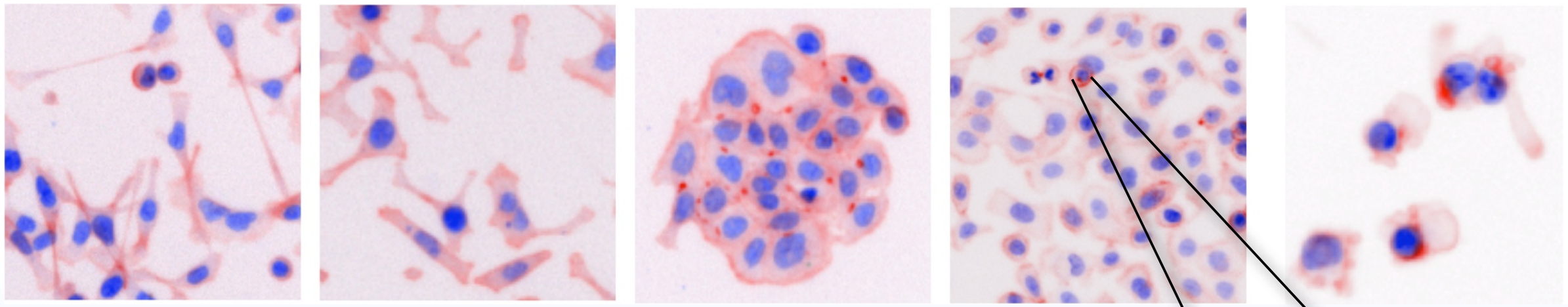
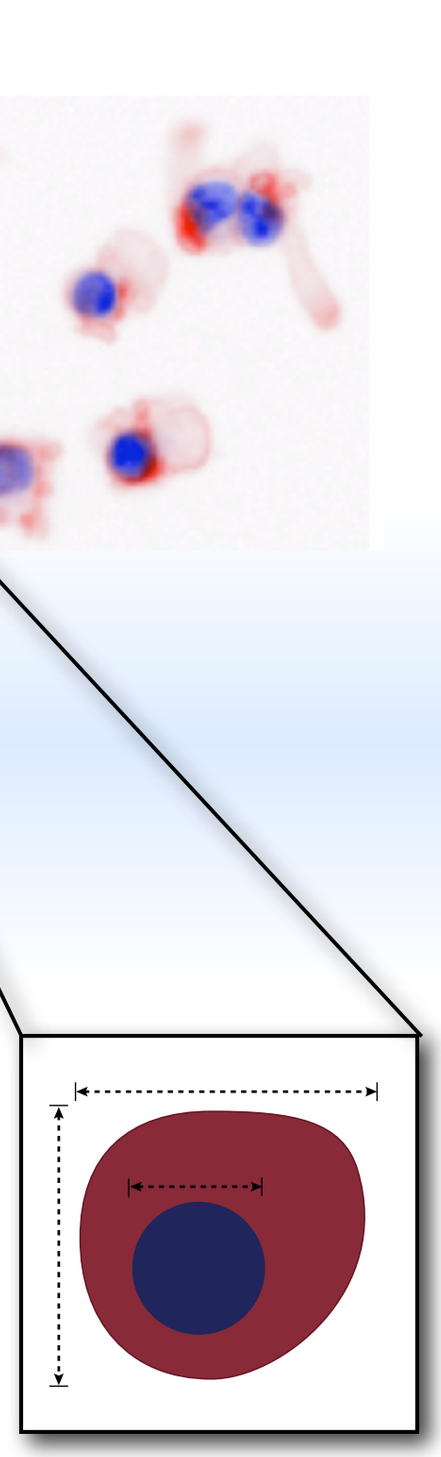


Extracting rich information from biological images

Anne E. 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

Carpenter lab

Broad Institute Imaging Platform

Taking on challenging image analysis and data mining projects

Anne Carpenter

Image assay development

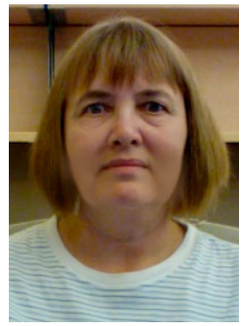
Apply image analysis methods to biological questions

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

Vebjørn Ljosa

Adam Fraser

Lee Kamentsky

Carolina Wählby

Students and postdocs

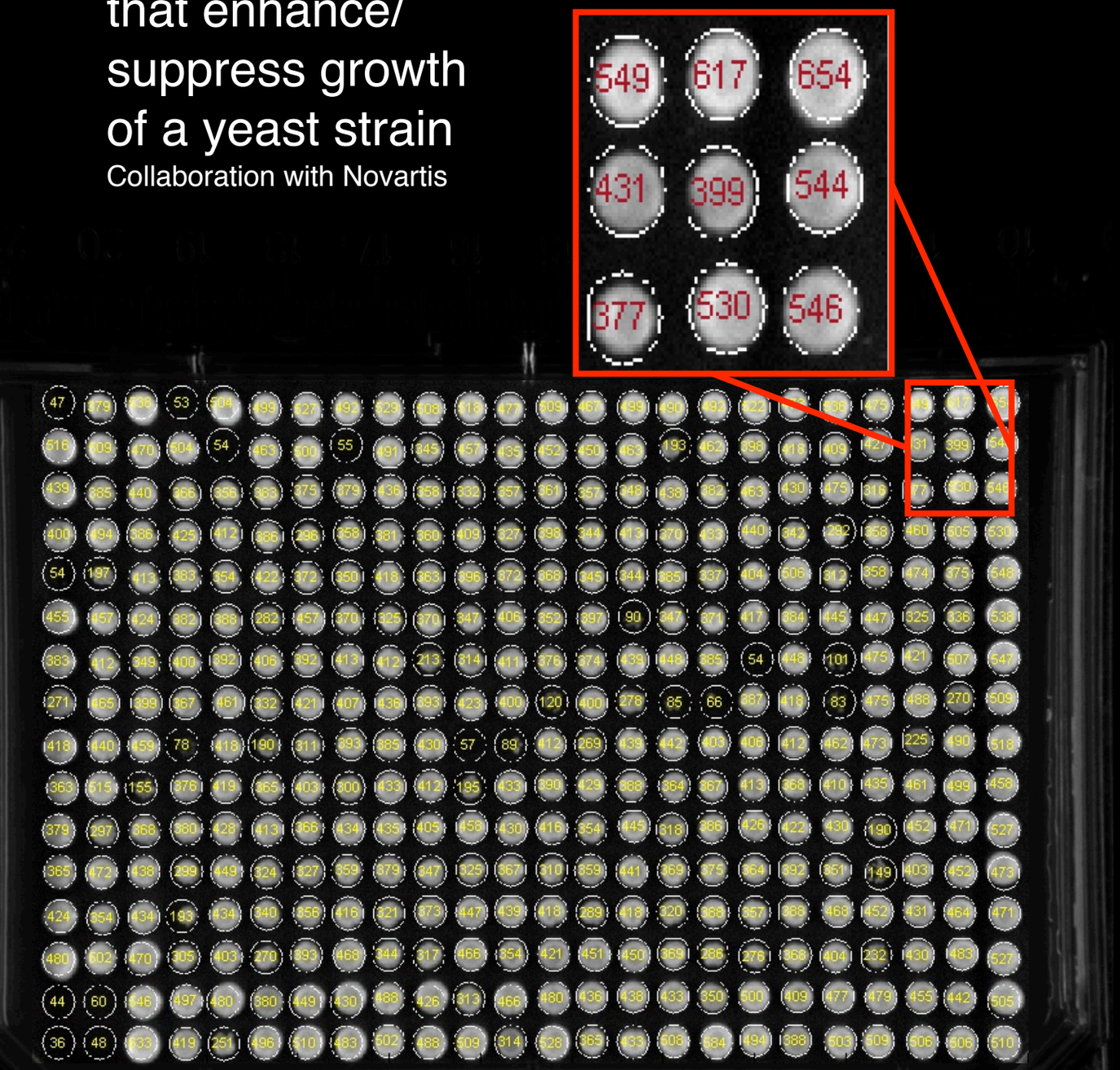
Imtiaz Khan

BROAD INSTITUTE IMAGING PLATFORM

Yeast patch growth:

Goal: identify chemicals or genetic knockouts that enhance/ suppress growth of a yeast strain

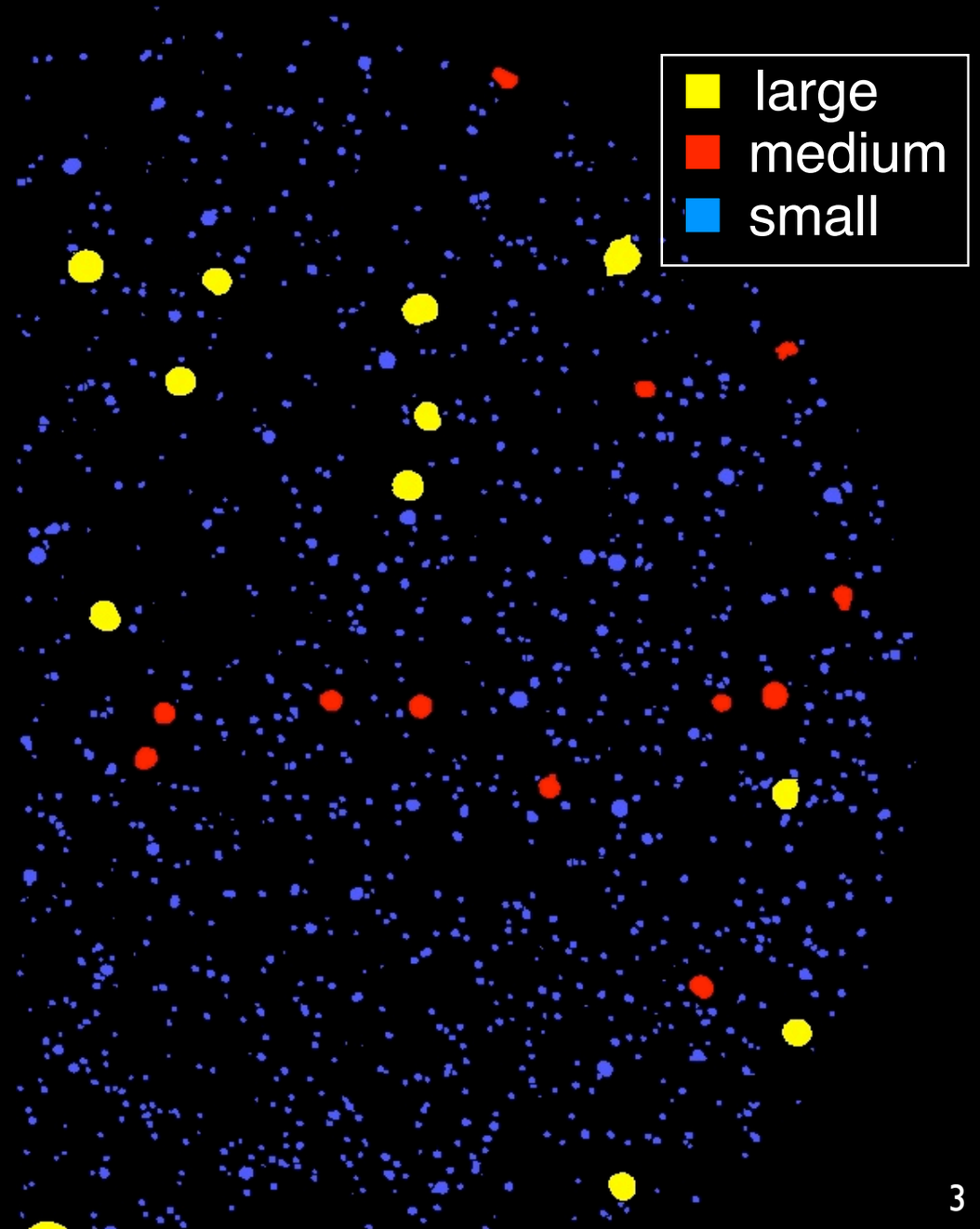
Collaboration with Novartis



Yeast colony size:

Goal: to understand pathways leading to drug-resistant yeast

Cowen, et al., Eukaryotic Cell, 2006



Images contain a wealth of information

Images: http://www.microscopyu.com

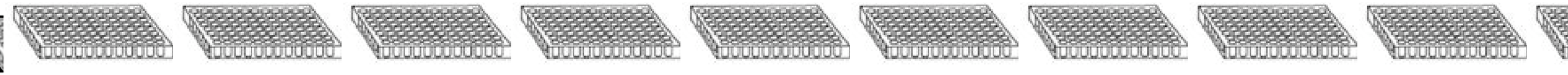
Screening to find genes and chemicals of interest

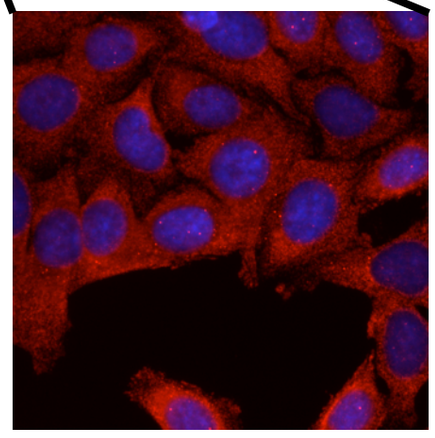
NIH: MLPCN

Biology research groups (Harvard, MIT, around the world)

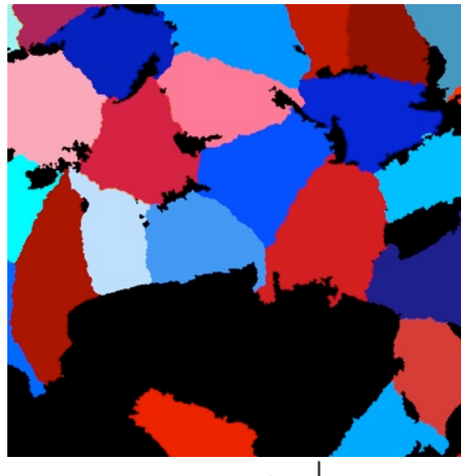
↓

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



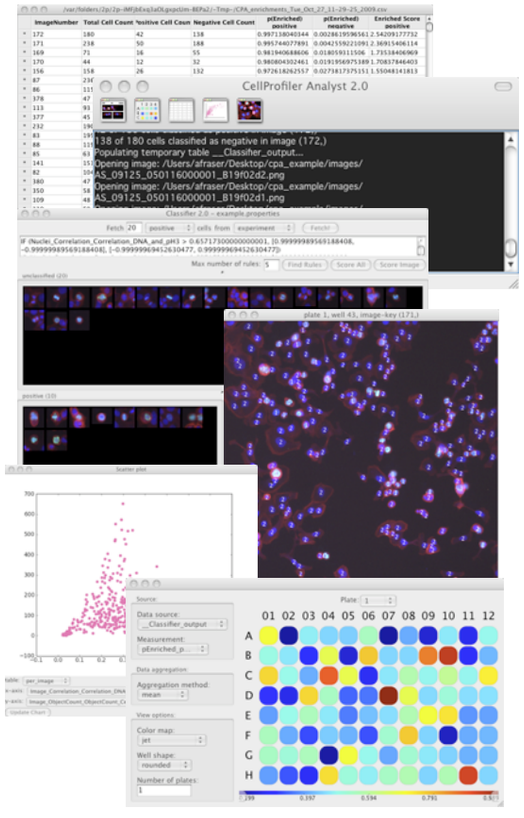


automated microscopy (any manufacturer)



→ Cell measurements

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



Data exploration & machine learning

Ray Jones

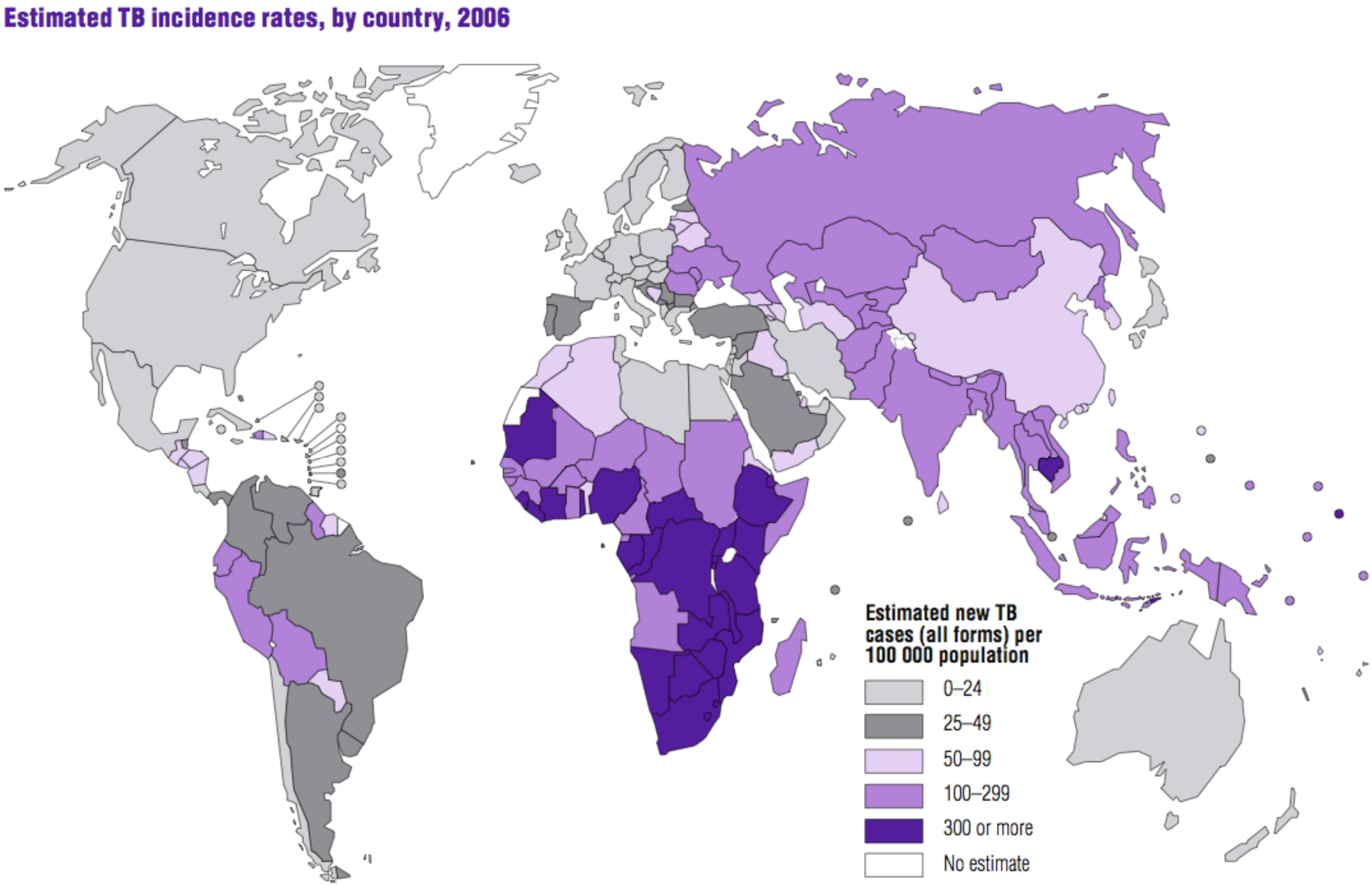
Anne Carpenter

BROAD INSTITUTE IMAGING PLATFORM

CellProfiler™ cell image analysis software

CellProfiler™ Analyst data exploration software

Case study: Tuberculosis

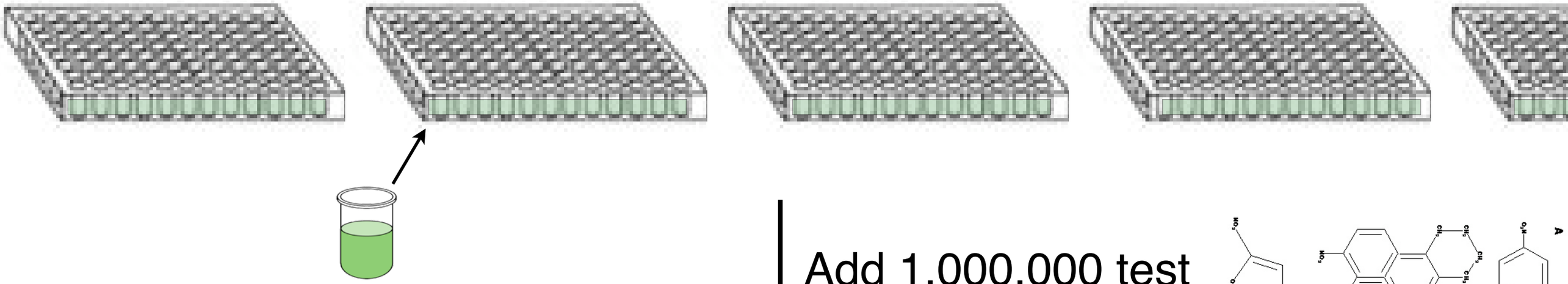


9.2 million new cases of tuberculosis in 2006
1.7 million deaths in 2006

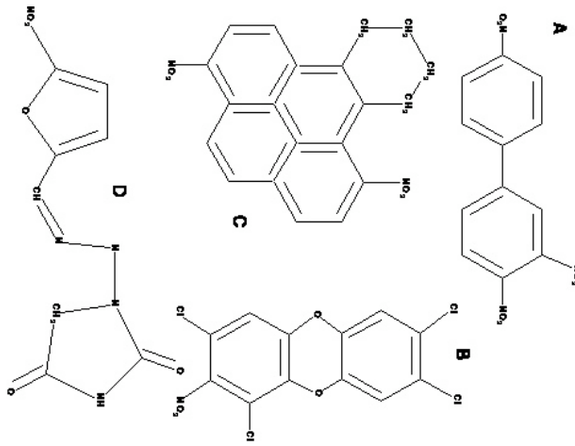
WHO Report, Global Tuberculosis Control 2008

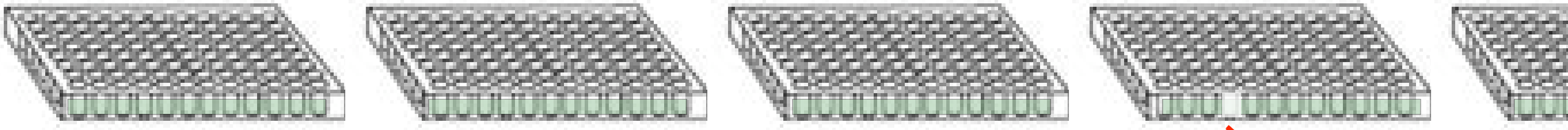
Traditional approach to find antibiotics

Put bacteria in individual wells of multi-well plates



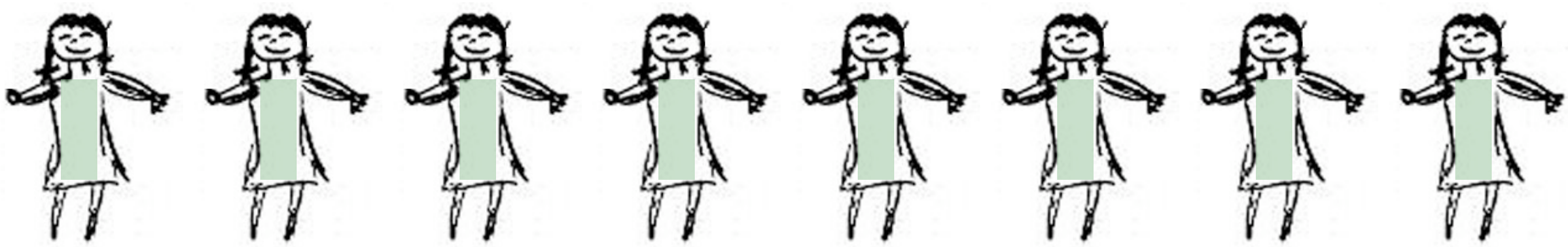
Add 1,000,000 test chemicals, each chemical in a different well



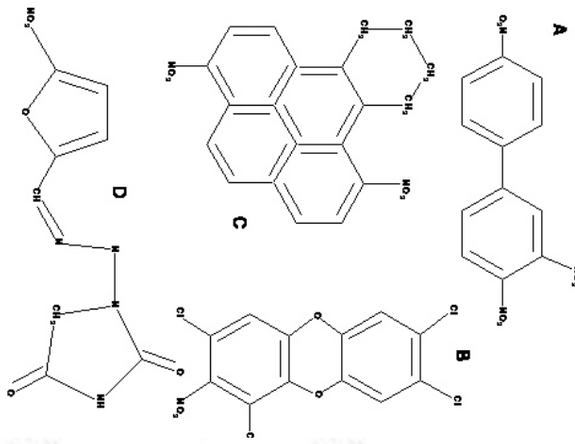


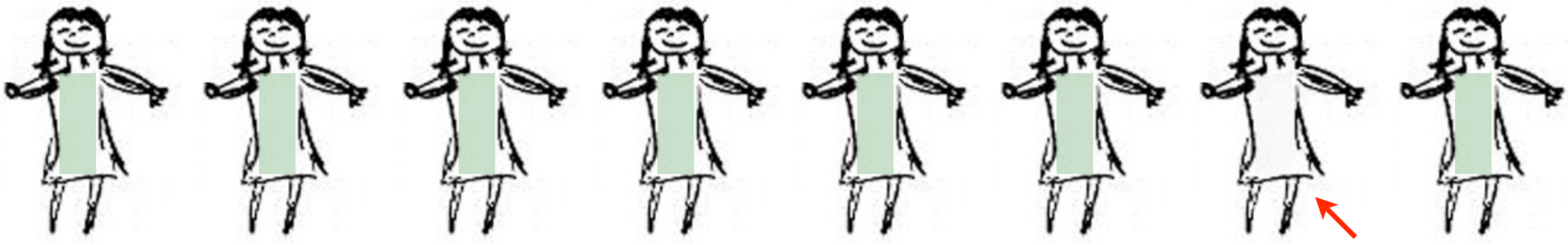
Measure amount of bacteria (fluorescence plate reader) and look for wells where bacteria died

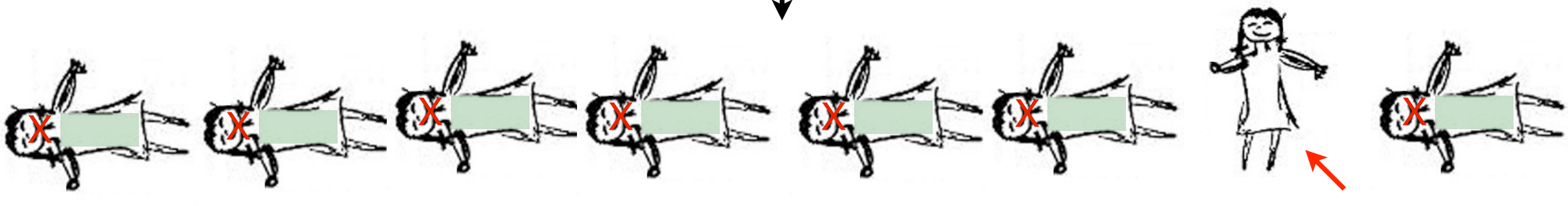
Alternate approach to find antibiotics (effective but non-ideal)



Add 1,000,000 test chemicals, each chemical in a different person







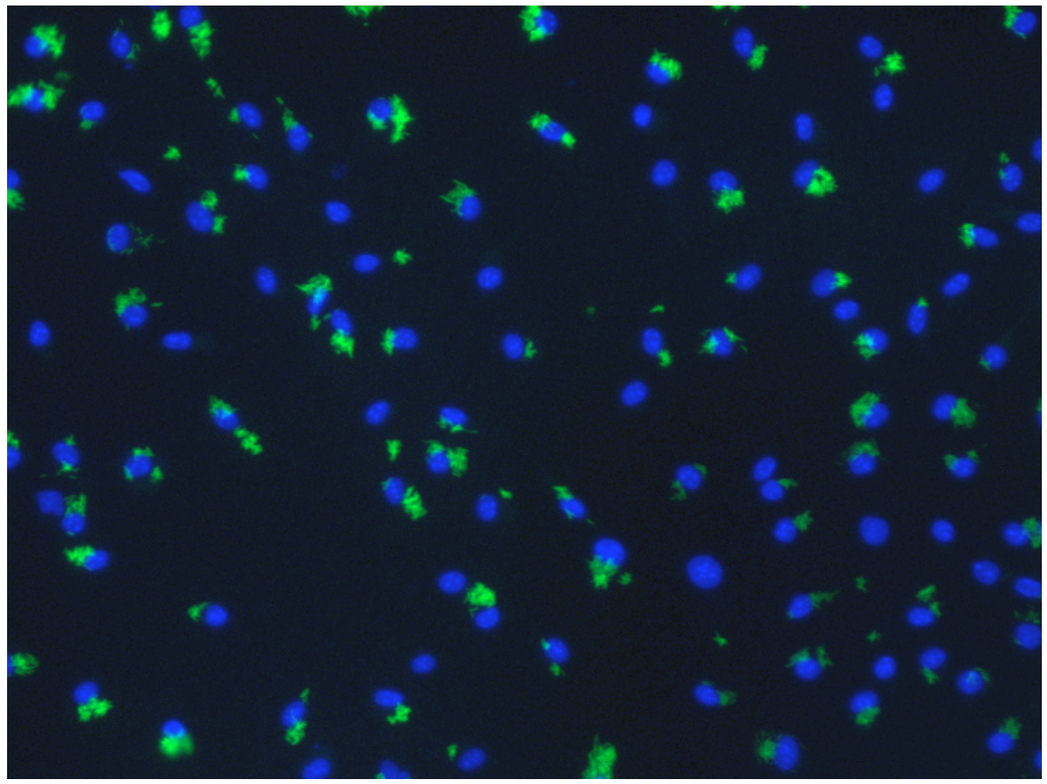
Search for tuberculosis treatments

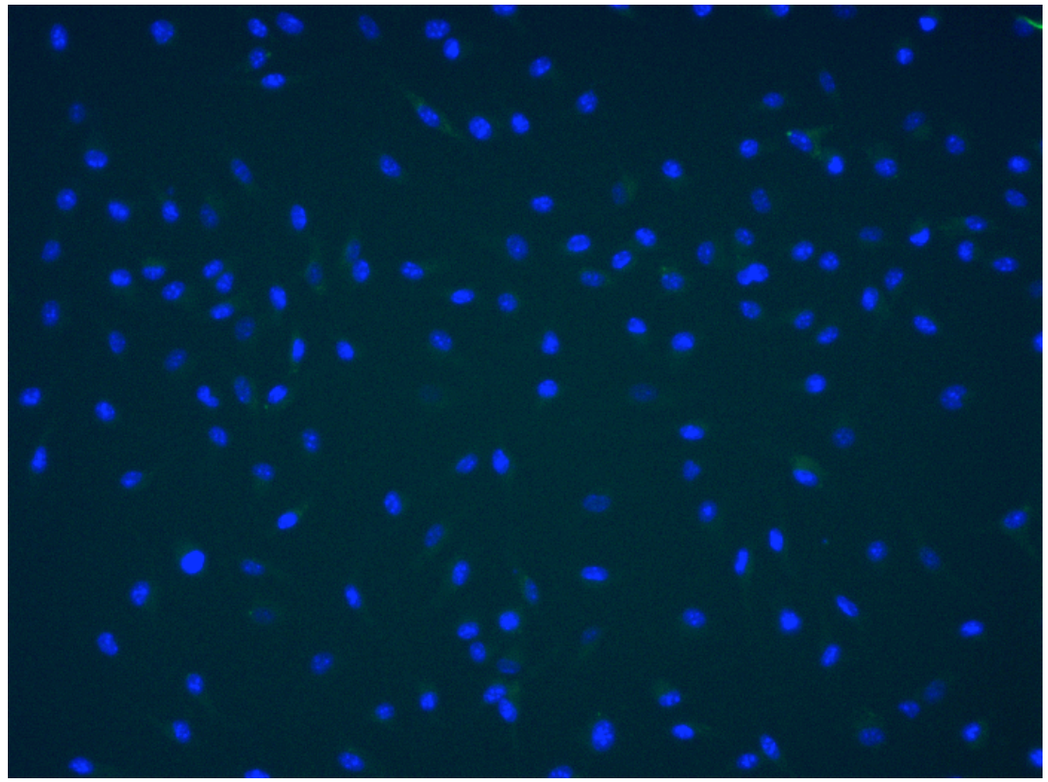
Without drug

With drug

human nuclei

tuberculosis bacteria





Martha Vokes

Mark Bray

BROAD INSTITUTE IMAGING PLATFORM

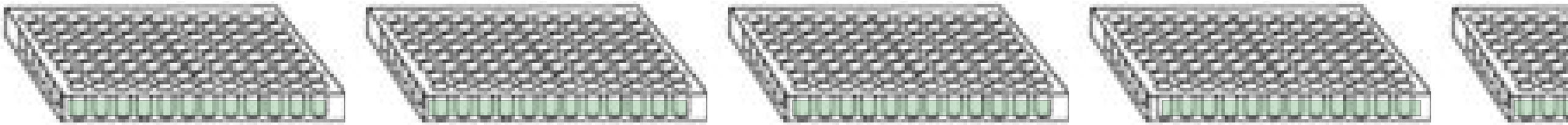
project in progress

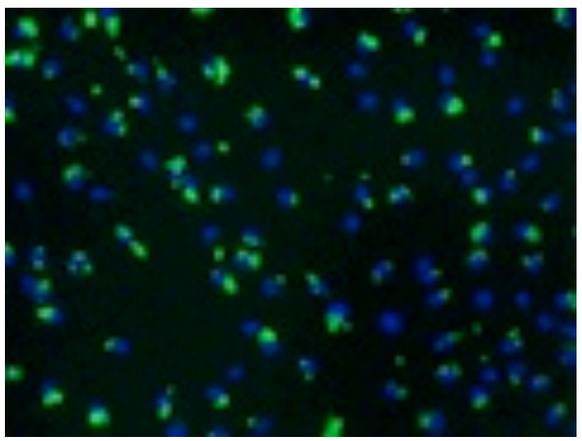
Deb Hung, Broad/MGH

Sarah Stanley, postdoc

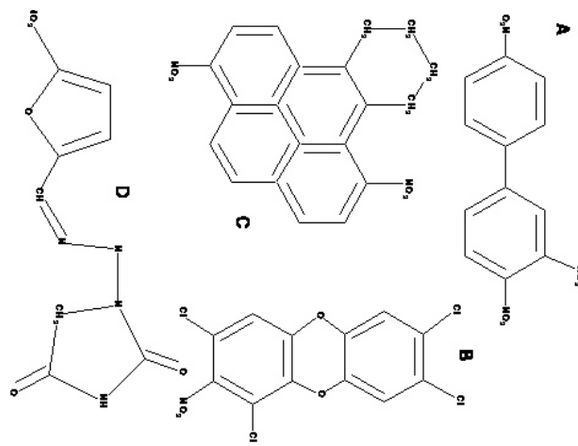
Search for tuberculosis treatments

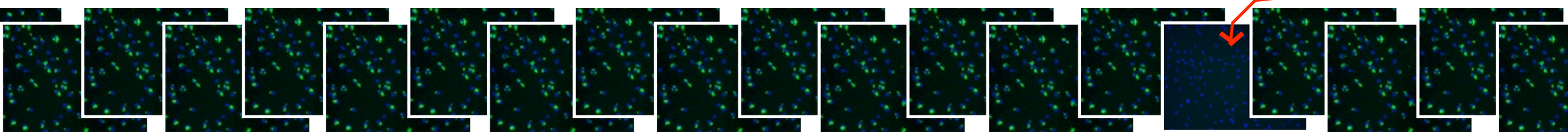
Put bacteria and human cells in individual wells of multi-well plates



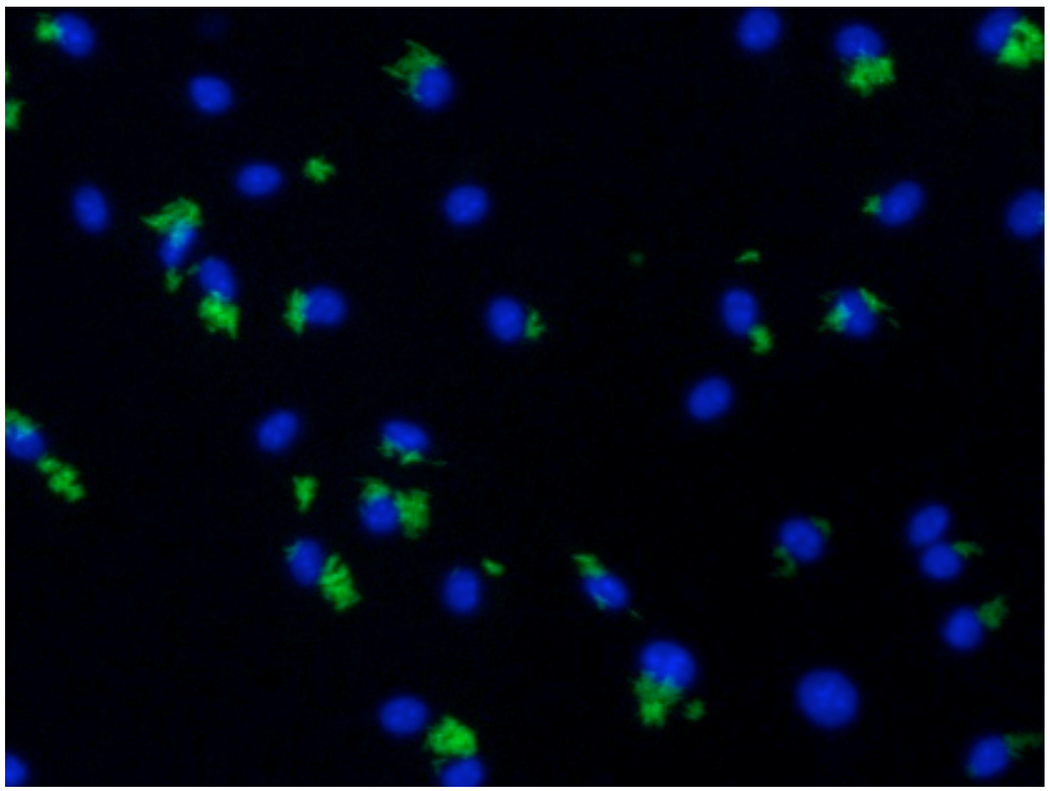


Add 20,000 test chemicals, each chemical in a different well

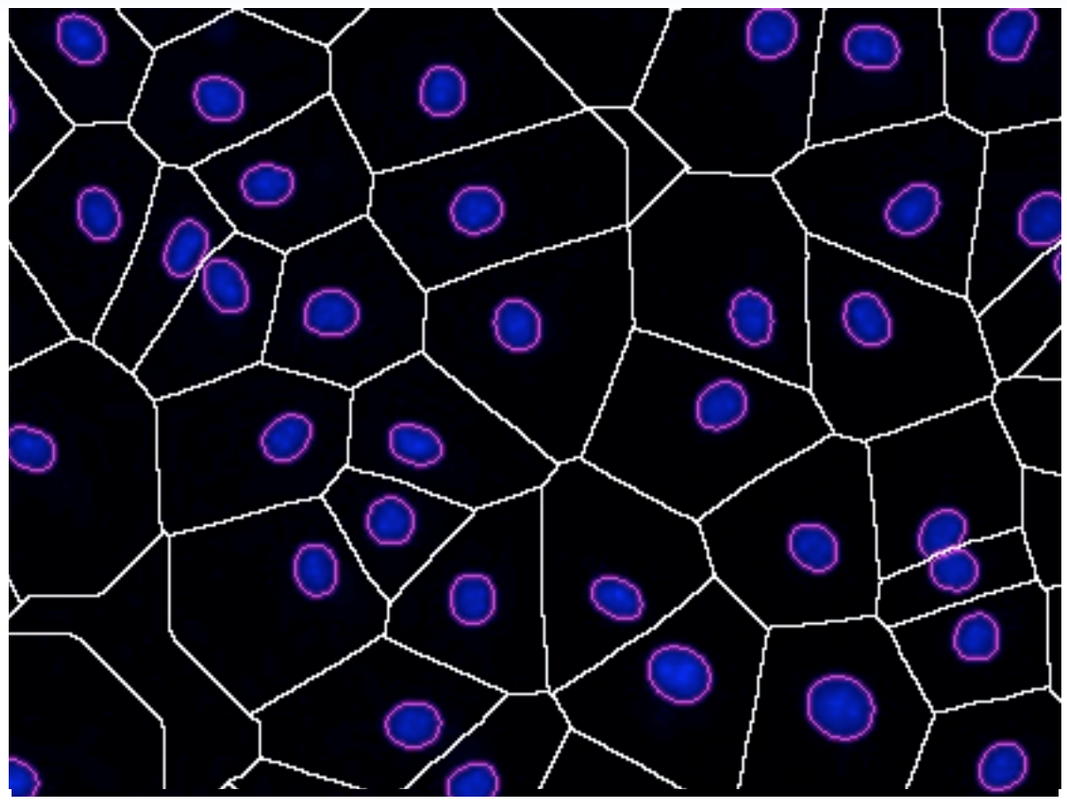




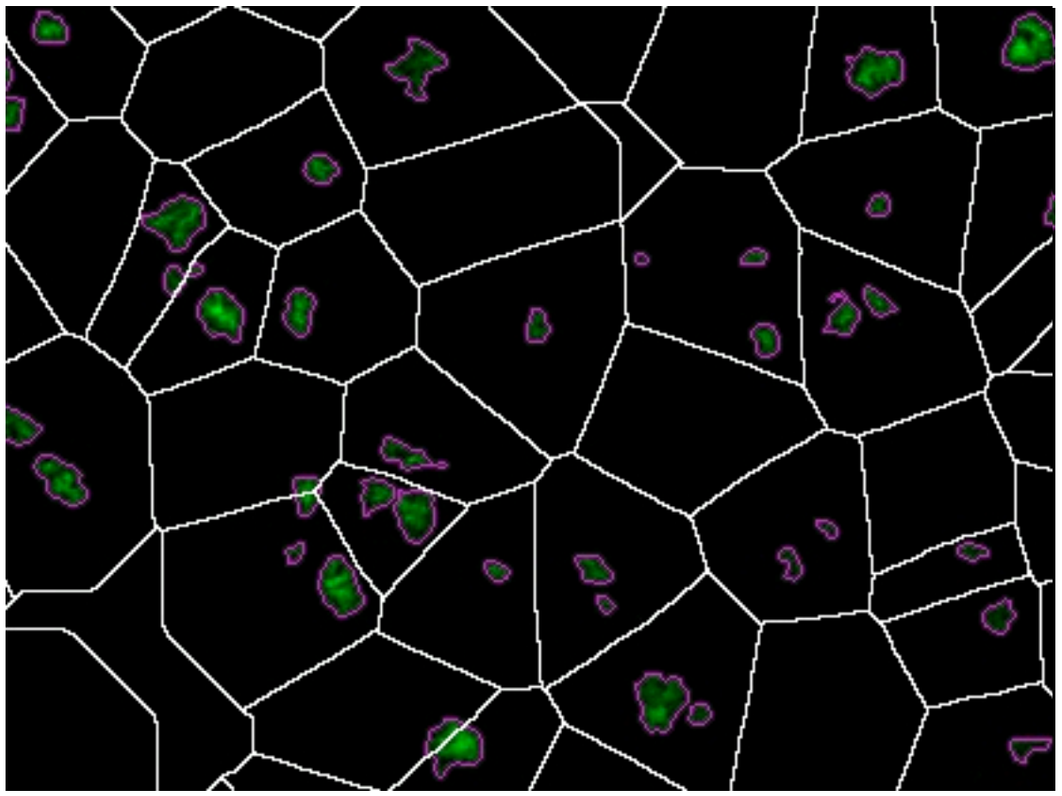
Automated image analysis



Find human nuclei



Find bacteria



Quantify the bacteria per human nucleus

Status: pursuing hits from small-scale bioactive compound screen + scale-up to 30,000 compounds

Martha Vokes

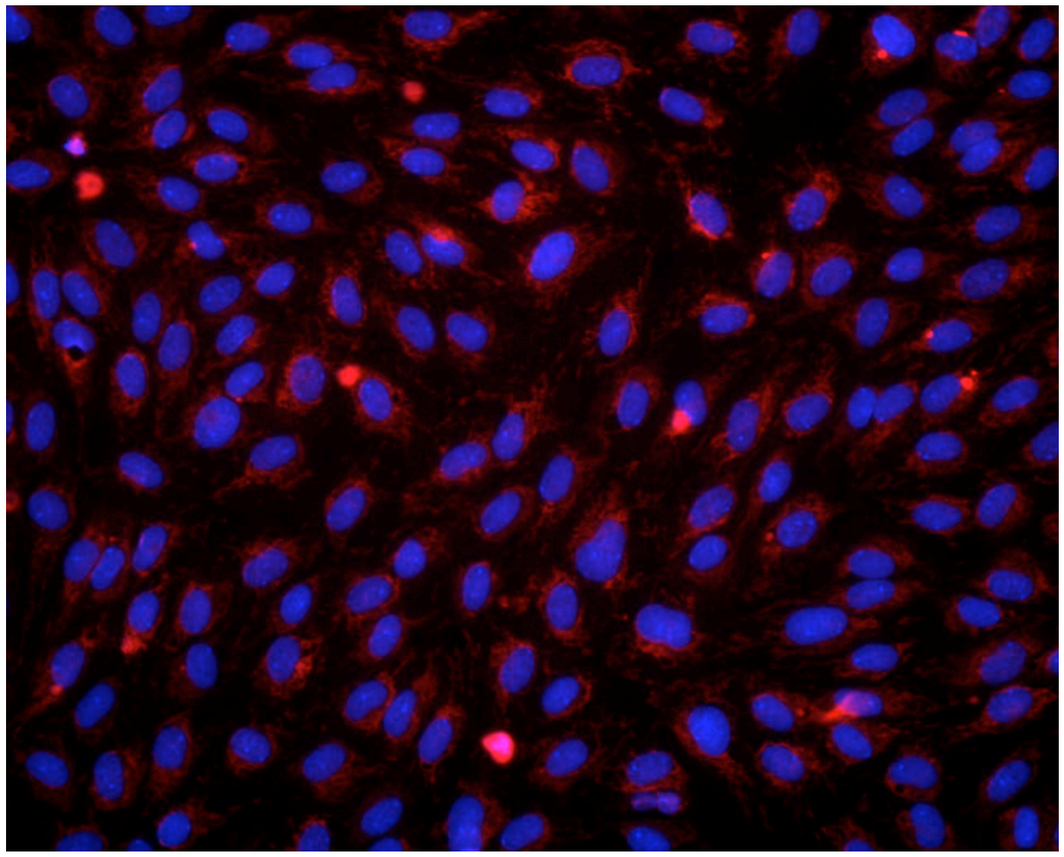
Mitochondrial abundance

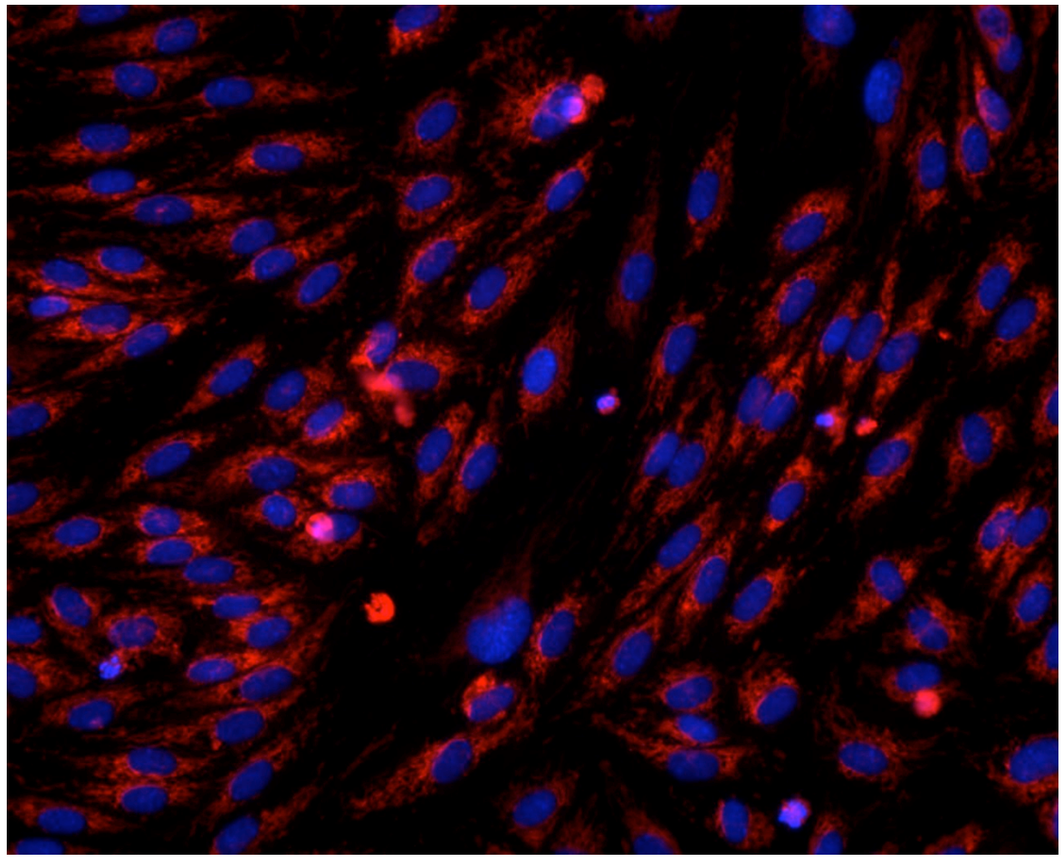
Negative control

Positive control

DNA

Mito-tracker





Ray Jones

Martha Vokes

Vamsi Mootha, Harvard Med/ MGH

Toshi Kitami, postdoc

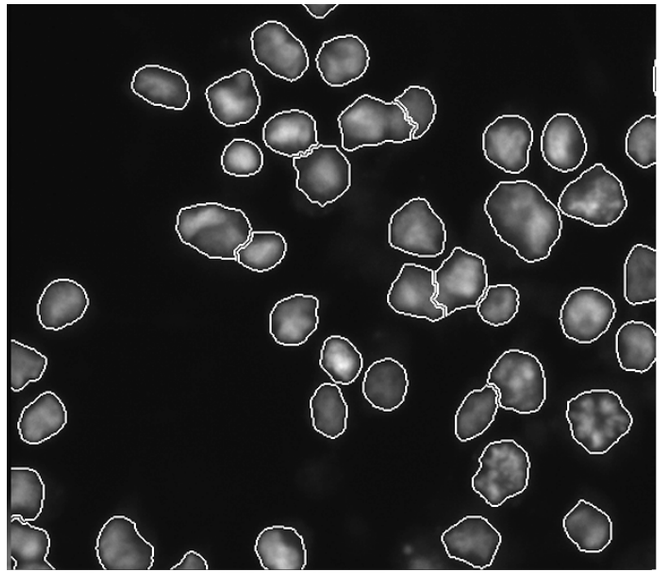
BROAD INSTITUTE IMAGING PLATFORM

project in progress

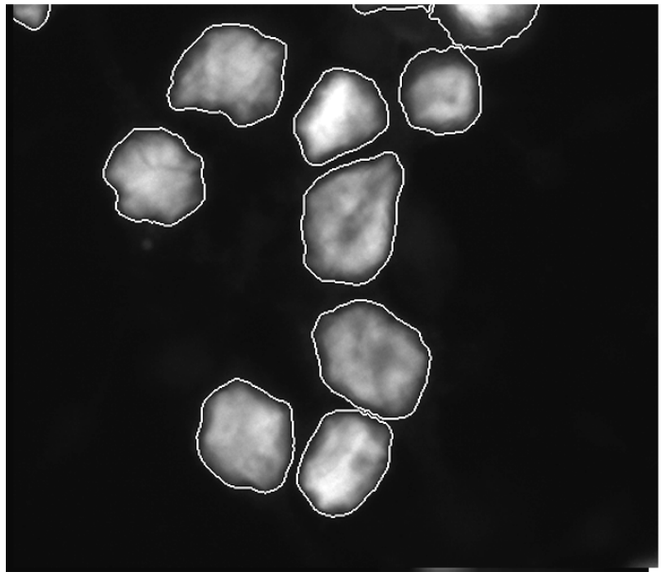
Polyploidization of megakaryocytes - AMKL (leukemia)

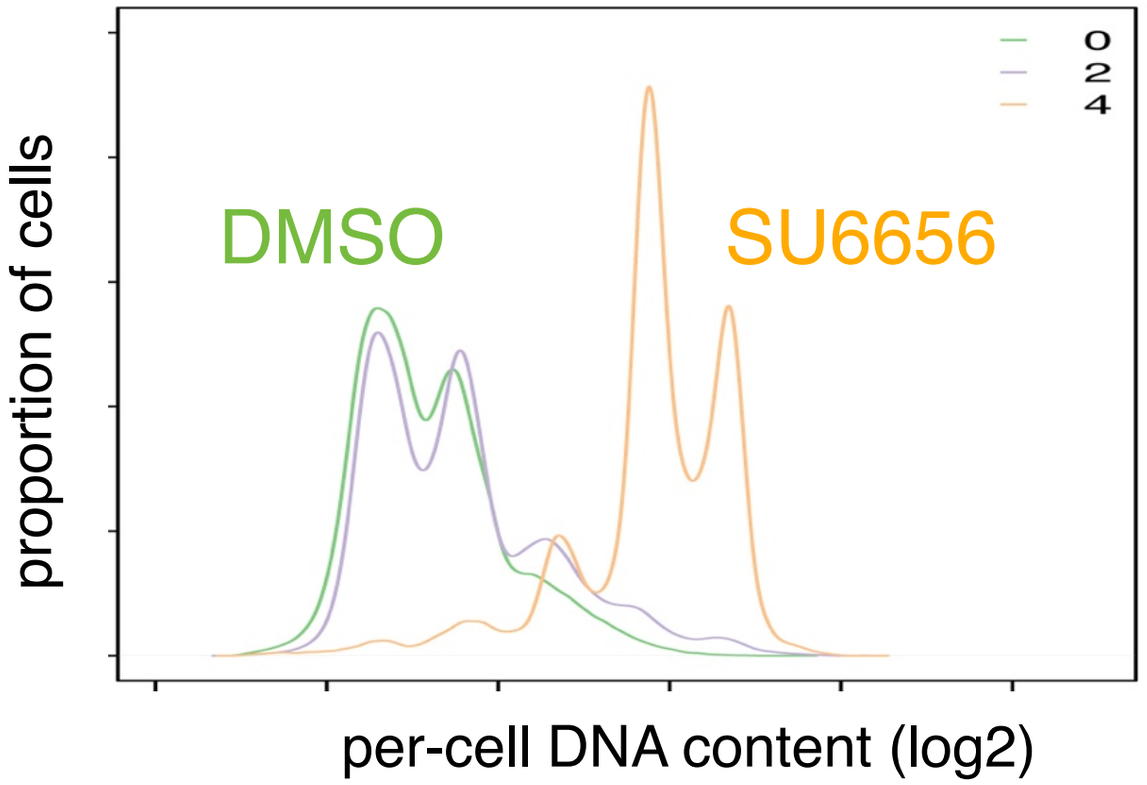
DNA stain, with outlines identifying the nuclei

DMSO
(negative control)



SU6656
(positive control)





project in progress

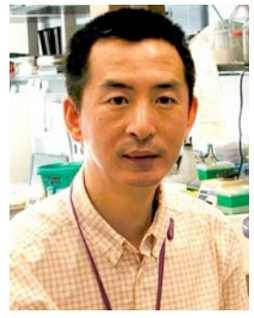
Status: *in vivo* testing of hits from screen of 10,000 compounds

Martha Vokes

Mark Bray

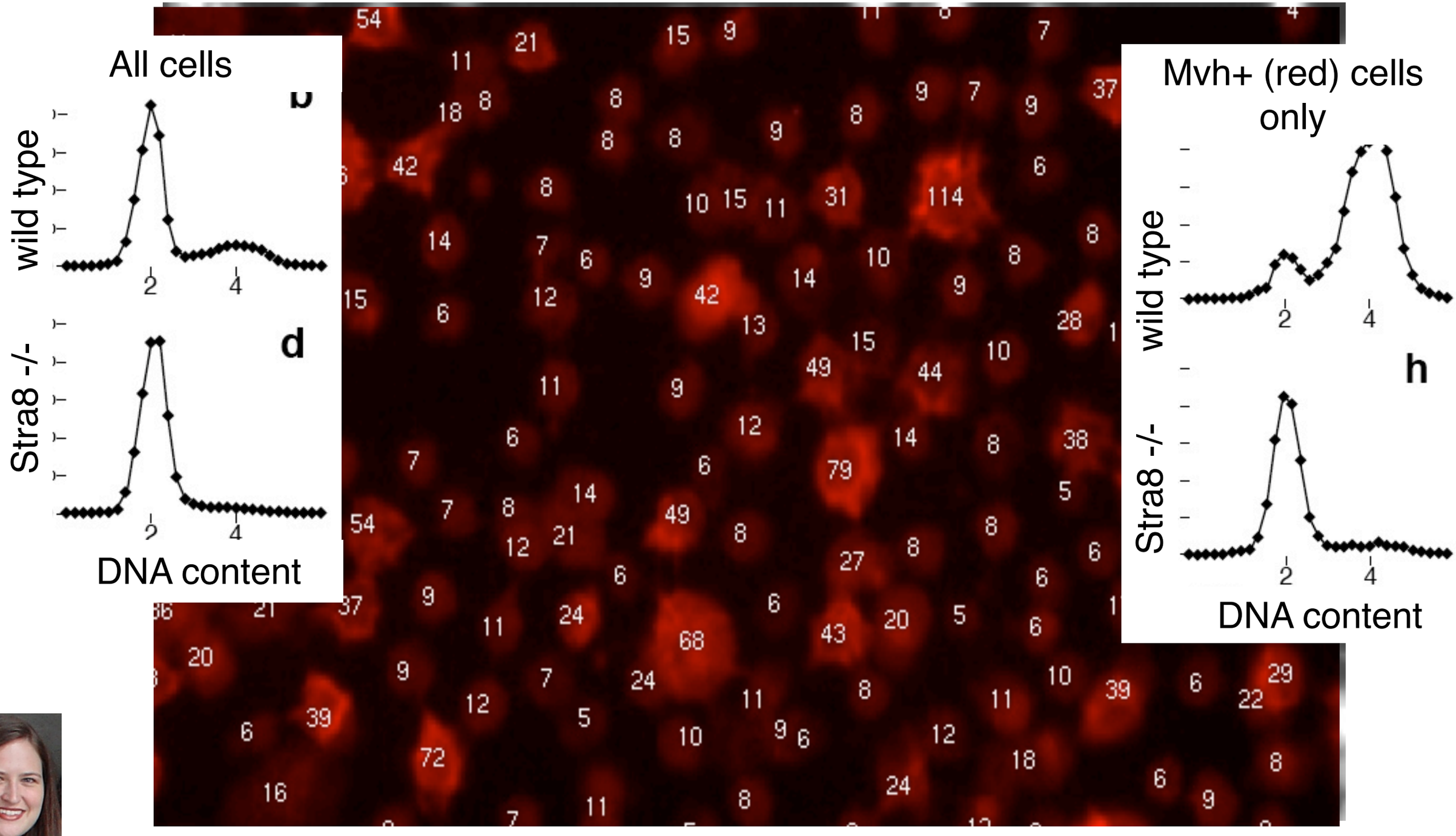
BROAD INSTITUTE IMAGING PLATFORM

John Crispino, Northwestern University



Jeremy Wen, postdoc

DNA and antibody staining intensity



Anne Carpenter

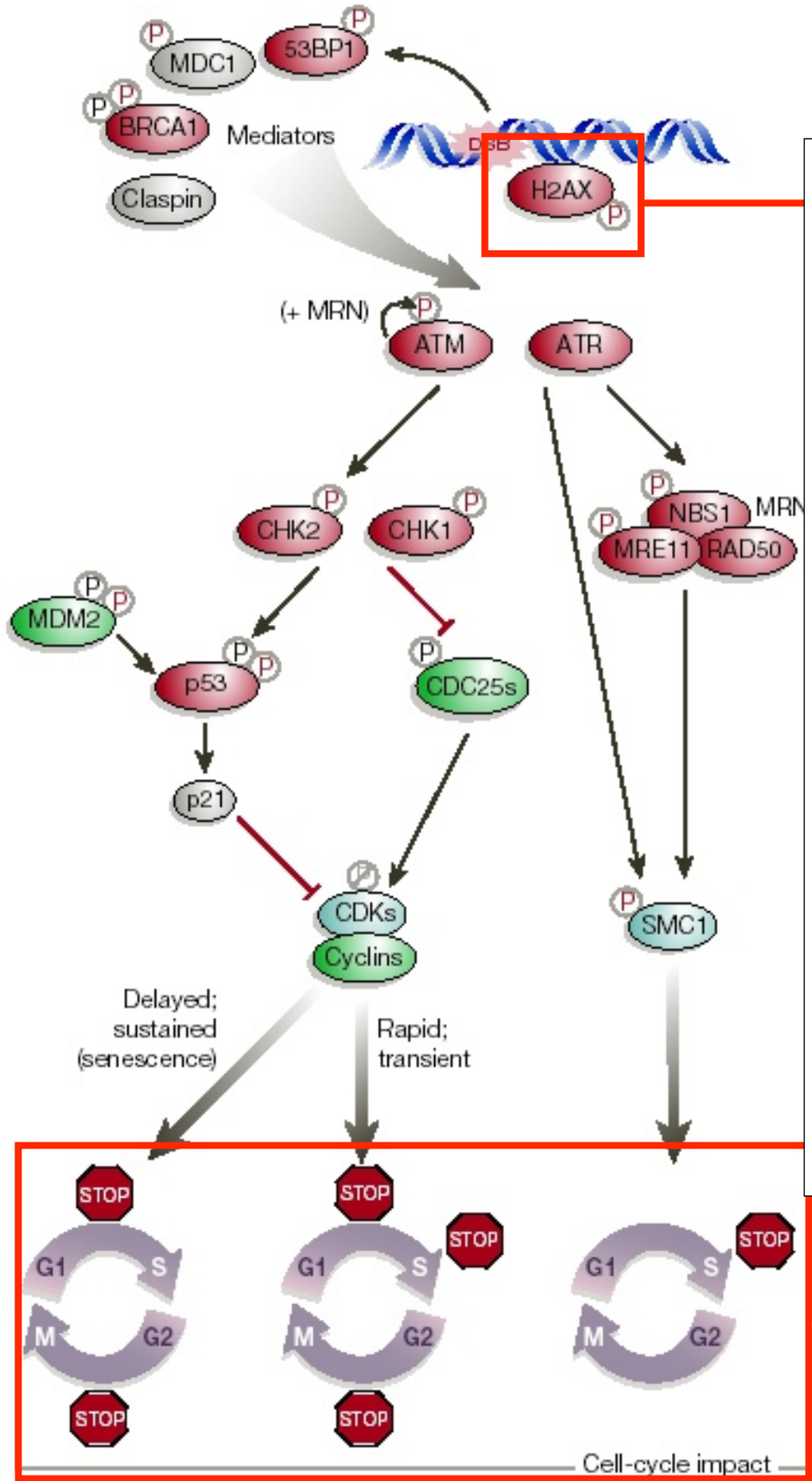
BROAD INSTITUTE IMAGING PLATFORM

Baltus, ... Carpenter, et al., Nature Genetics, 2006

David Page, Whitehead Institute

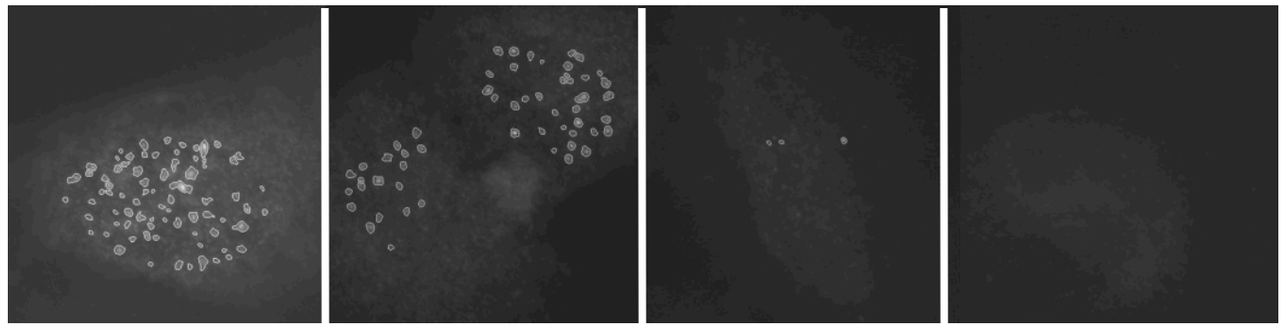
Andrew Baltus, postdoc

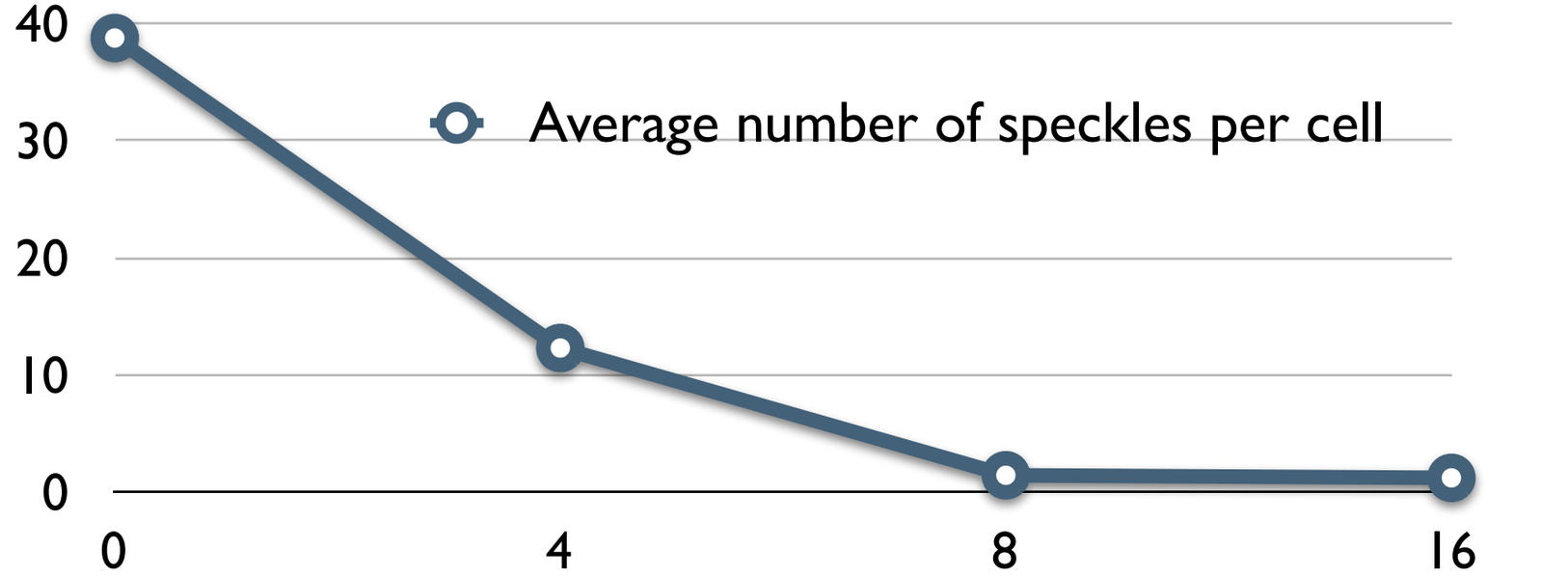
Screen genes/drugs for DNA damage response



Goal: find new drug targets in the DNA damage response pathway to improve efficacy and reduce side effects of existing treatments.

phospho-Histone 2AX

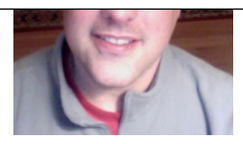




project in progress

Mike Yaffe, MIT

Michael Pacold, postdoc



Scott Floyd, postdoc

Novel antibiotics against *E. faecalis*

Control

Rescuing antibiotic

Bright-field

Ray Jones

Anne Carpenter

BROAD INSTITUTE IMAGING PLATFORM

Moy ...Carpenter ... Ausubel, et al. ACS Chem Bio 2009

Fred Ausubel, Harvard/ Mass. General Hospital

Terry Moy

Annie Lee Conery

Gang Wu

Novel antibiotics against *E. faecalis*

37,214 compounds and extracts in primary screen

↓

108 compounds and extracts retested positive for curing activity

80 known antibiotics, antibiotic analogs, or documented antimicrobial activity

28 other compounds extracts

↓

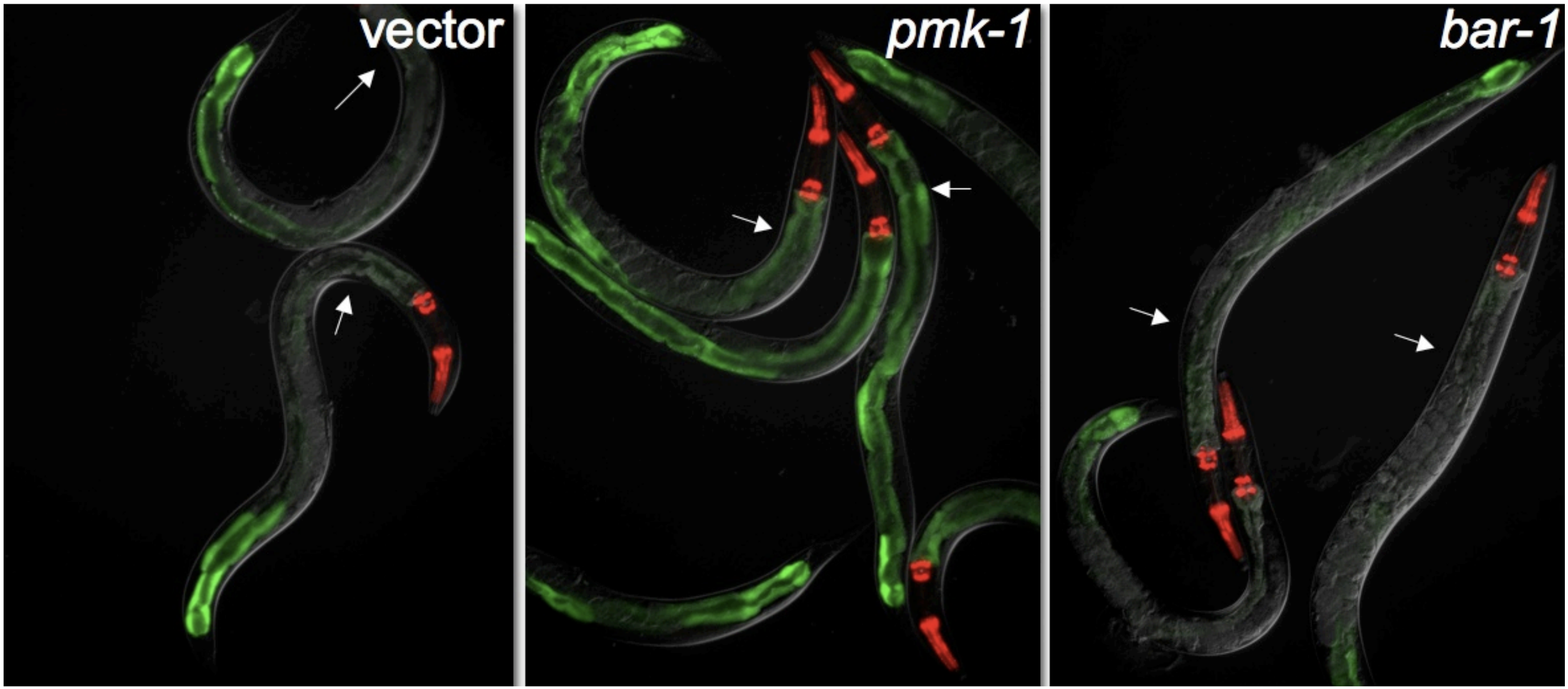
21 commercially available compounds selected for dose responses and MIC

↓

Six structural classes with “anti-infective” profile; i.e. cure infection but not antibiotics

Moy ...Carpenter ... Ausubel, et al. ACS Chem Bio 2009

Reporter expression in response to infection




Carolina Wählby


Kate Madden


Zihan Hans Liu

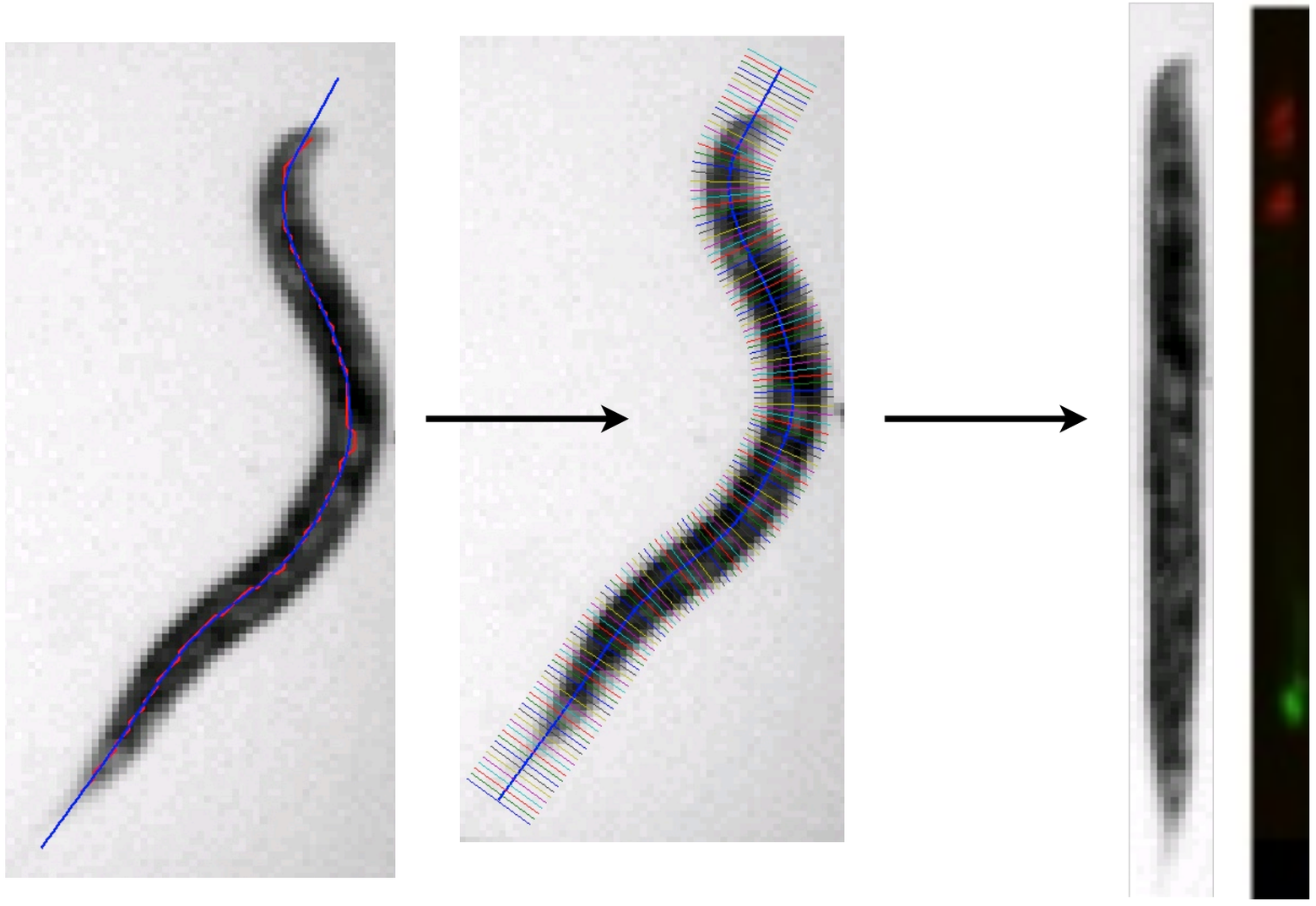
BROAD INSTITUTE IMAGING PLATFORM

project in progress


Javier Irazoqui

Reporter expression in response to infection

Goal: compare pattern of GFP along the length of the worm



Approach: ‘Straighten’ worms

Carolina Wählby

Kate Madden

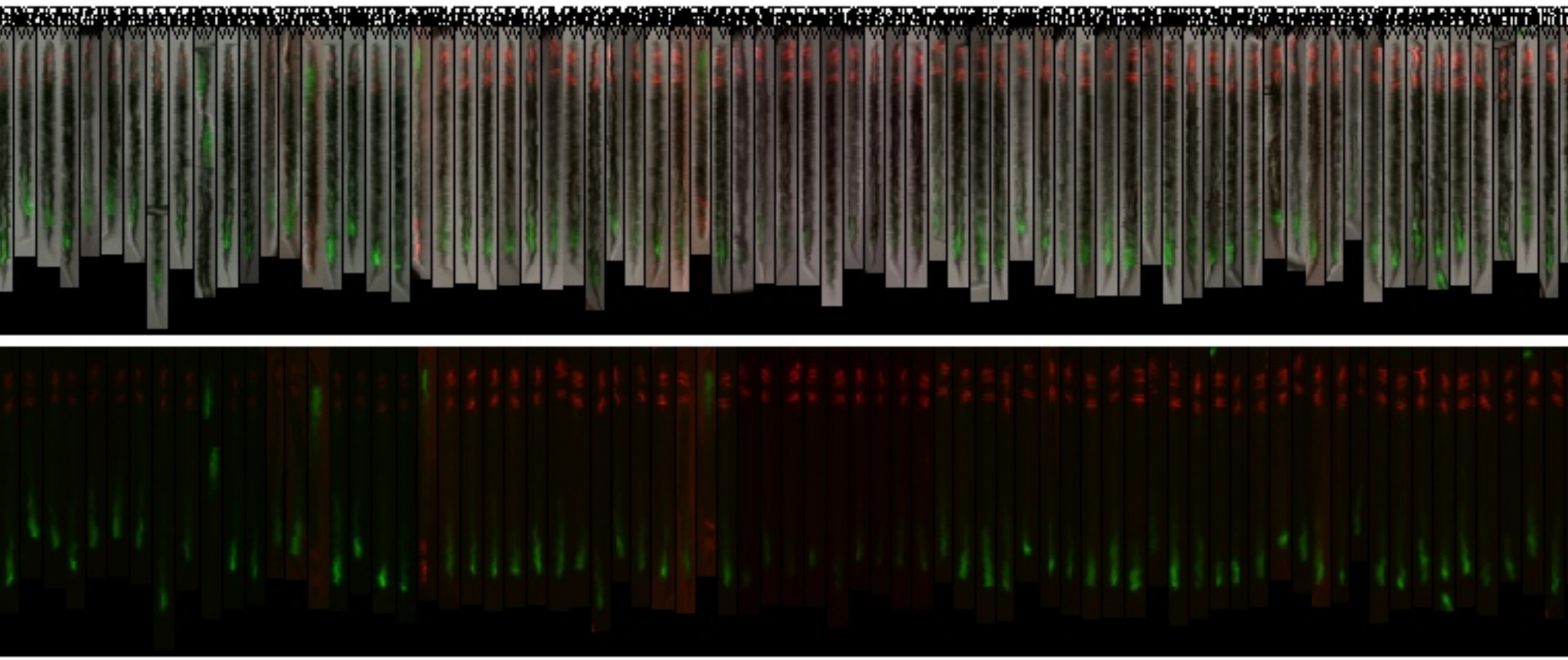
Zihan
Hans
Liu

Javier Irazoqui

BROAD INSTITUTE IMAGING PLATFORM

project in progress

Reporter expression in response to infection



Carolina Wählby

Kate Madden

Zihan Hans Liu

BROAD INSTITUTE IMAGING PLATFORM

project in progress

Javier Irazoqui

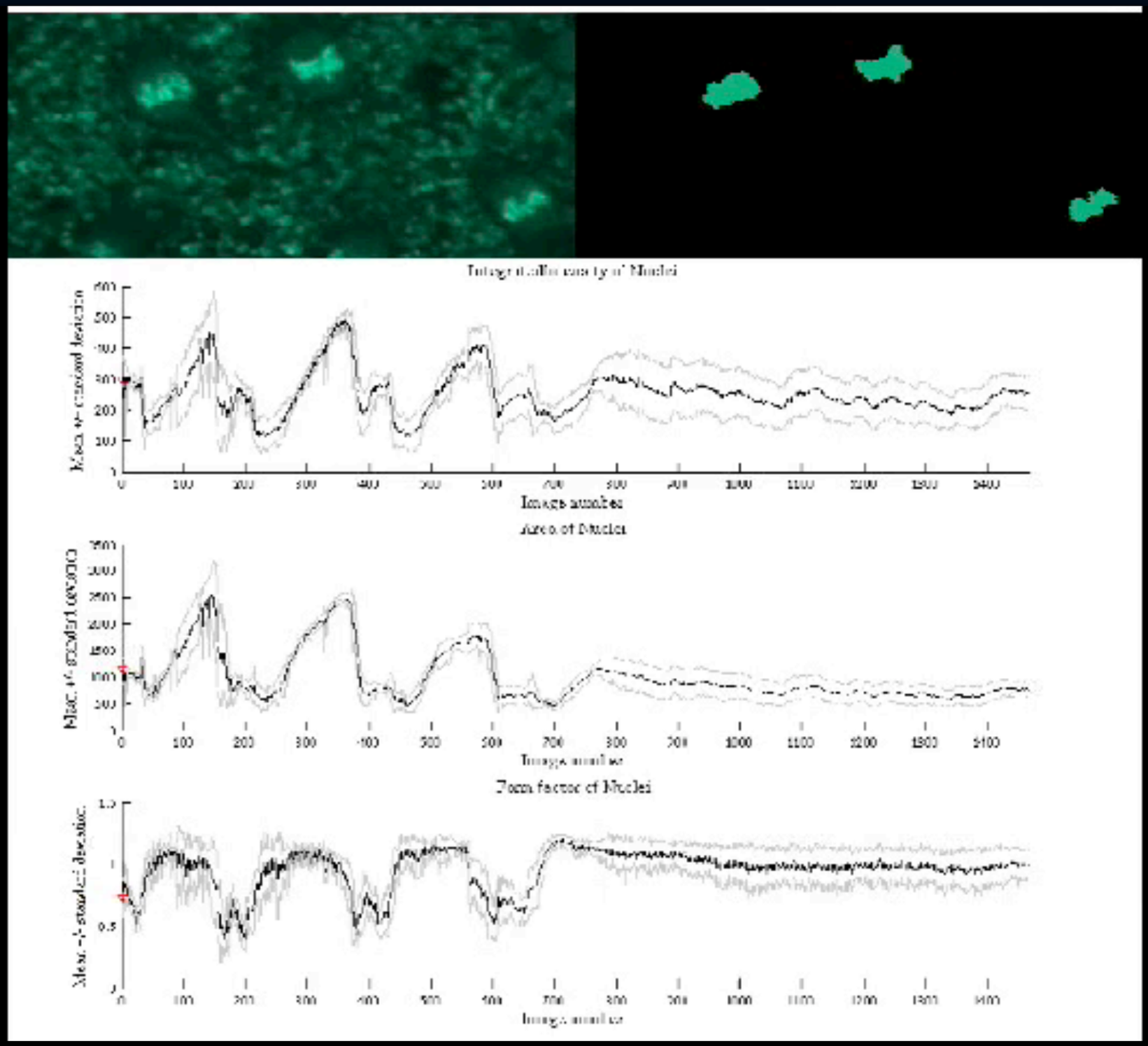
Extracting the wealth of information

movie from Victoria Foe,
Univ. Washington

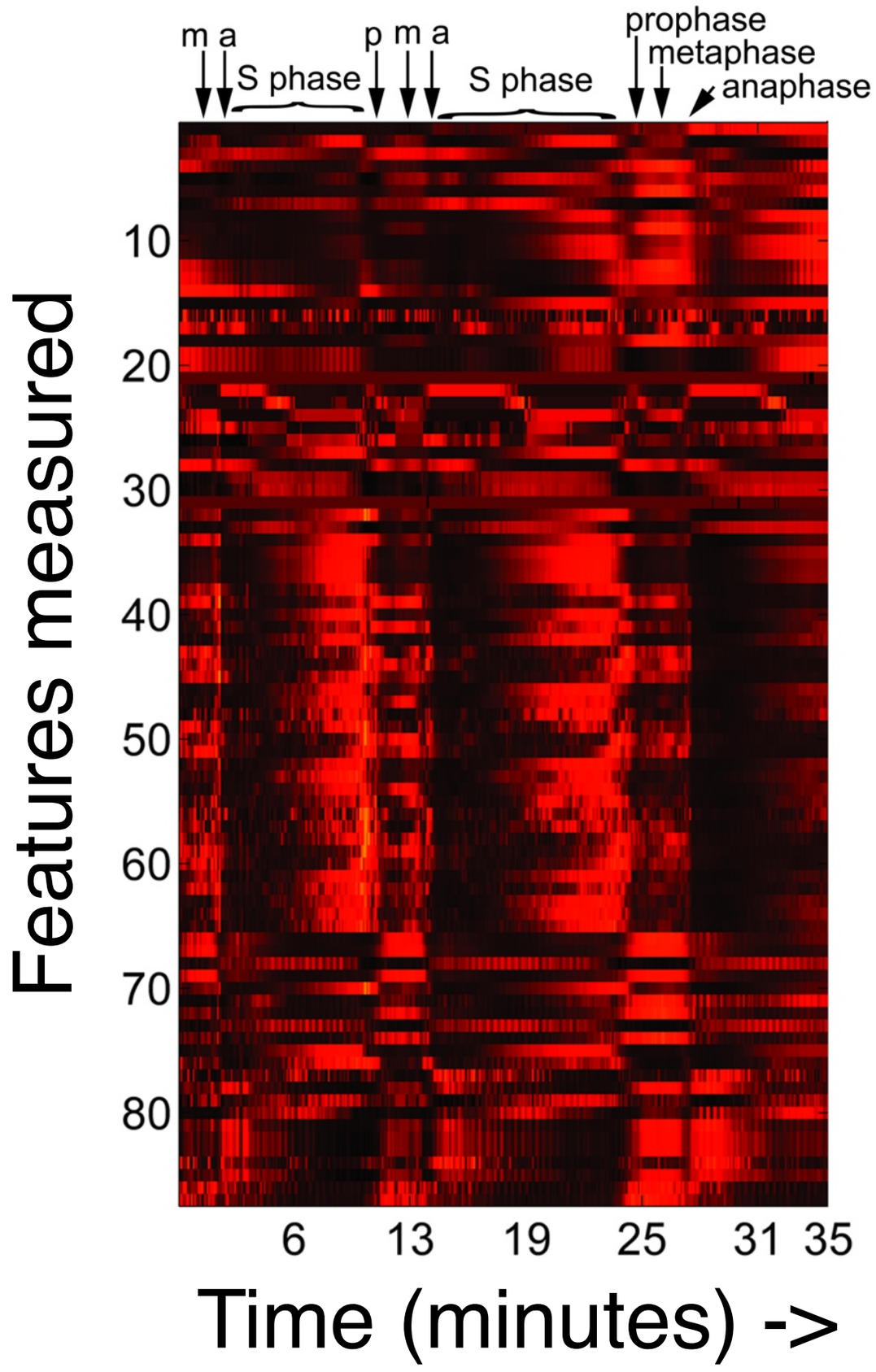
GFP content

Area

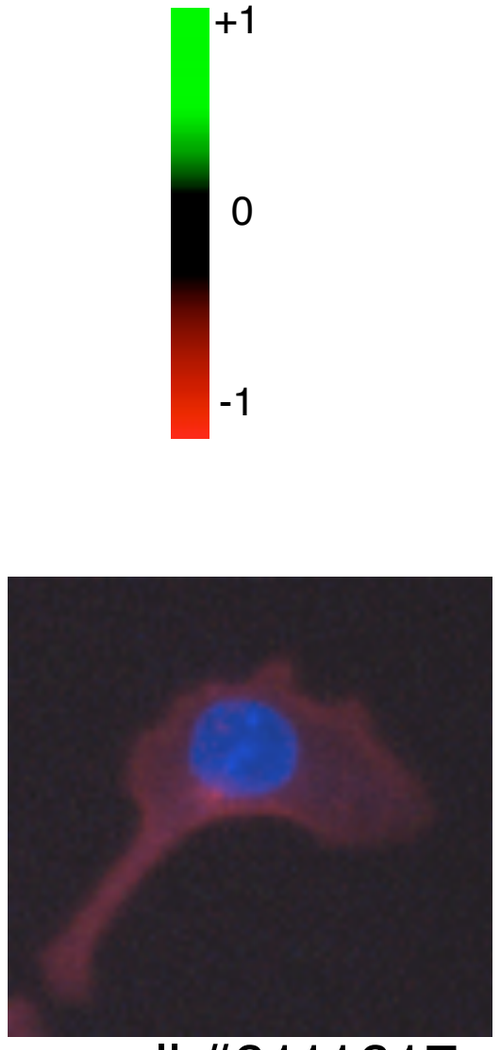
Shape



Extracting the wealth of information



Measure everything…ask questions later.



cell #6111617

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

Perlman, et al. Science 2004

Measure everything…ask questions later.

~500 features per cell: size, shape, staining intensity, texture (smoothness), etc.

Why?

(a) Several features may be necessary to score the phenotype
(b) Virtual secondary screens can help characterize hits
(c) Later re-screening for new phenotypes

Exploring multi-feature cell data

CellProfiler™ Analyst
data exploration software

CellProfiler Analyst - 2006_06_14_NIRHT

File Plot Classify Show Info Window

phospho-H3 intensity

DNA content

Cell Montage

Per_Image, sp

File

Histogram

File Plot Search Show Info

Histogram

Histogram

File Plot Search Show Info

Histogram

Entrez Gene: PPP2R2A protein phosphatase 2 (formerly 2A), regulatory subunit B (PR 52), alpha isoform [Homo sapiens]

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=Retrieve&do

Lit Searches OurImgAnalysis Personal YahooMail LabWebsites GoogleMaps Weather ExploreGeneList News CellVisualizerTheMo...

NCBI Entrez Gene

My NCBI [Sign In] [Register]

All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books OMIM

Search Gene for Go Clear

Limits Preview/Index History Clipboard Details

Display Full Report Show 5 Send to

All: 1 Current Only: 1 Genes Genomes: 1 SNP GeneView: 1

1: PPP2R2A protein phosphatase 2 (formerly 2A), regulatory subunit B (PR 52), alpha isoform [*Homo sapiens*]

Entrez Gene Home

updated 06-Oct-2006

GeneID: 5520

File

UniqueIDs	Total Obj Cl
24176	591
24167	497
36068	162
36066	153
39991	161
15985	126
15660	276
15659	174
15657	199
39200	175
39191	152
15000	95
18585	96

The RNAi Consortium @ Broad, *Moffat, et al. Cell 2006*; CellProfiler Analyst led by In Han Kang, Golland lab, MIT

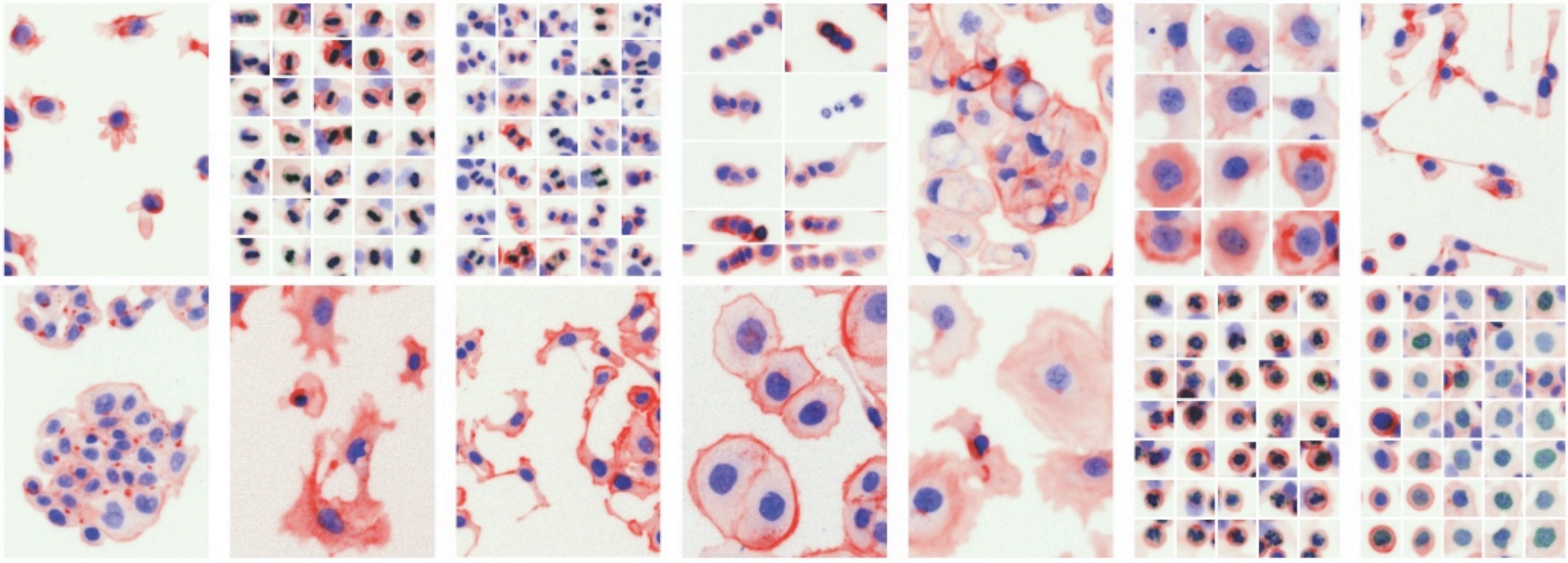
Measure everything…ask questions later.

~500 features per cell: size, shape, staining intensity, texture (smoothness), etc.

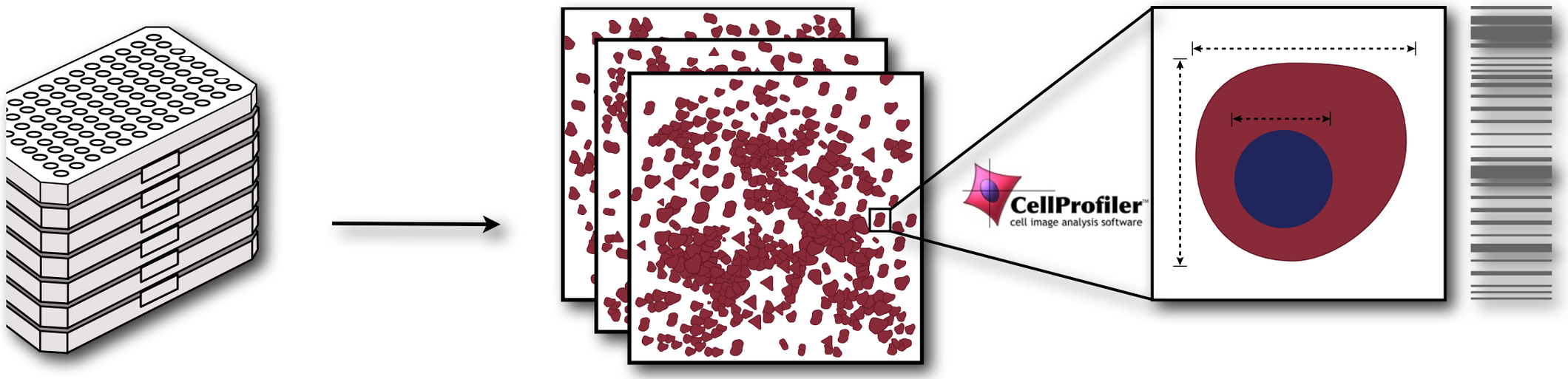
Why?

(a) Several features may be necessary to score the phenotype
(b) Virtual secondary screens can help characterize hits
(c) Later re-screening for new phenotypes
(d) The measurements required to score the phenotype of interest may not be known a priori

Challenging cellular phenotypes



Automated Cell Image Processing



Thousands of wells

10^4 images, 10^3 cells in each:
Total of 10^7 cells/experiment

Each cell with cytoprofile

Cytoprofile of 500+ features measured for each cell


Ray Jones


Anne Carpenter

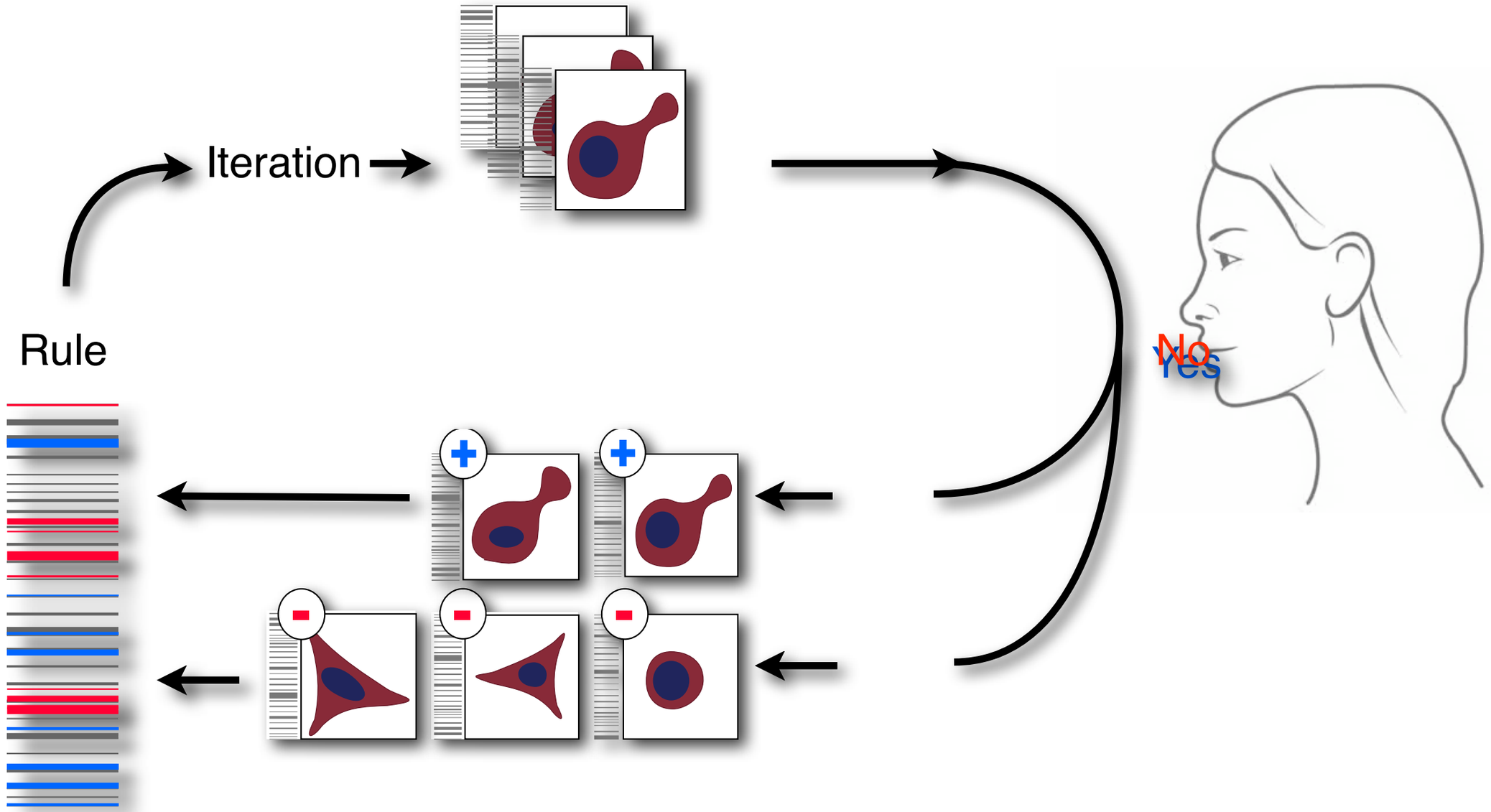
BROAD INSTITUTE IMAGING PLATFORM

Jones, Carpenter ... Sabatini, et al. PNAS 2009

Illustrations by Bang Wong, Nadav Kupiec, & Christopher Lewis

Iterative machine learning

System presents ~500 cells to biologists for scoring



System defines rule based on cytoprofile of scored cells


Ray Jones


Adam Fraser

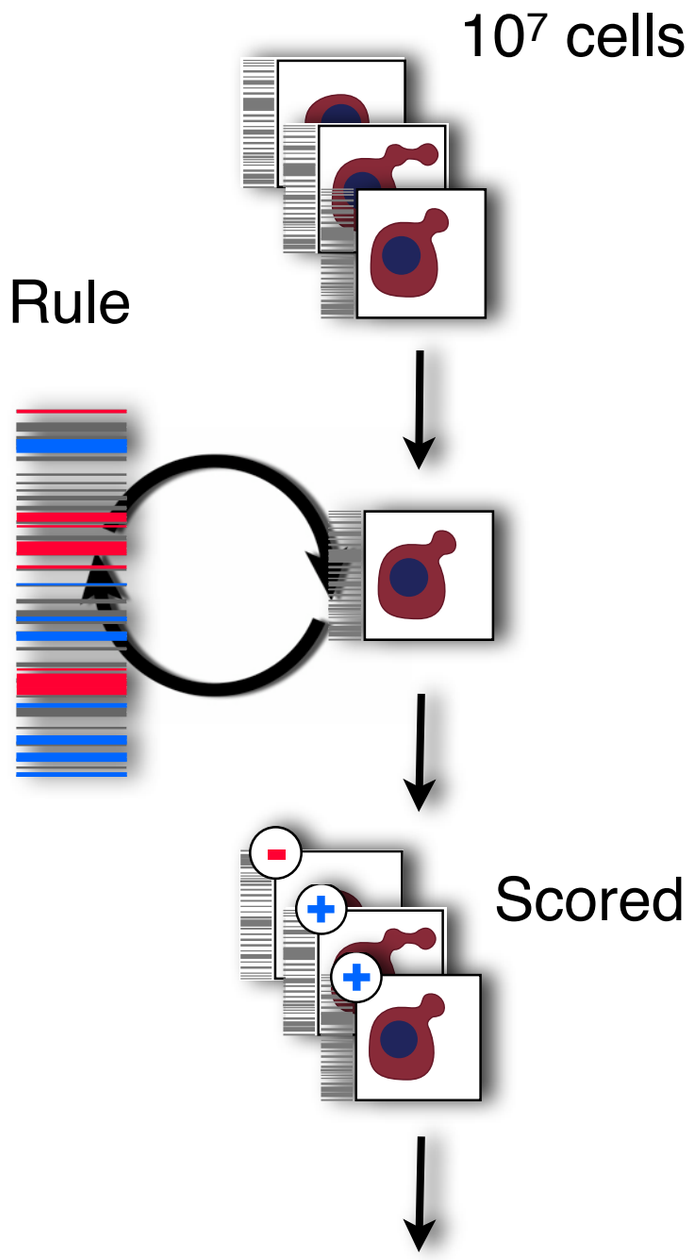
BROAD INSTITUTE IMAGING PLATFORM

CellProfiler™ Analyst
data exploration software

Based on:
- Boosting Image Retrieval (Tieu & Viola, 2000)
- GentleBoosting classifier (Friedman, et al. 1998)

Jones, Carpenter ... Sabatini, et al. PNAS 2009

Automated Scoring



CellProfiler™ Analyst
data exploration software

Scored cells are sorted by well:
identify samples with a high proportion of positive cells


Ray Jones

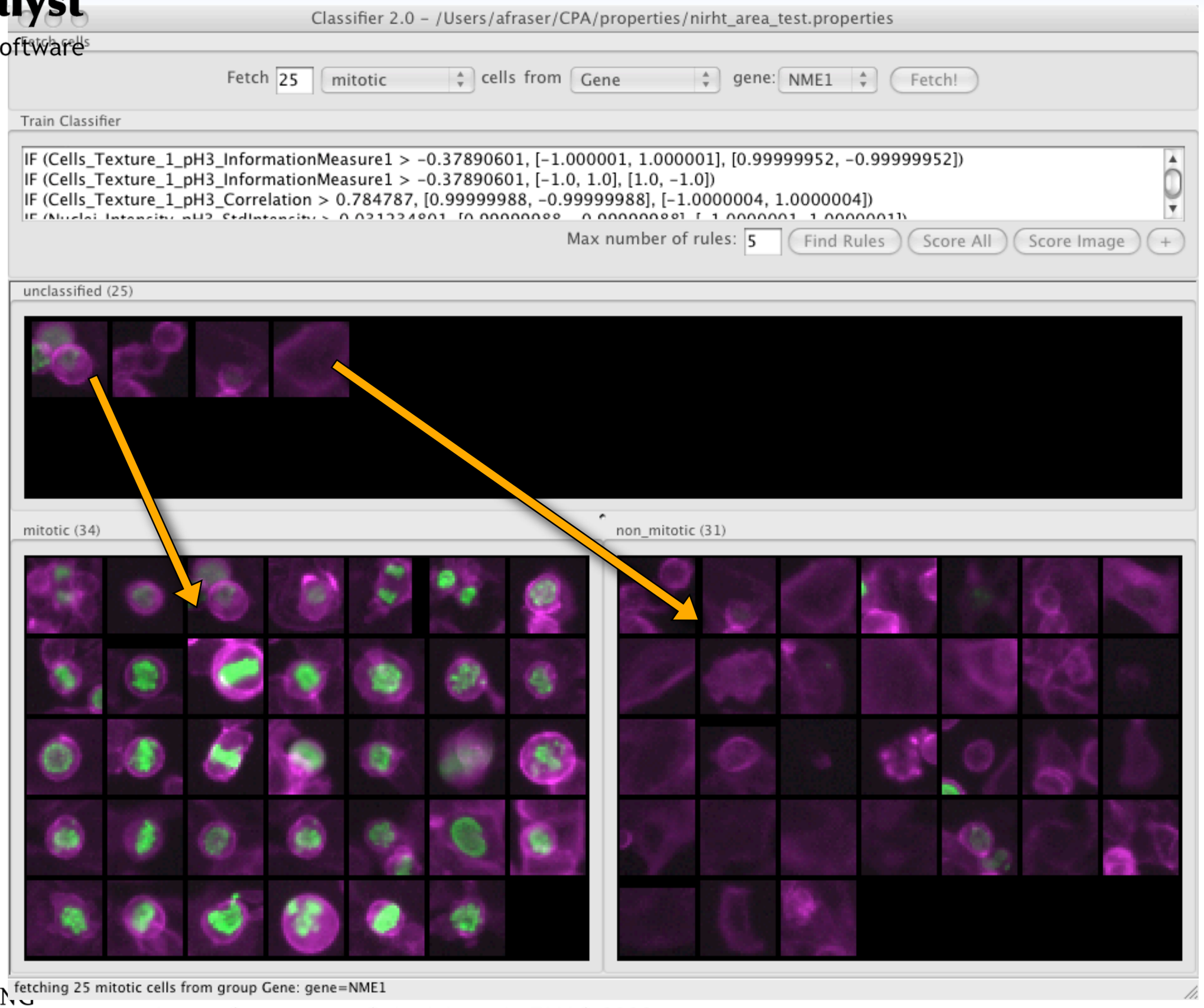

Adam Fraser

BROAD INSTITUTE IMAGING PLATFORM

Jones, Carpenter ... Sabatini, et al. PNAS 2009

CellProfiler Analyst

data exploration software



Ray Jones

Adam Fraser

BROAD INSTITUTE IMAGING PLATFORM

Jones, Carpenter ... Sabatini, et al. PNAS 2009

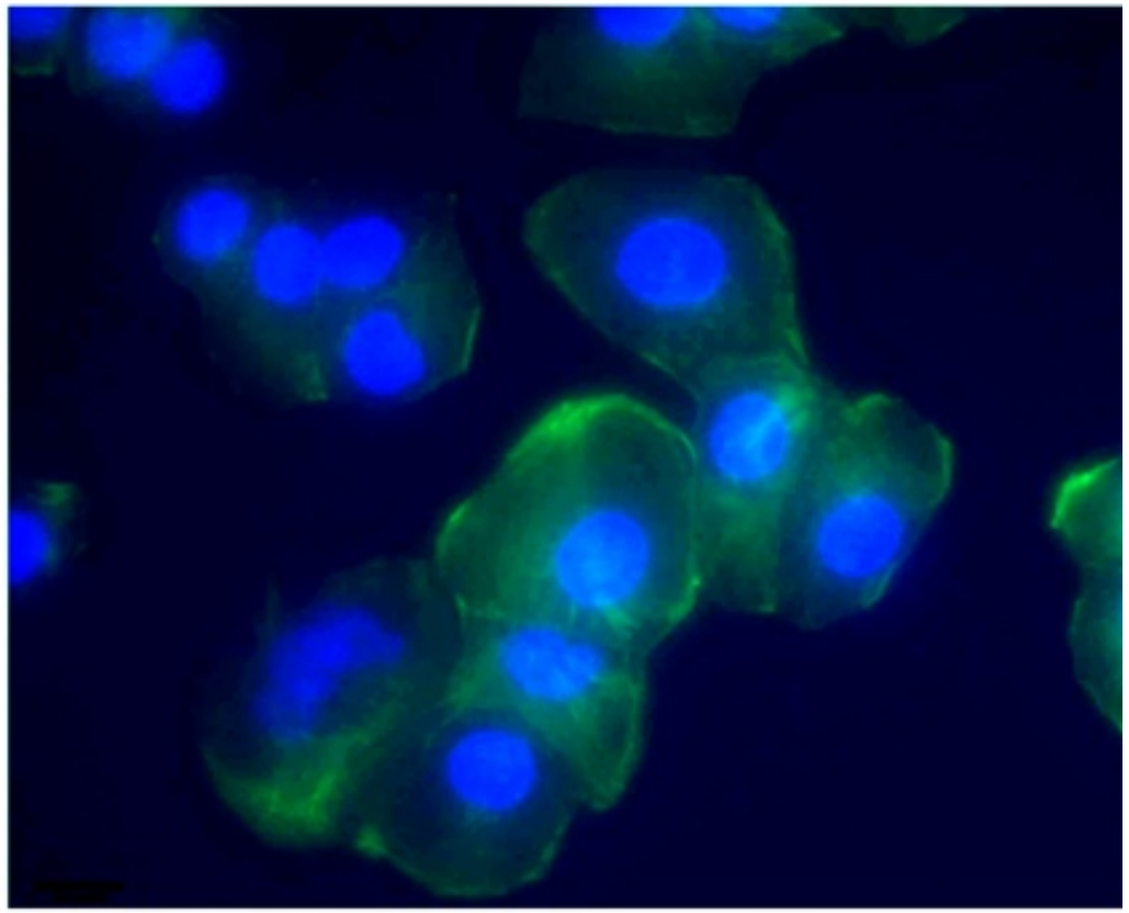
Breast cancer

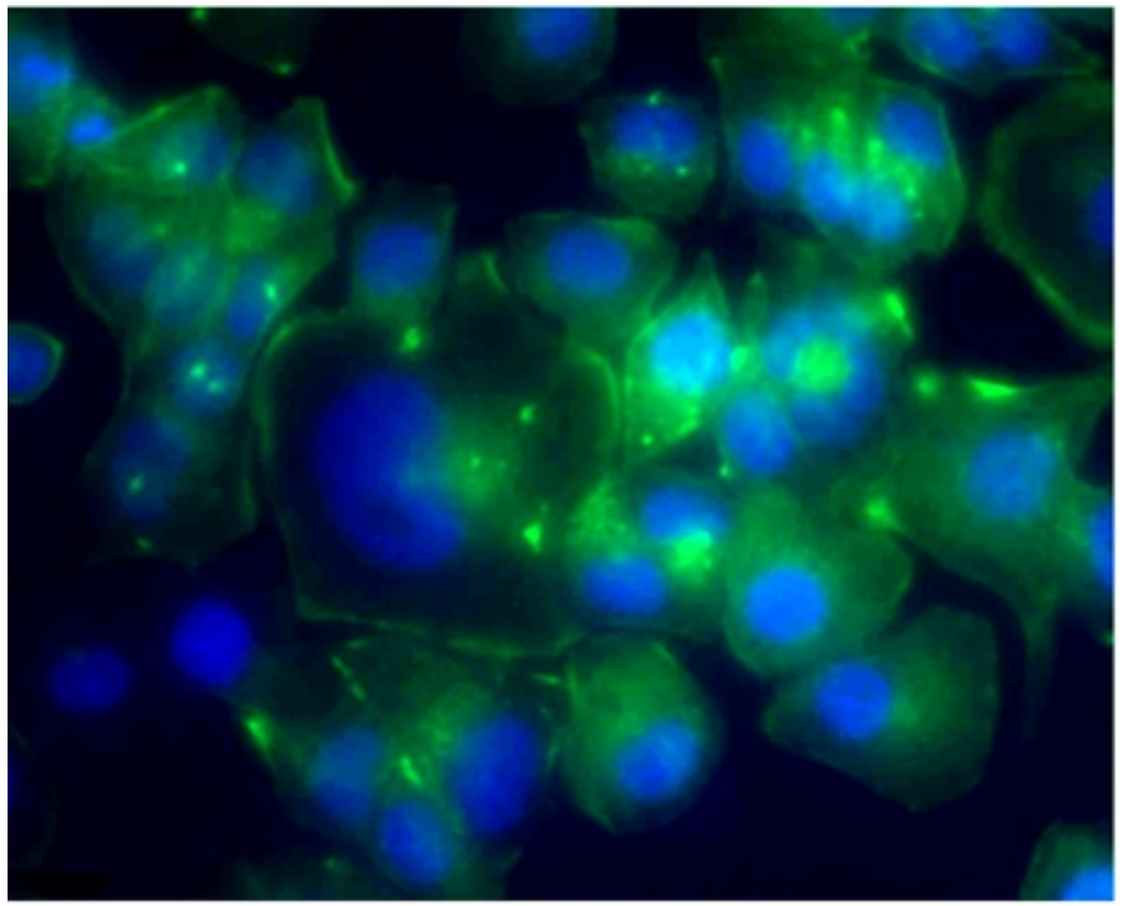
Control

+ Growth factor

DNA

Actin





Ray Jones

Anne Carpenter

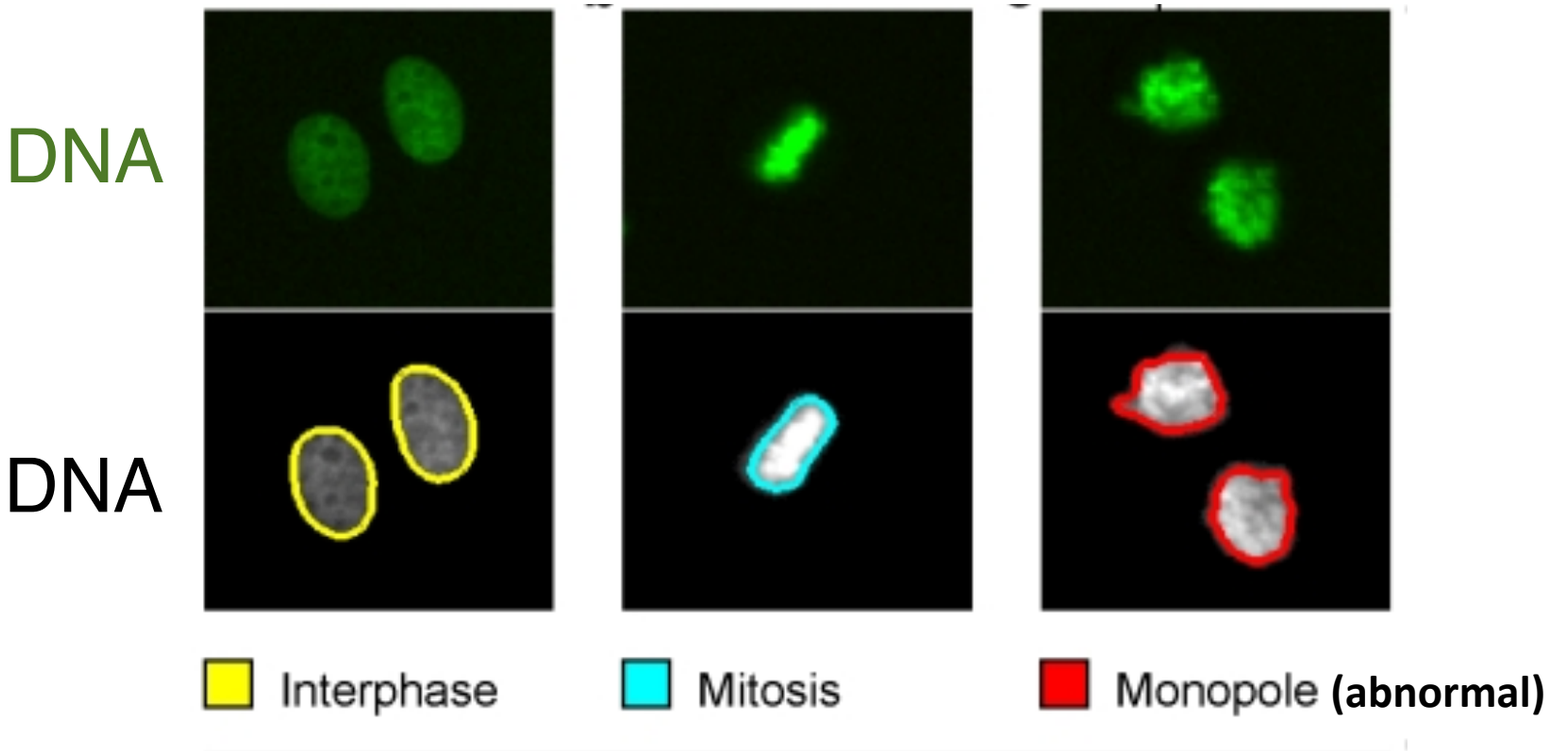
BROAD INSTITUTE IMAGING PLATFORM

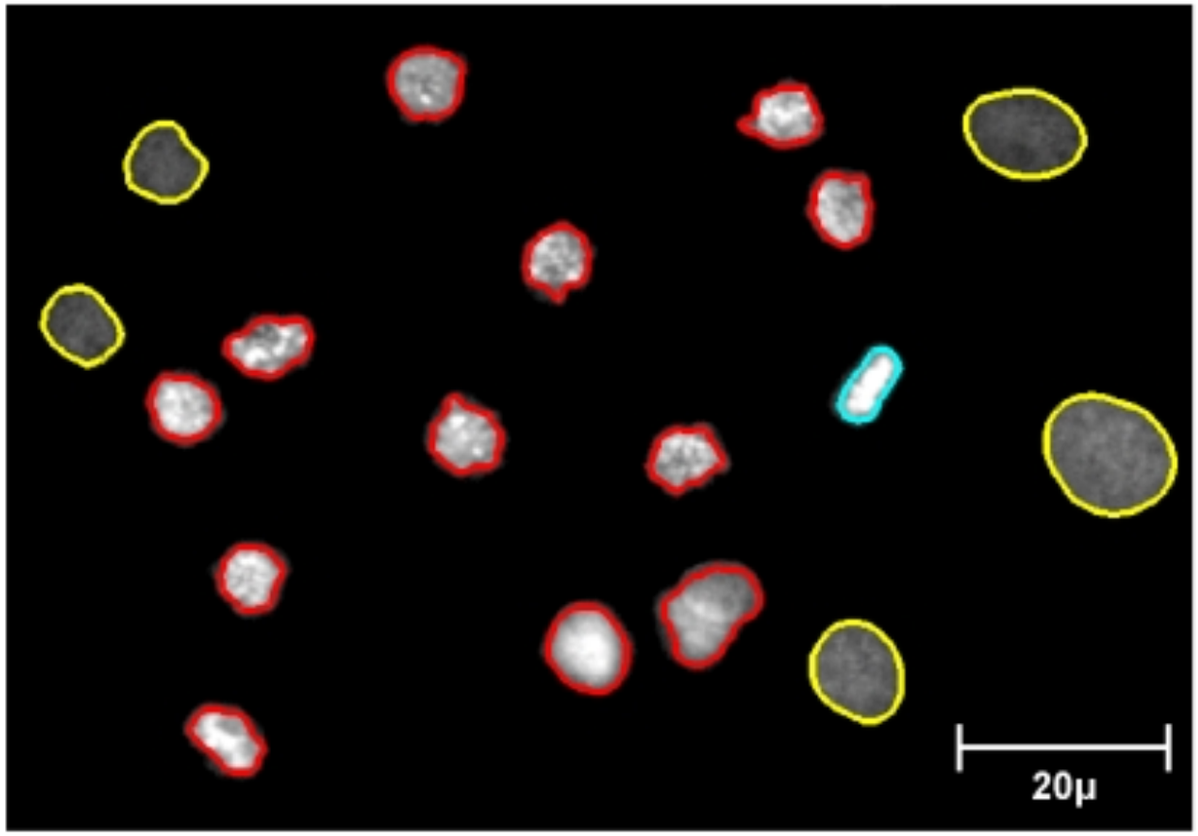
project in progress

Eric Lander, Broad Institute

Piyush Gupta, postdoc

Regulators of cell division





Ray Jones

Martha Vokes

BROAD INSTITUTE IMAGING PLATFORM

Tsui ... Carpenter ... Mitchison, et al. PLoS ONE, 2009

Tim Mitchison, Harvard Med.

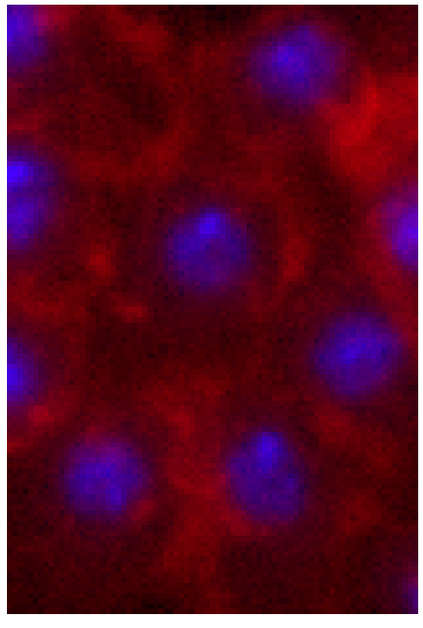
Tiao Xie

Melody Tsui

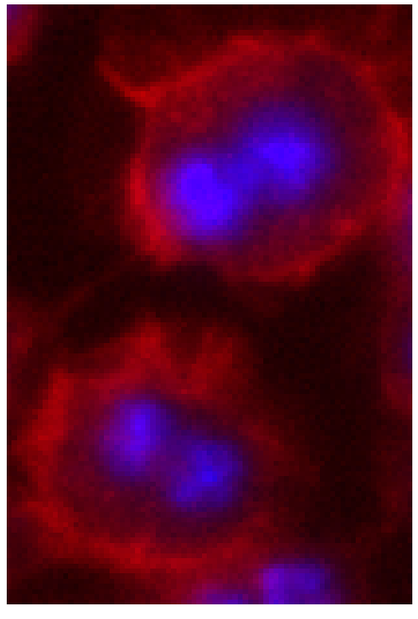
Regulators of cell division

DNA

Actin



Normal:
one nucleus
per cell



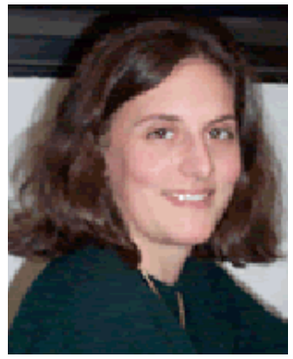
Abnormal:
two nuclei
per cell

Ray Jones

Martha Vokes

BROAD INSTITUTE IMAGING PLATFORM

Castoreno... Carpenter ... Eggert, Nature Chem Bio, 2010



Riki Eggert, Harvard Med.

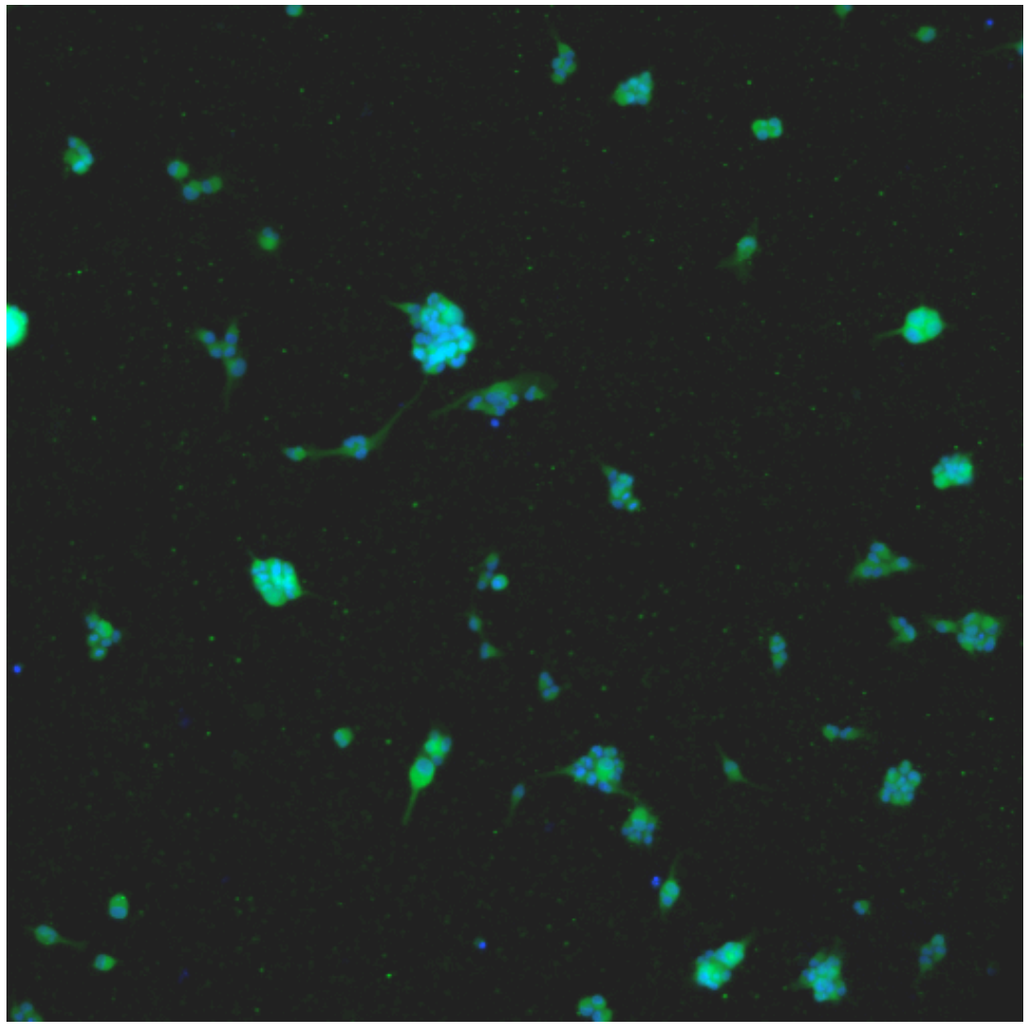
Adam Castoreno

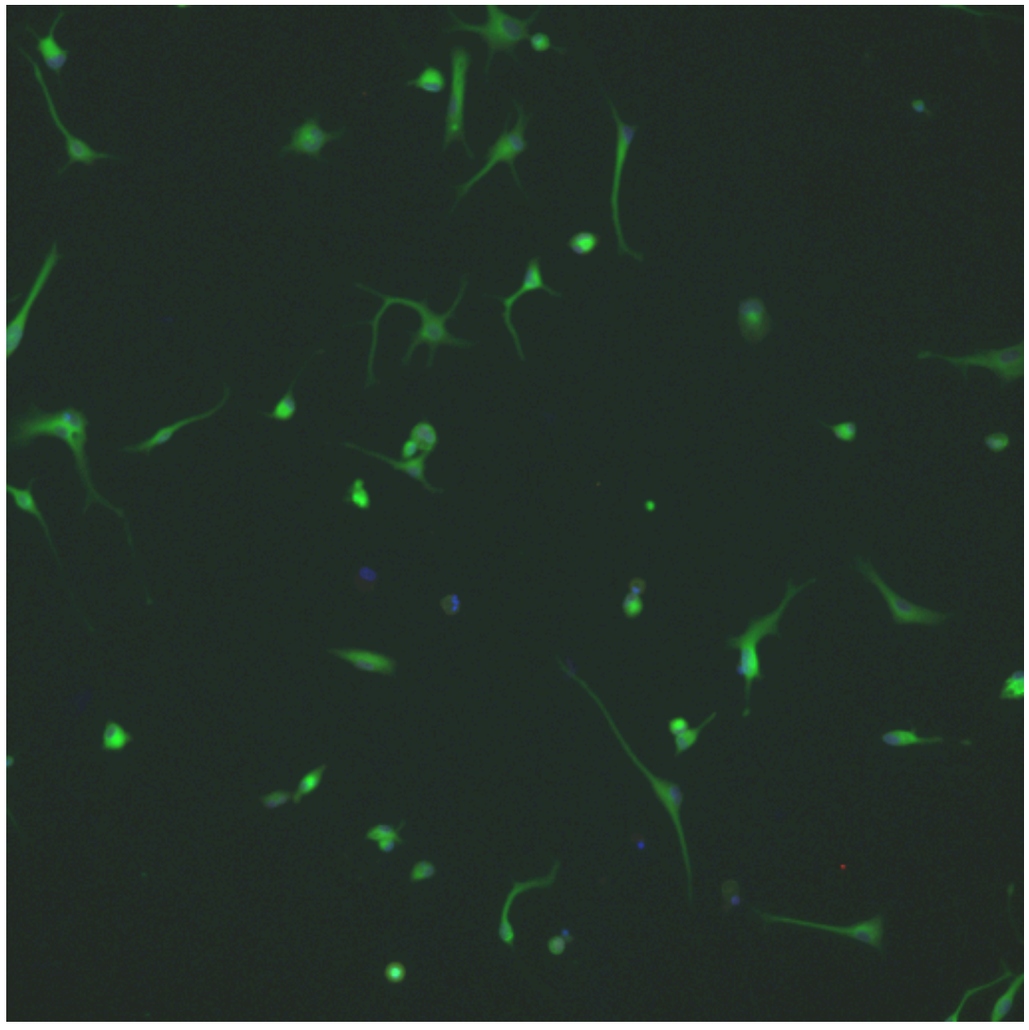
RNAi screen: glioblastoma proliferation & differentiation

Neurosphere phenotype

Flat, elongated phenotype

DNA / Tubulin





Martha Vokes

Mark Bray

BROAD INSTITUTE IMAGING PLATFORM

project in progress

David Sabatini, Whitehead Institute

Yakov Chud-novsky, postdoc

William Hahn, Broad Institute

Milan Chheda, postdoc

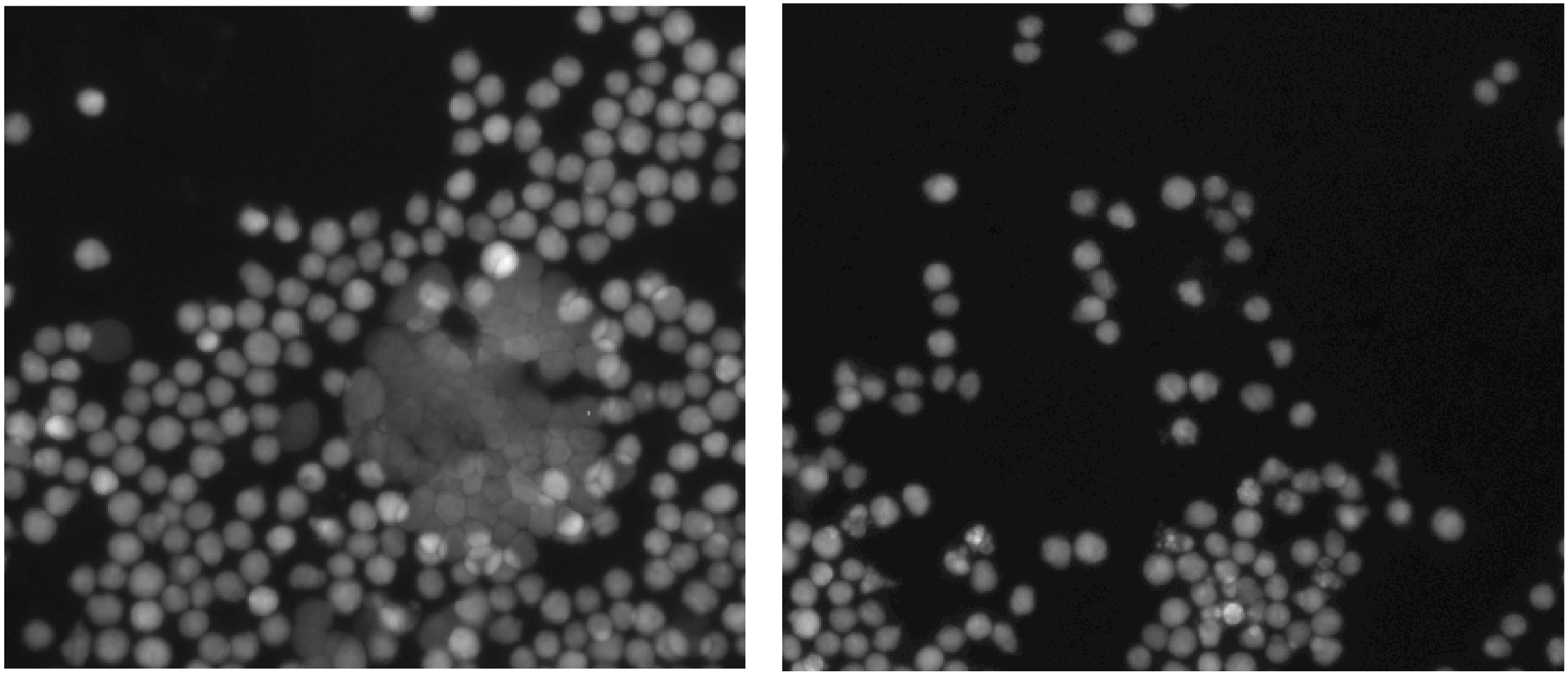
David Root, Broad Inst.

Leukemic & hematopoetic stem cells

Cobblestones

Differentiated hematopoietic cells

GFP



David Logan

BROAD INSTITUTE IMAGING PLATFORM

project in progress

Gary Gilliland, Brigham & Women's Hospital

David Scadden, Mass. General Hospital

Stuart Schreiber, Broad Institute

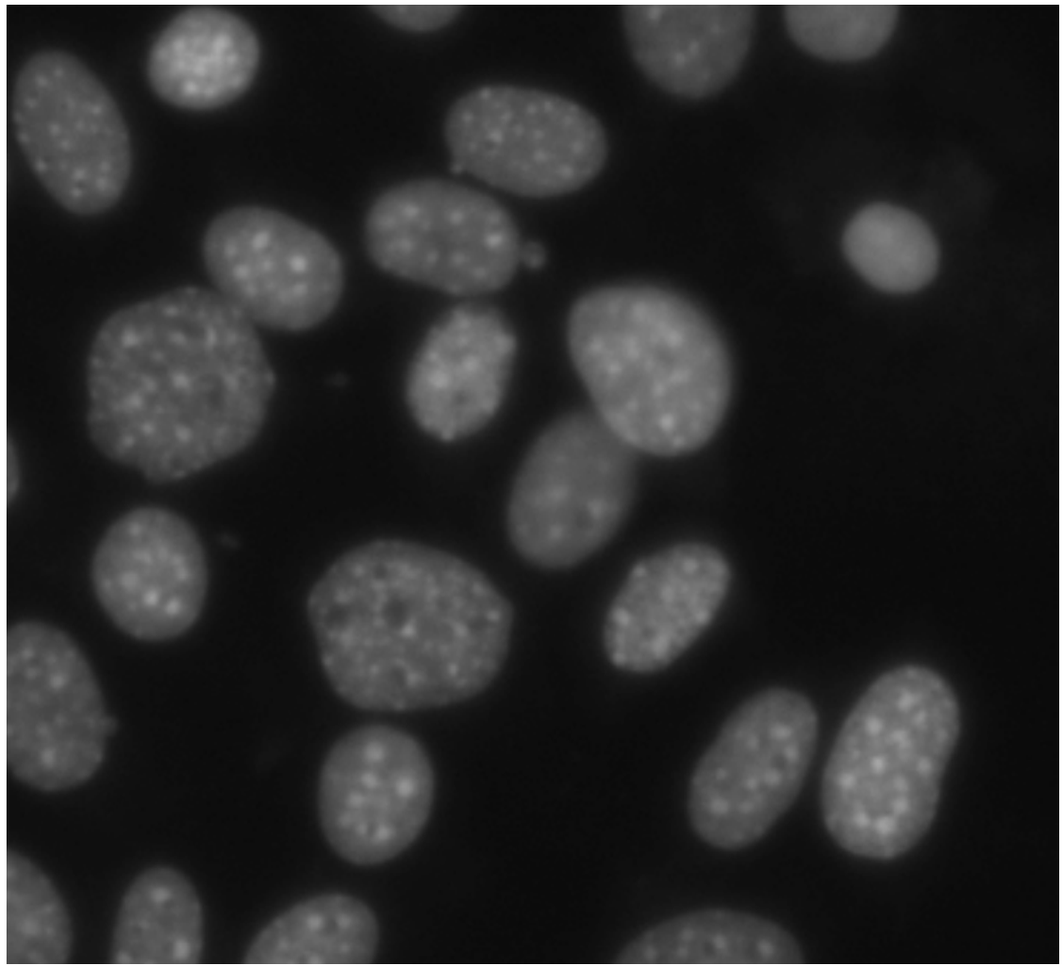
postdocs and students: Alison Stewart, Kimberly Hartwell, Peter Miller

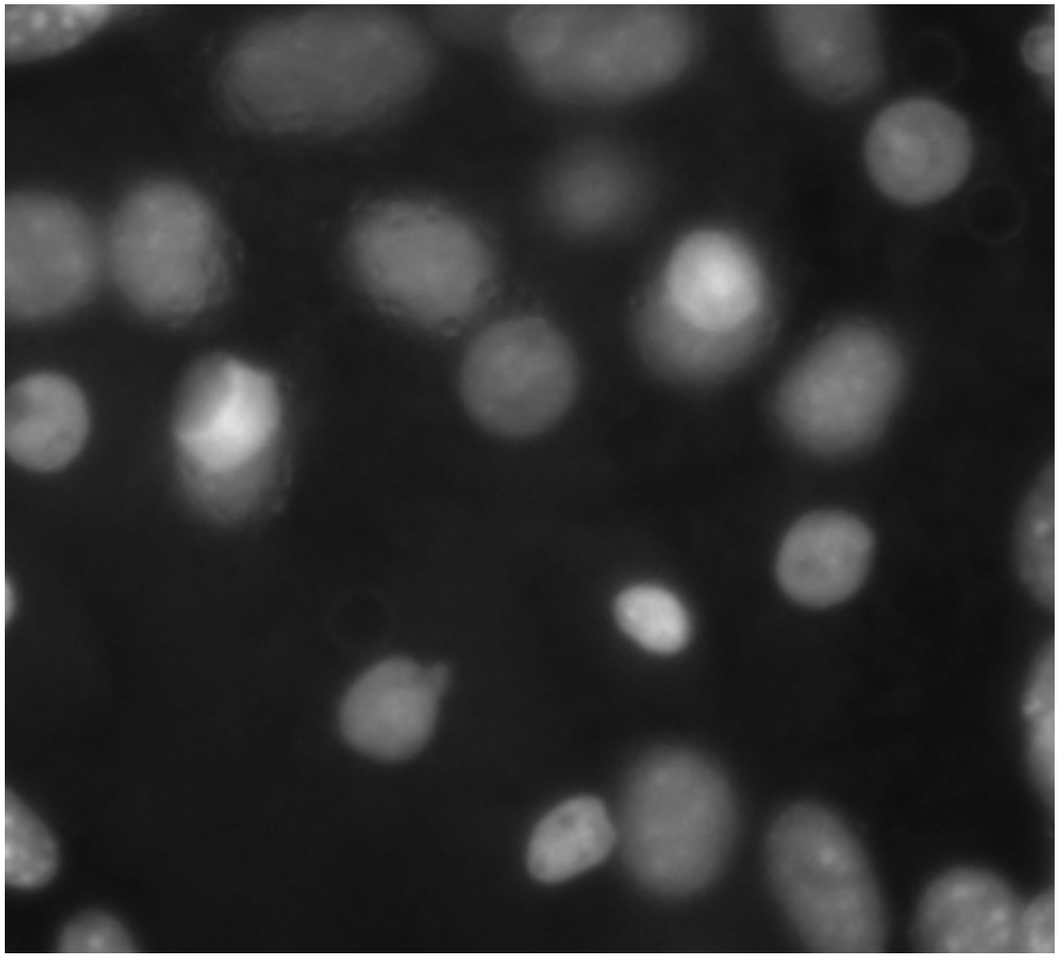
Hepatocyte proliferation

Control

Hepatocyte-enriched

DNA





Z’ factor for doubled hepatocyte count: 0.29

David Logan

BROAD INSTITUTE IMAGING PLATFORM

project in progress

Sangeeta Bhatia, MIT

Meghan Shan, student

Automatically extracting image-based phenotypes

Wild-type cells

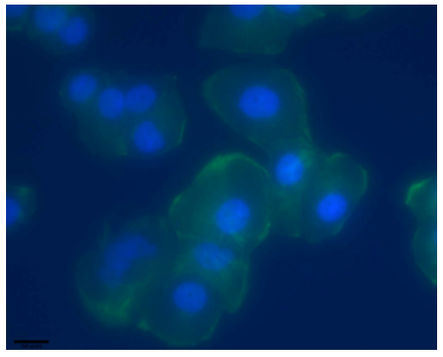


Image analysis

CellProfiler™ cell image analysis software

Detectable phenotypic difference?

Mutant cells (e.g., from GWAS)

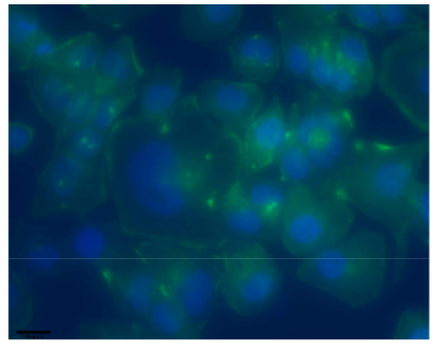
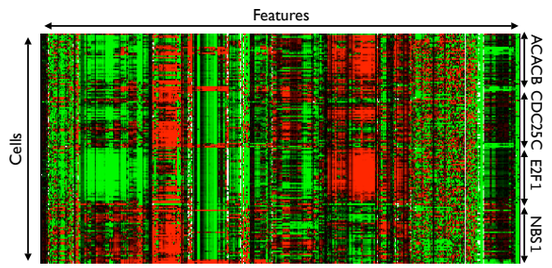
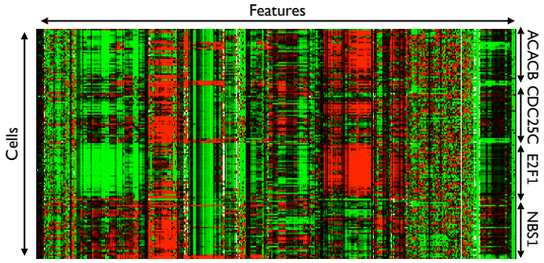


Image analysis

CellProfiler™ cell image analysis software





Identify mutant phenotype from image features
even if "invisible" to the human eye

Screen for chemicals that can revert mutant phenotype -> wild type

Ray Jones

Vebjorn Ljosa

Kate Madden

BROAD INSTITUTE IMAGING PLATFORM

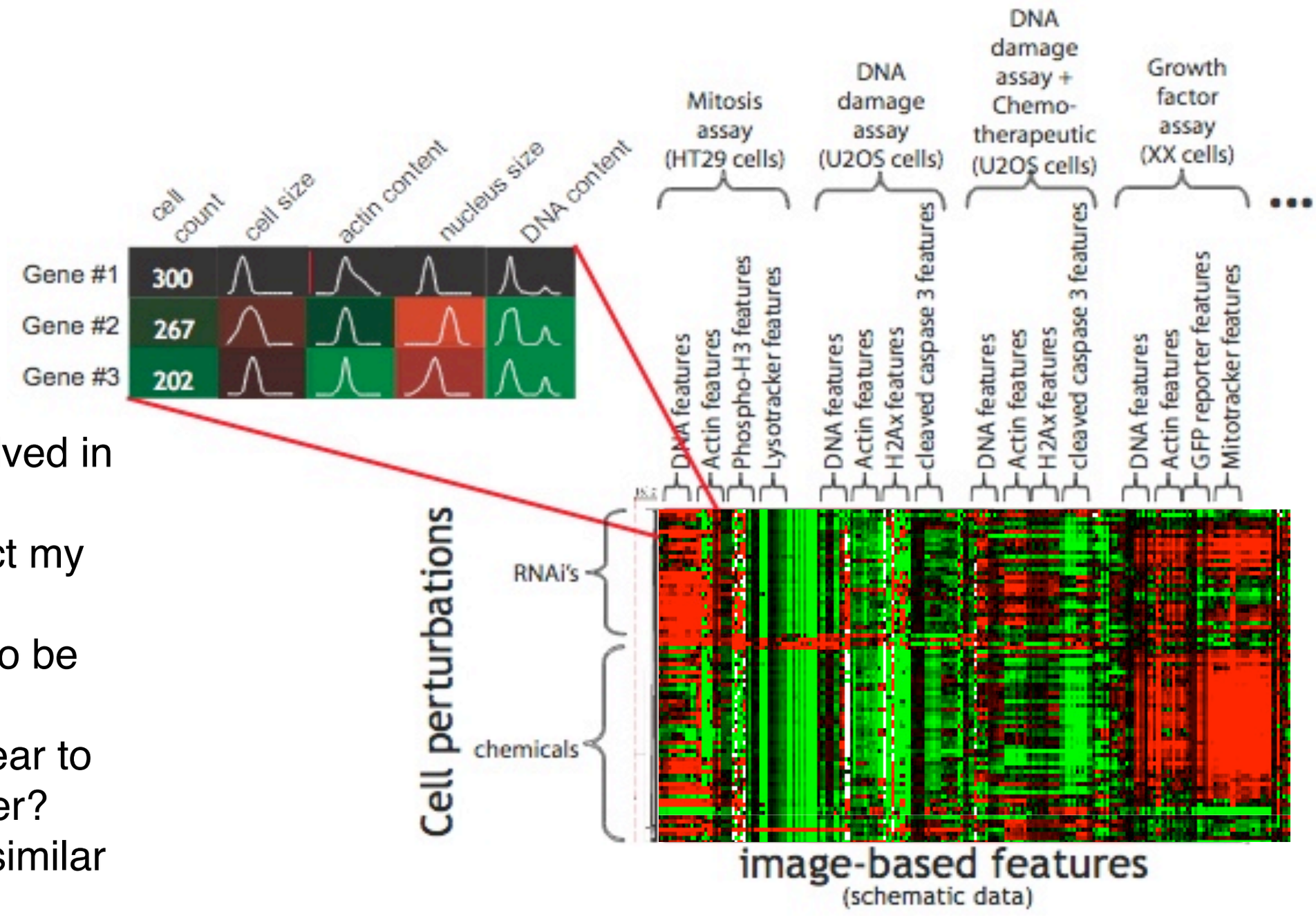
Measure everything…ask questions later.

~500 features per cell: size, shape, staining intensity, texture (smoothness), etc.

Why?

(a) Several features may be necessary to score the phenotype
(b) Virtual secondary screens can help characterize hits
(c) Later re-screening for new phenotypes
(d) The measurements required to score the phenotype of interest may not be known a priori
(e) The full spectrum of cellular responses to each treatment (even those not visible by eye) may be useful for data mining/machine learning/clustering…systems biology

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

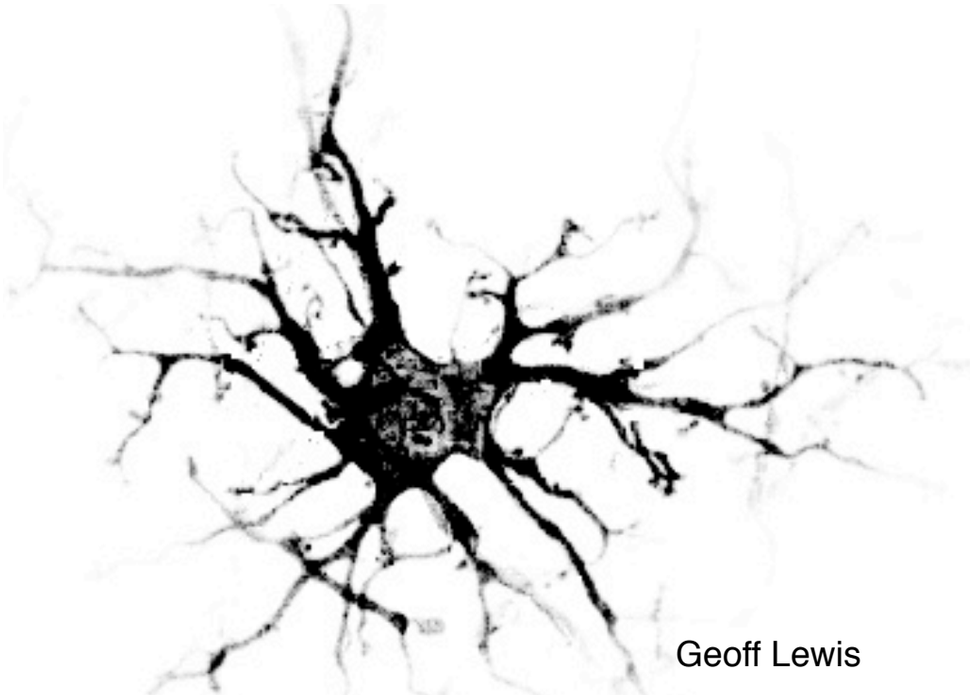


- What genes are involved in my phenotype?
- What chemicals affect my phenotype?
- What genes appear to be similar to each other?
- What chemicals appear to be similar to each other?
- What genes appear similar to a chemical?
- What phenotypes are coordinated?

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

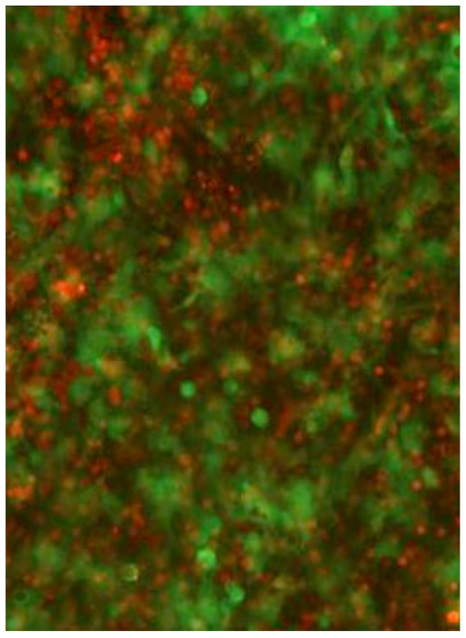
In progress

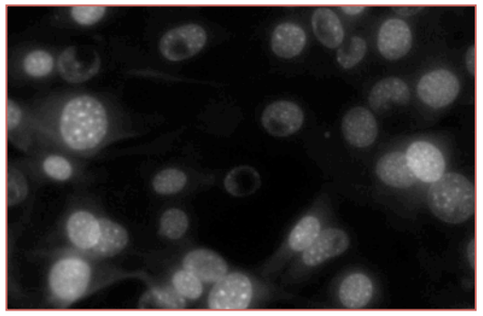
Neurons...

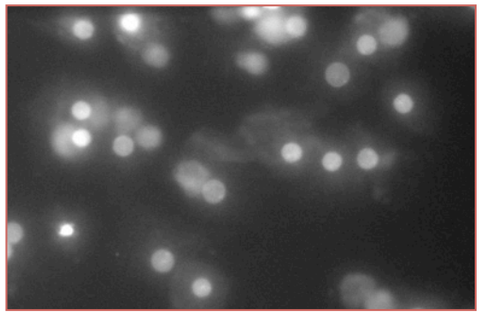


Geoff Lewis

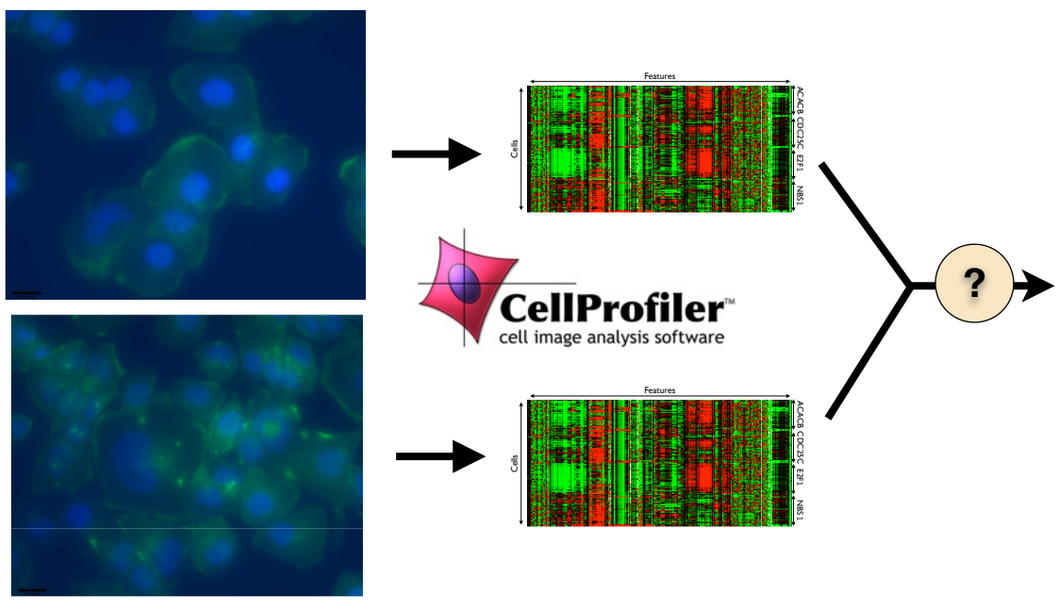
Co-cultured cell types...



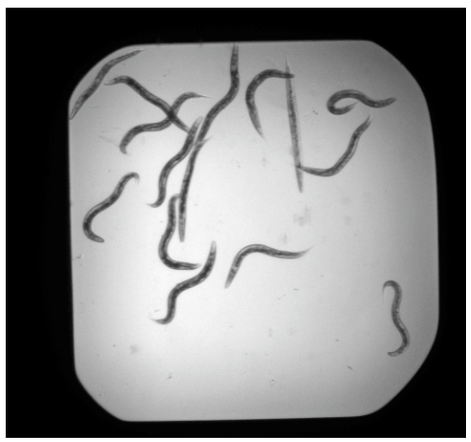


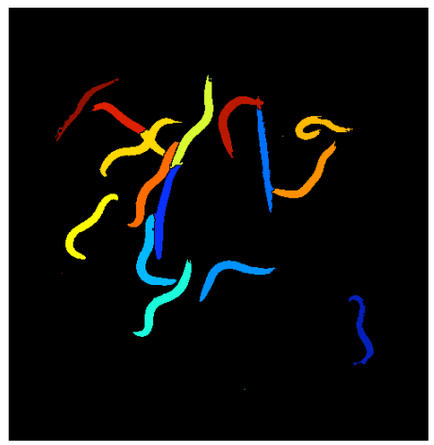


Unexpected phenotypes...



Organisms...



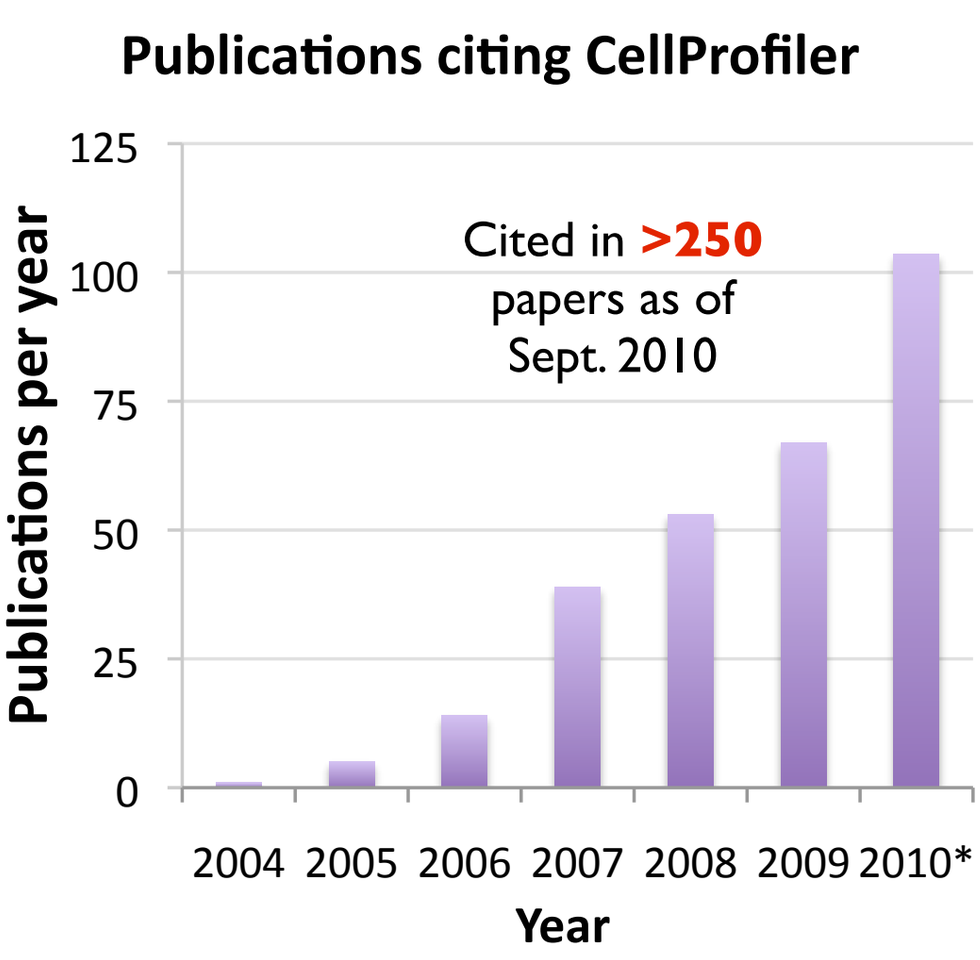


Time-lapse, 3D...

BROAD INSTITUTE IMAGING PLATFORM

The CellProfiler project

free, at www.CellProfiler.org



Selected high-throughput screens using CellProfiler	
Root lab, Cell, 2006	Screen for cell cycle regulators
Alon lab, Nature Methods, 2006	High-throughput analysis of protein dynamics
Neefjes lab, Nature, 2007	Screen for levels of Salmonella typhimurium infection
Raff lab, PLoS Biology, 2008	Screen for centriole duplication and mitotic PCM recruitment
Carpenter lab, PNAS 2009 & BMC Bioinformatics 2008	Screens for > 15 diverse phenotypes in human and *Drosophila* cells
Shokat lab, Cancer Cell, 2008	Screen for PI3K inhibitor resistance mutations in *S. cerevisiae*
Pelkmans lab, Nature, 2009	High-throughput infection assay
Ausubel lab, ACS Chem Bio, 2009	Screen for inhibitors of infection by *E. faecalis*

CellProfiler's is the 5th most-accessed Genome Biology paper of all time

Gratitude

BROAD INSTITUTE IMAGING PLATFORM

free, at www.cellprofiler.org :

CellProfiler™
cell image analysis software

CellProfiler™ Analyst
data exploration software

Anne E. Carpenter

Peggy Anthony
Mark Bray
Adam Fraser

Lee Kamentsky
Imtiaz Khan
Vebjørn Ljoså

David Logan
Kate Madden
Carolina Wählby

Contact:
anne@broadinstitute.org

This work has been supported by:

- NIH NIGMS R01 GM089652-01
- 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 collaborator at MIT's Computer Science/ Artificial Intelligence Laboratory (CSAIL)