Automated Annotation of Cell Identities in Dense Cellular Images

Shivesh Chaudhary[1], Sol Ah Lee[1], Yueyi Li[1], Dhaval S. Patel[1], Hang Lu[1,2*]

[1] School of Chemical & Biomolecular Engineering, Georgia Institute of Technology, Atlanta, Georgia, United States of America

[2] Petite Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, Georgia, United States of America

[*] Correspondence should be addressed to HL: hang.lu@gatech.edu.

Abstract

Assigning cell identities in dense image stacks is critical for many applications, for comparing data across animals and experiment conditions, and investigating properties of specific cells. Conventional methods are laborious, require experience, and could introduce bias. We present a generalizable framework based on Conditional Random Fields models for automatic cell identification. This approach searches for optimal arrangements of labels that maximally preserves prior knowledge such as geometrical relationships. The algorithm shows better accuracy and more robust handling of perturbations, e.g. missing cells and position variability, with both synthetic and experimental ground-truth data. The framework is generalizable across strains, imaging conditions, and easily builds and utilizes active data-driven atlases, which further improves accuracy. We demonstrate the utility in gene-expression pattern analysis, multi-cellular calcium imaging, and whole-brain imaging experiments. Thus, our framework is highly valuable to a wide variety of annotation scenarios including in zebrafish, Drosophila, hydra, and mouse brains.

Introduction

Biological name annotation of anatomical regions in images is a critical step in several domains, e.g. evolutionary and developmental phenotyping, spatial omics, and gene expression analysis. This is also the case for neuroscience: identifying brain regions, cells, cell types, etc. is a crucial step in image data analysis necessary for comparison across subjects, trials, experimental conditions, and facilitating the utility of existing knowledge about the system. However, the annotation task is typically challenging: reference atlases provide a static and often single view of the anatomy, while anatomical features vary across individuals and experimental conditions. Thus, manually matching data to atlas requires practice and is exceedingly laborious. Moreover, variations in experimental parameters such as exact resolution, orientation of animals during acquisition may not match the static atlases, making manual labeling infeasible. While computational pipelines have been proposed for identifying anatomical features in imaging data[1–8], most focus on naming coarse anatomical regions. As imaging of large structures is enabled by advanced microscopic techniques[9,10], the current bottleneck to generate interpretable data is, in part, due to methods to accurately annotate identities at cellular resolution in large image sets, and doing so under the constraints of biological variability.

For instance, cell identification in images is a critical component in many studies in *C. elegans* such as gene expression pattern analysis[11,12], lineage tracing[13], multi-cell calcium imaging[14] and whole-brain imaging[15–18]. Previous methods[11,19,20] focused on identifying sparsely distributed cells with stereotypical positions in young larvae animals. Tools for automatic and unbiased identification of cells in dense head ganglion do not exist. Further, all methods[11,19–22] for automatic annotation of cell identities in *C. elegans* are either registration-based or formulate a

linear assignment problem. Objective function in these methods minimizes the distances between cell specific features (such as positions of cells) in images and atlas. Thus, these methods only maximize extrinsic similarity[23] between images and atlas, which is sensitive to variability in cell positions. A better criterion for accurate annotation is to maximize intrinsic similarity, which is more robust against position noise[23,24] and inherently captures dependencies between cell label assignments. For instance, if region 1 is anterior and to region 2 in atlas and region 1 is assigned certain label, then the label available for region 2 is automatically dependent on region 1's assignment. Previous methods either do not optimize directly for such dependencies or only impose them indirectly as constraints in post-processing steps.

To directly optimize for dependencies between label assignment, we cast the cell annotation problem as a Structured Prediction problem[25–30] and build a fully connected Conditional Random Fields (CRF) model[31] to solve it. The model searches for, among all possible labeling arrangements, an optimal assignment to each cell that is maximally unbiased and most consistent with prior knowledge (e.g. label dependencies known in atlas). To maximize accuracy, we encode dependencies between all pairs of cells in the form of several positional relationship features in the model. These features include binary positional relationship features, the Gromov-Wasserstein discrepancy between cells in an image and the atlas[32,33], and an angular relationship feature. Using both synthetic data with realistic properties and manually annotated experimental ground-truth data, we demonstrate better performance of our method compared to previous methods for several tasks (up to 20% improvement).

There are several advantages of the CRF framework. First, the CRF framework can encode arbitrary order dependencies between labels. Additionally, ad hoc features can be added to the model to improve accuracy. We demonstrate this by incorporating spectral information, and landmarks in the model when such information was available. Second, the CRF framework is trainable algorithm[34,35] and can easily incorporate information from annotated data in the form of active data-driven atlas. We show that building such atlas is easy for our methods and requires cheap mathematical operations – simple averaging – thus making it computationally favorable for building atlas from thousands of images.

We show the utility of our approach in several contexts: determining gene expression patterns with no prior expectations, tracking activities of multiple cells during calcium imaging, and identifying cells in whole-brain imaging videos. For the whole-brain imaging experiment, our annotation framework enabled us to analyze the simultaneously recorded response of *C. elegans* head ganglion to food stimulus and identify two distinct groups of cells whose activities correlated with distinct variables – food sensation and locomotion.

Results

Structured prediction framework for automatic identification of neurons

Our annotation framework consists of 4 steps (Fig. 1, Supplementary Fig. 1, Supplementary Note 1). First, cells are automatically segmented using a Gaussian Mixture-based method; if available, cells with known identities (landmarks) are also detected in this step. Second, a head coordinate is generated by solving an optimization problem with considerations of the directional consistency of axes (Supplementary Note 1.3). With this coordinate system, we next

define cell specific features (unary potentials) and co-dependent features (pairwise potentials) in the model. The basic model uses several pairwise relationship features for all pairs of cells, including binary positional relationships, angular relationship, and the Gromov-Wasserstein discrepancy between cells in the image and an atlas. By encoding these features among all pairs of cells, our fully-connected CRF model accounts for label dependencies between each cell pair to maximize accuracy. Third, identities are automatically predicted for all neurons iteratively, taking into account neurons missing in the image stack (Supplementary Note 1.4). Duplicate assignments are handled by calculating a label-consistency score for each neuron, removing assignments with low scores (Supplementary Note 1.5) and re-running optimization. Lastly, identities predicted across each run are pooled to generate top candidate identities for each cell (Supplementary Note 1.6, Supplementary Video 1).

Identity assignment using intrinsic features outperforms other methods

Given the broad utility of image annotation, we envision our workflow to apply to a variety of problems where experimental constraints and algorithm performance requirements may be diverse. These use cases require our framework to be flexible and accurate. Furthermore, experimental data inherently contains perturbations that can affect annotation accuracy, e.g. deviation between cell positions in images and positions in atlas (position noise), different counts of cells in images and atlas due to missing cells in images (count noise), and presence or absence of landmarks with known identities (landmarks). Thus, we used two different kinds of data to tune the model and to assess accuracy: synthetic data generated from OpenWorm 3D atlas[36] (Supplementary Fig. 2a,b, Supplementary Fig. 3) and experimental data consisting of annotated ground-truth of 9 animals, with ~100 uniquely identified neurons. To tune the

126 features in the model, we compared prediction accuracy for several combinations of positional
127 relationship features. Among all co-dependent positional relationship features, the angular
128 relationship feature by itself or when combined with PA, LR, and DV binary position relationship
129 features performed best (Supplementary Fig. 4a).

130 While experimental data enables the assessment of prediction accuracy in real scenarios,
131 synthetic data enables us to dissect the effect of various perturbations independently. To assess
132 the effects of position noise and count noise on prediction accuracy, we simulated four
133 scenarios using the synthetic data (Supplementary Fig. 2c). In the absence of any perturbation,
134 relative positional relationship features predicted neuron identities with perfect accuracy, thus
135 demonstrating the suitability of co-dependent features and CRF framework for the annotation
136 task. We show that both position noise and count noise affect accuracy significantly
137 (Supplementary Fig. 2c,d) with position noise having a larger effect (compare scenarios 1-2 with
138 3-4, Supplementary Fig. 2c). Count noise is primarily caused by variability in the expression level
139 of the reporter used to identify cells, and limits in the computational methods to detect cells.
140 Results on both synthetic data and real data predict 10-15% improvement in prediction
141 accuracy can be attained by simply improving reagents and eliminate count noise
142 (Supplementary Fig. 2d). Another advantage of simulations using synthetic data is that it can be
143 used to obtain expected accuracy bounds by applying extreme case perturbations observed
144 empirically, thus setting an expectation on the performance of the method in real scenarios.
145 We obtained such bounds based on observed position noise of cells in experimental data
146 (Supplementary Fig. 2e). Indeed, the results for experimental data lied close to these bounds
147 (Supplementary Fig. 2f, Fig. 2 gray regions).

To account for missing cells, we developed a method that considers missing neurons as a latent state in the model (similar to hidden state CRF[37]) and predicts identities by marginalizing over latent states (Supplementary Note 1.6). Compared to the base case that assumes all cells are present in an image, simulating missing neurons significantly increased the prediction accuracy (Fig. 2a left) on experimental data; the top 5 candidate labels generated by this method attained ~80% average accuracy (Fig. 2a right), similar to synthetic data (Supplementary Fig. 2f).

Another way to improve the cell identification accuracy is to use landmarks. These landmarks act as additional constraints on the optimization while the algorithm searches for the optimal arrangement of labels for non-landmark cells. We showed using both experimental data (Fig. 2b) and synthetic data (Supplementary Fig. 4b), randomly chosen landmarks increased prediction accuracy by ~10-15%. It is possible that strategic choices of landmarks could further improve accuracy.

Next, we compared our method against registration method popular for automatic cell annotation[11,19–22] (Fig. 2c, Supplementary Note 2). Remarkably for both experimental and synthetic data, relative positions in the CRF framework performs the best (Fig. 2d, Supplementary Fig. 5a). Further, the superiority of the CRF framework using relative positions is insensitive to the position noise level in the synthetic data (Supplementary Fig. 5b). This has important practical implications as neuron positions being highly variable across individual animals has been shown[38], and confirmed with our datasets (Supplementary Fig. 6a,b). Because cell positions on average can vary by more than the distance to their tenth nearest neighbor (Supplementary Fig. 6b), we expect that this variability introduces large matching errors in registration methods. In contrast, most pair-wise relationships are preserved despite the

variability of absolute positions (Fig. 2e, Supplementary Fig. 6c,d). Interestingly, combining registration using absolute positions with relative position features corrupts the annotation performance (Supplementary Fig. 5a), likely due to competing effects in the objective function.

Improvement in accuracy can be further achieved by incorporating information from annotated data via a data-driven atlas that better accounts for variability in experimental data. Building such data-driven atlas for our CRF framework is easy requiring only simple averaging operations (Supplementary Note 1.7). By using data-driven atlas, the accuracy improved significantly to 74% for the top label and 95% for the top 5 labels (Fig. 2f).

Cell annotation in gene expression pattern analysis

We next demonstrate the utility of our framework for gene expression analyses, which is important for many problems, e.g. mapping the molecular atlas of neurotransmitters[39,40], receptors[41], and neuropeptides[42]. Conventional methods e.g. screening a list of cell specific marker lines that overlap with the reporter are laborious and scale badly with the number of cells expressing the genes of interest and the number of new genes for which expression patterns are to be determined. Automatic cell annotation can considerably reduce manual efforts by generating a small list of candidate identities for each cell expressing the reporter. Subsequently, researchers can easily verify or prune the candidate list. To demonstrate this use case, we imaged a strain with multiple cells labeled with GFP and predicted candidate identities of each cell. Determining cell identities in this case is difficult due to large count noise along with position noise: since the full list of labels in the atlas is much bigger than few cells in the reporter strain (scenario 4, Supplementary Fig. 2c). Thus, several degenerate (equally probable)

191 solutions are possible. To avoid this, the reporter strain was crossed with a strain expressing
192 pan-neuronal red fluorescent protein (RFP), from which candidate identities were predicted.
193 Our framework accurately generated a candidate list for cells across all datasets (n = 21
194 animals); 85% of cells had true identities within the top 5 labels chosen by the framework. In
195 comparison, the candidate list generated by the registration method achieved only 61%
196 accuracy (Fig. 3).

197 Cell annotation in multi-cell functional imaging experiments

198 We next demonstrate the utility of our algorithm in another important application - annotating
199 cell identities in multi-cell calcium functional imaging *in vivo* (Fig. 4a). Automation in this case
200 dramatically reduces labor associated with cell annotation for many time points, across trials,
201 animals, and experiments. We used a strain carrying GFP in multiple cells as a proxy for GCaMP-
202 labeled strains (Fig. 4a). Given the known candidate list of labels that can be assigned (no count
203 noise), the configurational space is small, which makes the task easy (similar to scenario 3
204 Supplementary Fig. 2c). Indeed, our annotation framework identified neurons with high
205 accuracy (98%, n = 35 animals). In comparison, the registration method predicted identities
206 with lower accuracy (88%) even with the small label assignment space (Fig. 4b). In reality, some
207 neurons may be undetected in the data due to expression mosaicism or low calcium transients
208 (equivalent to scenario 4, Supplementary Fig. 2c). We simulated this case by randomly
209 removing up to a third of total neurons from the images and predicting identities of remaining
210 cells using the full label list (Fig. 4c, Supplementary Fig. 7). Even under these conditions, the
211 accuracy of our method remains high (88%) significantly outperforming registration method

(81%) (Supplementary Video 2). In practice, the performance can be further compensated for by using multiple frames from each video.

To further facilitate annotation accuracy, we explored the utility of landmarks. Landmarks can also help establish a coordinate system and guide post-prediction correction. Because the combinatorial space of potential landmarks is very large (~10^{14} for 10 landmarks out of ~200 cells in the head), we asked what properties landmarks should have. We found that landmarks distributed throughout the head or in lateral ganglion perform better in predicting identities of neurons in all regions of the brain (Supplementary Fig. 8, Methods). As a test case, we developed strains with spatially distributed, sparse neuronal landmarks using CyOFP (Supplementary Note 3), which by itself can assist researchers in cell identification tasks. When crossed with pan-neuronally expressing GCaMP/RFP reagents, the strains can be used for whole-brain imaging (Fig. 4d) by using only two channels. This has two advantages: CyOFP can be imaged "for free" while imaging GCaMP and RFP simultaneously, thus the landmarks providing a concurrent reference in all frames; this strategy also leaves other channels open for optogenetic manipulations and voltage imaging[43,44].

We next tested this strategy in a simple whole-brain imaging experiment. Isoamyl alcohol (IAA) is a well-known component of the bacterial metabolites that *C. elegans* senses and responds to[45–47]. We recorded neuronal responses to a step-change in IAA concentration using a microfluidic system (Supplementary Fig. 9). We observed both odor-specific responses and spontaneous activities (Fig. 4e). More importantly, neurons with algorithm-assigned identities demonstrate expected behavior. For instance, we identified the sensory neuron AWC, and detected an off-response to IAA, consistent with known AWC behavior. In addition, the

predicted interneurons (e.g. AVA, RIB, and AIB) also demonstrate previously known activity patterns[16].

We also tested worms' responses to periodic stimuli of a more complex and naturalistic input – supernatant of bacterial culture (Fig. 5, Supplementary Video 3). A periodic input (5 s On and 5 s Off for 8 cycles) entrains many neurons as expected, therefore allowing us to better separate the odor-elicited responses from spontaneous activities (Fig. 5a). We generated the candidate identities for all recorded neurons (Supplementary Fig. 10a). Notably, several highly entrained neurons were identified as sensory neurons known to respond to food stimuli [48–50] (Fig. 5c), some of which responded to the onset of the stimuli and some to the withdrawal of the stimuli (Fig. 5d). The power spectrum of these neurons showed a strong frequency component at 0.1 Hz as expected (Fig. 5b).

Next, to examine the latent dynamics in the whole-brain activities during the entire experiment, we used traditional Principal Component Analysis (PCA) and Sparse Principal Component Analysis (sPCA)[51]. The overall dynamics are low-dimensional with top 3 traditional PCs capturing 70% of the variance (Supplementary Fig. 10b). In comparison, while the top 3 sparse PCs (SPCs) explain 43% of the variance in the data, they enable meaningful interpretation of the latent dynamics by eliminating mixing of activity profiles in PCs (Fig 5e). SPC1 shows a systematic decline of the signals, presumably related to photobleaching of the fluorophores; both SPC2 and SPC3 illustrate spontaneous activities with different temporal dynamics (Fig 5e). With automatic annotation, we were able to identify cell classes belonging to each SPC (Supplementary Fig. 10c). We then analyzed the relationship between motion and neuron activities. In our microfluidic device, the animals are not fully immobilized. By tracking

landmarks on the body; we observed propagating waves along the body (Fig 5f, Supplementary Fig. 10d, Supplementary Video 4). Interestingly, cells participating in SPC2 showed significantly higher mutual information with motion than any other component (Fig. 5g). Examining the connection between activities of neurons that drive SPC2 and animal motion demonstrates that these neurons are indeed correlated or anti-correlated with the motion we detected (Fig. 5h); notably, command interneurons such as AVA, RIM, and motor neurons such as VA, DA correlate well with backward motion (Fig. 5h). Cross-correlation analysis between motion and neuron activities showed that neurons are activated ahead of motion (Fig. 5i); when lag is added to the neuron activities, the mutual information of SPC2 neurons with motion is maximum at the same delay observed in the cross-correlation (Supplementary Fig. 10e). These experiments together demonstrate the power of the approach, which enabled previously difficult simultaneous analyses of several sensory, inter-, and motor neurons' activities to natural food stimulus. Thus, automatic identity prediction enabled meaningful interpretation of the whole-brain data.

Framework is broadly applicable to wider conditions

Another important advantage of the CRF framework is its flexibility to incorporate additional information to improve the identification accuracy, by simply adding new terms in the objective function without disturbing the weights of existing features. Here we demonstrate this idea by using the recently developed NeuroPAL[38] that provides a unique chromatic code to each neuron (Fig. 6a). This code was included as a unary potential in the model. Using manually curated ground-truth data, we compared different methods. These methods included different orthogonal feature combinations, as used by previous approaches, thus providing insights into

which features perform best (Fig. 6b, Supplementary Note 2). Unsurprisingly, registration performs poorly (with or without color information); color alone is not sufficient, and color combined with spatial features improves the accuracy (whether registration or relative position is used). Notably, the best performing model uses relative position features in combination with color and without registration term (Fig. 6b, Supplementary Fig. 11a), achieving 76% accuracy for the top-label prediction. Further, for 88% of the neurons, the true identity is within the top three labels. A further improvement in the model accuracy was achieved by using data-driven atlas to account for biological variability in both the positional relationships and color (Fig. 6c, Supplementary Fig. 12). Using the data-driven atlas significantly improves the accuracy to >87% (top labels); more than 95% of the neurons have their true identities in the top three labels chosen by the model. We expect that more datasets for the atlas will continue to improve the accuracy.

Lastly, we show that our model is equipped to work with realistic complex scenarios of animals imaged in different orientations, often not rigid rotations (Fig. 6d). Identifying cells in these cases is challenging: manual annotation using the 2D-atlas[52] is not possible since it lacks left-right information; further, due to low-z sampling of image stacks, segmented positions of cells along z-axis are noisier. These challenges can be addressed by using the data-driven atlas. We imaged and manually annotated seven animals in microfluidic devices with varying degrees of orientations to test the methodology. With data-driven atlas built from animals imaged laterally, the prediction accuracy of top labels was 63% (76% for top three labels) which are reasonable for practical purposes. Accuracy was further improved when the atlas was updated

combining the data from animals imaged in rotated orientations: 65% for top labels (80% for top three labels) (Supplementary Fig. 11b, Supplementary Fig. 13).

Discussion

Annotating anatomical features and cellular identities in biological images are important tasks for many applications. Here, we demonstrated our CRF framework is suitable for fluorescently labeled cells in 3D images for many applications. Using both ground-truth experimental data of whole-brain image stacks and synthetic data generated from atlas, we showed that our framework is more accurate compared to existing approaches. We demonstrated using real examples how the pipeline can be used for analysis of gene expression pattern for instance, and for neuron identification from dense multi-cell or whole-brain imaging experiments. Further, our CRF framework significantly speeds up the cell identification compared to manual labeling while reducing bias.

With the pipeline, we address several challenges. There is ample evidence that anatomy varies from individual to individual, and from condition to condition. This variability, or position noise, is a major source of roadblock in effectively applying previous methods to annotate the whole-brain recording data. Because our framework leverages intrinsic similarity[23], it performs better than registration methods in handling position noise (Supplementary Fig. 5). Further, CRF formulation is more accurate in handling count noise i.e. cases of missing or undetectable cells in images (Fig. 4c, Supplementary Fig. 7), because the missing neurons do not upset the relationships among the detectable neurons in the CRF formulation while missing neurons introduces large uncertainty in registration methods. Lastly, the CRF method predicts identities

with sufficient accuracy for different postural orientations of the worms often seen in our microfluidic experiments. We expect that this superiority is maintained for any data that have relational information preserved, this is the case virtually in all biological samples where tissues are connected by matrix, such as in other whole-brain recordings or for registration of fixed tissues.

Building and using data-driven atlases in the pipeline is simple and yet highly effective. We expect that data from more animals, different orientations, age, and imaging techniques will further improve the generalizability. Since building such data-driven atlas for our framework requires only cheap mathematical operations (Supplementary Note 1), incorporating more data is quite simple and easily scalable. In contrast, other methods may require simultaneous or batch-wise registration of multiple images to one reference; this would require solving multiple constrained regression problems on increasingly large data sets, thus rendering them computationally unscalable.

While we only considered pairwise features in the current formulation, feature functions with arbitrary dependency can be included in the model that may further improve prediction accuracy[53,54]. Advances in structured energy minimization field[53,55,56] will enable tackling the increased complexity of combinatorial optimization in these cases. Our workflow is the first application of Structured Prediction framework, borrowing techniques from metric object matching literature for annotation in biological images. Log-linear parameterization in our framework makes the model a member of the exponential families[57]; thus, the objective function in our framework has striking similarities with the entropy-regularized optimal transport objective functions[32,58]. Therefore, improvements in speed can be achieved by

342 borrowing fast optimization techniques developed in optimal transport literature. Advances in
343 these fields will continue to improve the method development in image analysis.

344 We anticipate that by using our generalizable formulation, similar pipelines can be set up to
345 annotate more image sets in other organisms and build atlases. Data in many anatomical
346 annotation problems (e.g. brain atlas construction, registering images from different
347 modalities, comparing animals or related species to one another for developmental studies)
348 share a similar property, in that the anatomical features of interest maintain a cohesion from
349 sample to sample. This underlining cohesion lends itself to the CRF framework. As we have
350 shown, the pipeline is extremely flexible in incorporating new information. Thus, framework
351 should be easily modifiable catering to the data demands in other organisms including features
352 besides landmarks and spectral information such as cellular morphology and expected cellular
353 activities (e.g. calcium transients). Because the only inputs to our framework are segmented
354 anatomical regions in images and positional relationships among them, information already
355 available in data across organisms[1,6,59,60], the framework proposed here should be generally
356 useful for many problems in model organisms such as *Drosophila*[59,61], zebrafish[6], mammalian
357 brains[1,60]. Besides fluorescence, the pipeline should also be able to work with data from other
358 modalities including EM, live imaging, and fluorescence imaging from cleared tissues.

359

Online Methods

Reagents

For all experiments, animals were cultured using standard techniques[62]. A detailed list of strains used is provided in Supplementary Note 4.

Imaging

All imaging was performed using either a Perkin Elmer spinning disk confocal microscope (1.3 NA, oil objective) or Brucker Opterra II Swept field confocal microscope (Plan Fluor ELWD air objective) at 40x magnification, with an EMCCD camera.

To acquire data used for framework validation and comparison against other methods (Fig. 2), gene expression pattern analysis (Fig. 3), multi-cell calcium imaging (Fig. 4), imaging landmark strain (Fig. 4) and NeuroPAL imaging (Fig. 6), animals were synchronized to L4 stage and were imaged in an array device[63]. A single 3D stack was acquired with either 0.5 µm or 1 µm spacing between z-planes and 10 ms exposure time (except for NeuroPAL strain where exposure times of different channels were chosen based on the guidelines provided in NeuroPAL manuals[38]).

Whole-brain functional recording data while providing chemical stimulus were acquired using a microfluidic device designed for applying chemical stimulation[64] to the nose-tip of the animal. Here image stacks were acquired with 1 µm spacing between z-planes and 10 ms exposure for each z-plane. This enabled recording videos at 1.1 volumes/s while imaging two channels simultaneously (GCaMP and RFP). Animals were synchronized to Day-1 adult stage.

379 Generating synthetic data for framework tuning and comparison against other methods

380 Synthetic data was generated using the freely available 3D atlas at OpenWorm[36]. Atlas available
381 at Worm Atlas[52] was not used as it provides only a 2D view. To mimic the conditions
382 encountered in experimental data, two perturbations were applied to the 3D atlas
383 (Supplementary Fig 4). First, due to inherent biological variability, positions of cells observed in
384 images do not exactly match the positions in atlas. Thus, position noise was applied to each cell
385 in the atlas sampled from a normal distribution with fixed variance. Thus, the position of the i^{th}
386 cell p_i in synthetic data was determined as $p_i = p_{i,atlas} + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma^2)$. Here $p_{i,atlas}$ is the
387 position of the cell in the atlas. To determine the variance σ^2, we quantified the variance of cell
388 positions observed in experimental data (Supplementary Fig. 3a,c,e) using the strain with
389 neuronal landmarks. We calculated the 25th percentile and 75th percentile of the variance
390 across all cells across all animals (n = 31) to define the lower bound and upper bound position
391 noise observed in experimental data. However, this variability cannot be directly applied to the
392 atlas due to different spatial scales. Thus, we applied the 25th or the 75th percentile of the
393 variance of cell positions to the atlas scaled by the inter-cell distances in atlas (Supplementary
394 Fig. 3b,d,f,g,h) to define lower bound and upper bound noise to be applied to the atlas.

395 Second, although there are 195-200 neurons in head ganglion in *C. elegans,* only 100-130 cells
396 were detected in images. Remaining cells are not detected either due to low-expression levels
397 of fluorophores or segmentation methods to resolve densely packed cells. This increases the
398 complexity of determining the labels of cells. To illustrate, matching 195 cells in an image to
399 195 cells in the atlas is easier as only one or very few possible configurations of label
400 assignments exist that maximally preserves the positional relationships among cells. In contrast,

in the case of matching 100 cells in an image to 195 cells in atlas, many possible labeling arrangements may exist that equally preserve the positional relationships among cells. Thus, to simulate this case, randomly selected cells were marked as missing and identities were predicted for remaining cells only. Since no prior information was available on which regions of the head ganglion had more cells missing, we selected the missing cells uniformly across brain regions. Finally, bounds on prediction accuracy (shown as gray regions in Fig. 2 and Supplementary Fig. 2) were obtained as the average prediction accuracy across runs obtained on synthetic data by applying lower bound and upper bound position noise.

Generating ground-truth data for framework tuning and comparison against other methods

NeuroPAL reagents OH15495 and OH15500 were used to generate ground-truth data. 3D image stacks were acquired following the guidelines provided in NeuroPAL manual[38]. Identities were annotated in image stacks using the example annotations provided in NeuroPAL manual. Individual channel image stacks were read in MATLAB, gamma and contrast were adjusted for each channel individually so that the color of cells in the RGB image formed by combining the individual channels match as much as possible (perceptually) the colors of cells in NeuroPAL manuals. To annotate identities in the 3D stack, Vaa3D software was used[65].

Comparison against other methods

Detailed description of the methodology used for each method that the CRF framework was compared against is provided in Supplementary Note 2. Note, for fair comparisons, standard 3D atlas was used by all methods as the reference (including CRF framework) for defining positions of cells (used by registration methods) and for defining positional relationships among cells (used by the CRF framework).

Simulations for choosing landmark locations

Landmarks (cell with known identities) improve prediction accuracy by constraining the optimization problem as it forces the CRF framework to choose optimal labels for all cells such that they preserve their positional relationships with the cells with fixed identities. However, choosing an optimal set of landmarks is difficult. This is because the combinatorial space of choosing landmarks is huge (~10^{14} for 10 landmark cells out of 195 in head ganglion). Simulating each such combination and predicting identities is not computationally tractable. Thus, we asked which regions of the brain landmark cells should lie in. We divided the head ganglion region into three groups: anterior group consisting of anterior ganglion, middle group consisting of lateral, dorsal and ventral ganglion, and posterior group consisting of retrovesicular ganglion. Two hundred runs were performed for each group with 15 randomly selected landmarks in each run. We constrained the landmarks cells to lie in a specific group and assessed how well the landmarks in that group perform in predicting the identities of cells in other regions. Overall, landmarks in anterior and posterior groups performed badly in predicting identities of cells in posterior and anterior groups respectively. Landmarks in the middle group and landmarks spatially distributed throughout the head performed equally (Supplementary Fig. 8). We chose landmarks spatially distributed throughout the head due to practical advantages: spatially distributed landmarks can be easily identified manually in image stacks thus can be used as input to the CRF framework. In contrast cells in middle group are densely packed and may not be identified easily. We tested this using several reporter strain with GFP labeled cells. Further, landmarks should be reliably expressed across animals, should have known and verified expression patterns and should label neither too few cells (not useful)

nor too many cells (difficult identification). Thus, we chose *unc-47, gcy-32 and gcy-8* reporters for labeling landmarks.

Whole-brain data analysis

All videos were processed using custom software in MATLAB for automatic segmentation and tracking of nuclei in whole-brain image stacks. Tracks for nuclei with minor tracking errors were corrected in post-processing steps. Tracks with large tracking errors were dropped from the data.

Segmentation - Neurons were automatically segmented in image stacks using a gaussian mixture model based segmentation technique. Briefly here, a 3D gaussian mixture model is fitted to the intensity profiles in image stacks using expectation-maximization algorithm. The number of components in the model and the ellipsoidal shape of each component determines the number of nuclei segmented and their shapes.

Tracking – Custom software was used for tracking cells. Briefly, segmented nuclei at each timepoint in image stacks are registered to a common reference frame and temporally nearby frames to produce globally and locally consistent matching. Based on these matchings, consistency constraints such as transitivity of matching were imposed in the post-processing step to further improve tracking accuracy. A custom MATLAB GUI was used to quickly and manually inspect the accuracy of tracking. Tracks of cells with minor tracking errors were resolved using semi-automated method.

Cell identification – Identities were predicted using the CRF framework with positional features (Supplementary Note 1) and data-driven atlas. Landmarks cells with known identities were

identified in the CyOFP channel were provided as input to the framework to achieve higher accuracy.

Identification of stimulus tuned neurons – To identify stimulus tuned neurons, the power spectrum of activities of all cells within the stimulus application window (100 s – 180 s) was calculated using "fft" function in MATLAB. Cells that showed significant power (> 0.08) at 0.1 Hz were selected. This criterion identified all cells except two with low response amplitude to the stimulus however the response could be manually seen in the video. Thus, these cells were manually selected.

PCA and Sparse PCA – Principal Component analysis (PCA) of neuron activity time-series data was performed using in-built functions in MATLAB. Sparse Principal component analysis (SPCA) was performed using freely available MATLAB toolbox [66].

Neuron activities correlation to animal motion – To ascertain that the motion of the worm in device has signatures of wave-propagation in freely moving animals, we looked for phase shift in the velocity of the different regions of the animal in the device (similar to phase shift in curvature of body parts of animals seen in freely moving animals[67]). To calculate the velocity, displacement of randomly selected cells along the anterior-posterior axis of the animal was calculated (Supplementary Video 4) based on the tracking of cells. Cell displacements were smoothed using Savitzky-Golay filter. Subsequently, velocity of each cell was calculated by differentiating the displacement of each cell.

Mutual information (MI) of the obtained velocity signal was calculated with 1) stimulus tuned neurons, 2) neurons with significant weights in sparse principal components 1-3, and 3)

487 remaining cells. MI analysis requires estimating the joint probability density of velocity and
488 neuron activity. We used the kernel density estimation method to do so that uses Gaussian
489 kernel with bandwidth parameters (that specify the variance of gaussian kernel) set to [0.05,
490 0.05]. Cells grouped in SPC2 always had the largest mutual information with velocity regardless
491 of the choice of the bandwidth parameter.

492 Statistical Analysis

493 Standard statistical tests were performed using Paired Comparisons App in OriginPro 2020.
494 Details regarding the tests (sample size, significance, method) are reported in figure legends.
495 Following asterisk symbols are used to denote significance level throughout the manuscript - *
496 ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$). Wherever significance level not indicated implies not
497 significantly different (n.s).

498 Code and Data availability

499 Code and data used in this study can be accessed at
500 https://github.com/shiveshc/CRF_Cell_ID.git. This repository contains the following 1) All code
501 and individual components necessary for using CRF framework to annotate cells in new data,
502 visualize results, and build new atlases based on annotated data 2) Code to reproduce results
503 for comparison shown against other methods in this study, and 3) all raw datasets used in this
504 study as well as human annotations created for those datasets except whole-brain imaging
505 datasets.

Acknowledgements

The authors acknowledge the funding support of the U.S. NIH (R21DC015652, R01NS096581, R01GM088333) and the U.S. NSF (1764406 and 1707401) to HL. Some nematode strains used in this work were provided by the Caenorhabditis Genetics Center (CGC), which is funded by the NIH (P40 OD010440), National Center for Research Resources and the International *C. elegans* Knockout Consortium.

Author Contributions

SC and HL designed the algorithm, experiments and methods. SC, SL collected whole-brain imaging data. SC, SL, YL and DSP developed strain with neuronal landmarks; SC, SL, YL, and HL analyzed the data. SC and HL wrote the paper.

Competing Interests statement

The authors declare no competing interest.

References

1. Chen, Y. *et al.* An active texture-based digital atlas enables automated mapping of structures and markers across brains. *Nat. Methods* **16,** 341–350 (2019).
2. Iqbal, A., Khan, R. & Karayannis, T. Developing a brain atlas through deep learning. *Nat. Mach. Intell.* **1,** 277–287 (2019).
3. Peng, H. *et al.* BrainAligner: 3D registration atlases of Drosophila brains. *Nat. Methods* **8,** 493–498 (2011).
4. Pacheco, D. A., Thiberge, S. Y., Pnevmatikakis, E. & Murthy, M. Auditory Activity is Diverse and Widespread Throughout the Central Brain of Drosophila. *bioRxiv* (2019).
5. Randlett, O. *et al.* Whole-brain activity mapping onto a zebrafish brain atlas. *Nat. Methods* **12,** 1–12 (2015).
6. Ronneberger, O. *et al.* ViBE-Z: A framework for 3D virtual colocalization analysis in zebrafish larval brains. *Nat. Methods* **9,** 735–742 (2012).
7. Lovett-Barron, M. *et al.* Ancestral Circuits for the Coordinated Modulation of Brain State. *Cell* **171,** 1411–1423.e17 (2017).
8. Mann, K., Gallen, C. L. & Clandinin, T. R. Whole-Brain Calcium Imaging Reveals an Intrinsic Functional Network in Drosophila. *Curr. Biol.* **27,** 2389–2396.e4 (2017).
9. Liu, T.-L. *et al.* Observing the cell in its native state: Imaging subcellular dynamics in multicellular organisms. *Science (80-.).* **360,** (2018).
10. Voleti, V. *et al.* Real-time volumetric microscopy of in vivo dynamics and large-scale samples with SCAPE 2.0. *Nat. Methods* **16,** 1054–1062 (2019).
11. Long, F., Peng, H., Liu, X., Kim, S. K. & Myers, E. A 3D digital atlas of C. elegans and its application to single-cell analyses. *Nat. Methods* **6,** 667–72 (2009).
12. Murray, J. I. *et al.* Automated analysis of embryonic gene expression with cellular resolution in C. elegans. *Nat. Methods* **5,** 703–709 (2008).
13. Bao, Z. *et al.* Automated cell lineage tracing in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci.* **103,** 2707–2712 (2006).
14. Iwanir, S. *et al.* Irrational behavior in C. elegans arises from asymmetric modulatory effects within single sensory neurons. *Nat. Commun.* **10,** (2019).
15. Schrödel, T., Prevedel, R., Aumayr, K., Zimmer, M. & Vaziri, A. Brain-wide 3D imaging of neuronal activity in Caenorhabditis elegans with sculpted light. *Nat. Methods* **10,** 1013–1020 (2013).
16. Kato, S. *et al.* Global Brain Dynamics Embed the Motor Command Sequence of Caenorhabditis elegans. *Cell* **163,** 656–669 (2015).

17. Venkatachalam, V. *et al. Pan-neuronal imaging in roaming Caenorhabditis elegans. Proceedings of the National Academy of Sciences of the United States of America* **113,** (2016).

18. Nguyen, J. P. *et al.* Whole-brain calcium imaging with cellular resolution in freely behaving Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U. S. A.* 33 (2015). doi:10.1073/pnas.1507110112

19. Long, F., Peng, H., Liu, X., Kim, S. & Myers, G. Automatic Recognition of Cells (ARC) for 3D images of C. elegans. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **4955 LNBI,** 128–139 (2008).

20. Aerni, S. J. *et al.* Automated cellular annotation for high-resolution images of adult Caenorhabditis elegans. in *Bioinformatics* **29,** (2013).

21. Scholz, M. *et al.* Predicting natural behavior from whole-brain neural dynamics. *bioRxiv* 445643 (2018). doi:10.1101/445643

22. Toyoshima, Y. *et al.* An annotation dataset facilitates automatic annotation of whole-brain activity imaging of C. elegans. *bioRxiv* 698241 (2019). doi:10.1101/698241

23. Bronstein, A. M., Bronstein, M. M. & Kimmel, R. Rock, paper, and scissors: Extrinsic vs. intrinsic similarity of non-rigid shapes. in *Proceedings of the IEEE International Conference on Computer Vision* (2007). doi:10.1109/ICCV.2007.4409076

24. Bronstein, A. M., Bronstein, M. M. & Kimmel, R. Topology-invariant similarity of nonrigid shapes. *Int. J. Comput. Vis.* **81,** 281–301 (2009).

25. Bakir, G. H. *et al. Predicting Structured Data (Neural Information Processing)*. (The MIT Press, 2007).

26. Nowozin, S. Structured Learning and Prediction in Computer Vision. *Found. Trends® Comput. Graph. Vis.* **6,** 185–365 (2010).

27. Caelli, T. & Caetano, T. Graphical models for graph matching: Approximate models and optimal algorithms. *Pattern Recognit. Lett.* **26,** 339–346 (2005).

28. Rusu, R. B., Holzbach, A., Blodow, N. & Beetz, M. Fast geometric point labeling using conditional random fields. *2009 IEEERSJ Int. Conf. Intell. Robot. Syst.* 7–12 (2009). doi:10.1109/IROS.2009.5354763

29. Kappes, J. H. *et al.* A Comparative Study of Modern Inference Techniques for Structured Discrete Energy Minimization Problems. *Int. J. Comput. Vis.* **115,** 155–184 (2015).

30. Litany, O., Remez, T., Rodola, E., Bronstein, A. & Bronstein, M. Deep Functional Maps: Structured Prediction for Dense Shape Correspondence. in *Proceedings of the IEEE International Conference on Computer Vision* **2017–Octob,** 5660–5668 (2017).

31. Lafferty, J., McCallum, A. & Pereira, F. C. N. Conditional random fields: Probabilistic

models for segmenting and labeling sequence data. *ICML '01 Proc. Eighteenth Int. Conf. Mach. Learn.* **8,** 282–289 (2001).

32. Solomon, J., Peyré, G., Kim, V. G. & Sra, S. Entropic metric alignment for correspondence problems. in *ACM Transactions on Graphics* **35,** (2016).

33. Mémoli, F. Gromov-Wasserstein Distances and the Metric Approach to Object Matching. *Found. Comput. Math.* **11,** 417–487 (2011).

34. Caetano, T. S., McAuley, J. J., Cheng, L., Le, Q. V. & Smola, A. J. Learning graph matching. *IEEE Trans. Pattern Anal. Mach. Intell.* **31,** 1048–1058 (2009).

35. Taskar, B., Guestrin, C. & Koller, D. Max-margin Markov networks. *Adv. Neural Inf. Process. Syst. 16 - NIPS'03* 25–32 (2003). doi:10.1.1.129.8439

36. Szigeti, B. *et al.* OpenWorm: an open-science approach to modeling Caenorhabditis elegans. *Front. Comput. Neurosci.* **8,** (2014).

37. Quattoni, A., Wang, S., Morency, L.-P., Collins, M. & Darrell, T. Hidden conditional random fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **29,** 1848–1853 (2007).

38. Yemini, E. *et al.* NeuroPAL: A Neuronal Polychromatic Atlas of Landmarks for Whole-Brain Imaging in C. elegans. *bioRxiv* 676312 (2019). doi:10.1101/676312

39. Gendrel, M., Atlas, E. G. & Hobert, O. A cellular and regulatory map of the GABAergic nervous system of *C. elegans*. *Elife* **5,** 1–38 (2016).

40. Pereira, L. *et al.* A cellular and regulatory map of the cholinergic nervous system of *C. elegans*. *Elife* **4,** (2015).

41. Vidal, B. *et al.* An atlas of Caenorhabditis elegans chemoreceptor expression. *PLOS Biol.* **16,** e2004218 (2018).

42. Bentley, B. *et al.* The Multilayer Connectome of Caenorhabditis elegans. *PLOS Comput. Biol.* **12,** e1005283 (2016).

43. Piatkevich, K. D. *et al.* Population imaging of neural activity in awake behaving mice. *Nature* **574,** 413–417 (2019).

44. Piatkevich, K. D. *et al.* A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters article. *Nat. Chem. Biol.* **14,** 352–360 (2018).

45. Chalasani, S. H. *et al.* Dissecting a circuit for olfactory behaviour in Caenorhabditis elegans. *Nature* **450,** 63–70 (2007).

46. L'Etoile, N. D. & Bargmann, C. I. Olfaction and odor discrimination are mediated by the C. elegans guanylyl cyclase ODR-1. *Neuron* **25,** 575–586 (2000).

47. Bargmann, C. I., Hartwieg, E. & Horvitz, H. R. Odorant-selective genes and neurons mediate olfaction in C. elegans. *Cell* **74,** 515–527 (1993).

48. Liu, H. *et al.* Reciprocal modulation of 5-HT and octopamine regulates pumping via feedforward and feedback circuits in C. elegans . *Proc. Natl. Acad. Sci.* **116,** 7107–7112 (2019).

49. Wakabayashi, T. *et al.* In vivo calcium imaging of OFF-responding ASK chemosensory neurons in C. elegans. *Biochim. Biophys. Acta - Gen. Subj.* **1790,** 765–769 (2009).

50. Zaslaver, A. *et al.* Hierarchical sparse coding in the sensory system of Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 1185–9 (2015).

51. Zou, H., Hastie, T. & Tibshirani, R. Sparse principal component analysis. *J. Comput. Graph. Stat.* **15,** 265–286 (2006).

52. Altun, Z. F. & Hall, D. H. Worm Atlas. *Wormatlas* (2009). doi:10.3908/wormatlas.1.14. Edited for the web by Laura A. Herndon. Last revision: April 30, 2012

53. Kohli, P., Ladický, L. & Torr, P. H. S. Robust higher order potentials for enforcing label consistency. *Int. J. Comput. Vis.* **82,** 302–324 (2009).

54. Najafi, M., Taghavi Namin, S., Salzmann, M. & Petersson, L. Non-associative higher-order markov networks for point cloud classification. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **8693 LNCS,** 500–515 (2014).

55. Komodakis, N. & Paragios, N. Beyond pairwise energies: Efficient optimization for higher-order mrfs. in *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009* **2009 IEEE,** 2985–2992 (2009).

56. Krähenbühl, P. & Koltun, V. in *Advances in Neural Information Processing Systems 24* (eds. Shawe-Taylor, J., Zemel, R. S., Bartlett, P. L., Pereira, F. & Weinberger, K. Q.) 109–117 (Curran Associates, Inc., 2011).

57. Wainwright, M. J. & and Jordan, M. I. Graphical Models, Exponential Families, and Variational Inference. *Found. Trends{®} Mach. Learn.* **1,** 1–305 (2007).

58. Nitzan, M., Karaiskos, N., Friedman, N. & Rajewsky, N. Gene expression cartography. *Nature* (2019). doi:10.1038/s41586-019-1773-3

59. Robie, A. A. *et al.* Mapping the Neural Substrates of Behavior. *Cell* **170,** 393–406.e28 (2017).

60. Kim, Y. *et al.* Mapping social behavior-induced brain activation at cellular resolution in the mouse. *Cell Rep.* **10,** 292–305 (2015).

61. Vaadia, R. D. *et al.* Characterization of Proprioceptive System Dynamics in Behaving Drosophila Larvae Using High-Speed Volumetric Microscopy. *Curr. Biol.* **29,** 935–944.e4 (2019).

62. Stiernagle, T. Maintenance of C. elegans. *WormBook : the online review of C. elegans biology* 1–11 (2006). doi:10.1895/wormbook.1.101.1

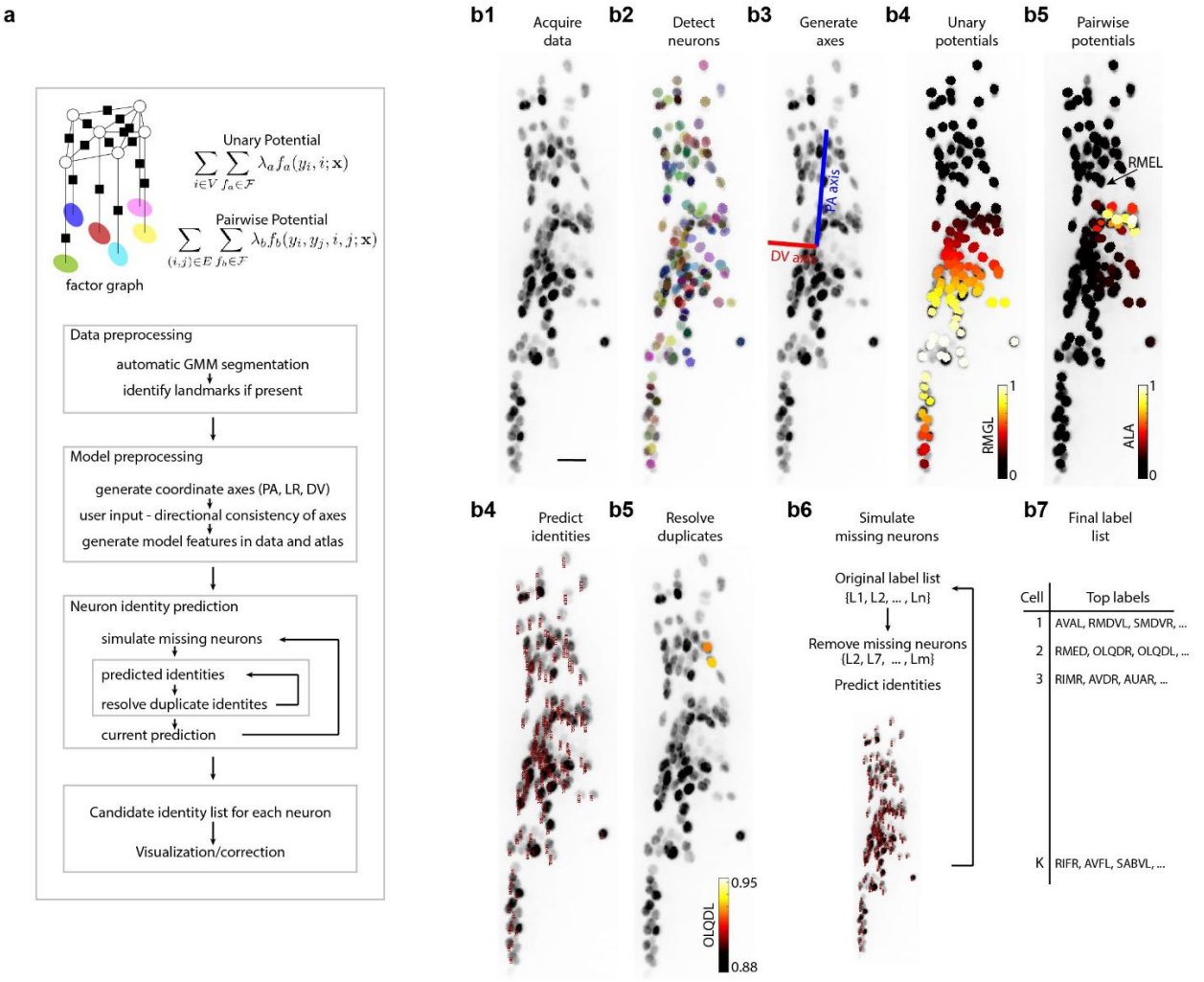
63. Lee, H. *et al.* A multi-channel device for high-density target-selective stimulation and long-term monitoring of cells and subcellular features in C. elegans. *Lab Chip* **14,** 4513–4522 (2014).

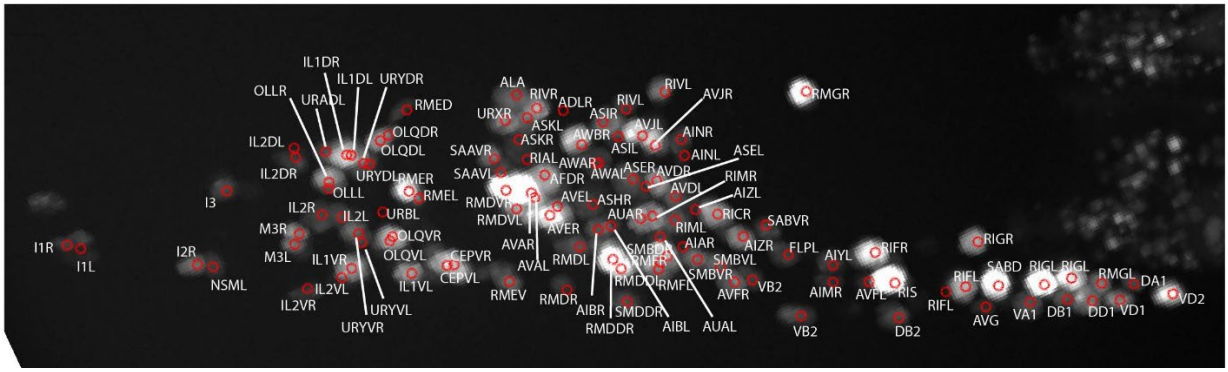
64. Cho, Y. *et al.* Multimodal Stimulation in a Microfluidic Device Facilitates Studies of Interneurons in Sensory Integration in C. elegans. *Small* **n/a,** 1905852 (2020).

65. Peng, H., Ruan, Z., Long, F., Simpson, J. H. & Myers, E. W. VAA3D enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets. *Nat. Biotechnol.* **28,** 348-U75 (2010).

66. Sjöstrand, K., Clemmensen, L. H., Einarsson, G., Larsen, R. & Ersbøll, B. SpaSM: A MATLAB toolbox for sparse statistical modeling. *J. Stat. Softw.* **84,** (2018).

67. Stephens, G. J., Johnson-Kerner, B., Bialek, W. & Ryu, W. S. Dimensionality and dynamics in the behavior of C. elegans. *PLoS Comput. Biol.* **4,** (2008).





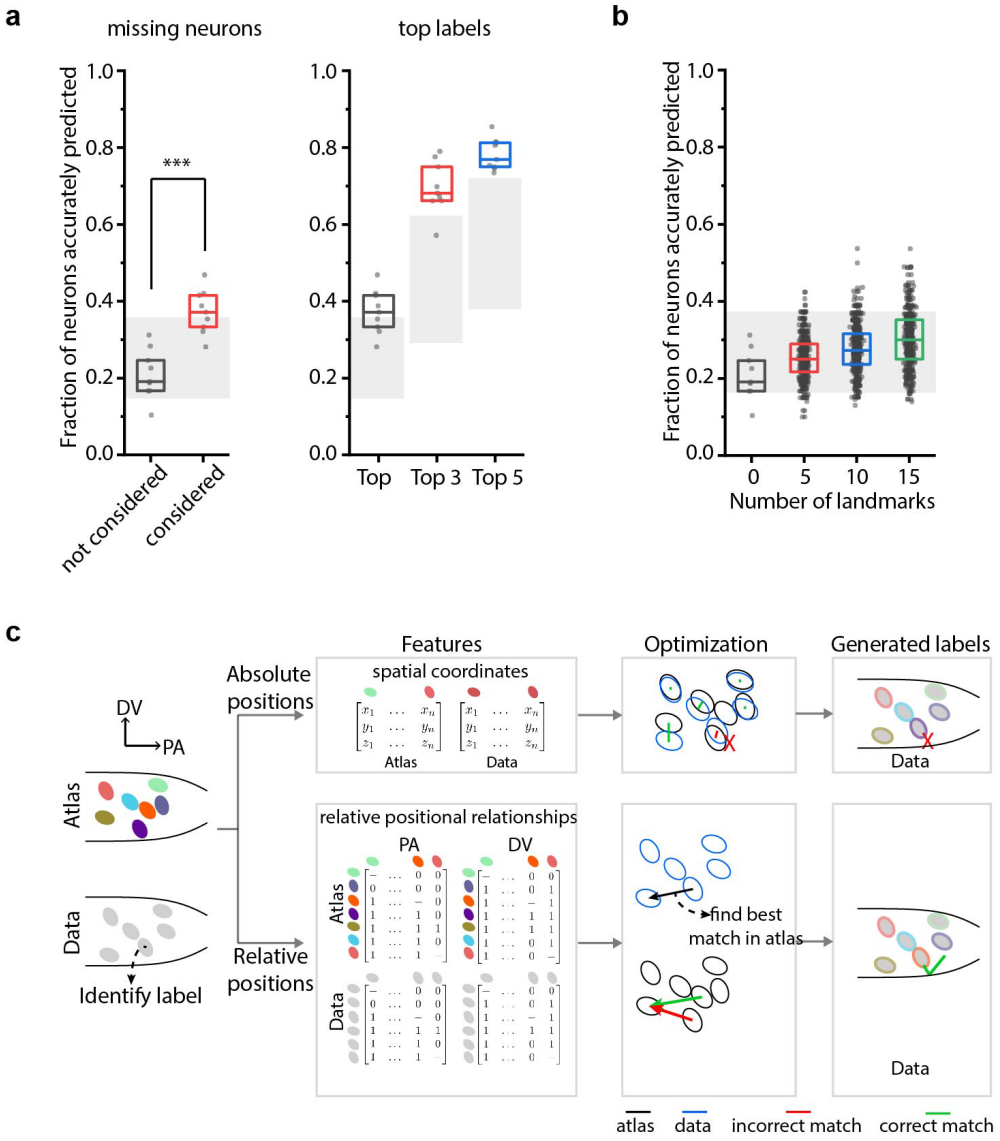
671

672

673 **Figure 1. CRF annotation framework automatically predicts cell identities in image stacks.**

674 a) CRF framework models a conditional joint probability distribution of labels assigned to
675 cells $P(\mathrm{y}/\mathbf{x})$. The underlying graph structure (factor graph) encodes two different kinds
676 of features – unary potentials (cell specific) and pairwise potentials (dependencies
677 between labels). The framework has four modules. 1) Data processing, 2) Preprocessing
678 for extracting features in the model, 3) Iterative prediction of neuron identities while
679 taking missing cells into account, and 4) Generating candidate name list for each cell.
680
681 b) Steps of CRF framework applied to neuron imaging in *C. elegans*. b1 - Max-projection of
682 a 3D image stack showing head ganglion neurons whose biological names (identities)
683 are to be determined. b2 – automatically detected cells (Methods) shown as overlaid
684 colored regions on the raw image. b3 – Coordinate axes are generated automatically
685 (Supplementary Note 1). b4 – an example of unary potentials showing the affinity of
686 each cell taking the label RMGL. b5 – an example of dependencies encoded by pairwise
687 potentials, shows the affinity of each cell taking the label ALA given the arrow-pointed
688 cell is assigned the label RMEL. b6 – identities are predicted by simultaneous
689 optimization of all potentials such that assigned labels maximally preserve the empirical
690 knowledge available from atlases. b7 – duplicate assignment of labels is handled using a
691 label consistency score calculated for each cell (Supplementary Note 1). b7 – the
692 process is repeated with different combinations of missing cells to marginalize over
693 missing cells (Supplementary Note 1). b9 – top candidate label list is generated for each
694 cell.
695
696 c) An example of automatically predicted identities (top picks) for each cell.

697



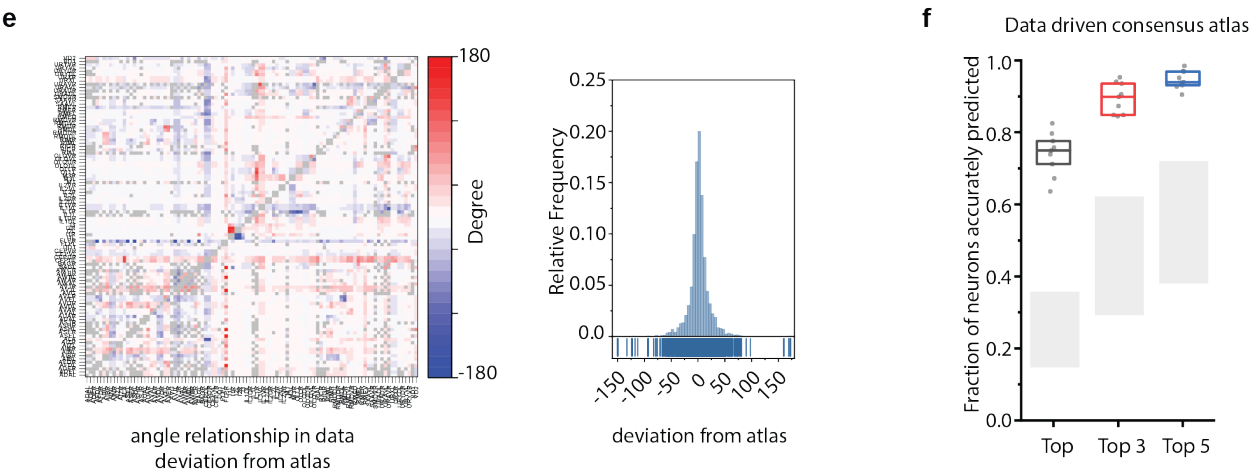


Figure 2. Annotation framework outperforms other approaches. Static OpenWorm atlas was used for predicting identities in all figures except f. Gray regions indicate bounds on prediction accuracy obtained using simulations on synthetic data.

a) (Left) Accounting for missing neurons improves prediction accuracy (n = 9 animals, *** p < 0.001, Bonferroni paired comparison test). (Right) True identities are in the top labels for majority of the neurons.

b) Landmarks cells improve prediction accuracy. For each condition (except no landmarks) 50 runs were performed for each ground-truth dataset (n = 9 animals) with randomly selected landmarks in each run. Thus, total n = 400-450 random combinations of landmarks tested for each condition.

c) Key difference between registration-based methods and our framework.

d) Prediction accuracy comparison across methods using ground truth experimental data (n=9, * p < 0.05, Bonferroni paired comparison test) and synthetic data (n = 190-200 runs, ***p < 0.001, Bonferroni paired comparison test). For synthetic data, random position and count noise applied in each run.

e) Pairwise positional relationships among cells are more consistent with atlas even though the absolute positions of cells vary across worms. (Left) average deviation of angular relationship measured in ground truth data (n=9) from the angular relationship in static atlas. (Right) distribution of all deviations in left panel (total of 8,516 relationships) is sparse and centered around 0 deviation, thus indicating angular relationships are consistent with atlas.

f) Using data-driven consensus atlas instead of static atlas improves prediction accuracy significantly. Results shown for experimental ground truth data (n = 9 animals).

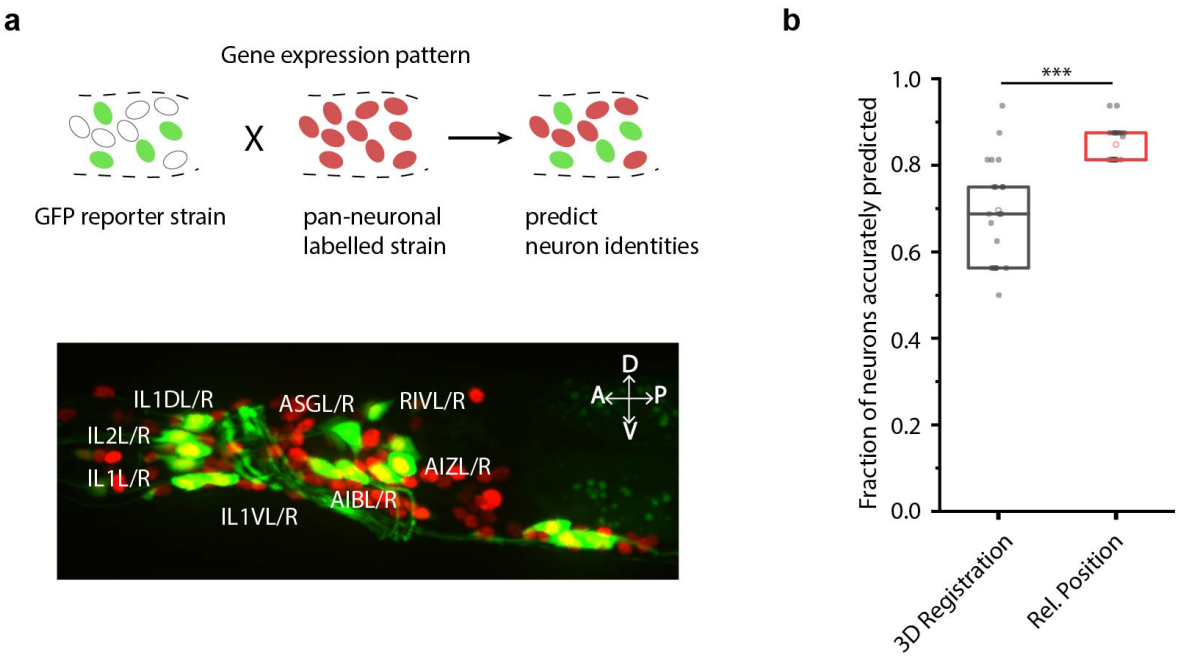
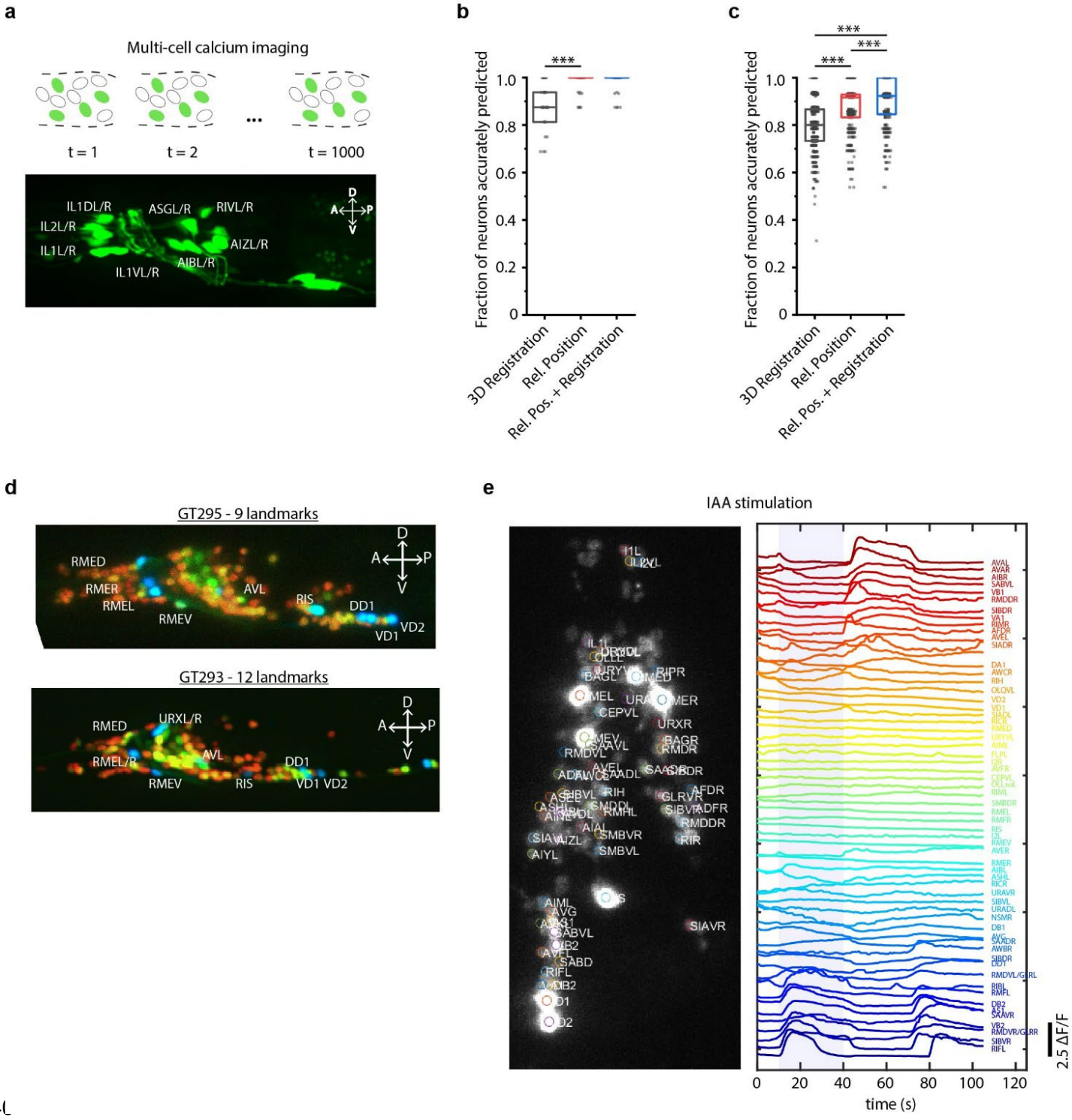


Figure 3. CRF framework predicts identities for gene expression pattern analyses.

a) (Top) Schematic showing a fluorescent reporter strain with GFP expressed in cells for which names need to be determined. Since no candidate labels are known *a priori* the reporter strain is crossed with a strain expressing RFP pan-neuronally and neuron labels are predicted. (Bottom) proxy strain carrying rab-3p(prom1)::2xNLS::TagRFP and odr-2p::GFP with 19 cells labeled with GFP was used to assess prediction accuracy.

b) CRF framework with relative position features outperforms registration method (n = 21 animals) (***$p < 0.001$, Bonferroni paired comparison test).



74(

741

Figure 4. Cell identity prediction in multi-cell calcium imaging experiments and landmark strain.

a) (Top) schematic showing automatic identification of cells in multi-cell calcium imaging for high-throughput analysis. (Bottom) proxy strain used with GFP labeled cells as an illustration of GCaMP imaging.

b) CRF framework outperforms registration method (n = 35 animals, *** $p < 0.001$, Bonferroni paired comparison test).

c) Prediction accuracy comparison for the case of missing cells in images (count noise). *** $p < 0.001$, Bonferroni paired comparison test. Total n = 700 runs were performed across 35 animals for each method with randomly selected cells removed in each run. For fair comparison, cells removed across methods were the same.

d) Max-projection of 3D image stacks showing CyOFP labeled landmark cells in head ganglion (pseudo-colored as cyan): animals carrying unc47p::CyOFP with 9 landmarks (top), and animals carrying [unc-47p::CyOFP; gcy-32p::CyOFP] with 12 landmarks (bottom).

e) (Left) max-projection of 3D image stack from whole-brain activity recording showing head ganglion cells and identities predicted by CRF framework (Top labels). Animal is immobilized in a microfluidic device channel and IAA stimulus is applied to the nose tip. T. (Right) GCaMP6s activity traces extracted by tracking cells over time in the same 108s recording and their corresponding identities. Blue shaded region shows IAA stimulation period.

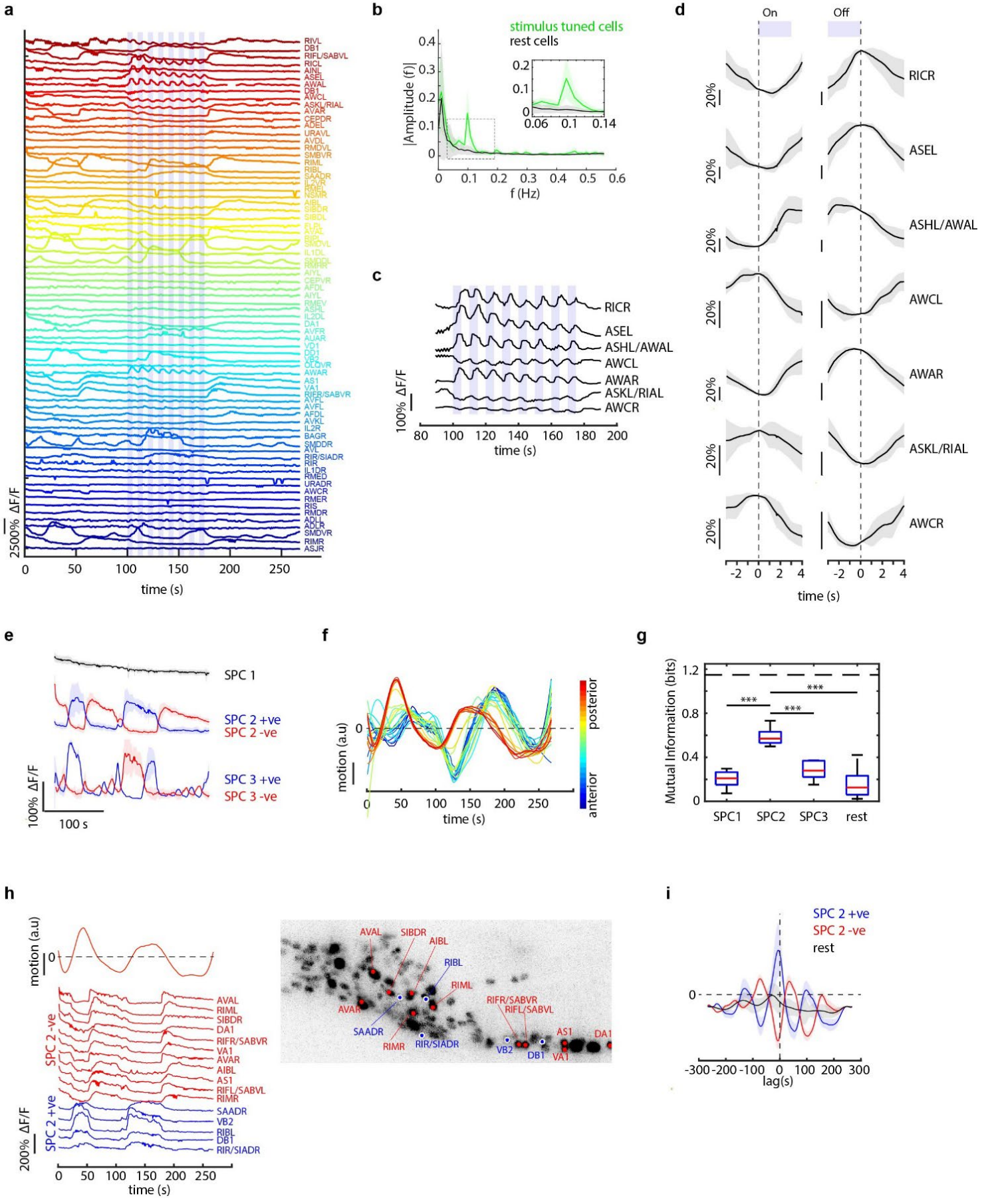


Figure 5. CRF framework identifies neurons representing sensory and motor activities in whole-brain recording.

a) GCaMP6s activity traces of 73 cells automatically tracked throughout a 278s long whole-brain recording and the corresponding predicted identities (top labels). Periodic stimulus (5 sec-on – 5 sec-off) of bacteria (*E. Coli* OP50) supernatant was applied starting at 100 s (shaded blue regions).

b) Power spectrum of neuron activity traces during the stimulation period for all cells. Cells entrained by 0.1 Hz periodic stimulus show significant amplitude for 0.1 Hz frequency component (green).

c) Activity traces of cells entrained by periodic stimulus shown for the stimulation period. Blue shaded regions indicate stimulus ON, unshaded region indicate stimulus OFF. Identities predicted by the framework are labeled.

d) Average ON and OFF responses of cells entrained by periodic stimulus across trials. The black line indicates mean and gray shading indicates ± s.e.m.

e) Average activities of neurons with significant non-zeros weights in the first 3 sparse principal components (SPCs). Activities within each component are stereotypical and different components show distinct temporal dynamics. Cells with positive weights (blue) and negative weights (red) in SPC2 and SPC3 showed anti-correlated activity. Out of the 67 non-stimulus-tuned cells, 19 had non-zero weights in SPC1, 16 cells had non-zero weights in SPC2 and 5 cells had non-zero weights in SPC3. Shading indicates mean ± s.e.m of activity.

f) Velocity (motion/second) traces of cells along anterior-posterior (AP) axis (blue to red) show phase shift in velocity indicating motion in device shows signatures of wave propagation.

g) Cells with non-zero weights in SPC2 show high mutual information with worm velocity compared to cells grouped in other SPCs (*** denotes $p < 0.001$, Bonferroni paired comparison test). Median (red line), 25^{th} and 75^{th} percentiles (box) and range (whiskers). Dashed line indicates entropy of velocity (maximum limit of mutual information between velocity and any random variable).

h) Activity traces of 16 cells (with significant non-zero weights) in SPC2 and corresponding identities predicted by the framework. Red traces for cells with negative weights in SPC2, blue traces for cells with positive weights in SPC2. Worm motion/second shown

808 on top. (Right) max projection of 3D image stack showing head ganglion neurons and
809 cells with positive weights (blue) and negative weights (red) in SPC2.

810

811 i) Cross-correlation analysis between velocity and cells with non-zero weights in SPC2
812 shows a strong correlation between neuron activities and velocity. In comparison, other
813 cells show low correlation.
814

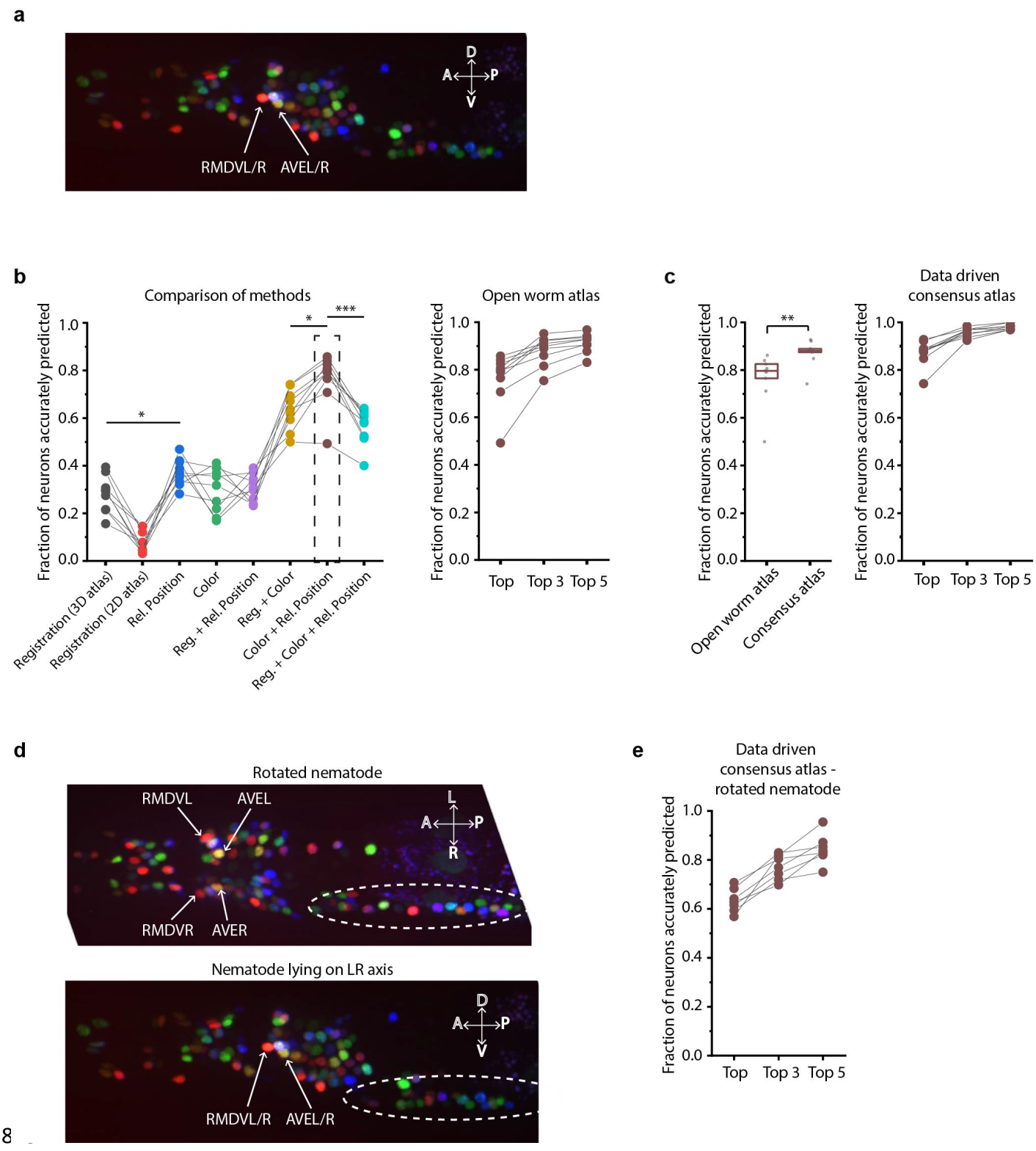


Figure 6. Annotation framework is generalizable and compatible with different strains and imaging scenarios.

a) A representative image (max-projection of 3D stack) of head ganglion neurons in NeuroPAL strain.

b) (Left) comparison of prediction accuracy for various methods that use different information. CRF framework that combines relative position features along with color information performs best (n = 9 animals, * $p < 0.05$, *** $p < 0.001$, Bonferroni paired comparison test). (Right) the best performing method predicts cell identities with high accuracy. OpenWorm static atlas was used for all methods to perform registration and to define positional relationship features among cells.

c) (Left) annotation framework can easily incorporate information from annotated data in the form of data-driven atlas, which improves prediction accuracy (** $p < 0.01$, Bonferroni paired comparison test). (Right) accuracy achieved by top labels.

d) An example image of head ganglion neurons in NeuroPAL strain for rotated animal (nematode lying on DV axis). In contrast, animal lying on the LR axis is shown below. The locations of RMDVL/R, AVEL/R cells in the two images are highlighted for contrasts. Dashed ellipses indicate positions of cells in retrovesicular ganglion, showing that the rotated animal is not rigidly rotated.

e) Top-label prediction accuracies for rotated animal. n = 7 animals.