



JUMP Info Hub

# Exploring gene function and morphology using JUMP Cell Painting Consortium data: The JUMP toolkit

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 apparatus, 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 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.

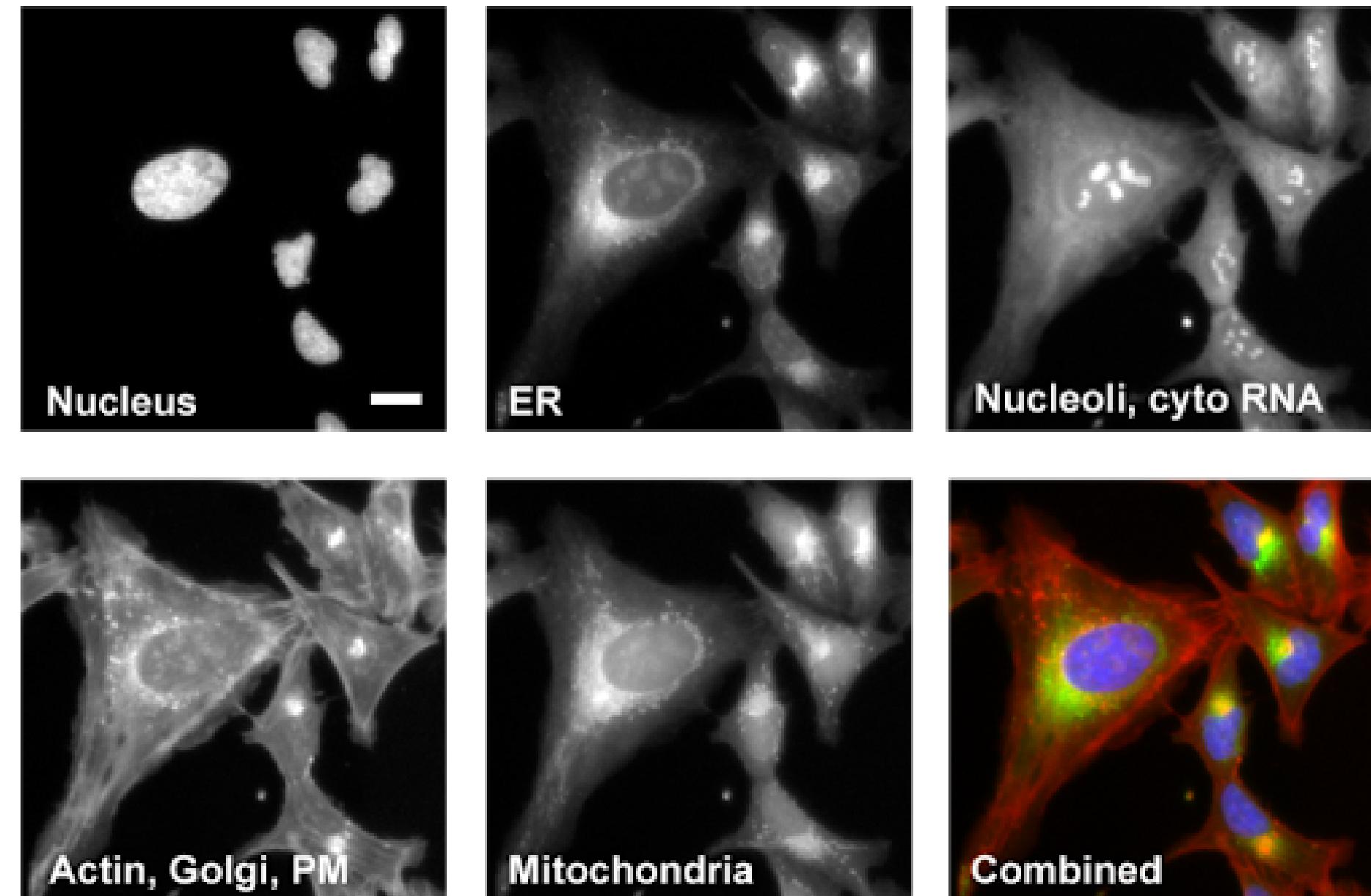
## Goals

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

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

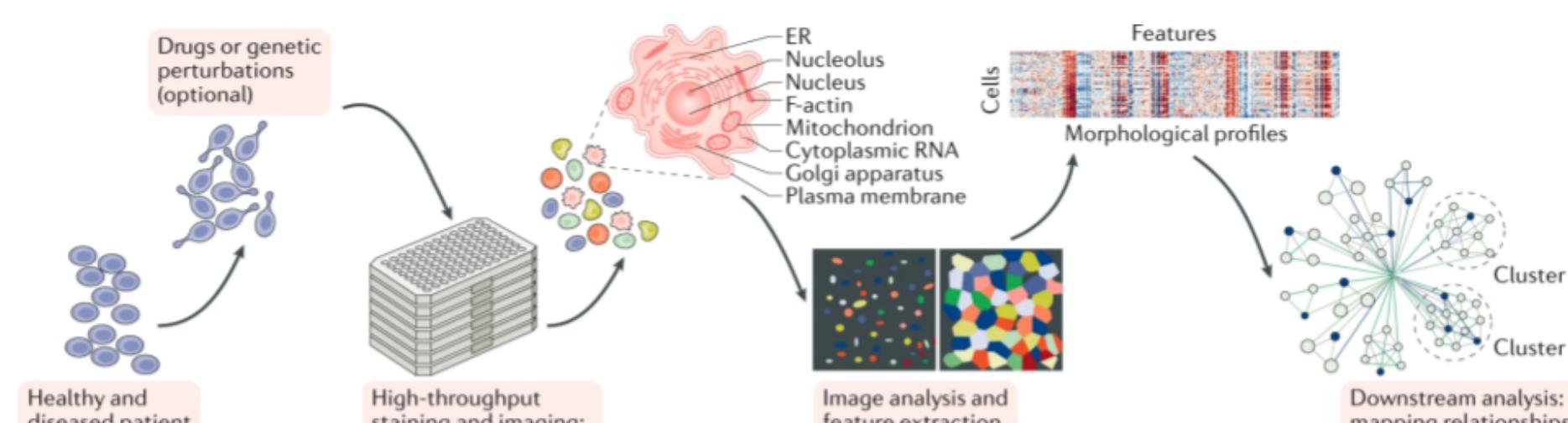
Build a stepping stone for a universal and accessible framework against which biologists can validate cell phenotypes.

We use data from the Cell Painting assay, in which 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.



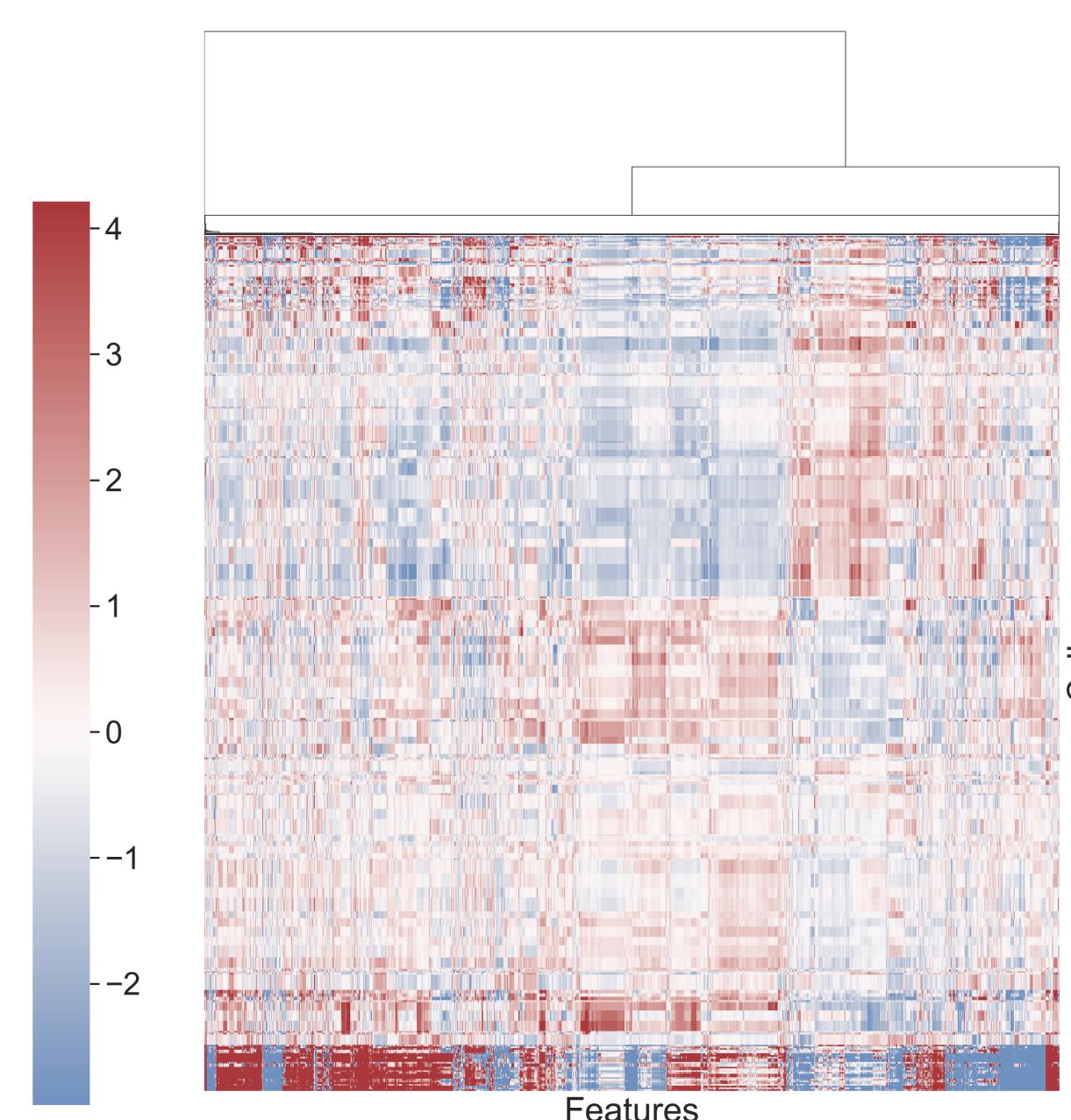
## QR code to this poster



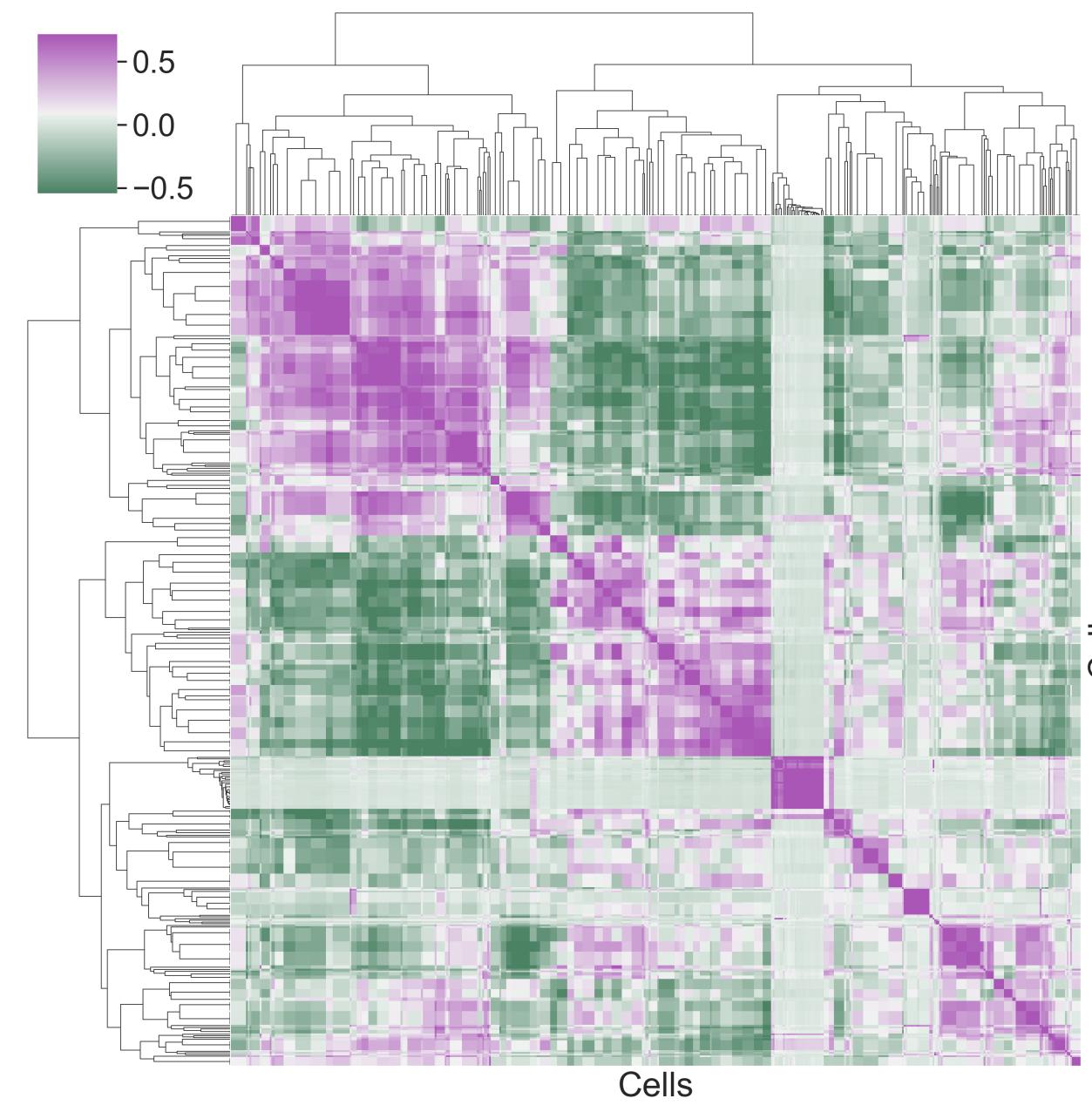
## We generated a reference dataset for cells and features that indicates clustered groups of genes

After applying batch correction, it becomes possible to query individual genes and find similar profiles. Precomputed distances for morphological profiles are made available.

## The JUMP consortium produced a massive set of morphological profiles

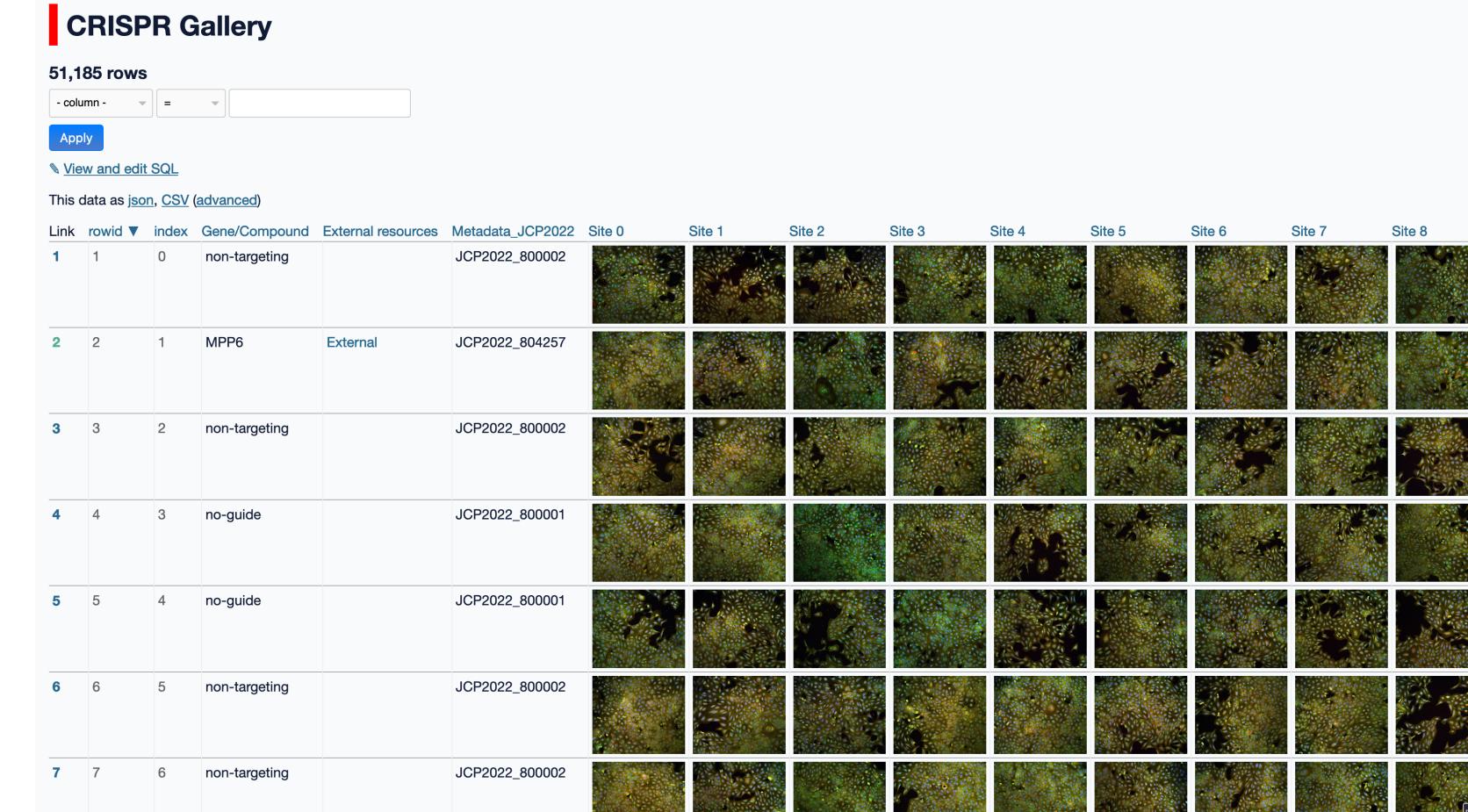


## We pre-calculated correlations between perturbations



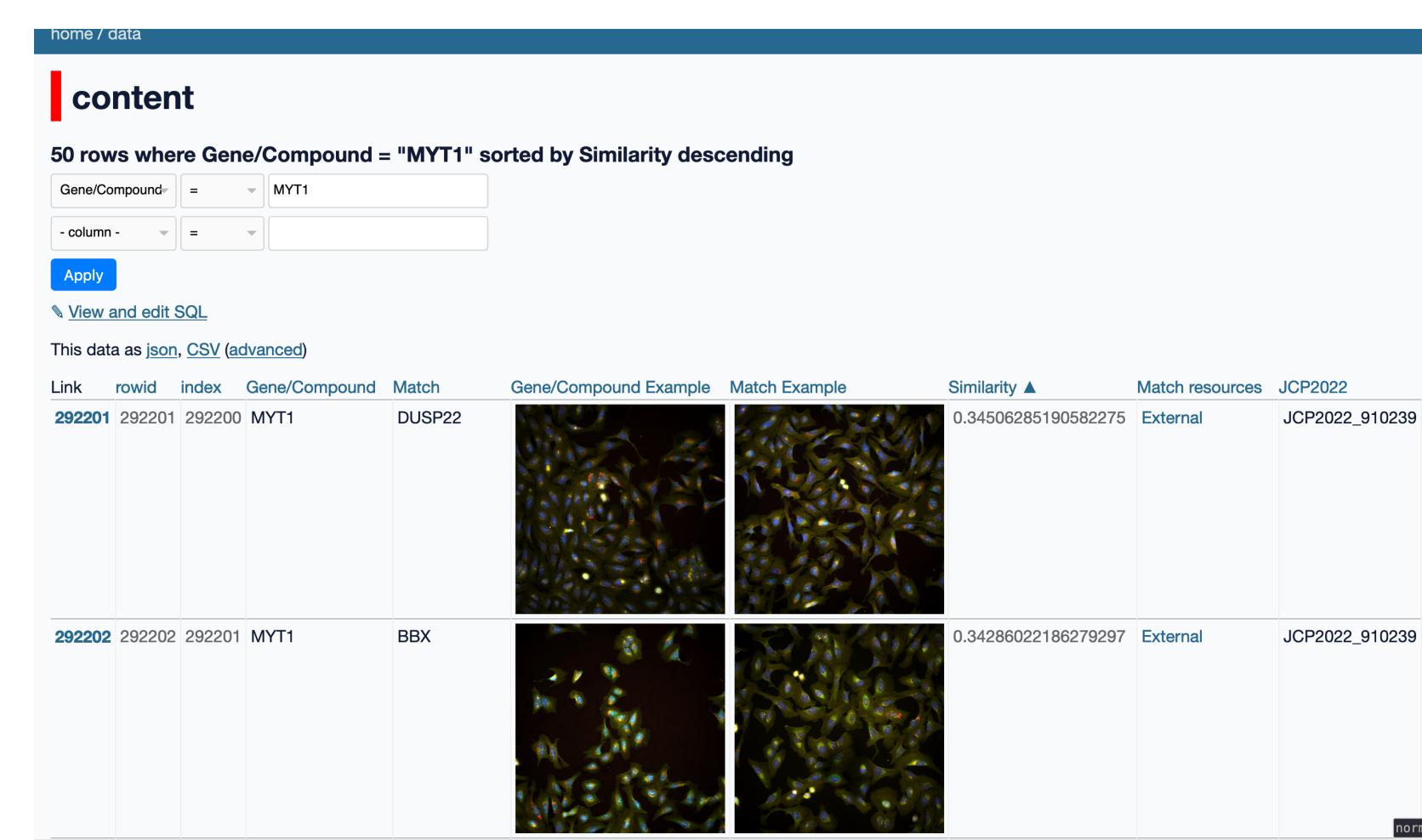
## We have also developed WebAssembly-based tools to access the data in multiple ways

A gallery to fetch all the available images for a given perturbation.



## Which other perturbations produce a phenotype similar to my gene of interest?

We developed an ecosystem tools for scientist to find the perturbations most similar to theirs.



## Which features are the most significant for my gene of interest?

Statistical values of all features for a given perturbation.

Link	rowid	Index	Cell region	Feature	Channel	Suffix	Feature significance	Median	JCP2022 ID	Gene/Compound example image	Phenotypic activity	Gene/Compound	Reso
1	1	0	Cells	AreaShapeArea	0.0645098819882662	-2.20339513244603	0.0001913973534894518	SP8	JCP2022_803801		0.0001913973534894518	SP8	NCB
2	2	1	Cells	AreaShapeArea	0.0881443545222803	0.455999903140033	0.0001913973534894518	SP8	JCP2022_803801		0.08475285780135741	LDHC	NCB
3	3	2	Cells	AreaShapeArea	0.01968881301581573	-1.942212839126597	0.0001913973534894518	SP8	JCP2022_800609		0.0001913973534894518	ABBS	NCB
4	4	3	Cells	AreaShapeArea	0.01259880042915173	-1.395169732160616	0.0001913973534894518	SP8	JCP2022_804862		0.0203395084320874	PORV4	NCB
5	5	4	Cells	AreaShapeArea	0.017951803042014	-0.25338030287216197	0.0001913973534894518	SP8	JCP2022_802076		0.0001913973534894518	EF402	NCB
6	6	5	Cells	AreaShapeArea	0.017951803042014	-0.25338030287216197	0.0001913973534894518	SP8	JCP2022_802076		0.0001913973534894518	EF402	NCB

## A standard analysis workflow has the following steps:

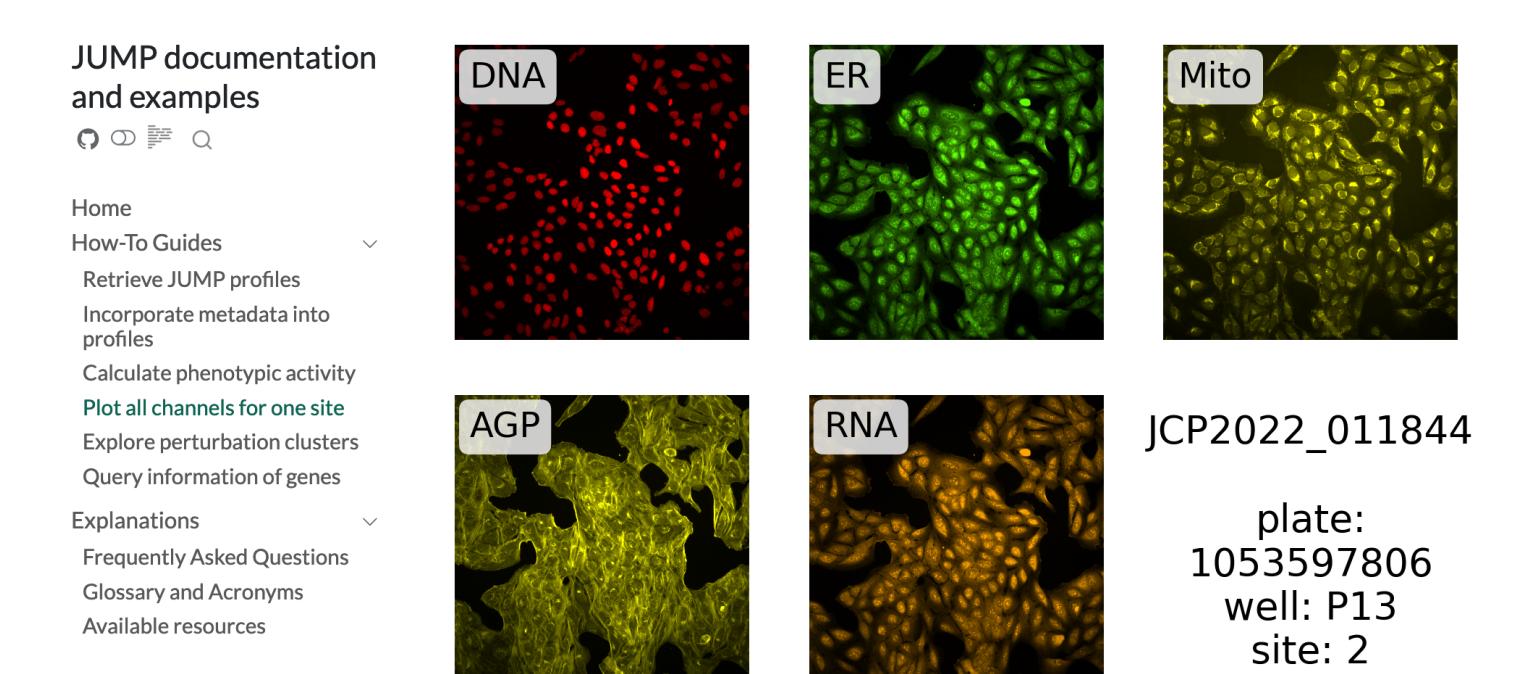
- Find the most correlated and anticorrelated genes.
- Find the features that show highest variance between these correlated/anticorrelated candidates.
- Use these feature to guide comparisons between perturbed cells and negative controls.
- Fetch images for these perturbations for inspection
- Retrieve additional annotations from existing databases.

## We provide libraries for data scientists and developers

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

- broad\_babel: Find the basic metadata for all perturbations
- jump\_portrait: Fetch images using perturbation identifiers.
- cp\_measure: is a new and experimental tool to extract CellProfiler features directly from images.

Examples of these tools and other workflows are available on the JUMP Hub.



## Available resources

Dataset	ORF	CRISPR
Description	Gene overexpression	Gene knock-out
Genes ranking	broad.io/orf	broad.io/crispr
Features	broad.io/orf_feature	broad.io/crispr_feature
Gallery	broad.io/orf_gallery	broad.io/crispr_gallery

## Conclusions

All data and tools for programmatic and manual access to the data are made available so people can explore and train models (Chandrasekaran et al. 2021). Refer to [broad.io/jump](https://broad.io/jump) for more information.

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

Our querying systems can help both biologists and data 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>.