Review Article

Towards comprehensive cell lineage reconstructions in complex organisms using light-sheet microscopy

Fernando Amat* and Philipp J. Keller*

Janelia Farm Research Campus, Howard Hughes Medical Institute, 19700 Helix Drive, Ashburn, Virginia, 20147, USA

Understanding the development of complex multicellular organisms as a function of the underlying cell behavior is one of the most fundamental goals of developmental biology. The ability to quantitatively follow cell dynamics in entire developing embryos is an indispensable step towards such a system-level understanding. In recent years, light-sheet fluorescence microscopy has emerged as a particularly promising strategy for recording the *in vivo* data required to realize this goal. Using light-sheet fluorescence microscopy, entire complex organisms can be rapidly imaged in three dimensions at sub-cellular resolution, achieving high temporal sampling and excellent signal-to-noise ratio without damaging the living specimen or bleaching fluorescent markers. The resulting data-sets allow following individual cells in vertebrate and higher invertebrate embryos over up to several days of development. However, the complexity and size of these multi-terabyte recordings typically preclude comprehensive manual analyses. Thus, new computational approaches are required to automatically segment cell morphologies, accurately track cell identities and systematically analyze cell behavior throughout embryonic development. We review current efforts in light-sheet microscopy and bioimage informatics towards this goal, and argue that comprehensive cell lineage reconstructions are finally within reach for many key model organisms, including fruit fly, zebrafish and mouse.

Key words: cell lineage reconstruction, cell tracking, data visualization, embryonic development, light-sheet microscopy, quantitative developmental biology.

Introduction

Following the dynamic behavior of every cell at every point in time and space throughout the development of entire complex organisms is one of the central goals of developmental biology (Megason & Fraser 2007; Keller *et al.* 2008; Khairy & Keller 2011; Tomer *et al.* 2012). Comprehensive reconstructions of cellular dynamics and cell lineage information are indispensable for systematically dissecting functional relationships in the developmental building plan, understanding the morphological development of complex tissues and entire organisms, quantitatively and comparatively analyzing mutant phenotypes, correlating gene expression and cell fate decisions, testing biophysical models of the physical forces acting in development and, ultimately, formulating and testing models of the entire developing embryo. In the long-term perspective, the systematic reconstruction and correlation of cell lineage information for individuals of the same species as well as across species borders may furthermore provide key insights into the fundamental quantitative rules underlying developmental building plans.

In order to realize the automated reconstruction of cellular dynamics, however, a combination of key advances in *in vivo* fluorescence light microscopy, computational image processing, image data management and data visualization are needed to generate and efficiently analyze the large amount of information required for system-level studies of development. Figure 1 shows a generic pipeline for such experiments and analyses (please see Box 1 for a definition of technical terms). Briefly, the specimen is recorded *in vivo* for the maximum duration possible without causing damage to the fluorescent markers (owing to photo-bleaching) or to the specimen itself (owing to photo-toxic effects). Achieving good physical coverage, spatial resolution and temporal sampling is
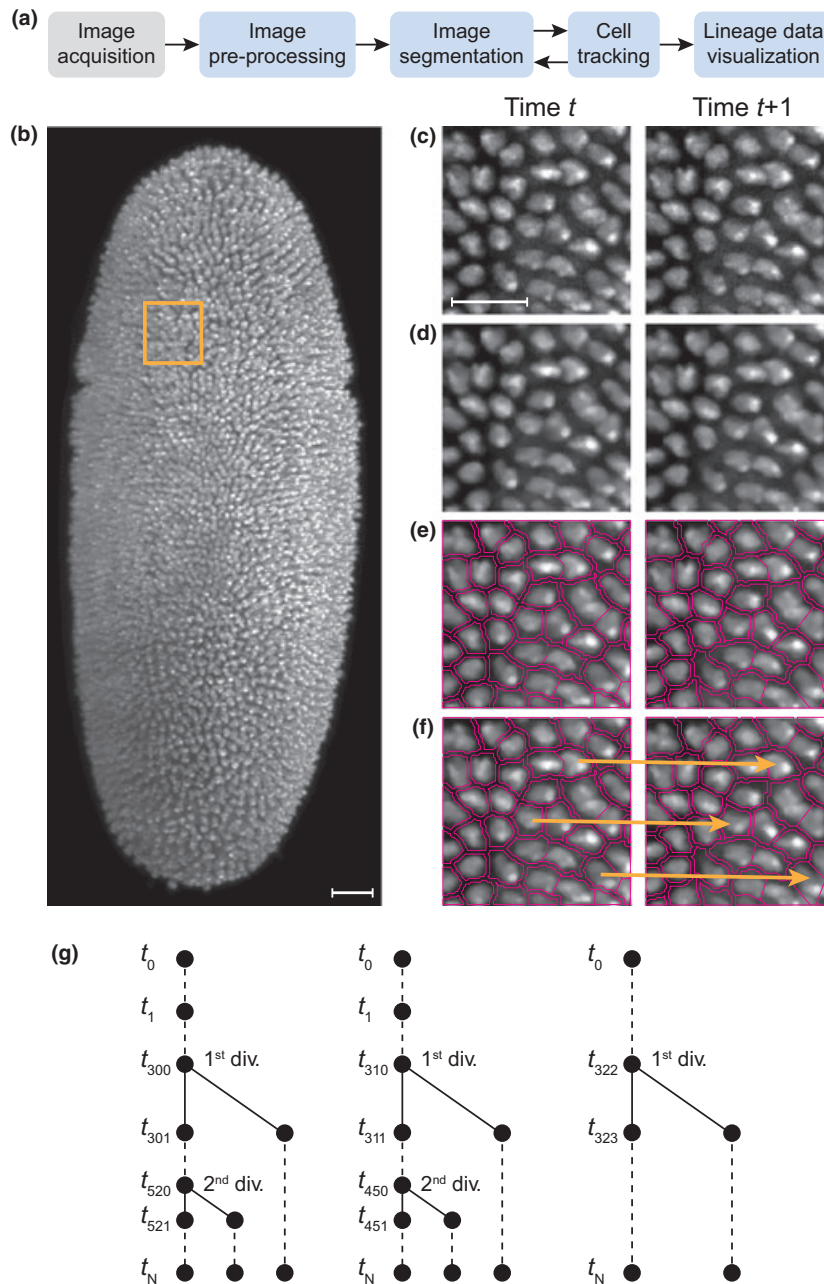
Fig. 1. Pipeline for cell lineage reconstructions. (a) Block diagram comprising the main steps required to obtain cell lineage information from time lapse microscopy data. Pre-processing refers to any image processing tasks required to enhance the datasets for the purpose of accurate segmentation and cell tracking. Segmentation refers to spatial coherence while tracking refers to temporal coherence. Although image segmentation and cell tracking are often seen as separate steps, both tasks can benefit from each other, considering the high spatial correlation between adjacent images in time (indicated by a bidirectional arrow). Panels (b–g) provide examples for each step, using 2D images for illustrative purposes. The same pipeline can be applied to 3D images. (b) Maximum-intensity projection of a raw 3D image dataset showing a nuclei-labeled *Drosophila* embryo recorded with SiMView light-sheet microscopy. (c) Enlarged view of the region indicated by the orange box in (a). Two consecutive time points are shown. (d) Applying a median filter to the images shown in panel (c) removes shot noise and represents a commonly-used image pre-processing step in light microscopy. (e) Segmentation of the images shown in panel (d) provides estimated cell boundaries (pink). (f) Cell tracking involves identification of corresponding nuclei in subsequent time points (orange arrows). (g) Abstraction of the segmentation and tracking results obtained from the processing steps illustrated in panels (c–f) allows visualizing and analyzing the lineage information. The cell lineages are constructed by concatenating the pairwise associations shown in panel (f) across multiple time points and through cell divisions. Scale bars: 25 μm (b), 10 μm (c). Credits: Panel (g) was reprinted from Tomer *et al.* (2012), Copyright (2012), with permission from Macmillan Publishers Ltd.

crucial to reliably capture and resolve cell migration and cell division events across the entire embryo. In the resulting datasets, cell boundaries and/or locations need to be identified for every cell in the embryo and at every time point (segmentation) and associated with the correct object in the next time point (tracking). Since complex multicellular organisms typically comprise many tens of thousands of cells already in early developmental stages, automated computational approaches are required to perform these tasks and extract quantitative information from the recorded images that can then be analyzed for new biological insights.

In the following sections, we will discuss in more detail state-of-the-art approaches for each of these steps, from image acquisition to image analysis and data visualization. Although there are still many challenges (Keller *et al.* 2008; McMahon *et al.* 2008; Olivier *et al.* 2010), we argue that complete cell lineage reconstructions in complex multicellular organisms are within reach in coming years.

Light-sheet microscopy

Light-sheet microscopy has emerged as a powerful technology that provides substantially improved performance over confocal and point-scanning two-photon microscopy in several parameters crucial for long-term *in vivo* imaging of complex multicellular specimens (Keller & Dodt 2011; Tomer *et al.* 2011). The central idea behind light-sheet microscopy is to illuminate a thin volume section of the sample from the side, using a thin sheet of light or a rapidly scanned pencil beam perpendicular to the axis of fluorescence detection (Fig. 2) (Siedentopf & Zsigmondy 1903; Voie *et al.* 1993; Fuchs *et al.* 2002; Huisken *et al.* 2004; Keller *et al.* 2008). Light-sheet microscopy provides intrinsic optical sectioning by illuminating only the part of the sample that is in focus of the detection system. Exposure of the specimen to laser light is thus greatly reduced. Positioning the detection objective perpendicular to the illuminated plane allows recording an image of the entire illuminated plane with a camera-based detection system in a single step. Fast three-dimensional (3D) imaging is performed by moving the sample through the light sheet or by quickly displacing the light sheet and detection optics.

Light-sheet microscopy has four main advantages over conventional imaging approaches. First, photo-bleaching and photo-toxic effects are substantially reduced, which is of particular importance for long-term *in vivo* imaging under physiological conditions. For example, light-sheet microscopy has been used to image *Drosophila* (Tomer *et al.* 2012) and early zebra-fish (Keller *et al.* 2008) embryogenesis at a temporal
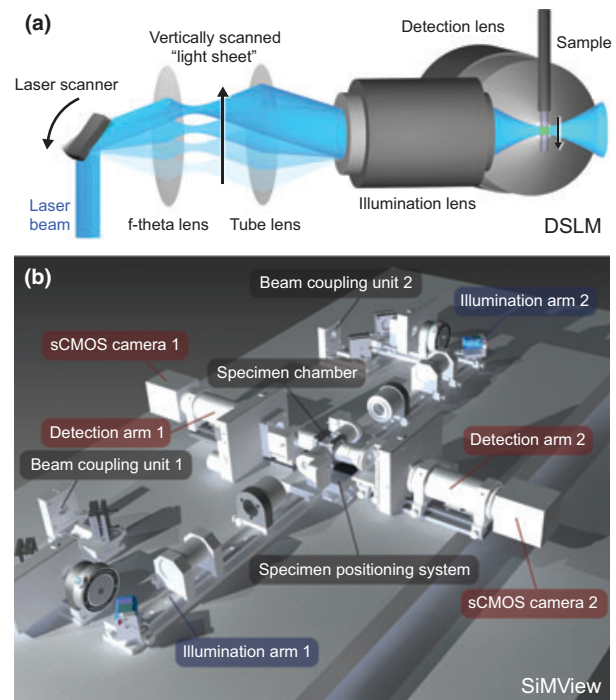

Fig. 2. Scanned light-sheet microscopy. (a) The illustration shows the principle behind Digital Scanned Laser Light Sheet Fluorescence Microscopy (DSLM). The f-theta lens converts the tilting movement of the scan mirror into a vertical displacement of the laser beam. The tube lens and the illumination objective focus the laser beam into the specimen, which is positioned in front of the detection lens. The laser beam thus illuminates the specimen from the side and excites fluorophores along a single line. Rapid scanning of a thin volume and fluorescence detection at a right angle to the illumination axis provides an optically sectioned image. (b) Computer model of the opto-mechanical implementation of a light sheet microscope for simultaneous multiview imaging (SiMView). The opto-mechanical modules of the instrument consist of two illumination arms for fluorescence excitation with scanned light sheets (blue), two fluorescence detection arms equipped with sCMOS cameras (red) as well as beam-coupling modules, specimen chamber and the specimen positioning system (grey). Credits: Panel (a) was reprinted from Keller *et al.* (2008), Copyright (2008), with permission from AAAS. Panel (b) was reprinted from Tomer *et al.* (2012), Copyright (2012), with permission from Macmillan Publishers Ltd.

resolution that enables comprehensive cell tracking. Second, camera-based fluorescence detection greatly speeds up image acquisition over point-scanning volumetric imaging. The speed bottleneck in light-sheet microscopy is effectively determined by the performance of the camera and the electronics required for data transfer and storage. Third, light-sheet microscopy provides an exceptionally high signal-to-noise ratio (SNR), owing to the long pixel dwell times arising from parallelized signal-detection with CCD or sCMOS detectors. Finally, light-sheet microscopes are relatively

inexpensive (a very simple, yet powerful system can be built for around $50 000 and less) and several ongoing efforts aim at providing open source software for recording and processing light-sheet microscopy data (Eliceiri *et al.* 2012). Moreover, the first commercial light-sheet microscopes have become available recently.

A key challenge in light-sheet microscopy as well as any other light-based microscopy technique is the limited physical penetration depth of light at physiological wavelengths in biological tissues. In addition, optical aberrations and the increase in light sheet thickness as a result of light scattering can lead to substantial variation of the point spread function (PSF) across the specimen. However, the core design principles of light-sheet microscopy can be extended in multiple ways to effectively address these issues. Combining light-sheet microscopy with multi-photon excitation substantially improves sample penetration and physical coverage of the specimen (Palero *et al.* 2010; Truong *et al.* 2011; Tomer *et al.* 2012). In sequential multiview imaging, multiple complementary views of the specimen are recorded along different directions to increase physical coverage of the specimen (Huisken *et al.* 2004; Swoger *et al.* 2007; Keller *et al.* 2008, 2010; Preibisch *et al.* 2010). This latter approach, however, introduces a trade-off between physical coverage and imaging speed and disrupts the spatio-temporal continuity of the recording when imaging fast dynamic processes. In contrast, light-sheet microscopes designed with multiple detection and illumination arms enable multiview imaging without the need for sample rotation and without a reduction of temporal sampling (Tomer *et al.* 2012; Krzic *et al.* 2012). By combing such optical multi-lens designs with two-photon excitation the truly simultaneous acquisition of four complementary views of the specimen can be realized, providing close to optimal physical coverage even for large non-transparent specimens (Tomer *et al.* 2012). Finally, the axial extent of the point-spread-function can be significantly reduced by using Bessel beam illumination in combination with structured illumination or multi-photon excitation (Planchon *et al.* 2011; Gao *et al.* 2012).

All of these recent developments lead to complementary improvements in the quality of the data recorded with light-sheet microscopy. The SiMView light-sheet microscopy implementation introduced above provides imaging capabilities indispensable for cell lineage reconstructions: Figure 3 shows a SiMView light-sheet microscopy dataset of *Drosophila* embryonic development recorded at a rate of 175 million voxels per second (Tomer *et al.* 2012). In these types of experiments, 3D image stacks of the entire embryo are acquired simultaneously from four different directions every 30 s, generating several gigabytes of image data per time point over a period of approximately 20 h. Over the course of a single experiment, several terabytes of image data are recorded, which provide detailed information on the cellular dynamics of tens of thousands of cells in the developing embryo. Figure 4 shows a manual proof-of-principle reconstruction of several neuroblast and epidermoblast cell lineages from such a SiMView recording. Scaling these types of analyses (McMahon *et al.* 2008; Swoger *et al.* 2011; Tomer *et al.* 2012) to the whole-embryo level, that is, realizing the goal of complete cell lineage reconstructions for entire complex multi-cellular organisms, will rely critically on the development of new automated computational approaches with extremely low error rates.

Computational approaches to cell lineage reconstruction

Generally, three main computational tasks are involved in cell lineage reconstructions: image pre-processing, cell segmentation and cell tracking (Fig. 1). Image pre-processing refers to any image processing task required to improve the SNR, image contrast or resolution of the recordings to the extent necessary to facilitate the other two main steps. Typical pre-processing steps are image registration to fuse multiview datasets (Preibisch *et al.* 2010), deconvolution (Temerinac-Ott *et al.* 2012; Tomer *et al.* 2012) and filtering (Perona & Malik 1990). Segmentation refers to any partitioning or grouping of the voxels in each 3D volume based on whether they belong to the same cell or not. Tracking refers to any partitioning or grouping of the voxels between two consecutive volumes in time based on whether they belong to the same cell or not. Thus, segmentation returns image clusters in space, whereas tracking returns image clusters in time. Once a complete cell lineage reconstruction has been obtained, the recorded terabytes of image information can be synthesized very efficiently using a tree structure as shown in Figure 4. Every node in the tree corresponds to a cell and contains information extracted from the image such as size, position, movement speed, gene expression levels, etc.

Both segmentation and tracking are long-standing key problems in image processing, computer vision and engineering research fields (Stone *et al.* 1999; Russ 2011), and it is thus not feasible to provide a comprehensive review of all approaches presented through decades of research in the following sections. We will therefore primarily focus on cell tracking
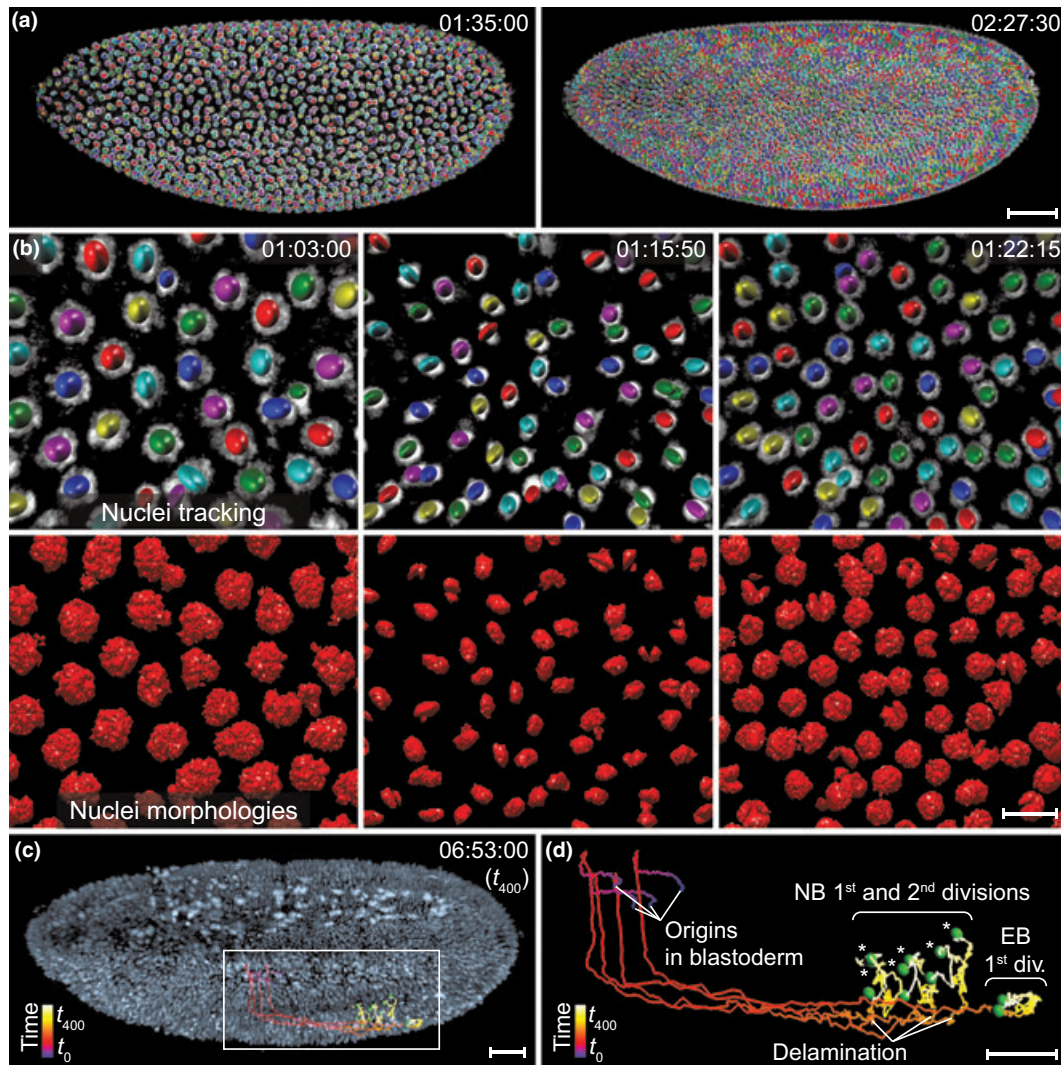
Fig. 3. Segmentation and cell tracking in the early *Drosophila* embryo. (a) Quantitative reconstruction of nuclei dynamics in the syncytial blastoderm. Global nuclei tracking in the entire *Drosophila* syncytial blastoderm. Raw image data from light-sheet microscopy was superimposed with automated tracking results using sequential Gaussian mixture model approach. Images show snapshots before the 12th mitotic wave and after the 13th mitotic wave (using a random color scheme in the first time point), which is propagated to daughter nuclei using tracking information. (b) Enlarged view of a reconstructed embryo in panel (a) with nuclei tracking information (left) and morphological nuclei segmentation (right). (c) SiMView recording of a histone-labeled *Drosophila* embryo superimposed with manually reconstructed lineages of three neuroblasts and one epidermoblast for 120–353 min after fertilization (time points 0–400 min); track color encodes time. (d) Enlarged view of tracks highlighted in (c). Green spheres show cell locations at time point 400. Asterisks mark six ganglion mother cells produced in two rounds of neuroblast division. NB, neuroblast; EB, epidermoblast. Scale-bars: 50 μm (a), 10 μm (b), 30 μm (c,d). Credits: Figures reprinted from Tomer *et al.* (2012), Copyright (2012), with permission from Macmillan Publishers Ltd.

approaches tested in the context of cell lineage reconstructions using light microscopy. We refer the reader to (Khairy & Keller 2011) for a detailed review on image pre-processing and segmentation approaches for light-sheet microscopy datasets. Finally, it should be noted that, although segmentation and tracking are traditionally seen as separate problems, both tasks can be interleaved, since information from an improved segmentation will simplify tracking and vice versa (Kausler *et al.* 2012).

Cell tracking

As discussed in the previous section, advanced light-sheet microscopes provide exceptionally good performance (high spatio-temporal resolution, high SNR, good physical coverage, etc.) for the purpose of systematic cell lineage reconstructions in complex multicellular organisms. Nevertheless, many challenges must be addressed to achieve faithful cell tracking. First, using typical fluorescent marker strategies
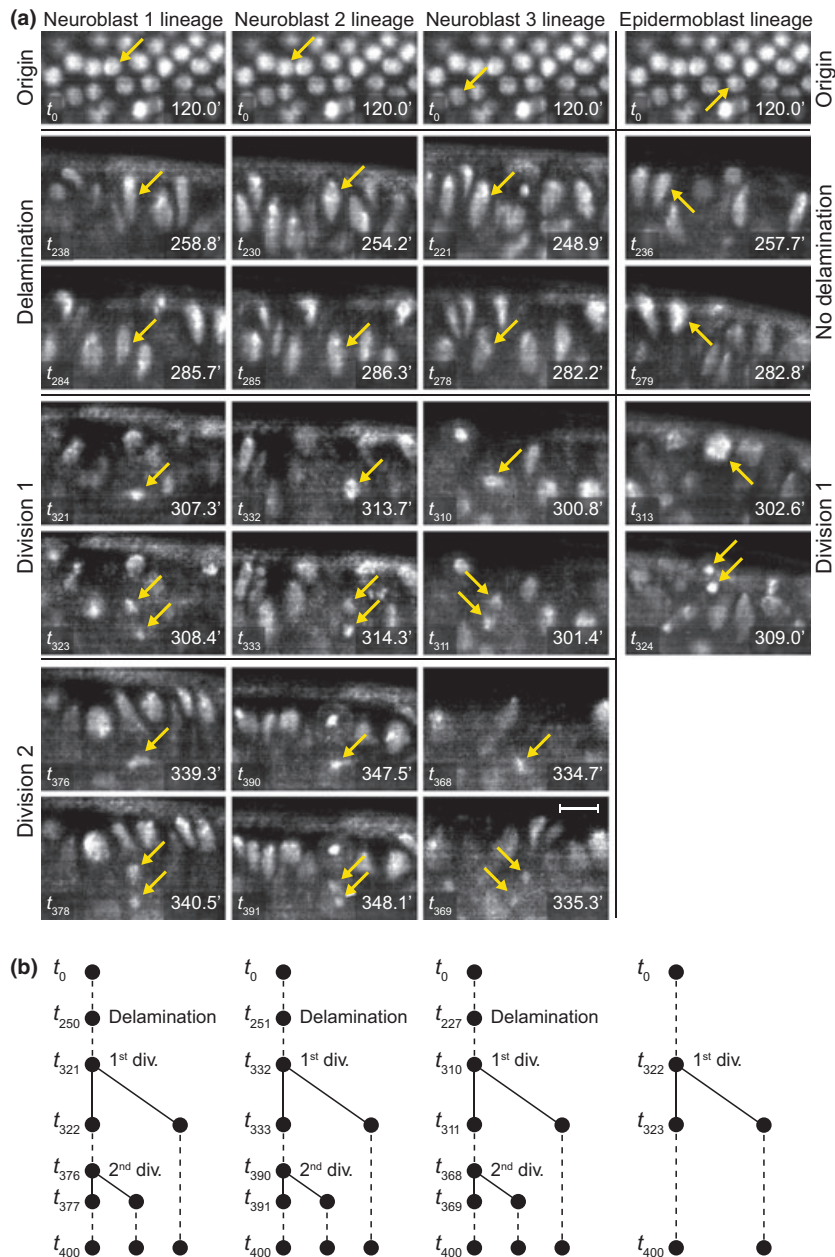
Fig. 4. Reconstructing neuroblast and epidermoblast lineages in the *Drosophila* embryo. (a) Raw optical slices from SiMView recording for key events in the lineage reconstructions visualized in Figure 3c. Optical slices indicate blastoderm origins, delamination, first cell division and second cell division for three neuroblasts, as well as blastoderm origin and first cell division for one epidermoblast. Yellow arrows indicate the locations of the nuclei of the tracked cells. The appearance of stripes in the raw data arises from the column gain variability typically encountered in first generation sCMOS cameras (such as the Andor Neo detector used in this recording). The SiM-View processing pipeline contains a module for measuring column gain factors and correcting these stripes. (b) Lineage trees for the neuroblast/epidermoblast lineage reconstructions visualized in Figure 3c (1st div. = first division, 2nd div. = second division). Four blastoderm cells and their respective daughter cells were manually tracked from time point 0 to 400 (120–353 min post fertilization, 35 s temporal resolution), using Imaris (Bitplane) and ImageJ (http://rsbweb.nih.gov/ij/). Tracks start in the blastoderm (time point 0). The neuroblasts delaminate between time points 227 and 251, and subsequently produce ganglion mother cells in two division cycles (first cycle between time points 310 and 332, second cycle between time points 368 and 390). The epidermoblast remains in the outer cell layer and divides once at time point 313. Manual tracking was performed until time point 400 for all cells. Scale-bar: 10 μm (a). Credits: Figure reprinted from Tomer *et al.* (2012), Copyright (2012), with permission from Macmillan Publishers Ltd.

involving ubiquitously-expressed labels with nuclear or membrane localization, most of the tens of thousands of nuclei/cells in the specimen look very much alike. Each nucleus or cell occupies only a small number of voxels in the volume and they are typically very close to each other, especially in advanced developmental stages. As an analogy, imagine following tens of thousands of cars of the same exact model and color in a traffic jam. Separating them from one another at each time point and keeping track of the correct identity of each vehicle over time proves challenging, owing to the lack of visual cues. The closer the objects are to each other, the more challenging the problem becomes. Second, cells usually divide at different time points, which greatly increases the number of possible linking hypotheses that need to be considered and leads to a non-constant number of tracking targets over time (Fig. 5). In a temporally well-sampled dataset, cell divisions are typically rare events. This complicates their systematic detection owing to the risk of introducing false positives. At the same time, however, the correct identification of divisions is the single most important step needed to correctly reconstruct complete cell lineages. Third, it is important to keep in mind the requirement of scalability of the computational approaches, since typically on the order of tens of thousands of objects need to be tracked over thousands of time points. Finally, compared to other tracking applications, considerably higher accuracy is required to achieve a satisfactory result, since any single tracking mistake will affect an entire branch of the cell lineage tree. For example, even with an accuracy of 99.9% in linking nuclei between consecutive time points (and thus a random error rate of only 0.1%), 10% of all cell lineages are affected by a tracking mistake after only 100 time points, which corresponds to <1 h of development in a typical *Drosophila* SiMView recording. None of the approaches presented in the following paragraphs are close to this level of accuracy and the problem is thus still effectively unsolved. However, the advances discussed here indicate that, in coming years, a complete reconstruction of development will be within reach for several key biological model organisms.

Contour evolution methods

Nuclei tracking algorithms can be divided into the following two categories: contour evolution and data association (Fig. 6). Contour evolution methods assume a contour (or segmentation) is available for each nucleus at the initial time point. If temporal resolution is high enough with respect to the time scales involved in the cellular dynamics, those initial contours can be used as an initial segmentation for the next time point (Fig. 6). The correct contours for the current time point are then obtained by minimizing an energy function. Once the solution is found for time t, the procedure is repeated sequentially for time $t + 1$. Two key advantages of such methods are that they perform segmentation and tracking simultaneously and can also be applied to membrane labels. However, sufficiently high temporal sampling is a key prerequisite. As a rule of thumb, nucleus displacements between consecutive time points should be less than the nucleus diameter to guarantee spatial overlap. This condition depends on many factors and the required time interval can range, for example, from 30 s in early *Drosophila* embryogenesis (Tomer *et al.* 2012) to several minutes in mouse embryonic development (Nowotschin *et al.* 2010).

Non-parametric contour evolution

Within the contour evolution methods, two main subcategories can be distinguished: non-parametric and

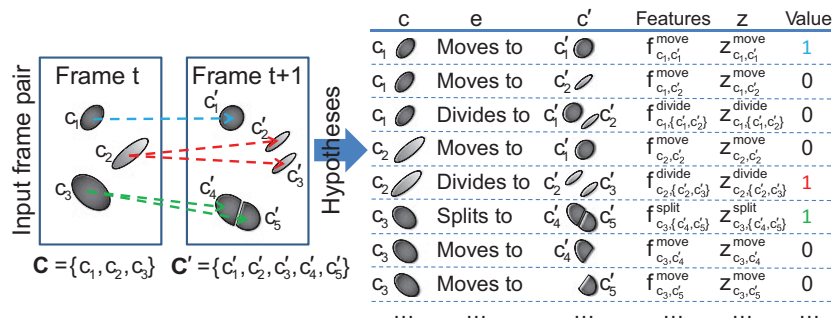| c	e	c'	Features	z	Value
c_1	Moves to	c_1'	$f_{c_1,c_1'}^{move}$	$z_{c_1,c_1'}^{move}$	1
c_1	Moves to	c_2'	$f_{c_1,c_2'}^{move}$	$z_{c_1,c_2'}^{move}$	0
c_1	Divides to $c_1' c_2'$		$f_{c_1,\{c_1',c_2'\}}^{divide}$	$z_{c_1,\{c_1',c_2'\}}^{divide}$	0
c_2	Moves to	c_1'	$f_{c_2,c_2'}^{move}$	$z_{c_2,c_2'}^{move}$	0
c_2	Divides to $c_2' c_3'$		$f_{c_2,\{c_2',c_3'\}}^{divide}$	$z_{c_2,\{c_2',c_3'\}}^{divide}$	1
c_3	Splits to $c_4' c_5'$		$f_{c_3,\{c_4',c_5'\}}^{split}$	$z_{c_3,\{c_4',c_5'\}}^{split}$	1
c_3	Moves to	c_4'	$f_{c_3,c_4'}^{move}$	$z_{c_3,c_4'}^{move}$	0
c_3	Moves to	c_5'	$f_{c_3,c_5'}^{move}$	$z_{c_3,c_5'}^{move}$	0
...	...	...	...	...	...

Fig. 5. Linking hypothesis between consecutive time points. Toy example: two sets of object candidates, and a small subset of the possible association hypotheses. One particular interpretation of the scene is indicated by colored arrows (left) or equivalently by a configuration of binary indicator variables z (rightmost column in table). Credits: Figure reprinted from Lou & Hamprecht (2011), Copyright (2011).
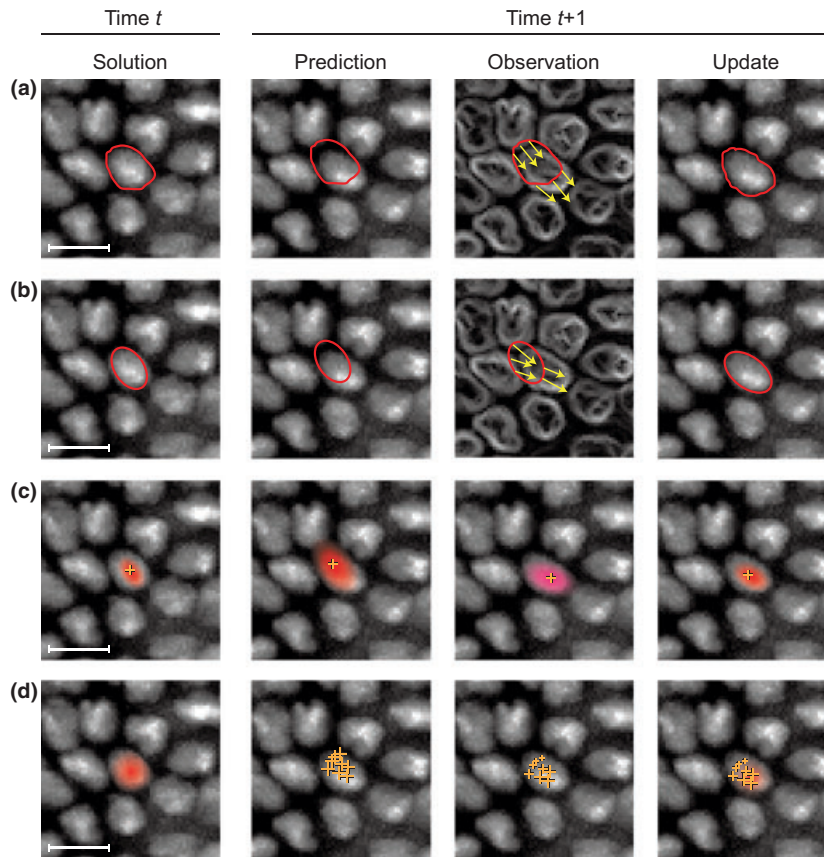
Fig. 6. Tracking algorithms. Illustration of different tracking algorithms. Rows correspond to different algorithms, columns show common conceptual steps. Images show a sub-region of the recording from Figure 1c. (a) Non-parametric contour evolution method. Each nucleus is segmented using a set of points defining a closed contour. The contour at time $t + 1$ is predicted based on the solution at time t and a cell motion model (second column). Based on the image gradient, each of the points of the contour is then pulled to an edge (third column, yellow arrows). Thereby, the prediction is updated to properly fit the nucleus at time $t + 1$. This scheme is sequentially performed over consecutive time points to track all objects. (b) Parametric contour evolution method. The same principles as in (a) apply. The only difference is that the contour is defined by a parametric shape (in this case an ellipsoid), which constrains the possible set of contours. (c) Kalman filter. The position of each nucleus is defined by a centroid (mean, orange cross) and a measure of the uncertainty of its location (covariance, red gradient ellipsoid). The measure of uncertainty is a key difference with respect to methods in (a) and (b). The mean and covariance define a normal distribution that needs to be updated for every time point. As in (a) and (b), the position at time $t + 1$ is predicted based on the solution at time t using a cell motion model. This prediction is compared to the likelihood that a nucleus is located at a certain point based on image features (third column, pink gradient ellipsoid), in order to obtain the final solution (fourth column, red gradient ellipsoid). (d) Particle filter. This is a generalization of the Kalman filter to extend the method beyond normality assumptions (process cannot be modeled just with a mean and covariance). A set of particles (orange crosses) is associated with each nucleus. The weighting of these particles (cross size) indicates how likely they are to represent the center of the nucleus. Each of the particle positions is updated as in (c). Then, centroid and uncertainty are updated using a weighted histogram of all particles. Scale-bars: 5 μm (a–d).

parametric methods. The first category encompasses classic image processing methods such as level-sets and active contours (Sethian 1999; Mosaliganti *et al.* 2009; Delgado-Gonzalo *et al.* 2012). Briefly, the contour of each object is described as a set of boundary voxels without any specific shape. The boundary is found by minimizing an energy function that balances shape constraints, such as average curvature, with image constraints, such as absolute intensity or intensity gradients. Li *et al.* presented work on cell lineage reconstruction from two-dimensional (2D) images of stem cells (Li *et al.* 2008) using phase contrast time-lapse microscopy data in different cell populations with different densities (from 100 to 5000 cells per frame). They achieved on average 90% of pairwise linkage accuracy between consecutive time points and 88% cell division detection accuracy, which translated to a fraction of 68% correct lineages. The authors also developed fast algorithms in order to be able to process each frame with thousands of cells in under a

minute. One of the main advantages of non-parametric approaches is that cell divisions are directly integrated into the framework, since curves or surfaces are allowed to merge and split if the image content seems to indicate such geometrical arrangements. Unfortunately, when objects are too densely packed, they tend to merge too much and additional heuristics are required to correct this bias (Li *et al.* 2008). Another potential drawback of this sequential approach is that errors can accumulate over time, since time point *t* is used as an initialization for time point *t* + 1. For this reason, most of the related computational approaches include some mechanism that attempts to correct obvious mistakes before analyzing the next time point.

Parametric contour evolution

As the name indicates, parametric approaches define the surface (or contour) of each object explicitly using a set of parameters. Since nuclei have a blob-like shape, an ellipsoid (nine parameters in 3D) is a good compromise between shape information and degrees of freedom (Tomer *et al.* 2012). In the case of membrane markers, where shapes can be much more irregular, spherical harmonics can be used to parameterize irregular closed surfaces with a few dozen parameters (Khairy *et al.* 2008). The advantage of this family of approaches versus non-parametric ones is that it is typically easier to find the optimal set of values to fit each shape to an object in the image. Unfortunately, it is considerably harder to handle topological changes (such as cell divisions), since the framework does not directly allow changing the number of objects from time point to time point. We previously presented a method that models each time point as a mixture of Gaussians (Tomer *et al.* 2012), that is, each nucleus in the volume is an ellipsoid, and the maximum likelihood solution is found by variational methods (Bishop 2007). In order to handle cell divisions, a machine learning classifier based on image features was used to detect when a single Gaussian volume contained multiple nuclei. In that case, the Gaussian was split into two ellipsoids to account for both daughter cells. The method was tested in early stages of *Drosophila* embryogenesis to track over 3000 cells for 140 time points (Fig. 3). The reported pairwise linkage accuracy between consecutive time points was 94% and the cell division detection accuracy was also 94%, which provided approximately 70% of correct lineages through two cycles of mitotic waves. The algorithm was implemented on a general purpose graphics processing unit (GPGPU), allowing 100-fold speed up and the possibility of tracking and segmenting more than 6000 nuclei in under a minute. As is the case with non-parametric methods, individual Gaussians tend to encompass multiple nuclei when these are too close to each other, owing to the effectively missing intensity separation between the nuclei in such cases.

Parametric contour evolution approaches can also be seen as a subset of state-space models. Classical tracking approaches such as Kalman filters (Kalman 1960) and particle filters (Doucet *et al.* 2010) fall into this category. In the case of nuclei labels, the state can include variables such as position, speed, intensity, cell phase, etc. The main idea is to define a set of values that describe a state for each existing nucleus at a given time point and recursively update the state for each time point based on the previous estimate (Markovian assumption). The algorithm proceeds in two steps: prediction and update. The prediction step tries to guess where the object will be in the next time point based on a motion model supplied by the user. The update step adjusts the guessed state from the prediction based on the observed image data in the new time point. However, in general, the update and prediction steps do not have a closed analytical form (the Kalman filter is an exception) and computationally expensive probabilistic methods have to be employed to find a solution. Lack of computational scalability to thousands of objects has been one of the main reasons why these methods have not been applied to cell lineage reconstructions yet, since heuristics or restrictive independence hypotheses between objects are required in order to avoid exponentially large global state spaces. Another reason is the fact that cellular motion models in complex multicellular organisms are hard to define in mathematical terms and can lead to guesses in the prediction step too remote from the true state, effectively making the update lose the object. Generally, Interacting Multiple Models (IMM) are used in order to try to model cellular dynamics without increasing complexity. Briefly, a set of possible motion models are described, and the algorithm then switches between these models according to a set of transition probabilities and observations at different time points. However, if switches between dynamics occur too fast (such as during mitosis), the algorithm might lack behind and the prediction step might not be sufficiently accurate. Meijering *et al.* (2012) and Smal *et al.* (2008) proved the power of these methods in particle tracking for fluorescent microscopy, especially in low SNR environments, since models can be very flexible and can easily incorporate *a priori* knowledge about a particular dataset and the noise characteristics in the data.

Data association

Tracking methods of the second class, which we here refer to as data association, approach the tracking problem from a combinatorial perspective and they do not integrate segmentation and tracking in a single step as contour evolution methods do. Given a set of objects at time t (obtained by some segmentation method) and a set of objects at time $t + 1$, these methods try to find the best possible match between the two sets. Thus, they only solve the linkage problem between time points. Figure 5 illustrates the typical hypotheses considered in this scenario for each object at time t: displacement, division, death, birth and merging or splitting due to errors in the segmentation. The total number of hypotheses to consider for each object is large and one has to be careful in designing the algorithm to avoid exponential complexity growth of the combinatorics, considering that thousands of objects need to be matched. A simple first step towards avoiding this effect is to limit the number of possible matches per object. Since nuclei have a maximum velocity and the temporal resolution of the data is known, a sphere can be drawn around each cell defining the maximum displacement. Any object at time $t + 1$ outside this sphere will not be considered for matching.

The program StarryNite represents one of the first successful computational attempts in this area, reconstructing *Caenorhabditis elegans* cell lineages up to the 350-cell stage (Bao *et al.* 2006). The authors matched objects using spatial nearest neighbors and some *ad hoc* rules to handle cell divisions, such as customized shape and intensity descriptors during mitotic events. The approach is very fast and successful during the initial stages, where linkage accuracy higher than 99% is achieved up to the 194-cell stage. However, the accuracy of this approach decreases rapidly for later stages when cell density increases, dropping to 97% at the 350-cell stage. Recently, the open-source software NucleiTracker4D by Giurumescu *et al.* (2012) increased the accuracy to >99% at the 350-cell stage by combining automatic tracking with a user-friendly Graphical User Interface (GUI), which allows curating ∼4,000 nuclei tracking linkages per day. Thus, a complete accurate *C. elegans* lineage reconstruction can be obtained in three weeks.

The idea of nearest neighbor and shape descriptors can be formalized and extended using well-established graph-matching algorithms and convex optimization techniques. Bise *et al.* (2011) and Kausler *et al.* (2012) presented work using this formalism. The main idea is sketched schematically in Figure 5. Briefly, a cost f is assigned to each possible hypothesis for each object.

Then, the assignment that minimizes the added cost of all selected hypotheses needs to be found while ensuring compatibility of these hypotheses. For example, a cell cannot die and divide at the same time, so only one of these hypotheses can be selected at any given time. The advantage of these methods is that neighboring objects share hypotheses, and decisions made in one object affect the set of possible decisions made in a neighboring object. Thus, these algorithms effectively use spatial contextual information to make linking decisions. Another advantage is that they can be formulated as convex linear integer problems, which have been studied for decades in the optimization literature (Schrijver 1998) and freely available efficient solvers exist for academic institutions (such as the ILOG CPLEX Optimization Studio, IBM). Kausler *et al.* (2012) used this framework in early *Drosophila* stages and tuned the model to tolerate a large number of false detections due to autofluorescence. They tracked approximately 256 cells over 40 time points that comprised two mitotic waves, achieving 96% accuracy in tracking and 92% accuracy in cell division detection. The total run time was on the order of minutes.

Learning the best match

Aside from the list of hypotheses, one of the key considerations in graph-matching algorithms is the definition of costs f, since these effectively determine the optimal solution. However, the optimal solution in the model will not agree with the true solution unless the costs are set appropriately. Many studies manually define and tune the costs based on observations of the person analyzing the data (Bao *et al.* 2006; Jaqaman *et al.* 2008). The most common choice is some type of weighted average between cell displacement, shape descriptors and/or expression levels. If temporal sampling is high enough, displacement is certainly the most useful feature. For more complex scenes, Lou and Hamprecht presented a machine learning approach in order to learn the optimal cost values based on annotated data (Lou & Hamprecht 2011). The idea is to define a cost as a weighted linear combination of all possible features and to use a machine learning algorithm similar to support vector machines to calculate the optimal weight for each feature. As with all machine learning techniques, the caveat is that the training set has to be representative of all types of dynamics present in the data. This task can demand a significant effort in 3D+t datasets of complex multicellular organisms. Moreover, when exploring the data not all the dynamics might be known *a priori* and the training set might thus miss certain types of dynamics.

Following similar concepts, Huh *et al.* (2011) and Liu *et al.* (2012) presented a machine learning approach specifically target at detecting mitotic events based on temporal and image features. For example, when nuclei divide, the chromatin becomes compacted and fluorescence levels increase. Again, mitosis is typically a rare event, which makes it hard to identify with high confidence, but it is the most important event in the context of cell lineage reconstructions. The authors showed that the machine learning approach improved cell division detection accuracy from 88 to 96% in 2D phase contrast time-lapse microscopy datasets of stem cells.

Global tracking: beyond pairwise matching

Up to this point, we discussed approaches to matching individual objects between two consecutive time points. The same idea can be extended to the matching of partial lineages (known as *tracklets* in the computer science literature), where the end of one partial lineage and the beginning of another partial lineage are in neighboring time points (Jaqaman *et al.* 2008; Bise *et al.* 2011). In this scenario, the matching cost *f* applies between two partial lineages instead of two objects. This adds temporal contextual information, making the approach more global and robust. One of the main advantages is that the number of tracklets is much smaller than the number of cells. Thus, algorithm complexity and run-time is lower. However, defining the matching cost might be more complex and it will require larger training datasets if using machine learning approaches. Finally, the tracklets can be constructed with any of the algorithms reviewed in this manuscript for tracking in consecutive time points. The user simply needs to set the parameters in a conservative manner, such that only obviously associated nuclei are linked together. Bise *et al.* (2011) presented work based on tracklets, improving their tracking accuracy by 28% with respect to their previous work in the same 2D stem cell phase-contrast microscopy time-lapse datasets.

Visualization and editing

Complementing the automated computational approaches to synthetizing image data into digital reconstructions of cellular dynamics, visualization and editing tools are required to study specific biological processes and to validate and correct errors in the computational results. A good approach to data visualization is essential to interpreting the data and determining the main sources of errors in the computational reconstruction, which in turn allows improving the

algorithms. Previous projects that required sorting through large amounts of data, such as connectomics datasets (Bock *et al.* 2011; Peng *et al.* 2011), cellular atlases of model organisms (Long *et al.* 2009) or complete cell lineage reconstructions, showed that interacting with the data can often become the main bottleneck. In other words, it is crucial to identify the most suitable strategy to visualize terabytes of 3D+*t* data, and to superimpose the computational results with raw data such that important information does not become occluded and results can be efficiently edited and annotated.

Very efficient interfaces exist for 2D+*t* datasets, since 2D images can be easily stacked or displayed as movies without losing information. Cordeli (2004) and Winter *et al.* (2011) provide freely available software tools for effective visualization and editing of cell tracking and cell lineage results. However, the extension to 3D+*t* is not straight-forward for two main reasons. First, four-dimensional (4D) datasets are difficult to visualize efficiently in a 3D world. Second, the amount of data easily increases by several orders of magnitude, which pushes the performance limits even of the most advanced computer workstations. The different approaches described in the following paragraphs try to solve these problems using different models (Fig. 7).

Megason developed an open source project, GoFigure 2 (Megason 2009), which is specifically aimed at the analysis of cell lineages using 3D time-lapse light microscopy (Fig. 7j). The system uses a database to store and retrieve all lineage information (contours, shapes, volumes, linkages, etc.) and defines a hierarchy between objects in order to make data navigation more user-friendly. 2D contours for each nucleus or membrane are grouped into 3D meshes. These 3D meshes are grouped into tracks and each set of tracks is grouped into a lineage. Murray *et al.* (2006) developed an open-source software for 3D lineages named AceTree, which has been used in combination with StarryNite to build complete cell lineages for *C. elegans* (Fig. 7i). The software Simi Biocell is also specifically designed for manual cell lineage reconstruction using a graphical interface (Schnabel *et al.* 1997).

Vaa3D (Peng *et al.* 2010) (Fig. 7a) and Icy (Chaumont *et al.* 2012) (Fig. 7b–h) offer powerful visualization engines for 3D+*t* datasets with plug-in frameworks that allow users to develop their own tools for interacting with the data. These engines also offer the option of overlapping and interacting with 3D geometrical objects that can be used to represent segmentation data and tracking information between time points. Both tools are open source and freely available for all platforms. These approaches are less specialized since their visualization engines can be used in the context
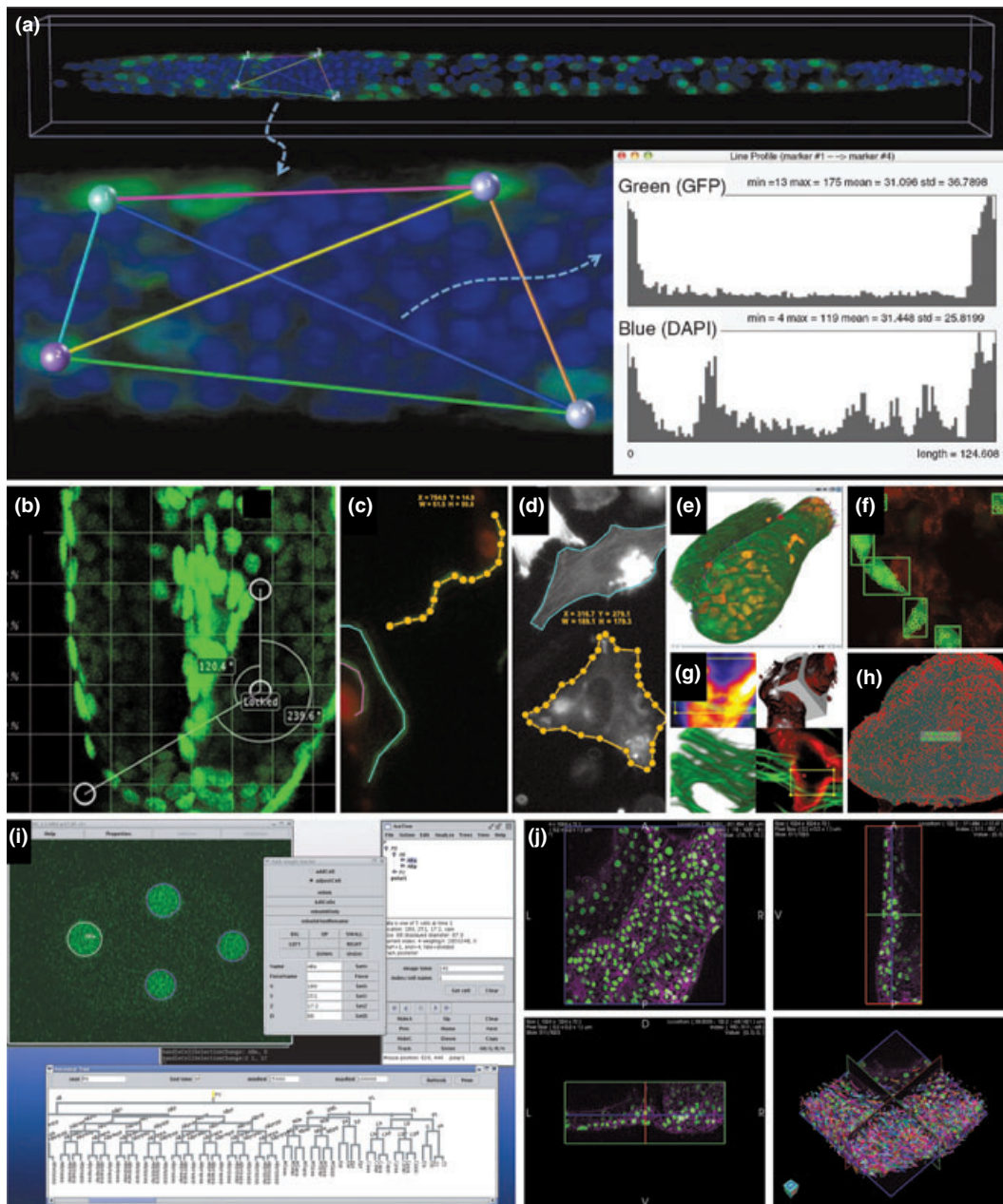
Fig. 7. Visualization and editing software. (a) Segmentation, annotation and quantitative measurement of gene expression levels in a *Caenorhabditis elegans* confocal microscopy image with Vaa3D. (b–h) Displaying microscopy images, segmentation masks and tracking information with the Icy visualization interface. (i) Editing cell lineages with AceTree. The panel shows the "Editing", "Lineage" and "Add One" windows. (j) Screenshot of manual segmentation and tracking tools in GoFigure 2. Credits: Panel (a) was reprinted from Peng *et al.* (2010), Copyright (2010), with permission from Macmillan Publishers Ltd. Panels (b–h) were reprinted from Chaumont *et al.* (2012), Copyright (2012), with permission from Macmillan Publishers Ltd. Panel (i) was reprinted from Murray *et al.* (2006), Copyright (2006), with permission from Macmillan Publishers Ltd. Panel (j) was kindly provided by and used with permission from Sean Megason (Harvard Medical School).

of many different applications. Unfortunately, no plug-ins exist yet to handle entire cell lineages. Tools such as Imaris (Bitplane) (Fig. 3c) and Volocity (Perkin Elmer) represent commercial alternatives. Both feature 3D+*t*

visualization engines and incorporate specific segmentation and lineage editing tools in 3D that enable the reconstruction of partial lineages in localized areas in time and space (McMahon *et al.* 2008). However,

these software packages tend to load all data into the memory of the workstation at once, which substantially limits the size of the dataset that can still be processed.

Originating from the connectomics field, where massive amounts of microscopy data need to be segmented and tracked, the open-source platform Collaborative Annotation Toolkit for Massive Amounts of Image Data (CATMAID) is designed to navigate, share and collaboratively annotate very large image datasets of biological specimens (Saalfeld *et al.* 2009). The interface is inspired by Google Maps, which stores data at different levels of resolution and loads only the currently analyzed spatio-temporal region into memory, such that datasets of unlimited size can be browsed efficiently. Currently, CATMAID is being extended to support time-lapse image data as well. One of its most important features is the possibility of editing and visualizing the same dataset by multiple users at the same time through a browser connecting to a centralized server that handles most of the computational load.

Sharing tools and ground truth annotations of key datasets will be crucial for the future progress of the entire field (Cardona & Tomancak 2012). At the same time, it is important to note that almost each microscopy technique and each model system are accompanied by their own specific challenges and optimal results can only be obtained by customizing and adapting approaches to the respective data.

Conclusions and future work

We discussed the main steps required for systematic large-scale analyses of cellular dynamics and cell lineages in complex multi-cellular organisms. The success of such studies requires that each step in the pipeline, from microscopy, via image processing to data visualization, is carefully designed and optimized for the respective task. As far as data collection is concerned, light sheet-based microscopy is emerging as the technology of choice for the problem at hand. Low photo-toxicity and photo-damage, combined with high recording speeds and high SNR, make light sheet microscopes the ideal tool for *in toto* imaging of the development of complex multicellular organisms at single-cell resolution. Image data quality will improve even further in the near future, as a result of ongoing efforts towards the development of new strategies for addressing optical aberrations in complex specimens, improving optical penetration and increasing spatio-temporal resolution with minimal energy load on the specimen.

Several computational processing and visualization tools designed specifically for the systematic reconstruction and quantitative analysis of cell lineage information are under development. High computational accuracies can already be achieved with existing methods in specific scenarios, and further efforts in computational tool development are needed to finally achieve the accuracy required to perform complete lineage reconstructions for different model systems and throughout their development. The most important metric that has to be minimized is the proof-reading time required to maximize the quality of reconstructions comprising millions of database entries. One of the major challenges towards this goal is the efficient handling and visualization of the vast amount of information in the raw microscopy recordings and the resulting computational reconstructions.

Considering the performance of state-of-the-art techniques and the current trajectory of their development, we believe that complete cell lineage reconstructions in complex multicellular organisms, such as the fruit fly, zebrafish and mouse, are within reach in coming years. The availability of such data and technology will open the door to fundamentally new approaches and questions in quantitative developmental biology.

Acknowledgments

Box 1

Summary of technical terms

Point spread function: mathematical description of the image formed by a microscope when the observed object can be considered a single point in space. The point spread function (PSF) characterizes the resolution of the microscope.

Dwell time: time interval, over which the microscope detects and integrates signal in the currently recorded volume element. For example, in point-scanning confocal or two-photon microscopy, the dwell time corresponds to the amount of time the laser beam illuminates the focal volume corresponding to a single pixel in the final image, before moving on to the next volume element. Longer dwell times lead to higher signal-to-noise ratio, but also reduce speed and increase photo-damage.

Multi-photon fluorescence microscopy: optical microscopy technique that uses a non-linear fluorescence excitation

mode to achieve optical sectioning as well as deeper penetration into biological tissues.

Structured illumination: optical microscopy technique that uses patterns of light for specimen illumination. Two common types include incoherent structured illumination, which allows enhancing image contrast in light-scattering samples, and coherent structured illumination, which can be used to increase resolution beyond the diffraction limit.

Bessel beam: beam with a central peak surrounded by a concentric ring system. The central peak of the Bessel beam is thinner than the Gaussian beam typically used in light-sheet microscopy. When suppressing the contribution of the Bessel beam's ring system to the recorded image, for example, by multi-photon excitation or structured illumination, a Bessel beam light-sheet microscope can achieve higher axial resolution than a conventional light-sheet microscope.

Image registration: computational task of aligning two or more images with respect to each other by finding common features.

Deconvolution: computational task of correcting for the blurring effect arising from the point spread function of the microscope.

Segmentation: computational task of associating pixels in the same image that represent the same object.

Tracking: computational task of associating pixels or objects across different time points.

Parametric shape: mathematical description of the shape of an object based on an analytical formulation with few parameters. For example, an ellipsoid is a parametric shape in 3D with nine parameters.

Non-parametric shape: mathematical description of the shape of an object based on an exhaustive instead of an analytical formulation. For example, enumerating all the voxels in an image that belong to an ellipsoid.

Contour: type of non-parametric shape for 2D closed objects, such as cell membranes and nuclei. In this case, the user enumerates all the points along the boundary of the object in order to describe it.

Energy function: in image processing, this refers to a mathematical equation to model the problem at hand. The extremum (minimum or maximum) of this equation should correspond to the correct biological solution.

Level sets: non-parametric shape representation. The shape is described by all the points equal to a given value (usually zero) of a mathematical function, which allows great flexibility with respect to the type of shapes that can be represented.

Active contours or snakes: computational technique for fitting contours to images using an energy minimization approach.

Image feature: any information that can be extracted from the image and that can be represented by a single real number in order to compare its value in different regions of interest.

Machine learning classifier: mathematical function that, based on some input such as image features, predicts the correct output for a given task. For example, deciding if a cell is dividing or not. The function has free parameters that can be adjusted using examples given by the user (training data), effectively learning to model the given task.

Support vector machine: specific type of machine learning classifier, which has become very popular due to its ease of use and its applicability to a large spectrum of problems.

References

Bao, Z., Murray, J. I., Boyle, T., Ooi, S. L., Sandel, M. J. & Waterston, R. H. 2006. Automated cell lineage tracing in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **103**, 2707–2712.

Bise, R., Zhaozheng, Y. & Kanade, T. 2011. Reliable cell tracking by global data association. *ISBI* , 1004–1010.

Bishop, C. M. 2007. *Pattern Recognition and Machine Learning*. Springer, New York.

Bock, D. D., Lee, W.-C. A., Kerlin, A. M., Andermann, M. L., Hood, G., Wetzel, A. W., Yurgenson, S., Soucy, E. R., Kim, H. S. & Reid, R. C. 2011. Network anatomy and in vivo physiology of visual cortical neurons. *Nature* **471**, 177–182.

Cardona, A. & Tomancak, P. 2012. Current challenges in open-source bioimage informatics. *Nat. Methods* **9**, 661–665.

Chaumont, F. D., Dallongeville, S., Chenouard, N., Hervé, N., Pop, S., Provoost, T., Meas-Yedid, V., Pankajakshan, P., Lecomte, T., Montagner, Y. L., Lagache, T., Dufour, A. & Olivo-Marin, J.-C. 2012. Icy: an open bioimage informatics platform for extended reproducible research. *Nat. Methods* **9**, 690–696.

Cordeli, F. 2004. Manual Tracking plug-in for ImageJ. Available from URL: http://rsbweb.nih.gov/ij/plugins/track/track.html.

Delgado-Gonzalo, R., Chenouard, N. & Unser, M. 2012. Fast parametric snakes for 3D microscopy. *ISBI* 852–855.

Doucet, A., Freitas, N. D. & Gordon, N. 2010. *Sequential Monte Carlo Methods in Practice*. Springer, New York.

Eliceiri, K. W., Berthold, M. R., Goldberg, I. G., Ibáñez, L., Manjunath, B. S., Martone, M. E., Murphy, R. F., Peng, H., Plant, A. L., Roysam, B., Stuurmann, N., Swedlow, J. R., Tomancak, P. & Carpenter, A. E. 2012. Biological imaging software tools. *Nat. Methods* **9**, 697–710.

Fuchs, E., Jaffe, J., Long, R. & Azam, F. 2002. Thin laser light sheet microscope for microbial oceanography. *Opt. Express* **10**, 145–154.

Gao, L., Shao, L., Higgins, C. D., Poulton, J. S., Peifer, M., Davidson, M. W., Wu, X., Goldstein, B. & Betzig, E. 2012. Noninvasive imaging beyond the diffraction limit of 3D dynamics in thickly fluorescent specimens. *Cell* **151**, 1370–1385.

Giurumescu, C. A., Kang, S., Planchon, T. A., Betzig, E., Bloomekatz, J., Yelon, D., Cosman, P. & Chisholm, A. D.

2012. Quantitative semi-automated analysis of morphogenesis with single-cell resolution in complex embryos. *Development* **139**, 4271–4279.

Huh, S., Ker, D. F. E., Bise, R., Mei, C. & Kanade, T. 2011. Automated mitosis detection of stem cell populations in phase-contrast microscopy images. *TMI* **30**, 586–596.

Huisken, J., Swoger, J., Del Bene, F., Wittbrodt, J. & Stelzer, E. H. K. 2004. Optical sectioning deep inside live embryos by selective plane illumination microscopy. *Science* **305**, 1007–1009.

Jaqaman, K., Loerke, D., Mettlen, M., Kuwata, H., Grinstein, S., Schmid, S. L. & Danuser, G. 2008. Robust single-particle tracking in live-cell time-lapse sequences. *Nat. Methods* **5**, 695–702.

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82**, 35–45.

Kausler, B. X., Schiegg, M., Andres, B., Lindner, M., Köthe, U., Leitte, H., Wittbrodt, J., Hufnagel, L. & Hamprecht, F. A. 2012. A discrete chain graph model for 3D+t cell tracking with high misdetection robustness. *ECCV* 144–157.

Keller, P. J. & Dodt, H. U. 2011. Light sheet microscopy of living or cleared specimens. *Curr. Opin. Neurobiol.* **22**, 138–143.

Keller, P. J., Schmidt, A. D., Santella, A., Khairy, K., Bao, Z., Wittbrodt, J. & Stelzer, E. H. 2010. Fast, high-contrast imaging of animal development with scanned light sheet-based structured-illumination microscopy. *Nat. Methods* **7**, 637–642.

Keller, P. J., Schmidt, A. D., Wittbrodt, J. & Stelzer, E. H. K. 2008. Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. *Science* **322**, 1065–1069.

Khairy, K. & Keller, P. J. 2011. Reconstructing embryonic development. *Genesis* **49**, 488–513.

Khairy, K., Reynaud, E. & Stelzer, E. 2008. Detection of deformable objects in 3D images using Markov-Chain Monte Carlo and spherical harmonics. *Med. Image Comput. Comput. Assist. Interv.* **11**, 1075–1082.

Krzic, U., Gunther, S., Saunders, T. E., Streichan, S. J. & Hufnagel, L. 2012. Multiview light-sheet microscope for rapid in toto imaging. *Nat. Methods* **9**, 730–733.

Li, K., Miller, E. D., Chen, M., Kanade, T., Weiss, L. E. & Campbell, P. G. 2008. Cell population tracking and lineage construction with spatiotemporal context. *Med. Image Anal.* **12**, 546–566.

Liu, A., Li, K. & Kanade, T. 2012. A Semi-Markov model for mitosis segmentation in time-lapse phase contrast microscopy image sequences of stem cell populations. *IEEE Trans. Med. Imaging* **31**, 359–369.

Long, F., Peng, H., Liu, X., Kim, S. K. & Myers, E. 2009. A 3D digital atlas of *C. elegans* and its application to single-cell analyses. *Nat. Methods* **6**, 667–672.

Lou, X. & Hamprecht, F. A. 2011. Structured learning for cell tracking. *NIPS* 1296–1304.

McMahon, A., Supatto, W., Fraser, S. E. & Stathopoulos, A. 2008. Dynamic analyses of Drosophila gastrulation provide insights into collective cell migration. *Science* **322**, 1546–1550.

Megason, S. G. 2009. In toto imaging of embryogenesis with confocal time-lapse microscopy. *Methods Mol. Biol.* **546**, 317–332.

Megason, S. G. & Fraser, S. E. 2007. Imaging in systems biology. *Cell* **130**, 784–795.

Meijering, E., Dzyubachyk, O. & Smal, I. 2012. Methods for cell and particle tracking. *Methods Enzymol.* **504**, 183–200.

Mosaliganti, K., Smith, B., Gelas, A., Gouaillard, A. & Megason, S. 2009. Cell tracking using coupled active surfaces for nuclei and membranes. *Insight J.* Available at URL: http://www.insight-journal.org/browse/publication/642.

Murray, J. I., Bao, Z., Boyle, T. J. & Waterston, R. H. 2006. The lineaging of fluorescently-labeled *Caenorhabditis elegans* embryos with StarryNite and AceTree. *Nat. Protoc.* **1**, 1468–1476.

Nowotschin, S., Ferrer-Vaquer, A. & Hadjantonakis, A. K. 2010. Imaging mouse development with confocal time-lapse microscopy. *Methods Enzymol.* **476**, 351–377.

Olivier, N., Luengo-Oroz, M. A., Duloquin, L., Faure, E., Savy, T., Veilleux, I., Solinas, X., Debarre, D., Bourgine, P., Santos, A., Peyrieras, N. & Beaurepaire, E. 2010. Cell lineage reconstruction of early zebrafish embryos using label-free nonlinear microscopy. *Science* **329**, 967–971.

Palero, J., Santos, S. I., Artigas, D. & Loza-Alvarez, P. 2010. A simple scanless two-photon fluorescence microscope using selective plane illumination. *Opt. Express* **18**, 8491–8498.

Peng, H., Long, F., Zhao, T. & Myers, E. 2011. Proof-editing is the bottleneck of 3D neuron reconstruction: the problem and solutions. *Neuroinformatics* **9**, 103–105.

Peng, H., Ruan, Z., Long, F., Simpson, J. H. & Myers, E. W. 2010. V3D enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets. *Nat Biotech.* **28**, 348–353.

Perona, P. & Malik, J. 1990. Scale-space and edge detection using anisotropic diffusion. *TPAMI* **12**, 629–639.

Planchon, T. A., Gao, L., Milkie, D. E., Davidson, M. W., Galbraith, J. A., Galbraith, C. G. & Betzig, E. 2011. Rapid three-dimensional isotropic imaging of living cells using Bessel beam plane illumination. *Nat. Methods* **8**, 417–423.

Preibisch, S., Saalfeld, S., Schindelin, J. & Tomancak, P. 2010. Software for bead-based registration of selective plane illumination microscopy data. *Nat. Methods* **7**, 418–419.

Russ, J. C. 2011. *The Image Processing Handbook*, 6th edn. CRC Press.

Saalfeld, S., Cardona, A., Hartenstein, V. & Tomančák, P. 2009. CATMAID: collaborative annotation toolkit for massive amounts of image data. *Bioinformatics* **25**, 1984–1986.

Schnabel, R., Hutter, H., Moerman, D. & Schnabel, H. 1997. Assessing normal embryogenesis in *Caenorhabditis elegans* using a 4D microscope: variability of development and regional specification. *Dev. Biol.* **184**, 234–265.

Schrijver, A. 1998. *Theory of Linear and Integer Programming*. Wiley, New York.

Sethian, J. A. 1999. *Level Set Methods and Fast Marching Methods*. Cambridge University Press, Cambridge.

Siedentopf, H. & Zsigmondy, R. 1903. Über Sichtbarmachung und Größenbestimmung ultramikroskopischer Teilchen, mit besonderer Anwendung auf Goldrubingläser. *Ann. Phys.* **315**, 1–39.

Smal, I., Meijering, E., Draegestein, K., Galjart, N., Grigoriev, I., Akhmanova, A., van Royen, M. E., Houtsmuller, A. B. & Niessen, W. 2008. Multiple object tracking in molecular bioimaging by Rao-Blackwellized marginal particle filtering. *Med. Image Anal.* **12**, 764–777.

Stone, L. D., Corwin, T. L. & Barlow, C. A. 1999. *Bayesian Multiple Target Tracking*. Artech House, London.

Swoger, J., Muzzopappa, M., Lopez-Schier, H. & Sharpe, J. 2011. 4D retrospective lineage tracing using SPIM for zebrafish organogenesis studies. *J. Biophotonics* **4**, 122–134.

Swoger, J., Verveer, P., Greger, K., Huisken, J. & Stelzer, E. H. 2007. Multi-view image fusion improves resolution in three-dimensional microscopy. *Opt. Express* **15**, 8029–8042.

Temerinac-Ott, M., Ronneberger, O., Ochs, P., Driever, W., Brox, T. & Burkhardt, H. 2012. Multiview deblurring for 3-D images from light-sheet-based fluorescence microscopy. *TIP* **21**, 1863–1873.

Tomer, R., Khairy, K., Amat, F. & Keller, P. J. 2012. Quantitative high-speed imaging of entire developing embryos with simultaneous multiview light-sheet microscopy. *Nat. Methods* **9**, 755–763.

Tomer, R., Khairy, K. & Keller, P. J. 2011. Shedding light on the system: studying embryonic development with light sheet microscopy. *Curr. Opin. Genet. Dev.* **21**, 558–565.

Truong, T. V., Supatto, W., Koos, D. S., Choi, J. M. & Fraser, S. E. 2011. Deep and fast live imaging with two-photon scanned light-sheet microscopy. *Nat. Methods* **8**, 757–760.

Voie, A. H., Burns, D. H. & Spelman, F. A. 1993. Orthogonal-plane fluorescence optical sectioning: three-dimensional imaging of macroscopic biological specimens. *J. Microsc.* **170**, 229–236.

Winter, M., Wait, E., Roysam, B., Goderie, S. K., Ali, R. A. N., Kokovay, E., Temple, S. & Cohen, A. R. 2011. Vertebrate neural stem cell segmentation, tracking and lineaging with validation and editing. *Nat. Protocols* **6**, 1942–1952.