Quantitative semi-automated analysis of morphogenesis with single-cell resolution in complex embryos

Claudiu A. Giurumescu1, Sukryool Kang2, Thomas A. Planchon3, Eric Betzig3, Joshua Bloomekatz1, Deborah Yelon1, Pamela Cosman2 and Andrew D. Chisholm1,*

SUMMARY
A quantitative understanding of tissue morphogenesis requires description of the movements of individual cells in space and over time. In transparent embryos, such as C. elegans, fluorescently labeled nuclei can be imaged in three-dimensional time-lapse (4D) movies and automatically tracked through early cleavage divisions up to ~350 nuclei. A similar analysis of later stages of C. elegans development has been challenging owing to the increased error rates of automated tracking of large numbers of densely packed nuclei. We present Nucleitracker4D, a freely available software solution for tracking nuclei in complex embryos that integrates automated tracking of nuclei in local searches with manual curation. Using these methods, we have been able to track >99% of all nuclei generated in the C. elegans embryo. Our analysis reveals that ventral enclosure of the epidermis is accompanied by complex coordinated migration of the neuronal substrate. We can efficiently track large numbers of migrating nuclei in 4D movies of zebrafish cardiac morphogenesis, suggesting that this approach is generally useful in situations in which the number, packing or dynamics of nuclei present challenges for automated tracking.

KEY WORDS: Cell lineage, Nuclear tracking, Computer image analysis, Cell migration, C. elegans, Zebrafish

INTRODUCTION
Tracking individual cells in developing embryos or tissues is important for understanding the molecular basis of differentiation and tissue morphogenesis at single-cell resolution. Through tracking, one gains access to a wide array of information on cell behavior, including the dynamics of cell division timing, ancestry, migratory paths and gene expression profiles. The quantitative and potentially comprehensive nature of such data is of increasing importance in systems-level analyses of development (Megasone and Fraser, 2007). Development of robust and accurate methods for tracking cells is becoming increasingly important for analysis of data sets in neuroscience (Higginbotham et al., 2011), stem cell research (Glauche et al., 2009) and cell biology (Arhel et al., 2006; Meijering et al., 2009). Automated tracking of cells or particles has become widespread in situations in which the signal-to-noise ratio (SNR), speed and sparseness of the objects permit reliable segmentation and tracking. However, many biological samples consist of large numbers of closely packed objects with low SNR, posing challenges for purely automated tracking.

The rapid development and near-transparent nature of C. elegans embryos and larvae has long allowed cell divisions to be followed manually using differential interference contrast (DIC) microscopy of live animals (Sulston and Horvitz, 1977; Sulston et al., 1983). Cell lineages can be traced using tools such as Simi Biocell (Schnabel et al., 1997). However, the low SNR of DIC images has precluded automated tracking in all but the earliest stages (Hamahashi et al., 2005). Localized fluorescent labels such as histone-GFP fusions, with their higher SNR, have recently made computer-based nuclear identification and tracking possible. Automated segmentation and tracking of histone-GFP-labeled nuclei in early C. elegans embryos has greatly facilitated analysis of early embryonic cell lineages (Bao et al., 2006; Santella et al., 2010), mutant phenotypes (Boeck et al., 2011), and gene expression patterns (Murray et al., 2008). However, tracking nuclei in later embryonic development, here defined as after the 350-cell stage, has been more challenging. The increasing density of nuclei and their semi-synchronous cell divisions lead to increasing error rates in automated analysis of embryos with >350 nuclei. Moreover, the SNR currently possible in conventional confocal 4D imaging is limited by the need to avoid phototoxicity in long time-lapse movies.

To track nuclear movements quantitatively beyond the 350-cell stage and through morphogenetic stages, we developed a new software approach. This approach has two key aspects. First, we seamlessly integrate automated search and manual curation, whereby the user confirms nuclear identifications at each time point prior to searching at the next time point. Thus, each automated search starts with a confirmed set of nuclei, minimizing propagation of false positives and other errors in the lineage tree. Second, as most nuclei do not move more than a predictable distance in the time intervals used in C. elegans 4D movies, we confine searches within a defined radius of each previously confirmed nucleus, greatly reducing tracking errors. Our approach is efficient and adaptable: we are able to track nuclei in 4D movies from several microscopy platforms, including a newly developed optical sectioning technology (Planchon et al., 2011), and from nuclear GFP movies of zebrafish development. Our semi-automated approach allows single-cell-level tracking of nuclei in complex samples in which the errors incurred in fully automated lineaging make automatic analysis challenging.

MATERIALS AND METHODS
4D movie acquisition with laser scanning confocal microscopy
In the C. elegans movies reported here we used the HIS-72::GFP (zuIs178 V) strain RW10029 (Ooi et al., 2006) to label nuclei. The RW10029 strain

1Division of Biological Sciences, Section of Cell and Developmental Biology, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA.
2Department of Electrical and Computer Engineering, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA.
3Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147, USA.

*Author for correspondence (chisholm@ucsd.edu)

Accepted 20 August 2012
was fully viable and fertile over several months of imaging. To acquire 4D data sets using laser scanning confocal microscopy, we followed previously published approaches (Murray et al., 2006). We used Zeiss LSM510 or LSM700 instruments equipped with a 100× NA 1.46 oil immersion objective. During acquisition, we employ a large pinhole size to improve the fluorescence signal, and a pixel dwell time of 1.26 μs to reduce laser exposure. We acquire z-stacks of 64 × 35 × 30 μm³ with a voxel resolution of 0.125 × 0.125 × 0.85 μm³ every minute for 480 minutes of development (Fig. 1A), starting at the 4- to 6-cell-stage embryo. We selected these imaging parameters to ensure embryo viability while maintaining image quality suitable for automated analysis. Different sublineages show different HIS-72::GFP fluorescence intensity, with body muscle D, endodermal E, and germline P4 sublineages having the lowest signal (not shown). Because the nuclear concentration and intensity of HIS-72::GFP increases through cleavage divisions (notably ~90 minutes after the 4-cell stage), we adjusted imaging parameters to maintain a constant SNR over the time course of acquisition (Fig. 1B). We acquired 4D data sets from 18 embryos imaged on a Zeiss LSM510 and two embryos imaged on a Zeiss LSM700. Phototoxicity due to fluorescence imaging was strongly temperature dependent: embryos above 24°C arrested during time-lapse confocal imaging, whereas embryos below 24°C were fully viable under identical conditions. We developed a Peltier-based cooling device to regulate specimen temperature, controlled by the confocal software.

Imaging using Bessel beam microscopy
We also acquired *C. elegans* 4D movies using Bessel beam plane illumination microscopy (Planchnon et al., 2011). Plane illumination microscopy is an optical microscopy method that permits optical sectioning (permitting 3D imaging); its main advantages are its low phototoxicity and its fast imaging capability. Plane illumination microscopy using Bessel beam excitation also improves the axial resolution, allowing imaging of full development while maintaining sample viability. Embryos were attached to polylysine-coated coverslips and oriented vertically in an imaging chamber containing M9 medium at 20-22°C. Volumes of 66 × 48 × 50 μm³ with a voxel resolution of 0.133 × 0.133 × 0.5 μm³ were acquired every minute for 480 minutes using the Bessel beam in linear excitation mode and 0.8 NA excitation and detection objectives. Five HIS-72::GFP embryos were recorded on the Bessel beam microscope.

Zebrafish nuclear GFP 4D movies
Movies of zebrafish endocardial development were obtained using transgenic line *Tg(fli1a:negfp)* (Roman et al., 2002). All animal work followed approved protocols of the University of California San Institutional Animal Care and Use Committee. We mounted transgenic embryos on coverslip bottom dishes (MatTek) in AquaPor low-melting temperature agarose (National Diagnostics) and cultured them in E3 media (Nüsslein-Volhard and Dahm, 2002). We performed time-lapse imaging at 28.5°C using a Leica SP5 confocal microscope with a 40× water immersion objective, capturing 40 optical slices at 1-minute intervals for 20 minutes. After imaging, we removed embryos from the imaging dishes and cultured them overnight to confirm wild-type heart morphology.

Tracking software and data analysis
Our cell tracking software, NucleiTracker4D, is implemented in MATLAB (MathWorks, Natick, MA, USA) and is freely available as an open source project at Sourceforge (https://sourceforge.net/projects/nucleitracker4d/). A user manual is provided on Sourceforge. We compute SNR in 4D data sets as follows. First, each stack is rendered isometric by replicating missing z-slices. Using the user-accepted nuclear positions, we treat all pixels inside the spherical mask defining nuclei as signal, and those outside as noise. We compute the mean signal over all the nuclei and mean noise over the remaining stack space. The SNR in dB is:

$$SNR(t) = 10\log_{10} \left( \frac{\text{mean Signal}(t)^2}{\text{mean Noise}(t)} \right), \quad (1)$$

where t is the stack time. Other data analysis and statistical methods were implemented in MATLAB.

RESULTS
A local search approach to tracking nuclei in 4D data sets
Our goal was to develop methods for tracking of all nuclei in the *C. elegans* embryo through morphogenetic stages, using 4D movies of embryos in which all nuclei have been labeled with histone::GFP (see Materials and methods) (Fig. 1C; supplementary material Movie 1). We developed a combination of automated tracking and manual curation to generate 4D models of the nuclei that can be easily analyzed computationally (Fig. 1C). Tracking of cell nuclei in a developing embryo involves several steps. First, nuclei must be recognized at a given time point, t. Second, nuclei identified at the next time point, t+1, must be linked to nuclei at time t. In the developing *C. elegans* embryo, nuclei at t have four possible fates at t+1: movement within the search radius; movement over a longer distance; division; or death. As cell deaths are easily recognized (see below), the crux of the problem is to distinguish whether a nucleus in a new position at time t+1 is a pre-existing nucleus that has moved beyond the threshold distance, or...
a daughter nucleus generated by division. Our approach does not involve de novo segmentation of the image into nuclei. Rather, we begin with the user defining the nuclei in the starting image, and confirming or correcting nuclear identifications at subsequent time points. Tracking the nuclei of a new sample embryo starts with the user naming the nuclei at the first time frame. We typically start at points. Tracking the nuclei of a new sample embryo starts with the user defining the nuclei in the starting image, and involve de novo segmentation of the image into nuclei. Rather, we begin with the user defining the nuclei in the starting image, and find the new position where the convolved signal is maximum (Fig. 2A). If more than one position within the search space shows a maximum convolution value, we choose the one closest to the position at time \( t \). If the distance between the position at \( t \) and at \( t+1 \) is less than \( R \), the new position is automatically accepted, and the user simply confirms that the automatic tracking is correct. Otherwise, the user is prompted for manual curation of the nucleus. Use of a search box of size proportional to nuclear radius was motivated by the observation that very early in development, nuclei are large and move around over long distances. At later stages, when nuclear radii are smaller, nuclei are more packed, hence their movement is constrained. In late development, large nuclei (e.g. E or P4 progeny) are also constrained; therefore, starting at \( \sim 250 \) minutes, we set the search space for nuclei in the C, E and P4 sublineages as \( \min[\text{radius}, 1.25R_{t+1}] \), where \( R_{t+1} \) is the displacement of that nucleus from the previous time point. These and other tracking parameters can be optimized depending on the particular sample. The tracking algorithm seeks a single nearby object at \( t+1 \) based on information at \( t \). Thus, for cells that divided in the time between \( t \) and \( t+1 \), the tracking algorithm will only identify one daughter nucleus. The other daughter can be curated manually, or identified using a ‘cell division detection’ algorithm that has been described elsewhere (Kang et al., 2012). Most cell divisions in the data sets presented here were curated manually.

**Effects of nuclear brightness, nuclear radius and cell death**

Our convolution maximum approach is most successful when the nuclear fluorescence signal is uniform within a nucleus and neighboring nuclei have similar intensity. This approach is slightly less reliable when fluorescence intensity differs significantly between neighboring nuclei. For the HIS-72::GFP marker used here, nuclei in the D, P4 and E lineages show fainter fluorescence than those of the AB, MS or C lineages. Thus, after convolution, the position of a faint nucleus is shifted towards one of the brighter neighboring nuclei if the latter is close enough (Fig. 2B). However, as these faint nuclei signals account for only 42 of 58 live nuclei, and only some of these 42 reside near more intense AB or MS nuclei, mis-positioning accounts for \(<5\%\) of curation events.

Although previous nuclear tracking methods search for an appropriate nuclear radius during segmentation (Bao et al., 2006), we found that the efficiency of our convolution maximum approach was increased by pre-specifying nuclear radii. In wild-type *C. elegans* embryos, nuclear radius decreases during development in a predictable way such that nuclei at any given round of division of a blastomere have similar radii (e.g. all AB16 cells have similar radii) (Fig. 2C). The local search algorithm for tracking is...
insensitive to slight decreases (but not to increases) in the nuclear radius, i.e. the predicted nucleus could have a radius slightly smaller by as much as 0.25 μm than the actual radius. The additional computation required to obtain a precise nuclear radius is therefore unnecessary for tracking. In mutants, or in non-*C. elegans* samples (see below), the nuclear radius model can be defined to allow more or less variation in nuclear size. Cells undergoing programmed cell death (apoptosis) display a distinctive pattern of histone-GFP expression in 4D movies (Fig. 2D; supplementary material Movie 2). Cells fated to undergo apoptosis are smaller than their sisters, although their nuclei are similar in size (Sulston et al., 1983; Hatzold and Conradt, 2008). Our procedure follows dying nuclei at their initial radius and represents them as a red circle of constant radius. When fluorescence of the dying cell drops below a threshold that would not interfere with the tracking of neighboring cells, the cell is marked as dead and is removed from the digital image.

**An integrated graphical user interface for tracking, curation and visualization**

Using the graphical user interface (GUI), users can specify the combination of algorithms to use for tracking and cell division and the cell naming convention to be used. If no tracking or cell division algorithms are specified, the program defaults to local search for tracking. The GUI relies on two sources of information for its operation: the user-curated list of nuclei at time \( t \) and the estimated list of nuclei at time \( t+1 \). The latter is derived from the former using the tracking and cell division algorithms specified by the user. The GUI prompts for user curation any objects that are not deemed correct by the algorithms; the user can also curate minor mistakes in automation, such as re-centering a virtual nucleus to a better center, so that the next step of tracking starts with a more reliable position. As our semi-automated approach is different in conception from fully automated tracking, a direct comparison of error rates is not meaningful. However, a comparison of the error rates of the automated part of our software with those of the initial and revised versions of Starrynite (Bao et al., 2006; Santella et al., 2010) indicates that it performs with comparable accuracy (supplementary material Fig. S1).

The key advantage to our approach is that error propagation is minimized by repeated curation as the number and packing of nuclei increase beyond 350 cells. At lower cell numbers or packing than 350 cells, our approach is slightly slower than the Starrynite algorithms, but yields a lineage that is completely accurate as judged by the user. A full 4D movie (300 z-stacks at 1 stack/minute, then 60 stacks taken every two minutes) has a cumulative total of 105,000 nuclear objects. About 22% of these accrue up to the 350-cell stage. Given the reliability of automated tracking in the early stages, one day of curation is sufficient to reach the 350-cell stage (270 minutes).

Timing of cell divisions was also highly reproducible, with cell cycles displaying standard deviations of ~4-7% of the mean (not shown), consistent with previous observations (Bao et al., 2008). Despite the invariance of *C. elegans* embryonic lineages and division timing, previous studies have reported greater variability in the relative positions of cells during embryogenesis (Schnabel et al., 1997; Schnabel et al., 2006). This variation might in part be due to differing degrees of mechanical compression of embryos during imaging (Hench et al., 2009). We therefore compared the development of embryos imaged under slight compression (beam mount, laser scanning microscopy confocal imaging) with that of embryos imaged in the uncompressed state using a novel structured illumination technology, Bessel beam imaging (Planchnon et al., 2011). Bessel beam microscopy uses scanning of an annular beam projected to the rear pupil of the excitation objective, which itself is orthogonal to a second detection objective. In Bessel beam imaging, the specimen is mounted on a vertically oriented polylysine-coated slide within a buffer chamber, reducing or eliminating compression. Bessel beams permit illumination with extremely thin light sheets, resulting in increased xyz resolution with non-invasive levels of illumination. Bessel beam illumination allowed high resolution 4D movies of histone-GFP labeled *C. elegans* embryos to be acquired through embryonic development (supplementary material Movie 5).

Early in gastrulation, a subset of nuclei undergo coordinated rotation as a loose grouping, counterclockwise around the long axis as viewed from the posterior. We confirmed this overall rotation in compressed embryos (supplementary material Fig. S2A, Movie 4) and found that rotation was reduced but not eliminated in uncompressed (Bessel-imaged) embryos (supplementary material Fig. S2B). The gastrulation stage rotation process, therefore, is not purely a consequence of mechanical compression. To quantitatively analyze the effects of compression on cellular contacts, we defined nuclear nearest neighbors (NNs; supplementary material Fig. S3) using Delaunay triangulation. Delaunay triangulation entails selecting sets of four nuclear ‘points’ such that no other point is contained within the tetrahedron defined by these four points. Thus, each nuclear point will share the edges of the tetragons made up by itself and its nearest neighbors. An equivalent geometrical approach, Voronoi decomposition, has been applied to embryos of up to 150 cells (Hench et al., 2009), and our results are broadly consistent with that analysis (e.g. supplementary material Fig. S3B). Although this metric of ‘instantaneous nearest neighbors’ (INNs) might overestimate actual cell-cell contacts, it captures potential local interactions without imposing an arbitrary distance threshold. We define lifetime NNs (LNNs) as the cumulative number of INNs that a nucleus has over its lifetime; to normalize for the effects of differing cell lifetimes, we divide LNNs by the cell lifetime to yield the ‘LNN rate’.

**Cellular neighborhoods and the effects of compression**

Our analysis confirms that the *C. elegans* embryonic lineage is highly invariant in the wild type; we found only a single instance of an extra division in the E lineage among the 25 wild-type embryos analyzed (20 analyzed to 350-cell stage, five to comma stage). Timing of cell divisions was also highly reproducible, with cell cycles displaying standard deviations of ~4-7% of the mean (not shown), consistent with previous observations (Bao et al., 2008). Despite the invariance of *C. elegans* embryonic lineages and division timing, previous studies have reported greater variability in the relative positions of cells during embryogenesis (Schnabel et al., 1997; Schnabel et al., 2006). This variation might in part be due to differing degrees of mechanical compression of embryos during imaging (Hench et al., 2009). We therefore compared the development of embryos imaged under slight compression (beam mount, laser scanning microscopy confocal imaging) with that of embryos imaged in the uncompressed state using a novel structured illumination technology, Bessel beam imaging (Planchnon et al., 2011). Bessel beam microscopy uses scanning of an annular beam projected to the rear pupil of the excitation objective, which itself is orthogonal to a second detection objective. In Bessel beam imaging, the specimen is mounted on a vertically oriented polylysine-coated slide within a buffer chamber, reducing or eliminating compression. Bessel beams permit illumination with extremely thin light sheets, resulting in increased xyz resolution with non-invasive levels of illumination. Bessel beam illumination allowed high resolution 4D movies of histone-GFP labeled *C. elegans* embryos to be acquired through embryonic development (supplementary material Movie 5).
determined by its position in the embryo, its lifetime and whether it or its neighbors actively migrate. For example, cells with the most LNNs include internal endodermal cells, such as Eala, with 87.6±6.3 LNNs over its lifetime of 83 minutes (Fig. 4A,B; supplementary material Movie 6), accumulated at a rate of 1.06±0.09 NNs min⁻¹. By contrast, an exterior cell, such as ABprpppap, encounters 29.9±3.7 NNs over its lifetime of 55 minutes, at a rate of 0.55±0.07 NNs min⁻¹. Cells that migrate over long distances encounter more LNNs than do cells that remain within the same neighborhood. For example, the somatic gonadal precursors Z1 and Z4, which undergo long-range migrations from the head to the mid body, accumulate LNNs faster than cells in comparatively static neighborhoods, such as Eprpp (Fig. 4C,D).

Finally, only cells that are NNs for a specified minimum time are counted, and are multiplied by the number of frames in which they are NNs, giving a measure of ‘persistent NNs’ (PNNs) per cell. Internal cells, such as the Eaxa cells, have many PNNs (17-18, as determined by confocal data sets). However, this is largely accounted for by their lifetime; when PNNs are normalized to lifetime, the PNN rate of Eaxa is low (~0.2 PNNs min⁻¹). PNNs were weakly but significantly anti-correlated with migration: cells that migrate effectively tend to have fewer PNNs (supplementary material Fig. S3A).

Next, we addressed whether compression affected cellular neighborhoods by comparing the INNs and LNNs of all cells that were completely imaged in confocal and Bessel embryos. The global mean of INNs increased through development, with an essentially identical profile in both imaging conditions (supplementary material Fig. S3B). LNNs for each nucleus were also highly correlated between compressed and uncompressed embryos (r = 0.79) (Fig. 4B), as was the rate of accumulation of LNNs (not shown). To address whether the composition of these neighborhoods varied significantly between compressed and uncompressed data sets, we analyzed the common nearest neighbors (CNNs, common to confocal embryos, common to Bessel embryos, and common to both) (supplementary material Fig. S3C). CNNs of confocal-imaged embryos showed high pairwise correlation (Spearman r = 0.87-0.89). Bessel-imaged embryos were slightly less correlated with each other (r = 0.79-0.84) and significantly less correlated with confocal-imaged embryos (r = 0.67-0.73) (supplementary material Fig. S3D). Compressed embryos displayed higher reproducibility in neighborhoods at most time points (supplementary material Fig. S3E). In conclusion, uncompressed embryos show slightly more embryo-to-embryo variation in their cellular neighborhoods than do compressed embryos, suggesting that compression constrains embryonic nuclei to a more predictable geometry.
Coordinated ventral neuroblast movements during ventral cleft closure and epidermal enclosure

Our primary interest in developing our nuclear tracking tool was to examine morphogenetic movements in the late embryo, a period in which the number and density of nuclei precluded the use of fully automated tracking algorithms. Here, we present the first single-cell resolution description of the movements of neural precursors in the mid-body following gastrulation and up to epidermal enclosure. After gastrulation, the coordinated movement of the ventral mesoderm creates a cleft on the ventral surface that closes from posterior to anterior (Sulston et al., 1983). Cleft closure involves the convergent movement of multiple ventral neuroblasts (VNBs) leading to the formation of a VNB monolayer between 290 and 350 minutes (ABpxp-derived cells; Fig. 3D,E, blue; supplementary material Movie 7); this sheet subsequently provides the substrate for epidermal enclosure movements (ABpxa-derived; Fig. 3D,E, red). The ventral blast cells mostly generate neurons, but also make some epidermal and muscle cells; for simplicity, we refer to them as ventral neuroblasts.

VNB movements in cleft closure

Immediately flanking the posterior ventral cleft are three pairs of AB<sub>128</sub> cells: ABp(r)pappa, ABp(U)r)pappp and ABp(U)r)pappp. After ~200 minutes, these precursors move to converge with their left-right counterparts, beginning with the posterior pair ABp(U)r)pappp. Between 225 and 232 minutes, these six cells divide in the eighth round of AB divisions to generate the cleft-closing VNBs (Fig. 5A, red cells; supplementary material Movie 8); for brevity, we refer to these VNBs as A1-6L and A1-6R; see supplementary material Table S1 for a full list of VNB abbreviations. We found that other VNBs could be grouped into anteroposterior sets that underwent coordinated movements during cleft closure and epidermal enclosure. More lateral to the A VNBs are rows of VNBs that we term the B (yellow), C (blue), and D (dark blue) neuroblasts; these do not undergo extensive movements until epidermal closure. The different VNB groups might have distinct molecular identities; for example, CeNeuroD/CND-1 is expressed in the C, D and E VNB groups but not in the majority of A and B VNBs (Murray et al., 2012).

Eph signaling is required for the coordinated movement of A VNBs

Several pathways have been implicated in closure of the gastrulation cleft, including signaling via the Eph receptor VAB-1 (George et al., 1998). To define more precisely the role of VAB-1 in VNB movements, we tracked VNB movements in vab-1 mutants and compared them with those in the wild type. We find that in vab-1 mutants the parents of the A VNBs display dramatic defects in closure. These defects resulted from delays and stalling at specific points rather than an overall slower migration rate (Fig. 5B,C). Thus, VAB-1 signaling might be involved in interactions between the VNBs and specific cells in their environment, rather than in general adhesion or motility.

Substrate movements during epidermal enclosure

Following cleft closure, the VNBs remain fairly static until epidermal enclosure begins at ~350 minutes. During epidermal enclosure, a striking set of coordinated movements rearrange the progeny of the VNBs into a longitudinal row of ventral midline ganglia and neurons (supplementary material Movie 9). The B, C and D rows of VNBs converge on the ventral midline, concurrent with the displacement of A VNB progeny to the anterior and posterior. The anterior A VNBs give rise to the excretory cell and neurons in the retrovesicular ganglion; the posterior A VNBs generate the neuron PVT and a mixture of rectal, valve and muscle cells. The E VNBs (Fig. 6, cyan) undergo strikingly rapid long-range movements during early epidermal enclosure (Fig. 6A). Ventral cord motor neurons, such as the DDs, arise from C-rank VNBs, and intercalate at the ventral midline such that left-hand (odd numbered) neurons migrate anteriorly to right-hand (even) neurons, and migrate in concert with closure of the ventral pocket by epidermal P cells (Fig. 6B). Finally, a large number of anterior (head) neurons also undergo intricate migrations concomitant with late epidermal enclosure; for example, a group of approximately ten anterior ventral neurons and support cells form bilateral streams that converge towards the anterior ventral midline from 330 to 380 minutes (Fig. 6C). In conclusion, our analysis reveals the highly dynamic nature of the neuronal substrate during epidermal enclosure, suggesting close coordination of substrate and epidermal movement.
Tracking patterns of cell movement during zebrafish endocardial morphogenesis

To address the generality of our local search and tracking algorithm, we tested it on 4D movies of GFP-labeled nuclei in the developing zebrafish heart. Transgenic expression of nuclear localized EGFP driven by the fli1a promoter allows labeling of nuclei during endocardial tube formation (Roman et al., 2002). Endocardial morphogenesis in zebrafish begins with the migration of endocardial progenitors from bilateral positions in the anterior lateral plate mesoderm to the midline at ~14 hours post-fertilization (hpf) (Bussmann et al., 2007). Once at the midline, endocardial cells translocate in a posterior direction between 17 and 20 hpf. Endocardial cells then undergo a leftward migration at ~20 hpf to form a ventrally extending tube by 22 hpf. The patterns of cell movement underlying these morphogenetic changes remain poorly understood.

We imaged endocardial movement beginning at 17 hpf, immediately after the fusion of the bilateral populations at the ventral midline and at the beginning of its posterior movement (Fig. 7A; supplementary material Movie 10). Despite the ellipsoidal morphology of the endocardial nuclei, we found that our local search algorithm performed well with few modifications, using a spherical convolution mask and a search box based on a constant nuclear radius. Nuclei move extensively during this stage (mean speed 1.7 μm/second, 1-minute intervals), up to 50 μm during the 20-minute movie. Nuclei were also tightly packed; as no endocardial nuclei divided in this short movie, most curation events involved resolving potential collisions of nuclei. Using our tracking approach, we were able to visualize the migration of >130 endocardial nuclei over a 20-minute period (Fig. 7A).

Our analysis revealed the utility of NucleiTracker4D software for visualization of a diversity of patterns of individual endocardial nuclei movement. For example, we easily observed the rapid posterior migration of most endocardial nuclei at the midline (Fig. 7B), consistent with prior studies (Bussmann et al., 2007). At the same time, we were able to observe nuclei moving in the anterior direction (Fig. 7B), as well as nuclei with net leftward or rightward trajectories (Fig. 7C). Notably, quantification of overall displacement vectors in the depicted example revealed a net leftward flow of cells around a central core of cells that moved slightly rightward (Fig. 7C). Coupling individual cell tracks with Delaunay triangulation allowed us to evaluate the degree of correlation between the movement patterns of neighboring endocardial nuclei (Fig. 7D,E). For instance, within the example shown, neighboring nuclei showed significant correlations in their trajectories and speeds (Fig. 7D). However, some endocardial nuclei moved in the opposite direction from their nearest neighbors, suggesting that they actively migrate rather than being passively swept along by other cells (Fig. 7E). Thus, this approach provides opportunities for future studies that might lead to a deeper understanding of
migration patterns underlying morphogenetic movements during endocardial tube formation.

**DISCUSSION**

We have presented a straightforward solution to the problem of tracking large numbers of nuclei in a complex and dynamic embryo. By integrating automated nuclear tracking with user curation we achieve both speed and accuracy. To our knowledge, the resulting data sets are the first complete lineages of single *C. elegans* embryos through morphogenetic stages. The accuracy of the original manual lineage analyses is a testament both to the power of direct observation and to the extreme invariance of the *C. elegans* embryonic lineage. However, our approach does not simply trace genealogical relations but allows quantitative analysis of many other previously inaccessible morphogenetic parameters.

Additional refinement of our approach is clearly possible. For example, as closely packed nuclei tend to move less, the size of the search space for nuclei could be adapted to suit local nuclear packing. Likewise, the use of a spherical mask may not always be optimal; an ellipsoidal mask could be used in cases in which nuclei are more variable in proportion. The open source nature of our software should allow modular additions and improvements by the community. However, the major limitation in achieving accurate automatic tracking from long 4D data sets is the need to compromise image quality (both in xyzr resolution and in SNR) to preserve specimen viability. Recent developments in 4D microscopy might be able to overcome this limit (Fischer et al., 2011; Wu et al., 2011). We find that the low light levels possible using Bessel beam illumination (Planchon et al., 2011) allow far higher resolution 4D movies to be obtained, without detectable phototoxicity. Unlike most other microscopy platforms, in which the embryo is slightly compressed between a cover slip and a slide, in the Bessel microscope the embryo is attached only on one side to a polylysine cover slip, and is otherwise suspended, uncompressed, in the buffer chamber. The higher resolution of nuclear structures possible with Bessel microscopy means that nuclei no longer appear as simple spherical blobs, but adopt a wider range of shapes. Although in theory such variability in nuclear form could make use of a spherical convolution mask less valid, in practice the improved resolution means that nuclear tracking accuracy is comparable to that in our LSM confocal data sets. The lack of compression possible in Bessel beam imaging might be important in certain experiments. However, uncompressed embryos can assume a wide variety of orientations, making visual inspection of embryonic processes occasionally challenging. We find that the slight compression of bead mounting allows embryos to develop in stereotyped ways that largely preserve the neighborhood relationships of nuclei in uncompressed embryos. For many purposes, these technical advantages of compression may be very useful.

The dynamic rearrangements of cells in the *C. elegans* embryo remain relatively poorly understood, with the exception of the ingression movements of gastrulation (Lee and Goldstein, 2003; Lee et al., 2006; Harrell and Goldstein, 2011). Following gastrulation, the most prominent morphogenetic movements are those of ventral cleft closure and subsequent epidermal enclosure. Our analysis has defined the individual cell movements that contribute to gastrulation cleft closure. The subsequent rearrangement of substrate cells during epidermal enclosure has not previously been described. Our analysis indicates that epidermal movements are precisely and extensively correlated with movements in the underlying neuronal substrate. Initial movement of the epidermal leading cells correlates with the rapid migrations of the E rank neuroblasts and their progeny. Closure of the ventral pocket is accompanied by extensive convergence and intercalation of the B, C and D rank neuroblasts, which generate the ventral nerve cord and ventral ganglia. Indeed, descendants of the A3, B3 and C3 VNBs have recently been shown to play specific roles in bridging the ventral pocket P9/10 cells during enclosure (Ikegami et al., 2012). Finally, we find that enclosure of the anterior epidermis is accompanied by extensive anterior flow of neurons and neuronal support cells. Interestingly, substrate movements play an active force-generating role in epithelial spreading during *Drosophila* dorsal closure (Solon et al., 2009). It will be important to address the functional and mechanical relationship of substrate and epidermal movements: are they independent, dependent or interdependent?

Our nuclear tracking algorithm has yielded, to our knowledge, the first complete descriptions of cellular positions during *C. elegans* epidermal enclosure. Although tracking nuclei does not give a complete picture of cell contacts or movements, it is important for interpreting other kinds of dynamic data, such as those involving cell surfaces or junctions. Despite recent advances,
automated segmentation of surface or junction markers remains challenging (Blanchard et al., 2009; Aigouy et al., 2010; Fernandez et al., 2010; Luengo-Oroz et al., 2011); nuclear information can provide a framework for such efforts. By combining the nuclear tracking information presented here with data on cell contacts, it might eventually be possible to generate predictive models of the cellular dynamics and mechanical forces at play during C. elegans embryonic morphogenesis.

Acknowledgements

We thank Zhirong Bao for continued helpful discussions on automated nuclear tracking. We thank members of the Jin and Chisholm laboratories for comments, advice and encouragement.

Funding

C.A.G. was supported by a University of California San Diego (UCSD) Neurobiology Training Grant [T32 NS007220]; and by a Ruth S. Kirschstein National Research Service Award [F32 GM090652]. Work in D.Y.’s laboratory was supported by a March of Dimes Research Grant [1- FY11-493] and by an American Recovery and Reinvestment Act (ARRA) supplement to A.D.C. Deposited in PMC for release after 12 months.

Competing interests statement

The authors declare no competing financial interests.

Supplementary material

Supplementary material available online at http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.086256/-/DC1

References


A

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>True positives (TP)</td>
<td>False negatives (FN)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>False positives (FP)</td>
<td>True negatives</td>
<td></td>
</tr>
</tbody>
</table>

**True Positive Rate (TPR)**

\[ TPR = \frac{TP}{TP + FN} \]

**Positive Predictive Value (PPV)**

\[ PPV = \frac{TP}{TP + FP} \]

**Error**

\[ Error = 1 - TPR \times PPV \]

---

**Fig. S1. Error rates in NucleiTracker 4D and other automated lineage software.**

(A) The confusion matrix and error metrics. (B) NucleiTracker4D performance. We used NucleiTracker4D to track nuclei over 99 to 115 minutes from the four-cell stage in three confocal data sets: the original data set from Bao et al. (Bao et al., 2006) (blue); one Zeiss LSM 510 movie (red); one dataset obtained from Zeiss LSM 700 (green) using NucleiTracker4D. NucleiTracker4D displays both a high true positive rate (TPR) and positive predicted value (PPV) across all data sets, resulting in very low cumulative error. (C) Starrynite v2 (Santella et al., 2010) performance. We first optimized the program parameters for each data set using curated data for the first 50 minutes (opt) and then ran the program with these optimized parameters up to 99 minutes.
Fig. S2. The effect of compression on embryonic rotational movement and nuclear displacement. (A) The relative rotation of a given precursor and its descendants, expressed as degrees around the long axis of the embryo as viewed from the anterior pole. Angle of the precursor at 52 minutes is set as 0°. The angle of the center of gravity of the descendants is plotted over time, as is the average of the AB⁸ groups (excluding ABala, which lies at the anterior pole and movements of which are not well captured by this metric). Note the clockwise rotation of most cell groups except ABalp and ABara between 150 and 180 minutes. In general, posterior AB cells rotate more than anterior AB cells. AB⁸, MS and C color codes as in Movie 4. (B) Analysis of embryonic rotations in uncompressed embryos. Unlike confocal imaging, Bessel imaging does not constrain the embryo to a stereotypical orientation. We therefore rotated and aligned Bessel data sets to match the long axes and left-right symmetry of confocal imaged data sets, prior to analysis of embryonic rotation.
**Fig S3.** Nuclear nearest neighbors (NNs) and the effect of compression. (A) Negative correlation of persistent NNs with displacement effectiveness in a single confocal-imaged embryo. The net displacement of a nucleus throughout its lifetime is the vector from its position at birth to its position prior to division. The total displacement is the sum of stepwise migrations made by a nucleus throughout its lifetime. The displacement effectiveness (DE) is defined as the ratio of net to total displacement. (B) Evolution of the global mean number of instantaneous nearest neighbors (INNs) through embryonic development for wild-type compressed (blue) and wild-type uncompressed (Bessel imaging, red) embryos. At each time point the number of nearest neighbors of all nuclei is averaged for the entire embryo. Error bars indicate s.e.m. across the nuclei present in the embryo at a particular time. Compressed embryos display slightly higher mean INNs at late time points; c.f. Figure 5C of Hench et al. (Hench et al., 2009). (C) Set description of common nearest neighbors for a specific nucleus. The set of common nearest neighbors (CNNs) is defined as the LNNs for a specific nucleus consistently found in all data sets. A nucleus must be neighboring for at least a threshold minimum time to be counted as a NN. For each nucleus we make all possible pairwise comparisons among embryos imaged in confocal (C1, C2...n=3) or Bessel (B1, B2...n=3) conditions. (D) Typical correlations of CNN values (threshold=2 minutes) between individual compressed (blue) and uncompressed (red) embryos. Compressed embryos are more highly correlated with each other than are uncompressed embryos. (E) Uncompressed (Bessel) imaged embryos display lower embryo-embryo reproducibility in nearest neighborhoods compared with confocal imaged embryos. Reproducibility is defined as the ratio (CNN/LNN), averaged over the whole embryo at each time point.
Table S1. Shorthand names for ventral neuroblasts from ~220 to 280 minutes

<table>
<thead>
<tr>
<th>E, D right</th>
<th>C right</th>
<th>B right</th>
<th>A right</th>
<th>A left</th>
<th>B left</th>
<th>C left</th>
<th>E, D left</th>
</tr>
</thead>
<tbody>
<tr>
<td>E0R</td>
<td>ABarappapa</td>
<td>RMDDR/x-SMBDR</td>
<td>B-1R</td>
<td>ABprppapa</td>
<td>SMDDR/AIYR</td>
<td>B-1L</td>
<td>ABplpappa</td>
</tr>
<tr>
<td>E1R</td>
<td>ABppapaap</td>
<td>exc socket G1/DB3</td>
<td>B0R</td>
<td>ABppapaap</td>
<td>SMBDR/AIYR</td>
<td>B0L</td>
<td>ABplpappaap</td>
</tr>
<tr>
<td>E1L</td>
<td>ABppapaap</td>
<td>exc duct DB1</td>
<td>E0L</td>
<td>ABarappapa</td>
<td>RMDDL/x-SMBDL</td>
<td>E1L</td>
<td>ABppapaap</td>
</tr>
<tr>
<td>C1R</td>
<td>ABprppapa</td>
<td>x/RIMR</td>
<td>B1R</td>
<td>ABprppapa</td>
<td>excglR/AVKRI</td>
<td>B1L</td>
<td>ABplpappaap</td>
</tr>
<tr>
<td>C1L</td>
<td>ABplppapaap</td>
<td>x/RIML</td>
<td>E2L</td>
<td>ABarappapa</td>
<td>x/SMBVL</td>
<td>C1L</td>
<td>ABplppapaap</td>
</tr>
<tr>
<td>C2R</td>
<td>ABprppapaap</td>
<td>AIAR/DB7</td>
<td>B2R</td>
<td>ABprppapaap</td>
<td>RIH/AVL</td>
<td>A2L</td>
<td>ABplpappap</td>
</tr>
<tr>
<td>C2L</td>
<td>ABplppapaap</td>
<td>AIAL/DB6</td>
<td>B2L</td>
<td>ABplpappap</td>
<td>AVKL/excglL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D1R</td>
<td>ABprppapaap</td>
<td>SABVR, RIFR/DA1</td>
<td>B3R</td>
<td>ABprppapaap</td>
<td>RIRAVG</td>
<td>A3L</td>
<td>ABplpappaap</td>
</tr>
<tr>
<td>D1L</td>
<td>ABplppapaap</td>
<td>SABVL, RIFR/SABD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D2R</td>
<td>ABprppapaap</td>
<td>DA3/DA6</td>
<td>B4R</td>
<td>ABprppapaap</td>
<td>vir/virR</td>
<td>A4L</td>
<td>ABplppappaap</td>
</tr>
<tr>
<td>D2L</td>
<td>ABplppapaap</td>
<td>DA2/DA4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3R</td>
<td>ABprppapaap</td>
<td>Y/DA7</td>
<td>B5R</td>
<td>ABprppapaap</td>
<td>RIRAVG</td>
<td>A5L</td>
<td>ABplppappaap</td>
</tr>
<tr>
<td>C3L</td>
<td>ABplppapaap</td>
<td>RIGL/DD1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4R</td>
<td>ABprppapaap</td>
<td>Y/DA7</td>
<td>B6R</td>
<td>ABprppapaap</td>
<td>vir/virR</td>
<td>A6L</td>
<td>ABplppappaap</td>
</tr>
<tr>
<td>C4L</td>
<td>ABplppapaap</td>
<td>DA9/DA6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C6R</td>
<td>ABprppapaap</td>
<td>PVCRL/LUAL/PHAR</td>
<td>B6R</td>
<td>ABprppapaap</td>
<td>spike/hyp10</td>
<td>B6L</td>
<td>ABplppapaap</td>
</tr>
<tr>
<td>C6L</td>
<td>ABplppapaap</td>
<td>PVCVL/LUAL/PHAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>