Few-Shot Bioassay Prediction with Cell Painting for Drug Discovery

Son Ha

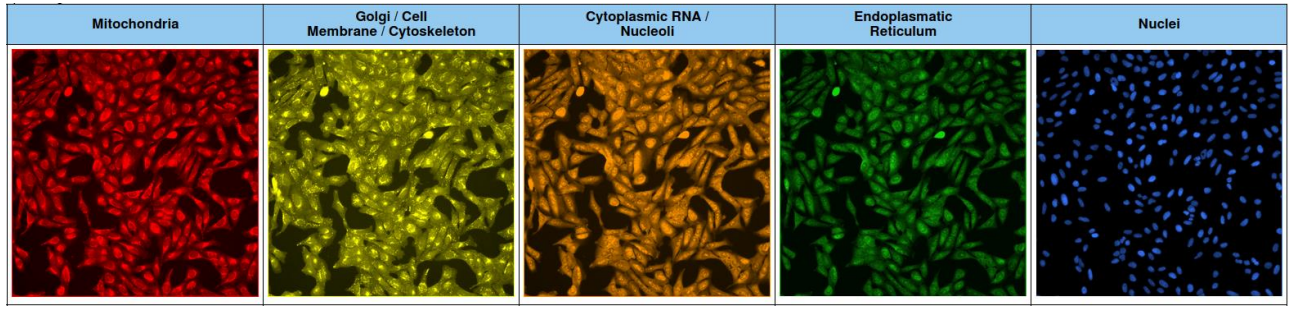
JOHANNES GUTENBERG UNIVERSITÄT MAINZ

JG|U

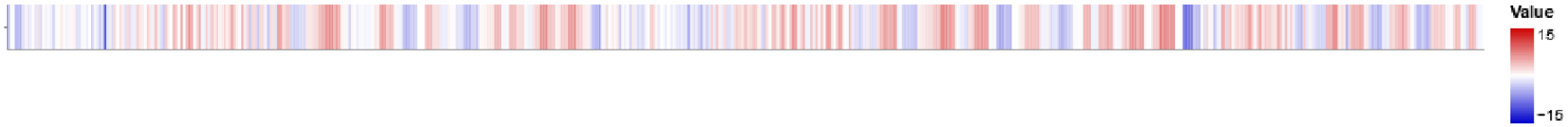
CELL PAINTING ASSAY

- Morphological profiling: Perturb cells with a desired compound. Then capture cell's **morphological changes** after being stained by different dyes.

→ Unbiased and multiparametric way to **represent compounds**.



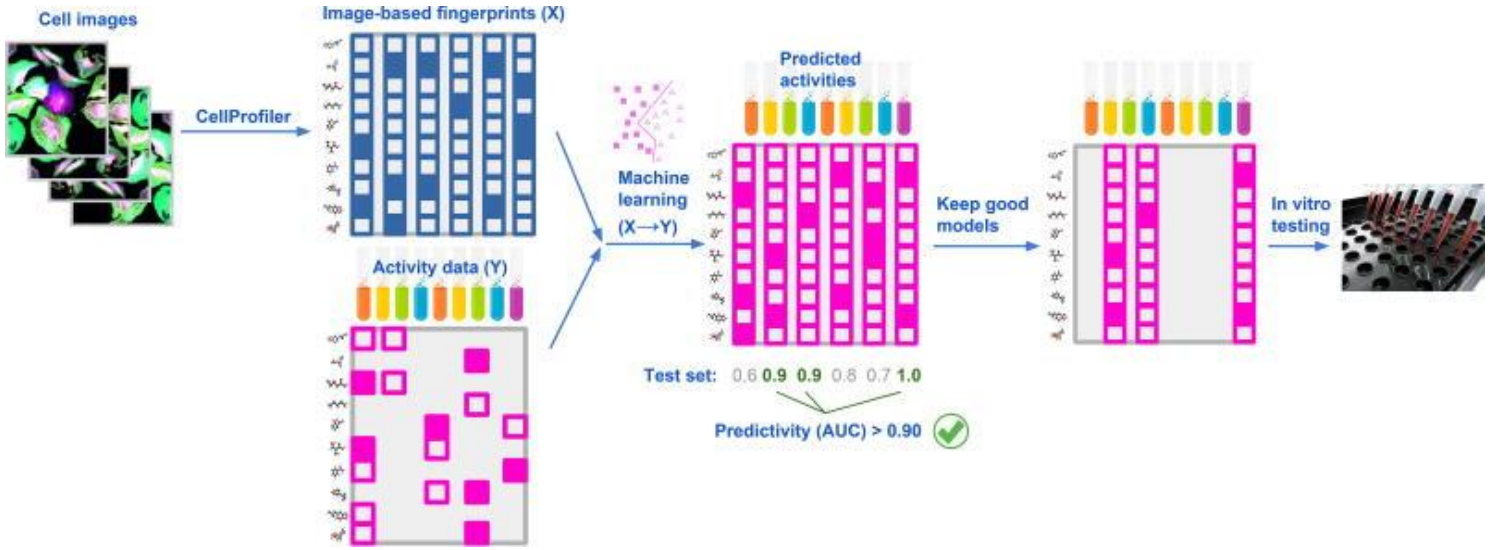
Cell Profiler extracts numerical values from images. MPI pipeline calculates morphological features per test compound



1. Bray MA, Gustafsdottir SM, Rohban MH, Singh S, Ljosa V, Sokolnicki KL, Bittker JA, Bodycombe NE, Dancík V, Hasaka TP, Hon CS, Kemp MM, Li K, Walpita D, Wawer MJ, Golub TR, Schreiber SL, Clemons PA, Shamji AF, Carpenter AE. A dataset of images and morphological profiles of 30 000 small-molecule treatments using the Cell Painting assay.

2. Aishvarya Tandon's rdkit UGM 2020 (https://github.com/rdkit/UGM_2020/blob/master/LightningTalks/AishvaryaTandon_DeepPainting_RDKitUGM2020_FlashTalk.pdf)

PREDICTION OF BIOASSAYS WITH CP



- Simm et al. (2018): Janssen collaboration project in 2 Drug Discovery campaigns.
- Pipeline that uses **features calculated from microscopy images** for **multitask prediction of bioassays**, which further helps identify hits.

→ Increased hit rates by **60- and 250-fold** over that of the initial project assays

→ Increased chemical **structure diversity** of the hits

1. Simm J, Klambauer G, Arany A, Steijaert M, Wegner JK, Gustin E, Chupakhin V, Chong YT, Vialard J, Buijnsters P, Velter I, Vapirev A, Singh S, Carpenter AE, Wuyts R, Hochreiter S, Moreau Y, Ceulemans H. Repurposing High-Throughput Image Assays Enables Biological Activity Prediction for Drug Discovery

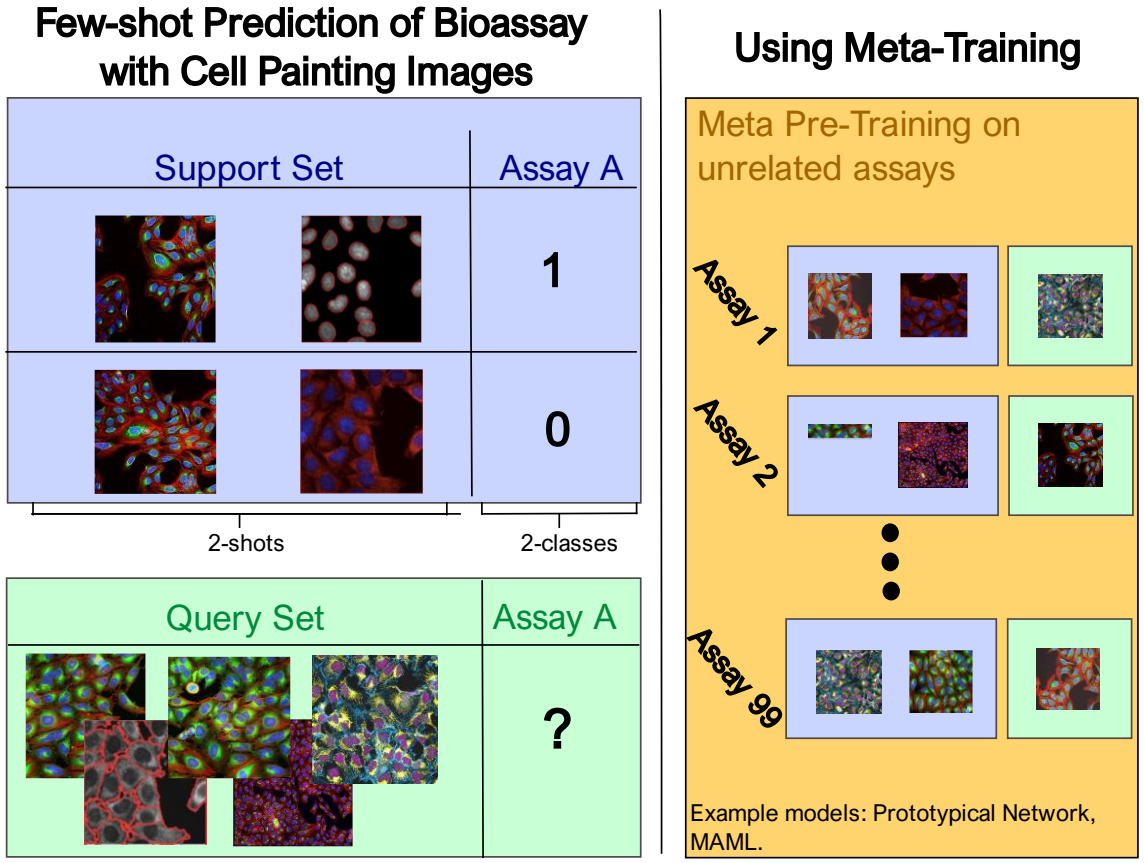
CP IN FEW-SHOT LEARNING

- Prediction of bioassays is an important tasks for Drug Discovery.
- However, it is common that there are only few data points available for each bioassay.

→ **Few-shot-learning**: Deep Learning in a low-data regime.

- Idea: Pre-train the model with a large dataset of **unrelated assays**. That knowledge can be **transferred gainfully** to the prediction of our desired low-data assay.

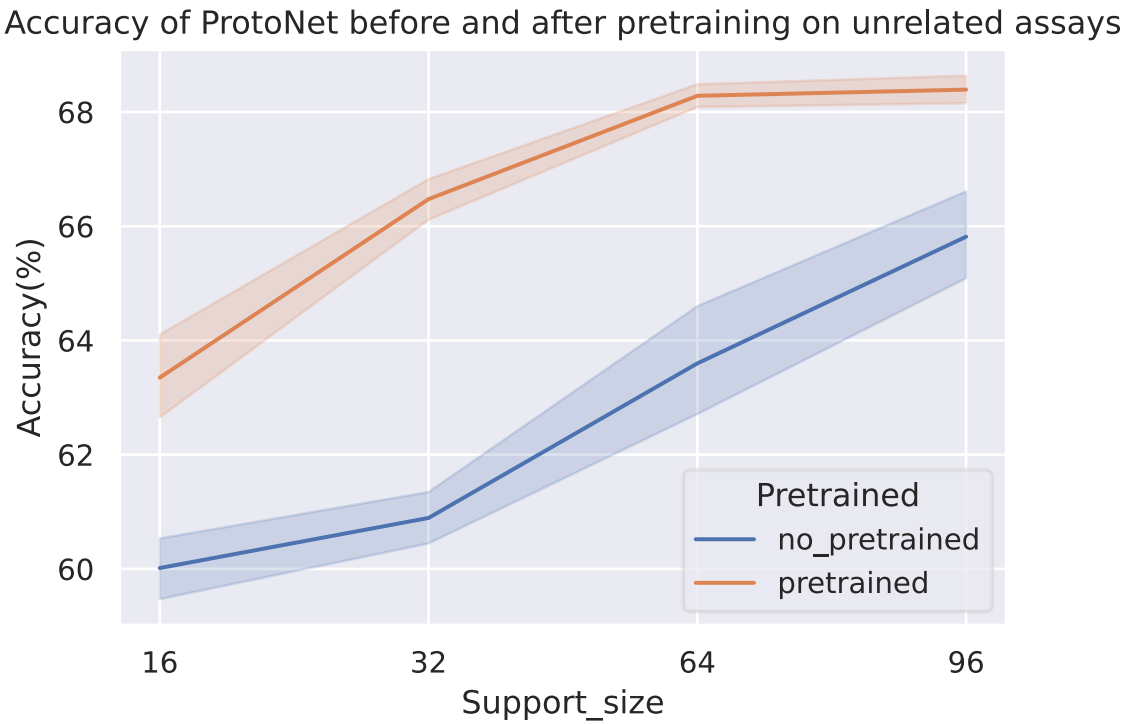
- *Will this work?*
- *Comparison of different fsl methods?*



TRANSFER OF KNOWLEDGE FROM UNRELATED ASSAYS

- Task: Prediction of one bioassay (CHEMBL1794311) using CP features.
- Accuracies of this model are recorded as size of the support set increases.
- Model: Prototypical Network (Meta-Learning family)

- At different support set (available data) size, pretraining on unrelated assays results in an increase in accuracy.



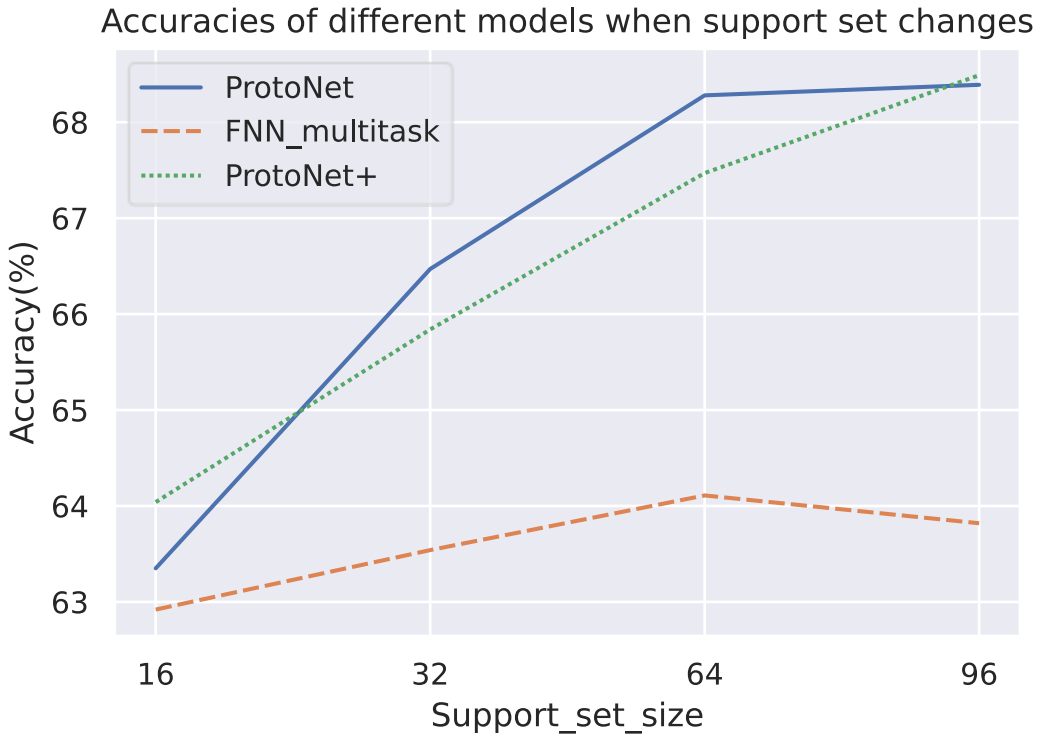
COMPARISON OF DIFFERENT FSL METHODS

- Task: Prediction of one bioassay (CHEMBL1794311) using CP features.
- Accuracies of 3 models are recorded as size of the support set increases.
- Models:

  Prototypical Network (Meta-Learning family)

  Fully connected NN (Multitask)

→ ProtoNet from the Meta-Learning family outperforms multitask training with fully connected neural network for FSL on this assay.



Thank you for listening!

JOHANNES GUTENBERG
UNIVERSITÄT MAINZ

JG|U

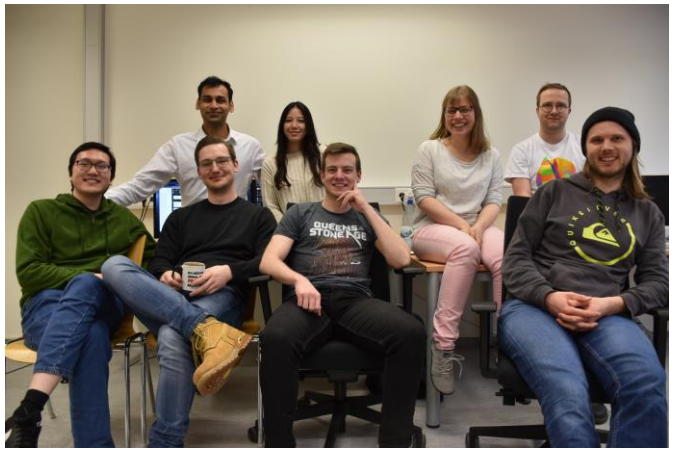
ACKNOWLEDGEMENTS

Supervisors:

Prof. Dr. Paul Czodrowski, Dr. Steffen Jaensch, Dr. Lorena Freitas

My colleagues at:

- CzodrowskiLab
- Max-Planck Institute of Molecular Physiology
- Janssen Pharmaceutical
- Lab of Assist.-Prof. Mag. Dr. Günter Klambauer
- AIDD Project



MAX-PLANCK-INSTITUT FÜR MOLEKULARE PHYSIOLOGIE

JKU JOHANNES KEPLER UNIVERSITÄT LINZ

Janssen

COMPARISON OF DIFFERENT FSL METHODS

- Multiple runs are repeated and all of the accuracies are recorded
- Boxplot provides a closer look at the interquartile range of all accuracies for prediction of this assay.

→ Multitask model's accuracies are more spread out that those of 2 ProtoNet variants.