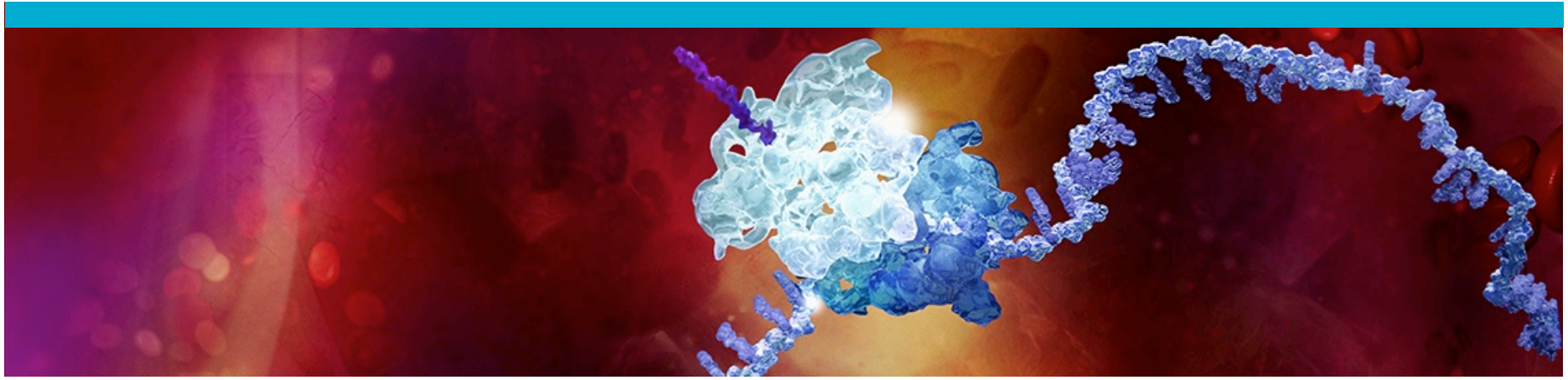
AstraZeneca
IMED Biotech Unit

# Machine Learning for Smarter Drug Discovery

**Claus Bendtsen**

Spring 2018



# Drug Discovery Today

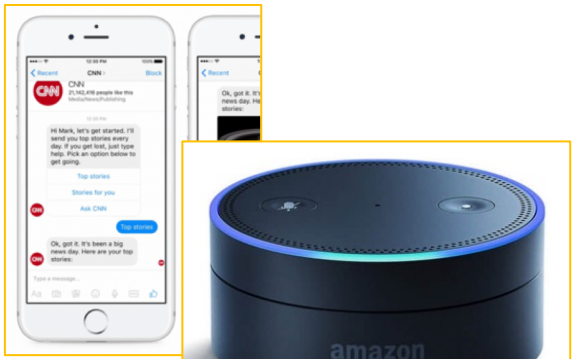
## An industry perspective of Today's Challenge


We need to become **better**, **faster** & **cheaper**

Source: PhRMA profile 2016

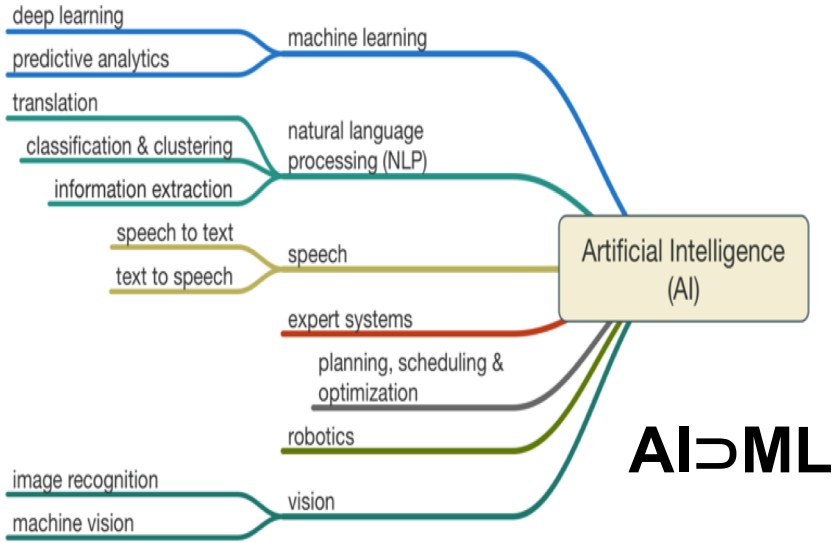
# Machine Learning & AI can help us doing so

## … by simulating human behavior (intelligently)





> ***"AI is the new electricity…just as electricity transformed industry after industry 100 years ago, I think AI will do the same."***
>
> *Andrew Ng, founder of Google Brain & Coursera, now at Baidu*



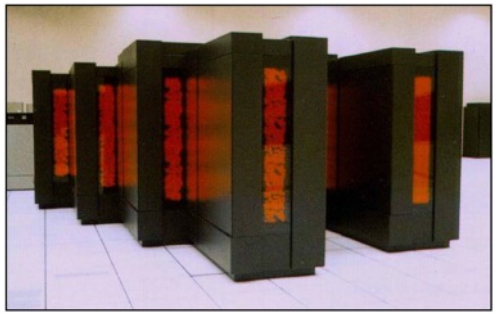

**AI⊃ML⊃DL**

# Why now?

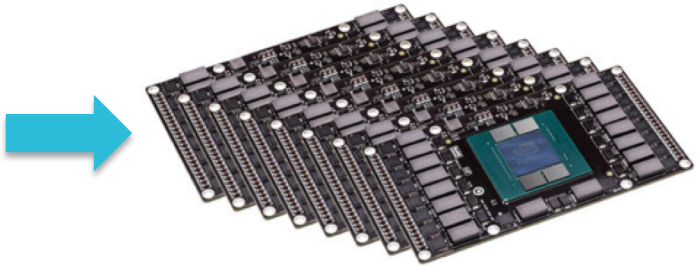
## Advances in HW and SW have transformed what is possible

**~1 million times more compute power**

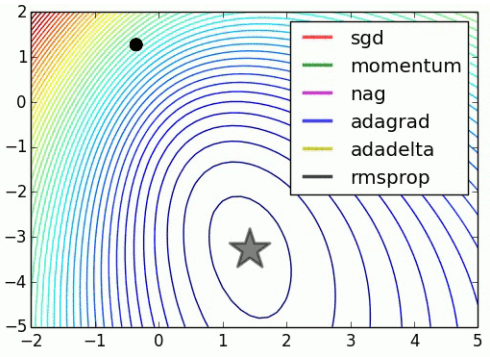
**Algorithmic advancements**



**Late 1990's**
~1 unit = $800K



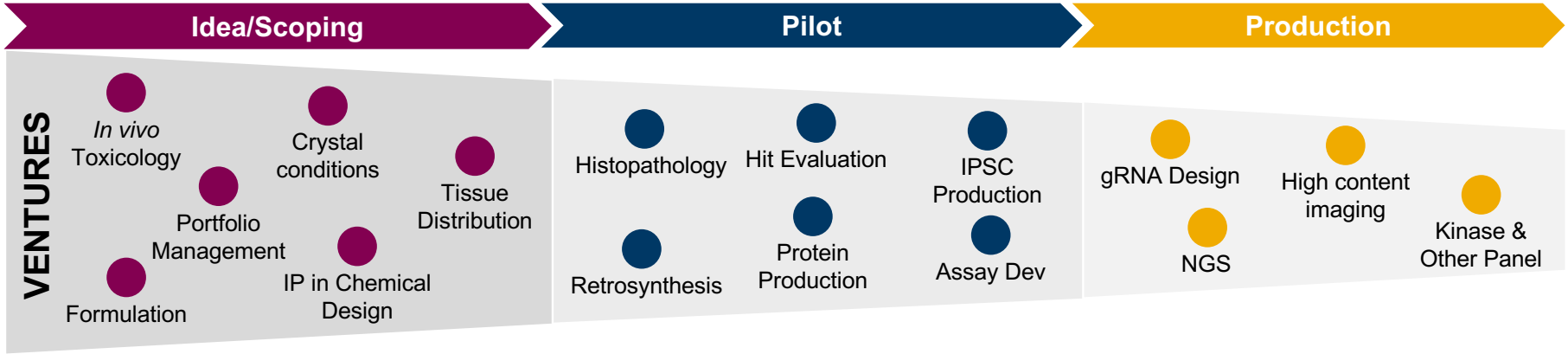
**Now**
~1 unit = $1

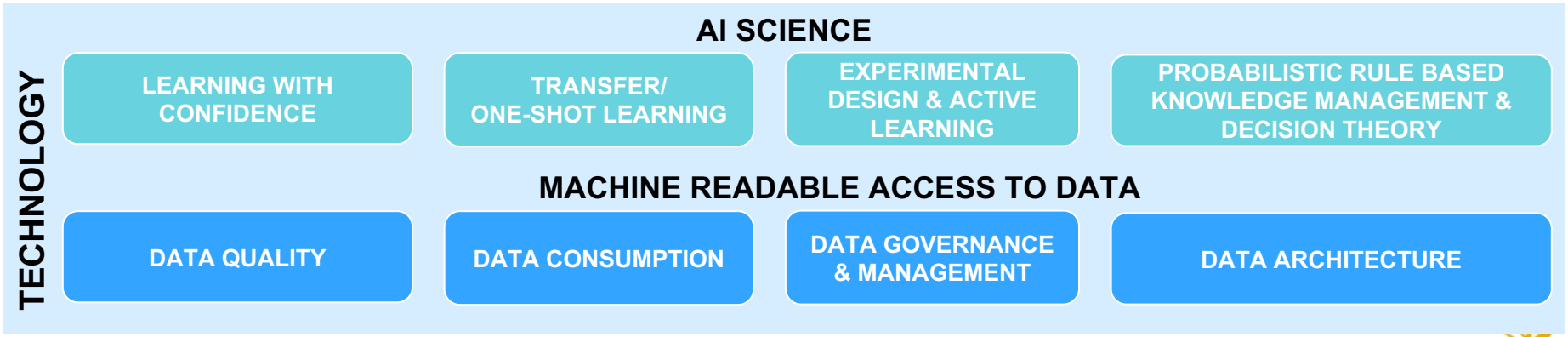



Source: http://hduongtrong.github.io

# Our approach

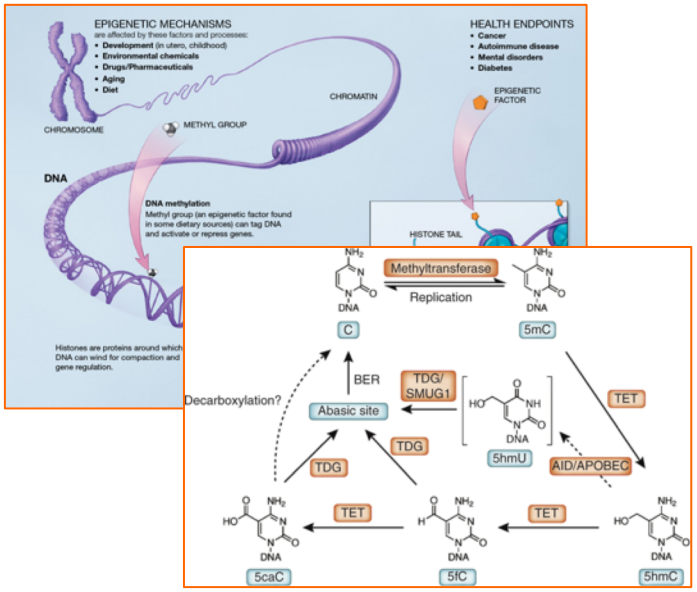
## ... to transforming drug discovery with ML & AI



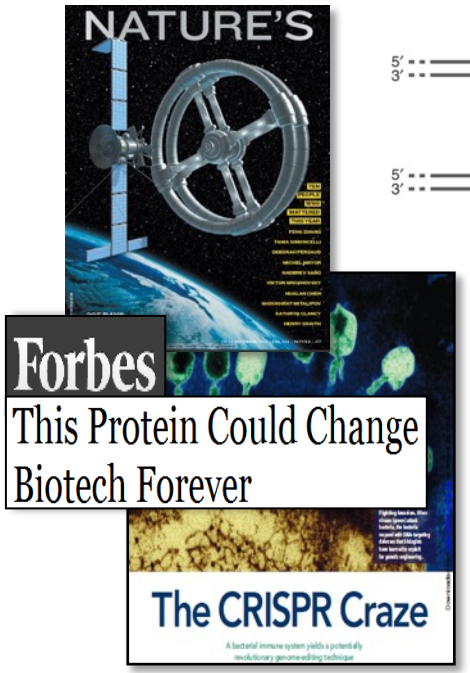
# Becoming BETTER with AI

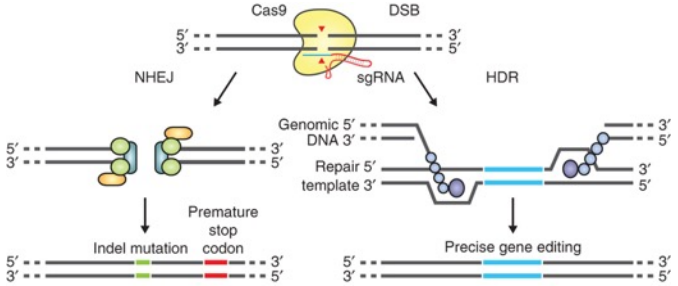
## Examples from epigenetics of Heart Failure and CRISPR gRNA design

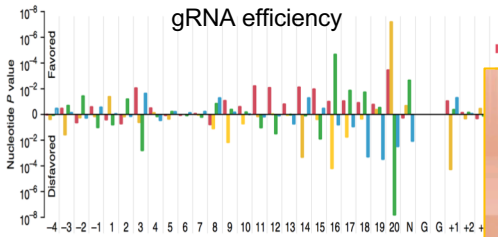


# Becoming BETTER with AI

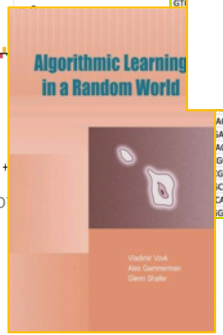
## Examples from epigenetics of Heart Failure and CRISPR gRNA design



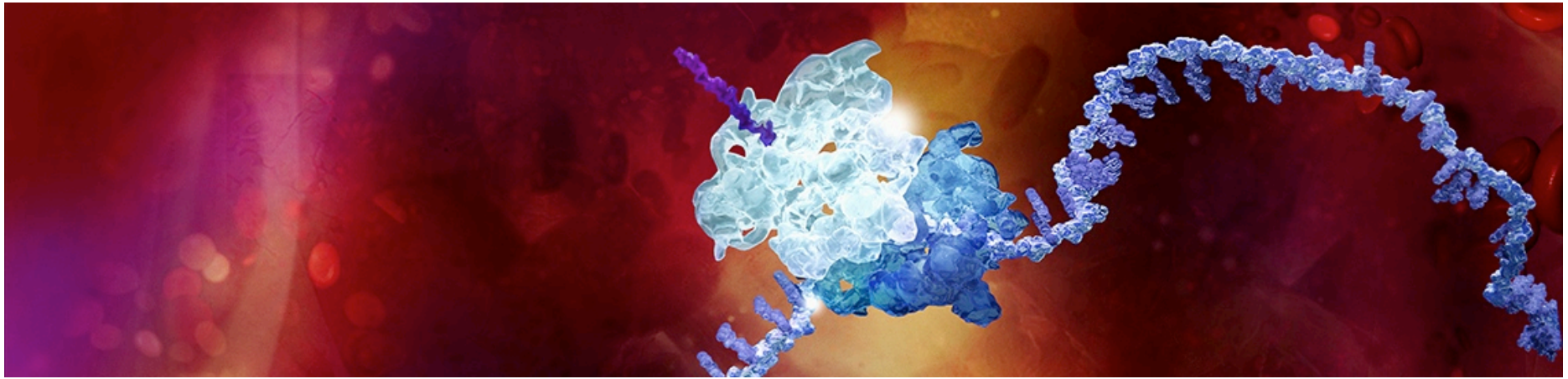




Doench et al. (2014) Nature Biotech, doi:10.1038/nb



(

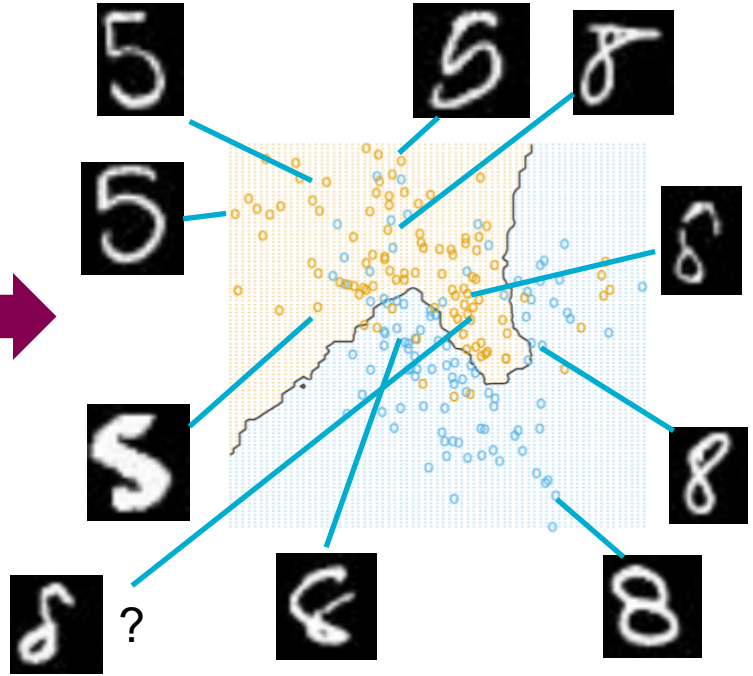
# Conformal Prediction – why do we care?



# Machine Learning - Traditional way

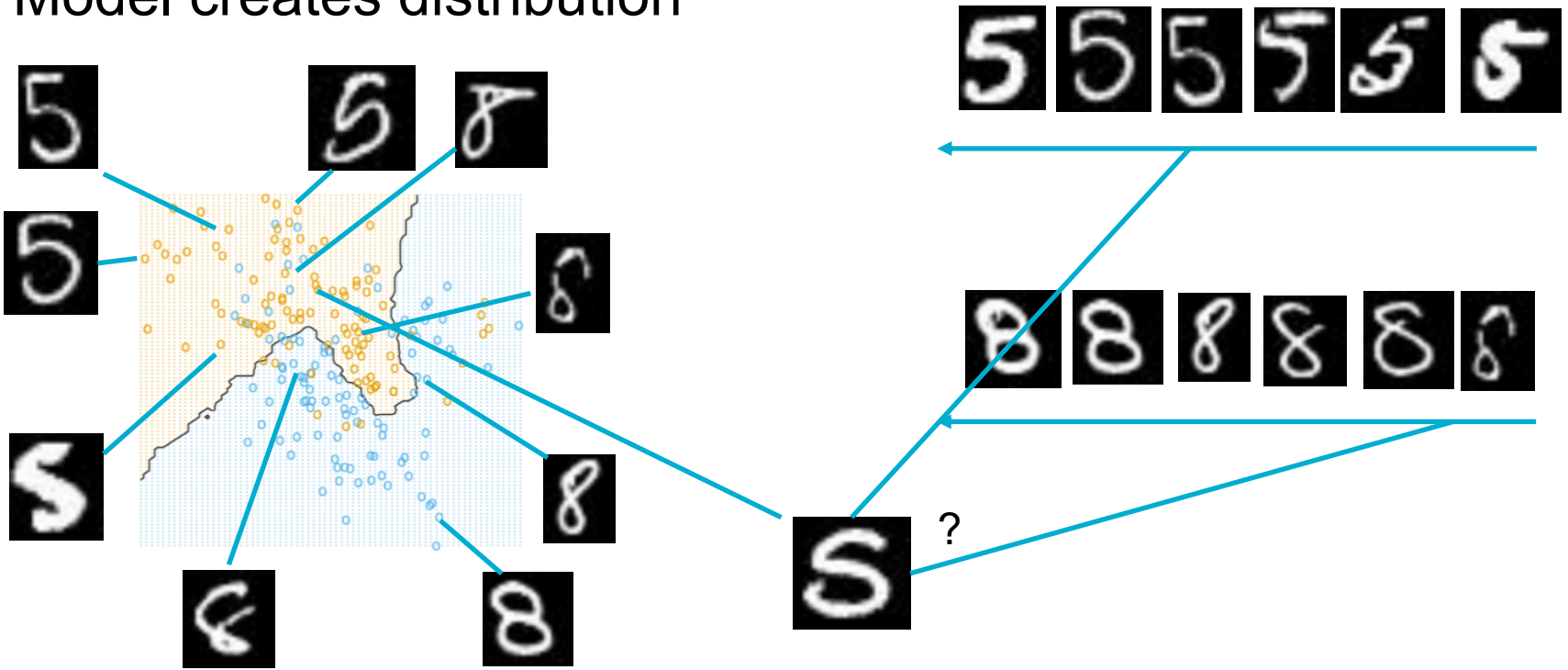
Training set

Model

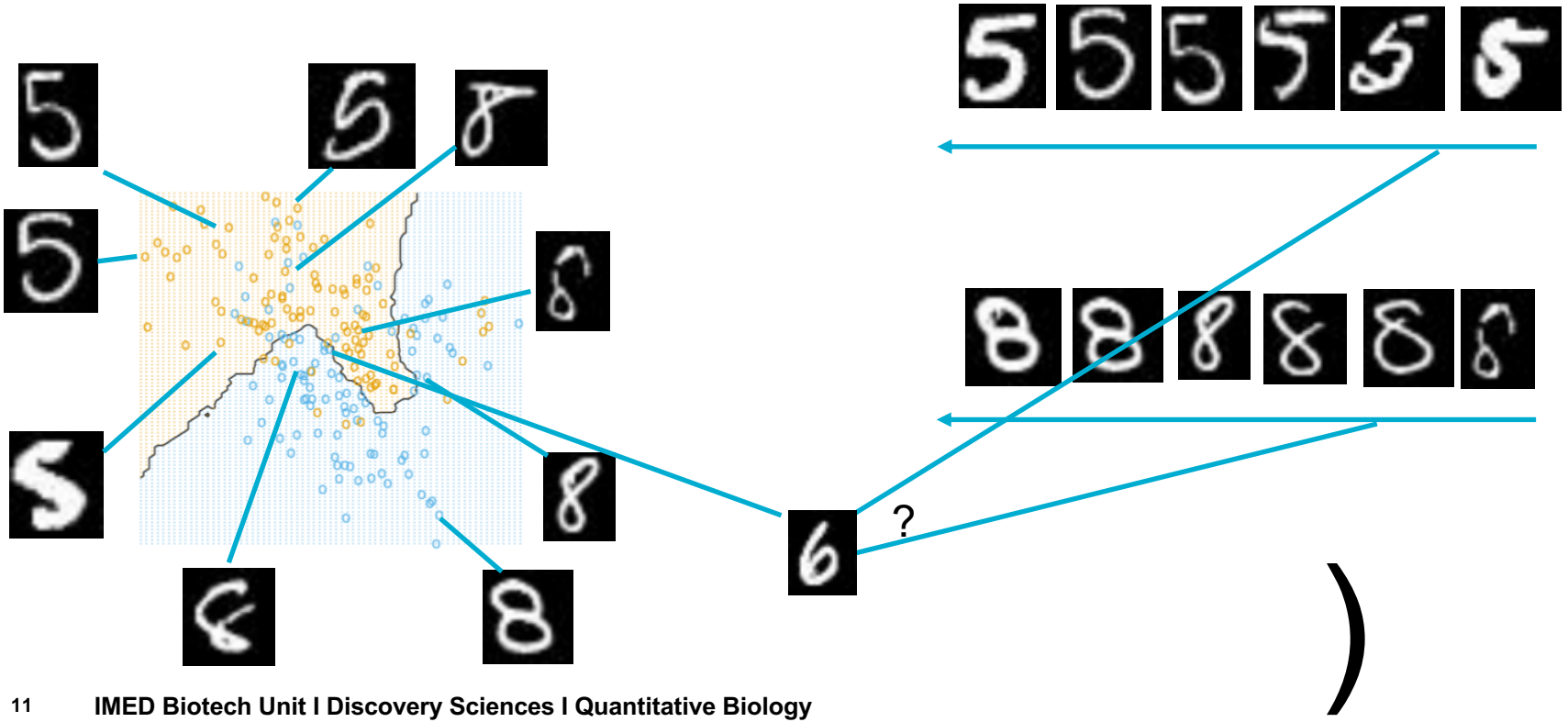


# Machine Learning – conformal prediction

## Model creates distribution



# Machine Learning – conformal prediction for the unseen

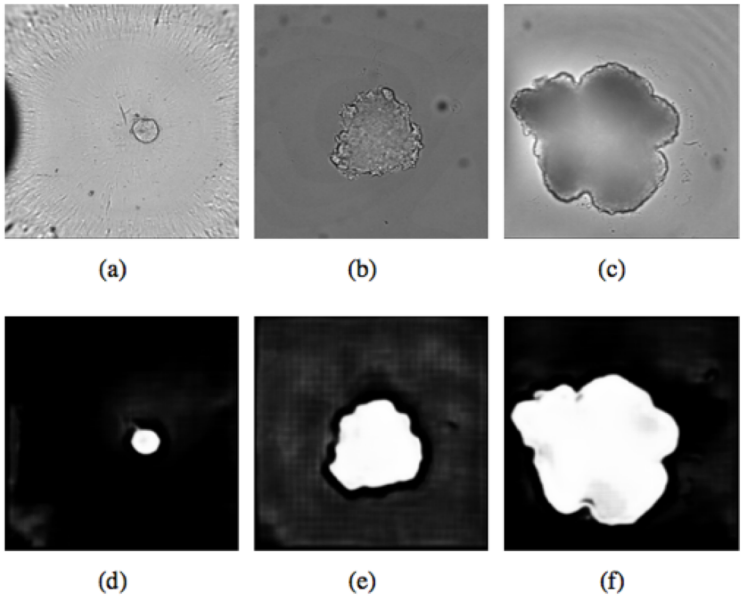


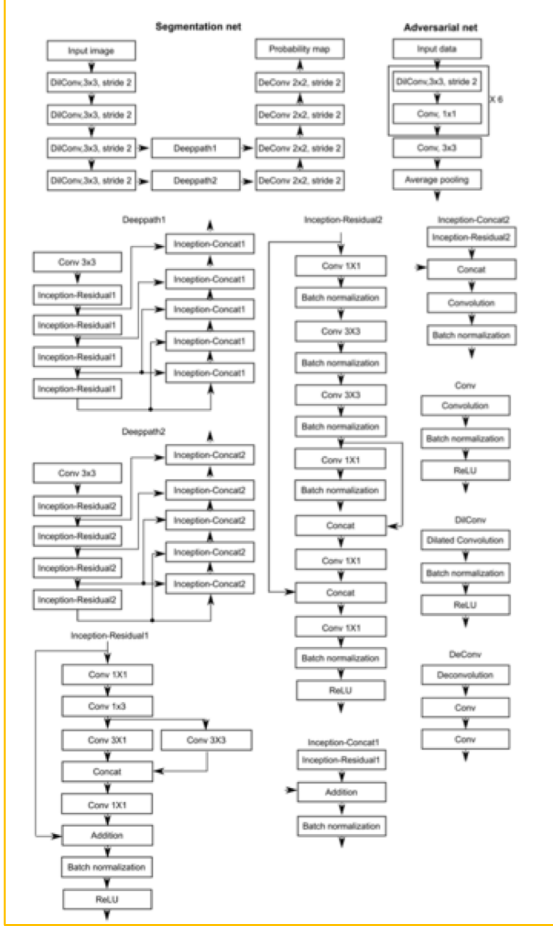
# Becoming BETTER

## Using GANs for spheroid segmentation

Variation in spheroid appearance make traditional solutions ineffective

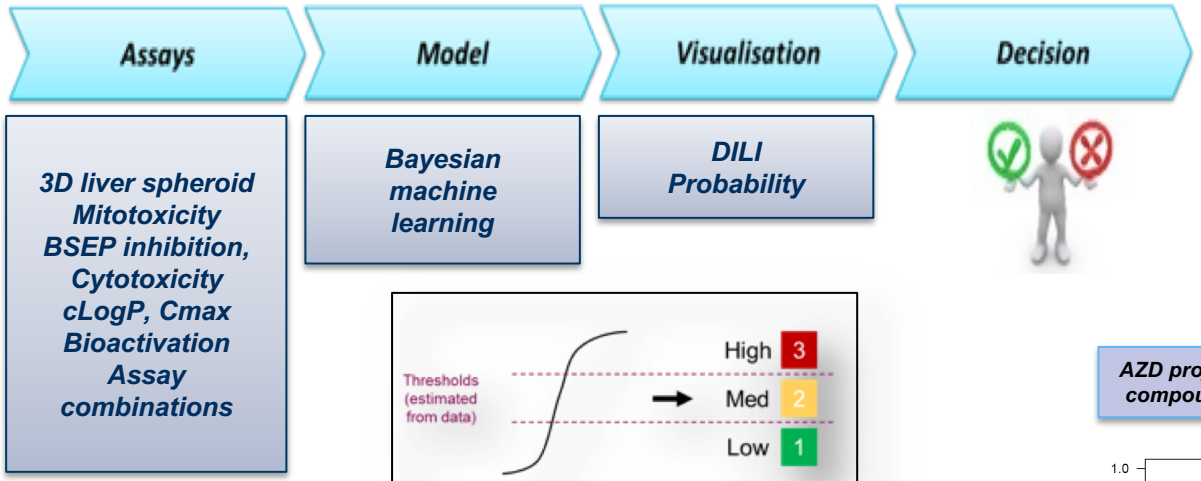
Deep networks output weighted prediction maps that can more easily be segmented



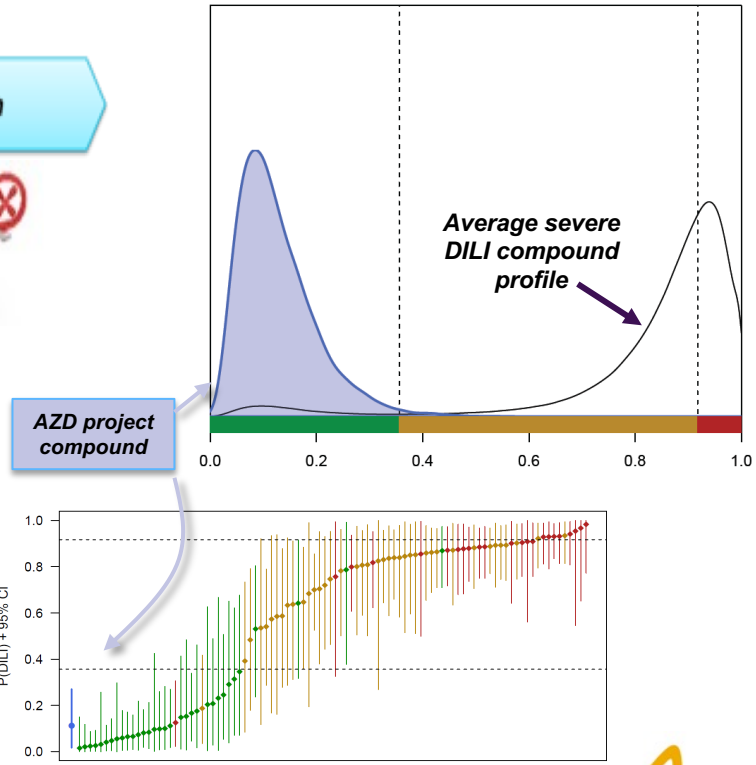
# Becoming BETTER at assessing DILI risk

## by integrating data across a multitude of assays



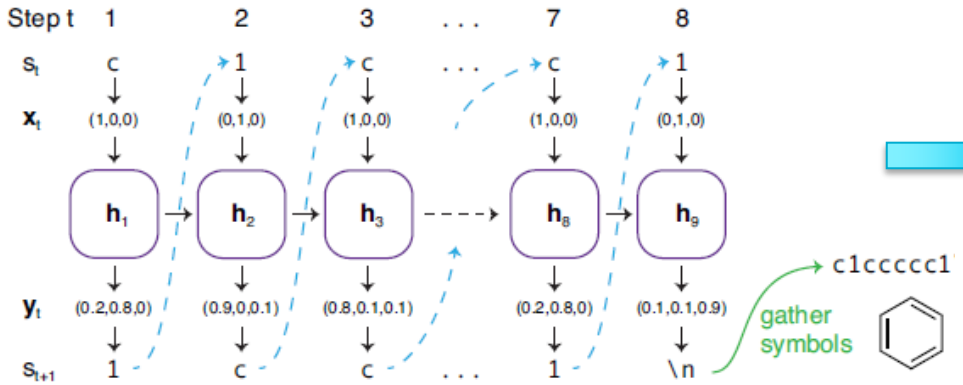


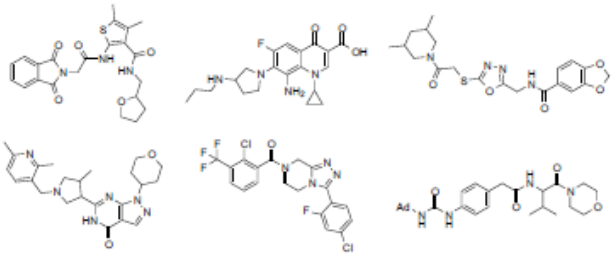


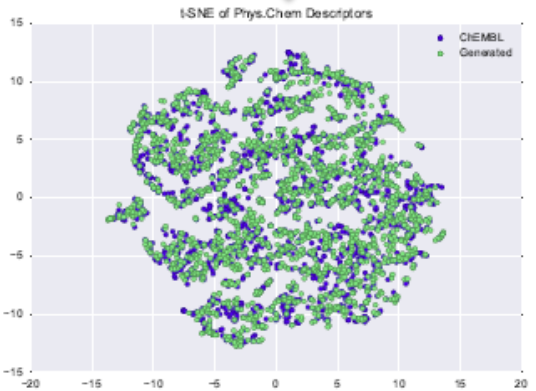


# Becoming BETTER

## through RNN based automated de-novo molecule design



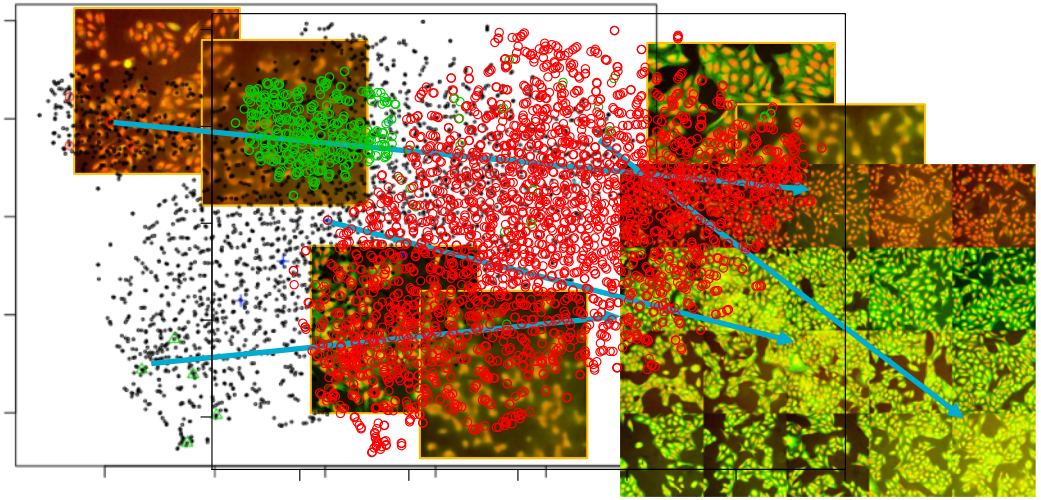


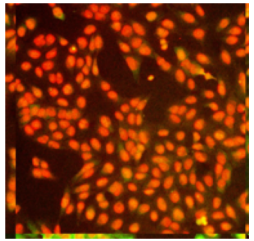

| # | $pIC_{50}$ | Train. | Test | Gen. mols. | Reprod. | EOR |
|---|---|---|---|---|---|---|
| 1 | > 8 | 1239 | 1240 | 128,256 | 28% | 66.9 |
| 2 | > 8 | 100 | 1240 | 93,721 | 7% | 19.0 |
| 3 | > 9 | 100 | 1022 | 91,034 | 11% | 35.7 |

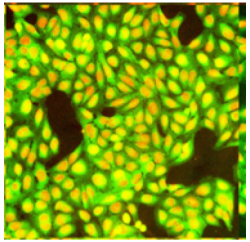
Reproducing known actives in the Plasmodium test set. EOR: Enrichment over random.

# Becoming FASTER with AI

## Through unsupervised learning for hit identification







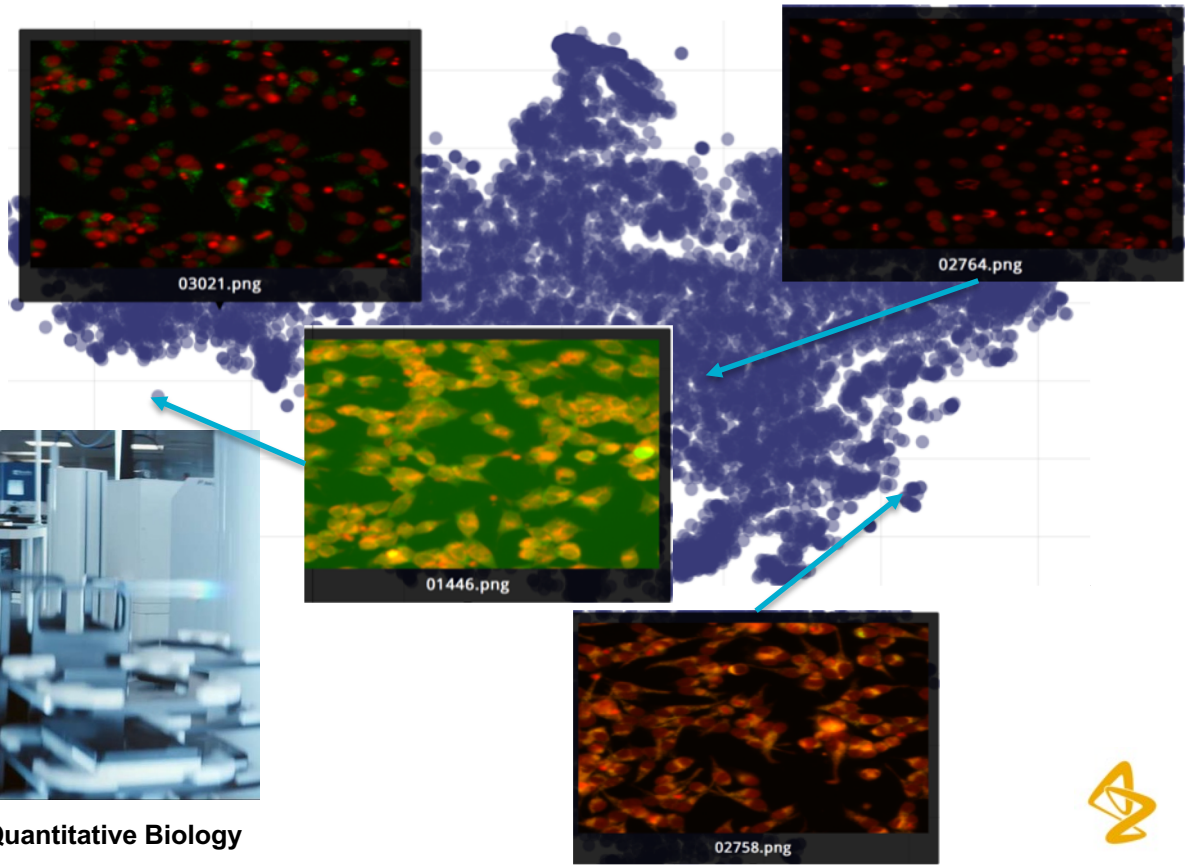
Deep CNN autoencoder → Manifold Learning (t-SNE) → Deep CNN classifier

# Becoming FASTER

## … through transfer learning for HTS

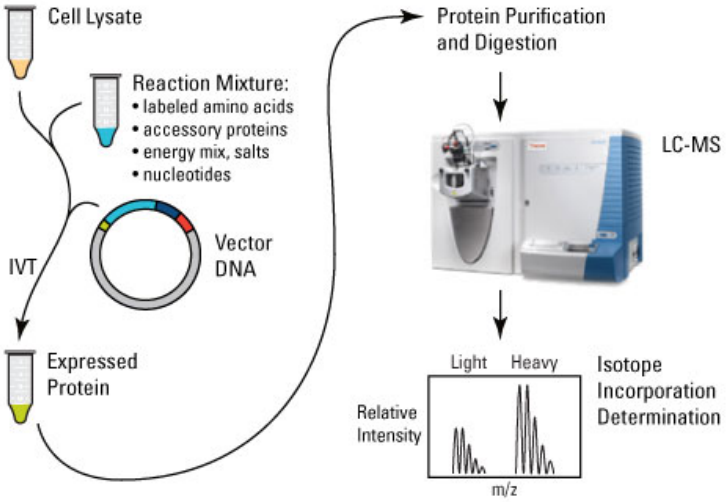
Using CNN pre-trained on general image data for microscopy to abandon screen specific training efforts

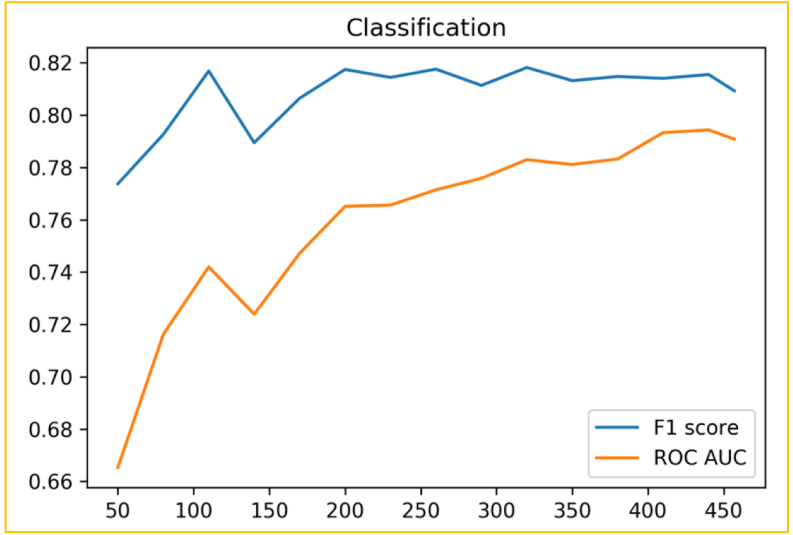
Hit ID with limited effort & identification of artifacts




# Becoming FASTER
## by optimizing vector design

- Protein production is a bottleneck in early discovery
- Multiple construct and conditions are typically explored to deliver a quality product

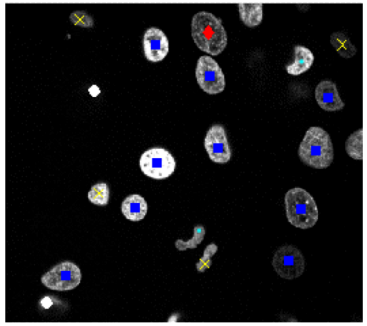



- Using ML we can predict yield from plasmid design
- Together with a recommender like system this allows our scientists to deliver more protein to time (and at lower cost)

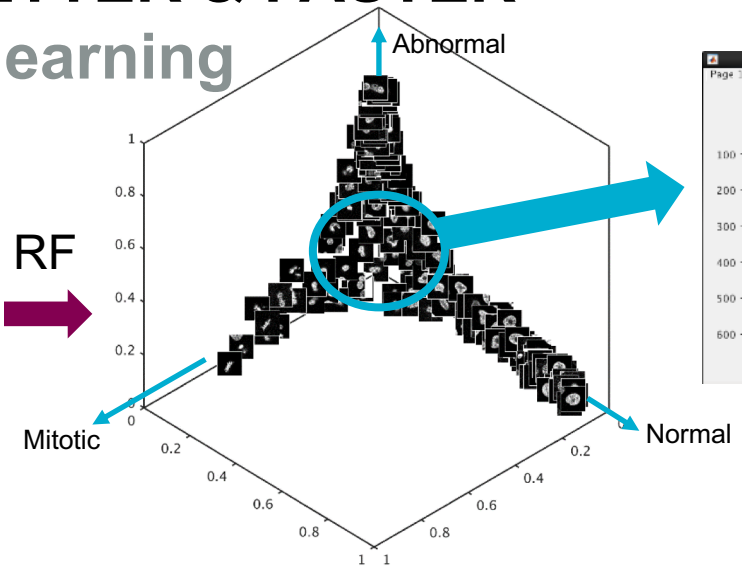
# Becoming BETTER & FASTER

## using active learning

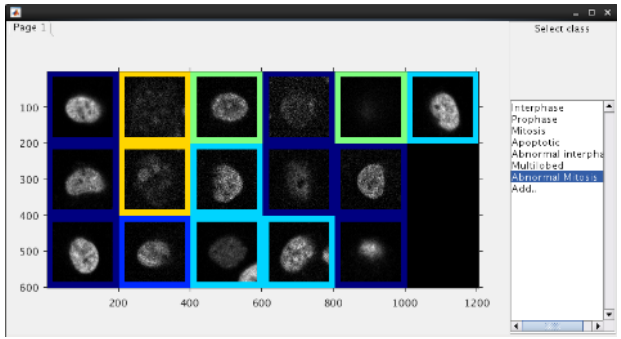


Automated cell tracking, but identification of subtle phenotypes requires time-consuming manual annotation

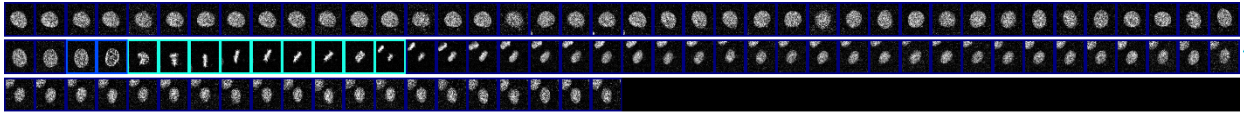
RF

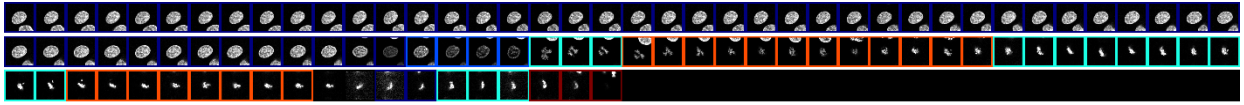



Choose the most informative cells to efficiently expand training set

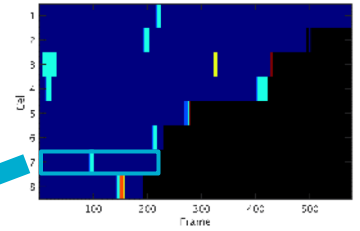


time

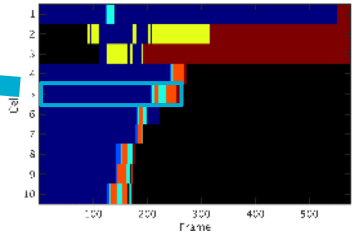


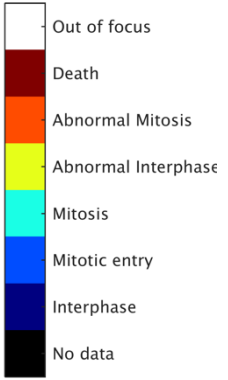


Control



Treated





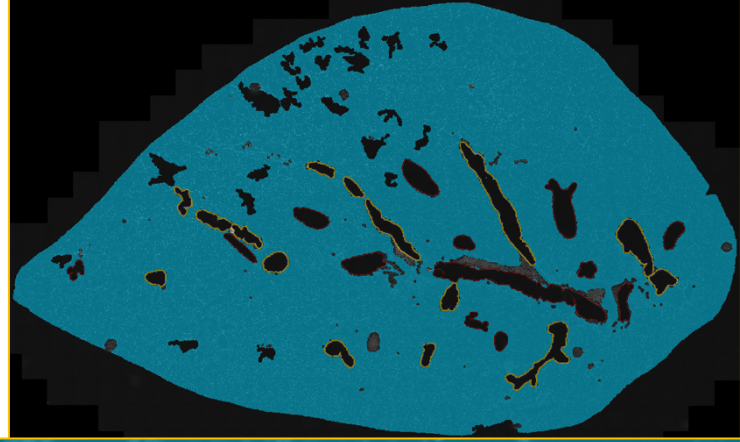
# Becoming CHEAPER

## by automating tissue segmentation

Understanding tissue distribution of inhaled therapies requires accurate physiological classification from tissue samples.

Using a combination of machine learning and rule based artificial intelligence leads to robust segmentation of fluorescence microscopy images of lung tissue.

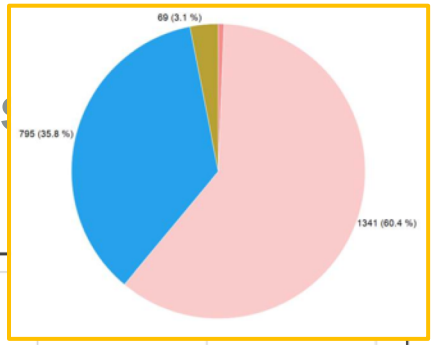
Saving resources in pathology

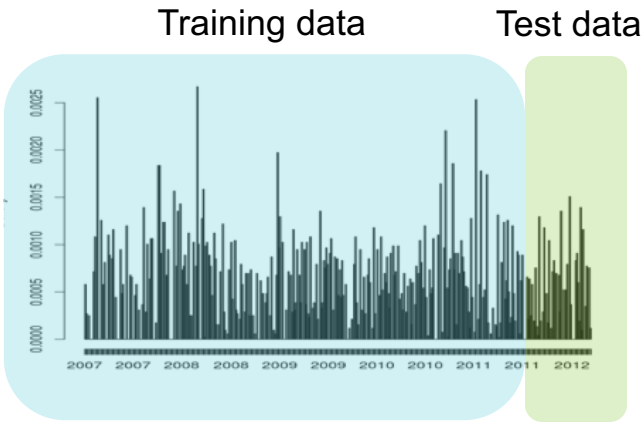


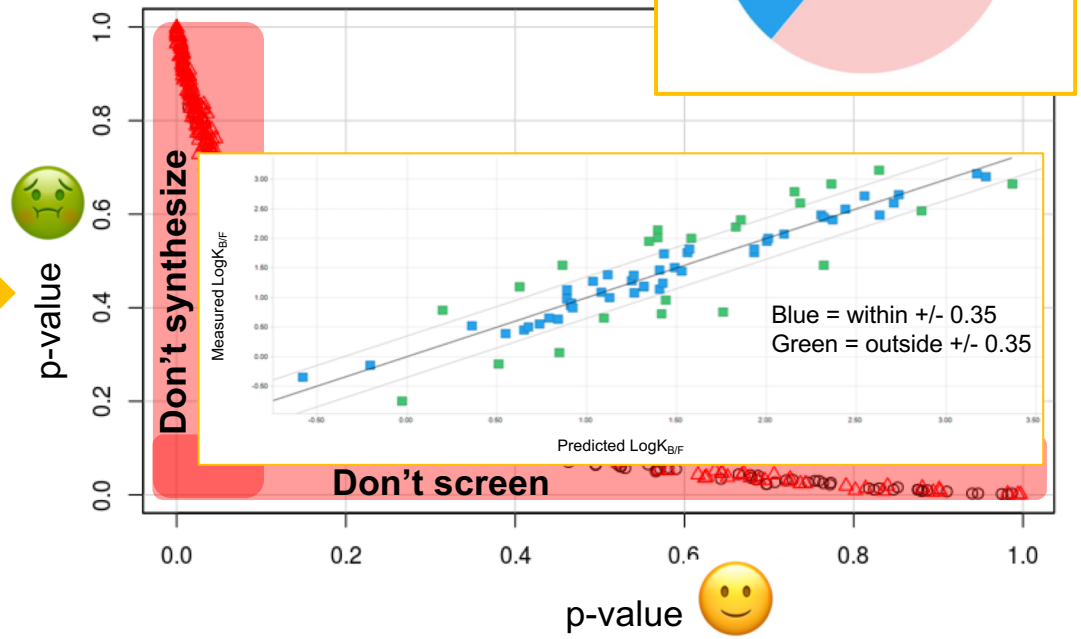
# Becoming CHEAPER

## … by focusing our chemistry synthesis efforts

## 36% Reduction in compounds screened



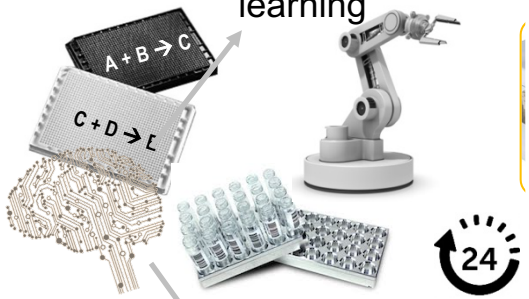



# Becoming BETTER, FASTER and CHEAPER with AI

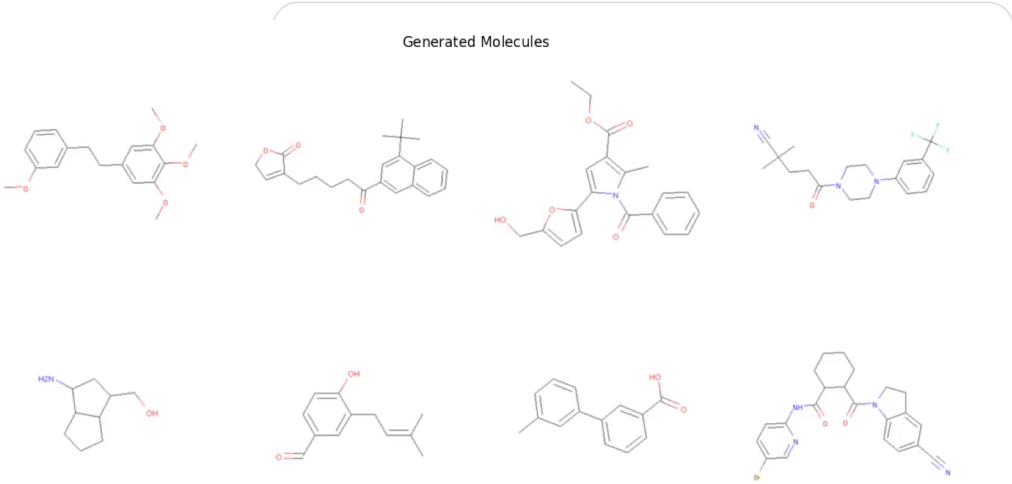
## By automating drug discovery

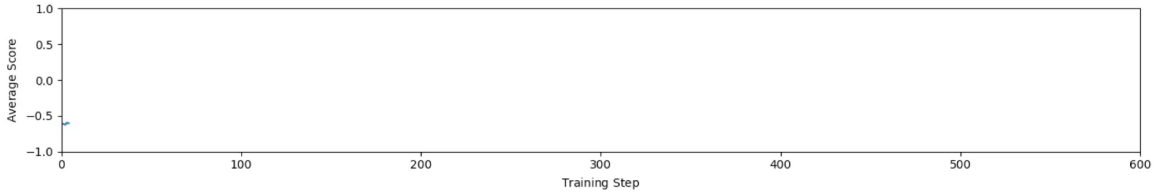
**Solution:** Structure generation and learning




*Fully automated synthesis and pur*

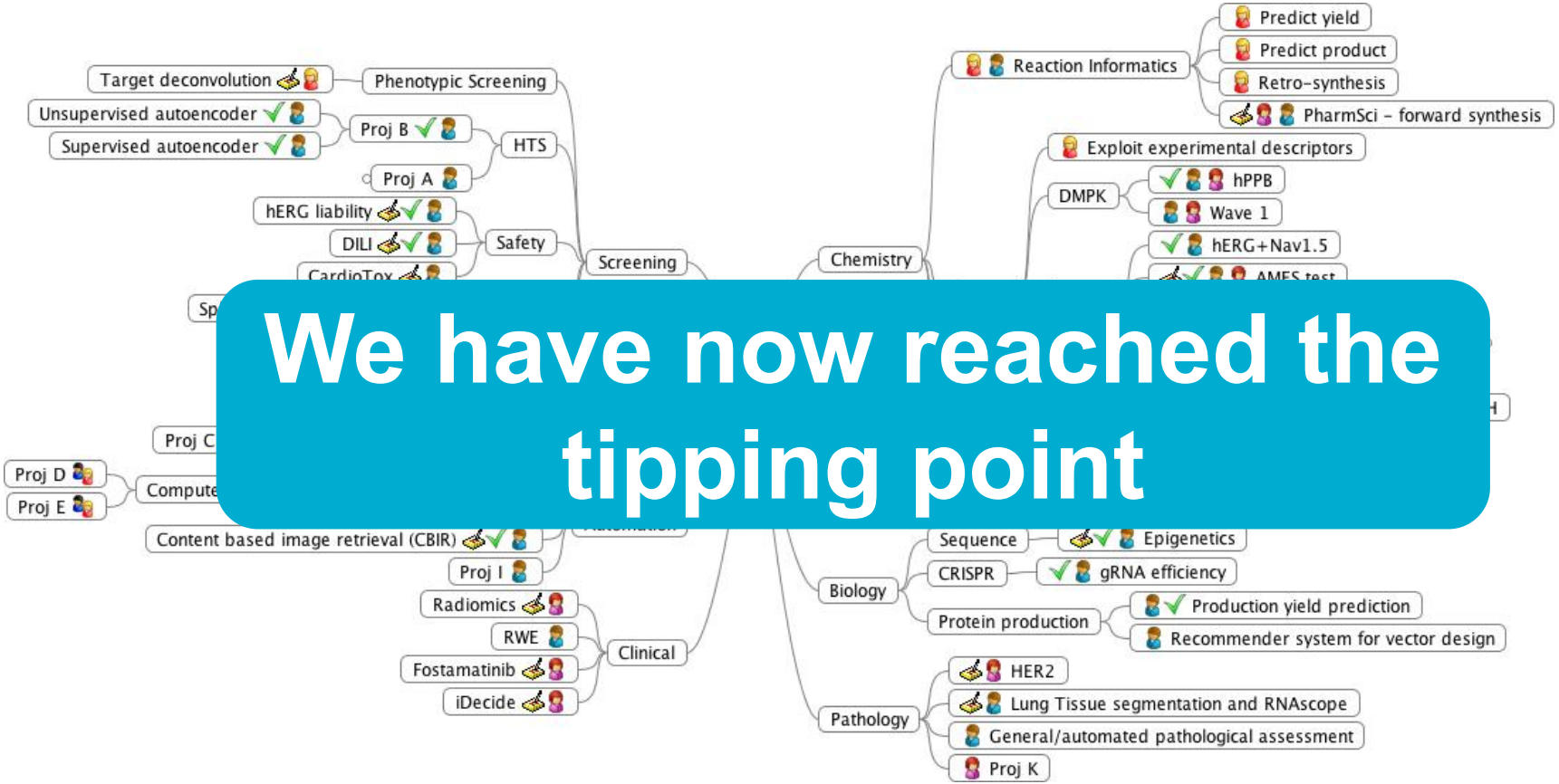
Progress towards the target profile




Olivecrona et al. https://arxiv.org/abs/1704.07555

# With a culture embracing what ML & AI can do…




**We have now reached the tipping point**

# Q&A


Acknowledgements:

- QuBi – esp. Lars Carlsson, Stan Lazic, Aurelie Bornot, Yinhai Wang, Johan Karlsson, Adam Corrigan & Mike Firth
- Ola Engkvist, Thiery Kogej & HD team
- Ralph Knöll & team
- IMED ISG team, HIT ID Futures team & Futures 2.0 team
- HASTE, DASDARD, EXCAPE, Vovk & Gammerman