1. Collins, T.J. *Biotechniques* **43** Suppl, 25–30 (2007).
2. Carpenter, A.E. *et al. Genome Biol.* **7**, R100 (2006).
3. Shamir, L., Delaney, J.D., Orlov, N., Eckley, D.M. & Goldberg, I.G. *PLoS Comput. Biol.* **6**, e1000974 (2010).
4. Huang, K. & Murphy, R.F. in *Proc. of the 2004 2nd IEEE International Symp. on Biomedical Imaging*, 1139–1142 (2004).
5. Shamir, L. *et al. Source Code Biol. Med.* **3**, 13 (2008).
6. Bhattacharya, A. *et al. Proc. of the Fifth IEEE International Conf. on Data Mining*, 50–57 (2005).
7. Nanni, L. & Lumini, A. *Artif. Intell. Med.* **43**, 87–97 (2008).
8. Cruz-Roa, A., Caicedo, J.C. & González, F.A. *Artif. Intell. Med.* **52**, 91–106 (2011).
9. Loo, L.H. *et al. J. Cell Biol.* **187**, 375–384 (2009).
10. Fuchs, F. *et al. Mol. Syst. Biol.* **6**, 370 (2010).

Annotated high-throughput microscopy image sets for validation

To the Editor: Choosing among algorithms for analyzing biological images can be a daunting task, especially for nonexperts. Software toolboxes such as CellProfiler[1,2] and ImageJ[3] make it easy to try out algorithms on a researcher's own data, but it can still be difficult to assess whether an algorithm will be robust across an entire experiment based on the small subset of images that is practical to examine or annotate. Even if controls are available, a pilot high-throughput experiment may be insufficient to show that an algorithm will robustly identify rare phenotypes and handle the experimental artifacts that will invariably be present in a high-throughput experiment. It is therefore useful to know that a particular algorithm has proven superior on several similar image sets. The performance comparisons presented in papers that introduce new algorithms are often not very helpful for assessing this because each study typically relies on a different test image set (often to the advantage of the proposed algorithm), the algorithms compared may not be the ones the researcher is most interested in and the authors may not have implemented other algorithms as optimally as their own. Although biologists should always also validate algorithms on their own images, it would be useful if developers would quantitatively test new algorithms against a publicly available established collection of image sets. In this way, objective comparison can be made to other algorithms, as tested by the developers of those algorithms. We see a need for such a collection of image sets, together with ground truth and well-defined performance metrics.

Here we present the Broad Bioimage Benchmark Collection (BBBC), a publicly available collection of microscopy images intended as a resource for testing and validating automated image-analysis algorithms. The BBBC is particularly useful for high-throughput experiments and for providing biological ground truth for evaluating image-analysis algorithms. If an algorithm is sufficiently robust across samples to handle high-throughput experiments, low-thoughput applications also benefit because tolerance to variability in sample preparation and imaging makes the algorithm more likely to generalize to new image sets.

Each image set in the BBBC is accompanied by a brief description of its motivating biological application and a set of ground-truth data against which algorithms can be evaluated. The ground truth sets can consist of cell or nucleus counts, foreground and background pixels, outlines of individual objects, or biological labels based on treatment conditions or orthogonal assays (such as a dose-response curve or positive- and negative-control images). We describe canonical ways to measure an algorithm's performance so that algorithms can be compared against each other fairly, and we provide an optional framework to do so conveniently within CellProfiler. For each image set, we list any published results of which we are aware.

The BBBC is freely available from http://www.broadinstitute.org/bbbc/. The collection currently contains 18 image sets, including images of cells (*Homo sapiens* and *Drosophila melanogaster*) as well as of whole organisms (*Caenorhabditis elegans*) assayed in high throughput. We are continuing to extend the collection during the course of our research, and we encourage the submission of additional image sets, ground truth and published results of algorithms.

Vebjorn Ljosa, Katherine L Sokolnicki & Anne E Carpenter

Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts, USA.
e-mail: ljosa@broad.mit.edu or anne@broadinstitute.org

1. Carpenter, A.E. *et al. Genome Biol.* **7**, R100 (2006).
2. Kamentsky, L. *et al. Bioinformatics* **27**, 1179–1180 (2011).
3. Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. *Nat. Methods* **9**, 671–675 (2012).