Deciphering tissue morphodynamics using bioimage informatics

Alexandre C. Dufour\textsuperscript{1,3}, Anneliene H. Jonker\textsuperscript{2} and Jean-Christophe Olivo-Marin\textsuperscript{1,3}

\textsuperscript{1}Institut Pasteur, Bioimage Analysis Unit, 25-28 rue du Docteur Roux, Paris, France
\textsuperscript{2}INSERM US14, 96 rue Didot, Paris, France
\textsuperscript{3}CNRS, UMR 3691, 25-28 rue du Docteur Roux, Paris, France

In recent years developmental biology has greatly benefited from the latest advances in fluorescence microscopy techniques. Consequently, quantitative and automated analysis of this data is becoming a vital first step in the quest for novel insights into the various aspects of development. Here we present an introductory overview of the various image analysis methods proposed for developmental biology images, with particular attention to openly available software packages. These tools, as well as others to come, are rapidly paving the way towards standardized and reproducible bioimaging studies at the whole-tissue level. Reflecting on these achievements, we discuss the remaining challenges and the future endeavours lying ahead in the post–image analysis era.

This article is part of the themed issue 'Systems morphodynamics: understanding the development of tissue hardware'.

1. Introduction

The past decade has witnessed remarkable progress in the field of developmental biology, in part due to the revolution in microscopy imaging of living systems \cite{1}. Recent advances in microscopy techniques, such as multi-photon laser scanning microscopy (MSLM) and selective plane illumination microscopy (SPIM) \cite{2–7} offer an unprecedented insight into the complexity of development, giving access to different facets of cell behaviour within their surrounding tissue microenvironment \cite{8}. Thanks to the continuous improvement and generalization of these imaging systems (either commercially or via collaborative or open developments \cite{4}), numerous mechanisms underlying tissue development can now be observed in a reproducible manner, opening the way to robust phenotype-based analyses, as well as biophysical and biomechanical modelling of morphogenesis. Meanwhile, these developments have rapidly engendered a substantial surge in the volume of imaging data produced in routine, easily reaching several Terabytes per experiment. To make sense out of this colossal amount of information, manual inspection of, for example, cell shape and motility patterns at the tissue level is no longer a viable option, let alone statistically relevant quantification across multiple samples and experimental conditions. Consequently, the challenge of quantitative developmental biology has reached an analytical bottleneck, calling for the development of computerized analysis tools able to transform these raw data into biologically interpretable information in a statistically robust and automated manner. The bioimage informatics community is already hard at work on these issues, actively and continuously investigating novel strategies and developing new algorithms to optimize every step of the so-called ‘information workflow’, from smart acquisition to storage, visualization, enhancement and quantification, with the ultimate goal to facilitate biological interpretation. The impact of these algorithmic developments on developmental biology has become particularly visible in recent years, notably thanks to the latest advances in computing architectures able to handle and process massive...
multi-dimensional datasets, together with the progressively systematic release of computer codes through open collaborative software platforms [9].

This review discusses the problems, solutions and remaining challenges in the quest for quantitative tissue analytics from the point of view of bioimage informatics, with a particular focus on the primordial tasks of cell segmentation and tracking, for they are crucial first steps to decipher the mechanisms by which cells form and shape tissue throughout development. While an extensive review of cell segmentation and tracking methods would fall beyond the scope of this review, here we specifically focus on methods that have been effectively applied to the problem of tissue analysis. The remainder of this review is organized as follows. In a first step, we recognize the importance and impact of image quality on the subsequent analysis, and discuss current developments to enhance bioimaging data after acquisition. Next, we tackle the issue of cell segmentation in the specific case of dense three-dimensional tissue. We then address the question of cell tracking, noting the computational challenge of tracking several thousands of cells over time simultaneously. Finally, we review a number of open-software solutions implementing some of the methods discussed here, and conclude by exploring the remaining bottlenecks and future challenges towards truly quantitative tissue biology.

2. Improving bioimaging data during and after acquisition

While it is well established in the bioimage informatics community that image quality has a dramatic impact on its analysis, the notion of image quality remains quite subjective to many outside the field. In the digital imaging world, the quality of an image is generally measured using mathematical criteria such as the so-called signal-to-noise ratio (SNR), as well as other descriptors that locally characterize the image ‘texture’ (leading to measures of, for example, contrast or homogeneity) [10]. More recently, an image-centric measure known as the SSIM (for structural similarity index measure) was devised to measure the quality of an image in comparison to some reference or ‘ground-truth’ image (the availability of which is problematic in microscopy). While many of these measures coincide with our visual, human appreciation of what is a ‘good’ image, they are rather poor indicators of whether the image is actually suitable for subsequent (computerized) analysis [11]. More often than not, the acquisition parameters should be defined and optimized alongside the analysis pipeline itself in order to ensure proper quantification. In some instances, this optimization process may additionally benefit from re-thinking the sample preparation itself, ensuring that the acquisition and analysis are well adapted to the biological question (e.g. a nuclear reporter does not permit optimal cell segmentation, but is better suited for cell counting, notably in dense tissue).

After acquisition, image quality can usually be improved using a range of signal processing methods known as denoising tools. Unfortunately, fluorescence microscopy images often exhibit numerous different artefacts due to the obligatory compromise between image quality and cell viability, and therefore require denoising algorithms that are tailored to each and every case. For instance, excitation of fluorescent dyes can cause toxicity and bleaching, imposing heavy constraints on exposure time and therefore limiting global contrast. Moreover, insufficient excitation (as well as uneven fluorescent labelling) may induce gaps and non-homogeneities within the structures of interest, thus considerably hampering their extraction. Finally, imaging in thick tissue is usually affected by light scattering, causing the SNR to progressively decrease as the illumination source traverses the sample. In most instances, so-called ‘pre-processing’ algorithms can be used to reduce these artefacts in order to facilitate the subsequent analysis steps.

Commonly used pre-processing methods for tissue imaging can be grouped into one of four categories:

1. **Fusion** methods combine data from multiple acquisitions of the same sample, thereby reducing the background noise or the impact of light scattering through the tissue. These methods range from simple line or frame averaging to more complex multi-view fusion of varying sample poses, notably in SPIM [4,7,12–14].

2. **Deconvolution** methods aim at improving image resolution by reducing the artefacts and aberrations introduced by the optical device, defined by its point-spread function (PSF). The PSF not only depends on the imaging technique, but also varies in space and depth, notably in thick or scattering samples, thus raising a significant challenge for the signal processing community [15–17].

3. **Contrast enhancement** methods are used to homogenize signal distribution over the entire dataset (examples include image normalization and histogram equalization). They can be particularly useful in a three-dimensional context to correct the progressive intensity decrease caused by light scattering in deeper sections of the sample [18–20].

4. **Filtering** algorithms are used to reduce image noise (locally or globally) using either conventional low-pass or median filters, or by specifically enhancing or correcting heterogeneities of the structures of interest based on local texture information [18,19,21–24].

Image pre-processing can be of crucial importance to facilitate the subsequent analysis steps, so long as the chosen method is adapted to the specimen and optical device at hand. Nevertheless, it is worth stressing that these techniques apply some transformation that directly affects the pixel values (and thereby the dynamic range) of the original dataset. In some instances, such pre-processed datasets may no longer be usable to extract biological information that directly depends on carefully calibrated intensity values (e.g. protein expression levels). For such sensitive applications, great care should be taken in ensuring that the data are knowingly and coherently adjusted.

3. Cell and tissue segmentation methods

Image segmentation is one of the cornerstones of digital image analysis, and describes the process of separating an image into different meaningful parts or *segments*. Novel methods are constantly being developed to tackle the new and ever more complex challenges raised by novel or diversifying bioimaging techniques. Cell segmentation therefore remains today one of the most active research topics in the bioimage informatics community, even after decades of investigation [25,26].
In the context of tissue analysis, cell segmentation methods fall into one of two categories. In studies concerned with the global localization of cells within tissue (and eventually their lineage throughout morphogenesis), there is only minor interest in extracting the actual shape of each individual cell. Therefore, the cell segmentation problem is reduced to that of finding the cell nuclei, which is arguably simpler in tissue and model organisms where nuclei appear well separated [27,28] (the problem becomes considerably more complex as tissue density increases [18,29–31]). In other studies where either the cell shape or the local neighbourhood is required (for instance to characterize cell–cell interaction and intercalation), the full cell outline must be extracted, and therefore imposes that the cell membranes be imaged. In contrast to nuclear labelling, fluorescent signal is usually less homogeneous along the cell membrane (notably due to poor resolution and light scattering), therefore segmentation is particularly more challenging even after suitable pre-processing [32,33], and often requires the help of nuclear localization to initialize the extraction process [18,20,31].

For nuclear segmentation, a straightforward approach lies in intensity-based pixel classification, where each pixel is classified as being part of the nucleus according to some intensity threshold (either global or adapted to the local context [30]). This simplistic approach is fast and efficient on highly contrasted datasets, but quickly becomes error-prone as the SNR decreases. Alternative solutions have been proposed to combine intensity with size thresholds [18], or use multi-scale analysis [28], perhaps followed by region-growing approaches including watershed [13] and active contours [34,35]. Nevertheless, a common bottleneck arises in highly dense tissue, where nuclei that do not appear separated are either discarded or merged, leading to under-segmentation. Separating clustered objects is still an open problem with no universal solution, although several specialized methods have been proposed, notably based on watershed [27], gradient flow [36,37], super-pixel grouping [29] or the concept of ‘lines-of-sight’ [30].

Cell segmentation in a tissue context is a comparatively much more delicate endeavour, due to the extreme proximity and high density of the microenvironment. Methods in this category fall into one of two categories, depending on their starting point:

(1) **Region-growing approaches** start by detecting the centre of each cell and applying a region-growing approach to reach the cell membrane, which is particularly useful when a nuclear segmentation is already available [18,19]. However in the absence of a nuclear marker, the central region of the cell can be inferred from local-intensity minima in the membrane signal. Region growing is typically achieved using the watershed approach [20,23,28,32,38–41], where the image is considered as a topographic relief map that is iteratively flooded from every initial seed iteratively until the edges of the water basins meet (defining the so-called watersheds). While the watershed is particularly sensitive to noise, the final membranes are also not always optimally placed, notably on low-resolution datasets where considerable amounts of data are missing [32]. A popular alternative for cell segmentation lies in deformable models (also known as active contours) [18,24,31,35], where a contour is initialized around each seed (nucleus) and is attracted towards the cell membrane by forces resulting from the minimization of a cost functional with various parameters controlling the behaviour of the contour. Deformable models are more flexible in comparison to other approaches, at the expense of slightly increased computation times, although efficient implementations are available [35,42].

(2) **Direct approaches** rely primarily on the membrane signal, and rather consider the cell membranes as a network that is to be extracted from the image [33,43]. These approaches typically start from an intensity-based analysis, followed either by a polygonal fitting procedure [33] or morphological operations followed by surface reconstruction and local refinement [43], not unlike active contours.

The number and diversity of segmentation methods clearly highlight the current lack of a universal, fully automated solution. While each method is usually better suited for a given application, it is worth pointing out the growing interest for machine learning strategies, where the user is allowed to intervene at different steps of the workflow to correct and teach the algorithm when it errs [24]. The growing influence of machine learning in numerous areas of science (and notably in image recognition [44]) is opening a new avenue for future developments in this field, and provides a potential alternative to build a truly generic segmentation tool, as illustrated by the open-source Ilastik software (www.ilastik.org).

4. Tracking cells in context

As we shall discuss in this section, cell tracking is an essential and necessary step towards the understanding of development over time, through different cellular events, such as proliferation, differentiation and migration.

Cell tracking is the process of following the position of each and every cell of interest within its environment over the course of time (typically, a time-lapse acquisition), with the ultimate goal to extract spatio-temporal features such as deformation, migration, intercalation and lineage tree [45]. Similarly to cell segmentation, manual cell tracking is a particularly tedious undertaking, especially on massive three-dimensional sequences comprising thousands of cells and their progeny. Cell tracking has therefore been and continues to be a topic of substantial interest in the bioimage informatics community [46], as illustrated by the organization of a recent community challenge [47]. Traditionally, tracking methods are classified into two major categories: (i) **association** methods, where all the cells are first segmented in the entire sequence, then the tracks are built from these detections in a subsequent step, and (ii) **evolution** methods, where each cell is sequentially segmented and propagated from frame to frame (typically using template-matching approaches or active contours). In developmental biology, however, the landscape is slightly different. Indeed the computational burden induced by deformable models for segmentation tasks renders their application to tracking even more challenging. Instead, the vast majority of tracking methods rely on association, although in three major ways:

(1) Frame-to-frame association consists of assigning to each cell in the first frame a cell in the subsequent frame, based on some prior knowledge or hypothesis of
motion. The simplest prior is to assign the spatially closest cell in the next frame, but other criteria have been used to improve or refine this association, based either on the global motion of the observed tissue [23] or other information such as shape, intensity and prior knowledge on cell behaviour [31].

(2) Graph-based approaches represent the tracking problem in the form of an oriented graph where vertices represent detected cells, while edges represent all the possible links between the vertices over time [33,39,40]. The association problem is then solved by a graph-optimization procedure to extract the most plausible temporal pathway of cells through the graph.

(3) Global optimization strategies have been investigated to segment and track cells simultaneously within the entire spatio-temporal four-dimensional (3D+time) hypervolume in a single step, based on spatio-temporal morphological operators [21] (by analogy with kymograph analysis, this would be equivalent to tracking cells in two dimensions by extracting three-dimensional tubes within a 2D+time hyperstack). Such approaches are efficient on small datasets, but do not scale properly on conventional workstations.

In the vast majority of cases, the efficacy of association-based cell tracking methods is highly dependent on the previous (segmentation) step. A cell that is not properly detected can quickly add to the computational burden of the tracking algorithm, which must rely on more rules and strategies to determine whether the absence of a cell at a given location is due to erroneous segmentation, or rather due to a natural biological process (e.g. cell death, local change in tissue dynamics, etc.). Eventually, excessively strict rules that produce a perfect tracking in a specific situation may inadvertently limit the interpretation of tissue dynamics to only a subset of the (biologically) plausible scenarios. It is therefore crucial to keep a reasonable parameter tradeoff to prevent data over-fitting, which is best solved by returning to and improving the segmentation step.

5. Towards reproducible research via open software resources

Ingrained in the principles of reproducible research is the verification and reusability of experimental results, facilitating the conception of future experiments and preventing wasting time in reinventing the wheel and/or building upon false conclusions. While the importance of the availability of raw data is now openly recognized, such current practices in the bioimaging field are still far behind other scientific domains [48], while discussions between journal editors and publishers on developing best-practice guidelines to address this issue are still in their infancy [49]. In recent years however, the bioimage informatics community has been actively pushing to promote reproducibility in both the software and hardware aspects, by releasing analysis algorithms and equipment blueprints online [4,50,51], as well as ground-truth data in the form of publicly available datasets and software for benchmarking purposes [47].

Several such developments have directly targeted the developmental biology community, which we summarize in table 1. Each solution has its own strengths and weaknesses, which generally stem from the biological application driving its development. For instance, general-purpose software platforms typically provide a wide range of analysis tools to analyse bioimaging data (including tissue), although they may exhibit lower efficacy in specific contexts where specialized (though less generic) solutions have been developed. While this review does not provide an exhaustive list of such solutions, we focus here on some of the most recent or active developments in the field of epithelial and embryonic development, most of which are available as open-source platforms:

- Icy [52] and ImageJ/Fiji [53] are typical illustrations of such general-purpose, community-driven bioimaging platforms providing a large variety of tools and frameworks for data acquisition, visualization and analysis in biology, with a particular focus on extensibility via community-contributed plugins, scripts and protocols. They have been successfully applied to developmental biology studies through various extensions [14,18,20,54]. For example, one such extension for the Icy platform studies the spatial organization of the embryonic mouse heart, by combining a filtering approach to increase membrane signal acquired in confocal microscopy, and a segmentation tool based on deformable models to extract both the nuclei and the cell membranes [18,55].

- MARS-ALT [38] is among the first reported freely available software dedicated for the segmentation and lineage tracing of cells in three-dimensional time-lapse imaging data for developmental biology studies, and more specifically in the context of plant growth in the Arabidopsis thaliana model plant. The pipeline starts with an algorithm to fuse multi-angle confocal three-dimensional imaging stacks. The cells are then segmented using the watershed algorithm, and tracked over time using nonlinear registration of the consecutive time points, followed by a graph-based matching algorithm. It is worth pointing out that this software, although freely available, is not open-source, and therefore does not fully comply with reproducible research principles.

- Automated cell morphology extractor (ACME) is a software developed for the automated reconstruction of cell membranes in three dimensions [23]. The software provides an all-in-one solution for dense tissue that sequentially applies data-filtering for membrane structure enhancement, followed by the segmentation of cells in dense tissues. It was specifically developed for zebrafish data (illustrated on neuroectoderm and paraxial mesoderm data) obtained via confocal and two-photon microscopy.

- Focusing more particularly on the dynamics of epithelial tissue during development, the EDGE4D software [43] has been proposed to segment and track cells across the complex process of epithelial folding during Drosophila gastrulation. The algorithm combines nuclear and membrane signals to provide superior segmentation to its aforementioned counterparts in the challenging context of multi-photon microscopy.

- MorphoGraphX [41] is a platform developed for the quantification of four-dimensional (three-dimensional time-lapse) analysis of growth dynamics in plant morphogenesis, focused on computational efficiency and realistic visualization through the use of graphics processing units (GPU). This software provides specific routines to segment and
Table 1. Overview of selected image analysis software platforms used in developmental biology and morphogenetic studies. Release date indicates the date on which the described pipeline was made available (this does not necessarily coincide with the initial release of the enclosing software). Acronyms (in alphabetical order): AC, active contours; CS, cell segmentation; GBM, graph-based matching; IT, intensity threshold; ME, membrane enhancement; MSP, multi-scale products; MVF, multi-view fusion; NLR, nonlinear registration; NNA, nearest-neighbour association; NS, nuclear segmentation; W, watershed.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>language</td>
<td>Python</td>
<td>Java</td>
<td>C++</td>
<td>C++ , R</td>
<td>C++ , ITK</td>
<td>R</td>
<td>C++</td>
<td>Matlab, Java (Icy)</td>
</tr>
<tr>
<td>structures</td>
<td>membranes</td>
<td>nuclei</td>
<td>membranes</td>
<td>membranes</td>
<td>membranes</td>
<td>membranes</td>
<td>nuclei</td>
<td>membranes</td>
</tr>
<tr>
<td>segmentation</td>
<td>MVF</td>
<td>ME</td>
<td>ME</td>
<td>ME</td>
<td>CS (W)</td>
<td>CS (W)</td>
<td>NS (MSP + IT)</td>
<td>CS (W)</td>
</tr>
<tr>
<td>tracking</td>
<td>NLR + GBM</td>
<td>n.a.</td>
<td>n.a.</td>
<td>NNA</td>
<td>manual</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>licence</td>
<td>Cecill-C</td>
<td>BSD</td>
<td>n.a.</td>
<td>n.a.</td>
<td>MPI-CBG</td>
<td>Apache v2</td>
<td>CC BY</td>
<td></td>
</tr>
<tr>
<td>support</td>
<td>INRIA (France)</td>
<td>NIH (USA)</td>
<td>n.a.</td>
<td>MPI-PBR (Germany)</td>
<td>HHMI (USA)</td>
<td>MRC (UK)</td>
<td>CR UK (UK)</td>
<td></td>
</tr>
</tbody>
</table>
track the outer cell layer of the embryo (represented as an unstructured graph), by leveraging the power of existing open-source packages including Cimg (www.cimg.eu) and ITK (www.itk.org). It has mainly been applied in the context of Arabidopsis [56,57], but has been shown to work with other models, notably Drosophila [41].

TissueMiner is a framework to quantify epithelial cell dynamics in living tissues over prolonged acquisitions [19]. It specifically includes tools to generate a database from cell segmentation and tracking results, and provides a number of quantitative spatio-temporal descriptors of epithelial morphogenesis, and was applied to study the Drosophila pupal wing. The interface is based on the R platform and provides utilities to query and visualize the database in an efficient manner.

Real-time accurate cell shape extractor (RACE) [28] is a high-performance image analysis framework for automated three-dimensional cell segmentation, with a particular focus on massive SPIM datasets. It provides the necessary software to extract nuclei and cell shapes from terabyte-sized images, and therefore constitutes a step towards scalability and high-throughput studies, as illustrated by examples on Drosophila, zebrafish and mouse embryos. Additionally, a companion tracking software was developed to study large-scale embryonic development [29].

EpiTools is another open-source toolkit for the study of developing epithelial tissue [20]. Based on Matlab, it provides graphical interfaces to segment and track the contour of cells in 3D+time confocal series of membrane-labelled cells. It also closely integrates with the Icy platform [52] by providing specific extensions to open and visualize the segmentation and tracking results within Icy. The tool is specifically developed to study the dynamics of epithelial growth and morphogenesis, and has been applied to the Drosophila wing disc.

6. Discussion

Joint developments in both the bioimaging and bioimage informatics fields are progressively revolutionizing how developmental biologists apprehend fundamental questions related to morphogenesis and tissue dynamics. High-quality imaging data can now be acquired from a wide range of experimental models and conditions, with remarkable precision, longevity and reproducibility, slowly leading us into the era of Big (Imaging) Data. Concomitantly, computerized algorithms and dedicated software able to handle, visualize and analyse such colossal amounts of data have rapidly flourished, bringing an ever more quantitative insight into the numerous processes underlying morphogenesis. These developments (and the new to come) are dramatically accelerating the pace at which new biological hypotheses can be tested and validated, and are therefore becoming an essential part of the standard toolkit in developmental biology.

It is frequently argued that the exponential increase in imaging data produced routinely (notably via multi-view SPIM devices, but not only) is raising the issue (both technical and financial) of data storage, management and curating. Meanwhile, the development of open-hardware imaging solutions is progressively reducing the overall cost of imaging systems and thereby of data production. Without giving up on reproducible research, it is becoming reasonable to question whether long-term storage of terabyte-sized datasets is effectively worth the effort, since new imaging data can be produced rapidly and reliably in reasonable time (while benefiting from latest advances in, for example, imaging techniques and fluorescent probes). Potentially, the progressive integration of imaging hardware and software through engineering will allow us to speculate on the future emergence of task-focused, bench-top equipment able to perform imaging and quantitative analysis in a single and transparent step, thereby circumventing the issue altogether.

As the remainder of this special issue illustrates, the road to a comprehensive understanding of the mechanisms underlying morphogenesis is still long and challenging, with quantitative imaging and analysis representing only a minimal (though necessary) first step. Considerable effort is currently invested into developing in silico, computational models of embryonic or plant development based on mechanical and physical rules (e.g., [58]). Now comes the equally crucial challenge of inferring mechanistic descriptions from image-based information, calling for proper biophysical and biomechanical modelling of the various observable facets of development. The bioimage informatics community has recently started to advance in this direction, with the appearance of generative models of embryo and organ formation based on image data [59], paving the way towards a new interdisciplinary research endeavor at the interface between optics, signal processing, mathematics and physics.

Competing interests. We have no competing interests.

Funding. A.C.D and J.-C.O.-M. are funded by the Institut Pasteur. A.H.J. is funded by the INSERM.

References


Downloaded from http://rstb.royalsocietypublishing.org/ on May 7, 2017
31. Barnes N. 2010 Publish your computer code: it is good enough. Nature 467, 753. (doi:10.1038/467753a)
33. de Chaumont F et al. 2012 Icy: an open bioimage informatics platform for extended reproducible...


