

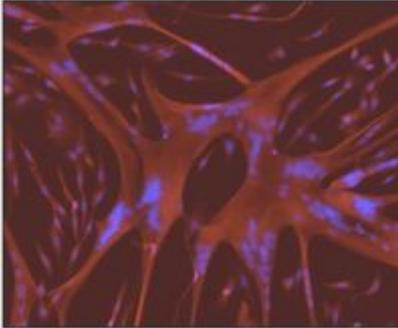
Image-based cell phenotyping with deep learning

Juan C. Caicedo PhD
Schmidt Fellow



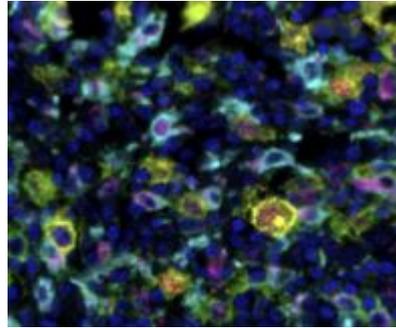
Images can be quantified for all kinds of phenotypes

Muscle structure



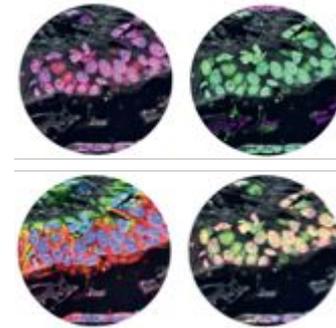
David Thomas

Patient biopsy tissue



Margaret Shipp/Scott Rodig

Image Mass Spec

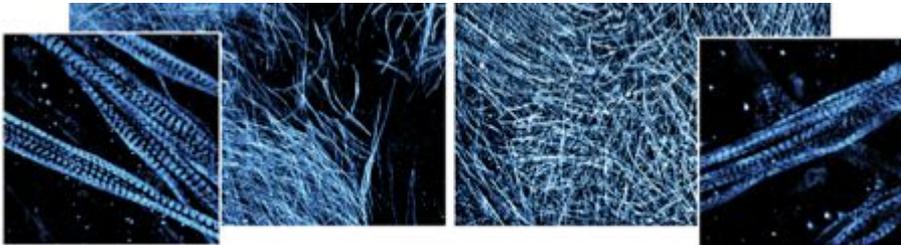


Michael Angelo

Muscle structure

Control human iPS

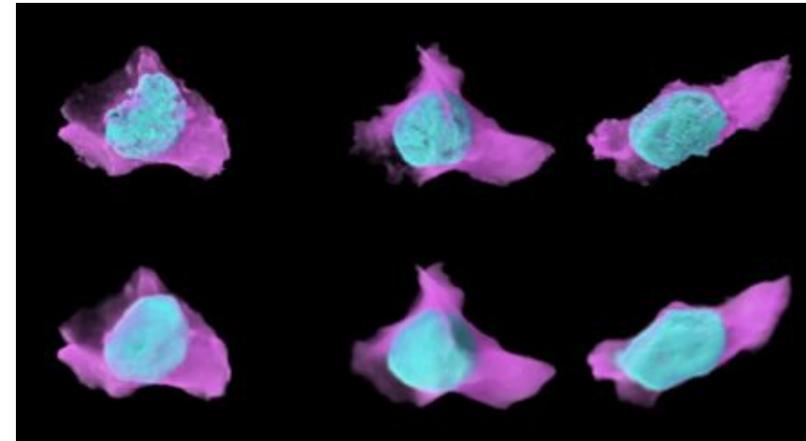
Isogenic Duchenne-like iPS



Olivier Pourquie

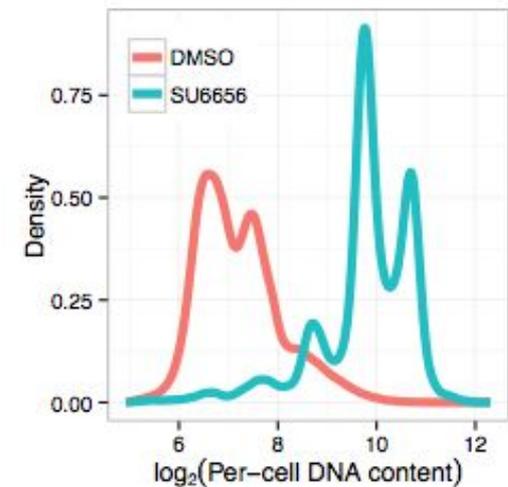
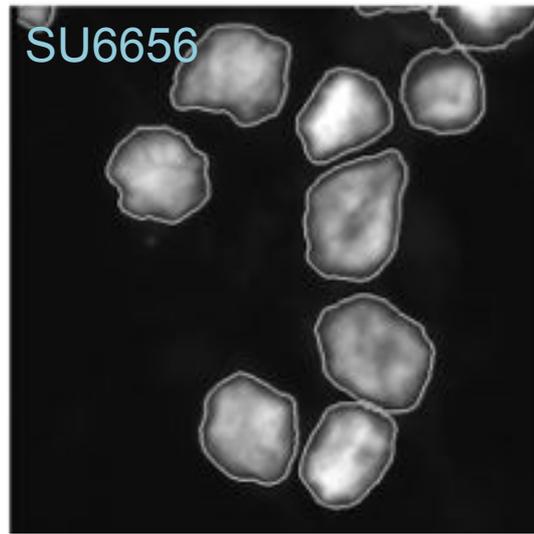
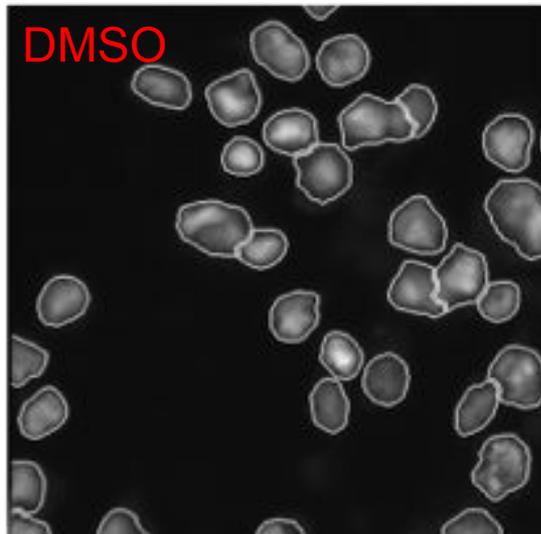
3D

Allen Institute for Cell Science



Screen for specific phenotypes using images

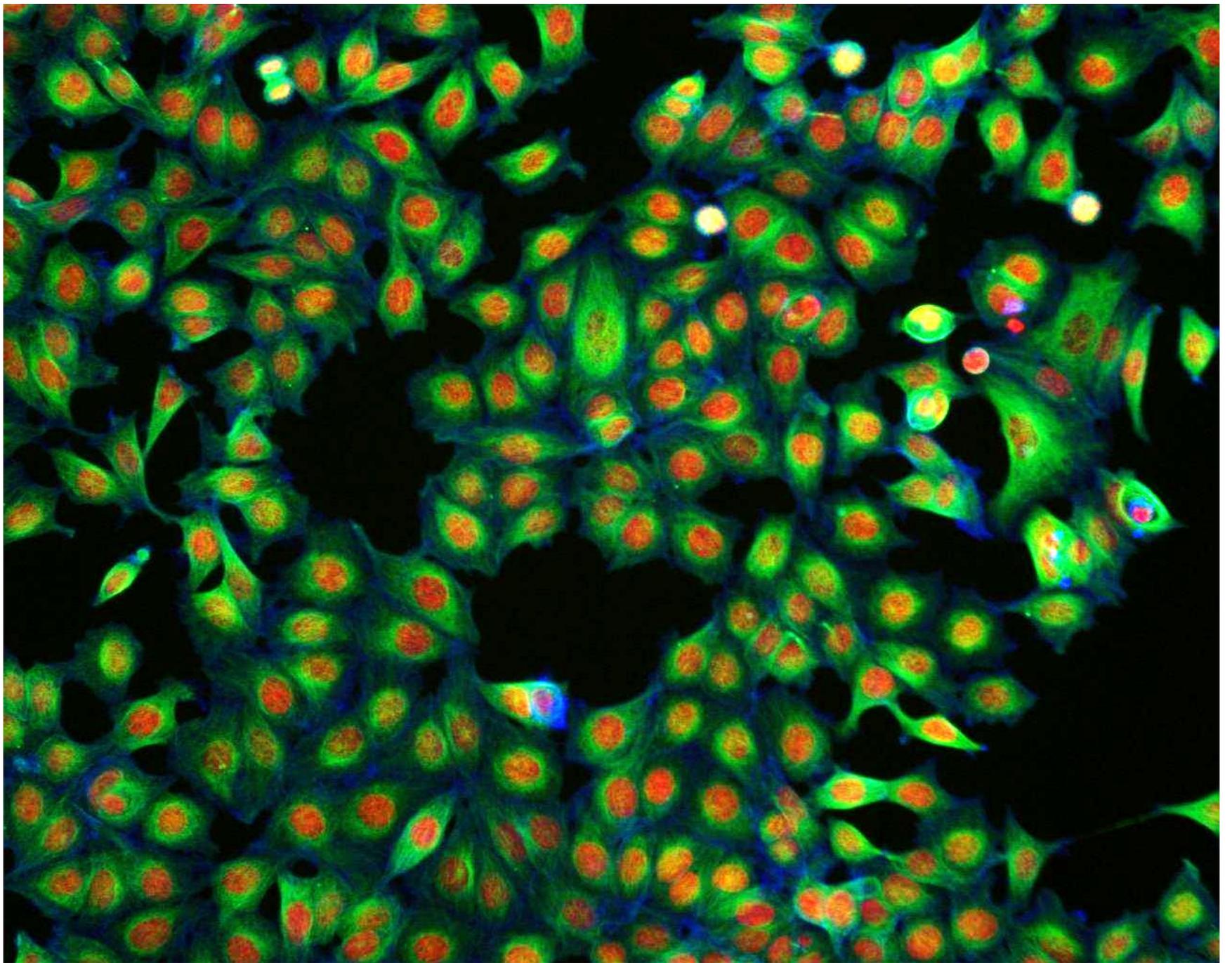
Treatment for AMKL (leukemia)

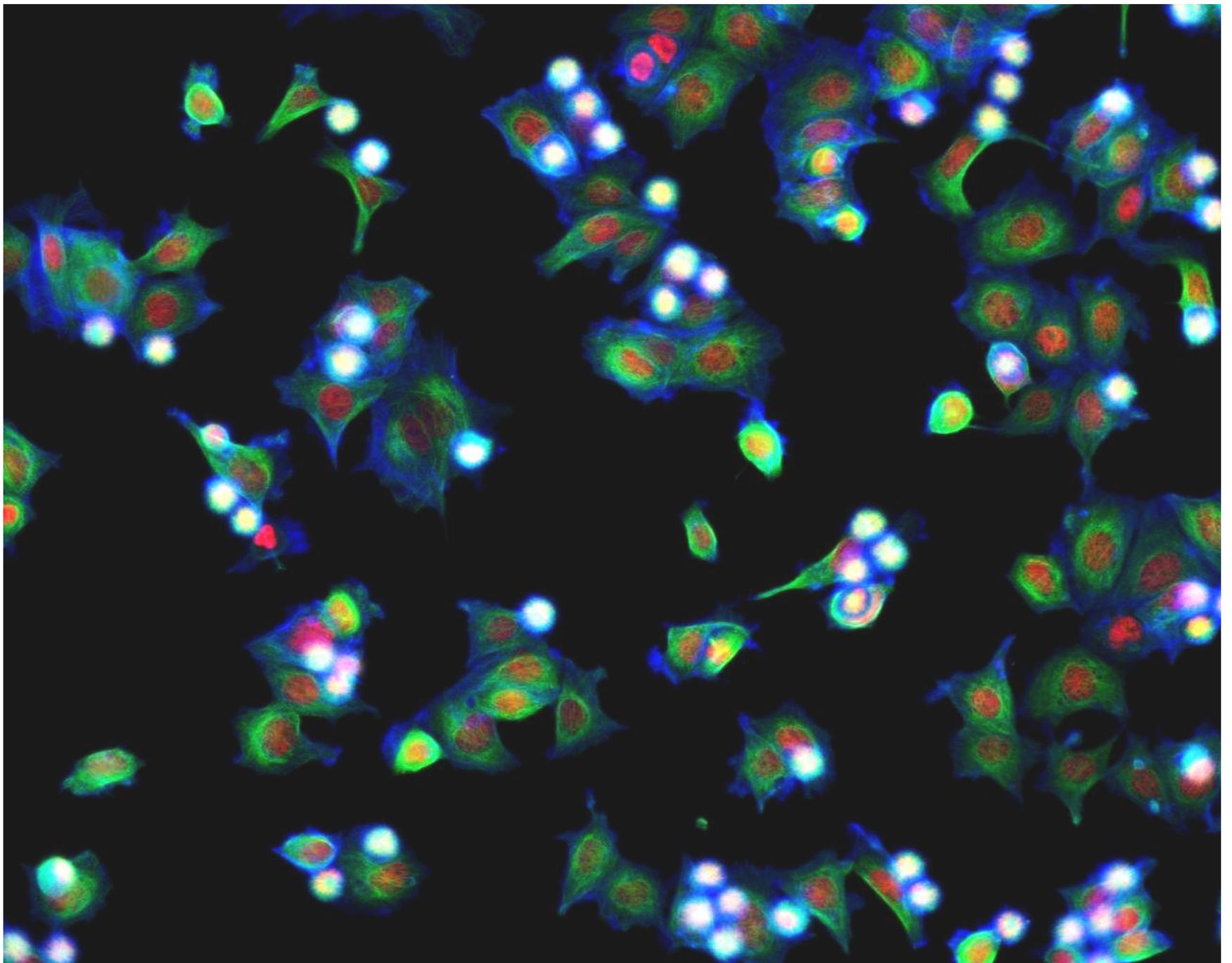


DNA stain with outlines identifying the nuclei

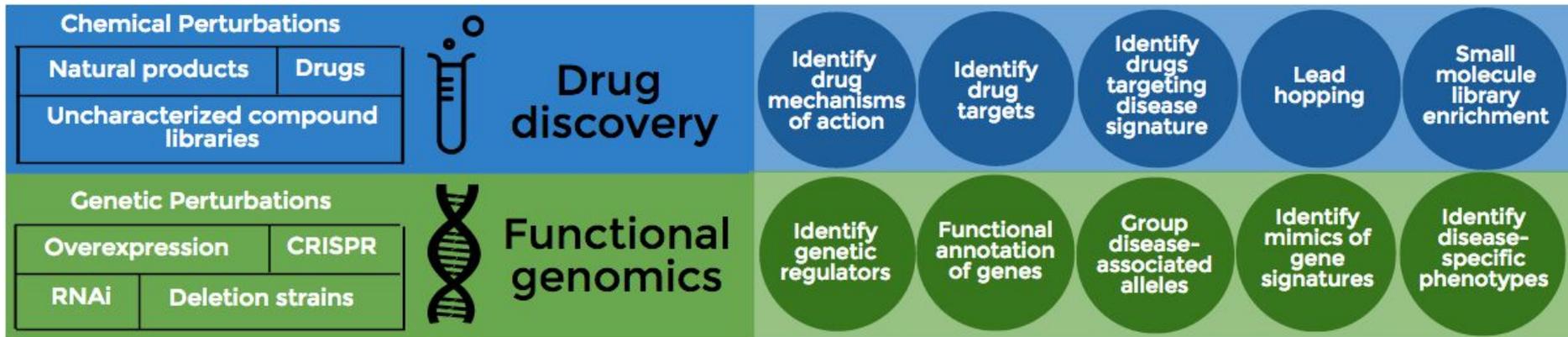
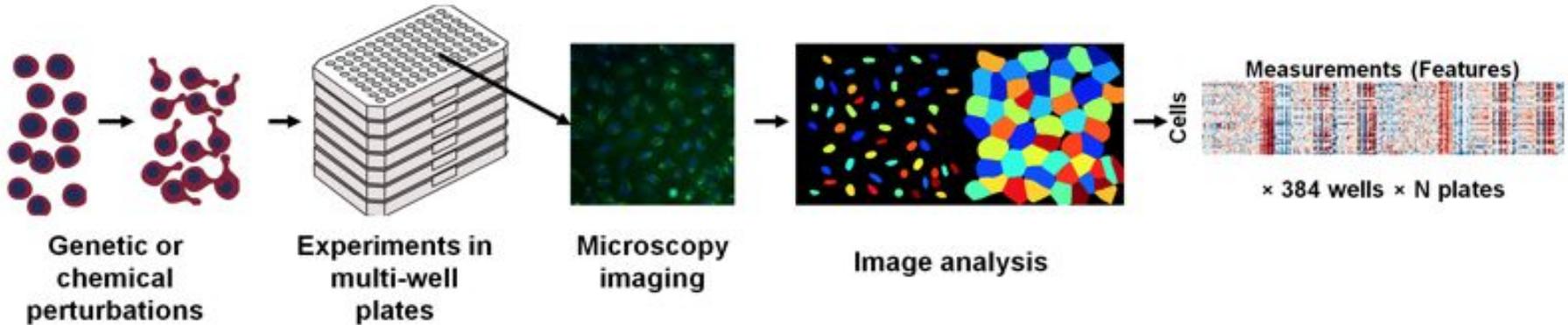
Clinical trials underway for Alisertib in adults with AMKL.

Wen Q, et al. (2012). Cell 150(3):575-89





What is image-based profiling?



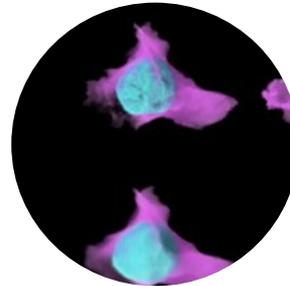
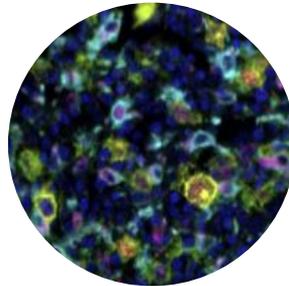
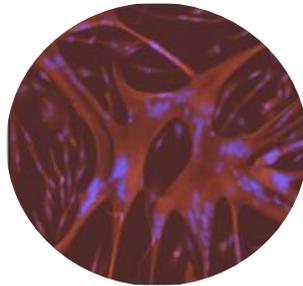
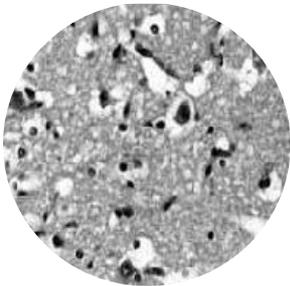
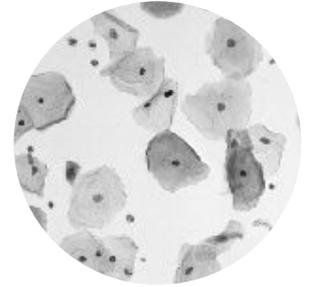
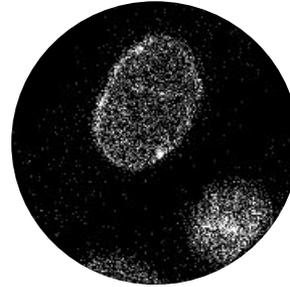
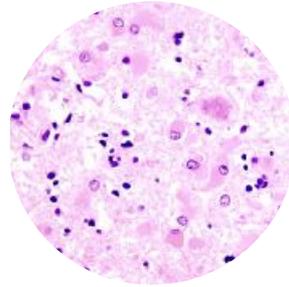
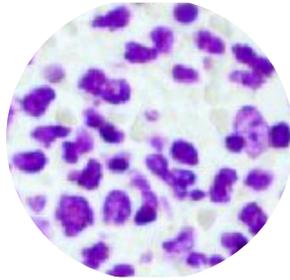
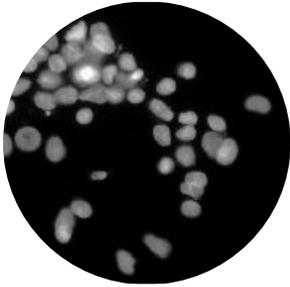
Caicedo J.C., Singh S., Carpenter A. "Applications of Image-Based Profiling of Perturbations". *Current Opinion in Biotechnology* - 2016.

ML for image-based profiling - Overview

1. Cell segmentation
2. Single-cell representation learning

1. Cell segmentation

Diversity of cell segmentation problems



Cell segmentation as face detectors



Dataset and challenge

Featured Prediction Competition

2018 Data Science Bowl

Find the nuclei in divergent images to advance medical discovery



Passion. Curiosity. Purpose.

\$100,000

Prize Money

Booz
Allen

Booz Allen Hamilton · 739 teams · 8 months ago

Presented by

Booz | Allen | Hamilton & kaggle

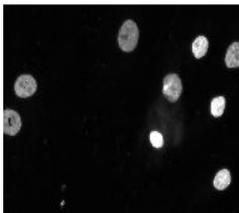


3
months

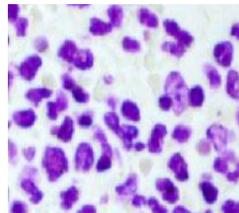
3,634
teams

65,333
experiments

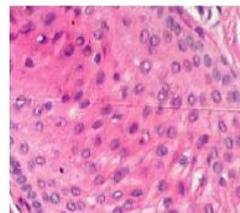
Small
fluorescent



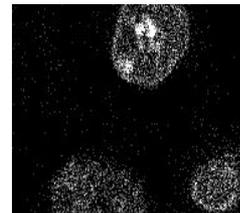
Purple
tissue



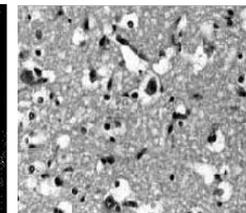
Pink and
purple tissue



Large
fluorescent

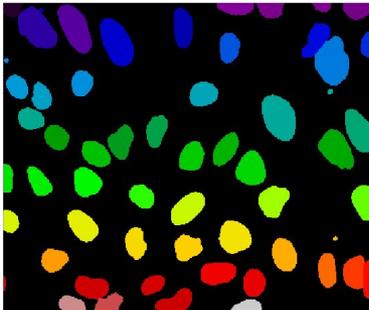
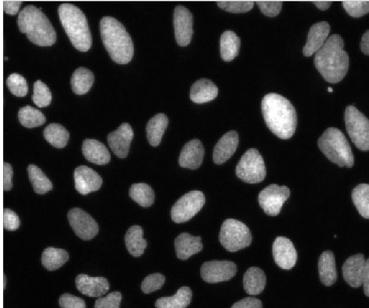


Grayscale
tissue

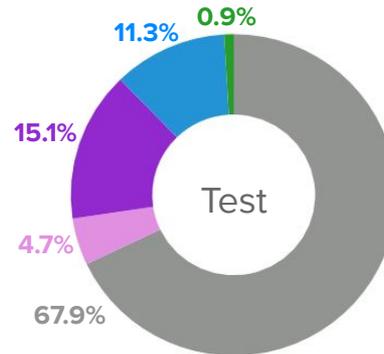
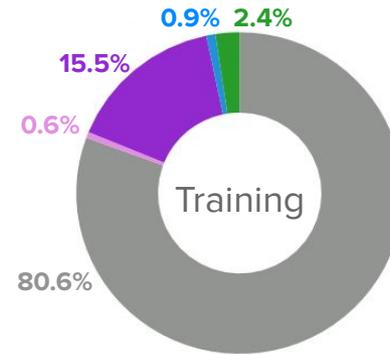


Data Science Bowl — Organization

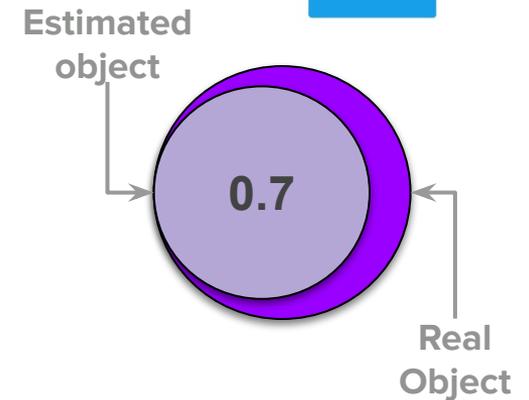
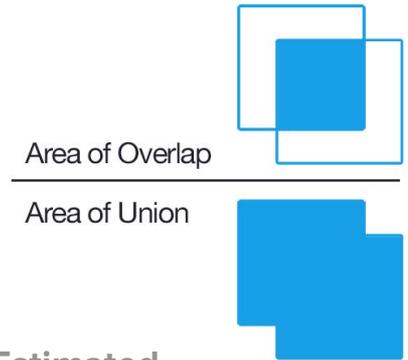
Create annotated dataset



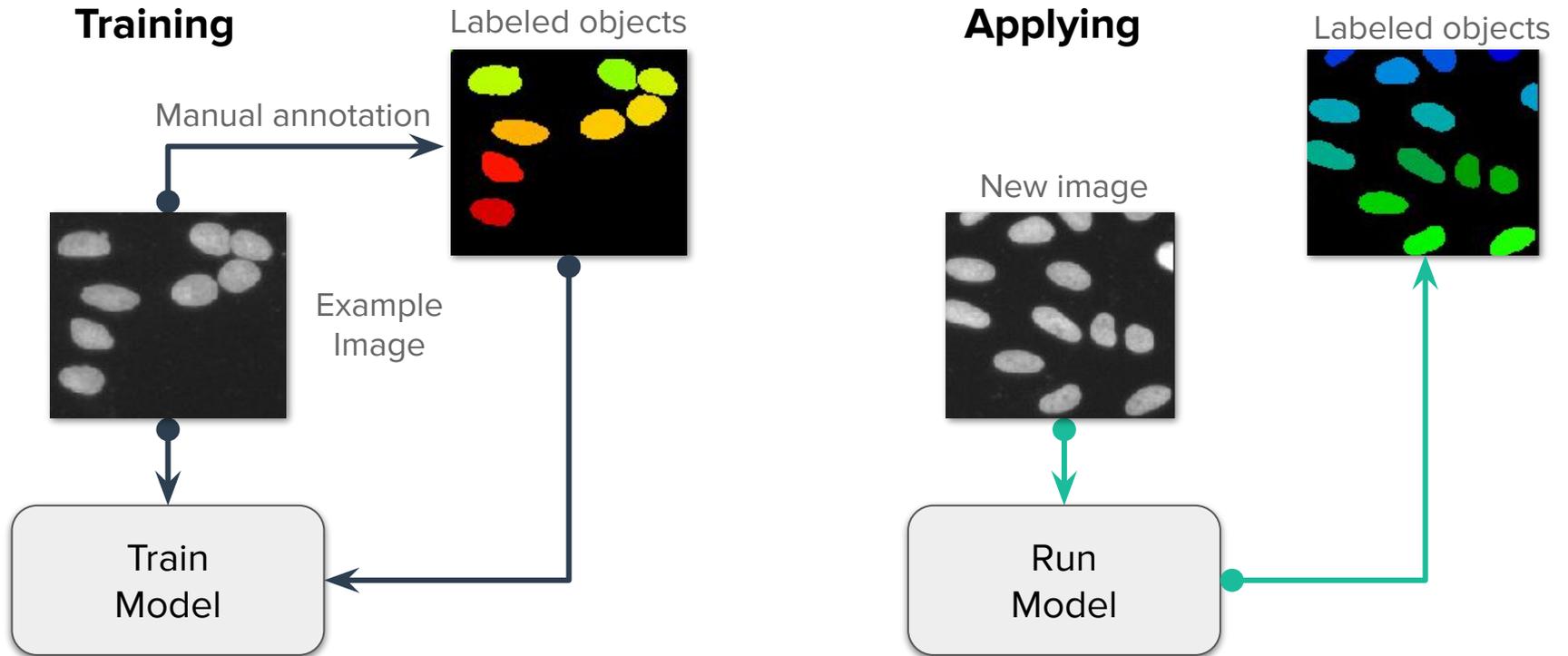
Split training and test



Define metric of success



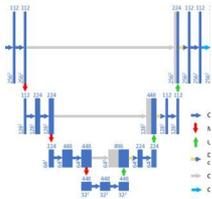
Training models for segmentation



Diversity of models for image segmentation

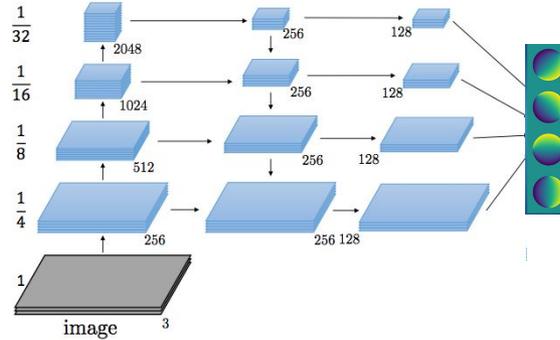
4 folds, 8 TTAs (flips/rotations)

- A DPN-92-UNet-softmax
- A ResNet34-UNet-softmax
- V InceptionResnetV2-UNet-softmax
- V DenseNet121-UNet-Softmax
- S DenseNet169-FPN-Softmax
- S ResNet152-FPN-sigmoid
- S ResNet101-FPN-sigmoid
- A DPN-92-UNet-sigmoid



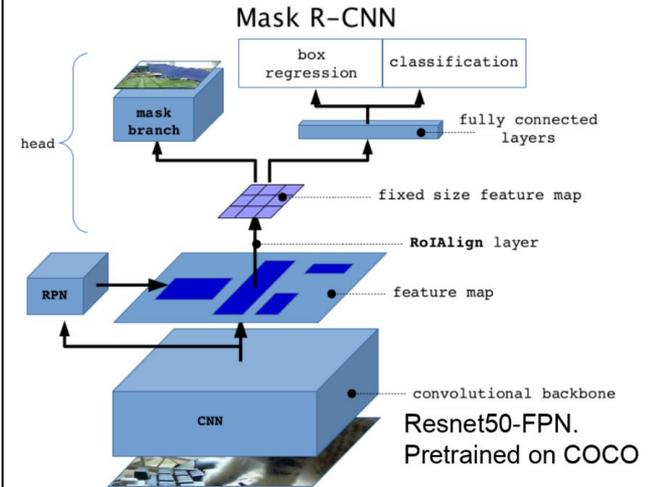
[ods.ai] topcoders
1st place

U-Nets



jacobkie
2nd place

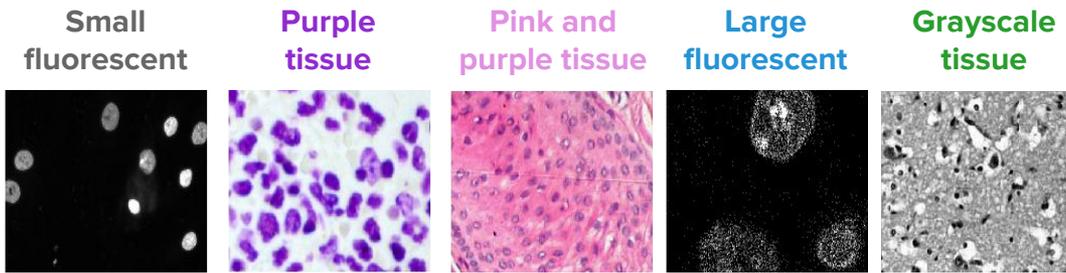
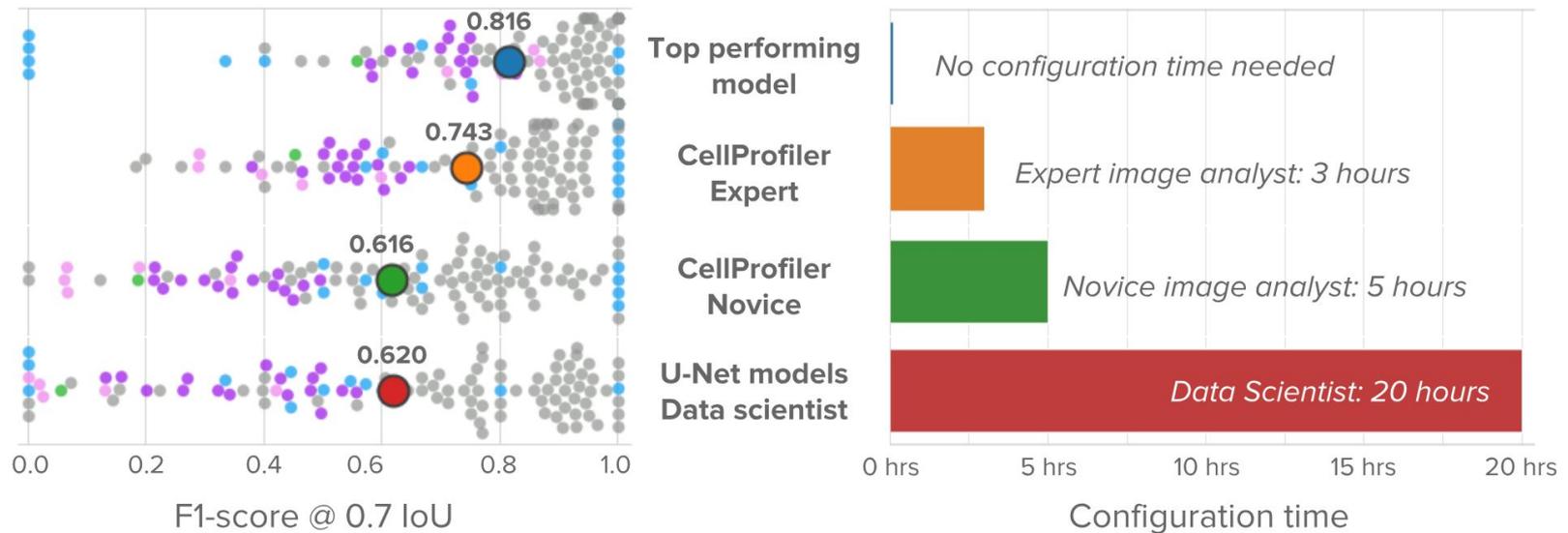
Feature Pyramid Nets



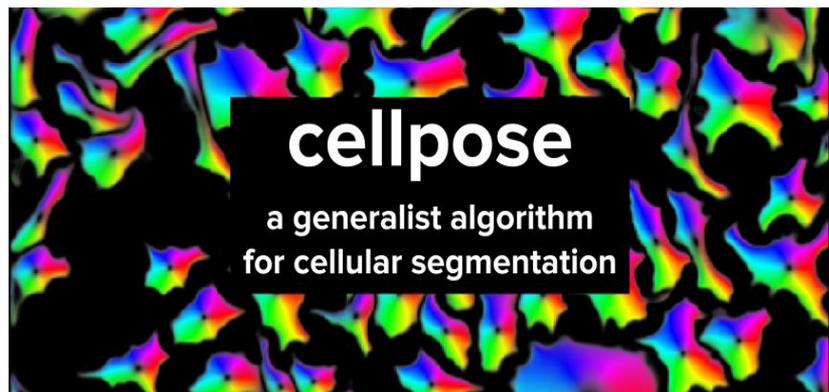
Deep Retina
3rd place

Mask RCNN

A single model improves accuracy and reduces time

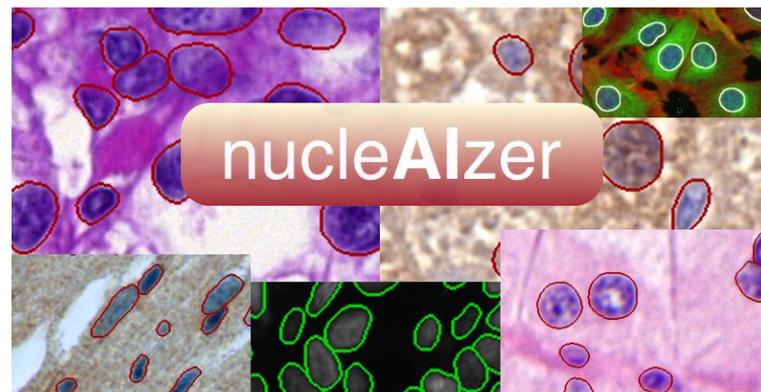


Existing tools for generic cell segmentation



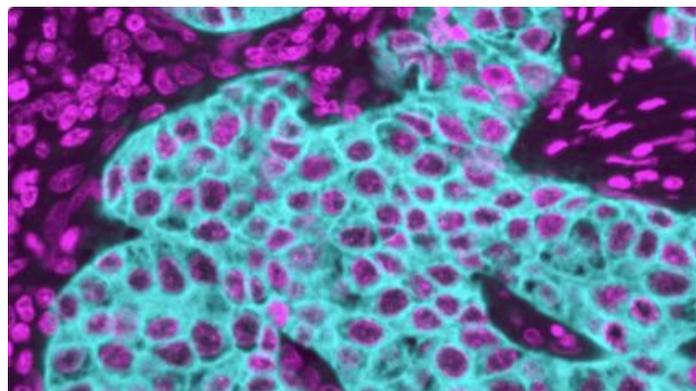
Stringer et al. 2020, Nat Meth

U-Nets



Hollandi et al. 2020, Cell Systems

Mask RCNN



Greenwald et al. 2021, BioRxiv

Feature Pyramid Nets

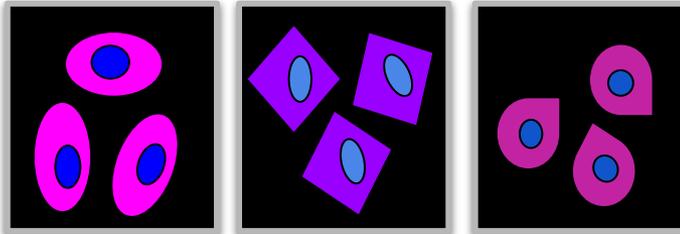
Open challenges

- Collecting larger publicly available datasets
- Learning from few examples with active learning

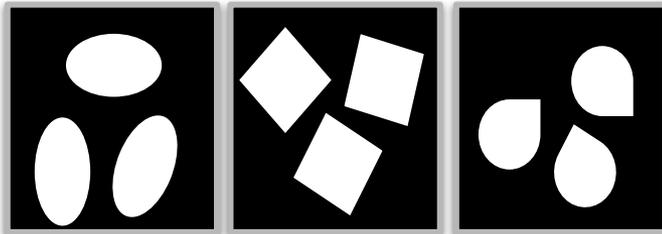
2. Single cell representation learning

Representation learning for image-based profiling

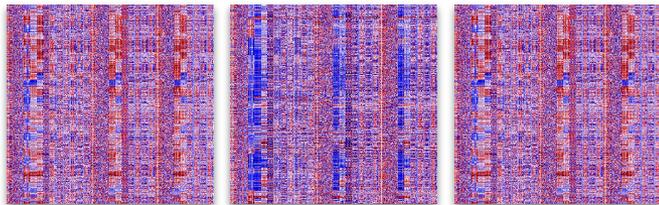
1. Raw images



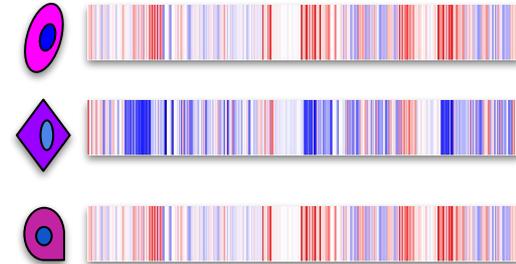
2. Segmented images



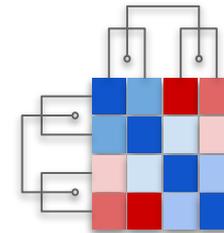
3. Single-cell feature matrices



4. Population profiles of treatments



5. Downstream statistical analysis

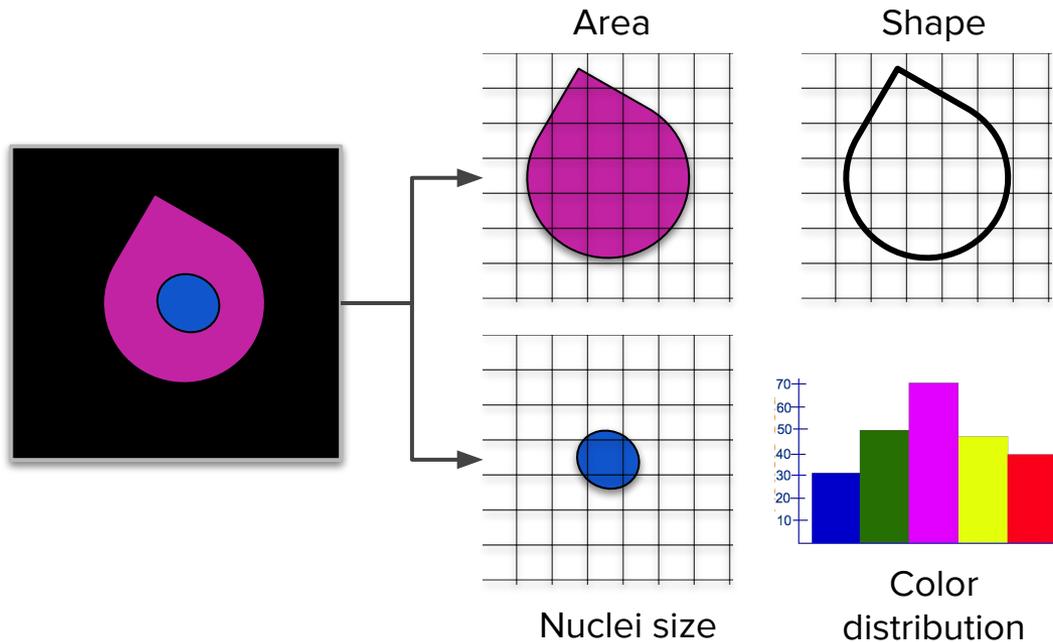


Are treatments significantly different / effective?

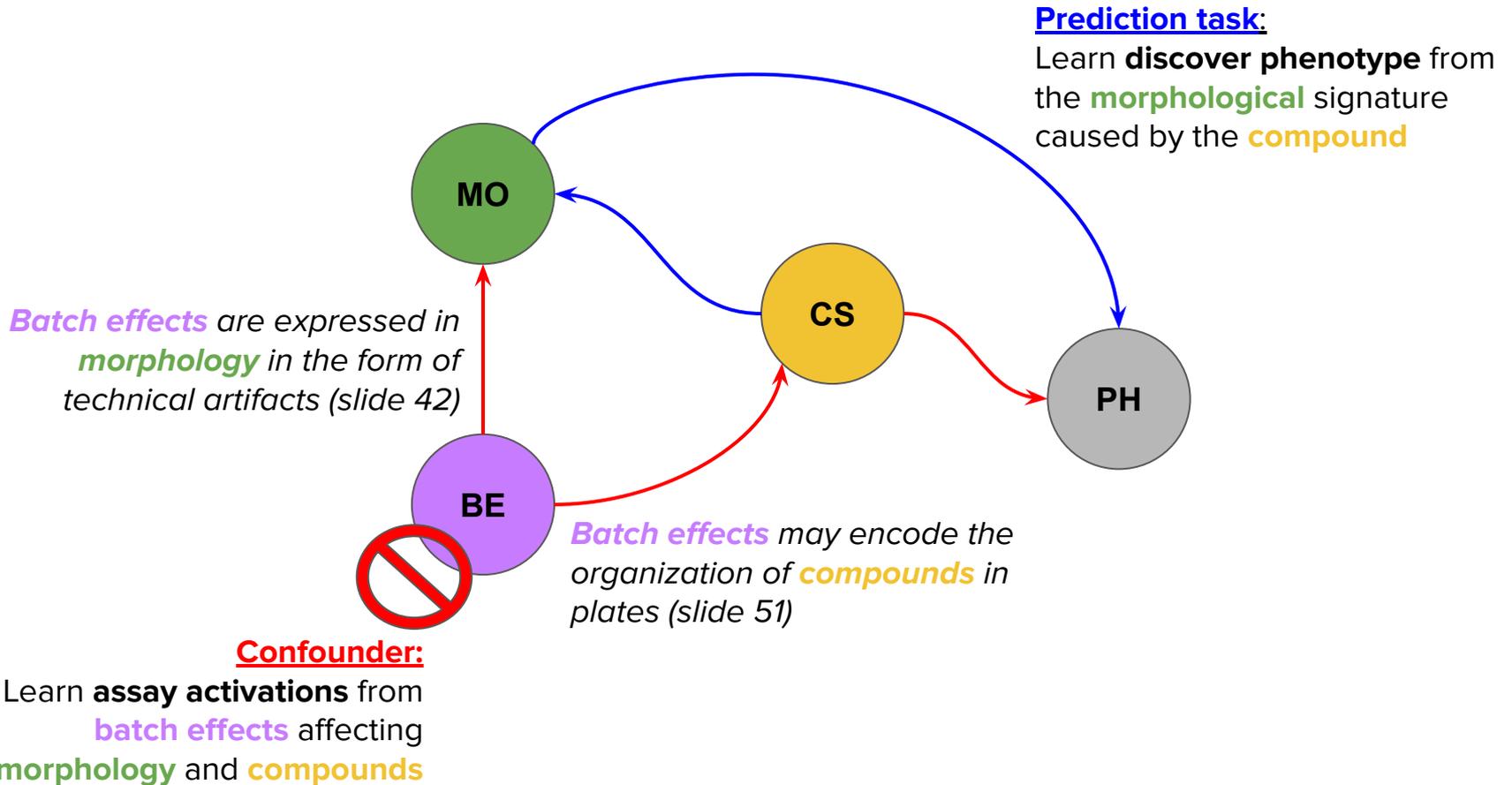
Classical approach to measuring cell morphology

Engineer measurements

Define and compute useful properties



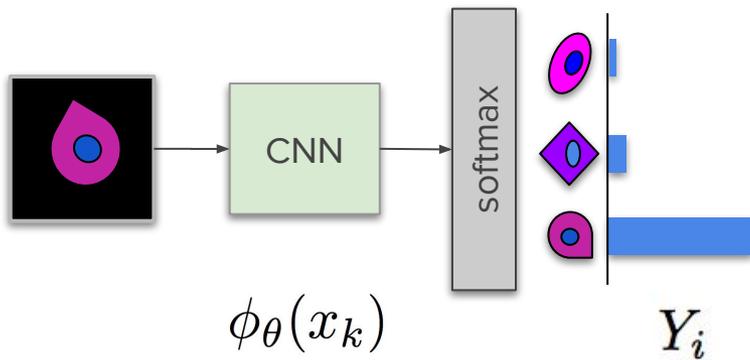
Perturbation experiments



Weakly supervised learning of single-cell feature embeddings

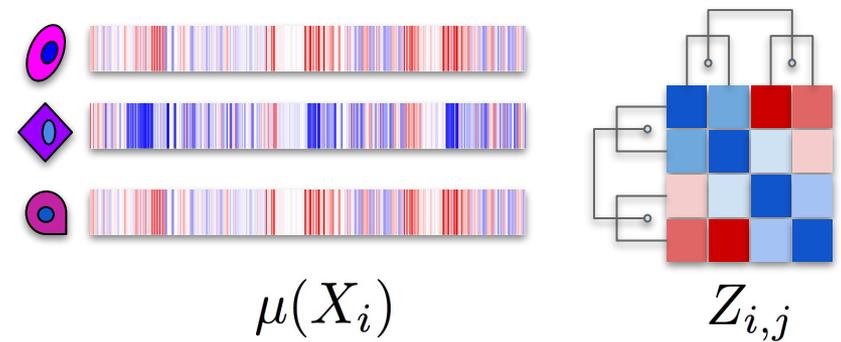
Auxiliary task:

Single-cell treatment classification

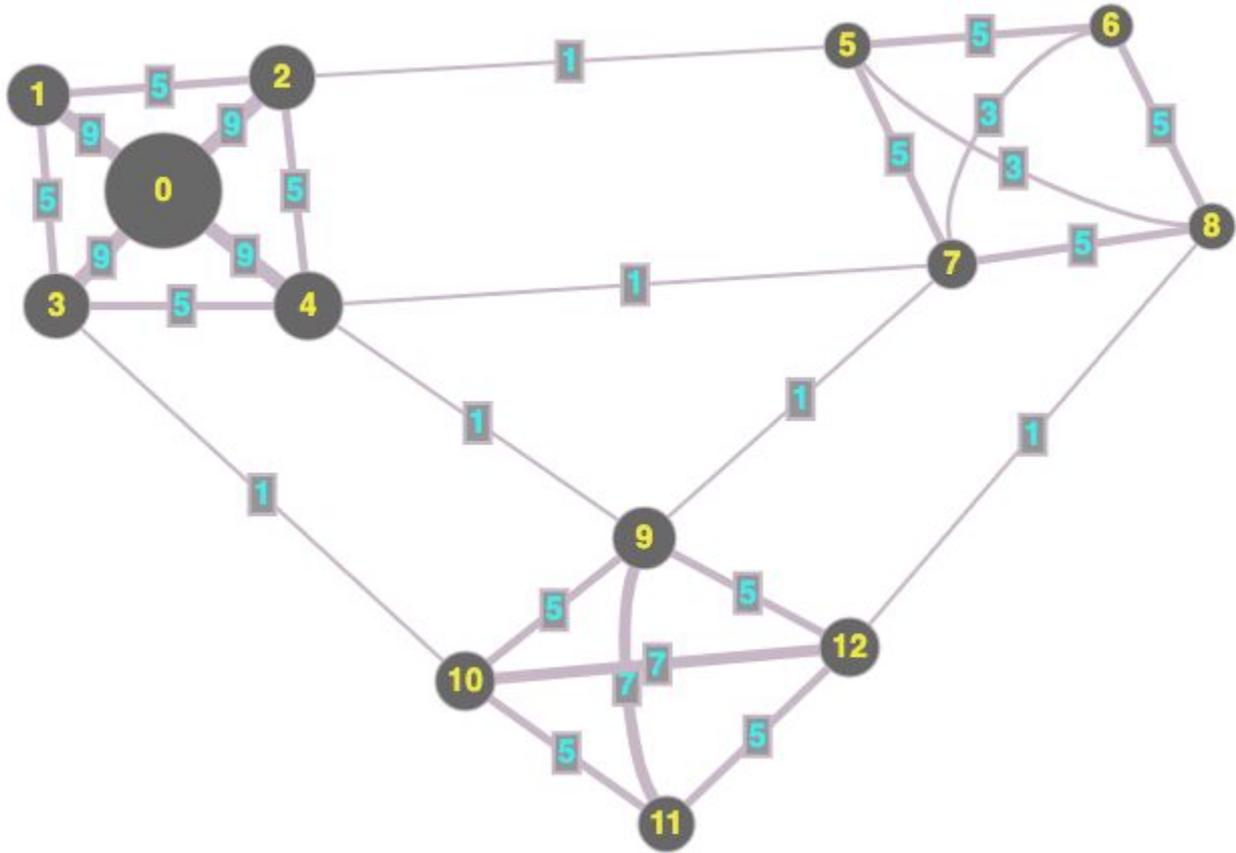


Main goal:

Treatment-level profiling



Evaluating Image-Based Profiling



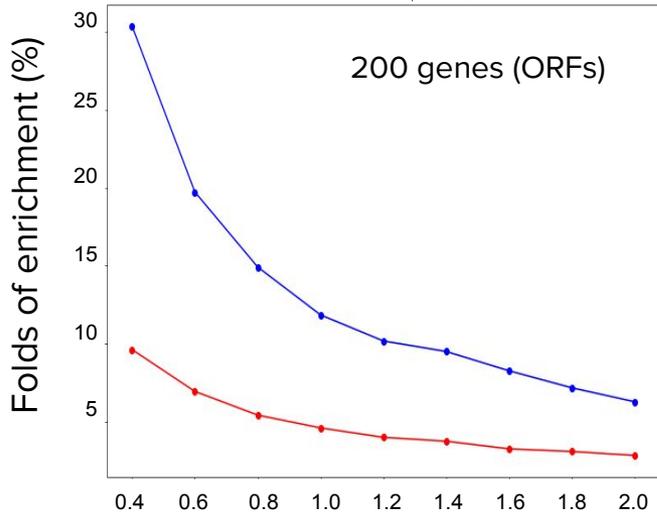
13 data points

78 pairwise connections

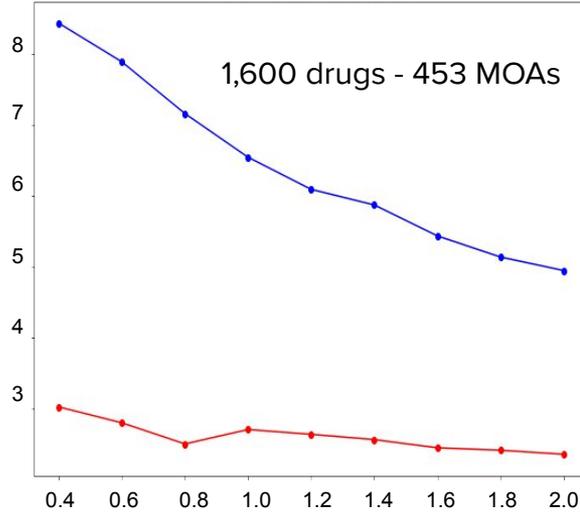
Learned representations improve profiling performance

Number of folds of enrichment for top connections to have the same MOA/pathway vs. rest of the connections

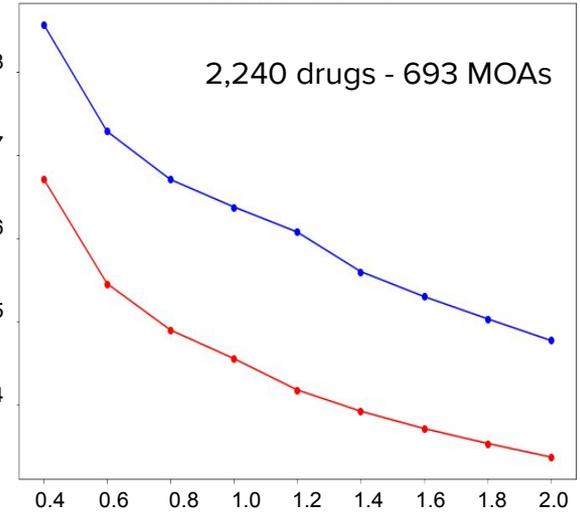
 Gene over-expression
BBBC037



 Compound screen
BBBC022



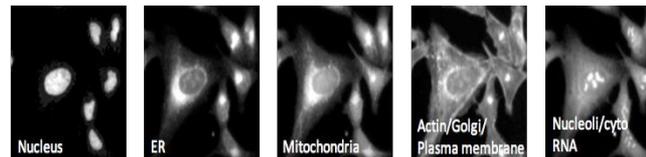
 Compound screen
BBBC036



% top connections

Deep learning

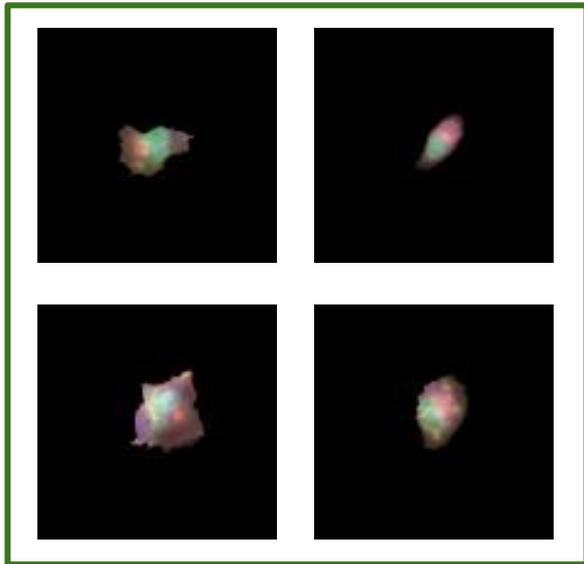
Classical features



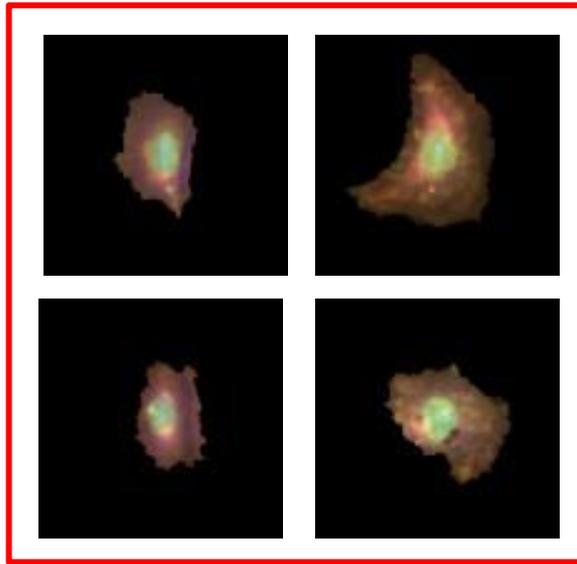
Determining variant impact

EGFR_p.S645C

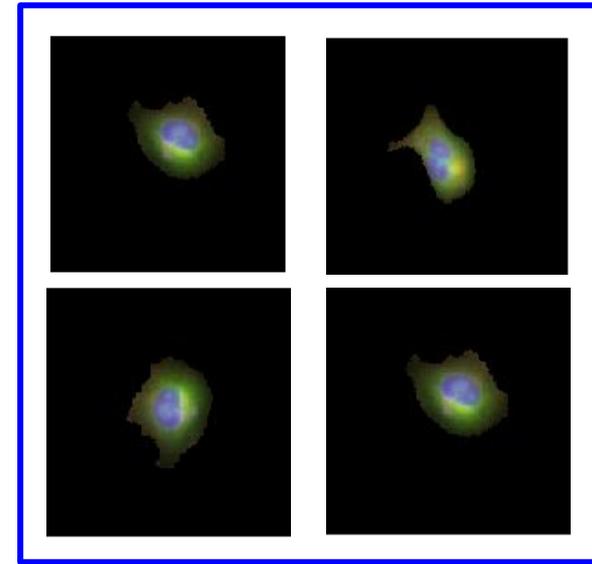
Control



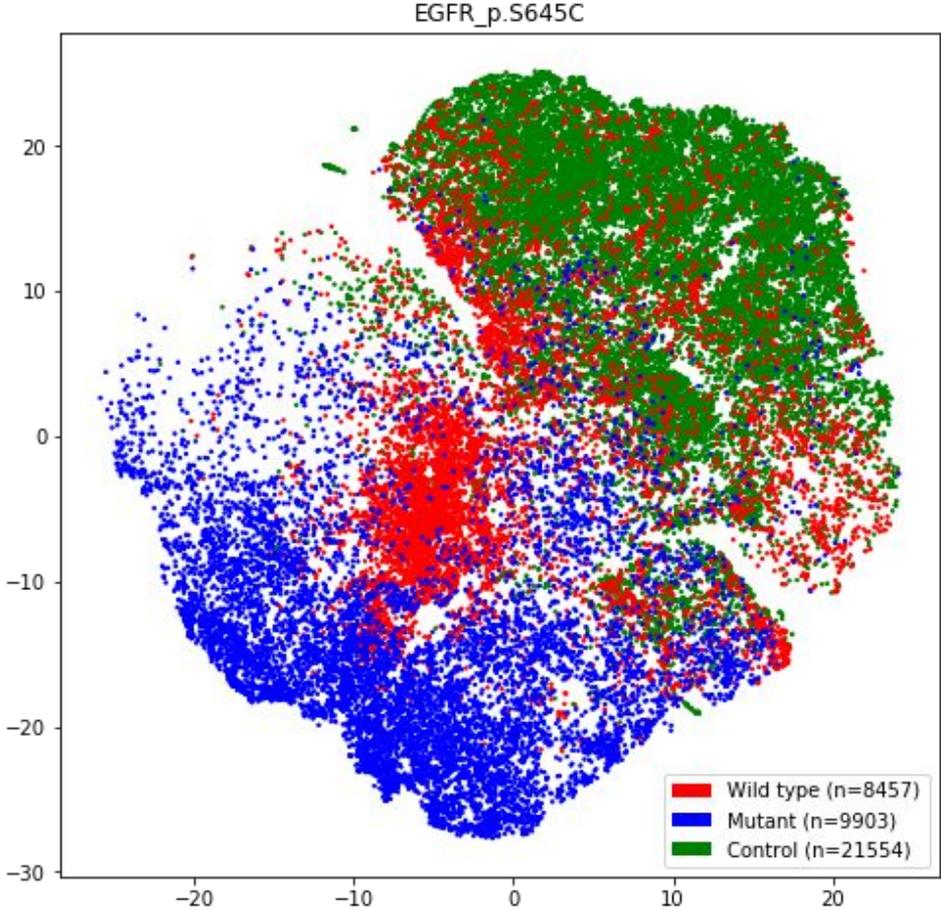
EGFR Wild Type



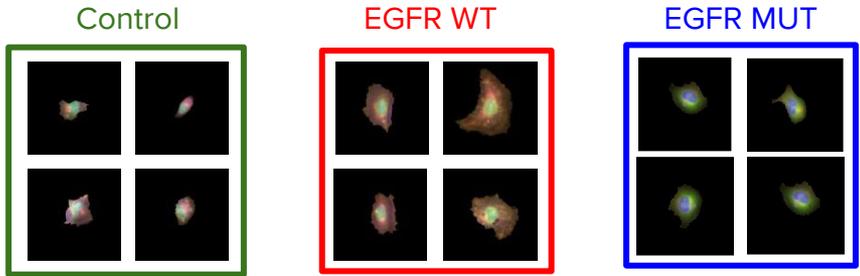
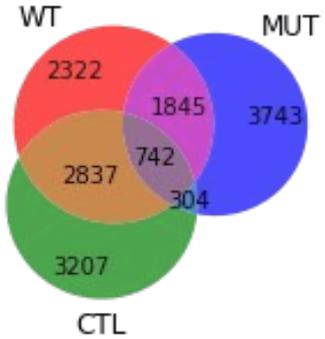
EGFR Mutant



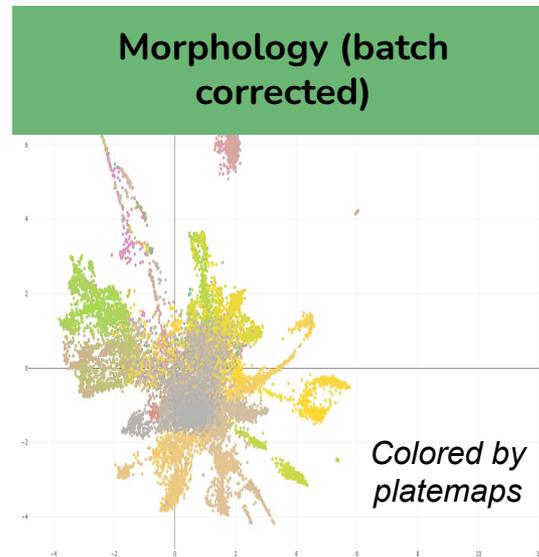
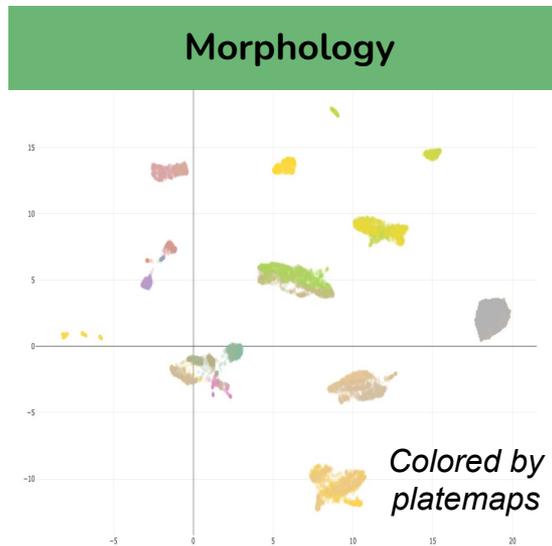
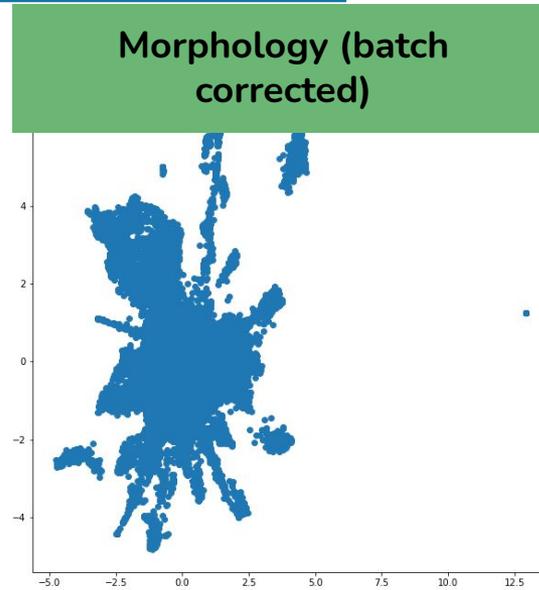
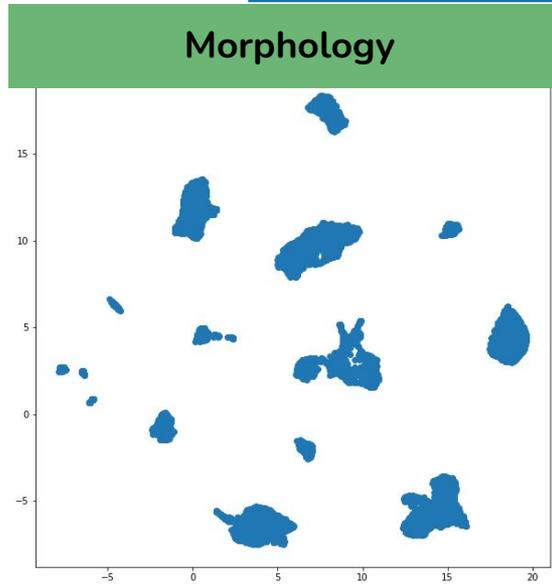
Determining variant impact



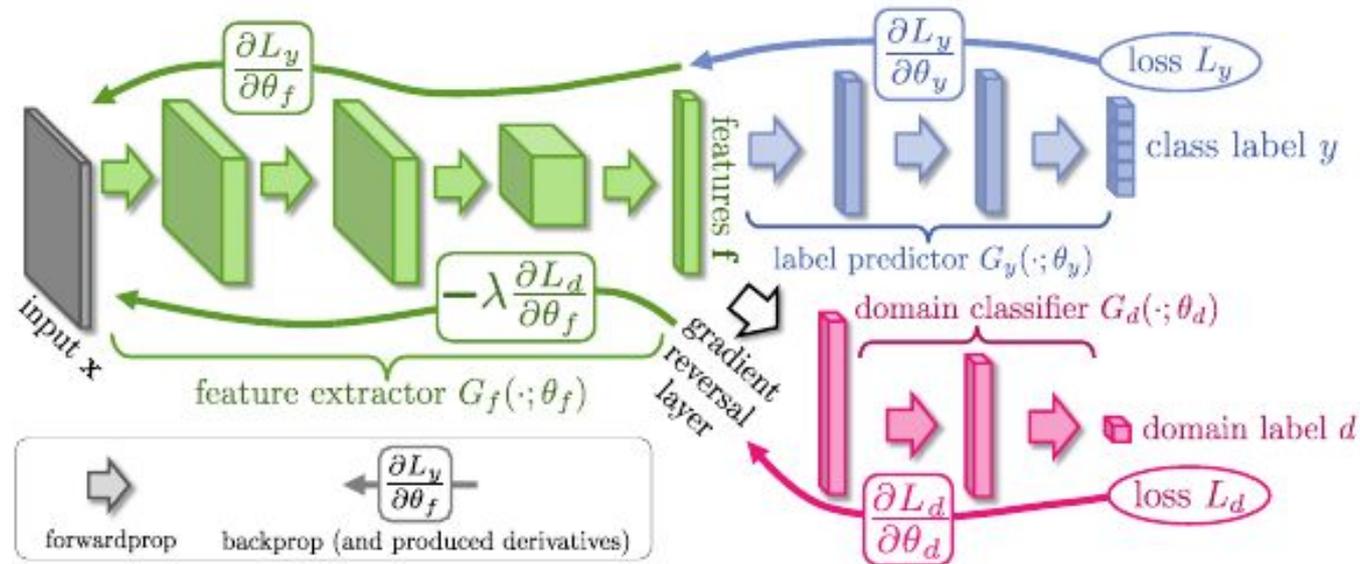
Variant impact: 66.9%



Correcting for batch effects

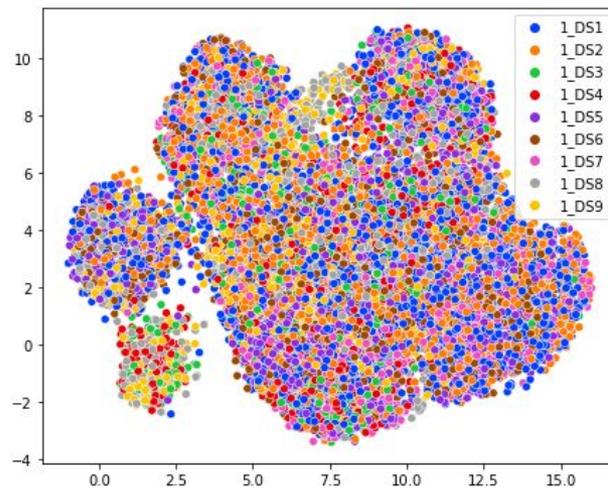
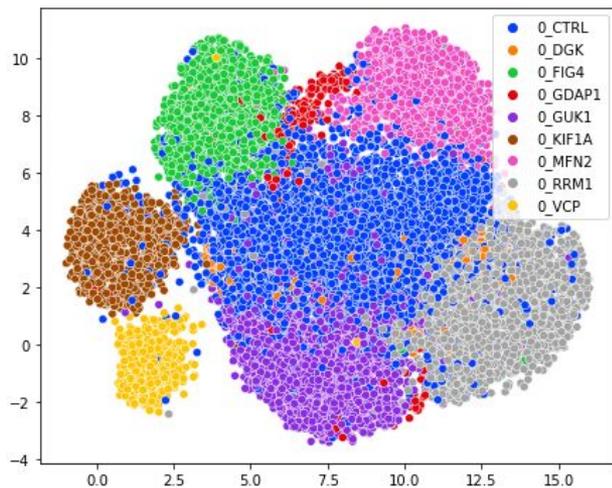


Correcting for batch effects — Gradient Reversal Layer

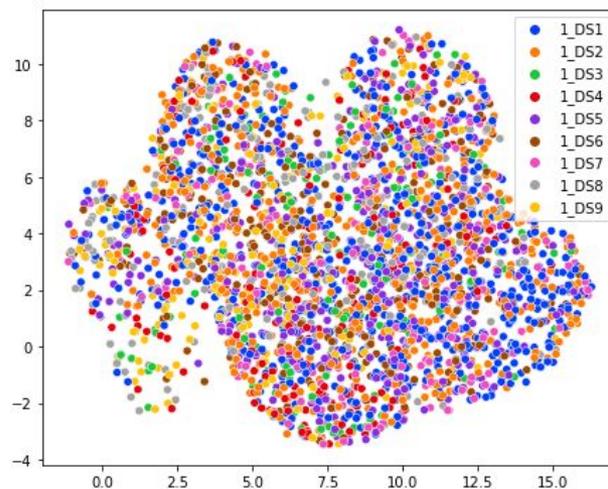
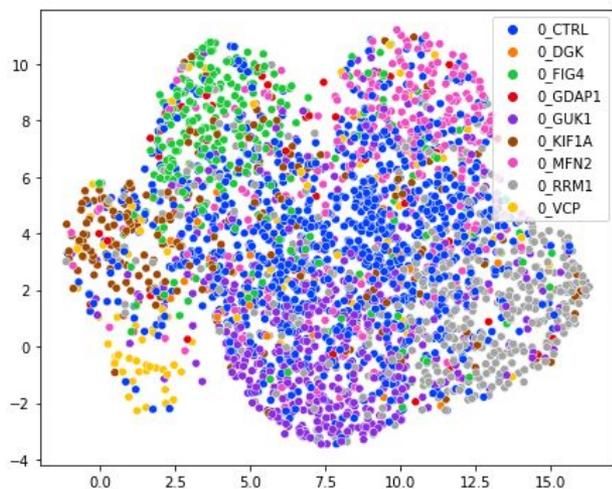


Minimizing batch effect confounders

ResNet18GRL - training



ResNet18GRL - validation



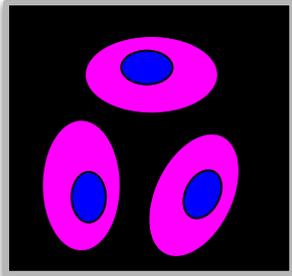
Open challenges

- Domain adaptation and batch effect correction
- Explainable models and interpretation capabilities

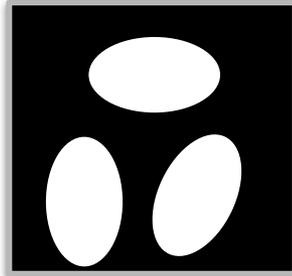
Image-based Profiling

Extracting information from biomedical images

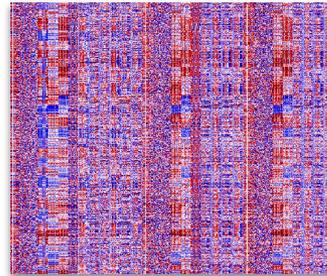
1. Raw images



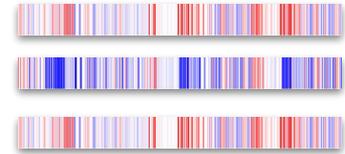
2. Single cells



3. Feature extraction



4. Aggregated profiles



1. Raw diamonds



2. Pieces of diamond



3. Diamond polishing



4. Jewelry

Thank you!