6-2015

WormGUIDES: An Interactive Single Cell Developmental Atlas and Tool for Collaborative Multidimensional Data Exploration.

James C. Schaff
University of Connecticut School of Medicine and Dentistry

William A. Mohler
University of Connecticut School of Medicine and Dentistry

Follow this and additional works at: https://opencommons.uconn.edu/uchcre_articles
Part of the Life Sciences Commons

Recommended Citation
Schaff, James C. and Mohler, William A., "WormGUIDES: An Interactive Single Cell Developmental Atlas and Tool for Collaborative Multidimensional Data Exploration." (2015). UCHC Articles - Research. 290.
https://opencommons.uconn.edu/uchcre_articles/290
WormGUIDES: an interactive single cell developmental atlas and tool for collaborative multidimensional data exploration

Anthony Santella1†, Raúl Catena1,2†, Ismar Kovacevic1, Pavak Shah1, Zidong Yu1,3, Javier Marquina-Solis4, Abhishek Kumar4,5, Yicong Wu5, James Schaff7, Daniel Colón-Ramos4, Hari Shroff5, William A. Mohler6,7* and Zhirong Bao1*

Abstract

Background: Imaging and image analysis advances are yielding increasingly complete and complicated records of cellular events in tissues and whole embryos. The ability to follow hundreds to thousands of cells at the individual level demands a spatio-temporal data infrastructure: tools to assemble and collate knowledge about development spatially in a manner analogous to geographic information systems (GIS). Just as GIS indexes items or events based on their spatio-temporal or 4D location on the Earth these tools would organize knowledge based on location within the tissues or embryos. Developmental processes are highly context-specific, but the complexity of the 4D environment in which they unfold is a barrier to assembling an understanding of any particular process from diverse sources of information. In the same way that GIS aids the understanding and use of geo-located large data sets, software can, with a proper frame of reference, allow large biological data sets to be understood spatially. Intuitive tools are needed to navigate the spatial structure of complex tissue, collate large data sets and existing knowledge with this spatial structure and help users derive hypotheses about developmental mechanisms.

Results: Toward this goal we have developed WormGUIDES, a mobile application that presents a 4D developmental atlas for Caenorhabditis elegans. The WormGUIDES mobile app enables users to navigate a 3D model depicting the nuclear positions of all cells in the developing embryo. The identity of each cell can be queried with a tap, and community databases searched for available information about that cell. Information about ancestry, fate and gene expression can be used to label cells and craft customized visualizations that highlight cells as potential players in an event of interest. Scenes are easily saved, shared and published to other WormGUIDES users. The mobile app is available for Android and iOS platforms.

(Continued on next page)
Conclusion: WormGUIDES provides an important tool for examining developmental processes and developing mechanistic hypotheses about their control. Critically, it provides the typical end user with an intuitive interface for developing and sharing custom visualizations of developmental processes. Equally important, because users can select cells based on their position and search for information about them, the app also serves as a spatially organized index into the large body of knowledge available to the C. elegans community online. Moreover, the app can be used to create and publish the result of exploration: interactive content that brings other researchers and students directly to the spatio-temporal point of insight. Ultimately the app will incorporate a detailed time lapse record of cell shape, beginning with neurons. This will add the key ability to navigate and understand the developmental events that result in the coordinated and precise emergence of anatomy, particularly the wiring of the nervous system.

Keywords: Visualization, C. elegans, Morphogenesis, Neurons, Single-cell analysis

Background

Biologists must increasingly analyze and interpret the behavior of single cells in complex, spatially variant tissue. Microscopy improvements now make it possible to produce highly detailed records of complex structures as they develop. It is possible to perform in toto imaging of metazoan development [1–4] with the resolution necessary to follow individual cells. Increasingly sophisticated stem cell and tissue engineering applications also motivate a need to follow cell movements and tissue morphogenesis at the single cell level. Image analysis algorithm development supports these efforts by allowing increasingly automated distillation of these 4-dimensional datasets into curated records of tracked cells [4–7]. However, manually navigating hundreds of cell tracks over hours is a daunting task. Developmental events often involve large numbers of cells that are inherently embedded in a complex spatio-temporal context. Meaningful information is only implicit in the undifferentiated mass of tracks produced by cell tracking. Software tools have been developed to aid in viewing images along with annotation in order to validate or edit the results of image analysis [8–11]. There remains however, a need for tools focused on navigating this wealth of cell identity and position information in order to see and understand events of interest amidst the clutter of complex unrelated events. In addition, spatio-temporal data about cells does not exist in a void. There is extensive prior knowledge available about developmental events and gene expression that can be better understood when placed in a spatio-temporal context. Systematic analysis of developmental regulation [12] adds to the volume and complexity of this data. Anatomical atlases in adult C. elegans [13, 14] and brain atlases in other models e.g., [15, 16] lack temporal information and are usually focused on presenting a single self-contained set of data. Tools are needed to enable the exploration of complex 4D records of multicellular systems alongside contextual data. A spatial data infrastructure for complex tissues would provide a systematic method of exploring and collating developmental information in a unified framework, analogous to the role of GIS and associated data infrastructure in revolutionizing access to and the use of geographical information.

C. elegans is an approachable point of entry in attacking these problems of data visualization and interpretation. Popular because of its simplicity, C. elegans has an invariant lineage; in each embryo the same number of cells are born and assume the same functions. In every embryo the same cells play out tightly conserved movements and shape changes over the eight hours between fertilization and when the embryo begins to move within the egg shell [17–19]. At around 7 h, the comma stage, elongation begins and a tail becomes visible. Uniquely, the adult neural structure has been mapped out at the synaptic level [20], meaning the end-state of neural developmental processes is known. The 670 embryonic cell divisions, 113 programmed cell deaths, and 302 neurons that go into an adult worm provide a fixed landscape on which developmental data can be superimposed. In addition, since an embryonic cell exists over a relatively short period of time, embryonic cell names provide an approximate 4D location, making it possible to connect existing and new knowledge to the spatio-temporal context of that individual cell.

The goal of WormGUIDES is to enable the visual exploration of this information by providing an interface for examining single-cell level records of development. The invariant lineage of C. elegans allows these records to become a spatial index into community knowledge bases. Detailed positional information is thus linked to community knowledge: gene expression, web pages and databases. This portal creates an easy-to-navigate and more unified view of development. Although the adult neural structure of C. elegans is known, and full time-lapse records of cell positions are available, much less is understood about how developing neurons coordinate position and growth to generate adult neural structures. The expectation is that easy visualization, customization, and spatial search will enable a deeper understanding of
Implementation

The WormGUIDES mobile app was developed to address the need for data exploration in 4D datasets. Developed in Java for Android and Objective-C for iOS, WormGUIDES is open source, with the source code available at http://www.wormguides.org/open-source-software (Android) and https://bitbucket.org/raulcatena/iworm/ (iOS). For the end user, the app and underlying data can be downloaded and installed for free through the Google Play Store and Apple App Store.

Figure 1 outlines the information collected into WormGUIDES. The core of WormGUIDES is a record of the location of every cell at every minute during embryonic development. This reference model has been built from the measured nuclear positions of every cell in an individual hemaphrodite embryo between the four cell and comma stage with 599 tracked cells (including dead cells that remain visible) in the last frame. Cell tracking was performed with Starrynite [5, 21]. Cell tracking results were edited and validated via systematic manual comparison with images, and automated comparison of cell positions measured in 2 additional embryos to flag positional outliers as potential errors for manual inspection.

This detailed record of development is complemented by extensive pre-existing knowledge regarding fate and gene expression. A wide range of information is available online and through C. elegans specific websites. Embryonic cell names provide a way to spatially index this information and link to it from the WormGUIDES app. Key information about cell fate and gene expression is integrated into the app directly. The fate of each cell in

![Fig. 1 WormGUIDES vision and input data. The goal of WormGUIDES is to enable novel visual exploration of development by providing an intuitive interface for exploring single cell level records of embryonic development in 4D. WormGUIDES also provides a spatial index into community knowledge bases by enabling users to quickly search for information on cells they have identified via 4D exploration. Detailed cell positions and (in future releases) cell morphology are combined with available knowledge to create a convenient way to navigate views of development. The expectation is that this ability to understand cells in the context of the embryo will enable a deeper understanding of how complex processes, particularly neural development, unfold.](image-url)
the C. elegans embryo is known, and this knowledge is summarized in the standard parts list [22]. Partial single cell level expression information is available for many genes through the WormBase [23] website. Gene expression information is an ideal example of information that, although inherently spatio-temporal, is typically not presented that way. Rather, it has been available as curated lists of cell names with known expression. WormGUIDES contains the detailed positions of these cells over their lifetimes and this information has been integrated to enable the exploration of the spatio-temporal distribution of expression. Valuable insights into the control of a process can be gained by cross referencing expression with the stages of the process, allowing users to easily screen known molecular actors that might contribute to the event of interest.

The app presents nucleus position and identity over time in a multi-touch interface (Fig. 2a). Developmental time is controlled with a scroll bar, while spatial navigation is controlled by standard dragging and pinching gestures. When a feature of interest is identified, surrounding cells can be queried for their identity. Tapping on any cell displays its systematic name in a pop up dialog. For terminal cells, a functional name and parts list description are also displayed (Fig. 2b). Options in the pop up allow several community specific websites to be queried as well as general web searches to be performed for the selected cell’s name. This allows the app to serve as an aggregate spatial index of existing knowledge about single-cells during C. elegans development.

A key part of the app interface is the search panel (Fig. 2c). The goal of this panel is to allow easy composition of complex visualizations that highlight specific subsets of cells. As an example, the color scheme in Fig. 2a colors terminal cells based on their fate. Colors follow the tissue color conventions of WormAtlas [24], a community anatomy knowledge base. Sliders explicitly control the type of search executed: systematic names, functional cell names, parts list descriptions or gene expression search. Labels can be propagated to the ancestors and descendants of these cells as well. The resulting search applies a new color to the cells that match the search terms. Multiple searches can be layered to highlight these cellular features within the global context of the entire embryo.

The scene sharing menu is designed to facilitate communication between researchers (Fig. 2d). Once a custom visualization has been created a user can share it with others. This can be done by sending a screen shot by email, or saving the screen shot as a figure for reference. Alternatively, and most powerfully, the WormGUIDES scene configuration can be shared as a URL that encodes the coloring and camera position in the scene. This allows other app users to generate their own fully configurable instance of the original visualization within the WormGUIDES app.

Results and Discussion
We demonstrate the use of the WormGUIDES app by creating visualizations highlighting the key features of the app and their ability to highlight important biological structures. The default color scheme in the app is lineage based (Fig. 3a). Each of six founder cells is assigned a different color which is applied to all its descendants, thus coloring the embryo based on clonal descent. Searching based on cell fate simplifies highlighting key structures and their arrangement at particular developmental stages. Coloring a set of cells with a given fate allows the arrangement and relative positions of these cells to be followed over time. Neuronal subtypes and ancestors are highlighted in Fig. 3b, motor neurons in green, sheath and socket cells in red, interneurons in blue and sensory neurons in yellow. Each category was highlighted by performing a keyword search against part list descriptions. The cells at comma stage have a distinct spatial arrangement with a clear correspondence to the final structures, such as the ventral cord (central posterior cluster in green); about two hours earlier the structures are significantly less distinct (Fig. 3c). Neighboring cells, potential players in events, can be queried for their names with a tap.

Gene expression (Fig. 3d) based coloring is a key instance of integrating external information into a unified spatial framework. When a gene search is executed, a query is made for the WormBase page for that gene and the expression field on the page is parsed for a list of embryonic cell names. These cells are used as the search result. Results often give a useful sense of early embryonic expression. The results of a pha-4 search, for example, appropriately color the pharynx primordium and the gut cells. Drawing on live data means results will inherently reflect updates as more systematic gene expression information [25] is integrated into WormBase. This transplantation of expression information back into a spatial context is possible because of the fixed spatial relationships between cells in C. elegans, and provides a detailed, interactive visualization that would be all but impossible without the benefit of WormGUIDES.

Each search result can be seen as a layer of information visualized as colors. In general layers from different results may overlap. This is indicated in the app by striping nuclei with all colors that apply to that cell. Overlapping search results in Fig. 3e illustrate striping. Hypodermal cells are colored with three distinct search methods: part list description search for the term ‘hypodermal’, systematic name search for 3 sub-lineages that are heavily hypodermal and gene search for nhr-25 associated with hypodermal fate. Overlap, or its lack, between these searches can be easily observed and used to guide investigation of individual cell identities.
Future work

Significant additions to WormGUIDES functionality are planned. A key priority is a desktop version of WormGUIDES. This will allow browsing of raw image data as well as cellular morphology and take advantage of additional screen space to provide more sophisticated tools for data exploration. Additional windows and data management tools will make it possible to...
view and manipulate online search results and use them to control visualization in more flexible ways. Another key feature of the desktop version will be the ability to add and navigate arbitrary user generated annotations of events or structures either manually or by importing data tagged with cell names or spatio-temporal positions.

The gradual addition of cell morphology data to WormGUIDES a major focus of ongoing efforts. Fig. 4 briefly presents the strategy for characterizing cell morphology and some preliminary results toward this goal. In the short term, the morphology of neuronal cells is our focus, though our approach, and ultimate ambition, is general. Our strategy (Fig. 4a) is to analyze embryos from a number of strains, each of which uses promoters to label different subsets of neurons. This sparse labeling aids clear time-lapse imaging, and greatly simplifies the segmentation of cell shape. Nuclear positions are tracked in each embryo and these are used to align results to a unified coordinate system allowing a synthesized atlas of neuronal shape to be created.

Figure 4b,c demonstrates this process for one strain labeled with *lim-4*:GFP (mglsl19[lim-4::gfp, pRF4]; ujls113 [Ppie-1::H2B::mCherry, Pnhr-2::HIS-24::mCherry, unc-119(+)]). At the comma stage *lim-4*:GFP labels at least 8 pairs of left-right symmetric cells that are part of the nerve ring. Cell identities (Fig. 4b) were established by lineaging using Starrynite. Time lapse data of cell bodies were acquired with a Dual-view Inverted Selective Plane Illumination Microscopy (diSPIM) system [26, 27] and segmented semi-automatically in Imaris based on image intensity. *lim-4*:GFP expressing cells were tracked, temporal alignment was manually assigned and cell positions in each frame were used to compute a best fit (rigid plus scaling) transformation between the *lim-4*:GFP embryo...
and WormGUIDES nuclear positions. Fig. 4c shows the lim-4::GFP cells superimposed on the WormGUIDES embryo in a series of frames covering 40 min of development before twitching commences. Nuclei are rendered as small semi-transparent spheres to avoid obscuring the cell bodies. Over the sequence of frames it is possible to see the bundle of processes that make up the nerve ring extending dorsally.

As the lim-4::GFP proof of concept data suggests, neural cell morphology integrated into the app will add an extra dimension of information, highlighting when changes in morphology occur, and the sequence of interactions that assemble cells into the structures revealed by the adult wiring diagram. Ongoing work will scale up the processing of neuronal markers, including automating membrane marker segmentation [28], and finalize algorithmic details of assembling these data into a unified atlas.

Alignment of data from many embryos will require a more systematic understanding of variability in cell position over time. Previous systematic study of the embryo up to the 350 cell stage suggests it is reasonable to represent nuclear position as unimodal, with limited variability [29]. Collection of additional late stage data to assess variability is in progress. Ultimately, cell positions within the app that are based on a single embryo may be replaced by an aggregate consensus model. An explicit representation of observed variability is a likely feature of the desktop atlas. Eventual analysis and integration of a male worm is also possible.

Additional alignment issues will need to be addressed to characterize post-twitching development, which can
be imaged using diSPIM but will require the development of methods for straightening and aligning twisted embryos. The tightly packed elongated embryo presents a more difficult problem than the normal curved pose of the adult worm [30]. Methods are under development to straighten the embryo using a combination of junctional and nuclear markers to semi-automatically extract a 3D model of the worm’s body in each frame and unwrap it into a consistent straightened coordinate system [31].

Conclusions
The WormGUIDES app provides users with the ability to explore and navigate developmental events in their spatiotemporal context and cross reference these events against the accumulated knowledge amassed by the worm community. By simplifying exploration of the developing embryo, WormGUIDES should make it easier to understand the coordinated motions of cells as they unfold in a spatiotemporal context. This in turn should make it easier to correlate these events with gene expression and other information in order to develop testable hypotheses about mechanistic control of development. The visualization customization interface within the app allows any user to investigate embryonic events they care about, regardless of what those events are. As it stands, the WormGUIDES app provides a useful tool for navigating development in complex tissue and examining the coordinated behavior of cells over time. We hope it will find use as both a reference for the study of C. elegans development and as an educational tool. These applications are not unique to the worm community. The kinds of knowledge available to the worm community are becoming increasingly standard as the availability of genomic and other big data increases for other systems. Spatially localizing this information will become possible in other models as paradigms develop for handling cell-level correspondence between individuals in organisms with variable cell lineages [32]. The fundamental types of data, interactive tasks and challenges of alignment to reference data addressed by WormGUIDES are universal and our solutions should be applicable to other organisms.

Availability and requirements
Project name: WormGUIDES
Project home page: http://www.wormguides.org/ Operating systems: Android, iOS- Programming languages: Java, Objective-C License: GNU GPL

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
DCR HS ZB and WAM conceived WormGUIDES and contributed to the design of app features. AS, RC, WAM designed the app interface. RC and WAM implemented the iOS and Android apps respectively. JS contributed to rendering embryos on the android system. JMS provided testing and input into the app design. AK, YW, HS, IK and PS contributed to microscopy and strain selection. AS, ZY and IK contributed to lineage editing and quality control. All authors read and approved the manuscript.

Acknowledgments
Thanks to Amelia White, Zhuo Du, Thomas Flott, Teagan Marshall, Nhan Nguyen and Ryan Christensen for advice, imaging, segmentation and lineage editing assistance. Thanks to Oliver Hobert’s lab for providing worms with the irm-4::GFP transgene. We also thank the Research Center for Minority Institutions program and the Institute of Neurobiology of the University of Puerto Rico for providing a meeting and brainstorming platform. This work was partially conducted at the Marine Biological Laboratories at Woods Hole, under a Whitman research award (to D.A.C.-R, Z.B., W.A.M. and H.S.). This work was supported by the Intramural Research Programs of the NIH National Institute of Biomedical Imaging and Bioengineering and by NIH grants U01 HD075602 and R24OD016474.

Author details
1 Developmental Biology Program, Sloan-Kettering Institute, New York, NY, USA. 2 Present Address: Institute for Molecular Life Sciences, University of Zurich, Zürich, Switzerland. School of Energy and Power Engineering, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu 212003, China. 4 Program in Cellular Neuroscience, Neurodegeneration and Repair, Department of Cell Biology, Yale University School of Medicine, New Haven, CT, USA. 5 Section on High Resolution Optical Imaging, National Institute of Biomedical Imaging and Bioengineering, NIH, Bethesda, MD, USA. 6 Dept. of Genetics & Dev Biology, University of Connecticut Health Center, Farmington, CT, USA. 7 Center for Cell Analysis and Modeling, University of Connecticut Health Center, Farmington, CT, USA.

Received: 16 February 2015 Accepted: 23 May 2015

Published online: 09 June 2015

References
1. Amat F, Lemon W, Mossing DP, McDole K, Wan Y, Branson K, et al. Fast, accurate reconstruction of cell lineages from large-scale fluorescence microscopy data. Nat Methods. 2011;11(9):951–8.
2. Keller PJ, Schmidt AD, Wittbrodt J, Stelzer EH. Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. Science. 2008;322(5904):1065–9.
3. Wu Y, Ghitani A, Christensen R, Santella A, Du Z, Rondeau G, et al. Inverted selective plane illumination microscopy (SPIM) enables coupled cell identity lineageing and neurodevelopmental imaging in Caenorhabditis elegans. Proc Natl Acad Sci U S A. 2011;108(43):17708–13.
4. Giurumescu CA, Kang S, Planchon TA, Betzig E, Bloomer J, Chen Y, et al. Quantitative semi-automated analysis of morphogenesis with single-cell resolution in complex embryos. Development. 2012;139(22):4271–9.
5. Santella A, Du Z, Bao Z. A semi-local neighborhood-based framework for probabilistic cell lineage tracing. BMC Bioinformatics. 2014;15:217.
6. Mace DL, Weisdepp P, Gevitzman L, Boyle T, Waterston RH. A high-fidelity cell lineage tracing method for obtaining systematic spatiotemporal gene expression patterns in Caenorhabditis elegans. G3 (Bethesda). 2013;3(5):851–63.
7. Heid PJ, Voss E, Soll DR. 3D-DIAsemble: a computer-assisted system for reconstructing and motion analyzing in 4D every cell and nucleus in a developing embryo. Dev Biol. 2002;245(2):329–47.
8. Peng H, Ruan Z, Long F, Simpson JH, Myers EW. VSD enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets. Nat Biotechnol. 2010;28(4):348–53.
9. Henriksson J, Hench J, Tong YG, Johansson A, Johansson D, Burglin TR. Endrov: an integrated platform for image analysis. Nat Methods. 2013;10(6):454–6.
10. Boyle TJ, Bao Z, Murray JJ, Ayara CL, Waterston RH. AceTree: a tool for visual analysis of Caenorhabditis elegans embryogenesis. BMC Bioinformatics. 2006;7:275.
11. Saalfeld S, Cardona A, Hartenstein V, Tomancak P. CATMAD: collaborative annotation toolkit for massive amounts of image data. Bioinformatics. 2009;25(15):1984–6.
12. Du Z, Santella A, He F, Tiongson M, Bao Z. De novo inference of systems-level mechanistic models of development from live-imaging-based phenotype analysis. Cell. 2014;156(1–2):359–72.
13. Long F, Peng H, Liu X, Kim SK, Myers E. A 3D digital atlas of C. elegans and its application to single-cell analyses. Nat Methods. 2009;6(9):667–72.

14. Szegedi B, Gleeson P, Vella M, Khaynulin S, Palyanov A, Hokanson J, et al. OpenWorm: an open-science approach to modeling Caenorhabditis elegans. Front Comput Neurosci. 2014;8:137.

15. Oh SW, Harris JA, Ng L, Winslow B, Cain N, Mihalas S, et al. A mesoscale connectome of the mouse brain. Nature. 2014;508(7495):207–14.

16. Peng H, Chung P, Long F, Qu L, Jenett A, Seeds AM, et al. BrainAligner: 3D registration atlases of Drosophila brains. Nat Methods. 2011;8(6):493–500.

17. Chisholm AD, Hardin J. Epidermal morphogenesis, WormBook : the online review of C. elegans biology. 2005. p. 1–22.

18. Roh-Johnson M, Shemer G, Higgins CD, McClellan JH, Werts AD, Tulu US, et al. Triggering a cell shape change by exploiting preexisting actomyosin contractions. Science. 2012;335(6073):1232–5.

19. Poji C, Tiongson M, Moore JL, Santella A, Bao Z. Actomyosin-based self-organization of cell internalization during C. elegans gastrulation. BMC Biol. 2012;10:94.

20. White JG, Southgate E, Thomson JN, Brenner S. The structure of the nervous system of the nematode Caenorhabditis elegans. Philos Trans R Soc Lond Ser B Biol Sci. 1986;314(1165):1–340.

21. Bao Z, Murray JJ, Boyle T, Ooi SL, Sandel MJ, Waterston RH. Automated cell lineage tracing in Caenorhabditis elegans. Proc Natl Acad Sci U S A. 2006;103(8):2707–12.

22. Wood WB. The Nematode Caenorhabditis elegans. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory; 1988.

23. Harris TW, Baran J, Bieri T, Cabunoc A, Chan J, Chen WJ, et al. WormBase 2014: new views of curated biology. Nucleic Acids Res. 2014;42(Database issue):D789–93.

24. WormAtlas [http://www.wormatlas.org]

25. Moore JL, Du Z, Bao Z. Systematic quantification of developmental phenotypes at single-cell resolution during embryogenesis. Development. 2013;140(5):926–74.

26. Peng H, Long F, Liu X, Kim SK, Myers EW. Straightening Caenorhabditis elegans images. Bioinformatics. 2008;24(2):234–42.

27. Heckscher ES, Long F, Layden MJ, Chuang YH, Manning L, Richart J, et al. Atlas-builder software and the eNeuro atlas: resources for developmental biology and neuroscience. Development. 2014;141(12):2524–32.