technische universität dortmund

MAX PLANCK INSTITUTE OF MOLECULAR PHYSIOLOGY
mpi dortmund

Deep Painting

an amalgamation of Cell Painting and Machine Learning

Aishvarya Tandon, PhD student
CzodrowskiLab | Waldmann Group

- aishvarya.tandon@tu-dortmund.de
- aishvarya.tandon@mpi-dortmund.mpg.de

RDKit
Open-Source Cheminformatics and Machine Learning

RDKit UGM 2020
6th to 8th October 2020



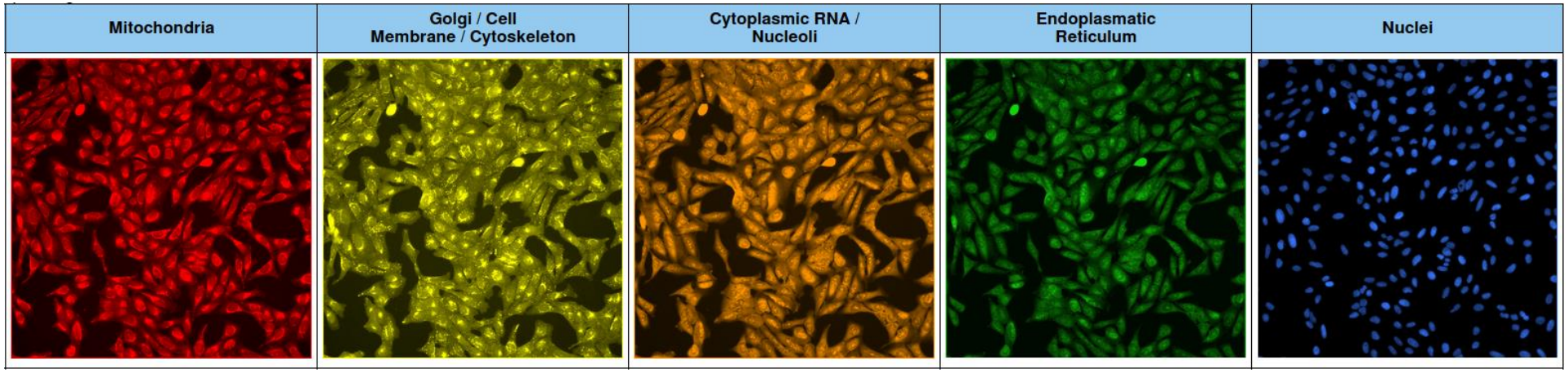


A quick introduction to Cell Painting Assay

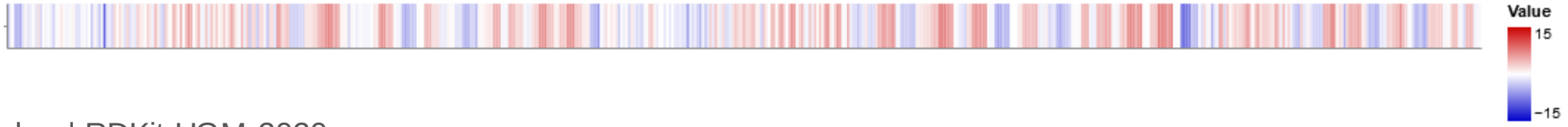
- Developed by the Carpenter group from Broad Institute (Bray et. al 2016)
- This assay involves staining of different cellular components by different dyes with distinct excitation-emission range
- Result: an unbiased, multiparametric, high-throughput and high-resolution image-based cell assay where changes in cells' morphology caused by different treatments are quantitatively recorded

Good to read: Bray, M., Singh, S., Han, H. *et al.* Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat Protoc* **11,** 1757–1774 (2016).

The MPI Cell Painting Dataset



CellProfiler extracts numerical features from microscope images, MPI pipeline calculates morphological features per test compound.



The MPI Cell Painting Dataset: About

- 25,000 measurements of 14,000 compounds
- 4000 reference compounds, 10,000 internal research compounds

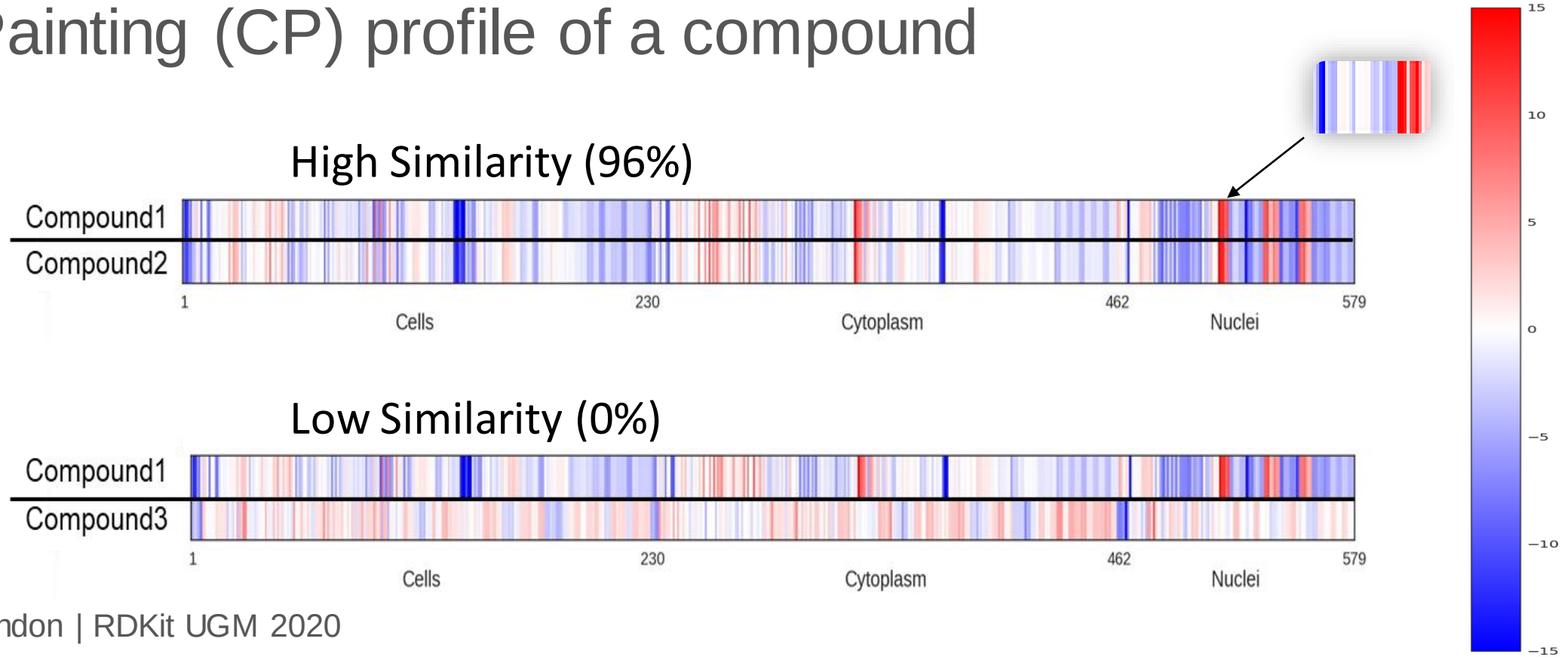
Compound_Id	Container_Id	Well_Id	Producer	Conc_uM	Is_Ref	Activity	Toxic	Pure_Flag	Rel_Cell_Count	...	Median_Nuclei_Texture_SumVariance_Ph_golgi_10_00	Median_Nuclei_Texture_SumVariance_Ph_golgi_3_00
394146	394146:01:08	394146:01:08_10.00	PAHL	10.0	False	38.3	False	Ok	87	...	-0.423	-0.636
393986	393986:01:09	393986:01:09_10.00	PAHL	10.0	False	13.6	False	Ok	97	...	0.556	0.312
406439	406439:01:03	406439:01:03_10.00	PAHL	10.0	False	12.1	False	Ok	107	...	1.123	0.761
406430	406430:01:04	406430:01:04_10.00	PAHL	10.0	False	7.9	False	Ok	104	...	2.020	1.071
406435	406435:01:03	406435:01:03_10.00	PAHL	10.0	False	6.4	False	Ok	103	...	0.621	-0.216
393982	393982:01:10	393982:01:10_10.00	PAHL	10.0	False	6.2	False	Ok	96	...	1.107	0.628
406437	406437:01:04	406437:01:04_10.00	PAHL	10.0	False	5.0	False	Ok	110	...	-1.030	-0.897

Metadata (data about data)

579 most relevant Cell Painting features

The MPI Cell Painting Dataset: CP Profile

Heatmaps are the most convenient way to visualize the entire Cell Painting (CP) profile of a compound



Ongoing Project: Deep Painting

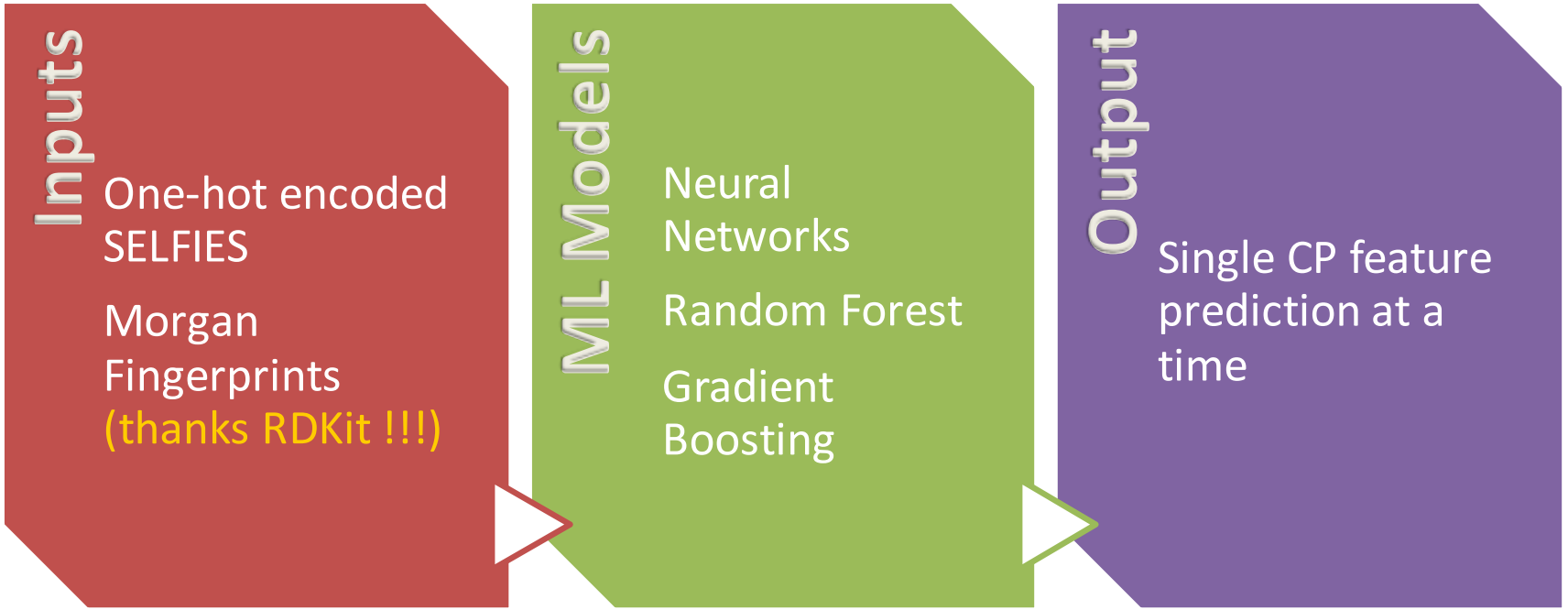
- Develop machine learning models which can help predicting the CP profile of a compound

(Few) Applications of the Cell Painting assay

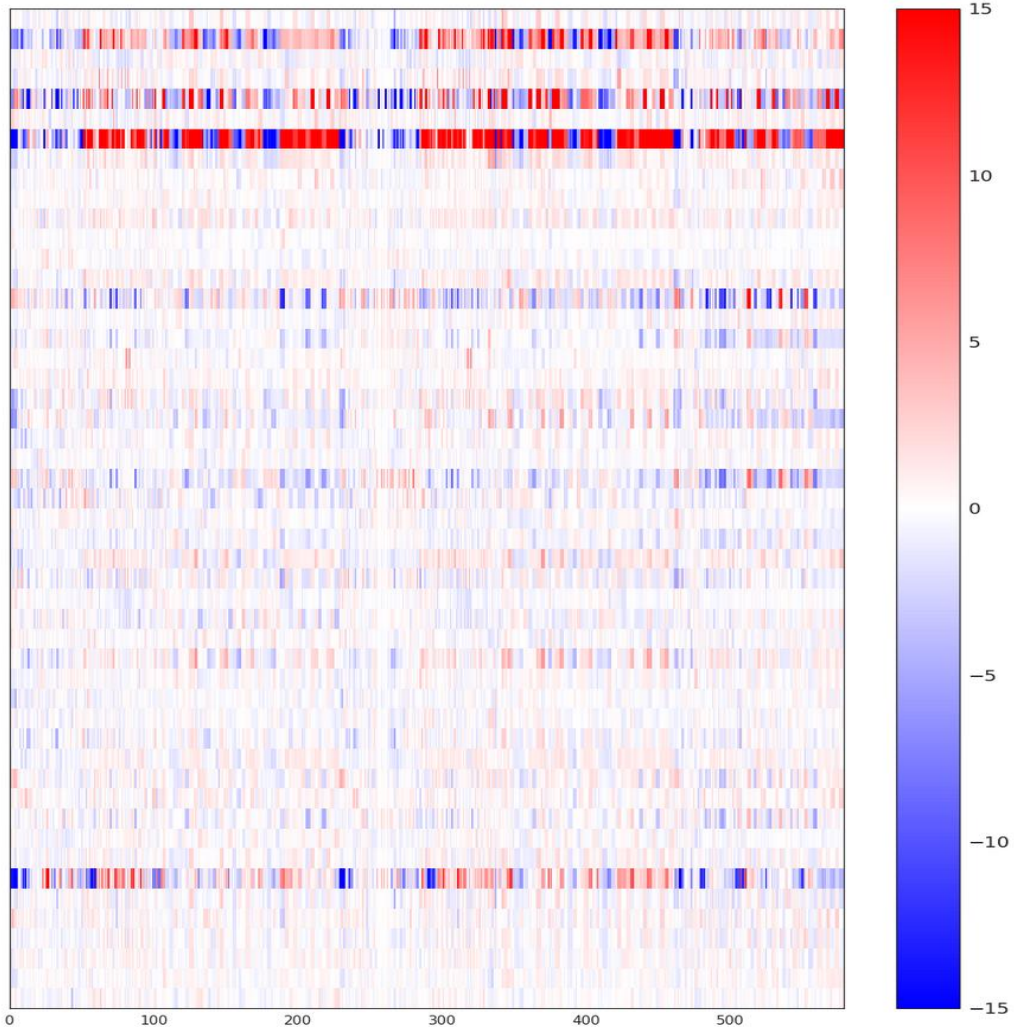
- Performing various types of clustering
- Exploring compounds with similar CP profiles
- Bioactivity prediction
- Exploration of chemical space using CP profiles

for deep learning enthusiasts
;)

Models

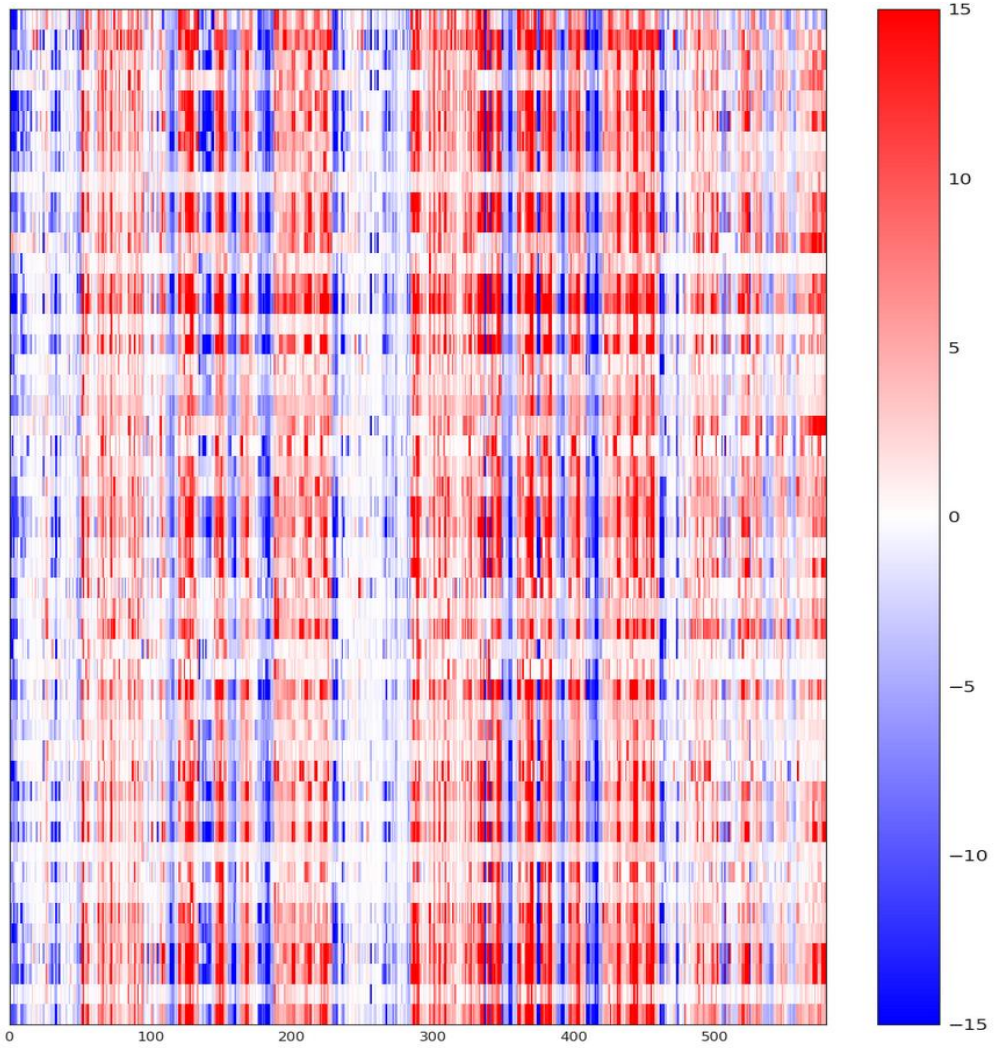

Inputs
One-hot encoded SELFIES
Morgan Fingerprints (thanks RDKit !!!)
ML Models
Neural Networks
Random Forest
Gradient Boosting
Output
Single CP feature prediction at a time

Original



CP Profile Prediction (so far)

Predicted



Outlook

Prediction of Cell Painting profiles is challenging:

- Biochemical assay
- Most of the compounds have low-induction value

technische universität dortmund

Acknowledgements

My supervisors:

Prof. Dr. Paul Czodrowski and Prof. Dr. Herbert Waldmann

My mentors:

Dr. Axel Pahl

Dr. Jose-Manuel Gally

My colleagues at:

CzodrowskiLab

Max-Planck Institute of Molecular Physiology

COMAS

Special thanks to:

Dr. Michael Beck