Machine learning-based development of cytotoxicity flags within the Tox21/ToxCast assays

Getachew Tedla[1], Alexandre Borrel[1], Agnes Karmaus[1], Victoria Hull[1], Kimberly T. To[1], J. Todd Auman[1], D.G. Allen[1], N. Kleinstreuer[2]

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

Background and Purpose

The abundance of in vitro assays performed as part of the Tox21 and ToxCast high-throughput screening programs includes data measured in multiple cell lines and technologies. However, the validity of some reported activities is uncertain due to cytotoxicity effects. Chemicals display potentially non-selective bioactivity near the cytotoxic concentration. Therefore, distinguishing specific from non-specific activities is important. For this purpose, the ToxCast/Tox21 program included 91 cytotoxicity assays that can help identify assay results that are above the cytotoxicity cutoff. We developed an approach to flag reported bioactivities in the ToxCast/Tox21 assays that are above the chemical-specific concentrations that cause cytotoxicity.

Methods

We mapped cytotoxicity assays to paired mechanistic ToxCast/Tox21 assays, using cell type parameters and direct pairwise cytotoxic and non-cytotoxicity assay mapping. Then we explored strategies used to determine the cytotoxicity cutoff, i.e., where reported bioactivities could be due to cell stress or death. We cross-checked bioactivity response curves with the minimum, median, or maximum cytotoxicity cut-offs. We compared this approach with the cytotoxicity burst approach developed by Judson et al. by computing z-score values and identifying chemical-assay pairs that were above or below z-score thresholds. Finally, for records without matching cytotoxicity data, we used machine learning to build quantitative structure–activity relationship (QSAR) models that predict the cytotoxicity cutoff for chemical-assay combinations.

Results

We mapped ToxCast/Tox21 mechanistic assays to cytotoxicity assays using 28 common cell types and 3,226 common chemicals. Out of 119,253 ToxCast/Tox21 positive assay records, we identified 54,656 records with matching cytotoxicity assays and 64,597 records remain without a directly matching cytotoxicity assay. The matched records constituted 486 unique ToxCast/Tox21 assay endpoints and 2,601 unique chemicals. Our assessment indicated that up to 60% (32,874 records) of the assay endpoints had reported AC50 concentrations above matching cytotoxicity AC50 values. Most bioactivity records flagged by the burst approach were also flagged by our approach, and we identified several curve responses showing limitations of both approaches. Finally, we developed QSAR models for 7,641 records where we had cytotoxicity assays mapped but the chemicals were not tested for cytotoxicity. Different machine-learning models were tested, with the better ones having an R^2 of approximately 0.6 for the assays modeled.

Conclusions

Our results indicate that numerous chemicals have activities for Tox21/ToxCast assay endpoints at concentrations higher than the concentrations reported for cytotoxicity. We used various approaches to describe the number of assay endpoints reported above AC50 concentrations that cause cell death. In addition, we are currently developing a robust method that combines multiple

approaches and curve-response quality information to assess the reliability or non-reliability of chemical-assay endpoint pairs. This project was funded in whole or in part with federal funds from the NIEHS, NIH under Contract No. HHSN273201500010C.

Keywords: cytotoxicity, computational Toxicology and Data Integration, QSAR, ToxCast/Tox21