

## Open-Source Software For Scoring Complex Image-Based Assays

M.A. Bray\* and A.E. Carpenter\*

\* Broad Institute Imaging Platform, 7 Cambridge Center, Cambridge, MA 02141

As automated microscopes become widespread, researchers investigating cells, organisms, or materials are increasingly interested in extracting quantitative information from images. Traditionally, researchers gained tremendous insight from images through qualitative visual inspection or small-scale manual image quantification. However, the image acquisition rate currently attainable easily outstrips the ability to visually inspect them. Automated image analysis software can produce rich, quantitative results from hundreds of thousands of experimental samples, while simultaneously speeding analysis, reducing subjectivity and increasing statistical power.

To enable rapid development and execution of automated image analysis routines for biological images, we have developed two open-source software projects for image analysis and data mining and released them to the broader scientific community: CellProfiler for image analysis [1,2] and CellProfiler Analyst [3,4] for the exploration and analysis of high-dimensional data from images, both available from [www.cellprofiler.org](http://www.cellprofiler.org) (FIG. 1).

By using CellProfiler to chain together individual image processing modules, researchers have created complex analysis pipelines to accomplish a broad range of projects without requiring algorithm development for each new application. Such projects include identification of a wide variety of biological objects, ranging from yeast colonies, neurons and stem cell co-cultures, to whole organism studies of *C. elegans* worms (FIG. 2). Sub-compartments of each object can be identified (for example, the nucleus and cytoplasm of cells) if appropriate stains are used. Multiple features of each object and associated sub-compartments are then measured, including size, shape, intensity and texture of each color channel in the original image. Additional modules measure image features relevant for quality control.

For complex or subtle assays, it is often not clear which features are suited for classifying the cells or other biological objects as “positive” or “negative” for the phenotype of interest in the experiment. CellProfiler Analyst provides a powerful machine learning tool that can nevertheless be used by beginners to classify objects of interest based on multiple features. The combination of these analytical and classification tools enable most image-based assays to be scored automatically and quantitatively.

### References

- [1] M.R. Lamprecht et al., *Biotechniques*. 42 (1) (2007) 71.
- [2] A.E. Carpenter et al., *Genome Biol.* 7 (2006) R100.
- [3] T.R. Jones et al., *BMC Bioinformatics*. 9 (1) (2008) 482.
- [4] T.R. Jones et al., *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 1826.

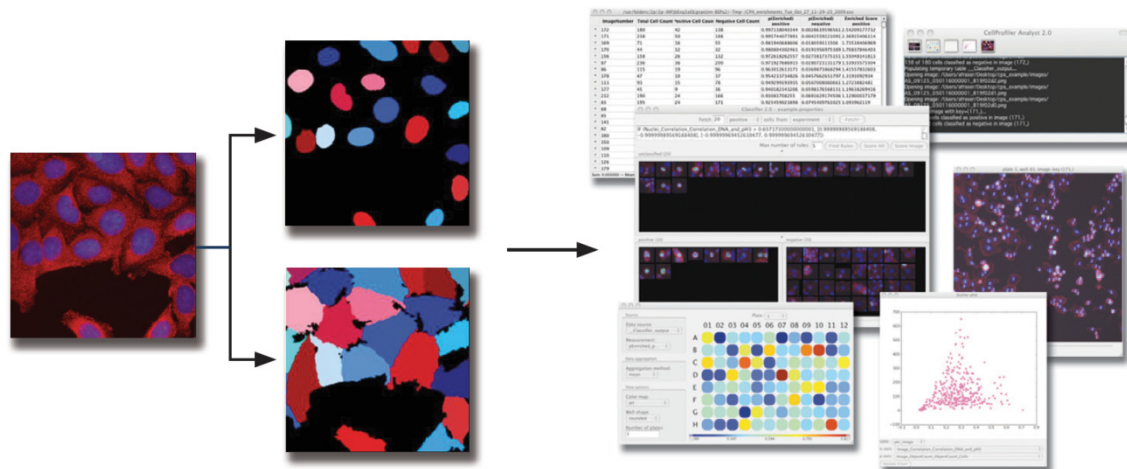


FIG. 1. Schematic of workflow in CellProfiler (image analysis) and CellProfiler Analyst (data analysis and image exploration).

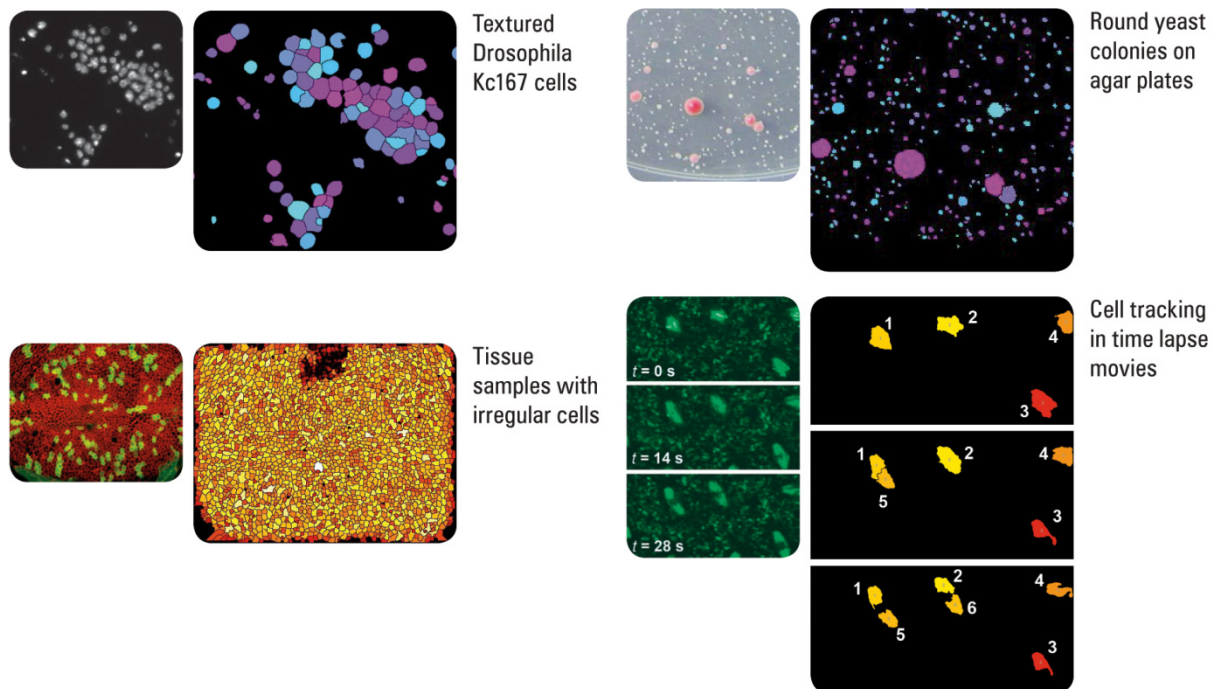


FIG. 2. Examples of typical image-based assays. For each panel, the left image shows the raw image from fluorescence or bright field microscopy; the right image shows the identified objects to be used for downstream quantification.