bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Integrating Cell Morphology with Gene Expression and Chemical Structure to Aid Mitochondrial Toxicity Detection

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 Rd, CB2 1EW, Cambridge, United Kingdom

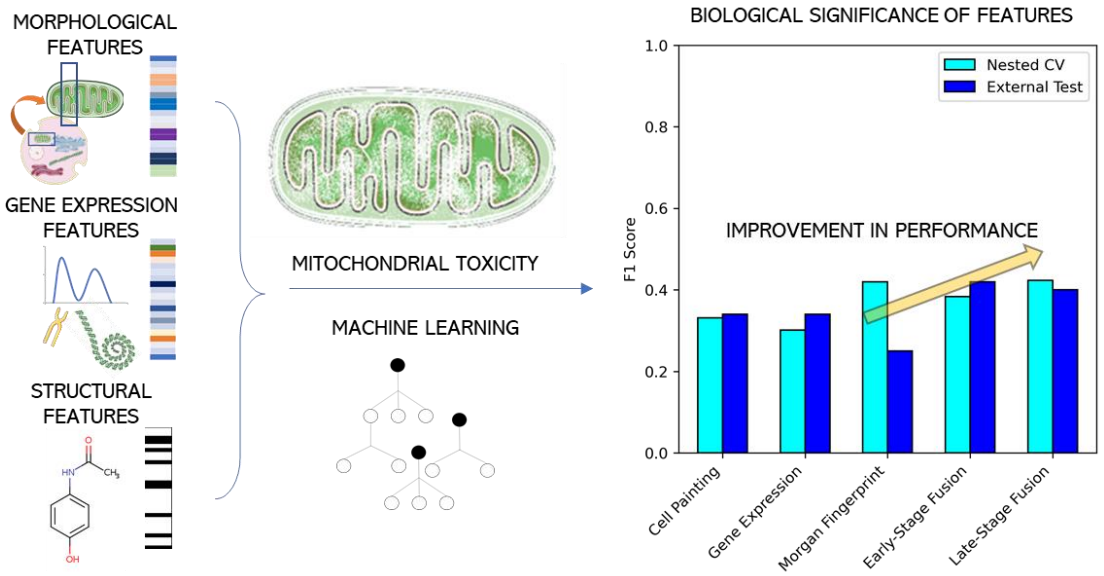
[2]Department of Pharmaceutical Biosciences and Science for Life Laboratory, Uppsala University, Box 591, SE-75124, Uppsala, Sweden

Mitochondrial Toxicity; Cell Painting; Toxicity Prediction; Cell Morphology

* Email: ab454@cam.ac.uk, ola.spjuth@farmbio.uu.se

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

GRAPHICAL ABSTRACT



bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ABSTRACT

Mitochondrial toxicity is an important safety endpoint in drug discovery. Models based solely on chemical structure for predicting mitochondrial toxicity are currently limited in accuracy and applicability domain to the chemical space of the training compounds. In this work, we aimed to utilize both -omics and chemical data to push beyond the state-of-the-art. We combined Cell Painting and Gene Expression data with chemical structural information from Morgan fingerprints for 382 chemical perturbants tested in the Tox21 mitochondrial membrane depolarization assay. We observed that mitochondrial toxicants significantly differ from non-toxic compounds in morphological space and identified compound clusters having similar mechanisms of mitochondrial toxicity, thereby indicating that morphological space provides biological insights related to mechanisms of action of this endpoint. We further showed that models combining Cell Painting, Gene Expression features and Morgan fingerprints improved model performance on an external test set of 236 compounds by 60% (in terms of F1 score) and improved extrapolation to new chemical space. The performance of our combined models was comparable with dedicated *in vitro* assays for mitochondrial toxicity; and they were able to detect mitochondrial toxicity where Tox21 assays outcomes were inconclusive because of cytotoxicity. Our results suggest that combining chemical descriptors with different levels of biological readouts enhances the detection of mitochondrial toxicants, with practical implications for use in drug discovery.

Mitochondrial Toxicity; Cell Painting; Gene Expression; Cell Morphology; Machine Learning

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

INTRODUCTION

Drug-induced mitochondrial toxicity[1] is being increasingly recognized as a contributor to late-stage withdrawals[2] causing cardiotoxicity[3], drug-induced liver toxicity[4] and diseases related to ageing such as Parkinson.[5] Mitochondrial toxicity can be directly or indirectly caused by combinations of multiple mechanisms (as shown in Figure 1) which makes predicting mitochondrial toxicity challenging.[6] A common direct cause of mitochondrial dysfunction is uncoupling of the electron transport chain from ATP synthesis or accumulation of calcium in mitochondria causing an increase in Reactive Oxygen Species (ROS) leading to oxidative stress and damaging mitochondrial DNA (mtDNA).[7] Indirect effects of drugs on cells such as inhibition of fatty acid β-oxidation, uncoupling of oxidative phosphorylation, the opening of the membrane permeability transition pore, and disruption of mtDNA synthesis and translation have also been shown to cause mitochondrial toxicity.[7]

Risk of mitochondrial toxicity in drug discovery can be indicated either via experimental methods (such as the Glu/Gal assay[8]) or using predictive methods trained on data from in vitro assays. Dedicated assays often use HepG2 cells to detect mitochondrial toxicants[9] and mostly fluorescent dyes (Mito-MPS[10], DiOC6[11], rhodamine-123[12], MitoTracker Orange[13], TMRM[12], TMRE[12], JC-1[14]). These assays capture proxy endpoints, for example, membrane depolarisation, which are quite heterogenous and not in absolute concordance to the term "mitochondrial toxicity". Another example, in the Glu/Gal assay, the ratio of IC_{50} values in different cultures is not always easy to translate to *in vivo* effects. Further, each fluorescent dye has its limitation[15], for example, JC1 is sensitive to membrane depolarization but disadvantaged by its poor water solubility and low signal-to-background window**.** Imaging assays, such as the Apredica HepG2 mitochondrial membrane potential and mitochondrial mass assays[16] measure the average cell intensities for mitochondria from high content imaging, where the average

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

intensity of mitochondria was used to define mitochondrial membrane potential, and the total intensity is used to define mitochondrial mass.[16] This makes defining and detecting mitochondrial toxicity a challenging task in itself.

Previous approaches to computationally predict mitochondrial toxicity have to a large extent been based on predicting mitochondrial membrane depolarisation using chemical structure (Supplementary Figure S1, Supplementary Table S1) and machine learning methods including Support Vector Machines[17], Random Forest models[18,19], and naïve Bayes classifier[20]. Using molecular descriptors or structural fingerprints, the best models showed a balanced accuracy between 0.74 to 0.86 as reported by Zhao et. al.[18] However, Zhao *et al.* showed that extrapolation to new structural space is difficult and accuracy inside the models' applicability domain was significantly higher when compared with out-of-domain compounds.[18]

Since mitochondrial toxicity can be characterised by a multitude of mechanisms[3], it has been challenging to assemble sufficient data that can sustain computational methods able to extrapolate to new chemical space. Together with the fact that *in vitro* assays for mitochondrial toxicity are demanding and with varying degree of reliability, there is a clear need for advancements in the field.[21] In recent years, hypothesis-free data on cell lines has become available on a much larger scale, both publicly and in company repositories. In this work, we explore how data from Cell Morphology in addition to Gene Expression can improve the detection of mitochondrial toxicity. To the best of our knowledge, this is the first study that presents predictive models for mitochondrial toxicity *in vitro* assays based on integrated data derived from two types of hypothesis-free data and chemical structure.

The LINCS L1000 gene expression technology developed by Broad Institute (described in Figure 2) captures changes in 978 landmark genes, and large scale data before and after treating different human cell lines to FDA-approved drugs and small molecules is now available on a

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

sufficiently broad base to be useful for modelling.[22] Gene Expression features have been used in predicting *in vitro* cell viability[23], drug proteins targets[24], organ level toxicity such as hepatotoxicity[25], nephrotoxicity and cardiotoxicity[26]. The Gene Ontology initiative aims to unify gene and gene product attributes in a classification effort that will provide functional interpretation of gene expression data which, in our case, helps better generalise pathways of mitochondrial toxicity.[27] The Cell Painting assay (described in Figure 2) is a relatively recent technology developed by the by Broad Institute and is used to capture cellular morphological changes in image data from genetic or chemical perturbations.[28,29] Microscopic images are processed to obtain over 1700 measures of cellular and organelle changes such as morphology, texture and intensity. Cell Painting features have been previously used in predicting *in vitro* toxicity such as cytotoxicity[30], bioactivity endpoints[31] and mechanism of action[32], cell health phenotypes[33], drug-protein targets[34], antiviral drug discovery[35] as well as organ level toxicity such as drug-induced liver toxicity[36]. However, the predictivity of high-dimensional biological features for safety- or efficacy-related endpoint needs to be established in each case, which for *in vitro*-to-*in vivo* extrapolation (for example from mitochondrial toxicity to liver injury caused by the former) is not a trivial exercise.[37,38]

With the availability of high throughput hypothesis-free data from cell profiling technologies, we are presented with new opportunities to improve detection of mitochondrial toxicity. In this work, we use Cell Painting and Gene Expression features to extrapolate the applicability domain of structure-based models to the new chemical space. While Gene Expression data is easier to directly interpret[39], in this work we put particular emphasis on exploring and interpreting the biological significance and applicability of Cell Painting features that contain information about mitochondrial toxicity.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

RESULTS

Data for *in vitro* mitochondrial toxicity was collected from the Tox21 assay[40] for mitochondrial membrane potential disruption summary assay (AID 720637).[41] Image-based morphological features from the Cell Painting assay were collected from Bray *et al*.[29] Gene Expression features were extracted from the LINCS L1000 dataset as pre-processed by the Ma'ayan Lab.[39] A combined dataset was assembled to be used for model development (henceforth referred as "training data") that contained 404 distinct compounds (62 mitotoxic, 320 nontoxic and 22 inconclusive) that contained both Cell Painting features and Gene Expression features. An external test set was assembled that comprised a total of 236 distinct compounds (39 mitotoxic and 197 nontoxic) from Hemmerich *et al*.[19] who compiled various assays relevant to toxicity of mitochondrial function, binding and inhibition, where no compound overlapped with the training data. Both datasets covered drugs over a wide range of ATC code distribution at the top level 327 drugs (training data) and 103 drugs (external test set) as shown in Supplementary Figure S2.

Mitochondrial toxicants are similar in morphological space

We analysed if mitochondrial toxicants were more similar to each other in morphological space than toxicants to non-toxicants, which could be a prerequisite for the use of this readout space for the detection of mitochondrial toxicity. This was done by comparing the median values of 5 highest Tanimoto similarity coefficients and the absolute value of median of 15 most positively and 15 most negatively Pearson correlation statistic values for Cell Painting features. As shown in Figure 3, we found that mitotoxic compounds are considerably different from non-toxic compounds in morphological space (median Pearson correlation of 0.08 vs 0.01, t-test independent p value=3.3e-20). However, they remain distinguishable in structural space (median Tanimoto Similarity of Morgan fingerprints 0.22 vs 0.19, t-test independent

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

samples p value=6.3e-03). We conclude that morphological space can discriminate between mitochondrial toxicants and non-toxicants, and that this readout space is more able to discriminate between both classes of compounds than chemical fingerprints on the dataset analysed here.

Cell Painting features cluster mitochondrial toxicants to identify different mechanisms of mitochondrial toxicity

We firstly analysed morphological readout space with respect to the ability to differentiate different mechanisms of action (MOA) for mitochondrial toxicity. We performed feature selection on the initial 1,729 features (see *Methods Section*) which selected 110 Cell Painting features and visualised the morphological space using Principal Component Analysis (PCA). As shown in Figure 4, compound clusters emerged, which were related to mitochondrial toxicity (for further details see Supplementary Table S2). In particular, Cluster I (Figure 4) comprises several microtubule destabilisers such as fenbendazole, parbendazole, mebendazole, that belong to the benzimidazole class[42,43,44] together with structurally dissimilar compounds, namely rotenone and paclitaxel, both of which are known mitochondrial toxicants as well as microtubule destabilizers.[45,46] Supplementary Figure S3, shows that cell painting phenotypes for six microtubule disruptor drugs (Cluster I: albendazole, colchicine, mebendazole, paclitaxel, parbendazole and podophyllotoxin) reveals alterations at the nuclear level, depicted by nuclear fragmentation as well as multinucleated cells, vacuolation of the endoplasmic reticulum, redistribution of the mitochondria and cytoskeleton destabilisation. We found ouabain and digoxin in Cluster II (Figure 4) have similar mechanisms for mitochondrial injury as inhibitors of the plasma membrane Na^+ pump, which can lead to impaired mitochondrial Ca^{2+} retention, increased ROS production and reduced mitochondrial membrane potential.[47,48] Cluster III (Figure 4) consists of statins, namely lovastatin and simvastatin, which are known

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to inhibit the synthesis of mevalonate, a precursor of ubiquinone that is vital to the mitochondrial respiratory chain, and causes oxidative stress.[49] Compounds in Cluster IV (Figure 4), namely mevastatin (a statin), raloxifene (a selective estrogen receptor modulator) and prazosin (an alpha blocker), form again a cluster that is rather diverse with respect to chemical structures, and primary pharmacology/indication areas. However, those compounds are all known to induce apoptotic signalling cascades which trigger the release of cytochrome c into the cytosol.[50,51,52] This causes depolarization in the mitochondrial membrane leading to mitochondrial injury. Overall, our findings show that cell morphology readouts from the Cell Painting assay can cluster several modes of action of mitochondrial toxicants, such as the disruption of microtubules, increased ROS production and oxidative stress.

Cell Painting features are correlated to Gene Expression features

We used 62 known mitotoxic compounds to calculate Pearson's correlation between the selected 110 Cell Painting features, and 10 Gene Expression features related to unfolded protein response, endoplasmic reticulum stress, T cell apoptotic process and side of the membrane which represent biological processes which from prior knowledge are known to be related to mechanisms mitochondrial toxicity.[53,54] We found significant correlations for specific Cell Painting features with these Gene Expression features as shown in Figure 5 (further details on biological significance in Supplementary Table S3). We found that Gene Expression features corresponding to unfolded protein response and endoplasmic reticulum stress were most positively correlated to "Cytoplasm_AreaShape_FormFactors". Form factors indicate how perfectly circular an object is which corresponds to the rounding up of cells due to apoptosis and could be indicative of cell death (caused by ER stress or the unfolded protein response which induce cell death, like many other stress responses[55]). Gene Expression features related to unfolded protein response were negatively correlated to

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

"Cells_Texture_DifferenceVariance_RNA_10_0" which calculates the image variation in a normalized co-occurrence matrix and could correspond to various secondary processes following ER stress (including a reduction in transcription, but also reduced translation, caspase activation, apoptosis, etc.[56]). The Gene Expression feature "side of membrane", which is a parent to the cytoplasmic side of mitochondrial outer membrane was found most positively correlated to "Nuclei_Granularity_1_RNA" and most negatively correlated to "Cytoplasm_Correlation_Costes_DNA_Mito". An increase or decrease in granularity of cytoplasmic RNA, in the proximity of the nucleus, might indicate the formation of RNA inclusion bodies or RNA processing while the correlation between DNA and mitochondria object could correlate to DNA fragmentation and heterogeneity in mitochondrial content. Hence, we conclude that Cell Painting features contain information of biological significance related to pathways of mitochondrial membrane depolarisation.

Cell Painting and Gene Expression enables training of accurate and interpretable models for detecting mitochondrial toxicity

As the utility of chemical structure in detecting mitochondrial toxicity was previously explored by Hemmerich *et. al.*[19], our work focussed on comparing individual Cell Painting features and Gene Expression features with respect to their ability to detect mitochondrial toxicants. We used positive predictive values (PPV) and F1 scores from single decision tree classifiers trained on individual features (see *Methods*) to detect a signal for mitochondrial toxicants and provide a biological interpretation of these feature spaces. We found that Cell Painting features related to granularity, intensity, location, and radial distribution of mitochondrial objects over the three compartments (cells, cytoplasm and nuclei) had high predictivity for mitochondrial toxicity (median PPV grouped by compartment, channel, and feature group above 0.70; Supplementary Figure S4). We next more closely considered the

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

feature value distribution for individual features with high PPV for mitotoxicity (Supplementary Figure S5*)* For example, "Cells_Intensity_MaxIntensityEdge_Mito" (PPV=0.83). Compounds toxic to mitochondria evenly affects edge of the mitochondria object. Since this measurement is at the edge of the segmented object, it indicates a loss of membrane integrity. Another feature, "Cells Intensity MADIntensity Mito" (PPV=0.8) is a measurement of statistical dispersion which measures the standard deviation and median absolute deviation (MAD) of pixel intensity values while being robust to outliers. For MitoTracker Deep Red used in Cell Painting assay, this might indicate a variation of intensities among fragments of the mitochondrial membrane, resulting from loss of membrane integrity. "Cells Granularity 1 RNA" (PPV=0.56) reveals information present of pixel 1 in the RNA channel where certain mitotoxic compounds also have significantly lower feature values. An increase or decrease in granularity of cytoplasmic RNA might indicate formation of RNA inclusion bodies or RNA processing. Further attempted biological interpretations for some features (knowing that this is not a trivial process) are shown in Figure 6 and Supplementary Table S4.

Gene Expression features with high PPV could be classified as either causing mitochondrial membrane depolarisation or as an effect of mitochondrial toxicity (as shown in Figure 7 and Supplementary Table S5). Features such as endoplasmic reticulum unfolded protein response (PPV 0.92) and activation of signalling protein activity involved in unfolded protein response (PPV 1.00) have previously been attributed to ER related effects such as protein folding, oxidative stress and ER stress.[55] Such effects are linked to each other and toxins affecting the same can depolarise the mitochondrial membrane affect the movement of mitochondria on microtubules and eventually regulate apoptosis.[57] Features such as external side of plasma membrane (PPV 1.00), side of membrane (PPV 1.00), autophagic vacuole membrane (PPV 1.00), negative regulation of T cell activation (PPV 0.86) are related to processes of cell proliferation, cell cycle arrest as well as apoptosis that causes oxidative stress and cell death

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

which can cause mitochondria to depolarise.[58] The GO Cellular Component endritic plasma membrane (PPV 0.88) and the Biological Process oocyte development (PPV 0.88) are greatly affected by mitochondrial dysfunction as neurons are mitochondria dependent cells[59] while oocyte development requires optimal energy production and is highly dependent on mitochondrial function for the same.[60] Hence we conclude that a significant number of Cell Painting and Gene Expression features showing a high predictivity to mitochondrial membrane depolarization are also interpretable in the mechanistic roles either causing mitochondrial toxicity or being a consequence of the same.

Fusion models accurately detects mitochondrial toxicity and expands the applicability domain

We finally established predictive models for mitochondrial toxicity based on 3 models using Cell Painting features, Gene Expression features and Morgan fingerprints and another 2 combinations thereof in early- and late-stage fusion. Early-stage fusion appended all three features into a single vector while late-stage fusion averaged the probabilities of the three individual models. We used a Random Forest model with repeated nested cross validation on a training data of 382 compounds (out of which 62 have mitotoxic annotations) and validated using an external dataset of 236 compounds (39 mitotoxic) where test compounds, although run across various assay conditions) were structurally diverse and generally dissimilar to the training data (details shown in Supplementary Figure S6**).** Figure 8 shows median performance from nested-cross validations and external validations (for further results see Supplementary Table S6 and for an overview of results from each fold of nested-cross validation see Supplementary Figure S7).

Fusion models combining Cell Painting features, Gene Expression features, and Morgan fingerprints exhibited significantly higher F1 scores on the external dataset (early-stage fusion:

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

0.42, late-stage fusion: 0.40) in detecting mitotoxicity than models using only Morgan fingerprints (0.25). The drop in F1 score of models using only Morgan fingerprints from 0.42 in repeated nested cross-validation to 0.25 in the external test (Figure 8) shows that Morgan fingerprints lack extrapolation power to novel chemical space. Although the training dataset was different and larger in previous work by Hemmerich et. al[19] focusing on purely chemical structure data (1412 compounds vs 382 compounds here), the results are hence not directly comparable; our late-stage fusion model based had slightly higher sensitivity (0.82 vs. 0.79) which implied improved ability to detect mitochondrial toxicants in the external test set. As shown in Figure 9a, the success of fusion models is underlying in the fact that in morphological space, mitotoxic compounds in the external test set were more morphologically similar to mitotoxic compounds in the training set while no such correlation was present among the images of non-toxic compounds. Finally, late-stage fusion was more sensitive to the toxic class in the external test set compared to early-stage models (0.82 vs 0.62) while the balanced accuracy remained the same (0.68). Given the importance of detecting mitotoxicants in practice, higher sensitivity of a model is likely advantageous in practical situations even at identical balanced accuracy.

We next analysed in more detail the predictions of mitochondrial toxicants in the external test set with our models as shown in Figure 9b (with further details shown in Supplementary Table S7). The model using Morgan fingerprints could correctly classify only 8 out of the 39 mitochondrial toxicants in the external test set; these compounds were at a low structural distance to mitotoxic compounds in the training set (Supplementary Figures S6 and S8). The model using only Cell Painting features could extrapolate well into structurally diverse compounds and correctly predict 29 out of 39 mitotoxic compounds in the external test set but failed when the distance to morphological space was high for example, with compounds 71145-03-4 (Figure 9b), while the same compound was correctly predicted by the model using only

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Morgan fingerprints (which was explicable due to lower structural distance to training data). The late-stage fusion model correctly predicted 32 out of 39 mitotoxic compounds, combining information from both spaces, out of which 4 mitochondrial toxicants were neither correctly predicted by the model using only Cell Painting features, nor by the model using only Morgan fingerprints. Among them were betulinic acid, ketoconazole (which inhibits oxidative phosphorylation[61]) and fluoxetine (which inhibits oxygen consumption and lowers mitochondrial ATP[62]) shown in Figure 9b. These examples demonstrate the synergistic effect of the late-stage fusion model, using information from both the cell morphological as well as the chemical fingerprint space.

Late-stage fusion models accurately detects mitochondrial toxicity of Tox21 compounds labelled as inconclusive

Next, we compared predictions using the 5 models above for the 22 compounds from the data with inconclusive Tox21 assay outcomes due to high cytotoxicity either in the mitochondrial depolarization assay or in the cell viability assay.[63] Literature analysis revealed (further details in Supplementary Table S8) that 4 of the 22 compounds (loratadine, progesterone, ticlopidine and tyrphostin A25) previously have been shown to not cause mitochondrial damage (in fact, progesterone[64] and tyrphostin A25[65] reduce oxidative stress and repair oxidative damage). Another 10 compounds showed some mitochondrial toxicity, such as ketoconazole (inhibitor of oxidative phosphorylation[61]), diflunisal (uncoupler of oxidative phosphorylation[66]), daidzein and fipronil (increase ROS causing mitochondrial depolarization[67,68]). The mitochondrial toxicity for the remaining 8 compounds could not be elucidated further from the literature. Mitochondrial toxicity, like any other compound effect, is concentration- dependent, and the literature evidence compiled as well as the Cell Painting assays whose data was used in this work might hence use different concentrations. Also, the

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

cell line/biological system considered in the literature evidence, the Tox21 assays and the Cell Painting assay can also be very different. With respect to concentration, we explored to what extent the data used here would be predictive for the mitotoxicity endpoint considered, on a relative scale for the different input parameters used in our models; while for the cell line used it has been shown before that Cell Painting yields similar biological phenotypes for different cell lines without cytochemistry protocols requiring specific cell-type-specific optimization.[69] The latter indicates the predictions from the Cell Painting assay may be applicable in detecting toxicity mitochondrial toxicity in another cell line or biological system.

For the 14 compounds for which mitotoxicity annotations were found, Morgan fingerprints correctly predicted toxicity only 1 out of 10 toxic compounds and correctly predicted the non-toxic nature of 3 out the 4 non-toxic compounds (Supplementary Table S9). Thus, Morgan fingerprints showed only very low sensitivity on this dataset. The best performing model, late-stage fusion (averaging predictions from all three models using Cell Painting, Gene Expression features and Morgan fingerprints) however correctly predicted toxicity for all (10 out of 10) mitochondrial toxicants, and correctly predicted the non-toxic nature of 3 out the 4 non-toxic compounds (that is, the increased sensitivity does not come at a cost of a large false positive rate, given that the latter stayed constant between both models). Overall, the late-stage fusion model could hence extrapolate to the morphological space of these inconclusive compounds and detect mitochondrial toxicity even when Tox21 assays reported inconclusive outcomes due to cytotoxicity.

Sensitivity of fusion models are on par with dedicated *in vitro* mitochondrial toxicity assays

Finally, we compared the performance of our models to detect mitochondrial toxicity with dedicated *in vitro* assays. Hallinger *et al.* compared various high throughput screening assays

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and a respirometric screening assay to detect known mitochondrial toxicants (as shown in Supplementary Figure S1 and Supplementary Table S1).[70] When comparing 60 reference chemicals to existing Tox21 assays, they found RSA to be most predictive (balanced accuracy 0.90), while the Tox21 mitochondrial membrane potential assay was also highly predictive (balanced accuracy 0.87). However, respirometric screens have lower throughput than Tox21 assays and are not suitable for screening a large number of compounds. Among other assays they compared were high content imaging assays, where the Apredica HepG2 mitochondrial membrane potential and mitochondrial mass assays[16] were found to be comparatively less predictive (balanced accuracy 0.78 and 0.65, respectively). Although the 236 compounds in the external test set in our study are not the same as these 60 reference chemicals, from a numerical performance comparison we found that our late-stage fusion model achieved higher sensitivity in the external test set (0.82 in our study vs 0.37 in Apredica MitoMass vs 0.8 in RSA) in detecting mitotoxic compounds with comparable balanced accuracies (0.68 in our study vs 0.65 in Apredica MitoMass). The added advantage in using Cell Painting is that it is a comparatively inexpensive single screen that can also be used simultaneously for multiple endpoints for which it is found to be predictive. Hence, we can conclude that the late-stage fusion model based on Cell Painting, Gene Expression and chemical structural data compares well with respect to its predictive power for mitotoxicity to many dedicated assays for this purpose (although precise numerical values cannot be compared due to the different data sets used).

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

DISCUSSION

Mitochondrial toxicity is a leading cause of late-stage drug withdrawals[2] and numerous drugs such as amiodarone, doxorubicin, statins (e.g cerivastatin, simvastatin) and valproic acid have been shown to induce adverse effects *via* mitochondrial dysfunction. Mitochondrial toxicity can be caused by multiple mechanisms and prediction using only chemical structural fingerprints has been shown to be difficult, with respect to extrapolation to novel chemical space, where low model sensitivity has been regularly observed.

To the best of our knowledge, in this work we present the first study combining hypothesis-free high throughput Cell Painting and Gene Expression features with structural fingerprints to predict mitochondrial toxicity. In this work, we confirmed that Cell Painting readouts can discriminate mitotoxic and non-mitotoxic compounds and are able to cluster mitotoxic compounds with shared mode of action (including compounds with inconclusive assay outcomes in Tox21 due to excessive cytotoxicity) in morphological space. This indicates that Cell Painting features are able to detect similarity with respect to both mode of action and mitochondrial toxicity, also in situations of large differences in chemical space. Further, we showed that Cell Painting features correlate to Gene Expression features, which are related to mechanisms of mitochondrial toxicity. We trained late-stage fusion models, which are averaging the results from the Cell Painting, Gene Expression and Morgan fingerprint models, and found those fusion models to be better predictive of mitochondrial toxicity, when extrapolating to new chemical space of an external validation set, compared to the model using only Morgan fingerprints. Also compared to dedicated mitochondrial high content imaging assays our late-stage fusion model showed favourable sensitivity. Given that Cell Painting readouts can be used for multiple purposes, this supports their use also for the prediction of a mitochondrial toxicity endpoint.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Using hypothesis-free data, such as Cell Painting and Gene Expression data, in machine learning models can hence be used to detect toxicity (here mitochondrial toxicity), as well as to help understand modes of toxicity, also in situations where this is not possible based on chemical structure alone. From a predictive modelling perspective, by combining high predictivity of fingerprints in areas of structural space close to the training set with better generalizability of Cell Painting features at greater distances to the training set, such models can contribute to extending the applicability domain of the overall model.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

METHODS

In this work, we followed the workflow as displayed in Supplementary Figure S9 for dataset curation, feature selection and model architecture.

Mitochondrial toxicity dataset preparation

Data for *in vitro* mitochondrial toxicity, used as the endpoint in this study, was collected from the Tox21 assay[40] for mitochondrial membrane potential disruption (MMP) summary assay (AID 720637).[41] This summary assay combines results from a mitochondrial membrane potential (MMP) assay[71] (AID 720635) and a cell viability counter screen[72] (AID 720634) into a binary assay hit call.

For the Tox21 MMP assay, a water-soluble mitochondrial membrane potential sensor was used to evaluate chemically induced mitochondrial toxicity. In healthy cells, this dye accumulates in the mitochondria with red fluorescence. However, should the potential collapse, the dye is no longer able to accumulate in the mitochondria and remains in monomers giving a green fluorescence from the cytoplasm. The cytotoxicity was tested in the same assay well as the mitochondrial potential using a counter cell viability screen. The viability of the cells in the culture was determined by measuring the amount of ATP present. Thus, the summary assay considers not only triplicate runs of the ratio (red/green) readout in the MMP assay but also each fluorescence channel separately, as well as the cytotoxicity results.[63] This helps differentiate compounds that decreased MMP from those inducing high cytotoxicity (thus labelled inconclusive) although caution must be exercised in interpreting the inconclusive compounds as mitochondrial dysfunction may also have caused excessive cell death. Hence, in our models, inconclusive compounds were removed and for the remaining compounds, mitochondrial toxicity labels were assigned as per assay hit calls from the Tox21 assay.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Multiple endpoints, such as mitochondrial membrane depolarisation, can be related to mitochondrial toxicity as can be increased ROS or alteration of energy homeostatic, especially if the membrane potential is depolarised. Hemmerich *et al.*[19] compiled various mitochondrial membrane potential as well as additional assays relevant to mitochondrial toxicity (mitochondria, mitochondria potential and mitochondria complex) from various sources, including CHEMBL, PubChem and 246 drugs tested by Zhang *et al.*[17] into a single dataset related to mitochondrial function, binding and inhibition consisting of 824 mitochondrial toxicants and 4937 non-toxic compounds. To evaluate our models, we further used an external test set using compounds (that do not appear in the training data) from this dataset.

Cell Painting features

Image-based morphological features were extracted from the Cell Painting assay experiments in Bray *et al*.[29] These experiments contained perturbations from 30000 chemicals (around 10,000 small molecules, 2200 drugs and 18000 novel compounds from diversity-oriented synthesis) using DMSO neutral control, USO2 cells in 384-well plates in 5 channels staining eight cellular organelles: nucleus, endoplasmic reticulum, F-actin cytoskeleton, Golgi apparatus, plasma membrane, mitochondria, cytoplasmic RNA and the nucleoli. We obtained consensus morphological features for each compound using the following procedure similar to Lapins *et al*.[32] For each plate, the average feature value from the DMSO plates was subtracted from the perturbation's average feature value. Next, we calculated the median feature value for each compound and dose combination. For replicates, we used the median feature values for doses that were within one standard deviation of the mean dose. The concentration was also included as a feature. Features known to be noisy and generally unreliable as recommended by Way *et al.* were removed.[73] When using Cell Painting features, we avoided compounds having low cell count (step 1, Supplementary Figure S9) in Cell Painting images by removing

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

compounds using "Cells Number Object Number" lesser than -15. Thus, consensus morphological profiles consisting of 1729 numerical features were obtained.

Gene Expression features

The Gene Ontology initiative aims to unify gene and gene product attributes in a classification effort that will provide biological and functional interpretation of gene expression data.[27] They also ensure that genes are consistently annotated across different available datasets. The Gene Expression features used in this work have been derived from transcriptomic data from LINCS L1000. LINCS L1000 gene expression technology profiles changes in 978 landmark genes on perturbations of compounds for a variety of human cell lines.[22] In this work, Gene Ontology transformed Gene Expression features were extracted from the http://maayanlab.net/SEP-L1000/#download which contained 4438 annotated Gene Expression features corresponding to 19803 distinct compounds.[39] The authors used quantile-normalized gene expression profiles from the LINCS L1000 dataset for all replicates of each compound. For each compound, the strongest signatures were used irrespective of the cell line, concentration, or time point which minimizes the number of features required. Gene expression signatures for each compound perturbation was computed using the Characteristic Direction (CD) method[74] on 978 measured hallmark genes. Further, they computed enrichment p-values for each CD signature in the space of all genes against gene set libraries (including biological processes, cellular components, and molecular functions, as well as other gene set libraries accessible from the Enrichr tool[75]) using an extension of the CD technique called Principal Angle Enrichment Analysis (PAEA).[76] We used these annotations for each GE-perturbation combination for further analysis.

Dataset curation and collation

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

We calculated the intersection between mitochondrial toxicity and Cell Painting and Gene Expression datasets described above using standard InChI calculated using RDKit.[77] For conflicting replicates, we considered a compound toxic, if it was detected to be mitotoxic at least once (since in such situations evidence for mitochondrial toxicity, at least under some conditions, exists). We obtained 830 distinct compounds (161 mitotoxic, 61 inconclusive and remaining non-toxic) from the mitochondrial toxicity dataset overlapped with Gene Expression features and a total of 513 distinct compounds (82 mitotoxic and 27 inconclusive) in overlap with Cell Painting features. Similarly, we found 404 distinct compounds (62 mitotoxic and 22 inconclusive) in overlap with both Cell Painting features and Gene Expression features. For the external test set, after adding required annotations of Cell Painting features and Gene Expression features, removing compounds with low cell count, and ensuring no compounds from this was used in feature selection or training our models, a total of 236 distinct compounds (39 mitotoxic and remaining nontoxic) remained in the external test set.

Structural fingerprints

For modelling purposes, we used Morgan fingerprints which contain structural information about compounds and have been successfully used before for toxicity prediction.[78] The MolVS standardizer, an open-source tool based on RDKit[77], was used to standardize (including tautomer standardization) and canonicalize SMILES of the parent molecules.[79] This involved sanitization, normalisation, greatest fragment chooser, charge neutralisation, tautomer enumeration, and canonicalization as implemented in the MolVS tool and described in the MolVS standardizer. We calculated Morgan fingerprints of radius 2 and 2048 bits from standardized SMILES using RDKit.[77]

Feature selection

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

For each of the Cell Painting features and Gene Expression features, standardized values for compounds in the training set were separately subjected to three statistical tests, namely, two-sample Kolmogorov Smirnov test (KS test)[80], Mann–Whitney U test[81] (MWU test) and Point-Biserial correlation[82] (PBS correlation). While the Random Forest algorithm employed for modelling (see below) is in principle able to select features, this is not always successful which made us compare different explicit feature selection methods in parallel with inputting all features into the models subsequently. This method in our experience led to less overfitting while still being interpretable and also able to extrapolate to the external test set compared to other methods (Principal Component Analysis, Maximum Relevance — Minimum Redundancy or using all features, see Supplementary Table S10 for modelling results when comparing different feature selection methods).

After removing the inconclusive compounds (step 2, Supplementary Figure S9), the feature selection (step 3, Supplementary Figure S9) was performed on Cell Painting features and Gene Expression annotations for the remaining compounds. For cross-validation of models, the overlap of the top 40 negatively and 40 positively correlated features from the MWU test and PBS correlation and top 40 correlated features from the KS test were selected for further modelling. For evaluating the external test where more data was available for training, we selected the top 25 correlated features from each test (both positively and negatively for MWU and PBS) and obtained 110 Cell Painting features and 102 Gene Expression features.

Comparing class separation and visualization of compounds in morphological space

For a comparison of intra-class (Toxic vs Toxic) and inter-class (Toxic vs Nontoxic) in morphological space, we used for 486 compounds (85 mitotoxic) for which Cell Painting annotations were available. We randomly resampled the majority class (non-toxic compounds) to match the number of samples of the minority class to ensure our comparisons are equivalent.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Then we visualised mean Tanimoto similarity, median positive image correlation (considering only positive Pearson correlations) and absolute median image correlation (considering the absolute value of median both positive and negative Pearson correlations) for various values of k in k-nearest neighbours in four quartiles of the distribution for intra- and inter-class pairwise distributions (Supplementary Figure S10*)*. We found better separation between intra- and inter-class pairwise when using the absolute value of the median values from the most 15 positively and 15 negatively pairwise Pearson correlations of Cell Painting features. For visualizing the same in structural space, we used the median of 5 highest pairwise Tanimoto similarity of Morgan Fingerprints.[83,84] The methodology was followed when comparing test to train set distances as defined in subsection *"Extrapolation to New Structural/Morphological Space"*.

For visualization of compounds in morphological space, we analysed the 110 selected Cell Painting features on 513 distinct compounds (85 mitotoxic and 27 inconclusive). We normalized 110 selected Cell Painting features and performed Principal Component Analysis using DataWarrior[85] which compared to other nonlinear methods is more interpretable.

Correlation between Cell Painting and Gene Expression features and their positive predictive values

To determine the correlation between selected Cell Painting features and Gene Expression features for compounds exhibiting mitochondrial toxicity, we used Pearson correlation using pandas python package.[86] Comparing the negative logarithmic p-value and the effect size, we determined which Cell Painting features were significantly correlated to specific Gene Expression features related to unfolded protein response, endoplasmic reticulum stress, T cell apoptotic process, side of membrane etc.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Random Forests are not able to detect feature importance when several features are correlated as the Gini index tends to dilute over different features in different trees. To evaluate an individual feature's importance, we used the positive predictive value (PPV) from single decision tree classifiers trained on individual Cell Painting features. These classifiers were trained on 486 compounds (85 mitotoxic) having max depth one and two leaf nodes on our dataset for each feature. The tree hence determines an optimal threshold per feature to distinguish mitochondrial toxic compounds from non-toxic compounds. The mean PPV of Cell Painting features having PPV>0 was grouped by compartment (Cells, Cytoplasm, and Nuclei), channel (AGP, Nucleus, ER, Mito, Nucleolus/Cytoplasmic RNA), and feature group (Correlation, Granularity, Intensity, Radial Distribution, Texture). The predictive value of individual Gene Expression features was computed in a similar manner using decision tree classifiers on 768 compounds (161 mitotoxic).

Types of features' combinations used

Here we employed 5 types of models having different input features, combinations thereof as well as model ensembling. Initially, Cell Painting features, Gene Expression features and Morgan fingerprints were used separately as features for three separate models. As shown in Supplementary Figure S11, an early-stage models fused Cell Painting, Gene Expression and Morgan fingerprints by appending the features into a single vector while another a late-stage fusion model averaged the probabilities of the three models using only Cell Painting, Gene Expression and Morgan fingerprints respectively into a single probability value.

Model generation and evaluation

382 compounds (62 mitotoxic) from the mitochondrial toxicity data having both Cell Painting and Gene Expression annotations were used for modelling. Random Forest models were trained using scikit-learn.[87]

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

As shown in Supplementary Figure S12, a grid search with 4-fold cross-validation revealed 100 trees were optimal for achieving high balanced accuracies; hence we set Random Forests parameters with 100 trees. The nodes were let to expand until all leaves were pure or until all leaves contained less than a minimum of 2 samples that are required to split an internal node. A minimum of 1 sample was required to be at a leaf node and the number of features to consider when looking for the best split was set as the square root of total features. The consistent performance is most likely as Random Forests are usually robust against overfitting. As shown in Supplementary Figure S13, we used 4-fold nested cross-validation; inside the outer loop, a 4-fold stratified splitting divided the data into a training set (75%), on which feature selection was performed and the remaining into a test set (25%). Inside the inner loop, a Random Forest model with parameters as above was trained on the training set using 4-fold stratified cross-validation. For each model, to account for class imbalance, we tuned the threshold of probability to determine the cut-off for toxicity labels having maximum value for Youden's J statistic (J = True Positive Rate – False Positive Rate). The Youden index is frequently used to detect an optimal threshold to be used as a criterion for classifying subjects without biasing the model towards one class. Thus, the predictions can be used to fully exploit the model giving equal weights to sensitivity and specificity without favouring one of them. From combined results of the out-of-fold data from cross-validation, we chose the threshold of probability with the largest Youden's J statistic value. This threshold was then used for the test set (hence the test set was not used directly while selecting the optimal threshold). The entire process of nested cross-validating was repeated 50 times; we evaluated our models on the distribution and median of the performance metrics from all 200 test sets. The models overall trained with reasonable training time and threshold balancing ensured that overfitting on an unbalanced dataset could be avoided.

External model evaluation

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

For the external test set, we trained 5 Random Forest models for each feature/combination on our dataset using 4-fold cross-validation and the optimal threshold was determined similarly from the combined out-of-fold data. The model was then retrained on the entire dataset and used to predict the external test set with the threshold previously determined.

Evaluation metrics

F1 scores of the minority class (mitotoxic compounds), precision of the minority class (mitotoxic compounds), sensitivity, specificity, Balanced Accuracy (BA), Area Under Curve-Receiver Operating Characteristic (AUC-ROC), Area Under Curve-Precision Recall (AUCPR), and Mathew's correlation constant (MCC) were used to assess model performance as implemented in scikit-learn python package.[87] Often in a toxicity prediction problem with unbalanced data, the number of nontoxic compounds far outweigh the number of mitotoxic compounds and improvement in the prediction of the mitotoxic compounds (minority class) is desired.[88] Here particular metrics such as sensitivity and AUCPR are useful and less likely to exaggerate model performance. For comparing model predictions to true values in the external test set, F1 scores and precision of the minority class and the sensitivity of the model were used as they focus on the minority class (mitotoxic compounds) being detected by the model.

Extrapolation to new structural/morphological space

To evaluate if our models can extrapolate to novel chemical space (either in structural space or in morphological space) we defined for each compound in the external test set two parameters:

(1) Structural distance to the training set:

$$Structural\ distance\ = median\ (x_1\ , x_2\ \dots , x_{15}\ )$$

where x_k= pairwise Tanimoto distances in decreasing magnitude, where,

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$Tanimoto\ distance\ = 1 - Tanimoto\ similarity$$

and (2) Morphological distance to the training set:

$$Morphological\ distance\ = 1 - Morphological\ similarity$$

where,

$$Morphological\ similarity\ = abs\ (\ median(\ a_1\ , a_2 \dots a_{15}\ , b_1\ , b_2 \dots b_{15}))$$

and a_k = positively and b_k = negatively pairwise Pearson correlations in decreasing magnitude.

The distances were defined same as the subsection *"Statistical Analysis of Cell Painting features and Gene Expression features"*. The structural distance was defined as the median of the five lowest Tanimoto distances[89] between Morgan fingerprints of the test compound and the compounds in the training dataset of the same activity annotation. The morphological distance was defined as the one minus the absolute value of median of 15 most positively and 15 most negatively pairwise Pearson correlations (using selected Cell Painting features) of the test compound and the compounds in the training dataset of the same activity annotation. In this manner, we could evaluate if true positives from test sets for each model lie in relatively distant structural or morphological space to their training space.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475326; this version posted January 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

CODE AVAILABILITY

We released the python code for our models which are publicly available at *https://git.io/JDGyc*

DATA AVAILABILITY

The training dataset (used for nested cross validation) and the external test set used in this study are released in Supplementary Table S11 and S12.

REFERENCES

1. Dykens, J. A., Marroquin, L. D. & Will, Y. Strategies to reduce late-stage drug attrition due to mitochondrial toxicity. *Expert Rev. Mol. Diagn.* **7**, 161–175 (2007).

2. Dykens, J. A. & Will, Y. The significance of mitochondrial toxicity testing in drug development. *Drug Discov. Today* **12**, (2007).

3. Varga, Z. V., Ferdinandy, P., Liaudet, L. & Pacher, P. Drug-induced mitochondrial dysfunction and cardiotoxicity. *Am. J. Physiol. Heart. Circ. Physiol.* **309,** H1453–H1467 (2015).

4. Begriche, K., Massart, J., Robin, M. A., Borgne-Sanchez, A. & Fromenty, B. Drug-induced toxicity on mitochondria and lipid metabolism: Mechanistic diversity and deleterious consequences for the liver. *J. Hepatol.* **54**, 773–794 (2011).

5.Will, Y., Shields, J. E. & Wallace, K. B. Drug-induced mitochondrial toxicity in the geriatric population: Challenges and future directions. *Biology (Basel)*. **8**, 32 (2019).

6. Meyer, J. N. *et al.* Mitochondria as a target of environmental toxicants. *Toxicol. Sci.* **134**, 1–17 (2013).

7. Chan, K., Truong, D., Shangari, N. & O'Brien, P. J. Drug-induced mitochondrial toxicity. *Expert. Opin. Drug. Metab. Toxicol.* **1**, 655–669 (2005).

8. Eakins, J. *et al*. A combined *in vitro* approach to improve the prediction of mitochondrial toxicants. *Toxicol. Vitr.* **34**, 161–170 (2016).

9. Kamalian, L. *et al.* The utility of HepG2 cells to identify direct mitochondrial dysfunction in the absence of cell death. *Toxicol. Vitr.* **29**, 732–740 (2015).

10. Sakamuru, S. *et al*. Application of a homogenous membrane potential assay to assess mitochondrial function. *Physiol. Genomics* **44**, 495–503 (2012)

11. Salvioli, S., Ardizzoni, A., Franceschi, C. & Cossarizza, A. JC-1, but not DiOC6(3) or rhodamine 123, is a reliable fluorescent probe to assess ΔΨ changes in intact cells: Implications for studies on mitochondrial functionality during apoptosis. *FEBS Lett.* **411**, 77–82 (1997).

12. Scaduto, R. C. & Grotyohann, L. W. Measurement of mitochondrial membrane potential using fluorescent rhodamine derivatives. *Biophys. J.* **76**, 469–477 (1999).

13. Sugrue, M. M., Wang, Y., Rideout, H. J., Chalmers-Redman, R. M. E. & Tatton, W. G. Reduced mitochondrial membrane potential and altered responsiveness of a mitochondrial membrane megachannel in p53-induced senescence. *Biochem. Biophys. Res. Commun.* **261**, 123–130 (1999).

14. Reers, M., Smith, T. W. & Chen, L. B. J-Aggregate Formation of a Carbocyanine as a Quantitative Fluorescent Indicator of Membrane Potential. *Biochemistry* **30**, 4480–4486 (1991).

15. Wills, L. P. The use of high-throughput screening techniques to evaluate mitochondrial toxicity. *Toxicology* **391**, 34–41 (2017).

16. Shah, I. *et al.* Using toxcast^{TM} data to reconstruct dynamic cell state trajectories and estimate toxicological points of departure. *Environ. Health Perspect.* **124**, 910–919 (2016).

17. Zhang, H. *et al.* In silico prediction of mitochondrial toxicity by using GA-CG-SVM approach. *Toxicol. Vitr.* **23**, 134–140 (2009).

18. Zhao, P. *et al.* In silico prediction of mitochondrial toxicity of chemicals using machine learning methods. *J. Appl. Toxicol.* jat.4141 (2021).

19. Hemmerich, J., Troger, F., Füzi, B. & F.Ecker, G. Using Machine Learning Methods and Structural Alerts for Prediction of Mitochondrial Toxicity. *Mol. Inform.* **39**, (2020).

20. Zhang, H. *et al.* Development of novel prediction model for drug-induced mitochondrial toxicity by using naïve Bayes classifier method. *Food Chem. Toxicol.* **110**, 122–129 (2017).

21. Sanuki, Y., Araki, T., Nakazono, O. & Tsurui, K. A rapid mitochondrial toxicity assay utilizing rapidly changing cell energy metabolism. *J. Toxicol. Sci.* **42**, 349–358 (2017).

22. Subramanian, A. *et al*. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437-1452.e17 (2017).

23. Szalai, B. *et al.* Signatures of cell death and proliferation in perturbation transcriptomics data—from confounding factor to effective prediction. *Nucleic Acids Res.* **47**, 10010–10026 (2019).

24. Hizukuri, Y., Sawada, R. & Yamanishi, Y. Predicting target proteins for drug candidate compounds based on drug-induced gene expression data in a chemical structure-independent manner. *BMC Med. Genomics* **8**, 1–10 (2015).

25. Carbonell, P., Lopez, O., Amberg, A., Pastor, M. & Sanz, F. Hepatotoxicity prediction by systems biology modeling of disturbed metabolic pathways using gene expression data. *ALTEX - Altern. to Anim. Exp.* **34**, 219–234 (2017).

26. Liu, R., Yu, X. & Wallqvist, A. Using Chemical-Induced Gene Expression in Cultured Human Cells to Predict Chemical Toxicity. *Chem. Res. Toxicol.* **29**, 1883–1893 (2016).

27. Hill, D. P., Smith, B., McAndrews-Hill, M. S. & Blake, J. A. Gene Ontology annotations: What they mean and where they come from. *BMC Bioinformatics* **9**, S2 (2008).

28. Wawer, M. J. *et al.* Toward performance-diverse small-molecule libraries for cell-based phenotypic screening using multiplexed high-dimensional profiling. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 10911–10916 (2014).

29. 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).

30. Seal, S., Yang, H., Vollmers, L. & Bender, A. Comparison of Cellular Morphological Features and Molecular Fingerprints for the Prediction of Cytotoxicity- And Proliferation-Related Assays. *Chem. Res. Toxicol.* **34**, 422–437 (2021).

31. Hofmarcher, M., Rumetshofer, E., Clevert, D. A., Hochreiter, S. & Klambauer, G. Accurate Prediction of Biological Assays with High-Throughput Microscopy Images and Convolutional Networks. *J. Chem. Inf. Model.* **59**, 1163–1171 (2019).

32. Lapins, M. & Spjuth, O. Evaluation of Gene Expression and Phenotypic Profiling Data as Quantitative Features for Predicting Drug Targets and Mechanisms of Action. *bioRxiv* 580654 (2019) doi:10.1101/580654.

33. Way, G. P. *et al.* Predicting cell health phenotypes using image-based morphology profiling. *Mol. Biol. Cell* **32**, 995–1005 (2021).

34. Trapotsi, M.-A. *et al.* Comparison of Chemical Structure and Cell Morphology Information for Multitask Bioactivity Predictions. *J. Chem. Inf. Model.* **61**, 1444–1456 (2021).

35. Rietdijk, J. *et al*. A phenomics approach for antiviral drug discovery. *BMC Biol.* **19**, 1–15 (2021).

36. Chavan, S., Scherbak, N., Engwall, M. & Repsilber, D. Predicting Chemical-Induced Liver Toxicity Using High-Content Imaging Phenotypes and Chemical Features: A Random Forest Approach. *Chem. Res. Toxicol.* **33**, 2261–2275 (2020).

37. Bender, A. & Cortés-Ciriano, I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 1: Ways to make an impact, and why we are not there yet. *Drug Discov. Today.* **26,** 511–524 (2021).

38. Bender, A. & Cortes-Ciriano, I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 2: a discussion of chemical and biological data. *Drug Discov. Today.* ***26,*** 1040–1052 (2021).

39. Wang, Z., Clark, N. R. & Ma'ayan, A. Drug-induced adverse events prediction with the LINCS L1000 data. *Bioinformatics* **32**, 2338–2345 (2016).

40. Richard, A. M. *et al*. The Tox21 10K Compound Library: Collaborative Chemistry Advancing Toxicology. Chem. Res. Toxicol. **34**, 189–216 (2021).

41. AID 720637 - qHTS assay for small molecule disruptors of the mitochondrial membrane potential: Summary - PubChem BioAssay Summary https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=720637 (accessed Apr 15, 2021).

42. Brum, A. M. *et al.* Connectivity Map-based discovery of parbendazole reveals targetable human osteogenic pathway. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 12711–12716 (2015).

43.Guerini, A. E. *et al.* Mebendazole as a candidate for drug repurposing in oncology: An extensive review of current literature. *Cancers*. **11**, 1284 (2019).

44. Dogra, N., Kumar, A. & Mukhopadhyay, T. Fenbendazole acts as a moderate microtubule destabilizing agent and causes cancer cell death by modulating multiple cellular pathways. *Sci. Rep.* **8**, 11926 (2018).

45. McCormick, B., Lowes, D. A., Colvin, L., Torsney, C. & Galley, H. F. MitoVitE, a mitochondria-targeted antioxidant, limits paclitaxel-induced oxidative stress and mitochondrial damage *in vitro*, and paclitaxel-induced mechanical hypersensitivity in a rat pain model. *Br. J. Anaesth.* **117**, 659–666 (2016).

46. Maldonado, E. N., Patnaik, J., Mullins, M. R. & Lemasters, J. J. Free tubulin modulates mitochondrial membrane potential in cancer cells. *Cancer Res.* **70**, 10192–10201 (2010).

47. Li, Q., Pogwizd, S. M., Prabhu, S. D. & Zhou, L. Inhibiting Na+/K+ ATPase can impair mitochondrial energetics and induce abnormal Ca2+ cycling and automaticity in guinea pig cardiomyocytes. *PLoS One* **9**, e93928 (2014).

48. Ozdemir, T. *et al.* Ouabain targets the unfolded protein response for selective killing of HepG2 cells during glucose deprivation. *Cancer Biother. Radiopharm.* **27**, 457–463 (2012).

49. Abdoli, N., Heidari, R., Azarmi, Y. & Eghbal, M. A. Mechanisms of the statins cytotoxicity in freshly isolated rat hepatocytes. *J. Biochem. Mol. Toxicol.* **27**, 287–294 (2013).

50. Jánosi, J. *et al.* Mevastatin-induced apoptosis and growth suppression in U266 myeloma cells. *Anticancer Res.* **24**, 1817–1822 (2004).

51. Morishima, S., Shibata, M. A., Ohmichi, M. & Otsuki, Y. Raloxifene, a selective estrogen receptor modulator, induces mitochondria-mediated apoptosis in human endometrial carcinoma cells. *Med. Mol. Morphol.* **41**, 132–138 (2008).

52. Liao, C. H., Guh, J. H., Chueh, S. C. & Yu, H. J. Anti-angiogenic effects and mechanism of prazosin. *Prostate* **71**, 976–984 (2011).

53. Bhat, T. A. *et al.* Endoplasmic reticulum-mediated unfolded protein response and mitochondrial apoptosis in cancer. *Biochim. Biophys. acta. Rev. cancer* **1867**, 58–66 (2017).

54. Vringer, E. & Tait, S. W. G. Mitochondria and inflammation: Cell death heats up. *Front. Cell Dev. Biol.* **7**, 100 (2019).

55. Malhotra, J. D. & Kaufman, R. J. ER stress and Its functional link to mitochondria: Role in cell survival and death. *Cold Spring Harb. Perspect. Biol.* **3**, 1–13 (2011).

56. Szegezdi, E., Logue, S. E., Gorman, A. M. & Samali, A. Mediators of endoplasmic reticulum stress-induced apoptosis. *EMBO Rep.* **7**, 880 (2006).

57. Liu, D. *et al.* Intercellular mitochondrial transfer as a means of tissue revitalization. *Signal Transduct. Target Ther.* **6**, 1–18 (2021).

58. Arciuch, V. G. A., Elguero, M. E., Poderoso, J. J. & Carreras, M. C. Mitochondrial regulation of cell cycle and proliferation. *Antioxidants Redox Signal.* **16**, 1150–1180 (2012).

59. Schmidt, B. Z. *et al. in vitro* acute and developmental neurotoxicity screening: an overview of cellular platforms and high-throughput technical possibilities. *Arch. Toxicol.* **91**, 1–33 (2017).

60. Babayev, E. & Seli, E. Oocyte mitochondrial function and reproduction. *Current Opinion in Obstet. Gynecol.* **27**, 175–181 (2015).

61. Rodriguez, R. J. & Acosta, D. Inhibition of mitochondrial function in isolated rat liver mitochondria by azole antifungals. *J. Biochem. Toxicol.* **11**, 127–131 (1996).

62. Charles, E. *et al.* The antidepressant fluoxetine induces necrosis by energy depletion and mitochondrial calcium overload. *Oncotarget* **8**, 3181–3196 (2017).

63. Attene-Ramos, M. S. *et al.* Profiling of the Tox21 Chemical Collection for Mitochondrial Function to Identify Compounds that Acutely Decrease Mitochondrial Membrane Potential. *Environ. Health Perspect.* **123**, 49–56 (2015).

64. Irwin, R. W. *et al.* Progesterone and estrogen regulate oxidative metabolism in brain mitochondria. *Endocrinology* **149**, 3167–3175 (2008).

65. Sagara, Y., Ishige, K., Tsai, C. & Maher, P. Tyrphostins protect neuronal cells from oxidative stress. *J. Biol. Chem*. **277**, 36204–36215 (2002).

66. Chávez, E., Bravo, C., Gil, H. A. & Reyes, P. A. Ionophoretic-like action of diflunisal. *Life Sci.* **37**, 1491–1498 (1985).

67. Jin, S., Zhang, Q. Y., Kang, X. M., Wang, J. X. & Zhao, W. H. Daidzein induces MCF-7 breast cancer cell apoptosis via the mitochondrial pathway. *Ann. Oncol.* **21**, 263–268 (2010).

68. Seydi, E., Mehrpouya, L., Sadeghi, H., Rahimi, S. & Pourahmad, J. Toxicity of fipronil on rat heart mitochondria. *Toxin Rev.* 1–9 (2019) doi:10.1080/15569543.2019.1700382.

69. Willis, C., Nyffeler, J. & Harrill, J. Phenotypic Profiling of Reference Chemicals across Biologically Diverse Cell Types Using the Cell Painting Assay. *SLAS Discov.* **25**, 755–769 (2020).

70. 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).

71. AID 720635 - qHTS assay for small molecule disruptors of the mitochondrial membrane potential - PubChem. https://pubchem.ncbi.nlm.nih.gov/bioassay/720635. (accessed Apr 15, 2021)

72. AID 720634 - qHTS assay for small molecule disruptors of the mitochondrial membrane potential - cell viability - PubChem. https://pubchem.ncbi.nlm.nih.gov/bioassay/720634. (accessed Apr 15, 2021)

73. Way G. 1. Blocklist Features - Cell Profiler. https://figshare.com/articles/dataset/Blacklist_Features_-_Cell_Profiler/10255811. (accessed Apr 11, 2021).

74. Clark, N. R. *et al.* The characteristic direction: A geometrical approach to identify differentially expressed genes. *BMC Bioinformatics* **15**, 1–16 (2014).

75. Chen, E. Y. *et al.* Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 1–14 (2013).

76. Clark, N. R. *et al.* Principle Angle Enrichment Analysis (PAEA): Dimensionally reduced multivariate gene set enrichment analysis tool. in *Proceedings - 2015 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2015* 256–262 (Institute of Electrical and Electronics Engineers Inc., 2015). doi:10.1109/BIBM.2015.7359689.

77. RDKit: Cheminformatics and Machine Learning Software http://www.rdkit.org. (Accessed 22 June 2021)

78.Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **50**, 742–754 (2010).

79. Swain, M. MolVS: Molecule Validation and Standardization — MolVS 0.1.1 documentation. https://molvs.readthedocs.io/en/latest/ (2016). (accessed Apr 15, 2021).

80. Massey, F. J. The Kolmogorov-Smirnov Test for Goodness of Fit. *J. Am. Stat. Assoc.* **46**, 68–78 (1951).

81. Wald, A. & Wolfowitz, J. On a Test Whether Two Samples are from the Same Population. *Ann. Math. Stat.* **11**, 147–162 (1940).

82. Kornbrot, D. Point Biserial Correlation. in *Encyclopedia of Statistics in Behavioral Science* (John Wiley & Sons, Ltd, 2005). doi:10.1002/0470013192.bsa485.

83. Holliday. Grouping of Coefficients for the Calculation of Inter-Molecular Similarity and Dissimilarity using 2D Fragment Bit-Strings. *Comb. Chem. High Throughput Screen.* **5**, (2002).

84. Bender, A. *et al.* How similar are similarity searching methods? A principal component analysis of molecular feature space. *J. Chem. Inf. Model.* **49**, 108–119 (2009).

85. Sander, T., Freyss, J., Von Korff, M. & Rufener, C. DataWarrior: An open-source program for chemistry aware data visualization and analysis. *J. Chem. Inf. Model.* **55**, 460–473 (2015).

86. pandas 1.3.1 documentation.

https://pandas.pydata.org/pandas-docs/stable/reference/index.html. (accessed Jul 29, 2021).

87. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12,** http://scikit-learn.sourceforge.net. (2011).

88. Krawczyk, B. Learning from imbalanced data: open challenges and future directions. *Prog. Artif. Intell.* **5**, 221–232 (2016).

89. Bajusz, D., Rácz, A. & Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.* **7**, 1–13 (2015).

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, and Trinity Henry Barlow (Trinity College) for providing funding for this study. S.S. gratefully acknowledge discussions with Prof. Michele Vendruscolo (University of Cambridge). 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). Cartoons in Figures 1, 2, 4-7 were created with Bioicons (https://bioicons.com) which compiled images from the Database Center for the life sciences/TogoTV (https://togotv.dbcls.jp/en/pics.html) and Servier (https://smart.servier.com)

AUTHOR INFORMATION

All authors have given approval to the final version of the manuscript.

Affiliations

Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Rd, CB2 1EW, Cambridge, United Kingdom

Srijit Seal, Maria-Anna Trapotsi, Hongbin Yang, Andreas Bender.

Department of Pharmaceutical Biosciences and Science for Life Laboratory, Uppsala University, Box 591, SE-75124, Uppsala, Sweden

Jordi Carreras-Puigvert, Ola Spjuth.

Contributions

S.S. designed and performed exploratory data analysis, implemented, and trained the models. S.S, J.C.P and O.S analysed the biological interpretation of morphological features. S.S, M.A.T and H.Y contributed to analysing the results of models. S.S. wrote the manuscript with extensive discussions with O.S and J.C.P. A.B. and O.S. supervised the project. All the authors (S.S., J.C.P, M.A.T, H.Y, O.S and A.B) reviewed, edited, and contributed to discissions on the manuscript.

Corresponding authors

* Dr Andreas Bender, Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom.

ab454@cam.ac.uk

*Ola Spjuth, Department of Pharmaceutical Biosciences and Science for Life Laboratory, Uppsala University, Box 591, SE-75124, Uppsala, Sweden

ola.spjuth@farmbio.uu.se

ETHICS DECLARATIONS

Competing interests

The authors declare no competing interests.

SUPPLEMENTARY INFORMATION

Supporting Tables (XLSX)

Supporting Figures (PDF)

ABBREVIATIONS

AUC-ROC, Area Under Curve- Receiver Operating Characteristic; AUC-PR, Precision-Recall Area Under Curve; AGP, Actin, Golgi, Plasma membrane; BA, Balanced Accuracy; BMF, Bayesian Matrix Factorisation; DMSO, Dimethyl sulfoxide; DNA, Deoxyribonucleic acid; FN: Number of False Negatives; FP: Number of False Positives; KS, Kolmogorov-Smirnov; MACCS, Molecular ACCess System ;MCC, Mathew's Correlation constant; MMP, Mitochondrial Membrane Potential; MOA, Mechanism of Action; mtDNA, mitochondrial DNA; PCA, Principal component analysis; ROS, Reactive Oxygen Species; RSA, Respirometric Screening Assay; RF, Random Forest; RNA, Ribonucleic acid; SMILES, Simplified Molecular-Input Line-Entry System ; TN: Number of True Negatives; TP: Number of True Positives

FIGURES

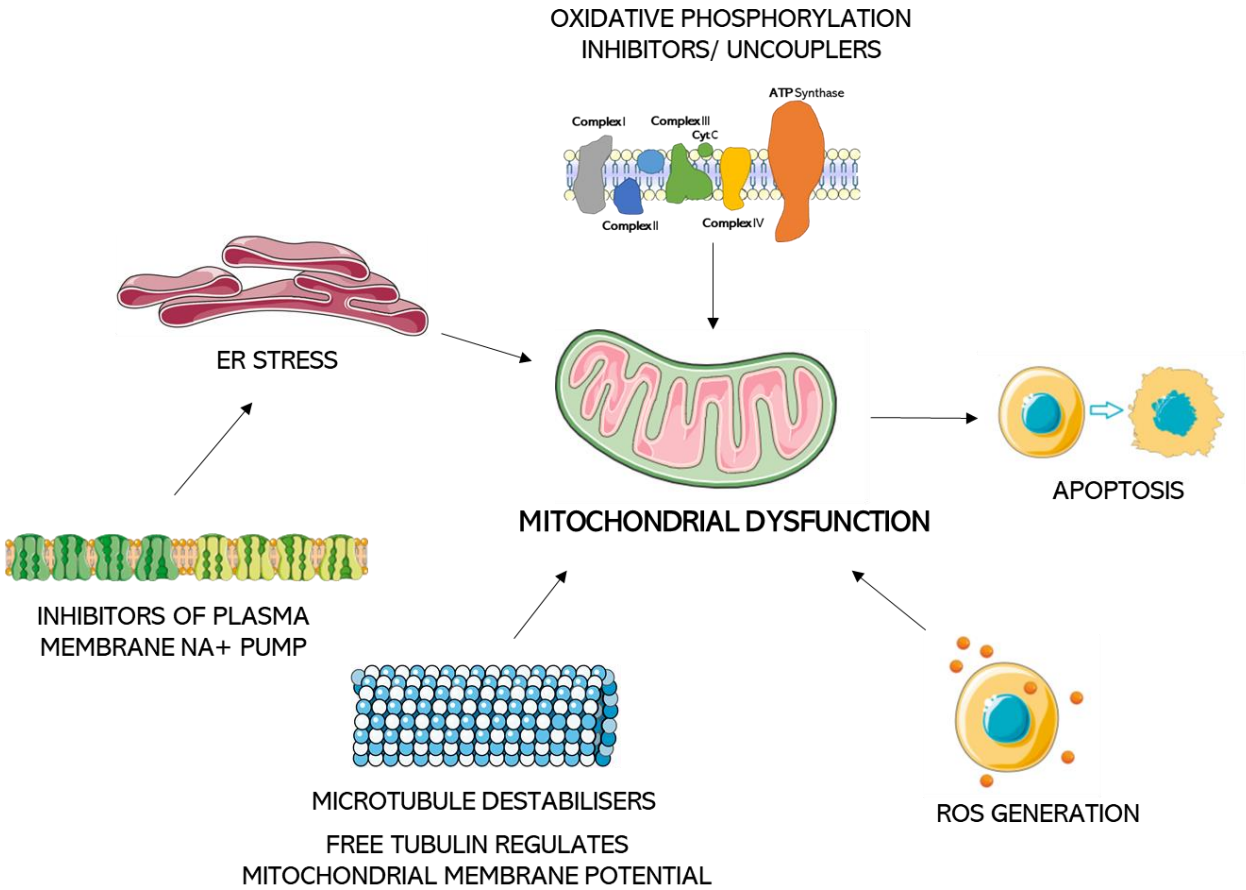


Figure 1. Major mechanisms of mitochondrial toxicants. Toxicants act on multiple pathways to exhibit mitochondrial toxicity, mostly inhibition of mitochondrial respiratory chain or uncoupling of oxidative phosphorylation, oxidative stress from responses including generation of reactive oxygen species (ROS), microtubule disruption and ER stress from various responses including inhibition of Na^+ pumps etc.

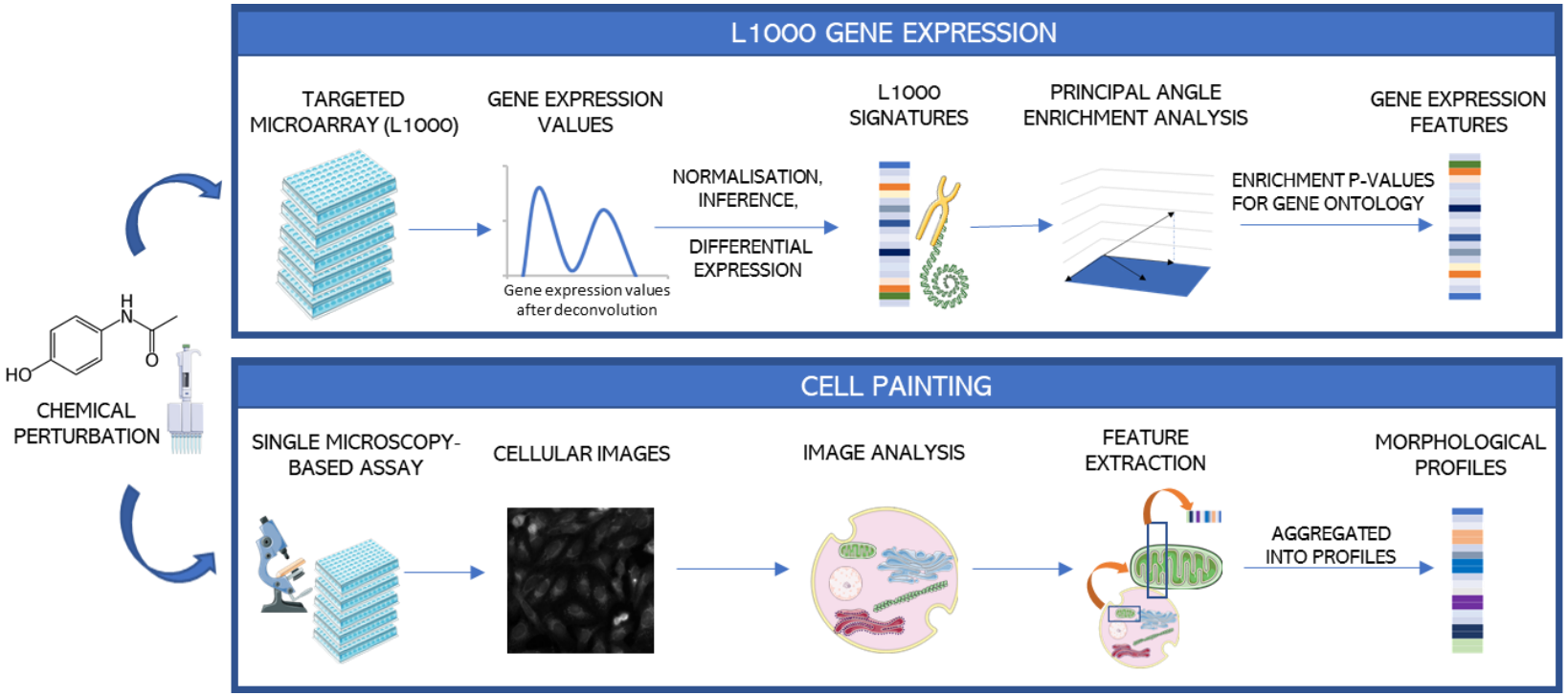


Figure 2. Overview of the workflow employed in this study. L1000 technology for Gene Expression and Cell Painting Technology for cell morphology statistics. The LINCS L1000 gene expression technology profiles changes in 978 landmark genes before and after chemical perturbations on different human cell lines. Raw unprocessed flow cytometry data from Luminex is converted to quantile-normalized gene expression profiles for all replicates of each compound. We use Gene Expression data from Wang *et al*[39], who for each compound, computed the strongest gene expression signature using the Characteristic Direction (CD) method and computed enrichment p-values for each CD signature in the space of all genes against gene set libraries using Principal Angle Enrichment Analysis (PAEA). The Cell Painting assay, on the other hand, captures cellular morphological changes in the form of numerical statistics which are converted from microscopic image data of cells treated with chemical perturbations.

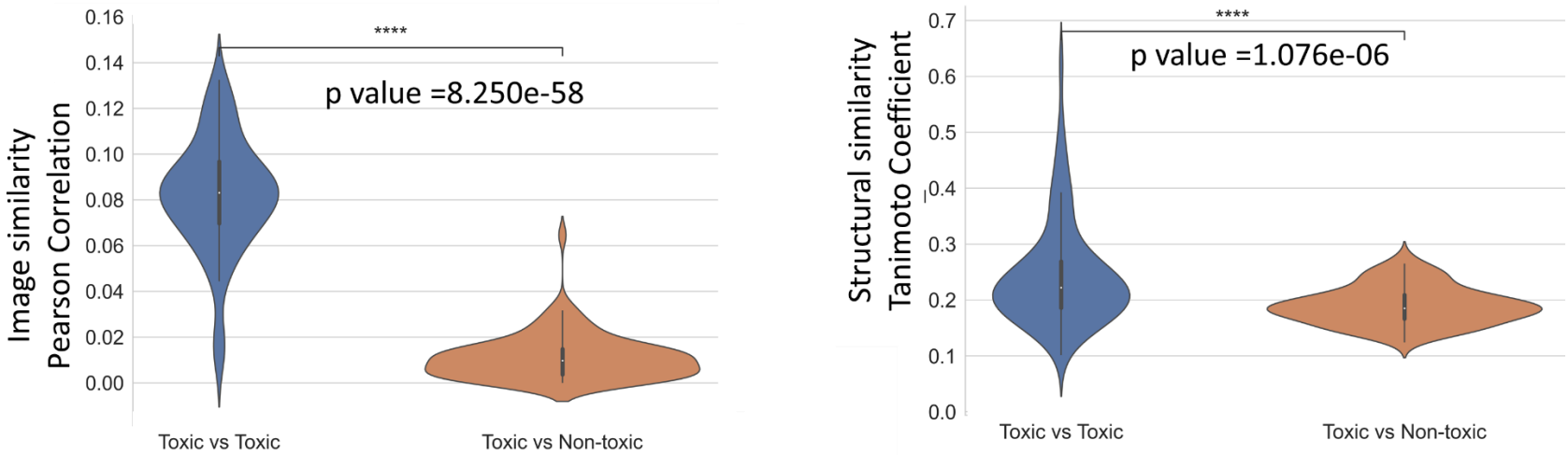


Figure 3. Intra- and inter-class pairwise similarity for 486 compounds (85 mitotoxic)
Median of five highest Tanimoto similarity coefficients of Morgan fingerprints and (B) Absolute value of the median of fifteen most positively and fifteen most negatively Pearson correlation effect sizes of selected 110 Cell Painting features for mitochondrial toxic and non-toxic compounds.
Mitotoxic compounds considerably vary from non-toxic compounds in morphological space (median Pearson correlation of 0.140 vs 0.038, t-test independent p value=3.301e-20) while also varying in structural space (median Tanimoto Similarity of Morgan fingerprints 0.208 vs 0.183, t-test independent samples p value= 6.329e-03).

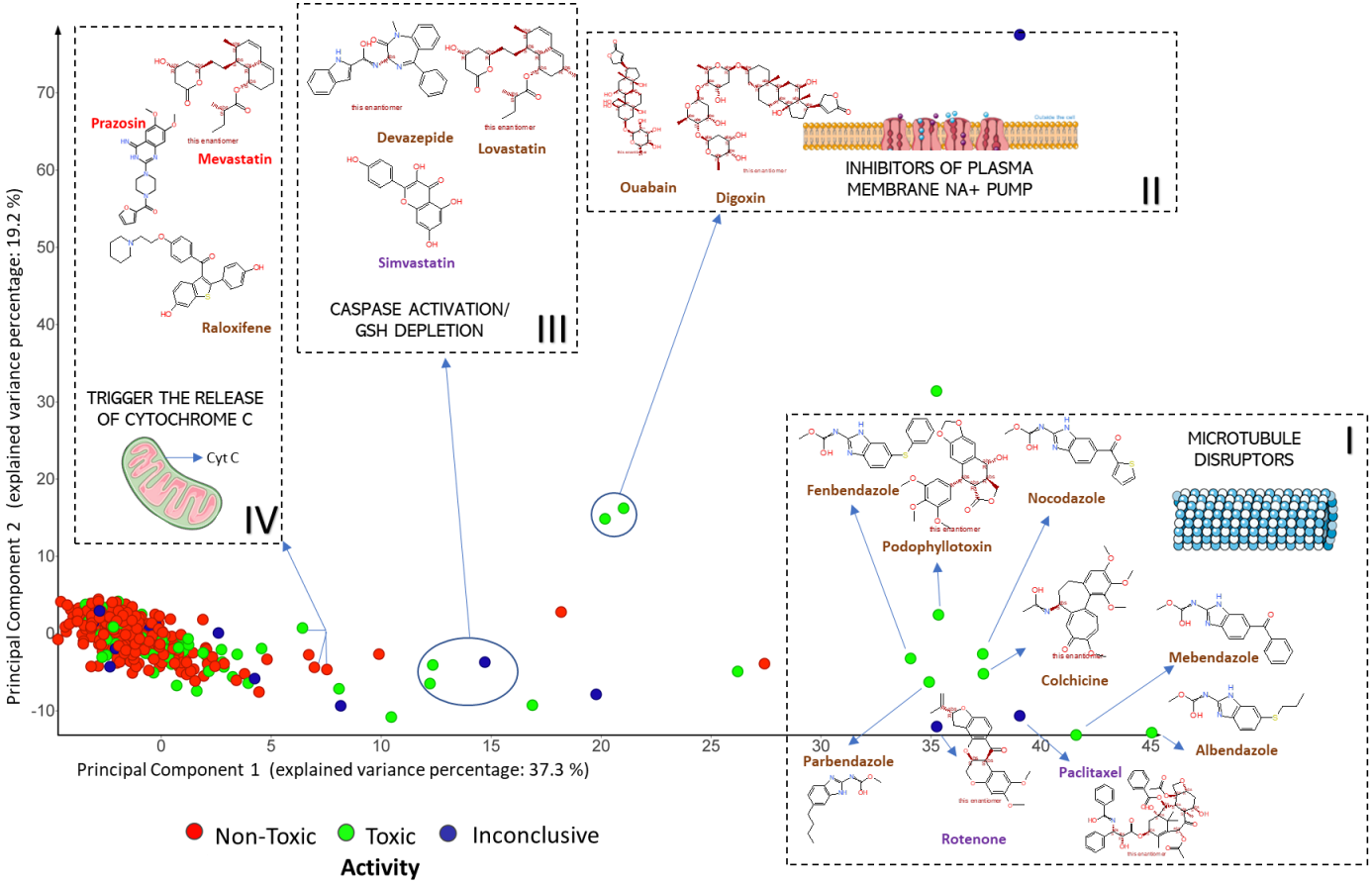


Figure 4. Principal Component Analysis of 542 compounds in 110-dimensional Cell Painting feature space. Certain compounds clustered further away from the distribution of majority of compounds having similar mechanisms of actions such as of microtubule destabilizers, or compounds inducing apoptotic signalling cascades, compounds causing oxidative stress due to GSH depletion or those that are inhibitors of plasma membrane Na+ pump (all of which reduce mitochondrial membrane potential). Cluster I (microtubule destabilisers): rotenone, albendazole, parbendazole, mebendazole, nocodazole, fenbendazole, colchine, paclitaxel and podophyllotoxin; Cluster II (inhibitors of plasma membrane Na+ pump): ouabain and digoxin; Cluster III (caspase activation and GSH depletion): devazepide, lovastatin, simvastatin; and Cluster II (trigger the release of cytochrome c into the cytosol): mevastatin, prazosin, and raloxifene.

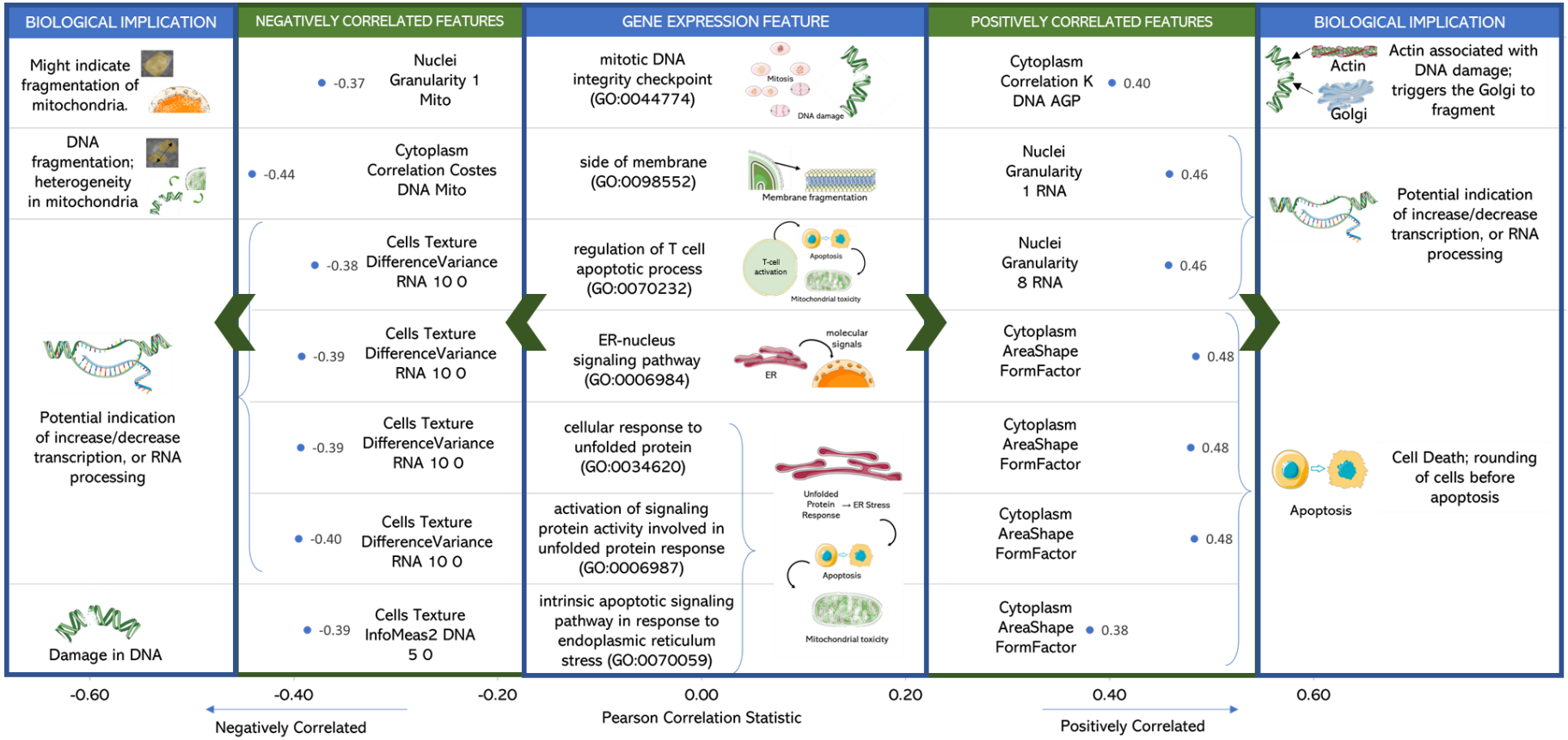


Figure 5. Computational significance and biological implication in mitochondrial toxicity of the Cell Painting features that are most positively or negatively correlated to Gene Expression descriptors particularly unfolded protein response and endoplasmic reticulum stress (RNA variance and cell area shape), T cell apoptotic processes (mitochondrial granularity and DNA fragmentation) and side of the membrane (RNA granularity and heterogeneity in mitochondria). Further details in Supplementary Table S3.

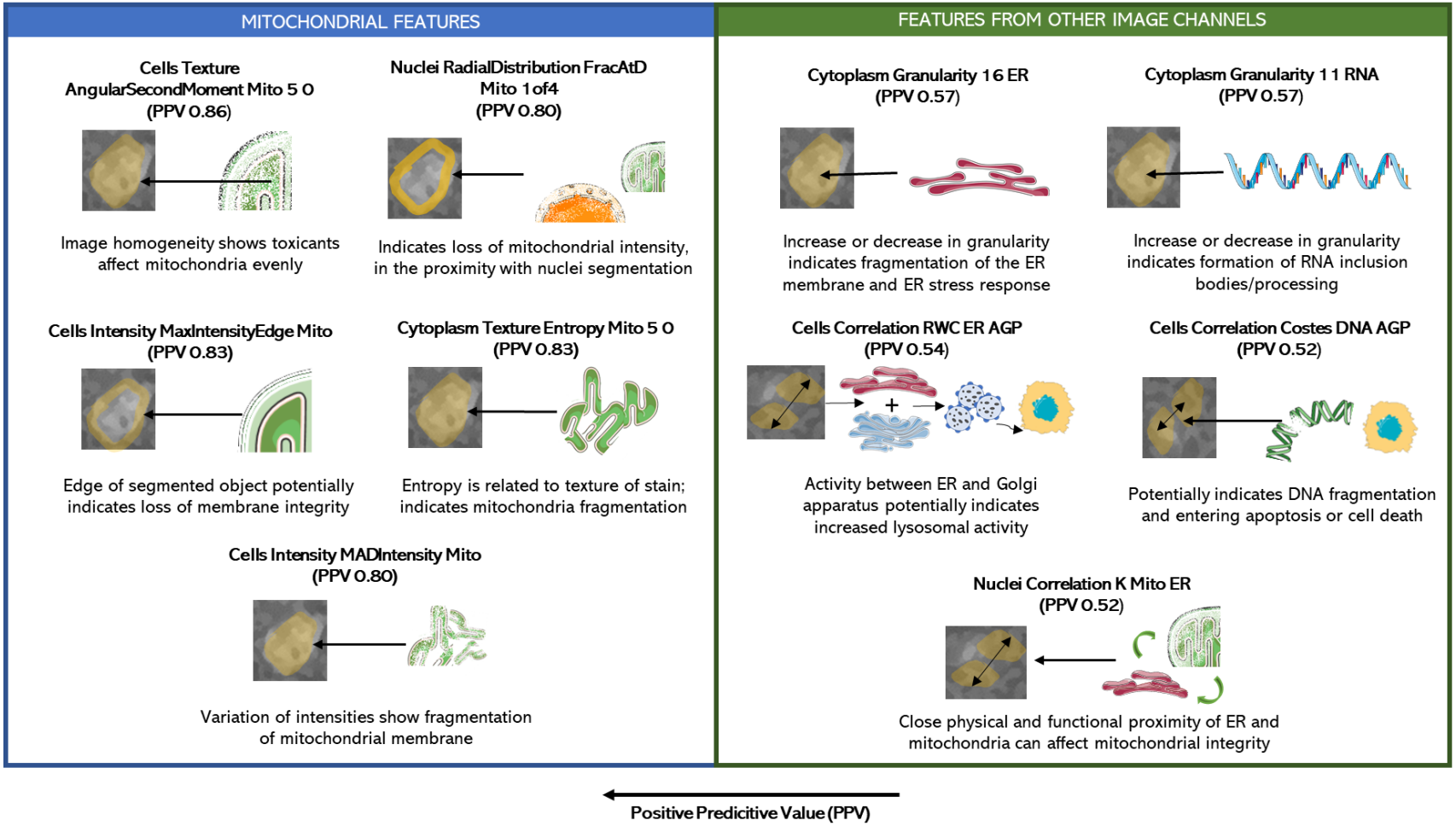


Figure 6. Biological implication of mitochondrial toxicity translated from the computational image statistics of Cell Painting features. Features were mainly related to edge intensity of cells (possibly related to integrity of cell wall), radial distribution and intensity in mitochondria (related to mitochondrial death) and granularity features (related to cell death and amount of information contained in cellular images). Further details in Supplementary Table S4. AGP: Actin Golgi Plasma membrane, DNA: Deoxyribonucleic acid, ER: Endoplasmic Reticulum, Mito: Mitochondria, RNA: Ribonucleic acid

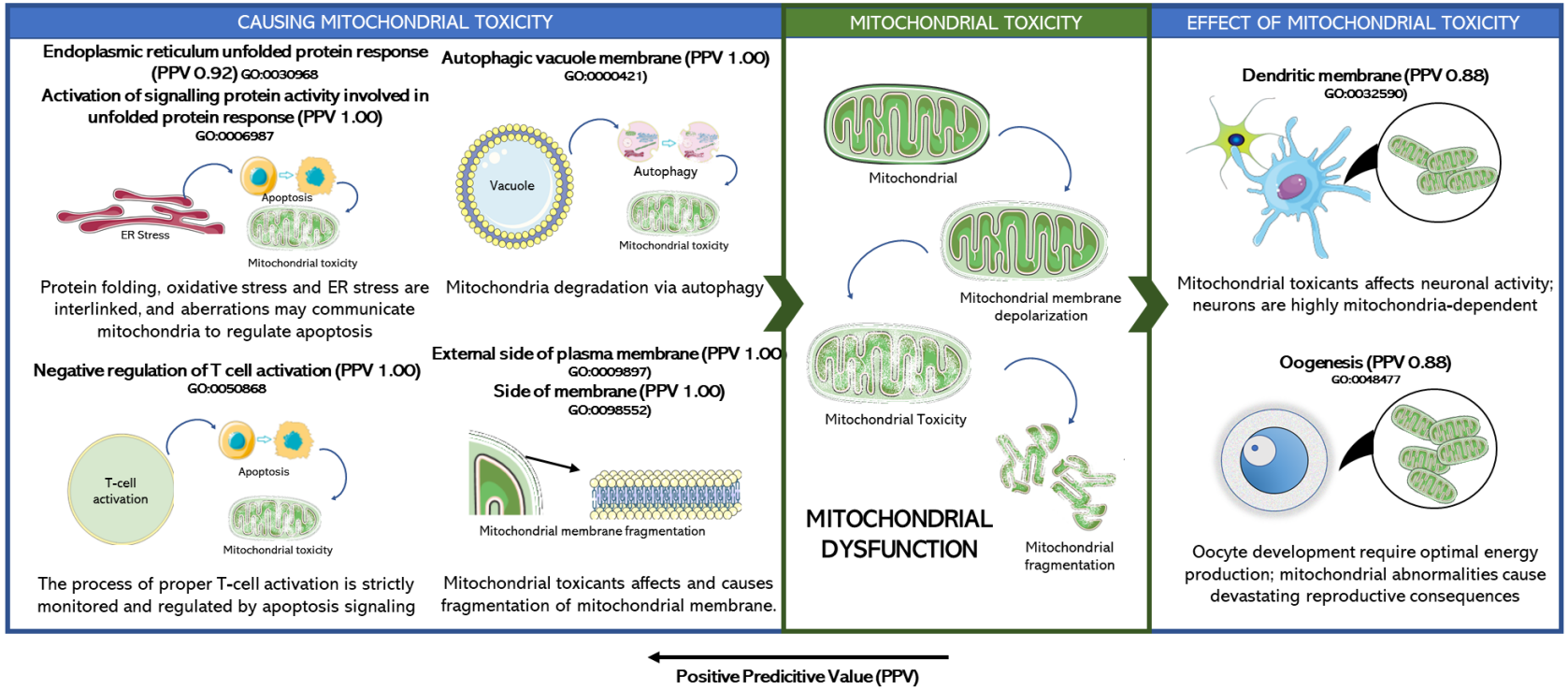


Figure 7. Biological implication of mitochondrial toxicity translated from the Gene Expression features. Features causing mitochondrial toxicity mainly related to unfolded protein response (possibly related to ER stress) and plasma membrane (related to membrane depolarisation). Some effects of mitochondrial toxicity were also captured by Gene Expression features such as oogenesis and dendritic plasma membrane; both processes are heavily mitochondria dependent. Further details in Supplementary Table S5.

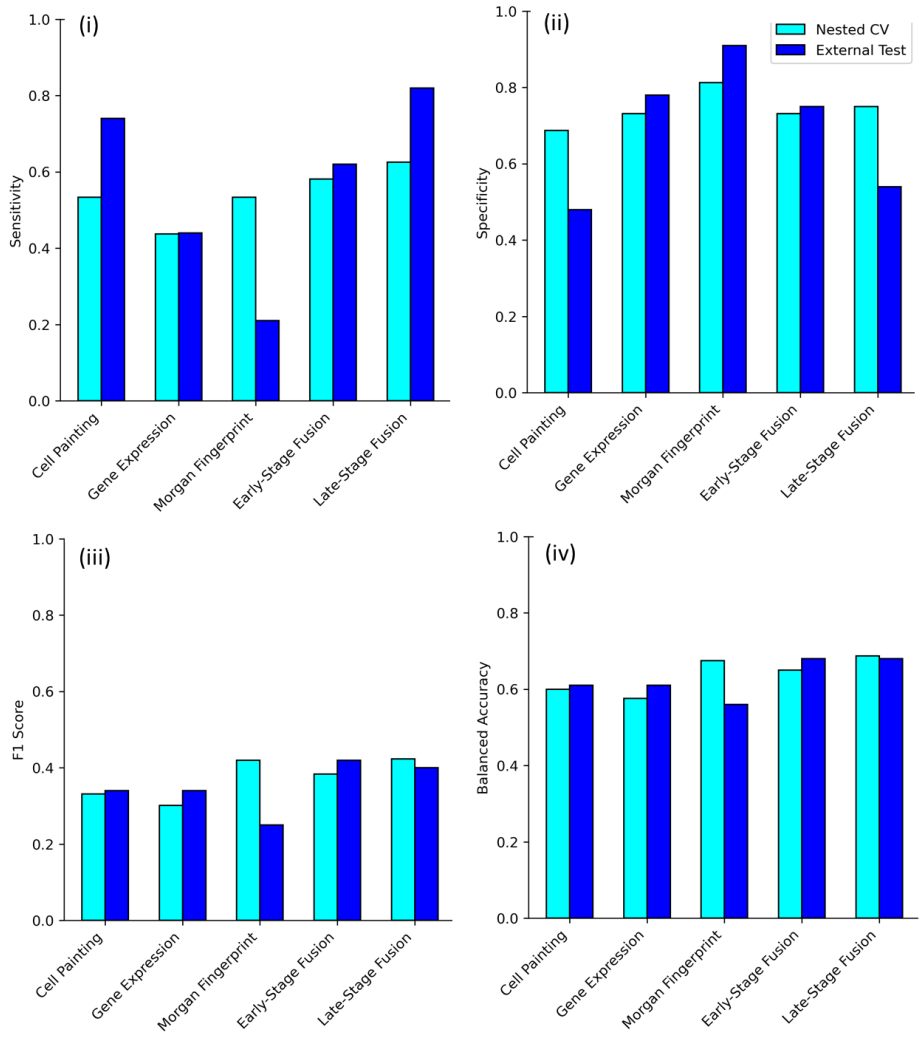
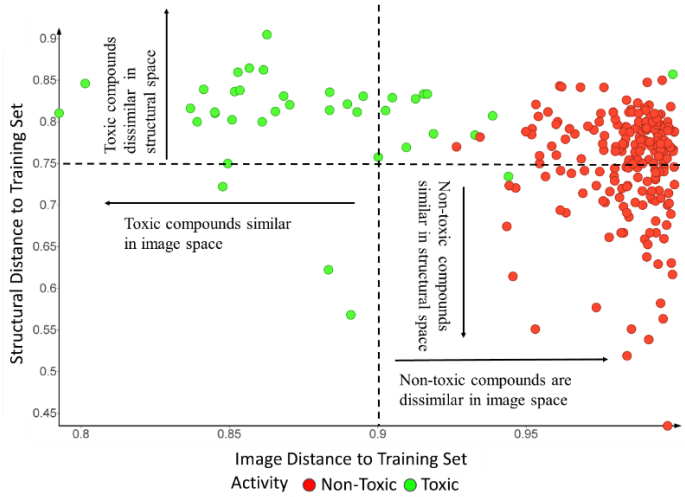
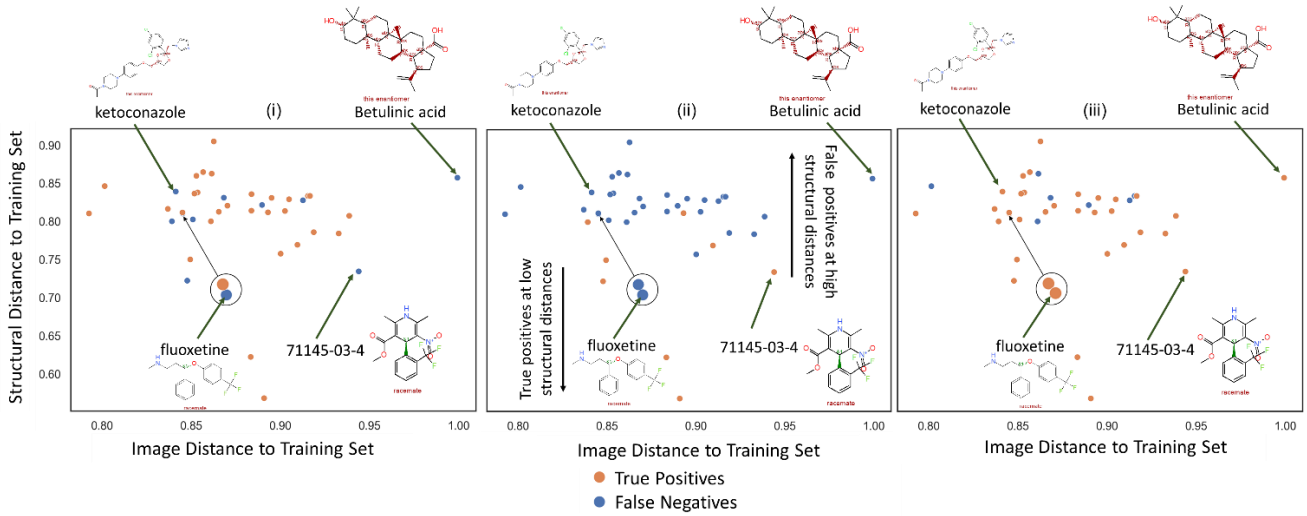


Figure 8. Evaluation metrics, namely from (i) sensitivity, (ii) specificity, (iii) F1 score and (iv) balanced accuracy for five models from (a) Nested CV (median of repeated nested cross validations) and (b) external test set. Early-stage fusion and Late-stage fusion models combining all three feature sets of Cell Painting, Gene Expression and Morgan have higher F1 score for compounds exhibiting mitochondrial toxicity and extrapolate well into new structural space in external test sets compared to models using Morgan fingerprints where F1 Score performance falls by 60% (0.25 to 0.40 in absolute terms).



(a)



(b)

Figure 9. (a) Most mitotoxic compounds are similar in image space for training and test sets, however, non-toxic compounds in the external test set were dissimilar to non-toxic compounds in training set in the image space. Further, toxic compounds are often dissimilar in structural space, indicating the need for fusion models.
(b) Structural and morphological distance for mitotoxic compounds in external test set to the training set for models using (i) Cell Painting features, (ii)Morgan fingerprints and the (iii) Late-stage fusion models. Morgan fingerprints failed to correctly classify mitotoxic compounds (eg. betulinic acid) at high structural distances while models using Cell Painting features could extrapolate well into structurally diverse compounds. The late-stage fusion models correctly classified mitotoxic compounds (eg. 71145-03-4 or methyl 2,6-dimethyl-5-nitro-4-[2-(trifluoromethyl)phenyl]-1,4-dihydropyridine-3-carboxylate, ketoconazole and fluoxetine) in both diverse morphological and structural space where individual models failed demonstrating the synergistic effect of the features spaces.