Development of Optical Tools and Techniques Toward a Functional Connectomic Understanding of *C. elegans*

by

Danielle Marie Orozco Cosio

B.S., Massachusetts Institute of Technology (2014)

Submitted to the Department of
Brain and Cognitive Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2022

Author . . . . . .
Department of Brain & Cognitive Science
May 5, 2021

Certified by. . . . . .
Edward S. Boyden
Y. Eva Tan Professor of Brain and Cognitive Sciences
Thesis Supervisor

Accepted by . . . . . .
Mark Harnett
Graduate Officer, Department of Brain and Cognitive Sciences

Development of Optical Tools and Techniques Toward a Functional Connectomic Understanding of *C. elegans*

by

Danielle Marie Orozco Cosio

Submitted to the Department of Brain and Cognitive Sciences
on April 29, 2022, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
In Neuroscience

Abstract

Optical methods to study *C. elegans* behavior and neural activity are popular and well-established in the field, but many of the most commonly used optical tools leave much to be improved upon. In this research I sought to establish the utility of new optical tools in *C. elegans* to address the shortcomings of commonly used ones. First, I present the properties and demonstrate the functionality of an improved near-infrared negative calcium ion indicator, NIR-GECO2, for imaging olfactory stimulated and optogenetically evoked neural activity. Next, I present the properties and demonstrate the functionality of a tool to strategically arrange GCaMP in clusters, STARC, to enable imaging of compartmentalized neural activity in regions dense with neural projections. Finally, I present a novel computational and RNA fluorescent in-situ hybridization-based method for unique identification of *C. elegans* neurons which allows for experiments to be performed in any *C. elegans* strain and can be flexibly applied to new applications and organisms. When appropriately combined with existing methods, these tools and techniques enable experiments that can push the field of *C. elegans* systems neuroscience towards a functional connectomic understanding of the neural control of the animal's behavior.

Thesis Supervisor: Ed Boyden
Title: Y. Eva Tan Professor of Brain & Cognitive Science

Acknowledgements

I could not have done any of this without standing on the shoulders of my friends and collaborators in the Boyden lab, of which there are far too many to name. Thank you so much for the hours of brainstorming, venting, being silly, and just hanging out together. It's made this experience one I'll always treasure. I want to specifically thank Orhan Çeliker, for being right there alongside me through this whole journey, especially when I need some last-minute support in lab or just to talk about our lives. And I especially want to thank my advisor, Ed Boyden, for your undying spirit of exploration, your adamant insistence on independence as a researcher, and your infinite offers of any other ways you could help. I joined this lab specifically to grow in those directions, and I feel like an entirely different person from when I started six years ago. A lot of that is also due to the amazing friends I've made and deepened in the course of this program. A few I've lived with, a few I've made silly projects with, several I've worked with, and probably all I've cried in front of. As my friend Joyce Wang says, "the real PhD is the friends we made along the way"

Finally, and most importantly, I could not have come even a little close to where I am today without the unrelenting support from my family. To my parents and siblings, thank you so much for teaching me to give my all to something I love and care about. Through late nights finishing homework, commiserating on our lows and being there to celebrate our highs, and through example, you've taught me to see and appreciate all I am capable of. And to my husband, David Orozco Cosio, and son, David Armando, (and my emotional support dog Luna): you're there every day reminding me of what's important. You pick up slack when I'm overwhelmed, and take me out of my head when I need it most. Coming home to you guys and remembering why I work so hard and what I'm working toward, and that we're going to get there together, is the best part of every day and what always keeps me going.

Contents

1 Introduction

1.1 The Role of Optical Methods for Calcium Imaging and Connectomics in *C. elegans* Systems Neuroscience

Optical methods of calcium imaging are among the most commonly used methods for observing neural activity across large numbers of neurons. They are also increasingly used in conjunction with behavioral paradigms in many model organisms in neuroscience. An understanding of the underlying connective structure that underlies neurons' observed neural activity is required to enable analysis of how information is represented and transformed in these ensembles as animals perform complex tasks. These two pieces of information have a currently unique intersection in C. elegans, a transparent organism with a compact neural system whose 302 neurons enable a wide variety of complex behaviors, whose canonical connectome is already mapped and molecularly well-characterized, and in which a multitude of molecular tools and optical methods is available.

Since the mid-70s, calcium ion (Ca^{2+}) level indicators have been recognized as highly valuable tools that provide a proxy measure for neural activity. Upon the broad availability of genetically encoded Ca^{2+} indicators (GECIs) starting in the early 2000s, this technique became a staple of neuroscience, with improved versions of the most commonly used GECI, GCaMP, being developed every few years (Nakai *et al.,* 2001, Zhang *et al.,* 2021). In *C. elegans*, a transparent animal with mainly ramped potentials and a few neurons showing action potentials (Liu *et al.* 2018, Jiang *et al.* 2021), the last several versions of GCaMP are fast and sensitive enough to capture subtle and spatially distinct Ca^{2+} transients (both within and between cells) reflecting neural computation. These tools are the default for observing

neural activity in the animal and have been used in several configurations for a broad diversity of circuit study in worm systems neuroscience. Cytosolic expression in small numbers of neurons has been used to observe correlated and compartmentalized neural activity, co-expression with optogenetic tools has been used to show functional and behavioral causation, and nuclear localized expression has become increasingly popular with the advent of whole-brain imaging in the model system (Lin *et al.* 2022).

Unfortunately, there are several issues specific to *C. elegans* that arise with the use of blue light required as the optimal excitation wavelength in these most commonly used indicators. One of the more dramatic is that of phototoxicity. Prolonged exposure to visible light, and especially short wavelength blue or UV light has several detrimental effects to worm health, including the induction of oxidative stress and an unfolded protein response, a decreased lifespan, and stunted growth in development (De Magalhaes Filho *et al.,* 2018). Most relevant to the present work, however, is its effect as an aversive stimulus (Edwards *et al.,* 2008, Ward *et al.,* 2008). Blue light has been shown to elicit a robust negative phototaxis response in worms, which has specifically been used in behavioral assays and whose sensory responses have been investigated (Lee and Aschner 2016, Yemini *et al.,* 2021); it has also been shown to slow feeding and drive expulsion of ingested material (Sando *et al.*, 2021). This is sufficient reason to believe that exposures to the wavelengths necessary for excitation of GCaMP may stoke motor and neural activity responses that could confound the results of a behavioral or whole brain imaging experiment. The use of mutants to avoid these negative consequences could necessitate further behavioral and imaging controls to avoid skepticism of genetic reliability of the strain and generalizability of results, but it is an accepted solution

(Edwards *et al.* 2008). Additional disadvantages to using blue-light shifted Ca^{2+} indicators are the high levels of background fluorescence emitted by the worm intestine (Fig 1A, 1D, 2B), and the inability to co-express with the certain optogenetic tools. CoChR, a blue-light activated optogenetic tool is often used in worms to test the causal relationship between a neuron and a behavior, but if researchers would like to combine optical stimulation and imaging, they switch to a red light-activated optogenetic tool such as Chrimson to observe evoked neural activity with GCaMP (Schild & Glauser, 2015, Klapoetke *et al.* 2014).

As mentioned earlier, whole-brain imaging is an increasingly popular technique in *C. elegans,* and is becoming understood to be a necessary strategy for understanding how global neural activity drives sensation and behavior. Global patterns of neural dynamics collected in such experiments have been associated with chemosensory responses, thermosensory responses, specific locomotor patterns, behavioral states, and wakefulness (Yemini *et al.,* 2021, Nguyen *et al.,* 2016, Hallinen *et al.,* 2021, Venkatachalam *et al.,* 2016, Gonzales *et al.,* 2019). However, early studies had been unable to perform comparisons across animals or narrow down specific circuit dynamics because of an inability to identify the complete set of imaged neurons. Solutions to this problem have been introduced as genetically encoded deterministic fluorescence maps in a small number of available worm strains (Toyoshima *et al.,* 2020, Yemini *et al.,* 2021). However, concerns about the phenotypical, behavioral, and reproductive health of these strains may slow their widespread adoption as a default worm model for whole-brain imaging. Increasing attention is also being given to efforts to characterize the physical connectome in further detail via electron microscopy and brainbow applications, along with efforts in understanding its relationship to electrical and

neuromodulatory networks and how they all integrate between their informational modalities to produce observed activity and behavior (Moyle *et al.,* 2021, Brittin *et al.,* 2021, Ruach *et al.,* 2022, Flavell & Gordus 2022). Regrettably, the same method that enables whole-brain imaging (nuclear localization of Ca^{2+} reporters) currently excludes the possibility of monitoring activity in neural processes (especially the densely organized nerve ring), where these studies indicate a large amount of compartmentalized computation involved in sensory integration may take place. Nevertheless, this growing integration of interdisciplinary and multiscale approaches will lead to more complete datasets that will allow for highly predictive models of neural activity, chemical dynamics, and behavior.

1.2 Current Methods of Cell Identification in C. elegans

A feature that makes *C. elegans* unique among model organisms is its known connectome of highly stereotyped neural number, identity, and connectivity (White *et al.,* 1986). This has made it possible to assign cell classes to neurons that can reliably be found in equal number and the same approximate location in every worm. The two most common methods used to identify neurons are optical methods that persist from the techniques first used to define those classes: differential interference contrast (DIC) microscopy and fluorescent labeling; these can be utilized in combination to identify a fluorescently marked unknown neuron or small set of neurons (Emmons *et al.* 2021). Using DIC microscopy to identify neurons is time-consuming, manual, and requires a deep knowledge of location and morphology of *C. elegans* neurons to perform. Fosmid reporter strains created by the Hobert lab fluorescently label cholinergic, GABAergic, and glutamatergic neurons can be crossed with a strain with unknown marked neurons to identify the expression pattern of the marker

(Emmons *et al.,* 2021). Neither of these techniques, however, are particularly useful for identifying the full set of neurons in a whole-brain imaging application.

The Hobert and Iino labs have created fully labeled fluorescent neuronal marker strains specifically to address the issue of neuronal cell class ID in whole-brain imaging (Yemini *et al.,* 2021, Toyoshima *et al.,* 2020). While they effectively solve this problem, each of these approaches still involve considerable knowledge of specific neuronal properties and some manual annotation, which takes years and hours to acquire and perform, respectively. They are also designed under tight constraints, where the fluorescent markers are predetermined and require several specific sets of optical filters. While the most commonly used channels for activity imaging and fluorescent labeling are left open to be used at the investigators' discretion, this means that the label fluorophores are less commonly used filter sets, and all work must be done in the genetic background of worm strains that have large numbers of existing genetic manipulations. This lack of flexibility extends to applicability to other organisms and applications; these strains took a lot of time and manual trial and error to generate, and while lessons learned in this process could surely be applied to similar efforts in other organisms or specific cases, generating transgenic animals along a similar strategy would take comparably as long, or longer.

In summary, several technical improvements can be made to improve the experimental flexibility and reliability in *C. elegans* systems neuroscience, as is the aim of this thesis. Specifically, alternatives to blue light shifted Ca^{2+} indicators, methods for monitoring compartmentalized activity in the nerve ring, and more flexible and robust methods of neural cell ID to facilitate the comparison of whole-brain imaging analyses between animals

and to elucidate the relationship between neural functional dynamics and the physical connectome.

1.3 Overview of Thesis

Optical methods to study *C. elegans* neural activity are popular and well-established in the field, but many of the most commonly used optical tools leave much to be improved upon. In this research I sought 1) to establish the utility of new optical tools in *C. elegans* to address the shortcomings of commonly used ones, and 2) to develop a flexible method of *C. elegans* cell type identification (ID) that has potential to easily generalize to other organisms and be applied not only for whole brain imaging, but also for developmental, cell fate and specification, and other types of experiments. . When appropriately combined with existing methods, these tools and techniques could push the field of *C. elegans* neuroscience towards a functional connectomic understanding of the neural control of the animal's behavior.

In Chapter 2, I present the functionality of an improved, near-infrared, genetically-encoded calcium ion indicator in *C. elegans* neurons. Use of this tool to monitor neural activity removes the need to bathe the animals in blue light during long-term activity imaging experiments and allows for use of blue light-activated optogenetic tools to be utilized during activity imaging, among other advantages.

In Chapter 3, I present the functionality of GCaMP stochastically arranged in clusters in *C. elegans* neurons. This tool allows for monitoring activity in neurite compartments in multiple neurons and may serve well in tightly crowded spaces like the nerve ring.

In Chapter 4, I introduce a novel, RNA fluorescent in-situ hybridization (FISH)-based method for unique identification of *C. elegans* neurons. This method allows for

behavioral, genetic, and activity imaging experiments to be performed flexibly in strains of a lab's own creation, while also accelerating the practice of neuronal cell ID.

2 An Improved Near-InfraRed, Genetically-Encoded Calcium Ion Indicator (NIR-GECO2) for in vivo Imaging in *C. elegans*

2.1 Introduction

Collaborators and I have demonstrated that near-infrared (NIR) excitation and emission (>650nm) calcium ion (Ca^{2+}) indicators have the advantages of reduced phototoxicity, reduced cross-talk with optogenetic actuators, and decreased scattering and absorption in tissues compared to green or red genetically encoded fluorescent indicators (Qian *et al.,* 2020). However, in contrast to fluorescence imaging of intracellular Ca^{2+} transients using genetically encoded Ca^{2+} indicators (GECIs) based on green and red fluorescent proteins (GFP and RFP, respectively), efforts to develop NIR GECIs are at a relatively nascent state (Chen *et al.,* 2013, Dana *et al.,* 2016, Dana *et al.,* 2019, Qian *et al.,* 2019, Subach *et al.,* 2019). One of the only available NIR GECIs, a molecule designated NIR-GECO1, was engineered by genetic insertion of the Ca^{2+}-responsive domain calmodulin (CaM)-RS20 into the protein loop close to the biliverdin binding site of mIFP (Qian *et al.,* 2019). NIR-GECO1 provides a robust inverted fluorescence response (a fluorescence decrease upon Ca^{2+} increase) in response to Ca^{2+} concentration changes in cultured cells, primary neurons, and acute brain slices. We sought to improve NIR-GECO1, using directed evolution to increase brightness and dynamic range in response to calcium concentration ($\Delta F/F_0$). We then demonstrated the ability to image pan-neuronal

spontaneous and light-evoked activity using the resulting NIR-GECO variants in vivo in small organisms, including *C. elegans*.

In *C. elegans*, there are several species-specific potential advantages of using a NIR fluorescent protein-based tool as opposed to green or red fluorescent protein-based tools for imaging neuronal activity. These potential advantages include:

1) reducing the need to expose worms to blue light, a known aversive stimulus (Edwards *et al.,* 2008, Ward *et al.,* 2008), ,
2) reducing non-indicator intestinal background that occurs when imaging at a 488nm excitation wavelength,
3) enabling use of powerful blue-light activated optogenetic tools (such as CoChR) to evoke activity.

By using NIR Ca^{2+} indicators for *C. elegans* neural activity imaging, many of the assumptions and accommodations regularly made to complete these experiments with GCaMP variants (as outlined in 1.1) could be rigorously tested.

Details concerning the development and characterization of the new NIR-GECO variants can be found in the Qian et al. 2019 published paper. This chapter will discuss my primary contribution in establishing their use in *C. elegans* in vivo imaging.

2.2 Materials & Methods

Worms were cultured and maintained following standard protocols (Brenner *et al.,* 1974). The genes of NIR-GECO2, NIR-GECO2G, HO1, CoChR, and jGCaMP7s for expression in C. elegans were codon-optimized using SnapGene codon-optimization tool and synthesized by GenScript. Transgenic worms expressing NIR-GECO2G and jGCaMP7s

pan-neuronally or NIR-GECO2 in AVA and CoChR-GFP in ASH were generated by injecting the plasmids tag-168::NLS-NIR-GECO2(G)-T2A-HO1(or NIR-GECO2) and tag-168::NLS-jGCaMP7s or plasmids flp-18::NIR-GECO2-T2A-HO1, sra-6::CoChR-SL2-GFP, and elt-2::NLS-GFP into N2 background worms, respectively, picking those with the strongest expression of green fluorescence (in neurons for the pan-neuronal strain and in the gut for optogenetic strain). NLS sequence used in this experiment was PKKKRKV.

Hermaphrodite transgenic worms were picked at L4 stage of development and put onto NGM plates with freshly seeded OP50 lawns 12 to 24 hours before experiments, with or without 100-μM all-trans-retinal (ATR) for optogenetic experiments. Worms were mounted on 2% agarose pads on microscope slides, immobilized with 5 mM tetramisole, covered by a coverslip, and imaged using a Nikon Eclipse Ti inverted microscope equipped with a confocal spinning disk (CSU-W1), a 40×, 1.15 NA water-immersion objective, and a 5.5 Zyla camera (Andor, Belfast, Northern Ireland), controlled by NIS-Elements AR software. To acquire data shown in Fig 1, the fluorescence of NIR-GECO2 was imaged with 640-nm excitation provided by a 41.9-mW laser and a 685/40-nm emission filter; jGCaMP7s/GFP fluorescence was imaged with a 488-nm excitation provided by 59.9-mW laser and a 525/50-nm emission filter. Optogenetic stimulation was performed with 488-nm illumination at 20 mW/mm2. For 200 mM NaCl stimulation, worms were imaged using the same optical setup as above, using a microfluidic device that was described previously (Chronis *et al.,* 2007).

For brightness and SBR comparison of NLS-jGCaMP7s and NIR-GECO2, as shown in Fig 2A, 2B and 2D, the plasmids tag168::NLS-jGCaMP7s and tag168::NIR-GECO2-T2A-

HO1 were mixed at a ratio of 1:1 before injection. NLS-jGCaMP7s was imaged with 488-nm excitation at a power of 17.2 mW/mm2, and a 525/50-nm emission filter. NIR-GECO2 was imaged with 640-nm excitation at a power of 12.6 mW/mm2,and a 660LP emission filter. All other instrument settings were the same for NLS-jGCaMP7s and NIR-GECO2. The data for brightness from each ROI were averaged by ROI area. SBR was obtained via dividing fluorescence intensity from neurons by averaged autofluorescence from the intestine area. The imaging conditions in Fig 2C and 2E are the same as those in Fig 1. The data for SNR of NLS-jGCaMP7s and NLS-NIR-GECO2 were quantified from spontaneously active neurons. SNR was calculated by dividing fluorescence change associated with the peak of a graded potential by the standard deviation (SD) of the baseline fluorescence over the 2-second period immediately before the increase.

All images in the manuscript were processed and analyzed using either ImageJ (NIH) or NIS-Elements Advanced Research software (Nikon). Traces and graphs were generated using GraphPad prism 8, Origin (OriginLab, Wellesley, Massachusetts, USA), and Matlab. Data are presented as mean ± SD or mean ± standard error of the mean (SEM) as indicated.

2.3 Results

Two variants were derived by applying directed evolution (Qian *et al.,* 2019) to NIR-GECO1 as a template, with approximately equivalent fluorescence spectral profile, peak maxima, extinction coefficient, quantum yield, and pK_a. NIR-GECO2G showed a 50% improvement in brightness, a higher Ca^{2+} affinity (dissociation constant (K_d) = 480nM), and 3.7-fold improvements in sensitivity to Ca^{2+} concentrations in HeLa cells compared to NIR-GECO1 (K_d = 885nM). NIR-GECO2 was 25% dimmer than NIR-GECO1, but showed an

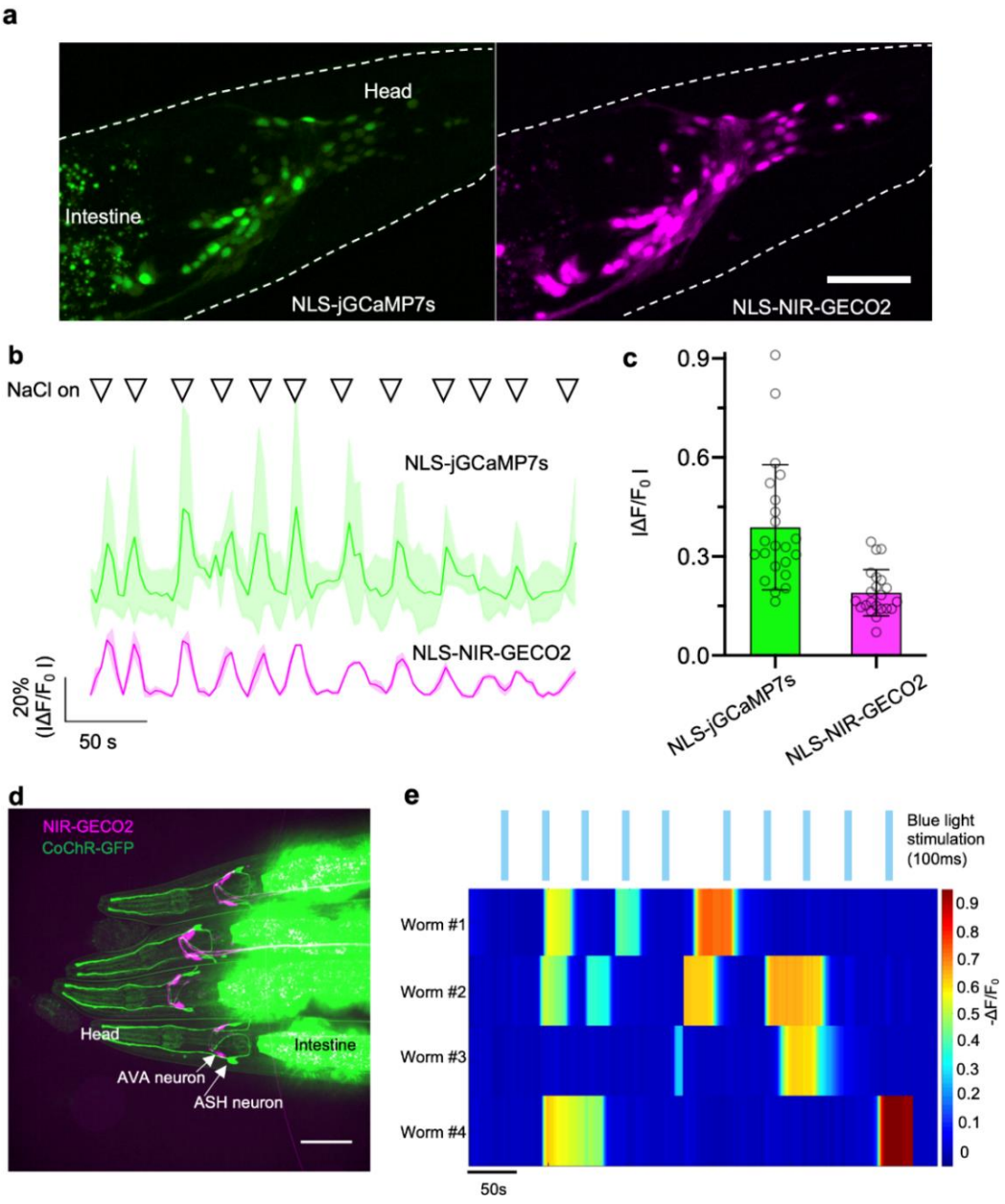

Figure 1: NIR-GECO2 can be used to detect microfluidic chemical stimulation and is spectrally compatible with blue light optogenetic stimulation in *C. elegans* in vivo. A) Left, fluorescent image of neurons expressing NLS-jGCaMP7s. Right, fluorescent image of neurons expressing cytosolic NIR-GECO2. Images from a single worm, representative of more than 3 worms, both under tag-168 promoter. Scale bar, 50 µm. This strain was used to obtain data for B and C. B) Traces of GCaMP (green) and NIR-GECO2 (pink) in response to 200 mM NaCl delivered in a microfluidic chamber. Solid lines represent averaged data from 3 neurons each. Shaded areas are SD. Triangles above the traces indicate the starting time points of stimulation (20 second duration). C) Quantitative fluorescence changes of NLS-jGCaMP7s and NLS-NIR-GECO2 in b (n = 36 peaks from 3 neurons). D) Fluorescence image of 4 *C. elegans* expressing CoChR (green) in the ASH neurons (under sra-6 promoter) and NIR-GECO2 (pink) in the AVA neurons (under flp-18 promoter). The merged image is shown. This strain is used to obtain the data shown in E. E) Individual traces of NIR-GECO2 fluorescence in an AVA neuron under blue light illumination (20 mW/mm2, λex = 488-nm laser light, 100 ms duration), indicated by blue bars above activity data. Figure and description adapted from Qian *et al.,* 2021.

even higher Ca^{2+} affinity (K_d = 331nM) and 3.6-fold improvements in sensitivity to Ca^{2+} concentrations.

We tested whether NIR-GECO2 (chosen for its higher Ca^{2+} affinity) would be suitable for in vivo imaging of neuronal activity by creating non-integrated lines of *C. elegans*. Because C. elegans is unable to synthesize heme de novo, its internal biliverdin concentration is quite low. We therefore co-expressed heme oxygenase-1 (HO-1) to increase the conversion of heme into biliverdin, injecting nuclear localized (NLS)-NIR-GECO2-T2A-HO1 and NLS-jGCaMP7s as a co-injection marker, both in plasmids under the pan-neuronal tag-168 promoter in extrachromosomal arrays. The resulting transgenic worms exhibited bright nuclear localized fluorescence from both NIR-GECO2 and jGCaMP7s (Fig 1A).

In worms expressing both NLS-jGCaMP7s and NIR-GECO2, the resting state cellular brightness of NIR-GECO2 was approximately 1.5-fold higher than that of jGCaMP7s (Fig 2A), and the signal-to-background ratio (SBR; i.e., the ratio of fluorescence emitted from neurons to autofluorescence from the intestine area) of NIR-GECO2 was approximately 5-fold larger than that of jGCaMP7s (Fig 2B). SBRs had no significant differences at different imaging depths in the worm (Fig 2D).

For functional imaging of NIR-GECO2, microfluidic chips (Chronis *et al.,* 2007) were used to deliver a high osmotic strength stimulus (200 mM NaCl) to individual worms, and the fluorescence was imaged simultaneously in the NIR and green fluorescence channels. Following exposure to a high concentration of NaCl, we detected synchronous but opposing fluorescent changes for jGCaMP7s (fluorescence increases) and NIR-GECO2

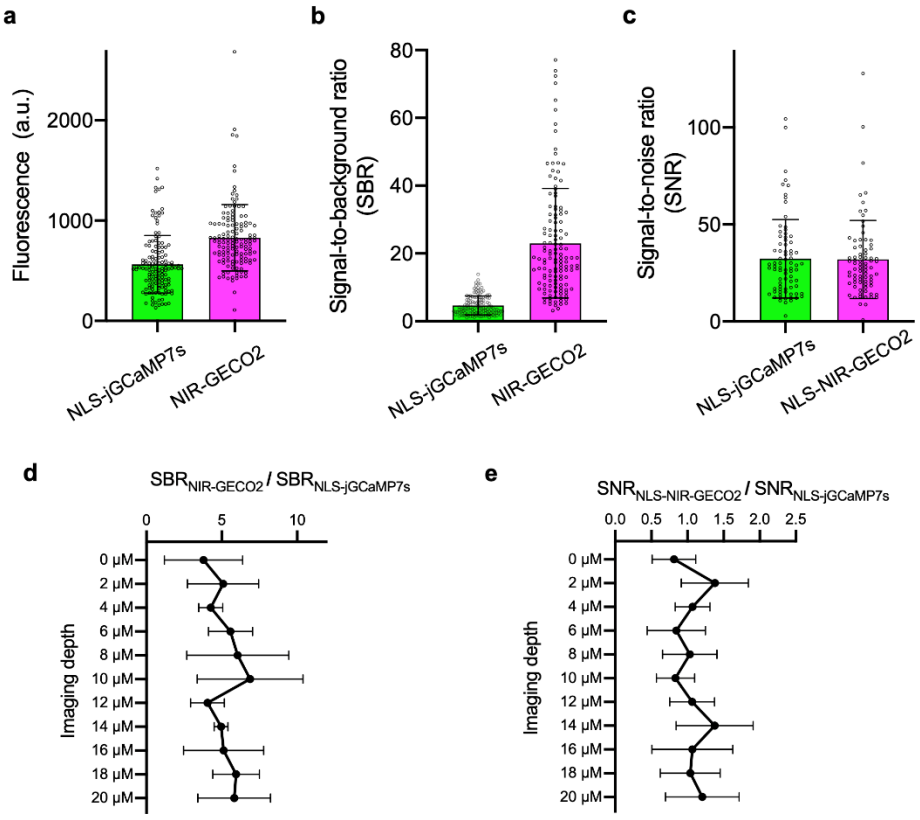

Figure 2: NIR-GECO2 has similar fluorescent properties and improved signal-to-background compared to jGCaMP7s in *C. elegans*. A) Fluorescence intensity of NLS-jGCaMP7s and NIR-GECO2 in C. elegans neurons at resting state. Fluorescence was normalized to the same excitation intensity (n = 132 ROIs from 5 worms; data are shown as mean ± SD) B) SBR of NLS-jGCaMP7s and NIR-GECO2 in neurons of C. elegans at resting state (n = 132 ROIs from 5 worms; data are shown as mean ± SD). SBR was obtained via dividing the fluorescence intensity from neurons by the averaged autofluorescence from the intestine area. C) SNR of NLS-jGCaMP7s and NLS-NIR-GECO2 quantified from spontaneously active neurons (n = 78 ROIs from 4 worms; data are shown as mean ± SD). SNR was calculated by dividing the fluorescence change associated with the peak of a graded potential by the SD of the baseline fluorescence over the 2-second period immediately before the increase. D) The ratio of $SBR_{NIR\text{-}GECO2}$ to $SBR_{NLS\text{-}jGCaMP7s}$ at different imaging depths (n = 5 worms; data are shown as mean ± SD). E) The ratio of $SNR_{NLS\text{-}NIR\text{-}GECO2}$ to $SNR_{NLS\text{-}jGCaMP7s}$ at different imaging depths (n = 4 worms; data are shown as mean ± SD). NIR-GECO2 (without NLS) and NLS-jGCaMP7s were used for the experiments in A, B, and D; NLS-NIR-GECO2 and NLS-jGCaMP7s were used for the experiments in C and E. Figure and description adapted from Qian *et al.,* 2021.

(fluorescence decreases) (Fig 1B). Quantitative analysis of 36 graded potentials from 3 neurons showed that the $-\Delta F/F_0$ of NIR-GECO2 was about half of the $\Delta F/F_0$ of jGCaMP7s following NaCl stimulation ($\Delta F/F_0 = 0.39 \pm 0.19$ (SD) for jGCaMP7s; $-\Delta F/F_0 = 0.19 \pm 0.07$ for NIR-GECO2; Fig 1C). We also quantified signal-to-noise ratio (SNR) of jGCaMP7s and NIR-GECO2 from spontaneously firing neurons in C. elegans. NIR-

GECO2 and jGCaMP7s exhibited similar SNRs (Fig 2C), and neither GECI showed a substantial advantage as a function of imaging depth, consistent with what we found with SBRs (Fig 2E).

We next tested simultaneous optical stimulation and imaging of neural activity using co-expression of the blue light-sensitive channelrhodopsin CoChR (Klapotke *et al.,* 2014) and NIR-GECO2. NIR-GECO2-T2A-HO1 was expressed in the AVA neuron (a command interneuron required for backward locomotion; Chalfie *et al.,* 1985), and CoChR-GFP was expressed in the directly upstream ASH sensory neuron. Imaging transgenic worms with confocal microscopy revealed successful expression of both plasmids (Fig 1D); blue light stimulation of CoChR in ASH neurons was feasible during simultaneous imaging of fluorescent decreases lasting tens of seconds to a few minutes in NIR-GECO2 fluorescence ($-\Delta F/F_0$ of 30% to 90%) in the downstream AVA interneurons (Fig 1E). In contrast to activity evoked through exposure to chemical stimuli, observed patterns of activity in each of the four worms' AVA neurons are highly variable; it is possible that detected activity is spontaneous. Nonetheless, these data indicate that NIR-GECO2 and CoChR expressed in parallel provide a spectrally compatible all-optical method to interrogate hierarchical circuits in *C. elegans*.

2.4 Discussion

In summary, we have developed 2 improved NIR fluorescent Ca^{2+} indicators designated NIR-GECO2 and NIR-GECO2G. Of the 2, NIR-GECO2 has higher response amplitudes but dimmer fluorescence compared to NIR-GECO1, based on characterization in neurons and HeLa cells. In contrast, NIR-GECO2G is improved relative to NIR-GECO1

in terms of both overall cellular brightness (approximately 50% brighter than NIR-GECO1) and sensitivity (up to an approximately 3.7-fold improvement in $-\Delta F/F_0$ relative to NIR-GECO1 for single action potentials). As we have demonstrated in this work, these improvements make the new variants particularly useful for imaging Ca^{2+} dynamics in small model organisms. Specifically, NIR-GECO2 offers comparable sensitivity to jGCaMP7s in *C. elegans*. With their NIR excitation and emission, improved sensitivity, and ability to be subcellularly targeted (e.g., with an NLS), NIR-GECO2 and NIR-GECO2G should prove useful for multicolor and multi-compartment imaging when combined with other fluorescent probes and light-activated neural actuators.We have also demonstrated several of the material advantages of using NIR-GECO2 in *C. elegans* when compared to jGCaMP7s. As listed in the introduction of this chapter, using NIR-GECO2 (1) eliminates exposure of the animals to aversive blue light, (2) reduces the amount of intestinal background fluorescence (5-fold improvement in SBR compared to jGCaMP7s), and (3) is spectrally compatible with blue-light activated optogenetic tools.. Additionally, use of a negative indicator (high fluorescence is present in low calcium ion environments), like NIR-GECO1, might be of use not only for detecting changes in activity, but also for tracking nuclei locations during whole-brain imaging, rather than employing a second, non-dynamic fluorescent signal for tracking purposes, which is the most common current solution (Emmons *et al.*, 2021). The rationale for this concept is that a majority of neurons in the interrogated circuit will be inactive, and therefore fluorescent for a majority of the imaging time, and thus able to be detected and tracked. For a positive indicator like those in the GCaMP series, which fluoresces when Ca^{2+} concentrations are high, GCaMP fluorescence

will not be seen in most neurons, most of the time. This creates the effect of neurons appearing in the dataset only when strongly activated. While this is advantageous for being able to detect activity signals, a second, constitutively fluorescent marker must also be expressed in the same locations to track neuron locations. Neurons labeled with a negative reporter would disappear during few and far-between periods of strong activation, meaning that computational methods for tracking nuclei would need to be thoughtfully designed to appropriately track neurons that occasionally disappear from imaging data. Computational efforts towards tracking transient fluorescent signals from GECIs in mice and hydra has been described by Lagache *et al.* 2021; their Elastic Motion Correction and Concatenation approach could be one way to test whether the appearing or disappearing nuclear localized GCaMP or NIR-GECO2 signal, respectively, are sufficient to track locations of neural nuclei in freely moving *C. elegans* without the use of a non-dynamic fluorescent signal.

Even with the improvements described in this work, NIR GECIs still face challenges including lower brightness, slower kinetics, and faster photobleaching compared to the state-of-art green and red fluorescent GECIs. For these reasons, it remains challenging to use NIR-GECO2 and NIR-GECO2G to image Ca^{2+} dynamics with single-cell resolution in rodents. While the off-kinetics of NIR-GECOs are very similar to that of GCaMP6s or jGCaMP7s, the on-kinetics of NIR-GECO series are slower than that of GCaMP series. These slower on-kinetics may mean that closely spaced Ca^{2+} peaks, which could be resolved with a GCaMP variant, will appear to merge into a single peak with NIR-GECO variants. It also remains to be tested whether any increased thermal energy from the prolonged use of infrared light has adverse effects for *C. elegans* or other organisms expressing NIR-GECO2,

and whether expression of the detector has any effect on the overall health of integrated lines, including physiology, development, reproduction, and behavior. Finally, photobleaching of NIR-GECO2(G) was not a major limitation for the 1-photon imaging experiments reported here, but may be a concern for long duration 1-photon imaging experiments or other types of imaging techniques (such as photoacoustic imaging or swept confocally-aligned planar excitation (SCAPE) microscopy) where strong illumination power is required. Overcoming these challenges will undoubtedly require further directed molecular evolution and optimization of the NIR-GECO series or the possible development of alternative NIR GECI designs based on brighter and more photostable NIR FP scaffolds.

2.5 Contributions

Yong Qian performed mutagenesis, molecular cloning, imaging, and all other procedures involved in directed evolution and characterization of the NIR-GECO2 variants. I generated the transgenic lines and acquired structural and functional imaging data for the characterization of NIR-GECO2 variants in vivo in *C. elegans*. Orhan Çeliker performed imaging data analysis.

3 STochastic Arrangement of Reagents in Clusters (STARC) Enables Neural Compartment Activity Imaging in *C. elegans*

3.1 Introduction

The use of fluorescent calcium ion (Ca^{2+}) indicators in a neuron-dense space poses a challenge to monitoring single-cell resolution changes in neural activity. In many organisms, it is common practice to express the indicator in a subset of neurons, just to enable single-cell resolution level analysis. *C. elegans* is a particularly interesting counterexample. Because of the small total number of neurons, whole-brain imaging is increasingly expected in systems-level work (Kato *et al.,* 2015, Nguyen *et al.* 2016, Linder *et al.* 2016, Venkatachalam *et al.* 2016, Nguyen *et al.* 2017, Halinen *et al.,* 2021, Yu *et al.,* 2021), despite the fact that unlike other organisms' neuropil, worm head ganglia are crowded with cell bodies, with few intervening neurites (except in the nerve ring). The accepted solution of imaging a nuclear localized Ca^{2+} indicator (e.g., NLS-GCaMP) is made less useful by the fact that nuclear localized indicators experience slowed dynamics in comparison to their cytosolically-expressed counterparts (Shemesh *et al.,* 2020). Worse, there is mounting evidence that in both *C. elegans* and other organisms, important computations may be happening in neurites far from the cell body (Hendricks *et al.,* 2012, Donato *et al.,* 2019, Moyle *et al.,* 2021, Brittin *et al.,* 2021, Ruach *et al.,* 2022). In *C. elegans*, an area of interest for localized computation is known as the nerve ring, where parallel projections from interneurons (implicated in such cognitive functions as sensory integration, decision-making, and learning) make contact before many projecting

further or forming synapses to other areas of the worm. This means that nuclear localized Ca^{2+} indicators, the only tools currently enabling whole-brain imaging with single-cell resolution in a neurally dense environment like *C. elegans*, is likely insufficient for capturing several computationally relevant changes in Ca^{2+} concentration, even when expressed and imaged pan-neuronally.

We developed a method for spatially discretizing neural activity reporters by concentrating them into puncta and distributing those puncta throughout the neuron, a strategy we've termed "stochastic arrangement of reagents in clusters" (STARC). STARC-based GCaMP6f has been analyzed for brightness, sensitivity, and kinetics with results indicating minimal differences from cytosolic GCaMP6f (Fig 3). Work has been done to demonstrate its usefulness in cultured cells and in vivo in mice, zebrafish, and *C. elegans*. In this chapter, I demonstrate the functionality of STARC-GCaMP in *C. elegans* neurons and the protocol for later expansion and neurite tracing. This tool allows for monitoring activity in multiple neurons in neurite compartments and holds potential to do so in tightly crowded spaces like the nerve ring.

3.2 Materials & Methods

Worms were cultured and maintained following standard protocols (Brenner *et al.,* 1974). The genes of GCaMP7f, mCherry, and all STARC variant sequences, including w11, w14, and O3-33 (all unpublished variants described in this thesis), for expression in *C. elegans* were codon-optimized using SnapGene codon-optimization tool, then synthesized and cloned by Epoch. Transgenic worms expressing variants of STARC-GCaMP7f pan-neuronally were generated by co-injecting the plasmids tag-168::(STARC variant)-GCaMP7f

and tag-168::mCherry or injecting the single plasmid tag-168:: (STARC variant)-GCaMP7f-T2A-mCherry into the gonads of N2 background worms, picking those with the strongest expression of red fluorescence in neurons.

Hermaphrodite transgenic worms were picked at L4 stage of development and put onto NGM plates with freshly seeded OP50 lawns 12 to 24 hours before experiments. Worms were mounted on 5% low-melt agarose pads on microscope slides, immobilized with 5 mM tetramisole, covered by a coverslip, and imaged using a Nikon Eclipse Ti inverted microscope equipped with a confocal spinning disk (CSU-W1), a 40×, 1.15 NA water-immersion objective, and a 5.5 Zyla camera (Andor, Belfast, Northern Ireland), controlled by NIS-Elements AR software. For 1:500 diacetyl stimulation, worms were imaged using the same optical setup as above, using a microfluidic device that was described previously (Chronis *et al.,* 2007).

All images in the manuscript were processed and analyzed using either ImageJ (NIH) or NIS-Elements Advanced Research software (Nikon). Traces and graphs were generated using Matlab.

3.3 Results

Several variants of STARC were designed by Or Shemesh and Chanyang Linghu by fusing the sequences of different self-assembling protein fragments with varying geometries to the sequence of GCaMP6f. Characterization of one variant was performed by comparing 908 puncta from 4 cultured hippocampal neurons expressing STARC- GCaMP6f to 908 ROIs of comparable sizes from 6 cultured hippocampal neurons expressing cytosolic GCaMP6f in vitro. The ROIs were analyzed for brightness, change in fluorescence (dF/F_0)

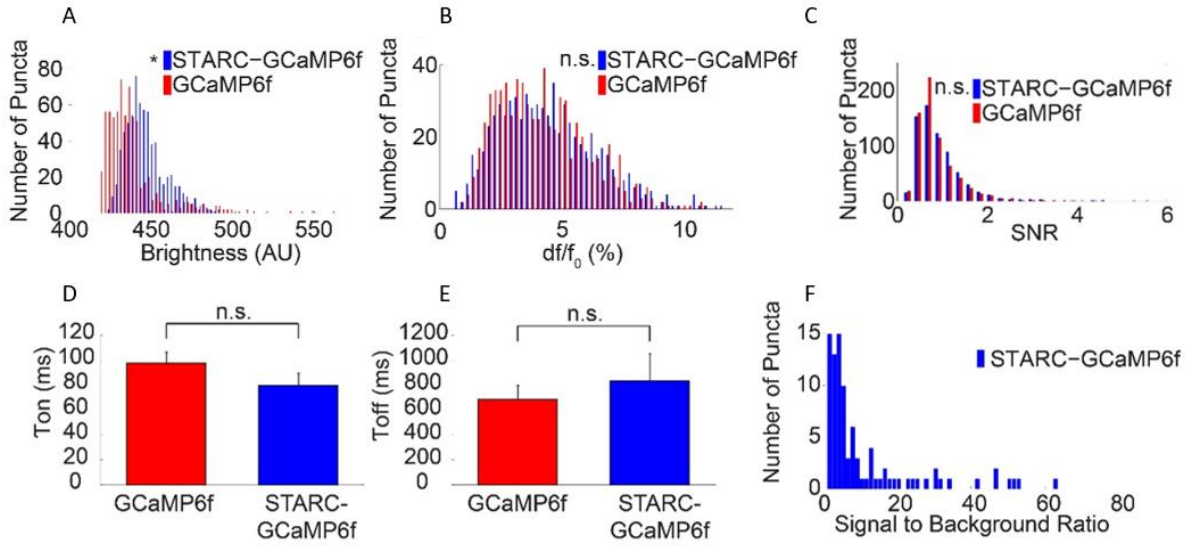

Figure 3: STARC-GCaMP6f has similar fluorescent and kinetic properties and improved brightness compared to cytosolic GCaMP6f in cultured hippocampal neurons. A) Brightness in STARC-GCaMP6f ROIs was significantly higher than in GCaMP6f ROIs according to Kolmogorov-Smirnov statistical tests. The comparisons of df/f_0 during an action potential (B), signal to noise ratio (SNR, C), average fluorescent rise time (T_{on}, D), and average fluorescent decay time (T_{off}, E) show no significant differences in Kolmogorov-Smirnov (histograms) or Wilcoxon rank sum (bar charts) tests. N=908 puncta from 4 cells expressing STARC-GCaMP6f and 908 ROIs of comparable size from 6 cells expressing cytosolic GCaMP6f in vitro (A-E). F) Average signal to background ratio following an action potential (SBR) for STARC-GCaMP6f. N=100 puncta in 4 cells. Figures courtesy of Or Shemesh and Changyang Linghu, unpublished.

during an evoked action potential, signal to noise ratio (SNR, defined as the magnitude of fluorescence change caused by a single action potential divided by the standard deviation of the baseline fluorescence), and rising/falling kinetics (T_{on}, T_{off}) in vitro. No significant differences from cytosolic GCaMP6f were found via Kolmogorov-Smirnov or Wilcoxon rank sum tests for any of these properties, except in brightness, which was higher in STARC-GCaMP6f ROIs (Fig 3, A-E).

When testing whether STARC-GCaMP would be suitable for in vivo imaging of neuronal activity, we found that different geometric configurations of the molecule resulting in puncta of various sizes were best suited to different model systems. We therefore created

non-integrated lines of *C. elegans* for several STARC variants. To determine whether a variant of STARC was useful, the number, size, and location of puncta were evaluated. Three STARC variants were viable, meaning they displayed many small puncta that migrated away from the cell body to far-reaching areas of the neurites. We specifically sought to identify strains where puncta were observed in the far-reaching dendrites of the ciliated neurons, which extend projections to the nose tip of the worm. These variants, in order of suitability based on above criteria, were w11 (uses a 24-subunit cubic cage complex with octahedral symmetry; Lai *et al.*, 2014), w14 (uses a fiber complex with infinite potential subunit growth; data not shown), and 03-33 (uses a 24-subunit octahedral cage complex with octahedral symmetry; King *et al.*, 2012). The w11 and 03-33 variants are shown in Figure 4 A and B, respectively, and were used to test functional imaging.

For functional imaging of STARC-GCaMP, spontaneous and stimulus-evoked activity were evaluated. For spontaneous activity, the fluorescence of STARC-GCaMP puncta in the nose tip were imaged during immobilization with 5mM tetramisole (Fig 4C). For stimulus-evoked activity, microfluidic chips (Chronis *et al.,* 2007) were used to deliver a 10-second appetitive stimulus (1:500 diacetyl) to individual worms, and fluorescence was imaged in head puncta (Fig 4D). Diacetyl was chosen as a chemical stimulus with a robust electrical response in the neuron AWA (mediated by odr-10; Sengupta *et al.,* 1996), which extends a dendrite into the nose tip, where we were seeking puncta. Increases in fluorescence were observed in both experimental setups. Changes in fluorescence were not uniform across puncta in spontaneous or stimulated experimental conditions, even when it seemed that puncta were physically near each other. Ex-vivo tracing of neurites was not performed,

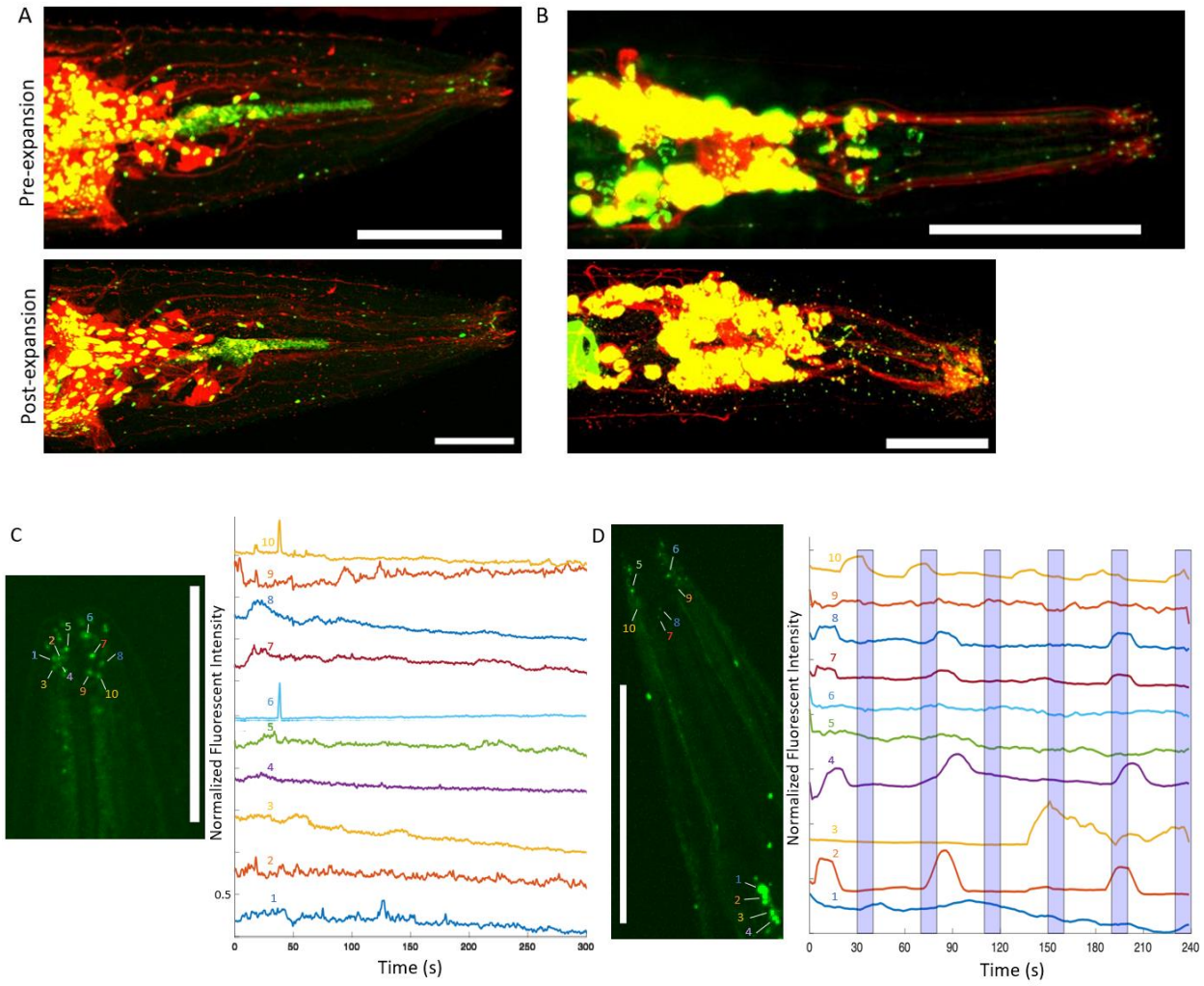

Figure 4: STARC-GCaMP6f variants detect spontaneous activity and are compatible with use in microfluidic chambers in *C. elegans* in vivo. Representative fluorescent image of neurons expressing STARC variant w11 (A, 24-subunit cubic cage protein complex with octahedral symmetry), and STARC variant 03-33 (B, 24-subunit octahedral cage protein complex with octahedral symmetry), pre- and post-expansion. C) Representative traces show spontaneous activity in puncta shown at left (variant 03-33 STARC-GCaMP6f). D) Representative traces in microfluidic chamber, during stimulation with 1:500 diacetyl in M9, in puncta shown at left (variant w11 STARC-GCaMP6f). Blue bars indicate the time points of stimulation (10s pulses with 30s rest in between). Mean fluorescent intensity in ROIs surrounding puncta were normalized by maximum value. All scale bars are 50μm. Puncta in C and D may be in the distal dendrites of any of the ciliated neurons, which includes AWA, responsible for sensing diacetyl (see Inglis *et al.* 2005-2018 for a complete list)

but due to the locations of the puncta, we may assume that the neurons whose activity is recorded are among the ciliated neurons whose dendrites extend to the nose tip (Inglis *et al.* 2005-2018 contains a complete list). While we did not observe robust stimulus-evoked

activity in the selected puncta, these data indicate that STARC-GCaMP can provide a detailed and dynamic perspective on spontaneous activity in *C. elegans* neurites in multiple experimental conditions.

We finally developed a methodological workflow for the purpose of interrogating specific circuits in C. elegans. This consists of the following steps:

1) expression of STARC-GCaMP (we used pan-neuronal expression vectors, though ideal experiments would drive expression in 2-5 neurons classes of interest)
2) live imaging of spontaneous or stimulated neural activity
3) recovery and fixation of worms in separate tubes to track worm identity
4) expansion and antibody staining, and
5) neurite tracing to the relevant cell body for cell ID.

The need to keep track of an individual worm through imaging and expansion sometimes led to the loss of the worm. Worms were most often lost in step 3 after microfluidic imaging because of the need to track the worm through liquid outflow passages or in step 4 where the *C. elegans* expansion protocol required dozens of washes in sometimes viscous liquid in large volumes (1mL) compared to the size of a single worm. Overall, this pipeline resulted in a worm retention rate of approximately 65%.

3.4 Discussion

In summary, we have developed several STARC variants that have demonstrated success in clustering GCaMP6 in cultured hippocampal neurons and in *C. elegans*. STARC variants w11 and 03-33 have demonstrated function in detecting spontaneous changes in Ca^{2+} concentration in vivo in *C. elegans* both during immobilization via chemical reagents

and mechanically, in a microfluidic chamber. There is no difference in kinetic properties or SNR of GCaMP6 when expressed cytosolically or in STARC, though STARC-GCaMP6f showed significantly brighter peak fluorescent signal. This increased brightness is likely due to the close configuration of several GCaMP molecules when self-assembled into clusters by the fused STARC sequences. Cytosolic GCaMP molecules, in contrast, are likely to diffuse relatively evenly throughout the cell. A smaller number of molecules in a comparable physical space, therefore results in a slightly dimmer signal. This result is encouraging for the applicability of identified STARC variant sequences to be applied to new fluorescent tools which have not been optimized for brightness to the same degree that GCaMP6 has, and may be too dim for practical use without clustering.

STARC-GCaMP also provides several solutions particularly useful for Ca^{2+} activity imaging in *C. elegans*. The current standard for whole-brain Ca^{2+} imaging in C. elegans is pan-neuronally expressed NLS-GCaMP. Because indicator kinetics are slightly slowed when localized to the nucleus, and compartmentalization is likely in this organism, NLS-GCaMP is ill-suited to capture Ca^{2+} dynamics that underlie important computations. Unfortunately, the distribution of puncta in STARC-GCaMP is not sparse enough in the *C. elegans* soma to serve as a replacement for NLS-GCaMP pan-neuronally. As can be seen in Figure 4A-B, STARC-GCaMP clusters in the soma are more ubiquitous and larger than in neurites, resembling cytosolic GCaMP, and in order to visualize sparse puncta in the neurites, brightness levels must be set so that clusters in the soma are oversaturated. However, STARC-GCaMP is likely well-suited for interrogating compartmentalized Ca^{2+} activity in *C. elegans* circuits as evidenced by the transportation of puncta to the far-reaching tips of

dendrites in the ciliated neurons. When expressed more sparsely, this could provide insight to computation in spatially constrained areas such as the nerve ring. For example, STARC-GCaMP could be expressed in the interneuron class AIY, in which calcium transients have been observed in the neurites, but not in the soma (Clark *et al.,* 2006, Chalasani *et al.,* 2007), and candidate neurons such as AFD and AWC that contribute to these transients in response to temperature and odor stimuli. Another example experiment may be expression in the RIA interneuron and SMD motor neuron classes. These are known to communicate during head bending (Hendricks *et al.,* 2012, Liu *et al.,* 2018), potentially in such a way that the dorsal and ventral SMD motor neurons communicate motor commands upstream to the RIA interneurons to aid in identification of self-generated movements (Donato *et al.,* 2019). Investigating these relationships with STARC-GCaMP could provide the opportunity to capture compartmentalized Ca^{2+} dynamics in feedforward and feedback directions in multiple neurons in the circuit, at multiple points along the neurites, unlike current methods which have limited fluorescent imaging to a single neuron, forcing guesses as to pre- or post-synaptic activity.

Finally, several improvements can be made to the design and methodology for STARC-GCaMP in C. elegans. Other self-assembling peptide sequences may be tested for even smaller and more sparsely distributed puncta in the soma, which would enable the use of STARC-GCaMP pan-neuronally in *C. elegans*, though we believe the variants characterized here are sufficient for current applications. We have also noted that different self-assembling peptide sequences are best suited to different models and organisms (not shown). It also remains to be seen whether integrated worm lines (yet to be created) expressing any STARC

variants display issues in the overall health of the worm, including physiology, development, reproduction, and behavior. One of the most significant drawbacks of the methodology is the attrition rate associated with tracking individual worms for this protocol. There are several interventions now possible to improve the worm retention rate during steps 3 and 4 of the experimental workflow. The first, and simplest, is to choose live imaging paradigms in which it is easy to recover the worm. Because we performed stimulated microfluidic imaging in silicone chips, our only choice for worm recovery was to follow the worm through egress channels of the chip system and hope to capture it with waste liquid before washing and fixation. Imaging on a slide with a chemical immobilization agent made worm recovery very simple, as would a freely-behaving imaging setup. This work was also completed with early versions of the expansion for *C. elegans* (ExCel) protocol (our latest published version is Yu *et al.,* 2022), which has since been improved upon. The newest protocol (in preparation for publication by Yangning Lu and Chi Zhang) involves almost half as many wash steps, which were the steps with highest potential for losing track of a worm. By implementing both of these changes we believe worm retention would be greatly improved.

3.5 Contributions

Or Shemesh and Changyang Linghu designed STARC variants and performed molecular cloning, imaging, and all other procedures involved in generation and characterization of the STARC-GCaMP6f variants in vitro. I generated transgenic lines and acquired structural and functional imaging data for the characterization of STARC-GCaMP6f variants in vivo in *C. elegans*. Jay Yu performed expansion of worms, and Orhan Çeliker performed imaging data analysis.

4 Endogenous Barcoding (EnBarc): A Novel, Customizable Approach for Cell Identification

4.1 Introduction

Whole-brain Ca^{2+} imaging has become an increasingly popular method in *C. elegans* systems neuroscience. Understandably so, as the organism boasts a tractable 302 neurons, sorted into only 118 classes, and a known canonical connectome (White *et al.,* 1986). With connectivity data, a tractable number of cells, and molecular tools to enable whole-brain imaging, one might imagine the possibility of being able to fully describe how global neural activity drives behavior. However, analyses thus far have been largely limited to the same techniques used in larger animals in which only a subset of the neural population is recorded, where single cell data is consolidated into dimensionally reduced analyses rather than piloting analyses which take into account the completeness of a *C. elegans* dataset (Kato *et al.,* 2015, Nguyen *et al.* 2016, Linder *et al.* 2016, Venkatachalam *et al.* 2016, Nguyen *et al.* 2017, Halinen *et al.,* 2021, Yu *et al.,* 2021). This is partially due to the technological difficulty of assigning cellular identity to whole-brain images of the tightly packed worm head. Traditional methods of neural ID require discernment of neuronal morphology, familiarity with variable cell body position, and several hours for manual annotation of a single worm.

Yemini *et al.* (2019) offer a solution in their NeuroPAL strain: a deterministic "brainbow" approach which identifies neuron class by the differential expression of fluorescent reporters. However, this limits investigators to use a highly specialized strain of one lab's design, in which the expression of large fluorescent proteins driven by over 40 different promoters has an unknown impact on behavior. This limits the advantage *C. elegans*

researchers enjoy of quickly and flexibly being able to test their hypotheses on strains of their own design. Further, many *C. elegans* researchers have informally expressed concerns about the "dumpy" morphological phenotype, slow growth rate, and the implications of these for the worms' overall health and behavioral reliability. Still, this is the only solution currently available to annotate a whole-brain Ca^{2+} activity dataset with neuronal cell ID.

There are numerous other potential applications for methods that enable accessible, flexible, and reliable methods of cell ID in C. elegans that don't require the use of crosses or expression of exogenous genes. These include studies of dynamic gene expression patterns in adulthood and throughout development as well as studies of cell fate and specification. Our goal in this project is to introduce a strategy to identify all 118 neuron classes in a way that:

1. does not require a long researcher training process to obtain individual knowledge of morphology and position of all neuronal classes,
2. does not require any specific set of optical filters or equipment to apply the technique,
3. could be used in strains researchers designed for their own behavioral, genetic, and/or imaging experiments without performing crosses,
4. can be easily customized to accommodate researchers' resources or for special cases of cell ID,
5. uses analysis code that is accessible even to researchers without a strong computer science background, and
6. could be adapted to different model organisms for other applications of cell ID.

In the work described herein, we made significant steps to achieve this by detecting differential expression of endogenous RNAs via hybridized chain reaction RNA fluorescent in situ hybridization (HCR-RNA-FISH) in fixed *C. elegans*. This is a technique we conceptualize as reading out the "endogenous barcode" (EnBarc) of each cell type.

Work for this technique is ongoing, and a detailed protocol and final results will be completed and shared after my defense in a separate publication. This chapter outlines the strategy taken to complete this project and preliminary data that lead us to believe that the work will be completed shortly.

4.2 Materials & Methods

Worms were cultured and maintained following standard protocols (Brenner *et al.,* 1974). Strains used in HCR-RNA-FISH screening, single-round multiplexed HCR-RNA-FISH, and gel-embedded multi-round HCR-RNA-FISH were N2 (https://cgc.umn.edu/strain/N2), AML32 (https://cgc.umn.edu/strain/AML32), or OH15265 (https://cgc.umn.edu/strain/OH15265), ordered from the Caenorhabditis Genetics Center (CGC).

Custom HCR probe sets, amplifiers, and required buffers were ordered from Molecular Instruments (MI, https://store.molecularinstruments.com/new-bundle/rna-fish). Detection stage and Amplification stage protocols for whole-mount *C. elegans* larvae from MI were followed for probe hybridization and amplification for all HCR-RNA-FISH experiments (https://files.molecularinstruments.com/MI-Protocol-RNAFISH-Nematode-Rev7.pdf). To aid in spinning down worms in viscous liquids, up to 500uL of the next step's less-viscous buffer (Probe Wash Buffer or SSCT) was added as needed.

For HCR-RNA-FISH screening of single gene HCR probe sets, mixed-age populations of hermaphrodite worms were collected, fixed, and cuticles reduced as described in Yu et al. 2022. Worm cuticles were partially digested by incubating in 1:800 Proteinase K in PBS for 15m at room temperature. HCR-RNA-FISH was performed in 1.5mL tubes and imaging was done in 96-well glass-bottom plates.

For single-round and multi-round multiplexed HCR-RNA-FISH, mixed-age populations of hermaphrodite worms were collected in 1.5mL tubes with M9 and placed on ice for 1h. Chilled worms were fixed in pre-chilled 4% paraformaldehyde in PBS at 4°C for 4h, washed and stored in PBS. Worms were embedded in a hydrogel for cuticle digestion with gel-anchored Proteinase K as described in Lu and Zhang et al., 2022. HCR-RNA-FISH was performed on gels as above in 24-well glass-bottom plates, where imaging was also performed. For multi-round experiments, HCR amplifiers were removed by the application of toehold reversal strands at 10x concentration of the concentration of amplifier. Reversal strands and reversible amplifiers were provided upon custom request by Molecular Instruments and performed according to MI's suggested protocol of incubation in 5x SSCT for 1h and subsequent washes.

All experiments were imaged using a Nikon Eclipse Ti inverted microscope equipped with a confocal spinning disk (CSU-W1), a 40×, 1.15 NA water-immersion objective, and a 5.5 Zyla camera (Andor, Belfast, Northern Ireland), controlled by NIS-Elements AR software.

All images were processed and analyzed using either ImageJ (NIH) or NIS-Elements Advanced Research software (Nikon). Traces and graphs were generated using GraphPad prism 8, Origin (OriginLab, Wellesley, Massachusetts, USA), and Matlab.

For SNR comparisons, three measurements were taken from each of three worms in three independent experiments, for a total of 27 measurements per gene probe set. ROIs were chosen by manual scoring of nuclei as "positive" or "negative" based on visible presence of HCR-RNA-FISH puncta signal. Average fluorescent intensity from an ROI drawn around a "positive" nucleus was divided by average fluorescence from an ROI of equal size around a "negative" nucleus to generate internal-internal SNR, or an ROI of equal size placed arbitrarily outside the worm's outline to calculate internal-external SNR.

4.3 Results

EnBarc for cell ID involves three phases. First, combinatorial cell class barcode candidates are generated by use of a genetic algorithm to search an extraordinarily large, public single-cell RNA sequencing data set. Second, the final gene set to determine class barcodes is chosen by experimental screening of HCR-RNA-FISH probes, and neural barcodes in fixed *C. elegans* are detected by sequential rounds of 3-color multiplexed HCR-RNA-FISH. Finally, barcode readout and probabilistic cell ID labels are produced by semi-automated analysis code. In this results section, each phase will be described theoretically and progress toward its completion will be presented.

4.3.1 Computational Barcode Generation

In order to identify neuron classes by their mRNA expression, we sought to generate short endogenous barcodes for each class. An endogenous barcode differs between cell

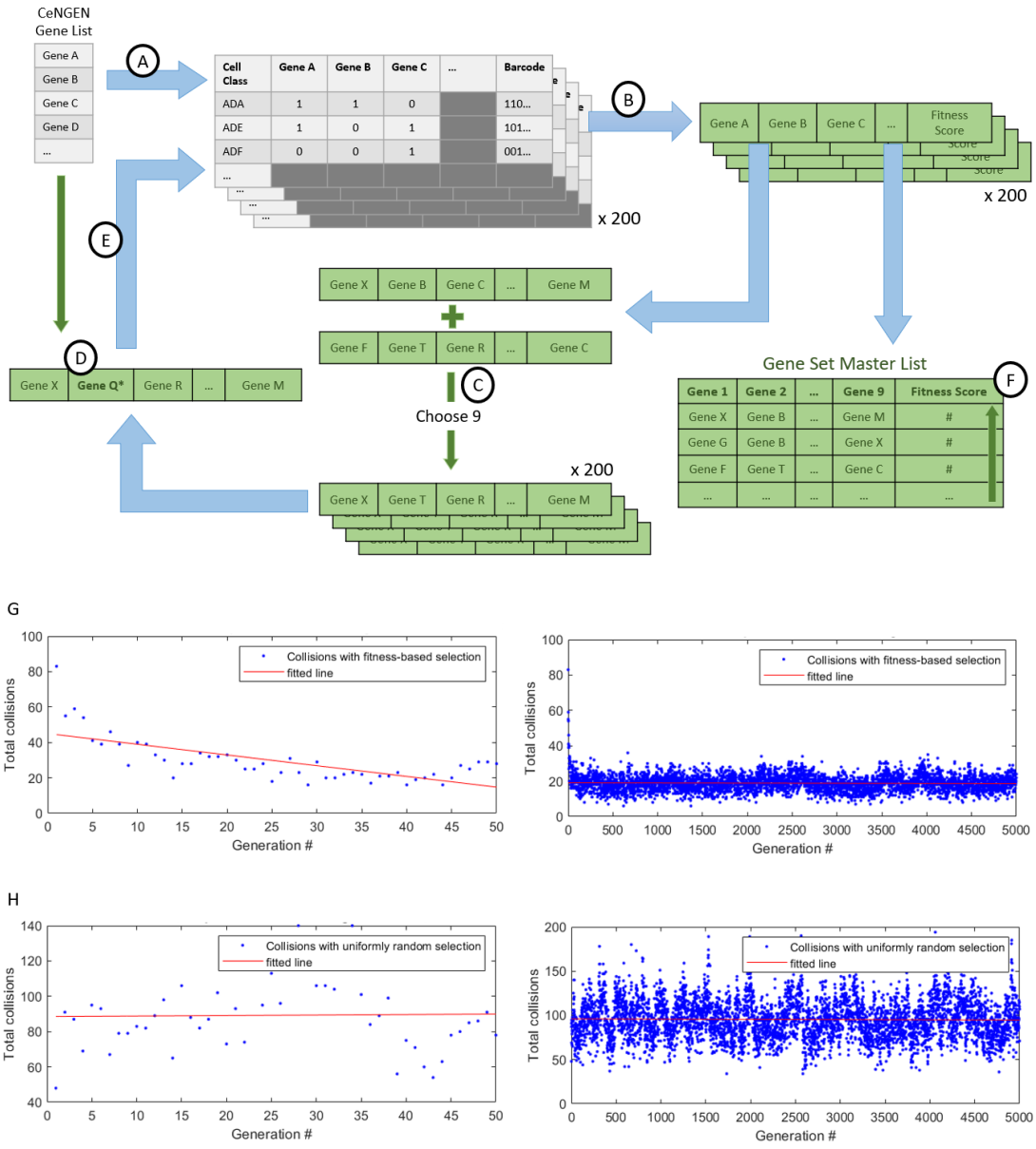

Figure 5: EnBarc gene selection algorithm for barcode determination. A) Randomly select 200 sets of 9 genes from CeNGEN gene list, generate cell class barcodes for each set. B) Calculate fitness score for each gene set by considering number of collisions and all-zero barcodes. Fitness score is saved alongside gene set in "master list" of all evaluated gene sets and serves as probability for the gene set to be recombined. C) Choose gene sets with probability decided by fitness, pair at random, and recombine to generate "offspring" set. Recombination was done by choosing 9 genes (without replacement) from the list of unique genes in two paired "parent" gene sets to produce offspring gene sets. 200 offspring were produced per "generation". D) "Spontaneous mutation" occurs with 10% probability for each offspring gene set, implemented by replacing one gene in the set with a randomly selected gene from the CeNGEN gene list. E) Repeat steps B-D, calculating and recording fitness, recombining the best gene sets, and introducing random mutations at a low probability for approximately 2000 generations. F) The master list of all evaluated gene sets is sorted by fitness to present the best gene sets for generating cell class barcodes. G-H) Representative total collision results in each generation of the gene selection algorithm, using fitness-selective recombination (G) or random recombination (H), run for 50 generations (left) and 5,000 generations (right).

classes and consists of a string of "0"s and "1"s denoting absence or presence, respectively, of detected mRNA for an ordered list of genes (Fig 5A). The list of genes, and therefore the barcodes, should be as short as possible to simplify experimental readout. In order to allow for 118 unique barcodes in combinatorial space and maximize the information gained in each 3-color round of HCR, we chose to use a set of 9 genes to form barcodes. This was because 6 genes (read out over two 3-color rounds of multiplexed HCR-RNA-FISH, Fig 7) would not allow for 118 barcodes (2^6=64 unique combinations), but another full round (for a total of three 3-color multiplexed HCR-RNA-FISH detection rounds) would suffice (2^9=512 unique combinations) and provide some extra expression information to robustly make cell ID calls.

Given the barcode length (9), we next had to assemble data and from which we would choose the genes to produce the barcodes. We selected sets of 9 genes from the publicly available CeNGEN single cell RNA sequencing data set (Taylor 2021) as a basis for expected presence or absence of mRNA, and incorporated approximate neuron location data from the OpenWorm NeuroML *C. elegans* Connectome model, c302 (Gleeson 2018; accessed via https://github.com/openworm/CElegansNeuroML). Our most important constraint in searching for an ideal gene set to determine barcodes was that two neighboring neurons should not have the same barcode, a phenomenon we termed a "collision." Neurons at least one neighbor away as determined by the model's location data could be assigned the same barcode, because location information would be considered to make final ID calls (Fig 11D). Another result we looked to avoid was "all zero" barcodes, or barcodes which indicated that none of the 9 genes in the barcode gene list were expected to be

Table 1: Top 30 "fittest" gene sets and the most frequently observed genes (highlighted).

Rank	Fitness Score	Gene 1	Gene 2	Gene 3	Gene 4	Gene 5	Gene 6	Gene 7	Gene 8	Gene 9
1	29	*unc-103*	*zfh-2*	*ddl-2*	*inx-7*	*acc-4*	*inx-19*	*ceh-32*	*zag-1*	*inx-2*
2	31	inx-2	acc-4	inx-1b	rig-6	zag-1	inx-7	zfh-2	nsy-7	vab-3
3	33	dkf-2	goa-1	glr-4	nsy-7	inx-7	acc-4	unc-7	unc-8	zag-1
4	33	zfh-2	unc-47	nlp-14	inx-2	nsy-7	unc-17	inx-7	unc-7	inx-1b
5	33	inx-2	inx-1a	inx-7	zig-5	unc-46	nlp-6	zfh-2	zag-1	nsy-7
6	34	inx-1a	zfh-2	glr-1	unc-7	unc-5	ntr-2	ceh-18	inx-7	zag-1
7	34	inx-2	unc-8	nhr-67	ceh-32	inx-1b	zag-1	unc-47	inx-7	pdfr-1
8	34	sra-11	ceh-18	zag-1	nsy-7	inx-7	cog-1	inx-1a	acc-4	dkf-2
9	34	nsy-7	inx-7	cam-1	eat-4	inx-18a	unc-7	acc-4	inx-1b	inx-2
10	34	sgk-1	unc-103	rig-6	zfh-2	inx-1b	nsy-7	inx-7	inx-2	inx-18a
11	34	unc-17	unc-103	inx-7	zfh-2	unc-46	inx-1a	nsy-7	unc-8	zag-1
12	34	inx-1b	ceh-19	unc-47	inx-2	inx-7	cam-1	acc-4	zfh-2	nsy-7
13	35	nsy-7	ahr-1	inx-1a	glr-4	zfh-2	inx-14	inx-7	zag-1	inx-2
14	35	inx-7	ceh-18	zag-1	unc-7	inx-19	inx-1a	rig-6	egl-4	inx-2
15	35	inx-18a	zfh-2	ceh-32	nlp-6	cho-1	inx-7	zag-1	glr-2	inx-1b
16	35	ceh-32	nsy-7	unc-62	zfh-2	inx-7	inx-2	inx-14	ceh-18	zip-4
17	35	unc-8	unc-7	zag-1	unc-47	inx-7	inx-2	inx-1a	mig-5	zfh-2
18	35	unc-7	ceh-32	inx-2	ser-1	grd-8	inx-7	flp-18	zfh-2	inx-1a
19	35	cam-1	unc-7	ser-1	glr-4	inx-1b	osm-6	npr-1	inx-2	zag-1
20	35	inx-1b	ags-3	mir-124	inx-2	glr-5	zag-1	ceh-32	cam-1	zfh-2
21	36	ceh-32	nlp-3	nlp-6	zag-1	zfh-2	inx-1a	srv-32	inx-2	inx-7
22	36	unc-7	acc-4	npr-1	zfh-2	inx-2	inx-1a	zag-1	ser-2	nmr-1
23	36	inx-2	unc-103	inx-7	unc-7	glr-2	flp-18	ceh-18	rig-5	inx-1b
24	36	tmc-1	nsy-7	ceh-14	unc-17	unc-7	zfh-2	inx-7	unc-47	inx-2
25	36	acc-4	inx-1b	acr-5	inx-2	ceh-18	zfh-2	zag-1	rig-6	nsy-7
26	36	egl-4	inx-2	inx-7	acc-4	nsy-7	cct-4	inx-18a	ztf-3	zfh-2
27	36	ceh-13	inx-1b	ceh-32	inx-7	zag-1	kel-8	zfh-2	cam-1	inx-2
28	36	zfh-2	inx-7	ceh-32	zag-1	inx-2	nsy-7	srg-13	dkf-2	inx-1b
29	36	unc-7	wrk-1	inx-2	acc-4	zfh-2	glr-4	ceh-30	npr-1	nsy-7
30	36	cam-1	inx-7	inx-1b	zig-1	pdfr-1	zag-1	unc-8	zfh-2	unc-7

detected in a neuron. This was experimentally undesirable, as it would be difficult to discern an expected lack of fluorescence from an experimentally-faulty lack of fluorescence.

With these data and design constraints in place, we were next tasked with finding a way to search the combination space of possible 9-gene sets, though there were two significant

challenges. First, this search space was massive. With 3200 genes of expression data available in the CeNGEN dataset, there were 9.6e25 possible unique gene set combinations. Our relatively powerful desktop computer with 10 cores and 64 gigabytes of RAM computes approximately 8,000 combinations per second, but even if we could compute at a hundred times that speed, it would take 380 billion years to compute through all possible combinations. Another salient challenge was that a data set of genes and their binarized expression is discrete (rather than continuous) and thus non-differentiable, so we can't apply commonly used methods of gradient descent to optimize gene choice. In other words, knowing how well one gene set combination uniquely identifies cells will not smoothly lead us to the next best gene set iteratively until we find a universally optimal set. Worse, inclusion of an individual sub-optimally performing gene can defeat the entire set.

To address these challenges of an extremely large, discrete search space, we chose to evaluate the fitness of a comparatively small number of randomly chosen 9-gene sets to determine barcodes, then apply a what is known as a genetic algorithm to "evolve" better gene sets by "recombination" of the best-performing sets in each "generation", or round of iteration of this algorithm. The design and computational implementation of this algorithm is described in high detail in my colleague, Orhan' Çeliker 's, graduate thesis (Çeliker, 2021); a brief summary of the algorithm is illustrated in Fig 5A-F. This algorithm converges toward its best fitness scores in 50-200 generations; it does not converge if parent gene sets are recombined at random instead of with probability assigned by their fitness (Fig 5G-H). The average total collisions for barcodes in a gene set when using fitness-selective recombination after 5,000 generations was 19.07 with a minimum of 6; when using random recombination

the average total collisions for barcodes in a gene set after 5,000 generations was 96.01 with a minimum of 34.

4.3.2 Experimental Barcode Detection

The above described method for barcode generation results in several gene sets with comparable suitability in regard to number of collisions (Table 1). This means that we were presented with several theoretically ideal options for a single set of 9 genes that could, when

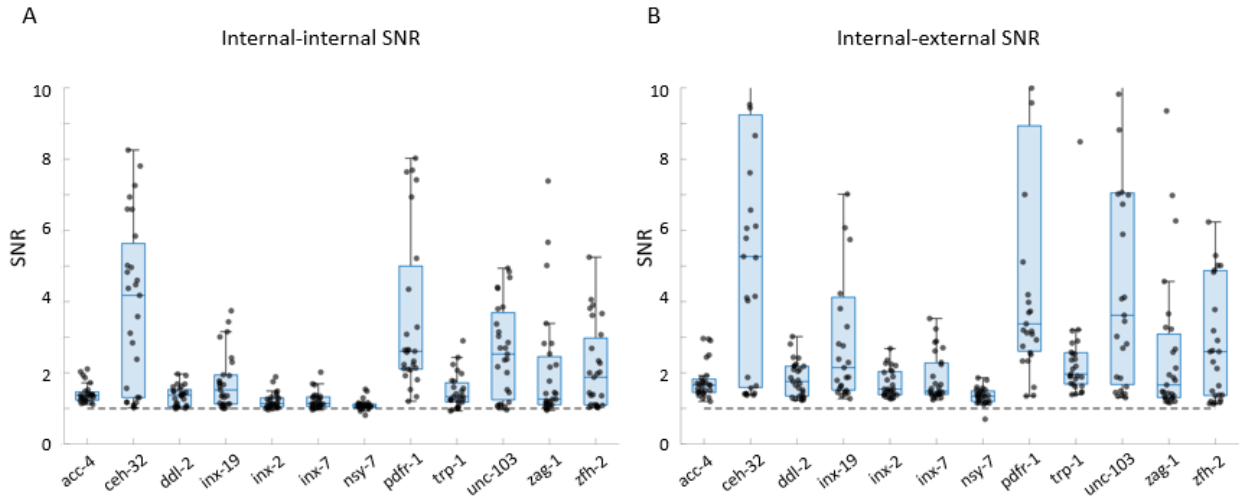

Figure 6: Signal-to-noise ratio analysis in candidate genes used to define endogenous barcodes. Box-and-whisker plots of two types of SNR calculated for each gene considered for use in multiplexed HCR-RNA-FISH for whole-brain cell ID. SNR was calculated by drawing an ROI around a nucleus manually scored as "positive", and dividing fluorescent signal by fluorescent signal in an ROI of equal size, around a nucleus manually scored as "negative" (A, internal-internal SNR) or arbitrarily positioned outside of the body of the worm (B, Internal-external SNR). Measurements from 3 different "positive" neurons per worm and 3 different worms were taken from 3 experimental replications, for a total of 27 measurements per gene, for each type of SNR.

detected as mRNA in *C. elegans* neurons, combinatorically reveal cell ID. However, the suitability of a gene set in experimental terms has a different set of constraints. Namely, the ability to determine the presence or absence of mRNA in a neuron was of highest priority when selecting the final protocol gene set. We therefore sought to select the final set of 9 genes by the best quantified signal to noise ratio (SNR) for mRNA detection via HCR-RNA-

FISH for individual genes. Fortunately, the selection algorithm also resulted in considerable overlap in individual genes present between theoretically most fit gene sets (Table 1, highlight).

We have performed HCR-RNA-FISH on 12 genes thus far to assess their SNR and subjective quality of staining. There were several factors considered to select genes to include in this screen for computational detectability, including frequency of appearance in the 30 fittest gene sets and appearance in the 3 fittest. First, we counted the number of times an individual gene appeared in the 30 theoretically most fit gene sets; inx-7 and zfh-2 appeared in 25 sets, inx-2 appeared in 24 sets, and zag-1 appeared in 10 sets out of the fittest 30 gene sets and were selected for SNR analysis. These genes were also present in the top two fittest sets, as were nsy-7, acc-4 and ceh-32, which were also present in 16, 12 and 9 of the top 30 sets, respectively. The genes unc-103, inx-19, and ddl-2 were selected despite a low frequency in the top 30 sets (appearing in 4, 2, and 1 set, respectively) because they completed the top fittest set. A few other genes were included in this screen because of their frequent appearance in top sets when the genetic selection algorithm was run on an earlier, smaller single-cell RNA sequencing dataset.

We would continue to analyze SNR until all individual genes comprising a gene set were screened, and we could test our analysis code's ability to consistently and accurately determine presence or absence of mRNA. This will allow us to make a recommendation on a lower SNR limit to exclude genes when customizing an EnBarc gene set for detection with HCR-RNA-FISH in any other application (other cell types, other organisms, etc.). The results of SNR analysis for the 12 genes we have screened thus far are shown in Figure 6.

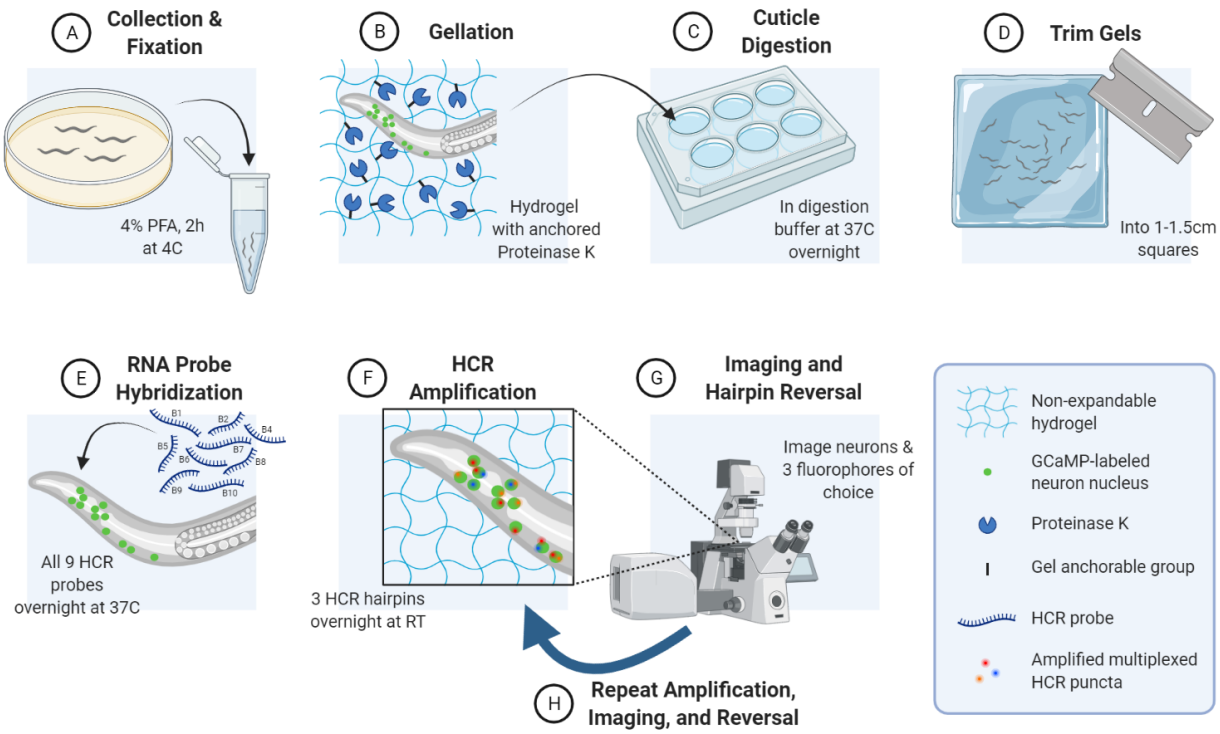

Figure 7: EnBarc experimental protocol for barcode detection. A) Collect worms by washing from growth plates and pipetting into 1.5mL tubes. Wash with M9 until bacteria are no longer visible. Freeze pellet on ice for 1h, then fix in 4% paraformaldehyde at 4°C for 2h. Wash with M9. This can be stored for up to 2 weeks. B) Prepare a gelation chamber as described in Yu et al. 2022. Prepare non-expanding hydrogel solution by adding pre-chilled worms, water, 2x gel stock, AcX, TEMED, Proteinase K as described in Lu et al., 2022. Add APS to begin gelation, pipette thoroughly, and transfer to gelation chamber. Incubate for 30m at 37°C to complete gelation. C) Remove gels from chambers, wash away un-anchored Proteinase K, and incubate in digestion buffer in a 6-well plate, spinning at 100 RPM at 37°C overnight. D) Wash digestion buffer from gels, check worm density, and trim gels to sizes that contain sufficient numbers of worms for the given experiment. Final gels should be no larger than 1.5cm. E) Complete RNA probe hybridization as described in protocols by Molecular Instruments (MI). We hybridize all 9 amplification sequences to be used in the multi-amplification round experiment in a single step, ordered commercially through MI. F) Complete HCR amplification for three of the hybridized probes as described in protocols by MI, using fluorophores of your choice according to settings available in your imaging equipment. We used AlexaFluor 546, 594, and 647, ordered commercially through MI. G) Capture high quality volumetric images of worms in channels and resolution that will resolve neuron nuclei and puncta of your chosen fluorophores. We used a spinning disc confocal imaging with a water immersion 40X objective, taking volumetric stacks of images in slices 1µm or thinner. Imaging using the lowest laser power and exposure times possible to detect your HCR-RNA-FISH puncta will result in better stripping results. Apply toehold reversal probes as described in protocols by MI. H) Repeat HCR amplification, imaging, and reversal steps (F &G) until all FISH probes have been amplified by HCR and imaged. We perform detection in three rounds of 3-color multiplexed HCR, but this experimental protocol can be adjusted to accommodate any at least two-color imaging setup with sufficient power to resolve neural nuclei by adjusting the number of colors multiplexed and the number of rounds of staining, imaging, and stripping.

Fluorescent signal in nuclei manually scored to be positive was compared against fluorescent signal from nuclei manually scored to be negative (internal-internal SNR) or against locations

outside of the worm's body (internal-external SNR). SNR varied widely in most genes screened; this was because HCR-RNA-FISH staining had a "hit or miss" nature with the protocol used, where some experimental replications had more or less internal and external signal or noise than others. One reason for noise may be non-neuronal signal for genes that are expressed in other tissues, such as muscle or intestine. One way to avoid choosing genes with this non-neuronal signal would be to cross-reference gene set output with expression profiles from non-neuronal tissues.

Genes with higher averages and wider ranges were considered to be more ideal. The gene nsy-7 was found to have the worst SNR (low average and low standard deviation) and subjectively low signal in images; sets using this gene were excluded from consideration to set final experimental barcodes. From the fittest set, ceh-32, inx-19, unc-103, and zfh-2 had internal-external SNR averages above 2. Unfortunately, three of the four genes most frequently seen in gene sets were among those with lower internal-external SNR. We chose to proceed with testing the full detection protocol with this gene set, as a way of testing our analysis code's ability to determine presence or absence of mRNA and potentially define a lower limit for SNR. If we were unable to consistently make a presence or absence call for a specific common gene, we could then eliminate gene sets containing the gene in choosing further genes to screen for stain quality.

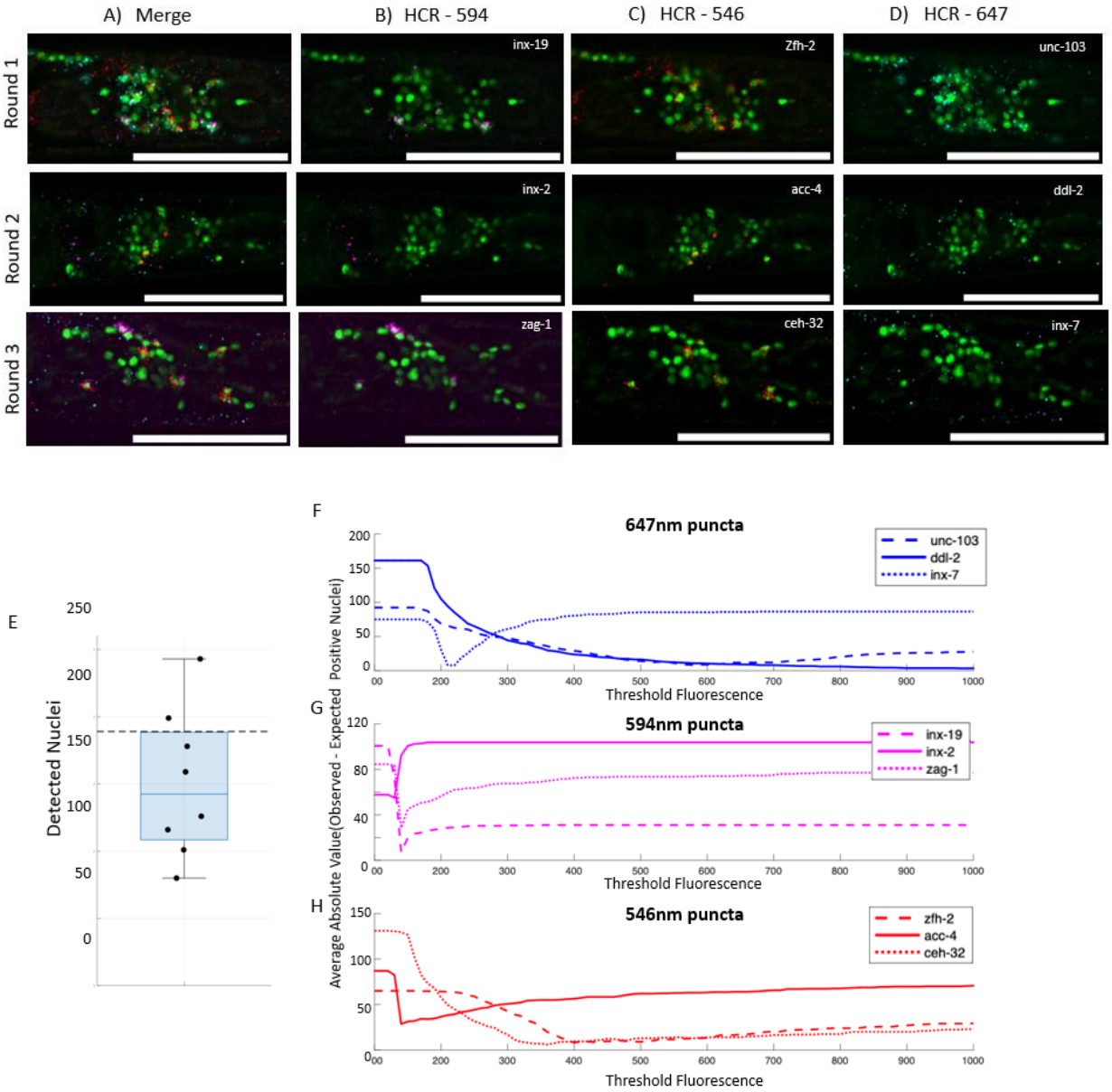

Figure 8: Single-round multiplexed HCR-RNA-FISH. A-D) Representative single slice from volumetric images taken from three-gene multiplexed HCR-RNA-FISH using the EnBarc neural ID genes. A) Each multiplexed "round" of HCR-RNA-FISH for three genes is shown in a separate worm, tiling the entirety of the barcode genes across the three worms and staining conditions. Pan-neuronal nuclear GCaMP (green) is shown together with HCR-RNA-FISH puncta using AlexaFluor 594 (B), AlexaFluor 546 (C), and AlexaFluor 647 (D) fluorophores in each experiment. E) Box-and-whisker plot showing the number of nuclei detected in each worm head (n=8). Dotted line at 189 indicates the expected number of nuclei in the head. F) The average absolute value of the expected number of nuclei positive for a gene (normalized to the number of detected nuclei) subtracted from the observed number of nuclei marked positive for that gene. A value of zero indicates that the number of detected positive nuclei is equal to the number of expected positive nuclei. Blue, pink, and red lines correspond to the fluorophore used in the three rounds of the experiment. N=3 for unc-103, inx-19, zfh-2, ddl-2, inx-2, and acc-4; n=2 for inx-7, xag-1, and ceh-32. All scale bars are 50µm.

To read out the barcodes, we performed multi-round, multiplexed HCR-RNA-FISH. Three genes, detected by three different fluorophores, would be detected in a single round, along with a pan-neuronal nuclear signal. Fluorescent signal would then be stripped and re-stained in a total of three rounds to detect presence or absence of all 9 genes in the barcode for each neural nucleus. The final protocol for this is illustrated in Figure 7.

We took a stepwise approach toward validating this complex protocol. First, we arranged HCR-RNA-FISH individual gene probes into groups of three and stained for three genes at a time with three different fluorophores for amplification in a single round, in different populations of worms for each round (Fig 8A-D). This showed that we indeed did see differential staining for each gene around different nuclei. In these experiments, an average of 150.3 ± 54.7 nuclei in the head were detected, out of an expected 189 head neuron nuclei (Fig 8E). We then tested toehold-mediated reversal for one amplifier (Fig 9A). Fluorescent intensity in HCR-RNA-FISH puncta was significantly higher in samples before application of the toehold reversal protocol than in images after removal (Fig9B). This experiment showed that punctate signal that was clustered around neuronal nuclei could be disassociated from their point of binding. Together, these experiments served as a proof of principal for the protocols to 1) multiplex and read out these specific probes and fluorophores and 2) reverse binding of amplifiers in general for multi-round staining. Once we showed that we had protocols that enabled these basic functions of the full workflow illustrated in Figure 7 in isolation, we continued on to multi-round multiplexed HCR-RNA-FISH to complete the procedural proof of principal.

Multi-round multiplexed HCR-RNA-FISH experiments piloting the entire experimental protocol were performed semi-successfully. As can be seen in Figure 10, some differential expression can be seen by observing single slices. However, a small number of issues caused less than ideal results in this proof of concept experiment. First, it may be

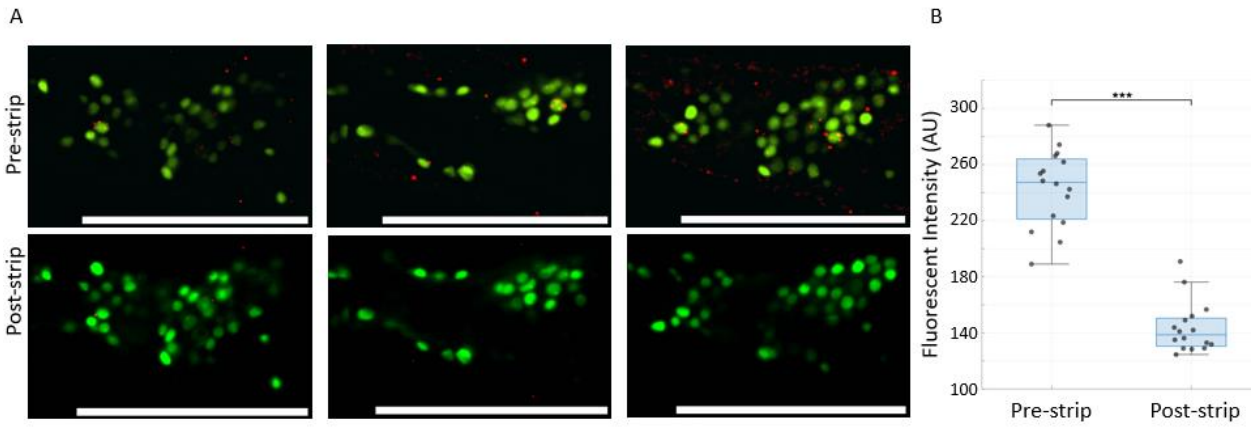

Figure 9: Toehold mediated reversal of HCR-RNA-FISH. A) Representative single slices from the volumetric image of one worm after HCR-RNA-FISH staining (top) and after using toehold reversible hairpins (bottom; AlexaFluor 546 fluorophores; B2 amplifier sequence). All scale bars are 50μm. B) Quantification of fluorescence of HCR-RNA-FISH puncta in manually selected "positive" puncta before and after toehold mediated reversal of staining. N=16 neurons in 2 worms; values normalized by imaging intensity. P<<0.001 in two-tailed paired t-test.

observed that the pan-neuronal nuclear GCaMP signal is significantly reduced as compared to the same signal in worms of the same strain in the single round multiplexed HCR-RNA-FISH experiment (Fig 8) and the toehold mediated reversal experiment (Fig 9). The next clear issue with this experiment is the fact that the HCR-RNA-FISH failed in the 546nm channel in the third round of barcode readout, resulting in null expression information for inx-19. Finally, although not visible in this figure, observation of the data indicates that stripping of HCR-RNA-FISH signal may be incomplete between rounds.

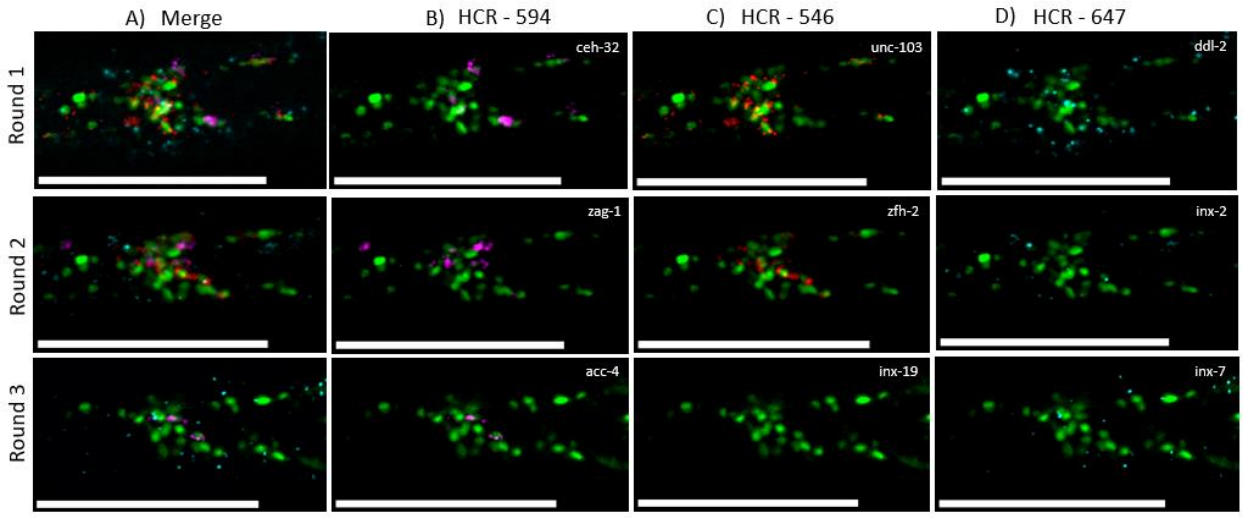

Figure 10: EnBarc experimental proof of concept via multi-round multiplexed HCR-RNA-FISH. Representative single slice from volumetric images taken from multi-round three-gene multiplexed HCR-RNA-FISH using the EnBarc neural ID genes. A) One multiplexed "round" of three genes is shown in a single worm, revealing the entirety of the barcode genes. Pan-neuronal nuclear GCaMP signal (green) is shown together with HCR-RNA-FISH puncta with AlexaFluor 594 (B), AlexaFluor 546 (C), and AlexaFluor 647 (D) fluorophores in each round of imaging in a single worm. Note that nuclear GCaMP signal is reduced in this experiment as compared to Figures 8 and 9, and signal was not observed for inx-19 in the 546nm channel during the third round of imaging. All scale bars are 50μm.

4.3.2 Semi-Automated Barcode Readout and Probabilistic Cell ID

After volumetric images are taken for each round of HCR-RNA-FISH to record EnBarc gene set expression, these volumes are evaluated by a semi-automated analysis code that pre-processes image data, reads presence or absence of each gene and outputs a table of probabilities representing the likelihood of a neuron belonging to each cell class (Fig 11). Likelihoods for a given segmented neuron N to belong to a given cell class C are calculated by Bayesian inference (Fig 11D). The prior probability that neuron N belongs to class C is calculated by first aligning the segmented neurons to approximate neuron location data from the OpenWorm NeuroML *C. elegans* Connectome model, c302 (Gleeson et al., 2018; accessed via https://github.com/openworm/CElegansNeuroML) via 3D point cloud alignment. 3D Gaussian distributions were centered on cell class canonical locations, and the

distance of the segmented neuron N from a cell class location C determined the prior probability of N belonging to C according to the distribution. This belief was updated by multiplying the prior to the probability of observing gene A given that neuron N did in fact

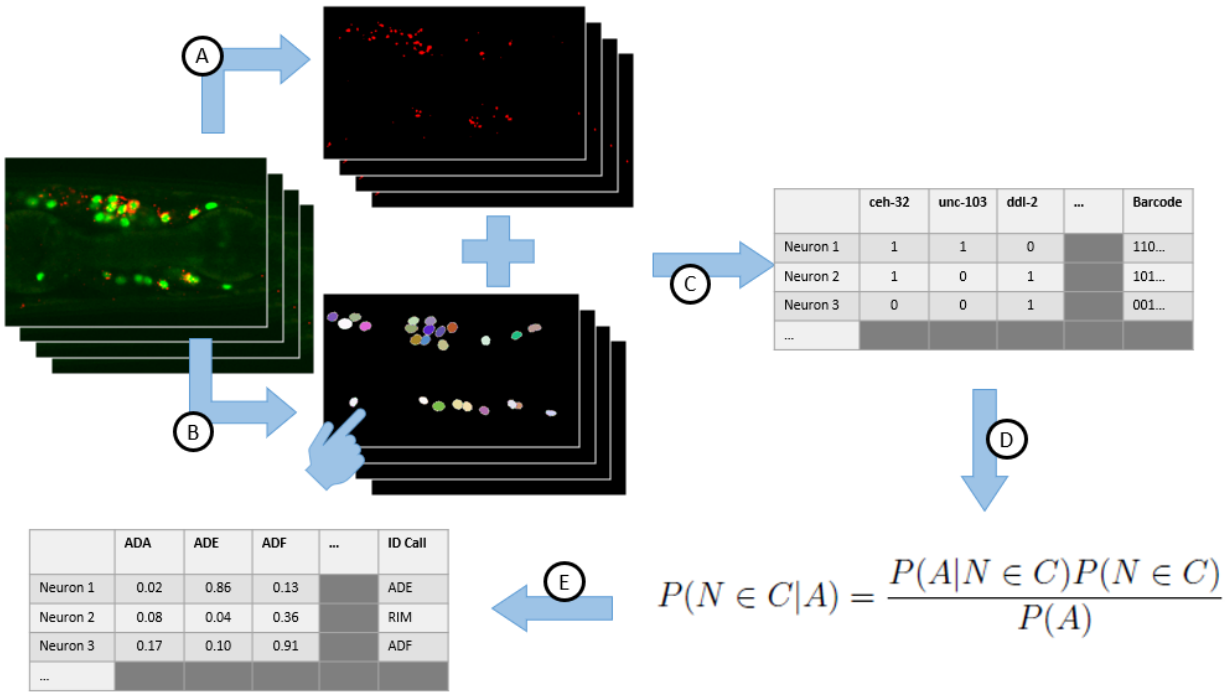

Figure 11: EnBarc semi-automatic process for barcode readout and cell ID. Given a volume of a worm and the associated fluorescent signal corresponding to neuronal nuclei and HCR-RNA-FISH puncta for three genes, (A) A threshold is applied to generate a 3D binary image denoting puncta and non-puncta in the volume. This is used as a mask which can be applied to the HCR-RNA-FISH channel that was used to generate it; a watershed algorithm is then applied to this image to find centroids of the puncta. Centroid location in 3D space is stored with the gene identity. B) The StarDist Python library is used to segment the volumetric nuclei from the GCaMP fluorescent image. Each 3D nucleus is individually labeled and stored with 3D centroid location and information about which pixels are included in its volume. A researcher scrolls through the fluorescent image and segmented labels, manually removing any false nuclei. C) Labelled nucleus and 3D puncta centroid location information are overlaid within imaging rounds. The genes of corresponding puncta centroids that fall within the volume of labeled nuclei are recorded as being observed in that nucleus. A border of specified pixel size may be added to nuclear label perimeters to make the process of assigning puncta more permissive. The three segmented nucleus label files are aligned between rounds using 3D point cloud alignment. Full barcodes reflecting presence of absence of a gene are recorded for all labelled nuclei that can be aligned between all three rounds. D) A Bayesian belief update equation is calculated for each labeled nucleus with barcode information. For each neuron class C, the probability that a labelled nucleus N belongs to that class is updated given the measured observation of gene A. This is calculated by first multiplying the probability of observing gene A given that the nucleus N belongs to class C (from the CeNGEN single cell seq data) by the prior probability that the nucleus N belongs to class C (based on proximity of N to C when aligned to an atlas of cell class locations). This value is divided by the overall probability of observing gene A, a noise term which can be calculated or measured. E) These calculations result in a probability between 0 and 1 that each nucleus belongs to each class. The final ID call for each nucleus is given to the class with highest probability.

belong to class C and dividing by the independent probability of observing gene A (a noise term that could be calculated or measured). Final cell ID calls are given for each nucleus detected and aligned between all three volumes; the call for each nucleus is determined by the cell class with the highest calculated probability of a given segmented neuron being a member.

There are several ways we may use this analysis pipeline to evaluate the effectiveness of the EnBarc methodology above. The first is by checking the number of neurons marked as "positive" (assigned "1" in barcode readout) in the data against the number we expect to see from the CeNGEN dataset. Table 2 contains these results for 8 worms from the imaging data of the single-round multiplexed RNA-FISH experiments. In these data, 150.3 ± 54.7 nuclei out of the 189 expected head nuclei were detected per image. The number of expected positive nuclei for each gene based on CeNGEN expression data was adjusted for each image proportionally to the number of detected to expected nuclei. For these results, fluorescent intensity thresholds for each gene were determined by minimizing the absolute sum of adjusted expected – detected positive nuclei (the rightmost column of Table 2). This results in very low measured average differences in expected and observed counts of positive nuclei in summary statistics, but inconsistent stain quality between images may result in thresholds at once too low for some images and too high for others. An accurate number of segmented nuclei marked positive also does not necessarily mean that the correct nuclei are marked positive, and a study of final ID calls in complete multi-round experiments will be required to fully determine the best method of thresholding.

Table 2: Expected and Detected Positive Nuclei per EnBarc Gene.

Gene	Fluorescent Intensity Threshold	Expected Positive Nuclei (CeNGEN)	Average Expected Positive Nuclei (Adjusted for Detected Nuclei)	Average Detected Positive Nuclei	Average Expected (Adjusted) - Detected Positive Nuclei
unc-103	602	57	39.7	40.3	0.7
inx-19	139	45	31.3	34.0	2.7
zfh-2	433	96	66.7	66.7	0.0
ddl-2	900	0	0.0	3.7	3.7
inx-2	133	121	103.3	112.0	8.7
acc-4	145	87	74.3	73.7	-0.7
inx-7	214	101	86.5	87.0	0.5
zag-1	138	90	77.0	69.5	-7.5
ceh-32	399	36	30.5	31.0	0.5

We can also use this pipeline to evaluate the robustness of our 9-gene EnBarc system for neural ID in *C. elegans*. We simulated the results of how prior probabilities and final cell ID calls change with the addition of information from RNA expression data over 1000 trials. Ground truth data for neural nucleus location was generated by adding random noise to canonical nucleus locations within controlled ranges chosen to reflect the amount of maximum displacement that can be reasonably expected, as reported in previous studies of neural position variation (Yemini *et al.* 2021, Toyoshima *et al.,* 2020). As has been noted in these other studies, we found that cell ID calls cannot reliably be made from this information alone, although recently improved atlases using identification strains may make improvements to this ability in the future (Yemini *et al.,* 2021; Toyoshima *et al.,* 2020; Skuhersky *et al.,* 2021). When nucleus locations vary within a maximum of 2.5μm from canonical locations, the accuracy of ID calls based on location fell by over 40% (Fig 12A). This is approximately equal to the smallest axis of the 70% Gaussian ellipsoids generated by

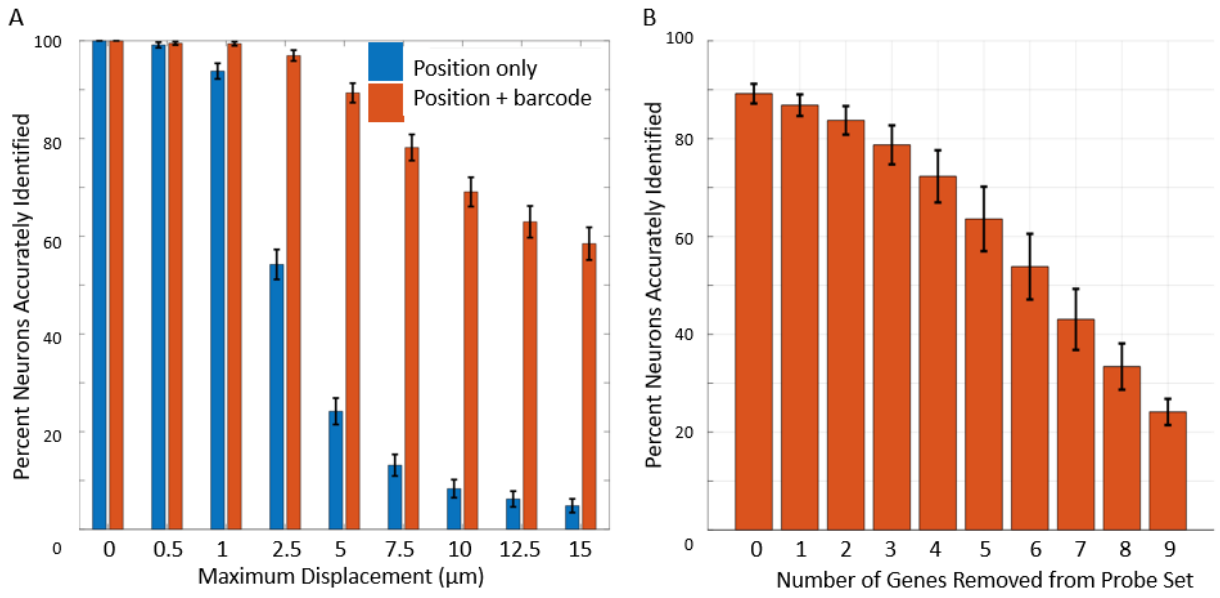

Figure 12: EnBarc cell ID compensates for locational variability and is robust to experimental error in simulated data. Cell ID calls were made in simulated data. Ground truth for nucleus location was generated by adding random noise in a uniform distribution to each neuron's canonical location. A) Percent of neurons accurately identified in simulations where random noise of several values of maximum displacement were added to canonical neuron coordinates. Cell ID calls were made by taking into account positional information only (blue) or both position and barcode information (red). N=1000 simulations in each condition; error bars are SD. B) Percent of neurons accurately identified when gene information is missing. N= 1000 simulations; 5µm of maximum random positional noise; error bars are SD. Left/right and dorsal/ventral identities were taken into account in final accuracy calls, potentially decreasing the overall accuracy.

Toyoshima *et al.,* describing variability in cell position, implying that it is near the minimum displacement we should reasonably expect in real data. At a maximum simulated displacement of 15µm from canonical locations (near maximum reported radii for cells that move dramatically, Toyoshima *et al.,* 2020), location data alone results in 5% of accurate ID calls. When high quality barcode data is accounted for, 54-98% of simulated ID calls are accurate within reasonably expected ranges of positional variation (2.5-15µm maximum displacement, Fig 12A, red bars). We also find that a 9-gene barcode is robust to experimental error. In 1000 simulations of data with locational noise within 5µm, removing one gene from the barcode did not significantly impact accuracy of identified neurons in

comparison to 9-gene barcodes, regardless of the gene whose information was missing from the data (Fig12B). These initial simulations results are encouraging, especially when we consider that locational variation is not uniform as simulated, so only a small number of neurons are likely to have large displacements from canonical locations as opposed to a majority which will be within 5-10μm of where they are expected. Additionally, these simulations were extraordinarily strict in scoring ID calls, penalizing incorrect left/right and dorsal/ventral calls which would easily be disambiguated in practice by quick quality assessment.

4.4 Discussion

4.4.1 Work to be Completed

There are a few experiments that still must be performed in order to complete this work. The most notable are a set of full protocol experiments (multi-round multiplexed HCR-RNA-FISH) in neurotransmitter reporter strains. We plan to perform this experiment in reporter strains for dopamine (marks 8 hermaphrodite neurons of 3 classes, BZ555, https://cgc.umn.edu/strain/BZ555), serotonin (marks 11 hermaphrodite neurons of 6 classes, OH12495, https://cgc.umn.edu/strain/OH12495), and GABA (marks 26 hermaphrodite neurons from 6 classes, CZ13799, https://cgc.umn.edu/strain/CZ13799). We will implement our full three-round protocol in these strains and compare the accuracy of cell ID via hand annotation and via EnBarc. This will serve as a set of positive controls for the method as a whole.

Negative controls are also a notable factor that are missing here and will be necessary to complete for submission of this project to a journal. We performed preliminary HCR-

RNA-FISH experiments in established worm strains that would normally serve as negative controls for some of the genes in our barcode set, but these failed as true negative controls (not shown). This was due to the nature of both the strains and the HCR strategy for RNA-FISH. The strains in question contained large deletions of the genes for which they were chosen as negative controls, but each strain still contained fragments of the deleted gene sequence. The way that HCR is designed is such that a set of FISH probes, not a single probe, is provided by Molecular Instruments (MI), which will hybridize at 12-40 unique sites in the sequence provided for a custom probe set order (Choi *et al.,* 2018). When checked against the deletion strains, the probe sets delivered to us contained between 3 and 15 probes that hybridized to non-deleted regions of the knocked-out genes. We are in contact with MI to remedy this, and will request probe sets that exclude the probes that hybridize to non-deleted gene fragments in the knockout strains; additionally, we will suggest that they have a mechanism to submit details of negative controls planned with their reagents to prevent confusion and experimental delay in the future.

We also plan to complete additional multi-round multiplexed HCR-RNA-FISH experiments. This thesis currently shows only one proof of concept result, admitting a handful of points to improve upon, and we plan to do exactly that. To address the issue of low nuclear GCaMP signal, we will begin with a newly ordered population of worms of the same strain, revisit detailed experiment notes to see if any subtle protocol changes could have contributed to the change in nuclear signal, and make use of a fluorophore-conjugated antibody that will further amplify nuclear signal. Difficulties in fully reversing amplification in early rounds of the experimental readout protocol were also suspected. In anecdotal

experience gleaned from my labmates in the Boyden lab also working on multi-round multiplexed punctum imaging in gel-embedded samples, I have been advised that imaging parameters may play a role. High laser power and long exposure times may cause fluorophores to become anchored to the reactive chemical makeup of the hydrogel. This indicates that using lower laser power and shorter exposure times may ameliorate the issue of cross-round signal. Alternatively, increasing incubation time with dissociation strands or the relative concentration of dissociation strands to amplifier may also solve the problem. Finally, the reason for failure of the inx-19 HCR-RNA-FISH in the third round of the present proof of concept is unknown and inconsistent with previous results; we don't expect this to be a persistent issue, as we suspect it was the result of human error.

Completion of further proof of concept experiments with improved fluorescent signal will build a body of evidence from which we can complete statistical analyses of the accuracy, redundancy, and ease of use of our computational and experimental systems. We will evaluate the number of nuclei segmented and assigned a final ID call, assess whether any classes are more likely to be assigned than others, and whether ID calls can be made with incomplete experimental information to test the results of our present simulations. We will also be able to report the range and average confidence values for final ID calls and make comparisons to existing solutions for *C. elegans* neural cell ID.

4.4.2 Potential Pitfalls

Even upon completion of the above experiments in preparation for publication, there are a number of potential pitfalls researchers seeking to use the EnBarc system in *C. elegans* may encounter. While a system to ID neurons by cell class has many applications, a popular

aim will be to apply this to whole brain imaging experiments. We are optimistic for the application of this technique to such projects, but it remains to be seen how well neural nuclei can be aligned between live activity and fixed datasets. Additionally, the experimental protocol does take 5 days to complete once worms have been fixed. The length of this protocol and the initial time investment to learn to work with hydrogels may deter researchers from adopting the method.

4.4.3 Utility and Further Customization of EnBarc

Nonetheless, we believe that this method has benefits that outweigh the potential pitfalls:

1. As with any method, a period of learning is to be expected, and the time to learn multi-round multiplexed HCR-RNA-FISH for neuron ID via EnBarc is much shorter than the time it would take a researcher to obtain proficiency in hand-annotation.
2. The use of commercially available reagents from MI makes the fluorescent multiplexing customizable to a lab's existing microscopy equipment, including filter sets.
3. This technique could be used in strains researchers designed for their own behavioral, genetic, and/or imaging experiments (as opposed to using integrated cell ID strains such as NeuroPAL).
4. EnBarc gene sets can be newly generated with improved single cell RNA sequencing data or for special cases of cell ID, such as a gene set utilized for exclusively head or tail neurons.

We also plan to package computational code from this work in a virtual machine accessible by web browser, making it effortless to design new EnBarc gene sets and analyze imaging data. This will solve a major problem in introducing new protocols, which is often a long learning curve in setting up, debugging, and learning to use researcher-generated code.

Another major benefit of this method is its adaptability. The gene set selection algorithm could use RNA sequencing data and cell class location data for other organisms entirely and produce an EnBarc strategy for those with minimal changes to the code. This could apply cell ID to imaging projects of any type, as long as the input data is available. The experimental readout of an EnBarc gene set is also customizable. In this thesis we outline a strategy that uses HCR-RNA-FISH because of its quality of staining with low experimental noise in *C. elegans* as compared to other RNA-FISH methods, but any conceivable way of detecting mRNA sequences could be used to read out EnBarc barcodes. As technological advances emerge, the EnBarc workflow is adaptable to incorporate them, potentially in ways that make the technique even more powerful. For example, labs that are able to perform In-Situ RNA Sequencing (Lee *et al.,* 2014) could conceivably use that technique to generate their barcode readout datasets and still be able to feed it into the same analysis code (equipped with appropriate cell class location and gene expression information) to annotate their data by cell class.

4.5 Contributions

I designed the EnBarc theoretical framework with equal contribution from Orhan Tunç Çeliker and conceptual contributions from Jay Yu. The computational gene selection algorithm and analysis pipelines were developed by Dr. Çeliker with contributions by

undergraduate research assistant Hannah Chen. HCR-RNA-FISH screening, single-round, and multi-round multiplexed experiments were designed, performed, and imaged primarily by me, with advice in troubleshooting by Yangning Lu, Chi Zhang, and Konstantinos Kagias. Steps of the experimental protocol, including the gelled cuticle digestion, were contributed by Dr. Zhang and Dr. Lu. Dr. Çeliker and I collaborated on imaging data analysis strategy; all code was written by Dr. Çeliker.

7 References

Brenner S. The genetics of Caenorhabditis elegans. Genetics. 1974;77:71–94. pmid:4366476

Brittin, C.A., Cook, S.J., Hall, D.H. et al. A multi-scale brain map derived from whole-brain volumetric reconstructions. Nature 591, 105–110 (2021). https://doi.org/10.1038/s41586-021-03284-x

Çeliker Orhan Tunç (2021). Automated cellular identity assignment in C. elegans using differential gene expression (Doctoral thesis). Massachusetts Institute of Technology, Cambridge, MA, USA.

Chalasani SH, Chronis N, Tsunozaki M, Gray JM, Ramot D, Goodman MB, Bargmann CI. Dissecting a circuit for olfactory behaviour in Caenorhabditis elegans. Nature. 2007 Nov 1;450(7166):63-70. doi: 10.1038/nature06292. Erratum in: Nature. 2008 Jan 3;451(7174):102. Erratum in: Nature. 2016 May 5;533(7601):130. PMID: 17972877.

Chalfie M, Sulston J E, White J G, Southgate E, Thomson J N, Brenner S. The neural circuit for touch sensitivity in C. elegans. J. Neurosci. 1985

Chen T-W, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature. 2013;499:295–300. pmid:23868258

Choi HMT, Schwarzkopf M, Fornace ME, Acharya A, Artavanis G, Stegmaier J, Cunha A, Pierce NA. Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development, 2018.

Chronis N, Zimmer M, Bargmann CI. Microfluidics for in vivo imaging of neuronal and behavioral activity in Caenorhabditis elegans. Nat Methods. 2007;4:727–731. pmid:17704783

Clark DA, Biron D, Sengupta P, Samuel AD. The AFD sensory neurons encode multiple functions underlying thermotactic behavior in Caenorhabditis elegans. J Neurosci. 2006 Jul 12;26(28):7444-51. doi: 10.1523/JNEUROSCI.1137-06.2006. PMID: 16837592; PMCID: PMC6674189.

Dana H, Mohar B, Sun Y, Narayan S, Gordus A, Hasseman JP, et al. Sensitive red protein calcium indicators for imaging neural activity. Elife. 2016;5:e12727. pmid:27011354

Dana H, Sun Y, Mohar B, Hulse BK, Kerlin AM, Hasseman JP, et al. High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. Nat Methods. 2019;16:649–657. pmid:31209382

De Magalhaes Filho CD, Henriquez B, Seah NE, Evans RM, Lapierre LR, Dillin A. Visible light reduces C. elegans longevity. Nat Commun. 2018 Mar 2;9(1):927. doi: 10.1038/s41467-018-02934-5. PMID: 29500338; PMCID: PMC5834526.

Donato A, Kagias K, Zhang Y, Hilliard MA. Neuronal sub-compartmentalization: a strategy to optimize neuronal function. Biol Rev Camb Philos Soc. 2019 Jun;94(3):1023-1037. doi: 10.1111/brv.12487. Epub 2019 Jan 4. PMID: 30609235; PMCID: PMC6617802.

Edwards SL, Charlie NK, Milfort MC, Brown BS, Gravlin CN, Knecht JE, Miller KG. A novel molecular solution for ultraviolet light detection in Caenorhabditis elegans. PLoS Biol. 2008 Aug 5;6(8):e198. doi: 10.1371/journal.pbio.0060198. PMID: 18687026; PMCID: PMC2494560.

Emmons SW, Yemini E, Zimmer M. Methods for analyzing neuronal structure and activity in Caenorhabditis elegans. Genetics. 2021 Aug 9;218(4):iyab072. doi: 10.1093/genetics/iyab072. PMID: 34151952; PMCID: PMC8864745.

Flavell SW, Gordus A. Dynamic functional connectivity in the static connectome of Caenorhabditis elegans. Current Opinion in Neurobiology, Volume 73, 2022, 102515, ISSN 0959-4388, https://doi.org/10.1016/j.conb.2021.12.002.

Gleeson Padraig, Lung David, Grosu Radu, Hasani Ramin and Larson Stephen D. 2018c302: a multiscale framework for modelling the nervous system of Caenorhabditis elegansPhil. Trans. R. Soc. B3732017037920170379. http://doi.org/10.1098/rstb.2017.0379 (repository with relevant data can be accessed at https://github.com/openworm/CElegansNeuroML)

Gonzales DL, Zhou J, Fan B, Robinson JT. A microfluidic-induced C. elegans sleep state. Nat Commun. 2019 Nov 6;10(1):5035. doi: 10.1038/s41467-019-13008-5. PMID: 31695031; PMCID: PMC6834590.

Hallinen KM, Dempsey R, Scholz M, Yu X, Linder A, Randi F, Sharma AK, Shaevitz JW, Leifer AM. Decoding locomotion from population neural activity in moving C. elegans. Elife. 2021 Jul 29;10:e66135. doi: 10.7554/eLife.66135. PMID: 34323218; PMCID: PMC8439659.

Hendricks, M., Ha, H., Maffey, N. et al. Compartmentalized calcium dynamics in a C. elegans interneuron encode head movement. Nature 487, 99–103 (2012). https://doi.org/10.1038/nature11081

Inglis PN, Ou G, Leroux MR, et al. The sensory cilia of Caenorhabditis elegans. In: WormBook: The Online Review of C. elegans Biology [Internet]. Pasadena (CA): WormBook; 2005-2018. Available from: https://www.ncbi.nlm.nih.gov/books/NBK19729/

Jiang J, Su Y, Zhang R, Li H, Tao L, Liu Q. C. elegans enteric motor neurons fire synchronized action potentials underlying the defecation motor program. Nat Commun 13, 2783 (2022). https://doi.org/10.1038/s41467-022-30452-y

King NP, Sheffler W, Sawaya MR, Vollmar BS, Sumida JP, André I, Gonen T, Yeates TO, Baker D. Computational design of self-assembling protein nanomaterials with atomic level accuracy. Science. 2012 Jun 1;336(6085):1171-4. doi: 10.1126/science.1219364. PMID: 22654060; PMCID: PMC4138882.

Klapoetke NC, Murata Y, Kim SS, Pulver SR, Birdsey-Benson A, Cho YK, Morimoto TK, Chuong AS, Carpenter EJ, Tian Z, Wang J, Xie Y, Yan Z, Zhang Y, Chow BY, Surek B, Melkonian M, Jayaraman V, Constantine-Paton M, Wong GK, Boyden ES. Independent optical excitation of distinct neural populations. Nat Methods. 2014 Mar;11(3):338-46. doi: 10.1038/nmeth.2836. Epub 2014 Feb 9. Erratum in: Nat Methods. 2014 Sep;11(9):971. PMID: 24509633; PMCID: PMC3943671.

Kato, S., Kaplan, HS., Schrödel, T., Skora, S., Lindsay, TH., Yemini, E., Lockery, S., Zimmer, M. Global Brain Dynamics Embed the Motor Command Sequence of Caenorhabditis elegans. Cell, 2015.

Lagache T, Hanson A, Pérez-Ortega JE, Fairhall A, Yuste R. Tracking calcium dynamics from individual neurons in behaving animals. PLoS Comput Biol. 2021 Oct 8;17(10):e1009432. doi: 10.1371/journal.pcbi.1009432. PMID: 34624016; PMCID: PMC8528277.

Lai, YT., Reading, E., Hura, G. et al. Structure of a designed protein cage that self-assembles into a highly porous cube. Nature Chem 6, 1065–1071 (2014). https://doi.org/10.1038/nchem.2107

Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Yang JL, Ferrante TC, Terry R, Jeanty SS, Li C, Amamoto R, Peters DT, Turczyk BM, Marblestone AH, Inverso SA, Bernard A, Mali P, Rios X, Aach J, Church GM. Highly multiplexed subcellular RNA sequencing in situ. Science. 2014 Mar 21;343(6177):1360-3. doi: 10.1126/science.1250212. Epub 2014 Feb 27. PMID: 24578530; PMCID: PMC4140943.

Lee KH, Aschner M. A Simple Light Stimulation of Caenorhabditis elegans. Curr Protoc Toxicol. 2016;67:11.21.1-11.21.5. Published 2016 Feb 1. doi:10.1002/0471140856.tx1121s67

Lin, A., Witvliet, D., Hernandez-Nunez, L. et al. Imaging whole-brain activity to understand behaviour. Nat Rev Phys 4, 292–305 (2022). https://doi.org/10.1038/s42254-022-00430-w

Linder, Ashley Nicole. Whole-Brain Dynamics Underlying Spontaneous Behavior in C. elegans. Unpublished doctoral dissertation, 2017.

Liu M, Sharma AK, Shaevitz JW, Leifer AM. Temporal processing and context dependency in Caenorhabditis elegans response to mechanosensation. Elife. 2018 Jun 26;7:e36419. doi: 10.7554/eLife.36419. PMID: 29943731; PMCID: PMC6054533.

Liu Q, Kidd PB, Dobosiewicz M, Bargmann CI. C. elegans AWA Olfactory Neurons Fire Calcium-Mediated All-or-None Action Potentials. Cell. 2018 Sep 20;175(1):57-70.e17. doi: 10.1016/j.cell.2018.08.018. Epub 2018 Sep 13. PMID: 30220455.

Moyle, M.W., Barnes, K.M., Kuchroo, M. et al. Structural and developmental principles of neuropil assembly in C. elegans. Nature 591, 99–104 (2021). https://doi.org/10.1038/s41586-020-03169-5

Nakai, J., Ohkura, M. & Imoto, K. A high signal-to-noise Ca2+ probe composed of a single green fluorescent protein. Nat Biotechnol 19, 137–141 (2001). https://doi.org/10.1038/84397

Nguyen JP, Shipley FB, Linder AN, Plummer GS, Liu M, Setru SU, Shaevitz JW, Leifer AM. Whole-brain calcium imaging with cellular resolution in freely behaving Caenorhabditis elegans. Proc Natl Acad Sci U S A. 2016 Feb 23;113(8):E1074-81. doi: 10.1073/pnas.1507110112. Epub 2015 Dec 28. PMID: 26712014; PMCID: PMC4776509.

Nguyen JP, Linder AN, Plummer GS, Shaevitz JW, Leifer AM. Automatically tracking neurons in a moving and deforming brain. PLoS Comput Biol. 2017 May 18;13(5):e1005517. doi: 10.1371/journal.pcbi.1005517. PMID: 28545068; PMCID: PMC5436637.

Qian Y, Piatkevich KD, Mc Larney B, Abdelfattah AS, Mehta S, Murdock MH, et al. A genetically encoded near-infrared fluorescent calcium ion indicator. Nat Methods. 2019;16:171–174. pmid:30664778

Qian Y, Cosio DMO, Piatkevich KD, Aufmkolk S, Su W-C, Celiker OT, et al. (2020) Improved genetically encoded near-infrared fluorescent calcium ion indicators for in vivo imaging. PLoS Biol 18(11): e3000965. https://doi.org/10.1371/journal.pbio.3000965

Ruach R, Ratner N, Emmons SW, Zaslaver A. The synaptic organization in the C. elegans neural network suggests significant local compartmentalized computations. bioRxiv 2021.12.30.474568; doi: https://doi.org/10.1101/2021.12.30.474568

Sando SR, Bhatla N, Lee EL, Horvitz HR. An hourglass circuit motif transforms a motor program via subcellularly localized muscle calcium signaling and contraction. Elife. 2021 Jul 2;10:e59341. doi: 10.7554/eLife.59341. PMID: 34212858; PMCID: PMC8331187.

Schild LC, Glauser DA. Dual Color Neural Activation and Behavior Control with Chrimson and CoChR in Caenorhabditis elegans. Genetics. 2015 Aug;200(4):1029-34. doi: 10.1534/genetics.115.177956. Epub 2015 May 28. PMID: 26022242; PMCID: PMC4574232.

Sengupta P, Chou JH, Bargmann CI. odr-10 encodes a seven transmembrane domain olfactory receptor required for responses to the odorant diacetyl. Cell. 1996 Mar 22;84(6):899-909. doi: 10.1016/s0092-8674(00)81068-5. PMID: 8601313.

Seth R. Taylor, Gabriel Santpere, Alexis Weinreb, Alec Barrett, Molly B. Reilly, Chuan Xu, Erdem Varol, Panos Oikonomou, Lori Glenwinkel, Rebecca McWhirter, Abigail Poff, Manasa Basavaraju, Ibnul Rafi, Eviatar Yemini, Steven J. Cook, Alexander Abrams, Berta Vidal, Cyril Cros, Saeed Tavazoie, Nenad Sestan, Marc Hammarlund, Oliver Hobert, David M. Miller, Molecular topography of an entire nervous system, Cell, Volume 184, Issue 16, 2021, Pages 4329-4347.e23, ISSN 0092-8674, https://doi.org/10.1016/j.cell.2021.06.023.

Skuhersky M, Wu T, Yemini E, Boyden E, Tegmark M. Toward a More Accurate 3D Atlas of C. elegans Neurons. bioRxiv 2021.06.09.447813; doi: https://doi.org/10.1101/2021.06.09.447813

Subach OM, Barykina NV, Anokhin KV, Piatkevich KD, Subach FV. Near-Infrared Genetically Encoded Positive Calcium Indicator Based on GAF-FP Bacterial Phytochrome. Int J Mol Sci. 2019;20. pmid:31315229

Taylor SR, Santpere G, Weinreb A, Barrett A, Reilly MB, Xu C, Varol E, Oikonomou P, Glenwinkel L, McWhirter R, Poff A, Basavaraju M, Rafi I, Yemini E, Cook SJ, Abrams A, Vidal B, Cros C, Tavazoie S, Sestan N, Hammarlund M, Hobert O, Miller DM 3rd. Molecular topography of an entire nervous system. Cell. 2021 Aug 5;184(16):4329-4347.e23. doi: 10.1016/j.cell.2021.06.023. Epub 2021 Jul 7. PMID: 34237253; PMCID: PMC8710130.

Toyoshima, Y., Wu, S., Kanamori, M. et al. Neuron ID dataset facilitates neuronal annotation for whole-brain activity imaging of C. elegans. BMC Biol 18, 30 (2020). https://doi.org/10.1186/s12915-020-0745-2

Venkatachalam, V., Ji, N., Wang, X., Clark, C., Mitchell, J.K., Klein, M., Tabone, C.J., Florman, J., Ji, H., Greenwood, J., et al. Pan-neuronal imaging in roaming Caenorhabditis elegans. Proc Natl Acad Sci, 2016

Ward A, Liu J, Feng Z, Xu XZ. Light-sensitive neurons and channels mediate phototaxis in C. elegans. Nat Neurosci. 2008;11(8):916-922. doi:10.1038/nn.2155

White, J.G., Southgate, E., Thomson, J.N., and Brenner, S. The structure of the nervous system of the nematode Caenorhabditis elegans. Philosophical Transactions of the Royal Society of London Biological Sciences, 1986.

Yemini E, Lin A, Nejatbakhsh A, Varol E, Sun R, Mena GE, Samuel ADT, Paninski L, Venkatachalam V, Hobert O. NeuroPAL: A Multicolor Atlas for Whole-Brain Neuronal Identification in C. elegans. Cell. 2021 Jan 7;184(1):272-288.e11. doi: 10.1016/j.cell.2020.12.012. Epub 2020 Dec 29. PMID: 33378642.

Yu CJ, Barry NC, Wassie AT, Sinha A, Bhattacharya A, Asano S, Zhang C, Chen F, Hobert O, Goodman MB, Haspel G, Boyden ES. Expansion microscopy of C. elegans. Elife. 2020 May 1;9:e46249. doi: 10.7554/eLife.46249. PMID: 32356725; PMCID: PMC7195193.

Yu CJ, Orozco Cosio DM, Boyden ES. ExCel: Super-Resolution Imaging of C. elegans with Expansion Microscopy. Methods Mol Biol. 2022; 2468:141-203. doi: 10.1007/978-1-0716-2181-3_9. PMID: 35320565.

Yu X, Creamer MS, Randi F, Sharma AK, Linderman SW, Leifer AM. Fast deep neural correspondence for tracking and identifying neurons in C. elegans using semi-synthetic training. Elife. 2021 Jul 14;10:e66410. doi: 10.7554/eLife.66410. PMID: 34259623; PMCID: PMC8367385.

Zhang Y, Rózsa M, Liang Y, Bushey D, Wei Z, Zheng J, Reep D, Broussard GJ, Tsang A, Tsegaye G, Narayan S, Obara CJ, Lim JX, Patel R, Zhang R, Ahrens MB, Turner GC, Wang SSH, Korff WL, Schreiter ER, Svoboda K, Hasseman JP, Kolb I, Looger LL. Fast and sensitive GCaMP calcium indicators for imaging neural populations. bioRxiv 2021.11.08.467793; doi: https://doi.org/10.1101/2021.11.08.467793