Extracting rich information from biological images to tackle world health problems

Anne E. Carpenter, Ph.D.
Images contain a wealth of information
Cells or organisms 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

  - CellProfiler™
    - Cell image analysis software
  - CellProfiler Analyst
    - Data exploration software
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

[Image of a yeast patch growth experiment]
Case study: Tuberculosis

Estimated TB incidence rates, by country, 2006

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

Try to kill **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

mouse nuclei
tuberculosis bacteria

Martha Vokes
Mark Bray

Deb Hung, Broad/MGH
Sarah Stanley, postdoc
Amy Barczak, postdoc

project in progress
Search for tuberculosis treatments

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

Status: pursuing hits that prevent bacterial infection/expansion but do NOT kill the bacterium directly

Add 10,000 bioactive chemicals, each chemical in a different well
Search for tuberculosis treatments

Without drug

With drug

mouse nuclei

tuberculosis bacteria

Status: pursuing hits that prevent bacterial infection/expansion but do NOT kill the bacterium directly

project in progress
Polyploidization of megakaryocytes - AMKL (leukemia)

DNA stain, with outlines identifying the nuclei

DMSO (negative control)

SU6656 (positive control)

Status: recommended clinical testing of AURKA inhibitor for AMKL

Wen, ..., Carpenter ... Crispino, et al. Cell 2012
Cancer radiation treatment

Mike Yaffe’s lab at MIT’s Center for Cancer Research (Scott Floyd & Michael Pacold, postdocs) & The RNAi platform/TRC at the Broad Institute

- Response to DNA damage
- Recovery from DNA damage
- p-H2AX foci

Mitochondrial abundance

Vamsi Mootha’s lab at Harvard Medical School/Mass General Hospital (Toshi Kitami, postdoc)

- Negative control
- Positive control
- DNA
- Mitotracker

S6K signaling

Anne Carpenter & David Sabatini at the Whitehead Institute for Biomedical Research (Rob Lindquist, student)

- Control
- dTOR RNAi
- Blue = DNA
  - Green = actin
  - Red = p-S6

Lindquist ... Carpenter, Genome Research, 2011
Screens to identify genes & chemicals: increasing physiological relevance

Cell-based assays with complex morphologies

Co-cultured mixtures of cell types

Whole organisms

Image-based profiling: extracting ‘signatures’
Breast cancer

Control + Growth factor

DNA
Actin

Ray Jones
Anne Carpenter
Eric Lander, Broad Institute
Piyush Gupta, postdoc

project in progress
“Cytological profile”: collection of measurements describing the appearance of a cell

Perlman, et al. Science 2004
User-friendly machine learning

fetching 25 mitotic cells from group Gene: gene=NME1
Challenging cellular phenotypes
Regulators of cell division

DNA

DNA

- Interphase
- Mitosis
- Monopole (abnormal)

Tsui ... Carpenter ... Mitchison, et al. PLoS ONE, 2009
Regulators of cell division

Normal: one nucleus per cell

Abnormal: two nuclei per cell

Ray Jones  Martha Vokes  Riki Eggert  Adam Castoreno

Castoreno... Carpenter ... Eggert, Nature Chem Bio, 2010
RNAi screen: glioblastoma proliferation & differentiation

Neurosphere phenotype

Flat, elongated phenotype

DNA / Tubulin

Martha Vokes
Mark Bray

David Sabatini, Whitehead Institute
Yakov Chudnovsky, postdoc
William Hahn, Broad Institute
Milan Chheda, postdoc
David Root, Broad Inst.
Screens to identify genes & chemicals: increasing physiological relevance

Cell-based assays with complex morphologies

Co-cultured mixtures of cell types

Whole organisms

Image-based profiling: extracting ‘signatures’
Co-cultured cell systems

Two or more cell types cultured together in order to maintain physiological conditions

Necessary

• Many primary cell types lose their physiological functions when grown in isolation

Challenging

• Culture conditions are difficult to optimize and less robust

• Need to distinguish the cell type of interest from the co-cultured cells, ideally without using additional cellular stains
Hepatocyte proliferation using human primary liver cells co-cultured with fibroblasts

DNA

Control

Hepatocyte-enriched

Z’ factor = 0.29 (positive control = 2x hepatocytes)

Status: identified chemicals that stimulate primary human hepatocyte proliferation

Shan, ... Carpenter, Bhatia, Nature Chem Bio, in press
Studying liver-stage malaria using human primary liver cells

project in progress
Leukemic & hematopoetic stem cells (HSCs/LSCs) using mouse primary HSCs or LSCs co-cultured with stromal cells

Co-cultured LSCs and stroma

LSC channel only: live, no DNA stain

Status: identified drugs that preferentially kill leukemic cells

David Logan

Co-cultured LSCs

Todd Golub, Broad Institute
Ben Ebert, HMS, Dana Farber
Gary Gilliland, Brigham & Women's Hospital
David Scadden, Mass. General Hospital
Stuart Schreiber, Broad Institute

postdocs and students: Kimberly Hartwell, Alison Stewart, Peter Miller, et al.

project in progress
Screens to identify genes & chemicals: increasing physiological relevance

- Cell-based assays with complex morphologies
- Co-cultured mixtures of cell types
- Whole organisms

Image-based profiling: extracting ‘signatures’
C. elegans High-Throughput Screening Facility

Consistent sample preparation & imaging at 4-6 plates per hour
96-well agar or 384-well liquid culture

Media Dispensing
Worm sorting & dispensing (COPAS BioSort)
Image acquisition on automated microscope

Ausubel & Ruvkun labs

Annie Lee-Conery  Gang Wu  Eyleen O’Rourke  Terry Moy  Jonah Larkins-Ford
Novel anti-infectives against *E. faecalis*

*C. elegans* + *E. faecalis* = death

*C. elegans* + *E. faecalis* + rescuing drug = survival

Brightfield

(SYTOX stain not shown)

37,214 compounds and extracts

80 known antibiotics and analogs

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

Ray Jones
Anne Carpenter

Fred Ausubel, Harvard/Mass. General Hospital
Terry Moy
Annie Lee Conery
Gang Wu

Moy ... Carpenter ... Ausubel, et al. ACS Chem Bio 2009
Screens for anti-infectives

Goal: score worm viability based on shape, after detangling the worm clusters

Z’ factor = 0.27
(Z’ factor for visual scoring = 0.33)

Riklin-Raviv, Ljosa, Conery, Ausubel, Carpenter, Golland, Wählby: MICCAI 2010

Wählby, Riklin-Raviv, Ljosa, Conery, Golland, Ausubel, Carpenter: ISBI 2010

Wahlby ...Carpenter, Nature Methods 2012
Reporter expression in response to infection

$Z'$ factor $= 0.21$

Wahlby ... Carpenter, Nature Methods 2012
Regulators of fat accumulation/metabolism

Goal: Quantify Oil Red O staining

wild type

daf-2

Status: identified genes where RNAi alters fat metabolism

project in progress
Screens to identify genes & chemicals: increasing physiological relevance

- Cell-based assays with complex morphologies
- Co-cultured mixtures of cell types
- Whole organisms

Image-based profiling: extracting ‘signatures’
Latent information in image-based experiments

- thousands of features per cell
- thousands of cells per sample

“Cytological profile”: collection of measurements describing the appearance of a cell

Perlman, et al. Science 2004
Automatically extracting image-based phenotypes

Wild-type cells

Perturbed cells

Detectable phenotypic difference?

Extract features

Extract features

Identify difference in phenotype from image features

even if “invisible” to the human eye

Screen for chemicals that can revert mutant phenotype -> wild type
Cells from patient without mental illness

Detectable phenotypic difference?

Cells from patient with bipolar disorder

Cells treated with hepatotoxic drug

Cells treated w/ non-hepatotoxic analog

Wild-type cells

Cells treated with RNAi against HDAC1, 2, 3...

Postdoctoral position available!

Ray Jones  Vebjorn Ljosa  Kate Sokolnicki  Auguste Genovesio  Shantanu Singh

Bruce Cohen, McLean Hosp.  Rakesh Karmacharya

Todd Golub, Broad Institute  David Thomas

Stuart Schreiber, Broad Institute  Angela Koehler
In order to detect phenotypic differences

• Biology
  – The relevant biological difference must be present (cell type, culture conditions, etc)

• Labels
  – We must have stained the samples so that the differences are visible

• Image features
  – Feature extraction algorithms must capture the differences

• Similarity measure
  – We must compare the features effectively
Comparing profiling methods for predicting MOA (mechanism of action, for small molecules)

(A) means  (B) KS statistic  (C) normal vector of SVM hyperplane  (D) GM  (E) factor analysis

12 mechanisms of action

103 treatments

453 features

profile values

103 treatments
distance calc.
cosine angle

103 treatments

mechanism-of-action prediction

1NN

103 treatments

image feature extraction

cells

profile calc.

103 treatments

project in progress

Vebjorn Ljosa
Rob ter Horst

IMAMILY
In order to detect phenotypic differences

• Biology
  – The relevant biological difference must be present (cell type, culture conditions, etc)

• Labels
  – We must have stained the samples so that the differences are visible

• Image features
  – Feature extraction algorithms must capture the differences

• Similarity measure
  – We must compare the features effectively
Maximize information in profiling experiments

Cell-painting assay

6 stains, 5 channels imaged, revealing:
nucleus, nucleoli, actin, golgi, plasma membrane, mitochondria, ER
Profiling more diverse compounds

1838 bioactives
2029 MLPCN
17560 DOS
Profiling more diverse compounds

203 out of 1,600 compounds yielded a phenotype
Can profiling be used to identify similar compounds in a large-scale experiment?

- Profile ~200 “query” compounds from ~35 screens
- Identify similar-performing compounds from a library of 10,000 commercially available compounds + 20,000 novel compounds (DOS)
- Use imaging and gene expression for profiling
• Downloaded >38,000 times
• Launched >90,000 times/year
• Cited in >600 papers
• One of the Top 10 most-accessed papers of all time in Genome Biology
• Winner of Bio-IT World's IT & Informatics Best Practices Award in 2009

As of December, 2012
ProtocolNavigator: reproducible research
Paramorama: optimizing image processing

Roy Ruddle, Univ. Leeds
Hannes Pretorius
Helpful resources

www.cellprofiler.org -> Tutorials

Introduction to the Quantitative Analysis of Two-Dimensional Fluorescence Microscopy Images for Cell-Based Screening

Vebjorn Ljosa, Anne E. Carpenter

Box 1. Resources for further exploration

Box 2. Practicalities
Gratitude

Many thanks to our many biology collaborators

Mark Bray
Lee Kamentsky
Vebjørn Ljoså
David Logan

Shantanu Singh
Matthew Veneskey
Carolina Wählby

Recent major funding for this work provided by:

• NIH NIGMS R01 GM089652 (Carpenter)
• NIH NIGMS R01 GM095672 (Wahlby)
• NSF CAREER award (Carpenter)
• Human Frontiers in Science Program (Carpenter)

anne@broadinstitute.org

Gratitude

free, at www.cellprofiler.org: