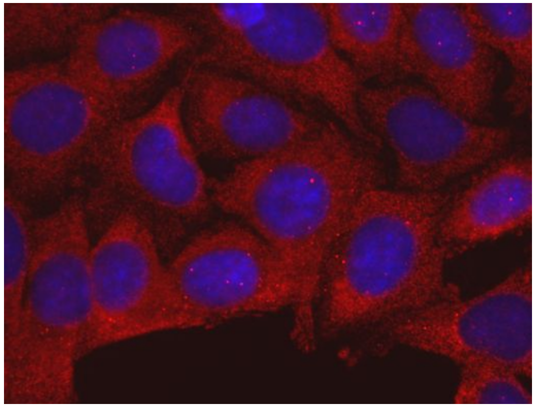
CellProfiler™
cell image analysis software

Rapid, high-content genome-wide assays using cell microarrays

Anne E. Carpenter, Ph.D.
David M. Sabatini Lab
Whitehead Institute for Biomedical Research

What are all the genes doing?



Cell size, count & morphology

Nuclear size & morphology

Nucleolar size, count & morphology

Cell cycle distribution (DNA content)

Upregulation of transcription

Apoptosis

Amount & distribution of signaling lipids

Upregulation of translation

Mitochondrial size & morphology

Localization of signaling proteins

Phosphorylation of signaling proteins (S6, S6K, Akt)

Actin content/ morphology

Protein content

SYSTEMATIC GENOME-WIDE SCREENS OF GENE FUNCTION

Anne E. Carpenter and David M. Sabatini

Nature Reviews Genetics 5:11-22 (2004)

Technologies to quickly determine gene function

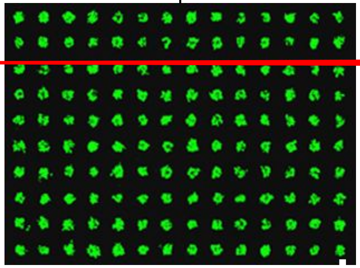
Create a set of RNA interference reagents, one for each gene

Create spots of cells, each treated with a different RNAi reagent

Stain cells for a specific phenotype

Automated image collection

"Living cell microarray"



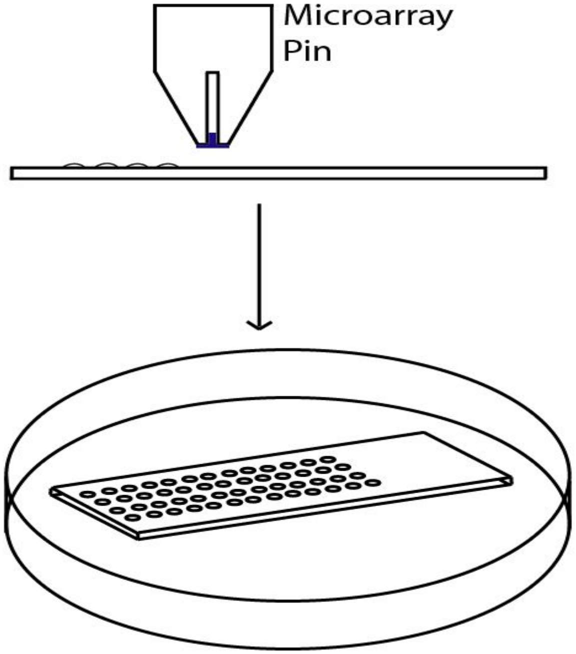
every cluster of cells has a different gene knocked-down

Automated image analysis with CellProfiler

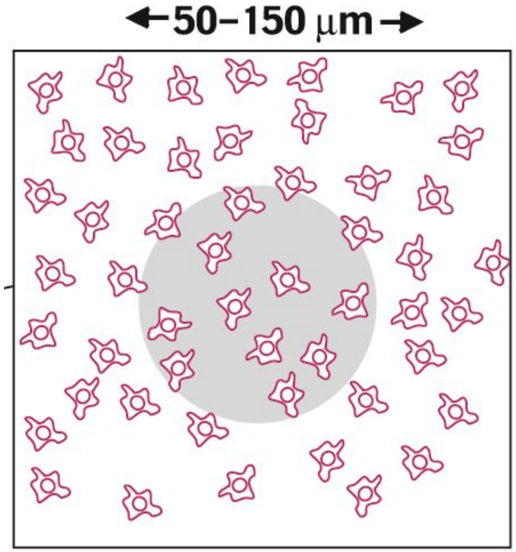
Data analysis

Determine the phenotypic effects of knocking down every gene in the genome

Living cell microarray technology

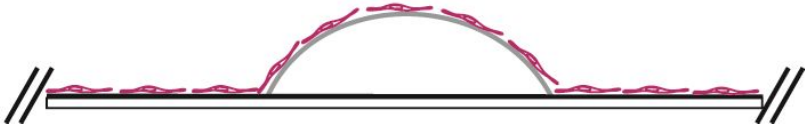


5,600 spots
per slide



top view

side view:



cDNA expression → **gain of function**

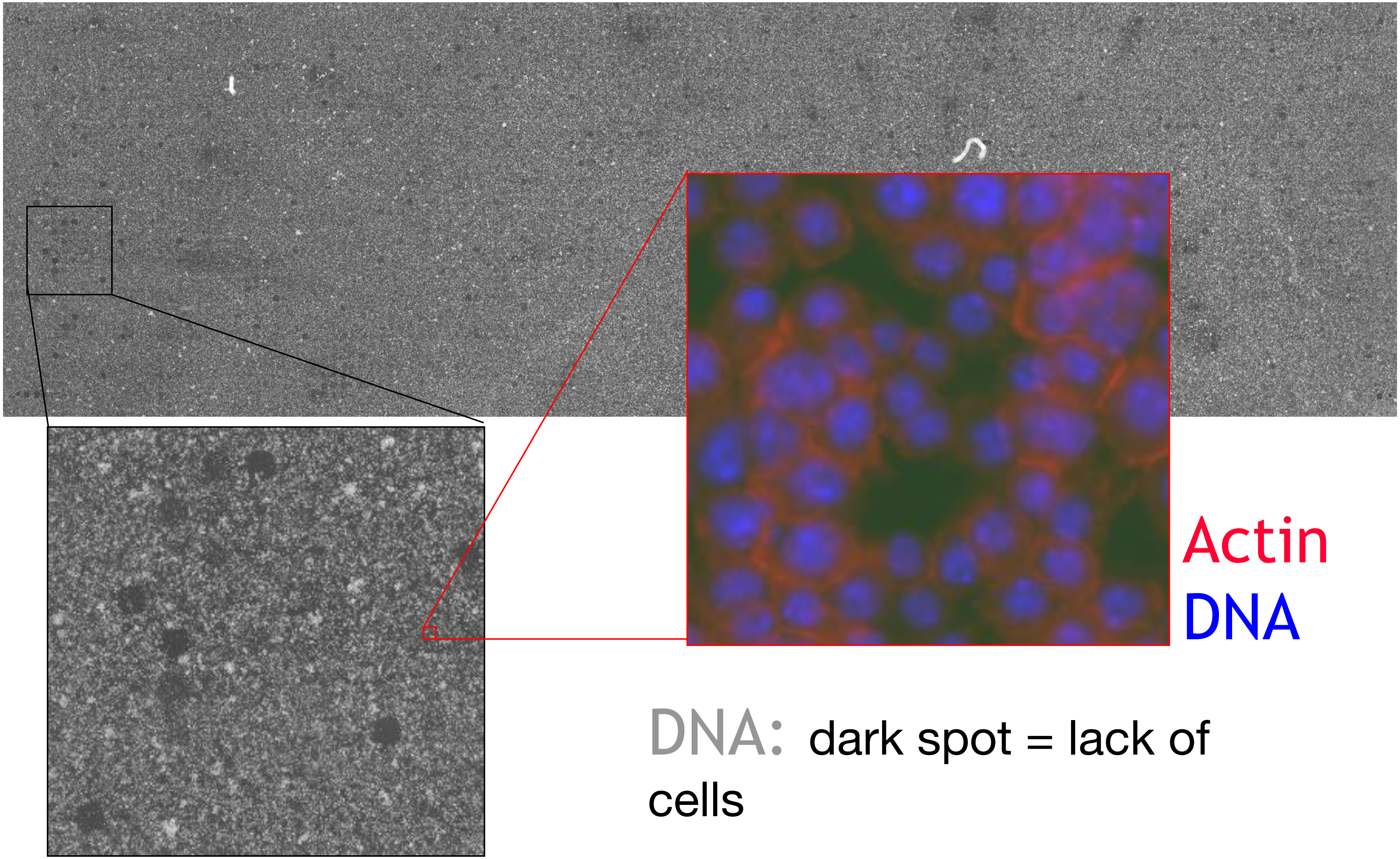
RNA interference → **loss of function**

small molecules → **chemical genetics/ drug discovery**

Fast
Cheap
Requires little reagent/ few cells
Uniform - can see subtle effects
Synthetic genetic interactions easy
High content screening/western blots

Ziauddin & Sabatini: Nature, 411:107-110 (2001)
Bailey, et al. Nature Methods (2006)

Genome-wide screens in Drosophila

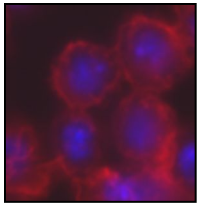


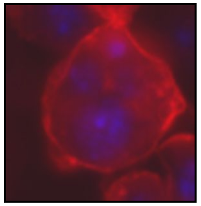
DNA: dark spot = lack of cells

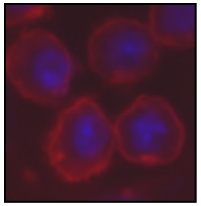
technique described in Wheeler…Sabatini, Nature Methods 2004
reviewed in Wheeler, Carpenter, Sabatini, Nature Genetics suppl., June 2005

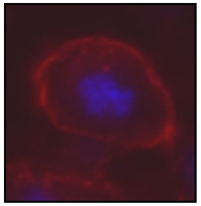
How can we measure cells automatically?

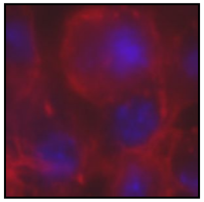
Result: hundreds of thousands of cell images

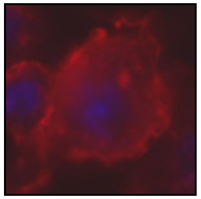

GFP

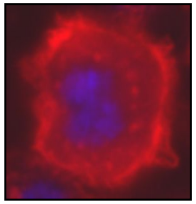

PPV

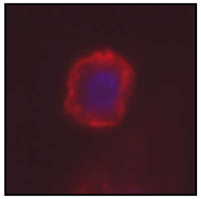

dTOR


PP2A CatSub


CG9006


String


Cyclin A


rpL12

…plus ~20,000 more images

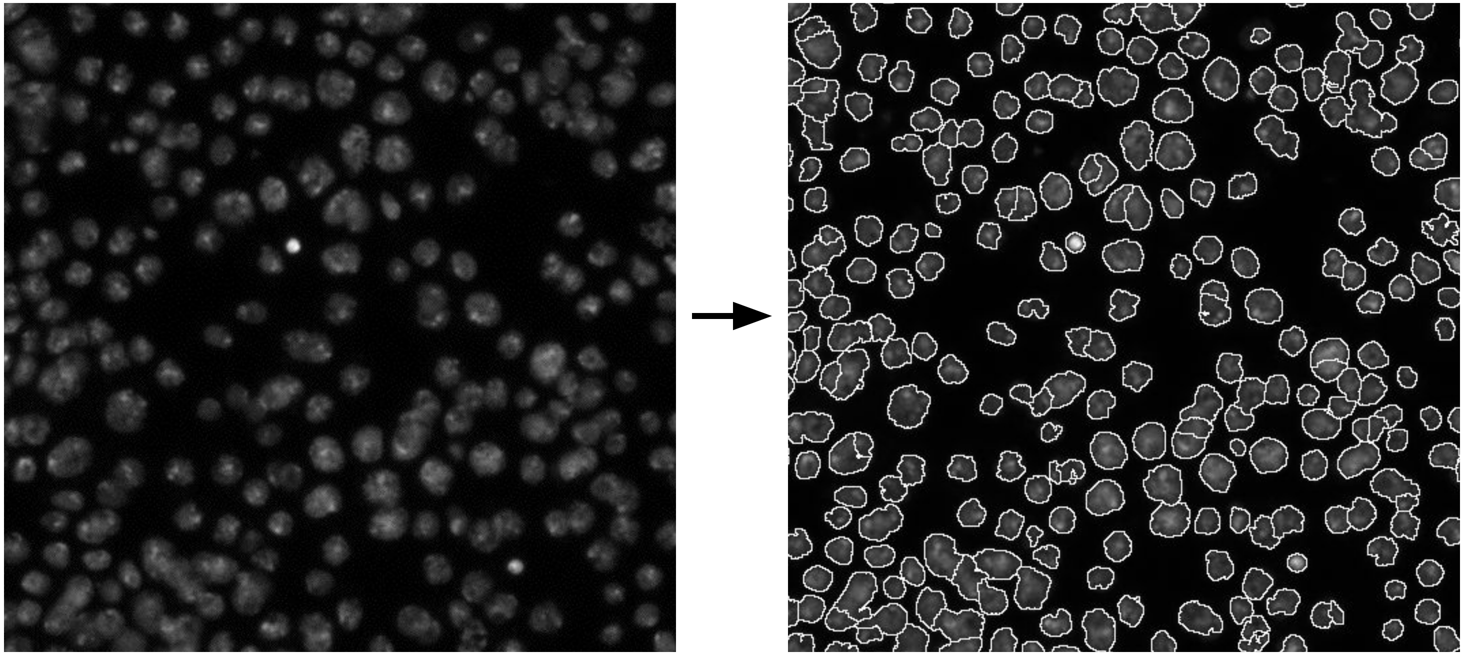
We want to know quantitatively and automatically: size, shape, intensity, texture, overlap of colors, etc. for *every* cell in *every* image.

- less tedious, less biased, quantitative

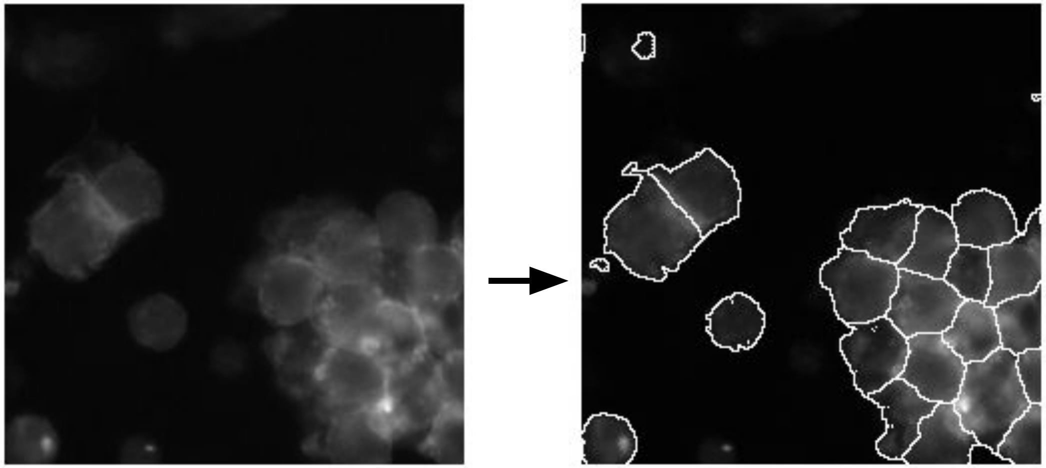
Sophisticated algorithms needed

Drosophila Kc167 cells

DNA (nuclei)



Actin (cell edges)



Jones, Carpenter & Golland (2005) ICCV Workshop on Computer Vision for Biomedical Image Applications

The CellProfiler project

Free!

CellProfiler™
cell image analysis software

Runs on Mac/PC/Unix, plugs into Matlab, can make use of cluster computing

Image file types: tif, jpg, bmp, gif, cur, dib, hdf, ico, pbm, pcx, pgm, png, ppm, ras, stk, xwd, avi

Allows quantitative analysis of various cell phenotypes in thousands of images (high-throughput experiments, time lapse, etc.)

Usable by cell biologists without programming knowledge

Modular design allows custom image analysis modules to be added

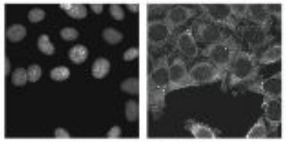
Whitehead Institute
for Biomedical Research

Anne E. Carpenter
Whitehead Institute for Biomedical Research: Laboratory of David Sabatini

Thouis R. Jones
MIT Computer Sciences/ Artificial Intelligence Laboratory: Laboratory of Polina Golland

MIT

Typical CellProfiler pipeline:



original images

Image processing modules

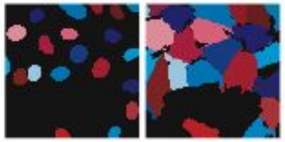
processed images

Illumination correction modules

illumination-corrected images

Object identification modules

identified objects (nuclei and cells)



Measurement modules

measurements for every cell in every image (number, location, size, shape, intensity, texture) can be analyzed by:

1. built-in CellProfiler data tools
2. exporting to spreadsheet
3. exporting to database
4. analyzing in MATLAB

Measurable cell features

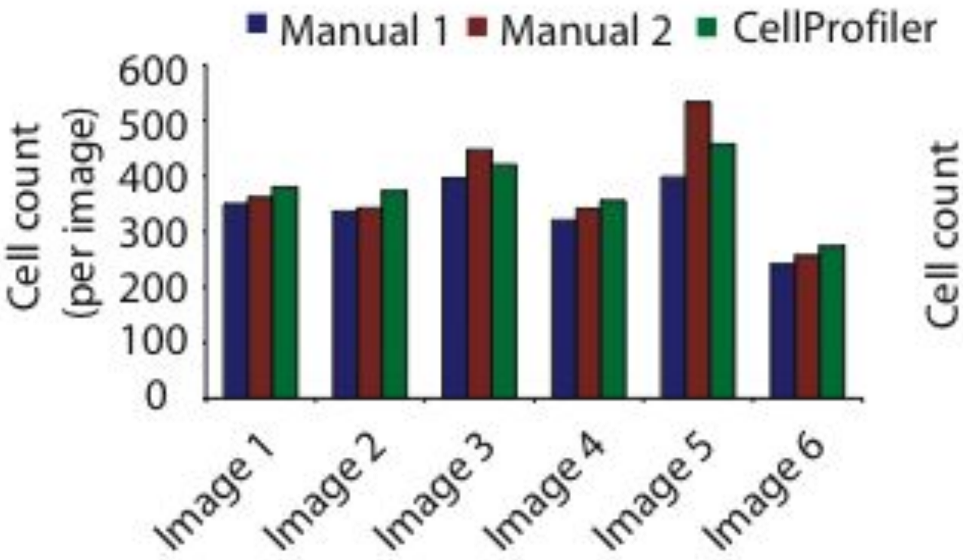
- Location: X,Y
- Cell count
- Object count within cells (e.g. speckles within nucleus)
- Neighbors
- Size
- Shape

For all colors

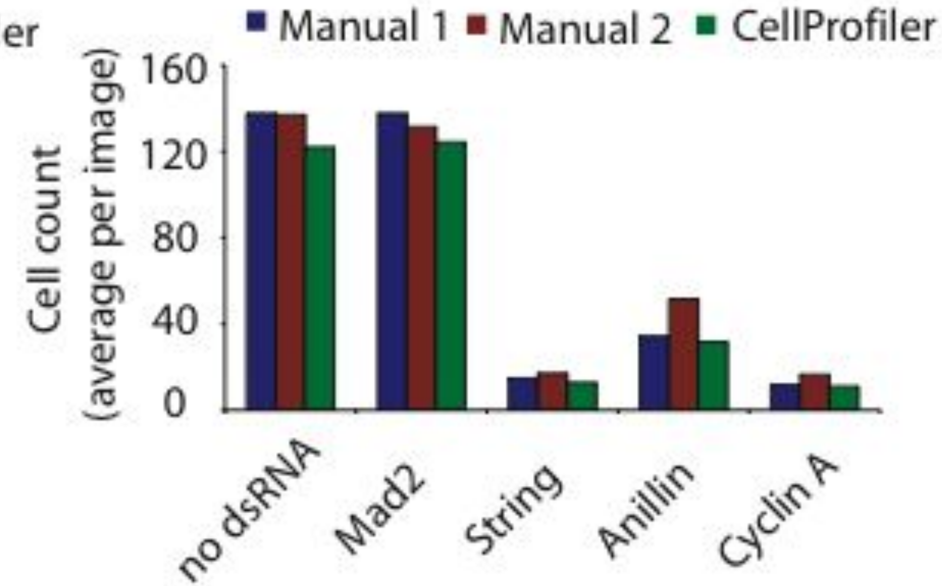
- Intensity (of entire object and of the edge of the object)
- Texture
- Correlation between different colors

Cell count validation

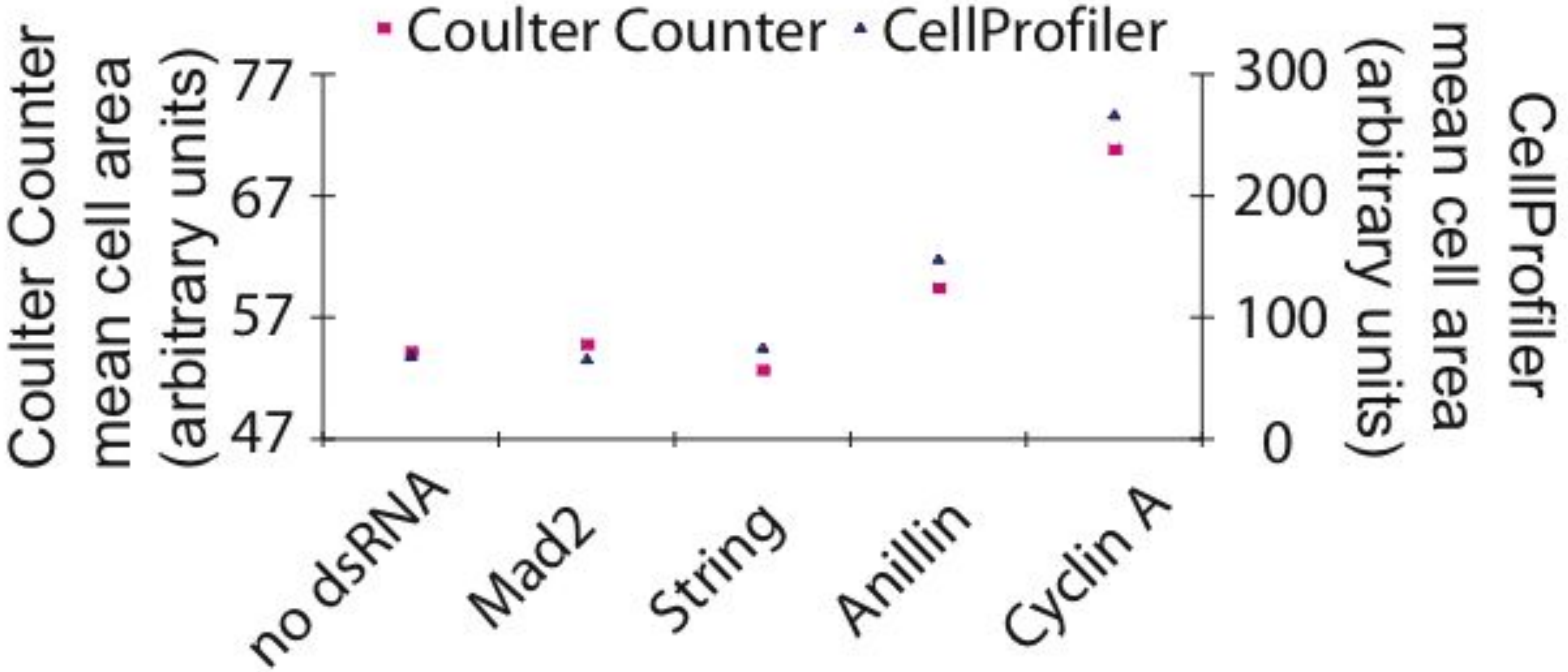
Human HT29



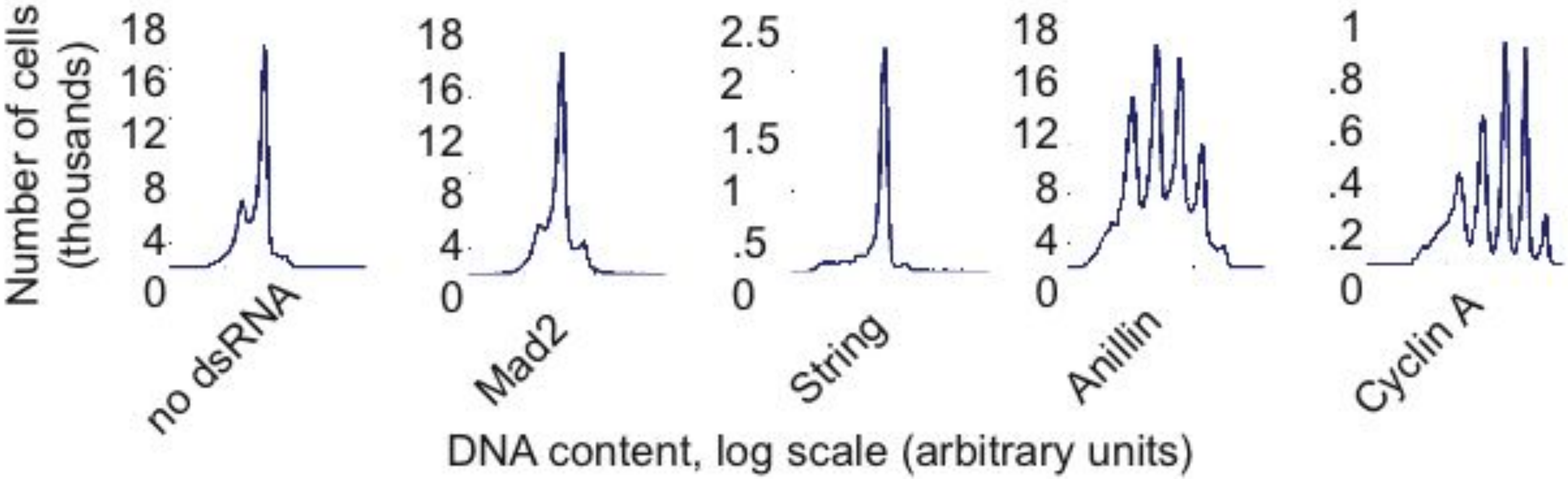
Drosophila



Cell area validation

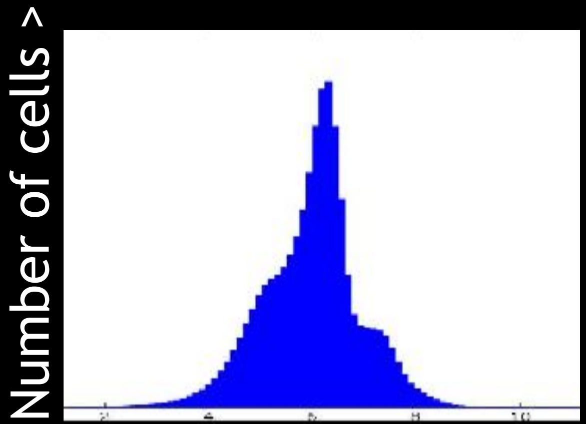


Validation for DNA content (cell cycle)-Drosophila

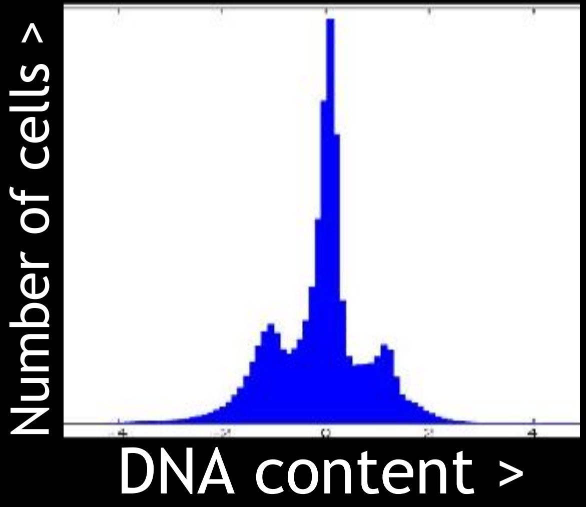


Slide scale normalization

Per-nucleus DNA content:



Normalized by local median:



Field-of-view illumination correction

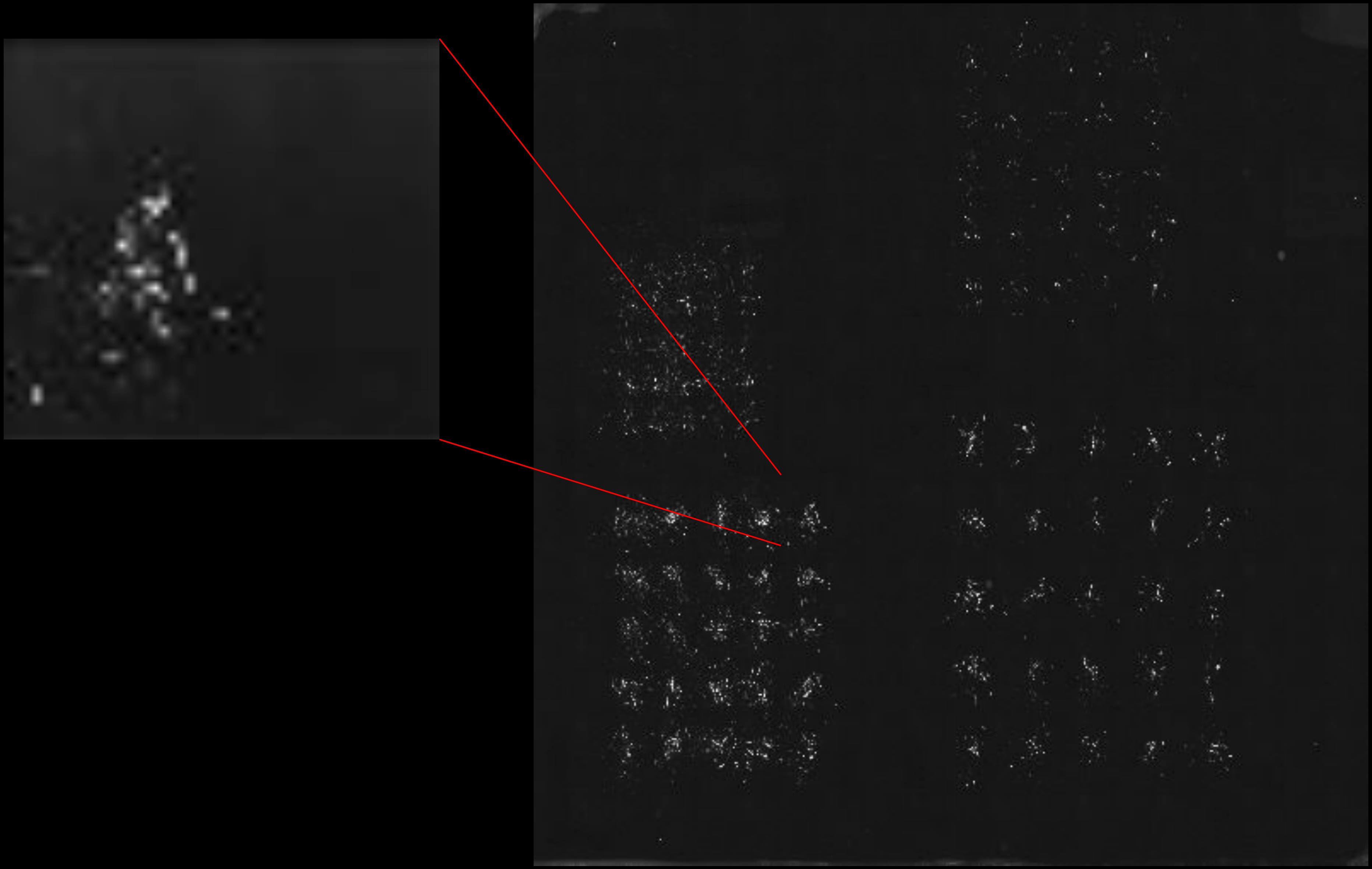
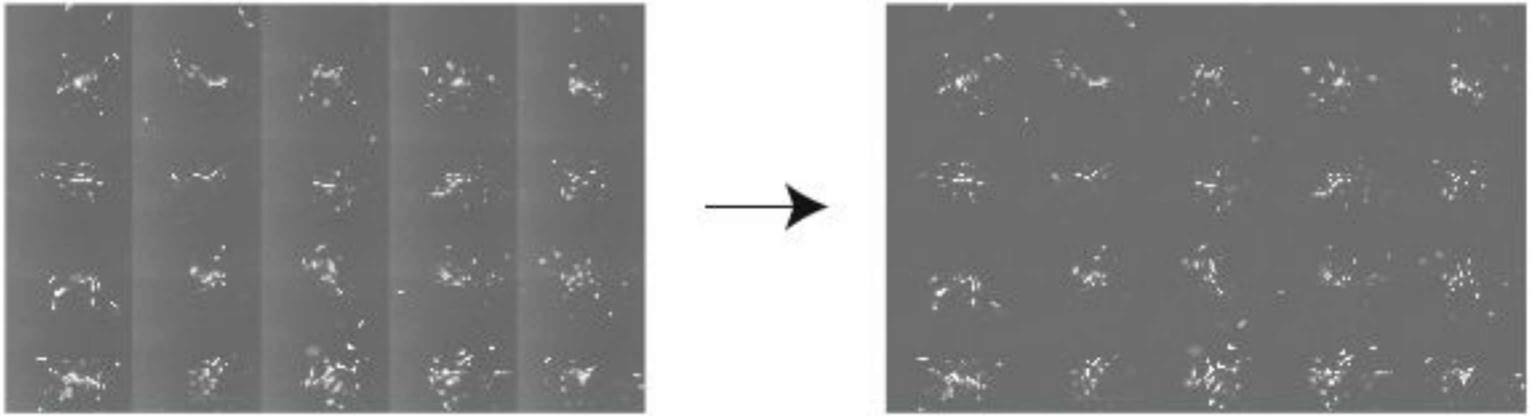
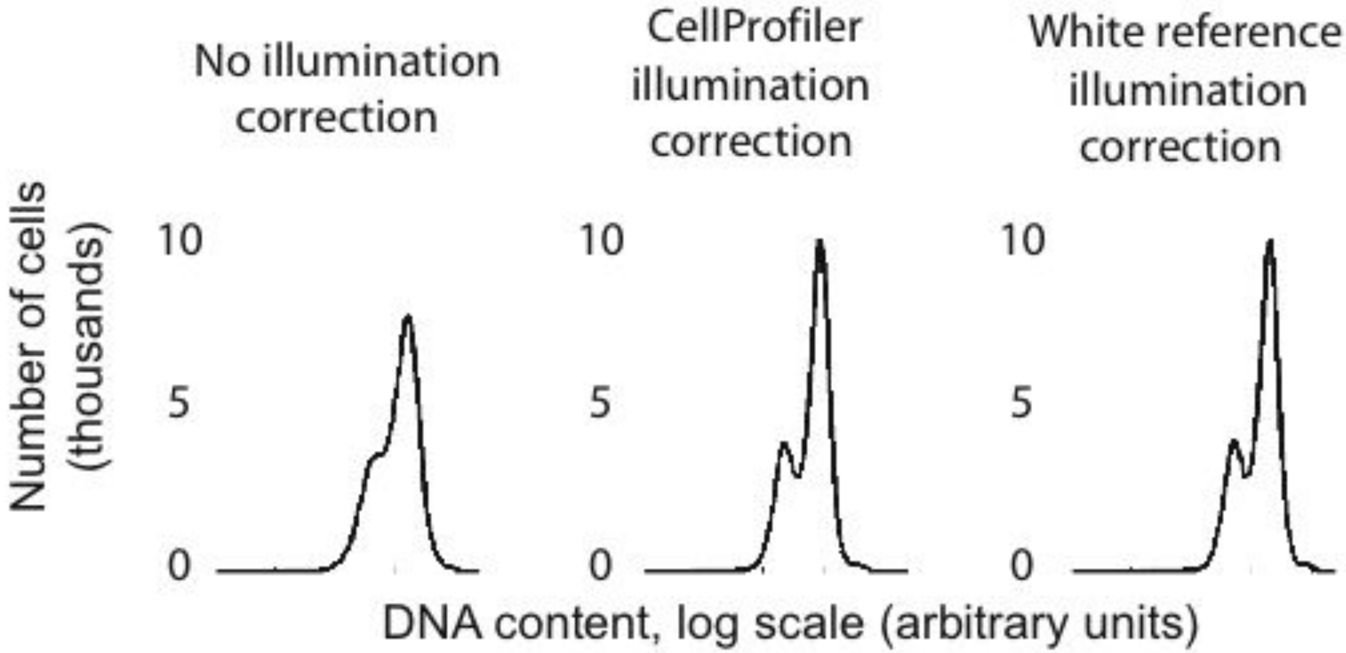


image from Steve Bailey, Sabatini lab

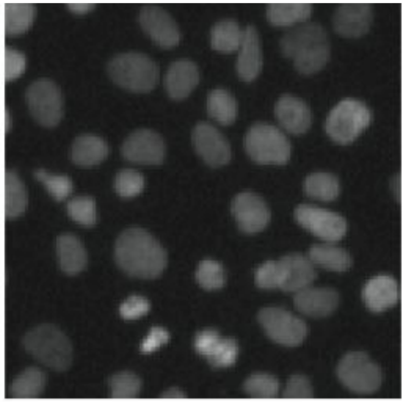
Field of view illumination correction



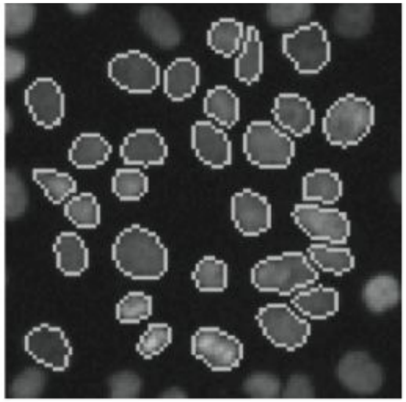


Validation for DNA content

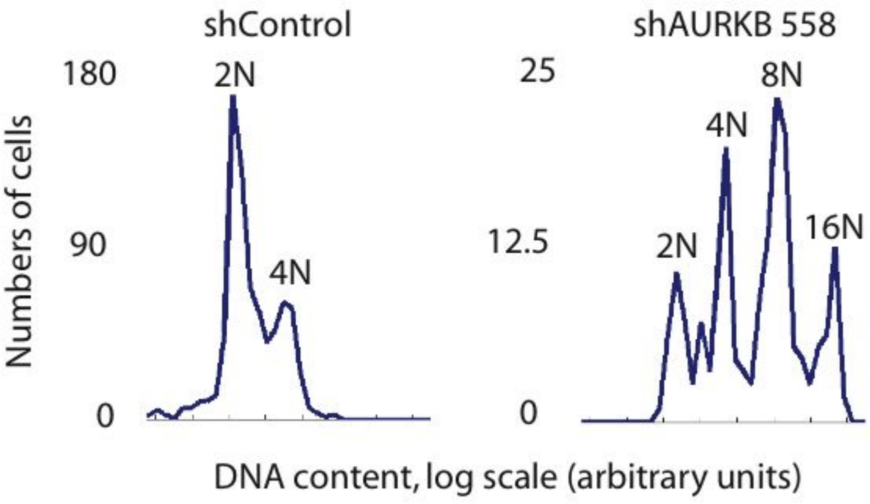
Human HT29



Nuclei image

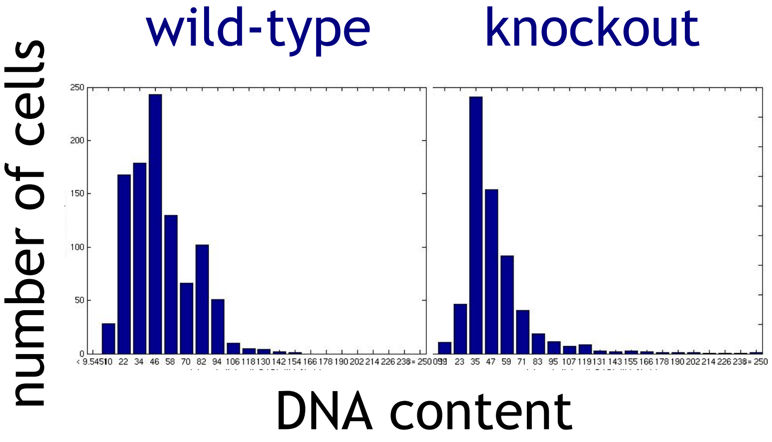


CellProfiler-outlined nuclei



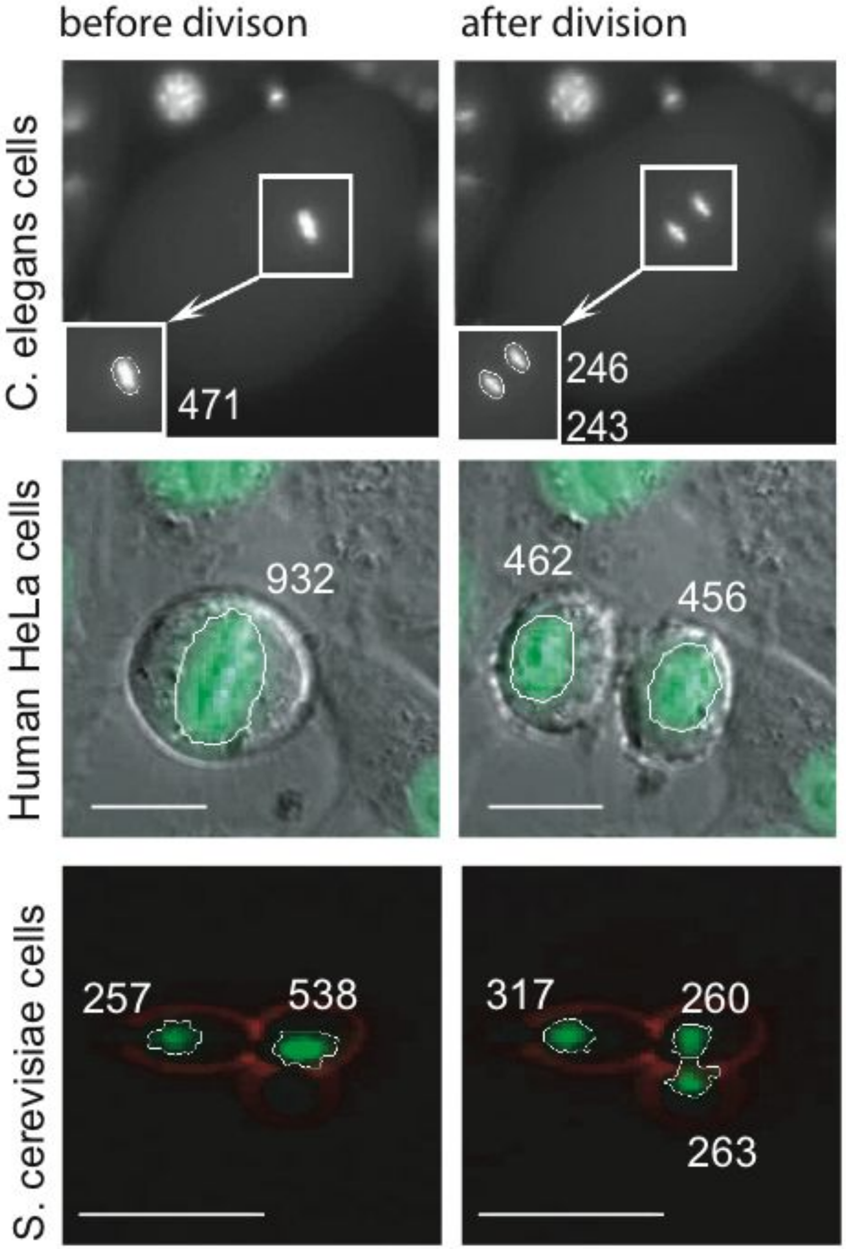
Moffat, et al, Cell, in press

mouse



images from Andrew Baltus,
Page lab, Whitehead Institute

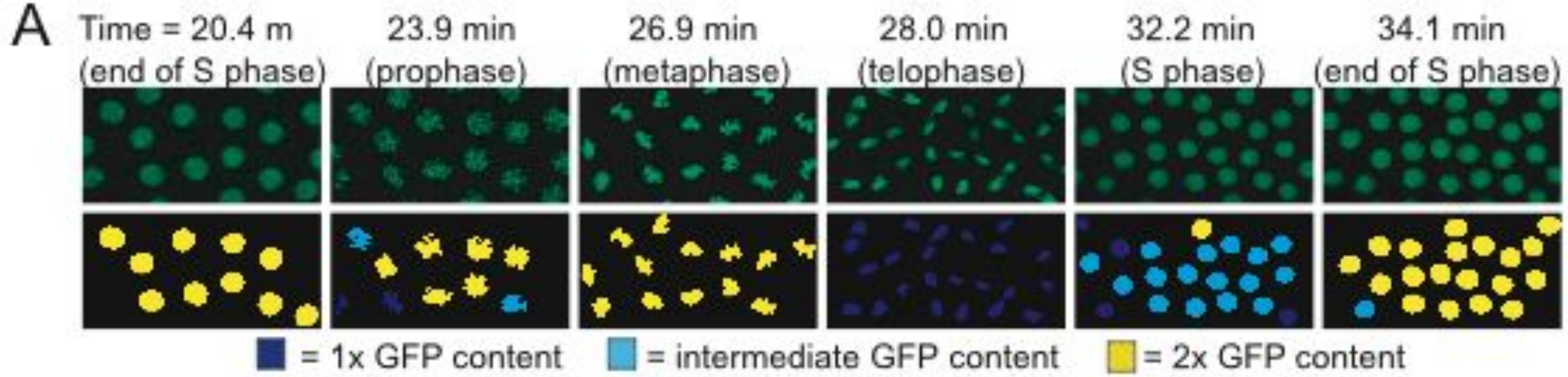
Validation for DNA content



Time lapse movies of Drosophila embryos

Goal: identify nuclei & measure morphology & GFP content

movie from Victoria Foe, Univ. Washington



Time lapse movies of Drosophila embryos

Goal: identify nuclei & measure morphology & GFP content

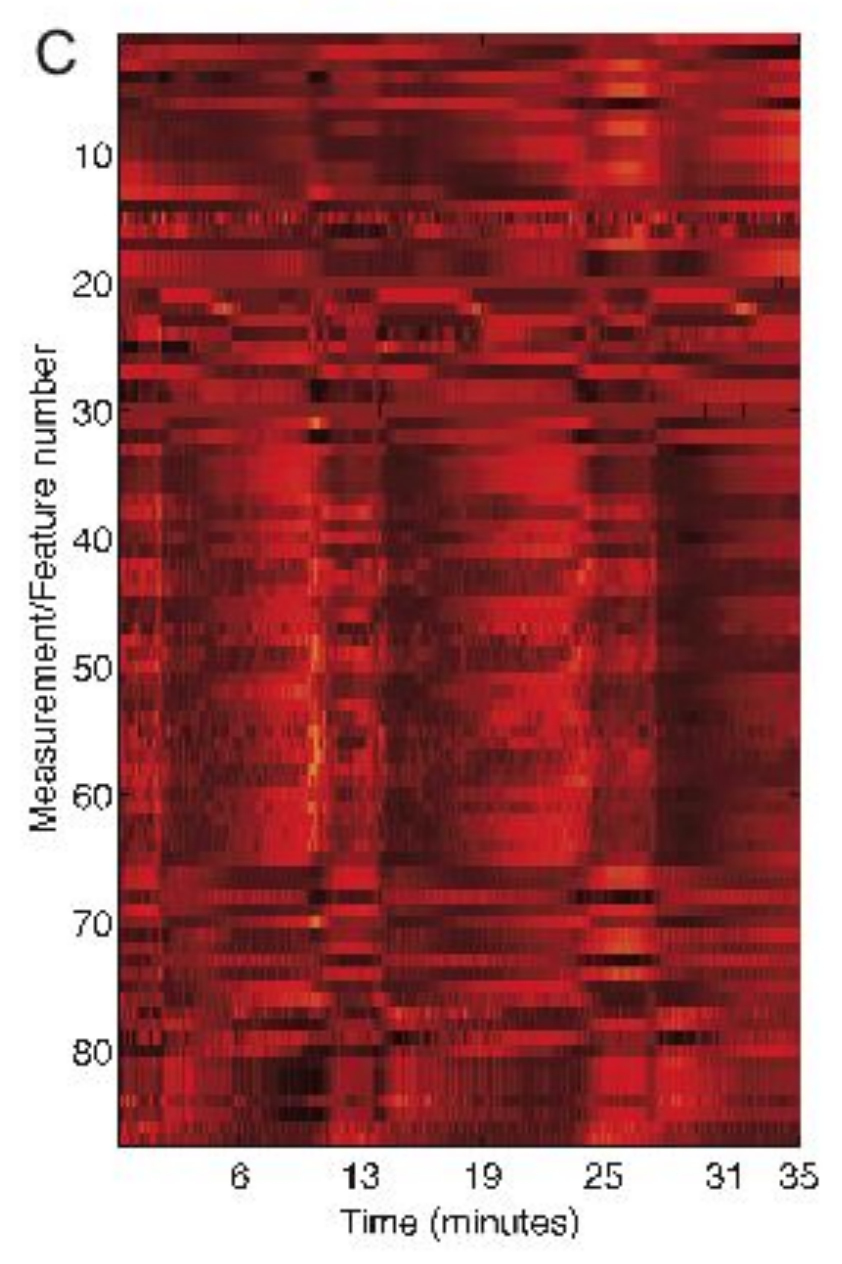
movie from Victoria Foe,
Univ. Washington

GFP
content

Area

Shape

Time lapse movies of Drosophila embryos



Antibody staining intensity - Mouse tissue

Goal: score cells as positive or negative for the red-stained Mvh

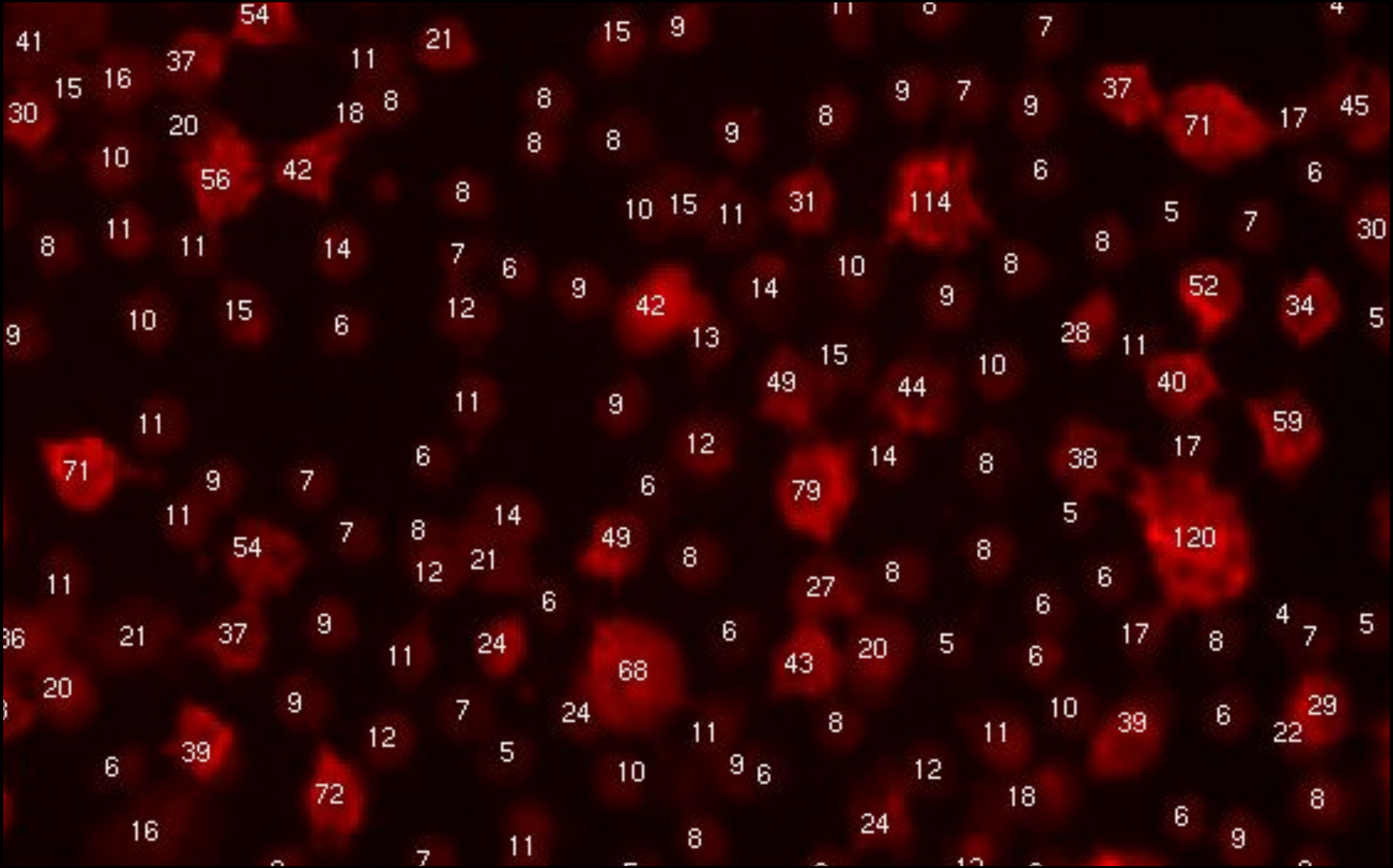


image from Andrew Baltus, Page lab, Whitehead

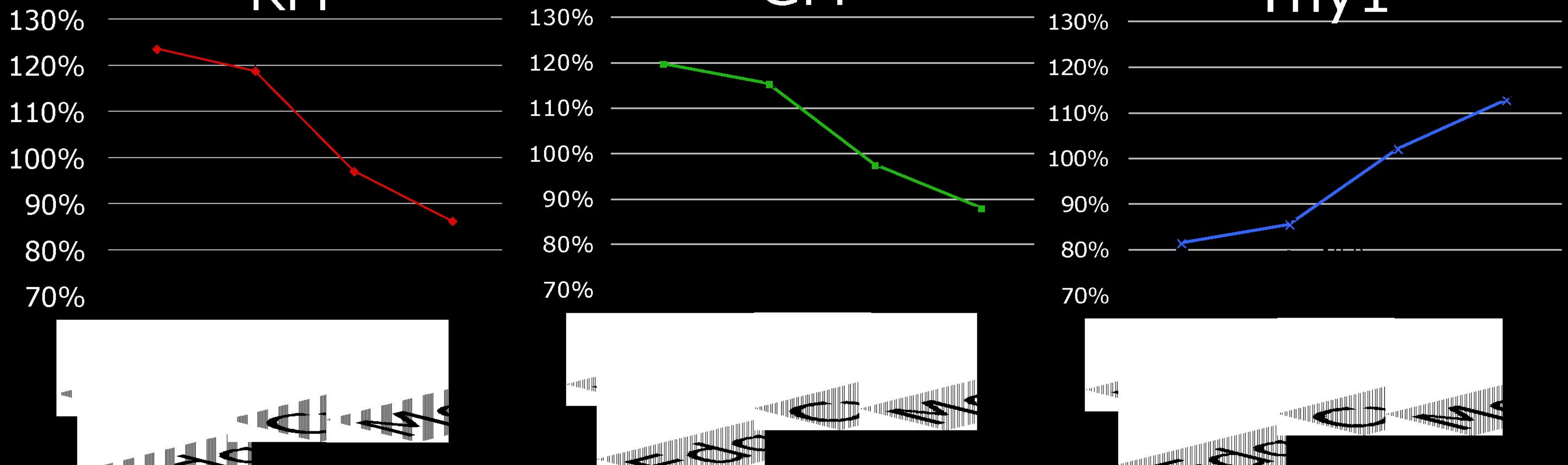
Membrane localization

Goal: quantify the localization of proteins

RFP

GFP

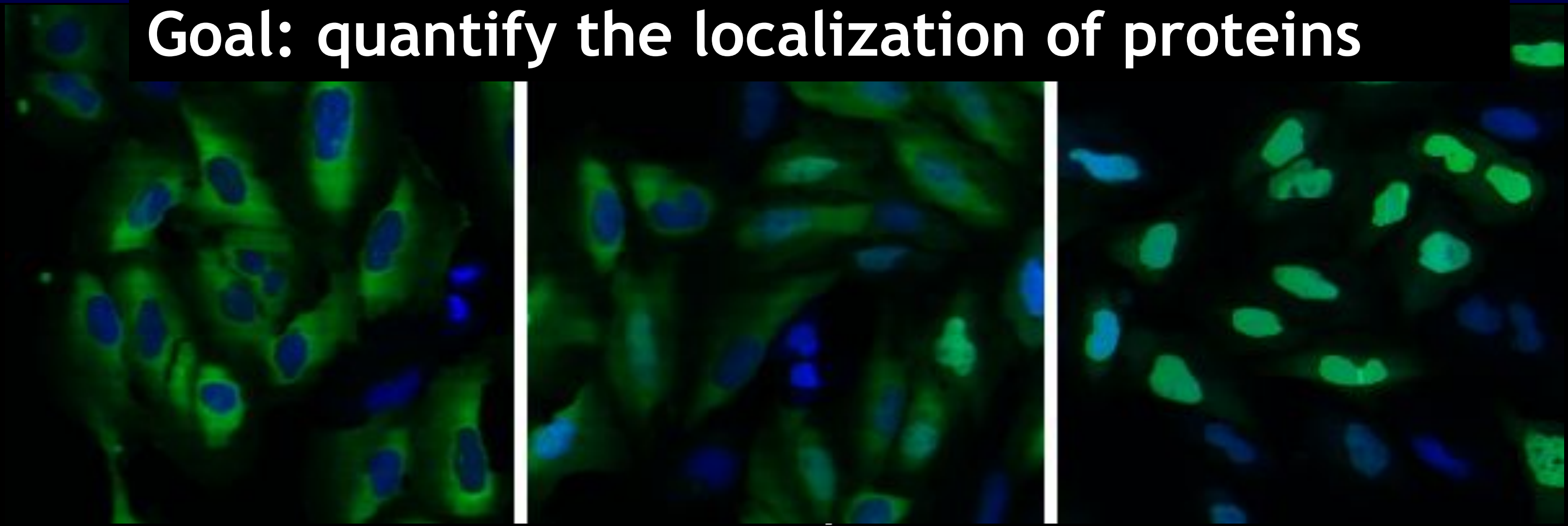
Thy1

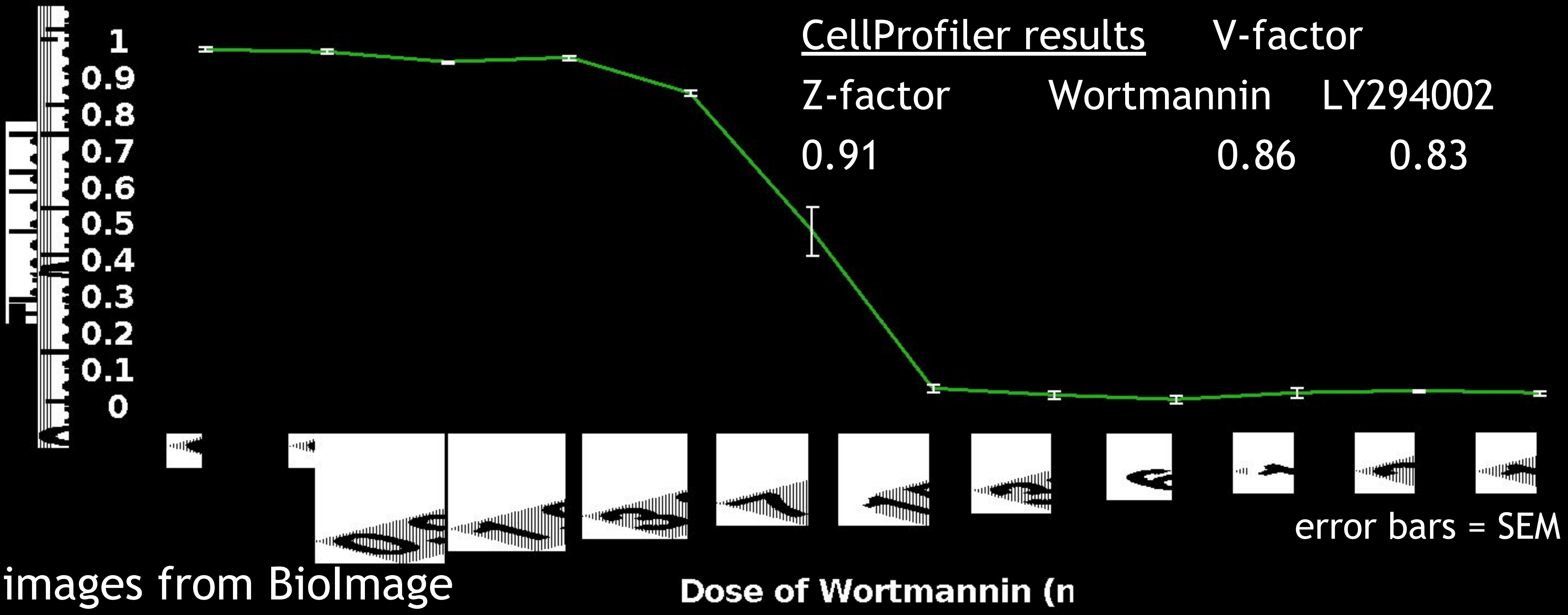


Bailey, Ali, Carpenter, Higgins, Sabatini, Nature Methods

Cytoplasm-nucleus translocation assay

Goal: quantify the localization of proteins





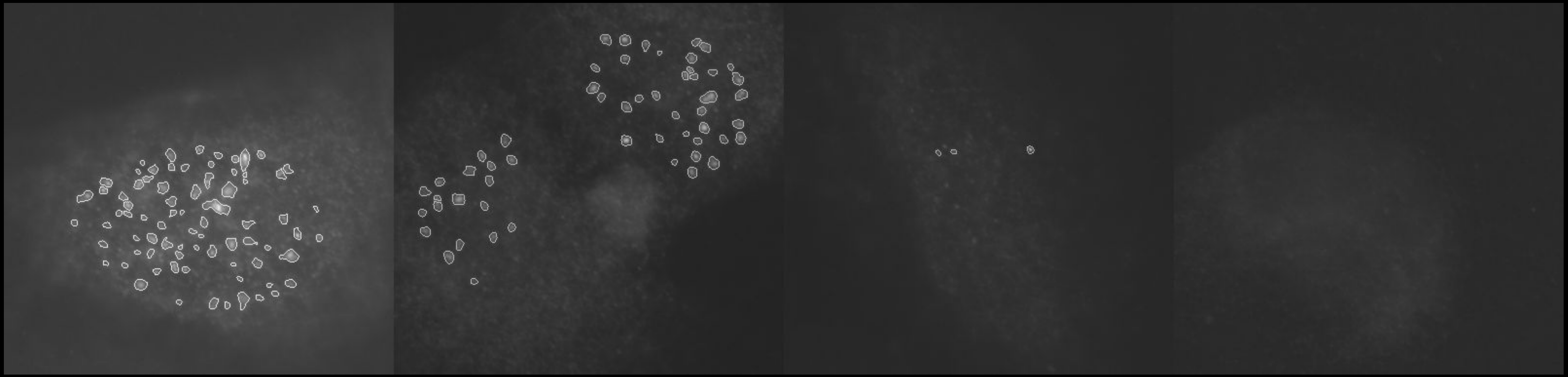
CellProfiler results	V-factor	
Z-factor	Wortmannin	LY294002
0.91	0.86	0.83

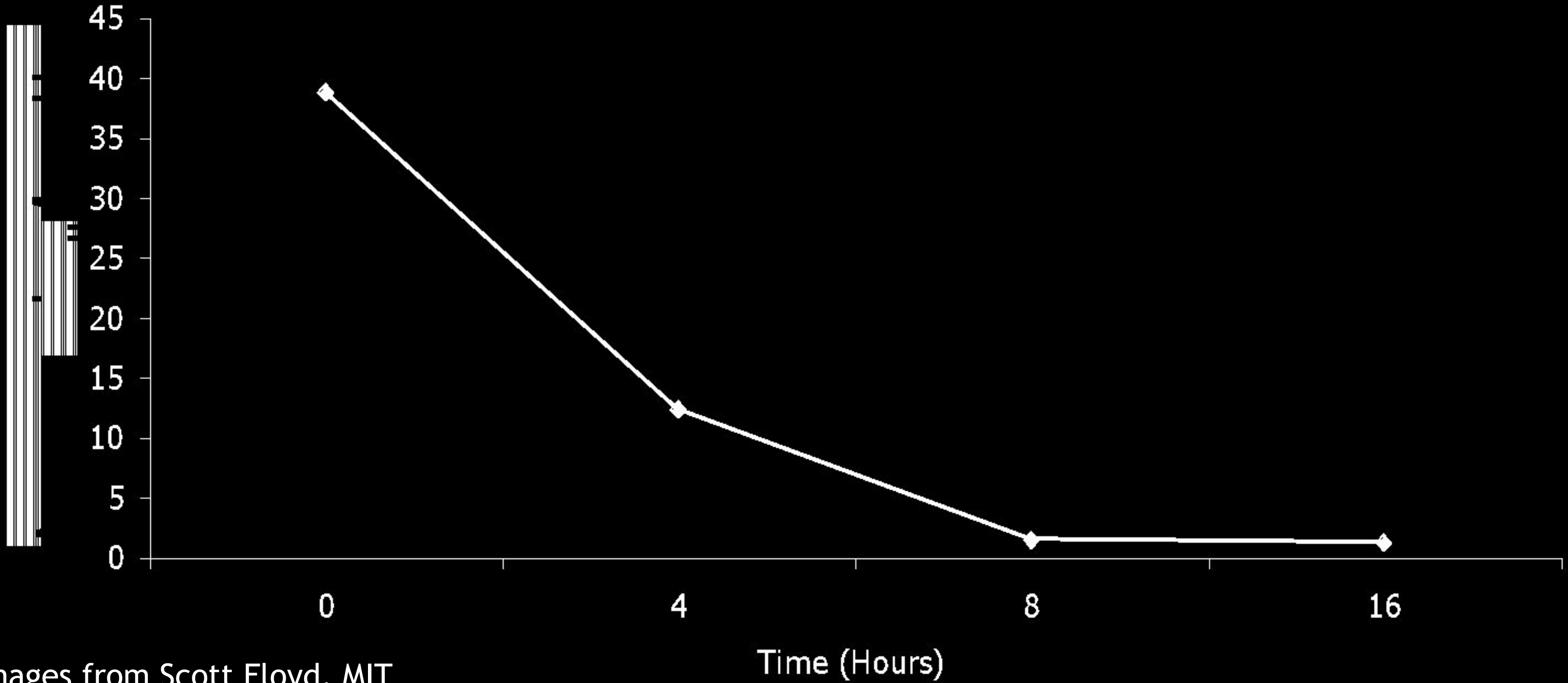
error bars = SEM

images from BioImage

Speckle-counting assay

Goal: count and measure phospho-Histone2AX speckles





images from Scott Floyd, MIT

First genome-wide screen: in progress

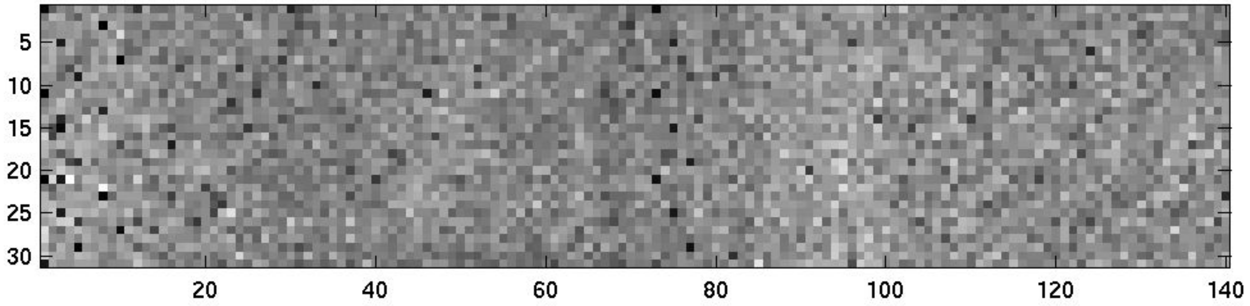
DNA staining: cell count
cell cycle distribution
chromatin texture
nuclear size
nuclear morphology
Actin staining: cell size
cell morphology
actin content
actin texture
phospho-H3: p-H3 amount
p-H3 localization

Every gene can be screened in a single experiment using four microscope slides!

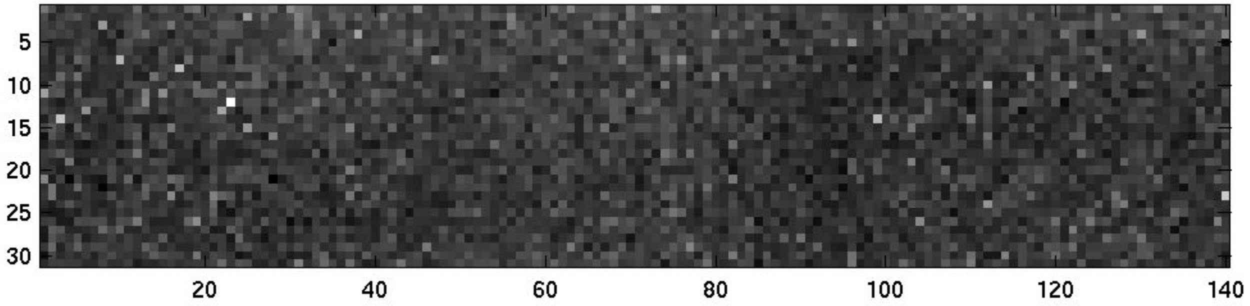
Data analysis: Population measures

black = low values, white = high values

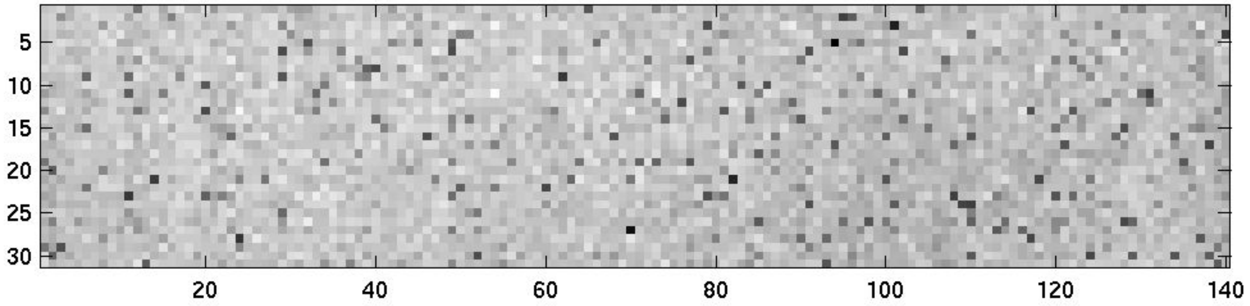
Cell count



Average cell area



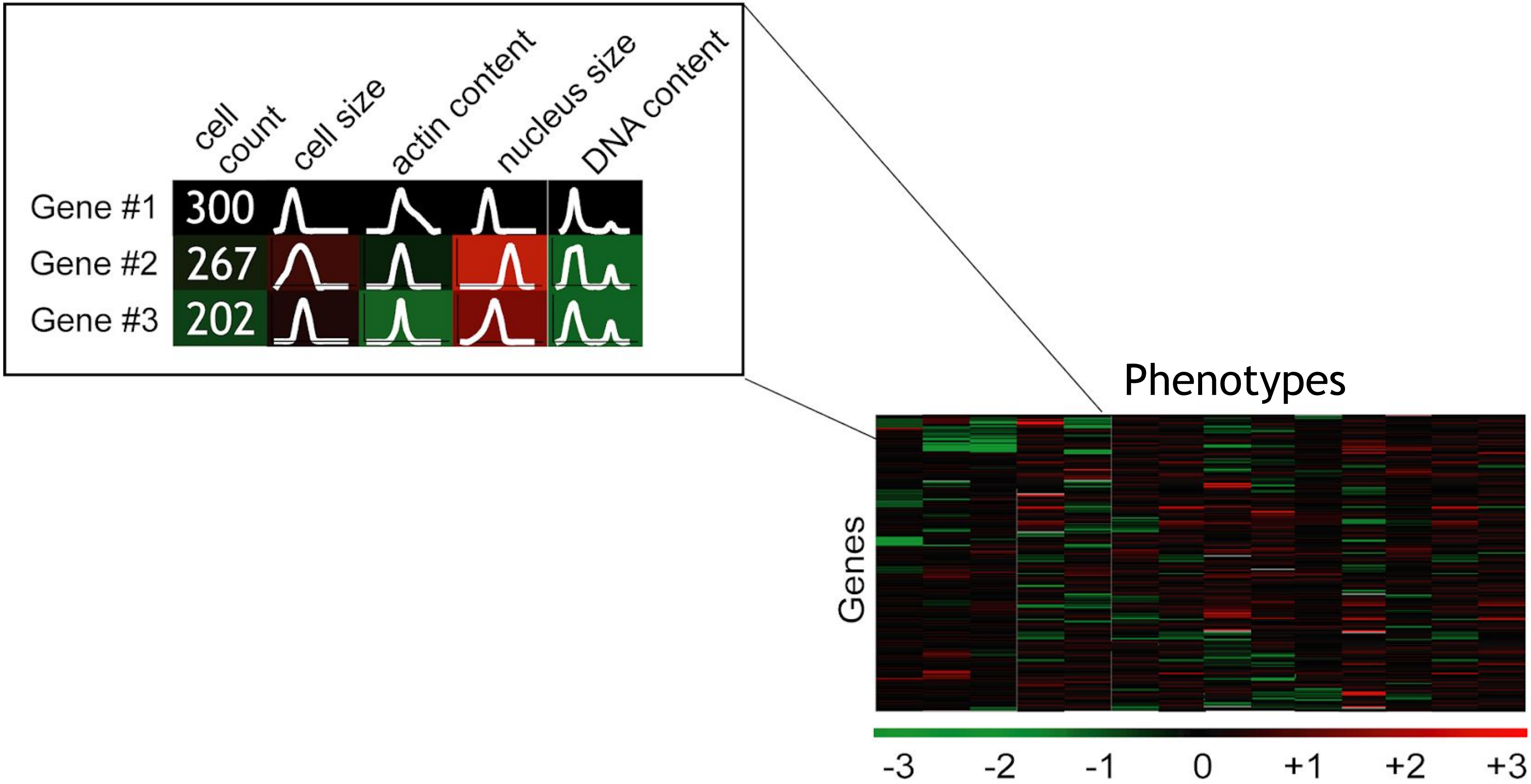
Correlation between actin and phospho-Akt staining



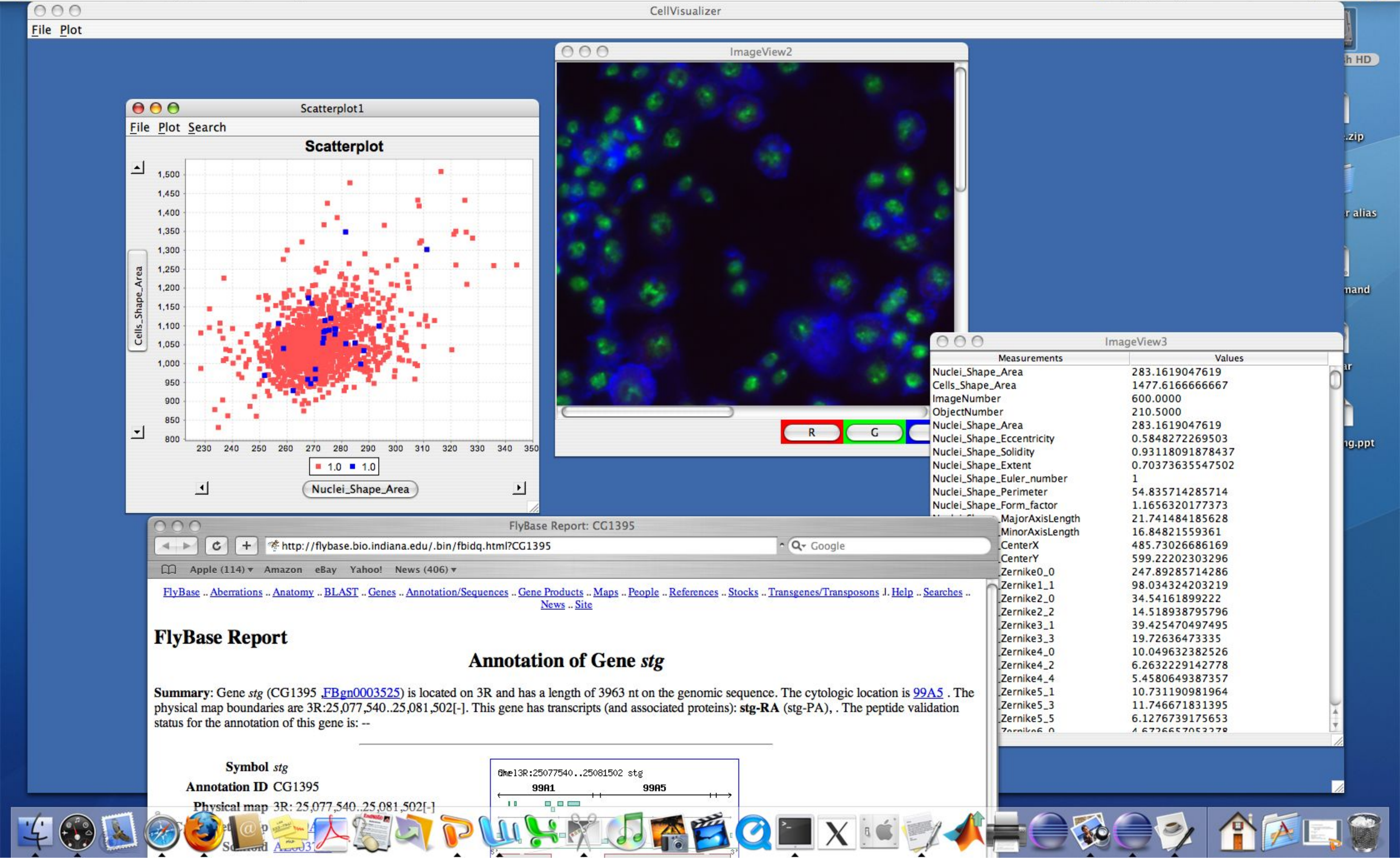
Data analysis: Population measures

True high-content data set produced by multi-parameter phenotypic analysis

Discovering the function of undescribed genes



In progress: data exploration with CellVisualizer



Measurements	Values
Nuclei_Shape_Area	283.1619047619
Cells_Shape_Area	1477.6166666667
ImageNumber	600.0000
ObjectNumber	210.5000
Nuclei_Shape_Area	283.1619047619
Nuclei_Shape_Eccentricity	0.5848272269503
Nuclei_Shape_Solidity	0.93118091878437
Nuclei_Shape_Extent	0.70373635547502
Nuclei_Shape_Euler_number	1
Nuclei_Shape_Perimeter	54.835714285714
Nuclei_Shape_Form_factor	1.1656320177373
_MajorAxisLength	21.741484185628
_MinorAxisLength	16.84821559361
_CenterX	485.73026686169
_CenterY	599.22202303296
_Zernike0_0	247.89285714286
_Zernike1_1	98.034324203219
_Zernike2_0	34.54161899222
_Zernike2_2	14.518938795796
_Zernike3_1	39.425470497495
_Zernike3_3	19.72636473335
_Zernike4_0	10.049632382526
_Zernike4_2	6.2632229142778
_Zernike4_4	5.4580649387357
_Zernike5_1	10.731190981964
_Zernike5_3	11.746671831395
_Zernike5_5	6.1276739175653

http://flybase.bio.indiana.edu/.bin/fbidq.html?CG1395

FlyBase .. Aberrations .. Anatomy .. BLAST .. Genes .. Annotation/Sequences .. Gene Products .. Maps .. People .. References .. Stocks .. Transgenes/Transposons .. Help .. Searches .. News .. Site

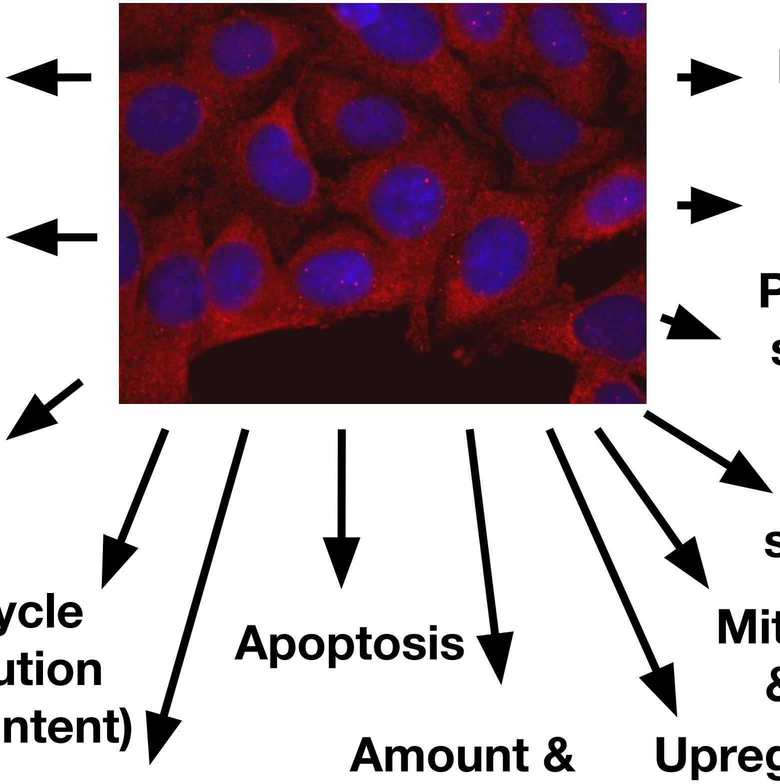
FlyBase Report

Annotation of Gene *stg*

Summary: Gene *stg* (CG1395 ,FBgn0003525) is located on 3R and has a length of 3963 nt on the genomic sequence. The cytologic location is 99A5 . The physical map boundaries are 3R:25,077,540..25,081,502[-]. This gene has transcripts (and associated proteins): **stg-RA** (stg-PA), . The peptide validation status for the annotation of this gene is: --

Symbol *stg*
Annotation ID CG1395
Physical map 3R: 25,077,540..25,081,502[-]

What are all the genes doing?



Cell size, count & morphology

Nuclear size & morphology

Nucleolar size, count & morphology

Cell cycle distribution (DNA content)

Upregulation of transcription

Apoptosis

Amount & distribution of signaling lipids

Upregulation of translation

Mitochondrial size & morphology

Localization of signaling proteins

Phosphorylation of signaling proteins (S6, S6K, Akt)

Actin content/ morphology

Protein content

SYSTEMATIC GENOME-WIDE SCREENS OF GENE FUNCTION

Anne E. Carpenter and David M. Sabatini

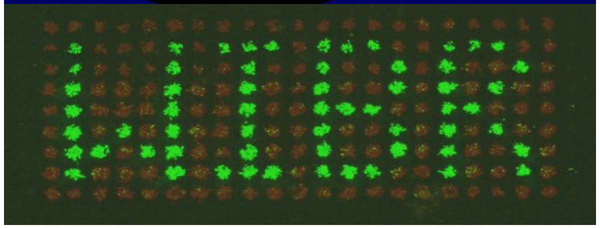
Nature Reviews Genetics 5:11-22 (2004)

Thanks to…

David M. Sabatini Laboratory

Anne Carpenter
Colin Clarke
Xana (Maria) Frias
David Guertin
Kalyani Guntur
Thouis R. Jones
Mike Lamprecht
Wyman Li
Susan Ma
Jason Moffat
Kathleen Ottina
Tim Peterson
Dos Sarbassov
Yasemin Sancak
Shomit Sengupta
Joon-Ho Sheen
Carson Thoreen
Doug Wheeler

Whitehead Institute
for Biomedical Research



CellProfiler™
cell image analysis software
www.cellprofiler.org

Created by:
Anne E. Carpenter and Thouis R. Jones

In the laboratories of:
David M. Sabatini and Polina Golland

at:
the Whitehead Institute and MIT

with help from:
Michael Lamprecht, Colin Clarke, In Han Kang, Ola Friman, Steve Lowe, Joo Han Chang, Susan Ma

CellProfiler has been supported by funding from:

- Novartis postdoc fellowship from the Life Sciences Research Foundation
- Merck/MIT Computational & Systems Biology postdoc fellowship
- Society for Biomolecular Screening Small Grant Award
- DOD Tuberous Sclerosis Complex Grant
- Sabatini Laboratory