United States Environmental Protection Agency

Revisiting and updating chemical categorization with new approach methodologies: *Lessons learned*

US EPA in collaboration with Health Canada, Environment Climate Change Canada

APCRA
November 11, 2022

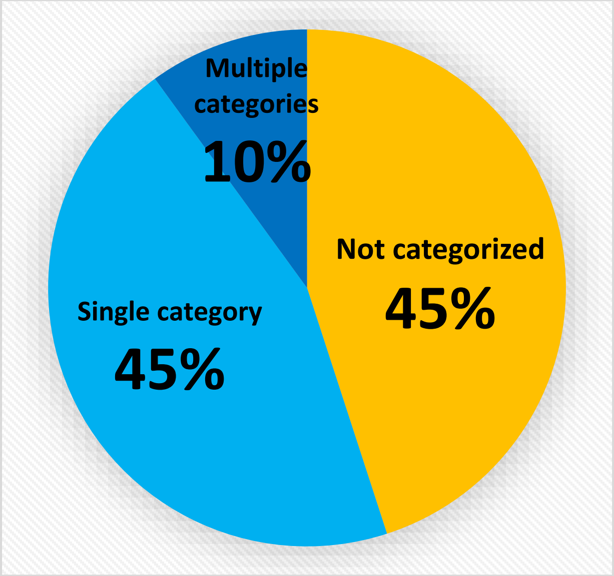
Disclaimer: The views expressed in this presentation are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency. This presentation has not been reviewed for policy and is not for distribution.

Team members

[Accelerating the Pace of Chemical Risk Assessment (APCRA)]

- US EPA
  - Dan Chang
  - Kellie Fay
  - Kristan Markey
  - Martin Phillips
  - Grace Patlewicz
  - Ann Richard
  - Gino Scarano
  - Mahmoud Shobair
  - Ryan Lougee
  - Ellery Saluck (summer intern)
- Environment & Climate Change Canada (ECCC)
  - John Prindiville
  - Cristina Inglis
- Health Canada
  - Mark Lewis
- ILS
  - Kamel Mansouri

Background: Chemical categorization



- "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
- 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.
- Almost half of all New Chemical inventories across regulatory jurisdictions cannot be categorized using NCC or ECOSAR.
- Some chemicals fall into multiple categories.

Case study objectives

- ***How do we incorporate new approach methodologies (NAMs) and cheminformatic approaches to assist in identifying new chemical categories?***
- ***Can we use a classified consensus Mode-of-Action (cMOA) dataset (supervised learning) to develop a robust classification model to discriminate between narcotic (N) and specific-acting (S) chemicals for aquatic (fish) toxicity?***

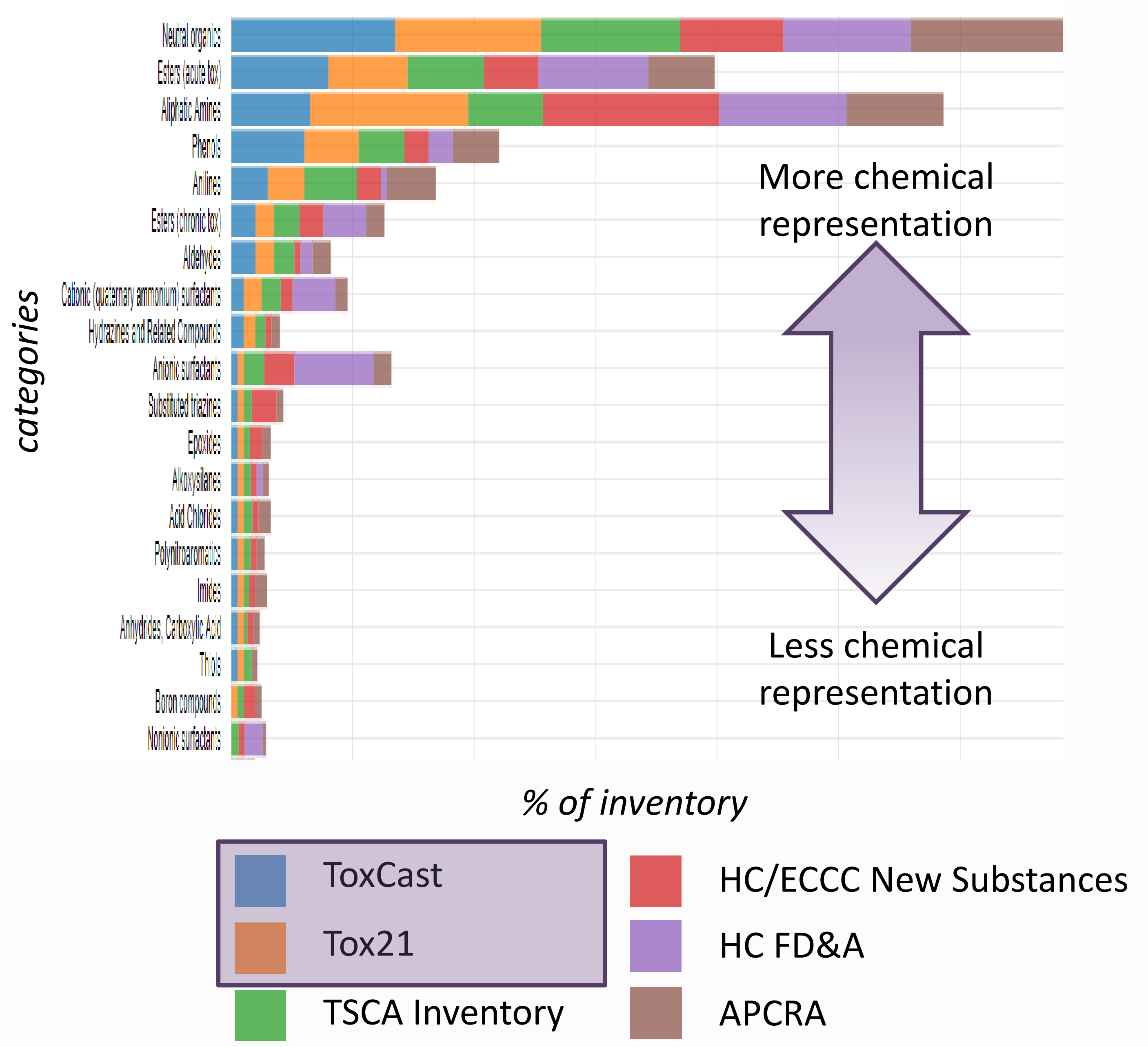
Key case study accomplishments

1. Identified the landscape of regulatory and bioassay chemical inventories based on New Chemical Category (NCC) definitions and ToxPrints (TxP).
2. Developed of a robust N/S classification model for aquatic toxicity.
3. Identified known limitations regarding unclassified cMOA chemicals.
4. Suggested targeted use of NAM information (*i.e.*, use of specific assay data) with Chemotype enrichment workflows.

Office of Research and Development
Center for Computational Toxicology and Exposure

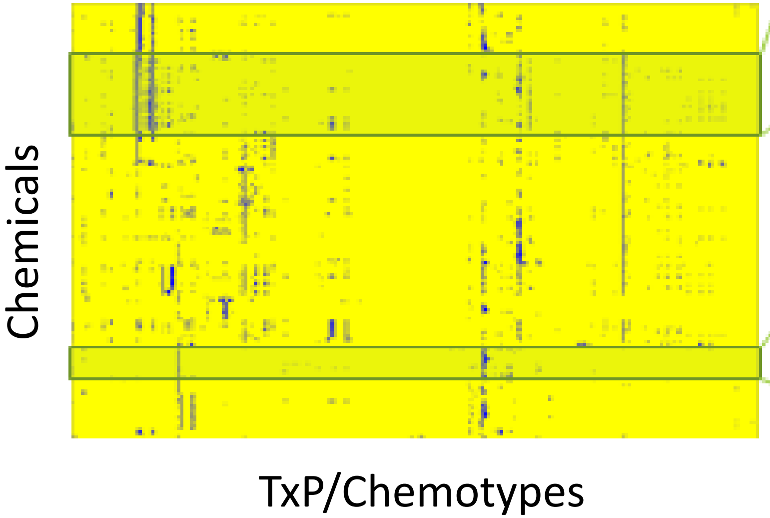
1. Identification of the chemical landscape

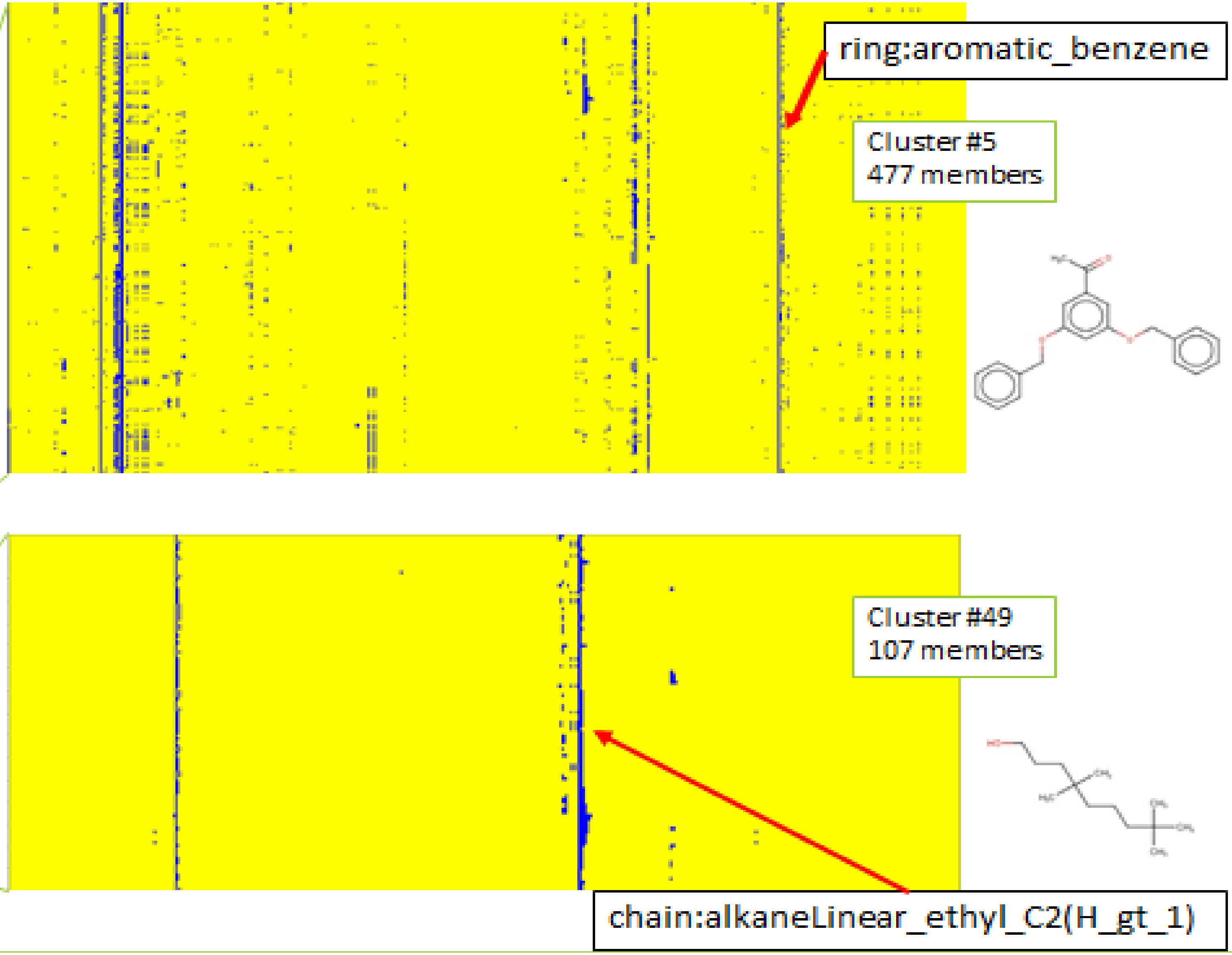
- Can we use existing NAM data rather than traditional *in vivo* data?
- 57 NCC chemical categories are based on structure, phys chem properties and existing *in vivo* data.
- ToxCast and Tox21 HTS data overlaps well with existing regulatory chemical inventories.
- ***For TSCA and APCRA inventories, ToxCast and Tox21 may be appropriate potential surrogates for NAM-based information in larger categories – Neutral Organics, Phenols, etc.***



TSCA Neutral Organics

- TxP analysis with hierarchical clustering
- Identified unique clusters of chemistries/representations
- Potential implications for risks/endpoints





Heatmap of Neutral Organics based on ToxPrint features, hierarchical clustering

2. Classification model for aquatic toxicity

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

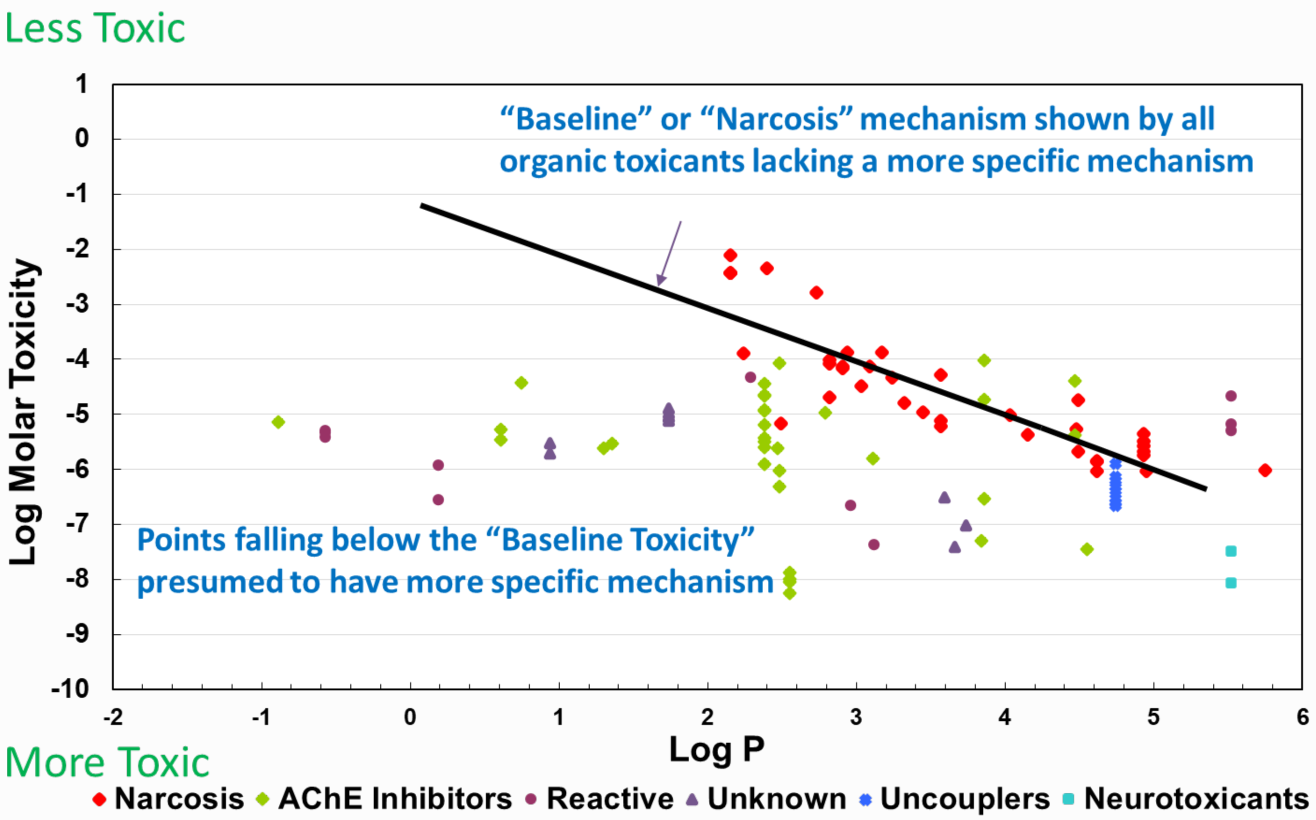
Applications of chemical categorization include first tier assessment efforts and read across from structurally similar analogs – ECOSAR

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

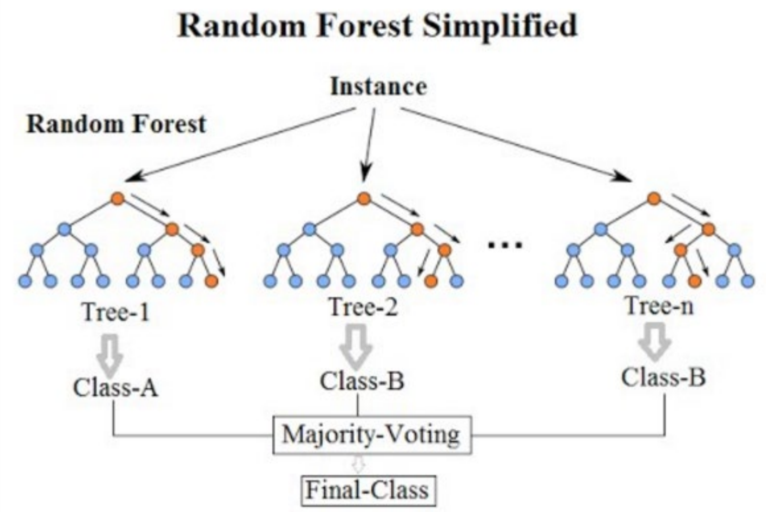


1. ***Regulators consider MOA information to determine the size of assessment factors***
2. ***Can we develop a viable model to use as a tool to discriminate specific-acting & narcotic MOAs for potential category development?***

United States Environmental Protection Agency

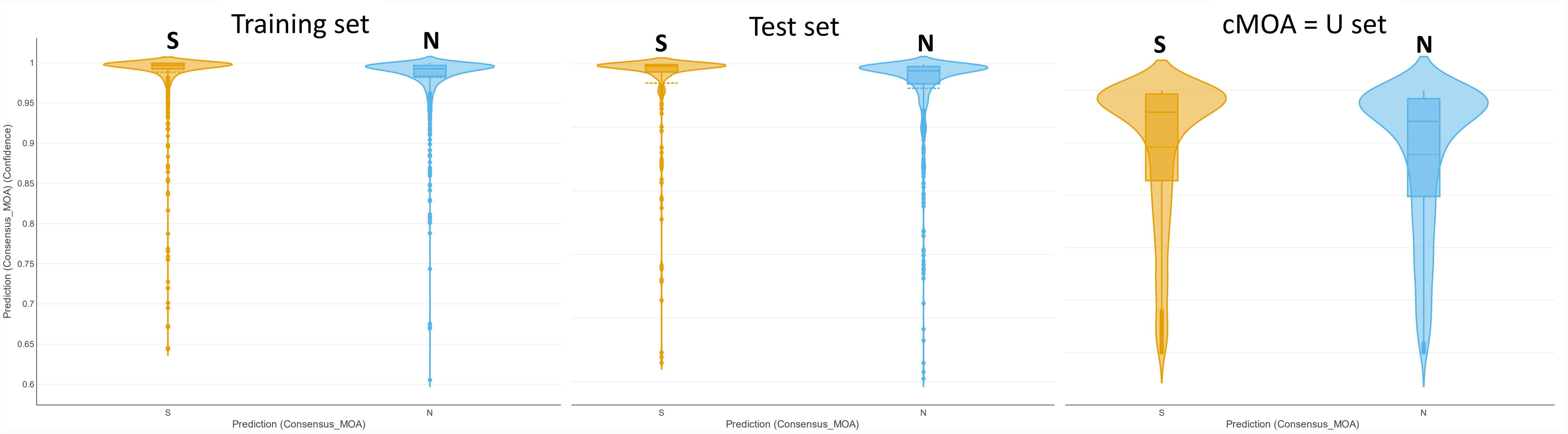
TxP model details

- **Dataset: supervised learning via Consensus MOA (cMOA) dataset**
  - EnviroTox Database: Aquatic toxicity *in vivo* dataset with a consensus call based on 4 structure based models (Health and Environmental Sciences Institute (HESI). 2019. EnviroTox Database & Tools. Version 1.1.0 Available: http://www.envirotoxdatabase.org/)
- **Features: ToxPrints**
  - 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.
- **Method: Random Forest (Boosted Gradient Method)**
  - 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/ a cMOA = N or S)
  - 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}



https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d

Distribution of prediction confidence [0,1] by (N,S) class



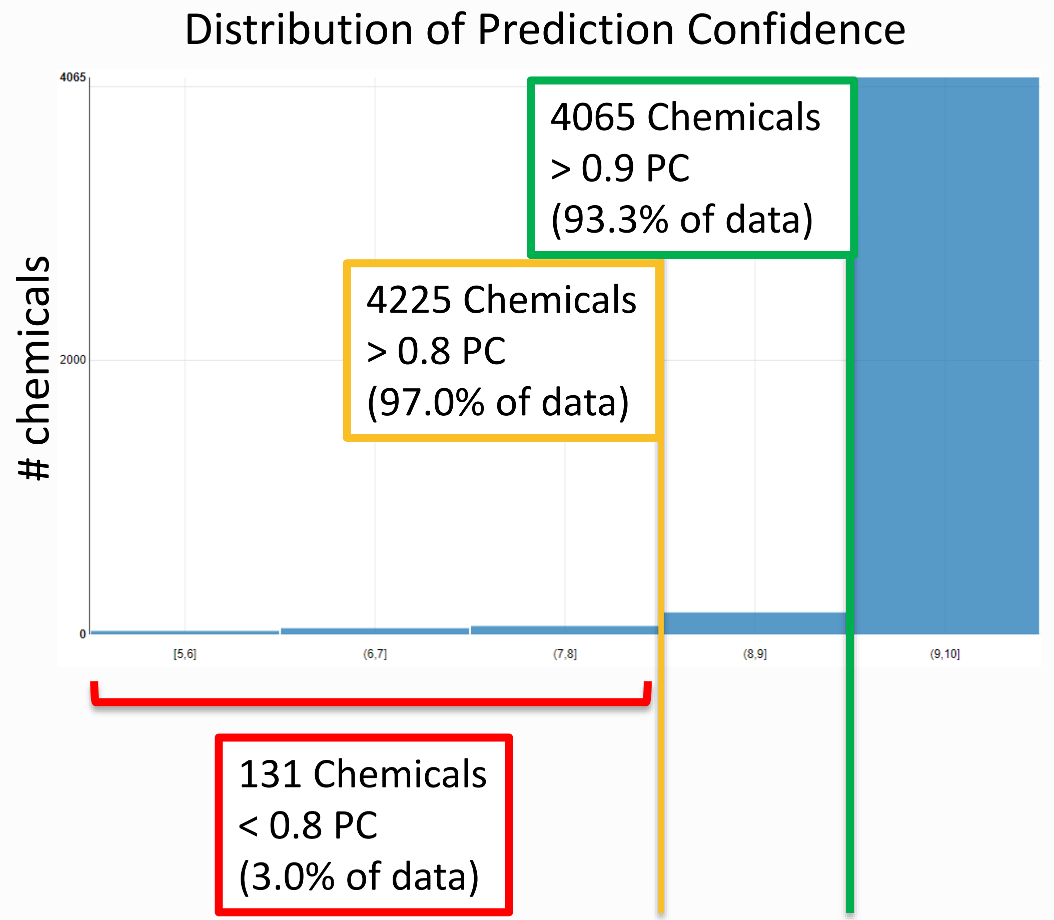
Training Set
Median: 0.999, 0.993
Mean: 0.988, 0.982

Test Set
Median: 0.996, 0.989
Mean: 0.970, 0.962

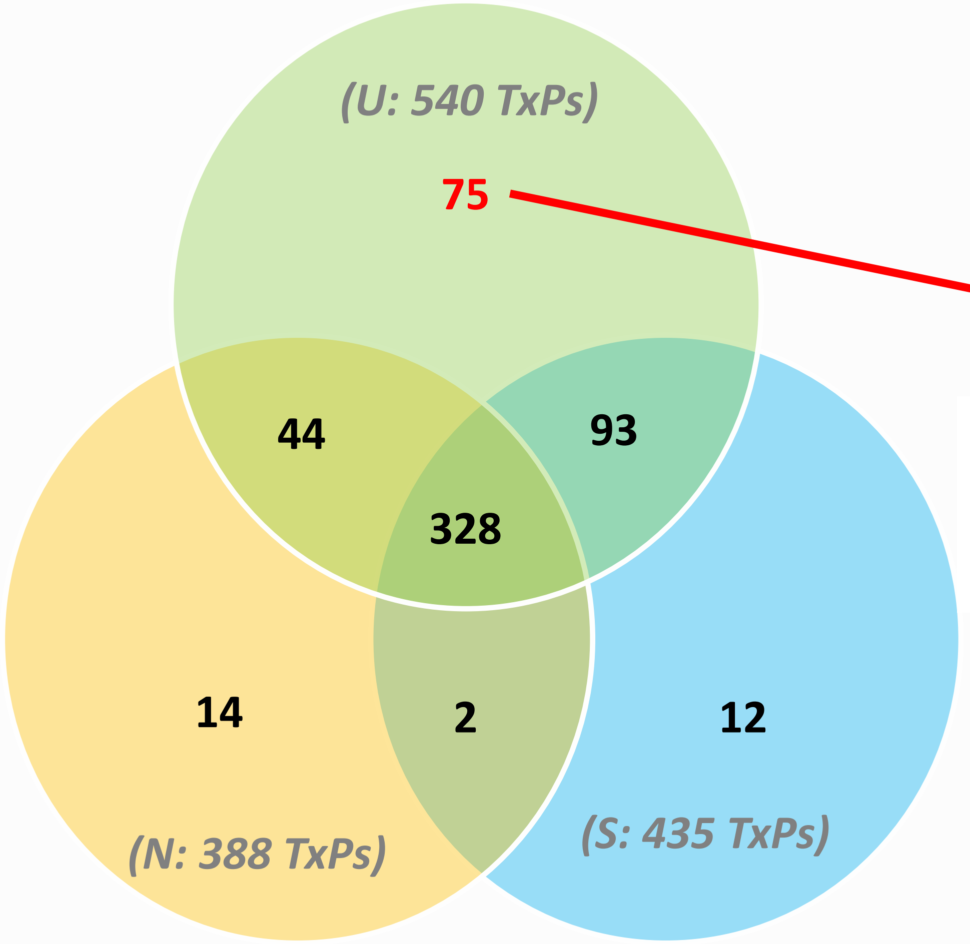
Unclassified Set
Median: 0.958, 0.941
Mean: 0.892, 0.877

Prediction confidence across the cMOA = N or S

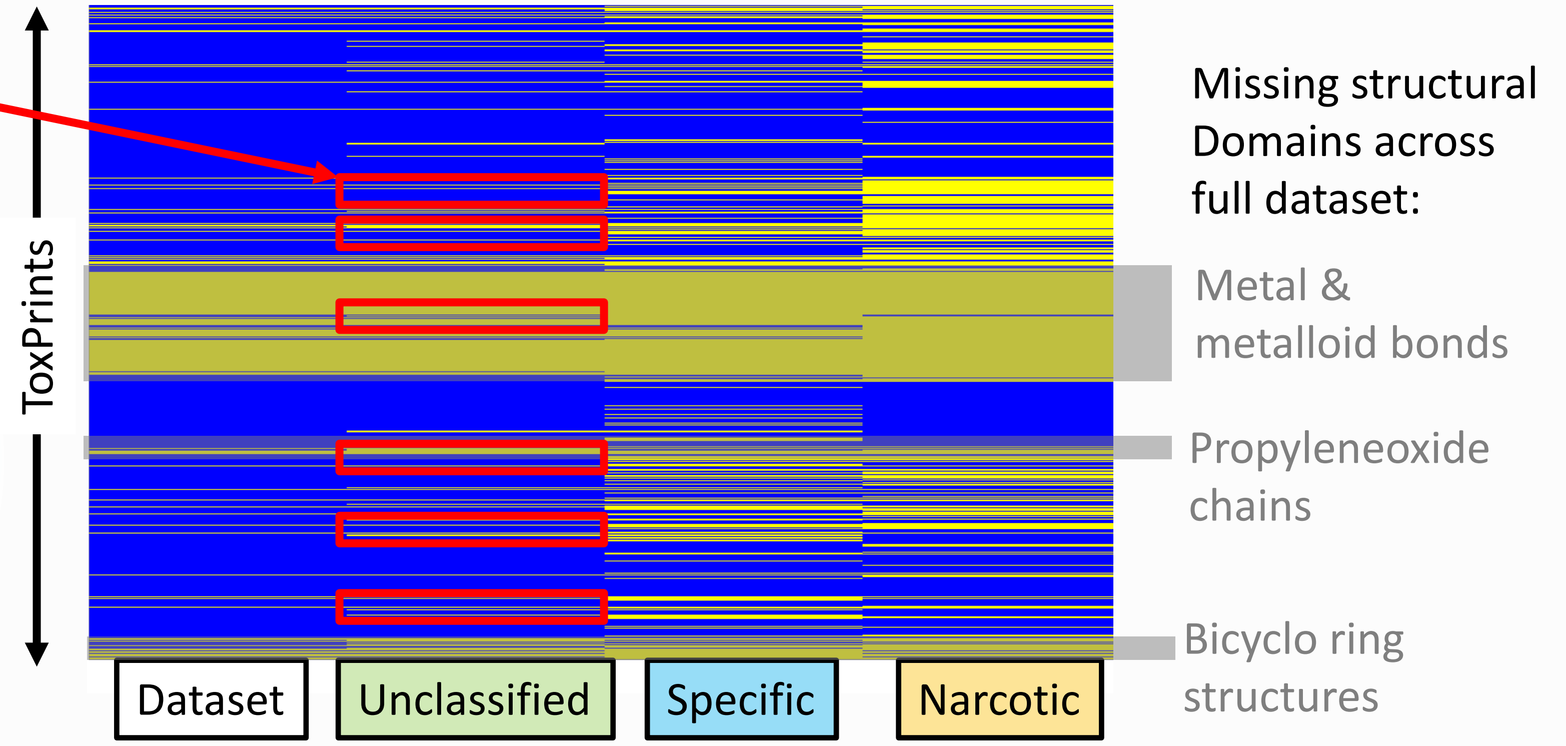
- Distribution of prediction confidence (PC) tends to be > 0.8 for the classified data (cMOA = N or S)
- Model has fewer # misclassifications in S
  - Misclassifications for 93 cMOA confidence = 2, and 12 with 1,3 scores (recall 3>2>1 for confidence)
  - **~46% of the misclassifications can be attributed to the chemicals with PC < 0.8**
  - ~67% of the misclassification can be attributed to chemicals with PC < 0.88



3. Analysis of unclassified consensus MOA chemicals: Characterization of TxP coverage per consensus MOA class



Heatmap representation of ToxPrints

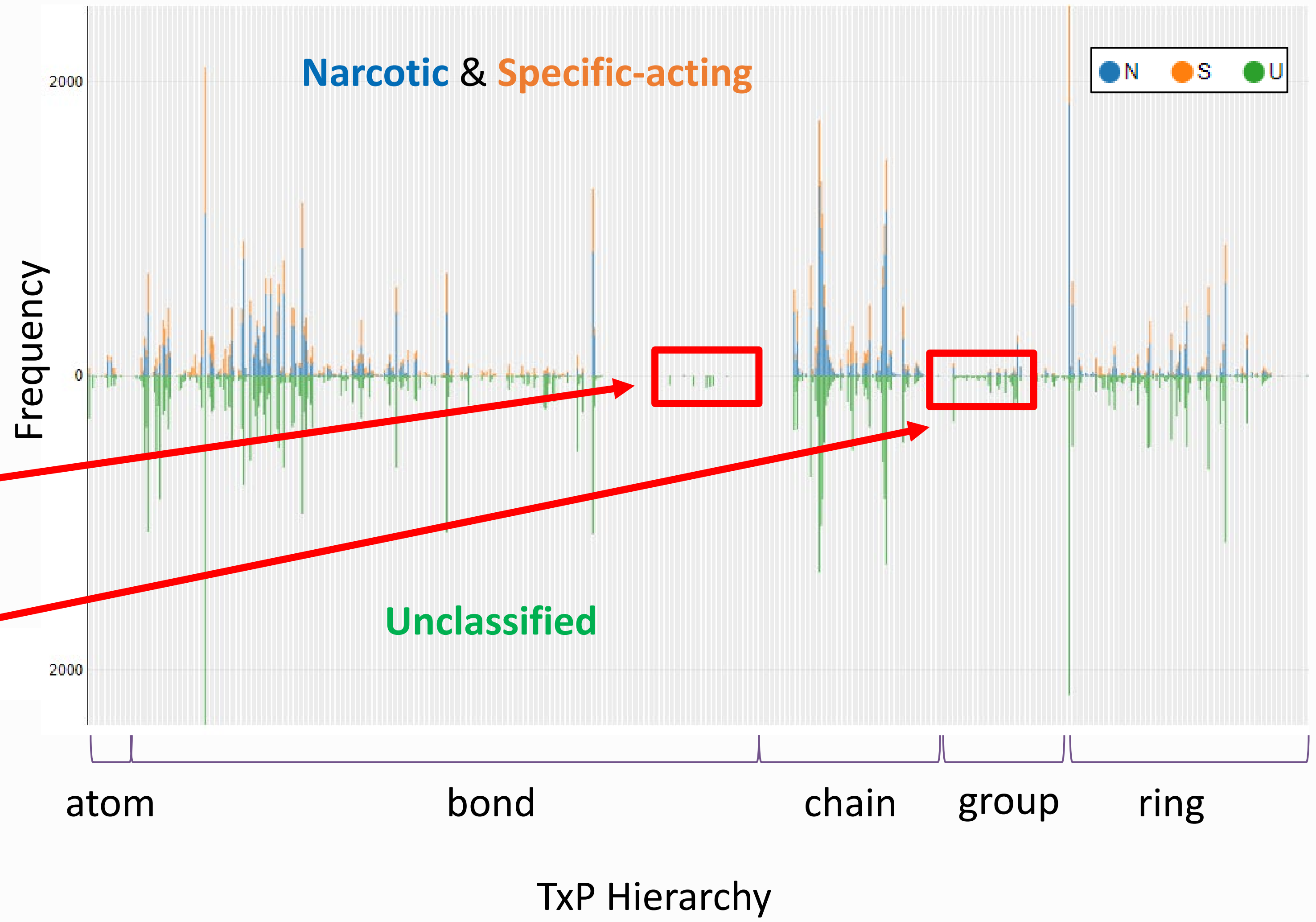


ToxPrints: Dataset > Unclassified > Specific-acting > Narcotic

Unique TxPs in the unclassified set

- ~7x more unique features in U (than in N or S)
- Could explain the lower prediction confidence in N/S classification of the U set
- 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

Frequency of TxPs per consensus MOA class

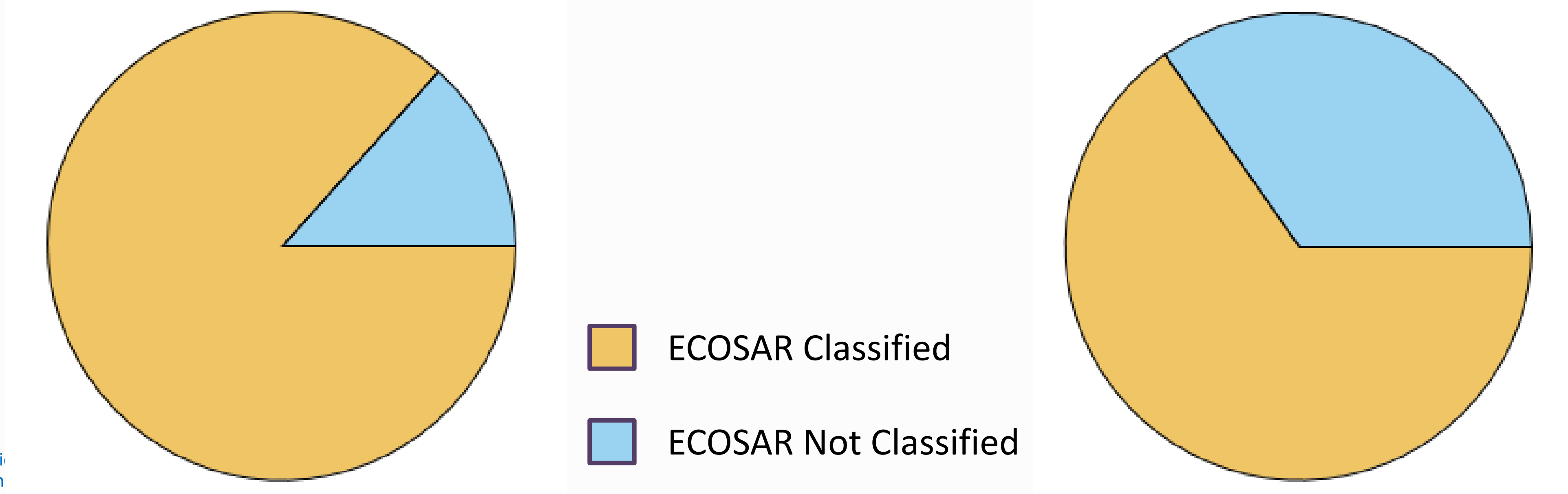


TxP model predicted MOAs of the EnviroTox unclassified set

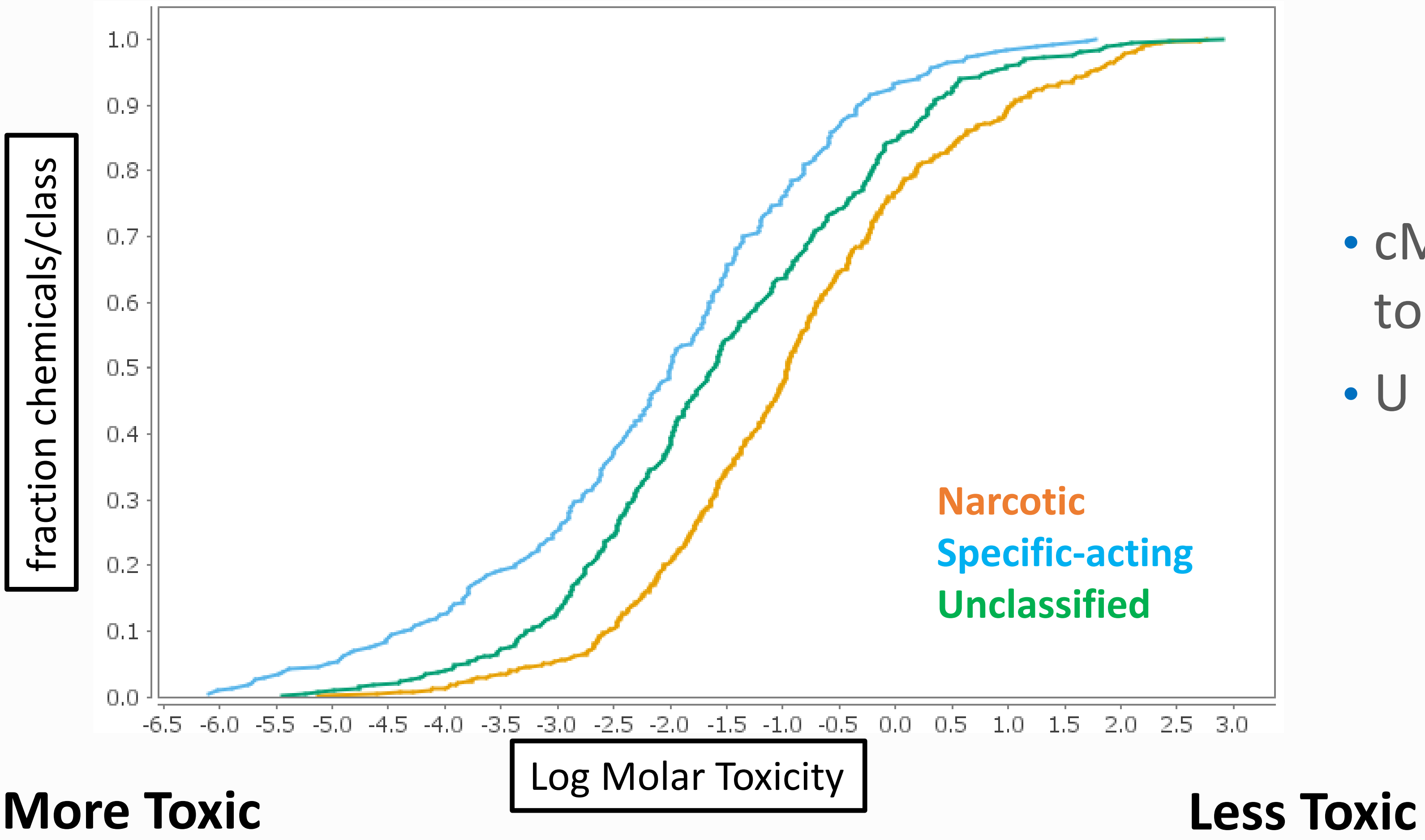
- 674 subset chemicals in the EnviroTox dataset that had low confidence or ambiguous consensus (unclassified)
- Applied TxP model to the unclassified set and compared predictions to ECOSAR classification

361 predicted as Narcotic

313 predicted as Specific-acting



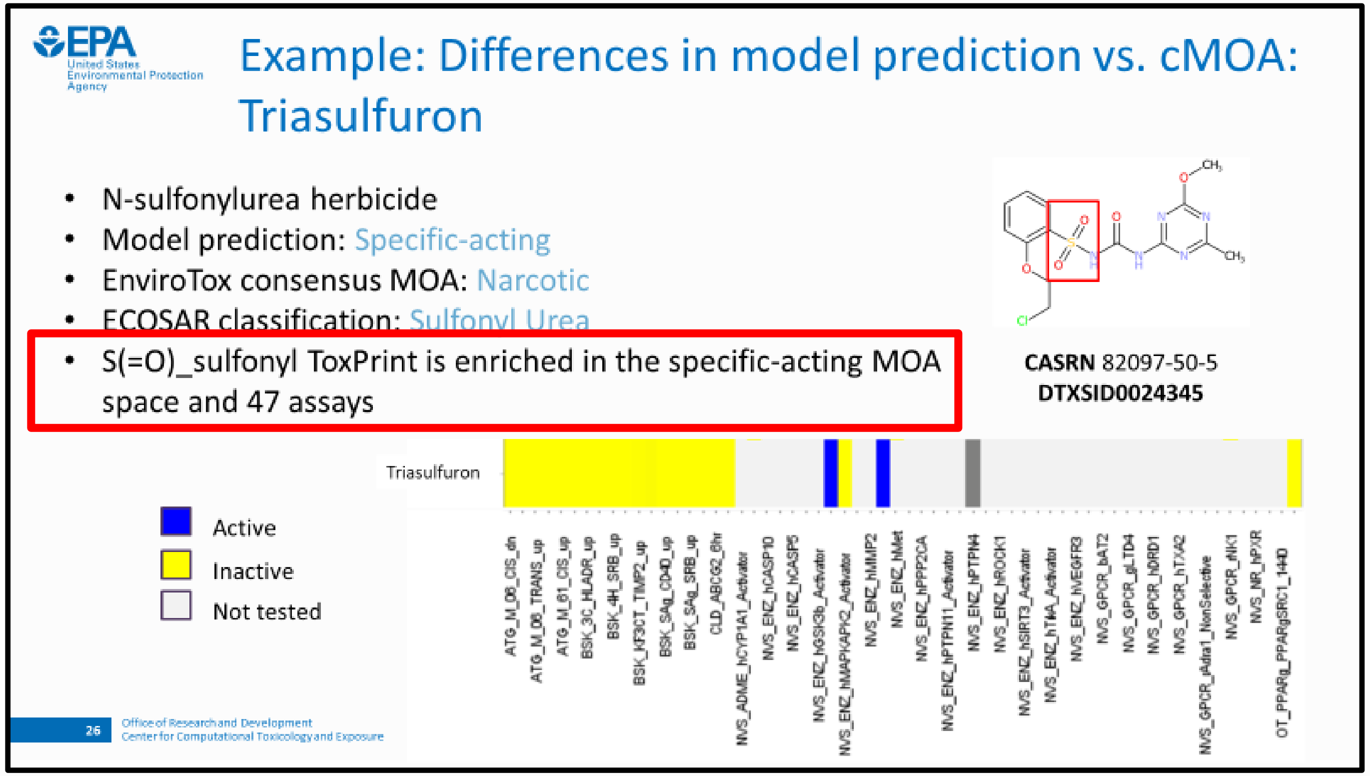
Cumulative distribution function: Log molar toxicity, (LC50, 96h, FISH) for cMOA classes (N,S,U)



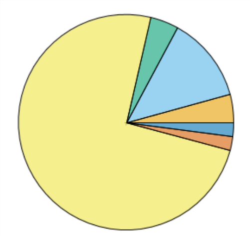
- cMOA classification is sufficient to discriminate N,S
- U presents some challenges

4. Targeted use of NAM information and Txp/Chemotypes

- Use chemotype enrichments to inform potential NAM data streams
- Example: sulfonyl TxP enrichments across NovaScreen (NVS) assay platform
- "…assays measure chemical binding to nuclear receptors, G-protein-coupled receptors (GPCR), transporters, and ion channels, and enzymatic inhibition or activation for a range of proteins including kinases, phosphatases, CYP450s, proteases, and histone deacetylases."
- Identified 47 assays due to sulfonyl TxP enrichment



Assay platform identification:



ATG | BSK | CLD | NVS | OT | TOX21

United States Environmental Protection Agency

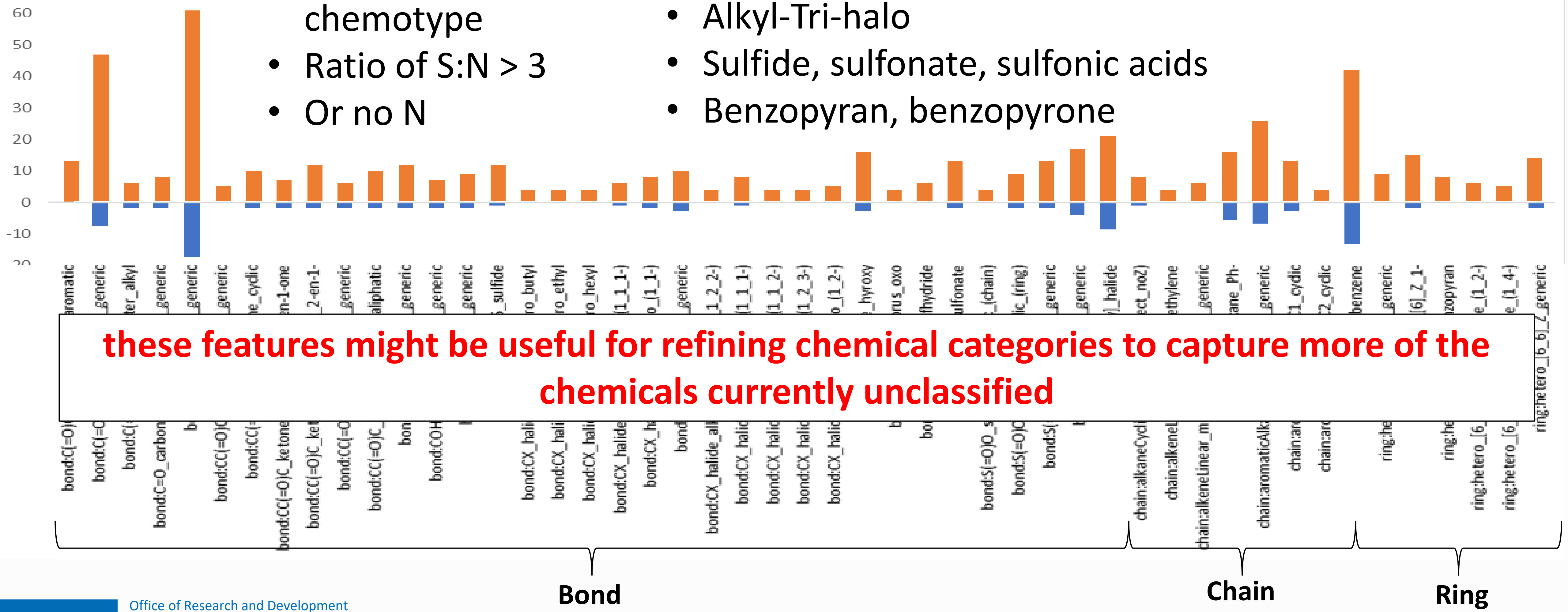
Enriched TxPs: Unclassified chemicals, TxP model predicted specific-acting

Criteria:

- ≥ 3 chemicals per chemotype
- Ratio of S:N > 3
- Or no N

Results:

- Ketones
- Alkyl-Tri-halo
- Sulfide, sulfonate, sulfonic acids
- Benzopyran, benzopyrone



these features might be useful for refining chemical categories to capture more of the chemicals currently unclassified

United States Environmental Protection Agency

Summary and Final Steps

- Developed a robust structural TxP model
  - Good N/S classification
  - Challenges in unclassified chemistries
- Investigated model predictions to inform ECOSAR subset of unclassified chemicals
  - Some unclassified chemicals predicted as potentially specific-acting MOAs
  - Identified primary chemotypes for specific-acting MOAs
- Explored 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
- Developing a manuscript on the existing TxP model and analyses

Thank you!