

# Statistical Atlas of *C. elegans* Neurons

Erdem Varol<sup>2,3,4</sup>, Amin Nejatbakhsh<sup>1,3,4</sup>, Ruoxi Sun<sup>1,3,4</sup>, Gonzalo Mena<sup>7</sup>,  
Eviatar Yemini<sup>5</sup>, Vivek Venkatachalam<sup>6</sup>, Oliver Hobert<sup>5</sup>, and Liam  
Paninski<sup>1,2,3,4</sup>

<sup>1</sup> Department of Neuroscience, Columbia University

<sup>2</sup> Department of Statistics, Columbia University

<sup>3</sup> Grossman Center for the Statistics of Mind

<sup>4</sup> Zuckerman Institute, Center for Theoretical Neuroscience

<sup>5</sup> Department of Biological Sciences, Columbia University, New York, NY 10025

<sup>6</sup> Department of Physics, Northeastern University, Boston, MA 02115

<sup>7</sup> Department of Statistics and Data Science Initiative, Harvard University, Boston,  
MA 02138

**Abstract.** Constructing a statistical atlas of neuron positions in the nematode *Caenorhabditis elegans* enables a wide range of applications that require neural identity. These applications include annotating gene expression, extracting calcium activity, and evaluating nervous-system mutations. Large complete sets of neural annotations are necessary to determine canonical neuron positions and their associated confidence regions. Recently, a strain of *C. elegans* ("NeuroPAL") has been introduced to assign correct identities to all neurons in the worm via a deterministic, fluorescent colormap. This strain has enabled efficient and accurate annotation of worm neurons. Using a dataset of 10 worms, we propose a statistical model that captures the latent means and covariances of neuron locations, with efficient optimization strategies to infer model parameters. We demonstrate the utility of this model in two critical applications. First, we use our trained atlas to automatically annotate neuron identities in *C. elegans* at the state-of-the-art rate. Second, we use our atlas to compute canonical correlations between neuron positions, thereby determining covariance in neuron placement. The code to replicate the statistical atlas is distributed publicly at <https://github.com/amin-nejat/StatAtlas>.

## 1 Introduction

Imaging-based atlases of human and animal brains have enabled the principled and standardized means of hypothesis testing in a wide variety of domains [3,20,7,10,13,12,15]. Common procedures that atlases enable are the registration of population samples to a common space [23], discriminating pattern differences across samples [2], segmentation into regions of interest [4], and regularizing complex Bayesian models [17]. Importantly, atlases enable the formation of large-scale population studies due to their ability to gather high-dimensional data into a commensurate space.

*C. elegans* is a widely studied model organism with a simple nervous system that consists of 302 neurons in the adult hermaphrodite [21]. Its simplicity and

stereotypy have enabled highly-reproducible experimental settings which have been crucial in elucidating neuroscientific hypotheses. Furthermore, to date, *C. elegans* is the only animal whose connectome is completely mapped [21,9,6]. Despite this atlas of connectivity, there have been no large-scale attempts at quantifying the variability of the neuron positions. This is due to the limited number of samples available from electron micrograph reconstructions and an inability to identify neuronal identities via position alone [22]. The recent introduction of NeuroPAL, a strain for complete neural identification in *C. elegans*, has enabled efficient and precise annotation of neuron positions in multiple worms.

Using the NeuroPAL dataset of all head and tail neurons from 10 worms, we propose a latent multivariate statistical model that captures the canonical positions and covariances of *C. elegans* neurons. The observed neurons were captured by fluorescent volumetric imaging. These were then modeled as a multivariate sample, drawn from a latent distribution subjected to a random affine transformation. Given this statistical model, we infer the canonical means and covariances of all neurons present in the head and tail of the worm, yielding a novel positional statistical atlas. To improve our statistical atlas with additional, incompletely annotated worms, we propose a semi-supervised approach for cell-identification. Using our trained atlas, we demonstrate the ability to automatically identify neurons in out-of-sample worms with more than 85% accuracy in the head and 94% accuracy in the tail. These accuracies represent the current state of the art improving the accuracies reported in [11] and [20]. Furthermore, we demonstrate an additional application of our atlas to obtain a canonical correlation analysis of neuronal positions, which sheds light on the structural organization of neurons and their potential connections to genetic lineages.

## 2 Data and pre-processing

To construct the statistical atlas of *C. elegans* neurons, we used volumetric images of both heads and tails from 10 worms (strain OH15262). All worms were imaged on a Zeiss LSM 880 confocal with 32 detector channels and the following laser lines: 405nm, 488nm, 561nm, and 633nm. Volumetric resolution was approximately (X,Y,Z):  $0.2 \mu\text{m} \times 0.2 \mu\text{m} \times 0.8 \mu\text{m}$ . Images were acquired with four color channels, corresponding to the NeuroPAL fluorophores: mTagBFP2, CyOFP1, mNeptune2.5, and TagRFP-T [22]. See figure 1 for a representative maximum intensity projection from a head sample. The volumetric images were subsequently annotated by an expert to denote the approximate center for each neuron and its corresponding identity. In total, 240 neurons were annotated in each worm, 195 from the head and 45 from tail. The remaining neurons from the midbody were not imaged for this study.

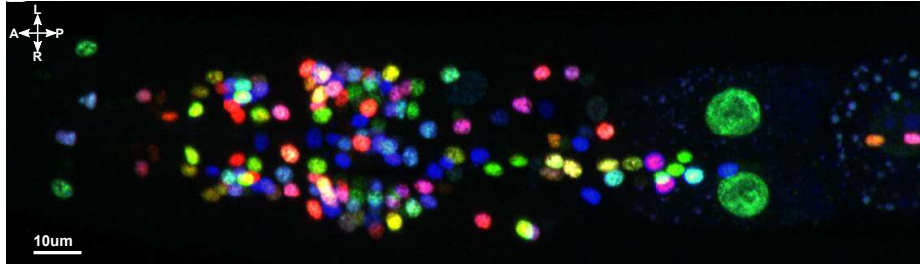


Fig. 1: Deterministic coloring of *C. elegans* neurons, in a NeuroPAL strain, enables the complete neural identification across a population of worms

### 3 Method

Due to variability in illumination and the pose of the worm when imaged, observed neuron positions and their exact color balance may vary across imaged worms. This presents a significant challenge when attempting to obtain correspondence between worms to infer the identities of neurons. Therefore, to normalize the random variability that occurs across different worms, prior to identifying neurons in any given microscopy image, we estimate a statistical atlas of neuron positions and colors.

The approach we take resembles the joint expectation-maximization alignment of point sets technique of [8], with several important differences discussed below. The dataset we are modeling consists of a collection of point sets: each worm corresponds to one point set, with each point in the set corresponding to the position and color of a single detected neuron. We model each of these positions and colors as samples from a statistical atlas that is common across worms. Each neuron  $i$  has a corresponding mean and covariance in this atlas, denoted as  $\mu_i$  and  $\Sigma_i$ , respectively. After drawing all the positions and colors for a given worm  $j$  we apply a random affine transformation (parametrized by a matrix  $\beta_j$  and translation vector  $\beta_j^0$ ). Finally, since the order of neurons in each point set is arbitrary, we scramble the identities of the neurons with a random permutation, parameterized by a permutation matrix  $P_j$ . This generative model is summarized in Figure 2. See also [3] for a related model (without the alignment term, and with an inference approach that differs from the methods we describe below).

We build on the methods in [8] to infer the parameters of this generative model (i.e., the means and covariances of the statistical atlas, the random transformations, and the random permutations), in a completely unsupervised fashion, using a three-way expectation-maximization procedure. However, in our dataset, we have access to fully annotated neuron detections. We take advantage of this supervised data to simplify the inference problem.

Now we can describe our model in detail. Neuron positions are three-dimensional, and there are three color channels in this dataset (given our three neuron-specific fluorophore channels, we discard the pan-neuronal TagRFP-T channel as unin-

formative); therefore, if we use  $\mathbf{w}_{i,j}$  to denote the appended position and color vector of the  $i$ -th neuron in worm  $j$  (as output by the detection step described in the previous section), then  $\mathbf{w}_{i,j} \in \mathbb{R}^6$ . Each of these observed  $\mathbf{w}_{i,j}$  vectors has a corresponding latent vector  $\mathbf{z}_{i,j}$  in the aligned atlas space. We model this latent vector as a Gaussian,

$$\mathbf{z}_{i,j} \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \quad (1)$$

with means  $\boldsymbol{\mu}_i \in \mathbb{R}^6$  and covariances  $\boldsymbol{\Sigma}_i \in \mathbb{S}_+^6$  that do not depend on the worm index  $j$ . We model the covariance  $\boldsymbol{\Sigma}_i$  with block structure of the form  $\boldsymbol{\Sigma}_i = \begin{bmatrix} \boldsymbol{\Sigma}_{\text{position}}^i & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\text{color}}^i \end{bmatrix}$ , since position and each color are independently varying.

Now the latent vectors  $\mathbf{z}_{i,j}$  in the atlas space and observed data  $\mathbf{w}_{i,j}$  extracted from the imaged worm  $j$  are connected by a worm-specific random affine transformation and permutation. We denote the intermediate affine-transformed variables as  $\mathbf{x}_{i,j}$ :

$$\mathbf{x}_{i,j} = \mathbf{z}_{i,j} \boldsymbol{\beta}_j + \boldsymbol{\beta}_j^0, \quad (2)$$

with  $\boldsymbol{\beta}_j$  a  $6 \times 6$  matrix (with a similar block structure as  $\boldsymbol{\Sigma}_i$ ) and  $\boldsymbol{\beta}_j^0 \in \mathbb{R}^6$ . then we obtain  $\mathbf{w}_{i,j}$  by scrambling the labels via the permutation  $p_j$  (corresponding to a permutation matrix  $\mathbf{P}_j$ ):

$$\mathbf{w}_{i,j} = \mathbf{x}_{p_j(i),j}. \quad (3)$$

The summary of the generative process is illustrated in figure 2 by combining the permutation operation and the transformation together as a latent function  $f(\cdot)$ . Note that this model permits partial and variable observations of neurons across different animals if we let the permutation matrix to be unbalanced (not square) indicating the existence of neurons that are not observed in individual animals.

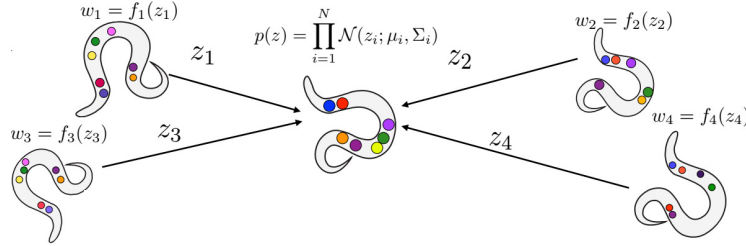


Fig. 2: Schematic of the generative model of neuron position and color expression. First we draw a position and color for each neuron  $i$  from a distribution with mean  $\boldsymbol{\mu}_i$  and covariance  $\boldsymbol{\Sigma}_i$ ; then, to create the observed data  $\mathbf{w}_{i,j}$  (the color and position of the  $i$ -th neuron of the  $j$ -th worm) we apply a random affine transformation and a random permutation encoded by  $f(\cdot)$

Given these modeling assumptions, for a dataset of  $m$  worms and  $n_j$  detected neurons in each worm, we can express the likelihood as:

$$P(\mathbf{w}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{P}, \boldsymbol{\beta}, \boldsymbol{\beta}^0) = \prod_{j=1}^m \prod_{i=1}^{n_j} \frac{e^{-(1/2)(\mathbf{w}_{i,j} - \boldsymbol{\mu}_{\mathbf{P}_{i,j}}\boldsymbol{\beta}_j - \boldsymbol{\beta}_j^0)(\boldsymbol{\beta}_j\boldsymbol{\Sigma}_{\mathbf{P}_{i,j}}\boldsymbol{\beta}_j^T)^{-1}(\mathbf{w}_{i,j} - \boldsymbol{\mu}_{\mathbf{P}_{i,j}}\boldsymbol{\beta}_j - \boldsymbol{\beta}_j^0)^T}}{(2\pi)^{d/2} \det((\boldsymbol{\beta}_j\boldsymbol{\Sigma}_{\mathbf{P}_{i,j}}\boldsymbol{\beta}_j^T))^{1/2}} \quad (4)$$

Since the term  $\sum_j \sum_i (1/2) \log \det((\boldsymbol{\beta}_j\boldsymbol{\Sigma}_{\mathbf{P}_{i,j}}\boldsymbol{\beta}_j^T))$  is permutation invariant, we can write it as  $\sum_j \sum_i (1/2) \log \det((\boldsymbol{\beta}_j\boldsymbol{\Sigma}_i\boldsymbol{\beta}_j^T))$  and thus the maximum likelihood estimate (MLE) for our generative model involves optimizing the negative log-likelihood:

$$\underset{\mathbf{P}, \boldsymbol{\beta}, \boldsymbol{\beta}^0, \boldsymbol{\mu}, \boldsymbol{\Sigma}}{\text{minimize}} \sum_{j=1}^m \sum_{i=1}^{n_j} (\mathbf{w}_{i,j} - \boldsymbol{\mu}_{\mathbf{P}_{i,j}}\boldsymbol{\beta}_j - \boldsymbol{\beta}_j^0)(\boldsymbol{\beta}_j\boldsymbol{\Sigma}_{\mathbf{P}_{i,j}}\boldsymbol{\beta}_j^T)^{-1}(\mathbf{w}_{i,j} - \boldsymbol{\mu}_{\mathbf{P}_{i,j}}\boldsymbol{\beta}_j - \boldsymbol{\beta}_j^0)^T + \log \det((\boldsymbol{\beta}_j\boldsymbol{\Sigma}_i\boldsymbol{\beta}_j^T)) \quad (5)$$

### 3.1 Optimization

To infer the parameters of the generative model, we take an iterative block-coordinate descent approach, similar to [8]: we fix  $(\mathbf{P}, \boldsymbol{\beta}, \boldsymbol{\beta}_0)$  (with  $\mathbf{P}$  abbreviating the collection of permutations  $\mathbf{P}_j$  for all worms  $j$ , and similarly for  $\boldsymbol{\beta}, \boldsymbol{\beta}_0$ ) and solve for  $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , then fix  $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  and solve for  $(\mathbf{P}, \boldsymbol{\beta}, \boldsymbol{\beta}_0)$ . Below are the update steps for each of these blocks.

**Inference of the statistical atlas parameters  $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ :** Let  $\mathbf{P}_j \in \mathcal{P}^{n \times n}$  denote the permutation matrix,  $\mathbf{W}_j = [\mathbf{w}_{1,j}^T \dots \mathbf{w}_{n,j}^T]^T \in \mathbb{R}^{n \times d}$  denote the row stacked features of the neurons of the  $j$ th worm, and let  $\boldsymbol{\mu} = [\boldsymbol{\mu}_1^T \dots \boldsymbol{\mu}_n^T] \in \mathbb{R}^{n \times d}$  denote the row stacked neuron means. The generative model can be written in matrix form as:  $\mathbf{W}_j = \mathbf{P}_j\boldsymbol{\mu}\boldsymbol{\beta}_j + \mathbf{1}\boldsymbol{\beta}_j^0 + \mathbf{E}$  where  $\mathbf{E}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\beta}_j\boldsymbol{\Sigma}_{\mathbf{P}_{i,j}}\boldsymbol{\beta}_j^T)$  denotes the row stacked uncertainty terms.

Since  $\mathbf{P}_j^T \mathbf{P}_j = \mathbf{I}$  because  $\mathbf{P}$  is a permutation matrix and assuming that  $\boldsymbol{\beta}_j$  is a non-degenerate transformation, its inverse exists and can be used to write the system as:  $\mathbf{P}_j^T \mathbf{W}_j \boldsymbol{\beta}_j^{-1} - \mathbf{1}\boldsymbol{\beta}_j^0 \boldsymbol{\beta}_j^{-1} = \boldsymbol{\mu} + \mathbf{V}$  where  $\mathbf{V}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_i)$  is a term to quantify uncertainty.

This equation can be used to infer  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  in closed form by computing the first and second moments of  $\mathbf{V}$ :

$$\boldsymbol{\mu}^* = \frac{1}{m} \sum_{j=1}^m \mathbf{P}_j^T \mathbf{W}_j \boldsymbol{\beta}_j^{-1} - \mathbf{1}\boldsymbol{\beta}_j^0 \boldsymbol{\beta}_j^{-1} \quad (6)$$

$$\boldsymbol{\Sigma}_i^* = \frac{1}{m} \sum_{j=1}^m (\mathbf{P}_{j,i}^T \mathbf{W}_j \boldsymbol{\beta}_j^{-1} - \boldsymbol{\beta}_j^0 \boldsymbol{\beta}_j^{-1} - \boldsymbol{\mu}_i)^T (\mathbf{P}_{j,i}^T \mathbf{W}_j \boldsymbol{\beta}_j^{-1} - \boldsymbol{\beta}_j^0 \boldsymbol{\beta}_j^{-1} - \boldsymbol{\mu}_i) \quad (7)$$

**Algorithm 1** Train statistical neuron atlas

---

**Input:**  $\{\mathbf{w}_{i,j}\}$  (colors and positions) and  $\{\mathbf{P}_{i,j}\}$  (neuron correspondences) for  $j = 1, \dots, m$  worms in a training set and  $i = 1, \dots, n$  neurons,  $\epsilon$  (convergence tolerance)

**Initialization**  $k < \frac{n}{2}$  (number of outliers)

Select random worm  $j \sim \text{Unif}[m]$

Set means as neuron centers of worm  $j$ :  $\boldsymbol{\mu}_i^0 \leftarrow \mathbf{w}_{i,j}$  for  $i = 1, \dots, n$

Set covariances as identity:  $\boldsymbol{\Sigma}_i^0 \leftarrow \mathbf{I}_6$  for  $i = 1, \dots, n$

- 1: **while** Not converged **do**
- 2:      $t \leftarrow t + 1$
- 3:     **for**  $j=1, \dots, n$  **do**
- 4:         Solve alignment of  $j$ th worm to atlas  $\{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$  using equations 9 and 10
- 5:     **end for**
- 6:     Update  $\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t$  using equations 6
- 7:     Check convergence  $\|\boldsymbol{\mu}^t - \boldsymbol{\mu}^{t-1}\|_F \leq \epsilon$  and  $\|\boldsymbol{\Sigma}^t - \boldsymbol{\Sigma}^{t-1}\|_F \leq \epsilon$
- 8: **end while**
- 9: **return** Statistical atlas of neuron colors and positions  $\{\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t\}$ .

---

**Inference of the transformation terms  $\boldsymbol{\beta}, \boldsymbol{\beta}_0$ :** We can infer the transformation and translation terms  $\boldsymbol{\beta}, \boldsymbol{\beta}_0$  by solving a weighted linear regression problem with a Mahalanobis norm for each neuron quantified by the covariance terms,  $\boldsymbol{\Sigma}_i$ :

$$\underset{\boldsymbol{\beta}_j^{-1}, \boldsymbol{\beta}_j^0}{\text{minimize}} \sum_{i=1}^{n_j} (\mathbf{P}_{j,i}^T \mathbf{W}_j \boldsymbol{\beta}_j^{-1} - \boldsymbol{\beta}_j^0 \boldsymbol{\beta}_j^{-1} - \boldsymbol{\mu}_i) \boldsymbol{\Sigma}_i^{-1} (\mathbf{P}_{j,i}^T \mathbf{W}_j \boldsymbol{\beta}_j^{-1} - \boldsymbol{\beta}_j^0 \boldsymbol{\beta}_j^{-1} - \boldsymbol{\mu}_i)^T. \quad (8)$$

This system admits a fixed point iteration that yields the global minimum [8]. First, the closed form solution for  $\boldsymbol{\beta}_j^0 \boldsymbol{\beta}_j^{-1}$  is given by:

$$\boldsymbol{\beta}_j^0 \boldsymbol{\beta}_j^{-1*} = \left( \sum_{i=1}^n (\mathbf{P}_{j,i}^T \mathbf{W}_j \boldsymbol{\beta}_j^{-1} - \boldsymbol{\mu}_i) \boldsymbol{\Sigma}_i^{-1} \right) \left( \sum_{i=1}^n \boldsymbol{\Sigma}_i^{-1} \right)^{-1} \quad (9)$$

To analytically solve for  $\boldsymbol{\beta}_j^{-1}$ , we utilize the fact that  $\text{vec}(ABC) = (C^T \otimes A) \text{vec}(B)$  where  $\text{vec}(\cdot)$  denotes the vectorization operation and  $\otimes$  denotes Kronecker product. This yields the following vectorized closed form update for  $\boldsymbol{\beta}_j^{-1}$ :

$$\text{vec}(\boldsymbol{\beta}_j^{-1*}) = \left( \sum_{i=1}^n (\boldsymbol{\Sigma}_i^{-1} \otimes (\mathbf{P}_{j,i}^T \mathbf{W}_j)^T (\mathbf{P}_{j,i}^T \mathbf{W}_j)) \right)^{-1} \left( \sum_{i=1}^n \text{vec}((\mathbf{P}_{j,i}^T \mathbf{W}_j)^T (\boldsymbol{\beta}_j^0 \boldsymbol{\beta}_j^{-1} + \boldsymbol{\mu}_i) \boldsymbol{\Sigma}_i^{-1}) \right) \quad (10)$$

**Permutation inference:** Lastly, we can solve for the doubly-stochastic matrix,  $\mathbf{P}_j$  by setting up a  $n \times n$  transport matrix  $\mathbf{D}$  where

$$\mathbf{D}_{u,v} = (\boldsymbol{\mu}_u \mathbf{W} \boldsymbol{\beta}_j + \boldsymbol{\beta}_j^0 - \mathbf{W}_{j,v}) \boldsymbol{\Sigma}_u^{-1} (\boldsymbol{\mu}_u \mathbf{W} \boldsymbol{\beta}_j + \boldsymbol{\beta}_j^0 - \mathbf{W}_{j,v})^T \quad (11)$$

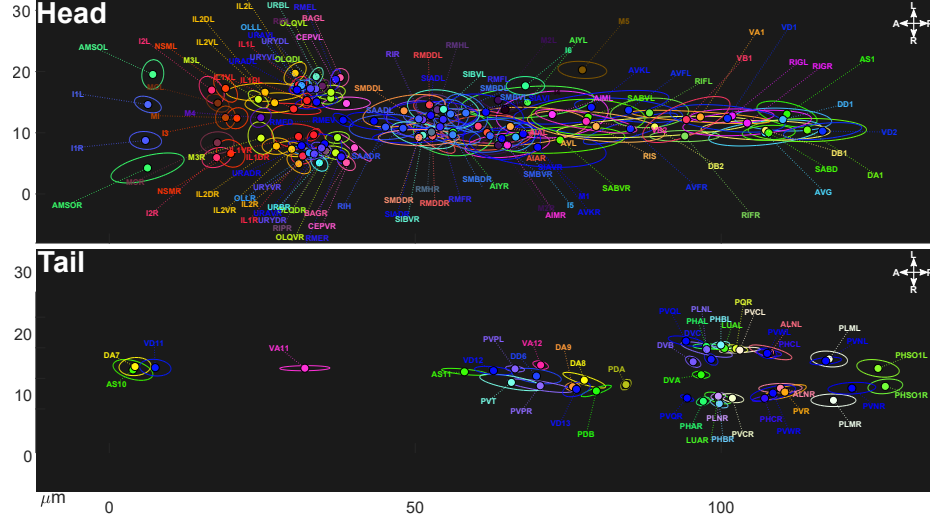


Fig. 3: A statistical atlas of *C. elegans* neuron positions and their NeuroPAL colors (including several glial cells) assembled from both the head and tail of 10 worms. Ellipses represent in-plane neuronal positional covariance at 95% probability. Saturation and brightness for the NeuroPAL colors of all cells in the atlas have been uniformly amplified for visibility. We provide a ventral view, excluding the lateral and dorsal ganglia, for visibility (see supplement for a lateral view that includes these ganglia). The figure is best viewed electronically.

and obtaining  $\mathbf{P}_j$  through the solving the entropic optimal transport problem [16] using the Sinkhorn-Knopp algorithm [18] which minimizes the following objective:

$$\mathbf{P}_j^* = \arg \min_{p \in \mathcal{P}} \sum_{u,v} p_{u,v} \mathbf{D}_{u,v} - \gamma p_{u,v} \log p_{u,v} \quad (12)$$

Further details of permutation inference for neuron identification can be found in [14].

### 3.2 Statistical atlas of neuron positions and colors

If we have access to several annotated worms, meaning that we have access to both the neural detections  $\mathbf{W}$  and their corresponding identities,  $\mathbf{P}$ , we can infer the transformation terms  $\{\beta, \beta_0\}$  as well as the parameters of the statistical atlas  $\{\mu, \Sigma\}$  using the procedure outlined in algorithm 1.

In words, algorithm 1 operates in the following way. First, the targeted inference parameters are initialized using the neuron centers and colors for a random worm. Then, the remaining worms are affinely aligned to the hypothetical atlas by solving the linear system for  $\{\beta_j, \beta_j^0\}$  in equation 10. The means and covariances of the aligned neurons are then used to update the atlas parameters of  $\mu$

and  $\Sigma$ . This procedure is iteratively repeated until convergence. The resulting statistical atlas consisting of the mean stereotypical neuron positions and colors and their covariances can be seen in Fig. 3.

## 4 Applications

**Automatic/semi-automatic neuron identification:** Automated neural detection and identification in *C. elegans* [1] [11] [19] enable neural-level hypotheses and facilitate large-scale, high-throughput analyses across worm populations. These features include neural activity traces, measuring the intensity and presence of fluorescent biomarkers under various experimental conditions, and assessment of mutant-driven neural displacements from canonical positioning. To infer neural identity, we use the trained statistical atlas’s positional means as a source point cloud, computing the Mahalanobis distance of out-of-sample worm neurons using the atlas covariances in equation (11). We then employ the iterative closest point algorithm [5] using iterative updates of equations (10), (9) and (12) to match atlas neurons to the out-of-sample neurons. This procedure yields a leave-worm-out automatic neuron identification accuracy of 88% with 85% in the head and 94% in the tail, with most misidentifications occurring in the dense ventral and retro-vesicular ganglia (Fig. 4A). This result is superior to the previously reported automatic neuron identification accuracy in [20]. Note that in this study, Toyoshima et al. use a larger set of animals (311 worms), but a partial coverage of the head neurons (on average 80% of 196 head neurons) and of these, the accuracy of the first rank identity is 62%. We further incorporate a semi-supervised approach (described in the supplementary), wherein the user can provide a number of annotations, thus improving algorithmic accuracy to over 95% for both the head and tail using 20 and 4 manual annotations, respectively (Fig. 4C-D).

**Positional canonical correlation:** Using the 10 annotated worm samples, we align all animals to the common statistical atlas space. We then perform a canonical correlation analysis between pairs of neurons using the displacement vectors of individual worms’ neurons relative to their corresponding means. Analyzing the top canonical correlation for each pair of neurons, we obtain figure 4B. The blockwise patterns displayed here demonstrate that there is a strong intra-ganglion correlation of neuron placements in both the anterior and posterior pharyngeal bulbs. Additionally, we observe anti-correlation between neurons in the anterior ganglion with respect to those in the lateral ganglion, suggesting the role of the nerve ring (a dense collection of neuropil connecting the majority of worm neurons) in separating these two ganglia. Finally, we compute the correlation between observed neuron displacements and the invariant lineage of worm neurons. We find that lineal relationships between neurons, their divisional distance from a common precursor cell, may account for 14% and 35% of their relative positional displacements, for the head and tail respectively; al-

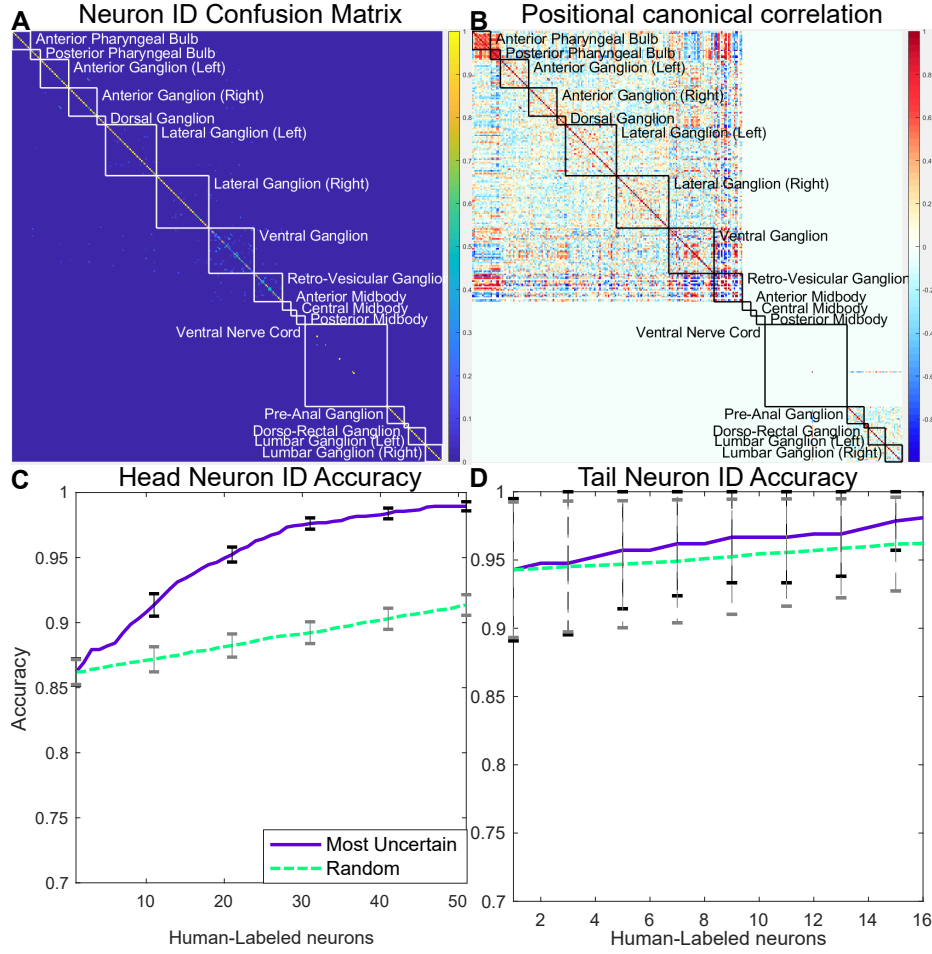


Fig. 4: **A:** Neuron identification accuracy confusion matrix, organized by ganglia, **B:** Positional-displacement canonical correlations of neurons after their alignment to the atlas space. **C,D:** Semi-supervised improvement of accuracy in the head (**C**) and tail (**D**), solid lines show the results of prioritizing annotation for the most uncertain neurons (i.e. neurons with the highest identification error rates), dashed lines show the results for random ordering.

though, further work and a larger sample of worms are necessary to rule out false correlations due to overfitting and/or bending artifacts.

**Conclusion:** The proposed statistical atlas of *C. elegans* has a wide variety of applications in addition to the ones demonstrated above. Atlases can quickly be derived for all larval developmental stages, including dauer and male worms.

These atlases could then be used to model neural repositioning throughout development.

The current atlas utilized an affine alignment model since the worms that were studied exhibited a straight posture. In future work, it is possible to generalize the alignment model to include higher-order deformations to account for postural variability.

Automatic segmentation of neurons and neural activity (or dynamic changes in biomarker intensities), extracted from time-series data, are two other prominent directions of research that will be beneficial to the computational neuroscience community. Such work could further be extended to freely-moving worms by deforming the statistical neural atlas. Lastly, analysis of color alterations as well as cell losses and/or duplications in mutant worms, combined with inferences from the invariant lineage, could be used to model lineal mutant effects consistent with the observed disruptions, thus elucidating genes involved in patterning the nervous system.

## Acknowledgments

The authors acknowledge the following funding sources. Paninski Lab: NSF NeuroNex Award DBI-1707398, The Gatsby Charitable Foundation, NIBIB R01 EB22913, DMS 1912194. Hobert Lab: Howard Hughes Medical Institute, NIH (5T32DK7328-37, 5T32DK007328-35, 5T32MH015174-38, and 5T32MH015174-37), Venkatachalam Lab: Burroughs Wellcome Fund Career Award at the Scientific Interface.

## References

1. Aerni, S.J., Liu, X., Do, C.B., Gross, S.S., Nguyen, A., Guo, S.D., Long, F., Peng, H., Kim, S.S., Batzoglu, S.: Automated cellular annotation for high-resolution images of adult *caenorhabditis elegans*. *Bioinformatics* **29**(13), i18–i26 (2013)
2. Ashburner, J., Friston, K.J.: Voxel-based morphometry—the methods. *Neuroimage* **11**(6), 805–821 (2000)
3. Bubnis, G., Ban, S., DiFranco, M.D., Kato, S.: A probabilistic atlas for cell identification. *arXiv preprint arXiv:1903.09227* (2019)
4. Cabezas, M., Oliver, A., Lladó, X., Freixenet, J., Cuadra, M.B.: A review of atlas-based segmentation for magnetic resonance brain images. *Computer methods and programs in biomedicine* **104**(3), e158–e177 (2011)
5. Chetverikov, D., Svirko, D., Stepanov, D., Krsek, P.: The trimmed iterative closest point algorithm. In: *Object recognition supported by user interaction for service robots*. vol. 3, pp. 545–548. *IEEE* (2002)
6. Cook, S.J., Jarrell, T.A., Brittin, C.A., Wang, Y., Bloniarz, A.E., Yakovlev, M.A., Nguyen, K.C.Q., Tang, L.T.H., Bayer, E.A., Duerr, J.S., Bulow, H.E., Hobert, O., Hall, D.H., Emmons, S.W.: Whole-animal connectomes of both *caenorhabditis elegans* sexes. *Nature* **571**(7763), 63–71 (2019)
7. Dickie, D.A., Shenkin, S.D., et al.: Whole brain magnetic resonance image atlases: a systematic review of existing atlases and caveats for use in population imaging. *Frontiers in neuroinformatics* **11**, 1 (2017)

8. Evangelidis, G.D., Horaud, R.: Joint alignment of multiple point sets with batch and incremental expectation-maximization. *IEEE transactions on pattern analysis and machine intelligence* **40**(6), 1397–1410 (2018)
9. Jarrell, T.A., Wang, Y., Bloniarz, A.E., Brittin, C.A., Xu, M., Thomson, J.N., Albertson, D.G., Hall, D.H., Emmons, S.W.: The connectome of a decision-making neural network. *Science* **337**(6093), 437–44 (2012)
10. Jones, A.R., Overly, C.C., Sunkin, S.M.: The allen brain atlas: 5 years and beyond. *Nature Reviews Neuroscience* **10**(11), 821–828 (2009)
11. Kainmueller, D., Jug, F., Rother, C., Myers, G.: Active graph matching for automatic joint segmentation and annotation of *c. elegans*. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 81–88. Springer (2014)
12. Lein, E.S., Hawrylycz, M.J., et al.: Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007)
13. Mazziotta, J., Toga, A., et al.: A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (icbm). *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **356** **1412**, 1293–322 (2001)
14. Mena, G., Varol, E., Nejatbakhsh, A., Yemini, E., Paninski, L.: Sinkhorn permutation variational marginal inference. In: *Symposium on Advances in Approximate Bayesian Inference*. pp. 1–9 (2020)
15. Oh, S.W., Harris, J.A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A.M., et al.: A mesoscale connectome of the mouse brain. *Nature* **508**(7495), 207–214 (2014)
16. Peyré, G., Cuturi, M., et al.: Computational optimal transport. *Foundations and Trends® in Machine Learning* **11**(5-6), 355–607 (2019)
17. Saxena, S., Kinsella, I., Musall, S., Kim, S.H., Meszaros, J., Thibodeaux, D.N., Kim, C., Cunningham, J., Hillman, E., Churchland, A., et al.: Localized semi-nonnegative matrix factorization (locanmf) of widefield calcium imaging data. *bioRxiv* p. 650093 (2019)
18. Sinkhorn, R., Knopp, P.: Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics* **21**(2), 343–348 (1967)
19. Tokunaga, T., Hirose, O., Kawaguchi, S., Toyoshima, Y., Teramoto, T., Ikebata, H., Kuge, S., Ishihara, T., Iino, Y., Yoshida, R.: Automated detection and tracking of many cells by using 4d live-cell imaging data. *Bioinformatics* **30**(12), i43–i51 (2014)
20. Toyoshima, Y., Wu, S., Kanamori, M., Sato, H., Jang, M.S., Oe, S., Murakami, Y., Teramoto, T., Park, C., Iwasaki, Y., Ishihara, T., Yoshida, R., Iino, Y.: An annotation dataset facilitates automatic annotation of whole-brain activity imaging of *c. elegans*. *BioRxiv* p. 698241 (2019)
21. White, J.G., Southgate, E., Thomson, N.J., Brenner, S.: The structure of the nervous system of the nematode *caenorhabditis elegans*. *Philos Trans R Soc Lond B Biol Sci* **314**(1165), 1–340 (1986)
22. Yemini, E., Lin, A., Nejatbakhsh, A., Varol, E., Sun, R., Mena, G.E., Samuel, A.D., Paninski, L., Venkatachalam, V., Hobert, O.: Neuropal: A neuronal polychromatic atlas of landmarks for whole-brain imaging in *c. elegans*. *BioRxiv* p. 676312 (2019)
23. Zitova, B., Flusser, J.: Image registration methods: a survey. *Image and vision computing* **21**(11), 977–1000 (2003)