Gene module discovery by cytological profiling of RNAi-perturbed cells

Anne E. Carpenter[1], Thouis R. Jones[2], Polina Golland[2], David M. Sabatini[3], et al.

[1]Broad Institute, Cambridge, MA 02142. [2]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02142. [3]Department of Cell Biology, Whitehead Institute for Biomedical Research, Cambridge, MA 02142.

Many well-studied biological pathways were originally identified by observation of a mutant phenotype (that is, an observable trait) in a model organism. Even if the biological significance of the phenotype was unknown, identification of the genes related to the phenotype formed the basis of a genetic module for further study. Identifying genes linked to an unusual or uncharacterized phenotype is a relatively unbiased way to open up new biological fields of study and characterize the functions of previously uncharacterized genes in the genome.

We therefore rapidly identified multiple gene modules in human cells based on phenotypes induced by RNA interference (RNAi, a method of genetic perturbation) in a single experiment that was originally designed to identify mitotic regulators. In this approach, biologists train a computer to recognize cellular phenotypes of interest (often quite subtle) in fluorescence microscopy images. Machine learning algorithms then distinguish cells of interest based on each cell's cytological profile, its rich set of image cytometry-measured features including size, shape, intensity and texture. Genes within the same genetic module often produce cells with a particular rare phenotype (often less than 1 in a 100,000 cells in a large screen).

We rapidly scored millions of individual cells for ~15 phenotypes, yielding multiple predicted gene modules that we are now validating and pursuing. The methods developed for this project are incorporated into the soon-to-be-released, open-source software, CellVisualizer (www.cellvisualizer.org), which allows biologists to conveniently score high-throughput screens for complex and subtle visual phenotypes.