hot topic

High-Content Screening: Practical Advice, Tools and Resources

By Anne E. Carpenter, Ph.D.
Director, Imaging Platform
Broad Institute of Harvard and MIT

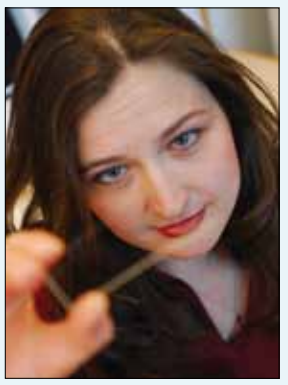
This is the second article in our two-part series on high-content screening (HCS). In this article, Dr. Carpenter gives practical advice on how to get started and what to keep in mind as you move forward. She will also describe her group's open-source software project, CellProfiler, winner of the 2009 Bio-IT World Best Practices award in IT and Informatics. Readers who want to learn more about current research in the field are also referred to the August 2010 issue of the ***Journal of Biomolecular Screening****—a special issue devoted to HCS, and the August 2010 issue of* ***SBS News****.*

1) How do I choose software for my high-content screening project?

If you are working with a standard cell type and a standard phenotype, I would generally advise that you start by testing the commercial software that comes with your high-throughput microscope. This software is usually a convenient package, integrated with a database, that stores images from your microscope and is tailored to specific, popular assays. Using this software offers a low barrier of entry that often allows people to rapidly analyze their phenotype.

2) What if existing software doesn't work for my phenotype of interest?

Probably the next step is to work with experts in the software that is available to you, to make sure you are using the software to its best capabilities. Often a simple tweak can dramatically improve results. If results are still not optimal, most microscopes allow you to export images in a standard format that can be read by other software, so the next option is to try other high-throughput imaging software available to you, such as my group's open-source CellProfiler (see box, right).


Anne E. Carpenter, Ph.D.

If existing software fails and new algorithms are truly needed, you will need to enlist the help of an image analysis expert. Be sure to work with the computer scientist very closely and iteratively to ensure that he or she understands the characteristics and goals of the experiment; because computer scientists often struggle to appreciate the inherent variability in biological experiments, working with someone who has prior experience with screening is very helpful.

If you work with local computer scientists to develop a new algorithm for your assay, they may be very excited about building a stand-alone piece of software to carry out the functions they have developed. By all means, encourage them instead to add their

(continued on page 7)

CELLPROFILER: an open-source project

CellProfiler is free, open-source software designed to enable biologists without training in computer vision or programming to quantitatively measure phenotypes from thousands of images automatically. The author and Thouis (Ray) Jones published the first paper detailing CellProfiler in 2006 and they continue to develop it to suit the needs of the high-content screening community, of which they are a part. The software has so far been cited in more than 200 papers and its own debut publication is currently *Genome Biology*'s eighth most popular paper of all time. In 2009, CellProfiler won the Bio-IT World Best Practices Award in IT & Informatics. The software project received several initial small grants, including an academic grant from SBS in 2004, and this year the project is being supported in part by the National Institutes of Health (R01 GM089652-01).

In CellProfiler, researchers perform the many steps of image analysis, such as illumination correction, object identification and measurement, by placing modules for each step into a pipeline. Then, the settings for each module are adjusted to suit the cell/organism type and phenotype of interest. For help getting started, there is an online Q & A forum, and the Broad Institute Imaging Platform offers training workshops to the scientific community. Although written for automated high-throughput screening, the software is also very popular for low-throughput experiments where quantification of visible phenotypes is needed: 70 percent of the papers citing the use of CellProfiler involve experiments with less than 100 images.

Efforts to increase interoperability among biological imaging software have gained momentum in the past few years. Most notably for the CellProfiler community is its recent interfacing with the popular BioFormats project (www.loci.wisc.edu/software/bio-formats), developed in Kevin Eliceiri's group at the University of Wisconsin-Madison (www.loci.wisc.edu) and by Glencoe Software as part of the Open Microscopy Environment project (www.openmicroscopy.org). Because of this connection, CellProfiler can now read 75 file formats, including most of those produced by the major low- and high-throughput microscopy manufacturers. There is also a major effort underway to restructure the popular ImageJ software such that it can more readily interface with other software projects, such as CellProfiler (www.imagejdev.org).

Student Internships

(continued from page 1)

nucleus of interested companies to partner with. Now we are alerting our academic members to make them aware that these internship opportunities exist. Interested student members of SLAS can apply at no charge; students who don't belong to SLAS will pay $25 to apply and become members in the process.

In addition to internship opportunities, hiring organizations will provide any stipends and/or room and board provisions that may be available. For candidates who are offered internships outside of their hometowns, SLAS will pay for travel (one trip) to and from their internship assignments (this does not include daily commuting). Internships will run for eight to 10 weeks in the summer of 2011. For more information, see SLAS.org. ❊

From the SLAS President

(continued from page 1)

SLAS recognizes the importance of emerging talent and this next generation of thought leaders. Online and on-site programs are tailored exclusively to assist them in gaining professional ground. The SLAS Board of Directors is especially excited about the potential offered by the new SLAS internship program (see page one), and the fact that students interested in participating in the SBS 17th Annual Conference in Orlando now are eligible to apply for SLAS Student Travel Awards.

At the other end of the continuum, our veteran members are honored and appreciated for their years of experience (retired professionals may maintain memberships in *both* sections for just $50). Soon, members who actively contribute to SLAS may qualify for the new "membership by contribution" provision, which will allow complimentary membership for the relevant calendar year.

Collaborations

The mutual interests of our now combined membership are already taking root in the dynamic exchange of participation and information between the LabAutopedia wiki and the Assay Guidance Manual (AGM) wiki, which had their roots in ALA and SBS members respectively. These wikis are online collaboration tools through which multiple contributors create and develop (within certain limits) the content. Thanks to the SLAS IT team, crossover participation is steadily increasing via trans-wiki inclusion. Many AGM items are mirrored live in LabAutopedia, enabling readers to cross dimensions seamlessly as they follow the threads of their individual information searches.

Literally and metaphorically, these examples speak to the mission and goals of the new SLAS—to provide forums for education and information exchange to encourage the study of, and improve the science and practice of, laboratory automation and screening. Dedicated to our expanding scientific bandwidth, we remain mindful of individual goals and priorities. Our association is taking shape as a practical yet future-focused force. Thank you for sticking with us and providing the knowledge, experience, expertise and enthusiasm that will take us to the next level. ❊

High-Content Screening

(continued from page 5)

algorithm to an existing software toolbox such as ImageJ or CellProfiler. This ensures ease of use and cross-platform compatibility, as well as longevity and broad dissemination of the software. If you want the new algorithm to make an impact in the scientific community, picture what will happen to a stand-alone piece of software five years after the graduate student leaves the laboratory!

My own research group is focused on developing new image analysis and data mining methods for screening, and we enjoy working closely with assay developers from around the world as a project develops. This interaction enables sample preparation and image acquisition to be coordinated with image analysis, improving the overall data quality in a project. Our goal in collaborative projects is to develop methods and implement them in CellProfiler so that biologists can run the analysis themselves and take full ownership of image analysis. Again, putting tools directly in the hands of the researcher who understands the goals and characteristics of the experiment yields higher quality data than passing images and data back and forth between biologists and computer scientists.

3) How challenging is it to learn to set up customized image analysis for high-throughput screening?

It is far easier than it used to be. Whereas 10 years ago custom analyses required the direct help of a computer scientist, modern software in this field is putting more power into the hands of biologists. Many commercial packages offer some sort of macro-recording or script-writing function that is easier to learn than true programming languages. Our software, CellProfiler, was designed from the ground up to be modular and flexible, to enable customization by biologists without requiring programming skills.

4) What about machine learning—how hard is it to learn to use?

Biologists are often surprised to find that machine learning, while it sounds intimidating, has made the process of image analysis for high-throughput screening far easier than it used to be. If you are able to tweak your image analysis settings to identify the cellular compartments that are labeled in your experiments, and you can generate a large, generic set of measurements from each cell, then you are ready for the machine-learning step.

Machine learning requires less image processing knowledge than designing a particular measurement to quantify a phenotype of interest. We created a tool, CellProfiler Analyst's "Classifier" function, to enable biologists to readily use machine learning, which is very powerful, to identify subtle and rare phenotypes in images. Biologists are able to learn to use "Classifier" in only a few minutes, by visually sorting cells in their experiment according to the phenotype they display. The computer then iteratively learns the phenotype with feedback from the biologist and finally scores the entire experiment.[1,2]

For some experiments where the phenotype is fairly penetrant in the image, you can skip the step of accurately identifying cells and their subcompartments, and instead use a machine-learning method that examines arbitrary neighborhoods within images. One of the best-tested algorithms for this is the open-source WND-CHARM, developed in Ilya Goldberg's group at NIH.[3] We plan to make this approach available in CellProfiler soon. Overall, these machine-learning methods are making image analysis easier, not more complicated.

5) What should I read to understand image analysis better?

In addition to a review article on image-based screening[4], Vebjorn Ljosa and I just wrote a tutorial to familiarize biologists with image analysis for screening.[5] This open-access article covers the basics and provides guidance on practical tips for sample preparation and image acquisition. It also has an extensive list of resources in the field: review articles, societies, conferences, workshops, websites, discussion groups, journals, books and image analysis software. ❊

REFERENCES

1. T. R. Jones et al., *Proc Natl Acad Sci USA* 106, 1826 (Feb 2, 2009).

2. T. R. Jones et al., *BMC Bioinformatics* 9, 482 (2008 Nov 15, 2008).

3. N. Orlov et al., *Pattern Recognition Letters* 29, 1684 (Aug, 2008).

4. A. E. Carpenter, *Nat Chem Biol* 3, 461 (Aug, 2007).

5. V. Ljosa, A. E. Carpenter, *PLoS Computational Biology* 5, e1000603 (Dec, 2009).