Cheminformatics and Data Mining Approaches for Exploring the Alternatives Testing Landscape: Case Studies in ToxCast and Skin Sensitization

United States Environmental Protection Agency

Ann Richard[1], Grace Patlewicz[1], Jeremy Fitzpatrick[2], Chris Grulke[1], Indira Thillainadarajah[3], Antony Williams[1]

[1]U.S. EPA, NCCT, RTP, NC, USA; [2]ORISE Fellow; [3]SEE Program

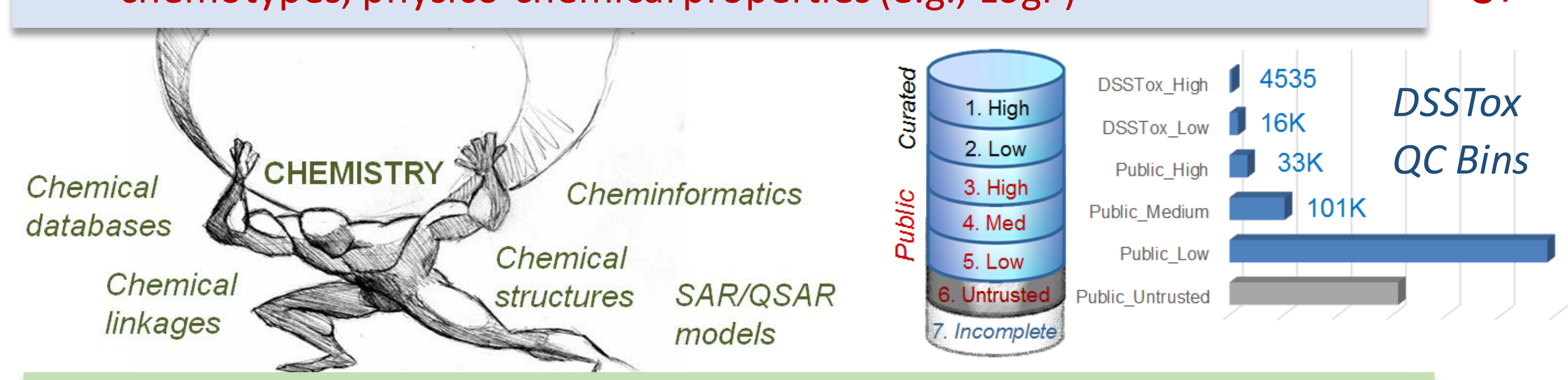
Science Context

Cheminformatics approaches and structure-based rules are being used to evaluate and explore the ToxCast chemical landscape and associated high-throughput screening (HTS) data. We have shown that the library provides comprehensive coverage of the knowledge domains and target inventories of interest to the Agency. Building on this work, we illustrate how ToxPrint chemotypes (CTs) - an objective, transparent, and reproducible set of chemical substructural features defining local chemical neighborhoods - can be used to profile and probe chemical bioactivity enrichment patterns across various chemical "activity" sets and in relation to the *in vitro* data landscape. Lastly, we critically evaluate some common structure-based assumptions associated with the induction of skin sensitization (SS), that serve as key inputs into Integrated Approaches to Testing & Assessment (IATA) and Adverse Outcome Pathways (AOP) for this well-studied endpoint.

Approach

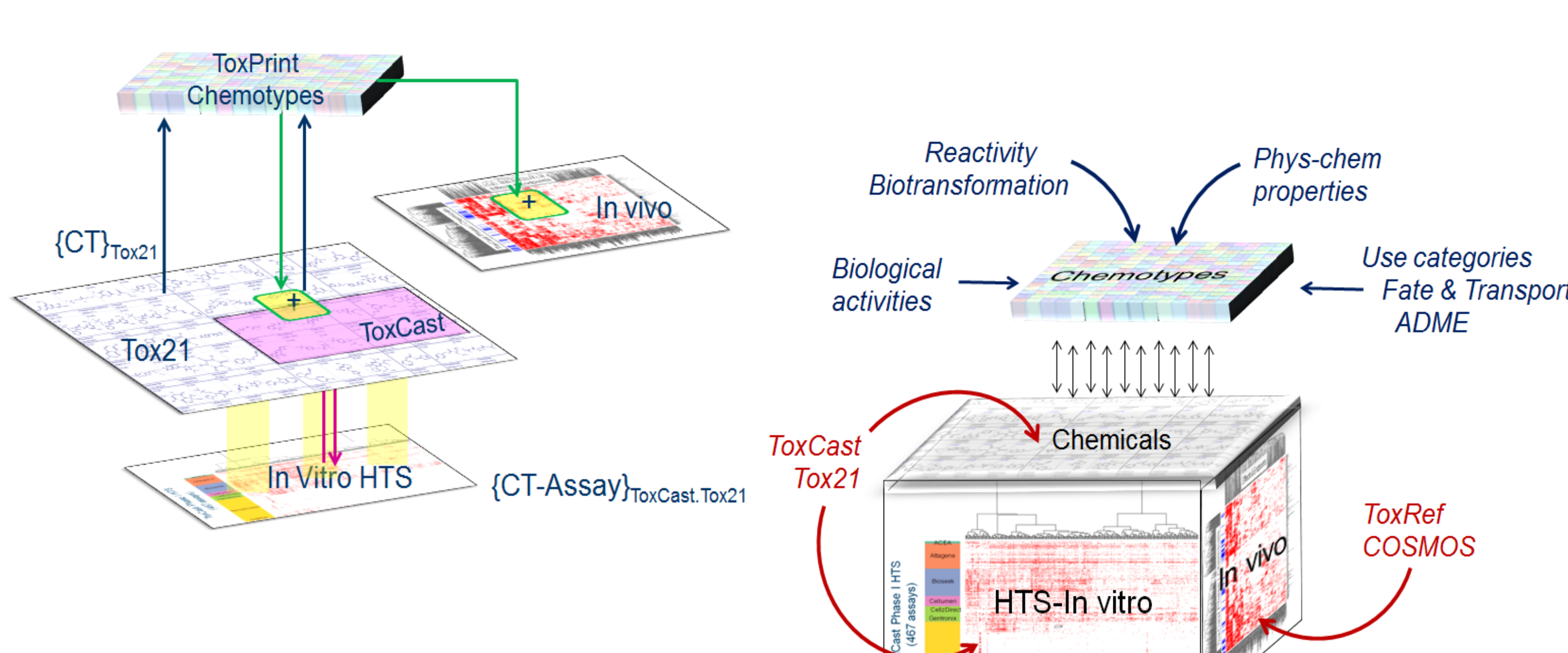
Deliverables thru FY16

1) Build ToxCast chemical library, generate structure-representations, survey contents → *Richard et al., CRT, 2016*;
2) Build chemical registration system and significantly expand EPA DSSTox database to create high quality chemical structure resource → increase from 25K to 730K chemicals, assign to 6 "curation QC" bins
3) Register structure inventories of high interest to EPA programs → e.g., ToxCast, Tox21, ToxRef, TSCA, HPV, Hydrofracking, ECHA datasets, etc.
4) Generate predicted features and properties for DSSTox structures for use in EPA research programs → QSAR-ready files, ToxPrint chemotypes, physico-chemical properties (e.g., LogP)



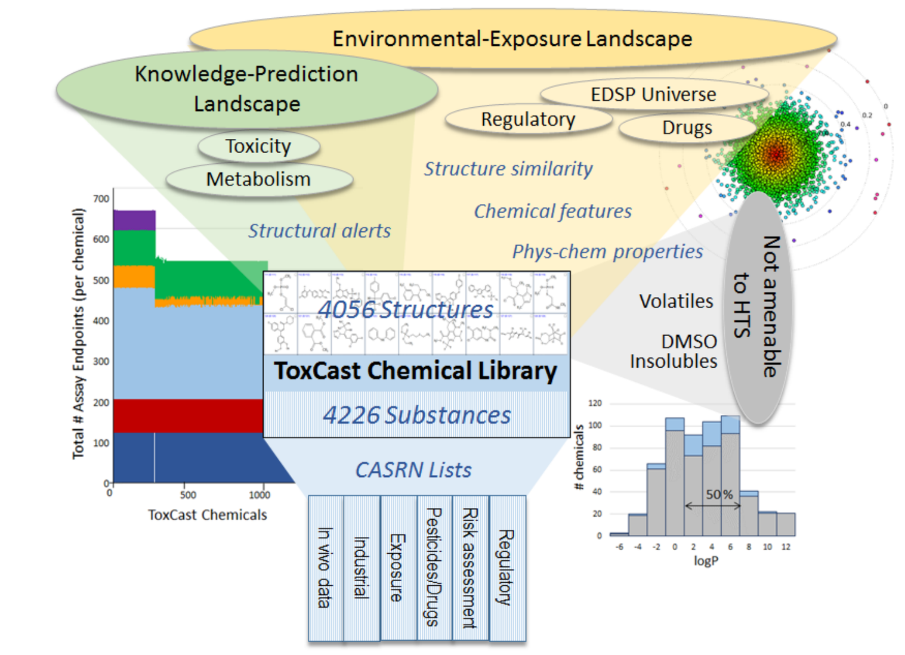
FY17-FY18

5) Compute enrichments of ToxPrint CTs in activity subsets to characterize activity within defined local chemistry domains (e.g., HTS actives, *in vivo* actives, functional category, analytical QC "fails", etc.);
6) Develop workflows to support use of ToxPrint CTs in modeling;
7) Expand coverage of chemical inventories & data sets of interest (toxicity, activity, usage) for use in structure-based data mining and modeling in support of toxicity assessments.

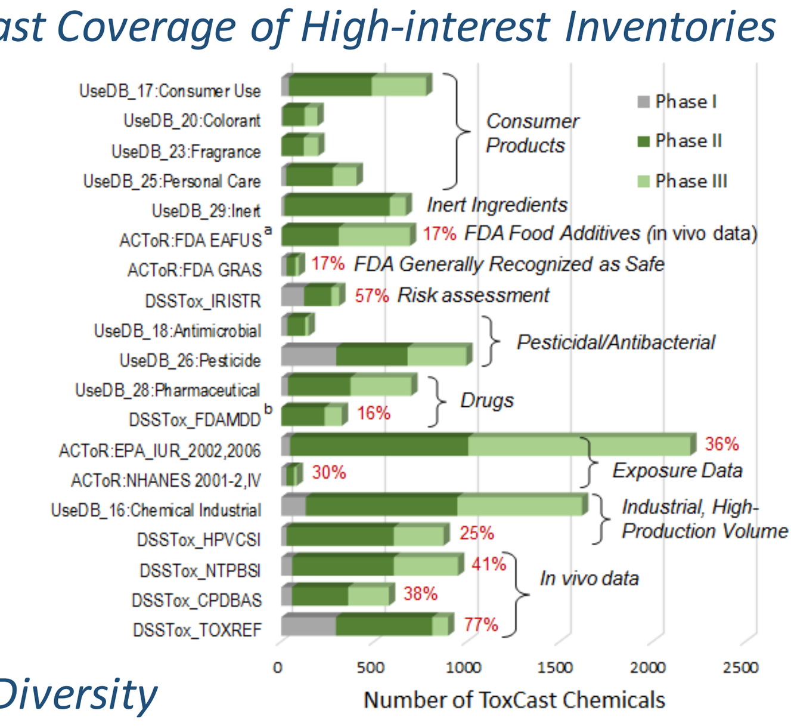


Preliminary Results

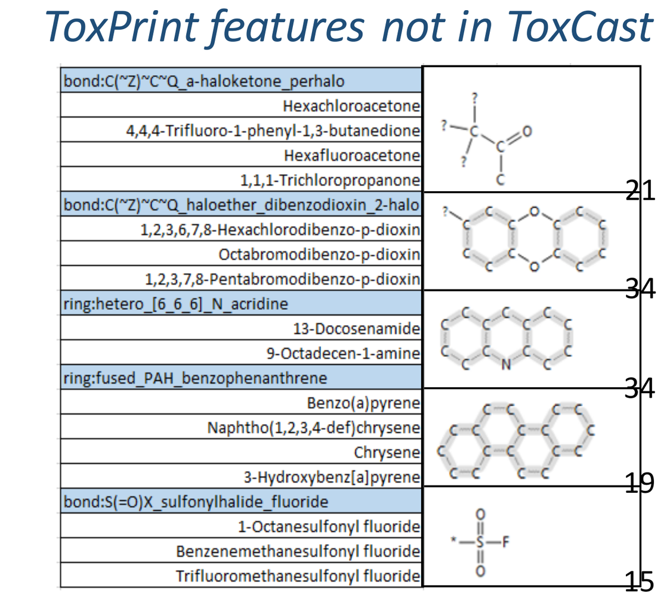
ToxCast Chemical Landscape: Evaluating Coverage, Diversity, & Fit-for-purpose



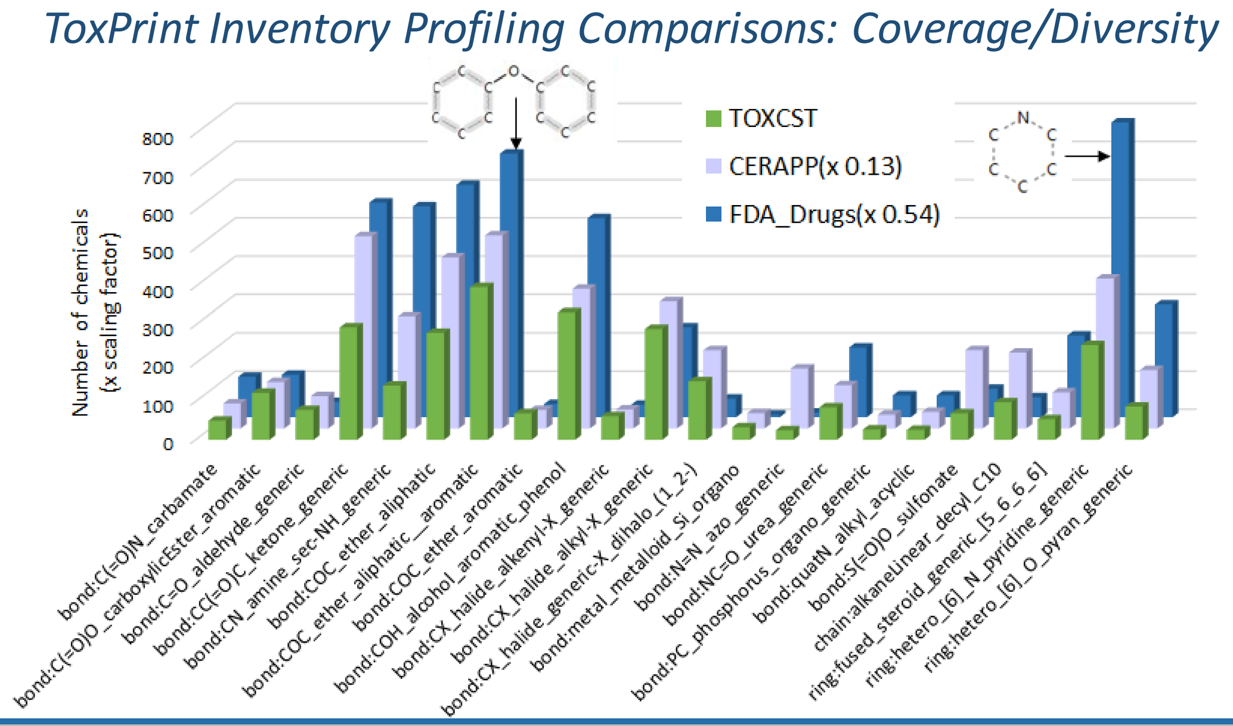
ToxCast Coverage of High-interest Inventories

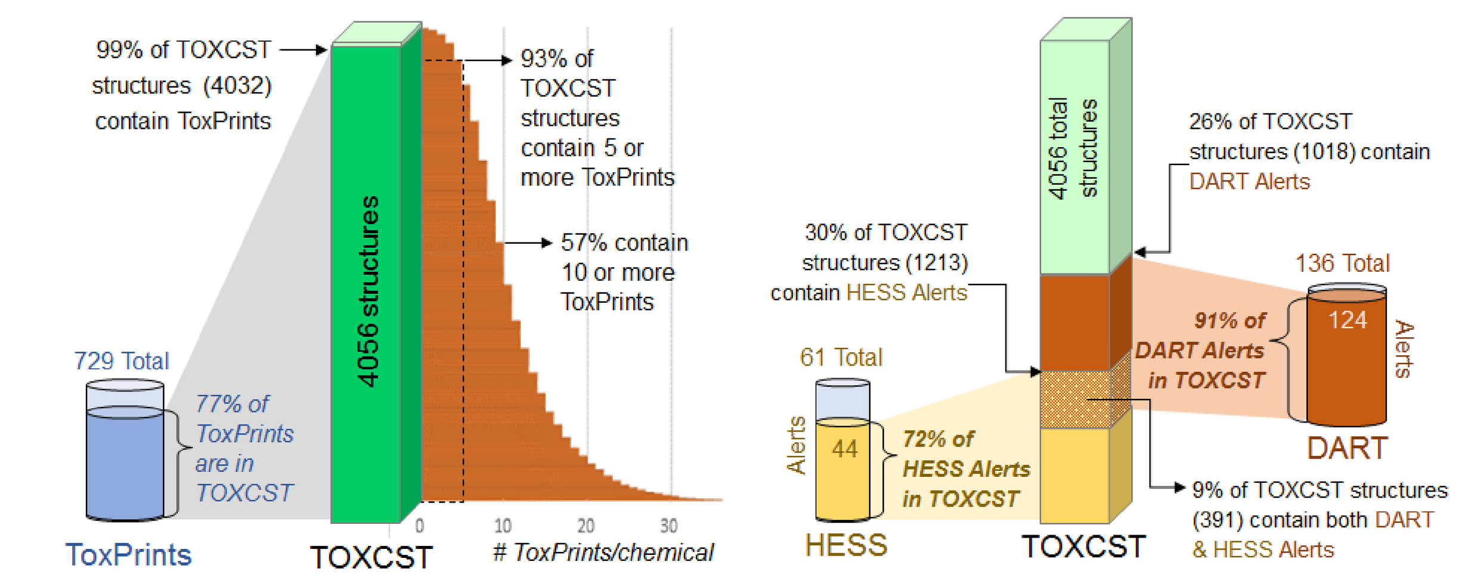


ToxPrint features not in ToxCast



ToxPrint Inventory Profiling Comparisons: Coverage/Diversity

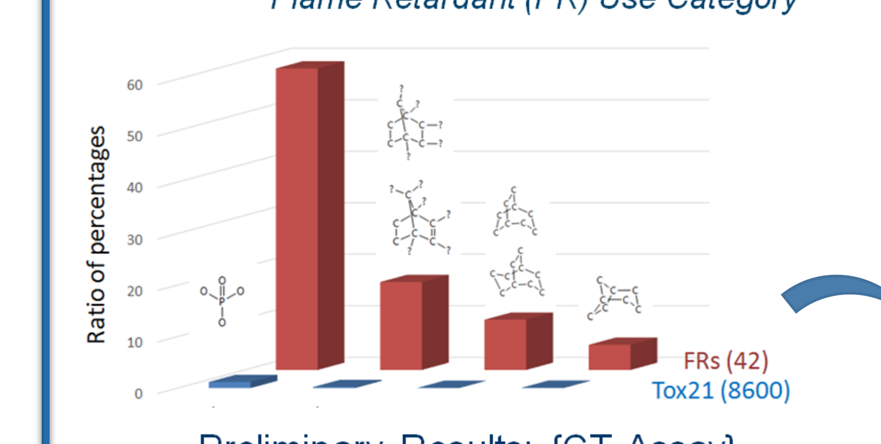




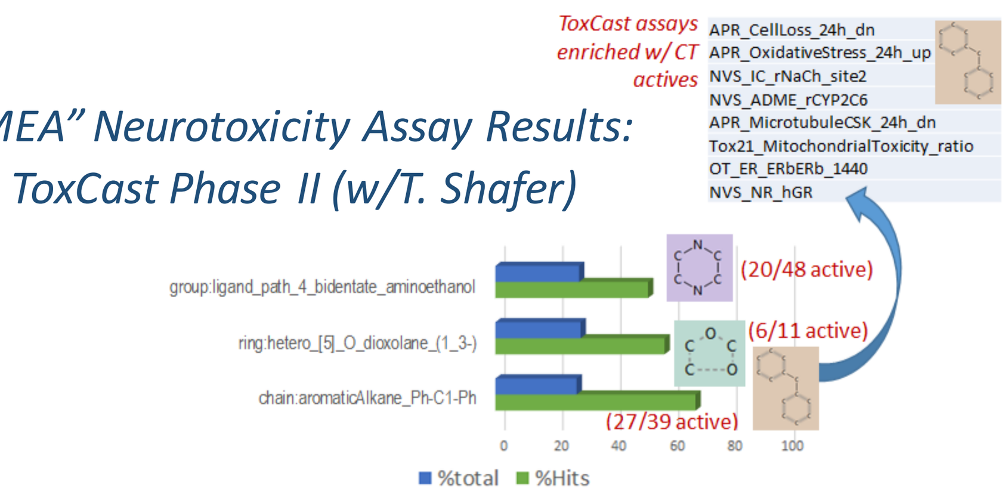
ToxCast Coverage of ToxPrints & Historical Structure-alerts for Toxicity

ToxPrint Chemotype Enrichments in "Activity" Space

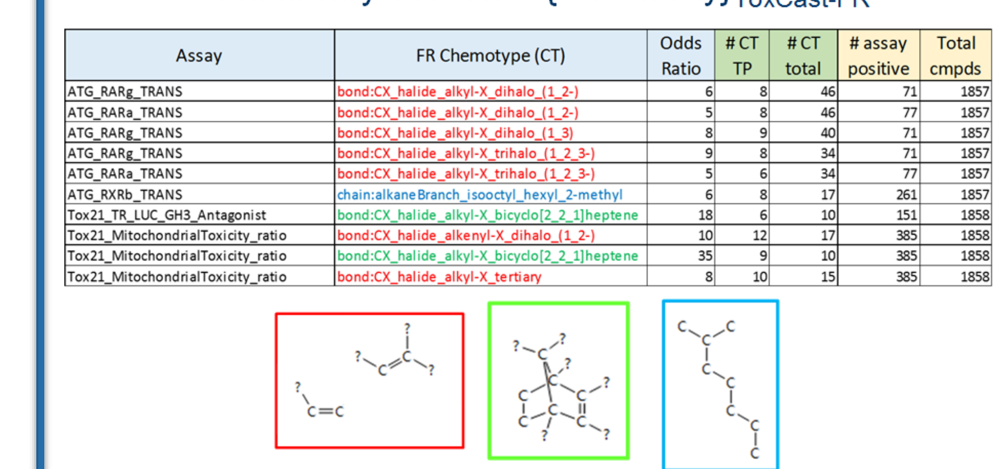
Chemotype Enrichment, e.g.
Flame Retardant (FR) Use Category



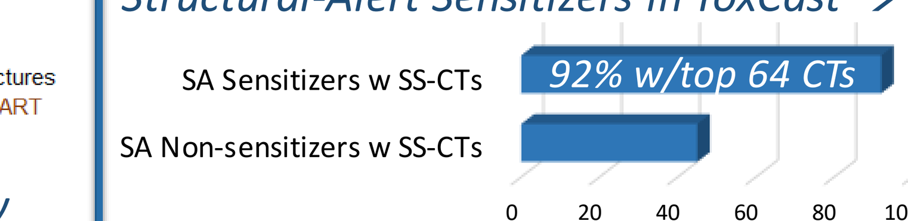
"MEA" Neurotoxicity Assay Results: ToxCast Phase II (w/T. Shafer)



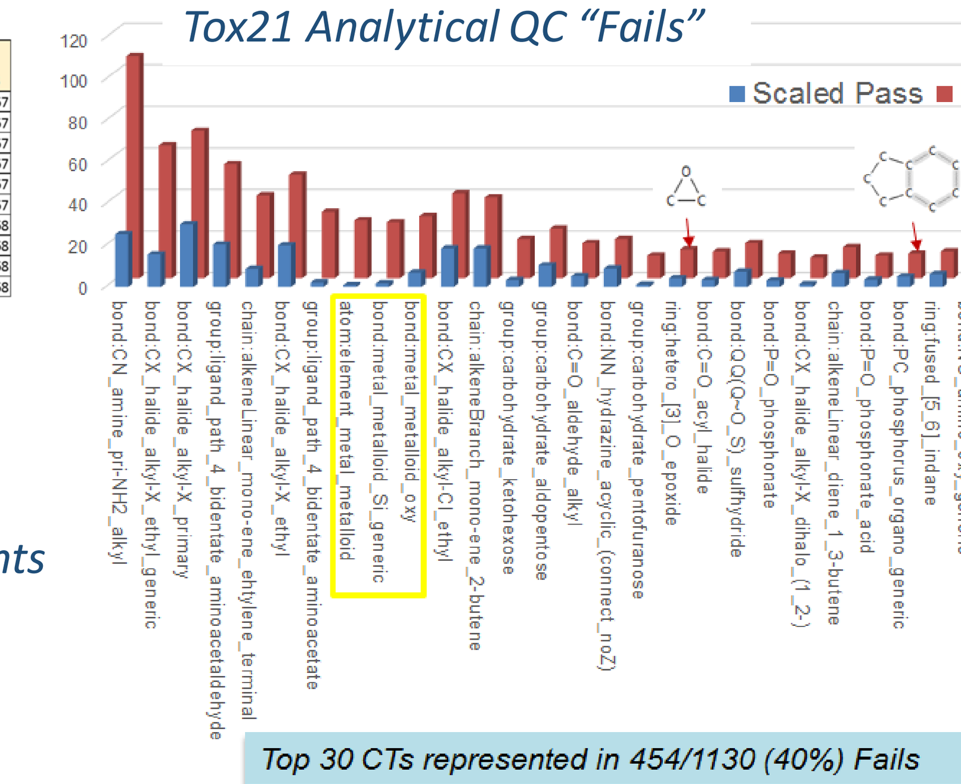
Preliminary Results: {CT-Assay}$_{ToxCast\text{-}FR}$



Structural-Alert Sensitizers in ToxCast → ToxPrints



Tox21 Analytical QC "Fails"



Critical Evaluation of AOP & IATA* Assumptions in Predicting Skin Sensitization

I. Are MW & LogP Thresholds for Skin Penetration Valid?

- Compounds that can not penetrate the skin are assumed not to be sensitizers;
- Molecular Weight (MW) < 500 or Log octanol/water partition coeff (LogP) > 1 thresholds *assumed* for penetration to occur;
- Few reported skin sensitizers exceed these thresholds

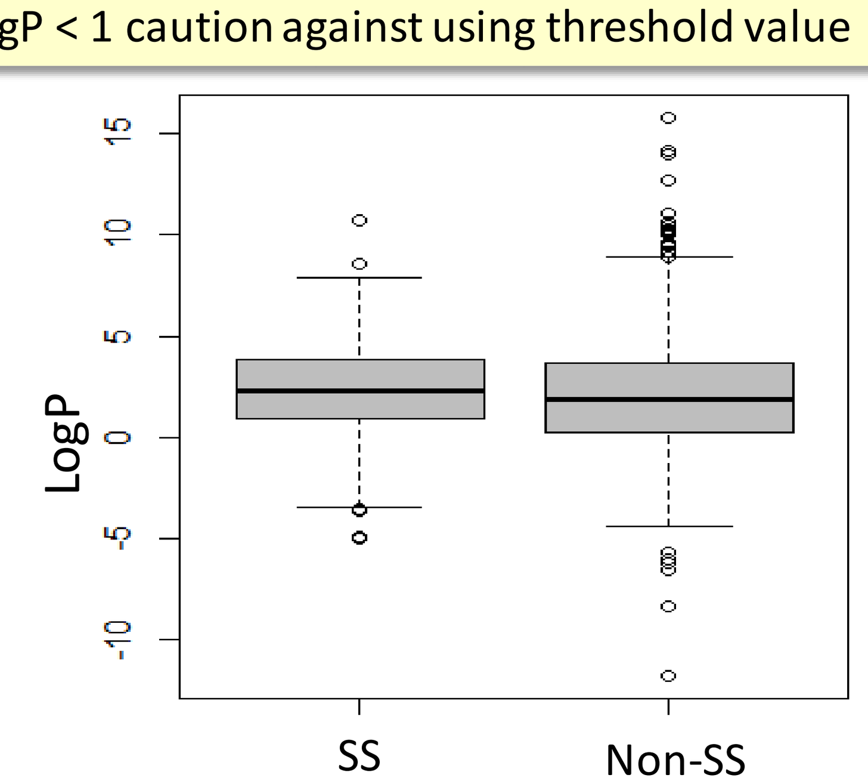
→ *Systematic search of a large body of sensitization data collected under the EU REACH regulation and available through OECD eChemPortal was used to test validity of MW and LogP thresholds as conditions for skin penetration and, thus,*

Skin Sensitization Information From ECHA

	MW > 500		MW ≤ 500	
	# Cmpds	% Cmds	# Cmpds	% Cmpds
Sensitizers	33	17%	735	27%
Non-Sensitizers	164	83%	1972	73%
Total Compounds	197		2707	

→ Compounds with MW > 500 should **NOT** be automatically ruled out from assessment.

→ Although the majority of 1482 reported skin sensitizers have LogP > 1, finding significant numbers of skin sensitizers with LogP < 1 caution against using threshold value



II. Can non-animal tests correctly identify "pre- or pro- haptens" (i.e., indirectly acting sensitizers)?

- Pre-haptens are activated abiotically outside the skin, mainly by autoxidation;
- Pro-haptens are activated in skin, mainly by metabolic mechanisms.

→ *An EURL ECVAM dataset of 127 substances having results from LLNA and 3 non-animal test methods was reviewed. Reaction chemistry structure-based rules were used to identify pre-hapten and pro-hapten subsets, and predictivity of the 3 non-animal tests was examined.*

28/28 indirectly acting sensitizers were positive in 1 or more tests; 6/6 pro-haptens were positive in at least 1 cell based test

→ Careful consideration needs to be paid to anticipated reaction chemistry before testing, but existing non-animal test methods were found to be fit for purpose for detecting in direct acting sensitizers.

* *Adverse Outcome Pathways & Integrated Approaches to Testing and Assessment*

Anticipated Products/Impacts

- Expanded coverage of chemical inventories (toxicity, activity, usage) for use in structure-based data mining and modeling in support of toxicity assessments;
- Knowledge-based repository of ToxPrint CT associations, including enrichments in ToxCast/Tox21 HTS assays, use categories, etc.
- Automated workflows to support use of ToxPrint CTs in "read-across" and data-mining;

→ *A knowledge-based cheminformatics layer to effectively integrate multiple data streams in support of robust, chemical safety assessments*

References

- Richard A, Judson R, Houck K, Grulke C, Volarath P, Thillainadarajah I, Yang C, Rathman J, Martin M, Wambaugh J, Knudsen T, Kancherla J, Mansouri K, Patlewicz G, Williams A, Little S, Crofton K, Thomas R. The ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology. *Chem. Res. Toxicol.* 2016, 29(8):1225-1251. doi: 10.1021/acs.chemrestox.6b00135
- Yang C, Tarkhov A, Marusczyk J, Bienfait B, Gasteiger J, Kleinoeder T, Magdziarz T, Sacher O, Schwab CH, Schwoebel J, Terfloth L, Arvidson K, Richard A, Worth, A, Rathman J. New publicly available chemical query language, CSRML, to support chemotype representations for application to data mining and modeling. *J. Chem. Inf. Model*. 55:510-28, 2015.
- Fitzpatrick JM, Roberts DW, Patlewicz G. Is skin penetration a determining factor in skin sensitisation potential and potency? Refuting the notion of a LogKow threshold for Skin Sensitisation. *J. Appl. Toxicol.* 2016, doi: 10.1002/jat.3354
- Fitzpatrick JM, Roberts DW, Patlewicz G. What determines skin sensitisation potency–myths, maybes and realities. The 500 molecular weight cut-off – An updated analysis. *J. Appl. Toxicol.* 2016, doi: 10.1002/jat.3348
- Patlewicz G, Casati S, Basketter D, Asturiol D, Roberts DW, Lepoittevin J-P, Aschberger K. Can currently available non-animal methods detect pre and pro haptens relevant for skin sensitization? *Reg Toxicol Pharmacol* 2016, pii: S0273-2300(16)30228-8. doi: 10.1016/j.yrtph.2016.08.007