Anne Carpenter

Quantifying complex and subtle phenotypes in images for high content screening

In high content screening, the rate-limiting step is often developing an automated image analysis protocol to score cell images for unusual phenotypes. We have developed open-source software for accurately measuring hundreds of morphological features of every cell in every image in large-scale screens (CellProfiler, www.cellprofiler.org).   Many groups are using this flexible software to analyze previously intractable image-based phenotypes in cells and even in whole organisms like C. elegans. Recently, we have developed a supervised machine-learning system that uses the biologist's expertise to learn to score complex phenotypes based on combinations of cellular features. In this approach, biologists spend a few hours showing the computer examples of a cellular phenotype of interest in fluorescence microscopy images. Machine learning algorithms then automatically score all cells in all images, based on each cell's measured features. This focus on individual cells allows scoring rare and non-penetrant phenotypes. We used these methods to rapidly score 14 unusual cellular phenotypes induced by RNA interference in human cells in a single experiment that was originally designed to identify mitotic regulators. We have released these methods in the open-source software package CellProfiler Analyst.