Data science in cell imaging
Lecture 1: course overview

“The Great Wave off Kanagawa”, by Hokusai, ~1830 (Source: Wikipedia)

All slides are open under the cc-by license

PPTX slides available here

Welcome to the course data science in cell imaging!

- Assaf Zaritsky, SISE, BGU
- Lab of computational cell dynamics
- Email: assafzar@gmail.com
- Office: 96-308

Course's theme

This course will review the state-of-the-art in visualizing, processing, integrating and mining massive cell image data sets, **deciphering complex patterns and turning them into new biological knowledge**. It will include a mix of computational approaches (e.g., machine learning, computer vision) applied to bio-imaging data.

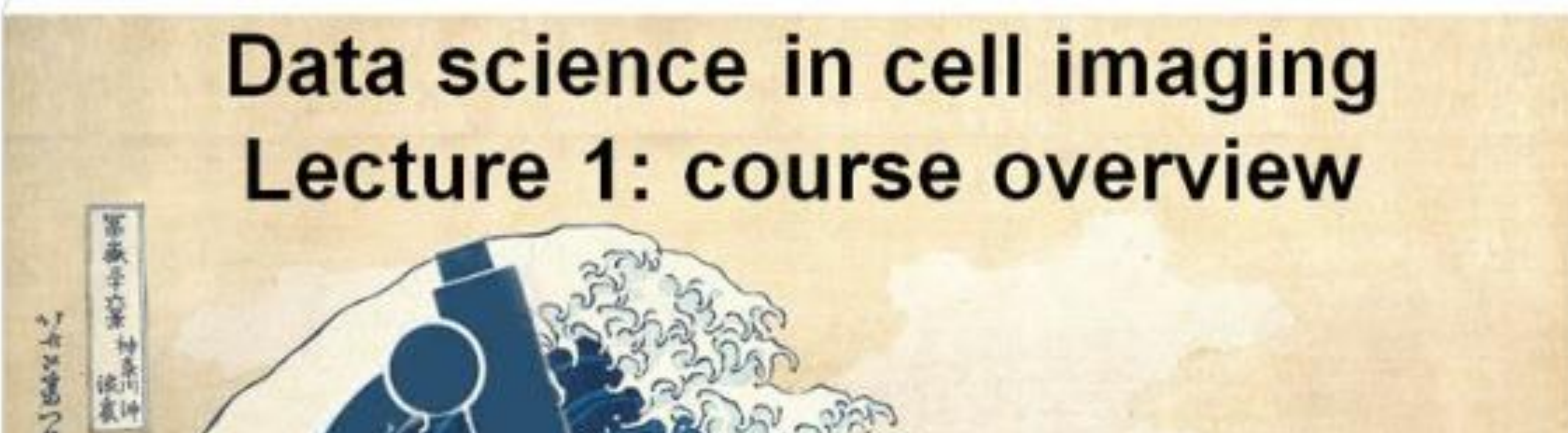
Before we start: quick Q&A

Q	**A**
Why are you speaking English?	Science is communicated in English (also useful for industry)
Do I need any background in biology?	No. I assume no prior knowledge in biology.
Do I need computational background?	Yes. I am encouraging students from diverse background to join, I will try to keep it as simple and as intuitive as possible. But this course main target audience are computational students.
Obligatory attendance?	No! But you'll find it hard to follow if you do not attend
I am looking for an easy course	Look elsewhere

Assaf Zaritsky
@AssafZaritsky

Preparing slides for the first lecture in my brand new course on Data Science in Cell Imaging starting next week @bengurionu!

Very curious to see if I can get computational grad students (with no BIO background) interested in the interface of computation and cell biology :-)

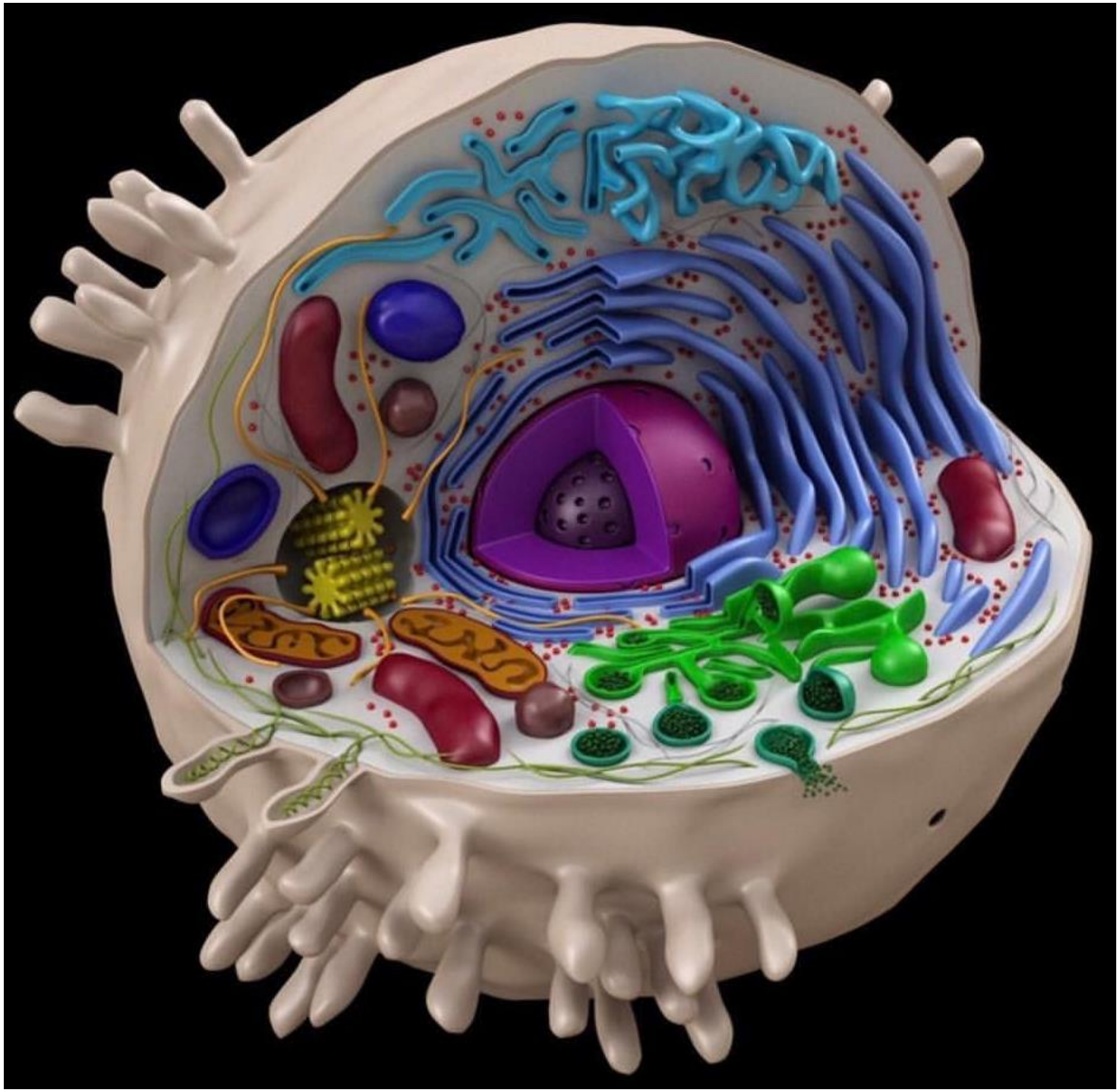


Today

- Introduction
- Course objectives and admin.
- Course (tentative) schedule and overview

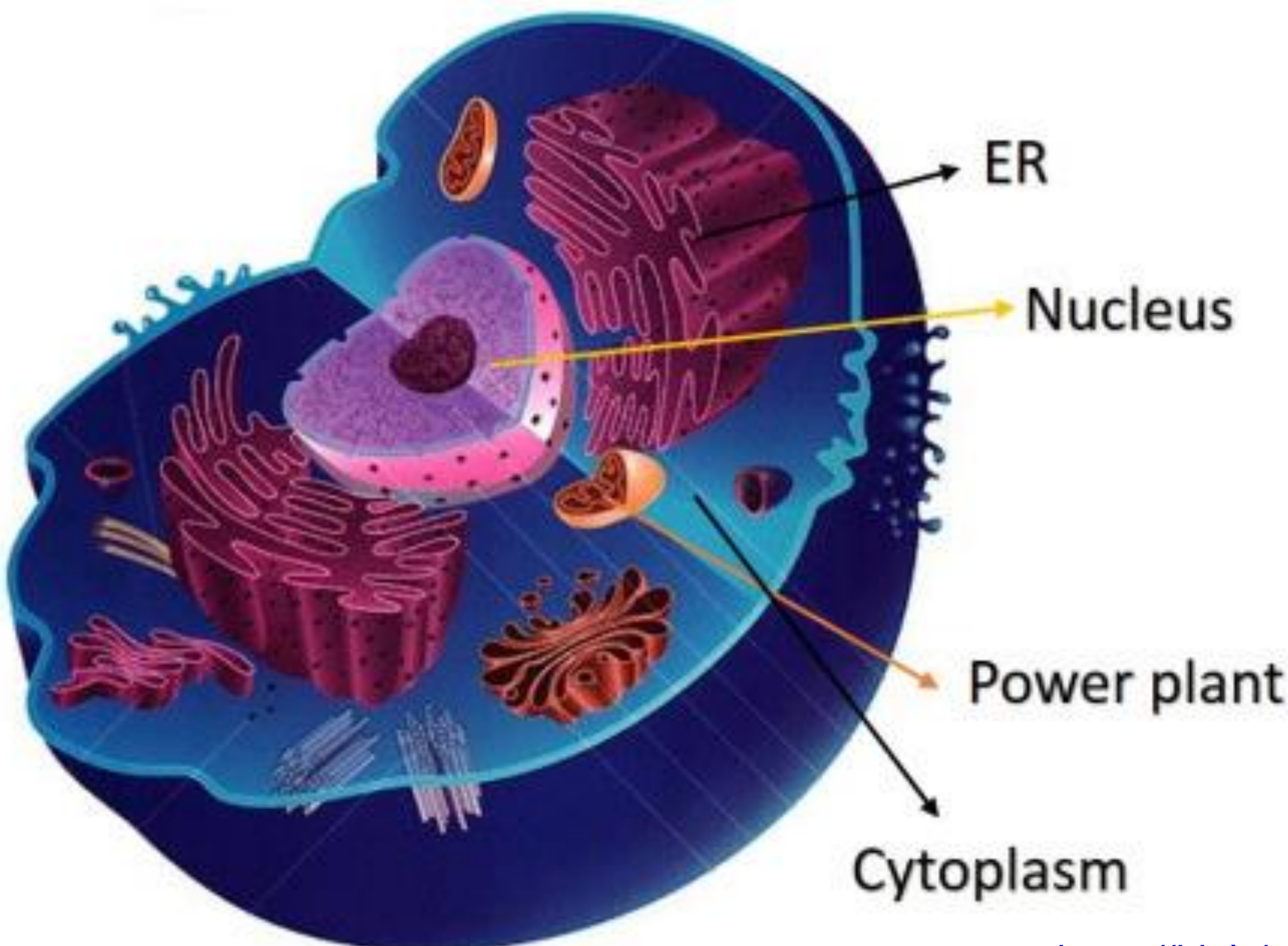
Broad introduction

Cells are the fundamental unit of structure and function in organisms



Source: https://bit.ly/2uVBE05

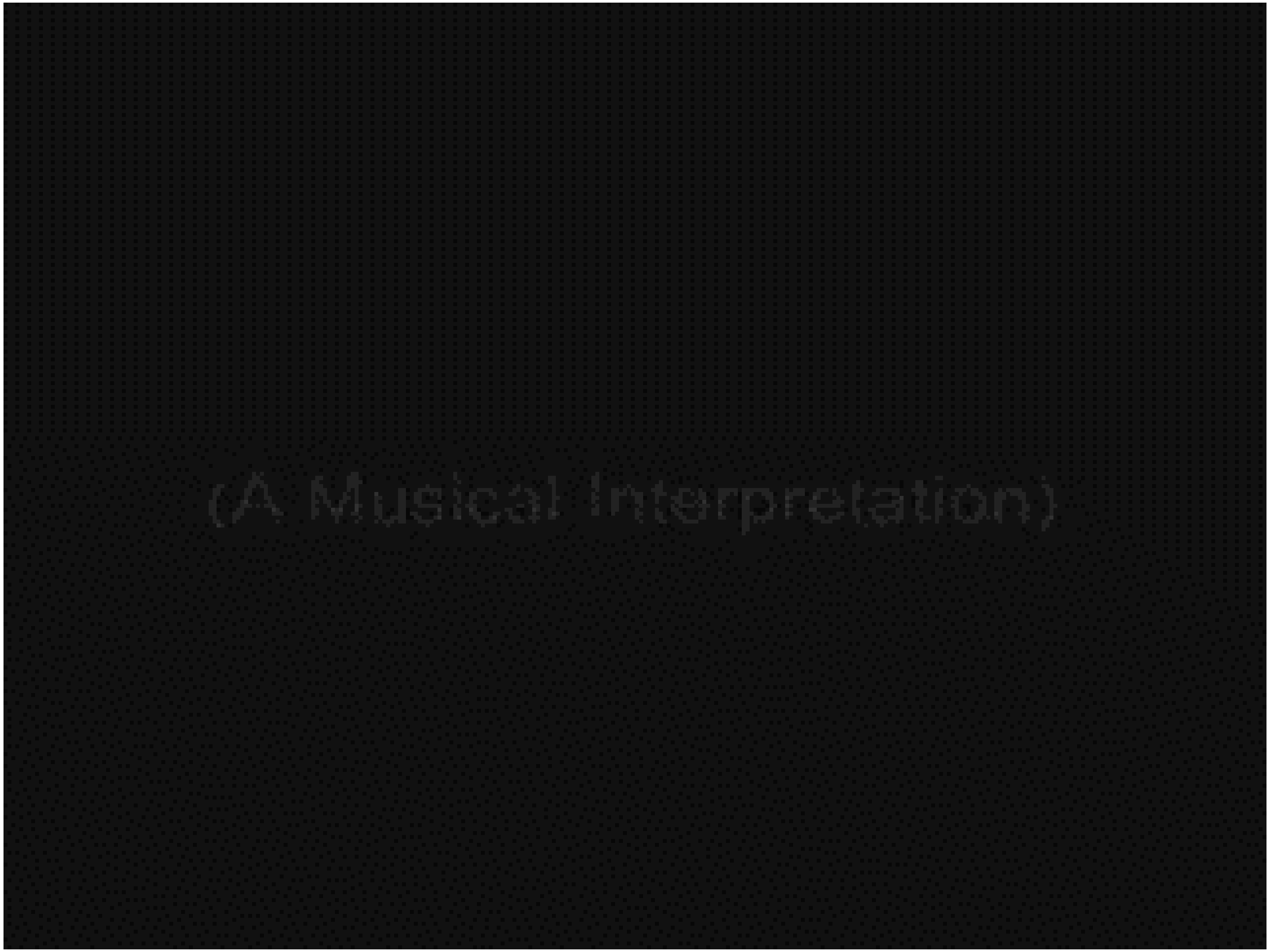
Proteins are molecular machines that make our cells tick



Source: https://bit.ly/2UDmGKT

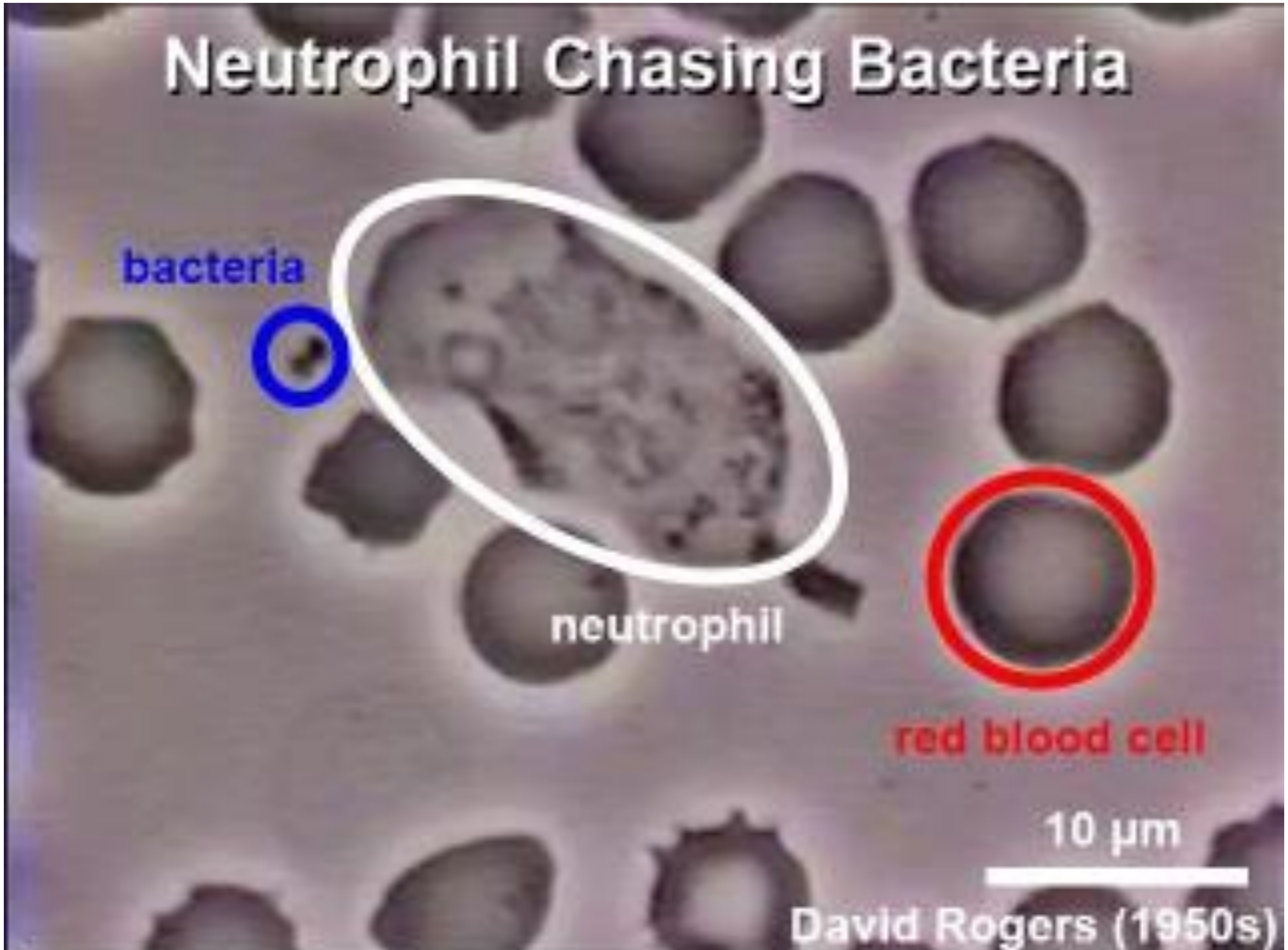
Neutrophil (white blood cell) chasing bacteria

sensing, information processing, decision making



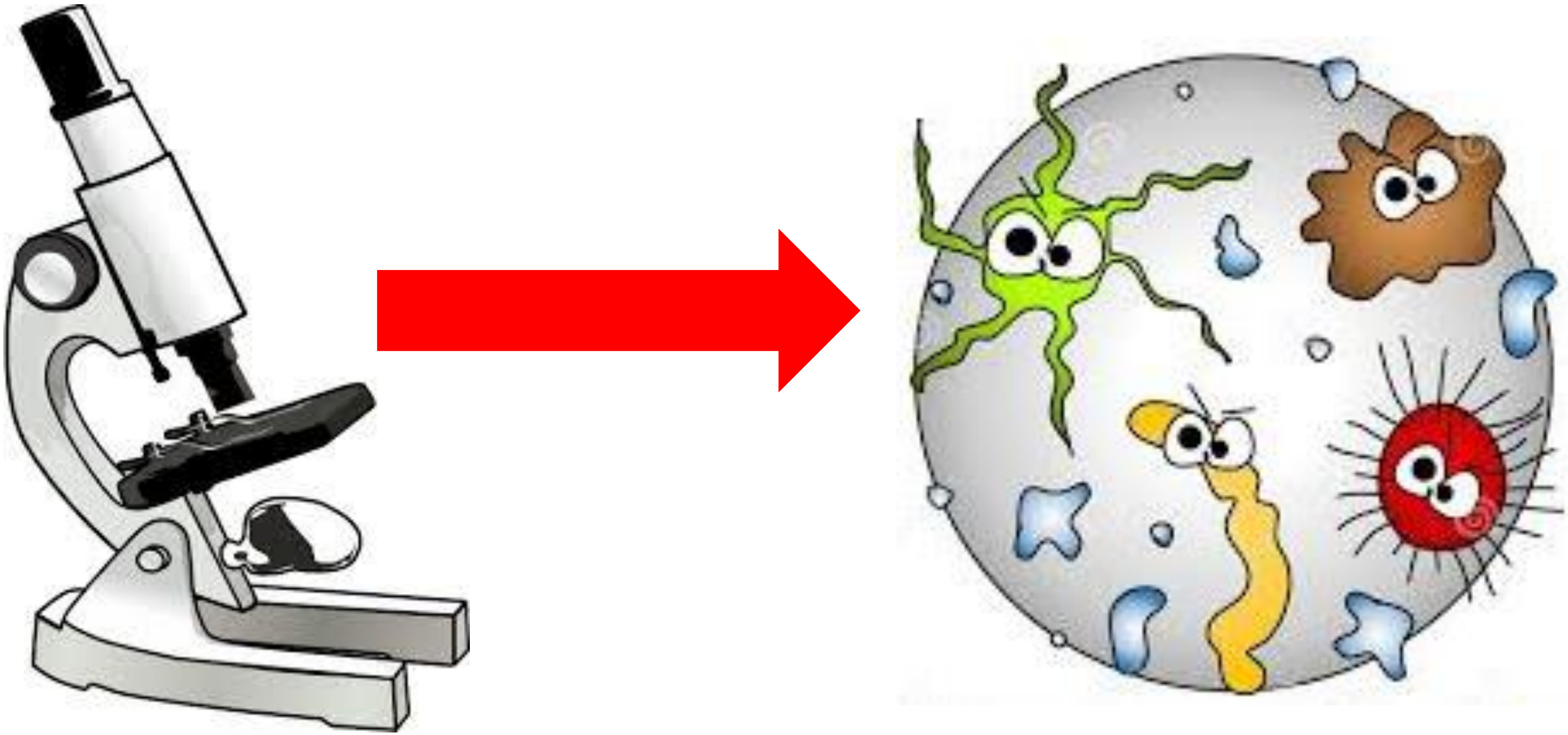
Original movie made in the 1950s by the late David Rogers at Vanderbilt University

Sensing, information processing, decision making



Original movie made in the 1950s by the late David Rogers at Vanderbilt University

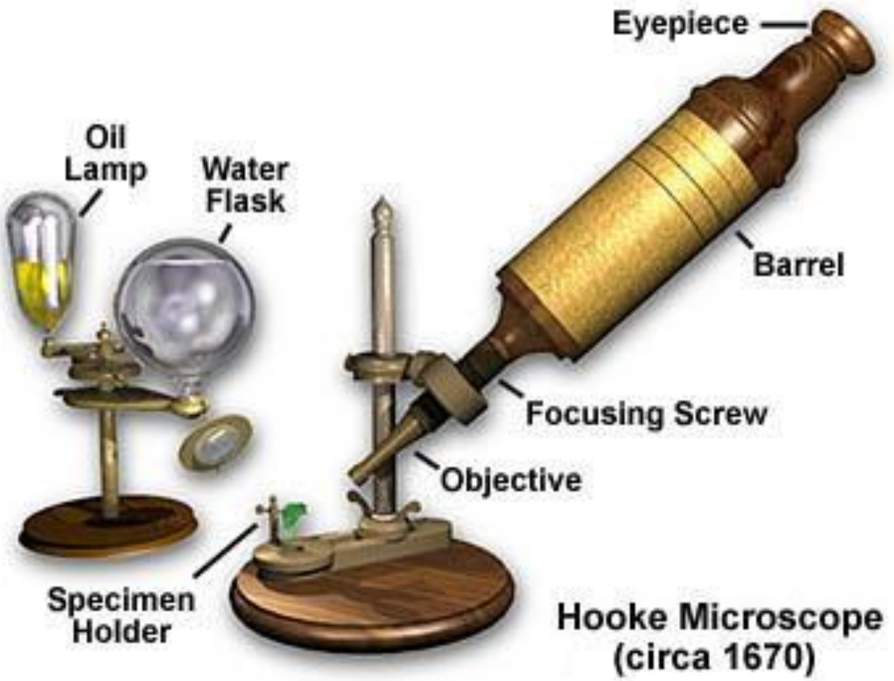
Seeing is believing



John Clarke (1639)

Robert Hooke, “the father of microscopy” (1665)

The term “Cell” was suggested due to the resemblance of plant cells to cells of a honeycomb. (These were actually dead cells)

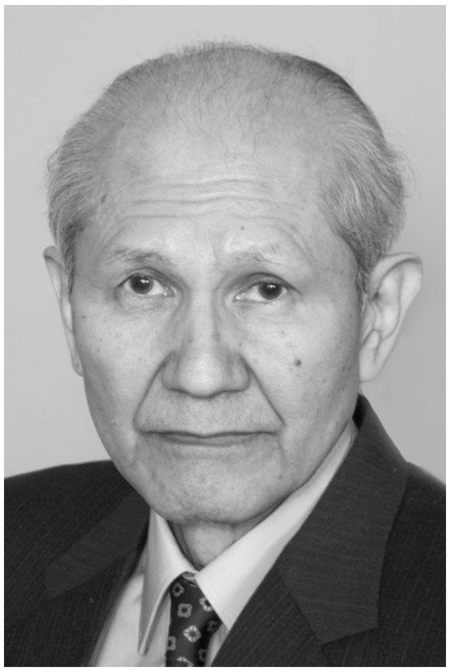


Hooke Microscope (circa 1670)

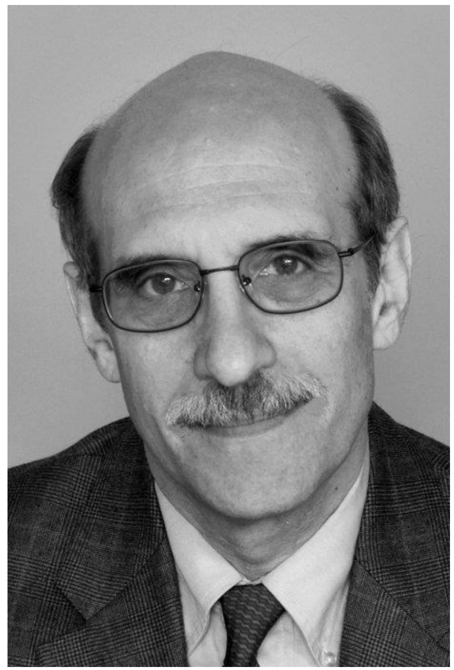
Natalie Elia

Green fluorescent protein (GFP)

Nobel Prize 2008, Chemistry



Osamu Shimomura



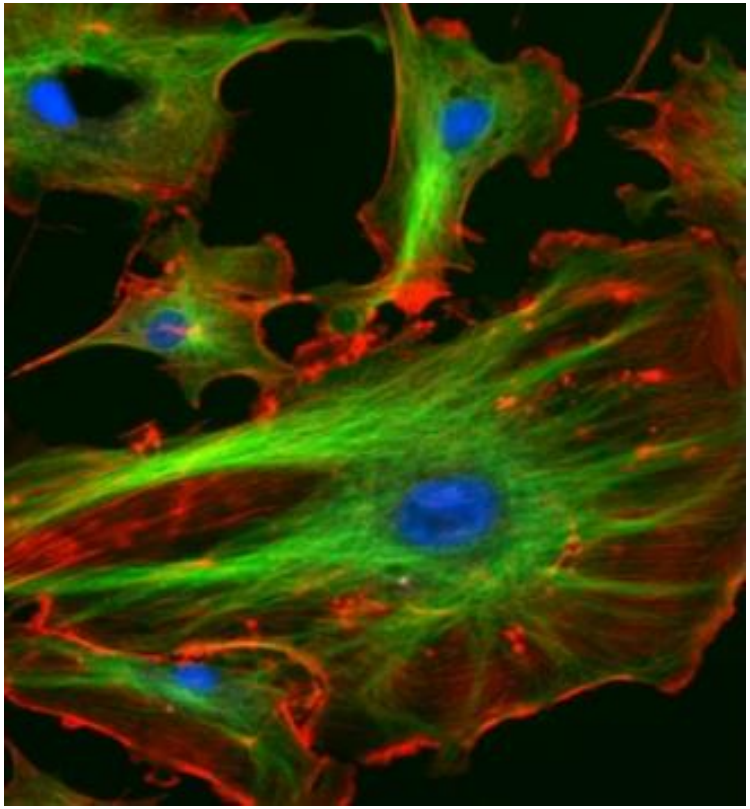
Martin Chalfie

Roger Y. Tsien

“for the discovery and development of the green fluorescent protein, GFP.”

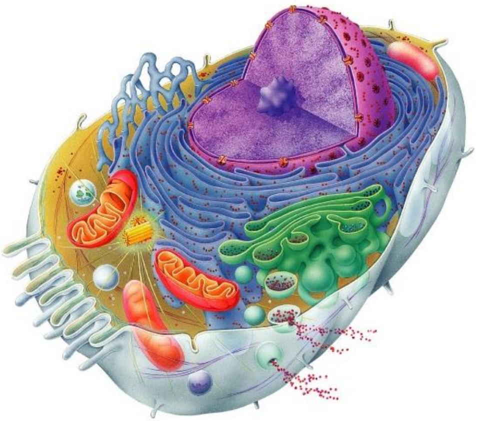
“Seeing” molecules with fluorescence microscopy

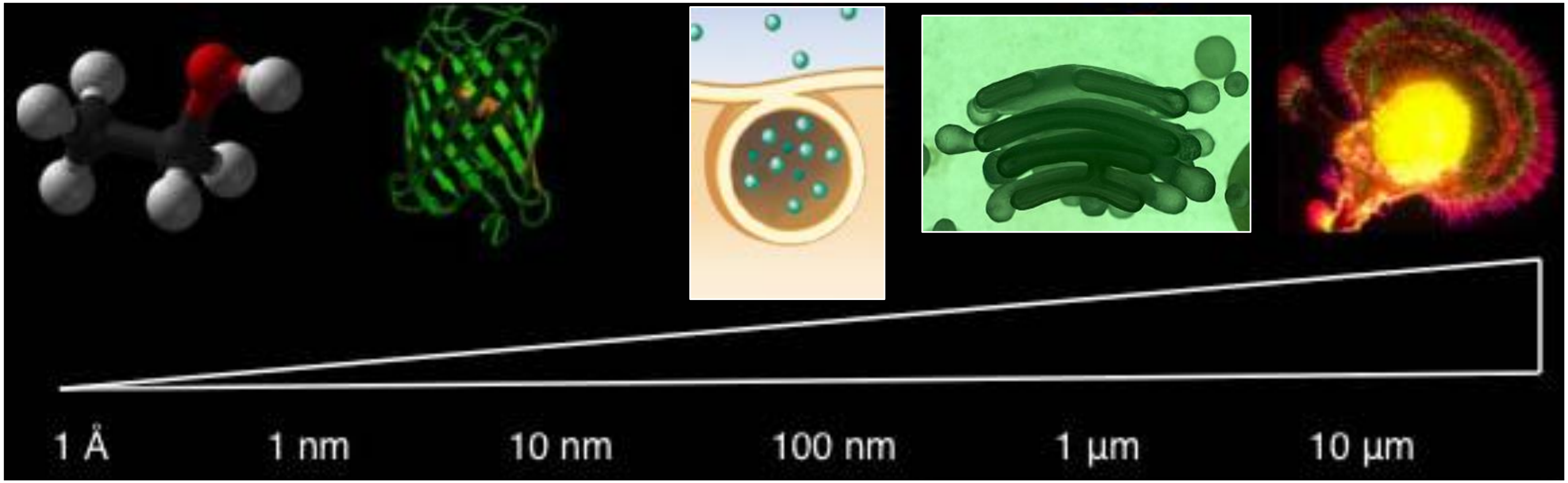
Green fluorescent protein,
Nobel prize in Chemistry (2008)



Source: https://bit.ly/2I19AB6

The cellular “ruler”



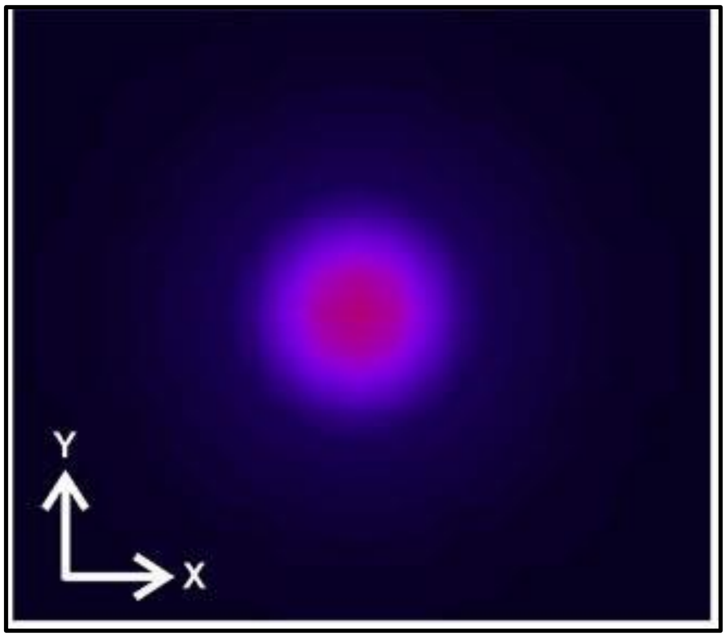


Light microscopy

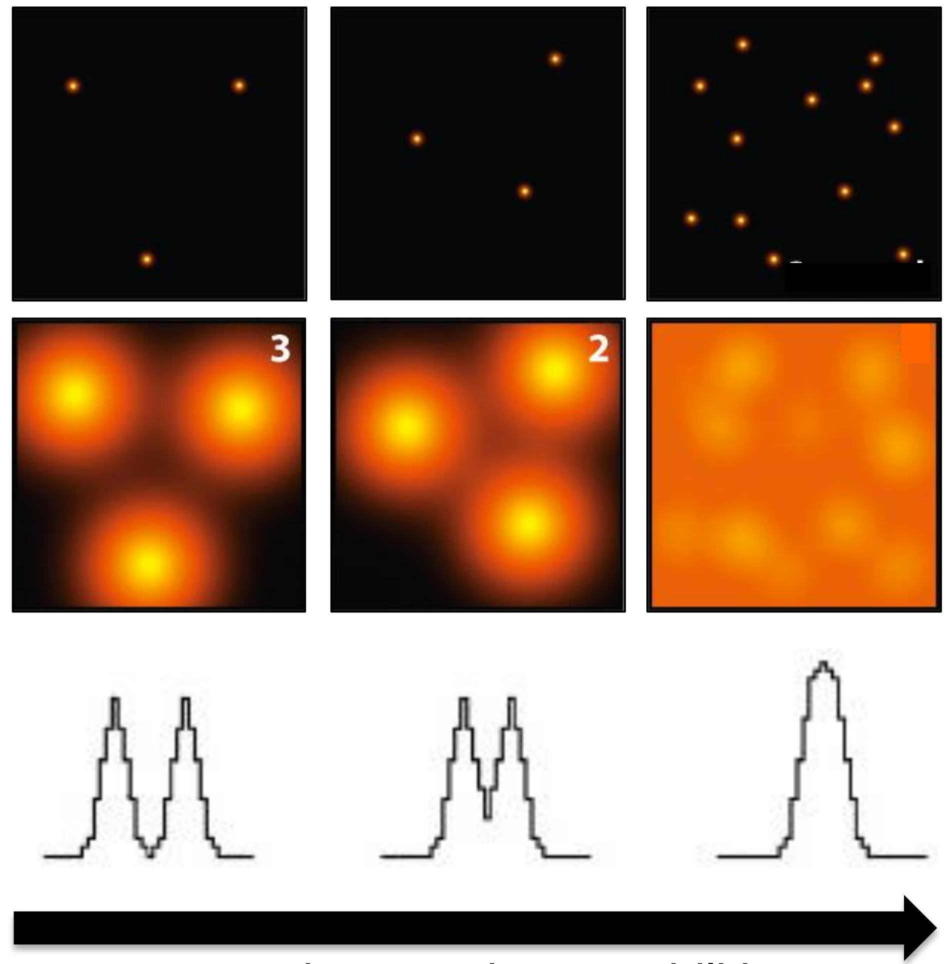
Natalie Elia

Light microscopy:

Resolution and PSF (point spread function)



XY resolution 250 nm



Decreased separation capabilities

“Super” resolution

Natalie Elia

Fluorescence super resolution microscopy

Nobel Prize 2014, Chemistry

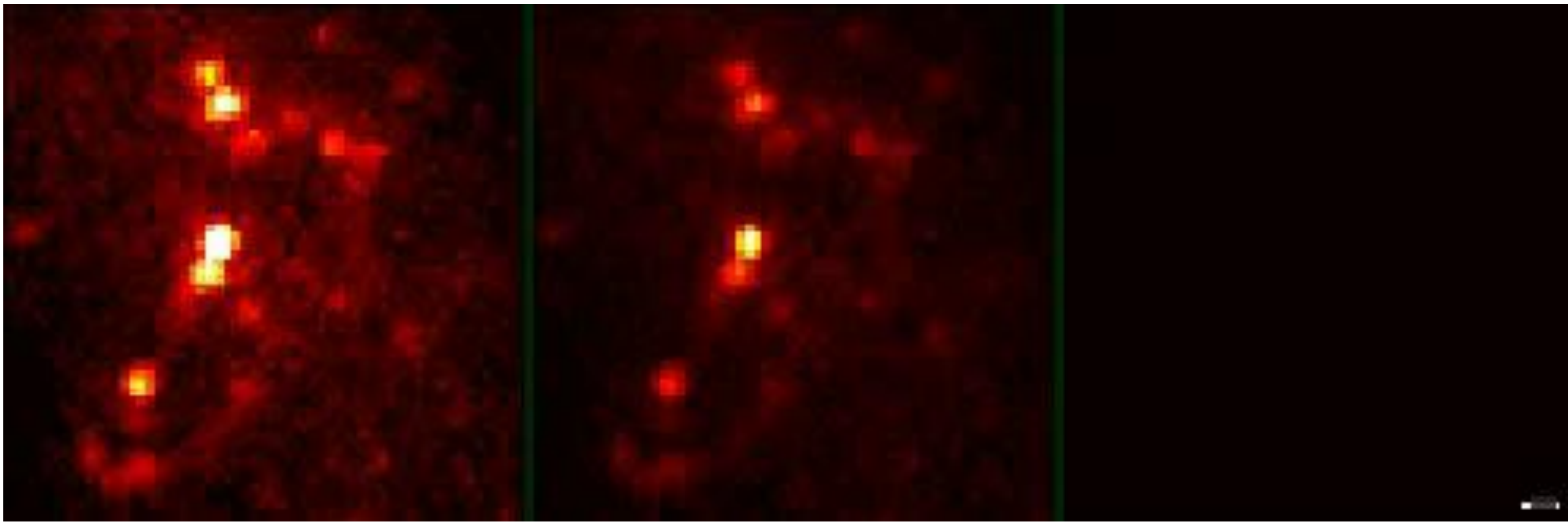
Eric Betzig

William Moerner

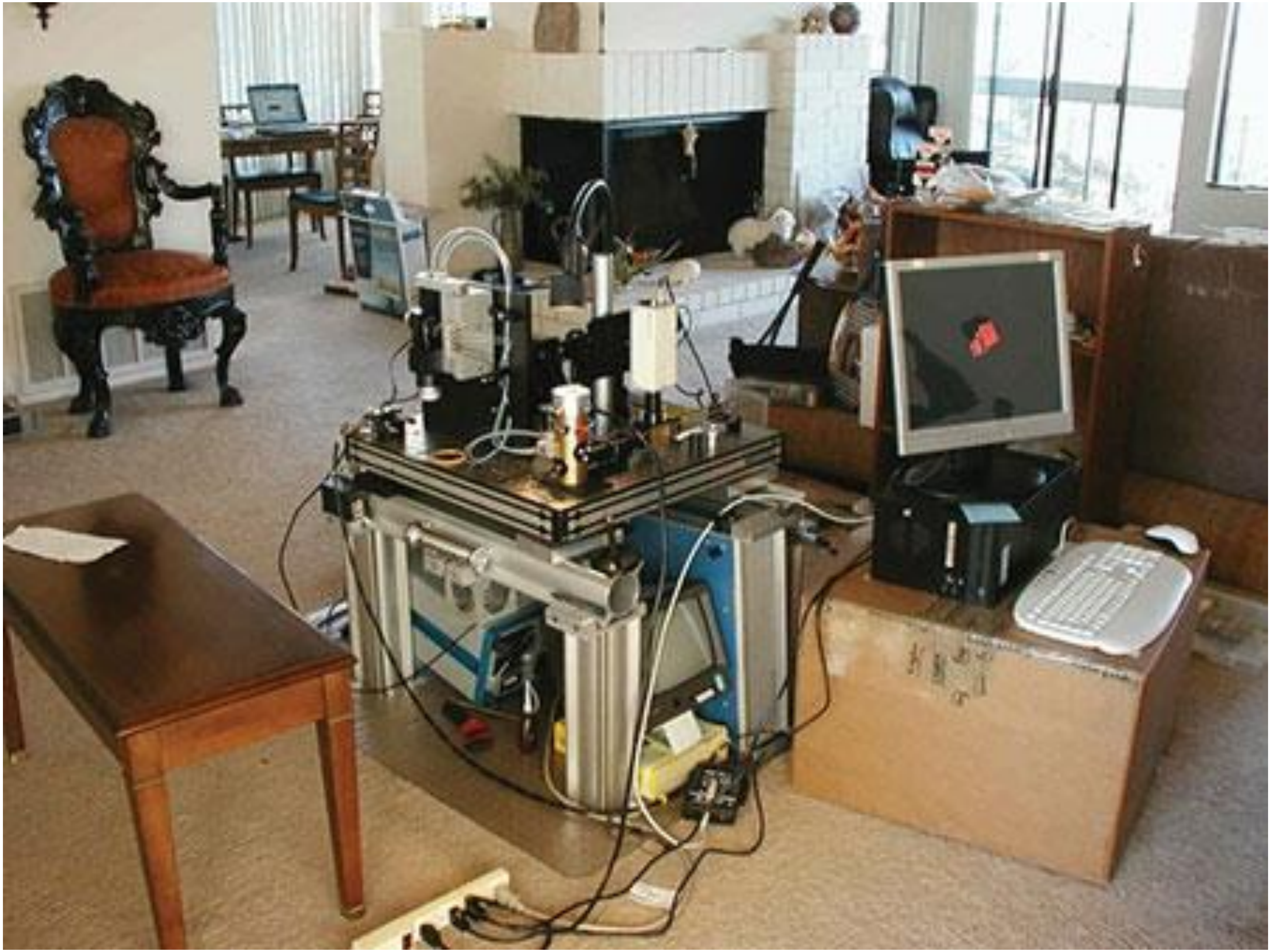
Stefan Hell

“for the development of super-resolved fluorescence microscopy”

Photoactivated localization microscopy (PALM)

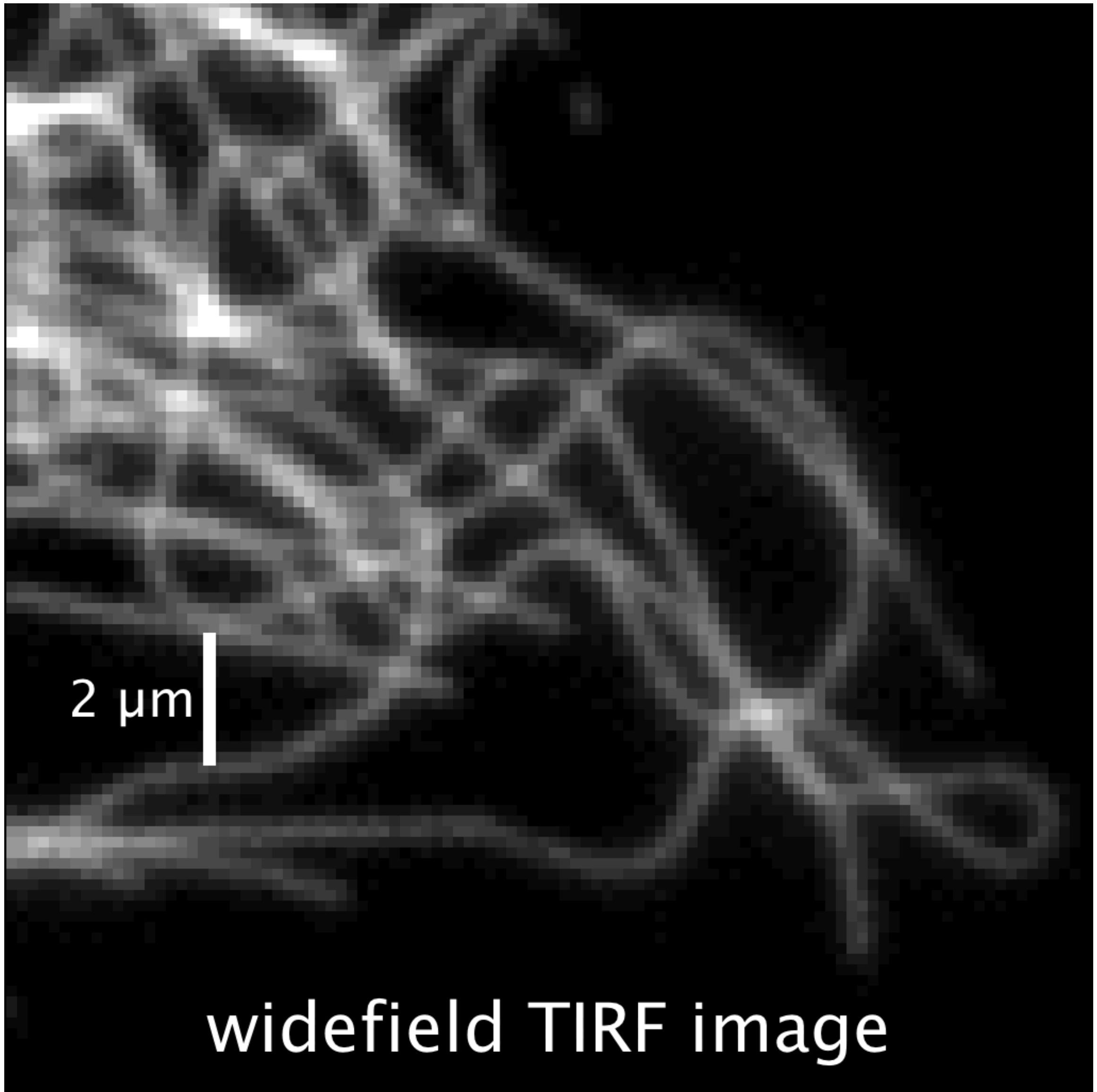


Betzig (2006)



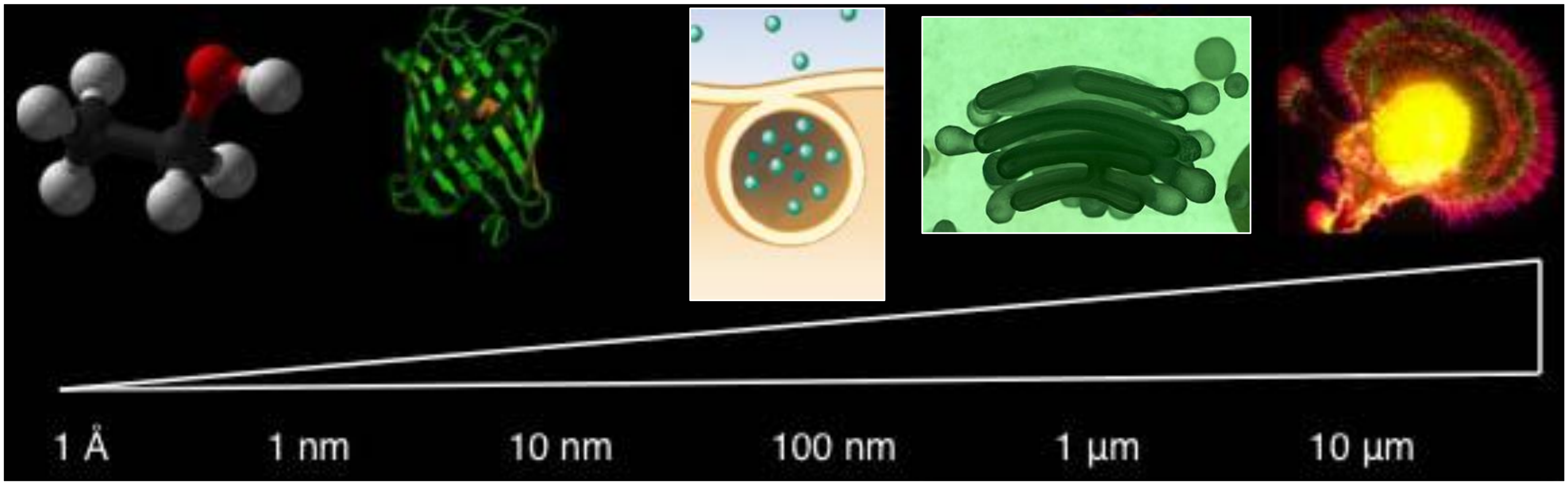
Source: https://www.nobelprize.org/prizes/chemistry/2014/betzig/biographical/

Stochastic optical reconstruction microscopy (STORM)



@christlet

The cellular “ruler” and resolution



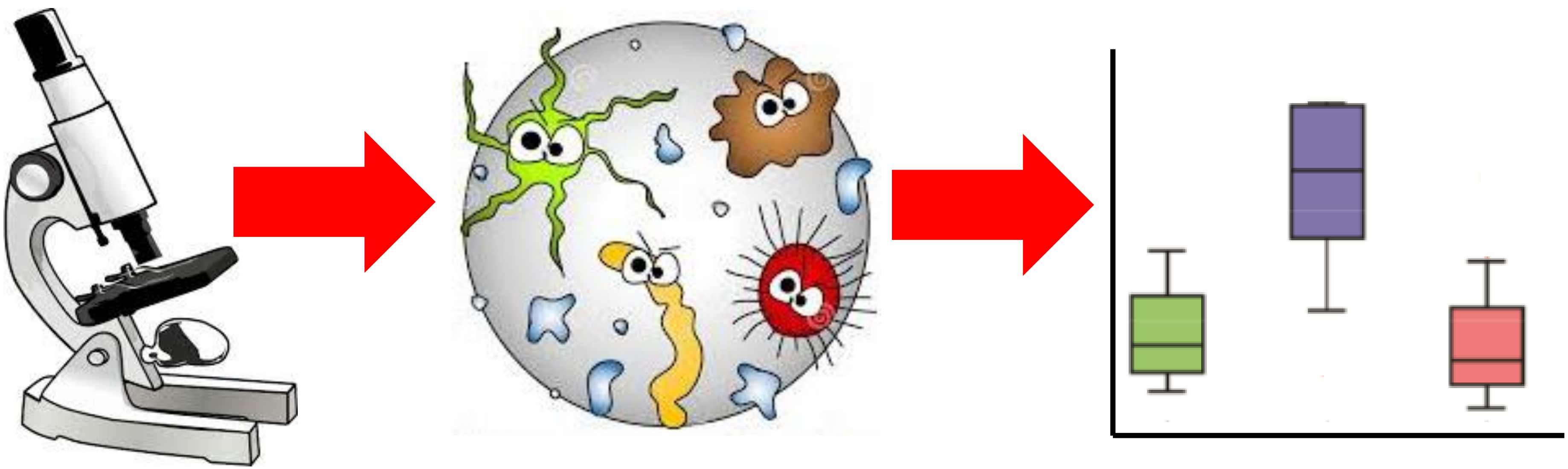
Light microscopy

Super resolution

Electron microscopy

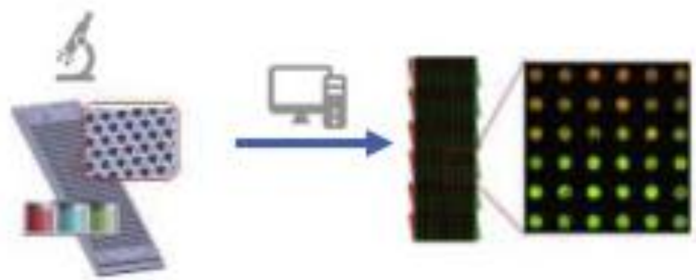
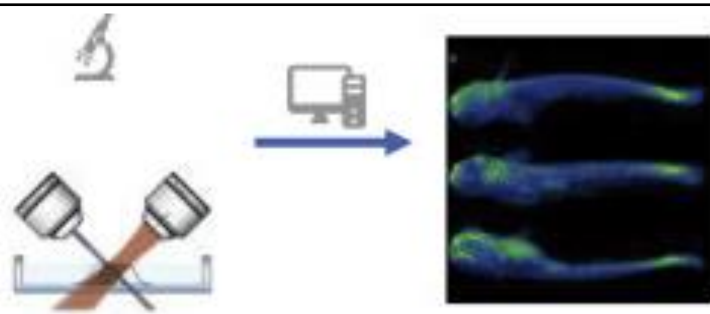
Natalie Elia

“Seeing is believing, quantifying is convincing”



Sbalzarini (2016)

Big data in cell imaging

Imaging technique		Production rate
Single molecule localization microscopy		~10 GB/h-1 TB/h
High-content Image-based screening		< 100 GB/h
Light-sheet microscopy		1-10 TB/h

High-end DNA sequencing ~100 GB/h
YouTube upload rate ~1 TB/h

Ouyang & Zimmer (2017)

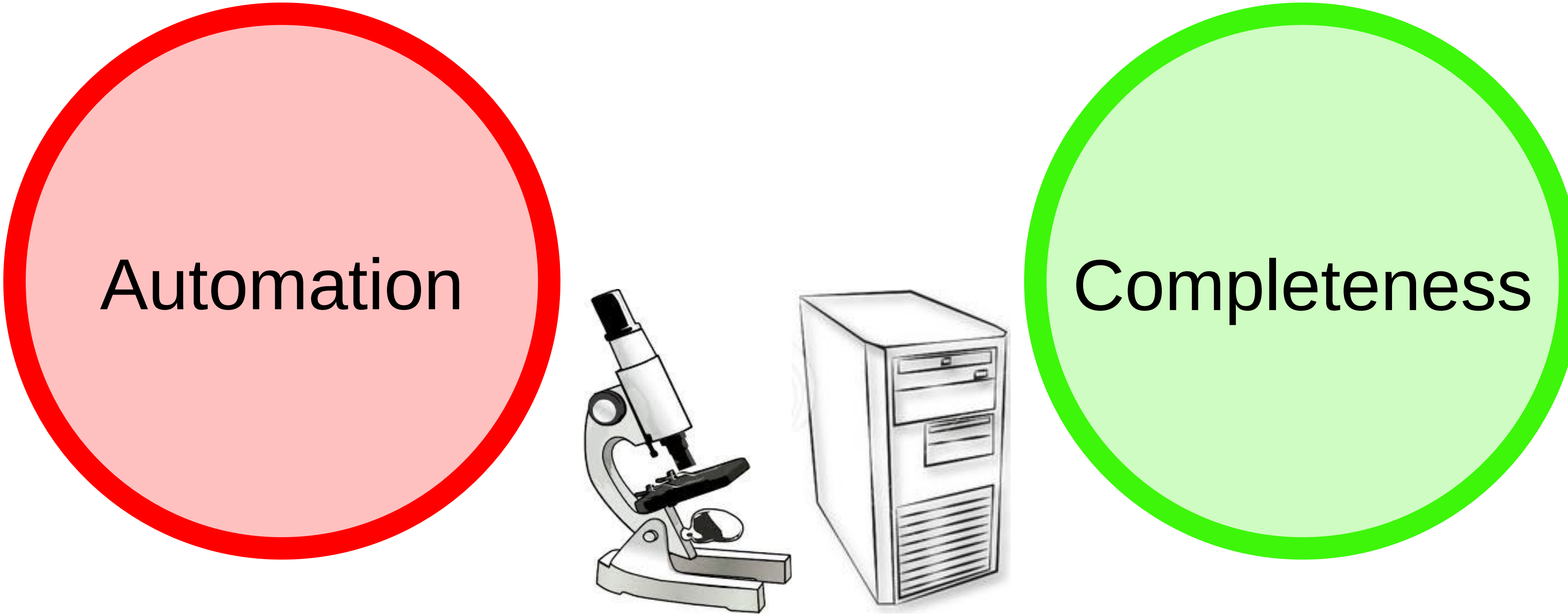
Bioimage analysis tools

ImageJ
Image Processing and Analysis in Java

CellProfiler
cell image analysis software

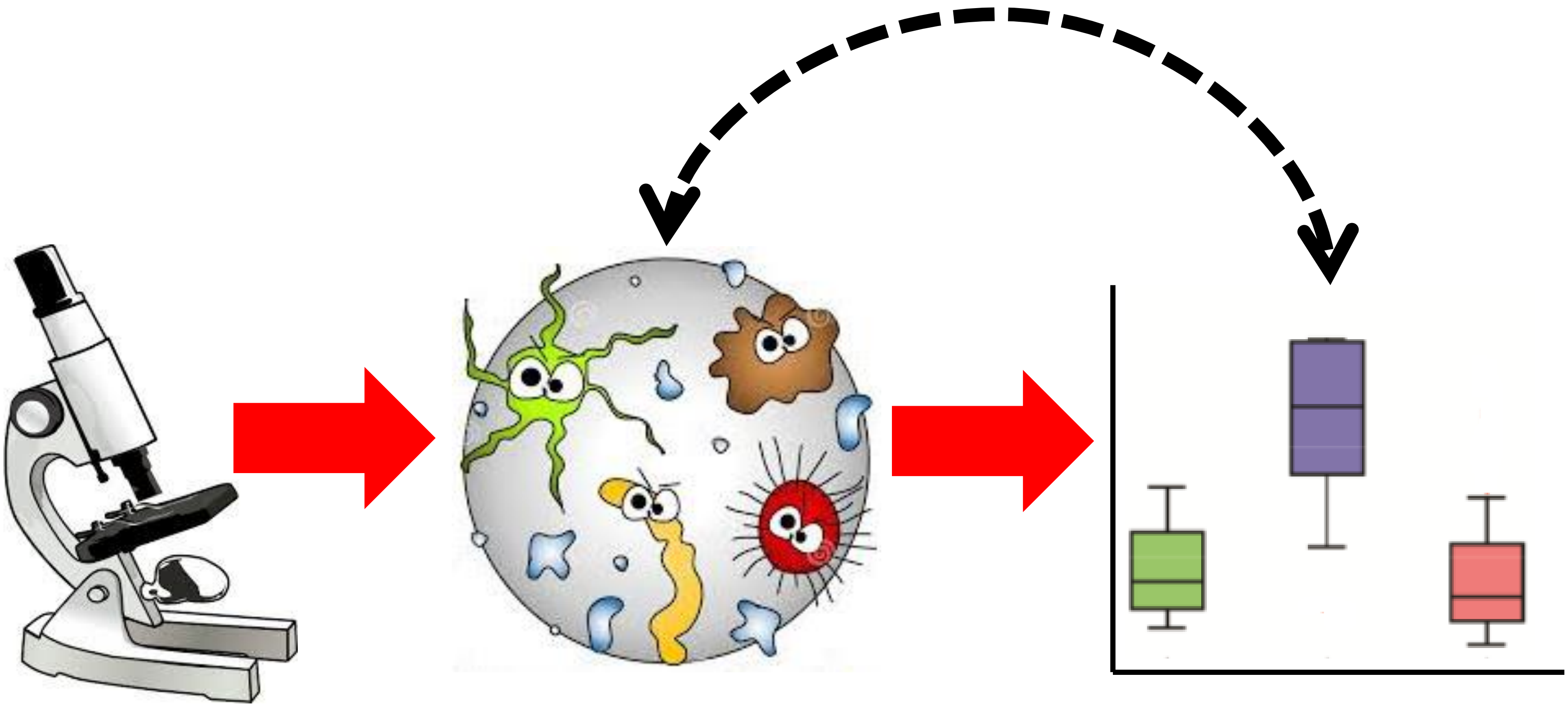
FEI
part of Thermo Fisher Scientific

BITPLANE
SCIENTIFIC SOLUTIONS



Danuser (2011)

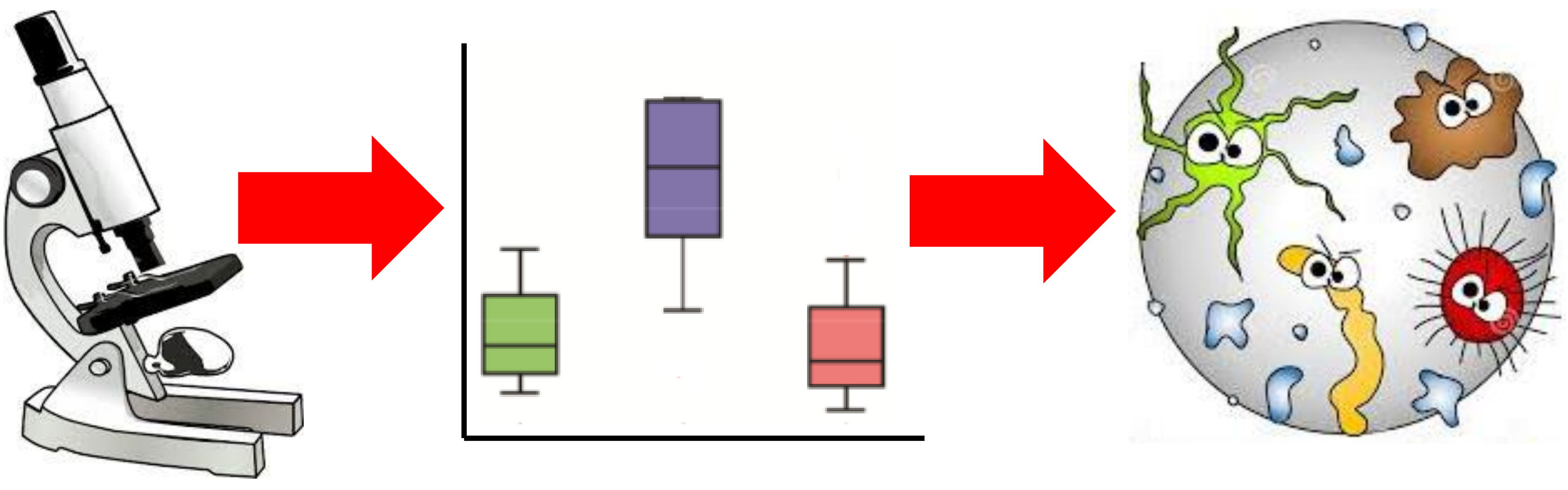
Quantifying the invisible (and then, sometimes, seeing it)

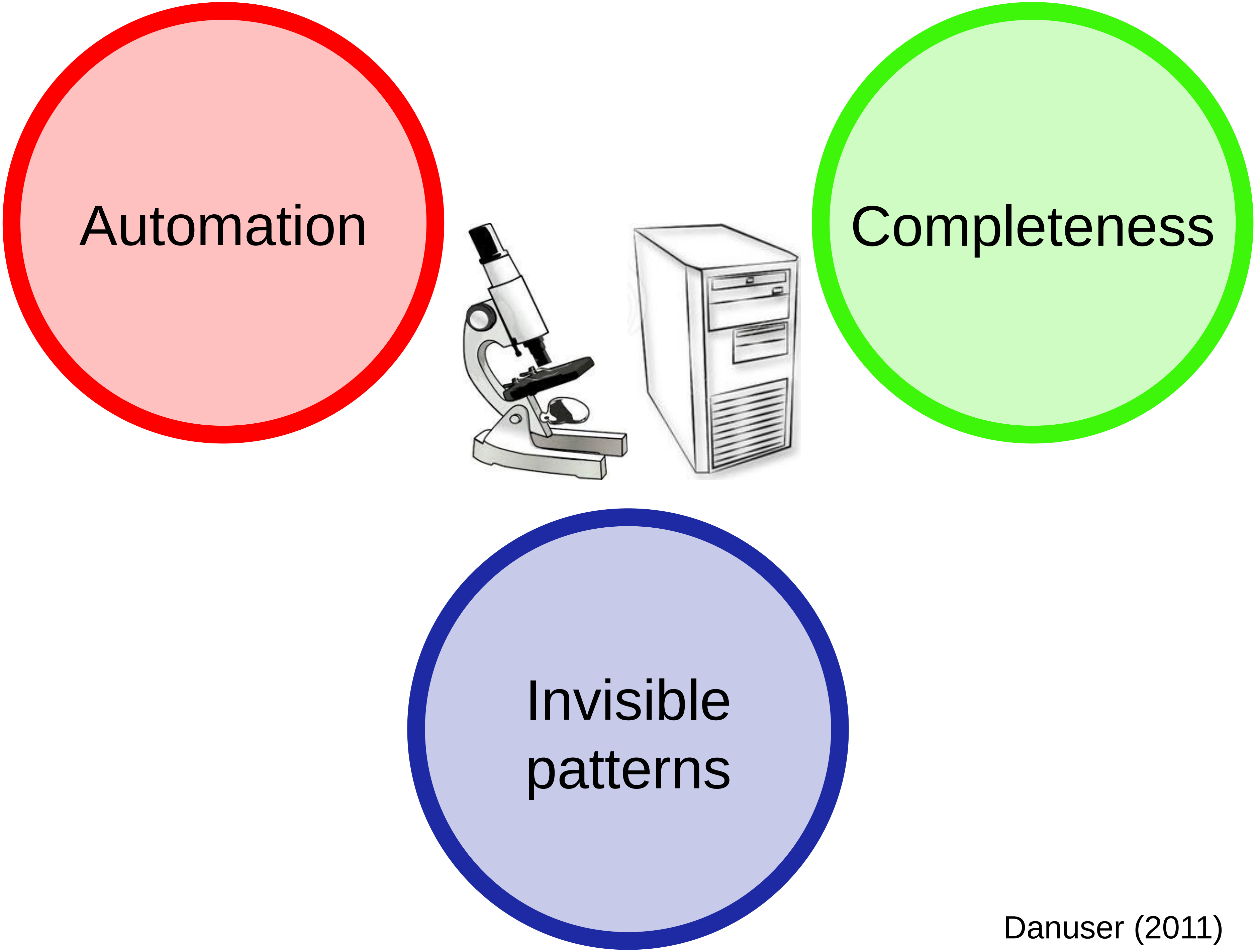


Sbalzarini (2016)

Quantifying the invisible
(and then, sometimes, seeing it)

Domain knowledge





Danuser (2011)

Lab of computational cell dynamics (and the focus of this course)

Motivated by fundamental questions in cell biology our lab produces biological insights along with specialized analytic tools that reveal hidden patterns in dynamic cell imaging data

Why dynamics?



- What direction is the ball moving?
- What direction are the players running?
- Who is fastest?
- What collective strategy leads to a winning?

Live cell imaging

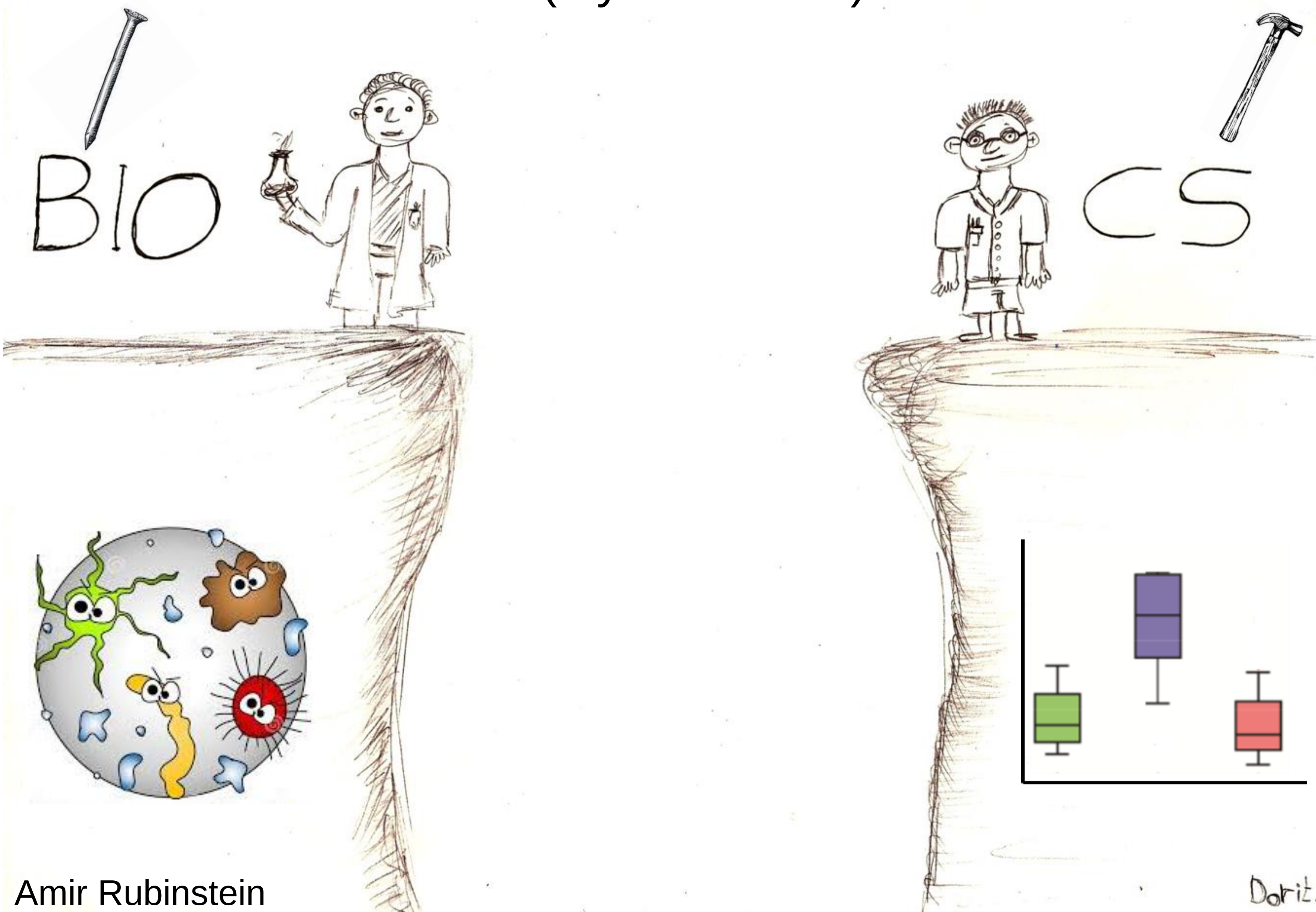



Tim Fessenden, @timisstuck

Course objectives and admin.

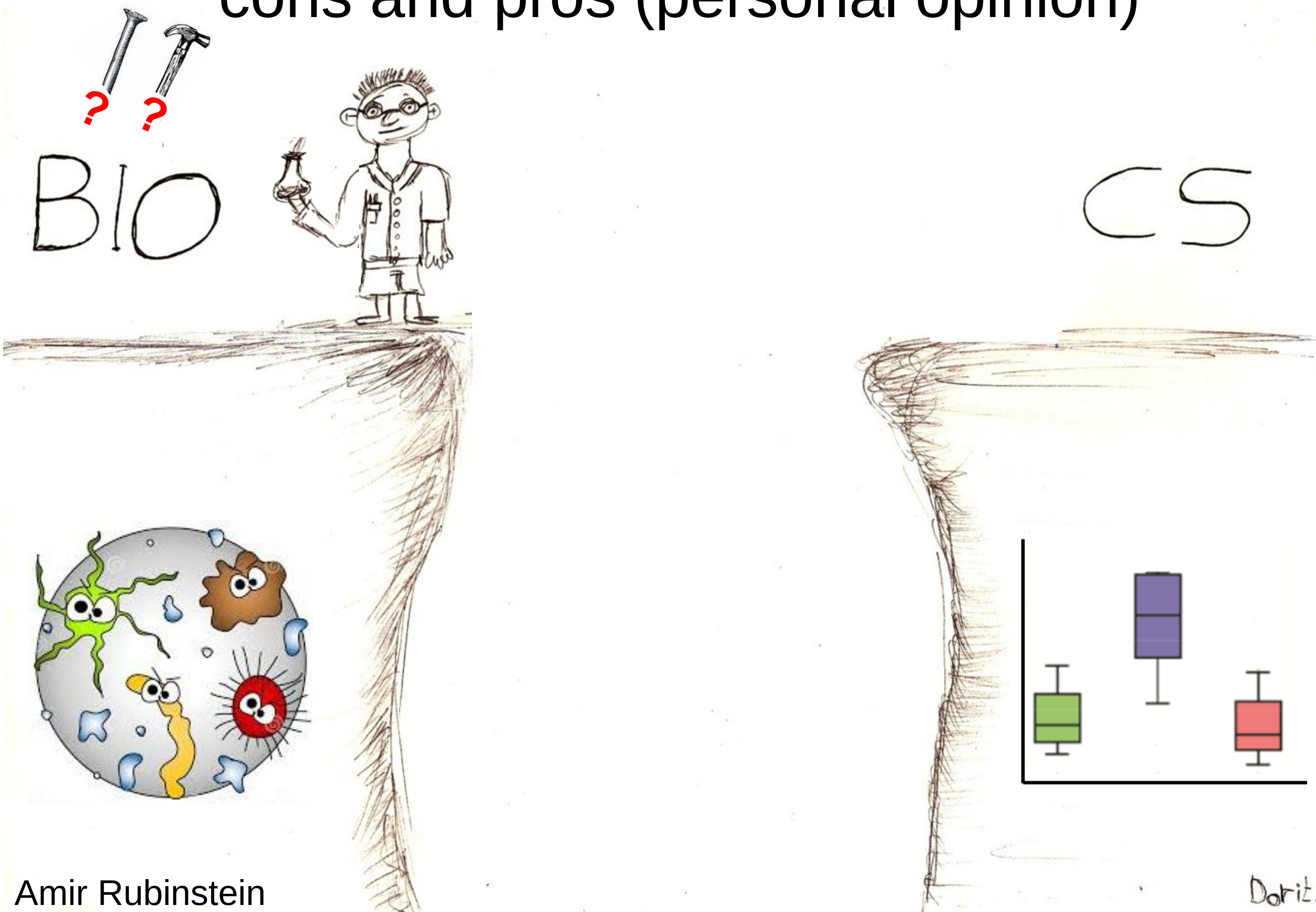
Course objective

(my motivation)



Amir Rubinstein

Experiments and computation:
cons and pros (personal opinion)



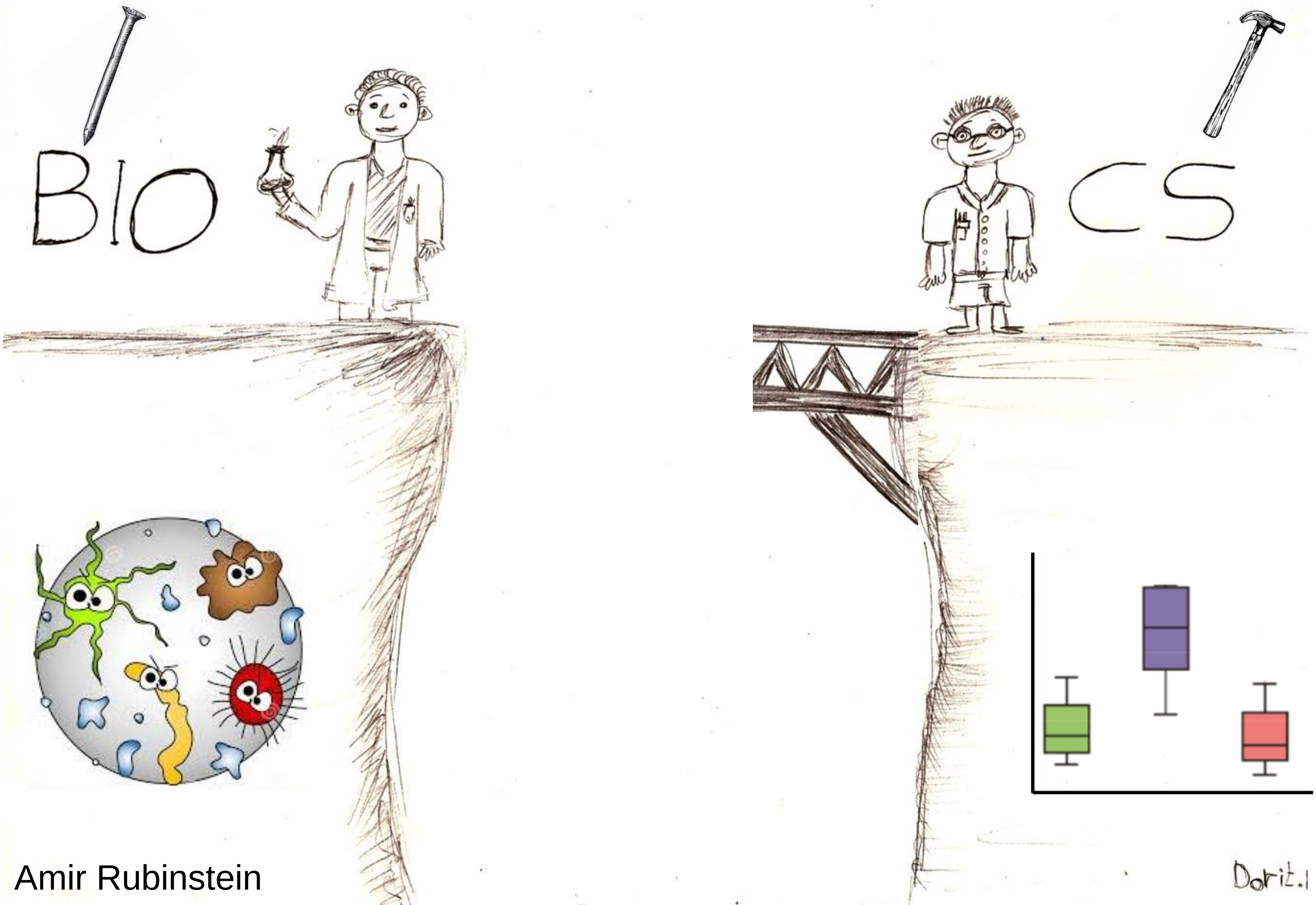
Amir Rubinstein

Bioimage informatics (Wikipedia)

- Bioimage informatics is a subfield of bioinformatics and computational biology
- The use of computational techniques to analyze bioimages at large scale and high throughput
- The goal is to obtain useful knowledge out of complicated and heterogeneous image and related metadata

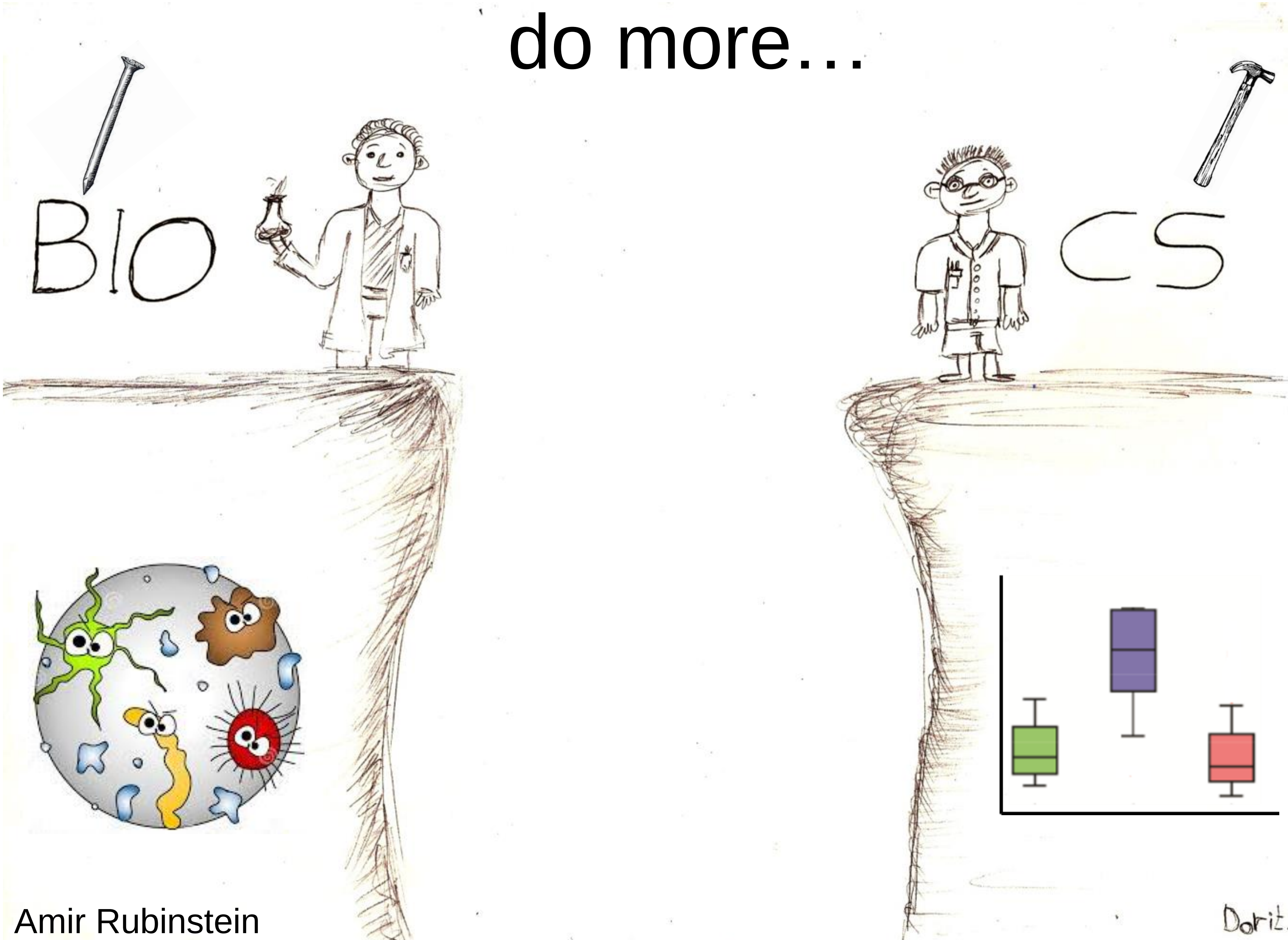
“The ultimate goal is to **replace human labor as much as possible with computer calculations**, so that **biologists can focus fully on formulating high-level hypotheses and designing increasingly sophisticated experiments**, while improving the objectivity and reproducibility of these experiments.” (Meijering et al., 2016)

Bioimage informatics



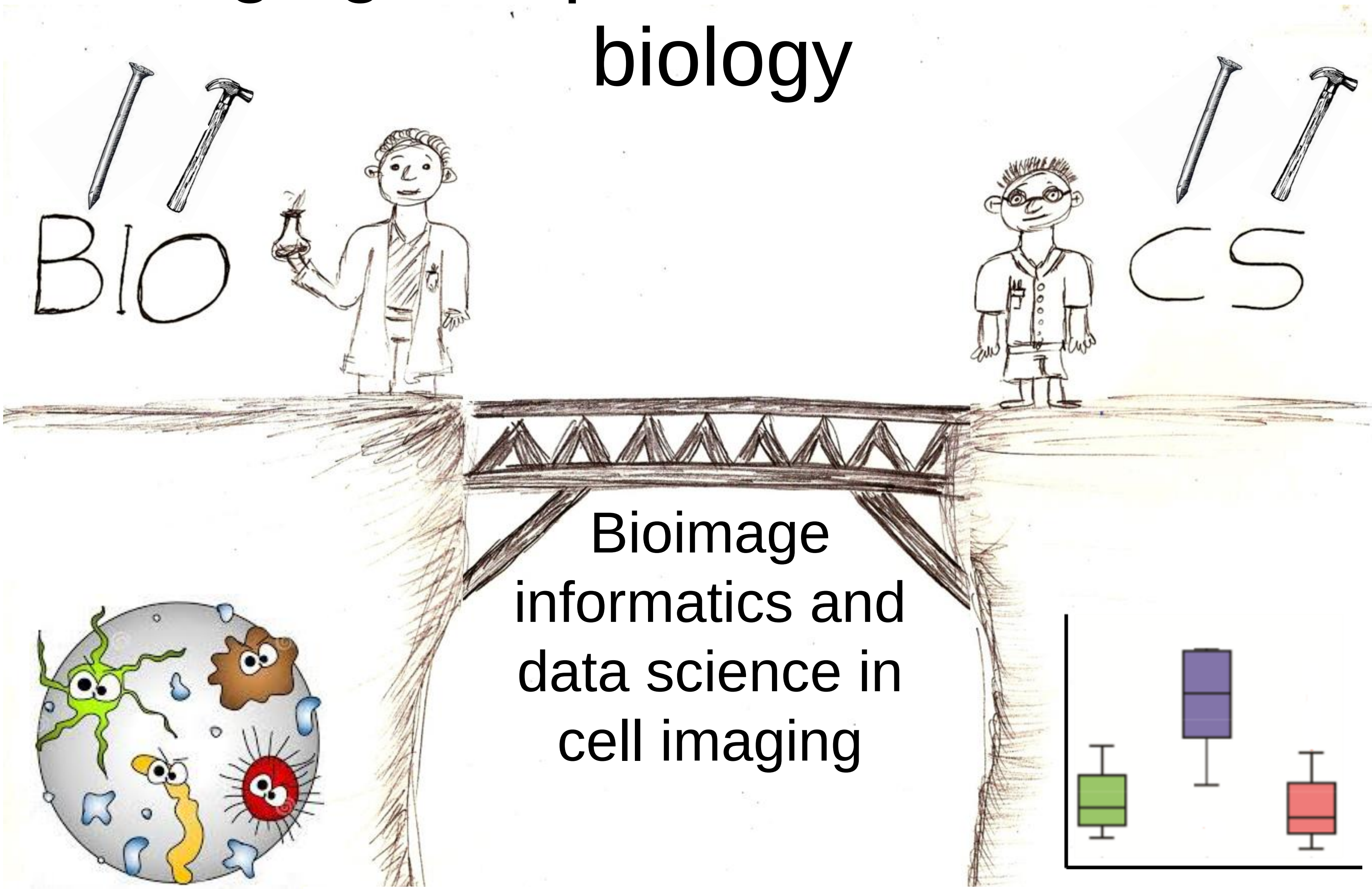
Amir Rubinstein

We (computational scientists) can do more…



The interplay between development of quantitative tools ("hammers“) and identifying open important questions in cell biology ("nails“)

Bridging computer science and cell biology



Bioimage informatics and data science in cell imaging

Read: https://www.nature.com/news/cell-biologists-should-specialize-not-hybridize-1.20277

Amir Rubinstein

I think the field is ready…

Overview of strategic priorities

Promote the centrality of cell biology to **diversify membership**, expand partnerships with **adjacent disciplines** and societies, and help **others see themselves as cell biologists**

How did I get here?

- B.Sc., computer science, BGU
- M.Sc., computer science (AI), BGU
- Algorithms engineer (> 5 years, multiple companies)
- Ph.D., computer vision, TAU
  - and then I discovered cell biology…
  - meaningful and exciting research in an emerging field - opportunity to make an impact!
  - Beautiful pictures and videos!
- Postdoc at a medical school, UTSW
- Back home at BGU since Oct. 2018 ☺

Course description

- Presenting a broad set of (microscopy-driven) **biological questions** and the diverse set of **computational tools** to solve them
- Not going deep into the biology nor into the computational aspects (unless students push for it)
- Computational aspects include techniques from computer vision, machine learning, time series analysis, network algorithms, etc.
- Full disclosure: topics selection is biased toward my personal interests (and research))

General info and requirements 1

- Open for all SISE and adjacent departments (e.g., CS, EE)
- Students from other departments are welcomed (and even encouraged!)
- Lectures will be held in English (why?)

General info and requirements 2

- Background in math and programming is required. Prior knowledge in machine learning and/or computer vision is highly recommended, but not necessary.
- No prior biological knowledge is required; all background will be covered in the lectures.
- The interdisciplinary, terminology, and culture will make the course quite challenging…

Grades

- Each student will present a paper/s – 20%
- Semester long research project – 80%
  - In groups (1-2 students, maybe 3?)
  - Open data exploration / collaboration / tool building/validation/comparison
  - "Freestyle" (more on that later)
  - Written report
  - Each participant will clearly articulate their contribution
- Grades (maybe) according to a pre-determined distribution + ranking

A warning!

- This is the first round of this course
- It is a first of a kind. No textbook or resources beyond (mostly recent) papers that I will pick…
- Please refer to it as a pilot - are you feeling adventurous? ☺

Another warning!

- It is not easy to dive into the complex domain of cell biology
- This is not going to be an easy course
- Are you feeling adventurous? ☺

Course (tentative) schedule and overview

Tentative topics list (we are probably not going to cover all!)

- Cell biology and microscopy (Natalie Elia, BGU)
- Bioimage informatics (Ofra Golani, WIS - tentative)
- High content single cell phenotypic profiling
- Sharing and reusing cell image data: public data repositories, harmonization, integration and fusion
- Enhancing cell image quality with deep learning
- Generative models for cell structure with deep learning
- Classifying cell state with deep learning
- 3D image analysis
- Visualization
- Colocalization

Tentative topics list (we are probably not going to cover all!)

- Quantifying cell motility: from the intraceullular to the multicellular scales
- Computer vision in cell imaging (Tammy Riklin Raviv, BGU - tentative)
- Quantifying intercellular communication, Quantifying causality with fluctuations analysis (no perturbation)
- Speckle fluorescent microscopy
- Importing ideas from systems biology (intro: Tal Shay, BGU)
- Crowd sourcing
- High content modeling
- Medical imaging applications

Tentative schedule

#Session	Date	Topics
1	11.3	Introduction to data science in cell imaging
2	18.3	Cell biology and microscopy (Natalie Elia)
3	25.3	High content single cell phenotypic profiling, Sharing and reusing cell image data: public data repositories, data harmonization, integration and fusion
4	1.4	Enhancing cell image quality with deep learning, generative models for cell structure with deep learning
5	22.4	Classifying cell state with deep learning, colocalization
6	6.5	Quantifying cell motility: from the intraceullular to the multicellular scales 1
7	13.5	computer vision in cell imaging (Tammy Riklin Raviv), applications in medical imaging
8	20.5	Assaf away
9	3.6	Importing ideas from systems biology (intro: Tal Shay), Crowd sourcing
10	10.6	Bioimage informatics (Ofra Golani)
11	17.6	Quantifying intercellular communication, Quantifying causality with fluctuations analysis (without perturbations)
12	24.6	3D image analysis, Speckle fluorescent microscopy, High content modeling

While image processing / computer vision is not the focus of this course, you'll have to figure out on your own how to work with images and follow the relevant litrature..

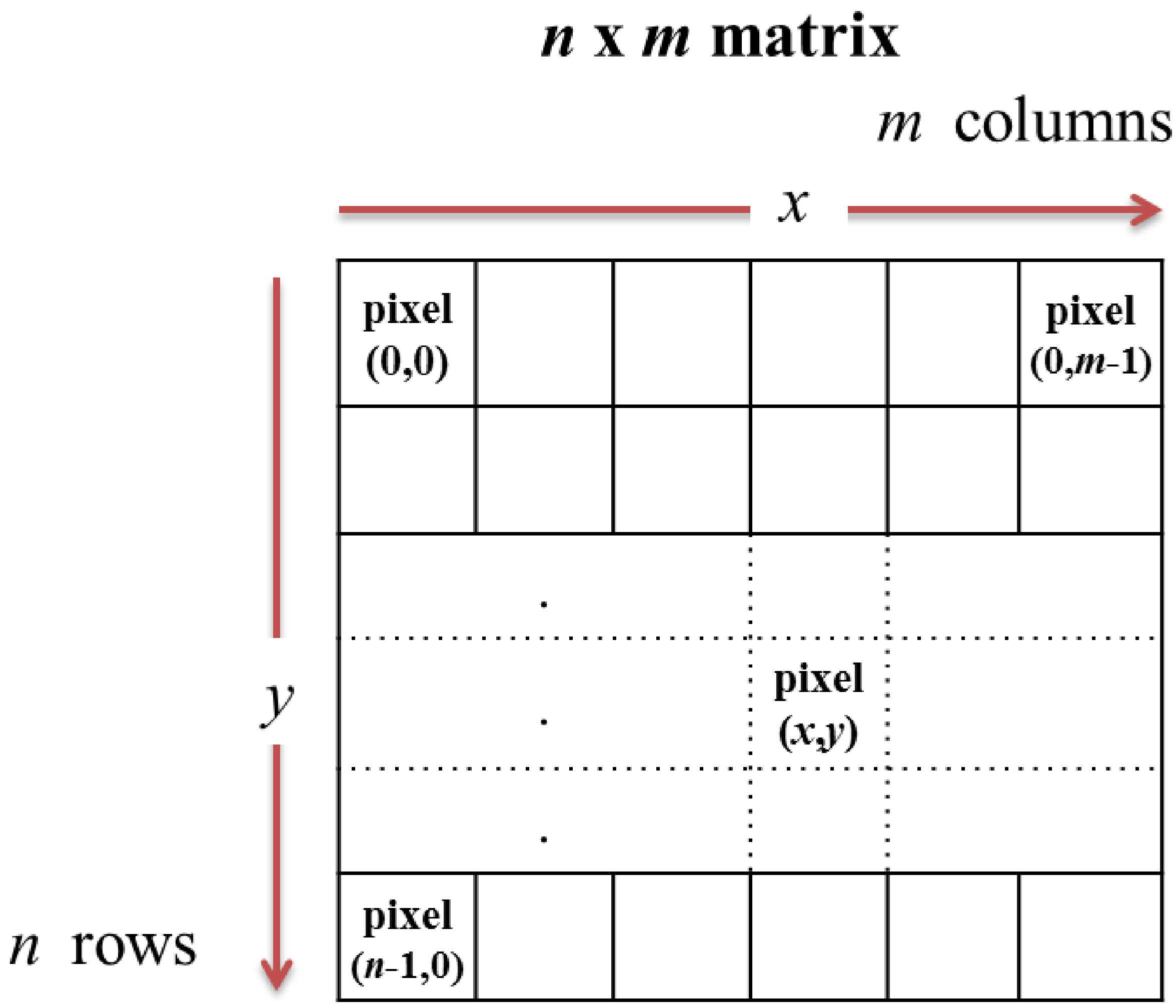
I'll now go through some brief background that can fit non-computational students

We'll hear more throughout the course (optics – Natalie Elia, computer vision – Tammy Riklin Raviv, Bioimage informatics – Ofra Golani, and me ☺)

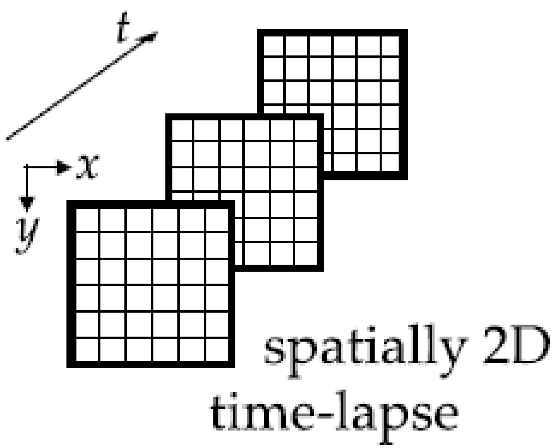
What is an image?
(very basic slides adapted from my intro to CS course...)

Basic model of a digital image

Pixel: each element M[x, y] of the image



- A 2D image is encoded as a n-by-m matrix M
- Videos

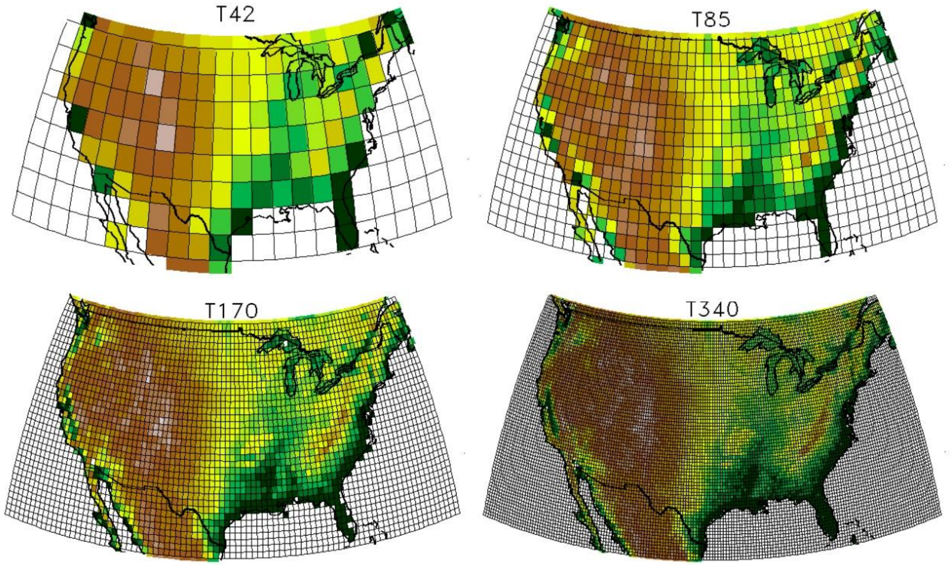


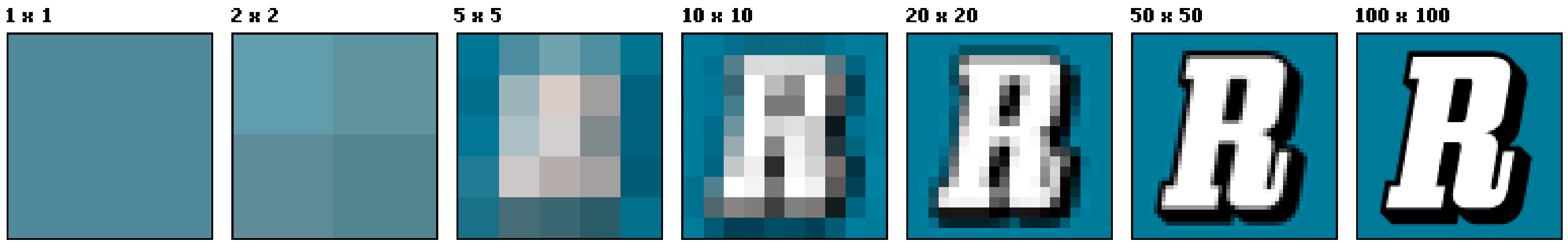
- 3D
- Multiple channels..

Resolution

- Resolution is the capability of the sensor to observe or measure the smallest object clearly with distinct boundaries.
- Resolution depends on the physical size of a pixel.

Higher resolution = more pixels per area = **lower** pixel size.

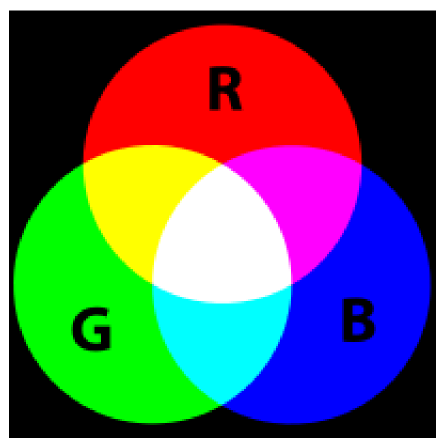


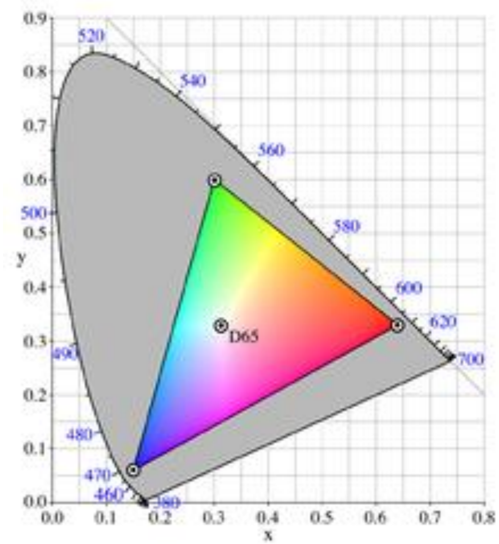


Source: wikipedia

Color images (RGB)

- For standard (RGB) color images, M[x, y] is a triplet of values, representing the red, green, and blue components of the light intensity at the pixel.
- In microscopy, each channel (does not have to be 3) labels a different intracellular structure.

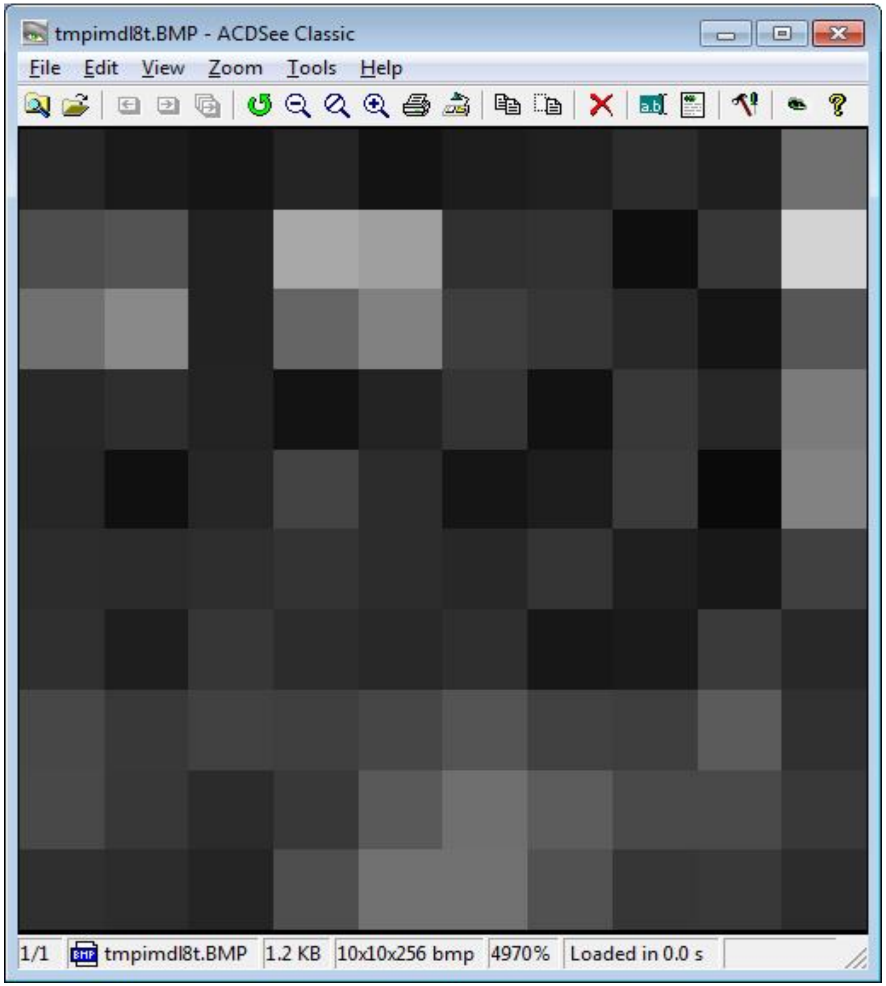




(images from Wikipedia)

Gray scale image

- 8 bits per pixel (2^8=256 gray levels): 0 = black, 255 = white



38,	26,	21,	36,	19,	28,	33,	44,	31,	112,
77,	83,	34,	168,	159,	48,	50,	14,	55,	211,
112,	137,	34,	101,	129,	62,	54,	40,	21,	86,
41,	46,	35,	19,	35,	52,	18,	57,	39,	123,
38,	16,	38,	67,	45,	21,	29,	59,	10,	130,
45,	43,	46,	51,	44,	39,	53,	31,	24,	64,
47,	30,	54,	45,	40,	46,	23,	26,	58,	40,
71,	57,	66,	63,	70,	84,	65,	62,	91,	49,
72,	55,	43,	57,	90,	111,	92,	73,	74,	56,
47,	45,	36,	78,	114,	113,	81,	54,	57,	44

Bit Depth

- Number of bits per pixel.

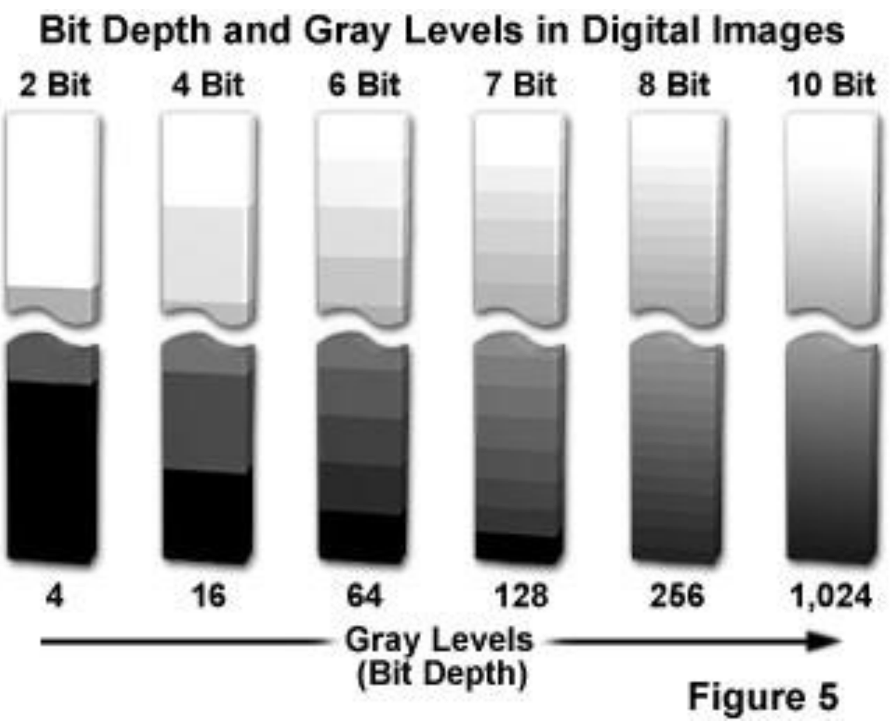


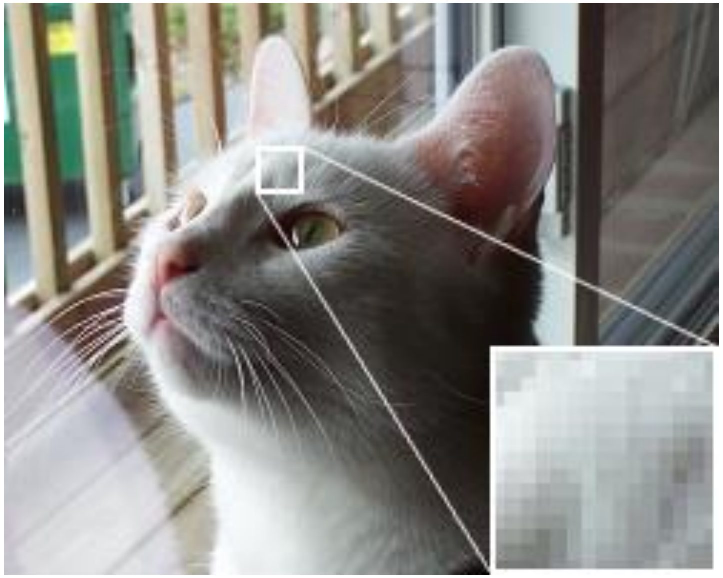
Image from:
http://micro.magnet.fsu.edu/

- A human observer sees at most a few hundreds shades of gray
- Higher bit depths images: typically for automated analysis by a computer.

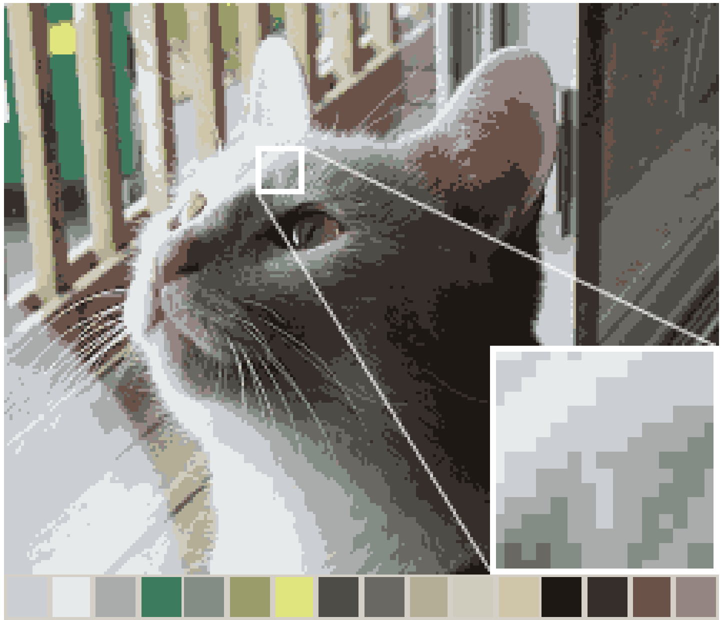
Image Quantization

Number of bits per pixel

24 bit RGB

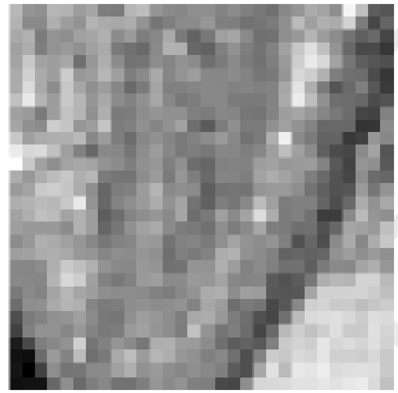


16 colors



Note that both images have the same pixel & spatial resolution

Why is it hard for a computer to interpret images?



Any guesses as to what this image is (or is part of)?

```
[183, 148, 143, 181, 178, 156, 165, 126, 123, 181, 189, 148, 139, 135, 142]
[178, 138, 138, 175, 158, 147, 179, 171, 168, 173, 147, 117, 127, 139, 139]
[168, 176, 123, 147, 142, 161, 165, 176, 140, 154, 125, 136, 169, 122, 99]
[161, 173, 127, 147, 154, 144, 161, 170, 122, 113, 105, 138, 177, 175, 104]
[164, 200, 162, 158, 167, 111, 151, 174, 140, 115, 117, 141, 133, 146, 140]
[170, 208, 171, 158, 204, 152, 158, 182, 159, 126, 138, 169, 134, 147, 157]
[171, 219, 170, 140, 199, 166, 143, 157, 134, 106, 123, 164, 129, 129, 133]
[192, 232, 180, 156, 200, 181, 149, 168, 140, 130, 144, 167, 135, 120, 125]
[160, 154, 122, 157, 194, 181, 145, 175, 122, 124, 141, 148, 144, 144, 128]
[165, 145, 140, 205, 197, 150, 157, 197, 124, 128, 144, 133, 145, 162, 111]
[199, 160, 172, 174, 175, 184, 136, 156, 125, 108, 135, 145, 133, 121, 129]
[234, 215, 218, 193, 159, 129, 104, 137, 135, 118, 141, 156, 135, 122, 126]
[254, 199, 160, 142, 163, 168, 163, 147, 140, 127, 144, 151, 127, 144, 109]
[173, 152, 173, 188, 199, 175, 182, 124, 117, 116, 141, 154, 122, 150, 126]
[154, 163, 200, 206, 197, 172, 142, 102, 124, 128, 155, 180, 138, 142, 139]
```

Computer vision
(old) slides adapted from Lior Wolf, TAU & Facebook AI Research

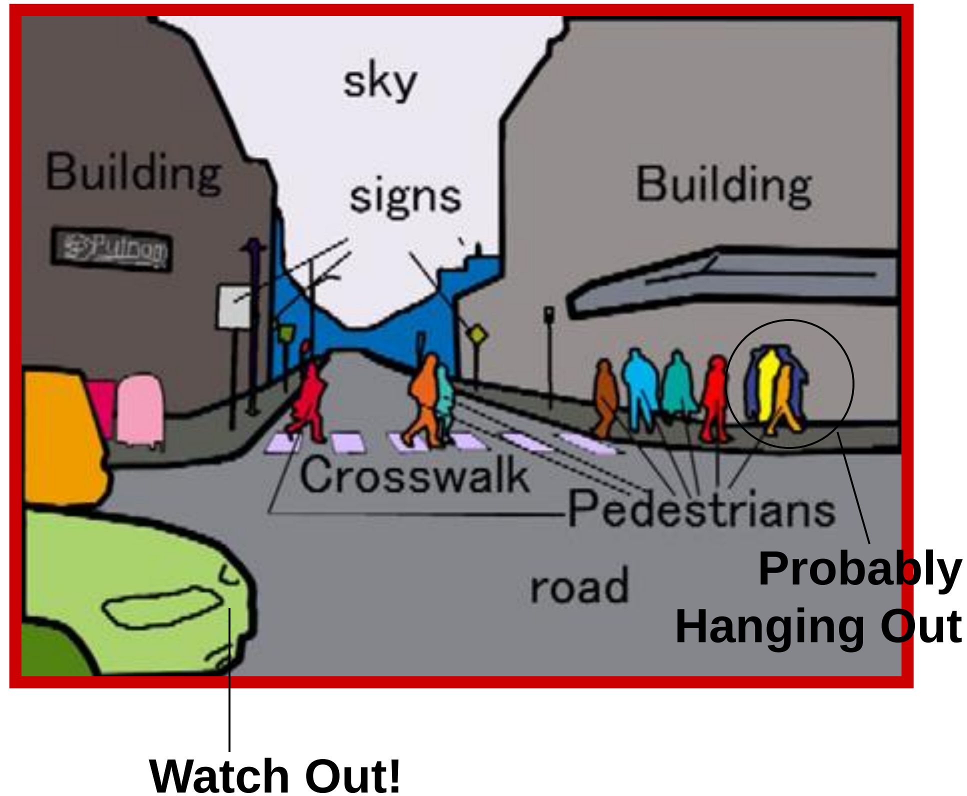
I spy a mousehole, two butterflies,
An empty green bottle, a jar full of eyes;
A dragon cauldron, a scroll that's rolled,
WIZARD, a spider, and a wand made of gold.

Credit: www.scholastic.com

Example tasks

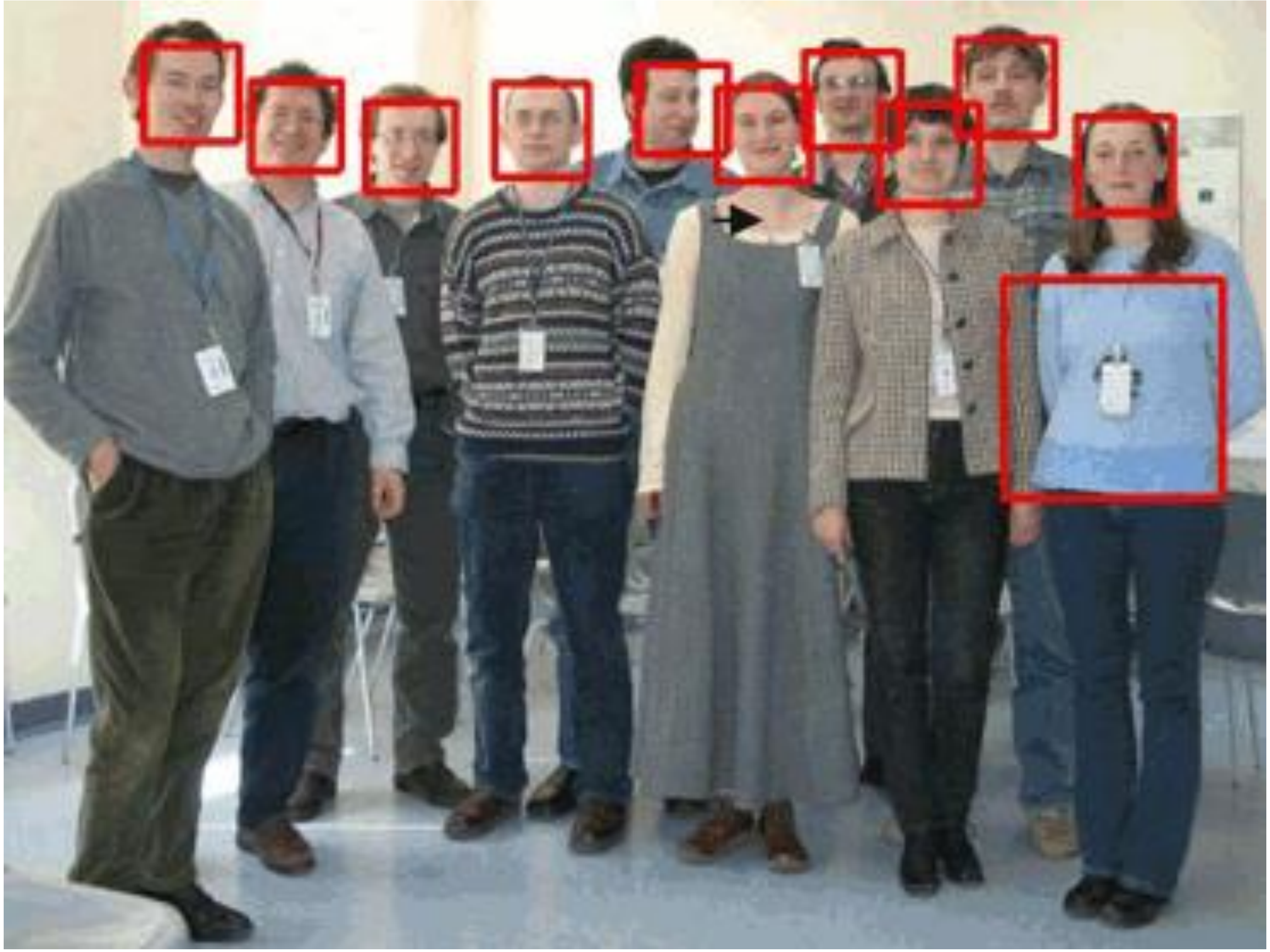
- Street scene understanding
- Face detection
- Face identification
- Image categorization
- Pose estimation
- Motion and behavior analysis

Street scene understanding



Credit: Stan Bileschi, CBCL

Face detection

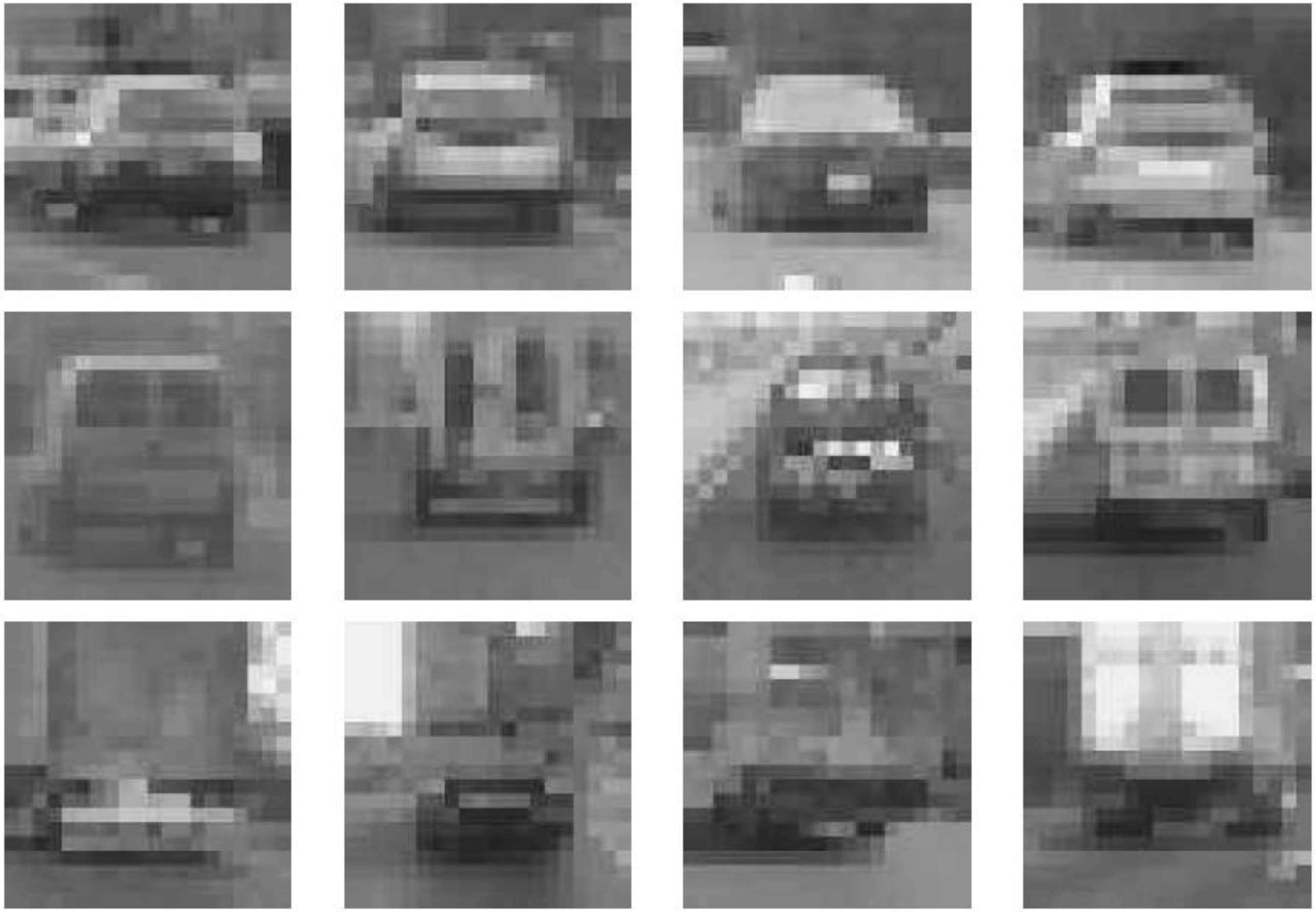


Credit: Intel Technology Journal, Volume 09, Issue 01

Face identification

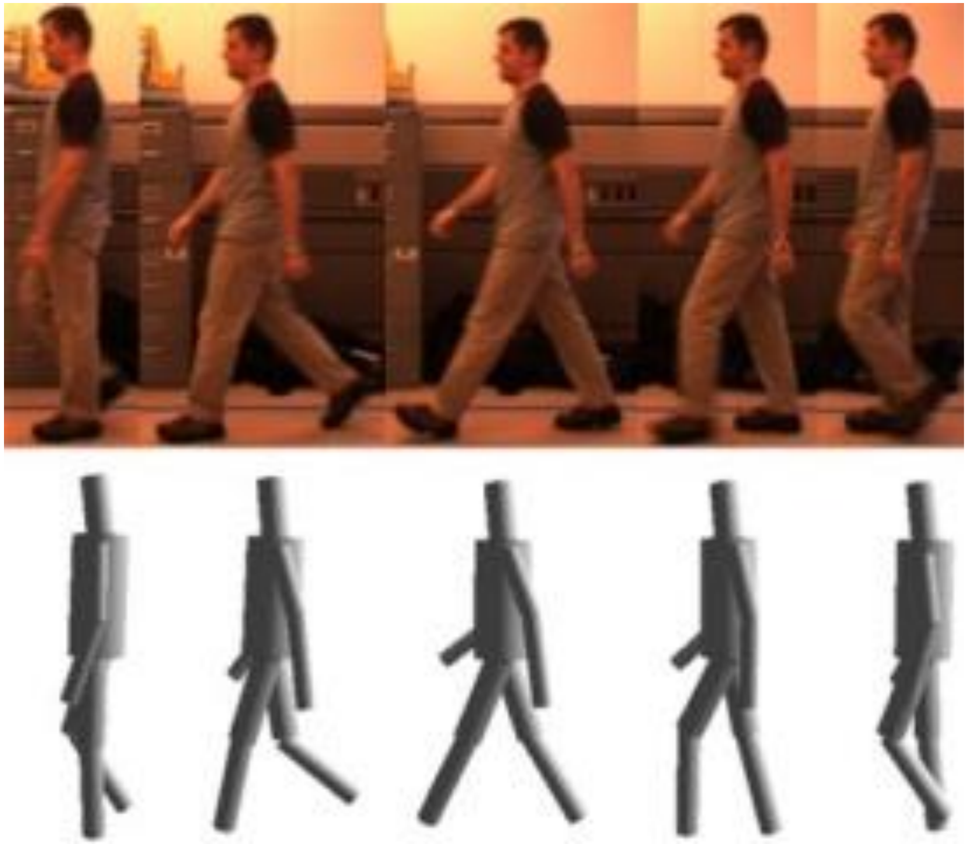
Credit: www.security-lab.com

Image categorization



Images courtesy of Mobileye

Pose estimation



Credit: Thomas Izo, CSAIL

Motion recognition

The "Birmingham Royal Ballet"

Ballet turn:

Input video:

Output:

Credit: Eli Shechtman and Michal Irani

Challenges

View point variation

Michelangelo 1475-1564

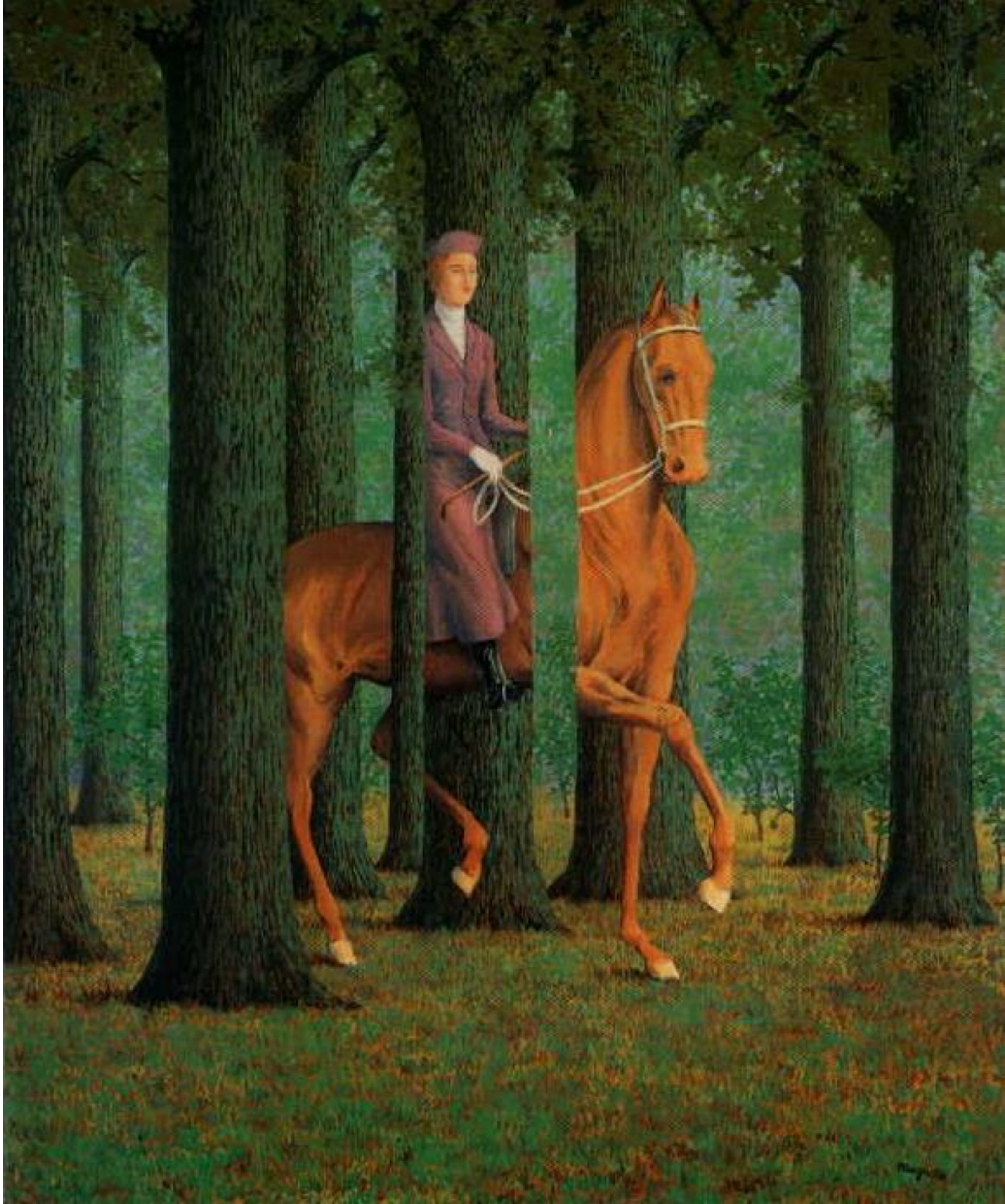
slide by Fei Fei,
Fergus & Torralba

illumination

slide credit: S. Ullman

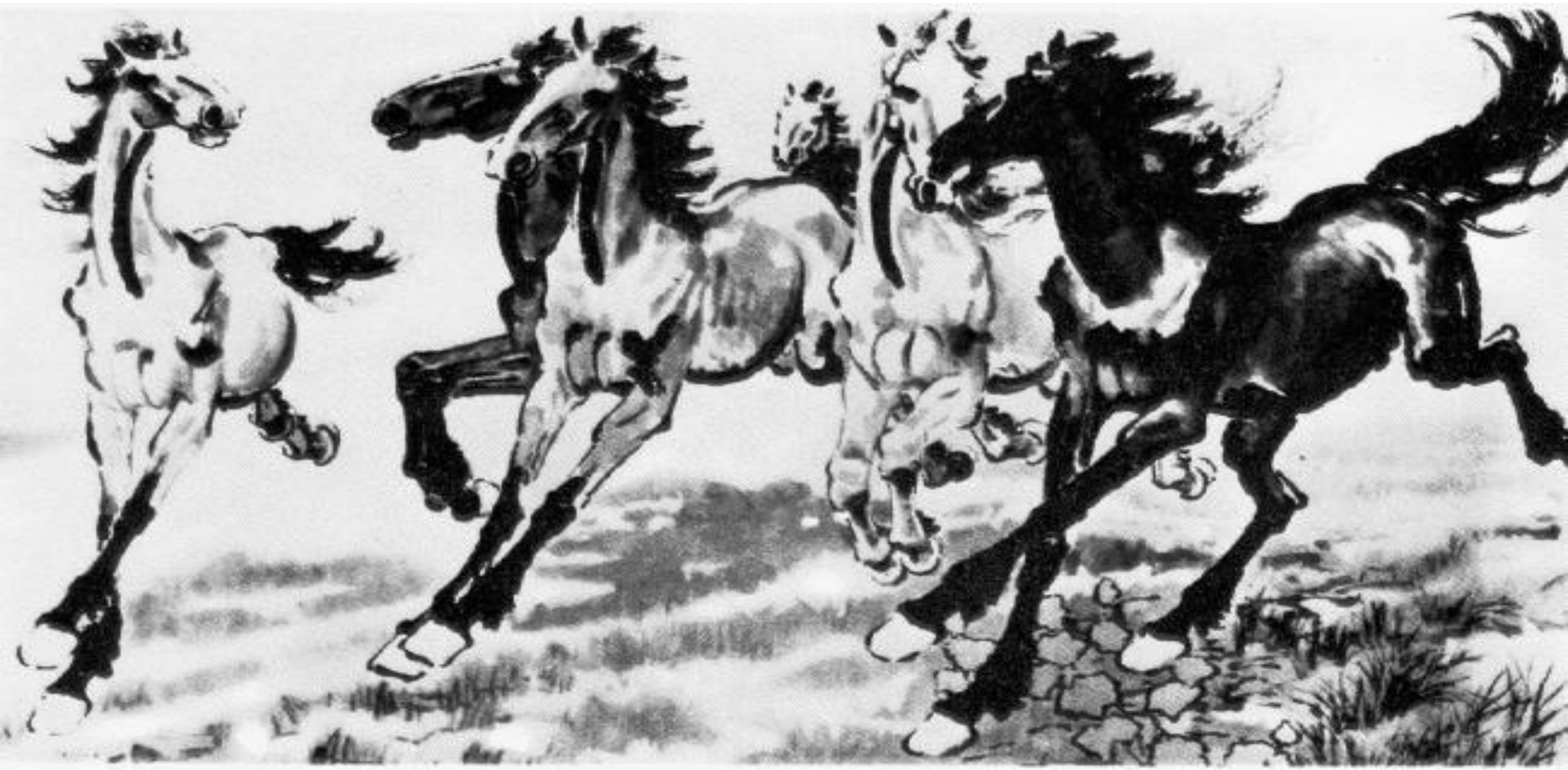
occlusion

slide by Fei Fei, Fergus & Torralba



Magritte, 1957

deformation



Xu, Beihong 1943

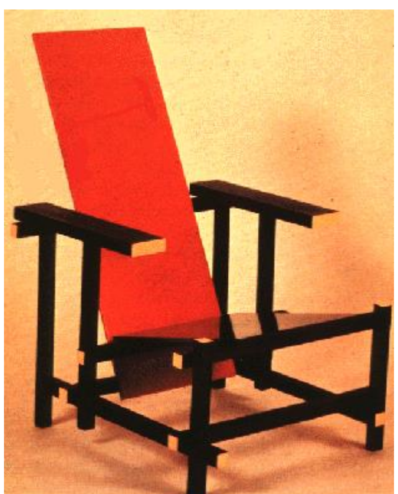
slide by Fei Fei, Fergus & Torralba

background clutter

Klimt, 1913

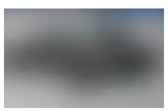
slide by Fei Fei, Fergus & Torralba

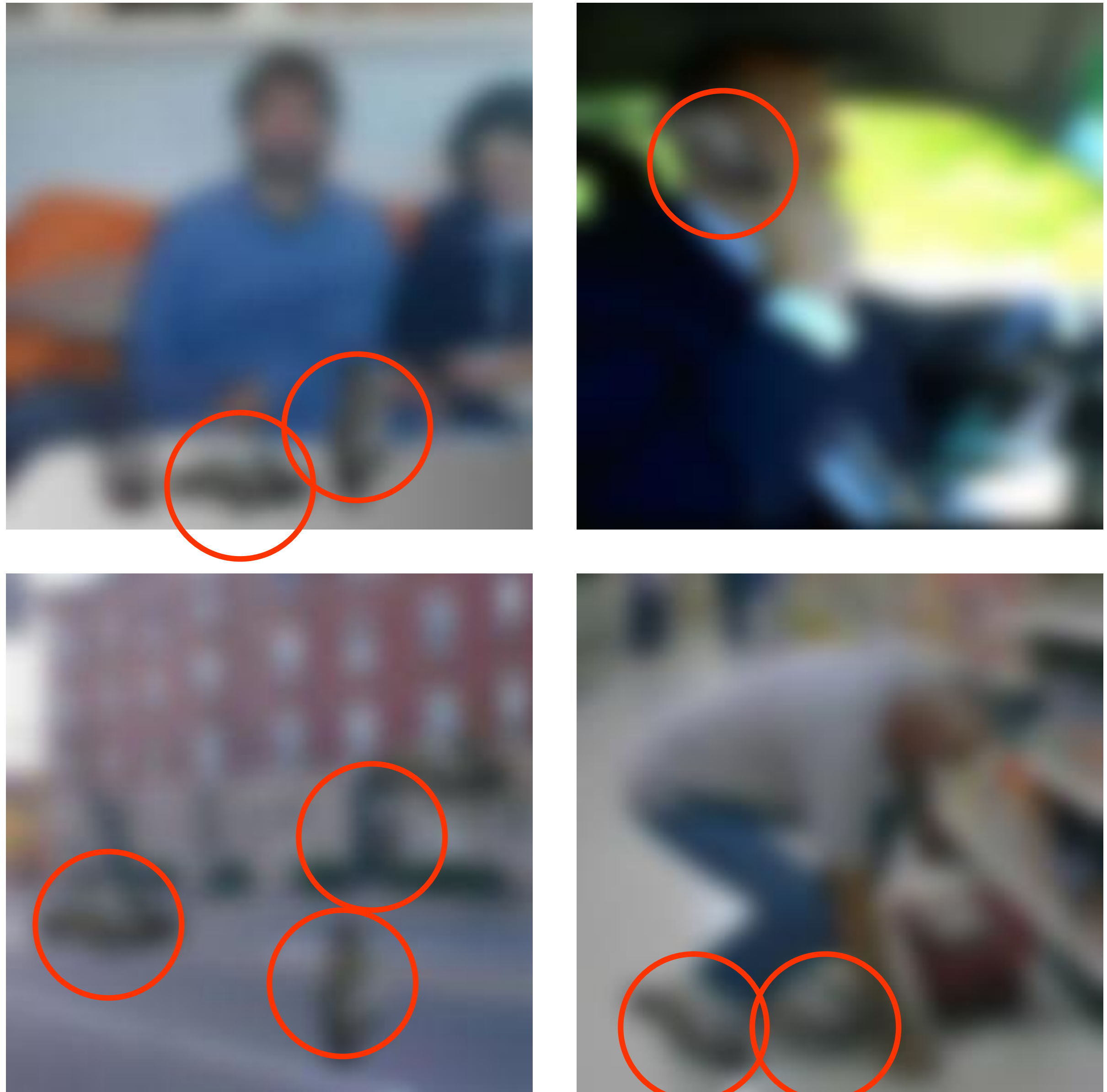
object intra-class variation



slide by Fei-Fei, Fergus & Torralba

local ambiguity





slide by Fei-Fei, Fergus & Torralba

the world behind the image (context)

Credit: A. Efros

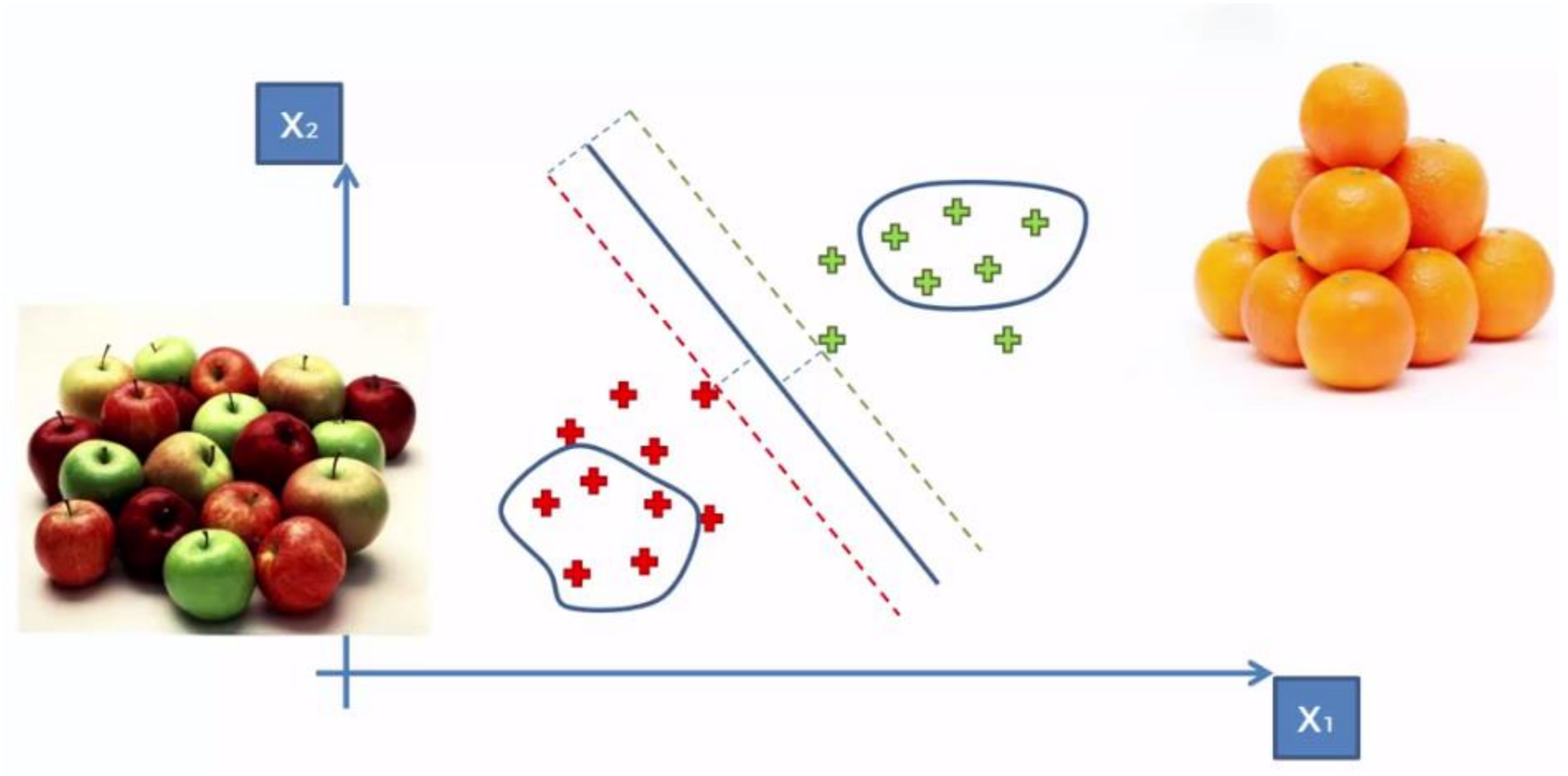
The general idea behind machine learning (the main driver behind modern computer vision)

How would you tell apart?

Source: https://bit.ly/2V2bJn2

Supervised machine learning

data → features → training → model



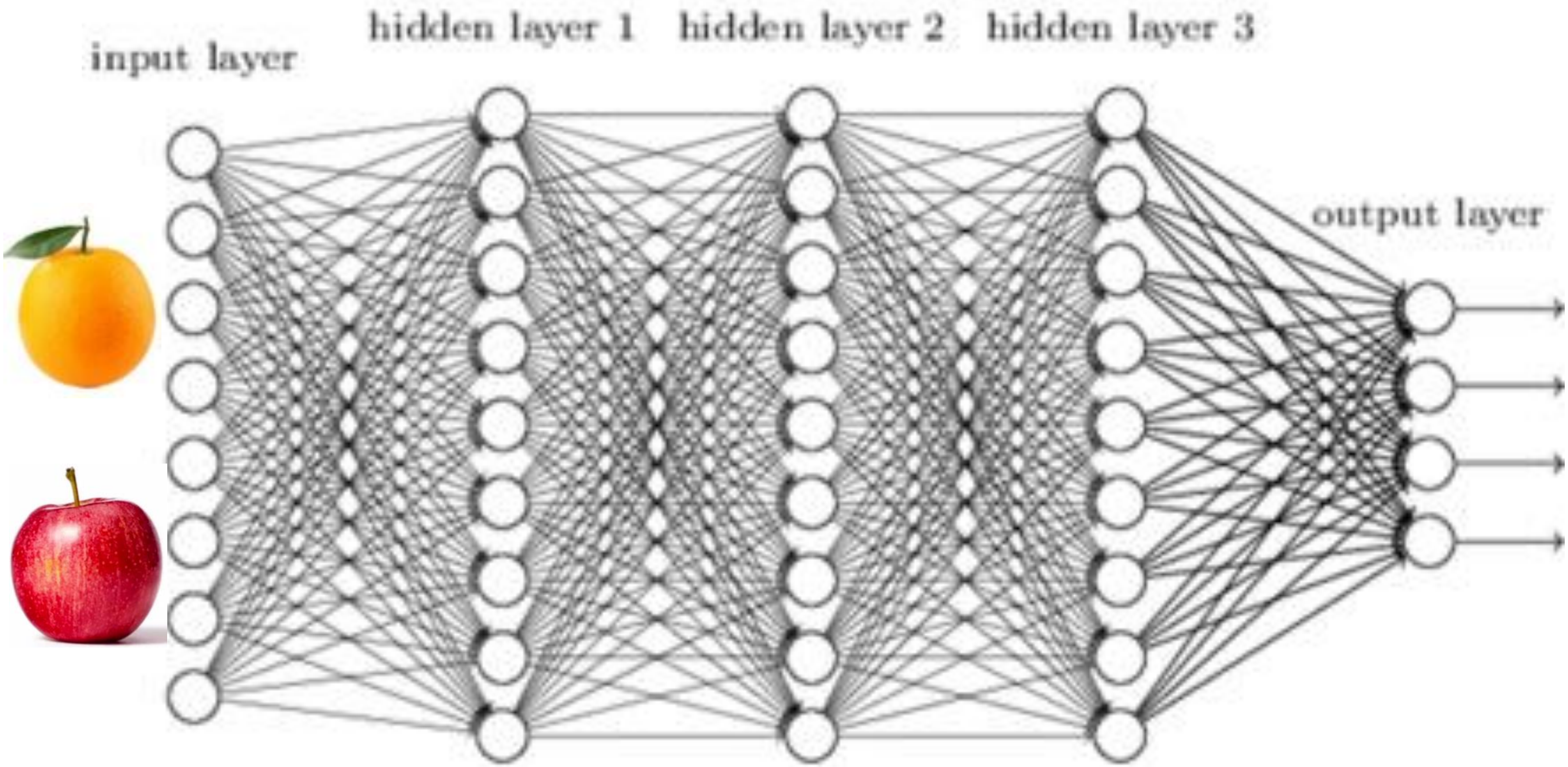
Credit: https://bit.ly/2UZVXch

Unsupervised machine learning

data → features → clustering

Source: https://bit.ly/2IJN0NT

Deep learning
Automated feature engineering



Source: https://bit.ly/2GMZou6

Current state of computer vision

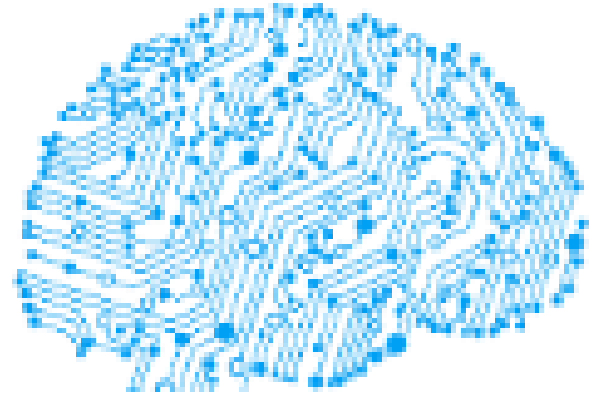
(Caveat: personal view, I am not an expert)

- Improved object detection, recognition, segmentation
- Image captioning
- Generative Adversarial Networks: seeing is no longer believing (many examples)
- Applications:
  - Biometrics
  - Autonomous vehicles
  - Manufacturing
  - Medical imaging - surpassing human experts!

Cell biology (microscopy) is behind, in terms of the application of modern computer vision – an opportunity!

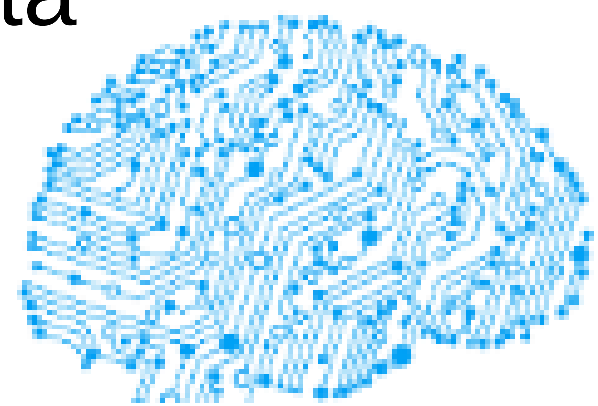
Most computer vision tasks

$>$



Identifying patterns in complex cell image data

$<$



Research outcome

Algorithmic elegance and performance

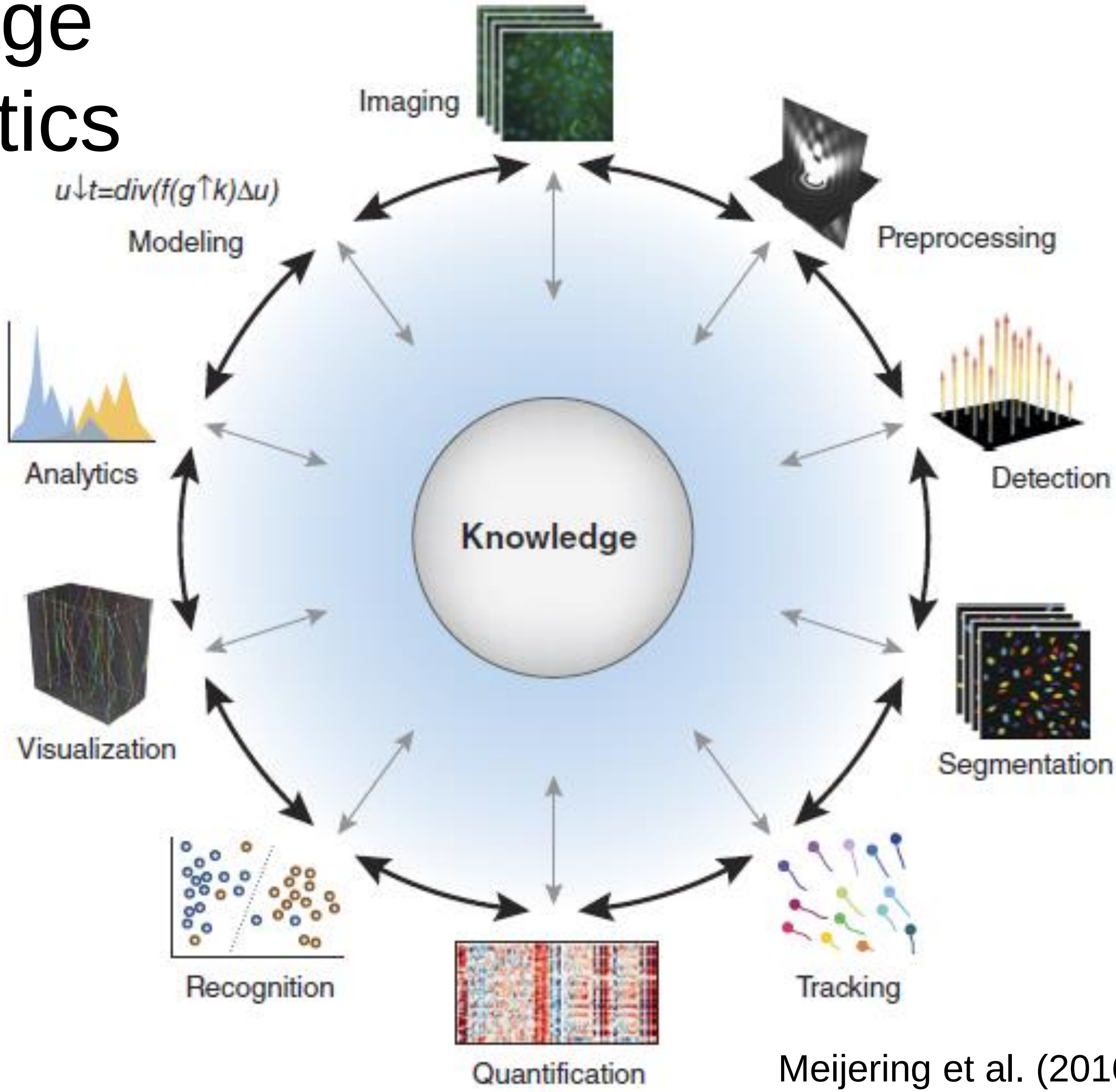
versus

discovery!

What is special about cell image data? - discussion

Brief browsing through <u>some</u> of the course highlights

Bioimage informatics



Meijering et al. (2016)

Tradeoffs in image acquisition

Spatial resolution

Temporal resolution

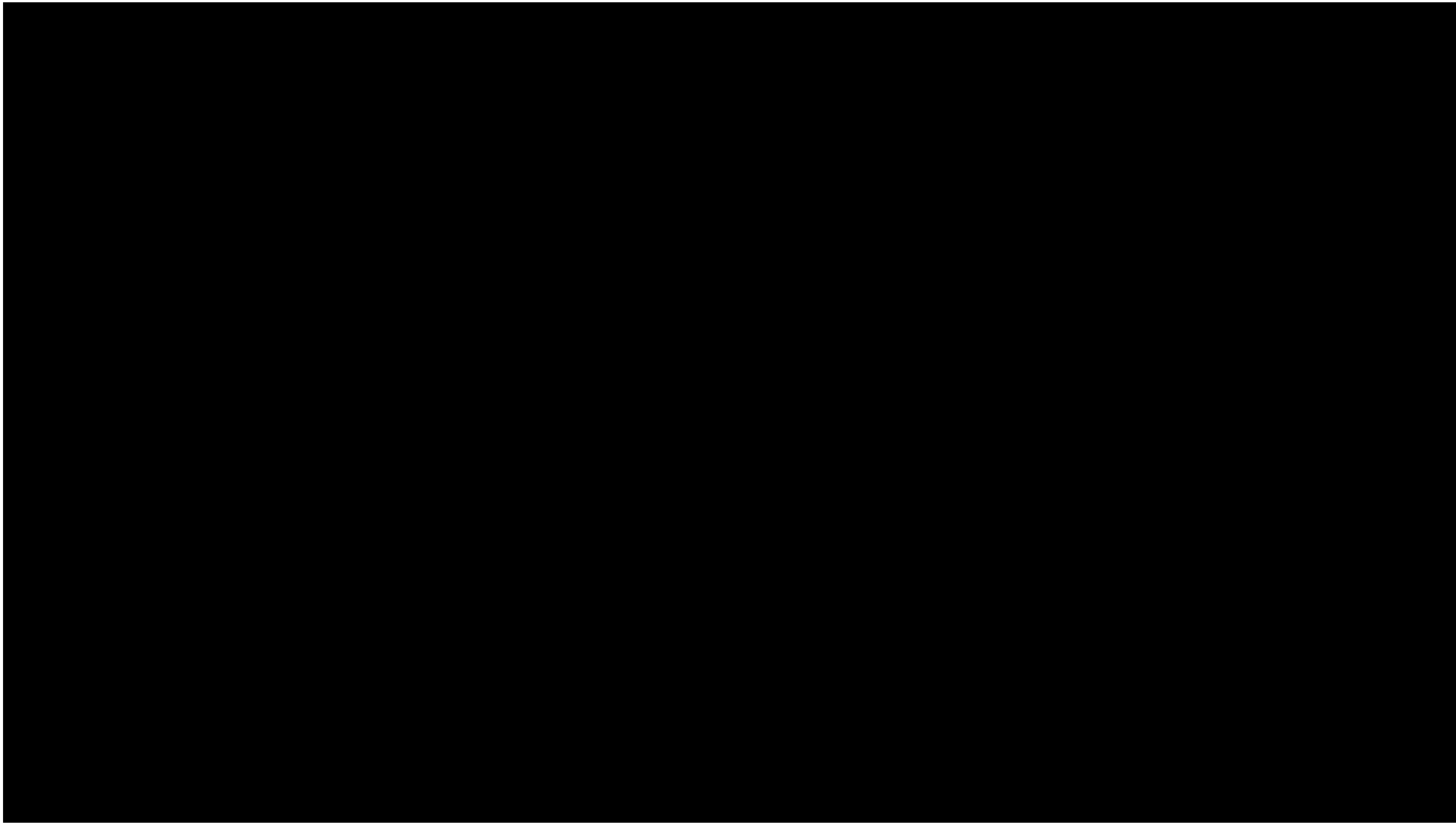
Bleaching and phototoxicity

Signal to noise

Preprocessing

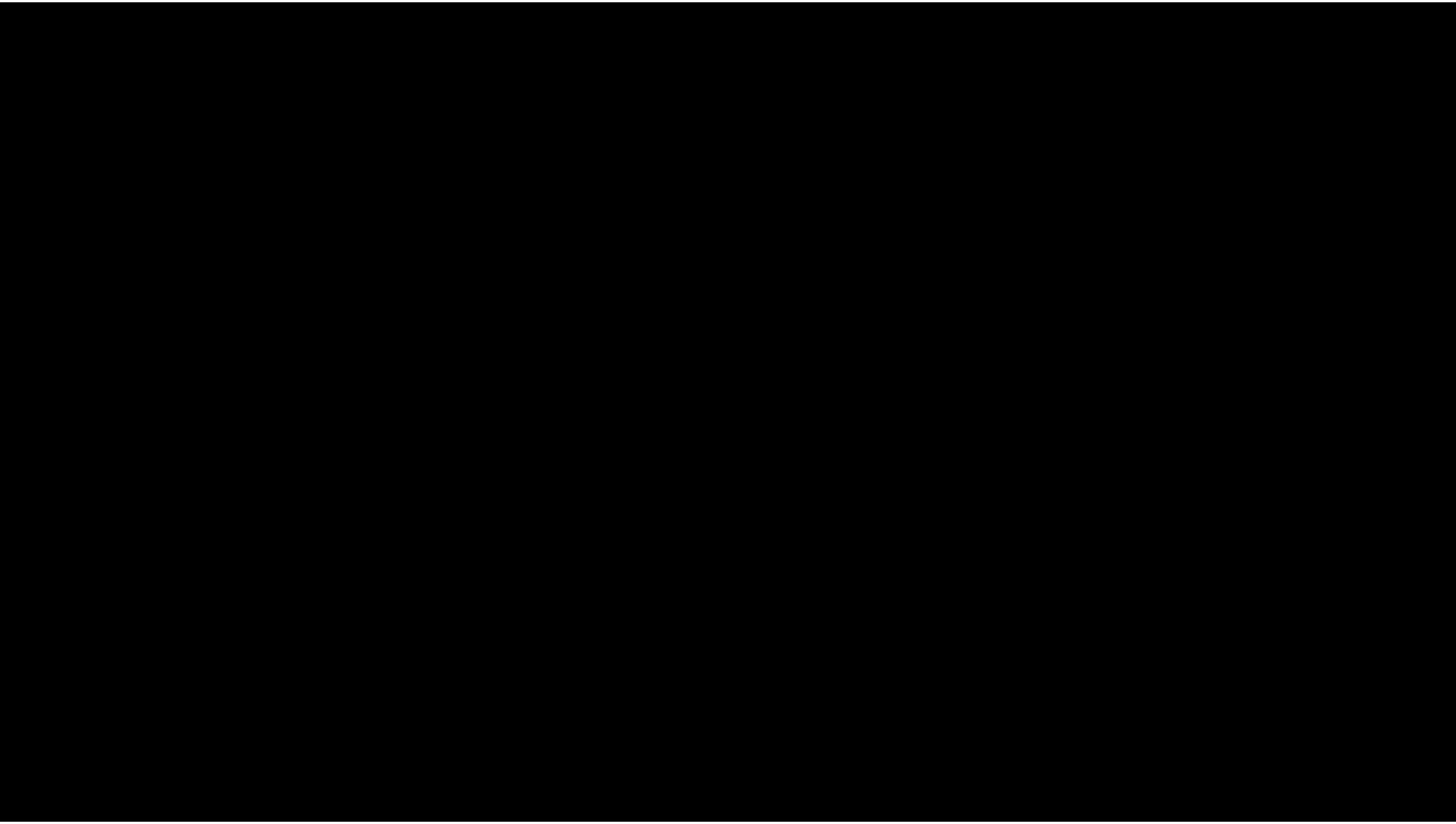
- For example image restoration / deconvolution
- Using **general** assumptions to perform the restoration
- Idea: leverage the specific experimental data, without prior assumptions

Content aware image restoration



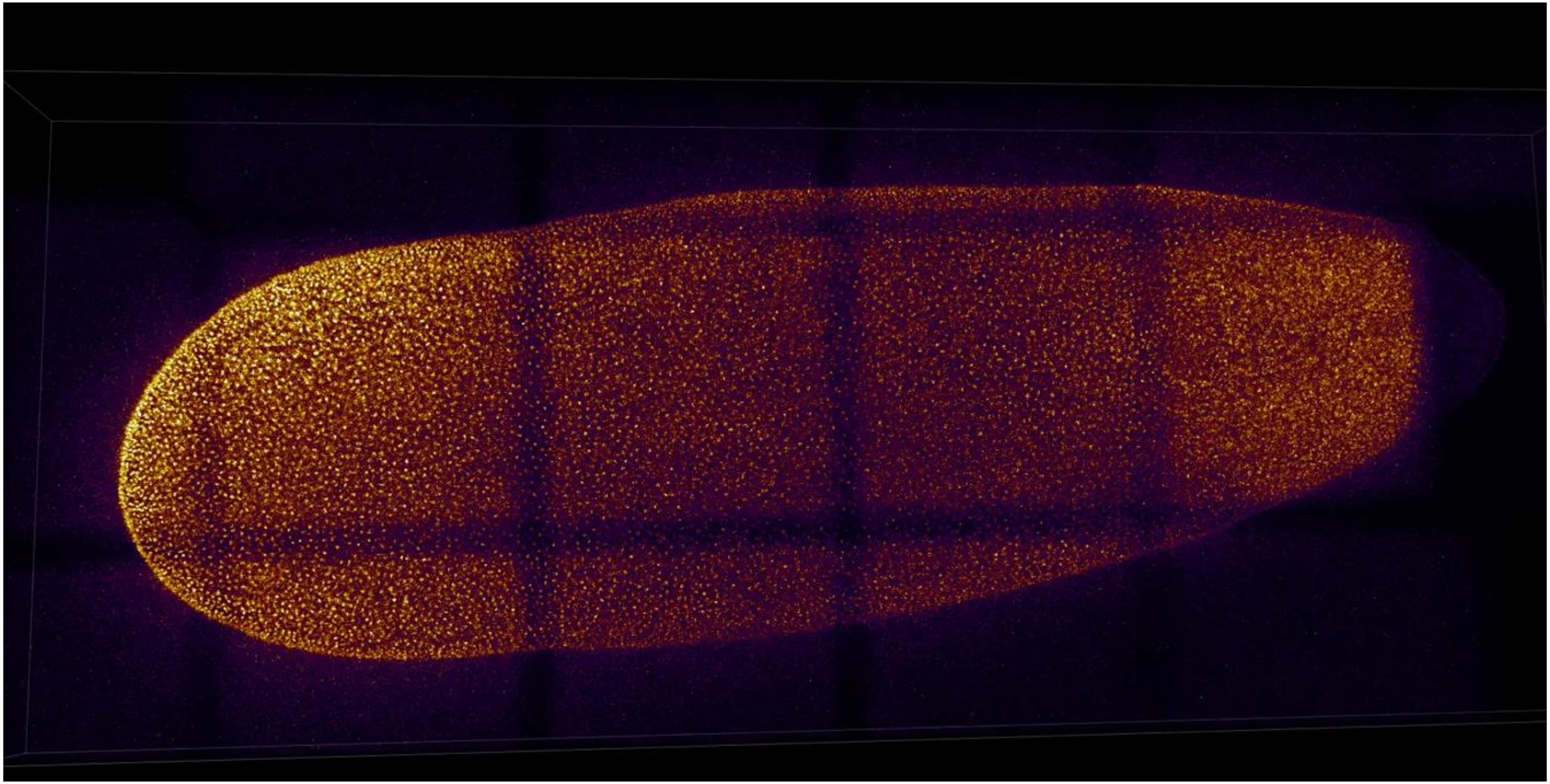
Weigert et al. (2018)

Content aware image restoration



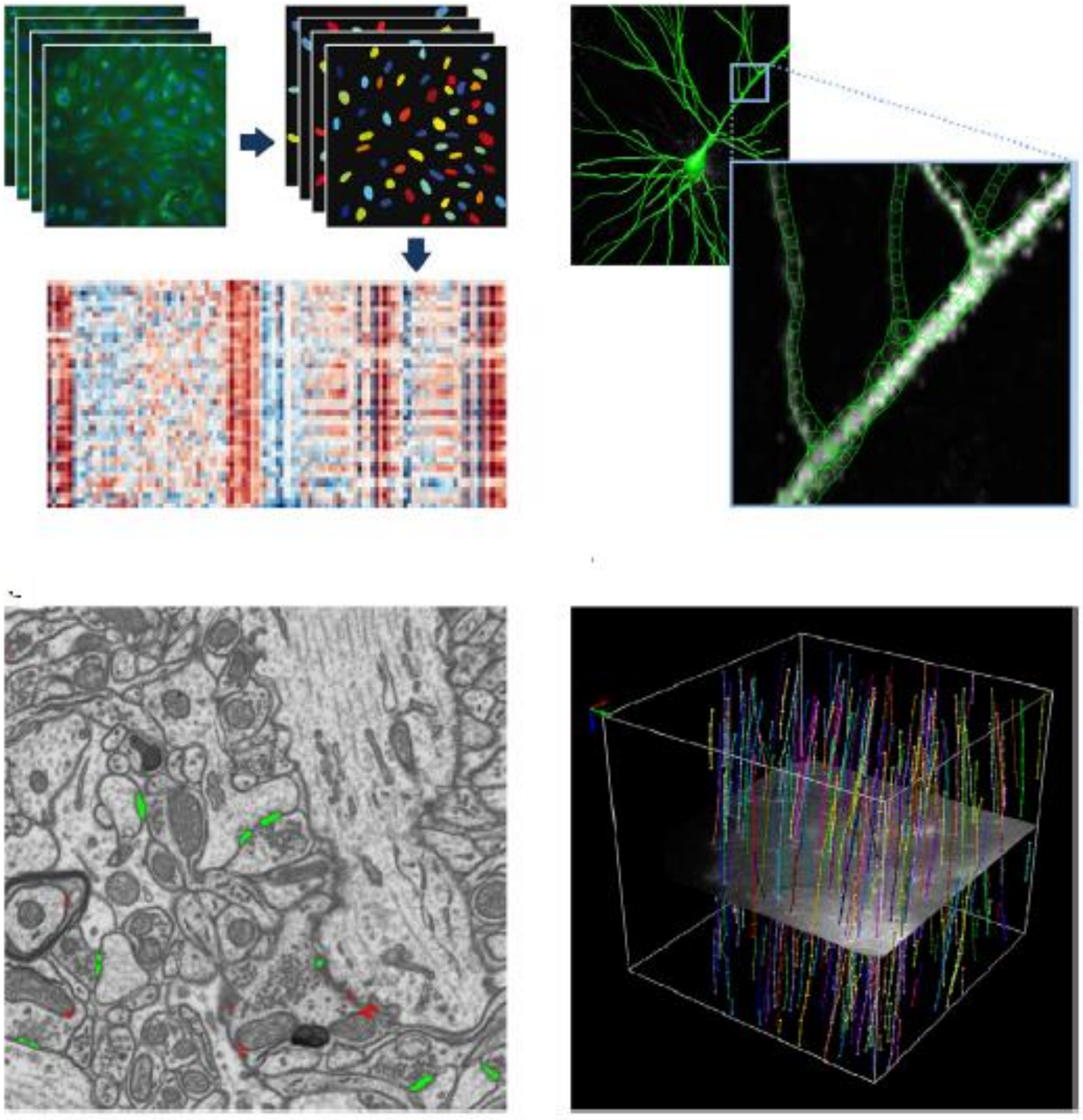
Weigert et al. (2018)

Content aware image restoration



Weigert et al. (2018)

Bioimage analysis applications



Meijering et al. (2016)

Bioimage analysis tools

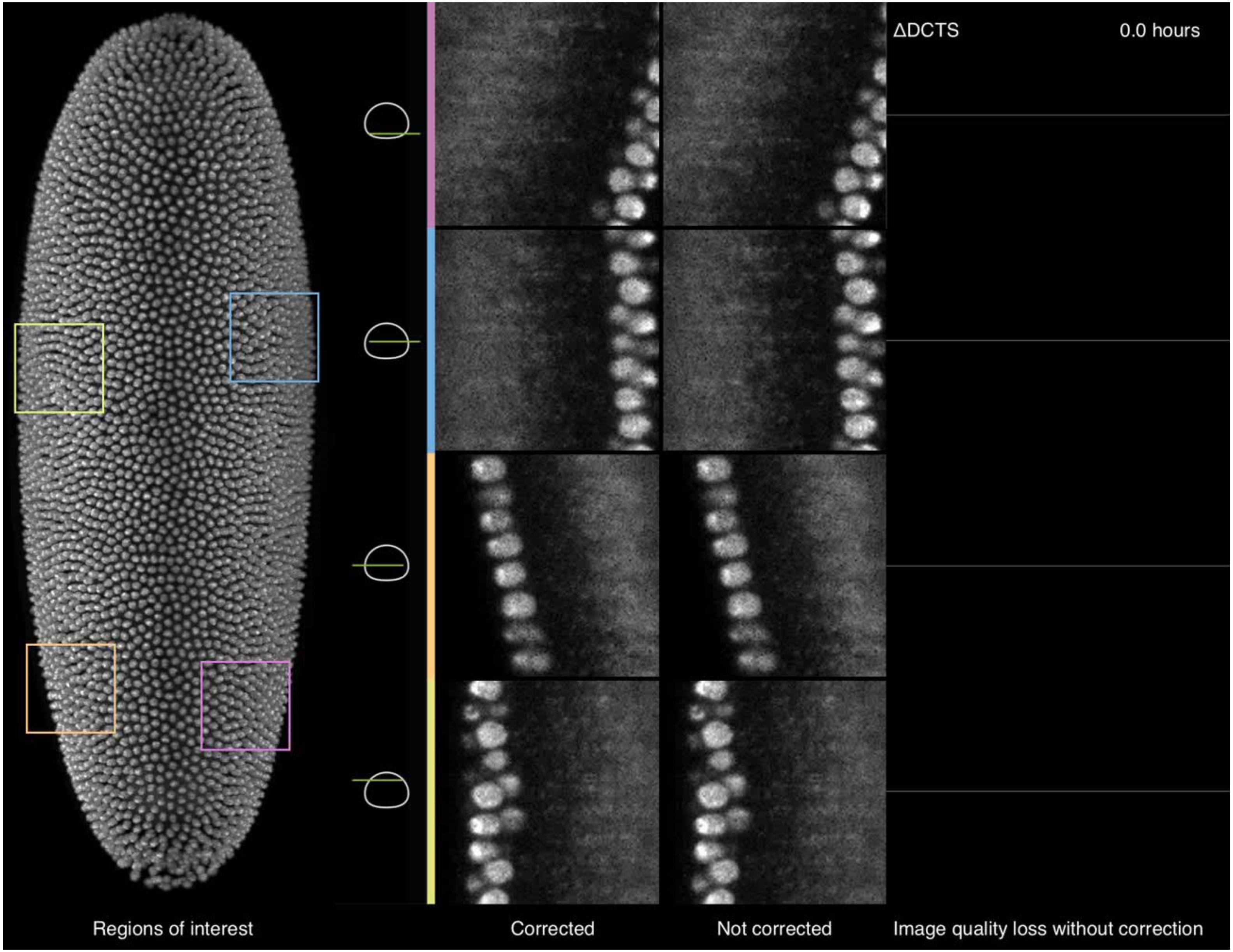
ImageJ
Image Processing and Analysis in Java

CellProfiler
cell image analysis software

FEI
part of Thermo Fisher Scientific

BITPLANE
SCIENTIFIC SOLUTIONS

Adaptive (event-driven) microscopy

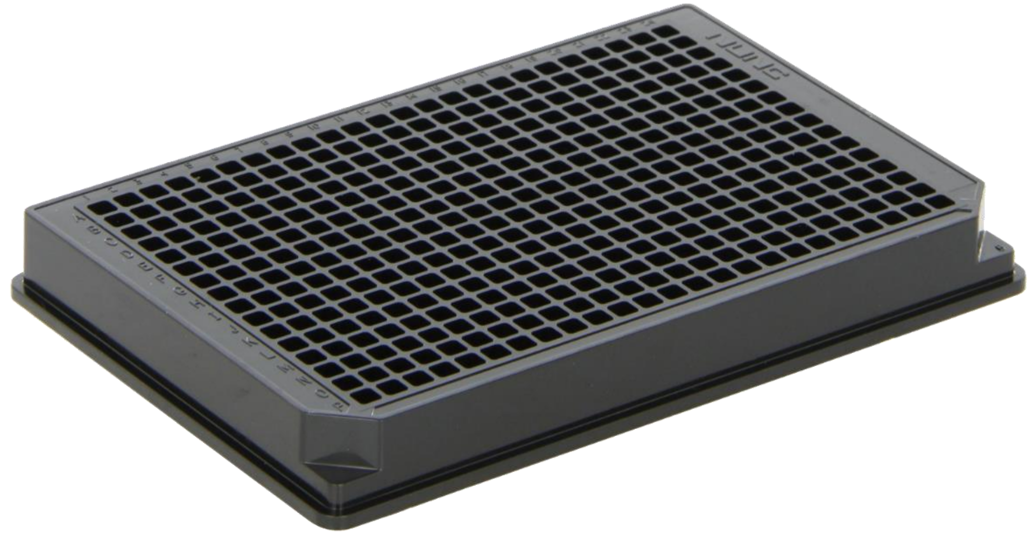


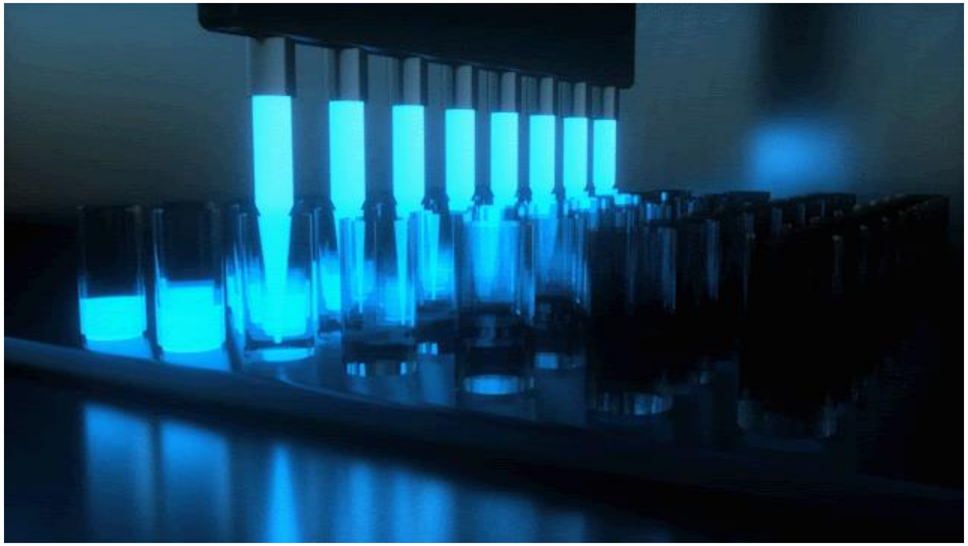
Royer et al. (2016)

Cell morphology is a marker for its functional state

Discovering drugs in high throughput

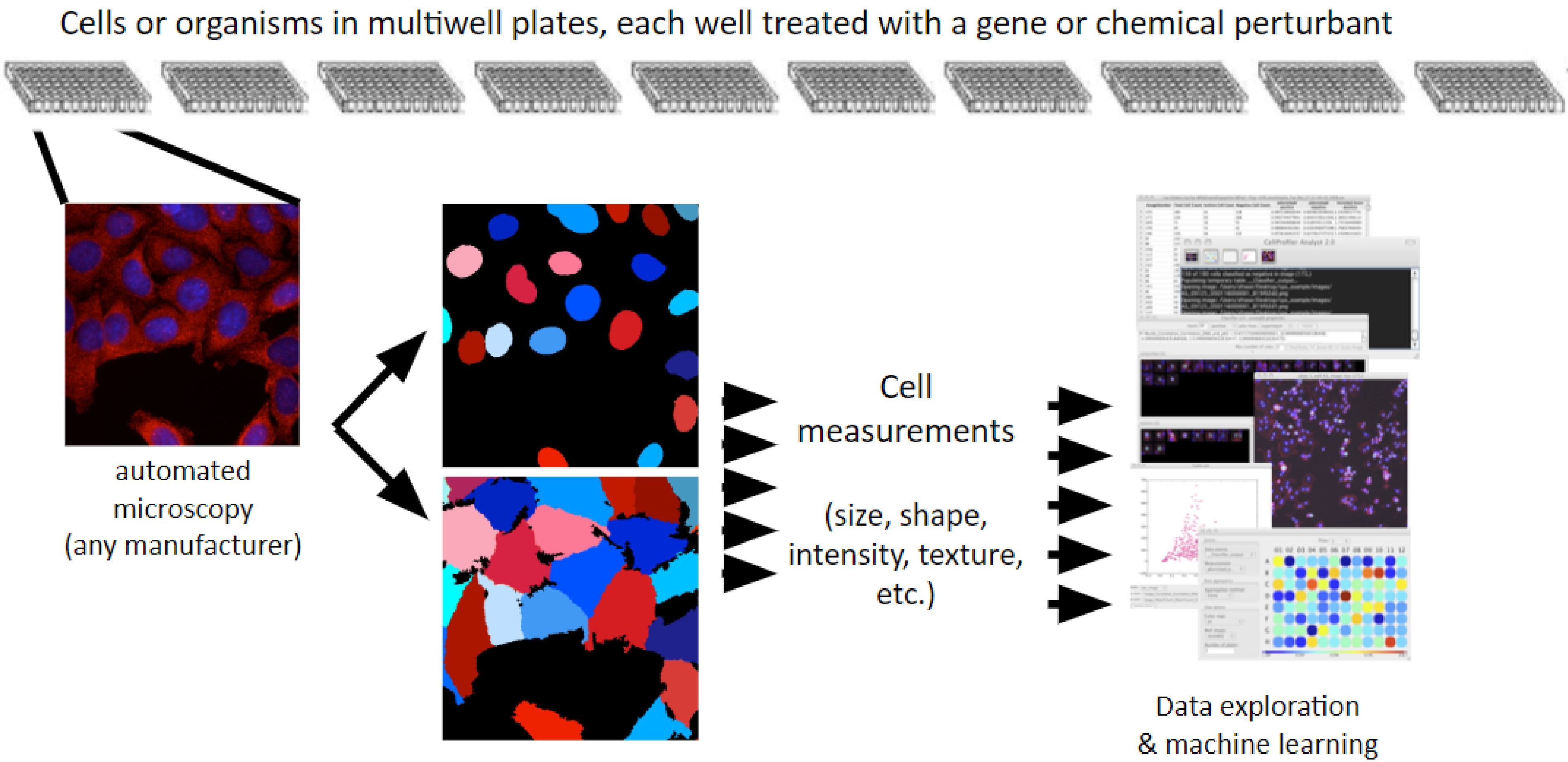
X 384 =





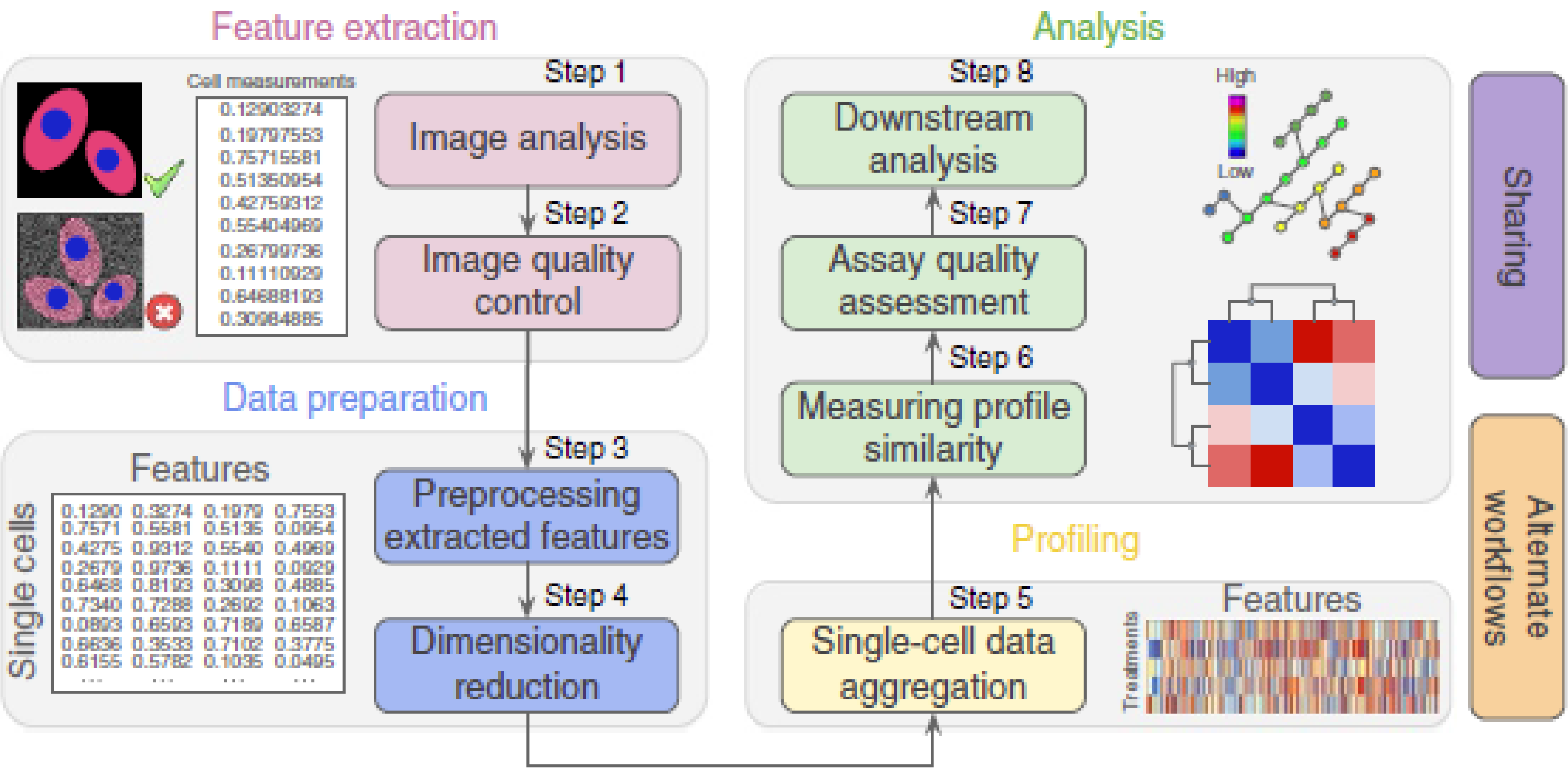
Anne Carpenter, Image: Nalgene; video: Chemistry World

Large scale imaging experiments



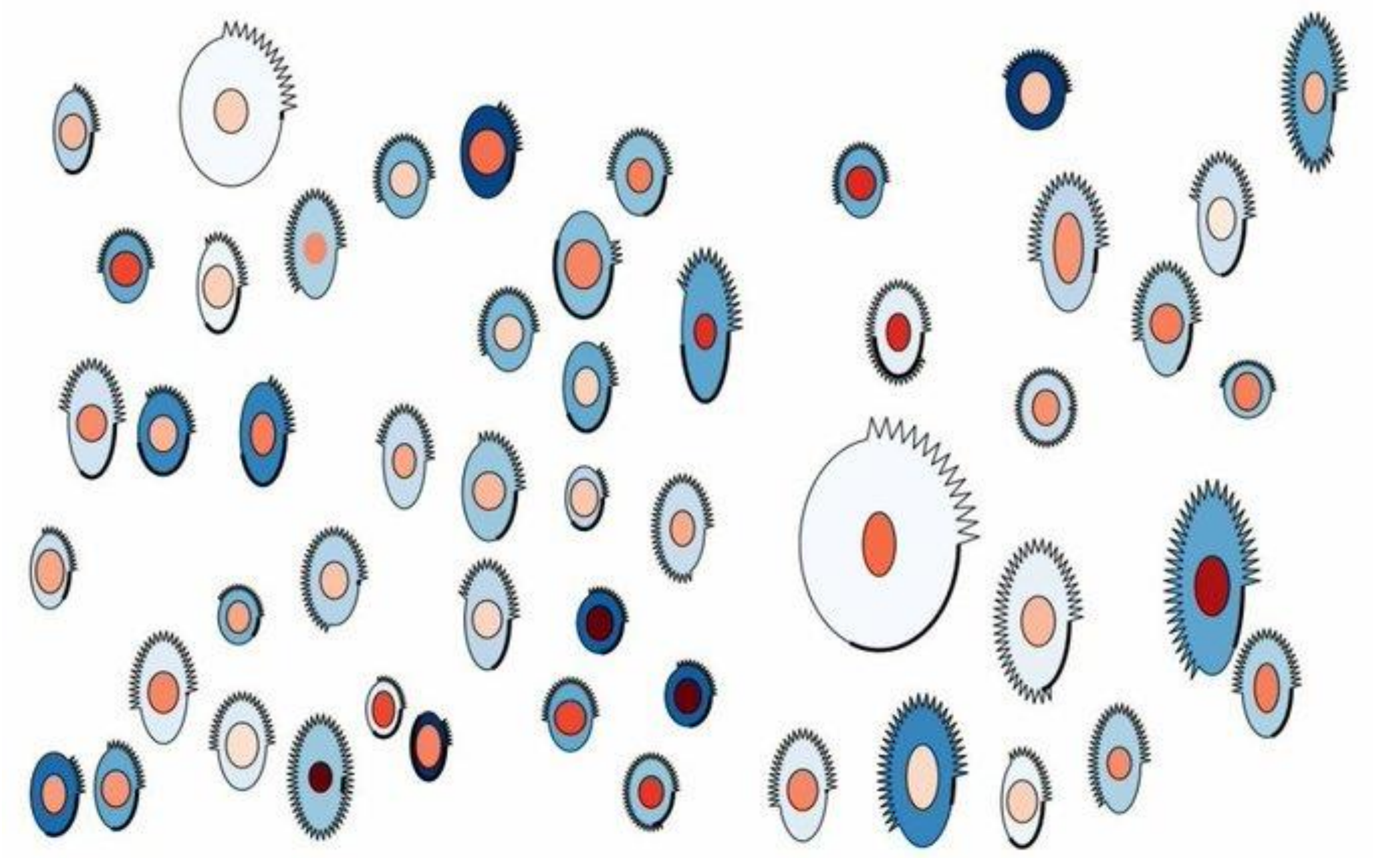
Anne Carpenter

High content single cell phenotypic profiling



Caicedo et al. (2017)

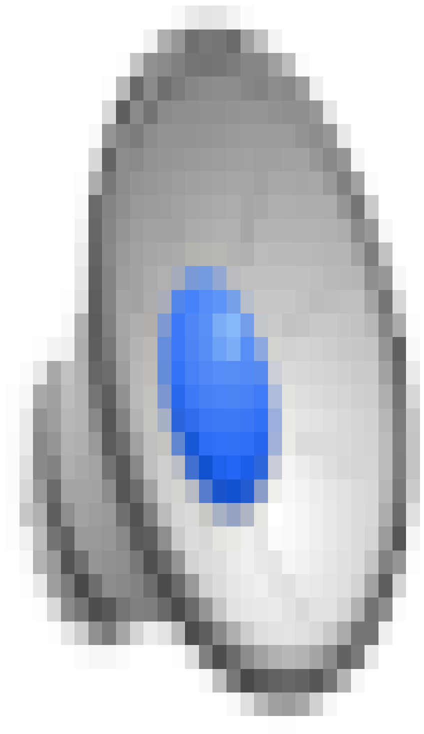
Visualization



Sailem et al. (2015)

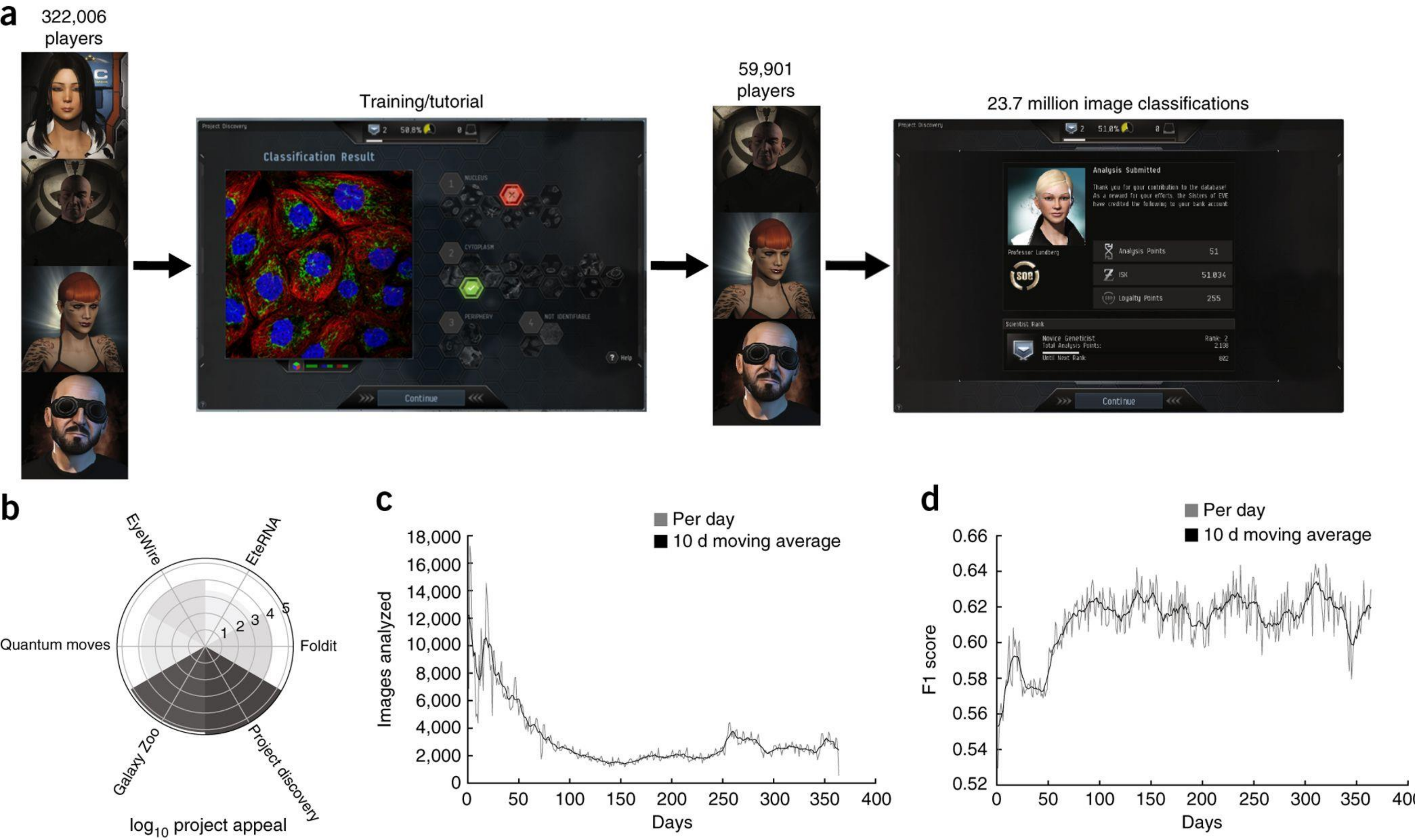
Cell shape in 3D

Immune cell in 3D



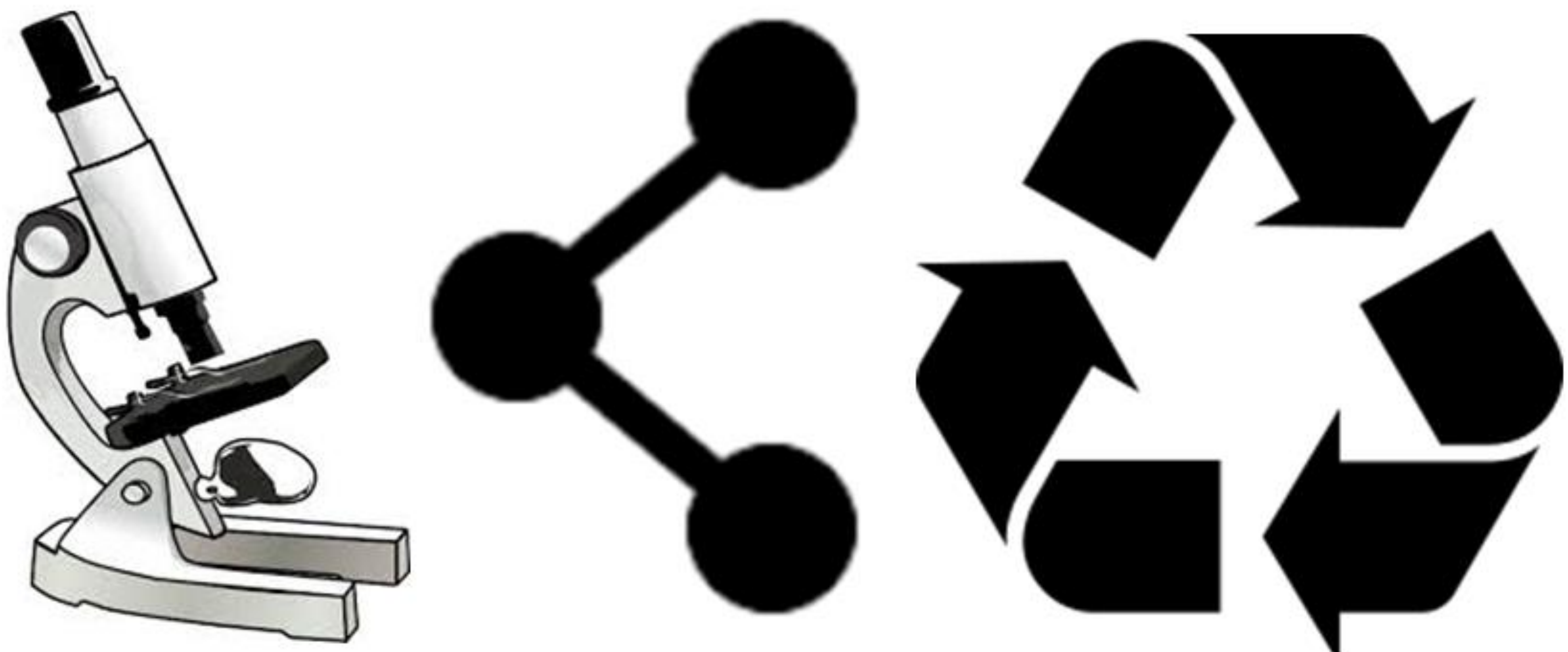
Renkawitz et al. (2019)

Crowd source for annotations..

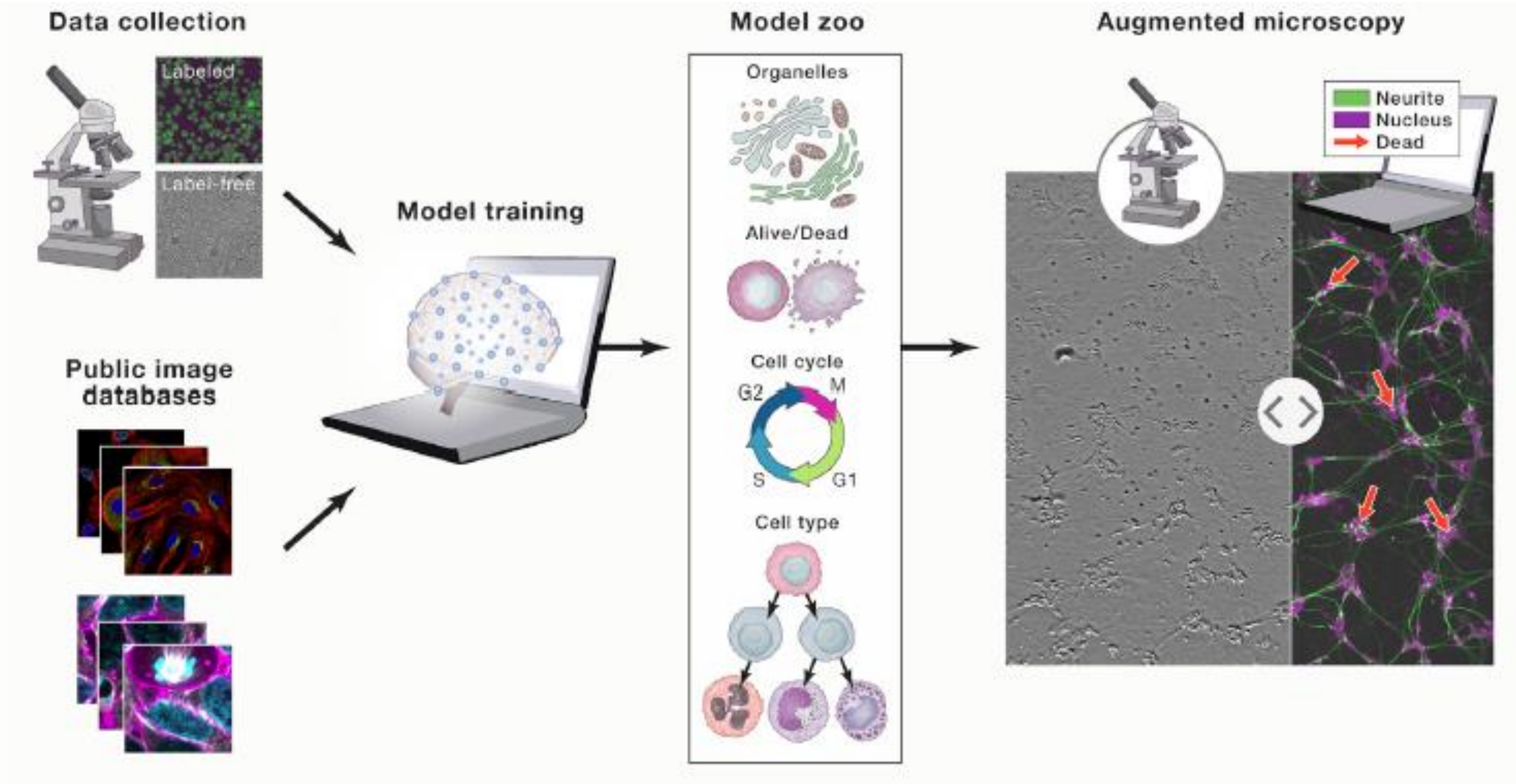


Sullivan and Winsnes (2018)

Sharing and reusing cell imaging data

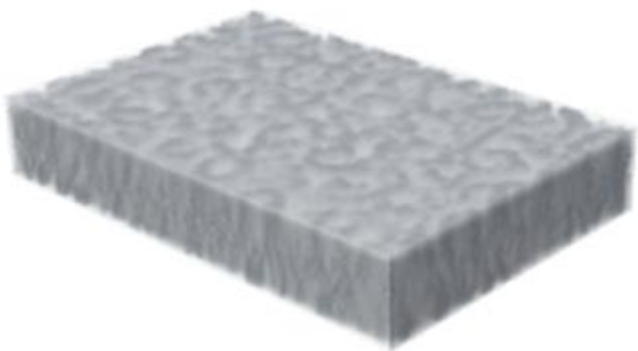


A future of augmented microscopy



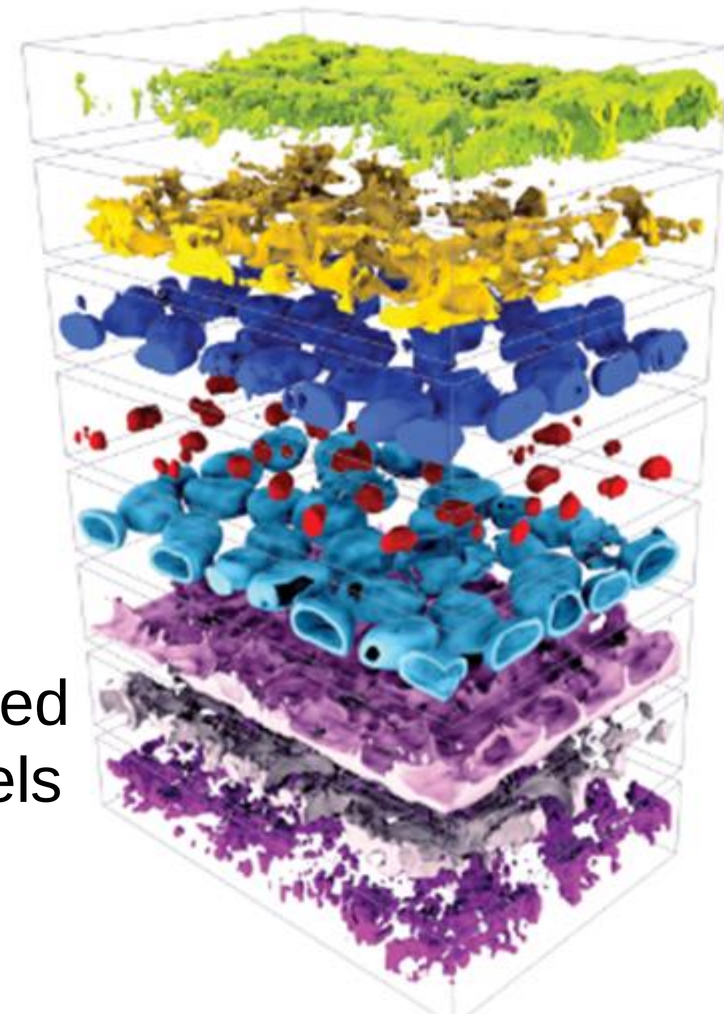
Sullivan and Lundberg (2018)

Label-free images contain information on the molecular organization of the cell



Bright field

Predictive models

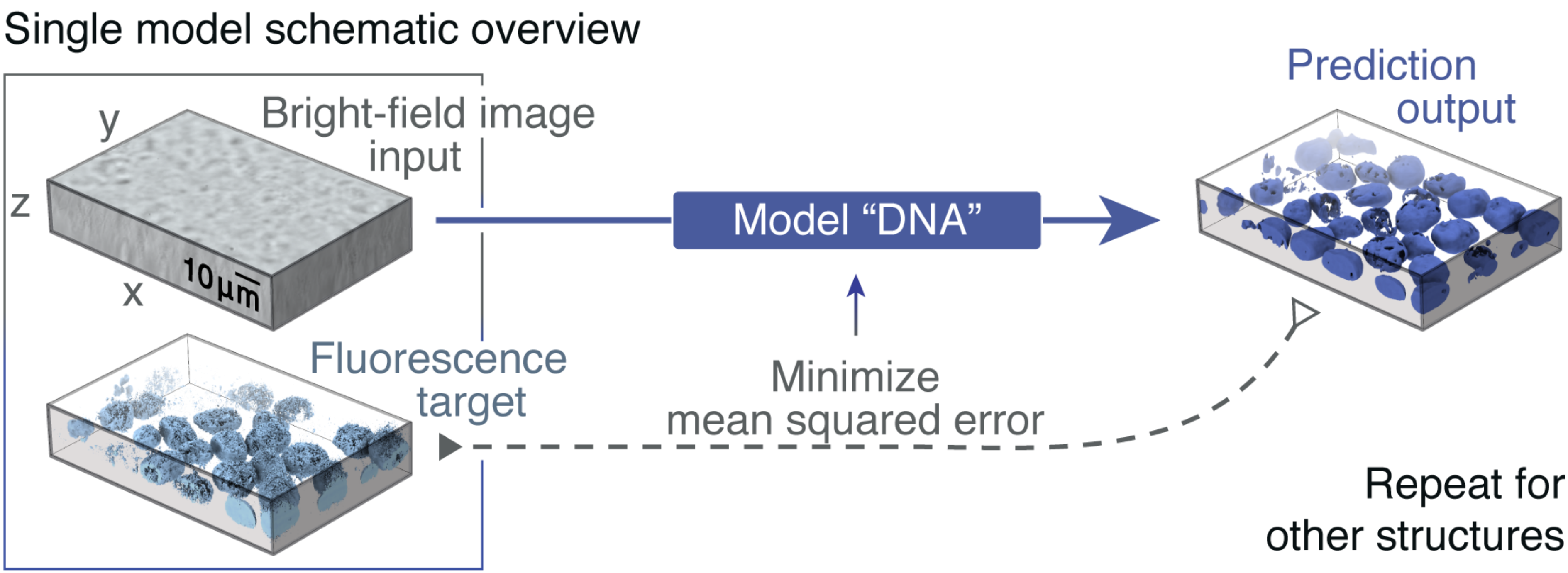


Predicted channels

Ounkomol et al. (2018)

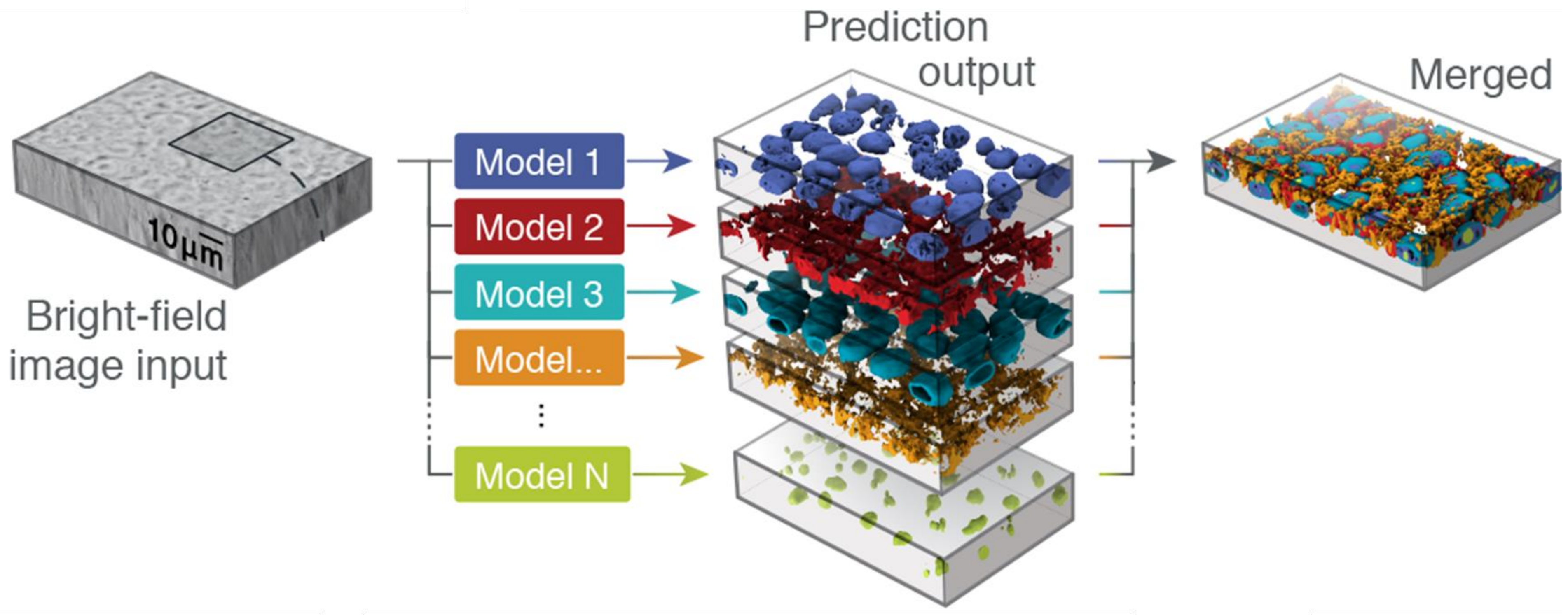
Christiansen et al. (2018)

Unstructured-to-structured information with supervised models



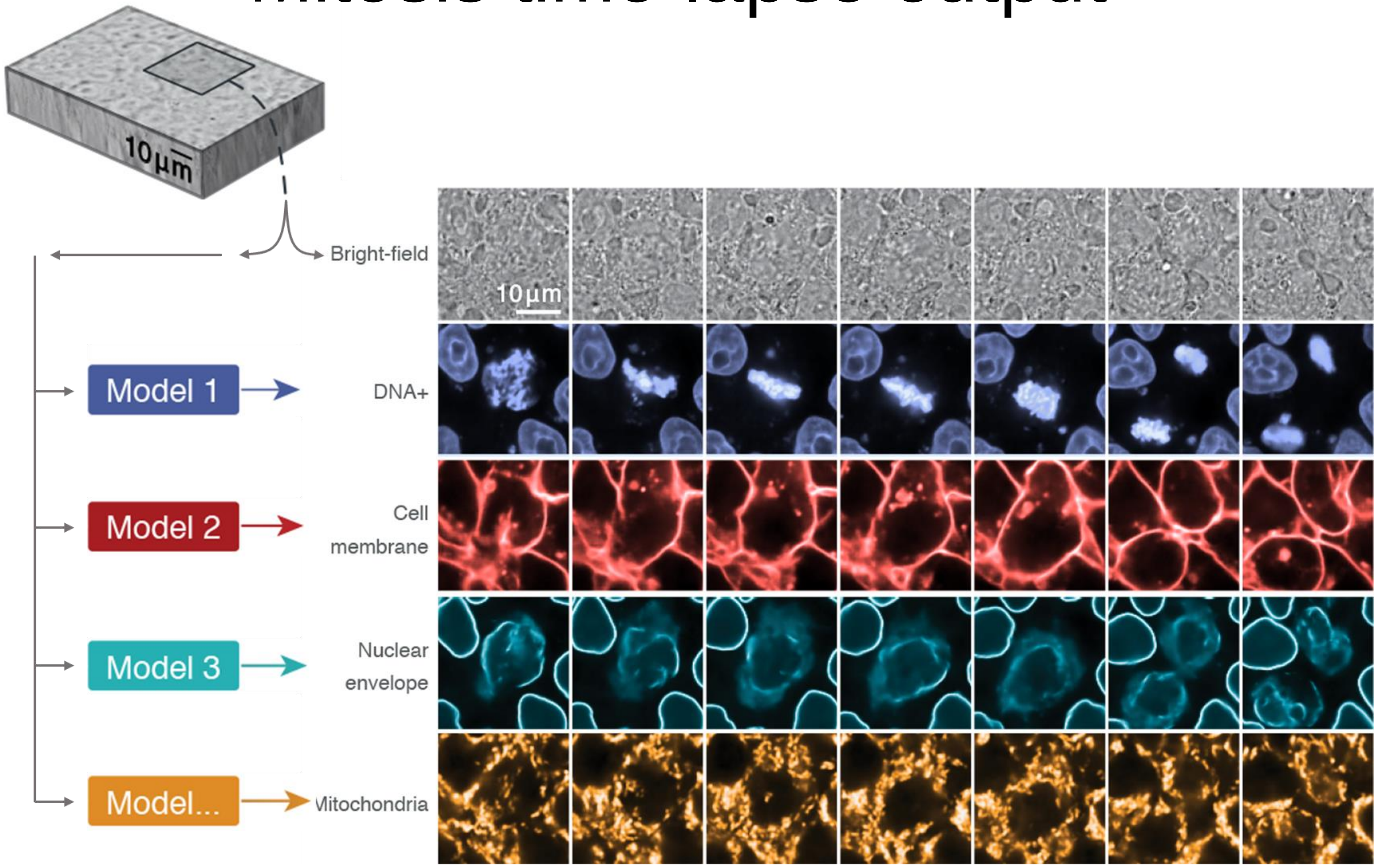
Ounkomol et al. (2018)

Combining multiple models



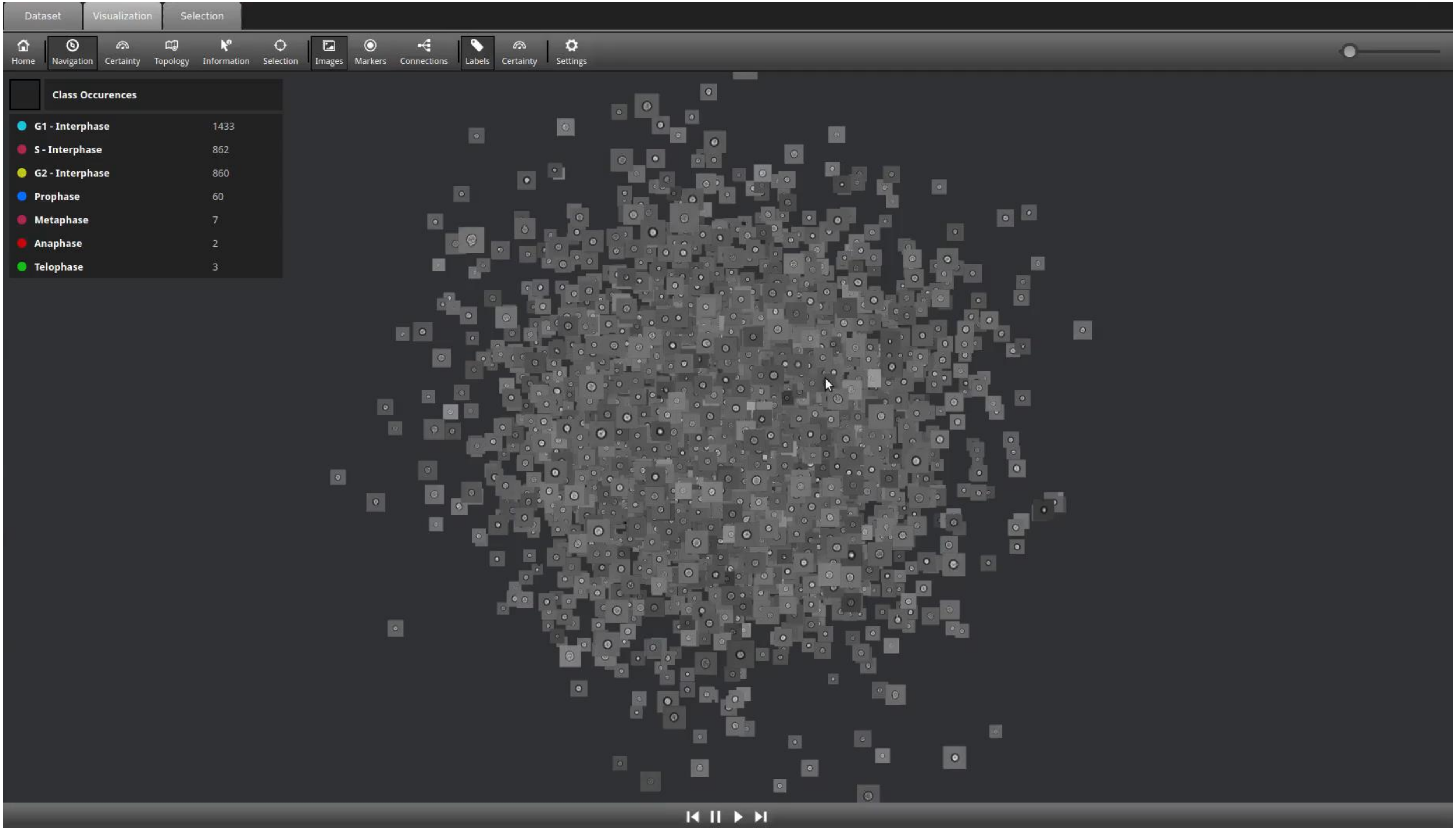
Greg Johnson

Mitosis time-lapse output



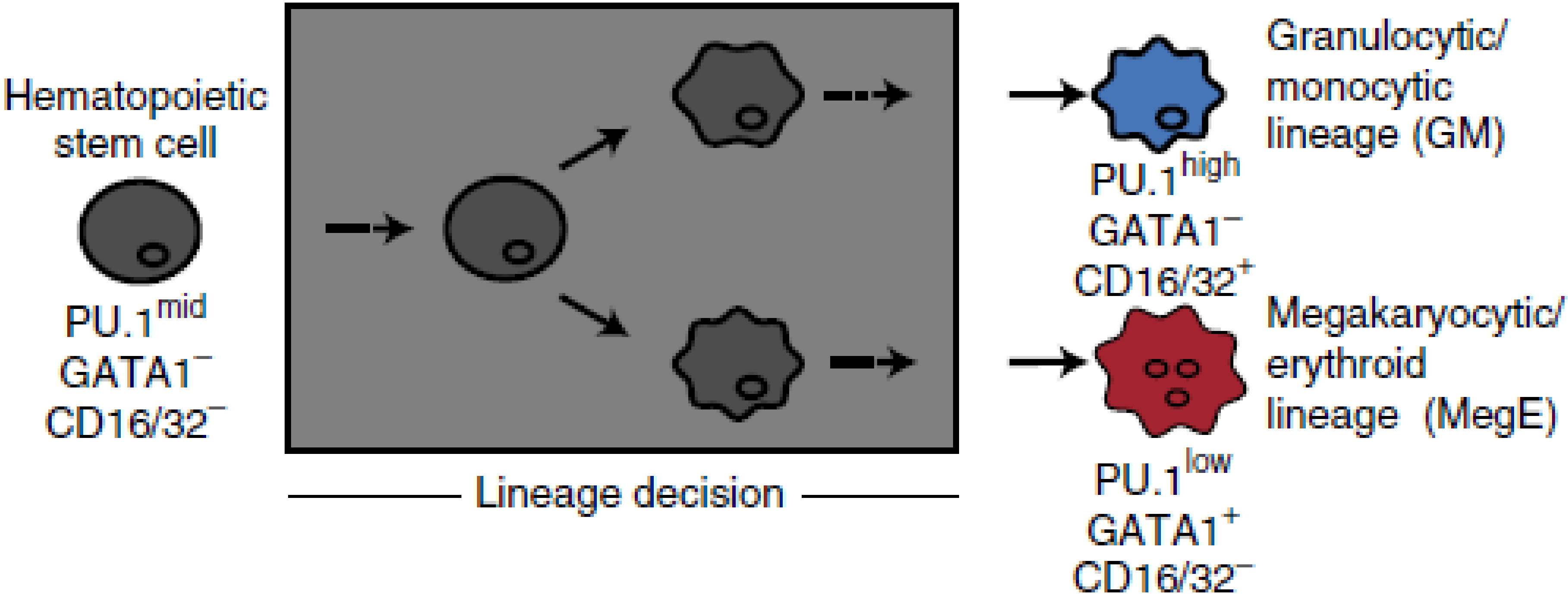
Greg Johnson

Predicting cell cycle



Eulenberg and Köhler et al. (2017)

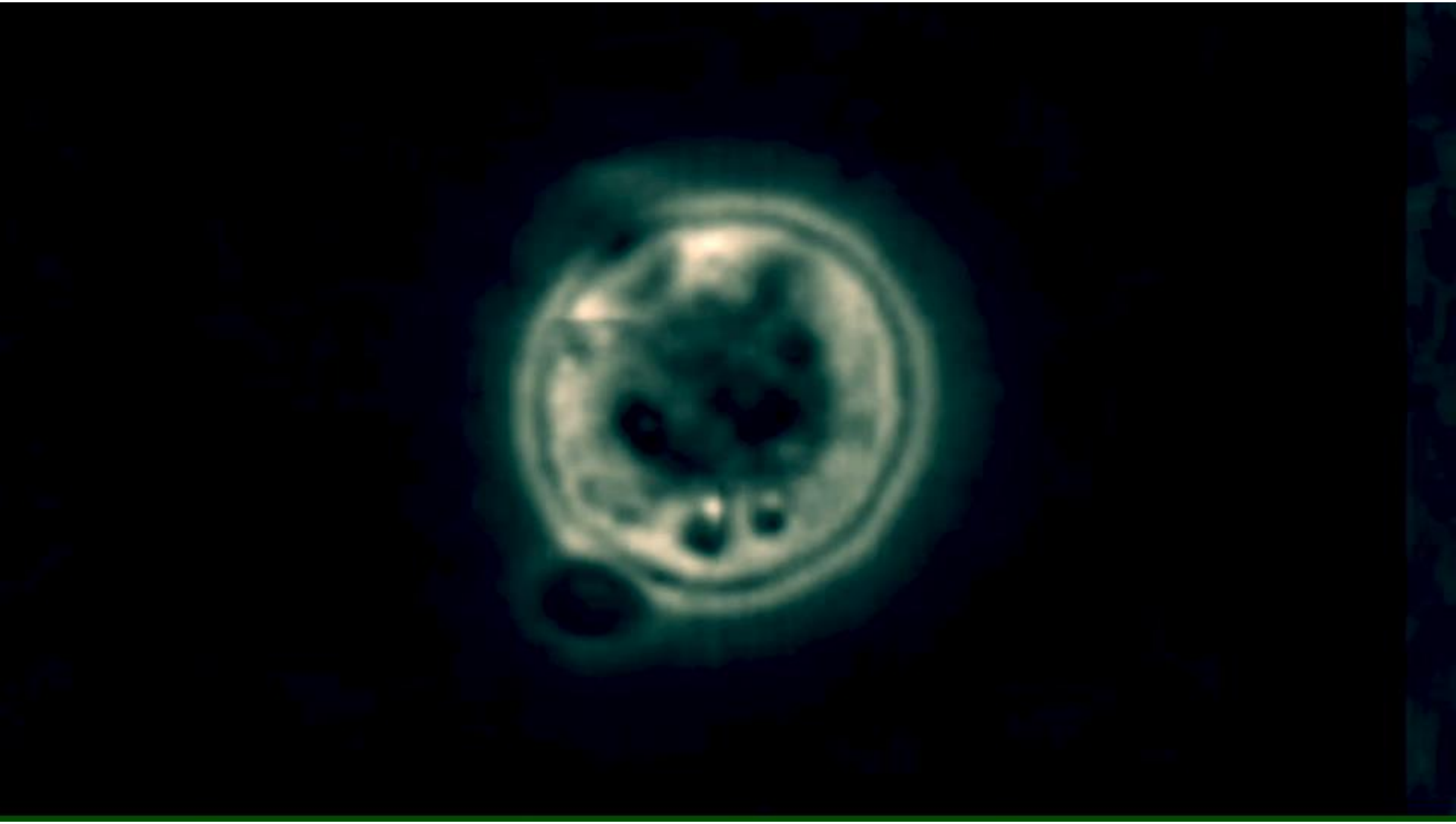
Predicting lineage choice



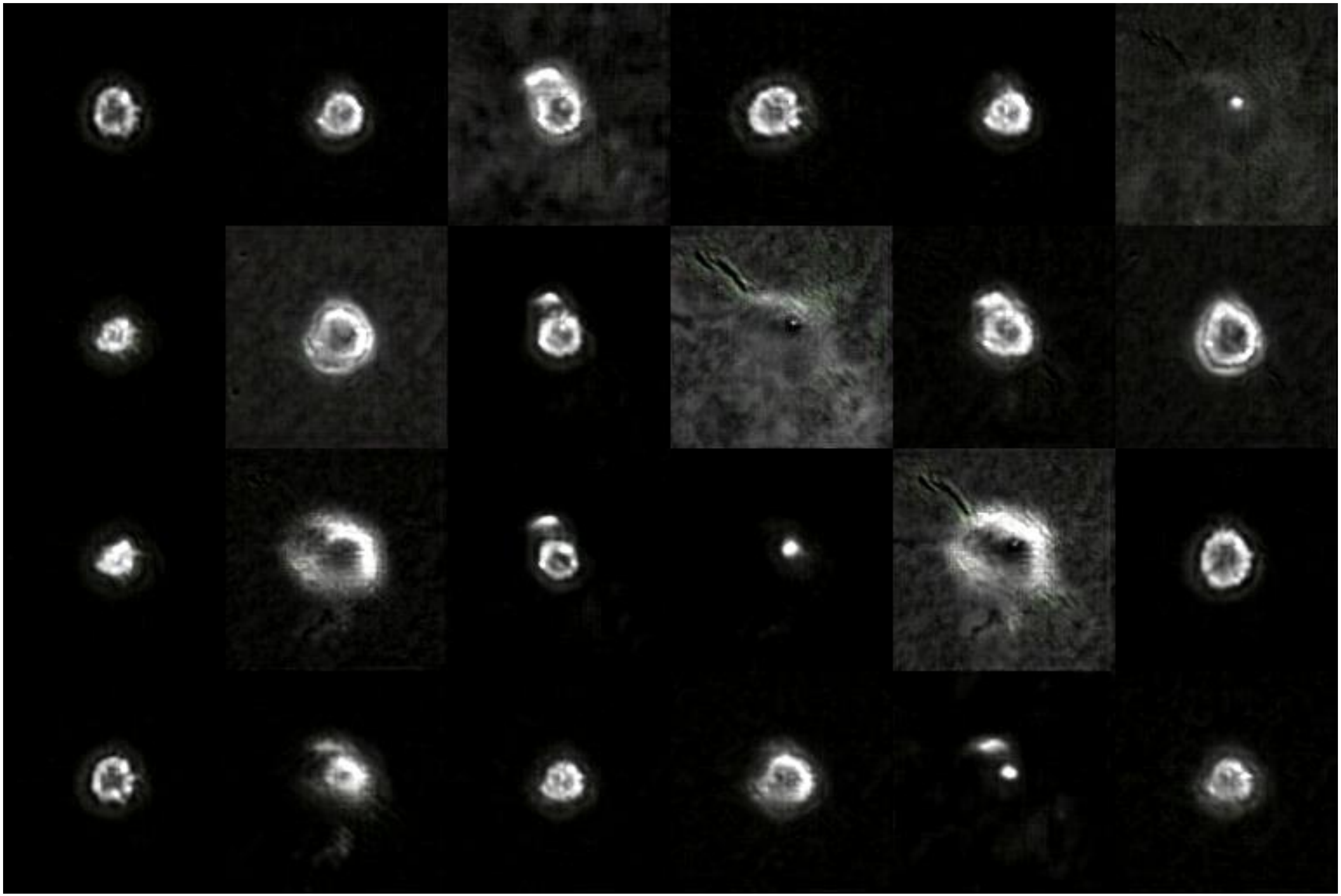
Buggenthin and Buettner et al. (2017)

Can we predict cancer cell functional states from live label-free cell images?

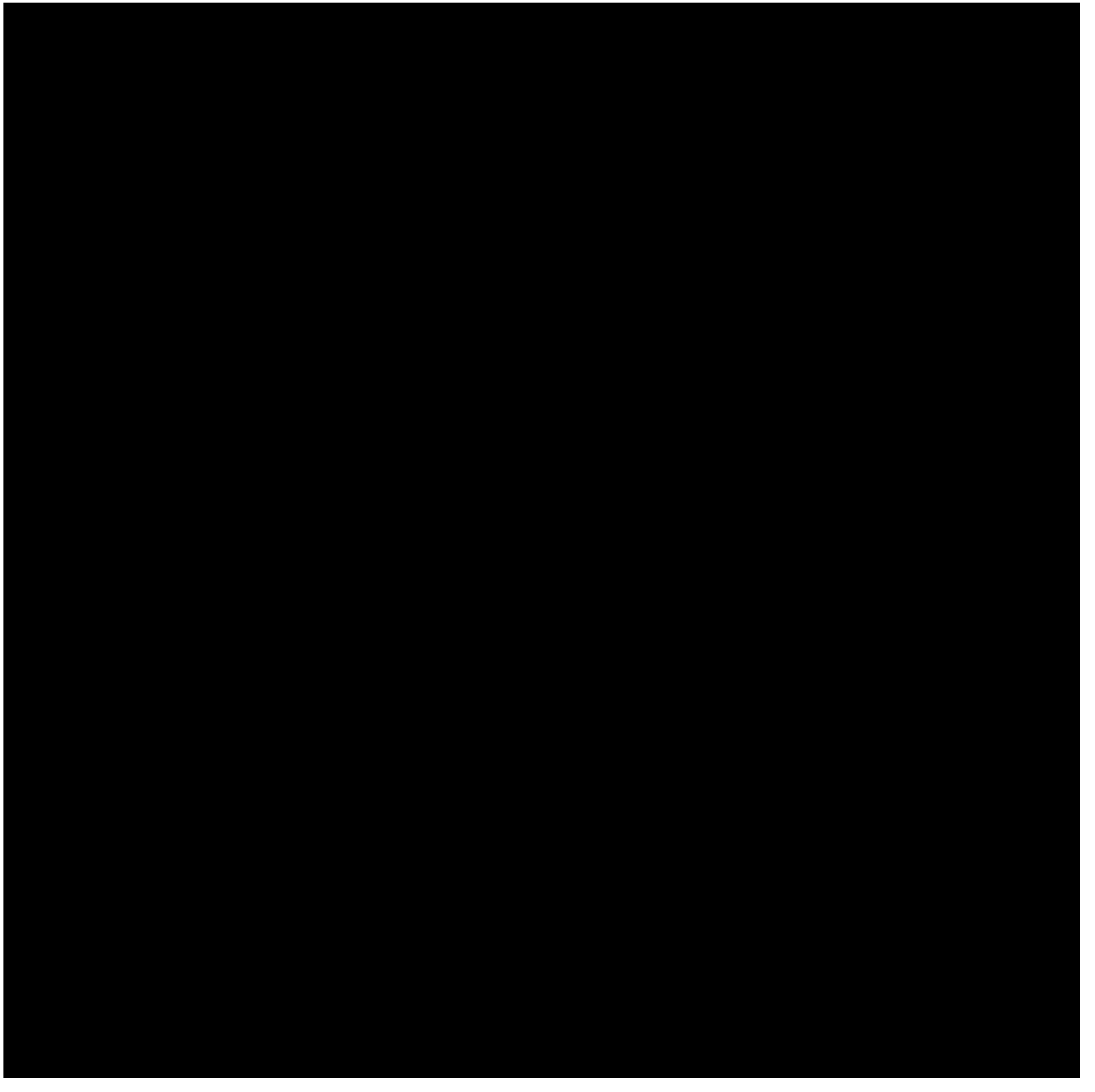
Interpretable machine learning for melanoma classification



Morphing cells

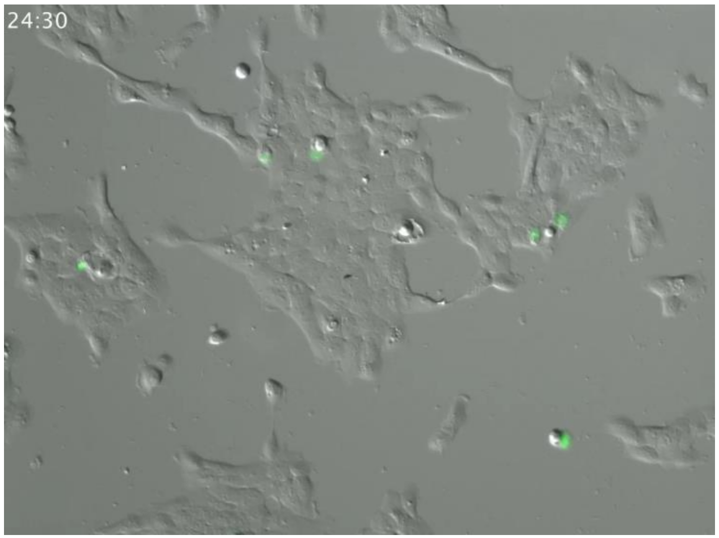


“Synthetic” cell transitioning



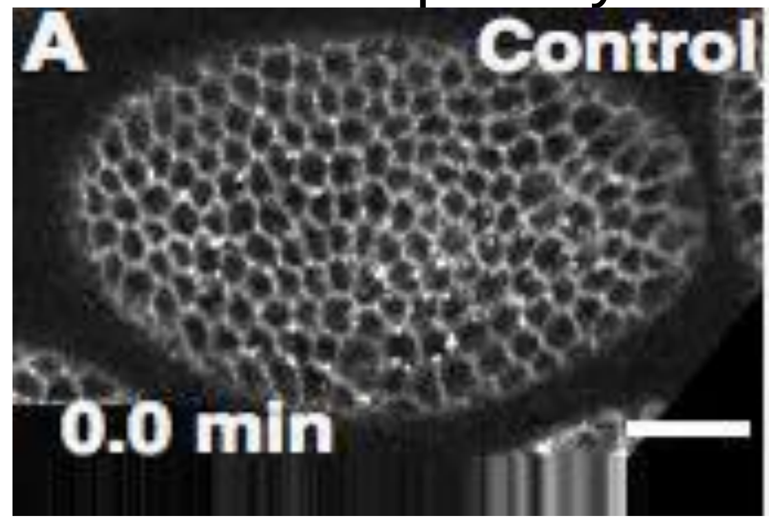
Collective cell behavior emerges from single cell behavior and cell-cell communication

Collective cell death



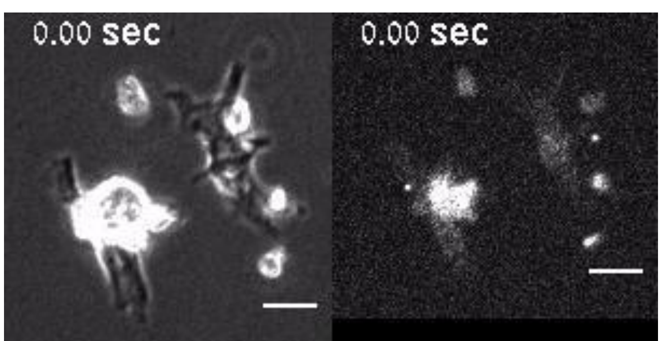
Overholtzer lab

Planar cell polarity



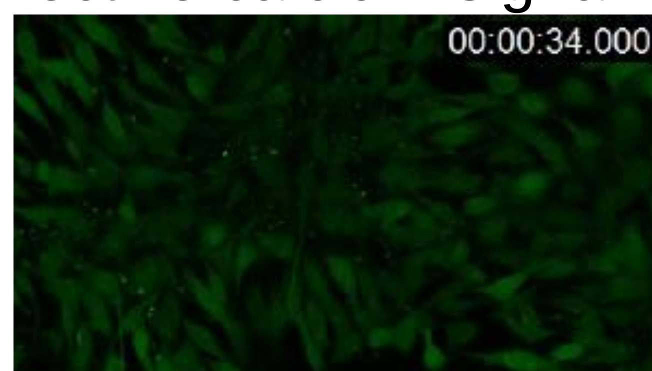
Barlan et al. (2017)

Synchronized cardiac cells



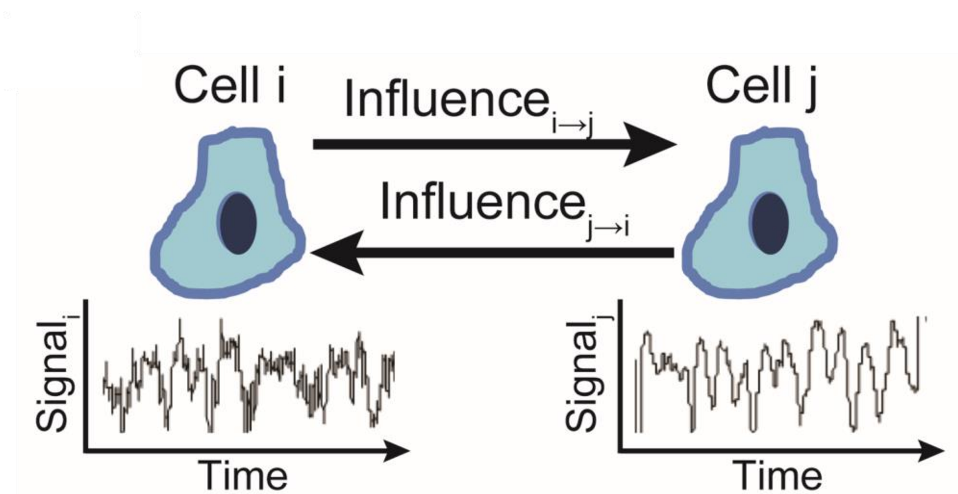
Nitsan et al. (2016)

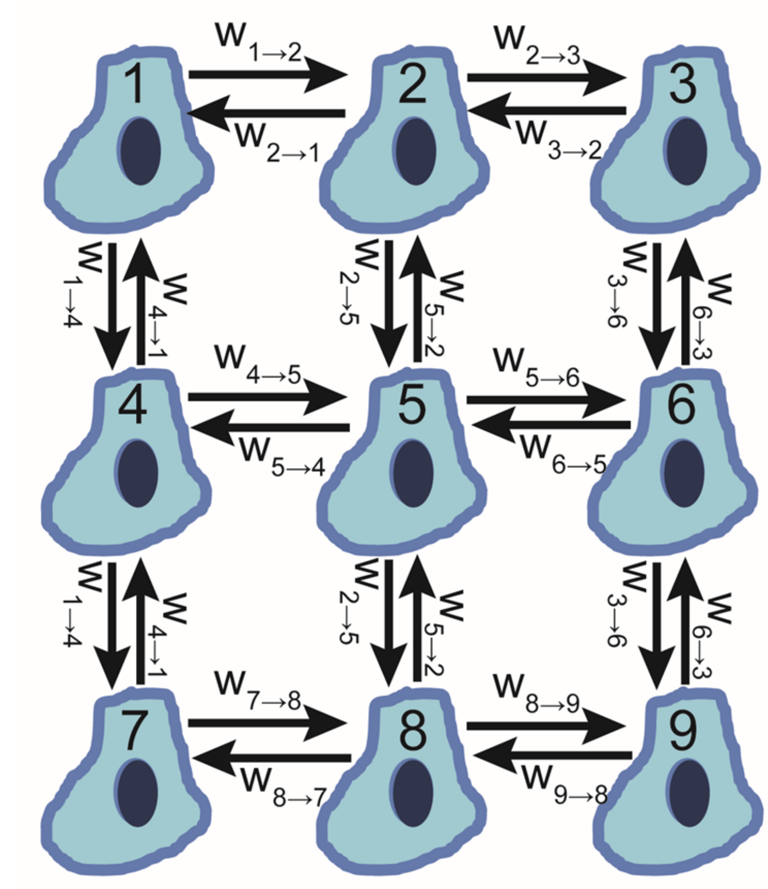
Collective calcium signaling



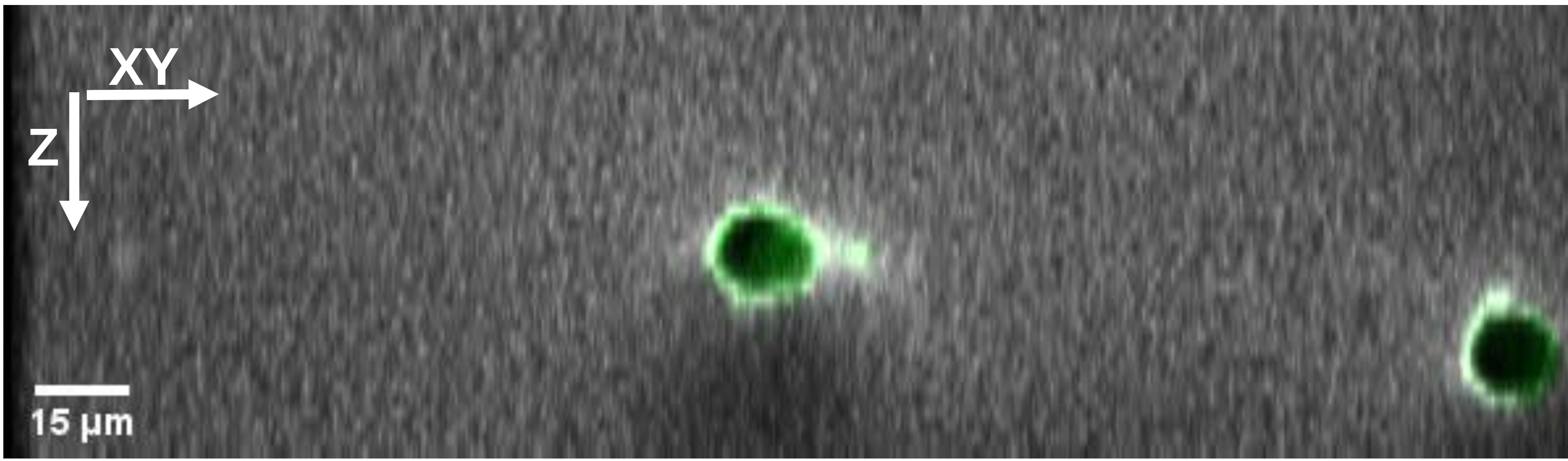
Sun et al. (2012)

Modeling networks of spatial influence at the single cell resolution



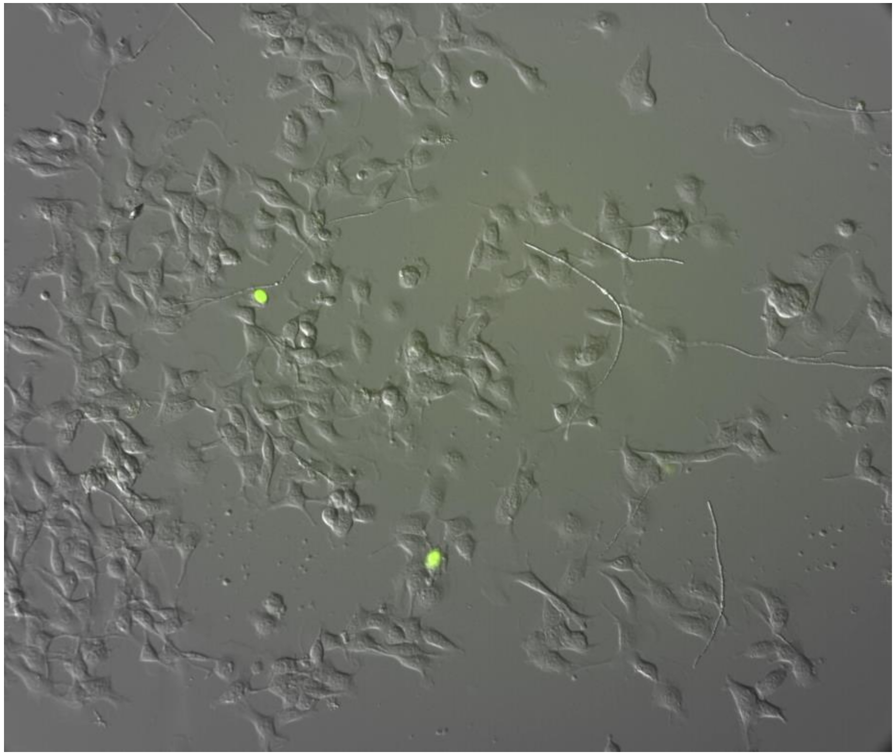


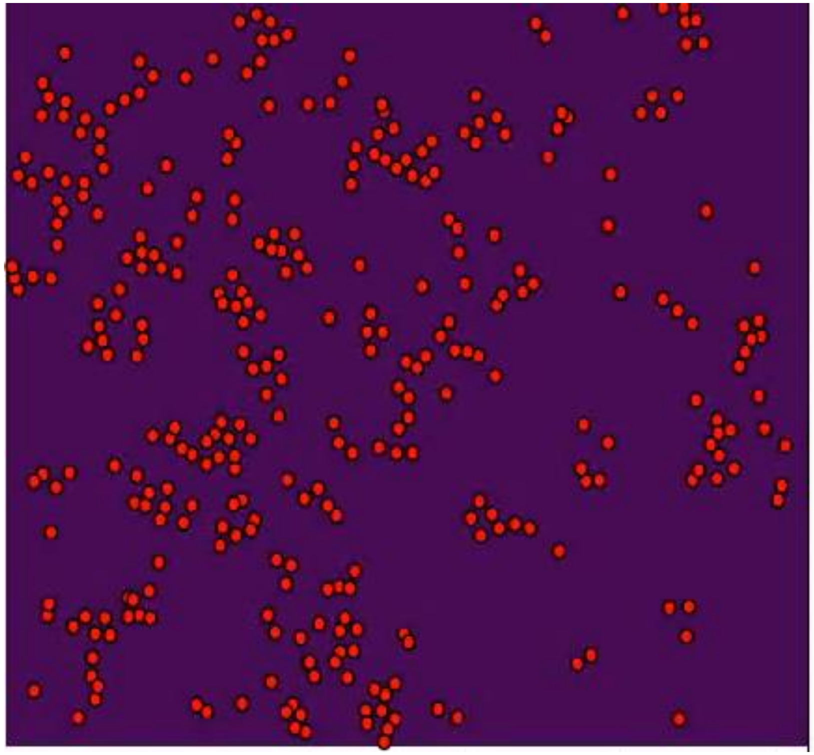
Quantifying cell-ECM and cell-ECM-cell interactions



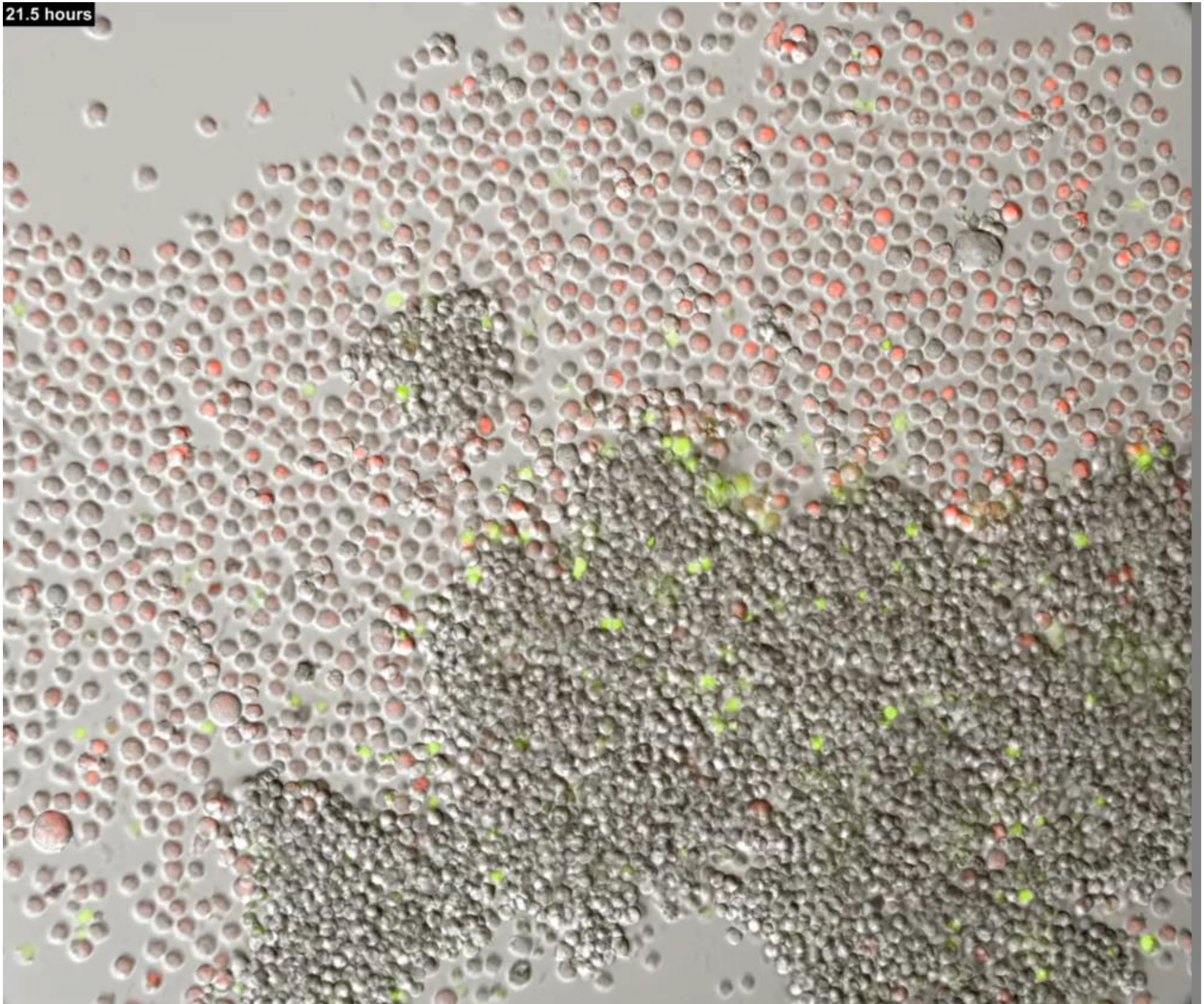
Assaf Nahum

Combining quantitative imaging and simulations for spatiotemporal characterization of collective cell death



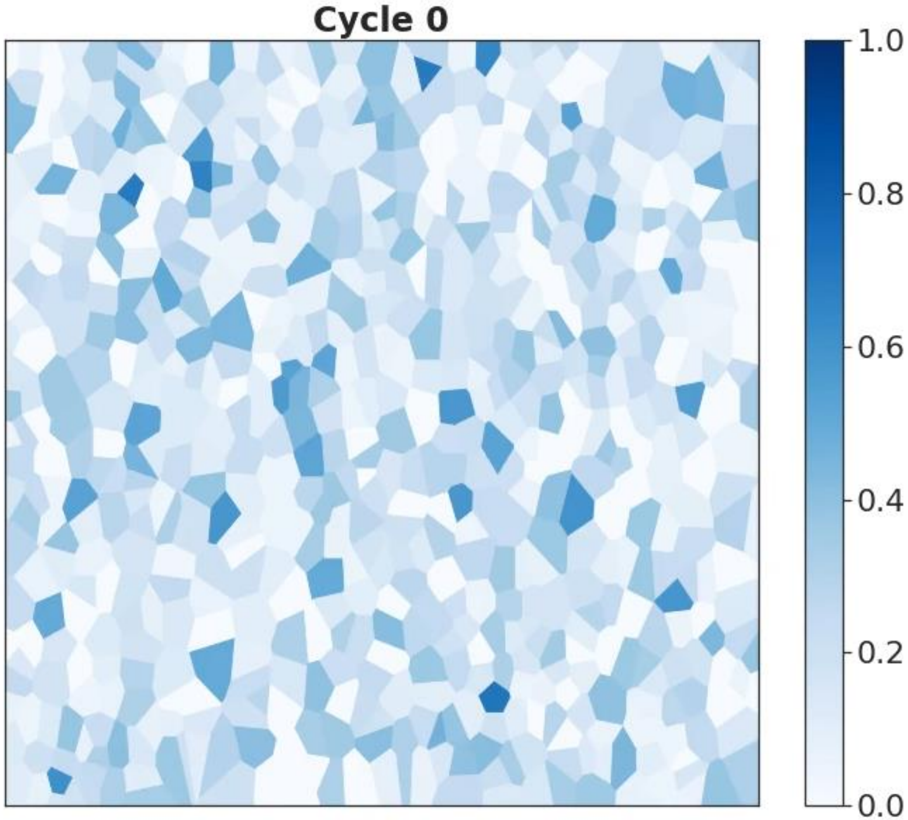


Chen Galed

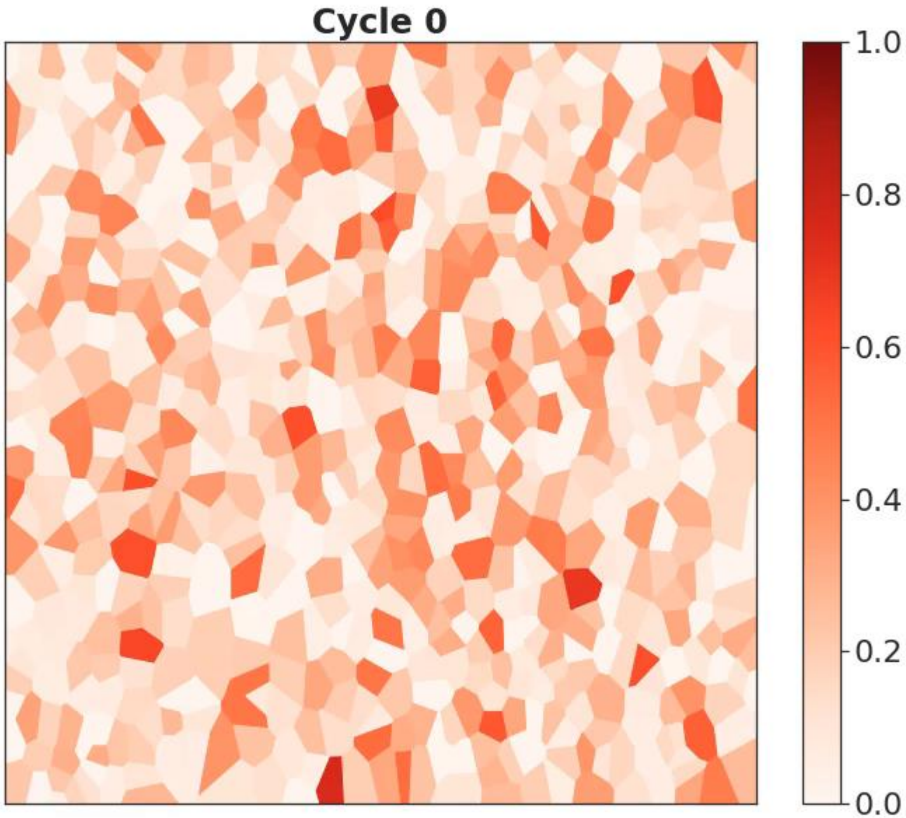

21.5 hours

Emergence of synchronized multicellular mechanosensing from spatiotemporal integration of heterogeneous single-cell information transfer

Transmission score

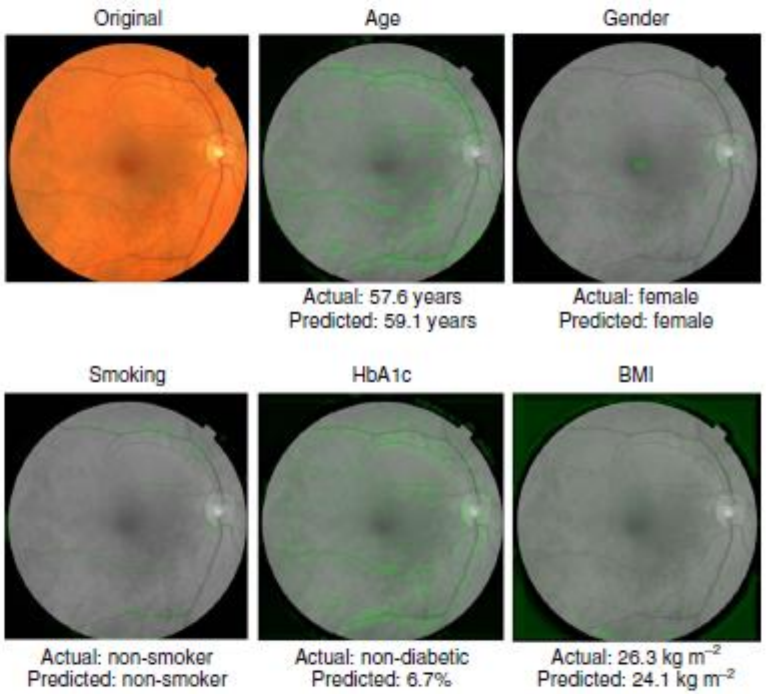


Receiver score

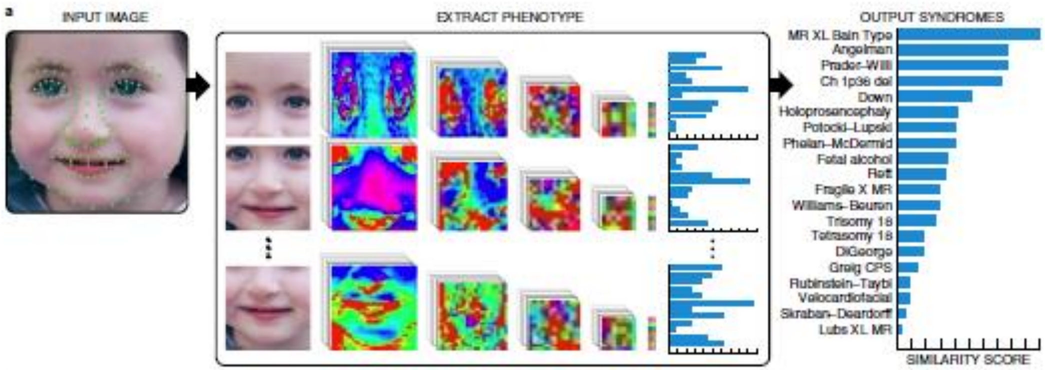


Amos Zamir

Medical applications



Poplin, Varadarajan, Peng and Webster et al. (2018)



Gurovich et al. (2019)

Research project types

- Data mining and integration in public repositories
- Collaborative bioimage analysis projects
- Tool building projects

Open resources for your research projects

- Image data resource (IDR), Williams et al. (2017), Image Data Resource: a bioimage data integration and publication platform https://idr.openmicroscopy.org/
- The Allen Institute of Cell Science
- The Human Protein Atlas
- Branda Andrews datasets http://sites.utoronto.ca/andrewslab/data.shtml

Open resources for your research projects

- Bray et al. (2017), A dataset of images and morphological profiles of 30 000 small-molecule treatments using the Cell Painting assay. Data, https://github.com/gigascience/paper-bray2017
- Pascual-Vargas et al. (2017), RNAi screens for Rho GTPase regulators of cell shape and YAP/TAZ localisation in triple negative breast cancer Data via IDR
- Pizzagalli et al (2018), Leukocyte Tracking Database, a collection of immune cell tracks from intravital 2-photon microscopy videos (via figshare)

Specific projects – switch to confidential presentation ;-)