stimulation experiments in awake patients, Penfield divided the brain into an array of functional areas that included motor, sensory, and speech areas we still refer to today.[3–5] With these boundaries in place, scientists began mapping the organization of cells within areas. By systematically stimulating different areas within the motor and sensory cortices, Penfield demarcated groups of cells that corresponded to different body parts. This work among others defined the body maps referred to today as the homunculus.[3,6–9] Since then, scientists have established maps all over the brain including retinotopic maps in the visual system,[10,11] tonotopic maps in the auditory system,[12–14] and cognitive maps in the hippocampus.[15–17] Each map provides a new lens with which to understand how the structure of the brain leads to

function. The following dissertation is a continuation of this effort. With the advent of large-scale volumetric electron microscopy, scientists are building wiring diagrams of the brain that map all the connections between cells, referred to as connectomes. I begin this dissertation with an introduction to why and how these connectome maps are useful. Chapters 1 through 3 describe three distinct efforts to understand the function of the nervous system through the development and analysis of connectomes. Finally, I conclude with a discussion chapter on some of the promising directions, groundbreaking advancements, and exciting applications of connectomics.

What's in a map?

Maps are commonly used as tools to facilitate movement from point A to point B, yet they are also rich with insights into how a given place operates. From the size and density of streets alone, one gets a sense of how people travel through any given metropolis. Add in labels to the map and one can begin to discern the function of different neighborhoods such as residential versus business districts. Streets filled with restaurants suggest areas with higher evening traffic, whereas universities and schools are probably busier during the day. These types of observations about the structural organization of a space not only provide an overall roadmap but also provide insight into how a city functions and how people move and live within the structural bounds of their city.

This type of framework is not limited to the study of cities and geographies and has proven to be incredibly valuable to the study of the nervous system. The nervous system is a vast and complex network made up of millions of cells that communicate through billions of connections and multiple modalities. The roadmap of the nervous system is made up of thousands of miles of cables, referred to as axons, that connect one neuron to the next through synapses. In fact, one cubic millimeter of the mouse cortex, roughly the size of a grain of sand, contains 200,000 cells, four kilometers of axon, and 523 million synapses.[18] Unraveling the complexity of these dense

systems and defining the structural organization of neural networks will be immensely valuable in our pursuit of understanding the nervous system.

Building a roadmap of the nervous system will orient our understanding around how information moves throughout the network, the relative diversity of cell-types within a system, and how different connectivity motifs facilitate certain computations over others. Precise network architectures will accelerate discovery by providing scientists with detailed hypotheses on the functional role of neuronal circuits and how they impact the behavior and overall well-being of different organisms. Several neurodegenerative diseases, such as Parkinson's and Amyotrophic Lateral Sclerosis, are defined by the loss of a particular cell-type in the nervous system.[19–24] Understanding the connectivity patterns of these cells and the specific circuits they form will direct researchers towards therapeutic avenues to recover those missing connections. Furthermore, architectures, connectivity patterns, and learning rules derived from this type of data may inspire the next generation of artificial neural networks and pioneer a new field of biomimetic engineering.

Building the map

To map the structural organization of the nervous system, scientists need to reconstruct the wiring diagram at a high enough resolution to make out individual synapses and identify cell-types, whilst also covering large enough spans of tissue to encompass entire circuits. With the recent advancements in volumetric electron microscopy (vEM), neuroscientists can now map the nervous system at a scale and resolution far greater than ever before. Large-scale vEM provides an opportunity to unravel the structural organization of entire neural circuits and study the morphology and connectivity of cells at a groundbreaking scale. This broad mapping of synaptic connections using vEM techniques is referred to as 'connectomics'. The first connectome was collected in C. elegans in 1986.[25] Decades later, bolstered by advancements in tissue staining and

sectioning,[26–34] high-throughput imaging,[28,29,34,35] gpu capabilities, and deep learning backed image segmentation,[36–38] we are witnessing a new wave of connectomics. Many vEM datasets have been collected from different species and different areas of the nervous system (Fig. 1). In the past few years alone, notable connectomes include an entire fly brain,[39,40] two fly ventral nerve cords,[33,41–44] a cubic millimeter of mouse visual cortex[18] and a comparable sample from human temporal cortex,[45] with many more on the horizon.

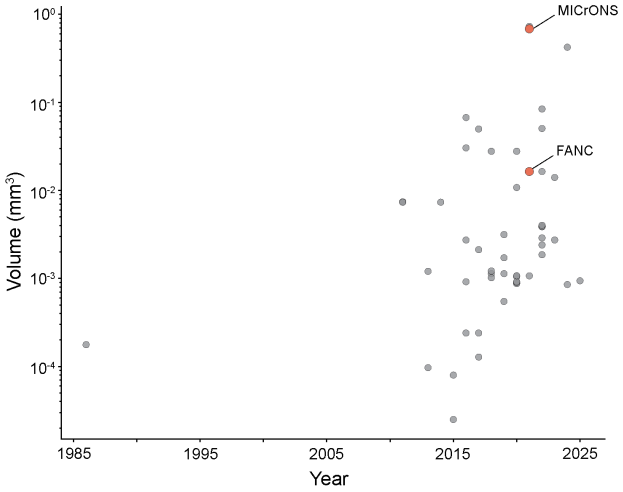


Figure 1. Growth of published connectome datasets in the last 40 years. This dissertation delves into two datasets; the millimeter cubed dataset from the mouse visual cortex (MICrONS), and the Drosophila female adult nerve cord (FANC), (red points).

As datasets are released to the public, the call continues for more: more samples, more species, and more data. While the opportunities and analyses from these ever-growing vEM datasets seem endless, there are still several significant hurdles that limit both the impact and the efficacy of the connectome revolution among the wider scientific community. Building connectome maps is still a long and arduous process with many fickle steps that involve lots of people and can be financially burdensome if not prohibitive. Briefly, the process of building these circuit maps involves: tissue processing, tissue sectioning/milling, imaging, alignment, segmentation, and analysis. While we have made meaningful improvements at each step, efforts to optimize workflows, cut costs, facilitate scientific collaboration, and develop frameworks for analysis are necessary for this field to reach its fullest potential.

One of the largest bottlenecks in the pipeline today is the manual correction of the automated segmentation, a process referred to as proofreading. Proofreading today is a largely manual process where expert anatomists follow the neurites of individual cells, adding any processes that were missed and removing extraneous processes that had been falsely merged by the initial segmentation. Further, some of the most common circuit analyses are dependent on a level of cell identification. Many of our current methods for morphologies cell-identification in the nervous system rely on near complete reconstructions and thus require proofread cells.[46–49] Tools and methods that could either accelerate proofreading or be impervious to reconstruction completeness would greatly advance the accessibility of these datasets.

To address this hurdle, the first chapter of this dissertation will present a novel framework for cell identification that bypasses the need for complete reconstructions, is adaptable to multiple cell-typing schemas, and is computational inexpensive to run across petabyte-scale vEM datasets. In this chapter I propose an expansion of our current understanding of cell-types by presenting that the morphology and connectivity at the cell-body alone contains more cell-type specific information than previously appreciated. I emphasize that this is not an assertion for a specific definition of cell-type rather a proposed framework for cell identification that can incorporate changes as our understanding of cell-types evolve. Ultimately, chapter one is best accompanied by a collection of works that utilize the presented method and highlight its applicability for scientific discovery.[18,50–52]

Utilizing the map

Just as people must move within the structural bounds of their environment, computations in the brain must occur within the structural bounds of neural circuits. In that respect, one of the burgeoning impacts of connectomics is the ability to untangle the wiring diagram of a given circuit

to form precise and specific functional predictions. The wiring diagram alone does not explain the function of a circuit. However, it does limit the search space of possibilities and steer the design and prioritization of experiments that can probe function more directly. The dense sampling from connectome datasets provides a more holistic lens into the diversity of cell-types within a given circuit, the relative abundance of these cell-types, and the connectivity patterns between them. Even with the incredible genetic toolkit available in fly neuroscience, it is reported that only 40% of cell-types had been described before the fly brain connectome.[39,53] As such, studies describing the diversity of cells, unique wiring rules, and persistent connectivity features provide a pivotal foundation for future findings to build off.[33,40,41,44,50–52,54,55]

On its own, a single connectome is a static wiring diagram that represents the circuitry from one animal at a single point in time. Scientists have brought these circuits to life through theoretical modeling and connectome constrained simulations to contextualize how the architecture of a circuit gives rise to specific behaviors. A leaky integrate and fire model constrained by connectome of the entire fruit fly brain successfully predicted the activation of different sensorimotor pathways in response to well-known odorants. Constrained by circuit structure alone, researchers used the model to predict an interaction between sugar and water pathways, which they confirmed with calcium and behavioral experiments.[56] Other studies layered biophysical processes and other dynamic properties on top of connectome constrained circuits to probe how neural activity gives rise to behaviors such as locomotion in C. elegans.[57,58] Furthermore, researchers have developed connectome inspired deep networks that recapitulated optic flow in the fruit fly. These models not only reproduced the functional profile of some well-studied cell types but also provided verifiable predictions for the role for the remaining cell-types within the circuit.[59] These examples highlight how the intersection of connectomics and theoretical

modeling can accelerate the discovery of circuit mechanisms underlying specific neural computations and behaviors.

Alongside the theoretical efforts, experimentalists utilized the connectome to make direct connections between the structure of specific circuits and their functional role in cognitive or behavioral tasks. The benefit of these comprehensive wiring diagrams is that they minimize the theoretical search space, guide hypotheses, and shape experimental designs. The study of fly navigation in the central complex has become a quintessential example of this type of research. The first comprehensive connectome of the central complex[60,61] enabled researchers to model and interpret decades of experimental studies on how neurons track the fly's angular orientation and maintain an internal compass,[62,63] mechanisms for goal-directed navigation,[64] and multisensory integration.[61] Recent work comparing the central complex across insects has begun to tease apart which features of this circuit are conserved across species and how the circuit may have evolved to support diverging navigation strategies.[65–67]

Leveraging the genetic tools and connectome datasets available in fly neuroscience, Chapter 2 explores the underlying mechanisms for spatially targeted behavior in response to tactile stimulation. The study reveals how the adult fly leg's tactile circuit yields specific, testable predictions about spatial information transmission. After experimentally validating these predictions, I present a four-layer circuit explaining how the tactile circuit enables spatially targeted behaviors. In chapter 3, I delve into the chemosensory circuit of the adult fly leg and present an alternative circuit structure that maintains spatial information. Despite the stark structural differences between the tactile and chemosensory circuits from the leg, I propose a framework for how spatial information may still be maintained within the chemosensory system and the ethological relevance for such a mapping.

References

1. Harlow, J. M. Recovery from the passage of an iron bar through the head. *Hist. Psychiatry* **4**, 274–281 (1993).

2. Harlow, J. M. Passage of an Iron Rod Through the Head. *J. Neuropsychiatry Clin. Neurosci.* **11**, 281–283 (1999).

3. Penfield, W. & Jasper, H. *Epilepsy and the Functional Anatomy of the Human Brain*. xv, 896 (Little, Brown & Co., Oxford, England, 1954).

4. Penfield, W. & Boldrey, E. Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain* **60**, 389–443 (1937).

5. PENFIELD, W. & PEROT, P. THE BRAIN'S RECORD OF AUDITORY AND VISUAL EXPERIENCE1: A FINAL SUMMARY AND DISCUSSION. *Brain* **86**, 595–696 (1963).

6. Paul P. Broca. Loss of Speech, Chronic Softening, and Partial Destruction of the Anterior Left Lobe of the Brain. *Bull. Société Anthropol.* **2**, 235–238 (1861).

7. *Brodmann's Localisation in the Cerebral Cortex*. (Springer US, Boston, MA, 2005). doi:10.1007/b138298.

8. Fritsch, G. & Hitzig, E. On the electrical excitability of the cerebrum. (1870).

9. Penfield, W. & Rasmussen, T. *The Cerebral Cortex of Man; a Clinical Study of Localization of Function*. xv, 248 (Macmillan, Oxford, England, 1950).

10. Engel, S. A., Glover, G. H. & Wandell, B. A. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb. Cortex* **7**, 181–192 (1997).

11. Wandell, B. A. & Winawer, J. Imaging retinotopic maps in the human brain. *Vision Res.* **51**, 718–737 (2011).

12. Clopton, B. M., Winfield, J. A. & Flammino, F. J. Tonotopic organization: Review and analysis. *Brain Res.* **76**, 1–20 (1974).

13. Ehret, G. & Romand, R. *The Central Auditory System*. (Oxford University Press, 1997).

14. Saenz, M. & Langers, D. R. M. Tonotopic mapping of human auditory cortex. *Hear. Res.* **307**, 42–52 (2014).

15. O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map*. (Clarendon Press ; Oxford University Press, Oxford : New York, 1978).

16. Tolman, E. C. Cognitive maps in rats and men. *Psychol. Rev.* **55**, 189–208 (1948).

17. Wikenheiser, A. M. & Schoenbaum, G. Over the river, through the woods: cognitive maps in the hippocampus and orbitofrontal cortex. *Nat. Rev. Neurosci.* **17**, 513–523 (2016).

18. Consortium, Mic. *et al.* Functional connectomics spanning multiple areas of mouse visual cortex. 2021.07.28.454025 Preprint at https://doi.org/10.1101/2021.07.28.454025 (2021).

19. Sreedharan, J. & Brown Jr, R. H. Amyotrophic lateral sclerosis: Problems and prospects. *Ann. Neurol.* **74**, 309–316 (2013).

20. Robberecht, W. & Philips, T. The changing scene of amyotrophic lateral sclerosis. *Nat. Rev. Neurosci.* **14**, 248–264 (2013).

21. Bruijn, L. I., Miller, T. M. & Cleveland, D. W. UNRAVELING THE MECHANISMS INVOLVED IN MOTOR NEURON DEGENERATION IN ALS. *Annu. Rev. Neurosci.* **27**, 723–749 (2004).

22. Hadzipasic, M. *et al.* Selective degeneration of a physiological subtype of spinal motor neuron in mice with SOD1-linked ALS. *Proc. Natl. Acad. Sci.* **111**, 16883–16888 (2014).

23. Hirsch, E. C., Jenner, P. & Przedborski, S. Pathogenesis of Parkinson's disease. *Mov. Disord.* **28**, 24–30 (2013).

24. Dauer, W. & Przedborski, S. Parkinson's Disease: Mechanisms and Models. *Neuron* **39**, 889–909 (2003).

25. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode Caenorhabditis elegans. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **314**, 1–340 (1986).

26. Song, K., Feng, Z. & Helmstaedter, M. High-contrast en bloc staining of mouse whole-brain and human brain samples for EM-based connectomics. *Nat. Methods* **20**, 836–840 (2023).

27. Hua, Y., Laserstein, P. & Helmstaedter, M. Large-volume en-bloc staining for electron microscopy-based connectomics. *Nat. Commun.* **6**, 7923 (2015).

28. Hayworth, K. J. *et al.* Imaging ATUM ultrathin section libraries with WaferMapper: a multi-scale approach to EM reconstruction of neural circuits. *Front. Neural Circuits* **8**, (2014).

29. Denk, W. & Horstmann, H. Serial Block-Face Scanning Electron Microscopy to Reconstruct Three-Dimensional Tissue Nanostructure. *PLOS Biol.* **2**, e329 (2004).

30. Kasthuri, N. *et al.* Saturated Reconstruction of a Volume of Neocortex. *Cell* **162**, 648–661 (2015).

31. Knott, G., Rosset, S. & Cantoni, M. Focussed Ion Beam Milling and Scanning Electron Microscopy of Brain Tissue. *J. Vis. Exp. JoVE* 2588 (2011) doi:10.3791/2588.

32. Xu, C. S. *et al.* Enhanced FIB-SEM systems for large-volume 3D imaging. *eLife* **6**, e25916 (2017).

33. Phelps, J. S. *et al.* Reconstruction of motor control circuits in adult Drosophila using automated transmission electron microscopy. *Cell* **184**, 759-774.e18 (2021).

34. Yin, W. *et al.* A petascale automated imaging pipeline for mapping neuronal circuits with high-throughput transmission electron microscopy. *Nat. Commun.* **11**, 4949 (2020).

35. Mahalingam, G. *et al.* A scalable and modular automated pipeline for stitching of large electron microscopy datasets. *eLife* **11**, e76534 (2022).

36. Wu, J., Silversmith, W. M., Lee, K. & Seung, H. S. Chunkflow: hybrid cloud processing of large 3D images by convolutional nets. *Nat. Methods* **18**, 328–330 (2021).

37. Januszewski, M. *et al.* High-precision automated reconstruction of neurons with flood-filling networks. *Nat. Methods* **15**, 605–610 (2018).

38. Popovych, S. *et al.* Petascale pipeline for precise alignment of images from serial section electron microscopy. *Nat. Commun.* **15**, 289 (2024).

39. Dorkenwald, S. *et al.* Neuronal wiring diagram of an adult brain. *Nature* **634**, 124–138 (2024).

40. Schlegel, P. *et al.* Whole-brain annotation and multi-connectome cell typing quantifies circuit stereotypy in Drosophila. 2023.06.27.546055 Preprint at https://doi.org/10.1101/2023.06.27.546055 (2023).

41. Azevedo, A. *et al.* Connectomic reconstruction of a female Drosophila ventral nerve cord. *Nature* **631**, 360–368 (2024).

42. Lesser, E., Moussa, A. & Tuthill, J. C. Peripheral anatomy and central connectivity of proprioceptive sensory neurons in the Drosophila wing. 2025.05.29.656810 Preprint at https://doi.org/10.1101/2025.05.29.656810 (2025).

43. Takemura, S. *et al.* A Connectome of the Male Drosophila Ventral Nerve Cord. *eLife* **13**, (2024).

44. Marin, E. C. *et al.* Systematic annotation of a complete adult male Drosophila nerve cord connectome reveals principles of functional organisation. *eLife* **13**, (2024).

45. Shapson-Coe, A. *et al.* A connectomic study of a petascale fragment of human cerebral cortex. 2021.05.29.446289 Preprint at https://doi.org/10.1101/2021.05.29.446289 (2021).

46. Gouwens, N. W. *et al.* Classification of electrophysiological and morphological neuron types in the mouse visual cortex. *Nat. Neurosci.* **22**, 1182–1195 (2019).

47. Scorcioni, R., Polavaram, S. & Ascoli, G. A. L-Measure: a web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies. *Nat. Protoc.* **3**, 866–876 (2008).

48. Kanari, L. *et al.* Objective Morphological Classification of Neocortical Pyramidal Cells. *Cereb. Cortex* **29**, 1719–1735 (2019).

49. Costa, M., Manton, J. D., Ostrovsky, A. D., Prohaska, S. & Jefferis, G. S. X. E. NBLAST: Rapid, Sensitive Comparison of Neuronal Structure and Construction of Neuron Family Databases. *Neuron* **91**, 293–311 (2016).

50. Schneider-Mizell, C. M. *et al.* Inhibitory specificity from a connectomic census of mouse visual cortex. *Nature* **640**, 448–458 (2025).

51. Gamlin, C. R. *et al.* Connectomics of predicted Sst transcriptomic types in mouse visual

cortex. *Nature* **640**, 497–505 (2025).

52. Bodor, A. L. *et al.* The Synaptic Architecture of Layer 5 Thick Tufted Excitatory Neurons in the Visual Cortex of Mice. 2023.10.18.562531 Preprint at https://doi.org/10.1101/2023.10.18.562531 (2023).

53. Schlegel, P. *et al.* Whole-brain annotation and multi-connectome cell typing of Drosophila. *Nature* **634**, 139–152 (2024).

54. Winding, M. *et al.* The connectome of an insect brain. *Science* **379**, eadd9330 (2023).

55. Verasztó, C. *et al.* Whole-body connectome of a segmented annelid larva. *eLife* **13**, (2024).

56. Shiu, P. K. *et al.* A Drosophila computational brain model reveals sensorimotor processing. *Nature* **634**, 210–219 (2024).

57. Shlizerman, E. Driving the connectome by-wire: Comment on “What would a synthetic connectome look like?” by Ithai Rabinowitch. *Phys. Life Rev.* **33**, 25–27 (2020).

58. Kim, J., Florman, J. T., Santos, J. A., Alkema, M. J. & Shlizerman, E. Modular integration of neural connectomics, dynamics and biomechanics for identification of behavioral sensorimotor pathways in Caenorhabditis elegans. 724328 Preprint at https://doi.org/10.1101/724328 (2025).

59. Lappalainen, J. K. *et al.* Connectome-constrained networks predict neural activity across the fly visual system. *Nature* **634**, 1132–1140 (2024).

60. Zheng, Z. *et al.* A Complete Electron Microscopy Volume of the Brain of Adult Drosophila melanogaster. *Cell* **174**, 730-743.e22 (2018).

61. Hulse, B. K. *et al.* A connectome of the Drosophila central complex reveals network motifs suitable for flexible navigation and context-dependent action selection. *eLife* **10**, e66039 (2021).

62. Fisher, Y. E., Lu, J., D’Alessandro, I. & Wilson, R. I. Sensorimotor experience remaps visual input to a heading-direction network. *Nature* **576**, 121–125 (2019).

63. Turner-Evans, D. B. *et al.* The Neuroanatomical Ultrastructure and Function of a Biological Ring Attractor. *Neuron* **108**, 145-163.e10 (2020).

64. Mussells Pires, P., Zhang, L., Parache, V., Abbott, L. F. & Maimon, G. Converting an allocentric goal into an egocentric steering signal. *Nature* **626**, 808–818 (2024).

65. Heinze, S. Variations on an ancient theme — the central complex across insects. *Curr. Opin. Behav. Sci.* **57**, 101390 (2024).

66. Honkanen, A., Adden, A., da Silva Freitas, J. & Heinze, S. The insect central complex and the neural basis of navigational strategies. *J. Exp. Biol.* **222**, jeb188854 (2019).

67. von Hadeln, J. *et al.* Neuroarchitecture of the central complex of the desert locust: Tangential neurons. *J. Comp. Neurol.* **528**, 906–934 (2020).

Perisomatic Ultrastructure Efficiently Classifies Cells in Mouse Cortex

Leila Elabbady[1,2], Sharmishtaa Seshamani[1], Shang Mu[,3], Gayathri Mahalingam[1], Casey Schneider-Mizell[1], Agnes L Bodor[1], J. Alexander Bae[,3], Derrick Brittain[1], JoAnn Buchanan[1], Daniel J. Bumbarger[1], Manuel A. Castro[3], Sven Dorkenwald[1,3], Akhilesh Halageri[3], Zhen Jia[3], Chris Jordan[3], Dan Kapner[1], Nico Kemnitz[3], Sam Kinn[1], Kisuk Lee[3], Kai Li[3], Ran Lu, Thomas Macrina[3], Eric Mitchell[3], Shanka Subhra Mondal[3], Barak Nehoran[3], Sergiy Popovych[3], William Silversmith[3], Marc Takeno[1], Russel Torres[1], Nicholas L Turner[3], William Wong[3], Jingpeng Wu[3], Wenjing Yin[1], Szi-chieh Yu[3], The MICrONS Consortium[1,3,4], H. Sebastian Seung[3], R. Clay Reid[1], Nuno Maçarico Da Costa[1], Forrest Collman[1]*

1 Allen Institute for Brain Science, Seattle, WA
2 University of Washington, Seattle, WA
3 Princeton Neuroscience Institute, Princeton University, NJ
4 Department of Neuroscience, Baylor College of Medicine, Houston, TX

Mammalian neocortex contains a highly diverse set of cell types. These types have been mapped systematically using a variety of molecular, electrophysiological and morphological approaches.[1–4] Each modality offers new perspectives on the variation of biological processes underlying cell type specialization. Cellular scale electron microscopy (EM) provides dense ultrastructural examination and an unbiased perspective into the subcellular organization of brain cells, including their synaptic connectivity and nanometer scale morphology. It also presents a clear challenge for analysis to identify cell-types in data that contains tens of thousands of neurons, most of which have incomplete reconstructions.[5] To address this challenge, we present the first systematic survey of the somatic region of all cells within a cubic millimeter of cortex using quantitative features obtained from EM. This analysis demonstrates a surprising sufficiency of the perisomatic region to identify cell-types, including types defined primarily based on their connectivity patterns. We then describe how this classification facilitates cell type specific connectivity characterization and locating cells with rare connectivity patterns in the dataset.

Foreword

The following manuscript was published in Nature on April 9, 2025 as part of the MICrONS Project. To access the online version visit: https://doi.org/10.1038/s41586-024-07765-7.

Introduction

Electron microscopy volumes provide a unique perspective on neural circuits by enabling dense tracing of individual axons, dendrites and synaptic connections. Recent progress in data acquisition and dense segmentation have dramatically grown the capability to acquire large-scale datasets.[5–11] The size of these volumes allow for large numbers of cells to be analyzed with reconstructions of entire dendrites and local axons of mammalian neurons. However, it raises the challenge of accurately classifying tens or hundreds of thousands of cells. Doing so is necessary for many basic investigations, from co-registering neurons, to studying specific cell populations (including neuronal and non-neuronal cells), or characterizing cell type specific connectivity at scale. Just as experimental systems require genetic tools to provide inexpensive access to rare cell populations that would otherwise be difficult to study with non-selective techniques, large-scale electron microscopy requires computational tools to provide inexpensive access to specific cell types to facilitate further analyses. Existing methods for automated cell-typing based on morphology or connectivity often necessitate near complete axonal or dendritic reconstructions.[2,12–14] Such reconstructions require manual correction to the segmentation, often referred to as proofreading, which is prohibitively time consuming at scale. This means classifying cells based on specific output connectivity profiles is difficult in the dataset. Moreover, after proofreading, a single neuron reconstruction contains thousands of synaptic targets to identify (Figure 1). A method that could identify cell-types in the dataset in a way that is insensitive to

changes in proofreading and truncation is therefore of high utility, both to automate the classification of targets and to help guide proofreading to cells that have connectivity patterns of interest.

Here we describe a fast, scalable and computationally inexpensive approach which can address these problems. We first analyzed the somatic region of nearly 100,000 cells in the MICrONs dataset,[5] a cellular compartment which contains morphological and connectivity based biological properties that, as we will present, differentiate cell-types. By analyzing only the somatic region of a cell, our analysis was robust to segmentation errors, unique per cell, and thus insensitive to most proofreading changes. We included well-established features known to differentiate cells, such as somatic size and cortical depth, as well as novel features whose cell-type distinctions are less well recognized such as nuclear folding and soma synapse density. We further developed an unsupervised approach to describe the fine scale morphology of the perisomatic region of inhibitory cells, and demonstrate that it varies across major inhibitory subclasses. With these features, we address the need for dataset wide cell-type labels outlined above, by training a hierarchical classifier to identify basic cell classes across the entire dataset. Lastly, we demonstrate the utility of perisomatic features to facilitate the targeted search for rare cell types across a dataset. This method is already being used to reveal fundamental aspects of cell type specific wiring of mammalian cortex.[5,15–17]

Results

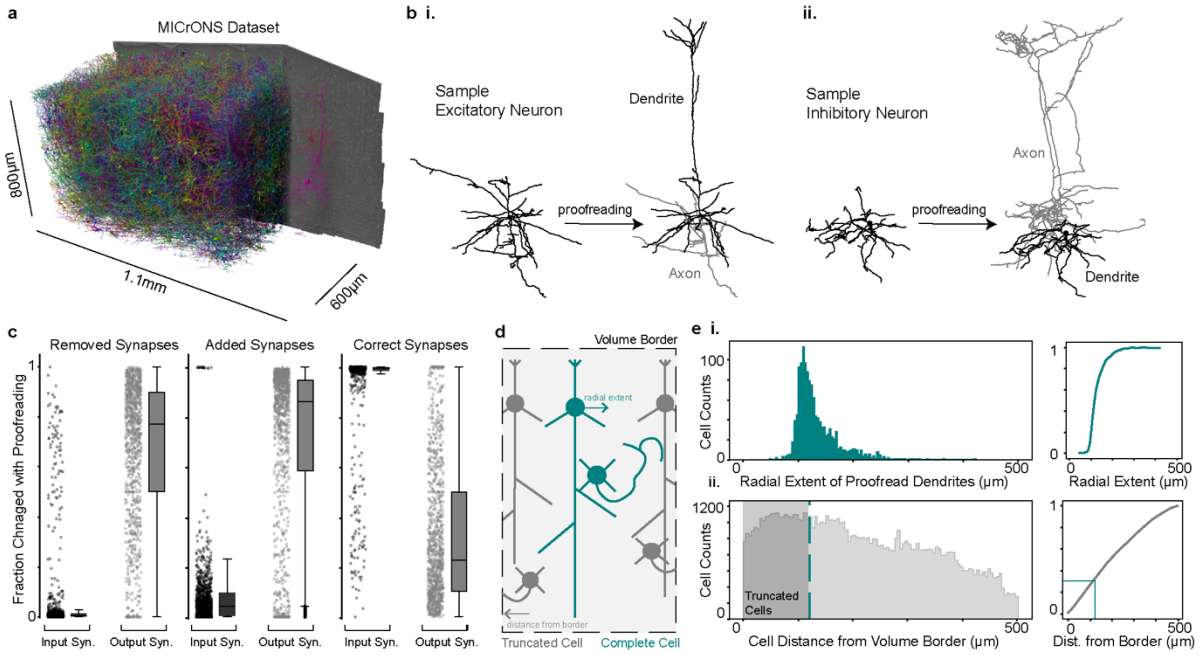


Figure 1: Large-scale automated segmentations necessitate proofreading insensitive cell classifications. a) Rendering of a small fraction of neurons from the MICrONS dataset (1.4mm x 800µm x 800µm) covering all layers of cortex and multiple visual areas, with 1,207 rendered and then cutaway to reveal the full morphology of 2 selected neurons on the right portion of the dataset. **b)** Example neuronal morphologies before and after proofreading, **i)** excitatory neuron and **ii)** inhibitory neuron. **c)** Fraction of input and output synapses removed (left), added (middle) and maintained (right) after proofreading for 1,347 neurons. For all box plots, center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; outliers not shown (visible in the adjacent scatter plots). **d)** Neurons near the volume borders will have truncated morphologies. **e) i)** Histogram of the radial extent of dendrites from a sample of 1,347 proofread neurons[15] (left) and the cumulative distribution of those cells (right). **ii)** Histogram of the minimum distance from a volume border for all high quality nuclear detections (n=94,010) (left) and the cumulative distribution of those distances (right). The portion of cells which are less than the median radial extent (33% of cells) is indicated with teal shading and teal lines.

Quality of Neuronal Arbor Segmentation

We analyzed the larger portion of the MICrONS dataset, a 1.4mm x 800µm x 800µm volumetric serial section EM dataset from mouse visual cortex,[5] that contains a dense segmentation of cells, a nucleus segmentation, and dataset wide synapse detection (Fig. 1a).[7,18] This dataset includes 94,010 high quality nuclear detections enclosed within the boundaries of the volume (see methods) and spans cortical Layer 1 to the white matter. For most cells, high quality cellular

segmentation requires proofreading to clean and complete the reconstructions, particularly axons (Fig. 1b-c). Most false mergers are of axonal fragments, leading most outputs of unproofread axons to be incorrect (Fig. 1c). When axonal proofreading is invested in a cell, it creates an elaborate object to analyze with thousands of postsynaptic targets. In order to analyze the cell-type specific connectivity pattern of that single reconstructed cell (examples in Fig. 5-6), each of those postsynaptic cells requires a cell type label. Dendrites on the other hand are quite precise, as 75-95% of the 1,000 to 15,000 synapses detected on reconstructed axons can be mapped to their soma in the MICrONs dataset with more than 99% accuracy (Fig. 1c). However, many of these targets have incomplete reconstructions themselves because even for mm^3 scale volumes about a third the cells are close enough to the edge to have their dendrites truncated (Fig. 1d-e). This level of truncation across cells, whether due to segmentation errors or proximity to the volume border, led us to investigate alternative methods for cell-typing that would be insensitive to a cell's dendritic and axonal reconstruction status.

Perisomatic Features Across Cortical Cells

The somatic region of the cell is an attractive location for such a method because the automated reconstruction of the somatic region is typically precise and complete, with few changes during proofreading, if any (Fig. 2a). Moreover, the soma also has unique biological processes occurring within it, which led us to investigate if information within the perisomatic region could enable cell classification.

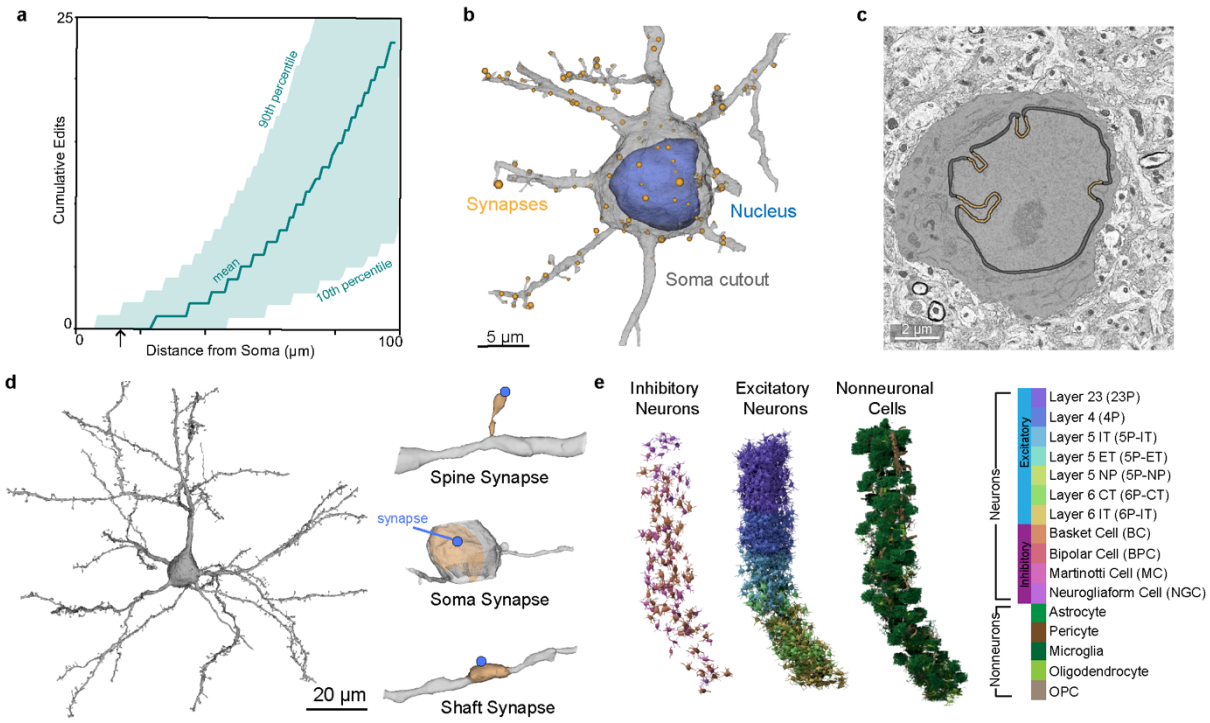


Figure 2: Perisomatic region of cortical cells. a) A measure of the distance from the soma for each edit that was made to the segmentation during proofreading of 1,347 cells. Average noted by the teal line, area chart notes the 10th and 90th percentile across all cells. Arrow notes 15 microns. **b)** Example cell demonstrating the extent of mesh information used to extract somatic, nuclear and synapse features. All cell meshes were restricted to 15 microns from the center of the nucleus. **c)** Representative example of nuclear infolding in a single electron microscopy image. The soma is highlighted in gray, black represents the nuclear envelope and orange are the areas classified as infolded based on the shrink wrap method (see Methods). **d)** Example cell demonstrating the extent of mesh information used to extract postsynaptic features (left) and three example postsynaptic shapes (PSS) (right). All synapses included in the PSS analyses were within 60 microns from the center of the nucleus. **e)** A 3D rendering of the somatic cutouts from all the cells from a 100 micron column that was densely reconstructed for which manual labels were given. Cells rendered are organized by their cell class and colored by their cell subclass according to the color scheme displayed.

We extracted geometric properties of the nucleus and soma within 15 microns from the center of the nucleus (Fig. 2b). For nuclei, this included volume, surface area, and depth from the pial surface. The 3D nuclear segmentations provide a detailed view of an interesting feature of neuronal nuclei, their tendency to form infoldings of their membranes, sometimes also referred to as invaginations. We quantified the fraction of nucleus membrane area that was within an infolding (Fig. 2c) (see methods). Similar geometric properties were calculated for the somatic region including the total volume, surface area, the ratio of the nucleus volume to the soma volume, and

the distance from the centroid of the nucleus to the centroid of the soma. We also measured the number and surface density of synaptic inputs detected on the somatic region of the cell. Together these somatic and nucleus features represent a feature space that was extracted for most cells (75% of nuclei detections, see methods). For a subset of neurons, we also analyzed the nano-scale structure of the postsynaptic compartments, what we are terming a "post synaptic shape" (PSS) (see methods) within 60μm of the nucleus center (Fig. 2d).

We used a densely reconstructed and manually annotated column of 1,619 cells across all layers of primary visual cortex as the reference dataset for all subsequent analyses (Fig. 2e) (see methods).[5,15] This column included excitatory neurons (1,115), inhibitory neurons (143), and non-neurons (361) with expert annotated labels for cellular classes and neuronal subclasses (Excitatory: Layer 2/3, Layer 4, Layer 5 inter-telencephalic (IT), near-projecting (NP) and extra-telencephalic (ET), Layer 6 IT and cortico-thalamic (CT) and Inhibitory: Martinotti/non-Martinotti cell (MC), Basket cell (BC), Bi-polar cell (BPC) and neurogliaform cell (NGC), Non-neurons: astrocyte, oligodendrocyte precursor cell (OPC), oligodendrocyte, microglia, pericyte) (Fig. 2e). While we used the above cell-typing scheme throughout our analyses, it should be noted that our approach can incorporate alternative labels as cell-type definitions evolve.

To investigate the efficacy of different features in separating cells, we plotted the variability of individual features and trained classifiers to distinguish cells at different levels of granularity. Nucleus features alone were sufficient to separate neurons from non-neuronal cells as non-neuronal cells had smaller nuclei compared to neuronal cells (cross-validated classification accuracy 90%, Extended Data Table 1). Each non-neuronal cell class exhibited a distinct range and consistency in their nucleus volume (Fig. 3a.i) and thus a nucleus-only classifier was able to identify non-neuronal subclasses with a cross-validated accuracy of 94% (Extended Data Table 1).

Nucleus features of excitatory neurons recapitulated expected laminar organization, wherein the borders between layer 2/3 (L23), layer 4 (L4), layer 5 (L5), and layer 6 (L6) are all demarcated by shifts in the distribution and variation of nucleus volumes. (Fig. 3a.i). Inhibitory cells, on the other hand, had less striking laminar patterns, but had a wider variation of nucleus volumes, overlapping with excitatory cells, with the exception of the larger layer 5 excitatory neurons (Extended Data Fig. 1).

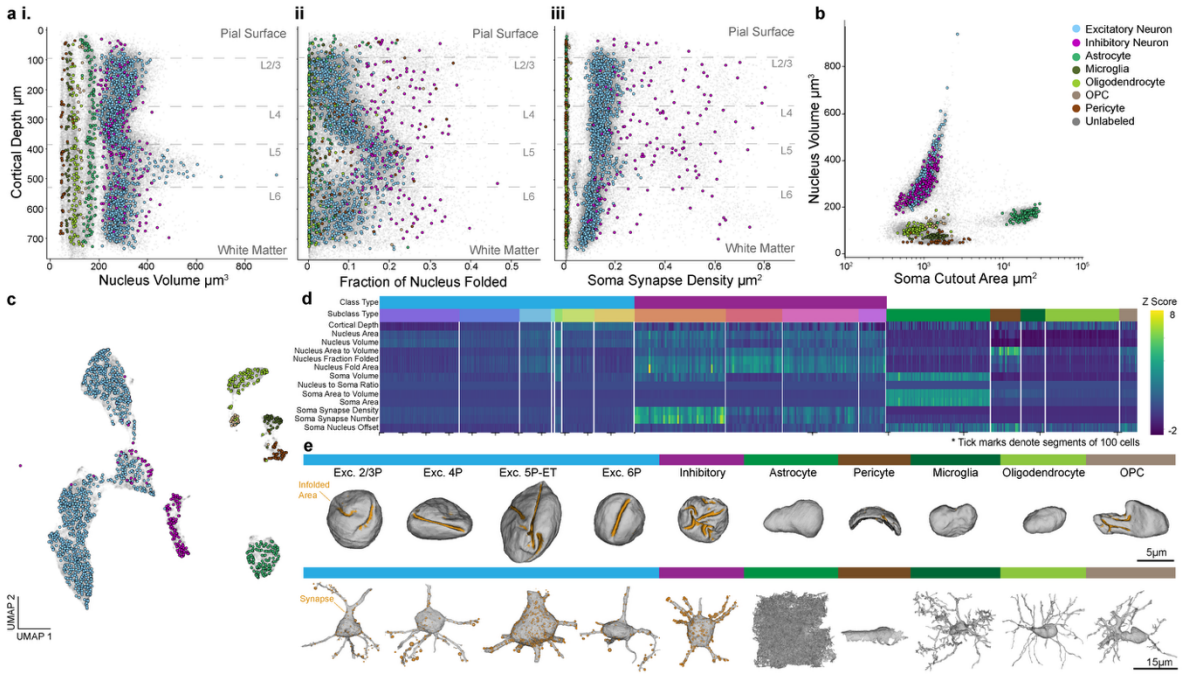


Figure 3: Variations of nucleus and somatic features show stark laminar and cell-class based distinctions. **a) i)** Nuclear volume µm^3 **ii)** fraction of nuclear membrane infolded and **iii)** somatic synapse density µm^2 plotted against distance from the pial surface in microns. Cortical layer boundaries are noted by the dashed lines. **b)** Somatic surface cutout area in µm^2 (within 15µm from the nuclear center) plotted against nuclear volume µm^3. **c)** 2D UMAP embedding of all neuronal and non-neuronal cells inferred from somatic features, nuclear features and cortical depth. **d)** Z-scored feature matrix representing all the somatic and nuclear features on the manually labeled cells from the cortical column. Cells are organized by their annotated subclass. Dashed marks along the x axis denote segments of 100 cells (1115 excitatory neurons, 143 inhibitory neurons, 361 non-neurons). For all plots, manually labeled cell classes are represented in color (1,619) and unlabeled examples in light gray (n=92,391). **e)** 3D mesh renderings of representative examples of different neuronal and non-neuronal cell classes. In the top row, nuclei displayed with the folded surface area highlighted in orange. Corresponding cell bodies displayed in the bottom row with somatic synapses in orange. Sphere size corresponds to predicted synapse size from the synapse detection model.[5]

The fraction of membrane inside an infolding also varied depending on depth (Fig. 3a.ii, 3e). Layer 2/3 neurons had smooth nuclear membranes, whereas there was a clear gradient of infolding within layer 4. All layer 5 excitatory cells had high degrees of infolding, despite the notable diversity of cell types and sizes within that population.[19] Infolding dropped again in layer 6 (Fig. 3a.ii). On the other hand, inhibitory nuclei had 15-30% of their membrane within an infolding, regardless of their position within cortex. This made them quite distinct from excitatory neurons in layer 1, 2/3, 4 and 6 of cortex, but highly similar to those in layer 5 (Fig. 3a.ii, Extended Data Fig. 1). Non-neuronal cells generally did not have infoldings, though microglia, OPCs and oligodendrocytes had less spherical and convex shapes (Fig. 3e). Pericytes had the smallest overall volumes with shapes dominated by their close apposition to the vascular walls (Fig. 3e).

Two features alone, nucleus volume and soma cutout area, revealed a striking separation between the major cell classes found in the brain (Fig. 3b). Neurons were separated from all non-neuronal classes while each non-neuronal subclass occupied distinct portions of this 2-dimensional space. The large surface area measurement for astrocytes was explained by the high density of their processes near the soma. Moreover, the high prevalence of segmentation mergers of pericytes with cortical vasculature resulted in variability in their soma size features as represented by the range in soma cutout area (Fig. 3b). Including the somatic features along with the nucleus features, we trained a classifier to distinguish neurons, non-neurons and erroneous segmentations from each other with a cross validated accuracy of 95.6%, and a classifier on non-neuronal classes with 97.5% accuracy (Extended Data Table 1).

Excitatory neurons showed a consistent synapse density that varied in a linear fashion with depth through the cortical volume. There was a notable increase in variation in layer 5 that correlated with the three subclasses found there with ET cells having larger synapse densities, NP

cells with low synapse densities and IT cells in between (Fig. 3a.iii., 3e). Inhibitory cells had much larger density of somatic innervation than excitatory cells, but also have a much wider degree of variation, reflecting previously recorded diversity of inhibitory subclasses (Fig. 3a.iii, 3e).[18,20–23]. Classifier accuracy for excitatory subclasses was high (90% Extended Data Table 1), with most confusion surrounding IT cells located at laminar borders, general locations expert annotators can disagree. Notably, accuracy was high across the layer 5 cell types (99% for NP, 85% for IT, and 87% for ET).

To gain a broader understanding of the perisomatic feature landscape, we computed a low-dimensional embedding based on both nucleus and somatic features (Fig. 3c). Consistent with the diversity observed in individual features, low dimensional embeddings of the feature space across all the cells in the dataset (gray, n = 92,391) reflected the variations observed from the manually labeled cortical column (cell class colors, n = 1,619). Non-neuronal cell classes occupied distinct areas of the feature space whereas excitatory neurons were primarily organized by cortical layers (Extended Data Fig. 1). Inhibitory neurons were largely restricted to distinct clusters within this space, with some cells overlapping with cortical layer 5 cells due to the increase in nuclear infolding in those excitatory neurons. Although there were broad differences in the nucleus and somatic features between the major interneuron subclasses (accuracy of 90%) (Fig, 3d, Extended Data Table 1), upon inspection, it was clear that the ultrastructure the perisomatic regions of inhibitory neuron was diverse in ways that the soma and nucleus features did not capture (Fig.4a).

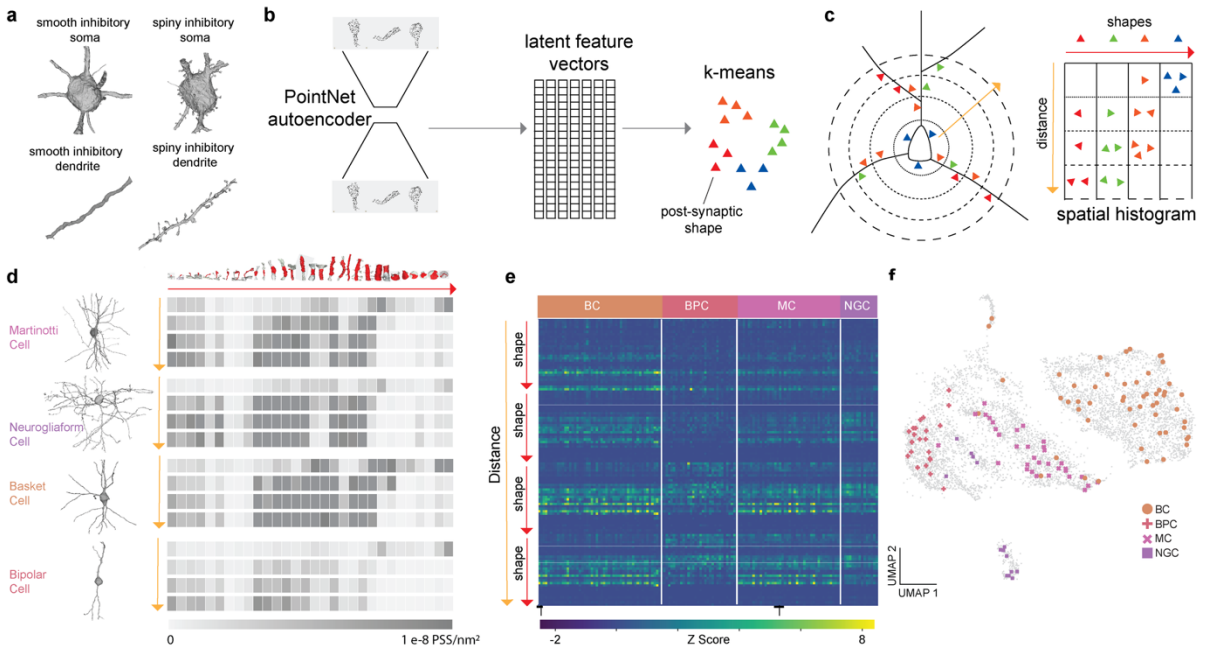


Figure 4: Post Synaptic Shape (PSS) Features. a) Inhibitory neurons elicit large variability in ultrastructural morphology. **b)** Procedure for building a PSS dictionary model. The set of shapes is used to train a PointNet autoencoder which learns a latent feature vector of a fixed size (1024). This autoencoder is then applied to all shapes in the dictionary to generate a set of latent feature vectors. K-means with K = 30 is applied to this to obtain a set of cluster centers for binning the shapes. **c)** For each cell, the PSS are binned by shape type and distance from the soma (4 bins) from 0 - 60 microns with 15 micron bin sizes. The resulting histogram is a 2D histogram shown above with the shapes in the x direction and distances in the y direction. **d)** Examples of 60 micron cutouts of the 4 predicted inhibitory subclasses with their spatial histograms shown as heatmaps. The top row shows the shape of the cluster center of each of the 30 clusters. In each heat map, darker boxes indicate higher values. **e)** Z-scored feature matrix representing the distance binned PSS features on the manually labeled inhibitory cells from the cortical column (n=143). Cells are organized by their annotated subclass. Dashed marks along the x axis denote segments of 100 cells **f)** 2D UMAP of all the inhibitory neurons (n=6,805) inferred after concatenating nucleus, soma and PSS features, cortical column cells in color and dataset wide inhibitory neurons in gray.

Proximal PSSs Across Inhibitory Neurons

Proximal inhibitory branches varied in caliber and surface texture, from smooth and uniform to covered in small spine-like protrusions (Fig. 4a). To take advantage of this, we developed a method to summarize these fine shape statistics of the proximal arbor.[24] We used the automated synapse detections to identify areas on the dendrites where changes in fine structure (e.g. spines) may occur. This approach gave us the added advantage of combining information

about synaptic innervation with the fine morphological structure of the postsynaptic neuron surrounding any given synapse.

We computationally segmented the compartment on the postsynaptic side of each synapse, which we refer to as the "Post Synaptic Shape" (PSS).[24] This shape, computed as a variable sized mesh, represented either a portion of the soma, the shaft of a dendrite, or a spiny protrusion, though it could also be an axon or axon-initial segment. To model the diversity of these shapes, we needed to be able to quantitatively compare them. We therefore trained a PointNet auto-encoder that allowed us to generate a fixed size representation of each shape (Fig. 4b).

To measure the distribution of shapes present in a cell, we collected 236,000 PSSs from a variety of neurons and applied a two dimensional reduction in order to visualize their distribution. This resulted in a continuous latent space where PSS objects of similar morphological character were closer together (Extended Data Fig. 2). We summarized this PSS space into a 30 dimensional histogram using k-means, to describe the distribution of shapes within a cell (see methods).

We observed that the location of the PSS could further distinguish between cells. For example, while spiny protrusions were most often found on the dendrites of cells, some cells also had them on the soma (Fig. 4a, Extended Data Fig. 3). Therefore, we took a second step to summarize a cell's distribution of PSSs by adding distance from the nucleus center. For distance binning we used four 15 micron bins between 0 to 60 microns from the soma (see methods). Combining the shape and distance binning resulted in a 120 dimensional spatial shape histogram (Fig. 4c), that summarizes information about the spatial organization of dendritic shapes and synapse densities near the soma, similar to a multi-dimensional Sholl analysis.[25] There were clear visual differences in the spatial histograms of different cell types (Fig. 4d-e, Extended Data Fig. 3). For example, a Martinotti cell had a greater density of synapses onto small protrusions on its

proximal dendrites than the basket or bipolar cell, but similar numbers to the neurogliaform cell. However, the neurogliaform cell had very few synapses on its soma, whereas the Martinotti has many, both onto smaller protrusions and smoother compartments of its somatic compartment (Fig. 4d).

We extracted these features on the vast majority of putative inhibitory neurons in the dataset (as predicted by perisomatic features, see methods). Appending these features to the soma features increased accuracy of the inhibitory subclass classifier to 94% (Extended Data Table 1). This was also reflected in the 2D UMAP embedding where inhibitory subclasses were more separable (Fig. 4f).

Dataset Wide Classification

To enable dataset wide classification, we developed a hierarchical model that used a cascade of classifiers to sort cells at increasingly finer distinctions. Classifiers were integrated into a comprehensive model where individual cells are sorted down the hierarchical tree (Fig. 5a). We utilized our collection of classification models (support vector machines or multilayer perceptrons) on different feature sets: 1 - nucleus only, 2 - nucleus and soma, 3 - nucleus, soma and PSS and found an optimal combination of classifiers which predicted cell types labeled within the column with an overall accuracy of 91% (Fig. 5b, Extended Data Table 1). All classifiers were trained using the labels from the manually labeled cortical column (see methods for more detail). Notably, this provided classifications for 88% of the cellular objects in the dataset (94,010/106,761 cells). To further validate this classification, we randomly sampled 100 cells from each subclass predicted by the hierarchical model and had anatomical experts assess the labels (Extended Data Fig. 4). For many classes, the average classification in this validation was consistent with performance accuracy within the column. The lower validation accuracy in the inhibitory subclasses as well as

5P-ET and 5P-NP was related to the sparse sample sizes in the training data from the column. The largest single confusion between types here was between adjacent layers of similar pyramidal classes, where strict laminar boundaries separating manual classes is less confident. This demonstrates that these features are indeed useful for separating cell-types based on local somatic reconstructions of cortical cells, consistent with the structure of the low dimensional embedding (Fig. 5c). Furthermore, predictions of cell density and overall cell counts per subclass across the dataset (Extended Data Fig. 5) corroborate the sampling rates we would expect from previous studies.[26–29] Importantly, this approach can be adapted to accommodate new cell-type labels derived from more detailed or expansive studies of the dataset, creating a scalable platform for extending labels derived on smaller numbers of cells to dataset wide coverage. For example, we have trained models based on the unsupervised clustering labels of morphological and connectivity properties of the same column cells as described in Schneider-Mizell 2023 (Extended Data Fig. 6).

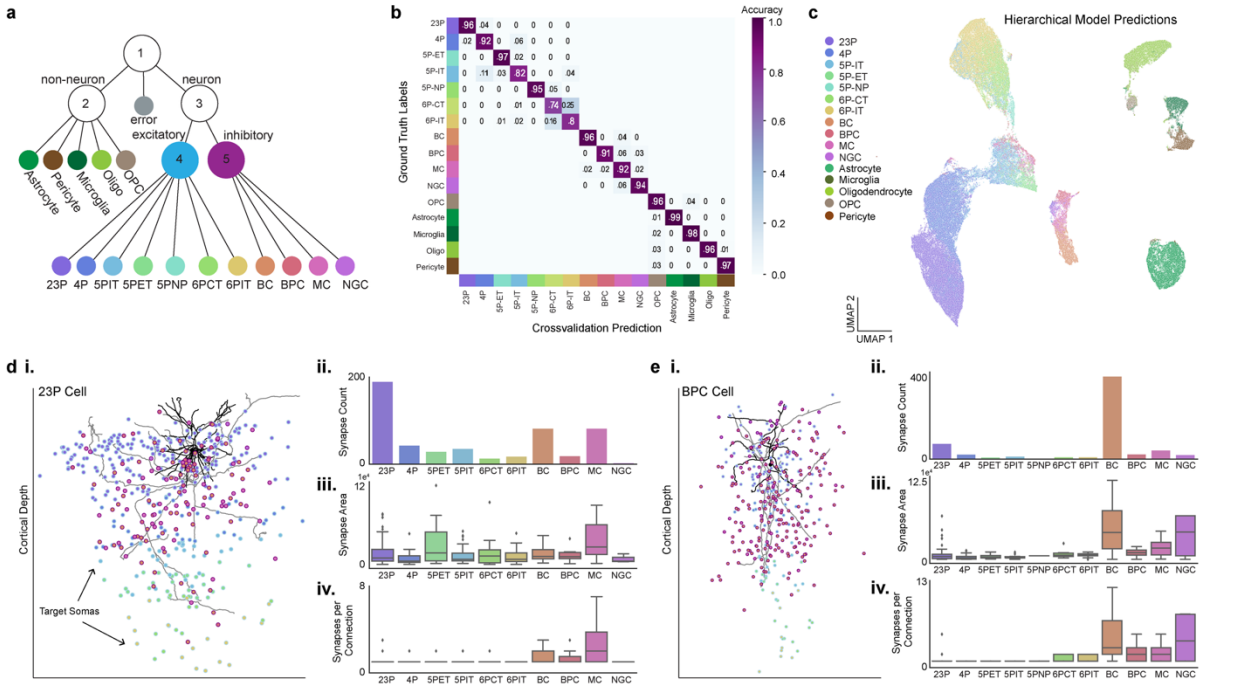


Figure 5: Hierarchical predictions enable dataset wide circuit analyses. a) Diagram of the hierarchical model framework used to predict neuronal and non-neuronal subclasses using a set of 5 classifiers. Nucleus and soma features alone were used for models 1-4. PSS features were added to predict inhibitory subclasses in model 5 (Extended Data Table 1). **b)** Confusion matrix of the cross validation performance for all cells within the manually labeled column. Note that classifiers for excitatory neurons, inhibitory neurons, and non-neurons were trained separately (Models 2,4,5 in panel a). The confusion rate between these classes can be seen in Extended Data Fig. 4. **c)** 2D UMAP embedding inferred from depth, nucleus, and soma features of all cells in the dataset colored by the hierarchical model predictions. **d) i)** 2D rendering of a representative 23P cell morphology, dendrite in black and axon in gray. Points represent the somatic position of all downstream target cells colored by the hierarchical model subclass prediction. **ii)** synapse count **iii)** total synapse area and **iv)** number of synapses per connection displayed by the model predicted subclasses illustrating the local targeting profile of this individual cell. **e)** Similar information as in d but for an inhibitory bipolar cell that is predicted to preferentially target basket cells. This unique population of bipolar cells has been further characterized in Schneider-Mizell 2023. For all box plots, center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range, outliers shown.

Dataset wide classifications enable a range of subsequent analyses. The typical axon of a well-proofread neuron has hundreds or thousands of postsynaptic targets.[5] To quantify the cell-type specific connectivity of such cells, each of those targets should have a cell-type label. Doing so manually is a practical bottleneck in analyzing these data. With these predictions, scientists can analyze the most numerous postsynaptic targets, the weight of these synapses with respect to predicted synapse area, and the number of synapses between proofread cells and cell subclasses across thousands of synapses (Fig. 5d-e). For example, a given layer 2/3 pyramidal neuron made the most synapses onto other 23P neurons (Fig. 5d). However, when we looked at the total predicted synapse size, 5P-ET neurons receive some of the largest average synapses. On the other hand, some examples were more surprising than this 23P cell. For example, bipolar cells (which largely overlap with a VIP subclass) have been described as the only disinhibitory specialist interneuron class, and are described as making synapses primarily onto SST cells (which are thought to overlap with the Martinotti Cell definition used here).[30–33] Although the dataset contains cells consistent with that view, a companion study on extensively proofread cells identified a collection of disinhibitory multipolar neurons which exhibit strong targeting preferences for basket

cells.[15] This unique connectivity profile is observed in the dataset wide classifications as well (Fig. 5e).

Efficient Search For Rare Cells

Studying the connectivity patterns of cell types requires identifying many example cells of a particular connectivity profile. With more than 70,000 neurons sampled across a millimeter scale there should be many examples of any individual cell type. However, locating those examples can be challenging for rare subclasses because of their infrequent appearance and the need for axonal proofreading in order to use connectivity to suggest their subclass.

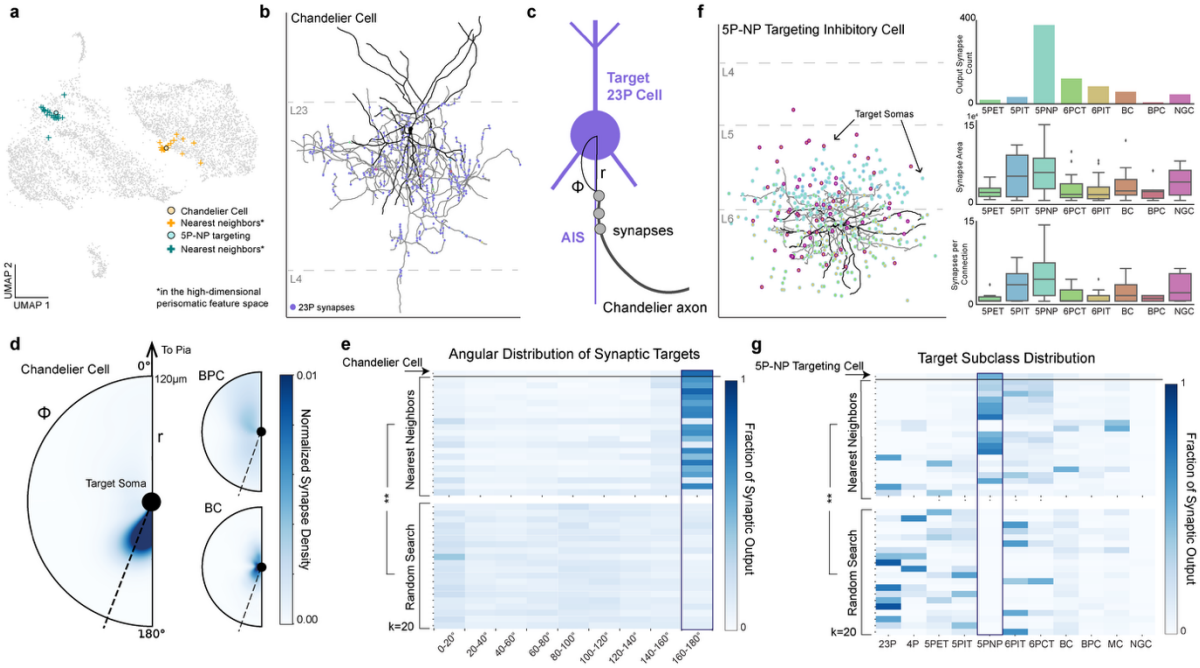


Figure 6: Perisomatic feature space enables more efficient search for unique cells. a) 2D UMAP embedding highlighting a chandelier cell (orange dot), a 5P-NP targeting cell (blue dot), and their respective 20 nearest neighbors (+) in the perisomatic feature space. Note: UMAP non-linearly distorts feature space, so not all nearest neighbors appear closest in the plot. **b)** Example proofread chandelier cell in L23 (dendrite in black, axon in gray). Output synapses marked along the axon and colored by subclass prediction. Note the characteristic vertical chains of synapses onto 23P cells. **c)** Chandelier cells are characterized by their preference to synapse onto the axon initial segment (AIS) of target cells.[34,35] Here quantified by measuring the angle between the target soma and the synapse (ϕ) and the distance from the soma (r). **d)** Heatmap illustrating the angle and distance distribution of the chandelier cell shown in b as well as two non Chandelier inhibitory examples. Color denotes the normalized synapse density for

each cell. Synapses that had an angle >160° were considered onto the AIS of the target cell (shown by the dotted line). **e)** Angular distribution histogram of the chandelier cell (top row), 20 nearest neighbors in the perisomatic feature space, and 20 random inhibitory cells ($p<0.00001$). **f)** Example cell that preferentially targets rare 5P-NP subclass (dendrite in black, axon in gray), points represent target cell soma locations colored by predicted subclass. Output synapse counts reflect strong preference to 5P-NP cells. For all box plots, center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range, outliers shown. **g)** Fraction of output connectivity onto neuronal subclasses of the 5P-NP targeting cell (top row), 20 nearest neighbors in the perisomatic feature space, and 20 random inhibitory cells ($p<0.00001$). All p-values reported and asterisks note significance by two-tailed Fisher Exact Test.

Given that the major inhibitory neuron subclasses differ in their connectivity profiles, we already had some evidence that connectivity profile correlates with the perisomatic features we extracted (Fig. 4e), but we conjectured they could be useful for finding rarer types with specific connectivity patterns for which we did not yet have labels. One well known rare cell type in mouse visual cortex is the chandelier cell, which synapses onto the axon initial segment (AIS) of excitatory neurons.[20,34–38] We used a single proofread chandelier cell to see if we could facilitate finding other cells like it using the perisomatic features. We selected the top 20 nearest neighbors of the perisomatic feature space (Fig. 6a) and assessed what fraction of them were chandelier cells based on their connectivity profiles after cleaning them of false mergers and modest axonal extension (see methods). The chandelier cell's connectivity profile is easy to recognize, both from its morphology where it makes vertical strings of synapses (Fig. 6b), and the unique targeting of synapses onto axon initial segment (AIS) of excitatory neurons. Because the AIS is located just below the soma of excitatory cells in the cortex, the angular distribution of synapses relative to somatic targets can be used as a spatial proxy for AIS targeting (Fig. 6c-d). A histogram of the angular distribution of synapses relative to the target soma demonstrates that 16 of 20 the nearest neighbor cells have connectivities consistent with chandelier cells (Fig. 6d-e). In contrast, none of the 20 random interneurons we sampled from the inhibitory neurons in the dataset, or any of the 143 interneurons sampled in the column were chandelier cells, reflecting a significant enrichment ($p<0.00001$ by two-tailed Fisher exact test). Based on this success, we tried to find more examples

of cells with a less well known connectivity profile. We selected an undescribed but proofread interneuron which made the majority of its synapses in layer 5 onto 5P-NP neurons, despite those pyramidal neurons themselves being rare and with few input synapses (Fig. 6f).[39] Picking the top 20 nearest neighbors of this cell, we found 13 cells which made at least 30% of their synaptic targets onto 5P-NP cells (based on classifier predictions). This stands in contrast to 0 of the 20 random interneurons we sampled, or 2 cells of the other 143 sampled in the column, again a significant enrichment ($p<0.00001$) (Fig. 6g). This application demonstrates how these perisomatic features can facilitate the search of rare cell types in the cortex.

Discussion

Our analysis of the perisomatic region of cells in the mouse visual cortex demonstrates that a surprising amount of cell-type information can be extracted from the somatic regions of brain cells. Our approach has already been used to characterize the connectivity of distinct types of layer 5 Martinotti cells,[16] the inter-related connectivity motifs of layer 5 thick tufted cells (5P-ET) and the surrounding inhibitory sub-network,[17] and to confirm the connectivity profiles of interneurons outside the manually labeled column.[15] Future work in this dataset and others will leverage iterations of dataset wide cell classifiers to discover novel aspects of cell-type specific wiring of cortical circuits. Other cell classification approaches have been applied to this dataset, including unsupervised clustering of morphological features, and supervised approaches based on morphological graphs.[40,41] All these approaches have focused on smaller subsets of the data that contained higher quality or complete reconstructions, reducing their effective coverage in the datasets to less than half the cells.

The breadth of cells in large-scale electron microscopy data makes it an attractive modality to study cell-types. Our approach provides an example of how computational methods are an important practical tool for directing study to small subsets of cells within large datasets. This is particularly clear within diverse and rare inhibitory cells (Fig. 5-6), but similar questions arise among glial sub-types. One such example is the difference between OPCs and premyelinating oligodendrocyte cells which are thought to be differentiated OPCs that are in transitional states to oligodendrocytes.[42] The structural diversity of cells predicted as OPCs within the low dimensional embedding space (Fig. 5c) suggests that searching within the perisomatic feature space, as illustrated in Fig. 6, could be used to facilitate scientific discovery across brain cell types. More broadly, some of the features described here can be measured with other techniques, such as x-ray tomography or light microscopy, and can be used to distinguish cells into different subclasses in a manner similar to what has been presented here.

Many studies of anatomical diversity of cortical cells have focused on the diversity of dendritic and somatic morphologies, axonal projection patterns, and synaptic connectivity patterns.[3,4,21,43] Fewer studies have focused on differences of somas,[44–46] particularly quantitative studies of the 3D ultrastructure of the soma with large single cell sample sizes across all layers of cortex. Laminar differences in cell body size distributions are well known, and serve as the basis for cyto-architectural definitions of layers, which corresponds with shifts in cell-type distributions, specifically excitatory ones.[47] For example, pyramidal Layer 5 ET projection neurons are characterized by their large somas. This likely reflects differing demands for gene expression and metabolic load.[48] Also, 5P-NP neurons have been recognized before as having smaller rounder somas on average.[39,49] Further, anecdotal descriptions of variations in nucleus in-folding have been reported, though only in two dimensions within a narrower range of types.[50,51] Intriguingly,

differences in nucleus infoldings have been reported to be modulated by activity in some brain areas.[52] The results we present here not only corroborate the variability of these features across cells but suggest that the perisomatic region holds greater cell-type specific information than previously recognized.

There are a few limitations to this work that should be kept in mind when interpreting its results. Our most detailed analysis has only been completed on one dataset which comes from a single animal. That said, some patterns are consistent with what was found in smaller published dataset from layer 2/3,[18,20,53] and the basic patterns found in these features across mouse visual cortex are reproduced in a second smaller dataset (Extended Data Fig. 7). Our approach is not the final word in cell-type predictions in this dataset, or large-scale EM in general, and there are a number of dimensions of potential improvement. Further, cell-type labels will continue to evolve as more cells are classified by either human experts or quantitative methods with increasing specificity and sophistication. In particular, our validation results are consistent with there being a larger diversity of inhibitory cells than exist within the column thus expanding the number of class labels could improve performance. However, we think the dataset wide framework we have presented here will continue to be valuable, as we expect any new labels to only be available for a subset of cells. Finally, this model does not use all the information present at the soma of neurons. For example, the detailed ultrastructure visible in the imagery is not fully utilized. Other methods have utilized the underlying imagery of cells to distinguish cell types, either through detection of more subcellular organelles like cilia or by using imagery more directly to define abstract embeddings.[54–56]

Beyond the somatic region, there are a large variety of studies have shown how local features visible in the ultrastructure contain information that encode information about cell types,

including neurotransmitters of fly synapses, identity of neuromodulatory axons, or cutouts of local dendrite and axons.[54,57] These results all support the view that large-scale quantitative measurements of ultrastructure provides a rich basis for identifying cellular properties of cells. The efficacy of these approaches provides a roadmap for how to develop a scalable platform for leveraging local features of cells to infer cell-type classifications. Beyond neuroscience, this approach illustrates how large-scale ultrastructural imaging of cells can facilitate the study of diverse and rare cell populations if paired with appropriate quantitative analysis.

References

1. Tasic, B. *et al.* Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018).

2. Gouwens, N. W. *et al.* Classification of electrophysiological and morphological neuron types in the mouse visual cortex. *Nat. Neurosci.* **22**, 1182–1195 (2019).

3. Scala, F. *et al.* Phenotypic variation of transcriptomic cell types in mouse motor cortex. *Nature* **598**, 144–150 (2021).

4. Kepecs, A. & Fishell, G. Interneuron Cell Types: Fit to form and formed to fit. *Nature* **505**, 318–326 (2014).

5. Consortium, Mic. *et al.* Functional connectomics spanning multiple areas of mouse visual cortex. 2021.07.28.454025 Preprint at https://doi.org/10.1101/2021.07.28.454025 (2021).

6. Yin, W. *et al.* A petascale automated imaging pipeline for mapping neuronal circuits with high-throughput transmission electron microscopy. *Nat. Commun.* **11**, 4949 (2020).

7. Macrina, T. *et al.* Petascale neural circuit reconstruction: automated methods. 2021.08.04.455162 Preprint at https://doi.org/10.1101/2021.08.04.455162 (2021).

8. Phelps, J. S. *et al.* Reconstruction of motor control circuits in adult Drosophila using automated transmission electron microscopy. *Cell* **184**, 759-774.e18 (2021).

9. Motta, A. *et al.* Dense connectomic reconstruction in layer 4 of the somatosensory cortex. *Science* **366**, eaay3134 (2019).

10. Dorkenwald, S. *et al.* Neuronal wiring diagram of an adult brain. *bioRxiv* 2023.06.27.546656 (2023) doi:10.1101/2023.06.27.546656.

11. Shapson-Coe, A. *et al.* A connectomic study of a petascale fragment of human cerebral cortex. 2021.05.29.446289 Preprint at https://doi.org/10.1101/2021.05.29.446289 (2021).

12. Kanari, L. *et al.* Objective Morphological Classification of Neocortical Pyramidal Cells. *Cereb. Cortex* **29**, 1719–1735 (2019).

13. Scorcioni, R., Polavaram, S. & Ascoli, G. A. L-Measure: a web-accessible tool for the

analysis, comparison and search of digital reconstructions of neuronal morphologies. *Nat. Protoc.* **3**, 866–876 (2008).

14. Costa, M., Manton, J. D., Ostrovsky, A. D., Prohaska, S. & Jefferis, G. S. X. E. NBLAST: Rapid, Sensitive Comparison of Neuronal Structure and Construction of Neuron Family Databases. *Neuron* **91**, 293–311 (2016).

15. Schneider-Mizell, C. M. *et al.* Cell-type-specific inhibitory circuitry from a connectomic census of mouse visual cortex. 2023.01.23.525290 Preprint at https://doi.org/10.1101/2023.01.23.525290 (2023).

16. Gamlin, C. R. *et al.* Integrating EM and Patch-seq data: Synaptic connectivity and target specificity of predicted Sst transcriptomic types. 2023.03.22.533857 Preprint at https://doi.org/10.1101/2023.03.22.533857 (2023).

17. Bodor, A. L. *et al.* The Synaptic Architecture of Layer 5 Thick Tufted Excitatory Neurons in the Visual Cortex of Mice. 2023.10.18.562531 Preprint at https://doi.org/10.1101/2023.10.18.562531 (2023).

18. Turner, N. L. *et al.* Reconstruction of neocortex: Organelles, compartments, cells, circuits, and activity. *Cell* **185**, 1082-1100.e24 (2022).

19. Kim, E. J., Juavinett, A. L., Kyubwa, E. M., Jacobs, M. W. & Callaway, E. M. Three Types of Cortical Layer 5 Neurons That Differ in Brain-wide Connectivity and Function. *Neuron* **88**, 1253–1267 (2015).

20. Schneider-Mizell, C. M. *et al.* Structure and function of axo-axonic inhibition. *eLife* **10**, e73783 (2021).

21. Kubota, Y., Karube, F., Nomura, M. & Kawaguchi, Y. The Diversity of Cortical Inhibitory Synapses. *Front. Neural Circuits* **10**, 27 (2016).

22. Callaway, E. M. Inhibitory Cell Types, Circuits and Receptive Fields in Mouse Visual Cortex. in *Micro-, Meso- and Macro-Connectomics of the Brain* (eds. Kennedy, H., Van Essen, D. C. & Christen, Y.) (Springer, Cham (CH), 2016).

23. Sultan, K. T. & Shi, S.-H. GENERATION OF DIVERSE CORTICAL INHIBITORY

INTERNEURONS. *Wiley Interdiscip. Rev. Dev. Biol.* **7**, 10.1002/wdev.306 (2018).

24. Seshamani, S. *et al. Automated Neuron Shape Analysis from Electron Microscopy*. http://arxiv.org/abs/2006.00100 (2020) doi:10.48550/arXiv.2006.00100.

25. Sholl, D. A. Dendritic organization in the neurons of the visual and motor cortices of the cat. *J. Anat.* **87**, 387-406.1 (1953).

26. Keller, D., Erö, C. & Markram, H. Cell Densities in the Mouse Brain: A Systematic Review. *Front. Neuroanat.* **12**, (2018).

27. Rudy, B., Fishell, G., Lee, S. & Hjerling-Leffler, J. Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Dev. Neurobiol.* **71**, 45–61 (2011).

28. Kim, Y. *et al.* Brain-wide Maps Reveal Stereotyped Cell-Type-Based Cortical Architecture and Subcortical Sexual Dimorphism. *Cell* **171**, 456-469.e22 (2017).

29. Cragg, B. G. The density of synapses and neurones in the motor and visual areas of the cerebral cortex. *J. Anat.* **101**, 639–654 (1967).

30. Kapfer, C., Glickfeld, L. L., Atallah, B. V. & Scanziani, M. Supralinear increase of recurrent inhibition during sparse activity in the somatosensory cortex. *Nat. Neurosci.* **10**, 743–753 (2007).

31. Pfeffer, C. K., Xue, M., He, M., Huang, Z. J. & Scanziani, M. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nat. Neurosci.* **16**, 1068–1076 (2013).

32. Lee, S., Kruglikov, I., Huang, Z. J., Fishell, G. & Rudy, B. A disinhibitory circuit mediates motor integration in the somatosensory cortex. *Nat. Neurosci.* **16**, 1662–1670 (2013).

33. Pi, H.-J. *et al.* Cortical interneurons that specialize in disinhibitory control. *Nature* **503**, 521–524 (2013).

34. Defelipe, J., Hendry, S. H. C., Jones, E. G. & Schmechel, D. Variability in the terminations of GABAergic chandelier cell axons on initial segments of pyramidal cell axons in the monkey sensory-motor cortex. *J. Comp. Neurol.* **231**, 364–384 (1985).

35. Fairén, A. & Valverde, F. A specialized type of neuron in the visual cortex of cat: A Golgi

and electron microscope study of chandelier cells. *J. Comp. Neurol.* **194**, 761–779 (1980).

36. Somogyi, P., Freund, T. F. & Cowey, A. The axo-axonic interneuron in the cerebral cortex of the rat, cat and monkey. *Neuroscience* **7**, 2577–2607 (1982).

37. Jones, E. G. Varieties and distribution of non-pyramidal cells in the somatic sensory cortex of the squirrel monkey. *J. Comp. Neurol.* **160**, 205–267 (1975).

38. Peters, A., Proskauer, C. C. & Ribak, C. E. Chandelier cells in rat visual cortex. *J. Comp. Neurol.* **206**, 397–416 (1982).

39. Sorensen, S. A. *et al.* Correlated Gene Expression and Target Specificity Demonstrate Excitatory Projection Neuron Diversity. *Cereb. Cortex* **25**, 433–449 (2015).

40. Celii, B. *et al.* NEURD: automated proofreading and feature extraction for connectomics. 2023.03.14.532674 Preprint at https://doi.org/10.1101/2023.03.14.532674 (2023).

41. Weis, M. A. *et al.* Large-scale unsupervised discovery of excitatory morphological cell types in mouse visual cortex. 2022.12.22.521541 Preprint at https://doi.org/10.1101/2022.12.22.521541 (2023).

42. Buchanan, J. *et al.* Oligodendrocyte precursor cells ingest axons in the mouse neocortex. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2202580119 (2022).

43. Gouwens, N. W. *et al.* Integrated Morphoelectric and Transcriptomic Classification of Cortical GABAergic Cells. *Cell* **183**, 935-953.e19 (2020).

44. Attili, S. M., Silva, M. F. M., Nguyen, T.-V. & Ascoli, G. A. Cell numbers, distribution, shape, and regional variation throughout the murine hippocampal formation from the adult brain Allen Reference Atlas. *Brain Struct. Funct.* **224**, 2883–2897 (2019).

45. Luengo-Sanchez, S. *et al.* A univocal definition of the neuronal soma morphology using Gaussian mixture models. *Front. Neuroanat.* **9**, (2015).

46. Fariñas, I. & DeFelipe, J. Patterns of synaptic input on corticocortical and corticothalamic cells in the cat visual cortex. I. The cell body. *J. Comp. Neurol.* **304**, 53–69 (1991).

47. Gilman, J. P., Medalla, M. & Luebke, J. I. Area-Specific Features of Pyramidal Neurons—a Comparative Study in Mouse and Rhesus Monkey. *Cereb. Cortex N. Y. NY* **27**,

2078–2094 (2017).

48. Tasic, B. *et al.* Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018).

49. Kim, E. J., Juavinett, A. L., Kyubwa, E. M., Jacobs, M. W. & Callaway, E. M. Three Types of Cortical Layer 5 Neurons That Differ in Brain-wide Connectivity and Function. *Neuron* **88**, 1253–1267 (2015).

50. Peters, A. & Kara, D. A. The neuronal composition of area 17 of rat visual cortex. I. The pyramidal cells. *J. Comp. Neurol.* **234**, 218–241 (1985).

51. Georgiou, C. *et al.* A subpopulation of cortical VIP-expressing interneurons with highly dynamic spines. *Commun. Biol.* **5**, 1–15 (2022).

52. Wittmann, M. *et al.* Synaptic Activity Induces Dramatic Changes in the Geometry of the Cell Nucleus: Interplay between Nuclear Structure, Histone H3 Phosphorylation, and Nuclear Calcium Signaling. *J. Neurosci.* **29**, 14687–14700 (2009).

53. Dorkenwald, S. *et al.* Binary and analog variation of synapses between cortical pyramidal neurons. *eLife* **11**, e76120 (2022).

54. Dorkenwald, S. *et al.* Multi-Layered Maps of Neuropil with Segmentation-Guided Contrastive Learning. 2022.03.29.486320 Preprint at https://doi.org/10.1101/2022.03.29.486320 (2022).

55. Ott, C. M. *et al.* Nanometer-scale views of visual cortex reveal anatomical features of primary cilia poised to detect synaptic spillover. 2023.10.31.564838 Preprint at https://doi.org/10.1101/2023.10.31.564838 (2023).

56. Zinchenko, V., Hugger, J., Uhlmann, V., Arendt, D. & Kreshuk, A. MorphoFeatures for unsupervised exploration of cell types, tissues, and organs in volume electron microscopy. *eLife* **12**, e80918 (2023).

57. Eckstein, N. *et al.* Neurotransmitter Classification from Electron Microscopy Images at Synaptic Sites in Drosophila Melanogaster. 2020.06.12.148775 Preprint at https://doi.org/10.1101/2020.06.12.148775 (2023).

Methods

MICrONS Dataset

This dataset consists of a 1.4mm x 800µm x 800µm volumetric serial section EM dataset from mouse visual cortex of a male P87 mouse. The dataset covers all layers of cortex and spanning primary visual cortex and two higher visual areas. The dataset has been described in detail elsewhere.[5] Briefly, two photon imaging was performed on the mouse, which was subsequently prepared for electron microscopy. The specimen was then sectioned and imaged using transmission electron microscopy.[6] The images were then stitched, aligned, and processed through a deep learning segmentation algorithm, followed by manual proofreading.[5–7,15]

Cortical Column

In this manuscript we leveraged proofreading and labels that were done as part of a separate study of a 100µm columnar region of primary visual cortex within the larger dataset.[15] For clarity to the reader and completeness we are repeating some aspects of the methods that define that column here.

Column Selection:

The column borders were found by manually identifying a region in the primary visual cortex that was as far as possible from both the volume boundaries and the boundaries with higher order visual areas. A 100 x 100 µm box was placed on layer 2/3 and was extended along the y axis of the dataset. While analyzing the data, it was observed that deep layer neurons had apical dendrites that were not oriented along the most direct pia-to-white-matter direction, and thus adapted the definition of the column to accommodate these curved neuronal streamlines. Using a

collection of layer 5 ET cells, points were placed along the apical dendrite to the cell body and then along the primary descending axon towards white matter. The slant angle was computed as two piecewise linear segments, one along the cortical depth to lower layer 5 where little slant was observed, and one along the direction defined by the vector averaged direction of the labeled axons.

Using these boundaries and nucleus centroids,[5] all cells were identified inside the columnar volume. Coarse cell classes (excitatory, inhibitory, and non-neuronal) were assigned based on brief manual examination and rechecked by subsequent proofreading and confusion with early versions of the classifiers described here. To facilitate concurrent analysis and proofreading, all false merges were split connecting any column neurons to other cells (as defined by detected nuclei) before continuing with other work.

Proofreading:

Proofreading was performed by five expert neuroanatomists using the Connectome Annotation Versioning Engine[58,59] and a modified version of Neuroglancer.[60] Proofreading was aided by on-demand highlighting of branch points and tips on user-defined regions of a neuron based on rapid skeletonization (https://github.com/AllenInstitute/Guidebook). This approach quickly directed proofreader attention to potential false merges and locations for extension, as well as allowed a clear record of regions of an arbor that had been evaluated.

For dendrites, all branch points were checked for correctness and all tips to see if they could be extended.

False merges of simple axon fragments onto dendrites were often not corrected in the raw data, since they could be computationally filtered for analysis after skeletonization (see below).

Detached spine heads were not comprehensively proofread, and previous estimates place the rate of detachment at approximately 10-15%.

For inhibitory axons, axons were "cleaned" of false merges by looking at all branch points. Axonal tips were extended until either their biological completion or data ambiguities, particularly emphasizing all thick branches or tips that were well-suited to project to new laminar regions. For axons with many thousand synaptic outputs, some but not all tips were followed to completion once major branches were cleaned and established. For smaller neurons, particularly those with bipolar or multipolar morphology, most tips were extended to the point of completion or ambiguity. Axon proofreading time differed significantly by cell type not only because of differential total axon length, but axon thickness differences that resulted in differential quality of auto segmentations, with thicker axons being of higher initial quality. Typically, inhibitory axon cleaning and extension took 3-10 hours per neuron.Expert neuroanatomists further labeled excitatory and inhibitory neurons into subclasses. Layer definitions were based on considerations of both cell body density (in analogy with nuclear staining) supplemented by identifying kinks in the depth distribution of nucleus size near expected layer boundaries.

Cell Labeling:

For excitatory neurons, the categories used were: Layer 2/3-IT, Layer 4-IT, Layer 5-IT, Layer 5-ET, Layer 5-NP, Layer 6-IT, and Layer 6-CT cells. Layer 2/3 and upper Layer 4 cells were defined on the basis of dendritic morphology and cell body depth. Layer 5 cells were similarly defined by cell body depth, with projection subclasses distinguished by dendritic morphology following Gouwens, Sorenson, and Berg[2] and classical descriptions of thick (ET) and thin-tufted (IT) cells. Layer 5 ET cells had thick apical dendrites, large cell bodies, numerous spines, a pronounced apical tuft, and deeper ET cells had many oblique dendrites. Layer 5 IT cells had more

slender apical dendrites and smaller tufts, fewer spines, and fewer dendritic branches overall. Layer 5 NP cells corresponded to the "Spiny 10" subclass described in Gouwens, Sorenson, and Berg; these cells had few basal dendritic branches, each very long and with few spines or intermediate branch points. Layer 6 neurons were defined by cell body depth, but only some cells were able to be labeled as IT or CT by human experts. Layer 6 pyramidal cells with stellate dendritic morphology, inverted apical dendrites, or wide dendritic arbors were classified as IT cells. Layer 6 pyramidal cells with small and narrow basal dendrites, an apical dendrite ascending to Layer 4 or Layer 1, and a myelinated primary axon projecting into white matter were labeled as CT cells.

Basket cells were recognized as cells which made more than 20% of their synaptic inputs onto the soma or proximal dendrites of cells. Neurogliaform cells were recognized by having a low density of output synapses, and boutons that had often had synaptic vesicles but no post-synaptic structures. Bipolar cells were labeled by having only 2 or 3 primary dendrites, and primarily making synapses with other inhibitory neurons. Note, the Martinotti/non-Martinotti subclass label was given to cells that have previously been described in the literature to primarily target the distal dendrites of excitatory neurons without exhibiting hallmark features of bi-polar or neurogliaform cells.

Due to high levels of proofreading in the column, there were very few errors thus the training set was augmented with manually labeled errors from the entire dataset.

Proofreading and Truncation Analysis

For every proofread cell in the cortical column (described above) we compared the cellular volume of the initial reconstruction from the automated segmentation to the cleaned and completed

reconstruction. To measure the precision connectivity for each cell we noted the number of synapses that got removed with proofreading, the number of synapses that were added, and the number of synapses that were maintained with each cell before and after proofreading.

To estimate the likelihood of truncation, we measured the distribution of dendritic extents from the proofread column cells. For each cell we measured the radial distance of each input synapse from the cell's soma. The radial extent of a given cell was considered the distance of the 97th percentile input synapse. From this distribution we used the median value of 121 microns as a threshold for dendritic truncation, although closer to 250 microns would be required to guarantee no truncation for any cell. For the rest of the cells in the dataset, we measured the distance of the soma from the volume borders in x and z. The overlap in these distributions relates to the probability of truncation, leading to our conclusion that roughly one third of the cells have some degree of dendritic truncation.

Generating Nucleus and Soma Features

We analyzed nuclei using the results of a deep neural network segmentation,[5] extracted the mesh using marching cubes and obtained the largest component of the detected mesh. Nuclear features were then extracted on the remaining meshes. These features included, nucleus volume, nucleus area, the area to volume ratio, nucleus surface area within an infolding, the fraction of the total surface area within an infolding, and cortical depth (measured as the distance from the pial surface). Nucleus fold features were extracted by creating a shrink wrapped[47] mesh for each nucleus mesh. We then calculated the distance of each vertex on the nucleus mesh from the shrink-wrapped mesh. After visual inspection of cells across all the reported subclasses, any vertex further than 150 nm was considered within an infolding.

For each nucleus detection the somatic compartment was identified as the ID in the segmentation which surrounded >80% of the nucleus. Somatic segmentations (downloaded at 64x64x40 nm resolution) went through a heuristic cleaning procedure to remove missing slices of data and incorrectly merged fragments. Since each soma was matched to its corresponding nucleus, 15 microns surrounding the nucleus' center of mass was cut out from the dense segmentation and converted into a binary mask. 15 microns was chosen due to the high quality of the segmentation (Fig. 2a) and it was large enough to encompass the entire soma of all cells from the smallest glial cell to the largest 5P-ET neuron. Binary dilation by 5 voxels in 3d was performed, followed by filling of all holes, and then binary erosion of 3 voxels. The resulting binary mask was meshed using marching cubes and connected component analysis was run on the result. 5 voxels was deemed an appropriate dilation to remove merged fragments without creating additional holes in the mesh. The largest connected component mesh was retained, and any disconnected components were dropped. Somatic features were extracted for all nuclear detections that were not cut off by the volume boundary (see Filtering procedure). These somatic features included soma area, soma volume, the area to volume ratio, the number of synapses on the somatic cutout, and the soma synapse density. Using both the somatic and nucleus meshes, we calculated the ratio between the nucleus volume and soma volume and the offset between the two, measured as the euclidean distance between nuclear center of mass and soma center of mass.

Filtering procedure:

There were 133,580 nuclear detections in the dataset and the filtering procedure consisted of three steps. Firstly, any detected objects less than 25μm^3 were filtered out as errors as these largely consisted of small fragments of nucleoli. Second, after identifying the segment IDs within a 15μm bounding box around each nucleus, if over 20% of these IDs corresponded to error ID 0,

they were filtered out. The majority of these error cases were cells close to the volume border or areas in the volume with higher segmentation errors such as those near blood vessels. Thirdly, cells that were predicted as errors based on the object classifier of the hierarchical model described below (Fig. 5a) were also removed from analysis. This resulted in a final set of 94,010 cells, neuronal and non-neuronal.

Feature normalization:

Due to differences in section thickness during sample preparation, we noticed abrupt shifts in nucleus and soma size features along the sectioning axis (Z plane). This presumably is due to changes in section thickness across the dataset. To account for these abrupt and systematic shifts we binned the entire dataset by the longest length scale for which there didn't appear to be systematic shifts in the distribution in the z plane (800 nm) and normalized each feature value by the average within each Z bin.

For 2D UMAP embeddings and training of the classifiers it was important to place all features in approximately similar scales. For this reason, we independently Z-scored each feature across all cells and used that as the input for classifier training as well as the UMAP embeddings in Fig. 3-6.

Generating PSS Features

Around each synapse, we extracted a 3500 nm region to obtain the synapse region mesh. We experimented with region cutouts between 1000 to 5000 nm, however smaller cutouts led to ambiguities in the main shaft identification and thereby produced errors in the subsequent skeletonization. At 3500 nm the skeletons were more stable and segmenting as expected. This mesh was then segmented using the CGAL surface segmentation algorithm[61] which splits regions

based on differences in thickness. We adapted our previously developed method[24] to identify the PSS region by using a local skeleton calculated from the synapse region mesh, rather than a precomputed whole cell mesh. This allowed us to adapt this method for cells in the dataset without the need for proofreading.

Given a cell for which all PSS have been extracted within a 60 micron radius from the nucleus center, the objective was to build a descriptor that encapsulates the various properties of the PSS. In particular, we aim to capture two of these properties: the type of shape of the PSS and the distance of the PSS from the soma. Moreover, since different cells can have different numbers of shapes (synapses), we needed a fixed size representation for each cell. To capture shape information, a dictionary of all shape types was built using a dictionary dataset from 236,000 PSSs from a variety of neurons.[24] These shapes were rotationally normalized and used to train a pointnet autoencoder[62] to learn a latent representation of size 1024. The high dimensional latent space spanning all these shapes is a continuous space (Extended Data Fig. 3) which was used to generate a Bag of Words model[30] for the shapes. To ensure we were sampling the entire embedding space, we performed K-means clustering with K=30 to estimate cluster centers. We manually re-ordered the bin centers for visualization purposes from shapes representing small spines, to those representing longer spines, to dendritic shafts of different shapes, and finally somatic compartments. The top row of Fig. 4c shows the shape in the dictionary that is closest to each of these cluster centers. For distance from the soma, we split the 60 micron radius around the nucleus center into four 15 micron radial bins (Fig. 4c). All PSS were then binned according to their shape and distance properties to generate a histogram of counts. Initially we extracted PSS from within 120 micron radius. However, upon inspection of the normalized histograms and the 2D UMAP embedding space, the additional radial bins did not increase our differentiability and did increase

truncation effects near the dataset thus we reduced the radius to 60 microns. Finally, this histogram was Z-scored and then added to the rest of the features as input to classifiers and the UMAP embedding Fig. 4 and Fig. 6.

Hierarchical Model Training and Validation

Hierarchical Framework:

We defined an object as the segmentation associated with a predicted nucleus[5] from which nucleus, soma, and PSS features could be extracted. A hierarchical framework was designed to predict the cell type of any such object (Fig. 5c). To begin, there were 106,761 nuclear segmentations that passed the first two filters described above (see filtering procedure). The first level in the hierarchy predicted whether an object was a neuron (72,158), non-neuron (21,856), or an error (12,751). All objects predicted as errors were excluded from all subsequent analyses except for the hierarchical model evaluation. Non-neuronal cells were then classified as one of the following: astrocyte (7,850), microglia (2,638), oligodendrocyte (7,020), oligodendrocyte precursor cells (OPC) (1,703), or pericyte (2,645). For neurons, cells were predicted as either excitatory (64,195) or inhibitory (7,963) followed by a separate subclass classifier for each class type. Excitatory subclasses: Layer 2/3 pyramidal (19,735), Layer 4 pyramidal (14,777), Layer 5 IT (7,949), Layer 5 ET (2,215), Layer 5 near projecting (NP) pyramidal (970), Layer 6 IT (11,734), Layer 6 CT pyramidal (6,815). After extracting PSS features from all predicted inhibitory neurons, a subset of neurons (n=1,158) that were actually excitatory clearly separated from the rest of the cells in the perisomatic feature space (with PSS features). This was expected due to known differences in proximal dendrite morphology between inhibitory and excitatory neurons. These neurons were then passed through the excitatory neuron classifier and labeled as excitatory for all subsequent analyses with a final set of 6,805 inhibitory cells with the following subclass counts:

Basket cells (3,239), Bipolar cells (997), Martinotti/non-Martinotti cells (1,992), and Neurogliaform cells (571).

Training:

Soma and nucleus features were extracted from the 3D mesh of all objects and PSS features were extracted for all neurons predicted as inhibitory. For each level of the hierarchy, multiple classifiers were trained using either nucleus only, nucleus and soma features, or nucleus, soma, and PSS features. Within each level of the hierarchy, classifiers were trained using the cells and labels from the manually annotated cortical column. Due to the sparsity of some of the cell classes, we augmented the training set in the following ways: 470 errors were added from within and around the column for the object model, 11 proofread 5P-NP cells and 250 proofread 5P-ET cells were added to train the excitatory subclass model.

For each classifier, model type was chosen using a randomized grid search for the following models: Support Vector Machine SVM with a linear kernel, SVM with a radial basis function kernel, Nearest Neighbors, Random Forest Classifier, Decision Tree and Neural Network. For each type, 50 models were trained with varying parameters and the top performing model was chosen. Individual models were further optimized using 10-fold cross validation evaluated based on accuracy and F1 score (a measure for precision and recall). Training, and test examples were held consistent across models for direct performance comparison within each level.

Model Performance and Validation:

The hierarchical model was defined as the sequential combination of the best performing classifiers at each level. To see the performance of the all different feature sets at each level of the hierarchy please see Extended Data Table 1. The overall performance of the hierarchical model

was measured with a test set that involved manual inspection of 100 examples of each of the neuronal and non-neuronal subclasses as well as errors. This resulted in a test set of 1700 cells. Cross validation and test performance for the hierarchical model are reported below (Extended Data Fig.4). Note that all scores reported are the weighted accuracy based on the sampling rate of each class within the column.

The top level of the hierarchy (the object model), distinguished neurons from non-neurons as well as erroneous detections. The cross validated accuracy score on the column was 96% with a test score of 97%. The second level of the model simply distinguished excitatory from inhibitory neurons. Here, the column cross validated accuracy score was 94% and the test set was 93%. Overall, across all subclasses, the hierarchical model on the column had a cross validated accuracy of 91% and a dataset wide test set accuracy of 82%.

Chandelier Cell Identification

Chandelier cells are characterized by their unique axo-axonal synapses onto the AIS of target pyramidal cells. As there were no chandelier cells within the densely reconstructed column, we sought to test if the perisomatic feature space would facilitate an enriched dataset wide search for these cells. After identifying and proofreading a chandelier cell, we selected the top 20 nearest neighbors by euclidean distance using a KDTree search of the perisomatic feature space (nucleus, soma, and PSS features) after z-score normalization of each feature across cells. We also selected 20 random cells from the predicted inhibitory neurons. For each of these 40 cells, we proofread the reconstructions to ensure that there were no extraneous neurites attached and extended the axon until there were at minimum 100 output synapses. On average the 20 nearest neighbors had 590 output synapses attached and the random cells had 809 synapses attached.

To quantify whether a given cell was a chandelier or not, we measured the angle (ϕ) and the distance (r) between every output synapse and the soma of the postsynaptic cell (Fig. 6c). A synapse with an angle value of 0° would be considered straight above the target soma whereas an angle of 180° would be right below. Due to variations in axon directionality with respect to the pial surface, we determined that synapses with angle values between 160-180° and within 60 microns of the soma were considered on the AIS of the target soma. In fact, because the specificity of chandelier targeting is so high, the density of synapse angle distributions alone was enough to identify other chandelier cells (Fig. 6e). Upon inspection of the proofread 20 nearest neighbors, we determined that cells with over 40% of their synapses within 160-180° were chandelier cells. The average normalized density for the identified cells was 62% as compared to 8% for the non chandelier cells. A two-tailed Fisher Exact Test was performed to test significance between the random cell population and the nearest neighbors.

Inhibitory Neuron Output Targeting

After characterizing a single 5P-NP targeting cell, we applied a similar strategy to the one above to search for more neurons in the dataset that had a similar connectivity pattern. We selected the top 20 nearest neighbors by euclidean distance in the perisomatic feature space using KDTree search. These cells were proofread to remove false mergers and extend the axon to include at minimum 100 synapses. It should be noted that there were 5 cells where the axons could not be extended due to volume boundaries or segmentation errors so they were replaced with the 5 nearest cells. On average the 20 nearest neighbors had 448 synapses attached.

To quantify whether a cell preferentially targeted 5P-NP neurons, we measured the fraction of total output that targeted different predicted subclasses. Cells that output over 30% of their synapses onto 5P-NP cells were considered to have this rare connectivity preference. A two-tailed

Fisher Exact Test was performed to test significance between the random cell population and the nearest neighbors.

Predicted Subclass Densities

To measure the predicted cell densities per subclass across the MICrONS dataset, we divided the dataset into 50μm^2 bins in the XZ plane. Within each bin we calculated the number of cells for each subclass and scaled that to a mm^2 to facilitate direct comparisons to reported densities in the literature.

Dataset 2

The second dataset covers a millimeter square cross-sectional area, and 50 microns of depth within the primary visual cortex of a P49 male mouse.[18,20,53] The largest available segmentation spans Layer 2/3 of the cortex through to Layer 6. After applying the nuclear detection model[18] and filtering out all nuclear objects below 25μm^3 and cells that were cut off by the volume border (see Filtering procedure above), 1,944 cells were used for the analysis. Class type of each cell was labeled manually and used as ground truth. Due to thinness of the volume, much of the distal cell morphologies were cut off and thus subclass type labeling was not possible. Nuclear and somatic mesh cleaning as well as feature extraction and normalization followed the same procedures outlined above.

Methods References

58. Dorkenwald, S. *et al.* CAVE: Connectome Annotation Versioning Engine. 2023.07.26.550598 Preprint at https://doi.org/10.1101/2023.07.26.550598 (2023).

59. Dorkenwald, S. *et al.* FlyWire: Online community for whole-brain connectomics. *Nat. Methods* **19**, 119–128 (2022).

60. Maitin-Shepard, J. google/neuroglancer: (2021) doi:10.5281/zenodo.5573293.

61. Ilker O. Yaz and Sebastien Loriot. CGAL 5.4.1 - Triangulated Surface Mesh Segmentation: User Manual. https://doc.cgal.org/latest/Surface_mesh_segmentation/index.html.

62. Qi, C. R., Su, H., Mo, K. & Guibas, L. J. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. Preprint at https://doi.org/10.48550/arXiv.1612.00593 (2017).

63. Fei-Fei, L. & Perona, P. A Bayesian hierarchical model for learning natural scene categories. in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* vol. 2 524–531 vol. 2 (2005).

Acknowledgements and Support

We thank the Allen Institute for Brain Science founder, P. G. Allen, for his vision, encouragement and support. HSS acknowledges support from the National Institutes of Health (NIH) BRAIN Initiative RF1 MH129268, U24 NS126935, and RF1 MH123400, as well as assistance from Google. FC, NMdC and CR acknowledge support from NIH RF1 MH125932 and from NSF NeuroNex 2 award 2014862. LE acknowledges support from NIH RF1 MH117808. This work was supported by the Intelligence Advanced Research Projects Activity via Department of Interior/Interior Business Center contract numbers D16PC00004, D16PC0005, 2017-17032700004-005 and 2020-20081800401-023. The US Government is authorized to reproduce

and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Intelligence Advanced Research Projects Activity, ODNI, Department of Interior/Interior Business Center or the US Government.

Author Contributions

LE, SS, and FC extracted the perisomatic features and analyzed all the data presented above. SM, GM, LE, and FC developed the nucleus detection model used to identify cells in the dataset. AB, NMdC, CMSM, JB, LE and FC provided expert labels on the cortical column dataset and the model validation sets. LE, SS, and FC created all the figures and wrote the manuscript. AH, ALB, BN, CJ, CMSM, DB, DJB, DK, EM, FC, GM, HSS, JAB, JB, JW, KL, KL, LE, MAC, MT, NK, NLT, NMdC, RCR, RL, RT, SCY, SD, SK, SM, SP, SSM, TM, WS, WW, WY, ZJ created the structural MICrONS65 dataset. SD, CMSM, DB, CJ, AH, WS, NK, MC, FC developed the analysis infrastructure referred to as CAVE.

Competing Interest Statement

T. Macrina, K. Lee, S. Popovych, D. Ih, N. Kemnitz, and H. S. Seung declare financial interests in Zetta AI. The remaining authors declare no competing interests.

Additional Information Statement

Reprints and permissions information is available at www.nature.com/reprints.

Data Availability Statement

All datasets described in the manuscript are publicly available at https://microns-explorer.org/ and https://bossdb.org/. The cell-type predictions presented in the manuscript are made available as an open-access data tool accompanying the MICrONS dataset

Code Availability Statement

Code developed for the nucleus and soma feature extraction pipeline can be found at:

https://github.com/AllenInstitute/Extract_Somatic_Features.

Code developed for the spine and PSS extraction pipeline can be found at: https://github.com/AllenInstitute/featureExtractionParty/, and

https://github.com/AllenInstitute/pss_extraction_pipeline.

Code for the analyses presented in the manuscript can be found at: https://github.com/AllenInstitute/Perisomatic_Based_CellTyping.

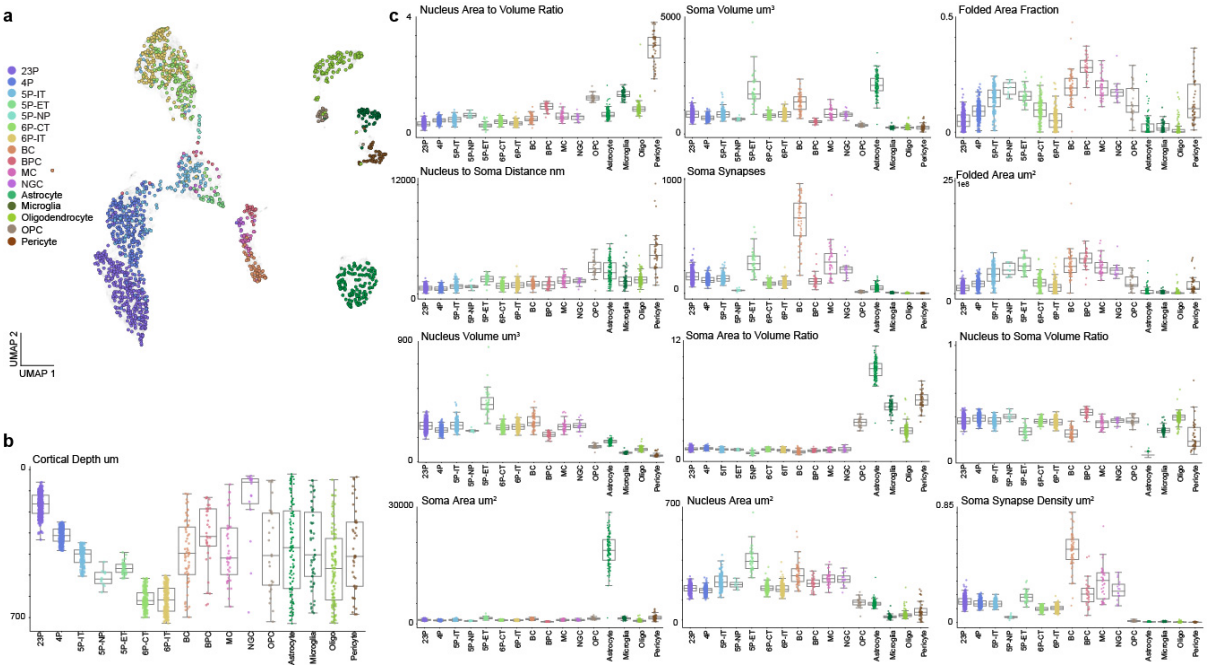
Extended Data

	Nucleus Column Cross-Validation Accuracy	Soma + Nucleus Column Cross-Validation Accuracy	PSS + Soma + Nucleus Column Cross-Validation Accuracy
1. Object Class	0.90	**0.96**	—
2. Nonneuron Class	0.94	**0.98**	—
3. Neuron Class	0.95	**0.99**	—
4. Excitatory Subclass	0.88	**0.90**	—
5. Inhibitory Subclass	0.63	0.90	**0.94**

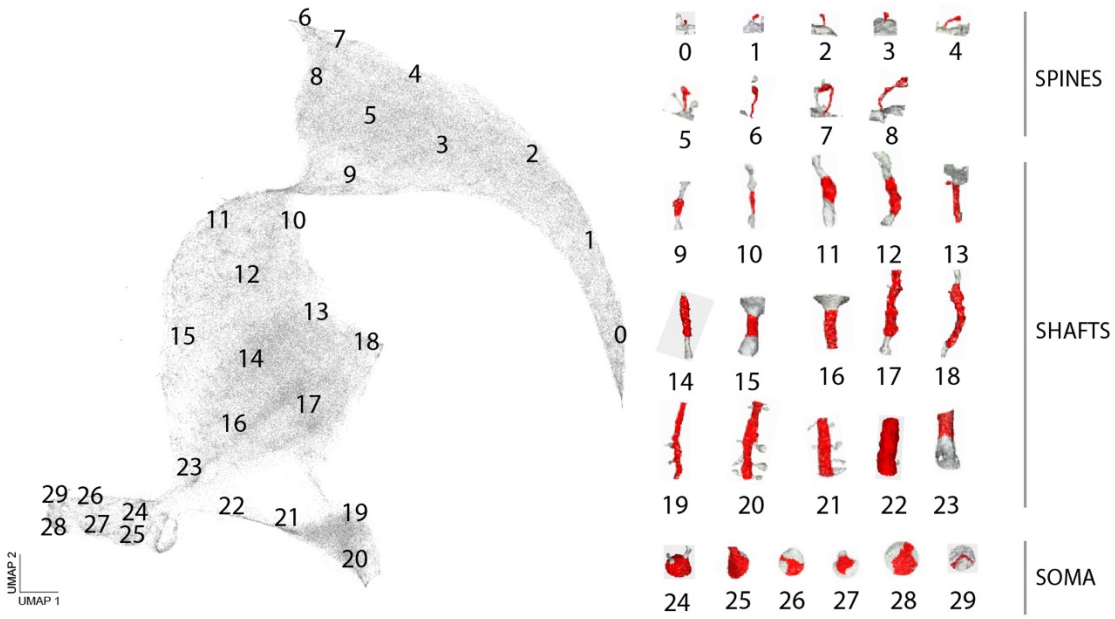
Bolded Classifiers used in the Hierarchical Model

	Column Cross-Validation Accuracy	Dataset Wide Validation Accuracy
Hierarchical Model	**0.91**	**0.82**

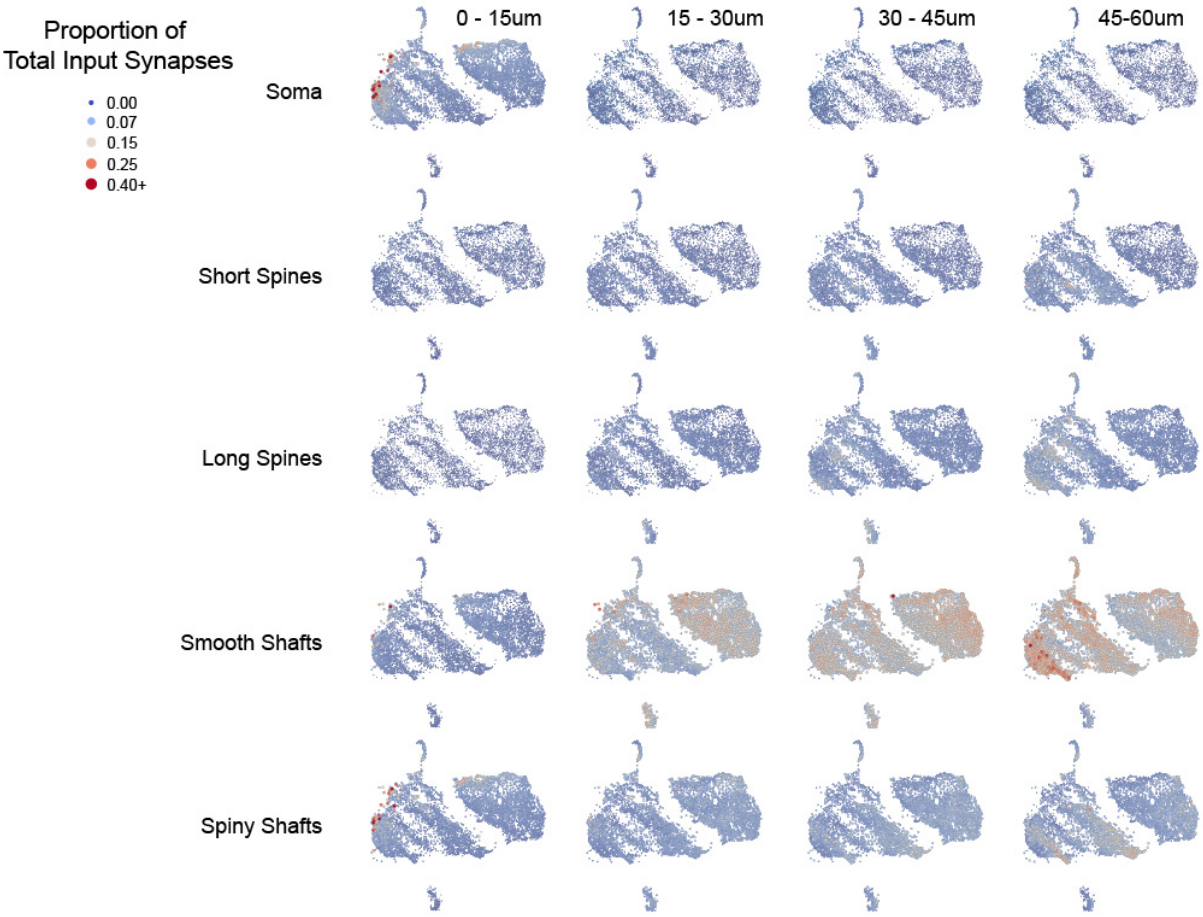
Extended Data Table 1: Cross validation accuracy scores for individual classifiers at each level of the hierarchical model with differing input features. Each row corresponds to the corresponding numbers in the diagram in Fig. 5A. All training examples were held consistent between features sets for appropriate model comparisons. Classifiers with the highest accuracy score at each level were included in the hierarchical model (shown in bold). The overall hierarchical model performance on the column and the dataset wide validation set (see methods) is reported at the bottom.



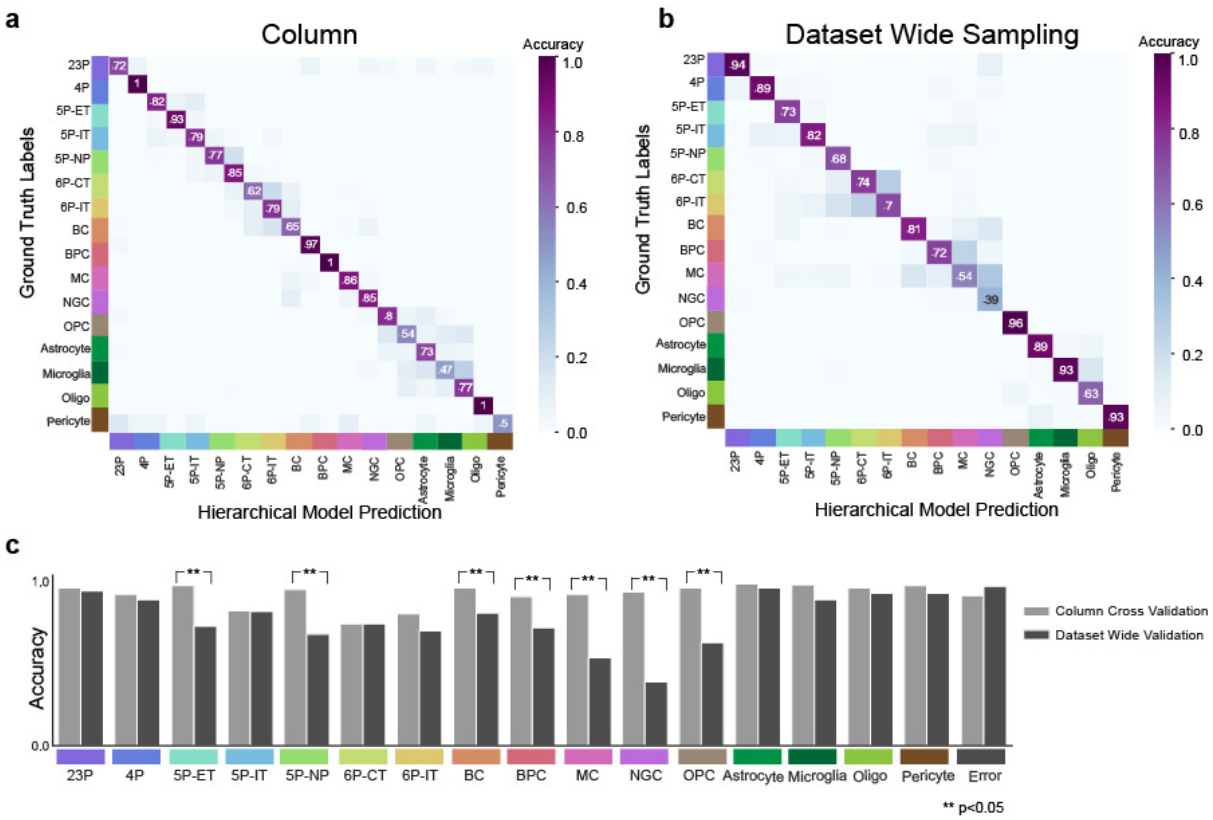
Extended Data Figure 1: Neuronal and non-neuronal subclass distribution of individual soma and nucleus features. a) 2D UMAP embedding of all neuronal and non-neuronal cells inferred from somatic features, nuclear features and cortical depth. Manually labeled cellular subclasses are represented in color (1,619) and unlabeled examples in light gray (n=92,391). b) Distribution and variation of cortical depth of all cells from the manually labeled column dataset. c) Distribution and variation of nucleus and somatic features of all cells from the column dataset. For all box plots, center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range, outliers not shown. Individual cells, including outliers, are shown in the overlaid swarm plots. Color denotes human assigned subclass labels.



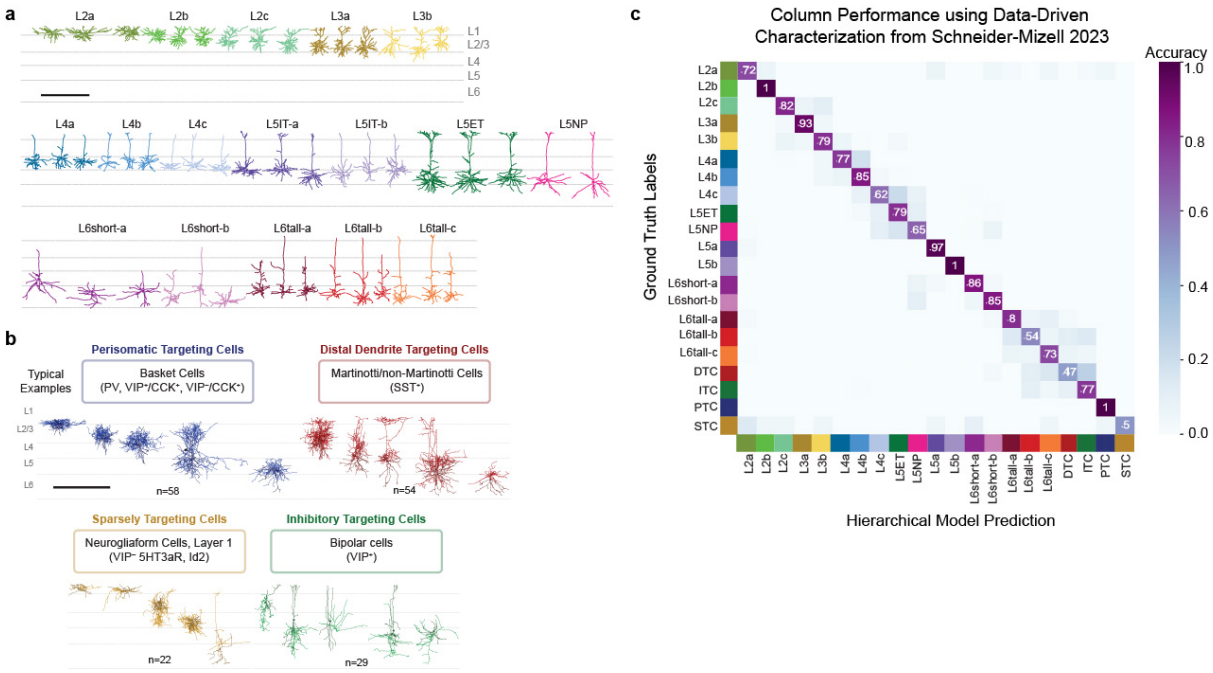
Extended Data Figure 2: PSS embedding space organized by postsynaptic ultrastructural morphologies. 2D UMAP embedding of all shapes in the PSS Dictionary. The numbers indicate the bin centers mapped in this 2D space and the corresponding PSS meshes on the right show the shape associated with each bin center. Bins 1-8 range in spine shapes, Bins 9-23 are shaft shapes and Bins 24-29 are soma shapes.



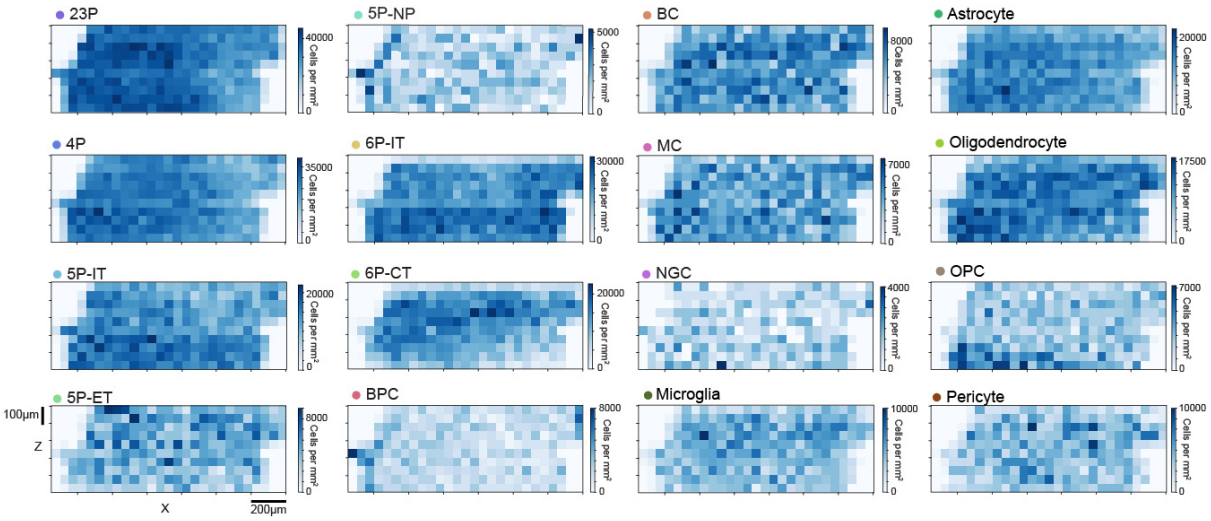
Extended Data Figure 3: Inhibitory neuron subclasses exhibit spatial patterns to PSS distributions. The UMAP embedding of all the perisomatic features, including PSS features, across all inhibitory cells, colored with respect to what fraction of that cell's input (within the 60µm cutout) comes from what PSS/distance bin. PSS shape bins were simplified from 29 bins to 5 broad categories to simplify the visualization (bins 0-4: short spines, 5-8: long-spines, 9-18+23: smooth shafts, 19-22: spiny shafts, 24-29: soma). This visualization gives insight into how different cells in different parts of this embedding space receive varying amounts of input onto different shapes within different spatial zones of the perisomatic area. Cells on the far left hand side of the embedding, where in general bipolar type neurons were found, have larger fractions of their inputs near the soma, including dendritic shafts which are more irregular in shape ("spiny shafts"), and smooth shaft inputs farther away where the dendrites begin to elaborate. Basket cells on the right hand side of the side of the embedding are dominated by somatic inputs and smooth shaft inputs which are more evenly distributed spatially. The island at the bottom that is dominated by neurogliaform cells is characterized by having relatively fewer somatic inputs, but an increasing amount of shaft and spiny input at distal dendrites.



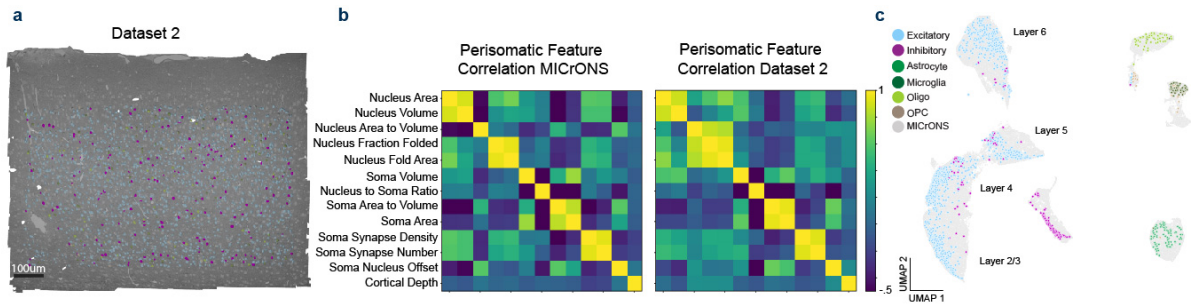
Extended Data Figure 4: Classifier validation. a) Confusion matrix of hierarchical model performance for all cells within the manually labeled column after training. **b)** Confusion matrix of hierarchical model performance on a dataset wide sample of 100 cell predictions from each subclass. **c)** Comparison of column cross validation vs. dataset wide model performance, asterisk notes significance by two-tailed Fisher Exact Test.



Extended Data Figure 5: Cell densities across the dataset by cellular subclass. Predicted cell densities per mm^2 for each subclass across the entire dataset in the XZ plane. Each square represents 50 $micron^2$ and color denotes the density scaled per mm^2. Note due to the approximate 1 mm depth of cortex, these values are also roughly densities per mm^3. They roughly agree with densities of cells estimated from light microscopy stereology of subclasses,[26] usually utilizing histochemical markers or genetic tools. For some subclasses, there is not a 1-1 to alignment between the definitions of types in this study and the usual molecular markers used in those studies, as molecular markers are not directly measurable in this electron microscopy volume.



Extended Data Figure 6: Perisomatic feature based classification utilized with different cell-type labels. a) Alternative excitatory subclass labels in the column from Schneider-Mizell et al 2023, based on unsupervised clustering of dendritic and synaptic features rather than manual human expert calls. Labels on the clusters were inferred based on the overlap with expert labels and cortical depth, with finer distinctions added when necessary (i.e. L4a, L4b, L4c). **b)** Alternative inhibitory subclass labels from Schneider-Mizell et al. 2023 in the column based on unsupervised clustering of their output connectivity statistics. These subclasses (Perisomatic targeting, Distal Targeting, Sparsely Targeting and Inhibitory Targeting) likely largely but not completely align with broad molecular distinctions made amongst inhibitory cells, based on reviews of the literature where molecular and output connectivity has been measured in the same cells. **c)** A confusion matrix of a hierarchical model retrained to utilize these subclass labels for excitatory neurons vs inhibitory neurons rather than human expert labels. Cross validation performance on the excitatory (67%) and inhibitory (85%) subclass models was lower than the expert labels, due primarily to the fine grained distinctions made amongst layer 4 and 6 types. The confusion matrix shown here is the output of the final model trained on all samples from the column.



Extended Data Figure 7: Basic perisomatic feature patterns maintained across a second dataset from a different animal. a) A cutout of a second dataset, which covers layer 2/3 to 6 of cortex, but is only 50µm thick. Somas contained within this volume (n=1,944) were analyzed in a manner identical to the larger dataset and soma, nucleus and PSS features were extracted. Excitatory nuclei highlighted in light blue and inhibitory nuclei in magenta. **b)** Feature to feature Pearson correlations exhibit similar correlation structure between the two datasets. **c)** A joint UMAP of the perisomatic features with the MICrONS dataset data shown in gray, and the smaller dataset covered by manually identified cell classes overlaid. In general, the same overall patterns and degree of separation amongst layers and cell classes was observed. Note: pericytes were manually excluded from this dataset due to the lower quality of nucleus and somatic segmentations. Extensive detailed subclass cell type validation is not possible in this dataset due to the truncation of axons and dendrites.

Sensorimotor circuits transform a somatotopic map of the fly leg into targeted grooming actions

Leila Elabbady[1,2], Grant Chou[2], Anne Sustar[2], Andrew Cook[2], Forrest Collman[3]*, John C. Tuthill[2]*

[1]Neuroscience Graduate Program, University of Washington, Seattle, WA, USA
[2]Department of Neurobiology and Biophysics, University of Washington, Seattle, WA, USA
[3]Allen Institute for Brain Science, Seattle, WA, USA
* Co-senior, corresponding authors

Animals continuously monitor their body surfaces to detect and remove debris or parasites. Effective grooming requires that tactile inputs from specific body regions be transformed into precisely targeted motor actions, but the neural circuits that support this sensorimotor transformation remain poorly understood. Here, we combine genetic tools and connectomics to elucidate a central somatotopic map of the *Drosophila* leg. We show that the axonal projections of leg touch receptors within the fly's ventral nerve cord (VNC) are organized along the same cardinal axes as the developing leg. Somatotopically-organized bristle axons target a specific class of developmentally-related local interneurons, which imbricate the leg map with overlapping receptive fields of different shapes and sizes. These second-order interneurons target distinct pools of premotor interneurons, which in turn synapse directly onto motor neurons that control leg muscles. Optogenetic activation of second-order interneurons elicits spatially targeted grooming of specific leg regions, consistent with our spatial receptive field predictions based on the connectome. Together, our results reveal a four-layer circuit that transforms a somatotopic map of the fly leg into spatially targeted grooming behaviors.

Introduction

Humans and other animals must constantly monitor the surface of the body to detect and remove unwelcome intrusions. A fly landing on a person's knee may deflect a hair, which triggers

tactile sensory neurons to fire. These signals are transmitted into the spinal cord, where they are transformed across layers of interneurons into patterns of spikes in motor neurons, which move a hand to scratch the leg. Studies in cats and turtles have demonstrated that these animals adapt their scratching movements to reach the site of stimulation[1–3]. This suggests that central circuits are organized to elicit targeted movements in response to activation of specific touch receptors. However, the complexity of vertebrate tactile circuits and the sparseness of previous tracing methods have made it challenging to understand how sensorimotor circuits transform sensory signals into spatially targeted grooming behaviors.

A common organizational structure found in early sensory circuits, which may help to simplify such sensorimotor computations, is the topographic map[4]. The axons of tactile sensory neurons from specific parts of the body often project into specific regions of the nervous system, and axons from neighboring regions may exhibit similar morphology and connectivity. In some cases, these sensory maps are preserved in downstream circuits, as in the mammalian somatosensory and visual cortices[5–8]. Understanding the structure of sensory maps is an important prerequisite for deciphering how patterns of sensory neuron activity are transformed into precise motor actions.

In insects, the sense of touch is mediated by tactile bristles distributed across the body[9–11]. Each bristle is innervated by a single mechanosensory neuron, which fires action potentials when the bristle is deflected by external forces (Fig. 1A). Bristles are extremely sensitive, responding to deflections less than 100 nm[12]. Insects rely on bristles to detect external objects in the environment or debris on their bodies. In *Drosophila*, mechanical or optogenetic stimulation of tactile bristles elicits avoidance reflexes and/or spatially targeted grooming at the site of stimulation[11,13–17]. Some of these behaviors are maintained in headless flies, suggesting that the basic circuitry for spatially

targeted grooming is contained within the fly ventral nerve cord (VNC), the invertebrate analog of the spinal cord[13,16,18]. Fly grooming is modular and hierarchical: a dirty fly will first clean its eyes and head before proceeding to more posterior body regions like the thorax and abdomen[15,19–22]. Neurons that elicit certain grooming modules (e.g., head, wings, antenna) have been identified[15,23], but less is known about the neural mechanisms that underlie spatial targeting of grooming movements within a module.

Axons from leg bristles project into the VNC which, like the spinal cord, is organized into neuropil compartments that sense and control specific body parts, including the legs, wings, thorax, and abdomen[11,24–26]. Past work using dye fills of single bristle neurons has revealed that their axons are stereotyped across individuals and suggested the existence of a topographic map within the leg neuropil[24,25,27,28]. However, because each leg has many hundred bristles, the precise organization of the leg map in the VNC remains unknown. Electrophysiological recordings have identified a subset of VNC interneurons that integrate signals from multiple bristle neurons[11]. Yet the circuits that integrate leg bristle signals and transform them into spatially targeted motor commands remain poorly understood.

Advancements in high throughput electron microscopy and automated image segmentation have resulted in the collection of large volumetric datasets that enable comprehensive cell reconstruction and synapse identification. These datasets, referred to as connectomes, enable the study of structural wiring diagrams and how circuit architecture may facilitate the function. Although there exist multiple connectome datasets of the *Drosophila* brain and VNC[29–33], it remains a challenge to link these connectomes to the fly's body and peripheral nervous system.

Here, we use a connectome dataset of the *Drosophila* Female Adult Nerve Cord (FANC)[31,34,35] to investigate how tactile information is mapped in the VNC, from the sensory

neurons in the leg through the VNC to the motor neurons that innervate specific leg muscles. We first combined genetic and connectomic tools to elucidate the central somatotopic map of the fly leg. We found that the spatial map of bristle axons in the VNC matches the somatotopic organization of the larval imaginal disc from which the leg develops. We then reconstructed and analyzed how populations of VNC interneurons sample the leg tactile map. Our results reveal a four-layer neural architecture, from leg bristles to motor neurons. Second-order neurons imbricate the leg map into overlapping receptive fields. These second-order neurons target distinct pools of third-order neurons which then target leg motor neurons. Optogenetic activation of second-order interneurons from different regions of the map drove spatially targeted grooming of specific leg regions, consistent with our receptive field predictions from the connectome. Overall, our results elucidate the organization of central circuits in the fly VNC that transform peripheral tactile signals into spatially-targeted behaviors.

Results

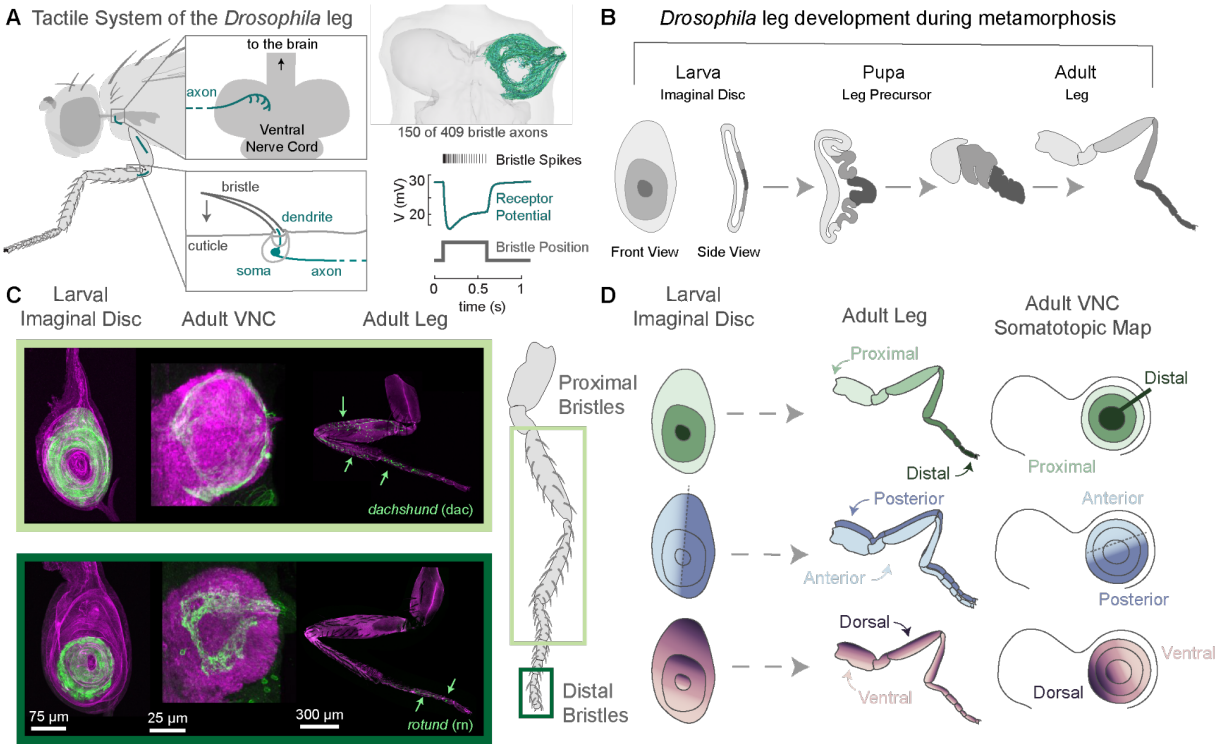


Figure 1: Somatotopy of the leg is maintained in the VNC and recapitulates the larval leg imaginal disc. A) A bristle neuron is located at the base of each sensory hair on the leg. The dendrite is stretched upon deflection of the hair (bottom left). Bristle axons project to the ventral nerve cord (VNC) (top left). We reconstructed 409 bristle axons from the front left leg of an adult female fly (top right). **B)** The larval leg imaginal disc develops into the adult leg. **C)** Bristle neurons that express the proximal leg precursor *dachshund* (dac) during development (top). Bristle neurons that express a distal leg precursor *apterous* (ap) during development (bottom). Confocal images show maximum intensity projections of cells in the larval leg imaginal pJFRC7-20XUAS-IVS-mCD8::GFP (green) and an antibody against phalloidin (magenta). Bristle neurons in the leg and VNC were labeled with mcd8::GFP (green) and an antibody against the neuropil marker bruchpilot (magenta). **D)** The somatotopic map of the leg in the VNC recapitulates the somatotopic map of the leg in the larval imaginal disc during development. The proximal to distal axis is mapped along the peripheral to central axis (top). The anterior leg maps to the anterior portion of the VNC leg neuropil and the posterior leg maps onto the posterior leg neuropil (middle). The dorsal leg maps to the area intersecting the anterior to posterior border, while the ventral leg corresponds to axons that remain within either the anterior or posterior region (bottom).

Leg somatotopy in the VNC recapitulates imaginal disc somatotopy

The front leg of *Drosophila melanogaster* is covered by more than 400 mechanosensory bristles, with the highest density on the more distal leg segments[9]. To understand how tactile information from the leg is mapped in the VNC, we reconstructed 409 bristle axons from the left front leg in a volumetric electron microscopy dataset of a *Drosophila* female adult nerve cord

(FANC)[31,34,35]. We identified bristle axons based on their morphology and projection patterns into the left front leg neuromere – the region of neuropil corresponding to the left front leg (see Methods). As a population, bristle axons fan out to cover the ventral surface of the VNC; however, each bristle axon innervates a small region within the VNC neuropil. Bristle axons exhibit a range of morphologies (Extended Data Fig.1). While most axons terminate within the same region of the neuropil (e.g. anterior or posterior) there are a subset of axons that branch across the midline in the shape of a hockey stick (Extended Data Fig. 1). Across the population, axons with similar morphologies project to similar locations within the VNC neuropil. This structure motivated us to determine the relationship between the location of bristles on the leg and their axonal projections into the VNC.

We developed a genetic strategy to label bristles on specific sections of the leg by restricting the expression of a bristle GAL4 line with transcription factors expressed during development. During metamorphosis, each fly leg develops from an imaginal disc that expresses specific transcription factors defining the three cardinal leg axes (anterior/posterior (A/P), dorsal/ventral (D/V), proximal/distal (P/D) (Figure 1)[36–39]. We used a recombinase driven by different transcription factors to turn on Gal4 expression, thus labeling bristle cell bodies on the leg and their axons in the VNC. For example, bristle neurons that express *dachshund (dac)* during development end up in the proximal leg and project their axons to the outer edges of the VNC neuropil (Fig. 1C). Distal leg bristles are labeled by *apterous (ap)* or *rotund (rn)*, and their axons project into the center of the neuromere (Fig. 1C, Extended Data Fig. 2). Thus, we concluded that the proximal-distal axis of the leg is mapped in concentric rings around the VNC neuropil, with distal bristles at the center and proximal bristles along the outer edges (Fig. 1C, Extended Data Fig. 2).

In a similar manner, using the transcription factor *hedgehog*, we found that the bristle cell bodies on the anterior leg project their axons into the anterior VNC, while posterior bristles project their axons to the posterior neuropil (Extended Data Fig. 2). Cells in the dorsal leg that express *decapentaplegic* (dpp) extend their axons to cross the A/P border. Ventral leg bristle neurons that express *midline* (mid) do not cross this border (Extended Data Fig. 2). In other words, axons that enter the VNC anteriorly terminate anteriorly and vice versa. Notably, the mapping of bristle axons in the VNC recapitulates the leg map in the larval imaginal disc[37]. Similar to the AP compartment boundary in the imaginal disc[40,41], the A/P axis in the VNC is defined by a stark branch point, at which each axon projects either anterior or posterior. The D/V and P/D axes are more gradual, so we defined the position of each axon along a gradient relative to the population (Fig. 1D). The striking similarity between the leg maps in the VNC and the imaginal disc suggests that the adult leg develops in coordination with the postembryonic restructuring of the VNC neuropil.

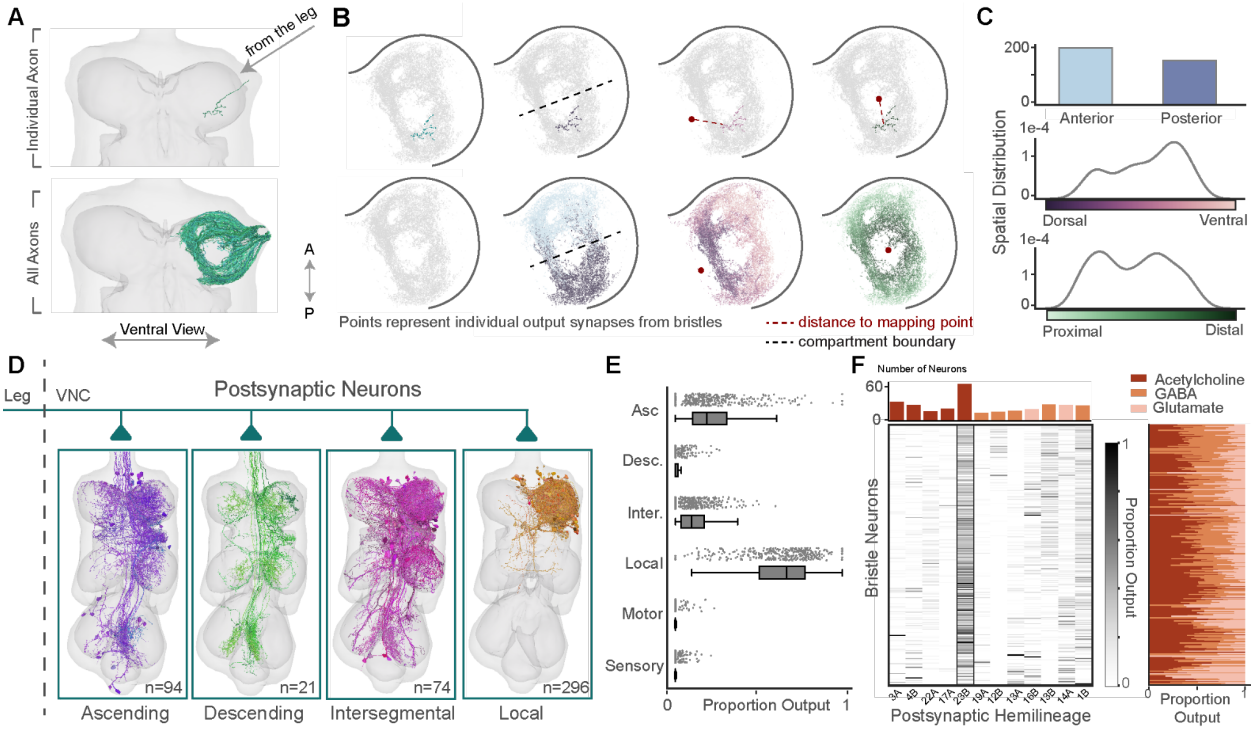


Figure 2: Bristle neurons across the leg preferentially target local 23B neurons in the VNC. A) A single bristle axon from the left front leg (top), out of a population of 409 bristle axons reconstructed from the FANC EM

dataset, including axons from left front leg nerve, VProN, and DProN[42] (150/409 axons shown for clarity in the bottom panel). **B)** Output synapses from the single bristle axon shown in panel **A** that arches posteriorly to the AP compartment boundary, colored by the average synapse distance (d) from the DV and PD mapping points respectively (see Methods) (top). Output synapses from all the reconstructed axons colored by their anterior or posterior annotation and the average synapse distance for each individual axon along the DV and PD axes (bottom). **C)** Predicted distribution of bristle axons along the three spatial axes. **D)** Top classes of postsynaptic partners to bristle axons: Ascending n=9, Descending n=21, Intersegmental n=74, and Local n=296. **E)** Proportion output for each bristle axon onto all classes of postsynaptic partners. **F)** Proportion output for each bristle axon (rows) onto VNC neurons from different developmental hemilineages (columns) (heatmap). Number of unique cells of each hemilineage across the postsynaptic partner population (bar chart top). Proportion output for each bristle axon (rows) onto postsynaptic partners that release acetylcholine, glutamate, or GABA. Neurotransmitter type assigned based on hemilineage classification for each postsynaptic partner(stacked bar chart, righ[43]. For all box plots, center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; outliers not shown.

A predicted axon map recapitulates the distribution of bristles along the leg

Based on the leg map defined above, we developed three mapping rules to predict the peripheral location of bristle axons in the FANC connectome. We defined bristles from the anterior portion of the leg as the axons that arch anteriorly upon entering the VNC from the leg nerves, whereas cells located on the posterior leg arch posteriorly in the VNC. To recapitulate the graded distribution along the PD axis (Fig. 1C-D), we placed a mapping point in the center of the left leg neuropil and calculated the average synaptic distance between each axon and the center mapping point (Fig. 2A-B top row, see Methods). Axons that were closer to this mapping point were estimated to be more distal on the leg compared to those further from the mapping point. We used a similar approach for the DV axis with a different mapping point to more accurately represent the pattern described in Figure 1 (see Methods). The spatial predictions for each leg bristle qualitatively matched the patterns observed in genetic labelling experiments (Fig. 2B, bottom row). They also recapitulated the expected nonuniform anatomical distribution of bristles along the leg (Fig. 2C)[44].

Bristle axons target local excitatory neurons from the 23B hemilineage

We next used the connectome to analyze the connectivity between bristle axons and downstream neurons in the VNC. Based on automated synapse predictions[31], each bristle axon

makes on average 550 output synapses in the VNC and receives on average 77 input synapses (Extended Data Fig. 3). VNC neurons downstream of bristle axons are divided into five broad morphological classes: ascending, descending, intersegmental, local, and motor neurons (Fig. 2D, see Methods). On average, local neurons receive the largest proportion of bristle synapses (63%), followed by ascending (22%) and intersegmental neurons (12%) (Fig. 2E). Only ~1% of bristle synapses are onto other sensory neurons. Descending neurons receive less than 2% of bristle synapses, and most bristles make zero synapses onto motor neurons (Fig. 2E).

Most neurons in the VNC develop from 33 postembryonic stem cell hemilineages. Cells from the same developmental hemilineage share broad morphological features, typically release the same neurotransmitter[43,45,46], and may perform similar functions[47]. Using morphological criteria, we classified the developmental hemilineage of each VNC neuron that received input from leg bristles (see Methods). The strongest downstream targets of bristle axons are neurons from hemilineage 23B (Fig. 2F). 23B interneurons receive on average 25% of each bristle axon's synaptic output (Fig. 2F, heatmap). Not only are 23B neurons the strongest downstream target, but cells from this hemilineage are the most frequent postsynaptic target of leg bristles (61 cells; Fig. 2F, bar chart). Thus, we hypothesized that 23B neurons, as a population, represent a map of the leg and that individual 23B neurons integrate tactile signals from specific regions of this somatotopic map.

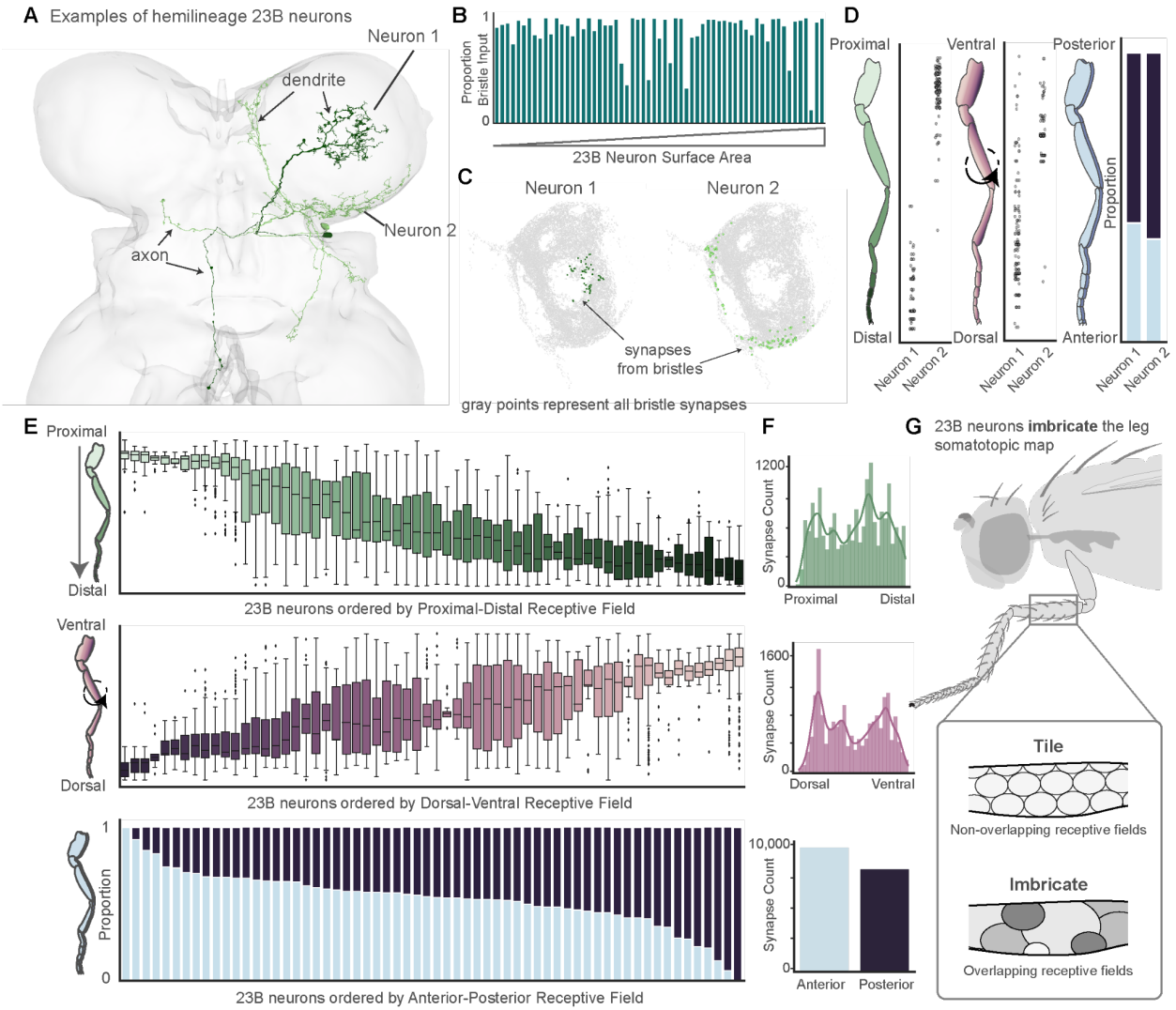


Figure 3: 23B neurons imbricate the leg map in distinct overlapping receptive fields. A) Two example 23B neurons highlighted in different colors. Dendritic and axonal segments denoted by the arrows. **B)** Proportion of all sensory input from bristle axons for each 23B neuron, bars ordered by surface area. **C)** Input synapses from bristle axons onto Neuron 1 (brown, n=76) and Neuron 2 (orange, n=136) as compared to all the output synapses from bristle axons (gray). **D)** Receptive field predictions for example Neuron 1 and Neuron 2. Receptive field for the PD axis (left), DV axis (middle), and AP axis (right). **E)** Receptive fields along the PD axis (top), DV axis (middle), AP axis (bottom) for each individual 23B neuron. Individual points represent input synapses from bristle axons and the y axis represents where on the leg each presynaptic bristle axon originates. **F)** Number of bristle input synapses onto all 23B neurons from different areas of the leg along the three spatial axes. For all box plots, center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; outliers not shown. **G)** 23B neuron receptive fields on the leg imbricate the somatotopic map into overlapping receptive fields, as compared to a non-overlapping tiling pattern.

23B neurons are selective for tactile sensory input

Because they are the top postsynaptic partner of leg bristles, we focused our analysis on 23B interneurons, which release the predominantly excitatory neurotransmitter acetylcholine[43]. Of

the 61 23B neurons we reconstructed, 56 are local, meaning that their synaptic inputs are restricted to the front left leg neuromere. Four are intersegmental and receive synaptic inputs from multiple leg neuropils and one has an ascending axon that projects to the brain. All 23B cells have a soma located on the dorsal surface of the VNC and their neurites fasciculate together as they enter the neuropil. All 23B cells possess extensive pre and postsynaptic arbors close to the ventral surface of the neuromere (Fig. 3A). Regardless of size or morphology, 23B neurons receive on average 40% of their total synaptic input from sensory axons, 85% of which comes from bristle axons (Fig. 3B). This suggests that most 23B neurons are specialized for tactile sensing.

23B neurons imbricate the somatotopic map of the fly leg

Despite the fact that all 61 23B neurons receive input from leg bristle axons, the dendritic arbors of each 23B neuron within the front left leg neuromere are highly variable (Fig. 3A). Based on this diversity, we hypothesized that individual 23B neurons receive input from bristle neurons at different locations on the leg. To quantify this location for each 23B neuron we used the somatotopic mapping approach described above (Fig. 2). Each 23B neuron receives input synapses from a selection of bristle axons (Fig. 3A, C). Based on our somatotopic mapping, each bristle axon represents a single location on the leg along the three cardinal axes. Therefore, we represented each bristle input synapse onto a 23B neuron by the location on the leg of the presynaptic bristle axon. We refer to the distribution of input synapses along each axis as the receptive field for each 23B neuron (Fig. 3C-D). Individual receptive fields varied substantially as some neurons received input exclusively from proximal or distal bristle axons (Fig. 3C). Overall, the receptive fields of 23B neurons covered the entire somatotopic map of the leg across all three axes (Fig. 3D-E). Similar to pebbles on a riverbed, 23B neurons *imbricate* the somatotopic leg map by covering the space with overlapping receptive fields of different sizes and shapes (Fig. 3G).

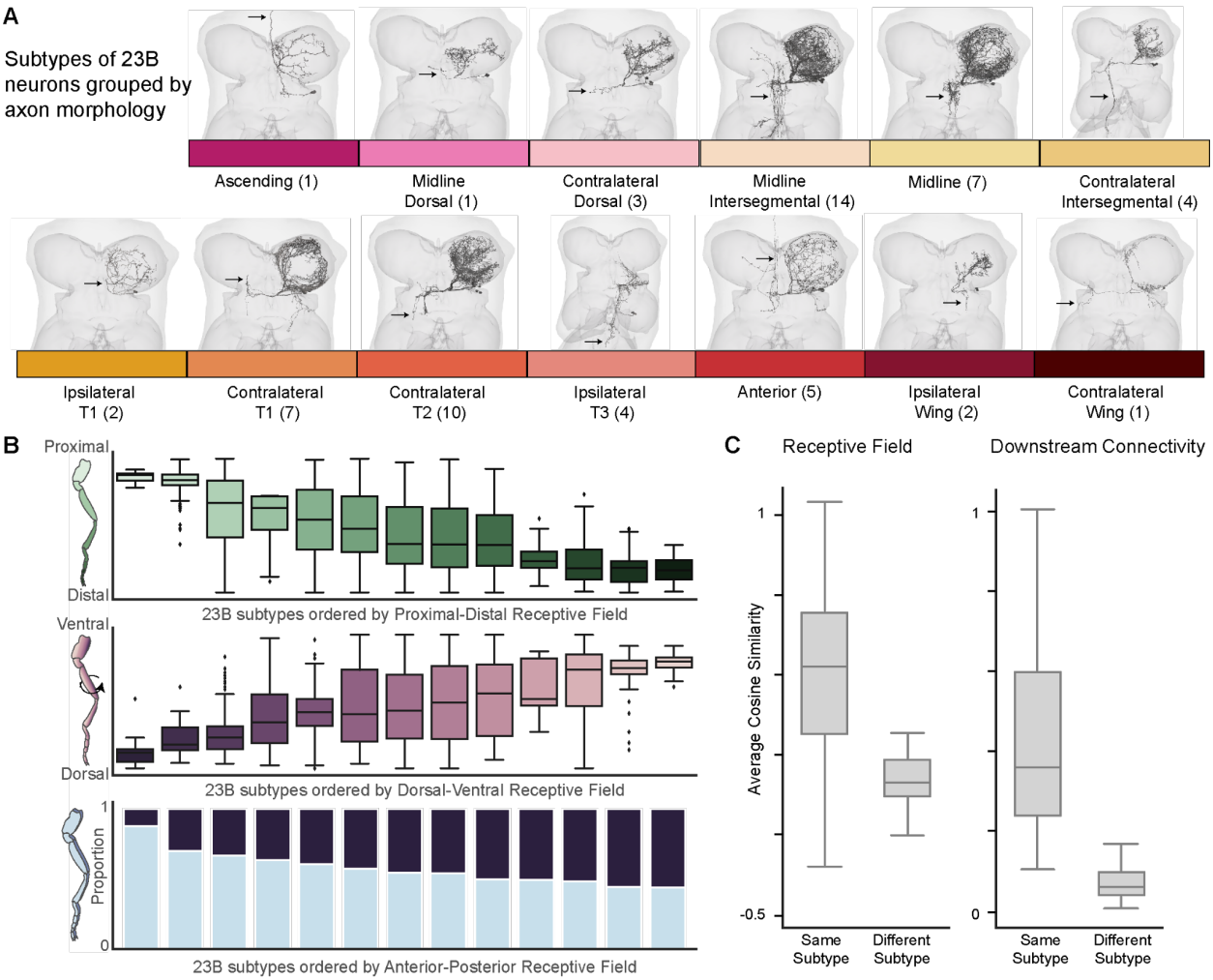


Figure 4: 23B subtypes exhibit similar morphology, receptive fields, and downstream connectivity. A) 23B neuron morphologies organized and labeled by the axonal projection patterns (arrows indicate axon location). Ascending (1), Club (1), Dorsal (3), Midline Intersegmental (14), Midline (7), Contralateral Intersegmental (4), Ipsilateral T1 (2), Contralateral T1 (7), Contralateral T2 (11), Ipsilateral T3 (3), Anterior (5), Ipsilateral Wing (2), and Contralateral Wing (1). **B)** Average receptive fields of 23B subtypes along the three cardinal axes. 23B subtypes ordered by their receptive fields. **C)** Cosine similarity of individual 23B neurons relative to other 23B neurons within and between subtypes according to receptive field (left) and downstream connectivity (right). For all box plots, center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; outliers not shown.

23B neurons organized by axonal projection patterns

While 23B neurons possess similarities in their overall morphology and the proportion of tactile input, they exhibit distinct axonal projection patterns to other regions of the VNC. Moreover, we found that the axons of 23B neurons bundle together in their projections to different target regions. Thus, we reasoned that the distinct axonal morphologies of 23B neurons could be

a useful means to group them into subtypes. Grouping 23B cells by the projection pattern of their axons resulted in 13 subtypes (Fig. 4A). Each subtype had between 1-14 neurons. While they were grouped solely by axonal projection, we noticed that 23B neurons within the same subtype had similar dendritic arbors. To quantify this similarity, we represented each 23B neuron by the mean receptive field value in each of the three cardinal axes and calculated the cosine similarity within and between subtypes. We observed that receptive fields were more similar within than across subtypes (Fig. 4B-C). Furthermore, the downstream connectivity of 23B neurons was more similar within subtypes (Fig. 4C). This is notable considering that synapses on the axonal projections make up only 17% of 23B output synapses. This means that despite the overlap of dendritic arbors within the left front leg neuromere, 23B neurons from different subtypes target distinct postsynaptic neurons. From these similarities in morphology, receptive field, and postsynaptic targeting, we hypothesized that distinct 23B subtypes function as distinct sensorimotor modules.

Testing connectome-derived predictions of 23B neuron receptive fields

We used optogenetics to test the behavioral function of 23B subtypes. We hypothesized that if 23B neurons are specialized for localizing tactile stimuli, the fly's behavioral responses to activating these cells would reflect their spatial receptive fields. We identified two genetic driver lines that specifically label distinct 23B subtypes with contralateral T1 (SS04746) and midline intersegmental (R21B10) neurons (Fig. 5A). We used SPARC to sparsely label the axons of individual 23B neurons in ~20 different VNCs for each genetic driver line (Extended Data Fig. 5). These sparse labeling experiments confirmed that the two driver lines label different subpopulations of 23B neurons (Fig. 5A, Extended Data Fig. 5).

We calculated a connectome-derived receptive field prediction for each genetic driver line. Both SS04746 and R21B10 had six 23B neurons labeled in each neuromere. To predict the

cumulative receptive field of these six neurons, we iteratively sampled six neurons from the connectome weighted by the subtype proportions outlined above (Extended Data Fig. 4, see Methods). For each sampled subset of 23B neurons, we summed the bristle input from these cells to predict the aggregate receptive field for each driver line (see Methods). From these calculations, we predicted that activation of the 23B neurons in SS04746 would correspond to proximal bristle activation and thus elicit proximally targeted grooming. Conversely, the 23B neurons in R21B10 flies received input from distal bristles and thus we hypothesized that activation of these neurons would elicit distal grooming (Fig. 5B). Along the DV axis, we predicted that activating 23B neurons in SS04746 flies would lead to more ventrally targeted grooming compared to 23B neurons in R21B10 flies. Finally, we predicted there would be little to no difference along the AP axis (Fig. 5B). Based on these predictions, we refer to SS04746 as *proximal 23B neurons* and R21B10 as *distal 23B neurons*.

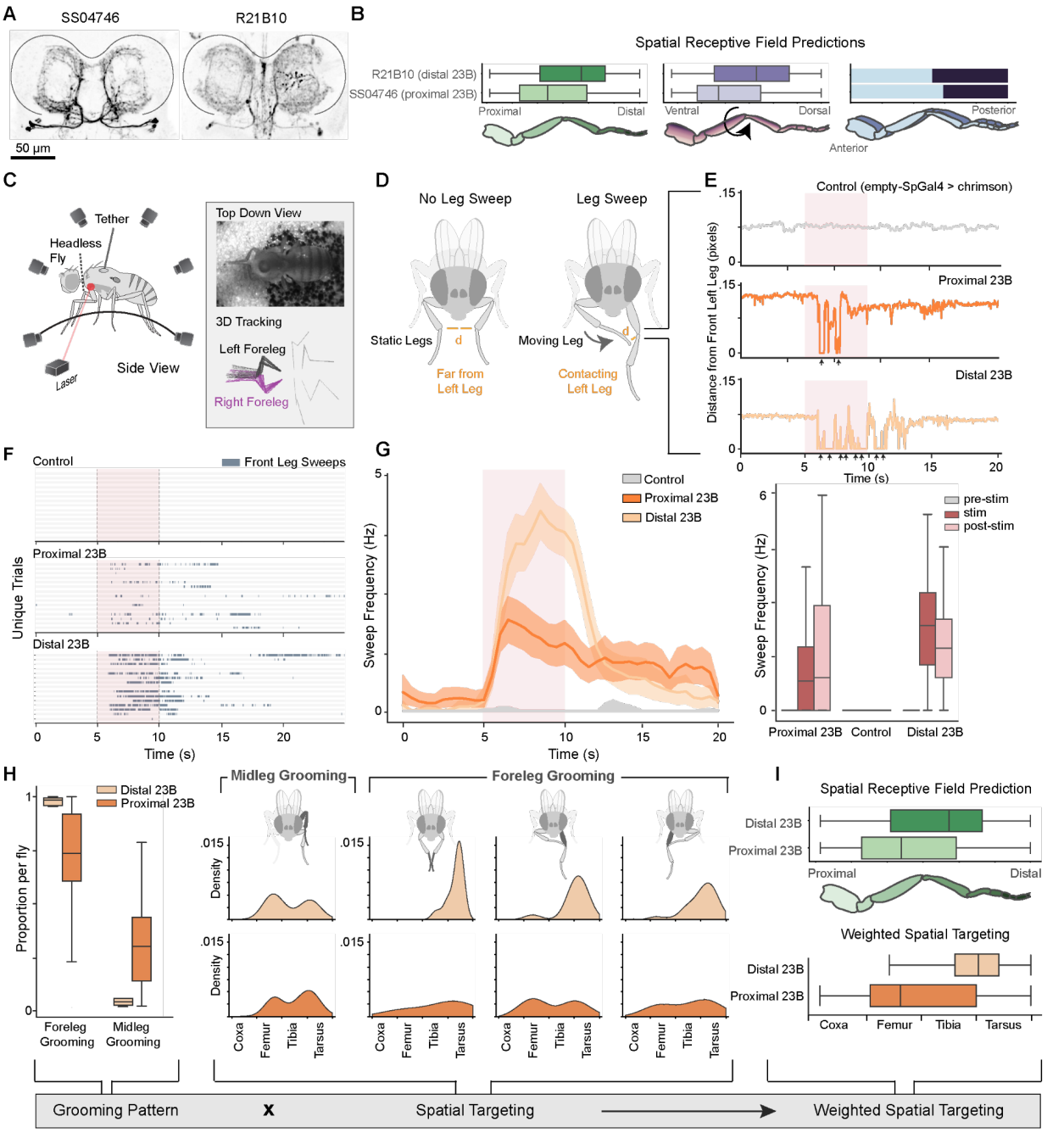


Figure 5: Optogenetic activation of 23B subtypes drives distinct and spatially targeted grooming. A) Confocal images show labeling of 23B neurons in the front leg neuropils for two genetic driver lines: SS04746 (left) and R21B10 (right). Neurons labeled with mcd8::GFP (black) (sparsely labeled VNCs in Extended Data Fig. 4). **B)** Receptive field predictions for each line across all three cardinal axes (see Methods). Each line is labeled by the predicted receptive field along the proximal-distal axis. **C)** Experimental setup. Headless flies were tethered and positioned on a spherical treadmill. Red laser stimulation was directed to the body-coxa joint of the left front leg. Behavioral recording and joint tracking was collected from video data from six cameras (inset top) and tracked with DeepLabCut[68] and Anipose[48]. Bottom inset shows leg movements from one sweep (see Methods). **D)** Individual leg sweeps during grooming were identified as consecutive time points with two legs in close proximity and moving at a minimum velocity (see Methods). **E)** Example trials for empty-SpGal4 flies (control, gray), proximal 23Bs (dark

orange), and distal 23Bs (light orange). Distance from the left front leg to the nearest leg over time. Black arrows indicate individually detected sweeps of the left leg. **F)** Leg sweep ethogram with 15 random trials from empty-SpGal4 flies (control, top), proximal 23B flies (middle) and distal 23B flies (bottom). Each row represents an individual trial across time (seconds). Color represents whether the fly was engaged in leg sweeping (blue) or not (gray). **G)** Average sweep frequency (Hz) over time in seconds for control (empty-SpGal4) flies, proximal 23B flies (dark orange), and distal 23B flies (light orange). Distribution of sweep frequency for each line before the stimulus (prestim), during the stimulus (stim), and after the stimulus (poststim). For all box plots, center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; outliers not shown. **H)** Proportion of all leg sweeps using front leg grooming or middle leg grooming for each fly (left). Spatial distributions of the first contact point location for a subset of leg sweeps from each grooming pattern (right). **I)** Weighted spatial targeting in response to proximal and distal 23B activation compared to the connectivity-derived receptive fields presented in panel b.

Optogenetic stimulation of 23B neurons drives spatially-targeted grooming

Previous studies have shown that bristle activation in headless flies elicits spatially targeted grooming[13,16,18], suggesting that local VNC circuits are sufficient to support this behavior. To eliminate the contribution of descending input from the brain, we optogenetically stimulated 23B neurons in headless flies while tracking their behavior with 3D pose estimation[48]. We tethered headless flies, positioned them on a spherical treadmill, and recorded their behavior with six cameras. We targeted a red laser at the body-coxa joint of the front left leg to activate 23B neurons in the left front leg neuromere (Fig. 5C). To quantify spatial targeting of grooming behavior, we identified leg sweeps as consecutive time points where two legs were in contact and moving at a minimum velocity (Fig. 5D, individual sweeps noted by the black arrows). Upon 23B activation, sweep frequency increased in both experimental lines (Fig. 5E-G, Supplemental Video 1-2). The flies continued to sweep the left leg for several seconds after the stimulus terminated. On the other hand, control flies lacking CsChrimson[49] expression (empty-SpGal4) did not respond (Supplemental Video 3) (SS04746; 11 flies, 98 trials, R21B10; 8 flies, 73 trials, empty-SpGal4; 10 flies, 80 trials; Fig. 5E-G).

Activation of 23B subtypes elicits different spatially targeted grooming patterns

We observed two common grooming patterns in response to 23B activation: grooming the left front leg with the contralateral right front leg (front leg grooming) and grooming the left front leg with the ipsilateral left middle leg (middle leg grooming). Distal 23B activation elicited predominantly front leg grooming (96% front leg, 4% middle leg) while proximal 23B activation elicited both front leg and middle leg grooming (68% front leg, 32% middle leg; Fig. 5H). Thus, the activation of the different 23B subtypes triggered different grooming patterns.

We also wanted to determine if these different grooming patterns were spatially targeted, and how this compared to the predicted receptive field location of each driver line. For all instances of middle leg grooming, the flies brought the left middle leg forward to rub the stationary left front leg (Fig. 5H, left). We observed more variability in front leg grooming so we subdivided these instances into three categories (Fig. 5H, see Methods).

To measure the spatial specificity of each grooming pattern, we annotated the first contact position of individual sweeps (see Methods). Activating proximal 23B neurons produced front leg grooming, during which flies contacted the proximal femur of the targeted leg (Fig. 5H bottom). On the other hand, distal 23B activation triggered front leg grooming of the distal portion of the leg, i.e., the tibia and tarsus (Fig. 5H, top). Proximal 23B activation also elicited middle leg grooming of the distal femur tibia and tarsus, while distal 23B activation triggered middle leg grooming of the middle of the femur. Because flies from the two experimental groups did not use these grooming strategies equally, we multiplied the spatial targeting of each pattern by its prevalence to calculate a weighted spatial targeting (Fig. 5I, bottom). We observed that proximal 23B activation elicited grooming of the proximal leg, targeting the middle of the femur (Fig. 5I).

Distal 23B activation elicited grooming more distally, at the tibia-tarsus joint (Fig. 5I). These spatial patterns were consistent with our receptive field predictions based on the connectome (Fig. 5I, top).

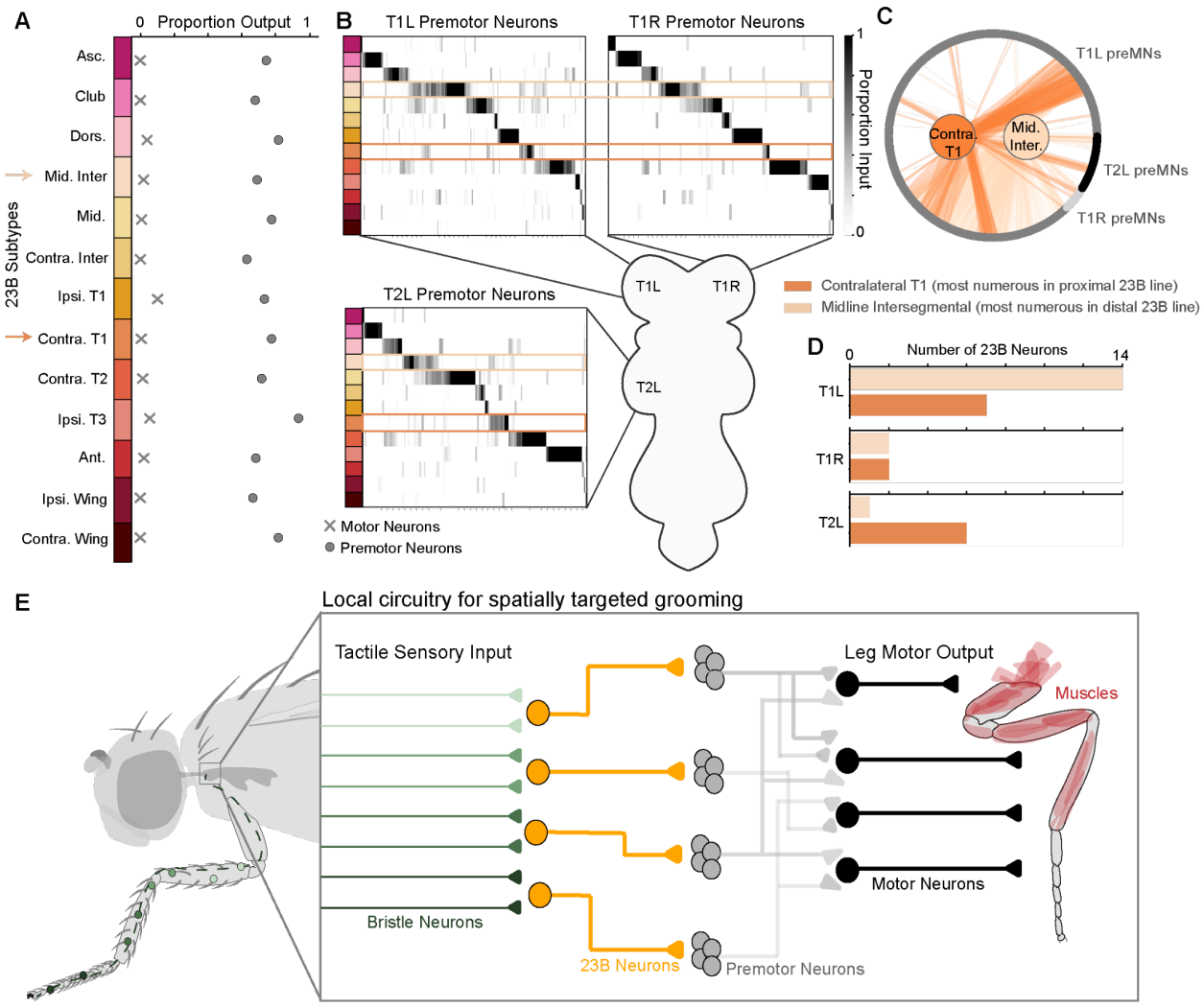


Figure 6: 23B subtypes synapse onto distinct leg premotor pools. A) Proportion of total synaptic output from 23B neurons onto motor (x) and premotor neurons (o). 23B neurons ordered and colored by subtype. **B)** Selectivity of 23B subtypes for left middle leg premotor neurons (T2L), left front leg premotor neurons (T1L), and right front leg premotor neurons (T1R). Colored boxes highlight Midline Intersegmental and Contralateral T1 as the most numerous subtypes in the distal and proximal grooming lines respectively. **C)** Contralateral T1 and Midline Intersegmental subtype connectivity onto T1L, T1R and T2L premotor neurons (preMNs). **D)** Number of 23B neurons from each subtype that contact T1L, T1R, and T2L premotor neurons. **E)** The local four-layer circuit. First-order bristle neurons form a tactile leg map. Second-order 23B neurons imbricate the leg map into overlapping receptive fields and target distinct premotor neuron pools. Premotor neurons recruit leg motor neurons to elicit spatially targeted grooming.

23B neurons do not directly contact leg motor neurons

Activation of both 23B driver lines elicited front leg grooming, however the precise leg movements differed in their spatial targeting (Fig. 5H). We therefore wanted to understand how the activation of different 23B subtypes could produce distinct leg movements. In the fly's front leg, 18 leg muscles are controlled by 71 uniquely identifiable motor neurons[31]. If different[31] 23B neurons produce distinct movements of the same leg, we might expect a difference in their synaptic connectivity onto leg motor neurons. We classified the downstream targets of 23B neurons and the proportion of 23B synapses onto each class type. Other than two cells (both projecting locally to the left front leg neuromere), 23B neurons rarely synapse on leg motor neurons, (1% synaptic output, Fig. 6A, Extended Data Fig. 5A). Thus, it is unlikely that 23B neurons directly recruit different leg motor neurons to produce distinct grooming patterns.

23B subtypes contact distinct pools of premotor neurons

We next quantified the proportion of 23B target neurons that were premotor. We defined premotor neurons as any neuron that was presynaptic to any motor neuron in the VNC[35]. We further classified each premotor neuron by the motor neurons it targets (e.g., left front leg, right front leg). We found that 75% of 23B synaptic output was onto premotor neurons across the VNC (Fig. 6A). In our experiments, we observed that the flies moved the left front leg, right front leg and left middle leg in response to front left leg 23B activation, thus we focused on these three premotor populations for subsequent analyses.

If our hypothesis is correct that different subtypes of 23B neurons elicit distinct grooming patterns, then we would expect them to contact distinct populations of premotor neurons. To test this, we measured the proportion of input from each 23B subtype onto individual premotor neurons. We found that across the three leg neuropils (T1L, T1R, T2L), many premotor neurons

received input from only one 23B subtype (Fig. 6B). While there was some degree of overlap, each 23B subtype synapsed onto a mostly unique set of premotor neurons. This supports the hypothesis that subtypes of 23B neurons recruit distinct motor patterns through distinct premotor populations.

Focusing on the two subtypes of 23B neurons we tested with optogenetics experiments, we observed that distal and proximal 23B neurons contact premotor neurons in three leg neuropils (T1L, T1R, T2L), though the specific populations differ across neuropils (Fig. 6B-C). While both 23B subtypes primarily synapse onto left front leg premotor neurons, six of seven proximal neurons make strong connections (12% of their premotor synaptic output) onto left middle leg premotor neurons in T2L (Fig. 6D, Extended Data Fig. 6B). These results are consistent with our finding that optogenetic stimulation of contralateral 23B neurons produced frequent middle leg grooming (Fig. 5H).

Taken together, we propose that spatially targeted grooming is mediated by a four-layer circuit from tactile sensory neurons to motor neurons (Fig. 6E). First-order tactile sensory neurons target local interneurons that belong to hemilineage 23B. These second-order interneurons imbricate the leg map into overlapping receptive fields and target distinct pools of third-order premotor neurons. Premotor neurons then drive dynamic patterns of leg movement through excitation and inhibition of leg motor neurons. Although not shown in the circuit schematic in Figure 6E, we note that the two middle layers exhibit dense recurrent connectivity, which may support grooming dynamics[35].

Discussion

A somatotopic map of the fly leg

In this study, we used genetic labeling to determine that tactile bristles from the fly's left front leg form a somatotopic map in the VNC. Notably, this map matches the somatotopic

organization of the leg imaginal disc during development in several key respects. The development of the leg is regulated by graded expression of transcription factors. The PD and DV axes of the leg are established by genes like *apterous* and *decapentaplegic*, while *wingless* and *hedgehog* expression establish the "compartment boundary" along the anterior/posterior axis[40,41]. In the VNC leg neuropil, we found that bristle axons are also organized along a gradient in the PD and DV axes, where the projection of each axon is slightly offset relative to its neighbor. Yet the AP axis was divided by bristle axons that branches either anteriorly or posteriorly, as if separated by a compartment boundary at the center of the left front leg neuromere. In summary, we observe striking similarities between the spatial organization of the leg imaginal disc and the topographic projections of bristle axons in the VNC neuropil.

While the differentiation of the leg imaginal disc occurs well before bristle neurons have developed[50,51], it is possible that similar molecular factors regulate the temporal differentiation of sensory cells, axon guidance, and development into the nervous system. Recent studies tracing bristle neuron growth and development from the locust antenna suggest that bristle axons enter the nerve tract in order of differentiation[52]. Distal neurons, which differentiate first, enter the nerve tract and are surrounded by more proximal neurons as they grow towards the central nervous system. This results in distal neurons occupying the central region of the tract and proximal neurons concentrically wrapping themselves around the periphery. This topography of the PD axis is consistent with our findings in the fly VNC and previous work tracing bristle neurons on the head[22,53]. Investigations into the underlying mechanisms and exact timing of sensory axon development in the VNC will be necessary to elucidate how the leg bristle map is established in the fly VNC. Beyond the leg, other precursor structures such as the wing, haltere, and antennal imaginal discs, may also contribute to the creation of somatotopic maps in the adult fly nervous

system.

Limitations

Our connectome results come from one dataset of a female adult nerve cord (FANC). However, the general distribution of bristle axons and the strong downstream connectivity onto 23B neurons is maintained in a connectome dataset from the male adult nerve cord (MANC)[54]. Due to the variable and immutable state of bristle neuron reconstructions in MANC, a direct comparison was not possible. However, we were able to identify all of the different 23B subtypes in MANC (not shown). This suggests the circuitry is stereotyped across flies and not sexually dimorphic. Similar to conclusions from a comparison of multiple fly brain connectomes[30], we expect that the overall structure of the bristle sensorimotor circuit is similar across individuals, while the precise connectivity between individual neurons may vary. The consistency between the light-level morphologies described here and the connectome morphologies supports this view, as does the fact that predictions based on the connectome of one fly were validated in behavioral experiments done on other flies. With the recent availability of connectomes of the full central nervous system[33,55], future analyses may also elucidate how the connectivity to and from the brain affects grooming dynamics.

Grooming behavior

Previous studies mapping the tactile receptive fields of interneurons in the fly and locust proposed that the tactile circuit is composed of diverging streams of tactile information[56,57]. Our results confirm this hypothesis through the dense reconstruction of the tactile circuit from one leg. We observed that the population of 23B neurons imbricate the leg with distinct yet overlapping receptive fields. After classifying 23B neurons by their axonal projection patterns, we found that neurons of the same subtype contact similar downstream targets and that these different subtypes

contact distinct premotor populations across leg neuropils in the VNC. In other words, nearby bristle signals form diverging streams of tactile information that feed into distinct sensorimotor modules. In the spinal cord and the brain, modular motor circuits are found across species and provide a structural scaffold for controlling flexible behavior[58–62]. Here we propose that distinct 23B subtypes work in concert to activate different populations of premotor neurons that in turn activate motor neurons to elicit targeted grooming responses.

Are grooming circuits for other body parts similarly organized? Previous work in the fly antennal grooming circuit focused on a class of brain interneurons that they refer to as B2. Interestingly, the B2 cells also develop from hemilineage 23B[15]. B2 neurons are strong downstream targets of antennal mechanosensory neurons and, similar to our findings, optogenetic activation of B2 neurons increased antennal grooming. These similarities suggest that the structural and functional organization of grooming circuits in the fly may be repeated across body segments. If so, how do these circuits interact, for example when bristles are activated all over the body of a fly? Past work has shown that flies groom their bodies with a stereotyped and hierarchical pattern, starting with the head and proceeding to the legs and abdomen[15,19,20]. Furthermore, several studies have described command-like neurons that elicit grooming of different body segments[15,18,23]. If subtypes of 23B interneurons imbricate each body segment, future investigations into the interactions between 23B neurons and these command-like neurons may provide insight into the neural mechanisms that underlie the hierarchical organization and coordination of grooming behavior.

Beyond grooming, bristle activation can elicit movements such as walking, uncoordinated leg movements and kicking[13,16]. Here we focused on two 23B subtypes for which we were able to identify specific genetic driver lines. In the future, it will be interesting to explore the range of

actions produced by activation of other 23B subtypes, as well as their natural activity patterns during grooming behavior. Here we predicted the receptive fields for 23B neurons (Fig. 3), but this approach could be used to define the receptive field of any neuron downstream of bristle axons. Characterization of other interneurons within the tactile circuitry of the VNC will help define the degree to which touch signals diverge to distinct sensorimotor modules and whether 23B neurons are necessary for all spatially targeted behaviors.

From sensory input to motor output

In our behavioral experiments, we observed several categories of front leg grooming, suggesting that the spatial location of the tactile stimulus dictates the movement of the leg and patterns of muscle contraction. Our analysis of the four-layer sensorimotor circuit suggests that the distinct premotor connectivity of 23B subtypes is important for producing spatially targeted grooming. If bristle neurons can be equated to the pixels of the somatosensory space, we propose that different 23B subtypes sample the leg space to drive the appropriate, spatially-targeted behavioral response.

While we have outlined a simplified four-layer circuit, the connectivity of premotor circuitry onto motor neurons is very complex[31,35]. Understanding how tactile stimuli elicit dynamic motor patterns will require recordings of activity dynamics in 23B neurons and downstream cells during behavior. Dynamic modeling of the connectome may also reveal new insights, as this approach has reproduced the functional role of previously characterized cells and revealed the function of uncharacterized circuits[63–65]. For example, future studies that simulate the tactile circuitry could compare how motor neuron outputs change as a function of which premotor pools are activated and the influence of proprioceptors as a proxy for limb position. While we focus on the excitatory 23B neurons in this study, the second strongest target of bristle axons were inhibitory

neurons of hemilineage 1B (Fig. 2F). More work is needed to understand how inhibitory signals sculpt spatiotemporal processing of tactile signals. More generally, our work establishes a model circuit within the fly nerve cord to explore how transient sensory stimuli (e.g, touching a leg) produce sustained and dynamic patterns of motor activity.

References

1. Sherrington, C. S. Decerebrate Rigidity, and Reflex Coordination of Movements. *J. Physiol.* **22**, 319–332 (1898).

2. Stein, P. S. G. & Grossman, M. L. Central program for scratch reflex in turtle. *J. Comp. Physiol.* **140**, 287–294 (1980).

3. Mortin, L. I. & Stein, P. S. Spinal cord segments containing key elements of the central pattern generators for three forms of scratch reflex in the turtle. *J. Neurosci.* **9**, 2285–2296 (1989).

4. Kaas, J. H. Topographic Maps are Fundamental to Sensory Processing. *Brain Res. Bull.* **44**, 107–112 (1997).

5. Penfield, W. & Boldrey, E. Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain* **60**, 389–443 (1937).

6. Narici, L. *et al.* Neuromagnetic somatosensory homunculus: A non-invasive approach in humans. *Neurosci. Lett.* **121**, 51–54 (1991).

7. Engel, S. A., Glover, G. H. & Wandell, B. A. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb. Cortex* **7**, 181–192 (1997).

8. Wandell, B. A. & Winawer, J. Imaging retinotopic maps in the human brain. *Vision Res.* **51**, 718–737 (2011).

9. Held Jr., L. I. Bristle patterning in Drosophila. *BioEssays* **13**, 633–640 (1991).

10. Newland, P. L. & Burrows, M. Processing of tactile information in neuronal networks controlling leg movements of the Locust. *J. Insect Physiol.* **43**, 107–123 (1997).

11. Tuthill, J. C. & Wilson, R. I. Parallel Transformation of Tactile Signals in Central Circuits of Drosophila. *Cell* **164**, 1046–1059 (2016).

12. Walker, R. G., Willingham, A. T. & Zuker, C. S. A Drosophila Mechanosensory Transduction Channel. *Science* **287**, 2229–2234 (2000).

13. Vandervorst, P. & Ghysen, A. Genetic control of sensory connections in Drosophila. *Nature* **286**, 65–67 (1980).

14. Corfas, G. & Dudai, Y. Adaptation and fatigue of a mechanosensory neuron in wild-type Drosophila and in memory mutants. *J. Neurosci.* **10**, 491–499 (1990).

15. Hampel, S., Franconville, R., Simpson, J. H. & Seeds, A. M. A neural command circuit for grooming movement control. *eLife* **4**, e08758 (2015).

16. Medeiros, A. M., Hobbiss, A. F., Borges, G., Moita, M. & Mendes, C. S. Mechanosensory bristles mediate avoidance behavior by triggering sustained local motor activity in Drosophila melanogaster. *Curr. Biol.* **34**, 2812-2830.e5 (2024).

17. Li, J. *et al.* A Defensive Kicking Behavior in Response to Mechanical Stimuli Mediated by Drosophila Wing Margin Bristles. *J. Neurosci.* **36**, 11275–11282 (2016).

18. Zhang, N. & Simpson, J. H. A pair of commissural command neurons induces Drosophila wing grooming. *iScience* **25**, 103792 (2022).

19. Hampel, S., McKellar, C. E., Simpson, J. H. & Seeds, A. M. Simultaneous activation of parallel sensory pathways promotes a grooming sequence in Drosophila. *eLife* **6**, e28804 (2017).

20. Seeds, A. M. *et al.* A suppression hierarchy among competing motor programs drives sequential grooming in Drosophila. *eLife* **3**, e02951 (2014).

21. Mueller, J. M., Ravbar, P., Simpson, J. H. & Carlson, J. M. Drosophila melanogaster grooming possesses syntax with distinct rules at different temporal scales. *PLOS Comput. Biol.* **15**, e1007105 (2019).

22. Eichler, K. *et al.* Somatotopic organization among parallel sensory pathways that promote a grooming sequence in Drosophila. *eLife* **12**, RP87602 (2024).

23. Yoshikawa, S., Tang, P. & Simpson, J. H. Mechanosensory and command contributions to the *Drosophila* grooming sequence. *Curr. Biol.* **34**, 2066-2076.e3 (2024).

24. Murphey, R. K., Possidente, D. R., Vandervorst, P. & Ghysen, A. Compartments and the topography of leg afferent projections in Drosophila. *J. Neurosci.* **9**, 3209–3217 (1989).

25. Newland, P. L., Rogers, S. M., Gaaboub, I. & Matheson, T. Parallel somatotopic maps of gustatory and mechanosensory neurons in the central nervous system of an insect. *J. Comp. Neurol.* **425**, 82–96 (2000).

26. Tsubouchi, A. *et al.* Topological and modality-specific representation of somatosensory information in the fly brain. *Science* **358**, 615–623 (2017).

27. Schrader, S. & Merritt, D. j. Central projections of Drosophila sensory neurons in the transition from embryo to larva. *J. Comp. Neurol.* **425**, 34–44 (2000).

28. Newland, P. L. Morphology and somatotopic organisation of the central projections of afferents from tactile hairs on the hind leg of the locust. *J. Comp. Neurol.* **312**, 493–508 (1991).

29. Scheffer, L. K. *et al.* A connectome and analysis of the adult Drosophila central brain. *eLife* **9**, e57443 (2020).

30. Schlegel, P. *et al.* Whole-brain annotation and multi-connectome cell typing of Drosophila. *Nature* **634**, 139–152 (2024).

31. Azevedo, A. *et al.* Connectomic reconstruction of a female Drosophila ventral nerve cord. *Nature* **631**, 360–368 (2024).

32. Takemura, S. *et al.* A Connectome of the Male Drosophila Ventral Nerve Cord. *eLife* **13**, (2024).

33. Bates, A. S. *et al.* Distributed control circuits across a brain-and-cord connectome. 2025.07.31.667571 Preprint at https://doi.org/10.1101/2025.07.31.667571 (2025).

34. Phelps, J. S. *et al.* Reconstruction of motor control circuits in adult Drosophila using automated transmission electron microscopy. *Cell* **184**, 759-774.e18 (2021).

35. Lesser, E. *et al.* Synaptic architecture of leg and wing premotor control networks in Drosophila. *Nature* **631**, 369–377 (2024).

36. Lewis, E. B. A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565–570 (1978).

37. Lecuit, T. & Cohen, S. M. Proximal–distal axis formation in the Drosophila leg. *Nature* **388**, 139–145 (1997).

38. Kojima, T. The mechanism of Drosophila leg development along the proximodistal axis. *Dev. Growth Differ.* **46**, 115–129 (2004).

39. Ruiz-Losada, M., Blom-Dahl, D., Córdoba, S. & Estella, C. Specification and Patterning of Drosophila Appendages. *J. Dev. Biol.* **6**, 17 (2018).

40. Sturtevant, M. A., Biehs, B., Marin, E. & Bier, E. The spalt gene links the A/P compartment boundary to a linear adult structure in the Drosophila wing. *Dev. Camb. Engl.* **124**, 21–32 (1997).

41. Beira, J. V. & Paro, R. The legacy of Drosophila imaginal discs. *Chromosoma* **125**, 573–592 (2016).

42. Court, R. *et al.* A Systematic Nomenclature for the Drosophila Ventral Nerve Cord. *Neuron* **107**, 1071-1079.e2 (2020).

43. Lacin, H. *et al.* Neurotransmitter identity is acquired in a lineage-restricted manner in the Drosophila CNS. *eLife* **8**, e43701 (2019).

44. Bryant, P. J. Chapter 2 Determination and Pattern Formation in The Imaginal Discs Of *Drosophila*. in *Current Topics in Developmental Biology* (eds. Moscona, A. A. & Monroy, A.) vol. 8 41–80 (Academic Press, 1974).

45. Eckstein, N. *et al.* Neurotransmitter classification from electron microscopy images at synaptic sites in Drosophila melanogaster. *Cell* **187**, 2574-2594.e23 (2024).

46. Lacin, H. & Truman, J. W. Lineage mapping identifies molecular and architectural similarities between the larval and adult Drosophila central nervous system. *eLife* **5**, e13399 (2016).

47. Harris, R. M., Pfeiffer, B. D., Rubin, G. M. & Truman, J. W. Neuron hemilineages provide the functional ground plan for the Drosophila ventral nervous system. *eLife* **4**, e04493 (2015).

48. Karashchuk, P. *et al.* Anipose: A toolkit for robust markerless 3D pose estimation. *Cell Rep.* **36**, (2021).

49. Klapoetke, N. C. *et al.* Independent optical excitation of distinct neural populations. *Nat. Methods* **11**, 338–346 (2014).

50. von Kalm, L., Fristrom, D. & Fristrom, J. The making of a fly leg: A model for epithelial morphogenesis. *BioEssays* **17**, 693–702 (1995).

51. Mangione, F. *et al.* Co-option of epidermal cells enables touch sensing. *Nat. Cell Biol.* **25**, 540–549 (2023).

52. Boyan, G. & Ehrhardt, E. From bristle to brain: embryonic development of topographic projections from basiconic sensilla in the antennal nervous system of the locust Schistocerca gregaria. *Dev. Genes Evol.* **234**, 33–44 (2024).

53. Calle-Schuler, S. A., Santana-Cruz, A. E., Kmecová, L., Hampel, S. & Seeds, A. M. A comprehensive mechanosensory connectome reveals a somatotopically organized neural circuit architecture controlling stimulus-aimed grooming of the Drosophila head. 2025.05.19.654894 Preprint at https://doi.org/10.1101/2025.05.19.654894 (2025).

54. Marin, E. C. *et al.* Systematic annotation of a complete adult male Drosophila nerve cord connectome reveals principles of functional organisation. *eLife* **13**, (2024).

55. Berg, S. *et al.* Sexual dimorphism in the complete connectome of the Drosophila male central nervous system. 2025.10.09.680999 Preprint at https://doi.org/10.1101/2025.10.09.680999 (2025).

56. Burrows, M. Local circuits for the control of leg movements in an insect. *Trends Neurosci.* **15**, 226–232 (1992).

57. Burrows, M. & Newland, P. L. Correlation between the receptive fields of locust interneurons, their dendritic morphology, and the central projections of mechanosensory neurons. *J. Comp. Neurol.* **329**, 412–426 (1993).

58. Bizzi, E., D'Avella, A., Saltiel, P. & Tresch, M. Book Review: Modular Organization of Spinal Motor Systems. *The Neuroscientist* **8**, 437–442 (2002).

59. Bizzi, E., Mussa-Ivaldi, F. A. & Giszter, S. Computations Underlying the Execution of Movement: A Biological Perspective. *Science* **253**, 287–291 (1991).

60. Giszter, S. F., Mussa-Ivaldi, F. A. & Bizzi, E. Convergent force fields organized in the frog's spinal cord. *J. Neurosci.* **13**, 467–491 (1993).

61. Tresch, M. C., Saltiel, P. & Bizzi, E. The construction of movement by the spinal cord. *Nat. Neurosci.* **2**, 162–167 (1999).

62. Gentner, R. & Classen, J. Modular Organization of Finger Movements by the Human Central Nervous System. *Neuron* **52**, 731–742 (2006).

63. Shiu, P. K. *et al.* A Drosophila computational brain model reveals sensorimotor processing. *Nature* **634**, 210–219 (2024).

64. Lappalainen, J. K. *et al.* Connectome-constrained networks predict neural activity across the fly visual system. *Nature* **634**, 1132–1140 (2024).

65. Pugliese, S. M. *et al.* Connectome simulations identify a central pattern generator circuit for fly walking. 2025.09.12.675944 Preprint at https://doi.org/10.1101/2025.09.12.675944 (2025).

66. Dorkenwald, S. *et al.* CAVE: Connectome Annotation Versioning Engine. 2023.07.26.550598 Preprint at https://doi.org/10.1101/2023.07.26.550598 (2023).

67. Moore, R. J. D. *et al.* FicTrac: A visual method for tracking spherical motion and generating fictive animal paths. *J. Neurosci. Methods* **225**, 106–119 (2014).

68. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).

Materials and Methods

Sample preparation for confocal imaging of imaginal discs

For confocal imaging of imaginal discs (Fig. 1, Extended Data Fig. 2), we crossed flies carrying the Gal4 driver to flies carrying pJFRC7-20XUAS-IVS-mCD8::GFP. Prothoracic leg imaginal discs were dissected from third instar larvae in PBS, and fixed for 20 minutes in 4% paraformaldehyde in PBS at room temperature. Discs were washed and permeabilized 3x in 0.2% Triton X-100 in PBS (PBST) over 1 hour, then incubated in 1:50 phalloidin for 1 hour at room temperature. The discs were rinsed 3x with PBS over 1 hour, then mounted in VectaShield. We acquired z-stacks on a confocal microscope (Olympus FV1000).

Sample preparation for confocal imaging of VNCs

For confocal imaging of mcd8::GFP-labeled neurons in the VNCs (Fig. 1, Extended Data Fig. 2), we dissected the VNC from 2-day old female adults in PBS. We fixed the VNC in a 4% paraformaldehyde PBS solution for 20 min and then rinsed the VNC in PBS three times. We put the VNC in blocking solution (5% normal goat serum in PBST) for 20 min, then incubated it with a solution of primary antibodies (chicken anti-GFP antibody, 1:50; rabbit anti-dsRed 1:500; anti-brp mouse for nc82 neuropil staining, 1:50) in blocking solution for 24 hours at room temperature. At the end of the first incubation, we washed the VNC with PBS with 0.2% Triton-X (PBST) three times over two hours, then incubated the VNC in a solution of secondary antibody (anti-chicken-Alexa 488 1:250; anti-rabbit-Alexa 568 1:250; anti-mouse-Alexa 633 1:250) dissolved in blocking solution for 24 hours at room temperature. Finally, we washed the VNC in PBST three times, once in PBS, and then mounted on a slide with Vectashield (Vector Laboratories). We acquired z-stacks of each VNC on a confocal microscope (Olympus FV1000).

We aligned the morphology of the VNC to a female VNC template in ImageJ with the Computational Morphometry Toolkit plugin (CMTK32; http://nitrc.org/projects/cmtk).

Sample preparation for confocal imaging of bristles on legs

For confocal imaging of mcd8::GFP-labeled bristles in legs (Figure 1, Extended Data Fig. 2), we selected prothoracic legs from 2-day old female adults while the flies were anesthetized with CO2. We immediately fixed the legs in 4% formaldehyde in PBS with 0.2% Triton-X for 20 min and rinsed them in PBS three times over 30 minutes. We mounted the legs in VectaShield and acquired z-stacks on a confocal microscope (Olympus FV1000).

Fly Transgene	**Full genotype**	**Source**	**Identifier**
UAS flp (x)	P{w\[+mC\]=UAS-FLP.Exel}1, y\[1\] w\[1118\]	Bloomington	RRID:BDSC 8208
UAS flp (II)	y[1] w[*]; P{w[+mC]=UAS-FLP.D}JD1	Bloomington	RRID:BDSC 4539
LexAop>stop>mcd8::GFP	y\[1\] w\[*\]; +; P{w\[+mC\]=lexA(stop.FRT)mCD8.GFP}3	Bloomington	RRID:BDSC 57588
R38B08-LexA	w[*]; R38B08-LexA / CyO; TM6b/MKRS	Gift from Janelia	n/a
LexAop-mcd8::GFP	P{13XLexAop2-mCD8::GFP}attP40/CyO	Bloomington	RRID:BDSC 32205
dac-GAL4	dac-GAL4[P7d23]	Gift from Victor Hatini (Tufts)	
hh-GAL4	y\[1\] w\[*\]; Mi{Trojan-GAL4.0}hh\[MI10526-TG4.0\]/TM3, Sb\[1\] Ser\[1\]	Bloomington	RRID:BDSC 67493
wg-GAL4	w\[*\]; P{w\[+mW.hs\]=GAL4-wg.M}MA1	Bloomington	RRID:BDSC 4918

ap-GAL4	y\[1\] w\[1118\]; P{w\[+mW.hs\]=GawB}ap\[md544\]/CyO	Bloomington	RRID:BDSC 3041
Dll-GAL4	P{w\[+mW.hs\]=GawB}Dll\[md23\]/CyO	Bloomington	RRID:BDSC 3038
dpp-GAL4	w\[*\]; wg\[Sp-1\]/CyO; P{w\[+mW.hs\]=GAL4-dpp.blk1}40C.6/TM6B, Tb\[1\]	Bloomington	RRID:BDSC 1553
rn-GAL4	w\[1118\]; P{w\[+mW.hs\]=GawB}rn\[GAL4-5\]/TM3, P{ry\[+t7.2\]=ftz-lacC}SC1, ry\[RK\] Sb\[1\] Ser\[1\]	Bloomington	RRID:BDSC 7405
mid-GAL4	w[*]; P{w[+mW.hs]=GawB}NP2113 / CyO	Kyoto DGGR	104093
LexAop>stop>CsChrimson (II)	13XLexAop2>dsFRT>CsChrimson-mVenus in su(Hw)attP5	Gift from Yoshi Aso, Janelia	
LexAop>stop>CsChrimson (III)	13XLexAop2>dsFRT>CsChrimson-mVenus in attP2	Gift from Yoshi Aso, Janelia	
UAS-CsChrimson	w[1118]; P{y[+t7.7] w[+mC]=20XUAS-IVS-CsChrimson.mVenus}attP40	Bloomington	RRID:BDSC 55135
UAS-mcd8::GFP	P{pJFRC7-020XUAS-IVS-mCD8::GFP}attP2	Gift from Rubin Lab, Janelia	
R21b10-GAL4	w[1118]; P{y[+t7.7] w[+mC]=R21B10-GAL4}attP2	Bloomington	RRID:BDSC 49295
ss04746 split GAL4	w[1118]; P{y[+t7.7] w[+mC]=R77C10-p65.AD}attP40; P{y[+t7.7] w[+mC]=VT026010-GAL4.DBD}attP2	Bloomington	RRID:BDSC 88151
empty split-Gal4	w[1118]; P{y[+t7.7] w[+mC]=p65.AD.Uw}attP40	Bloomington	RRID:BDSC 79603

	; P{y[+t7.7] w[+mC]=GAL4.DBD.Uw}attP2		
UAS-phiC31	P{UAS-phiC31}attP18; Star/CyO; Pri/TM6B	Gift from Rachel Wilson	
SPARC2 CsChrimson (intermediate)	TI{20XUAS-SPARC2-I-Syn21-CsChrimson::tdTomato-3.1}CR-P40	Bloomington	RRID: BDSC 84144
SPARC2 CsChrimson (sparse)	TI{20XUAS-SPARC2-S-Syn21-CsChrimson::tdTomato-3.1}CR-P40	Bloomington	RRID:BDSC 84145

Reagent	**Source**	**Identifier**
Mouse anti-Bruchpilot antibody	Developmental Studies Hybridoma Bank	RRID:AB_2314866
Chicken GFP polyclonal antibody	Thermofisher PA1-9533	RRID:AB_1074893
Rabbit DsRed Polyclonal Antibody	Takara Bio 632496	RRID:AB_10013483
Goat anti-mouse secondary antibody, Alexa Fluor 633 conjugate	Thermofisher A-21050	RRID:AB_141431
Goat anti-Chicken IgG, Alexa Fluor 488	Thermofisher A-11039	RRID:AB_2534096
Goat anti-Rabbit IgG, Alexa Fluor 568	Thermofisher A-11011	RRID:AB_143157
Alexa Fluor Phalloidin 647	Thermofisher A22287	n/a
Vectashield mounting medium	Vector Labs H-1000	n/a

Bristle neuron reconstruction

409 tactile mechanosensory axons were reconstructed from the front left leg in a connectome dataset of the female adult nerve cord (Fig. 2, Extended Data Fig. 1)[31,34,35]. Reconstruction, referred to as proofreading, was completed using Neuroglancer, an interactive software for visualizing, editing, and annotating 3D volumetric data. Proofreading entailed two types of edits; splitting off neurites that did not belong to the cell of interest and merging segments of the neuron that were falsely missed by the automated segmentation. All edits and annotations to these neurons are hosted and accessible on the connectome annotation versioning engine (CAVE) platform[66]. 394 of the reconstructed axons entered the VNC through the Leg Nerve, eight from the ventral prothoracic nerve and seven from the dorsal prothoracic nerve. A small number (<20) of axons could not be reconstructed due to irreconcilable segmentation errors.

Spatial mapping in FANC

To project the spatial axes of the leg map onto the bristle axons in FANC, three mapping rules were applied. The first was that each axon was classified as either anterior or posterior based on whether the axon morphology branched anteriorly or posteriorly upon entering the VNC (Fig. 2B). The DV and PD axes were quantified along a gradient to reflect the distribution observed from the genetic labeling experiments (Fig. 1). For each axis, a mapping point was placed within the neuropil and the distance of every synapse from that point was calculated. To account for spatial outliers, we normalized the distribution of distances along each axis by the 1st and 99th percentile. The relative spatial prediction of each axon was the average synaptic distance from each reference point (Fig. 2B).

Analysis of circuit connectivity

To reduce the presence of weak connections and the likelihood of false positive synapse detections, connections with fewer than three synapses between pairs of neurons were filtered out of all analyses, similar to past work[29,30]. We proofread all downstream targets of the bristle neuron and 23B neuron populations that met this synapse threshold.

We classified each neuron by class (local, intersegmental, ascending, descending, sensory or unknown). We defined local cells as VNC interneurons with inputs limited to the left front leg neuromere, whereas intersegmental cells received input from multiple neuropils. Ascending neurons had a soma in the VNC and projected up through the neck connective. Descending neurons did not have a soma in the VNC and consisted of axons that projected down from the neck connective. We defined sensory cells as afferent axons incoming from the peripheral neurons. Finally, we labeled neuronal fragments that could not be reconnected to the larger arbor as Unknown. Synapses that belonged to an 'unknown' object were also filtered out of all analyses (6% of the total connectivity).

We classified all VNC neurons in the tactile circuit by developmental hemilineage. Cells within a hemilineage are born from the same post embryonic stem cell and share morphological features, neurotransmitter expression, and broad functional roles within the VNC[43,46,47]. We assigned hemilineage identity based on soma location, fasciculation bundle into the VNC and dendritic and axonal morphology and projection patterns[46,54]. We then inferred neurotransmitter identity from the hemilineage classification based on previously published experiments[43,47]. Less than 1%[43,47]neurons could not be classified into a specific hemilineage and were filtered out of any analyses that depended on this labeling (Fig. 2F).

23B subtype classification

We reconstructed 62 23B neurons downstream of bristle neurons from the left front leg in the FANC connectome. This included 58 from the left front leg neuropil, 3 from the left wing neuropil that extended into the left front leg neuromere. We classified 23B neurons into subtypes based on the axonal projection pattern (Fig. 4). For example, 23B neurons in the left front leg neuromere with an axon that projected to the front right leg neuromere were considered Contralateral T1 neurons. 23B neurons that projected to the left wing neuropil were labeled as Ipsilateral Wing neurons and so on (Fig. 4A) Axons from neurons of the same subtype bundled together in the VNC. Therefore, in cases where neurons had axons with an ambiguous projection pattern, we classified them based on the axons they bundled with.

Receptive field calculation

Based on the spatial mapping methods outlined above, we mapped a single location on the leg for each bristle axon and its output synapses (Fig. 2B). For each 23B neuron, we selected all the input synapses from bristle axons (Fig. 3C). The receptive field along each cardinal axis was represented as the distribution of spatial locations as they were mapped to the presynaptic bristles (Fig. 3D-E). If for example a 23B neuron received input from three bristles axons that we had mapped to the ventral proximal area of the leg, the receptive field would be represented by the distribution of input synapses from those three axons. The same method was applied to each 23B neurons (Fig. 3E).

SPARC labeling of 23B neurons

To classify the axonal projection patterns of individual 23B neurons labeled by our two experimental lines, we crossed *UAS-PhiC31; ss04746-split-GAL4* or *UAS-PhiC31; R21b10-GAL4* females to males carrying the intermediate or sparse variants of SPARC2 CsChrimson (Extended

Data Fig. 4). We dissected, fixed, stained, and imaged the VNCs as described above. Neurons were classified by manual inspection of the image stacks based on the morphology and projection pattern of the axon. (Extended Data Fig. 4)

Connectome derived spatial targeting prediction

Based on the proportions derived from our sparsely labelled VNCs (Extended Data Fig. 4C), we sampled a subset of 23B neurons and summed the bristle input from these cells to predict the aggregate receptive field for that set of neurons. For example, for SS04746, there were six neurons labeled in each neuromere so we sampled six neurons with a sampling rate weighted by the proportion of subtypes present in the SPARC2 experiments (pie chart in Extended Data Fig. 4). The aggregate receptive field from this set of six neurons was considered one simulated RF. We then simulated 100 RFs to create the average RF for each experimental line.

Optogenetic experiments

Optogenetic experiments were performed on adult female flies that were raised on 35mM in 95% EtOH ATR for 1-3 days, were 2-5 days old, de-winged, and fixed to a rigid tether (0.1 mm thin tungsten rod) with UV glue (KOA 300). These flies were placed onto a spherical foam ball (weight: 0.13 g; diameter: 9.08mm) suspended by air within a dark arena. A red laser (638 nm; 1200 Hz pulse rate; 30% duty cycle, Laserland) was focused on the thorax-coxa joint of the left front leg (Fig. 5C). Optogenetic activation experiments were conducted on flies in which different subtypes of 23B neurons expressed CsChrimson, as well as flies with an empty-SpGal4 (**Table 1**). Trials were 20 seconds in duration and consisted of five seconds prestimulus, five second with the laser flickering on/off at 5Hz, and 10 second post stimulus (Fig. 5E). During each trial, the behavior each fly was recorded with 6 high-speed cameras (300 fps; Basler acA800-510 μm; Balser AG) and the movement of the ball was recorded at 30 fps with a camera (FMVU-03MTM-CS) and

processed using FicTrac[67]. The 3D positions of each leg joint were determined by using DeepLabCut[68] and Anipose[48] (Fig. 5C-D). Kinematic analyses were performed in a custom Python script.

Leg sweep detection

We used the 3D joint positions to detect contacts between legs (Fig. 5C-D). The automated tracking detected the following joints for each leg of the fly: body-coxa, coxa-femur, femur-tibia, tibia-tarsus, and the tarsus tip. We interpolated vectors between the joints of individual legs to represent the legs in 3D space. We defined contacts as individual frames where two legs were in close proximity to one another. The distance threshold we used to classify contacts varied between flies to account for diurnal variability in camera calibration settings, however they all ranged between 0.13-0.17 pixel distance. We defined leg sweeps as consecutive frames with a contact detection between the same two legs. At least one of the legs had to be moving at a minimum velocity of 2 mm per second to be considered a valid leg sweep (Fig. 5D-E). We added the velocity condition to exclude moments when the fly idly stood with two legs in contact. Finally, to account for noise from the binary contact detection, we merged individual sweeps that were separated by three or less frames (Fig. 5D-E).

Spatial targeting and contact point annotation

To define the spatial targeting of each grooming pattern we needed the exact contact point location between legs. Since we tracked joint positions and not entire leg segments, we annotated the contact points for a subset of frames that could then be measured relative to our interpolated legs. To do this we defined the first point of contact as the first frame of each individual leg sweep. We then divided first contacts by grooming pattern based on the legs involved; sweeps between the left front leg and the middle front leg were considered middle leg grooming, sweeps between

the two front legs were considered front leg grooming (Fig. 5H). We sampled first contact frames for each grooming pattern across the two populations of experimental flies: Middle leg grooming SS04746 (53), middle leg grooming R21B10 (42), Front leg grooming SS04746 (32), Front leg grooming R21B10 (67). All frames were annotated by a person blind to genotype using the point annotation software Anivia. We annotated the contact point location across all six camera views for each frame. Due to the variability in front leg grooming we also annotated the category of front leg grooming. We defined Category 1 as both front legs towards the midline. Category 2 was when flies brought the right leg over to the left side and contacted an extended left leg. Category 3 when flies brought the left front leg over to the right and contacted an extended right leg (Fig. 5H).

To compare contact point locations relative to the leg in 3D space, we triangulated the annotated contact points into the same space. This was done by importing the calibration settings for each respective trial and running the tracking process described above. To determine the spatial location of the contact we measured the closest point on the interpolated legs to the annotation point. We defined the spatial targeting profile as the distribution of leg locations contacted for each grooming pattern (Fig. 5H).

Figure Reference	**Genotype**
Figure 1C	Leg imaginal discs: w[1118]; dac-Gal4 / +; UAS-mcd8::GFP / + w[1118]; +; rn-Gal4 / UAS-mcd8::GFP Adult vnc and leg: w[1118]; dac-Gal4 / UAS-flp;LexAop>stop>mcd8::GFP / R38B08-LexA w[1118] ; LexAop>stop>mcd8::GFP / UAS-flp; rn-Gal4 / R38B08-LexA
Figure 5A	w[1118]; +; ss04746 split GAL4 / UAS-mcd8::GFP

	w[1118]; +; R21B10-GAL4/ UAS-mcd8::GFP
Figure 5E-H	Control w[1118]; P{y[+t7.7] w[+mC]=p65.AD.Uw}attP40 / P{y[+t7.7] w[+mC]=20XUAS-IVS-CsChrimson.mVenus}attP40; P{y[+t7.7] w[+mC]=GAL4.DBD.Uw}attP2 / + Proximal 23B w[1118]; P{y[+t7.7] w[+mC]=R77C10-p65.AD}attP40 / P{y[+t7.7] w[+mC]=20XUAS-IVS-CsChrimson.mVenus}attP40 ; P{y[+t7.7] w[+mC]=VT026010-GAL4.DBD}attP2 / + Distal 23B P{UAS-phiC31}attP18 / w[1118]; TI{20XUAS-SPARC2-S-Syn21-CsChrimson::tdTomato-3.1}CR-P40; R21B10-GAL4 / +
Extended Data Fig. 2C	Leg imaginal discs: w[1118]; R38B08-LexA / +; LexAop-mcd8::GFP / + (no expression) w[1118]; +; hh-Gal4 / UAS-mcd8::GFP w[1118]; mid-Gal4 / +; UAS-mcd8::GFP / + w[1118]; +; dpp-Gal4 / UAS-mcd8::GFP w[1118]; dac-Gal4 / +; UAS-mcd8::GFP / + w[1118]; +; rn-Gal4 / UAS-mcd8::GFP w[1118]; ap-Gal4 / +; UAS-mcd8::GFP / + w[1118]; dac-Gal4 / +; UAS-mcd8::GFP / + w[1118]; +; rn-Gal4 / UAS-mcd8::GFP Adult vnc and leg: w[1118]; R38B08-LexA / +; LexAop-mcd8::GFP / + w[1118]; hh-Gal4 / UAS-flp; LexAop>stop>mcd8::GFP / R38B08-LexA w[1118]; mid-Gal4 / UAS-flp; LexAop>stop>mcd8::GFP / R38B08-LexA w[1118] ; LexAop>stop>mcd8::GFP / UAS-flp; dpp-Gal4 / R38B08-LexA w[1118]; dac-Gal4 / UAS-flp;LexAop>stop>mcd8::GFP / R38B08-LexA w[1118] ; LexAop>stop>mcd8::GFP / UAS-flp; rn-Gal4 / R38B08-LexA w[1118]; ap-Gal4 / UAS-flp; LexAop>stop>mcd8::GFP / R38B08-LexA
Extended Data Fig. 4A	P{UAS-phiC31}attP18 / w[1118]; P{y[+t7.7] w[+mC]=R77C10-p65.AD}attP40/ TI{20XUAS-SPARC2-S-Syn21-CsChrimson::tdTomato-3.1}CR-P40; P{y[+t7.7]

	w[+mC]=VT026010-GAL4.DBD}attP2 / +
Extended Data Fig. 4B	P{UAS-phiC31}attP18 / w[1118]; TI{20XUAS-SPARC2-S-Syn21-CsChrimson::tdTomato-3.1}CR-P40; R21B10-GAL4 / +

Acknowledgements and Support

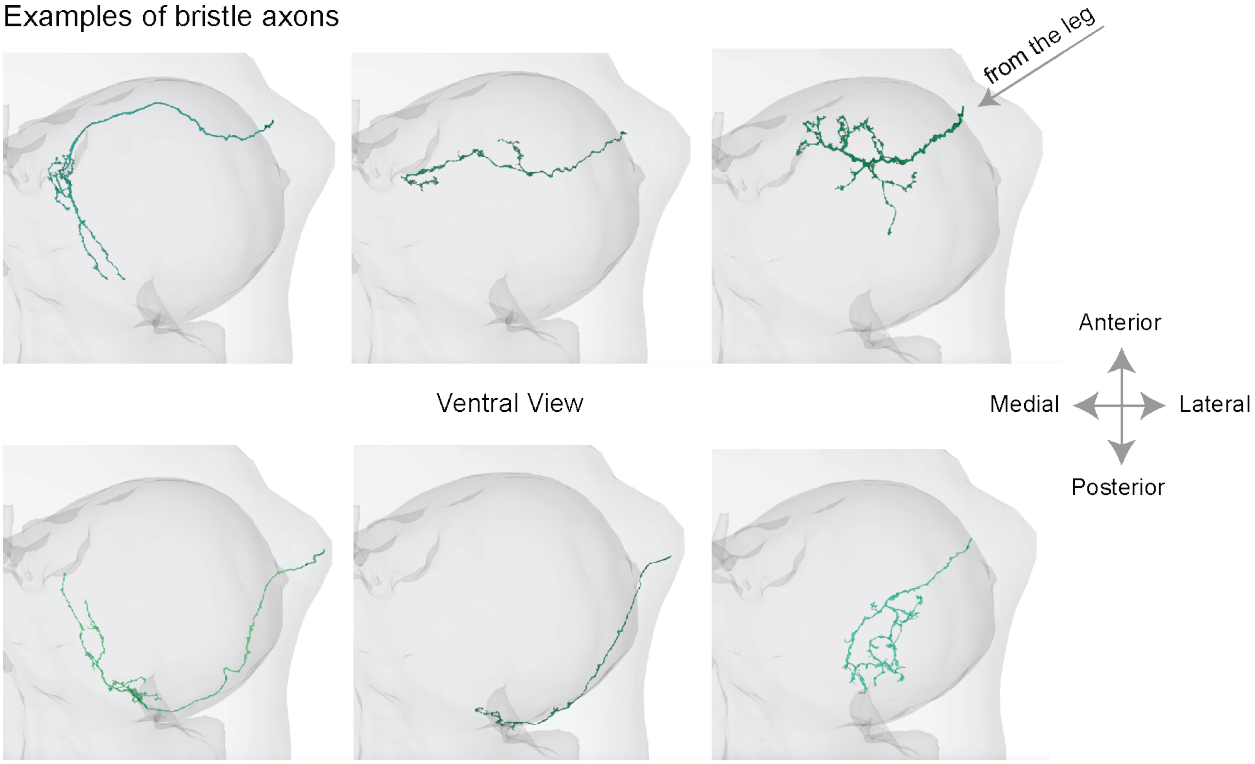
We thank members of the Tuthill Lab for technical assistance and continual feedback on the manuscript. We thank Elizabeth C. Marin and Haluk Lacin for their assistance with hemilineage identification for cells in FANC. We thank Casey Schneider-Mizell, Sven Dorkenwald, and Ben Pedigo for their thoughtful feedback on the manuscript, and Andy Seeds and Steffi Hampel for helpful discussions. We thank Igor Siwanowicz for sharing his blender model and his help understanding bristle distributions on the leg. We thank Kiet Tran for assistance proofreading neurons. L.E was supported by the Ruth L. Kirchstein Fellowship (F31NS134135) from the National Institute of Health. Other support was provided by the Allen Institute for Brain Science, the National Institutes of Health grants R01NS102333, R01NS128785, and U19NS104655, a Searle Scholar Award, a Klingenstein-Simons Fellowship, a Pew Biomedical Scholar Award, a McKnight Scholar Award, a Sloan Research Fellowship, the New York Stem Cell Foundation, and a UW Innovation Award to J.C.T. J.C.T is a New York Stem Cell Foundation – Robertson Investigator.

Declaration of Interest Statement

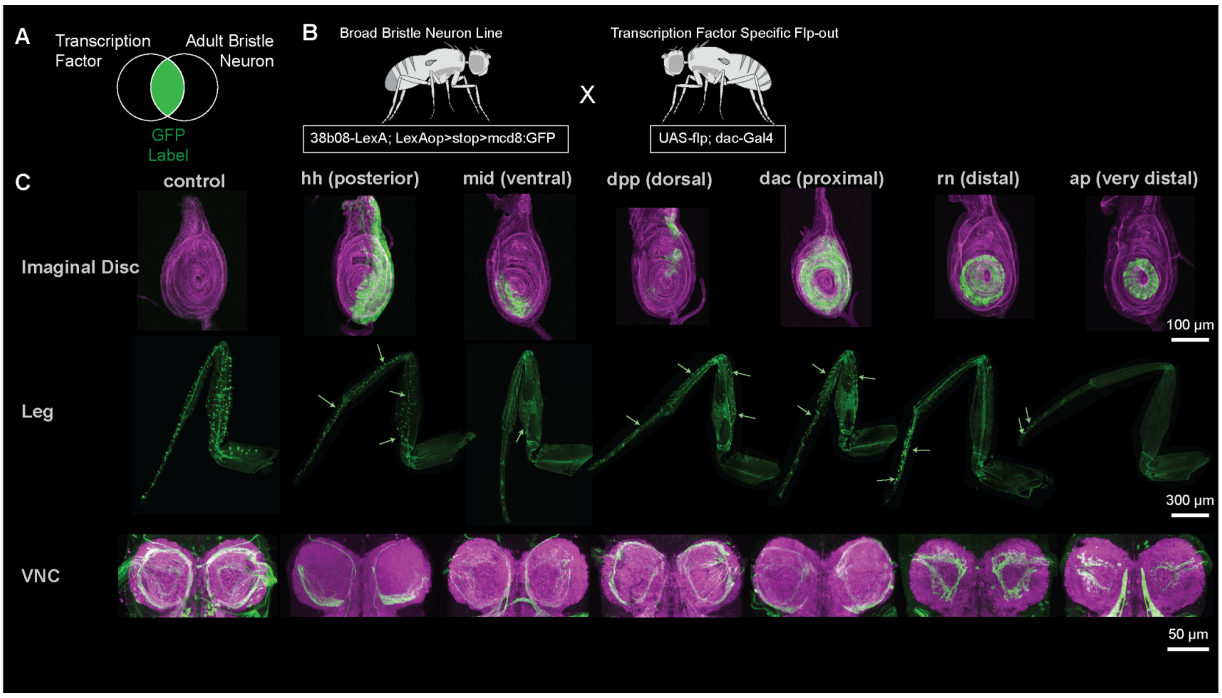
J.C.T. is a member of *Current Biology*'s advisory board.

Extended Data

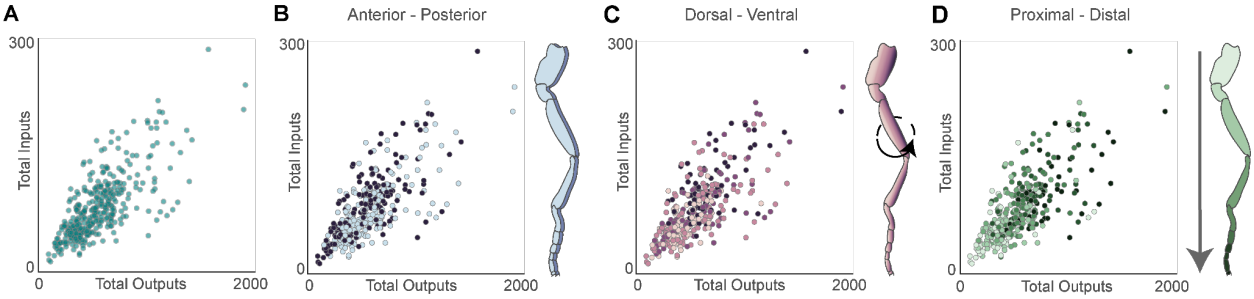
Examples of bristle axons



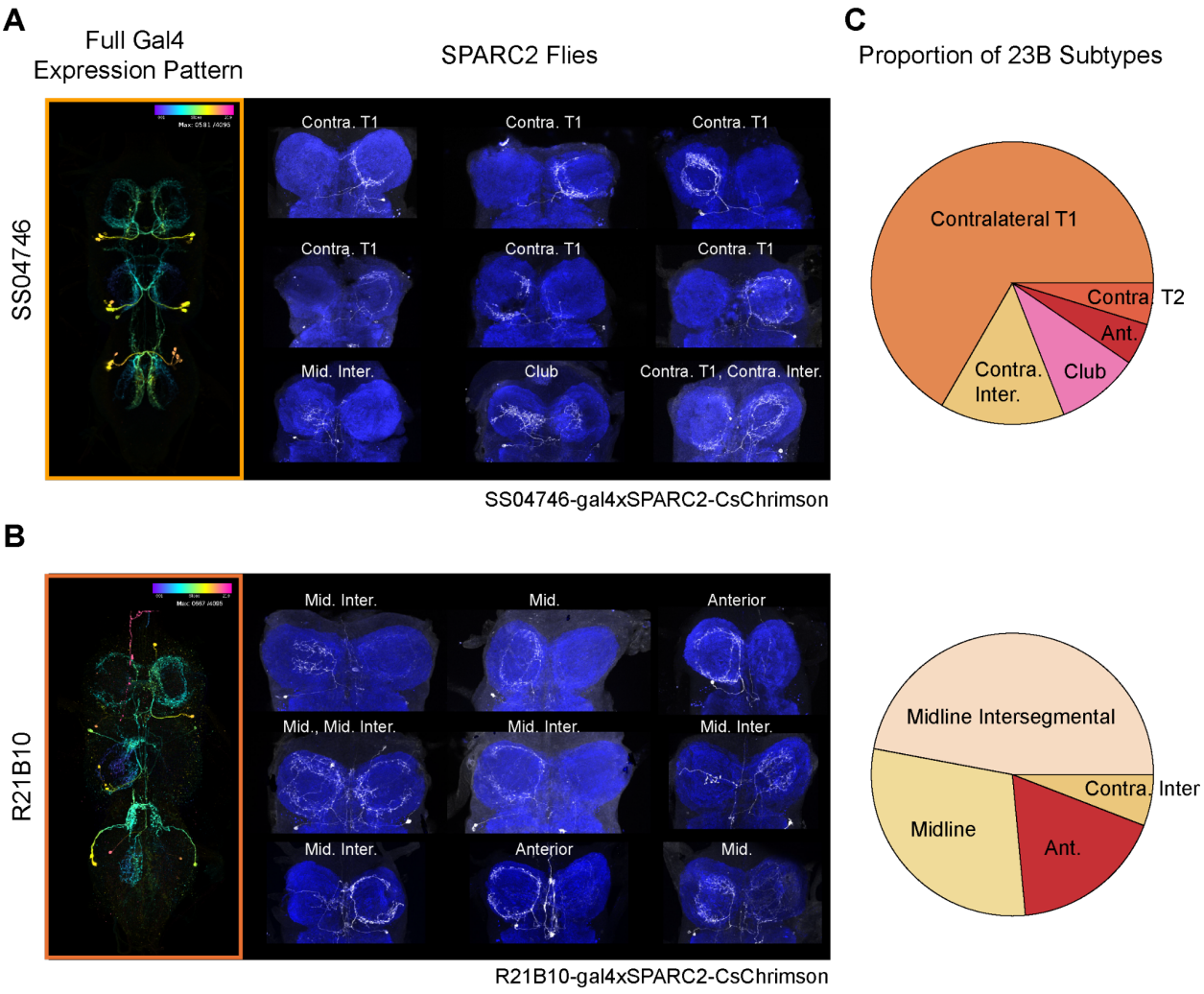
Extended Data Figure 1: Bristle axons vary in morphology. Individual bristle axon morphologies. Three bristle axons that branch anteriorly (top row), and three that branch posteriorly (bottom row). Axons that cross the anterior to posterior border (left), axons that do not cross (middle), and axons that project closer to the center of the left leg neuromere (right).



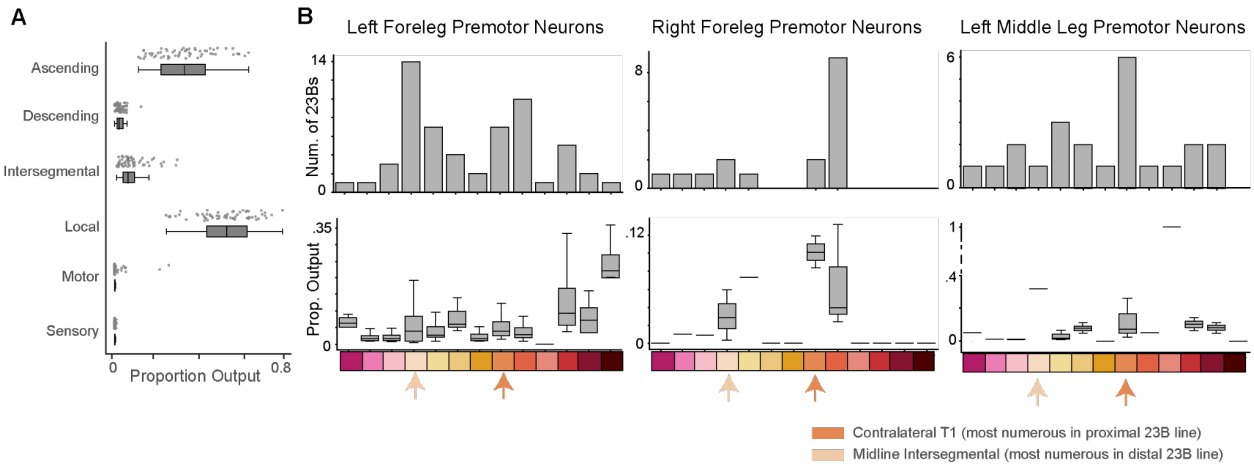
Extended Data Figure 2: GFP expression of bristle neurons driven by coexpression of different transcription factors in the larval leg imaginal disc, leg, and VNC. A) For each line, only bristle cells that express a specific transcription factor will be labeled with GFP. **B)** Example genetic cross. **C)** Shown are maximum intensity projections of cells in the larval leg imaginal pJFRC7-20XUAS-IVS-mCD8::GFP (green) and an antibody against phalloidin (magenta). Bristle neurons in the leg and VNC were labeled with mcd8::GFP (green) and an antibody against the neuropil marker bruchpilot (magenta), green arrows indicate a sample of labeled bristle neurons. From left to right: all bristle neurons labeled by R38B08-LexA alone, bristle neurons that coexpressed *hedgehog* (hh), *midline* (mid), *decapentaplegic* (dpp), *dachshund* (dac), *rotund* (rn), and *apterous* (ap) during metamorphosis.



Extended Data Figure 3: Synaptic input and output counts do not vary somatotopically. A) Number of input and output synapses for each reconstructed bristle axon (teal). Colored by the predicted spatial location on the leg along the **B)** anterior-posterior axis (r^2=4.64e-05), **C)** dorsal-ventral axis (r^2=0.05), **D)** proximal-distal axis (r^2=0.30).



Extended Data Figure 4: Experimental lines SS04746 and R21B10 label different 23B subtypes. A) VNC expression of SS04746 (top) and R21B10 (bottom) from the Janelia library (left). **B)** Example VNCs from sparsified line SS04746-gal4xSPARC2-CsChrimson (top) and R21B10-gal4xSPARC2-CsChrimson (bottom). 23B neurons in the VNC were labeled with mcd8::GFP (white) and an antibody against the neuropil marker bruchpilot (blue). Each neuron was classified by axonal projection pattern. **C)** Proportion of different 23B subtypes in SS04746 (n=21) and R21B10 (n=17).



Extended Data Figure 5: 23B subtypes connectivity onto premotor neurons in T1L, T1R, and T2L. **A)** Proportion of 23B output connectivity onto different neuron classes in the VNC. **B)** 23B subtype connectivity onto premotor pools for the left front leg (T1L), right front leg (T1R), and the left middle leg (T2L). The bar graph represents the number of 23B neurons of each subtype that contact any premotor neurons within each leg neuropil. Boxplots represent the proportion of 23B output synapses onto premotor neurons within each leg neuropil. Color bars represent different 23B subtypes, from left to right: Ascending, Club, Dorsal, Midline Intersegmental, Midline, Contralateral Intersegmental, Ipsilateral T1, Contralateral T, Contralateral T2, Ipsilateral T3, Anterior, Ipsilateral Wing, and Contralateral Wing. Arrows indicate the most prominent subtype in the proximal (SS04746) and distal (R21B10) grooming lines. For all box plots, center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; outliers not shown.

Extended Data Video 1: Proximal 23B activation in headless flies. Example trial for optogenetic activation of proximal 23B neurons (SS04746) expressing CsChrimson. Each trial was 20 seconds in duration, five seconds prestimulus, five seconds with the laser flickering on/off at 5Hz, and 10 seconds post stimulus.

Extended Data Video 1: Distal 23B activation in headless flies. Example trial for optogenetic activation of distal 23B neurons (R21B10) expressing CsChrimson. Each trial was 20 seconds in duration, five seconds prestimulus, five seconds with the laser flickering on/off at 5Hz, and 10 seconds post stimulus.

Extended Data Video 3: Laser activation of empty-SpGal4 in headless flies. Example trial for laser activation of empty-SpGal4 flies with no CsChrimson expression. Each trial was 20 seconds in duration, five seconds prestimulus, five seconds with the laser flickering on/off at 5Hz, and 10 seconds post stimulus.

Is the spatial location of chemosensory stimuli preserved in neural circuits of the fly ventral nerve cord?

All animals use chemosensation to help them maneuver through the world. Yet, how they respond to a given chemical stimulus depends on the context: what chemical was perceived and where it was perceived. In the fly, chemosensory neurons that respond to different stimuli are distributed on the leg. Here we analyzed the synaptic wiring of the leg chemosensory system in the fly ventral nerve cord (VNC) to elucidate chemical modality and spatial location are mapped at the synapse level. We used a volumetric electron microscopy dataset of the entire VNC to reconstruct 106 chemosensory axons from the front left leg of an adult female fly. We clustered these neurons by morphology and identified five distinct groups that projected to specific regions within the fly VNC. Chemosensory neurons exhibited "like-to-like" connectivity wherein neurons synapsed onto other neurons from the same cluster. Neurons from the same cluster also shared largely distinct postsynaptic partners compared to neurons from other clusters. This suggests that chemosensory neurons of the same type converge onto similar neurons while distinct chemosensory neurons feed into distinct parallel circuits. We then mapped the spatial location of chemosensory neurons. We observed that neurons of the same type clustered together while neurons from different types occupied distinct overlapping bands along the proximal-to-distal axis of the leg. Overall, this study provides a circuit mechanism for how flies disentangle what chemical was perceived and where it was perceived on the leg.

Foreword

In Chapter 2 we examined how tactile information is organized in the fly ventral nerve cord. Focused on mechanosensory neurons at the base of hairs all over the fly leg, we used the connectome to predict specific receptive fields of downstream neurons. We showed that the

activation of local neurons with different receptive fields elicited distinct and spatially targeted grooming responses that matched our connectome predictions. While each bristle hair on the fly is innervated by one tactile mechanosensory neuron, a subset of these hairs are considered chemosensory bristle hairs which are innervated by several chemosensory neurons in addition to the mechanosensory neuron at the base.[1,2] Similar to our initial predictions in Chapter 2, previous studies suggest that these chemosensory neurons are also organized somatotopically in the nervous system.[3] As such, the following chapter will examine the extent to which chemosensory axons from the fly leg preserve spatial information with respect to their connectivity and sensory modality.

Introduction

Across the animal kingdom, chemosensation is essential for critical behaviors such as feeding, foraging, protection against toxic compounds courtship and reproduction.[4–6] Defined as the ability to detect chemical compounds, it encompasses sensations like taste, olfaction, and nociception among others. In vertebrates, chemosensory neurons are located all over the body and differentially express receptors that respond to different chemical compounds, thereby giving us the ability to distinguish substances as sweet, bitter, salty, sour, umami, spicy and so on.[7–10] Similarly, insects have what are referred to as chemosensory sensilla that are found all over the body. In Drosophila, these sensilla are found at the base of hairs in the labellum, interior pharynx, distal legs and anterior wing.[10–12] These chemosensory hairs each contain a pore at the tip that is innervated by two to four chemosensory neurons that project their axons in the central nervous system (Figure 1).[1,2] Critically, each chemosensory neuron at a given sensilla expresses different receptors and thereby distinguishes different qualities of taste or pheromones. This mosaic of chemosensory receptor expression in the fly mediates a vast repertoire of behaviors in response to

the world around them. For example, flies elicit appetitive behaviors towards sweet and low salt compounds, whereas they will exhibit avoidance behaviors in response to bitter and high salt compounds.[13–15] Further, the ability to detect different pheromones has been shown to aid in behaviors such as mate selection and prevent cross-species courtship.[16–18] This brings into question how chemosensory information is mapped in the nervous system and how the structure of the circuit facilitates variability in motor responses to different chemosensory stimuli.

In the nervous system, evidence suggests that gustatory neurons, a subclass of chemosensory neurons, are organized topographically by taste modality (e.g. sweet versus bitter). In the brain, neurons expressing sweet receptors project to different areas of the subesophageal ganglion (SOG) than those that express bitter receptors.[3,19] High resolution volumetric electron microscopy (vEM) from the adult brain suggests that neurons of the same taste modality are also more laterally connected receiving on average 39% of their synaptic inputs from other gustatory neurons of the same type.[20] This sensory-to-sensory connectivity has been shown in other sensory systems such as fly olfaction, where olfactory neurons receive on average 40% of their synaptic input from other olfactory neurons projecting to the same glomerulus,[21–23] and mechanosensation where mechanosensory neurons that are closer together in the periphery are more connected synaptically.[24,25] It is thought that this interconnectedness of sensory neurons may be one way for the nervous system to amplify similar signals, repress opposing signals, modulate the gain, or sensitivity to specific compounds.[26–28] Although some qualities of taste are mapped separately within the nervous system (particularly sweet versus bitter) many GRNs express multiple receptors across overlapping populations.[5,19] This variability is thought to mediate increased sensory discernibility and behavioral flexibility within the nervous system.[13,29–31] Investigations in the brain suggest that gustatory neurons of different types contact distinct postsynaptic partners.[32] This

divergence of downstream targets is thought to underlie the behavioral differences observed in response to different gustatory stimuli.

Beyond taste modality, the somatotopic location of where the chemical is sensed on the body can also influence the types of actions and behaviors that are elicited in response. For example, mosquitos that sense sucrose on one side of their body will reliably turn towards the direction of the detected sugar.[33] Similarly, activation of GRNs from different areas on the body elicit different behaviors.[34] The same study that mapped gustatory neurons by taste modality also showed that these neurons were organized by body segment.[3] For example, within the area of the SOG specific to sweet sensation, the neuropil can be further divided into axons projecting from the proboscis, mouthparts and the leg. Further, these different populations of sweet neurons have been shown to have distinct downstream connectivity suggesting a circuit mechanism for spatially distinct behavior.[20] In fact, previous studies suggest that even decapitated flies respond to chemical compounds suggesting that some level of chemosensory mapping must be maintained within the ventral nerve cord (VNC).[35] Therefore, an investigation into the spatial topography of chemosensation within the VNC would provide insight into how the nervous system organizes sensory information to elicit the appropriate behavior in response to variable stimuli.

In the following chapter, we sought to elucidate the topographic organization of leg chemosensory axons in the VNC of an adult female fly. While previous studies suggest that axons from different body segments are separated in the brain (e.g. mouth versus leg),[3] we explored whether the connectivity of these cells suggest a finer map of the leg or if the nervous system integrates information across the entire leg. Here we reconstructed the circuitry of 106 chemosensory neurons from the fly foreleg and we organized these axons into putative clusters based on morphology and synaptic connectivity. We then analyzed the connectivity with respect

to postsynaptic neurons in the circuit and described a potential organization of chemosensory neurons that would preserve spatial and modality specific information from the leg within the VNC. In summary we present a map of chemosensory information that suggests that flies can not only distinguish between different types of chemosensory compounds on the leg but also where on the leg it was sensed.

Results

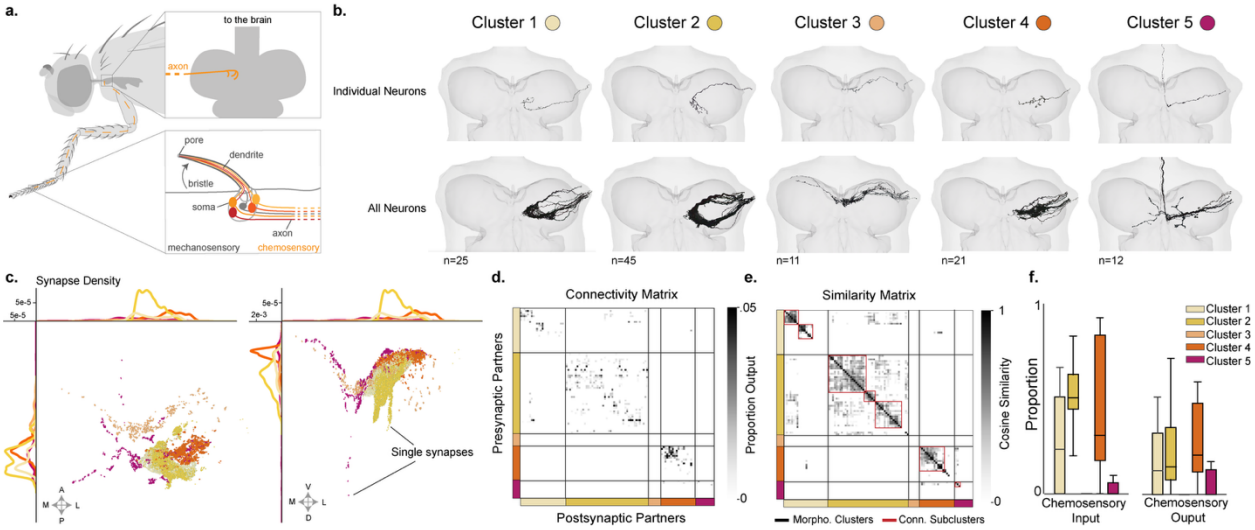


Figure 1: Chemosensory neurons are more synaptically connected to other chemosensory neurons with similar morphology. a) Leg chemosensory neurons are located at the base of some bristle hairs. The dendrite projects to a pore at the tip of the hair (bottom inset) and the axon projects into the VNC (top inset). **b)** Chemosensory axons in the VNC (n=106) clustered into five groups based on morphology (NBLAST, see methods). **c)** Output synapse positions from each morphology cluster occupied distinct areas of the VNC as viewed along the anterior-posterior axis (left) and ventral to dorsal axis (right). **d)** Synaptic connectivity between chemosensory axons (symmetric rows and columns); axons ordered by morphology cluster, color bar represents the proportion of all synaptic output onto individual chemosensory axons. **e)** Cosine similarity of connectivity matrix in d. Morphology clusters outlined in black, connectivity based subclusters outlined in red. Color bar represents the cosine similarity value with 0 denoting no similarity and 1 denoting exact similarity. **f)** Proportion of all input synapses from chemosensory axons for neurons in each morphology cluster (left). Proportion of all output synapses onto chemosensory axons for neurons in each morphology cluster (right). For all box plots, center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; outliers not shown.

Morphology and Sensory Connectivity

To investigate how chemosensory information from the leg is organized in the nervous system, we traced the circuitry of afferent axons in the VNC and their subsequent postsynaptic targets. The cell bodies and dendrites of chemosensory neurons are located at the base of select chemosensory hairs on the leg and the axon projects into the VNC (Fig. 1a). We reconstructed 106 chemosensory axons from the front left leg using a public vEM dataset of the female adult nerve cord (FANC) that contains morphological, ultrastructural, and synaptic information for all the neurites within the VNC.[36–38] The neurons were identified as chemosensory based on their unique morphologies and ventral projection patterns (see methods).[10–12] Chemosensory neurons exhibit distinct shapes and branching patterns within the VNC thus we used NBLAST to further subdivide these axons into five clusters based on morphology (Fig. 1b, see methods).[39]

As expected from previous work, all the chemosensory neurons projected to the ventral surface in the VNC. However, axons with different morphologies projected to distinct regions within the chemosensory neuropil. Neurons in Cluster 1, 2, and 3 projected to the midline of the VNC near the border between the front left and right leg neuropil, however they differed in the shape or their axons and the location along the anterior to posterior axis. Cluster 3 neurons were closest to the anterior portion of T1L (near the neck connective) whereas neurons from Cluster 1 and 2 were closer to the center of T1L. Neurons in Cluster 4 also projected in the center of T1L but terminated more laterally than those in Clusters 1 and 2. Finally, neurons in Cluster 5 all projected straight across the center of T1L but were distinct in that all of these neurons ascended to the brain through the neck connective (Fig. 1b).

Using the automated synapse predictions[36–38] we observed that this subdivision of the chemosensory neuropil was maintained at the synaptic level as the synapse locations of neurons from different morphological clusters occupy different areas in 3D space (Fig. 1c). Based on previous investigations of chemosensory neurons in the fly brain[20] we expected that neurons would be more synaptically connected with sensory neurons of similar morphology and this is exactly what we observed. Chemosensory neurons in three of the five clusters were most connected to other neurons from the same morphology cluster (Fig. 1d-e). Neurons in cluster 3 and 5 scarcely made synapses onto other chemosensory neurons, however the neurons in cluster 5 all project to the brain and thus they may be synaptically connected outside the bounds of this dataset. It should be noted that while the connectivity of chemosensory neurons to one another largely aligned with the morphological clustering, there were instances where highly connected subclusters emerged within a given morphology cluster (Fig. 1e). Furthermore, this like-to-like connectivity made up a large portion of chemosensory neuron input. Neurons in cluster 2 received on average 50% of their synaptic input from other neurons within cluster 2 (Fig. 1f). This proportion input was much higher than other sensory neurons in the leg such as the tactile bristles presented in Chapter 2 (<10%) and proprioceptors from the femoral chordotonal organ (<1%).[40]

Downstream Connectivity Patterns

There were four main classes downstream of chemosensory neurons: ascending neurons that projected up to the brain, intersegmental neurons which received input from multiple neuropil in the VNC, local neurons which only received input within T1L, and descending axons projected down from the brain (Fig. 2a). Across the entire population of chemosensory neurons, ascending and intersegmental neurons were the most numerous partners, with a small subset of local neurons and a few descending neurons. This was reflected in the relative strength of these connections, as

chemosensory neurons received the largest proportion of their synaptic input from intersegmental and ascending neurons. Neurons in Cluster 4 were an exception in that they received 33% of their inputs from descending neurons as compared to the other clusters which received on average only 9% descending input.

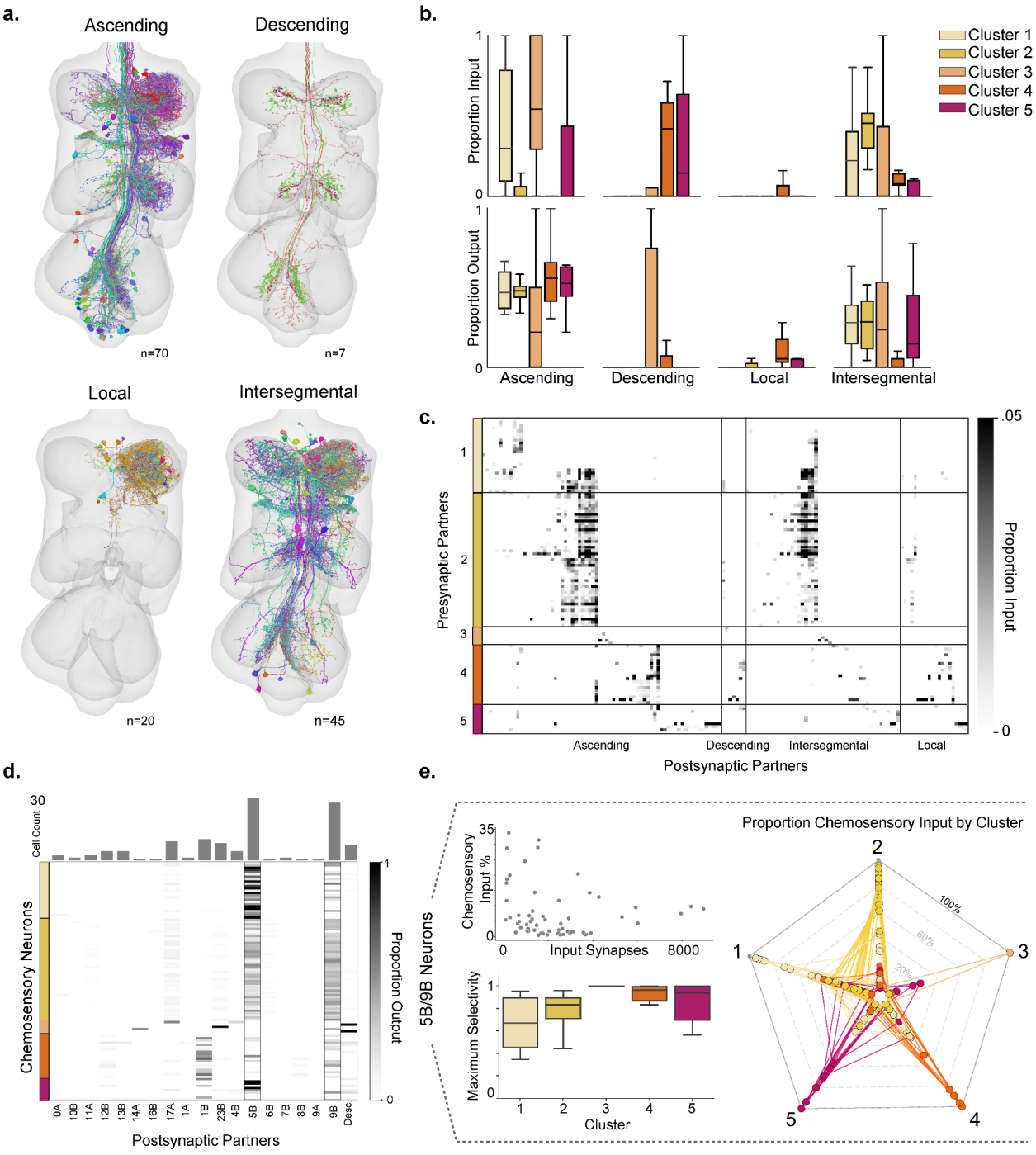


Figure 2: Leg chemosensory axons target distinct populations of ascending and intersegmental neurons. a) Chemosensory axons targeted four classes of non-sensory neurons; Ascending (n=70), Descending (n=7), Local (n=20), Intersegmental (n=45). **b)** Proportion of total synaptic input from each class of neurons (top). Proportion of total synaptic output onto each class of neurons (bottom). **c)** Connectivity between chemosensory neurons ordered by morphology cluster (rows) and postsynaptic partners ordered by class. Colorbar represents proportion input (see methods). **d)** Connectivity between chemosensory neurons ordered by morphology cluster (rows) and postsynaptic partners grouped by their developmental hemilineage. Colorbar represents the proportion of synaptic output for each axon onto each hemilineage. **e)** Total number of input synapses for individual 5B and 9B neurons plotted against the percentage of their total synaptic input that comes from chemosensory neurons from the left foreleg (top left). Maximum selectivity for individual 5b and 9B neurons. Neurons were sorted by the morphology cluster for which they received the most input (see methods) (bottom left). Proportion input of 5B and 9B neurons from different morphology clusters. Radial lines represent percent input in increments of 20%, corners represent distinct morphology clusters, markers represent individual 5B or 9B neurons (right). For all box plots, color represents morphology clusters, center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; outliers not shown.

Looking downstream, 50% of chemosensory synapses targeted ascending neurons alone. Intersegmental, local and descending neurons received 23%, 3%, and 3% respectively. This was in stark contrast with connectivity patterns of mechanosensory neurons in Chapter 2 and other sensory neurons from the leg, all of which largely contact local T1L neurons.[40,41] Note that neurons from cluster 3 were distinct in that 33% of their synapses onto descending neurons. Not only did neurons within the same morphology cluster exhibit similar cell class preferences, many of them synapsed onto the same postsynaptic neurons (Fig. 2c).

Neurons in the VNC are commonly classified by their developmental hemilineage as it is thought that neurons from the same hemilineage release the same neurotransmitters and may share functional roles.[42,43] Thus we sought to see if chemosensory neurons targeted specific hemilineages. To begin answering this question we classified each ascending, intersegmental, and local neuron by the neurodevelopmental hemilineage.[42,43] 5B and 9B neurons, both intersegmental and ascending, were the most numerous and the strongest downstream partners across chemosensory neurons, with 31 and 29 neurons receiving an average of 33% and 22% of the output respectively (Fig. 2d). Unlike the mechanosensory neurons in Chapter 2 which most strongly

targeted cholinergic neurons, both 5B and 9B are thought to be inhibitory, with the third strongest cell-type downstream of chemosensory neurons, 1B, also being inhibitory.[43]

Although over 60% of chemosensory output targeted intersegmental neurons that had dendrites spanning all leg and wing neuropils, we wanted to investigate if the downstream targets were selective to the type of chemosensory neuron they received input from. To do this, we focused on the 5B/9B neuronal population to limit cell-type variability whilst capturing the majority of the synaptic connectivity in the circuit. 5B/9B neurons received between 100-8,000 synapses with an average of 8% of their total synaptic input from chemosensory neurons from the left leg alone (Fig 2e, top). Note that these neurons receive inputs from multiple neuropil and thus the overall proportion of chemosensory input may be higher.

5B/9B neurons that received most of their inputs from chemosensory neurons in cluster 4 or cluster 5, were more selective in that they received little to no input from neurons in other clusters (Fig 2e, bottom). On the other hand, if a neuron received most of its chemosensory input from cluster 1 or 2, it received input from other neurons as well (Fig 2e, bottom). In fact, most of the cross-talk was between neurons in Clusters 1 and 2 (Fig. 2e, right). Few neurons received input from Cluster 3 neurons though there was one neuron that was exclusively postsynaptic to Cluster 3 (Fig. 2e bottom and right).

Somatotopic Mapping

In the brain, chemosensory neurons are organized somatotopically and by modality.[3,19] We wanted to investigate if the somatotopy of chemosensory neurons on the leg was maintained in the VNC i.e. could we further subdivide leg chemosensory neurons along the proximal-distal axis. To do this we applied a similar framework to mapping as described in Chapter 2 (see methods).

Based on spatially restricted labeling of chemosensory sensilla on the leg, we observed that chemosensory neurons that were more distal on the leg, projected closer to the middle of T1L in the VNC (Extended Data Fig. 2). On the other hand, neurons that were more proximal projected further away from the midline. For this reason, we used a reference line along the middle of T1L and measured the distance of every synapse from a chemosensory neuron to that line (Fig. 3a). Synapses from individual axons occupied limited areas of the chemosensory neuropil (Extended Data Fig. 2), thus neurons with an average synaptic distance closer to the reference line were predicted to be more distal on the leg than those with synapses further away (Fig. 3a). From this mapping we predicted that axons from different morphology clusters originated from overlapping yet distinct portions of the leg. Neurons from Clusters 4 and 5 were predicted to be the most distal on the leg, with neurons from Clusters 1 and 2 predicted to be more proximal (Fig. 3b). Finally, neurons from Cluster 3 were predicted to be the most proximal with no overlap in the spatial distributions of the other chemosensory neurons (Fig. 3b).

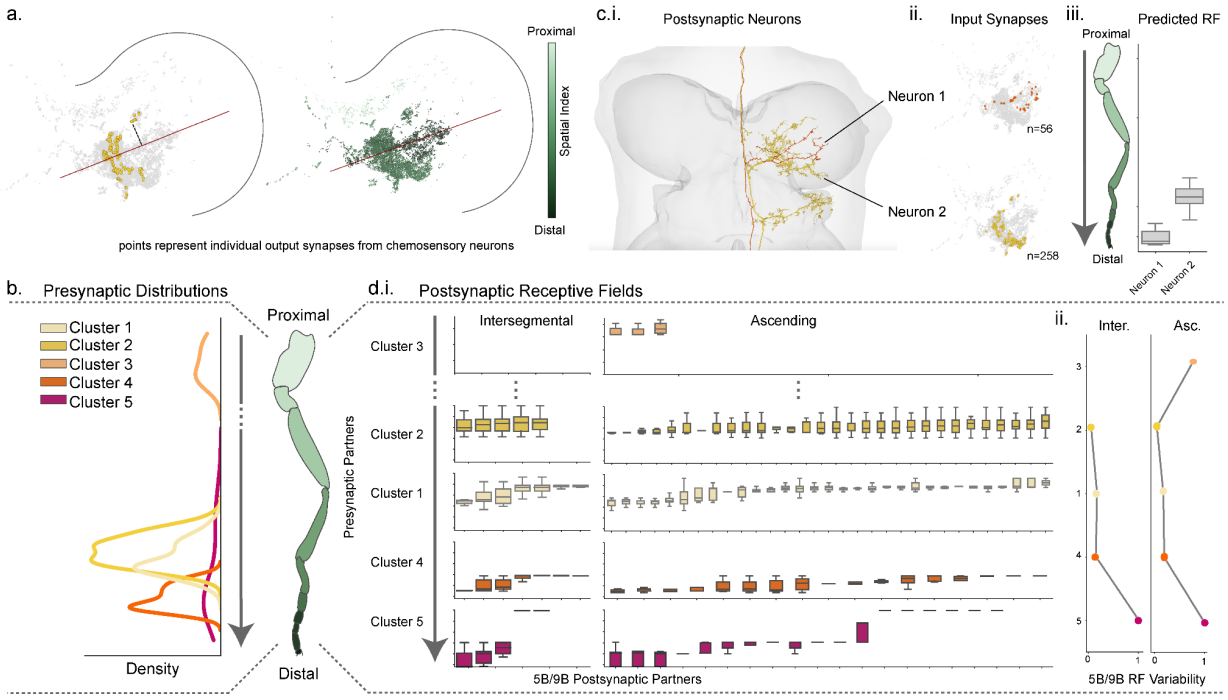


Figure 3: Somatotopy of chemosensory neurons on the leg. a) Output synapses from reconstructed chemosensory neurons in the VNC (gray points). Output synapses from one example chemosensory neuron from Cluster 2 (yellow points). Distance from an example synapse to the reference line (red) shown in dashed black line (left). All output synapses colored by the average synapse distance from the reference line (red) for each chemosensory axon (right) For each axon, the closer the average distance from the reference line the more distal the predicted position on the leg. **b)** Predicted distribution of chemosensory neurons along the proximal to distal axis of the leg, neurons grouped by morphology cluster. **c) i)** Receptive fields of postsynaptic 5B or 9B neurons. Two example neurons from cluster 4 (Neuron 1) and cluster 2 (Neuron 2). **ii)** Chemosensory synapses into Neuron 1 (top) and Neuron 2 (bottom). **iii)** Predicted receptive field along the proximal to distal axis of the leg based on the average synapse distance of all input synapses for each neuron. **d) i)** Receptive field for each 5B or 9B neuron (Intersegmental left, Ascending right) that receives input from chemosensory axons ordered and colored by morphology clusters. Morphology clusters ordered by average position on the leg in descending order. **ii)** Normalized receptive field variability among 5B and 9B neurons within each morphology cluster with 1 representing the most variability. For all box plots, center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; outliers not shown.

To investigate whether these differences in predicted spatial distribution were maintained within the circuit we mapped that spatial receptive field of individual 5B/9B neurons based on the chemosensory input that they received (see methods). Neurons that received synaptic input from predicted distal chemosensory neurons were predicted to have receptive fields that were more distal compared to those that received input from proximal chemosensory neurons (Fig 3c-d.i.). Based on the different spatial distributions of the morphology clusters we wanted to probe the spatial variability of post-synaptic neurons within a cluster. For example, if a neuron receives input from a given Cluster, does it have a similar receptive field to other neurons downstream of the same Cluster or do different neurons have different receptive fields? By comparing the predicted receptive fields within a morphology cluster, we found that input from Cluster 2 neurons spatially homogenous across the intersegmental and ascending populations (Fig. 3d). On the other hand, neurons downstream of Cluster 5 received input from spatially distinct subsets of Cluster 5 neurons (Fig. 3d). Overall, input from neurons in Clusters 3 and 5 were the most spatially variable, Clusters 1 and 4 received less variable input and input from Cluster 2 had little to no variability (Fig. 3d.ii).

Discussion

In this study we explored the variations in morphology and synaptic connectivity of chemosensory neurons from the left foreleg of an adult female fly. We reconstructed 106 axons in the FANC connectome dataset and organized them into 5 clusters based on morphological similarity. We found that synapses from these morphology clusters occupied different areas within the chemosensory neuropil and chemosensory neurons were more interconnected with neurons in the same morphology clusters suggesting like-to-like sensory connectivity. The postsynaptic partners of these axons were also distinct, however there are similarities in overall connectivity patterns. For example, ascending and intersegmental neurons received the most synaptic input across four of the five morphology clusters and the most numerous and strongest downstream cell-type to these sensory neurons were 5B and 9B neurons, which are both thought to be inhibitory. While the input to these 5B/9B neurons was broadly selective for specific morphology clusters, the input was more intermixed for cells that were postsynaptic to Clusters 1 or 2.

Finally, we investigated the spatial mapping of leg chemosensory axons. Based on synapse location, we predicted that neurons from different morphology clusters would be located in distinct but overlapping bands along the proximal to distal axis of the leg. Four of the five clusters were predicted to be on the distal leg whereas one cluster was predicted to project from a more proximal portion of the leg. Utilizing this spatial mapping, we quantified the receptive fields of postsynaptic 5B and 9B neurons and found that neurons postsynaptic to Clusters 1,4 and 5 had more variable receptive fields than those postsynaptic to Cluster 1 or 2 suggesting that the spatial resolution of input from these chemosensory neurons would be higher in the former compared to the latter.

From morphology to modality

Extensive research has been done on the morphology and projection patterns of chemosensory neurons that respond to different chemical modalities such as sweet, bitter, water, and pheromones among others.[4,5,44] However, much of that work utilized genetic labeling techniques that either label sparse subsets or entire populations and thus the exact morphologies of all chemosensory neurons with respect to modality is not known. Here we had the unique opportunity to reconstruct individual neurons across the chemosensory population from the front left leg nerve with 4nm resolution and thereby characterized each neuron by its specific morphology and projection pattern. Due to the mismatch in resolution and the lack of receptor expression data we could not definitively classify neurons by modality. That said, the similarities in morphology, like-to-like interconnectedness, and downstream connectivity lead us to believe that the clusters we present here separate well known chemosensory types.

For example, Cluster 2 likely contains pheromone neurons due to the shape and sheer number of these axons. Furthermore, the downstream connectivity of Cluster 2 separates these neurons into two subclusters which may represent a previously described distinction between cholinergic and glutamatergic pheromone neurons on the leg.[18] Neurons in Cluster 1 may contain sweet sensing neurons however the variability in morphology and connectivity suggests that this cluster contains neurons from multiple modalities. Similarly, some of the neurons in Cluster 4 exhibited a straight and symmetric morphology previously attributed to water sensing neurons,[45] however the variability within this cluster included neurons with morphologies that have not been characterized before. All of the neurons in Cluster 5 ascend to the brain and thus based on previous studies we believe that many of these neurons are bitter or sweet sensing neurons.[45] This is supported by variability in the downstream targets within this cluster, however due to the low

synapse counts in the VNC it is likely that these axons are more distinguishable based on their synapses in the brain. Overall, similarities in connectivity between clusters may suggest that neurons of the same modality are found within different morphology clusters or that certain postsynaptic neurons are sensitive to a broader signal such as the valence of the stimulus e.g. whether a given stimulus is aversive or appetitive.[3]

Cluster 3 neurons were different from the rest of the chemosensory neurons in almost every respect. These neurons were labeled as chemosensory for their thin axons, ventral projections within the leg neuropil, they entered the VNC with other chemosensory neurons from the leg nerve and have been previously characterized as such in a connectome dataset of the male adult nerve cord.[46,47] However, their morphology was distinct in that they projected along the anterior edge of leg neuropil to the contralateral neuropil while the rest of the neurons projected closer to the middle. While our spatial mapping predicted that most chemosensory neurons are found in the distal leg, as corroborated by the literature, the spatial prediction for Cluster 3 neurons was far more proximal. Moreover, the rest of the chemosensory neurons projected in a similar direction to the distal mechanosensory neurons described in Chapter 2. On the other hand, Cluster 3 neurons projected to a similar area of the leg neuropil as the anterior proximal mechanosensory neurons.

The synaptic connectivity profile of these neurons was also different from the rest of the chemosensory population. Firstly, they did not exhibit any sensory-to-sensory connectivity whereas other clusters had as high as 50% of their synaptic input from other chemosensory neurons. Notably, Cluster 3 neurons made on average 33% (Fig. 2b) of their synaptic outputs directly onto descending neurons as compared to the other clusters which made little to no synapses onto this class of neurons. This direct connection to descending input from the brain may suggest that these neurons are involved in shorter sensorimotor circuits compared to the other clusters

where most output synapses were onto ascending neurons e.g. that information is sent up to the brain before it reaches the motor circuits in the VNC. Lastly, many of the postsynaptic neurons that received input from Cluster 3 did not receive input from other chemosensory types further suggesting that Cluster 3 neurons provide some unique sensory signal as compared to the rest.

Spatial somatotopy of chemosensory neurons

In this study we sought to understand how chemosensory information is organized in the fly VNC with respect to both the chemosensory modality and the spatial position of the stimulus. At first glance, all the chemosensory neurons project to a similar region of the leg neuropil, they contact the same types of postsynaptic neurons, and these postsynaptic neurons are largely intersegmental or ascending neurons with dendrites in every leg neuropil (unilateral and bilateral). This level of convergence and integration across neuropil might suggest that the spatial resolution of chemosensory signals is at the level of body segments e.g. legs, mouthparts, proboscis etc.[3] An alternate hypothesis could be that spatial information is carried by the mechanosensory neuron at the base of each chemosensory hair. However, we found that mapping synapse location was critical to uncovering a finer spatial map of the leg in several ways. With exact synapse position we were able to see that different groups of axons made synapses within limited regions of the chemosensory neuropil (Fig. 1c). Moreover, despite the homogeneity in cell-type targeting, the individual neurons postsynaptic to each cluster of chemosensory neurons were distinct (Fig. 2c). Our synapse based leg mapping approach predicted that the neurons from different clusters project from distinct but overlapping bands along the leg with the vast majority of axons predicted to come from the distal leg (Fig. 3b). This corroborated anatomical studies of the distribution of different types of chemosensory neurons on the leg.[10,48,49] Finally, by quantifying the receptive fields of

individual postsynaptic neurons, we predicted that there is in fact a finer map of the leg represented within the organization of chemosensory neurons and their downstream targets (Fig. 3).

Does the fly need to distinguish the spatial location of chemosensory stimuli? For example, if a mosquito turns right upon detecting sucrose on its right legs, does the behavior change if it was detected on the proximal tarsus versus the distal tarsus? Just as flies can localize mechanical stimuli to remove pests or dust, one might imagine that a finer resolution of chemosensory information would aid in the removal of chemical irritants as well. In fact, the same chemosensory stimulus can result in distinct behaviors from an insect depending on the spatial location of the stimulus.[48] While our results posit that the wiring diagram of the chemosensory circuit maintains a spatial map of the leg, further investigations into the functional role of this mapping are necessary to understand how this spatial organization directs the behavior of the fly.

References

1. Stocker, R. F. The organization of the chemosensory system in Drosophila melanogaster: a rewiew. *Cell Tissue Res.* **275**, 3–26 (1994).

2. Montell, C. A Taste of the Drosophila Gustatory Receptors. *Curr. Opin. Neurobiol.* **19**, 345–353 (2009).

3. Wang, Y. *et al.* Anatomical, physiological and molecular properties of Martinotti cells in the somatosensory cortex of the juvenile rat. *J. Physiol.* **561**, 65–90 (2004).

4. Dethier, V. G. Chemoreception and the Behavior of Insects. in *Survey of Biological Progress* (ed. Glass, B.) vol. 3 149–183 (Elsevier, 1957).

5. Herrero, P. Fruit fly behavior in response to chemosensory signals. *Peptides* **38**, 228–237 (2012).

6. Philyaw, T. J., Titos, I., Cummins, P. N., Rodan, A. R. & Rothenfluh, A. Drosophila Cocaine Avoidance is Mediated by Peripheral Bitter Gustatory Neurons. 2022.06.22.497211 Preprint at https://doi.org/10.1101/2022.06.22.497211 (2022).

7. Matsumoto, I., Ohmoto, M. & Abe, K. Functional diversification of taste cells in vertebrates. *Semin. Cell Dev. Biol.* **24**, 210–214 (2013).

8. Pallante, L. *et al.* On the human taste perception: Molecular-level understanding empowered by computational methods. *Trends Food Sci. Technol.* **116**, 445–459 (2021).

9. Policarpo, M., Baldwin, M. W., Casane, D. & Salzburger, W. Diversity and evolution of the vertebrate chemoreceptor gene repertoire. *Nat. Commun.* **15**, 1421 (2024).

10. Singh, R. N. Neurobiology of the gustatory systems of Drosophila and some terrestrial insects. *Microsc. Res. Tech.* **39**, 547–563 (1997).

11. Newland, P. L., Rogers, S. M., Gaaboub, I. & Matheson, T. Parallel somatotopic maps of gustatory and mechanosensory neurons in the central nervous system of an insect. *J. Comp. Neurol.* **425**, 82–96 (2000).

12. Pollack, G. S. & Balakrishnan, R. Taste sensilla of flies: Function, central neuronal projections, and development. *Microsc. Res. Tech.* **39**, 532–546 (1997).

13. Lee, Y., Moon, S. J. & Montell, C. Multiple gustatory receptors required for the caffeine response in Drosophila. *Proc. Natl. Acad. Sci.* **106**, 4495–4500 (2009).

14. McDowell, S. A. T., Stanley, M. & Gordon, M. D. A molecular mechanism for high salt taste in *Drosophila*. *Curr. Biol.* **32**, 3070-3081.e5 (2022).

15. Moon, S. J., Köttgen, M., Jiao, Y., Xu, H. & Montell, C. A Taste Receptor Required for the Caffeine Response In Vivo. *Curr. Biol.* **16**, 1812–1817 (2006).

16. Coyne, J. A., Crittenden, A. P. & Mah, K. Genetics of a Pheromonal Difference Contributing to Reproductive Isolation in Drosophila. *Science* **265**, 1461–1464 (1994).

17. Ferveur, J.-F., Cobb, M., Boukella, H. & Jallon, J.-M. World-wide variation inDrosophila melanogaster sex pheromone: behavioural effects, genetic bases and potential evolutionary consequences. *Genetica* **97**, 73–80 (1996).

18. Luo, Y., Talross, G. J. S. & Carlson, J. R. Function and evolution of Ir52 receptors in mate detection in *Drosophila*. *Curr. Biol.* **34**, 5395-5408.e6 (2024).

19. Thorne, N., Chromey, C., Bray, S. & Amrein, H. Taste Perception and Coding in Drosophila. *Curr. Biol.* **14**, 1065–1079 (2004).

20. Engert, S., Sterne, G. R., Bock, D. D. & Scott, K. Drosophila gustatory projections are segregated by taste modality and connectivity. *eLife* **11**, e78110 (2022).

21. Tobin, W. F., Wilson, R. I. & Lee, W.-C. A. Wiring variations that enable and constrain neural computation in a sensory microcircuit. *eLife* **6**, e24838 (2017).

22. Horne, J. A. *et al.* A resource for the Drosophila antennal lobe provided by the connectome of glomerulus VA1v. *eLife* **7**, e37550 (2018).

23. Schlegel, P. *et al.* Information flow, cell types and stereotypy in a full olfactory connectome. *eLife* **10**, e66018 (2021).

24. Hampel, S. *et al.* Convergence of distinct subpopulations of mechanosensory neurons onto a neural circuit that elicits grooming. 2020.06.08.141341 Preprint at

https://doi.org/10.1101/2020.06.08.141341 (2020).

25. Eichler, K. *et al.* Somatotopic organization among parallel sensory pathways that promote a grooming sequence in Drosophila. *eLife* **12**, RP87602 (2024).

26. French, A. S. *et al.* Dual Mechanism for Bitter Avoidance in Drosophila. *J. Neurosci.* **35**, 3990–4004 (2015).

27. Benton, R. The neurobiology of gustation in insect disease vectors: progress and potential. *Curr. Opin. Insect Sci.* **20**, 19–27 (2017).

28. Seada, M. A., Ignell, R., Al Assiuty, A. N. & Anderson, P. Functional Characterization of the Gustatory Sensilla of Tarsi of the Female Polyphagous Moth Spodoptera littoralis. *Front. Physiol.* **9**, (2018).

29. Dahanukar, A., Lei, Y.-T., Kwon, J. Y. & Carlson, J. R. Two Gr Genes Underlie Sugar Reception in Drosophila. *Neuron* **56**, 503–516 (2007).

30. KWON, J. Y., DAHANUKAR, A., WEISS, L. A. & CARLSON, J. R. A map of taste neuron projections in the Drosophila CNS. *J. Biosci.* **39**, 565–574 (2014).

31. Slone, J., Daniels, J. & Amrein, H. Sugar Receptors in Drosophila. *Curr. Biol.* **17**, 1809–1816 (2007).

32. Shiu, P. K. *et al.* A Drosophila computational brain model reveals sensorimotor processing. *Nature* **634**, 210–219 (2024).

33. Pappas, L. G. & Larsen, J. R. Gustatory mechanisms and sugar-feeding in the mosquito, Culiseta inornata. *Physiol. Entomol.* **3**, 115–119 (1978).

34. Dethier, V. G. *The Hungry Fly: A Physiological Study of the Behavior Associated with Feeding*. 489 (Harvard U Press, Oxford, England, 1976).

35. Yanagawa, A., Guigue, A. M. A. & Marion-Poll, F. Hygienic grooming is induced by contact chemicals in Drosophila melanogaster. *Front. Behav. Neurosci.* **8**, 254 (2014).

36. Phelps, J. S. *et al.* Reconstruction of motor control circuits in adult Drosophila using automated transmission electron microscopy. *Cell* **184**, 759-774.e18 (2021).

37. Azevedo, A. *et al.* Connectomic reconstruction of a female Drosophila ventral nerve cord. *Nature* **631**, 360–368 (2024).

38. Lesser, E. *et al.* Synaptic architecture of leg and wing premotor control networks in Drosophila. *Nature* **631**, 369–377 (2024).

39. Costa, M., Manton, J. D., Ostrovsky, A. D., Prohaska, S. & Jefferis, G. S. X. E. NBLAST: Rapid, Sensitive Comparison of Neuronal Structure and Construction of Neuron Family Databases. *Neuron* **91**, 293–311 (2016).

40. Lee, S.-Y. J., Dallmann, C. J., Cook, A., Tuthill, J. C. & Agrawal, S. Divergent neural circuits for proprioceptive and exteroceptive sensing of the Drosophila leg. *Nat. Commun.* **16**, 4105 (2025).

41. Lesser, E., Moussa, A. & Tuthill, J. C. Peripheral anatomy and central connectivity of proprioceptive sensory neurons in the Drosophila wing. 2025.05.29.656810 Preprint at https://doi.org/10.1101/2025.05.29.656810 (2025).

42. Harris, R. M., Pfeiffer, B. D., Rubin, G. M. & Truman, J. W. Neuron hemilineages provide the functional ground plan for the Drosophila ventral nervous system. *eLife* **4**, e04493 (2015).

43. Lacin, H. *et al.* Neurotransmitter identity is acquired in a lineage-restricted manner in the Drosophila CNS. *eLife* **8**, e43701 (2019).

44. Scott, K. Gustatory Processing in Drosophila melanogaster. *Annu. Rev. Entomol.* **63**, 15–30 (2018).

45. Kim, H., Kirkhart, C. & Scott, K. Long-range projection neurons in the taste circuit of Drosophila. *eLife* **6**, e23386 (2017).

46. Marin, E. C. *et al.* Systematic annotation of a complete adult male Drosophila nerve cord connectome reveals principles of functional organisation. *eLife* **13**, (2024).

47. Takemura, S. *et al.* A Connectome of the Male Drosophila Ventral Nerve Cord. *eLife* **13**, (2024).

48. Joseph, R. M. & Heberlein, U. Tissue-Specific Activation of a Single Gustatory Receptor

Produces Opposing Behavioral Responses in Drosophila. *Genetics* **192**, 521–532 (2012).

49. Ling, F., Dahanukar, A., Weiss, L. A., Kwon, J. Y. & Carlson, J. R. The Molecular and Cellular Basis of Taste Coding in the Legs of*Drosophila*. *J. Neurosci.* **34**, 7148–7164 (2014).

Methods

Morphology Assignment

Morphological assignments were based on characterization of both previous studies in the literature as well as a collection of sparsely labeled cells that we collected. Based on these two sources, chemosensory neurons were defined as thin axons that entered the VNC through the front left leg nerve and projected along the ventral surface. All our chemosensory annotations were corroborated by similar annotations completed from a public dataset of the male adult nerve cord.[46,47] This resulted in 106 axons which were subsequently reconstructed. Reconstruction involved manual editing of the automated segmentation to remove any falsely merged segments and add any false split branches. Due to the variability in shape and branching patterns among chemosensory neurons, we used NBLAST to agnostically cluster the neurons by morphology.[39] Briefly, this method uses the skeletonized representation of neurons to compute a distance matrix based on the position and branching pattern of each neuron. We then applied Agglomerative Clustering to achieve five distinct morphological clusters.

Cell Identification

All neurons were assigned three levels of labels, flow class, hemilineage, and putative neurotransmitter. The flow class assignment denotes the overall projection pattern of each individual cell. Ascending neurons have a cell body around the VNC and send at least one projection to the brain through the neck connective. Descending neurons consisted of sole axons that descended through the neck connective. Intersegmental neurons had a local soma, local axon, and had a dendritic arbor that spanned more than one neuropil e.g. dendritic arbors in T1L and T2L. Local neurons were similar, but their dendritic arbor was restricted to only one neuropil.

Finally, all axons incoming from peripheral nerves were labeled as sensory neurons. Hemilineage labels were assigned by morphology and neurite bundle. These annotations were cross referenced with those assigned in another public dataset from the male ventral nerve cord to ensure consistent labeling.[46,47] Finally, a putative neurotransmitter was assigned based on previous literature that experimentally quantified neurotransmitters released for each hemilineage.[43]

Connectivity Analyses

All connectivity-based analyses utilized the predicted synapse counts and positions made available with the FANC dataset.[36–38] Due to the homogeneity in synapse size, the raw synapse count between any two neurons was used as the connection weight. To minimize the impact of false positive synapses and very weak connections, all connections were filtered with a synapse threshold of >= 4 synapses. Note that for all the instances where proportion input or output was calculated, the total unfiltered synapse count was used in the denominator to more closely reflect the true connection weight.

Similarity matrices were calculated using the cosine similarity of the weighted connectivity matrix. All matrix rows were ordered by morphology cluster and cosine similarity (Fig. 1d-e, Fig. 2c-d). If the columns in a matrix represented chemosensory neurons, they were also ordered by morphology cluster and cosine similarity (Fig. 1d-e). In matrices where the columns represented postsynaptic targets, columns were ordered by cell class (ascending, descending etc.), hemilineage, and presynaptic cluster preference (i.e. the cluster it received the most input from) (Fig. 2c-d).

Selectivity Index

To calculate the selectivity index of 5B/9B neurons (Fig. 2e) each neuron was categorized based on the chemosensory cluster that it received the most input from. For example, a neuron that

received 0.6, 0.2, 0.2 input from clusters 2, 4 and 5 respectively would fall under cluster 2. The reported distributions of selectivity accounted for the maximum selectivity of each neuron (Fig. 2e bottom) while the breadth and variability of chemosensory input was reported separately (Fig. 2e right).

Spatial Mapping

To map the proximal to distal somatotopy of the leg in the VNC we adapted the methods described in Chapter 2. A reference line was placed in the middle of the left foreleg neuropil, T1L, that ran parallel to the angle from the leg nerve towards the middle of the VNC that is observed by all bristle neuron axons (chemosensory and tactile). For each chemosensory neuron, we calculated the distance between the reference line and each individual synapse. To ensure that individual axons referred to only one location on the leg, each axon was represented by the average distance of its synapses. All distances were capped by the 5th and 95th percentile to account for potential outliers. As all the distances are calculated in voxel space and represented a relative position, we then normalized the distances by the 95th percentile so that neurons with a value closer to 1 represented the most proximal neurons and values closer to 0 represented more distal neurons.

Receptive Fields

Receptive fields of postsynaptic neurons were calculated with respect to the proximal-to-distal axis of the leg. For each postsynaptic neuron we calculated the weighted spatial prediction of all the chemosensory input synapses. Thus, the receptive field was represented by the distribution of input synapses multiplied by the predicted spatial location on the leg of each sensory axon. For example, a neuron that received 10 input synapses from an axon with a spatial index of

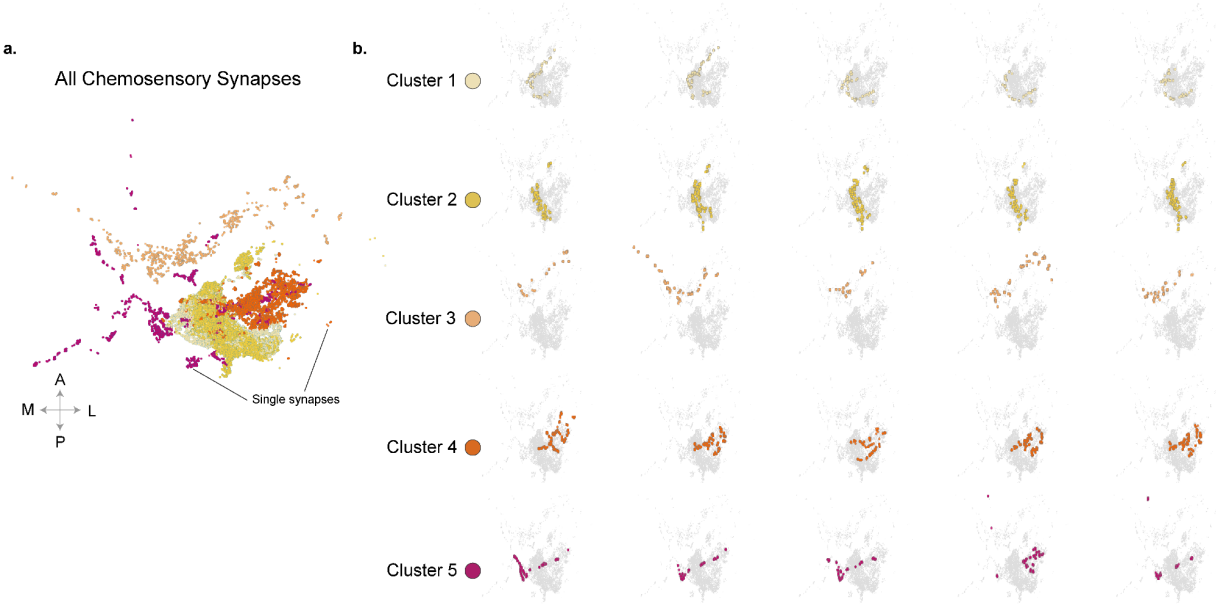
0.5 and 20 synapses from an axon with a spatial index of 0.1 had a receptive field of 10(0.5) + 20(0.1).

To probe the spatial variability within clusters the receptive fields reported in Fig. 3d were filtered for within cluster connectivity only. This means that a neuron that received input from multiple chemosensory clusters was represented once in each cluster and the receptive field was limited to the synaptic input from that cluster alone.

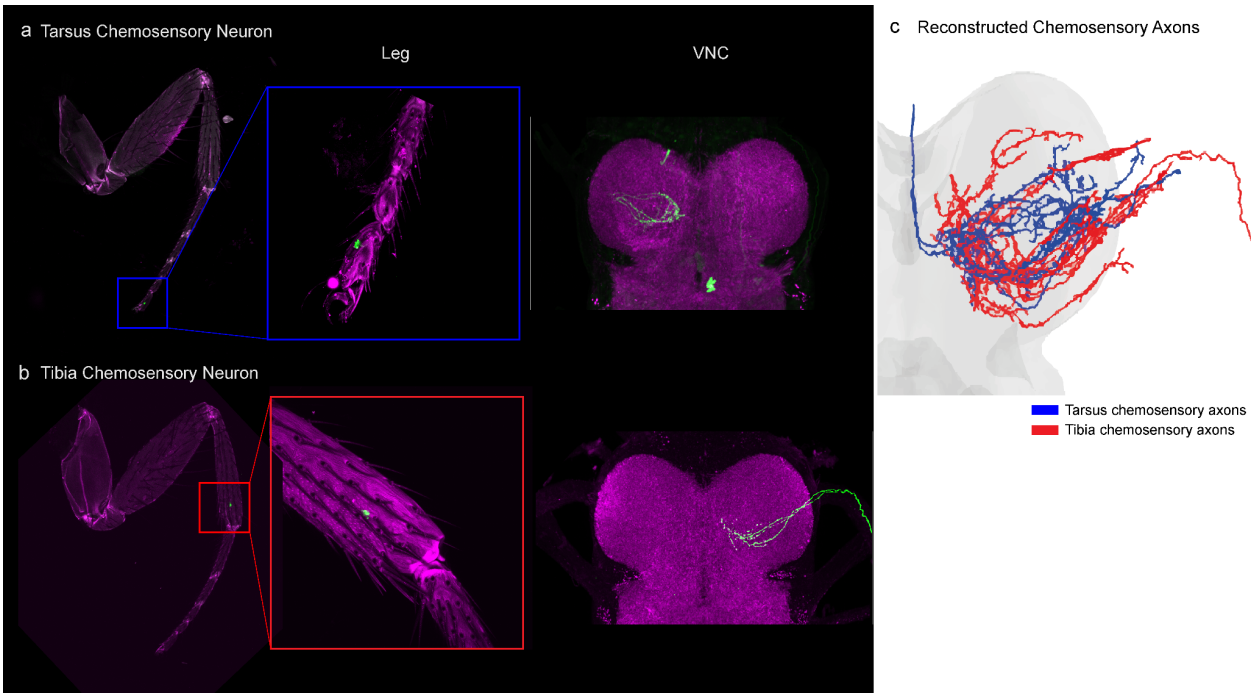
Receptive Field Variability

To measure the RF variability downstream of different chemosensory clusters we took the mean spatial index of the receptive field of each postsynaptic neuron and calculated the standard error of the mean across all postsynaptic neurons within a chemosensory cluster. To compare the SEM across clusters we normalized each value by the maximum variation for intersegmental and ascending neurons respectively (Figure 3dii).

Extended Data



Extended Data Figure 1: Synapses from individual neurons are spatially limited within the chemosensory neuropil. a) Points represent all output synapses from chemosensory neurons colored by morphology cluster. **b)** Each plot represents synapses from one example neuron (colored points) compared to all output synapse points (gray). Five example neurons shown for each cluster.



Extended Data Figure 2: Chemosensory neurons from the tarsus project centrally relative to tibia chemosensory neurons. a) Confocal image of the fly foreleg leg stained with rhodamine (magenta), chemosensory neurons labeled with GFP (green) from the fifth tarsal segment. (left). Inset represents 40X zoom of the same image. VNC from the same fly, GFP expressing chemosensory axon (green), and an antibody against the neuropil marker bruchpilot (magenta)(right). **b)** Confocal image of the fly foreleg leg stained with rhodamine (magenta), chemosensory neurons labeled with GFP (green) from the distal tibia (left). Inset represents 40X zoom of the same image. VNC from the same fly, GFP expressing chemosensory axon (green), and an antibody against the neuropil marker bruchpilot (magenta)(right) **c)** Axon reconstructions from a composite of 16 VNCs with single axon GFP labeling. Axons from tarsus chemosensory neurons (blue) and tibia chemosensory neurons (red).

Conclusions and Future Directions

In this dissertation, I present three investigations of neural circuit structure. In the first chapter, I explore the cell-type diversity of circuits in the mouse visual cortex. At its core, this work presents an efficient, scalable solution in response to a technical hurdle – automated cell-type predictions for large-scale datasets that are agnostic to proofreading. Yet it also brings into question the definition of a cell-type. Mammalian neuronal cell-types have long been described by the structure of dendritic arbors, axonal projection patterns, and cell-type specific targeting.[1–9] All features which require near complete reconstruction of any given cell. Here I describe a method that utilizes information around the cell-body alone and predicts commonly accepted cell-types with 91% accuracy. Notably, this view of the cell-body enables efficient search for rare cell-types like the Chandelier cell, a cell-type classically defined by its unique post-synaptic targeting patterns.

From these findings, I propose a new framework. One that builds on the study of entire cell morphologies and acknowledges that cell-type identity may be evident at the compartment level. A framework that recognizes that small segments of a cell contain unique morphological and ultrastructural characteristics that signal something about the cell's role within the circuit. Alternative methods that identify cell-type from small segments of dendrite or axon and identify neurotransmitter from electron microscopy images of individual synapses corroborate this idea that cell-type identity is expressed at multiple levels.[10,11]

Our efforts to identify and describe cell-types are ultimately efforts to characterize the functional role of cells within the system. In this vein, I investigate this relationship between structure and function more directly in Chapter 2. Utilizing the genetic toolkit of the fly,

connectome analyses, and behavioral experiments, I examine how sensory information is transmitted through the fly nerve cord to produce spatially targeted behavior. Knowing that flies could accomplish spatially targeted grooming,[12–18] I reconstructed the tactile bristle neurons in the fly leg and all their downstream targets. Based on the circuit structure, I hypothesize that a specific class of interneurons encode spatial information from the leg. I show that upon activation of these interneurons, the spatially targeted grooming patterns of the fly match the connectome derived receptive fields. Finally, I propose that these interneurons form parallel sensorimotor pathways that may facilitate this spatially targeted behavior. In Chapter 3, I go on to examine the leg chemosensory circuit and postulate how an entirely different circuit structure might also relay spatial information in the fly VNC. While I delve into the biological value and implications of these two circuits below, it is important to note that the speed and precision with which these insights were made were enabled by the connectome.

Just as the first human genome required immense time, collaboration and funding from around the world. Just as the creation of bioinformatics as a field was an attempt to make sense of the genome and this new type of data. I view connectomics as the beginning of a fundamentally new way of doing neuroscience. The initial datasets referenced in this dissertation were made possible by the incredible resilience and motivation of hundreds of scientists and the myriad of federal and private funding sources that enabled the research. I cannot stress enough that many of these datasets are made free to the public; to scientists, to classrooms, to anyone interested in exploring the brain. These datasets are an example of how supporting large-scale, collaborative, moonshot projects can result in exceptionally rich resources that keep on giving - to the scientific community and beyond. As such, it is with pride that I note that the insights in this body of work are a mere drop in the sea of knowledge we have gained from mapping the nervous system and

will continue to gain in the decades to come. With eyes pointed forward, I outline below some of the promising directions and exciting applications of connectome research.

Comparative Connectomics

As the number and diversity of available connectomes has increased (Introduction Fig. 1), researchers are delving into comparative connectomics; examining the extent to which features of neural circuits are present across connectomes and which features are unique per individual. This is evident within the C. elegans and fly neuroscience communities as multiple connectomes from the same species are already available. Studies comparing adult connectomes in C. elegans reported that the more synapses between pairs of neurons, the more stable those edges were across animals. These strong connections accounted for 42% of the edges but 72% of all synapses.[19] Similarly, a comparison of the first two fly brain connectomes reported that strong connections (16% of edges and 79% of all synapses) were 90% likely to be present across both brains.[20] In both C. elegans and the fly, weaker connections were more variable between individuals.[19,20] One potential hypothesis from this is that the strongest connections provide a rudimentary roadmap of the nervous system and that the rest of the connections offer the flexibility to tune, develop, and refine circuits based on an animal's experience. Future research utilizing connectome constrained simulations, targeted perturbation experiments, and comparing larger numbers of connectomes will allow us to test hypotheses like this one and probe the degree to which connectomes are hardwired.

The question of connectome variability extends beyond a single species. We know from the study of evolution and bioinformatics that we share 98% of our genetic code with non-human primates.[21] To what degree is this similarity maintained in the wiring diagram of the nervous system and how does that support our cognitive and behavioral variability? Insects and small

marine animals are a current boon in the study of comparative connectomics across species.[20,22–27] Their complex behavioral repertoires offer interesting biological questions while their small nervous systems are more tractable to analyze. In Drosophila, the central complex is a well-studied area of the brain that facilitates fly heading direction and navigation.[28–32] Interestingly, the central complex is conserved across insects including beetles, butterflies and locusts despite the fact that they all use different navigation strategies.[22] Comparing the wiring diagrams of these central complexes may elucidate which aspects of the circuit are maintained across species and which aspects have evolved to support the divergence in navigation behaviors. These different nervous systems are all incredibly powerful, flexible, and plastic networks that solve complex problems and facilitate life in a range of environments. If we understand the benefits and limitations of different wiring patterns and how certain aspects of a circuit support specific behaviors, we can apply those elements to solve complex problems in our lives today.

Biomimicry of Neural Architectures

Biomimicry is a growing field that utilizes models, systems, and designs from nature to help solve complex problems. These solutions range from pollinating machines that mimic the vibrations of bumblebees and other pollinator strategies,[33] to healable materials that repair themselves after a break like latex forming plants,[34] to sustainable textiles produced by a process that recapitulates the creation of spider silk.[35] As such, can we take inspiration from the architectural solutions found in the nervous system to better design systems more generally? In chapters 2 and 3 I analyzed the structure of two sensory systems that originated from hairs on the left foreleg of the fly. In the tactile system, hundreds of bristle neurons fanned out across the ventral surface of the VNC to create a map of the leg that recapitulates the anatomical distributions (Chapter 2, Fig. 2c). The strongest downstream partners of these tactile neurons were local

excitatory neurons that imbricated the leg map with distinct but overlapping receptive fields (RFs) (Chapter 2, Fig. 3). In turn these excitatory neurons fed into distinct premotor populations across the VNC (Chapter 2, Fig. 5). From an engineering perspective, the degree of overlap in the RFs of the local excitatory neurons may seem redundant. However, based on the behavioral experiments presented in Chapter 2 (Fig. 4), I believe that this redundancy in RFs results in multiple representations of the leg, each independently facilitating different sensorimotor circuits. These parallel streams of overlapping spatial information in turn increase the flexibility in the system and ultimately the behavioral repertoire with which the fly can respond to tactile stimulation.

Conversely, the chemosensory system was highly convergent. Unlike the tactile neurons, the strongest downstream partners of chemosensory neurons were inhibitory intersegmental neurons, many of which projected to the brain. Chemosensory neurons that were morphologically similar converged onto the same downstream neurons. Furthermore, those downstream neurons received input from similar chemosensory neurons in other body segments suggesting even more convergence within the system. Again, contrasting the highly local tactile circuit, chemosensory information was in the brain by the second synapse in the circuit, suggesting a role for centralized or higher order processing of chemosensory signals.

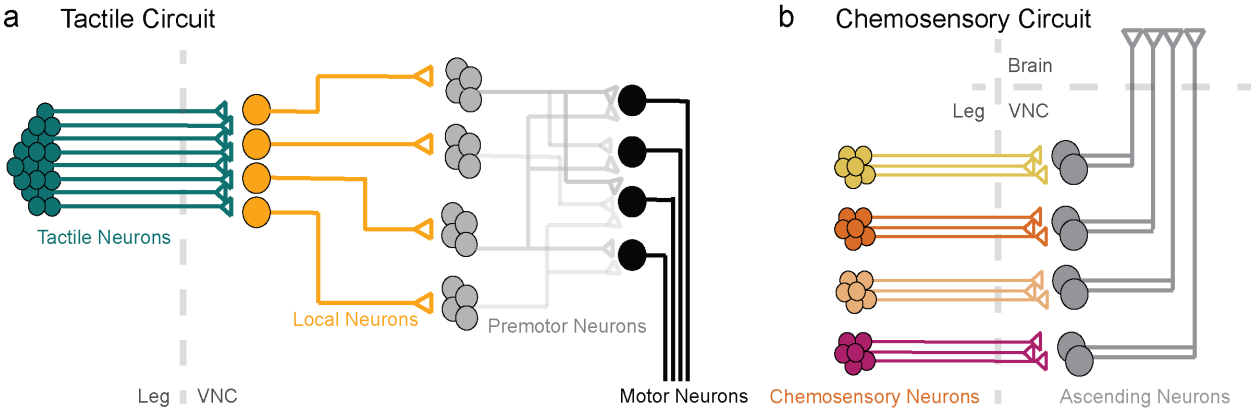


Figure 1: Tactile and chemosensory circuit architectures from the fly leg. a) Local neurons in the VNC receive tactile sensory input from and leg and form parallel sensorimotor circuits. **b)** Chemosensory sensory neurons of the same type converge onto the same ascending neurons in the VNC.

Here we have two systems that originate from the same sensilla. While the structure of the tactile system seems optimized for independent modularity, the structure of the chemosensory circuit pools information from all over the body to a centralized area in the brain. Two different circuit structures for two different sensory systems. This difference makes sense as the tactile and chemosensory systems offer distinct ethological value to the fly. The former facilitates the fly's interaction with the physical world. A modular tactile circuit enables flexible limb movement to overcome physical obstacles, remove pests, or clean different body parts. On the other hand, the chemosensory system detects gustatory and pheromone signals that direct the fly towards appetitive signals and away from aversive signals. Here the fly moves its body more holistically (towards or away from the stimulus), thus pooling sensory information from around the body may be a more efficient network to discern appetitive versus aversive. This contrast in neural circuit structure and function could inspire a more curated approach to network design. Inspired by existing neural circuit architectures, engineers could optimize circuit design for the function and priorities of any given system. Example applications include the design of artificial neural

networks trained to excel in a specific task or brain-computer interfaces aimed to recover loss of function in disease.

Connectomics in health and disease

Beyond basic understanding, connectomics offers a new lens in the study of health and disease. Many of our current therapeutics are broad acting such as brain computer interfaces that act upon entire neuronal populations or drugs that increase global levels of particular chemicals. It is here that the groundbreaking resolution of connectome wiring diagrams can be utilized to design next generation therapies with increased target specificity. Several neurodegenerative diseases such as Amyotrophic Lateral Sclerosis (ALS) and Parkinson's disease result in the loss of specific cell-types in the brain.[36–40] In ALS, motor neurons that project from the motor cortex down to the spinal cord gradually degenerate, leading to complete paralysis of the patient. One of the early markers of this disease is the loss of inhibitory synapses onto motor neurons.[41] However, we have yet to understand the cell-type specificity of this synapse loss and whether the circuits are altered more generally. The dense nature of connectome wiring diagrams offers an opportunity to measure changes in overall synapse density, cell-type specific targeting, and potential rewiring patterns as a function of disease. Comparing the wiring diagrams of healthy versus diseased tissue will not only elucidate which connections are lost but will also provide insight into the structural and morphological changes that occur within diseased circuits. In Alzheimer's, studies suggest that synapse dysfunction in Layer 5 cortical cells occurs well before the physical loss of those connections.[42,43] Utilizing the nanometer resolution of vEM connectomes, scientists can study how these synapses are altered structurally and whether these changes are ubiquitous or if they occur at specific synapses in the Layer 5 circuit. These few examples highlight how comparative connectomics can aid in neuropathological research.

Multimodal connectomics

As with any method, connectomics comes with its set of limitations and drawbacks. Time is one of the inherent limitations of connectome data as they represent a snapshot in time of a dynamic and plastic network. To address this, scientists are developing multimodal techniques that pair electron microscopy with methods like calcium imaging, correlative imaging, x-ray imaging, and expansion microscopy. Foundational datasets from mouse and zebrafinch now provide electron microscopy ultrastructure, along with calcium imaging from neurons, and behavioral measurements during a range of naturalistic behaviors.[44–46] With calcium imaging, researchers gain insight into the functional activity of cells that add context to when and how cells respond within a network.[47] Correlative light and electron microscopy (CLEM) combines molecular insights from light microscopy with the ultrastructural and synaptic connectivity from EM. CLEM enables scientists to utilize results from the light microscopy images to efficiently target EM images of specific subsets of the data. Furthermore, scientists can label proteins associated with specific cell-types, synaptic plasticity, learning and memory, and neurodegenerative diseases to better contextualize the structural circuit within its dynamic environment.[48–50]

Despite the recent growth in dataset size, the volume and scale of connectome data is still technological constrained. Xray nano-holography is one method that contextualizes how high resolution circuits fit within the larger organism. In smaller organisms like Drosophila (larva and adult) XNH has connected motor neurons to muscles in the body, sensory neurons in the periphery to the nerve cord, and the interaction between the nervous system and internal organs in the animal.[51,52] In larger organisms where whole brain connectomes are currently infeasible, correlative XNH and EM studies have verified the surprising degree to which fine ultrastructure of circuits can be measured with XNH alone.[51,53] Groundbreaking advancements in expansion

microscopy have enabled synapse resolution imaging with widely accessible light microscopes, molecular annotation of ultrastructure within connectomes, and automated segmentation of larger and larger datasets.[54] These developments aim to enable whole-brain connectomes in larger organisms such as mice and humans in the near future.

A technical note

In response to the growing demand for connectomes and the increasing robustness of the computational methods required to collect and analyze these datasets, we are already seeing the formation of several private companies to address this need. The outsourcing of different steps of the pipeline such as segmentation and proofreading will ultimately increase the accessibility and diversity of available connectomes. No longer will the collection and curation of connectome data be limited to the few groups who can accomplish the entire pipeline in house. Just as labs across the world send samples out to companies dedicated to sequencing and protein synthesis, I expect a similar rise in third party organizations to support connectome data collection and analysis.

References

1. Gouwens, N. W. *et al.* Classification of electrophysiological and morphological neuron types in the mouse visual cortex. *Nat. Neurosci.* **22**, 1182–1195 (2019).
2. Kanari, L. *et al.* Objective Morphological Classification of Neocortical Pyramidal Cells. *Cereb. Cortex* **29**, 1719–1735 (2019).
3. Scorcioni, R., Polavaram, S. & Ascoli, G. A. L-Measure: a web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies. *Nat. Protoc.* **3**, 866–876 (2008).
4. Costa, M., Manton, J. D., Ostrovsky, A. D., Prohaska, S. & Jefferis, G. S. X. E. NBLAST: Rapid, Sensitive Comparison of Neuronal Structure and Construction of Neuron Family Databases. *Neuron* **91**, 293–311 (2016).
5. Defelipe, J., Hendry, S. H. C., Jones, E. G. & Schmechel, D. Variability in the terminations of GABAergic chandelier cell axons on initial segments of pyramidal cell axons in the monkey sensory-motor cortex. *J. Comp. Neurol.* **231**, 364–384 (1985).
6. Fairén, A. & Valverde, F. A specialized type of neuron in the visual cortex of cat: A Golgi and electron microscope study of chandelier cells. *J. Comp. Neurol.* **194**, 761–779 (1980).
7. Jones, E. G. Varieties and distribution of non-pyramidal cells in the somatic sensory cortex of the squirrel monkey. *J. Comp. Neurol.* **160**, 205–267 (1975).
8. Peters, A. & Kara, D. A. The neuronal composition of area 17 of rat visual cortex. I. The pyramidal cells. *J. Comp. Neurol.* **234**, 218–241 (1985).
9. Sorensen, S. A. *et al.* Correlated Gene Expression and Target Specificity Demonstrate Excitatory Projection Neuron Diversity. *Cereb. Cortex* **25**, 433–449 (2015).
10. Dorkenwald, S. *et al.* Multi-layered maps of neuropil with segmentation-guided contrastive learning. *Nat. Methods* **20**, 2011–2020 (2023).
11. Eckstein, N. *et al.* Neurotransmitter classification from electron microscopy images at synaptic sites in Drosophila melanogaster. *Cell* **187**, 2574-2594.e23 (2024).

12. Vandervorst, P. & Ghysen, A. Genetic control of sensory connections in Drosophila. *Nature* **286**, 65–67 (1980).

13. Corfas, G. & Dudai, Y. Habituation and dishabituation of a cleaning reflex in normal and mutant Drosophila. *J. Neurosci.* **9**, 56–62 (1989).

14. Seeds, A. M. *et al.* A suppression hierarchy among competing motor programs drives sequential grooming in Drosophila. *eLife* **3**, e02951 (2014).

15. Hampel, S., Franconville, R., Simpson, J. H. & Seeds, A. M. A neural command circuit for grooming movement control. *eLife* **4**, e08758 (2015).

16. Hampel, S., McKellar, C. E., Simpson, J. H. & Seeds, A. M. Simultaneous activation of parallel sensory pathways promotes a grooming sequence in Drosophila. *eLife* **6**, e28804 (2017).

17. Yoshikawa, S., Tang, P. & Simpson, J. H. Mechanosensory and command contributions to the *Drosophila* grooming sequence. *Curr. Biol.* **34**, 2066-2076.e3 (2024).

18. Medeiros, A. M., Hobbiss, A. F., Borges, G., Moita, M. & Mendes, C. S. Mechanosensory bristles mediate avoidance behavior by triggering sustained local motor activity in Drosophila melanogaster. *Curr. Biol.* **34**, 2812-2830.e5 (2024).

19. Witvliet, D. *et al.* Connectomes across development reveal principles of brain maturation. *Nature* **596**, 257–261 (2021).

20. Schlegel, P. *et al.* Whole-brain annotation and multi-connectome cell typing of Drosophila. *Nature* **634**, 139–152 (2024).

21. King, M.-C. & Wilson, A. C. Evolution at Two Levels in Humans and Chimpanzees. *Science* **188**, 107–116 (1975).

22. Heinze, S. Variations on an ancient theme — the central complex across insects. *Curr. Opin. Behav. Sci.* **57**, 101390 (2024).

23. Honkanen, A., Adden, A., da Silva Freitas, J. & Heinze, S. The insect central complex and the neural basis of navigational strategies. *J. Exp. Biol.* **222**, jeb188854 (2019).

24. von Hadeln, J. *et al.* Neuroarchitecture of the central complex of the desert locust: Tangential

neurons. *J. Comp. Neurol.* **528**, 906–934 (2020).

25. Verasztó, C. *et al.* Whole-body connectome of a segmented annelid larva. *eLife* **13**, (2024).
26. Jokura, K., Jasek, S., Niederhaus, L., Burkhardt, P. & Jékely, G. Neural Connectome of the Ctenophore Statocyst. 2025.06.26.661250 Preprint at https://doi.org/10.1101/2025.06.26.661250 (2025).
27. Randel, N. *et al.* Neuronal connectome of a sensory-motor circuit for visual navigation. *eLife* **3**, e02730 (2014).
28. Seelig, J. D. & Jayaraman, V. Neural dynamics for landmark orientation and angular path integration. *Nature* **521**, 186–191 (2015).
29. Wolff, T., Iyer, N. A. & Rubin, G. M. Neuroarchitecture and neuroanatomy of the Drosophila central complex: A GAL4-based dissection of protocerebral bridge neurons and circuits. *J. Comp. Neurol.* **523**, 997–1037 (2015).
30. Green, J. *et al.* A neural circuit architecture for angular integration in Drosophila. *Nature* **546**, 101–106 (2017).
31. Turner-Evans, D. *et al.* Angular velocity integration in a fly heading circuit. *eLife* **6**, e23496 (2017).
32. Fisher, Y. E., Lu, J., D'Alessandro, I. & Wilson, R. I. Sensorimotor experience remaps visual input to a heading-direction network. *Nature* **576**, 121–125 (2019).
33. Homepage. *BloomX* https://www.bloomx.ag/.
34. CompPair | Bringing repairability to composites. *CompPair* https://www.comppair.ch.
35. Spintex | Silk Reimagined. *Spintex* https://www.spintex.co.uk.
36. Sreedharan, J. & Brown Jr, R. H. Amyotrophic lateral sclerosis: Problems and prospects. *Ann. Neurol.* **74**, 309–316 (2013).
37. Bruijn, L. I., Miller, T. M. & Cleveland, D. W. UNRAVELING THE MECHANISMS INVOLVED IN MOTOR NEURON DEGENERATION IN ALS. *Annu. Rev. Neurosci.* **27**, 723–749 (2004).

38. Robberecht, W. & Philips, T. The changing scene of amyotrophic lateral sclerosis. *Nat. Rev. Neurosci.* **14**, 248–264 (2013).

39. Hirsch, E. C., Jenner, P. & Przedborski, S. Pathogenesis of Parkinson's disease. *Mov. Disord.* **28**, 24–30 (2013).

40. Dauer, W. & Przedborski, S. Parkinson's Disease: Mechanisms and Models. *Neuron* **39**, 889–909 (2003).

41. Mora, S. & Allodi, I. Neural circuit and synaptic dysfunctions in ALS-FTD pathology. *Front. Neural Circuits* **17**, (2023).

42. Chen, C. *et al.* Disrupted Maturation of Prefrontal Layer 5 Neuronal Circuits in an Alzheimer's Mouse Model of Amyloid Deposition. *Neurosci. Bull.* **39**, 881–892 (2022).

43. Buskila, Y., Crowe, S. E. & Ellis-Davies, G. C. R. Synaptic deficits in layer 5 neurons precede overt structural decay in 5xFAD mice. *Neuroscience* **254**, 152–159 (2013).

44. Consortium, Mic. *et al.* Functional connectomics spanning multiple areas of mouse visual cortex. 2021.07.28.454025 Preprint at https://doi.org/10.1101/2021.07.28.454025 (2021).

45. Lueckmann, J.-M. *et al.* ZAPBench: A Benchmark for Whole-Brain Activity Prediction in Zebrafish. Preprint at https://doi.org/10.48550/arXiv.2503.02618 (2025).

46. Svara, F. *et al.* Automated synapse-level reconstruction of neural circuits in the larval zebrafish brain. *Nat. Methods* **19**, 1357–1366 (2022).

47. Ding, Z. *et al.* Functional connectomics reveals general wiring rule in mouse visual cortex. *Nature* **640**, 459–469 (2025).

48. Begemann, I. & Galic, M. Correlative Light Electron Microscopy: Connecting Synaptic Structure and Function. *Front. Synaptic Neurosci.* **8**, (2016).

49. Han, X. *et al.* Multiplexed volumetric CLEM enabled by scFvs provides insights into the cytology of cerebellar cortex. *Nat. Commun.* **15**, 6648 (2024).

50. Friedrichsen, K. *et al.* Reconstructing neural circuits using multiresolution correlated light and electron microscopy. *Front. Neural Circuits* **16**, (2022).

51. Kuan, A. T. *et al.* Dense neuronal reconstruction through X-ray holographic nano-tomography. *Nat. Neurosci.* **23**, 1637–1643 (2020).

52. Azevedo, A. *et al.* Connectomic reconstruction of a female Drosophila ventral nerve cord. *Nature* **631**, 360–368 (2024).

53. Collins, L. T., Huffman, T. & Koene, R. Comparative prospects of imaging methods for whole-brain mammalian connectomics. *Cell Rep. Methods* **5**, 100988 (2025).

54. Tavakoli, M. R. *et al.* Light-microscopy-based connectomic reconstruction of mammalian brain tissue. *Nature* **642**, 398–410 (2025).

Acknowledgements

In many respects I feel like I fell into neuroscience, graduate school and connectomics. Despite my own attempts, there was never a grand plan or longstanding goal, just a series of small decisions taken along the way. For that, I am eternally grateful for every person who inspired me, mentored me, and nudged me in this direction. This path has neither been clear nor straightforward, but I am grateful to have been on a journey full of challenge, excitement, and so much growth.

Thank you for the University of Washington, the Graduate Program in Neuroscience, the Neurobiology and Biophysics Department, and the Allen Institute for Brain Science for the support and opportunity to pursue this degree. Thank you to the National Institute of Health and the following grants for supporting my work: NIH RF1MH125932, NIH RF1MH117808, and NIH F31NS134135. This research would not have been possible without federal support and so for the love of anything good - let's keep funding science! Thank you to the Washington chapter of the ARCS Foundation and the Weintraub Committee at the Fred Hutch for supporting me and recognizing my contributions as a scientist.

To my committee – Adrienne Fairhall, Bing Brunton, Osama Ahmed, Rachel Wong, and Uygar Sumbul. Thank you for all the time and dedication you have given me. Your invaluable guidance and your contributions to my work have been instrumental to the success of my project and my growth as a researcher.

To John, thank you for pushing me to be the scientist I want to be, in every circumstance. For teaching me to embrace my curiosity and run with it. And for setting an example of what it means to balance work with one's mental, physical, and emotional health. Sometimes a case of powder fever hits harder than the flu…and that's ok.

To Forrest, thank you for your patience and compassion as a colleague, a manager, and now an advisor. You have exemplified what it means to be a 'you first' mentor. Your honesty and thoughtfulness are two of many attributes I hope to emulate in the future. Thank you for always pushing me to think objectively, approach life confidently, and to genuinely consider what door I want to walk through.

Thank you to all the members of the Tuthill lab. You are a big reason I joined this lab. I was picking a lab amidst the isolation of the pandemic and I was craving community. Thank you for being that community and so much more. You have made the lab an inspiring, creative, and an overall fun place to work. I will miss the energy of this group.

To the EM Connectomics team - Wow. You have been busting one record after the next and I have been so lucky to work with such an incredible and innovative group of people. In my eyes, you are all the epitome of science. Thank you for teaching me that the question is never 'is this possible' but 'how can we make it possible'.

To Stephen Smith, your excitement for life and learning is infectious. You not only kept me interested in a career in science but you showed me that it could be fun, exciting, and exhilarating! You have taught me that basking in the wonders of the world is a choice one makes every day and you made it look way too enjoyable not to opt in.

To Synapse Biology, you were the best team anyone could wish for. You've set the standards very high for what a group of colleagues can (should) be and I am so grateful that I joined the workforce with such creative, collaborative and caring people.

Thank you to Ms. Olek, who spent many a free period patiently showing me that I could not only pass IB Chem but that I could - maybe - become a scientist if I ever chose to. To Professor Goddard who gently guided me away from doing everything, everywhere, all at once and inspired me to take the hard road as long as I was learning and opening doors for myself. To Dr. Anne Churchland who took a chance on me and offered me my first scientific research opportunity. To Professor Keane who taught me that brains could physically hurt from thought but that critical thinking was a skill worth the effort.

To my friends who took me out of my bubble – Grad school can feel very isolating and grad school during a pandemic felt even more so. Thank you for feeling close at all times. For riding the waves with me, cheering me on in the highs and pulling me up in the lows. Thank you for always taking me seriously while also preventing me from taking myself too seriously. I have loved growing alongside you all, and now that this meaty chapter is over, I can't wait to see where we go next.

To my in-laws, thank you for your endless support, career advice, and all the wonderful adventures along the way. Ann, I never imagined I would be lucky enough to have such an engaged and excited mother-in-law. Your enthusiasm and interest around my work have inspired me as a scientist and a creative. Making art together and transforming the science with clay, beads, wires, and paint has been incredible fun and I hope we continue this joint venture for years to come.

To my parents Engy and Tarek and to my brother Hassan. I'm not even sure where to start. To my dad, thank you for being a constant inspiration. You instilled in us a persistent drive to learn, to innovate and to impact. Three traits I live by, and which got me here today. Thank you for always supporting me, whatever the venture. Now we can be 'not real' doctors together. To my mom, superwoman! When you do something, you don't just do it well – you set an entirely new standard, you stick the landing, and you make it look easy. Thank you for your endless patience and wisdom in life. Most importantly of course - thank you for teaching us that, figuratively and literally, life is better when you to stop to smell, admire, and eat the roses. To Hassan, I could not imagine a better partner from the start. I feel very lucky that we had so much fun as kids and got to explore so much of the world together as teenagers. Thank goodness you turned out to be such a cool, wise and caring adult because your presence and perspective have made all the difference.

To Ian, I genuinely would not be here without you. Thank you for your patience, your perspective, and your all-around brilliance. Thank you for listening to my ideas and helping me brainstorm; for motivating me to focus and insisting I take breaks. Thank you for being my number one hype man

- always. For the hidden jalapeno chips and for getting rid of all the beans. To properly thank you, I may need to write another dissertation, and I don't think either of us is ready for that. Instead, please remind me to laud your efforts on our next hike, during our next culinary experiment, or any of the many life adventures we have ahead of us. Cheers to us, we made it!

I know that I am on record for saying that it takes a certain kind of stupid to go to graduate school - and I still think it does. But now that I've joined the club, I must say I've learned that it also comes with wondrous excitement, adventurous fun, insatiable curiosity, and a persistence for the truth. Not to mention a lifetime network of inspiring, intelligent, and accomplished colleagues.

Finally, to the trees. For slowing it down, putting it in perspective, and filling it with wonder.