From Lots of Pictures to Lots of Numbers to Lots of Answers:

Advances in quantitative high-content imaging

Dr. Erin Weisbart
Senior Computational Scientist, Lead Image Analyst
Cimini Lab, Imaging Platform
Broad Institute of MIT and Harvard

BROAD INSTITUTE

Overview

Why images?

Advances in software
data science
machine learning

Importance of public data
open-source software
cross-discipline collaboration

Slides available at http://broad.io/Weisbart_WACD_2023

Why images?

Biological images are informative

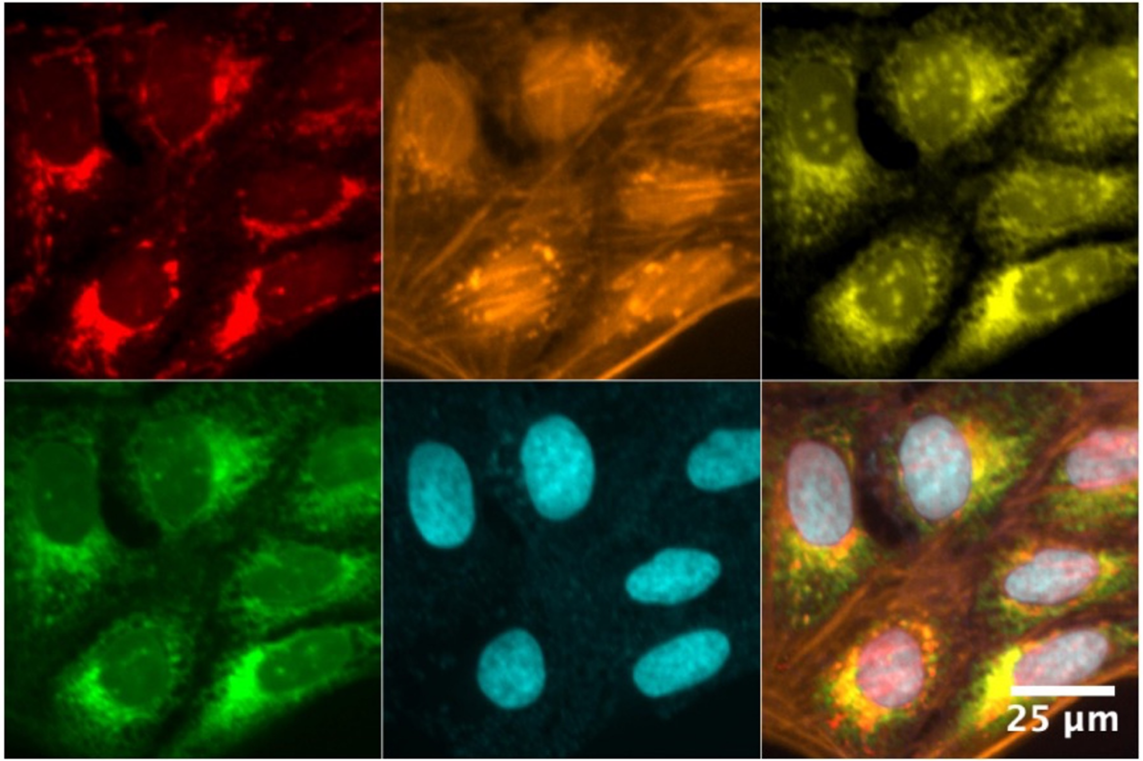
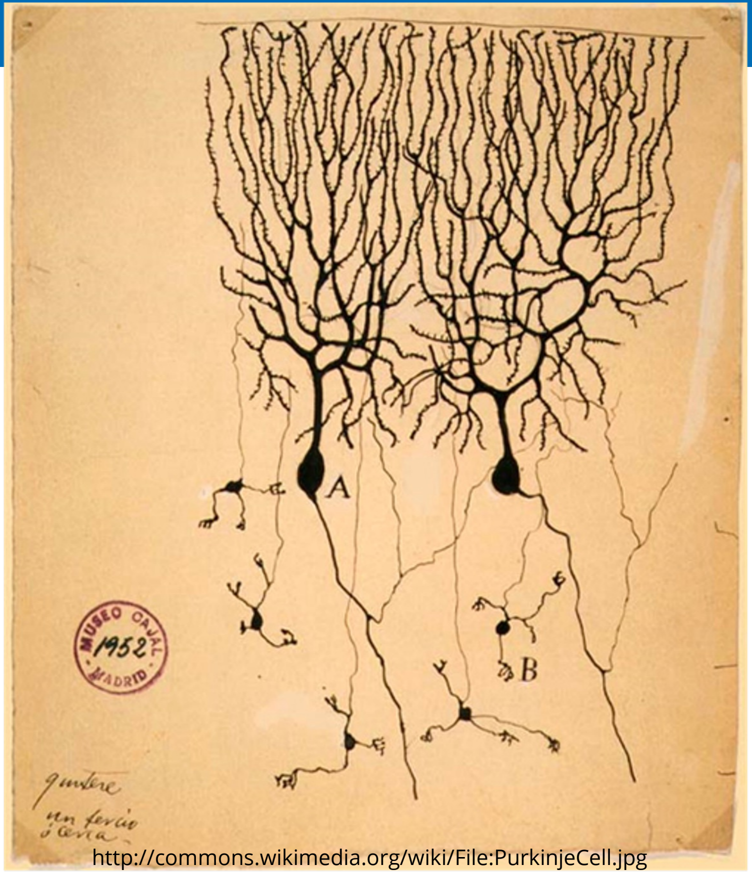


Image data can be:

- Single cell resolution
- Interaction between molecules
- Live cell data
- Changes in expression level **and** behavior
- …
- Quantitative

Microscopy’s history is mostly qualitative



http://commons.wikimedia.org/wiki/File:PurkinjeCell.jpg

Drawing of Purkinje cells (A) and granule cells (B) from pigeon cerebellum, 1899. Instituto Santiago Ramón y Cajal, Madrid, Spain

My history of image quantitation

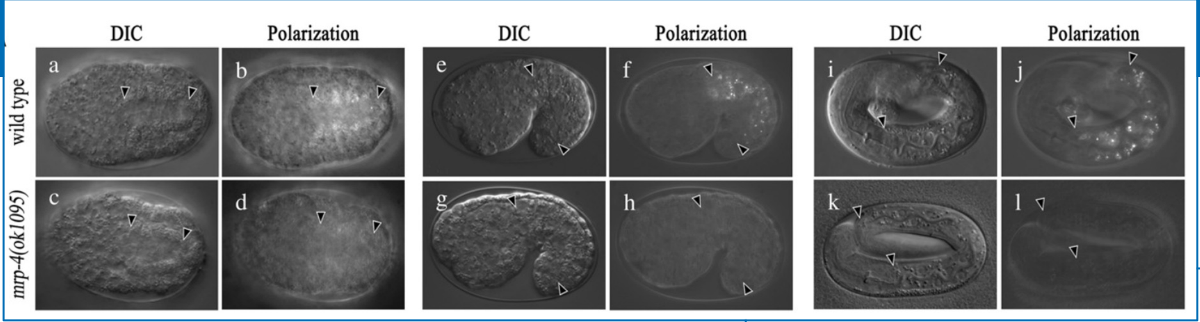


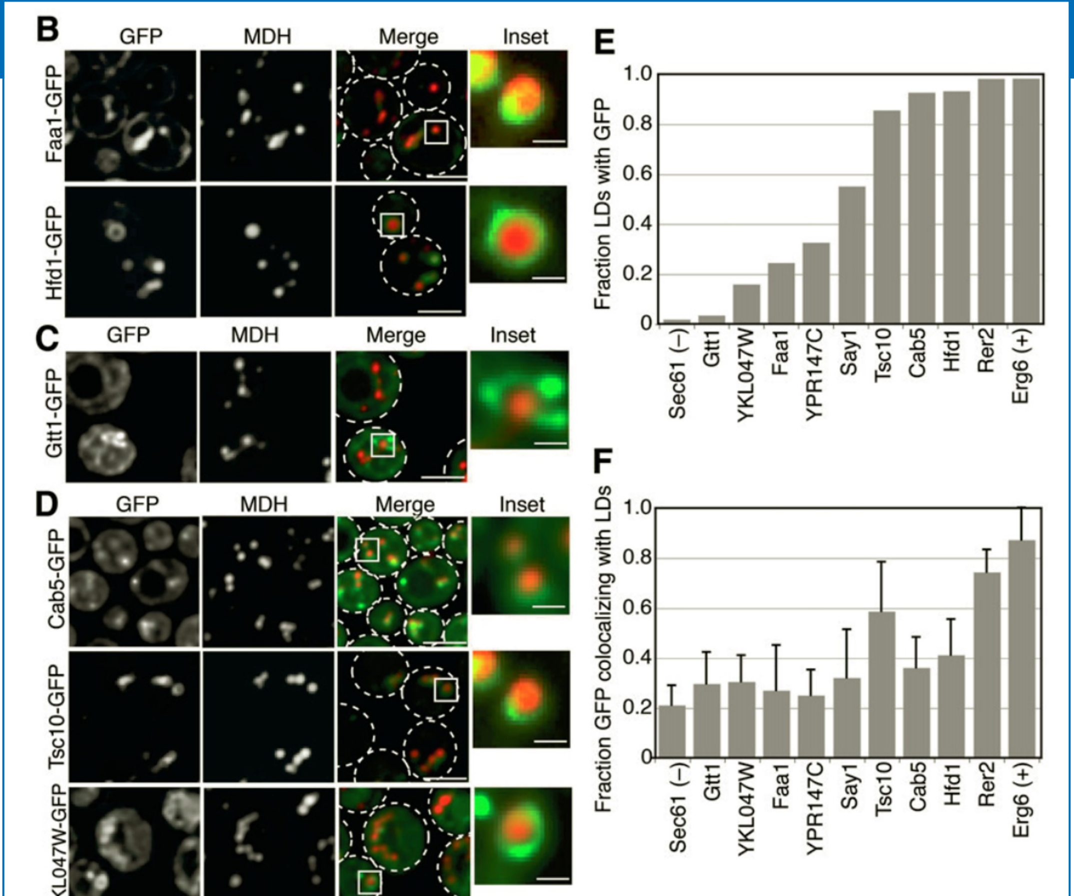
TABLE 1

Analysis of birefringent gut granules

Genotype	% of embryos lacking birefringent material in intestinal cells (*n*)	% of larvae lacking birefringent material in intestinal cells (*n*)
Wild type[a]	0 (>100)	0 (>100)
ABC transporters		
mrp-4(cd8)	93 (99)	20 (46)
mrp-4(ok1095)	96 (164)	12 (182)
mrp-4(RNAi)	93 (390)	14 (227)
wht-2(RNAi)	52 (673)	5 (137)
Double mutants		
mrp-4(cd8); wht-2(RNAi)	97 (107)	81 (89)
mrp-4(ok1095); wht-2(RNAi)	100 (109)	89 (61)
Mosaic RNAi		
rrf-1(pk1471)[a]	0 (96)	0 (47)
rrf-1(pk1471); mrp-4(RNAi)	53 (45)	36 (22)
rrf-1(pk1471); wht-2(RNAi)	3 (119)	0 (26)
rsd-2(pk3307)[a]	0 (84)	0 (45)
rsd-2(pk3307); mrp-4(RNAi)	0 (63)	0 (40)

Currie, E., King, B., Lawrenson, A. L., Schroeder, L. K., Kershner, A. M., & Hermann, G. J. (2007). Role of the Caenorhabditis elegans multidrug resistance gene, mrp-4, in gut granule differentiation. *Genetics*, *177*(3), 1569–1582. https://doi.org/10.1534/genetics.107.080689

My history of image quantitation



Currie, E., Guo, X., Christiano, R., Chitraju, C., Kory, N., Harrison, K., Haas, J., Walther, T. C., & Farese, R. V. (2014). High confidence proteomic analysis of yeast LDs identifies additional droplet proteins and reveals connections to dolichol synthesis and sterol acetylation [S]. *Journal of Lipid Research*, *55*(7), 1465–1477. https://doi.org/10.1194/jlr.M050229

Imaging Platform at Broad Institute

Anne E. Carpenter
Senior Director of the Imaging Platform, Institute Scientist

Shantanu Singh
Senior Group Leader

Beth Cimini
Associate Director for Bioimage Analysis

Software to turn Images → Numbers → Answers

CellProfiler™
cell image analysis software

CellProfiler Analyst™
data exploration software

Distributed Something™

PyCytominer

PIXIMI™

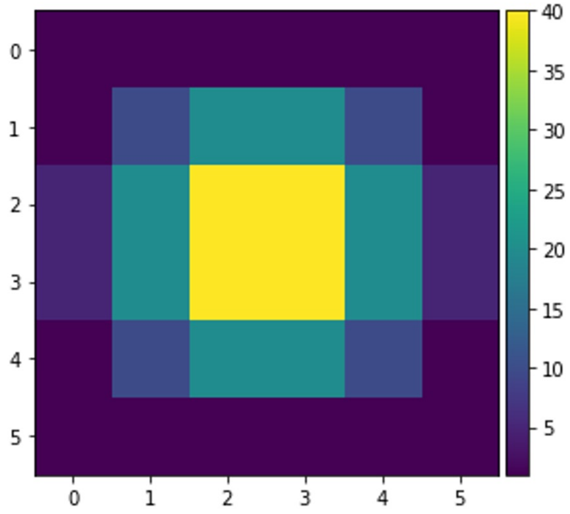
How has software and data science changed imaging?

Digital images are arrays

```
skimage.io.imshow(a)
```

```
/usr/local/lib/python3.8/site-packages/skimag
e; displaying image with stretched contrast.
  lo, hi, cmap = _get_display_range(image)
```

```
<matplotlib.image.AxesImage at 0x125202070>
```



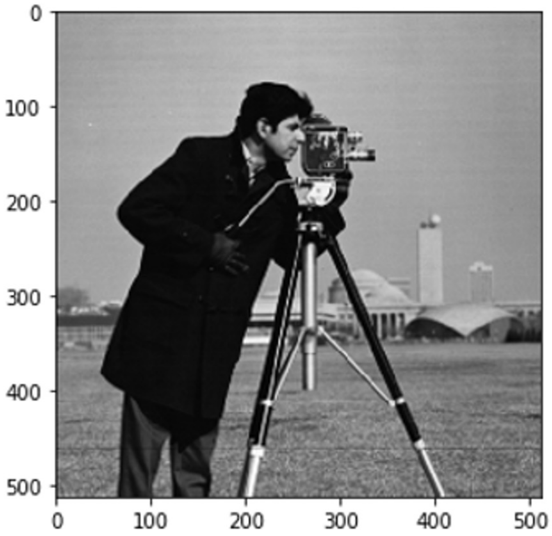

```
a
```

```
array([[ 1,  1,  1,  1,  1,  1],
       [ 1, 10, 20, 20, 10,  1],
       [ 5, 20, 40, 40, 20,  5],
       [ 5, 20, 40, 40, 20,  5],
       [ 1, 10, 20, 20, 10,  1],
       [ 1,  1,  1,  1,  1,  1]])
```

Digital images are arrays

```
skimage.io.imshow(b)
```

<matplotlib.image.AxesImage at 0x12552cca0>




```
b
```

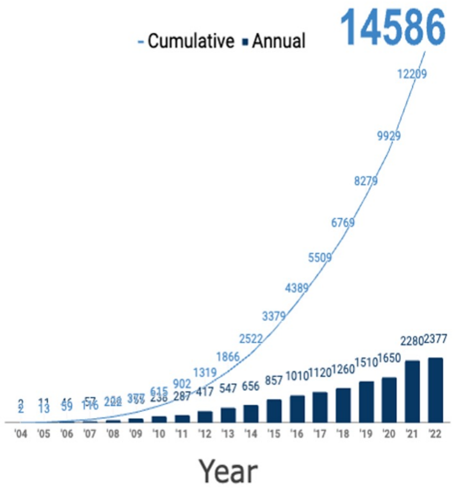
```
array([[156, 157, 160, ..., 152, 152, 152],
       [156, 157, 159, ..., 152, 152, 152],
       [158, 157, 156, ..., 152, 152, 152],
       ...,
       [121, 123, 126, ..., 121, 113, 111],
       [121, 123, 126, ..., 121, 113, 111],
       [121, 123, 126, ..., 121, 113, 111]], dtype=uint8)
```

CellProfiler™
cell image analysis software

Open-source software for image analysis

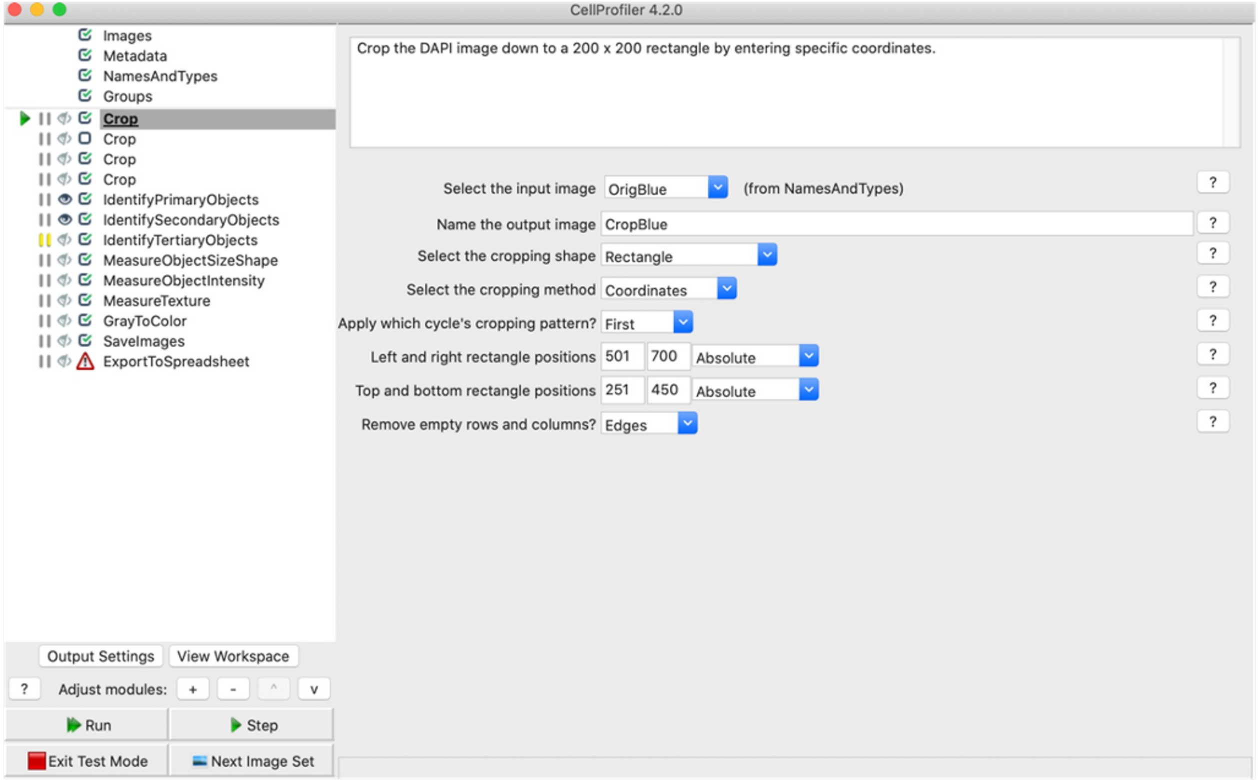
- Free and open-source; Windows, Mac, Linux
- Ranked **most flexible** and **usable** in independent analysis (*Wiesmann et al.)*
- Cited in **2,000+** papers per year
- Used in **7/10** top pharma companies
- In the **Top 10** most popular papers in Genome Biology

Publications citing CellProfiler



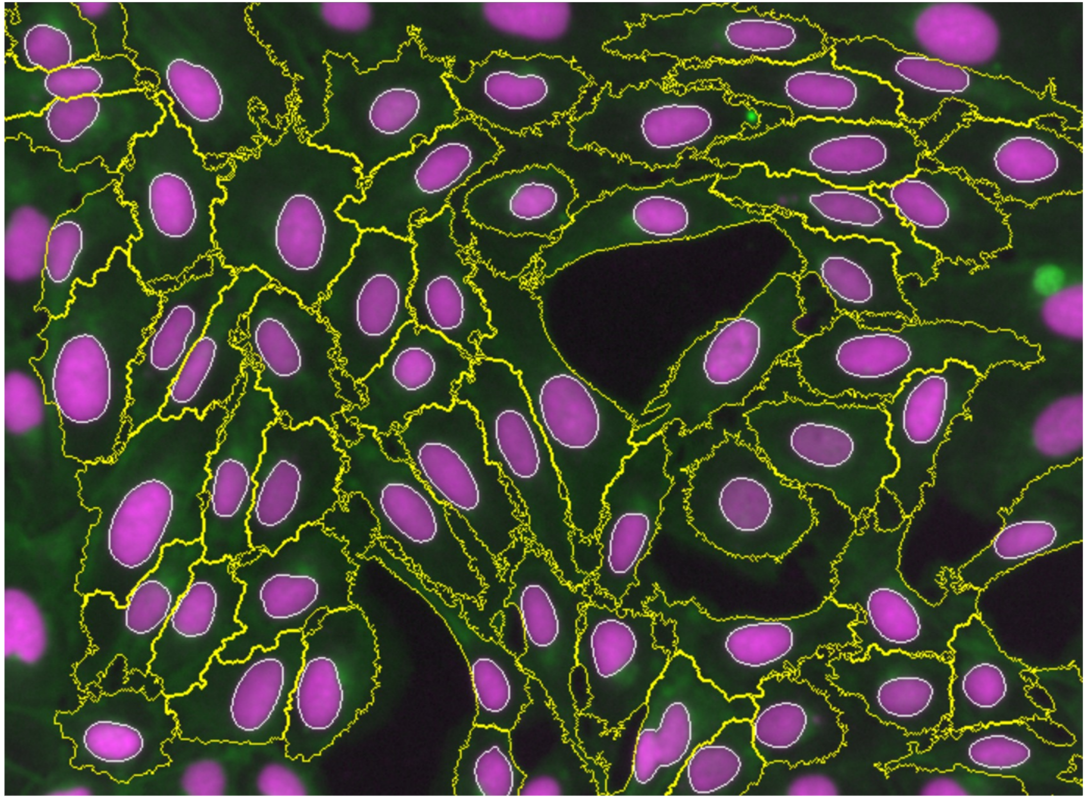
Anne Carpenter | Ray Jones | Lee Kamentsky | Allen Goodman | Claire McQuin | Beth Cimini | David Stirling | Alice Lucas | Nodar Gogoberidze

CellProfiler Interface



- 94 “modules”
- 930 settings
- Total library is 121K lines of code, 37K lines of documentation (23% documentation)

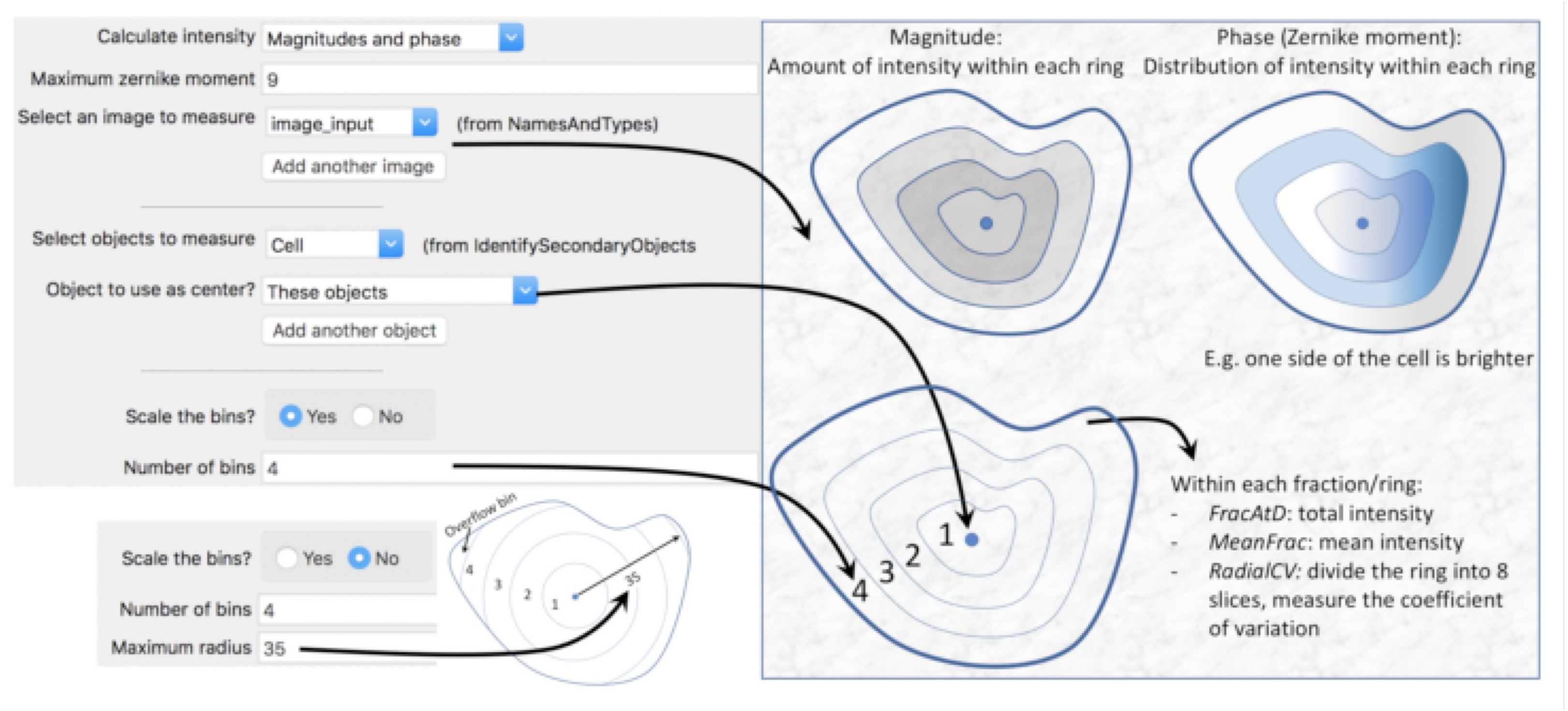
Identify Objects



Measure object … intensity

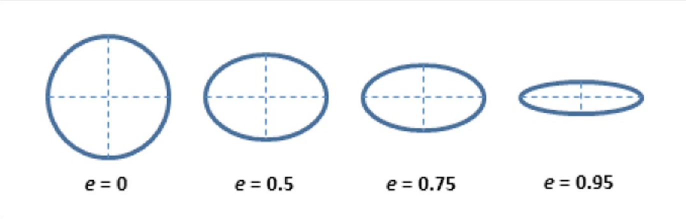
- Integrated intensity: Sum of the pixel intensities within an object
- Mean, median, standard deviation intensities
- Maximal and minimal pixel intensities
- Lower/Upper quartile of the intensity
- Object intensities may be measured from any channel, not just the channel used to identify the object

Measure object … intensity distribution

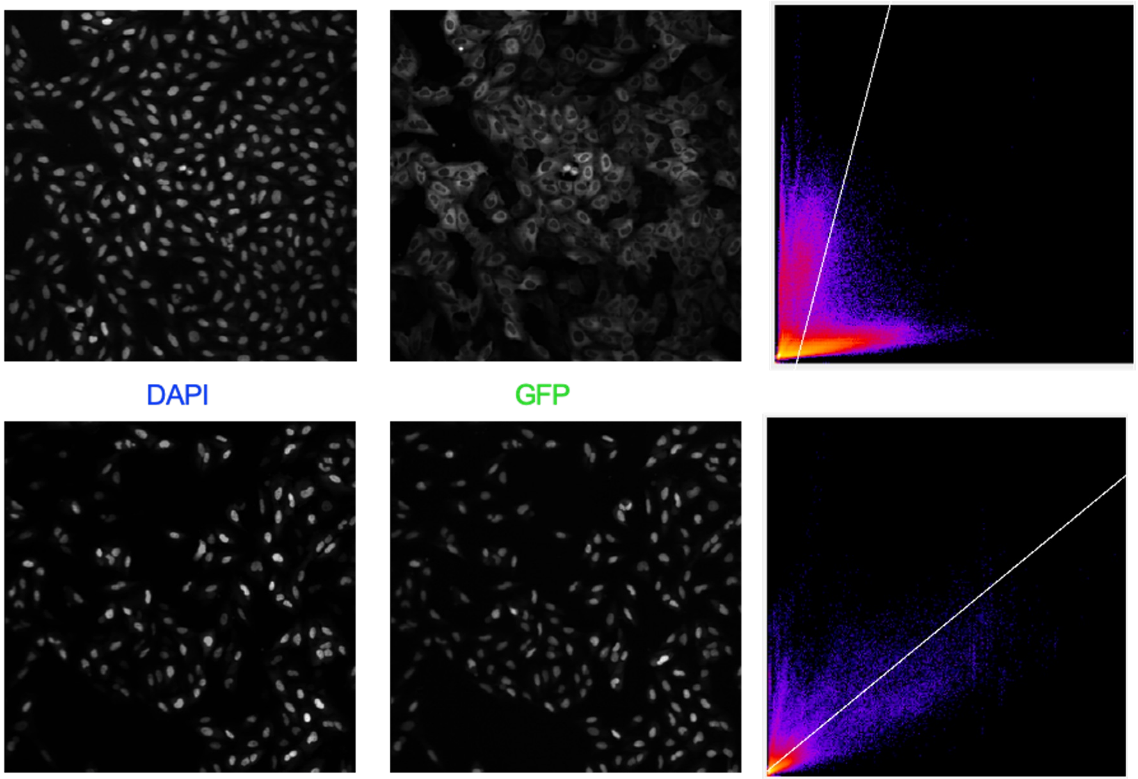


Measure object ... size and shape

- Area
- Perimeter
- Eccentricity (circle = 1, line = 0)
- MajorAxisLength
- MinorAxisLength
- Orientation
- Solidity

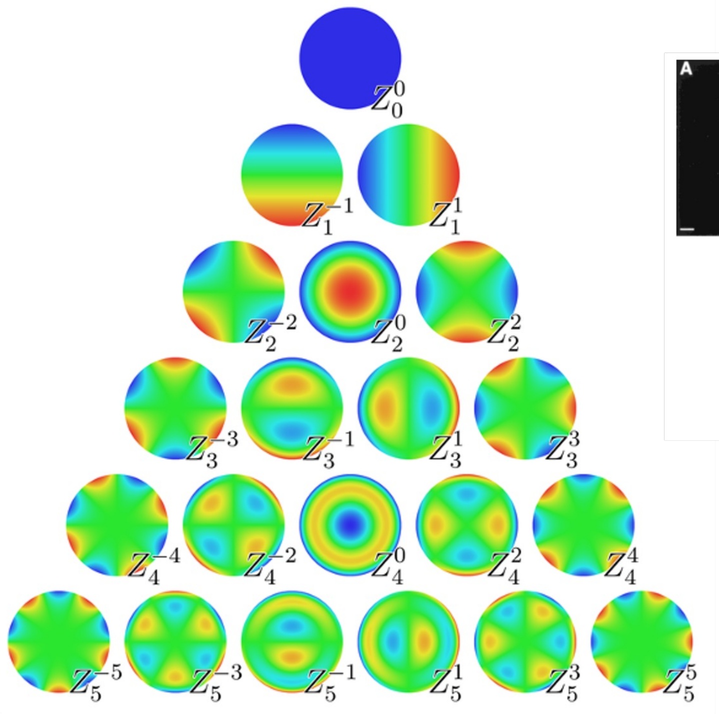


Measure object … colocalization

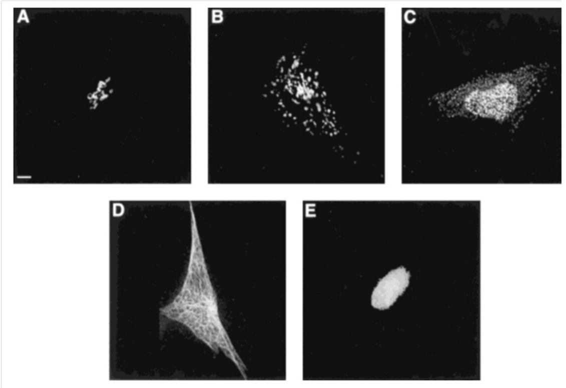


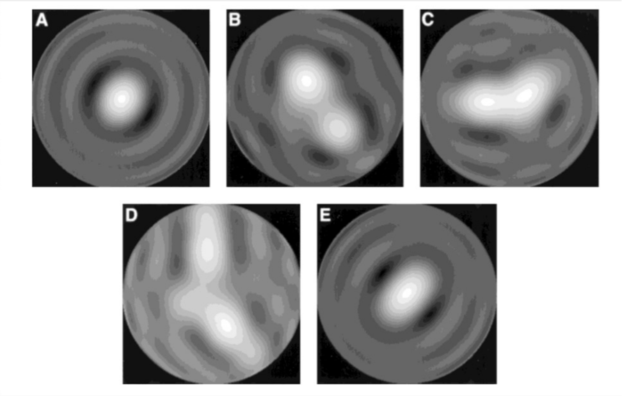
FIJI, Coloc 2- https://imagej.net/Coloc_2

Measure object … Zernike polynomials



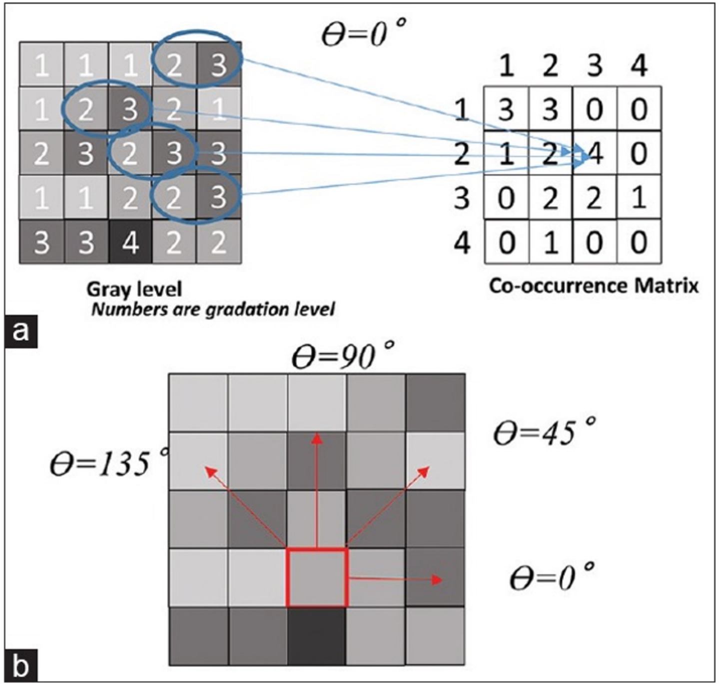
https://en.wikipedia.org/wiki/File:Zernike_polynomials2.png

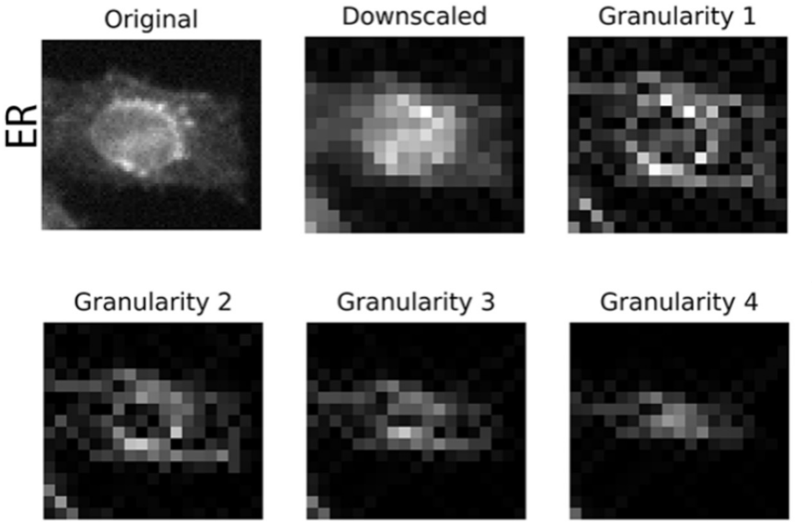




Boland, M.V., Markey, M.K., Murphy, R.F. 1998

Measure object … texture, granularity

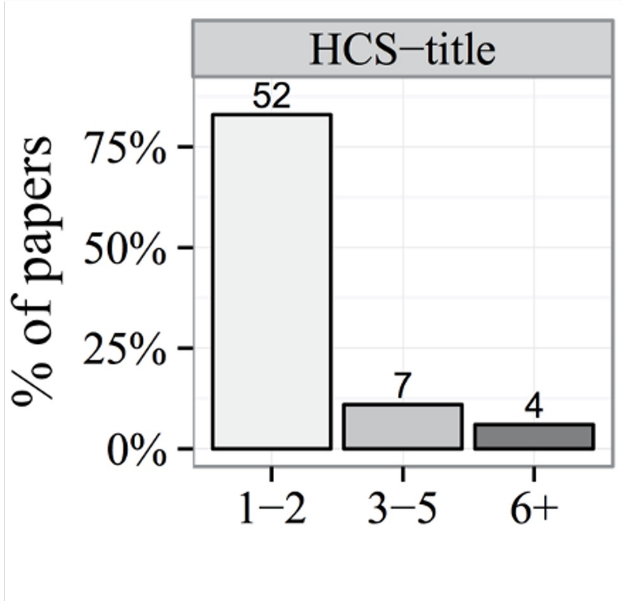




Saito A, Numata Y, Hamada T, et al. J Pathol Inform. 2016;7:36

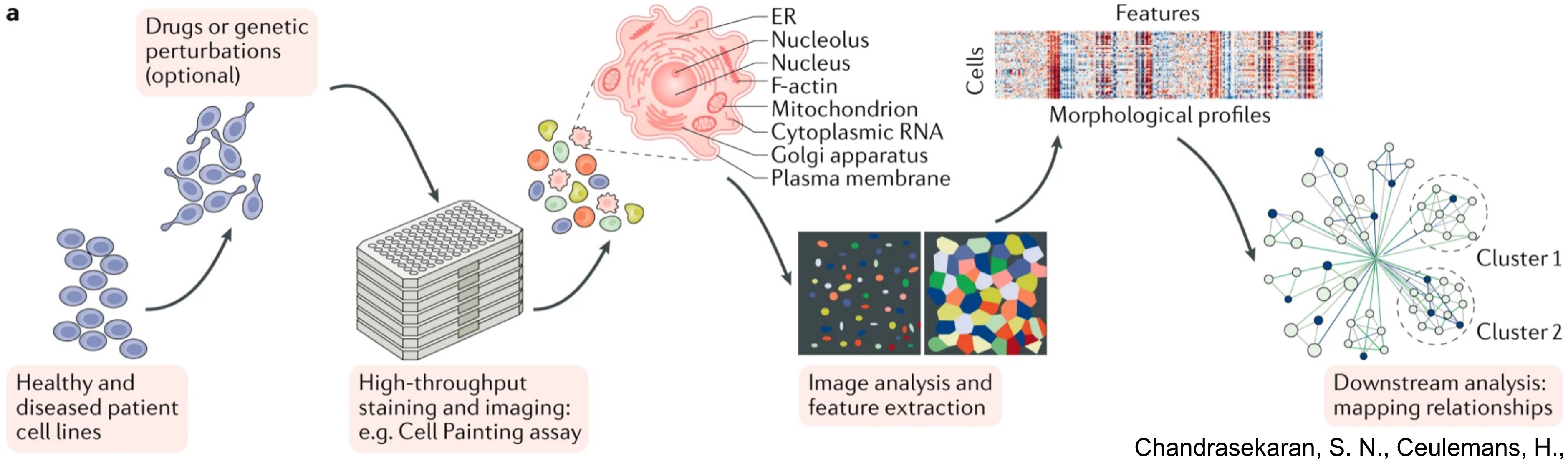
Most high-content screens are not data rich

60-80% of “high-content” studies use only 1 or 2 cellular features



Singh, Carpenter, Genovesio. (2014) J Biomol Screen 19(5):640-650.

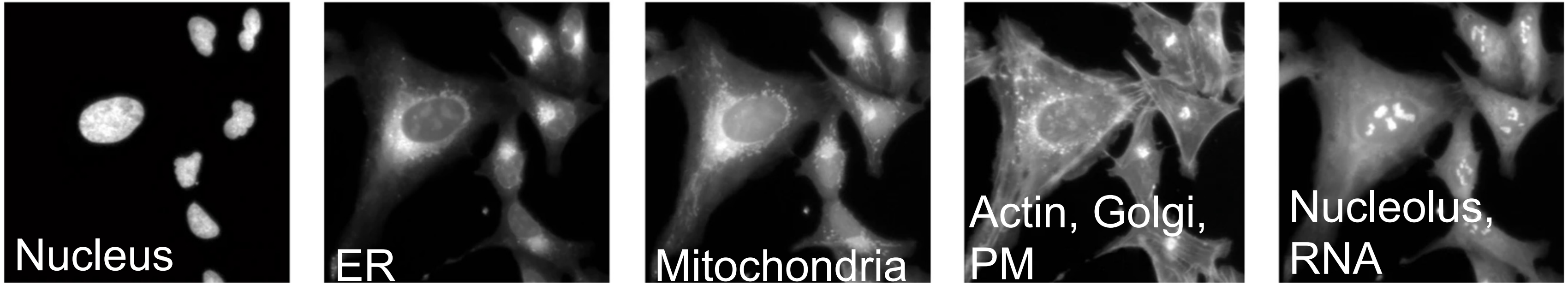
Image-based (Morphological) Profiling



Chandrasekaran, S. N., Ceulemans, H., Boyd, J. D., & Carpenter, A. E. (2021). Image-based profiling for drug discovery: due for a machine-learning upgrade? *Nature Reviews. Drug Discovery*, *20*(2), 145–159. https://doi.org/10.1038/s41573-020-00117-w

Image-based Profiling with Cell Painting

Cell Painting Assay



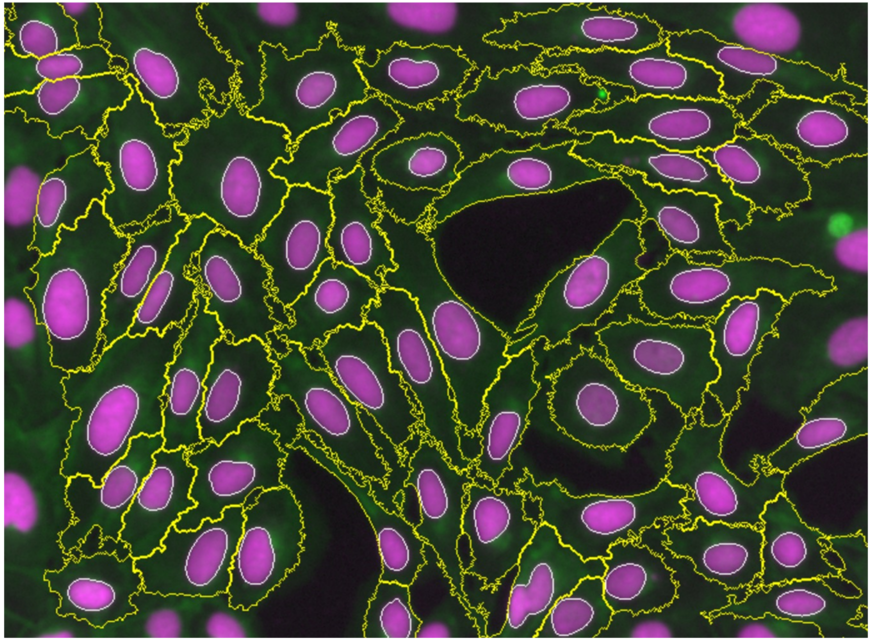
6 stains, 5 channels, 8 organelles/compartments

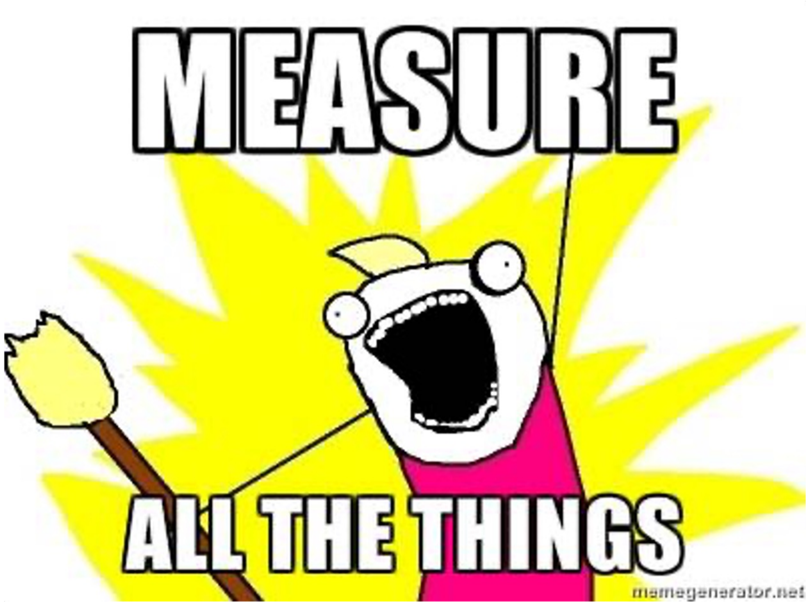
½ to ¼ cost of mRNA profiling (L1000)

Cimini, B. A., Chandrasekaran, S. N., Kost-Alimova, M., Miller, L., Goodale, A., Fritchman, B., Byrne, P., Garg, S., Jamali, N., Logan, D. J., Concannon, J. B., Lardeau, C.-H., Mouchet, E., Singh, S., Shafqat Abbasi, H., Aspesi, P., Jr, Boyd, J. D., Gilbert, T., Gnutt, D., … Carpenter, A. E. (2023). Optimizing the Cell Painting assay for image-based profiling. *Nature Protocols*. https://doi.org/10.1038/s41596-023-00840-9

Image-based Profiling with Cell Painting

Identify nuclei, cells, cytoplasm



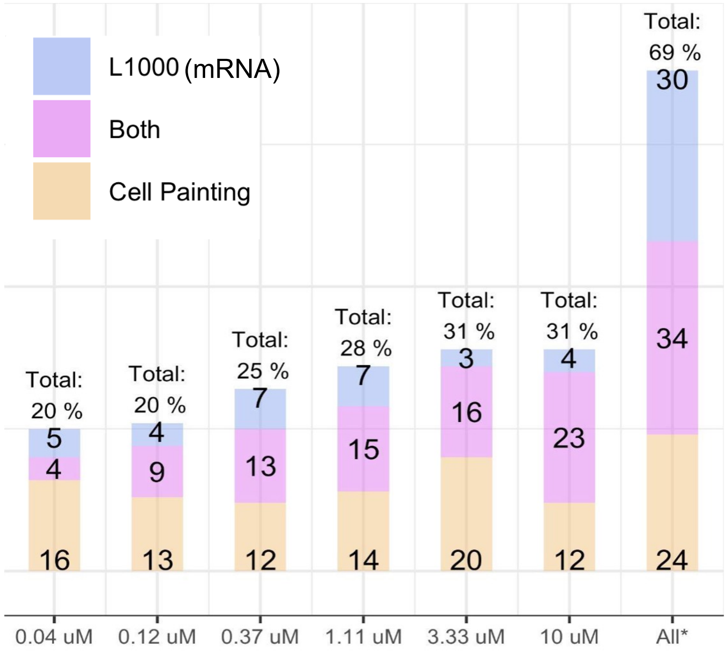


Cell Painting used for:

- predicting small molecule activity
- characterizing gene alleles within cells
- predicting cellular drug resistance
- distinguishing between disease states
- predicting outcomes of other assays
- ...

Cell Painting complements mRNA profiling

Unsupervised analysis: number (and %) of MOA classes whose compounds are self-similar



Way, G. P., Natoli, T., Adeboye, A., Litichevskiy, L., Yang, A., Lu, X., Caicedo, J. C., Cimini, B. A., Karhohs, K., Logan, D. J., Rohban, M. H., Kost-Alimova, M., Hartland, K., Bornholdt, M., Chandrasekaran, S. N., Haghighi, M., Weisbart, E., Singh, S., Subramanian, A., & Carpenter, A. E. (2022). Morphology and gene expression profiling provide complementary information for mapping cell state. *Cell Systems*, *13*(11), 911–923.e9. https://doi.org/10.1016/j.cels.2022.10.001

Cell Painting complements protein profiling

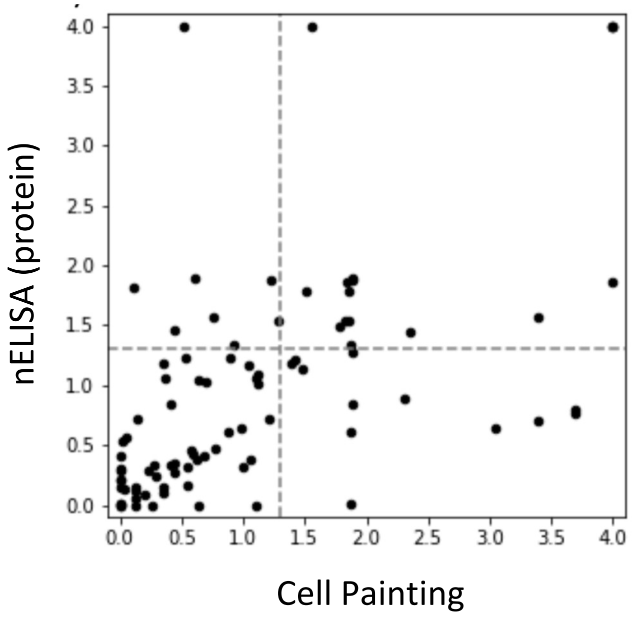
nELISA by Nomic bio

A next-generation, massively parallel ELISA

Current capabilities:

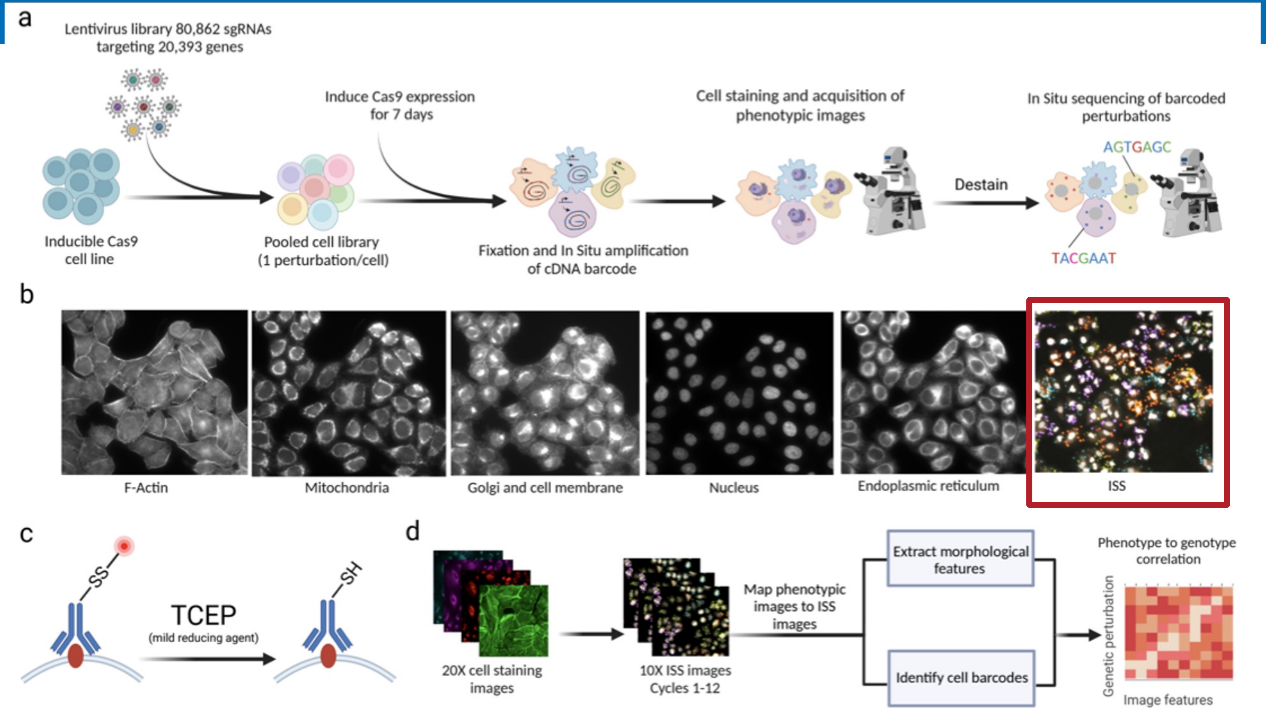
- 191-plex secretome panel

Predicting the mechanism of action for compounds



Dagher, M., Ongo, G., Robichaud, N., Kong, J., & Rho, W. (2023). nELISA: A high-throughput, high-plex platform enables quantitative profiling of the secretome. *bioRxiv*. https://www.biorxiv.org/content/10.1101/2023.04.17.535914.abstract

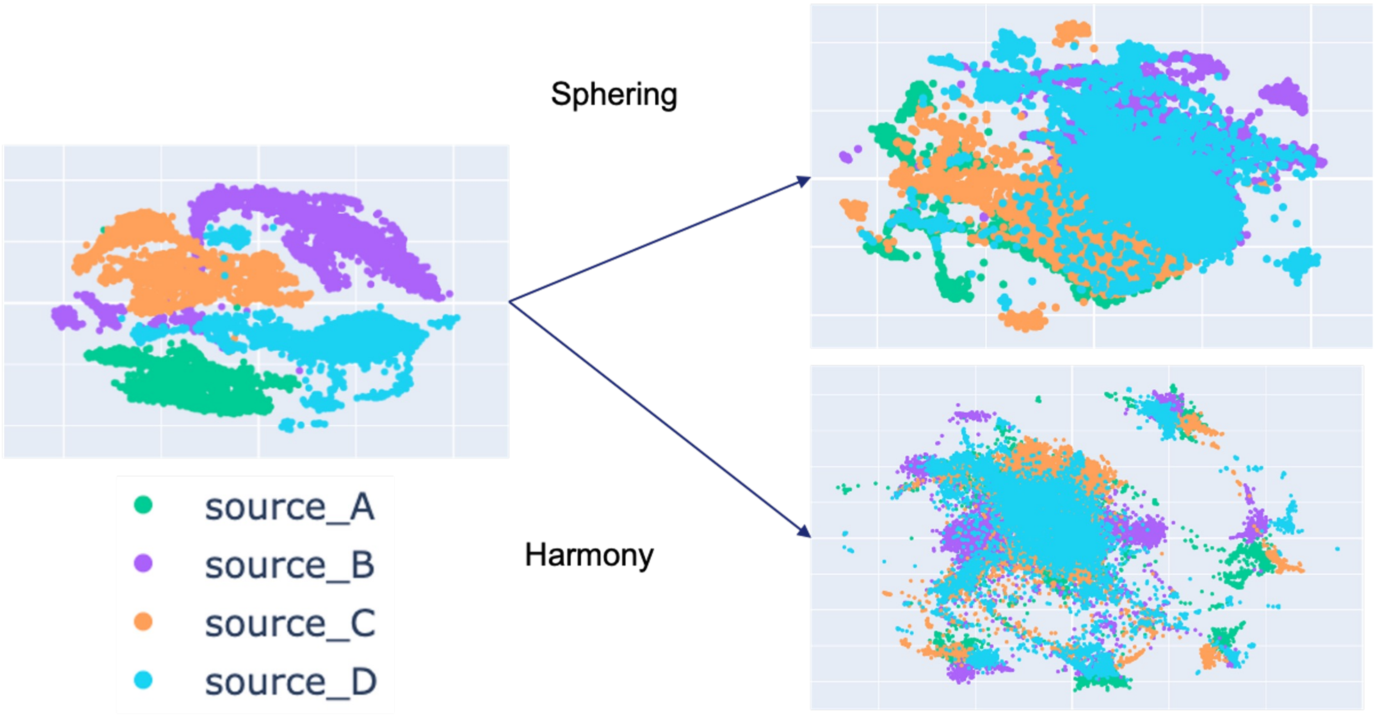
Pooled Cell Painting (PERISCOPE)



cost:
~$0.001/cell

Ramezani, M.*, Bauman, J.*, Singh, A.*, **Weisbart, E.***, Yong, J., Lozada, M., Way, G. P., Kavari, S. L., Diaz, C., Haghighi, M., Batista, T. M., Pérez-Schindler, J., Claussnitzer, M., Singh, S., Cimini, B. A., Blainey, P. C., Carpenter, A. E., Jan, C. H., & Neal, J. T. (2023). A genome-wide atlas of human cell morphology. In bioRxiv (p. 2023.08.06.552164). https://doi.org/10.1101/2023.08.06.552164

Bringing tools from other -omics into morphological profiling: Harmony



Arevalo, J., van Dijk, R., Carpenter, A. E., & Singh, S. (2023). Evaluating batch correction methods for image-based cell profiling. *bioRxiv : The Preprint Server for Biology*. https://doi.org/10.1101/2023.09.15.558001

How can we bring quantitative bioimaging to everyone?

FAIR data principles

Findable

- (meta)data are assigned a globally unique and persistent identifier
- data are described with rich metadata
- metadata clearly and explicitly include the identifier of the data it describes
- (meta)data are registered or indexed in a searchable resource

Accessible

- (meta)data are retrievable by their identifier using a standardized communications protocol
- the protocol is open, free, and universally implementable
- the protocol allows for an authentication and authorization procedure, where necessary
- metadata are accessible, even when the data are no longer available

Interoperable

- (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- (meta)data use vocabularies that follow FAIR principles
- (meta)data include qualified references to other (meta)data

Reusable

- meta(data) are richly described with a plurality of accurate and relevant attributes
- (meta)data are released with a clear and accessible data usage license
- (meta)data are associated with detailed provenance
- (meta)data meet domain-relevant community standards

Cell Painting Gallery on RODA

Registry of Open Data on AWS

Cell Painting Gallery

bioinformatics | biology | cancer | cell biology | cell imaging | cell painting | chemical biology | computer vision | csv | deep learning | fluorescence imaging | genetic | high-throughput imaging | image processing | image-based profiling | imaging | machine learning | medicine | microscopy | organelle

Description

The Cell Painting Gallery is a collection of image datasets created using the Cell Painting assay. The images of cells are captured by microscopy imaging, and reveal the response of various labeled cell components to whatever treatments are tested, which can include genetic perturbations, chemicals or drugs, or different cell types. The datasets can be used for diverse applications in basic biology and pharmaceutical research, such as identifying disease-associated phenotypes, understanding disease mechanisms, and predicting a drug's activity, toxicity, or mechanism of action (Chandrasekaran et al 2020). This collection is maintained by the Carpenter–Singh lab and the Cimini lab at the Broad Institute. A human-friendly listing of datasets, instructions for accessing them, and other documentation is at the corresponding GitHub page about the Gallery.

Update Frequency

Typically when an associated publication is posted on biorxiv

License

CC0 1.0 Universal (CC0 1.0) Public Domain Dedication, but please do cite the corresponding publication for each dataset, as listed here.

Documentation

https://github.com/broadinstitute/cellpainting-gallery

Resources on AWS

Description

Cell Painting data, comprising fluorescence microscopy cell images (TIFF), extracted features (CSV), and associated metadata (CSV and TXT).

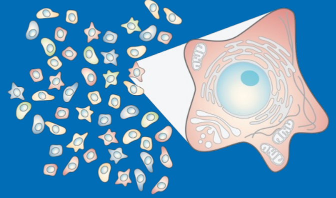
18 complete datasets

13 in-progress datasets

671.1 TB data

JUMP in Cell Painting Gallery

Joint Undertaking for Morphological Profiling

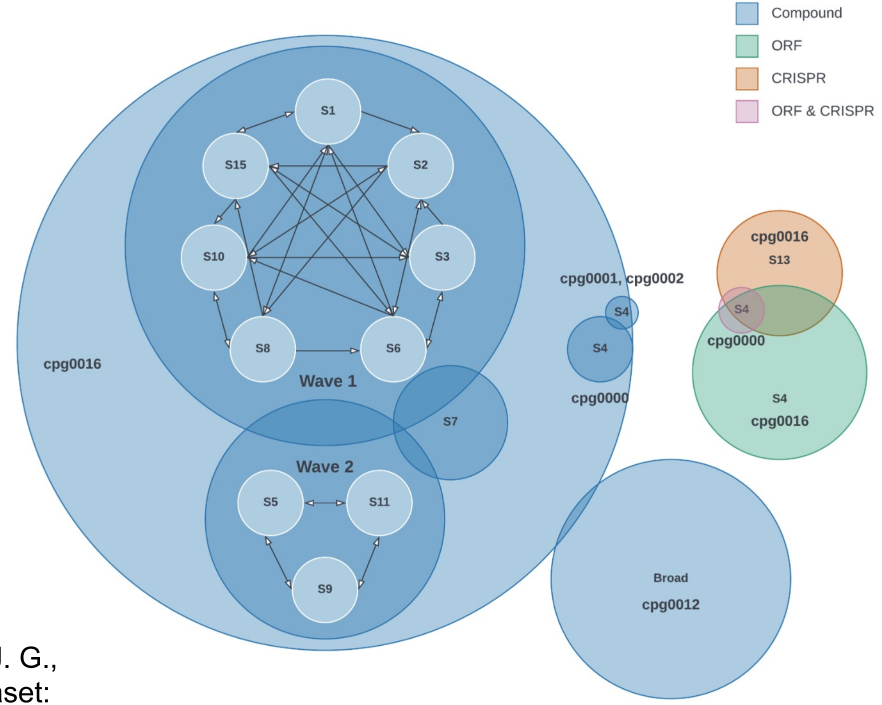


13,000 overexpressed genes

8,000 CRISPR knockdowns

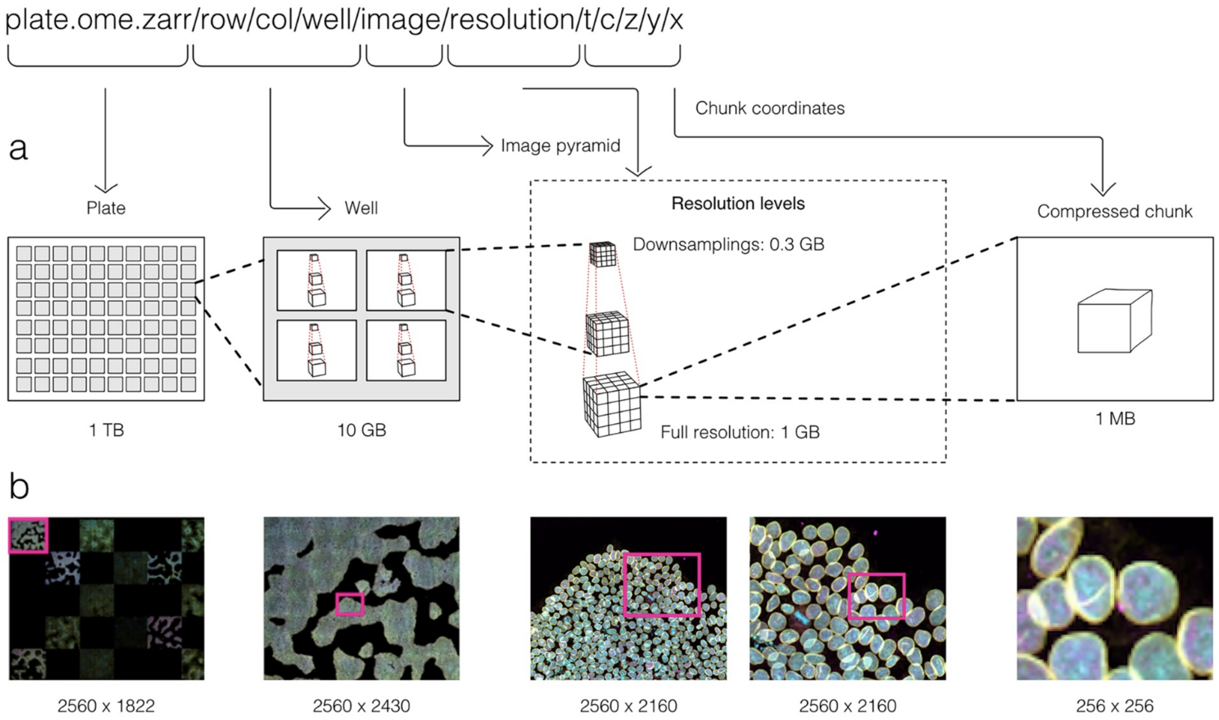
116,000 (+ 30,000) small molecules

Raw images, classical features, deep-learning features



Chandrasekaran, S. N., Ackerman, J., Alix, E., Michael Ando, D., Arevalo, J., Bennion, M., Boisseau, N., Borowa, A., Boyd, J. D., Brino, L., Byrne, P. J., Ceulemans, H., Ch'ng, C., Cimini, B. A., Clevert, D.-A., Deflaux, N., Doench, J. G., Dorval, T., Doyonnas, R., … Carpenter, A. E. (2023). JUMP Cell Painting dataset: morphological impact of 136,000 chemical and genetic perturbations. In *bioRxiv* (p. 2023.03.23.534023). https://doi.org/10.1101/2023.03.23.534023

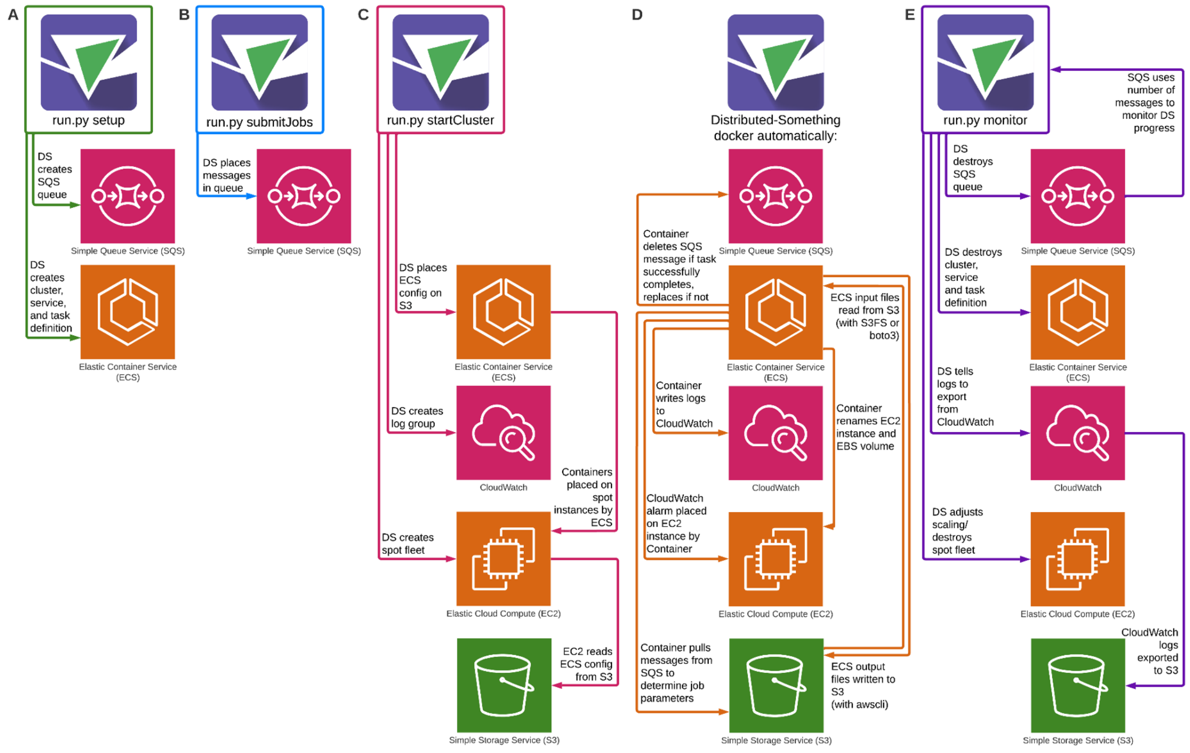
NGFF: OME-Zarr



Moore, J., Basurto-Lozada, D., Besson, S., Bogovic, J., Bragantini, J., Brown, E. M., Burel, J.-M., Casas Moreno, X., de Medeiros, G., Diel, E. E., Gault, D., Ghosh, S. S., Gold, I., Halchenko, Y. O., Hartley, M., Horsfall, D., Keller, M. S., Kittisopikul, M., Kovacs, G., … Swedlow, J. R. (2023). OME-Zarr: a cloud-optimized bioimaging file format with international community support. *Histochemistry and Cell Biology*. https://doi.org/10.1007/s00418-023-02209-1

Distributed-Something

Unlimited compute power/storage, low computational expertise

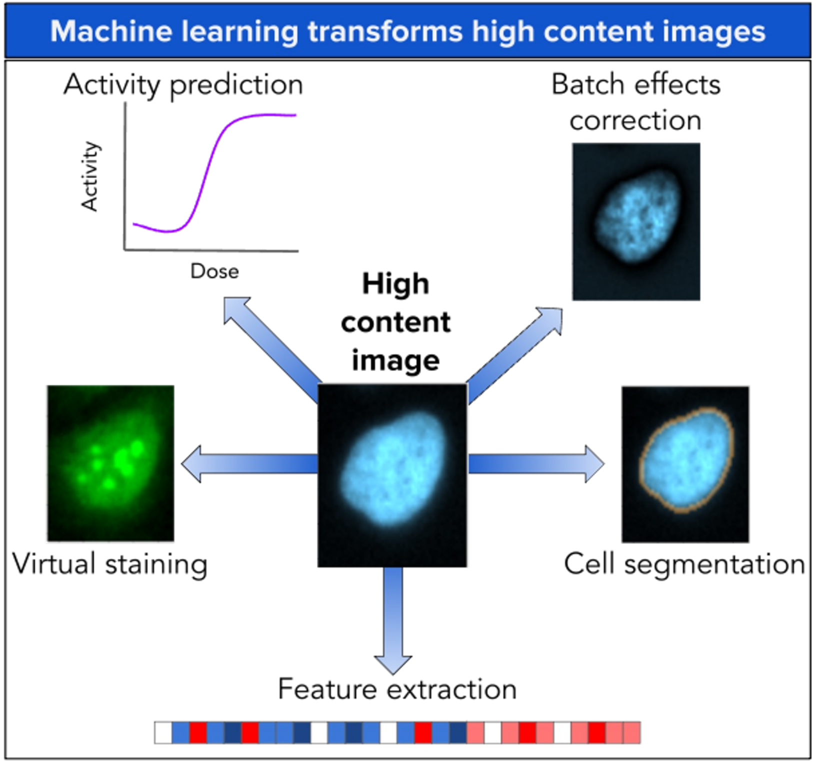


Weisbart, E., & Cimini, B. A. (2023). Distributed-Something: scripts to leverage AWS storage and computing for distributed workflows at scale. *Nature Methods*. https://doi.org/10.1038/s41592-023-01918-8

How has deep learning changed imaging?

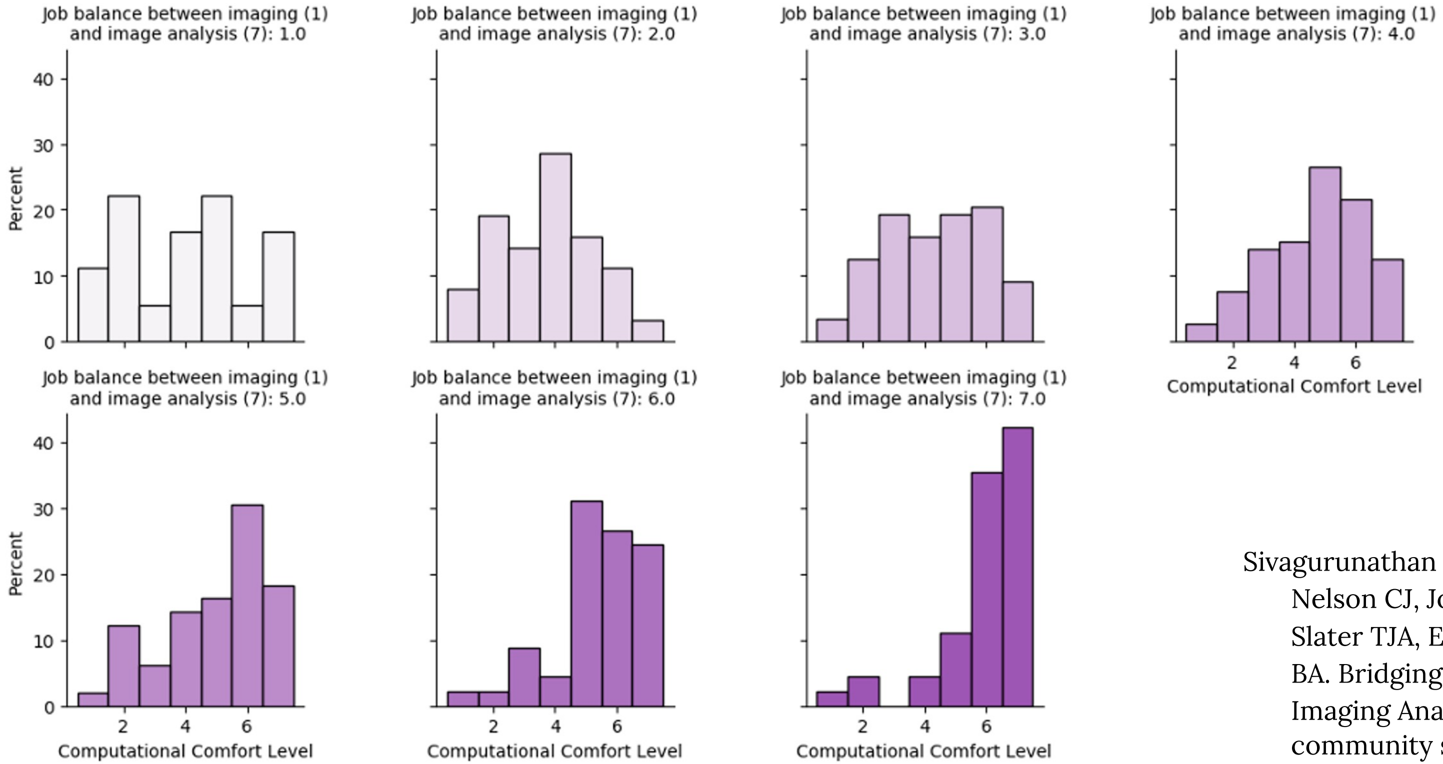
Deep learning can help us get data from bioimages in many different ways

Caicedo and Cimini, arxiv 2023



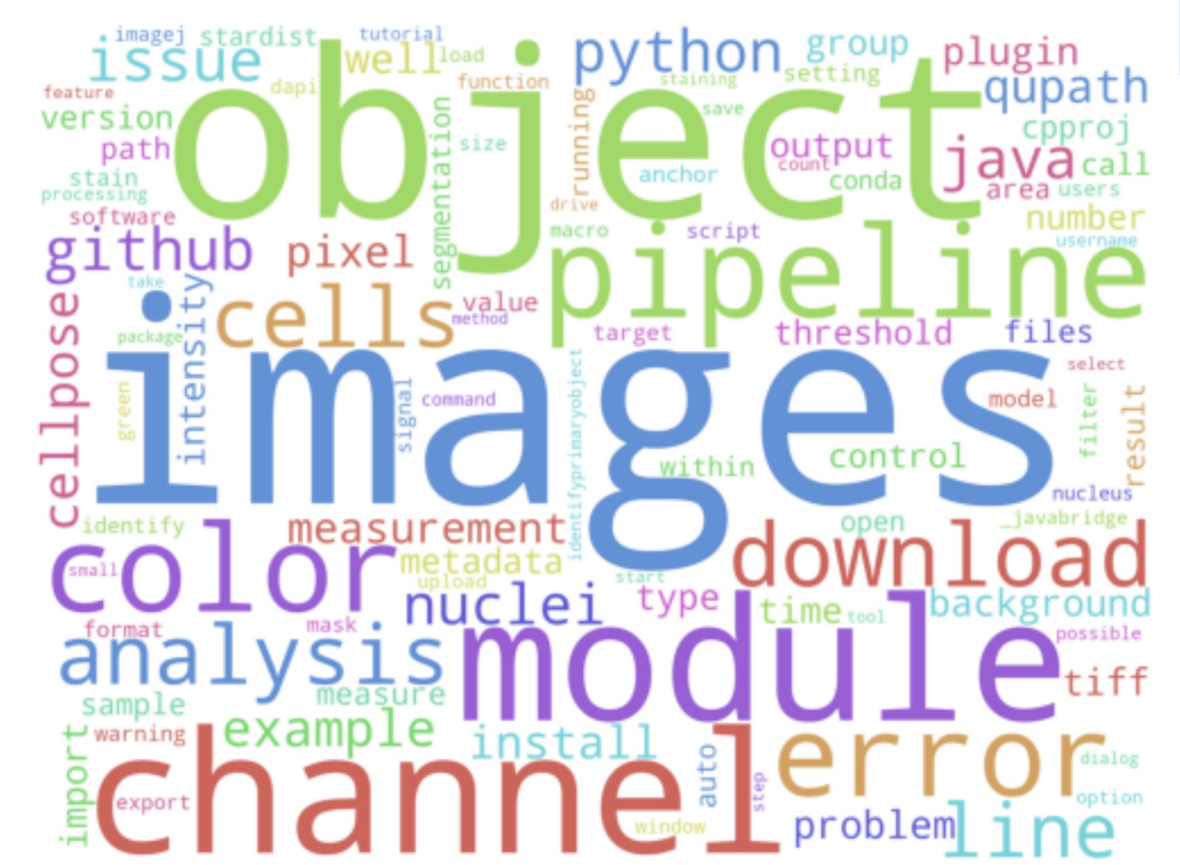
How can we bring everyone into the deep learning revolution?

Scientists generating data are less computationally comfortable



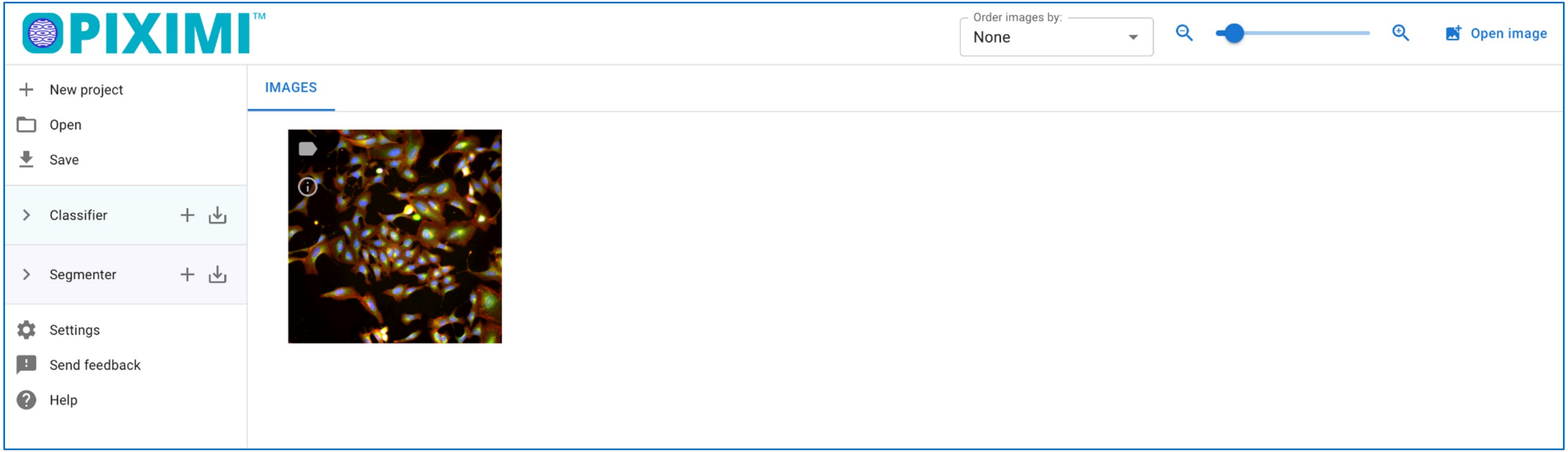
Sivagurunathan S, Marcotti S, Nelson CJ, Jones ML, Barry DJ, Slater TJA, Eliceiri KW, Cimini BA. Bridging Imaging Users to Imaging Analysis - A community survey. *J Microsc.* 2023 Sep 20.

Simplify CellProfiler plugin installation

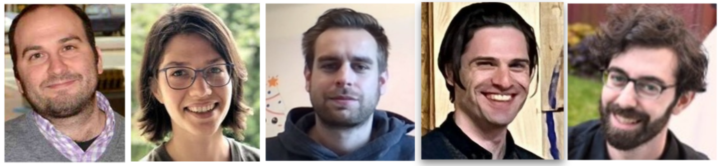


Weisbart, E.*, Tromans-Coia, C.*, Diaz-Rohrer, B.*, Stirling, D.R.*, Garcia-Fossa, R.*, Senft, R.A.*, Hiner, M.C., de Jesus, M.B., Eliceiri, K.W., & Cimini, B.,A. (2023) CellProfiler plugins - an easy image analysis platform integration for containers and Python tools. J. Microsc. https://doi.org/10.1111/jmi.13223

PIXIMI: Images to Discovery



https://www.piximi.app/
https://github.com/piximi/piximi



Allen Goodman | Alice Lucas | Levin Moser | Nodar Gogoberidze | Andréa Papaleo

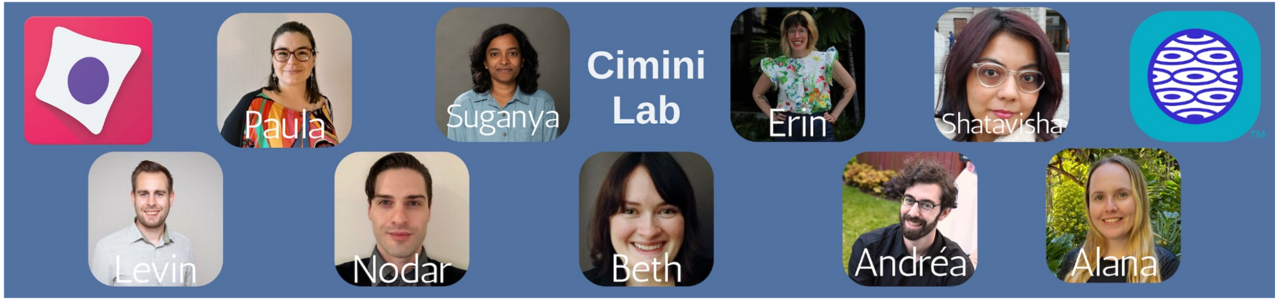
Acknowledgements

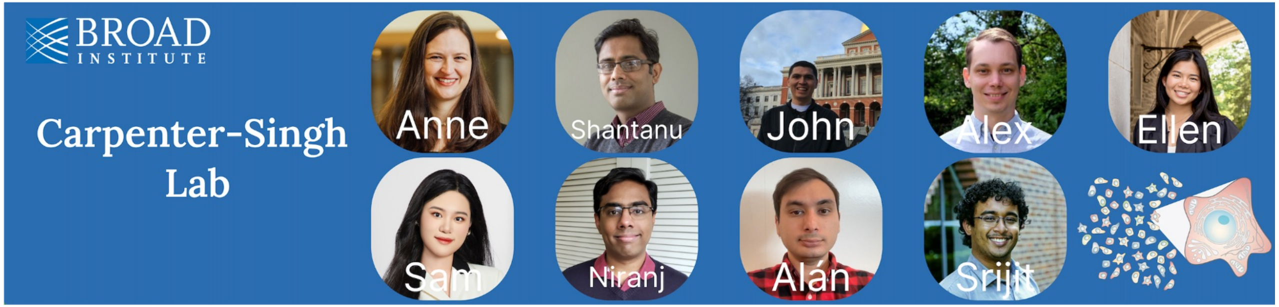
Cimini lab members

Beth Cimini
Shatavisha Dasgupta
Nodar Gogoberidze
Alana MacMahon
Levin Moser
Eddy Onyango
Andréa Papaleo
Suganya Sivagurunathan
Erin Weisbart

Carpenter-Singh lab members

Anne Carpenter
Shantanu Singh
John Arevalo
Niranj Chandrasekaran
Sam Chen
Alexander Kalinin
Ank Kumar
Alana MacMahon
Alán Muñoz González
Srijit Seal
Ellen Su





Recent major funding for this work provided by:

- CZI Imaging Scientist Fellowship
- NIH NIGMS: MIRA R35 GM122547
- CZI Software Fellows program
- NIH NIGMS: P41 GM135019
- Massachusetts Life Sciences Center

BROAD INSTITUTE