

A 3D digital atlas of *C. elegans* and its application to single-cell analyses

Fuhui Long^{1,3}, Hanchuan Peng^{1,3}, Xiao Liu², Stuart K Kim² & Eugene Myers¹

We built a digital nuclear atlas of the newly hatched, first larval stage (L1) of the wild-type hermaphrodite of *Caenorhabditis elegans* at single-cell resolution from confocal image stacks of 15 individual worms. The atlas quantifies the stereotypy of nuclear locations and provides other statistics on the spatial patterns of the 357 nuclei that could be faithfully segmented and annotated out of the 558 present at this developmental stage. We then developed an automated approach to assign cell names to each nucleus in a three-dimensional image of an L1 worm. We achieved 86% accuracy in identifying the 357 nuclei automatically. This computational method will allow high-throughput single-cell analyses of the post-embryonic worm, such as gene expression analysis, or ablation or stimulation of cells under computer control in a high-throughput functional screen.

Despite the detailed knowledge of the anatomy of the nematode *C. elegans*¹ as well as its determined cell lineage^{2,3}, the mapped connectivity of its nervous system^{4,5} and its sequenced genome^{6,7}, we still lack a three-dimensional (3D) digital atlas of positions of nuclei in any postembryonic stage. Such an atlas has several potential applications. First, it provides us with previously unavailable quantitative knowledge about the degree of stereotypy of the positions of nuclei and the specific spatial relationships between different cells. Second, the atlas can serve as a standard template; we can compare any 3D image of a wild-type *C. elegans* to the atlas and extract the identities of individual nuclei using an automated approach. This is essential for high-throughput analysis of cellular information such as gene expression at single-cell resolution. Such an analysis provides much richer information than does analysis of expression data from a DNA microarray experiment^{8,9} as DNA microarrays reveal average expression from tissue or from the entire individual but not the expression in an individual cell.

Prior to this study, the anatomy of *C. elegans* has been described qualitatively by images with a text description or two-dimensional sketches¹⁰. Early efforts using electron microscopy analyses have resulted in detailed views of the anatomy¹⁰ and even a connectivity graph of the nervous system^{4,5}, but to date, manual or automated segmentation of the fine structure of such an ultra-high-resolution image stack has not been demonstrated. Whereas

one might contemplate carrying out such a manual segmentation for a single worm, doing so for enough worms to deliver statistical information on the location of nuclei is effectively impractical.

Our method for automatically analyzing individual cells in postembryonic worms complements the similar capability developed previously for the embryo^{11,12}. However, the computational problem is completely different. The previous method consists of tracking nuclei as they divide *in vivo* using known cell lineage information. In this work, we identified nuclei *in situ* without the help of temporal or lineage information. Moreover, our method is general and could be applied to other stereotypic systems, such as later post-embryonic worm stages or the fruit fly embryonic nervous system¹³. Here we present the method we used to build the atlas, followed by analyses of the model to confirm that it is well-constructed and recapitulates known biology. Then we describe our automated approach for using the atlas to annotate cell identities in images of first larval stage (L1) worms and demonstrate its accuracy.

RESULTS

Building a 3D digital atlas

We used DAPI (4,6-diamidino-2-phenylindole) to stain the nuclei of all 558 cells. We used a *myo-3::GFP* transgene to label the nuclei of the 81 body wall muscle cells and 1 depressor muscle cell. These nuclei were fiducial markers, used by our manual and automated approach to annotate cells. We used a gene encoding monomeric Cherry protein (mCherry) driven by a promoter from a gene of interest to reveal expression in a set of target cells. We collected 3D images of *C. elegans* at the L1 stage using a Leica confocal microscope (Fig. 1a) with a $\times 63$ oil lens and x - y and z dimension sampling set at 0.116 μm and 0.122 μm per pixel, respectively.

To build a standard digital atlas, we first computationally straightened the curved worm body in the 3D image into a rod shape¹⁴ (Fig. 1b, Supplementary Fig. 1 and Supplementary Video 1). The method first detects the principal curve or 'backbone' that represents the anterior-posterior line that passes from head to tail through the center of a straightened worm (Supplementary Fig. 1b). It then generates one-pixel separated planes orthogonal to the backbone and restacks them along a straight anterior-posterior line, making the planes parallel to each

¹Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia, USA. ²Department of Developmental Biology, Stanford University Medical Center, Stanford, California, USA. ³These authors contributed equally to this work. Correspondence should be addressed to E.M. (myersg@janelia.hhmi.org).

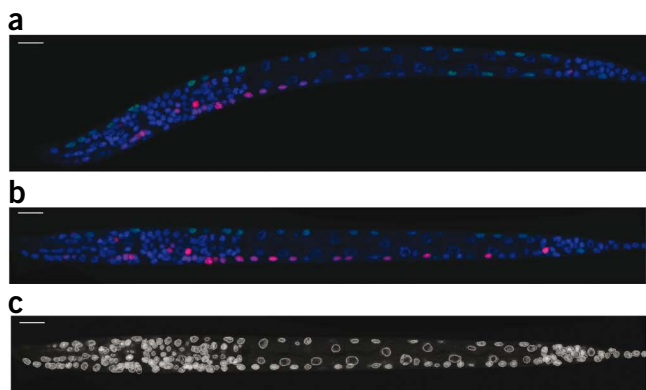


Figure 1 | Automatic processing of a 3D image of *C. elegans*. (a) A two-dimensional (2D) slice of a 3D image. Blue, DAPI; green, nuclear localization signal (NLS)-GFP expressed from the *myo-3* promoter; red, mCherry regulated by a promoter of interest (in this example, expression is in some ventral motor neurons and neurons in the nerve ring). (b) The same 2D slice after worm body straightening. (c) The segmentation result of the DAPI channel of same 3D image, with the same 2D slice as in a. Scale bars, 10 μm .

other (Supplementary Fig. 1c). Because each local transform is a rigid rotation, the resolution of the straightened image is merely a function of the subpixel interpolation error when a straight line is rotated in space. Thus, the overall resolution loss is naturally minimized.

We next developed an automatic approach to segment each individual nucleus in the 3D image of the straightened worm (Fig. 1c and Supplementary Fig. 2). First, our method applies 3D median filtering ($3 \times 3 \times 3$ pixels) followed by Gaussian filtering (s.d. = 1) to reduce noise so that the intensity distributions in nuclear regions become less variable. It then fills any intensity holes in the nuclei, which are typically nucleoli that are unstained by DAPI (Supplementary Fig. 2b). Next, our method uses adaptive thresholding to detect local background, generating a location-dependent foreground mask of nuclei or clusters of nuclei (Supplementary Fig. 2c). After that, it applies the distance transform, which computes the distance of a foreground pixel from the nearest background pixel, converting the binary foreground mask into a gradient image. An initial segmentation of nuclei is then generated from the gradient image using the 3D watershed algorithm^{15,16} (Supplementary Fig. 2d). Finally, to handle the small number of over- and under-segmented regions, we developed both rule- and training-based methods to merge and split regions. The rule-based method uses the statistical information of the segmented regions to predict regions of wrong segmentation and then uses rules defined on shape, size and intensity of typical nuclear regions to merge or split regions. The training-based method trains a support vector machine classifier^{17,18} using the intensity, size and shape of nuclei to determine whether a region should be further split or merged. Splitting and merging of regions are iterated until the classifier predicts that additional merging or splitting of a given region is unwarranted (Supplementary Fig. 2e,f). Overall, the training-based approach produces slightly better results.

Then, we manually validated, corrected and annotated the segmented nuclei using cell name conventions of the

C. elegans community. For this purpose, we had previously developed a 3D annotation and visualization tool called a volume-object image annotation system¹⁹ (VANO; Supplementary Fig. 3), which permits one to edit any observed errors in the segmentation and to enter a name for every segmented region or nucleus. Our manual annotation is based on the morphology and relative spatial positions of cells¹⁰ (Online Methods). As we used GFP to highlight the 81 body wall muscle cells and 1 depressor cell in a separate channel, we annotated these nuclei first. We then used these muscle cells as fiducial markers to identify additional nuclei whose spatial relationships with respect to these marker nuclei are stable. We added the newly annotated nuclei to the marker set and repeated this process. By doing so, we manually annotated approximately 357 nuclei in each image with high confidence (Supplementary Note and Supplementary Fig. 4).

Individual *C. elegans* images differ in size and orientation; therefore, the final step was to register or map all of the stacks into the same canonical space so that the positions of their nuclei are comparable. For this purpose, we computed the ‘median’ or ‘centroid’ of the collection of K stacks, say C , for which the sum

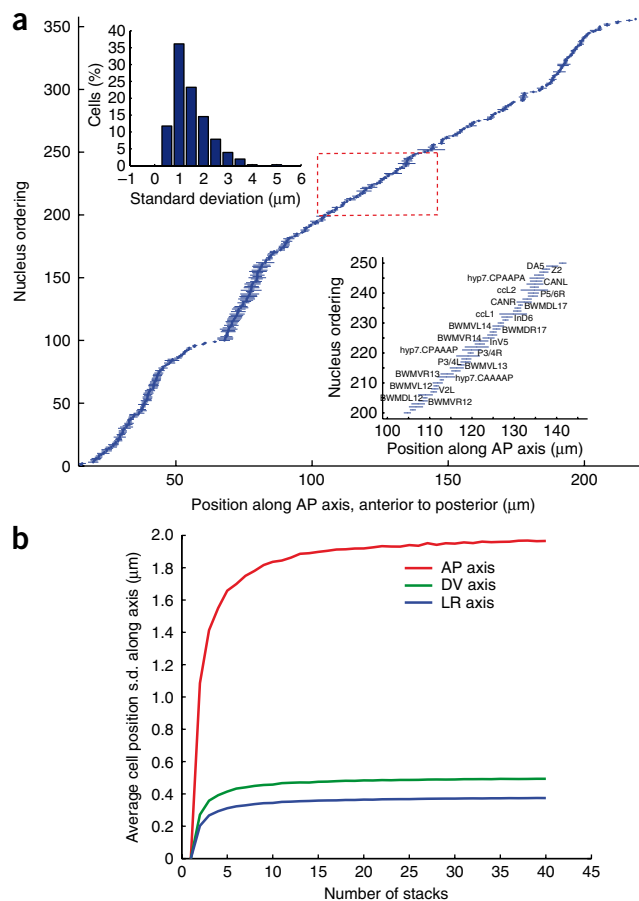


Figure 2 | Statistics of nuclear positions. (a) The mean (dots) and standard deviation (lines) of the locations of 357 nuclei along the AP axis computed from 15 randomly selected images of L1 hermaphrodites. The vertical axis is the ordering of the nuclei sorted according to their mean locations along the AP axis. Bottom inset, names of a subset of the 357 nuclei marked by the red box. Top inset, distribution of the standard deviation of nuclear locations of all the 357 nuclei. (b) The average standard deviation of nucleus positions along the indicated dimensions as functions of the number of stacks used to build the atlas.

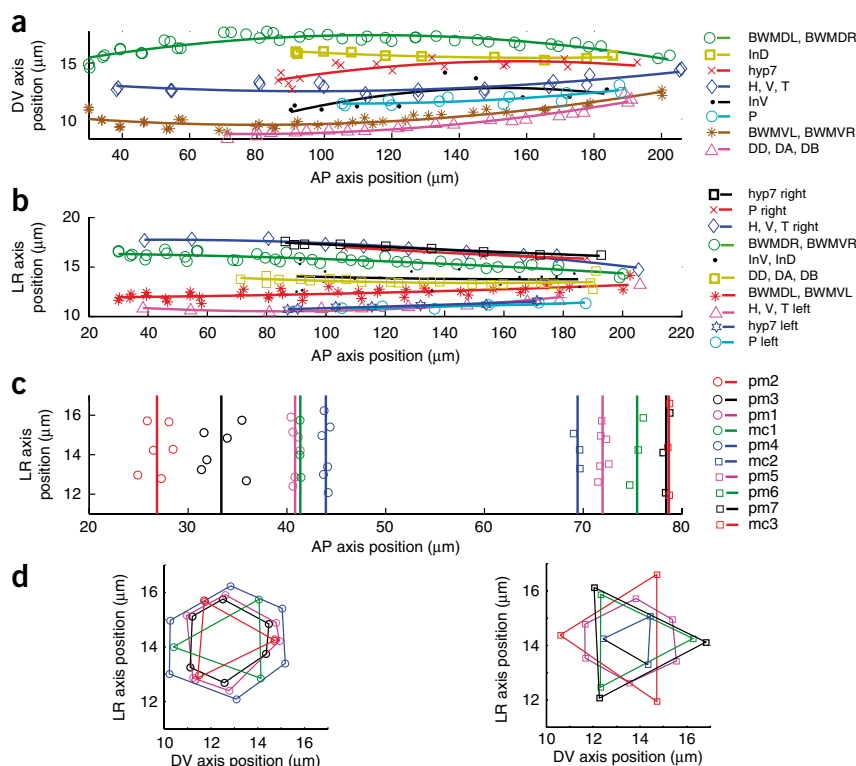


Figure 3 | The nuclear spatial location patterns for different cell types. **(a,b)** Nuclear locations of the four bundles of body wall muscle cells (BWMDL, BWMDR, BWML and BWMLR) and most of the trunk cells including dorsal and ventral intestinal cells (InD and InV), trunk hypodermal cells (hyp7), H, V, T and P cells as well as ventral motor neurons (DD, DA and DB), projected onto the AP-DV plane **(a)** and AP-LR plane **(b)**. For better visualization, we performed quadratic polynomial fitting for each cell type. **(c,d)** Nuclear locations of the 7 rings of pharyngeal muscle cells (pm) and the 3 rings of marginal cells (mc) in the head, projected onto AP-LR **(c)** and DV-LR **(d)** plane. On the AP-LR plane, vertical lines show the mean locations of each ring along the AP axis. On the DV-LR plane, nuclei of the same ring are connected in lines. In **d**, pm2, pm3, pm1, mc1 and pm4 rings are shown on the left; mc2, pm5, pm6, pm7 and mc3 rings are shown on the right.

To determine the minimum number of stacks needed to build the atlas, we tested how the statistics of nucleus positions changed as the number of stacks increased. For this purpose, we randomly chose K stacks, with K ranging from 5 to 40, and

computed the average standard deviation of nucleus positions along AP, DV and LR axes for each K value. To make the statistics independent of the stacks chosen, we repeated this process 200 times, each time with different subsets of the stacks and obtained an average curve (**Fig. 2b**). The average standard deviation of cell positions along AP, DV, and LR axes increased quickly with K and then tapered off with only an inconsequential and asymptotically limited increase after $K = 15$ stacks. This confirmed the stability of our computational approach and justified using 15 stacks for the atlas.

The atlas permits more sophisticated analysis on nuclear locations. For instance, we quantitatively modeled and visualized the spatial patterns of nuclei within and across different types of cells, such as the four bundles of body wall muscle cells (BWML, BWMLR, BWMDL and BWMDR), the intestinal cells (InD and InV), the hypodermal cell hyp7, the blast cells H, V and T, the P cells, and the ventral motor neurons DD, DA and DB, using quadratic polynomial curves through the mean cell nucleus positions (**Fig. 3a,b**). The spatial distributions of pharyngeal muscle cell and marginal cell nuclei in the head (**Fig. 3c,d**) form 7 rings of pharyngeal muscle cell nuclei and 3 rings of marginal cell nuclei projected onto the AP-LR plane. The spacing of nuclei within each ring was quite consistent with earlier qualitative descriptions of worm anatomy^{1,10} (see also <http://www.wormatlas.org/>).

We also analyzed the invariant spatial relationships between nuclei along the AP, DV and LR dimensions. For this purpose, we built a graph in which each nucleus was a vertex and there was a directed edge from nucleus u to nucleus v if nucleus u was always in front of nucleus v in the dimension under consideration (**Fig. 4** and **Supplementary Fig. 6**). Note that we also applied transitive reduction to the graphs, meaning that if nucleus u was always in front of nucleus v and nucleus v was always in front of nucleus w , then we removed the transitively inferable edge from nucleus u to w . Such graphs show the

of the squared differences of the 357 or so nuclei centers between C and every other stack was minimal. We then used C as the reference stack to map every other stack to it via an affine transform. The final atlas is the ensemble of the transformed nuclei positions for each named nucleus in the coordinate system of the reference stack (**Supplementary Video 2**).

Statistical analysis of the atlas

We analyzed the mean and standard deviation of the center positions of each nucleus along the anterior-posterior (AP; **Fig. 2a**), dorsal-ventral (DV; **Supplementary Fig. 5a**) and left-right (LR; **Supplementary Fig. 5b**) axes (**Supplementary Table 1**). The distribution of standard deviation along the AP axis showed that 77% of the cells had standard deviations of cell positions less than 2 μm (**Fig. 2a**). The average standard deviation of the position of cell nuclei along the AP axis was 1.87 μm, which is about 72% of the average nuclear diameter. This recapitulates that cell nuclei in the L1 stage have positions that are highly stereotyped and provides a quantitative estimate of the stereotypy of each cell. Note that this estimate is an upper bound as some of the observed variations may be due to imperfect staging, straightening, registration and so on.

Several cell nuclei had a standard deviation of their position that was more than twice that of other nuclei along a given axis (see **Supplementary Note** for a list of these cells). These hyper-variable nuclei included the nuclei of hypodermal cell hyp7, the intestinal nuclei, and the hermaphrodite-specific neuron and coelomocyte cell nuclei, among others. hyp7 is a large syncytium with 23 nuclei that are free to move within the cell relative to each other. The number of intestinal cells varies between 19 and 21 in individual worms, and the location of these cells is variable depending on cell number. Finally, the hermaphrodite-specific neuron and coelomocyte cells start to migrate shortly after hatching. So the results from the atlas agree with known biology.

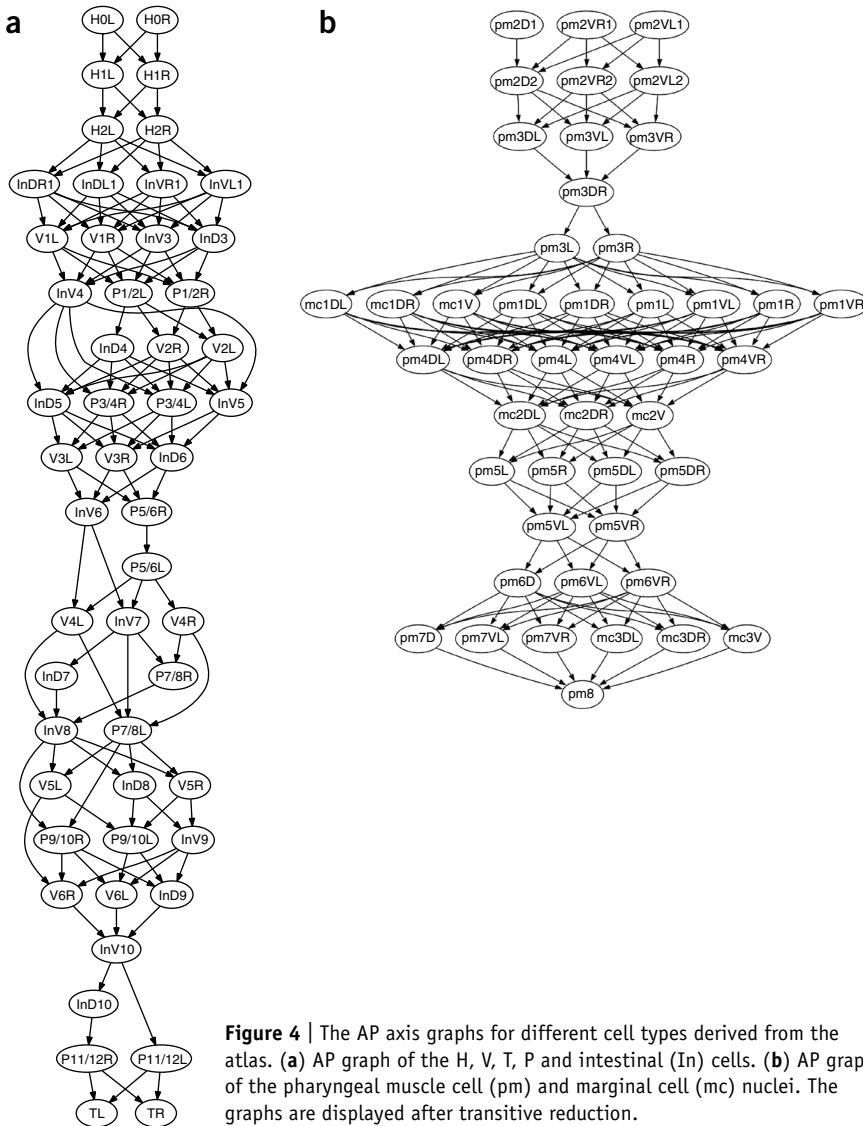


Figure 4 | The AP axis graphs for different cell types derived from the atlas. **(a)** AP graph of the H, V, T, P and intestinal (In) cells. **(b)** AP graph of the pharyngeal muscle cell (pm) and marginal cell (mc) nuclei. The graphs are displayed after transitive reduction.

3.23 μm , which could be used to identify them. Additionally, *hyp7* and V nuclei were also larger, with average diameters of 2.88 μm and 2.90 μm , respectively. None of the nuclei were considerably smaller than average.

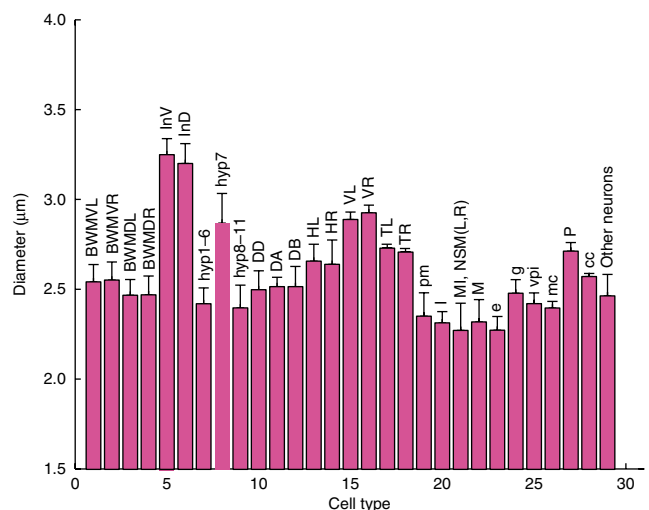
Automated annotation of nuclei

With an atlas in hand, we developed an automated approach that replaced the manual annotation of cell identities in the analysis of potentially thousands of stacks of worm lines, each expressing mCherry from a different target gene's promoter. The motivation was to permit medium- to high-throughput analysis of cellular information at single-cell resolution. The procedure for a newly acquired image stack was (i) automated straightening; (ii) automated segmentation of the DAPI, GFP and mCherry channels; (iii) optional manual curation of the segmentation; (iv) optional manual preannotation of 'problematic' cell nuclei; (v) automated registration and annotation of the atlas cell nuclei; (vi) optional curation of the annotation; and (vii) automated extraction of cellular information for each cell (for example, expression in the red fluorescence (mCherry) channel). A Matlab implementation of all the automated steps took less than 1 h to process one stack with a 2.3 GHz central processing unit. Our pipeline implemented in C is under development and preliminary timings indicate the computation speed will be about 10 times faster. For the manual steps, it takes about 1 h to curate the segmentation of one stack and another two hours to curate the annotation. In contrast, it takes about 3 d to process one stack completely manually.

statistically verified invariance of the relative cell position, especially among cells of particular tissues in the body plan, and can only be built given an atlas constructed from many worm observations.

In addition to recording the position of each nucleus, we also estimated the volume and diameter of every nucleus in the atlas (**Fig. 5**). The average diameter of a nucleus at the L1 stage was 2.58 μm , but intestinal cells were much bigger, with an average diameter of

Figure 5 | Nuclear sizes for different cell types. BWMVL and BWMVR, body wall muscle ventral left and right bundles; BWMDL and BWMDR, body wall muscle dorsal left and right bundles; InV and InD, intestine ventral and dorsal cells; *hyp*, hypodermal cells; DD, DA and DB, ventral motor neurons; HL and HR, H cell left and right bundles; VL and VR, V cell left and right bundles; TL and TR, T cell left and right; pm, pharyngeal muscle; I, pharyngeal motor neurons; MI and NSM(L,R), pharyngeal motor interneuron and secretory motor neuron; M, pharyngeal motor neuron; e, pharyngeal epithelial cells; g, pharyngeal gland cells; vpi, pharyngeal intestinal valve cells; mc, marginal cells in pharynx; P, P cells; cc, coelomocyte; and other neurons include BDU(L,R), ALM(L,R), CAN(L,R), Q(L,R), AVG, SABD, SABV(L,R), RIG(L,R), RIF(L,R), PVT, PVP(L,R), PVQ(L,R), PHA(L,R), PHB(L,R), LUA(L,R), PVC(L,R), ALN(L,R), PHsh(L,R), PLM(L,R), PVR, DVA and DVC.



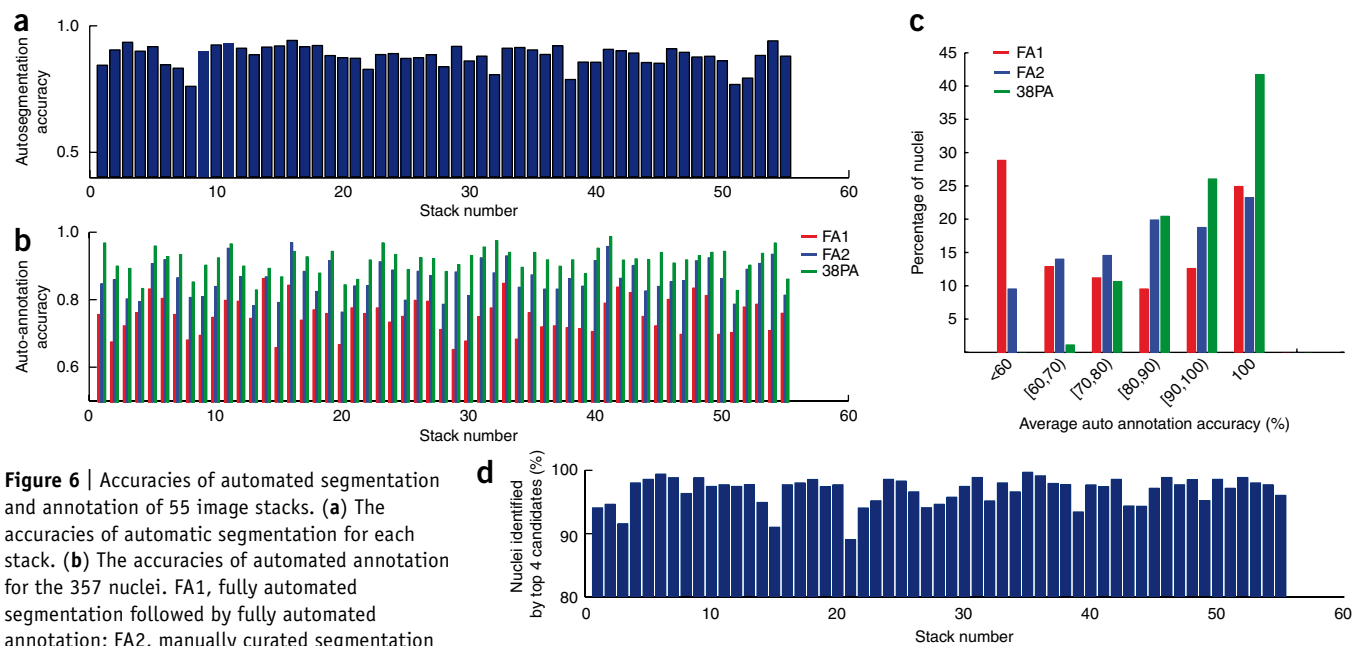


Figure 6 | Accuracies of automated segmentation and annotation of 55 image stacks. **(a)** The accuracies of automatic segmentation for each stack. **(b)** The accuracies of automated annotation for the 357 nuclei. FA1, fully automated segmentation followed by fully automated annotation; FA2, manually curated segmentation followed by fully automated annotation; and 38PA, manually curated segmentation followed by automated annotation of 319 nuclei in each stack with the remaining 38 nuclei preannotated manually. **(c)** Percentages of nuclei falling into different annotation accuracy ranges for fully automatic annotation of 357 nuclei (FA1 and FA2) and for automatic annotation of 319 nuclei (38PA). '[', not less than; ')', less than. **(d)** Percentages of the 357 nuclei whose identities could be determined by the top 4 candidates for each stack.

The computational problem posed by step *v* is as follows. We are given the K (for example, $K = 15$) registered 'template' images of the atlas for which 357 nuclei have been annotated and a 'subject' image S in which approximately 558 nuclei have been segmented and have yet to be annotated. The problem is to 'match' the 357 nuclei in the atlas and a subset of the nuclei in S . First, we annotated the 82 'marker' nuclei (81 body wall muscle cells and 1 depressor cell) stained in the green fluorescence channel by simultaneous registration and matching using a random sample consensus (RANSAC)-like approach²⁰. More specifically, we registered S to the reference stack C through many trials. In each trial, we selected 4 pairs of non-coplanar corresponding marker nuclei centers and computed an affine transform that maps S to C . We then used a bipartite matching algorithm²¹ to find the best match between the 82 marker nuclei of C and S under the given transform, that is, that minimizes the sum of the Euclidean distance between corresponding nucleus centers. The trial with the smallest distance produced the best annotation of the marker nuclei. Second, we took the affine transformation T_{GFP} that minimized the difference between S and C with respect to all 82 GFP-labeled nuclei and used the now-annotated marker nuclei in S to triangulate the remaining cell nuclei in the DAPI channel (Online Methods). We computed the normalized distance between a nonmarker nucleus to be annotated and its nearest marker nuclei (one anterior and one posterior nucleus in each of the four muscle bundles) and used this metric to find the best bipartite matching between nuclei in S and a subset of those in the atlas (Online Methods). We then examined the AP, DV and LR relationship in our initial labeling of S to see whether there was any assignment that seemed to substantially conflict with the invariants of the AP, DV and LR graphs derived from the atlas²². If so, we assumed the most conflicted assignment was erroneous and

reran the bipartite matching above but this time prohibiting the conflicting match. We iterated this process until the level of conflict could not be improved.

We tested the automated annotation on a new set of 55 confocal image stacks, which we manually annotated as well to assess the accuracy of the automated approach. First, we segmented the ~558 cells in each stack using our automated approach. The average accuracy of the automated segmentation was 89% (Fig. 6a). Errors mainly occurred in the head, in which nuclear density was very high and the axial resolution was not high enough to resolve the boundary between neighboring nuclei. Among the 55 stacks, 6 of them had notably lower signal-to-noise ratio and segmentation accuracies were lower.

We then ran automated annotation without any intervening manual curation of the segmentation. As errors in automated segmentation tend to induce errors in the ensuing automated annotation, we also manually corrected the segmentation errors in these stacks using VANO and then applied automated annotation. The average annotation accuracy in this case was 86% for all 357 cells in all 55 stacks (Fig. 6b). We found that the nuclei of body wall muscle cells, P cells, H, V and T cells, intestinal cells and most of the ventral motor neurons (D) were annotated with greater than 80% accuracy. There were 38 cell nuclei whose annotation accuracies were lower than 60%. Some were in the pharynx, where cell density is very high, and others had variable positions. If with the help of additional information such as cell morphology and size, we pre-annotated these 38 cell nuclei, then the automated annotation on the remaining 319 cells in step *v* became more accurate (Fig. 6b). Overall, the average accuracy improved to 92%. The percentages of cells that had higher annotation accuracy also appreciably improved (Fig. 6c).

Another way to use the automated annotator is to provide for each nucleus s a small list of k candidate identities sorted

according to their likelihood for which we used $w(s,t)$ (the score of matching nucleus t in the templates to nucleus s in S ; Online Methods) as a proxy. This presents to the user the nuclei in the atlas giving the top k candidates for each s in order of score. We considered such a list to be informative if the correct answer was in the list. We identified the percentage of lists that we deemed informative for each of the 55 stacks when $k = 4$ candidates (Fig. 6d). The average rate was 97%; in other words, the top four candidates determined by our approach included the correct identity of almost every nucleus.

DISCUSSION

The quantitative stereotypy of nuclear locations and the invariance of their relative locations in this atlas can be used to examine hypotheses about the anatomy of a worm. In our view, the most important application of the atlas is that it allows us to automatically process a new image stack and identify cells in the lineage of an L1 without human intervention, with sufficient accuracy to consider high-throughput studies. This cannot be achieved without a digital nuclear atlas and the automated nuclear annotation approach.

We expect the atlas can be used in many additional applications. It may be used to measure gene expression patterns at single-cell resolution and to quantitatively characterize the molecular expression signature of each cell. It may help detect mutants by identifying nuclear locations and gene expression that differ from those of wild-type *C. elegans*. It may be used to automatically differentiate hermaphrodites from male worms. As our software becomes faster and accuracy improves, we could place the software onboard a microscope and automatically direct the laser ablation or stimulation of channelrhodopsin- or halorhodopsin-expressing cells. This could allow high-throughput studies of worm cell function.

An important limitation is that currently, the L1 atlas includes only 357 nuclei. The remaining nuclei are mostly neurons in the nerve ring, and a small number of hypodermal cells, arcade cells and socket cells in the head. The nuclei are very dense in this region of the worm, presenting difficulties for manual annotation. Additional fluorescent fiducial markers, analogous to the GFP labeling of the body-wall muscle cells, could help us to resolve identities of nuclei in this difficult region and provide training data for automated segmentation and annotation. Another difficulty in resolving nuclei in this region is that standard confocal microscopy does not have sufficiently high resolution. It is our intuitive estimate that a factor of two or greater improvement in resolution would be sufficient and there is some hope that stimulated emission depletion microscopy and selective plane illumination microscopy, or an optimally cleared worm preparation would suffice. Once we resolve the identities of these cells, our computational pipeline can be directly applied to generate the complete atlas.

Our methodology can be applied directly to other developmental stages, and this work is currently in progress. From an algorithmic point of view, we developed a system that can identify objects whose positions are stereotypic given an atlas that is in effect a set of registered and labeled training data. Given that many early developmental body plans are stereotypic, we expect that these methods may be useful in gene expression studies of development.

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/naturemethods/>.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank A. Fire for providing reagents and advice. The work of X.L. and S.K.K. was funded by the Ellison Medical Foundation and the US National Institutes of Health. X.L. was also funded by the Larry L. Hillblom Foundation. The work of F.L., H.P. and E.M. was funded by Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

F.L., H.P. and E.M. developed the pipeline and related algorithms, built the atlas, performed the statistical analysis and wrote the manuscript. X.L. and S.K.K. prepared the worm assays and images, manually annotated the nuclei and helped prepare the manuscript.

Published online at <http://www.nature.com/naturemethods/>.

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>.

- Riddle, D.L., Blumenthal, T., Meyer, B.J. & Priess, J.R. *C. elegans* II. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA, 1997).
- Sulston, J.E., Schierenberg, E., White, J.G. & Thomson, J.N. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64–119 (1983).
- Sulston, J.E. & Horvitz, H.R. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.* **56**, 110–156 (1977).
- White, J.G., Southgate, E., Thomson, J.N. & Brenner, S. The structure of the nervous system of the nematode *C. elegans*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **314**, 1–340 (1986).
- Chen, B.L., Hall, D.H. & Chklovskii, D.B. Wiring optimization can relate neuronal structure and function. *Proc. Natl. Acad. Sci. USA* **103**, 4723–4728 (2006).
- Stein, L.D. *et al.* The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol.* **1**, E45 (2003).
- Reece-Hoyes, J.S. *et al.* A compendium of *Caenorhabditis elegans* regulatory transcription factors: a resource for mapping transcription regulatory networks. *Genome Biol.* **6**, R110 (2005).
- Kim, S. *et al.* A gene expression map for *C. elegans*. *Science* **293**, 2087–2092 (2001).
- Wang, J. & Kim, S. Global analysis of dauer gene expression in *Caenorhabditis elegans*. *Development* **130**, 1621–1634 (2003).
- Hall, D.H. & Altun, Z.F. *C. elegans* atlas. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA, 2007).
- Bao, Z. *et al.* Automated cell lineage tracing in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **103**, 2707–2712 (2006).
- Murray, J.I. *et al.* Automated analysis of embryonic gene expression with cellular resolution in *C. elegans*. *Nat. Methods* **5**, 703–709 (2008).
- Pearson, B.J. & Doe, C.Q. Regulation of neuroblast competence in *Drosophila*. *Nature* **425**, 624–628 (2003).
- Peng, H., Long, F., Liu, X., Kim, S. & Myers, E. Straightening *Caenorhabditis elegans* images. *Bioinformatics* **24**, 234–242 (2008).
- Vincent, L. & Soille, P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.* **13**, 583–598 (1991).
- Beucher, S. & Meyer, F. The morphological approach to segmentation: the watershed transformation. in *Mathematical Morphology in Image Processing* (ed., Dougherty, E.R.) 433–482 (Marcel Dekker, New York, 1993).
- Vapnik, V. *The Nature of Statistical Learning Theory* (Springer-Verlag, New York, 1995).
- Peng, H., Long, F. & Ding, C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**, 1226–1238 (2005).
- Peng, H., Long, F. & Myers, G. VANO: a volume-object image annotation system. *Bioinformatics* **25**, 695–697 (2009).
- Fischler, M.A. & Bolles, R.C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **24**, 381–395 (1981).
- Cormen, T.H., Leiserson, C.E., Rivest, R.L. & Stein, C. *Introduction to algorithms* 2nd edn. (MIT Press and McGraw-Hill, Cambridge, Massachusetts, USA, 2001).
- Long, F., Peng, H., Liu, X., Kim, S. & Myers, E. Automatic recognition of cells (ARC) for 3D images of *C. elegans*. *Lect. Notes Comput. Sci.* **4955**, 128–139 (2008).

ONLINE METHODS

Manual annotation. To build the atlas, we manually annotated nuclear identities based on the morphology and relative spatial positions of cells in *C. elegans* qualitatively described in earlier literatures and the WormAtlas website (<http://www.wormatlas.org>)¹⁰ (**Supplementary Note**). We assessed the accuracy of our manual annotation and further improved it using three parallel approaches. In the first approach, we manually annotated three stacks, each twice independently. We found that over 98.5% nuclei were assigned the same names in independent annotation trials on different days. In the second approach, we annotated some worms carrying mCherry reporters whose expression has been well studied, including reporters for *elt-7*, *pal-1*, *cnd-1*, *die-1* and *pha-4*. The expression patterns of these reporters based on our annotation were consistent with their cell- and tissue-specific patterns reported in previous literature. In the third approach, we computed the standard deviation of the positions of nuclei annotated with the same identities across different image stacks. We then pinpointed those outliers with large standard deviation (bigger than twice of the standard deviation of other nuclei along a given axis) and corrected potential annotation errors if there were any. Note that position-variable cells were identified using other cues such as their morphology, size, and locations relative to marker nuclei. This bootstrapping strategy was repeated until we were highly confident that the identities of nuclei were correct. Note that since the atlas was built on the statistical analysis of multiple 3D worm images, it is robust to potential annotation errors if there were any.

Matching score in automatic annotation of nuclei in the DAPI channel. For each nonmarker nucleus t in the DAPI channel, we found the posterior-most marker nucleus in each body wall muscle cell bundle that is anterior to t in all templates, and the anterior-most such marker that is posterior to t in all templates, if they exist. We called this set B_t or the AP-axis ‘bracketing markers’

for t . For each bracketing marker b , we computed the mean $\mu_{b,m}(t)$ and standard deviation $\sigma_{b,m}(t)$ over the K stacks of its distance to nucleus t along each dimension $m \in \{\text{AP, DV, LR}\}$. The score of matching nucleus t in the templates to nucleus s in S , $w(s,t)$ is the average number of standard deviation $\sigma_{b,m}(t)$ that the distance, $d_{b,m}(s)$, between s and each bracket marker b (in S) differs from the mean distance $\mu_{b,m}(t)$ for t , that is,

$$w(s,t) = \frac{1}{3|B_t|} \sum_{b,m} \frac{|d_{b,m}(s) - \mu_{b,m}(t)|}{\sigma_{b,m}(t)}.$$

This criterion is used to find the best bipartite matching between the nuclei in S and a subset of those in the atlas and serves as an initial annotation of the nonmarker nuclei.

Other computational methods. Descriptions of worm body backbone detection, hollow-shaped nuclei pattern filling, adaptive thresholding, watershed algorithm, and region merging/splitting for nuclei segmentation, VANO interface, affine transform in building the atlas, details on computing AP, DV and LR dimension graphs and adding spatial constraints to automatic nuclei annotation are available in the **Supplementary Note**.

Data and software. The 3D digital atlas of the 357 nuclei is available in two forms. A list of the mean and standard deviation of the position of each nucleus is available in **Supplementary Table 1**. A ‘point-cloud’ of the atlas (**Supplementary Data 1**) can be displayed and queried using the software V3D (H.P., unpublished data) we developed for microscopy image data processing and visualization (**Supplementary Video 3**). Both V3D and the annotation tool VANO are available as a free download at <http://penglab.janelia.org/proj/v3d> and <http://penglab.janelia.org/proj/vano>. The Matlab code of this pipeline, called CellExplorer, along with a sample dataset, is available as **Supplementary Data 2**.