hot topic

High-Content Screening: FAQs about Trends, Technologies, and Outcomes

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

This is the first of a two-part series on high-content screening (HCS). In this article, Dr. Carpenter responds to questions about where the field is headed and why you should consider using this new technology. In the second article, which will run in the October 2010 issue of SBS News, she will give practical advice on how to get started and what to keep in mind. 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 referred to the August issue of the Journal of Biomolecular Screening—a special issue devoted to HCS.

Interest in high-value, high-throughput methods to assist in drug discovery and pathway analysis continues to grow in all sectors—the pharmaceutical industry, biotechnology companies, and academia. I am often asked about trends in high-content screening (HCS), a field that has grown steadily over the past 10 years and has become a common tool in drug discovery. Also known as image-based screening or microscopy-based screening, the field began with the development and refinement of robotic microscopes and has flourished in recent years with increasingly powerful image-analysis software.

As head of an academic research group developing novel image analysis algorithms and data mining methods for image-based screens, I am often asked about the benefits of HCS, as well as current and future trends in HCS. Here, I respond to some frequently asked questions.

1) Why the enthusiasm for image-based screening?

The quality and quantity of information that can be extracted from a single sample interrogated by imaging is much higher than almost any other modality. The richness of information stems from several sources. Several fluorescence stains (antibodies or dyes) can usually be multiplexed in the same sample—routinely, four to six different cellular components can be labeled fluorescently, and bright-field images can also be acquired.

Furthermore, even a single fluorescent stain can reveal dozens of different phenotypes of interest. The sizes, shapes, intensities, and textures of each dye's target may be relevant to the physiological process being interrogated in the experiment. Multiple variations in a stain's pattern may be distinguishable from each other and relevant for characterizing the perturbations in each sample. It is not unusual, for example, to use a single DNA stain to measure a dozen different biologically interpretable phenotypes, in addition to hundreds of other raw features useful for machine learning.

Imaging is also powerful because it allows the phenotype of each individual cell to be determined. When a phenotype is very rare, it is an absolute necessity to be able to identify and count individual cells with the phenotype; imaging permits this approach as opposed to whole-population modalities where rare cells are lost in the noise of a large population of non-responding cells.

Imaging also allows identification of phenotypes in a relevant subset of the cell population. For example, some assays require measuring a phenotype from only one of a physiologically relevant mixture of several cell types, such as identifying leukemic stem cells cultured with bone marrow stroma, or primary hepatocytes cultured with fibroblasts.

High-content screens are conducted in living cells or whole organisms, which is a more physiological environment than biochemical assays. Properly managed, imaging is non-destructive, so cells can be imaged over time, revealing dynamic events in living cells.

Perhaps most importantly, HCS allows us to directly interrogate a phenotype of interest, rather than selecting biochemical proxies presumed to be in the pathway relevant to the phenotype of interest. Thus, a specific protein target need not be identified in advance. This is particularly powerful for those disease areas where we haven't a clue what pathways are involved, which is especially the case for many mental illnesses.

2) I think I can use microscopy as the readout for my phenotype... Should I?

Imaging also has its downsides. It should not be used if you have an alternative, faster plate-based approach to gain the same information from each sample. Image acquisition is still a bottleneck; microscopy for each plate takes tens to hundreds of times longer than a plate reader.

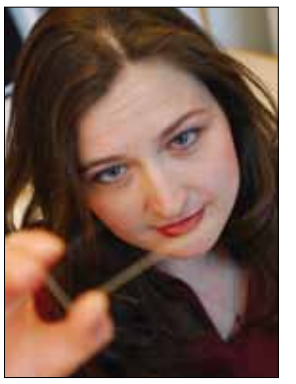
Image-based assay development is often challenging, as the cell density, staining conditions, and image acquisition steps must be carefully optimized. And image analysis can be a real challenge as well, requiring many rounds of optimization and often a computational cluster for processing. Typically, if you do not intend to multiplex an imaging assay or extract more than a single phenotype from the images you collect, it would be better to design an assay that can be done in a bulk plate-reader format instead. That said, our collaborators often begin with a single phenotype they would like to measure, but once they become aware of the wealth of phenotypes that can be simultaneously measured by imaging, they often design very sophisticated multiplexed imaging assays that are worth the effort. By adding relevant stains they essentially end up conducting a suite of secondary screens in parallel with the primary screen.

3) Where is high-content screening headed?

Bigger and better! "Bigger" in the sense that we are definitely seeing trends towards screening increasing numbers of samples. These can be chemical compound libraries as well as genetic libraries—for example, RNA interference libraries, gene overexpression libraries, and panels of genetic mutants.

In pharmaceutical companies, image-based screening is steadily moving from solely low-throughput secondary assays to medium-to-high-throughput primary screens. Researchers are seeing the value in extracting rich information out of a screen in the first pass so that the initial leads are already well characterized and of high quality.

For academics, image-based screens were already quite popular because funding considerations limited the scale of screens anyway. Thus, it was worth developing a sophisticated imaging screen to maximize rich information content at the primary screen stage. In the academic world, screens of more than 100,000 samples were almost unheard of a few years ago—but now, screens of a library of more than 300,000 chemical compounds are performed and funded through the NIH's Molecular Libraries Probe Production Centers Network (MLPCN) program (http://mli.nih.gov). This trend seems likely to continue as the new NIH Director, Francis Collins, has made it clear that high-throughput technologies are one of the five "areas of particular promise" in biomedical research today.[1] Several high-profile, high-throughput programs have already been announced, such as the program for Large Scale Production of Perturbagen-Induced Cellular Signatures (RFA-RM-10-003) and the stimulus package-funded NIH Director's Opportunity for Research in Five Thematic Areas (RFA-OD-10-005).


Anne E. Carpenter, Ph.D.

"Better" in that image acquisition and image analysis steps have become much more robust in the past five years, and data consistency is much improved as a result. Furthermore, researchers are taking much better advantage of the abilities of imaging to perform more sophisticated experiments—increasing the complexity of phenotypes measured; more complex cell cultures and even whole organisms; more involved imaging techniques such as time-lapse or three-dimensional imaging; and more sophisticated computational techniques to extract phenotypes of interest.

The result: leads emerging from a typical high-content screen today are better characterized, more specific, and of higher quality than those from a typical screen five years ago. For example, the European MitoCheck project just published results from their astonishing RNA interference screen of the entire genome, where three siRNAs were tested for each gene, in triplicate, with time-lapse movies for each sample spanning two full days at 30-minute intervals.[2] The 190,000 movies from this experiment will be mined for biological insights for years to come.

4) When will automated high-content screens be feasible in whole organisms?

Screens in whole organisms such as *Caenorhabditis elegans* nematodes, *Danio rerio* zebrafish, and the plant *Arabidopsis thaliana* are becoming quite common, but are rarely analyzed by automated means. Robotic sample preparation is often employed; however, visual inspection of live samples is often utilized as opposed to automated image acquisition and automated image analysis, and currently, both steps are challenging.

There have been some very impressive successes for zebrafish.[3,4] Here at the Broad Institute, Dr. Carolina Wählby is currently working on a suite of algorithms for analyzing images of *C. elegans* samples.[5,6] It is clear there is huge demand for flexible and robust whole-organism algorithms. ❊

REFERENCES
1. F. S. Collins, *Science* 2010;327: 36.
2. B. Neumann et al., *Nature* 2010;464: 721.
3. J. Rihel et al., *Science* 2010;327: 348.
4. D. Kokel et al., *Nat Chem Biol* 2010; 6: 231.
5. C. Wählby et al., *Resolving clustered worms via probabilistic shape models. Proceedings of ISBI 2010, IEEE Internatl. Symp. on Biomedical Imaging: From Nano to Macro*, Rotterdam, Netherlands.
6. T. Riklin-Raviv et al., *Morphology-Guided Graph Search for Untangling Objects: C. Elegans Analysis.* Accepted for publication in *Proceedings of MICCAI 2010, the 13th International Conference on Medical Image Computing and Computer Assisted Intervention.*