Discerning Gene and Drug Function from High-Throughput Imaging and Morphological Profiling

Beth Cimini, Shantanu Singh, and Anne Carpenter

As computational analysis of high-throughput imaging screens has advanced and matured, so has our understanding that cell's morphological phenotypes have much to teach us about how genes and compounds function in cells. After using staining techniques such as the Cell Painting assay (Bray et al 2016), features of cell size, shape, intensity, and texture can be extracted by tools like CellProfiler (McQuin et al 2018); once extracted, these features can be built into a "profile" that allows the researcher to group chemicals by mechanism of action, associate genes into functional pathways (even learning novel pathway-pathway interactions), and even help determine which drugs can be used on a given disease (reviewed in Caicedo et al 2017). While traditionally this is done by comparing changes in the per-well average measurements after various treatments, incorporating measurements of the heterogeneity of treated cells can increase the ability to correctly group similar treatments (Rohban et al, 2019). In the absence of conventional segmentation and feature extraction, features extracted by deep networks on cropped treated single cells can also learn to classify treatments successfully (Caicedo et al 2018).

While the features and relationships learned from such experiments are valuable in and of themselves and can answer questions within a given experiment, the features extracted can also be used for other purposes. High content imaging assays can be used to aid target enrichment by learning relationships between lower throughput functional assays and features extracted from imaging assays that are more easily scaled up. Such relationships can then serve as classifiers to create "virtual screens" that increase hit rates of selected drugs by as much as 50-250X (Simm et al 2018). These advancements may substantially bring down the time and/or cost of drug discovery, making research more efficient and creating a "Rosetta stone" that allows association and comparison across many kinds of data.

Bray et al, Nature Protocols 2016. doi: http://dx.doi.org/10.1038/nprot.2016.105
McQuin et al, PLoS Biology 2018. doi: http://doi.org/10.1371/journal.pbio.2005970
Caicedo et al, Nature Methods 2017. doi: http://doi.org/10.1038/nmeth.4397
Rohban et al, Nature Communications 2019. doi: http://doi.org/10.1038/s41467-019-10154-8
Caicedo et al, 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: http://doi.org/10.1101/293431
Simm et al, Cell Chemical Biology 2018. doi: http://doi.org/10.1016/j.chembiol.2018.01.015

This work was partially funded by NIH R35 GM122547 to Anne Carpenter.