Data science in cell imaging
Lecture 4: deep learning in microscopy

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

Our perspective on data science in cell imaging is out – read it!

© 2021. Published by The Company of Biologists Ltd | Journal of Cell Science (2021) 134, jcs254292. doi:10.1242/jcs.254292

The Company of Biologists

PERSPECTIVE

SUBJECT COLLECTION: IMAGING

Data science in cell imaging

Meghan K. Driscoll[1,*] and Assaf Zaritsky[2,*]

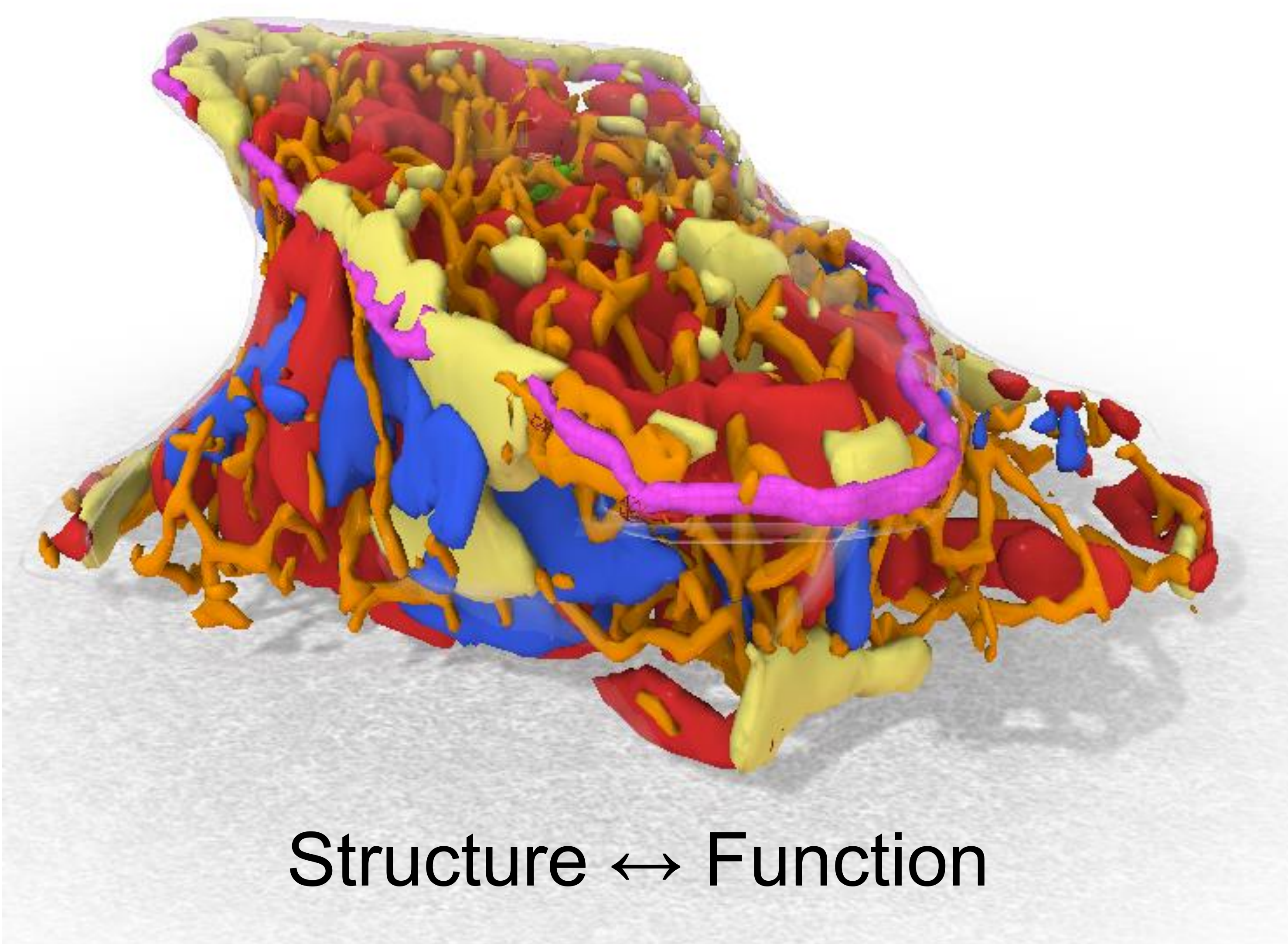
https://jcs.biologists.org/content/134/7/jcs254292

Access via Moodle (or email me)

Driscoll & Zaritsky (2021)

Look at a cell and know what it is doing

What it did



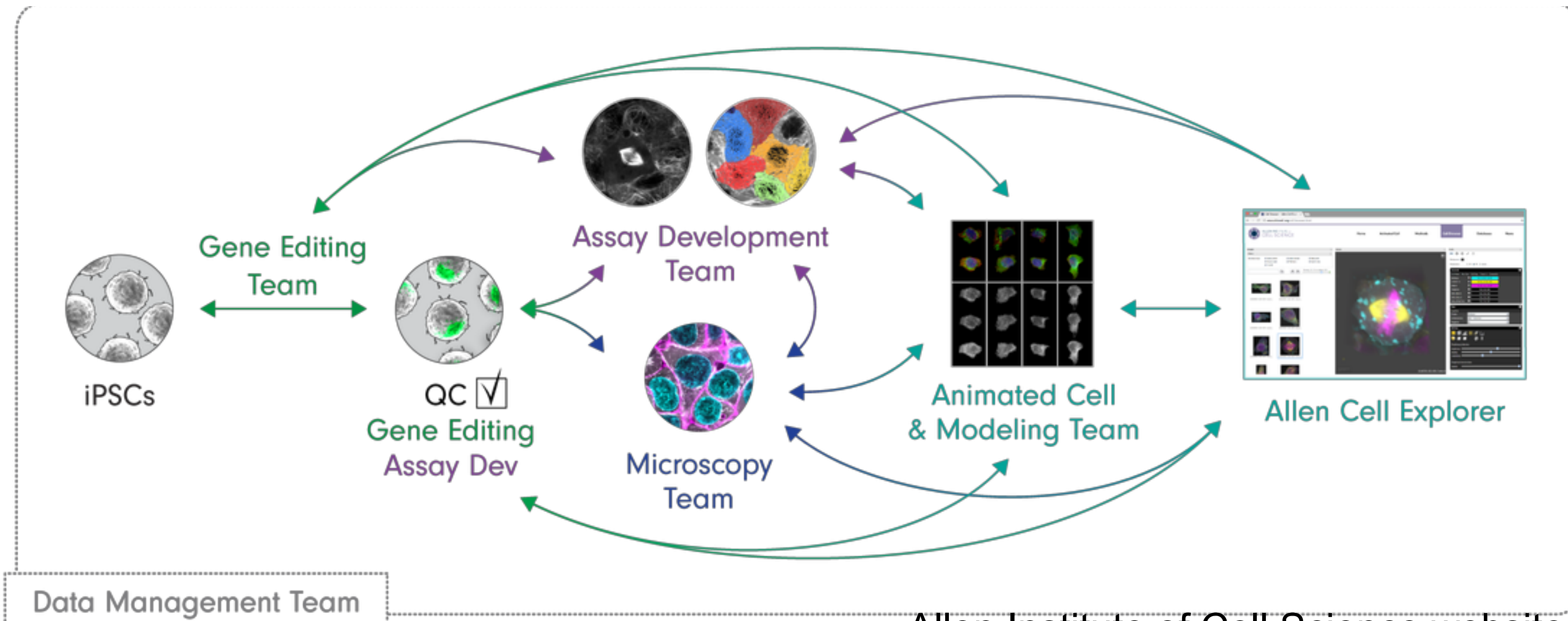
What it will do

Slide adapted from Susanne Rafelski, Allen Institute of Cell Science

Allen institute of cell science

Our Mission

The mission of the Allen Institute for Cell Science is to create dynamic and multi-scale visual models of cell organization, dynamics and activities that capture experimental observation, theory and prediction to understand and predict cellular behavior in its normal, regenerative, and pathological contexts.



Allen Institute of Cell Science website

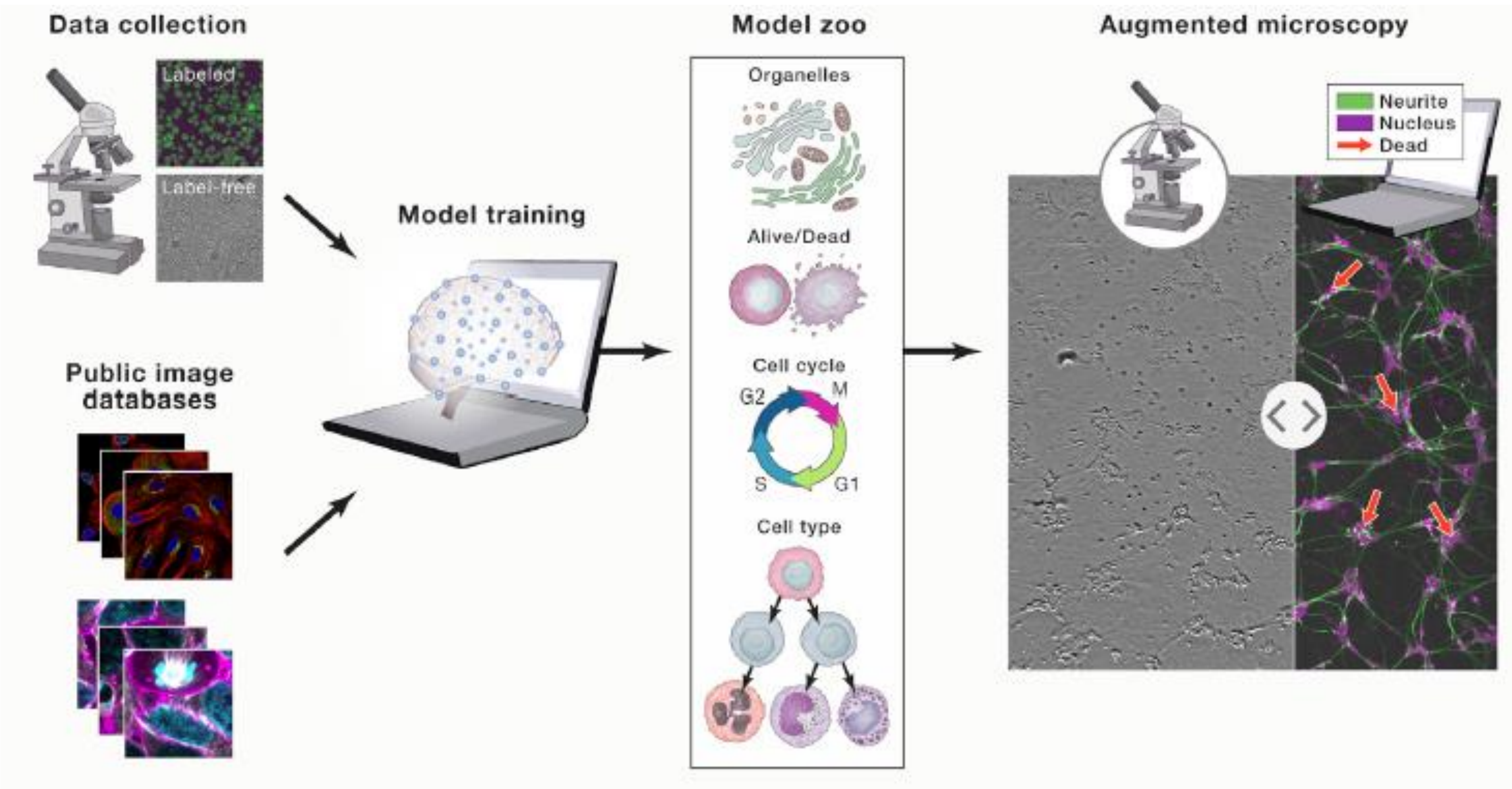
Allen institute of cell science

- Overview https://www.allencell.org/what-we-do.html
- Visualization https://www.allencell.org/visual-guide-to-human-cells.html
- 3D cell viewer https://www.allencell.org/3d-cell-viewer.html
- Cell feature explorer https://bit.ly/355Lq1j
- Publicly available cell lines, tools, data, code!
- Research projects https://www.allencell.org/cell-research-projects.html

The holy grail!

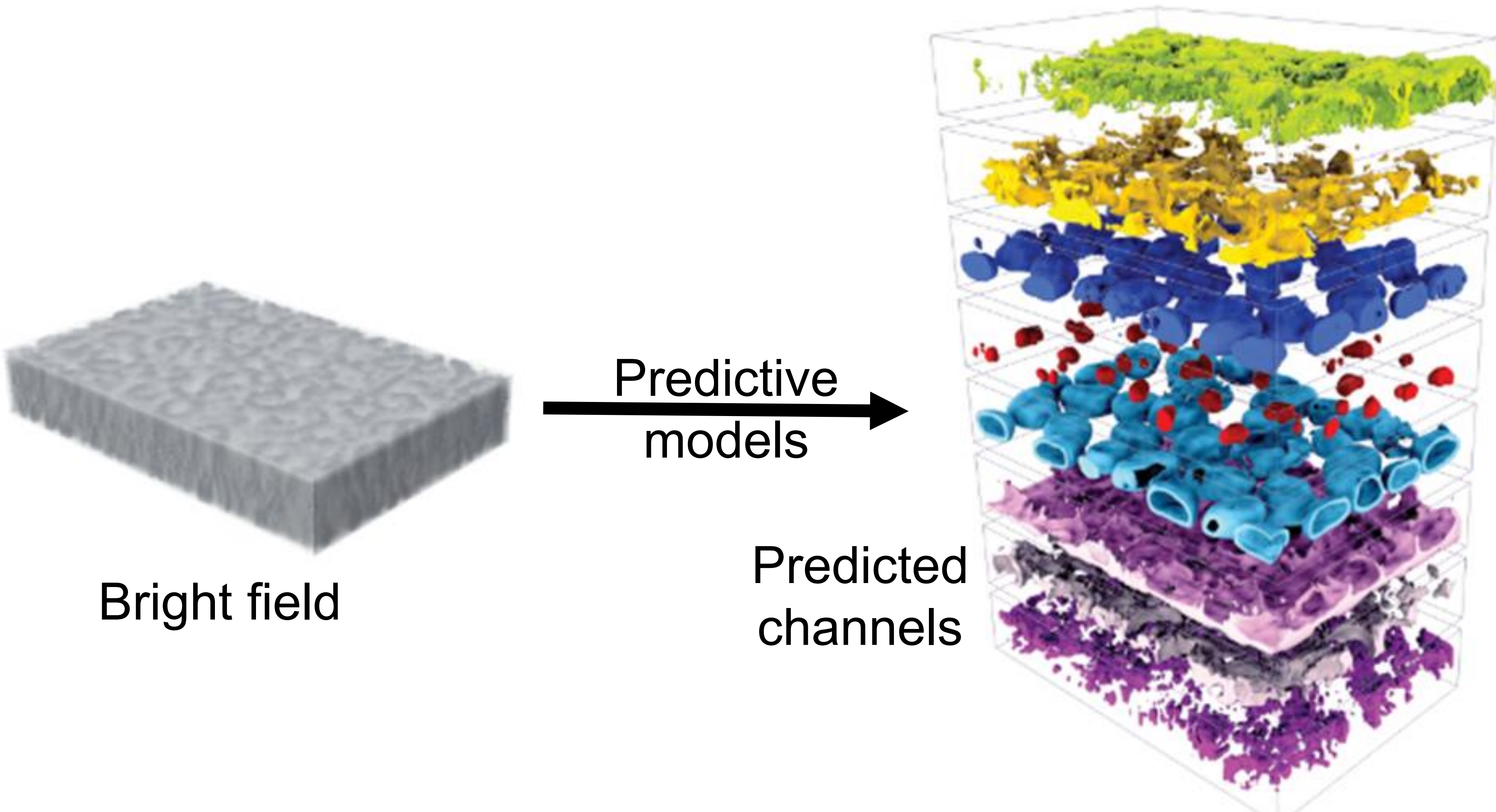
Can we train a generative model for accurate fluorescent imaging from label-free (transmitted light) imaging?

A future of augmented microscopy



Sullivan and Lundberg (2018)

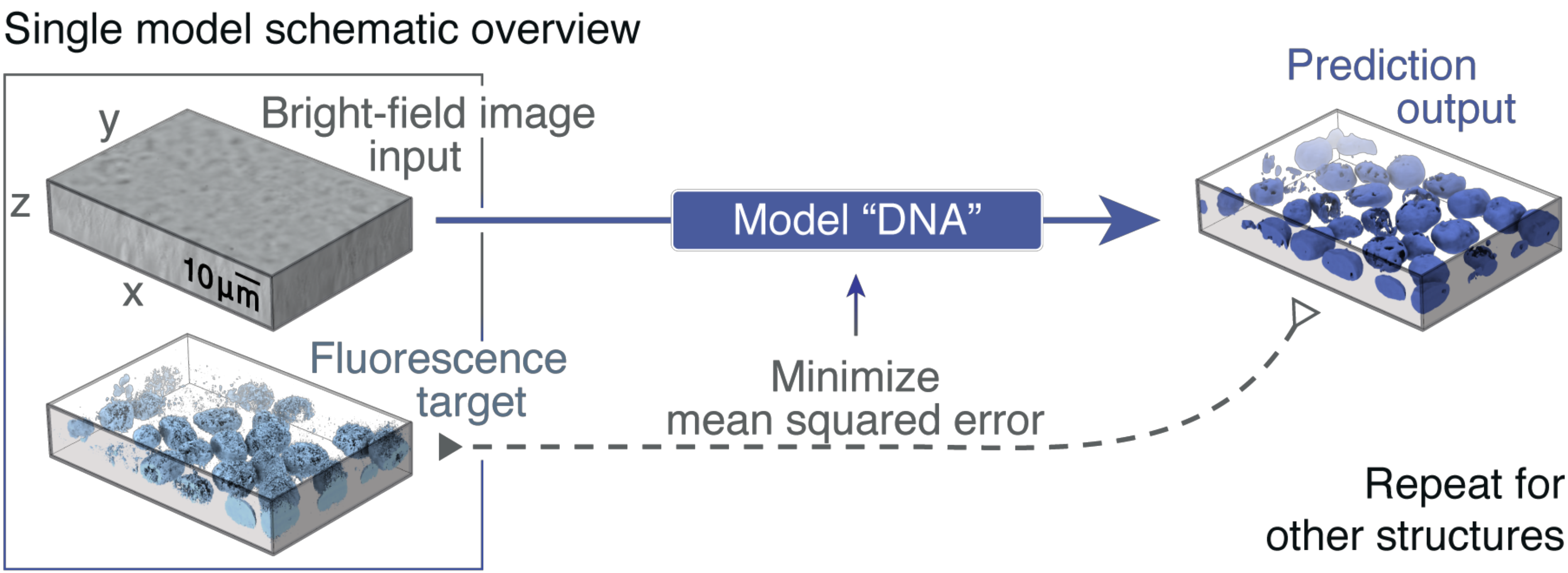
Label-free images contain information on the molecular organization of the cell!



Ounkomol et al. (2018)

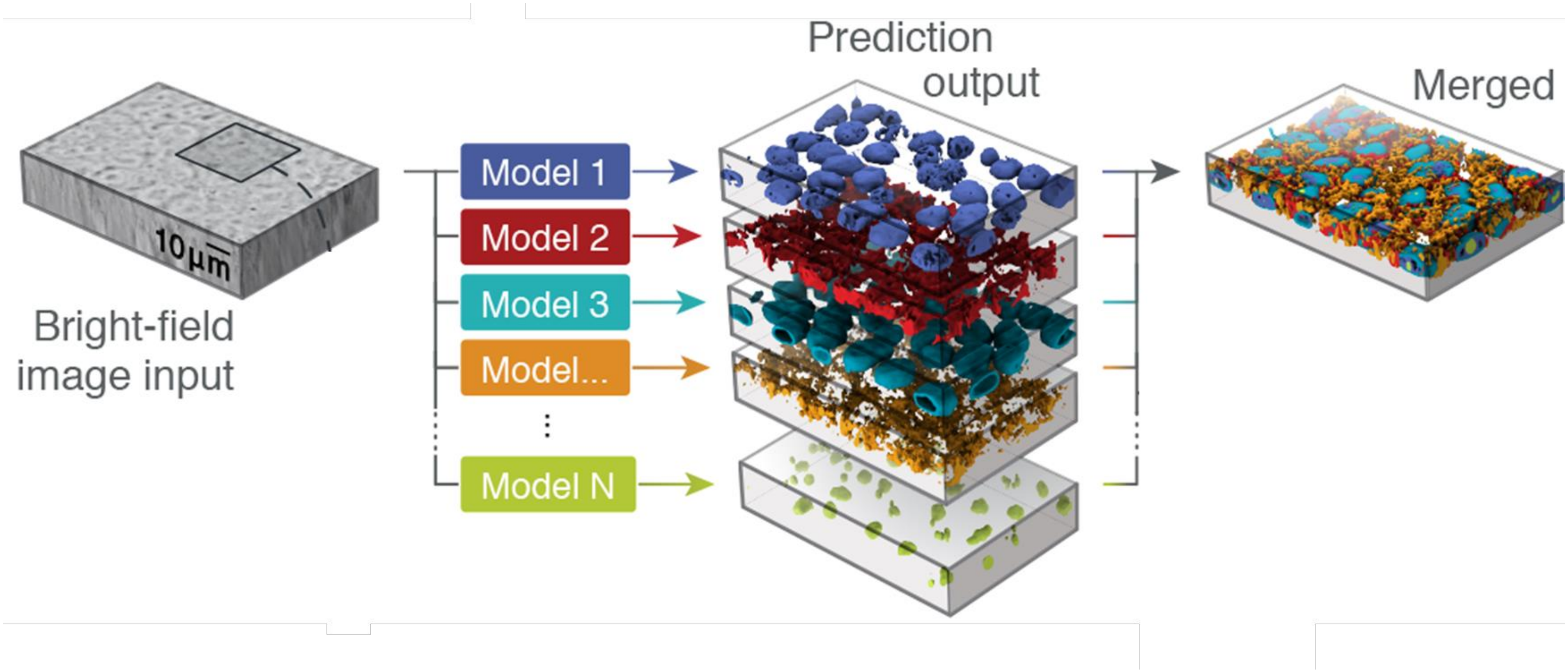
Christiansen et al. (2018)

Unstructured-to-structured information with supervised models



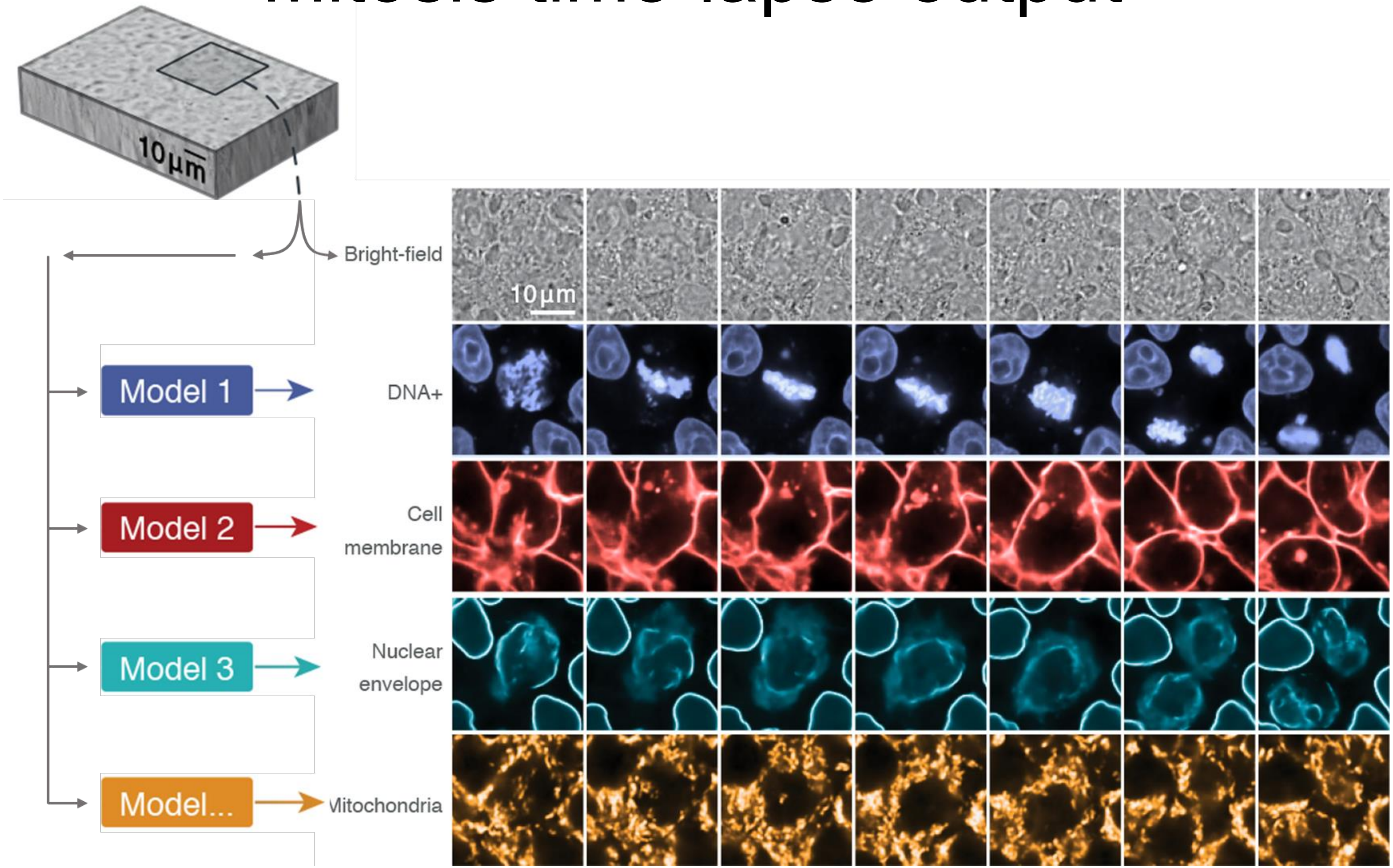
Ounkomol et al. (2018)

Combining multiple models



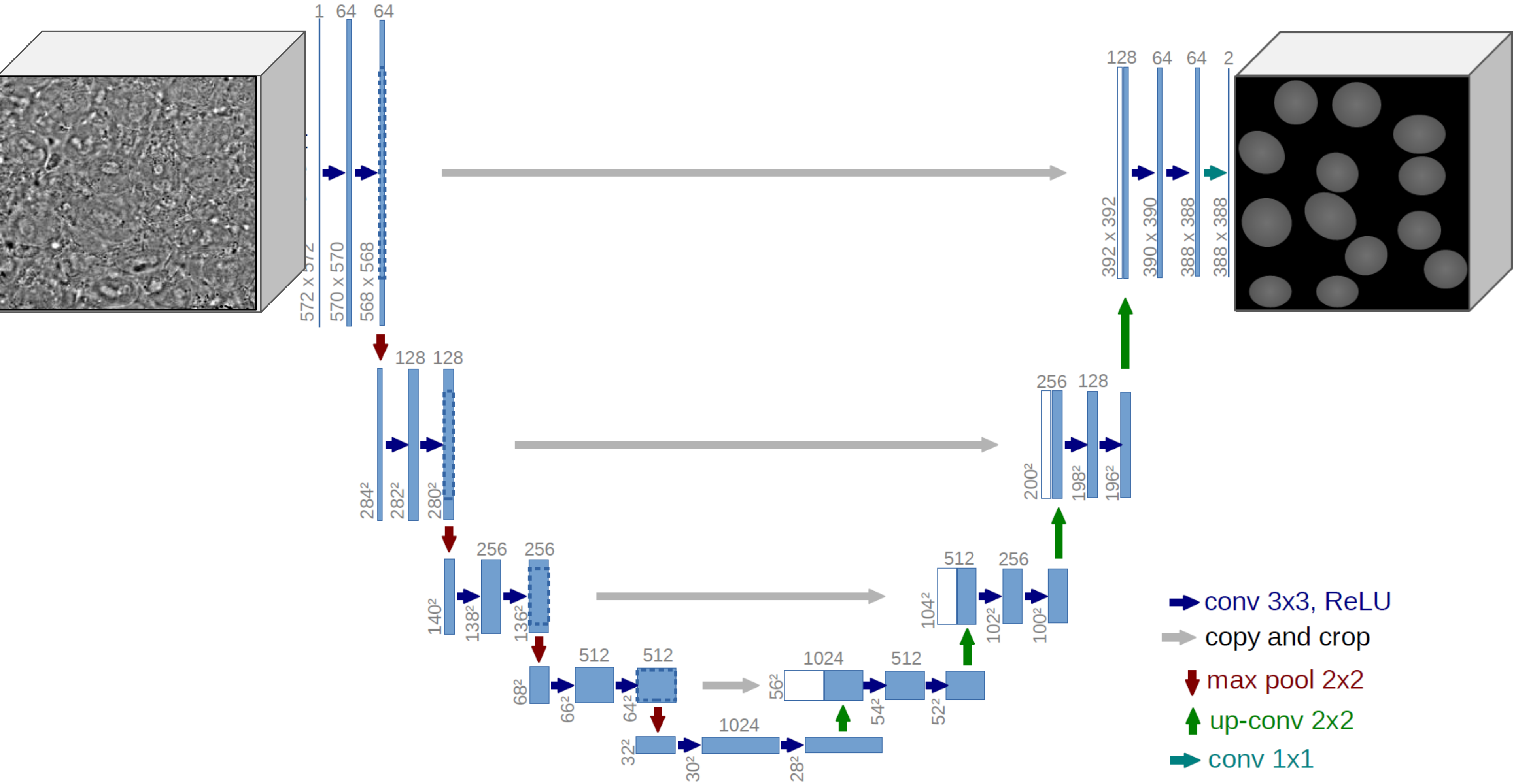
Ounkomol et al. (2018)

Mitosis time-lapse output



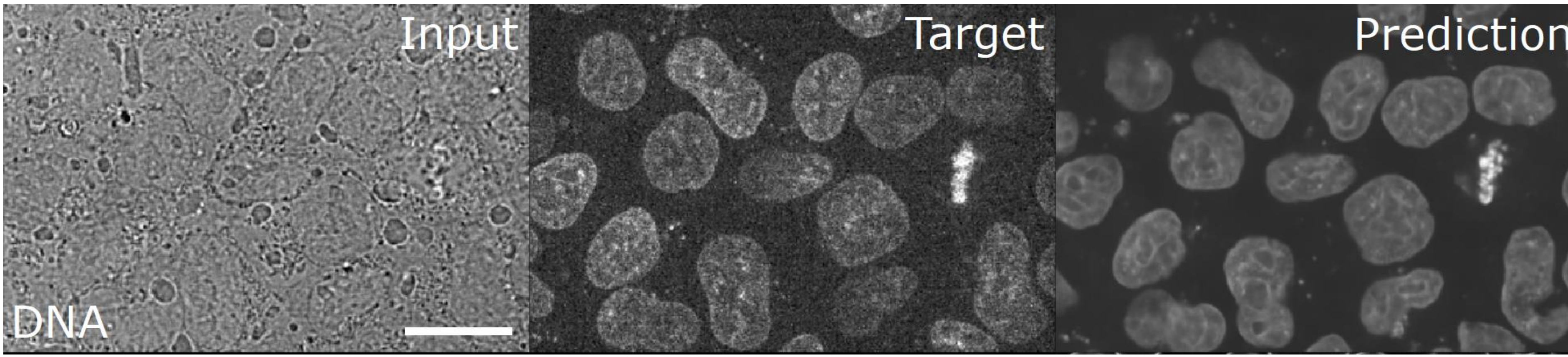
Ounkomol et al. (2018)

U-Net architecture



Greg Johnson, Ronneberger et al. (2015)

Predicting DNA localization from transmitted light

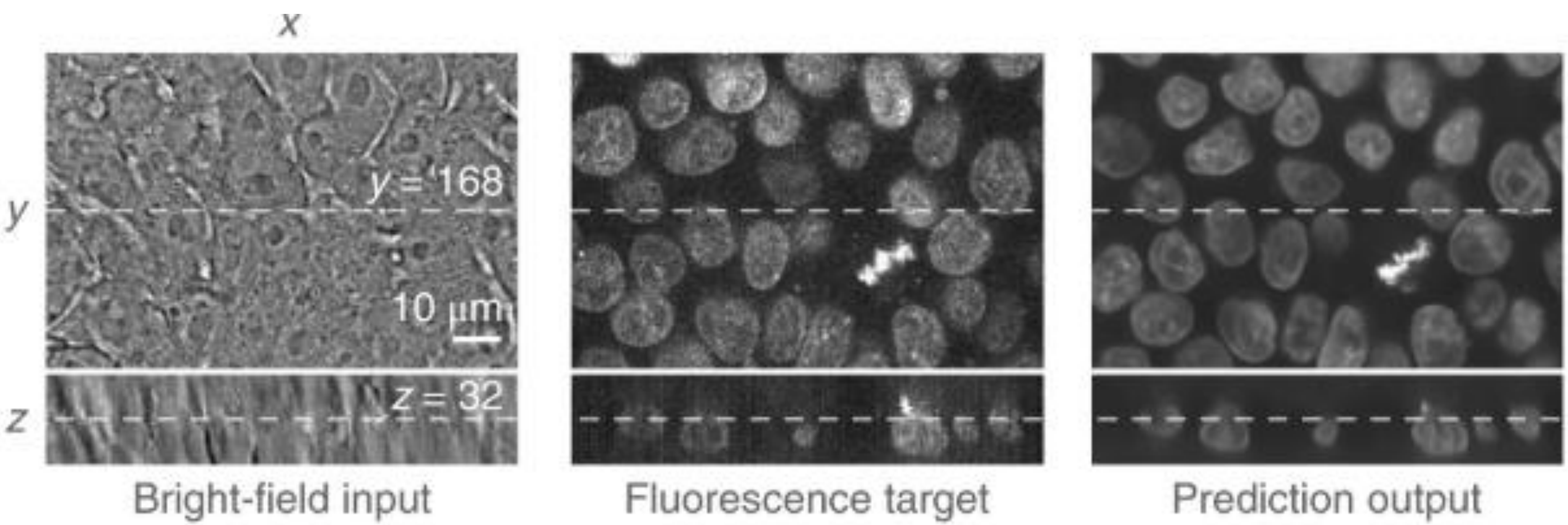


Good pixel-wise correlation between (3D) "true" and predictions. Less noise in predictions.

(assessing and improving similarity/distance measures could be a course project)

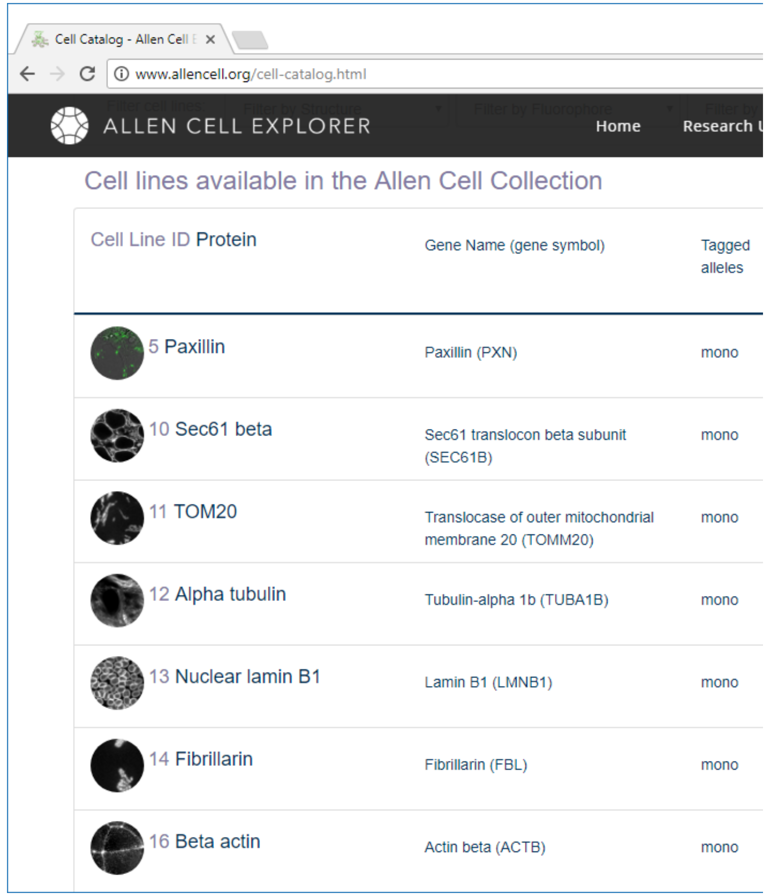
Greg Johnson

Predictions are 3D



Greg Johnson

What about other structures?



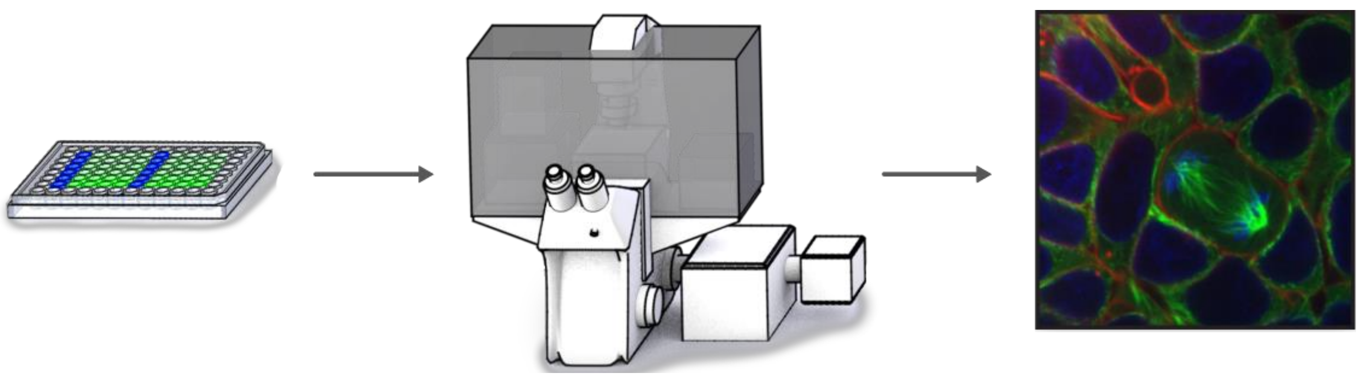
Cell Catalog - Allen Cell E

www.allencell.org/cell-catalog.html

ALLEN CELL EXPLORER Home Research

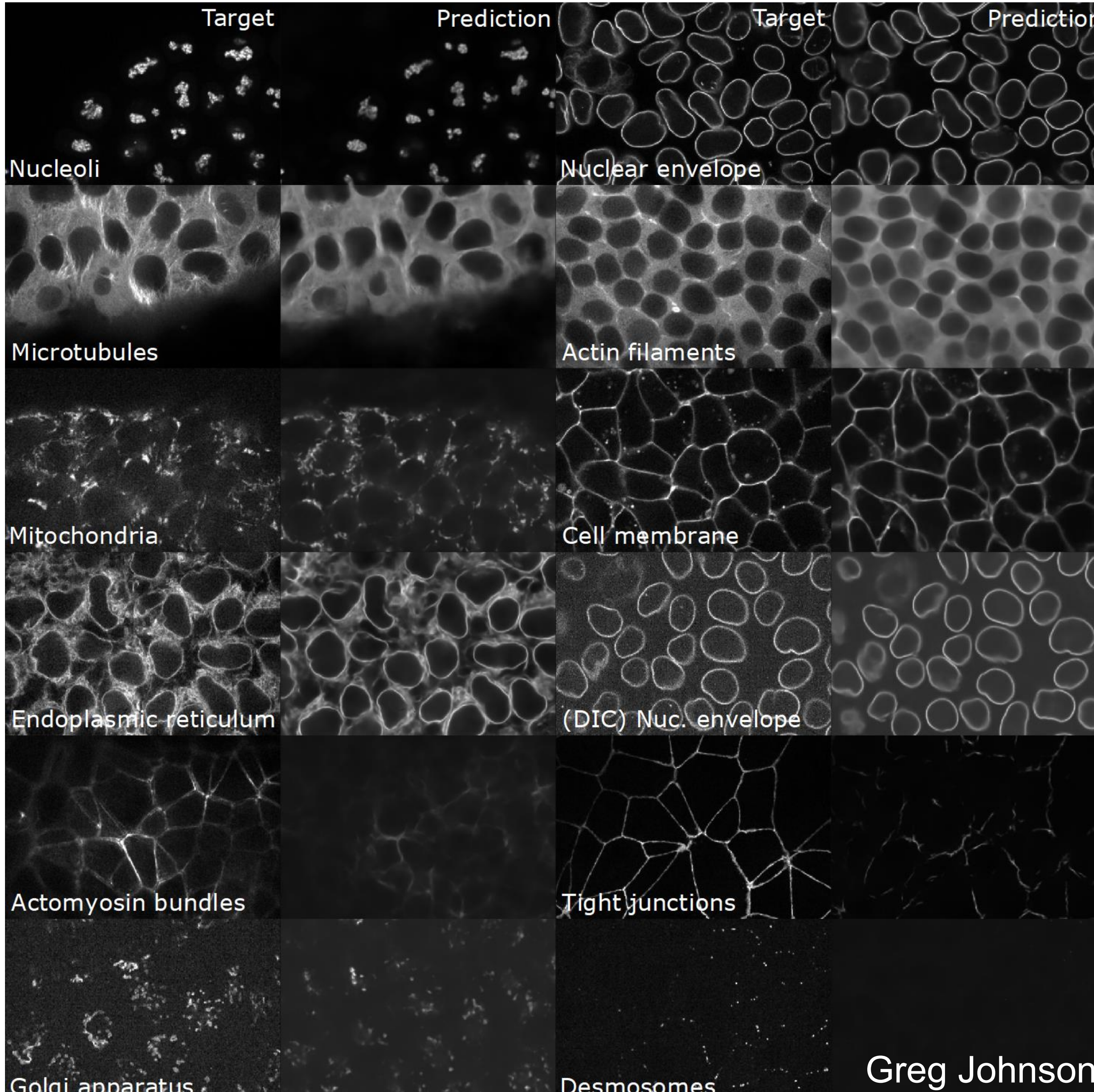
Cell lines available in the Allen Cell Collection

Cell Line ID Protein	Gene Name (gene symbol)	Tagged alleles
5 Paxillin	Paxillin (PXN)	mono
10 Sec61 beta	Sec61 translocon beta subunit (SEC61B)	mono
11 TOM20	Translocase of outer mitochondrial membrane 20 (TOMM20)	mono
12 Alpha tubulin	Tubulin-alpha 1b (TUBA1B)	mono
13 Nuclear lamin B1	Lamin B1 (LMNB1)	mono
14 Fibrillarin	Fibrillarin (FBL)	mono
16 Beta actin	Actin beta (ACTB)	mono



Greg Johnson

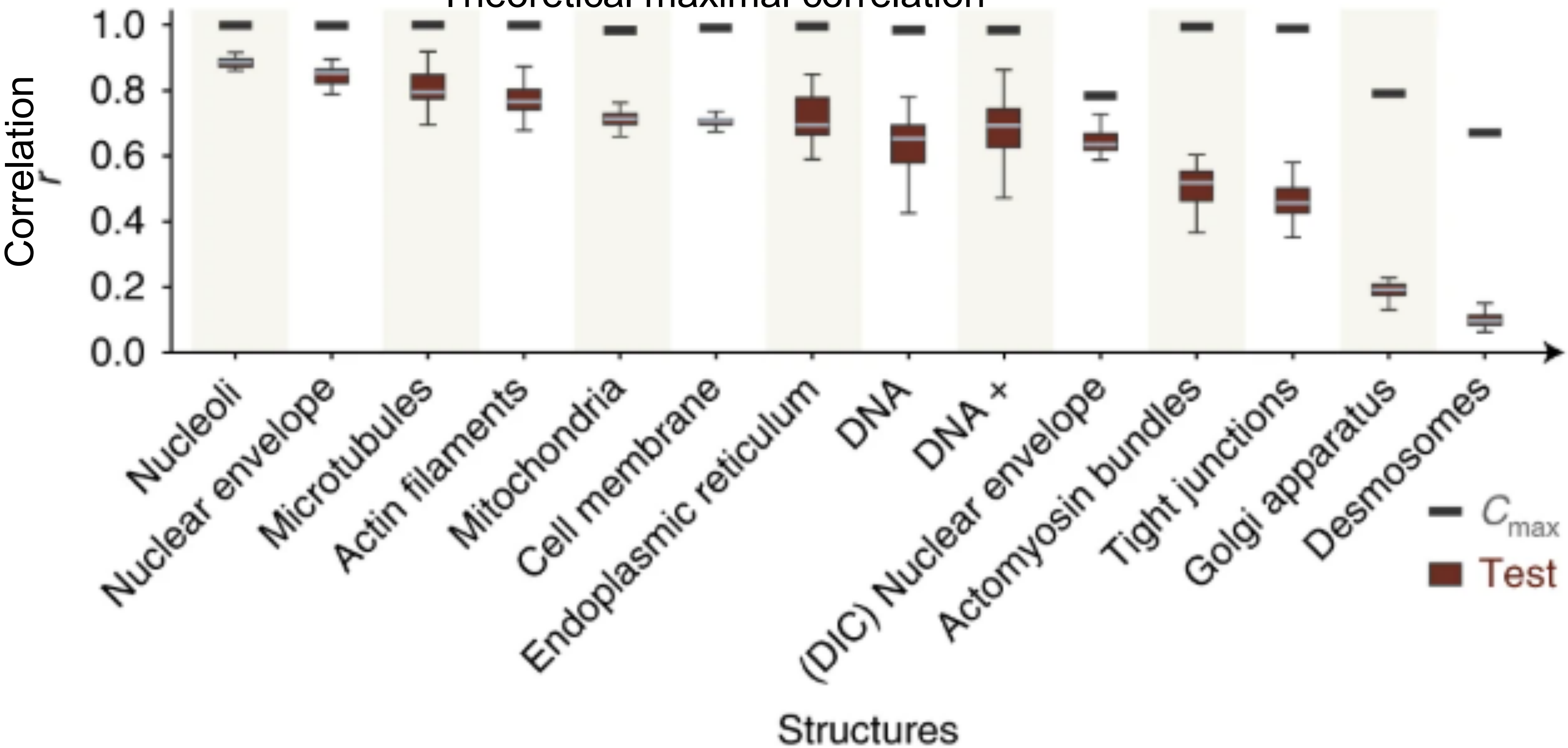
Results



Greg Johnson

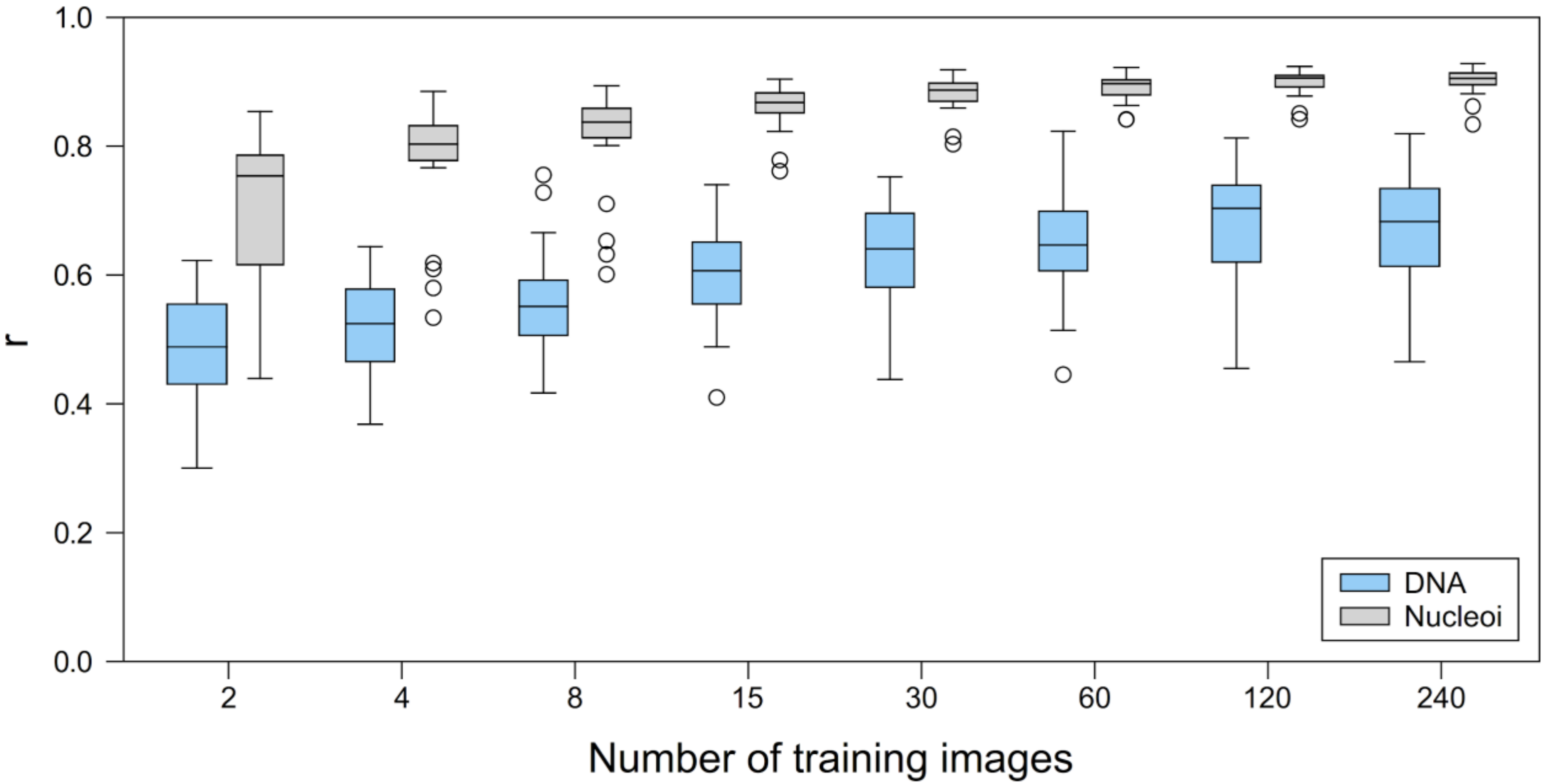
Predictions performance

Theoretical maximal correlation



Greg Johnson

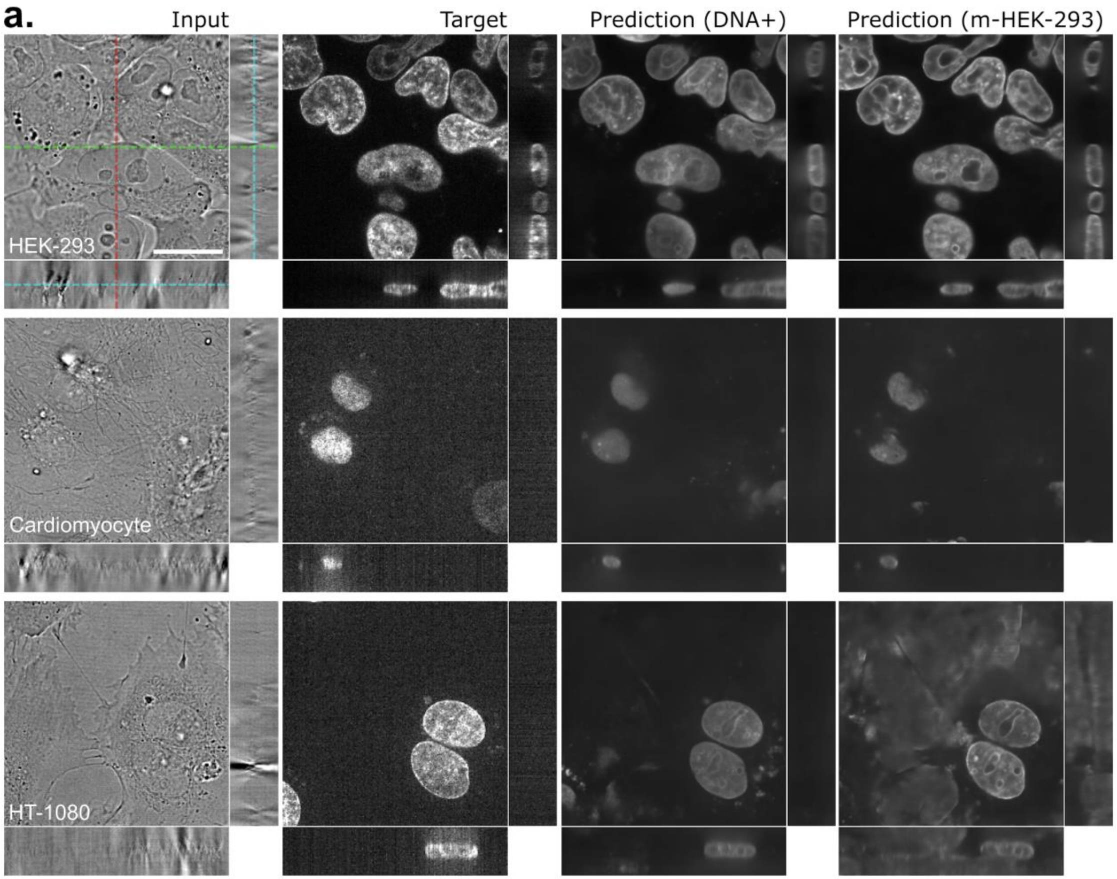
Models can be trained with fairly small datasets

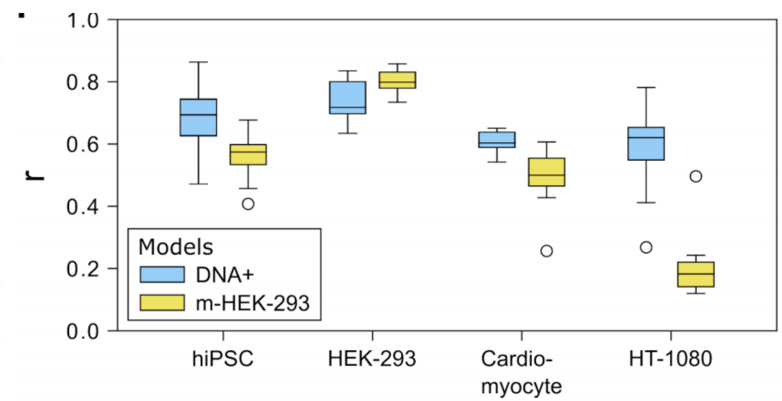


Greg Johnson

Generalization to other cell types

Always perform best for data similar to the training





Greg Johnson

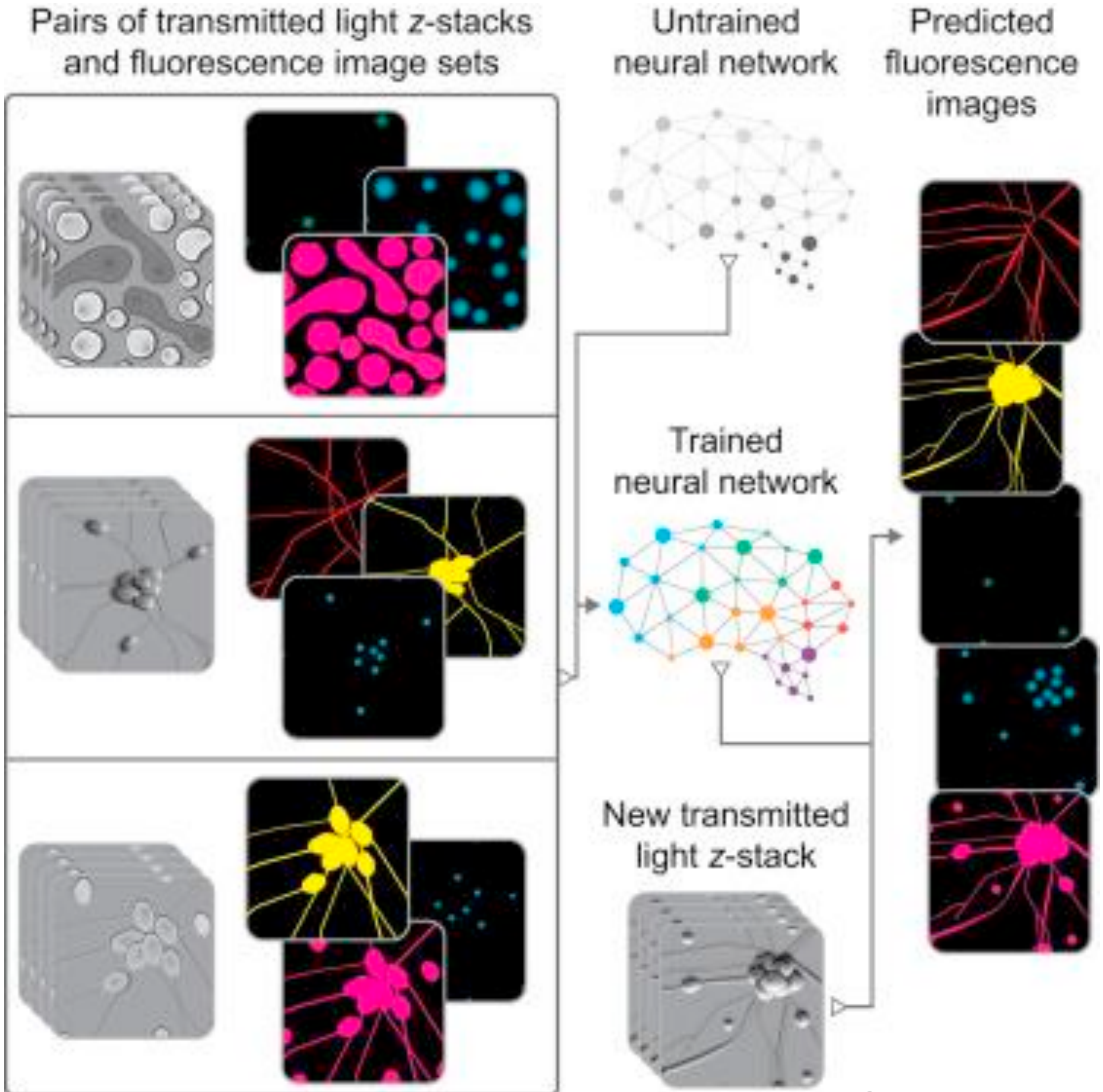
Summary

- Great for hypothesis generation!
- Be very careful regarding “hallucinated” pixel intensities
- Generalization? effects of perturbations / different cell systems / imaging

In silico labeling: predicting fluorescent labels in unlabeled images

- Fluorescence microscopy images can be predicted from transmitted-light z stacks
  - Cell nuclei, live/dead, cell type, organelle type
- 7 fluorescent labels were validated across three labs, modalities, and cell types
- Multi-task learning
- Transfer learning: new labels can be predicted using minimal additional training data

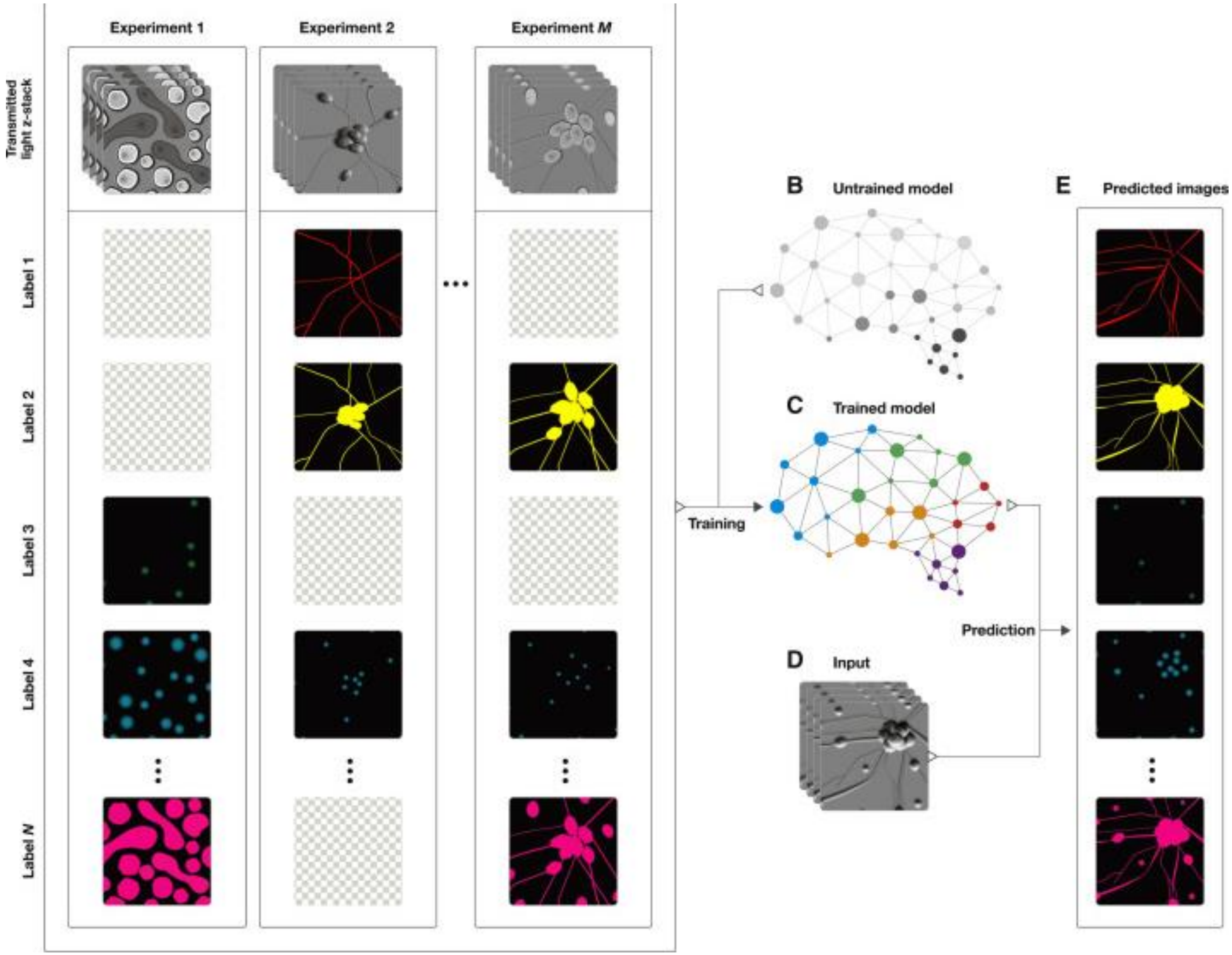
Christiansen et al. (2018)



Christiansen et al. (2018)

Multi-task learning

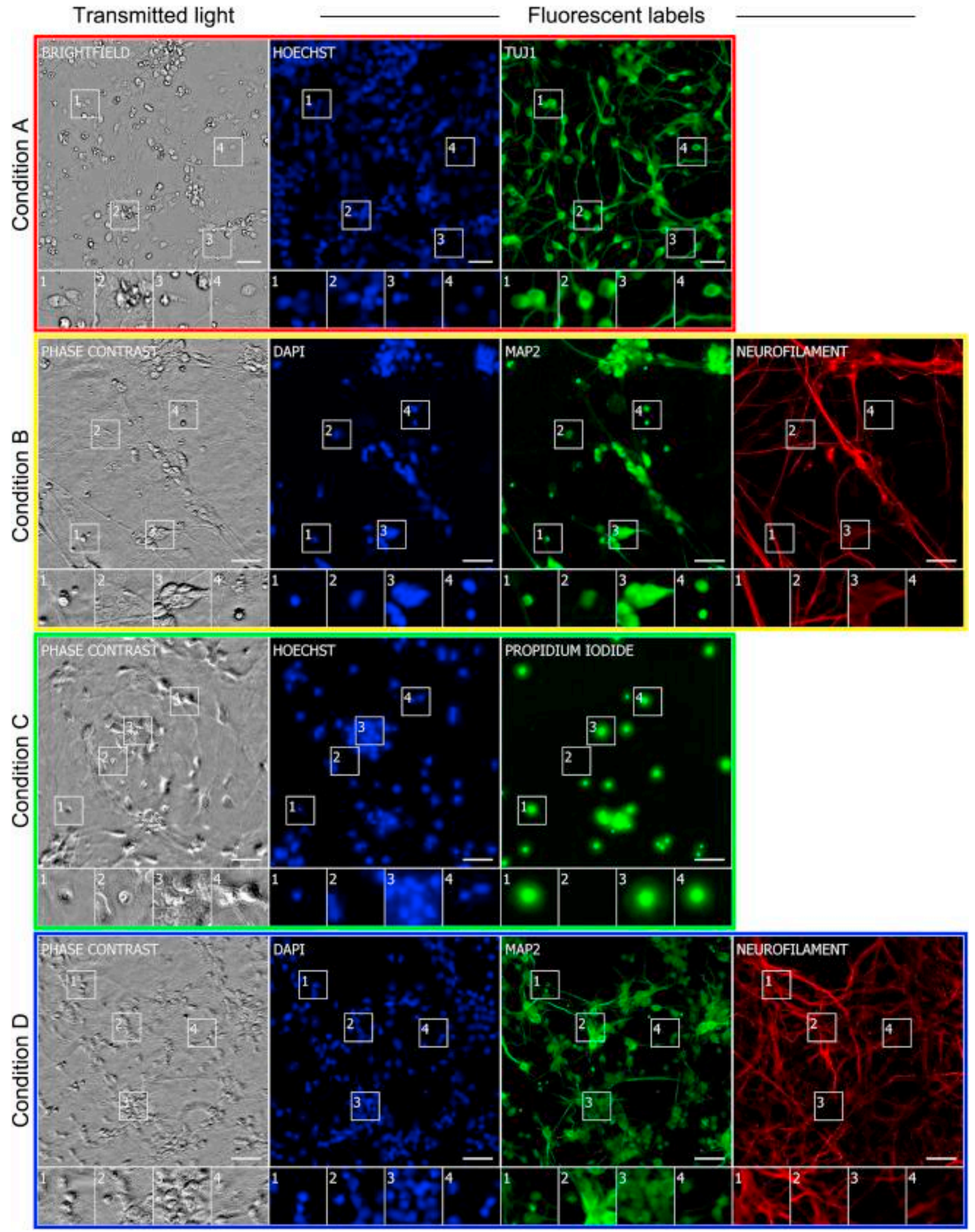
Learned abstractions can be reused across multiple tasks



Different cell types, labels, labs!

Christiansen et al. (2018)

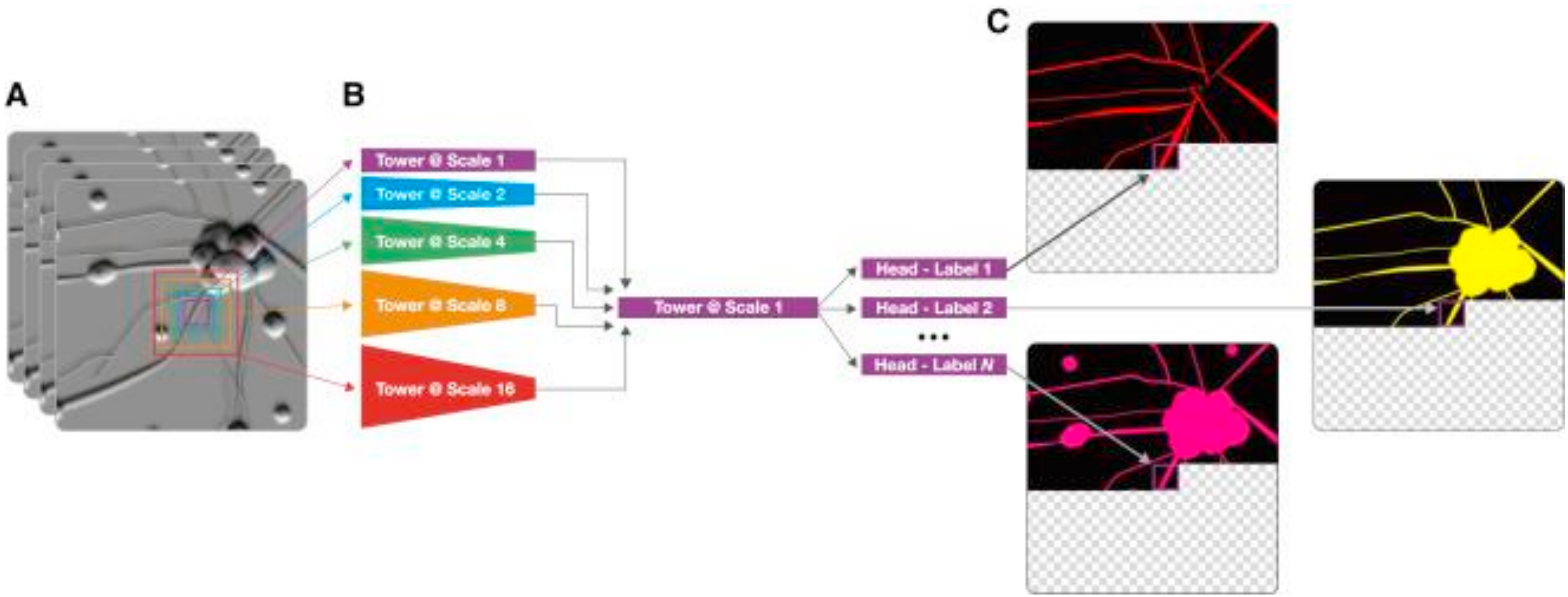
Training data



Christiansen et al. (2018)

Multi-scale learning

learning the spatial interpolation

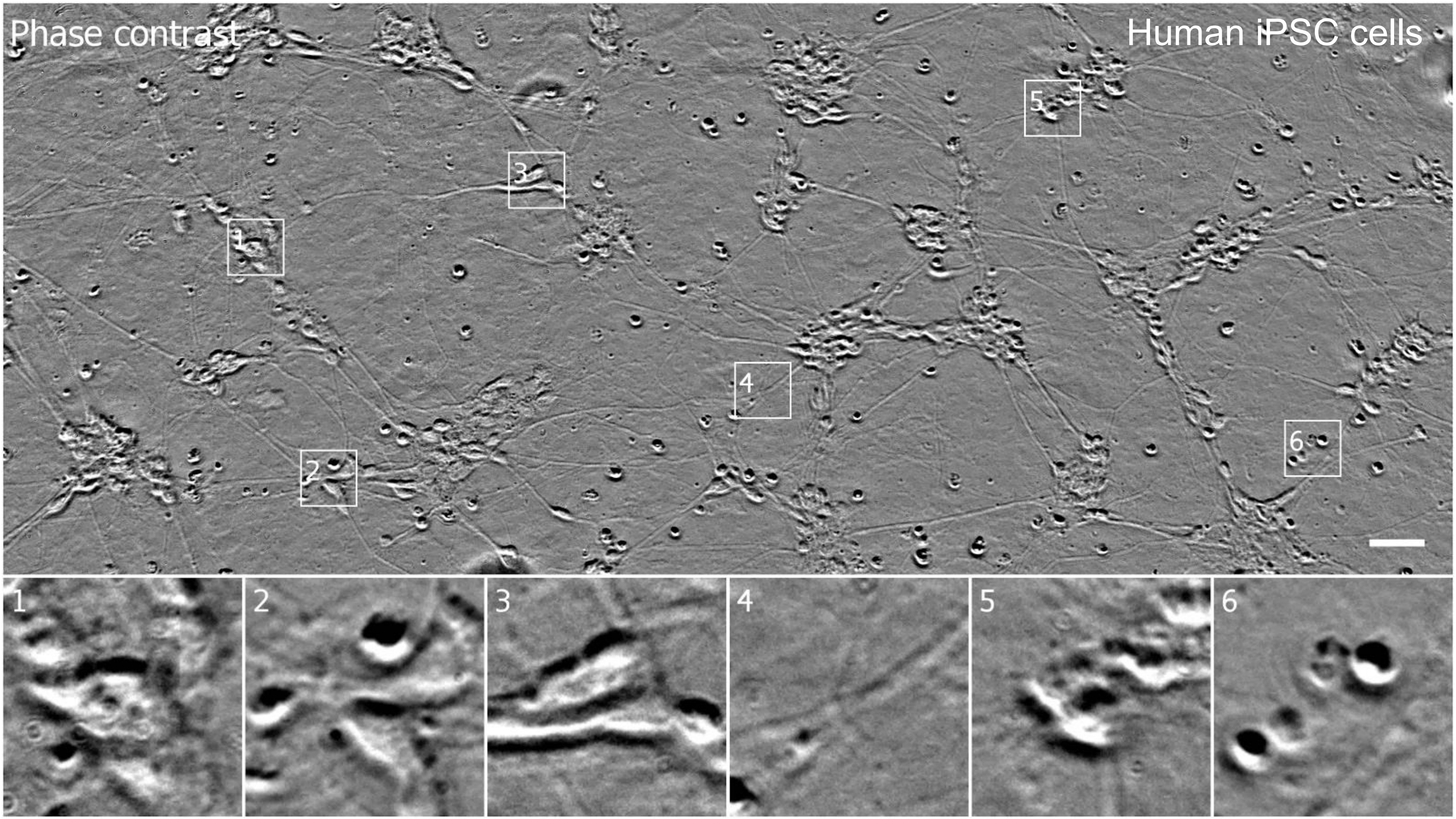


Inception modules (Szegedy et al., 2015) optimized with Google Hypertune (Golovin et al., 2017) to design the network architecture

Blogposts with simple explanations of inception networks here and here

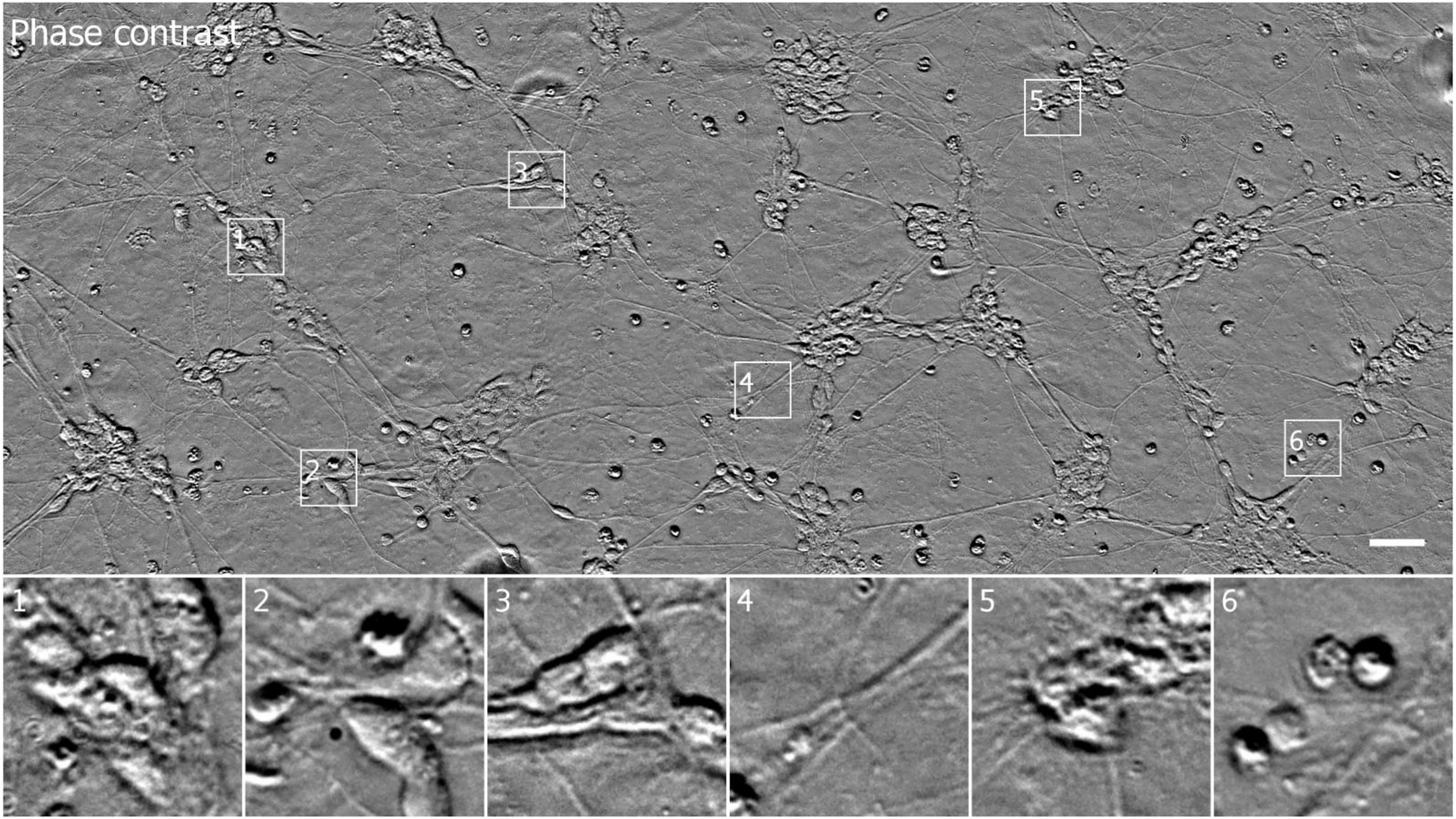
Christiansen et al. (2018)

Input: phase contrast imaging



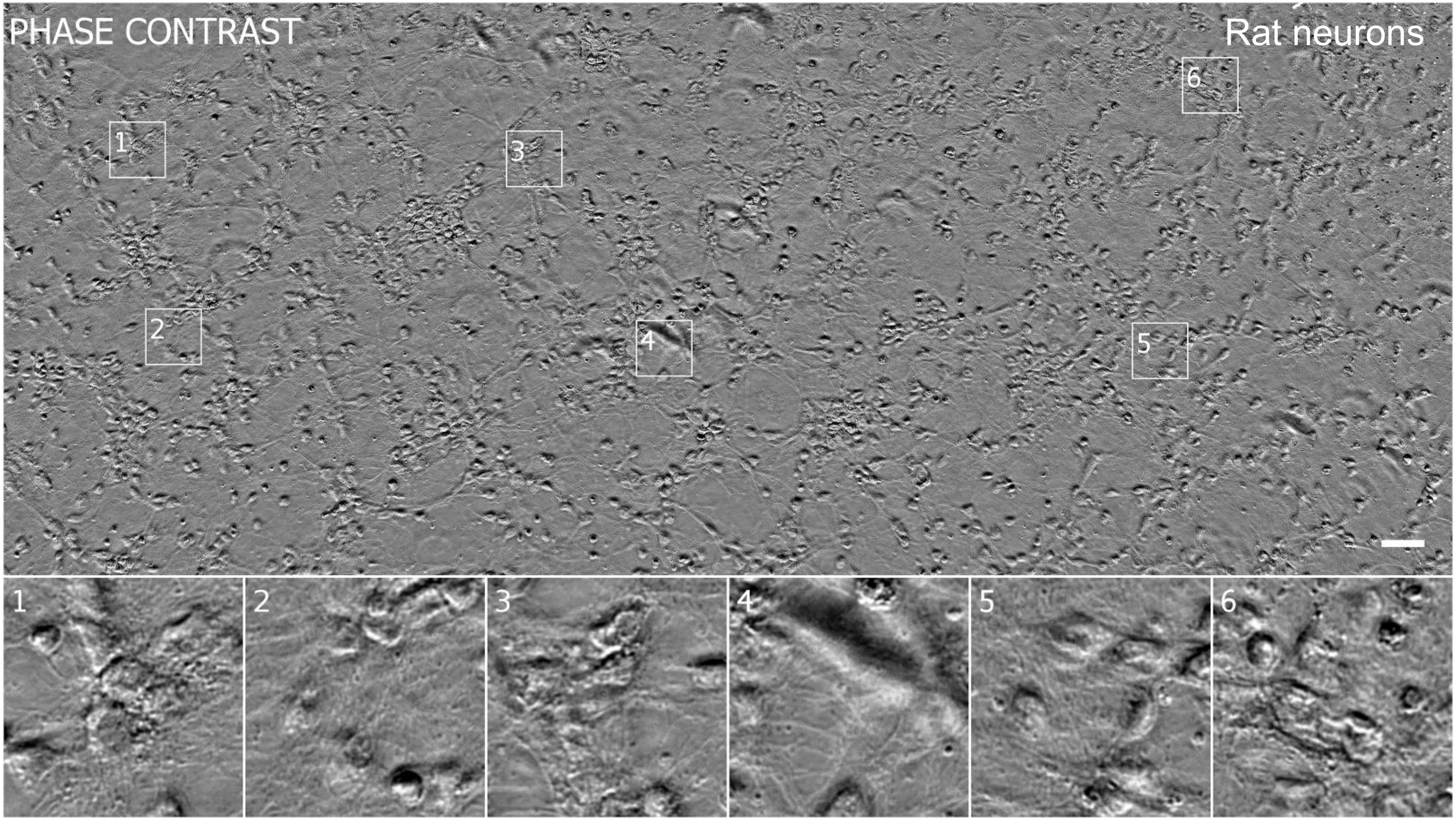
Eric Christiansen

Predictions



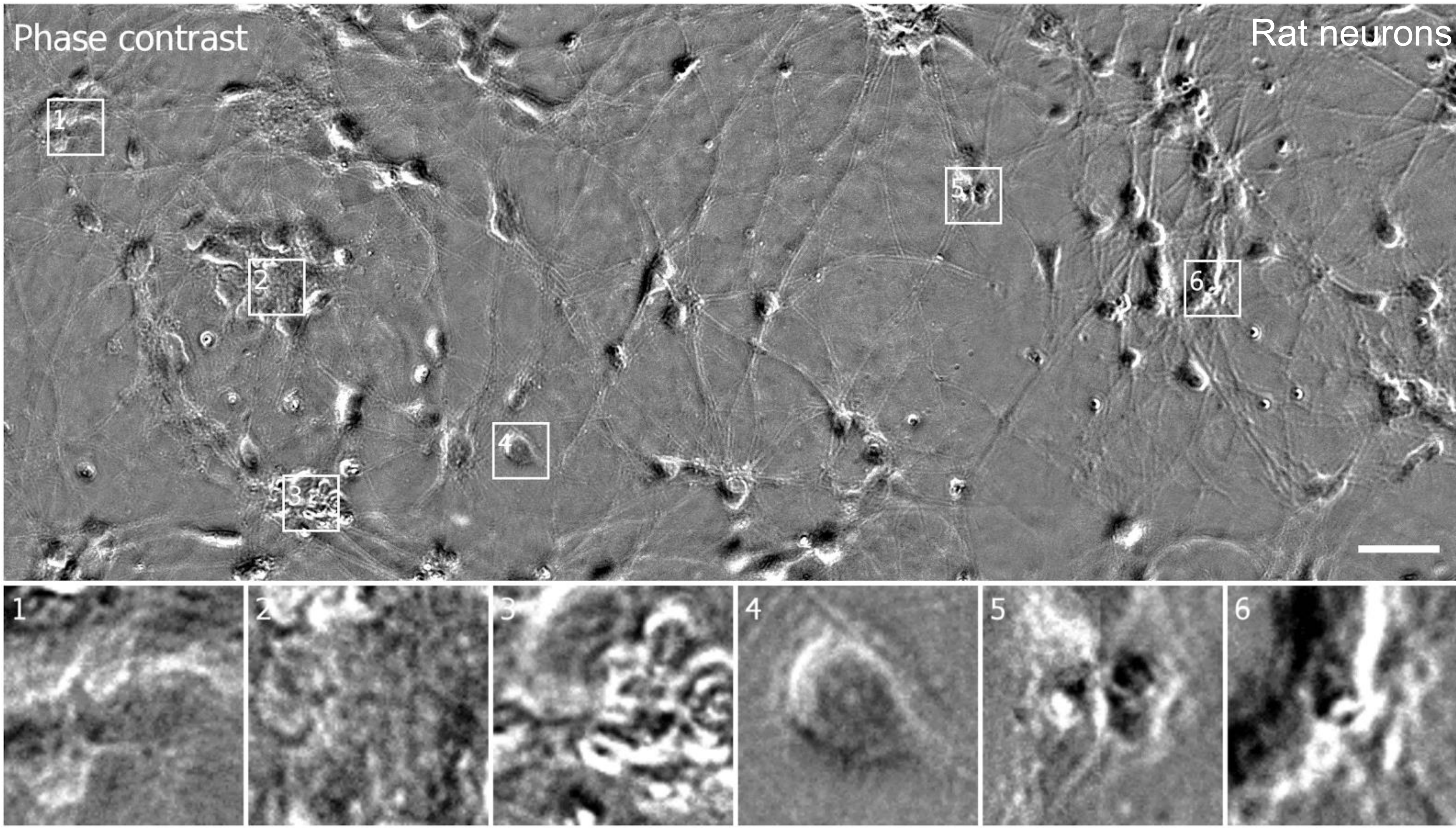
Eric Christiansen

Predictions



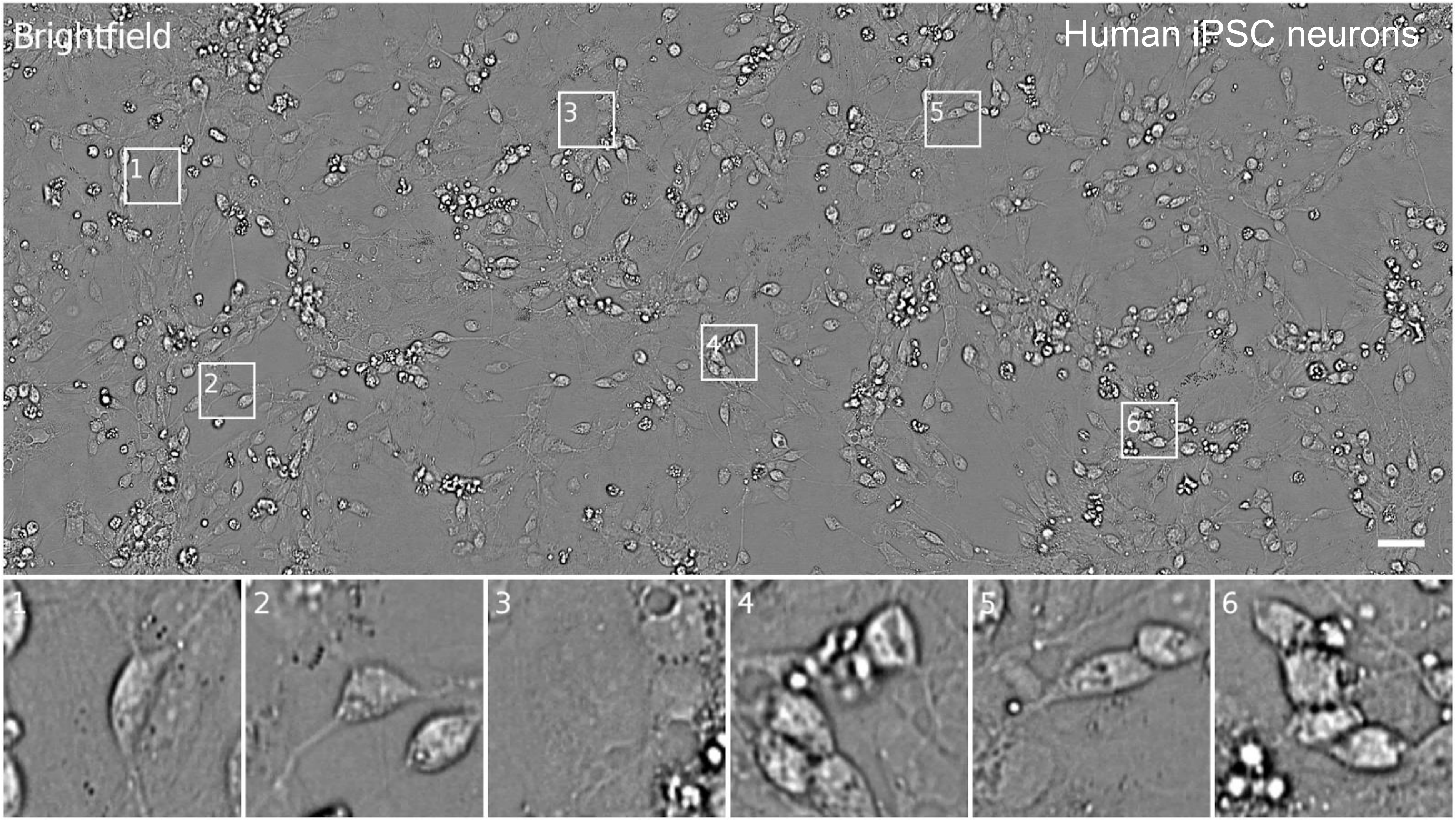
Eric Christiansen

Predictions



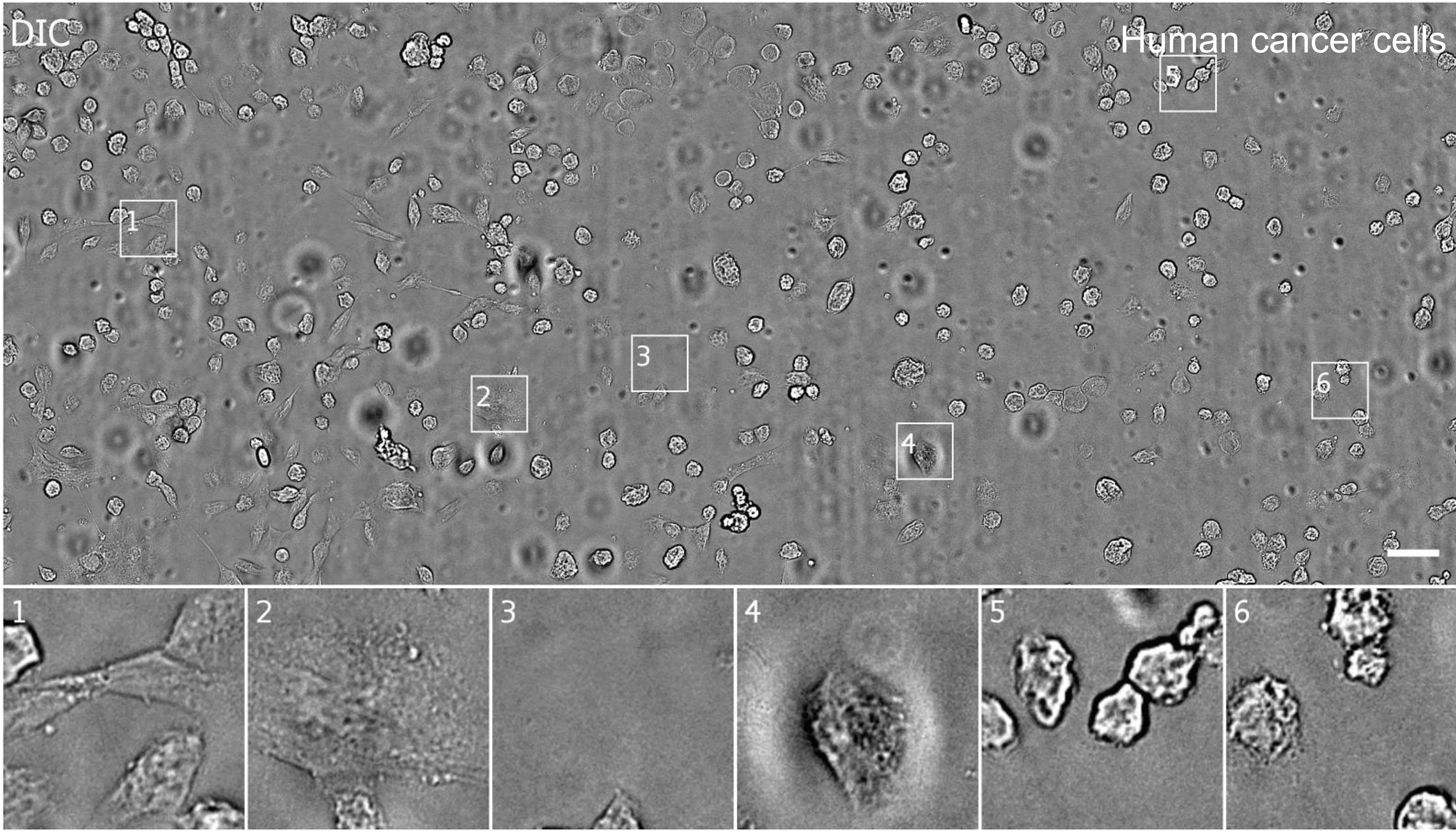
Eric Christiansen

Predictions



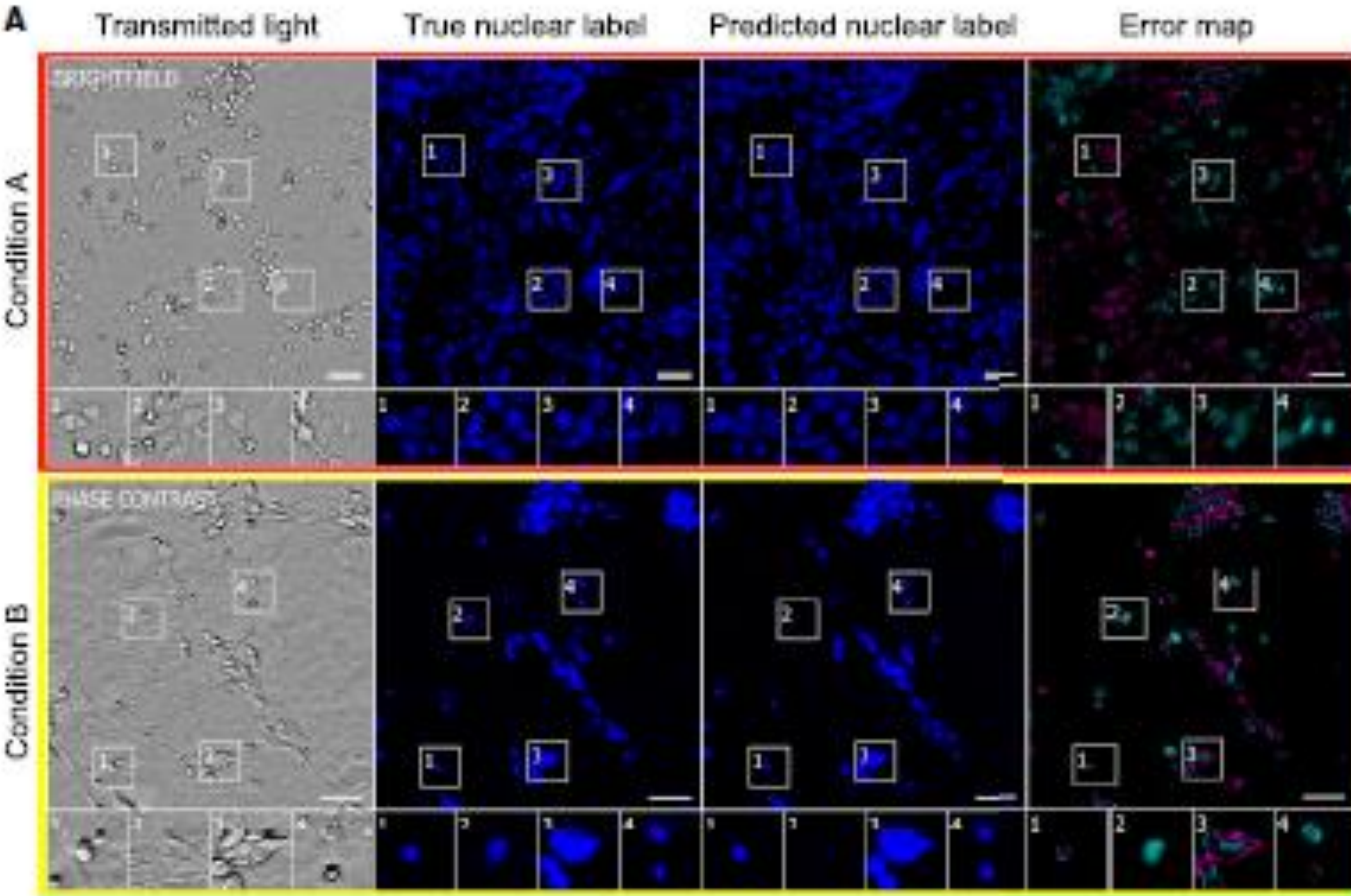
Eric Christiansen

Predictions



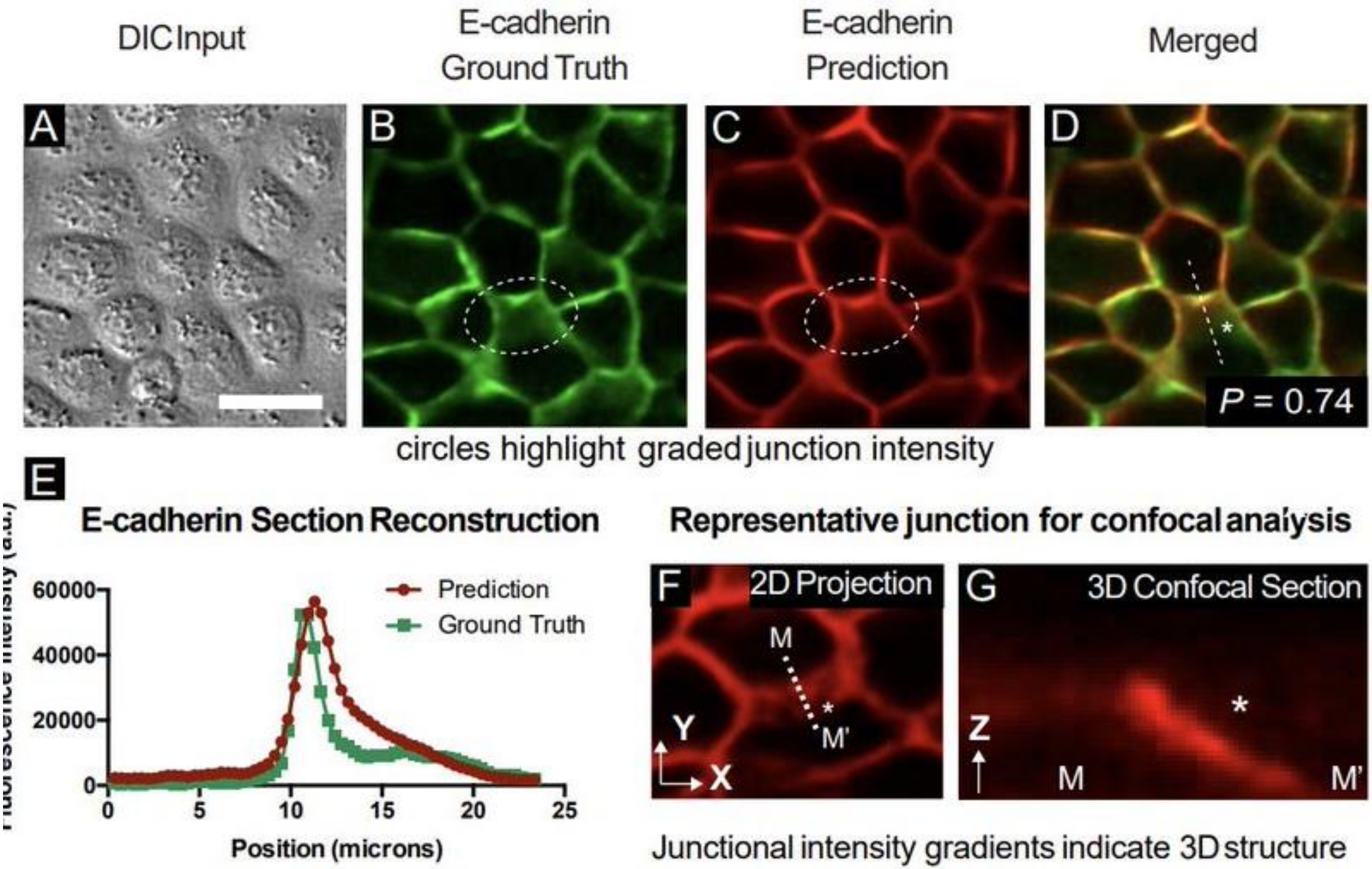
Eric Christiansen

Predicting nuclei (DAPI/Hoechst)



Christiansen et al. (2018)

Validations of downstream analysis



Could be picked as a student presentation

LaChance and Cohen (2020)

Examples of cross modality image mapping with deep learning

- PhaseStain: phase-to-histology (Rivenson, Liu, Wei, et al., 2019)
- Label free to physical cell properties (Guo, Yeh, Folkesson, at al., 2020)
- Dual labeling (Kölln et al. 2020)
- Fiducial markers to protein localization (Kobayashi et al. 2021)

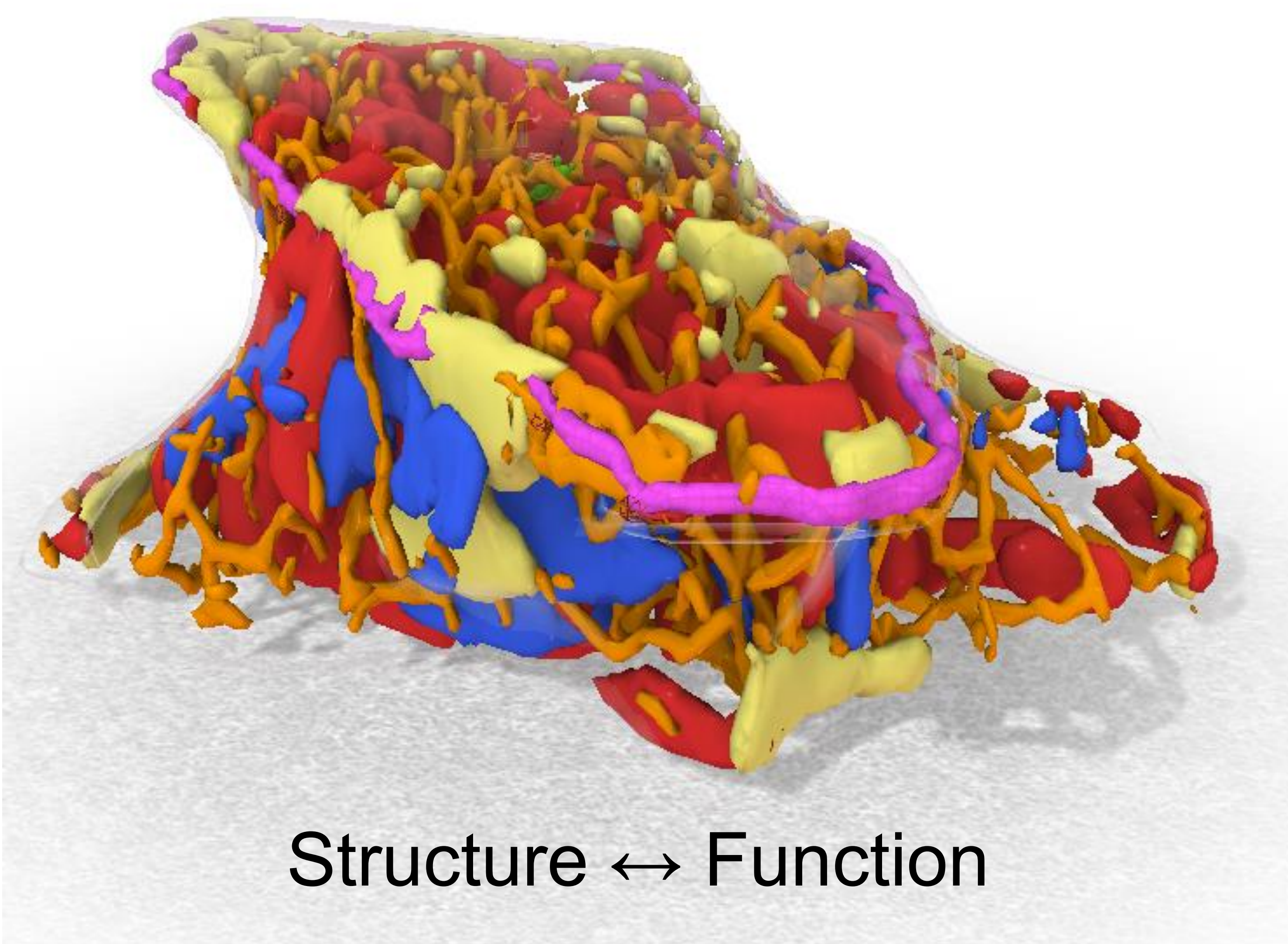
Could be picked as a student presentation

Other (non DL based) generative models for cell organization

- Robert Murphy’s lab, CMU, http://www.andrew.cmu.edu/user/murphy/
- Could be picked as a student/s presentation

Look at a cell and know what it is doing

What it did



What it will do

Slide adapted from Susanne Rafelski, Allen Institute of Cell Science

Pick a paper/s and schedule yourself for class presentation

- 20% of your grade
- In pairs
- From the list I distributed (or ask for my approval)
  - New/ish papers that I think are important/interesting (or ones that I did not have a chance to read)
- 10 minutes presentation + 5 discussion, strict timing!
- Iterative preparation: Yishaia (the TA) → me
  - Grade will be determined based on the full process.
- Public (ppt) slides on the course webpage with your name on it (unless you disagree)

Prepare yourself…

- 20% of your grade
- These papers are LONG (compared to CS papers) and include cryptic biological terms
- Focus on the important stuff in the course’s context: idea, methodology, results, impact, limitations
- Reading tip: you can ignore many biological/experimental details
  - Example: specific molecule names. Very important, but less in the context of our course

Suggested presentation structure

- Background and context in relation to what we learn in class
- Main contribution and results
- Highlight specific results / ideas you find interesting
- Critic / weakness / limitations
- Personal opinion