Revisiting and updating chemical categorizations using chemical fingerprint and high-throughput screening data

Abstract 309

Kellie A. Fay[1], Kamel Mansouri[2], John Prindiville[3], Grace Y. Patlewicz[4], Mark Lewis[5], Ann Richard[4], Mahmoud Shobair[4], Ellery Saluck[6], Daniel T. Chang[4]

[1]USEPA OCSPP/OPPT, Washington, DC, USA. [2]NIEHS NTP, Durham, NC, USA. [3]Environment and Climate Change Canada, Ottawa, ON, Canada. [4]USEPA ORD/CCTE, Durham, NC, USA. [5]Health Canada, Ottawa, ON, Canada. [6]Student Intern, USEPA OCSPP/OPPT, Washington, DC, USA

BACKGROUND & OBJECTIVES

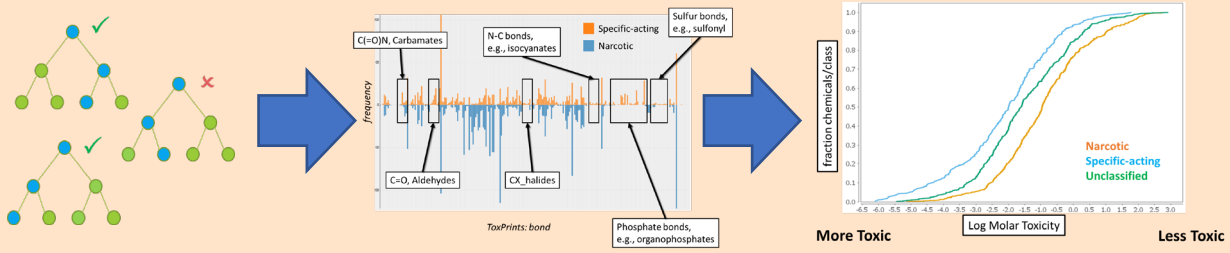
- Traditional approaches to chemical categorization are based on accumulated data and past decisional precedents.
- Many new chemicals across various regulatory jurisdictions cannot be categorized using existing *in silico* models and methods.
- ***How do we incorporate new approach methodologies (NAMs) and cheminformatic approaches to assist in identifying new chemical categories (or classes), and to create more robust models at predicting chemical toxicity?***
- ***Primary focus of this work: Identification of narcotic (N) and specific-acting (S) chemicals for aquatic (fish) toxicity using a consensus Mode-of-Action (cMOA) classification dataset.***

APPROACH

- Classify narcotic and specific-acting mechanisms for a set of ~7000 ToxCast chemicals based on a consensus Mode-of-Action (cMOA) methodology[4] developed by Kienzler *et al.*, 2019.
- Use classified cMOA data to develop predictive models based on ToxPrint (TxP) chemotypes.
- Identify and use targeted dichotomized NAM bioassays hit calls to improve characterization and comparisons with existing Envirotox database aquatic toxicity data.

RESULTS

- Development of a robust N/S classification model for aquatic toxicity.
- Known limitations regarding unclassified cMOA chemicals were identified.
- Chemotype enrichment suggests targeted use of NAM information – suggested use of specific assay data.



SUMMARY/IMPACTS

- Increase the available chemical space of EnviroTox w/ cMOA classifications.
- Develop a robust N/S classification structural ToxPrint based model.
- Identify challenges in unclassified cMOA chemicals – i.e., metal & metalloids, as well as amino acids and polydentate ligands .
- Using chemotype enrichments to identify potential bioassays with bioactivity to provide support of NAM data in category development.

For more information, contact: Daniel T Chang, (chang.daniel@epa.gov)

This work does not reflect EPA policy.

Acknowledgments: ILS (contractor support) and Kristan Markey (USEPA)

Revisiting and updating chemical categorizations using chemical fingerprint and high-throughput screening data

BACKGROUND & OBJECTIVES

Primary focus of this work: Identification of narcotic (N) and specific-acting (S) chemicals for aquatic (fish) toxicity using a classified consensus Mode-of-Action (cMOA) dataset.

"A chemical category is a group of chemicals whose physicochemical and human health and/or ecotoxicological properties and/or environmental fate properties are likely to be similar or follow a regular pattern, usually as a result of structural similarity." - OECD

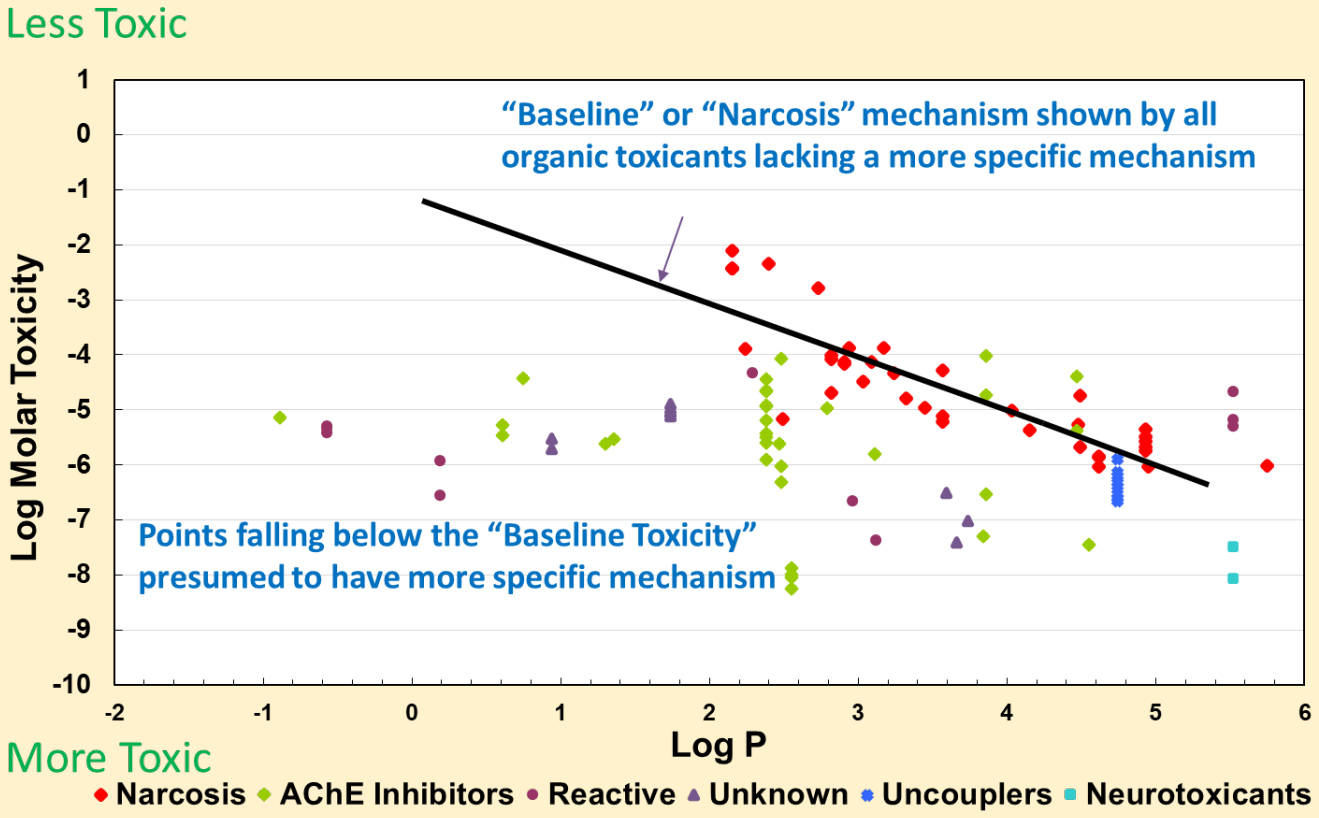
Applications of chemical categorization include first tier assessment efforts and read across from structurally similar analogs – ECOSAR[1]

US EPA ECOSAR chemical classifications

- Class-based SAR to predict aquatic toxicity
- Classification scheme identifies excess toxicity
- Estimates **acute** and **chronic toxicity** based on accumulated data and past decisional precedents

Acute Effects:	Chronic Effects:
Fish 96-hr LC_{50}	Fish ChV
Daphnid 48-hr EC_{50}	Daphnid ChV
Algae 72/96-hr EC_{50}	Algae ChV

- Profiler in OECD QSAR Toolbox

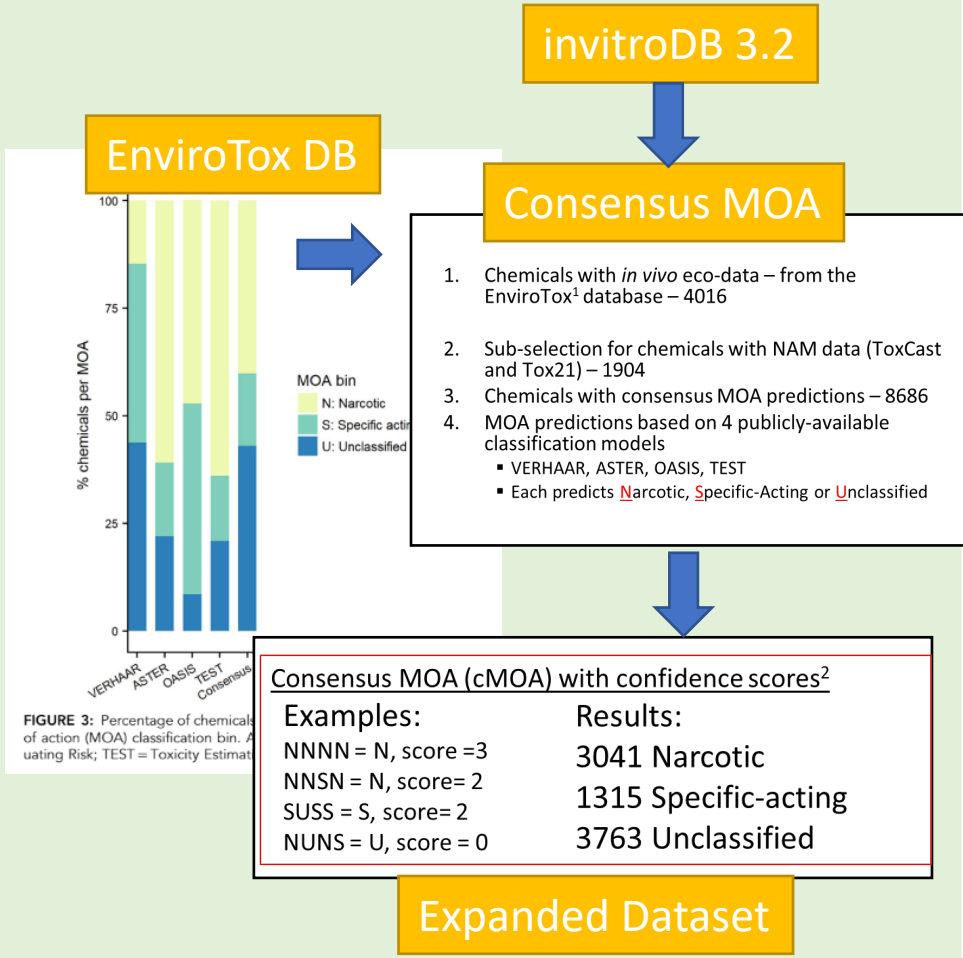


Regulators consider MOA information to determine the size of assessment factors

Revisiting and updating chemical categorizations using chemical fingerprint and high-throughput screening data

APPROACH – Datasets, Classification and Fingerprints

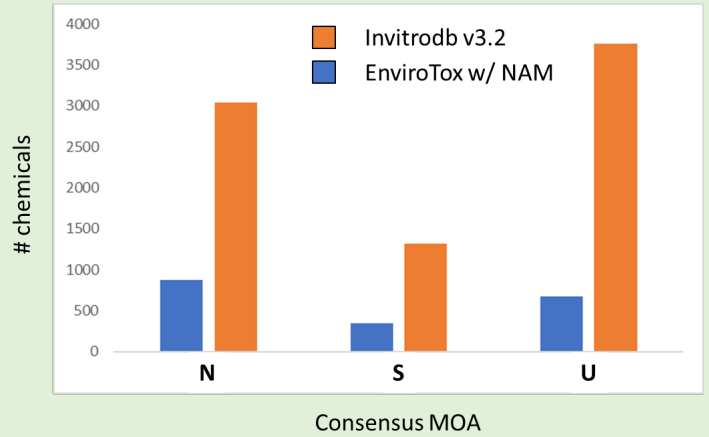
- Current NAM datasets: ToxCast/Tox21 Inventory[2]
  - All ToxCast/Tox21 chemical data is publicly available
  - ToxCast has data on over 4500 chemicals from a broad range of sources
  - Tox21 has screened over 8500 chemicals on over 80 assays
- *Aquatic toxicity In vivo* datasets: EnviroTox database[3]
  - Establish confidence in applying Threshold of Toxicological Concern (TTC) concepts in an eco (multispecies) concept.
  - >91k aquatic toxicity records, >4k chemicals, >1500 species
  - Includes data from ECHA (REACH), USEPA ECOTOX & Pesticide, METI, FET, AiiDA
- Classify data: Consensus Mode-of-Action (cMOA)[4]
  - Establish confidence/performance of several MOA models through a consensus approach: TEST, OASIS, ASTER and Verhaar
  - Towards a more harmonized approach to MOA classification models
  - Differences exist across the classification models used
- Fingerprint/feature set: ToxPrints[5]
  - 729 chemical features
  - Chemically interpretable
  - Coverage of diverse chemistry includes scaffolds, functional groups, chains, rings, bonding patterns and atom-types
  - Survey of ToxPrint chemotypes across ToxCast chemical space[5] (Richard *et al.*, 2016)
  - Provides a link to the High-Throughput Screening (HTS) assay through chemical structure archetypes



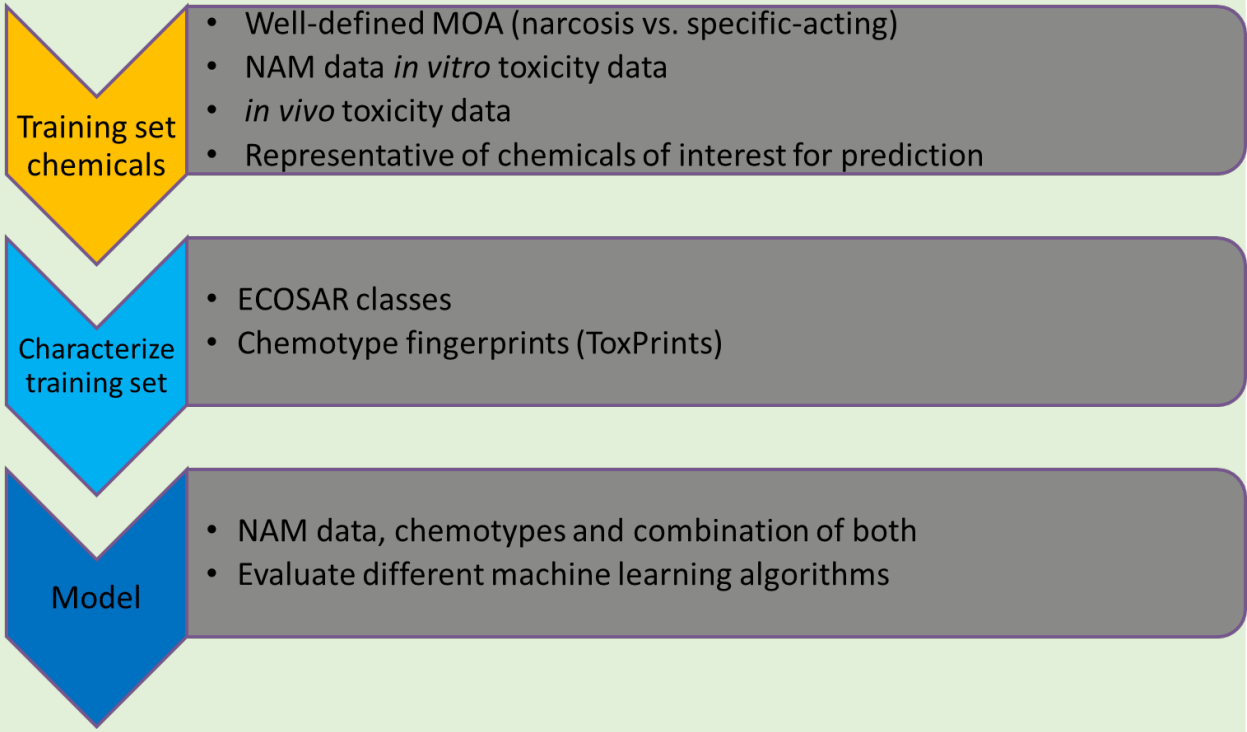
Revisiting and updating chemical categorizations using chemical fingerprint and high-throughput screening data

APPROACH – Model development

- Additional 6215 chemicals with NAM data and cMOA calls
  (compared to Envirotox db: **U**nclassified, **S**pecific-acting, **N**arcotic)



- Random Forest (Boosted Gradient Method) provided the best model results:
  - Split data into 80% training and 20% hold out (test) sets
  - Hyperparameter tuning with 5-fold cross validation, square-root sampling, etc.
  - Training set: “balanced” down-sampled subset (2104 chemicals w/ cMOA = **N**arcotic or **S**pecific-acting)

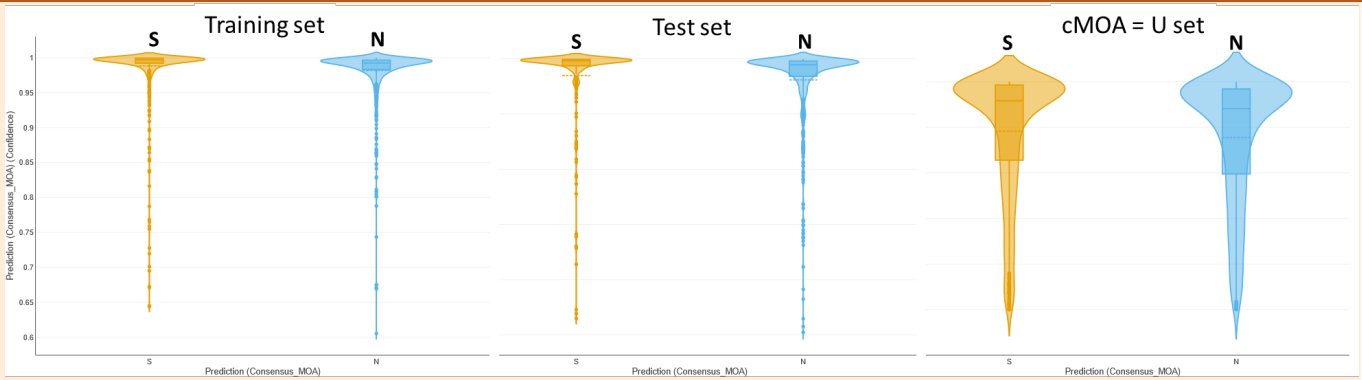


Revisiting and updating chemical categorizations using chemical fingerprint and high-throughput screening data

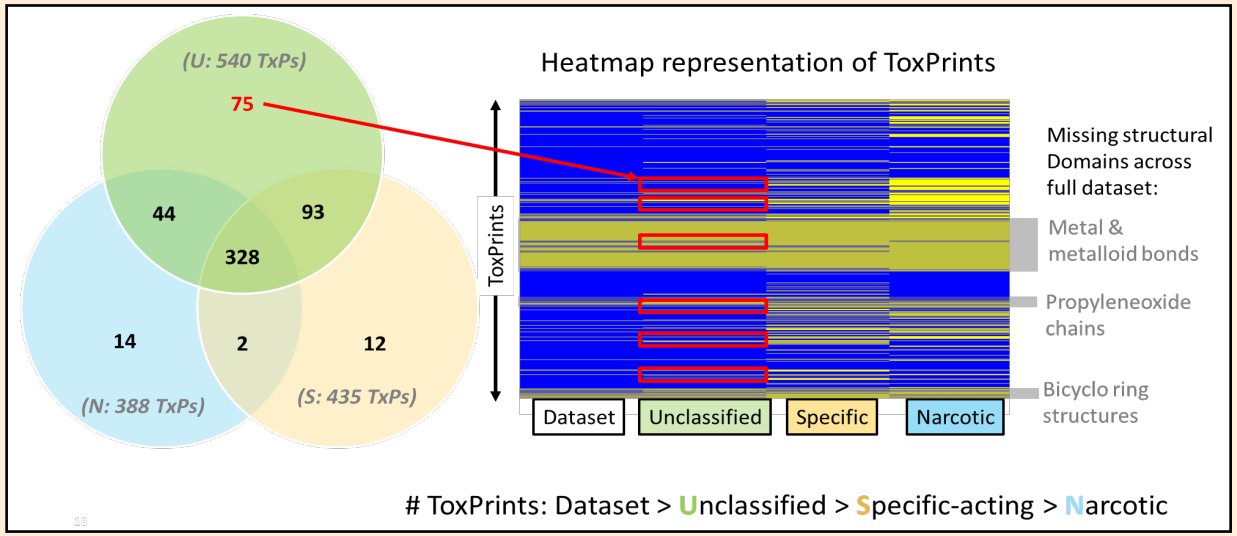
MAIN RESULTS

- High accuracy in both training and test sets
  training = 99.7%; test = 95.8%)
- Total Accuracy on all N + S data set = 97.6%
  (4356 cMOA = N or S)
- Across all N + S chemicals -> 105 chemicals misclassified:
  - 24 F_{pos}{predicted S}
  - 81 F_{neg}{predicted N}
- Lower prediction confidence in N/S classification of the U set *may* be attributable to applicability domain issues

Prediction Confidence across the Training, Test and Unclassified Sets



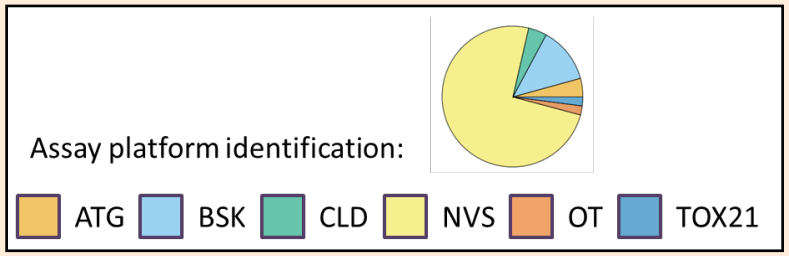
Training Set	Test Set	Unclassified Set
Median: 0.999, 0.993	Median: 0.996, 0.989	Median: 0.958, 0.941
Mean: 0.988, 0.982	Mean: 0.970, 0.962	Mean: 0.892, 0.877



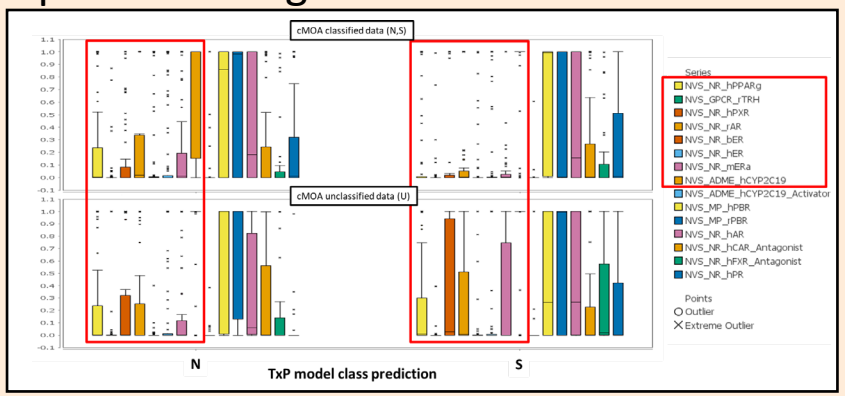
- Good overlap of existing ToxPrint (TxP) features between all 3 cMOA classes: Unclassified, Specific-acting, Narcotic
- Potential applicability domain issues for Unclassifed cMOA
  - ~7x more unique features in U (than in N or S)
  - Potential for additional categories based on structure:
    - 2 atom TxPs (metal group III)
    - 38 bond TxPs (metalloid: silane and siloxanes…)
    - 8 chain TxPs (ethyleneoxide alkanes C10 – C20)
    - 19 group TxPs (amino acids, polydentate ligands)
    - 8 ring TxPs

Revisiting and updating chemical categorizations using chemical fingerprint and high-throughput screening data

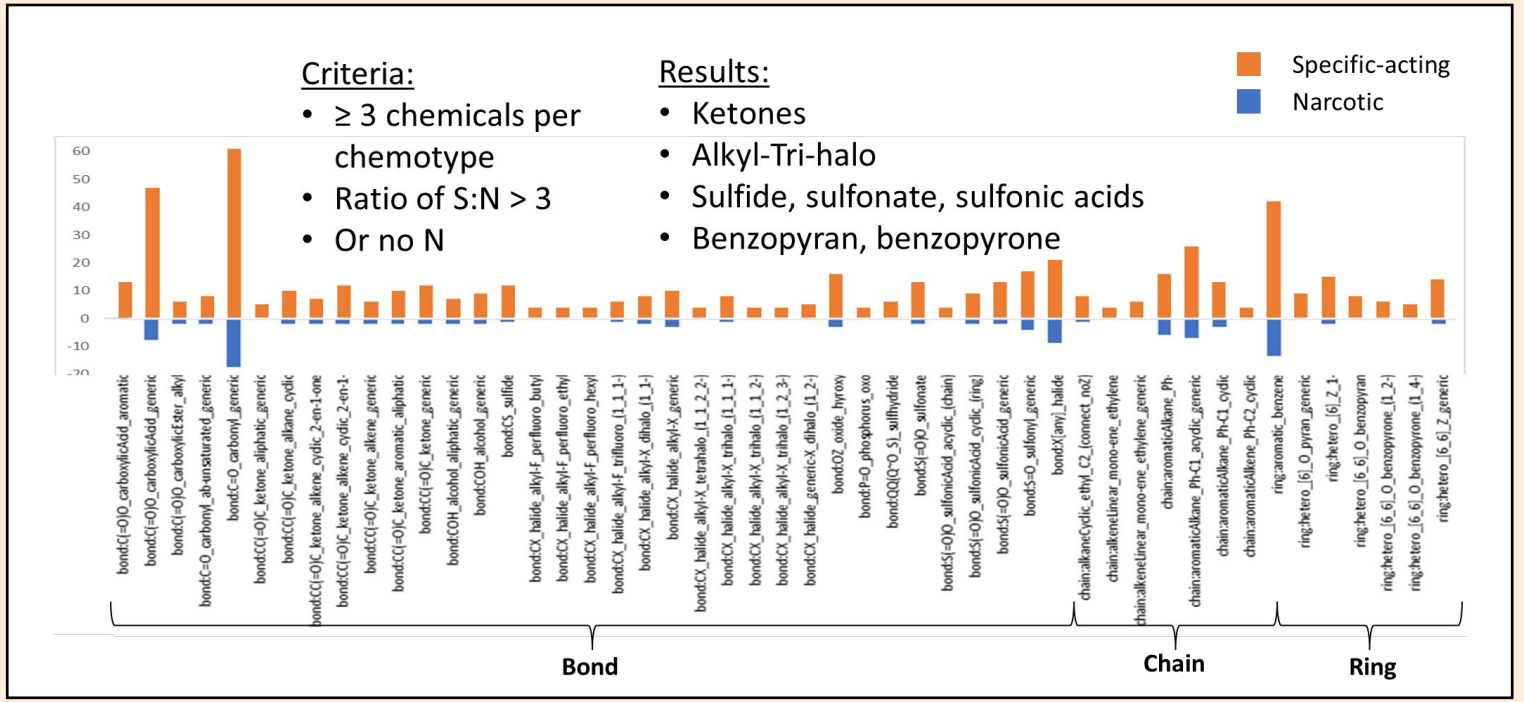
- Chemotype enrichment workflow identified ToxPrint (TxP) features that might be useful For refining chemical categories and Potential NAM assays
- Example S(=O)_sulfonyl ToxPrint is enriched in the specific-acting MOA space and 47 assays

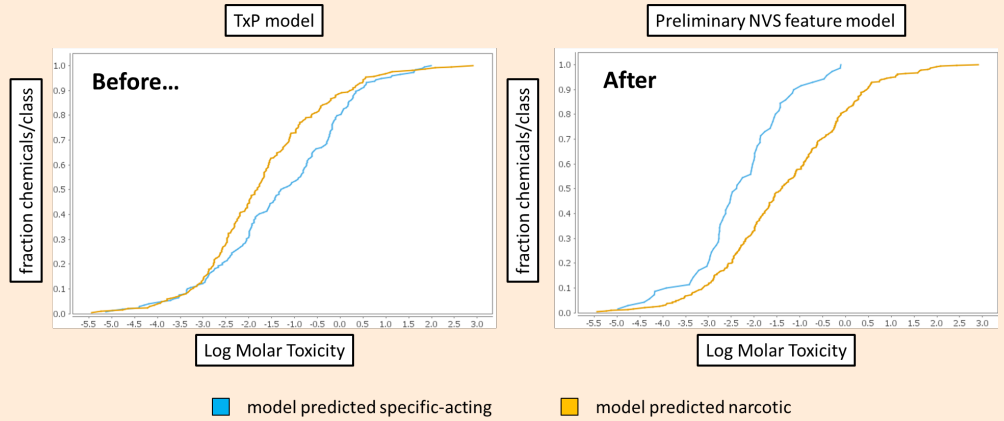


- Potential to use the specific assay platfoms to Improve existing classification of unclassified cMOA chemicals



MAIN RESULTS





Revisiting and updating chemical categorizations using chemical fingerprint and high-throughput screening data

IMPACT/SIGNIFICANCE

- Increased the available chemical space of EnviroTox w/ cMOA classifications
- Developed a robust structural TxP model
  - Robust N/S classification
  - Challenges in unclassified chemistries
- Investigated model predictions to inform ECOSAR preliminary set of unclassified chemicals
  - Majority of unclassified chemicals predicted to have a specific acting MOA
  - Identified primary chemotypes for specific-acting MOAs
- Continued work to explore methods to fold in NAM data streams
  - Using chemotype enrichments to identify potential bioassays with bioactivity to provide support of NAM data in category development

REFERENCES

[1] The Ecological Structure Activity Relationship (ECOSAR) Class Program, Version 2.0 Available: https://www.epa.gov/tsca-screening-tools/ecological-structure-activity-relationships-ecosar-predictive-model

[2] ToxCast and Tox21 data available through invitroDBv3.2 https://www.epa.gov/chemical-research/exploring-toxcast-data-downloadable-data

[3] Health and Environmental Sciences Institute (HESI). 2019. EnviroTox Database & Tools. Version 1.1.0 Available: http://www.envirotoxdatabase.org/

[4] Kienzler *et al.*. Environ Toxicol and Chem. 2019, 38(10) 2294-2304

[5] Richard *et al.*, Chem. Res. Toxicol. 2016, 29(8) 1225 – 1251; Strickland *et al.*, Arch Toxicol. 2018 92(1) 487 – 500; Wang *et al.*, Environment International 2019, 126 377 – 386