

# Chapter 17: Bioimage Informatics for Systems Pharmacology

Fuhai Li, Zheng Yin, Guangxu Jin, Hong Zhao, Stephen T. C. Wong\*

NCI Center for Modeling Cancer Development, Department of Systems Medicine and Bioengineering, The Methodist Hospital Research Institute, Weil Medical College of Cornell University, Houston, Texas, United States of America

**Abstract:** Recent advances in automated high-resolution fluorescence microscopy and robotic handling have made the systematic and cost effective study of diverse morphological changes within a large population of cells possible under a variety of perturbations, e.g., drugs, compounds, metal catalysts, RNA interference (RNAi). Cell population-based studies deviate from conventional microscopy studies on a few cells, and could provide stronger statistical power for drawing experimental observations and conclusions. However, it is challenging to manually extract and quantify phenotypic changes from the large amounts of complex image data generated. Thus, bioimage informatics approaches are needed to rapidly and objectively quantify and analyze the image data. This paper provides an overview of the bioimage informatics challenges and approaches in image-based studies for drug and target discovery. The concepts and capabilities of image-based screening are first illustrated by a few practical examples investigating different kinds of phenotypic changes caused by drugs, compounds, or RNAi. The bioimage analysis approaches, including object detection, segmentation, and tracking, are then described. Subsequently, the quantitative features, phenotype identification, and multidimensional profile analysis for profiling the effects of drugs and targets are summarized. Moreover, a number of publicly available software packages for bioimage informatics are listed for further reference. It is expected that this review will help readers, including those without bioimage informatics expertise, understand the capabilities, approaches, and tools of bioimage informatics and apply them to advance their own studies.

This article is part of the “Translational Bioinformatics” collection for *PLOS Computational Biology*.

## 1. Introduction

The old adage that a picture is worth a thousand words certainly applies to the identification of phenotypic variations in biomedical studies. Bright field microscopy, by detecting light transmitted through thin and transparent specimens, has been widely used to investigate cell size, shape, and movement. The recent development of fluorescent proteins, e.g., green fluorescent protein and its derivatives [1], enabled the investigation of the phenotypic changes of subcellular protein structures, e.g., chromosomes and microtubules, revolutionizing optical imaging in biomedical studies. Fluorescent proteins are bound to specific proteins that are uniformly located in relevant cellular structures, e.g., chromosomes, and emit longer wavelength light, e.g., green light, after exposure to shorter wavelength light, e.g., blue light. Thus, the spatial morphology and temporal dynamic activities of subcellular protein structures can be imaged with a fluorescence microscope - an optical microscope that can specifically detect emitted fluorescence of a specific wavelength [2]. In current image-based studies, five-dimensional (5D) image data of thousands of cells (cell populations) can be acquired: spatial (3D), time lapse (1D), and multiple fluorescent probes (1D).

With advances to automated high-resolution microscopy, fluorescent labeling, and robotic handling, image-based studies have become popular in drug and target discovery. These image-based studies are often referred to as the High Content Analysis (HCA) [3], which focus-

es on extracting and analyzing quantitative phenotypic data automatically from large amounts of cell images with approaches in image analysis, computation vision and machine learning [3,4]. Applications of HCA for screening drugs and targets are referred to as High Content Screening (HCS), which focuses on identifying compounds or genes that cause desired phenotypic changes [5–7]. The image data contain rich information content for understanding biological processes and drug effects, indicate diverse and heterogeneous behaviors of individual cells, and provide stronger statistical power in drawing experimental observations and conclusions, compared to conventional microscopy studies on a few cells. However, extracting and mining the phenotypic changes from the large scale, complex image data is daunting. It is not feasible to manually analyze these data. Hence, bioimage informatics approaches were needed to automatically and objectively analyze large scale image data, extract and quantify the phenotypic changes to profile the effects of drugs and targets.

Bioimage informatics in image-based studies usually consists of multiple analysis modules [3,8,9], as shown in Figure 1. Each of the analysis tasks is challenging, and different approaches are often required for the analysis of different types of images. To facilitate image-based screening studies, a number of bioimage informatics software packages have been developed and are publicly available [9]. This chapter provides an overview of the bioimage informatics approaches in im-

**Citation:** Li F, Yin Z, Jin G, Zhao H, Wong STC (2013) Chapter 17: Bioimage Informatics for Systems Pharmacology. *PLoS Comput Biol* 9(4): e1003043. doi:10.1371/journal.pcbi.1003043

**Editors:** Fran Lewitter, Whitehead Institute, United States of America and Maricel Kann, University of Maryland, Baltimore County, United States of America

**Published:** April 25, 2013

**Copyright:** © 2013 Li et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The research is supported by NIH R01 LM008696, NIH R01 CA121225, NIH R01 LM009161, NIH R01 AG028928, NIH U54CA149169 and CPRIT RP110532. The funders had no role in the preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: stwong@tmhs.org

## What to Learn in This Chapter

- What automated approaches are necessary for analysis of phenotypic changes, especially for drug and target discovery?
- What quantitative features and machine learning approaches are commonly used for quantifying phenotypic changes?
- What resources are available for bioimage informatics studies?

age-based studies for drug and target discovery to help readers, including those without bioimage informatics expertise, understand the capabilities, approaches, and tools of bioimage informatics and apply them to advance their own studies. The remainder of this chapter is organized as follows. Section 2 introduces a number of practical screening applications for discovery of potential drugs and targets. Section 3 describes the challenges and approaches for quantitative image analysis, e.g., object detection, segmentation, and tracking. Section 4 introduces techniques for quantification of segmented objectives, including feature extraction, phenotype classification, and clustering. Section 5 reviews a number of prevalent approaches for profiling drug effects based on the quantitative phenotypic data. Section 6 lists major, publicly available software packages of bioimage informatics

analysis, and finally, a brief summary is provided in Section 7.

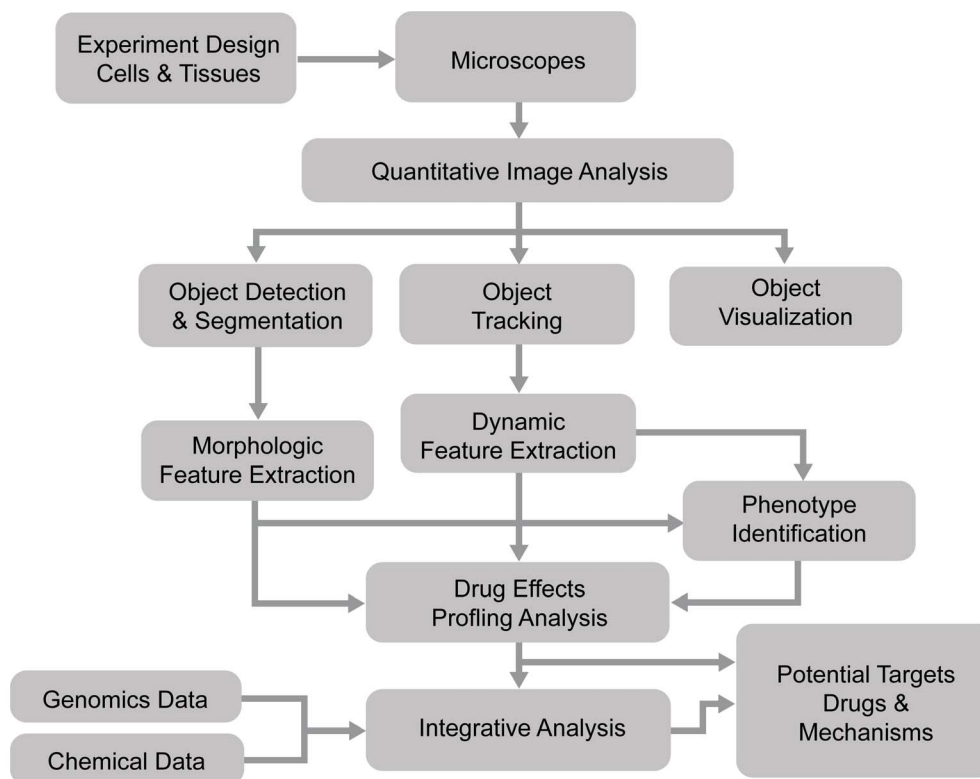
## 2. Example Image-based Studies for Drug and Target Discovery

There are a variety of image-based studies for discovery of drugs, targets, and mechanisms of biological processes. A good starting point for learning about bioimage informatics approaches is to study practical image-based studies, and a number of examples are summarized below.

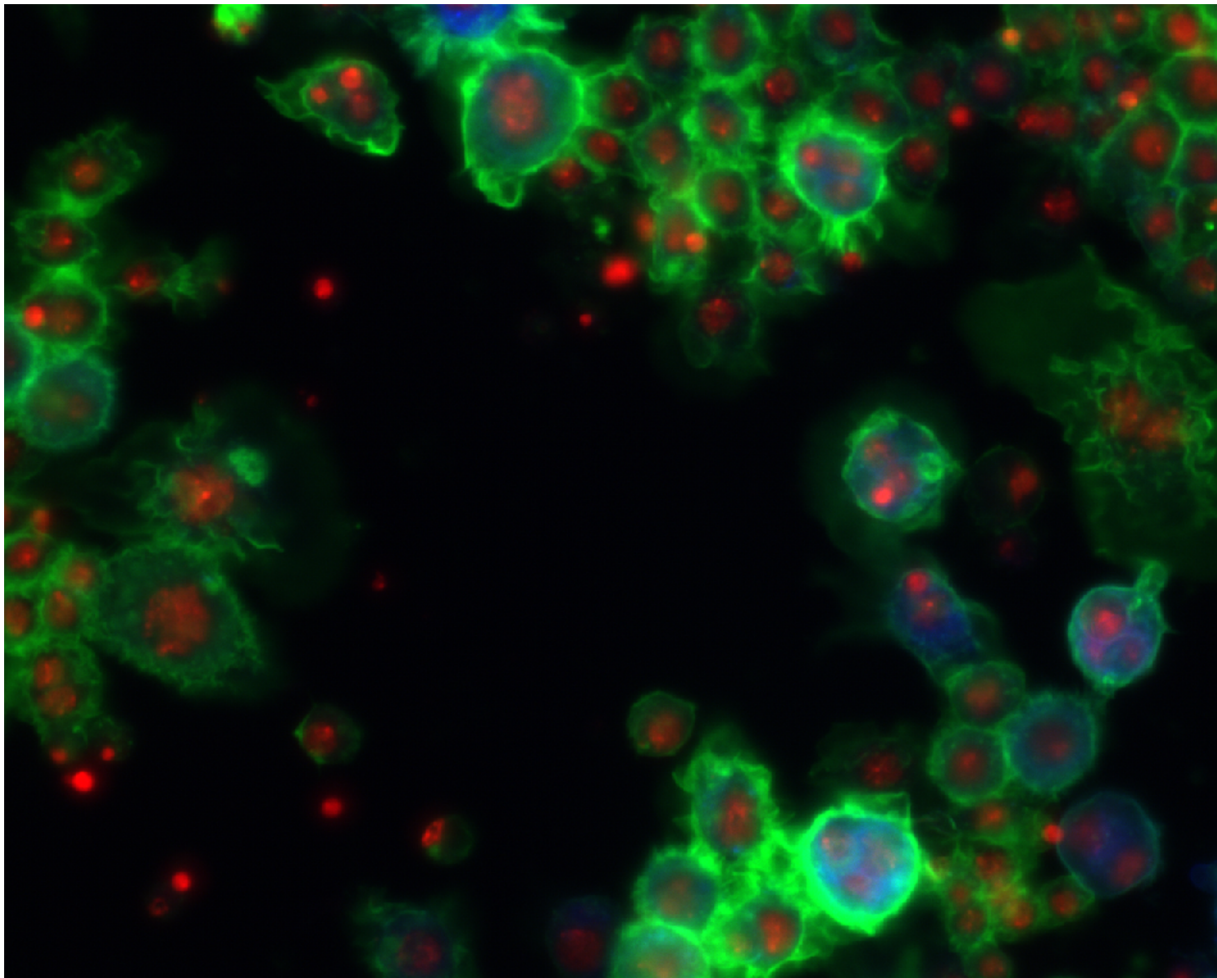
### 2.1 Multicolor Cell Imaging-based Studies for Drug and Target Discovery

Fixed cell images with multiple fluorescent markers have been widely used for drug and target screening in scientific

research. For example, the effects of hundreds of compounds were profiled for phenotypic changes using multicolor cell images in [10–12]. Hundreds of quantitative features were extracted to indicate the phenotypic changes caused by these compounds, and then computational approaches were proposed to identify the effective compounds, categorize them, characterize their dose-dependent response, and suggest novel targets and mechanisms for these compounds [10–12]. Moreover, phenotypic heterogeneity was investigated by using a subpopulation based approach to characterize drug effects in [13], and distinguish cell populations with distinct drug sensitivities in [14]. Also in [15,16], the phenotypic changes of proteins inside individual *Drosophila* Kc167 cells treated with RNAi libraries were investigated by using high resolution fluorescent microscopy, and bioimage informatics analysis was applied to quantify these images to identify genes regulating the phenotypic changes of interest. Figure 2 shows an image of *Drosophila* Kc167 cells, which were treated with RNAi and stained to visualize the nuclear DNA (red), F-actin (green), and  $\alpha$ -tubulin (blue). Freely available software packages, such as CellProfiler [17], Fiji



**Figure 1. The flowchart of bioimage informatics for drug and target discovery.**  
doi:10.1371/journal.pcbi.1003043.g001



**Figure 2. A representative image of *Drosophila* Kc167 cells treated with RNAi.** The red, green, and blue colors are the DNA, F-actin, and  $\alpha$ -tubulin channels.  
doi:10.1371/journal.pcbi.1003043.g002

[18], Icy [19], GCELLIQ [20], and PhenoRipper [21] can be used for the multicolor cell image analysis.

## 2.2 Live-cell Imaging-based Studies for Cell Cycle and Migration Regulator Discovery

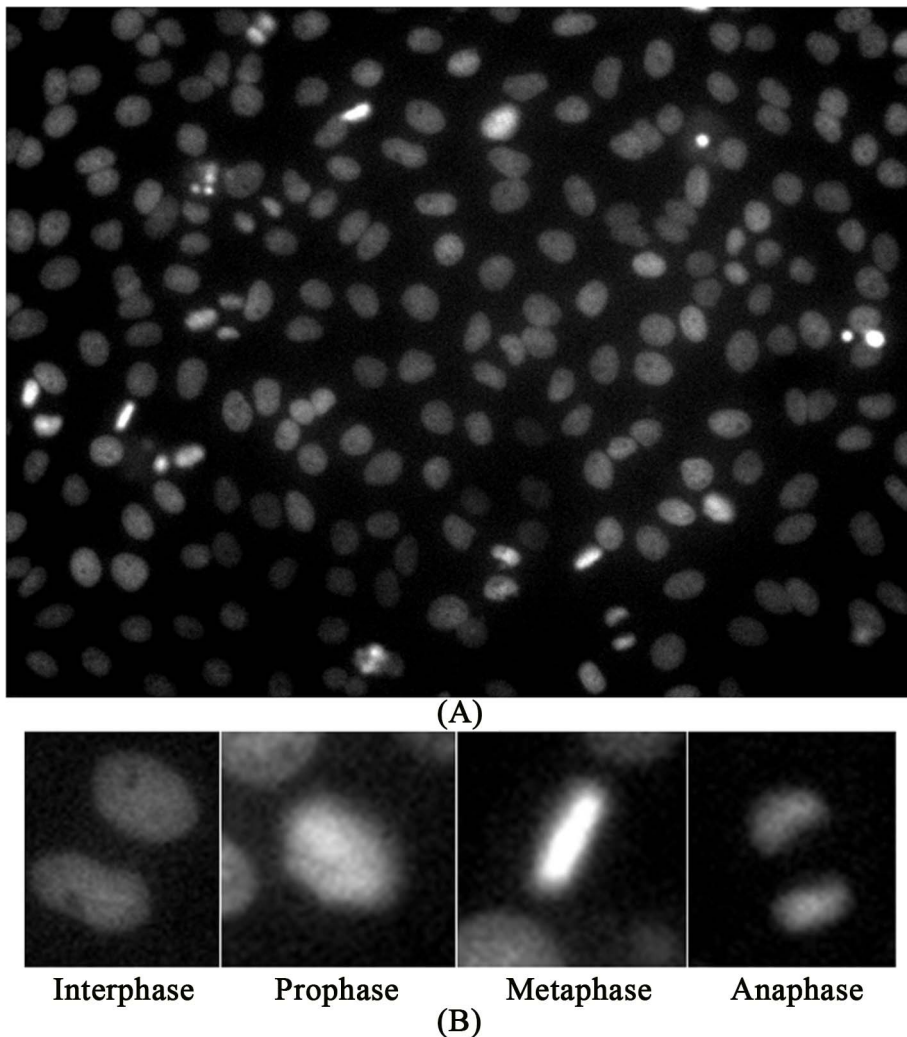
Two hallmarks of cancer cells are uncontrolled cell proliferation and migration. These are also good phenotypes for screening drugs and targets that regulate cell cycle progression and cell migration in time-lapse images. For example, out of 22,000 human genes, about 600 were identified as related to mitosis by using live cell (time-lapse) imaging and RNAi treatment in the MitoCheck project ([www.mitocheck.org](http://www.mitocheck.org)) [22,23]. The project is now being expanded to study how these identified genes work together to regulate cell mitosis, in which mistakes can lead to cancer, in the MitoSys (systems biology of mitosis) project (<http://www.mitosys.org/>). Also, live cell imaging of HeLa cells was used to discover drugs and

compounds that regulate cell mitosis in [24,25]. Moreover, the time-lapse images of live cells were used to study the dynamic behaviors of stem cells in [26,27] and predict cell fates of neural progenitor cells using their dynamic behaviors in [28]. Figure 3 shows a single frame of live HeLa cell images and the images of four cell cycle phases: interphase, prophase, metaphase, and anaphase [25]. The publicly available software packages for time-lapse image analysis include, for example, the plugins of CellProfiler [17], Fiji [18], BioimageXD [29], Icy [19], CellCognition [23], DCELLIQ [30], and TLM-Tracker [31].

## 2.3 Neuron Imaging-based Studies for Neurodegenerative Disease Drug and Target Discovery

Neuronal morphology is illustrative of neuronal function and can be instructive toward the dysfunctions seen in neurodegenerative diseases, such as Alzheimer's and Parkinson's disease [32,33]. For

example, the 3D neuron synaptic morphological and structural changes were investigated by using super-resolution microscopy, e.g., STED microscopy, to study brain functions and disorders under different stimulations [34–36]. Also other advanced optical techniques were proposed in [37,38] to image and reconstruct the 3D structure of live neurons. Figure 4 shows an example of 2D neuron image used in [39]. In [40], neuronal degeneration was mimicked by treating mice with different dosages of A $\beta$  peptide, which may cause the loss of neuritis, and drugs that rescue the loss of neurites were identified as candidates for AD therapy. Figure 5 shows an example of neurites and nuclei images acquired in [40]. To quantitatively analyze neuron images, a number of publicly available software packages have been developed, for example, NeurophologyJ [41], NeuronJ [42], NeuriteTracer (Fiji plugin) [43], NeuriteIQ [44], NeuronMetrics [45], NeuronStudio



**Figure 3. Examples of HeLa cell nuclei and cell cycle phase images.** (A) A frame of HeLa cell nuclei time-lapse image sequence; (B) Example images of four cell cycle phases.  
doi:10.1371/journal.pcbi.1003043.g003

[46,47], NeuronJ [42], NeuronIQ [39,48], and Vaa3D [49,50]. A review of software packages for neuron image analysis was also reported in [51].

#### 2.4 *Caenorhabditis elegans* Imaging-based Studies for Drug and Target Discovery

*Caenorhabditis elegans* (*C. elegans*) is a common animal model for drug and target discovery. Consisting of only hundreds of cells, it is an excellent model to study cellular development and organization. For example, the invariant embryonic development of *C. elegans* was recorded by time-lapse imaging, and the embryonic lineages of each cell were then reconstructed by cell tracking to study the functions of genes underpinning the development process [52–54]. Moreover, an atlas of *C. elegans*, which quantified the nuclear locations and statistics on their spatial patterns in devel-

opment, was built based on the confocal image stacks via the software, CellExplorer [55,56]. In addition, CellProfiler provides an image analysis pipeline for delineating bodies, and quantifying the expression changes of specific proteins, e.g., clec-60 and pharynx, of individual *C. elegans* under different treatments [57].

These examples have demonstrated diverse cellular phenotypes in different image-based studies. To quantify and analyze the complex phenotypic changes of cells and sub-cellular components from large scale image data, bioimage informatics approaches are needed.

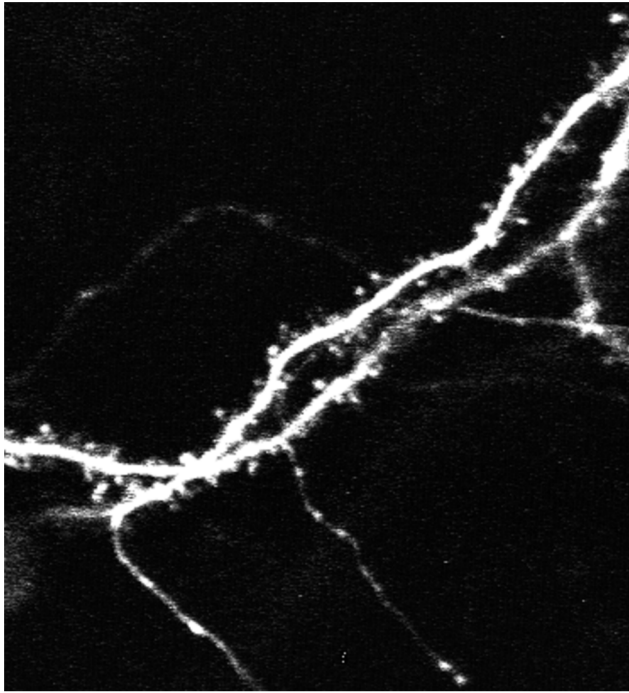
### 3. Quantitative Bioimage Analysis

After image acquisition, phenotypic changes need to be quantified for characterizing functions of drugs and targets.

Due to the large amounts of images generated, it is not feasible to quantify the images manually. Therefore, automated image analysis is essential for the quantification of phenotypic changes. In general, the challenges of quantitative image analysis include object detection, segmentation, tracking, and visualization. The word ‘object’ in this context means the object captured in the bioimages, e.g., the nucleus and cell. The following sections will introduce techniques used to address these challenges.

#### 3.1 Object Detection

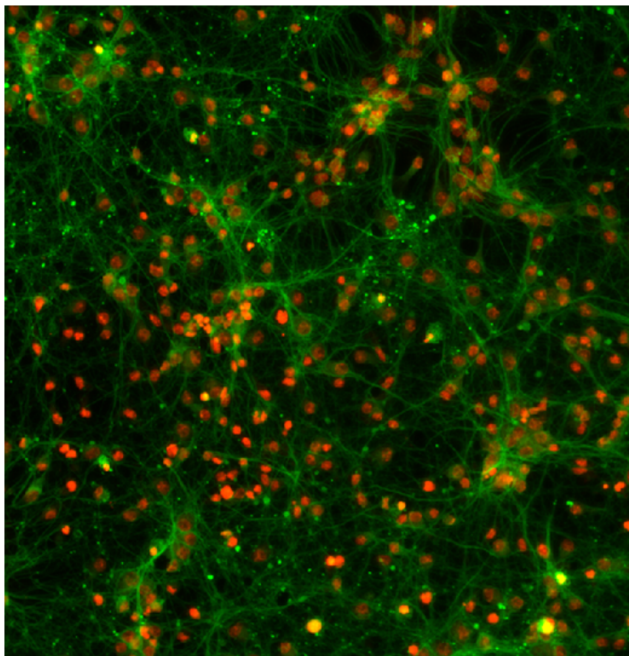
Object detection is to detect the locations of individual objects. It is important, especially when the objects cluster together, to facilitate the segmentation task by providing the position and initial boundary information of individual objects. Based on the shape of objects, two



**Figure 4. A representative 2D neuron images.** The bright spots near the backbones of neurons are the dendritic spines.  
doi:10.1371/journal.pcbi.1003043.g004

categories of object detection techniques are developed: blob structure detection, e.g., particles and cell nuclei, and tube structure detection, e.g., neurons, blood vessels.

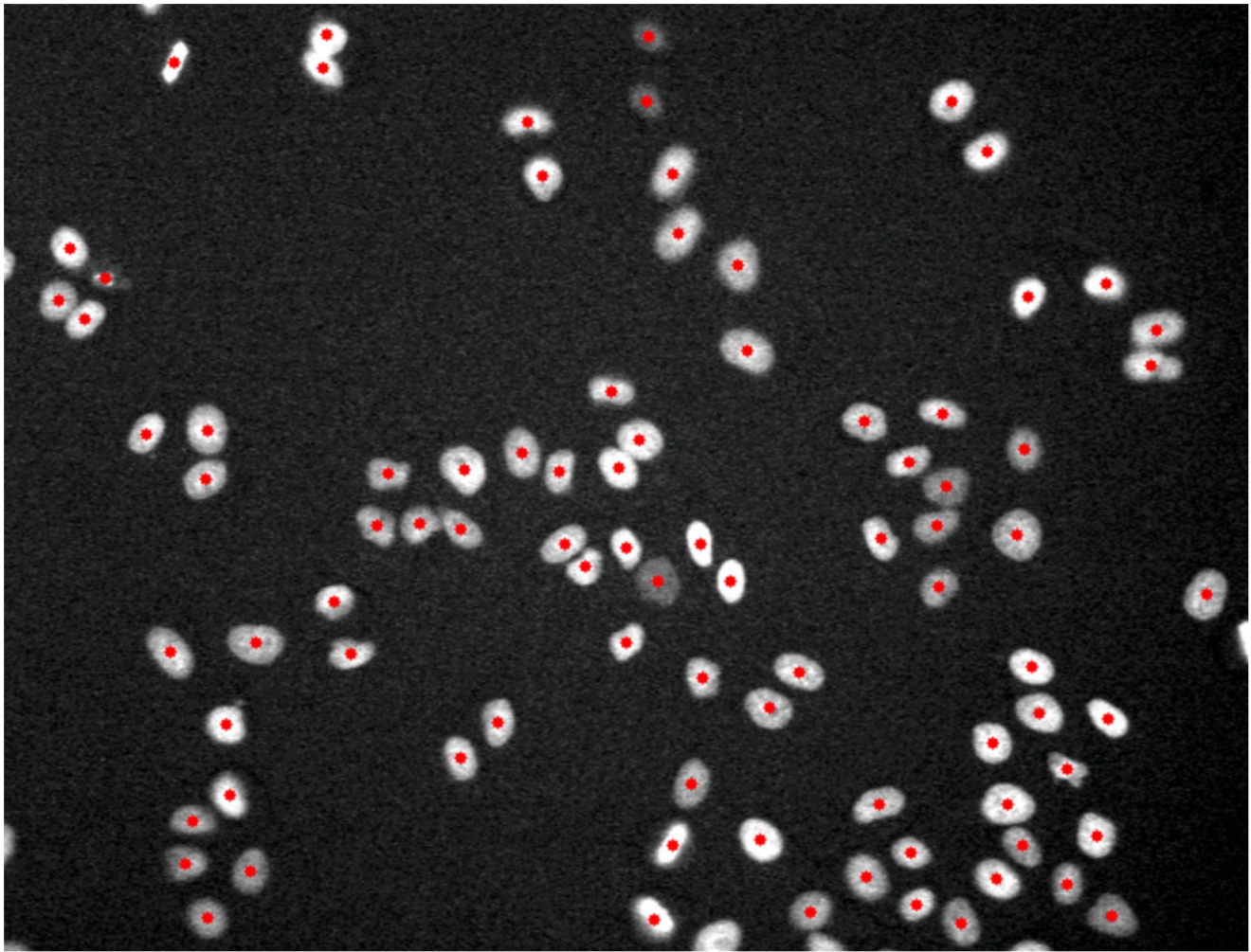
The shape information of blob objects can be used to detect the centers of objects using distance transformation [58]. The concavity of two touching objects would cause two local maxima in the distance



**Figure 5. A representative image of neurites.** Red indicates nuclei and green represents neurites.  
doi:10.1371/journal.pcbi.1003043.g005

image, such that thresholding or seeded watershed can be employed to the distance image to detect and separate the touching blob objects [59]. The intensity information is also often used for blob detection. Blob objects usually have relatively high intensity in the center, and relatively low intensity in the peripheral regions. For example, the Laplacian-of-Gaussian (LOG) filter is effective [60–63] to detect blob objects based on the intensity information. After LOG filtering, local maximum response points often correspond to centers of blob objects, as shown in Figure 6. Moreover, the intensity gradient information is also used for blob detection. For example, in [64] the intensity gradient vectors were smoothed by using the gradient vector flow approach [65] so that the smoothed gradient vectors continuously point to the object centers. Consequently, the blob object centers can be detected by following the gradient vectors [64]. In addition, the boundary points of blob objects with high gradient amplitude can be used to detect their centers, based on the idea of Hough Transform [66]. For example, in [67] an iterative radial voting method was developed to detect such object centers based on the boundary points. In brief, the detected boundary points vote the blob center with oriented kernels iteratively, and the orientation and size of the kernels are updated based on the voting results. Finally, the maximum response points in the voting image are selected as the centers of objects. The advantage of this method is that it can detect the centers of objects with noise appearance [67]. The distance transform and the intensity gradient information also can be combined for the object detection [68]. For other blob objects with complex appearances, the machine learning approaches based on local features [69,70] can also be used for object detection [71,72], as in the Fiji (trainable segmentation plugin) [18] and Ilastik [73].

Tubular structure detection is based on the premise that the intensity remains constant in the direction along the tube, and varies dramatically in the direction perpendicular to the tube. To find the local direction of tube center lines, the eigenvector corresponding to the minimum and negative eigenvalue of Hessian matrix was proposed in [44,74]. Center line points can be characterized by their local geometric attributes, i.e., the first derivative is close to zero and the magnitude of second derivatives is large in a direction perpendicular to tube center line [42,44,74]. After the center line point



**Figure 6. An example of blob-structure (HeLa cell nuclei) detection.** The red spots indicate the detected centers of objects.  
doi:10.1371/journal.pcbi.1003043.g006

detection, a linking process is needed to connect these center line points into continuous center lines based on their direction and distance. For example, in NeuronJ, Dijkstra's shortest-path was used based on the Gaussian derivative features to detect the neuron's centerline between two given points on the neuron [42]. Figure 7 provides an example of neurite images, and Figure 8 shows the corresponding centerline detection results [44] based on the local Gaussian derivative features. In addition to the approaches based on Gaussian derivatives, there are other tubular structure detection approaches. For example, four sets of kernels (edge detectors) were designed to detect the neuron edges and centerlines [75], and super-ellipsoid modeling was designed to fit the local geometry of blood vessels [76].

Moreover, machine learning-based tubular structure detection is a widely

used method. For example, blood vessel detection in retinal images is a representative tubular structure detection task with the supervised learning approaches [77,78]. In these methods, the local features, e.g., intensity and wavelet features, of an image patch containing a given pixel are calculated, and then a classifier is trained using these local features based on a set of training points [77,78]. A good survey of blood vessel (tube structure) detection approaches in retinal images was reported in [79]. For more approaches and details of tubular structure detection, readers should refer to the aforementioned neuron image analysis software packages.

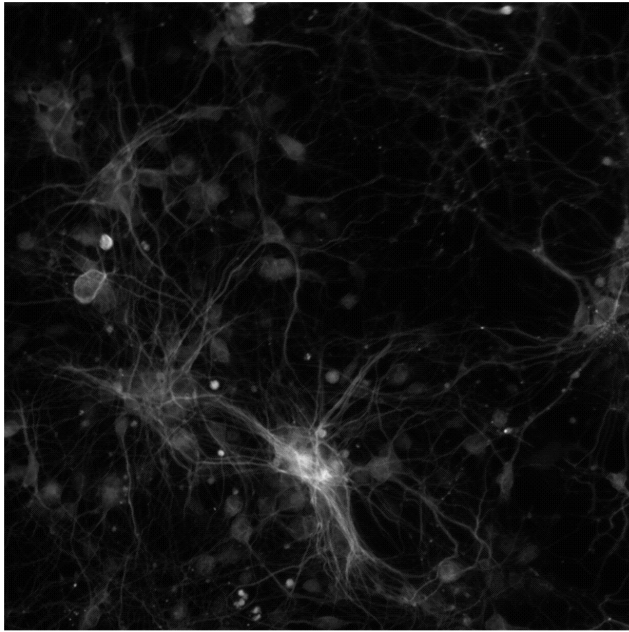
In summary, blobs and tubes are the dominating structures in bioimages. The detection results provide the position and initial boundary information for the quantification and segmentation processes. In other words, the segmentation process tries to delineate boundaries of objects

starting from the detected centers or centerlines of objects. Without the guidance of detection results, object segmentation would be more challenging.

### 3.2 Object Segmentation

The goal of object segmentation is to delineate boundaries of individual objects of interest in images. Segmentation is the basis for quantifying phenotypic changes. Although a number of image segmentation methods have been reported, this remains an open challenge due to the complexity of morphological appearances of objects. This section introduces a number of widely used segmentation methods.

Threshold segmentation [80] is the simplest method:  $T(I) = \begin{cases} 1; & t_2 > I(x,y) > t_1 \\ 0; & \text{otherwise} \end{cases}$ , where  $I(x,y)$  is the image, and  $t_1$  and  $t_2$  are the intensity thresholds. As an extension of the thresholding method, Fuzzy-C-



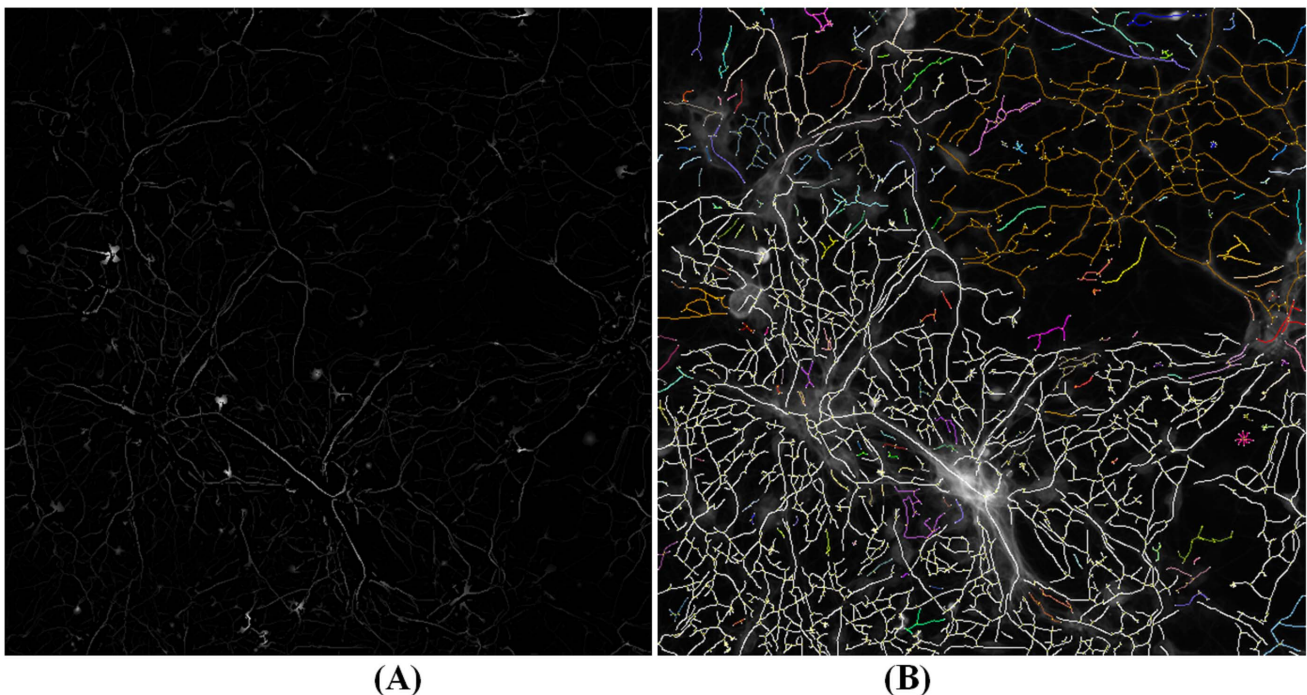
**Figure 7. A representative neurite image for centerline detection.**  
doi:10.1371/journal.pcbi.1003043.g007

Means [81] can be used to separate images into more regions based on intensity information. These methods could divide the image into objects and background, but fail to separate the object

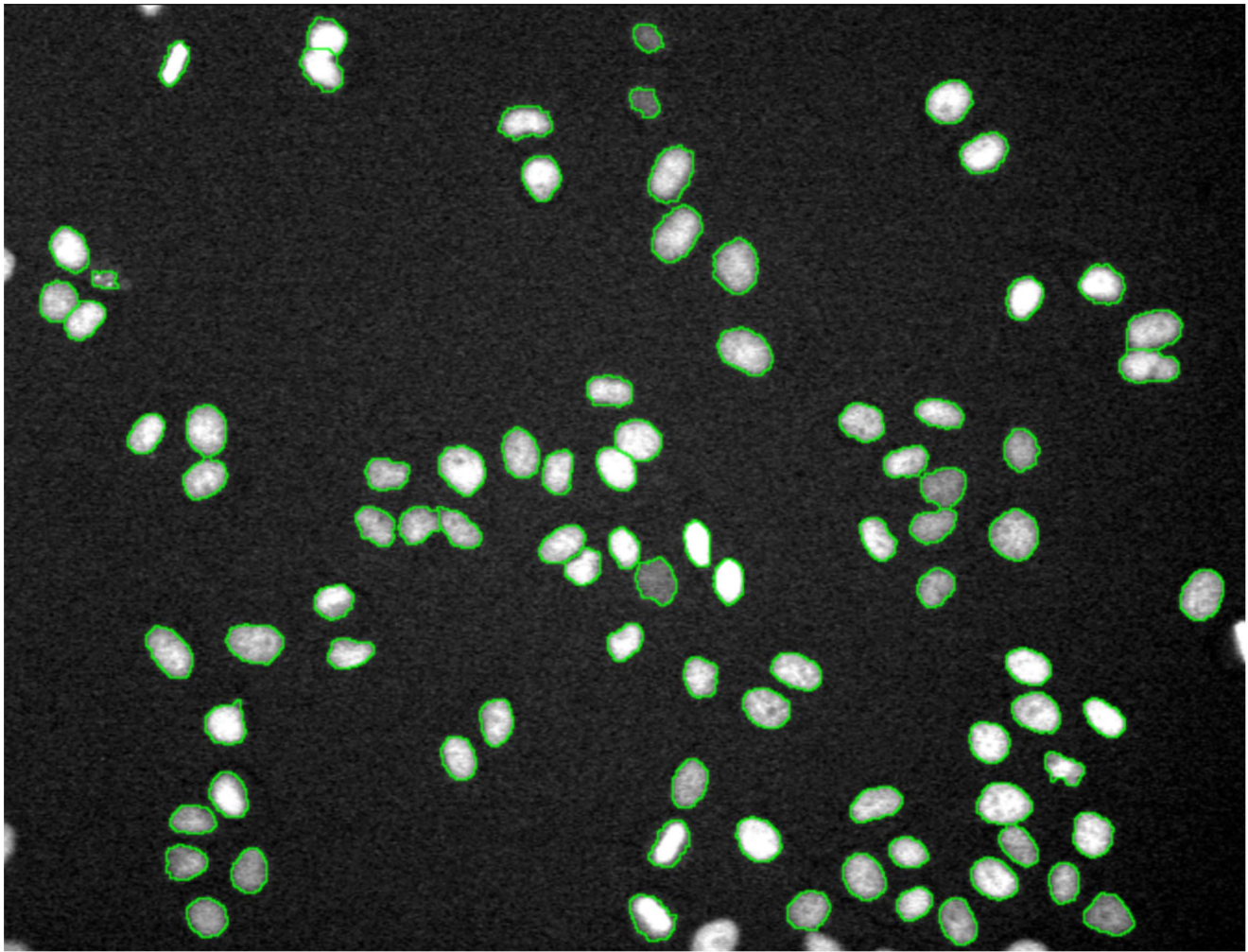
clumps (i.e., multiple objects touching together). Watershed segmentation and its derivatives are widely used segmentation methods. They build object boundaries between objects on the

pixels with local maximum intensity, which act like dams to avoid flooding from different basins (object regions) [82]. To avoid the over-segmentation problem of the watershed approach, the marker-controlled watershed (or seeded watershed) approach, in which the floods are from the ‘marker’ or ‘seed’ points (the object detection results), was proposed [68,83–85]. Figure 9 illustrates the segmentation result of HeLa cell nuclei using the seeded watershed method based on the cell detection results.

Active contour models are another set of widely used segmentation methods [86–90]. Generally, there are two kinds of active contour models: boundary-driven and region-competition models. In the boundary-driven model, the contours’ (boundaries of objects) evolution is determined by the local gradient. In other words, the boundary fronts move toward the outside (or inside) quickly in the regions with low intensity variation (gradient), and slowly in the regions with high gradient (where the boundaries are). When great intensity variation appears inside cells, or the boundary is weak, this method often fails [91]. Instead of using gradient information, the region-competition model makes use of the intensity similarity



**Figure 8. An example of neurite centerline detection.** (A) The centerline confidence image obtained by using the local Gaussian derivative features. Higher intensity indicates higher confidence of pixels on the centerlines. (B) The neurite centerline detection result image. Different colors indicate the disconnected branches.  
doi:10.1371/journal.pcbi.1003043.g008



**Figure 9. An example of HeLa nuclei segmentation using the seeded watershed algorithm.** The green contours are the boundaries of nuclei.

doi:10.1371/journal.pcbi.1003043.g009

information to separate the image into regions with similar intensity. Region competition-based active contour models could solve the weak boundary problem; however, they require that the intensity of touching objects is separable [87]. To implement these active contour models, level set representation is widely used [92]. Level set is an  $n+1$  dimensional function that can easily represent any  $n$  dimensional shape without parameters. The inside regions of objects are indicated by using positive levels, and outside regions are represented using negative levels. For this implementation, the initial boundary (zero level) is required, and the signed distance function is often used to initialize the level set function [92,93]. To evolve the level set functions (grow the boundaries of objects), the following two equations are classical models. The first equation is often called geodesic active contour (GAC) [86], and the second one is often named the Chan

and Vese active contour (CV) [87].

$$\frac{d}{dt}\psi = \alpha(\nabla g \cdot \nabla \psi) + g(\kappa + c)|\nabla \psi|$$

(GAC level set evolution equation),

$$\frac{d}{dt}\psi = \delta_\varepsilon(\psi) \left[ \mu\kappa - \nu - \lambda_1(I - c_1)^2 + \lambda_2(I - c_2)^2 \right]$$

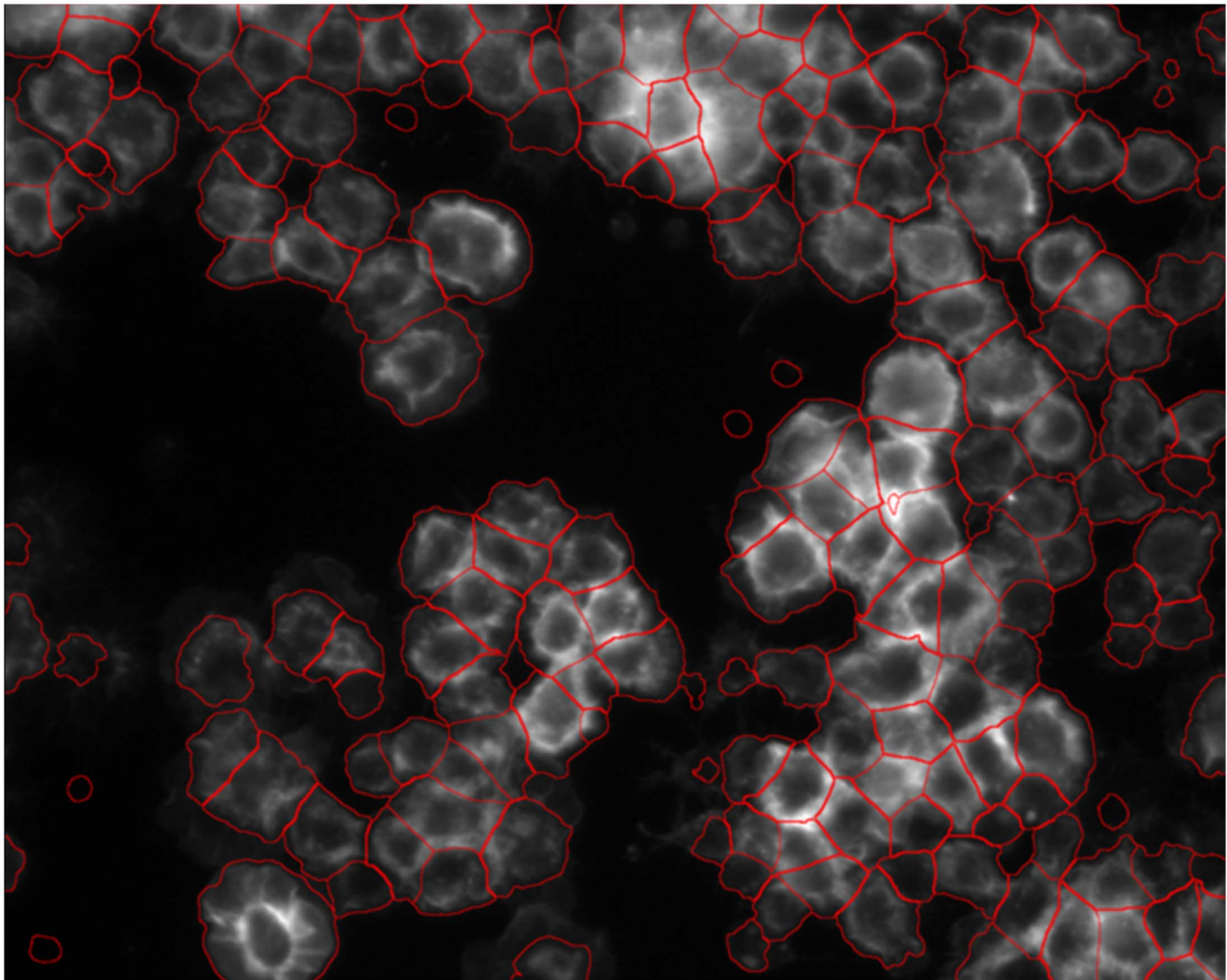
(CV level set evolution equation),

where  $\psi$  denotes the level-set function, and  $g$  indicates the gradient function,  $\nabla$  is the gradient operator,  $c$ ,  $c_1$ , and  $c_2$  are constant variables.  $\delta_\varepsilon(x) = \frac{1}{\pi} \frac{\varepsilon}{\varepsilon^2 + x^2}$  is an approximation of the Dirac function to indicate the boundary bands), which is the derivative function of Heaviside function denoting inside/outside regions of objects:

$$H(x) = \frac{1}{2} \left[ 1 + \frac{2}{\pi} \arg \tan \left( \frac{x}{\varepsilon} \right) \right], \text{ and the curvature term, } \kappa = \text{div} \left( \frac{\nabla \psi}{|\nabla \psi|} \right) = \frac{\psi_{xx}\psi_y^2 - 2\psi_x\psi_{xy}\psi_y + \psi_x^2\psi_{yy}}{(\psi_x^2 + \psi_y^2)^{3/2}}$$

indicates the local smoothness of boundaries, and 'div' is the divergence operation. Figure 10 demonstrates the segmentation result using GAC level set approach. An additional segmentation method, Voronoi segmentation [94], first defines the centers of objects and then constructs the boundaries between two objects on the pixels, from which the distances are the same to the two centers. In CellProfiler, the Voronoi segmentation method was extended by considering the local intensity variations in the distance metric to achieve better segmentation results [95]. This method is fast and generates level set comparable





**Figure 10. An example of segmentation of *Drosophila* cell images using the level set approach.**  
doi:10.1371/journal.pcbi.1003043.g010

results. Graph cut segmentation method views the image as a graph, in which each pixel is a vertex and adjacent pixels are connected [63,96,97]. It ‘cuts’ the graph into several small graphs from the regions where adjacent pixels have the most different properties, e.g., intensity.

Different from the aforementioned segmentation approaches, local feature and machine learning-based segmentation approaches are implemented, for example, in Fiji (trainable segmentation plugin) [18] and Ilastik [73]. Users can interactively select the training sample pixels/voxels or small image patches conveniently, and then classifiers are automatically trained based on the features of the training pixels or voxels (or patches) to predict the classes, e.g., cells or background, of the pixels or voxels (or patches) in a new image. The image patches could be a circle or square neighbor regions of a given point, and also

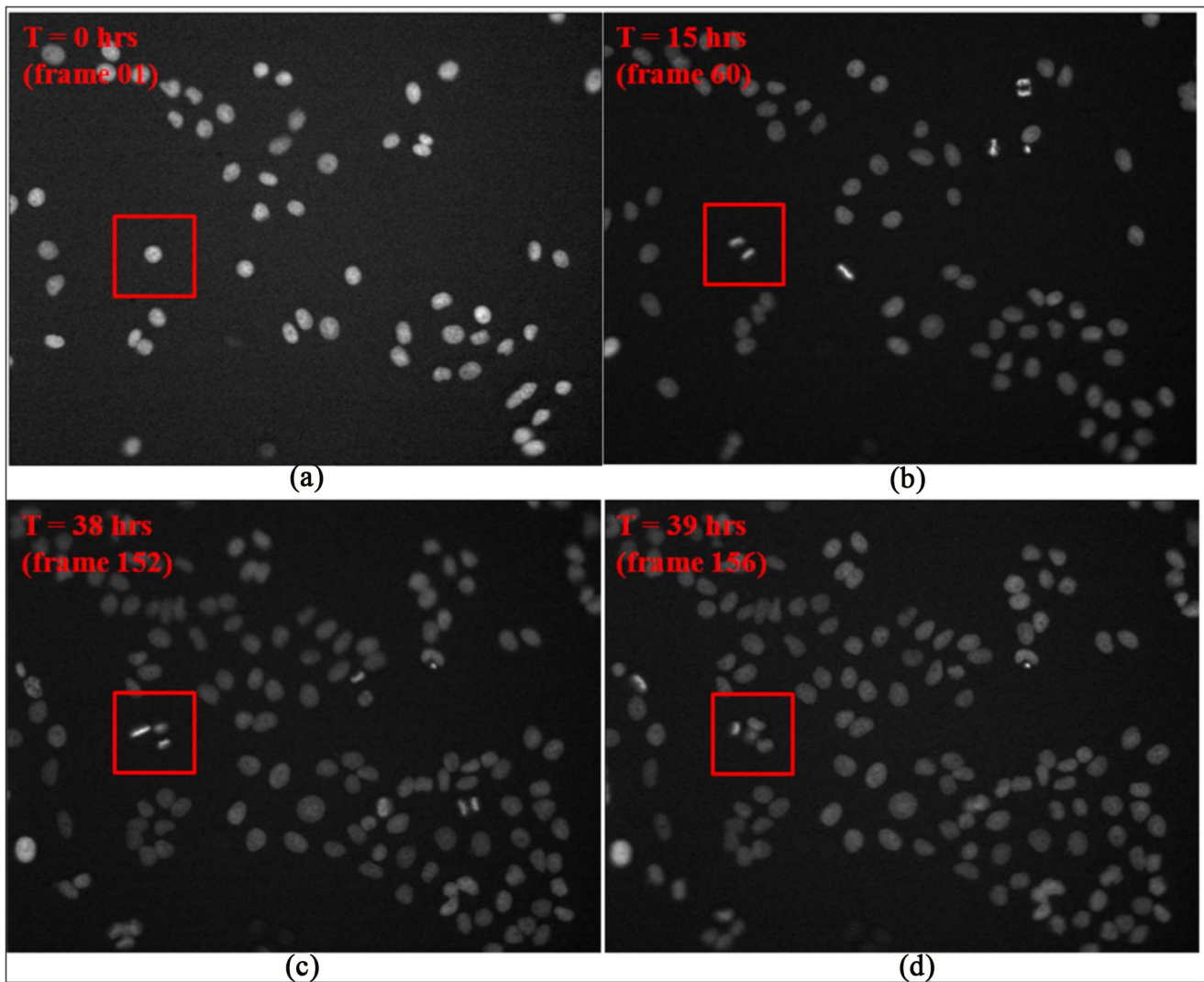
could be regions (superpixel) obtained by the clustering analysis. For example, Simple Linear Iterative Clustering (SLIC) made use of the intensity and coordinate information of pixels to separate the image into uniformly sized and biologically meaningful regions [98,99], and then the machine learning approaches were used to identify the regions of interest, e.g., boundary superpixels, for object segmentation [99].

### 3.3 Object Tracking

To study the dynamic behaviors and phenotypic changes of objects over time (e.g., cell cycle progression and migration), object tracking using time lapse image sequences is necessary. Figure 11 shows a HeLa cell’s division process in four frames at different time points, and Figures 12 and 13 show the examples of cell migration trajectories and cell lineages reconstructed from the time-lapse images of

HeLa cells [30]. Object tracking is a challenging task due to the complex dynamic behaviors of objects over time. In general, cell tracking approaches can be classified into three categories: model evolution-based tracking, spatial-temporal volume segmentation-based tracking, and segmentation-based tracking.

In the model evolution based tracking approaches, cells or nuclei are initially detected and segmented in the first frame, and then their boundaries and positions evolve frame by frame. Some tracking techniques in this category are mean-shift [100] and parametric active contours [88,101]. However, neither mean-shift nor parametric active contours can cope well with cell division and nuclei clusters. Though the level set method enables topological change, e.g., cell division, it also allows the fusion of overlapping cells. Extending these methods to cope with



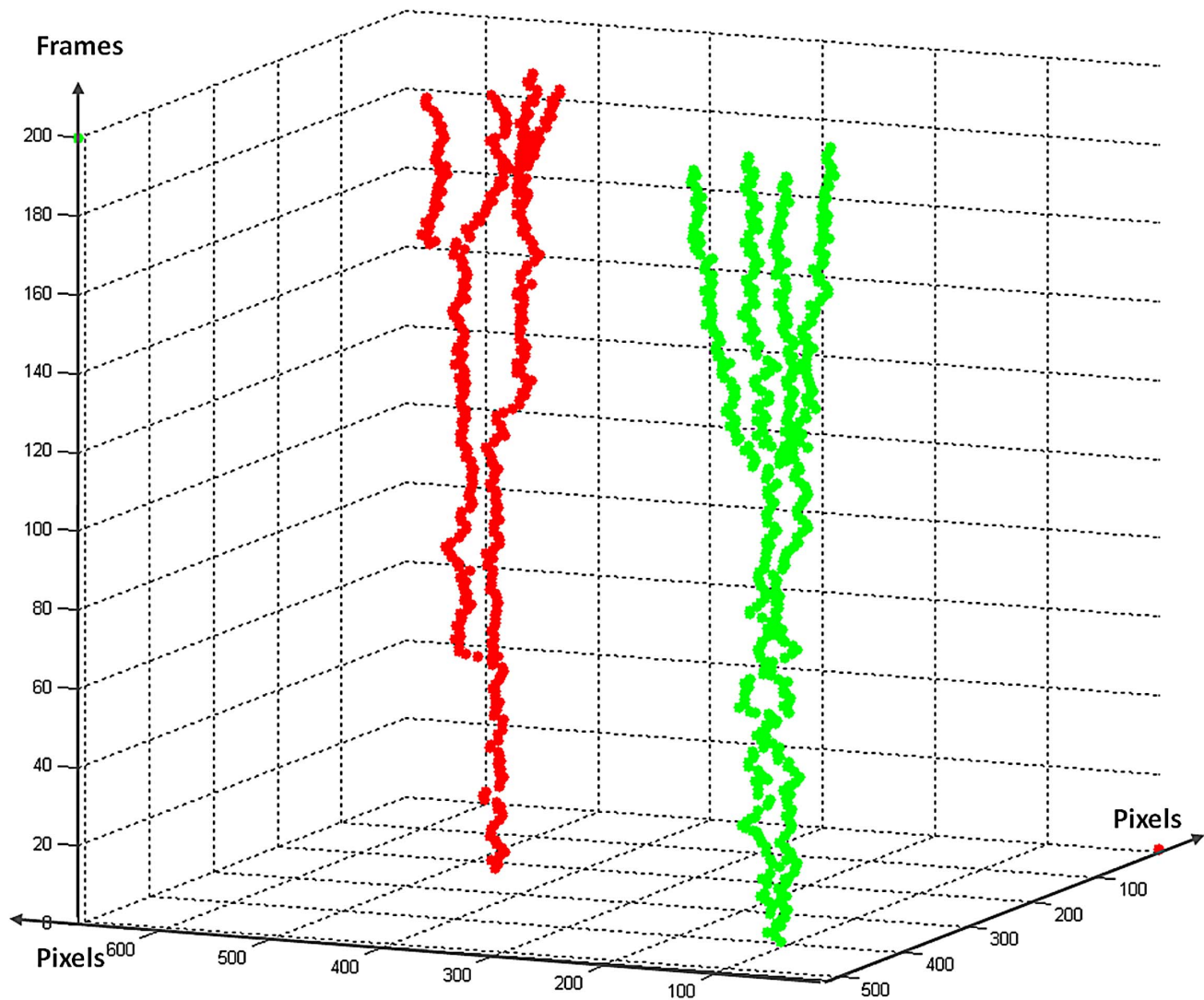
**Figure 11. Time-lapse images indicating cell cycle progression.** The cell in the red square in the first frame (A) divided into two cells in frame 60 (B). The descendent cells divided again in frame 152 and 156 respectively as shown in the red squares in (C) and (D).  
doi:10.1371/journal.pcbi.1003043.g011

these tracking challenges is nontrivial and increases computation time [90,102–104]. For example, the coupled geometric active contours model was proposed to prevent object fusion by representing each object with an independent level set in [105], and this was further extended to the 3D cell tracking in [90]. The other approach explicitly blocking the cell merging is to introduce the topology constraints, i.e., labeling objects regions with different numbers or colors. For example, the region labeling map was employed in [27,106] to deal with the cell merging, and planar graph–vertex coloring was employed to separate the neighboring contours. From that four separate level set functions could easily deal with cell merging [107] based on the four-color theorem [108,109]. For the spatial-tempo-

ral volume segmentation based tracking, 2D image sequences were viewed as 3D volume data (2D spatial+temporal), and the shape and size constrained level set segmentation approaches were applied to segment the traces of objects, and reconstruct the cell lineage in [110–112].

For detection and segmentation-based tracking, objects are first detected and segmented, and then these objects are associated between two consecutive frames, based on their morphology, position, and motion [30,113–115]. The tracking approaches are usually done fast, but their accuracy is closely related to detection and segmentation results, similarity measurements, and association strategies. The cell center position, shape, intensity, migration distance, and spatial context information were used as similar-

ity measurements in [113,115]. For the association approaches, the overlap region and distance based method was employed in [114], in which objects in the current frame were associated with the nearest objects in the next frame. Then the false matches, e.g., many-to-one or one-to-many, were further corrected through the post processing. Different from the individual object association above, all segmented objects were simultaneously associated by using the integer programming optimization in [113,116]:  $\mathbf{x}^* = \max_{\mathbf{x} \in \{0,1\}^N} \mathbf{S}\mathbf{x}$ , *s.t.*  $\mathbf{A}\mathbf{x} \leq 1$ , where  $\mathbf{A}\mathbf{x} \leq 1$  restricts that one object can be associated to one object at most,  $\mathbf{A}$  is an  $(m+n) \times N$  matrix, and the first  $m$  rows correspond to  $m$  objects in frame  $t$ , and the last  $n$  rows denote objects in frame  $t+1$ .  $N$  is the



**Figure 12. Examples of cell migration trajectories.** Different colors represent different trajectories.  
doi:10.1371/journal.pcbi.1003043.g012

number of all possible associations among objects in frame  $t$  and frame  $t+1$ .  $\mathbf{S}$  is a  $1 \times N$  similarity matrix, and  $\mathbf{S}(j) = S(\mathbf{c}_{kt+1} | \mathbf{c}_{it})$ . For the unmatched cells, e.g., the new born or new entered cells, a linking process is usually needed to link them to the parent cells or as a new trajectory. This optimal matching strategy was also used to link the object trajectory segments in [27] to link the broken or newly appearing trajectories.

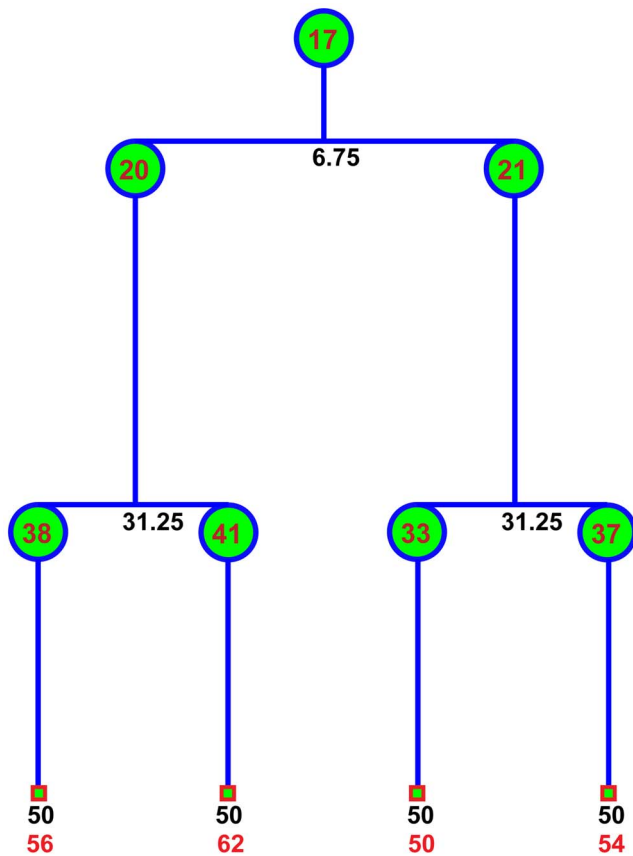
As an alternative to frame-by-frame association strategies, Bayesian filters, e.g., Particle filter and Interacting Multiple Model (IMM) filters [117,118], are also used for object tracking. The goal of these filters is to recursively estimate a model of object migration in an image sequence. Generally, in the Bayesian methods, a state vector,  $\mathbf{x}_t$ , is defined to indicate the characters of objects, e.g., position,

velocity, and intensity. Then, two models are defined based on the state vector. The first is the state evolution model,  $\mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_{t-1}) + \varepsilon_t$ , where  $\mathbf{f}_t$  is the state evolution function at time point,  $t$ , and  $\varepsilon_t$  is a noise, e.g., Gaussian noise, which describes the evolution of the state. The other is the observation model,  $\mathbf{z}_t = \mathbf{h}_t(\mathbf{x}_{t-1}) + \eta_t$ , where  $\mathbf{h}_t$  is the map function, and  $\eta_t$  is the noise, which maps the state vector into observations that are measurable in the image. Based on the two models and Bayes' rule, the posterior density of the object state is estimated as follows:  $p(\mathbf{x}_t | \mathbf{z}_{1:t}) \propto p(\mathbf{z}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{z}_{1:t-1})$ , and  $p(\mathbf{x}_t | \mathbf{z}_{1:t-1}) = \int p(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} | \mathbf{z}_{1:t-1}) d\mathbf{x}_{t-1}$  where the  $p(\mathbf{z}_t | \mathbf{x}_t)$  is defined based on the observation model, and the  $p(\mathbf{x}_t | \mathbf{x}_{t-1})$  is defined based on the state evolution model. The basic principle of particle filter is to approximate the posterior density by a set of samples

(particles) being stochastically drawn, and it had been employed for object tracking in fluorescent images in [119–121]. In some biological studies, the motion dynamics of objects are complex. Therefore, one motion model might not be able to describe object motion dynamics well. The IMM filter is employed to incorporate multiple motion models, and the motion model of objects can be transitioned from one to another in the next frame with certain probabilities. For example, the IMM filter with three motion models, i.e., random walk, first-order, and second-order linear extrapolation, was used for 3D object tracking in [118], and for 2D cell tracking in [27].

### 3.4 Image Visualization

Most of the aforementioned software packages provide functions to visualize 2D images and the analysis results. However,



**Figure 13. Examples of cell lineages constructed by the tracking algorithm.** The black numbers are the time of cell division (hours). The bottom red numbers indicate the number of traces, and the numbers inside circles are the labels of cells in that frame.  
doi:10.1371/journal.pcbi.1003043.g013

for higher dimensional images, e.g., 3D, 4D (including time), and 5D (including multiple color channels), visualization is challenging. Fiji [18], Icy [19], and BioimageXD [29], for example, are the widely used bioimage analysis and visualization software packages for higher dimensional images. In addition, NeuronStudio [46,47] is a software package tailored for neuron image analysis and visualization. Farsight [122] and vaa3D [123] are also developed for analysis and visualization of 3D, 4D, and 5D microscopy images. For developing customized visualization tools, the Visualization Toolkit (VTK) is a favorite choice (<http://www.vtk.org/>) as it is open source and developed specifically for 3D visualization. ParaView (<http://www.paraview.org/>) and ITK-SNAP (<http://www.itksnap.org/>) are the popular Insight Toolkit (ITK) (<http://www.itk.org/>) and VTK based 3D image analysis and visualization software packages.

This section has introduced a number of major methods for object detection, segmentation, tracking, and visualization in bioimage analysis. These analyses are essen-

tial and provide a basis for the following quantification of morphological changes.

## 4. Numerical Features and Morphological Phenotypes

### 4.1 Numerical Features

To quantitatively measure the phenotypic changes of segmented objects, a set of descriptive numerical features are needed. For example, four categories of quantitative features, measuring morphological appearances of segmented objects, are widely used in imaging informatics studies for object classification and identification, i.e., wavelets features [124,125], geometry features [126], Zernike moment features [127], and Haralick texture features [128]. In brief, Discrete Wavelet Transformation (DWT) features characterize images in both scale and frequency domains. Two important DWT feature sets are the Gabor wavelet [129] and the Cohen–Daubechies–Feauveau wavelet (CDF9/7) [130] features. Geometry features describe the shape and texture features of the individual cells, e.g., the maximum value, mean value, and stan-

dard deviation of the intensity, the lengths of the longest axis, the shortest axis, and their ratio, the area of the cell, the perimeter, the compactness of the cell ( $compactness = perimeter^2 / 4\pi * area$ ), the area of the minimum convex image, and the roughness ( $area\ of\ cell / area\ of\ convex\ shape$ ). The calculation of Zernike moments features was introduced in [131]. First, the center of mass of the cell image was calculated, then the average radius for each cell was computed, and the pixel  $p(x, y)$  of the cell image was mapped to a unit circle to obtain the projected pixel as  $p(x', y')$ . Then Zernike moment features were calculated based on the projected image  $I(x', y')$ . The Haralick texture features are extracted from the gray-level spatial-dependence matrices, including the angular second moment, contrast, correlation, sum of the squares, inverse difference moment, sum of the average, sum of the variance, sum of entropy, entropy, difference of the variance, difference of entropy, information measures of correlation, and maximal correlation coefficient [132]. More descriptions and calculation programs about these Subcellular Location Features (SLF) and SLF-based machine learning approaches for image classification can be found at: <http://murphyllab.web.cmu.edu/services/SLF/features.html>.

### 4.2 Phenotype Identification

Although these numerical features are informative to describe the phenotypic changes, it can be difficult to understand these changes in terms of visual and understandable phenotypic changes. For example, the increase or decrease of cell size can be understood; however, it is not clear what the physical meaning of the increase or decrease is for certain wavelet features. Therefore, transforming the numerical features into biologically meaningful features (phenotypes) is important. This section introduces a number of widely used phenotype identification approaches.

**4.2.1. Cell cycle phase identification.** In cell cycle studies, drug and target effects are indicated by the dwelling time of cell cycle phases, e.g., interphase, prophase, metaphase and anaphase. Additional cell cycle phases, e.g., Prometa-, Ana 1-, Ana 2-, and Telophases, were also investigated in [133] and [23,134]. After object segmentation and tracking, cell motion traces can be extracted, as shown in Figure 14, and then the automated cell cycle phase identification is needed to calculate the dwelling time of individual cells on different phases.

Cell cycle phase identification can be viewed as a pattern classification problem. The aforementioned numerical features, and a number of classifiers can be used to identify the corresponding phases of individual segmented cells, e.g., support vector machine (SVM) [115,133,135], K-nearest neighbors (KNN), and naïve Bayesian classifiers [114]. However, the classification accuracy is often poor for cell cycle phases appearing for a short time, e.g., prophase and metaphase, due to the unbalance of sample size compared to interphase, and the segmentation bias. Fortunately, the cell cycle phase transition rules, e.g., from interphase to prophase, and from prophase to metaphase, can be used to reduce identification errors. Thus, a set of cell cycle phase identification approaches based on the cell tracking results were proposed to achieve high identification accuracy. This problem is often formulized as follows, and as shown in Figure 15. Let  $\mathbf{x} = (x_1, x_2, \dots, x_T)$  denote a cell image sequence of length  $T$ . Each cell image is represented by a numerical feature vector  $\varphi(x_i) \in \mathcal{R}^d$  (using the aforementioned numerical features). Let  $\mathbf{y} = (y_1, y_2, \dots, y_T)$  represent the corresponding cell cycle phase sequence that needs to be predicted. Based on the cell cycle progression rules, for example, the variation of nuclei size and intensity were used as an index to identify the mitosis phases of cells in [25], and Hidden Markov Modeling (HMM) was used to identify the cell cycle phases in CellCognition [23]. In brief, the transition possibility from one phase to the other was learned from the training data of cell cycle progressions, which could improve the accuracy of cell cycle phase identification. As an extension of HMM, Temporally Constrained Combinatorial Clustering (TC3), which is an unsupervised learning approach for cell cycle phase identification, was designed and combined with Gaussian Mixture Model (GMM) and HMM to achieve robust and accurate cell cycle identification results in [134]. Also, in [133] Finite State Machine (FSM) was employed to check the phase transition consistency and make corrections to the error cell cycle phases predicted by using SVM classifier [115]. Moreover, the cell cycle phases could be identified during the segmentation and linking process in the spatiotemporal volumetric segmentation-based tracking methods [110–112].

**4.2.2 User defined phenotype, identification, and classification.** In certain image-based studies, cells may not

have an intrinsic phenotype, e.g., cell cycle phases, but may exhibit unpredicted and novel phenotypes caused by experimental perturbations, e.g., drugs or RNAi treatments. These phenotypes are often defined by well-trained biologists to characterize drug and target effects [16]. Figure 16 shows images of *Drosophila* cells with three defined phenotypes: Normal, Ruffling and Spiky [136].

In large scale screening studies, however, it is subjective and time-consuming for biologists to uncover novel phenotypes from millions of cells. Thus, automated discovery of novel phenotypes is important. For example, an automated phenotype discovery method was proposed in [20]. In brief, a GMM was constructed first for the existing phenotypes. Then the quantitative cellular data from new cellular images were combined with samples generated from the GMM, and the cluster number of the combined data was estimated using gap statistics [137]. Then, clustering analysis was performed on the combined data set, in which some of the cells from the new cellular images were merged into the existing phenotypes, and the clusters that could not be merged by any existing phenotype classes were considered as new phenotype candidates. After the phenotypes are defined, classifiers can be built conveniently based on the training data and the numerical features for classifying cells into one of the predefined phenotypes. However, it is tedious to manually collect enough training samples of the rare and unusual phenotypes. To solve this challenge, an iterative machine learning based approach was proposed in [138]. First, a tentative rule (classifier) was determined based on a few samples of a given phenotype, and then the classifier presented users a set of cells that were classified into the phenotype based on the tentative rule. Users would then manually correct the classification errors, and the corrections are used to refine the rule. This method could collect plenty of training samples after several rounds of error correction and rule refinement [138].

This section introduced numerical feature extraction, phenotype identification, and classification. These analyses provide quantitative phenotypic change data for identifying candidate targets and drug hits that cause desirable phenotypic changes. The following section will describe approaches to analyze the quantitative phenotypic profile data for drug and target identification.

## 5. Multidimensional Profiling Analysis

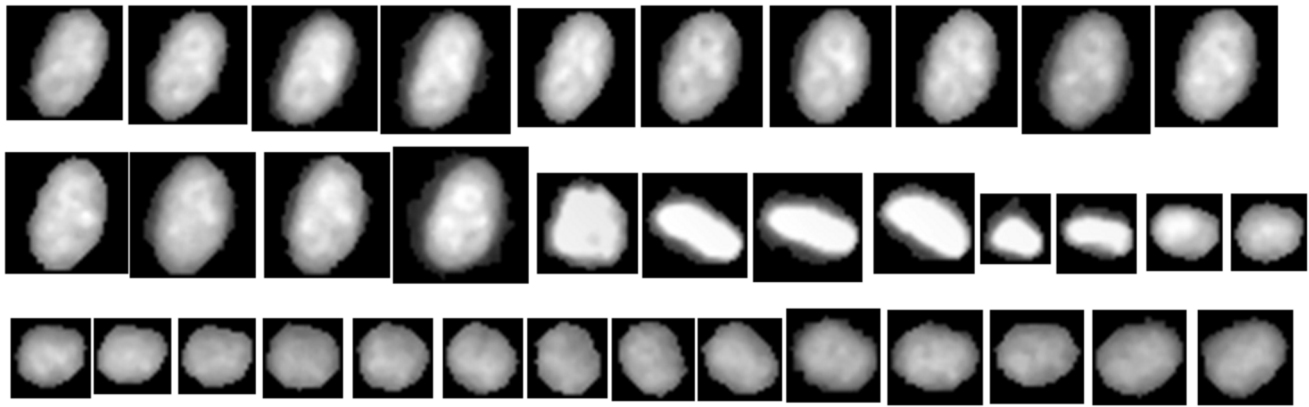
The aim of profiling analysis is to characterize the functions of drugs and targets, divide them into groups with similar phenotypic changes, and identify the candidates causing desired phenotypic changes. To help analyze and organize these multidimensional phenotypic profile data, some publicly available software packages have been designed, for example, CellProfiler Analyst (<http://www.cellprofiler.org/>) and PhenoRipper (<http://www.phenoripper.org/>). In addition, KNIME (<http://www.knime.org/>) is a publicly available pipeline and workflow system to help organize different data flows. It also provides connections to bioimage analysis software packages, e.g., Fiji [18] and CellProfiler [9], and enables users to conveniently build specific data analysis pipelines in KNIME. This section describes some prevalent approaches in analyzing quantitative phenotypic profile data.

### 5.1 Clustering Analysis

Clustering analysis is to divide experimental perturbations, e.g., drugs, RNAis, into groups that have similar phenotypic changes. As clustering analysis approaches, e.g., Hierarchical Clustering [139] and Consensus Clustering [140], are well established, their technical details will not be discussed here. In addition to the aforementioned software, Cluster 3.0 (<http://www.falw.vu/~huik/cluster.htm>) and Java TreeView (<http://jtreeview.sourceforge.net/>) are two additional easy-to-use clustering analysis software packages available in public domain.

### 5.2 SVM-based Multivariate Profiling Analysis

SVM classifier was employed for analyzing the multivariate drug profiles in [141]. To measure the phenotypic change caused by drug treatments, the cell populations harvested from the drug-treated wells were compared with cells collected from the control wells (no drug treatment). The difference between the control and drug treatment was indicated by two factors that are the outputs of the SVM classifier. One is the accuracy of classification, which indicates the magnitude of the drug effect. The other is the normal vector (d-profile) of the hyperplane separating the two cell populations, which indicates the phenotypic changes caused by the drug. Figure 17 illustrates the idea; the yellow arrow is the d-profile indicating the direction of drug effects in the



**Figure 14. A segment of cell cycle procession sequence.** Four cell cycle phases, interphase, prophase, metaphase, and anaphase, appear in order. doi:10.1371/journal.pcbi.1003043.g014

phenotypic feature space. Drugs with similar d-profiles were found to have the same functional targets, and thus it could be used to predict functions of new drugs or compounds.

### 5.3 Factor-based Multidimensional Profiling Analysis

In the set of numerical features, some are highly correlated within groups but poorly correlated with features in other groups. One possible explanation is that the features in one group measure a common biological process, such as increase or decrease of nuclei size. The challenge using these numerical features directly is that biological meanings of certain phenotypic features are often vague. It is thus difficult to explain the phenotypic changes represented by these numerical features as aforementioned. To remove the redundant features and make the biological meanings of numerical

features explicitly clear, factor analysis was employed in [12]. The basic principle of factor analysis is to determine the independent common ‘traits’ (factors). Mathematically it is formulated by the following equation.

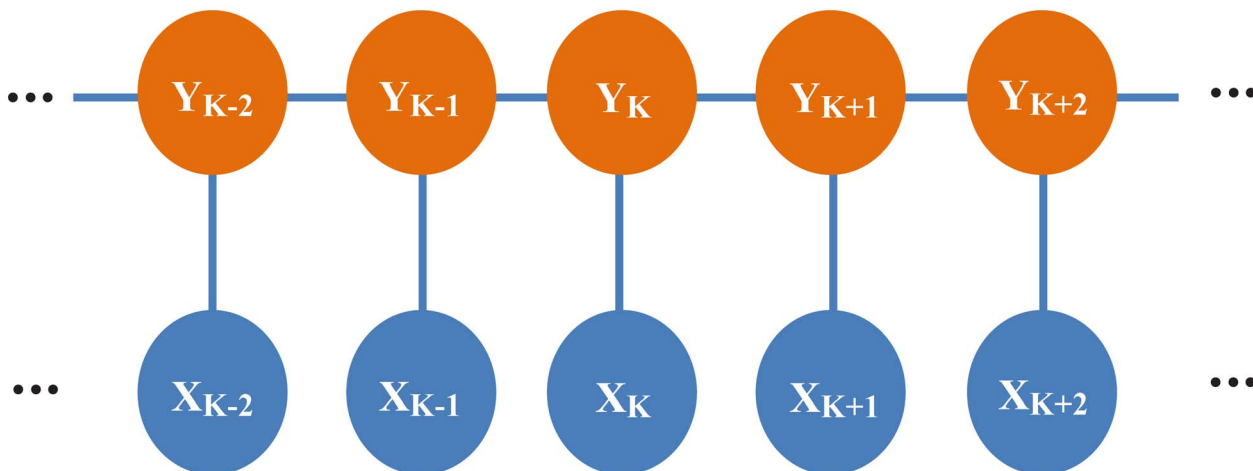
$$\begin{bmatrix} x_{11}, x_{12}, \dots, x_{1n} \\ x_{21}, x_{22}, \dots, x_{2n} \\ \dots \\ x_{m1}, x_{m2}, \dots, x_{mn} \end{bmatrix} = X_{mn} \\ = \mu_{mn} + L_{mk} F_{kn} + \varepsilon_{mn}$$

where  $\mu_{mn}$  is the mean value of each row,  $F_{kn}$  denotes the  $k$  factor, and the  $L_{mk}$  is the loading matrix, which is the coordinates of the  $n$  samples in the new  $k$ -dimensional space. In other words,  $k$  factors are independent and are the underlying biological processes that regulate the phenotypic changes. For example, six factors

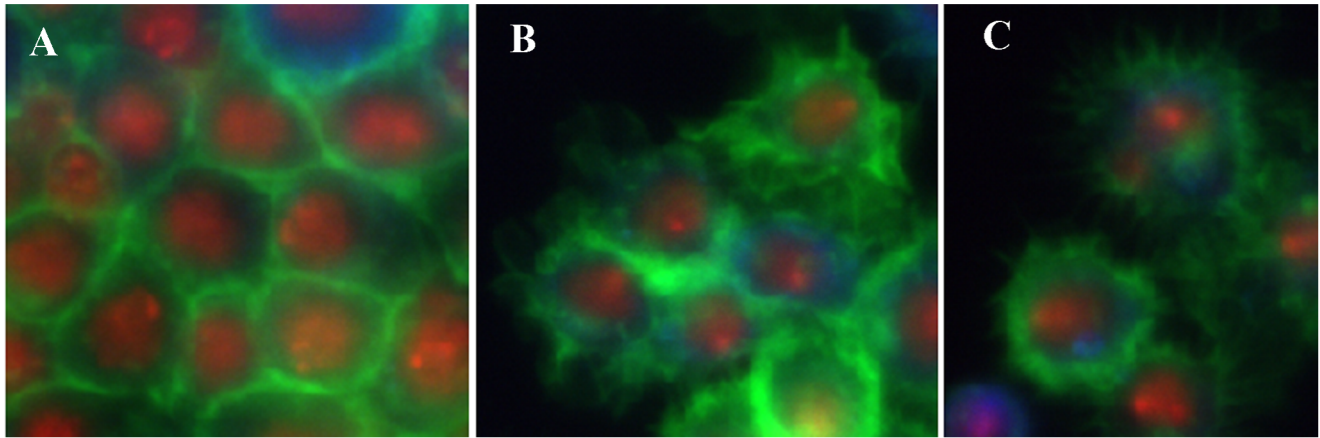
representing nuclei size, DNA replication, chromosomal condensation, nuclei morphology, Edu texture, and nuclei ellipticity, were obtained through factor analysis in [12].

### 5.4 Subpopulation-based Heterogeneity Profiling Analysis

In image-based screening studies, heterogeneous phenotypes often appeared within a cell population, as shown in Figures 2 and 16, which indicated that individual cells responded to perturbations differently [142]. However, the heterogeneity information was ignored in most screening studies. To better make use of the heterogeneous phenotypic responses, a subpopulation based approach was proposed to study the phenotypic heterogeneity for characterizing drug effects in [13], and distinguishing cell populations with distinct drug sensitivities in [14]. The basic principle of the subpopulation based



**Figure 15. The graphical representation of cell cycle phase identification.** doi:10.1371/journal.pcbi.1003043.g015

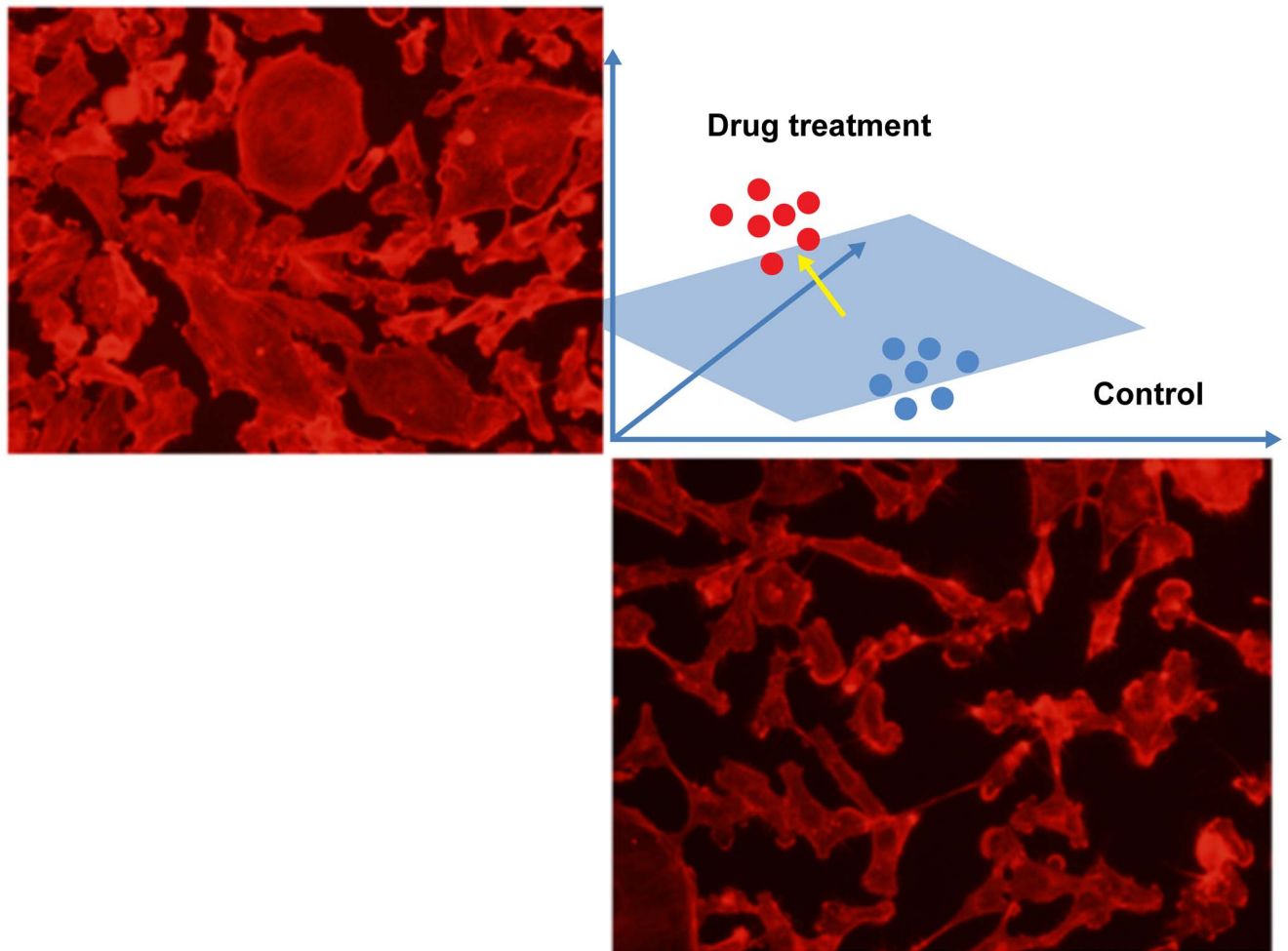


**Figure 16. A representative image of *Drosophila* cells with three phenotypes: (A) Normal, (B) Ruffling and (C) Spiky phenotypes.**  
doi:10.1371/journal.pcbi.1003043.g016

method is to characterize the phenotypic heterogeneity with a mixture of phenotypically distinct subpopulations. This idea

was implemented by fitting a GMM in the numerical space, and each model component of the GMM represents a distinct

subpopulation. To profile the effects of perturbations, cells collected from perturbation conditions were first classified into



**Figure 17. An illustration of drug profiling using the normal vector of hyperplane of SVM.** The red and blue spots indicate the spatial distribution of cells in the numeric feature space. The yellow arrow represents the normal vector of the hyperplane (the blue plane). The top left and bottom right (MB231 cell) images are from drug treated and control conditions respectively.  
doi:10.1371/journal.pcbi.1003043.g017

**Table 1.** List of publicly available bioimage informatics software packages.

Name	Link	Basic Functions
ImageJ	<a href="http://rsb.info.nih.gov/ij/">http://rsb.info.nih.gov/ij/</a>	General image analysis with rich plugins
Fiji (A distribution of ImageJ)	<a href="http://fiji.sc/">http://fiji.sc/</a>	Bioimage analysis with rich plugins
CellProfiler	<a href="http://www.cellprofiler.org/">http://www.cellprofiler.org/</a>	Bioimage analysis with rich analysis pipelines
CellProfiler Analyst	<a href="http://www.cellprofiler.org/">http://www.cellprofiler.org/</a>	Screening data analysis with machine learning approaches
Icy	<a href="http://icy.bioimageanalysis.org/index.php">http://icy.bioimageanalysis.org/index.php</a>	Bioimage analysis
BioimageXD	<a href="http://www.bioimagexd.net/">http://www.bioimagexd.net/</a>	3D Bioimage analysis and Visualization
PhenoRipper	<a href="http://www.phenoripper.org">http://www.phenoripper.org</a>	Bioimage analysis for rapid exploration and interpretation of bioimage data in drug screening
FarSight	<a href="http://www.farsight-toolkit.org/wiki/Main_Page">http://www.farsight-toolkit.org/wiki/Main_Page</a>	Dynamic Biological Microenvironments from 4D/5D Microscopy Data
Vaa3D	<a href="http://penglab.janelia.org/proj/v3d/V3D/About_V3D.html">http://penglab.janelia.org/proj/v3d/V3D/About_V3D.html</a>	Bioimage visualization and analysis
Cell Analyzer	<a href="http://penglab.janelia.org/proj/cellexplorer/cellexplorer/What_is_Cell_Explorer.html">http://penglab.janelia.org/proj/cellexplorer/cellexplorer/What_is_Cell_Explorer.html</a>	<i>C. elegans</i> image analysis
AceTree and StarryNite	<a href="http://starrynite.sourceforge.net/">http://starrynite.sourceforge.net/</a>	<i>C. elegans'</i> embryo cell tracking and lineage reconstruction
Ilastik	<a href="http://www.ilastik.org/">http://www.ilastik.org/</a>	Image classification and segmentation
Image Quantitators (ZFIQ, DCELLIQ, GCELLIQ, NeuriteIQ, NeuronIQ)	<a href="http://www.methodisthealth.com/bbposoftware">http://www.methodisthealth.com/bbposoftware</a>	A set of image analysis software packages for cell tracking in time-lapse images, and RNAi cell, neuron, neurite and Zebrafish image analysis
CellCognition	<a href="http://cellcognition.org/software/cecogalyzer">http://cellcognition.org/software/cecogalyzer</a>	Cell tracking in time-lapse image analysis
TLMTracker	<a href="http://www.tlmtracker.tu-bs.de/index.php/Main_Page">http://www.tlmtracker.tu-bs.de/index.php/Main_Page</a>	Cell tracking in time-lapse image analysis
NeuronJ	<a href="http://www.imagescience.org/meijering/software/neuronj/">http://www.imagescience.org/meijering/software/neuronj/</a>	Neurite Tracing and Quantification
NeurphologyJ	<a href="http://life.nctu.edu.tw/~microtubule/neurphologyJ.html">http://life.nctu.edu.tw/~microtubule/neurphologyJ.html</a>	Neuron image analysis
NeuronStudio	<a href="http://research.mssm.edu/cnic/tools-ns.html">http://research.mssm.edu/cnic/tools-ns.html</a>	Neuron image analysis
CellOrganizer	<a href="http://cellorganizer.org/">http://cellorganizer.org/</a>	Synthetically model and simulate fluorescent microscopic cell images
SimuCell	<a href="http://www.simucell.org">http://www.simucell.org</a>	Synthetically model and simulate fluorescent microscopic cell images
PatternUnmixer	<a href="http://murphylab.web.cmu.edu/software/PatternUnmixer2.0/">http://murphylab.web.cmu.edu/software/PatternUnmixer2.0/</a>	Model fundamental sub-cellular patterns
μManager	<a href="http://valelab.ucsf.edu/~MM/MMwiki/">http://valelab.ucsf.edu/~MM/MMwiki/</a>	Control of automated microscopes
ScanImage	<a href="http://openwiki.janelia.org/wiki/display/ephus/ScanImage%2C+Ephus%2C+and+other+DAQ+software">http://openwiki.janelia.org/wiki/display/ephus/ScanImage%2C+Ephus%2C+and+other+DAQ+software</a>	Control of automated microscopes
OME	<a href="http://www.openmicroscopy.org/site">http://www.openmicroscopy.org/site</a>	Image Database Software
Bisque	<a href="http://www.bioimage.ucsb.edu/bisque">http://www.bioimage.ucsb.edu/bisque</a>	Image Database Software
OMERO.searcher	<a href="http://murphylab.web.cmu.edu/software/searcher/">http://murphylab.web.cmu.edu/software/searcher/</a>	Content-based bioimage search
KNIME	<a href="http://www.knime.org/example-workflows">http://www.knime.org/example-workflows</a>	Workflow system for data analytics, reporting and integration

doi:10.1371/journal.pcbi.1003043.t001

one of the subpopulations, and then the portions of cells belonging to each subpopulation were calculated as features to further characterize the effects of perturbations. For more details, please refer to [13,14].

## 6. Publicly Available Bioimage Informatics Software Packages

A number of commercial bioimage informatics software tools e.g., GE-InCellAnalyzer [143], Cellomics [144], Cellumen [145], MetaXpress [146], BD Pathway [147] have been developed and are widely used in pharmaceutical companies, and academic institutions. In addition to the commercially available software packages,

there are a number of publicly available bioimage informatics software packages [9], which provide even more powerful functions with cutting-edge algorithms and screening-specific analysis pipelines. For the convenience of finding these popular software packages, they are listed in Table 1. It is difficult to summarize all of their capabilities and functions because many of them are designed for flexible bioimage analysis with a set of diverse plugins and function modules, e.g., Fiji, CellProfiler, Icy, and BioimageXD. The software selection for specific applications is also non-trivial, and the best way might be to check their websites and online documents. In addition to the bioimage informatics software packages, there are other software

packages, including the microscope control software for image acquisition (μManager and ScanImage) and image database software (OME, Bisque and OMERO-searcher). Also, certain cellular image simulation software packages, e.g., CellOrganizer and SimuCell, provide useful insights into the organizations of proteins of interest within individual cells. These software packages represent the prevalent directions of bioimage informatics research, thus their websites and features are worth checking.

## 7. Summary

With the advances of fluorescent microscopy and robotic handling, image-



based screening has been widely used for drug and target discovery by systematically investigating morphological changes within cell populations. The bioimage informatics approaches to automatically detect, quantify, and profile the phenotypic changes caused by various perturbations, e.g., drug compounds and RNAi, are essential to the success of these image-based screening studies. In this chapter, an overview of the current bioimage informatics approaches for systematic drug discovery was provided. A number of practical examples were first described to illustrate the concepts and capabilities of image-based screening for drug and target discovery. Then, the prevalent bioimage informatics techniques, e.g., object detection, segmentation, tracking and visualization, were discussed. Subsequently, the widely used numerical features, phenotypes identification, classification, and profiling analysis were introduced to characterize the effects of drugs and targets. Finally, the major publicly available bioimage informatics software packages were listed for future reference. We hope that this review provided sufficient information and insights for readers to apply the approaches and techniques of bioimage informatics to advance their research projects.

## 8. Exercises

**Q1.** Understand the principle of using green fluorescent protein (GFP) to label the chromosome of HeLa cells.

**Q2.** Download a cellular image processing software package, then download some cell images, and use them as examples to perform the cell detection, segmentation, and feature extraction, and provide the analysis results.

**Q3.** Download a time-lapse image analysis software package, then download some time-lapse images, and use them as examples to perform cell tracking, and cell cycle phase classification, and provide the analysis results.

**Q4.** Download a neuron image analysis software package, then download some neuron images, and use them as examples to perform dendrite and spine detection, and provide the analysis results.

**Q5.** Implement the watershed and level set segmentation methods by using ITK functions (<http://www.itk.org/>) and test them on some cell images.

Answers to the Exercises can be found in Text S1.

## Supporting Information

**Text S1** Answers to Exercises. (DOCX)

## Acknowledgments

This paper summarizes over a decade of highly productive collaborations with many colleagues worldwide. The authors would like to acknowledge their collaborators, in particular, Norbert Perrimon, Jeff Lichtman, Bernardo Sabatini, Randy King, Junying Yuan, and Tim Mitchison from Harvard Medical School; Alexei Degterev and Eric Miller from Tufts University; Weiming Xia from Boston VA Medical Center and Boston University, Jun Lu from Stanford University; Chris Bakal from Institute of Cancer Research, Royal Cancer Hospital, U.K.; Yan Feng of Novartis Institutes of Biomedical Research; Shih Fu Chang of Columbia University; Marta Lipinski from the University of Maryland at Baltimore; Jinwen Ma from Peking University of China; Liang Ji from Tsinghua University of China; Myong Hee Kim of EWha Womans University, Korea; Yong Zhang from IBM Research; and Guanglei Xiong from Siemens Corporate Research. The raw image data presented in this paper were mostly generated from the labs of our biological collaborators. We would also like to thank our colleagues at the Department of Systems Medicine and Bioengineering, The Methodist Hospital Research Institute for their discussions, notably Xiaofeng Xia, Kemi Cui, Zhong Xue, and Jie Cheng, as well as former members including Xiaowei Chen, Ranga Srinivasan, Peng Shi, Yue Huang, Gang Li, Xiaobo Zhou, Jingxin Nie, Jun Wang, Tianming Liu, Huiming Peng, Yong Zhang, and Qing Li. We would also like to thank James Mancuso, Derek Cridebring, Luanne Novak, and Rebecca Danforth for proofreading and discussion.

## Further Reading

- Taylor DL (2010) A personal perspective on high-content screening (HCS): from the beginning. *J Biomol Screen* 15(7): 720–725.
- Shariff A, Kangas J, Coelho LP, Quinn S, Murphy RF (2010) Automated image analysis for high-content screening and analysis. *J Biomol Screen* 15(7): 726–734.
- Dufour A, Shinin V, Tajbakhsh S, Guillen-Aghion N, Olivo-Marin JC, et al. (2005) Segmenting and tracking fluorescent cells in dynamic 3-D microscopy with coupled active surfaces. *IEEE Trans Image Process* 14: 1396–1410.
- Danuser G (2011) Computer vision in cell biology. *Cell* 147(5): 973–978.
- Murray JI, Bao Z, Boyle TJ, Boeck ME, Mericle BL, et al. (2008) Automated analysis of embryonic gene expression with cellular resolution in *C. elegans*. *Nat Methods* 5: 703–709.
- Rodriguez A, Ehlenberger DB, Dickstein DL, Hof PR, Wearne SL (2008) Automated three-dimensional detection and shape classification of dendritic spines from fluorescence microscopy images. *PLoS ONE* 3: e1997. doi:10.1371/journal.pone.0001997.
- Bakal C, Aach J, Church G, Perrimon N (2007) Quantitative morphological signatures define local signaling networks regulating cell morphology. *Science* 316(5832): 1753–1756.
- Neumann B, Walter T, Hériché JK, Bulkescher J, Erfle H, et al. (2010) Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature* 464: 721–727.
- Allan C, Burel JM, Moore J, Blackburn C, Linkert M, et al. (2012) OMERO: flexible, model-driven data management for experimental biology. *Nat Methods* 9: 245–253.
- Eliceiri KW, Berthold MR, Goldberg IG, Ibanez L, Manjunath BS, et al. (2012) Biological imaging software tools. *Nat Methods* 9: 697–710.

## Glossary

- **Cellular phenotype:** A cellular phenotype refers to a distinct morphological appearance or behavior of cells as observed under fluorescent, phase contrast, or bright field microscopy.
- **Green fluorescent protein (GFP):** GFP is used as a protein reporter by attaching to specific proteins, and exhibiting bright green fluorescence when exposed to light in the blue to ultraviolet range.
- **Fluorescence microscope:** A fluorescence microscope is an optical microscope that uses higher intensity light source to excite a fluorescent species in a sample of interest.
- **High content analysis (HCA):** HCA focuses on extracting and analyzing quantitative phenotypic data automatically from large amounts of cell images with automated image analysis, computer vision and machine learning approaches.
- **High content screening (HCS):** Applications of HCA for screening drugs and targets are referred to as HCS that aims to identify compounds or genes that cause desired phenotypic changes.
- **RNA interference (RNAi):** RNAi is a biological process, in which RNA molecules inhibit gene expression, typically by causing the destruction of specific mRNA molecules.
- **Automated image analysis:** Automated image analysis aims to quantitatively analyze images automatically by computer programs with minimal human interventions.
- **Object detection:** Object detection is to automatically detect locations of objects of interest in images.
- **Blob structure detection:** Blob structure detection is to detect positions of objects of interest that have circle, sphere like structures, e.g., nuclei and particles.
- **Tube structure detection:** Tube structure is to detect centerlines of objects that have long tube like structures, e.g., neuron dendrite and blood vessel.
- **Object segmentation:** Object segmentation is to automatically delineate boundaries of objects of interest in images.
- **Object tracking:** Object tracking is to identify the motion traces of objects of interest in time-lapse images.
- **Feature extraction:** Feature extraction is to quantify the morphological appearances of segmented objects by calculating a set of numerical features.
- **Phenotype classification:** Phenotype classification is to assign each segmented object into a sub-group that has distinct phenotypes from other sub-groups.
- **Cell cycle phase identification:** Cell cycle phase identification is to automatically identify the corresponding cell cycle phase that a given cell is in according to its morphological appearances.

## References

1. Tsien RY (1998) The green fluorescent protein. *Annu Rev Biochem* 67: 509–544.
2. Lichtman JW, Conchello JA (2005) Fluorescence microscopy. *Nat Methods* 2: 910–919.
3. Shariff A, Kangas J, Coelho LP, Quinn S, Murphy RF (2010) Automated image analysis for high-content screening and analysis. *J Biomol Screen* 15: 726–734.
4. Danuser G (2011) Computer vision in cell biology. *Cell* 147: 973–978.
5. Taylor DL (2010) A personal perspective on high-content screening (HCS): from the beginning. *J Biomol Screen* 15(7): 720–725.
6. Abraham VC, Taylor DL, Haskins JR (2004) High content screening applied to large-scale cell biology. *Trends Biotechnol* 22: 15–22.
7. Giuliano KA, DeBiasio RL, Dunlay RT, Gough A, Volosky JM, et al. (1997) High-content screening: a new approach to easing key bottlenecks in the drug discovery process. *J Biomol Screen* 2: 249–259.
8. Peng H (2008) Bioimage informatics: a new area of engineering biology. *Bioinformatics* 24: 1827–1836.
9. Eliceiri KW, Berthold MR, Goldberg IG, Ibanez L, Manjunath BS, et al. (2012) Biological imaging software tools. *Nat Methods* 9: 697–710.
10. Yarrow JC, Feng Y, Perlman ZE, Kirchhausen T, Mitchison TJ (2003) Phenotypic screening of small molecule libraries by high throughput cell imaging. *Comb Chem High Throughput Screen* 6: 279–286.
11. Perlman Z, Slack M, Feng Y, Mitchison T, Wu L, et al. (2004) Multidimensional drug profiling by automated microscopy. *Science* 306: 1194–1198.
12. Young DW, Bender A, Hoyt J, McWhinnie E, Chirn G-W, et al. (2008) Integrating high-content screening and ligand-target prediction to identify mechanism of action. *Nat Chem Biol* 4: 59–68.
13. Slack MD, Martinez ED, Wu LF, Altschuler SJ (2008) Characterizing heterogeneous cellular responses to perturbations. *Proc Natl Acad Sci U S A* 105: 19306–19311.
14. Singh DK, Ku C-J, Wichaidit C, Steinger RJ, Wu LF, et al. (2010) Patterns of basal signaling heterogeneity can distinguish cellular populations with different drug sensitivities. *Mol Syst Biol* 6: 369.
15. Bakal C, Lindner R, Lense F, Heffern E, Martin-Blanco E, et al. (2008) Phosphorylation networks regulating JNK activity in diverse genetic backgrounds. *Science* 322: 453–456.
16. Bakal C, Aach J, Church G, Perrimon N (2007) Quantitative morphological signatures define local signaling networks regulating cell morphology. *Science* 316: 1753–1756.
17. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, et al. (2006) CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* 7: R100.
18. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, et al. (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9: 676–682.
19. de Chaumont F, Dallongeville S, Chenouard N, Herve N, Pop S, et al. (2012) Icy: an open bioimage informatics platform for extended reproducible research. *Nat Methods* 9: 690–696.
20. Yin Z, Zhou X, Bakal C, Li F, Sun Y, et al. (2008) Using iterative cluster merging with improved gap statistics to perform online phenotype discovery in the context of high-throughput RNAi screens. *BMC Bioinformatics* 9: 264.
21. Rajaram S, Pavie B, Wu LF, Altschuler SJ (2012) PhenoRipper: software for rapidly profiling microscopy images. *Nat Methods* 9: 635–637.
22. Neumann B, Walter T, Heriche JK, Bulkescher J, Erfle H, et al. (2010) Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature* 464: 721–727.
23. Held M, Schmitz MH, Fischer B, Walter T, Neumann B, et al. (2010) CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging. *Nat Methods* 7: 747–754.
24. Shi Q, King RW (2005) Chromosome nondisjunction yields tetraploid rather than aneuploid cells in human cell lines. *Nature* 437: 1038–1042.
25. Sigoillot FD, Huckins JF, Li F, Zhou X, Wong ST, et al. (2011) A time-series method for automated measurement of changes in mitotic and interphase duration from time-lapse movies. *PLoS ONE* 6: e25511. doi:10.1371/journal.pone.0025511
26. Miki T, Lehmann T, Cai H, Stolz DB, Strom SC (2005) Stem cell characteristics of amniotic epithelial cells. *Stem Cells* 23: 1549–1559.
27. Li K, Miller ED, Chen M, Kanade T, Weiss LE, et al. (2008) Cell population tracking and lineage construction with spatiotemporal context. *Med Image Anal* 12: 546–566.
28. Cohen AR, Gomes FL, Roysam B, Cayouette M (2011) Computational prediction of neural progenitor cell fates. *Nat Methods* 7: 213–218.
29. Kankaanpaa P, Paavola L, Tiitta S, Karjalainen M, Paivarinne J, et al. (2012) BioImageXD: an open, general-purpose and high-throughput image-processing platform. *Nat Methods* 9: 683–689.

30. Li F, Zhou X, Ma J, Wong TCS (2010) Multiple Nuclei Tracking Using Integer Programming for Quantitative Cancer Cell Cycle Analysis. *IEEE Trans Med Imaging* 29: 96–105.
31. Klein J, Leupold S, Biegler I, Biedendieck R, Munch R, et al. (2012) TLM-Tracker: software for cell segmentation, tracking and lineage analysis in time-lapse microscopy movies. *Bioinformatics* 28: 2276–2277.
32. Segal M (2005) Dendritic spines and long-term plasticity. *Nat Rev Neurosci* 6: 277–284.
33. Hyman BT (2001) Molecular and anatomical studies in Alzheimer's disease. *Neurologia* 16: 100–104.
34. Ding JB, Takasaki KT, Sabatini BL (2009) Supraresolution imaging in brain slices using stimulated-emission depletion two-photon laser scanning microscopy. *Neuron* 63: 429–437.
35. Nagerl UV, Willig KI, Hein B, Hell SW, Bonhoeffer T (2008) Live-cell imaging of dendritic spines by STED microscopy. *Proc Natl Acad Sci U S A* 105: 18982–18987.
36. Carter AG, Sabatini BL (2004) State-dependent calcium signaling in dendritic spines of striatal medium spiny neurons. *Neuron* 44: 483–493.
37. Duemani Reddy G, Kelleher K, Fink R, Saggau P (2008) Three-dimensional random access multiphoton microscopy for functional imaging of neuronal activity. *Nat Neurosci* 11: 713–720.
38. Iyer V, Hoogland TM, Saggau P (2006) Fast functional imaging of single neurons using random-access multiphoton (RAMP) microscopy. *J Neurophysiol* 95: 535–545.
39. Cheng J, Zhou X, Miller E, Witt RM, Zhu J, et al. (2007) A novel computational approach for automatic dendrite spines detection in two-photon laser scan microscopy. *J Neurosci Methods* 165: 122–134.
40. Ofengeim D, Shi P, Miao B, Fan J, Xia X, et al. (2012) Identification of small molecule inhibitors of neurite loss induced by Abeta peptide using high content screening. *J Biol Chem* 287: 8714–8723.
41. Ho SY, Chao CY, Huang HL, Chiu TW, Charoenkwan P, et al. (2011) NeuropologyJ: an automatic neuronal morphology quantification method and its application in pharmacological discovery. *BMC Bioinformatics* 12: 230.
42. Meijering E, Jacob M, Sarria JC, Steiner P, Hirling H, et al. (2004) Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images. *Cytometry A* 58: 167–176.
43. Pool M, Thiemann J, Bar-Or A, Fournier AE (2008) NeuriteTracer: a novel ImageJ plugin for automated quantification of neurite outgrowth. *J Neurosci Methods* 168: 134–139.
44. Xiong G, Zhou X, Degtrev A, Ji L, Wong STC (2006) Automated neurite labeling and analysis in fluorescence microscopy images. *Cytometry Part A* 69A: 494–505.
45. Narro ML, Yang F, Kraft R, Wenk C, Efrat A, et al. (2007) NeuronMetrics: software for semi-automated processing of cultured neuron images. *Brain Res* 1138: 57–75.
46. Rodriguez A, Ehlenberger DB, Dickstein DL, Hof PR, Wearne SL (2008) Automated three-dimensional detection and shape classification of dendritic spines from fluorescence microscopy images. *PLoS ONE* 3: e1997. doi:10.1371/journal.pone.0001997
47. Wearne SL, Rodriguez A, Ehlenberger DB, Rocher AB, Henderson SC, et al. (2005) New techniques for imaging, digitization and analysis of three-dimensional neural morphology on multiple scales. *Neuroscience* 136: 661–680.
48. Zhang Y, Zhou X, Witt RM, Sabatini BL, Adjeroh D, et al. (2007) Dendritic spine detection using curvilinear structure detector and LDA classifier. *Neuroimage* 36: 346–360.
49. Peng H, Ruan Z, Atasoy D, Sternson S (2010) Automatic reconstruction of 3D neuron structures using a graph-augmented deformable model. *Bioinformatics* 26: i38–i46.
50. Peng H, Long F, Myers G (2011) Automatic 3D neuron tracing using all-path pruning. *Bioinformatics* 27: i239–i247.
51. Meijering E (2010) Neuron tracing in perspective. *Cytometry A* 77: 693–704.
52. Boyle TJ, Bao Z, Murray JI, Araya CL, Waterston RH (2006) AceTree: a tool for visual analysis of *Caenorhabditis elegans* embryogenesis. *BMC Bioinformatics* 7: 275.
53. Bao Z, Murray JI, Boyle T, Ooi SL, Sandel MJ, et al. (2006) Automated cell lineage tracing in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* 103: 2707–2712.
54. Sarov M, Murray JI, Schanze K, Pozniakovski A, Niu W, et al. (2012) A genome-scale resource for in vivo tag-based protein function exploration in *C. elegans*. *Cell* 150: 855–866.
55. Liu X, Long F, Peng H, Aerni SJ, Jiang M, et al. (2009) Analysis of cell fate from single-cell gene expression profiles in *C. elegans*. *Cell* 139: 623–633.
56. Long F, Peng H, Liu X, Kim SK, Myers E (2009) A 3D digital atlas of *C. elegans* and its application to single-cell analyses. *Nat Methods* 6: 667–672.
57. Wahlby C, Kamensky L, Liu ZH, Riklin-Raviv T, Conery AL, et al. (2012) An image analysis toolbox for high-throughput *C. elegans* assays. *Nat Methods* 9: 714–716.
58. Borgfors G (1986) Distance transformations in digital images. *Computer Vision, Graphics, and Image Processing* 34: 344–371.
59. Wahlby C, Sintorn I, Erlandsson F, Borgfors G, Bengtsson E (2004) Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections. *J Microsc* 215: 67–76.
60. Lindeberg T (1993) Detecting salient blob-like image structures and their scales with a scale-space primal sketch: a method for focus-of-attention. *Int J Comput Vision* 11: 283–318.
61. Lindeberg T (1998) Feature detection with automatic scale selection. *Int J Comput Vision* 30: 79–116.
62. Byun J, Verardo MR, Sumengen B, Lewis GP, Manjunath BS, et al. (2006) Automated tool for the detection of cell nuclei in digital microscopic images: application to retinal images. *Mol Vis* 12: 949–960.
63. Al-Kofahi Y, Lassoued W, Lee W, Roysam B (2010) Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Trans Biomed Eng* 57: 841–852.
64. Li G, Liu T, Tarokh A, Nie J, Guo L, et al. (2007) 3D cell nuclei segmentation based on gradient flow tracking. *BMC Cell Biology* 8: 40.
65. Xu C, Prince JL (1998) Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing* 7: 359–369.
66. Duda RO, Hart PE (1972) Use of the Hough transformation to detect lines and curves in pictures. *Commun ACM* 15: 11–15.
67. Parvin B, Yang Q, Han J, Chang H, Rydberg B, et al. (2007) Iterative voting for inference of structural saliency and characterization of sub-cellular events. *IEEE Trans Image Process* 16: 615–623.
68. Lin G, Adiga U, Olson K, Guzowski JF, Barnes CA, et al. (2003) A hybrid 3D watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks. *Cytometry A* 56: 23–36.
69. Lienhart R, Maydt J (2002) An extended set of Haar-like features for rapid object detection. pp. I-900–I-903. Vol. 1. Proceedings of the 2002 International Conference on Image Processing.
70. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. I-511–I-518. Vol. 1.
71. He W, Wang X, Metaxas D, Mathew R, White E (2007) Cell segmentation for division rate estimation in computerized video time-lapse microscopy. 643109–643109.
72. Jiang S, Zhou X, Kirchhausen T, Wong ST (2007) Detection of molecular particles in live cells via machine learning. *Cytometry A* 71: 563–575.
73. Sommer C, Strachle C, Kothe U, Hamprecht FA (2011) Ilastik: interactive learning and segmentation toolkit. pp. 230–233. 2011 IEEE International Symposium on Biomedical Imaging; 30 March–2 April 2011.
74. Steger C (1998) An unbiased detector of curvilinear structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20: 113–125.
75. Al-Kofahi KA, Lasek S, Szarowski DH, Pace CJ, Nagy G, et al. (2002) Rapid automated three-dimensional tracing of neurons from confocal image stacks. *IEEE Transactions on Information Technology in Biomedicine* 6: 171–187.
76. Tyrrell JA, di Tomaso E, Fuja D, Ricky T, Kozak K, et al. (2007) Robust 3-D modeling of vasculature imagery using superellipsoids. *IEEE Trans Med Imaging* 26: 223–237.
77. Soares JVB, Leandro JGG, Cesar RM, Jelinek HF, Cree MJ (2006) Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. *IEEE Trans Med Imaging* 25: 1214–1222.
78. Staal J, Abramoff MD, Niemeijer M, Viergever MA, van Ginneken B (2004) Ridge-based vessel segmentation in color images of the retina. *IEEE Trans Med Imaging* 23: 501–509.
79. Fraz M, Remagnino P, Hoppe A, Uyyanonvara B, Rudnicka A, et al. (2012) Blood vessel segmentation methodologies in retinal images - a survey. *Comput Methods Programs Biomed* 108(1): 407–433.
80. Otsu N (1978) A threshold selection method from gray level histogram. *IEEE Transactions on System, Man, and Cybernetics* 8: 62–66.
81. Dunn JC (1973) A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics* 3: 32–57.
82. Vincent L, Soille P (1991) Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13: 583–598.
83. Beucher S (1992) The watershed transformation applied to image segmentation. *Scanning Microscopy International* 6: 299–314.
84. Meyer F, Beucher S (1990) Morphological segmentation. *Journal of Visual Communication and Image Representation* 1: 21–46.
85. Wahlby C, Lindblad J, Vondrus M, Bengtsson E, Björkstén L (2002) Algorithms for cytoplasm segmentation of fluorescence labelled cells. *Analytical Cellular Pathology* 24: 101–111.
86. Casselles V, Kimmel R, Sapiro G (1997) Geodesic active contours. *International Journal of Computer Vision* 22: 61–79.
87. Chan T, Vese L (2001) Active contours without edges. *IEEE Transactions on Image Processing* 10: 266–277.
88. Zimmer C, Labryère E, Meas-Yedid V, Guillén N, Olivo-Marin J (2002) Segmentation and tracking of migrating cells in videomicroscopy with parametric active contours: a tool for cell-based drug testing. *IEEE Trans Med Imaging* 21: 1212–1221.
89. Yan P, Zhou X, Shah M, Wong ST (2008) Automatic segmentation of high-throughput RNAi fluorescent cellular images. *IEEE Trans Inf Technol Biomed* 12: 109–117.
90. Dufour A, Shinin V, Tajbakhsh S, Guillen-Aghion N, Olivo-Marin JC, et al. (2005) Segmenting and tracking fluorescent cells in dynamic 3-D microscopy with coupled active surfaces. *IEEE Transactions on Image Processing* 14: 1396–1410.
91. Caselles V, Kimmel R, Sapiro G (1997) Geodesic active contours. *International Journal of Computer Vision* 22: 61–79.

92. Osher S, Sethian JA (1988) Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of Computation Physics* 79: 12–49.
93. Chunming L, Chenyang X, Changfeng G, Fox MD (2005) Level set evolution without re-initialization: a new variational formulation. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*; 20–25 June 2005. pp. 430–436. Vol. 1.
94. Aurenhammer F (1991) Voronoi diagrams - a survey of a fundamental geometric data structure. *ACM Comput Surv* 23: 345–405.
95. Jones T, Carpener A, Golland P (2005) Voronoi-based segmentation of cells on image manifolds. *Lecture Notes in Computer Science*: 535–543.
96. Shi J, Malik J (2000) Normalized cuts and image segmentation. *IEEE Trans Pattern Anal Mach Intell* 22: 888–905.
97. Felzenszwalb PF, Huttenlocher DP (2004) Efficient graph-based image segmentation. *Int J Comput Vision* 59: 167–181.
98. Radhakrishna A, Shaji A, Smith K, Lucchi A, Fua P, et al. (June, 2010) SLIC superpixels. Technical report 149300, EPFL.
99. Lucchi A, Smith K, Achanta R, Lepetit V, Fua P (2010) A fully automated approach to segmentation of irregularly shaped cellular structures in EM images. *Proceedings of the 13th International Conference on Medical Image Computing and Computer-Assisted Intervention. Med Image Comput Comput Assist Interv* 13(Pt 2): 463–471.
100. Debeir O, Ham PV, Kiss R, Decaestecker C (2005) Tracking of migrating cells under phase-contrast video microscopy with combined mean-shift processes. *IEEE Trans Med Imaging* 24: 697–711.
101. Zimmer C, Olivo-Marin JC (2005) Coupled parametric active contours. *IEEE Trans Pattern Anal Mach Intell* 27: 1838–1842.
102. Yang F, Mackey MA, Ianzini F, Gallardo G, Sonka M (2005) Cell segmentation, tracking, and mitosis detection using temporal context. *Med Image Comput Comput Assist Interv* 8(Pt 1): 302–309.
103. Buniyak F, Palaniappan K, Nath SK, Baskin TL, Gang D (2006) Quantitative cell motility for in vitro wound healing using level set-based active contour tracking. *Proc IEEE Int Symp Biomed Imaging* 2006 April 6: 1040–1043.
104. Dzyubachyk O, van Cappellen WA, Essers J, Niessen WJ, Meijering E (2010) Advanced level-set-based cell tracking in time-lapse fluorescence microscopy. *IEEE Trans Med Imaging* 29: 852–867.
105. Bo Z, Zimmer C, Olivo-Marin JC (2004) Tracking fluorescent cells with coupled geometric active contours. *IEEE International Symposium on Biomedical Imaging*; 15–18 April 2004. pp. 476–479. Vol. 1.
106. Li K, Miller ED, Weiss LE, Campbell PG, Kanade T (2006) Online tracking of migrating and proliferating cells imaged with phase-contrast microscopy. *Conference on Computer Vision and Pattern Recognition Workshop*. New York City, New York. pp. 65.
107. Nath SK, Palaniappan K, Buniyak F (2006) Cell segmentation using coupled level sets and graph-vertex coloring. *Proceedings of the 9th International Conference on Medical Image Computing and Computer-Assisted Intervention. Med Image Comput Comput Assist Interv* 9 (Pt 1): 101–108.
108. Appel K, Haken W (1977) Every planar map is four colorable part I. Discharging. *Illinois Journal of Mathematics*: 429–490.
109. Appel K, Haken W, Koch J (1977) Every planar map is four colorable part II. Reducibility. *Illinois Journal of Mathematics*: 491–567.
110. Padfield DR, Rittscher J, Sebastian T, Thomas N, Roysam B (2006) Spatio-temporal cell cycle analysis using 3D level set segmentation of unstained nuclei in line scan confocal fluorescence images. *3rd IEEE International Symposium on Biomedical Imaging*; 6–9 April 2006. pp. 1036–1039.
111. Padfield DR, Rittscher J, Roysam B (2008) Spatio-temporal cell segmentation and tracking for automated screening. *5th IEEE International Symposium on Biomedical Imaging*; 14–17 May 2008. pp. 376–379.
112. Padfield D, Rittscher J, Thomas N, Roysam B (2009) Spatio-temporal cell cycle phase analysis using level sets and fast marching methods. *Medical Image Analysis* 13: 143–155.
113. Al-Kofahi O, Radke RJ, Goderie SK, Shen Q, Temple S, et al. (2006) Automated cell lineage construction: a rapid method to analyze clonal development established with murine neural progenitor cells. *Cell Cycle* 5: 327–335.
114. Chen X, Zhou X, Wong STC (2006) Automated segmentation, classification, and tracking of cancer cell nuclei in time-lapse microscopy. *IEEE Transactions on Biomedical Engineering* 53: 762–766.
115. Harder N, Mora-Bermudez F, Godinez WJ, Ellenberg J, Eils R, et al. (2006) Automated analysis of the mitotic phases of human cells in 3D fluorescence microscopy image sequences. *Med Image Comput Comput Assist Interv* 9: 840–848.
116. Li K, Chen M, Kanade T (2007) Cell population tracking and lineage construction with spatiotemporal context. *Med Image Comput Assist Interv* 10: 295–302.
117. Blom HAP (1984) An efficient filter for abruptly changing systems. *Proceedings of 23rd IEEE Conference on Decision and Control* 23: 656–658.
118. Genovesio A, Liedl T, Emiliani V, Parak WJ, Coppey-Moisand M, et al. (2006) Multiple particle tracking in 3-D+t microscopy: method and application to the tracking of endocytosed quantum dots. *IEEE Trans Image Process* 15: 1062–1070.
119. Smal I, Draegestein K, Galjart N, Niessen W, Meijering E (2007) Rao-blackwellized marginal particle filtering for multiple object tracking in molecular bioimaging. *Proceedings of the 20th International Conference on Information Processing in Medical Imaging. Kerkrade, The Netherlands: Springer-Verlag*.
120. Smal I, Niessen W, Meijering E (2006) Bayesian tracking for fluorescence microscopic imaging; 6–9 April 2006. pp. 550–553.
121. Godinez WJ, Lampe M, Worz S, Muller B, Eils R, et al. (2007) Tracking of virus particles in time-lapse fluorescence microscopy image sequences. 12–15 April 2007. pp. 256–259.
122. Luisi J, Narayanaswamy A, Galbreath Z, Roysam B (2011) The FARSIGHT trace editor: an open source tool for 3-D inspection and efficient pattern analysis aided editing of automated neuronal reconstructions. *Neuroinformatics* 9: 305–315.
123. Peng H, Ruan Z, Long F, Simpson JH, Myers EW (2010) V3D enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets. *Nat Biotechnol* 28: 348–353.
124. Manjunath B, Ma W (1996) Texture features for browsing and retrieval of image data. *IEEE Trans Pattern Anal Mach Intell* 18: 837–842.
125. Zhou X, Wong STC (2006) Informatics challenges of high-throughput microscopy. *IEEE Signal Processing Magazine* 23: 63–72.
126. Chen X, Zhou X, Wong ST (2006) Automated segmentation, classification, and tracking of cancer cell nuclei in time-lapse microscopy. *IEEE Trans Biomed Eng* 53: 762–766.
127. Boland M, Murphy R (2001) A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells. *Bioinformatics* 17: 1213–1223.
128. Haralick R (1979) Statistical and structural approaches to texture. *Proceedings of IEEE* 67: 786–804.
129. Manjunatha BS, Ma WY (1996) Texture features for browsing and retrieval of image data. *IEEE Trans Pattern Anal Mach Intell* 18: 837–842.
130. Cohen A, Daubechies I, Feauveau JC (1992) Bi-orthogonal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics* 45: 485–560.
131. Zernike F (1934) Beugungstheorie des schneidenerfarhens unseiner verbesserten form, der phasenkontrastmethode. *Physica* 1: 689–704.
132. Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics* 6: 610–620.
133. Harder N, Mora-Bermudez F, Godinez WJ, Wunsche A, Eils R, et al. (2009) Automatic analysis of dividing cells in live cell movies to detect mitotic delays and correlate phenotypes in time. *Genome Res* 19: 2113–2124.
134. Zhong Q, Busetto AG, Fededa JP, Buhmann JM, Gerlich DW (2012) Unsupervised modeling of cell morphology dynamics for time-lapse microscopy. *Nat Methods* 9: 711–713.
135. Wang M, Zhou X, Li F, Huckins J, King WR, et al. (2008) Novel cell segmentation and online SVM for cell cycle phase identification in automated microscopy. *Bioinformatics* 24: 94–101.
136. Wang J, Zhou X, Bradley PL, Chang SF, Perrimon N, et al. (2008) Cellular phenotype recognition for high-content RNA interference genome-wide screening. *J Biomol Screen* 13: 29–39.
137. Yan M, Ye K (2007) Determining the number of clusters using the weighted gap statistic. *Biometrics* 63: 1031–1037.
138. Jones TR, Carpenter AE, Lamprecht MR, Moffat J, Silver SJ, et al. (2009) Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. *Proc Natl Acad Sci U S A* 106: 1826–1831.
139. Young DW, Bender A, Hoyt J, McWhinnie E, Chirm GW, et al. (2008) Integrating high-content screening and ligand-target prediction to identify mechanism of action. *Nat Chem Biol* 4: 59–68.
140. Frise E, Hammonds AS, Celniker SE Systematic image-driven analysis of the spatial *Drosophila* embryonic expression landscape. *Mol Syst Biol* 6: 345.
141. Loo LH, Wu LF, Altschuler SJ (2007) Image-based multivariate profiling of drug responses from single cells. *Nat Methods* 4: 445–453.
142. Altschuler SJ, Wu LF (2010) Cellular heterogeneity: do differences make a difference? *Cell* 141: 559–563.
143. GE-InCellAnalyzer. <http://www.biocore.com/high-content-analysis/index.html>.
144. Cellomics. [http://www.cellomics.com/content/menu/About\\_Us/](http://www.cellomics.com/content/menu/About_Us/).
145. Cellumen. <http://www.cellumen.com/>.
146. MetaXpress. <http://www.moleculardevices.com/pages/software/metaxpress.html>.
147. BD-Pathway. [http://www.bdbiosciences.ca/bioimaging/cell\\_biology/pathway/software/](http://www.bdbiosciences.ca/bioimaging/cell_biology/pathway/software/).