# Biological Interpretation of Cell Painting and Gene Expression Features for Mitochondrial Toxicity Prediction

Srijit Seal[1], Jordi Carreras-Puigvert[2], Maria-Anna Trapotsi[1], Hongbin Yang[1], Ola Spjuth[2], Andreas Bender[1]
[1]*Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW*
[2]*Department of Pharmaceutical Biosciences and Science for Life Laboratory, Uppsala University, Box 591, SE-75124, Uppsala, Sweden*
ss2686@cam.ac.uk

## *Aim and Background*

**High-dimensional Cell Painting and L1000 gene expression** (Figure 1) **are versatile biological descriptors of a system.**

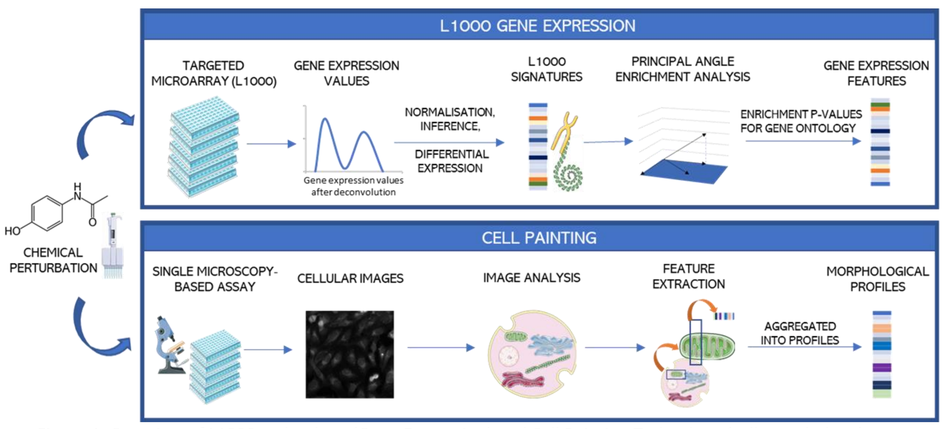



*Figure 1. Overview of L1000 technology (Gene Expression) and Cell Painting Technology (cell morphology)*

**Can integrating Cell Painting Profiles with Gene Expression and Chemical Structure** (as shown in Figure 2) **improve detection of mitochondrial toxicity?**

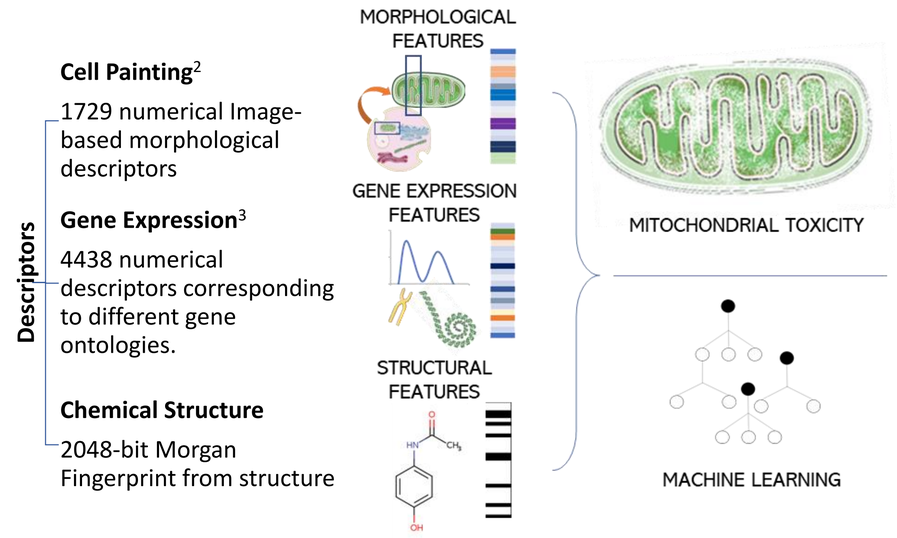



*Figure 2. Overview of machine learning strategy used in this study integrating cell morphology with Gene Expression and chemical structure for mitochondrial toxicity.*

**Mitochondrial toxicity** can be caused by different mechanisms as shown in Figure 3:

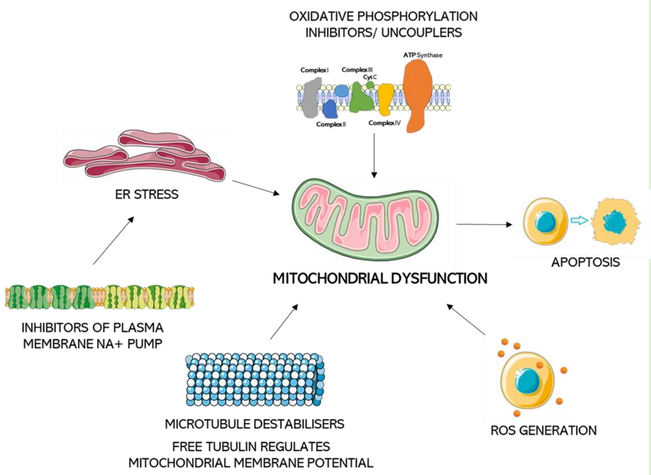
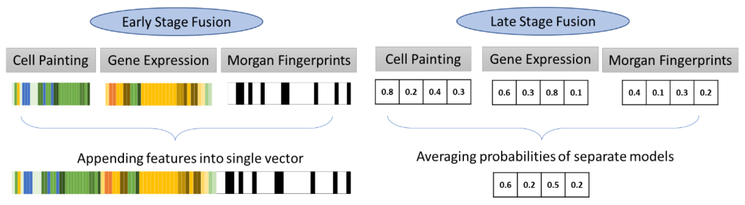



*Figure 3. Major mechanisms of mitochondrial toxicants. Toxicants act on multiple pathways to exhibit mitochondrial toxicity, mostly inhibition of mitochondrial respiratory chain or uncoupling of oxidative phosphorylation.*

## *Dataset and Methods*

***Dataset***
*in vitro* mitochondrial toxicity from Tox21 mitochondrial membrane potential disruption assay.

***Model Algorithm:*** Random Forest




**3 Individual Models**
- Cell Painting,
- Gene Expression
- Morgan fingerprints

**Early-stage models** fusing Cell Painting, Gene Expression and Morgan fingerprints into a single vector.

**Late-stage model** averaged predicted probabilities of the three individual models

***Evaluation***

***Nested Cross Validation*** 50 repeated 4-fold nested cross-validations on 382 compounds.
***External Test Set*** Models were evaluated on an external test set of 236 compounds

## *Morphological space can be a valuable feature space in detecting mitochondrial toxicity*

**Compounds clustered in morphological space having similar mechanisms of actions** which reduce mitochondrial membrane potential. (as shown in Figure 4):

- microtubule disruptors
- inhibitors of plasma membrane $Na^+$ pump
- Caspase activation/GSH depletion
- Trigger the release of cytochrome C

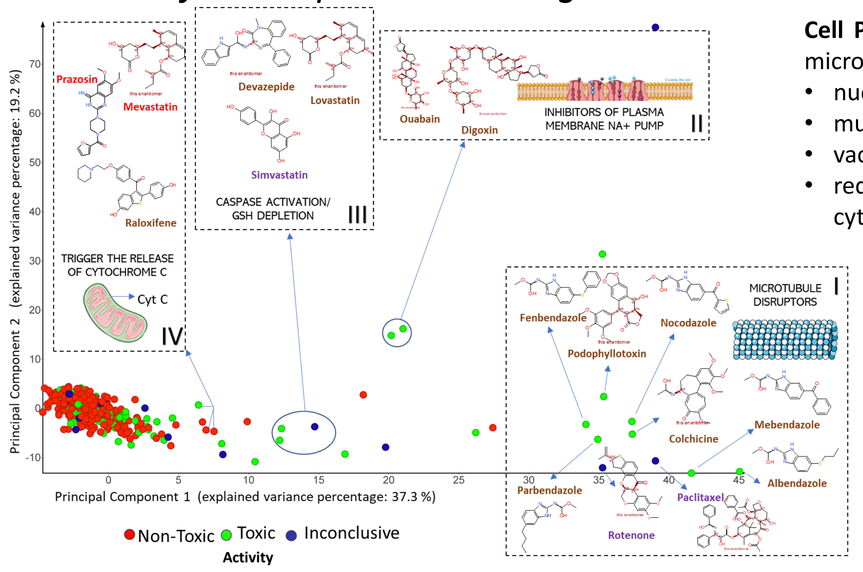



*Figure 4: Principal Component Analysis of 542 compounds in 110-dimensional Cell Painting feature space.*

**Mitochondrial Toxicants are More Similar in Morphological Space Compared to Structural Space** (as shown in Figure 5) **:**

Mitotoxic compounds considerably vary from non-toxic compounds in morphology space (median Pearson correlation of 0.140 vs 0.038).

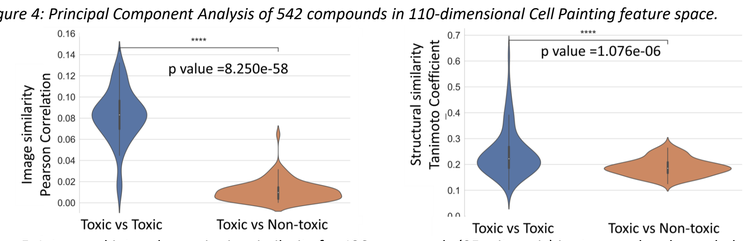



*Figure 5: Intra- and inter-class pairwise similarity for 486 compounds (85 mitotoxic) in structural and morphological space*

**Cell Painting phenotype reveals alterations** for microtubule disruptors (as shown in Figure 6)
- nuclear fragmentation
- multinucleated cells
- vacuolation of the endoplasmic reticulum
- redistribution of the mitochondria and cytoskeleton destabilisation

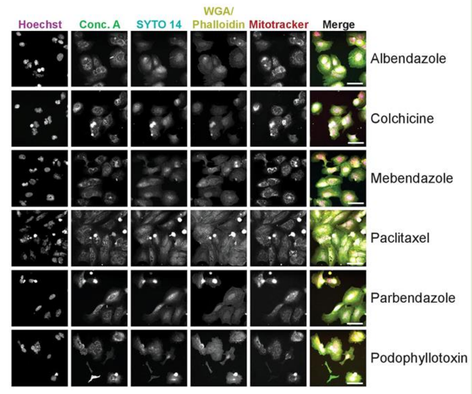



*Figure 6. Representative images of cells stained using the Cell Painting assay upon exposed to drugs Albendazole, Colchicine, Mebendazole, Paclitaxel, Parbendazole and Podophyllotoxin (microtubule disruptors that induce cytotoxicity). These images are publicly available through the Broad Bioimage Benchmark Collection (https://bbbc.broadinstitute.org/image_sets).*

## *Translating Computational features of the Cell Painting features and Gene Expression features to biological implication in mitochondrial toxicity*

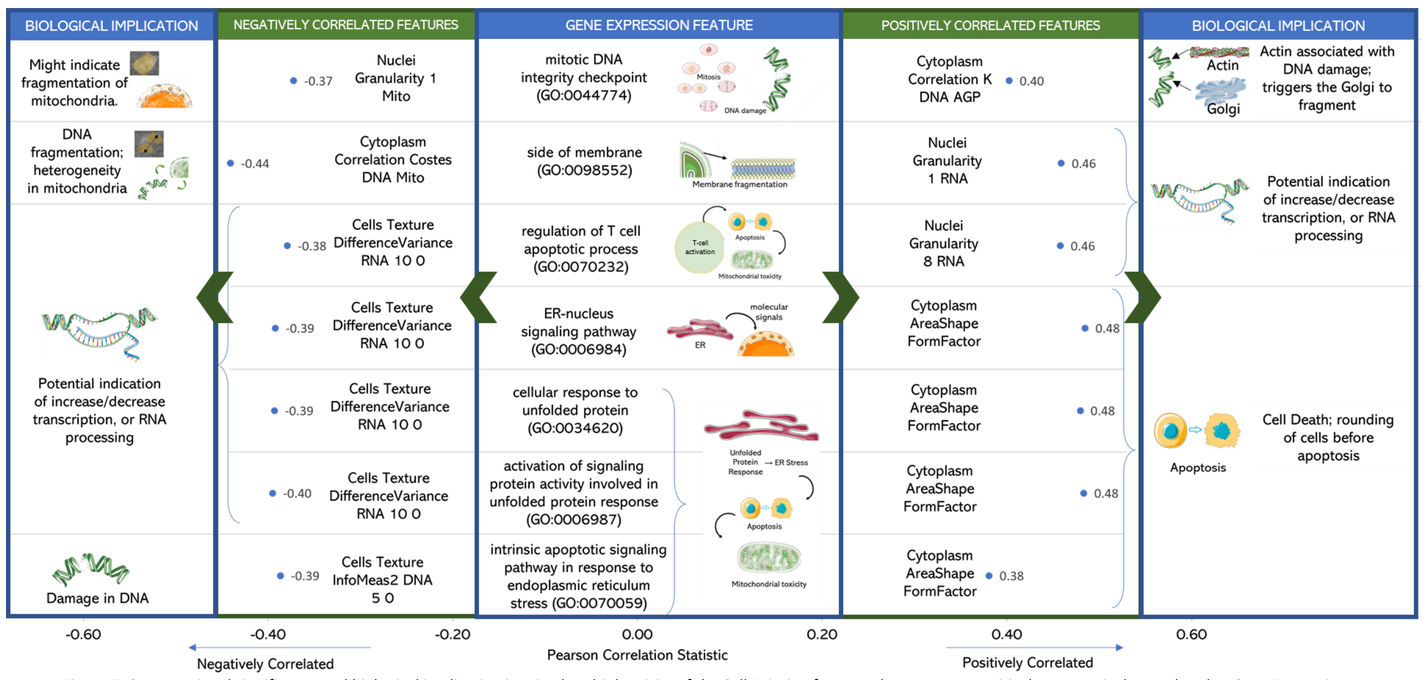



*Figure 7. Computational significance and biological implication in mitochondrial toxicity of the Cell Painting features that are most positively or negatively correlated to Gene Expression.*

**Cell Painting and Gene Expression features** as shown in Figure 7**:**

- Unfolded protein response and endoplasmic reticulum stress (RNA variance and cell area shape)
- T cell apoptotic processes (mitochondrial granularity and DNA fragmentation)
- Side of the membrane (RNA granularity and heterogeneity in mitochondria)

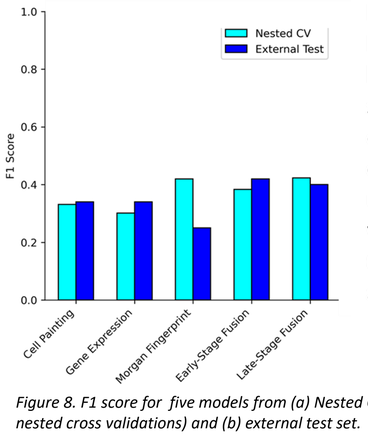
**Biological significance of Cell Painting features with respect to Mitochondrial Toxicity :**

- Edge intensity of cells (possibly related to integrity of cell wall)
- Radial distribution and intensity in mitochondria (related to mitochondrial death)
- Granularity features (related to cell death and amount of information contained in cellular images)

**Biological significance of Gene Expression features with respect to Mitochondrial Toxicity :**

- Unfolded protein response (possibly related to ER stress)
- Plasma membrane (related to membrane depolarisation).
- Some effects of mitochondrial toxicity were captured by Gene Expression features such as oogenesis and dendritic plasma membrane; both processes are heavily mitochondria dependent

## *Results in Predictive Performance*




**Early-stage fusion and Late-stage fusion models have higher F1 score** (as shown in Figure 8) in external test sets compared to models using Morgan fingerprints where F1 Score falls by 60% (0.25 to 0.40 in absolute terms)

*Figure 8. F1 score for five models from (a) Nested CV (median of repeated nested cross validations) and (b) external test set.*

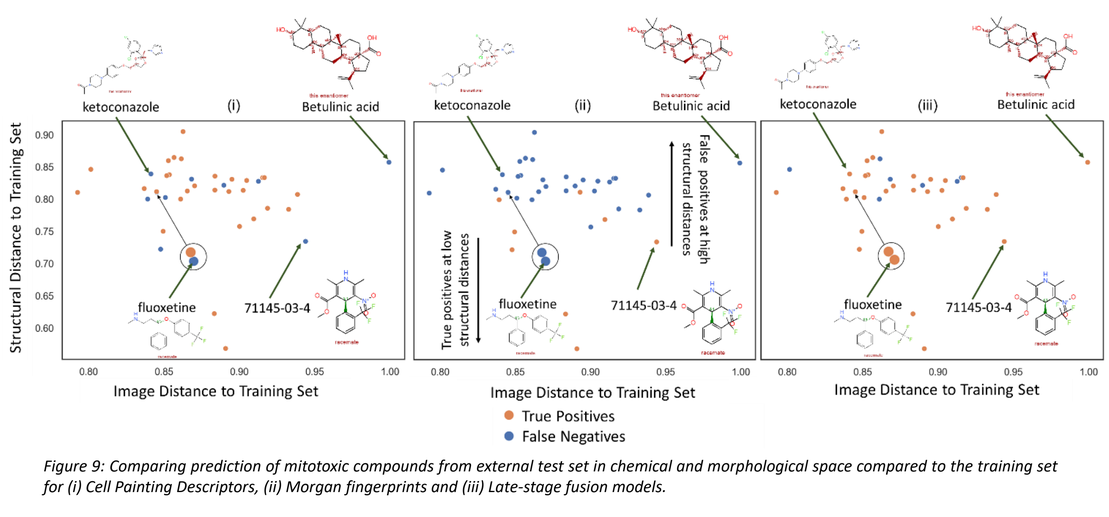



*Figure 9: Comparing prediction of mitotoxic compounds from external test set in chemical and morphological space compared to the training set for (i) Cell Painting Descriptors, (ii) Morgan fingerprints and (iii) Late-stage fusion models.*

**Comparison to Previous Machine Learning Models and Dedicated in-vitro Mitochondrial Toxicity Assays:**
In comparison, our method achieve higher sensitivity (0.82 in our study vs 0.37 in Apredica MitoMass[4]) with comparable balanced accuracies (0.68 in our study vs 0.65 in Apredica MitoMass[4]).

**Models could extrapolate well into new chemical spaces** (as shown in Figure 9):
- Morgan fingerprints correctly classify mitotoxic compounds at low Tanimoto distance to training set.
- Cell Painting descriptors extrapolate well into structurally diverse compounds.
- However, when the distance to morphological space was high, CP descriptors failed.

## *Significance and Conclusions*

- Mitochondrial toxicants significantly differ from non-toxic compounds in morphological space; clusters with similar mechanisms.
- Cell Painting features granularity features are highly predictive mitochondrial toxicity.
- Models combining Cell Painting, Gene Expression features and Morgan Fingerprints relatively improved detection (F1 Scores) of mitochondrial toxicants (by 60% from 0.25 to 0.40) compared to models using only structural features.
- Models extrapolated well into new chemical space.
- Finally, for detecting mitochondrial toxicants, these models using hypothesis-free features could perform with better sensitivity than some dedicated and hypothesis-based experimental high content imaging assays for mitochondrial toxicity.

**References:**
1. Seal, S. et al. Integrating Cell Morphology with Gene Expression and Chemical Structure to Aid Mitochondrial Toxicity Detection. bioRxiv 2022.01.07.475326 (2022) doi:10.1101/2022.01.07.475326.
2. Bray, M. A. et al. A dataset of images and morphological profiles of 30 000 small-molecule treatments using the Cell Painting assay. GigaScience. 6, 1–5 (2017).
3. Zhao, P. et al. In silico prediction of mitochondrial toxicity of chemicals using machine learning methods. J. Appl. Toxicol. jat.4141 (2021)
4. Hallinger, D. R., Lindsay, H. B., Friedman, K. P., Suarez, D. A. & Simmons, S. O. Respirometric screening and characterization of mitochondrial toxicants within the toxcast phase i and II chemical libraries. Toxicol. Sci. 176, 175–192 (2020).

**Acknowledgements:**
S.S. acknowledges the Cambridge Commonwealth, European and International Trust, Boak Student Support Fund (Clare Hall), Jawaharlal Nehru Memorial Fund, Allen, Meek and Read Fund, Trinity Henry Barlow (Trinity College), Clare Hall Boak Fund for providing funding for this study. OS acknowledges funding from Swedish Research Council under grant #2020-01865 and #2020-03731. This work was performed using resources provided by the Cambridge Service for Data Driven Discovery (CSD3) operated by the University of Cambridge Research Computing Service (www.csd3.cam.ac.uk), provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council (capital grant EP/P020259/1), and DiRAC funding from the Science and Technology Facilities Council (www.dirac.ac.uk).