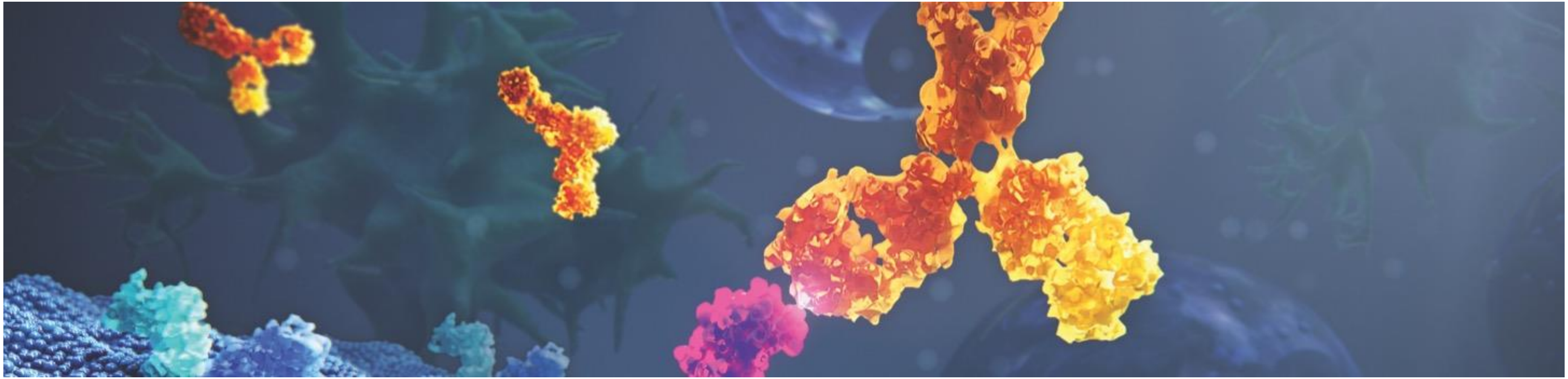
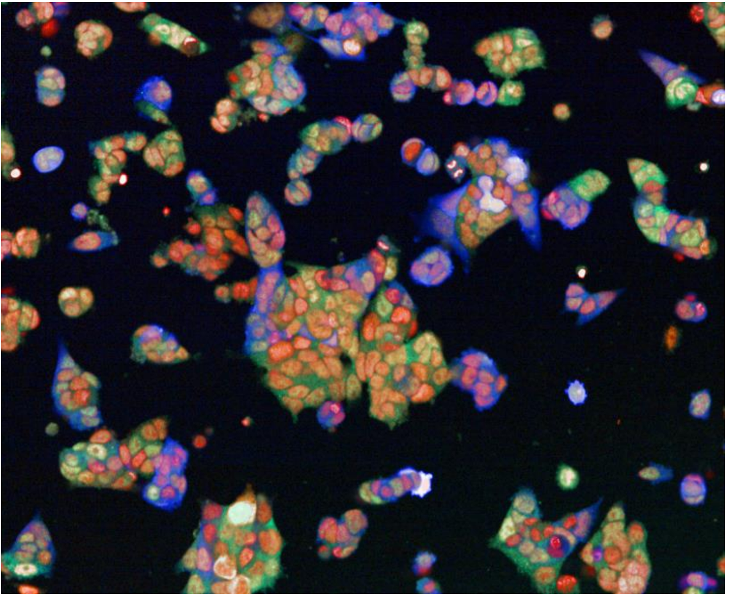
# A glance at image analytics at AstraZeneca

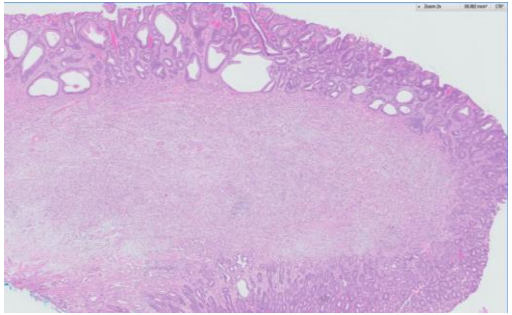
**Johan Karlsson (with contributions from Adam Corrigan & Yinhai Wang)**

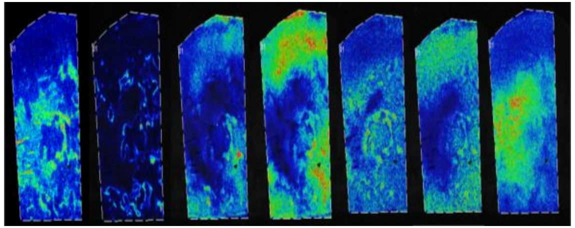
BigChem 2019

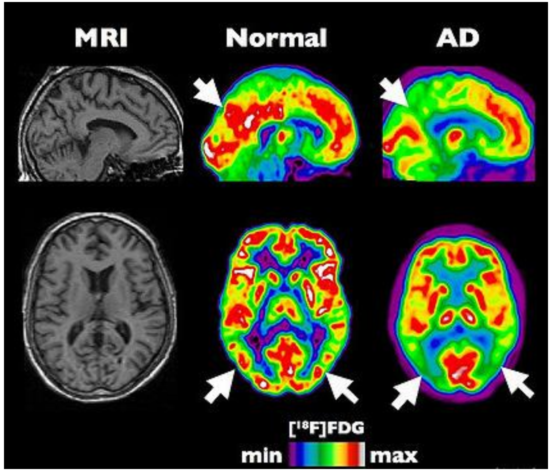


# Imaging in drug development is a diverse and rich area



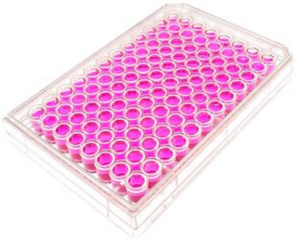


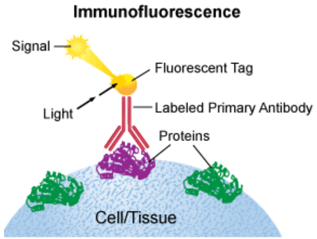


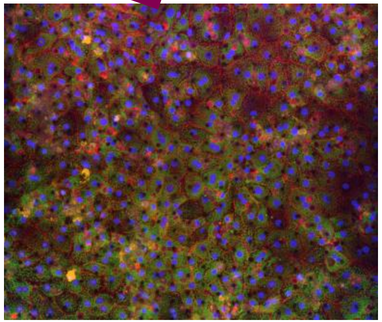


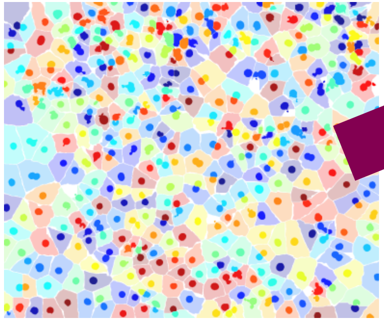
# Image analytics in preclinical drug development

# Typical Image Analytics Workflow



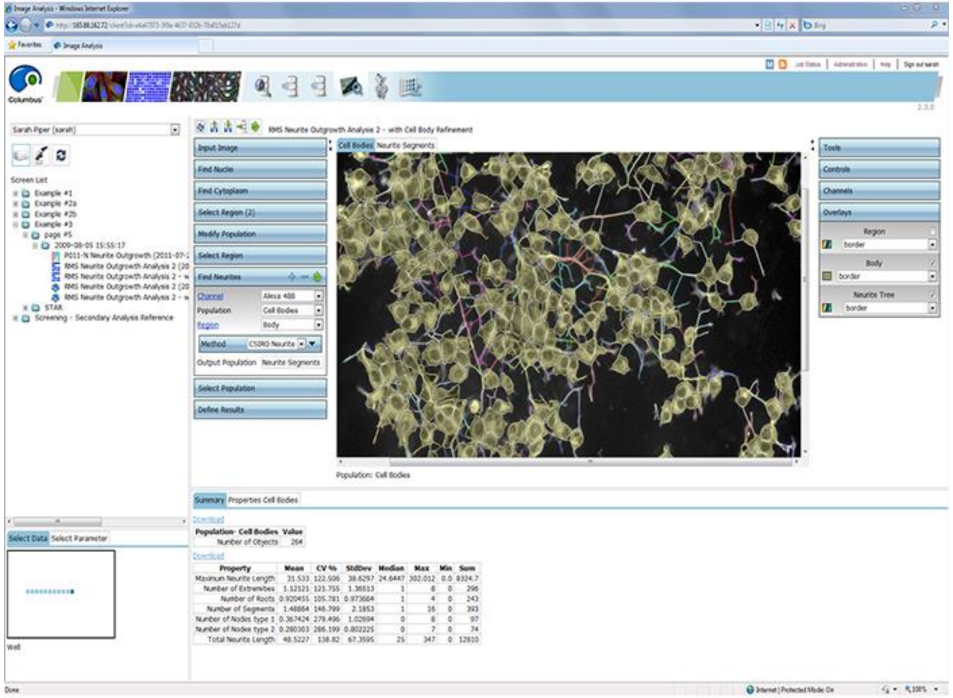


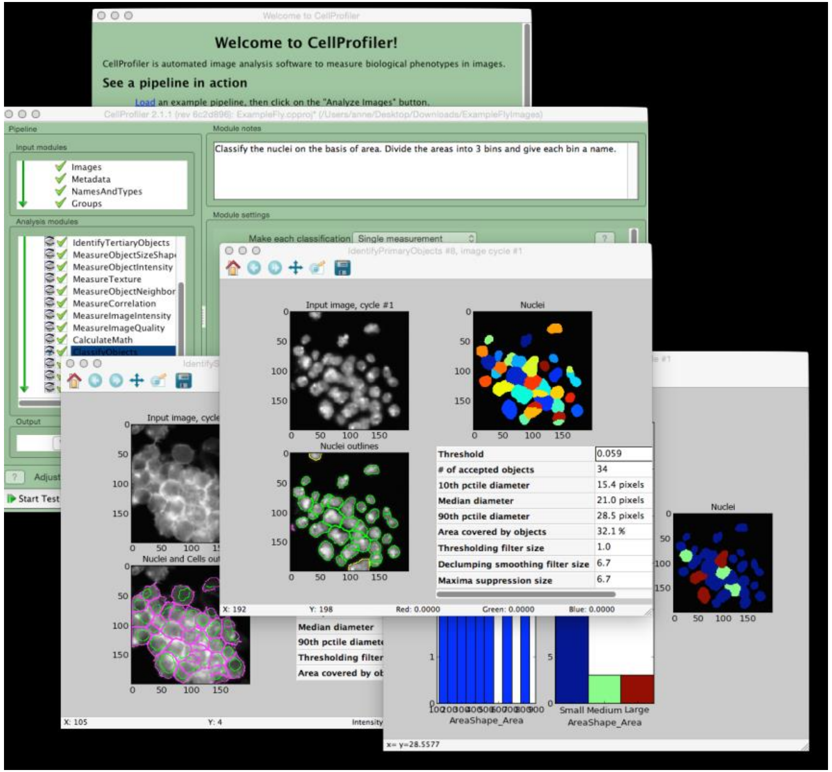





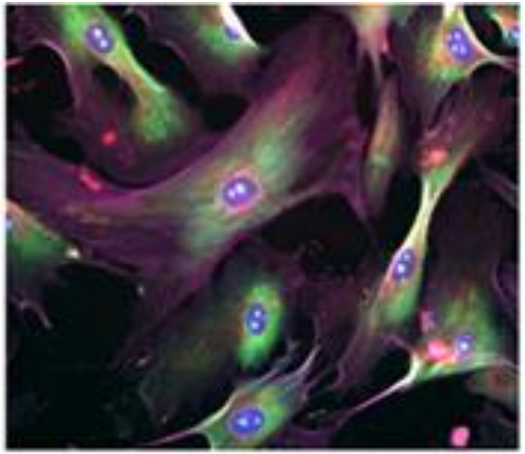
| Well | Field | HepCC3FieldFrac | HepZHeightSpread | HepTiltSpread | HepCellCount | HepATPBMean | HepATPBHighFrac | HepCC3Pos |
|---|---|---|---|---|---|---|---|---|
| A01 | 1 | 0.06 | 13.62280992 | 0.473103317 | 1455 | 35.64847258 | 0 | 0.041484716 |
| A01 | 2 | 0.05 | 12.09543754 | 0.385332613 | 1644 | 36.677698 | 0 | 0.029150824 |
| A01 | 3 | 0.05 | 12.46698843 | 0.412276633 | 1565 | 38.88030418 | 0.000668896 | 0.030100334 |
| A01 | 4 | 0.08 | 12.15373005 | 0.392444816 | 1475 | 44.13800275 | 0 | 0.058032555 |
| A01 | 5 | 0.07 | 14.14961318 | 0.484909507 | 1258 | 47.28197678 | 0 | 0.047658863 |
| A01 | 6 | 0.2 | 12.61694444 | 0.410657097 | 1369 | 42.50552717 | 0.00077821 | 0.182879377 |
| A01 | 7 | 0.08 | 12.50812583 | 0.424134195 | 1314 | 36.76156148 | 0.002417405 | 0.095890411 |
| A01 | 8 | 0.09 | 13.05970391 | 0.432667817 | 1427 | 35.8853677 | 0.003676471 | 0.105147059 |
| A01 | 9 | 0.06 | 15.32800383 | 0.523060336 | 1459 | 30.89599628 | 0 | 0.030612245 |
| | 10 | 0.06 | 11.02775962 | 0.372445543 | 1438 | 29.01598804 | 0 | 0.033650329 |
| | 1 | 0.05 | 13.68296575 | 0.510382515 | 1596 | 32.52823597 | 0 | 0.029761905 |

# Off-the-shelf: Columbus, CellProfiler, Visiopharm, Halo, ImageJ (and others)

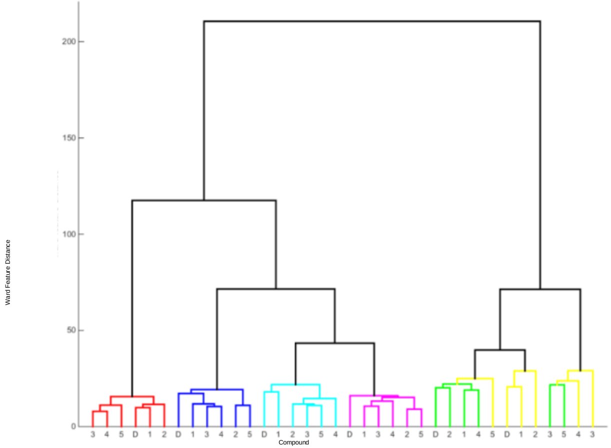




# Example



CellPainting Assay




Clustering analysis

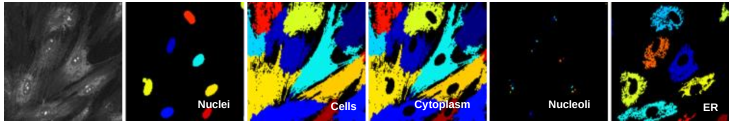
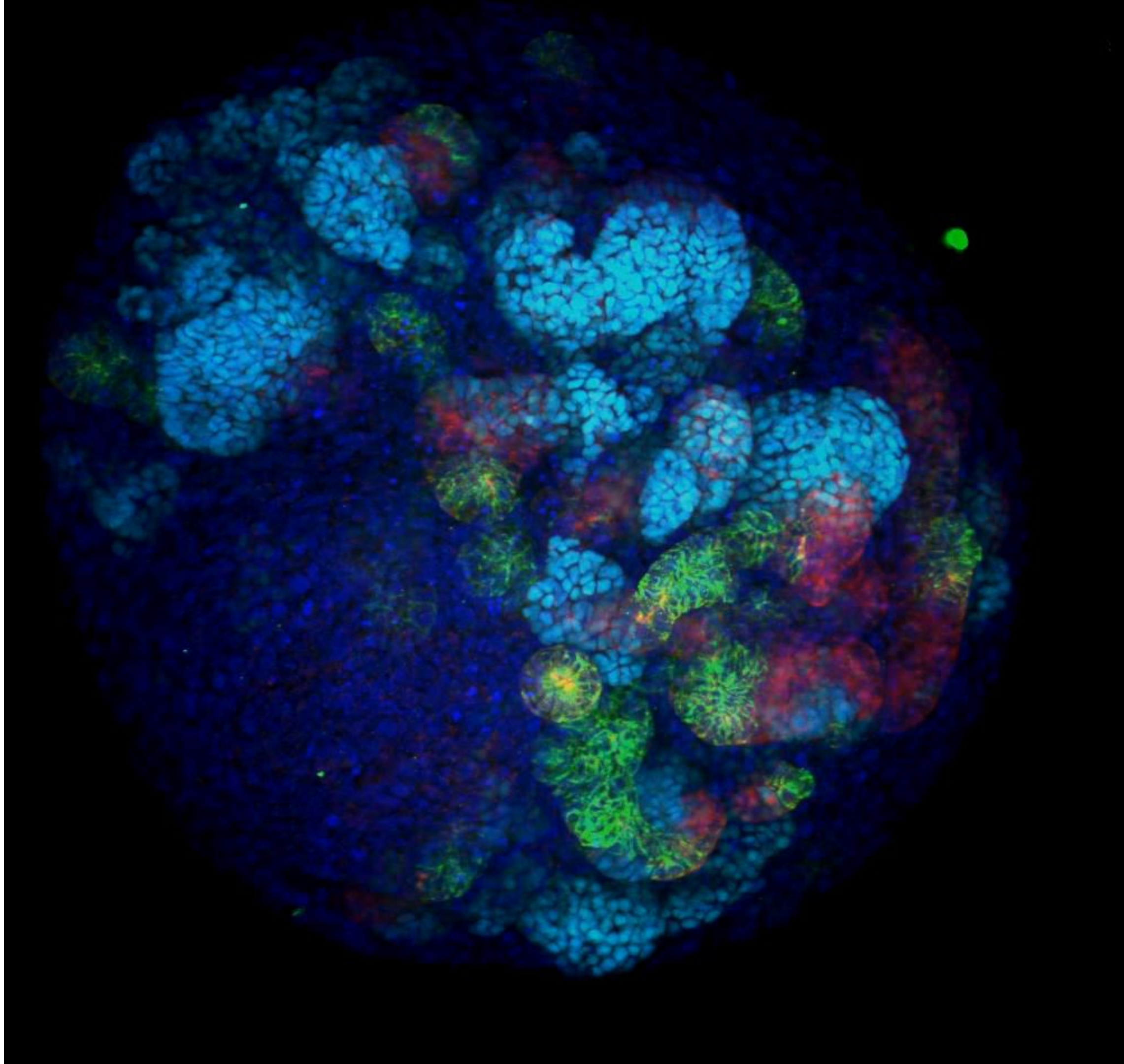



Figure 3: Cell Segmentation for Feature Extraction

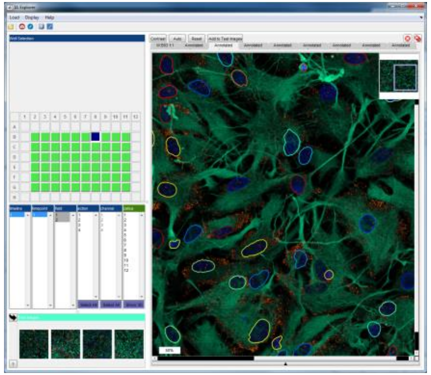
CellProfiler

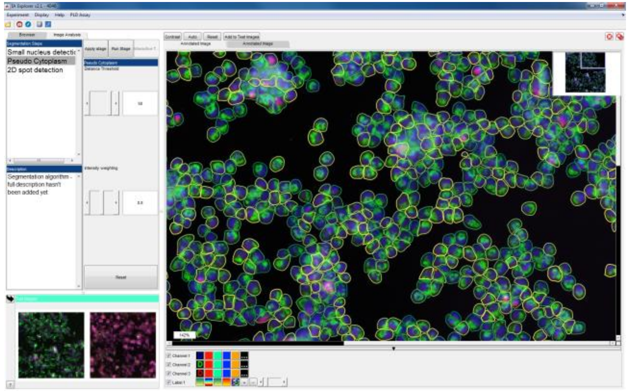
As experiments get more advanced, the need for image analytics beyong the standard workflows increases.

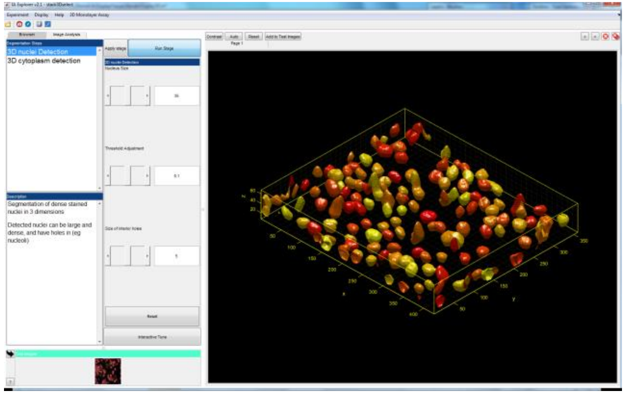
Example: quantitative evaluation of 3D multichannel images of complex organoids differentiated from stem cells.



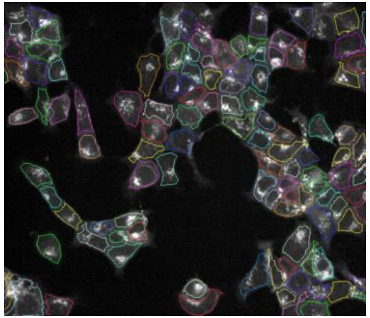
# Custom analysis not available in off-the-shelf tools: inhouse Matlab tool (below) + Python scripts

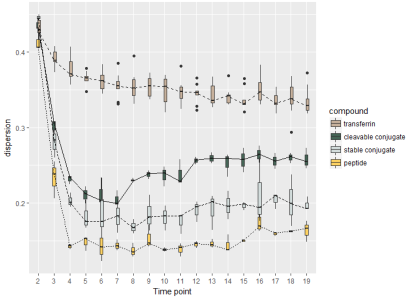




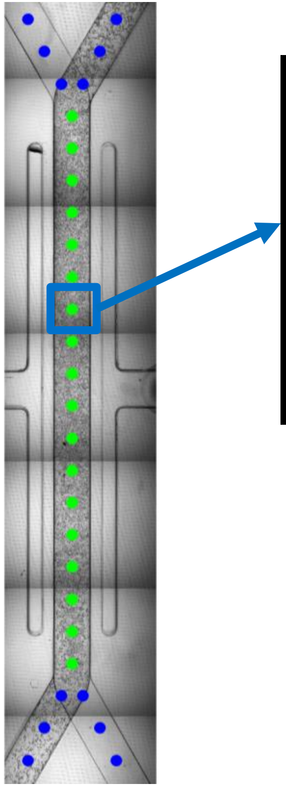



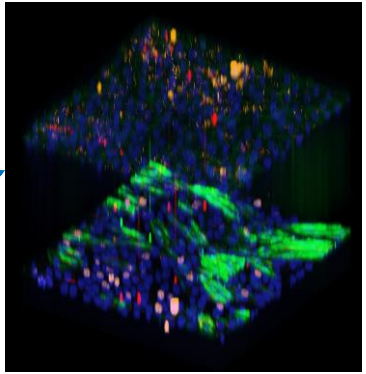

## Example: Aggregation patterns internalized cargo sub-cellular intensity





# Not only analysis if existing images: Using image analytics to control the microscope





4X scan of whole chip: automatically identify FOV for high magnification imaging

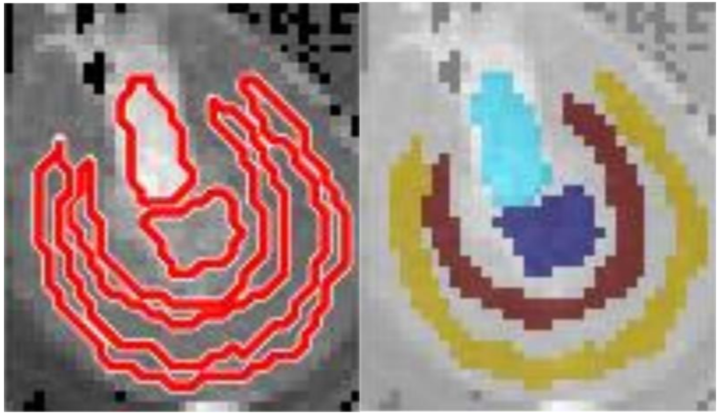
- Not known in advance where the sample should be imaged
- Low resolution first pass to map out sample
- Image processing to identify coordinates of interest
- Higher resolution second pass


- Potential to be applied across many experiments
- Intelligent microscope – automation of acquisition

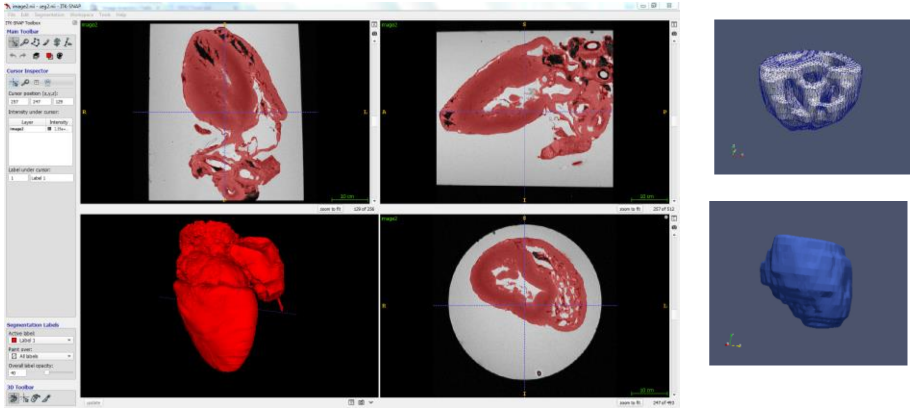
# *In/ex vivo* image analytics

Automation of manual work has the potential to save a lot of man hours as well as increase quality and reproducability, even with non-AI methods.

Registration based ROI segmentation in kidney MRI



Region growing based heart MRI segmentation for mesh simulation



# Nonbiological imaging: Quantification of capsules layers (CT)

Gel and film layers segmented

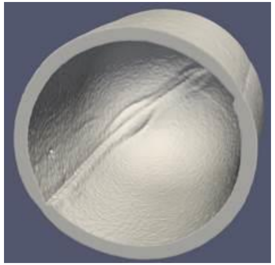
Thickness of both layers calculated everywhere

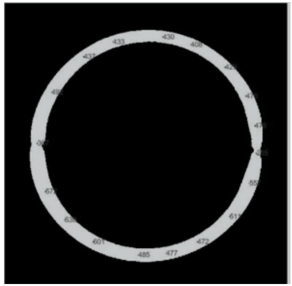
Projection to 2D thickness maps & detection of seam region

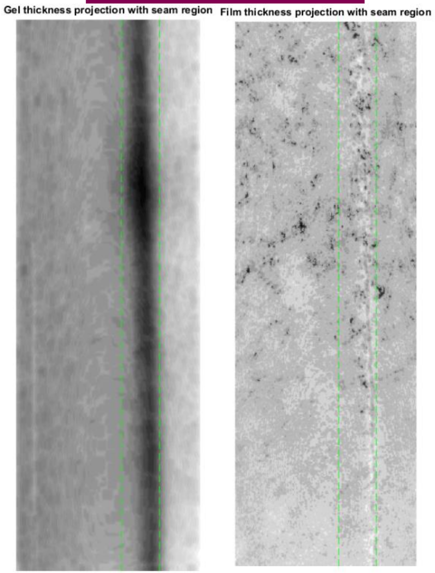
Feature extraction

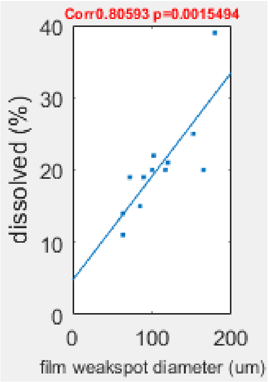
- Mean thickness
- Standard deviation
- Porosity
- Weakest point
- Etc

Correlation with dissolution

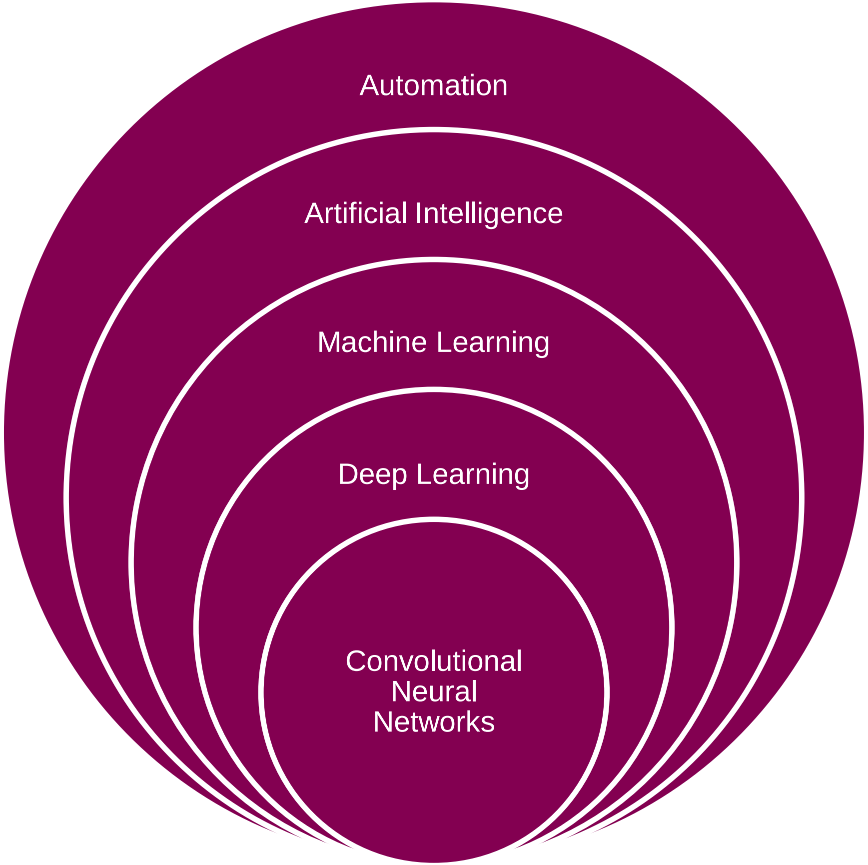






# AI, machine learning, deep learning


Automation
Artificial Intelligence
Machine Learning
Deep Learning
Convolutional Neural Networks

# Increasingly we turn to machine learning for image analytics

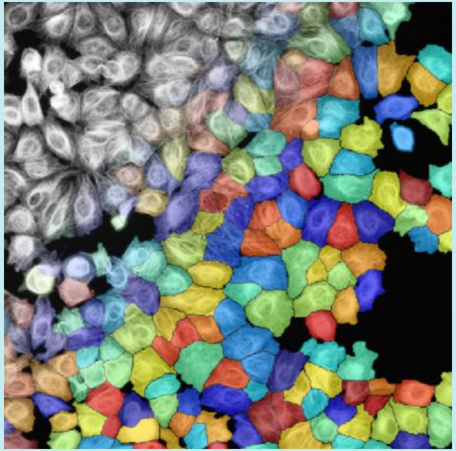


**Image Segmentation**

- Train machine to extract more information from fewer labels
- Single cell phenotypes

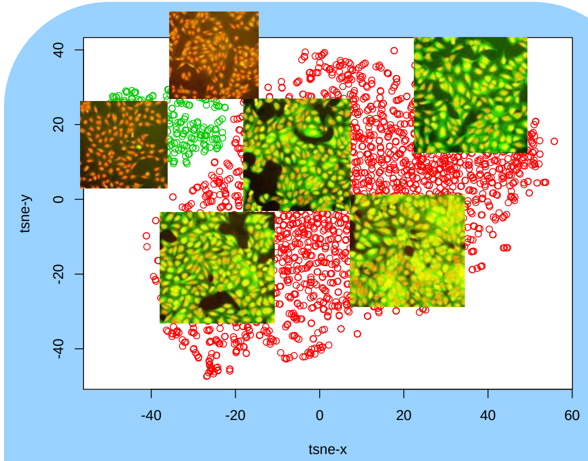



**Image Classification**

- Phenotypic screening
- Artefact Detection
- Translation prediction

**Dataset annotation**

- Image features
- Compound information
- Cell line metadata

Learn to pick out high level correlations

Specific problem solving

Machine Intelligence

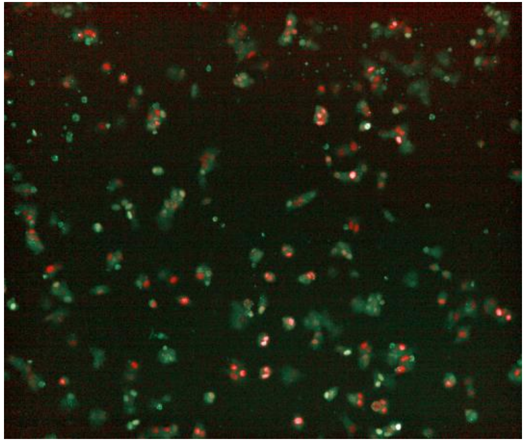
# Image Classification

Image classification with neural nets

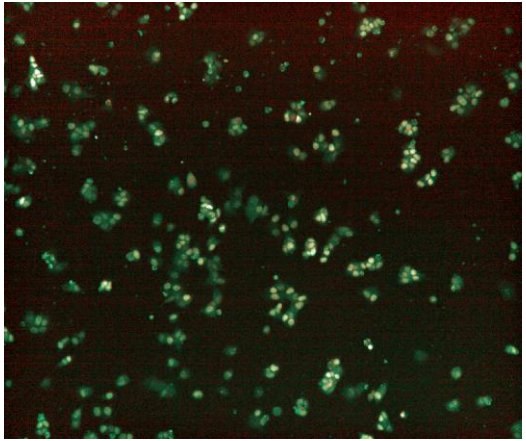
- Train autoencoder (without annotations)
- Train classifier network (using annotations)
- Transfer learning (when available dataset is small)
- Active learning (to annotate efficiently)

We do all of them.

# Typical classification task:
## Imaging Screen – identify compounds which cause desired effect



Negative Control



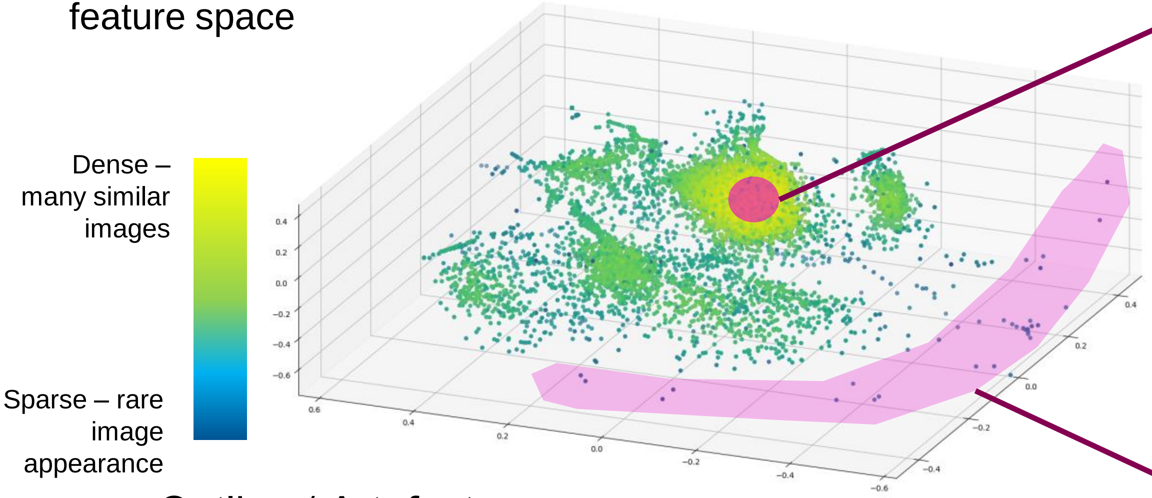
Positive control – ‘hit’

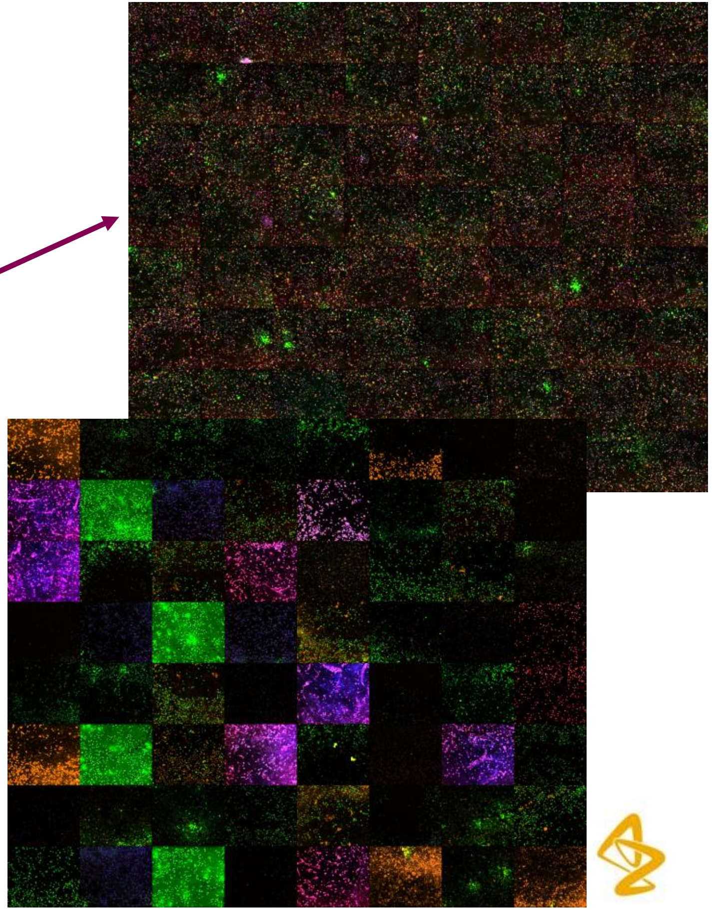
Image phenotype is not always so clear-cut, especially at the start of development

- Traditionally, assay development is iterative process:
  Experimental methods ⟷ Image Analysis
- At each round, how do we know the image analysis is capturing the treatment effect?
- Can deep learning be better and faster than us?

# Deep learning autoencoder for cell screening --Automatic Outlier Detection

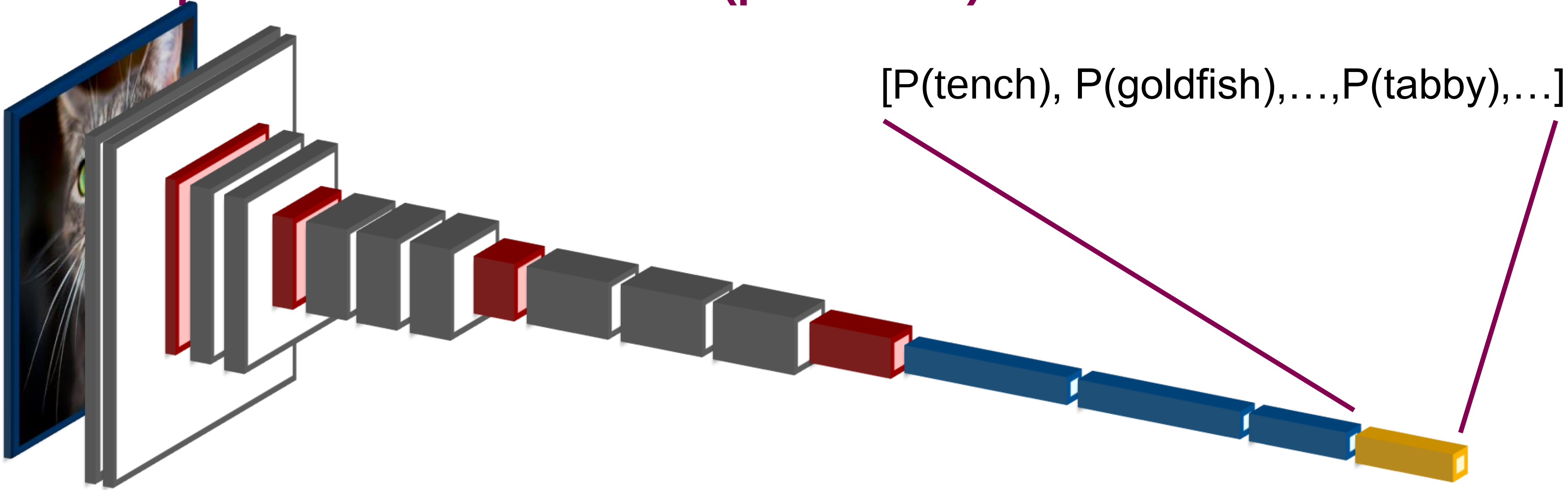
Deep Variational AutoEncoder (VAE) – unsupervised encoding of image appearance into spatial dimensions.
Points close together = similar images
Automatically identify odd images as low density regions in feature space



Outlier ≠ Artefact
(could also be uncommon phenotype)
⇒ Rapid manual annotation of latent feature space

# Image classification
# Example: VGG16 architecture (pretrained)

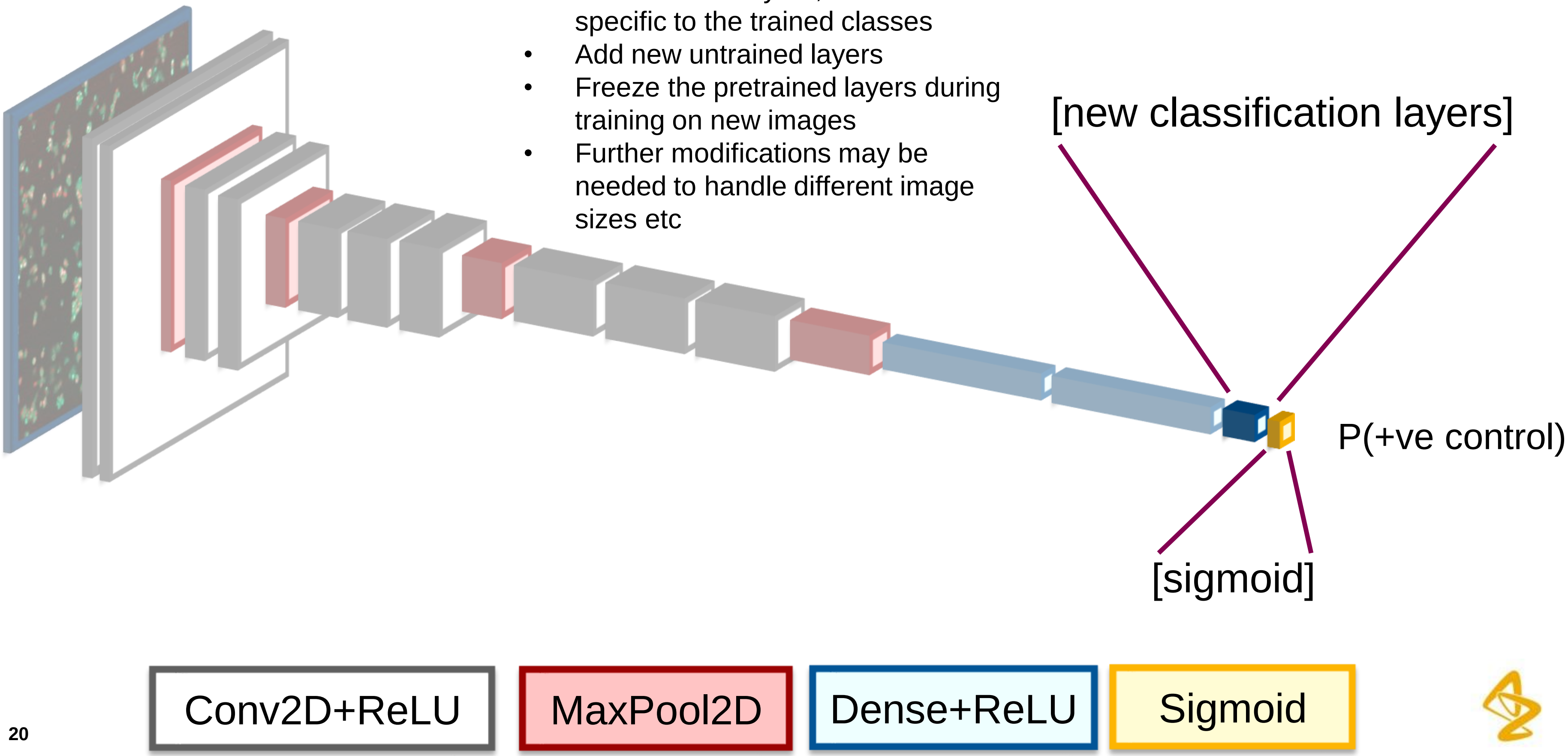



138’357’544 parameters
1’200’000 training images

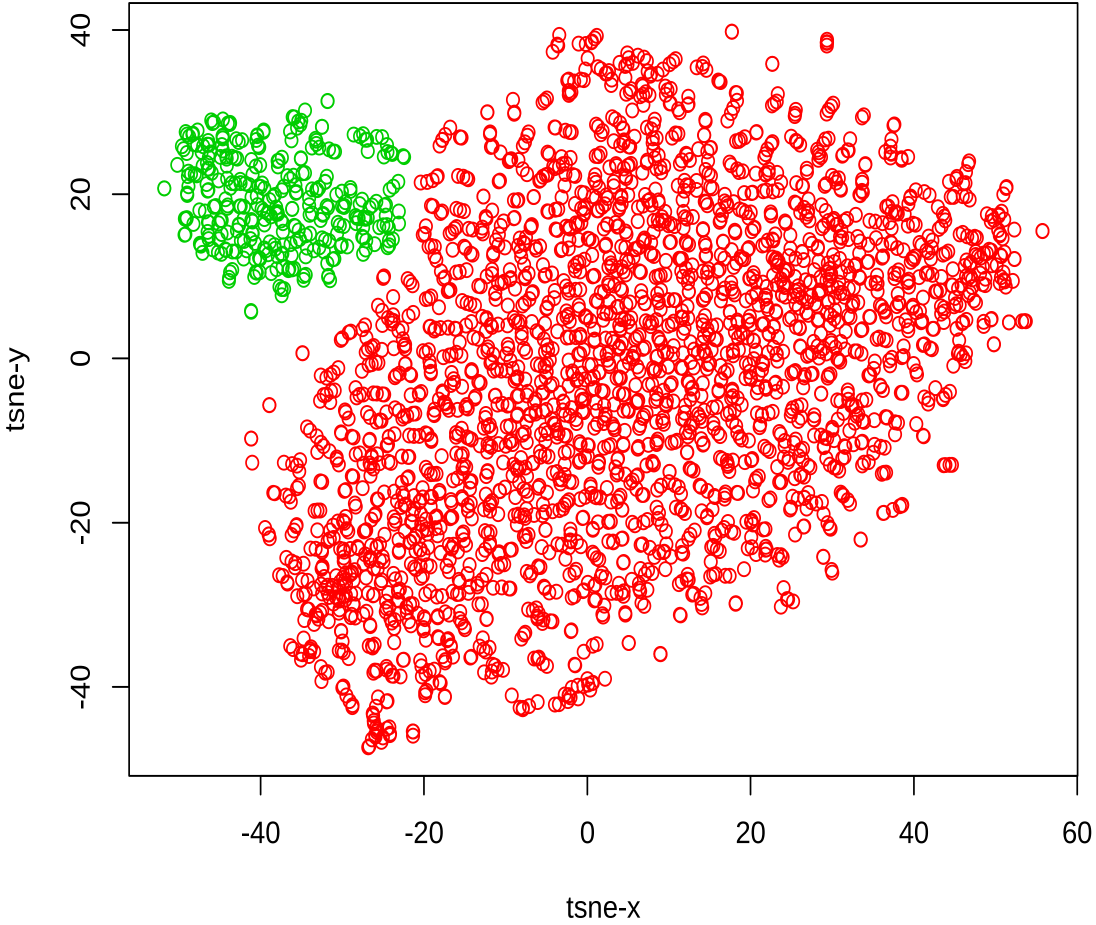
Conv2D+ReLU | MaxPool2D | Dense+ReLU | Softmax

# Transfer Learning (when available dataset it not large)

- Remove final layers, which are specific to the trained classes
- Add new untrained layers
- Freeze the pretrained layers during training on new images
- Further modifications may be needed to handle different image sizes etc

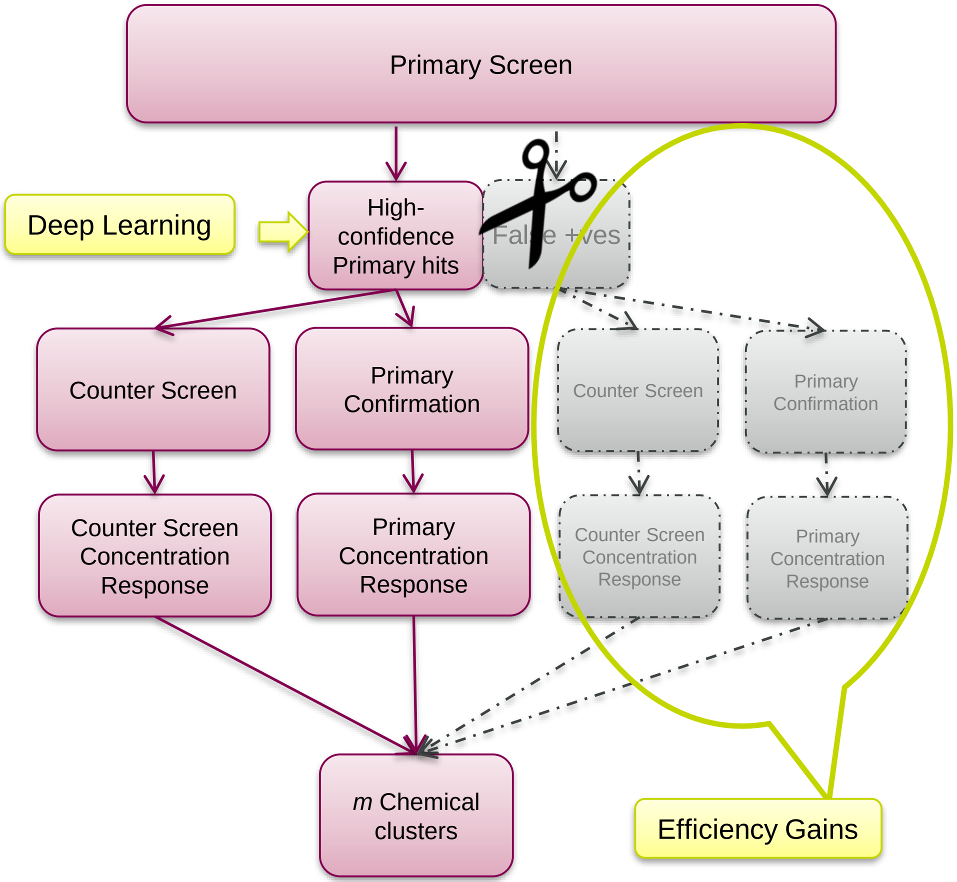


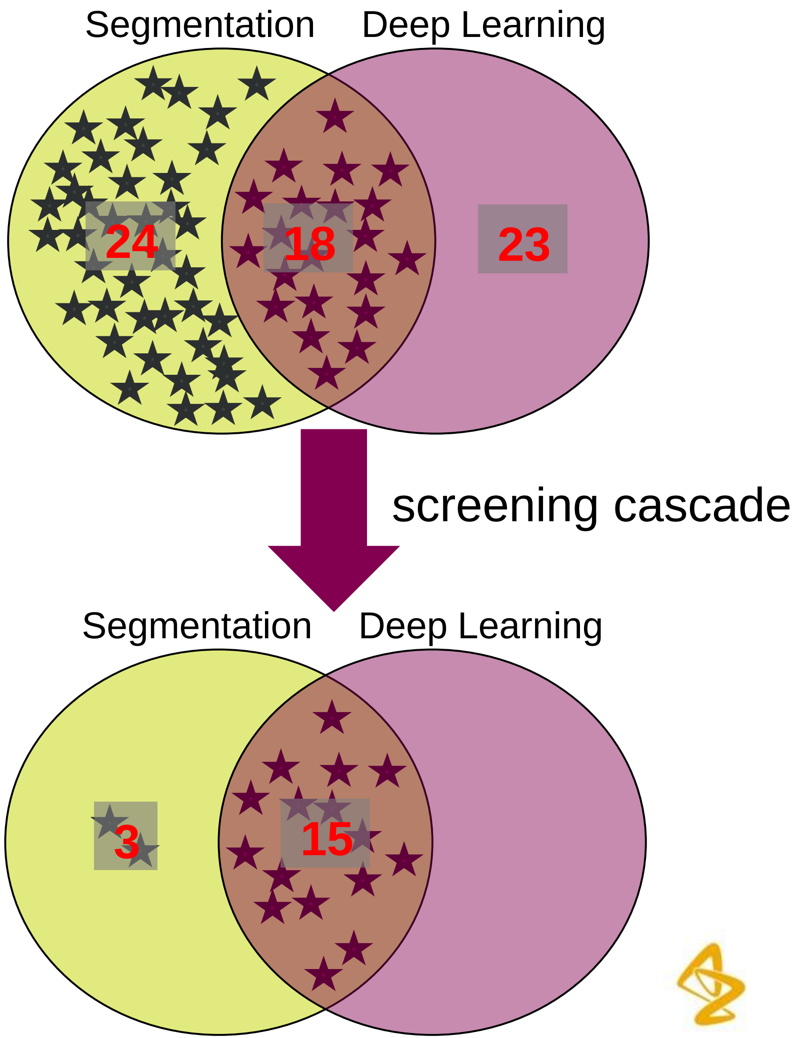
# Deep Learning assisted High Throughput Screening



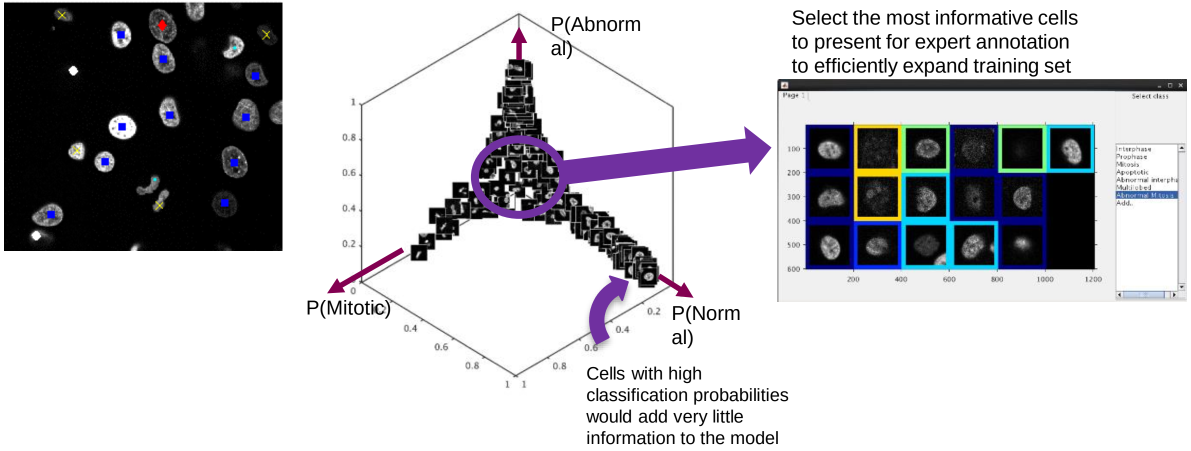

- Train deep learning classifier using controls as training data.
- Use it to select hits from screen instead of using standard hit criteria like fluorescence in nucleus above/below some threshold.

# Deep Learning assisted HTS



# Active Learning for Efficient Training
# Example: Mitotic Phenotype Classification

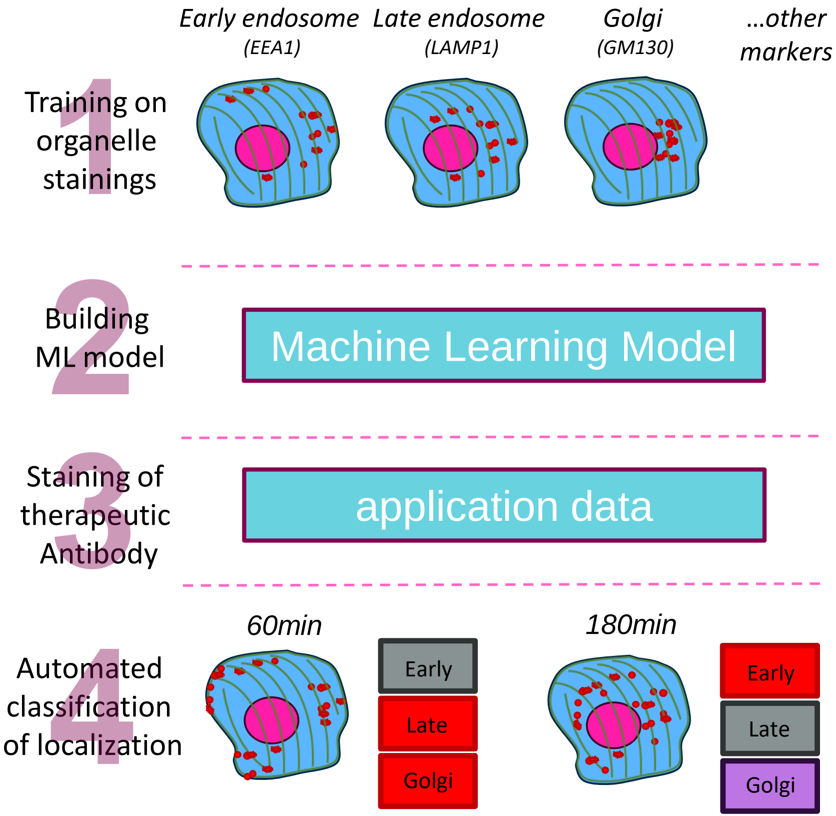


# Image Classification tasks we are currently looking into

# Classify antibody localization over time

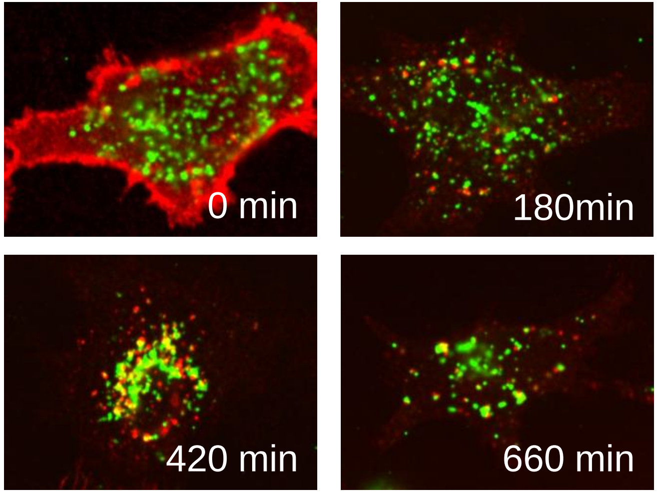
## Inspired by recent Kaggle HPA challenge for protein localization

**where we did not win, but placed in the top 5% ☺**




- Currently studying uptake of therapeutic antibodies requires multiple, complex co-stainings.

→ Goal: Machine Learning based classification

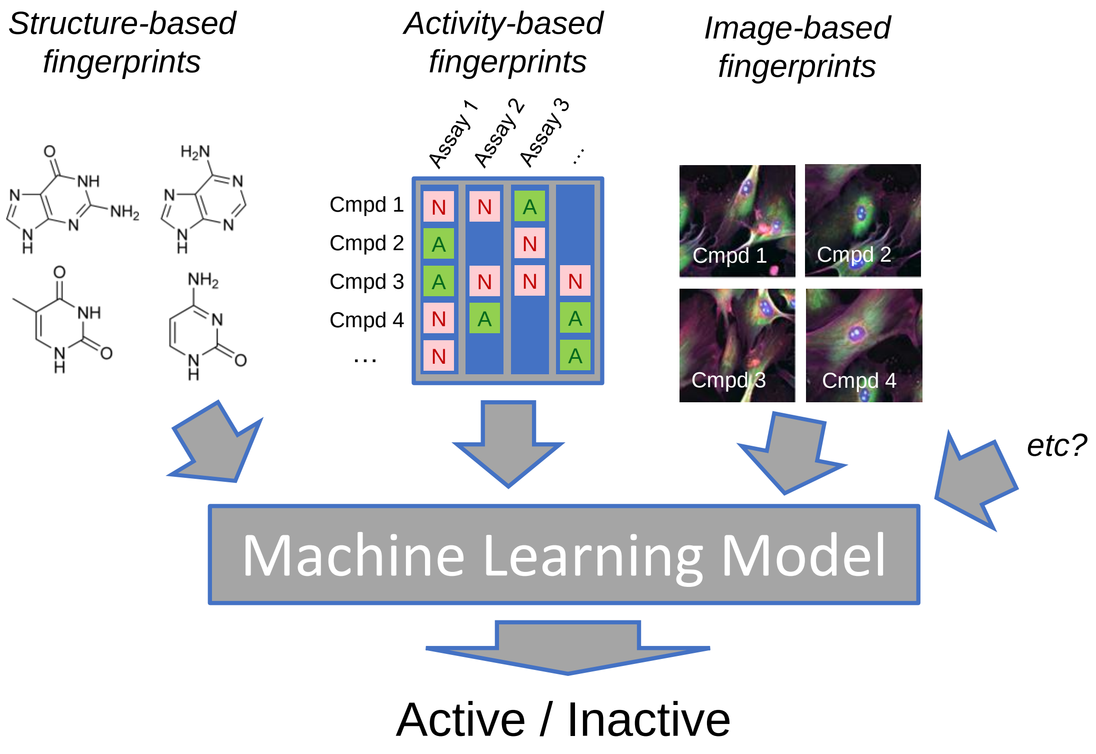


# Predicting compound activity using AI

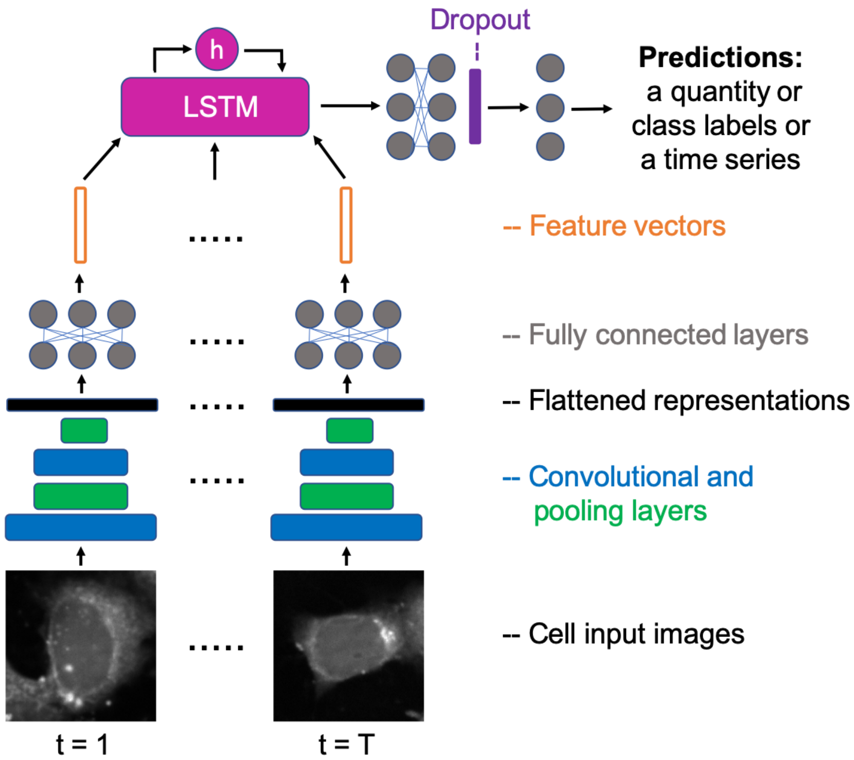
## Deep Learning for Drug Discovery, using Diverse and Sparse Data Sources

- Goal: Repurpose experimental data to predict compound activity in new assays

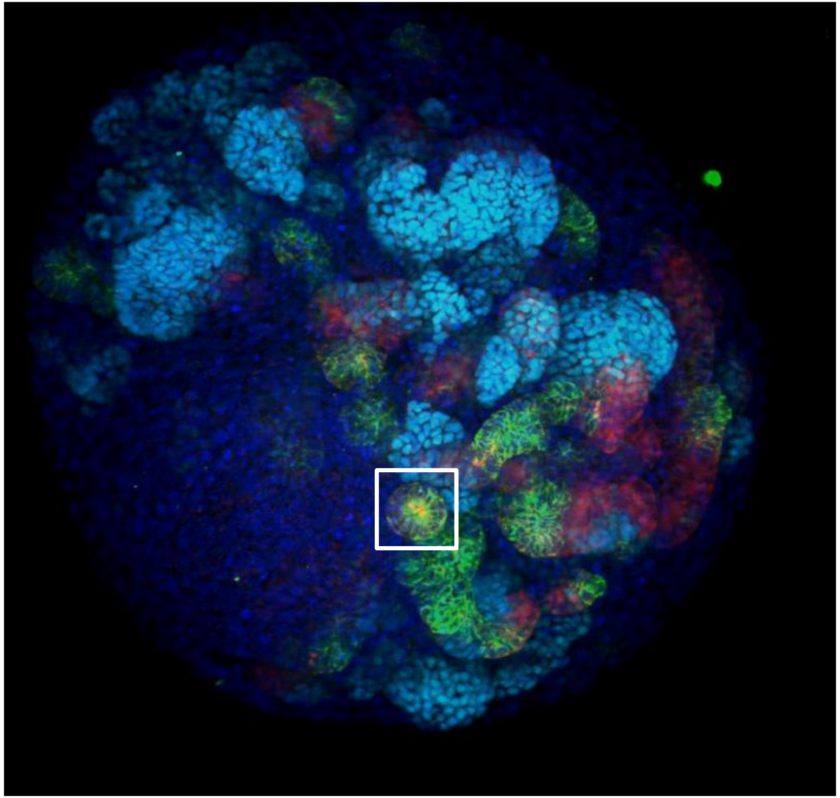
- Combine different data types
  - Structure-based fingerprints
  - Activity-based fingerprints
  - Image-based fingerprints
  - …



# Time series image data
# Predicting protein production following drug uptake



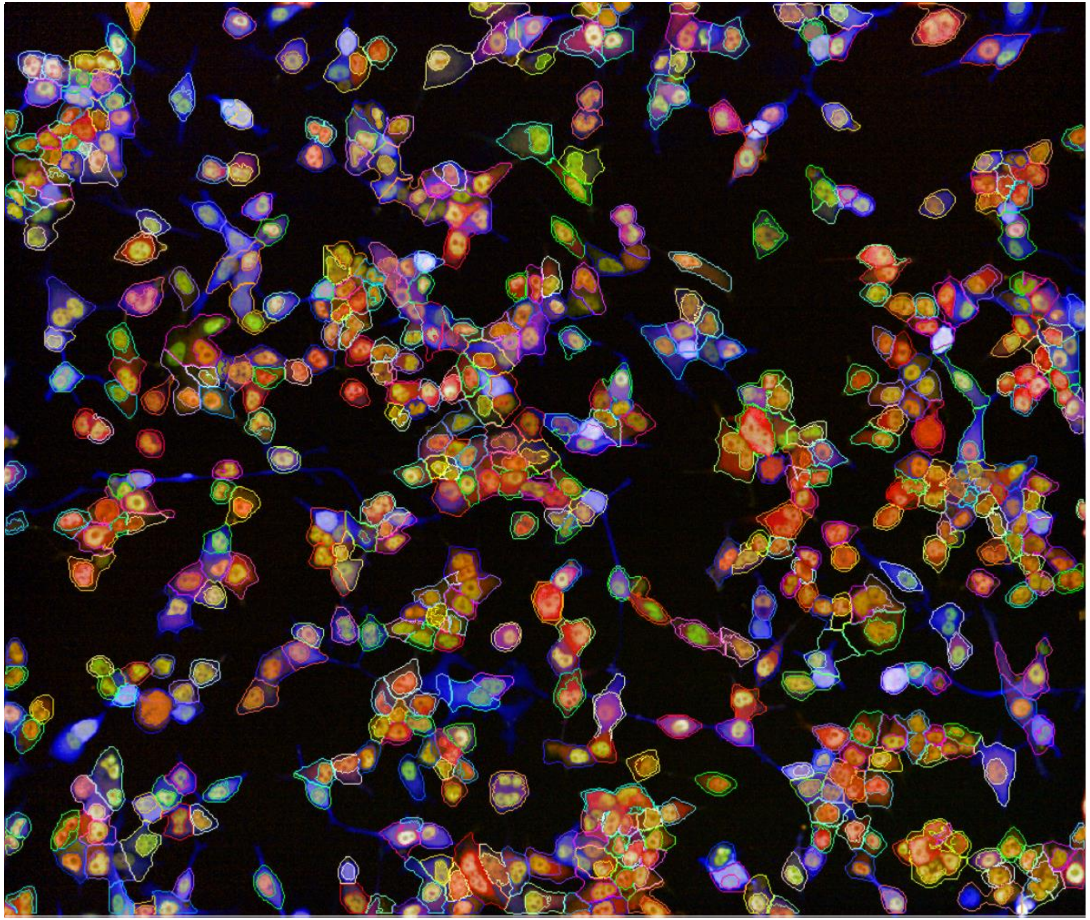
# One shot learning



The goal is to be able to tell the microscope
”Get me high resolution images of things like this”
Mark one example of something in one well and then get results from a stack of micro well plates.

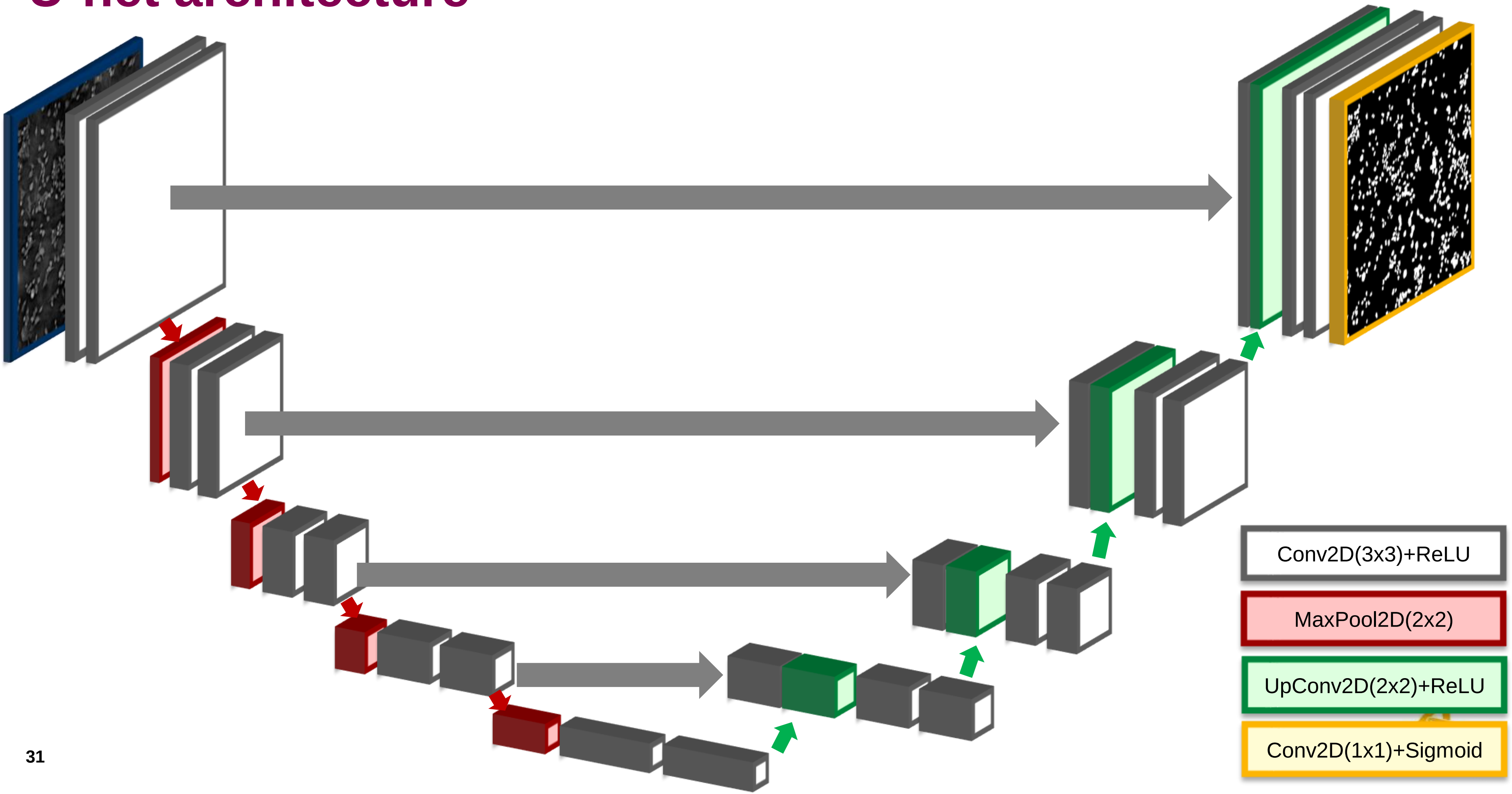
# AI Image Segmentation (and modality generation)

# Single Cell Analysis

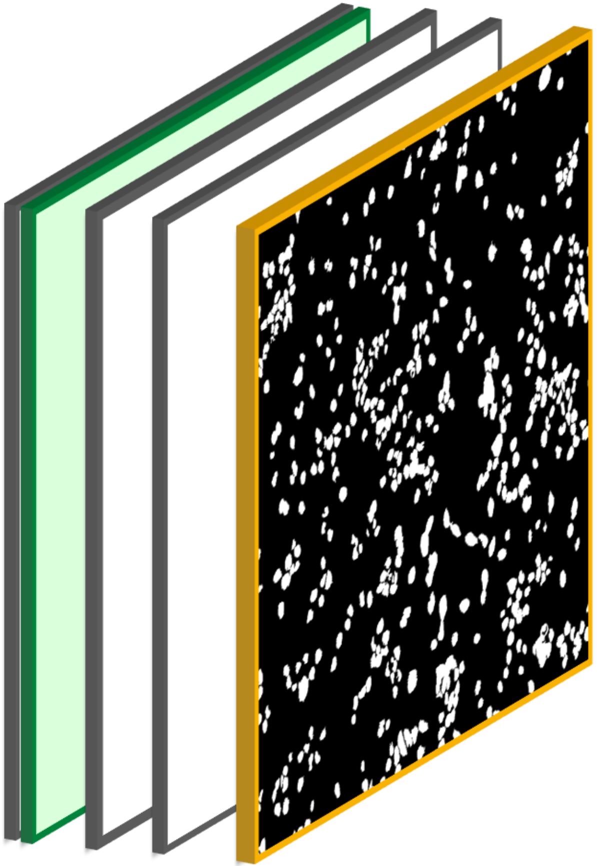


- Cellular images have lots of detail:
  - Cell behaviours
  - Cell types
  - Local environment
- Need to identify and segment cells to quantify rich information
- But don’t want to have to use a fluorescent marker for the nucleus and cytoplasm:
  - Cost of reagents
  - Preparation time
  - Uses up colour channels

# U-net architecture



# Final layers

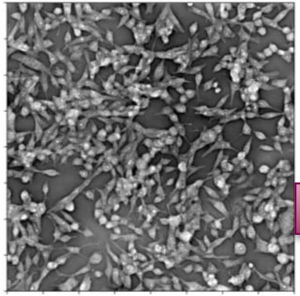


- Final activation depends on application
  - Sigmoid for binary classification
  - Softmax for multiple classes
  - No activation function for regression tasks – image to image

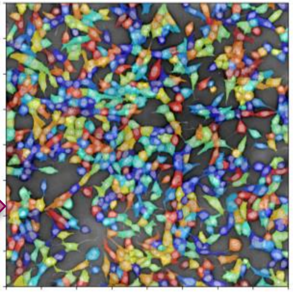
# Towards Label-Free Imaging in Cellular Screening using Deep Learning Segmentation

**Segmentation of nuclei and cells from phase contrast images**

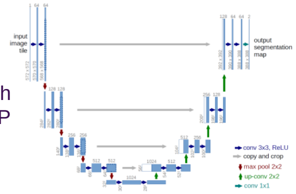
1) U-net prediction of nuclei and cell outlines from phase contrast



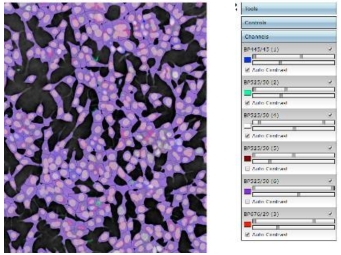
Neural network learns to find nuclei and cells using only the phase contrast image – label free



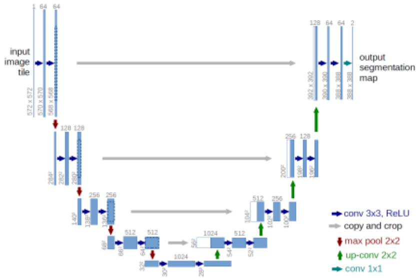
U-net trained on examples co-stained with DRAQ5 (nuclei) and BFP (cytoplasm) to provide training data




2) Build Columbus importer to integrate into workflow

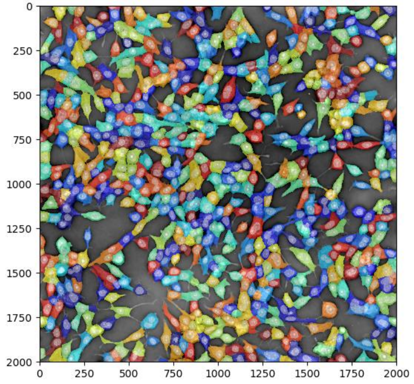


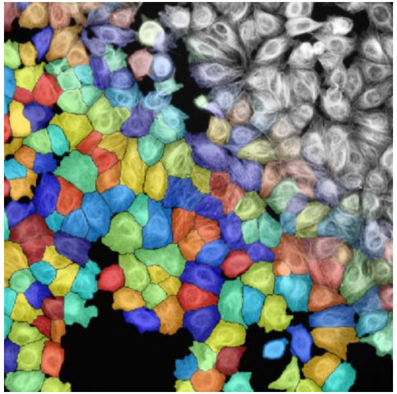
# Directly generalizable image segmentation workflow

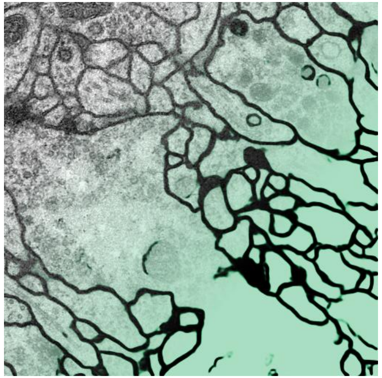


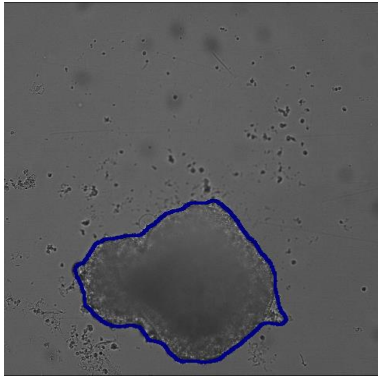

Training annotations can be:

- Another modality (e.g. a specific fluorescence marker)
- Manual annotation of images



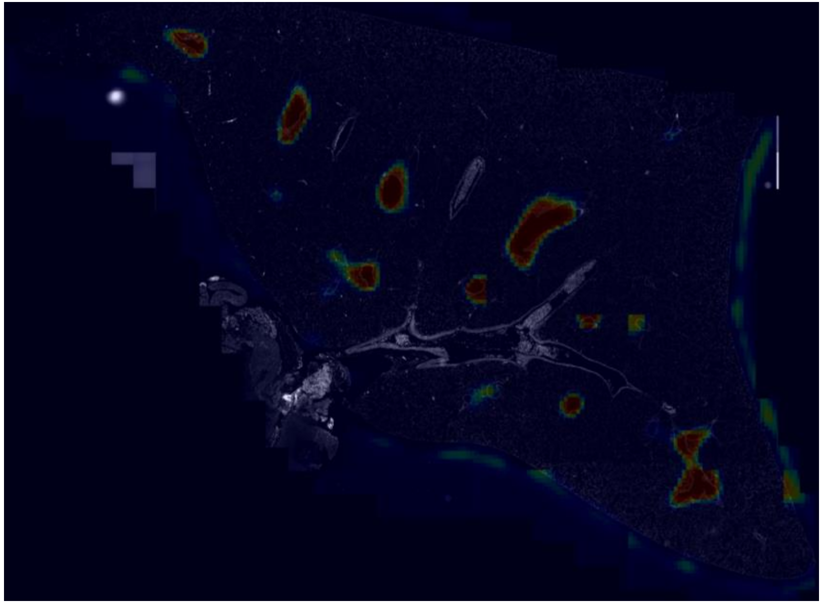




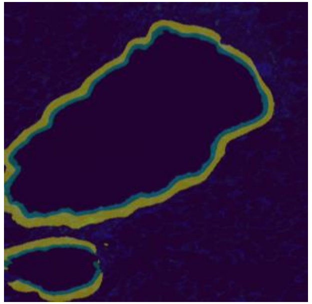
# Tissue segmentation AI analytics beyond the off-the-shelf tools

- Similarly for tissue analysis we use both off-the-shelf tools and custom analysis development.
- For example, for automatic quantification of compound quantities in distinct tissue types, tissue regions are classified, segmented, and quantified using deep learning, using confidence measures based on conformal prediction at each step.

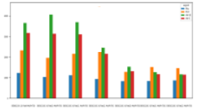
Find ROI in low resolution using
Fully Convolutional Neural Networks



Segment epithelial / subepithelial in ”medium” resolution with a segmentation network

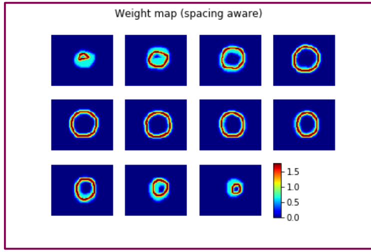


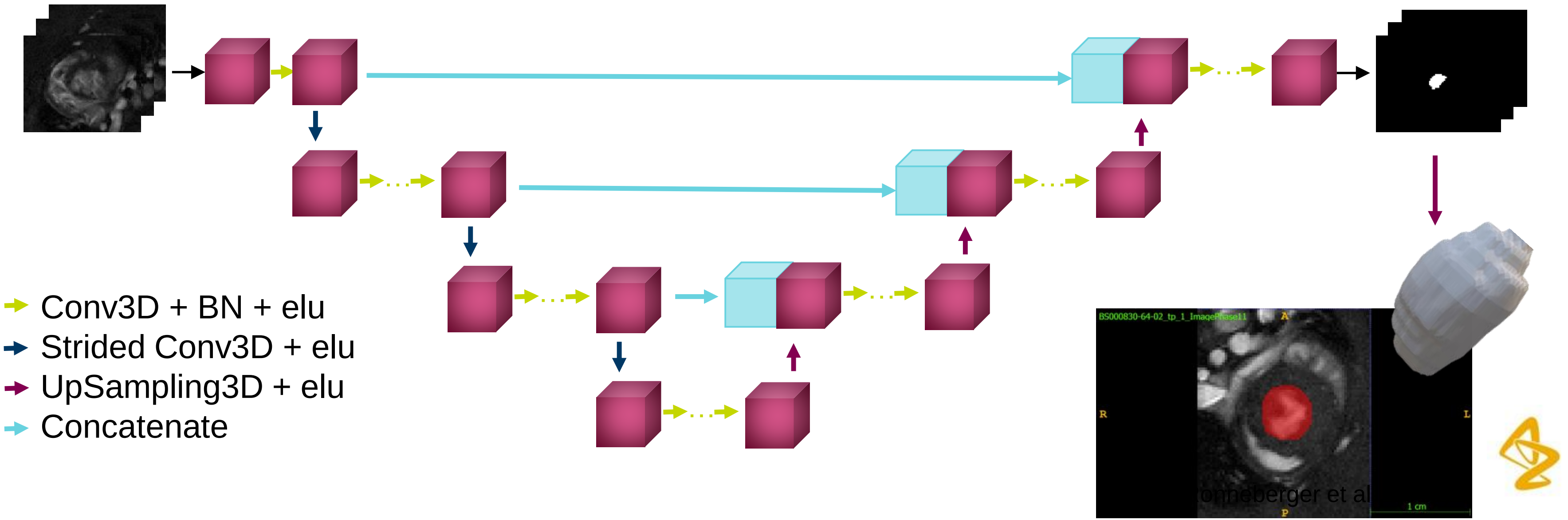
Quantify drug stain in highest resolution



# 3D MRI Deep Learning Segmentation

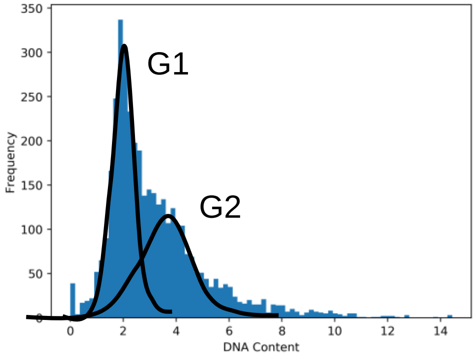
- Dataset of 189 3D MRIs of rat hearts –systole and diastole phases manually segmented
- **Goal:** Automatic segmentation and prediction of Ejection Fraction
- **Approach:** U-net inspired convolutional network adapted to work on 3D images

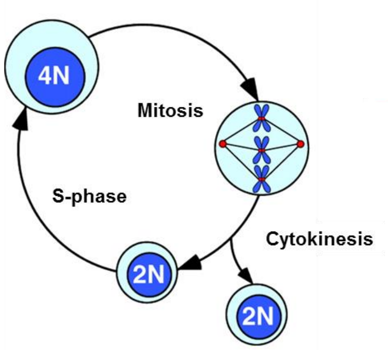


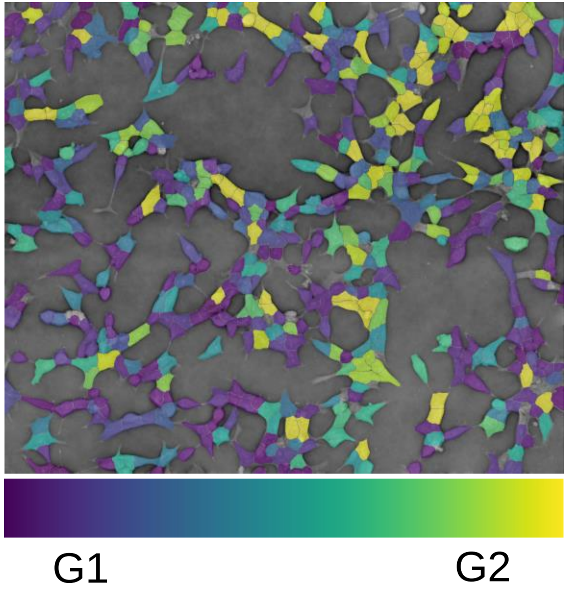



# Beyond segmentation – predict other property than segmentation mask

Fluorescent DNA Stain

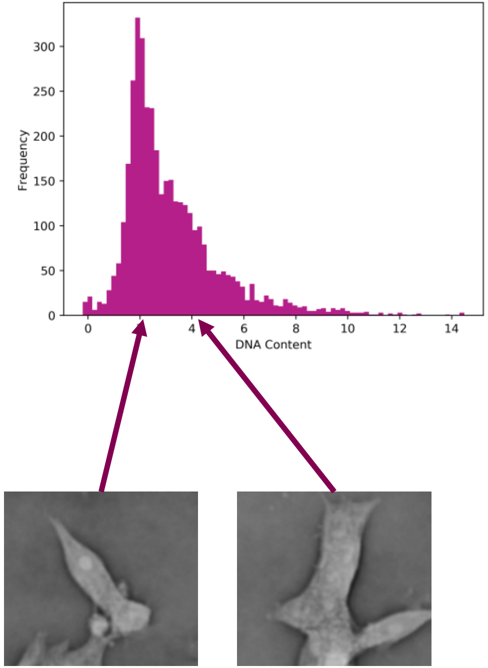


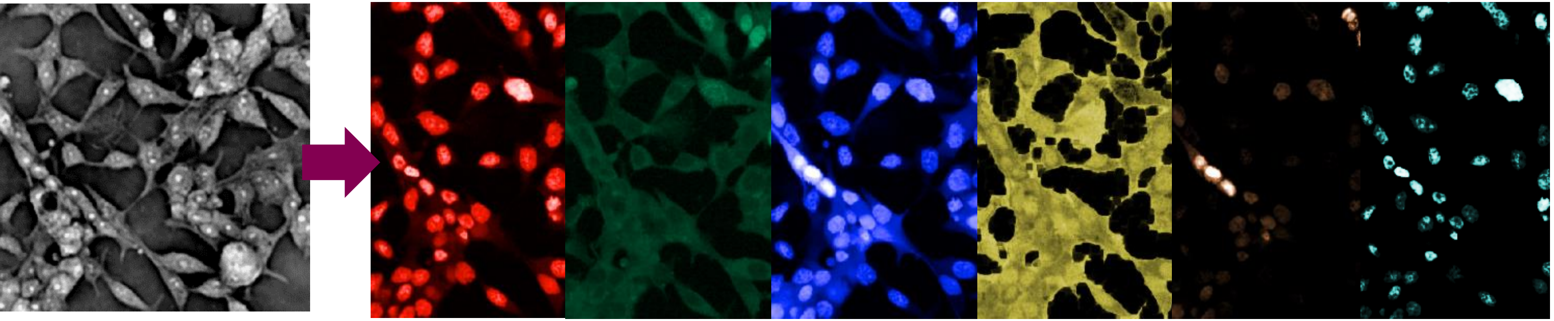

Transfer learning – repurpose pretrained network to predict new output at single cell level

Label free measurement



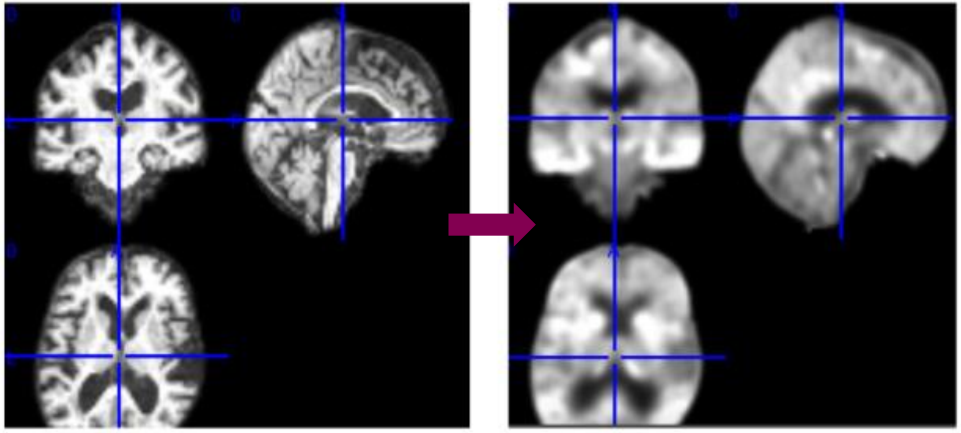
# Looking ahead: Digital Biomarker Panel

## Predict panel of biomarkers directly from label-free images



- ✓ **Reduce cost**
- ✓ Reserve microscope channels for **more relevant biology**
- ✓ **Live imaging**

- DNA
- Proliferation
- Cell Type
- Metabolism
- Cell Death
- Senescence
- Etc..



MRI PET

# Summary

- Image analytics is a critical element in many parts of the drug discovery and design process.
- Off the shelf tools are combined with inhouse custom R&D.
- Machine learning and especially deep learning (and especially especially CNNs) is increasingly the method of choice.
- AI is used both for automating existing processes and for making new things possible.

# Questions?