National Institute of Environmental Health Sciences
Division of Translational Toxicology

Machine Learning-based Development of Cytotoxicity Flags Within the Tox21/ToxCast Assays

A. Borrel[1], A.L. Karmaus[1*], B. Hill[1], K.T. To[1], V. Hull[1], G.Tedla[1], J.T. Auman[1], D.G. Allen[1*], N. Kleinstreuer[2]

[1]Inotiv, RTP, NC; [2]NIH/NIEHS/DTT/NICEATM, RTP, NC

Abstract 4383
Poster P168

Introduction

- In vitro high-throughput screening (HTS) assay data are a time- and cost-effective approach to provide mechanistic insights and predict toxicity for thousands of diverse chemicals. The US interagency collaborative Tox21 (Huang et al. 2016) and US EPA's ToxCast (Thomas et al. 2018) programs provide in vitro data for thousands of chemicals and assays (Feshuk 2023).
- Chemical effects in vitro may be confounded by overt cell stress and cytotoxicity, such that a decrease in viable cells could erroneously be attributed to a chemical's mechanistic effects. Integration of cytotoxicity assessment with assay endpoints can bolster confidence in the interpretation of assay outcomes.
- Many chemicals tested in ToxCast/Tox21 HTS assays lack directly relevant cytotoxicity data needed to ensure overt toxicity doesn't confound mechanistic outputs.
- Additionally, chemicals may have different potency for eliciting cytotoxicity across cell types and time trajectory.
- ToxCast/Tox21 cytotoxicity assays can be leveraged to define chemical-specific cytotoxic concentrations.
  - For example, the "burst" analysis integrates 91 cytotoxicity assays to define a cytotoxicity limit per chemical that may indicate where a "burst" of potentially nonselective bioactivity occurs (Judson et al. 2016).
- We propose a machine learning approach to predict chemical- and cell type-specific cytotoxicity concentrations to provide context for flagging nonspecific in vitro chemical-elicited bioactivity.

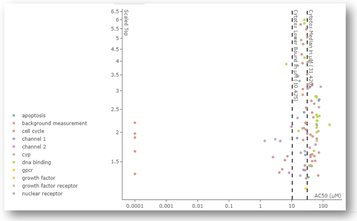
Objectives

Integrate cytotoxicity data to provide context for HTS bioactivity

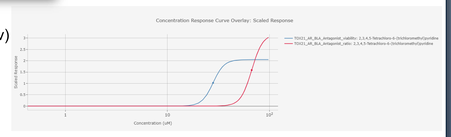
1. Identify bioactivity assay endpoints for which there are direct, concurrent cytotoxicity assessments.
2. For bioactivity assay endpoints without concurrent cytotoxicity data:
   - Leverage the 91 cytotoxicity assays that were used in the latest "burst" analyses to refine chemical-specific cytotoxicity predictions.
   - Group cytotoxicity assay endpoints by cell type and time point and apply multiple machine learning algorithms with optimized parameters to predict cell type specific cytotoxicity for each chemical.

Cytotoxicity Relative to Bioactivity

- The relative potency of cytotoxicity vs. endpoint-specific bioactivity can provide context as to whether the bioactivity may be confounded by overt toxicity, and thus considered a nonselective effect.



◀ **Figure 1. ToxCast assay bioactivity summary plot from US EPA's CompTox Chemicals Dashboard** (https://comptox.epa.gov/dashboard/) for 2,3,4,5-Tetrachloro-6-(trichloromethyl)pyridine (CASRN 1134-04-9) shows most bioactivity assay endpoint AC50 values are above the "burst" cytotoxicity limit values suggesting that overt toxicity drives most responses observed.



▶ **Figure 2. Concentration-response curve overlay from ICE** (https://ice.ntp.niehs.nih.gov) for bioactivity assay endpoint and its concurrent accompanying cytotoxicity readout revealing a more potent cytotoxic effect (AC50 is point on curve).

Machine Learning Workflow to Predict Cytotoxicity

1. Cytotoxicity assays were retrieved from ICE cHTS by filtering for those marked as "burst" by EPA's invitrodb v3.5. (91 assay endpoints).
2. Cytotoxicity assays were grouped by cell type and time point.
3. Per-chemical median AC50s were computed within each cytotoxicity assay group.
4. Chemicals were characterized by a set of molecular descriptors computed using RDKit (www.rdkit.org).
5. XGBoost and Random Forest models were tuned and developed with optimized parameters.
6. Machine learning models were used to predict cytotoxicity AC50s for chemicals; predictions were specific to chemical, cell type, and time point.

Figure 3. Machine learning workflow. To predict cell line, time point, an chemical-specific cytotoxicity AC50 concentrations that could be applied for contextualizing bioactivity assays in which there are no concurrent cytotoxicity assessments, machine learning approach was conducted.

Tiered Approach to Integrating Cytotoxicity

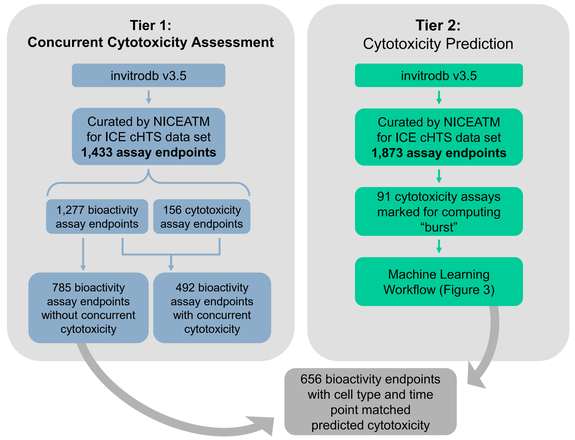


Figure 4. Tiered workflow for identifying cytotoxicity data. A two-tiered workflow, where tiering reflects priority for the source of cytotoxicity data, has been developed to work toward integrating cytotoxicity data for contextualizing bioactivity endpoints.

- The first tier manually maps cytotoxicity assays to concurrently measured bioactivity endpoints.
- The second tier focuses on using cytotoxicity assays used for "burst" cytotoxicity point calculations (Judson et al. 2016) to derive cell type and time point specific cytotoxicity predictions, per chemical.
- Future efforts will seek to develop and conduct a third tier which will address bioactivity endpoints evaluated in assays for which there is no cytotoxicity data available at all for the cell line.

Predicted Cell Type and Time Point-Specific Cytotoxicity

To predict cytotoxicity (predicted AC50 representing the point at which cytotoxicity is expected) for chemicals tested in assays without concurrent cytotoxicity measured, we developed a machine learning approach. Models were developed for specific subsets of assay conditions, namely cell line and time point (e.g., treatment duration). As a proof-of-concept, results are presented for two conditions are shown: HepG2 human hepatocarcinoma cells with 24 hrs treatment and HEK293T human embryonic kidney cells with 24 hrs of treatment.

Table 1: Summary of datasets used for machine learning

Cell Type	Time Point	Number of Active Chemicals	Number of Inactive Chemicals
HepG2	24 hrs	846	5680
HEK293T	24 hrs	746	5813

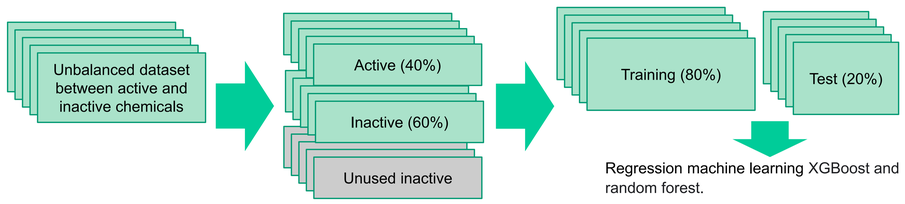


Figure 5. Undersampling approach. Given the unbalanced dataset (many more inactive chemicals than active chemicals in both of the cell type and time point limited cytotoxicity assay groupings), we implemented an undersampling approach. Briefly, random selection of inactive chemicals was conducted to achieve a ratio of 60% inactive to 40% active, which was repeated such that each final model is an aggregation of 20 undersampled models. The final chemical prediction is the mean of the 20 predictions for each chemical.

All predictions from the 20 undersampled models are averaged at the chemical level and the R^2 performance metric, where an R^2 value close to 1 indicates near-perfect prediction accuracy, are summarized in Tables 2 and 3. Each model was tuned using a Bayesian approach and a maximization of the performance criteria (R^2) for 100 iterations.

Table 2: HepG2 cells treated for 24 hrs - cytotoxicity prediction model performance

Machine Learning Model	Training Set R^2	Test Set R^2	10-fold Cross-Validation R^2
XGBoost	0.77	0.61	0.56 +/- 0.22
Random Forest	0.76	0.58	0.49 +/- 0.21

Table 3: HEK293T cells treated for 24 hrs - cytotoxicity prediction model performance

Machine Learning Model	Training Set R^2	Test Set R^2	10-fold Cross-Validation R^2
XGBoost	0.81	0.57	0.45 +/- 0.18
Random Forest	0.77	0.46	0.43 +/- 0.25

The machine learning approaches evaluated (XGBoost and Random Forest) achieved reasonable performance for the two proof-of-concept datasets. For our examples, XGBoost demonstrates better performance, but advanced machine learning techniques using deep learning or ensemble modeling should be considered in the future to strive for even better results.

Endpoints with Concurrent Cytotoxicity

Table 4: Examples of bioactivity assay endpoint mapping to concurrent cytotoxicity assay endpoint

aeid*	Bioactivity Assay Endpoint Name	aeid*	Cytotoxicity Assay Endpoint Name
1855	ACEA_AR_agonist_80hr	1850	ACEA_AR_agonist_AUC_viability
1856	ACEA_AR_antagonist_80hr	1857	ACEA_AR_antagonist_AUC_viability
2	ACEA_ER_80hr	1852	ACEA_ER_AUC_viability
1829	ArunA_Migration_hNC_dn	1826	ArunA_CellTiter_hNC_dn
1827	ArunA_Migration_hNP_dn	1825	ArunA_CellTiter_hNP_dn
2446	CCTE_Simmons_MITO_inhib_resp_rate_OCR_dn	2450	CCTE_Simmons_MITO_viability
2447	CCTE_Simmons_MITO_inhib_resp_rate_OCR_up	2450	CCTE_Simmons_MITO_viability
906	CEETOX_H295R_ESTRADIOL_dn	1664	CEETOX_H295R_MTT_cell_viability_dn
907	CEETOX_H295R_ESTRADIOL_up	1664	CEETOX_H295R_MTT_cell_viability_dn
914	CEETOX_H295R_TESTO_dn	1664	CEETOX_H295R_MTT_cell_viability_dn
915	CEETOX_H295R_TESTO_up	1664	CEETOX_H295R_MTT_cell_viability_dn
2037	CPHEA_Stoker_NIS_Inhibition_RAIU	2110	CPHEA_Stoker_NIS_Cytotoxicity
962	LTEA_HepaRG_CYP1A1_dn	1136	LTEA_HepaRG_LDH_cytotoxicity
963	LTEA_HepaRG_CYP1A1_up	1136	LTEA_HepaRG_LDH_cytotoxicity
1691	STM_H9_OrnCyssISnorm_ratio_dn	1858	STM_H9_Viability_norm
1690	STM_H9_OrnCyssISnorm_ratio_up	1858	STM_H9_Viability_norm
806	TOX21_AhR_LUC_Agonist	807	TOX21_AhR_LUC_Agonist_viability
767	TOX21_Aromatase_Inhibition	768	TOX21_Aromatase_Inhibition_viability
2047	TOX21_CAR_Agonist	2048	TOX21_CAR_Agonist_viabillity
2049	TOX21_CAR_Antagonist	2050	TOX21_CAR_Antagonist_viability

*aeid: assay endpoint identification numbers from invitrodb v3.5

Table 5: Distribution of active/inactive for matched concurrent cytotoxicity

	Cytotoxicity Active	Cytotoxicity Inactive
Bioactivity Endpoint Active	30,108	30,087
Bioactivity Endpoint Inactive	48,956	383,886

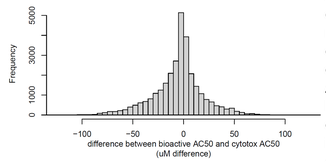


Figure 6. Difference between bioactive and concurrent cytotoxicity AC50s. Histogram of records where both bioactivity endpoint and cytotoxicity assay were active. The difference was computed as bioactivity AC50 minus cytotoxicity AC50. Negative values indicate "selective" bioactivity where bioactivity AC50 was less than cytotoxicity AC50 (18,069 records). Positive values indicate "non-specific" effects where bioactivity AC50 was greater than cytotoxicity AC50 (12,039 records).

Summary

- We have developed a tiered framework to integrate cytotoxicity with bioactivity assay endpoints enhancing the interpretation of in vitro assays and providing context for distinguishing between selective and confounded outcomes.
- Numerous chemicals elicit bioactivity in Tox21/ToxCast assay endpoints at AC50 concentrations higher than the cytotoxicity AC50s suggesting that those results may be non-specific secondary or cell stress-induced outcomes rather than specific bioactivity.
- We have mapped concurrent cytotoxicity readouts for 492 assay endpoints, allowing direct comparison of bioactivity potency against cytotoxicity. These comparisons will be integrated into future versions of concentration-response visualizations for the curated HTS data in the ICE Curve Surfer tool.
- Cell type- and time point-specific machine learning models were developed to predict cytotoxicity AC50s to provide context for assays without concurrent cytotoxicity data.
- Further refinement of our machine learning cytotoxicity predictive models is being conducted before they are integrated with bioactivity data to provide context and bolster confidence in assay outcome interpretation for identifying specific vs. non-specific/cytotoxicity-confounded bioactivities within these or other cell types.

Integrated Chemical Environment

References and Acknowledgments

Thomas R et al. 2018. ALTEX 35(2):163–168. https://doi.org/10.14573/altex.1803011.
Huang R et al. 2016. Nat Commun. 7:10425. https://doi.org/10.1038/ncomms10425.
Feshuk M et al. 2023. Front. Toxicol. 5:1275980. https://doi.org/10.3389/ftox.2023.1275980.
Judson R et al. 2016. Toxicol Sci 152:323–339. https://doi.org/10.1093/toxsci/kfw092.

The authors thank the National Institutes of Health (NIH) High-Performance Computing Group for providing resources to run the predicting models. This project was funded in part with federal funds from the NIEHS, NIH under Contract No. HHSN273201500010C. The authors declare there exists no conflict of interest.
*A.L. Karmaus is currently affiliated with Syngenta, Greensboro, NC.
*D.G. Allen is currently affiliated with the International Collaboration on Cosmetics Safety, New York.