A public database supporting evidence-based exposomics

Risa R. Sayre1,2, John F. Wambaugh1, Katherine Phillips3, Antony J. Williams1, Christopher M. Grulke1

1) U.S. Environmental Protection Agency, Office of Research and Development, National Center for Computational Toxicology, RTP NC 2) Oak Ridge Institute of Science and Education through an interagency agreement between the U.S. Department of Energy and EPA.

Pre-filtering categories 1 & 2
To avoid registering non-xenobiotic compounds, we created chemical structural classes to pre-filter chemicals from the identification workflow.

Addition of category 3b compounds
After finding a high number of false positives (>99%) in a PubMed search for “metabolite of [xenobiotic name]”, we used manually classified abstracts to build a natural language processing model (F1 = 0.98) to identify abstracts containing substrate/product pairs, or substance relationships. 74% of these transformation products were previously unregistered in DSSTox. To increase signal without adding noise, we registered only transformations observed at plausible exposure levels (and not rapidly transforming intermediates), which are linked to detection method and other metadata.

Method: Adding chemicals from EPA-relevant exposomics categories to DSSTox library

Method: Supporting NTA identification

Existing capability
Advanced mass- and formula-based searches in the Dashboard, including consideration of adducts5. Ranking of candidates utilizes predicted fragmentation patterns and metadata

+ New metadata from this project
- Structures grouped by multiple chemical lists of observed compounds in environmental and biological media support NTA
- Substance relationship mappings allow metadata aggregation (such as data source counts) from known transformation parents to their children, possibly improving proper identification of children (related families)

Discussion

Over 10,000 mappings of xenobiotic transformation relationships are being added to DSSTox, many of which are not currently registered in any metabolomics database. Developing methods to improve identification of these substances measured in human blood and their source categories supports active Agency research projects (e.g. for PFAS chemicals).

Registration of xenobiotics and observed transformation products based on dose levels demonstrated to yield a detectable (by a known method) amount of product in a particular species and medium in a chemical library
• allows development of exposure estimates
• identifies candidate substances and pathways to inform future high-throughput assay research to identify mechanisms

References
2) U.S. Food & Drug Administration. Pesticide Chemical Search.
3) U.S. EPA TSCA Chemical Substance Inventory. www.epa.gov/chem-substances.
4) www.accessdata.fda.gov/scripts/fdcc/?set=FoodSubstances
5) Risa R. Sayre l ORCID 0000-0002-6173-8202 l sayre.risa@epa.gov
6) Discussion

To group chemicals into these categories, a one-vs.-one linear support vector classifier was trained on the URLs of the top ten Google results for chemical names from manually curated and Dashboard-registered lists for categories 1 (from Rappaport et al 2014 supplement), 2a (from FDA Substances Added to Food), and 3a: pesticides, pharmaceutical active ingredients, and others (TSCA8, with overlapping names from other categories removed). The training set (n = 2640) was not restricted to compounds observed in blood. The overall F1 score on the validation set (n = 1320) of the model was 0.80.

This project was supported in part by an appointment to the Internship/Research Participation Program at the National Center for Computational Toxicology, U.S. Environmental Protection Agency, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and EPA.

This poster does not necessarily reflect U.S. EPA policy.