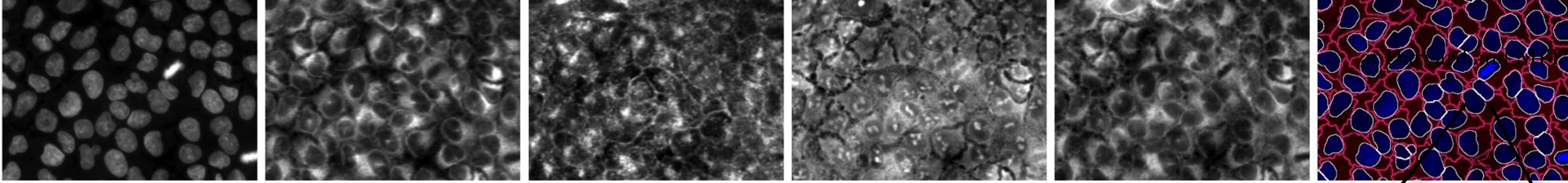
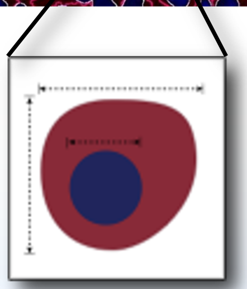




Advancing drug discovery and diagnostics through image-based AI

Anne E. Carpenter, PhD

BROAD INSTITUTE IMAGING PLATFORM

@DrAnneCarpenter
(*tweet all you like*)

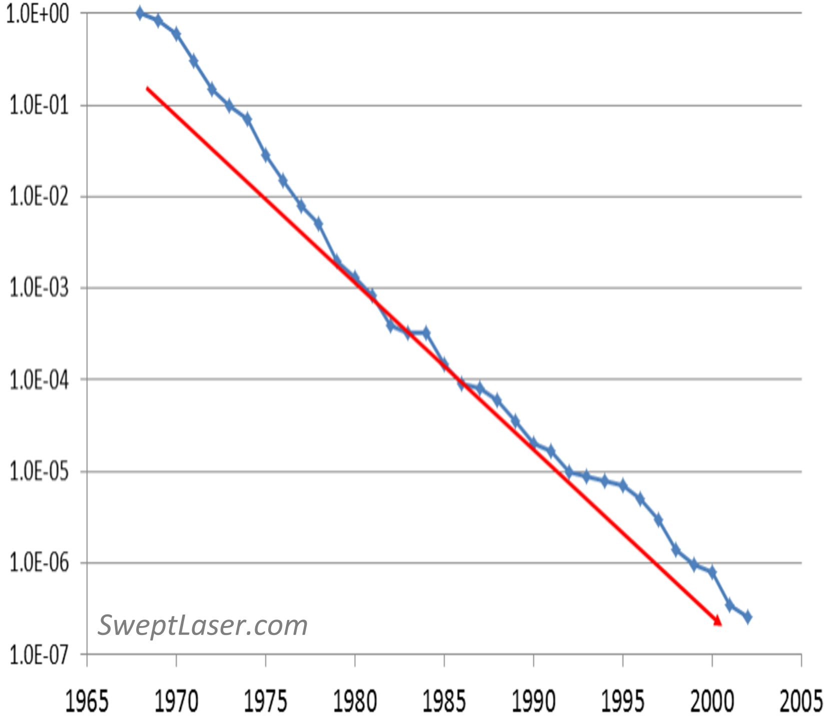
Download PDF of slides here: http://broad.io/CarpenterSlides

Give feedback on this talk: http://broad.io/CarpenterFeedback

Tale of two industries

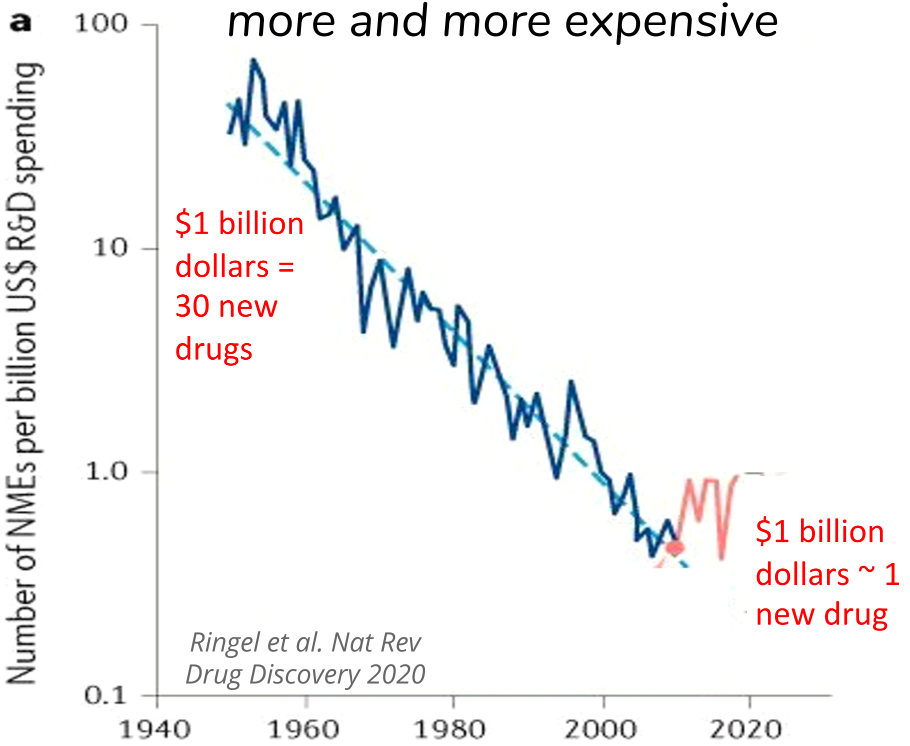
Moore's Law

Compute gets cheaper and cheaper



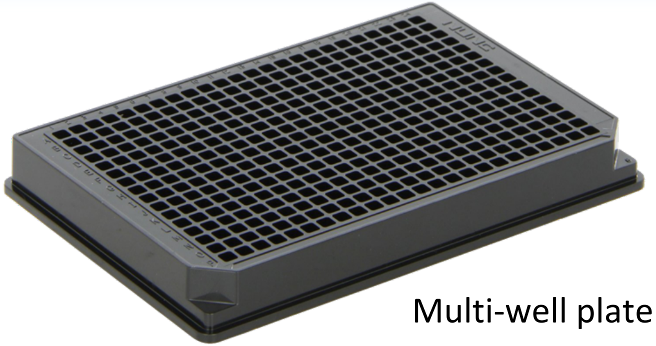
Eroom's Law

Discovering new medicines gets more and more expensive



Discovering drugs in high-throughput?

x 384 =



Multi-well plate

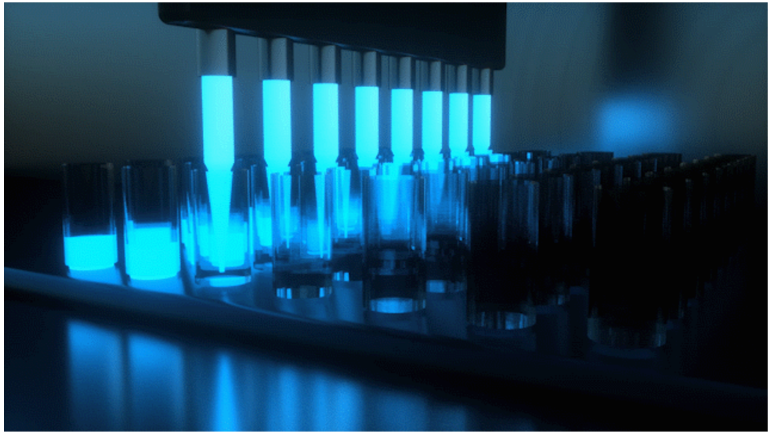
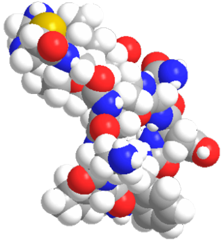


Image: Nalgene; video: Chemistry World

Drug discovery (we wish)

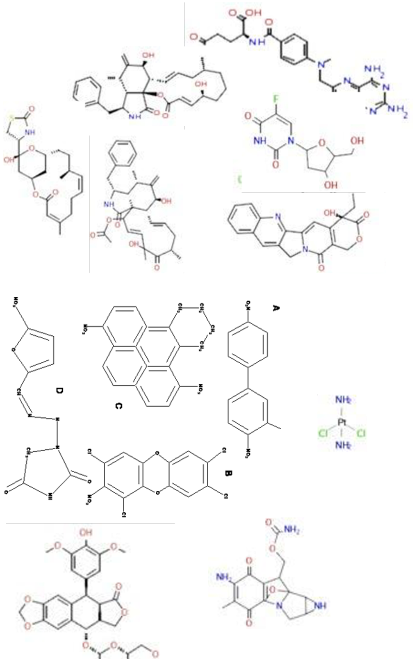
chemical + diseased human = healthy human



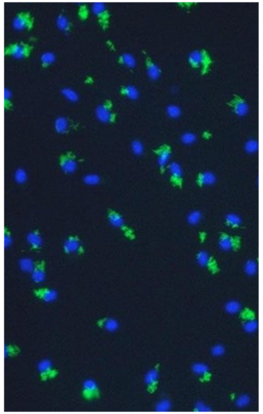
Drug discovery using cells as proxies

assay

millions of chemicals



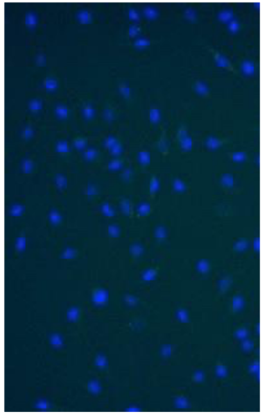
\+ "diseased" cells



cells
\+
tuberculosis bacteria

→

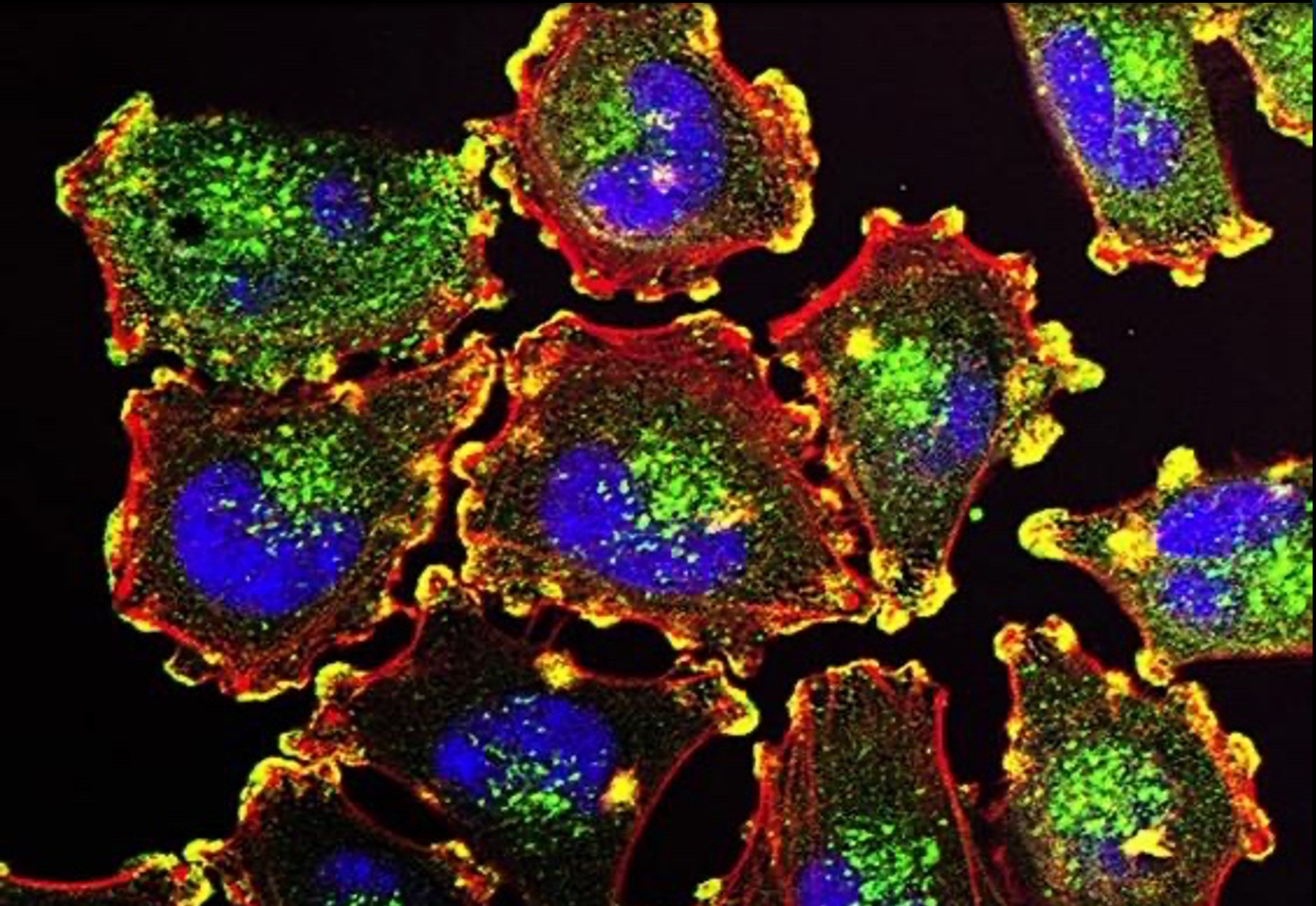
"healthy" cells



cells are alive!
\+
tuberculosis is gone!

readout (phenotype)

Images contain a wealth of information

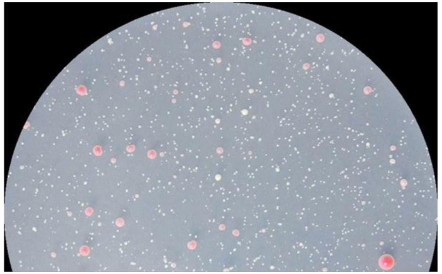


1. single-cell resolution
2. complex model systems
3. quantitative
4. multiplexed

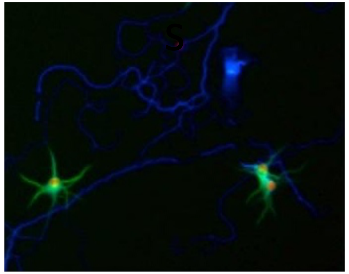
Image credit: Wikimedia

2. Complex models can be quantified by imaging

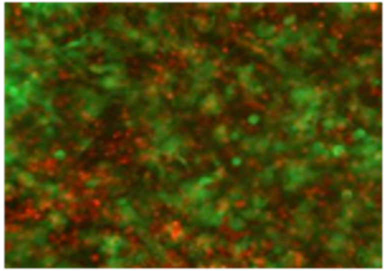
Yeast



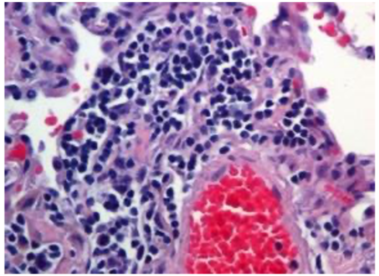
Neuron



Co-cultures

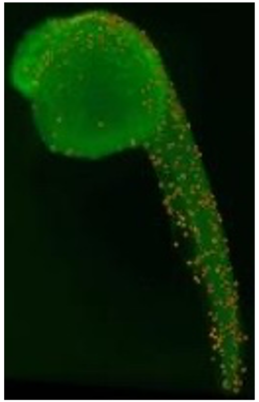


Tissue

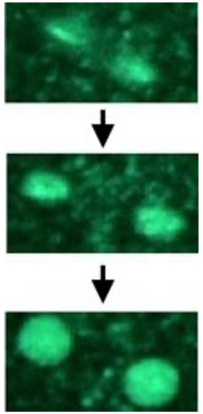


C. elegans

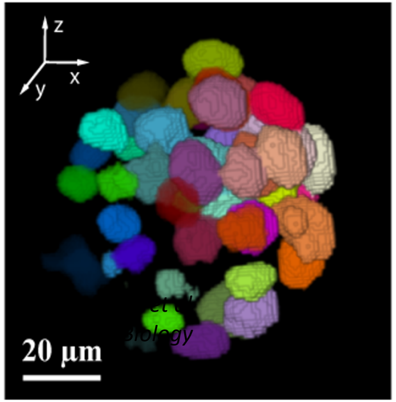
Zebrafish



Time-lapse

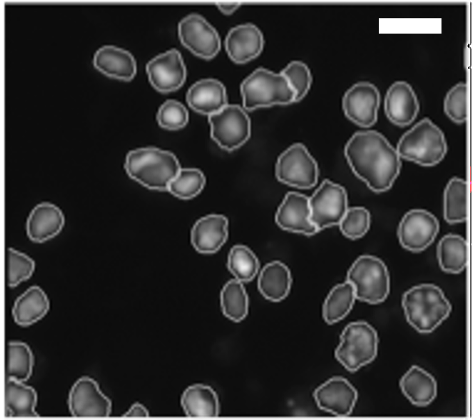


3D

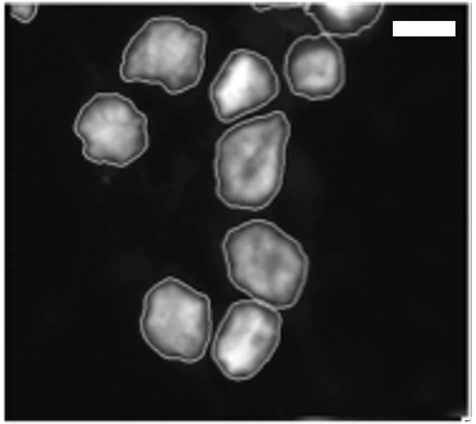


3. Image analysis is quantitative

DMSO
(negative control)

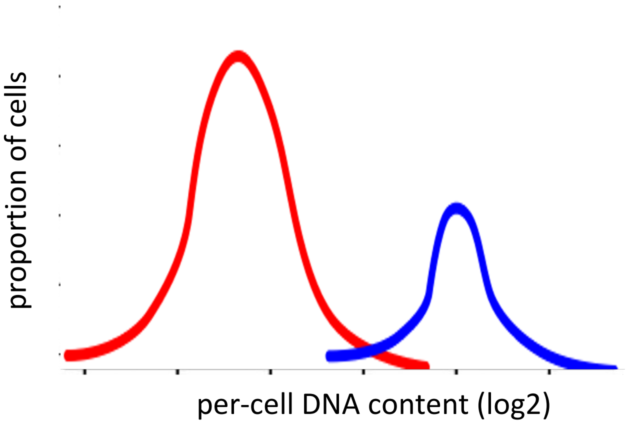


SU6656
(positive control)



Polyploidization of megakaryocytes - AMKL (leukemia)

Leukemic cells (stained nuclei, outlined by software)



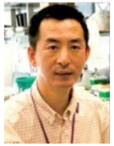
Clinical trial success:
Alisertib for myelofibrosis
Gangat et al. Clin Cancer Research 2019

Wen, et al. Cell 2012

BROAD INSTITUTE IMAGING PLATFORM

Mark Bray

Martha Vokes



Jeremy Wen, postdoc

John Crispino, Northwestern University

4. Imaging is highly multiplex: multiple stains

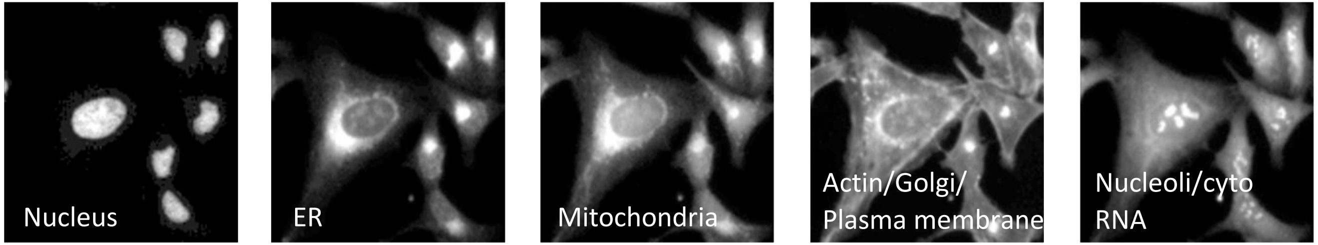
BROAD INSTITUTE IMAGING PLATFORM

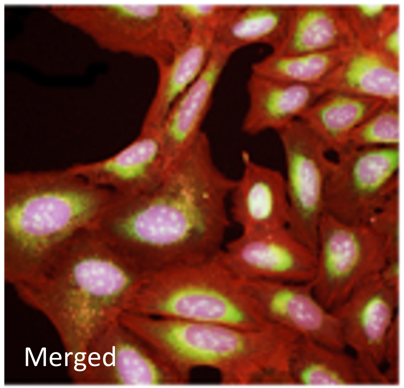

Kate Sokolnicki


Ray Jones

Cell Painting

6 stains, 5 channels imaged, revealing 8 constituents/organelles:






Sigrun Gustafsdottir


Paul Clemons


Aly Shamji


Stuart Schreiber, Broad Institute

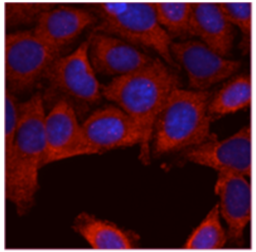
Gustafsdottir, et al. PLOS ONE 2013
Bray, et al. Nature Protocols 2016

Extracting features from images

Sample prep

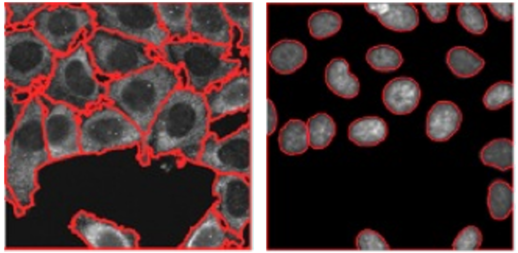
100's of 384-well plates

Microscopy



1000's of images/ plate; ~5 channels/image

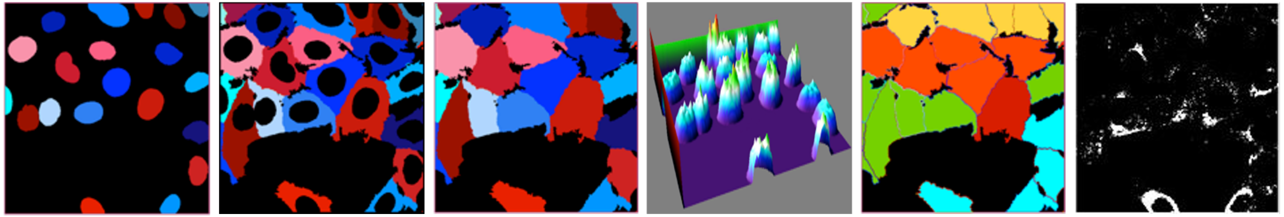
Image analysis: Segmentation



~500 cells/ image

Measure everything

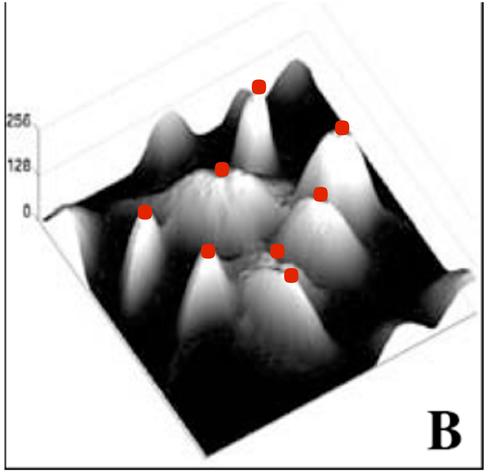
Counts, Shapes, Sizes, Intensities, Textures, Correlations, Relationships

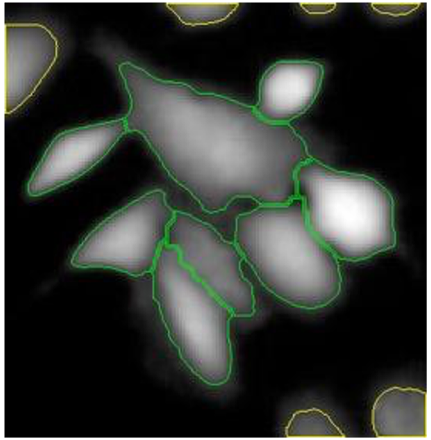


~1500 features / cell

CellProfiler™
cell image analysis software

- Powerful image analysis methods
- Point-and-click
- Biologist-focused help
- Flexible, modular, extensible
- Useful for both small and large experiments
- Supports reproducible research



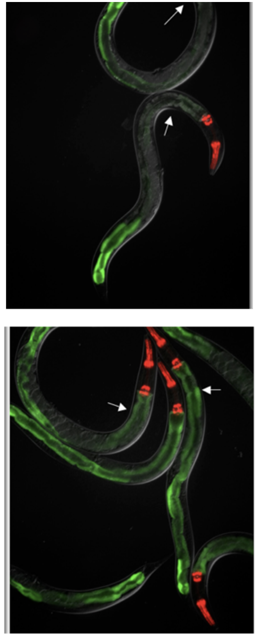


Images from Carolina Wahlby

- Free and open-source; Windows, Mac, Linux
- Cited in **2,000+** papers per year, **9,500+** total
- Ranked **most flexible** and **usable** in independent analysis (*Wiesmann et al.)*

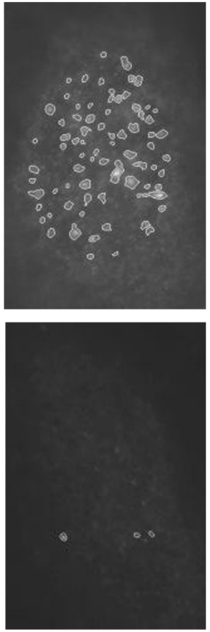
Screen all the things!

S. aureus infection



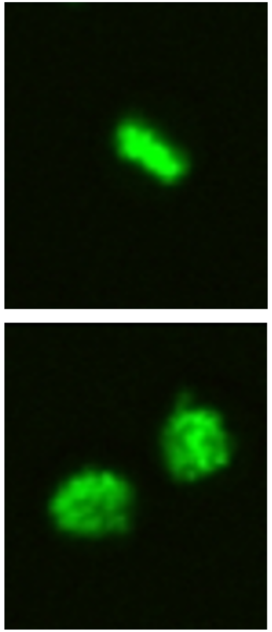
Ausubel/ Irazoqui labs

DNA damage



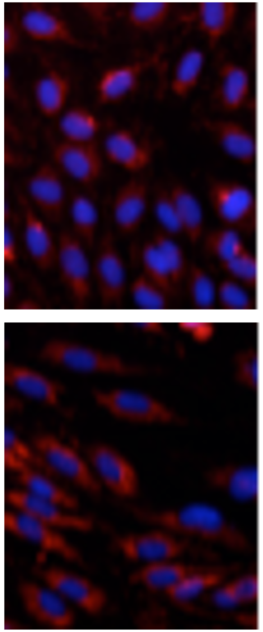
Yaffe lab

Mitosis



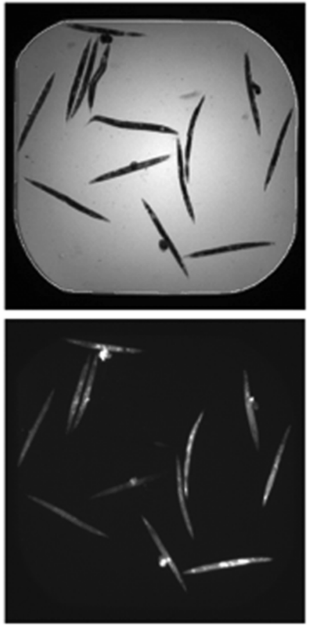
Mitchison lab

Mitochondrial abundance



Mootha lab

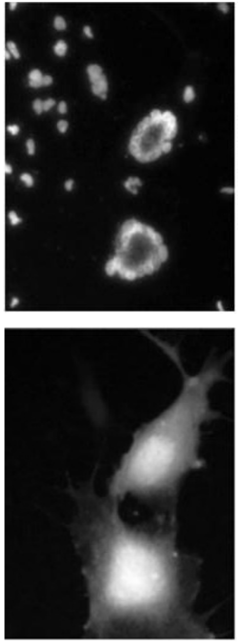
E. faecalis



Ausubel lab

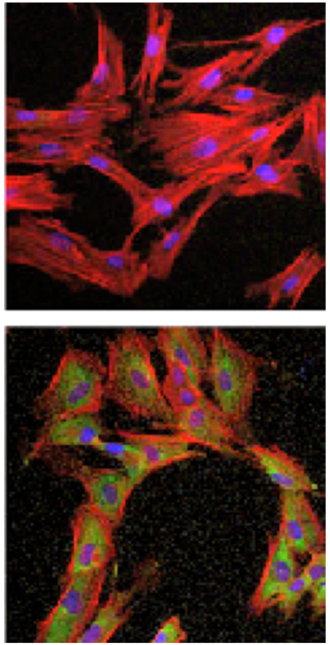
Screen all the things!

HIV neutralization



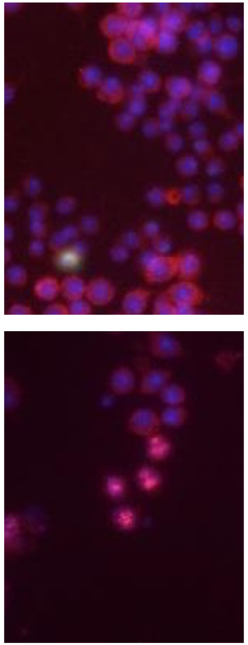
Fenyo lab

mTOR pathway activation



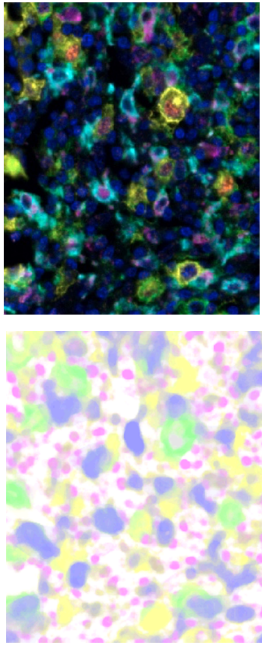
Sabatini lab

Meiosis



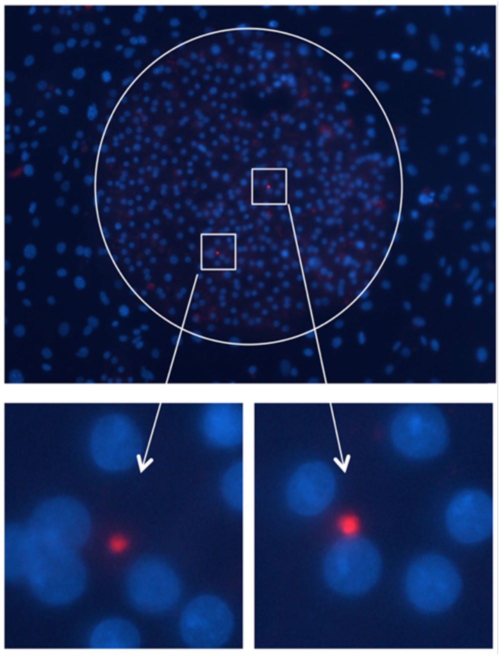
Orr-Weaver lab

Immune environment



Shipp/Rodig labs

Malaria



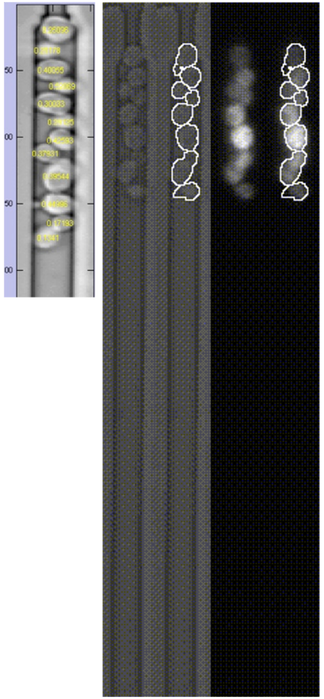
Bhatia lab

Screen all the things!

Lysosome transport, Gelfand lab

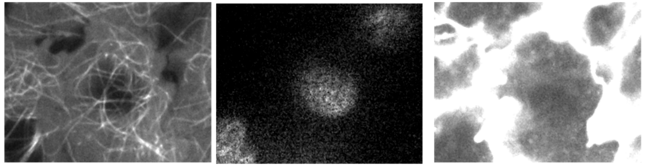
Jolly, et al. Cell Reports 2016

Yeast lineages, Weitz lab



Kim, et al. Proc. SPIE 2010

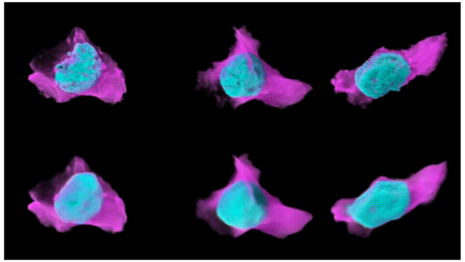
Protein localization, Allen Institute for Cell Science



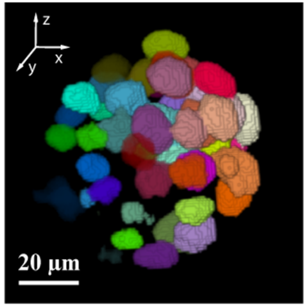
Cytoskeleton *(tubulin)*

DNA *(Hoechst)*

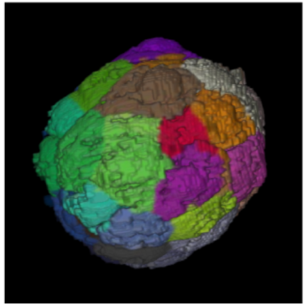
Membrane *(CellMask- membrane)*



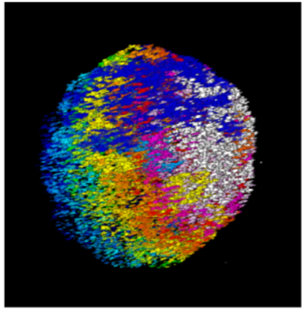
C. Segmented Nuclei



D. Segmented Cells



E. Segmented GAPDH

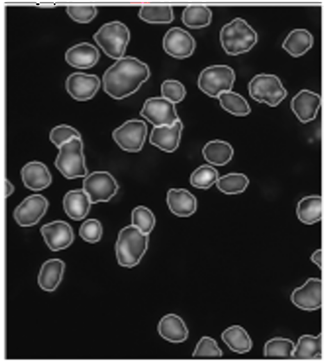


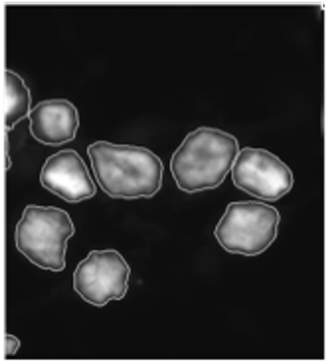
McQuin et al PLOS Biology 2018; Allen Institute for Cell Science

Impact on patients: successful clinical trials

BROAD INSTITUTE IMAGING PLATFORM

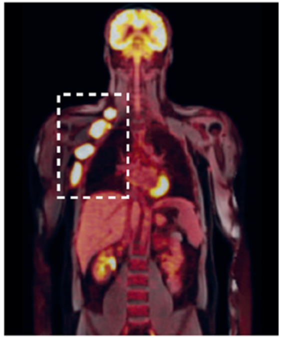
Myelofibrosis

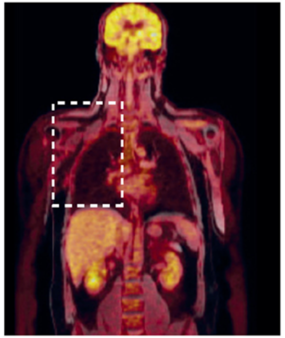




Crispino lab
(Gangat et al. Clin Cancer Research 2019)

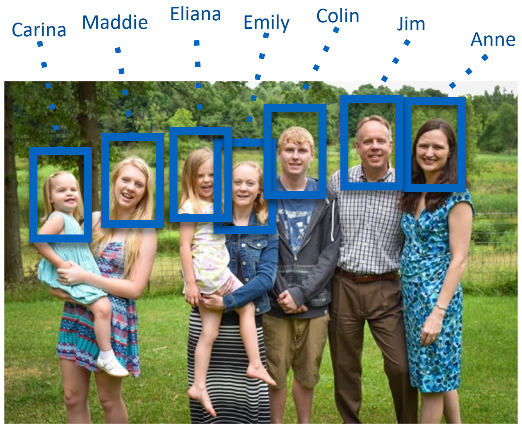
Leukemias &
Lymphomas



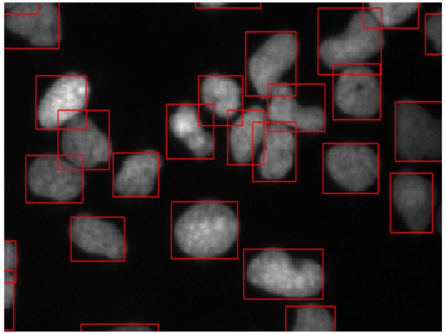


Vienna hospitals & medical institutes, ETH Zurich
(Snijder B et al. Lancet Haematology 2017)

Bringing deep learning to biologists

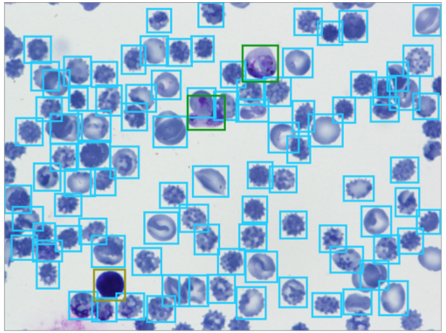


Finding objects
(e.g. nuclei)



Caicedo et al., Cytometry 2019
Caicedo et al., Nature Methods 2019
Data Science Bowl 2018

Classifying phenotypes
(e.g. stages of malaria life cycle)



Hung et al., CVPR workshop 2017

Anne Carpenter | **Allen Goodman** | **Kyle Karhohs** | **Jeanelle Ackerman** | **Beth Cimini** | **Minh Doan** | **Tim Becker** | **Shantanu Singh** | **Juan Caicedo** | **Jonathan Roth** | **Jane Hung**

What is deep learning?

Machine learning: builds an algorithm/model/set of rules that can predict outcomes (pixels / structures / phenotypes) given training examples to learn from

Deep learning: a type of machine learning using many layers of processing

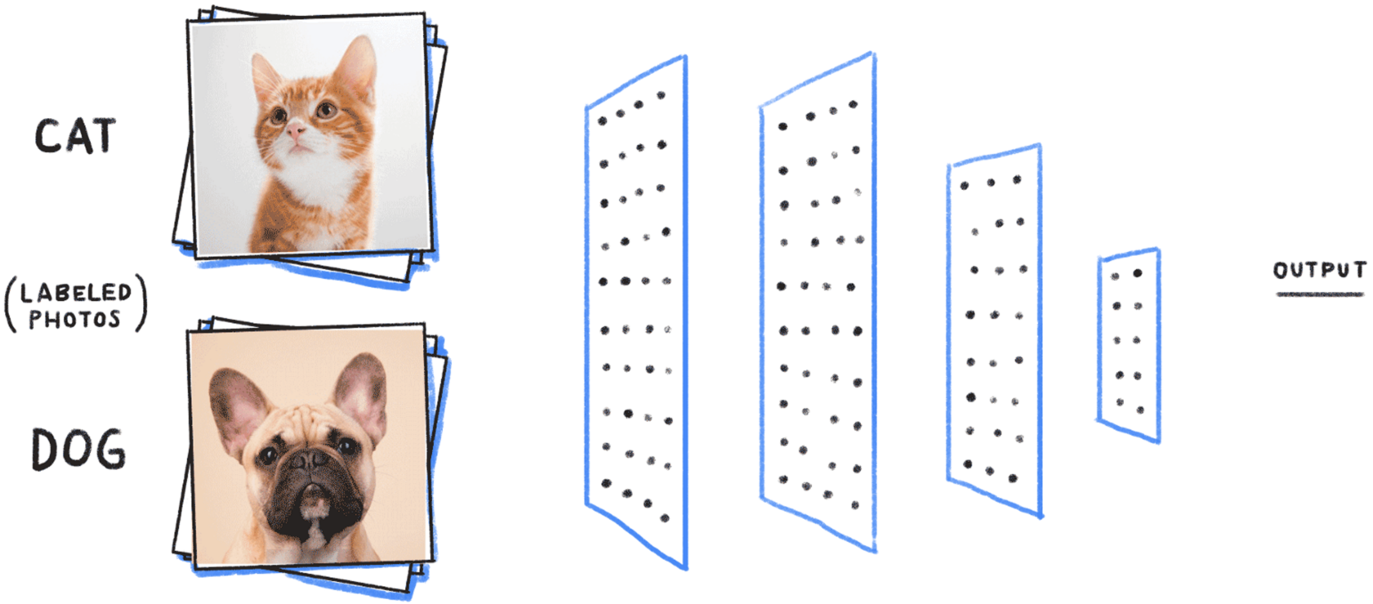
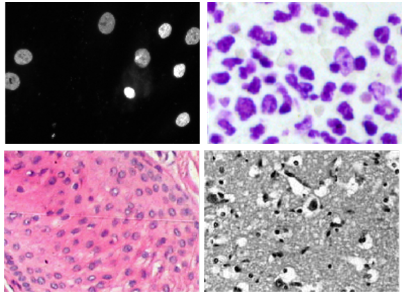


Image credit: becominghuman.ai

Deep learning excels at segmentation (finding things)

2018 Data Science Bowl: Towards a Universal Nucleus Finder



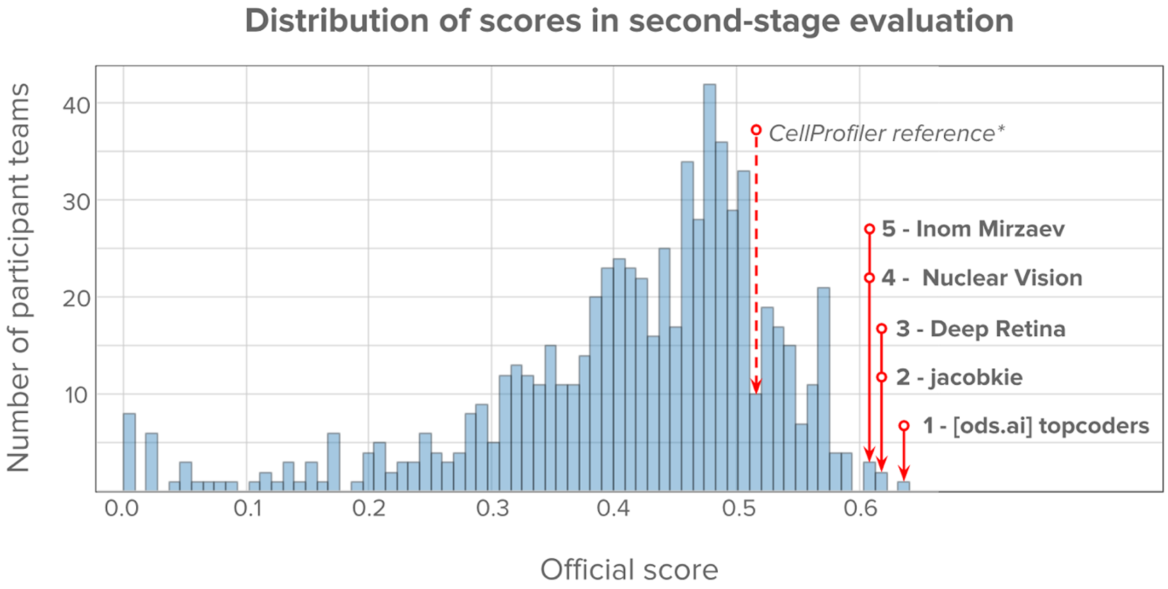
From Jan to Apr 2018
700 images
~37,333 nuclei
3,919 teams
65,333 submissions

Anne Carpenter | Allen Goodman | Kyle Karhohs | Jeanelle Ackerman | Beth Cimini | Minh Doan | Tim Becker | Shantanu Singh | Juan Caicedo

Caicedo et al., Nature Methods (2019)

Deep learning excels at segmentation (finding things)

2018 Data Science Bowl: Towards a Universal Nucleus Finder



Anne Carpenter	Allen Goodman	Kyle Karhohs	Jeanelle Ackerman	Beth Cimini	Minh Doan	Tim Becker	Shantanu Singh	Juan Caicedo

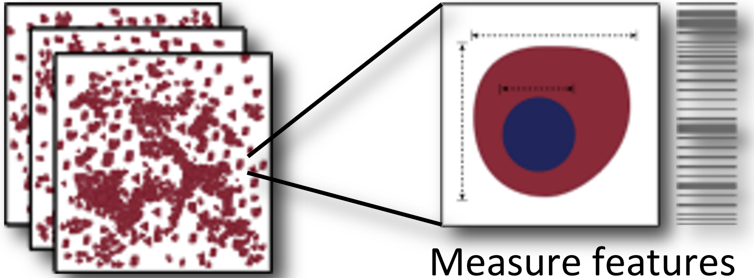
Caicedo et al., Nature Methods (2019)

Horvath lab: www.NucleAIzer.org

Phenotype classification using machine learning

(supervised classification)

Step 1: Identify & measure cells



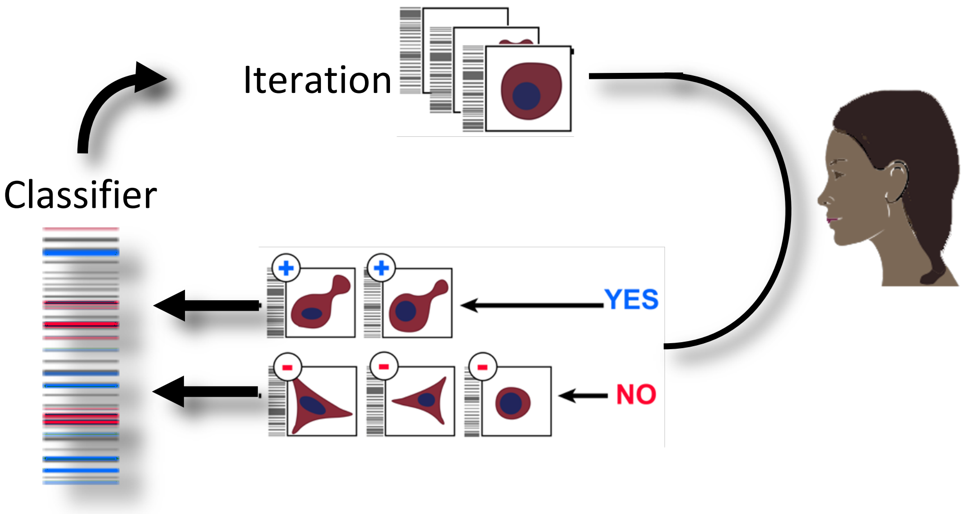
Original fields of view

Measure features for each cell

Phenotype classification using machine learning

Step 2: Train the classifier

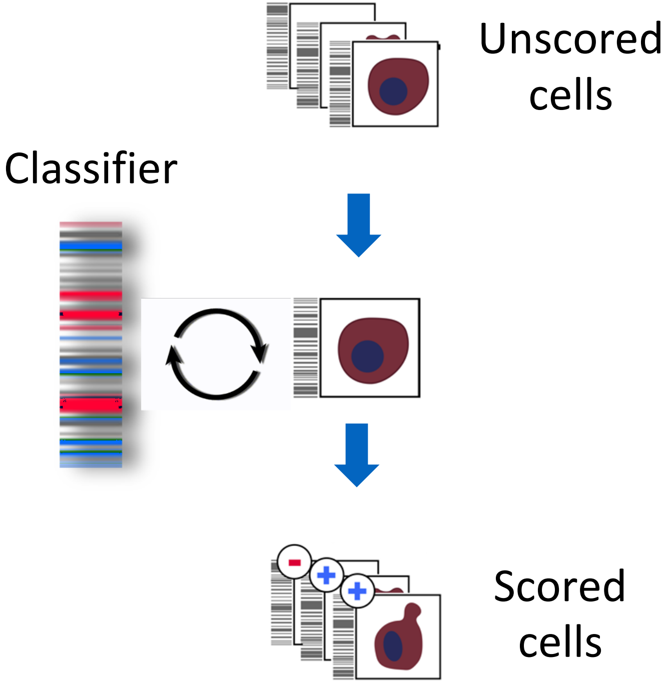
System presents a few dozen cells to the biologist for scoring



System defines classifier based on cell features

Phenotype classification using machine learning

Step 3: Score cells



User-friendly machine learning for phenotype classification

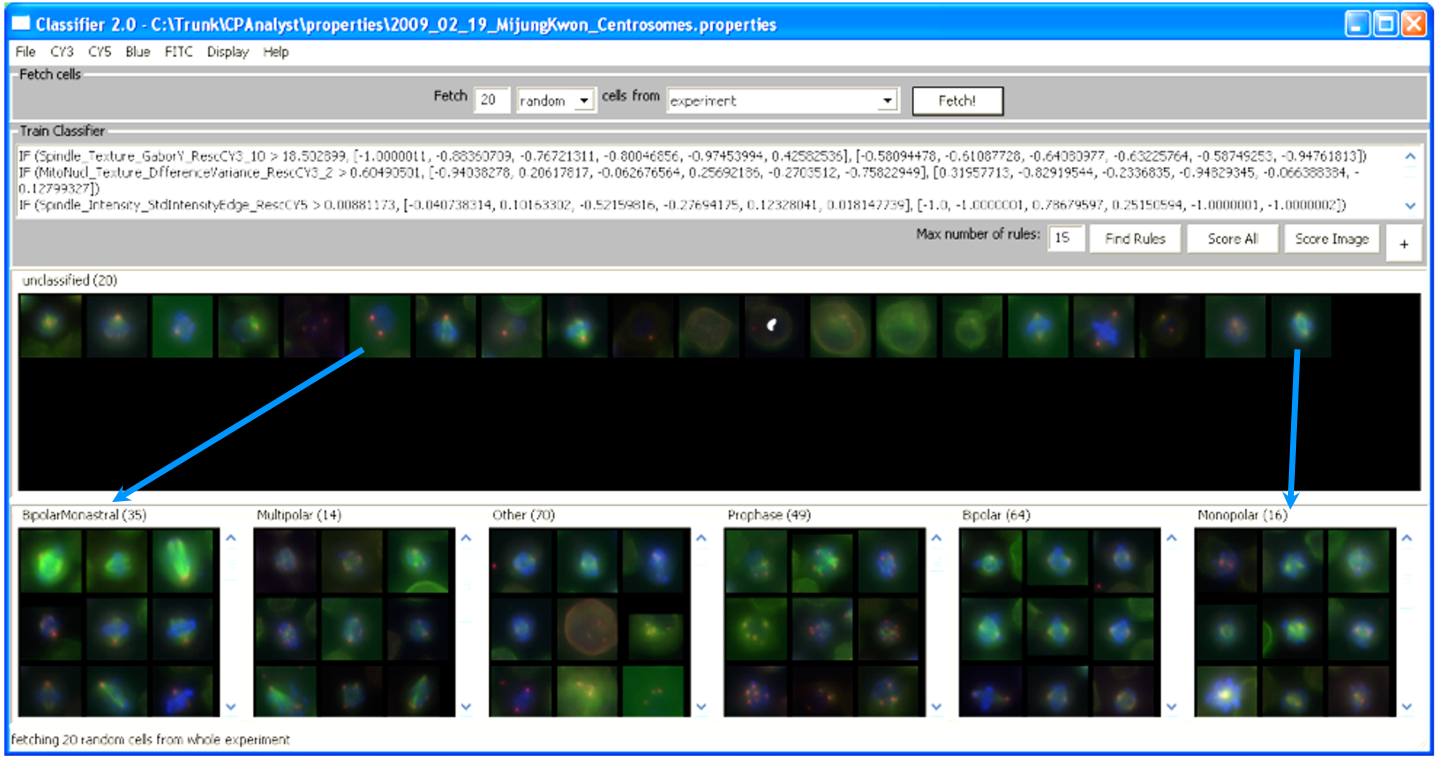
BROAD INSTITUTE IMAGING PLATFORM

CellProfiler Analyst
data exploration software

ADVANCED CELL CLASSIFIER

Coming soon!

Piximi: *deep learning, web-based app for classification*



Jones, et al. PNAS 2009; Dao et al. Bioinformatics 2016

A revolution in how we use images

Measure known phenotypes

Profile to characterize samples

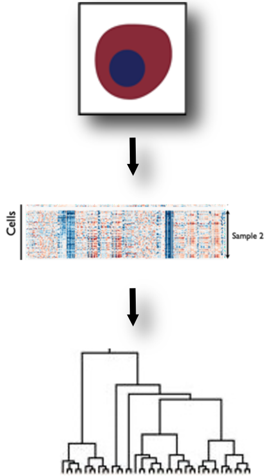
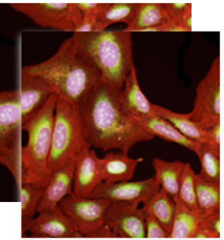
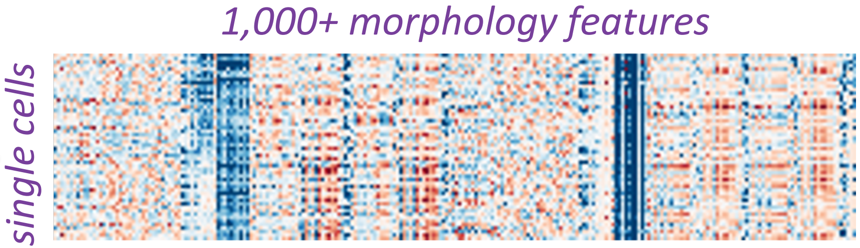
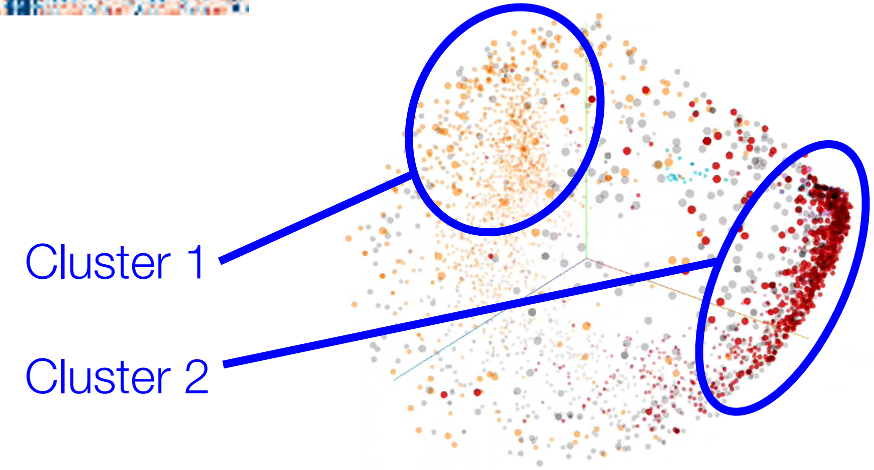


Image-based profiling

"Measure everything, ask questions later"

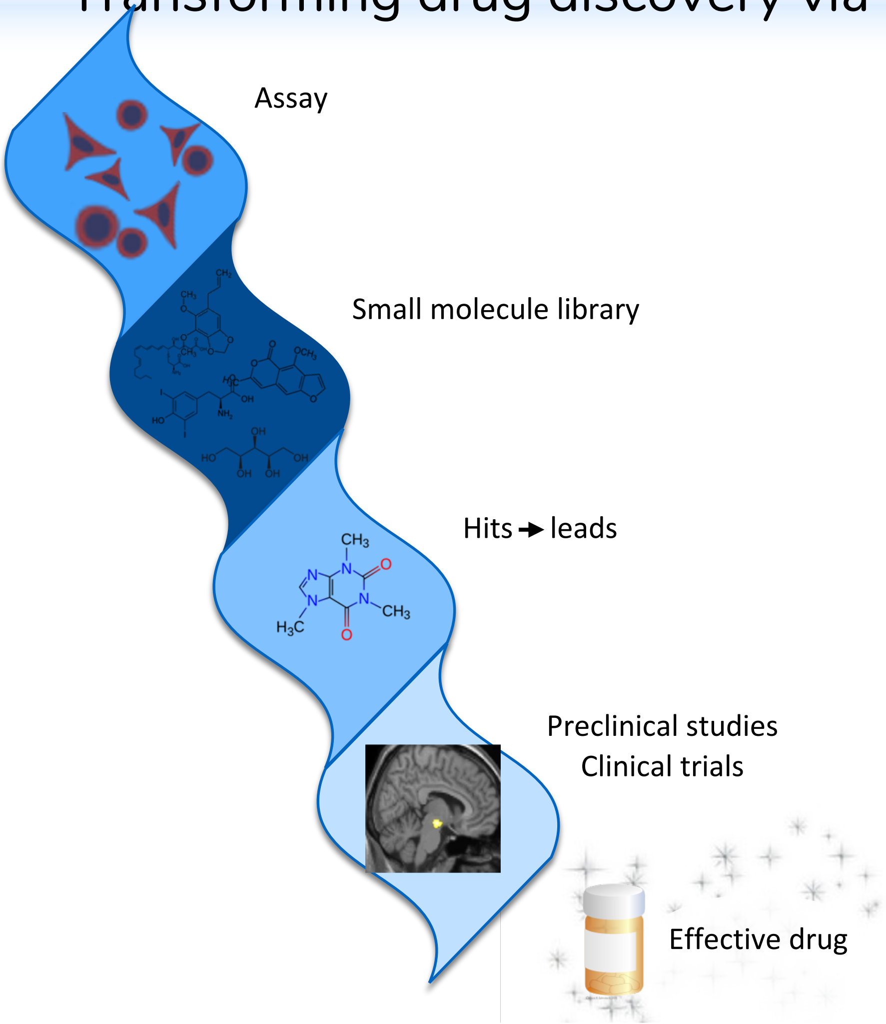






Transforming drug discovery via image-based profiling

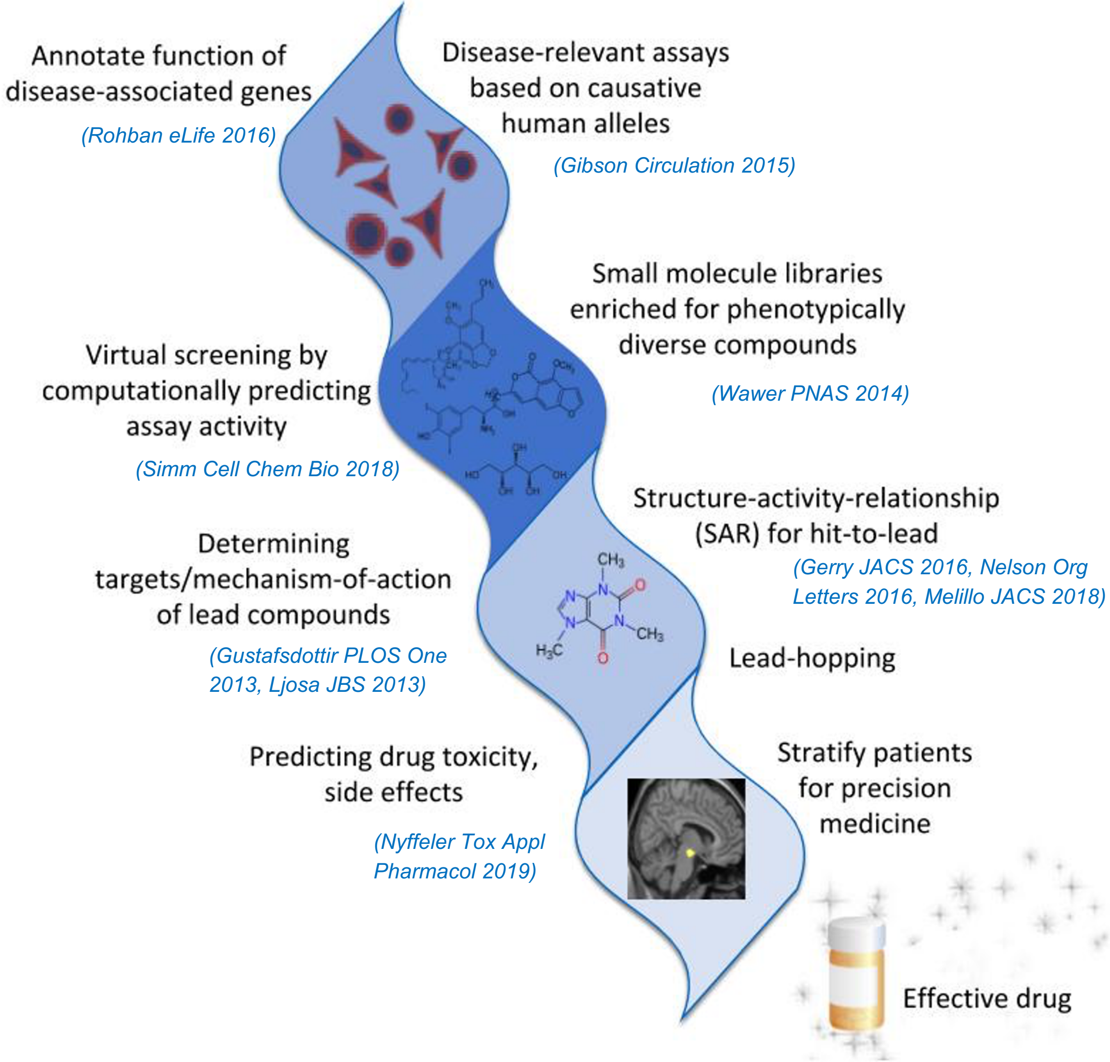
BROAD INSTITUTE IMAGING PLATFORM



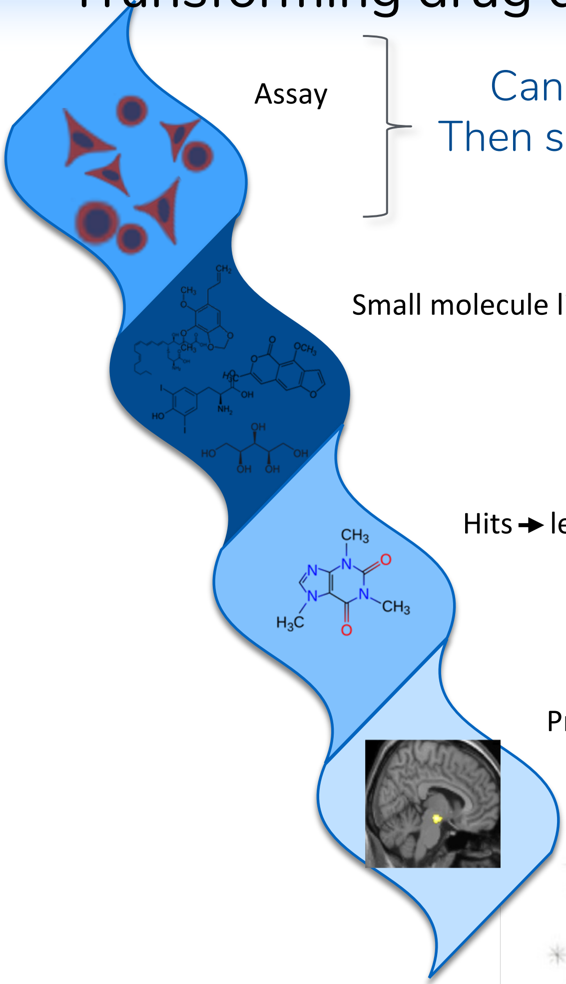
How might the rich information in images speed up the drug discovery process?

Image-based profiling: accelerating many steps of drug discovery

BROAD INSTITUTE IMAGING PLATFORM



Transforming drug discovery via image-based profiling



Assay

Can we identify signatures of disease?
Then screen drugs to reverse the signature?

Small molecule library

Hits → leads

Preclinical studies
Clinical trials

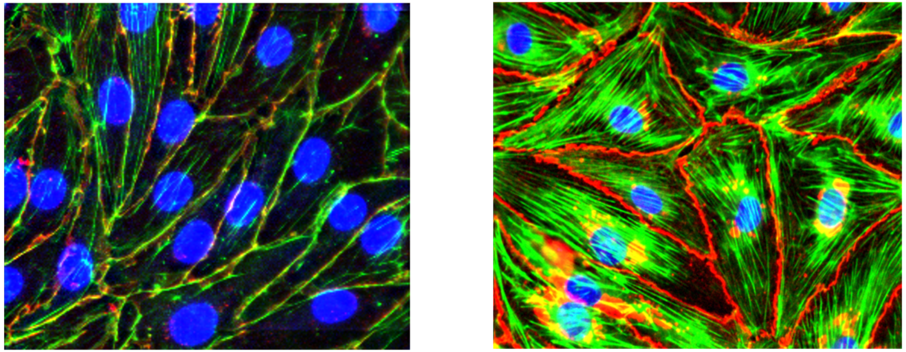
Effective drug

Image-based profiling can identify drugs for disease

+ drug?

Healthy | Disease (CCM knockdown)

VE Cadherin
Actin
DNA



Drug chosen as hits based on automated analysis outperformed those chosen by expert visual analysis.

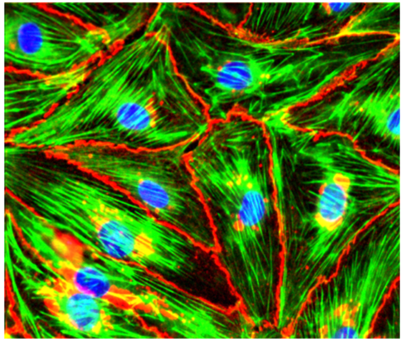
Now: 500+ disease models available for screening in parallel.

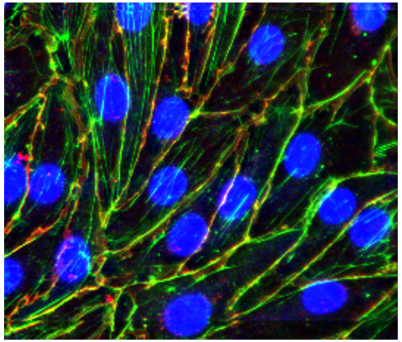
Disclosure: I serve on Recursion's Scientific Advisory Board

Gibson, et al. Circulation 2015

Impact on patients: in-progress clinical trials

(1) Cerebral cavernous malformation





University of Utah
(Gibson et al. Circulation 2015)

(2) GM2 Gangliosidosis (Tay-Sachs)

(3) Solid Tumors

(4) Neurofibromatosis Type 2

Recursion

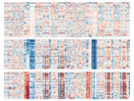
Disclosure: I serve on Recursion's Scientific Advisory Board

Disease states can be morphologically distinct

Cells from patient without mental illness

Features

Cells

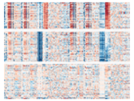


Detectable phenotypic difference?

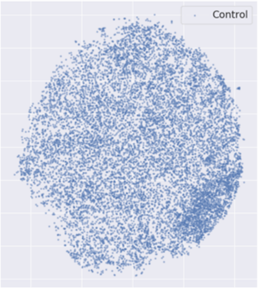
Cells from patient with a disorder

Features

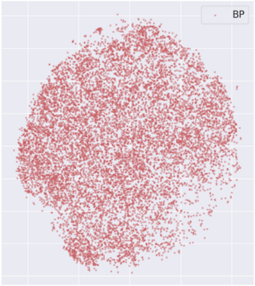
Cells

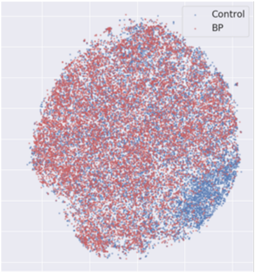


Controls



Bipolar



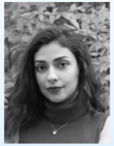


Project in progress

BROAD INSTITUTE IMAGING PLATFORM

Kyle Karhohs

Mohammad Rohban



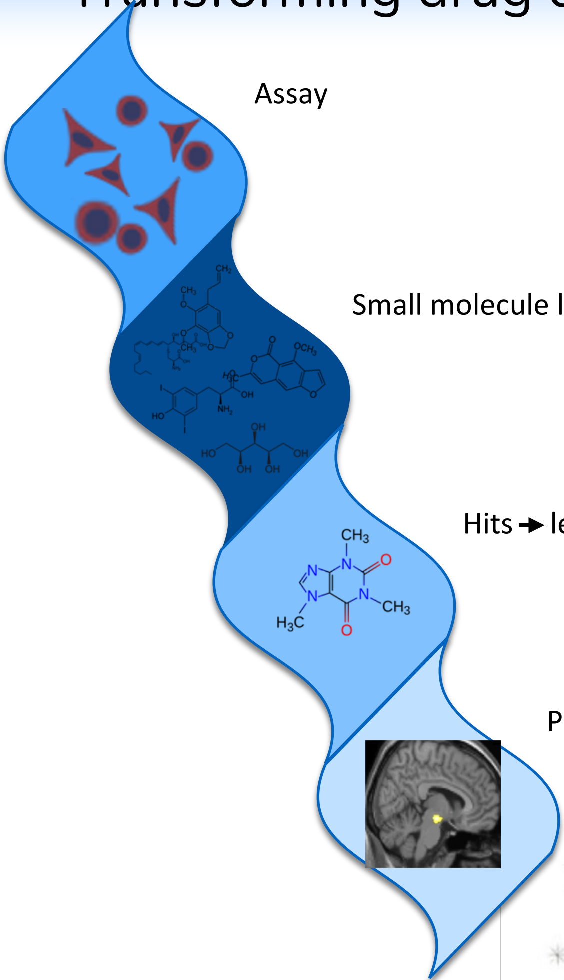
Marzieh Haghighi

Bruce Cohen, McLean Hospital

Donna McPhie

Transforming drug discovery via image-based profiling

BROAD INSTITUTE IMAGING PLATFORM



Assay

Small molecule library

Hits → leads

Preclinical studies

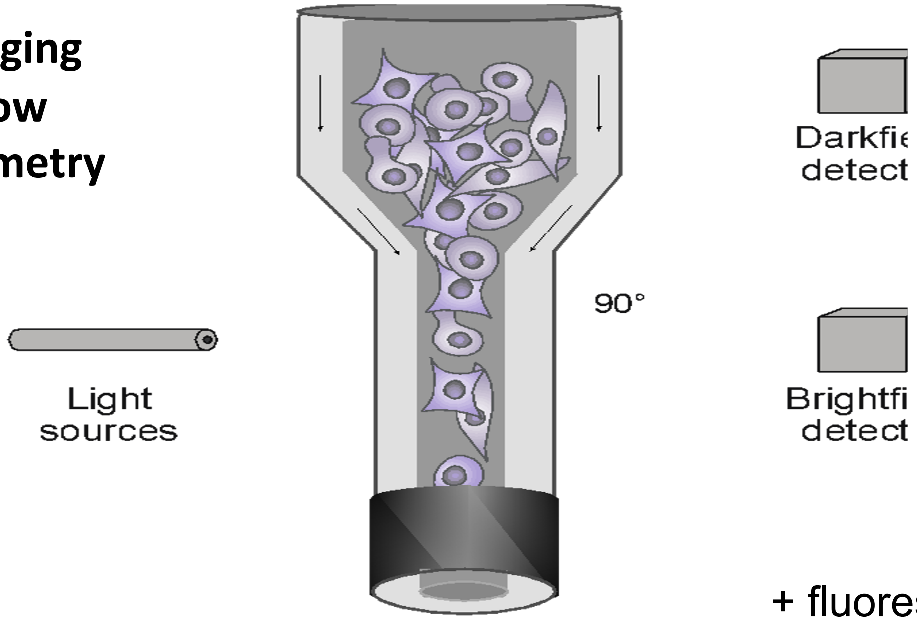
Clinical trials

Effective drug

Can we make better diagnostics?

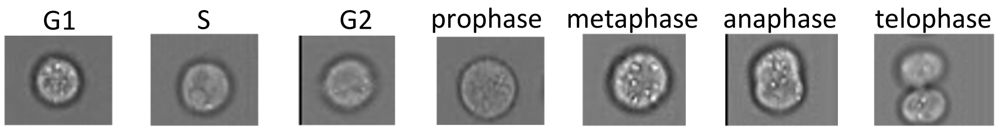
Imaging flow cytometry

Imaging flow cytometry



+ fluorescence channels

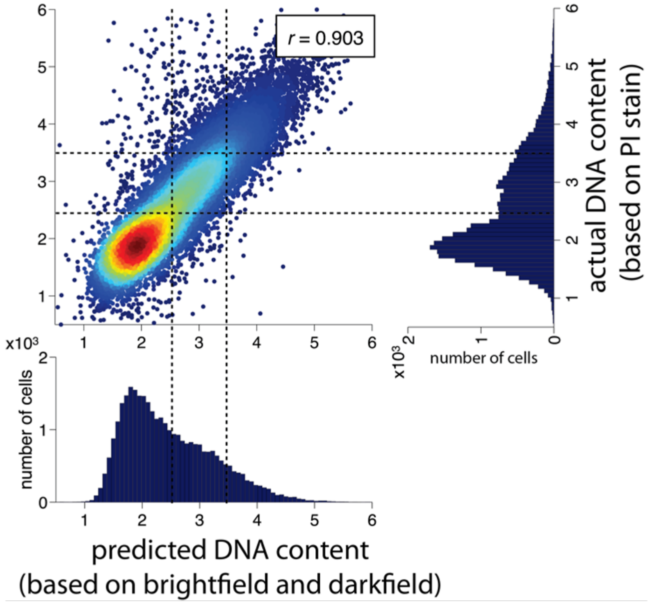
Hidden information: Classifying cell cycle stage label-free



actual phase

1	12394	1848	91	0	0	0	0
S	1578	5581	1445	0	0	0	0
i2	65	1211	7239	76	6	4	0
ro	0	9	218	368	11	0	0
:a	0	4	22	34	8	0	0
ia	0	0	5	5	2	3	0
lo	0	0	1	0	0	0	26

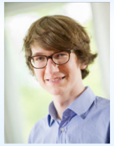
predicted phase



Blasi, et al. Nature Communications 2016
www.cellprofiler.org/imagingflowcytometry

BROAD INSTITUTE IMAGING PLATFORM


Holger Hennig


Thomas Blasi


Paul Rees, Swansea Univ.


Fabian Theis, Helmholtz Zentrum München


Andrew Filby, Newcastle Univ.

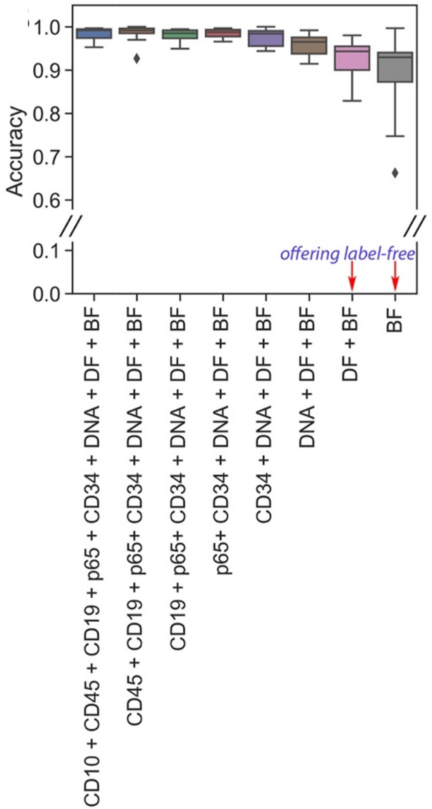

Huw Summers, Swansea Univ.

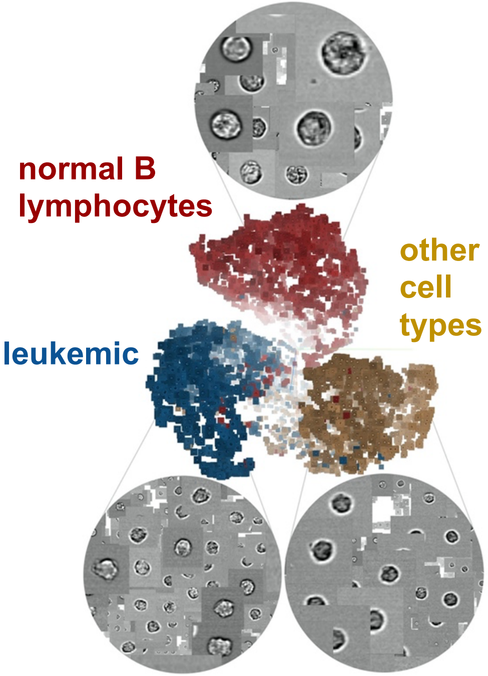
Detecting leukemia (ALL) in blood samples, label-free

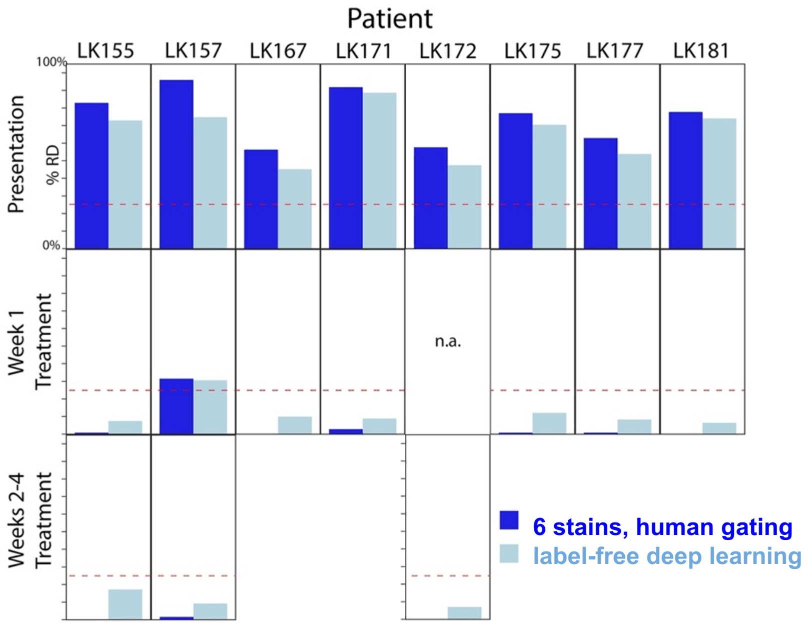
BROAD INSTITUTE IMAGING PLATFORM

Minh Doan *Marian Case Dino Masic Holger Hennig Claire McQuin Juan Caicedo Shantanu Singh Allen Goodman Olaf Wolkenhauer Huw D. Summers David Jamieson Frederik W van Delft Andrew Filby Paul Rees Julie Irving*

Can we replace a 4 to 10-antibody flow cytometry assay with images captured label-free?







Doan et al. Cytometry 2020

Lessons from implementing AI

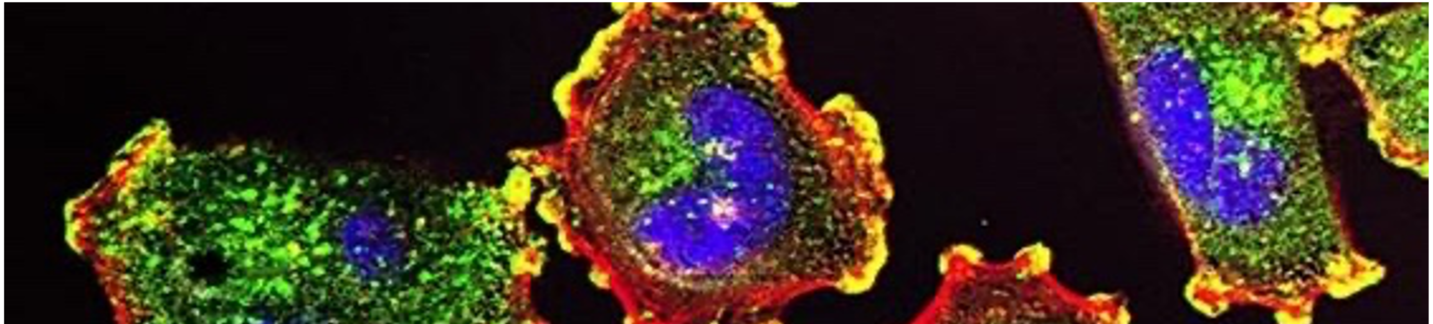
BROAD INSTITUTE IMAGING PLATFORM

What's needed to use AI well?

- **Identifying** situations appropriate for deep learning
- **Lots of examples** of right & wrong answers are needed for training
    - (for example, collecting manually-set parameters from users)
- Examples must come from diverse sources and conditions that **reflect reality** (what will be seen in the wild)
- Focus on **user needs** and workflow
- **Imagination**: willingness to trust AI can see more than we can (with controls!)

What will AI increase demand for?

- Instruments & reagents that produce **consistent** data are highly valued
- Large-scale, high-throughput experiments fuel advanced AI applications



Gratitude

Lab members

Hamdah Abbasi
Adeniyi Adeboye
Niranj Chandrasekaran
Yuen Ler Chow
Beth Cimini
Allen Goodman
Marzieh Haghighi
Nasim Jamali

Becki Ledford
Alice Lucas
Saketh Mynampati
Pearl Ryder
Shantanu Singh
David Stirling
Hillary Tsang
Gregory Way
Erin Weisbart

Many thanks to our many collaborators!

Recent major funding for this work provided by:

- NIH NIGMS P41 GM135019
- NIH NIGMS R35 GM122547 MIRA
- NIH NIGMS R01 GM089652
- Chan Zuckerberg Institute
- NSF CAREER award
- NSF/BBSRC Bio award (Carpenter/Rees)

Download PDF of slides here: http://broad.io/CarpenterSlides

Give feedback on this talk: http://broad.io/CarpenterFeedback

S.D.G.