Enhancing a Curated High-Throughput Screening Pipeline to Improve Toxicological Interpretations of In Vitro Data

V. D. Hull[1], B. N. Hill[1], A. Unnikrishnan[1], E. N. Reinke[1], N. C. Kleinstreuer[2], H. T. Hogberg[3]

[1]Inotiv, United States; [2]NIH/OD/DPCPSI, United States; [3]NIH/NIEHS/DTT/NICEATM, United States

Testing of thousands of chemicals in the ToxCast and Tox21 high-throughput screening (HTS) programs has generated millions of data points that can be difficult to analyze in detail. To increase confidence in HTS bioactivity calls, the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) has reviewed these data to generate a curated HTS (cHTS) data set, in which results are flagged based on chemical quality, curve fit, or potential technological interference. In this presentation, we will describe recent improvements to our curation process that have been made to accommodate major updates in concentration-response modeling in the U.S. Environment Protection Agency (EPA) ToxCast/Tox21 database (Feshuk et al. 2022).

All data were retrieved from the EPA invitrodb v4.2 database and analyzed using the tcpl R package. Representative curves were derived for each chemical–assay pair, and assays that were determined to have limited biological relevance were removed. Removals included, but were not limited to, background readouts used for assay quality control; individual channel readouts (only ratios were retained); and transcriptional activation assays analyzed inappropriately in the loss direction.

Remaining assays received "Flag-Omit" labels if assay performance or curve fit was deemed not biologically relevant. Examples included estimated points-of-departure that exceeded the tested range, mismatches between curve direction and the intended direction of analysis, or down-direction curves where the best-fit model was gain–loss. We also obtained quality control (QC) data on chemicals tested and applied them to the curation process. Curves were flagged as "QC-Omit" if the sample or chemical was graded as low concentration or low purity during the QC process. Finally, chemical–assay pairs were examined for potential technological interference, as luciferase inhibition and fluorescence quenching/autofluorescence can result in false bioactivity signals. Assays using these technologies were identified via assay metadata and literature review. Potential interferent chemicals were identified via Tox21 tests or predicted using quantitative structure–activity relationship models (Borrel et al. 2020). If a chemical was interferent for a technology and was tested in an assay using that technology, the assay–chemical pair was flagged with "Technological Interference." This curation results in at least one flag for 46% of the 1,865,056 assay-chemical pairs.

We also annotated cHTS data to mechanistic targets that facilitate linkage to modes-of-action and subsequently to toxicological outcomes of regulatory interest. All mechanistic target and mode-of-action terms were mapped using the Open Biological and Biomedical Ontology Foundry. We considered curve direction in this annotation, as an increased or decreased response can change the biological interpretation.

This rigorous curation process ensures that bioactivity calls are robust, interpretable, and biologically relevant. Updated cHTS data are available in NICEATM's Integrated Chemical Environment (ICE, https://ice.ntp.niehs.nih.gov/), an open-access resource for toxicologically relevant data and computational tools. This project was funded with federal funds from the NIEHS, NIH under Contract No. HHSN273201500010C. The views expressed above do not necessarily represent the official positions of any federal agency.