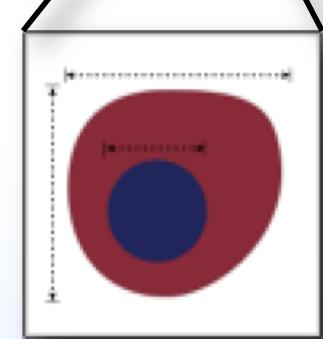


Open-source bioimage analysis software to accelerate drug discovery

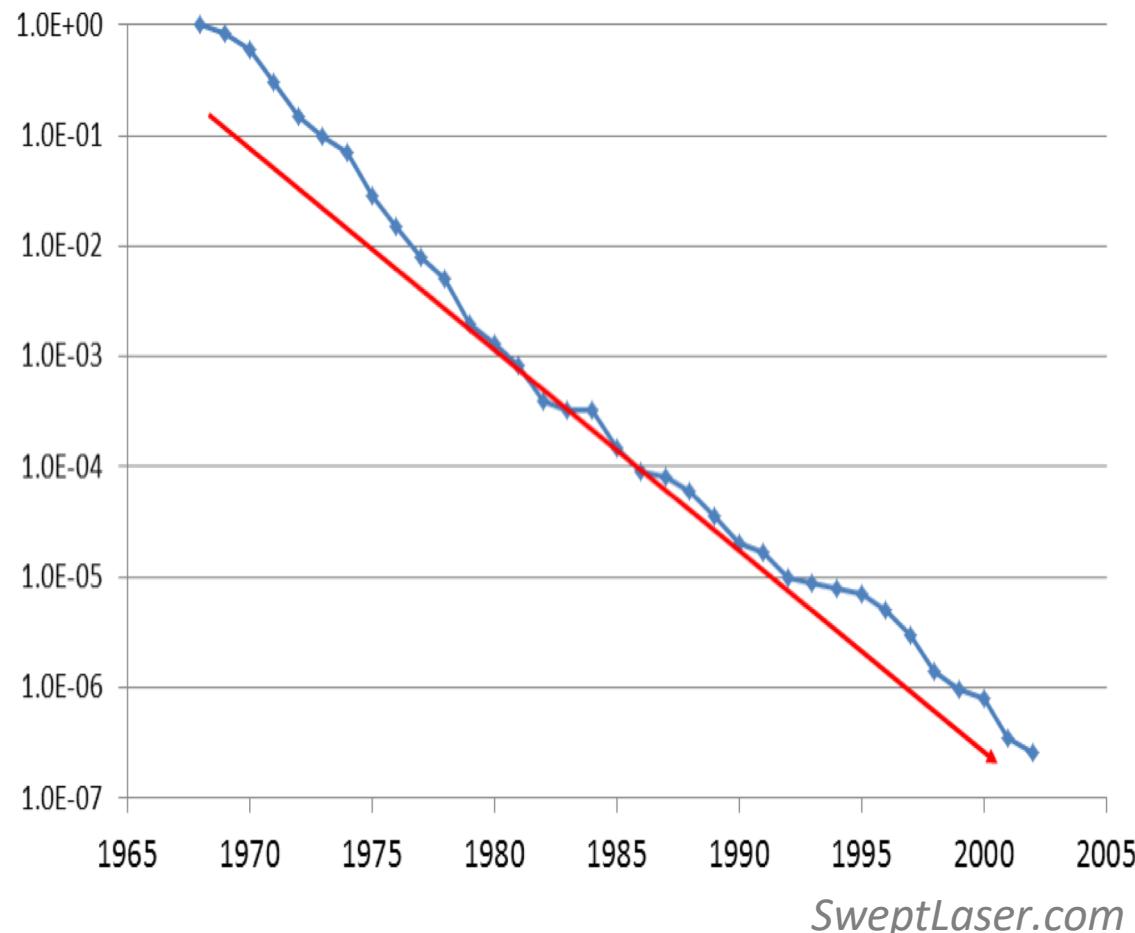


Anne E. Carpenter, PhD

Tale of two industries

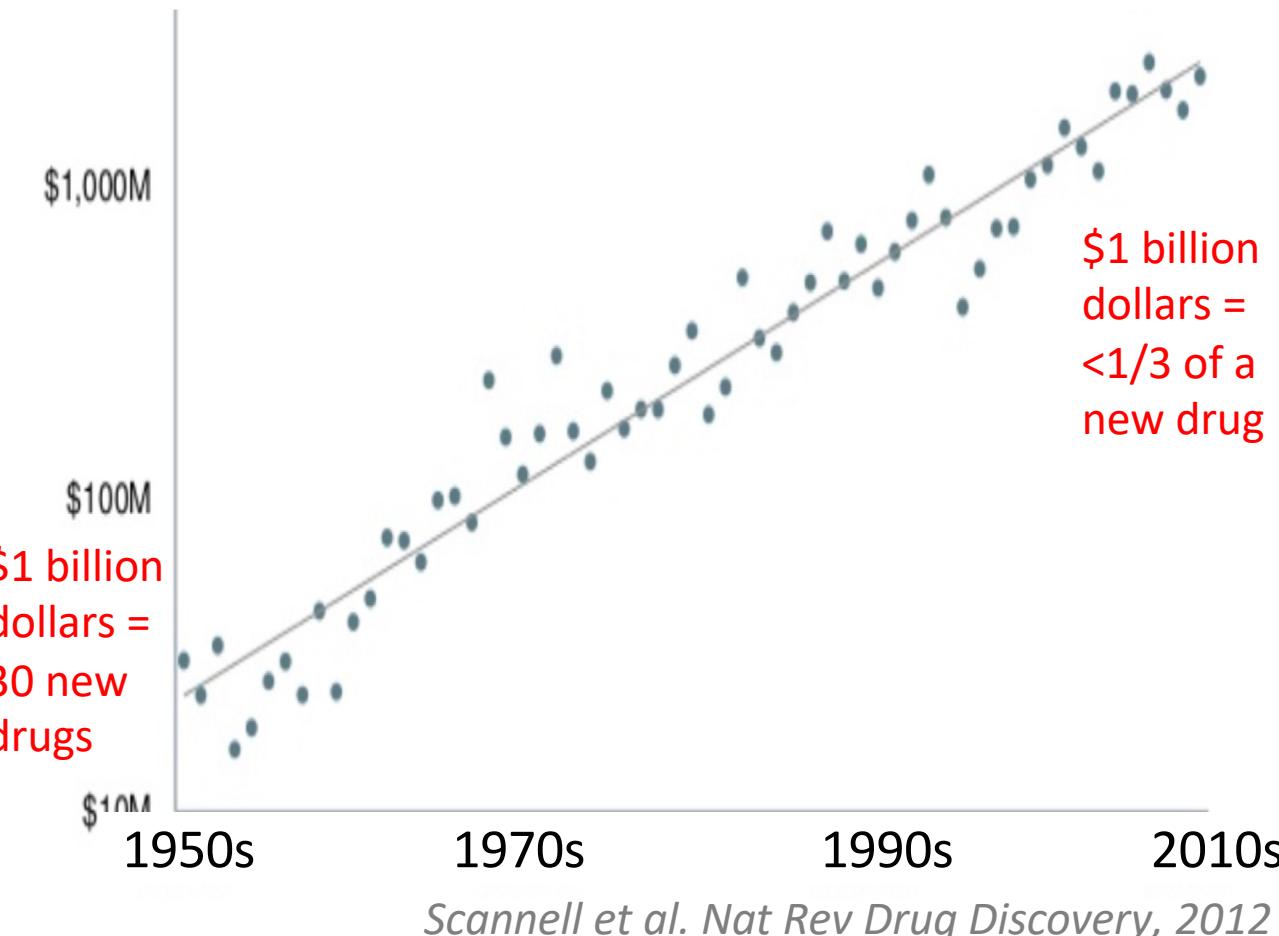
Moore's Law

Compute gets cheaper and cheaper

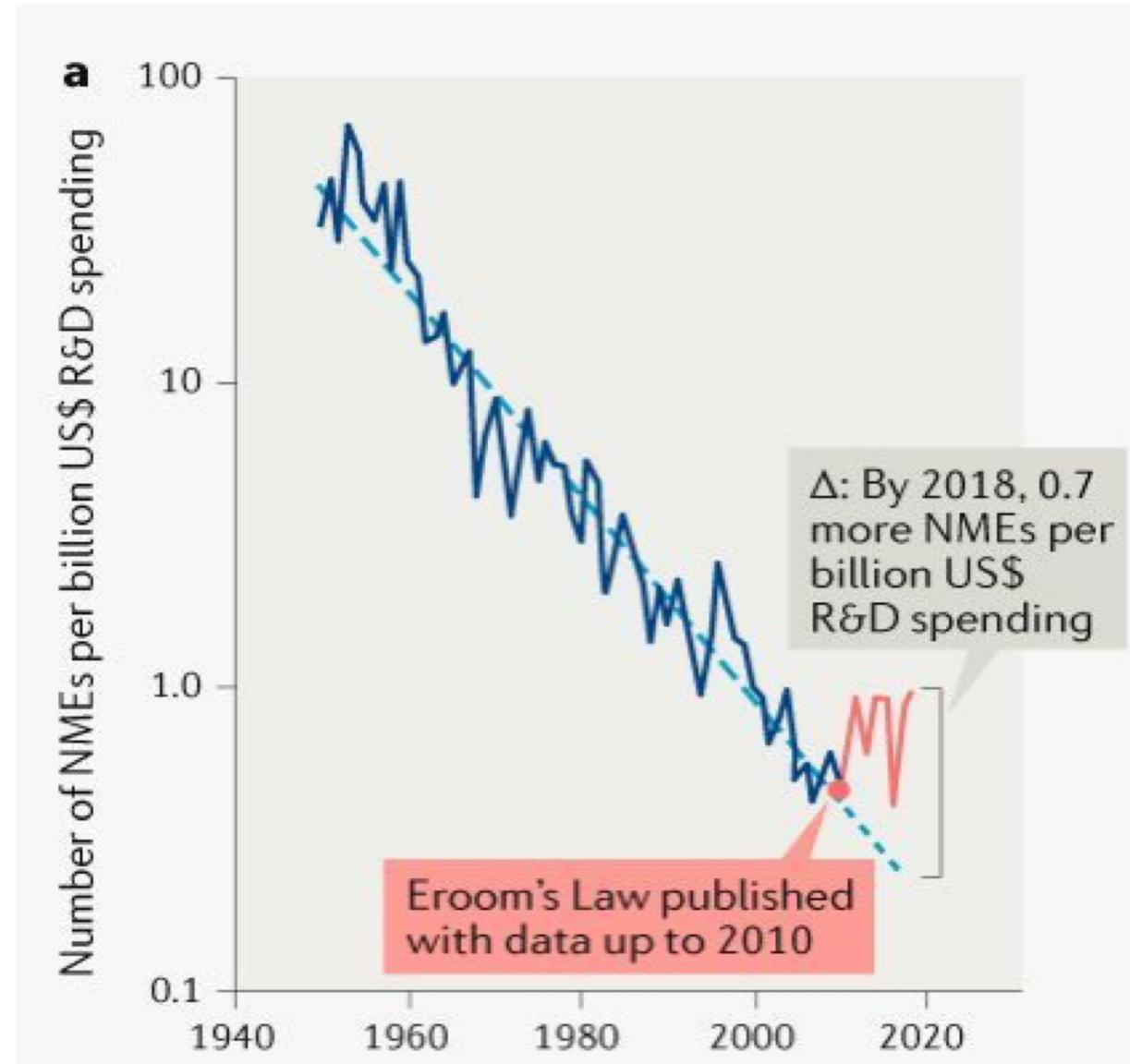


Eroom's Law

Discovering new medicines gets more and more expensive



Eroom's law of the pharma industry

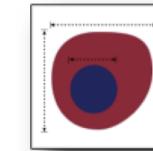


Ringel et al. Nat Rev Drug Discovery 2020

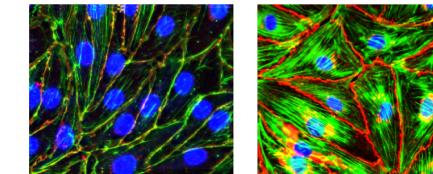
How can open bioimage analysis software power drug discovery?



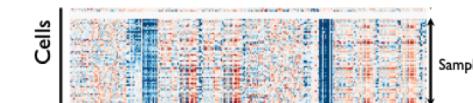
Bioimage analysis to measure cells:
CellProfiler



Machine learning to classify cells:
CellProfiler Analyst & Piximi



Libraries to profile cells:
cytominer and deepprofiler



Images contain a wealth of information

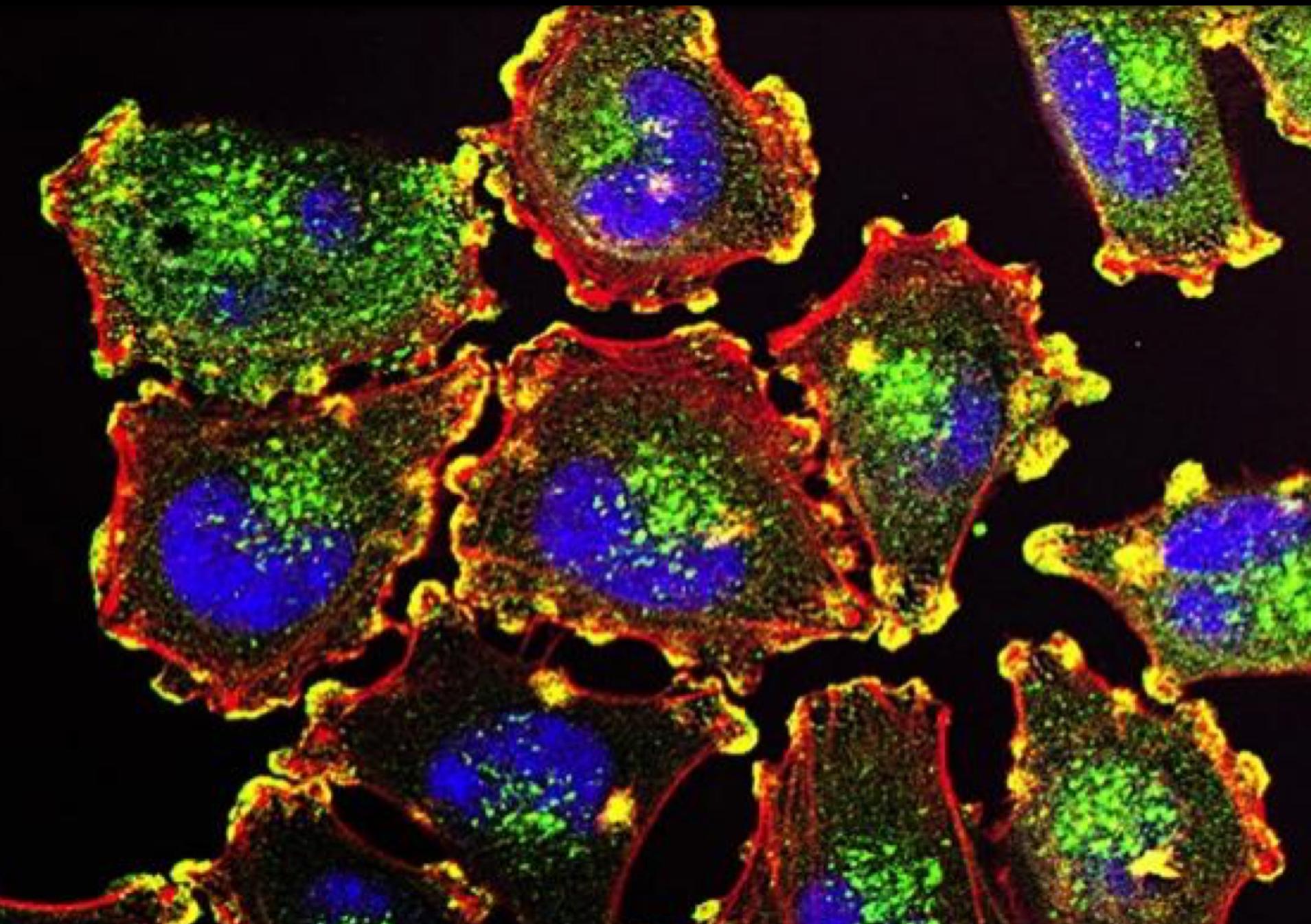
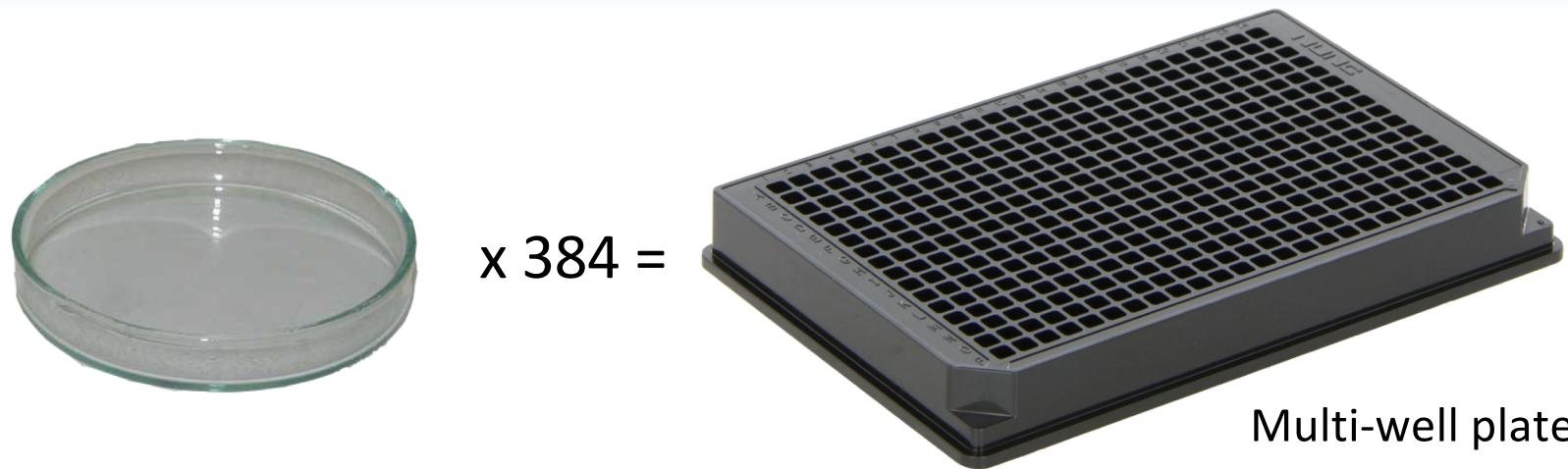


Image credit:
Wikimedia

Discovering drugs in high-throughput?



Multi-well plate

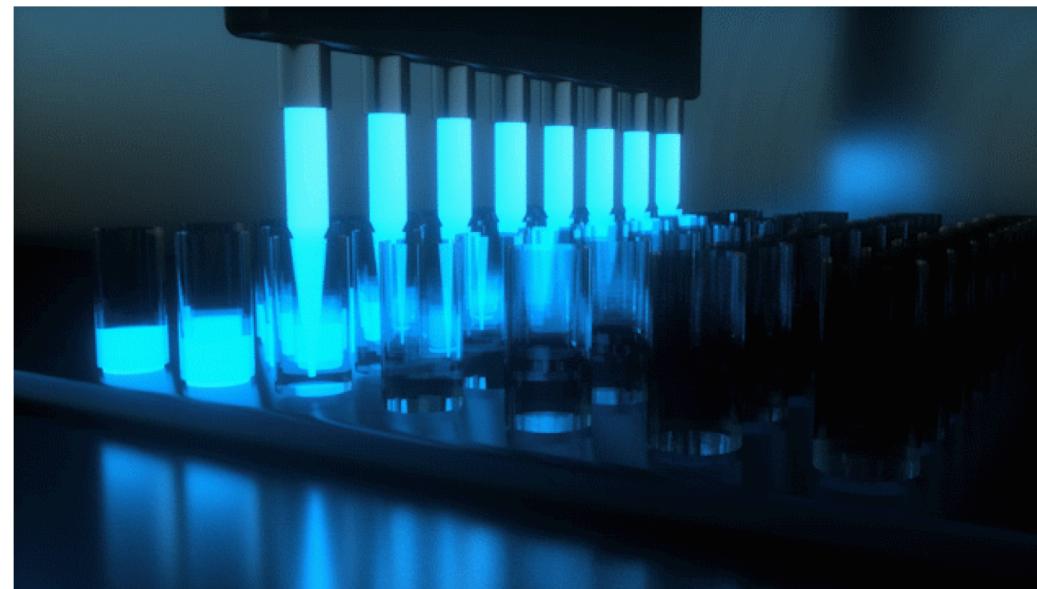


Image: Nalgene; video: Chemistry
World

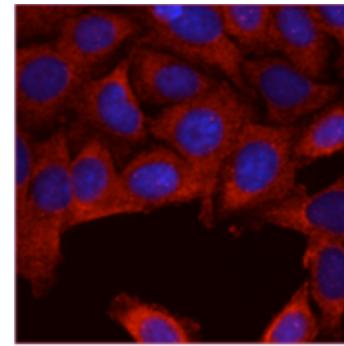
Extracting features from images

Sample prep



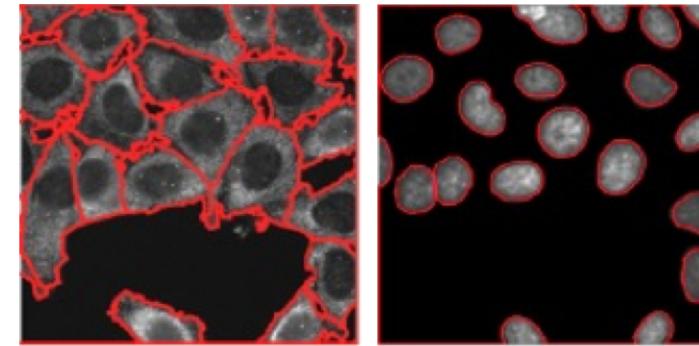
100's of 384-well plates

Microscopy



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

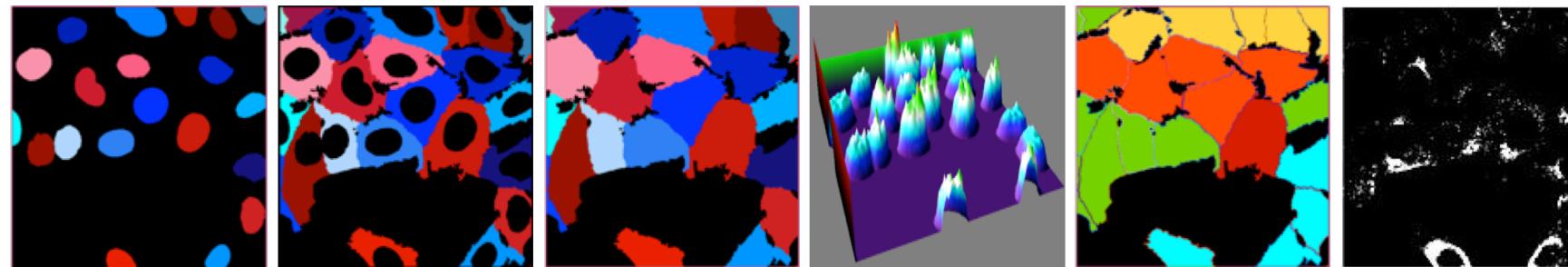
Image analysis:
Segmentation



~500 cells/ image

Measure everything

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

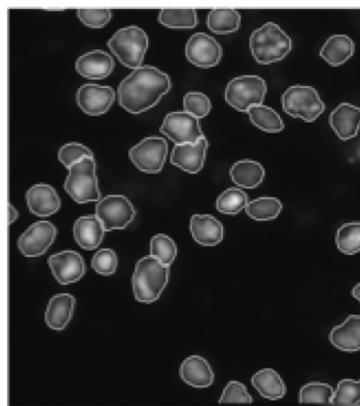


~1500 features / cell

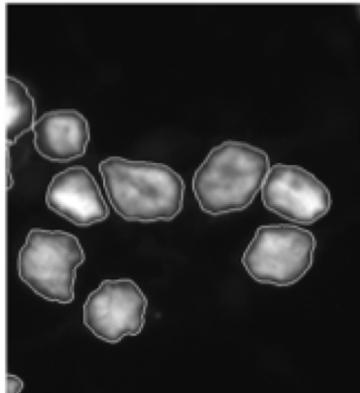
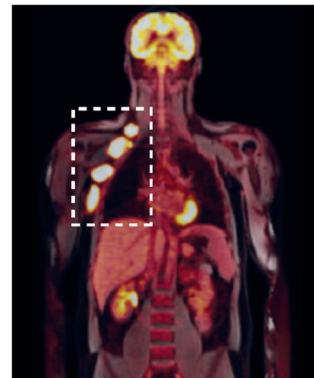
Impact on patients

Successful clinical trials

Myelofibrosis



Leukemias & Lymphomas

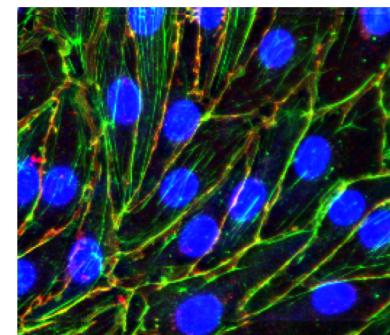
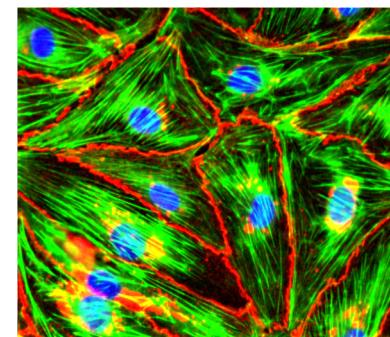


Crispino lab
(Gangat et al. *Clin Cancer Research* 2019)

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

In-progress clinical trials

(1) Cerebral cavernous malformation

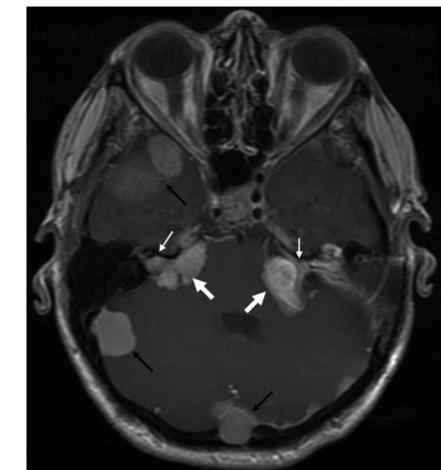


University of Utah
(Gibson et al.
Circulation 2015)

(2) GM2 Gangliosidosis
(Tay-Sachs)

(3) Solid Tumors

(4) Neurofibromatosis Type 2



Recursion



Disclosure: I serve on Recursion's Scientific Advisory Board

Behind the scenes: CellProfiler present, past, and future

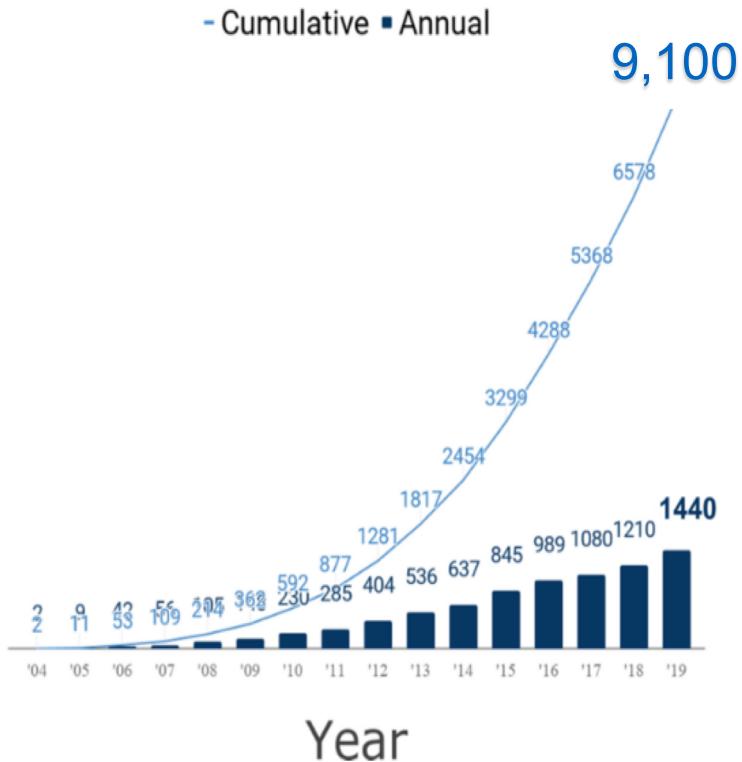
CellProfiler: Open-source image analysis software



CellProfiler™
cell image analysis software

- Free and open-source; Windows, Mac, Linux
- Cited in **2,000+** papers per year, **9,000+** total
- Used in **7/10** top pharma companies
- In the **Top 10** most popular papers in Genome Biology
- Ranked **most flexible** and **usable** in independent analysis (*Wiesmann et al.*)
- Written in python
- Generally created/supported by <1 person
- Developed by those intensively using it

Publications citing CellProfiler



Online Q&A! <http://forum.image.sc>



Ray Jones
2007-2008



Anne Carpenter
2003-2006



Allen Goodman
2015-2018



Lee Kamentsky
2009-2015



Beth Cimini
2017-2020



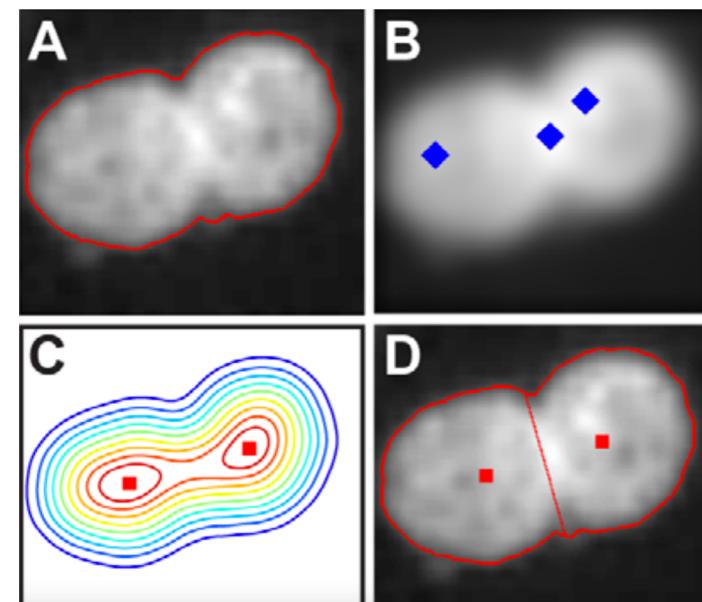
Claire McQuin
2017-2018



David Stirling
2019-2020

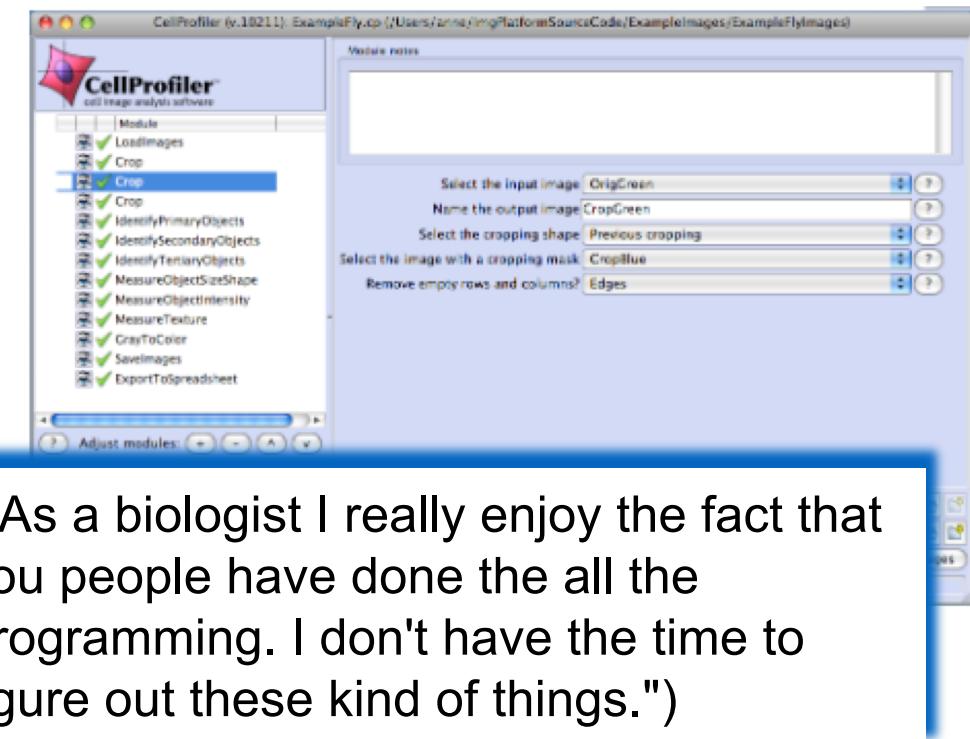
CellProfiler goals

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

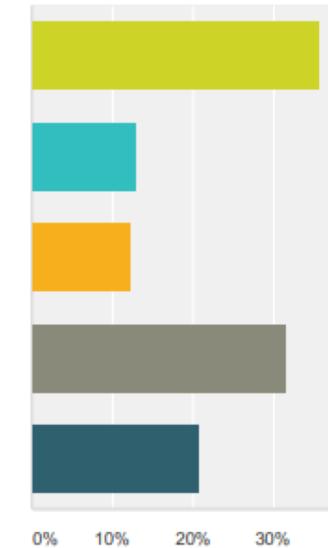


CellProfiler history

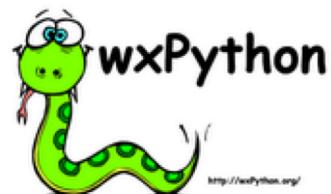
- 2003-5: wrote 1.0: MATLAB (pre-numpy!)
- 2005-6: public release/journal paper
- 2008-2010: wrote 2.0 in python:
72k lines of MATLAB (+ gui files) ->
62k lines of Python + 36k lines of tests
 - Why not MATLAB?
 - * waiting for bug fixes
 - * poor installation experience
 - * cost
 - Why not java?
 - * luck!
- 2017: 3.0 released (3D)
- 2020: 4.0 release (soon!)



Don't care (don't program)
Python
Java
No preference
Other



Thank you! Projects crucial for CellProfiler



PyInstaller
#!/bin/env python
import as
import sys



scikit-image
image processing in python



TensorFlow



PIL +



h5py

+



The HDF Group



virtualenv

Keeping software alive

- Make a choice about your target audience & serve them well:

Scientific community or just you and your lab?

Computationalists vs domain experts?

Contribute to existing tools/libraries

- Leverage your time

endless supply of requests for help/improvements

make your efforts public and searchable

“How to support open-source software and stay sane”

(Nature, Anna Nowogrodzki, July 2019)

“Case Study: Curing Disease with NumFOCUS Tools”

- Find funding

Track usage and survey users

(launches, executions, downloads, Q&A forum, forks)

“Challenges in funding and developing genomic software: roots and remedies”

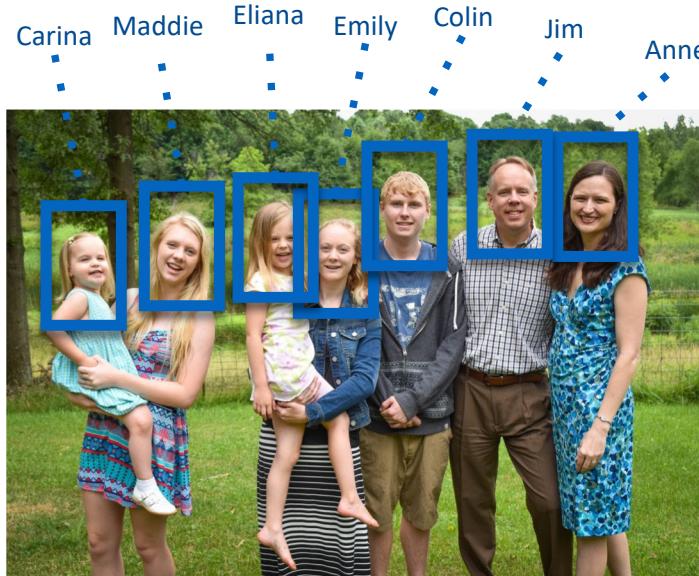
(Genome Biology, Adam Siepel, 2019)

“It’s sort of the difference between having insurance and having a GoFundMe when grandma goes to the hospital... it’s just not a nice way to live.” – Anne Carpenter

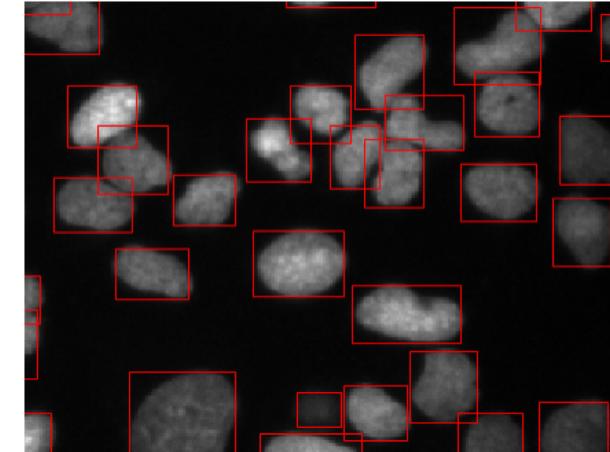
Chan
Zuckerberg
Initiative 


OPEN CODE = BETTER SCIENCE

Future: bringing deep learning to biologists

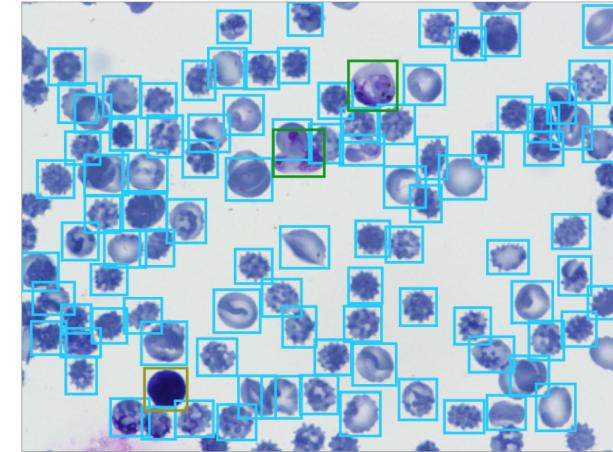


Finding objects
(e.g. nuclei)

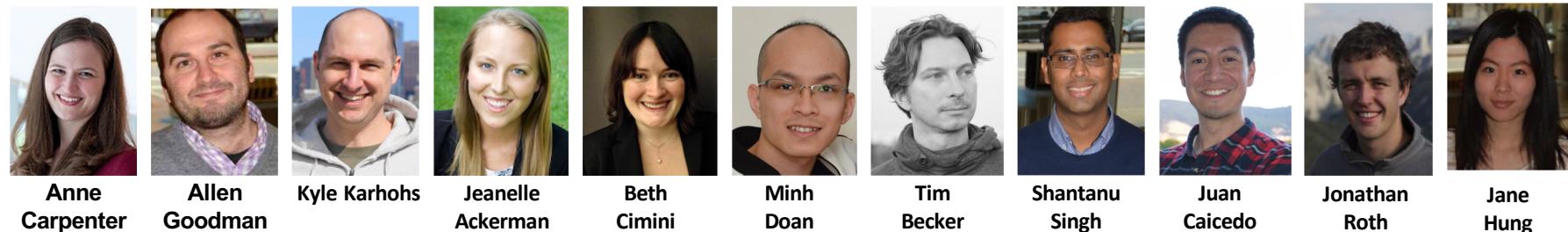


Caicedo et al., Cytometry 2019
Caicedo et al., Nature Methods 2019
Data Science Bowl 2018

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



Hung et al., CVPR workshop 2017

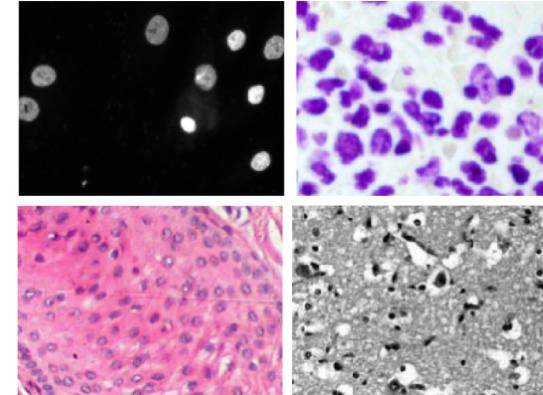


Deep learning excels at segmentation

2018 Data Science Bowl: Towards a Universal Nucleus Finder



kaggle



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

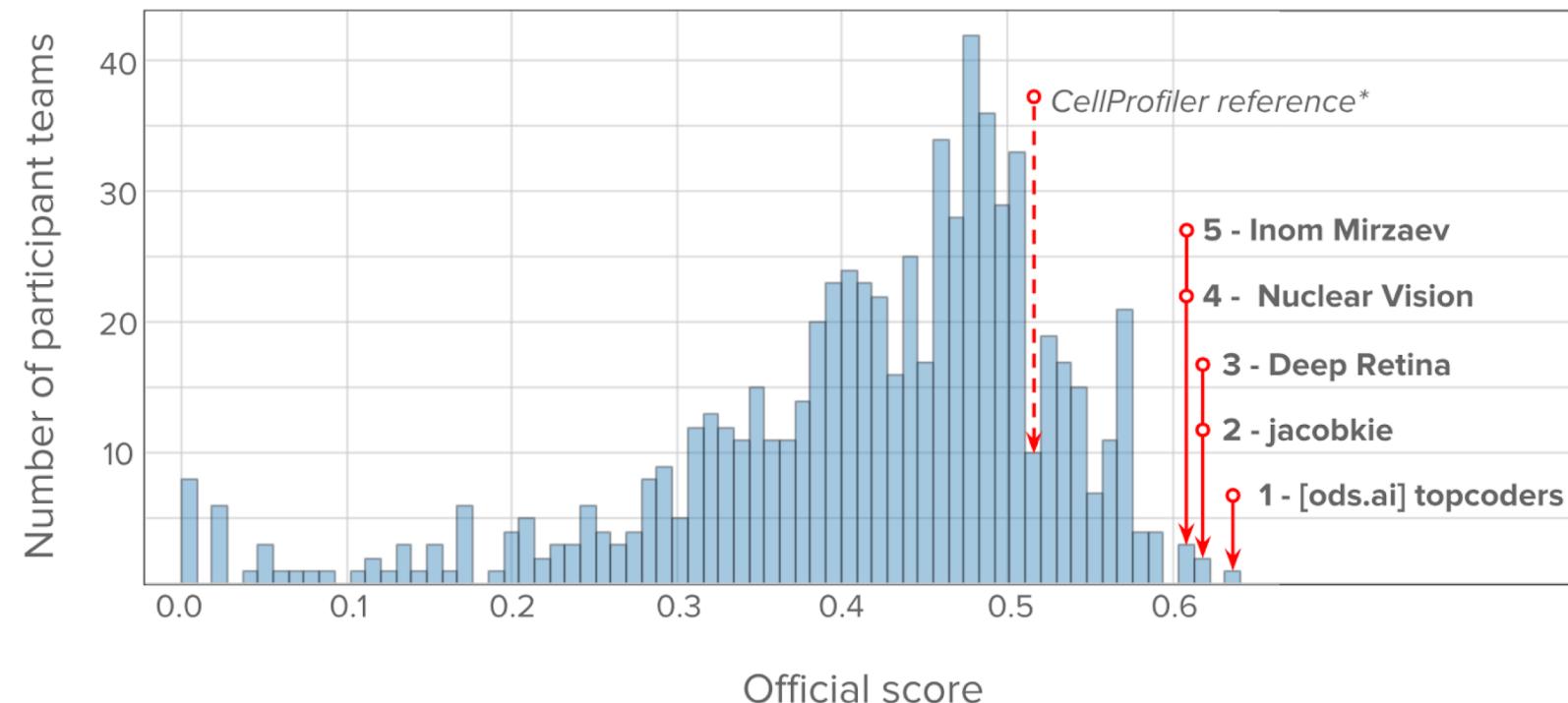
Caicedo et al., Nature Methods (2019)



Deep learning excels at segmentation

2018 Data Science Bowl: Towards a Universal Nucleus Finder

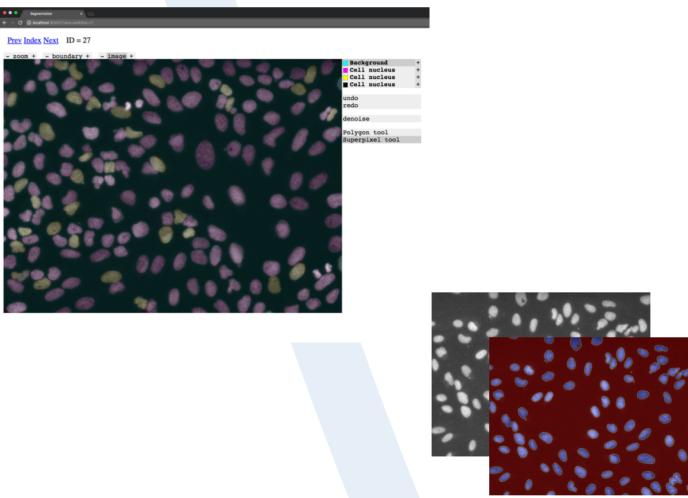
Distribution of scores in second-stage evaluation



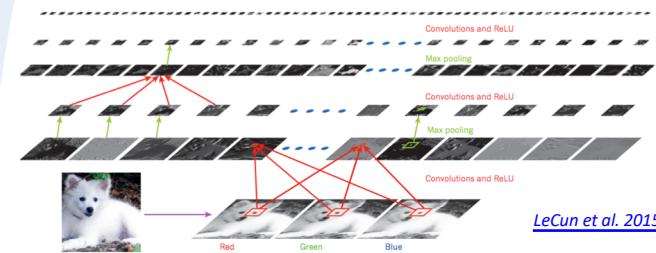
Horvath lab: www.NucleAlzer.org
Github: Keras-RCNN

Bringing deep learning to biologists: it takes a community

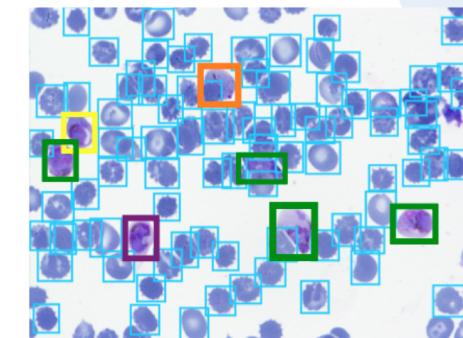
Annotation tools



Strategies and software to train models



Software to fine-tune a model



Public repository of ground-truth annotated images

Pipeline/data-flow vs. end-to-end?

Public repository of trained models

Software to run model on a set of images, user-friendly and at scale

Q & A





Kevin Eliceiri | ImageJ | Univ. Wisconsin & Anne Carpenter | CellProfiler | Broad Institute

To serve the cell biology community's growing need for sophisticated software,
we will catalyze:

- (1) deep learning based image processing
- (2) accessibility of image processing algorithms for biologists

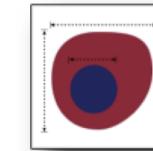
New! The Scientific Community Image Forum: <https://forum.image.sc>

Hosting in 2021! Images to Knowledge (I2K) conference Sept 23-24, 2021 Madison, Wisconsin

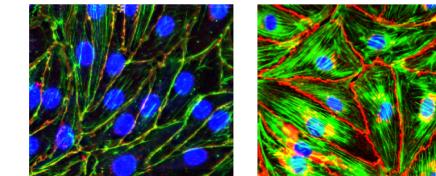
How can open bioimage analysis software power drug discovery?



Bioimage analysis to measure cells:
CellProfiler



Machine learning to classify cells:
CellProfiler Analyst & Piximi

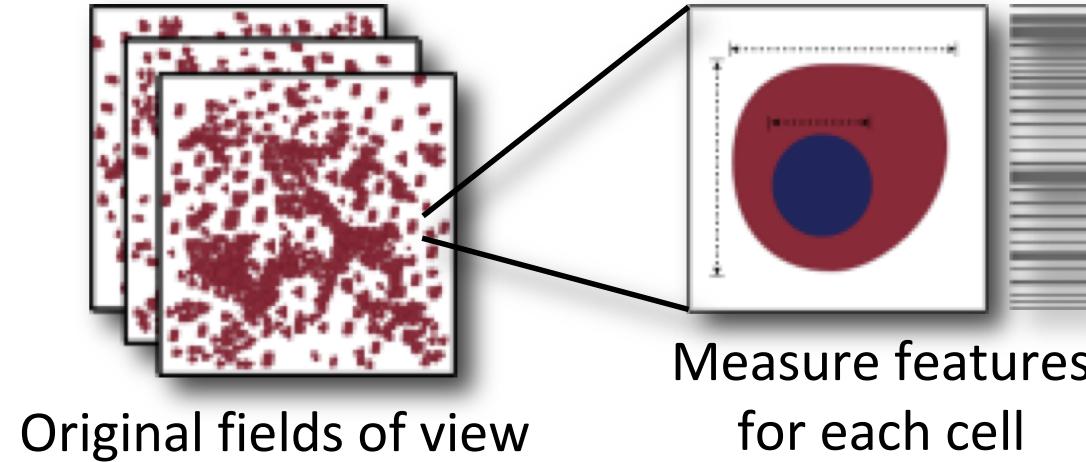


Libraries to profile cells:
cytominer and deepprofiler



Phenotype classification using machine learning (supervised classification)

**Step 1:
Identify &
measure
ROIs**



Phenotype classification using machine learning

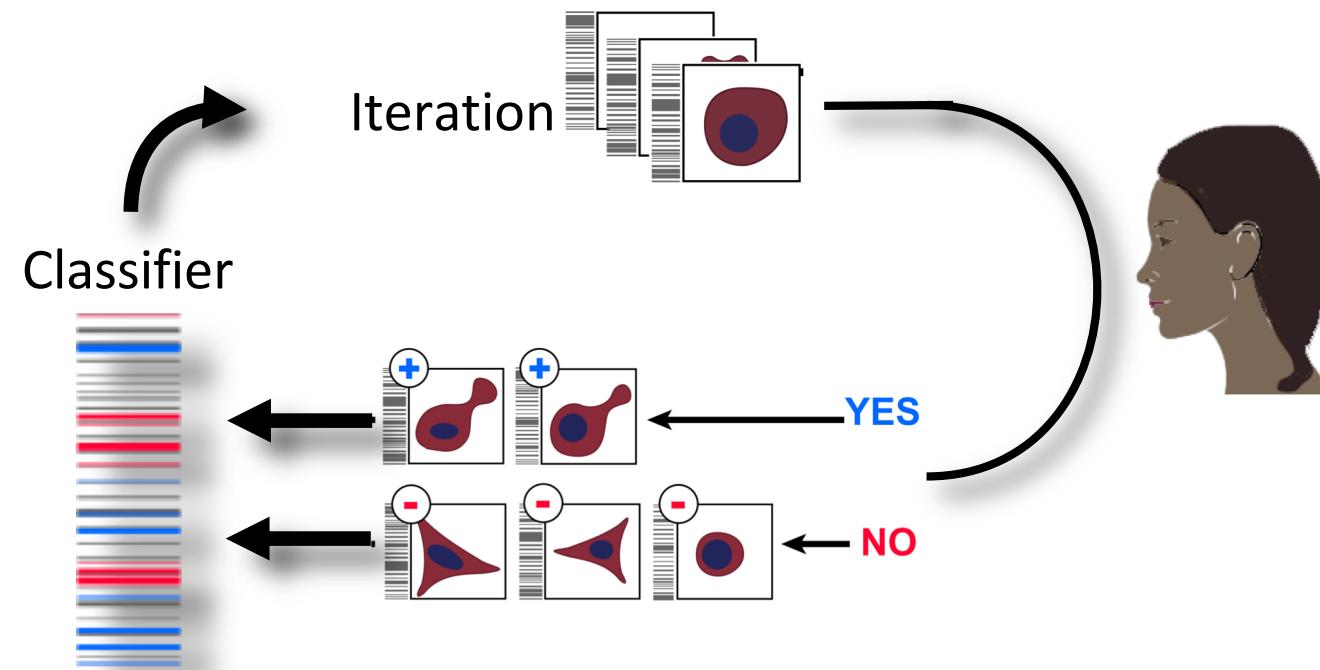


CellProfiler Analyst
data exploration software

Step 2: Train the classifier

(supervised classification)

System presents a few dozen cells
to the biologist for scoring



System defines classifier based on cell features



Adam Fraser



Ray Jones



Jane Hung

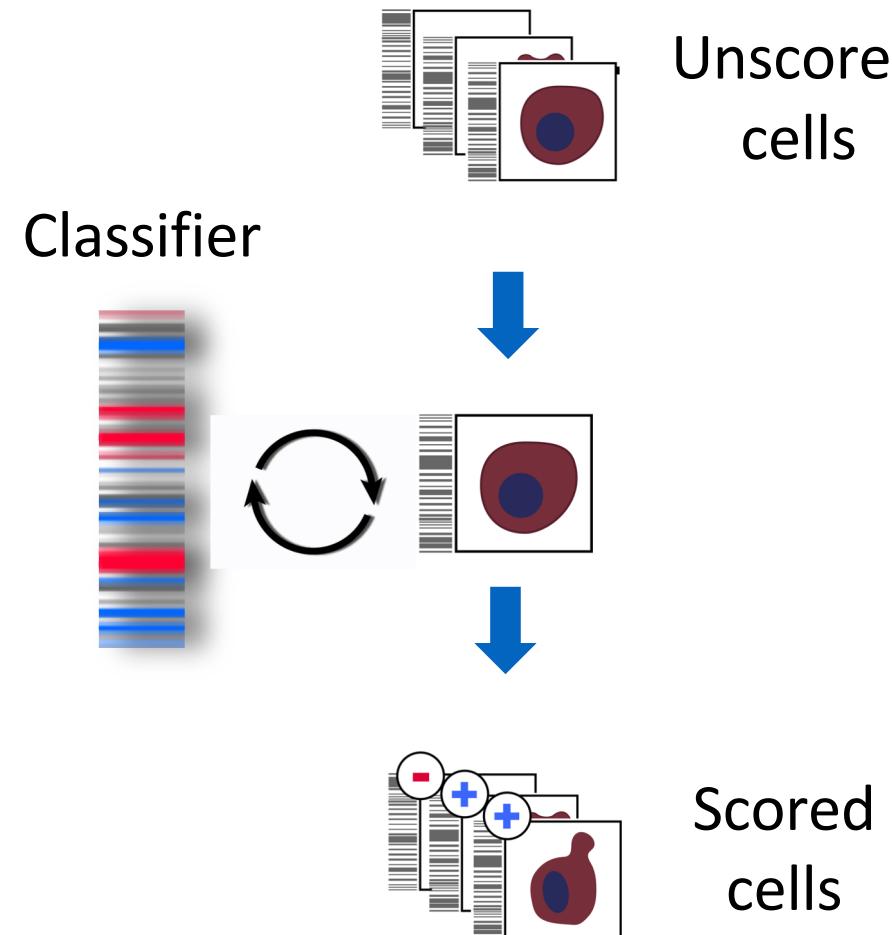


David Dao

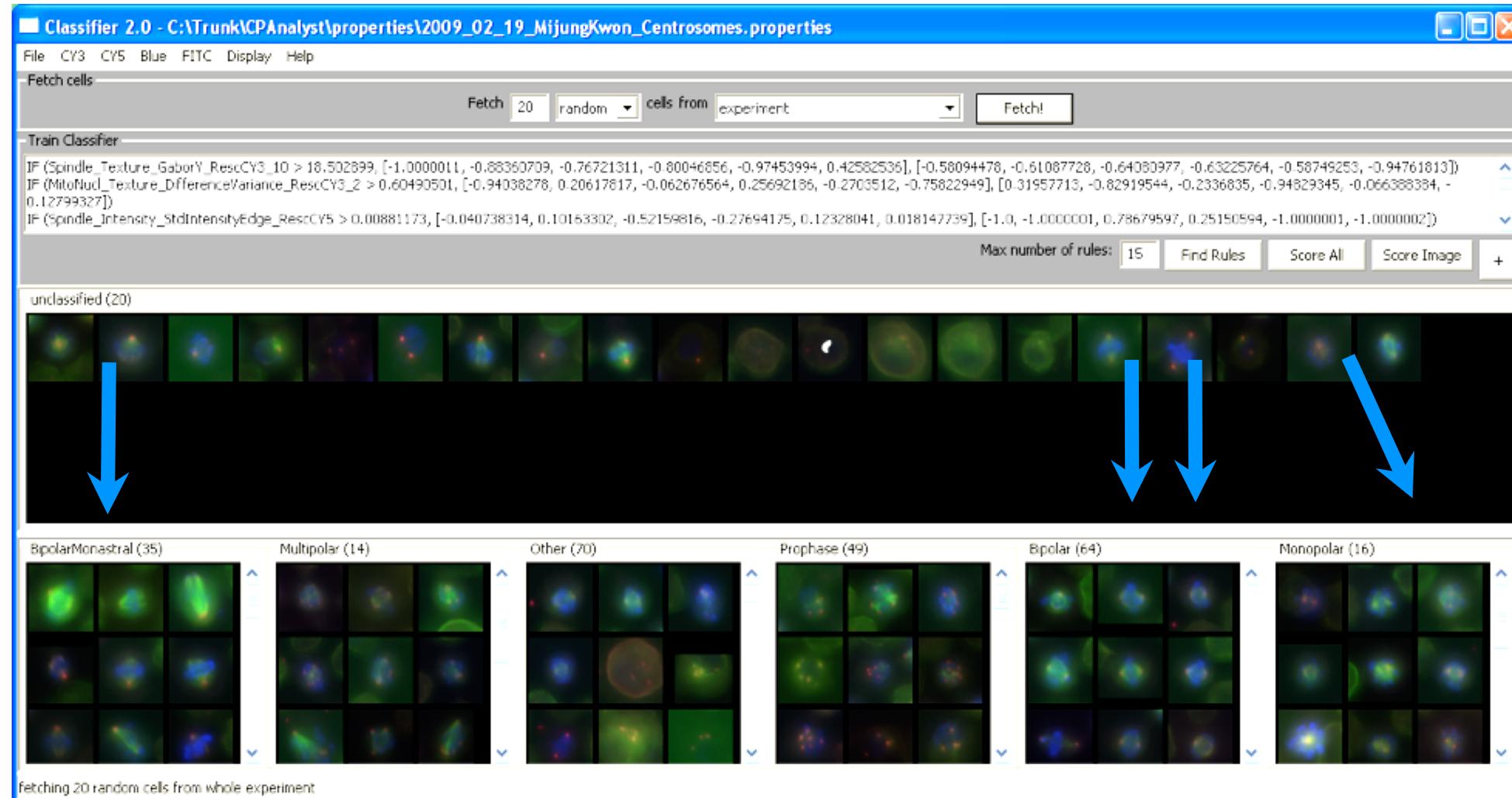
Phenotype classification using machine learning



Step 3: Score cells

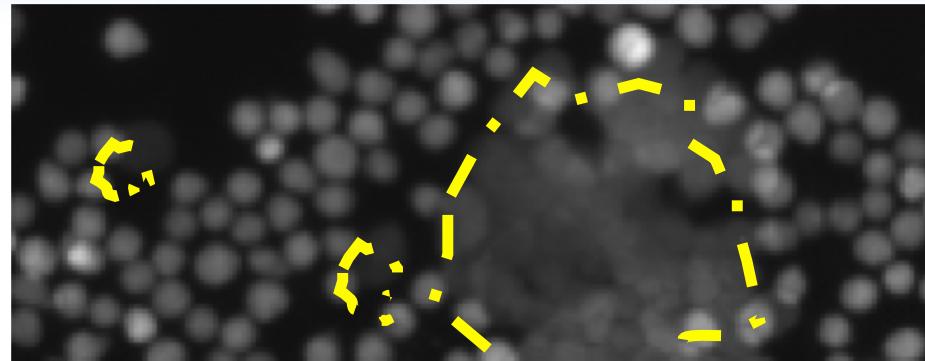


User-friendly machine learning for phenotype classification



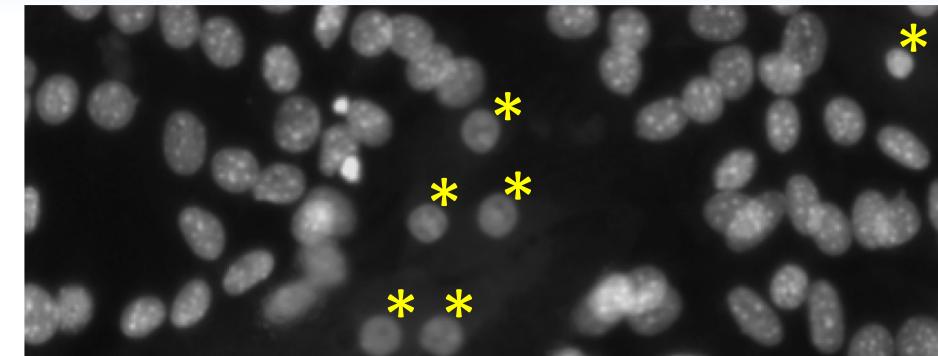
Jones, et al. PNAS 2009; Dao et al. Bioinformatics 2016

Measuring complex phenotypes



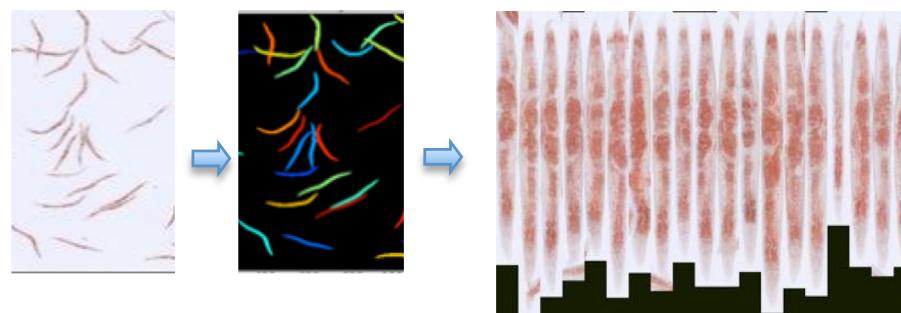
Identified drugs that reduce leukemic cell growth; lovastatin extends lifespan of mice given leukemic bone marrow cells

Hartwell, et al. Nature Chem Bio 2013 (Golub et al. labs)



Identified chemicals that stimulate primary human hepatocyte proliferation & differentiation

Shan, et al. Nature Chem Bio 2013 (Sangeeta Bhatia lab)



Identified genes whose knockdown alters fat metabolism

Wahlby, et al. Methods 2014 (Ruvkun lab)



Matched expert accuracy in detecting malaria stages in blood smears, using deep learning

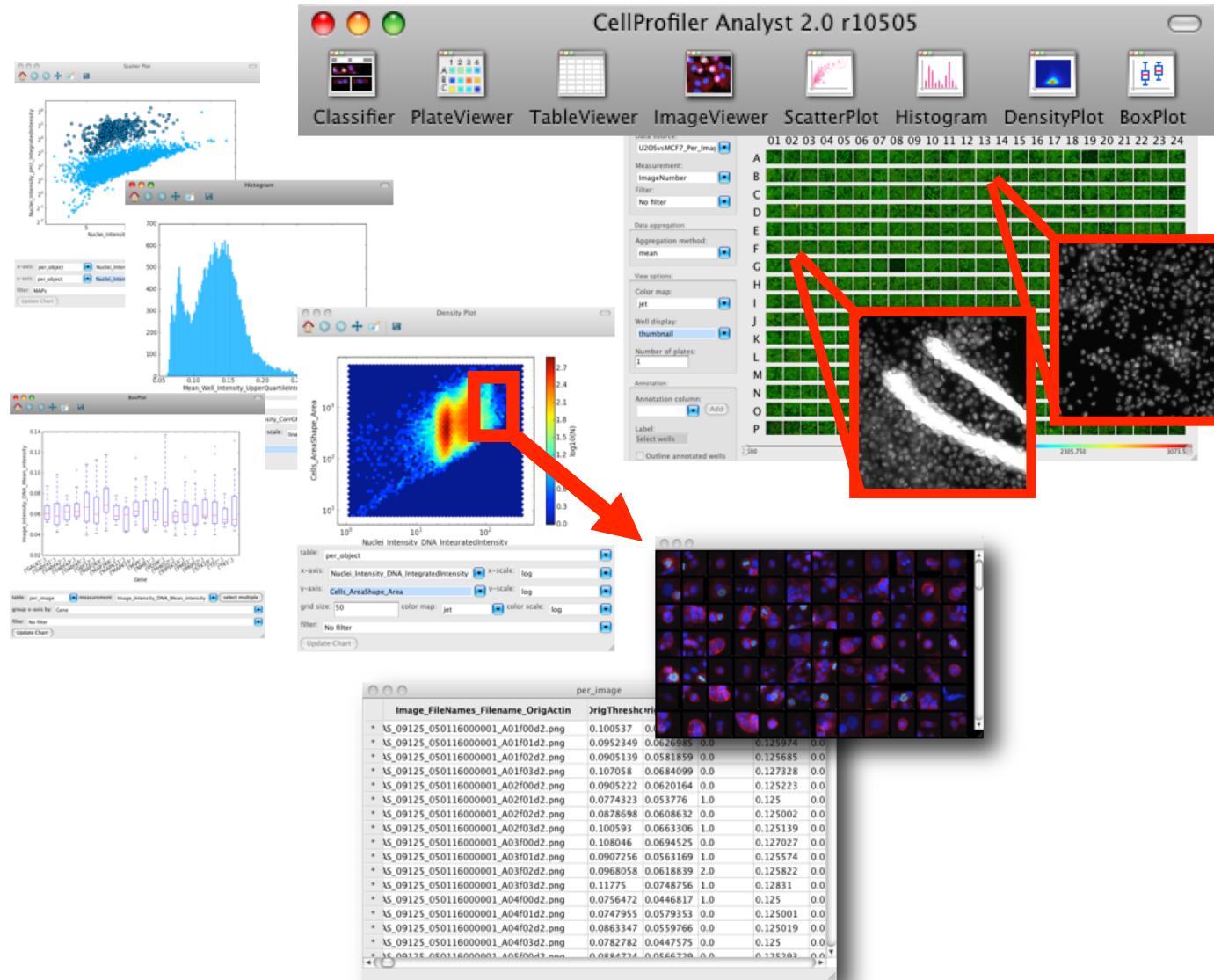
Hung, et al. in press BMC Bioinformatics (Marti lab)

Behind the scenes: CellProfiler Analyst present, past, and future (= Piximi)

CellProfiler Analyst history



- 2005-2008: wrote 1.0 in Java
- 2008: publicly release 1.0
- 2008 + 2009: published papers about CPA
- 2008-2009: wrote & released 2.0 in python
- 2016: upgraded & published paper about CPA



Future - Piximi: deep learning image classifier

To replace CellProfiler Analyst (Carpenter lab)
+ Advanced Cell Classifier (Horvath lab)



Allen Goodman David Dao

www.piximi.app
(under construction)

- Phenotype classifier for images
- **Web-based** 
- Uses **deep learning** algorithms 
- Interactive / intuitive user interface
- Source code: <https://github.com/piximi>
- Collaboratively developed!
- Rationale: easy installation (esp w TensorFlow) and platform independence



Christoph Friedrich
Lassi Paavolainen
Levin Moser
Bin Li
Ray Jones Csaba Molnar
Luca D'alessio
Rex Wang Nikolas Barkas
Frances Hubis Canchen Li
Ce Zhang Peter Horvath

Just Getting Started in bioimage analysis? Communities

NEUBIAS: <http://eubias.org/NEUBIAS>

BINA/Bioimaging North America: <https://www.bioimagingna.org>

Center for Open Bioimage Analysis (COBA): <http://nihcoba.org>

The Scientific Community Image Forum:

<https://forum.image.sc>

A screenshot of a web browser window showing the Image.sc forum. The address bar says "Search results for 'welcome im...'" and the URL is "forum.image.sc/search...". Below the address bar are standard Mac OS X window controls (red, yellow, green) and a "+" button. The main content area shows the "image.sc" logo and the text "Welcome to the Image.sc Forum!". It also includes a "Announcements" section and a "NEUBIAS Academy NEW live online bioimage analysis training events to follow" message.

iBiology video course:

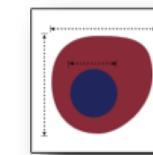
<https://www.ibiology.org/online-biology-courses/bioimage-analysis-course/>

A screenshot of a video player from the iBiology website. The title is "Bioimage Analysis Course: The Life Cycle of an Image Data Set". Below the title are social sharing icons for Facebook, Twitter, and Email, along with a "Share" button. The video frame shows two people, a woman and a man, in front of a background of biological images. The video player interface includes a play button, volume control, and "Watch later" and "Share" buttons.

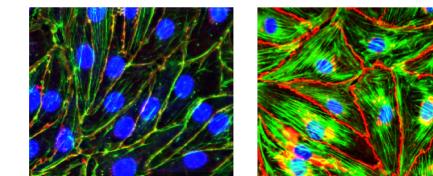
How can open bioimage analysis software power drug discovery?



Bioimage analysis to measure cells:
CellProfiler



Machine learning to classify cells:
CellProfiler Analyst & Piximi



**Libraries to profile cells:
cytominer and deepprofiler**



Morphological profiling – use images to create signatures of genes, compounds and diseases

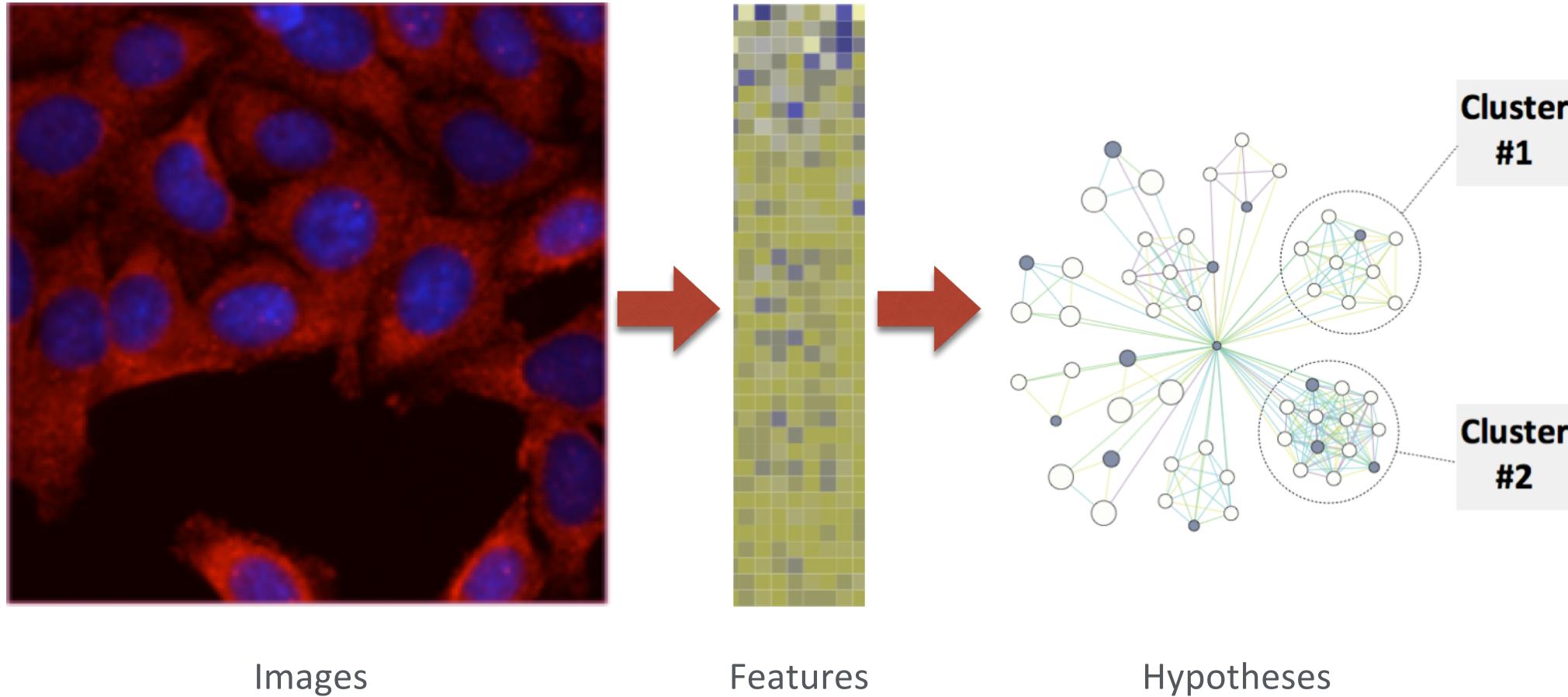
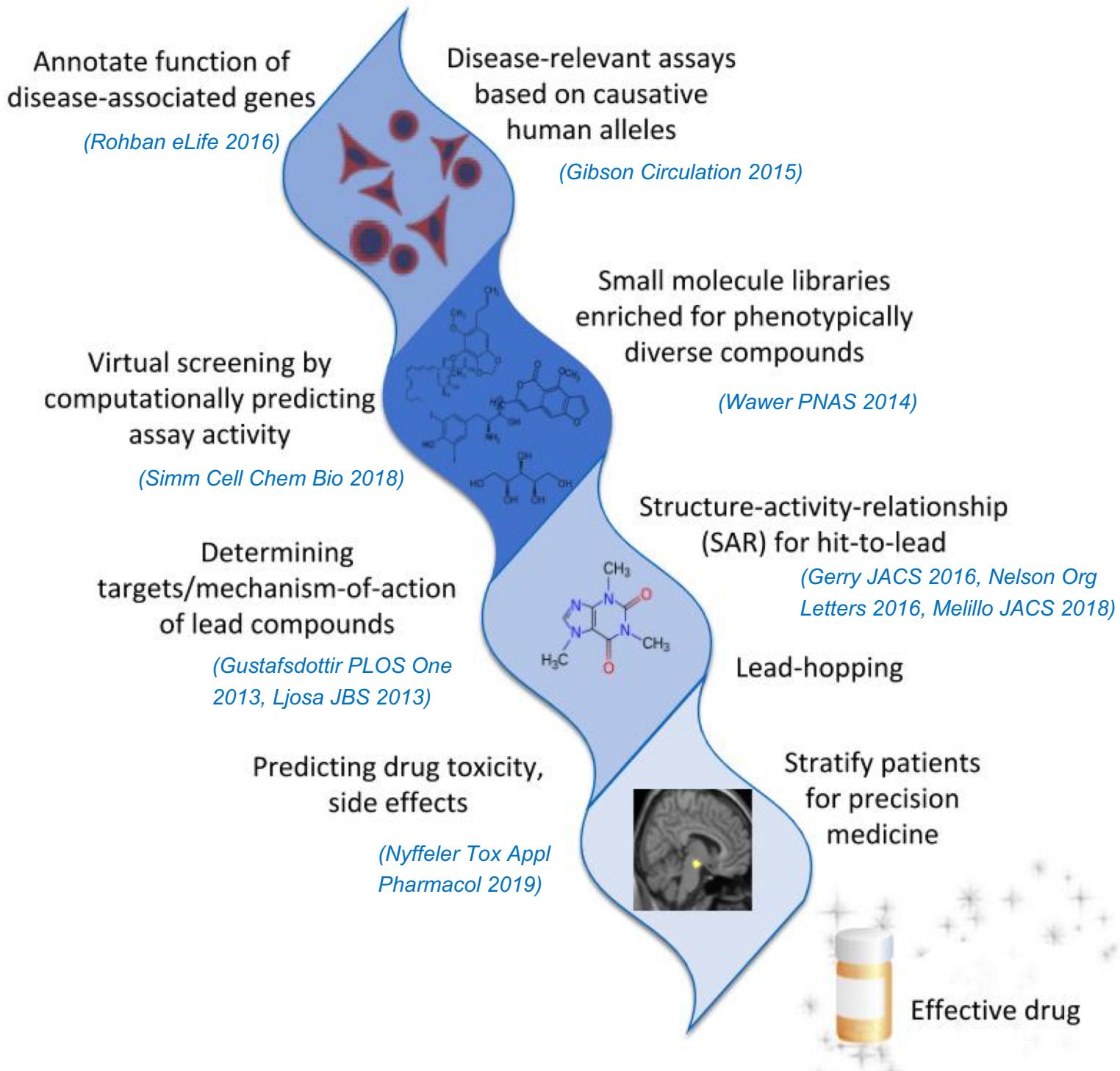


Image-based profiling: accelerating many steps of drug discovery



Transforming drug discovery via Cell Painting

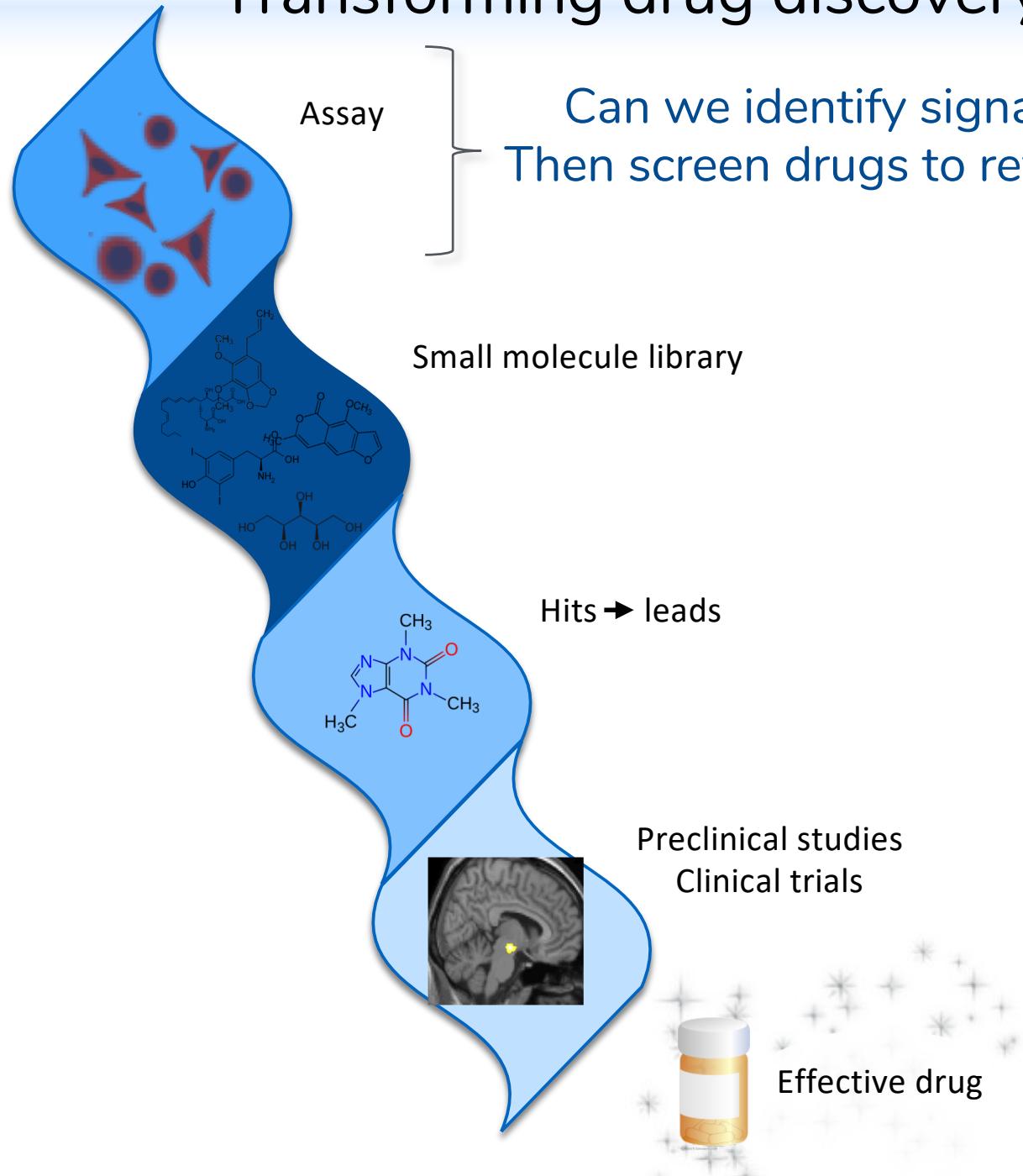
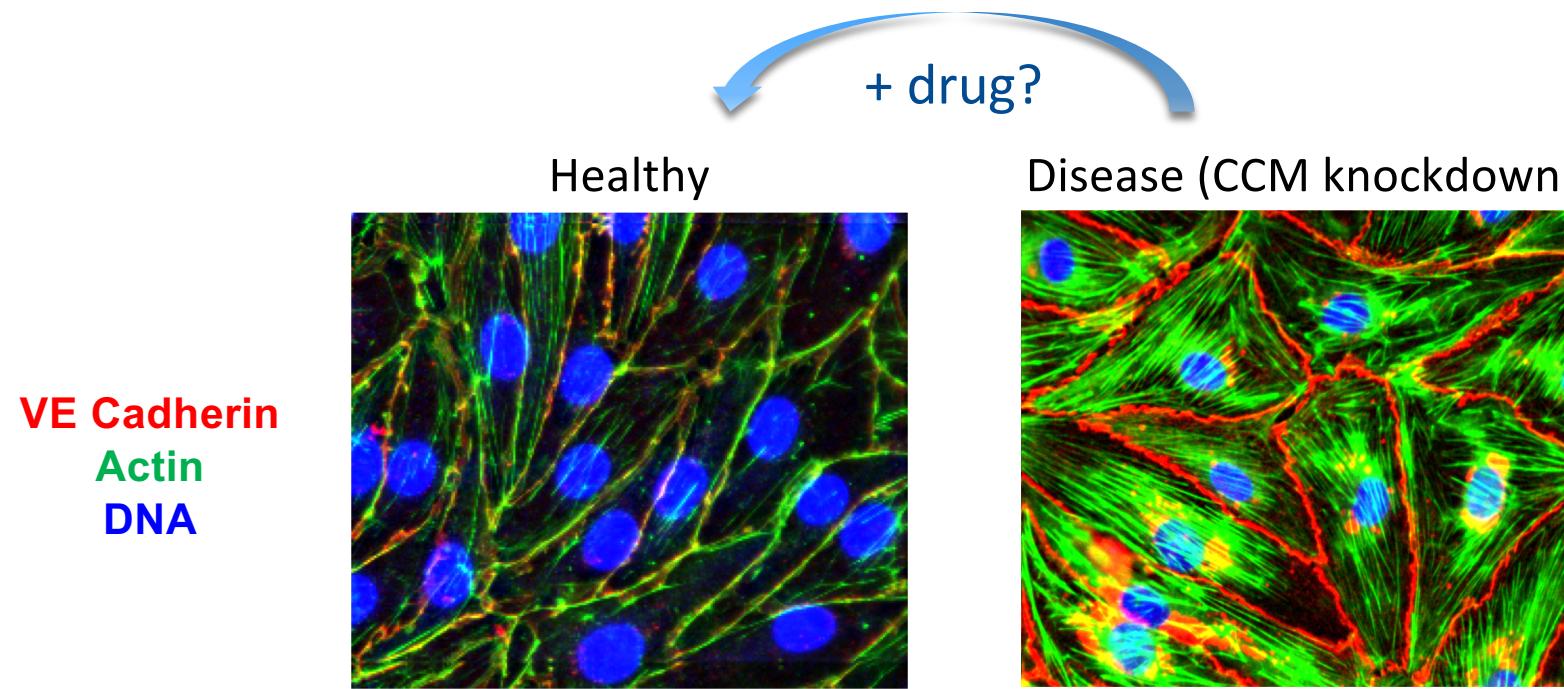


Image-based profiling can identify drugs for disease



*Disclosure: I serve on
Recursion's Scientific
Advisory Board*

Gibson, et al. Circulation 2015

Disease states can be morphologically distinct



Kyle Karhohs



Mohammad Rohban



Shantanu
Singh



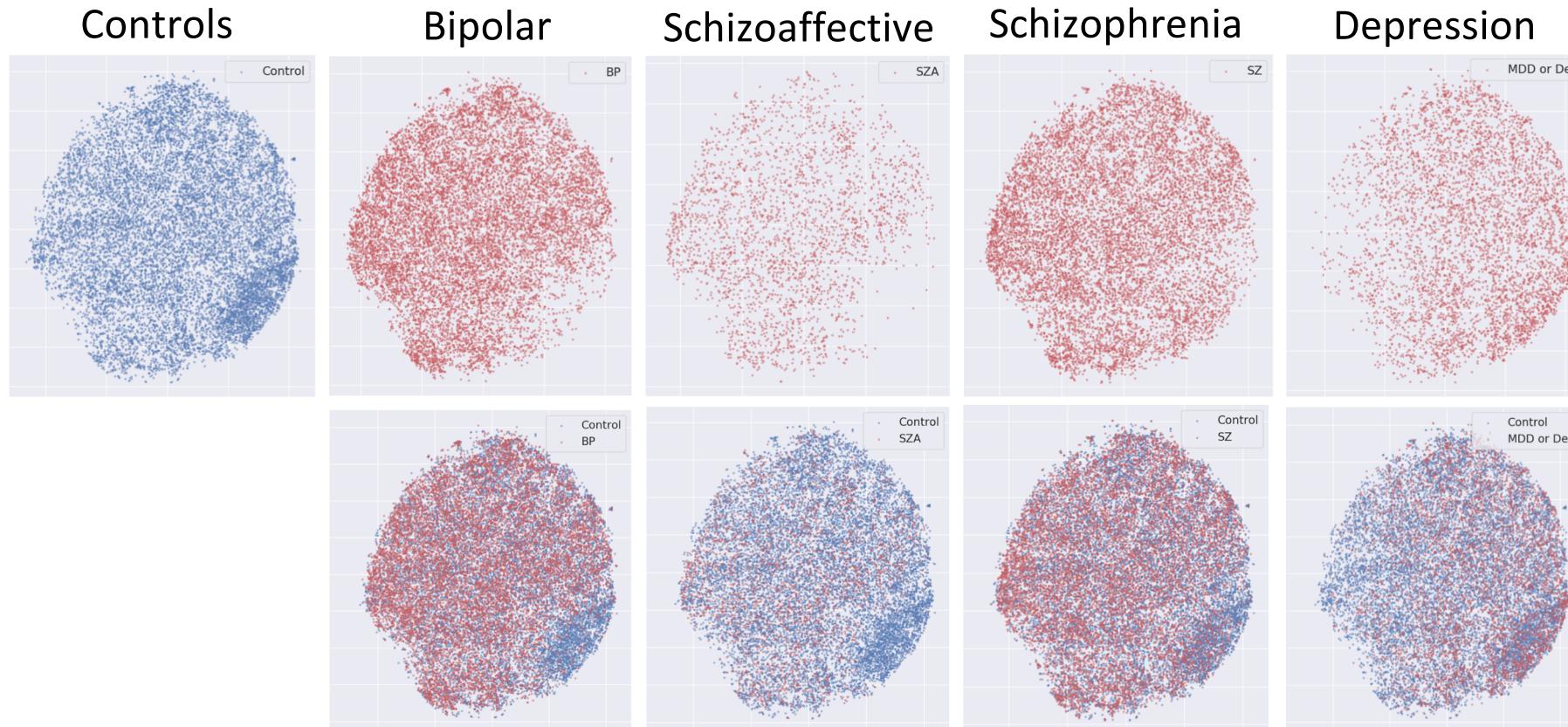
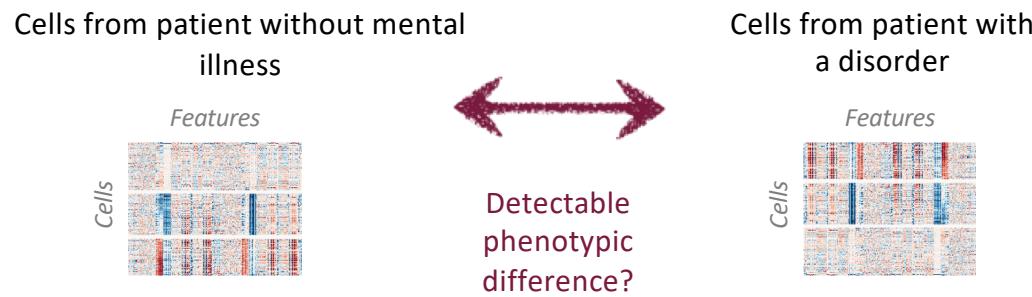
Marzieh
Haghghi



Bruce Cohen,
McLean
Hospital



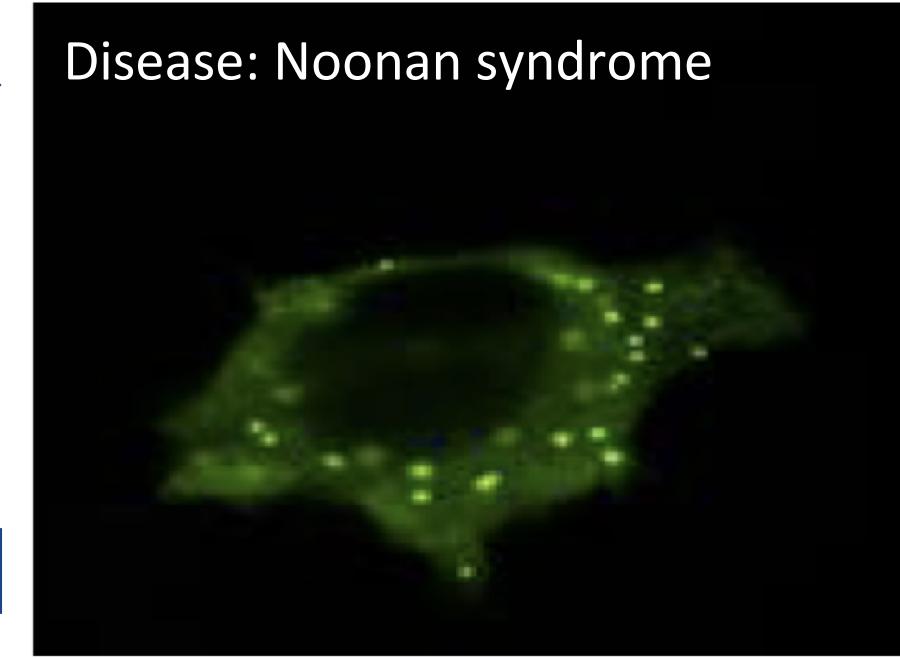
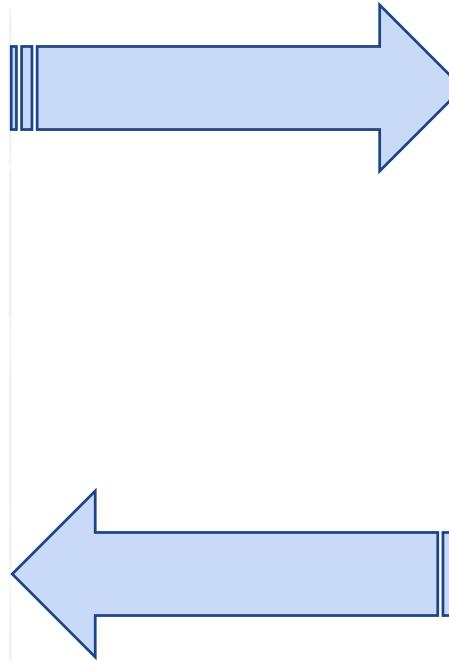
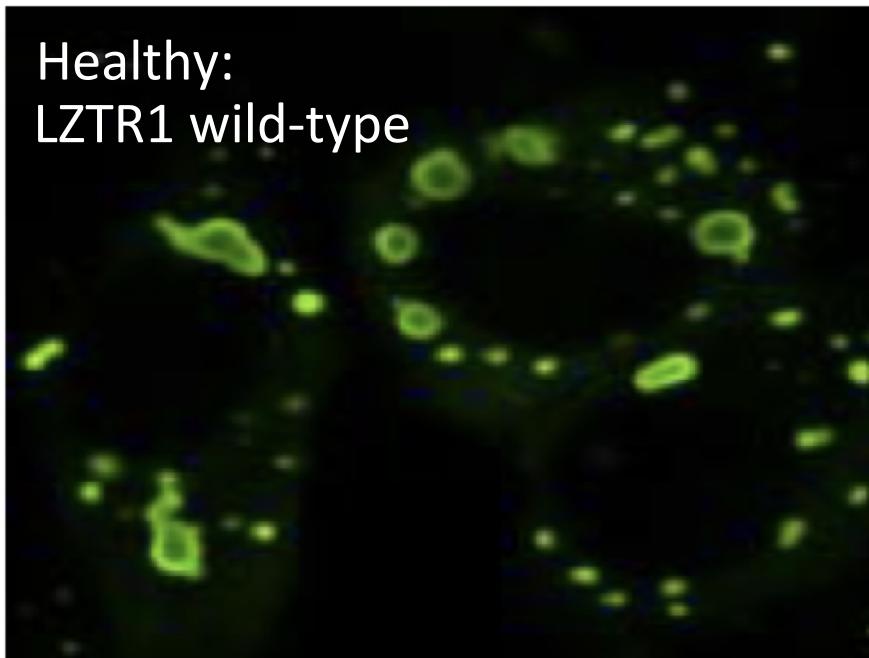
Donna McPhie



Project in progress

Identify disease-associated phenotypes using imaging

Step 1: Identify pathogenic phenotype



Step 2: Test existing drugs to reverse the phenotype, and genetic perturbations to find potential targets/pathways



Mikko
Taipale,
U Toronto



Jessica
Lacoste



Marzieh
Haghghi

Shantanu
Singh

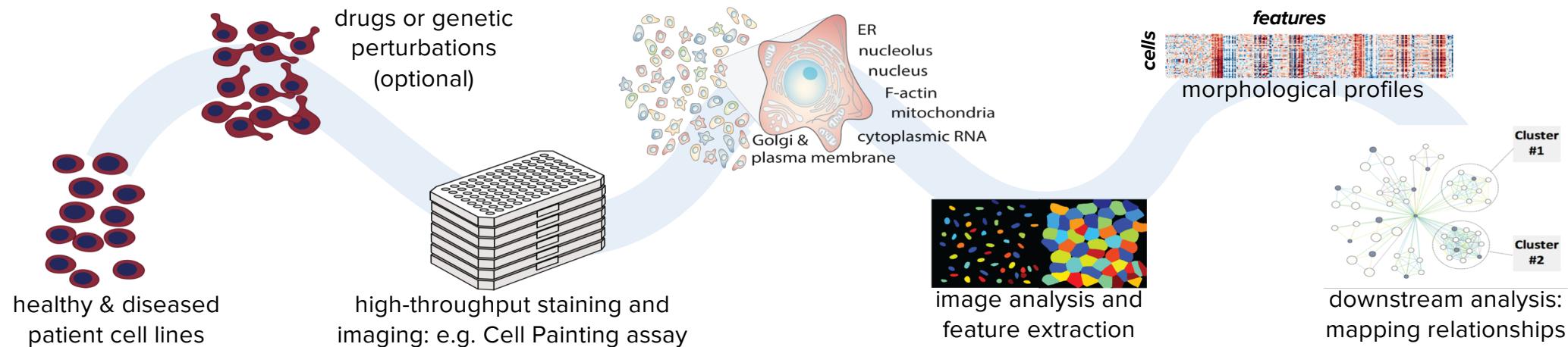


Marzieh
Haghghi

Shantanu
Singh

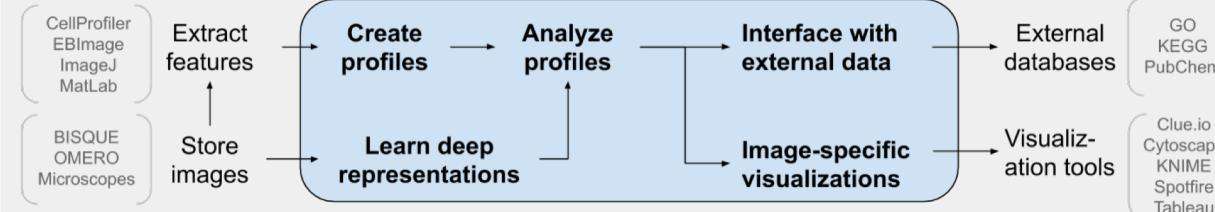
Behind the scenes: cytominer & deepprofiler present, past, and future

Software libraries for image-based profiling



Cytominer / Pycytominer

R/Python libraries for image-based profiling workflows

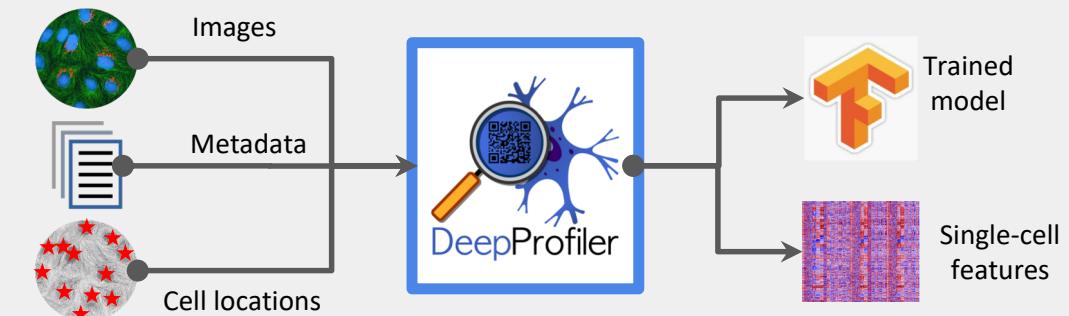


[github.com/cytomining/\[py\]cytominer](https://github.com/cytomining/[py]cytominer)

Shantanu Singh · Gregory Way · Allen Goodman
Tim Becker · Claire McQuin · Mohammad Rohban

DeepProfiler

Deep learning library for image-based profiling



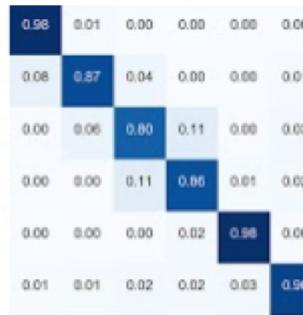
github.com/cytomining/DeepProfiler

Juan Caicedo · Santiago Benoit · Matthew Smith · Nikita Moshkov ·
Claire McQuin · Marzieh Haghghi · Allen Goodman · Shantanu Singh

More successes of image-based profiling

Applications reviewed in: Caicedo, Singh, Carpenter, Current Opin Biotech (2016) & Chandrasekaran, Ceulemans, Boyd, Carpenter (2020: coming soon!)
 Best practices reviewed in: Caicedo, et al. Nature Methods (2017)
 Annual conferences: CytoData, SBI2, Images 2 Knowledge (I2K)

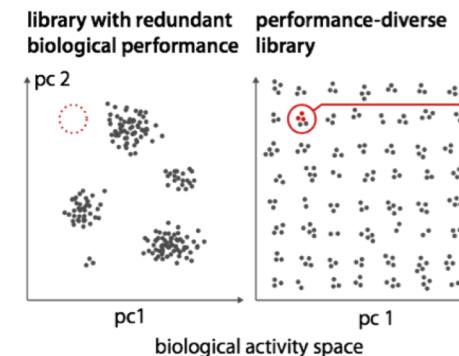
Label-free imaging flow cytometry diagnostics



classify cell cycle, red blood cell lesions, leukemic progression and response

Blasi, et al. Nature Comm. 2016
 Hennig, et al. Methods 2016
 Eulenberg, et al. Nature Comm. 2017
 Doan, et al. Trends Biotech 2018
 (+ work in progress)

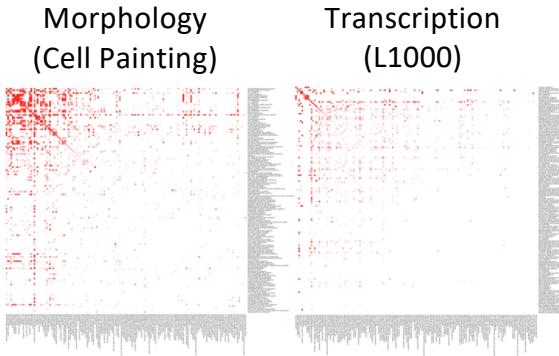
Creating performance-diverse screening libraries



improved hit rate
 50% increase vs randomly chosen compounds

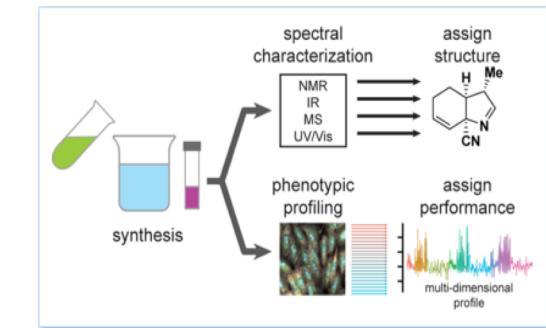
Wawer, et al. PNAS 2014

Identify drug targets/mechanisms of action



Ljosa, et al. JBS 2013
 Natoli, et al. in progress

Testing bioactivity for newly synthesized compounds



improved turnaround time
 from years to days

Gerry, et al. JACS 2016
 Nelson, et al. Org Lett 2016
 Melillo, et al. JACS 2018

How to get involved in biomedical data science/engineering?



Come work at the Broad!

*30+ open positions in data science/software:
www.broadinstitute.org/careers*

Work on a public data challenge!

Contribute to a code base!

Please.

Transitioning to computational work

Challenges

- learn different approach to problems, cultural practices
- learn and keep up with cutting-edge in two fields
- outside comfort zone / impostor syndrome
- funding/papers falling between the fields

Benefits

- major impact
- different work/life patterns
- collaborative
- improves communication skills
- major impact**



Software
Sustainability
Institute

Carpenter, *Bridging Domain and Data, Patterns* (2020)
<http://broad.io/DomainAndData>



Gratitude



Lab members

Hamdah Abbasi
Jeanelle Ackerman
Niranj Chandrasekaran
Beth Cimini
Allen Goodman
Marzieh Haghghi

Nasim Jamali
Becki Ledford
Shantanu Singh
David Stirling
Gregory Way
Erin Weisbart

Many thanks to our many collaborators!

Biologist wanting to learn image analysis?
Join our Postdoctoral Training Program in
Bioimage Analysis

Earned your PhD in Sweden?
Wallenberg postdoc fellowship at the
Broad Institute (due Nov)

Computational? Within 3 years of earning
PhD? Become a Schmidt Fellow at the Broad

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>

Chat / Qs: <http://broad.io/scipy2020chat>