Incorporating Ontologies into High Throughput Screening Assay Annotations to Increase Data Use and Interpretation

B.N. Hill[1], E.N. Reinke[1], J.T. Auman[1], A. Unnikrishnan[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

In vitro high throughput screening (HTS) assay data have increased our understanding of chemically mediated effects for thousands of chemicals and facilitated the development of computational approaches for testing and assessment. While HTS assays are valuable sources of mechanistic information, it is often not clear how these data relate to toxicological endpoints such as developmental toxicity and endocrine disruption. Here we present an annotation scheme for HTS assays that can provide biological context to the data and enable toxicological interpretation. Annotating the curated HTS data in the Integrated Chemical Environment (ICE: https://ice.ntp.niehs.nih.gov/) to mechanistic targets facilitates linkage to modes of action and subsequently to toxicological outcomes of regulatory interest, where possible. The scheme incorporates controlled vocabularies within the framework of ontologies to address two common challenges associated with large toxicological datasets: inconsistent terminologies and reporting. Annotation of datasets using controlled vocabularies and ontologies is also a critical aspect to ensuring data objects are findable, accessible, interoperable, and reusable (FAIR).

Methods

Technical and biological information from Tox21/ToxCast HTS assays was evaluated to infer toxicological endpoints and pathways. Annotation began with the evaluation of information provided in the U.S. Environmental Protection Agency's invitrodb v3.5, where examples of fields reviewed include "intended_target_family" and "biological_process_target". This mapping process was supplemented with expert review to annotate assay mechanistic targets to mode of action terms relevant to regulated toxicity endpoints (e.g., "vascularization" and "epigenetic process"). These terms were curated to ensure consistency and appropriateness of the annotation details. Previously, all mechanistic target terms were mapped to the NCI Metathesaurus. To harmonize HTS assay annotations across programs, we updated the mappings using the Open Biomedical and Biological Ontology (OBO) Foundry, a computationally interoperable database of biological sciences knowledge.

Results

Ontologies focusing on biological and toxicological processes, including the Gene Ontology knowledgebase, were chosen. This not only enabled refinement of assay mapping to mechanistic targets and modes of action but also provided structured relationships between these concepts using parent and child terminologies. We began with 153 previously mapped assay targets with NCI Metathesaurus terms and assigned OBO terms to them, and then expanded the number of mapped assays to include newly added datasets and targets. We envision these annotations will increase interoperability with other databases and facilitate harmonized reporting, including incorporation into Harmonized Template 201 developed by the Organisation for Economic Co-operation and Development.

Conclusions:

Overall, these efforts will increase accessibility and interpretation of HTS data to identify data gaps, to better inform chemical risk and hazard assessments, and to provide additional resources for investigations of regulatory-relevant endpoints. The curation and mapping of the HTS data upholds the FAIR principles and is intended to contribute high-quality, structured, mechanistically informative data to the broader data ecosystem via the ICE interface. This project was funded in whole or in part with federal funds from the NIEHS, NIH under Contract No. HHSN273201500010C.

Keywords: Alternatives to Animal Testing, Computational Toxicology, Mode of Action