# Mapping Biology With a Unified Representation Space for Genomic and Chemical Perturbations to Enable Accelerated Drug Discovery

*Aurora Blucher*, Safiye Celik*, James D. Jensen, James Taylor, Michael F. Cuccarese, Jacob C. Cooper, Jacob M. Rinaldi, Carl Brooks, Michael A. Statnick, Marta Fay, Nathan H. Lazar, Berton A. Earnshaw, Imran S. Haque*

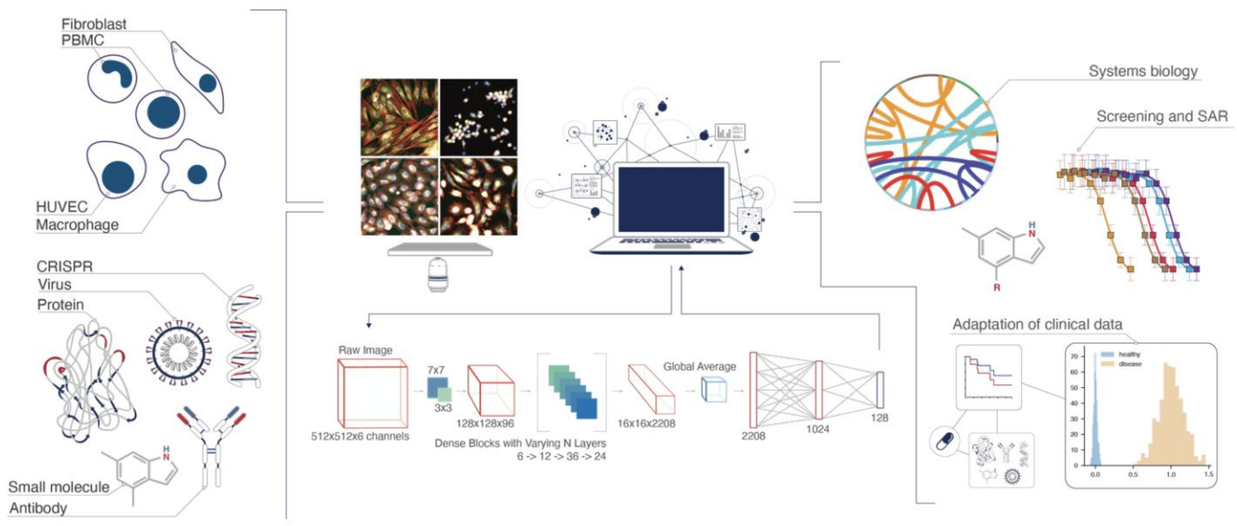
Recursion

## GENOTYPE TO PHENOTYPE: Recursion's arrayed screening platform captures a high-dimensional phenotypic readout from human cells at a massive scale.

### 1A. Genomic Tools for Drug Discovery

| Technology | Number of Genotypes | Number of Compounds | Dimensionality of readout |
|---|---|---|---|
| Single-target high-throughput screening | 1 protein target at a time | 1M+ | 1: Binary binding to target |
| DNA-Encoded Libraries | 1 protein target at a time | 1B+ | 1: Binary binding to target |
| Virtual drug screening | 1 protein target at a time | $10^6$-$10^9$ | 1: Binary binding to target |
| Pooled CRISPR scRNA | 20k+ | 1 per reaction | <20k |
| Protein design/evolution | 10k+ | 1 at a time | 1: Binary binding to target |
| Structural binding prediction (AlphaFold 2) | 100M + | 0 | 3D structure |
| RECURSION | ~18,000 | ~700k so far | 25M pixels reduced to 128D |

### 1B. Phenomics Screening Platform Overview




Various cell types (top left) are treated with a range of biological perturbants and treatments (bottom left), including CRISPR-based genetic modifications and small molecules.

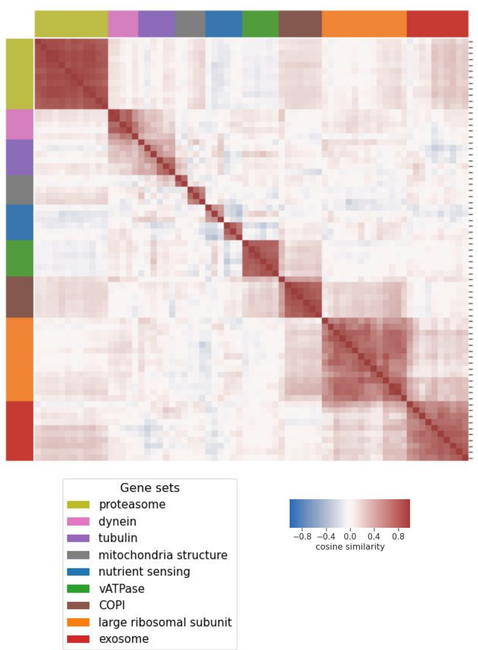
High-throughput fluorescence microscopy (middle-top) and deep-learning-enabled image featurization (middle-bottom) generate high-dimensional phenoprints that are used for interrogating a range of experimental questions.

Vector representations of millions of multi-channel fluorescence microscopy images generated using a proprietary analytics workflow based on an extension of a DenseNet-161 are analyzed (right) to map out gene-gene and gene-compound relationships, including protein complex membership, pathway regulation, target identification, and structure-activity relationship (SAR).

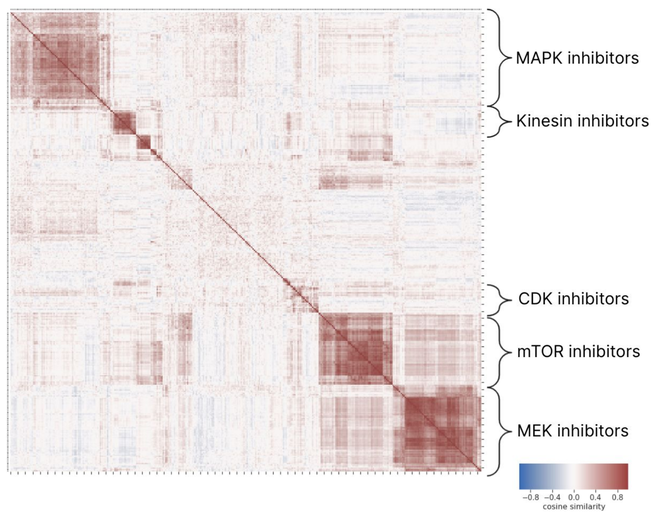
## PHENOTYPE TO GENOTYPE: Representations of gene knockouts (KO) and compounds reflect known and novel biology.

### 2A. Clustering gene-gene phenoprints recapitulates canonical biological pathway and gene sets.

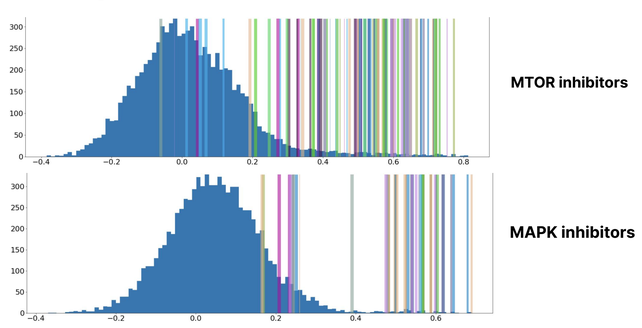
- Gene KO phenoprints are hierarchically clustered according to cosine similarity.
- Proteasome members form a distinct and strong phenosimilar cluster (yellow).
- Larger shared cluster across large ribosomal unit (orange) and exosome (red), which points to known collaborative functionality between those biological processes.




### 2B. Clustering gene-compound phenoprints captures known modes of mechanism of action and groups compounds together based on shared mechanism.

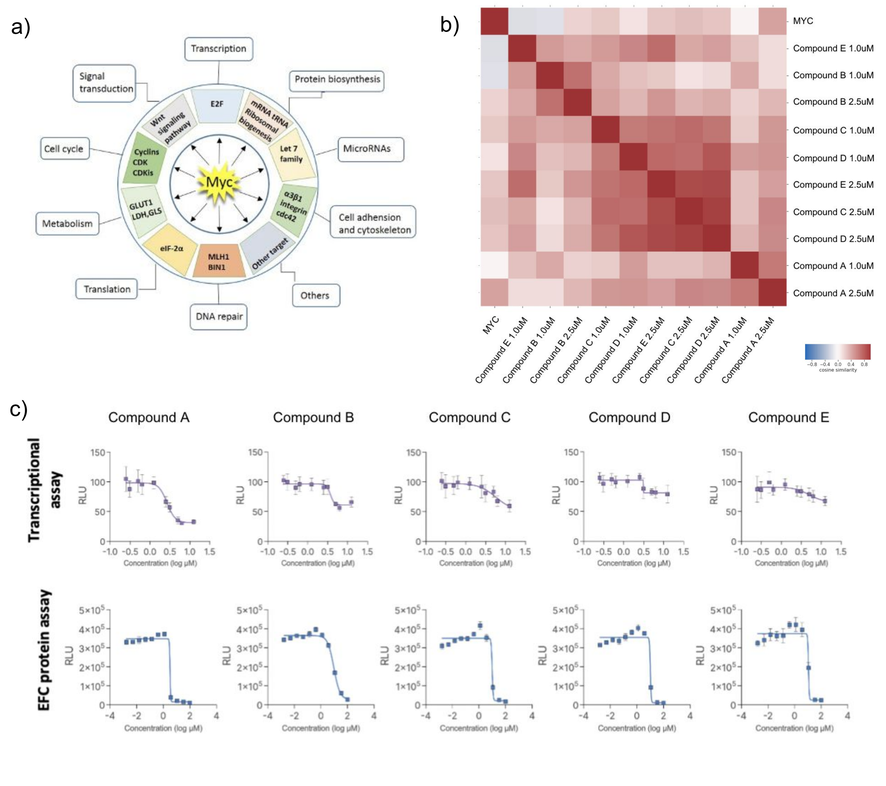



### 2C. Compounds are strongly cosine similar to the KO of their known targets.




### 2D. Platform identifies hits for classically undruggable targets.

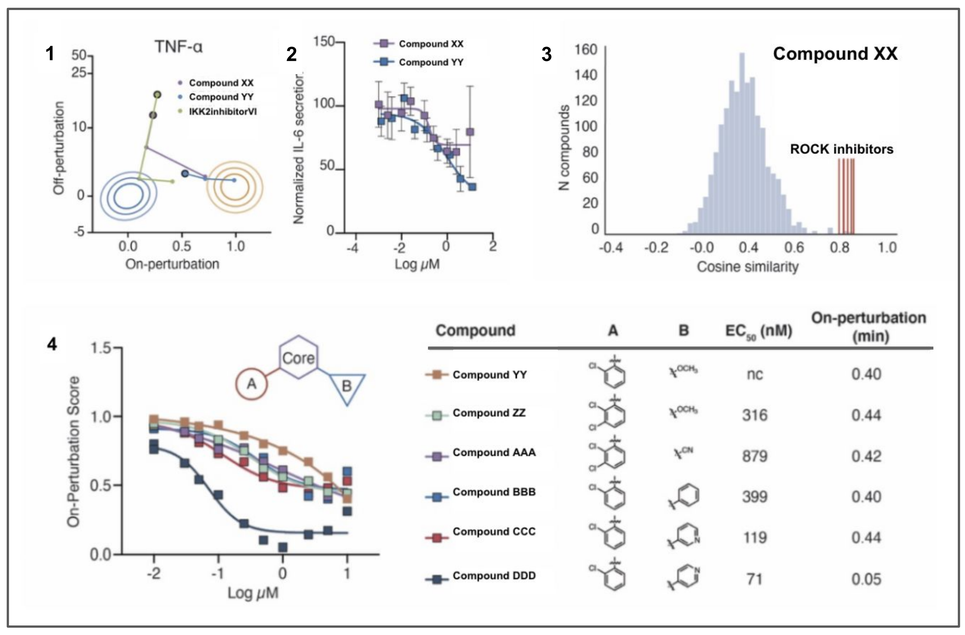
a) Gain-of-function alterations and amplifications in MYC have been identified in more than 50% of human cancers. MYC has remained an important undruggable target in oncology for decades.
b) New chemical entity hit molecules are identified through phenotypic similarity of library compounds to MYC.
c) Identified hits show verified activity in MYC transcriptional assay and c-MYC EFC protein turnover assay.



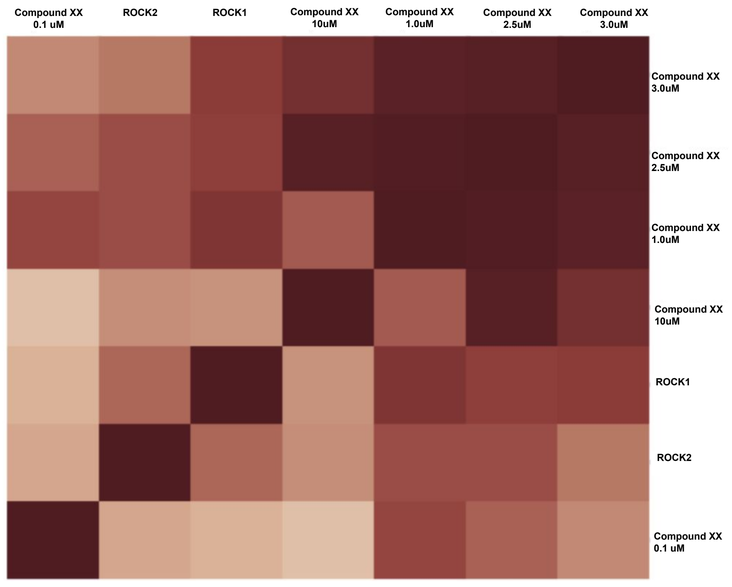

## PHENOTYPE TO CHEMOTYPE: Representations can be used to direct chemical space search in drug development.

### 3A. Hits identified through rescue screen translate to meaningful biological endpoints and direct targeted chemical searches.

1. Projections of compound response in the context of perturbation vector for TNF-α in HUVEC.
2. IL-6 secretion (HTRF) from HUVEC treated with 1 ng/mL TNF-α in the presence of Compound XX and Compound YY.
3. Distribution of cosine similarity of phenoprints of an annotated compound library to that of Compound XX. Red lines highlight ROCK inhibitors.
4. Projection of on-perturbation scores and EC50 values for each peripheral modification to the scaffold core (mean, n=6).




### 3B. Platform screened gene knockouts of ROCK1/2 show strong similarity to NCEs identified in rescue screens and targeted chemical search.






**References:**
Preprint: https://www.biorxiv.org/content/10.1101/2020.08.02.233064v2.full.pdf
S1 filing: https://www.sec.gov/Archives/edgar/data/1601830/000119312521089610/d89478ds1.htm
Myc figure: https://www.nature.com/articles/s41392-018-0008-7/figures/1