

# Morphological Data Analysis: From Descriptor Development to Predictive Modelling

Floriane Odje, Lisa-Marie Rolli, Andrea Volkamer

Data Driven Drug Design, Center for Bioinformatics, Saarland University



UNIVERSITÄT  
DES  
SAARLANDES

## Cell painting (CP) assay in a nutshell<sup>[1,2]</sup>

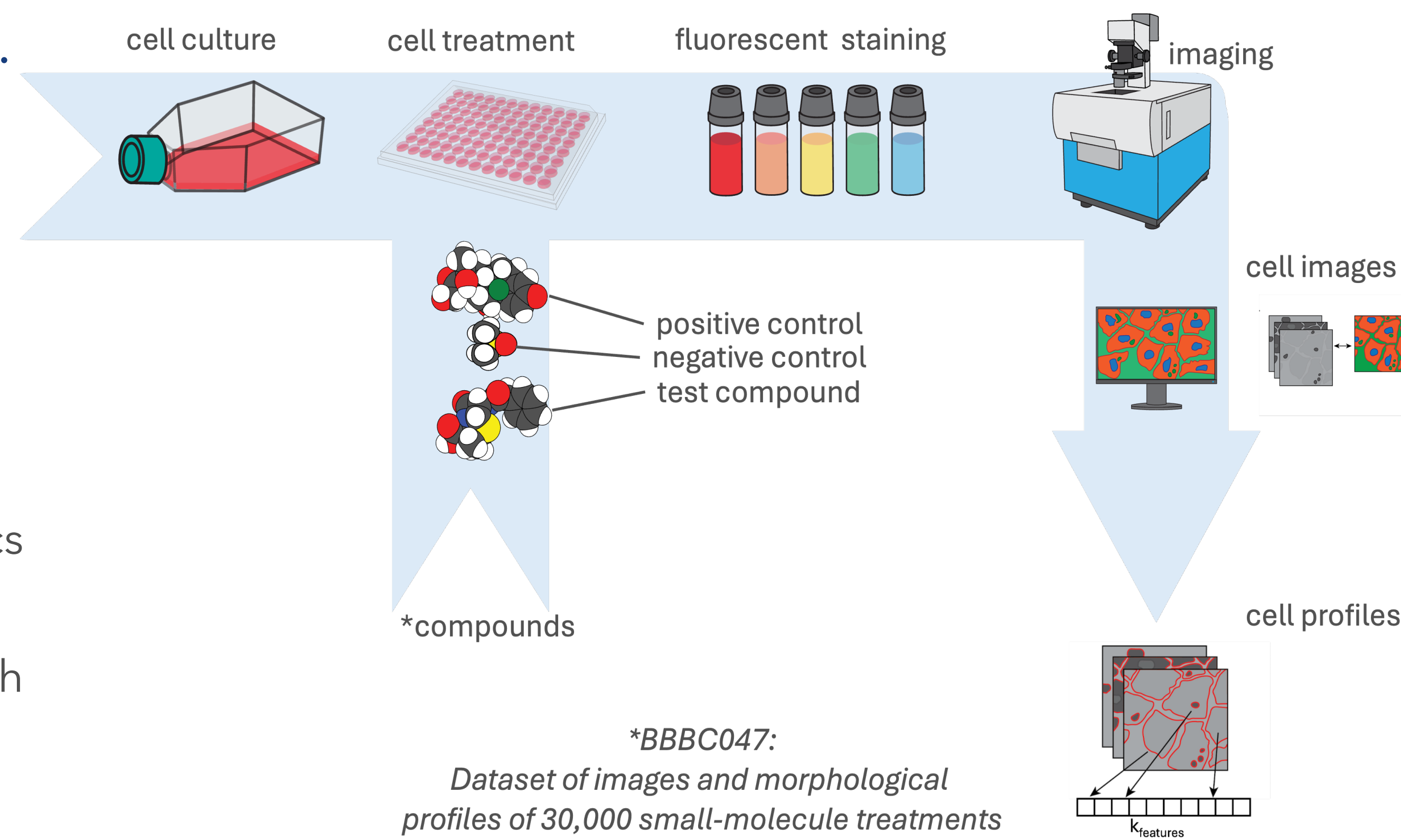
CP an image-based, high-throughput, phenotypic profiling method.

CP assay is a **powerful tool**, enabling rapid and **detailed insights** into molecular mechanisms **at the single-cell level**. Combined with machine learning and deep learning, CP can **help predict** compound activities and identify **modes of action**, **reducing** the need for **animal testing**.

**Process:** Capture detailed images of stained cells. Analyze cellular characteristics such as: size, spatial organization, and shape (phenotype).<sup>[3]</sup>

**Advantages:** Provides a comprehensive description of cells treated with compounds, can process many samples quickly (up to a full plate per day).

**Applications:** Drug discovery, toxicology, understanding cellular function.



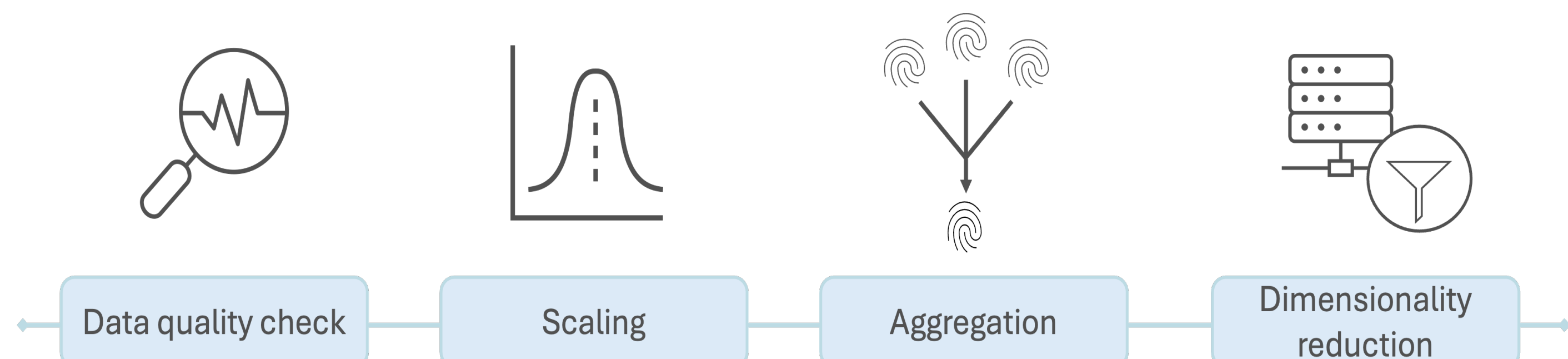
## Data Analysis<sup>[2]</sup>: Processing morphological fingerprint and Tox21<sup>[4]</sup> NR assays

### Dealing with outliers and different scales:

Outliers arise from lab environment, manipulation errors, cell death. Scaling ensures fair contribution of features.

**Aggregation:** From single cell to well to treatment level profile.

**Dimensionality reduction of profile features:** Keep the most relevant features for analysis.

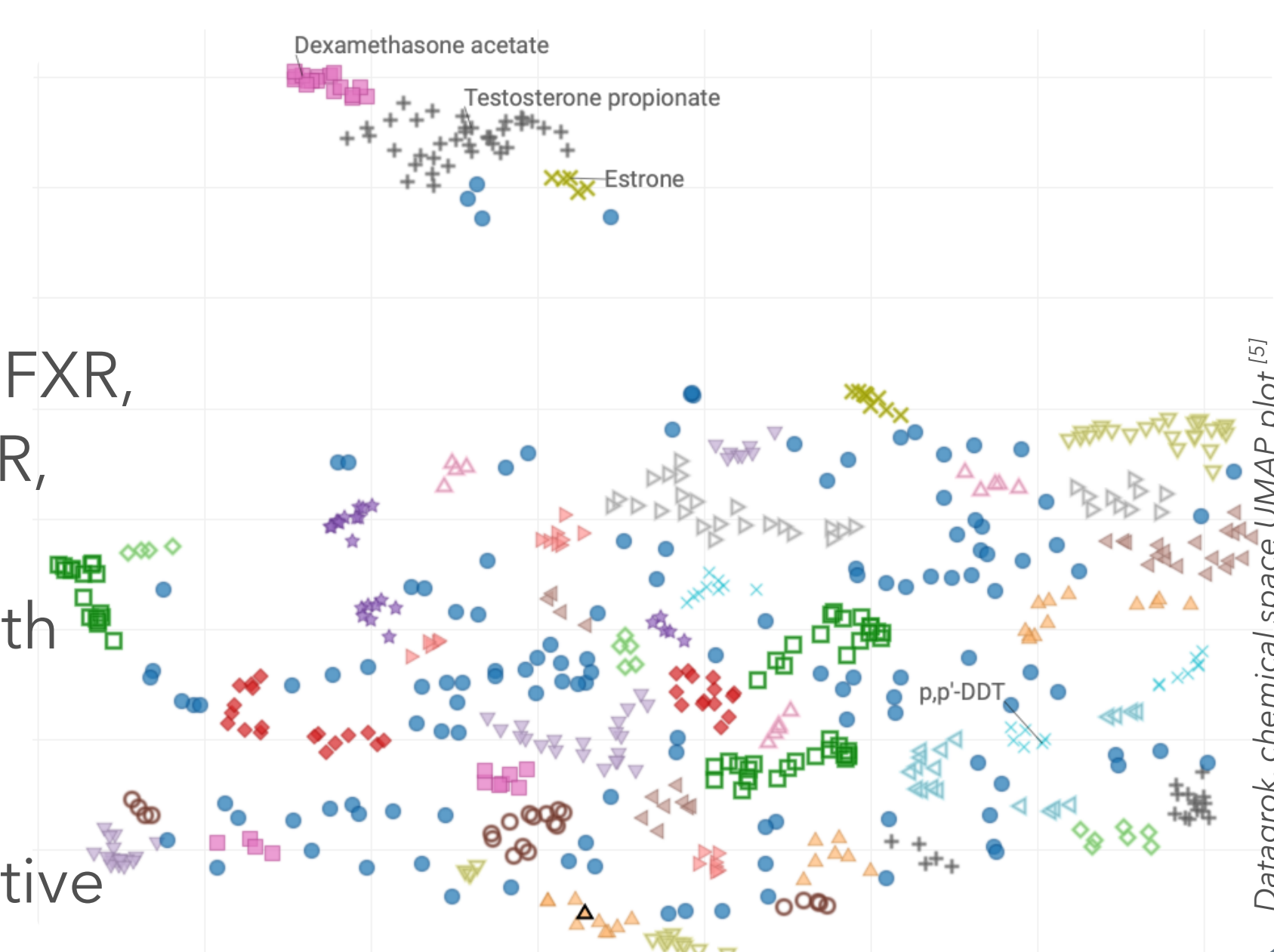


### Curation of Tox21 Nuclear Receptor (NR) assays for morphological fingerprint annotation:

A total of 8,948 molecules targeting approximately 14 nuclear receptors.

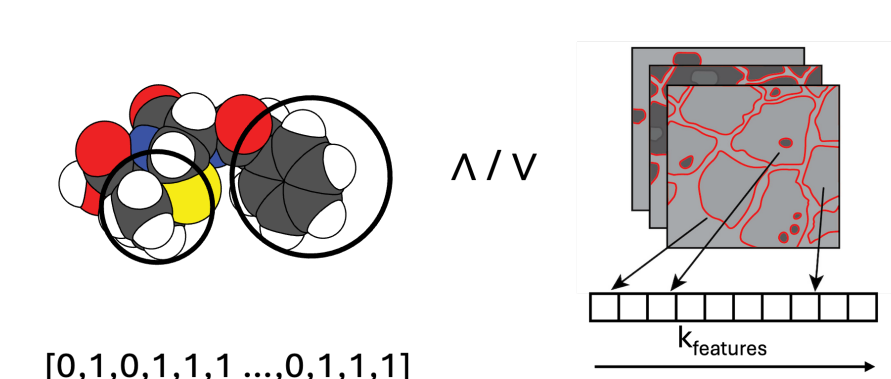
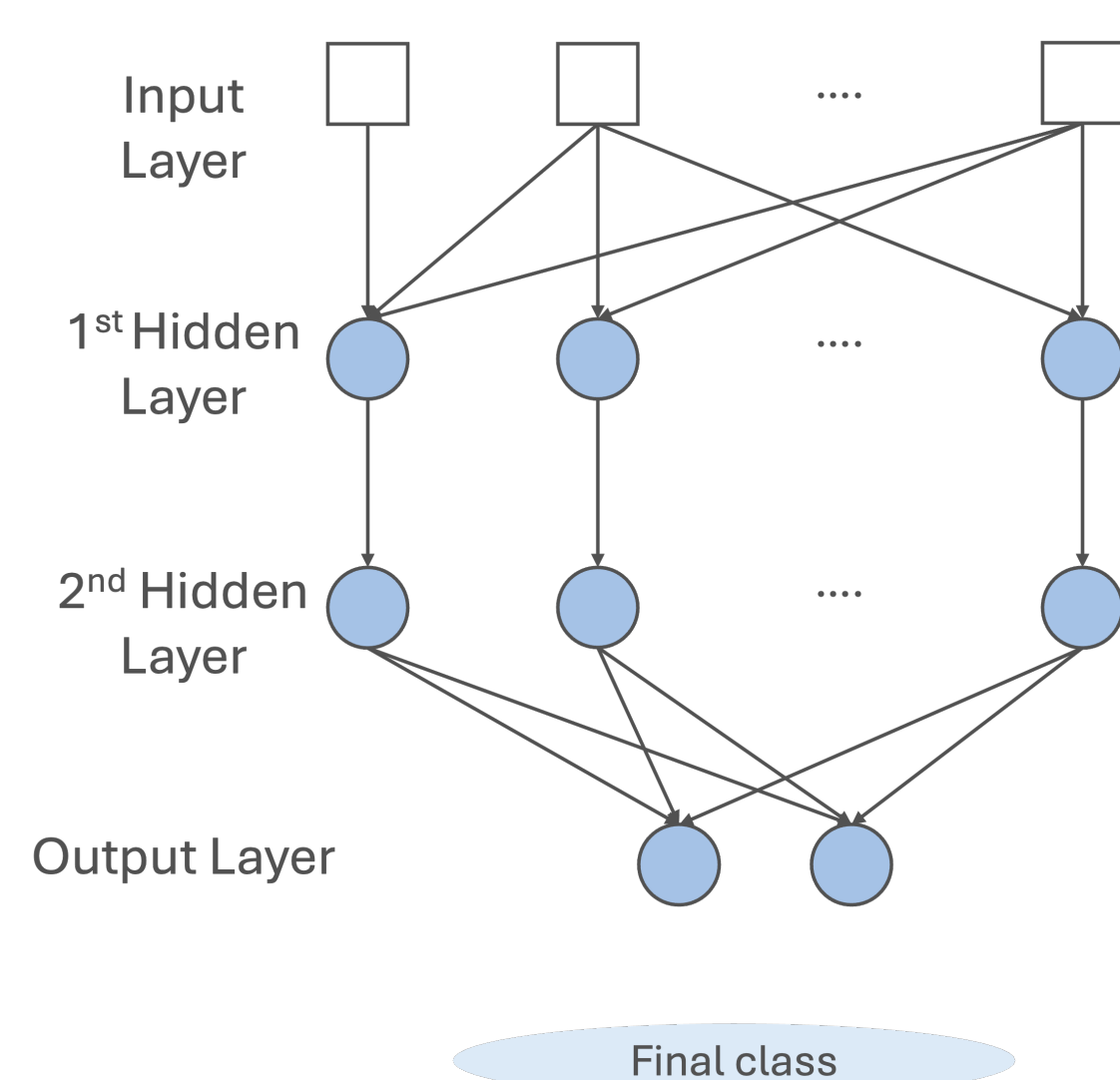
**Assay Focus:**  $\beta$ -lactamase reporter gene (19 total):

- 10 receptors : AR, ER $\alpha/\beta$ , FXR, PPAR $\delta$ , PPAR $\gamma$ , RXR, VDR, GR, PR
- 642 molecules overlap with those in the morphological dataset  
→ 324 inactive / 318 active



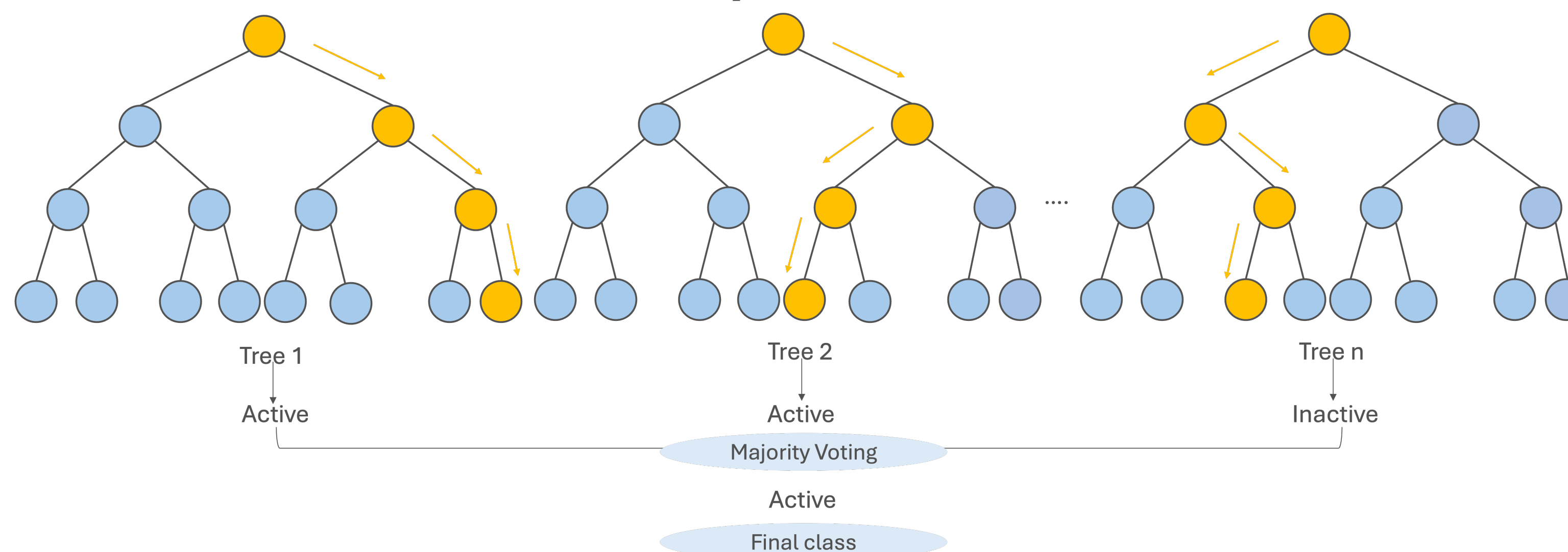
## Predictive modelling: Endocrine activity

### Multi Layer Perceptron (MLP)<sup>[7]</sup>:



**Data representation:** Both morphological and structural fingerprint<sup>[6]</sup> represented as vector.

### Random Forests (RF)<sup>[8]</sup>:



## Results

**Metric for evaluation:** Matthews correlation coefficient (MCC) range from [-1 (worst) to 1 (best)]

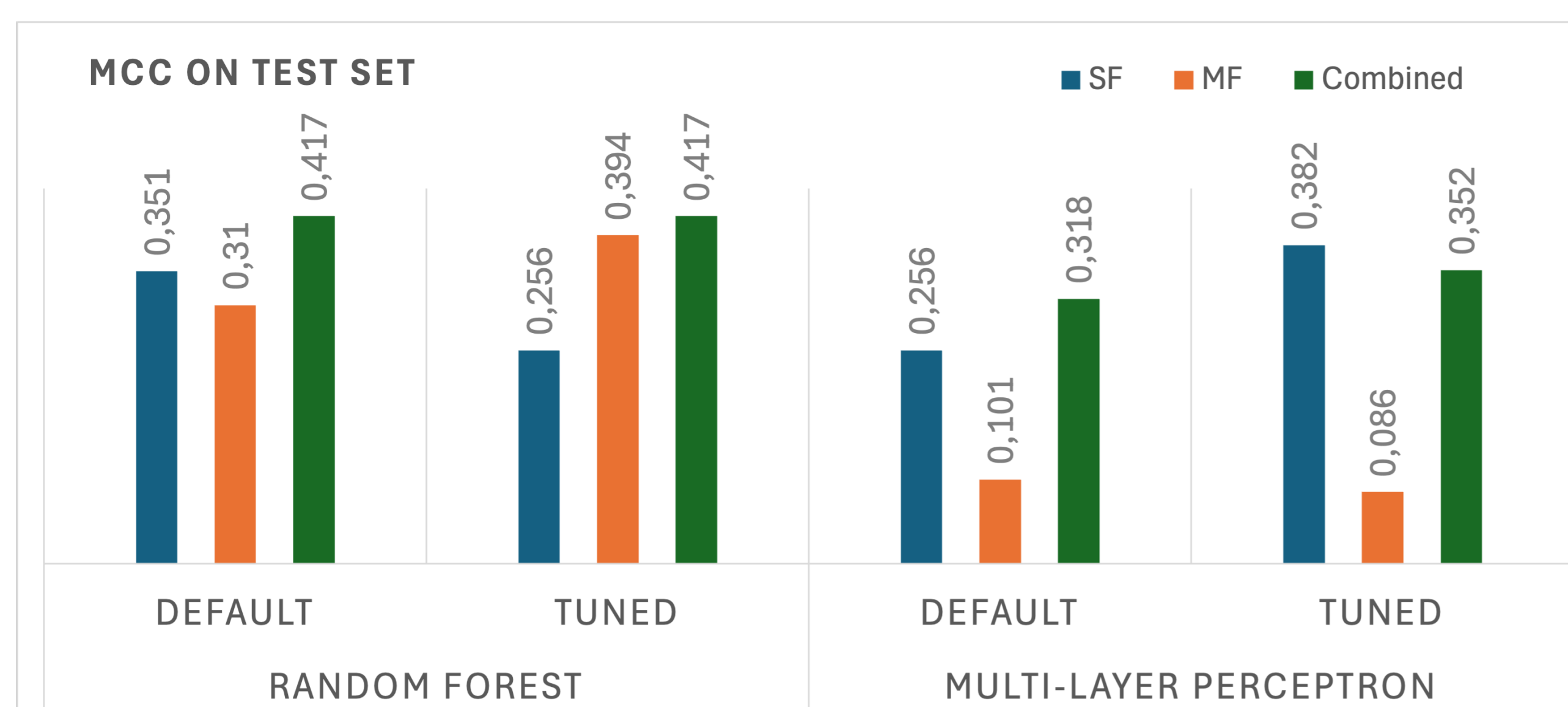
### Models:

Tuned vs default hyper-parameters

Random forest vs Multi Layer Perceptron

**Input:** Morphological fingerprint (MF) vs. structural fingerprint (SF) vs. combination of both

→ MF complement SF, helping identify common modes of action for molecules with different targets.<sup>[2,9,10]</sup>



GitHub Repo with tutorial

[1] Cimini, B. A., Chandrasekaran, S. N., Kost-Alimova, M., Miller, L., Goodale, A., Fritchman, B., ... & Carpenter, A. E. (2023). Optimizing the Cell Painting assay for image-based profiling. *Nature Protocols*, 18(7), 1981-2013.

[2] Odje, F., Meijer, D., Von Coburg, E., van der Hoof, J. J., Dunst, S., Medema, M. H., & Volkamer, A. (2023). Unleashing the potential of cell painting assays for compound activities and hazards prediction. *Frontiers in Toxicology*, 6, 1401036.

[3] Carpenter, A. E., Jones, T. R., Lamprecht, M. R., Clarke, C., Kang, I. H., Friman, O., ... & Sabatini, D. M. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology*, 7, 1-11.

[4] Tox21 NR assay, EPA.

[5] Datagrok, Inc. (n.d.). Retrieved from <https://datagrok.ai>

[6] Rogers, D., & Hahn, M. (2010). Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling*, 50(5), 742-754.

[7] Scikit-learn: MLP - Multi-Layer Perceptron Classifier. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.

[8] Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32.

[9] Schneidewind, T., Brause, A., Pahl, A., Burhop, A., Mejuch, T., Sievers, S., et al. (2020). Morphological profiling identifies a common mode of action for small molecules with different targets. *ChemBioChem*, 21(21), 3197-3207. doi:10.1002/cbic.202000381

[10] Von Coburg, E., & Dunst, S. (2023). The adverse outcome pathway for breast cancer: a knowledge management framework bridging biomedicine and toxicology. *Discover Oncology*, 14(1), 2023.

