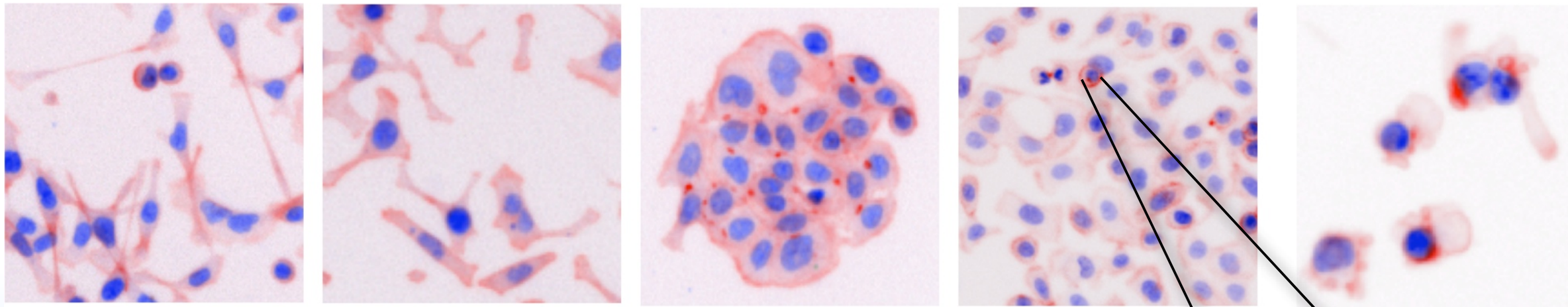
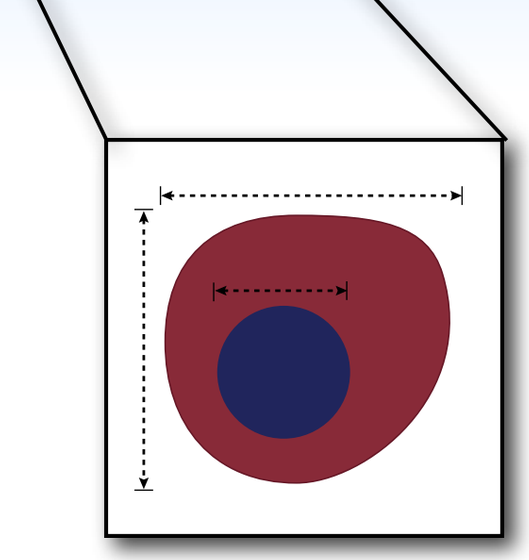


Data visualization for two-dimensional high-throughput microscopy screens

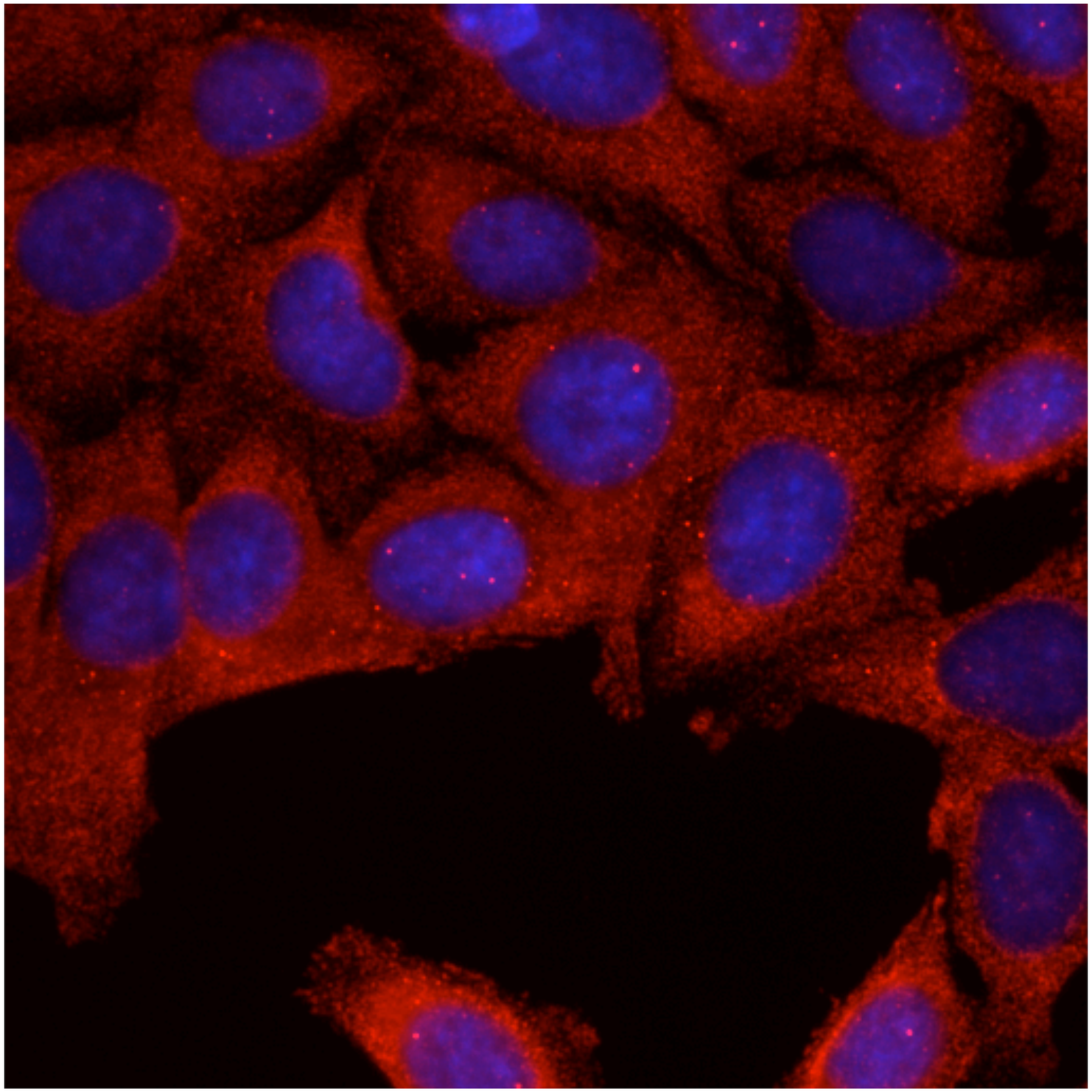
Anne Carpenter, Ph.D.

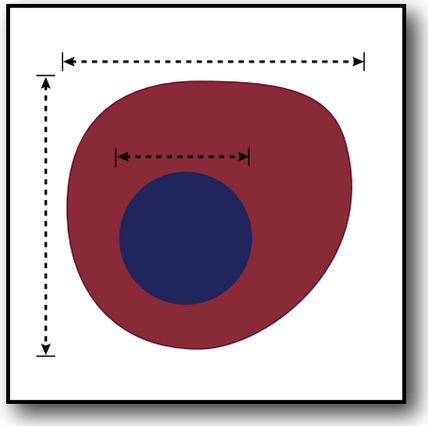


0.4233
54,454
45.777
0.6886
0.0055
6.9994
83.333
14.113
1.5567
0.0954 ...
0.5553

BROAD INSTITUTE IMAGING PLATFORM at Harvard and MIT

Visualizing 2D microscopy ~~images~~ image data



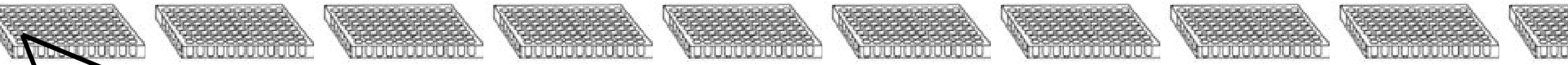


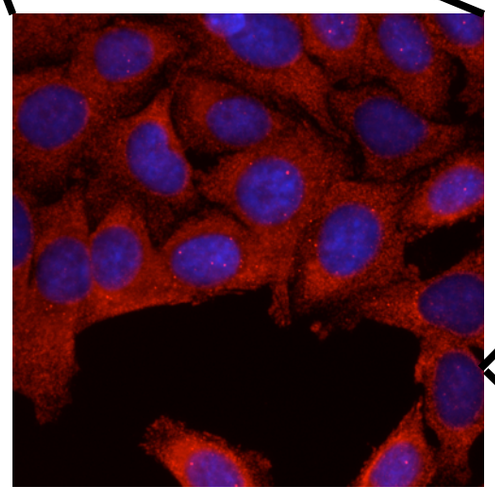
0.4233
54,454
45.777
0.6886
0.0055
6.9994
83.333
14.113
1.5567
0.0954
0.5553

"It is not surprising that the field is somewhat lacking in visual analysis tools that can gracefully and intuitively present rich image data representing possibly hundreds of attributes extracted from billions of cells."

-- Walter, et al., "Visualization of image data from cells to organisms", Nature Methods (in press)

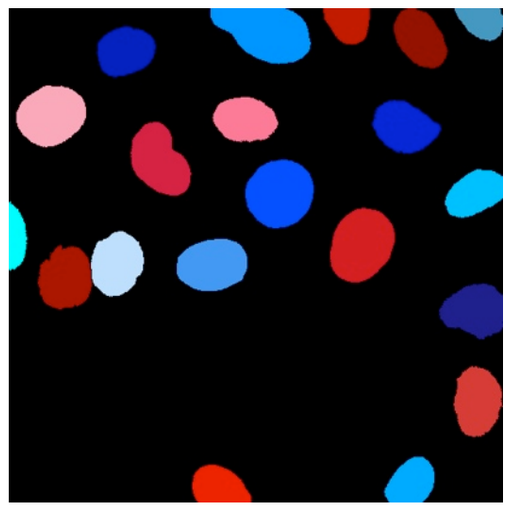
Cells or organisms in multiwell plates, each well treated with a gene or chemical perturbant

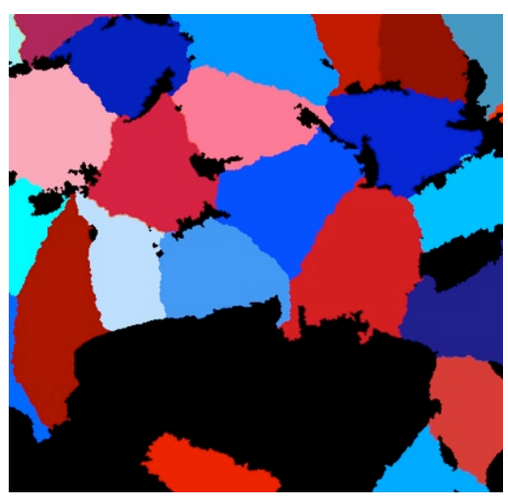




automated microscopy (any manufacturer)

thousands - millions of images



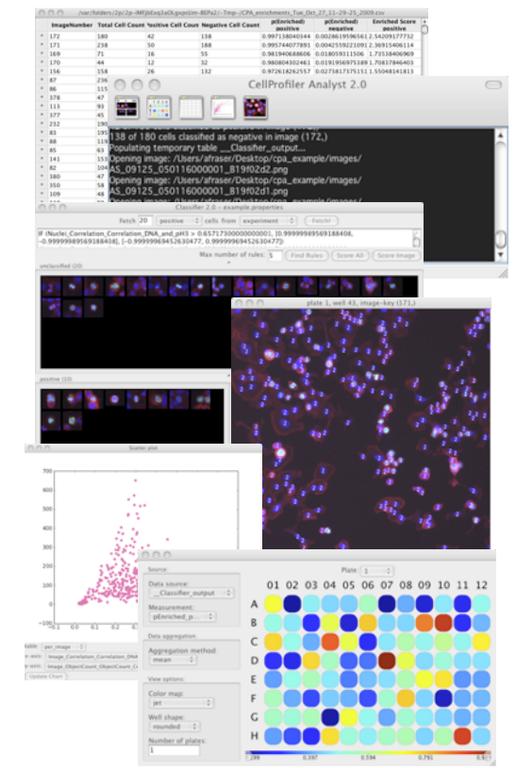


CellProfiler™ cell image analysis software

free, at www.cellprofiler.org

Cell measurements

(size, shape, intensity, texture, etc.)



Data exploration & machine learning

CellProfiler™ Analyst data exploration software

Ray Jones

BROAD INSTITUTE IMAGING PLATFORM

The team

Anne Carpenter

Image assay development

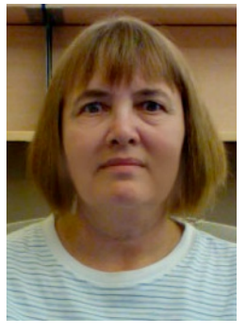
Apply image analysis methods to biological questions

David Logan

Mark Bray

Kate Madden

IT/Administration



Peggy (Margaret) Anthony

Algorithm development & software engineering

Develop & test new image analysis and data mining methods and create open-source software tools

Ray Jones

Vebjørn Ljoså

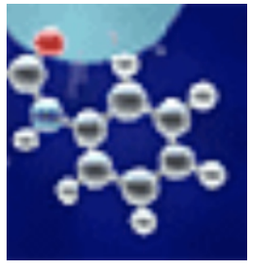
Adam Fraser

Lee Kamentsky

Carolina Wählby

Large-scale image-based science

knock down every gene in the genome by RNAi



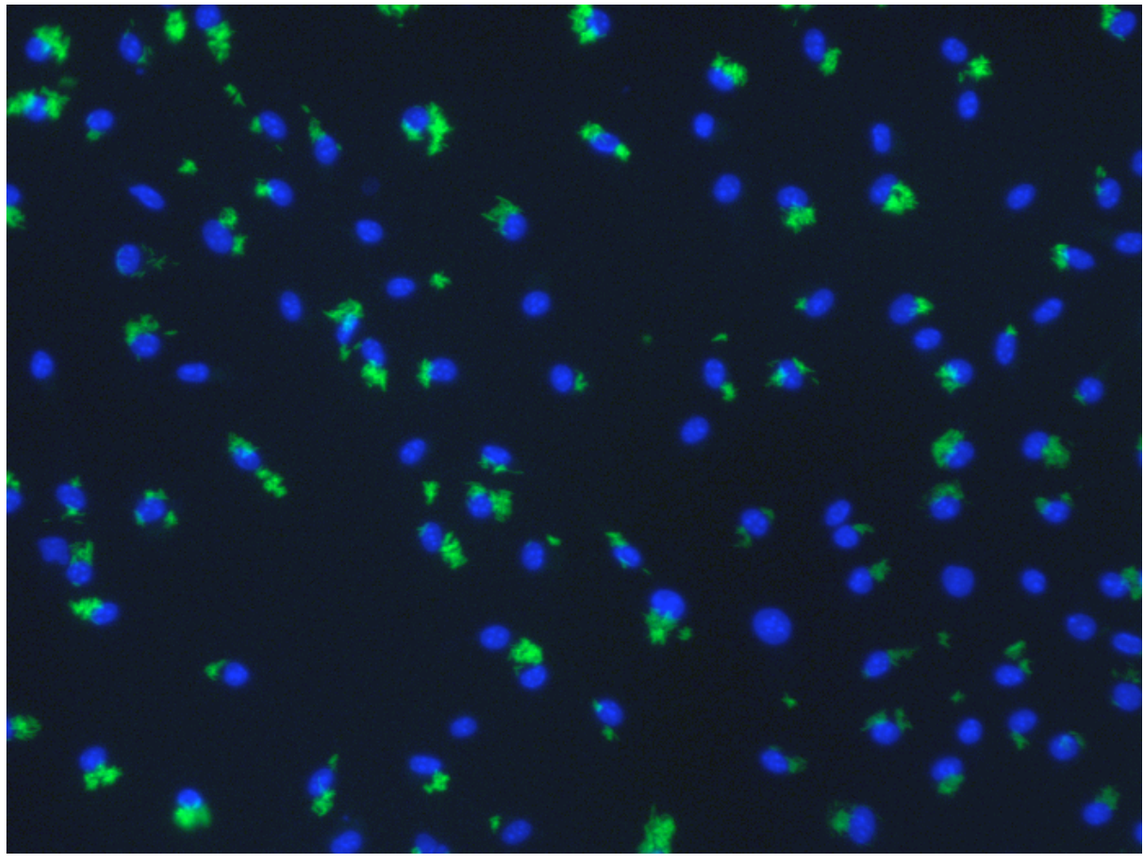
treat with small molecules from a chemical library

"look" for samples with unusual phenotype

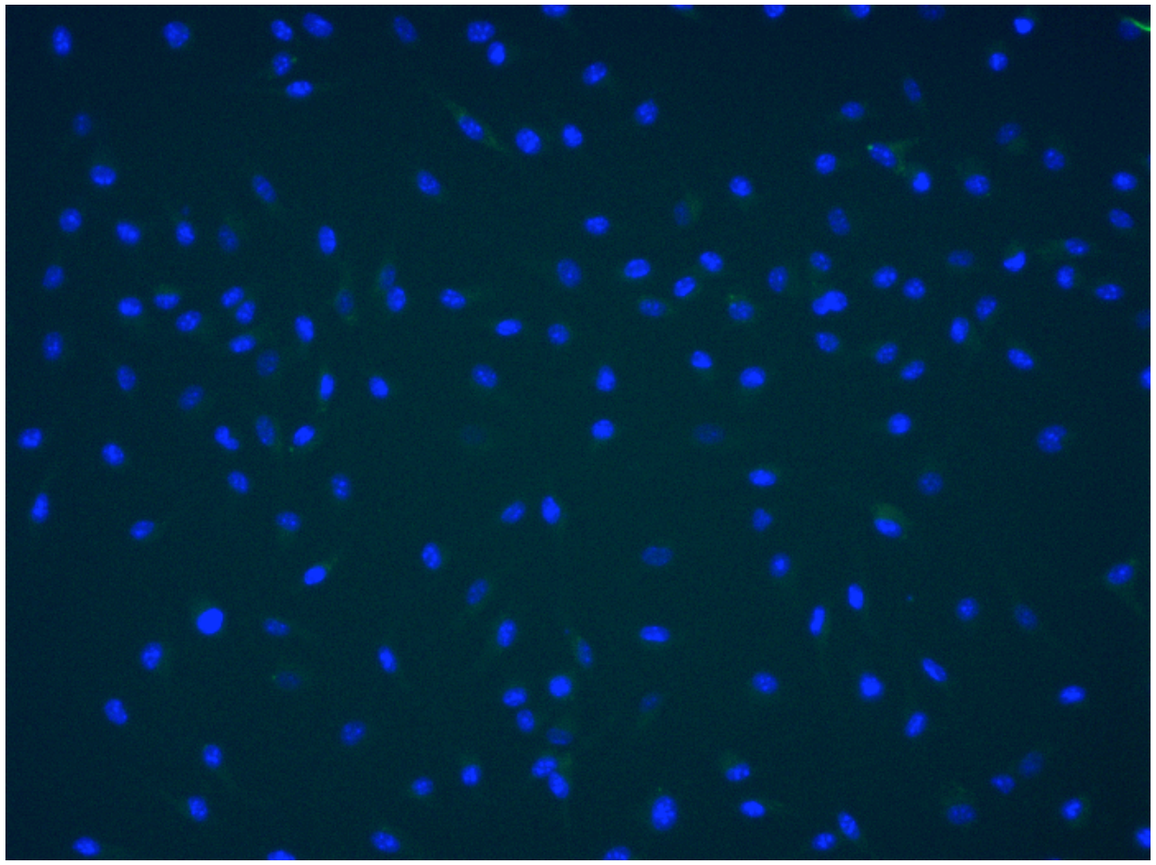
- to identify genes involved in a biological process
- to identify chemical probes to study the process
- to identify drugs to treat disease related to a process
- to identify processes we didn't know existed

Search for tuberculosis treatments

Without drug



With drug



blue = human cells

green = tuberculosis bacteria

Martha Vokes

Mark Bray

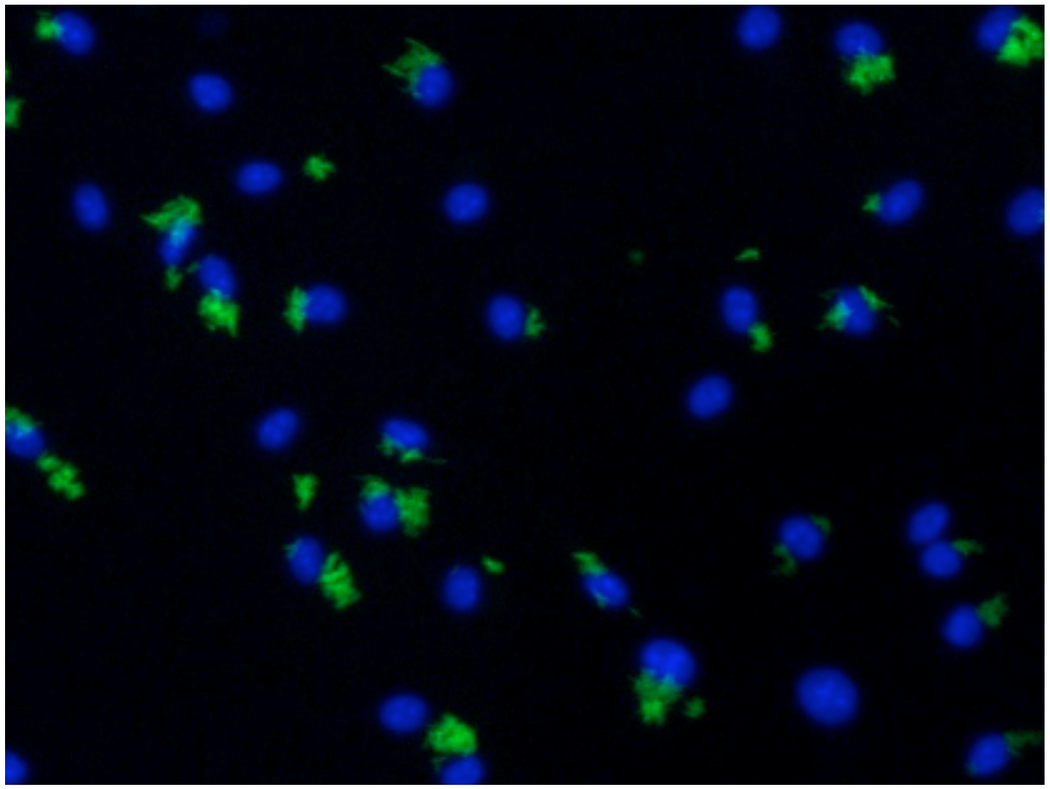
Deb Hung, Broad/MGH

Sarah Stanley, postdoc

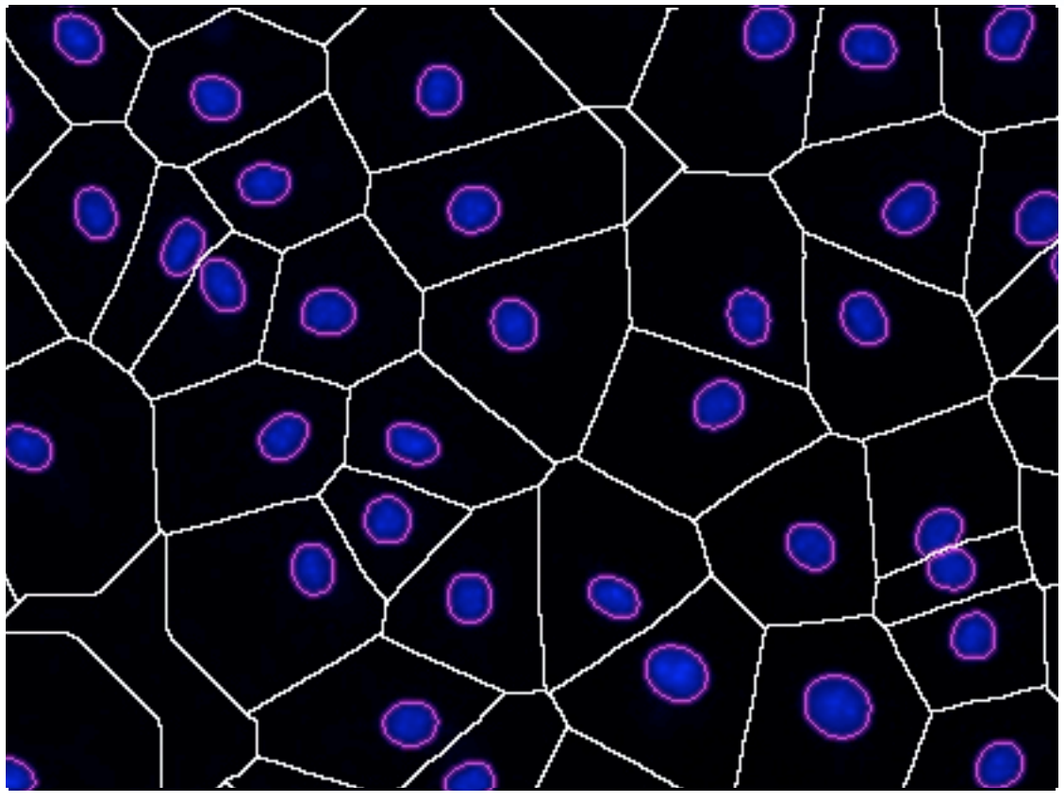
BROAD INSTITUTE IMAGING PLATFORM

project in progress

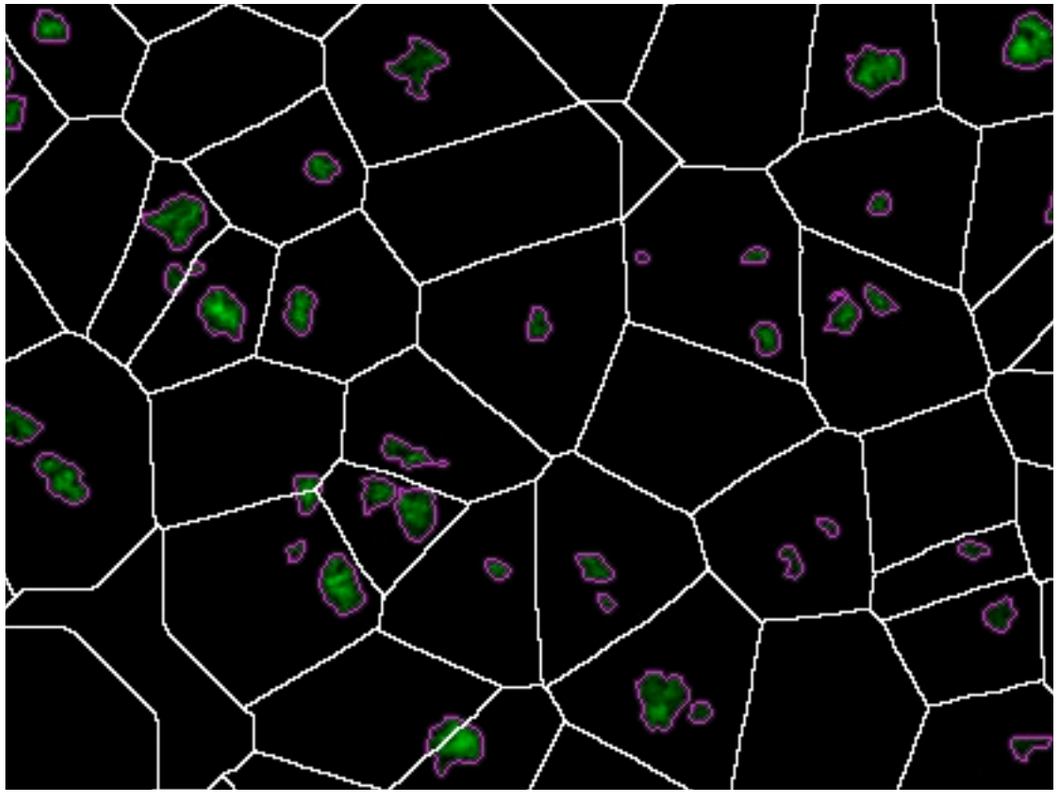
Automated image analysis



Find cells



Find bacteria

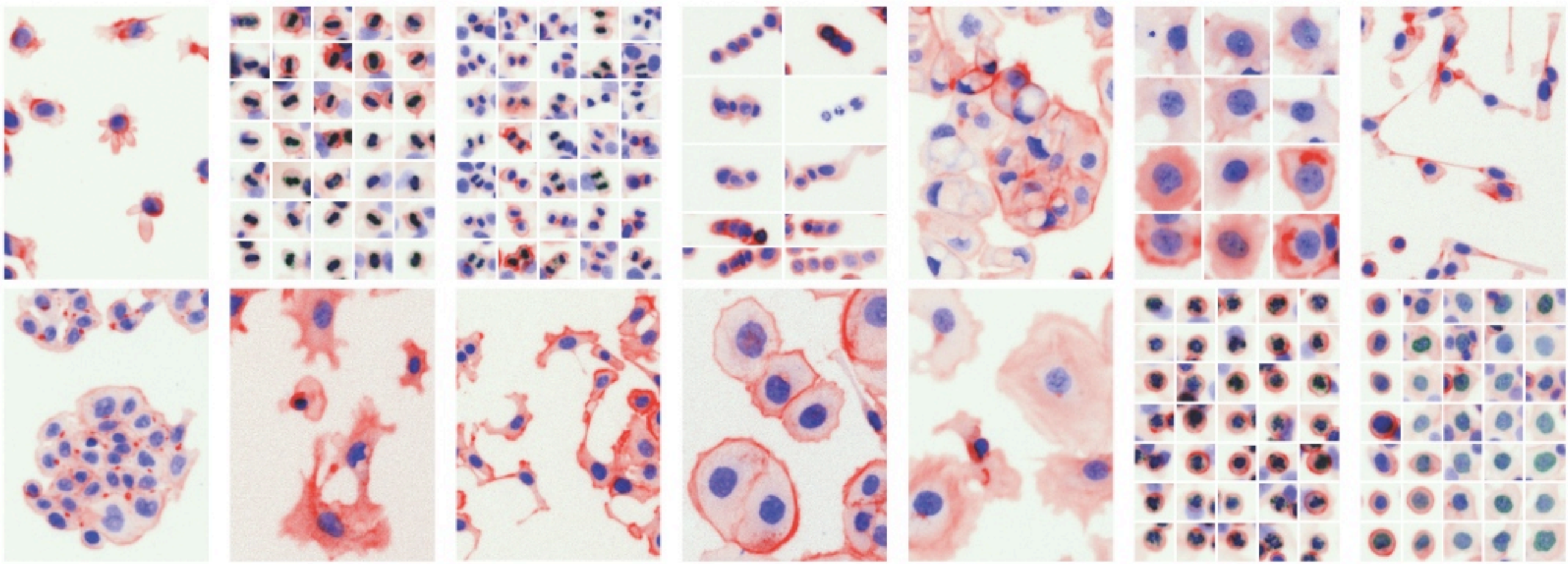


Count the number of bacteria per cell

CellProfiler™
cell image analysis software
free, at www.cellprofiler.org

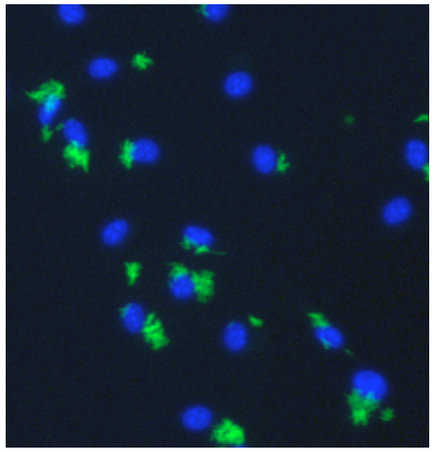
Martha Vokes
Image Assay Developer

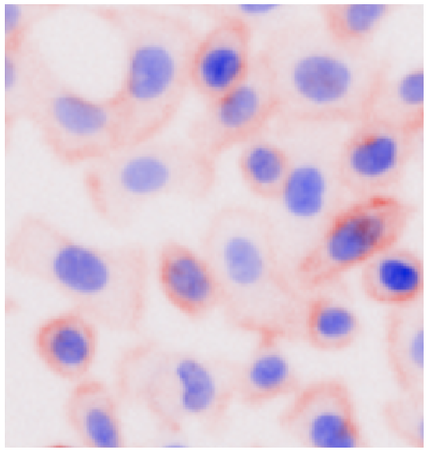
Challenging cellular phenotypes



Data visualization challenges

for 2D high-throughput microscopy



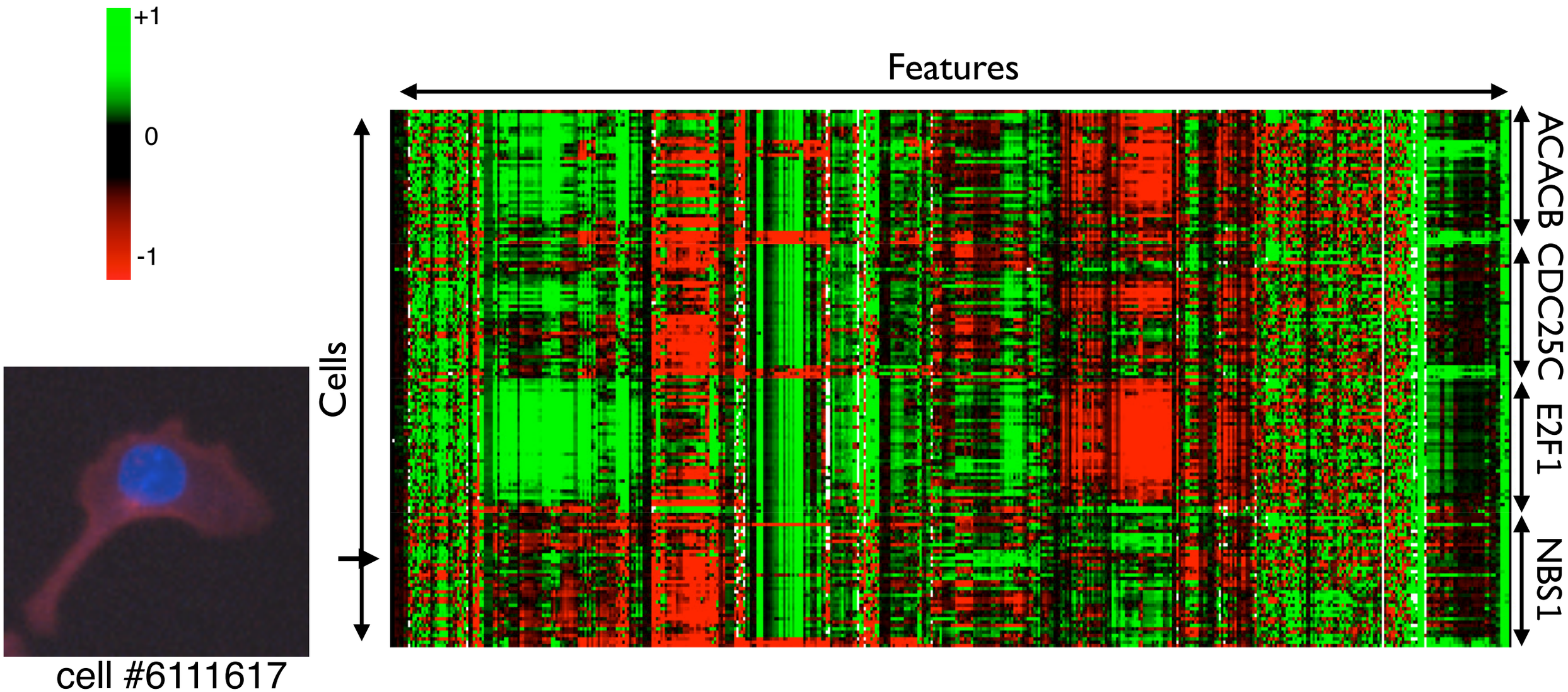


1. How can we assign each sample (= image) a numerical score?

2. How can we explore scored samples to narrow down to a small number of “hit” samples for further study?

3. How can we extract more information from these experiments, beyond a small list of “hit” samples?

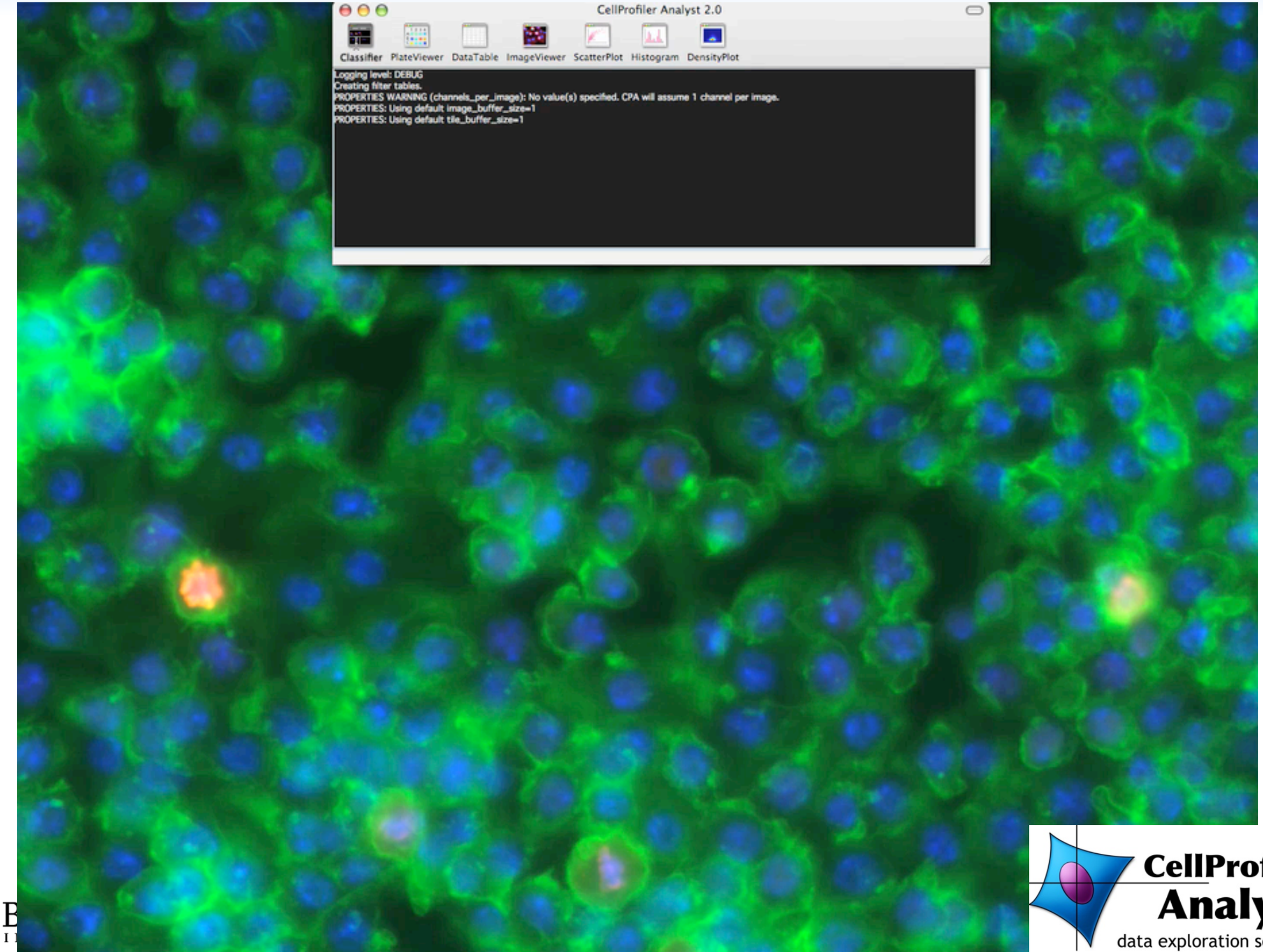
Measure everything...ask questions later.



"Cytological profile": collection of measurements describing the appearance of a cell

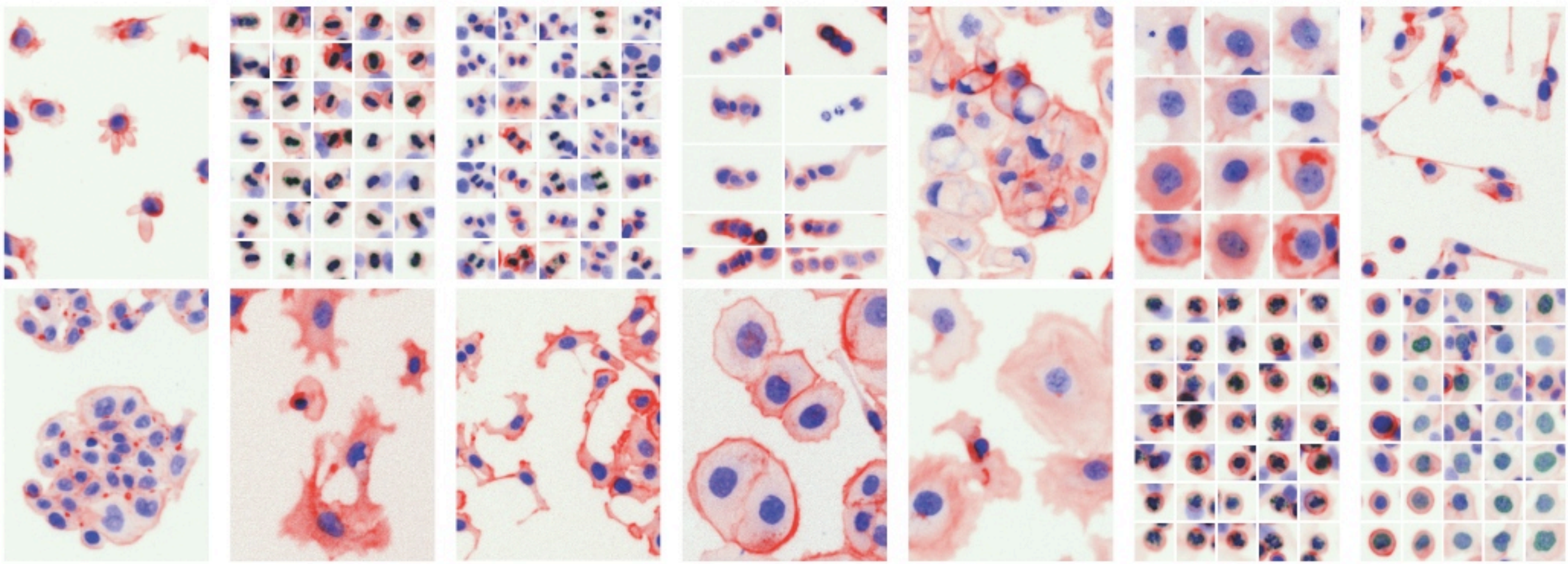
Data plot: Noa Shefi

Machine learning for complex phenotypes



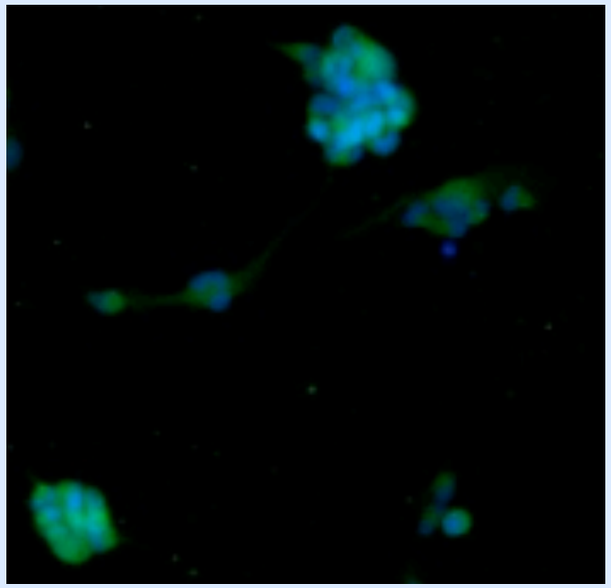
CellProfiler™ Analyst
data exploration software

Challenging cellular phenotypes

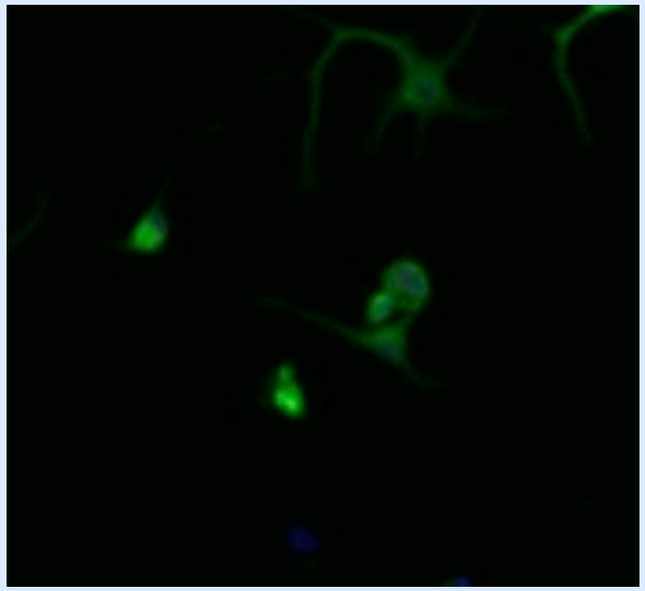


Glioblastoma proliferation & differentiation

Neurospheres

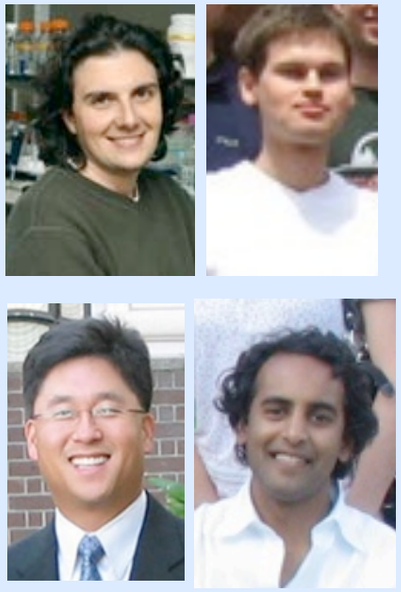


Adherent



Hoechst

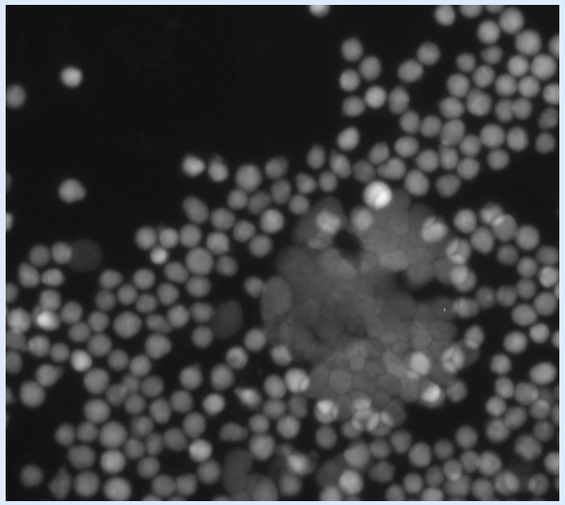
Tubulin



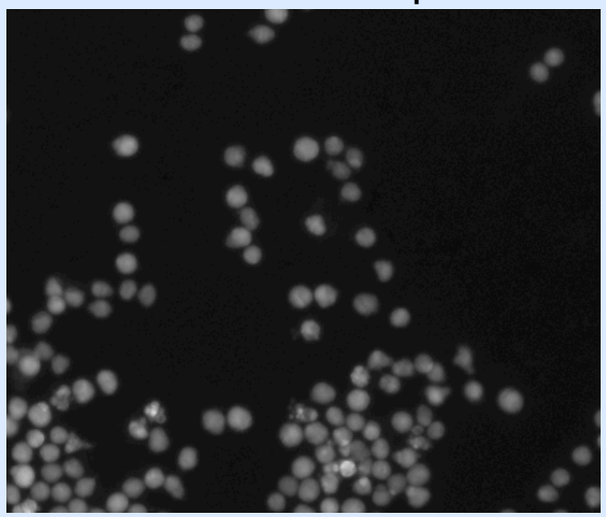
David Sabatini's lab at the Whitehead Institute (Yakov Chudnovsky, postdoc) and Bill Hahn's lab at Harvard Medical School and Dana-Farber Cancer Institute (Milan Chheda, postdoc) & the Broad RNAi platform (Dave Root, et al.)

Leukemic & hematopoetic stem cells

Cobblestones



Differentiated hematopoietic cells

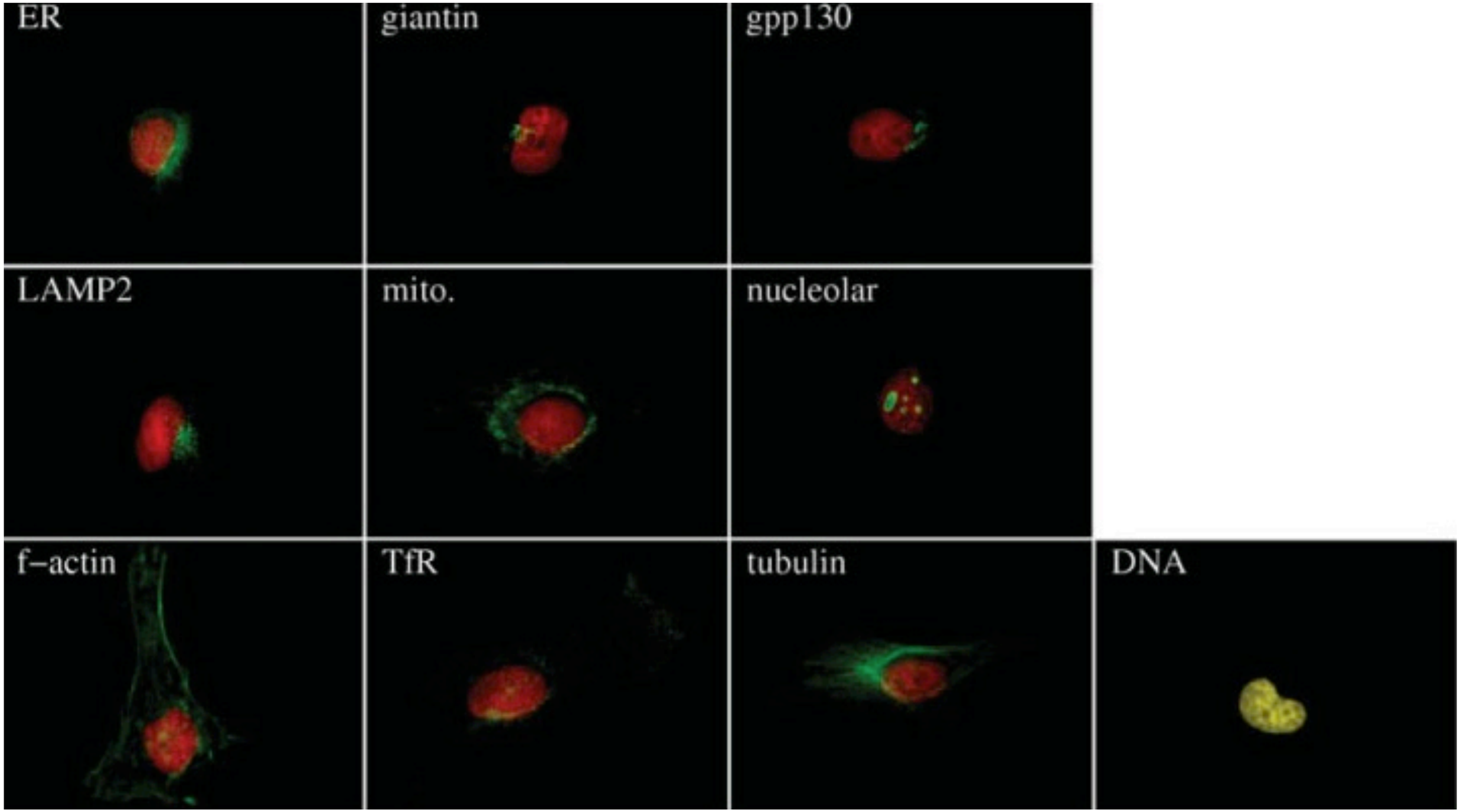


GFP

Gary Gilliland's lab at the Brigham and Women's Hospital and Harvard Medical School (Kimberly Hartwell & Peter Miller), Stuart Schreiber's lab at the Broad Institute (Alison Stewart, Shrikanta Chattopadhyay, et al.), David Scadden's lab at Massachusetts General Hospital (Siddhartha Mukherjee), Broad Chemical Biology

Scoring localization patterns

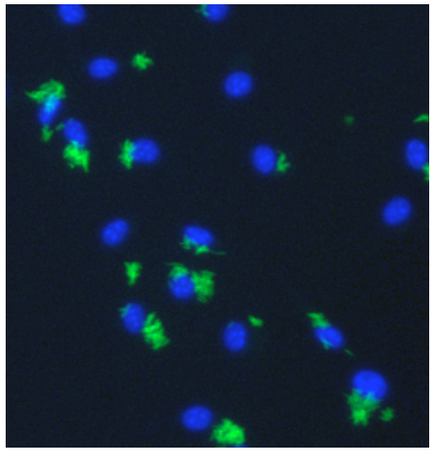
Murphy lab

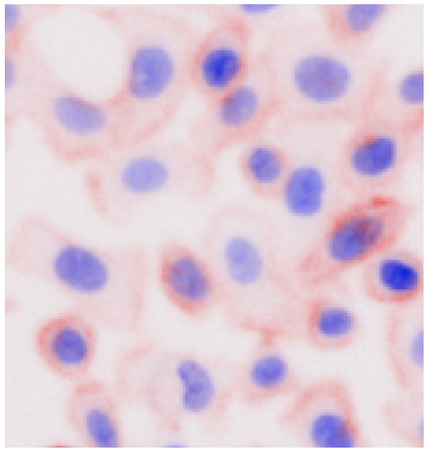


Chen, Velliste, Murphy. Cytometry (2006)

Data visualization challenges

for 2D high-throughput microscopy





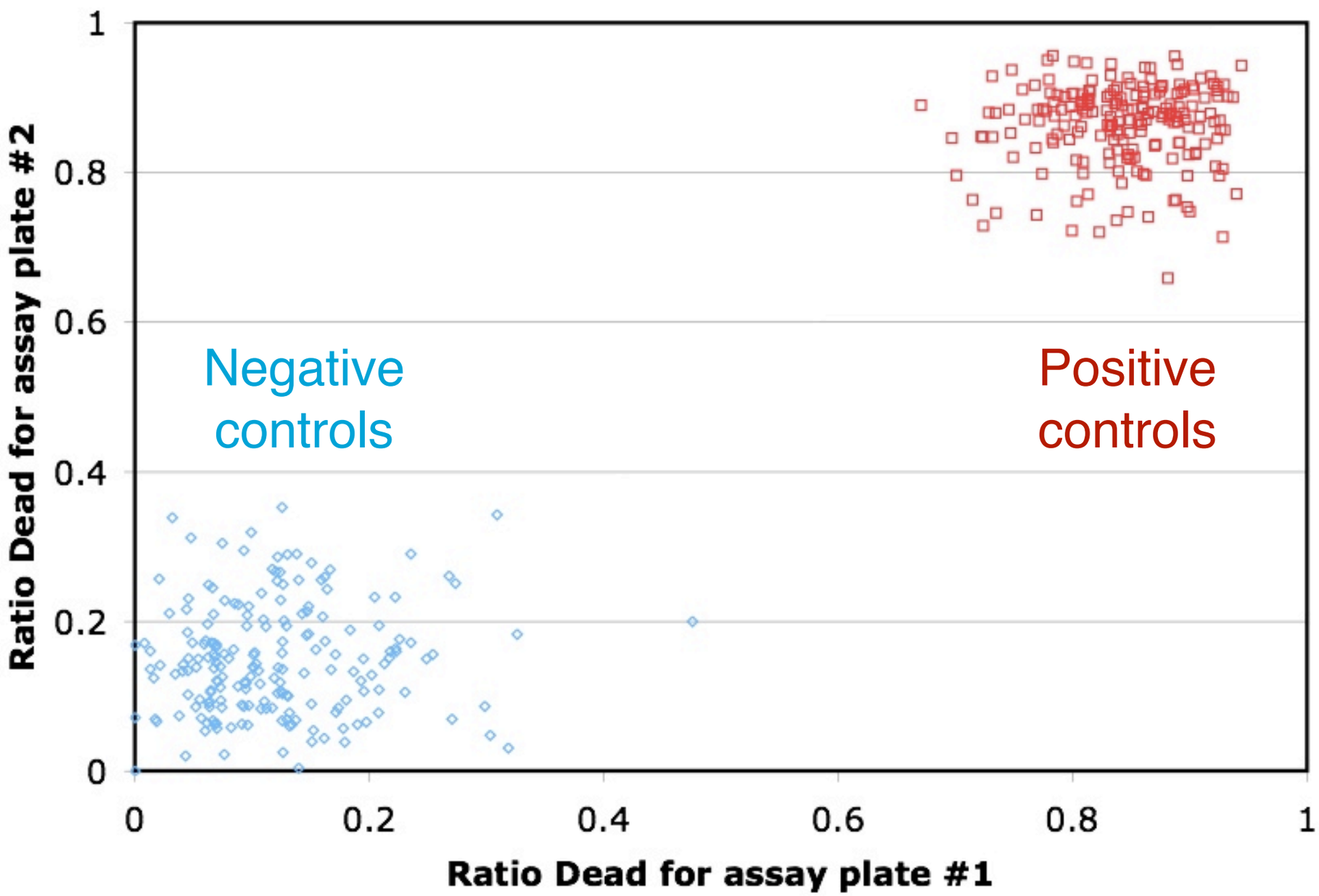
1. How can we assign each sample (= image) a numerical score?

2. How can we explore scored samples to narrow down to a small number of “hit” samples for further study?

3. How can we extract more information from these experiments, beyond a small list of “hit” samples?

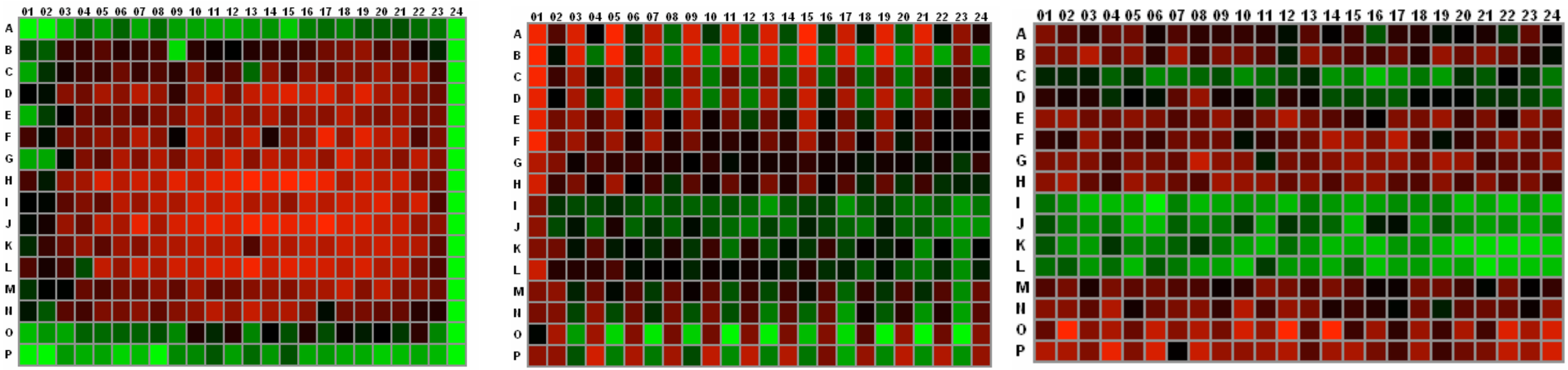
Quality control: do controls look good?

Simple scatterplot, one data point per image



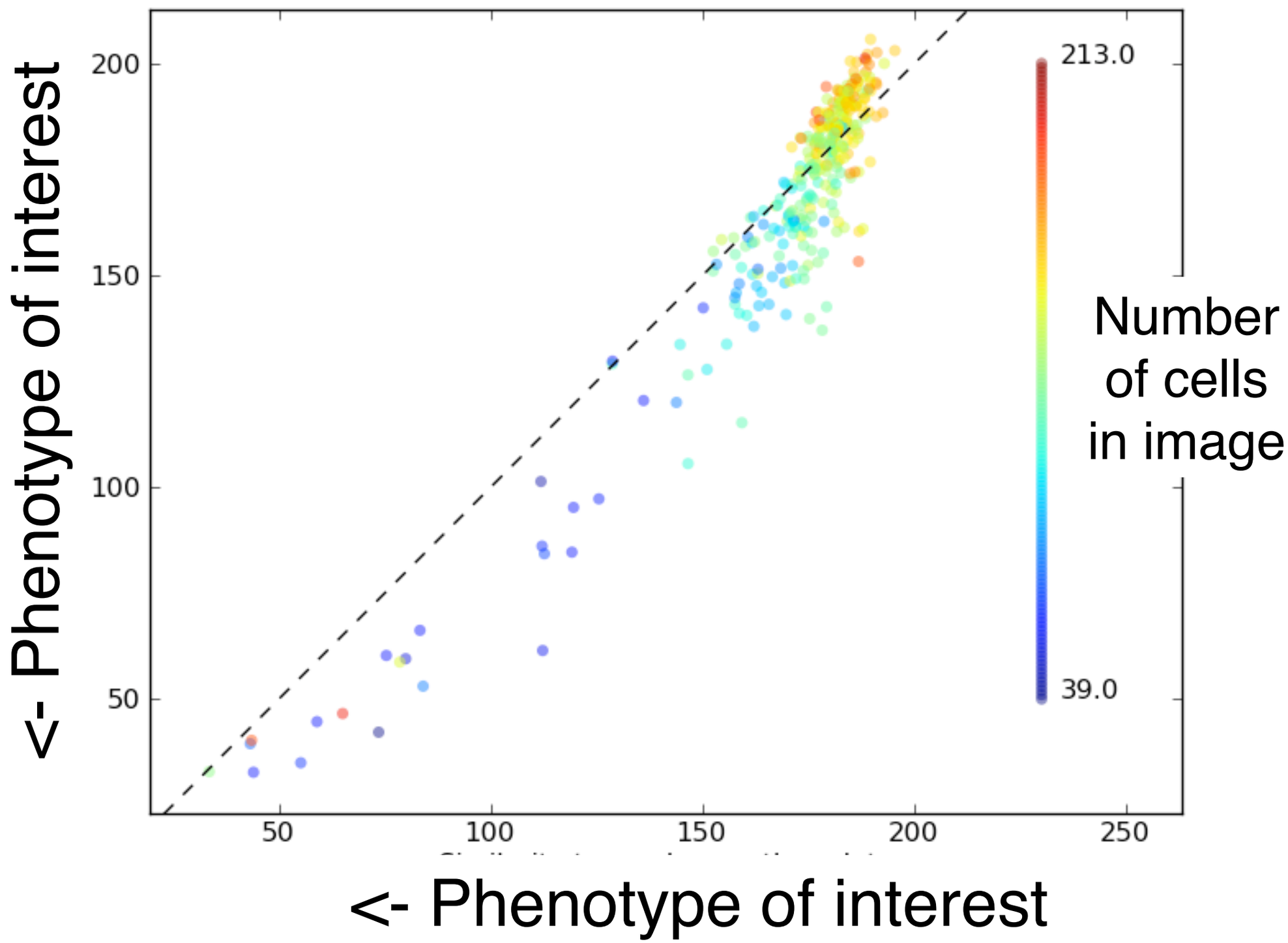
Then, encapsulate into a single metric of Assay Quality

Quality control: are there systematic artifacts?



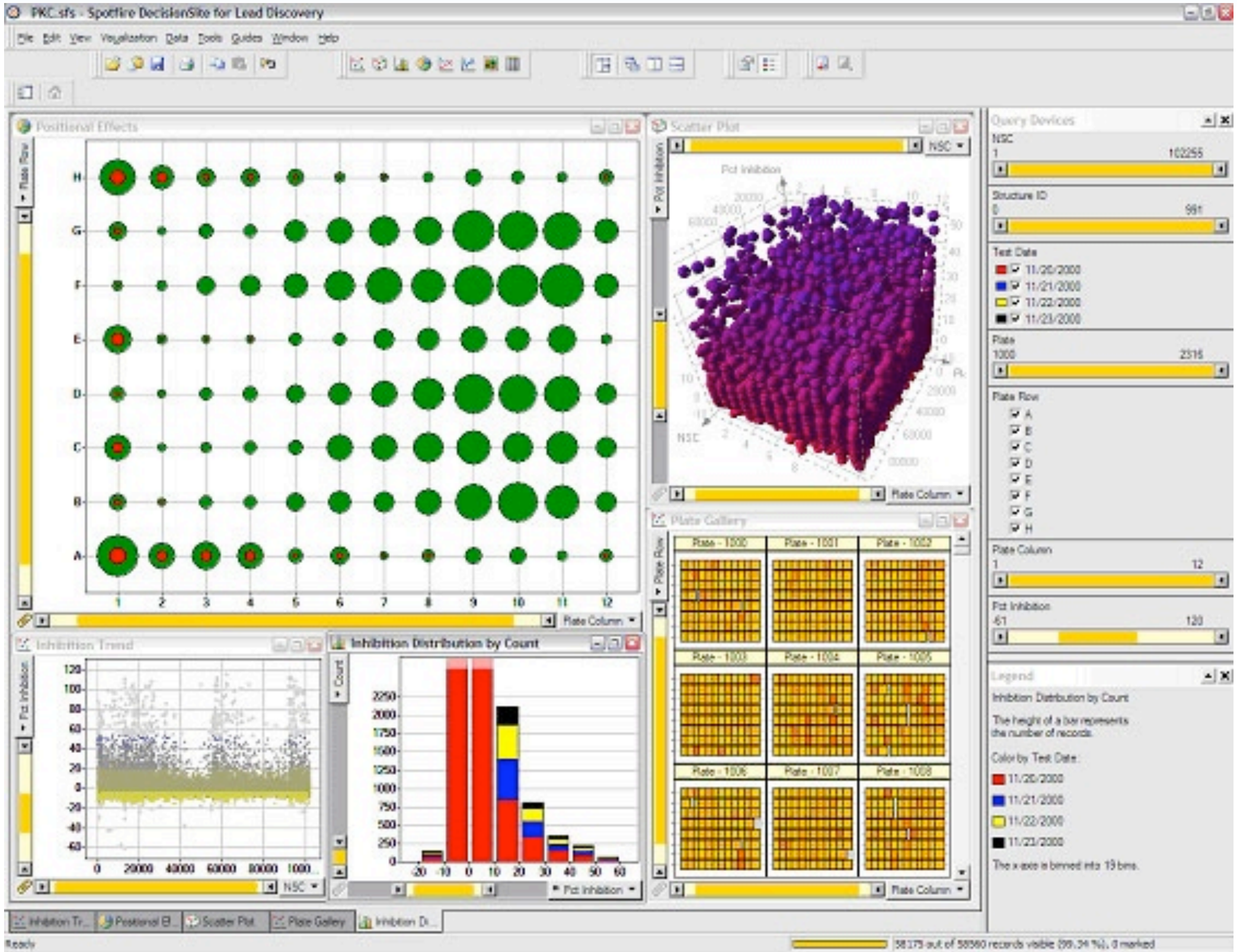
AcuityXpress images, courtesy of Ralph Garippa (Roche)

Quality control: is the phenotype confounded by other factors?



Vebjorn Ljosa and Thouis Jones, Broad Imaging Platform

Identifying hits



Spotfire image from Direct Effects.net

Quality control: are the hits real?

Source:

Data source: Screen1Feb2509run2_Per_Image

Measurements: Image_Coun...

Data Aggregation:

Aggregation method: cv%

View Options:

Color Map: jet

Well Shape: rounded

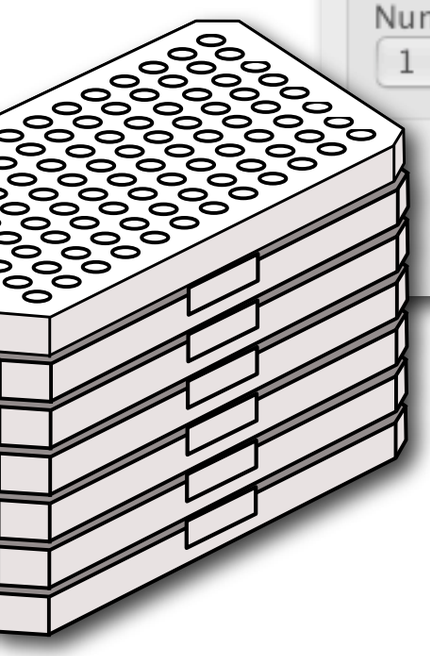
Number of Plates: 1

Plate: 2002-01-W01-02-01-CN00002412-Brerun

01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

A B C D E F G H I J K L M N O P

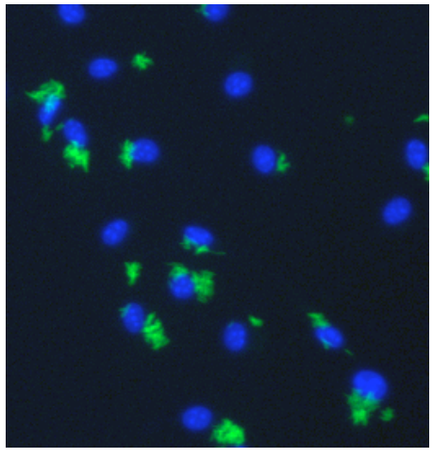
9.974 31.336 38.456

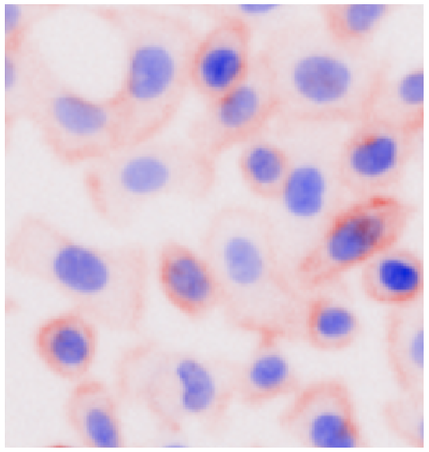


Adam Fraser

Data visualization challenges

for 2D high-throughput microscopy



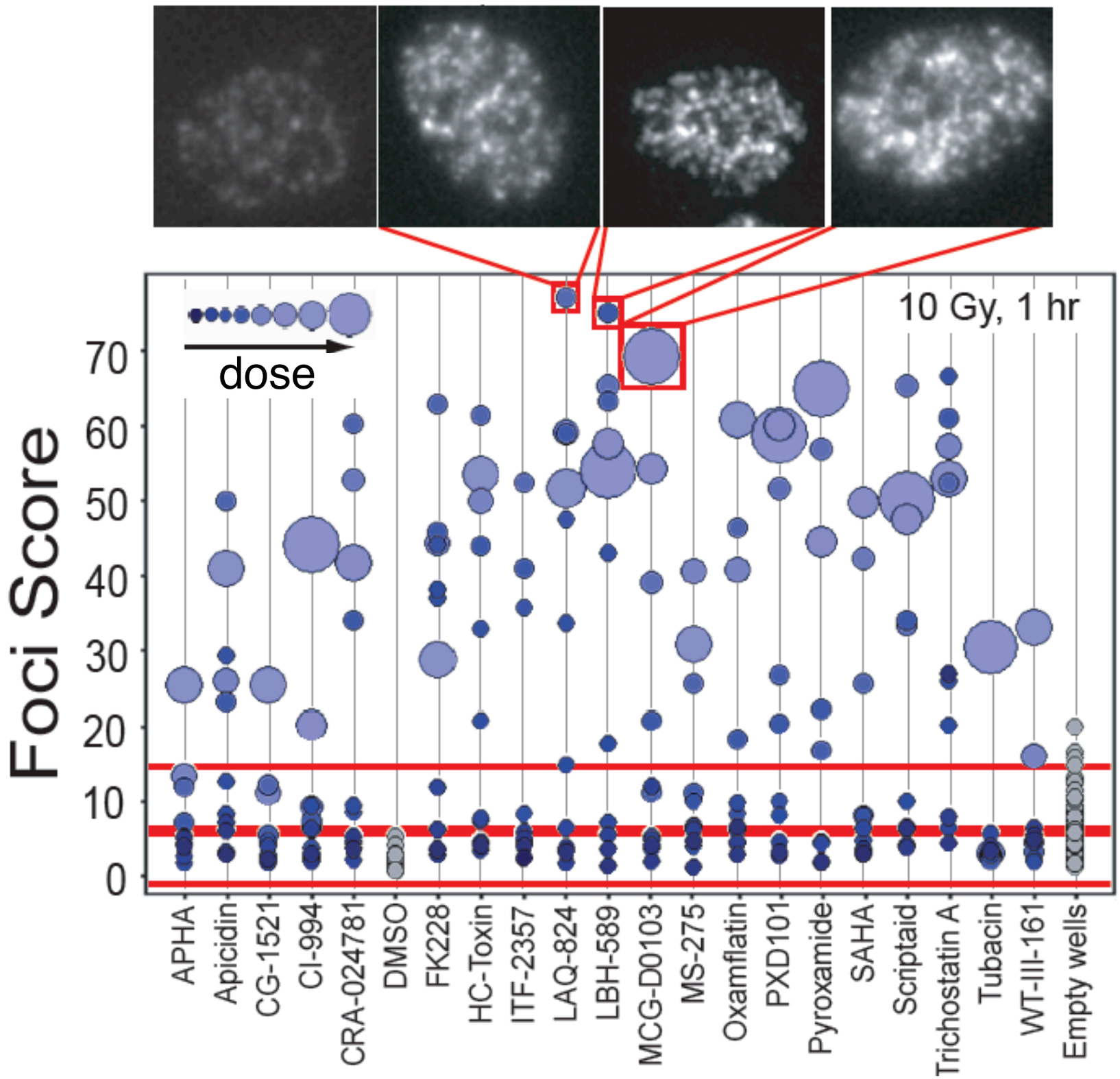


1. How can we assign each sample (= image) a numerical score?

2. How can we explore scored samples to narrow down to a small number of “hit” samples for further study?

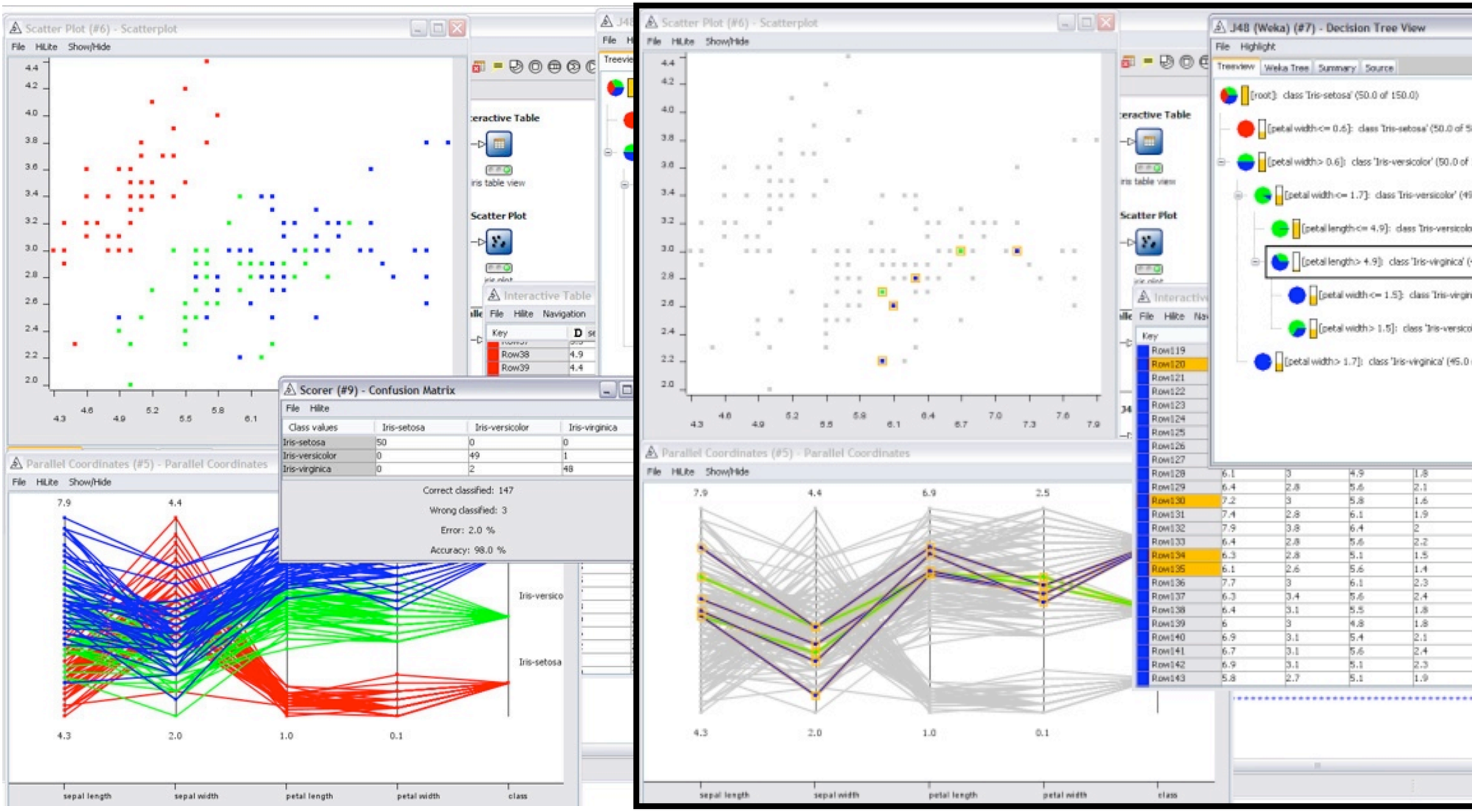
3. How can we extract more information from these experiments, beyond a small list of “hit” samples?

Exploring hits relative to other sources of data



Spotfire image, from Scott Floyd, MIT

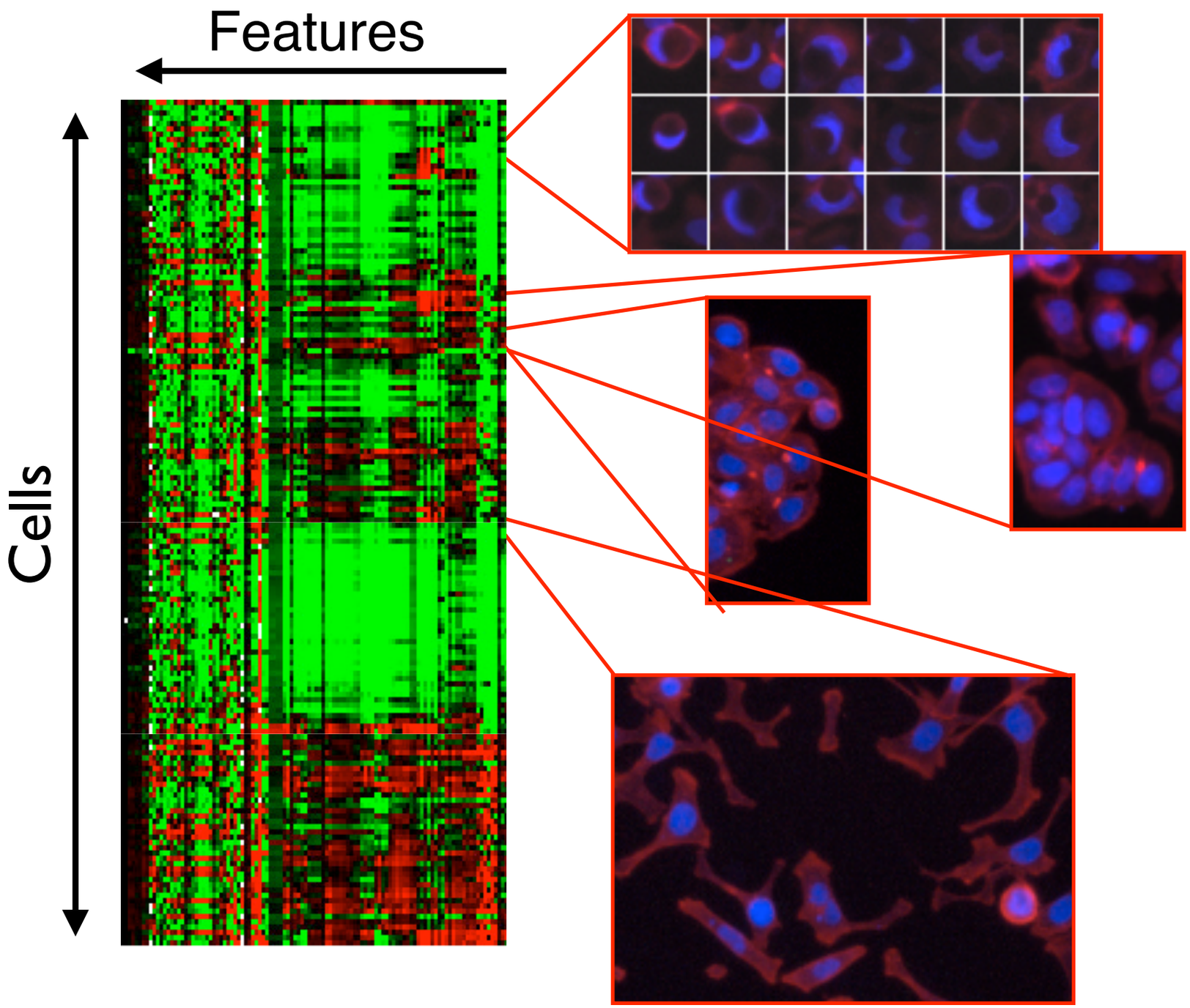
Exploring properties of hits



from www.KNIME.org

Unsupervised identification of novel phenotypes

Clustering of individual cells

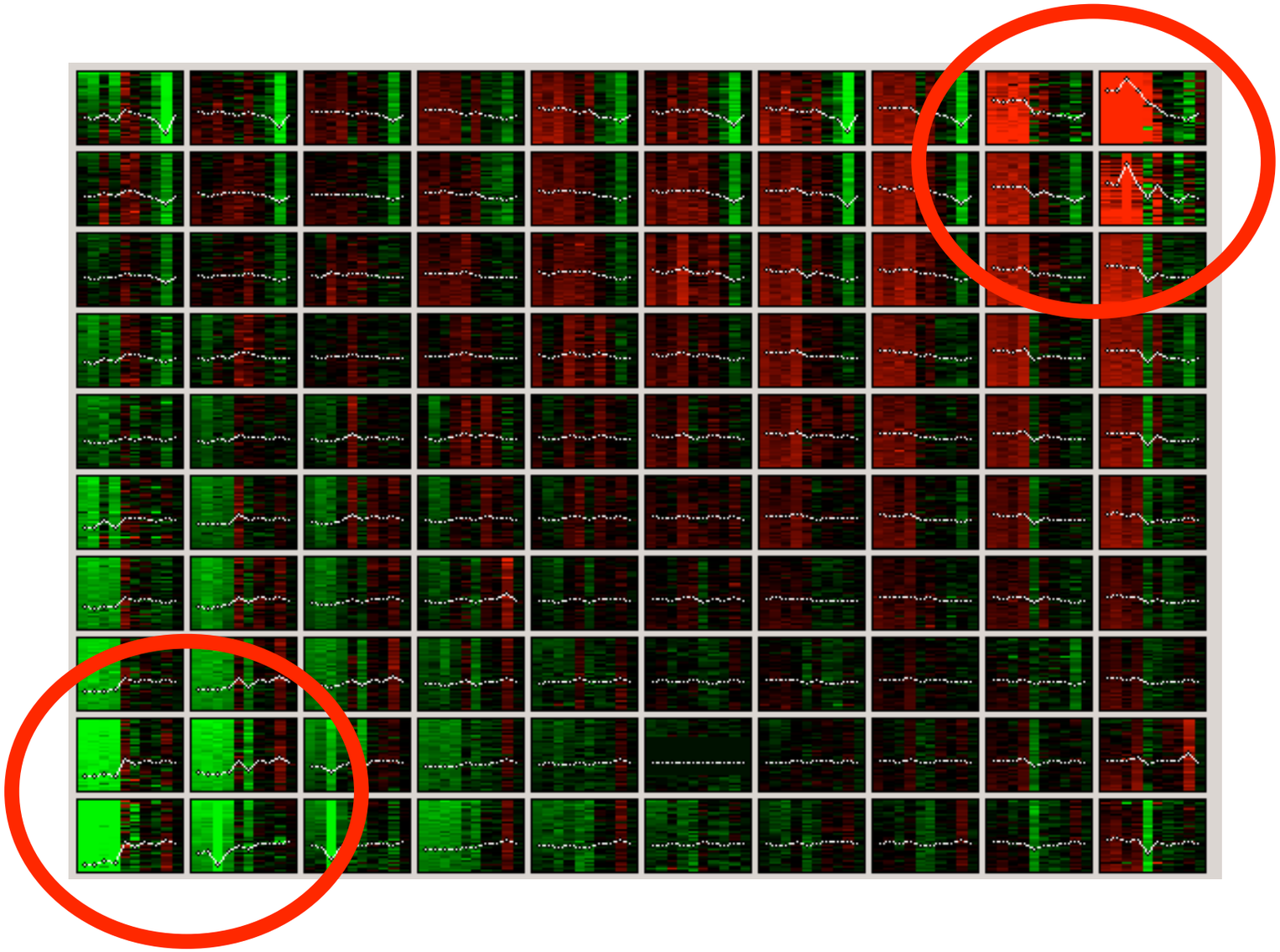


…plus, samples that produce distinct cytoprofiles but where cells look normal to our eye

Schematic

Unsupervised identification of novel phenotypes

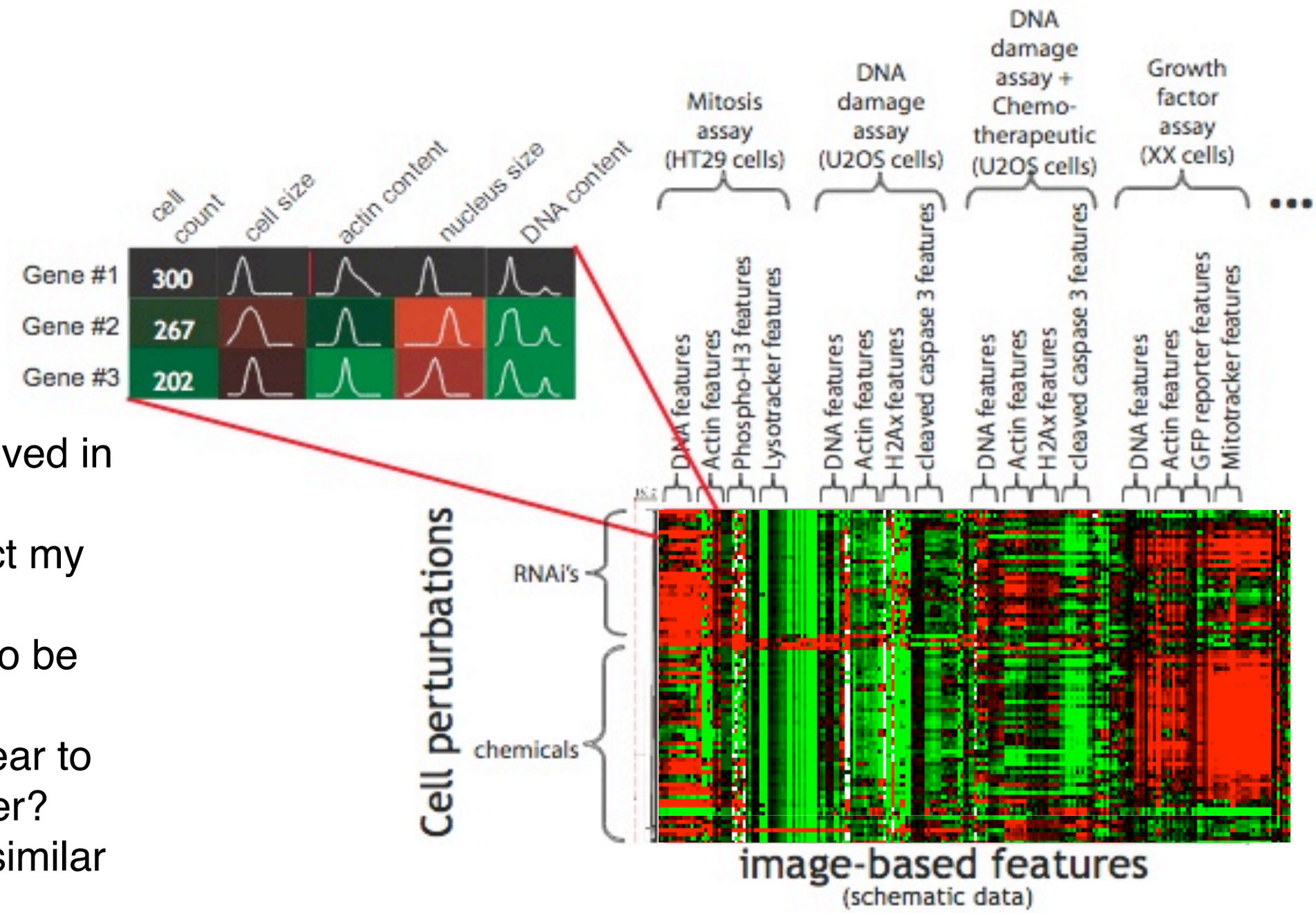
Self-organizing maps of samples



- **This reveals two compound-response profiles that are very different to the basal response**

AcuityXpress images, courtesy of Ralph Garippa (Roche)

Look for patterns/similarities/relationships among samples, phenotypes, and external data sources

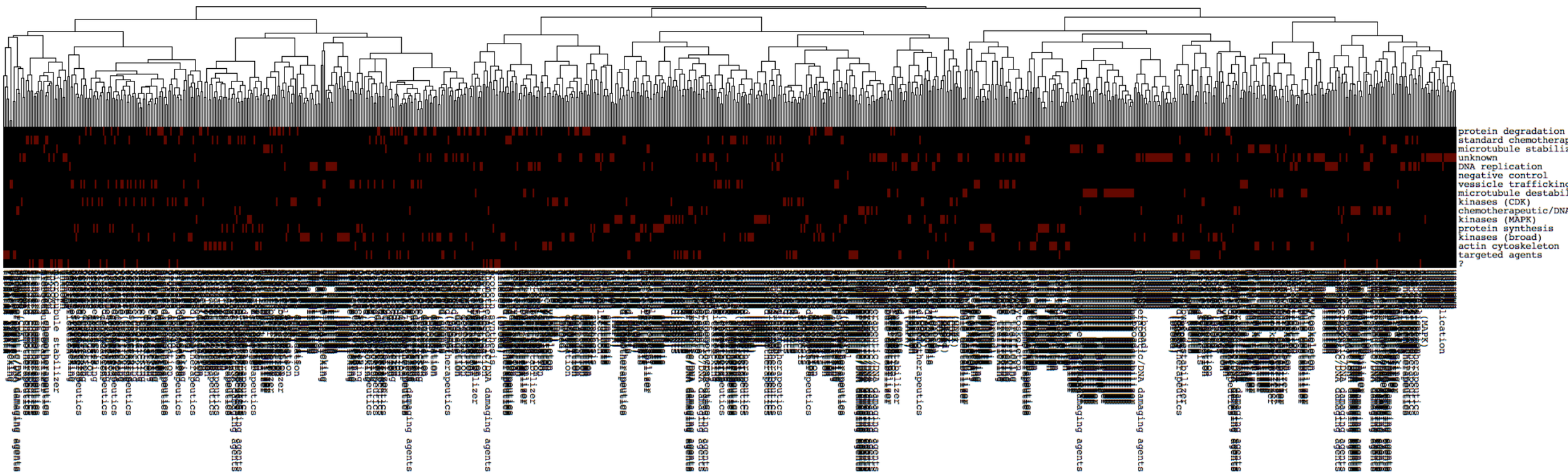


- What genes are involved in my phenotype?
- What chemicals affect my phenotype?
- What genes appear to be similar to each other?
- What chemicals appear to be similar to each other?
- What genes appear similar to a chemical?
- What phenotypes are coordinated?

- What relationships exist between my phenotype and known information about the samples (proteomics, transcriptional microarrays)?

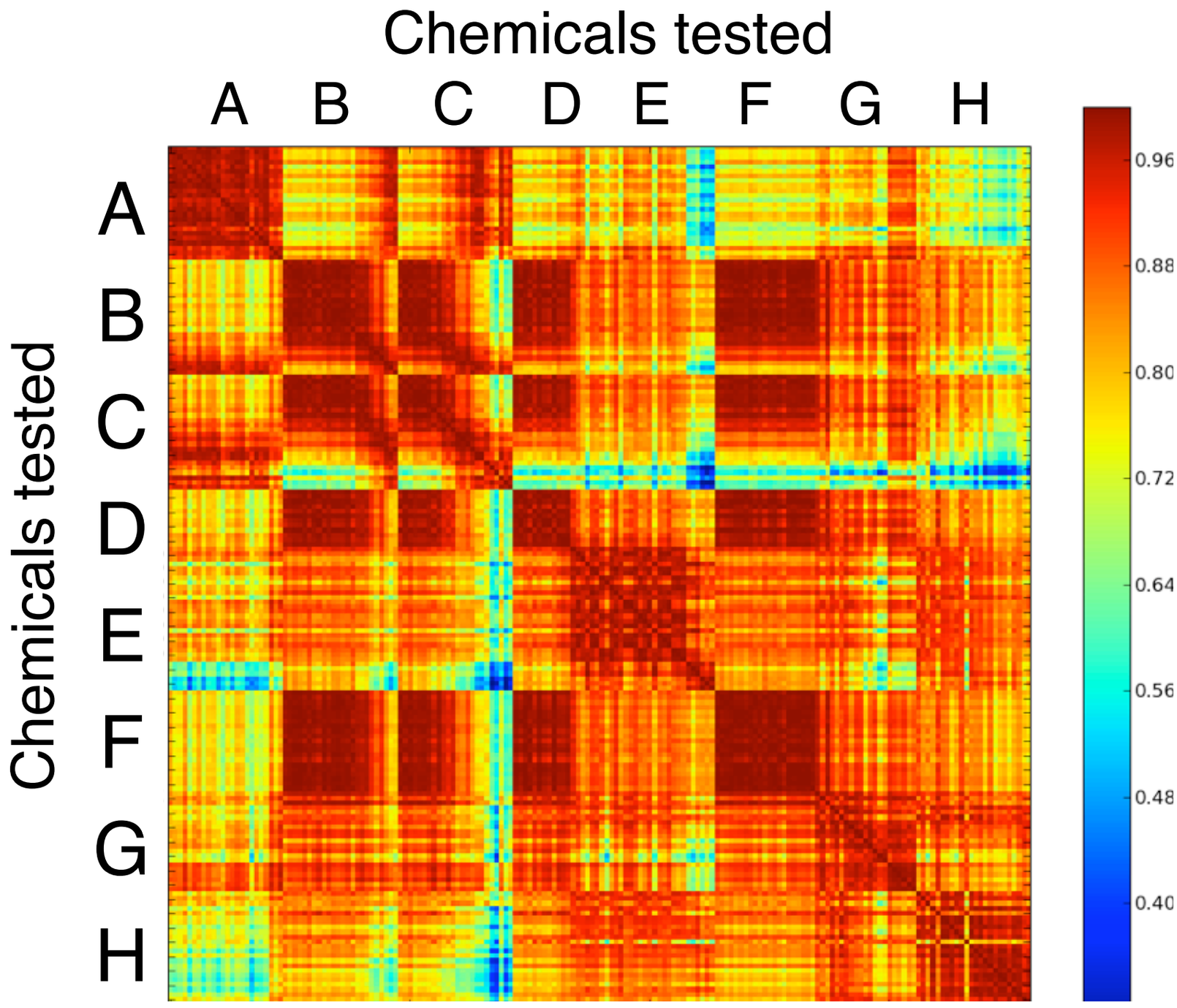
Exploring similarities between samples

Clustering



Exploring similarities between samples

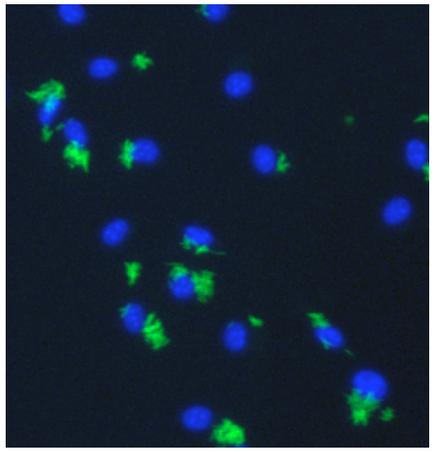
Clustering, with dose-response data

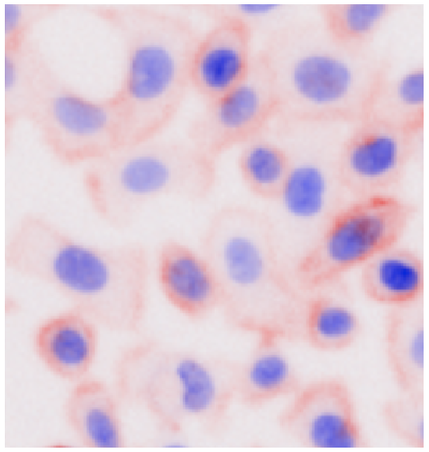


Color = pheno-type similarity

Data visualization challenges

for 2D high-throughput microscopy





1. How can we assign each sample (= image) a numerical score?

2. How can we explore scored samples to narrow down to a small number of “hit” samples for further study?

3. How can we extract more information from these experiments, beyond a small list of “hit” samples?

Collaborative opportunities

Broad Institute Imaging Platform

- Available for consulting on challenging image analysis and data mining projects (the Broad Institute is not-for-profit)

Broad Institute Chemical Biology Pipeline

- Apply to develop an assay and carry out high-throughput screening for a fee
- Funding is available for assays related to psychiatric disease: PsychHTS

NIH's MLPCN: Molecular Libraries Probe Production Centers Network

https://mli.nih.gov/mli

- **Assays are contributed/developed by the scientific community and screened (at no cost to the investigator) at a national MLPCN center, such as at the Broad Institute**
- Several programs provide funding for assay development prior to submitting an assay to the MLPCN (R01s, R03s, R21s, and "Fast Track" if an assay has already been developed via existing NIH funding)

Gratitude

free, at www.cellprofiler.org :

Anne E. Carpenter

Peggy Anthony
Mark Bray
Adam Fraser

Thouis (Ray) Jones
Lee Kamentsky
Vebjørn Ljosåa

David Logan
Kate Madden
Carolina Wählby

Contact:
anne@broadinstitute.org

This work has been supported by:

- The Broad Institute of Harvard and MIT
- Eli Lilly grant
- Society for Biomolecular Screening Small Grant Award
- L'Oreal for Women in Science fellowship
- DOD Tuberous Sclerosis Complex Grant
- Novartis fellowship from the Life Sciences Research Foundation
- Merck/MIT Computational & Systems Biology postdoc fellowship
- MIT EECS/Whitehead/Broad Training Program in Computational Biology

Many thanks to our many biology collaborators who provide images, and to Polina Golland, our computer science collaborator at MIT's Computer Science/Artificial Intelligence Laboratory (CSAIL)

S.D.G.