

# Improved structure, function, and compatibility for CellProfiler: modular high-throughput image analysis software

Lee Kametsky<sup>1</sup>, Thouis R. Jones<sup>1</sup>, Adam Fraser<sup>1</sup>, Mark-Anthony Bray<sup>1</sup>, David J. Logan<sup>1</sup>, Katherine L. Madden<sup>1</sup>, Vebjorn Ljosa<sup>1</sup>, Curtis Rueden<sup>2</sup>, Kevin W. Eliceiri<sup>2</sup>, and Anne E. Carpenter<sup>1\*</sup>

<sup>1</sup> Imaging Platform, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA 02142

<sup>2</sup> Laboratory for Optical and Computational Instrumentation, Univ. Wisconsin, Madison, Wisconsin, USA 53706

Associate Editor: Prof. Alfonso Valencia

## ABSTRACT

**Summary:** There is a strong and growing need in the biology research community for accurate, automated image analysis. Here, we describe CellProfiler 2.0, which has been engineered to meet the needs of its growing user base. It is more robust and user-friendly, with new algorithms and features to facilitate high-throughput work. ImageJ plugins can now be run within a CellProfiler pipeline.

**Availability and Implementation:** CellProfiler 2.0 is free and open source, available at <http://www.cellprofiler.org> under the GPL v. 2 license. It is available as a packaged application for Macintosh OS X and Microsoft Windows and can be compiled for Linux.

**Contact:** [anne@broadinstitute.org](mailto:anne@broadinstitute.org)

few to hundreds of thousands of images). A variety of cell types have been analyzed, including budding yeast, *Drosophila*, mouse, rat, and dozens of human cell types. The diverse measurements generated by CellProfiler provide raw material for machine-learning algorithms that can identify challenging phenotypes (Jones et al., 2009; Misselwitz et al., 2010; Ramo et al., 2009).

CellProfiler fills a unique role in the software landscape. It is a modular, high-throughput, open-source biological-image-analysis package, and it won the 2009 Bio-IT World Best Practices Award in IT & Informatics. CellProfiler 2.0 improves upon the design of the original version, resulting in professionally engineered software with improved usability and functionality, as well as integration with other open-source image-related software.

## 1 INTRODUCTION

CellProfiler is freely available, open-source software that enables researchers without training in computer programming to measure biological phenotypes quantitatively and automatically from thousands of images. With an interface designed by biologists and underlying algorithms developed by computer scientists, CellProfiler bridges the gap between advanced image-analysis algorithms and scientists who lack computational expertise. In the four years since its publication (Carpenter et al., 2006; Lamprecht et al., 2007), it has been rapidly adopted by the worldwide biological community and cited in more than 250 articles. Roughly half of its users are outside the United States. CellProfiler was initially designed for high-throughput image analysis but is often used for small-scale projects. This highlights the trend towards quantifying information in images regardless of experiment size.

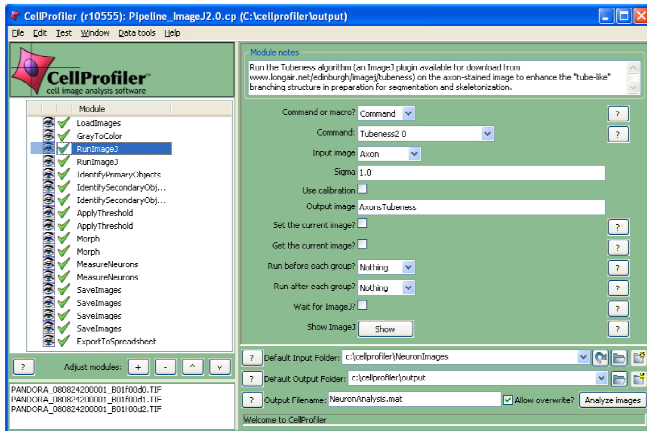
CellProfiler's interface lets researchers build customized chains of interoperable image-analysis modules to identify and measure biological objects and features in images. These modular pipelines can be saved and shared with colleagues. CellProfiler has been used to measure individual cells, colonies of cells, and whole organisms in a wide range of assays (e.g., counting cells, measuring staining intensities, and scoring complex phenotypes with machine learning) and at many experimental scales (from a

## 2 IMPROVEMENTS IN CELLPROFILER 2.0

*Robust infrastructure and interoperability:* We redesigned the software's infrastructure while porting it from the proprietary MATLAB language to the open-source Python language, making use of the high-performance scientific libraries NumPy and SciPy (Oliphant, 2007). While retaining the successful attributes of CellProfiler 1.0 (Supplemental Figure 1 and Supplemental Table 1), CellProfiler 2.0 compares favorably to CellProfiler 1.0 in terms of performance (Supplemental Figure 2) and features (Supplemental Table 2). Object-oriented design and professional software practices were integral to the porting effort, including version control, a continuous build process, and the development of an extensive validation suite.

CellProfiler 2.0 is designed to be extensible and interoperable; its plug-in interface allows outside developers to write and distribute new CellProfiler modules. We use Cython (<http://www.cython.org>) to implement computationally intensive algorithms, as well as bridge to precompiled libraries including Java via the Java Native Interface (JNI). The Java/Python bridge allows CellProfiler 2.0 to load nearly 100 image formats via the Open Microscopy Environment (OME) Consortium's Bio-Formats library (<http://www.loci.wisc.edu/software/bio-formats>). Because five percent of CellProfiler-citing papers also used ImageJ (<http://rsbweb.nih.gov/ij>), we built a bridge to run ImageJ macros in the context of a CellProfiler pipeline. In our own research, we have used third-party ImageJ plug-ins via CellProfiler to enhance neurites in images (Supplemental Figure 1A) and detect focal-planes in 3-dimensional images.

\*To whom correspondence should be addressed.



**Figure 1:** User interface for CellProfiler 2.0

**User-oriented improvements:** CellProfiler 2.0 has a much-enhanced user interface for editing pipelines (Figure 1), including drag-and-drop operations, context-sensitive menus, undo capabilities, user-friendly error reporting, and context-dependent warnings for mistakes in a pipeline's settings (Supplemental Figure 1B). A newly-designed test mode allows a researcher to step through a pipeline and repeatedly adjust settings (Supplemental Figure 1C) to optimize image analysis. Within each module, CellProfiler shows only those settings relevant to the user's existing choices, resulting in a concise and comprehensible display. Extensive context-dependent help guides users in choosing settings for their assay (Supplemental Figure 1D). Pipelines are now saved in a human-readable text format (Supplemental Data: Example CellProfiler 2.0 pipeline file).

**New and improved algorithms:** For neuron image analysis, CellProfiler 2.0 includes operations to enhance neurites and to measure their branching, and algorithms for neuron-specific metrics are in development. An updated time-lapse object-tracking module implements a recently developed algorithm based on a linear-assignment approach (Jaqaman et al., 2008). New morphological operations can find the convex hull of foreground objects and enhance dark holes in images. Illumination correction options now include spline fitting (Lindblad and Bengtsson, 2001), and thresholding options have been extended to partition intensities into three classes instead of the typical two. Other changes include an algorithm for more accurate operations on masked images (Knutsson and Westin, 1993), faster measurement of Zernike-based shape features (Supplemental Figure 2), and improved measurement of Gabor (Supplemental Figure 3) and Haralick texture features (Supplemental Table 3).

**Enhancements for high-throughput use:** CellProfiler can be run in batch mode: sets of images are partitioned between CellProfiler instances running on separate computing cores or cluster nodes in a distributed environment. In CellProfiler 2.0, images can be loaded via HTTP or located based on a comma-delimited text file containing image file locations, which might be generated by automated microscopes or laboratory information systems. Metadata about the images can also be loaded similarly. CellProfiler 2.0 has enhanced database capabilities and is now able to upload directly to MySQL or SQLite databases during image processing. CellProfiler 2.0's FlagImage module can exclude images from analysis based on measurements of image quality, such as blurriness and presence of debris. Images can be grouped for aggregate operations, such as illumination correction of images

on a per-plate basis or analysis of multiple time-lapse movies or three-dimensional image stacks. More detailed information on CellProfiler and high-throughput screening is available at <http://www.cellprofiler.org/hcs.html>.

**Future directions:** We will use the improved infrastructure and design of CellProfiler 2.0 as the basis for our future work. Where feasible, we will continue to leverage existing open-source projects to add functionality, such as software for workflow management (e.g., OMERO and KNIME) and classification of pixels or whole images by machine learning (e.g., Wndchrm and Ilastik). While supporting contributions from other developers, we will also develop novel algorithms for CellProfiler based on our ongoing research, including time-lapse and three-dimensional image analysis, metrics and corrections for assay quality control and performance evaluation, and algorithms for *C. elegans* image-based screens (Riklin-Raviv et al., 2010; Wählby et al., 2010).

## ACKNOWLEDGEMENTS

The authors thank members of their laboratories for contributing to the development of the software and this manuscript, especially Shravas Rao and Emily Schloff.

**Funding:** This work was supported by the National Institutes of Health [R01 GM089652-01 to AEC, RC2 GM092519-01 to KWE, and NIH RL1 HG004671, which is administratively linked to RL1 CA133834, RL1 GM084437, and UL1 RR024924].

**Conflict of Interest:** None declared.

## REFERENCES

- Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., Golland, P. and Sabatini, D.M. (2006) CellProfiler: image analysis software for identifying and quantifying cell phenotypes, *Genome Biol.* 7, R100.
- Jaqaman, K., Loerke, D., Mettlen, M., Kuwata, H., Grinstein, S., Schmid, S.L. and Danuser, G. (2008) Robust single-particle tracking in live-cell time-lapse sequences, *Nat Methods*, 5, 695-702.
- Jones, T.R., Carpenter, A.E., Lamprecht, M.R., Moffat, J., Silver, S.J., Grenier, J.K., Castoreno, A.B., Eggert, U.S., Root, D.E., Golland, P. and Sabatini, D.M. (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.
- Knutsson, H. and Westin, C.-F. (1993) Normalized and differential convolution: Methods for Interpolation and Filtering of incomplete and uncertain data, *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 515-523.
- Lamprecht, M.R., Sabatini, D.M. and Carpenter, A.E. (2007) CellProfiler: free, versatile software for automated biological image analysis, *Biotechniques*, 42, 71-75.
- Lindblad, J. and Bengtsson, E. (2001) A comparison of methods for estimation of intensity nonuniformities in 2D and 3D microscope images of fluorescence stained cells, *Proceedings of the 12th Scandinavian Conference on Image Analysis (SCIA)*, Bergen, Norway, 264-271.
- Misselwitz, B., Strittmatter, G., Periaswamy, B., Schlumberger, M., Rout, S., Horvath, P., Kozak, K. and Hardt, W.-D. (2010) Enhanced CellClassifier: a multi-class classification tool for microscopy images, *BMC Bioinformatics*, 11, 30.
- Oliphant, T.E. (2007) Python for Scientific Computing, *Comput. Sci. Eng.*, 9, 10-20.
- Ramo, P., Sacher, R., Snijder, B., Begemann, B. and Pelkmans, L. (2009) CellClassifier: supervised learning of cellular phenotypes, *Bioinformatics*, 25, 3028-3030.
- Riklin-Raviv, T., Ljosa, V., Conery, A.L., Ausubel, F.M., Carpenter, A.E., Golland, P. and Wählby, C. (2010) Morphology-Guided Graph Search for Untangling Objects: *C. elegans* Analysis. *Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. Beijing, China.
- Wählby, C., Riklin-Raviv, T., Ljosa, V., Conery, A.L., Golland, P., Ausubel, F.M. and Carpenter, A.E. (2010) Resolving clustered worms via probabilistic shape models. *IEEE International Symposium on Biomedical Imaging (ISBI): From Nano to Macro*. Rotterdam, The Netherlands.