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 (ORISE) Research Participant 3) U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, Computational Exposure Division, RTP NC

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 transformation products observed at plausible exposure levels (and not rapidly transforming intermediates), linked to detection method and other metadata.

Pre-Filtering for Non-Xenobiotic Compounds

A screenshot of a search for fatty acids in the CompTox Chemicals Dashboard.

Endogenous: will be observed in all human sera, regardless of exposures
Transformation product from xenobiotic, where did this come from?

Compounds observed in a pooled human blood sample
Compounds from pooled human serum samples were tentatively identified (in at least 2 out of 3 replicates) in GCxGC-MS and LC-QTOF SSA/NTA workflows (complete methods to be described in a future publication with Lesa Aylward [Summit Toxicology]) designed to filter out endogenous compounds. 22% of the compounds identified in the LC workflow were not registered in DSSTox (EPA’s Distributed Structure-Searchable Toxicity Database), most of which were endogenous.

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 transformation products observed at plausible exposure levels (and not rapidly transforming intermediates), linked to detection method and other metadata.

U.S. Environmental Protection Agency Office of Research and Development

existing capability
Advanced mass- and formulae-based searches in the Dashboard, including consideration of adducts. 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
• Observed substance relationships allow the aggregation of metadata (such as data source counts) from known transformation parents to their children, possibly simplifying the proper identification of those children.

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

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

References