Data science in cell imaging
Lecture 4: phenotypic screening

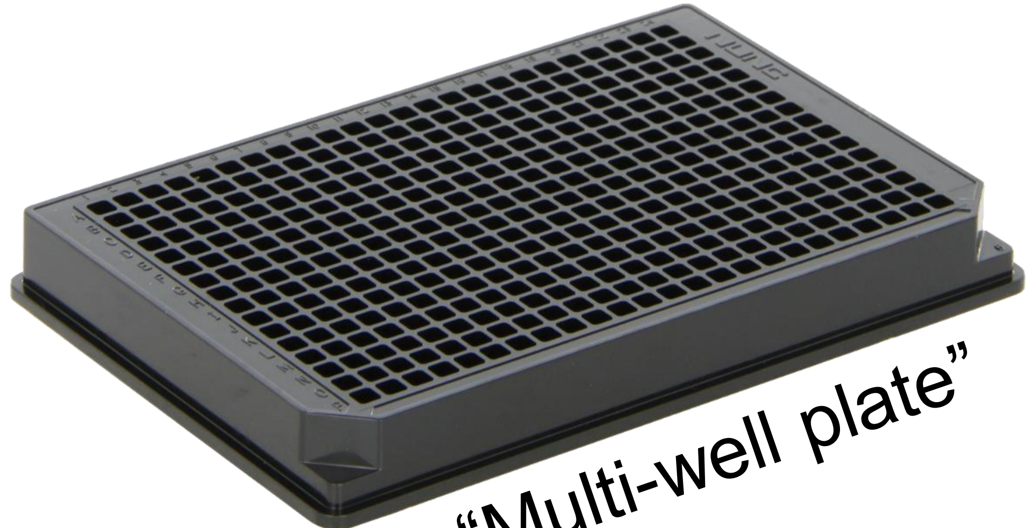
"The Great Wave off Kanagawa", by Hokusai, ~1830 (Source: Wikipedia)

Today

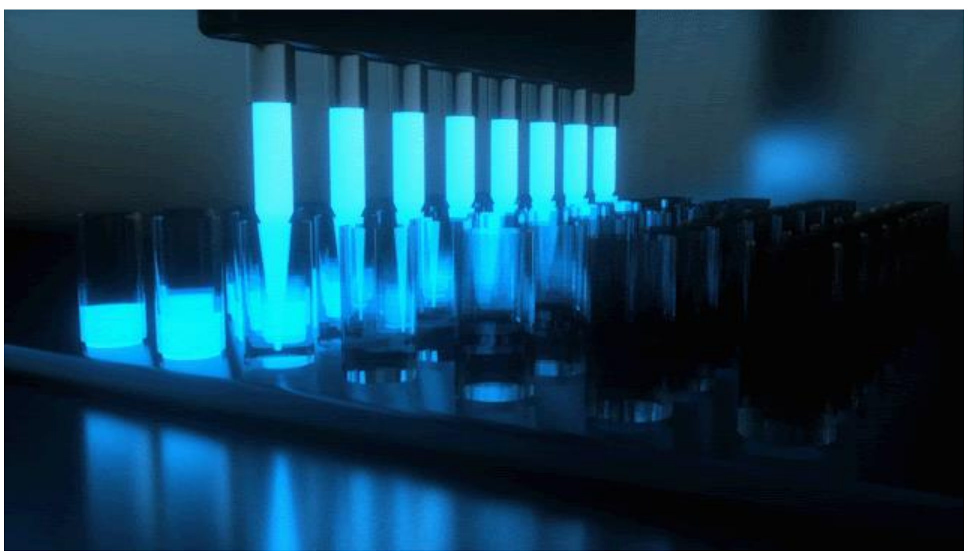
- High content cell profiling (slides adapted from Anne Carpenter)
- A few concrete examples

Discovering drugs in high throughput

X 384 =

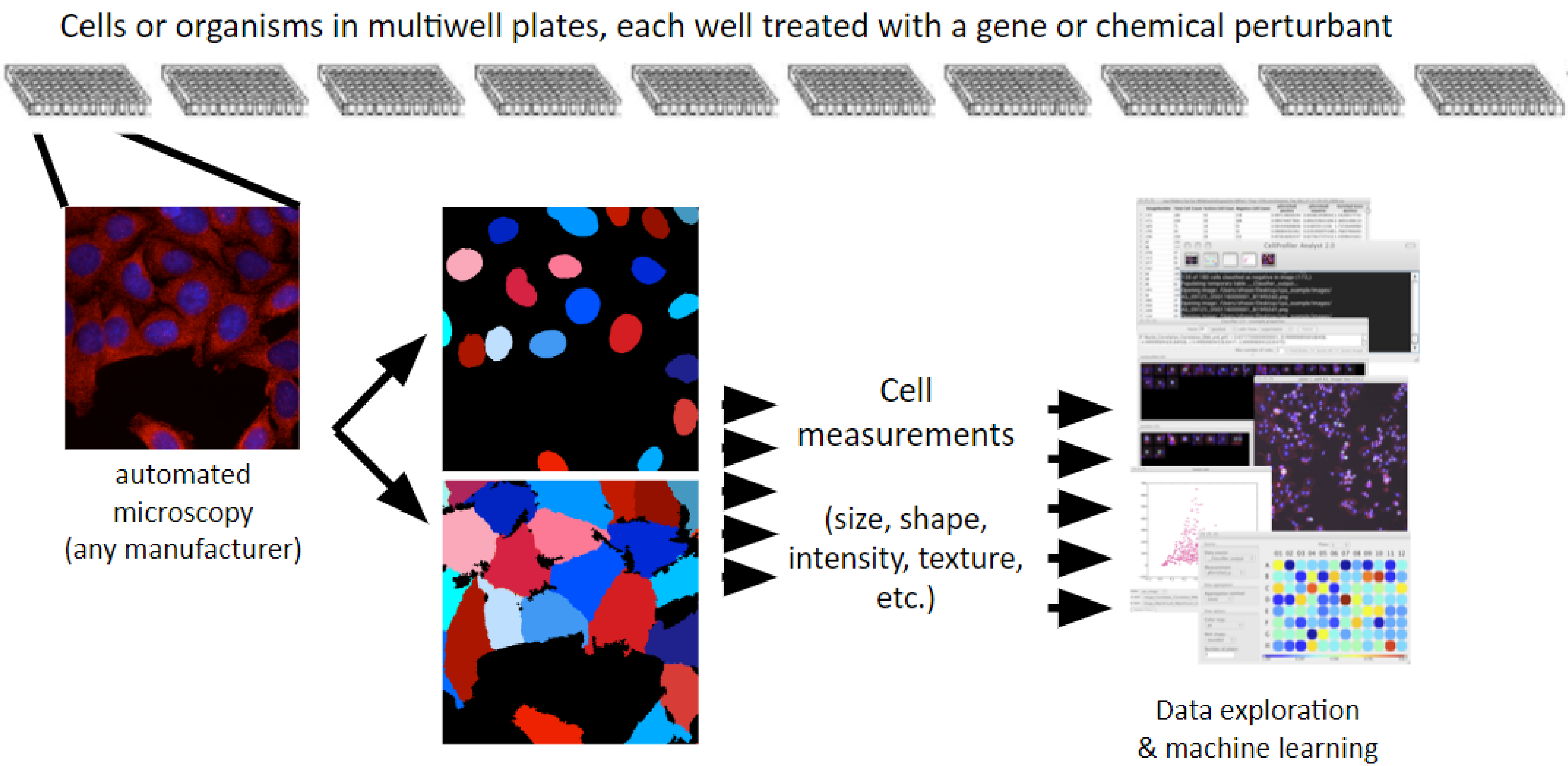


"Multi-well plate"



Anne Carpenter, Image: Nalgene; video: Chemistry World

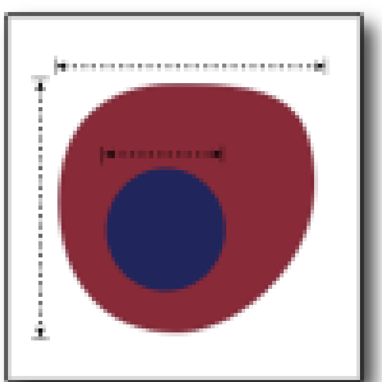
Large scale imaging experiments



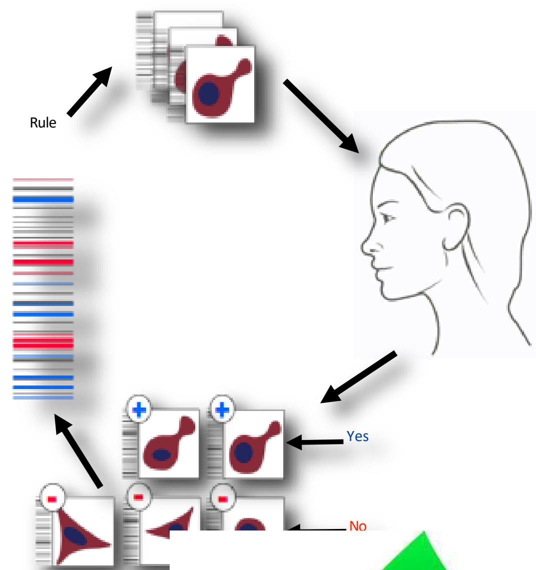
Anne Carpenter

Three waves of quantitative image analysis

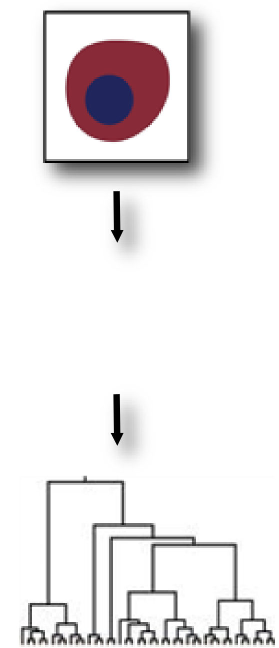
Measure known phenotypes



Train for known phenotypes



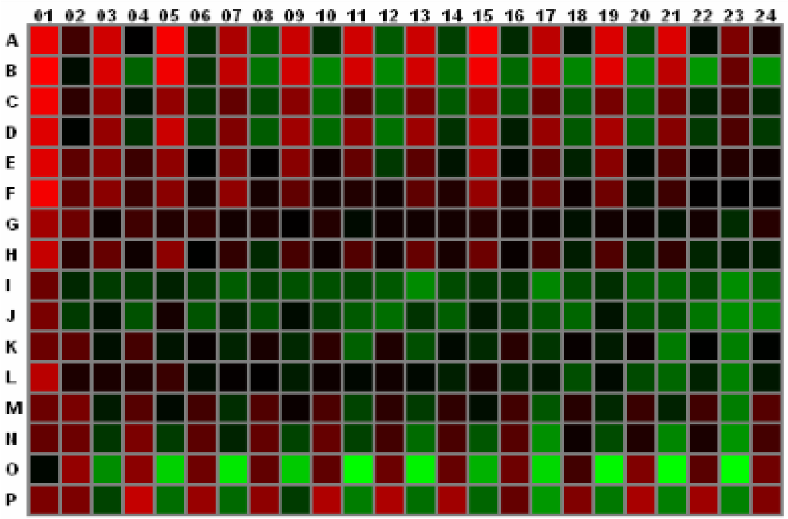
Discover new phenotypes

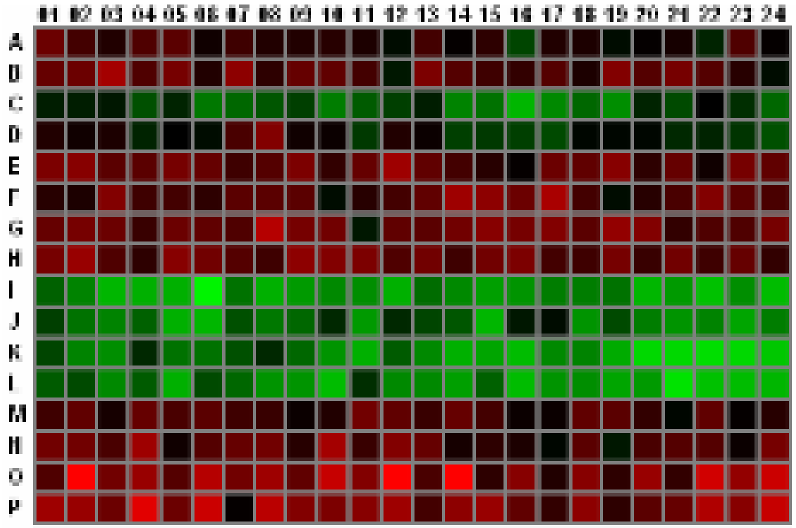


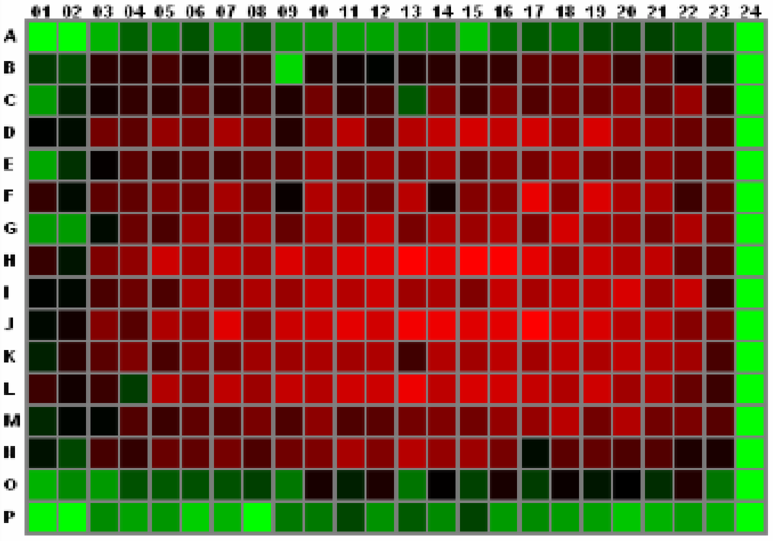
Anne Carpenter

Quality control (QC)

QC: are there systematic artifacts?

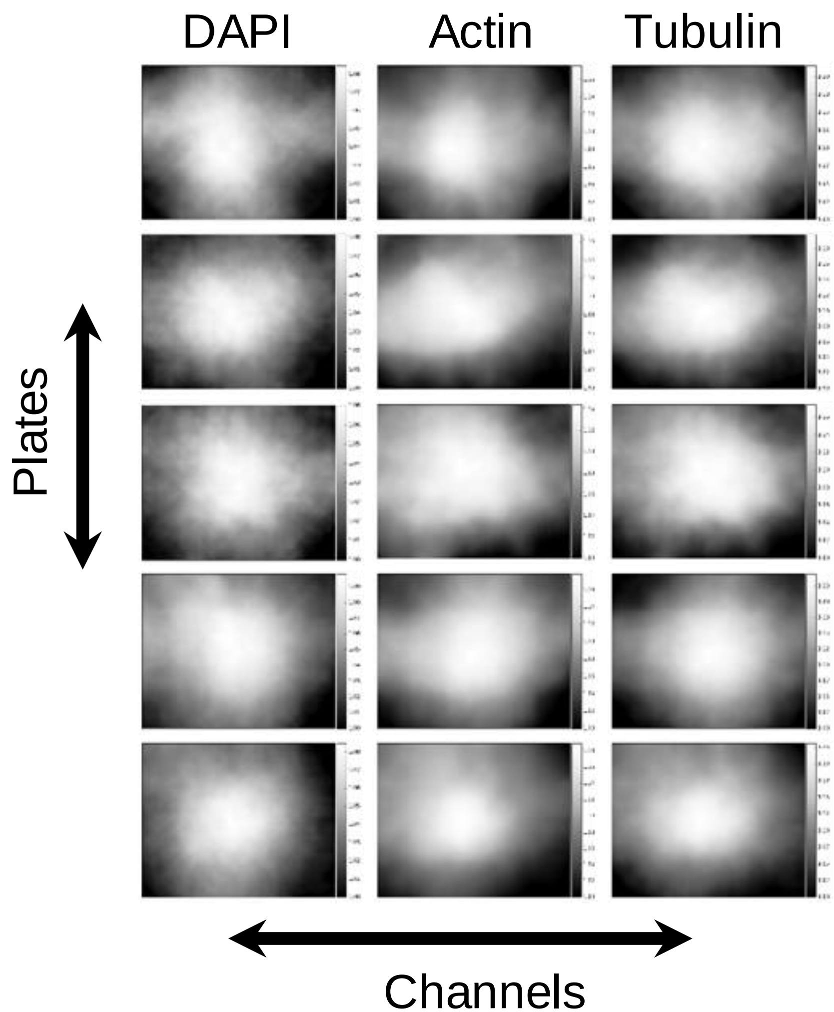






Anne Carpenter, AcuityXpress images, courtesy of Ralph Garippa

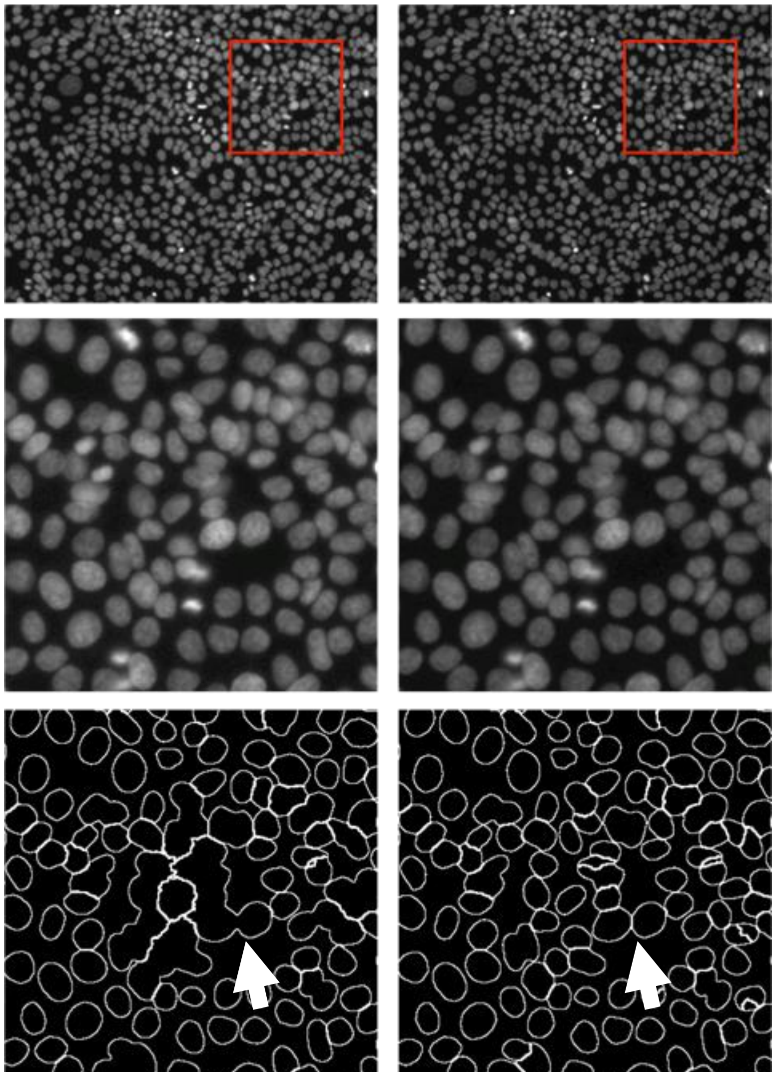
QC: Handling non-homogenous illumination across the image field



Anne Carpenter, from Singh et al. (2014)

QC: Handling non-homogenous illumination across the image field

Uncorrected | Corrected



Uncorrected

True mechanistic class		Act	Aur	Ch	DD	DR	Eg5	Epi	KI	MD	MS	PD	PS	Acc
Actin disruptors	Act	1						1					3	20 %
Aurora kinase inhibitors	Aur		12											100 %
Cholesterol-lowering	Ch			4				2						67 %
DNA damage	DD				8	1								89 %
DNA replication	DR				2	6								75 %
Eg5 inhibitors	Eg5						11				1			92 %
Epithelial	Epi			2				6						75 %
Kinase inhibitors	KI								5					100 %
Microtubule destabilizers	MD				1		2			11				79 %
Microtubule stabilizers	MS						1				8			89 %
Protein degradation	PD											7		100 %
Protein synthesis	PS												8	100 %

Predicted class (columns Act–PS)

Overall accuracy: 87 / 103 **84%**

Corrected

True mechanistic class		Act	Aur	Ch	DD	DR	Eg5	Epi	KI	MD	MS	PD	PS	Acc
Actin disruptors	Act	1						1					3	20 %
Aurora kinase inhibitors	Aur		12											100 %
Cholesterol-lowering	Ch			4				2						67 %
DNA damage	DD				9									100 %
DNA replication	DR					8								100 %
Eg5 inhibitors	Eg5						12							100 %
Epithelial	Epi							8						100 %
Kinase inhibitors	KI								5					100 %
Microtubule destabilizers	MD				1		2			11				79 %
Microtubule stabilizers	MS						1				8			89 %
Protein degradation	PD											7		100 %
Protein synthesis	PS												8	100 %

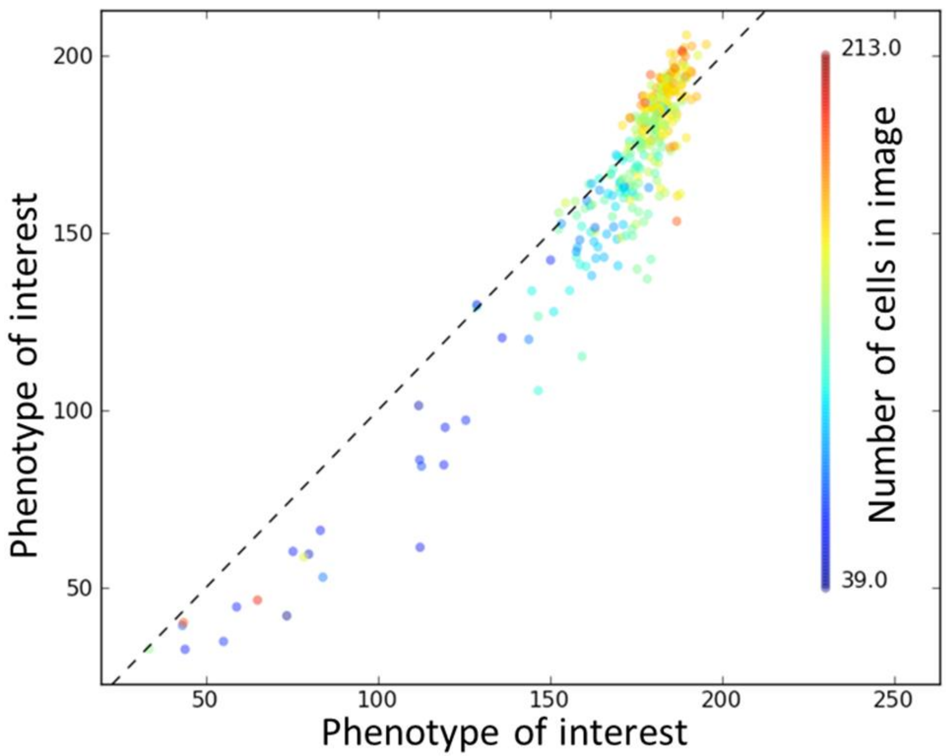
Predicted class (columns Act–PS)

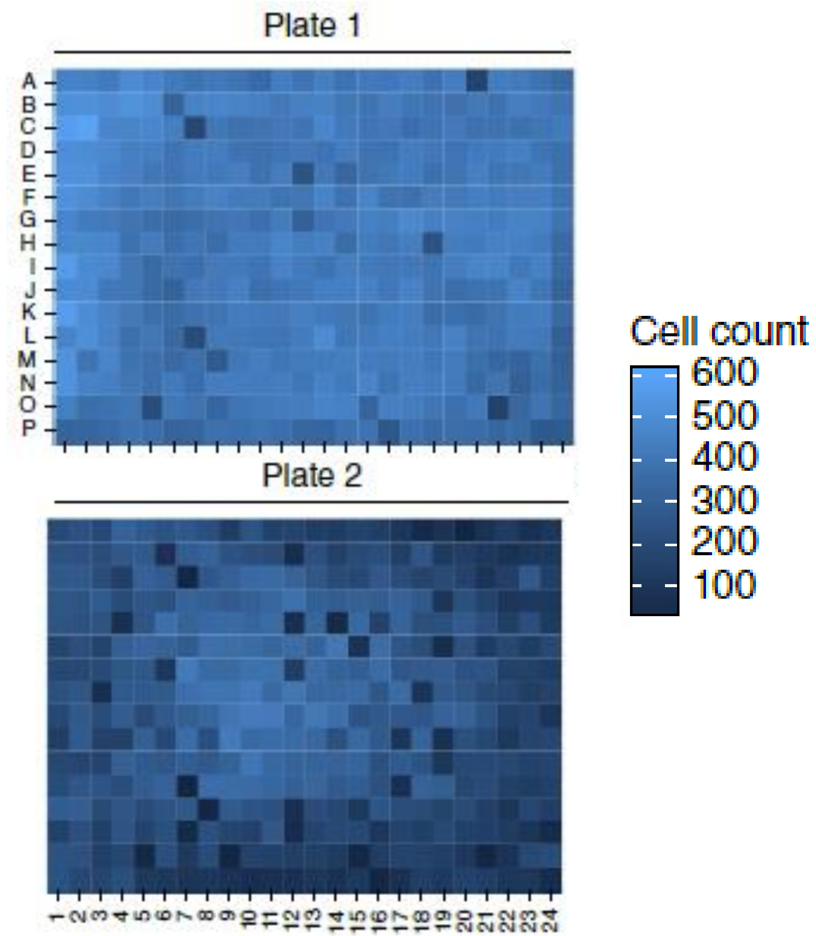
Overall accuracy: 93 / 103 **90%**

Anne Carpenter, from Singh et al. (2014)

QC: is the phenotype confounded by other factors?

e.g., cell density, cell cycle, cell microenvironment, …

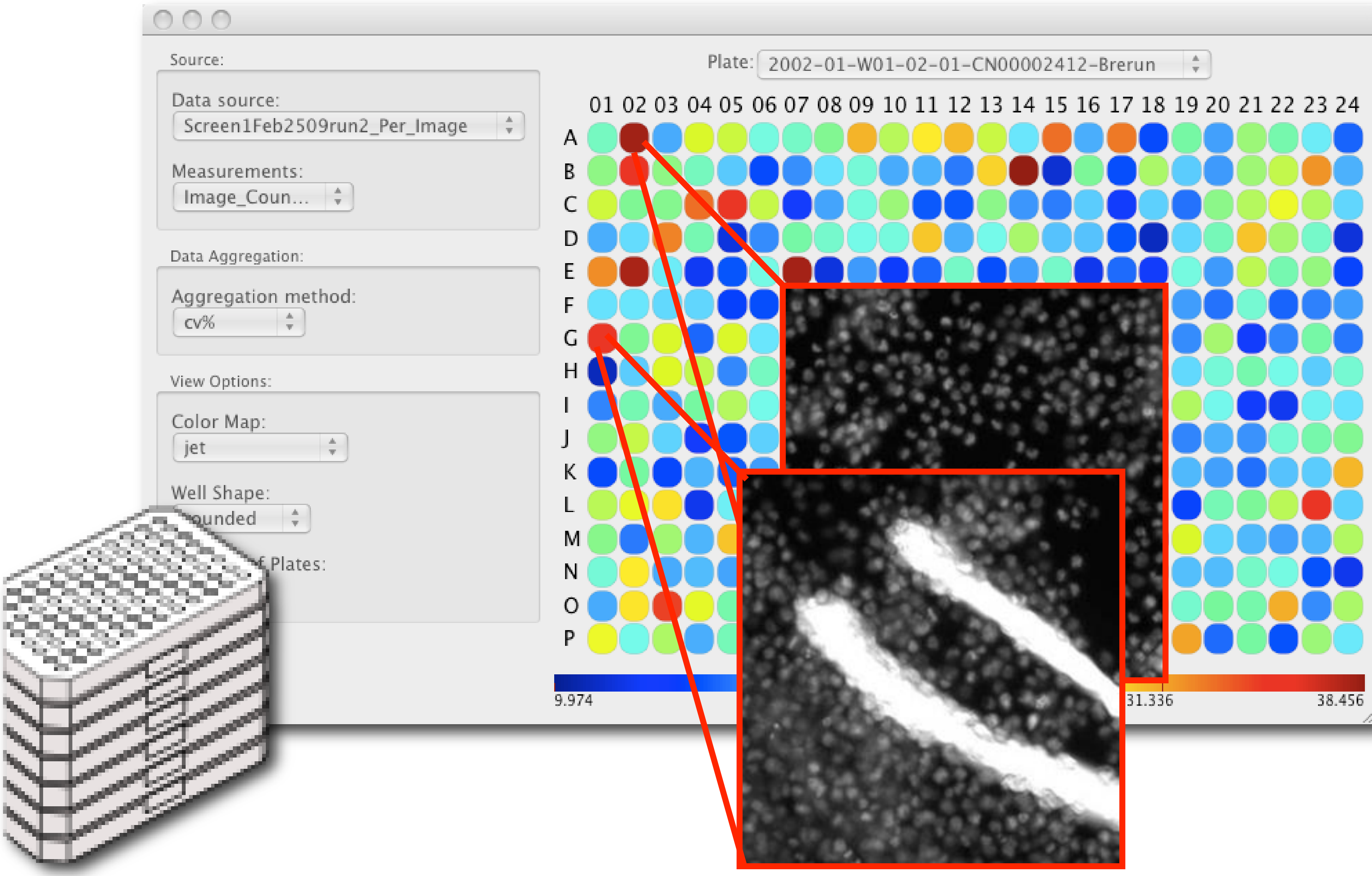




Anne Carpenter, Vebjorn Ljosa and Thouis Jones, Broad Imaging Platform

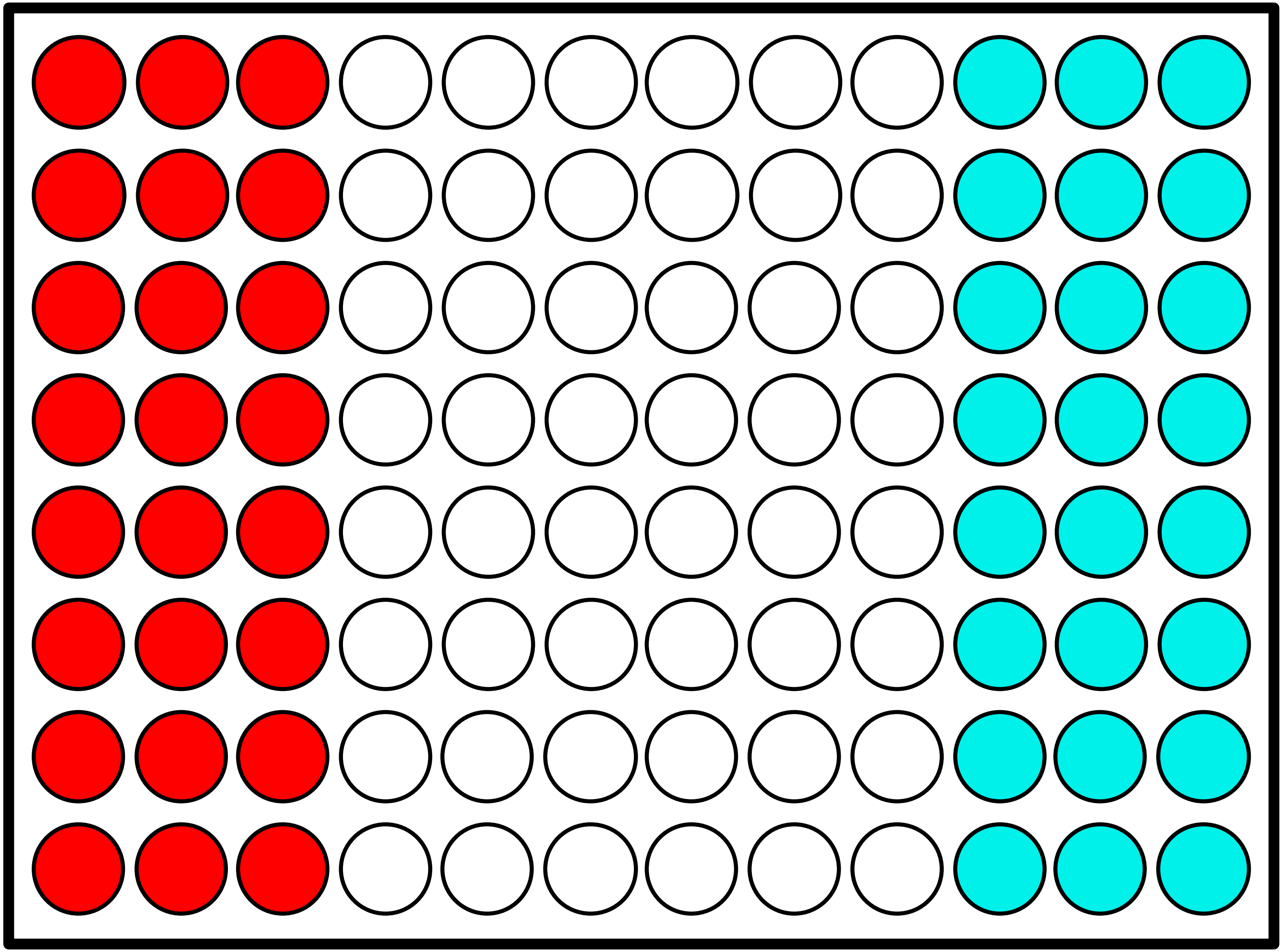
Caicedo et al. (2017)

QC: are the hits real?



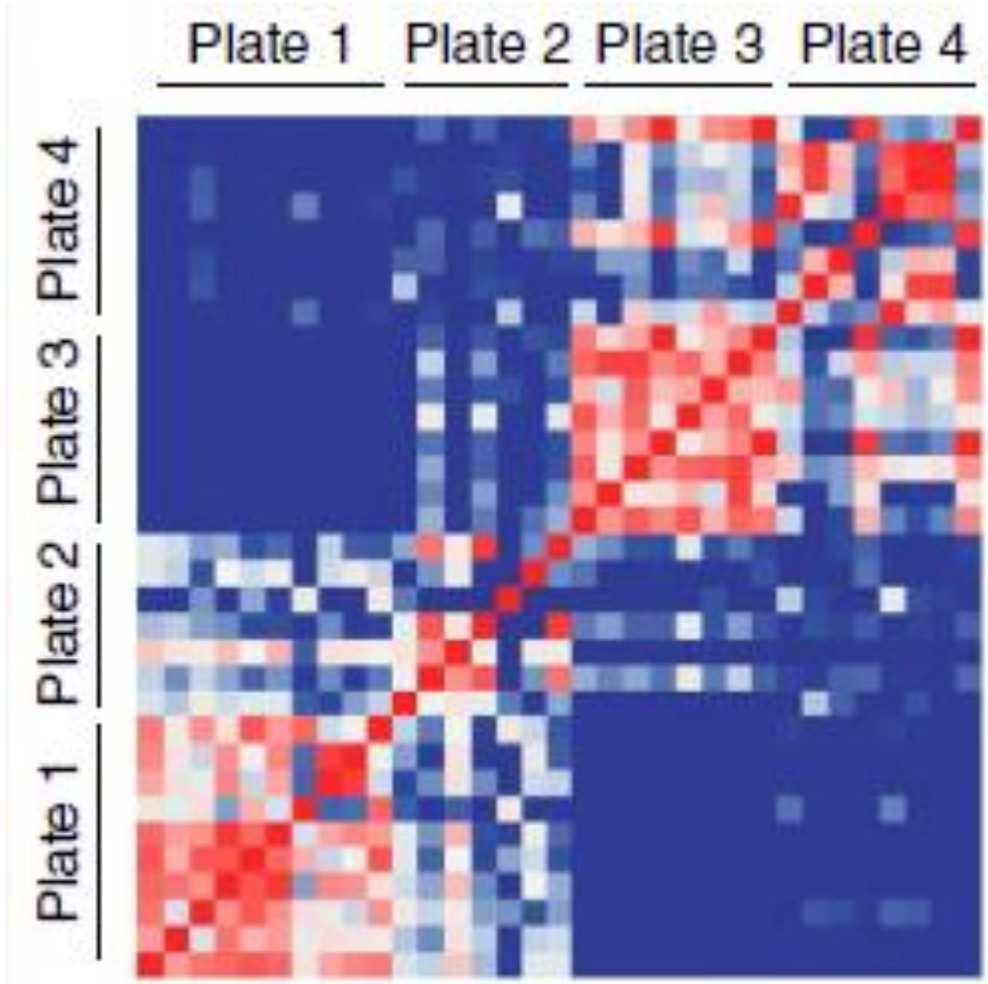
Anne Carpenter, from CellProfiler Analyst (www.cellprofiler.org)

Negative and positive controls



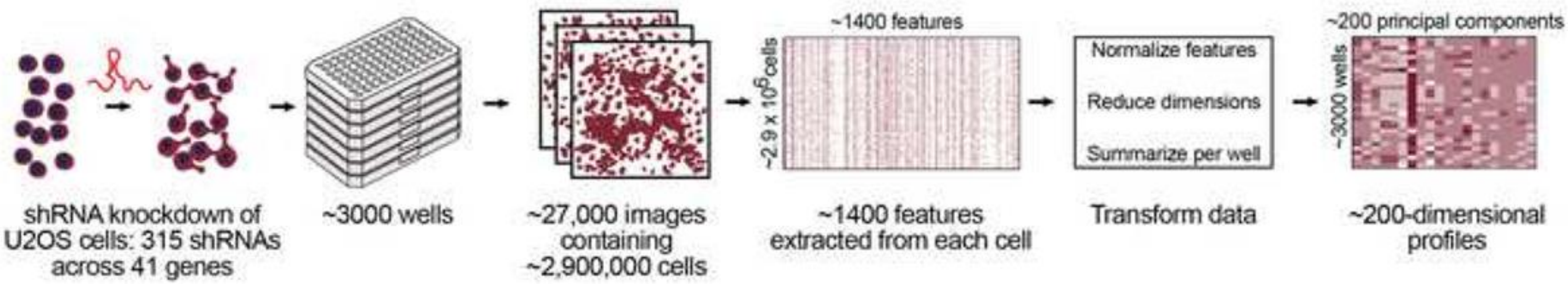
Batch effects

“non-biological factors in an experiment cause changes in the data produced by the experiment” (Wikipedia)



Caicedo et al. (2017)

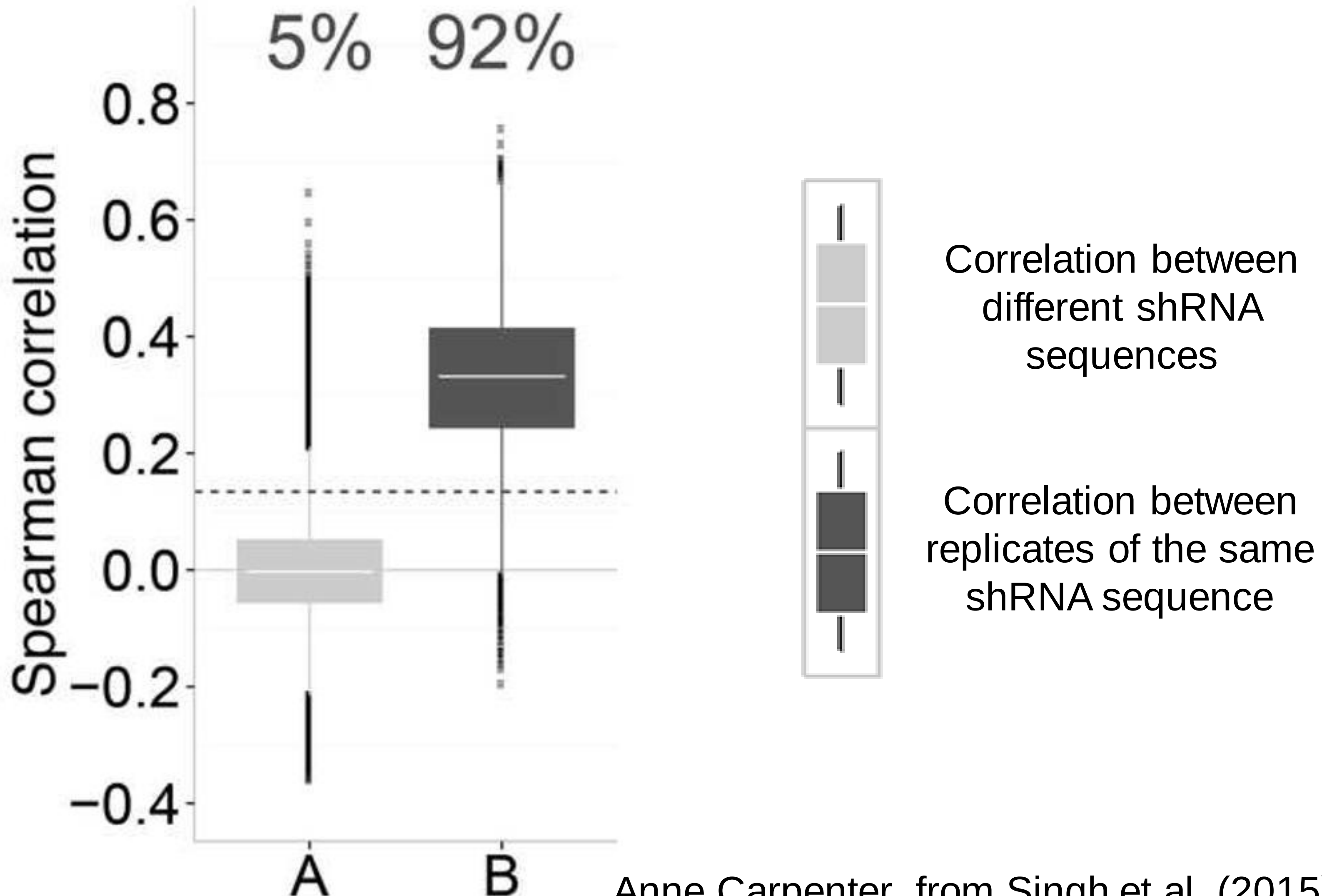
Off target effects
Demonstrated via shRNA screening



Test: 315 shRNAs against 41 genes in U2OS (human) cells

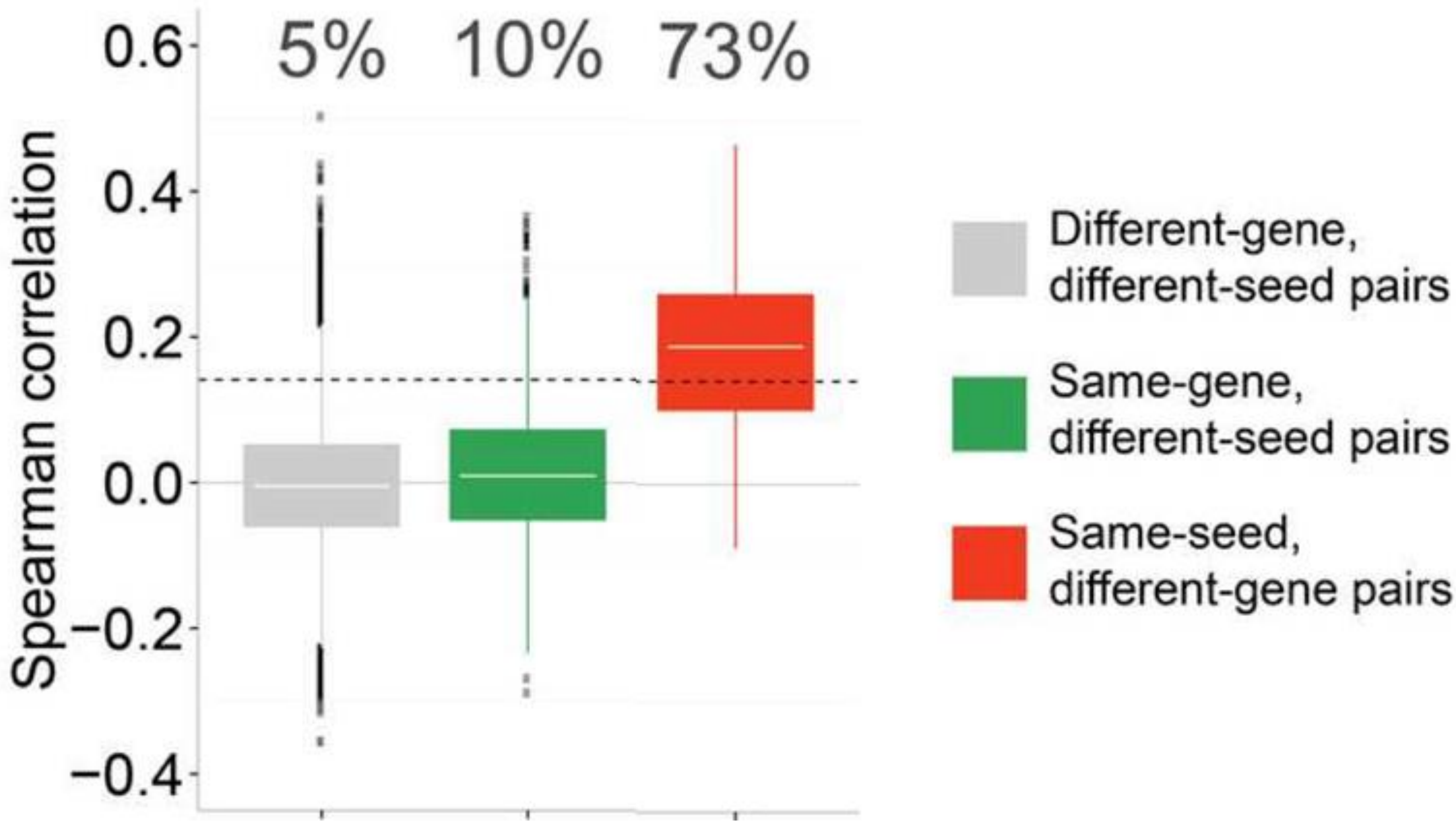
Anne Carpenter, from Singh et al. (2015)

shRNA profiles of screen hits are reproducible!



Anne Carpenter, from Singh et al. (2015)

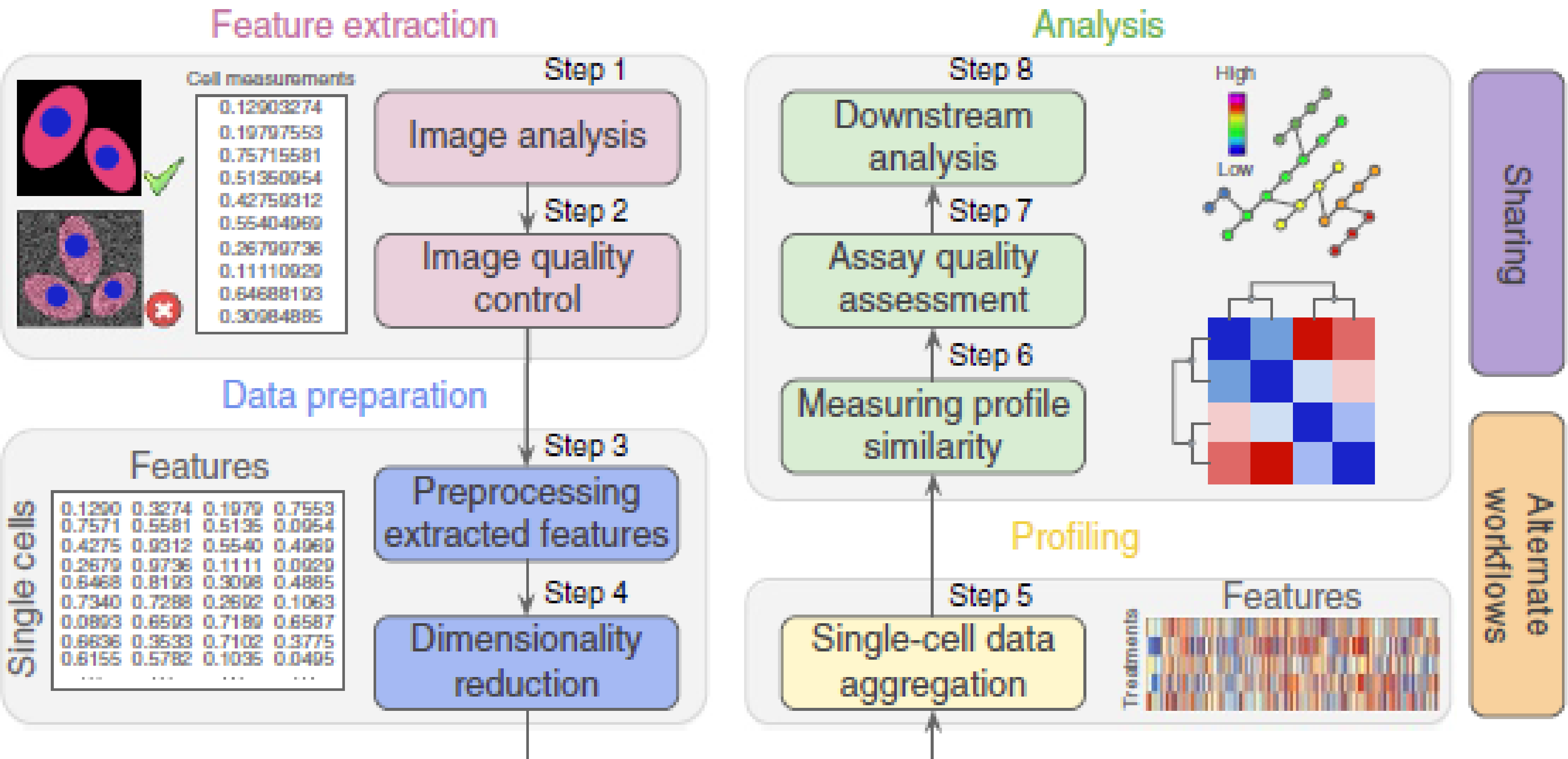
Off target effects dominate shRNA profiles!



Anne Carpenter, from Singh et al. (2015)

Back to high content single cell phenotypic profiling – the full pipeline

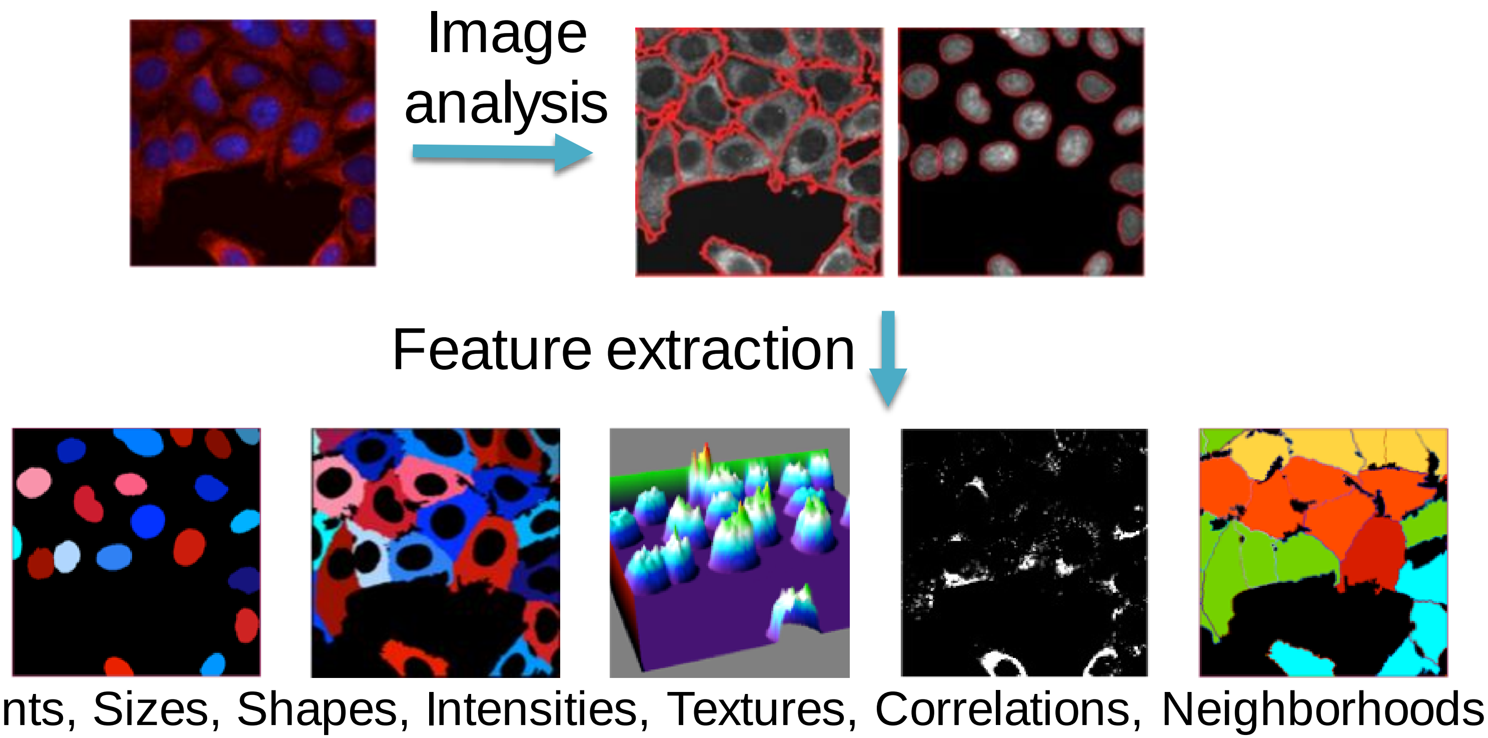
Cell profiling: "describing a population of cells as a rich collection of measurements"



Caicedo et al. (2017)

Image analysis
image → single cell features

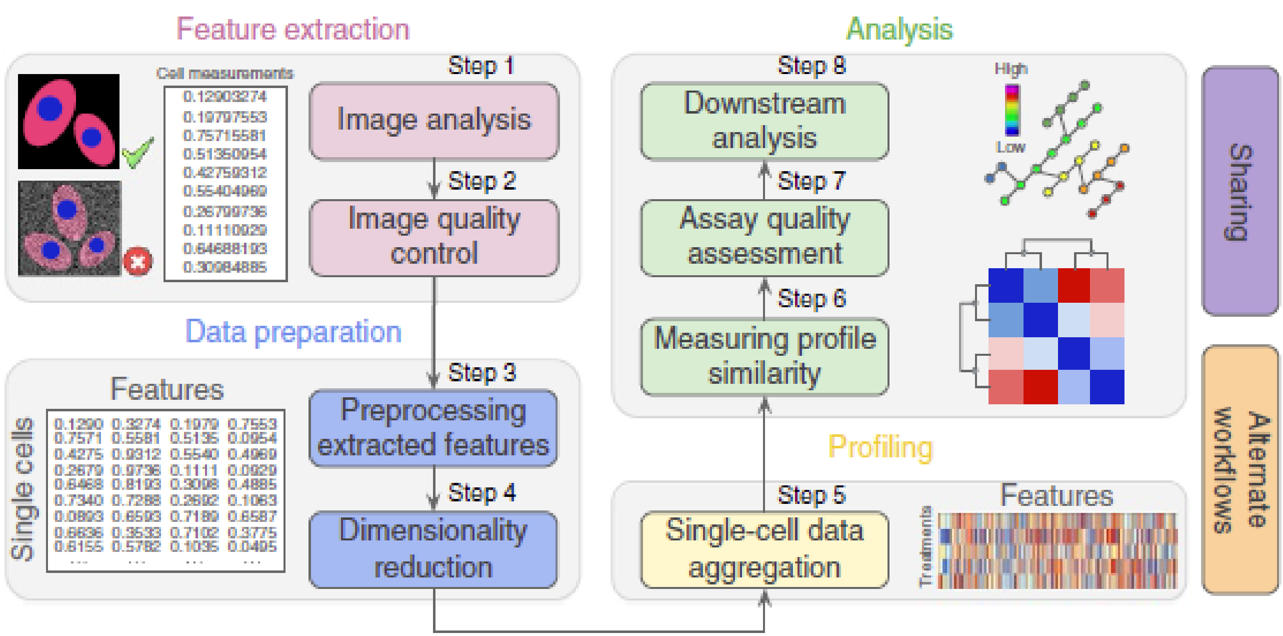
- Illumination correction
- Segmentation
- Tracking (for screens that include dynamics)



Anne Carpenter

(Automated) image quality control

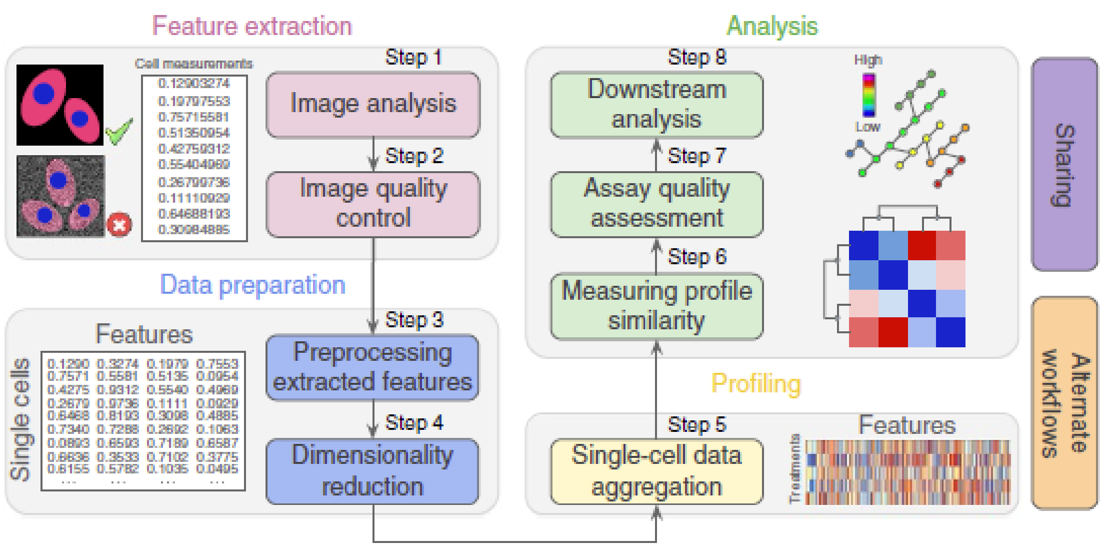
- Field of view: debris, saturation, focus
- Cell level quality control (outlier detection)



Caicedo et al. (2017)

Preprocessing extracted features

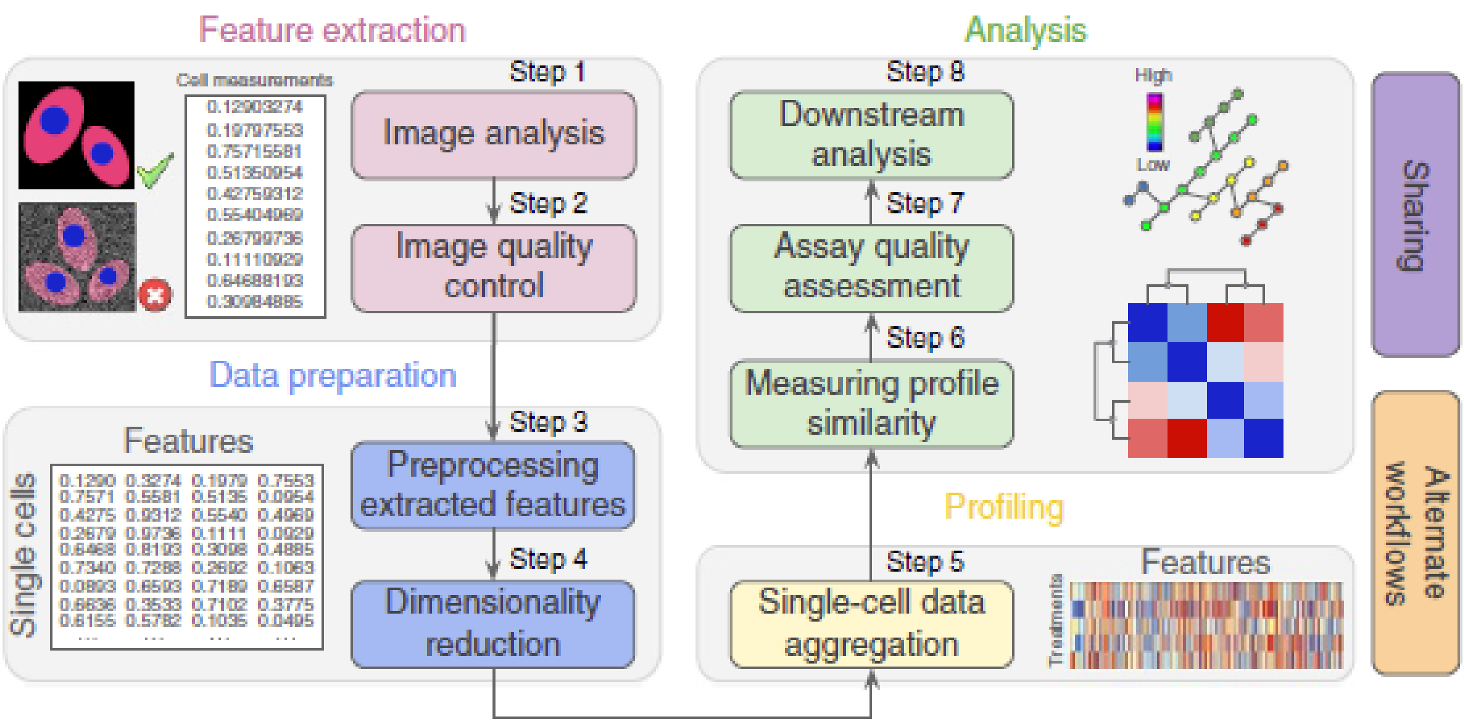
- Dealing with missing values
- Plate-layout-effect correction
- Batch-effect correction
- Feature transformation and normalization



Caicedo et al. (2017)

Dimensionality reduction

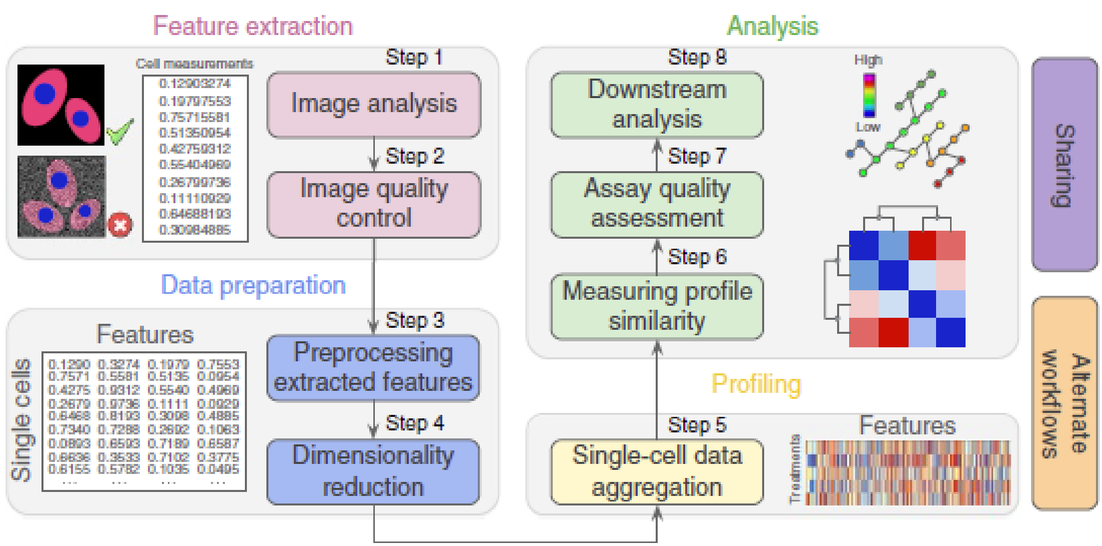
- Different ways of feature selection / dimensionality reduction



Caicedo et al. (2017)

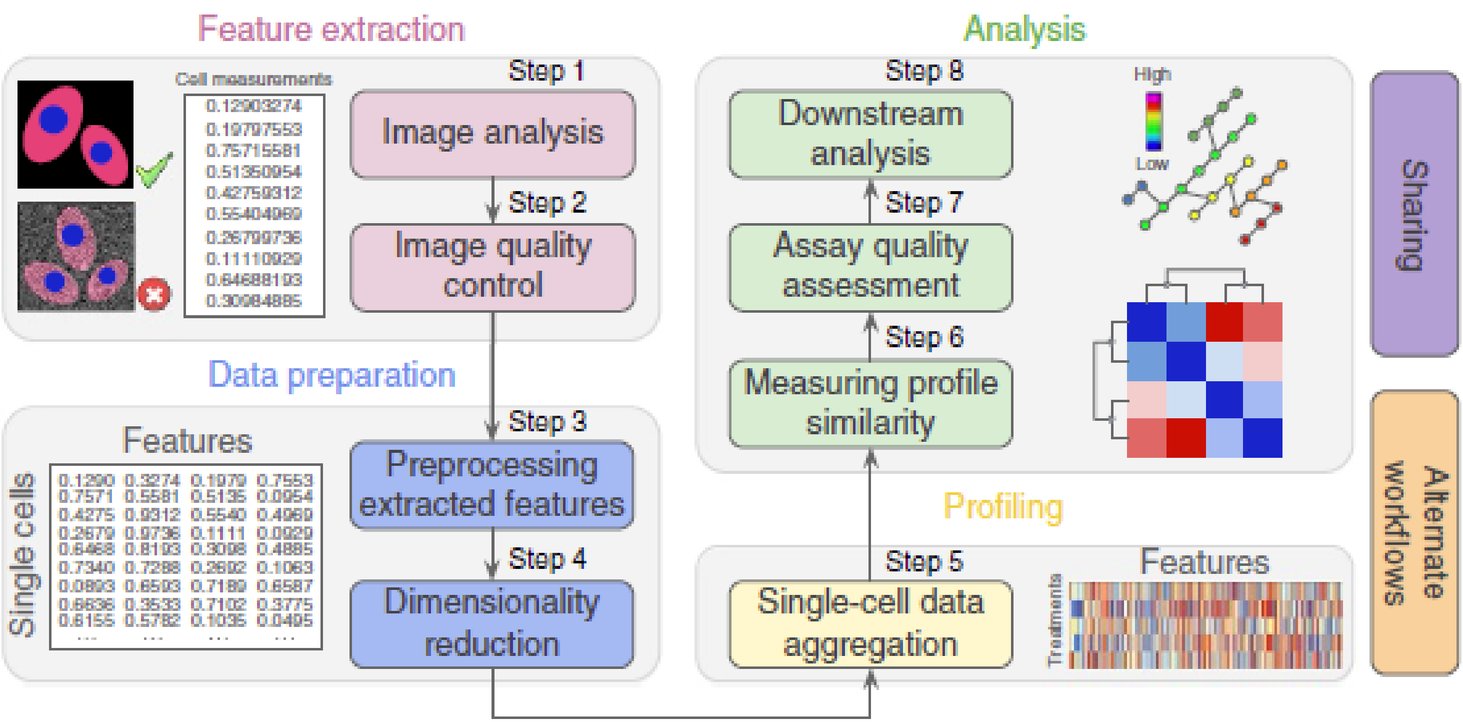
Single-cell data aggregation

- Comparing populations (profiling)
- Cell heterogeneity (different sub-populations of cells)
- Construct profiles at the level of images, fields of view, wells, or replicates



Caicedo et al. (2017)

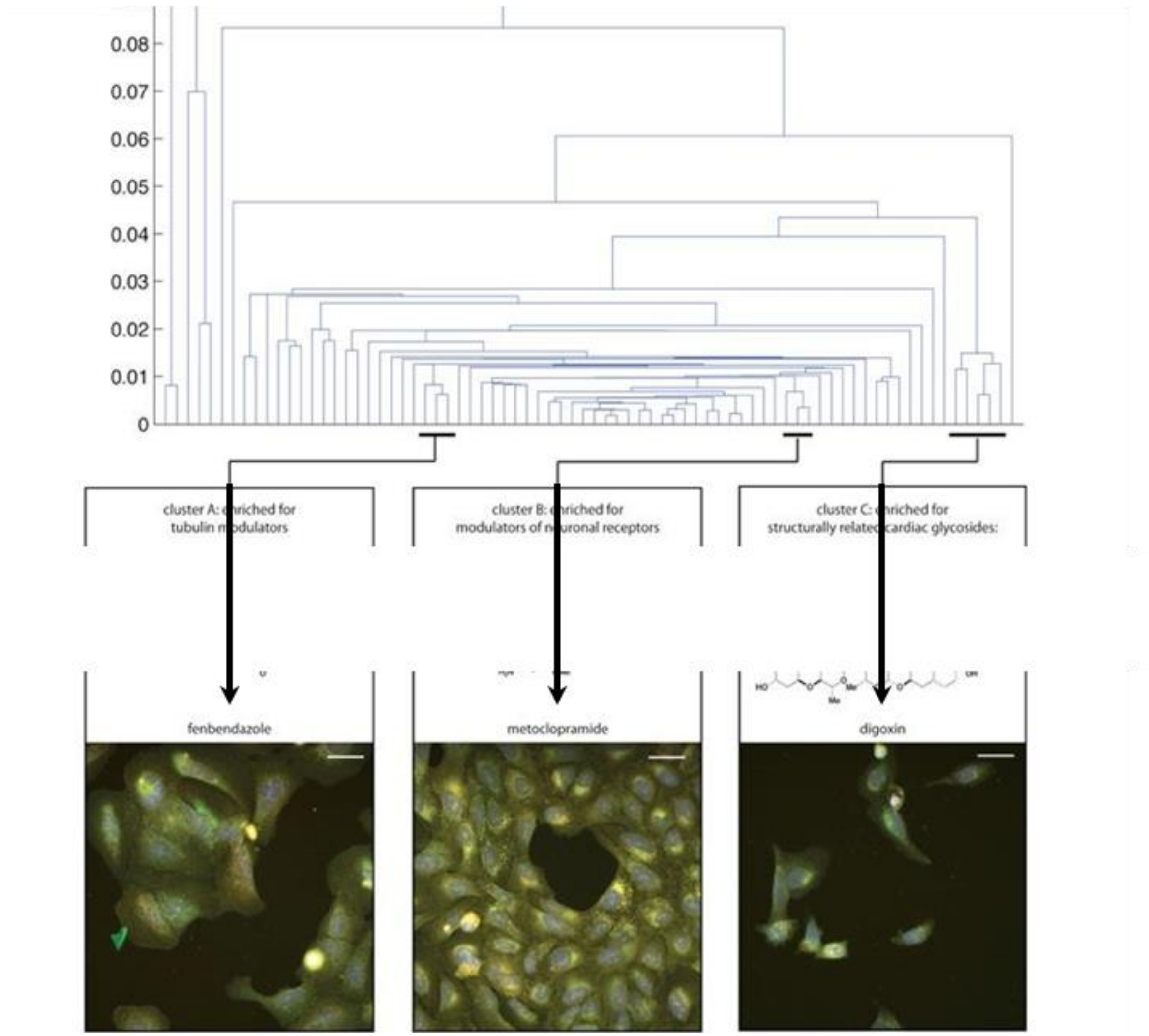
Measuring profile similarity
Assay quality assessment
Downstream analysis



Caicedo et al. (2017)

Interpreting high-dimensional phenotypes

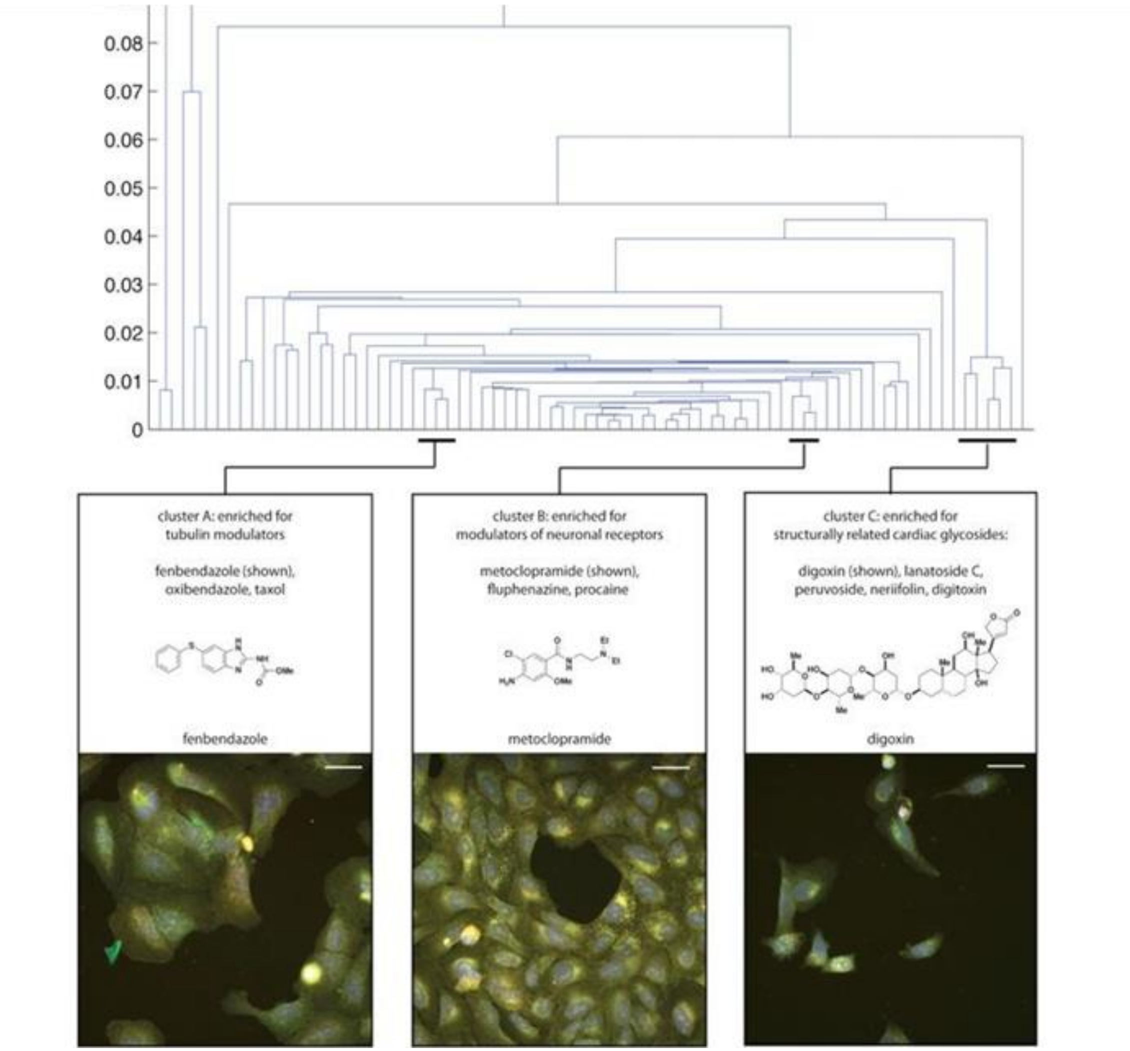
Look at images for each cluster



Anne Carpenter

Interpreting high-dimensional phenotypes

Look at annotations for each cluster



Anne Carpenter

Interpreting high-dimensional phenotypes

Look at features that distinguish clusters

	0	1	2	3	4	5	6
Cells_Texture_Gabor_Alexa568_10	98.684577	129.198993	130.007223	129.681593	129.317059	130.185317	127.858318
Cells_Texture_AngularSecondMoment_CellMask_3_0	0.134944	0.045163	0.040670	0.042798	0.042718	0.047332	0.045217
Cells_Texture_AngularSecondMoment_CellMask_5_0	0.094794	0.043140	0.039624	0.041798	0.041102	0.045364	0.042749
Nuclei_Texture_AngularSecondMoment_CellMask_3_0	0.142980	0.049463	0.047698	0.051781	0.046577	0.052370	0.047757
Nuclei_Texture_Gabor_Alexa568_10	98.988721	132.283091	135.577399	135.651436	133.083455	134.337589	130.468138
Cells_RadialDistribution_FracAtD_Hoechst_3of4	0.363300	0.404550	0.408073	0.403985	0.407065	0.408418	0.405231
Cytoplasm_Texture_InfoMeas2_CellMask_10_0	0.201429	0.840334	0.864921	0.865561	0.844166	0.766654	0.824663
Cytoplasm_Texture_InfoMeas2_CellMask_5_0	0.629889	0.862343	0.878067	0.880622	0.865987	0.852146	0.849805
Cells_Texture_Entropy_CellMask_3_0	2.956207	3.447352	3.511392	3.463545	3.488114	3.462852	3.460178
Cells_Texture_InverseDifferenceMoment_Hoechst_3_0	0.497116	0.403176	0.398306	0.403155	0.399352	0.395730	0.402702
Cells_Neighbors_NumberOfNeighbors_5	0.995991	0.029025	0.018101	0.006735	0.018800	0.027008	0.022804
Cytoplasm_AreaShape_Zernike_8_6	0.008789	0.012746	0.013128	0.013186	0.012873	0.012836	0.012480
Cells_Neighbors_PercentTouching_5	18.693193	0.194687	0.185990	-0.001881	0.164185	0.363572	0.129390
Nuclei_Neighbors_NumberOfNeighbors_1	0.779256	0.001808	-0.000646	0.001276	-0.000116	0.002775	-0.000143
Nuclei_Intensity_MassDisplacement_Alexa568	1.059710	0.376466	0.334523	0.303492	0.377752	0.310736	0.408470
Cells_Neighbors_NumberOfNeighbors_Adjacent	0.840963	0.003611	0.000905	-0.002544	-0.000132	0.007366	-0.000331
Nuclei_Texture_InverseDifferenceMoment_Hoechst_3_0	0.480997	0.393191	0.383865	0.385076	0.389208	0.378258	0.395047
Cells_Neighbors_PercentTouching_Adjacent	4.788959	-0.011521	0.004760	0.013783	0.010317	-0.042623	-0.006325
Nuclei_Neighbors_PercentTouching_1	4.124720	-0.017390	-0.001461	0.019684	0.009621	-0.054541	-0.004060
Cells_Correlation_Correlation_Hoechst_CellMask	0.445154	0.893111	0.894903	0.878988	0.900852	0.861685	0.901846
Cells_Intensity_MassDisplacement_Alexa568	0.998623	0.408807	0.409201	0.396260	0.404762	0.370103	0.416380
Cells_Texture_DifferenceEntropy_Hoechst_3_0	1.412808	1.662219	1.665551	1.649187	1.670052	1.688439	1.665959
Nuclei_Texture_AngularSecondMoment_CellMask_5_0	0.102415	0.047374	0.046753	0.051539	0.044798	0.050841	0.045112
Cells_Texture_Correlation_Alexa568_10_0	-0.239855	-0.376286	-0.381486	-0.366189	-0.387679	-0.384570	-0.384801
Cells_AreaShape_Zernike_3_1	0.020641	0.017346	0.016840	0.017188	0.016925	0.016582	0.017262
Nuclei_AreaShape_Zernike_3_1	0.021177	0.017543	0.017246	0.017813	0.017104	0.017072	0.017355
Nuclei_AreaShape_Zernike_1_1	0.052819	0.045796	0.045063	0.045380	0.045640	0.044408	0.045968
Cytoplasm_Correlation_Correlation_Hoechst_CellMask	0.199921	0.634145	0.677927	0.673406	0.642531	0.572587	0.623264

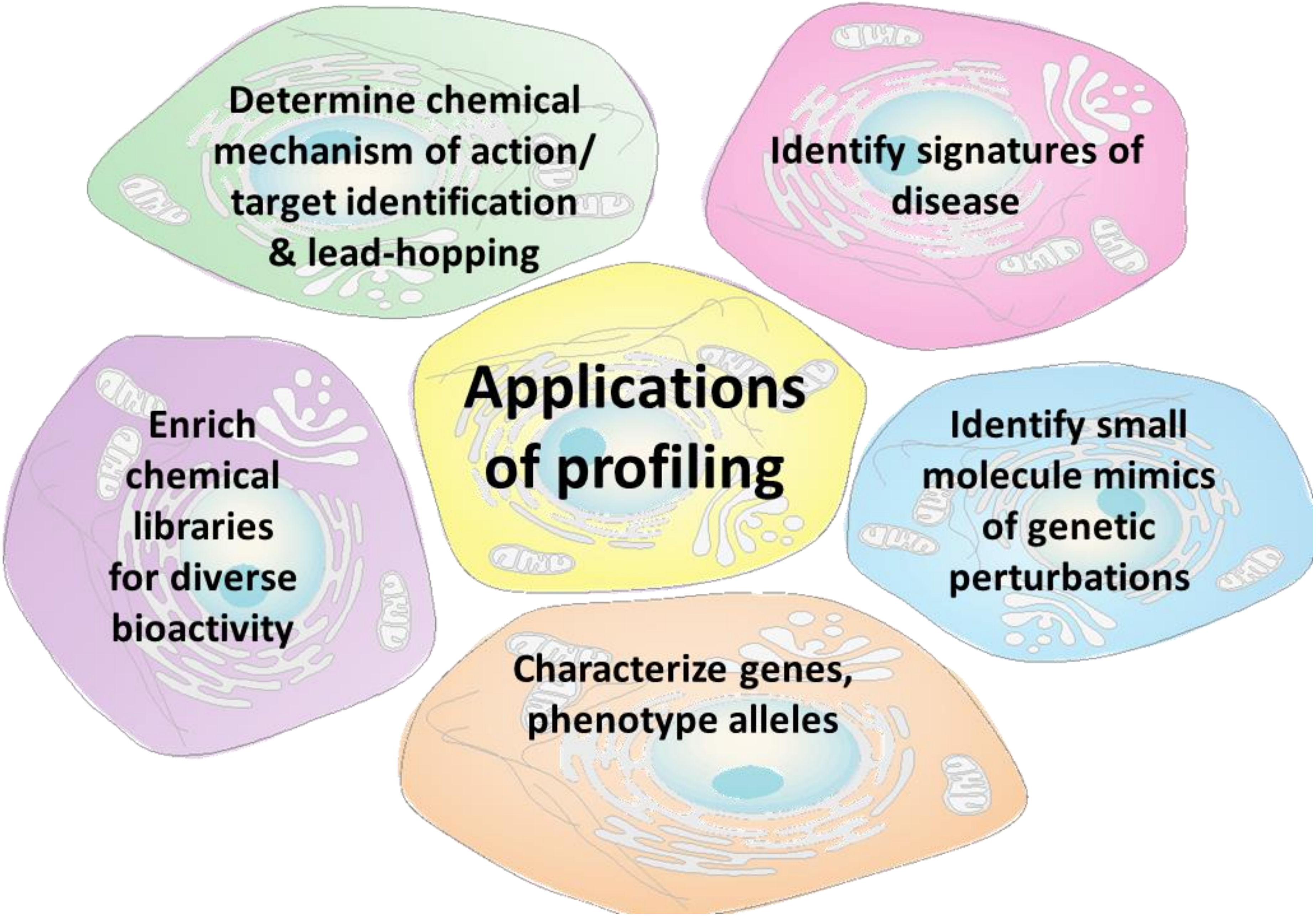
Anne Carpenter

Interpreting high-dimensional phenotypes

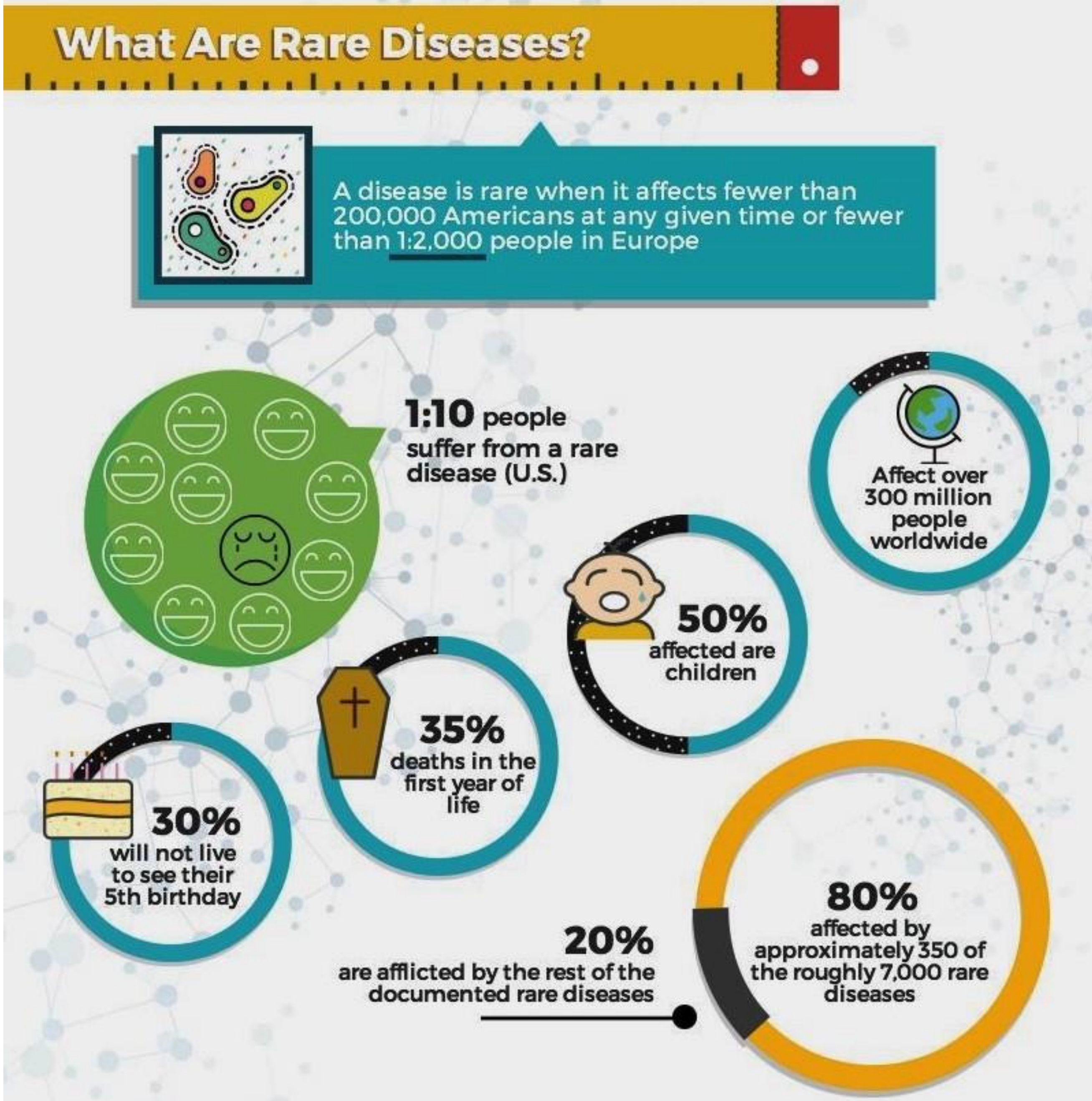
"Examining images or rank-ordered lists of features that distinguish individual profiles or clusters is tedious and lacks sensitivity for all but the most obvious of phenotypes, confirming that quantitative morphological profiling is more sensitive than the human visual system."

We'll elaborate on alternatives later today and during the course

Rohban et al. (2017)



Anne Carpenter



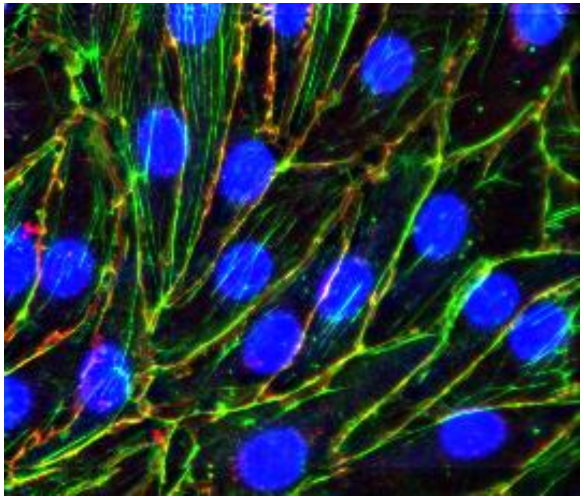
Anne Carpenter, Rare Genomics Institute

Image-based profiling can identify drugs for disease

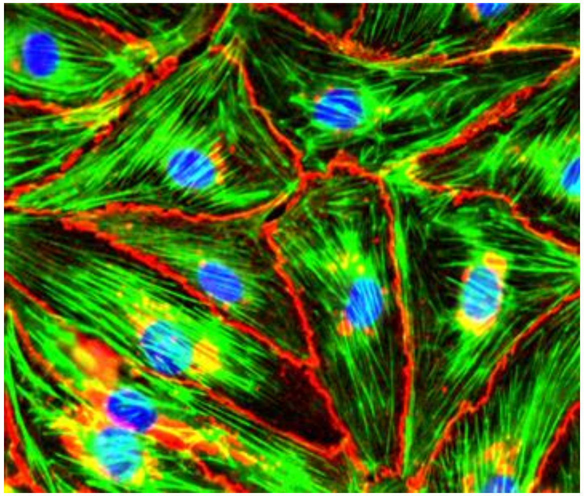
+ drug?

VE Cadherin
Actin
DNA

Healthy



Disease



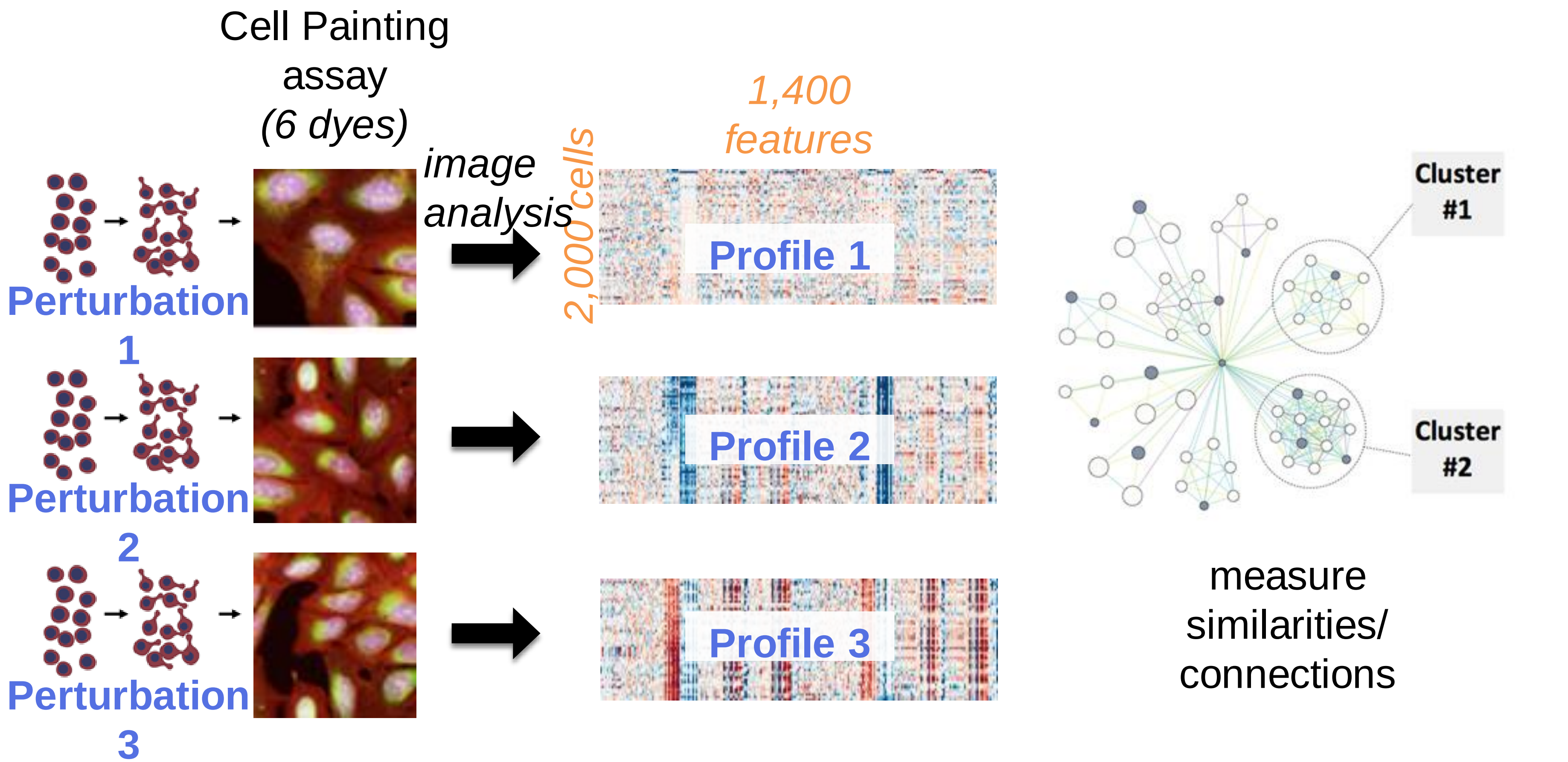
Cerebral cavernous malformation (CCM)

Drug chosen as hits based on automated analysis outperformed those chosen by expert visual analysis

Today: 300+ disease models available for screening in parallel

Anne Carpenter, Gibson, et al. (2015)

Signatures of genes, compounds and diseases



Anne Carpenter, Bray et al. (2016)

Identify small molecules mimicking genetic perturbations

Profile a
“query” gene

BRAF

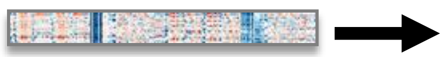
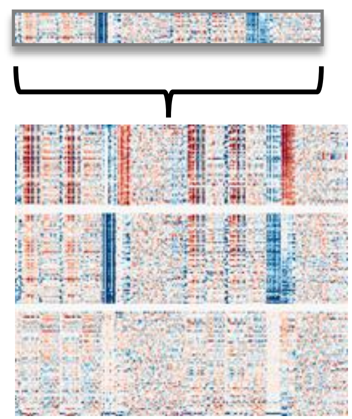
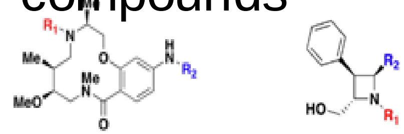


image
features

Identify similar-performing
compounds, based on
profile similarity



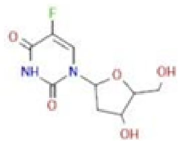
profiles for 30K
compounds



Anne Carpenter

Determine mechanisms of action (MoA)

Profile a
“query” compound



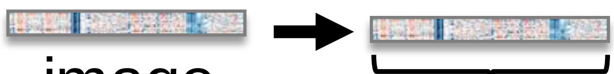
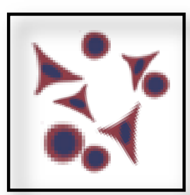
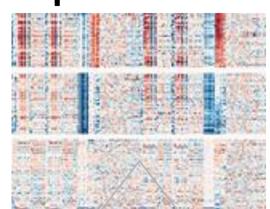
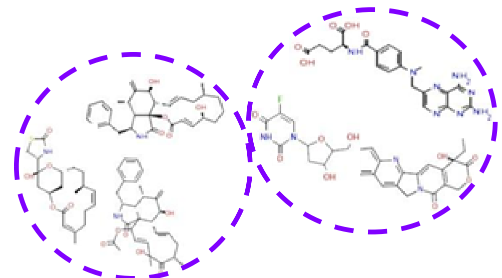


image
features





Compare to
profiles for well-
characterized
compounds



Goal: group
compounds into
clusters with
known compounds

Anne Carpenter, Gustafsdottir et al. (2013), Ljosa et al. (2013)

Lead hopping (the “opposite” of MoA)

Profile a “query” compound

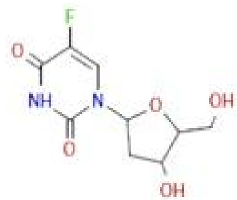
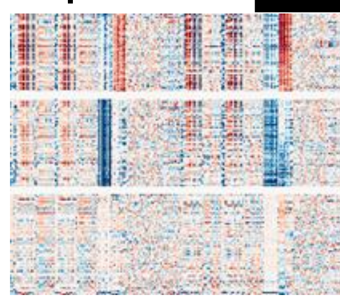
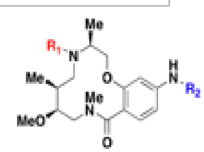


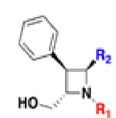
image features

Identify similar-performing compounds, based on profile similarity



Goal: novel chemical structures with desired phenotypic activity



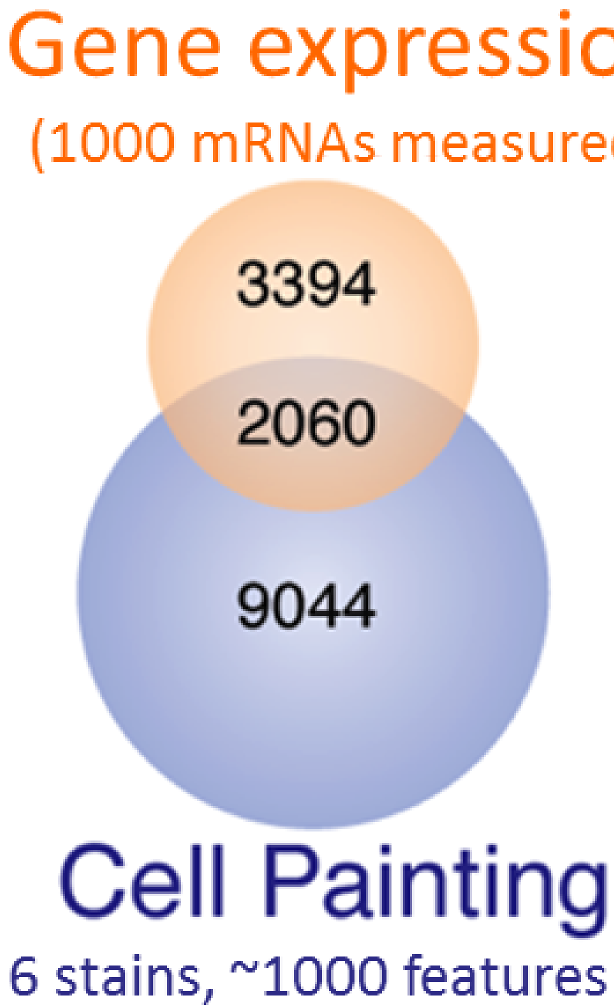


profiles for 30,000 compounds

Anne Carpenter

Profiling to enrich libraries

20,247 compounds profiled:



27% of compounds yield a detectable gene expression phenotype

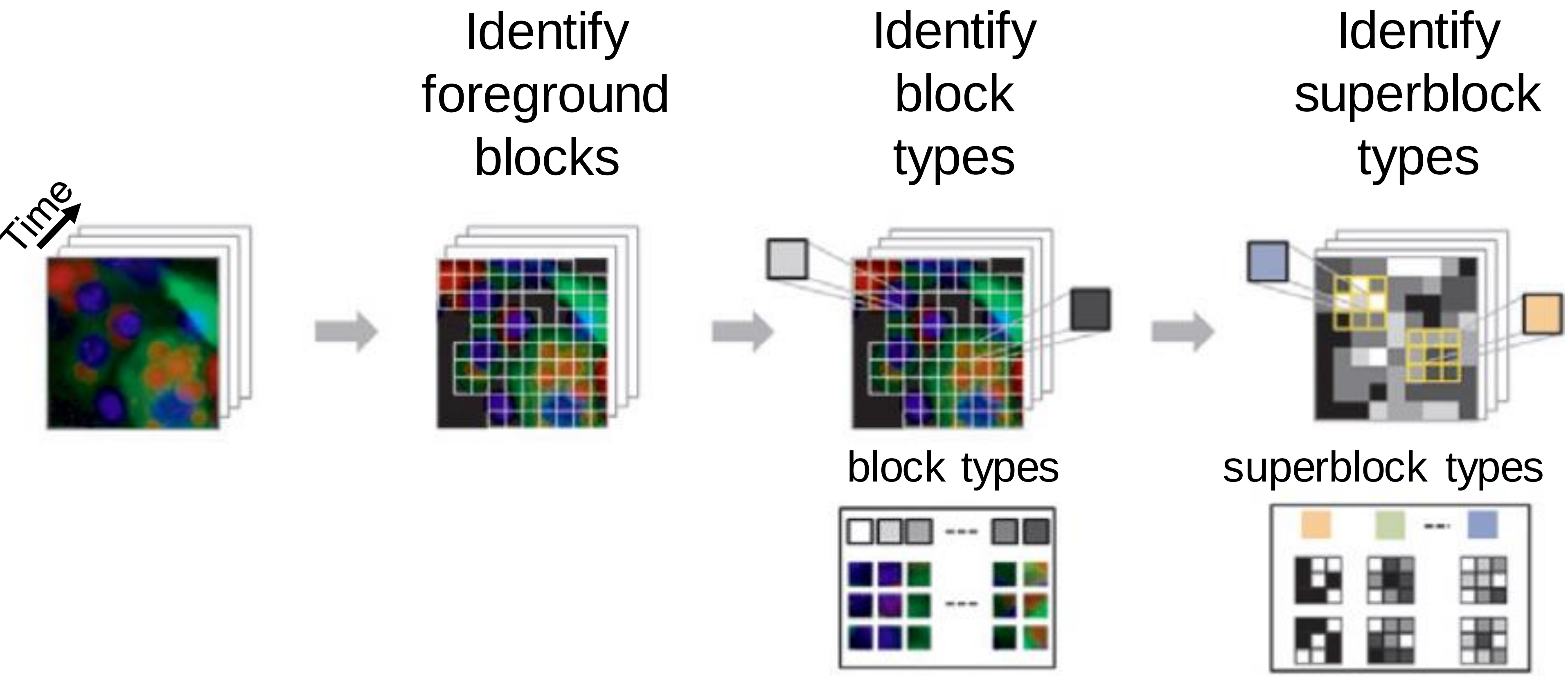
55% of compounds yield a detectable morphology phenotype

Anne Carpenter, Wawer et al. (2014)

Challenges (and opportunities) in phenotypic screening

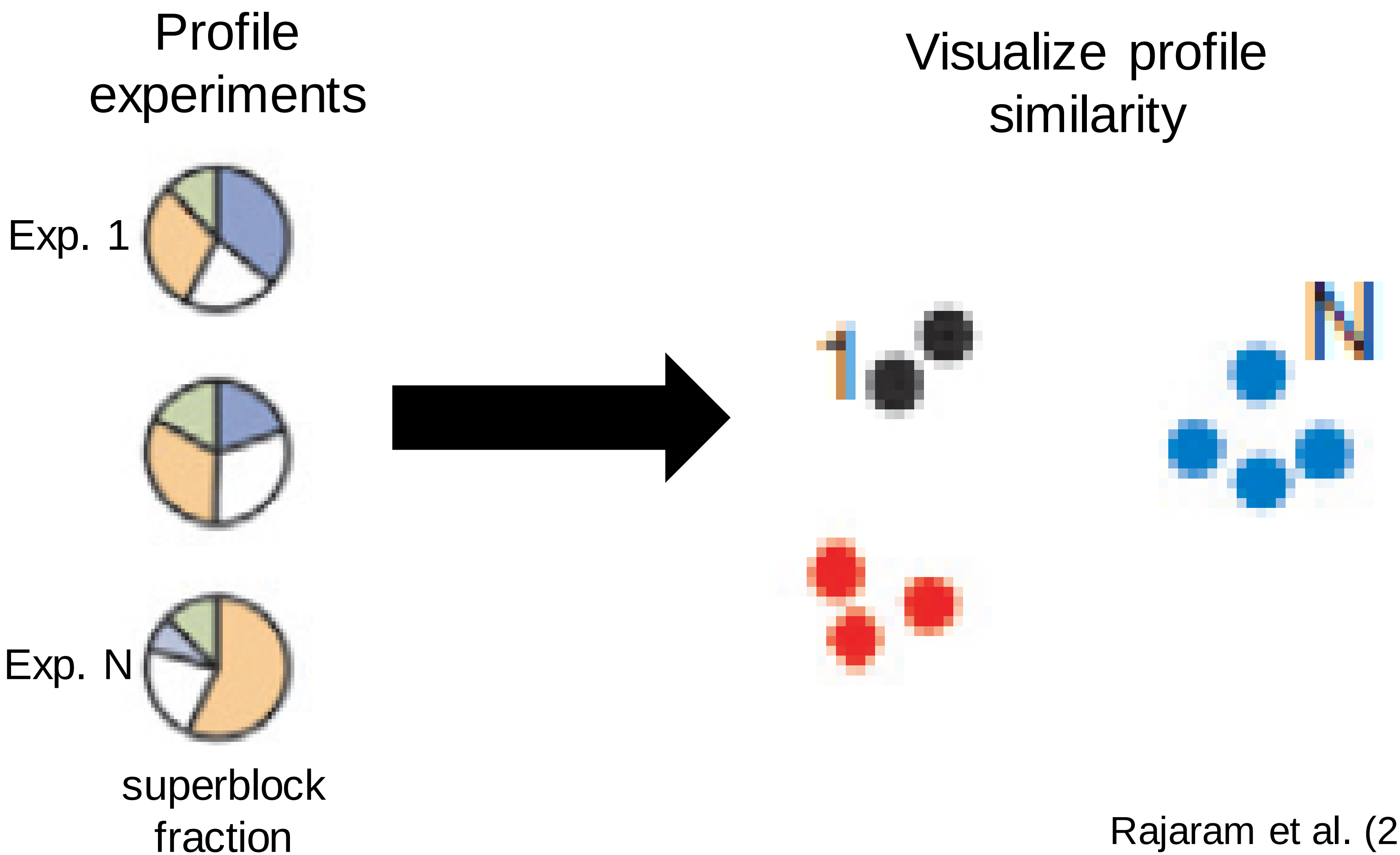
- Dealing with cell heterogeneity
- Defining better similarity measures between populations
- Interpretability
- 3D + microenvironment + time

Example: PhenoRipper, segmentation-free cell profiling



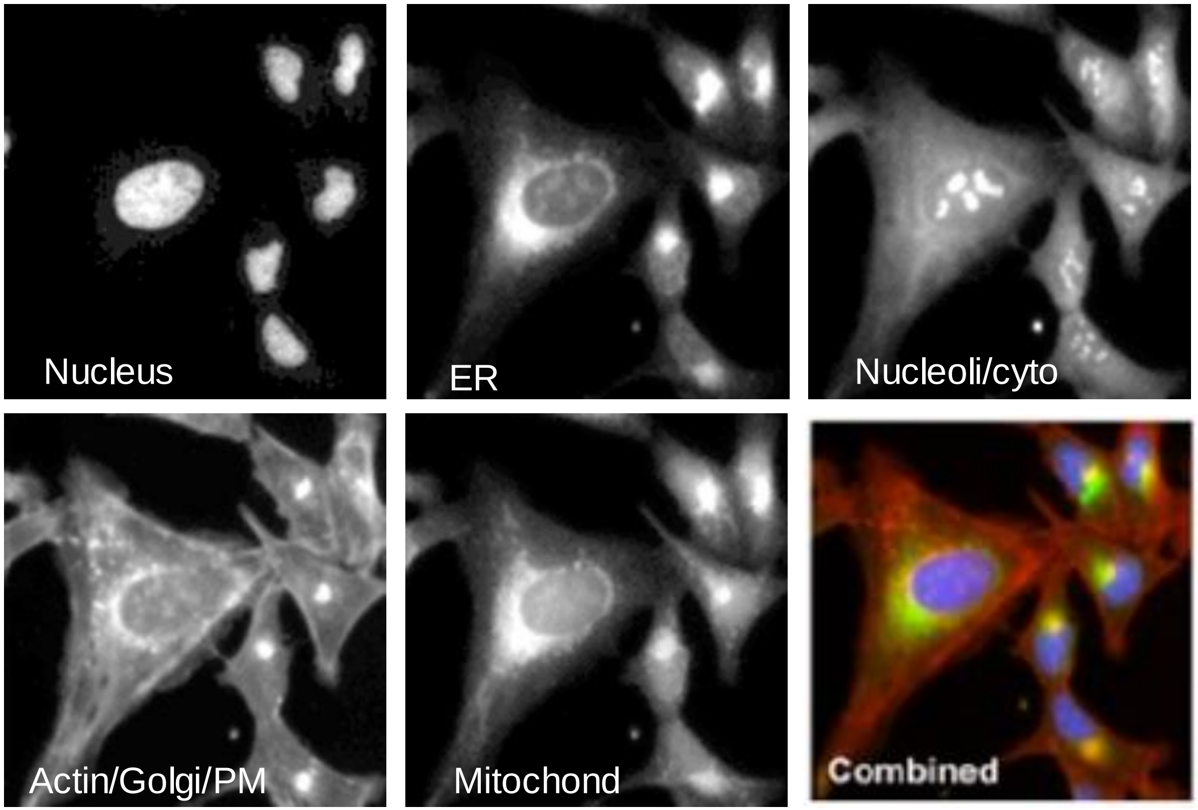
Rajaram et al. (2012)

PhenoRipper: segmentation-free cell profiling



Rajaram et al. (2012)

Example: cell painting



Cell Painting: 6 stains imaged in 5 channels reveal 8 cellular components

Extract signatures from each cell's image, then match these "profiles" to link drugs to genes to disease states

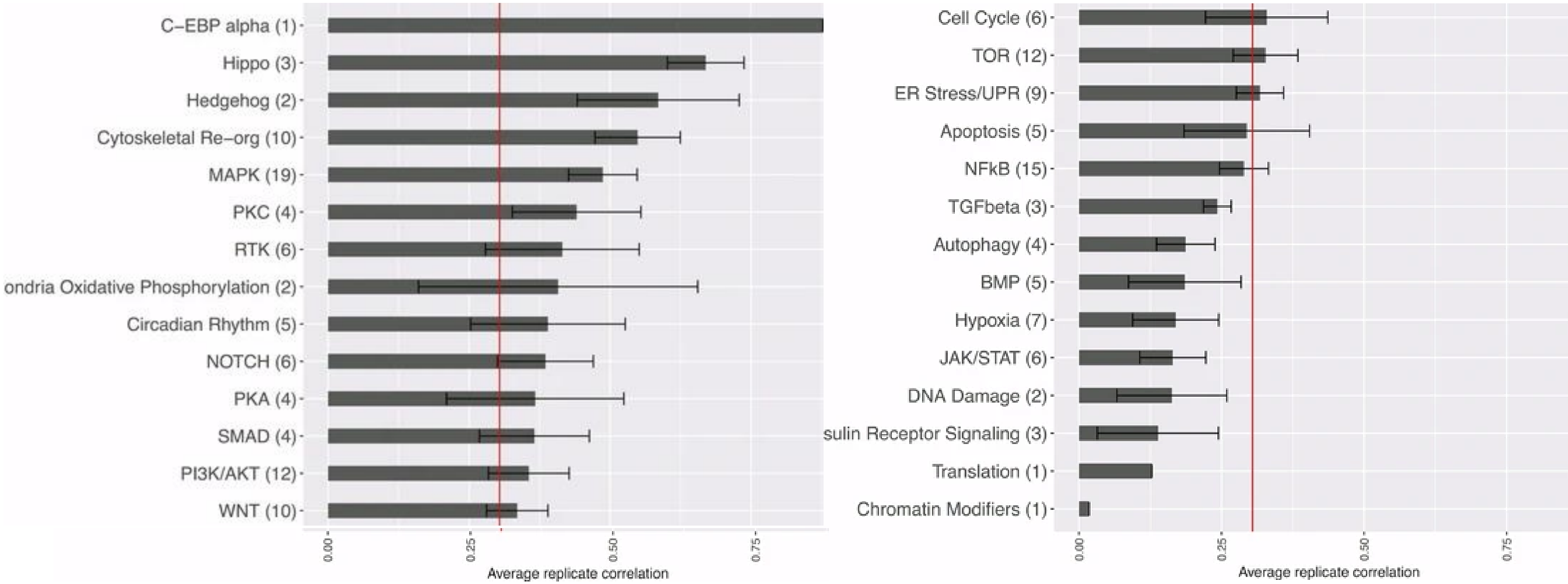
Example: match a drug to a CRISPR knockout to confirm a drug's target

Example: identify a signature in diseased patient cell lines and screen drugs to revert it

Anne Carpenter, Bray et al. (2016)

(Many) pathways can be interrogated by morphological profiling

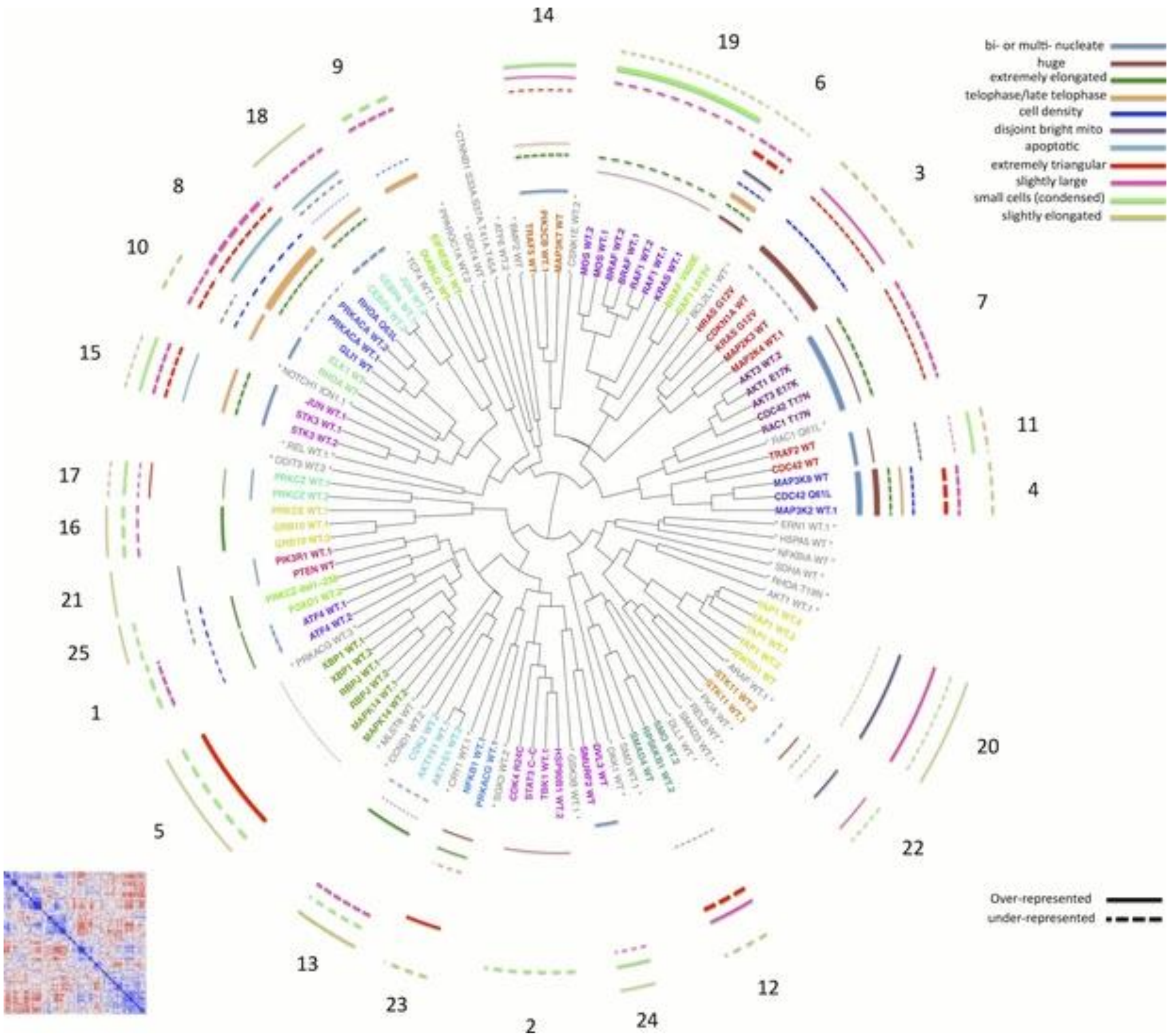
Overexpression screen



Correlation

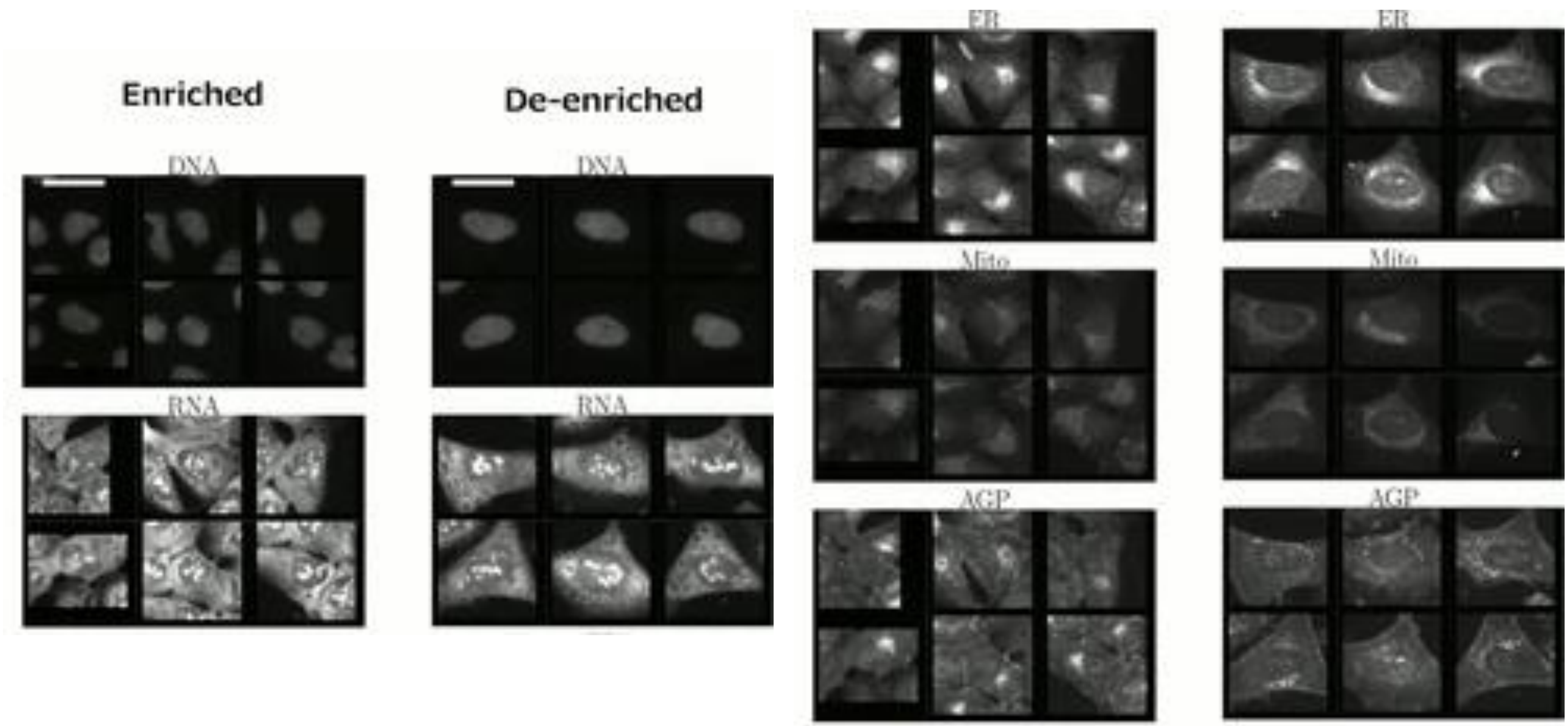
Rohban et al. (2017)

Morphological similarity captures known gene-gene relationships



Rohban et al. (2017)

Interpretation: sub-population clustering and visualization of each control-perturbed pair

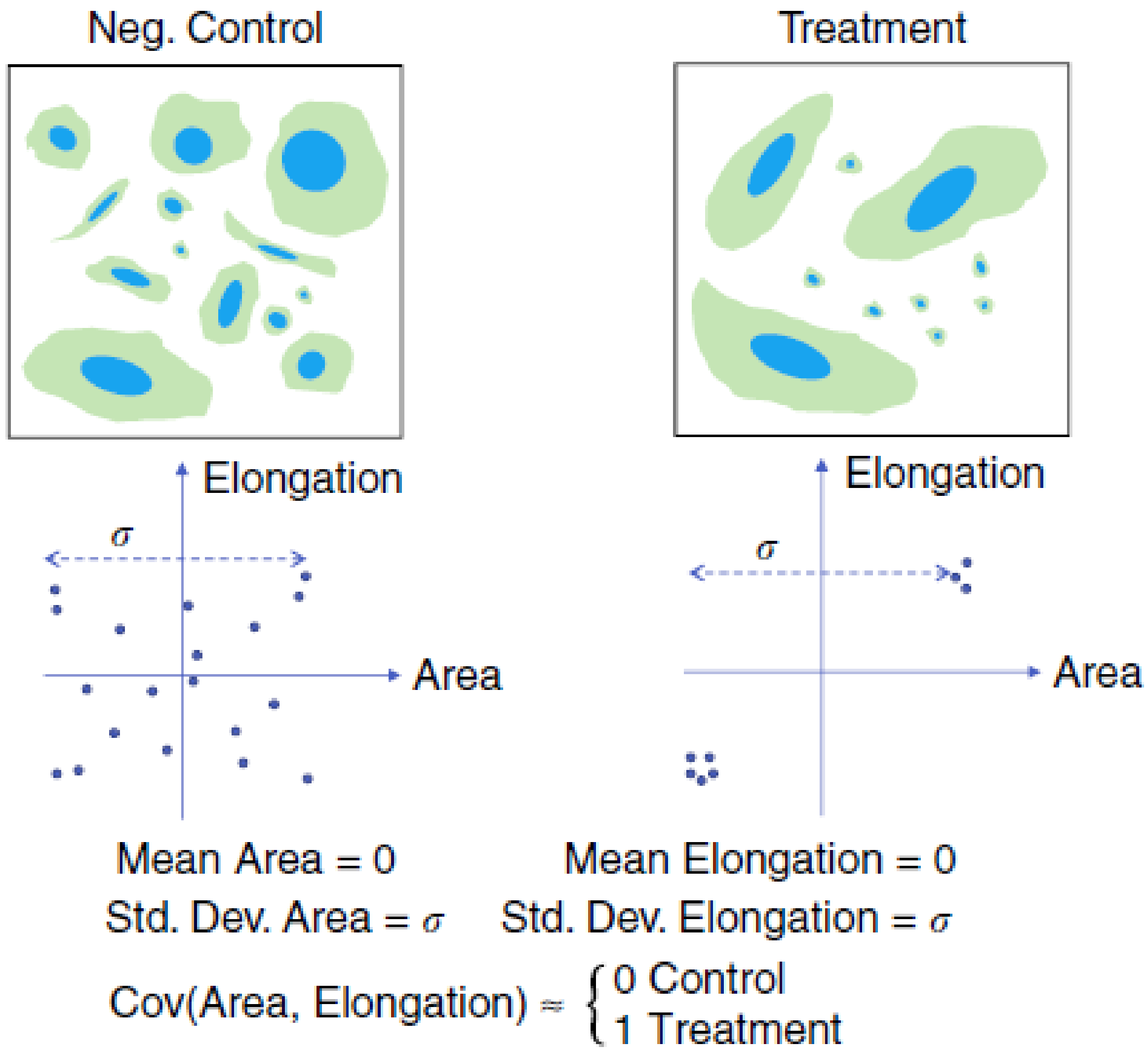


Rohban et al. (2017)

Cell painting – available datasets

- Bray et al. (2017), A dataset of images and morphological profiles of 30 000 small-molecule treatments using the Cell Painting assay. Data: https://github.com/gigascience/paper-bray2017
- Rohban et al. (2017) Systematic morphological profiling of human gene and allele function via Cell Painting. Data: http://idr.openmicroscopy.org/webclient/?show=screen-1751
- Gustafsdottir et al. (2013). Multiplex cytological profiling assay to measure diverse cellular states. Data: http://idr.openmicroscopy.org/webclient/?show=screen-1952

Cell painting: including dispersion and covariances to population averages



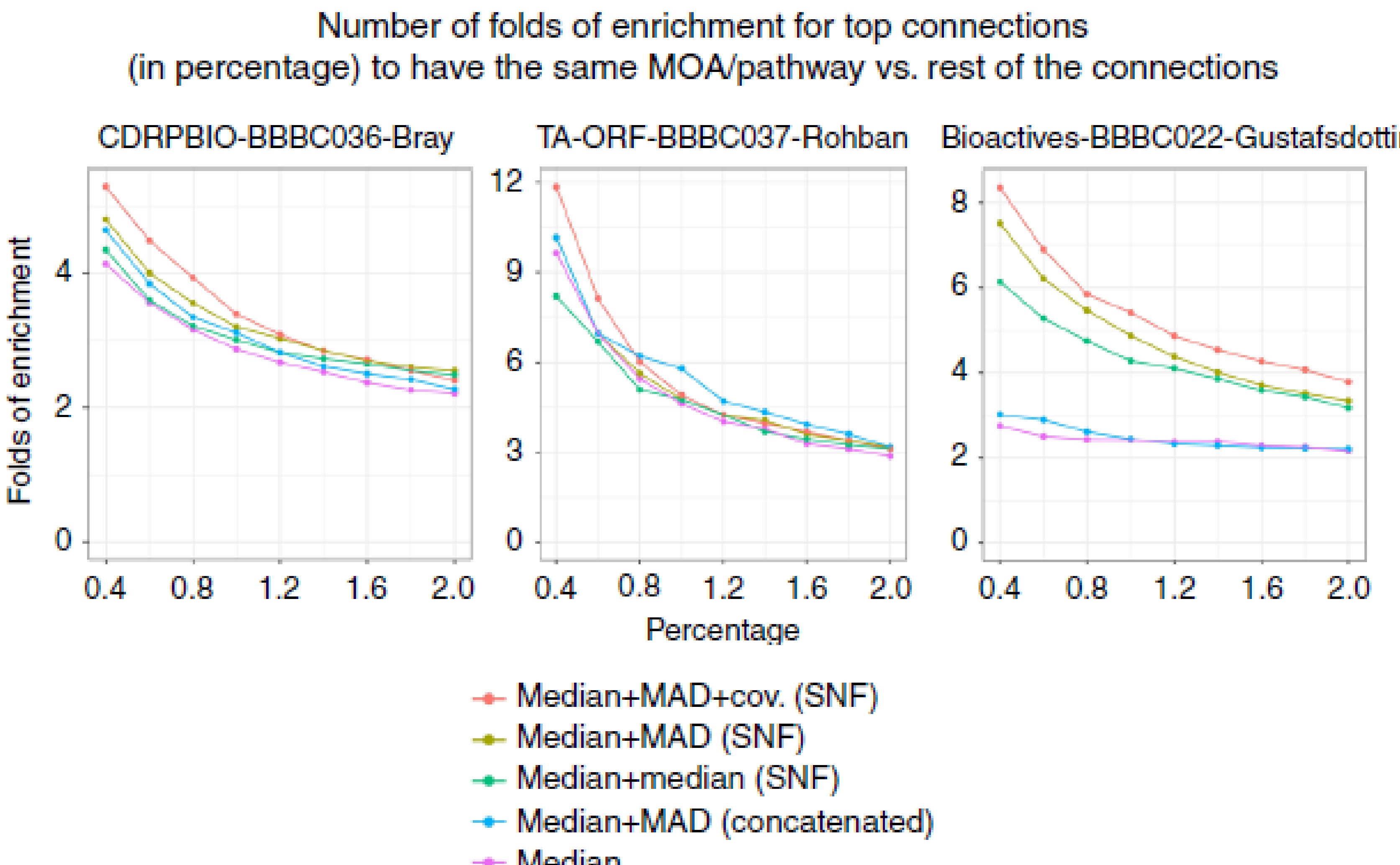
Rohban et al. (2019)

Fused profile similarities improves performance - validation

"Do pairs of cell populations that look most alike, according to the computed image-based profiles, have been treated with perturbations that are annotated as having the same mechanism of action (for compounds) or the same pathway (for gene overexpressions)?"

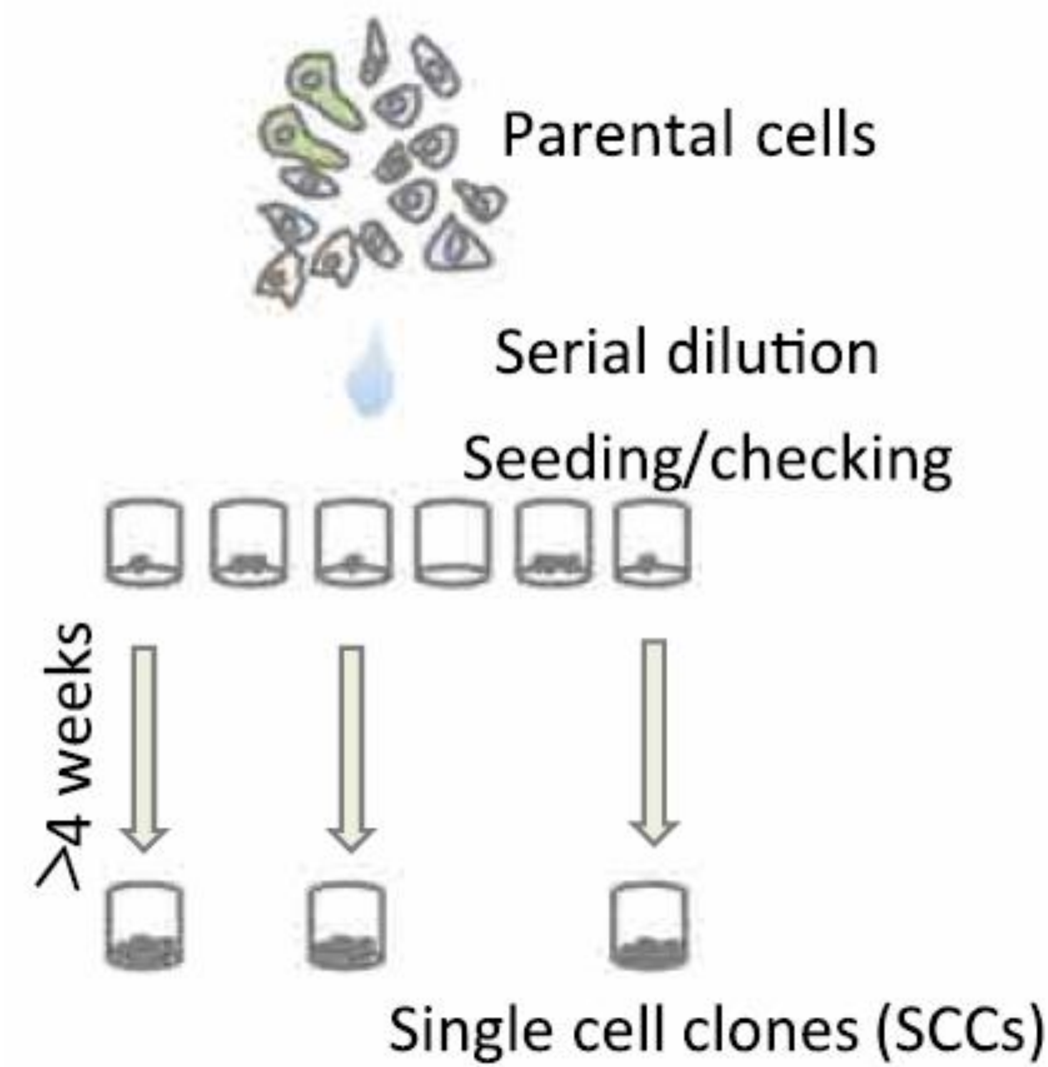
Rohban et al. (2019)

Fused profile similarities improves performance - validation

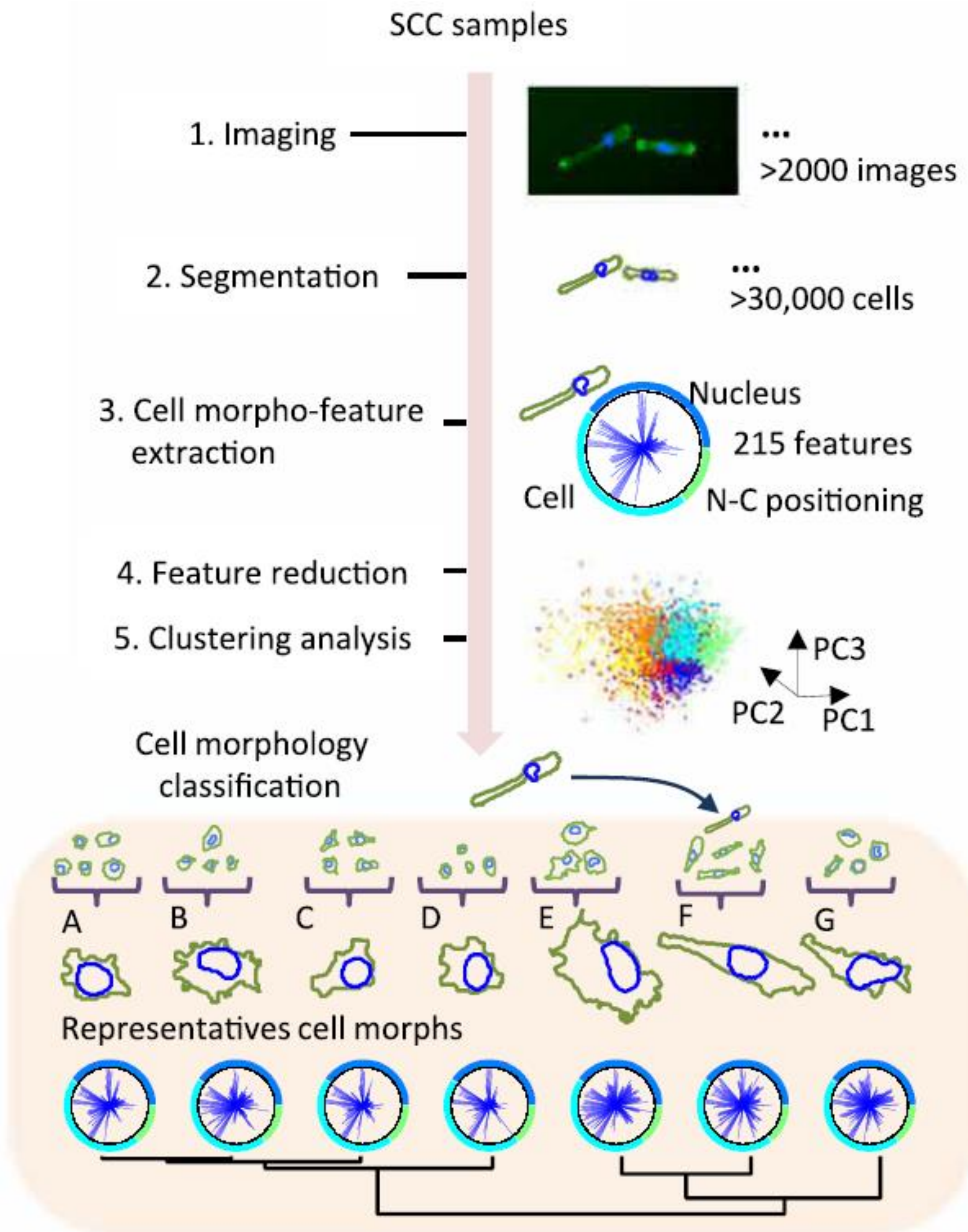


Rohban et al. (2019)

Example: single-cell morphology and metastatic potential



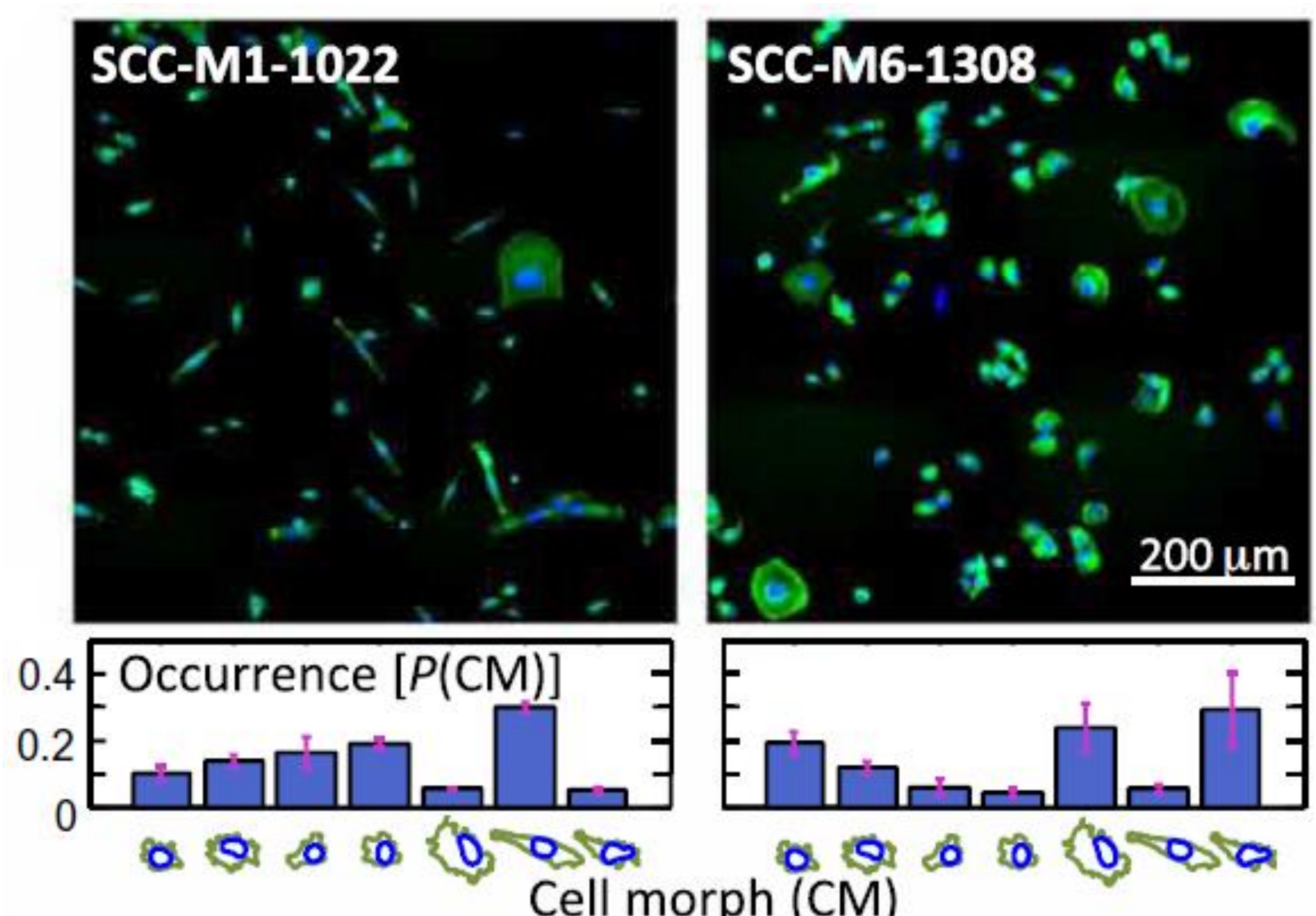
Wu et al. (2020)



SCC samples
1. Imaging
>2000 images
2. Segmentation
>30,000 cells
3. Cell morpho-feature extraction
Nucleus
215 features
Cell
N-C positioning
4. Feature reduction
5. Clustering analysis
PC3
PC2
PC1
Cell morphology classification
A
B
C
D
E
F
G
Representatives cell morphs

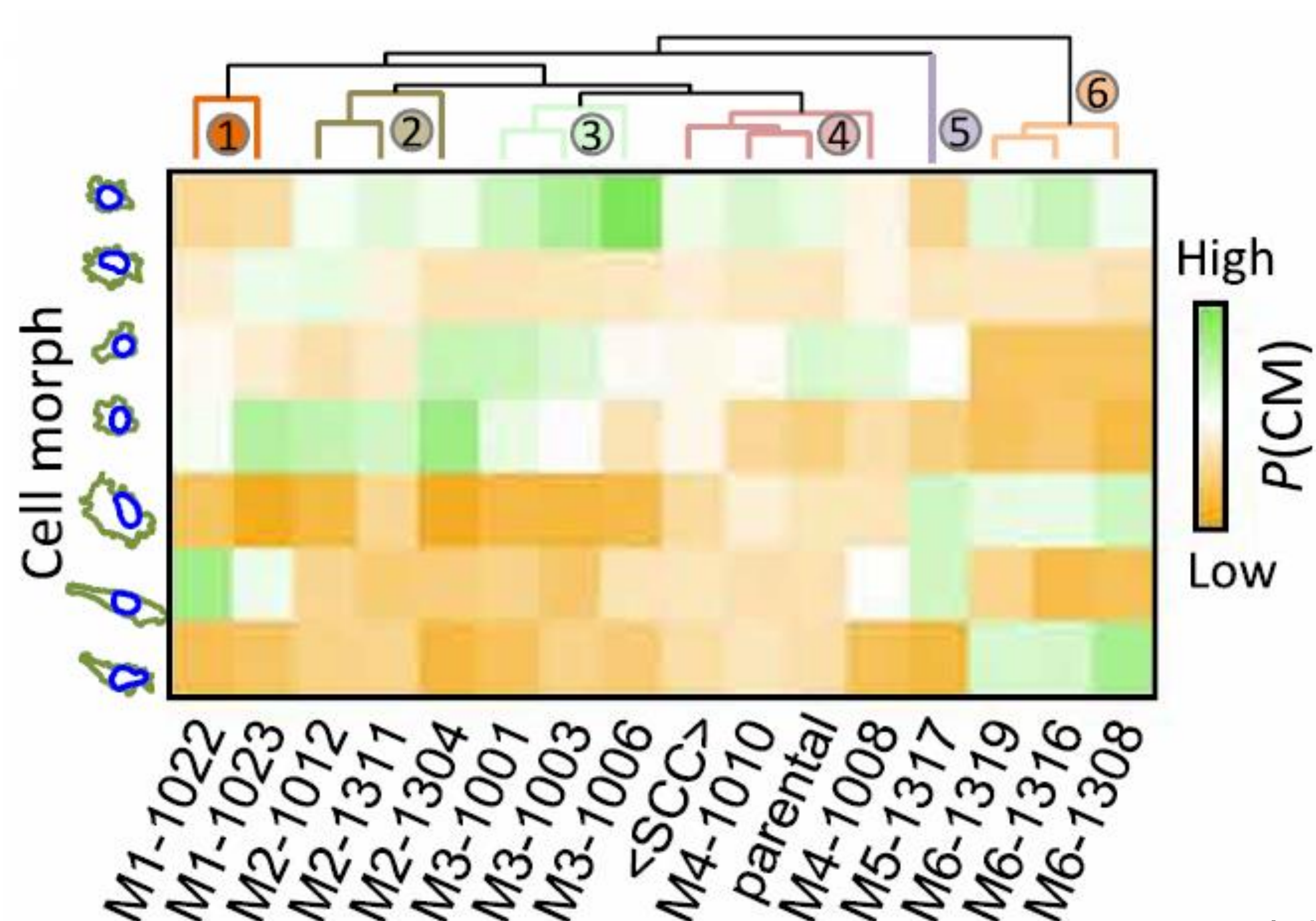
Wu et al. (2020)

Quantitative representation of cell populations



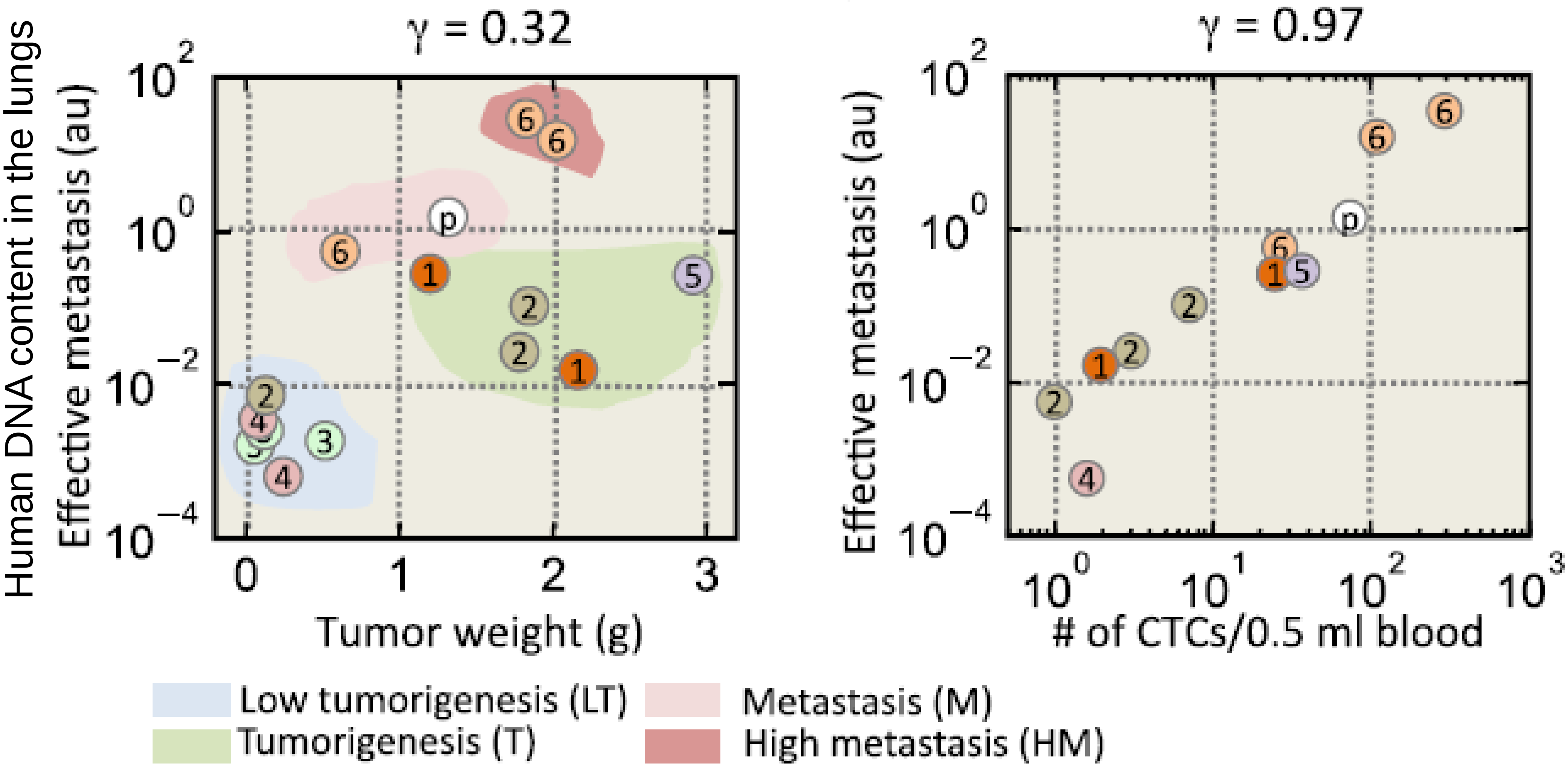
Wu et al. (2020)

Shape-clustering



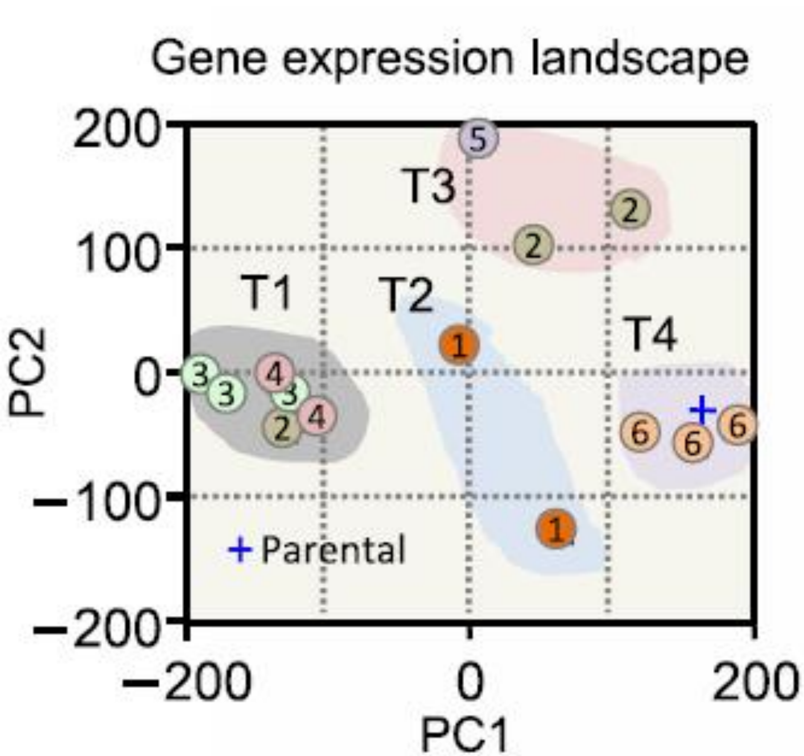
Wu et al. (2020)

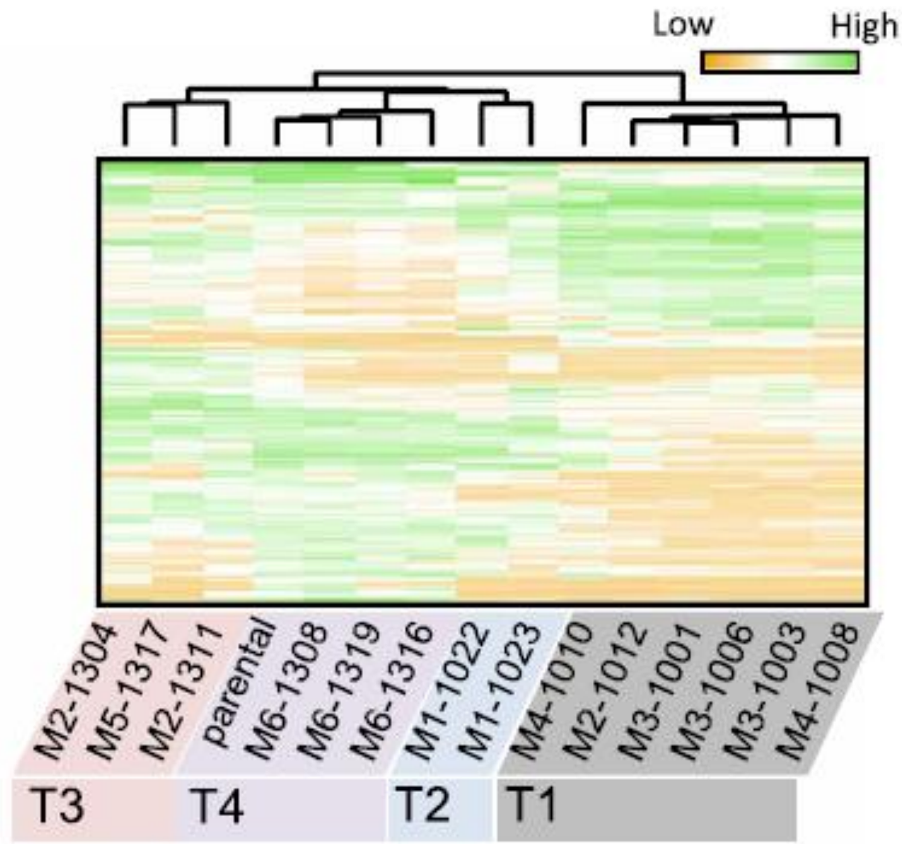
Morphological phenotypes in vitro and differential tumor progression in vivo



Wu et al. (2020)

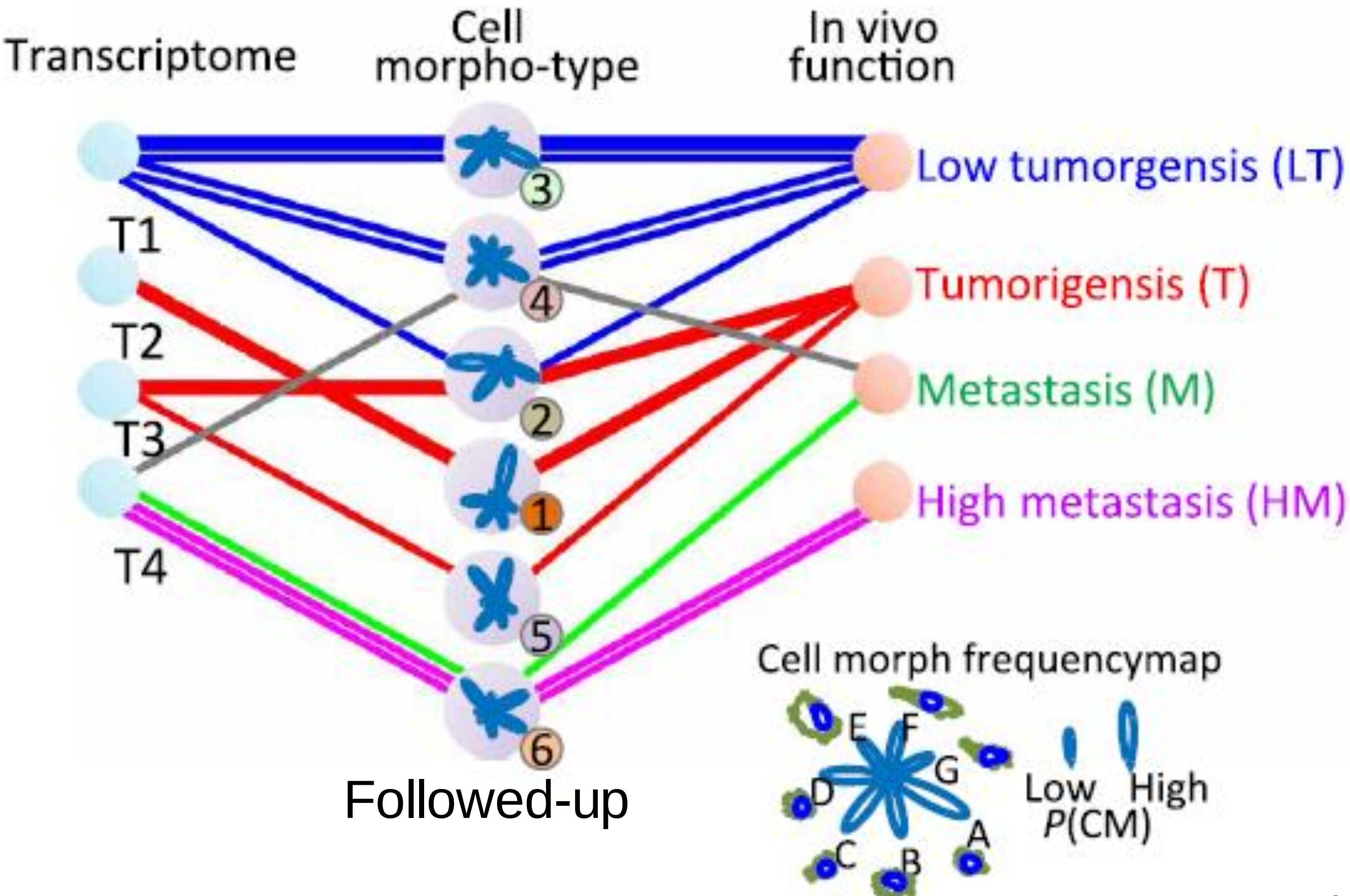
Morphology correlates with gene expression patterns





Wu et al. (2020)

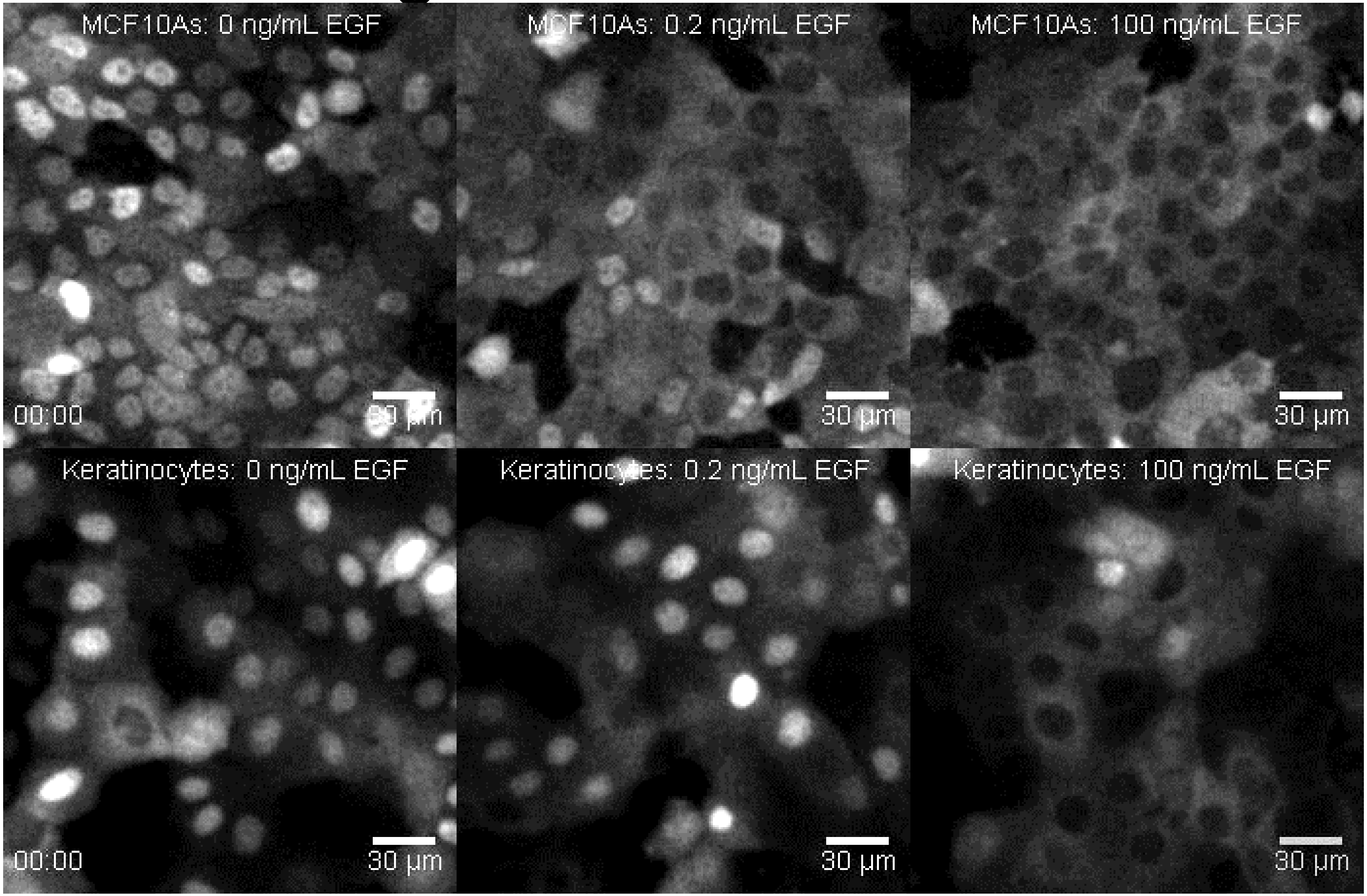
Morphology correlates with gene expression patterns



Followed-up

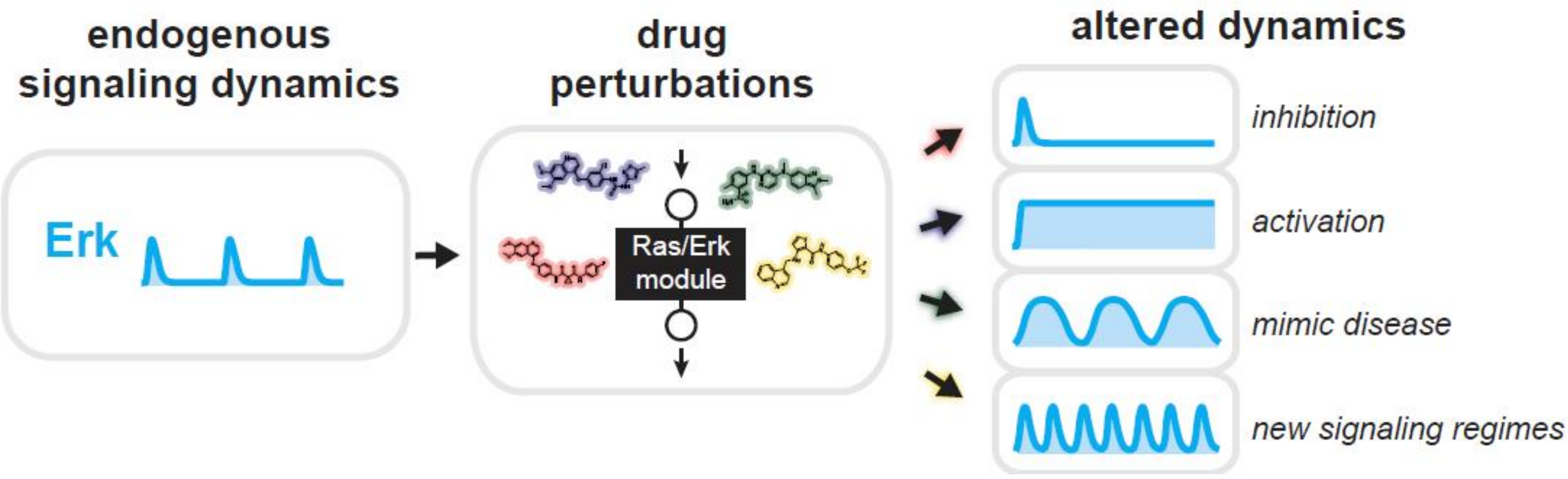
Wu et al. (2020)

Example: high-throughput screens using live-cell biosensor



Goglia et al. (2020)

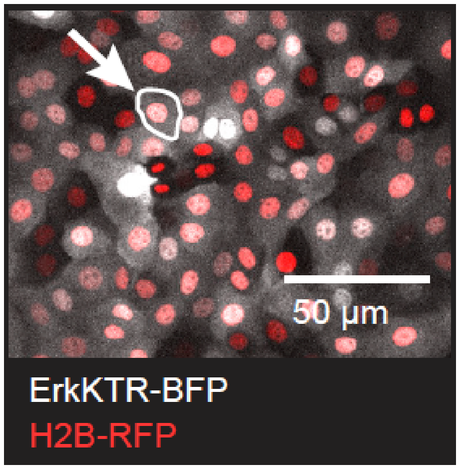
High-throughput screening for altered extracellular-regulated kinase (Erk) dynamics



Goglia et al. (2020)

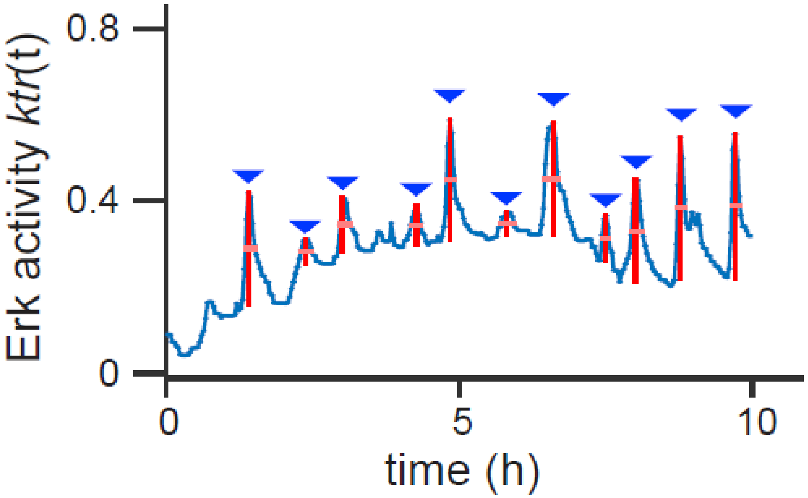
Imaging processing pipeline

KTR-H2B keratinocytes



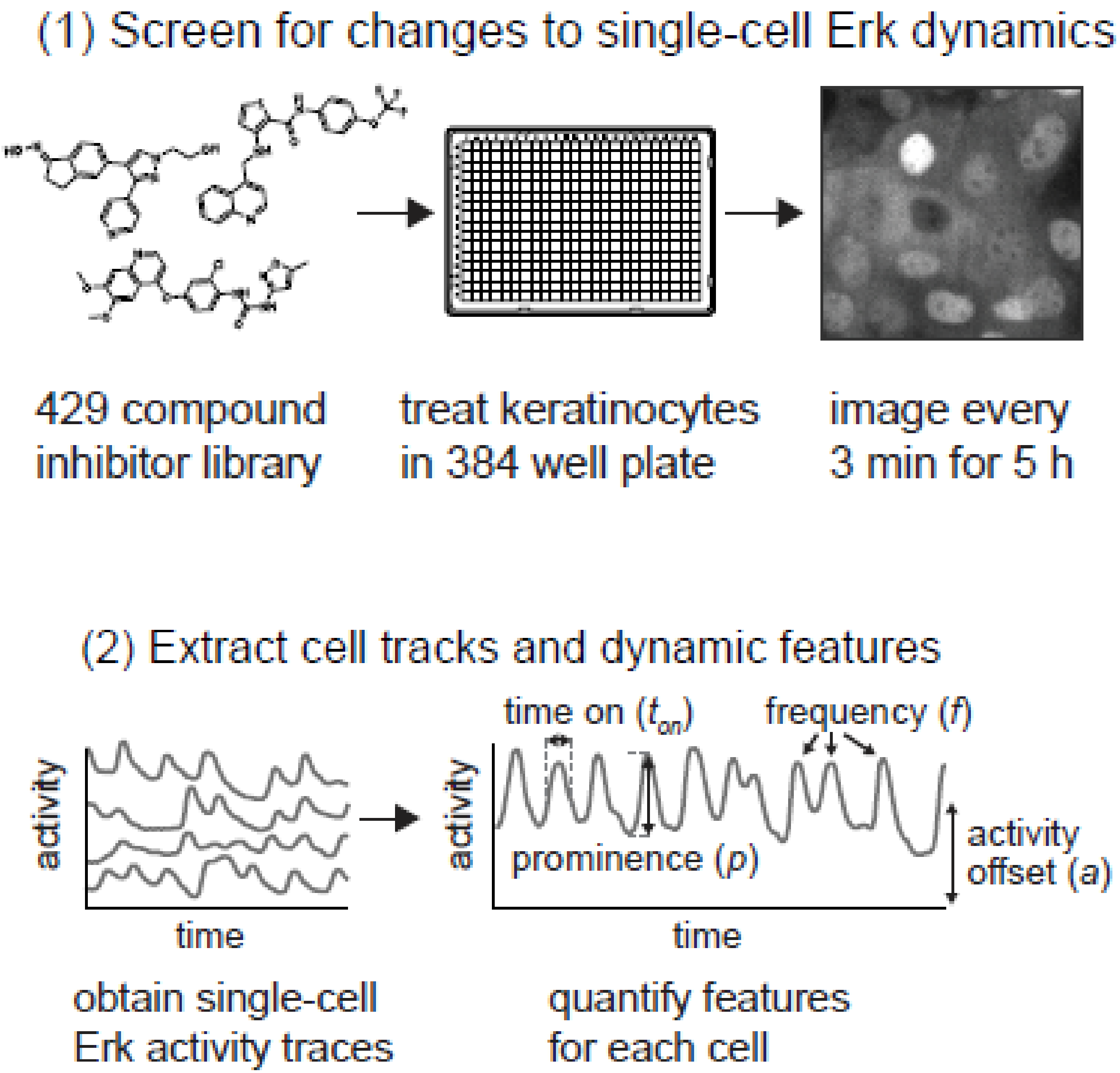
1. segment nuclei using TrackMate
2. measure nuclear ErkKTR-BFP
3. single-cell peak-finding to extract dynamic info.

single-cell Erk activity over time



Goglia et al. (2020)

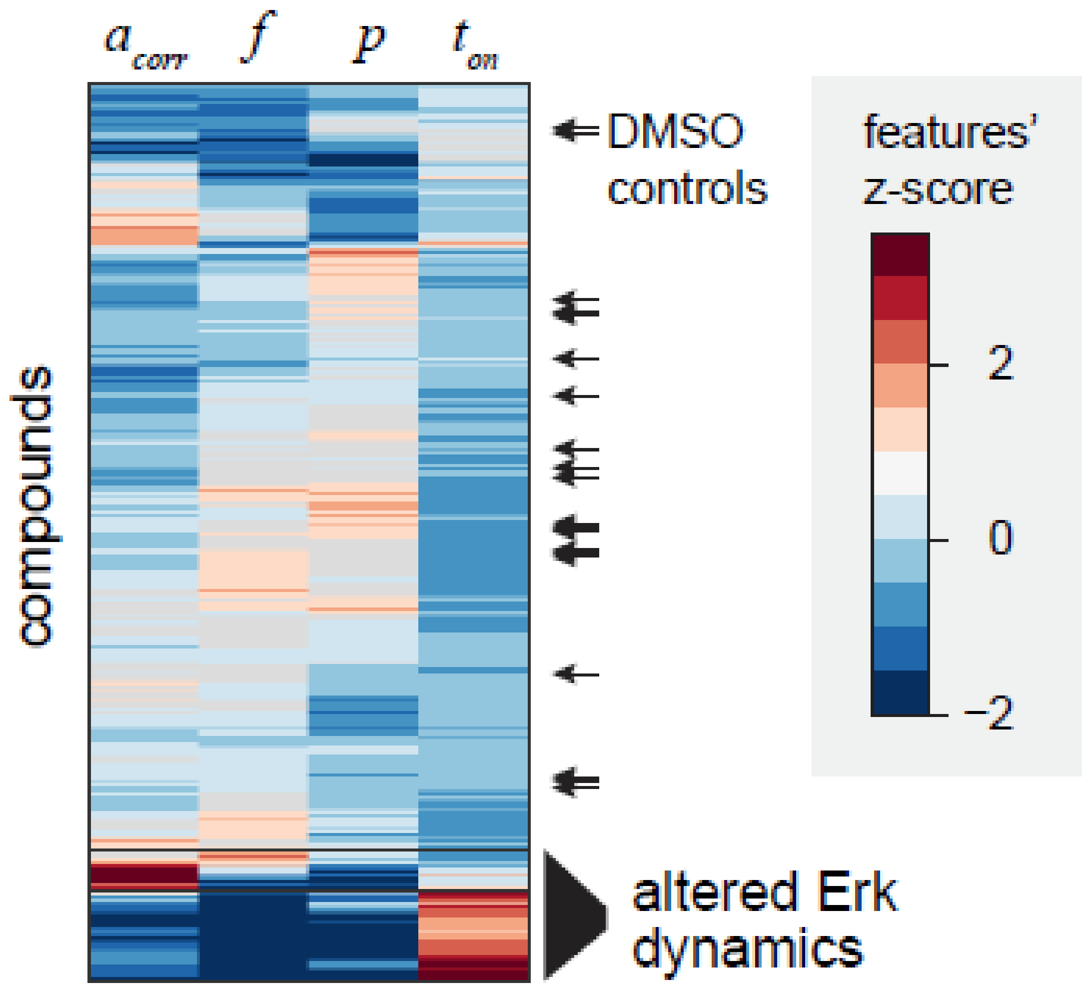
Screening



+ density
+ displacement

Goglia et al. (2020)

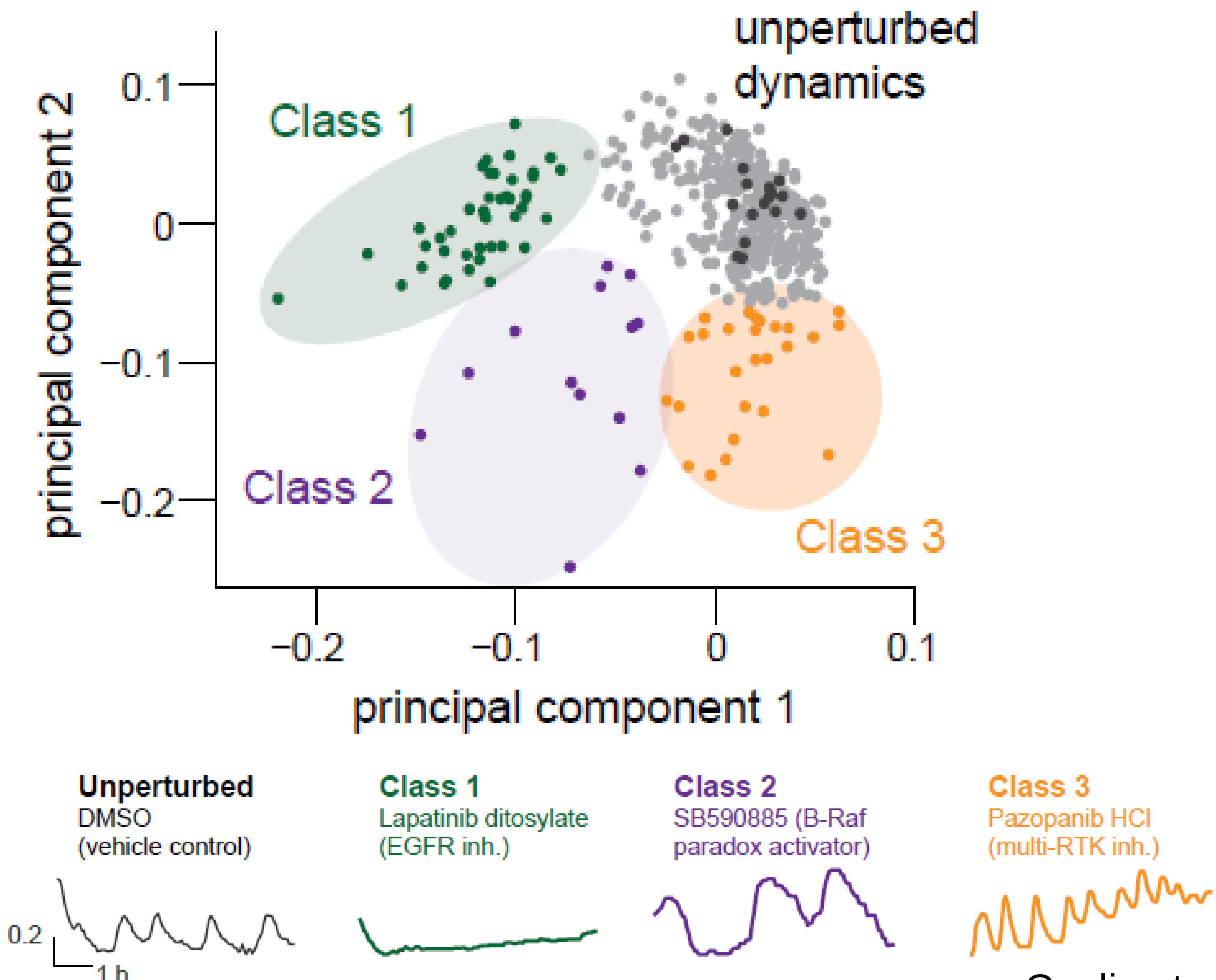
Screen results



429 kinase inhibitors, 80K cells (5 hours each at 3 min. intervals)

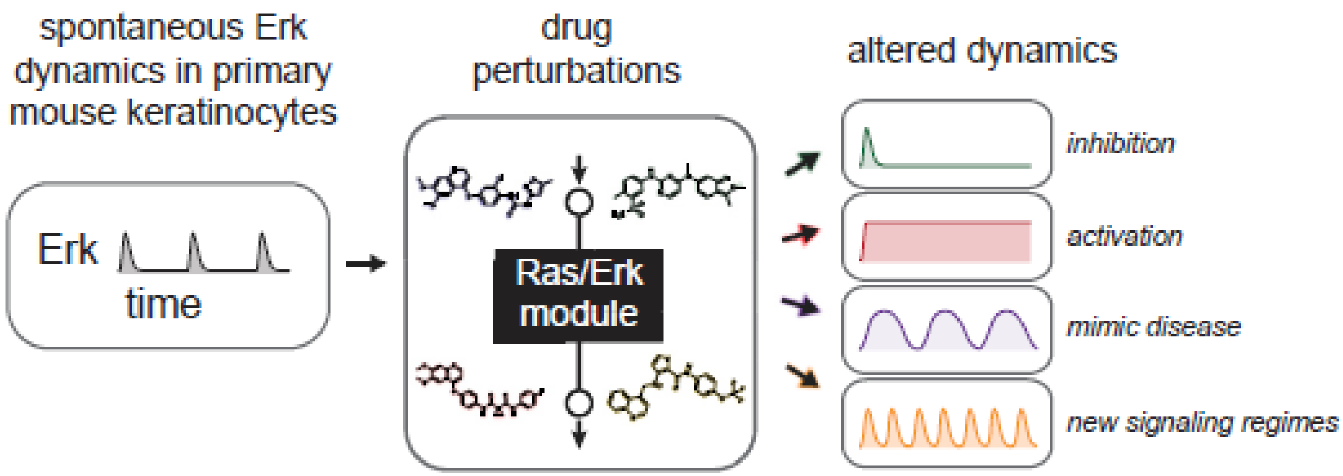
Goglia et al. (2020), raw data @IDR, processes data @Toettcher

Interpretation of "hits"



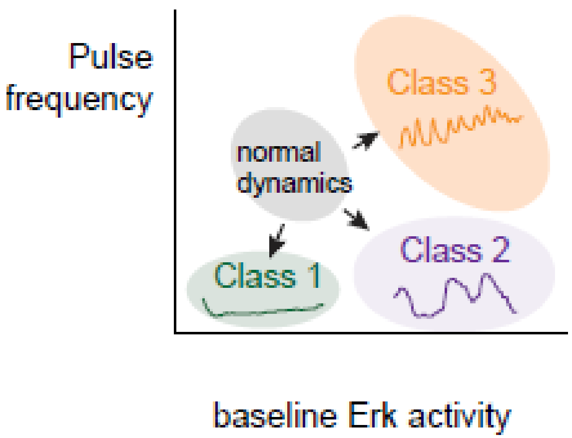
Goglia et al. (2020)

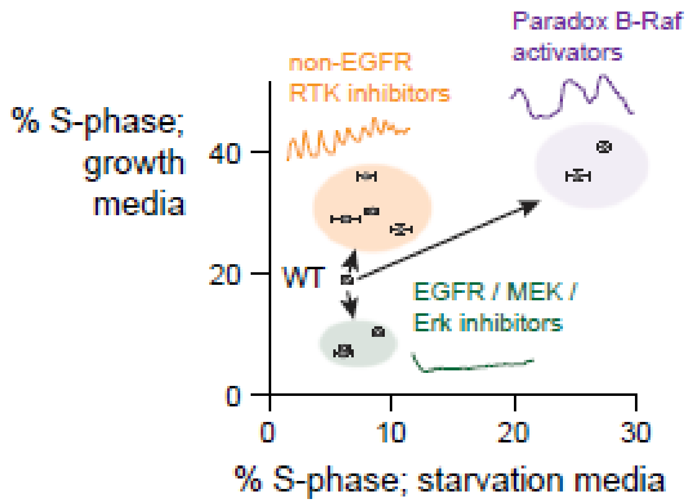
Conclusions (after follow-ups)



Kinase inhibitors alter Erk dynamics

Altered Erk dynamics regulate cell proliferation





Goglia et al. (2020)

Additional open (screening) datasets for your research projects

- Image data resource (IDR), Williams et al. (2017), Image Data Resource: a bioimage data integration and publication platform https://idr.openmicroscopy.org/
- Pascual-Vargas et al. (2017), RNAi screens for Rho GTPase regulators of cell shape and YAP/TAZ localisation in triple negative breast cancer Data via IDR
- Pizzagalli et al (2018), Leukocyte Tracking Database, a collection of immune cell tracks from intravital 2-photon microscopy videos (via figshare)
- Brenda Andrews lab resources: https://thecellvision.org/ , http://sites.utoronto.ca/andrewslab/data.shtml
- The Human Protein Atlas, https://www.proteinatlas.org/
- The Allen Institute of Cell Science