Unifying community-wide whole-brain imaging datasets enables robust automated neuron identification and reveals determinants of neuron positioning in *C. elegans*

Daniel Y. Sprague[1], Kevin Rusch[2], Raymond L. Dunn[1], Jackson M. Borchardt[1], Steven Ban[1], Greg Bubnis[1], Grace C. Chiu[1], Chentao Wen[3], Ryoga Suzuki[4], Shivesh Chaudhary[5], Hyun Jee Lee[5], Zikai Yu[5], Benjamin Dichter[6], Ryan Ly[7], Shuichi Onami[3], Hang Lu[5], Koutarou D. Kimura[4], Eviatar Yemini[2*], Saul Kato[1*]

[1]Department of Neurology, University of California San Francisco
[2]Department of Neurobiology, UMass Chan Medical School
[3]RIKEN Center for Biosystems Dynamics Research
[4]Graduate School of Science, Nagoya City University
[5]School of Chemical and Biomolecular Engineering, Georgia Institute of Technology
[6]CatalystNeuro, LLC
[7]Scientific Data Division, Lawrence Berkeley National Laboratory
*Correspondence: saul.kato@ucsf.edu, eviatar.yemini@umassmed.edu

Abstract

We develop a data harmonization approach for *C. elegans* volumetric microscopy data, still or video, consisting of a standardized format, data pre-processing techniques, and a set of human-in-the-loop machine learning based analysis software tools. We unify a diverse collection of 118 whole-brain neural activity imaging datasets from 5 labs, storing these and accompanying tools in an online repository called WormID (wormid.org). We use this repository to train three existing automated cell identification algorithms to, for the first time, enable accuracy in neural identification that generalizes across labs, approaching human performance in some cases. We mine this repository to identify factors that influence the developmental positioning of neurons. To facilitate communal use of this repository, we created open-source software, code, web-based tools, and tutorials to explore and curate datasets for contribution to the scientific community. This repository provides a growing resource for experimentalists, theorists, and toolmakers to (a) study neuroanatomical organization and neural activity across diverse experimental paradigms, (b) develop and benchmark algorithms for automated neuron detection, segmentation, cell identification, tracking, and activity extraction, and (c) inform models of neurobiological development and function.

Introduction

Whole-brain imaging experiments with single-neuron resolution (herein shortened to simply "whole-brain imaging") have undergone explosive growth since first demonstrated in the nematode *C. elegans*, a millimeter-sized worm, and the zebrafish *D. rerio* in 2013.[1,2] Since then, these methods have been widely adopted and advanced in the worm[3,4,5,6,7], zebrafish[8,9,10,11,12,13], and larval[14] and adult[15,16] fly communities. Moreover, there have been significant efforts and advances in neuron-resolution imaging of multiple and/or large brain regions in mammals, rapidly approaching whole-brain imaging, especially in mice.[17,18,19,20,21]

In *C. elegans*, whole-brain imaging datasets have enabled characterization of neural dynamics[3,6], functional connectivity[26,22,23,24], and the roles of individual neurons during behavior[7]. These studies leverage the property of eutely in this organism: each cell has a unique and stereotyped identity, consistent across every animal, that allows for data from individual neurons to be pooled and compared across multiple trials and animals. However, analyses of these experiments are bottlenecked by the need to determine the unique identities of each neuron in 3D volumetric recordings. Manual cell identification from fluorescent microscopy imagery is a notoriously difficult skill, requiring substantial expertise and labor. This task is particularly difficult for neurons labeled with nuclear localized fluorophores, which is typical for whole-brain recordings. We recently developed NeuroPAL[6], the first method where the unique identity of every single neuron can be distinguished by an invariant fluorescent color barcode in living animals at all developmental stages of both sexes.[25] NeuroPAL has greatly simplified the task of cell identification and has thus seen rapid adoption, with at least 6 labs[6,7,26,22,24] publishing whole-brain imaging datasets using these animals, and many more labs incorporating the system into their experimental protocols since its release in 2021.

Despite this innovation, neural identification remains a challenging task that requires expertise and many hours of manual work. In the past few years, researchers have proposed various algorithmic auto-identification approaches to attack this problem.[28,29,26,30,31,32] However, none of them have achieved widespread adoption, due at least in part to their incompatibility with different microscopy data formats and low performance on data acquired from different labs. Automatic approaches to the complementary problem of tracking neurons across video frames have achieved some generalized performance across various datasets[33,34], but so far there have not been efforts to perform similar training and benchmarking for automatic cell identification. In order to build automatic approaches that are robust, accurate, and generalizable, there is a critical need for a standardized format and compatible tools trained and benchmarked on a consolidated corpus of data that reflects the heterogeneity of microscopy equipment, experimental conditions, and protocols across labs.

To address this need, we take a data harmonization approach: a process of combining datasets from different sources and homogenizing them to produce a substantially larger data corpus that, in our case, minimizes non-biological inconsistencies across individual datasets while increasing the overall biological diversity of training and benchmarking data. Harmonization includes: i) aggregating the data, ii) converting it to a standardized format, iii) normalizing it, iv) handling duplicate and missing data, and v) pre-processing data to register it to a common space and coordinate system. Data harmonization is standard in many data science fields but has seen slower adoption in the life sciences[35]. Similar efforts to standardize data formats and build large corpuses of data have been essential in the development and benchmarking of many modern machine learning algorithms.[36,37,38]

We introduce WormID (wormid.org). This resource consists of: i) data harmonization tools including a standardized file format for both raw and processed data alongside related metadata that extends the existing Neurodata Without Borders (NWB) format, ii) pre-processing to align the color and

coordinate space of new datasets, and iii) open-source software to analyze whole-brain activity images. We also provide tutorials and documentation that enable researchers to easily incorporate these tools into their data pipelines. Finally, we provide a large online corpus of harmonized *C. elegans* whole-brain activity imaging and structural data that can be used for large-scale experimental analysis, neurobiological modeling, and algorithmic development. This corpus is stored in a popular community archive called the Distributed Archives for Neurophysiology Data Integration (DANDI), which serves as a repository for experimental neuroscience data from a variety of model organisms.[39]

By aggregating a diversity of datasets from multiple labs into a large data corpus, we achieve a substantial boost in the performance of three existing neural auto-identification algorithms, arguably moving into the regime of practical utility for the broader community of users. Furthermore, we mine this corpus to investigate the relationship between neural lineage, synaptic connectivity, and somatic positioning of *C. elegans* neurons to better understand the factors that drive the positioning of neurons in the adult worm.

This corpus and set of tools should be of wide utility to *C. elegans* researchers. We hope it will serve as a seed for continued community aggregation of brain imaging datasets and further the development and improvement of community data-analysis tools applicable across many model organisms.

Results

A standardized format for whole-brain *C. elegans* recordings enables data aggregation and algorithm interoperability

Current state-of-the-art whole-brain recordings of *C. elegans* typically consist of a combination of structural images that often use the NeuroPAL multi-channel fluorescent system to determine neuron identities (**Fig. 1a-b**) and time series images of neural activity acquired by using genetically-encoded activity sensors (e.g., GCaMP6s[40]) (**1c**). This imaging is performed either on immobilized worms (often constrained within a microfluidic chip to maximize image quality[1,3,41]) or on freely-moving worms.[4,5] To aid interpretation, herein we visualize whole-brain structural NeuroPAL images via i) an unrolled 'butterfly' plot of neuron positions that projects the 3D worm structure into a 2D plane (**Fig. 1a**), ii) a 2D projection plot of the NeuroPAL color space (**Fig. 1b**), and iii) 2D dorsal-ventral and lateral projection plots of the neurons (**Fig. 1b**). These visualizations facilitate quick comparisons of neuron color and position from different samples and fine-tuning of their global alignment.

All associated raw data and metadata is stored in the standardized NWB[42] file format with an additional extension that we developed, *ndx-multichannel-volume* (ndx = neurodata extension), to provide support for multi-channel volumetric recordings and *C. elegans* specific metadata (**Fig. 2a**). This extension is available in the NWB Extensions Catalog and is now the official NWB standard for data sharing of *C. elegans* whole-brain neural-activity imaging. NWB data is hierarchically organized with basic metadata stored at the file's root level, raw data stored in the 'acquisition module', and various processed experimental data stored in 'processing modules' (**Fig. 2b**). Individual NWB files contain a single experimental run for a single animal. These NWB files are then stored and accessed from the DANDI archive, where they receive a unique persistent digital object identifier (DOI) in accordance with the International Organization for Standardization (ISO).

We incorporated NWB *ndx-multichannel-volume* read and write functionality into two software tools. These independent software implementations both offer both user-friendly GUIs that analyze *C. elegans* NeuroPAL structural images and neural activity in immobilized worms (NeuroPAL software

https://github.com/Yemini-Lab/NeuroPAL_ID, **Fig. 2c** and eats-worm software https://github.com/focolab/eats-worm**, Fig. 2d**). This functionality can be straightforwardly incorporated into other data-analysis pipelines and software.

In **Table 1** we present a summary of the data we aggregated and harmonized into a corpus: 108 worms from six datasets acquired by five different labs, each with segmented neurons and human-labeled identities. This corpus can be mined for biological insights, training and benchmarking of machine-vision approaches, and neurobiological studies of structural and neural-activity time-series data. Each of these datasets is stored on DANDI and range from a few hundred megabytes to several terabytes (see **Methods** for dataset references). DANDI supports streaming from the cloud and allows users to selectively load data objects and data chunks, substantially reducing the local data storage and RAM requirements necessary to work with this data on a personal computer.

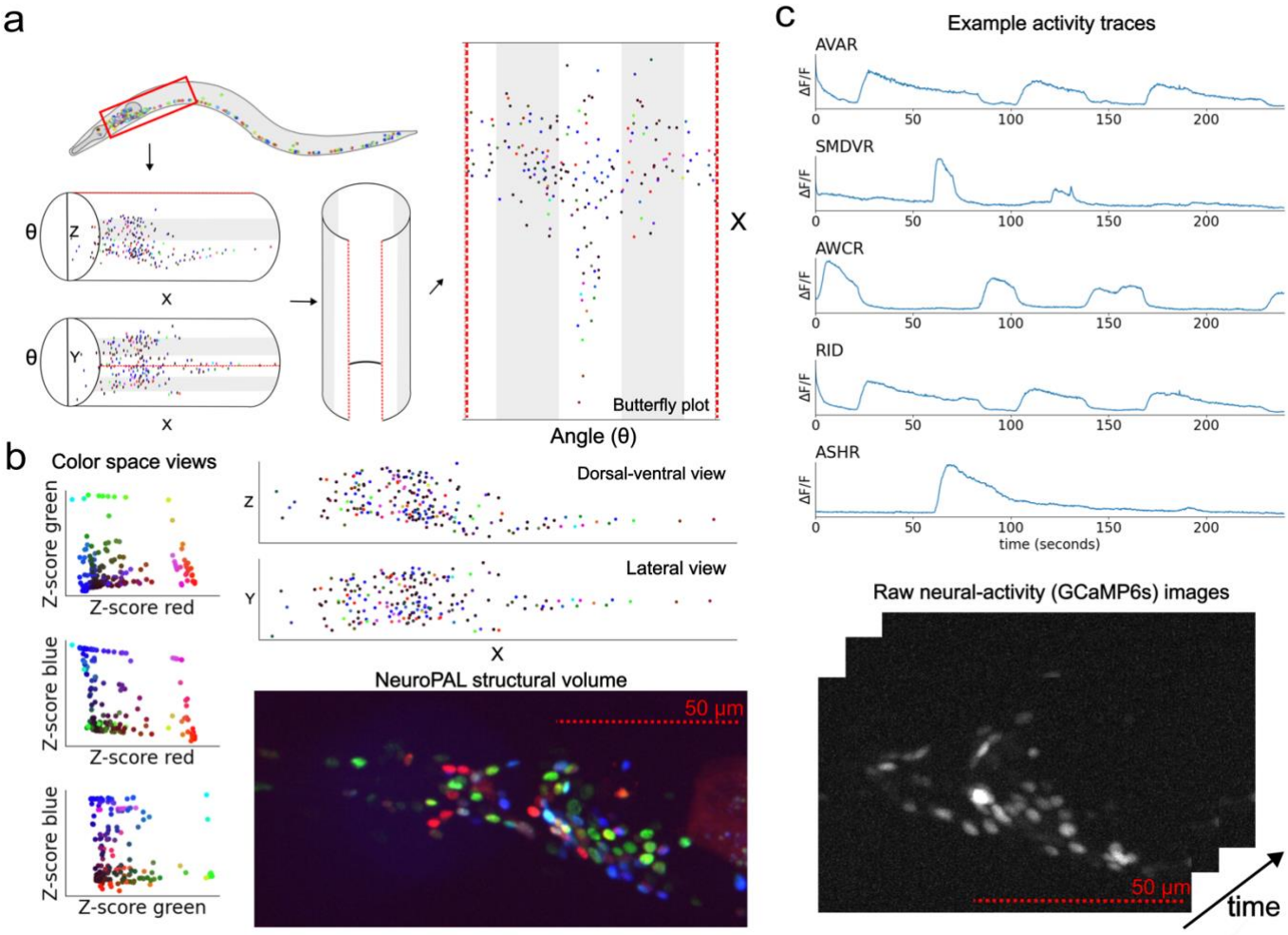


Figure 1: NWB file contents
(a) Illustration of a NeuroPAL worm. The head is highlighted by a red box. A butterfly plot visualizes the full 2D representation of the worm brain by projecting neurons onto the surface of the cylindrical body and then unrolling the cylinder. Neuron centers are colored using their composite NeuroPAL expression. (b) Visualizations of the raw NeuroPAL structural image and 2D projections of its RG, RB, GB color subspaces and XZ and XY projections of its neuron positions. Neuron centers are colored using their composite NeuroPAL expression. (c) Example activity traces for five neurons contained in the NWB file and of the raw neural-activity (GCaMP6s) images.

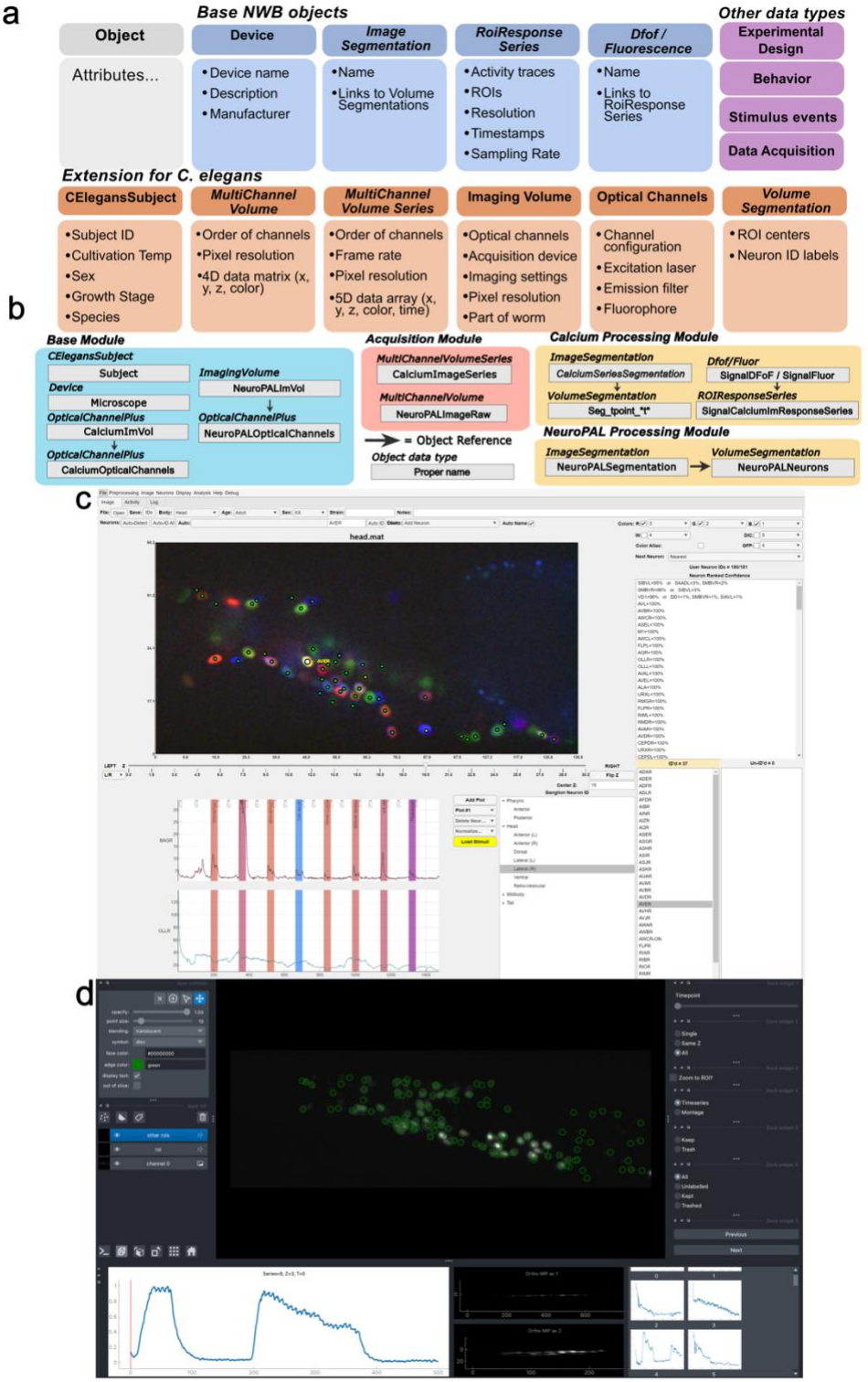


Figure 2: NWB schema and two software programs with NWB I/O support
(a) Names and content for objects used in *C. elegans* optophysiology NWB files. (b) File organization hierarchy of NWB files for *C. elegans* optophysiology. Modules are structured like folders within the root file in an HDF5-based hierarchy. (c-d) NeuroPAL ID software (c) and eats-worm software (d) GUIs with NWB I/O support for visualization and annotation of NeuroPAL structural images, neural segmentation and automated ID, and time-series of neural-activity with stimulus-presentation in immobilized worms.

Dataset #	# of Worms in the Dataset	Lab Code	NeuroPAL, GCaMP, or Both	# of Segmented Neurons (Avg)	# of ID Labels (Avg)
NP	10	NP	NeuroPAL	189-196 (193)	186-193 (190)
1	21	EY	Both	166-188 (177)	164-184 (175)
2	9	HL	NeuroPAL	113-125 (119)	58-69 (64)
3	9	KK	Both	149-163 (154)	149-163 (154)
4	38	SF	Both	29-96 (70)	29-96 (70)
5	21	SK1	Both	78-139 (111)	30-82 (48)
6	10	SK2	NeuroPAL	166-180 (173)	38-63 (49)
Summary	118	5 labs		29-196 (126)	29-193 (99)

Table 1: Summary of aggregated dataset characteristics
NP dataset comes from the original NeuroPAL paper[6], obtained as part of a collaboration between the labs of Oliver Hobert (Columbia University) and Aravinthan D.T. Samuel (Harvard University).

An updated atlas of the *C. elegans* hermaphrodite head

Our multi-lab data corpus allows data scientists to train and benchmark the performance of algorithms for automated neuron identification using datasets that reflect real-world diversity. In this section, we focus on the statistical atlas approach presented in Varol et al. 2020.[28] This approach was the first to take advantage of the color information provided by NeuroPAL and was presented alongside the original NeuroPAL work.[6] This neuron identification assignment algorithm was framed as a bipartite graph matching problem, with the goal of minimizing the total assignment cost using the well-known Hungarian algorithm.[43] Cost is calculated by comparing neuron position and color in the animal sample with the mean and covariance of neuron position and colors in a reference statistical atlas (see **Methods**). The original atlas presented in the paper was trained on 10 worms from the original NeuroPAL work. We retrain this atlas using the full multi-lab corpus that we present in this work, increasing the training set by over 10-fold. The statistical atlas generated by this approach serves the additional purpose of characterizing the mean and covariance of neuron position and color across the whole corpus of data.

In **Fig. 3** we present visualizations of the statistical atlas of neuron colors and positions trained on 104 of the 118 worms in our consolidated NWB/DANDI dataset as well as on the smaller dataset of 10 worms used in the original NeuroPAL paper (employing the Statistical Atlas algorithm in Yemini et al. 2021 and Varol et al. 2020[6,28]). 14 worms were omitted from the atlas due to large nonlinear deformities or obvious artifacts. With 104 worms, this represents, to the best of our knowledge, the most broadly trained statistical atlas for *C. elegans* neuron positions and NeuroPAL coloring available. By leveraging the diversity of the multi-lab corpus this atlas captures variability between individual worms,

strains, and lab-specific experimental conditions. This atlas can be used as a basis for automatic labeling algorithms and biological investigations of neuron positions and brain organization. This statistical atlas further complements detailed electron-microscopy (EM) based anatomical atlases with cellular structural detail and provides nearly 100 more animals in its corpus than the approximately 10 EM ones available.[44,45,46] Although our corpus lacks the synaptic connectivity found in the EM datasets, it provides the complementary functional activity that is not available from EM imaging.

WormID.org supplies links to the software, visualization tools, and datasets discussed previously in this paper. Furthermore, WormID.org provides links to the data corpus and related tools to work with whole-brain structural and activity images, convert datasets to NWB, and supplies tutorials and instructions for using these tools. Our aim is that this data standard, data corpus, and atlas of cell positions will be a continually evolving resource for the *C. elegans* neuroscience community, and eventually other model organisms.

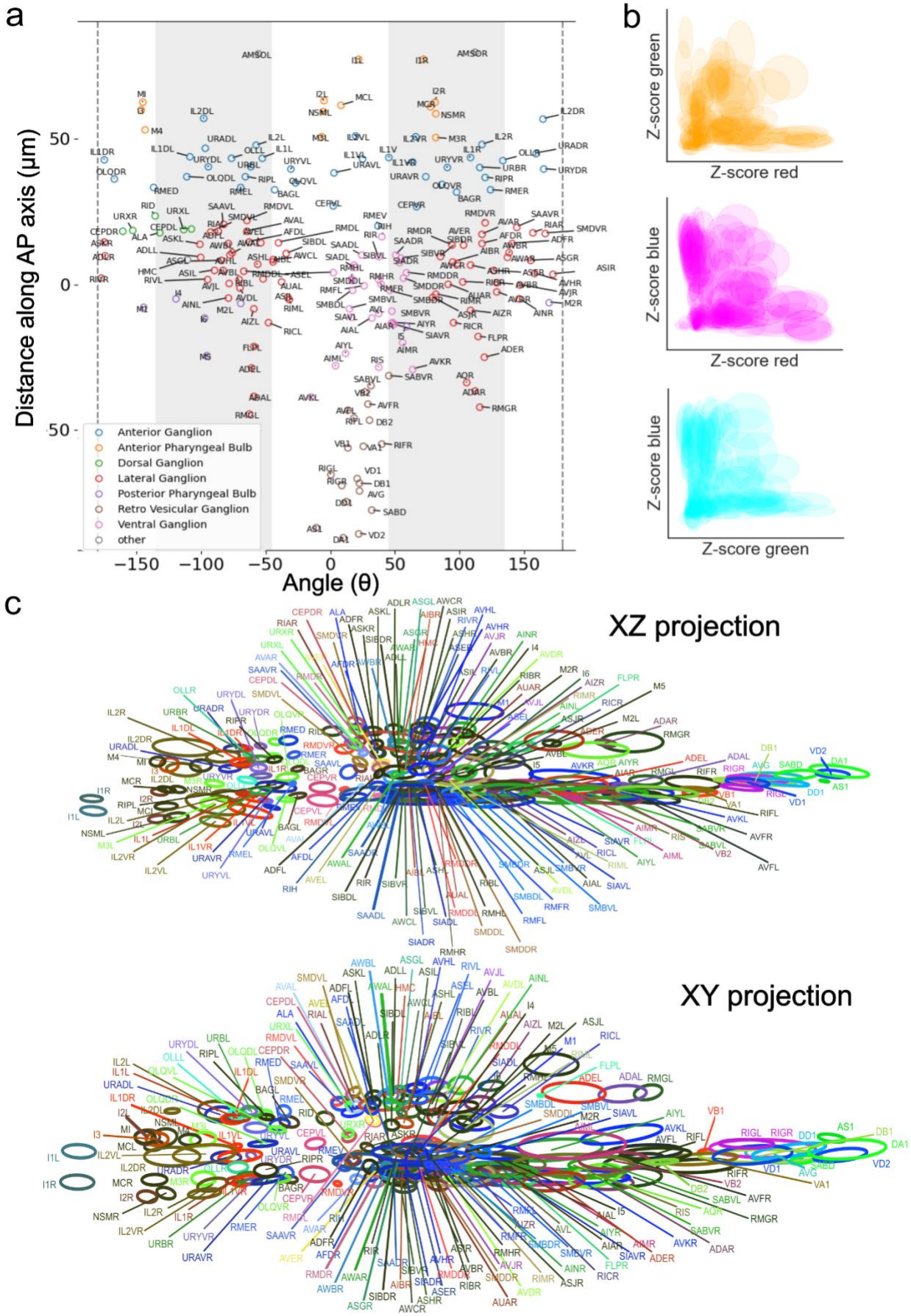
Analysis of biological factors in neuron positions

We statistically analyzed the spatial positions of *C. elegans* neuronal somas across individuals, strains, and lab conditions based on the mean and covariances in the statistical atlas. We focused on relative pairwise displacements rather than absolute positions because the absolute position of cells is dependent on positioning and deformation of the animal's body during recording and thus requires global alignment of all animals. Aligning multiple animals into identical positions is an imperfect task. In contrast, measuring pairwise cell positions does not require alignment and is relatively robust against animal positioning and deformation.

Before analyzing statistical properties on the neuron positions, we assessed the percentage of neurons that were labeled by humans in each dataset. We found that neurons in the ventral ganglion and retrovesicular ganglion were less commonly labeled than neurons in other ganglia. As is shown here and previously in Yemini et al. 2021[6], neurons in the ventral and retrovesicular ganglia exhibit high relative positional variability, which may explain why fewer of them were confidently labeled by researchers (**Fig. 4a**). For this reason, we explored several different factors hypothesized to contribute to the organization and variability of relative cell positions: i) gangliar boundaries (e.g., basal lamina and abutting tissue) which may restrict cell movement within the coelem, ii) synaptic connectivity which may impose energetic costs dependent on neuronal proximity, and iii) developmental-time and cell lineage effects whereby recently divided cells (i.e., sister cells) remain close together and more distant relatives (e.g., mother and grandmother cells) end up further apart.

Figure 3 (on next page): Multi-lab atlas of *C. elegans* neurons and their positional variability

(a) Butterfly plot showing the mean locations of neurons in the atlas colored by ganglion. (b) 2D color plots showing the distribution of neuron colors in the atlas. (b-c) Ellipses represent covariance (1 SD) and are centered at the mean for each neuron. (c) 2D projections of neuron positions and colors in the aligned atlas space. XZ projection (top) and XY projection (bottom). Ellipses are colored by the mean color, per color channel, for each neuron in the atlas.


a
Distance along AP axis (μm)
Angle (θ)
Anterior Ganglion
Anterior Pharyngeal Bulb
Dorsal Ganglion
Lateral Ganglion
Posterior Pharyngeal Bulb
Retro Vesicular Ganglion
Ventral Ganglion
other
b
Z-score green
Z-score red
Z-score blue
Z-score red
Z-score blue
Z-score green
c
XZ projection
XY projection

To test the first hypothesis, that gangliar boundaries regulate positional organization and variability, we measured the positional variability of neurons that are spatially close and compared pairs within the same ganglion to pairs straddling each other in different ganglia (see **Methods**). We found that neurons in the anterior pharyngeal bulb and neurons in the dorsal, lateral, and retrovesicular ganglia all exhibit significantly lower variability for pairs within the same ganglion compared to pairs in different ganglia. Conversely, for neurons in the anterior and ventral ganglia, we observed no significant difference between pairs in the same ganglion and pairs in different ganglia (**Fig. 4b**). We used an independent samples t-test to compare pairs within the same ganglion with those in different ganglia. Known anatomical features of the worm support this hypothesis: the pharynx is a muscular epithelial tube[47] that rigidly encases neurons; the remaining ganglia are separated by basal lamina that loosely restricts their boundaries[44]; finally, the anterior and ventral ganglia (and comparatively smaller retrovesicular ganglion) are completely bounded whereas all other ganglia are open at least at one end, and in White et al. 1986 it had been noted that tight cellular packing in these regions led to "slop", "uncertainty", and in live animals even "flipping" from side to side of the cells contained therein. This finding suggests that neural identification algorithms could be improved by a hierarchical approach, such as first predicting ganglia membership, then predicting neuron identities within each ganglion.

Next, we explored the relationship between somatic distance and synaptic connectivity. Overall, there was a very weak but statistically significant correlation between nuclear distance and synaptic weight for chemical synapses and no significance or detectable correlation for electrical synapses. However, we found that nearby neurons (mean distance < 40 μm) exhibit a wide range of chemical synaptic weights ranging anywhere from 0 to 70 synapses (with a median synaptic count of 3), whereas distant neurons (mean distance > 40 μm) have a maximum synaptic count of ~25 synapses (with a median count of 2) (**Fig. 4c**). This choice of distance cutoff was chosen by observing a distinct elbow at 40 μm in a 2D kernel density estimate plot of the scatter data (**Supplement 1)**. Our data suggests that neurons that are strongly wired together tend to be close to each other, although somatic proximity alone is not sufficient to imply strong connectivity. Recent findings in *C. elegans* have substantiated Peter's rule: neurons with larger colocalized axodendritic regions are more likely to form connections.[48] Our findings lend further support to this principle and suggest that close somatic or nuclear proximity also plays a role in determining neural connectivity.

Lastly, we explored the hypothesis that cell lineage is a determinant of adult cell positioning. Embryonic *C. elegans* are confined to a fixed volume within an eggshell approximately 50 μm in length and 30 μm in diameter.[49] After hatching, they grow over 4x in length from birth (~250 μm) to adulthood (over 1 mm), with an exponential expansion in their volume.[50,51] Sister cells are cells whose lineage differs only at the very last division. We hypothesized that animal growth should lead to both larger distances and higher variability between older sister cells that divided in the embryo, versus younger sister cells born much later at postembryonic larval stages of development. Surprisingly, we found no statistically significant correlation between the time of cell division and nuclear distance or between time of cell division and distance variability measured by SEM (**Fig. 4d**). In fact, most sisters remained within 15 μm of each other (~3 nuclei apart) at adulthood, regardless of when they were born. Strikingly, a substantial cohort of embryonic sisters ended up closer together at adulthood (< 8 μm) than those dividing at larval stages that occur more than 16 hours later (**Fig. 4d**). Our data rule out exponential postembryonic growth spurts as a major determinant of divergence and variability in neuron positions.

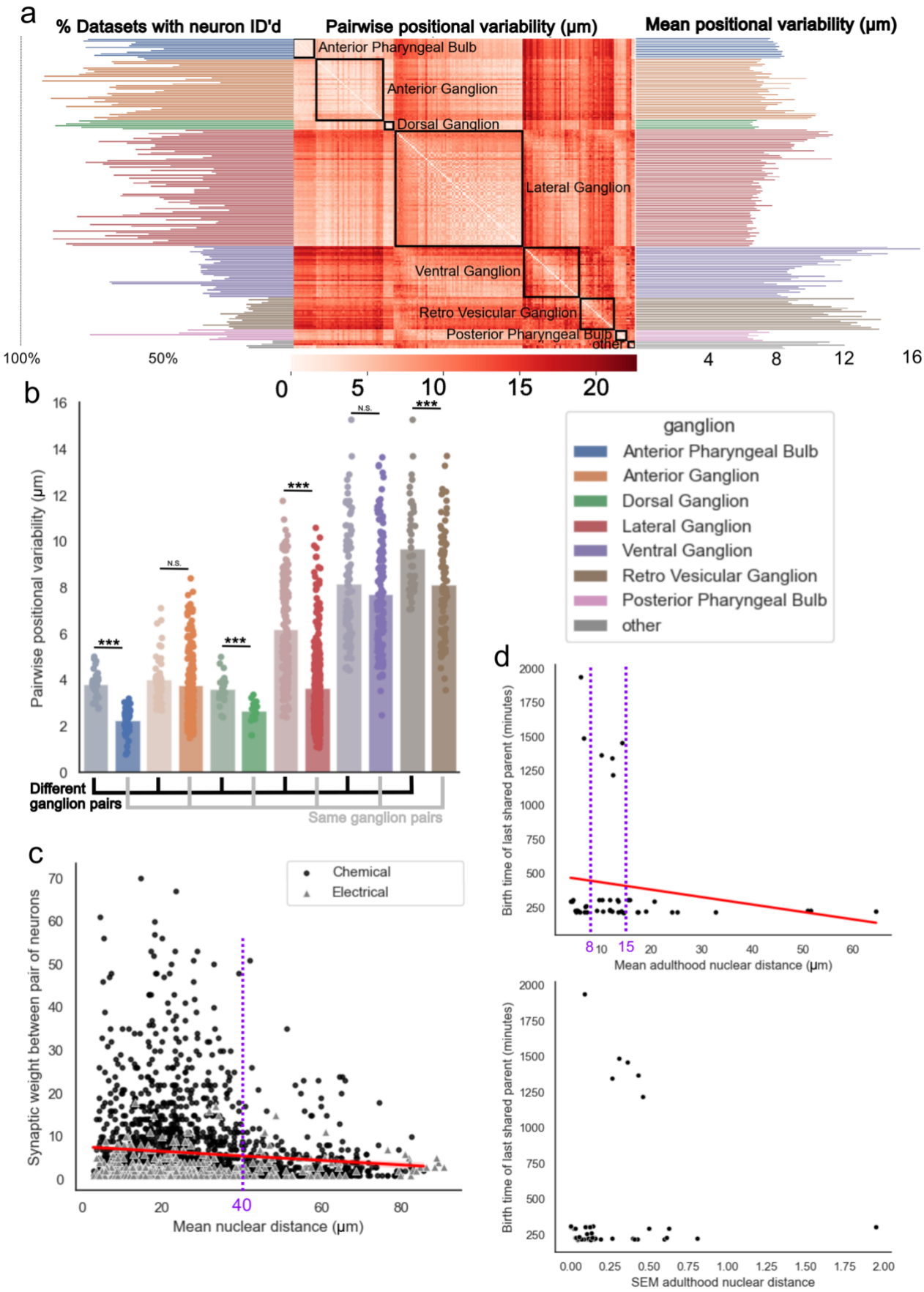
Neuron identification performance increases for all laboratories and all tested algorithms when trained on a harmonized multi-lab corpus

Our previously published statistical atlas algorithm ("StatAtlas") for automated neuron identification was trained on a homogenous dataset of 10 NeuroPAL worms.[6,28] Formerly, this 10 worm training set achieved average accuracies of 86% overall in head neurons that ranged from 50% for the ventral ganglion to 100% for the anterior pharyngeal ganglion. These accuracies facilitate neural identification, but in practice they require substantial verification and manual corrections, and thus necessitating significant time expenditure in these tasks. Moreover, the algorithm fails to generalize to datasets produced by other labs (**Fig. 5b**). We tested the performance of our previously published algorithm on each of the six aggregated datasets. Initial performance on these datasets ranged from ~21% to ~65% with an average of 41%, (**Supplement 2**). This substantial decrease in accuracy on datasets from different labs exposes the limitation of using single-lab training sets to produce tools intended for use by different labs with different instrumentation, experimental methods, and data acquisition pipelines.

To assess the performance benefits of using a large, harmonized corpus to train commonly-used automated neural-identification methods we tested two more popular algorithms: coherent point drift ("CPD")[52] and CRF_ID[26]. Coherent point drift is an untrained and unsupervised algorithm that: 1) globally aligns a sample point cloud of neurons to a reference atlas, then 2) locally matches points from the sample to their nearest neighbors in the atlas, and finally 3) identifies sample neurons (points) by their corresponding matches in the atlas. CRF_ID is a newer graph-based approach that identifies neurons using a combination of statistics from their individual features (e.g., absolute position and color) and pairwise relationships (e.g., displacement and angle relative to each other). There are no currently published benchmarks on the neural identification problem using CPD. Formerly, CRF_ID demonstrated a high accuracy of 83% when originally trained and tested solely on the HL dataset. Similar to StatAtlas, when testing the generalizability of the CPD and CRF_ID base models on the full WormID corpus we observed poor performance with an average overall accuracy of 39% and 59% respectively.

After inspecting recordings from multiple labs, we hypothesized that differences in color space may have negatively impacted algorithmic performance. Potential sources of color space variability include differences in microscope hardware, software and image settings, and configuration of the optical path. Anecdotally, in addition to these known sources of variability, researchers also typically adjust exposure, contrast, and other channel display parameters to make the composite rendered colors appear more like the images in the NeuroPAL reference manual.[53] In aggregate, this suggested that harmonizing the color space may aid automatic algorithms.

Figure 4 (on next page): Analyses of neuron positions, distances, and positional variability
(a) Left: percentage of datasets containing each labeled neuron, organized anterior-posterior within each ganglion. Middle: heatmap of the STD of pairwise positional distances between each pair of neurons across datasets. Right: Averaged sums of heatmap rows. Neurons with higher mean positional variability have less stereotyped positions within the worm body. (b) Pairwise positional variability by ganglia for 10 closest neighbors of each neuron, separating neuron pairs in the same ganglion from pairs in different ganglia. Anterior pharynx: effect size 95% CI [-1.79, -1.38] µm, $p = 2.5 * 10^{-21}, N_{same} = 53, N_{diff} = 36$; Dorsal: effect size 95% CI [-1.16, -0.72] µm, $p = 7.1 * 10^{-6}, N_{same} = 13, N_{diff} = 27$; Lateral: effect size 95% CI [-2.80, -2.38] µm, $p = 5.3 * 10^{-31}$, , $N_{same} = 319$, $N_{diff} = 159$; Retrovesicular: 95% CI effect size=[-1.97,-1.19]µm $p = 9.3 * 10^{-5}$, $N_{same} = 89$, $N_{diff} = 43$; Anterior: $p = 0.161$, $N_{same} = 176$, $N_{diff} = 50$; Ventral: $p = 0.252$, $N_{same} = 52$, $N_{diff} = 37$. (c) Relationship between pairwise neuron synaptic weights and their mean positional distance for chemical and electrical synapses. Chemical synapses: KendallTau $\tau = -0.036$, $p = 0.021$, Pearson $R = -0.098$, $p = 6.4 * 10^{-6}$, N=2119; Electrical synapses: KendallTau $\tau = 0.009$, $p = 0.80$, Pearson $R = 0.031$, $p = 0.51$, N=444. (d) Relationship between cell birth times and the mean and SEM of their nuclear positional distance in adulthood for sister cells. Mean: KendallTau $\tau = -0.144$, $p = 0.052$, Pearson $R = -0.161$, $p = 0.129$; SEM: KendallTau $\tau = 0.014$, $p = 0.845$, Pearson $R = 0.074$, $p = 0.487$.Most sisters are within 15 µm of each other in adulthood. More sisters that divide embryonically remain close together (<8 µm) than sisters that divide >16 hours later at postembryonic larval stages.

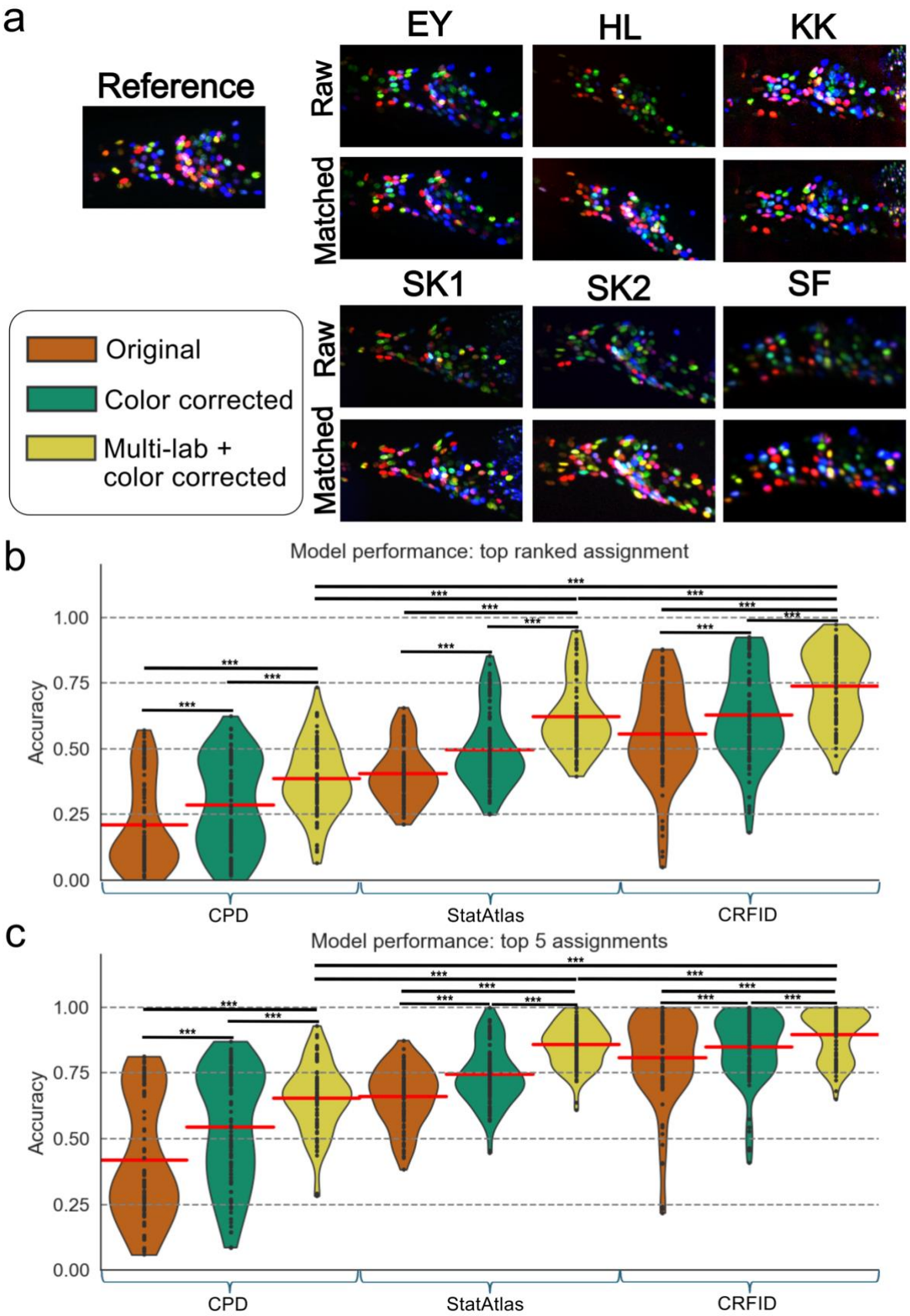

a
% Datasets with neuron ID'd
Pairwise positional variability (µm)
Mean positional variability (µm)
Anterior Pharyngeal Bulb
Anterior Ganglion
Dorsal Ganglion
Lateral Ganglion
Ventral Ganglion
Retro Vesicular Ganglion
Posterior Pharyngeal Bulb
other
100%
50%
0
5
10
15
20
4
8
12
16
b
Pairwise positional variability (µm)
N.S.
***
Different ganglion pairs
Same ganglion pairs
ganglion
Anterior Pharyngeal Bulb
Anterior Ganglion
Dorsal Ganglion
Lateral Ganglion
Ventral Ganglion
Retro Vesicular Ganglion
Posterior Pharyngeal Bulb
other
c
Synaptic weight between pair of neurons
Chemical
Electrical
40
Mean nuclear distance (µm)
d
Birth time of last shared parent (minutes)
8
15
Mean adulthood nuclear distance (µm)
SEM adulthood nuclear distance

We developed an approach to match the color histogram of a sample image to a reference histogram representing ideal coloring (see **Methods**). Histogram-matching the original small training set improved the accuracy of all three tested algorithms by an average of 8%, 9%, and 7% for CPD, statistical atlas, and CRF_ID respectively. It also qualitatively made composite color renderings better match the NeuroPAL reference manual, aiding users in annotating and correcting algorithmic predictions (**Fig. 5a,b,c**).

Given this success on the original small training set, we used the histogram-matched images to train a new atlas for the StatAtlas and CRF_ID algorithms on the full corpus of data. Test accuracy is reported using 5-fold cross-validation where each worm is tested against an atlas that was not trained on that worm. For CPD, we updated the algorithm to select the best template out of the full corpus (see **Methods** for further details). This led to significant improvement in accuracy across algorithms (**Fig. 5b, c)**, with an average improvement of 17%, 22%, and 18% that further raised average predictive accuracy from 22% to 39%, 41% to 62%, and 55% to 74% for CPD, StatAtlas, and CRF_ID respectively. This is equivalent to a ~1.3x, ~1.6x, and ~1.7x reduction in error rate. Accuracy reached as high as 95% for several individual datasets for both StatAtlas and CRF_ID. Furthermore, when considering the top 5 neural identity assignments (rather than just the top 1), the multi-lab models showed average accuracies of 65%, 86%, and 89% for CPD, StatAtlas, and CRF_ID respectively, with some datasets reaching 100% accuracy for both StatAtlas and CRF_ID (**Fig. 5c, Supplement 3**). In addition, we see similar improvements in accuracy across most datasets for StatAtlas and CRFID when training using all except one dataset and then testing on the left-out dataset (**Supplement 6**). This indicates that most of the benefits from retraining come from achieving a better representation of the full diversity across datasets, rather than capturing the specific nuances of any one dataset. This generalizability will enable labs to use these retrained algorithms out of the box rather than needing to do additional fine-tuning on their own data.

Differences in accuracy between datasets may have been caused by a variety of factors including poor initial alignment, optical quality, non-neuronal artifacts in the images, and nonlinear deformations of the worm body. Additionally, datasets with less neurons annotated had better automatic labeling accuracy, presumably because experimenters only labeled the easiest neurons to identify and left the hardest ones unannotated (**Supplement 4,5**).

Figure 5: Improvements in neural identification accuracy (on next page)
(a) Examples of raw and color-corrected (histogram matched) images from each lab and dataset. **(b)** Top ranked test accuracy for training set of original 10 reference worms with no color correction (orange), 10 reference worms with color correction (green), and the multi-lab corpus with color correction (yellow) for coherent point drift (CPD, left), the statistical atlas model (StatAtlas, middle), and the conditional random field model (CRF_ID, right). Algorithmic performance was evaluated using paired t-tests (where N=94 for each training set) to compare the performance of different atlases. Significance is reported using a Bonferroni correction with the convention of * for $p<0.05$, ** for $p<0.01$, and *** for $p<0.001$. **(c)** Same as b but using top 5 rank. Summary statistics and p values can be found in supplementary table 1.


a
Reference
EY
HL
KK
Raw
Matched
SK1
SK2
SF
Raw
Matched
Original
Color corrected
Multi-lab + color corrected
b
Model performance: top ranked assignment
Accuracy
1.00
0.75
0.50
0.25
0.00
CPD
StatAtlas
CRFID
c
Model performance: top 5 assignments
Accuracy
1.00
0.75
0.50
0.25
0.00
CPD
StatAtlas
CRFID

Discussion

Aggregation and harmonization of data from a variety of different sources is necessary to build a corpus for analytical methods and machine learning tools that generalize across the diversity of real-world data. In this work, we present a data harmonization pipeline for analyzing whole-brain structural and activity imaging in *C. elegans*. This pipeline includes data aggregation, conversion to a standardized file format, software for analyzing these standardized datasets, pre-processing approaches to align images and color spaces, and spatial registration of sample neuron point clouds to a common atlas.

We used this corpus to study potential biological factors that organize cell position in *C. elegans.* Specifically, we find that: i) restrictions in bounding tissue and gangliar space likely contribute to variability in neuron positions, ii) neurons with somatic distances less than ~40 μm of each other show higher synaptic connectivity, and iii) sister neurons that divide in the embryo can be found closer together at adulthood than ones dividing at larval stages more than 16 hours later. The positive relationship between synaptic connectivity and neuron somatic proximity thus augments the previously observed correlation of synaptic connectivity to axodendritic adjacency, termed Peter's Rule. Moreover, the close distances and low positional variability we measured for embryonically born sister neurons rules out exponential organismal growth as a major cause in driving neurons apart from each other during the establishment of the adult Bauplan.

We then used the corpus to train a machine learning tool to automate the intensive task of labeling cells in these datasets. This substantially boosted generalized performance across datasets from contributing labs for each tested algorithm, despite the variability in data from these different groups. Accuracy of auto-identification now approaches human performance for certain datasets using the StatAtlas or CRF_ID algorithms. In the future, our corpus can be used to incorporate neuronal shape- and size- based descriptors as well as dynamical time-series features to further improve neural identification algorithms.

The WormID.org tools and resources are readily applicable to new whole-brain structural and activity imaging datasets, and these new datasets can be easily added to the existing corpus. These tools streamline public data sharing to facilitate both open science and to satisfy data-sharing mandates. We hope this resource will continue to grow in size and breadth to enable the development and benchmarking of new machine learning tools and algorithms. Our analyses of cell features based on the full corpus of data can immediately be used to inform better feature selection and algorithms that continue to improve automated approaches for neuron-subtype identification in volumetric images. Additionally, the large corpus and trained statistical atlas can serve as a descriptive resource of the underlying neurophysiology of *C. elegans*. Moreover, our resources can be incorporated into computational neurobiology courses, such as the Neuromatch Academy (neuromatch.io)[54] to train the next global generation of neuroscientists on real-world datasets. As the community continues to develop new tools, this corpus will allow these new tools to be benchmarked for generalizable performance, spurring innovation.

As the community continues to scale up the generation of neural data and increasingly relies on machine learning analysis to tame this "big data", there is an ever-growing need to unify disparate datasets to produce verifiably robust, accurate, and generalizable analytical approaches. Harmonization efforts such as ours can significantly reduce the activation energy necessary for collaboration, data sharing, and the development of unified community-wide tools across labs. While some of the resources we created are specific to *C. elegans*, the framework and much of our toolkit can be applied to other model organism imaging communities.

Methods

Standardized file format - Neurodata Without Borders

NWB is an HDF5-based format built specifically for neurophysiology data and has emerged as the de facto standard for storing neurophysiology datasets with associated metadata for reuse and sharing. NWB provides object types for data and metadata including acquisition parameters, segmentation of 3D image regions, fluorescent time series (e.g., for neural activity), experimental design information, multichannel electrophysiology time series data, 3D images, stimulus events during an experiment, and behavioral data.[42]

The base NWB schema supports two-dimensional structural and time-series multi-channel images but did not originally support the type of five-dimensional (multi-channel, volumetric, time-series) data that is used in *C. elegans* whole-brain activity imaging or other metadata associated with these types of experiments. To solve this problem, we developed 'ndx-multichannel-volume' as a novel extension to the existing Neurodata Without Borders (NWB) standardized file format. More information and resources about NWB can be found at nwb.org.

Our extension adds new objects built off of existing ones in the schema to add and improve support for multi-channel, volumetric, time-series images and the metadata associated with those images as well as volumetric segmentation data and metadata fields specific to *C. elegans* such as cultivation temperature and growth stage. This extension and the datasets presented in this work represent the first applications of the NWB data format to *C. elegans* and have now been incorporated as the standard for this model organism. This extension is flexible, open-source, and can be continuously updated to incorporate new types of data for future experiments.

Storage on DANDI

Data and associated metadata were uploaded to the DANDI archive [RRID:SCR_017571] using the Python command line tool (https://doi.org/10.5281/zenodo.3692138). The data were first converted into the NWB format (https://doi.org/10.1101/2021.03.13.435173) and organized into a BIDS-like (https://doi.org/10.1038/sdata.2016.44) structure.

All datasets can be streamed or downloaded from the DANDI archive, available on WormID.org as well as these individual URLs:[55,56,57,58,59,60,61]
Original NeuroPAL: https://doi.org/10.48324/dandi.000715/0.240614.1942
EY: https://doi.org/10.48324/dandi.000472/0.240625.0454
HL: https://doi.org/10.48324/dandi.000714/0.240611.1954
KK: https://doi.org/10.48324/dandi.000692/0.240402.2118
SF: https://doi.org/10.48324/dandi.000776/0.240625.0015
SK1: https://doi.org/10.48324/dandi.000565/0.240625.0439
SK2: https://doi.org/10.48324/dandi.000472/0.240625.0450

Software systems with embedded NWB I/O

We present two software examples with user-friendly GUIs to interface with NWB datasets and run standard analysis pipelines for cell segmentation, identification, tracking, and extracting time-series of neural-activity traces annotated with any experimental stimuli presented. First, we present the NeuroPAL_ID software (**Figure 2c**) for visualization, annotation, neuronal segmentation and

identification, neural tracking, activity trace extraction and stimulus presentation data of volumetric NeuroPAL images and whole-brain activity. This software is pre-compiled for use on MacOS and Windows. The software is open-source, available from https://github.com/Yemini-Lab/NeuroPAL_ID/releases, and is written in MATLAB and Python. It has now been updated to include functionality described in this paper to enable histogram matching, color-corrected image visualization, and automated neural identification using the new Statistical Atlas. This software is written and managed by the Yemini Lab; further information can be found at https://www.yeminilab.com/neuropal.

Second, we present the eats-worm software for visualization, segmentation, and activity extraction of neural-activity time series from immobilized worms. Eats-worm similarly allows for manual verification and curation of the automatic segmentation and tracking algorithms. The tracking algorithm was optimized for tracking neurons across frames in immobilized worms, but there are currently efforts to extend this functionality to work for freely-moving worms as well. Eats-worm is written in Python and is built as a plugin to Napari, a popular 3D visualization tool. This software is written and managed by the Kato Lab; further information can be found at https://github.com/focolab/eats-worm.

Both software programs have embedded functionality to read and write NWB files. NWB I/O functionality enables a user to quickly run similar analyses on all of the datasets presented in this work without the need to develop specific pipelines to read in data from each dataset. Furthermore, this functionality can be easily embedded into MATLAB or Python-based analysis software.

Data acquisition

NeuroPAL structural volumes and neural activity time series volumes were acquired using the protocols outlined in Yemini et al. 2021.[6] After collection of these images, neurons were segmented and annotated according to the guidelines in the NeuroPAL manual.[53] Specific immobilization methods, microscope setup, and experimental protocols differ slightly between datasets. All datasets were taken using spinning disk confocal microscopes with XY resolution varying from 0.1604-0.54 µm/pixel and Z resolutions varying from 0.54-1.5 µm/pixel. XY resolution was the same for NeuroPAL structural images and neural-activity images (using GCaMP6s) for all datasets, but Z resolution varied from 0.54-3 µm/pixel. Z resolution is generally lower for neural-activity images due to limitations in optical sectioning with confocal microscopes. Lower Z resolution also helps to reduce the number of frames needed to record a full volume for a single time-point, to aid imaging at a higher temporal resolution. Most images were taken with the worm immobilized in a microfluidic chip with the exception of the KK dataset (where worms were semi-restricted in a microfluidic device) and the SF dataset (where worms were freely moving). The NWB files and the DANDI datasets that hold them contain metadata for the specific setup and conditions in each dataset. For published datasets, additional information can be found in the associated publications.[6,7]

After acquisition of NeuroPAL structural volumes and whole-brain activity time-series, images were segmented using various automatic segmentation algorithms ranging from classical computer vision approaches (e.g., template matching[64]) to deep neural network approaches. These were then manually verified. Ground truth annotations were done using a combination of existing automatic identification algorithms followed by manual corrections. Each neuron identity label was either explicitly annotated by experts or manually verified after algorithmic identification. Note that varying levels of completeness in labeling are due to the difficulty of this manual annotation task. For several datasets with lower image quality, even experts could only confidently label 30-50% of segmented neurons in the volume. For neural activity time-series, neuron centers were first tracked across images

using various algorithms and then manually verified by experts.[62,63] Fluorescence activity is then extracted from these tracked ROIs to obtain time series of neural-activity traces. Neurons in the NeuroPAL structural volume were then matched to the ROIs in the neural activity time-series to get labeled activity traces.

Datasets from various labs were converted to the NWB standardized file format using the ndx-multichannel-volume extension presented in this work. These files were then uploaded to the DANDI archive where they are now publicly accessible for data streaming, download, or online visualization.

Dataset	Microscope	Length of recording	Sample rate	Resolution (um/pixel)	Strain	Setup
NP_og	Zeiss LSM880 spinning disk confocal	~4 min	~4Hz	0.208 x 0.208 x 1.02	OH16230	Microfluidic chip
SF	Andor spinning disk confocal w/ Nikon ECLIPSE Ti microscope 40x water immersion	~15 min	1.7 Hz	0.54 x 0.54 x 0.54	Various	Freely moving
SK1	Leica DMi8 inverted spinning disk confocal, 40x WI, 1.1 NA	~25 min	1.04 Hz	0.1604 x 0.1604 x 1 (3 for calc images) OR 0.3208 x 0.3208 x 0.75 (2.5 for calc images)	FC121, FC128, OH16230	Microfluidic chip
SK2	Leica DMi8 inverted spinning disk confocal, 40x WI, 1.1 NA	~15 min	3.3 Hz	0.3208 x 0.3208 x 0.75 (1.5 for calc images)	OH16230	Microfluidic chip
KK	Nikon Eclipse Ti-U inverted spinning disk confocal, 40x 1.3 NA	~15-20	1.67	0.32 x 0.32 x 1.5 for both images	KDK92	Semi-restricted in microfluidic device
HL	Perkin Elmer spinning disk confocal 1.3 NA, 40x oil OR Brucker Opterra II swept field confocal 0.75 NA, 40x air	NA	NA	0.33 x 0.33 x 1	OH15495	Microfluidic device
EY	Spinning disk confocal	~4 min	~4 Hz	0.27 x 0.27 x 1.5	OH16230	

Butterfly plot

To produce the butterfly plot, we first manually found three orthogonal basis vectors to align neuron point clouds to a new cartesian coordinate space. To do so, we used human-guided affine transformation to roughly align these basis vectors to the anterior-posterior, dorsal-ventral, and left-right axes. The xyz coordinates of each neuron were projected into this new cartesian coordinate space and

then converted to cylindrical coordinates by the following equations. We plotted the new x and θ coordinates on a 2D plane to get the butterfly plots shown in (**Fig. 1a**, **Fig. 5a**). This projection is akin to flattening the positions of the neurons along the circumference of a cylinder of the worm body and then unrolling that cylinder into a flattened plane.

$$x_{cylinder} = -x_{new}$$
$$r = \sqrt{{y_{new}}^2 + {z_{new}}^2}$$
$$\theta = arctan2(y_{new}, z_{new})$$

Histogram matching

We modified the established approach of histogram matching to apply to 3-D volumetric, multi-channel data.[64] We created a reference histogram using the 10 worms from the original NeuroPAL work. This data is stored as uint16, so there are 65,536 possible values for each pixel. For each channel, we created a histogram counting the number of pixels within the bin edges, assigning each color value its own bin, and then averaged the values in each of these bins across the 10 images. Practically, these histograms were very similar across these 10 datasets, so the averaged histogram looked similar to each of the individual histograms.

To color match a new animal sample, we calculated a histogram for each channel. The number of bins for each channel histogram was equal to the maximum intensity value present in that channel in the image. Practically, this means that there are a different number of histogram bins for each channel in each image because images were collected at different bit depths and with varying levels of saturation.

$$bincount_{channel} = max(x_{channel})$$

We then calculated a cumulative density at each color value for both the sample and the reference. We created a lookup table M to associate each gray count value x in the sample to the color value in the reference with the closest cumulative density. Next we created a new matched image with each pixel transformed into the new color space using this lookup table as shown below.

$$M_{channel}(x) = \left(\left|cdf_{sample}(x) - cdf_{reference}(x')\right|\right)$$
$$MatchedImage_{channel}(i,j,k) = M_{channel}(A(i,j,k))$$

Color extraction

To extract the color values for the neurons in each image, we first calculated the mean and standard deviation of the pixel gray counts in each channel, and then converted each pixel value into its Z-score based on its gray count value. We then took a sample of a 3x3x1 grid of pixel values around each segmented neuron center in each channel. We use the median values of this 3x3x1 grid as the RGB values for that neuron center. Color values were extracted post-histogram matching when training or testing using histogram-matched images. For non-histogram-matched images, there are no additional color pre-processing steps beyond Z-scoring.

Positional variability analysis

We calculated pairwise positional variability by measuring the Euclidean distance between every pair of canonical head neurons across each structural volume when both neurons in that pair had a ground truth label. We then took the average and standard deviation of these distances for each neuron pair to find mean nuclear distance and pairwise positional variability, respectively. For these analyses we ignored pairs that are not present in at least 5 datasets. We used pairwise positional variability instead of absolute positional variability because absolute position is extremely sensitive to point-cloud realignment, which would make it hard to disaggregate natural positional variability from alignment errors; and furthermore, we are not interested in only how individual cells vary but rather how cells vary relative to each other. To get the mean positional variability for a given neuron, we averaged the mean pairwise distance for all pairs that contained that neuron.

Intra vs inter ganglion measures: for every neuron in the atlas, we found its n closest atlas neighbors and only measured pairs for its n closest neighbors. We then separated these pairings based on whether the two neurons in the pair are within the same ganglion or in two different ganglions. Note that pairs in different ganglions will appear twice: e.g. if one neuron in the pair is in the anterior ganglion and the other is in the lateral ganglion, the pair will be counted in the analysis for both the anterior ganglion and the lateral ganglion (**Fig. 4b**). Pairs within the same ganglion are only counted once. We compared this approach for n =1-20 (**Supplement 7**). For all numbers of neighbors there is higher positional variability for neighbors in different ganglia when compared to the same ganglion. The pattern stabilizes around n=7 and holds steady through n=20. Therefore, we selected 10 to use for n in our analysis.

Synaptic connection: synaptic weights between neuron pairs are derived from the whole-brain connectome of the adult hermaphrodite in Cook et al. 2019.[45]

Lineal distance: the cell lineage tree and associated birth times were taken from Sulston et al. 1982.[63] The last shared parent cell between two neurons is the most recent shared parent node in the lineal tree. We used the birth time of the last shared parent cell between two neurons as the lineal distance and explored the relationship between this lineal distance and mean pairwise nuclear distance (**Fig. 4d**). In this analysis, we focus only on sister cells: terminal cells that only divided from each other at the very last stage of their lineal tree.

Coherent Point Drift (CPD)

Coherent point drift has been a common algorithm for registering two similar point clouds to each other since its introduction in Myronenko and Song 2009[52]. CPD allows for both rigid and non-rigid point set registration. CPD models one point set as a set of GMM centroids that are fit to the second point set by maximizing the likelihood. GMM centroids are set to move coherently to preserve the structure of the point clouds. In the rigid case, the algorithm learns an affine transformation of the GMM centroid locations while in the non-rigid case, the algorithm learns a displacement function on the original centroid positions with an enforced regularization term to enforce smoothness. The objective function is optimized using an iterative EM optimization approach and yields both the aligned point set as well as an NxM correspondence probability matrix that represents the likelihood that each point n in set 1 corresponds to each point m in point set 2.

In this paper, we use the specific implementation of CPD used in Yu et al. 2021[34] First, rigid CPD is used to roughly align a test worm point cloud to a template point cloud. Then, non-rigid CPD is used to model non-linear deformations between the semi-aligned test and template. Neuron assignments

are then determined by creating a matrix of pairwise Euclidean distances between every neuron's position and color in the test and every neuron in the template in the aligned space. We then use the Hungarian algorithm on this distance matrix to find the optimal label assignments. To get 2nd ranked assignments, we assigned an infinite cost to each label assignment from the first pass and reran the Hungarian algorithm. We repeat this for the 3rd-5th order assignments.

Accuracy was calculated by counting the number of neurons whose algorithmic assignment was the same as the ground truth label and then dividing by the total number of neurons that have a ground truth label. Note that neurons without a ground truth label were not included in the accuracy metric but are still part of the cost matrix and received neuron assignments. Since there is no ground truth for these neurons we did not determine the accuracy of their label assignments.

The difference between the use of the 'original (10) worms' versus 'multi-lab corpus' for CPD is what set is included in possible options for the template. For the 'original' group, we compare every test set to each of the original 10 neuroPAL worms and report the accuracy for the template that has the highest average probability of correspondence after the rigid alignment step. Similarly, for the 'multi-lab corpus', each test worm is compared to each possible template worm in the whole multi-lab corpus and accuracy is reported similarly. The accuracy of CPD is highly sensitive to a good rough initial alignment and to similarity of the template and the test point cloud. The template with the highest average probability of correspondence is not necessarily the template that yields the highest accuracy, but it is the template that the algorithm has the highest confidence that it has found the 'correct' correspondence.

Statistical Atlas training and inference

Statistical atlases used for testing performance were trained using the algorithm described in Varol et al. 2020.[28] This algorithm uses a training set of neuron point clouds with both XYZ and RGB values and takes a block-coordinate descent approach where it iteratively learns affine transformation parameters to align the neuron point clouds, then updates the means and covariances of the positions and colors of each neuron until reaching convergence. This process generates mean and covariance parameters for each neuron as well as an aligned coordinate space for all the worms in the training set. The trained atlas consists of a list of neuron names alongside their associated means and covariances in the aligned position and color space.

We trained three atlases: the original atlas trained on just the original 10 NeuroPAL worms from Yemini et al. 2021[6], the color corrected atlas trained on these same 10 worms after histogram matching, and the multi-lab + color-corrected atlas which is trained on the full corpus of histogram-matched data.

For the original atlas, we tested every dataset in the full corpus without histogram matching. For the color-corrected atlas, we similarly tested every dataset on the full corpus of data with histogram matching. For the atlas trained on the full corpus, we use K-fold cross-validation. The corpus was split into five equally-sized groups. For each group, an atlas was trained on all datasets in the other four groups and performance was reported for the out-of-training set group. The 10 worms used to train the original and color-corrected atlas were included in the training for each of these five groups. These 10 worms were not used to report testing accuracy for any of the atlases (**Fig. 5**). The fully trained atlas presented in **Fig. 3** was trained using the 10 original worms and the full corpus of data presented in this work, without splitting it into groups. This full atlas was embedded into the autoID functionality of the NeuroPAL ID software shown in **Fig. 2**.

Neuron point clouds used for testing were pre-aligned by learning an affine transformation from each sample dataset to the aligned coordinates of the atlas based on a subset of the ground truth labeled

neurons in the sample. Briefly, assuming N neurons in the test sample and M neurons in the atlas, we calculated an NxM cost matrix using the Mahalanobis distance between each neuron center in the sample and each neuron distribution in the atlas. $\underline{x_i}$ represents the XYZRGB values of neuron i, while $\underline{\mu_j}$ and Σ_j represent the XYZRGB mean and covariance respectively for neuron j in the atlas.

$$Cost_{i,j} = (\underline{x_i} - \underline{\mu_j})^T \Sigma_j^{-1} (\underline{x_i} - \underline{\mu_j})$$

We then treated this cost matrix as a linear sum assignment problem. Label assignments (for neural identification) were calculated using the Hungarian algorithm.[43] 2nd-5th order ranked assignments and accuracy are calculated in the same way as described for CPD.

CRF_ID training and inference

CRF_ID atlases and inference are conducted using the algorithm described in Chaudhary et al. 2021[26]. This approach follows a probabilistic-graphical-model framework based on conditional random fields. The graph is defined by node features corresponding to unary measures for each neuron center such as position and color and edge features corresponding to pairwise measures for each pair of neurons such as distance, relative angle, or probability that one neuron is anterior to the other. After features are selected, a data-driven atlas is trained on a corpus of data to determine the average values for each of the measured features; then for a test worm, node and edge potentials are calculated based on comparison of each feature in the test worm to the atlas and infer the best global assignment of labels by maximizing an energy function using an approximate inference method. For the analysis in this work, we used the color information solely to define the node potentials, and the pairwise angle relationships only to define the edge potentials. Optimizing the weights of the node and edge features may result in a higher prediction accuracy.

We trained three atlases: the original atlas trained on just the original 10 NeuroPAL worms from Yemini et al. 2021[6], the color corrected atlas trained on these same 10 worms after histogram matching, and the multi-lab + color-corrected atlas which is trained on the full corpus of histogram-matched data. This training approach follows the same K-fold cross validation approach used for the Statistical Atlas method.

We use the roughly pre-aligned point clouds used in the Statistical Atlas algorithm as input to the CRF_ID algorithm to eliminate possible difference in the initial alignment step, which can dramatically change accuracy.

In practice, there are nearly always fewer detected neuron centers in a given image than total cells in the atlas. CRF_ID handles this by modeling a hidden variable $\boldsymbol{h} \in \{0,1\}^N$ where N is the number of neurons in the atlas. This variable specifies the probability that a given cell is missing in the image. Based on the number of cells in the test image, P cells are uniformly selected across different regions of the head and removed from the atlas. This process is repeated ~1000 times to sample multiple possible combinations of **h.** The top 1-5 predicted assignments are generated by compiling a list of the most frequent labels for each cell in the test image across all runs. Accuracy is reported in the same way as CPD and CRF_ID.

Optimizing the aforementioned energy function using an approximate inference method produces marginal distributions of label assignments for each cell. The top 1-5 predicted label assignments for each cell were generated by sorting the marginal probability of labels in a descending order. The label that resulted in the highest marginal probability was assigned as top 1.

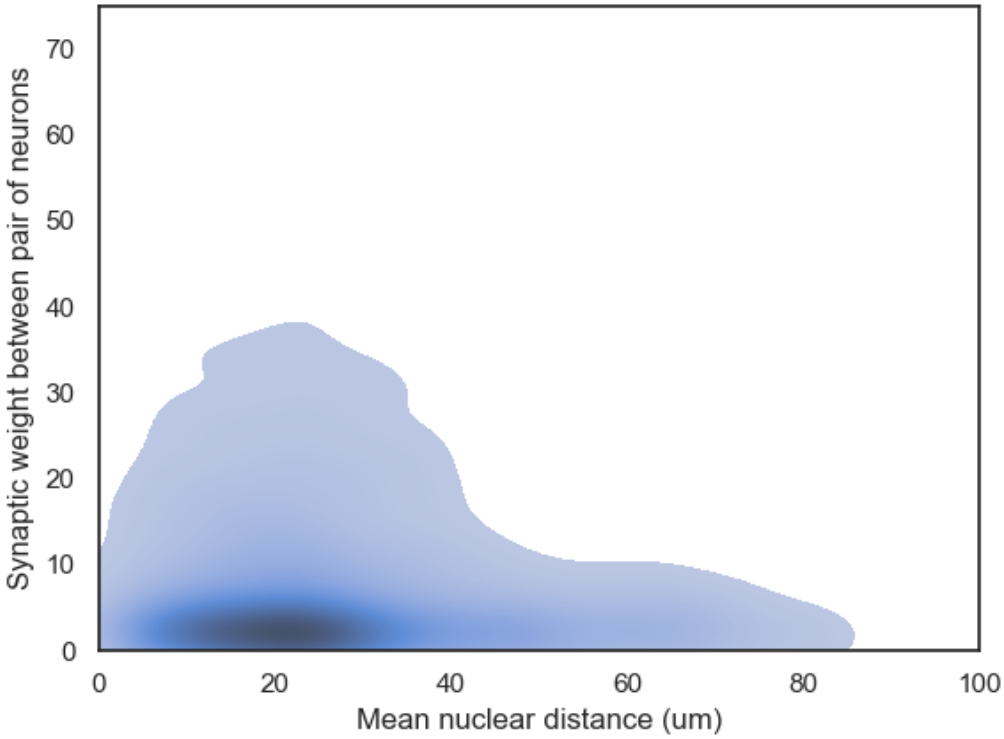
Author contributions

D.S. developed the NWB extension format, harmonization algorithms and code, and performed data analysis. K.R. updated the NeuroPAL ID software and performed data analysis. J.B., S.B. and R.D. developed the eats-worm software. H.J.L. and Z.Y. performed data analysis. D.S, K.R., G.B., R.D., J.B., G.C., C.W., R.S., S.C., H.L., and E.Y. performed collection, curation, and annotation of included datasets. C.W., R.S., S.O., and K.K. contributed to data conversion and the NWB extension format. B.D. and R.L. contributed to the NWB extension format. G.B. developed visualization approaches. D.S., E.Y. and S.K. designed the analysis, interpreted data, and wrote the manuscript. All authors reviewed the manuscript.

Acknowledgements

We thank Erdem Varol and Amin Nejatbakhsh for providing code and troubleshooting support to train the statistical atlas, and to label images of dense neurons used in figure panels that show atlases of neuron positions. We thank Adam Atanas and Steve Flavell for contributing their datasets. Research funding was provided by National Institute of General Medical Sciences of the National Institutes of Health 582 (#R35GM124735, SK), the Weill Institute for Neurosciences, and the Weill Neurohub. Development of eats-worm is supported by a napari Plugin Foundation grant from the Chan Zuckerberg Initiative. EY was funded by the Esther A. & Joseph Klingenstein Fund, the Simons Foundation, and the Hypothesis Fund.

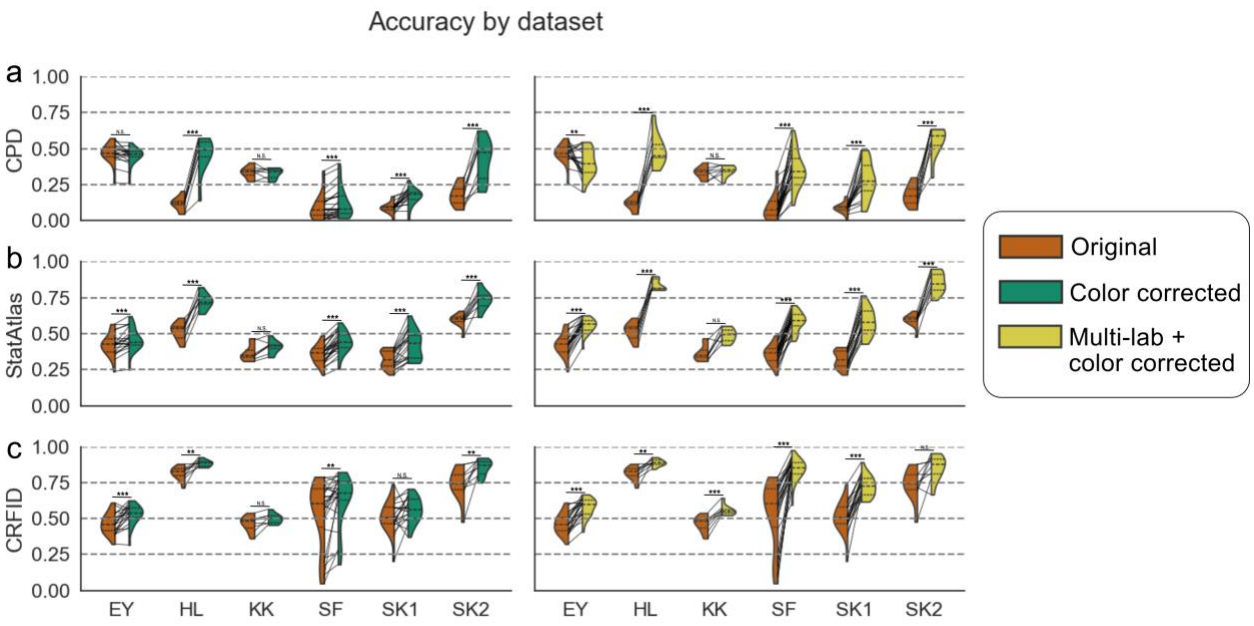
Supplemental Information



Supplement 1: KDE plot of mean nuclear distance and synaptic weight between pairs of neurons. Contour encompasses 95% of the cumulative mass of the data. Color shows density with higher density colored darker.

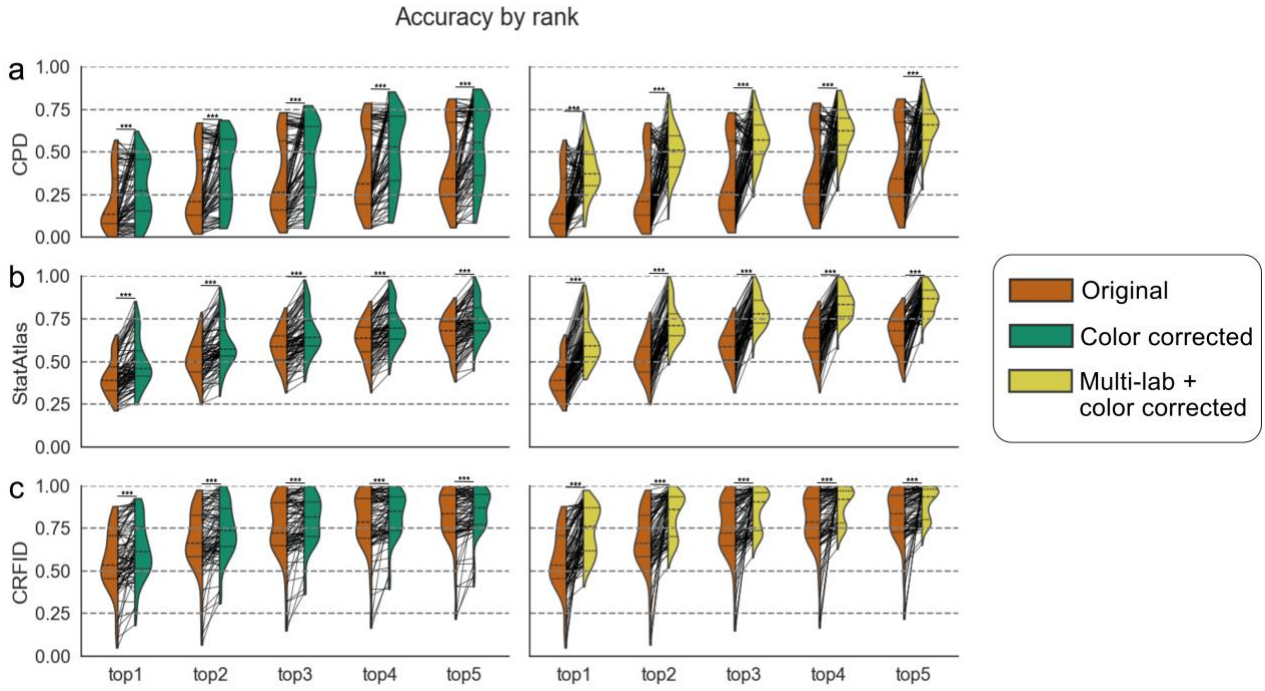
N=94	Top1 mean effect size [95% CI)	Top1 p-value	Top5 mean effect size [95% CI]	Top5 p-value
CPD: Original vs color corrected	7.66% [5.07%, 10.25%]	6.298e-08	12.63% [9.41%, 15.85%]	8.959e-12
CPD: Original vs multi-lab + color corrected	17.62% [13.60%, 21.63%]	9.881e-14	23.44% [18.71%, 28.17%]	4.119e-16
CPD: Color corrected vs. multi-lab + color corrected	9.96% [6.69%, 13.23%]	2.958e-08	10.81% [7.20%, 14.42%]	4.398e-08
StatAtlas: Original vs color corrected	9.05% [7.60%, 10.50%]	2.217e-21	8.30% [7.12%, 9.47%]	1.117e-24
StatAtlas: Original vs multi-lab + color corrected	21.57% [19.71%, 23.42%]	2.373e-40	19.76% [17.51%, 22.00%]	3.449e-31
StatAtlas: Color corrected vs. multi-lab + color corrected	12.52% [11.18%, 13.86%]	4.698e-33	11.46% [9.68%, 13.24%	3.172e-22
CRF_ID: Original vs color corrected	7.30% [5.09%, 9.51%]	2.829e-09	4.11% [2.19%, 6.02%]	4.948e-05
CRF_ID: Original vs multi-lab + color corrected	18.39% [15.16%, 21.62%]	3.222e-19	8.70% [5.53%, 11.86%]	3.839e-07
CRF_ID: Color corrected vs multi-lab + color corrected	11.09% [8.44%, 13.75%]	7.682e-13	4.59% [2.52%, 6.67%]	2.910e-05
Multi-lab + color corrected: CPD vs StatAtlas	23.61% [20.82%, 26.40%]	6.064e-30	20.51% [18.17%, 22.85%]	4.519e-31
Multi-lab + color corrected: CPD vs CRF_ID	35.24% [31.80%, 38.67%]	4.167e-36	24.17% [21.26%, 27.07%]	1.825e-29
Multi-lab + color corrected: StatAtlas vs CRF_ID	11.63% [9.09%, 14.16%]	1.492e-14	3.66% [2.52%, 4.79%]	5.908e-09

Supplementary table 1: Summary statistics for figure 5.



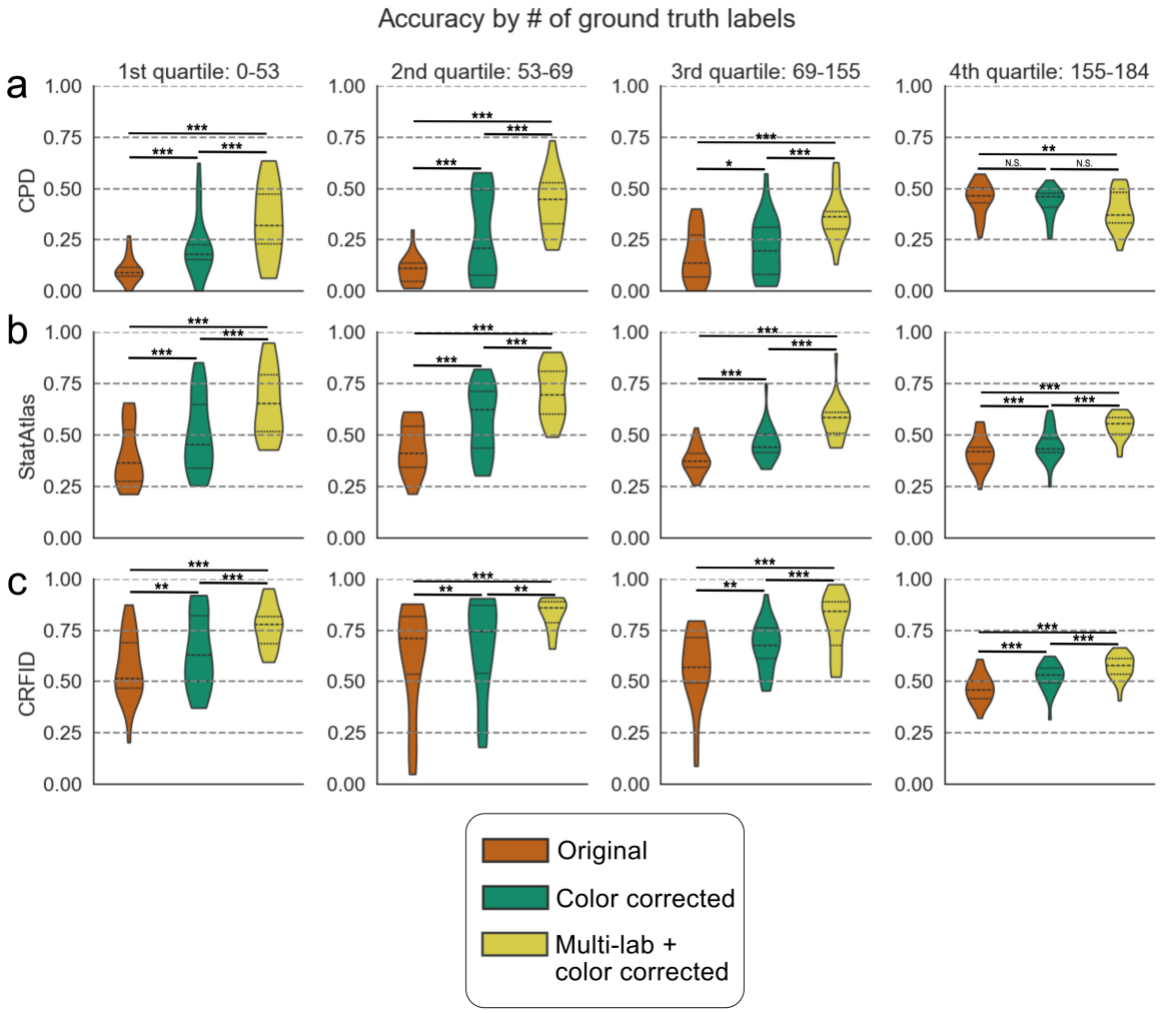
Supplement 2: Top accuracy by dataset
(a, b, c) Top ranked test accuracy for training set of original 10 reference worms with no color correction (orange), 10 reference worms with color correction (green), and the multi-lab corpus with color correction (yellow) broken out by model. Algorithmic performance was evaluated using paired t-tests (where N=94 for each training set) to compare the performance of different atlases. Significance is reported using a Bonferroni correction with the convention of * for p<0.05, ** for p<0.01, and *** for p<0.001. **(a)** Accuracy by dataset for CPD. EY dataset: original (10 worms) no color correction versus with color correction - mean effect size = -2.00%, 95% CI effect size = [-3.60%, -0.39%], p = 2.491e-02; original with no color correction versus multi-lab with color correction - mean effect size = -6.59%, 95% CI effect size = [-10.85%, -2.34%], p = 6.695e-03. HL dataset: original (10 worms) no color correction versus with color correction - mean effect size = 33.75%, 95% CI effect size = [27.57%, 39.93%], p = 6.815e-06; original with no color correction versus multi-lab with color correction - mean effect size = 37.66%, 95% CI effect size = [29.39%, 45.94%], p = 2.624e-05. KK dataset: original (10 worms) no color correction versus with color correction - mean effect size = -0.67%, 95% CI effect size = [-3.01%, 1.67%], p = 6.075e-01; original with no color correction versus multi-lab with color correction - mean effect size = 1.42%, 95% CI effect size = [-1.58%, 4.43%], p = 4.041e-01. SF dataset: original (10 worms) no color correction versus with color correction - mean effect size = 2.96%, 95% CI effect size = [1.83%, 4.09%], p = 2.671e-05; original with no color correction versus multi-lab with color correction - mean effect size = 26.09%, 95% CI effect size = [21.17%, 31.02%], p = 1.469e-10. SK1 dataset: original (10 worms) no color correction versus with color correction - mean effect size = 7.39%, 95% CI effect size = [5.08%, 9.70%], p = 5.516e-06; original with no color correction versus multi-lab with color correction - mean effect size = 19.55%, 95% CI effect size = [14.16%, 24.95%], p = 1.010e-06. SK2 dataset: original (10 worms) no color correction versus with color correction - mean effect size = 23.88%, 95% CI effect size = [17.33%, 30.44%], p = 6.968e-05; original with no color correction versus multi-lab with color correction - mean effect size = 37.46%, 95% CI effect size = [30.69%, 44.24%], p = 2.406e-06. **(b)** Accuracy by dataset for StatAtlas. EY dataset: original (10 worms) no color correction versus with color correction - mean effect size = 3.43%, 95% CI effect size = [1.76%, 5.10%], p = 6.888e-04; original with no color correction versus multi-lab with color correction - mean effect size = 12.85%, 95% CI effect size = [10.79%, 14.91%], p = 1.080e-10. HL dataset: original (10 worms) no color correction versus with color correction - mean effect size = 20.67%, 95% CI effect size = [16.96%, 24.38%], p = 5.871e-06; original with no color correction versus multi-lab with color correction - mean effect size = 31.95%, 95% CI effect size = [27.70%, 36.20%], p = 5.955e-07. KK dataset: original (10 worms) no color correction versus with color correction - mean effect size = 4.01%, 95% CI effect size = [1.70%, 6.31%], p = 1.403e-02; original with no color correction versus multi-lab with color correction - mean effect size = 12.22%, 95% CI effect size = [7.57%, 16.87%], p = 1.719e-03. SF dataset: original (10 worms) no color correction versus with color correction - mean effect size = 8.56%, 95% CI effect size = [7.01%, 10.11%], p = 6.380e-11; original with no color correction versus multi-lab with color correction - mean effect size = 22.85%, 95% CI effect size = [20.21%, 25.50%], p = 3.298e-15. SK1 dataset: original (10 worms) no color correction versus with color correction - mean effect size = 9.78%, 95% CI effect size = [6.80%, 12.75%], p = 3.856e-06; original with no color correction versus multi-lab with color correction - mean effect size = 25.88%, 95% CI effect size = [22.85%, 28.90%], p = 8.466e-13. SK2 dataset: original (10 worms) no color correction versus with color correction - mean effect size = 14.23%, 95% CI effect size = [10.56%, 17.91%], p = 4.361e-05; original with no color correction versus multi-lab with color correction - mean effect size = 26.06%, 95% CI effect size = [21.88%, 30.25%], p = 8.712e-07. **(c)** Accuracy by dataset for CRF_ID. EY dataset: original (10 worms) no color correction versus with color correction - mean effect size = 6.04%,95% CI effect size = [3.22%,

8.86%], p = 4.601e-04; original with no color correction versus multi-lab with color correction - mean effect size = 10.97%,95% CI effect size = [8.32%, 13.62%], p = 9.944e-08. HL dataset:
original (10 worms) no color correction versus with color correction - mean effect size = 6.94%,95% CI effect size = [3.56%, 10.32%], p = 4.708e-03; original with no color correction versus multi-lab with color correction - mean effect size = 7.28%,95% CI effect size = [4.05%, 10.50%], p = 2.776e-03. KK dataset: original (10 worms) no color correction versus with color correction - mean effect size = 3.61%,95% CI effect size = [1.13%, 6.10%], p = 2.958e-02; original with no color correction versus multi-lab with color correction - mean effect size = 9.42%,95% CI effect size = [6.72%, 12.12%], p = 3.264e-04. SF dataset: original (10 worms) no color correction versus with color correction - mean effect size = 9.70%,95% CI effect size = [3.94%, 15.46%], p = 2.897e-03; original with no color correction versus multi-lab with color correction - mean effect size = 31.28%,95% CI effect size = [23.27%, 39.28%], p = 5.119e-08. SK1 dataset: original (10 worms) no color correction versus with color correction - mean effect size = 5.15%,95% CI effect size = [0.17%, 10.13%], p = 5.838e-02; original with no color correction versus multi-lab with color correction - mean effect size = 21.75%,95% CI effect size = [17.54%, 25.97%], p = 4.768e-09. SK2 dataset: original (10 worms) no color correction versus with color correction - mean effect size = 11.27%,95% CI effect size = [5.15%, 17.38%], p = 6.775e-03; original with no color correction versus multi-lab with color correction - mean effect size = 10.91%,95% CI effect size = [3.73%, 18.08%], p = 1.803e-02.



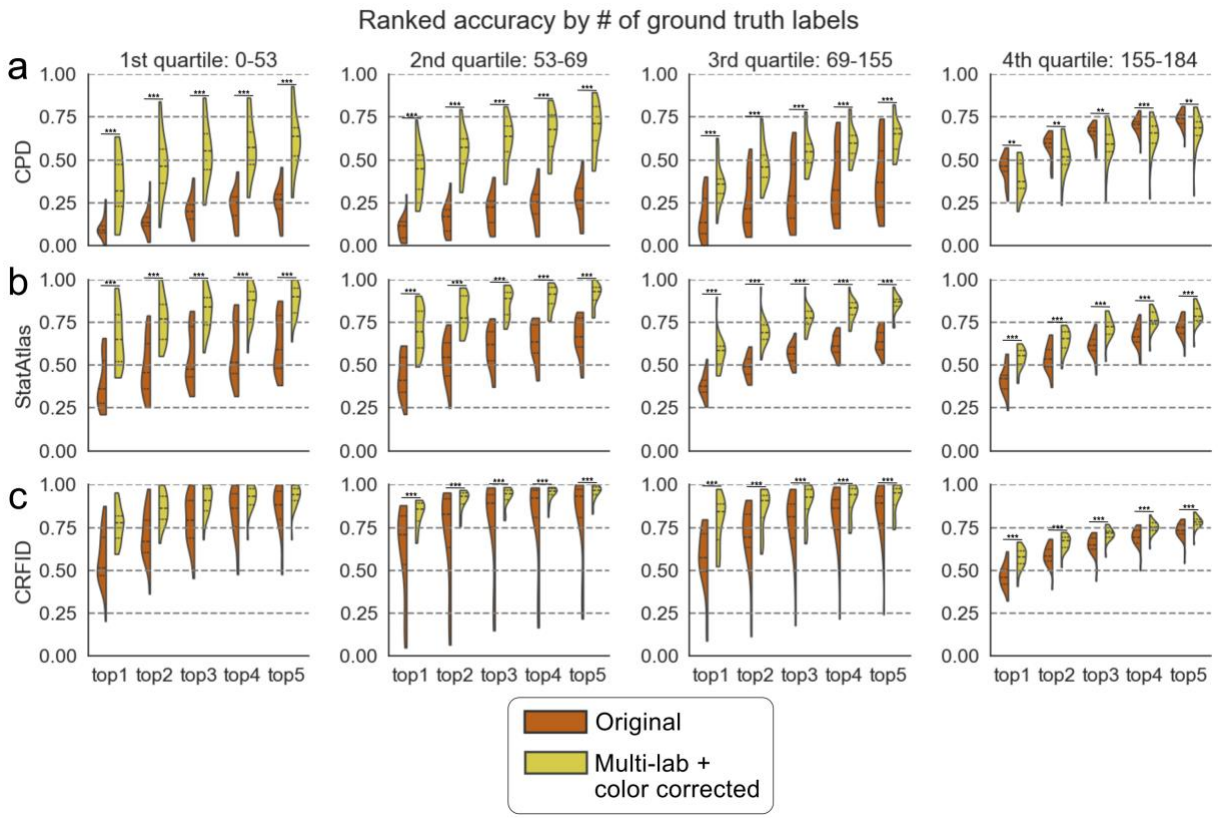
Supplement 3: Top n ranked assignment accuracy broken out by training group and model. Algorithmic performance was evaluated using paired t-tests (where N=94 for each training set) to compare the performance of different atlases. Significance is reported using a Bonferroni correction with the convention of * for p<0.05, ** for p<0.01, and *** for p<0.001. **(a)** Accuracy by rank for CPD. Original (10 worms) no color correction versus with color correction: top1 - mean effect size = 7.66%, 95% CI effect size = [5.07%, 10.25%], p = 6.298e-08; top2 - mean effect size = 10.28%, 95% CI effect size = [7.42%, 13.14%], p = 2.018e-10; top3 - mean effect size = 11.52%, 95% CI effect size = [8.55%, 14.49%], p = 1.252e-11; top4 - mean effect size = 12.34%, 95% CI effect size = [9.16%, 15.52%], p = 1.326e-11; top5 - mean effect size = 12.63%, 95% CI effect size = [9.41%, 15.85%], p = 8.959e-12**.** Original with no color correction versus multi-lab with color correction: top1 - mean effect size = 17.62%, 95% CI effect size = [13.60%, 21.63%], p = 9.881e-14; top2 - mean effect size = 21.80%, 95% CI effect size = [17.31%, 26.28%], p = 1.047e-15; top3 - mean effect size = 22.64%, 95% CI effect size = [18.03%, 27.25%], p = 6.539e-16; top4 - mean effect size = 23.21%, 95% CI effect size = [18.50%, 27.91%], p = 5.260e-16; top5 - mean effect size = 23.44%, 95% CI effect size = [18.71%, 28.17%], p = 4.119e-16 **. (b)** Accuracy by rank for StatAtlas. Original (10 worms) no color correction versus with color correction: top1 - mean effect size = 9.05%, 95% CI effect size = [7.60%, 10.50%], p = 2.217e-21; top2 - mean effect size = 9.01%, 95% CI effect size = [7.62%, 10.40%], p = 1.964e-22; top3 - mean effect size = 8.66%, 95% CI effect size = [7.44%, 9.89%], p = 8.658e-25; top4 - mean effect size = 8.83%, 95% CI effect size = [7.60%, 10.06%], p = 3.591e-25; top5 - mean effect size = 8.30%, 95% CI effect size = [7.12%, 9.47%], p = 1.117e-24. Original with no color correction versus multi-lab with color correction: top1 - mean effect size = 21.57%, 95% CI effect size = [19.71%, 23.42%], p = 2.373e-40; top2 - mean effect size = 21.37%, 95% CI effect size = [19.31%, 23.42%], p = 1.468e-36; top3 - mean effect size = 20.91%, 95% CI effect size = [18.98%, 22.85%], p = 6.920e-38; top4 - mean effect size = 20.84%, 95% CI effect size = [18.71%, 22.96%], p = 1.310e-34; top5 - mean effect size = 19.76%, 95% CI effect size = [17.51%, 22.00%], p = 3.449e-31. **(c)** Accuracy by rank for CRF ID. Original (10

worms) no color correction versus with color correction: top1 - mean effect size = 7.30%, 95% CI effect size = [5.09%, 9.51%], p = 2.829e-09; top2 - mean effect size = 6.45%, 95% CI effect size = [4.30%, 8.60%], p = 4.148e-08; top3 - mean effect size = 4.98%, 95% CI effect size = [3.00%, 6.96%], p = 2.710e-06; top4 - mean effect size = 4.38%, 95% CI effect size = [2.42%, 6.34%], p = 2.473e-05; top5 - mean effect size = 4.11%, 95% CI effect size = [2.19%, 6.02%], p = 4.948e-05. Original with no color correction versus multi-lab with color correction: top1 - mean effect size = 18.39%, 95% CI effect size = [15.16%, 21.62%], p = 3.222e-19; top2 - mean effect size = 14.97%, 95% CI effect size = [11.47%, 18.47%], p = 2.851e-13; top3 - mean effect size = 11.55%, 95% CI effect size = [8.16%, 14.94%], p = 1.101e-09; top4 - mean effect size = 9.89%, 95% CI effect size = [6.58%, 13.19%], p = 4.730e-08; top5 - mean effect size = 8.70%, 95% CI effect size = [5.53%, 11.86%], p = 3.839e-07.



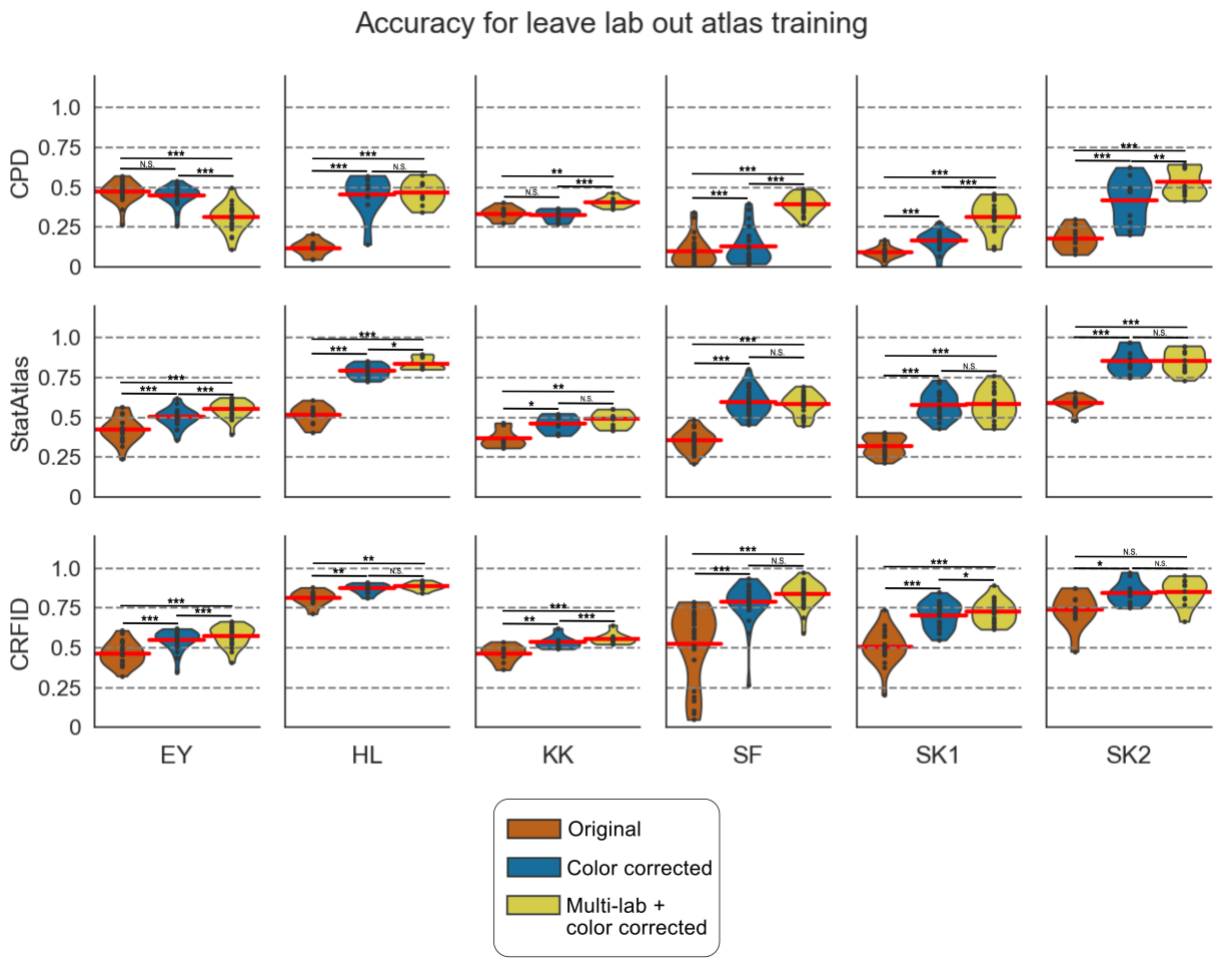
Supplement 4: Top ranked neural identification accuracy for datasets with varying numbers of ground truth labels. All statistics are conducted with a paired t-test. First quartile N=25, second quartile N=26, third quartile N=27, fourth quartile N=16. Quartiles are determined by the number of ground truth labels in each dataset. Significance is reported using a Bonferroni correction with the convention of * for p<0.05, ** for p<0.01, and *** for p<0.001. **(a)** Accuracy by quartile for CPD. 1st quartile: original (10 worms) no color correction versus with color correction - mean effect size = 11.11%, 95% CI effect size = [7.25%, 14.96%], p = 1.162e-05; original with no color correction versus multi-lab with color correction - mean effect size = 24.40%, 95% CI effect size = [18.20%, 30.61%], p = 1.124e-07; original with color correction versus multi-lab with color correction - mean effect size = 13.30%, 95% CI effect size = [7.85%, 18.74%], p = 9.123e-05. 2nd quartile: original (10 worms) no color correction versus with color correction - mean effect size = 17.21%, 95% CI effect size = [10.39%, 24.03%], p = 6.144e-05; original with no color correction versus multi-lab with color correction - mean effect size = 32.96%, 95% CI effect size = [27.42%, 38.49%], p = 7.032e-11; original with color correction versus multi-lab with color correction - mean effect size = 15.74%, 95% CI effect size = [8.48%, 23.00%], p = 3.335e-04. 3rd quartile: original (10 worms) no color correction versus with color correction - mean effect size = 4.58%, 95% CI effect size = [1.20%, 7.96%], p = 1.414e-02; original with no color correction versus multi-lab with color correction - mean effect size = 19.99%, 95% CI effect size = [14.47%, 25.51%], p = 3.175e-07; original with color correction

versus multi-lab with color correction - mean effect size = 15.40%, 95% CI effect size = [9.73%, 21.08%], p = 2.166e-05. 4th quartile: original (10 worms) no color correction versus with color correction - mean effect size = -1.72%, 95% CI effect size = [-3.26%, -0.19%], p = 3.829e-02; original with no color correction versus multi-lab with color correction - mean effect size = -5.96%, 95% CI effect size = [-9.80%, -2.12%], p = 5.834e-03; original with color correction versus multi-lab with color correction - mean effect size = -4.23%, 95% CI effect size = [-7.51%, -0.96%], p = 1.864e-02. **(b)** Accuracy by quartile for StatAtlas. 1st quartile: original (10 worms) no color correction versus with color correction - mean effect size = 11.01%, 95% CI effect size = [8.35%, 13.68%], p = 5.059e-08; original with no color correction versus multi-lab with color correction - mean effect size = 26.44%, 95% CI effect size = [23.56%, 29.32%], p = 1.265e-14; original with color correction versus multi-lab with color correction - mean effect size = 15.43%, 95% CI effect size = [12.38%, 18.47%], p = 1.443e-09. 2nd quartile: original (10 worms) no color correction versus with color correction - mean effect size = 13.54%, 95% CI effect size = [10.27%, 16.81%], p = 4.780e-08; original with no color correction versus multi-lab with color correction - mean effect size = 27.71%, 95% CI effect size = [25.22%, 30.20%], p = 2.244e-16; original with color correction versus multi-lab with color correction - mean effect size = 14.16%, 95% CI effect size = [11.53%, 16.80%], p = 4.772e-10. 3rd quartile: original (10 worms) no color correction versus with color correction - mean effect size = 8.22%, 95% CI effect size = [5.98%, 10.47%], p = 2.624e-07; original with no color correction versus multi-lab with color correction - mean effect size = 19.75%, 95% CI effect size = [16.14%, 23.36%], p = 2.053e-10; original with color correction versus multi-lab with color correction - mean effect size = 11.52%, 95% CI effect size = [9.18%, 13.87%], p = 1.536e-09. 4th quartile: original (10 worms) no color correction versus with color correction - mean effect size = 3.68%, 95% CI effect size = [2.18%, 5.19%], p = 7.996e-05; original with no color correction versus multi-lab with color correction - mean effect size = 12.84%, 95% CI effect size = [11.00%, 14.68%], p = 1.557e-12; original with color correction versus multi-lab with color correction - mean effect size = 9.16%, 95% CI effect size = [7.32%, 10.99%], p = 1.189e-09. **(c)** Accuracy by quartile for CRF_ID. 1st quartile: original (10 worms) no color correction versus with color correction - mean effect size = 7.22%, 95% CI effect size = [2.25%, 12.18%], p = 9.428e-03; original with no color correction versus multi-lab with color correction - mean effect size = 19.97%, 95% CI effect size = [15.62%, 24.32%], p = 8.294e-09; original with color correction versus multi-lab with color correction - mean effect size = 12.75%, 95% CI effect size = [8.00%, 17.51%], p = 2.936e-05. 2nd quartile: original (10 worms) no color correction versus with color correction - mean effect size = 5.53%, 95% CI effect size = [2.51%, 8.55%], p = 1.646e-03; original with no color correction versus multi-lab with color correction - mean effect size = 21.42%, 95% CI effect size = [11.84%, 31.00%], p = 2.412e-04; original with color correction versus multi-lab with color correction - mean effect size = 15.89%, 95% CI effect size = [7.27%, 24.50%], p = 1.563e-03. 3rd quartile: original (10 worms) no color correction versus with color correction - mean effect size = 10.91%, 95% CI effect size = [5.08%, 16.75%], p = 1.304e-03; original with no color correction versus multi-lab with color correction - mean effect size = 21.73%, 95% CI effect size = [15.50%, 27.96%], p = 5.751e-07; original with color correction versus multi-lab with color correction - mean effect size = 10.82%, 95% CI effect size = [8.39%, 13.25%], p = 9.590e-09. 4th quartile: original (10 worms) no color correction versus with color correction - mean effect size = 5.45%, 95% CI effect size = [2.90%, 8.00%], p = 3.574e-04; original with no color correction versus multi-lab with color correction - mean effect size = 10.63%, 95% CI effect size = [8.26%, 13.00%], p = 8.450e-09; original with color correction versus multi-lab with color correction - mean effect size = 5.18%, 95% CI effect size = [3.34%, 7.02%], p = 1.358e-05.



Supplement 5: Top n ranked neural identification accuracy for datasets with varying numbers of ground truth labels. All statistics are conducted with a paired t-test. First quartile N=25, second quartile N=26, third quartile N=27, fourth quartile N=16. Quartiles are determined by the number of ground truth labels in each dataset. Significance is reported using a Bonferroni correction with the convention of * for p<0.05, ** for p<0.01, and *** for p<0.001. **(a)** Accuracy by rank and quartile for CPD. 1st quartile: top1 - mean effect size = 24.40%, 95% CI effect size = [18.20%, 30.61%], p = 1.124e-07; top2 - mean effect size = 33.12%, 95% CI effect size = [27.39%, 38.86%], p = 1.250e-10; top3 - mean effect size = 34.51%, 95% CI effect size = [29.51%, 39.51%], p = 3.980e-12; top4 - mean effect size = 34.70%, 95% CI effect size = [29.53%, 39.87%], p = 6.885e-12; top5 - mean effect size = 35.63%, 95% CI effect size = [30.26%, 41.01%], p = 8.878e-12. 2nd quartile: top1 - mean effect size = 32.96%, 95% CI effect size = [27.42%, 38.49%], p = 7.032e-11; top2 - mean effect size = 38.79%, 95% CI effect size = [33.73%, 43.84%], p = 4.837e-13; top3 - mean effect size = 40.45%, 95% CI effect size = [35.58%, 45.33%], p = 1.003e-13; top4 - mean effect size = 42.43%, 95% CI effect size = [37.62%, 47.24%], p = 2.827e-14; top5 - mean effect size = 42.46%, 95% CI effect size = [37.67%, 47.25%], p = 2.595e-14. 3rd quartile: top1 - mean effect size = 19.99%, 95% CI effect size = [14.47%, 25.51%], p = 3.175e-07; top2 - mean effect size = 22.78%, 95% CI effect size = [16.36%, 29.20%], p = 4.413e-07; top3 - mean effect size = 23.65%, 95% CI effect size = [16.71%, 30.58%], p = 8.273e-07; top4 - mean effect size = 23.67%, 95% CI effect size = [16.39%, 30.96%], p = 1.708e-06; top5 - mean effect size = 23.43%, 95% CI effect size = [16.17%, 30.70%], p = 1.918e-06. 4th quartile: top1 - mean effect size = -5.96%, 95% CI effect size = [-9.80%, -2.12%], p = 5.834e-03; top2 - mean effect size = -6.32%, 95% CI effect size = [-10.24%, -2.41%], p = 4.364e-03; top3 - mean effect size = -6.82%, 95% CI effect size = [-10.77%, -2.86%], p = 2.630e-03; top4 - mean effect size = -6.70%, 95% CI effect size = [-10.16%, -3.23%], p = 9.588e-04; top5 - mean effect size = -6.46%, 95% CI effect size = [-9.99%, -2.92%], p = 1.601e-03. **(b)** Accuracy by rank and quartile for StatAtlas. 1st quartile: top1 - mean effect size = 26.44%, 95% CI effect size = [23.56%, 29.32%], p = 1.265e-14; top2 - mean effect size = 27.31%, 95% CI effect size = [24.42%, 30.20%], p = 6.743e-15; top3 - mean effect size = 26.80%, 95% CI effect size = [23.62%, 29.98%], p = 7.078e-14; top4 - mean effect size = 27.89%, 95% CI effect size = [23.77%, 32.02%], p = 6.028e-12; top5 - mean effect size = 26.03%, 95% CI effect size = [21.95%, 30.11%], p = 1.840e-11. 2nd quartile: top1 - mean effect size = 27.71%, 95% CI effect size = [25.22%, 30.20%], p = 2.244e-16; top2 - mean effect size = 27.48%, 95% CI effect size = [24.62%, 30.35%], p = 5.028e-15; top3 - mean effect size = 26.09%, 95% CI effect size = [23.60%, 28.58%], p = 7.857e-16; top4 - mean effect size = 25.60%, 95% CI effect size = [23.16%, 28.04%], p = 7.552e-16; top5 - mean effect size = 24.67%, 95% CI effect size = [21.88%, 27.46%], p = 2.711e-14. 3rd quartile: top1 - mean effect size = 19.75%, 95% CI effect size = [16.14%, 23.36%], p = 2.053e-10; top2 - mean effect size = 20.62%, 95% CI effect size = [16.74%, 24.50%], p = 3.608e-10; top3 - mean effect size = 21.18%, 95% CI effect size = [17.93%, 24.43%], p = 6.459e-12; top4 - mean effect size = 20.90%, 95% CI effect size = [17.37%, 24.44%], p = 4.479e-11; top5 - mean effect size = 21.27%, 95% CI effect size = [17.20%, 25.34%], p = 4.985e-10. 4th quartile:top1 - mean effect size = 12.84%, 95% CI effect size = [11.00%,

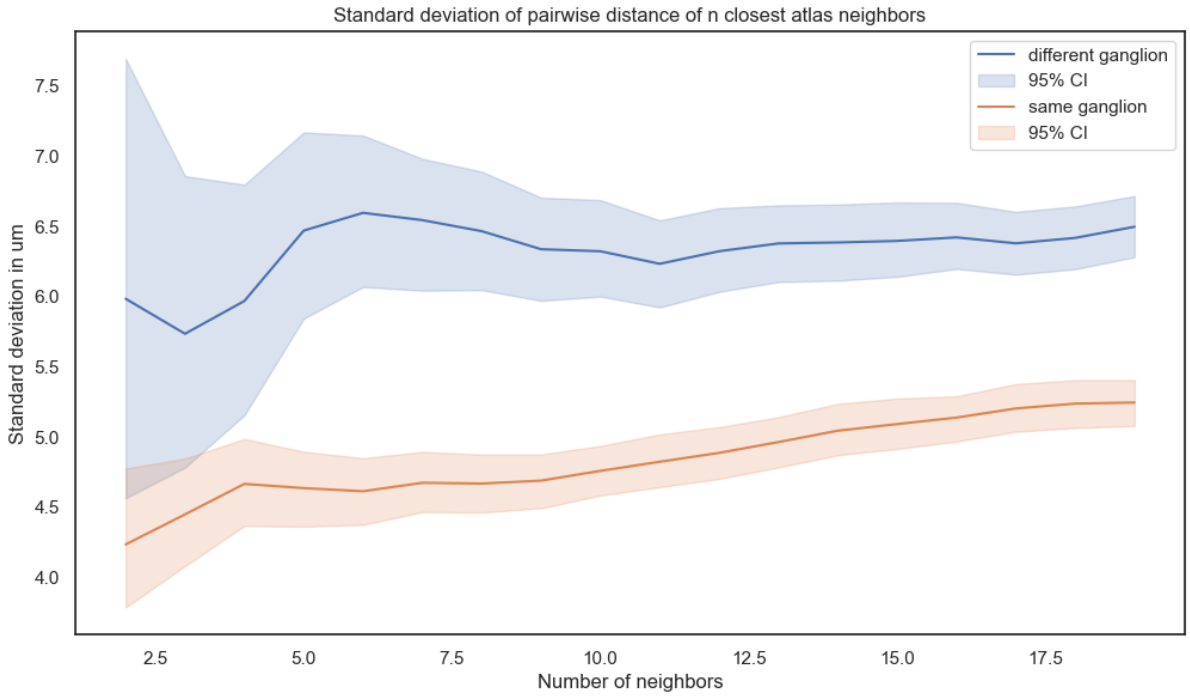
14.68%], p = 1.557e-12; top2 - mean effect size = 10.57%, 95% CI effect size = [8.72%, 12.42%], p = 8.887e-11; top3 - mean effect size = 10.05%, 95% CI effect size = [8.53%, 11.57%], p = 4.861e-12; top4 - mean effect size = 9.44%, 95% CI effect size = [7.84%, 11.03%], p = 4.475e-11; top5 - mean effect size = 7.53%, 95% CI effect size = [5.83%, 9.23%], p = 1.037e-08. **(c)** Accuracy by rank and quartile for CRF ID. 1st quartile: top1 - mean effect size = 19.97%, 95% CI effect size = [15.62%, 24.32%], p = 8.294e-09; top2 - mean effect size = 16.07%, 95% CI effect size = [12.47%, 19.67%], p = 1.329e-08; top3 - mean effect size = 11.56%, 95% CI effect size = [7.71%, 15.40%], p = 6.574e-06; top4 - mean effect size = 8.63%, 95% CI effect size = [4.70%, 12.55%], p = 2.912e-04; top5 - mean effect size = 7.24%, 95% CI effect size = [3.38%, 11.09%], p = 1.342e-03. 2nd quartile: top1 - mean effect size = 21.42%, 95% CI effect size = [11.84%, 31.00%], p = 2.412e-04; top2 - mean effect size = 20.18%, 95% CI effect size = [9.41%, 30.94%], p = 1.360e-03; top3 - mean effect size = 15.61%, 95% CI effect size = [4.96%, 26.26%], p = 8.945e-03; top4 - mean effect size = 14.52%, 95% CI effect size = [4.12%, 24.92%], p = 1.220e-02; top5 - mean effect size = 13.62%, 95% CI effect size = [3.71%, 23.53%], p = 1.343e-02. 3rd quartile: top1 - mean effect size = 21.73%, 95% CI effect size = [15.50%, 27.96%], p = 5.751e-07; top2 - mean effect size = 16.31%, 95% CI effect size = [9.15%, 23.47%], p = 1.789e-04; top3 - mean effect size = 12.68%, 95% CI effect size = [5.67%, 19.68%], p = 1.737e-03; top4 - mean effect size = 10.92%, 95% CI effect size = [4.19%, 17.64%], p = 4.178e-03; top5 - mean effect size = 9.21%, 95% CI effect size = [2.76%, 15.66%], p = 1.027e-02. 4th quartile: top1 - mean effect size = 10.63%, 95% CI effect size = [8.26%, 13.00%], p = 8.450e-09; top2 - mean effect size = 7.58%, 95% CI effect size = [5.54%, 9.62%], p = 2.031e-07; top3 - mean effect size = 6.54%, 95% CI effect size = [4.95%, 8.12%], p = 3.564e-08; top4 - mean effect size = 5.63%, 95% CI effect size = [4.15%, 7.10%], p = 1.420e-07; top5 - mean effect size = 4.87%, 95% CI effect size = [3.62%, 6.12%], p = 9.202e-08.



Supplement 6: Top ranked assignment accuracy for leave-one-lab-out training. For each dataset, we train using the full corpus of data excluding all data from the dataset being tested. We then report the test accuracy for the dataset that was left out. All statistics are conducted with a paired t-test. Significance is reported using a Bonferroni correction with the convention of * for p<0.05, ** for p<0.01, and *** for p<0.001. **(a)** Accuracy by dataset for CPD. EY: original (10 worms) no color correction versus with leave-lab-out with color correction - mean effect size = -2.00%, 95% CI effect size = [-3.60%, -0.39%], p = 2.491e-02; original with no color correction versus multi-lab with color correction - mean effect size = -15.97%, 95% CI effect size = [-19.26%, -12.68%], p =

7.764e-09; leave-lab-out with color correction versus multi-lab with color correction - mean effect size = -13.97%, 95% CI effect size = [-17.10%, -10.84%], p = 3.041e-08. HL: original (10 worms) no color correction versus with leave-lab-out with color correction - mean effect size = 33.75%,95% CI effect size = [27.57%, 39.93%], p = 6.815e-06; original with no color correction versus multi-lab with color correction - mean effect size = 34.68%,95% CI effect size = [31.18%, 38.18%], p = 6.960e-08; leave-lab-out with color correction versus multi-lab with color correction - mean effect size = 0.93%,95% CI effect size = [-6.29%, 8.15%], p = 8.137e-01. KK:
original (10 worms) no color correction versus with leave-lab-out with color correction - mean effect size = -0.67%,95% CI effect size = [-3.01%, 1.67%], p = 6.075e-01; original with no color correction versus multi-lab with color correction - mean effect size = 7.28%,95% CI effect size = [3.80%, 10.75%], p = 5.809e-03; leave-lab-out with color correction versus multi-lab with color correction - mean effect size = 7.95%,95% CI effect size = [5.26%, 10.63%], p = 8.722e-04. SF:
original (10 worms) no color correction versus with leave-lab-out with color correction - mean effect size = 2.96%,95% CI effect size = [1.83%, 4.09%], p = 2.671e-05; original with no color correction versus multi-lab with color correction - mean effect size = 29.87%,95% CI effect size = [27.07%, 32.67%], p = 2.330e-17; leave-lab-out with color correction versus multi-lab with color correction - mean effect size = 26.91%,95% CI effect size = [23.62%, 30.20%], p = 1.165e-14. SK1:
original (10 worms) no color correction versus with leave-lab-out with color correction - mean effect size = 7.39%,95% CI effect size = [5.08%, 9.70%], p = 5.516e-06; original with no color correction versus multi-lab with color correction - mean effect size = 22.22%,95% CI effect size = [17.35%, 27.08%], p = 3.335e-08; leave-lab-out with color correction versus multi-lab with color correction - mean effect size = 14.83%,95% CI effect size = [9.00%, 20.65%], p = 8.723e-05. SK2:
original (10 worms) no color correction versus with leave-lab-out with color correction - mean effect size = 23.88%,95% CI effect size = [17.33%, 30.44%], p = 6.968e-05; original with no color correction versus multi-lab with color correction - mean effect size = 35.29%,95% CI effect size = [31.00%, 39.58%], p = 7.940e-08; leave-lab-out with color correction versus multi-lab with color correction - mean effect size = 11.41%,95% CI effect size = [5.67%, 17.15%], p = 4.398e-03. **(b)** Accuracy by dataset for StatAtlas. EY: original (10 worms) no color correction versus with leave-lab-out with color correction - mean effect size = 7.63%,95% CI effect size = [5.54%, 9.72%], p = 6.863e-07; original with no color correction versus multi-lab with color correction - mean effect size = 12.85%,95% CI effect size = [10.79%, 14.91%], p = 1.080e-10; leave-lab-out with color correction versus multi-lab with color correction - mean effect size = 5.22%,95% CI effect size = [4.03%, 6.41%], p = 4.261e-08. HL: original (10 worms) no color correction versus with leave-lab-out with color correction - mean effect size = 27.40%,95% CI effect size = [23.34%, 31.46%], p = 1.361e-06; original with no color correction versus multi-lab with color correction - mean effect size = 31.95%,95% CI effect size = [27.70%, 36.20%], p = 5.955e-07; leave-lab-out with color correction versus multi-lab with color correction - mean effect size = 4.55%,95% CI effect size = [1.93%, 7.17%], p = 1.123e-02. KK: original (10 worms) no color correction versus with leave-lab-out with color correction - mean effect size = 9.27%,95% CI effect size = [4.06%, 14.49%], p = 1.268e-02; original with no color correction versus multi-lab with color correction - mean effect size = 12.22%,95% CI effect size = [7.57%, 16.87%], p = 1.719e-03; leave-lab-out with color correction versus multi-lab with color correction - mean effect size = 2.94%,95% CI effect size = [0.74%, 5.15%], p = 4.095e-02. SF: original (10 worms) no color correction versus with leave-lab-out with color correction - mean effect size = 24.01%,95% CI effect size = [20.07%, 27.95%], p = 7.939e-12; original with no color correction versus multi-lab with color correction - mean effect size = 22.85%,95% CI effect size = [20.21%, 25.50%], p = 3.298e-15; leave-lab-out with color correction versus multi-lab with color correction - mean effect size = -1.16%,95% CI effect size = [-4.18%, 1.87%], p = 4.601e-01. SK1: original (10 worms) no color correction versus with leave-lab-out with color correction - mean effect size = 25.56%,95% CI effect size = [23.05%, 28.08%], p = 3.666e-14; original with no color correction versus multi-lab with color correction - mean effect size = 25.88%,95% CI effect size = [22.85%, 28.90%], p = 8.466e-13; leave-lab-out with color correction versus multi-lab with color correction - mean effect size = 0.31%,95% CI effect size = [-1.52%, 2.15%], p = 7.414e-01. SK2: original (10 worms) no color correction versus with leave-lab-out with color correction - mean effect size = 26.04%,95% CI effect size = [21.35%, 30.72%], p = 2.302e-06; original with no color correction versus multi-lab with color correction - mean effect size = 26.06%,95% CI effect size = [21.88%, 30.25%], p = 8.712e-07; leave-lab-out with color correction versus multi-lab with color correction - mean effect size = 0.03%,95% CI effect size = [-1.57%, 1.63%], p = 9.738e-01. **(c)** Accuracy by dataset for CRFID. EY: original (10 worms) no color correction versus with leave-lab-out with color correction - mean effect size = 8.51%,95% CI effect size = [5.94%, 11.08%], p = 2.641e-06; original with no color correction versus multi-lab with color correction - mean effect size = 10.97%,95% CI effect size = [8.32%, 13.62%], p = 9.944e-08; leave-lab-out with color correction versus multi-lab with color correction - mean effect size = 2.46%,95% CI effect size = [1.33%, 3.60%], p = 4.055e-04. HL: original (10 worms) no color correction versus with leave-lab-out with color correction - mean effect size = 5.86%,95% CI effect size = [2.81%, 8.91%], p = 6.772e-03; original with no color correction versus multi-lab with color correction - mean effect size = 7.28%,95% CI effect size = [4.05%, 10.50%], p = 2.776e-03; leave-lab-out with color correction versus multi-lab with color correction - mean effect size = 1.42%,95% CI effect size = [0.26%, 2.58%], p = 4.977e-02. KK: original (10 worms) no color correction versus with leave-lab-out with color correction - mean effect size = 7.07%,95% CI effect size = [4.27%, 9.87%], p = 2.133e-03; original with no color correction versus multi-lab with color correction - mean effect size = 9.42%,95% CI effect size = [6.72%, 12.12%], p = 3.264e-04; leave-lab-out with color correction versus multi-lab with color correction - mean effect size = 2.35%,95% CI effect size = [1.76%, 2.93%], p = 1.356e-04. SF: original (10 worms) no color correction versus with leave-lab-out with color correction - mean effect size = 26.05%,95% CI effect size = [18.80%, 33.29%], p = 2.164e-07; original with no color correction versus multi-lab with color correction - mean effect size = 31.28%,95% CI effect size = [23.27%, 39.28%], p = 5.119e-08; leave-lab-out with color correction versus multi-lab with color correction - mean effect size = 5.23%,95% CI effect size = [1.22%, 9.24%], p = 1.705e-02. SK1: original (10 worms) no color correction versus with leave-lab-out with color correction - mean effect size = 19.66%,95% CI effect size = [15.19%, 24.13%], p = 5.897e-08; original with no color correction versus multi-

lab with color correction - mean effect size = 21.75%,95% CI effect size = [17.54%, 25.97%], p = 4.768e-09; leave-lab-out with color correction versus multi-lab with color correction - mean effect size = 2.09%,95% CI effect size = [0.66%, 3.52%], p = 1.015e-02. SK2: original (10 worms) no color correction versus with leave-lab-out with color correction - mean effect size = 10.61%,95% CI effect size = [3.88%, 17.34%], p = 1.521e-02; original with no color correction versus multi-lab with color correction - mean effect size = 10.91%,95% CI effect size = [3.73%, 18.08%], p = 1.803e-02; leave-lab-out with color correction versus multi-lab with color correction - mean effect size = 0.30%,95% CI effect size = [-3.51%, 4.11%], p = 8.847e-01.



Supplement 7: Standard deviation of pairwise distance of *n* closest atlas neighbors.

References

1. Schrödel, T., Prevedel, R., Aumayr, K. *et al.* Brain-wide 3D imaging of neuronal activity in *Caenorhabditis elegans* with sculpted light. *Nat Methods* 10, 1013–1020 (2013). https://doi.org/10.1038/nmeth.2637
2. Ahrens, M., Orger, M., Robson, D. *et al.* Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nat Methods* 10, 413–420 (2013). https://doi.org/10.1038/nmeth.2434
3. Kato, S., *et al.* Global Brain Dynamics Embed the Motor Command Sequence of *Caenorhabditis elegans. Cell* 163(3), 656-669 (2015). https://doi.org/10.1016/j.cell.2015.09.034
4. Venkatachalam, V., *et al.* Pan-neuronal imaging in roaming Caenorhabditis elegans. *PNAS* 113(8), E1082-E1088 (2015). https://doi.org/10.1073/pnas.1507109113
5. Nguyen, J.P., *et al.* Whole-brain calcium imaging with cellular resolution in freely behaving Caeonorhabditis elegans. *PNAS* 113(8), 1074-1081. (2015) https://doi.org/10.1073/pnas.1507110112.
6. Yemini, E., et al. NeuroPAL: A multicolor atlas for whole-brain neuronal identification in C. elegans. *Cell* 184(1), 272-288 (2021). https://doi.org/10.1016/j.cell.2020.12.012
7. Atanas, A., et al. Brain-wide representations of behavior spanning multiple timescales and states in C. elegans. *Cell* 186(19), 4134-4151 (2023). https://doi.org/10.1016/j.cell.2023.07.035
8. Portugues, R., Feierstein, C., Engert, F., Orger, M.B., Whole-brain activity maps reveal stereotyped distributed networks for visuomotor behavior. 81(6), P1328-1343 (2014). https://doi.org/10.1016/j.neuron.2014.01.019
9. Prevedel, R., Yoon, YG., Hoffmann, M. *et al.* Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy. *Nat Methods* 11, 727–730 (2014). https://doi.org/10.1038/nmeth.2964
10. Royer, L., Lemon, W., Chhetri, R. *et al.* Adaptive light-sheet microscopy for long-term, high-resolution imaging in living organisms. *Nat Biotechnol* 34, 1267–1278 (2016). https://doi.org/10.1038/nbt.3708
11. Kim, D., Kim, J., Marques, J. *et al.* Pan-neuronal calcium imaging with cellular resolution in freely swimming zebrafish. *Nat Methods* 14, 1107–1114 (2017). https://doi.org/10.1038/nmeth.4429
12. Vladimirov, N., Wang, C., Höckendorf, B. *et al.* Brain-wide circuit interrogation at the cellular level guided by online analysis of neuronal function. *Nat Methods* 15, 1117–1125 (2018). https://doi.org/10.1038/s41592-018-0221-x
13. Chen, X., *et al.* Brain-wide organization of neuronal activity and convergent sensorimotor transformations in larval zebrafish. *Neuron* 100(4), 876-890 (2018). https://doi.org/10.1016/j.neuron.2018.09.042
14. Lemon, W., Pulver, S., Höckendorf, B. *et al.* Whole-central nervous system functional imaging in larval *Drosophila*. *Nat Commun* 6, 7924 (2015). https://doi.org/10.1038/ncomms8924
15. Mann, K., Gallen, C.L., Clandinin, T.R., Whole-brain calcium imaging reveals an intrinsic functional network in *Drosophila*. *Current Biology* 27(15), 2389-2396 (2017). https://doi.org/10.1016/j.cub.2017.06.076
16. Aimon, S., *et al.* Fast near-whole-brain imaging in adult Drosophila during responses to stimuli and behavior. *PLoS Biology* 17(2), (2019). https://doi.org/10.1371/journal.pbio.2006732
17. Stirman, J., Smith, I., Kudenov, M. *et al.* Wide field-of-view, multi-region, two-photon imaging of neuronal activity in the mammalian brain. *Nat Biotechnol* 34, 857–862 (2016). https://doi.org/10.1038/nbt.3594
18. Skocek, O., Nöbauer, T., Weilguny, L. *et al.* High-speed volumetric imaging of neuronal activity in freely moving rodents. *Nat Methods* 15, 429–432 (2018). https://doi.org/10.1038/s41592-018-0008-0
19. Klioutchnikov, A., Wallace, D.J., Frosz, M.H. *et al.* Three-photon head-mounted microscope for imaging deep cortical layers in freely moving rats. *Nat Methods* 17, 509–513 (2020). https://doi.org/10.1038/s41592-020-0817-9
20. Zong, W., *et al.* Large-scale two-photon calcium imaging in freely moving mice. *Cell* 186(7), (2022). https://doi.org/10.1016/j.cell.2022.02.017

21. Manley, J., *et al.* Simultaneous, cortex-wide dynamics of up to 1 million neurons reveal unbounded scaling of dimensionality with neuron number. *Neuron*, (2024). https://doi.org/10.1016/j.neuron.2024.02.011.
22. Uzel, K., Kato, S., Zimmer, M., A set of hub neurons and non-local connectivity features support global brain dynamics in *C. elegans*. *Cell Current Biology* 32, 3443-3459 (2022). https://doi.org/10.1016/j.cub.2022.06.039
23. Flavell, S., Gordus, A., Dynamics functional connectivity in the static connectome of *Caenorhabditis elegans*. *Current opinions in Neurobiology* 73, (2022). https://doi.org/10.1016/j.conb.2021.12.002.
24. Randi, F., Sharma, A.K., Dvali, S. *et al.* Neural signal propagation atlas of *Caenorhabditis elegans*. *Nature* 623, 406–414 (2023). https://doi.org/10.1038/s41586-023-06683-4
25. Tekieli, T., *et al.* Visualizing the organization and differentiation of the male-specific nervous system of C. elegans. *Development (Cambridge, England)* 148,18 (2021). https://doi:10.1242/dev.199687
26. Chaudhary, S., Lee, S.A., Li, Y., Patel, D.S., Ly, H., Graphical-model framework for automated annotation of cell identities in dense cellular images. *elife* 10, (2021). https://doi.org/10.7554/eLife.60321.
27. Cecere, Z.T., et al. State-dependent network interactions differentially gate sensory input at the motor and command neuron level in *Caenorhabditis elegans*. Preprint at https://www.biorxiv.org/content/10.1101/2021.04.09.439242v3
28. Varol, E., et al. Statistical Atlas of C. elegans Neurons. *Medical Image Computing and Computer Assisted Intervention - MICCAI 2020* 12265, (2020).
29. Nejatbakhsh, A., Varol, E., Yemini, E., Hobert, O., Paninski, L., Probabilistic Joint Segmentation and labeling of C. elegans neurons. *Medical Image Computing and Computer Assisted Intervention - MICCAI 2020* 12265, (2020).
30. Skuhersky, M., Wu, T., Yemini, E. *et al.* Toward a more accurate 3D atlas of *C. elegans* neurons. *BMC Bioinformatics* 23, 195 (2022). https://doi.org/10.1186/s12859-022-04738-3
31. 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
32. Bubnis, G., Ban, S., DiFranco, M.D., Kato, S., A probabilistic atlas for cell identification. Preprint at https://arxiv.org/abs/1903.09227 (2019).
33. Wu, Y., Wu, S., Wang, X., Lang, C., Zhang, Q., Wen, Q., Xu, T., Rapid detection and recognition of whole brain activity in a freely behaving Caenorhabditis elegans. *PLoS Computational Biology*, (2022)
34. Yu, X., et al. Fast deep neural correspondence for tracking and identifying neurons in C. elegans using semi-synthetic training. *elife* 10, (2021). https://doi.org/10.7554/eLife.66410
35. Cheng, C., et al. A general primer for data harmonization. *Scientific Data* 11(1), (2024). https://doi.org/10.1038/s41597-024-02956-3
36. Deng, L,. The MNIST database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine* 29(6), 141–142 (2012).
37. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., ImageNet: A large-scale hierarchical image database, *IEEE Conference on Computer Vision and Pattern Recognition* 248-255 (2009). doi: 10.1109/CVPR.2009.5206848.
38. Lhoest, Q., et al. Datasets: A Community Library for Natural Language Processing. Preprint at https://arxiv.org/abs/2109.02846 (2021).
39. Halchenko , Y., *et al.* dandi/dandi-cli: 0.61.2. *Zenodo* (2024). https://doi.org/10.5281/zenodo.3692138
40. Chen, TW., Wardill, T., Sun, Y. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity.*Nature* **499**, 295–300 (2013). https://doi.org/10.1038/nature12354
41. Chronis, N., Zimmer, M., Bargmann, C.I., Microfluidics for in vivo imaging of neuronal and behavioral activity in Caenorhabditis elegans. *Nature methods* 4(9), 727-731 (2007). https://doi.org/10.1038/nmeth1075.
42. Rübel, O., et al. The Neurodata Without Borders ecosystem for neurophysiological data science. eLife 11, (2022). https://doi.org/10.7554/eLife.78362.

43. Kuhn, H.W., The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly* 2, 83-97 (1955). https://doi.org/10.1002/nav.3800020109.
44. White, J.G., Southgate, E., Thomson, J.N., Brenner, S., The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philosophical Transactions of the Royal Society Biological Sciences* 314(1165), (1986). https://doi.org/10.1098/rstb.1986.0056
45. Cook, S.J., Jarrell, T.A., Brittin, C.A. *et al.* Whole-animal connectomes of both *Caenorhabditis elegans* sexes. *Nature* 571, 63–71 (2019). https://doi.org/10.1038/s41586-019-1352-7
46. Witvliet, D., Mulcahy, B., Mitchell, J.K. *et al.* Connectomes across development reveal principles of brain maturation. *Nature* 596, 257–261 (2021). https://doi.org/10.1038/s41586-021-03778-8
47. Mango, S.E., The *C. elegans* pharynx: a model for organogenesis. *WormBook* (2007). https://doi/10.1895/wormbook.1.129.1, http://www.wormbook.org
48. Cook, S.J., Kalinski, C.A., Hobert, O., Neuronal contact predicts connectivity in the C. elegans brain. *Current Biology* 33(11), 2315-2320 (2023). https://doi.org/10.1016/j.cub.2023.04.071.
49. Riddle, D.L., et al. C. elegans II. 2nd edition. *Cold Spring Harbor Laboratory press* (1997).
50. Byerly, L., Cassada, R.C., Ruseell, R.L., The life cycle of the nematode *Caenorhabditis elegans*: I. Wild-type growth and reproduction. *Developmental Biology* 51(1), 22-33 (1976). https://doi.org/10.1016/0012-1606(76)90119-6.
51. Stojanovski, K., Großhans, H. & Towbin, B.D. Coupling of growth rate and developmental tempo reduces body size heterogeneity in *C. elegans*. *Nat Commun* **13**, 3132 (2022). https://doi.org/10.1038/s41467-022-29720-8
52. Myronenko, A., Song, X., Point-Set Registration: Coherent Point Drift. *Arxiv* (2009). https://doi.org/10.48550/arXiv.0905.2635
53. Yemini, E., NeuroPAL annotations manual. https://www.hobertlab.org/wp-content/uploads/2019/06/NeuroPAL-Reference-Manual-v1_small.pdf
54. Neuromatch. (2024). at Neuromatch.io
55. Yemini, E., Venkatachalam, V., Lin, A., Varol, E., Nejatbakhsh, A., Sprague, D., Samuel, A., Paninski, L., Hobert, O., NeuroPAL: Atlas of *C. elegans* neuron locations and colors in NeuroPAL worm (Version 0.240614.1942) [Data set]. *DANDI* archive (2023). https://doi.org/10.48324/dandi.000715/0.240614.1942
56. Yemini, E., Venkatachalam, V., Lin, A., Varol, E., Nejatbakhsh,h A., Sprague, D., Samuels, A., Paninski, L., Hobert, O., NeuroPAL microfluidic chip images and GCaMP activity (Version 0.240625.0452) [Data set]. *DANDI archive* (2023). https://doi.org/10.48324/dandi.000541/0.240625.0452
57. Chaudhary, S., Sprague, D., Lee, S.A., Li, Y., Patel, D.S., Lu, H., Segmented and labeled NeuroPAL structural images (Version 0.240611.1954) [Data set]. *DANDI archive* (2023). https://doi.org/10.48324/dandi.000714/0.240611.1954
58. Suzuki, R., Wen, C., Sprague, D., Onami, S., Kimura, K.D., Whole-brain spontaneous GCaMP activity with NeuroPAL cell ID information of semi-restricted worms (Version 0.240402.2118) [Data set]. *DANDI archive* (2023). https://doi.org/10.48324/dandi.000692/0.240402.2118
59. Atanas, A., et al., Brain-wide representations of behavior spanning multiple timescales and states in C. elegans (Version 0.240625.0022) [Data set]. *Dandi archive* (20230. https://doi.org/10.48324/dandi.000776/0.240625.0022
60. Dunn, R., Sprague, D., Kato, S., C. elegans whole-brain neuroPAL and immobilized calcium imaging (Version 0.240625.0439) [Data set]. *DANDI archive* (2023). https://doi.org/10.48324/dandi.000565/0.240625.0439
61. Sprague, D., Borchardt, J., Dunn, R., Bubnis, G., Kato, S., NeuroPAL volumetric images (Version 0.240625.0454) [Data set]. *DANDI archive* (2023). https://doi.org/10.48324/dandi.000472/0.240625.0454
62. Emmons, S.W., Yemini, E., Zimmer, M., Methods for analyzing neuronal structure and activity in *Caenorhabditis elegans*. *Genetics* 218(4), (2021). https://doi.org/10.1093/genetics/iyab072

63. Wen, C., et al. 3DeeCellTracker, a deep learning-based pipeline for segmenting and tracking cells in 3D time lapse images. *eLife* 10, (2021). https://doi.org/10.7554/eLife.59187.
64. Gonzales, R.C., Fittes, B.A., Gray-level transformations for interactive image enhancement. *Second conference on remotely manned systems*, 17-19 (1975).
65. Sulston, J.E., Schierenberg, E., White, J.G., Thomson, J.N., The embryonic cell lineage of the nematode Caenorhabditis elegans. *Developmental Biology* 100(1), 64-119 (1983). https://doi.org/10.1016/0012-1606(83)90201-4.