

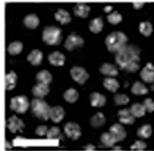
Carpenter Builds Open Source Imaging Software



Best Practices Winner: The Broad Institute of MIT and Harvard Project: CELLPROFILER
Category: IT & Informatics

By Kevin Davies
August 13, 2009 | Bio-IT World

July 21, 2009 | Anne Carpenter trained as a traditional cell biologist specializing in microscopy with no intention of writing image analysis software. “It wasn’t until I needed software to do something that existing commercial software couldn’t do that I became interested in writing software myself,” says Carpenter. The genesis of CellProfiler was “completely out of necessity.”

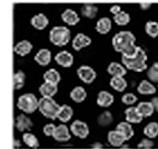


Drosophila:
Original image DNA

Carpenter found that the commercial software bundled with automated microscopes was good at measuring certain cell types, but little help measuring the size of *Drosophila* cells during her postdoc with David Sabatini at the Whitehead Institute. She came across some promising algorithms doing a literature search, but didn’t have any way of implementing them. “So I sent an email to the MIT computer science department asking if anyone could help out for a couple of hours a week.” A student named Thouis Jones agreed to help, and soon made it the subject of his Ph.D.

The satisfaction of developing useful software for the cell biology community persuaded Carpenter to abandon her postdoc project and focus on CellProfiler software development, training and implementation. “It became much more compelling to help dozens of other people working on image analysis for their projects versus doing my own,” she says.

One of those grateful beta testers was Scott Floyd, a cell biologist and physician at Beth Israel Deaconess Hospital. Floyd was screening for genes involved in cellular response to DNA damage in the search for drugs that could protect cancer patients against the side effects of radiation. He could recognize telltale increases in the speckled appearance of cell nuclei by eye, but struck out using commercial software.



Drosophila:
CP-outlined nuclei

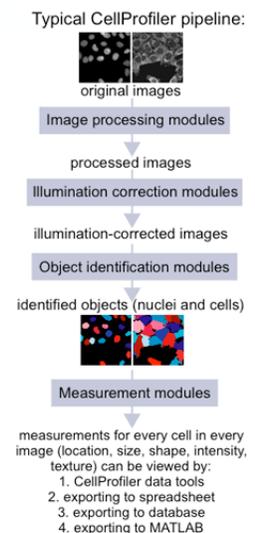
The software Carpenter built—CellProfiler—made its free open source debut in December 2005, and was detailed in *Genome Biology* in 2006. In January 2007, Jones and Carpenter established the Imaging Platform group at the Broad Institute, focusing on new algorithms and data analysis methods. From here, Carpenter can help dozens of researchers working on clinically relevant projects. “Everything we develop becomes open source, and the easiest way to get that out to the public is to put it into the CellProfiler interface.”

Profiler Packages

In contrast to the tedious and error-prone manual inspection of identifying specific cell shapes or morphology, CellProfiler’s easy point-and-click interface and modular structure allows operators to customize the workflow to a particular experiment—even computational novices. Researchers can build a “pipeline” of modules, each performing a set function on the images. This might be followed by measurements for each cell or for an entire image, such as size, location, and shape or the intensity and texture of the staining pattern within cells.

Carpenter’s team of computer scientists and biologists helps Broad colleagues test hundreds of thousands of samples to understand gene function and identify drug candidates. Her group operates “like a faculty research lab at any academic institution, but we are unique in having a very strong technology focus, and secondly, in being extraordinarily more collaborative than a typical faculty lab.”

CellProfiler comes into its own in the high-throughput analysis of images from robotic fluorescent light microscopes, such as those offered by companies like Cellomics, GE

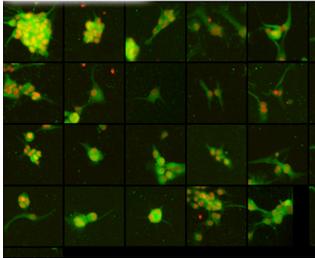


Healthcare, and PerkinElmer, essentially turning images into numbers. The software's strength lies in its flexibility and sophistication, which allow "accurate and rich measurements coming out of the cells." But Carpenter says the commercial packages still excel in their prepackaged convenience, and her team will recommend using commercial software when collaborators are screening a simple phenotype. "We only get involved when people are stumped on their project."

Maturity Level

Although CellProfiler has been gaining admirers for a few years, Carpenter only submitted for Bio•IT World's Best Practices competition once she was satisfied that the program had reached a certain level of maturity and popularity. Signs of maturity include the fact that the software was downloaded 300 times per month in 2008 and in total some 9000 times since its introduction, and has amassed more than 100 citations.

Positive cells identified with CP Analyst



Perhaps most important was "the killer application"—CellProfiler Analyst—which was submitted for publication in late 2008 and published in Proceedings of the National Academy of Sciences in early 2009. This tool looks at those measurements and performs machine-learning cell sorting. Says Carpenter: "You don't need to know anything about machine learning to use the software. It really just looks like a video game."

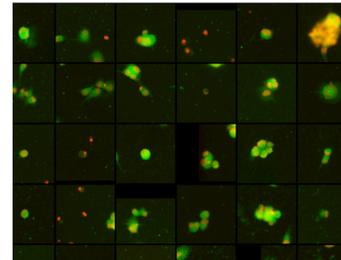
"We knew that would be a slam dunk popular tool for using CellProfiler data," she says. "Previously, if a biologist had a tough phenotype, they'd need six months writing a new algorithm. Here, provided we can find the cells in the image, we can use this machine learning. It typically takes a biologist anywhere from 1 hour to 1 day

of scoring cells by eye, and the computer has learned what they're looking for. So pretty much any phenotype we come across, we can score in a day."

CellProfiler has won many dedicated fans over the past few years. Michael Yaffe (Floyd's boss) calls CellProfiler "an indispensable component of a large-scale high-throughput screen" that "adds an entirely new dimension to analysis, leading to generation of a robust and novel dataset that will be extraordinarily useful for years to come."

Another satisfied user is Jhn McLaughlin, who runs a screening facility at Rigel Pharmaceuticals producing thousands of images weekly, and hasn't looked back since trying CellProfiler two years ago. "It had everything I needed," he says. McLaughlin likes the underlying Matlab platform, and its compatibility with a compute cluster, which is not found with all commercial packages. "My goal is to find drugs to cure disease, not learn (yet another) computer language," says McLaughlin.

Negative cells identified with CP Analyst



Carpenter's team is currently involved in numerous wide-ranging collaborations, from studying the genetic underpinnings of breast cancer with Eric Lander's group to improving the analysis of neuronal cell types, which she calls "challenging for the best algorithms." Other projects involve screening potential drugs for infectious diseases including tuberculosis in human cells, and whole-organism analysis of the nematode worm to develop novel antibiotics. On the technology side, her team is working to enable CellProfiler to do movie analysis and 3-D image analysis. "Right now, it's fairly impractical to collect large sets of 3-D images, but as that becomes more practical, we'll work on algorithms to study those images."