Characterizing chemosensory responses in C. elegans with multi-neuronal imaging

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Department of Physics

have examined a dissertation entitled

Characterizing chemosensory responses of *C. elegans* with multi-neuronal imaging

presented by Albert Lin

candidate for the degree of Doctor of Philosophy and hereby
certify that it is worthy of acceptance.

Signature _____

Typed name: Professor Aravinthan Samuel, Chair

Signature _____

Typed name: Professor Vinothan Manoharan

Signature _____

Typed name: Professor Mei Zhen

Signature _____

Typed name: Professor Steven Flavell

Signature _____
Digitally signed by Sandeep Robert
Datta, MD, PhD
Date: 2021.04.29 17:50:09 -07'00'

Typed name: Professor Sandeep Robert Datta

Signature _____

Typed name: Professor Cengiz Pehlevan

Date: April 29, 2021

Characterizing chemosensory responses of *C. elegans* with multi-neuronal imaging

A DISSERTATION PRESENTED
BY
ALBERT LIN
TO
THE DEPARTMENT OF PHYSICS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
PHYSICS

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
APRIL 2021

Thesis advisor: Professor Aravinthan D.T. Samuel

Albert Lin

Characterizing chemosensory responses of *C. elegans* with multi-neuronal imaging

Abstract

How do animals use their sensory neurons to perceive their chemical environment? How does sensory information get transformed by downstream circuits? And ultimately, how do animals generate behavior in response to chemical cues? We set out to understand how the entire *C. elegans* brain processes chemosensory information. To enable these studies, we developed genetic, experimental, and computational tools which allow for high-throughput acquisition of stimulus-evoked activity in populations of neurons. Observing the dynamics of the chemosensory neurons in response to a broad odor panel, we uncovered numerous previously unreported sensory responses and built a comprehensive picture of the way the chemosensory system in *C. elegans* encodes odor information. To understand whole-brain responses to chemosensory stimuli, we developed a system which labels every neuron in *C. elegans* with a multicolor landmark, allowing us to conduct pan-neuronal imaging with comprehensive neuronal identification. We found widespread stimulus-evoked activity across sensory neurons, interneurons, and motor neurons. We examined the relationship between functional activity and the *C. elegans* connectome, and built connectome-inspired artificial neural network models with pan-neuronal data. The methods we have developed enable new, whole-brain approaches to understanding sensorimotor transformations in *C. elegans*.

Contents

Listing of figures

Acknowledgments

It takes a village to do most anything, and (despite the "lone scientist" narrative so often regurgitated in classrooms) science is no exception. The work presented in this dissertation would not have been possible without the guidance, support, and contributions of many people.

First and foremost, I am incredibly grateful to my advisor, Dr. Aravinthan Samuel. Aravi is a remarkably generous and supportive mentor, who has given me the freedom to explore my scientific interests and has provided me with so many opportunities to grow professionally. He has guided me through the challenges of graduate school with constant optimism and encouragement, and I am grateful for all of the advice he has given me over the years. Aravi is a terrific role model: a successful scientist and educator who leads by example and puts the well-being of his people first. Working with him has been a genuine privilege.

I would like to thank my colleagues in the Samuel Lab, past and present, many of whom have become collaborators and friends: Vivek Ventakachalam, Guangwei Si, Katrin Vogt, Vladislav Susoy, Sahand Rahi, Albert Kao, Daniel Witvliet, Jessleen Kanwal, Luis Hernandez-Nunez, Jacob Baron, David Zimmerman, Helena Casademunt, Core Park, Ben Sorscher, Will Dorrell, Anson Sathaseevan, Alicia Chen, Greg Cain, Nicolas Tan, Raymond Valenzuela, and Dionne Clarke. They have made the lab an open, collaborative, and happy place. In particular, I would like to thank Vivek for mentoring me when I first joined the lab and for teaching me how to build everything from optics to tracking software, Guangwei for our many discussions about microfluidics and odor delivery, and Daniel for so generously answering all of my connectome questions. I would also like to thank Jess for helping to guide me through graduate school—I am grateful for her mentorship and friendship. It has been a pleasure to teach and work with the next generation of graduate students, Helena and Core, and the many undergraduates I have had the privilege to mentor in my time here, Will, Anson, Greg, Nic, and Ray. Their infectious enthusiasm for science constantly reminds me how incredible it is that we get to do the things we do. I thank Dionne for all of her administrative and logistical assistance—nothing in the lab would get done without her.

I am grateful to Lisa Cacciabaudo of the physics department for being such a friendly and helpful resource for all of the students, helping us navigate the department from our first day at Harvard all the way until graduation. I also acknowledge Jacob Barandes for instilling in all of us the importance of good teaching to our profession. I would also like to thank my undergraduate mentors, Thomas Gregor and Hernan Garcia, for teaching me how to be a scientist and for all of their advice.

I consider myself fortunate to have had the opportunity to work with so many great collaborators in my time in graduate school—much of the work presented in this dissertation has come out of

1

Introduction

THE BRAIN EVOLVED to integrate information and produce behaviors that help an animal inhabit the natural world. When faced with complex environments, the whole brain is engaged in many levels of activity, encompassing detection and interpretation of sensory inputs, decision-making, and motor planning and execution. To date, most experiments in neuroscience have focused on recording from single neurons, studying different neurons in different individuals and using these data to piece together the workings of specific circuits. This approach allows for the analysis of individual computations or stereotyped steps of neural processing, and has proven successful in mapping the circuits responsible for reflex behaviors in many animals. However, the more complex tasks that the brain does—identification, decision-making, learning, memory, and motor control—are not reflexes. In these cases, it is reasonable to expect that neurons from multiple circuits will be engaged, and that the activity of a given neuron will depend not just on the environment, but also on both the

animal's behavioral state and internal brain state. Single-neuron experiments cannot simultaneously monitor the state of the rest of the animal's brain, making it difficult to place the activity of a single neuron into the context of the whole brain.

In this dissertation, we sought to understand how the compact brain of the nematode *C. elegans* encodes olfactory information. We chose to study olfaction because it is a sensory modality which is inherently collective in its nature: animals employ a large number of sensory neurons to detect and identify chemical compounds. This makes olfactory coding a problem for which multi-neuronal imaging methods are ideally suited.

Recent technological advances are now allowing whole circuits, and even whole brains, to be recorded simultaneously in single individuals. Across model organisms, we now know that even simple stimuli engage large fractions of the nervous system, not just a small set of neurons. The ability to reliably capture ensemble-level neural activity with single-cell resolution has the potential to revolutionize our approach to understanding our neural computation. Extracting biological meaning from such high-dimensional data, however, has demanded novel experimental, computational, and theoretical approaches. In this introductory chapter, we will highlight the need for labeled multi-neuronal and pan-neuronal imaging approaches in neuroscience and describe the advantages of *C. elegans* as a model system for such work. To provide context for the work described in the chapters that follow, we will also give a brief overview of our current understanding of olfactory coding across model organisms, multi-neuronal imaging techniques, and whole-brain experiments that have been carried out in the worm.

1.1 The need for functional imaging of neuronal ensembles

Multi-neuronal and pan-neuronal imaging methods allow us to capture the activity of large portions of the brain simultaneously in the same individual. The simultaneous imaging of neuronal ensembles allows circuits to be experimentally interrogated at much higher throughput. Instead of carrying out experiments one neuron at a time in response to a controlled stimulus, for instance, one could collect a complete dataset containing all relevant neurons, dramatically increasing the rate at which single-neuron properties may be elucidated.

The advantages of multi-neuronal imaging extend beyond more efficient data acquisition. Circuits at different layers of processing or in different brain regions do not work in isolation. Rather, their activity is likely contingent on the activity of many other neurons across the brain, communicating across the entire system via synaptic connections or non-synaptic modulation. A complete understanding of such neural computations cannot be constructed from compartmentalized experiments which record from single neurons in isolation: large numbers of the neurons engaged in the task of interest must be simultaneously recorded. This is true at all layers of the brain, from sensory systems to motor circuits. Even relatively stereotyped behaviors, such as sensory responses or reflexes, are subject to variability across individuals. The biological meaning of a variable neuronal response cannot be recapitulated in the aggregate without the context provided by the activity patterns of other neurons. Circuit-level and whole-brain imaging are not just methods of improving the efficiency of experiments; they may be necessary to understand the systems-level computations that the brain performs. Multi-neuronal approaches do not supersede single-neuron approaches: fine-

grained analyses of individual neurons also provide critical information. The two must be coupled to build a complete picture of neural computation.

In nearly every multi-neuronal study conducted thus far, across organisms and across many sensory modalities and behavioral paradigms, we see that neural representations are far more widely distributed across the brain than one would naively expect. Unlike artificial control systems, in which computation is centralized and a single electronic circuit tends to have a well defined role, brains appear to compute in a distributed way, with neurons and neural circuits employed flexibly, playing different roles in different environmental and behavioral contexts.

The capacity to comprehensively record complete circuits from brains of small model organisms such as worms, flies, and fish arose with recent developments in optical imaging. Genetically encoded fluorescent indicators of neural activity have long been employed to monitor neurons in model organisms. Fast, high-throughput microscopes are now capable of rapid volumetric imaging, capturing the time-varying fluorescence of many neurons in the field of view with high spatial and temporal resolution. Because so many neurons are in the field of view, however, segmenting individual neurons and extracting fluorescence signals from them is a difficult image analysis problem. Additionally, to directly compare multi-neuronal data across individuals, it is necessary to identify neurons (putting names onto each cell). Labeling neurons with fluorescent markers of multiple colors can allow neurons of interest to be identified, but building such strains requires extensive knowledge of gene expression patterns in the organism of interest.

Grappling with these datasets also requires innovation in theory and computation. What causal meaning can we extract from mapping activity correlations between sensory inputs and the activ-

ity of many neurons? Are there principles of integrated brain function that shape these correlation maps? What constraints does anatomical wiring impose on brainwide activity patterns, and can those constraints help us understand brainwide computation? All of these experimental and theoretical challenges are difficult to address in large brains. In a small brain with a small number of neurons, however, we have a chance.

1.2 *C. ELEGANS* AS A MODEL SYSTEM FOR NEUROSCIENCE

In the 1960s, Sydney Brenner and his group developed the nematode *C. elegans* as a model system for developmental biology and neuroscience[1]. Today, *C. elegans* is studied in thousands of laboratories around the world. There are many advantages of working in *C. elegans*, including a short life cycle, optical transparency, and well developed genetic toolkit. Most individuals are self-fertilizing hermaphrodites, meaning a single individual can give rise to homozygous progeny and making genetic manipulations even more convenient. *C. elegans* was the first animal to have its entire genome sequenced[2], and decades of genetics work has advanced our understanding of the genetic basis of development and behavior. Genetics also give experimenters an extensive library of promoters we can employ to drive expression of transgenes in specific cells of interest.

C. elegans has an invariant number of somatic cells (959 cells in each individual), and exhaustive work has uncovered the developmental lineage of every one of these cells, from fertilization to adulthood[3,4,5]. As a corollary, every cell in the animal, including every neuron, has a name and can be identified, allowing for direct comparisons across individuals. The ability to identify neurons and

compare activity across individuals has been vital in *C. elegans* neuroscience.

The neurobiology of the worm

The compact nervous system of the nematode *C. elegans* makes it an ideal platform for systems neuroscience. The hermaphrodite *C. elegans* nervous system consists of 302 neurons, divided into 118 distinct classes (often bilaterally symmetric pairs). The majority of these neurons, about 200, are concentrated in the head of the animal, forming the brain of the worm (**Figure 1.1**). The remaining neurons form a ventral nerve cord, a structure akin to the spinal cord of a vertebrate, and a small cluster of neurons can be found in the tail of the animal. Male *C. elegans* have about 100 additional neurons in the tail, forming a second brain which controls the animal during mating[6].

Neurons in *C. elegans* are relatively unbranched when compared to the elaborate arbors of neurons in larger animals. Most neurons have only one or two neurites (also called "processes") projecting from the cell body. In many systems, neurites of neurons are either axons or dendrites, depending on whether they send or receive signals, respectively. In *C. elegans*, however, most neurites both send and receive signals. In the brain of the animal, neurites from many of the neurons come together to form an annular bundle called the nerve ring.

The synaptic connectivity of *C. elegans* has been completely mapped, providing researchers with a wiring diagram, or connectome[8,9]. Neurons in *C. elegans* can form both chemical synapses, where neurotransmitters are exchanged, and gap junctions, where electrical signals can be exchanged. From the connectome, we observe that the neurons in *C. elegans* are loosely organized into three layers: sensory neurons, interneurons, and motor neurons. This makes the nervous system of *C. elegans*
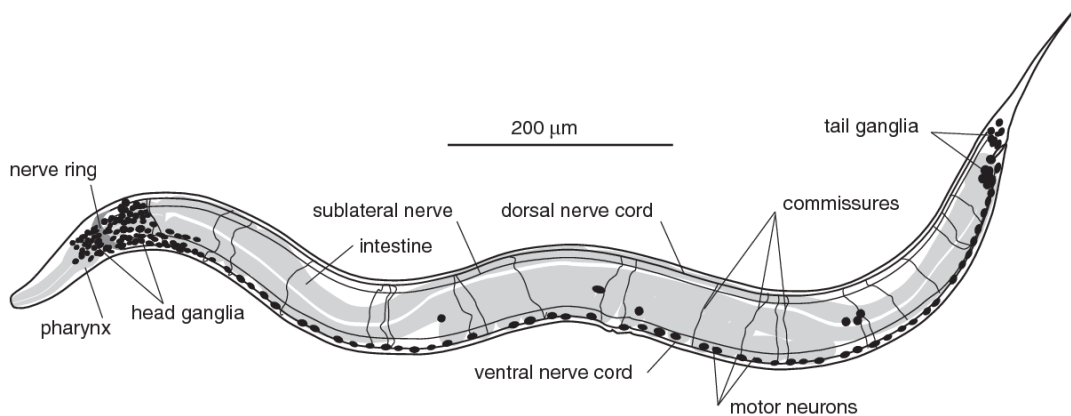
Figure 1.1: *C. elegans* and its nervous system. An adult hermaphrodite *C. elegans* is less than 1 mm in length, and contains 302 neurons (nuclei in black). The head ganglia contain about 200 neurons, which form the brain of the animal. The remaining neurons are distributed along the ventral nerve cord and in the tail. Figure from Hobson et al., 2017[7].

relatively shallow when compared to larger organisms. The connectome is a critical resource for *C. elegans* neuroscience, and will be discussed in greater depth in the next section. In addition to synaptic connections, neuronal activity *C. elegans* can be modulated by neuroendocrine signals[10,11].

The expression maps of the neurotransmitters acetylcholine, GABA, and glutamate have been determined in *C. elegans*[12,11,13,14,15]. Postsynaptic neurons receive neurotransmitter signals by either ionotropic or metabotropic receptors, but the map of neurotransmitter receptor expression is largely unknown, leaving our understanding of neurotransmitter communication networks incomplete.

When it comes to capturing neural dynamics, electrophysiological recordings are the most sensitive and have the highest temporal resolution. Such experiments in *C. elegans* have revealed the intrinsic electrical properties of its neurons. *C. elegans* expresses no voltage-gated sodium channels, and its neurons lack the sodium-triggered action potentials found in most insect and vertebrate

systems[16]. The first patch clamp recordings of *C. elegans* neurons found excitatory postsynaptic potentials (EPSPs)[17,18] and graded potentials[19], suggesting that the majority of electrical signals in the *C. elegans* nervous system were conducted passively. The lack of action potentials may be partially explained by the small size and high membrane resistance of *C. elegans*: many neurons experience a voltage change nearly simultaneously across the entire cell[19,20]. Neurons in do *C. elegans* express a large number of ion channel types, including voltage-gated potassium and calcium channels[21]. *C. elegans* muscle cells were discovered to use these channels to fire action potentials during muscle contraction[22]. Recent work has also found select neurons which fire action potentials mediated by these voltage-gated channels[23,24]. The extent of action potential signalling, and the role of action potentials in encoding information, are areas of active research.

However, it is impractical to record simultaneously from large numbers of neurons by patch clamping. The most common method of non-invasively extracting functional activity from individual neurons in *C. elegans* is to express genetically encoded calcium indicators such as GCaMP in neurons of interest. The time-varying fluorescence signal is a readout of the calcium dynamics of the labeled neuron. GCaMP recordings lack the temporal resolution and high sensitivity of electrophysiological recordings, but are much more flexible. Animals labeled with GCaMP can be imaged more easily while being presented with stimuli or while carrying out behaviors.

In *C. elegans*, some neurons have been shown to have different activity patterns in different parts of the same cell. Computationally relevant calcium dynamics may occur only in the nerve fibers and processes of many neurons[25,26,27]. In an animal that encodes the full range of its complex behaviors in only hundreds of neurons, the computational capacity of individual cells should not be

underestimated. The remarkable sophistication of individual cells in *C. elegans* is clearly demonstrated in its motor circuit. In larger animals, networks of spinal cord neurons give rise to rhythmic and organized movements[28,29,30,31]. In *C. elegans*, single motor neuron types encode the properties of networks of cells found in larger animals[32,33,34,35]. The multi-purpose functionality of many of the neurons in *C. elegans* is likely a result of evolutionary pressure, trying to maximize the computational ability of a such a small nervous system.

The *C. elegans* connectome

C. elegans was the first animal for which a connectome was generated, offering a near complete synapse-level map of its entire nervous system. Serial electron microscopy was employed to generate a high-resolution volumetric image of a complete animal, stitched together from multiple individuals (**Figure 1.2A**). The neurons were then manually traced through the volume and chemical and electrical connections were identified—a heroic feat with the methodology available in the 1980s[8].

An early achievement in worm systems neuroscience, an analysis of a complete circuit for a behavior, directly emerged from the connectome. Through systematic laser ablation and behavioral analysis, Chalfie et al. mapped the circuit for harsh touch sensitivity – a feedforward reflex that allows the worm to avoid anterior or posterior touches by rapid backward or forward movement – from sensory neurons to interneurons to motor neurons[32]. Since this early achievement, the connectome has provided an invaluable resource for mapping behavior to circuits[8,36,37]. With recent increases in the throughput of serial-section electron microscopy and the development of semi-automated or fully automated reconstruction algorithms, the connectomes of larger animals are now being re-

9
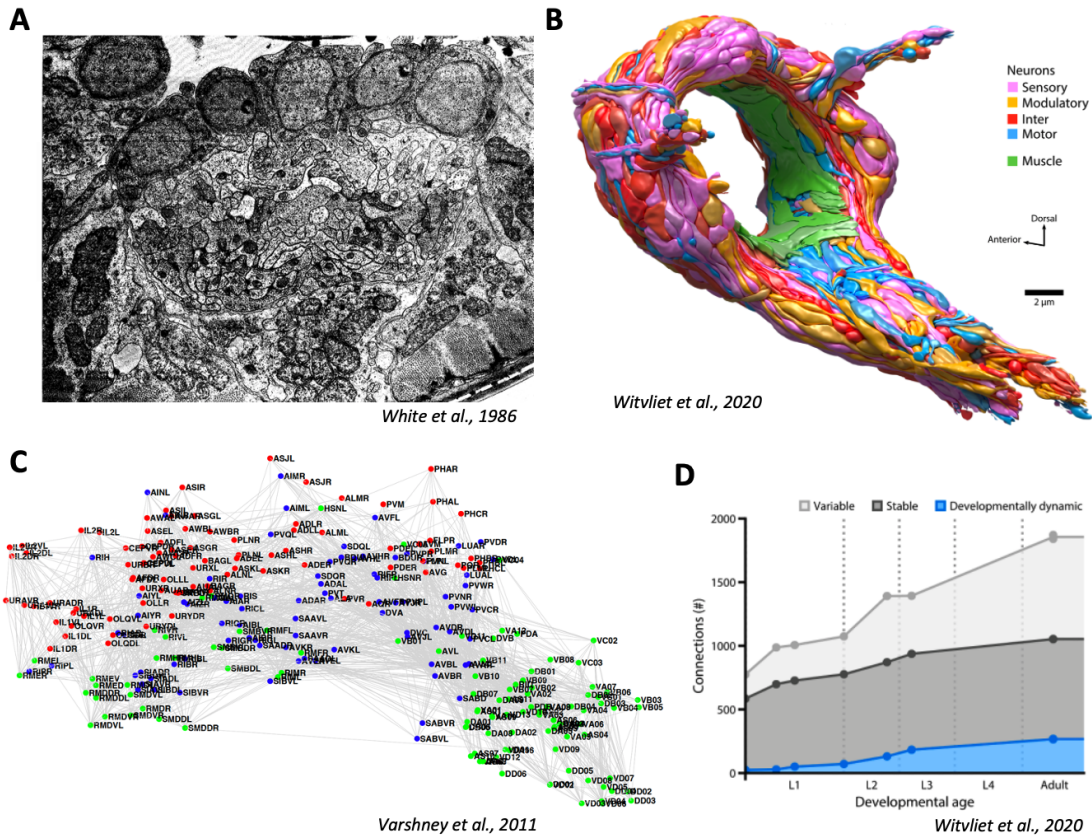
Figure 1.2: Mapping the connectome of *C. elegans*. (A) An EM image of a transverse section of the lateral ganglion in the head of the animal from the original John White dataset[8]. Several neural cell bodies are visible near the top of the image, with the neurites of dozens of neurons below them. **(B)** Neurons are reconstructed by tracing them through volumetric EM data. Here we see all of the neurons in the region near the nerve ring, in the brain of the animal[9]. **(C)** The topology of the *C. elegans* connectome, with sensory neurons labeled in red, interneurons labeled in blue, and motor neurons labeled in green, illustrating the shallow nature of the *C. elegans* nervous system. **(D)** The acquisition of multiple connectomes from different developmental time points have enabled comparative connectomics in *C. elegans*[9]. A small fraction of the connections are developmentally dynamic, reshaped as the animal grows. Other connections are stable—these exist across stages and across individuals. Up to 50% of the connections are variable across individuals. Panels adapted from White et al., 1986, Varshney et al., 2011, and Witvliet et al., 2020.

constructed including larval and adult *Drosophila* [38,39] and the larval zebrafish [40]. New technologies have enabled higher-throughput connectomics in *C. elegans* [9].

The first *C. elegans* connectome was reconstructed from partially overlapping segments from 4 individuals, which means that most synapses were reported with a sample size of $N = 2$ [8]. Despite the stereotypy of *C. elegans* development, however, the connectome is not identical across individuals. Recent work in which connectomes for 8 additional *C. elegans* were reconstructed has demonstrated significant variability in the adult connectome of *C. elegans*, with 43% of connections not conserved across adult individuals [9] (**Figure 1.2D**). Despite this variability, however, the strongly conserved connections do form a stable network across individuals.

The connectome is not sufficient to understand brain-wide dynamics. As studies of brain-wide activity repeatedly show, the same connectome can support functional correlations between neurons and across brain regions that change dramatically with environmental context and behavioral state. However, the connectome is a necessary framework for building mechanistic models of the brain.

The compact nature of the worm has made it the ideal subject for systems-level approaches in genetics and connectomics. More recently, *C. elegans* has also become a platform for systems neuroscience: the dissection of the motor circuit of *C. elegans* through multi-neuron and muscle activity recordings is a key example of the value of systems-level approaches [41,42,43,44,35]. In this work, we sought to extend these systems neuroscience approaches to study chemosensory systems, employing ensemble and pan-neuronal imaging to understand how the whole *C. elegans* brain processes olfactory information.

1.3 Olfactory coding in insects and vertebrates

Most organisms rely on the chemosensory modalities of sensation (smell and taste) to navigate the world, find food, and avoid dangerous environments. When compared to *C. elegans*, larger organisms such as insects and vertebrates employ a far larger number of neurons to detect olfactory stimuli.

The architecture of the olfactory system is broadly similar across insects and mammals[45,46,47]. The first, environment-facing layer consists of olfactory receptor neurons (ORNs), each of which nearly always expresses only one unique receptor type. This likely makes the strategies employed to encode odorant identity qualitatively distinct from those likely employed in *C. elegans*. These ORNs do not form synapse onto each other, though they can inhibit each other through ephaptic coupling[48]. ORNs converge on a number of glomeruli which can amplify, suppress, or modulate the tuning of ORN signals[49,50,51]. These glomeruli form brain regions dedicated to processing chemosensory information, the antennal lobe in insects and the olfactory bulb in vertebrates. From here, neurons project to higher brain regions responsible for learning and decision-making.

Both insect and mammalian olfactory systems employ olfactory receptors which are activated by multiple odorants, with each odorant binding to multiple receptors[52,53]. Each odorant can therefore activate a distinct subset of ORNs, allowing different odorants to be discriminated. This combinatorial receptor coding strategy, which allows a finite number of receptors to encode a very large number of odorants, has been characterized in larval *Drosophila*[54,55], adult *Drosophila*[56,57], and mouse[52,58]. In each case, it was found that some ORNs were more broadly tuned (responding to a

large number of stimuli), and some ORNs were more narrowly tuned (responding to a small number of stimuli). This may partly be due to evolutionary priorities—there is no *a priori* reason an animal's olfactory system must sample the entire chemical odorant space with equal precision. It is likely that there are certain compounds which are very important for an animal to identify, and other compounds in the environment whose detection is less critical to an animal's survival. A good example of this is a single ORN in *Drosophila* larva (one of only 21 in the entire animal) which is dedicated to detecting a pheromone released by a parasitoid wasp[59].

In addition to odor identity, an olfactory system must also be capable of identifying the concentration of an odorant. Typically, odorants at higher concentrations will activate ORNs more strongly, and activate a larger number of ORNs[60,61,55]. Odor intensity must then be decorrelated from odor identity in higher brain regions, though this process is still poorly understood.

Olfactory coding has been extensively studied in mouse[52,58], as well as both the larval and adult forms of the fruit fly *Drosophila melanogaster*[54,55,56,57]. Insects and mammals have olfactory systems which are broadly similar in structure[45,46,47]. The first, environment-facing layer consists of olfactory receptor neurons, each of which nearly always expresses only one unique receptor type. These receptors are activated by multiple odorants, and each odorant binds to multiple receptors[52,53]. Each odorant can therefore activate a distinct subset of sensory neurons, allowing different odorants to be discriminated. Odorants at higher concentrations will also activate a larger number of sensory neurons[60,61,55]. In insects and mammals, the olfactory receptor neurons converge onto a number of glomeruli, which synthesize the information from multiple sensory neurons before passing it on to higher processing regions[49,50,51]. There is spatial separation between the higher brain

13

regions responsible for innate behaviors (lateral horn in insects, amygdala in vertebrates) and learned behaviors (mushroom body in insects, piriform cortex in vertebrates)[51,62,63,64,65].

1.4   CHEMOSENSATION IN *C. ELEGANS*

C. elegans uses chemosensation primarily to find food and to avoid dangers, and is sensitive to both water-soluble and volatile chemical cues[66]. Behavioral experiments have identified sets of odorants which are behaviorally attractive or behaviorally repulsive to the animal[67], including compounds excreted by bacteria living in the environment *C. elegans* inhabits[68]. The majority of the work done to understand chemosensation in *C. elegans* has focused on single sensory neurons[69,70,71,72,73,74,75,76].

C. elegans is too small to establish a spatial gradient by differentially sensing odor concentration simultaneously in two places, as is common in larger animals. Instead, when navigating a chemosensory gradient, *C. elegans* employs a navigational strategy called the pirouette model[71], akin to the biased random walk in bacteria[77]. The animal crawls in a given direction for a certain period of time, then randomly changes its direction. The rate at which the animal changes direction is dependent on the concentration: if it is climbing up the gradient of an attractant on its forward bout (concentration experienced is increasing), it will be less likely to pirouette, and if it is going down the gradient (concentration experienced is decreasing), it is more likely to pirouette. This strategy requires *C. elegans* to be able to compare the current concentration of an odorant with recently experienced concentrations.

The primary chemosensory neurons project to the nose of the animal, where they are exposed to
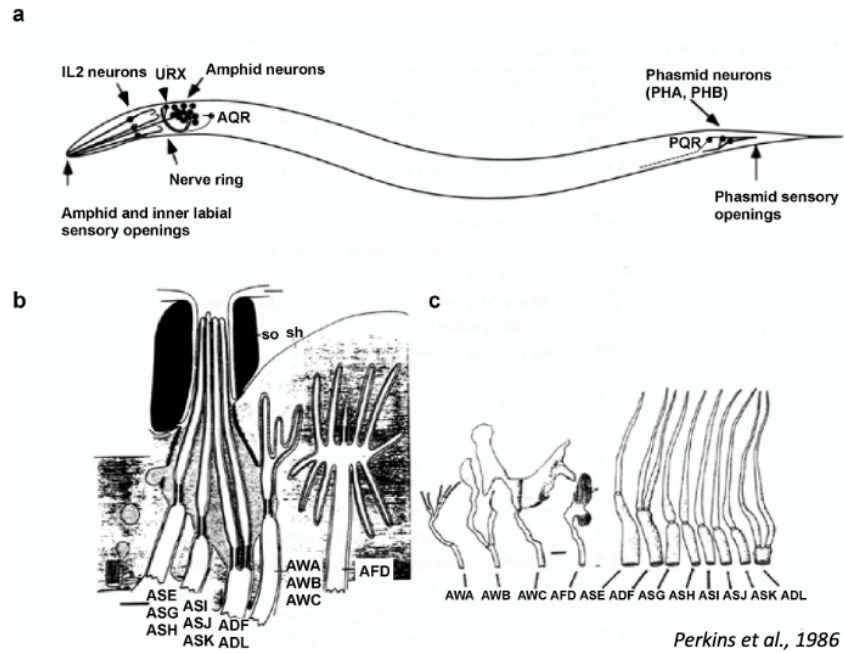
Figure 1.3: Anatomy of the chemosensory neurons in *C. elegans*. (A) The majority of the sensory neurons (amphids) are located in the head of the animal, and are exposed to the environment at the tip of the nose. Two pairs of phasmid sensory neurons are in the tail of the animal. **(B)** Structure of the nerve endings at the amphid pore in the nose. These include the 11 chemosensory neurons, and the thermosensory neuron AFD. **(C)** Detailed structure of the cilia of each of the amphid neurons. Adapted from Perkins et al., 1986, Bargmann et al., 2006[78,79].

the environment via two bilaterally symmetric amphid pores (**Figure 1.3**). Each of the neuron pairs forms a class which can be distinguished by morphology[8]. These 11 neuron classes are: AWA, AWB, AWC, ADF, ADL, ASH, ASE, ASG, ASI, ASJ, and ASK[78,79]. The nuclei of these chemosensory neurons are located in the lateral ganglia behind the nerve ring.

In large organisms such as insects and mammals, each olfactory receptor neuron expresses only one receptor type, with integration and further processing occurring in downstream levels of the network[45,46,47]. With fewer neurons, however, the *C. elegans* olfactory circuit must still solve similar computational problems. In *C. elegans*, every sensory neuron in expresses multiple receptor types—in total, about 1300 different receptors are thought to be expressed across the 11 chemosensory neurons[79,80]. Many of these neurons are also polymodal—in addition to olfactory stimuli, some detect gustatory stimuli such as pheromones, salts and amino acids, gases such as oxygen and CO_2, or temperature[79,81]. Additionally, we know from the connectome that many of these sensory neurons are wired to each other or receive feedback from motor neurons. This again contrasts with the olfactory systems of higher organisms, in which the first layer of olfactory neurons is largely feed-forward. Because of these characteristics, the *C. elegans* chemosensory system fits between the more distributed architecture we see in insect and vertebrate olfaction and the intracellular computations performed within a single chemotaxing bacteria.

Only one G-protein coupled receptor (GPCR) has been fully characterized in the context of *C. elegans* olfaction: ODR-10, which is expressed only in AWA and is known to respond to diacetyl[69,70]. Recent work has mapped the expression of a subset of the putative chemosensory GPCRs across the entire nervous system[80]. While this dataset is not exhaustive, we can use this information as an indi-

cation of the relative number of receptors expressed in each chemosensory neuron, and the degree of overlap in receptor expression between neurons.

ASH has been well characterized as a polymodal nociceptor, being activated by a wide range of repellent stimuli. It has been shown to be required for *C. elegans* to avoid high osmolarity, heavy metals, acidic environments, and aversive odorants, and is also necessary for touch avoidance at the nose of the animal[72,73,74,75,76].

The neurons AWA, AWB, and AWC are thought to be the primary neurons responsible for the detection of volatile compounds. Behavioral experiments have shown that ablation of either AWA or AWC destroys the ability of the animal to perform chemotaxis[67]. Experiments in which the receptor ODR-10 was expressed in AWB instead of AWA found that a previously attractive diacetyl stimulus became repulsive, suggesting that AWB activity may be linked to aversive stimuli[70]. AWB and AWC are also necessary for aversive olfactory learning[82].

ASE responds to salts and other water soluble compounds[71], and may also play a secondary role in chemotaxis to volatile odorants[72]. The ablation of the neurons ADF, ASG, ASI, ASJ, and ASK all degrade chemotaxis to a lesser extent, suggesting that the role they play in chemosensation is more minor[72]. That multiple neurons have to be ablated to completely abolish chemotaxis in response to certain stimuli also suggests the use of a distributed olfactory code in *C. elegans*. The neurons ASI, ASJ, ASK, and ADL are also known to detect ascaroside pheromones[83]. However, the specific role played by these neurons in olfaction remains largely unknown.

The chemosensory neurons in *C. elegans* synapse directly onto interneurons where behavioral decisions are being made. These downstream interneurons can be classified into two groups. The

first interneuron group consists of AVA, AVB, and AVD, command interneurons which control forward and backward locomotion[8,9]. This group receives the majority of its chemosensory inputs from ASH, and are responsible for reflexive avoidance behaviors in response to aversive stimuli. The AVA-AVB-AVD circuit can thus be thought of as analogous to brain regions such as the lateral horn and amygdala in higher organisms. The second interneuron group consists of AIA, AIB, AIY, AIZ, neurons have been shown to play roles in controlling goal-directed locomotion and learning behaviors[84,85,86,82]. This circuit receives inputs from all 11 chemosensory neuron classes, and can be considered as analogous to the brain regions in higher organisms which influence learning and decision-making, such as the mushroom body and piriform cortex.

Several studies have recorded the calcium activity of single chemosensory neurons in response to odorants, including AWB, AWC, and ASH responses to isoamyl alcohol[87], AWA responses to a panel of odorants[88], AWC responses to diacetyl[88], AWA, AWB, AWC, and ASE responses to benzaldehyde[89]. The latter results suggested that in response to benzaldehyde, some neurons (AWA and AWC) acted primary odorant-sensing neurons, recruiting the neurons AWB and ASE with classical neurotransmitter and neuropeptide signalling, respectively. A library of 15 single-neuron labeled lines was used to assess responses to isoamyl alcohol, diacetyl, and NaCl[90]. It has also been demonstrated that neuron AWA fires action potentials, and that these potentials may encode stimulus-specific features[24].

The neurons ASE and AWC are known to respond asymmetrically to certain chemosensory stimuli. The two ASE neurons are structurally symmetric but are functionally distinct, with ASEL and ASER expressing different sets of receptors[91]. ASER detects chloride and potassium ions, while

ASEL detects sodium ions[71]. The calcium activity of ASEL and ASER has been shown to be asymmetric in response to NaCl presentation[92,93]. AWC is stochastically asymmetric: in each worm, one neuron (either AWCL or AWCR) will adopt the identity AWC^{ON}, while its lateral pair will adopt the identity AWC^{OFF} [94]. The odorants 2-butanone and 2,3-pentanedione have been shown to elicit asymmetric responses in AWC, while other odorants such as isoamyl alcohol elicit symmetric responses.

Sensory adaptation has been observed in the neurons AWC, ASH, and ASE. When presented with a prolonged chemical stimulus on a timescale of minutes or hours, the neuron's activity will gradually be reduced[95,96,97,98,88]. In the neuron AWC, it has been shown that adaptation to one odorant, benzaldehyde, reduces the sensitivity of AWC to isoamyl alcohol, but does not reduce AWC's sensitivity to 4 other odorants: butanone, diacetyl, pyrazine, and 2,4,5-trimethylthiazole[95].

1.5 THEORETICAL MODELS OF OLFACTORY CODING

How does an ensemble of sensory neurons encode chemosensory information? An effective neural code must be reliable: a given odorant should elicit a similar sensory response every time it is encountered. The code must also be effectively decoded by downstream neurons. Evolution has tuned chemosensory systems to be maximally advantageous to the animal, which is generally interpreted to mean that chemosensory systems are robust, rapid, and efficient. The efficient coding hypothesis proposes that the goal of any sensory system is to encode the maximum amount of information in as few signals as possible[99].

Olfaction is fundamentally an object identification problem. Most of the environments organisms exist in contain thousands of different chemicals[100]. The animal must be able to detect the chemicals relevant to its life—chemicals that may indicate the presence of food, or a mate, or of a danger. To respond with the appropriate behavior, the animal must be able to distinguish one chemical from another.

In cases where detection of a specific molecule is critical to the animal, a single, narrowly tuned receptor could be dedicated to detecting that specific molecule. The sensory neuron(s) expressing this receptor then project directly to higher brain regions. This strategy is known as a labeled line[101,102,103]. In many cases, however, the animal must be capable of discriminating tens of thousands of chemicals relevant to its life—a set of chemicals called an odor space. This odor space can be different for different organisms, since an animal may only detect the odorants relevant to its ecological niche. For example, a pollinating insect would have an evolutionary advantage from being able to detect and discriminate the smells of many flowers, but a predator will not. An effective olfactory coding strategy is required to distinguish large numbers of odorants with a set of receptors at least an order of magnitude smaller in number.

One sensory encoding strategy is sparse coding, in which a small number of unique neurons is activated in response to a particular stimulus[104]. Experimental evidence for sparse coding has been found in insect and mammalian visual and auditory systems[105,106]. The advantages of a sparse code lie primarily in the simplicity of decoding the signal. However, requiring specific neurons narrowly tuned to specific stimuli limits the number of stimuli the sensory system can encode for.

In olfaction, most animals studied thus far employ a combinatorial coding strategy[54,55,56,57,52,58].

The advantage of a combinatorial code can be seen numerically. Given N sensory channels, and making the simplifying assumption that each of these channels is binary in its activity, the number of possible combinations is:

$$c = \sum_{n=1}^{N} 2^n,$$ (1.1)

a number which scales rapidly with N. Thus, combinatorial coding allows a finite number of receptors to encode a very large number of stimuli. In the context of olfaction, a combinatorial code requires several conditions. First, a single neuron (or single receptor) recognizes multiple odorants. Second, a single odorant is generally recognized by multiple neurons. Third, the set of neurons activated by a given odorant is unique [107]. A combination of these conditions results in a stimulus-specific, spatially distributed activity pattern, or spatial map, either directly at the sensory neuron level, or, as is more common in larger organisms, at the level of ganglia.

In addition to spatial coding, temporal signals may also play an important role in encoding olfactory information [102,108,109]. This is particularly true in cases where the animal must actively probe the chemical environment in a discrete manner (for example, a sniff), where the temporal order in which sensory neurons are activated can encode identity information, a strategy known as primacy coding [110,111].

Even within the same animal, different coding strategies may be used for different odorants. In many insect systems, it has been observed that while a combinatorial coding strategy encodes many odorants, particular stimuli are detected via labeled lines [103]. Generally these labeled lines detect

stimuli critical to the animal's ethology.

To efficiently encode olfactory stimuli, the number and tuning of receptors must also be optimized by evolution. The optimal number of receptors is expected to grow with the population size of olfactory sensory neurons, with the optimal distribution of receptors dependent on the environment[112].

How should the ensemble of chemosensory receptors be tuned? Experimental evidence thus far suggests a wide range of tuning properties within a given animal, with some broadly tuned neurons and some narrowly tuned neurons. However, it is difficult to rigorously define tuning without an exhaustive catalog of the animal's ethologically relevant odorants. Compressing such a high-dimensional odor space into a limited number of receptors is a key challenge. Using an information theoretic model, it has been shown that maximum theoretical coding efficiency (the most odors distinguishable with the fewest receptors) is reached when each receptor is broadly tuned, each responding to approximately 50% of the odor space[113]. However, this only holds if the responses of different receptors are uncorrelated with each other. It has been proposed that a receptor ensemble can achieve this by having disordered molecular affinities[114].

Theoretical work which models each sensory neuron in the ensemble as a nonlinear detector constrained by a limited response range has suggested that there may be an optimal sensitivity matrix, dependent both on the number of neurons and on the statistics of the odor environment[115]. In certain environmental limits, this will result in sparse tuning, while in other limits, it will result in broad tuning.

Measuring activity at the whole-brain level requires microscopic probes that can globally detect changes in electric fields, intracellular ion concentration, or neurotransmitter release. One of the most successful approaches has been to use proteins with activity-dependent fluorescence that are produced by the neurons being measured in transgenic animals. Genetically-encoded protein indicators have been developed for many aspects of neuronal activity, but nearly all whole-brain activity recordings to date have used GFP-based sensors of calcium dynamics. The quality and ease-of-use of genetically encoded calcium indicators (GECIs) such as GCaMP allow for stable, long-term imaging of large populations of cells throughout the brains of genetically accessible animals. One advantage of using an intracellular probe of neuronal activity like a calcium indicator – as opposed to a membrane-bound sensor of electrical activity, for example – is that it is easier to optically discriminate activity among large numbers of neighboring cells with touching membranes. To record the activity of multiple neurons using fluorescence measurements, we need microscopes that are capable of resolving cells throughout the brain volume while comprehensively sampling at the timescale of interesting behavioral and neuronal activity ($\sim$ 10 ms to $\sim$ 10 s). The most common approach to resolving large numbers of cells throughout a brain volume using fluorescence is to confine the excitation light to a subset of the imaging volume, and selectively capture in-focus light from that subset.

Conventional two-photon and confocal approaches rely on point scanning to image a brain volume. While point scanning affords advantages in optical resolution, it is necessarily slower and

precludes imaging many cells throughout large brain volumes on subsecond timescales. Confocal microscopy can be accelerated by simultaneously scanning many points at once in a focal plane using a 2D array of pinholes (a spinning disk confocal microscope). For example, spinning disk confocal microscopy has been used to record whole-brain activity from behaving worms at up to 10 volumes/second[116,117].

As our ability to perform whole-brain imaging increases, so does the problem of dealing with the enormous amount of data that it rapidly generates. Microscopes measuring whole-brain neuronal activity easily generate vast amounts of raw image data (up to 1 GB/s or more). These data must be reduced into compact time traces corresponding to the activity of discrete neurons or brain regions. Segregating the activity of individual neurons is particularly challenging when neurons and nerve fibers are densely packed in a brain volume or when neurons move relative to one another because of the animal's own movements. Continuously tracking neurons within the rapidly deforming brain of a moving *C. elegans* or *Drosophila* larva is as significant a challenge as acquiring the volumetric images in the first place. To complement the optical hardware that performs whole-brain imaging, image-processing algorithms that are both fast and accurate are being developed to meet the challenge of comprehensive neuronal segmentation[116,118,119].

1.7 Whole-brain imaging studies

Large-scale recording of large portions of the brain with neuron-level resolution has now become possible in several small model organisms: the nematode *Caenorhabditis elegans*, the larval zebrafish

Danio rerio[120,121,122], and the fruit fly *Drosophila melanogaster*, in both larval[123,124] and adult forms[125,126,127,128]. Other small animals like the hydra are being developed as models for whole-brain or whole-circuit approaches to behavior[129]. There has also been much recent work in rodent systems (mouse and rat) where large numbers of neurons are recorded in rich behavioral contexts[130,131]. These systems offer circuit-level dissections of behavior, but they do not yet allow comprehensive whole-brain recording.

The compact nervous system of *C. elegans* is ideal for whole-brain experiments. Small size allows nearly the entire worm brain to be rapidly imaged with single cell resolution using light microscopy – either the anterior brain or the posterior male "brain"[132]. Even with no external stimulus, a large proportion of brain neurons engage in coordinated activity in immobilized worms. When this whole-brain activity is projected onto a low-dimensional representation, brain dynamics resemble a limit cycle[133]. Portions of the limit cycle correspond to the activity of pre-motor interneurons known to be associated with locomotion direction, allowing epochs of fictive forward and backward movement to be inferred in stationary animals. The stereotyped brainwide activity patterns for forward/backward behavioral states have been interpreted to represent global commands that account for the majority of the variance in neural dynamics.

Forward and backward locomotion are slowly changing behavioral states, but muscle activity within each state occurs on faster time scales to drive rapid exploratory head bending and rhythmic body undulation[35]. Although the neurons that drive these movements operate at much faster time scales, they are directly modulated by other neurons with slowly changing activity that are correlated with forward/backward behavioral state changes. The activity and cross-modulation of neurons

across a hierarchy of time scales occurs in both moving and immobilized worms. Nested activity dynamics across time scales appears to be an organizing principle of the brain circuit, both in unrestrained and fictive behavior[134].

Capturing whole-brain dynamics in immobilized animals is sufficient for dissecting sensory inputs, but to understand how brains respond to motor feedback and produce complex behaviors, it is ideal to study brain and behavioral dynamics at the same time in the same animal.

Improvements in fast volumetric imaging and single neuron tracking now enable whole-brain recording in freely moving worms[116,117]. Consistent with recordings in immobilized worms, large numbers of neurons in the brain are correlated with forward and backward movement. But in freely moving worms, these correlations also exhibit substantial diversity with respect to additional quantifiable parameters of worm movement including velocity and curvature. Reliable decoding of all parameters of worm behavior from brain-wide activity requires large numbers of neurons, hinting at a more subtle and distributed neural code for the full dynamics of worm behavior[135]. Moreover, the correlation structure between certain pairs of neurons changed dramatically when freely-moving worms were immobilized. Thus, the neural dynamics of fictive behaviors in immobilized worms is measurably different from the corresponding neural dynamics in unrestrained worms, an important caveat when trying to understand a natural behavior by studying immobilized animals. Studies of the male tail during mating have uncovered many neurons which contribute to multiple sub-behaviors in different ways, leading to different correlation patterns throughout the circuit in different contexts. Functional correlations between neurons are not fixed, but explicitly depend on context and behavioral state[136].

Multi-neuronal and pan-neuronal imaging methods have become important tools in the arsenal of systems neuroscience. Common themes are emerging in studies of different animals across different behaviors, pointing to shared principles in brain-wide representations of sensory information and behavior[137] and reinforcing the need for large-scale recordings to fully understand the workings of the brain.

1.8 Dissertation overview

All animal brains use their sensory neurons to perceive the world around them. Sensory inputs are processed by the interneurons of the brain, ultimately generating behavioral decisions. In this dissertation, we set out to understand how the entire *C. elegans* brain represents, integrates, and processes olfactory information. Along the way, we developed genetic, experimental, and computational tools which enable high-throughput acquisition of stimulus-evoked activity in labeled neurons.

In Chapter 2, we developed an experimental setup which used custom hardware to present olfactory stimuli and record neural activity in *C. elegans*. We employed high-resolution fluorescence microscopy to image the brains of individual worms and a microfluidics system capable of precisely controlling the chemical environment experienced by these animals. In Chapter 3, we constructed a data analysis pipeline to extract calcium activity from this imaging data. We developed software capable of reliably and efficiently tracking multiple, densely-packed neurons in semi-immobilized animals. These methods have allowed us to robustly probe the *C. elegans* olfactory system in a high-throughput manner.

In Chapter 4, we employ these tools to understand how the chemosensory neuron ensemble in *C. elegans* represents odor information. We developed a transgenic line in which all chemosensory neurons in the worm can be simultaneously interrogated, and presented these animals with a broad panel of olfactory stimuli. We found that collectively, the sensory neurons act as a spatial map to encode odor identity, with each odorant activating a unique combination of sensory neurons. These neural representations of olfactory stimulus are robust: the neural responses to a single odor presentation are sufficient to identify the odorant. Ablation and mutant experiments indicate that these neural representations are not dependent on synaptic connections. We were also able to describe the roles of individual sensory neurons in olfactory coding, finding diverse dose response and tuning properties across neuron classes. Taken together, these results form a comprehensive picture of how *C. elegans* uses a small number of sensory neurons to encode both odorant identity and intensity.

Next, we wanted to uncover how odor information is transformed and represented among the downstream interneurons. To do this, we first needed a method of labeling every neuron in the *C. elegans* brain. In Chapter 5, we developed a transgenic worm in which all neurons in the animal can be identified with a deterministic multicolor barcode. We conducted labeled whole-brain imaging with these animals. We found that in the absence of stimulus, global brain activity was high-dimensional and varied across individuals. Stimulus presentation would push the brain into a specific state. Each chemosensory stimulus elicited a unique average brainwide activity pattern, spanning sensory neurons, interneurons, and motor neurons. These results demonstrate that stimulus-evoked activity is more widespread than previously thought, and suggest that odor identity information is encoded in different patterns of activity in the interneuron ensemble.

Pan-neuronal labeling has also allowed us to make direct comparisons between functional activity correlations and the anatomical wiring diagram of the worm. We found that functional activity did not significantly correlate with synapse counts, but we wanted to explore whether the brain-wide activity we observed reflected the structure of the connectome in other ways. In Chapter 6, we used our labeled whole-brain data to train a neural network model of the *C. elegans* nervous system constrained to connectome topology. We found that the anatomically-derived network was able to predict neural activity more accurately than naive networks of a similar size. This result demonstrates that on a basic level, the connectome puts measurable constraints on the brainwide activity patterns we observe in the animal.

Finally, we conclude with a discussion of all of these results, their implications, and promising directions for ongoing and future work. Looking forward, labeled pan-neuronal imaging will allow for careful examination of the correlations between interneuron and motor neuron activity, furthering our understanding of information flow from sensory input to motor output. The methods we have developed will enable new, whole-brain approaches to understanding sensorimotor transformations in *C. elegans*. ●

2

Methods for imaging and stimulus presentation

Vivek Venkatachalam, Guangwei Si, and Jessleen Kanwal contributed to the work described in this chapter.

To RECORD FROM THE BRAINS of intact animals while presenting them with chemical stimuli, we employed high-resolution fluorescence microscopy of individual animals placed in a microfluidics device capable of precisely controlling the chemosensory environment. In this chapter, we will discuss the development and operation of the experimental setup used to precisely and robustly probe chemosensory responses in *C. elegans*.

2.1 INTRODUCTION

The experimental *C. elegans* strains we used were genetically engineered so that the fluorescent calcium indicator GCaMP6s was expressed in either some or all of the neurons. GCaMP fluorescence

changes as a function of Ca^{2+} concentration, and is thus a useful readout of neuron activity. The experiments discussed in this thesis were conducted on animals in which a nuclear localization signal (NLS) was employed to restrict GCaMP to the nucleus of each labeled neuron. This was done to improve the separability of signals from nearby neurons. The animals discussed here also expressed activity-independent fluorescent markers, such as BFP, cyOFP, RFP, mNeptune, and mCherry. These markers can play two important roles. First, by labeling a subset of neurons in a specific color, we can identify the neurons more easily. Second, an activity-independent marker expressed in a neuron which also is expressing GCaMP can help to assess and correct motion artifacts or crosstalk between channels. To extract useful information from these animals, we used a spinning-disk confocal microscope to record from their brains with sufficient temporal and spatial resolution to distinguish individual neurons and quantify their activity.

In various studies, *C. elegans* have been imaged while freely crawling on agar plates, or while semi-immobilized in a variety of microfluidics devices[138,133,117,116]. Between these options, semi-immobilization in a microfluidics channel is less naturalistic. Indeed, it has been shown that locomotory circuit activity and feedback differ between immobilized and freely-moving animals[135].

Despite this, the microfluidics approach offers us several advantages when it comes to understanding chemosensory responses—responses which are unlikely to be dramatically changed by motor feedback. First, immobilizing the animal allows us to precisely control both the presentation and removal of chemosensory solutions, and to switch stimuli on the timescale of seconds. In contrast, in a freely moving experimental setup, chemosensory stimuli cannot be quickly removed from the animal's environment. Second, the higher spatial resolution we can obtain when imag-

ing an immobilized animal makes it easier to separate the signals of neighbors—a critical advantage when imaging a large number of neurons in a small region. Third, in an immobilized animal, we can acquire higher-resolution reference (landmark) volume separate from a lower-resolution activity movie. This high-resolution image can include more color channels, and be used to identify the neurons more accurately. Here, we will discuss in detail the experimental methods we developed to present *C. elegans* with chemosensory stimuli and record their neuronal responses.

2.2 Confocal microscopy

We employed a single-photon, spinning-disk confocal microscope to capture fluorescent images from intact *C. elegans*. The microscope was inverted to allow for easy access to the microfluidics device mounted on the stage. We employed a 488 nm laser to excite GCaMP *in vivo*, and used additional lasers to excite fluorescent proteins used as landmarks. To minimize crosstalk between channels, lasers were fired sequentially during multicolor recordings. We captured images with a 60x water-immersion objective with an NA of 1.2. Volumes were acquired using unidirectional scans of a piezo objective scanner. All fluorescence microscopy is a trade-off between spatial resolution, temporal resolution, laser power, and signal strength. We optimized two sets of imaging conditions, one set for activity imaging and another set for landmark imaging. Both sets of imaging conditions capture the region containing the majority of the neurons in the head of *C. elegans*, a volume of 112 μm by 56 μm by 30 μm.

In any given experiment, acquisition of a landmark volume precedes acquisition of an activity

movie. This volume, which contains sufficient color channels to capture all of the landmarking fluorophores, allows us to identify neurons of interest. The spatial resolution of these volumes is 0.5 μm x 0.5 μm x 1.5 μm/voxel, with the z-resolution of 1.5 μm set by the point spread function. This landmark can take up to 1 minute to acquire if the animal of interest contains multicolor labeling, thus requiring the animal to be immobilized (see Section 2.4.1 for details).

The activity movies were acquired at a high speed, with lower spatial resolution (1 μm x 1 μm x 1.5 μm/voxel). At this resolution, we could acquire volumes at a rate of 2.5 Hz in standard acquisition mode, or up to 4 Hz when the piezo was driven at higher speed using custom control software. Refer to Appendix A.1 for a listing of parts and filters used in this imaging setup.

2.3 MICROFLUIDICS DESIGN

To present animals with chemosensory stimuli in a controlled manner, we turned to microfluidics. We employed a modified version of a microfluidic system developed in our group which was capable of delivering multiple odors to *Drosophila* larvae[55]. The microfluidics chip is designed with an arbor containing delivery points for multiple stimuli, together with a buffer delivery point and two control switches, one for buffer and one for odor (**Figure 2.1A**).

At any given time, three flows are active: one of the control switches, the buffer blank, and one odor stimulus. The chip is designed to maintain laminar flow of each fluid, and the flow is split between a waste channel and an odor channel which flows past the animal's nose. The chip described here is designed to switch rapidly from one stimulus to the buffer. While the buffer is presented to

the animal, we can switch from one stimulus to another—a slower process due to the size of the arbor. Because the stimulus switch occurs away from the animal's nose, the animal experiences rapid switching from stimulus to buffer and then back to stimulus. Other chip designs employ a second arbor in place of the buffer, allowing for rapid switching between stimuli. After the flows pass the animal, they exit the chip via a waste port at atmospheric pressure. Waste is then removed with a vacuum.

We grafted the odorant delivery arbor to a *C. elegans* loading chamber similar to those designed by Chronis, et al.[138]. We designed a loading chamber suitable for adult *C. elegans*, a narrow channel 62 μm wide and 30 μm high, with a gently tapered end (**Figure 2.1B**). The tapered end serves as a guide to help hold the animal's nose in place without distorting the animal. However, since the body of a *C. elegans* worm is easily deformable, the tapering itself is insufficient to immobilize the animal. For standard experiments, we instead rely on a combination of gentle squeezing from the restricted height of the microfluidics channel and manual adjustment of hydraulic pressure to keep the worm in one place. We will discuss loading and immobilization in more detail in Section 2.4.1. In addition to the standard loading channel design, we designed a number of variants capable of recording juvenile, adult, and male *C. elegans*. Refer to Appendix A.3 for the dimensions of all of variants.

To increase experimental throughput and minimize the introduction of air bubbles, we designed extra height into the waste channel downstream of the animal. The larger cross-sectional area allows the animal to be flushed directly to the waste port at the end of an experiment, leaving the chip immediately ready for loading the next animal. If the need arises, however, it is possible to recover the
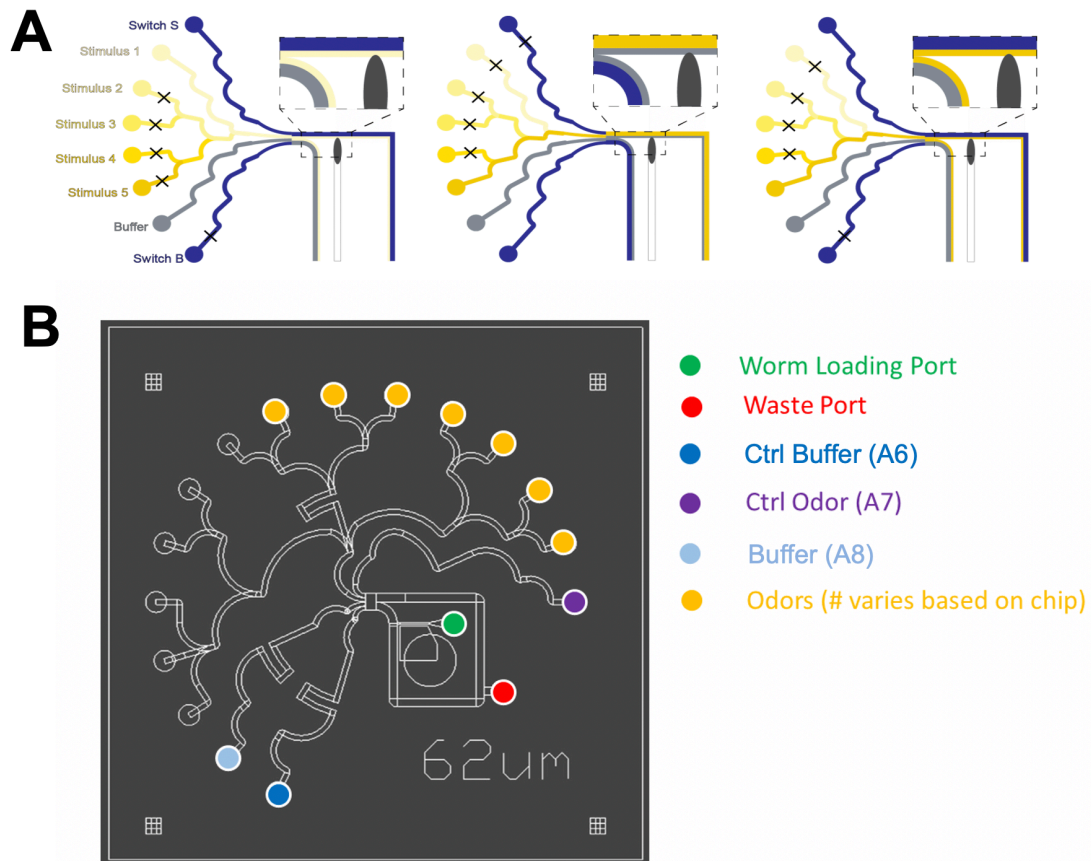
Figure 2.1: Microfluidics design for odor delivery. (A) A schematic of stimulus switching in the microfluidics chip. In the leftmost panel, Stimulus 1 is being presented to the animal. In the middle panel, the animal is experiencing a buffer blank. The buffer control switch has been opened and the odor control switch has been closed, pushing the odor stream away from the animal. By closing the Stimulus 1 switch and opening the Stimulus 5 switch, we allow the odor stream to change. In the rightmost panel, we switch back to odor delivery. The animal is now being presented with Stimulus 5. **(B)** The design of microfluidics chip used for adult *C. elegans*. The odor control switch port (purple) and the buffer control switch (blue) control the delivery of the stimuli (yellow) and buffer (light blue), respectively. The animal is loaded through the worm loading port (green), and fluids leave the chip via the waste port (red). The channels flowing to the waste port are double the height of the rest of the pattern.
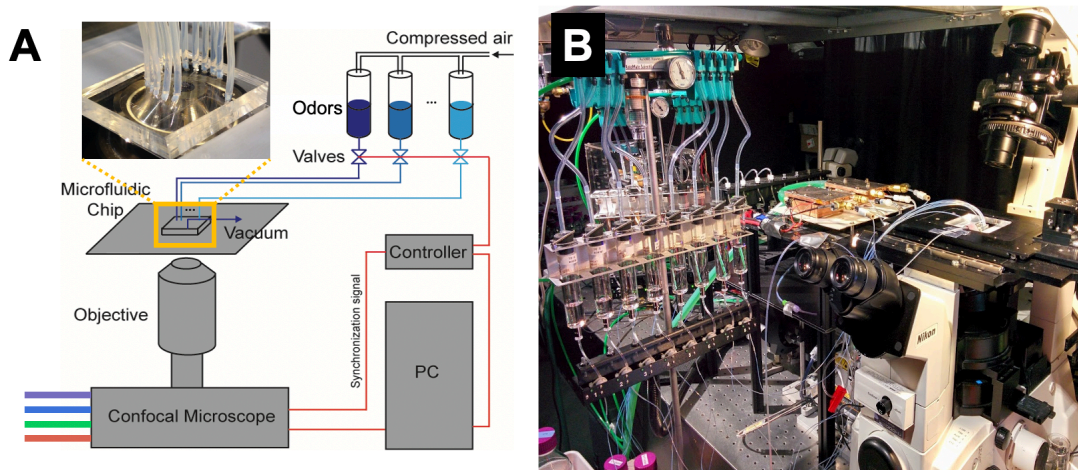
Figure 2.2: Odor manifold and confocal microscopy setup. (A) A schematic of the experimental setup. The odorant solutions are contained on the odor manifold in glass syringes under air pressure, and flow from these syringes is controlled by pinch valves. Tubing from the syringes leads to the ports on the microfluidics chip (image inset). The experimental sample is imaged by an inverted spinning-disk confocal microscope. A timing loop connects the microscope camera with the odor manifold controller. **(B)** An image of the experimental setup, with the odor manifold to the left and the confocal microscope to the right. Note the tubing from the syringes on the manifold to the microfluidics chip on the microscope stage.

experiment animal by sucking it back up the syringe and depositing it on an agar plate.

The microfluidics chips were manufactured in a cleanroom using standard soft lithography techniques. Refer to Appendix A.2 for microfluidics chip fabrication details.

To deliver stimulus and buffer solutions to the chips, we used an odor manifold consisting of a pinch-valve perfusion system which controlled the delivery of solutions held in glass syringes (**Figure 2.2**). Each syringe was connected to flexible silicone tubing fed through a pinch valve, allowing the valve to control flow from the syringe. The silicone tubing was then connected to more durable plastic tubing, which ran from the manifold to the ports on the microfluidics chip. To prevent air from getting into the chip, each syringe and tube were bled before being connected to the

chip. The syringes were pressurized to 4 psi, giving us a total flow rate through the chip of approximately 0.2 mL/min. The switch from odorant presentation to buffer presentation occurred within 20 ms. A microcontroller and custom MATLAB scripts were used to control the pinch valves and execute odorant patterns synchronously with the onset of camera recording[55].

Refer to Appendix A.4 for an odor manifold parts list.

2.4 EXPERIMENTAL DESIGN

In this section, we will discuss some of the scientific considerations that guided the design of the experiments we carried out for the projects in Chapters 4 and 5. Refer to Appendix A.5 for a detailed operations manual for our experimental setup.

2.4.1 WORM SELECTION, LOADING, AND IMMOBILIZATION

We conducted experiments on young adult animals which had not been starved for at least 2 generations. The animals were picked from their agar plate and washed in M9 buffer. The animal was loaded into this chamber nose-first by a syringe. As the animal enters the chip, we can manually adjust its orientation before it is pushed into its final position in the loading channel. The fluid pressure on the nose and tail of the worm is then manually equalized in order to minimize any motion of the animal driven by pressure gradients.

A combination of this pressure equalization and contact with the walls of the loading channel hold the worm loosely in place during imaging experiments. However, a worm is capable of wriggling about in this channel. During activity imaging exposure times are short and the tracking al-

gorithms (Chapter 3) are capable of following neurons through these movements. However, when long exposures are required, such as for multicolor landmark imaging, the animals must be completely immobilized. To do this, 2 mM tetramisole hydrochloride was added to the loading buffer to paralyze the animals.

We also prototyped a method for immobilizing animals using temperature, employing a dual-layer microfluidics device which circulated cold water above the loading channel. This approach did immobilize worms in the chip, but was more complex than drug immobilization. Cold temperature immobilization remains a viable option when paralytics like tetramisole cannot be employed.

2.4.2 Stimulus delivery patterns and adaptation

The odor manifold setup is capable of delivering any sequence of stimulus solutions and buffer blanks. The two key parameters in designing such an odor sequence are the length of the odor pulse and the length of time between odor pulses.

In *C. elegans*, it has been shown that the length of an odor delivery will change the calcium response of sensory neurons[88]. On short timescales (order seconds), the peak fluorescence tends to increase with odor delivery length. The dynamics of the neurons can also change as a function of odor delivery length. In the extreme limiting case, where an odor is presented and not removed, sensory neurons will eventually return to their baseline activity level. For our experiments across a broad odor panel, we were free to choose any odor delivery length, provided we used a consistent length of time across all conditions. After several trial experiments, we settled upon an odor delivery time of 10 seconds, a short odor pulse which avoids the effects of adaptation. We found that this

time period was sufficient to separate most ON responses (neuronal responses with the onset of stimulus delivery) from OFF responses (neuronal responses upon the removal of stimulus delivery). Note, however, that a 10 s pulse, when coupled with the timescale of GCaMP6s, is insufficient to distinguish the two bouts of activity of neurons with both an ON and an OFF response to a given odorant (i.e.: AWC and ASE to certain stimuli, such as 2-butanone).

The length of time between odor pulses would also have to be sufficiently long to avoid one odor delivery from potentially changing the response to the following odor delivery. We also want this time to be as short as possible, allowing for more odor presentations while minimizing photobleaching due to long imaging times. We conducted trial experiments with a range of odor pulse spacings, and found that odor pulses separated by at least 40 seconds of buffer blank would give independent sensory responses: the presentation of the first stimulus did not affect the dynamics of the second. We validated this spacing by comparing average responses to the first odor pulse to the average responses to sequential pulses, up to 8 odor pulses. We found no average trend in peak responses across odor pulses when spaced thus.

With these two timescales in hand, we constructed odor sequences as required by different experiments. To avoid confounding odor-evoked responses with blue light responses of some neurons in *C. elegans*, we began each experiment with 1 minute in which the animal is imaged but only a buffer blank is presented. Then, we began the odor delivery sequence, with 10 s odor pulses separated by 40 s buffer blanks. The odor pulses could either be distinct odorants or repeat deliveries of the same odorant. Under our imaging conditions, we found that an animal in the microfluidics chip could be imaged for up to 20 minutes before photobleaching is too severe for calcium imaging. However, it is

often necessary to adjust the position of an animal in the chip before then. We found experimental recordings of about 4-5 minutes to be ideal, with 3-6 odor deliveries per recording. The same animal could then be recorded from multiple times if necessary.

2.4.3 Odor choice and buffer preparation

As our odor perfusion system delivers stimuli in liquid form, all of the odorants employed in chemosensation experiments were water-soluble. The maximum concentrations we used (limited by the contamination concerns discussed below) were 10^{-4} dilutions for soluble odorants, and 200 mM for salts. We have also presented animals with pheromones and bacterial supernatant, and observed sensory neuron responses to both. For volatile odorants, any stored dilutions older than 3 weeks were remixed prior to conducting any experiments.

Initially, we diluted odorants in deionized water, as was done in *Drosophila* larva experiments[55]. However, to address concerns of osmolarity balance within and without the animal, we switched to diluting odorants in CTX buffer (5 mM KH_2PO_4/K_2HPO_4 pH 6, 1 mM $CaCl_2$ and 1 mM $MgSO_4$), with an osmolarity of 350 mOsm/L across stimulus and non-stimulus presentations. This osmolarity was maintained by adding an appropriate amount of sorbitol, adjusted as necessary to account for the osmolarity of any given chemosensory stimulus. We saw little qualitative difference between sensory neuron responses to odorants dissolved in water and odorants dissolved in CTX buffer[139].

M9 buffer (22 mM KH_2PO_4, 42 mM Na_2HPO_4, 85.5 mM NaCl, 1 mM $MgSO_4$) was also used for washing and loading animals into the microfluidics chip. An additional 2 mM tetramisole

hydrochloride was added to the M9 buffer when preparing animals for immobilized imaging.

2.4.4 Avoiding odor cross-contamination

In any chemosensation experiment, it is critical to avoid cross-contaminating different odorants, as animals are capable of detecting very low concentrations of chemicals, and many organic compounds will readily stick to plastic and glass surfaces. We took extensive precautions to ensure that the animals used in our experiments were only exposed to the intended stimuli. Each stimulus condition (given odor at a given concentration) was stored in its own glass bottle and delivered with its own glass syringe and tubing. Microfluidics chips were used for only a given set of stimulus conditions, and not reused for different stimuli. Gloves were used whenever the syringes and odors were being handled.

Special attention should be paid to the elastic tubing and the pinch valves. The elastic silicone tubing is more porous than harder plastic tubing, and small amounts of chemicals can leach through the elastic over time and contaminate the pinch valves. To mitigate this, the elastic tubing component was replaced more frequently than other parts of the syringe tubing. Additionally, we reserved the switches on the left side of the odor manifold for low concentration conditions, and used the switches on the right side of the odor manifold for high concentration conditions.

Additionally, odorants diluted in buffers containing salts have the potential to disable the experimental setup if improperly handled. Any spills or leaks on the odor manifold can damage the pinch valves. To avoid this, any spill was cleaned immediately, and the entire setup was regularly washed with deionized water. To prevent salt buildup in the tubing used to deliver odorants and buffer

blanks, every syringe and tube is washed at the end of an experiment by running pressurized deionized water through the syringe. It is also critical to avoid leaks which drip onto the objective and turret, as salt buildup within the microscope chassis is very difficult to clean. Additionally, even small amounts of liquid on the stage can cause the chip to move vertically as the piezo moves, resulting in z-drift in the recording. Therefore, before any new microfluidics chip is used on the microscope stage, it is first "proofed" by running pressurized deionized water through it. Likewise, any newly constructed syringes and tubing had their seals checked with pressurized deionized water prior to being loaded with odorants. Any chips with leaks due to cracks or poor bonding were discarded. The liquid expelled from the waste port must also not be allowed to build up to the point where it drips off of the top of the chip. We used a vacuum line attached to a glass pipette tip, which was secured to the microscope stage with putty.

2.5   DISCUSSION

The experimental odor delivery setup we developed for *C. elegans* has allowed us to characterize the neuronal responses of populations of animals to a broad panel of odorants in a high-throughput manner. The technology described in this chapter can be readily deployed to answer a number of other interesting questions in *C. elegans*, such as characterizing and quantifying the nature of adaptation in the chemosensory neurons or understanding changes in representation when two odorants are presented in rapid succession. Another interesting question we are in a position to pursue is the neuronal representation of mixtures of odorants. The experimental setup is capable of delivering

discrete mixtures, but the development of on-chip mixing will allow for much more sophisticated time-varying patterns of odor delivery[140]. Continuous changes in odor concentration as a function of time are more naturalistic than discrete odor pulses, and a gradual, continuous transition from 100% Odor A, 0% Odor B to 100% Odor B, 0% Odor A could shed light on shifts in neuronal representation, particularly if one odorant is attractive and the other is aversive.

The observed differences in neural activity between immobilized and freely crawling *C. elegans* suggest that a confined microfluidics channel may not be the best environment in which to study circuits in which motor feedback plays a critical role[135]. However, the precise control of the chemical environment experienced by an animal in this sort of chip makes it a valuable asset when pursuing questions of chemosensation. Looking forward to potential future whole-brain studies, it may make sense to combine immobilized microfluidics experiments like those described here with experiments in which the animal is allowed to navigate somewhat freely through an arena, allowing the animal freedom of motion at the expense of fine environmental control. ●

3

Computational methods for analyzing calcium imaging data

Vivek Venkatachalam, Amin Nejatbakhsh, Erdem Varol, and Liam Paninski contributed to the work described in this chapter.

RELIABLE EXTRACTION of calcium activity traces from volumetric recordings is a prerequisite for the study of neuronal ensembles. Recent advances in imaging techniques have enabled the capture of functional neuronal ensembles *in vivo* within a wide variety of animal models, including larval zebrafish, *Drosophila*, and *C. elegans*[133,141,121,125]. However, in larger organisms, whole-brain activity is often reported as fluorescence time traces within a given region of interest.

In contrast, the *C. elegans* brain is sufficiently small that the entire structure can be captured at single-neuron resolution. Extracting activity from these neurons, however, can be a challenge. The neurons in the head of the animal near the nerve ring are very densely packed, making it difficult to separate the calcium signals of neighboring neurons. Additionally, the *C. elegans* brain is

easily deformable, making it impossible to rigidly align images onto a brain atlas as is done in adult *Drosophila* and larval zebrafish. Instead, the neurons in *C. elegans* must be tracked in space as their relative position changes during a recording[117]. In this chapter, we will discuss our efforts to improve the accuracy and speed of neuron tracking in semi-immobilized animals.

3.1 Neuron tracking and fluorescence trace extraction

Even in the confined environment of a microfluidics chip, the worm can wriggle back and forth, distorting the neurons in the head of the animal. As a result, the relative location of neuronal nuclei can change over the course of an experiment. Confining GCaMP6s expression to nuclei makes each nucleus easier to identify, but the high density of labeled neurons in the heads of multi-neuronally and pan-neuronally labeled animals makes it difficult to separate nearby neurons using standard edge detection methods. Likewise, it is almost impossible to create genetically labeled strains in which GCaMP expression is uniform enough for baseline fluorescence levels to be equal across all labeled neurons. This precludes using standard peak detection methods to reliably segment neurons.

To solve this problem, we developed a hybrid method in which neurons were semi-automatically segmented in a single landmark frame, then tracked through time using a neighborhood correlation strategy.

3.1.1 Semi-automatic neuron segmentation

As a first pass to identify neuron centers, we use a straightforward combination of Gaussian filtering and peak detection to return putative neuron centers from each landmark volume. However, this step relies heavily on manual intervention to remove falsely labeled nuclei and to add nuclei that the peak detection algorithm missed, such as small or dim nuclei. To simplify this task, we built a GUI which allows users to navigate each 3D landmark image and click to add or remove neuron centers (**Figure 3.1A**). This GUI allows the user to toggle between multiple fluorescent channels and a maximum projection, allowing the user to take advantage of any fluorescent landmark labels in the strain. Complete labeling of all neuron centers is only necessary once for a given animal, even if multiple recordings have been made. The user then labels a small handful of widely spaced neurons (4-8) in the first frame of the activity recording. This small number of labeled neurons helps the tracking algorithm to compensate for any global motion or distortion that may have occurred in the animal between the landmark volume and the activity movie.

In addition to segmentation, the GUI allows neurons to be manually identified. The names the user applies are then associated with the activity traces of those neurons. For details on semi-automated cell identification in pan-neuronal multicolor animals, refer to Section 5.2.7.

3.1.2 Neighborhood correlation tracking of individual neurons

While the entire brain of the worm can distort significantly across large distances, the neighborhood immediately surrounding a neuronal nucleus of interest tends to remain consistent, with little local

46

deformation. Our image registration algorithm relies on this fact. Instead of attempting to identify neuron centers in every frame, we try to match the neighborhood surrounding the neuron center in the first frame to the most similar neighborhood in the following frame. We then return the center of the new neighborhood as the position of the neuron center in the next frame (**Figure 3.1B**).

We first employ this algorithm to map the neuron centers identified in the high-resolution landmark volume during the segmentation step onto the first frame of the activity movie, which is captured at a lower resolution. We then proceed to compare each frame of the movie to the next. Unlike previous tracking methods, which kept track of the relative positions of a neuron interest and a constellation of nearby neurons[117], the neighborhood correlation comparison is made independently for each neuron. While we lose some information about local deformations by not integrating information about how neighboring neurons are moving, we gain the ability to run the tracking of each neuron in a dataset as a parallel process, dramatically decreasing runtime. This also prevents a mistake in tracking one neuron from propagating to other nearby neurons. We run the tracking on a down-sampled version of the activity movie, also to improve runtime.

For a given neuron center, the tracking algorithm goes through the following steps:

1. Given the position of the given neuron center in the current frame, $n_t = (x_t, y_t, z_t)$, we identify the neuron's local 3D neighborhood N_t in that frame, the volume with dimensions $2a * 2b * 2c$, in the region spanned by $[x_t - a, x_t + a]$, $[y_t - b, y_t + b]$, and $[z_t - c, z_t + c]$.

2. We identify the naive center in frame $t + 1$, from where we begin our search for the neighborhood most similar to N_t. For the first frame of the movie, this point is adjusted by a distance-weighted average of the manually labeled neurons: $n'_{t+1} = (x_t + \Delta\Sigma wx^i, y_t + \Delta\Sigma wy^i, z_t + \Delta\Sigma wz^i)$. For any other frame, we simply take the naive center as the center of the previous frame, $n'_{t+1} = n_t = (x_t, y_t, z_t)$.
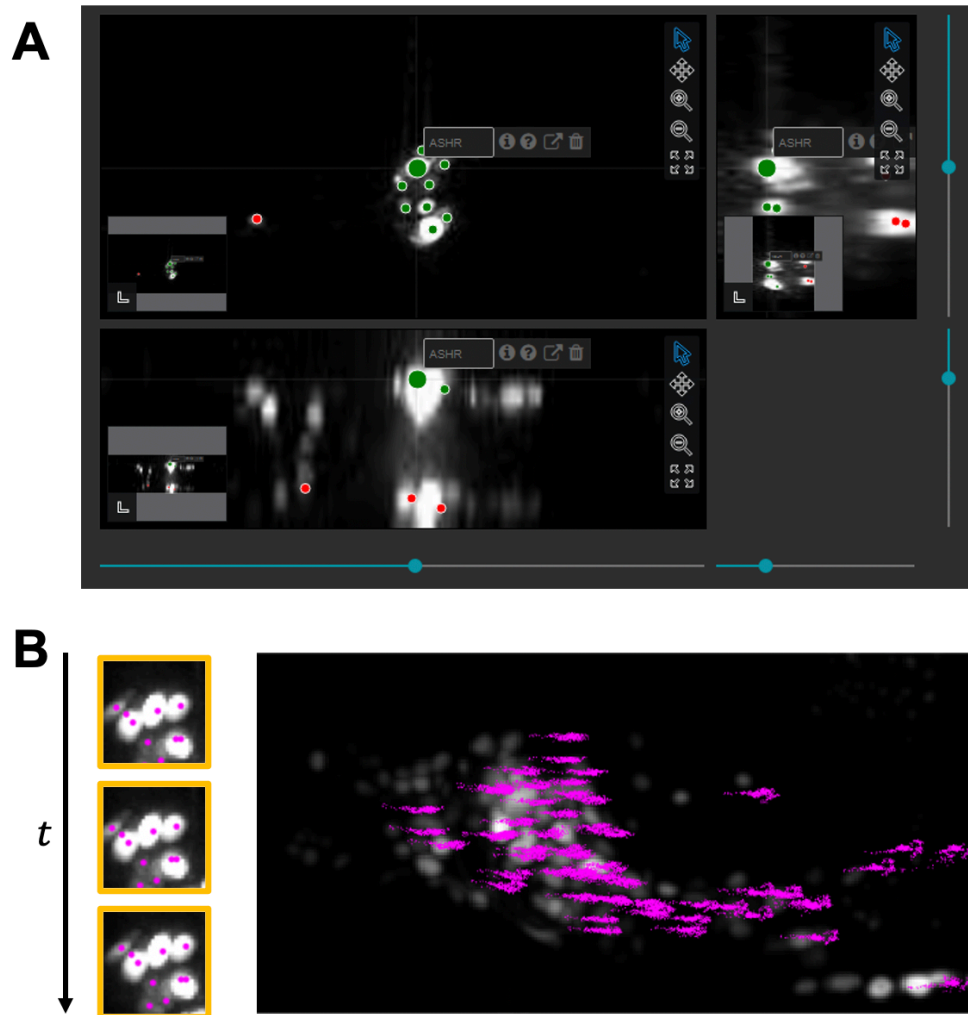
Figure 3.1: Neuron identification and neighborhood tracking. (A) A GUI for editing neuron centers prior to tracking. Neuron centers can be edited in three dimensions, and neuron names can be manually assigned. **(B)** We perform pairwise image registration (left) between the large 3D neighborhood N_t surrounding a given neuron center n_t and nearby neighborhoods in the following frame. The neighborhood N'_{t+1} best correlated with the original neighborhood N_t is identified, and the center of this new neighborhood is defined as the new neuron center n_{t+1}. This is then repeated for the duration of the activity movie. A subset of the tracked neuron centers in a representative pan-neuronally labeled animal is shown on the right. This animal was moving forward and backward in the microfluidics chip during imaging, motion that the algorithm has successfully captured.

3. Starting from the naive center n'_{t+1}, we perform image registration between the maximum intensity projections in x, y, and z of putative neighborhood N'_{t+1} and the previous neighborhood N_t, computing the pairwise correlation of these images. We then repeat this process, moving the putative center n'_{t+1} by 1 pixel per iteration until one of the following occurs:

   (a) The algorithm finds a putative neighborhood N'_{t+1} whose correlation with N_t exceeds the confidence threshold C (usually set at above 90%). This putative neighborhood is then defined as N_{t+1}.

   (b) The algorithm tests all putative neighborhoods within a maximum search radius r_{max} of the naive center n'_{t+1}, but failed to find a putative neighborhood whose correlation exceeds the confidence threshold C. The algorithm then returns the putative neighborhood with the highest correlation with N_t as N_{t+1}.

   (c) If no neighborhood is found with a correlation exceeding a minimum value, the neuron is considered lost in frame $t + 1$, likely either due to motion taking the neuron outside the region of interest. No center is reported, and the last reported neighborhood N_t is used as the basis of comparison for following frames ($t + 2, t + 3$, etc.).

4. The center of neighborhood N_{t+1} is defined as the neuron center in this frame, n_{t+1}.

5. Repeat until the end of the activity movie is reached.

The user can optimize the tracking parameters such as neighborhood size (a, b, c), maximum search radius r_{max}, and confidence threshold C for both accuracy and speed their type of data. The optimal parameters will be different for different imaging conditions, and are particularly sensitive to the temporal and spatial resolution of recordings.

3.1.3   EXTRACTING CALCIUM DYNAMICS

To extract calcium signals, we first map the positions of each tracked neuron center back onto the original-resolution volumetric images. We then extract fluorescence values from these images. We
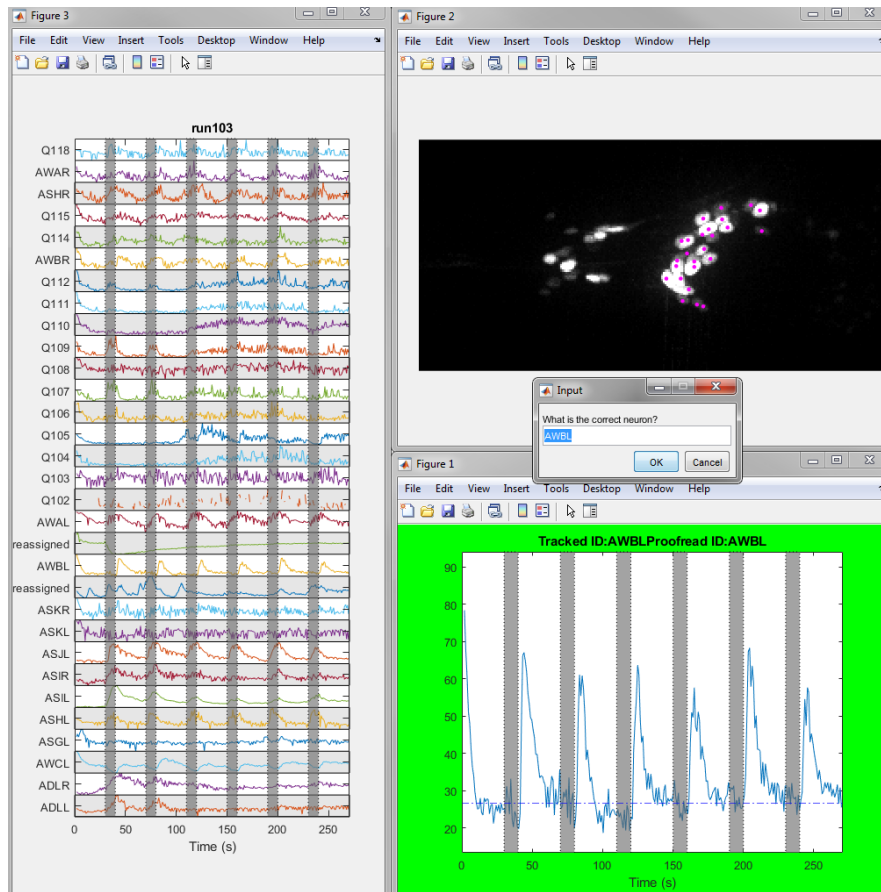
Figure 3.2: GUI for manual proofreading of calcium traces. This interface allows us to simultaneously assess the trace of a given neuron (lower right), traces of nearby neurons (left), and the spatial neuron track (upper right). Manual proofreading allows us to exclude bad tracks, manually adjust baseline fluorescence F_0, and correct mis-identified neurons.

identify a small volume around each neuron center, containing voxels whose fluorescence will be assigned to the neuron. This volume is set as 2 μm x 2 μm x 3 μm for our data, but can be adjusted by the user for the resolution of their data. We compute the mean of the 10 brightest pixels within this volume to extract a raw fluorescence trace $F_r(t)$. We then account for photobleaching by exponential detrending, giving us a clean fluorescence activity trace $F(t)$. We then identify the background fluorescence F_0 for each neuron, and report normalized neuron activity $\Delta F/F_0$.

3.1.4 MANUAL PROOFREADING OF TRACES

The manual proofreading step gives us the opportunity to improve the quality of the data by removing neurons which have been mistracked, adjust the computer-determined baseline fluorescence F_0, and to correct or add nuclear IDs. It also enabled us to remove traces which were contaminated by the signals of neighboring neurons. We built a GUI which presents the user with all relevant information about a given neuron to contextualize its trace, including a plot of its 3D spatial track and the activity its neighbors (**Figure 3.2**). The software then compiles all of the processed traces for a given individual into a single data structure.

For an operations manual for the complete code package, including detailed descriptions of the scripts, refer to Appendix B.1.

3.2 DISCUSSION

Neighborhood correlation with manual proofreading has given us high-precision tracks of even densely packed neurons, and has allowed us to greatly increase the throughput of neuron tracking

in whole-brain labeled *C. elegans*. It is now possible to extract clean signals from densely packed neurons in semi-immobilized animals with minimal human intervention.

However, many potential areas of improvement remain, particularly when attempting to track neurons in unconstrained animals. Currently, highly reliable tracking of densely-labeled neurons in freely moving *C. elegans* can only be accomplished with manual proofreading[136]. Furthermore, there is great interest in extracting signals from cytoplasmic GCaMP, a truer readout of neuron activity than nuclear-localized GCaMP. Being able to do this in a multi-neuronal context, however, requires the segregation and tracking of multiple neuron masks, each of which can be distorted in three dimensions. We are currently working to address these limitations by employing deep learning image tracking methods, and hope that these new approaches can further improve the throughput and quality of data processing for multi-neuronal labeled *C. elegans* lines. ●

4

Ensemble representations of olfactory stimuli in *C. elegans*

Vivek Venkatachalam, Greg Cain, Nicolas Tan, Raymond Valenzuela, Shanshan Qin, Cengiz Pehlevan, Wesley Hung, Min Wu, and Mei Zhen contributed to the work described in this chapter.

FOR MANY ANIMALS, chemosensation is the primary sensory modality through which they perceive the world. To detect and identify a wide range of chemical compounds, animals employ a large number of chemosensory neurons, making olfactory responses inherently collective responses. Thus, multi-neuronal imaging methods are ideal for understanding the neuronal basis of odor coding. Here, we studied the ensemble-level representation of odor identity in the nematode *C. elegans*. We recorded from all chemosensory neurons in the animal simultaneously while presenting a broad panel of olfactory stimuli in a highly controlled manner using a microfluidics device. Observing the dynamics of these neurons has allowed us to build a quantitative and comprehensive picture of the way the olfactory system in *C. elegans* consolidates and represents high-dimensional sensory infor-

mation. We found that collectively, the sensory neurons encode odor identity and intensity. We also described the roles of each of the individual sensory neurons in olfactory coding, finding diverse dose response and tuning properties across neuron classes.

4.1 INTRODUCTION

Nearly all organisms rely on olfaction to navigate the world, find food, and avoid dangerous environments. To do so, they rely on olfactory systems which are capable of determining the identity and concentration of large numbers of chemically diverse odorants. However, how the chemosensory system in the nematode *C. elegans* encodes this high-dimensional information is unknown.

Olfactory coding has been extensively studied in mouse [52,58], as well as both the larval and adult forms of the fruit fly *Drosophila melanogaster* [54,55,56,57]. Insects and mammals have olfactory systems which are broadly similar in structure [45,46,47]. The first, environment-facing layer consists of olfactory receptor neurons, each of which nearly always expresses only one unique receptor type. These receptors are activated by multiple odorants, and each odorant binds to multiple receptors [52,53]. Each odorant can therefore activate a distinct subset of sensory neurons, allowing different odorants to be discriminated. This combinatorial receptor coding strategy allows a finite number of receptors to encode a very large number of odorants. Odorants at higher concentrations will also activate a larger number of sensory neurons [60,61,55]. In insects and mammals, the olfactory receptor neurons converge onto a number of glomeruli, which synthesize the information from multiple sensory neurons before passing it on to higher processing regions [49,50,51]. There is spatial sep-

aration between the higher brain regions responsible for innate behaviors (lateral horn in insects, amygdala in vertebrates) and learned behaviors (mushroom body in insects, piriform cortex in vertebrates)[51,62,63,64,65].

When compared to *C. elegans*, however, these higher organisms employ a far larger number of neurons to detect and identify olfactory stimuli. The compact nervous system of *C. elegans* consists of 302 neurons, and the animal has just 11 pairs of chemosensory neurons[78,79]. Despite its small size, the *C. elegans* olfactory circuit is is capable of solving the same computational problems faced by higher organisms—guiding the animal towards food and away from dangers[66]. In *C. elegans*, every sensory neuron in expresses multiple receptor types[79,80]. Only one of these receptors (ODR-10), however, has been fully characterized in the context of olfaction[69,70]. Many of these neurons are also polymodal—in addition to olfactory stimuli, some detect gustatory stimuli such as pheromones, salts and amino acids, gases such as oxygen and CO_2, or temperature[79,81]. We also know from the anatomical wiring diagram of *C. elegans* that many of these sensory neurons are wired to each other or receive feedback from interneurons[8,9] (**Figure 4.1C**). This again contrasts with the olfactory systems of higher organisms, in which the first layer of olfactory neurons is largely feed-forward. Because of these characteristics, the *C. elegans* chemosensory system can be thought of as fitting somewhere between the more distributed architecture we see in insect and vertebrate olfaction and the intracellular computations performed within a single chemotaxing bacteria.

The multi-receptor sensory neurons in *C. elegans*, which synapse directly to interneurons where behavioral decisions are being made, may be thought of as analogous to both the olfactory receptor neurons and the glomeruli of higher organisms, their functions combined into a single cell. The

interneurons downstream of the chemosensory neurons can be classified into two groups (**Figure 4.1C**). The first interneuron group consists of AVA, AVB, and AVD, command interneurons which control forward and backward locomotion[8,9]. This group receives the most inputs from ASH, a nociceptive sensory neuron[72,73,74,76]. These three premotor interneurons are responsible for reflexive avoidance behaviors in response to aversive stimuli, and can thus be thought of as analogous to brain regions for innate responses such as the lateral horn and amygdala in higher organisms. The second interneuron group consists of AIA, AIB, AIY, AIZ, neurons have been shown to play roles in controlling goal-directed locomotion and learning behaviors[84,85,86,82]. These neurons may be analogous to the brain regions in higher organisms which influence learning and decision-making.

Behavioral experiments have identified odorants which are attractive or repulsive to the animal[67,68], and previous calcium work has characterized the responses of single chemosensory neurons to a handful of odorants[87,88,89,90,91,94,139]. However, a complete understanding of olfactory coding in *C. elegans* has not yet been achieved. How does *C. elegans* use its small ensemble of chemosensory neurons to encode information about odorants in the environment? Is the animal only capable of determining if an odorant is attractive or repulsive, or does it have sufficient information to identify odorants? And to what extent are the responses in the sensory ensemble shaped by communication between neurons?

We set out to build a comprehensive picture of the activity of the entire *C. elegans* chemosensory neuron ensemble in response to olfactory stimuli. We built a transgenic animal in which the 11 chemosensory neuron pairs could be identified and simultaneously recorded from, and employed microfluidics to deliver a broad panel of olfactory stimuli in a highly controlled manner at multiple

concentrations. We find that odor-evoked responses are more widespread than previously known, and that the chemosensory neuron ensemble acts as a spatial map to encode odorant identity: each odorant activates a unique combination of sensory neurons. These sensory representations of olfactory stimuli are sufficiently reliable for odorants to be identified from the neural responses to a single odor presentation. We find diverse response properties across the 11 chemosensory neuron classes: some neurons are broadly tuned, responding to a wide range of olfactory stimuli, while others are much more narrowly tuned, responding to only a few odorants in the panel. Dose response curves are also different for different odorants, even within the same neuron. Ablation and mutant experiments indicate that the sensory representations elicited by odorants are not dependent on synaptic connections. These results suggest that despite its small size, the *C. elegans* olfactory system can use ensemble representations to robustly determine both odor identity and odor intensity.

4.2 Results

4.2.1 Labeling and recording from the chemosensory neurons

We developed a new *C. elegans* imaging line, ZM10104, which enabled us to simultaneously record from and identify the 11 pairs of chemosensory neurons contained within the amphid chemosensory organs: AWA, AWB, AWC, ASE, ASG, ASH, ASI, ASJ, ASK, ADL, and ADF[79]. GCaMP6s expression was driven by the *ift-20* promoter, labeling all ciliated sensory neurons in the animal. GCaMP6s expression was localized to the nuclei of these neurons to allow signals from neighboring neurons to be better segmented. The relative position of these neurons can vary from individual to

individual[139]. To allow these neurons to be reliably identified, we used wCherry driven by the *gpc-1* as a sparse landmark, labeling the chemosensory neurons AWB, AWC, ASH, ASE, ASI, and ASJ (**Figure 4.1B**). To minimize the chances of incorrect identification, neuronal IDs for each odorant condition were reviewed by at least two individuals, and ambiguous neurons were omitted from the analyzed datasets.

Odor solutions were delivered using a microfluidic device adapted for *C. elegans* from a design for *Drosophila* larva[55,139]. The tapered end of the channel was positioned to allow odors in the delivery channel to flow past the nose of the animal (**Figure 4.1A**)[138]. Individuals were repeatedly presented with six 10 s odorant pulses separated by 40 s buffer blanks, and neuronal calcium activity was volumetrically captured at 2.5 Hz with a spinning disk confocal microscope (**Figure 4.1D**). We carried out experiments for over 70 conditions (23 odorants, each at multiple concentrations), with an average of 100 odor presentations across multiple individuals for each condition.

4.2.2 ODOR IDENTITIES ARE ENCODED BY THE SENSORY NEURON ENSEMBLE

We studied a broad panel of 23 odors, selected from a set of 122 odors previously studied in C. elegans[67,68]. We constructed an odor space based on the chemical descriptors of each previously studied odorant, and selected odors from different chemical families which span most of this space (**Figure 4.2A, Figure C.4**)[142]. We conducted experiments in which we presented each of these odorants to populations of animals at multiple concentrations, and recorded the responses of the 11 pairs of chemosensory neurons in response to each of these conditions (**Figure 4.2B-D, Figure C.2**).
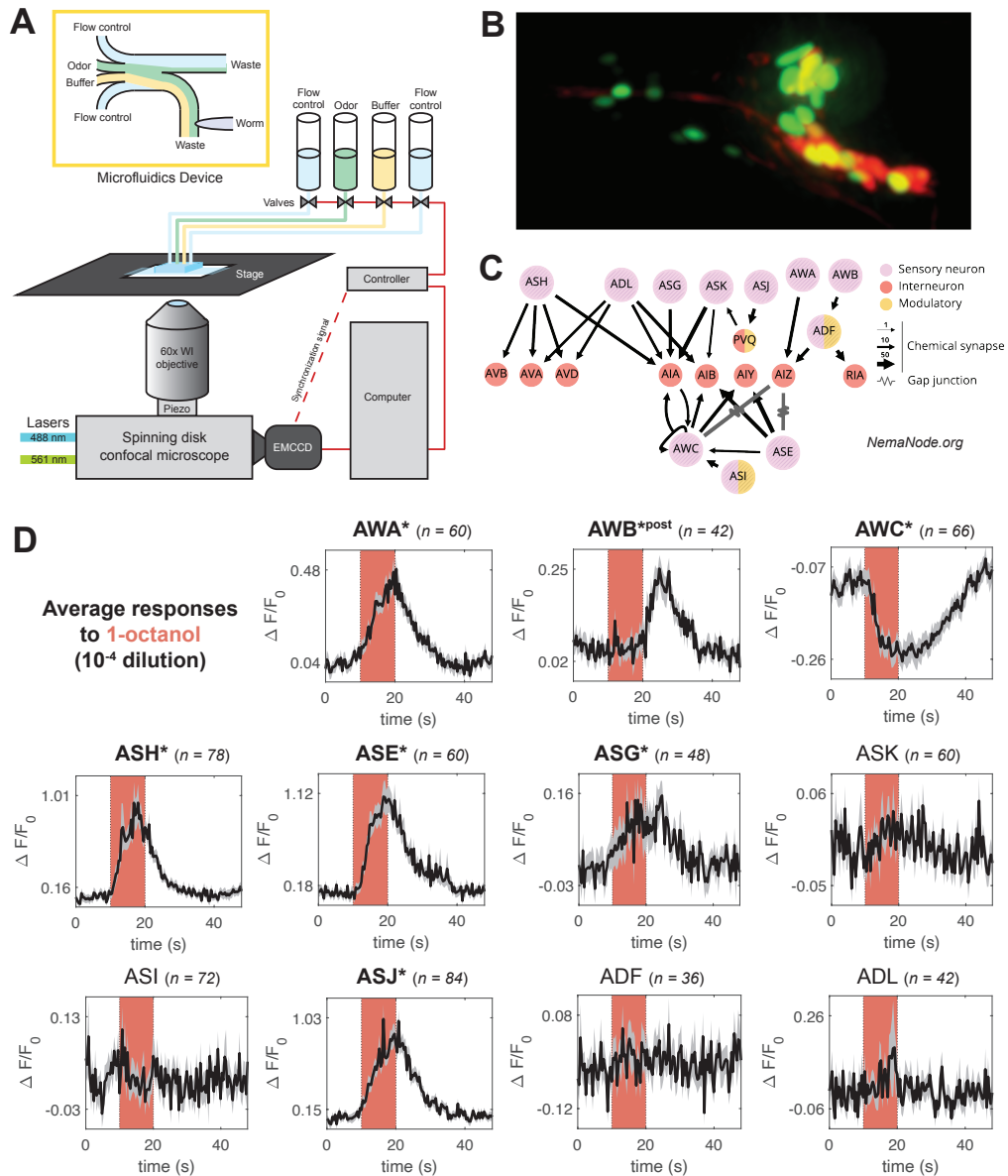
Figure 4.1: Labeling and recording from the chemosensory neurons. (A) Adult *C. elegans* were immobilized inside a microfluidics device and presented with odor solutions in a highly controlled manner. Each animal was volumetrically imaged at 2.5 Hz with a spinning disk confocal microscope during odor presentations. **(B)** The animals expressed nuclear-localized GCaMP6s in all ciliated sensory neurons. A sparsely expressed wCherry landmark allowed us to identify the 11 chemosensory neurons. **(C)** Reliable downstream partners of the 11 chemosensory neurons identified in the *C. elegans* connectome [8,9]. **(D)** Average neuronal activity traces of the 11 chemosensory neurons in response to a single odorant condition (1-octanol, 10^{-4} dilution). The 10 second odorant delivery period is indicated by the colored bar. Significant responses ($q \leq 0.01$) are marked with stars, with "post" indicating a significant response to stimulus removal.)

59
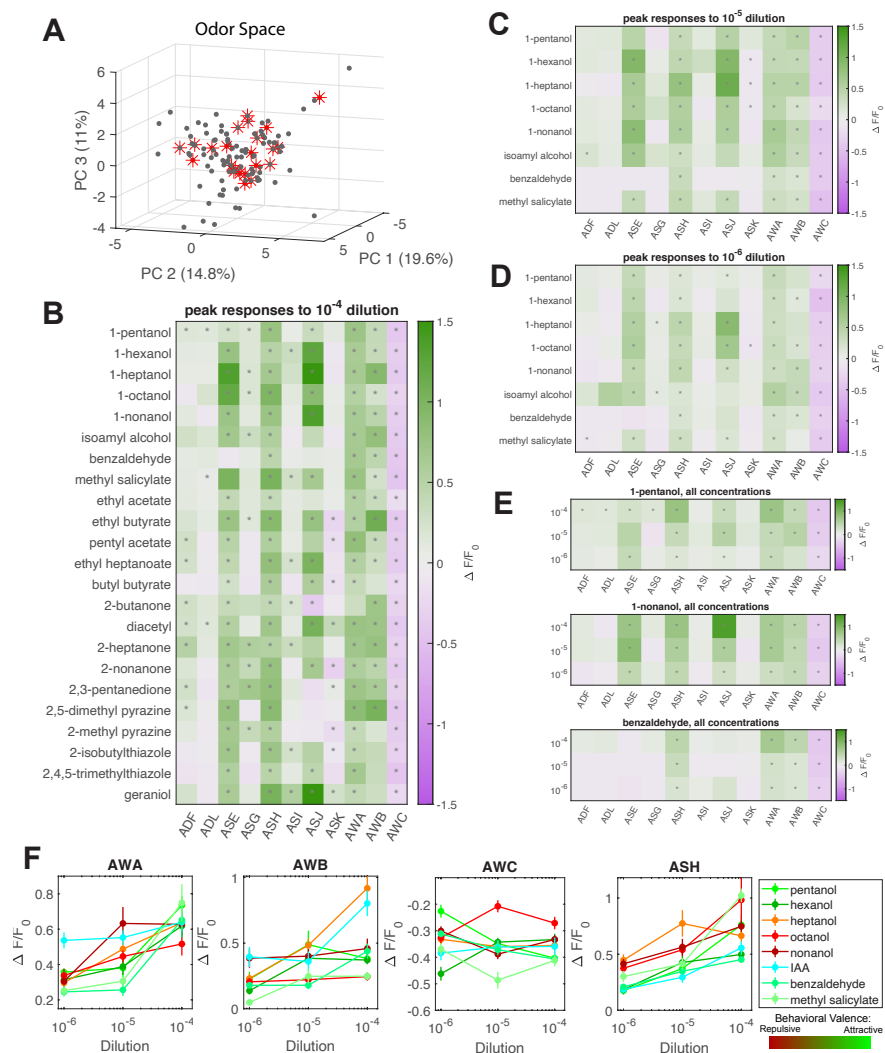
Figure 4.2: Ensemble responses to a broad panel of odorants. (A) An odor space constructed from the molecular descriptors of 122 odorants (gray) previously studied in *C. elegans*. We selected for our experiments a panel of 23 odorants (red) which span the odor space. Average peak responses of the 11 chemosensory neurons to odorants at **(B)** high concentration (10^{-4} dilution), **(C)** medium concentration (10^{-5} dilution), and **(D)** low concentration (10^{-6} dilution). Peaks were computed from a time window from the onset of odor delivery to 10 s after odor removal. Responses are reported as $\Delta F/F_0$, and significant responses ($q \leq 0.01$) are indicated with stars. Most odorants elicit significant responses from unique combinations of neurons. **(E)** Compiled responses to three representative odorants at multiple concentrations (1-pentanol, 1-nonanol, and benzaldehyde) show similar neural responses across concentration. The magnitude of neuron responses generally increases with increasing concentration, and for some conditions, additional neurons are recruited at high concentration. **(F)** Dose responses of AWA, AWB, AWC, and ASH are diverse, with distinct concentration-dependent curves in response to different odorants. ASH responses segregate by behavioral valence, with aversive odorants (orange-red) eliciting larger responses than attractive odorants (green-blue).

We found that a larger number of chemosensory neurons were activated by olfactory stimuli than previously known. For example, previous calcium imaging work had uncovered responses to isoamyl alcohol in the neurons AWA, AWB, and AWC, with the nociceptive neuron ASH being activated at high concentrations[87]. we recapitulated these responses to isoamyl alcohol, and identified novel reliable responses in neurons ASE and ASG. Across the 23 odorants, we largely recapitulated previously described responses to tested odorants, such as AWA to diacetyl and AWB, AWC, and ASE to 2-butanone[88,143,139]. We also uncovered many new significant responses (**Figure 4.2B**).

In the process of analyzing these data, we also made many neuron-specific observations. For instance, the neuron AWB is often activated upon the removal of an odorant (an OFF response), as can be seen in its response to 1-octanol (**Figure 4.1D**). For a small number of odorants, however, such as diacetyl and high-concentration isoamyl alcohol, we observe AWB being activated by the onset of the odor stimulus (an ON response). This is consistent with previous work which has demonstrated that AWB is capable of both ON and OFF responses in response to different concentrations of isoamyl alcohol[87] (**Figure C.3A**). Most of the chemosensory neurons have excitatory responses to stimulus, but some are inhibited by the onset of odor stimuli. Previous work has shown AWC to be inhibited by several odorants, including diacetyl, benzaldehyde, and 2-butanone[82,88,89,139]. We found that AWC is inhibited by every odorant in our panel (**Figure 4.2B**). We also observed inhibitory responses in ASK to a handful of odorants, such as ethyl butyrate and 2-nonanone (**Figure C.3B**), and an inhibitory response in ASJ to 2-butanone (**Figure C.3C**).

The majority of the sensory neuron responses we observed were symmetric—each left/right neuron pair had similar average responses. AWC is known to respond asymmetrically to some stimuli,

but because our transgenic animal does not express labels for AWC^{ON} and AWC^{OFF}, we cannot distinguish the two AWCs independent of their neural activity. ASE is known to have deterministic L/R asymmetric responses to salts[92,93]. We have observed significant asymmetric responses in ASE in response to a a handful of odorants, including ethyl heptanoate and butyl butyrate (**Figure C.3D**).

For most of the analyses presented in this study, we will be considering the peak responses of each sensory neuron class, averaged across L/R pairs. This is necessarily a coarse-grained view of chemosensory neuron activity, disregarding differences in dynamics, ON or OFF responses, and asymmetric responses. Additionally, neither intracellular activity in the neurites[25,26,27], nor any spiking signals which could encode stimulus information[24], is captured by our nuclear-localized GCaMP. Each of these details could contribute additional information to the downstream neurons. Thus, the information captured by the peak responses represents a lower bound on the information capacity of the chemosensory neuron ensemble.

Looking at the significant average peak responses to high-concentration odorants, we found that nearly every odorant in our panel reliably activated a unique combination of chemosensory neurons (**Figure 4.2B**). The chemosensory neuron ensemble is acting as a spatial map which contains sufficient information to encode odorant identities. Across odorants, there is significant overlap in significantly responding neurons, with neurons AWA, AWB, and AWC significantly responding to most of the odor panel. These results suggest that *C. elegans* may code odorant identities combinatorially, as is seen in higher organisms.

Another way of representing the neural activity of the ensemble of chemosensory neurons is

to compute the pairwise correlations of their average activity traces. These correlation maps are different for different odorants (**Figure C.5A-B**), supporting the conclusion that the ensemble activity is unique to each odorant.

4.2.3 ADDITIONAL NEURONS ARE RECRUITED WITH INCREASING CONCENTRATION

We presented animals with different concentrations of each odorant, spanning two to three orders of magnitude. Comparing chemosensory neuron responses for a given odorant across concentrations (**Figure 4.2B-D**), we see that for most significantly responding neurons, the response magnitude (as measured by peak fluorescence) increases as concentration increases. Neurons which are significant responders at low concentrations almost always remain significant responders at higher concentrations, indicating that neural representation of a given odorant is similar across concentrations. We see this in the compiled responses to three representative odorants: 1-pentanol, 1-nonanol, and benzaldehyde (**Figure 4.2E**). For 1-pentanol, the subset of neurons ASH, ASJ, AWA, and AWC is activated across all tested concentrations. For 1-nonanol, this subset contains neurons ASE, ASH, ASJ, AWA, AWB, and AWC, and for benzaldehyde, the subset consists of neurons ASH, AWA, AWB, and AWC. Note that for each of these three odorants, the sets of consistent responders are distinct from each other, suggesting that at low concentrations, odor identity can still be determined from ensemble responses. By considering the magnitude of the neural responses, it is also possible to decode the concentration of the odorant.

For many odorants, such as 1-pentanol, 1-octanol, and methyl salicylate, we observe recruitment of additional chemosensory neurons as concentration increases. In the case of 1-pentanol (**Figure**

4.2E), AWB is not significantly activated until the 10^{-5} dilution, and neurons ADF, ADL, and ASG are not significantly activated until the highest tested concentration, the 10^{-4} dilution.

Across all odorants, we see that at low concentrations neurons are less active, and at high concentrations neurons are more active and more reliably engaged. From these results, we can conclude that the chemosensory neuron ensemble encodes both odorant identity and odorant intensity. It is easier to identify an odorant at high concentrations due to the additional neurons engaged.

These conclusions are supported by phase trajectory analysis of the activity traces of the chemosensory neurons (**Figure C.5B**). Plotting low-dimensional representations of chemosensory ensemble activity in response to different stimuli in the same PC space, we see that the response trajectories of different odorants segregate. Response trajectories of different concentrations of the same odorant are aligned in direction but differ in magnitude.

4.2.4   COMPARING AVERAGE NEURAL REPRESENTATIONS OF ODORANTS

To assess the degree of similarity between odor representations, we clustered the odorants by their peak average neuronal responses (**Figure 4.3A**). Comparing these clusters to the chemical class of each odorant, we found that certain chemical classes, such as alcohols and ketones, have more similar neural representations within the class. Other chemical classes, such as esters, have more diverse neural representations.

Constructing a PC space from the standardized peak average neural responses, an alternate representation of neural space, we again see that classes such as alcohols and ketones are similar, while other classes are spread throughout the space (**Figure 4.3B**). From the loadings of the first three
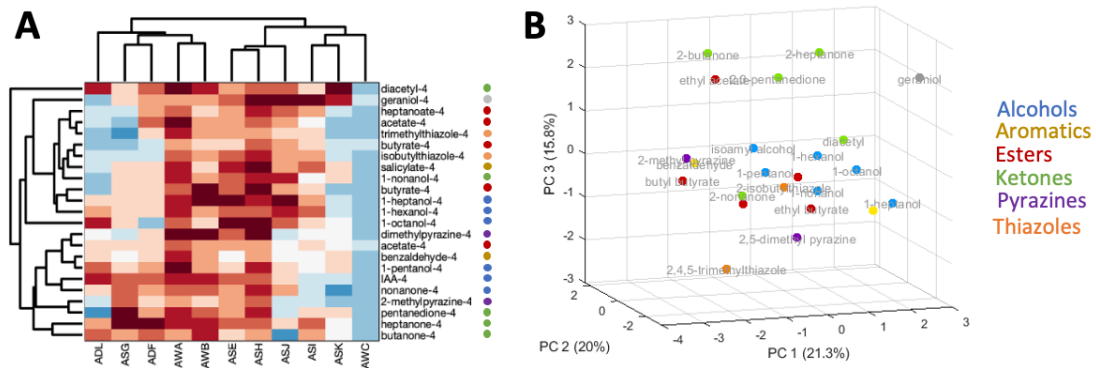
Figure 4.3: Odorants have distinct average neural representations. (A) Odorants (high concentration) clustered by their peak average neuronal responses. **(B)** A PC space built from standardized peak average neural responses. Chemical class is indicated by color. In both representations of neural space, we see that some groups of classes of odorants, such as alcohols and ketones, have more similar neural representations, while other classes of odorants, such as esters, have more diverse representations. Refer to **Figure C.6** for PC loadings.

principal components, we see that each of the PCs is broadly supported by the 11 chemosensory neuron classes, indicating that all neurons contribute to the animal's ability to distinguish odorants (**Figure C.6A**).

4.2.5 SINGLE-TRIAL RESPONSES ARE SUFFICIENT FOR ODOR IDENTIFICATION

While many neurons are reliably activated during odor presentation, the magnitude of their peak responses can vary across trials. Additionally, in response to the same stimulus, some neurons get weakly activated in some individuals but remain silent in others. Given these sources of variability, are the neuronal responses to a single trial (one odor pulse) sufficient to decode the identity of the odorant presented?

We compiled the single-trial peak responses of each odorant, imputing the data of missing neurons (Refer to **Appendix C.1.4** and **Figure C.7** for a detailed explanation of how missing data
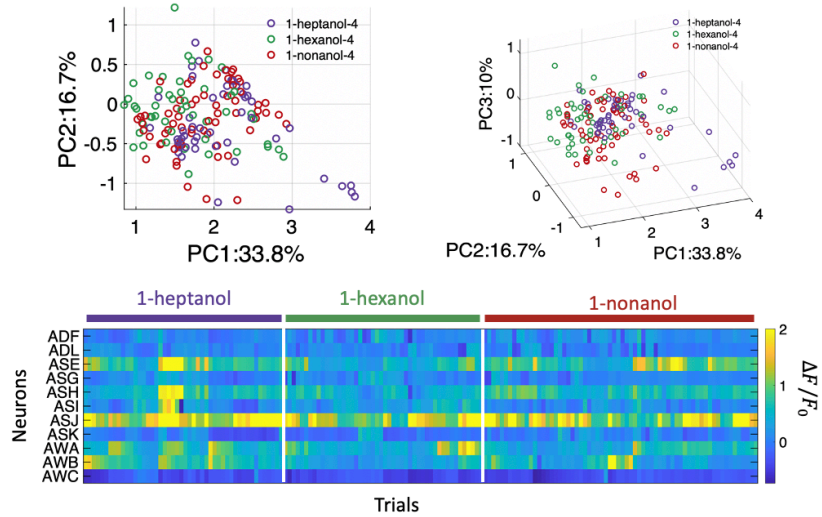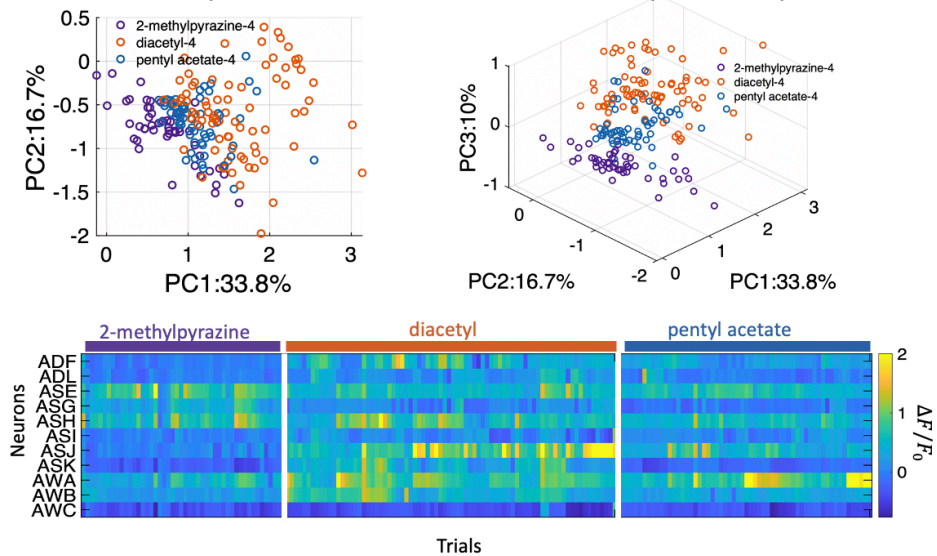
Figure 4.4: Representative comparisons of single-trial odorant responses. Representations of single-trial peak neural responses to sets of **(A)** three similar and **(B)** three dissimilar odorants. In each case, the PC space is constructed from the trial responses to only the three featured odorants. Single trial peak responses to the odorants is shown in the heatmap below, with data for missing neurons imputed (see **Appendix C.1.4** and **Figure C.7** for details). **(A)** We see that three similar odorants, the straight-chain alcohols 1-hexanol, 1-heptanol, and 1-nonanol, have more similar neural representations. **(B)** In contrast, three odorants of three distinct chemical classes, 2-methylpyrazine (a pyrazine), diacetyl (a ketone), and pentyl acetate (an ester), have more easily separable neural representations.
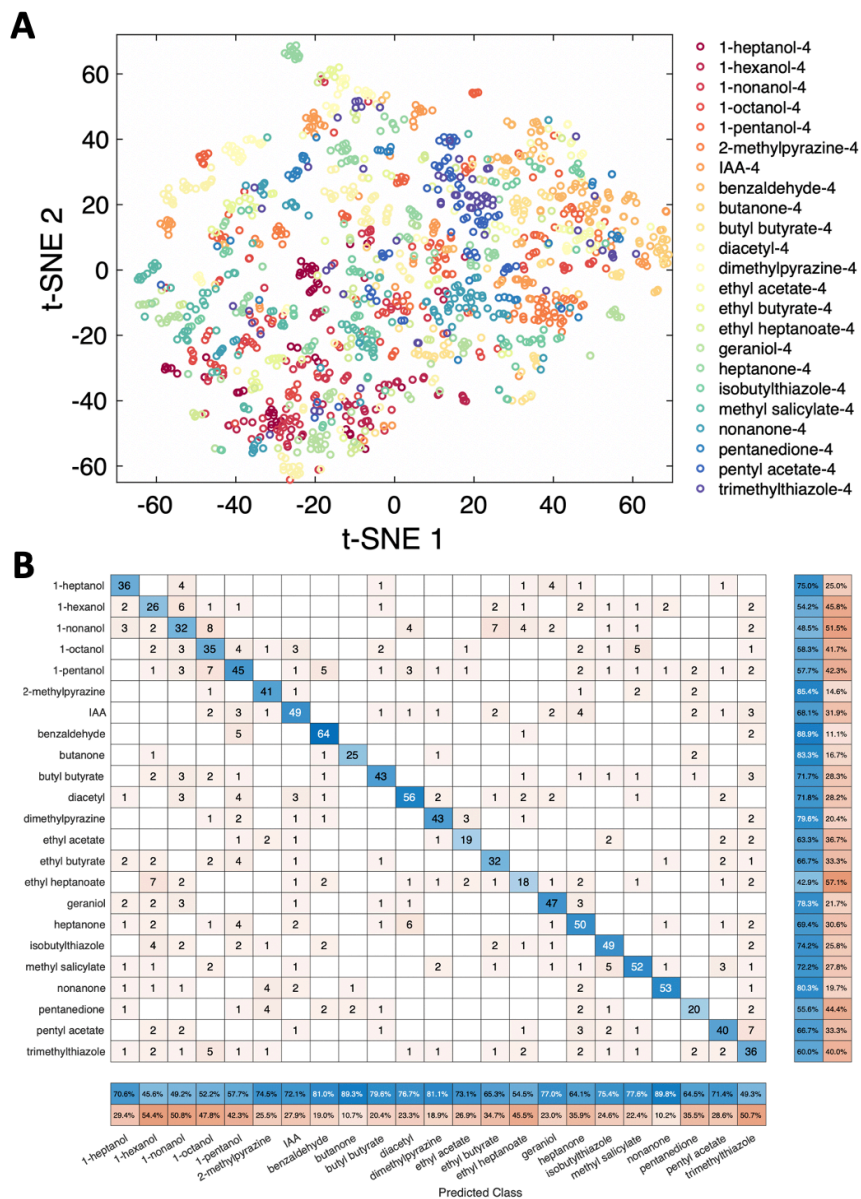
Figure 4.5: Odorants are distinguishable based on single-trial neural responses. (A) t-SNE representation of single-trial neural responses to all 23 odorants (high concentration). Responses to a given odorant generally cluster together. **(B)** Linear discriminability analysis of single-trial data. The majority of single trials are correctly classified by odorant delivered.

was imputed). We found that chemically similar odorants have more similar representations and are thus are more difficult to distinguish (**Figure 4.4A**), while chemically dissimilar odorants are more easily separable (**Figure 4.4B**). Given that over 40% of the variance is unexplained by the first three principal components, overlap in PC space does not necessarily mean that the odorants are indistinguishable. From a t-SNE representation of the single-trial responses to all 23 odorants in a low-dimensional space, we see that trials of the same odorant tend to group together, even if low-dimensional representations of different odorants overlap (**Figure 4.5A**).

To more rigorously test whether single-trial responses contain sufficient information to discriminate odorants, we first performed binary classification: attempting to separate the pooled single-trial peak responses of each odorant pair (**Figure C.8**). Both linear regression (**Figure C.8A**) and SVM (**Figure C.8B**) return low classification errors, demonstrating that the single-trial peak responses of any two odorants are linearly separable.

Given that all the pooled single-trial neural responses of any pair of odorants were separable, we then asked how well one could decode odor identity when the single-trial responses to all odorants were pooled together—a task significantly more challenging than binary classification. We built a multi-class classifier to perform linear discriminability analysis. The pooled single-trial data was randomly divided into a training set (90%) and validation set (10%).The classifier was trained on these data, then asked to predict odorant identity for each of the withheld single-trial neural responses in the validation set (refer to **Appendix C.1.5** for details). We found that the majority of single-trial responses were correctly classified (**Figure 4.5B**). From this, we conclude that the spatial map which encodes odor identity is robust enough that the responses to just a single odor pulse are sufficient to

reliably identify the presented odorant.

4.2.6 Diverse dose responses across neurons and odorants

In insect and mammalian systems, olfactory receptor neurons have been shown to have very similar dose response curves across neurons and across odorants[55,144,145]. We do not observe the same similarity in dose responses in *C. elegans*. Combining peak response data at multiple concentrations, we generated dose response curves for each chemosensory neuron in response to each of the odorants in the panel (**Figure 4.6**). We find diverse dose response curves across the chemosensory ensemble: the same neuron has different response curves to different odorants, and the same odorant elicits different response curves in different neurons.

Most of the odorants in the panel are attractive at low concentrations, but a handful, including 1-heptanol, 1-octanol, and 1-nonanol, are aversive at all concentrations (**Appendix C.1.3**)[67]. At high concentrations, nearly all odorants are behaviorally aversive. The sensory neuron ASH has been characterized as a nociceptive neuron, responding to aversive stimuli[72,73,74,76]. Consistent with this, every odorant in our panel elicits strong ASH responses when presented at high concentration (**Figure 4.2B**). The dose response curves of ASH in response to different odorants segregate by behavioral valence: aversive odorants such as octanol or nonanol elicit larger responses than the same concentration of attractive odorants, such as hexanol or isoamyl alcohol. This is particularly evident at low concentrations. We do not see this segregation between attractive and aversive odorant responses in the other 10 chemosensory neurons. ASH is unique in its connectivity, making connections with the command interneurons AVA, AVB, and AVD. We therefore expect that ASH alone
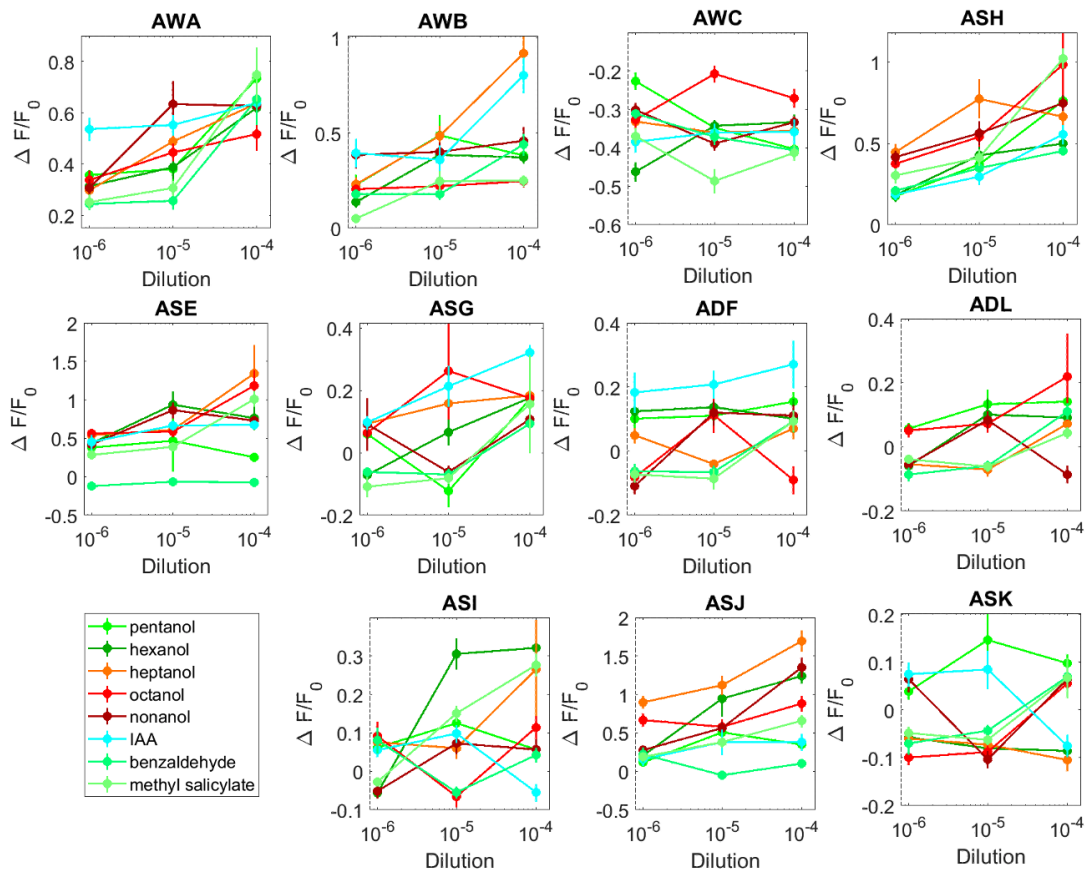
Figure 4.6: Dose responses of all 11 chemosensory neuron classes to a subset of the odorant panel. The low-concentration behavioral valence of odorants is indicated by color: aversive odorants are orange-red, and attractive odorants are green-blue.

encodes behavioral valence for reflexive behaviors.

For neurons such as AWA and AWB, we find distinct concentration-dependent curves in response to different odorants. For some odorants, the neuron's dynamic range is large across concentrations (such as AWA responding to 1-pentanol, or AWB responding to 1-heptanol). For other odorants, however, the same neuron's dynamic range across concentrations can be much smaller (such as AWA responding to isoamyl alcohol, or AWB responding to 1-octanol). Some neurons, such as AWC, ADF, or ASK, have very steep response functions, appearing to be activated or inhibited above a certain concentration and silent below it. Previous work has shown that when the ODR-10 receptor (natively expressed in AWA and sensitive to diacetyl) is instead expressed in AWB, a previously attractive concentration of diacetyl becomes repulsive to the animal[70], suggesting that AWB is activated by aversive stimuli. AWB has also been shown to be necessary for learning aversive olfactory cues[82]. From our dose response curves, we do not see the same segregation between attractive and aversive odorants in AWB as we do in ASH. This suggests that AWB may be broadly tuned to olfactory stimuli, rather than being strictly a nociceptive neuron.

The likeliest explanation for the diversity in dose responses lies in the expression of multiple receptor types in each neuron. A single neuron may use different receptors (or even different combinations of receptors) to sense different odorants, and may also be performing intracellular computations to synthesize the information from different receptors. Only one receptor has ever been fully characterized in *C. elegans* olfaction, ODR-10, which is localized to the neuron AWA and responds to diacetyl[69,70]. However, we know that AWA responds to many other odorants, and diacetyl activates other neurons. Until more is known about the sensitivities of individual receptors, it will be

difficult to extract insights about receptor tuning from the activity of the sensory neurons alone. The multi-receptor strategy also gives evolution the ability to tune the number and type of odorants each neuron responds to.

Some caveats about dose responses in *C. elegans*: we know that in *C. elegans*, the peak response of a neuron can change in a nonlinear fashion depending on the length of an odor presentation[88]. Additionally, because each of these sensory neurons expresses many receptor types, we cannot extract molecular parameters of a receptor from the dose response curve of a chemosensory neuron, as is commonly done in model systems in which only one receptor type is expressed, such as in insect or mammalian olfactory systems. Nor can we assume that the same receptor (or same set of receptors) is being activated when the neuron is presented with different odorants. Thus, these dose responses are best interpreted as a relative comparison of a neuron's responses to different odorants in the context of this experimental design.

4.2.7 Sensory neurons are broadly or narrowly tuned to olfactory stimuli

For sensory neurons engaged in olfaction, one can define tuning as the number and types of chemical compounds a given neuron responds to. In higher organisms, sensory neuron tuning is governed predominantly by the properties of the chemosensory receptor: the affinity of the receptor defines the set of chemical compounds the sensory neuron is activated by[55]. In the multi-receptor sensory neurons of *C. elegans*, however, tuning is a property of the cell.

We can visualize the tuning of each chemosensory neuron to olfactory stimuli by plotting the significant responses ($q \leq 0.01$) of each neuron in the odor space we constructed in **Figure 4.1A**.
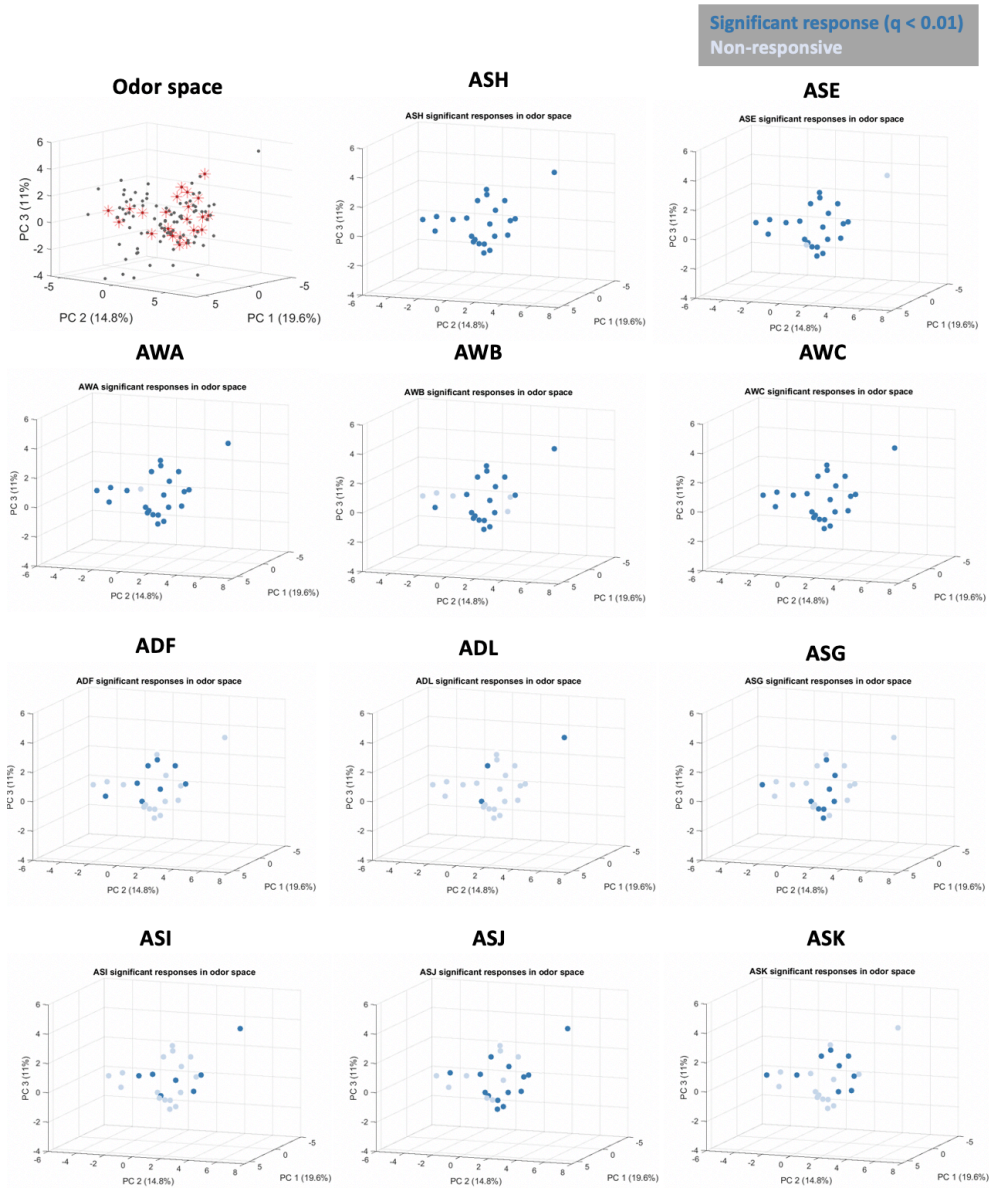
Figure 4.7: Chemosensory neuron tuning. Plotting the significant responses ($q \leq 0.01$) of each neuron in odor space (**Figure 4.1A**), we see that some neurons are activated by nearly the entire odorant panel (broadly tuned), while other neurons are activated by only small numbers of odorants (narrowly tuned). ASH, a known polymodal nociceptor, is broadly tuned. AWA, AWB, AWC, and ASE are also broadly tuned in their responses to odorants. Neurons ADF, ADL, ASG, ASI, ASJ, and ASK are more narrowly tuned, and are activated by distinct subsets of odorants. Refer to **Figure C.9** to visualize peak responses in odor space. Refer to **Figure C.4** for the odor space PC loadings.

We see that some neurons are activated by nearly the entire odorant panel (broadly tuned), while other neurons are activated by only small numbers of odorants (narrowly tuned) (**Figure 4.7**). ASH, a known polymodal nociceptor, is broadly tuned, being activated by all odorants in the panel at high concentration. This is unsurprising because at sufficiently high concentrations, all odorants should be behaviorally aversive. Since ASH is part of the reflexive avoidance circuit, we expect that it is not tuned to a particular part of odor space.

AWA, AWB, AWC, and ASE are also broadly tuned in their responses to odorants, each responding to nearly the entire odor panel at high concentrations. AWC responses are fairly uniform across the odorant panel. However, each of the neurons AWA, AWB, and ASE is most strongly activated by different groups of odorants. We can see this by projecting the average peak responses of each neuron into odor space (**Figure C.9**). AWA appears to be activated most strongly by ketones, AWB is activated more strongly by many esters, and ASE is most strongly activated by alcohols.

Neurons ADF, ADL, ASG, ASI, ASJ, and ASK appear to be more narrowly tuned, activated by small sets of odorants even at high concentration. These subsets are distinct for each of these narrowly tuned neurons. For instance, ASI responds significantly to 7 of 23 odorants in our panel, and ADL responds significantly to just 3 of 23. The sensitivity of each of these narrowly tuned neurons in odor space appears to be contiguous—the set of odorants it responds to contains chemically similar compounds. This suggests that these chemosensory neurons have been tuned to molecular properties, rather than specific molecules. Note that many of the neurons being classified here as narrowly tuned with respect to olfactory stimuli are known to play important roles in other forms of sensation. For instance, ASI is known to respond to tastants such as lysine and certain salts, ADF is

involved in oxygen sensing, and ADL is known to sense pheromones[79,81].

All sensory neurons synapse onto the interneuron group responsible for learned and goal-directed behaviors: AIA, AIB, AIY, and AIZ (**Figure 4.1C**). This includes both broadly tuned sensory neurons (AWA, AWB, AWC, and ASE), and narrowly tuned sensory neurons (ADF, ADL, ASG, ASI, ASJ, and ASK). These interneurons also receive an input from ASH onto AIA. The interneurons AIA, AIB, AIY, and AIZ synthesize the information from these sensory channels and inform behavioral decisions. Each set of sensory neurons could provide different information to these interneurons. The ASH signal indicates to the animal whether a stimulus is attractive or aversive. The broadly tuned sensory neurons could provide coarse-grained information about odor identity, identifying it perhaps as an alcohol or an ester. The narrowly tuned sensory neurons provide more detailed odor identity information by responding to small parts of odor space. Collectively, the broadly and narrowly tuned neurons generate the spatial map that allows odorants to be precisely and robustly identified.

Across the chemosensory neurons, tuning to olfactory stimuli (defined as the fraction of the odor panel which elicit significant responses) does not appear to be correlated with the number of GPCRs expressed[80] (**Figure C.10**). ADL the sensory neuron, which expresses the most chemoreceptor GPCRs, is sensitive to just 3 odorants in our panel. ASH expresses a large number of GPCRs and is also broadly tuned to our odorant panel, but ASK and ASJ, both with large numbers of GPCR types expressed, are both narrowly tuned. ASE, which appears to be broadly tuned in our study, expresses the smallest number of GPCRs. We also do not know how many of these GPCRs are used in olfaction. For instance, ADL is known to sense many pheremones, which may require

75

additional receptors. Additionally, tuning is a combination of receptor number and receptor properties. until we characterize more of these receptors, we cannot draw conclusions about tuning from the number of GPCRs expressed in each neuron.

4.2.8 Sensory representations are not dependent on synaptic connections

From the *C. elegans* connectome, we know that the chemosensory neurons are not just feed-forward. There exist lateral connections between sensory neurons, such as ASI to AWC and AWB to ADF[8,9]. There are also feedback connections from interneurons to some sensory neurons, such as AIA onto AWC. To assess the extent to which the neural representations of these odorants are shaped by interactions between neurons, we carried out a number of perturbative experiments.

We crossed the sensory neuron labeled line (ZM10104) with an *unc-13(s69)* synaptic transmission mutant[146], and presented these mutants with some of the odorants from our panel. For isoamyl alcohol and diacetyl, the same neurons significantly responded ($q \leq 0.01$) in the WT and mutant animals (**Figure 4.8A**). This indicates that, at least for these two odorants, synaptic connections do not shape the spatial map which encodes odorant identity. We do note, however, that the magnitudes of the average peak responses are generally lower in the mutants.

We also ablated the ASI neuron pair in the ZM10104 line, and presented the ablated animals with high-concentration 2-butanone. ASI is connected to AWC in the wiring diagram (**Figure 4.1C**), and both ASI and AWC are significantly activated by 2-butanone (**Figure 4.2B**). We found no difference in the set of significantly activated neurons between un-ablated (wt) and ablated animals, save the absence of ASI (**Figure 4.8A**). This indicates that ASI does not modulate the activity
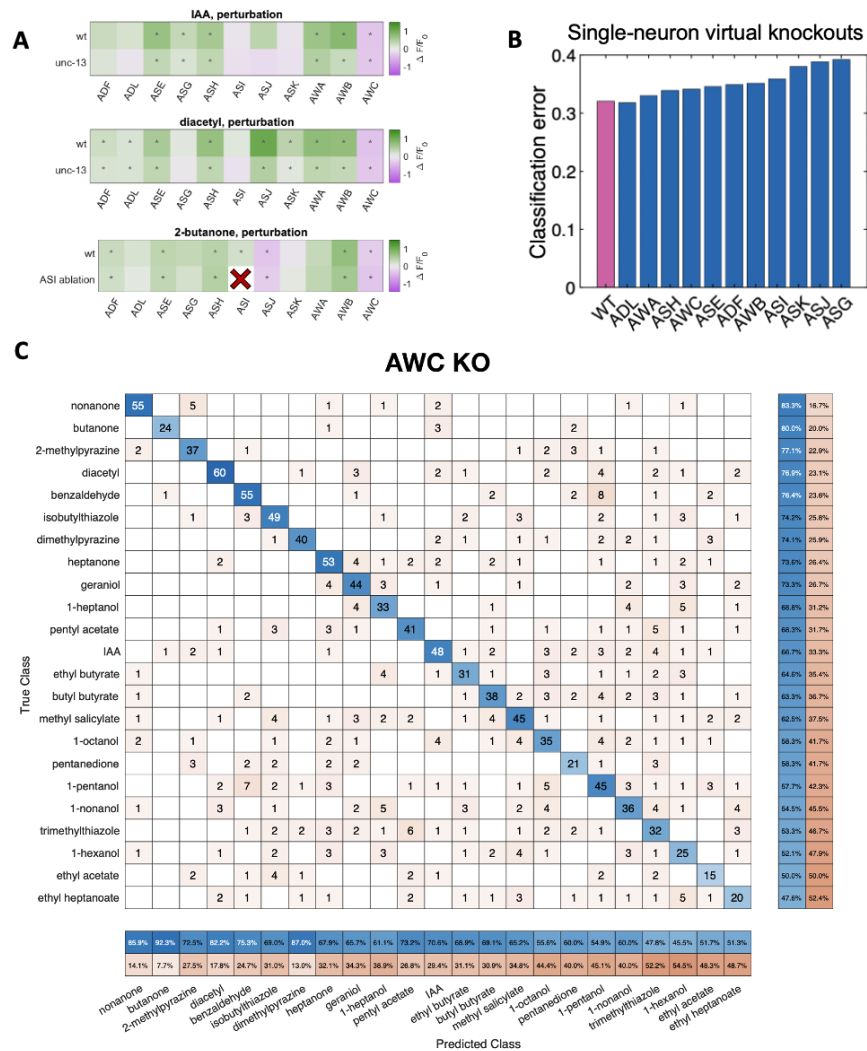
Figure 4.8: Impact of perturbations on odor representation. (A) We combined the sensory neuron imaging line (ZM10104) with an *unc-13(s69)* synaptic transmission mutant, and recorded neural activity during odor presentation. We found that for isoamyl alcohol and diacetyl, the same set of neurons were significantly responding ($q \leq 0.01$). This suggests that the ensemble responses are not being shaped primarily by synaptic connections. **(B)** We virtually removed each neuron from the dataset, and computed the average classification error for each virtual knockout. We find that classification accuracy remains close to wild type (all 11 neurons), but is degraded more severely by the removal of narrowly tuned neurons (ASI, ASK, ASJ, ASG) than by the removal of broadly tuned neurons. **(C)** Linear discriminability analysis of single-trial data with AWC virtually removed.

of AWC or any of the other chemosensory neurons during 2-butanone presentation.

Together, the results from ablations and the *unc-13* mutant suggest that the ensemble responses are not being shaped primarily by synaptic connections. A previous study suggested that neuropeptide signalling recruited the neuron ASE in response to benzaldehyde[89]. Our mutant experiments do not rule out the possibility that other extra-synaptic signals between the sensory neurons may help to shape the responses of the chemosensory ensemble to specific odorants.

4.2.9   Virtual neuron knockouts degrade classification accuracy

To understand the relative contribution of each chemosensory neuron to odor discriminability (**Section 4.2.5**), we performed virtual knockouts—removing each neuron class from the dataset and retraining the multi-class classifier on the remaining data. In these *in silico* perturbations, we find a minor decrease in classification accuracy when compared to wild type (**Figure 4.8B**). We note that classification accuracy is degraded more severely when narrowly tuned neurons such as ASI, ASK, ASJ, and ASG were removed from the dataset, than by the removal of broadly tuned neurons such as AWA, ASH, and AWC. Regardless of the neuron removed, no single-neuron virtual ablation severely degrades the performance of the multi-class classifier (**Figure 4.8C**).

4.3 Methods

4.3.1 Plasmids and crosses

To construct the ZM10104 imaging strain we created and then crossed two integrated lines, one expressing GCaMP and one expressing the wCherry landmark. The first of these lines, ADS700, was made by co-injecting *lin-15(n765)* animals with pJH4039 (*ift-20* GCaMP6s::3xNLS) and a *lin-15* rescuing plasmid. A stable transgenic line (hpEx3942) with consistent GCaMP expression in the chemosensory neurons was selected for integration, and transgenic animals were irradiated with UV light to integrate the transgenes into the genome. The resulting integrated line (aeaIs008) was then backcrossed four times against N2 wild type. The second line, ADS701, was similarly made by co-injecting *lin-15(n765)* animals with pJH4040 (*gpc-1* wCherry) and a *lin-15* rescuing plasmid. A stable transgenic line with good wCherry expression was selected for integration, and transgenic animals were irradiated with UV light to integrate the transgenes into the genome. The resulting integrated line (hpIs728) was then backcrossed four times against N2 wild type. To make ZM10104, ADS700 hermaphrodites were crossed with N2 males. Heterozygous aeaIs008/+ male progeny were then crossed with ADS701 hermaphrodites. F1 progeny were picked for wCherry expression, and F2 progeny were picked for both GCaMP6s and wCherry expression. The line was then homozygosed in the F3 generation.

The ADS707 mutant imaging line was created by crossing the ZM10104 line with EG9631, an *unc-13(s69)* mutant obtained from the CGC[146]. EG9631 hermaphrodites were crossed with

ZM10104 males. Heterozygous (aeaIs008/+; hpIs728/+; +/*unc-13*) F1 hermaphrodite progeny were selected by GCaMP6s and wCherry expression and wild type locomotion (*unc-13* is recessive). F2 progeny were picked for fluorescence and the *unc-13* uncoordinated phenotype. The line was then homozygosed for fluorescence in the F3 generation.

4.3.2 ODOR DELIVERY AND CONFOCAL MICROSCOPY

For these experiments, young adult *C. elegans* were loaded into microfluidics devices (Chapter 2). A single landmark image was captured in two channels (GCaMP6s, wCherry) for each animal. Then, for each odor condition (a given chemical at a given concentration), the animal was presented with six, 10 s pulses of odorant solution, separated with 40 s CTX buffer blanks. During odor presentation, the GCaMP6s channel was recorded at 2.5 Hz.

4.3.3 ANALYZING MULTI-NEURONAL RECORDINGS

The neurons in each activity recording were identified and then tracked through time using the neighborhood correlation method described in Chapter 3. The criteria for identifying each neuron class are described in Appendix C.1.2. Neurons which could not be unambiguiously identified were excluded from the dataset. All neuron tracks were then manually proofread to exclude mis-tracked neurons (Section 3.1.4). Activity traces were bleach corrected and reported in:

$$\frac{\Delta F}{F_0} = \frac{F(t) - F_0}{F_0}. \tag{4.1}$$

Normalization by baseline fluorescence F_0 allowed for direct comparisons within a given neuron class across L/R and across individuals. The baseline F_0 value was determined individually for every recorded neuron, set at the 5th percentile of the distribution of bleach-corrected fluorescence values, with the opportunity for manual correction.

We employed 2-tailed, paired t-tests to compare the mean signal during stimulus presentation with an unstimulated period of identical length within the same neuron. Neurons were tested for both ON and OFF responses. The p-values were corrected for multiple testing using FDR[147]. To test for asymmetric neuron responses, we used 2-tailed, two-sample t-tests (unpaired).

4.4 DISCUSSION

C. elegans uses a compact olfactory system consisting of just 11 pairs of chemosensory neurons to navigate its environment, find food, and avoid dangers. By recording from all of these neurons simultaneously while presenting animals with a broad panel of 23 odorants, we have been able to build a more complete picture of the activity of the chemosensory neuron ensemble in odor space.

We found that a sizeable fraction of the 11 chemosensory neurons responded to each olfactory stimulus, and that nearly every odorant in our panel reliably activated a unique combination of chemosensory neurons. This suggests that *C. elegans* may use the chemosensory neuron ensemble as a spatial map combinatorially code for both odorant identity and intensity. Our single-trial discriminability results show that despite the observed level of variability in neuron responses, sufficient information is encoded in this spatial map to correctly classify any odorant in our panel with over

80% accuracy from the responses of just a single odor pulse. Our mutant and ablation experiments suggest that synaptic communication and feedback do not play major roles in shaping the spatial map.

How do each of the 11 chemosensory neuron classes contribute to this spatial map? From the principal component analysis of average odorant responses, and from the *in silico* knockouts, we see that all chemosensory neurons contribute to odorant classification. However, some neurons are more important than others for this task. ASH detects aversive stimuli, but its activity does not seem to contribute much to odor identification. The neurons AWA, AWB, AWC, and ASE are active across most of our odorant panel, and their activity can provide very coarse odor identification. Their broad tuning suggests that their activity may be sufficient to drive chemotaxis behavior towards attractive stimuli or away from aversive stimuli. However, they do not contain enough information to robustly identify an odorant's identity. Narrowly tuned neurons such as ASK, ASJ, and ASG are each activated for only a small number of stimuli. From the perspective of odor identification, however, this is an advantage: if ASK is active, one could immediately narrow the number of possible odorants to a small fraction of the odor space. This is reflected in the virtual knockouts, where removing narrowly tuned neurons has a greater impact on classification accuracy than removing broadly tuned neurons.

An animal with only broadly tuned neurons intact would likely still be able to perform chemotaxis, but it may not be able to correctly classify odorants. An animal with only narrowly tuned neurons intact, however, would likely be unable to perform chemotaxis at all. This model is consistent with behavioral ablation experiments, which have shown that ablation of neurons such as ASK

result in minor degradation of chemotaxis ability, but ablation of neurons AWA and AWC destroy chemotaxis ability entirely[72].

The tuning properties of each of these chemosensory neurons in *C. elegans* is a property of the cell, not of a single receptor as it is in higher organisms. Evolution can tweak the number and types of receptors expressed in any given neuron, in addition to shaping the properties of each of the receptors themselves. Since the chemosensory neurons in *C. elegans* are thus theoretically more flexible, the diversity of dose responses we see is not unexpected. Despite this flexibility, we see in our data that many of the sensory neurons, including both broadly and narrowly tuned ones, seem to be sensitive to molecular properties (i.e.: particular chemical class(es)), rather than to a disparate set of molecules. As a field, we don't yet know enough about the chemosensory receptors in *C. elegans* to understand how multiple receptors combine to generate the tuning properties of the chemosensory neurons.

As mentioned previously, most of the analyses presented in this study were conducted with peak responses, a coarse-grained view of neural activity which leaves out many potentially important details, such as information encoded in dynamics, spikes, or asymmetries. The conclusions made in this study are thus based on a conservative lower bound on the information capacity of the chemosensory neuron ensemble. This study focused on pure odorants under conditions designed to avoid adaptation, but in its environment, most olfactory stimuli an adult *C. elegans* will experience likely come in the form of odor mixtures. How the animal represents the sensory input of naturalistic odor mixtures is an exciting question left for future work.

How does the information contained in these sensory representations get transformed into be-

havior? We have discussed how the downstream interneurons segregate into two clusters: a reflexive avoidance circuit consisting of the command interneurons AVA, AVB, and AVD which receives inputs predominantly from ASH; and a circuit for learned and goal-directed behaviors consisting of the interneurons AIA, AIB, AIY, and AIZ, which receives inputs from all of the chemosensory neurons. The former circuit is responsible for immediate action to avoid dangerous environments. As a nociceptor, makes sense for ASH to have a "direct line" to motor control, allowing the animal to initiate evasive action in response to a noxious stimulus. It is likely in the latter circuit that odorant identity and intensity are decoded by the animal. Previous studies have shown that these interneurons play a role in olfaction-driven behaviors, and may integrate information from the sensory neurons across multiple modalities[84,85,86,82]. It is in these interneurons that we may find processed odorant representations.

This study has uncovered reliable responses of the chemosensory neuron ensemble to a broad panel of olfactory stimuli, and has shown that these responses form a spatial map which contains sufficient information to robustly identify odorants. Whether, and how, the animal actually employs this information remains unknown. We hope future work can determine whether the downstream interneurons preserve odorant identity information, and how they transform these sensory representations to generate behavioral output in response to olfactory stimuli. ●

5

Whole-brain functional imaging with deterministic landmarking

Eviatar Yemini, Amin Nejatbakhsh, Erdem Varol, Ruoxi Sun, Gonzalo Mena, Liam Paninski, Vivek Venkatachalam, and Oliver Hobert contributed to the work described in this chapter.

DOWNSTREAM OF THE CHEMOSENSORY neuron ensemble, we find the interneurons. The *C. elegans* nervous system is topologically shallow, with only 1 or 2 interneuron layers separating most sensory neurons from motor neurons. The population of interneurons integrates chemosensory information with other inputs and generates behavioral decisions. We wanted to uncover how information encoded by the chemosensory neurons is transformed and represented in the downstream interneurons.

Only a small handful of single interneurons in *C. elegans* have been studied in the context of olfaction[84,85,86,82]. To build an understanding of odor representation among the interneuron population, we would like to be able to interrogate the entire population of interneurons simultaneously.

A pan-neuronal approach, in which all neurons in the animal are labeled with a calcium indicator such as GCaMP, would be ideal for such a study. A whole-brain labeled line would also give us access to the sensory and motor neuron activity. Previous pan-neuronal work in *C. elegans* has focused on whole-brain representations [133], but assigning cellular identities to neurons in whole-brain fluorescent images has proven to be a major challenge.

We aimed to span these two scales of inquiry—identifying reliable whole-brain responses to stimulus with single-neuron resolution. To do this, we had to develop a method of comprehensively identifying all of the neurons in the *C. elegans* brain. We achieved this by engineering a multicolor transgene called NeuroPAL (a Neuronal Polychromatic Atlas of Landmarks). NeuroPAL worms share a stereotypical multicolor fluorescence map of the hermaphrodite nervous system that allows comprehensive determination of neuronal identities. Neurons labeled with NeuroPAL do not exhibit fluorescence in the green, cyan, or yellow emission channels, allowing the transgene to be used with numerous reporters of gene expression or neuronal dynamics.

We employed NeuroPAL to capture brainwide activity in response to attractive and repulsive chemosensory cues, characterizing multimodal coding and novel neural asymmetries for stimuli. Comprehensive multicolor labels enabled us to average data across individuals, separating reliable stimulus-evoked responses from stimulus-independent activity. We found that chemosensory stimuli elicit widespread brainwide activity patterns, spanning sensory neurons, interneurons, and motor neurons. These patterns were unique to each stimulus, suggesting that odor identity information is preserved at the interneuron level. We also developed software which enables partially-automated determination of all neuronal identities based on color and positional information, and a

demixing method to separate signals from neighboring neurons.

5.1 INTRODUCTION

Whole-brain imaging and molecular profiling are widely used to study brain function, nervous-system development, and their molecular mechanisms [148,149,150,151]. One limitation in understanding whole-brain images is the difficulty of assigning unique identities to every neuron in a volume of densely packed and similarly labeled cells. Identifying neurons is a challenge even in small nervous systems like that of the nematode *C. elegans*. It is now possible in the worm to perform multi-neuronal functional imaging with single-cell resolution [133,152,117,153]. Multineuronal fluorescent reporters are also widely used to profile gene-expression patterns [154,155,156]. However, identifying neurons in either of these contexts is laborious and uncertain, requiring substantial expertise. The standard approach to identifying neurons is to cross-validate each expression pattern with a separate and sparsely-labeled landmark strain. Even so, many neurons lack well-established reporters, and it is not always possible to cross-validate every neuron of interest in a densely labeled volume even with a suitable landmark strain.

While the *C. elegans* nervous system is widely regarded as stereotyped, this stereotypy does not extend to the relative positions of cell bodies. An invariant color map of all neurons is thus needed to achieve comprehensive cell identification. Here, we leveraged the small size of the worm nervous system and its powerful genetics to develop a new method to identify all neurons in a whole-brain image with a single reagent. We describe the development of a transgene that we call NeuroPAL

(**Neuro**nal **P**olychromatic **A**tlas of **L**andmarks).

NeuroPAL is a combination of 41 selectively overlapping neuron-specific reporters, each of which expresses a subset of four distinctly-colored fluorophores. The NeuroPAL combination of reporters and colors provides a comprehensive color-coded atlas for the hermaphrodite nervous system. Our approach is fundamentally different from previously described "Brainbow" approaches [157,158,159]. In "Brainbow", multicolor labeling of the nervous system occurs when each neuron randomly expresses a subset of fluorophores. In NeuroPAL, each neuron expresses a stereotyped combination of fluorophores. NeuroPAL yields an invariant color map across individuals, where every neuron is uniquely identified by its color and position. We engineered NeuroPAL to be compatible with widely-used reporters for gene expression and neuronal activity. None of the NeuroPAL fluorophores emit in the spectral bands of green, cyan, or yellow fluorescent proteins. Thus, NeuroPAL can be co-expressed with numerous markers – GFP, CFP, YFP, mNeonGreen, or reporters of neuronal dynamics like GCaMP – without affecting its color map.

We measured the complete circuit-level response to a gustatory repellent (high concentration NaCl) and two olfactory attractants (2-butanone and 2,3-pentanedione). We observe stimulus-specific brainwide responses, and uncover novel asymmetric responses in sensory and interneurons. To facilitate the use of NeuroPAL, we provide an open-source software package that enables partially-automated identification of all neurons from whole-brain imaging data.

One of the major obstacles to adopting whole-brain imaging methods has been the difficulty of distinguishing neuronal identities. As demonstrated here, NeuroPAL now allows the *C. elegans* neuroscience community to annotate all neurons in whole-brain imaging datasets.

5.2 Results

5.2.1 Constructing the color palette for comprehensive landmarks

The *C. elegans* nervous system contains 302 neurons (organized into 118 different classes) distributed among 11 ganglia throughout the body[5,8]. The set of neurons in each ganglion are the same from animal to animal, but the relative location of cell bodies within each ganglion are variable. The largest ganglia contain around 30 neurons. We reasoned that roughly 30 unique colors would be needed to reliably identify all neurons in each ganglion, and thus all neurons in the nervous system. Three spectrally distinct fluorophores, distinguishable at four or more different levels (high, medium, low, and undetectable), yield at least 64 different colors. Thus, three carefully-chosen fluorophores should be enough to landmark the *C. elegans* nervous system (**Figure 5.1**).

We wanted our landmark reagent to be usable in animals that co-express transgenic reporters for gene expression or neuronal dynamics. The most popular fluorescent reporters include CFP, GFP/GCaMP, and YFP. We did not want the landmark fluorophores to contaminate emission signals from any of these reporters or vice-versa. Therefore, we sought fluorophores with unique excitation/emission profiles that also left free the cyan, green, and yellow emission bands. We tested a wide variety of fluorophores and found mTagBFP2, CyOFP1, mNeptune2.5, and TagRFP-T to be the best available candidates (**Figure 5.1A; Figure D.1**)[160,161,162,163]. By pseudo-coloring these fluorophores blue, green, red, and white respectively, their combinations generate RGB pseudo-colors (**Figure 5.1B-C**).
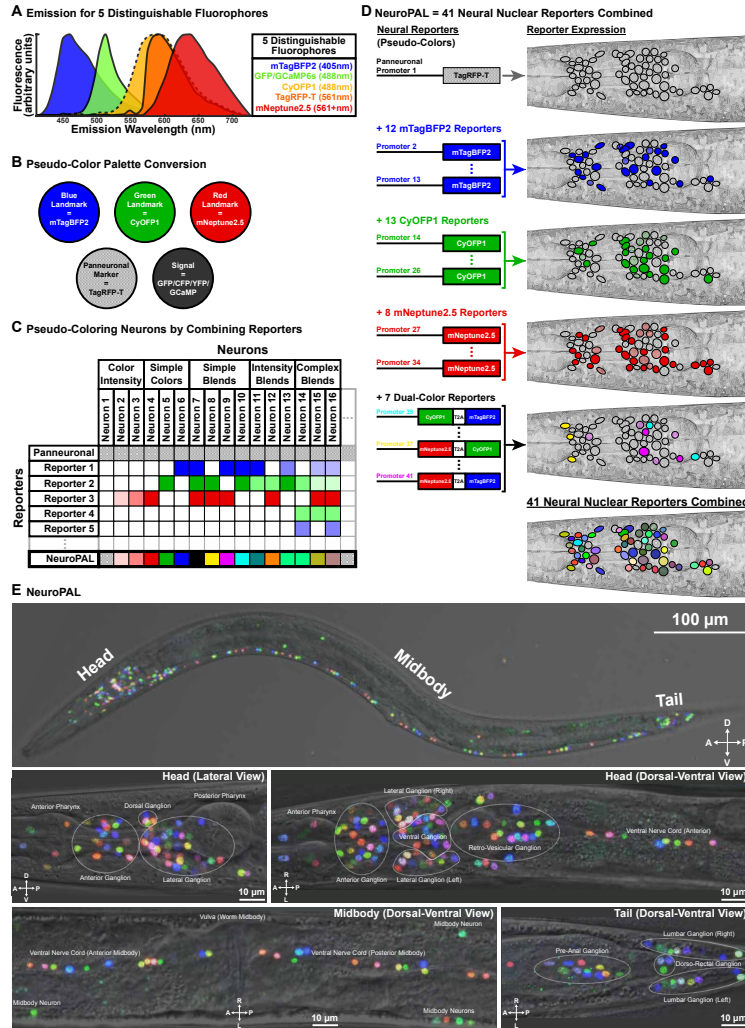
Figure 5.1: NeuroPAL methodology and images. (A) The emission for five distinguishable fluorophores. (B) Fluorophores are converted into pseudo colors to construct a primary color palette. Three fluorophores are designated as landmarks and pseudo colored to construct an RGB color palette: mNeptune2.5 is pseudo-colored red, CyOFP1 is pseudo-colored green, and mTagBFP2 is pseudo-colored blue. The fluorophore TagRFP-T is used as a panneuronal marker. The fluorophores GFP/CFP/YFP/GCaMP6s are reserved for reporters of gene expression or neural activity. TagRFP-T and GFP/CFP/YFP/GCaMP6s are visualized separately from the RGB landmarks to avoid confusion. (C) An example of how to stably pseudo color neurons. A set of reporters (rows), with stable neuronal expression (columns), are used to drive the fluorophores (table elements). NeuroPAL colors (last row) result from the combined patterns of reporter-fluorophore expression. (D) NeuroPAL scales this concept to 41 reporters that, in combination, disambiguate every neuron in *C. elegans* and, thus, generate, a single stereotyped color map across all animals). Seven of the NeuroPAL reporters use a self-cleaving peptide sequence (T2A) to simultaneously drive expression of two different colors. (E) Young adult NeuroPAL worms (*otIs669*) have a deterministic color map that remains identical across all animals. Each neuron is distinguishable from its neighbors via color.

90

Given the resolution limitations of light microscopy, we did not want fluorescence signals from neighboring cells to contaminate one another. To minimize spatial overlap in fluorescence emission, we localized fluorophore expression to cell nuclei via nuclear localization sequences (NLS) or histone H2B-tagging. We assigned one fluorophore, TagRFP-T, to act as a panneuronal label. To minimize variability in TagRFP-T expression, we constructed a synthetic ultra-panneuronal driver (UPN) by fusing the *cis*-regulatory elements of four different panneuronally expressed genes (**Table D.3.1**)[164]. The UPN delivered expression of TagRFP-T in 300 neurons of the nervous system but not the CAN cells.

5.2.2 Empirical assembly of the NeuroPAL transgene

Next, we sought to differentially express the remaining three fluorophores – mTagBFP2, CyOFP1, and mNeptune2.5 – to enable unique cellular identification (**Figure 5.1C-D**). We wanted each neuron to express a stable amount of each fluorophore, across animals. Furthermore, we wanted nearby cells in each ganglion to express visually distinguishable amounts of the fluorophores. These are the criteria for building a stereotyped cellular color map that permits unambiguous and comprehensive assignment of identity.

We began with a candidate list of 133 published neuronal reporters known to have differential gene expression patterns[165] (**Table D.3.1**). This candidate list was chosen to include both broad and narrowly expressed reporters. We partly verified reporter expression of candidates against a neural "checkerboard" expression pattern. This checkerboard pattern was constructed from four high-fidelity reporters with broad but heterogeneous expression driving green fluorescence (*acr-2::T-*

Sapphire; glr-4::T-Sapphire;lad-2::T-Sapphire;ocr-2::T-Sapphire;UPN::GFP. We evaluated the list of candidate reporters in groups of three by assigning each to a different landmark fluorophore and co-expressing them in the presence of the checkerboard pattern. The heterogeneous green fluorescence of the checkerboard helped verify the predicted expression of each candidate reporter. If a candidate reporter drove variable or weak expression, it was dropped from consideration. Thus we were left with a verified list of 101 neuronal reporters.

From our verified list of neuronal reporters (**Table D.3.1**), we proceeded empirically and itera- tively to build a transgene for comprehensive neuronal identification. We started with a small set of broadly expressed reporters that spanned most of the nervous system. We gradually co-expressed larger sets of reporters driving landmark fluorophores. In each iteration, we assessed which neurons could and could not be identified based on color and position, evaluating all transgenic siblings. To remove ambiguities, we expressed additional reporter-fluorophore constructs and re-assessed the results. We repeated these steps by trial and error until we found a suitable transgene that contained 41 different reporter-fluorophore fusions (**Table D.3.1**). The final transgene allowed us to disam- biguate every neuron in *C. elegans* based on its stereotyped color map (**Figure 5.1E; NeuroPAL Manuals: `https://www.hobertlab.org/neuropal/`**). We called this transgene NeuroPAL.

5.2.3 NEURONAL COLOR VERIFICATION AND PHENOTYPIC ASSESSMENT OF NEUROPAL STRAINS

We integrated the extrachromosomal NeuroPAL transgene into the genome of *C. elegans* using standard methods [166]. The brightest integrants (*otIs669*, *otIs670*, and *otIs696*) were outcrossed 8x

and exhibited stable expression for more than 100 generations. The color scheme of the NeuroPAL strains matched our expectations based on the combination of reporter-fluorophore fusions used in their construction. We verified the identity of each neuron by crossing the NeuroPAL integrants to 25 different GFP reporter lines with well-defined expression patterns (**Table D.3.1**). The position, color, and identity of all neurons were verified using predominantly two or more GFP reporter lines and multiple NeuroPAL integrants. We found that the NeuroPAL expression pattern was stable, robust, and stereotyped throughout the nervous system over hundreds of scored animals. A minor exception were four neuron classes that exhibited variable brightness (AVL, RIM, RIS, and PVW). This minor variability did not affect our ability to comprehensively identify all neurons. The stereotyped NeuroPAL color map permitted us to comprehensively identify all neurons at the L2 larval stage (**Figure D.2**) and developmental stages thereafter.

We assessed the general health of our NeuroPAL integrants (**Figure D.3**). All NeuroPAL integrants were able to generate progeny from either hermaphrodite or male parents. Thus, every integrant can be combined with other transgenic reporter lines using genetic crosses. We tested all NeuroPAL integrants with standard assays including brood size, growth, morphology, locomotion, and chemotaxis. The brightest integrant was *otIs669*. The integrant with locomotion and chemotactic behavior closest to wild-type, *otIs670*, is less bright but perhaps more suitable for behavioral analysis and calcium imaging. All NeuroPAL integrants are available at the *Caenorhabditis* Genetics Center (CGC).

5.2.4 Variability in neuronal cell body positions

The hermaphrodite nervous system is widely regarded as stereotyped. However, variability in the position of individual cell bodies makes it impossible to assign cell identities based on relative position alone (**Figure 2B**), underscoring the need for a reagent like NeuroPAL that disambiguates these identities based on genetic expression factors. Nevertheless, a probabilistic map of neuronal positions would be useful in many studies. To construct this probabilistic map, we globally aligned neurons in the head and tail of 10 young-adult NeuroPAL hermaphrodites (*otIs669*) of identical age, and measured the spatial coordinates of every neuron. This map revealed different amounts of variability in the positions of cell bodies across neuron types (**Figure 5.2C-D**).

5.2.5 Whole-brain activity imaging of gustatory and olfactory responses

A major challenge in analyzing pan-neuronal calcium imaging data in *C. elegans* has been determining neuronal identities using only their activity and positions[133,117,153]. To solve this problem, we combined NeuroPAL with the panneuronally-expressed calcium reporter GCaMP6s (strain OH16230). We then used multicolor imaging to comprehensively identify all neurons while recording their activity. We studied brainwide responses, in young-adult hermaphrodites, to a repulsive taste (160 mM NaCl) and two attractive odors (10^{-4} 2-butanone and 10^{-4} 2,3-pentanedione). These stimuli were delivered in chemotaxis buffer to the nose of the animal using a multichannel microfluidic device (**Figure 5.3A-B**)[55]. NaCl is primarily sensed by the left and right ASE neurons (ASEL and ASER), but with a stereotyped left/right asymmetry across animals[93]. ASEL and ASER
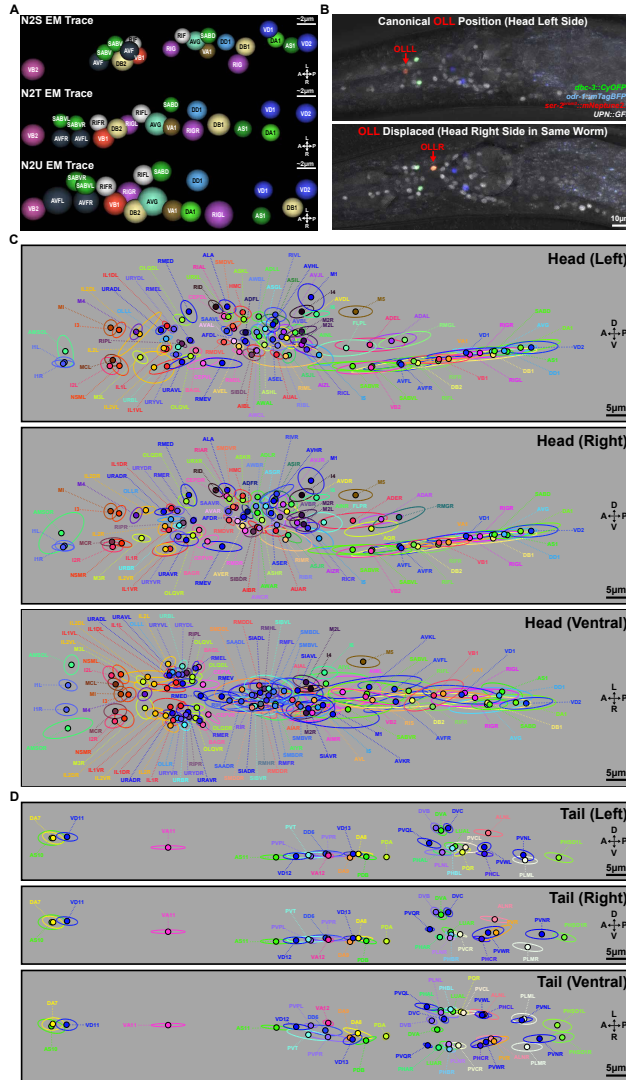
Figure 5.2: Neuron locations and their positional variability. (A) Neuron locations and variability, in the retrovesicular ganglion, taken from electron micrographs of three adult hermaphrodites N2S, N2T, and N2U [8]. (B) An example of substantial positional variability. The OLL left (OLLL) and right (OLLR) neurons, within a single animal, should share equivalent positions. Instead they show substantial anterior-posterior displacement relative to each other. The transgenic reporters and their pseudo colors are noted on the figure. (C,D) Canonical neuron locations (filled circles displaying the NeuroPAL colors) and their positional variability (encircling ellipses with matching colors) for all ganglia, as determined by NeuroPAL (otIs669). Positional variability is shown as the 50% contour for neuronal location (measured as a Gaussian density distribution), sliced within a 2D plane (**Text S1**). We show both the left-right and dorsal-ventral planes to provide a 3D estimation of positional variability. (C) Left, right, and ventral views of the head neuron positions. (D) Left, right, and ventral views of the tail neuron positions.

respond differently to upsteps and downsteps in salt concentration depending on step size[92]. 2-butanone and 2,3-pentanedione are sensed by the left and right AWC neurons (AWCL and AWCR), but with a stochastic asymmetry. In each worm, one neuron (either AWCL or AWCR) will adopt the identity AWC^{ON} and become more sensitive to 2-butanone while its lateral pair will adopt the identity AWC^{OFF} and become more sensitive to 2,3-pentanedione[94]. NeuroPAL differentially colors AWC^{ON} and AWC^{OFF}.

We subjected each animal (N=21) to the three chemical stimuli delivered in a randomized order (10s pulses spaced by 50s intervals). Using activity traces from all identified neurons, we assembled the mean brainwide response to each stimulus. Brainwide imaging revealed both known and novel neuronal responses to each stimulus (**Figure 5.3C**). As expected, upsteps in NaCl evoked stereotyped and distinct responses in the two ASE neurons. A few other sensory neurons have been implicated in detecting changes in NaCl concentration including ADF, ASG, ASH, ASI, ASJ[167,92,168]. We verified these stimulus-evoked responses (**Figure 5.3D**).

We uncovered novel responses in a surprisingly large number of sensory neurons and interneurons with strong correlations to the NaCl stimulus pulse (**Figure 5.3C-F**). Several neurons, such as ASH, were most strongly activated upon NaCl removal. Strikingly, we observed novel asymmetries in several other neuron responses to NaCl. AWA exhibited stereotyped L/R activity: the AWAL response was significantly larger than AWAR. AIY exhibited a stochastic correlation with the AWC ON/OFF identities: the AIY neuron on the same side as AWC^{OFF} showed a significantly larger response than its corresponding pair on the same side as AWC^{ON}.

Odors also evoked responses in a surprisingly large set of neurons throughout the brain (**Figure**
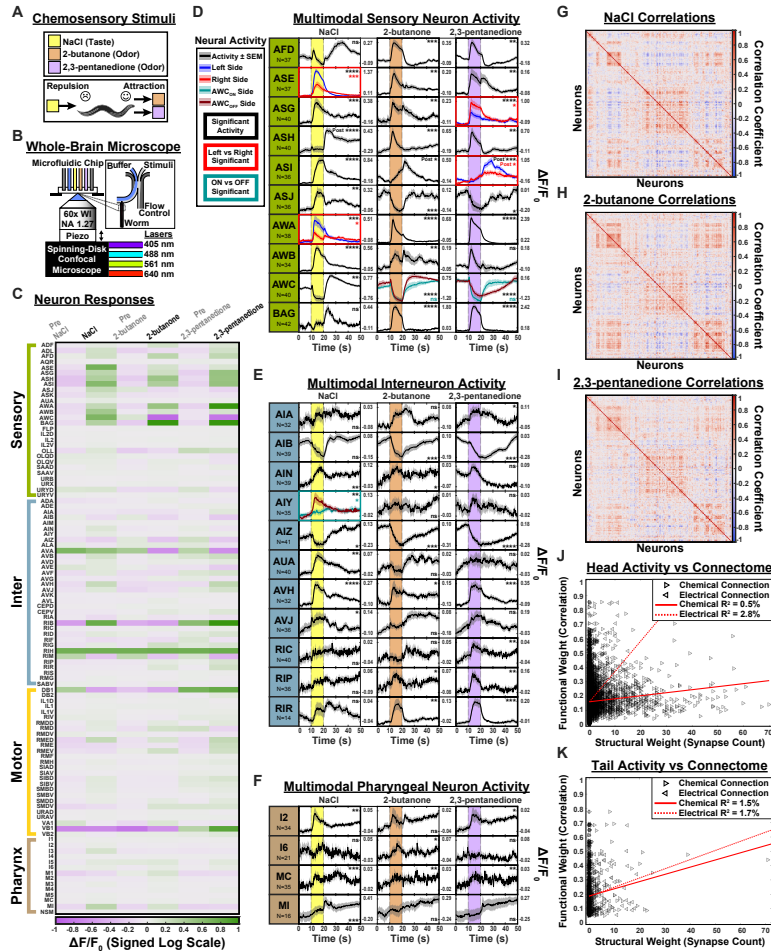
Figure 5.3: Whole-brain neuronal activity imaging of taste and odor responses. (A) *C. elegans* were subjected to three chemosensory stimuli: a repulsive taste (160 mM NaCl) and two attractive odors (10^{-4} 2-butanone and 10^{-4} 2,3-pentanedione). (B) Animals were immobilized inside a microfluidic chip. Stimuli were delivered in buffer. Each animal was imaged using a spinning disk confocal microscope with four excitation lasers. The NeuroPAL color map was imaged to identify all neurons. Thereafter, brainwide activity was recorded via the panneuronal calcium sensor GCaMP6s. (C) Peak neuronal activity, before and during stimulus presentation, for 109 head neuron classes. (D-F) Neuronal activity traces for selected (D) sensory neurons, (E) interneurons, and (F) pharyngeal neurons that responded to stimuli. The 10-second stimulus delivery period is indicated by the colored bar. Significant responses (p or q ≤ 0.05) are highlighted in bold boxes. Asymmetric left/right and AWC$^{ON/OFF}$-sided responses are indicated by colored boxes. "Post" = significant post-stimulus response. "ns" = no significant response. (F-I) Average pairwise correlations between 189 neurons for the 30 seconds following onset of (G) NaCl, (H) 2-butanone, and (I) 2,3-pentanedione. All three correlation maps are presented on the same axes, determined by clustering the full-time-course correlations. The sets of correlated and anti-correlated neurons differs for each stimulus presentation. (J-K) Comparison of functional activity to the connectome, we observe minimal correlation between synapse count and functional correlation between neuron pairs in the (J) head and (K) tail.

97

5.3C). Average chemosensory neuron responses to 2-butanone and 2,3-pentanedione were consistent with those observed in Chapter 2. As expected, 2-butanone and 2,3-pentanedione evoked stochastically asymmetric responses in AWC (**Figure 5.3D**). With 2,3-pentanedione, we also observed stereotyped L/R asymmetries in the ASG and ASI sensory neurons. The ASGL response was significantly smaller than ASGR. Upon stimulus removal, the ASIL response was significantly larger than ASIR. The set of neurons activated by all three stimuli was partly overlapping but distinct for each stimulus (**Figure 5.3C-F**). For example, ASJ was excited by NaCl but inhibited by both odors, whereas RIC was excited solely by 2,3-pentanedione.

Both salt and odors elicited responses across the synaptically-isolated pharyngeal nervous system, evoking significant activity in its sensory neurons, interneurons, and motor neurons (**Figure 5.3F**). The *C. elegans* pharynx employs a small and shallow network of only 20 neurons that is almost entirely separate from the remaining main nervous system. This small pharyngeal network synaptically-connects solely through its I1 interneurons and M1 motor neurons which communicate with the RIP interneurons of the main nervous system. Our results suggest that, despite its small size, the worm's pharyngeal network may encode its own representation of behavioral responses to chemosensory cues.

5.2.6 WHOLE-BRAIN NEURONAL DYNAMICS AND CONNECTIVITY

Exploring network-level dynamics, we computed pairwise correlations between the activity of all identified neurons from our whole-brain activity recordings. We found similar stimulus-specific neuronal correlations across individuals, but each stimulus generated its own correlation pattern

(**Figure 5.3G-I**). Even the two attractive odors produced distinct sets of neuronal correlations among and between sensory and interneurons. These results can be seen in brainwide neuronal trajectories through low-dimensional PCA-space (**Figure D.5C-E**). For instance, the AWB sensory neuron and its synaptically-connected interneuron partners AUA and AVH[8,37] exhibit distinct pairwise correlations depending on the stimulus (**Figure 5.3D-E**). Thus we find that brain dynamics are stimulus-specific.

While stimulus-evoked brainwide activity was similar across animals and across trials, stimulus-independent activity was highly variable across animals (**Figure D.5A**). In the absence of stimulus, the animal can be in any one of a large number of global brain states. Presentation of a chemosensory stimulus pushes the animal into a specific, stimulus-evoked brain state.

We asked whether a simple relationship exists between these functional correlations and the synaptic counts previously measured from the anatomical connectome[8,37]. To do so, we compared our correlation matrices of pairwise functional activity to the connectome matrix of pairwise synaptic connectivity. These matrices represent functional and structural measures of neuronal communication, respectively. We found low Pearson correlation between these functional and structural matrices. For electrical connectivity, $R^2 =2.8\%$ in the head and $R^2 =1.7\%$ in the tail, and for chemical connectivity, $R^2 =0.5\%$ in the head and $R^2 =1.5\%$ in the tail (**Figure 5.3J-K**). We tried multiple variations in our calculations (e.g., using ranked correlation metrics, log-scaling synaptic counts, and limiting functional activity to only stimulus or non-stimulus delivery periods) but these did not noticeably improve correlation between functional activity and the structural connectome.

5.2.7 SEMI-AUTOMATED NEURAL IDENTIFICATION

We developed an instruction guide to help researchers use NeuroPAL (**NeuroPAL Manuals: `https://www.hobertlab.org/neuropal/`**). This guide covers a variety of NeuroPAL-compatible microscope configurations and provides instructions on how to identify all neurons using the NeuroPAL color map. However, the manual annotation of neurons is laborious and time-consuming. To speed annotation, we developed a software pipeline that partially automates this task.

Our software pipeline uses three unsupervised algorithmic steps to semi-automatically annotate neuronal identities in NeuroPAL images (**Figure 5.4A**). First, we filter out non-neuronal fluorescence. Second, we detect the color and position of each neuron. Third, we compute a probabilistic estimate of each neuron's identity using a statistical atlas of NeuroPAL colors and positions (see **Variability in neuronal cell body positions**). Lastly, a graphical user interface (GUI) permits manual review and error correction of all steps in our unsupervised pipeline.

We evaluated the unsupervised neuronal identification performance of our pipeline. To do so, we cross validated its performance on previously annotated images and found our accuracy varied across ganglia (**Figure 5.4B-C**). Automated accuracy was 86% for the head and 94% for the tail. Accuracy largely depended on cell density; for example, our pipeline achieved high accuracy for all tail ganglia but lower accuracy for the much denser ventral ganglion in the head. Our software incorporates manual annotation of low-probability neuronal identities (computed in the *identify* step) to improve the estimated identities of the remaining unlabeled neurons. Adding eight manual annotations, on average, brings the head accuracy above 90% (**Figure 5.4B**).
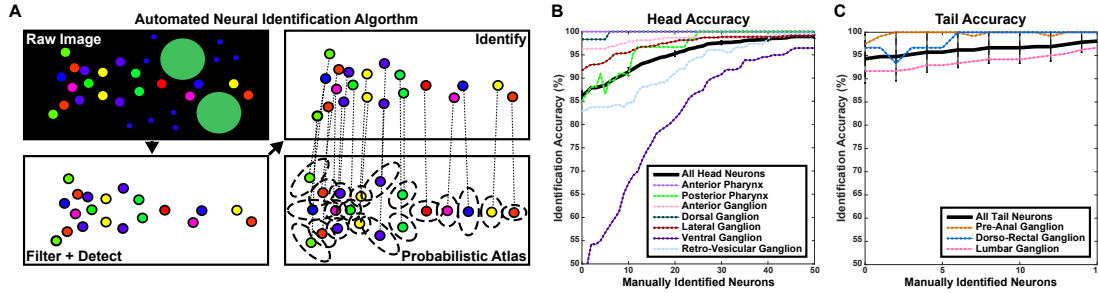
Figure 5.4: An algorithm for semi-automated neuronal identification and an algorithm to generate optimal-coloring solutions for cell identification (A-C) The algorithm used for automated neural identification. (A) Raw images are filtered to remove non-neuronal fluorescence and neurons are detected in the filtered image. Detected neurons are identified by matching them to a statistical atlas of neuronal colors and positions. (B,C) Automated neuronal identification accuracy begins at 86% for the head and 94% for the tail. Manually identifying eight neurons raises the head accuracy above 90%. Overall accuracy is displayed as a black line. Accuracy for each ganglion is displayed as a dotted, colored line (see legend). Many of the neurons and ganglia have high identification accuracy and confidence. The ventral ganglion is a problematic area, likely due to the high positional variance therein.

Our algorithm also provided a means of assessing the importance of color information in assigning cell identities. When we restricted the model to assign identities only on the basis of location, automated accuracy dropped to 50% for the head and 68% for the tail (**Text S1**). These results confirm a substantial improvement in accuracy with the color information provided by NeuroPAL.

5.3   Methods

5.3.1   Plasmids and injections

Fluorophores were ordered from IDT and/or cloned via standard techniques (Gibson, restriction-free, T4 ligase, or QuikChange™ mutagenesis) into the pPD95.62 Fire vector (a gift from Andrew Fire)[169,170]. The 1.4kb synthetic ultrapanneuronal driver (UPN) was generated by fusing *cis*-regulatory elements from four different panneuronally expressed genes: *unc-11^{prom8}*, *rgef-1^{prom2}*,

ehs-1^{prom7}, *ric-19^{prom6}* [164]. Fusion was done in a single quadruple PCR promoter fusion [171]. All cloned NeuroPAL reporters were made via PCR, gel purified, then inserted using standard techniques into the fluorophore vectors. To accommodate the large number of reporters and conserve space in extrachromosomal arrays, in place of plasmid backbones we used linear DNA amplified via PCR. Linear DNA has also been shown to improve expression levels [172]. Therefore, all injected NeuroPAL reporters were PCR amplified and gel purified to remove their vector backbone. Injection mixes consisted of complex arrays, with sheared bacterial DNA serving as spacers, to minimize potential crosstalk amongst reporters. Preliminary NeuroPAL strains were injected as complex arrays into *pha-1(e2123)* with *pBX[pha-1(+)]* to rescue at the selection temperature 25C [173]. Final NeuroPAL strains were injected into N2 without *pha-1(+)*. We used the *rab-3* reporter to drive panneuronal GCaMP6s expression. We noted that sensory neurons exhibited weaker GCaMP6s expression and thus supplemented the *rab-3* reporter with the *arrd-4* pansensory reporter. The panneuronal GCaMP reporters were injected as complex arrays into N2. Integrations were performed using gamma irradiation. All integrant worms were outcrossed 8x.

5.3.2 Fluorophore choice

The emission spectrum for published fluorophores can roughly be divided as blue, cyan, green, yellow, orange, red, and far red. GFP has a long emission tail; therefore, to avoid confusing its signal with NeuroPAL landmarks, we excluded cyan, green, yellow, and orange emitting proteins from our landmark choices. This left us with only three landmark emission channels: blue, red, and far red. The blue fluorophore EBFP2 is a GFP derivative and would preclude the ability to boost reporter

signal via cognate GFP staining[174]. mTagBFP2 is brighter alternative derived from TagRFP-T[160] and, for this reason, we chose it as our first NeuroPAL landmark. As a caveat, all red and far-red fluorescent proteins produce conformations with GFP-like excitation and emission and their mature fluorophores exhibit an equilibrium between this green and the preferred red emission states[163]. For this reason, we looked for bright red and far-red fluorophores with a high red-to-green emission ratio and considerable photostability. This yielded our next two landmark fluorophores: TagRFP-T and mNeptune2.5[163,162].

To expand the pool of candidates further we looked towards the large-stokes shift (LSS) fluorophores. LSS fluorescent proteins have a long gap between their excitation wavelengths and emission bands; for example, GFP is excited by 488nm and emits green while its LSS-derivative T-Sapphire is excited by 405nm and shares a nearly identical green-emission band as GFP[175]. In fact, wild-type (WT) GFP is a mixture of isoforms excited either by 405nm or 488nm. WT GFP was mutated to stabilize both isoforms independently, thus yielding T-Sapphire and enhanced GFP (EGFP – often abbreviated as simply GFP)[176]. Unfortunately, despite this mutagenesis, the commonly used variant of GFP retains considerable excitation at 405nm, ruling out both T-Sapphire and the yellow-emitting mAmetrine for use as landmarks[177]. This left only red and far-red LSS fluorophore candidates. Initially we tested mBeRFP[178] but found it was excited by a broad range of wavelengths, including those exciting TagRFP-T and mNeptune2.5. Furthermore, mTagBFP2 and GFP excitation wavelengths caused mBeRFP to produce a bright green emission, thus confounding our GFP signal channel. Consequently, we arrived at the newly-published CyOFP1 as one of our last available choices[161]. CyOFP1 has nearly identical emission to TagRFP-T but, its excitation wavelength

(488nm), fails to excite TagRFP-T; therefore, one can excite and visualize CyOFP1 independently of TagRFP-T, simply by using 488nm excitation. As a result of this brightness and photostability, we chose CyOFP1 as our fourth and final NeuroPAL landmark. Fortunately, mNeptune2.5's preferred excitation, 599nm, fails to excite either CyOFP1 or TagRFP-T. Unfortunately, TagRFP-T's preferred excitation wavelength, excites both CyOFP1 and mNeptune2.5. Thus three of our landmark fluorophores can be easily distinguished from each other and GFP but, the remaining one, TagRFP-T, suffers mild bleed through from both the CyOFP1 and mNeptune2.5 landmarks. For this reason we chose mTagBFP2, CyOFP1, and mNeptune2.5 to represent our primary color palette (**Figure D.1**). TagRFP-T was instead assigned to drive the panneuronal reporter. As a result, bleed through from the other two neural landmarks do not interfere with TagRFP-T's role in highlighting which cells are neurons. When testing our NeuroPAL landmarks we found that, despite claims of monomericity, mTagBFP2 and CyOFP1 showed aggregation when expressed at high concentration; therefore, we tamed their oligomerization by tethering them to the histone H2B.

5.3.3 Demixing multi-neuronal recordings of *C. elegans*

Even with nuclear localization, the neurons in the brain of the worm are very densely packed. Neighboring neurons are difficult to segment spatially, and the time-varying fluorescence of one neuron can bleed into the signal extracted from its neighbor.

To demix the signals of neighboring neurons, we developed a deformable non-negative matrix factorization (dNMF) method. We will be briefly summarizing the results of this work here, and in **Appendix D.2**. For a complete discussion of the dNMF method, refer to Nejatbakhsh et al.,

2020[119].

We tested the dNMF method on recordings of the tail of hermaphrodite *C. elegans* which expressed pan-neuronal nuclear-localized GCaMP. The worm's tail contains several ganglia, with densely-packed neurons, whose spatial footprints often overlap due to insufficient spatial resolution. Additionally, even neurons in separate ganglia can end up in sufficient proximity, due to microfluidic confinement or other imaging-setup induced deformations, such that their spatial footprints overlap. The spatial overlap represents a significant challenge, both for tracking individual neurons and demixing their signal. **Figure D.6** shows an example of the difficulty present when tracking and demixing neural activity signals from animals with spatially overlapping neural footprints in their recorded images. In this example, ROI tracking loses most of the signal from the LUAR and PLNR neurons and further mixes signals between the DA8/VD13, DVA/DVB, PHCR/PVWL, and PVNL/PVNR neurons. Normcorre+NMF performs better but loses nearly all signals from PLNR while also still mixing signals between the DVA/DVB, PHCR/PVWL, and PVNL/PVNR neurons. In comparison, dNMF recovers strong, independent signals from all ten neurons.

Open source code implementing this pipeline is available at `https://github.com/amin-nejat/dNMF`.

5.3.4 ANALYSIS OF WHOLE-BRAIN IMAGING DATA

To analyze the whole-brain imaging stimulus responses, we reviewed ASE responses to salt and AWC responses to the odors – the primary sensory neurons for these stimuli. Worms were marked as stimulus responsive if either their left or right neuron showed the published response to their corresponding stimuli (ASE excitation for NaCl and AWC inhibition for odors)[168,93]. 21 heads

responded to all three stimuli, providing strong internal controls to compare their circuit activity across all stimuli. Additionally, 21 tails were included, without response verification as the heads of these animals were not simultaneously imaged. Premotor interneurons (AVA, AVB, AVD, AVE, PVC) and motor neurons (AS, DA, DB, DD, VA, VB, VC, VD) show spontaneous cyclical activity and thus were excluded from significance testing. We used t-tests (2-tailed, paired) to compare the mean signal during stimulus presentation with an identical period immediately prior, within the very same neuron (a strong internal control). p-values were corrected for multiple testing using FDR[147]. Since GCaMP6s requires 488 nm excitation, we were unable to record a pre-stimulus period for blue-light responses. We reasoned that since worms habituate to light (as evidenced by our traces and previous reports[179]), the blue-light response could instead be tested in reverse by comparing the 10s immediately after lights on, to the 10s period thereafter. Our protocol of two minutes in the dark, just after the identification volume, ensures that the 488 nm GCaMP6s excitation laser is a sudden, strong stimulus, evoking an immediate aversive response. Light comparisons similarly employed a 2-tailed, paired t-test. To test asymmetric neuron responses we used 2-tailed, two-sample (unpaired) t-tests.

5.4 Discussion

Understanding the nervous system requires an integrated view of its underlying molecular, signaling, and functional networks. A primary bottleneck to mapping these networks across an entire brain has been the difficulty of reliably identifying neuronal types. Neuroscientists have long made

progress by focusing their attention on selected neurons in any given study. However, a means of comprehensive neural identification, coupled with existing brainwide imaging strategies, would permit unbiased and fully integrative approaches to entire brain networks.

We introduced NeuroPAL, a tool that allows researchers to use a single multicolor landmark strain to determine all neuron identities in the *C. elegans* nervous system. The fluorophores used to build NeuroPAL were chosen to have negligible emission overlap with CFP, YFP, and GFP/GCaMP, while maintaining compatibility with common illumination sources and filter sets. This allows NeuroPAL to be used in conjunction with many different types of fluorescent reporters for diverse studies of gene expression, nervous-system development, and brain activity.

To date, functional networks have been investigated by recording the activity of small subsets of labeled neurons. More recent work has allowed whole-brain activity imaging with cellular resolution [121,132,133,125]. However, the inability to reliably identify all neurons within whole-brain recordings has precluded a full picture with circuit-level details. Thus, principal component analysis (PCA) has been commonly employed to construct low-dimensional representations of brain dynamics in individual animals, but the lack of a common basis precluded animal-to-animal comparisons. Coupling NeuroPAL with whole-brain activity imaging methods, one can provide a unified view of network dynamics across animals without sacrificing circuit-level details.

Here, we used NeuroPAL for whole-brain imaging of panneuronal calcium dynamics in response to chemical stimuli. The three well-studied chemical stimuli that we used have been known to evoke activity in a small number of sensory neurons and interneurons. Our results show that the set of stimulus-evoked responses engage the nervous system far more broadly than previously realized,

extending across many sensory neurons and interneurons, even those in the synaptically-isolated pharyngeal nervous system. Surprisingly, we find that many neurons respond with distinct dynamics to the two olfactory attractants that we used (2-butanone and 2,3-pentanedione) and the one gustatory repellent (a high concentration of NaCl). Our data reveals new neural asymmetries in the nervous system, both deterministic left/right asymmetries and stochastic ON/OFF asymmetries, that also escaped previous analyses. Thus asymmetric activity in the worm nervous system is more widespread than in the selected and well-studied examples of the stereotyped ASEL/ASER pair and the stochastic AWC^{ON}/AWC^{OFF} pair.

Previous pan-neuronal imaging experiments in *C. elegans* concluded that in the absence of external stimulus, whole-brain activity could be represented as a low-dimensional limit cycle[133]. With no method of comprehensively identifying the neurons, however, it was not possible to compare activity across animals. Thus, representations of global brain activity were constructed from the data of one individual at a time, with qualitative similarities in representation observed across individuals. With comprehensive labeling of neurons, we were able to directly compare whole-brain activity across animals with single-neuron resolution. Global brain activity appears to be more complex than previously thought, and is highly variable across animals in the absence of stimulus. When the animal is exposed to stimulus, whole-brain activity is pushed towards a unique, stimulus-specific activity pattern.

Comprehensive neuronal identification has also enabled us to examine the relationship between whole-brain activity and the connectome. We compared the functional activity correlations in our data to the anatomical measures of synaptic weight, but found no strong correlations between them.

Unifying anatomical and functional views of the nervous system will require deeper understanding of the biological properties of synaptic communication and neuromodulation, as well as more sophisticated theoretical modeling approaches.

Ongoing work is extending the NeuroPAL colormap to other life stages of developing nematodes, the alternate dauer stage, and the male nervous system. NeuroPAL will accelerate ongoing work in understanding the development, molecular signaling, and brain-wide activity of the nematode nervous system. The rich set of network responses we uncovered has broad relevance in understanding sensorimotor processing. Even simple stimuli engage large portions of the worm nervous system, with different brain-wide neuronal correlations across different conditions. Generally, deterministic multicolor labeling strategies for cell-type identification in model organisms will enable new approaches in systems neuroscience. ●

6

Connectome-constrained modeling of whole-brain activity

Lu Mi, Richard Xu, Sridhama Prakhya, Helena Casademunt, and Srinivas Turaga contributed to the work described in this chapter.

IN OUR PAN-NEURONAL EXPERIMENTS, we observed reliable stimulus-evoked brainwide activity patterns. Many of these activated neurons were strongly correlated with each other. A common logical leap made by neuroscientists when circuit-cracking is as follows: two neurons with highly correlated activity are likely to be wired to each other. With comprehensively labeled whole-brain recordings, we could directly compare functional activity to the anotomical wiring diagram of *C. elegans*. The anatomical wiring diagram of C. elegans was first reconstructed via serial electron microscopy by John White and colleagues in 1986[8]. There are two types of anatomical connections between neurons: chemical synapses, where neurotransmitters may be exchanged, and gap junctions, which are direct electrical connections. Both types of connections were reported with some

measure of structural weight, a best guess as to the degree of contact between two neurons based on synapse number for chemical connections and contact patch area for electrical connections.

As described in the previous chapter, by plotting functional correlations against synapse count for all recorded neuron pairs, we can assess the extent to which the anatomical connectome explains the functional activity patterns we see. When we do these comparisons, we find that the r squared when comparing functional correlations to the structural weight of chemical connections is 0.005, and the r squared when repeating this comparison with electrical connections is 0.028 (Fig. 5.3J,K). In other words, neither the chemical connections nor the electrical connections in the connectome significantly explain any of the functional correlations we see in our data. Similarly low correlations have been observed in the tail of male *C. elegans* recorded during mating[136].

So, when it comes to functional correlations of neural activity, the connectome doesn't tell the whole story. Unifying these two views of the nervous system—anatomical and functional—is going to take more than a simple linear regression accounting for only first degree connections. There are many pertinent details which are left out in such a high-level comparison of function and structure. For instance, the properties of synaptic communication and neuromodulation in *C. elegans* are poorly understood. We don't know whether given chemical synapses are excitatory or inhibitory, or indeed whether the sign is context-dependent. Nor is either measure of structural weight (synapse count or contact patch area) likely a good proxy for synapse strength. There is also substantial evidence for extrasynaptic communication between neurons in *C. elegans*[11]. Finally, despite the stereotypy of *C. elegans* development, the connectome is not identical across individuals. The original connectome reconstructed by John White and colleagues was reconstructed from partially overlap-

ping segments from 4 individuals, which means that most synapses were reported with a sample size of $N = 2$ [8]. Recent work in which connectomes for 8 additional *C. elegans* were reconstructed has demonstrated significant variability in the adult connectome of *C. elegans*, with 43% of connections not conserved across individuals [9]. Even for strong, conserved connections, the number of synapses can vary. Since the available anatomical data comes from different individuals than those used for functional imaging, the synapse counts and connections in the available connectomes will only coarsely reflect the connectome of the individual being recorded.

With these caveats in mind, we can ask whether the functional activity we see in our pan-neuronal datasets reflect the known connectivity of the worm's nervous system. By employing the topology of the *C. elegans* wiring diagram and allowing the connection weights to be trained based on labeled functional activity data, we were able to build an anatomically-inspired neural network model capable of making reasonable predictions of neuron activity during chemosensation.

6.1 INTRODUCTION

Pan-neuronal functional activity measurements made at single-neuron resolution, when mapped onto the anatomical connectome, can guide the development of detailed computational models. However, from a network modeling perspective, these data are incomplete in several critical ways. Nuclear labels are incapable of reliably capturing the activity of some neurons, as nuclear localized GCaMP has slower dynamics and is less sensitive than cytoplasmically expressed GCaMP. Furthermore, neurons like RIA have been shown to have distinct calcium dynamics in different parts of the
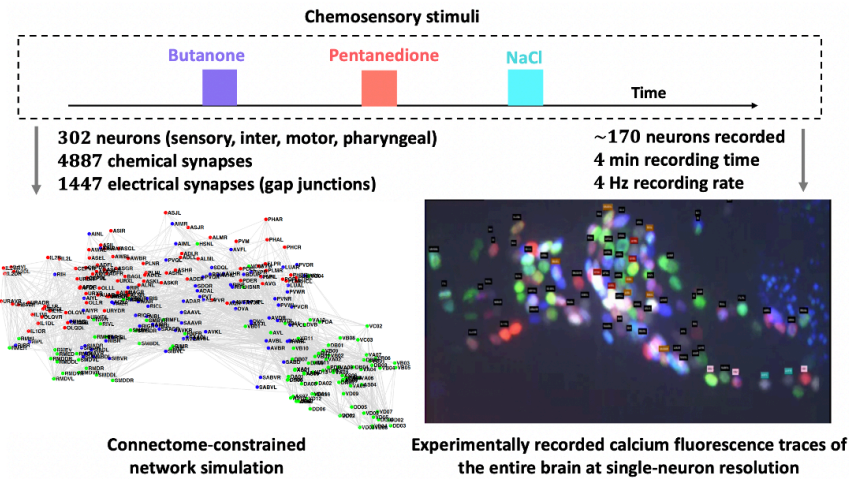
Figure 6.1: Anatomical and functional whole-brain data available in *C. elegans*. In the nematode *C. elegans* we have access to both anatomical (left) and functional (right) data for the entire brain at single-neuron resolution. The anatomical connectivity of *C. elegans* has been mapped, including both chemical and electrical connections between enruons[8,36]. The EM dataset from John White et al., 1986 was used in this work. Left panel adapted from Cook et al., 2019. Using the NeuroPAL animal (Chapter 5), we have functional activity data from the neurons in the head of the animal. Multi-color identification enables us to map functional activity onto the topology provided by the connectome[139].

cell, suggesting complex intracellular computations[25,26,27]. Despite the availability of maps of neurotransmmitter expression and neurotransmitter receptors in *C. elegans*[12,13,11], the signs of synapses are unknown. Further, the neural activity in many neurons is dominated by stimulus independent on-going background activity, which makes it challenging to fit deterministic nonlinear dynamical network models directly to calcium imaging data.

The anatomical considerations discussed previously suggest that while the existence of an edge in the connectome may be relied upon across individuals, the number of synapses is highly variable. This makes it difficult to predict neural function from structure because of the wide space of solutions that must be explored during optimization. Models that optimize functional performance are not necessarily unique.

To address these challenges, we built a connectome-constrained latent variable model (CC-LVM) of the neural activity dynamics. We use stochastic threshold linear dynamics to model the *C. elegans* nervous system, treating these dynamics as soft constraints on the neural activity. We model the underlying activity of each neuron as a latent variable, inspired by equations that model neuronal activity as voltage dynamics. We define the prior distributions for these latent variables by the calcium dynamics observed in the data. The chemical and electrical synaptic strengths in the model are learned, but the sparsity matrix is constrained by the connectivity given by the anatomical connectome: two nodes in the neural network can only interact directly if their corresponding neurons are connected in the *C. elegans* wiring diagram. The framework of variational auto-encoders (VAE) is used to train the model and infer underlying neuronal activity.

We applied the CC-LVM to our whole-brain calcium imaging data, which captured 170 of the 300 neurons in *C. elegans* as it responds to chemosensory stimuli. In principle, an accurate model of the nervous system constrained by incomplete activity measurements but complete connectivity measurements can enable accurate predictions of neural activity in neurons which were not recorded. We tested this hypothesis by using the CC-LVM to predict activities of neurons which were experimentally measured, but whose activity was withheld during model training.

We found that the CC-LVM predicted the activity of withheld neurons with greater accuracy than unconstrained neural networks of a similar size. This result demonstrates that, even without detailed biological information about the signs and strengths of individual connections, the topology of the connectome places measurable constraints on the brainwide activity patterns we experimentally observe.

114

6.2 Results

6.2.1 A connectome-constrained latent variable model

We built an artificial neural network which takes sensory information as input and generates predicted underlying activity and fluorescence traces for every neuron in the *C. elegans*. Each node corresponded to one of the neurons in *C. elegans*. To generate the inputs to this network, we first built a sensory encoder which generates a representation of the chemosensory stimuli experienced by the sensory neurons. The encoder is a deterministic, dense-layer network which takes as input the stimulus delivery sequence for the three compounds, $O_{\text{butanone}}(t)$, $O_{\text{pentanedione}}(t)$, $O_{\text{NaCl}}(t)$, and the fluorescence traces of the sensory neurons. This multilayer perceptron generates a representation of the chemosensory input $S_i(t)$ for all sensory neurons. Non-sensory neurons in the model do not receive this input ($S_n(t) = 0$).

To construct the underlying activity dynamics, we borrow from the electrophysiological equations which model neuron voltages. For each node in the neural network, the underlying activity was modeled as a stochastic non-spiking threshold linear leaky integrator with learned neuronal time constants and thresholds [180]. Each node of the network was coupled to other nodes by chemical synapses with learned signs and strength and electrical synapses with learned strengths. The existence of edges in the network were constrained by the *C. elegans* connectome annotated by John White and colleagues [8,36]: we allowed the trained weights $W_{ji}{}^{chem}$ and $W_{ji}{}^{elec}$ to be non-zero only where the connectome indicates the existence of an edge. Note that we only constrain the existence

of the edge by existence in the anatomical wiring diagram; we do not constrain the weights to be structural measures of weight, such as synapse count or contact patch area. The model first generates predicted underlying activity $V(t)$ for each neuron:

$$\tau_i \frac{dV_i(t)}{dt} + V_i(t) = \sum_{j=1}^{N} \left(W_{ji}{}^{chem}[V_j(t)]^+ + W_{ji}{}^{elec}[V_j(t) - V_i(t)] \right) + b_i + S_i(t) + \sigma_{V_i}\varepsilon_{V_i}(t),$$

(6.1)

$$\text{subject to } P[V(0), V(1), ..., V(T)|S(1), ..., S(T)] = P[V(0)] \prod_{t=1,...,T} P[V(t)|V(t-1), S(t)],$$

(6.2)

where $S(t)$ is the input, τ_i, $W_{ji}{}^{chem}$, $W_{ji}{}^{elec}$, b_i, and σ_{V_i} are trained parameters, and $V(t)$ is the output. The underlying activity is then converted into a fluorescence trace which can be directly compared to ground truth. Again, we employ an analogy to the mathematical relationship between voltage and calcium fluorescence:

$$\tau' \frac{dF_i(t)}{dt} + F_i(t) = \alpha_i \, softplus[V_i(t)] + \beta_i + \sigma_{F_i}\varepsilon_{F_i}(t),$$

(6.3)

$$\text{subject to } P[F(0), F(1), ..., F(T)|F(1), ..., F(T)] = P[F(0)] \prod_{t=1,...,T} P[F(t)|V(0), V(1), ..., V(t)],$$
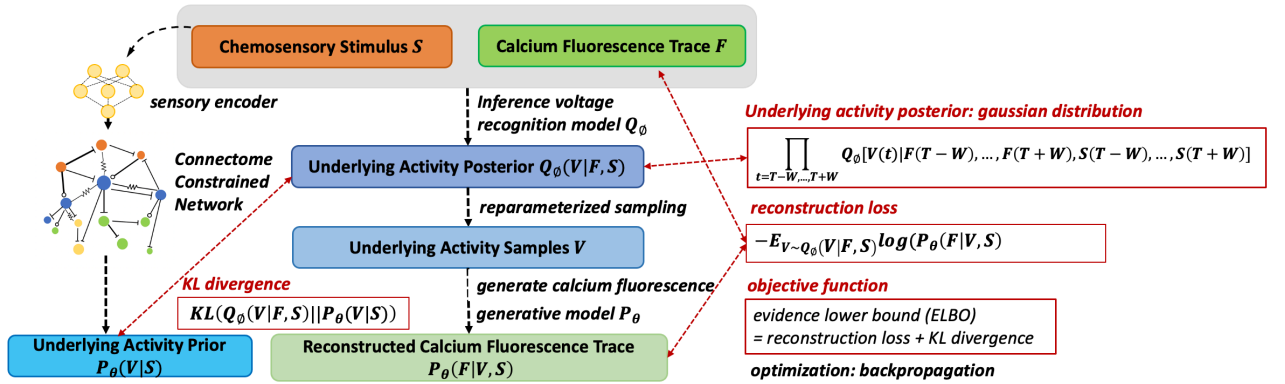
(6.4)

Figure 6.2: Schematic of the connectome-constrained model and variational auto-encoder. The connectome-constrained network (CC-LVM) consists of stochastic threshold linear integration of underlying activity dynamics. It takes as input the output of a deterministic sensory encoder which represents the chemosensory stimulus presented to the animals. The weights are constrained to be non-zero only for pairs of neurons where the connectome indicates the existence of synapses, and the electrical and chemical synapses, neuronal time-constants, and underlying activity-to-fluorescence parameters are optimized to fit the calcium imaging data. A variational auto-encoder (VAE) is used to fit the CC-LVM and to infer the posterior distribution over underlying neuron activity. The VAE is optimized with the ELBO loss through stochastic backpropagation.

where τ', α_i, β_i, and σ_{F_i} are trained, and $F(t)$ is the output. Together, the neural dynamics define a

prior distribution over likely underlying activity trajectories which are consistent with a given set of

weights, time constants, and thresholds (Equations 6.2 & 6.4). Importantly, the stochastic nature of

the dynamics, in the form of the neural activity noise terms $\varepsilon_{V_i}(t)$ and the imaging noise terms $\varepsilon_{F_i}(t)$,

allows for deviations from perfectly deterministic dynamics, allowing some leeway in our model

for variability in a single given neuron's activity and incomplete knowledge of the sensory inputs

driving the nervous system.

A latent variable model of this scale with a nonlinear generative model defined by the stochastic

dynamics is challenging to fit. To fit the model to the data, we first used the probabilistic inference

framework of a variational auto-encoder (VAE)[181] to train a black-box voltage inference network

($Q_\emptyset$) which generates a posterior distribution from the experimentally recorded calcium fluorescence traces. This convolutional inference network predicts a factorized Gaussian distribution over the underlying activity of all the neurons in the network, given sensory stimulus time courses, and calcium imaging measurements for a fraction of the neurons. The underlying activity posterior is then compared to the underlying activity prior generated by the connectome-constrained network, in order to train the parameters of the CC-LVM.

The entire VAE model was optimized with the ELBO (evidence lower bound) objective function L through stochastic backpropagation. The objective function includes both a reconstruction loss (first term) to maximize the data likelihood and a Kullback-Leibler (KL) divergence (second term) to minimize the distance between the prior and posterior distributions:

$$L = -E_{V \sim Q_\emptyset(V|F,S)} \log P_\theta(F|V,S) + KL(Q_\emptyset(V|F,S)||P_\theta(F|V,S)). \tag{6.5}$$

The resulting VAE thus has a biologically and biophysically plausible generative model of the neural dynamics of the *C. elegans* nervous system.

6.2.2 ACTIVITY PREDICTIONS ARE IMPROVED UNDER CONNECTOME CONSTRAINT

We applied the connectome-constrained latent variable model to the calcium imaging dataset collected for the NeuroPAL project [139]. This data captured the activity of a total of 170 neurons across 21 individuals in response to a panel of three chemosensory stimuli: NaCl, 2-butanone, and 2,3-pentanedione.
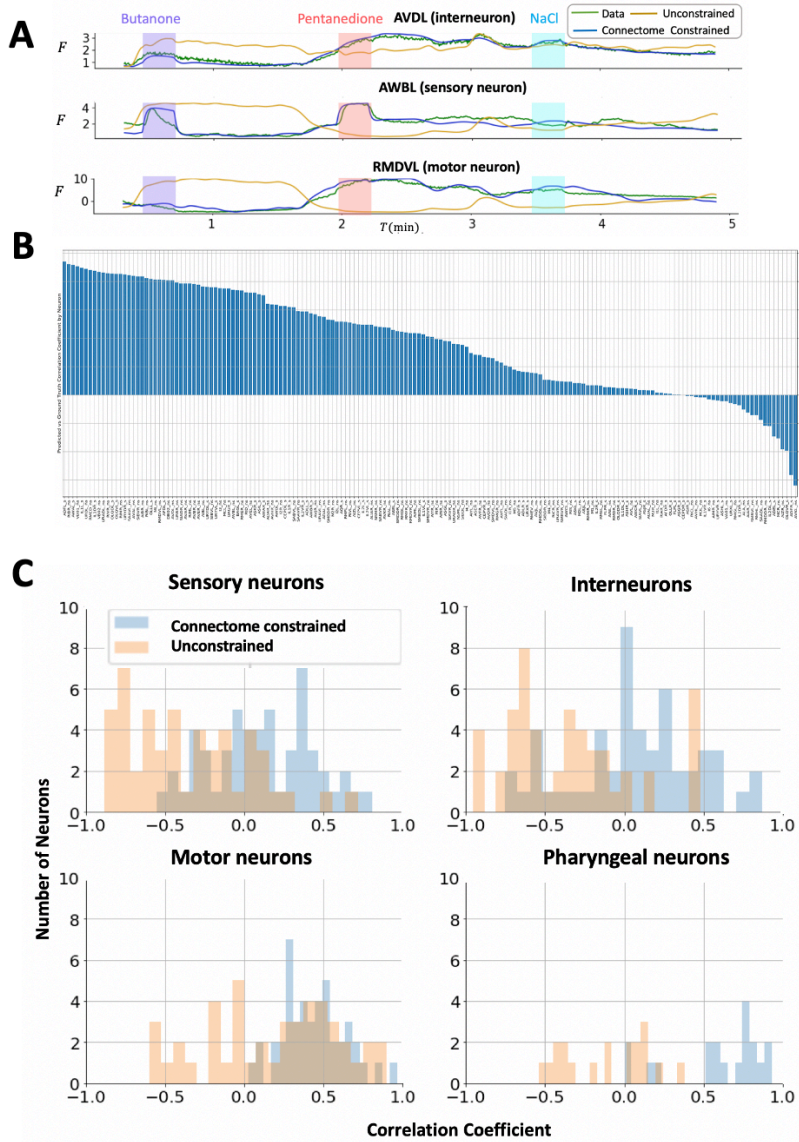
Figure 6.3: **Predicted activity of held out neurons. (A)** Representative comparisons of ground truth data (green), predicted activity from the connectome-constrained model (blue), and predicted activity from an unconstrained model (yellow) for neurons AVDL, AWBL, and RMDVL. **(B)** Correlations between the CC-LVM prediction and ground truth for all recorded neurons. Over 80% of the neuron activity predictions are positively correlated to ground truth. **(C)** Histograms comparing the correlations to ground truth for the connectome-constrained and unconstrained models for the different types of neuron: sensory, inter, motor, and pharyngeal. The CC-LVM makes predictions closer to ground truth across all neuron types.

We hypothesized that the CC-LVM, which incorporates the complete connectome but is trained on the activity of only a subset of the neurons, might nevertheless predict the activity of unmeasured neurons. To test this hypothesis, we conducted a hold-out evaluation, withholding a single neuron class (e.g.: AWAL and AWAR) from the model's training dataset and then evaluating the model's predictions. We compared the model's prediction neural activity both to the ground truth data and to a prediction made from an unconstrained latent variable model (**Figure 6.3A**). We repeated this process for all measured neuron classes (~ 100) and quantified the correlation coefficient between the predicted and measured neural activity. We found that our CC-LVM produced significantly better predictions than an unconstrained LVM with either a fully connected or randomly connected network topology.

The average correlation to ground truth in the connectome-constrained model is 0.479, superior to either randomly connected or fully connected models with the same number of nodes trained on the same data, which have average correlations of 0.05 or less (**Table 6.1**). Additionally, in the connectome-constrained model, the predicted activity of over 80% of the neurons are positively correlated to ground truth (**Figure 6.3B**). In contrast, randomly connected models are less consistent, often producing strongly anti-correlated predictions.

Dividing the neurons by type, into sensory, inter, and motor neurons and computing the average absolute correlations for each (**Table 6.1**), we see that the connectome-constrained model performs well on sensory neurons, though the null models tend to do decently here as well. One might expect the interneurons to be well predicted, since in the connectome they have on average the most connections. However, the interneuron class is predicted worst of the three classes, speaking to the

Model	All neurons	Sensory neurons	Interneurons	Motor neurons
Connectome-constrained LVM	0.479	0.432	0.173	0.631
Randomly connected LVM	0.043	0.131	0.002	0.081

Table 6.1: Average absolute correlations with ground truth for the connectome-constrained model compared to correlations from a representative randomly connected model, averaged across all neurons, sensory neurons, interneurons, and motor neurons.

degree of variability across individuals. The most astonishing gap in predictive performance is in the motor neurons, which are predicted well by a connectome-constrained model which was given only sensory information as input.

6.2.3 Performance is independent of the fraction of recorded neighbors

We can identify the upstream and downstream partners of any given neuron in *C. elegans* from the wiring diagram. However, not all of these partners were recorded in the imaging dataset, such as those in the body and tail of the worm. Naively, we might expect that the connectome constrained model performs better when the majority of the held out neuron's topological neighbors are part of the training dataset.

However, when we compare the holdout prediction performance with the fraction of neurons connected to the held out neuron which are experimentally captured, we find little correlation between prediction performance and either the fraction of input neurons recorded or the fraction of output neurons recorded (**Figure 6.4**). The neurons with a small fraction of their synaptic partners captured are just as likely to be well predicted as neurons with the majority of their synaptic partners
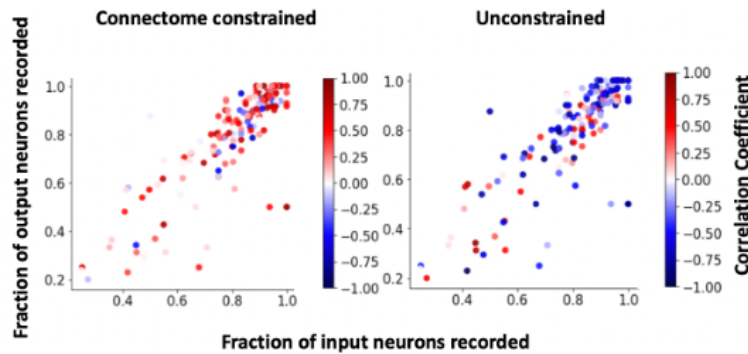
Figure 6.4: Correlation of predicted activity to ground truth for all neurons. (A) **Performance is independent of the fraction of neighboring neurons recorded.** Comparing the holdout prediction performance to the fraction of input neurons (upstream synaptic partners of the held out neuron) and fraction of output neurons (downstream synaptic partners of the held out neuron) recorded, we find no relationship (left). The unconstrained model (right), as expected given its random connectivity, also shows little correlation to the fraction of true neighbors recorded.

captured.

This result suggests that there is sufficient information stored in the network (either upstream or downstream or both) to reconstruct the activity of a neuron even if information from first-degree neighbors is missing. It may also be the case that for neurons whose activity is strongly correlated with chemosensory stimulus, of which we know from Chapter 4 that there are many, the number of recorded partners may be less relevant to predictive performance.

6.3 Discussion

Our results demonstrate that the knowledge of the connectome can aid *in silico* predictions of individual neurons whose activity were not measured, and suggests that information in the *C. elegans* brain is stored in a widely distributed manner. We can interpret the higher accuracy of the connectome-constrained model as confirmation that the global brain dynamics we observed in

Chapter 5 can be generated by a network of neurons wired together in the same topology as the connectome. The anatomical wiring diagram puts measurable constraints on the brainwide activity we observe in the experimental data.

It is not surprising that the network, fed inputs from a sensory encoder, recapitulates sensory neuron activity well. While the CC-LVM predicts interneurons better than a comparable uncon-strained model, overall accuracy is still poor here. Interneuron responses were the most variable of those captured in the pan-neuronal data. While some interneurons clearly showed stimulus evoked responses, the activity of many interneurons were independent of the stimuli being pre-sented (**Section 5.2.7**). These responses are likely from neurons not directly engaged in the task of chemosensation, and may reflect other variables, such as the animal's internal state. The strong performance of the CC-LVM when predicting motor neuron activity suggests that the network is capable of picking out the interneuron signals most relevant to motor output.

It is interesting to note that the trained weights in the model (W_{ji}^{chem} and W_{ji}^{elec}) are correlated (~ 0.3) with the anatomical weights derived from the wiring diagram (synapse count for chemical connections, and contact patch area for electrical connections). This is distinct from the question we asked in Section 5.2.6, when we compared the pairwise functional activity correlations directly to the anatomical weights. In the first comparison, we implicitly assumed that the activity correlations result from direct synaptic communication between the relevant neuron pairs. We have no way of distinguishing functionally relevant neuron pairs (neurons are directly communicating) from neuron pairs which happen to be correlated in activity but are not directly communicating. In the case of the CC-LVM, when training the model on the neural activity data, we are essentially asking

the model to return the most functionally relevant neuron pairs, and this information is encoded in the trained weights. These functionally relevant activity correlations are positively correlated with the anatomical weights in the connectome.

The accuracy of the connectome-constrained modeling approach is fundamentally limited by the fact that the reconstructed wiring diagram is not identical to the wiring diagram of the animal from which functional data was recorded. While many connections are robust across individuals, the existence of weak edges is often highly variable[9]. This means that the network topology used to construct models like this one might not be completely accurate. As the throughput of connectomics imaging continues to increase and automated neuron tracing methods become more accurate, one could imagine, in the near future, conducting functional imaging and then extracting the wiring diagram of the same individual. Such experiments would allow us to directly examine the relationship between anatomical variability and functional variability, and also more carefully dissect the relationship between functional correlations and anatomical weights.

Currently, we are still far away from making biologically testable predictions using this model. The work described in this chapter is ongoing, and we are expanding the training dataset by conducting pan-neuronal imaging experiments with some of the other odorants employed in Chapter 4. A more diverse set of odorants will broaden the input space, a more representative sample of the animal's natural chemosensory task than just three stimuli. Adding select command interneurons or motor neurons to the input space as a measure of the animal's behavioral state may also improve the predictive accuracy of the model.

Deriving causal understanding from high-dimensional whole-brain imaging data is a difficult

problem, one which has not yet been solved. The traditional circuit-cracking approach can be summarized as: (1) Identify activity correlations between neurons. (2) Systematically remove each relevant neuron or neuronal connection via perturbation (usually a mutant or ablation). (3) Observe changes in activity in the remaining neurons. (4) Derive causal relationships between neurons from these activity changes. This approach has proven remarkably successful for understanding small circuits, but it does not scale well to pan-neuronal experiments. It is simply not feasible to experimentally perturb every neuron which reliably responds to chemosensory stimuli. Nor is it likely that the removal of just one neuron would significantly impact the widespread brainwide activity patterns we've observed.

While it's unlikely that we will ever be able to build a fully realistic simulation of the entire *C. elegans* brain, we do hope that biologically-inspired models of neuronal networks like this one can be used to make "lower bound predictions." It is far easier to perform perturbative experiments *in silico* than *in vivo*, and these simulations could help us screen for neurons with particular functions. For instance, let us say that a population of 10 interneurons is reliably responding to a given chemosensory stimulus. From the wiring diagram, we know that each of these interneurons has large numbers of upstream and downstream synaptic partners. Feeding this data into a connectome-constrained model, we find that knocking out 3 of these neurons has a disproportionately large impact on predictive accuracy across the network. We could then focus our experimental efforts on perturbing those 3 interneurons. A tool which could help us identify key neurons in pan-neuronal data would help us move from correlation mapping and towards a causal understanding of information flow. ●

7
Conclusion

In this dissertation, we have detailed our efforts to understand how chemosensory information is encoded in the *C. elegans* brain, and how that information is processed by downstream circuits. We have developed experimental and computational methods which enable us to extract activity from large numbers of neurons *in vivo*, while presenting the animal with highly controlled chemosensory stimuli. We built transgenic animals in which neurons are landmarked with fluorescent labels, allowing us to reliably identify neurons and compare functional data directly with the anatomical connectome.

By characterizing chemosensory neuron responses to a broad panel of odorants, we built a picture of how the *C. elegans* olfactory system represents odor information. Despite its compact size, the olfactory system is capable of encoding odorant identity and intensity in a robust manner. We captured the activity of the entire *C. elegans* brain at single-neuron resolution, and found that large

fractions of the worm's brain were engaged in response to stimulus, and that these responses were stimulus-specific across sensory neurons and interneurons. Coupling this labeled pan-neuronal data with the topology of the *C. elegans* wiring diagram, we were able to build constrained neural-network models of the entire brain *in silico*. The projects described here have laid the groundwork for ongoing and future studies designed to further elucidate the principles of brainwide computation.

7.1 Revising our picture of *C. elegans* olfaction

Previous experimental work on olfaction in *C. elegans* focused on the roles of single neurons. ASH was described as a nociceptor, and AWA, AWB, and AWC were characterized as responding to specific odorants. Many studies have demonstrated complexity in the responses of these single neurons—complexities which may encode information about the olfactory environment. However, the advantages of *C. elegans* as a system (compact nervous system and neuron identification) have perhaps biased the field towards a view of these single neurons as labeled lines.

our results show that more of the chemosensory neurons reliably respond to volatile odorants than previously thought. We see that neurons like AWA, AWB, and AWC are broadly tuned, responding to nearly all of the odorants in our panel. Other chemosensory neurons appear to be narrowly tuned to olfactory stimuli. Together, the broadly and narrowly tuned neurons form a spatial map which provides sufficient information to discriminate odorants. These results suggest that *C. elegans* olfactory encoding is not limited to labeled lines—the responses of the chemosensory neuron ensemble can be interpreted as a distributed or combinatorial code of odorant identity.

We have demonstrated that the activity of the ensemble contains sufficient information to discriminate even chemically similar odorants. Whether the animal is capable of decoding this information, however, remains an open question. We hope that whole-brain recordings which capture interneuron and command neuron activity will help to address this question by elucidating differences in downstream neural activity across odor space. To better understand changes in neural representation, future experiments could employ a time-varying odor mixture stimulus. To correlate behavioral decisions to neural activity, one could employ freely-moving whole-brain imaging techniques to record animals in an arena chip as different odorants are presented.

Our theoretical analyses suggest that the responses of narrowly tuned neurons are critical to odor discrimination. Most of these narrowly tuned neurons have other primary roles. Of particular note are the neurons ASI, ASJ, ASK, and ADL, which are known to detect ascaroside pheromones. More work remains to be done to determine whether the olfactory responses we observe are a by-product of the receptors employed to detect ascarosides, or if olfaction engages a distinct subset of receptors within these neurons.

7.2   Looking forward: studying the interneurons

Our ability to identify all of the neurons in the animal is enabling scale-spanning approaches to *C. elegans* neuroscience: we can capture global brain activity with single-neuron resolution and directly compare activity across individuals. We now have experimental access to the entire input-output function of the brain, from sensory representations of stimulus to the motor commands

that drive behavior. Unpacking the computations that drive behavior, however, will be far from straightforward.

From our work, we've seen that the responses of sensory neurons to stimuli are largely deterministic. Activity in the interneurons and motor neurons, however, can be highly variable across trials and across individuals. Even in interneurons which have reliable stimulus-evoked activity, we find more variability than in the sensory neurons. This should not come as a surprise. It makes sense that sensory neurons respond similarly time and again to the same stimulus—their role is to encode information about the outside world. For interneurons and motor neurons which are engaged in reflex behaviors, it is also expected that a reflex-triggering stimulus will evoke a consistent response. But for any behavior more complex than a simple reflex, we cannot expect that the same set of stimuli will consistently give rise to the exact same actions, executed in the exact same way. The brain dynamically integrates sensory information (possibly from multiple modalities) with internal variables, memory, and motor feedback. These processes engage multiple circuits in the brain, and likely account for the variable responses we see at the interneuron layer across individuals.

C. elegans remains the ideal platform for understanding these computations, again because of the small size of its nervous system. Because the wiring diagram of the animal is so shallow, an average of only 2 to 3 synapses separate sensory neurons from motor neurons. We are currently conducting a series of new experiments in which pan-neuronally labeled animals are presented with repeated trials of the same stimulus. By reading out the activity of command and motor neurons, we can determine the animal's fictive behavioral response to each trial. Through careful identification of neuron populations correlated to different behavioral outcomes, we hope to gain neuron-resolution

insight into sensorimotor transformation.

One key limitation of studying interneurons in *C. elegans* is the observation that many of these neurons exhibit compartmentalized activity: in some neurons, calcium activity can only be captured from the neurites, and in others, different parts of the cell will have different dynamics at the same time. The role these intracellular differences in neurons have in performing computations remains unknown, but by limiting ourselves to nuclear-localized activity readouts, we may be missing a key piece of the puzzle. To address this, we are developing lines in which key interneurons are cytoplasmically labeled, and combining these labels with existing pan-neuronal lines. Recent work has also demonstrated the importance of behavioral context and motor feedback to brainwide activity patterns in *C. elegans*. As we move further away from the sensory layer and into the interneurons, we must take into account the role behavior can play in shaping the activity we observe by coupling whole-brain imaging with either fictive or true behavioral readouts.

7.3 Interpreting whole-brain activity in the context of the connectome

C. elegans is currently the only animal in which it is possible to make direct comparisons at the single-neuron level between whole-brain functional activity and anatomical connectivity. Thus far, comparisons have revealed very low correspondence between functional activity correlations and the synaptic weights in the connectome. Furthermore, while the connectome is largely fixed in its topology (at least on behavioral timescales), we have observed from functional studies that the brain is capable of engaging the same neurons in very different tasks. While connectomic data is undeniably

incredibly valuable, the topology of neuronal connections alone is insufficient to explain the activity of these biological networks. The connectome of an organism is a vital first step to understanding the anatomical features of a neuronal network, but will have to be coupled with metabotropic and transcriptomic maps to build a more complete picture of the structural constraints of information flow.

Determining which of the activity correlations observed in whole-brain data are causal is a key challenge for future work. The traditional circuit-cracking method of ablating neurons one by one and observing differences in activity and behavior is simply not practical when brainwide responses are so far-ranging and variable. Rather, this problem may have to be tackled through a combination of experiment and theory. It is far easier to perform perturbative experiments *in silico* than *in vivo*, and we hope that more sophisticated constrained neural network models will eventually provide sufficiently accurate predictions of neuron function to help screen for particular roles and to guide our experimental hand. ●

A

Supplemental materials for Chapter 2

A.1   Microscope components and filters

Part type	Manufacturer	Identifier
Microscope chassis	Nikon	Ti-E 2000-U inverted
Spinning disk	Yokogawa	CSU-X1
EMCCD camera	Andor	iXon DU-897
EMCCD camera	Andor	iXon Ultra 897
Objective	Nikon	CFI Plan Apochromat VC 60XC WI
Laser combiner	Andor	ALC-500
Laser	Andor	405 nm violet
Laser	Andor	488 nm blue
Laser	Andor	561 nm green
Laser	Andor	640 nm red

Part type	Manufacturer	Identifier
Filter	Semrock	FF01-445/45-25 blue bandpass
Filter	Semrock	FF02-525/40-25 green bandpass
Filter	Semrock	FF02-617/73-25 red bandpass
Filter	Semrock	F01-731/137-25 far red bandpass
Filter	Semrock	red-green dual bandpass

A.2   Microfluidics fabrication

Microfluidics chip patterns were designed in AutoCAD, and the photolithography masks were ordered from outputcity.com. Working in a cleanroom, we then transferred the patterns to silicon wafers via photolithography. Our chip design has two layers, requiring two cycles of photoresist application, UV exposure, and washing, with an alignment between the cycles.

Then, using standard soft lithography methods, the patterned silicon wafers were used as a stamp mold to fabricate the microfluidics devices from 10% polydimethylsiloxane (PDMS). Once the PDMS solidified, it was carefully cut from the silicon wafer (which can be reused), trimmed to size, and had access ports made with a biopsy punch. The devices were then plasma bonded to #1 thickness glass cover slips and cured at $65°$ C for at least 2 hours before use.

A.3 Loading channel dimensions of all microfluidics chip variants

Experiment animal(s)	Channel width (μm)	Channel height (μm)	Notes
Adult	62	32	
Adult (narrow)	62	27	Enforces L/R orientation
Young adult	55	30	Good for small mutants
L4	41	27	
Males	34	27	Works for both head and tail
L3	27	20	Difficult to fabricate

A.4 Parts list for odor manifold

Part	Manufacturer
Perfusion Pressure Kit	AutoMate Scientific
ValveLink 8.2 Pinch Valve System	AutoMate Scientific
Silicone tubing	AutoMate Scientific
Pinch valve fittings	AutoMate Scientific
LabJack UV-L3	LabJack
CB15 terminal board	LabJack
30 mL Luer lock glass syringe	VWR

Part	Manufacturer
Tygon FEP-Lined Tubing, Non-DEHP, 1/16"ID X1/8"OD	Cole-Parmer
Female Luer to 1/16" wide-bore hose barb, polycarbonate	Cole-Parmer
Microbore PTFE Tubing, 0.022"ID x 0.042"OD	Cole-Parmer
PTFE Tubing, 1/16"ID x 1/8"OD	Cole-Parmer

A.5 OPERATIONS MANUAL FOR IMAGING EXPERIMENTS WITH MICROFLUIDICS

SETUP

Power

1. Turn on the two power strips which power the laser sled, camera, and spinning disk.

2. Turn on the computer.

Odor Manifold Setup

1. With gloves on, place the odor syringes you will be using into the odor manifold. Check that the switches of the manifold are open (green lights).

2. With a gentle flossing motion, get the flexible tubing of each syringe into the corresponding valves. When this is done, manually shut the valves using the manifold control box. The valves should click and the lights should go red. If the valve is stuck closed, set the control box button to open (the light should go green). Then gently use a razor blade to open the valve. Then thread the flexible tubing of a syringe into the valve as usual. Try to close and open the valve. If the valve again refuses to open, rinse the valve with DI water, open it again with the razor blade and let dry before attempting use.

3. Stick the ends of each odor tube to the putty on the waste beaker. This is to ensure that in the event of a leak, the solutions drain to a waste container. However, ensure that the tip of the tube does not touch the bottom of the waste beaker, that the tips are not touching

each other, and that the tip itself is not in contact with the putty. Any one of these can cause contamination. If this does happen, cut an inch off the end of the tubing with a razor blade.

4. Pour DI water into the syringes labeled control water (A6) and control odor (A7). You can pour to about ¾ of the way full for these two syringes, as we use a lot of water in a series of experiments. Cap each with its corresponding metal stopper.

5. Pour CTX buffer with sorbitol (balanced to 350 mOsm) into the buffer syringe (A8). Fill about halfway. Be sure that the buffer does not leak, as this can contaminate the setup.

6. Pour the odors into the syringes. Make sure that you load the correct odor to the correct syringe. You should load approximately 5 ml of odor in each syringe. Cap each with its corresponding metal stopper. During experiments, do not let the odor level drop below 1 ml. At less than 1 ml, air pressure can be too variable to produce consistent flow.

7. Fit the retaining plate onto the manifold. The retaining plate can be loosened by loosening the nuts. After it is placed, tighten the nuts to finger tight. This does not have to be very tight, just enough to keep the caps in under pressure.

8. Turn the pressure on all used channels by turning the black switches on the top of the manifold from horizontal (closed) to vertical (open).

9. Check on the pressure dial that the system is pressurized (between 3-5 psi).

Microfluidics Setup

1. Choose a microfluidics chip–we will use a new chip every time we use a new stimulus sequence.

2. Before beginning to load, it might be a good idea to look at the chip briefly under 4x magnification and bright field light for obvious defects. To check for leaks, use a syringe full of water and gently flood the chip with water.

3. With Scotch tape (with tabs for easy removal), tape the chip down on the side of the stage, where we can plug in tubes without stressing the glass.

4. For each control tube, open the valve manually and bleed the tube into the waste container, until there is no more air in the tube and liquid is dripping out. Close the valve.

5. Plug in the three control tubes (Ctrl Water, Ctrl Odor, and Buffer/Water) into the chip, as specified by the chip's design.

6. Bleed the odor tubes in the same way, so there is no remaining fluid. Then plug them into the chip, in any order.

7. Position the waste tube on the chip above the waste port, not plugged in (which can cause pressure problems) but placed such that it sucks away waste coming out of the waste port.

MATLAB Configuration

1. Start MATLAB.

2. Define "all valves." For a single manifold, use the command:

   ```
   all_valves = {'A1','A2','A3','A4','A5','A6','A7','A8'}
   ```

3. Launch "OdorGUI."

4. Configure the odors and odor delivery program in OdorGUI, either manually, or by loading a saved preset using "Load Settings." If you want to save custom settings, use "Save Settings." Double check to make sure that the channels on OdorGUI reflect the odors in the manifold, both in identity and concentration. Give the experiment a name.

5. In Windows Explorer, create a new data folder, and point the OdorGUI program to the data folder.

6. Use the smset command, `smset('all_valves',0)`, to set computer control of valves to all closed, then toggle the valves open and closed to confirm that the computer has correct valve control. This has to be done every time we manually manipulate the valves.

7. When all the tubes are inserted to the chip, run the clean cycle and check that both the chip and the manifold are leak-free.

8. Ensure that the bars on the bottom match those of the programmed experiment—you can do this quickly by checking the number and color of the bars.

NIS Elements Configuration for R/G Landmark Stack

1. Start NIS Elements and configure it for the ID stack.

2. Select the GFP-RFP preset on NIS Elements.

3. Set laser power settings as appropriate for the experiment.

4. Check the following microscope settings: Under DU-897 Settings, No Binning, Readout Mode EM Gain 10 MHz at 14-bit, EM Gain Multiplier 100, Conversion Gain 1x, Temperature around -70 C. Under Manual Microscope Pad, Nosepiece is set to 60x. Under Filters and Shutters, AOTF, MasterALC, and CSU open, Filter set to Red/Green. Under tab Triggered Acquisition, Line 1 is 488 nm, Line 2 is 561 nm, Triggering Enabled. If any windows have been accidentally closed or missing, look under "View/Acquisition Controls."

5. Under the ND Acquisition tab, change the path to be the same as the path set for OdorGUI.

6. Under the "Z Series" tab, check that the volume setting is set on the Middle + Range mode. Step should be 1.5 um, Range 30 um for a total of 21 steps, Z Device should be "Triggered NIDAQ Piezo Z."

7. Under the "Time" tab. Ensure that the length of acquisition is consistent with the odor sequence used.

Experiment

Loading an Animal

1. With a syringe, pull up about 2 mL of the loading buffer (either M9 buffer or M9 buffer with 2 mM tetramisole). Then attach an elution filter to the syringe and pass the solution through the filter into a clean petri dish. This will help catch dust and other solids in the solution.

2. Inspect your loading syringe. If the glass tip is dusty, rinse it off with DI water.

3. Fill the loading syringe about halfway with loading buffer.

4. Using the dissection scope, pick a young adult animal off of the plate and place it in the loading buffer. You should be able to see the animal in the buffer by eye.

5. Suck the animal up the loading syringe. Do not do this aggressively—try to keep the animal as close to the tip of the glass as possible. This makes loading easier.

6. Before introducing the animal to the chip, check the chip again for air bubbles, and run a clean cycle, keeping the valves open once the cycle is complete. Having liquid flowing through the chip ensures that any small debris or air introduced during the loading process will flow downstream of the animal.

7. Plug the loading syringe into the loading port. Do this gently to avoid breaking the glass cover slip, while bracing the microfluidics chip with two fingers.

8. While watching the chip through the microscope binoculars under 4x magnification, begin gently applying pressure to the loading syringe until the animal appears in the loading pool. Gently apply slight positive and negative pressure until the animal rotates such that its head points in the direction of the loading channel.

9. Load the animal head first into the channel, such that the tip of its head protrudes out of the channel and in the flow. Use slight positive and negative pressure inputs to adjust the animal's position, and to keep it from drifting either forward or backward.

10. Once the animal is loaded, run the clean cycle and/or use partial vacuum pressure to clear the chip of air bubbles.

Running an Experiment

1. Center the head of the animal under 4x magnification.

2. Switch to the 60x objective by manually turning the objective turret. The 60x is a water immersion objective, so apply two drops of DI water to the objective before swinging it under the microfluidics chip. Be sure that the correction collar on the objective is set to 17.

3. Using binoculars, focus in 60x and center the animal's head.

4. Switch the microscope to confocal configuration by turning the safety interlock on the front panel from "Bino" to "Confocal." Turn the light port on the lower right of the microscope body from 1 to 5. Turn off any room or table lights, but keep the microscope bright field on to assist you in fine focus.

5. In NIS Elements, hit "Live" (green play button). You should be able to see an outline of the animal via bright field even if you do not see any fluorescence. Using the fine focus knob, move up and down in z until you see fluorescence. Once you do, turn off the microscope bright field, and center the neurons in the field of view.

6. In live view, go to the ROI drop-down menu and hit "Define ROI." Drag the box lengthwise to form a 512x256 rectangle parallel to the animal's head. Do not drag or position the box with the mouse.

Getting a Landmark Stack (1 time per animal)

7. Select the GFP-RFP preset tab.

8. Under the tab "DU-897 Settings," confirm that the Format for Live and Format for Capture are both set at "No Binning," and that the exposure is "1 frame."

9. On the "ND Acquisition" panel, go to the "Z Series" tab. Define a Home position by hitting the Home button with a red arrow. Then jump to the top and bottom positions in live view by double-clicking the green numbers on the right side of the box (+15, -15). Check to see if the animal is centered. If not, return to the home position by double clicking the red box on the right, and adjust the z-position of the center using fine focus. Redefine the Home position by clicking the Home button on the left, and check the top and bottom again. We want to bracket the entire head of the animals in the volume. When you are satisfied with the animal's position, return to the Home position by double clicking the red box on the right.

10. Check that the file path is correct, and name the file "idXYY."

11. Go to the "Time" tab in the ND Acquisition panel, and set the experiment time to 1 loop.

12. Hit the "Stop" button in NIS Elements to stop the live view.

13. In Elements, hit "Run Now" in the ND Acquisition panel. This will begin the experiment run and acquire an ID stack with a single time point.

Capturing an Experimental Movie (multiple times per animal)

14. In Elements, switch the preset to "GFP."

15. Under the tab "DU-897 Settings," confirm that the Format for Live and Format for Capture are both set at "Binning 2x2," and that the exposure is "1 frame." The frame size should now be 256x128. Set power on the 488 nm.

16. On the "ND Acquisition" panel, go to the "Z Series" tab. Define a Home position by hitting the Home button with a red arrow. Do not move the Z position before defining Home. It is important that the Home position for the experiment is identical to the Home position in the ID stack.

17. Go to the "Time" tab in the ND Acquisition panel, and ensure that the experiment time is long enough to capture the odor delivery sequence you have programmed (usually 4:30).

18. In both Elements and OdorGUI, double check that the path and filename are correct. The nomenclature we use for experiments is runXYY, where X is the animal number and YY is the two-digit run number. For instance, the first run of animal 2 will be "run201."

19. Hit the Stop button in Elements to stop the live view. This is important for the trigger signal to start correctly.

20. In the OdorGUI, hit "Run Experiment." The MATLAB program will wait for a signal from the camera to start concurrently with the acquisition.

21. In Elements, hit "Run Now" in the ND Acquisition panel. This will begin the experiment run. If something goes wrong, do not hit Abort. Wait for the experiment to finish and fix it on the next run.

22. When an experiment is complete, run a clean cycle. If you want to look at the movie, replay it and look for motion and activity to make note of. Use the volume button in the Elements viewer and switch it to Max Projection.

23. While the clean cycle is running, take the opportunity to make a quick check of the following:

    (a) Animal position. Is the animal still centered in the field of view? Is it centered in Z? If not, reposition the animal and/or stage.

    (b) Check the odor levels in the odor syringes, ensure that none are empty. If one is getting close to empty, put gloves on, turn off the pressures, refill it, then recap and turn the pressures back on. Do not let a syringe run completely dry, then it will require bleeding again.

    (c) Check that the odor tubes are still properly attached to the chip.

(d) Check that the waste bubble is small and the vacuum is appropriately positioned. If the waste starts to get in between the tape and the glass of the chip, we will have z-positioning problems and will have to wait for the setup to dry.

(e) If you saw little activity, switch to binocular mode ("Bino" on the interlock, Light port to 1) to check if: (i) The animal's nose is still sticking into the flow. If not, use the syringe to reposition the animal. (ii) There are any blocking air bubbles. If so, use clean cycle and/or partial vacuum to flush them.

(f) Make notes of any changes or observations.

24. After the clean cycle is complete, double check that Elements is stopped (not in live mode). Then repeat from step 18.

Removing an Animal

1. After you complete your experiments, it is time to remove it from the chip. Switch back to 4x magnification, and try the following strategies in order:

    (a) Strategy 1: Run a clean cycle, plug the vacuum line fully into the waste port, and flush the animal to the waste port. Apply positive pressure from the loading syringe if necessary.

    (b) Strategy 2: Suck the animal out using the loading syringe. If it comes free, unplug the syringe and empty it into the waste beaker.

    (c) Strategy 3: For large animals. Stick the vacuum line into the loading port, then stick the loading syringe (with water or buffer) into the waste port. Manually open all water channels, and apply gentle pressure with the loading syringe. This should force the animal back up into the vacuum.

    (d) Strategy 4: If the animal will only move forwards or is already in the chip, manually open all water channels, plug the vacuum line fully into the waste port, and push with the loading syringe until the animal is flushed.

2. Any animal removal may result in some air bubbles being introduced. Once you complete the extraction, run clean cycle until no more bubbles are present upstream of the loading chamber.

Teardown and cleanup

General Shutdown

1. Close all valves, and remove the last animal from the chip.

2. Exit NIS Elements.

3. Copy the data folder over to an appropriate place on the lab servers.

4. Put on gloves. Remove the chip from the center of the stage, and tape it back down on the side of the stage as done when setting it up.

5. Check that all valves are closed. Then one by one, carefully unplug each tube from the chip and attach it to the putty on the waste beaker, such that the tips do not touch each other, the waste liquid, or the putty.

6. Open all valves and let the excess liquids drain into the waste tube.

7. If running with CTX buffer, pour 10 mL DI water into any syringe that was loaded with a buffer solution during the experiment. Let the DI water drain.

8. With gloves on, gently remove the elastic tubing from each pinch valve.

9. After the liquids have drained, turn the pressures off and remove the retaining plate. Remove the caps and let the tubes air dry. Be sure to pull the caps out gently, straight up out of the glass syringe. Applying force at an angle may result in the glass syringe breaking. Remove the odor syringes from the manifold and put them in their storage racks (the three control syringes stay).

10. Discard the chip in glass disposal, or if you intend on using it again, make a note on the chip itself of which stimuli were used today.

11. At least once a week, clean the valves using a squirt bottle of DI water. Give the pinching area of each valve a long squirt of DI water, and let air dry. Do not spray water around the odor manifold in a random or uncontrolled manner—there are electronic components that do not react well to water.

12. Exit MATLAB, shut down the computer, and shut down both power strips which power the microscope.

B

Supplemental materials for Chapter 3

First time setup

1. Unpack the Analysis Package ZIP file somewhere on the local computer. This file should contain the following folders:

   - `matlab_tracking`: Contains the programs and dependencies for processing and tracking the data movies

   - `matlab_traces`: Contains the programs and dependencies for plotting, proofreading, and analyzing the data

   - `mat_vol_viewer`: Contains tools for visualizing any movies in .mat format

2. Add all these folders to your MATLAB path. Adding these folders to the path allows MATLAB to call these scripts from any location. Use a version of MATLAB 2018 or later.

3. Create a folder in which to put your analyzed data. This folder will contain your Data Status Excel file, and subfolders which will contain your analyzed data.

4. Open up `Analysis Package CTX\matlab_tracking\CTX_analyze.m` in MATLAB. This is the master analysis script, which calls all of the other programs. Open `CTX_load_excel.m`

in the same folder as well. In `CTX_load_excel.m`, and Step 0 of `CTX_analyze.m`, edit the code which identifies the computer you are using, and points to the appropriate location of the Data Status Excel file (`file_location`) and your version of the annotator (`annotator_root`). Optionally, you can set up this part of the code to run seamlessly on multiple machines.

Assessing the data

1. Copy the data to the local computer. Add the movies to the Data Status Excel file, and fill the following fields:

   - Raw Data Root: Location of the raw data on the local computer
   - Analyzed Data Root: Location where processed data will be placed
   - Animal: Give each animal a unique ID (ex: `[Stimulus]_###`)
   - ID stack: The landmark volume filename (ex: `id###`)
   - Run: The data run filename, (ex: `run###`)
   - Stimulus: Note the stimulus and the concentrations
   - Status of Steps: Set all squares to "FALSE", update them to "TRUE" when complete

   The MATLAB code will read metadata from the Excel file, and this metadata will tell the code where to look for data files and place analyzed files. The "TRUE" and "FALSE" flags are read as booleans by the code.

2. Using the NIS Elements viewer or your visualizer of choice, watch all the movies of a given animal. This viewer uses the same basic controls as NIS Elements on the microscope. Use the "Show Slices View" button to view the movie as three projections in different axes and set the Mode to Maximum Intensity Projection. To adjust the brightness, go to Window/Show LUTs Window, and use the lookup tables to adjust the contrast. You can also play around with the 3D "Volume View."

3. Watch the movies, taking qualitative notes. Pay close attention to the following:

- What is the motion of the animal like? If there is continuous motion in X and Y, this is likely still trackable, but make a note of it on the Data Status spreadsheet. If there is jerky, sudden motion in X and Y, or any drift in Z, the movie is probably not usable.

- What is the orientation of the animal? Is it a symmetric, lateral view? If so, make a note as to whether the left or right side of the animal is close to the objective. If the volume is not symmetric, estimate the angle of rotation and make a note of that.

- What is the activity like? If you don't see any activity, it might suggest something went wrong with the odor delivery.

Preprocessing the data

1. Open `CTX_analyze.m` in MATLAB. At the top of the Step 0 cell, input the animal ID, matching the animal ID in the Data Status Excel file. Also, input the strain of the animal (usually 'ZM10104'). With the cell highlighted, press Ctrl+Enter to run the cell. The MATLAB workspace should be populated with information about the animal, datasets, and the file and output locations.

2. Through the MATLAB analysis, you can exit the process and quit MATLAB whenever a cell is complete. The next time you start MATLAB, rerun the Step 0 cell with the appropriate animal ID, and then pick up where you left off.

3. In `CTX_analyze.m`, run the Step 1 cell (click in the cell to highlight it, then press Ctrl+Enter to run it). The scripts called in this cell convert the ND2 files of the datasets into several formats: raw TIFF files, .mat volumes at two compressions (16 bit and 8 bit), and JPEG images for the annotator. One of the reasons we do this is that the ND2 data format is proprietary to Nikon and is difficult to work with. It also extracts time and stimulus information from the movies and the stimulus text file. The times.json file contains the timestamps for each volume in the movie. You can view the two .mat volumes using the `mat_vol_viewer`. When this cell is done running, flip the "Volumes Made" field on the Excel sheet to "TRUE".

   The scripts in the Step 1 cell are described here. `convert_ND2_to_TIF_mat.m` generates TIFF volumes in both red and green channels from the ND2 movie. It also generates two 16

bit .mat volumes with some Gaussian smoothing (`gcamp_vol_16.mat`, `red_vol_16.mat`) and a times.json file in the `run###` folder in the local data folder.

The next script, `uint8_volume.m`, compresses the 16 bit .mat volumes into 8 bit .mat volumes, and saves these to the analysis folder. This compression reduces the online file size, and more importantly makes the tracking more efficient. The compression is done automatically with an algorithm which identifies the dynamic range which captures the most variance. By default this quantile target range is set to 0.9-0.9999, this can be adjusted manually if the 8 bit volumes look overly compressed.

The third script, `jpg_generator.m`, generates green, red, and maximum intensity projection JPEGs. These are for use in the annotator. Unlike TIFF files, you can't adjust the brightness of JPEGs ex post facto. By default, the brightness is set very high, as the annotator is mostly for identification. If you ever need to adjust this manually, change the denominator in the `process_image` function in the `jpg_generator` script (by default set to 1). The last script in this cell, `times_stimuli.m`, extracts stimulus identity and delivery times from the `run###.txt` file generated by the OdorGUI. It combines this information with the times.json file, and generates a `run###metadata.mat` file in the analysis folder.

Annotating the neurons

1. Add the preprocessed animal to the datasets.json file in the annotator folder. Create a new entry in the file:

```
    "S_001": {
  "id": "S_001",
  "shape_x": 256,
  "shape_y": 128,
  "shape_z": 21,
  "shape_c": 2,
  "shape_t": 5,
  "pixel_size_x": 2,
  "pixel_size_y": 2,
  "pixel_size_z": 8}
```

Fill the initial name and the id field to match the animal ID. Change the `shape_t` field to match the number of datasets this animal has, plus one for the landmark.

2. Start Web Server for Chrome, click CHOOSE FOLDER, and select your annotator folder. Then click on the Web Server URL to launch the annotator GUI, written in JavaScript. In the "Select a dataset" dropdown, select the animal, and the JPEG images should appear. If no datasets show up in the dropdown, there is probably a mistake in the datasets.json, either a missing comma or bracket.

3. We will now annotate the neurons in the head volume. The annotator has three windows: an XY projection, XZ projection, and YZ projection. Each window has four options, Selection (Default), Pan, Zoom In, Zoom Out, and Fit to Viewer. Use these to move about, but stay in Selection mode when making annotations. The sliders will move through the volume in X, Y, and Z, with the gray lines indicating the planes of view. The hotkeys for the annotator (active when the XY projection window is selected) are:

   - D: previous recording
   - F: next recording
   - C: up one frame in Z
   - V: down one frame in Z
   - E: previous color channel
   - R: next color channel

   We want to fully annotate the animal's landmark volume. The computer will then attempt to carry those annotations over to the other recordings, using a handful of manually labeled neurons in the remaining recordings. Make sure you start on the landmark volume. To annotate a neuron, move in Z in the green channel until you find a location that you think is close to the center of the nucleus of the neuron in X, Y, and Z. Double click to create an annotation. This should appear as a red dot. A red dot is an annotated but unidentified neuron. To label the neuron, click the red dot, and a text box will appear. Write down your best guess for the name of the neuron (already used labels appear in grey in the dropdown, while unused neurons appear in white). To delete a bad annotation, hit the trash can button. Annotate all of the neurons of interest.

148

4. When all the neurons are annotated in the landmark volume, decide on a handful of neurons (3-6) which you can ID with confidence. These neurons will be used by the computer to try and morph the annotations map from the landmark volume onto the other recordings, so ideally these neurons are spread widely in X, Y, and Z. Moving through the remaining recordings, annotate and label these landmark neurons (and only these landmark neurons). Make sure that for the remaining recordings, the same neurons are labeled. When this is complete for all remaining recordings for the animal, save the annotations in the annotator folder, overwriting the old annotations.json file. Flip the "ID-ed" field on all datasets for the animal on the Excel sheet to "TRUE".

TRACKING THE NEURONS AND GENERATING TRACES

1. Highlight the Step 3: Track cell and run (click in the cell to highlight it, then press Ctrl+Enter to run it). This cell will take the position data from the annotator, track all of the neurons across the datasets, pull out calcium activity traces, and generate rough plots. The tracking script, `CTX_track_neurons.m`, will generate plots at the end of each dataset, showing the results of the tracking. It will also continuously be outputting fit numbers for each frame—the higher these numbers are, the better the tracking is likely to be.

   The tracking code, `track_neurons.m`, works by taking a large neighborhood around each nucleus in the green channel, initially centered about the position in the annotator. Then the algorithm looks for the volume in each frame which best correlates to that initial volumes. You can adjust the size of the neighborhood and the search parameters in this script. This script generates a `run###_annotations.mat` file.

2. Once the tracking is complete for all of the datasets, the calcium activity is extracted from the movies using `generate_traces.m`, which sums the intensity over a small area around the center of the nucleus. The feature size volume and the number of retained pixels can be adjusted manually in the `generate_traces.m script`. This script generates two files, `run###annotations_gcamp.mat` and `run###_traces.mat`.

3. `plot_traces.m` does baseline and bleach correction, and then plots normalized traces from all of the annotated neurons, labeled by their tracked labels. This script generates a .fig of the raw traces in the dataset folder.

4. When the entire cell is complete, if there are no issues, flip the Tracked, Traces, and Plots fields on the Excel sheet to "TRUE".

PROOFREADING NEURONAL ACTIVITY

1. To proofread the traces, open the `dataset_proofreader.m` script in MATLAB. Set the dataset run ID and the analyzed root, the location on Dropbox where the processed data is stored. Then run the first cell, which loads the traces.mat and generates a proofread data structure (`run###_prfrd_data.mat`).

2. Run the second cell, which will plot all of the traces, labeled with their tracked IDs.

3. Running the third cell launches the Proofreader GUI. The controls for the GUI are as follows:

   - Left Arrow: Previous Neuron

   - Right Arrow: Next Neuron

   - E: Edit neuron identity (Enter the corrected neuron name in the popup)

   - B: Adjust the baseline F_0 value (The computer has made an automatic best guess at F_0. to adjust it up or down, input a positive or negative number in the popup)

   - G: Flag a trace as problem-free (green)

   - F: Unflag a trace (yellow)

   - D: Flag a trace as unusable (red)

   - S: Save

   - Esc: Save and exit

4. When proofreading the datasets, we are looking to exclude bad tracks (ex: missing frames, jumping to an incorrect neuron). We can also manually adjust the baseline F_0. More rarely, we will also have to correct mis-identified neurons. We will then flag the neurons as problem-free (green) or unusable (red). A green flag should be given to a neuron for which you have

high confidence that the identity is correct, the tracking was carried out correctly, and the baseline was correct. Note that a green flag does not have to do with activity. Neurons with no activity at all, but you are confident about the identity of, should still receive the green flag. Neurons you aren't sure about, or want to return to later, should remain yellow (hotkey F). Grounds for a red flag (hotkey D) can include:

- Highly unusual traces can indicate poor tracking (such as sudden spikes which do not look biological)
- Broken traces indicate that the track was lost
- Contaminated signal from a neighboring neuron
- Position suggests it might not have been a neuron at all

5. When you are done proofreading a set, save it exit the proofreader (Esc). This generates a `prfd_data.mat` file for that dataset. Flip the Proofread column on the Data Status spreadsheet to "TRUE". Repeat for all datasets for a given animal.

6. When you have completed proofreading for all of the datasets for a given animal, open the `avg_traces.m` script. Input the animal ID and line in Step 0, then run the Step 0 cell. Run the following 3 cells, which loads the proofread data, filters for green flags, and compiles the data from multiple runs into a single data structure.

7. The next cell (optional) allows you to manually inspect neurons by plotting their traces individually. Run the final cell to generate average and individual trace plots in the analysis data folder, and save a large `avg_data.mat` data structure.

C

Supplemental materials for Chapter 4

C.1   Supplemental methods

C.1.1   Worm maintenance

All worms used in this chapter were raised at 20° C on NGM plates, and fed OP50 *E. coli*[182]. Animals used for experiments were picked from plates which not been starved for at least 2 generations.

C.1.2   Identifying neurons in the ZM10104 strain

The ZM10104 strain used in this experiment expresses two fluorescent proteins: GCaMP6s driven by the *ift-20* promoter, and wCherry driven by *gpc-1*. GCaMP6s expression was localized to neuronal nuclei to minimize spatial overlap of neighboring neurons, and to make identification of neurons easier. The promoter *ift-20* drives GCaMP expression in all ciliated sensory neurons. Our neurons of interest, the chemosensory neurons, lie in the lateral ganglia, but note that this promoter drives expression in cells outside of the lateral ganglia. The wCherry landmark is expressed in the

cytoplasm of AFD, AWB, ASI, ASE, AWC, and ASJ. Note that it also is expressed in RIB, a neuron which is not labeled with GCaMP.

Relative positions are given in the orientation in **Figure C.1**, with the nose to the left, the tail to the right, dorsal top, and ventral bottom. Relative positions should be interpreted as "usually but not always," unless otherwise noted. Also note that overly compressing an animal in any direction will distort the relative positions. Before identifying neurons, it is important to identify the orientation of the animal in the recording by figuring out where the dorsal-ventral (DV) plane lies. This is most easily done by identifying the plane of bilateral symmetry. Once you have oriented yourself, you can begin the neuron identification process.

The easiest neurons to immediately identify in this strain are ASH, ASJ, and the anterior "triplet" of ASK, ADL, ASI. It is often effective to identify these neurons first, then work on the other neurons using the color landmarks and process of elimination. AWC and ASE should usually be in the neighborhood of ASH, and the four neurons AWA, AWB, ADF, and ASG are between ASH and the anterior triplet. These four neurons occasionally overlap. To avoid signal mixing, overlapping neurons were exluded from the dataset. For each odorant condition, neuronal identification was carried out independently by at least two individuals.
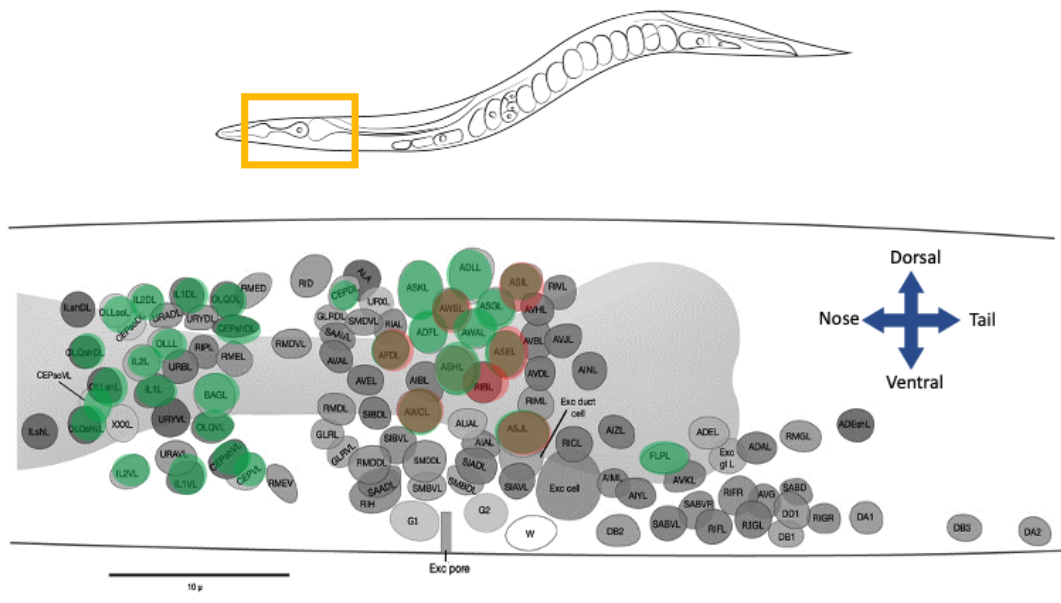
Figure C.1: Identifying neurons in the ZM10104 strain. The *ift-20* promoter drives GCaMP expression in the nulcei of ciliated sensory neurons. The nuclei of the chemosensory neurons are all posterior to the nerve ring. A red landmark is provided by cytoplasmic expression of wCherry in the neurons AFD, AWB, ASI, ASE, AWC, and ASJ.

Criteria for identifying each neuron class

Neuron	Color(s)	Relative Position	Notes
ASK	green	leftmost of the anterior triplet	large. do not confuse with URX, a small oblong neuron above ASK
ADL	green	part of the anterior triplet	larger than ASI
ASI	green & red	part of the anterior triplet	use color to distinguish from ADL
ASH	green & red	left of ASE, below AWA	bright, circular
ASE	green & red	right of ASH	smaller than ASH
AWC	green & red	variable. below ASH but can be to the left, directly below, or to the right	often oblong in shape
ASJ	green & red	tail end of the ganglion, bottom left	distance from AWC can vary
AWA	green	variable. usually above ASH	smaller than ASH, circular
AWB	green & red	variable. usually directly below the anterior triplet	small, dim, a bit oblong. use color to identify
ADF	green	usually left of AWA, AWB	dim
ASG	green	usually right of AWA, AWB	small, circular

C.1.3 LIST OF ODORANTS

Odorant	Chemical class	Behavioral valence (low conc.)
1-pentanol	alcohol	attractive
1-hexanol	alcohol	attractive
1-heptanol	alcohol	repulsive
1-octanol	alcohol	repulsive
1-nonanol	alcohol	repulsive
isoamyl alcohol	alcohol	attractive
geraniol	alcohol	attractive
benzaldehyde	aromatic	attractive
methyl salicylate	aromatic	attractive
ethyl acetate	ester	attractive
ethyl butyrate	ester	attractive
ethyl heptanoate	ester	attractive
pentyl acetate	ester	attractive
butyl butyrate	ester	attractive
diacetyl	ketone	attractive
2-butanone	ketone	attractive
2-heptanone	ketone	attractive
2-nonanone	ketone	repulsive
2,3-pentanedione	ketone	attractive
2,5-dimethylpyrazine	pyrazine	attractive
2-methylpyrazine	pyrazine	attractive
2-isobutylthiazole	thiazole	attractive
2,4,5-trimethylthiazole	thiazole	attractive

C.1.4 IMPUTING MISSING SINGLE-TRIAL RESPONSES

Across trials of all neurons and all conditions, about 20% of the neuron responses were either not captured, or excluded due to tracking mistakes or signal contamination issues. To compare single trials, we had to fill these missing data points in a reasonable and biologically motivated way.

For a given odorant and M trials, the peak responses of the $N = 11$ sensory neurons can be

compiled in a matrix $R \in \mathbb{R}^{N \times M}$. To impute the missing responses, we began by determining

the distribution of peak responses of every neuron to every odorant (**Figure C.7A**). Due to the

intrinsic correlation between the peak responses of different sensory neurons, the matrix R is low

rank. Sampling from these distributions, we employ a matrix completion algorithm to efficiently

impute the missing data points (**Figure C.7B**) [183,184].

Figure C.7

C.1.5 Computational methods for discriminability quantification

For binary classification of all odorant pairs, we used linear regression and a simple SVM (linear or

Gaussian kernel). To decode odor identity from the entire single-trial dataset, we built a multi-class

classifier. We concatenate all of the single-trial responses of the 23 odorants at high concentration.

Each trial is an 11-dimensional point, one dimension for every neuron class. Each point has an asso-

ciated label indicating the odorant identity. This data set was randomly divided into 10 parts, 9 of

which are used as a training set (90%) and one which is used as a validation set (10%).

We used the MATLAB function `fitcecoc` to fit a multi-class model which supports both SVM

and other classifiers. Mechanistically, this method reduces the problem of overall classification into

a sequence of binary classification problems. The performance was quantified by the classification

error, estimated using the `crossval` function. The confusion matrix was generated using the func-

tions `kfoldPredict` and `confusionchart`. The training is repeated 10 times, using each of the 10

parts of the datasets as the validation set, and the results were compiled.

For the *in silico* knockouts, we removed neurons from the training dataset, resulting, for example

in 10-dimensional responses when one neuron was removed. We trained the multi-class classifier as above.

C.1.6 STATISTICS, CODE, AND SOFTWARE

All statistical computations and image analaysis code were written and run in MATLAB using standard toolboxes, with the exception of the OME Bio-Formats API (used to read Nikon ND2 file formats)[185]. The tracking code and software will be made freely available upon publication of the manuscript.
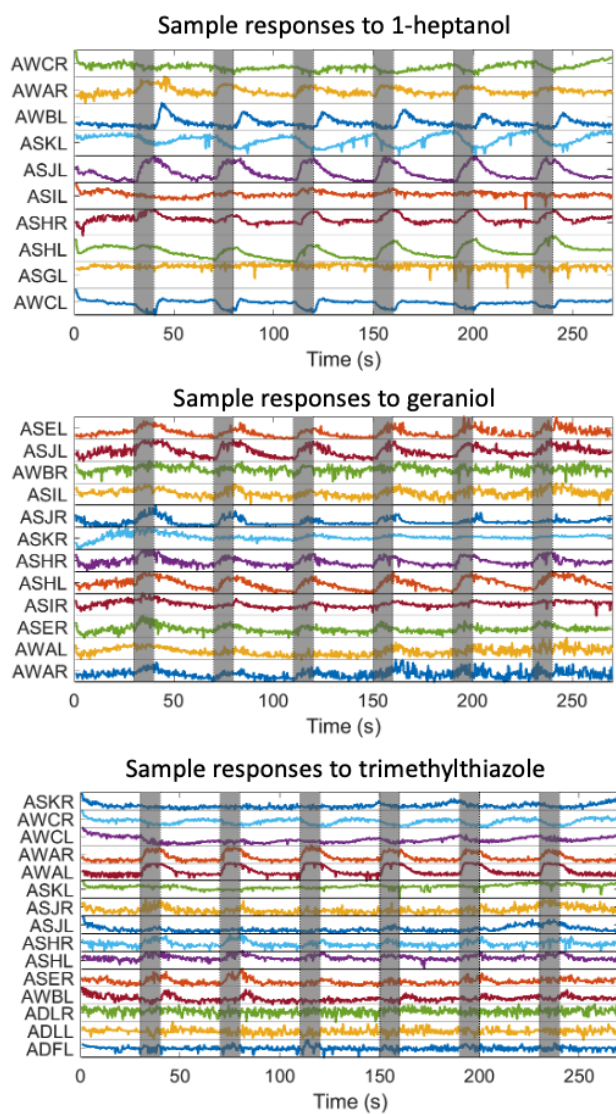
Figure C.2: Sensory neuron activity traces from representative experiments. Calcium activity traces from the chemosensory neurons captured in three representative experimental recordings in response to three different olfactory stimuli. This animal was presented with six 10-second pulses of 1-hexanol at 10^{-4} dilution (gray), separated by CTX buffer blanks. The animal was imaged for 270 seconds at a rate of 2.5 Hz.
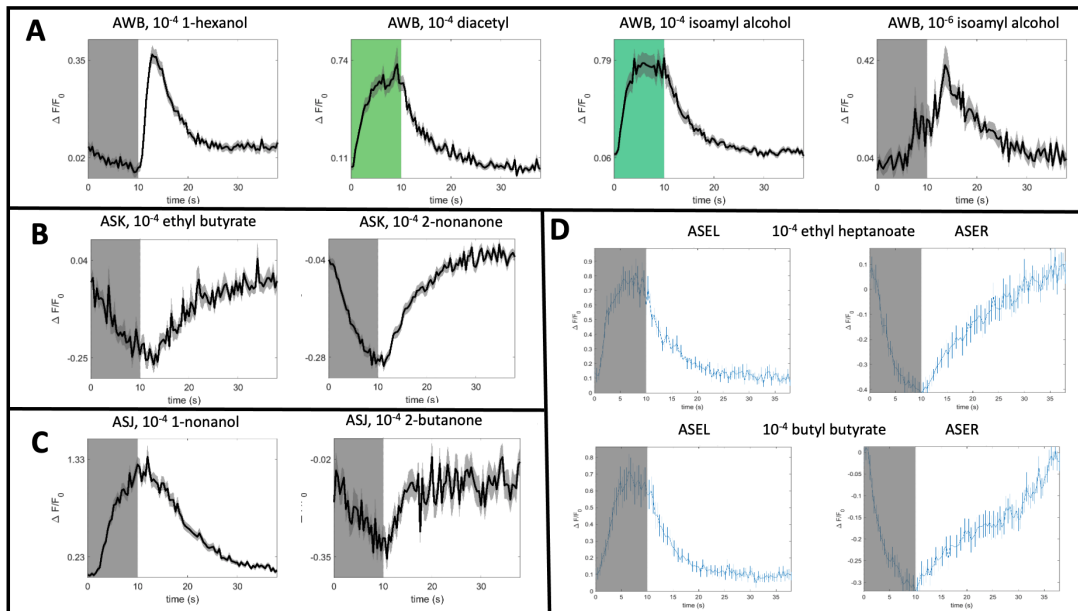
Figure C.3: Single neuron response observations. Average neuron responses, with odor delivery indicated by the gray/colored bar. **(A)** AWB is an OFF response for most stimuli, such as 1-hexanol, but is occasionally an ON response, as is the case for high concentration diacetyl. High concentration isoamyl alcohol elicits an ON response from AWB, but low concentration isoamyl alcohol elicits an OFF response. This has been previously observed in Yoshida et al., 2012[87]. **(B)** We observe inhibitory responses to some odorants in ASK. **(C)** ASJ has an excitatory response to some odorants, such as 1-nonanol, but has an inhibitory response to 2-butanone. **(D)** We observe L/R asymmetries in ASE in response to several odorants, such as ethyl heptanoate and butyl butyrate here.
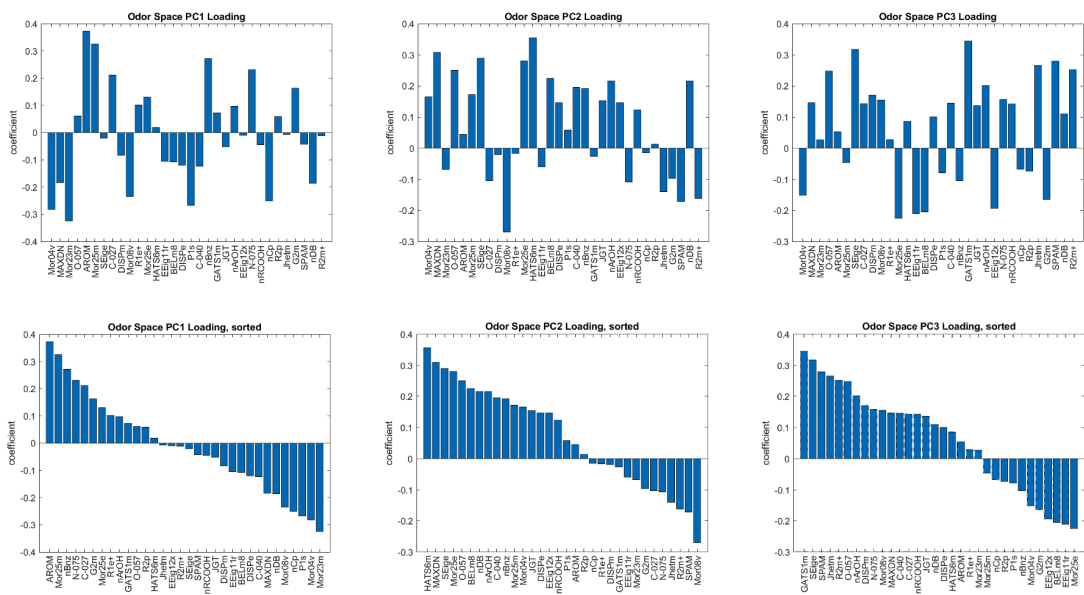
Figure C.4: Principal components of the odor space. The molecular descriptor loadings of the first 3 principal components of the *C. elegans* odor space, plotted on the same axes above, and sorted by coefficient below. The leading components of PC 1 are measures of aromaticity, and the leading components of PC2 are measures of electronegativity.
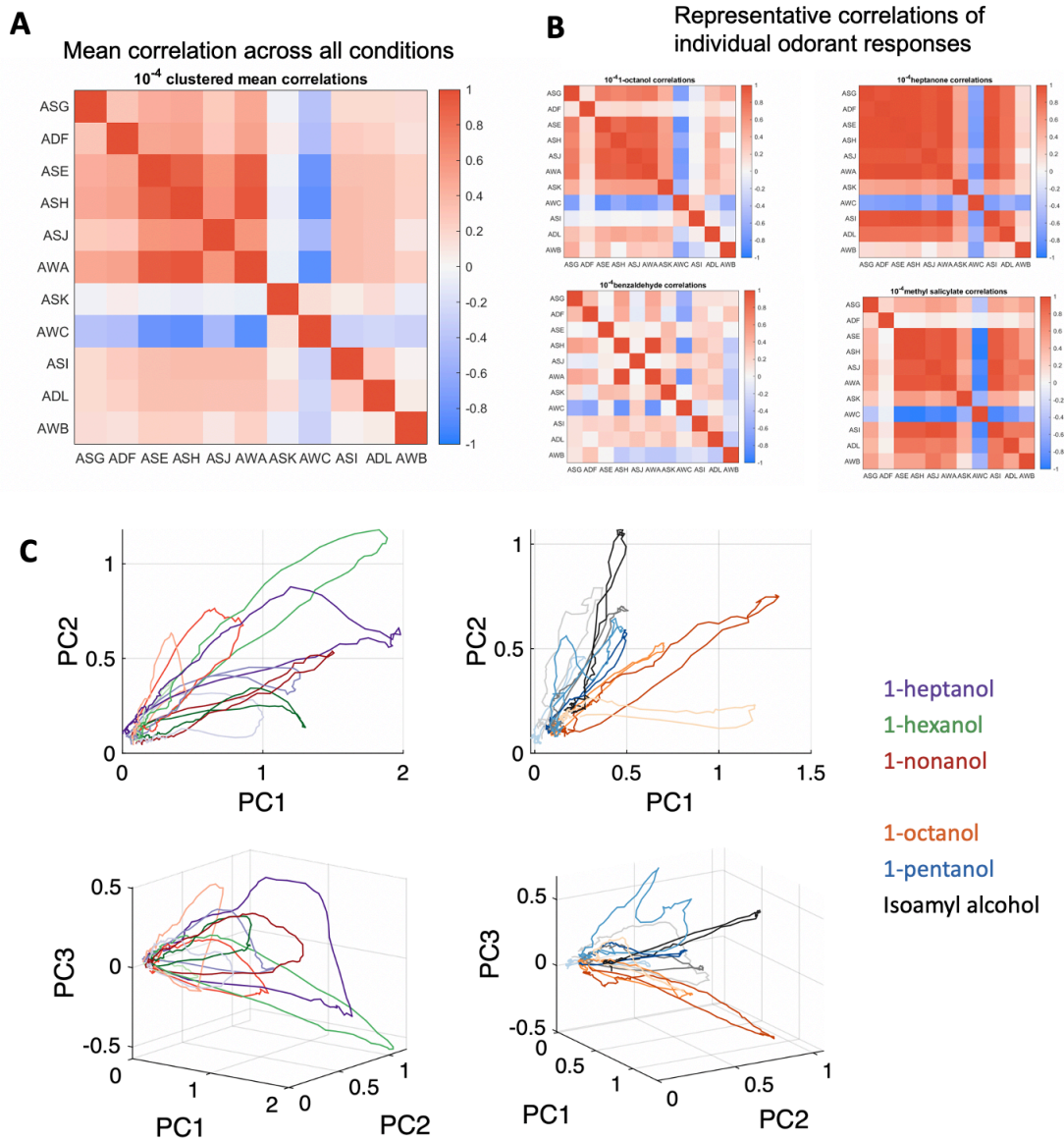
Figure C.5: Time trace correlations and phase trajectory analyses. (A) Average time trace correlation map of the 11 chemosensory neuron responses across all 23 odorants. **(B)** Selected average correlation maps of responses to four specific odorants (1-octanol, heptanone, benzaldehyde, and methyl salicylate), plotted on the same axes, show diverse response dynamics. **(C)** Phase trajectory plots of average neural activity for 1-heptanol, 1-hexanol, and 1-nonanol (left) and 1-octanol, 1-pentanol, and isoamyl alcohol (right). The shade of each color indicates concentration, with low concentration indicated by a light shade and high concentration indicated by a dark shade. Different concentrations of the same odorant tend to generate similar trajectories.
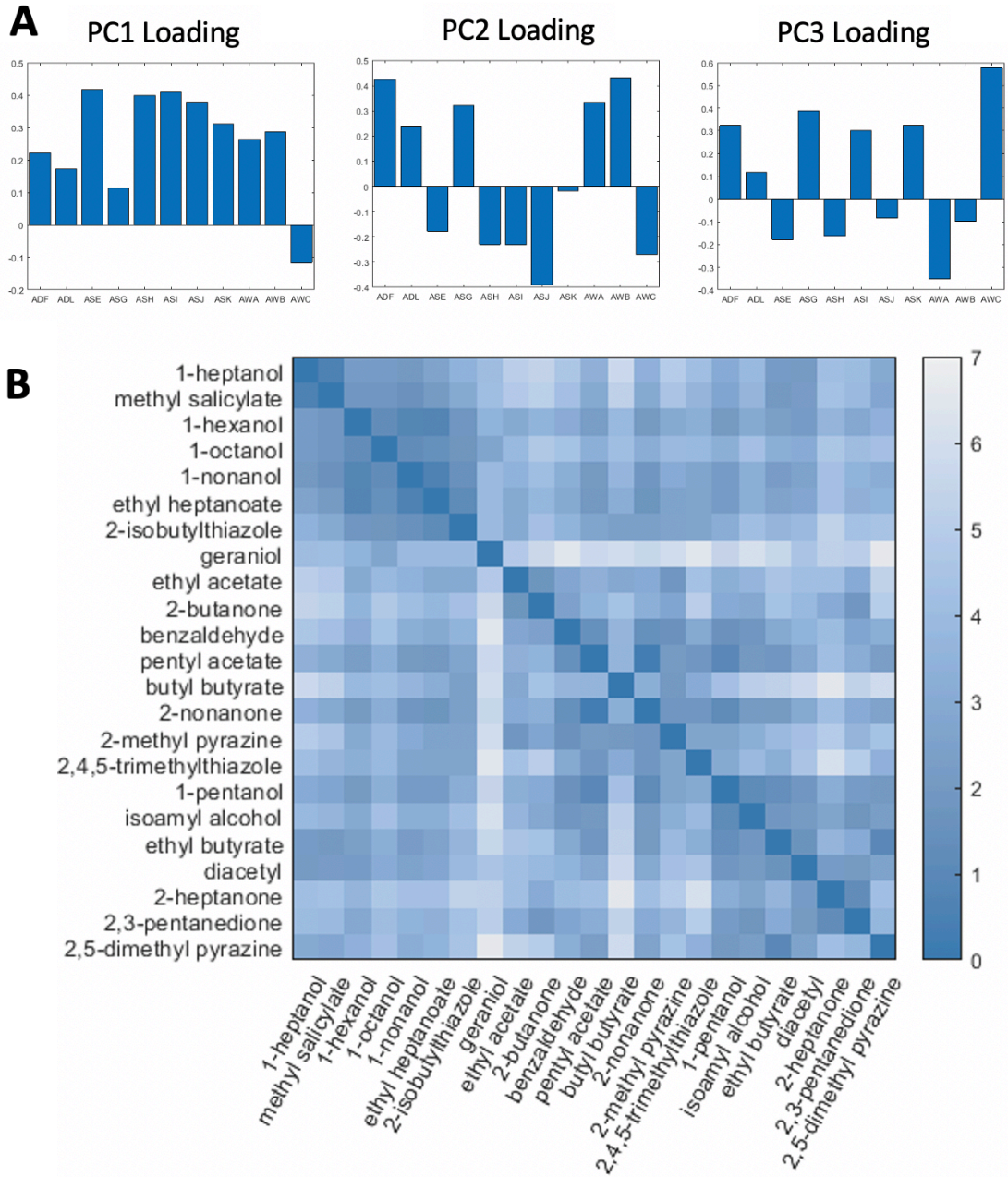
Figure C.6: Principal components of peak neural response space. (A) The loadings (in 11-dimensional neuron class space) of the first three principal components of standardized average peak neural response PC space in **Figure 4.3B. (B)** The pairwise distances between odorants in this PC space.

A Cumulative distribution of peak response to different odorants
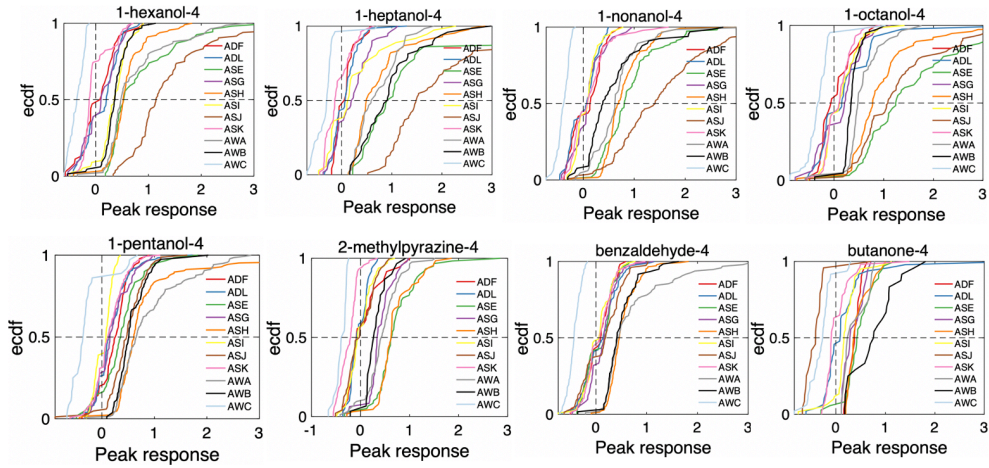
B Example of imputed peak responses for individual trial analysis
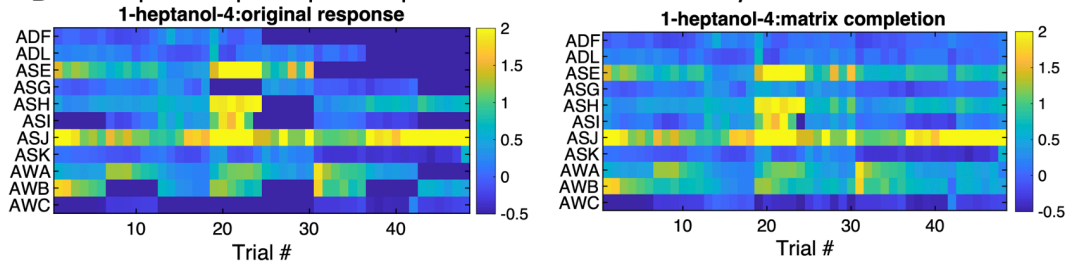
Figure C.7: Imputing missing data for single-trial analyses. Across trials of all neurons and all conditions, about 20% of the neuron responses were either not captured, or excluded for tracking or signal contamination issues. **(A)** To fill these missing data points when comparing single trials, we first determined the distribution of peak responses of every neuron to every odorant. **(B)** We randomly sampled these distributions to generate missing data points.
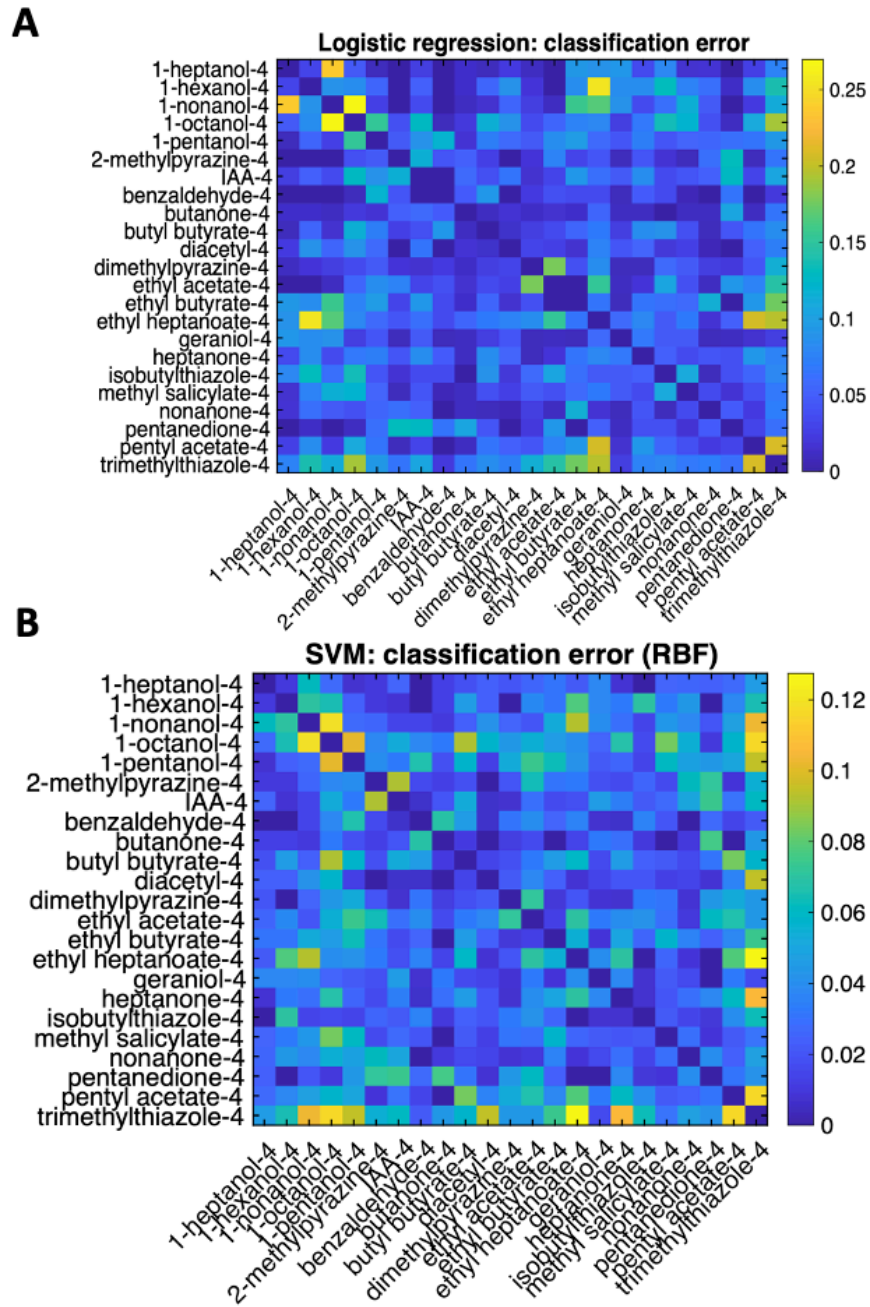
Figure C.8: Pairwise classification error of neural responses to odorants. Binary classification of all odorant pairs by **(A)** linear regression and **(B)** SVM. Both methods return very low classification errors, demonstrating that the single-trial peak responses of any two odorants are linearly separable.
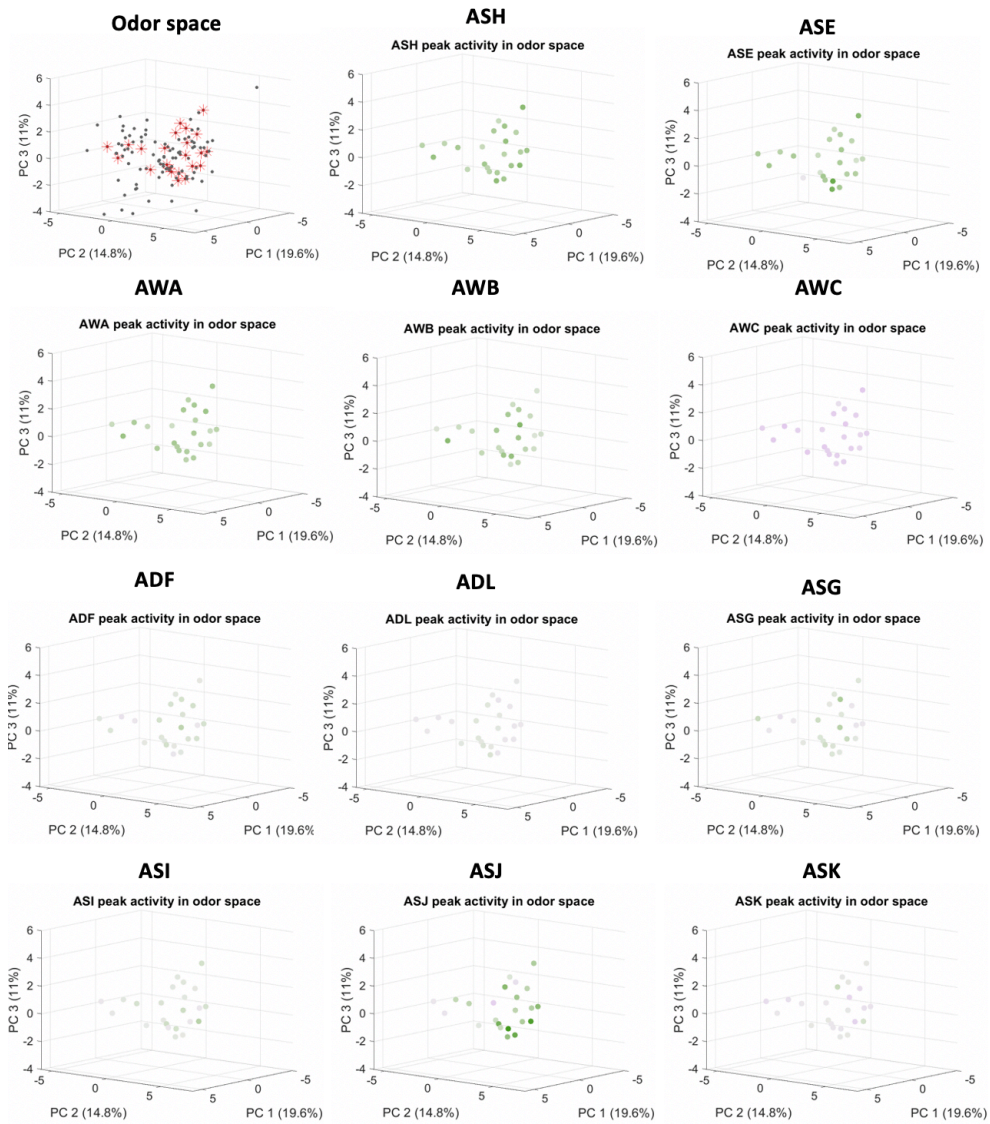
Figure C.9: Average peak responses plotted in odor space. Average peak responses of each of the 11 chemosensory neuron classes plotted in odor space (**Figure 4.1A**).
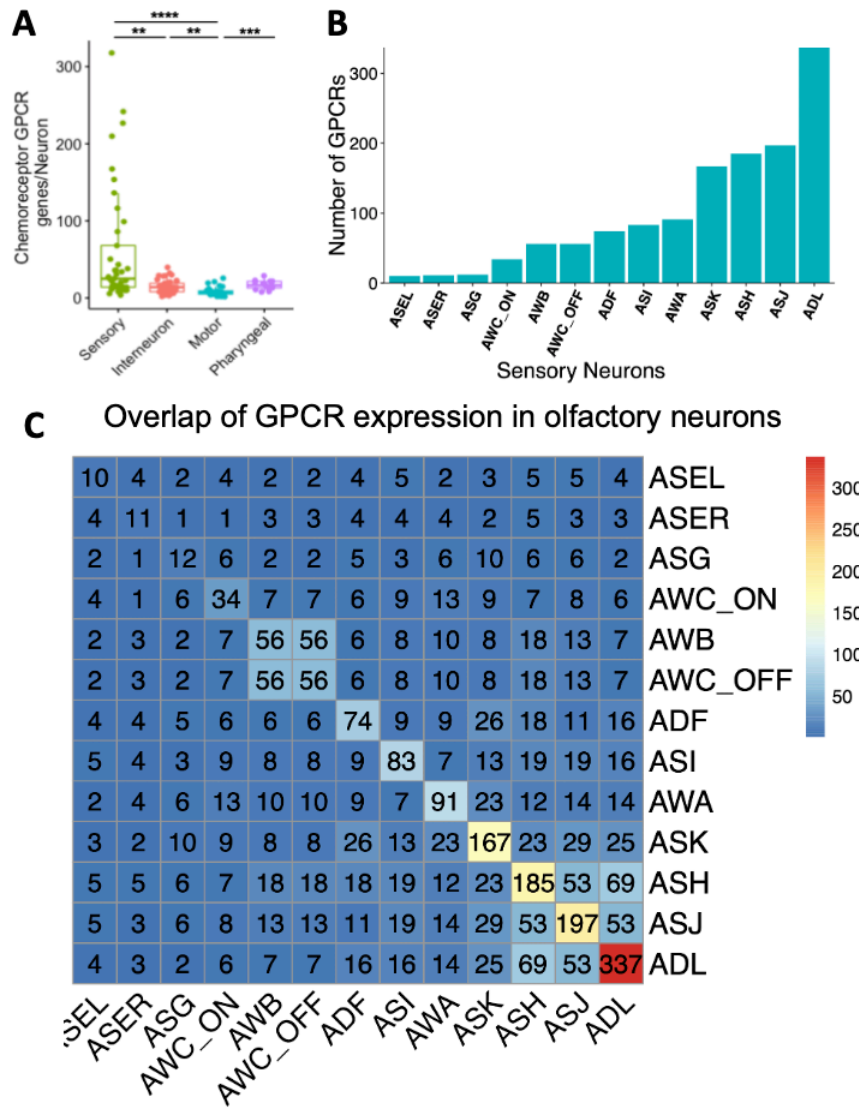
Figure C.10: **GPCR expression in the chemosensory neurons.** **(A)** Distribution of putative chemoreceptor GPCRs among the neuron classes (sensory, inter, motor, and pharyngeal neurons). **(B)** Number of GPCRs expressed in each of the 11 chemosensory neurons (AWC and ASE separated due to their asymmetries). **(C)** Number of overlapping GPCR types expressed in the chemosensory neurons. Data from Vidal et al., 2018.[80].

167

D

Supplemental materials for Chapter 5

D.1   Supplemental methods

D.1.1   Worm maintenance & phenotyping

All worms were raised at 20° C, on NGM plates, and fed OP50 *E. coli* as previously described[182], unless otherwise noted. Brood-size quantification, high-resolution behavioral phenotyping, dye-fill with DiO, chemotactic quadrant assays, and drop-test assays were performed using standard protocols[67,186,187,73,188].

D.1.2   Fluorophore, reporter, and mutant imaging

We imaged strains with a Zeiss LSM880, equipped with 7 laser lines: 405, 458, 488, 514, 561, 594, and 633 nm. Our standard configuration employed 405, 488, 561, and 633 nm to excite mTagBFP2, GFP/GCaMP + CyOFP1, TagRFP-T, and mNeptune2.5, respectively. The 8-color emission spectra was captured using strains that expressed each fluorophore individually. For these, we used the

LSM880's "lambda mode", employing its 32-channel spectral detector to capture color spectra from 391-727nm, at ~10nm color resolution – several fluorophores were imaged by exciting them with wavelengths below peak excitation and significantly increasing both the laser power and gain. To that end, for the 8-color emission spectra, we used: 405nm to excite mTagBFP, CFP, GFP, and Cy-OFP1; 488nm to excite YFP and mNeonGreen; and, 561nm to excite TagRFP-T and mNeptune2.5. All NeuroPAL reporter and mutant crosses were imaged with the same scope. When not performing a DIC overlay, gamma correction of ~0.5 was applied to images so as to improve color visibility. Occasionally, histograms were adjusted to balance colors for visibility. These image adjustments are necessary and suggested for NeuroPAL identification in order to deal with a variable range of GFP/CFP/YFP reporters and color alterations in mutant backgrounds.

D.1.3 STATISTICS, CODE, AND SOFTWARE

All statistics and code were run in MATLAB, using standard toolboxes, with the exception of the OME Bio-Formats API (used to read in Zeiss CZI and Nikon ND2 file formats)[185], and MathWorks FileExchange functions: munkres.m, and logsumexp.m. The code and software are freely available at: **https://github.com/amin-nejat/CELL_ID**, **https://github.com/Eviatar/Optimal_Coloring**, & **https://github.com/venkatachalamlab/NeuroPAL-traces/**.

D.2 DNMF SUPPLEMENTAL METHODS

The joint motion correction and signal extraction framework involves several steps. Let $\mathbf{Y}_t \in \mathbb{R}^d$ denote the d-pixel vectorized volumetric image at time $t = 1, \ldots, T$. We seek to decompose the

observations, $\mathbf{Y}_t$, into a factorization involving a time-varying deformation term, $\mathbf{f}_{\beta_t}$ that acts on a time-invariant canonical representation of k object shapes encoded by $\mathbf{A}$. The time-varying spatial signatures, $\mathbf{f}_{\beta_t}(\mathbf{A}) \in \mathbb{R}^{d \times k}$, are then multiplied by signal carrying coefficients $\mathbf{C}_t \in \mathbb{R}^k$. We also encourage model parameters to be "well-behaved" using regularization functions, $\mathcal{R}$. The resulting objective function is:

$$\min_{\mathbf{A},\mathbf{C},\beta} \sum_{t=1}^{T} \left\| \mathbf{Y}_t - \mathbf{f}_{\beta_t}(\mathbf{A})\mathbf{C}_t \right\|_2^2 + \mathcal{R}(\mathbf{A},\mathbf{C},\beta) \tag{D.1}$$

$$\text{s.t.} \mathbf{A}, \mathbf{C}_{1:T} \geq 0.$$

This formulation differs from standard NMF techniques[189] in that the spatial footprint term consists of a time invariant term, $\mathbf{A}$ and a time varying term, $\mathbf{f}_{\beta_t}$, which is a differentiable transformation parametrized by β_t, that deforms the canonical representation into the t-th time frame. β_t encapsulates the motion parameters and is usually low dimensional to avoid over-parameterization and overfitting. The regularization $\mathcal{R}(\cdot)$ further constrains the possible choice of spatial footprints, signal coefficients, and spatial deformations.

D.2.1 Spatial component parametrization

When we have strong prior information about the component shapes we can incorporate that into the model using an appropriate parameterization for the spatial footprints. Neural activity is most commonly imaged using cytosolic or nuclear-localized calcium indicators; nuclear-localized indi-

cators can be reasonably modelled using ellipsoidally-symmetric shape models. Specifically, we observed that the spatial component of the neurons in the videos analyzed here, of *C. elegans* imaged using nuclear-localized calcium indicators, can be well approximated using three-dimensional Gaussian functions. By taking advantage of this observation we can reduce the number of parameters in $\mathbf{A}$ from one parameter per pixel per component, to k 3D centers (3 parameters per each neuron) and k covariance matrices (6 parameters per each neuron using the Cholesky parameterization). Formally, we model the footprint of component k using a 3-dimensional Gaussian function with location parameters $\boldsymbol{\mu}_k \in \mathbb{R}^3$ and shape parameters $\boldsymbol{\Sigma}_k \in \mathbb{R}^{3 \times 3}$.

Under this new spatial model for $\mathbf{A} = \{\boldsymbol{\mu}_{1:K}, \boldsymbol{\Sigma}_{1:K}\}$, we modify the $\mathbf{f}_{\beta_t}$ function to match this parameterization to have $\mathbf{f}_{\beta_t}(\mathbf{A}) \in \mathbb{R}^{d \times k}$:

$$\mathbf{f}_{\beta_t}(\mathbf{A})[i, k] \approx \exp\left([\mathbf{p}_i - \beta_t \Psi(\boldsymbol{\mu}_k)]^T \boldsymbol{\Sigma}_k^{-1} [\mathbf{p}_i - \beta_t \Psi(\boldsymbol{\mu}_k)]\right), \tag{D.2}$$

where $\mathbf{p}_i$ is the 3D coordinate of the i-th pixel in the image. (Note that non-negativity of the spatial components is enforced automatically here.) Due to the differentiability of $\mathbf{f}_{\beta_t}$, it is straightforward to compute gradients with respect to β_t and $\boldsymbol{\Sigma}_k$.

D.2.2 Regularization: temporal continuity and plausible deformations

We employ two sets of regularizers to enforce a well-behaved model. To enforce smoothness of the temporal traces and motion trajectories in time we add a regularizer that penalizes discontinuities in the neural trajectories and signal coefficients. Specifically, we encourage the neural centers and signal

coefficients at neighboring time points to be close. The regularizer for this purpose is:

$$\mathcal{R}_T(\mathbf{C}, \boldsymbol{\beta}) = \lambda_\beta \sum_{t=0}^{T-1} \left\| \psi(\boldsymbol{\mu}_{1:K}) \boldsymbol{\beta}_{t-1} - \psi(\boldsymbol{\mu}_{1:K}) \boldsymbol{\beta}_t \right\|_F^2 + \lambda_C \sum_{t=0}^{T-1} \left\| \mathbf{C}_{t-1} - \mathbf{C}_t \right\|_F^2. \qquad \text{(D.3)}$$

In this formulation $\psi(\boldsymbol{\mu}_{1:K})$ is the quadratic transformation of the canonical neural centers.

When multiplied by $\boldsymbol{\beta}_{t-1}$ and $\boldsymbol{\beta}_t$ the result will be the neural centers at time $t-1$ and t respectively.

The term $\mathbf{f}_{\boldsymbol{\beta}_t}$ induces a deformable transformation of the pixel correspondences between time t and

the canonical representation $\mathbf{A}$.

In order to constrain this transformation to yield physically realistic deformations that respect

volumetric changes, we regularize the cost function using the determinant of the Jacobian of the

transformation term to encourage the Jacobian to be close to 1 and prevent the deformation from

contracting or expanding unrealistically. The Jacobian can be represented as: $\mathcal{J}_\beta(x_1, x_2, x_3)$, with

$\mathcal{J}_{ij} = \frac{\partial (\mathbf{f}_\beta)_i}{\partial x_j}$. Using the Jacobian, the regularizer is:

$$\mathcal{R}_{\mathcal{J}}(\boldsymbol{\beta}) = \lambda_{\mathcal{J}} \sum_{t=1}^{T} \sum_{i=1}^{j} \left(\det \mathcal{J}_{\beta_t}(x_i, y_i, z_i) - 1 \right)^2, \qquad \text{(D.4)}$$

where the Jacobian is evaluated on a grid.

D.2.3 OPTIMIZATION AND INITIALIZATION

All the variations of the dNMF cost function are optimized in the following way. To update β and $\mathbf{A}$ we use the `autograd` tool and `PyTorch` library to automatically compute gradients of the cost function and `Adam` optimizer to back-propagate the gradients. A forward pass of computation is evaluating the cost function with $\beta_{1:T}$ and $\mathbf{A}$ (in the fully parametric case, or $\beta_{1:T}$ (in the Gaussian case) as parameters. Note that for a fixed $\mathbf{C}$, all compartments of the cost function are differentiable with respect to the parameters.

To update $\mathbf{C}$ we use multiplicative updates as described in [116]:

$$\mathbf{C}_t \leftarrow \mathbf{C}_t \odot \frac{\mathbf{f}_{\beta_t}^T \mathbf{Y}_t + \lambda_C(\mathbf{C}_{t-1} + \mathbf{C}_{t+1})}{\mathbf{f}_{\beta_t}^T \mathbf{f}_{\beta_t} \mathbf{C}_t + 2\lambda_C \mathbf{C}_t}. \tag{D.5}$$

The key difference between these multiplicative updates from those found in [189] is that the parts of the derivatives of the temporal smoothness regularization terms $2\lambda_C \mathbf{C}_t$ and $\lambda_C(\mathbf{C}_{t-1} + \mathbf{C}_{t+1})$ appear in the denominator and numerator to promote smoothly varying signal.

One key advantage of the *C. elegans* datasets considered here is that we can reliably identify the locations of all cells in the field of view, using methods developed in [139]. Using the location of cells in the initial frame (for example) can tremendously aid the optimization of the objective D.1 for two main reasons. First, it serves as a very good initializer for the μ_k parameters for cell spatial footprints mentioned in section D.2.1. Second, we know a priori the correct number of cells to be demixed in the FOV. These two factors enable our framework to operate in a **semi-blind** manner towards the

deconvolution of neural signals of *C. elegans*, unlike fully blind deconvolution techniques such as e.g. PCA-ICA [190] or CNMF [191].

D.2.4 EVALUATION METRICS AND COMPARISONS

To evaluate the performance of the proposed method as well as the compared methods, we focus on several metrics that shed light both on the signal demixing capabilities of the methods as well as their ability to track objects in time. Namely we focus on two major metrics: **trajectory correlation**, which measures the ability of the deformation model to keep track of the observed motion, and **signal correlation**, which measures the demixing performance by comparing the correlation of demixed signal intensities relative to the ground truth. Specifically, these metrics can be expressed as:

Trajectory correlation:	**Signal correlation:**

$$\rho(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}) = \frac{\sum_{i,j,t}(\hat{\beta}_t^{ij} - \bar{\hat{\beta}})(\beta_t^{ij} - \bar{\beta})}{\sqrt{\sum_{i,j,t}(\hat{\beta}_t^{ij} - \bar{\hat{\beta}})^2}\sqrt{\sum_{i,j,t}(\beta_t^{ij} - \bar{\beta})^2}} \qquad \rho(\hat{\mathbf{C}}, \mathbf{C}) = \frac{\sum_{kt}(\hat{C}_{kt} - \bar{\hat{C}})(C_{kt} - \bar{C})}{\sqrt{\sum_{kt}(\hat{C}_{kt} - \bar{\hat{C}})^2}\sqrt{\sum_{kt}(C_{kt} - \bar{C})^2}}$$

Using these metrics, we evaluate the signal extraction performance of dNMF against two standard routines in calcium imaging. First, we compare against region of interest (ROI) tracking and pixel averaging within the ROI [117]. This method tracks the positions of cells across time and extracts signal by taking the average pixel intensity value in a pre-defined radial region around the tracking marker. We also compare against the routine of performing motion correction first and then signal extraction through NMF [191]. To replicate this routine in our experiments, we motion correct using Normcorre and then use the Gaussian cell shape parametrization version of NMF that is described in section D.2.1.
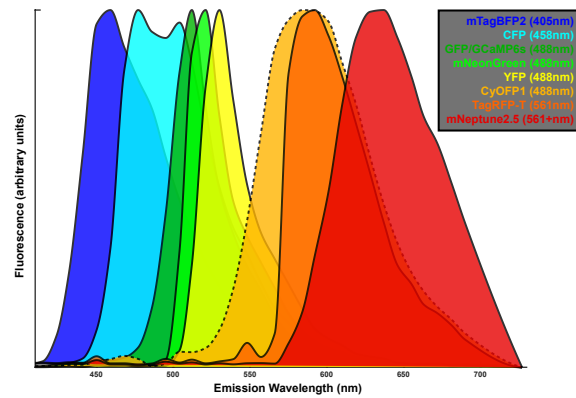
Figure D.1: Emission for NeuroPAL fluorophores and compatible signal fluorophores
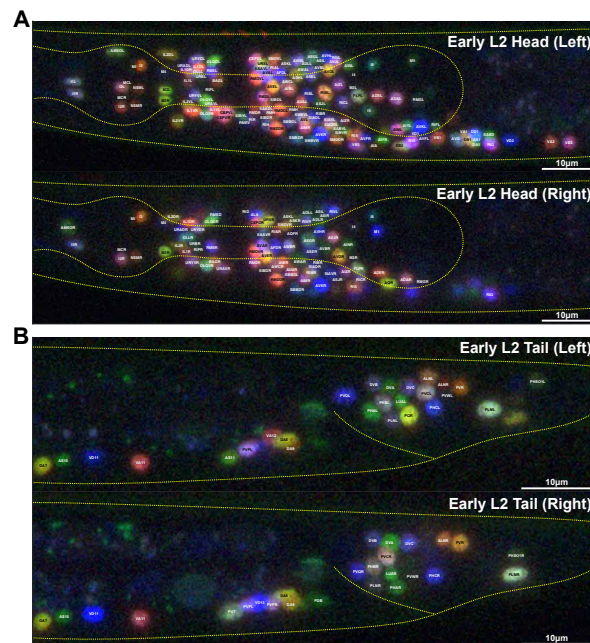
Figure D.2: NeuroPAL in the L2 larva. Neural identities for the head (**A**) and tail (**B**) of an early-L2, NeuroPAL worm. PDA, PVN, and RMF are not present since they are postembryonic neurons that differentiate later in development.
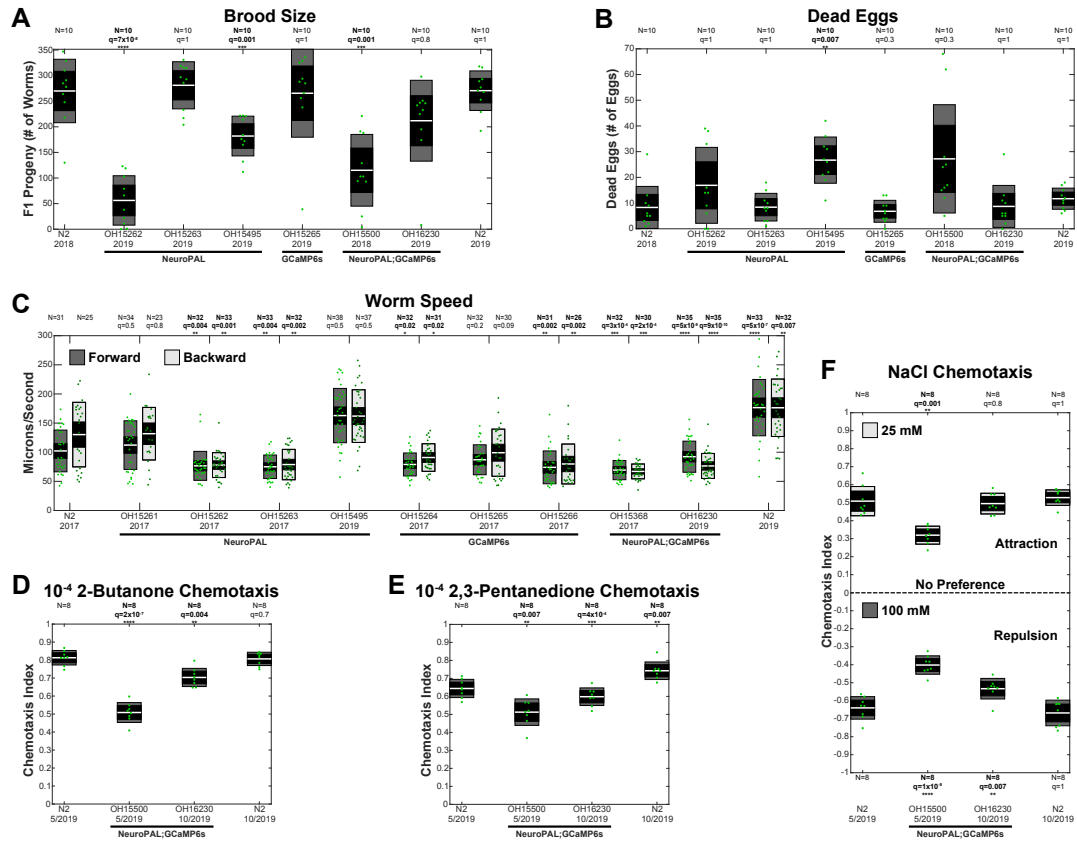
Figure D.3: NeuroPAL phenotype quantification. (A) Brood size. **(B)** Dead eggs. **C** Crawling speed. **(D-F)** Chemotaxis indices (displacement along gradient divided by distance traveled).
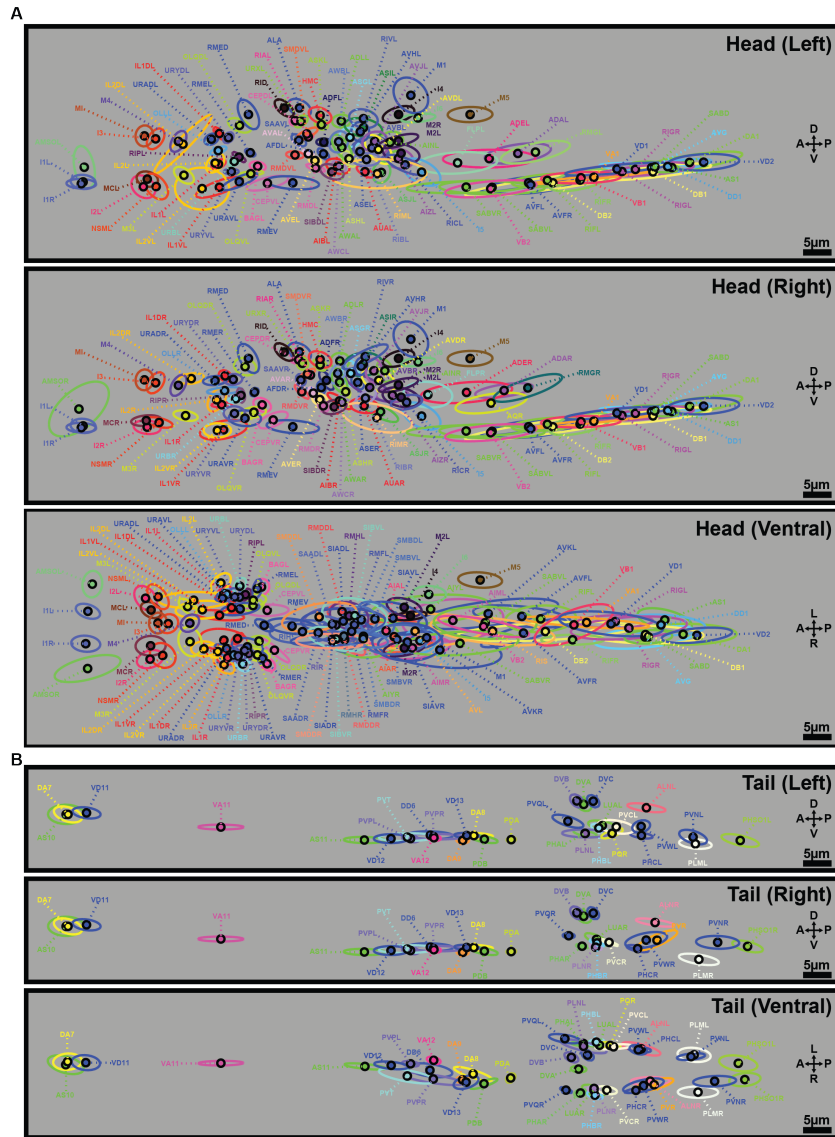
Figure D.4: Canonical neuron locations and their positional variability. Canonical neuron locations (filled circles with their NeuroPAL coloring) alongside their positional variability (encircling ellipses with matching color) for all ganglia in the head (**A**) and tail (**B**), as determined by NeuroPAL (*otIs669*), see **Supplement Data** for exact measurements. Positional variability is displayed as the 50% contour for neuronal location (measured as a Gaussian density distribution), sliced within a 2D plane; because we are restricted to a planar view, we show both the left-right and dorsal-ventral planes to provide a 3D estimation of the true contour bounding positional variability (**Supplement Text**). Left, right, and ventral views of neuron position variability is shown.
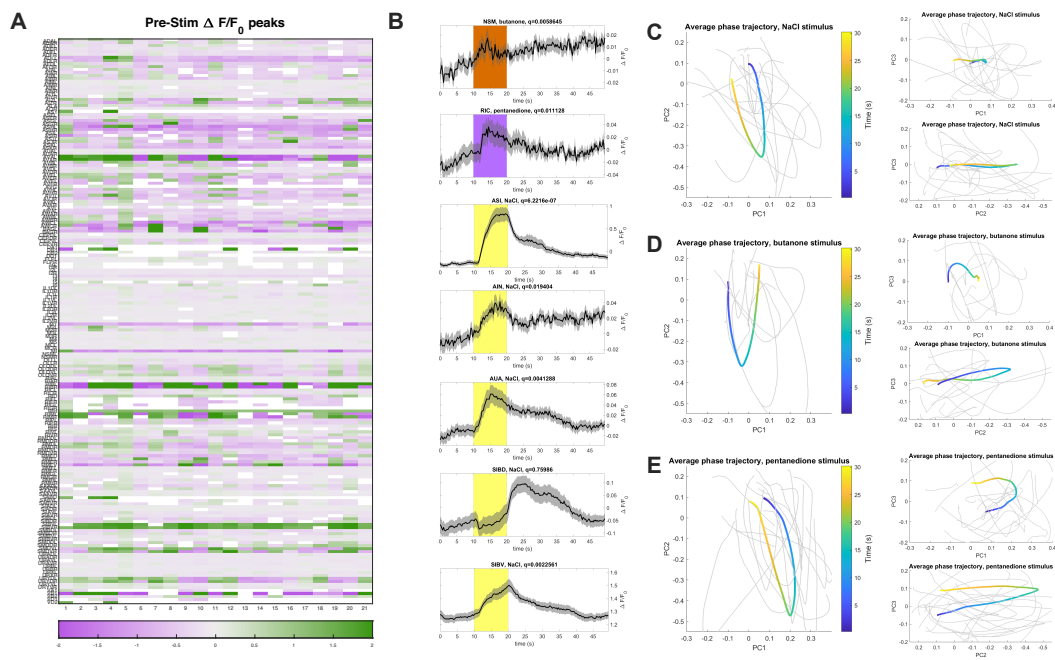
Figure D.5: Pre-stimulus calcium activity and phase trajectory analysis. (A) Peak $\Delta F/F_0$ values for neurons in 21 individuals in the minute prior to stimulation. White boxes indicate neurons which were not captured in that animal. **(B)** Additional neurons responsive to chemical stimulation. Average phase trajectories of neuronal representation during and immediately after **(C)** NaCl, **(D)** 2-butanone, and **(E)** 2,3-pentanedione, with selected individual phase trajectories shown in gray. The PC space is constructed from data from all individuals pooled together, with missing neurons imputed.
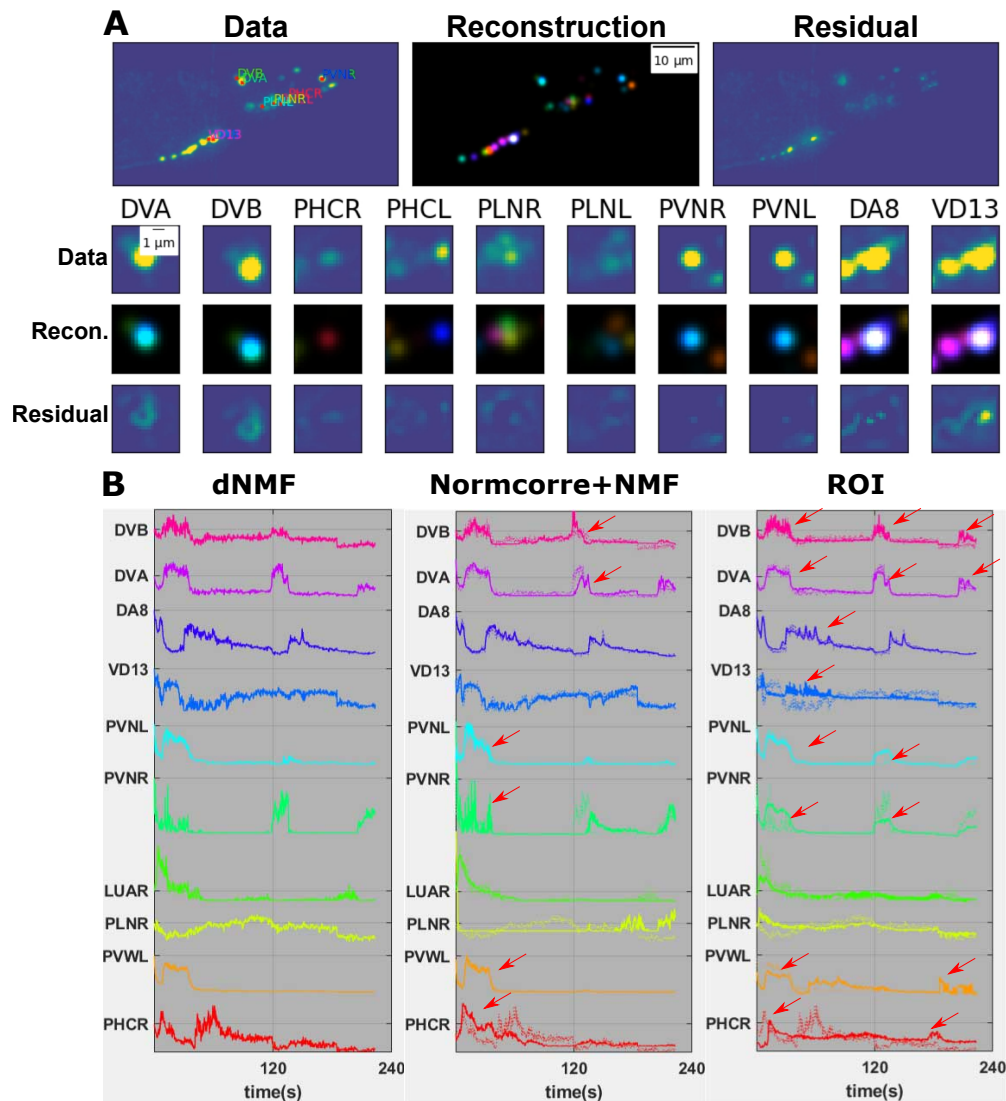
Figure D.6: Demixing neural calcium signals in semi-immobilized animals. (A) A representative z-axis maximum projected frame from a 4-minute long video of GCaMP6s neural activity in the tail of a pan-neuronally labeled animal. We focus on the signal from five pairs of spatially-neighboring neurons in the tail: DVA/DVB, PVNR/PVNL, PVWL/PHCR, PLNR/LUAR, and VD13/DA8. **(B)** Calcium signals extracted by dNMF (left), Normcorre[192] + NMF (middle), and ROI tracking and averaging (right). dNMF extracts uncoupled signals that demonstrate independent neural activity. The selected cells were chosen such that the signal recovered by ROI averaging is inconsistent with dNMF (quantified by having correlation smaller than 0.4). Normcorre + NMF partially mixes signals between both PHCR/PVWL and PVNL/PVNR around the 30-second mark and DVB/DVA around the 120-second mark (red arrows), and loses nearly all signal from PLNR, due to motion exhibited by the semi-immobilized animal. ROI averaging produces completely correlated signal (red arrows) between all of the labeled neurons, and loses most of the signal from LUAR and PLNR, due to overlap in their spatial footprints.

Table 1

D.4.1 NEUROPAL REPORTERS AND NEURON COLORS

The 41 reporters combined to create the NeuroPAL, their neural expression, and the resulting stereotyped color, per neuron class, ordered from anterior to posterior. NeuroPAL not only labels all 118 anatomically defined classes of the worm, but also a number of neuronal subclasses. For example, the stochastically-asymmetric, bilateral, AWC neuron pair were distinguishably colored by design. AWC-OFF expresses *srsx-3::CyOFP1*, thus giving it a green appearance relative to its ON equivalent. Similar design choices were made so as distinguish several neural subclasses, such as DA8 versus DA9 as well as VB1 versus VB2. Neuronal subtypes with differential coloring are listed separately (e.g., RMEL/R vs. RMED/V).

a = the neuron is missing expected coloring from at least one of the NeuroPAL reporters.

b = the neuron displays coloring that is not accounted for by any of the NeuroPAL reporters.

c = the neuron may appear either bright or dim in the NeuroPAL.

d = the neuron is solely marked by the panneuronal but may occasionally display dim blue coloring.

References

[1] A. K. Corsi, B. Wightman, and M. Chalfie, "A Transparent window into biology: A primer on Caenorhabditis elegans," *WormBook : the online review of C. elegans biology*, pp. 1–31, 2015.

[2] T. C. elegans sequencing Consortium, "Genome sequence of the nematode C. elegans: A platform for investigating biology," *Science (80-.).*, vol. 282, no. 5396, pp. 2012–2018, 1998.

[3] J. E. Sulston and H. R. Horvitz, "Post-embryonic cell lineages of the nematode, Caenorhabditis elegans," *Dev. Biol.*, vol. 56, no. 1, pp. 110–156, 1977.

[4] J. Kimble and D. Hirsh, "The postembryonic cell lineages of the hermaphrodite and male gonads in Caenorhabditis elegans," *Dev. Biol.*, vol. 70, no. 2, pp. 396–417, 1979.

[5] J. E. Sulston, "Neuronal cell lineages in the nematode Caenorhabditis elegans.," *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 48 Pt 2, pp. 443–452, 1983.

[6] M. M. Barr and L. R. Garcia, "Male mating behavior.," *WormBook : the online review of C. elegans biology*, pp. 1–11, 2006.

[7] R. J. Hobson, K. J. Yook, and E. M. Jorgensen, *Genetics of Neurotransmitter Release in Caenorhabditis elegans*□. Reference Module in Life Sciences, 2017.

[8] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner, "The structure of the nervous system of the nematode Caenorhabditis elegans," *Philosophical Transactions of the Royal Society of London B*, vol. 314, pp. 1–340, 1986.

[9] D. Witvliet, B. Mulcahy, J. K. Mitchell, Y. Meirovitch, D. K. Berger, Y. Wu, Y. Liu, W. X. Koh, R. Parvathala, D. Holmyard, *et al.*, "Connectomes across development reveal principles of brain maturation in C. elegans," *bioRxiv*, 2020.

[10] C. Li and K. Kim, "Neuropeptides.," *WormBook*, no. 212, pp. 1–36, 2008.

[11] M. A. Jobson, C. M. Valdez, J. Gardner, L. R. Garcia, E. M. Jorgensen, and A. A. Beg, "Spillover transmission is mediated by the excitatory GABA receptor LGC-35 in C. elegans," *Journal of Neuroscience*, vol. 35, no. 6, pp. 2803–2816, 2015.

[12] B. A. Bamber, A. A. Beg, R. E. Twyman, and E. M. Jorgensen, "The Caenorhabditis elegans unc-49 Locus Encodes Multiple Subunits of a Heteromultimeric GABA Receptor," *The Journal of Neuroscience*, vol. 19, pp. 5348–, 1999.

[13] M. Gendrel, E. G. Atlas, and O. Hobert, "A cellular and regulatory map of the GABAergic nervous system of C. elegans," *eLife*, vol. 5, 2016.

[14] L. Pereira, P. Kratsios, E. Serrano-Saiz, H. Sheftel, A. E. Mayo, D. H. Hall, J. G. White, B. LeBoeuf, L. R. Garcia, U. Alon, and O. Hobert, "A cellular and regulatory map of the cholinergic nervous system of C. Elegans," *Elife*, vol. 4, no. DECEMBER2015, pp. 1–46, 2015.

[15] E. Serrano-Saiz, R. J. Poole, T. Felton, F. Zhang, E. D. De La Cruz, and O. Hobert, "Modular control of glutamatergic neuronal identity in C. elegans by distinct homeodomain proteins," *Cell*, vol. 155, p. 659, oct 2013.

[16] C. I. Bargmann, "Neurobiology of the Caenorhabditis elegans genome," *Science (80-.).*, vol. 282, no. 5396, pp. 2028–2033, 1998.

[17] J. E. Richmond, W. S. Davis, and E. M. Jorgensen, "Unc-13 is required for synaptic vesicle fusion for neurotransmission," pp. 959–964, 1999.

[18] J. E. Richmond and E. M. Jorgensen, "One GABA and two acetylcholine receptors function at the C. elegans neuromuscular junction," *Nat. Neurosci.*, vol. 2, no. 9, pp. 791–798, 1999.

[19] M. B. Goodman, D. H. Hall, L. Avery, and S. R. Lockery, "Active currents regulate sensitivity and dynamic range in C. elegans Neurons," *Neuron*, vol. 20, no. 4, pp. 763–772, 1998.

[20] S. R. Lockery and M. B. Goodman, "The quest for action potentials in C. elegans neurons hits a plateau," *Nat. Neurosci.*, vol. 12, no. 4, pp. 377–378, 2009.

[21] O. Hobert, "The neuronal genome of Caenorhabditis elegans," *WormBook : the online review of C. elegans biology*, pp. 1–106, aug 2013.

[22] S. Gao and M. Zhen, "Action potentials drive body wall muscle contractions in Caenorhabditis elegans," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 6, pp. 2557–2562, 2011.

[23] J. E. Mellem, P. J. Brockie, D. M. Madsen, and A. V. Maricq, "Action potentials contribute to neuronal signaling in C. elegans," *Nat. Neurosci.*, vol. 11, no. 8, pp. 865–867, 2008.

[24] Q. Liu, P. B. Kidd, M. Dobosiewicz, and C. I. Bargmann, "C. elegans AWA Olfactory Neurons Fire Calcium-Mediated All-or-None Action Potentials," *Cell*, vol. 175, pp. 57–70.e17, sep 2018.

[25] M. Hendricks, H. Ha, N. Maffey, and Y. Zhang, "Compartmentalized calcium dynamics in a C. elegans interneuron encode head movement," *Nature*, vol. 487, no. 7405, pp. 99–103, 2012.

[26] M. H. Ouellette, M. J. Desrochers, I. Gheta, R. Ramos, and M. Hendricks, "A gate-and-switch model for head orientation behaviors in C. elegans," *bioRxiv*, vol. 2014, no. December, pp. 1–13, 2018.

[27] A. Cuentas-Condori, B. Mulcahy, S. He, S. Palumbos, M. Zhen, and D. M. Miller, "C. Elegans neurons have functional dendritic spines," *Elife*, vol. 8, pp. 1–23, 2019.

[28] O. J. Mullins, J. T. Hackett, J. T. Buchanan, and W. O. Friesen, "Neuronal control of swimming behavior: Comparison of vertebrate and invertebrate model systems," *Prog. Neurobiol.*, vol. 93, no. 2, pp. 244–269, 2011.

[29] O. Kiehn, "Development and functional organization of spinal locomotor circuits," *Curr. Opin. Neurobiol.*, vol. 21, no. 1, pp. 100–109, 2011.

[30] I. A. Rybak, K. J. Dougherty, and N. A. Shevtsova, "Organization of the mammalian locomotor CPG: Review of computational model and circuit architectures based on genetically identified spinal interneurons," *eNeuro*, vol. 2, no. 5, 2015.

[31] M. Falgairolle, J. G. Puhl, A. Pujala, W. Liu, and M. J. O'Donovan, "Motoneurons regulate the central pattern generator during drug-induced locomotor-like activity in the neonatal mouse," *Elife*, vol. 6, pp. 1–29, 2017.

[32] M. Chalfie, J. E. Sulston, J. G. White, E. Southgate, J. Nichol Thomson, and S. Brenner, "The neural circuit for touch sensitivity in Caenorhabditis elegans," *J. Neurosci.*, vol. 5, no. 4, pp. 956–964, 1985.

[33] Y. Zheng, P. J. Brockie, J. E. Mellem, D. M. Madsen, and A. V. Maricq, "<1-s2.0-S0896627300808491-main-1.pdf>," vol. 24, pp. 347–361, 1999.

[34] N. Cohen and T. Sanders, "Nematode locomotion: Dissecting the neuronal-environmental loop," *Curr. Opin. Neurobiol.*, vol. 25, pp. 99–106, 2014.

[35] M. Zhen and A. D. Samuel, "C. elegans locomotion: small circuits, complex functions.," *Curr. Opin. Neurobiol.*, vol. 33, pp. 117–126, 2015.

[36] L. R. Varshney, B. L. Chen, E. Paniagua, D. H. Hall, and D. B. Chklovskii, "Structural properties of the Caenorhabditis elegans neuronal network," *PLoS Comput. Biol.*, vol. 7, no. 2, 2011.

[37] S. J. Cook, T. A. Jarrell, C. A. Brittin, Y. Wang, A. E. Bloniarz, M. A. Yakovlev, K. C. Nguyen, L. T. Tang, E. A. Bayer, J. S. Duerr, H. E. Bülow, O. Hobert, D. H. Hall, and S. W. Emmons, "Whole-animal connectomes of both Caenorhabditis elegans sexes," *Nature*, vol. 571, pp. 63–71, jul 2019.

[38] L. K. Scheffer, C. S. Xu, M. Januszewski, Z. Lu, S. Y. Takemura, K. J. Hayworth, G. B. Huang, K. Shinomiya, J. Maitin-Shepard, S. Berg, J. Clements, P. M. Hubbard, W. T. Katz, L. Umayam, T. Zhao, D. Ackerman, T. Blakely, J. Bogovic, T. Dolafi, D. Kainmueller, T. Kawase, K. A. Khairy, L. Leavitt, P. H. Li, L. Lindsey, N. Neubarth, D. J. Olbris, H. Otsuna, E. T. Trautman, M. Ito, A. S. Bates, J. Goldammer, T. Wolff, R. Svirskas, P. Schlegel, E. R. Neace, C. J. Knecht, C. X. Alvarado, D. A. Bailey, S. Ballinger, J. A. Borycz, B. S. Canino, N. Cheatham, M. Cook, M. Dreher, O. Duclos, B. Eubanks, K. Fairbanks, S. Finley, N. Forknall, A. Francis, G. P. Hopkins, E. M. Joyce, S. Kim, N. A. Kirk, J. Kovalyak, S. A. Lauchie, A. Lohff, C. Maldonado, E. A. Manley, S. McLin, C. Mooney, M. Ndama, O. Ogundeyi, N. Okeoma, C. Ordish, N. Padilla, C. Patrick, T. Paterson, E. E. Phillips, E. M. Phillips, N. Rampally, C. Ribeiro, M. K. Robertson, J. T. Rymer, S. M. Ryan, M. Sammons, A. K. Scott, A. L. Scott, A. Shinomiya, C. Smith, K. Smith, N. L. Smith, M. A. Sobeski, A. Suleiman, J. Swift, S. Takemura, I. Talebi, D. Tarnogorska, E. Tenshaw, T. Tokhi, J. J. Walsh, T. Yang, J. A. Horne, F. Li, R. Parekh, P. K. Rivlin, V. Jayaraman, M. Costa, G. S. Jefferis, K. Ito, S. Saalfeld, R. George, I. A. Meinertzhagen, G. M. Rubin, H. F. Hess, V. Jain, and S. M. Plaza, "A connectome and analysis of the adult drosophila central brain," *eLife*, vol. 9, pp. 1–74, 2020.

[39] Z. Zheng, J. S. Lauritzen, E. Perlman, C. G. Robinson, M. Nichols, D. Milkie, O. Torrens, J. Price, C. B. Fisher, N. Sharifi, S. A. Calle-Schuler, L. Kmecova, I. J. Ali, B. Karsh, E. T. Trautman, B. J. A, P. Hanslovsky, G. S. X. E. Jefferis, M. Kazhdan, K. Khairy, S. Saalfeld,

R. D. Fetter, and D. D. Bock, "A Complete Electron Microscopy Volume of the Brain of Adult Drosophila melanogaster," *Cell*, vol. 174, p. 730–743.e22, 7 2018.

[40] D. G. C. Hildebrand, M. Cicconet, R. M. Torres, W. Choi, T. M. Quan, J. Moon, A. W. Wetzel, A. Scott Champion, B. J. Graham, O. Randlett, P. G. Scott, R. Portugues, I. H. Bianco, S. Saalfeld, A. D. Baden, K. Lillaney, R. Burns, J. T. Vogelstein, S. A. Franz, W.-C. A. Lee, W.-K. Jeong, J. W. Lichtman, and F. Engert, "Whole-brain serial-section electron microscopy in larval zebrafish," *Nature*, vol. 545, pp. 345–349, 5 2017.

[41] Z. V. Guo, A. C. Hart, and S. Ramanathan, "Optical interrogation of neural circuits in Caenorhabditis elegans," *Nat. Methods*, vol. 6, no. 12, pp. 891–896, 2009.

[42] T. Kawano, M. D. Po, S. Gao, G. Leung, W. S. Ryu, and M. Zhen, "An imbalancing act: Gap junctions reduce the backward motor circuit activity to bias C. elegans for forward locomotion," *Neuron*, vol. 72, no. 4, pp. 572–586, 2011.

[43] Q. Wen, M. D. Po, E. Hulme, S. Chen, X. Liu, S. W. Kwok, M. Gershow, A. M. Leifer, V. Butler, C. Fang-Yen, T. Kawano, W. R. Schafer, G. M. Whitesides, M. Wyart, D. B. Chklovskii, M. Zhen, and A. D. Samuel, "Proprioceptive coupling within motor neurons drives C. elegans forward locomotion.," *Neuron*, vol. 76, pp. 750–61, nov 2012.

[44] C. Fang-Yen, M. J. Alkema, and A. D. Samuel, "Illuminating neural circuits and behaviour in caenorhabditis elegans with optogenetics," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 370, no. 1677, 2015.

[45] J. G. Hildebrand and G. M. Shepherd, "Mechanisms of olfactory discrimination: Converging evidence for common principles across phyla," *Annu. Rev. Neurosci.*, vol. 20, pp. 595–631, 1997.

[46] B. W. Ache and J. M. Young, "Olfaction: Diverse species, conserved principles," *Neuron*, vol. 48, no. 3, pp. 417–430, 2005.

[47] L. M. Kay and M. Stopfer, "Information processing in the olfactory systems of insects and vertebrates," *Semin. Cell Dev. Biol.*, vol. 17, no. 4, pp. 433–442, 2006.

[48] C. Y. Su, K. Menuz, J. Reisert, and J. R. Carlson, "Non-synaptic inhibition between grouped neurons in an olfactory circuit," *Nature*, vol. 492, no. 7427, pp. 66–71, 2012.

[49] A. Ramaekers, E. Magnenat, E. C. Marin, N. Gendre, G. S. Jefferis, L. Luo, and R. F. Stocker, "Glomerular maps without cellular redundancy at successive levels of the Drosophila larval olfactory circuit," *Curr. Biol.*, vol. 15, no. 11, pp. 982–992, 2005.

[50] L. B. Vosshall and R. F. Stocker, "Molecular architecture of smell and taste in Drosophila," *Annu. Rev. Neurosci.*, vol. 30, pp. 505–533, 2007.

[51] C. Y. Su, K. Menuz, and J. R. Carlson, "Olfactory Perception: Receptors, Cells, and Circuits," *Cell*, vol. 139, no. 1, pp. 45–59, 2009.

[52] B. Malnic, J. Hirono, T. Sato, and L. B. Buck, "Combinatorial receptor codes for odors.," *Cell*, vol. 96, no. 5, pp. 713–23, 1999.

[53] M. De Bruyne and T. C. Baker, "Odor detection in insects: Volatile codes," *J. Chem. Ecol.*, vol. 34, no. 7, pp. 882–897, 2008.

[54] S. A. Kreher, D. Mathew, J. Kim, and J. R. Carlson, "Translation of Sensory Input into Behavioral Output via an Olfactory System," *Neuron*, vol. 59, no. 1, pp. 110–124, 2008.

[55] G. Si, J. K. Kanwal, Y. Hu, C. J. Tabone, J. Baron, M. Berck, G. Vignoud, and A. D. Samuel, "Structured Odorant Response Patterns across a Complete Olfactory Receptor Neuron Population," *Neuron*, vol. 101, pp. 950–962.e7, mar 2019.

[56] J. W. Wang, A. M. Wong, J. Flores, L. B. Vosshall, and R. Axel, "Two-photon calcium imaging reveals an odor-evoked map of activity in the fly brain," *Cell*, vol. 112, no. 2, pp. 271–282, 2003.

[57] E. A. Hallem and J. R. Carlson, "Coding of Odors by a Receptor Repertoire," *Cell*, vol. 125, no. 1, pp. 143–160, 2006.

[58] K. Nara, L. R. Saraiva, X. Ye, and L. B. Buck, "A large-scale analysis of odor coding in the olfactory epithelium," *J. Neurosci.*, vol. 31, no. 25, pp. 9179–9191, 2011.

[59] S. A. Ebrahim, H. K. Dweck, J. Stökl, J. E. Hofferberth, F. Trona, K. Weniger, J. Rybak, Y. Seki, M. C. Stensmyr, S. Sachse, B. S. Hansson, and M. Knaden, "Drosophila Avoids Parasitoids by Sensing Their Semiochemicals via a Dedicated Olfactory Circuit," *PLoS Biol.*, vol. 13, no. 12, pp. 1–18, 2015.

[60] K. Kajiya, K. Inaki, M. Tanaka, T. Haga, H. Kataoka, and K. Touhara, "Molecular bases of odor discrimination: Reconstitution of olfactory receptors that recognize overlapping sets of odorants," *J. Neurosci.*, vol. 21, no. 16, pp. 6018–6025, 2001.

[61] H. Spors, M. Wachowiak, L. B. Cohen, and R. W. Friedrich, "Temporal dynamics and latency patterns of receptor neuron input to the olfactory bulb," *J. Neurosci.*, vol. 26, no. 4, pp. 1247–1259, 2006.

[62] R. F. Stocker, M. C. Lienhard, A. Borst, and K. F. Fischbach, "Neuronal architecture of the antennal lobe in Drosophila melanogaster," *Cell Tissue Res.*, vol. 262, no. 1, pp. 9–34, 1990.

[63] D. L. Sosulski, M. Lissitsyna Bloom, T. Cutforth, R. Axel, and S. R. Datta, "Distinct representations of olfactory information in different cortical centres," *Nature*, vol. 472, no. 7342, pp. 213–219, 2011.

[64] S. Keshavarzi, J. M. Power, E. H. Albers, R. K. Sullivan, and P. Sah, "Dendritic organization of olfactory inputs to medial amygdala neurons," *J. Neurosci.*, vol. 35, no. 38, pp. 13020–13028, 2015.

[65] K. A. Bolding and K. M. Franks, "Recurrent cortical circuits implement concentration-invariant odor coding," *Science (80-.).*, vol. 361, no. 6407, 2018.

[66] P. Grewal and D. Wright, "Migration of Caenorhabditis elegans larvae towards bacteria and the nature of the bacterial stimulus," *Fundam. Appl. Nematol.*, no. 15, p. 159–166, 1992.

[67] C. I. Bargmann, E. Hartwieg, and H. Robert Horvitz, "Odorant-Selective Genes and Neurons Mediate Olfaction in C. elegans," tech. rep., 1993.

[68] S. E. Worthy, L. Haynes, M. Chambers, D. Bethune, E. Kan, K. Chung, R. Ota, C. J. Taylor, and E. E. Glater, "Identification of attractive odorants released by preferred bacterial food found in the natural habitats of C. elegans," *PLoS One*, vol. 13, no. 7, pp. 1–14, 2018.

[69] P. Sengupta, J. H. Chou, and C. I. Bargmann, "odr-10 Encodes a seven transmembrane domain olfactory receptor required for responses to the odorant diacetyl," *Cell*, vol. 84, no. 6, pp. 899–909, 1996.

[70] E. R. Troemel, B. E. Kimmel, and C. I. Bargmann, "Reprogramming chemotaxis responses: Sensory neurons define olfactory preferences in C. elegans," *Cell*, vol. 91, no. 2, pp. 161–169, 1997.

[71] J. T. Pierce-Shimomura, S. Faumont, M. R. Gaston, B. J. Pearson, and S. R. Lockery, "The homeobox gene lim-6 is required for distinct chemosensory representations in C. elegans," *Nature*, vol. 410, no. 6829, pp. 694–698, 2001.

[72] C. I. Bargmann and H. Robert Horvitz, "Chemosensory Neurons with Overlapping Functions Direct Chemotaxis to Multiple Chemicals in C. elegans," tech. rep., 1991.

[73] M. A. Hilliard, C. I. Bargmann, and P. Bazzicalupo, "C. elegans responds to chemical repellents by integrating sensory inputs from the head and the tail," Tech. Rep. 9, 2002.

[74] M. A. Hilliard, C. Bergamasco, S. Arbucci, R. H. Plasterk, and P. Bazzicalupo, "Worms taste bitter: ASH neurons, QUI-1, GPA-3 and ODR-3 mediate quinine avoidance in Caenorhabditis elegans," *EMBO J.*, vol. 23, no. 5, pp. 1101–1111, 2004.

[75] J. M. Kaplan and H. R. Horvitz, "A dual mechanosensory and chemosensory neuron in Caenorhabditis elegans," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 90, no. 6, pp. 2227–2231, 1993.

[76] Y. Sambongi, T. Nagae, Y. Liu, T. Yoshimizu, K. Takeda, Y. Wada, and M. Futai, "Sensing of cadmium and copper ions by externally exposed ADL, ASE, and ASH neurons elicits avoidance response in Caenorhabditis elegans," *Neuroreport*, vol. 10, no. 4, pp. 753–757, 1999.

[77] H. C. Berg, "Chemotaxis in bacteria," *Annu. Rev. Biophys. Bioeng.*, vol. 4, pp. 119–136, 1975.

[78] L. Perkins, E. Hedgecock, J. Thomson, and J. Culotti, "Mutant sensory cilia in the nematode C. elegans.," *Dev. Biol.*, vol. 117, pp. 456–487, 1986.

[79] C. I. Bargmann, "Chemosensation in *C. elegans*," *WormBook : the online review of C. elegans biology*, pp. 1–29, 2006.

[80] B. Vidal, U. Aghayeva, H. Sun, C. Wang, L. Glenwinkel, E. A. Bayer, and O. Hobert, *An atlas of Caenorhabditis elegans chemoreceptor expression*, vol. 16. 2018.

[81] A. Metaxakis, D. Petratou, and N. Tavernarakis, "Multimodal sensory processing in Caenorhabditis elegans," *Open Biol.*, vol. 8, no. 6, 2018.

[82] H. ick Ha, M. Hendricks, Y. Shen, C. V. Gabel, C. Fang-Yen, Y. Qin, D. Colón-Ramos, K. Shen, A. D. Samuel, and Y. Zhang, "Functional Organization of a Neural Network for

Aversive Olfactory Learning in Caenorhabditis elegans," *Neuron*, vol. 68, no. 6, pp. 1173–1186, 2010.

[83] P. T. McGrath and I. Ruvinsky, "A primer on pheromone signaling in Caenorhabditis elegans for systems biologists.," *Current Opinion in Systems Biology*, pp. 23–30, 2019.

[84] T. Wakabayashi, I. Kitagawa, and R. Shingai, "Neurons regulating the duration of forward locomotion in Caenorhabditis elegans," *Neurosci. Res.*, vol. 50, no. 1, pp. 103–111, 2004.

[85] S. H. Chalasani, N. Chronis, M. Tsunozaki, J. M. Gray, D. Ramot, M. B. Goodman, and C. I. Bargmann, "Dissecting a circuit for olfactory behaviour in Caenorhabditis elegans Gene-specific control of inflammation by TLR-induced chromatin modifications," *Nature*, vol. 451, pp. 6540–6540, nov 2008.

[86] P. A. Garrity, M. B. Goodman, A. D. Samuel, and P. Sengupta, "Running hot and cold: Behavioral strategies, neural circuits, and the molecular machinery for thermotaxis in C. elegans and Drosophila," *Genes Dev.*, vol. 24, no. 21, pp. 2365–2382, 2010.

[87] K. Yoshida, T. Hirotsu, T. Tagawa, S. Oda, T. Wakabayashi, Y. Iino, and T. Ishihara, "Odour concentration-dependent olfactory preference change in C. elegans," *Nat. Commun.*, vol. 3, 2012.

[88] J. Larsch, D. Ventimiglia, C. I. Bargmann, and D. R. Albrecht, "High-throughput imaging of neuronal activity in Caenorhabditis elegans," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, nov 2013.

[89] S. G. Leinwand, C. J. Yang, D. Bazopoulou, N. Chronis, J. Srinivasan, and S. H. Chalasani, "Circuit mechanisms encoding odors and driving aging-associated behavioral declines in Caenorhabditis elegans," *Elife*, vol. 4, no. September 2015, pp. 1–26, 2015.

[90] A. Zaslaver, I. Liani, O. Shtangel, S. Ginzburg, L. Yee, and P. W. Sternberg, "Hierarchical sparse coding in the sensory system of Caenorhabditis elegans.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 4, pp. 1185–9, 2015.

[91] S. Yu, L. Avery, E. Baude, and D. L. Garbers, "Guanylyl cyclase expression in specific sensory neurons: A new family of chemosensory receptors," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 94, no. 7, pp. 3384–3387, 1997.

[92] T. R. Thiele, S. Faumont, and S. R. Lockery, "Articles, Behavioral/Systems/Cognitive The Neural Network for Chemotaxis to Tastants in Caenorhabditis elegans Is Specialized for Temporal Differentiation," *Journal of Neuroscience*, vol. 23, pp. 11904–11911, 2005.

[93] H. Suzuki, T. R. Thiele, S. Faumont, M. Ezcurra, S. R. Lockery, and W. R. Schafer, "Functional asymmetry in Caenorhabditis elegans taste neurons and its computational role in chemotaxis," *Nature*, vol. 454, pp. 114–117, jul 2008.

[94] P. D. Wes and C. I. Bargmann, "C. elegans odour discrimination requires asymmetric diversity in olfactory neurons," *Nature*, vol. 410, no. 6829, pp. 698–701, 2001.

[95] H. A. Colbert and C. I. Bargmann, "Odorant-specific adaptation pathways generate olfactory plasticity in C. elegans," *Neuron*, vol. 14, no. 4, pp. 803–812, 1995.

[96] N. D. L'Etoile, C. M. Coburn, J. Eastham, A. Kistler, G. Gallegos, and C. I. Bargmann, "The cyclic GMP-dependent protein kinase EGL-4 regulates olfactory adaptation in C. elegans," *Neuron*, vol. 36, no. 6, pp. 1079–1089, 2002.

[97] G. Jansen, D. Weinkove, and R. H. Plasterk, "The G-protein γ subunit gpc-1 of the nematode C.elegans is involved in taste adaptation," *EMBO J.*, vol. 21, no. 5, pp. 986–994, 2002.

[98] M. A. Hilliard, A. J. Apicella, R. Kerr, H. Suzuki, P. Bazzicalupo, and W. R. Schafer, "In vivo imaging of C. elegans ASH neurons: Cellular response and adaptation to chemical repellents," *EMBO J.*, vol. 24, no. 1, pp. 63–72, 2005.

[99] H. Barlow, "Possible principles underlying the transformations of sensory messages.," *Sensory Communication*, pp. 217–34, 1961.

[100] C. Von Der Malsburg, "The what and why of binding: The modeler's perspective," *Neuron*, vol. 24, no. 1, pp. 95–104, 1999.

[101] J. S. Kauer and J. White, "Imaging and Coding in the Olfactory System," *Annu. Rev. Neurosci.*, vol. 24, pp. 963–979, mar 2001.

[102] G. Laurent, "Olfactory network dynamics and the coding of multidimensional signals," *Nat. Rev. Neurosci.*, vol. 3, no. 11, pp. 884–895, 2002.

[103] A. Haverkamp, B. S. Hansson, and M. Knaden, "Combinatorial codes and labeled lines: How insects use olfactory cues to find and judge food, mates, and oviposition sites in complex environments," *Front. Physiol.*, vol. 9, no. FEB, pp. 1–8, 2018.

[104] B. A. Olshausen and D. J. Field, "Sparse coding of sensory inputs," *Curr. Opin. Neurobiol.*, vol. 14, no. 4, pp. 481–487, 2004.

[105] W. E. Vinje and J. L. Gallant, "Natural Stimulation of the Nonclassical Receptive Field Increases Information Transmission Efficiency in V1," *J. Neurosci.*, vol. 22, no. 7, pp. 2904–2915, 2002.

[106] M. R. Deweese, M. Wehr, and A. M. Zador, "Behavioral/Systems/Cognitive Binary Spiking in Auditory Cortex," vol. 23, no. 21, pp. 7940–7949, 2003.

[107] V. Grabe and S. Sachse, "Fundamental principles of the olfactory code," *BioSystems*, vol. 164, pp. 94–101, 2018.

[108] A. T. Schaefer and T. W. Margrie, "Spatiotemporal representations in the olfactory system," *Trends Neurosci.*, vol. 30, no. 3, pp. 92–100, 2007.

[109] N. Uchida, C. Poo, and R. Haddad, "Coding and transformations in the olfactory system," *Annu. Rev. Neurosci.*, vol. 37, pp. 363–385, 2014.

[110] C. D. Wilson, G. O. Serrano, A. A. Koulakov, and D. Rinberg, "A primacy code for odor identity," *Nat. Commun.*, vol. 8, no. 1, 2017.

[111] D. Zwicker, "Primacy coding facilitates effective odor discrimination when receptor sensitivities are tuned," *PLoS Comput. Biol.*, vol. 15, no. 7, pp. 1–22, 2019.

[112] T. Tesileanu, S. Cocco, R. Monasson, and V. Balasubramanian, "Adaptation of olfactory receptor abundances for efficient coding," *Elife*, vol. 8, pp. 1–39, 2019.

[113] D. Zwicker, A. Murugan, and M. P. Brenner, "Receptor arrays optimized for natural odor statistics," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 20, pp. 5570–5575, 2016.

[114] K. Krishnamurthy, A. M. Hermundstad, T. Mora, A. M. Walczak, and V. Balasubramanian, "Disorder and the neural representation of complex odors: Smelling in the real world," *bioRxiv*, pp. 1–13, 2017.

[115] S. Qin, Q. Li, C. Tang, and Y. Tu, "Optimal compressed sensing strategies for an array of nonlinear olfactory receptor neurons with and without spontaneous activity," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 116, no. 41, pp. 20286–20295, 2019.

[116] J. P. Nguyen, A. N. Linder, G. S. Plummer, J. W. Shaevitz, and A. M. Leifer, "Automatically tracking neurons in a moving and deforming brain," *PLOS Computational Biology*, vol. 13, pp. 1–19, 05 2017.

[117] V. Venkatachalam, N. Ji, X. Wang, C. Clark, J. K. Mitchell, M. Klein, C. J. Tabone, J. Florman, H. Ji, J. Greenwood, A. D. Chisholm, J. Srinivasan, M. Alkema, M. Zhen, and A. D. T. Samuel, "Pan-neuronal imaging in roaming Caenorhabditis elegans.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, pp. E1082–8, feb 2016.

[118] A. Giovannucci, J. Friedrich, P. Gunn, J. Kalfon, B. L. Brown, S. A. Koay, J. Taxidis, F. Najafi, J. L. Gauthier, P. Zhou, B. S. Khakh, D. W. Tank, D. B. Chklovskii, and E. A. Pnevmatikakis, "CaImAn an open source tool for scalable calcium imaging data analysis," *eLife*, vol. 8, 2019.

[119] A. Nejatbakhsh, E. Varol, E. Yemini, V. Venkatachalam, A. Lin, A. D. Samuel, O. Hobert, and L. Paninski, "Extracting neural signals fromsemi-immobilized animals with deformablenon-negative matrix factorization," *MICCAI*, pp. 14–24, 2020.

[120] M. B. Ahrens, K. H. Huang, S. Narayan, B. D. Mensh, and F. Engert, "Two-photon calcium imaging during fictive navigation in virtual environments," *Front. Neural Circuits*, vol. 7, no. JUNE, 2013.

[121] M. B. Ahrens, M. B. Orger, D. N. Robson, J. M. Li, and P. J. Keller, "Whole-brain functional imaging at cellular resolution using light-sheet microscopy," *Nature Methods*, vol. 10, no. 5, pp. 413–420, 2013.

[122] Y. Mu, D. V. Bennett, M. Rubinov, S. Narayan, C. T. Yang, M. Tanimoto, B. D. Mensh, L. L. Looger, and M. B. Ahrens, "Glia Accumulate Evidence that Actions Are Futile and Suppress Unsuccessful Behavior," *Cell*, vol. 178, no. 1, pp. 27–43, 2019.

[123] W. C. Lemon, S. R. Pulver, B. Höckendorf, K. McDole, K. Branson, J. Freeman, and P. J. Keller, "Whole-central nervous system functional imaging in larval Drosophila," *Nature Communications*, vol. 6, no. May, 2015.

[124] D. Karagyozov, M. Mihovilovic Skanata, A. Lesar, and M. Gershow, "Recording Neural Activity in Unrestrained Animals with Three-Dimensional Tracking Two-Photon Microscopy," *Cell Reports*, vol. 25, no. 5, pp. 1371–1383, 2018.

[125] K. Mann, C. L. Gallen, and T. R. Clandinin, "Whole-Brain Calcium Imaging Reveals an Intrinsic Functional Network in Drosophila," *Current Biology*, vol. 27, no. 15, pp. 2389–2396, 2017.

[126] S. Aimon, T. Katsuki, T. Jia, L. Grosenick, M. Broxton, K. Deisseroth, T. J. Sejnowski, and R. J. Greenspan, "Fast near-whole–brain imaging in adult drosophila during responses to stimuli and behavior," *PLoS biology*, vol. 17, no. 2, p. e2006732, 2019.

[127] S. Deny, S. Ganguli, T. R. Clandinin, *et al.*, "Causal coupling between neural activity, metabolism, and behavior across the drosophila brain," *bioRxiv*, 2020.

[128] D. A. Pacheco, S. Y. Thiberge, E. Pnevmatikakis, and M. Murthy, "Auditory activity is diverse and widespread throughout the central brain of drosophila," *Nat Neurosci*, vol. 24, pp. 93–104, 2021.

[129] C. Dupre and R. Yuste, "Non-overlapping neural networks in hydra vulgaris," *Current Biology*, vol. 27, 2017.

[130] J. L. Chen, M. L. Andermann, T. Keck, N. L. Xu, and Y. Ziv, "Imaging neuronal populations in behaving rodents: Paradigms for studying neural circuits underlying behavior in the mammalian cortex," *J. Neurosci.*, vol. 33, no. 45, pp. 17631–17640, 2013.

[131] G. Hong and C. M. Lieber, "Novel electrode technologies for neural recordings," *Nature reviews, Neuroscience*, vol. 20, pp. 330–345, 2019.

[132] T. Schrödel, R. Prevedel, K. Aumayr, M. Zimmer, and A. Vaziri, "Brain-wide 3D imaging of neuronal activity in Caenorhabditis elegans with sculpted light," *Nature Methods*, vol. 10, no. 10, pp. 1013–1020, 2013.

[133] S. Kato, H. S. Kaplan, T. Schrödel, S. Skora, T. H. Lindsay, E. Yemini, S. Lockery, and M. Zimmer, "Global Brain Dynamics Embed the Motor Command Sequence of Caenorhabditis elegans," *Cell*, vol. 163, pp. 656–669, oct 2015.

[134] H. S. Kaplan, O. Salazar Thula, N. Khoss, and M. Zimmer, "Nested Neuronal Dynamics Orchestrate a Behavioral Hierarchy across Timescales," *Neuron*, vol. 105, no. 3, pp. 562–576, 2020.

[135] K. M. Hallinen, R. Dempsey, M. Scholz, X. Yu, A. Linder, F. Randi, A. Sharma, J. W. Shaevitz, and A. M. Leifer, "Decoding locomotion from population neural activity in moving c. elegans," *bioRxiv*, 2021.

[136] V. Susoy, W. Hung, D. Witvliet, J. E. Whitener, M. Wu, B. J. Graham, M. Zhen, V. Venkat-achalam, and A. D. Samuel, "Natural sensory context drives diverse brain-wide activity during c. elegans mating," *bioRxiv*, 2020.

[137] H. S. Kaplan and M. Zimmer, "Brain-wide representations of ongoing behavior: a universal principle?," *Current Opinion in Neurobiology*, vol. 64, pp. 60–69, 2020.

[138] N. Chronis, M. Zimmer, and C. I. Bargmann, "Microfluidics for in vivo imaging of neuronal and behavioral activity in Caenorhabditis elegans," *Nature Methods*, vol. 4, pp. 727–731, sep 2007.

[139] E. Yemini, A. Lin, A. Nejatbakhsh, E. Varol, R. Sun, G. E. Mena, A. D. Samuel, L. Paninski, V. Venkatachalam, and O. Hobert, "NeuroPAL: A Multicolor Atlas for Whole-Brain Neuronal Identification in C. elegans," *Cell*, vol. 184, pp. 272–288, 2021.

[140] G. Si, J. Baron, Y. Feng, and A. D. T. Samuel, "An olfactory pattern generator for on-demand combinatorial control of receptor activities," *in preparation*, 2021.

[141] R. Prevedel, Y.-G. Yoon, M. Hoffmann, N. Pak, G. Wetzstein, S. Kato, T. Schrödel, R. Raskar, M. Zimmer, E. S. Boyden, *et al.*, "Simultaneous whole-animal 3d imaging of neuronal activity using light-field microscopy," *Nature methods*, vol. 11, no. 7, pp. 727–730, 2014.

[142] R. Haddad, R. Khan, Y. K. Takahashi, K. Mori, D. Harel, and N. Sobel, "A metric for odorant comparison," *Nat. Methods*, vol. 5, no. 5, pp. 425–429, 2008.

[143] J. Larsch, S. W. Flavell, Q. Liu, A. Gordus, D. R. Albrecht, and C. I. Bargmann, "A Circuit for Gradient Climbing in C. elegans Chemotaxis," *Cell Rep.*, vol. 12, no. 11, pp. 1748–1760, 2015.

[144] H. Sass, "Sensory encoding of odor stimuli in Periplaneta americana," *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.*, vol. 107, p. 49–65, 1976.

[145] M. Meister and T. Bonhoeffer, "Tuning and Topography in an Odor Map on the Rat Olfactory Bulb," *J. Neurosci.*, vol. 21, pp. 1351–1360, feb 2001.

[146] A. Rose and D. Baillie, "Genetic organization of the region around unc-15 (i), a gene affecting paramyosin in caenorhabditis elegans," *Genetics*, vol. 96, no. 3, pp. 639–48, 1980.

[147] J. D. Storey, "A direct approach to false discovery rates," tech. rep., 2002.

[148] M. B. Ahrens and F. Engert, "Large-scale imaging in small brains," *Current Opinion in Neurobiology*, vol. 32, pp. 78–86, jun 2015.

[149] J. W. Lichtman and W. Denk, "The big and the small: Challenges of imaging the brain's circuits," *Science*, vol. 334, pp. 618–623, nov 2011.

[150] A. R. Jones, C. C. Overly, and S. M. Sunkin, "The allen brain atlas: 5 years and beyond," nov 2009.

[151] H. Zeng and J. R. Sanes, "Neuronal cell-type classification: Challenges, opportunities and the path forward," *Nature Reviews Neuroscience*, vol. 18, pp. 530–546, sep 2017.

[152] I. Kotera, N. A. Tran, D. Fu, J. H. Kim, J. Byrne Rodgers, and W. S. Ryu, "Pan-neuronal screening in Caenorhabditis elegans reveals asymmetric dynamics of AWC neurons is critical for thermal avoidance behavior," *eLife*, vol. 5, 2016.

[153] J. P. Nguyen, F. B. Shipley, A. N. Linder, G. S. Plummer, M. Liu, S. U. Setru, J. W. Shaevitz, A. M. Leifer, M. Liu, S. U. Setru, J. W. Shaevitz, and A. M. Leifer, "Whole-brain calcium imaging with cellular resolution in freely behaving C. elegans," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, pp. E1074—-81, jan 2015.

[154] M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher, "Green fluorescent protein as a marker for gene expression," *Science*, vol. 263, no. 5148, pp. 802–805, 1994.

[155] T. Boulin, J. F. Etchberger, and O. Hobert, "Reporter gene fusions.," *WormBook : the online review of C. elegans biology*, pp. 1–23, 2006.

[156] A. Fire, S. W. Harrison, and D. Dixon, "A modular set of lacZ fusion vectors for studying gene expression in Caenorhabditis elegans," *Gene*, vol. 93, no. 2, pp. 189–198, 1990.

[157] J. Livet, T. A. Weissman, H. Kang, R. W. Draft, J. Lu, R. A. Bennis, J. R. Sanes, and J. W. Lichtman, "Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system," *Nature*, vol. 450, pp. 56–62, nov 2007.

[158] B. Richier and I. Salecker, "Versatile genetic paintbrushes: Brainbow technologies," *Wiley Interdisciplinary Reviews: Developmental Biology*, vol. 4, pp. 161–180, mar 2015.

[159] T. A. Weissman and Y. A. Pan, "Brainbow: New resources and emerging biological applications for multicolor genetic labeling and analysis," feb 2014.

[160] O. M. Subach, P. J. Cranfill, M. W. Davidson, and V. V. Verkhusha, "An enhanced monomeric blue fluorescent protein with the high chemical stability of the chromophore," *PLoS ONE*, vol. 6, dec 2011.

[161] J. Chu, Y. Oh, A. Sens, N. Ataie, H. Dana, J. J. Macklin, T. Laviv, E. S. Welf, K. M. Dean, F. Zhang, B. B. Kim, C. T. Tang, M. Hu, M. A. Baird, M. W. Davidson, M. A. Kay, R. Fiolka, R. Yasuda, D. S. Kim, H. L. Ng, and M. Z. Lin, "A bright cyan-excitable orange fluorescent protein facilitates dual-emission microscopy and enhances bioluminescence imaging in vivo," *Nature Biotechnology*, vol. 34, pp. 760–767, jul 2016.

[162] N. C. Shaner, M. Z. Lin, M. R. McKeown, P. A. Steinbach, K. L. Hazelwood, M. W. Davidson, and R. Y. Tsien, "Improving the photostability of bright monomeric orange and red fluorescent proteins," *Nature Methods*, vol. 5, pp. 545–551, jun 2008.

[163] J. Chu, R. D. Haynes, S. Y. Corbel, P. Li, E. González-González, J. S. Burg, N. J. Ataie, A. J. Lam, P. J. Cranfill, M. A. Baird, M. W. Davidson, H. L. Ng, K. C. Garcia, C. H. Contag, K. Shen, H. M. Blau, and M. Z. Lin, "Non-invasive intravital imaging of cellular differentiation with a bright red-excitable fluorescent protein," *Nature Methods*, vol. 11, no. 5, pp. 572–578, 2014.

[164] N. Stefanakis, I. Carrera, and O. Hobert, "Regulatory Logic of Pan-Neuronal Gene Expression in C. elegans," *Neuron*, vol. 87, pp. 733–750, aug 2015.

[165] O. Hobert, L. Glenwinkel, and J. White, "Revisiting Neuronal Cell Type Classification in Caenorhabditis elegans," nov 2016.

[166] C. C. Mello, J. M. Kramer, D. Stinchcomb, and V. Ambros, "Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences," *The EMBO Journal*, vol. 10, no. 1, pp. 3959–3970, 1991.

[167] J. M. Gray, J. J. Hill, and C. I. Bargmann, "A circuit for navigation in Caenorhabditis elegans," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 9, pp. 3184–3191, 2005.

[168] S. H. Chalasani, S. Kato, D. R. Albrecht, T. Nakagawa, L. F. Abbott, and C. I. Bargmann, "Neuropeptide feedback modifies odor-evoked dynamics in Caenorhabditis elegans olfactory neurons," *Nature Neuroscience*, vol. 13, pp. 615–621, may 2010.

[169] D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison, and H. O. Smith, "Enzymatic assembly of DNA molecules up to several hundred kilobases," *Nature Methods*, vol. 6, no. 5, pp. 343–345, 2009.

[170] F. Van Den Ent and J. Löwe, "RF cloning: A restriction-free method for inserting target genes into plasmids," *Journal of Biochemical and Biophysical Methods*, vol. 67, pp. 67–74, apr 2006.

[171] O. Hobert, "PCR fusion-based approach to create reporter Gene constructs for expression analysis in transgenic C. elegans," *BioTechniques*, vol. 32, no. 4, pp. 728–730, 2002.

[172] J. F. Etchberger and O. Hobert, "Vector-free DNA constructs improve transgene expression in C. elegans," *Nature Methods*, vol. 5, p. 3, jan 2008.

[173] M. Granato, H. Schnabel, and R. Schnabel, "pha-1, a selectable marker for gene transfer in C.elegans," *Nucleic Acids Research*, vol. 22, no. 9, pp. 1762–1763, 1994.

[174] H. W. Ai, N. C. Shaner, Z. Cheng, R. Y. Tsien, and R. E. Campbell, "Exploration of new chromophore structures leads to the identification of improved blue fluorescent proteins," *Biochemistry*, vol. 46, pp. 5904–5910, may 2007.

[175] O. Zapata-Hommer and O. Griesbeck, "Efficiently folding and circularly permuted variants of the Sapphire mutant of GFP," *BMC Biotechnology*, vol. 3, 2003.

[176] T.-T. Yang, L. Cheng, and S. R. Kain, "Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein," Tech. Rep. 22, 1996.

[177] H. W. Ai, K. L. Hazelwood, M. W. Davidson, and R. E. Campbell, "Fluorescent protein FRET pairs for ratiometric imaging of dual biosensors," *Nature Methods*, vol. 5, pp. 401–403, may 2008.

[178] J. Yang, L. Wang, F. Yang, H. Luo, L. Xu, J. Lu, S. Zeng, and Z. Zhang, "mBeRFP, an Improved Large Stokes Shift Red Fluorescent Protein," *PLoS ONE*, vol. 8, jun 2013.

[179] J. Liu, A. Ward, J. Gao, Y. Dong, N. Nishio, H. Inada, L. Kang, Y. Yu, D. Ma, T. Xu, I. Mori, Z. Xie, and X. Z. Xu, "C. elegans phototransduction requires a G protein-dependent cGMP pathway and a taste receptor homolog," *Nature Neuroscience*, vol. 13, pp. 715–722, jun 2010.

[180] P. Dayan, L. F. Abbott, *et al.*, "Theoretical neuroscience: computational and mathematical modeling of neural systems," *Journal of Cognitive Neuroscience*, vol. 15, no. 1, pp. 154–155, 2003.

[181] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," *arXiv preprint arXiv:1401.4082*, 2014.

[182] S. Brenner, "The genetics of Caenorhabditis elegans.," *Genetics*, vol. 77, pp. 71–94, may 1974.

[183] E. Candès and B. Recht, "Exact matrix completion via convex optimization," *Commun. ACM*, vol. 55, no. 6, pp. 111–119, 2012.

[184] L. T. Nguyen, J. Kim, and B. Shim, "Low-Rank Matrix Completion: A Contemporary Survey," *IEEE Access*, vol. 7, pp. 94215–94237, 2019.

[185] M. Linkert, C. T. Rueden, C. Allan, J. M. Burel, W. Moore, A. Patterson, B. Loranger, J. Moore, C. Neves, D. MacDonald, A. Tarkowska, C. Sticco, E. Hill, M. Rossner, K. W. Eliceiri, and J. R. Swedlow, "Metadata matters: Access to image data in the real world," may 2010.

[186] D. L. Chase and M. R. Koelle, "Genetic analysis of RGS protein function in Caenorhabditis elegans," *Methods in Enzymology*, vol. 389, pp. 305–320, 2004.

[187] E. M. Hedgecock, J. G. Culotti, J. N. Thomson, and L. A. Perkins, "Axonal guidance mutants of Caenorhabditis elegans identified by filling sensory neurons with fluorescein dyes," *Developmental Biology*, vol. 111, no. 1, pp. 158–170, 1985.

[188] E. Yemini, T. Jucikas, L. J. Grundy, A. E. Brown, and W. R. Schafer, "A database of Caenorhabditis elegans behavioral phenotypes," *Nature Methods*, vol. 10, pp. 877–879, sep 2013.

[189] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in neural information processing systems*, pp. 556–562, 2001.

[190] E. A. Mukamel, A. Nimmerjahn, and M. J. Schnitzer, "Automated analysis of cellular signals from large-scale calcium imaging data," *Neuron*, vol. 63, no. 6, pp. 747–760, 2009.

[191] E. A. Pnevmatikakis, D. Soudry, Y. Gao, T. A. Machado, J. Merel, D. Pfau, T. Reardon, Y. Mu, C. Lacefield, W. Yang, *et al.*, "Simultaneous denoising, deconvolution, and demixing of calcium imaging data," *Neuron*, vol. 89, no. 2, pp. 285–299, 2016.

[192] E. A. Pnevmatikakis and A. Giovannucci, "Normcorre: An online algorithm for piecewise rigid motion correction of calcium imaging data," *Journal of neuroscience methods*, vol. 291, pp. 83–94, 2017.