Exploring gene function and morphology using JUMP Cell Painting Consortium data

Alán F. Muñoz and Anne Carpenter, in behalf of the JUMP Cell-Painting Consortium

Broad Institute of Harvard and MIT

Abstract

With the Cell Painting assay we quantify cell morphology using six dyes to stain eight cellular components: Nucleus, mitochondria, endoplasmic reticulum, nucleoli, cytoplasmic RNA, actin, golgi aparatus, and plasma membrane. After high-throughput fluorescence microscopy, image analysis algorithms then extract thousands of morphological features from each single cell's image. By comparing of these "profiles" we can can uncover new relationships among genetic and chemical perturbations.

The JUMP-CP Consortium (Joint Undertaking for Morphological Profiling-Cell Painting) released the first public high-throughput dataset with over 140,000 genetic and chemical perturbations (Chandrasekaran et al. 2023).

Here, we describe how this data can now be used to answer many biological questions. Researchers can pick any gene of interest and find what morphological phenotypes are induced when it is knocked-out or overexpressed and what genes produce a similar morphological profile when altered, uncovering functional relationships. Novel software tools developed for this dataset empower biologists to make discoveries of their own, and we show that mining this dataset can yield novel insights into current and relevant biological questions.

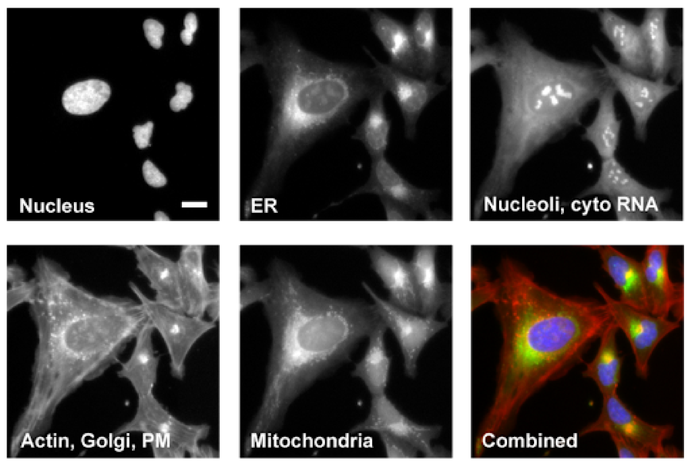
Goals

Device methods to interpret profile-based datasets to yield useful biological insight.

Develop a tool/workflow for biologists to discover genes that result in phenotypes similar to theirs.

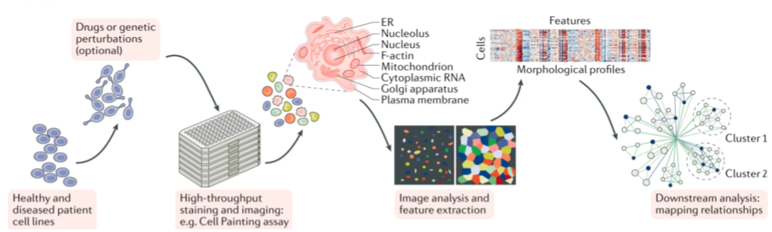
Build a stepping stone for a universal and accessible framework against which biologists can validate cell phenotypes.

We use data from the Cell Painting assay, in which eight cellular components are stained using six dyes and imaged in five channels



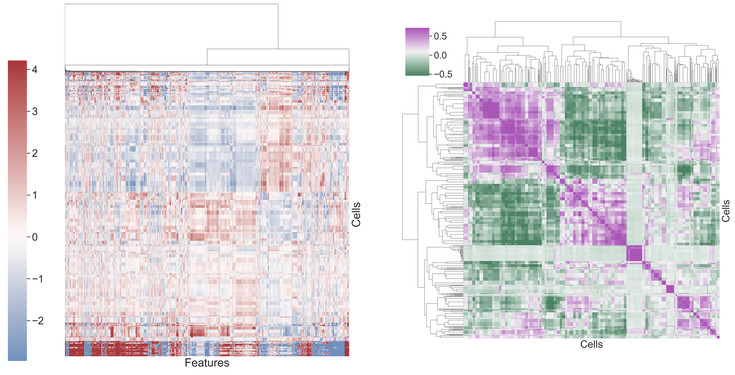
Morphological profiles were generated at a high-throughput scale

We generated and preprocessed a database composed of thousands of cell painting experiments.



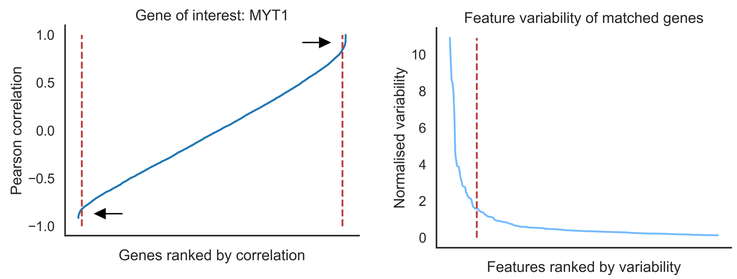
We generated a reference dataset for cells and features that indicates clustered groups of genes

After applying batch correction, it becomes possible to query individual genes and find similar profiles.



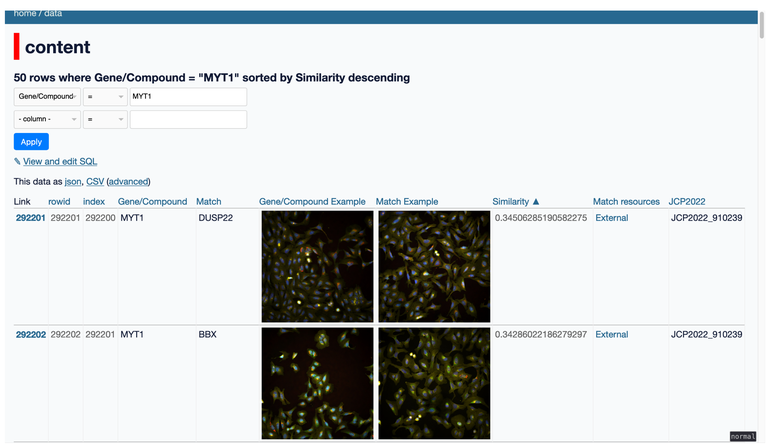
We then proceed to find morphological features that link genes closely

1. Find the most correlated and anticorrelated genes.
2. Find the features that show highest variance between these correlated/anticorrelated candidates.
3. Use these feature to guide comparisons between perturbed cells and negative controls.



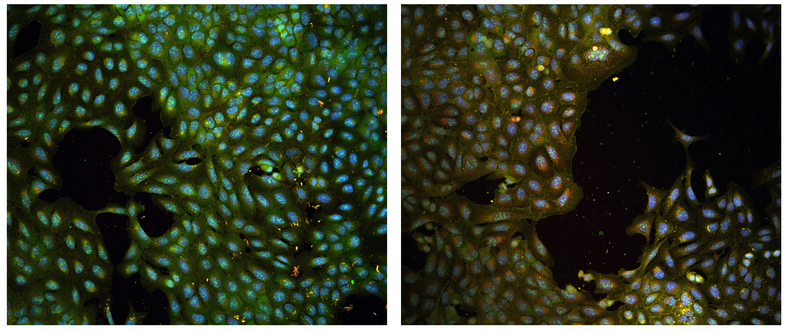
We provide web interface for data exploration

We made available multiple web assembly resources for scientist to access the morphology of genetic perturbations and other causes of similar morphologies.



Using these morphological features eases discovering novel insights

We compare images using tools that decompose the channels to focus on the most important features obtained from data mining



MYT1 deletion | Control

Ongoing research

Gene/compound	Phenotype/disease
MYT1/RNF41	Neuronal fate
MUC1	Cancer
PDE Inhibitors	Cancer treatment
CTDNEP1	Nuclear structure
MMP9	Alzheimer

Available resources

Dataset	Genes ranking	Features	Description
ORF	broad.io/orf	broad.io/orf_feature	Gene overexpression
CRISPR	broad.io/crispr	[WIP]	Gene knock-out
Compound	[WIP]	[WIP]	Chemical compounds

The data and tools for programatic and manual access to the data are made available so people can explore and train models (Chandrasekaran et al. 2021). Refer to broad.io/explore-jump for tools and broad.io/jump-cellpainting for more information.

Conclusions

The JUMP Cell Painting can serve as a resource to obtain candidate genes to find further insight on genes or proteins of interest.

High throughput analyses require biological expertise to provide novel insights, but provide an unprecedented opportunity to achieve a wholistic undersanding of the cell and human diseases.

Our querying systems can aid scientists to accelerate their biological discoveries by providing means to interpret features and listing genes with similar phenotypes

References

Chandrasekaran, Srinivas Niranj, Jeanelle Ackerman, Eric Alix, D. Michael Ando, John Arevalo, Melissa Bennion, Nicolas Boisseau, et al. 2023. "JUMP Cell Painting Dataset: Morphological Impact of 136,000 Chemical and Genetic Perturbations." bioRxiv. https://doi.org/10.1101/2023.03.23.534023.

Chandrasekaran, Srinivas Niranj, Hugo Ceulemans, Justin D. Boyd, and Anne E. Carpenter. 2021. "Image-Based Profiling for Drug Discovery: Due for a Machine-Learning Upgrade?" *Nature Reviews Drug Discovery* 20 (2): 145–59. https://doi.org/10.1038/s41573-020-00117-w.