Tools to reverse-engineer multicellular systems: case studies using the fruit fly

Qinfeng Wu, Nilay Kumar, Vijay Velagala and Jeremiah J. Zartman

Abstract
Reverse-engineering how complex multicellular systems develop and function is a grand challenge for systems bioengineers. This challenge has motivated the creation of a suite of bioengineering tools to develop increasingly quantitative descriptions of multicellular systems. Here, we survey a selection of these tools including microfluidic devices, imaging and computer vision techniques. We provide a selected overview of the emerging cross-talk between engineering methods and quantitative investigations within developmental biology. In particular, the review highlights selected recent examples from the Drosophila system, an excellent platform for understanding the interplay between genetics and biophysics. In sum, the integrative approaches that combine multiple advances in these fields are increasingly necessary to enable a deeper understanding of how to analyze both natural and synthetic multicellular systems.

Keywords: Microfluidics, Image processing, Machine learning, Deep learning, Imaging

Background
Answers to many human health challenges require an integrated systems-level understanding of the body [1]. Biocomplexity, the emergence of properties that are more than the sum of individual constituents, leads to profound implications on how to solve problems in regenerative medicine, cancer therapy, and personalized medicine [2]. This complexity spans multiple spatial scales from molecules, such as proteins and DNA, to cells, tissues, organs and organ systems. It requires a systems-level analysis to understand this complexity [3]. The general paradigm of systems research adopts an iterative approach, which usually involves transitioning from experiments to model formulation then to revision of original hypotheses (Fig. 1a) [4].

Genetic model systems, such as the worm—C. elegans, the zebrafish or the fruit fly—Drosophila melanogaster, serve as proof-of-principle platforms for developing tools to analyze multicellular systems or to test new techniques in forward-engineering living systems [5]. In particular, Drosophila enables genetic studies of how genes are regulated to control morphogenesis [6–8] and physiology [9]. It is an excellent system for studies that are at the crossroad of biophysics, information processing, and molecular and developmental biology. The fruit fly system provides many advantages, including cheap and easy husbandry, rapid life cycle, and many available genetic tools [5, 10–16]. These advantages contribute to the status of Drosophila as a premier model for reverse-engineering multicellular systems. Of note, several fundamental signaling pathways were first discovered in Drosophila, including Hedgehog [17], Notch [18] and Wingless pathways [19]. Therefore, Drosophila has been extremely crucial in biology and bioengineering researches in many areas and will surely continue to play a critical role in years to come [20].

Beyond fundamental research, Drosophila has been used to study many health challenges, including cancer [21–28], neurodegenerative disorders [29–31], infectious diseases [32], cardiac disease [33], aging and metabolic diseases [34], wound healing and organ regeneration [20, 35–38] (Fig. 1b). Drosophila disease models can accelerate the rate of therapeutic drug testing and discovery due to the availability of genetic tools and a genome that lacks redundancy [11, 39–41]. Thus, Drosophila has a proven track record for understanding the biocomplexity of multicellular systems.

Here, we review a selected set of engineering tools and methodologies that are broadly applicable to reverse-engineer organ development. As a case in point, we focus on selected examples centered on the
quantitative analysis of *Drosophila* (Fig. 1). This review highlights selected engineering advances that have led to the development of tools in the field of high-throughput and high-content screening: microfluidic devices, imaging technologies, and imaging analysis algorithms. Many novel and elegant engineering designs, such as various microfluidic devices and imaging modalities, have more precise manipulations and extract deeper insights from genetic systems, with a large breadth applied to the zebrafish, the fruit fly and the worm [42–45].

Rapid advances in machine learning and deep learning have greatly increased researchers’ ability to extract and analyze biological data. These tools are enabling increasingly quantitative characterization of fruit flies and other multicellular systems. Finally, the availability of many computational modeling tools (see, for example, reviews such as [46, 47]) has facilitated and accelerated the iterative cycle of hypothesis testing and revision (Fig. 1a). The review concludes with a perspective on current trends and future potential directions for reverse-engineering of multicellular systems.

**Microfluidic devices enable controlled imaging and perturbations of fruit fly development**

Microfluidic devices refer to systems that use channels with dimensions of tens to hundreds of micrometers to manipulate a small amount of fluids [48]. A big challenge in studying the fruit fly is how to accurately apply perturbations and manipulate its organs due to their small size. Microfluidic devices are an increasingly important technique for addressing this challenge. In the following section, we discuss how microfluidic devices were applied in representative individual studies and how they have contributed to the improvement of current experimental approaches.

**Sample preparation and immobilization**

Immobilization is a critical step to achieve high resolution imaging and precise manipulation for moving samples, such as *Drosophila* larvae. For example, to study the larval nervous system, researchers require the larva to be immobilized to image neuronal physiological activities. However, immobilization of larvae is difficult because of its digging and burrowing motion. Traditional immobilization techniques, such as tape or glue, still allow minor larval movement and reduce larval viability [49, 50]. Therefore, several strategies have been developed to immobilize samples. For example, Mondal et al. used a deformable membrane controlled by a water column to mechanically restrain larvae. The device allows them to image vesicle trafficking in the neurons of
**Drosophila, C. elegans, and zebrafish at high resolution** [51, 52]. Another chip designed by the same group immobilizes larvae by clamping the mouth region to reduce digging movement. There is an additional design that pneumatically immobilizes larvae and allows for automated larva loading, immobilization and unloading. Both methods achieved significant immobilization and resulted in high-resolution imaging of neural responses [53, 54]. Mechanical restraint achieves easy immobilization but leads to reduced viability and innate response to mechanical perturbation [53, 54].

Anesthesia is an alternative to mechanical immobilization. Heemskerk et al. developed an immobilization chamber that uses desflurane for anesthesia [55]. A newer design uses both CO\(_2\) and compression to immobilize larvae [56]. The chip also incorporates inputs for food feeding that allow for long-term (> 10 h) immobilization and imaging. Researchers were able to observe regenerative axonal growth up to 11 h of injury of the larva, demonstrating that CO\(_2\) did not affect the physiology of the larva in this study. An improved design uses coolant, instead of CO\(_2\), for anesthesia and immobilization (Fig. 2a). This technique enabled the imaging of in vivo mitochondria movement in axons with high resolution without affecting the larva physiology [57].

Orienting a multicellular sample during loading is a frequently encountered problem. To overcome this, Ardeshiri et al. employed a rotatable glass that can suck onto the head of the larva to rotate the larva [49, 58]. Another creative solution allows samples to be prepared on the cover glass first before the silicone slab is placed on top to form the channels of the device [59]. This design allows more flexible preparations, better orientations and wider accommodation of a variety of samples.
Microinjection
Delivery of genetic constructs into fly embryos requires precise microinjection. For perturbation studies, drugs/toxins must also be accurately introduced into fragile embryos. Due to the requirement of precise placement and the small volume of injection, microinjectors have become tools of choice. Several microfluidic devices have been created to miniaturize this technique and to surpass the reliability of manual injection. First, Delubac et al. designed a microfluidic system for automatic embryo loading, detection and injection [60]. The device retrieves and places the embryos in contact with the injector/needle. The injection begins when the system detects the embryo in front of the injector. This fully-automated process enables high-throughput screening of embryos and/or creation of transgenic Drosophila lines. However, there is no control as to how deep the injector can go. Later, Ghaemi et al. incorporated a long-taper needle and a micro-positioner to control the depth of injection (Fig. 2c) [61]. This system enables deep (up to 250 μm), highly-precise injections (a resolution of 5 μm) and low injection volumes (as low as 30 ± 10 pL) with minimum damage because of the tapered needle. The precise (position and volume) injection of toxins (NaN₃) into specific locations of the Drosophila embryo enables a detailed spatiotemporal study of how toxins affect embryo development [61].

Sorting, positioning and orienting of samples
One of the advantages of using Drosophila embryos is the high-throughput data collection enabled by the number of embryos that can be obtained at low cost. However, sorting, positioning and orienting of many embryos or other post-embryonic organs is a technical hurdle that needs to be addressed. Furlong et al. adopted the concept of fluorescence-activated cell sorting (FACS) and designed a device for sorting embryos expressing a fluorescent protein marker [62]. The device uses a robotic valve to separate the embryos into fluorescent and non-fluorescent samples. In 2004, Chen et al. presented a pressure-controlled microfluidic sorter for Drosophila embryos that directs the flow direction of embryos into different outlets [63]. The computer simulation and flow experiment with dye demonstrated the functionality of the device. Chen et al. improved the design to allow for high-speed sorting, enabled by a deflecting jet to change the movement of the object [64].

Bernstein et al. presented an early attempt to position and orient Drosophila embryos in batch for high-throughput microinjection. They designed a micro-assembly of protruded hydrophobic surfaces to achieve large-scale positioning and orienting of the embryos [65]. Embryos are flowed through the device and are immobilized when in contact with the hydrophobic surface. The designed achieved 95% immobilization rate and 40% alignment rate. They also presented a conceptual design of the high-throughput microinjection system that would work with the orientation array, still yet to be realized as a physical working model [66].

Lu and collaborators developed a series of array-based microfluidic devices for positioning and orienting Drosophila embryos. A first microfluidic array was designed to utilize passive hydrodynamics to trap, position and vertically orient Drosophila embryos (Fig. 2d) [67, 68]. The vertical orientation of the embryo allows the observation of dorsal-ventral patterning of proteins of interest. The device provided high-throughput dorso-ventral patterning data. Subsequently, the researchers modified the device to horizontally orient the embryo [69]. The Lu lab further improved the design to increase the loading efficiency to > 90% [70]. The new iteration also allows for anoxia perturbation of the embryos and potentially other forms of perturbation.

Multi-modal perturbations to organ systems
Spatiotemporal control over a range of perturbations (e.g. mechanical, chemical and electrical) on multicellular samples often requires multi-modal microfluidic device designs. Lucchetta et al. designed pioneering microfluidic devices to investigate how temperature regulates embryogenesis [71, 72]. The device generates a temperature step between the two compartments of a Drosophila embryo. This spatiotemporal perturbation of temperature created a way to understand the complex biochemical networks governing Drosophila embryogenesis [73]. Researchers have adopted this design and used it for other perturbations. For example, a similar design exerts spatiotemporal control of oxygen gradient on living embryos [74]. To accommodate various Drosophila samples and apply different kinds of chemical stimuli, Giesen et al. came up with a device that can immobilize a range of Drosophila organs and apply chemical stimulations [75]. The authors demonstrated the use of the device to perturb and image brain, leg and proboscis. They successfully measured calcium-based neuron responses to chemical stimuli at single-cell resolution using this device.

Zhang et al. devised a microfluidic system that applies millinewton-level mechanical stimuli to Drosophila larvae [76]. The system uses a pipette controlled by a robotic system to apply the mechanical stimulation. The robotic system significantly increases the accuracy and consistency of mechanical stimulation over manual operation. Another device that allows for precise mechanical perturbation of organs uses a diaphragm deflectable by pneumatic pressure to apply uniaxial compression on Drosophila wing disc (Fig. 2b) [77]. Using this device, Narciso et al. probed the genetic and mechanical
mechanisms of Ca²⁺ signaling in wing discs, a model organ for investigating signal transduction during organ growth. The device allows accurate mechanical stimulation of the wing disc, and it can be modified to accommodate other organoid-size systems and/or adding additional perturbations, such as electric stimulation [78].

Trends for microfluidic devices for multicellular systems
Microfluidic devices enable high-throughput analysis and perturbation with high spatiotemporal resolution. Recent efforts have combined functionalities that were traditionally achieved by multiple microfluidic devices into one design. For example, Shorr et al. invented a device that incorporates various automated operations of Drosophila embryo, including high-throughput automatic alignment, immobilization, compression, real-time imaging, and recovery of hundreds of live embryos [79]. These new devices have achieved multiplexing of various modalities, and allow for acceleration of research in developmental biology and multicellular systems [80].

The possibilities brought up by microfluidic devices are numerous and the development of new manufacturing technologies is helping the democratization of microfluidic devices as well. Computer-aided design (CAD) and simulation have greatly increased the accuracy and functionality of newly-designed devices [63, 64, 79]. 3D printing is enabling the customizable production of microfluidic chips [81, 82], as the resolution of those printers has improved significantly. 3D printers have brought down the cost of manufacturing and enabled the easy transfer of designs [80]. Other quick-fabrication techniques, such as hybrid-polyethylene-terephthalate laminate (PETL), are also lowering the barrier to entry for microfluidic devices [78, 83]. In addition, many universities are also providing training programs and have clean-room facilities that can support the adoption of microfluidic devices among new users [80]. Combined, these developments are encouraging the development of microfluidic devices with new applications in developmental biology and the synthetic biology of multicellular systems.

Three-dimensional imaging modalities enable the analysis of thick multicellular systems
Due to the larger scales involved, multicellular systems, including Drosophila tissues, require three-dimensional imaging techniques. An increasingly diverse range of imaging modalities is enabling researchers to investigate deeper into tissues. Recent improvements of fluorescence-based imaging modalities have increased imaging resolution, sample penetration and acquisition rate while reducing phototoxicity and photobleaching [84, 85]. In the meantime, other new imaging modalities, such as harmonic generation microscopy and micro-computed tomography (micro-CT), enable label-free imaging [86, 87] (Fig. 3a, b). In this section, we discuss variations of fluorescent imaging techniques and label-free imaging. We also cover the advantages and limitations of each imaging modality.

Confocal microscopy
Confocal microscopy uses a pinhole aperture to reject out-of-focus light to improve resolution and signal-to-noise ratio, compared to wide-field microscopy (Fig. 3c) [88]. Confocal microscopes can achieve a penetration depth of up to around 100 μm [89]. Confocal microscopy is divided into two main subcategories: laser scanning confocal microscopy and spinning disk confocal microscopy [89]. In laser scanning confocal microscopy, a single illumination spot is rastered across the field of view. The image acquisition rate is relatively low because of the point-by-point scanning system, especially when acquiring 3D stacks with multiple fluorescent channels from a sample. Because of the small focal point, laser scanning confocal microscopy can cause significant photobleaching and the specimen’s long-term viability is compromised due to phototoxicity [89]. Continuous efforts have resulted in significant increase of scanning speeds to lessen this limitation [90]. Alternatively, a spinning disk that contains many focus pinholes provides a multipoint scanning strategy that significantly increases the collection rate. This reduces photobleaching and improves specimen viability. However, this comes at a cost of reduced 3D-sectioning capability and resolution.

Light-sheet fluorescent microscopy
In light-sheet microscopy, only a single plane of focus is illuminated (Fig. 3b). The camera detects fluorescence from a direction perpendicular to the light-sheet. The scanning speed of a light-sheet fluorescent microscopy is 100–1000 times faster than that of laser scanning confocal microscope. These characteristics minimize both phototoxicity and photobleaching and enable long-term imaging experiments of 3D multicellular systems [84]. This advantage allows imaging of a beating heart of a zebrafish or imaging of whole Drosophila embryos with fast rates of acquisition [91]. For example, Drosophila embryos can complete normal development even after being irradiated for 11,480 images by a light-sheet microscope [92]. The limited illumination of the specimen also results in high signal-to-noise ratio.

Light-sheet microscopes are highly customizable and can be coupled with other imaging techniques and/or downstream computational processing. For example, Greiss et al. achieved single-molecule imaging in a living Drosophila embryo, which is highly opaque in later stages, with reflected light-sheet microscopy [93]. Tomer et al. built a simultaneous multiview light-sheet microscopy that
can acquire 175 million voxels per second (Fig. 3d) [94, 95]. Chhetri et al. developed isotropic multiview light-sheet microscopy for long-term imaging with double the penetration depth and 500-fold larger temporal resolution than previous design of light-sheet microscopes [96]. Aided by image segmentation and computational tracking, researchers reconstructed the geometry of the entire tissue and measured morphogenic dynamics during embryo development [97]. Lattice light-sheet microscopy, which results in an ultrathin light sheet, further increases the speed of image acquisition (scanning 200 to 1000 planes per second) with reduced phototoxicity [98].

Light-sheet microscopes can be constructed at relatively low cost, compared with other imaging technology setups. A great resource for building a customizable light-sheet microscope is an open hardware and software platform called OpenSPIM [99]. However, a significant challenge for light-sheet microscopes is how to process,
Multi-photon fluorescence microscopy

Multi-photon fluorescence microscopy relies on the simultaneous absorption of multiple photons to excite fluorophores (Fig. 3a). This process requires a high-energy laser concentrated at the laser focal point. Outside the focal point, the laser power is below the threshold required for two-photon excitation. This allows multi-photon microscopes to excite samples at a tiny volume around the focus point, thus reducing phototoxicity and extending the duration of in vivo imaging. The precise excitation at the focal point also improves the signal-to-noise ratio.

Multi-photon microscopes use near-infrared lasers with longer wavelengths (lower energy per photon) than lasers used in one-photon confocal microscopy. The near-infrared laser allows deeper penetration (2–3 times deeper for two-photon) into the sample, compared to confocal microscopy (Fig. 3d) [85]. The laser, because of the longer wavelength, also scatters less. Therefore, multi-photon microscopy provides good 3D sectioning capability for thick specimens. Researchers were able to image calcium dynamics in *Drosophila* adult brain in vivo in behavioral studies and odor-activated neuron response due to the deep penetration capability of two-photon microscopy, which is the most commonly used multi-photon microscopy [100–102]. Besides two-photon, three-photon microscopy has received increasing popularity because of its increased penetration and signal-to-noise ratio. For example, scientists have successfully imaged through adult mouse skulls at >500 μm depth using three-photon microscopy [103].

However, multi-photon microscopy has low acquisition rates due to the point scanning system and leads to accelerated photobleaching [104, 105]. Two-photon microscopy also causes autofluorescence of some chromophores, such as NAD(P)H, which can cause significant noise for image acquisition [106]. The cost is also significantly higher because of the more sophisticated laser, optics, mechanics, and maintenance required. Nevertheless, the improvement of functionality and the continuous reduction of costs will enable multi-photon laser scanning microscopy to be adopted by the wider research community. Multi-photon microscopy currently defines the upper limit of penetration depth in diffraction-limited microscopy [85].

Harmonic generation microscopy

The fluorescence microscopies discussed above have several innate shortcomings, such as photobleaching, phototoxicity, and the need to label the molecules [107]. Harmonic generation microscopy, on the other hand, achieves label-free imaging. Harmonic generation refers to the nonlinear optics phenomenon where multiple photons reach a molecule and generate a new photon without the presence of a fluorophore. For example, during second-harmonic generation, two identical incoming photons are combined to generate one outgoing photon with a wavelength of exactly half of the excitation beam (Fig. 3a).

The biggest advantage of harmonic generation microscopy is that it does not require labeling of the molecules of interest. Harmonic generation microscopy also substantially reduces photobleaching and phototoxicity because it does not rely on the excitation of fluorophores [108]. In addition, harmonic generation microscopy achieves deep penetration by using near-infrared wavelengths for the incident light. Harmonic generation microscopy has the ability to construct high-resolution three-dimensional images of several hundred microns of depth.

Harmonic generation provides additional structural information on molecular or supra-molecular order not easily detectable with fluorescence strategies. Second-harmonic generation is caused by materials that are noncentrosymmetric [109]. These materials include collagen fibril/fiber structure (type I and II fibrillar collagen), myofilaments, fibers, polarized microtubule assemblies, and muscle myosin (Fig. 3e) [87, 110–112]. Second-harmonic generation microscopy has been used to image developing muscle structures and the trachea system in 2nd-instar larva, and the lipid bodies in *Drosophila* cells [112, 113]. Researchers used second-harmonic generation microscopy to investigate the structure of *Drosophila* sarcomeres and visualize myocyte activity to study rhythmic muscle contraction [114, 115].

Third-harmonic generation occurs at structural interfaces with local transitions of the refractive index [116]. Third-harmonic generation was used to image lipid in *Drosophila* and mouse embryos. When coupled with second harmonic generation microscopy and two-photon imaging, one can explore the interactions between lipid, extracellular matrix and fluorescence-marked proteins (Fig. 3f) [113, 117–119]. Researchers used third-harmonic generation to visualize rhodopsin in the eye [120], and to measure the morphogenetic movement in *Drosophila* embryos by visualizing lipid droplets around cell nuclei and the interfaces of yolk structures [121]. Together, second- and third-harmonic generation microscopy modalities serve as powerful label-free imaging techniques.

Micro-CT

Micro-computed tomography (micro-CT), like traditional CT, uses X-rays to produce sectioning of a sample and uses computers to reconstruct the 3D...
morphology of the specimen [122]. Micro-CT produces images with microscopic resolution and avoids artifacts due to processing of samples used for fluorescence imaging [123]. Because insects are made of only soft tissues, they are ideal for micro-CT. With very simple contrast staining, micro-CT can produce quantitative, high-resolution, high-contrast volume images of Drosophila, bumblebee, etc. [86, 124]. Micro-CT has become increasingly popular and is used to study morphological changes in a broad range of Drosophila tissues (Fig. 3g), including the female reproductive tract [125], neuronal structures [126], urolithiasis studies of calcium oxalate deposition [127], and wings for computational aero-dynamic analysis [128].

The combination of multiple imaging modalities opens new possibilities to utilize the strengths while avoiding the limitations of individual techniques. For example, Truong et al. combined two-photon microscopy with light-sheet microscopy to implement two-photon-scanned light-sheet microscopy for Drosophila embryos [129]. This combination achieved twice the penetration of one-photon light-sheet microscopy and is more than ten times faster than two-photon laser scanning microscopy. Researchers also combined multi-photon microscopy with harmonic generation microscopy to construct a comprehensive picture of samples including both the fluorophore-labeled molecules and non-labeled structural molecules [130]. However, a major challenge for systems bioengineers is to process large datasets generated by these advanced imaging techniques. There is a critical need to automate the analysis of large datasets and to reduce high-dimensional data that includes information of molecular species and biophysical properties of cells through both space and time [131].

**Trends of imaging technologies for multicellular systems**

Besides the introduction of new imaging principles, existing imaging technologies are often combined for multiplexing of functionalities that further increases in performance [93–96, 98]. There is also a trend of democratization of imaging technologies, from the OpenSPIM project supporting the construction of customized light-sheet microscopes to mobile phone-based microscopy [99, 132–134]. The increase in acquisition speed and resolution encourages the advance of image analysis methods to handle the ever-increasing amount of data generated from analysis of multi-cellular systems with Drosophila providing a versatile system for proof-of-concept studies.

**Data-driven learning algorithms accelerate the quantitative analysis of multicellular systems**

The exponential increase in biological data acquisition rates challenges conventional analysis strategies [135]. Integration of advanced algorithms for bio-image analysis is thus highly desired. The result of a bio-image analysis pipeline can be as simple as quantification of fluctuations in cellular areas over time or as complex as a high-dimensional array of features of a Drosophila wing. In short, the goal of analysis is to convert images into arrays of numbers that are amenable to statistical evaluation. This helps create data-driven models or to validate predictions from phenomenological or mechanistic models. In this section, we discuss how both conventional machine-learning and deep-learning algorithms play critical roles in the analysis of multicellular systems, using selected examples focused on the fruit fly. In particular, we show how deep learning is rapidly emerging as a solution to accelerate the analysis of biological big data (Fig. 4a).

Machine-learning algorithms leverage training datasets to find features within the data to fulfill the task of either classification or prediction [136]. A feature is a measurable property or characteristic of a phenomenon within the image. Feature extraction can either be manual or embedded within the algorithm's architecture. Machine-learning algorithms are either supervised (requiring example input-output pairs to train the algorithm) or unsupervised (input data not annotated). Unsupervised learning algorithms, such as k-means clustering, perform poorly on noisy datasets and are frequently unsuited to bio-image analysis [137]. Therefore, supervised machine-learning algorithms are more commonly adopted for bio-image analysis (Fig. 5).

One of the major challenges in cellular tracking is obtaining high-quality segmentation masks of cells and separating regions of interest from noisy images at each time point. Non-machine-learning techniques, such as Otsu’s method [138] and P-tile method [139], are very sensitive to noise and do not produce good quality segmentation masks. An alternative approach is using region accumulation algorithms, such as watershed transformation [140] as implemented in EpiTools [141], where seed points are defined within the image and are iteratively grown to form the complete label [142]. However, these algorithms result in over-segmentation and require further manual processing.

In comparison, researchers have started using supervised machine learning based on pixel classifiers for image segmentation because of their versatility and robustness. Some of the most widely used algorithms in designing a pixel classifier are support vector machines [143], adaptive boosting (AdaBoost) [144] and random forest [145]. A number of open-source packages, such as CellProfiler [146], Ilastik [147], CellCognition [148], Phe-noRipper [149], Windchrm [150], Fiji [151] and EBImage [152], implement the above algorithms. However, the algorithms used in most of the existing packages require
selection of features by a user (Fig. 4b). Incorporating too many features slows down the implementation of the algorithm and makes them unsuitable for real-time quantification. Manual feature selection and extraction also increase the processing time for each image and hence make these algorithms unsuitable for big data processing.

To resolve these issues, researchers have started to use a class of machine learning algorithms called deep learning, which completely bypasses manual feature extraction.
Deep-learning techniques achieve higher accuracies than classical machine-learning methods. These algorithms rely on neural networks, where layers of neuron-like nodes mimic how human brains analyze information (Fig. 4c) [153]. Since deep learning is a relatively new concept in computer vision, its impact in the field of bio-image informatics is yet to be fully realized [154]. The architecture of neural networks automates the extraction of features, thus eliminating the need for feature selection (Fig. 5). Thus, deep-learning algorithms are suitable for processing large datasets as there is a significant reduction in computational time achieved by avoiding a separate task of feature extraction. Once trained, deep-learning algorithms can analyze data from new sources of bio-images.

Rapid development in processing capabilities and availability of packages, such as TensorFlow [155], Blocks and Fuel [156], Torch [157], Caffe [158] and MATLAB, are making deep-learning techniques widely accessible to the systems biology and bioengineering communities. Deep-learning algorithms generate more accurate segmentation masks in less time, compared to conventional supervised learning algorithms.

One of the most common deep-learning algorithms is convolutional neural network (CNN) [159]. In a CNN, every network layer acts as a detection filter for the presence of specific patterns in the data. The first layers in a CNN detect large patterns that can be recognized and interpreted relatively easily. Later layers detect increasingly smaller patterns that are more abstract. The last layer makes an ultra-specific classification by combining all the specific patterns detected by the previous layers. However, the usage of this class of algorithms is heavily restricted by the amount of training data available in biology. To overcome this problem, a modified...
full CNN called U-Net was created [160]. U-Net was used to segment cells in Drosophila first instar larva ventral nerve cord using only 30 training images, thus significantly reducing the size of training data required for conventional CNN. Duan et al. used CNN to identify and mark the heart region of Drosophila at different developmental stages [161]. The algorithm performs better than the conventional machine-learning algorithms (Fig. 4d).

Additional applications of deep learning for analyzing multicellular systems in Drosophila include image registration. For example, cultured samples often move during image acquisition. The movement, along with deformations within the tissue, makes spatial quantification of features a difficult task. Image registration for biological samples is a two-step process: a) segmentation to identify regions to be registered, and b) registration of the region of interest. Conventional machine-learning algorithms are not well-suited for this task as they often rely on manual identification of intensity-based features that vary over time. Liang et al. used deep learning to segment out the pouch from time-lapse movies of Drosophila wing discs that expresses GCaMP6, a genetically-encoded fluorescent sensor [162]. Segmentation and registering the wing disc is challenging due to the highly dynamic and stochastic Ca$^{2+}$ dynamics [162]. The full CNN architecture identifies high-level embedded patterns, which are sometimes impossible to identify and extract manually. Segmentation was followed by a modified traditional image registration approach for tracking the moving wing disc pouch. Similarly, a full CNN was also used with a novel non-rigid image registration algorithm to optimize and learn spatial transformations between pair of images to be registered (Fig. 4e) [163].

**Trends of data analysis techniques for multicellular systems**

In summary, data-driven learning algorithms, such as machine learning and deep learning, serve as powerful new techniques for image processing of multicellular systems such as Drosophila. These algorithms can be used to tackle complicated problems and reveal structure in data that is too big or too complex for the human brain to comprehend. One of the biggest challenges in using these algorithms is that they require extremely large datasets that are well-annotated to train the algorithm. To circumvent this challenge, researchers have been working on ways to train models more efficiently with less data. Advancements in transfer learning enable the deep learning to apply classification capabilities acquired from one data type to another data type, thus increasing its robustness [164]. However, there are several challenges that need to be overcome to fully unleash the power of deep learning in biological research. A significant challenge is to make these techniques accessible. Collaborations are required between computer vision researchers and biologists for developing general-use packages. Support and proper documentation standards are needed for maintaining new computational packages to enable researchers to benefit and more quickly adopt new algorithm methodologies.

**Concluding perspectives**

Systematic approaches that integrate advanced microfluidic devices, imaging acquisition, and machine learning are essential techniques for analyzing the development of multicellular systems. There is an emerging need and intensive focus toward accelerating the cycle of hypothesis generation and testing and interdisciplinary collaboration through the engineering of integrative experimental and computational pipelines (Fig. 1b). Significant progress is being made that combines device manufacturing, computer vision, statistical analysis with mechanical automation of time-consuming biological experiments by multidisciplinary teams [165, 166].

From the traditional fluorescence-based imaging to X-ray-based micro-CT, we are seeing a range of new imaging technologies being applied to multicellular systems, including genetic model systems such as Drosophila. Advances in traditional fluorescence-based imaging is also significantly increasing image-acquisition speed, penetration and signal-to-noise ratio [93, 95, 96, 102]. In the meantime, label-free imaging of structure and/or measurements of tissue mechanics is leading to broader applications [111, 167]. These imaging modalities further combine with other technologies to provide increasing imaging capabilities. An emerging bottleneck for automating multimodal imaging experiments is the need to develop capabilities for parallel imaging modules integrated with customizable multichannel microfluidic devices to image many biological samples at a time. This, in turn, will increase the need for data storage and management solutions for labs. The significant advances being made in acquisition speed and resolution also demands a paradigm shift of analysis methods to handle the gigabytes and terabytes of data that are generated per imaging session [94, 96]. These new trends are blurring the knowledge boundaries of different research disciplines and encouraging the collaboration of microfluidic device designers, imaging technicians and computer vision scientists.

With the large amount of image data generated from experiments, machine learning is becoming an integral part of bio-image analysis. Significant progress in terms of computational power and availability of open-source modeling languages like TensorFlow has made machine
learning accessible to cell and developmental biologists. Recently developed algorithms, based on the concept of transfer learning, has decreased the required sample sizes needed for training learning algorithms. For instance, U-Net required only 30 training images to analyze Drosophila larval neural cord, compared with hundreds of images needed for traditional CNN [160]. Algorithms that perform even faster than U-Net, such as context encoding networks, Mask R-CNN and Deeplabv3+, have also been proposed recently [168–170]. However, a domain expert is required to implement these techniques, because they require fine-tuning of parameters and hyperparameters within the network [171]. Currently, computer vision algorithms can handle a variety of tasks, including registration of dynamic imaging data, removal of obstructing elements in images, normalization of images, improvement of image quality, repair of data, and pattern discovery [172–174]. These algorithms will enable more robust and accurate quantification of images of multicellular systems.

Finally, computational models are an additional tool for reverse-engineering multicellular systems. They are often required to generate new insights for explaining emergent phenomena. They also systematize the process of hypothesis generation to close the iterative loop in reverse-engineering multicellular systems (Fig. 1a). For example, the interplay between mechanical forces, biochemistry and genetics governs how cells organize themselves into organs (as reviewed in [6]). These processes require computational models to integrate experimental data and reduce the complexity to identify underlying principles governing system behavior [175]. Historically, Drosophila provides an ideal playground for developing and testing computational models of many aspects of development including pattern formation [176–180], organ growth control [181] and morphogenesis [182].

Various methods have been used to model cell-based processes in Drosophila, with a significant focus on modelling cell mechanics during morphogenesis. These methods include cellular Potts models, vertex models, continuum models, viscoelastic models, subcellular element models and immersed boundary methods, to name a few. Interested readers are referred to several reviews that focus on computational model development and validation [46, 47, 183]. A key consideration in analyzing multicellular systems is the need to account for heterogeneity (reviewed in [184]) and multiple length-scales (reviewed in [185, 186]). Another challenge is to develop multiscale models of physiological activities under different timescales, from millisecond to hours ([187], reviewed in [185, 188–190]). Finally, the integration of inference tools that estimate the subcellular distribution of forces is enabling more direct comparisons between model predictions and quantified experimental image-based data (one such example includes [191]). A couple of recent reviews on inference tools include [192–194].

A future goal for the reverse engineering of multicellular system should be the integration of data acquisition and analysis as highlighted in this review with the development and validation of computational models to guide the analysis of multicellular systems into generalizable pipelines [46]. Because of the variability of the experimental data in biology, there is a need to integrate uncertainty into model development. A Bayesian probabilistic framework is one mathematical strategy that incorporates uncertainty quantification into the optimization processes [195]. A Bayesian probabilistic framework can be used as a tool for estimating the parameters required to run bioprocess simulations, using experimental data extracted from bio-image analysis. Using such frameworks for biological systems will help in the robust and accurate quantification of parameters involved in computational simulations. In conclusion, the integrative engineering analysis of multicellular systems, often with Drosophila and other genetic model systems paving the way, is now reaching an exponential phase of synergistic growth.

Abbreviations
AdaBoost: Adaptive boosting; CNN: Convolutional neural network; FACS: Fluorescence-activated cell sorting; Micro-CT: Micro-computed tomography; SEM: Subcellular element model

Acknowledgements
The authors thank Cody Narciso, Pavel Brodskiy, Megan Levis, Dharsan Soundararajan, Jamison Jangula, Ramezan Paravitorghabeh and an anonymous reviewer for feedback on earlier versions of the manuscript. The authors apologize for only being able to cite a subset of relevant examples and reviews due to space constraints.

Funding
The work in this manuscript was supported in part by NSF CBET-1553826 and NIH Grant R35GM124935 to JJZ.

Availability of data and materials
Not applicable.

Authors’ contributions
QW and JJZ devised the structure of the manuscript. QW, NK, VV, JJZ designed the figures, and wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

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

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
57. Chaudhury AR, Insolera R, Hwang R-D, Fridell Y-W, Collins C, Chronis N. On chip cryo-anesthesia of Drosophila larvae for high resolution in vivo imaging applications. Lab Chip. 2017;17:2303–22.
58. Aredshir R, Reza P. Lab-on-chips for manipulation of small-scale organisms to facilitate imaging of neurons and organs, 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016. p. 5749–52.
59. Fan A, Tofangchi A, Venecia MD, Saif T. A simple microfluidic platform for the partial treatment of insuspendable tissue samples with orientative control. Lab Chip. 2018;18:735–42.
60. Delubac D, Highley CB, Witzberger-Krajcovic M, Ayoob JC, Furbee EC, Minden JS, et al. Microfluidic system with integrated microinjector for automated Drosophila embryo injection. Lab on a Chip. 2012;12:4911–9.
61. Gaemari R, Arep P, Stosic A, Acker M, Raza Q, Jacobs JR, et al. A microfluidic microinjector for developmental and translational studies in Drosophila embryos. Lab Chip. 2017;17:8898–908.
62. Furlong EEM, Proft D, Scott MP. Automated sorting of live transgenic embryos. Nat Biotechnol. 2001;19:153–6.
63. Chen CC, Zappe S, Sahin O, Zhang XJ, Fish M, Scott M, et al. Design and operation of a microfluidic sorter for Drosophila embryos. Sensors Actuators B Chem. 2004;102:59–66.
64. Chen CC, Wang JS, Solgaard O. Micromachined bubble-jet cell sorter with multiple operation modes. Sensors Actuators B Chem. 2006;117:523–9.
65. Bernstein RW, Zhang X, Zappe S, Fish M, Scott M, Solgaard O. Characterization of fluidic microassembly for immobilization and positioning of Drosophila embryos in 2-D arrays. Sensors Actuators A Phys. 2004;114:191–6.
66. Bernstein RW, Scott M, Solgaard O. In: Ma Z, Jin G, Chen X, editors. BioMEMS – Micro and Nanotechnology: Fundamentals, Devices and Applications. 2016. p. 5749.
67. Lucchetta EM, Munson MS, Ismagilov RF. Characterization of the local embryonic patterning network perturbed in space and time using microfluidics. Nature. 2005;434:1134
68. Lucchetta EM, Levario TJ, Zhao C, Rouse T, Shvartsman SY, Lu H. An integrated platform for large-scale data collection and precise perturbation of live Drosophila embryos. Sci Rep. 2016;6:21366.
69. Lucchetta EM, Artner RW, Ismagilov RF. The Endo-siRNA pathway is essential for robust development of the Drosophila embryo. PLoS One. 2009;4:e7576.
70. Wang Z, Oppegaard SC, Eddington DT, Cheng J. Effect of localized hypoxia on Drosophila embryo development. PLoS One. 2017;12:e0185367.
71. van Giesen L, Neagu-Maier GL, Kwon JY, Sprecher SG. A microfluidics-based robotic micromanipulation system for Mechanotransduction studies of Drosophila wing discs. Biophys J. 2017;113:491–501.
72. Levario TJ, Zhao C, Rouse T, Shvartsman SY, Lu H. Microfluidic trap array for massively parallel imaging of Drosophila embryos. Nat Protoc. 2013;8:721–36.
73. Goyal Y, Levario TJ, Mattingly HH, Holmes S, Shvartsman SY, Lu H. Parallel imaging of Drosophila embryos for quantitative analysis of genetic perturbations of the Ras pathway. Dis Model Mech. 2017;10:923–9.
74. Levario TJ, Zhao C, Rouse T, Shvartsman SY, Lu H. An integrated platform for large-scale data collection and precise perturbation of live Drosophila embryos. Sci Rep. 2016;6:21366.
75. Zhang W, Oppegaard SC, Eddington DT, Cheng J. Effect of localized hypoxia on Drosophila embryo development. PLoS One. 2017;12:e0185367.
76. van Giesen L, Neagu-Maier GL, Kwon JY, Sprecher SG. A microfluidics-based platform for measuring neuronal activity in Drosophila chemosensory neurons. Nat Protoc. 2016;11:2389–400.
77. Zhang W, Sobolevski A, Li B, Rao Y, Liu X. An automated force-controlled robotic micromanipulation system for Mechatronautanation studies of Drosophila larvae. IEEE Trans Auto Sci Eng. 2016;13:789–97.
78. Narciso CE, Contento NM, Storey TJ, Hoelzel DJ, Zartman JJ. Release of applied mechanical loading stimulates intercellular calcium waves in Drosophila wing discs. Biophys J. 2017;113:491–501.
79. Levis MK, Kumar N, Apakian E, Moreno C, Hernandez U, Olives A, et al. Rapid fabrication of hybrid PETL-glass microfluidic devices for combined live imaging and multimodal perturbations of multicellular systems. Biomicrofluidics. 2019. In press.
80. Sherry AR, Mustafa Sonmez U, Minden JS, Le Duc P. High-throughput mechatronautanation in Drosophila embryos with mesofluidics. Lab on a Chip. 2019; Available from: https://pubs.rsc.org/en/content/articlelanding/2019/fe/c8ck01055b. [cited 18 Feb 2019].
116. Weigelin B, Bakker G-J, Friedl P. Third harmonic generation microscopy of NAD(P)H and Flavoprotein. Biophys J. 2002;82:2811–95.

117. Huang S, Heikal AA, Webb WW. Two-photon fluorescence spectroscopy and microscopy of NAD (P) H and Flavoprotein. Biophys J. 2002;82:2811–25.

118. Pena A-M, Fabre A, Tordjmann T, Combettes L, et al. Femtosecond pulse-induced modulation of morphogenetic movements in Drosophila embryos with second harmonic generation microscopy. Nat Methods. 2006;3:47–53.

119. Mehnert A, Jackway P. Image Segmentation by Using Thresholding Techniques. 2010;2:4.

120. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012;9:671–8.

121. Rajaram S, Pavie B, Wu LF, Altschuler SJ. PhenoRipper: software for rapidly profiling microscopy images. Nat Methods. 2012;9:635–42.

122. Held M, Schmitz MHA, Fischer B, Walter T, Neumann B, Olma MH, et al. CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging. Nat Methods. 2010;7:747–54.

123. Rajaram S, Pavie B, Wu LF, Altschuler SJ. PhenoRipper: software for rapidly profiling microscopy images. Nat Methods. 2012;9:635–42.

124. van Merriënboer B, Bahdanau D, Dumoulin V, Serdyuk D, Warde-Farley D, Abadi M, et al. Blocks and Fuel: Frameworks for deep learning. arXiv. 2019;13:33.

125. abadi. [cited 12 Dec 2018].

126. van Merriënboer B, Bahdanau D, Dumoulin V, Serdyuk D, Warde-Farley D, Abadi M, et al. Blocks and Fuel: Frameworks for deep learning. arXiv. 2019;13:33.
