

CellProfiler™: free, versatile software for automated biological image analysis

Michael R. Lamprecht¹, David M. Sabatini^{1,2}, and Anne E. Carpenter¹

¹Whitehead Institute for Biomedical Research, Cambridge, and ²Massachusetts Institute of Technology, Cambridge, MA, USA

BioTechniques 42:71-75 (January 2007)
doi 10.2144/000112257

Careful visual examination of biological samples is quite powerful, but many visual analysis tasks done in the laboratory are repetitive, tedious, and subjective. Here we describe the use of the open-source software, CellProfiler™, to automatically identify and measure a variety of biological objects in images. The applications demonstrated here include yeast colony counting and classifying, cell microarray annotation, yeast patch assays, mouse tumor quantification, wound healing assays, and tissue topology measurement. The software automatically identifies objects in digital images, counts them, and records a full spectrum of measurements for each object, including location within the image, size, shape, color intensity, degree of correlation between colors, texture (smoothness), and number of neighbors. Small numbers of images can be processed automatically on a personal computer, and hundreds of thousands can be analyzed using a computing cluster. This free, easy-to-use software enables biologists to comprehensively and quantitatively address many questions that previously would have required custom programming, thereby facilitating discovery in a variety of biological fields of study.

INTRODUCTION

One of the most powerful methods in biology is the visual analysis of a sample. While nothing can fully replace the expertise of a trained biologist, observing many samples by eye is time-consuming, subjective, and nonquantitative. Certain repetitive tasks in visual analysis are suitable for automation by collecting digital images and processing them with image analysis software. This liberates biologists for more interesting work and has several advantages over visual observations including speed, quantitative and reproducible results, and simultaneous measurement of many features in the image. Efforts to automate visual analysis in biology began several decades ago, but many aspects still need improvement (1).

While numerous commercial and free software packages exist for image analysis, many of these packages are designed for a very specific purpose, such as cell counting (2). Other packages are sold with accompanying hardware for image acquisition (e.g., yeast colony counters), but

these are expensive and do not allow measurement of features beyond those that are already built-in. Most commercial software is proprietary, meaning that the underlying methods of analysis are hidden from the researcher. At the other end of the continuum, some software packages are very flexible, especially for interactive analysis of individual images [e.g., Image-Pro Plus, MetaMorph®, and the open-source ImageJ/National Institutes of Health (NIH) Image (3)]. While users can program custom algorithms or record macros, these customized routines are challenging to adapt without knowing a programming language or interacting directly with the macro code.

The CellProfiler™ project was developed to address these software challenges by providing the scientific community with an easy-to-use open-source platform for automated image analysis. The compiled software is freely available for Macintosh®, PC, and Unix platforms at www.cellprofiler.org. It can accommodate adaptation to many biological objects and assays without requiring programming, due

to its modular design and graphical user interface. There are many existing software packages available for specific applications in biology, but CellProfiler accomplishes many of the same goals in one open-source program. We recently described CellProfiler's use for cell identification, cell size, intensity and texture of fluorescent stains, cell cycle distributions, and other features of individual cells in images (4). Here we describe its use for a variety of other applications such as yeast colony counting, grid analysis, wound healing, and other visually quantifiable assays.

MATERIALS AND METHODS

All of the image analysis in this paper used the freely available CellProfiler cell image analysis software. The pipelines and images for these examples, as well as others, are available for download (www.cellprofiler.org/examples.htm). The image of yeast colonies (Figure 1) is a plate of Hi90-strain cells plated on synthetic defined medium with 128 µg/mL fluconazole as previously described (5). Images of *Drosophila* Kc167 cells on cell microarrays (Figure 2, A–C) were prepared as described previously (4,6). Briefly, spots of double-stranded RNA (dsRNA) were printed onto plain slides, and cells were grown on these slides for 3 days before being fixed, stained with Hoechst 33342, and imaged. Images of yeast patches (Figure 2, D–G) were prepared by manually spotting cells (with a 96-well pinning device) onto agar plates containing galactose to induce the expression of α-synuclein and a gene of interest. The cells were grown for 2 days at 30°C prior to imaging (7). Images of green fluorescent protein (GFP)-labeled mouse tumors (Figure 3, A–C) are faces of a mouse lung lobe, dissected out at 8 weeks post-tail vein injection of an established metastatic human cancer cell line overexpressing a gene of interest as described (Kimberly Hartwell, unpublished data). Images of wound healing (Figure 3, E–G) were prepared using MDA-MB-435 cells and imaged at the time points indicated (Lynne Waldman, personal communication). A *Drosophila* wing imaginal disc from a third larval instar (Figure 3, H–J) was stained with

rhodamine-phalloidin to label F-actin, which is concentrated at points of cell-cell contact at the level of the adherens junction lattice (8).

RESULTS AND DISCUSSION

CellProfiler's main window allows the user to point and click to do most tasks, including the design of a new assay. The software uses the concept of a pipeline, which is a series of modules. Each module performs a specific task on the image or on identified objects (Figure 1A). A typical pipeline consists of loading the images, correcting for uneven illumination, identifying the objects, and then taking measurements on those objects. These modules can easily be added, removed, or rearranged within a pipeline. The resulting measurements can be viewed by (i) using CellProfiler's built-in viewing and plotting data tools; (ii) exporting in a tab-delimited spreadsheet format that can be opened in programs like Microsoft® Excel® and OpenOffice.org Calc; (iii) exporting in a format that can be imported into a database like Oracle® or MySQL® (MySQL, Cupertino, CA, USA); or (iv) opening in MATLAB® (Mathworks, Natick, MA, USA). An analysis can be done on one specific image, a group of images, or thousands of images by using a computing cluster.

CellProfiler bridges the gap between powerful computational methods and their practical application in

the biological laboratory. Computer scientists can prototype new computational methods and contribute them to the project, and then biologists can easily use these new additions in their work. Further, the functionality of existing modules can be enhanced by researchers with some programming experience, because the code is open-source, well-documented, and in a language that is relatively easy to understand. While most users will download the completely free, compiled version of CellProfiler, the CellProfiler Developer's version requires the software package MATLAB and its image processing toolbox.

As described in the manual, available at www.cellprofiler.org/linked_files/CellProfilerManual.pdf, CellProfiler already contains advanced object identification algorithms from the literature (4,9–16) and is open to adding new algorithms as described above. In object identification modules, users can rapidly select the best solution for their application using a Test Mode to see the results of various methods. In the following examples, we show the identification of objects by CellProfiler and select measurements for each. Note that the full spectrum of measurements, including many not often measured by biologists (17–19), can be recorded for each identified object, including location within the image, size, shape, color intensity, texture (smoothness), correlation between colors, and number of neighbors. Moreover, each broad category contains many different

specific measurements. For example, size includes area, perimeter, and major/minor axis length, and shape includes eccentricity (elongation), solidity, form factor, and 32 other shape-related measures.

Yeast Colonies

Counting colonies on agar plates and classifying them by size, color, or shape is tedious, time-consuming, and subjective. While complete systems for automated colony counting exist, they are more expensive and less flexible than using a digital camera or off-the-shelf flatbed scanner to acquire images for analysis by CellProfiler. The cost of this solution can be less than \$100. Furthermore, the algorithms in CellProfiler are accurate and adaptable, and unusual features of colonies, which commercial software and even the human eye cannot detect, can be measured (e.g., certain measures of texture and shape). After the initial analysis strategy has been established, plates can be analyzed automatically in large batches.

Here we show an example of yeast colonies (*Saccharomyces cerevisiae*) that were analyzed by CellProfiler (Figure 1B). In this analysis, the plates are automatically cropped to remove the edges, and individual colonies are identified, even when clumped (Figure 1C). Measurement modules then calculate measurements of interest for each individual colony. Any of the available measurements can then

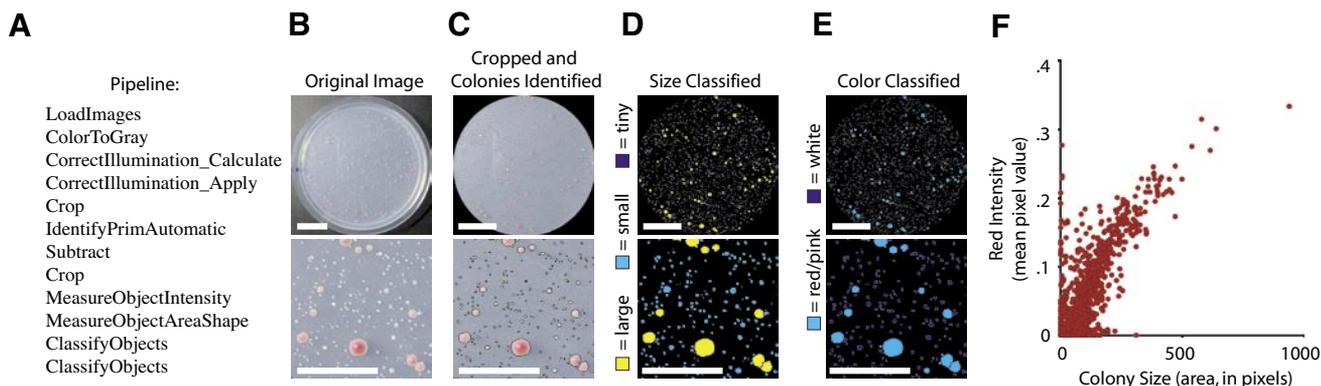


Figure 1. Yeast colonies growing on plates can be identified, measured, and classified. Scale bars, 2 cm (top) and 1 cm (bottom). (A) The pipeline of modules used for this analysis. (B) Original image of yeast colonies growing on an agar plate. (C) Image after automatic cropping of the plate and colony identification by CellProfiler, with individual colonies outlined in black. (D) Image with the identified colonies classified by size (see legend for color-coding). (E) Same as panel D, but classified by red intensity rather than size. (F) The red intensity (vertical axis) and size (horizontal axis) of each colony on the plate is plotted, revealing the relationship between these measurements.

be used to classify the colonies, for example, by size (Figure 1D), by color (Figure 1E), or by a combination of measurements, such as size and color. In this example, the apparent correlation between size-classified (Figure 1D) and color-classified (Figure 1E) yeast colonies is verified by a scatter plot of these two measurements (Figure 1F). Each class of colonies can be analyzed separately to allow the researcher to focus on classes of interest. This allows for addressing questions like “Are the red colonies larger than white colonies?” or “Do the larger colonies have more irregularly-shaped borders?” In this example, the colonies all display a smooth round phenotype, but the colony shape and texture of yeast strains with unusual morphology could be quantified using these methods.

Grid Analysis

Many experiments are designed in a grid format, such as cell microarrays, agar plates of yeast patches, and multiwell plates. In experiments with large or numerous grids, it is difficult to identify which reagent corresponds to a spot on this grid. CellProfiler can manually or automatically place a grid on an image and associate each grid location with an annotation such as a sample number or gene name. Each grid location can also be identified as an object, and all available measurements can be made on those objects.

For example, on a cell microarray, more than 5000 individual clusters of cells are grown on a single microscope slide (4,6,20–22). Each cluster has been treated with a perturbant, which could be a small molecule, an overexpression plasmid, or an RNA interference reagent. The goal is to determine which perturbants alter cells. When attempting to quantify even a simple phenotype such as the extent of cell death in a spot, it is difficult to determine which spot correlates to which reagent. In this example, CellProfiler places a grid on top of an image (Figure 2B) in a position defined by the user, who specifies the location of known control spots on the array and the number and size of rows and columns. The LoadText module allows the user to load a text file with corresponding sample information for each of the spots on the grid, and the DisplayGridInfo module allows this imported data to be assigned to each of the grid locations (Figure 2C). This quickly allows the user to associate a location in the grid with its reagent. For example, knockdown of the cytokinesis-related gene *CG10522* (*sticky*) shows an unusual large, bright-nuclei phenotype that is visible at low resolution (Figure 2C).

Plates of yeast patches also make use of grids (Figure 2E). Although large screens of yeast patches have been analyzed by eye (23), the number of these screens in progress is rapidly increasing, and visual analysis cannot keep up with the rapid pace at

which samples are being generated. Furthermore, quantitative analysis is much preferred, because subtle changes in growth can be identified and the screen can be analyzed statistically. Some software for this application has been developed, but none to our knowledge is fully automated, open-source, and flexible to new assays/unusual measurements.

Because thousands of plates are typically analyzed, the entire process of finding the grid and making measurements is performed automatically by CellProfiler. In the pipeline for this analysis, the images are cropped to remove the plate edges, and any yeast patches that are present are identified. Unlike the cell microarray example, a grid is then defined automatically based on the yeast patches that are identified. This allows for nonuniformity in the precise placement of the grid on the plate, to allow for experimental variation. For this function to work correctly, none of the outside rows or columns can be completely blank. This condition can be satisfied if most patches tend to grow well in the experiment or if control patches exist in two or more opposite corners. These yeast patches can be analyzed in their naturally identified shapes if the patch size or shape is of interest (Figure 2F). Alternately, a circle can be forced into the location of the identified objects to measure, for example, the intensity of each patch, which is a measure of growth (Figure 2G). Once the grid

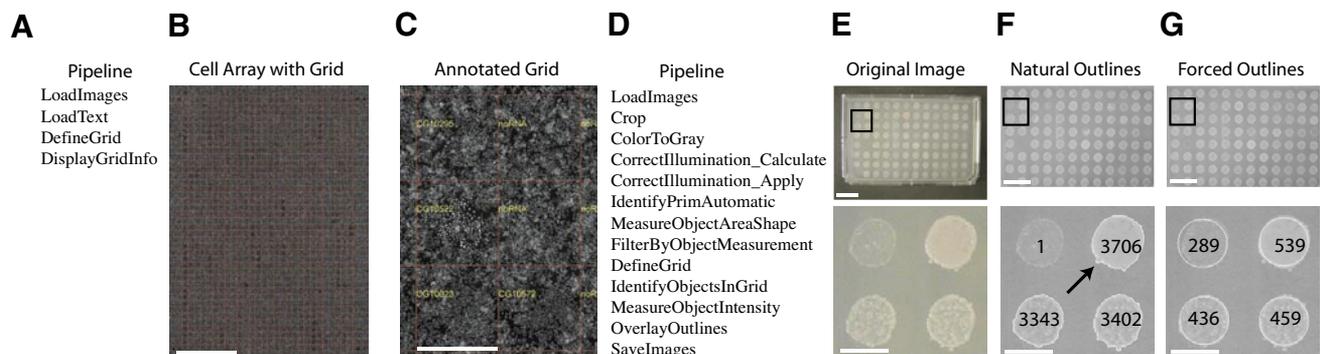


Figure 2. Grids of samples can be annotated and analyzed. (A) The pipeline of modules used for the analysis shown in panels B and C. (B) A living cell RNA interference microarray with 40 rows and 28 columns of spots, stained for DNA. Each spot contains a double-stranded RNA (dsRNA) that knocks down a particular gene. The grid placed on the image by CellProfiler is shown as red lines. noRNA is the control. Scale bar, 450 μ m. (C) Enlarged portion of panel B, with the annotations placed by CellProfiler shown in yellow. (D) The pipeline of modules used for the analysis shown in panels E–G. (E) Original image of yeast patches growing in a grid with 8 rows and 12 columns. Box indicates the region shown enlarged below. (F) Image showing the patches’ natural outlines determined by CellProfiler, including wiggly protrusions (arrow). The measured area of each patch is shown numerically on top of each patch, in pixels. (G) Image showing outlines of patches that were forced into a standard circular shape to measure the amount of growth in each patch, using intensity units. Scale bars for panel E–G: top, 20 mm; bottom, 5 mm.

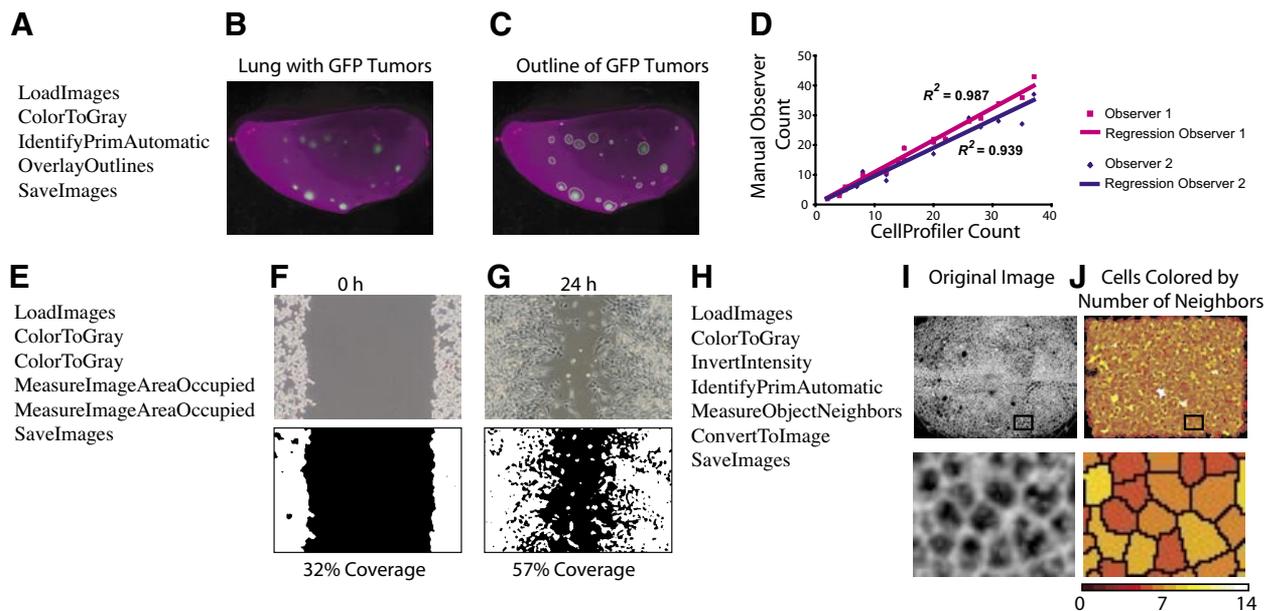


Figure 3. Identification and measurement of green fluorescent protein (GFP)-labeled tumors in mouse lungs, the wound healing assay, and *Drosophila* tissue topology. (A) The pipeline of modules used for the analysis shown in panels B–D. (B) Original image. (C) Image with tumors outlined by CellProfiler. (D) Tumors in a set of 20 images were counted by CellProfiler and by two researchers. The manual tumor count for each image (vertical axis) is plotted versus the CellProfiler count (horizontal axis), revealing strong concordance (R^2 value is shown). (E) The pipeline of modules used for the analysis shown in panels F and G. (F) At time point zero, the wound visible in the original image (top) is large and the cells present at the edges of the image cover a small percentage of the area of the image, as quantified by CellProfiler (bottom). (G) After 24 h, the wound has recovered due to cells migrating from the edges (top) and now is much smaller in measured size (bottom). (H) The pipeline of modules used for the analysis shown in panels I and J. (I) Original image of *Drosophila* epithelial cells growing in a sheet. Box indicates the region shown enlarged below. (J) Image showing cells identified and color-coded by CellProfiler based on how many neighbors they have. Box indicates the region shown enlarged below. Color scale indicates the number of neighbors.

and objects are identified, all available measurements can be calculated for each patch, including measures of growth (Figure 2, F–G, bottom).

Tumor Counting

When tumorigenic cells are labeled with GFP and injected into mice, the resulting tumors in the lungs are readily visible by fluorescence microscopy of the dissected lungs. Accurate, objective quantification of the number and size of the resulting tumors is necessary to understand the process of tumor metastasis (Kimberly Hartwell, unpublished data). The GFP signal from each tumor can be identified by CellProfiler (Figure 3, A–C). CellProfiler counts of identified tumors were comparable to counting by eye (Figure 3D). If the GFP brightness or the shape or texture/smoothness of the tumors is of interest, these measurements can also be recorded.

Wound Healing

The wound healing assay is a standard technique to determine the

migration of different cell types in different conditions. In this assay, a confluent monolayer of cells is wounded by scratching it with a pipet tip (24). The monolayer is then imaged at time points to record the size of the wound. In this example, the area of the images covered by cells is calculated by CellProfiler (Figure 3, E–G). While this is not a particularly challenging application, the structure of CellProfiler makes it simple to carry out this quantitative analysis for hundreds of thousands of images, enabling high-throughput screens. In addition, the shape characteristics of the wound border can be measured; for example, to distinguish between samples where all cells have steadily grown toward the middle versus samples where a few individual cells extend into the wound space.

Tissue Topology

In a developing tissue or at other sites of cell-cell contact (e.g., tumors and surrounding stromal cells), it is useful to determine the number

of neighbors each cell has to better understand the processes underlying the topology (25). CellProfiler can identify cells in tissues (Figure 3, H–J). In addition to typical measurements, the MeasureObjectNeighbors module can determine the number of cells neighboring each cell and record this measurement. The cells can then be color-coded by how many neighbors it has (Figure 3J), or the data can be exported to further analyze the topology of the tissue.

Summary

CellProfiler is a flexible platform that can automate the analysis of images to address a wide variety of biological questions. For many assays, described here and previously (4), it eliminates the tedium of repetitive visual analysis and produces rapid, quantitative, and accurate results. The modular design of the software provides an infrastructure for image analysis that is applicable to diverse assays. Its open-source code allows programmers to design and contribute new algorithms to the

project. It is our hope that CellProfiler will become a widely used platform, through which advanced algorithms are made conveniently available for automatic biological image analysis.

ACKNOWLEDGMENTS

We gratefully acknowledge contributions of images from researchers at the Harvard Medical School: Matt Gibson (Figure 3, H–J) and at the Whitehead Institute for Biomedical Research and Massachusetts Institute of Technology: Leah Cowen (Figure 1), Douglas B. Wheeler (Figure 2, A–C), Aaron Gitler (Figure 2, D–G), Kimberly A. Hartwell (Figure 3, A–D), and Lynne K. Waldman (Figure 3, E–G). We also thank the other members of the CellProfiler project software team for software development and lab members for helpful comments: Robert A. Lindquist, Shomit Sengupta, and David A. Guertin. This work was supported by a Merck/CSBi postdoctoral fellowship (A.E.C.), a Novartis fellowship from the Life Sciences Research Foundation (A.E.C.), a Society for Biomolecular Screening Academic grant (A.E.C.), Department of Defense (DOD) Tuberous Sclerosis Complex (TSC) research program grant no. W81XWH-05-1-0318-DS (D.M.S.), NIH grant no. R01 GM072555-01 (D.M.S.), and the Keck Foundation (D.M.S.).

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

REFERENCES

- Murphy, R.F., E. Meijering, and G. Danuser. 2005. Special issue on molecular and cellular bioimaging. *IEEE Trans. Image Process.* 14:1233-1236.
- Selinummi, J., J. Seppala, O. Yli-Harja, and J.A. Puhakka. 2005. Software for quantification of labeled bacteria from digital microscope images by automated image analysis. *BioTechniques* 39:859-863.
- Abramoff, M.D., P.J. Magalhaes, and S.J. Ram. 2004. Image processing with ImageJ. *Biophotonics Int.* 11:36-42.
- Carpenter, A.E., T.R. Jones, M.R. Lamprecht, C. Clarke, I.H. Kang, O. Friman, D.A. Guertin, J.H. Chang, et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* (In press).
- Cowen, L.E. and S. Lindquist. 2005. Hsp90 potentiates the rapid evolution of new traits: drug resistance in diverse fungi. *Science* 309:2185-2189.
- Wheeler, D.B., S.N. Bailey, D.A. Guertin, A.E. Carpenter, C.O. Higgins, and D.M. Sabatini. 2004. RNAi living-cell microarrays for loss-of-function screens in *Drosophila melanogaster* cells. *Nat. Methods* 1:127-132.
- Cooper, A.A., A.D. Gitler, A. Cashikar, C.M. Haynes, K.J. Hill, B. Bhullar, K. Liu, K. Xu, et al. 2006. α -Synuclein blocks ER-golgi traffic and Rab1 rescues neuron loss in Parkinson's models. *Science* 313:324-328.
- Gibson, M.C., A.B. Patel, R. Nagpal, and N. Perrimon. The emergence of geometric order in proliferating metazoan epithelia. *Nature* 2006 Aug 9; [Epub ahead of print].
- Jones, T.R., A.E. Carpenter, and P. Golland. 2005. Voronoi-based segmentation of cells on image manifolds. *ICCV Workshop on Computer Vision for Biomedical Image Applications*, p. 535-543.
- Jones, T.R., A.E. Carpenter, P. Golland, and D.M. Sabatini. Methods for high-content, high-throughput image-based cell screening. *MIAAB 2006 Workshop Proceedings* (In press).
- Malpica, N., C.O. de Solorzano, J.J. Vaquero, A. Santos, I. Vallcorba, J.M. Garcia-Sagredo, and F. del Pozo. 1997. Applying watershed algorithms to the segmentation of clustered nuclei. *Cytometry* 28:289-297.
- Meyer, F. and S. Beucher. 1990. Morphological segmentation. *J. Vis. Commun. Image Rep.* 1:21-46.
- Ortiz de Solorzano, C., E.G. Rodriguez, A. Jones, D. Pinkel, J.W. Gray, D. Sudar, and S.J. Lockett. 1999. Segmentation of confocal microscope images of cell nuclei in thick tissue sections. *J. Microsc. (Oxford)* 193:212-226.
- Wahlby, C. 2003. Algorithms for applied digital image cytometry, p. 75. In *Center for Image Analysis*. Uppsala University, Uppsala.
- Wahlby, C., I.M. Sintorn, F. Erlandsson, G. Borgefors, and E. Bengtsson. 2004. Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections. *J. Microsc.* 215:67-76.
- Vincent, L. and P. Soille. 1991. Watersheds in digital spaces—an efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.* 13:583-598.
- Boland, M.V., M.K. Markey, and R.F. Murphy. 1998. Automated recognition of patterns characteristic of subcellular structures in fluorescence microscopy images. *Cytometry* 33:366-375.
- Gabor, D. 1946. Theory of communication. *J. Instit. Electr. Engineer.* 93:429-441.
- Haralick, R.M., K. Shanmuga, and I. Dinstein. 1973. Textural features for im-
- age classification. *IEEE Trans. Syst. Man, Cybern.* SMC3:610-621.
- Bailey, S.N., D.M. Sabatini, and B.R. Stockwell. 2004. Microarrays of small molecules embedded in biodegradable polymers for use in mammalian cell-based screens. *Proc. Natl. Acad. Sci. USA* 101:16144-16149.
- Wheeler, D.B., A.E. Carpenter, and D.M. Sabatini. 2005. Cell microarrays and RNA interference chip away at gene function. *Nat. Genet.* 37(Suppl):S25-S30.
- Bailey, S.N., S.M. Ali, A.E. Carpenter, C.O. Higgins, and D.M. Sabatini. 2006. Microarrays of lentiviruses for gene function screens in immortalized and primary cells. *Nat. Methods* 3:117-122.
- Measday, V., K. Baetz, J. Guzzo, K. Yuen, T. Kwok, B. Sheikh, H. Ding, R. Ueta, et al. 2005. Systematic yeast synthetic lethal and synthetic dosage lethal screens identify genes required for chromosome segregation. *Proc. Natl. Acad. Sci. USA* 102:13956-13961.
- Yu, A.C., Y.L. Lee, and L.F. Eng. 1993. Astroglial in culture: I. The model and the effect of antisense oligonucleotides on glial fibrillary acidic protein synthesis. *J. Neurosci. Res.* 34:295-303.
- Classen, A.K., K.I. Anderson, E. Marois, and S. Eaton. 2005. Hexagonal packing of *Drosophila* wing epithelial cells by the planar cell polarity pathway. *Dev. Cell* 9:805-817.

Received 15 May 2006; accepted 7 July 2006.

Address correspondence to Anne E. Carpenter, Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142, USA. e-mail: carpenter@wi.mit.edu

To purchase reprints of this article, contact: Reprints@BioTechniques.com