Annotated high-throughput microscopy image sets for validation

Vebjorn Ljosa, **Katherine L Sokolnicki**, and **Anne E Carpenter**
Broad Institute of Massachusetts Institute of Technology and Harvard, 7 Cambridge Center, Cambridge, Massachusetts, USA.

To the editor: Choosing among alternative algorithms for analyzing biological images can be a daunting task, especially for non-experts. 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 testing on a few images. Even if positive controls or orthogonal assays are available for validation, a pilot experiment may be insufficient to show that an algorithm will be robust to the rare phenotypes and experimental artifacts that will invariably be present in the eventual high-throughput experiment. It is then useful to know that a particular algorithm has been superior on several similar image sets. The performance comparisons presented in papers that introduce new algorithms are often not very helpful, as each paper typically uses a different test image set (often to the advantage of the proposed algorithm), the algorithms chosen for comparison may not be the most appropriate for the application at hand, and the authors may not have configured or implemented the other algorithms as optimally as their own. To help guide biologists in their choices, it would be ideal for algorithm developers to quantitatively test new algorithms against a publicly available, established collection of image sets so 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 validation of automatic image-analysis algorithms. The BBBC focuses on high-throughput experiments and providing biological ground truth for evaluating image-analysis algorithms. Striving for the robustness across samples that is needed in high-throughput experiments benefits low-throughput applications as well because tolerance to variability in sample preparation and imaging makes an algorithm more likely to generalize to new image sets.

Each image set in the BBBC is accompanied by a brief description of the relevant biological application and a set of ground-truth data against which algorithms can be evaluated. The ground truth sets are of four kinds: nucleus or cell counts, foreground and background pixels, outlines of individual objects, and biological labels (e.g., 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

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 Broad Bioimage Benchmark Collection is freely available from http://www.broadinstitute.org/bbbc/. The collection currently contains 18 image sets, including images of cells (human and *Drosophila melanogaster*) as well as whole organisms (*Caenorhabditis elegans*) assayed in high throughput. We encourage the submission of additional image sets, ground truth, and published results of algorithms.

Acknowledgments

References

1. Carpenter AE, et al. Genome Biol. 2006; 7:R100. [PubMed: 17076895]

2. Kamentsky L, et al. Bioinformatics. 2011; 27:1179–80. [PubMed: 21349861]

3. Abràmoff M, Magalhaes P, Ram S. Biophotonics International. 2004; 11:36–42.