# Use of ToxPrint chemotypes for exploring chemical feature enrichments across the ToxCast chemical-assay landscape

A.M. Richard[1], R. Lougee[2], C. Grulke[1], K. Connors[3], K. Paul-Friedman[1], T. J. Shafer[4], J. Rathman[5], C. Yang[6], A. Williams[1]

[1]National Center for Computational Toxicology, US EPA, RTP, NC USA; [2]ORISE Student Trainee, US EPA, RTP, NC USA; [3]ORISE Post-Doctoral Fellow, RTP, NC USA; [4]National Health & Environmental Effects Research Laboratory, US EPA, RTP, NC USA; [5]Altamira, Columbus, OH, [6]Molecular Networks, GmbH, Erlangen, Germany

EPA's ToxCast chemical library spans diverse chemical use-types, functionalities, structures and features potentially relevant to toxicity and environmental exposure. However, this structural diversity, along with assay noise and low average hit rates across the varied ToxCast high-throughput screening (HTS) technologies and data sets, poses challenges to traditional, global structure-activity relationship (SAR) modeling approaches. A publicly available set of 729 ToxPrint chemotypes (CTs), representing a diverse set of chemical features spanning toxicity alerts, common scaffolds, and varied ring, bond, and atom types, is being used to reduce the dimensionality of the ToxCast inventory and codify local chemistry domains to amplify structure-activity signals within those domains. An automated CT-enrichment workflow (CTEW) has been developed, with preliminary functionality released via the EPA Chemistry Dashboard (https://comptox.epa.gov/). ToxPrints were computed for the entire DSSTox database (>700K structures), and CT-enrichment results were generated for >800 ToxCast/Tox21 HTS assays. Over 460 CTs are statistically enriched (Odds Ratio >3, Fisher's Exact p-value <0.05) in ≥ 1 assay, with 191 of these CTs enriched in ≥ 40 assays; 600 assays are enriched with ≥ 3 CTs, with 140 of these assays enriched in ≥ 40 or more CTs. These results offer a rich vein of chemical inferences to mine, such as CTs associated with promiscuous HTS activity. The CTEW has been applied to a variety of datasets, including: ToxCast assays grouped by target type (e.g., nuclear receptors) and by common detection technologies (e.g., autofluorescence); analytical QC "failed" Tox21 chemical samples; microelectrode array (MEA) neurotoxicity assay results where enriched CTs support mechanistic linkages to ToxCast ion channel assays enriched with the same CTs; and various activity subsets within the ToxRef vivo dataset. The approach offers an intuitive, flexible complement to traditional SAR methods, with results that are easily interpreted, anchored to visualizable chemical features, and that can productively guide more targeted SAR investigations. *Abstract does not reflect EPA policy.*