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)
- Apoptosis
- Amount & distribution of signaling lipids
- Upregulation of transcription
- Mitochondrial size & morphology
- Protein content
- Actin content/morphology
- Phosphorylation of signaling proteins (S6, S6K, Akt)
- Localization of signaling proteins

SYSTEMATIC GENOME-WIDE SCREENS OF GENE FUNCTION

Nature Reviews Genetics

Anne E. Carpenter and David M. Sabatini
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

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

- cDNA expression → gain of function
- RNA interference → loss of function
- small molecules → chemical genetics/ drug discovery

5,600 spots per slide

Drosophila: RNAi effective, reliable, and convenient

No interferon response

Less redundancy vs. mammals

Genome-wide libraries available

Genome-wide screens in Drosophila

One glass slide: 5600 spots of dsRNA, each knocking down a different gene.

DNA: dark spot = lack of cells

Actin

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

...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

CellProfiler™
cell image analysis software

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

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

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

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

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

For all colors
Cell count validation

Human HT29

Drosophila
Cell area validation

The diagram shows the mean cell area measured by Coulter Counter and CellProfiler with different conditions: no dsRNA, Mad2, String, Anillin, and Cyclin A. The y-axis represents the mean cell area (arbitrary units), and the x-axis lists the conditions.
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

- No illumination correction
- CellProfiler illumination correction
- White reference illumination correction

Number of cells (thousands)

DNA content, log scale (arbitrary units)
Validation for DNA content

Human HT29

Nuclei image

CellProfiler-outlined nuclei

DNA content, log scale (arbitrary units)

mouse

Wild-type

Knockout

images from Andrew Baltus, Page lab, Whitehead Institute

Moffat et al., Cell, in press
Validation for DNA content

C. elegans cells
- Before division: 471
- After division: 246, 243

Human HeLa cells
- Before division: 932
- After division: 462, 456

S. cerevisiae cells
- Before division: 257, 538
- After division: 317, 260, 263
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
Time lapse movies of Drosophila embryos
Antibody staining intensity - Mouse tissue

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

Goal: quantify the localization of proteins

Cytoplasm-nucleus translocation assay

Goal: quantify the localization of proteins

CellProfiler results

<table>
<thead>
<tr>
<th>V-factor</th>
<th>Z-factor</th>
<th>Wortmannin</th>
<th>LY294002</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.91</td>
<td>0.91</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
</tr>
<tr>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
</tr>
</tbody>
</table>

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

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

<table>
<thead>
<tr>
<th></th>
<th>cell count</th>
<th>cell size</th>
<th>actin content</th>
<th>nucleus size</th>
<th>DNA content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene #1</td>
<td>300</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene #2</td>
<td>267</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene #3</td>
<td>202</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Phenotypes
In progress: data exploration with CellVisualizer

**FlyBase Report**

**Annotation of Gene stg**

Summary: Gene *stg* (CG1395 [FlyBase](http://flybase.bio.indiana.edu/bin/fboq.html?CG1395)) is located on 3R and has a length of 3963 nt on the genomic sequence. The cytologic location is **99A3**. The physical map boundaries are 3R:25,075,540-25,081,5021. 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: 25,075,540-25,081,5021
What are all the genes doing?

- Cell size, count & morphology
- Nuclear size & morphology
- Cell cycle distribution (DNA content)
- Apoptosis
- Uptregulation of transcription
- Amount & distribution of signaling lipids
- Mitochondrial size & morphology
- Nuclear size, count & morphology
- Actin content/morphology
- Phosphorylation of signaling proteins (S6, S6K, Akt)
- Localization of signaling proteins
- Protein content

SYSTEMATIC GENOME-WIDE SCREENS OF GENE FUNCTION

Nature Reviews Genetics

Anne E. Carpenter and David M. Sabatini
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

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