Analysis of the ToxCast & Tox21 compound set using regulator-derived GHS toxicity annotations and *in silico*-derived protein-target descriptors

Chad H.G. Allen, Lewis H. Mervin and Andreas Bender

11th International Conference on Chemical Structures
28 May 2018

Overview

- Introduction:

  Motivation, prior work

- The Globally Harmonized System:

  Overview, utility

- Dataset collation:

  ToxCast dataset, protein target prediction, GHS annotation pipeline

- Dataset analysis:

  Toxicophore/GHS relationships, nearest neighbour distances in different descriptor spaces, linear discriminant analysis

- Modelling results:

  Predictive performance of Random Forests classification models

- Conclusions

Introduction

Motivation:
Need for toxicity data outstrips output of traditional toxicology

- There is increased regulatory demand for data on the safety of chemicals, e.g. the EU's REACH, the US's TSCA, Canada's CMP

- However, there is also increased regulatory/societal/economic pressure to "replace, reduce and refine" traditional *in vivo* toxicity studies

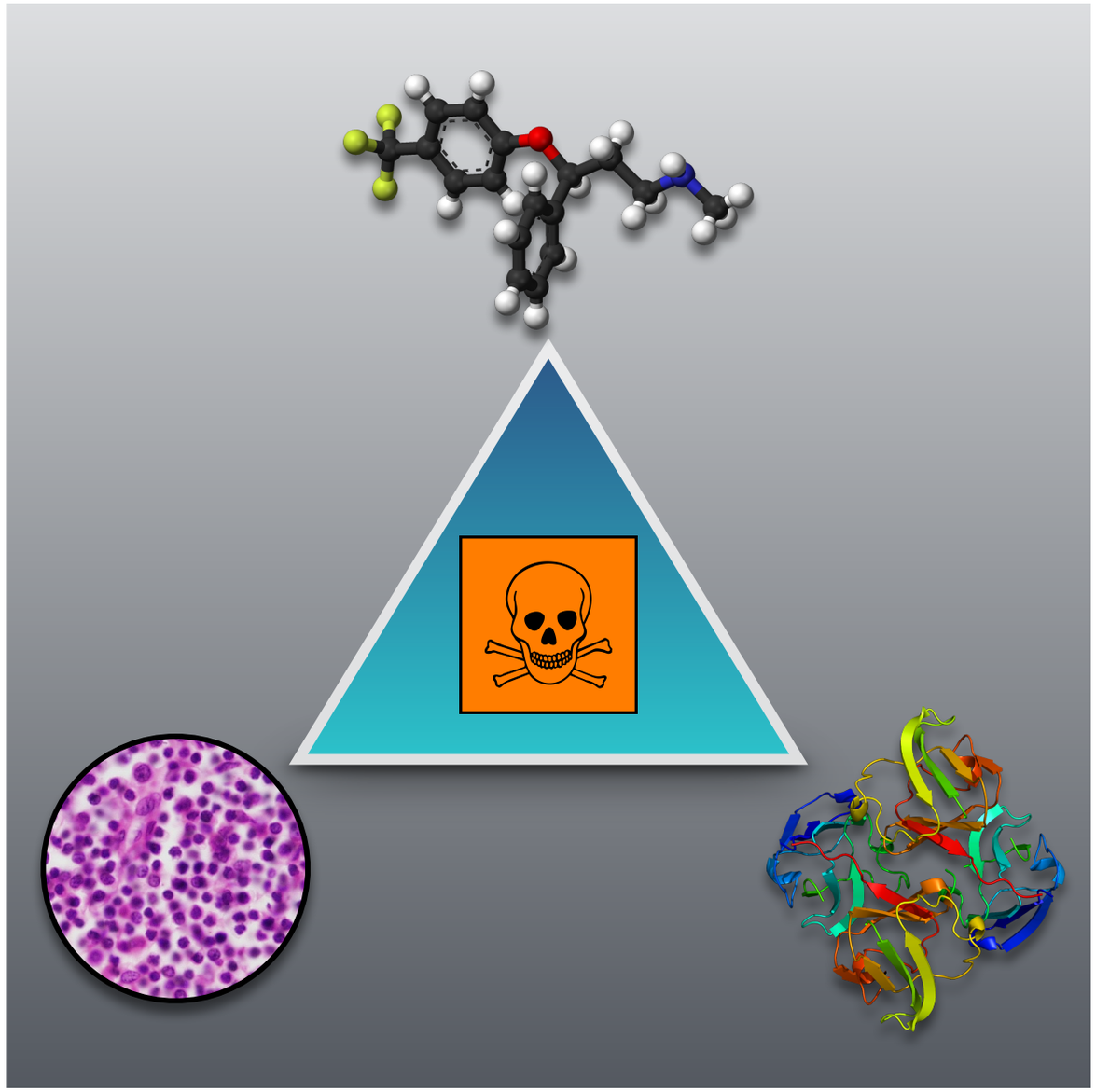
- Applying tradition toxicology techniques to the vast quantity of data-poor chemicals requiring evaluation is not practically possible – and translating animal results to humans is nontrivial

- Therefore, increased demand for novel approaches, including *in silico* methods

see e.g. Kavlock et al. *Chem. Res. Toxicol.,* **2018**, 31, 287–290

Prior work:

Integration of heterogenous data can improve performance of toxicity models



Prior work:
Integration of heterogenous data can improve performance of toxicity models

- Improved accuracy of acute oral rat toxicity QSAR models using in vitro HTS data previously demonstrated by A. Sedykh, A. Tropsha and colleagues[1]

- A tripartite, heterogeneous descriptor set for 367 compounds was comprised of:

  (a) chemical descriptors

  (b) descriptors derived from in vitro cell cytotoxicity dose–response data from a panel of human cell lines

  (c) protein target descriptors generated using an algorithm trained on 190,000 ligand–protein interactions from ChEMBL

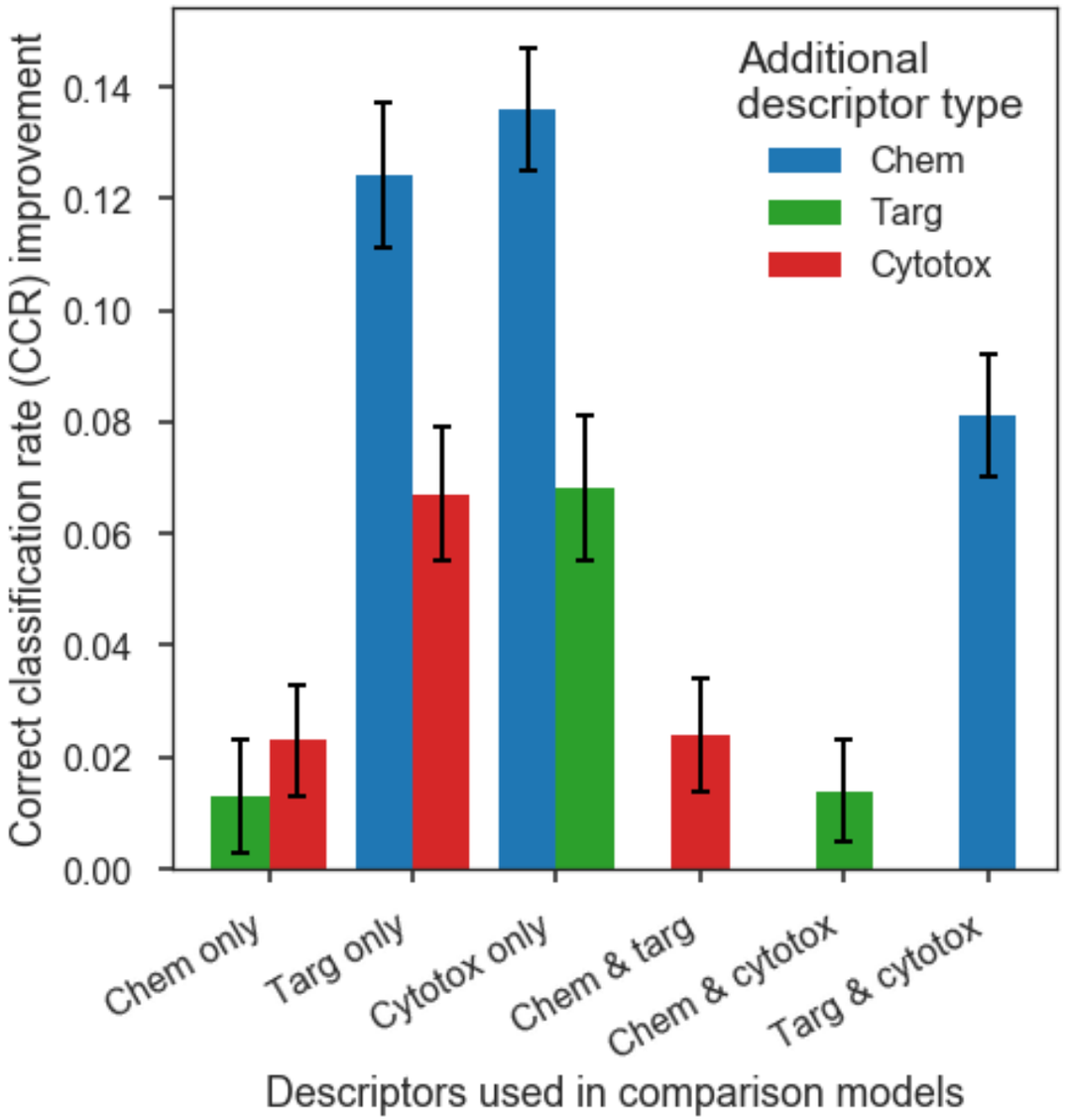
- This dataset was used to build Random Forests classification models, and the performance and interpretability of the models compared on successive integration of data types[2]

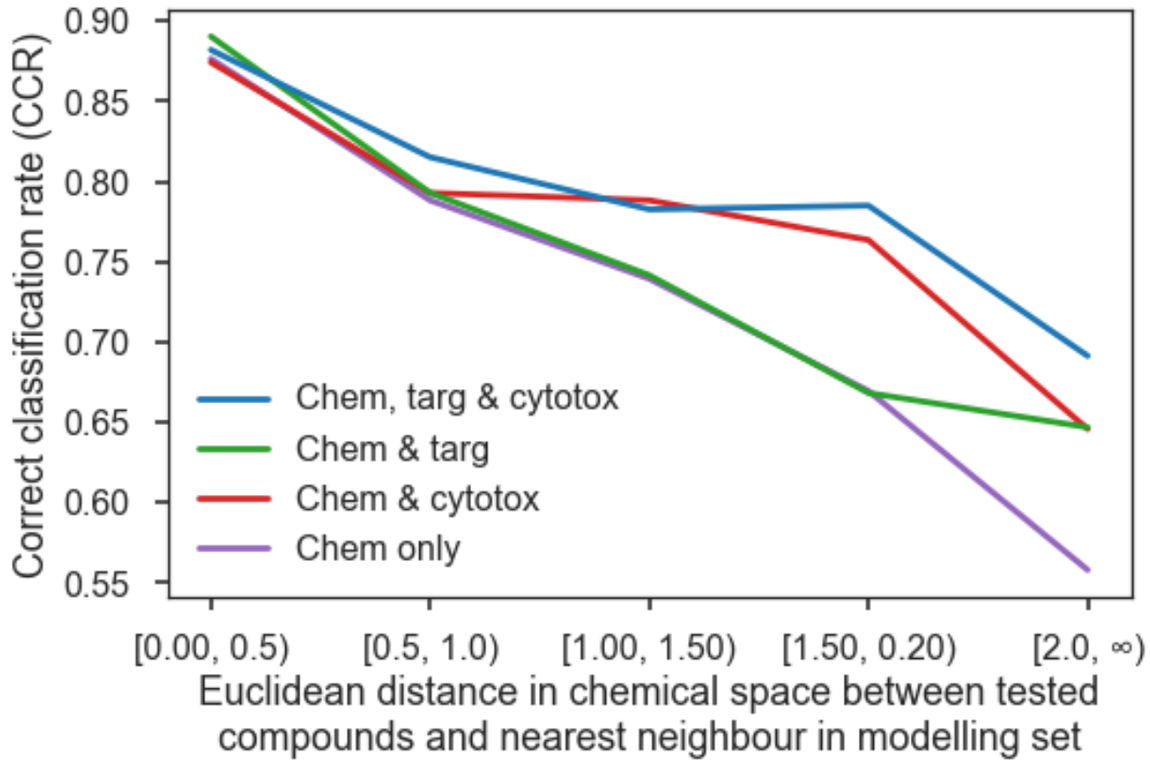
1. Sedykh et al. *E.H.P.,* **2011**, 119, 364–370

2. Allen et al. *Tox. Res.,* **2016**, *5*, 883-894

Prior work:

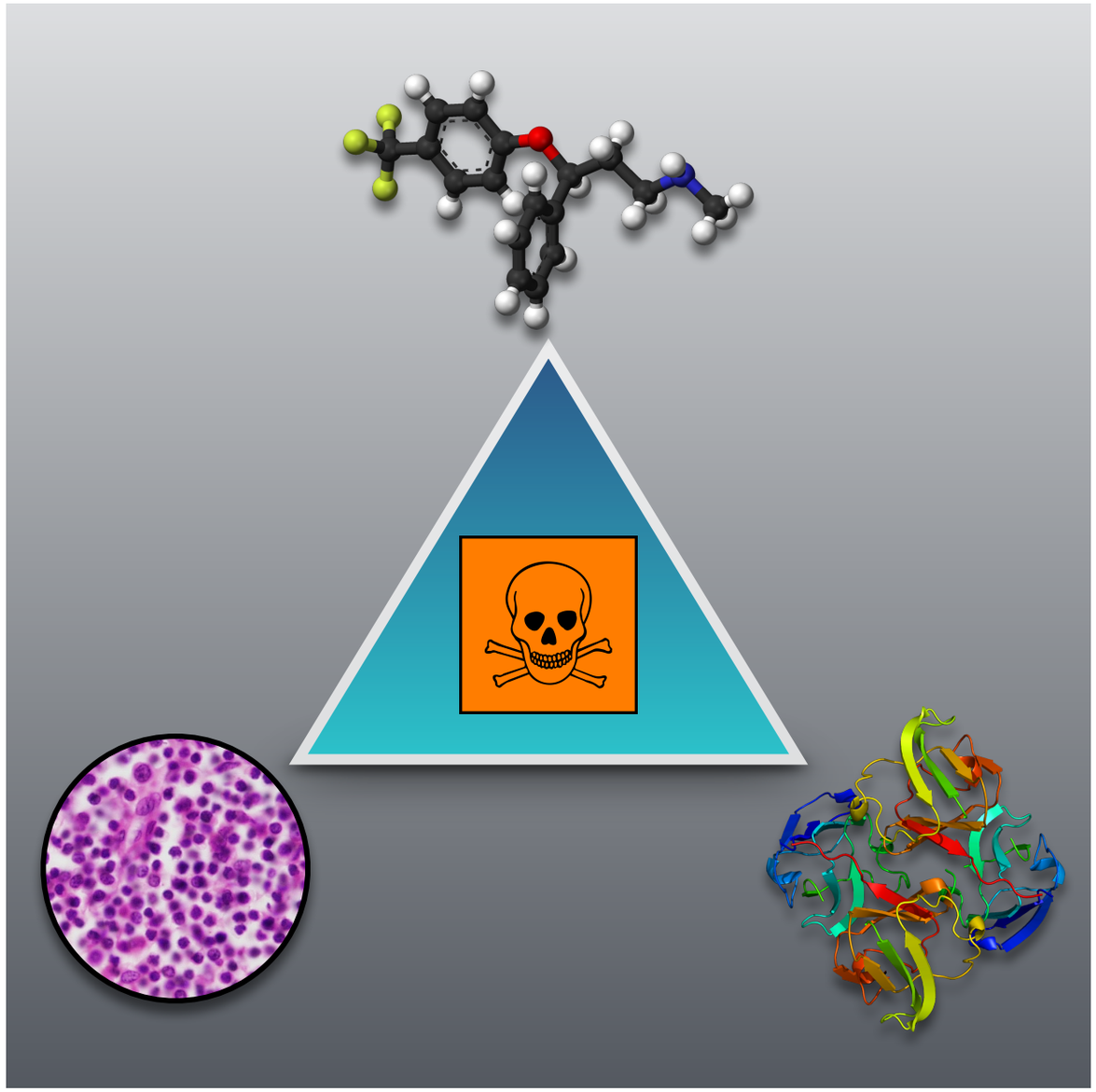
Integration of heterogenous data can improve performance of toxicity models



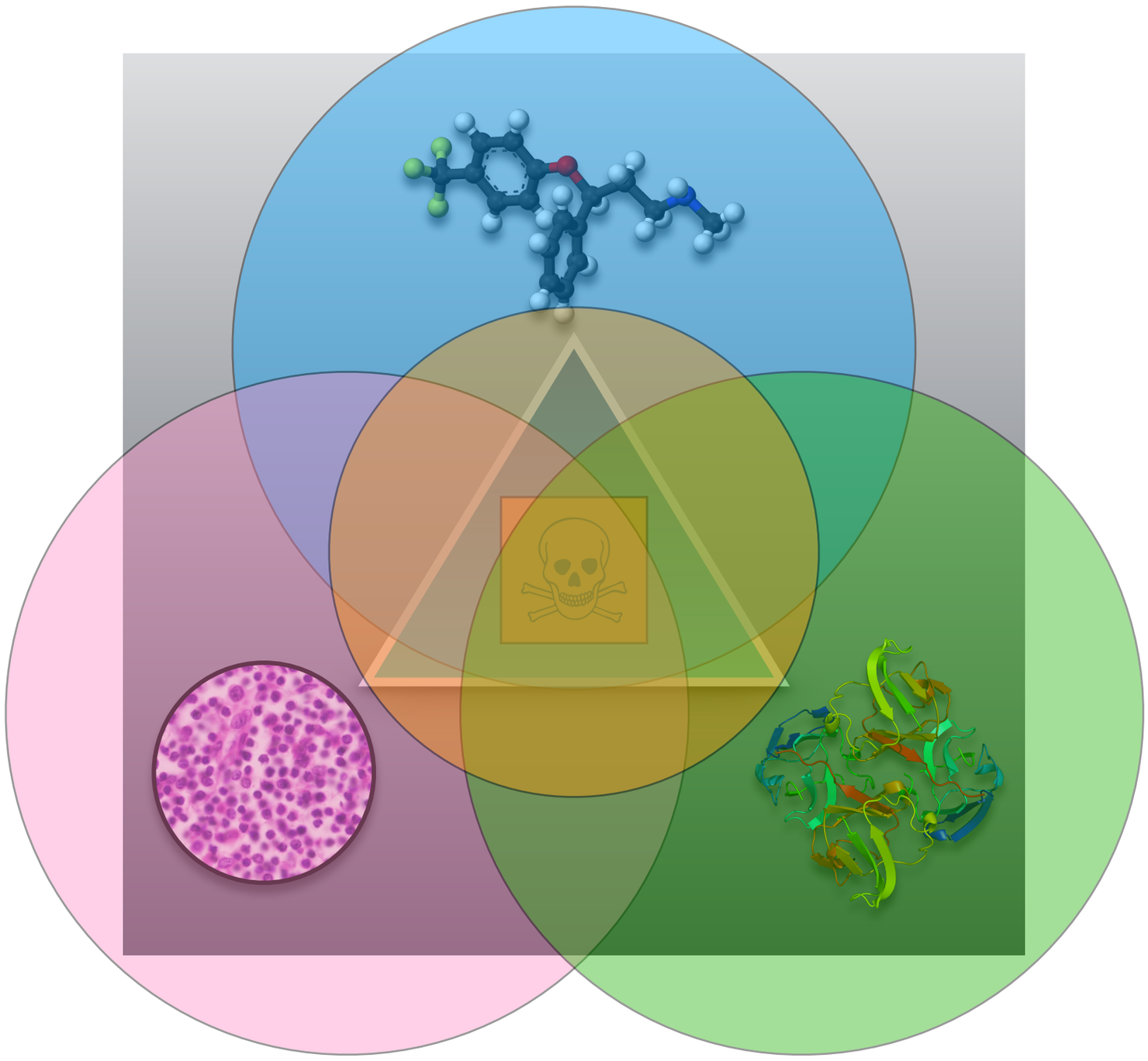


Allen et al. *Tox. Res.,* **2016**, *5*, 883-894

Practical difficulty:
Generating dataset with sufficient overlap of data domains



Practical difficulty:
Generating dataset with sufficient overlap of data domains



The Globally Harmonized System

GHS:
Hazard pictograms

Explosive

Flammable

Oxidizing

Compressed gas

Corrosive (chemically)

Toxic

Corrosive (to people!)

Irritant

Health hazard

Environmentally damaging

GHS:
Pictograms derived from Categories

	Category 1	Category 2	Category 3	Category 4	Category 5
Symbol	Skull and crossbones	Skull and crossbones	Skull and crossbones	Exclamation mark	*No symbol*
Signal word	Danger	Danger	Danger	Warning	Warning
Hazard statement:					
--Oral	Fatal if swallowed	Fatal if swallowed	Toxic if swallowed	Harmful if swallowed	May be harmful if swallowed
--Dermal	Fatal in contact with skin	Fatal in contact with skin	Toxic in contact with skin	Harmful in contact with skin	May be harmful in contact with skin
--Inhalation see Note	Fatal if inhaled	Fatal if inhaled	Toxic if inhaled	Harmful if inhaled	May be harmful if inhaled

Chapter 3.1: Acute Toxicity
In *Globally Harmonized System of Classification and Labelling of Chemicals (GHS)* [Online], 7th revised ed. https://www.unece.org/trans/danger/publi/ghs/ghs_rev07/07files_e.html (accessed April 2018)

GHS:
Categories derived from quantitative data

Exposure route	**Category 1**	**Category 2**	**Category 3**	**Category 4**	**Category 5**
Oral (mg/kg bodyweight) *See notes (a) and (b)*	ATE ≤ 5	5 < ATE ≤ 50	50 < ATE ≤ 300	300 < ATE ≤ 2000	2000< ATE≤ 5000 *See detailed criteria in Note (g)*
Dermal (mg/kg bodyweight) *See notes (a) and (b)*	ATE ≤ 50	50 < ATE ≤ 200	200 < ATE ≤ 1000	1000 < ATE ≤ 2000	
Gases (ppmV) *See notes (a), (b) and (c)*	ATE ≤ 100	100 < ATE ≤ 500	500 < ATE ≤ 2500	2500 < ATE ≤ 20000	*See detailed criteria in Note (g)*
Vapours (mg/l) *See notes (a), (b), (c), (d) and (e)*	ATE ≤ 0.5	0.5 < ATE ≤ 2.0	2.0 < ATE ≤ 10.0	10.0 < ATE ≤ 20.0	
Dusts and Mists (mg/l) *See notes (a), (b), (c) and (f)*	ATE ≤ 0.05	0.05 < ATE ≤ 0.5	0.5 < ATE ≤ 1.0	1.0 < ATE ≤ 5.0	

Chapter 3.1: Acute Toxicity
In *Globally Harmonized System of Classification and Labelling of Chemicals (GHS)* [Online], 7th revised ed. https://www.unece.org/trans/danger/publi/ghs/ghs_rev07/07files_e.html (accessed April 2018)

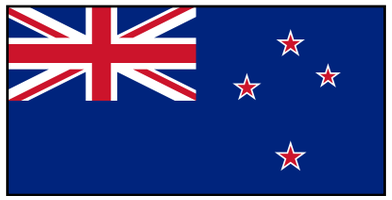
Public regulatory GHS data sources

safe work australia

5,577 classifications

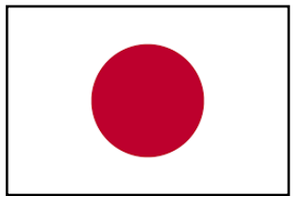
ECHA
EUROPEAN CHEMICALS AGENCY

141,475 classifications



Environmental Protection Authority
Te Mana Rauhī Taiao

5,450 classifications



nite
National Institute of Technology and Evaluation

3,967 classifications

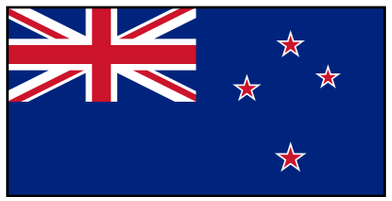
Public regulatory GHS data sources

safe work australia

5,577 classifications

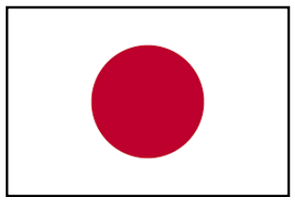
ECHA
EUROPEAN CHEMICALS AGENCY

141,475 classifications



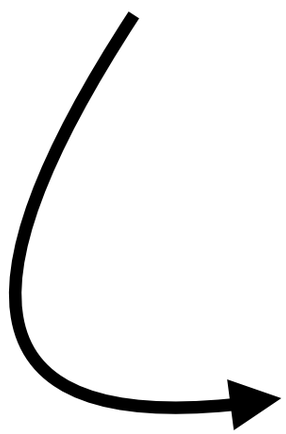
Environmental Protection Authority
Te Mana Rauhī Taiao

5,450 classifications



nite
National Institute of Technology and Evaluation

3,967 classifications



- “Harmonized” classifications (4,604)
- Industrial notifications (136,871)

Dataset collation

Dataset summary

- 3,336 compounds
- For each:
    - **qHTS assay data**: ToxCast/Tox21 assay data
    - **Chemical structures**: represented by physico-chemical descriptors from MOE and morgan fingerprints from RDKit
    - **Predicted targets**: “PIDGIN”-derived probabilities of ligand-protein bioactivity
    - **GHS acute toxicity classification**: Regulator-derived toxicity categories

Dataset:
qHTS assay data

- The US EPA has published data for the ToxCast chemical set for 821 assay endpoints freely available online – available as raw data and processed hit calls, AC_{50} values, etc.

- The dataset is designed with computational toxicology applications in mind.

- However, R. Thomas *et al.* (*Toxicol. Sci.*, **2012**, 128, 398–417) found that “the current ToxCast phase I assays and chemicals have limited applicability for predicting in vivo chemical hazards using standard statistical classification methods.”

USEPA. 2018. “ToxCast & Tox21 Summary Files” from invitrodb_v2
http://www2.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data (accessed Jan 2018)
Data released October 2015

Dataset:
Chemical structures

- 9011 substances in the ToxCast & Tox21 dataset:
    - Discarded those classed as “Mixture/Formulation”, “Polymer” or “Macromolecule”
    - Discarded those without a CAS number (required for looking up GHS data) or without a structure (where one could not be found from PubChem)
    - Standardised structures using ChemAxon’s Standardizer
- Resulting in 8539 structures
- From these, calculated 205 2D physico-chemical descriptors from CCG’s MOE, and 2048-bit (radius 2) Morgan fingerprints from RDKit

USEPA. 2018. “ToxCast & Tox21 Chemicals DSSTox Database” from DSSTox_20151019 http://www2.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data (accessed Jan 2018) Data released October 2015

Dataset:
Predicted targets

- In-house target prediction tool: “PIDGIN” (Prediction IncludiDinG INactivity)

- Random Forests with 100 trees, outputs either:
  - Platt-scaled probabilities of activity, *or*
  - Binary activity predictions for a given recall threshold

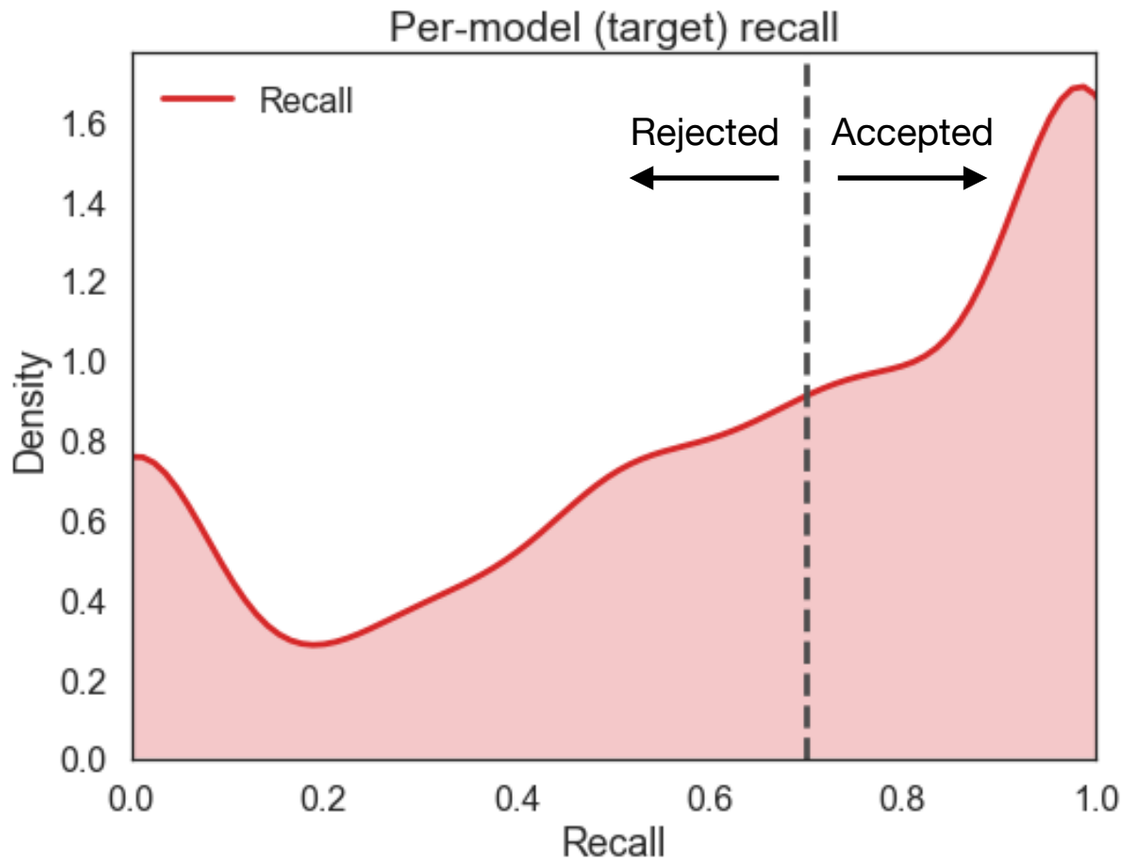
- Models for 3394 targets, trained on:
  - 19,918,879 bioactivities
    (extracted from PubChem and ChEMBL21)
  - 2,087,404 actives
  - 11,829,475 inactives

PIDGINv2
Available for download and use via GitHub – DOI:10.5281/zenodo.15984

Protein target model selection *via* performance estimation

PIDGIN comprises 3394 <u>independent</u> bioactivity models – each with their applicability domain.

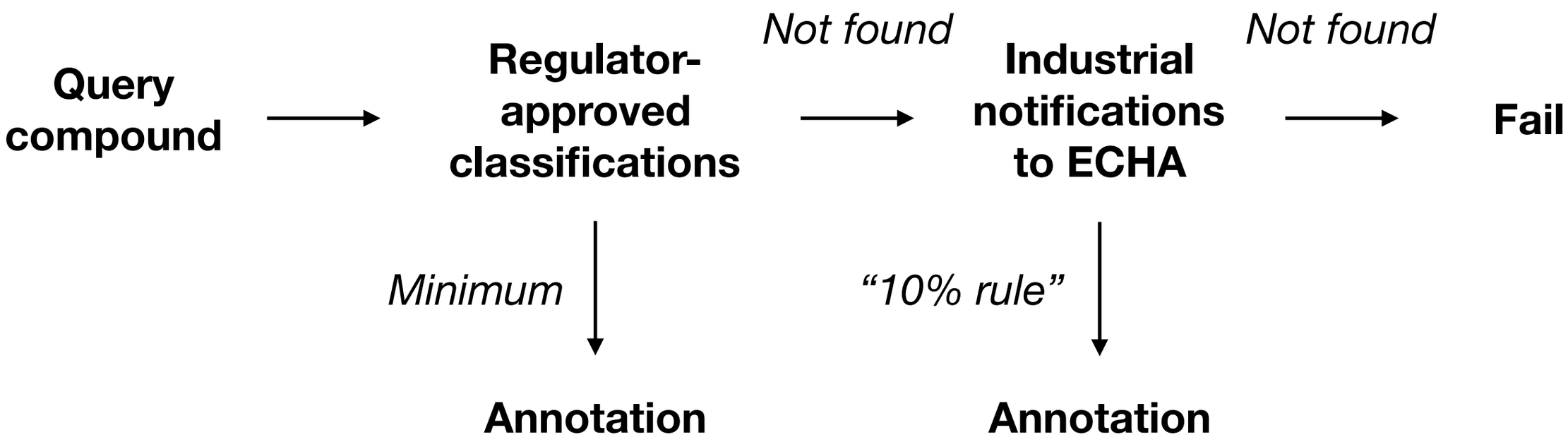
We can use input structures which also happen to be in the models' training sets to estimate each model's performance, rejecting all except the 800 that attain the required recall (here, 0.7)



In addition, we incorporate **known bio-activities** from PIDGIN's training data by setting the probability of interaction for these ligand-target pairs to be 1.

Dataset:

GHS acute toxicity classification



10% rule: Require for presence of a classification in >10% of industrial submissions before annotating a compound. This is the standard used by ECHA themselves for issuing a warning on their website.

Regulator-derived toxicity:
acute oral toxicity coverage

- 8539 structures from original ToxCast dataset, of which…
    - 3336 (39%) were classified, of which…
        - 920 (28%) in categories 1-3, labelled “toxic”
        - 1052 (32%) in category 4, labelled “harmful”
        - 245 (7%) in category 5, no label
        - 1119 (34%) were *implied nontoxic*

Implied nontoxicity

- There is no GHS acute toxicity category representing non-toxic – the least severe category has a lower limit
- However, regulatory classification and industrial submissions are expected to be “complete”
- Therefore, absence of an acute toxicity category in a set of purportedly complete GHS classifications should imply non toxicity

Chapter 3.1: Acute Toxicity
In *Globally Harmonized System of Classification and Labelling of Chemicals (GHS)* [Online], 7th revised ed. https://www.unece.org/trans/danger/publi/ghs/ghs_rev07/07files_e.html (accessed April 2018)

Dataset analysis

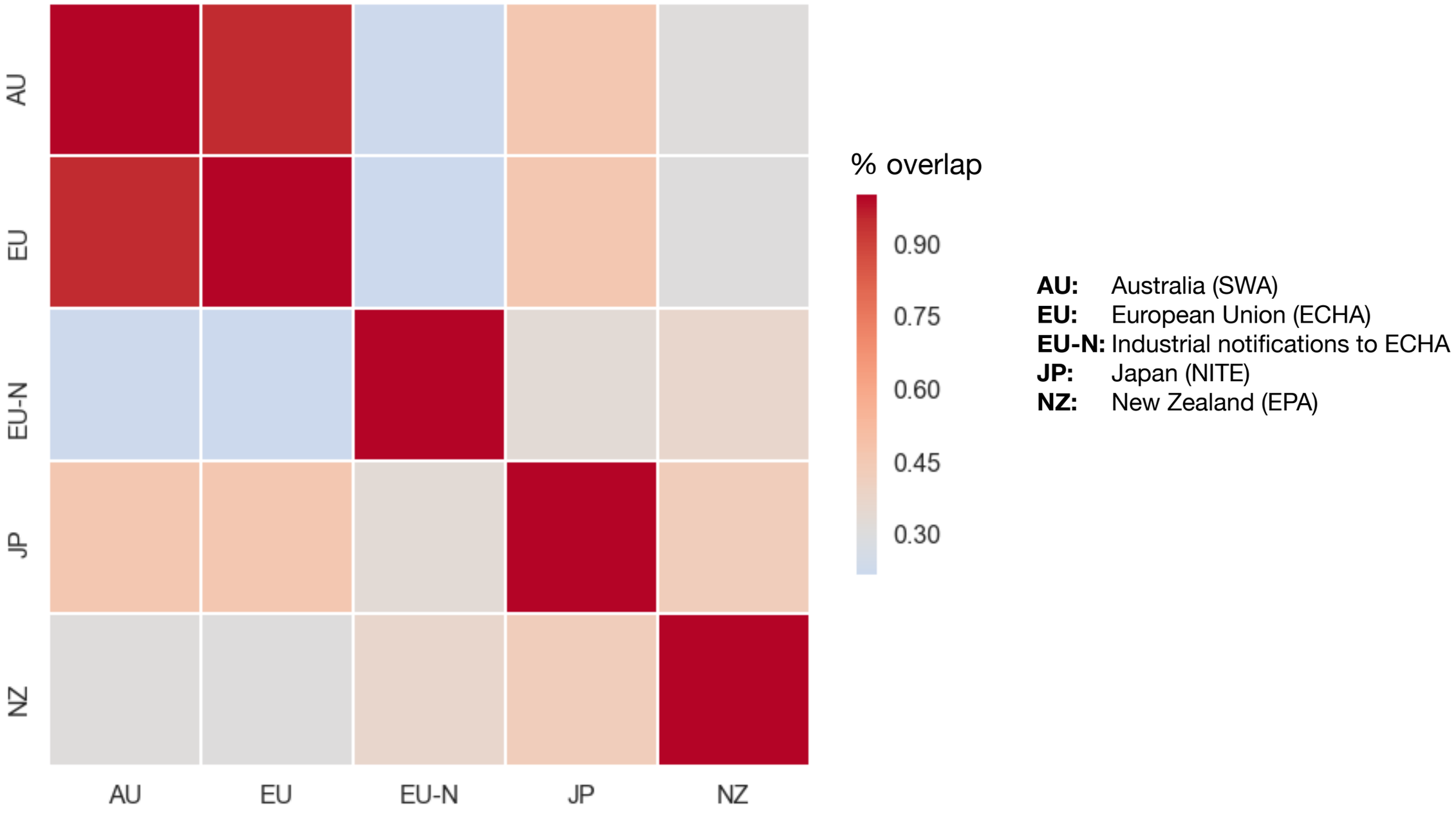
“Toxic” / “nontoxic” binning

- There are five GHS acute toxicity categories:

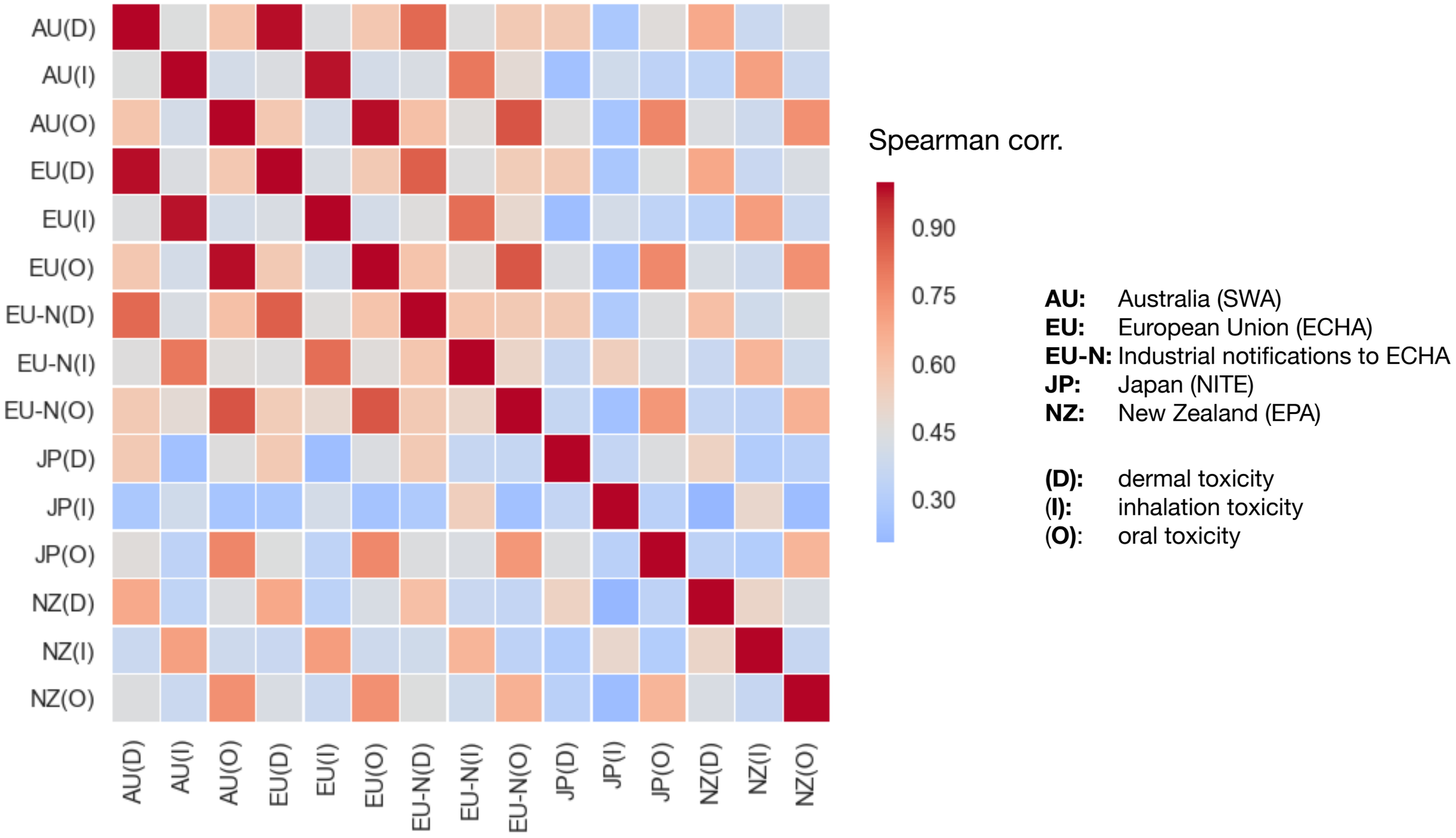
1-3:	($LD_{50} \leq 300$ mg/kg)	labelled “Toxic"
4:	(300 mg/kg < $LD_{50} \leq 2000$ mg/kg)	labelled “Harmful”
5:	(2000 mg/kg < $LD_{50} \leq 5000$ mg/kg)	not labelled

- When binary “toxic”/“nontoxic” classifications are utilised,
  - **“toxic”** includes to cat. 1-3
  - **“nontoxic**” includes cat. 5 and implied non-toxic (cat. 4 compounds are disregarded)

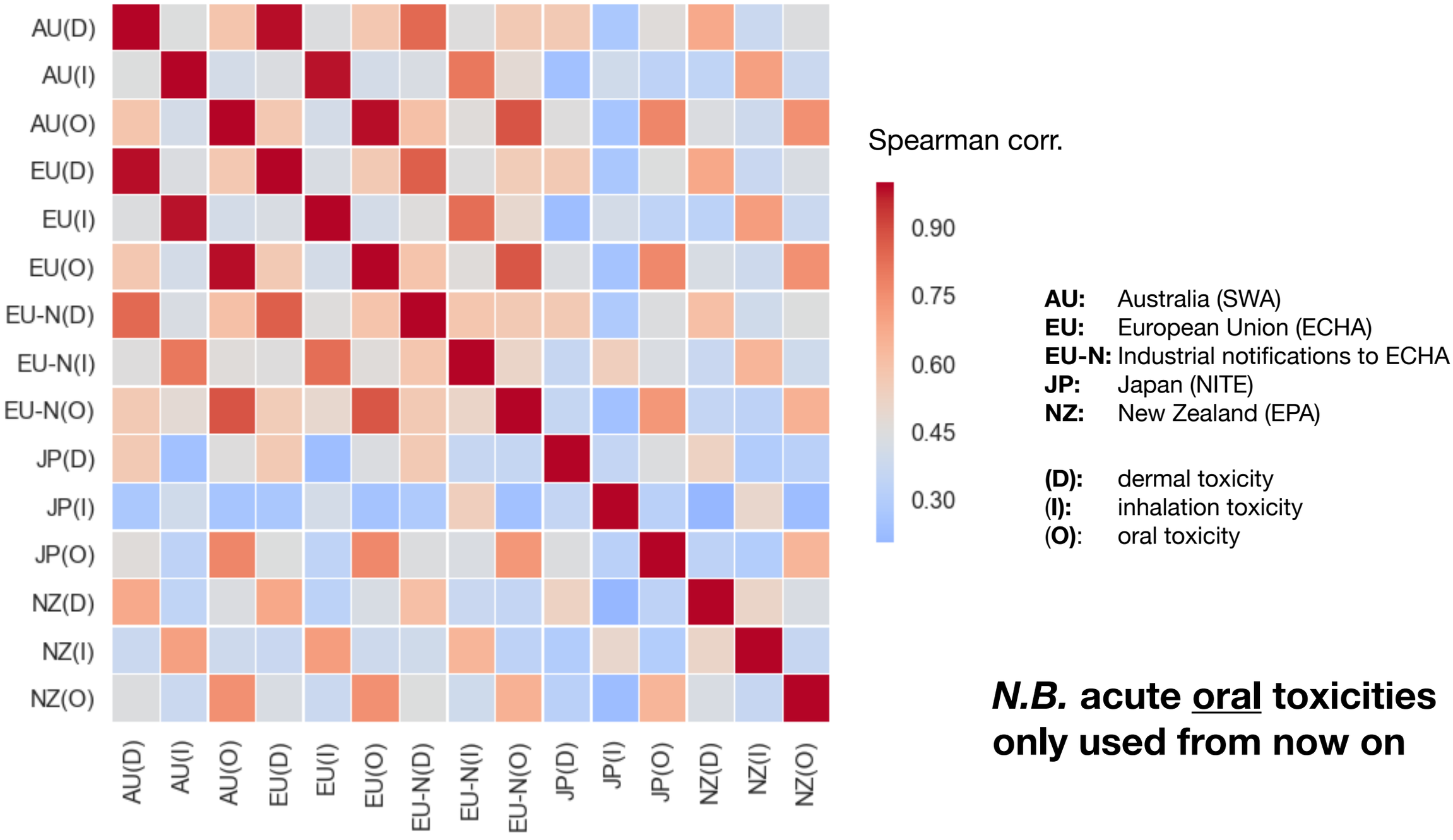
Most GHS data sources overlap by less than 60%



GHS classifications correlate with one another across different sources

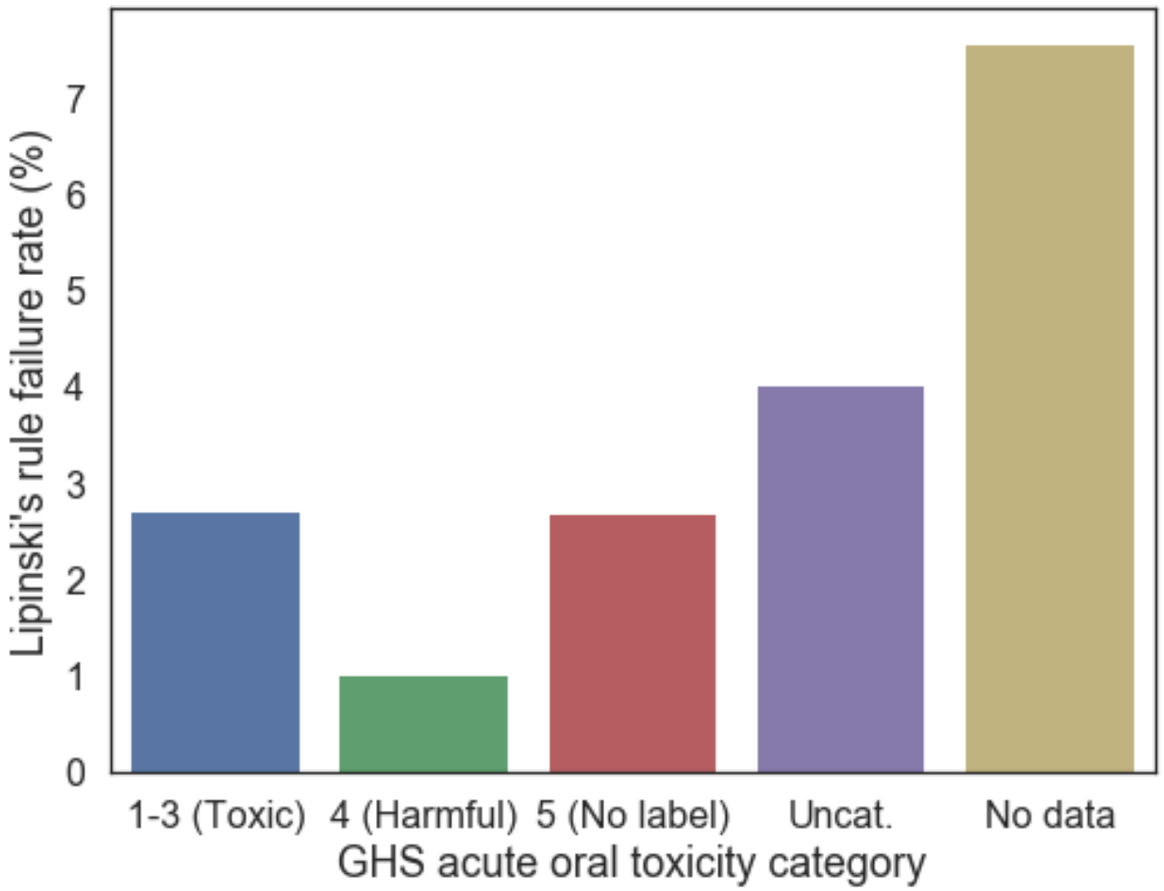


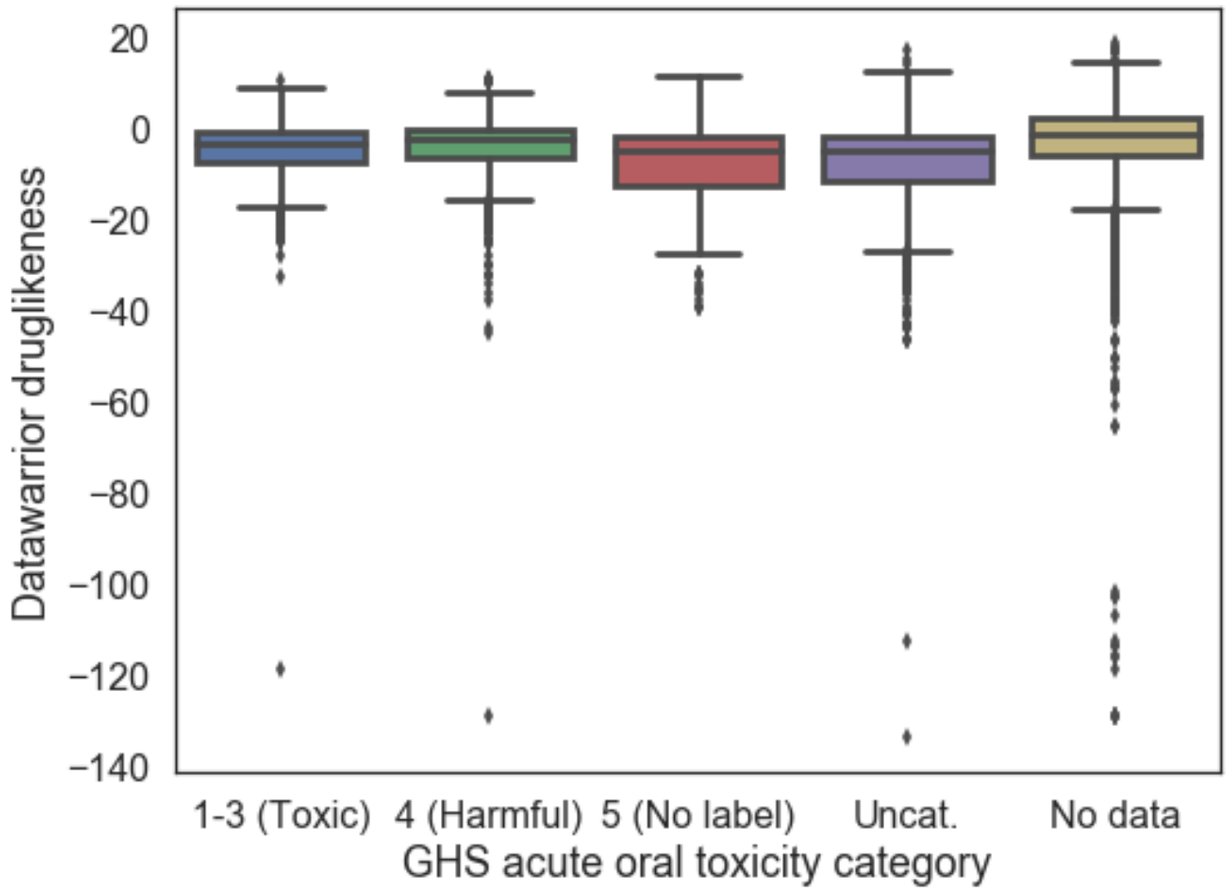
GHS classifications correlate with one another across different sources



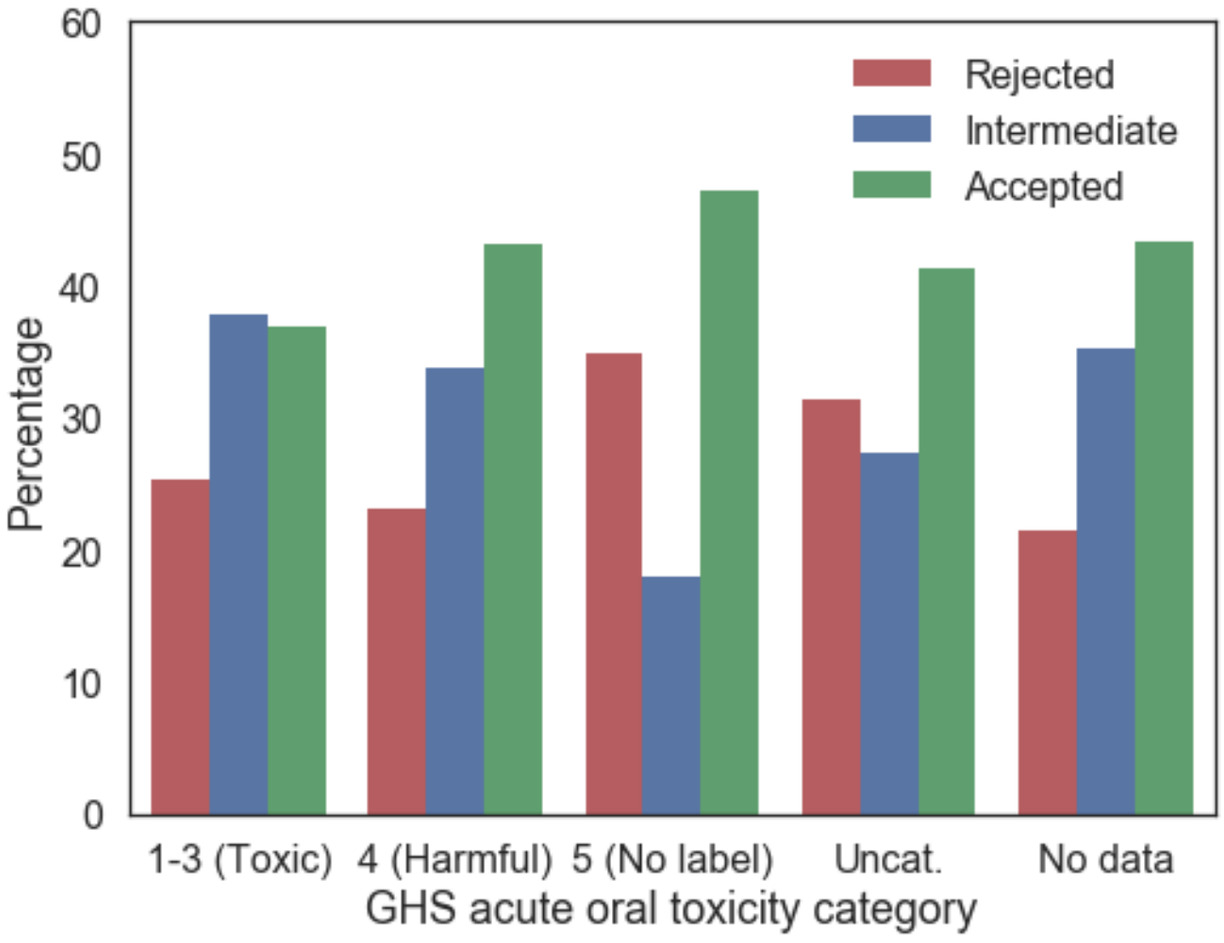
N.B. **acute <u>oral</u> toxicities only used from now on**

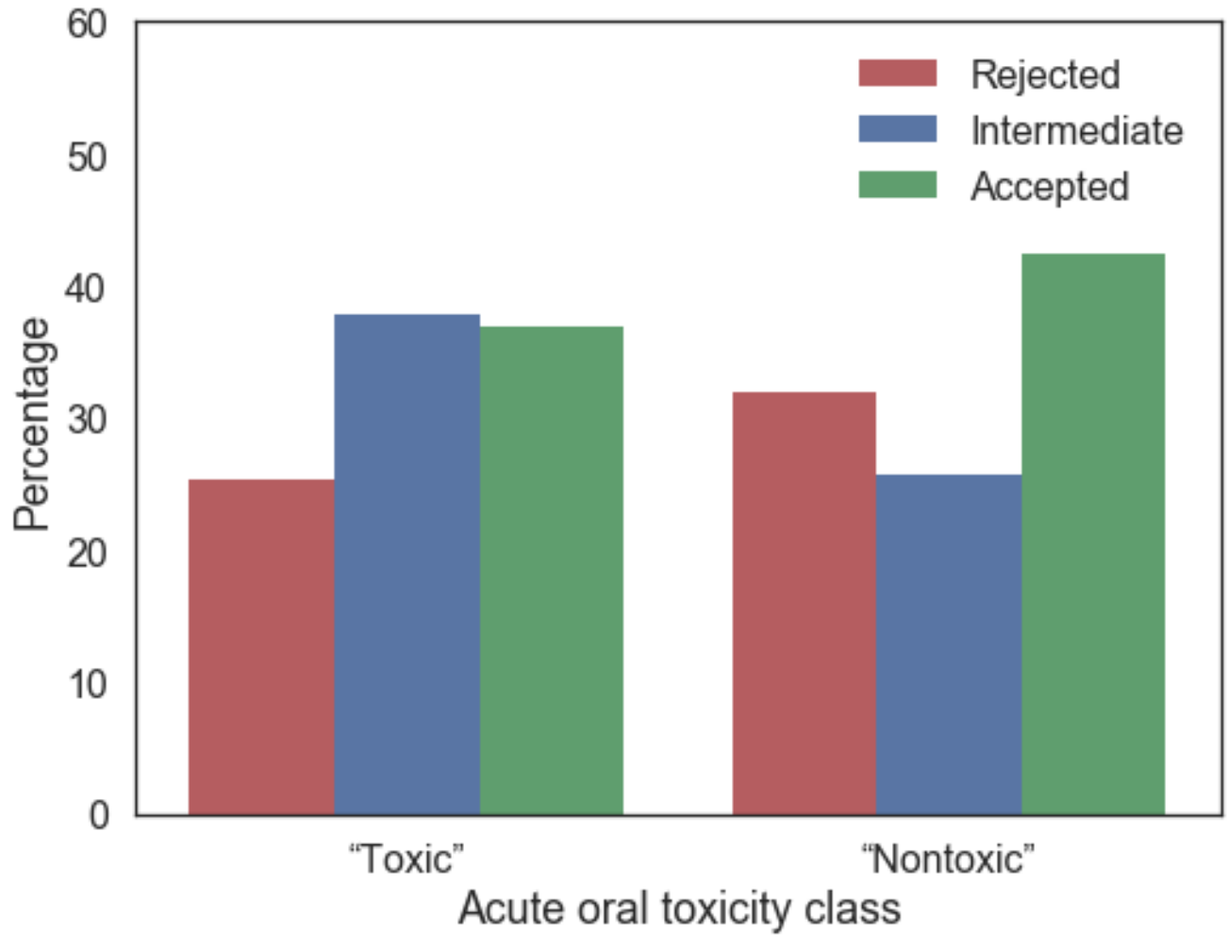
Bioavailability and druglikeness are largely independent of GHS category





FAFDrugs4 toxicophore screening does not catch GHS acute oral toxicities

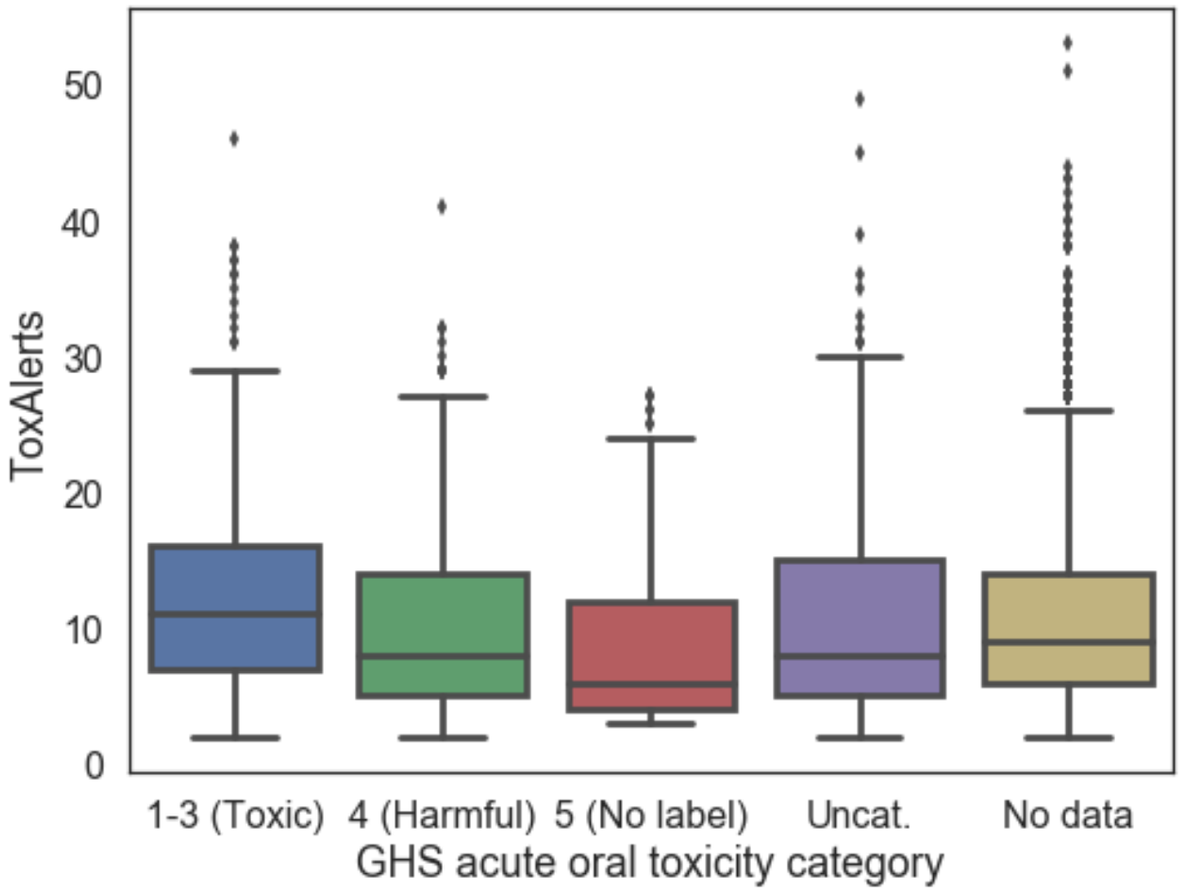




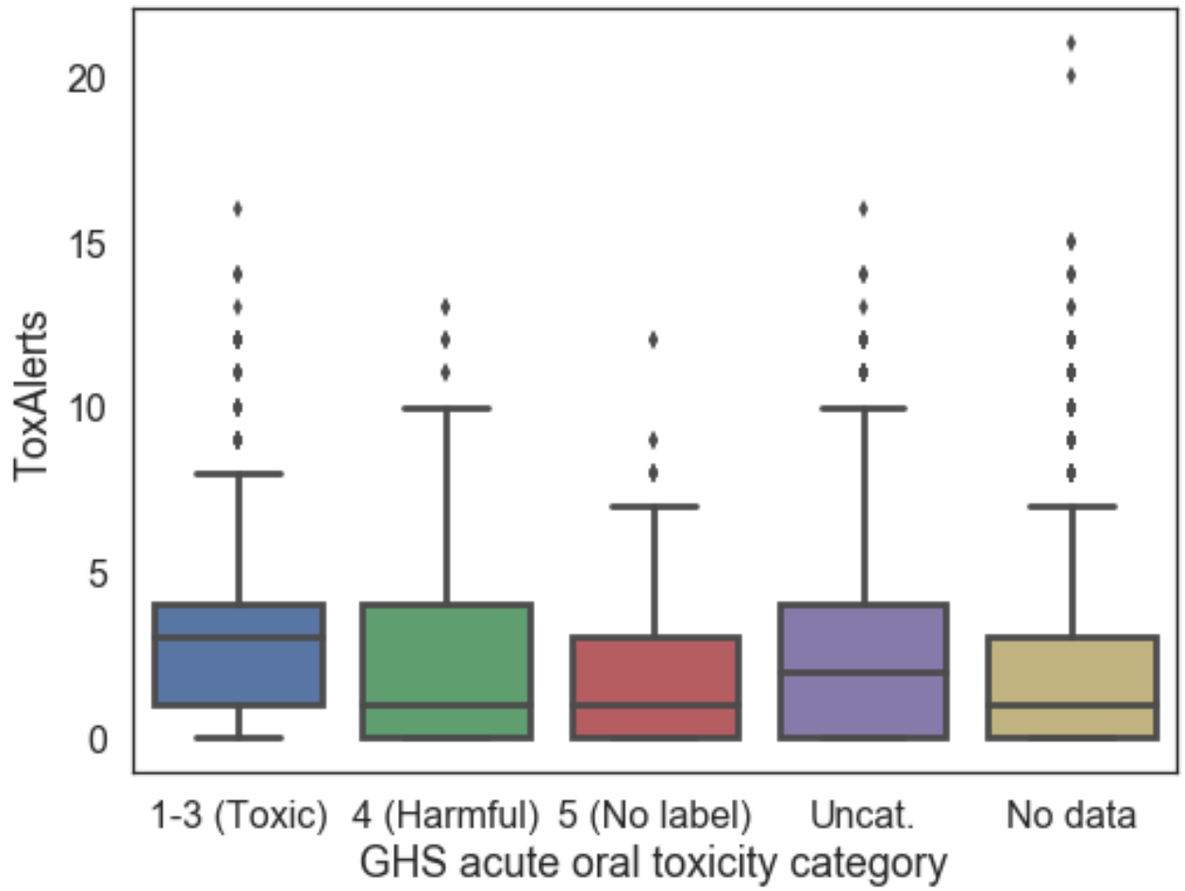
Lagorce et al. *Bioinformatics,* **2017**, 33, 3658-60

Number of ToxAlerts is also not a relevant screening metric

All toxicophore alerts

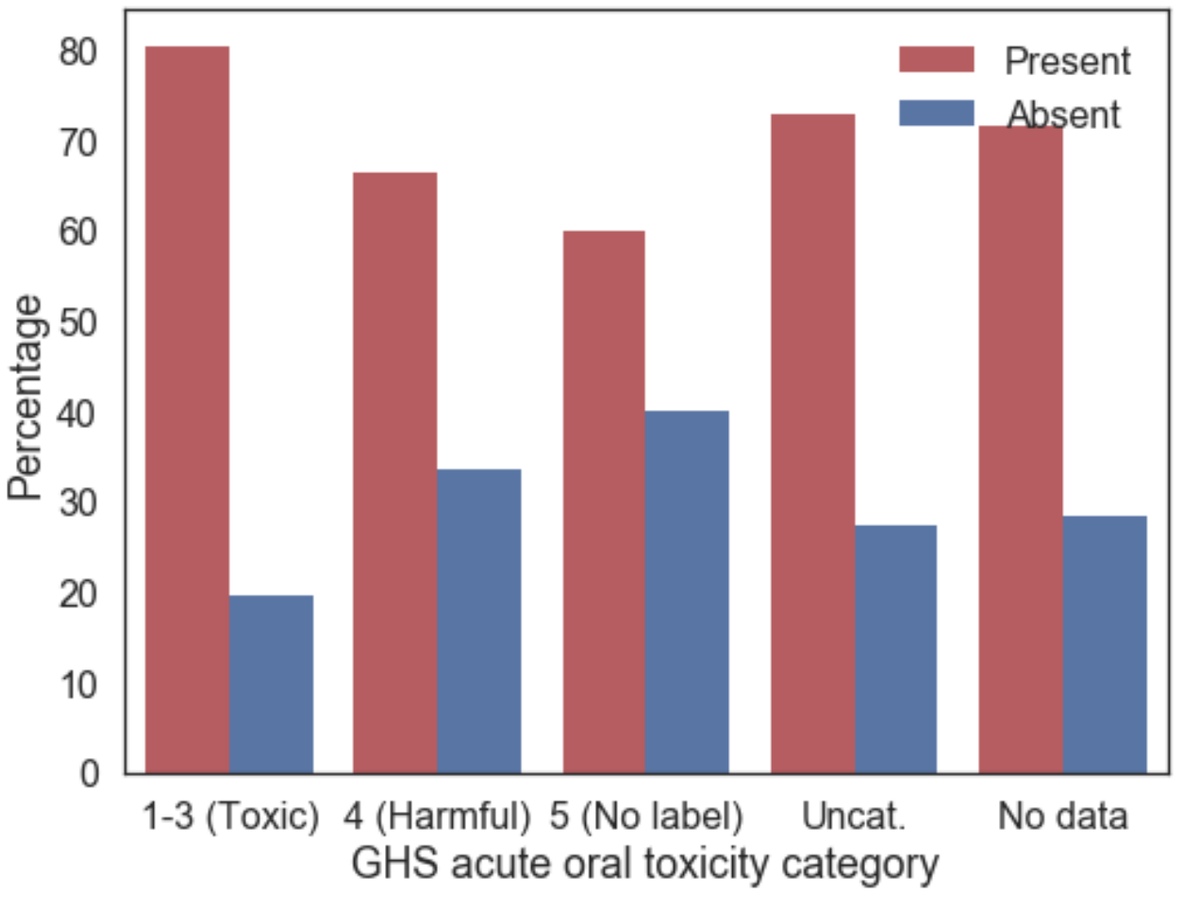


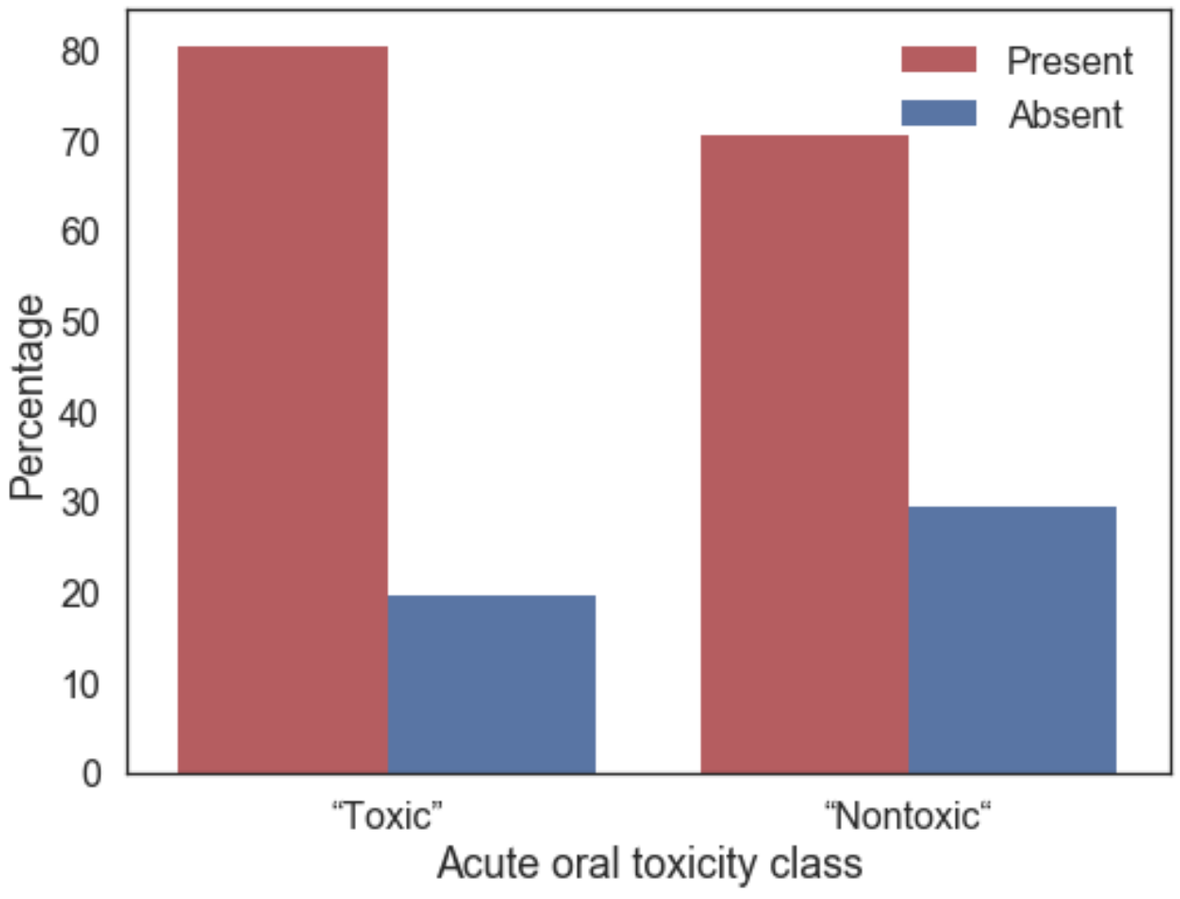
“Reactive, unstable, toxic” alerts



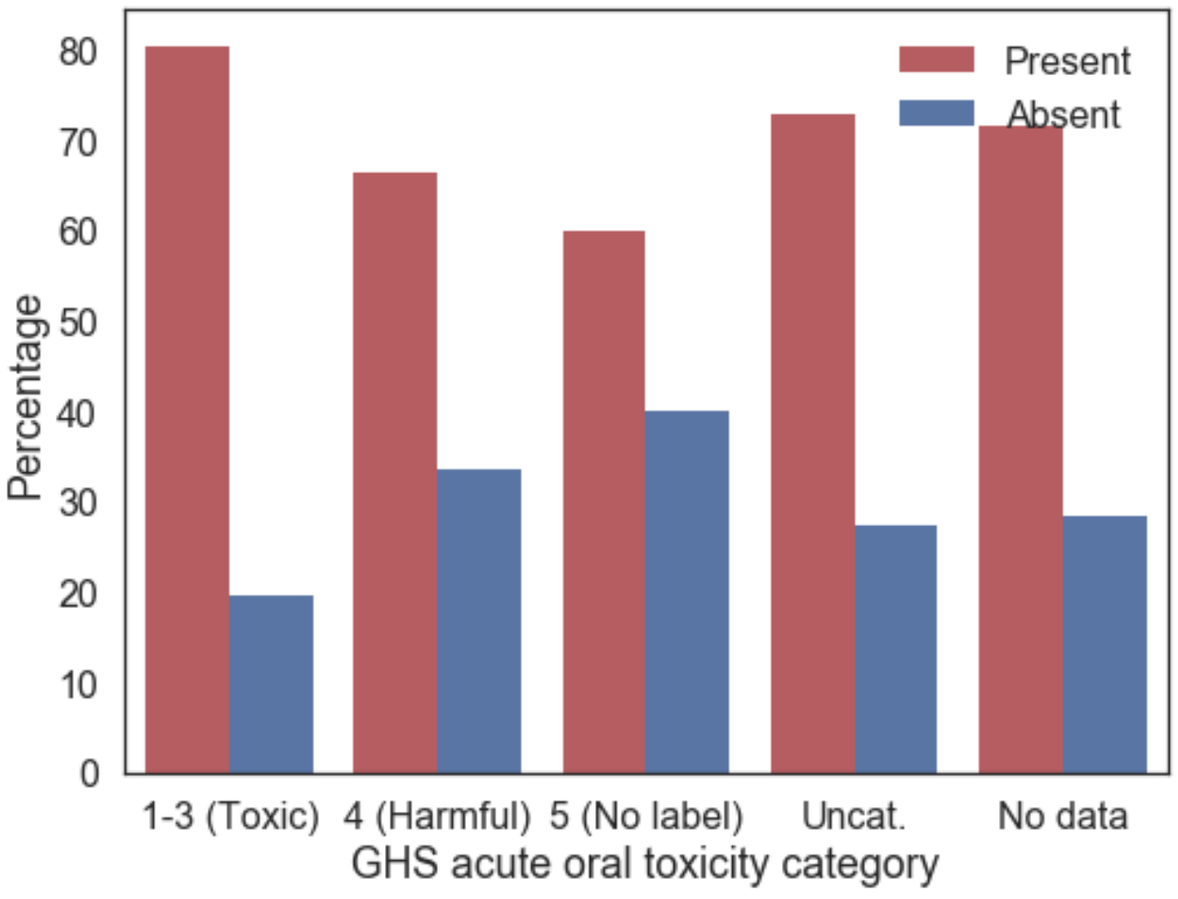
Sushko et al., *J. Chem. Inf. Model.*, **2012**, 52, 2310-6

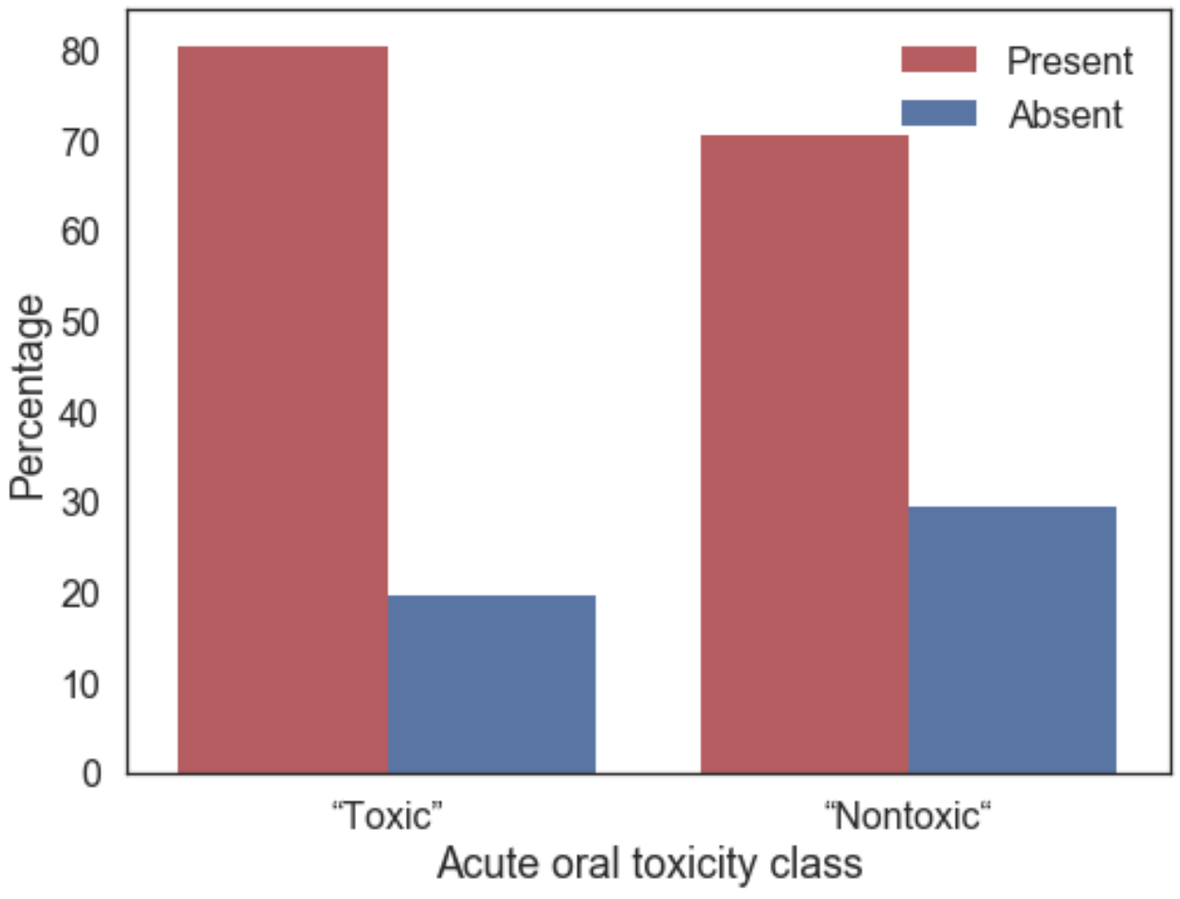
Presence/absence of any “reactive, unstable, toxic” ToxAlerts shows weak relationship with GHS acute oral toxicities





Presence/absence of any “reactive, unstable, toxic” ToxAlerts shows weak relationship with GHS acute oral toxicities

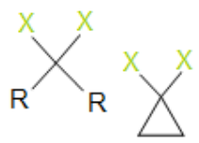
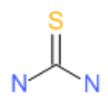
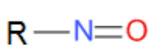
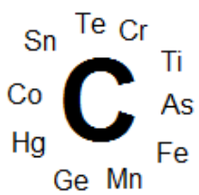




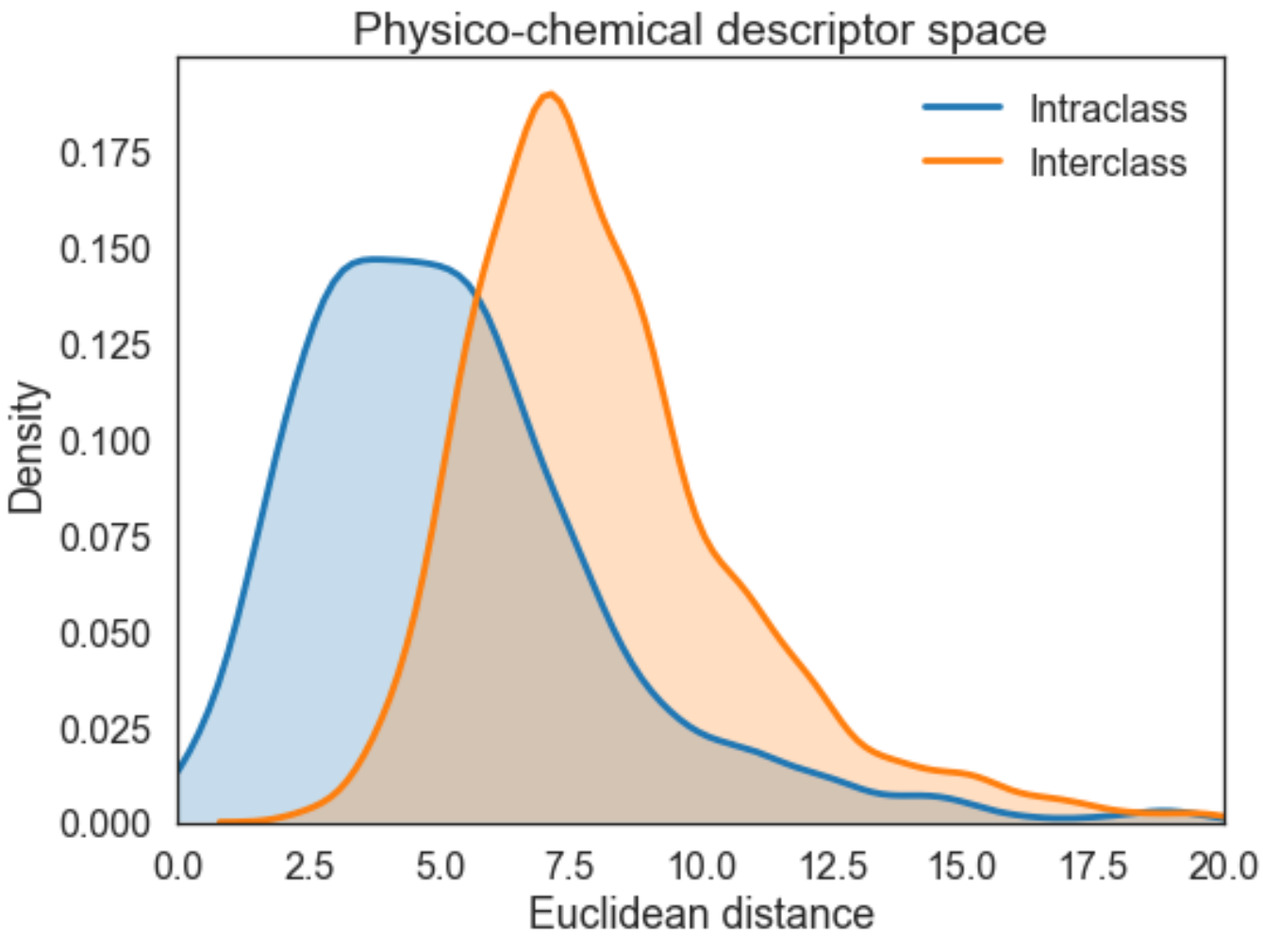
Odds ratio: 1.71

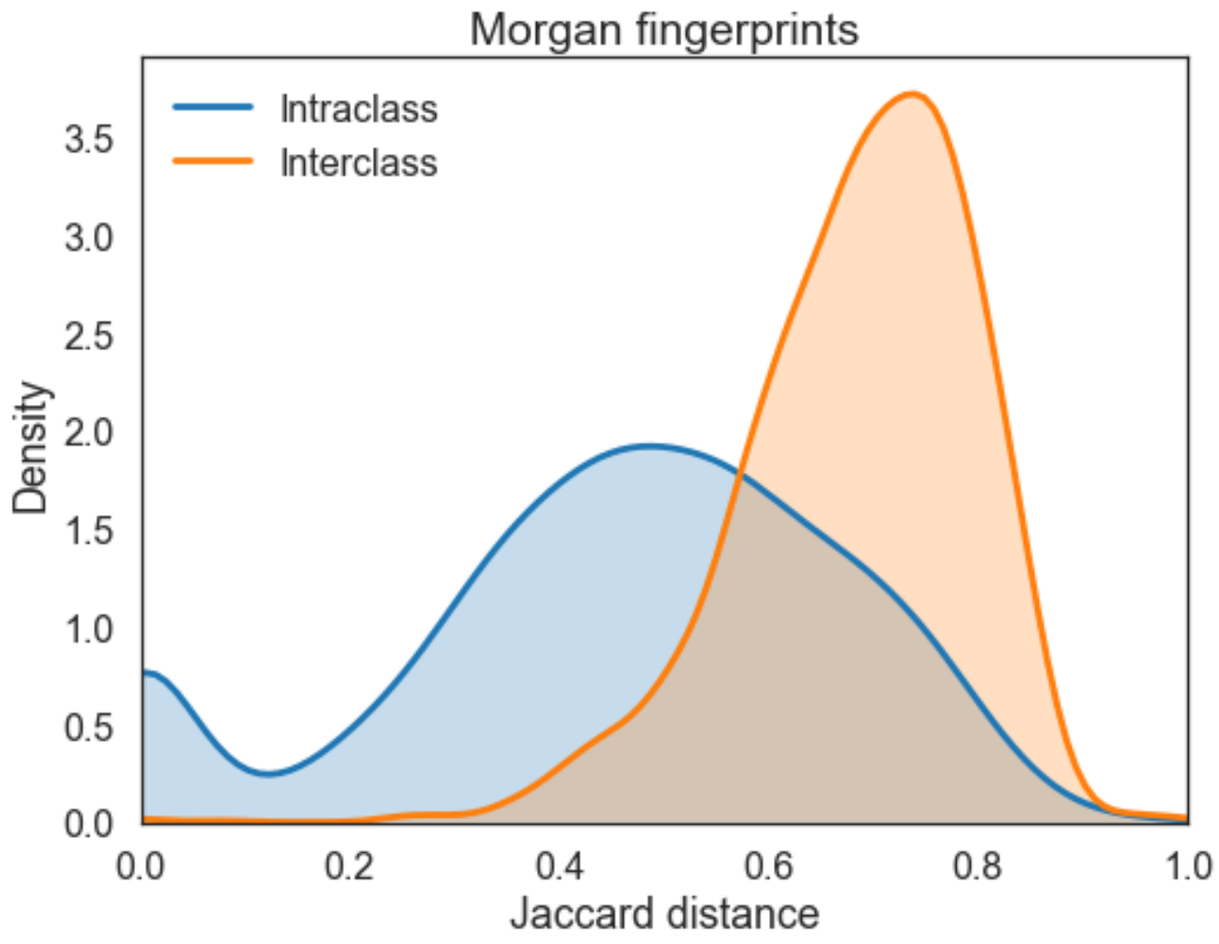
p value: 1.88×10^{-6}

Enrichment of "reactive, unstable, toxic" ToxAlerts (from nontoxic to toxic)

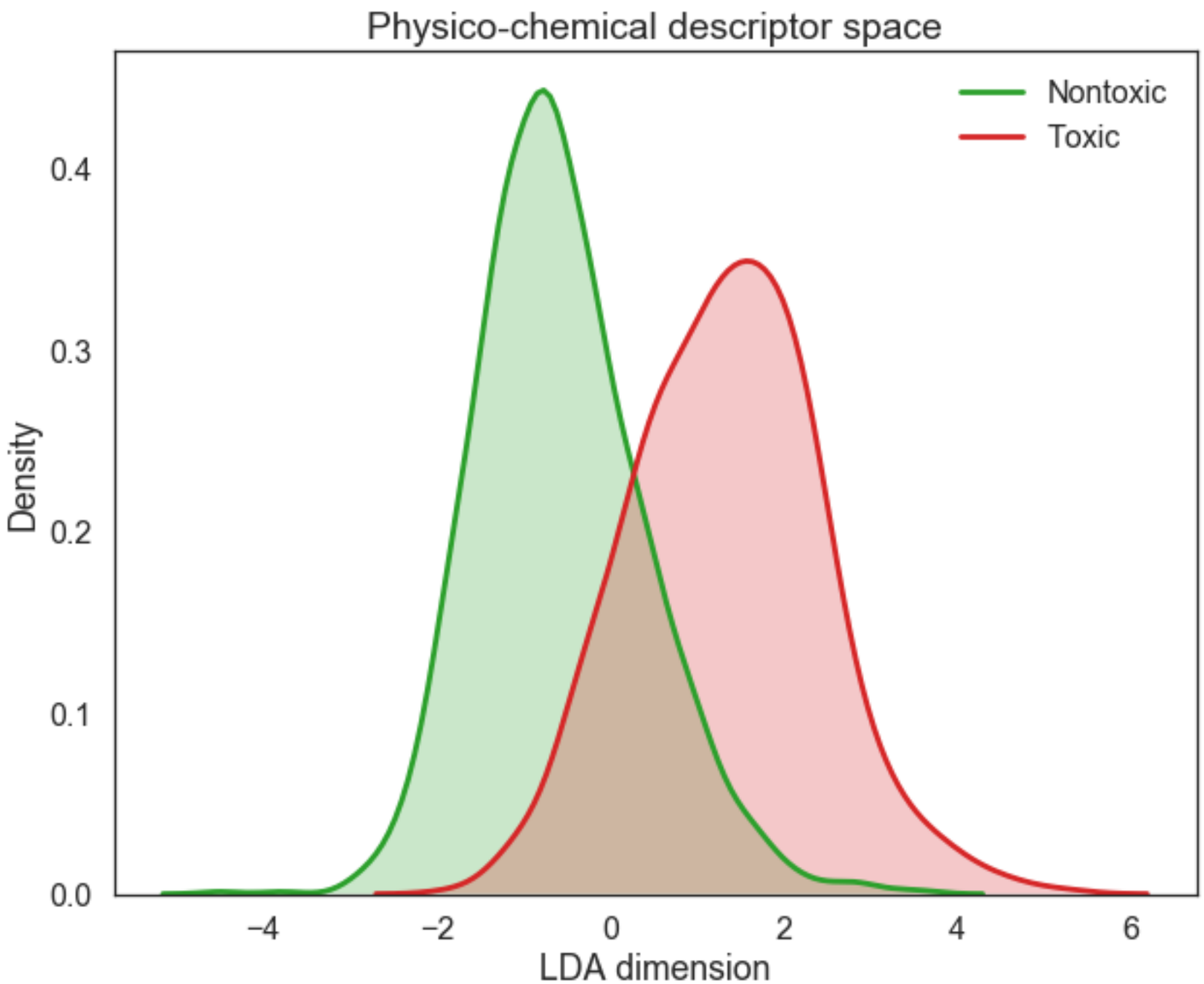
Structure	Alert ID	Name	Odds ratio	*p* value	Source
	TA1000	"double P=S and P=C bonds"	24.6	2.3×10^{-17}	"Filter to detect reactive, toxic and unstable compounds" – Enamine
	TA880	"gem-Dihalo propane and cyclopropane"	23.5	4.8×10^{-5}	"Toxic fragments in molecular structures" – Life Chemicals
	TA567, TA885, TA1075	"thioureas"	23.5	4.8×10^{-5}	"Reactive, unstable, and often toxic chemical groups" – ChemDiv etc.
	TA324, TA914, TA770, TA1087	"nitroso"	19.0	3.2×10^{-7}	"Exclusion criteria of the Maybridge Screening Collection database"
	TA574	"organometallic compounds"	7.6	1.6×10^{-5}	"Reactive, unstable, and often toxic chemical groups" – ChemDiv

Inter- and intra-class distances in chemical space

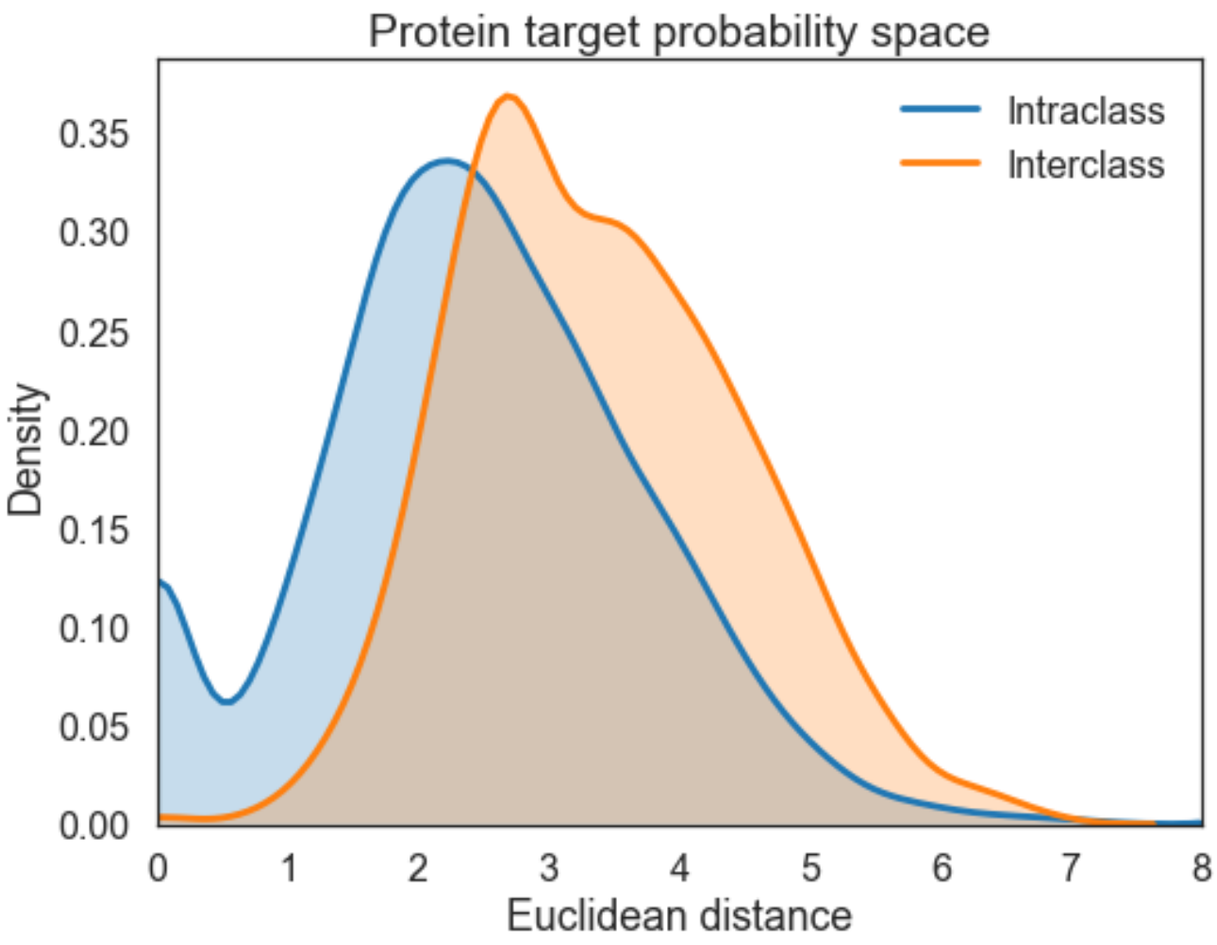


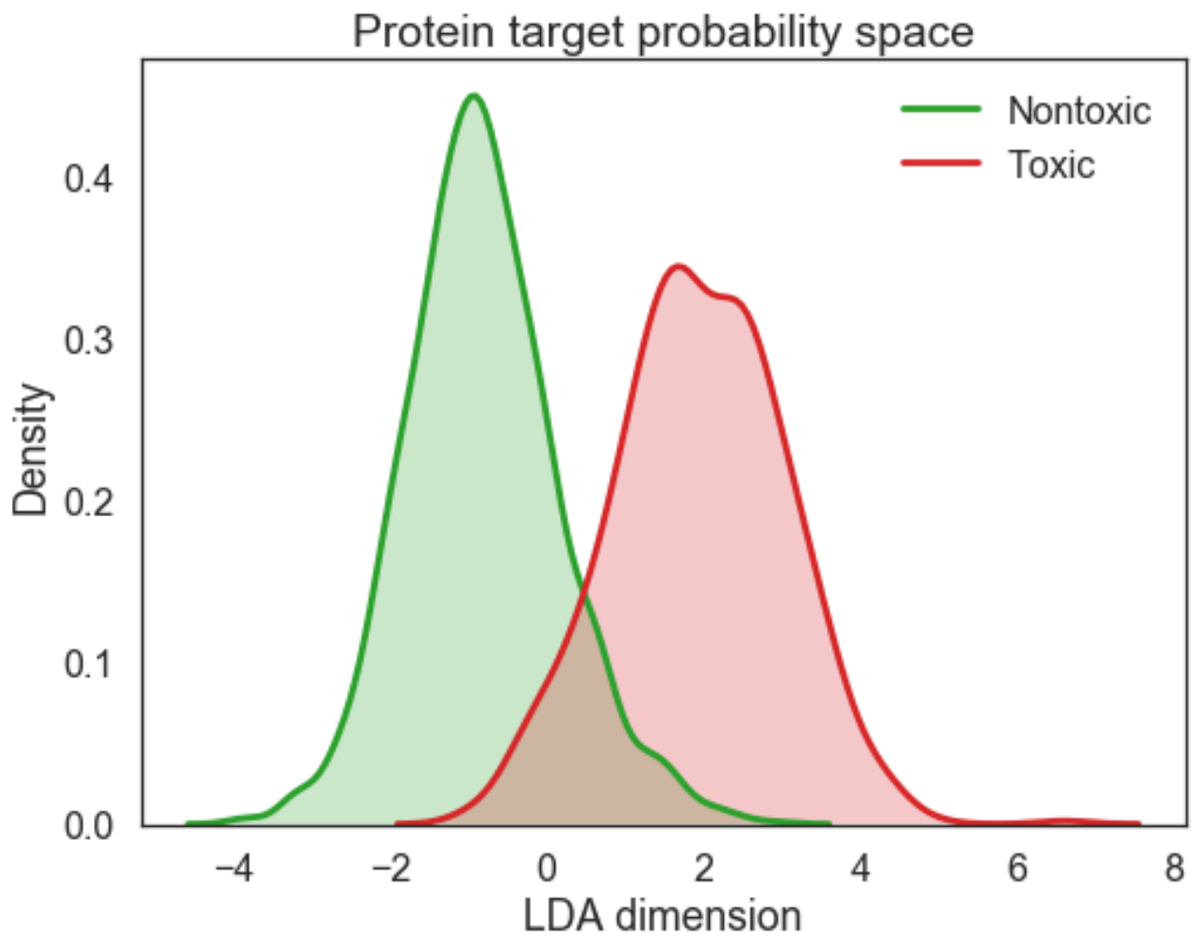


Linear Discriminant Analysis can partially separate classes in physico-chemical space



Inter- and intra-class distances in protein target space





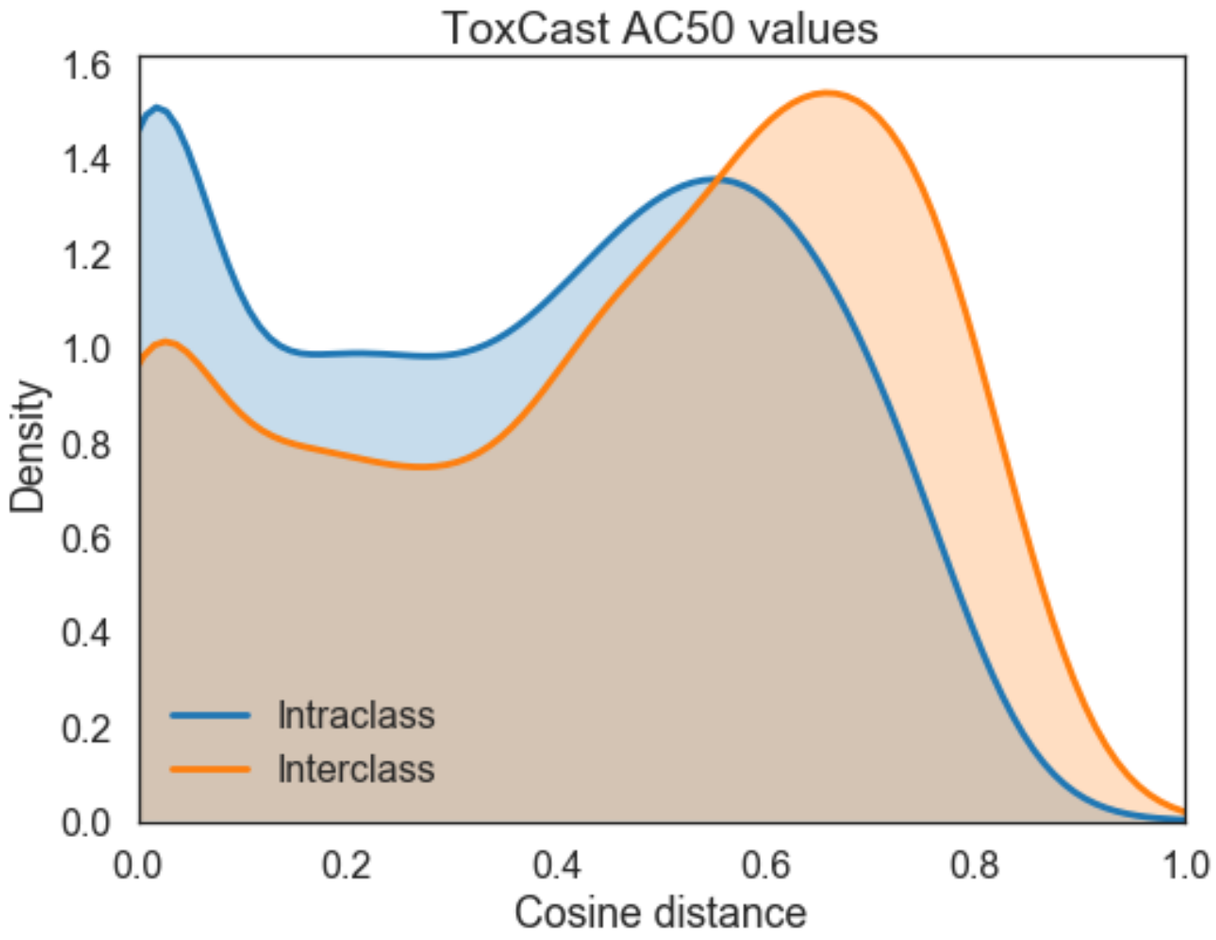
Enrichment of human target predictions (from nontoxic to toxic)

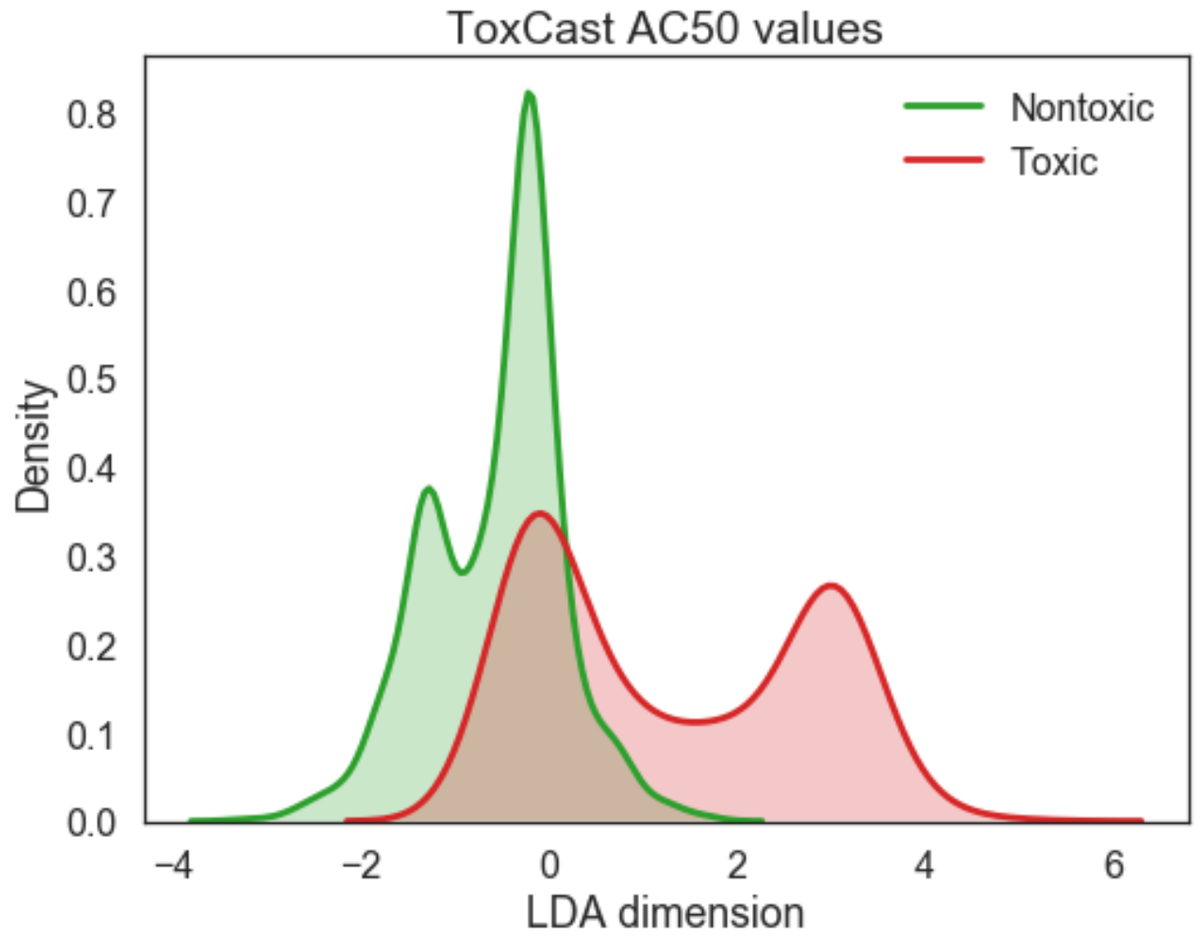
Uniprot	Protein	Odds ratio	*p* value	Associated pathologies
P07711	Cathepsin L1	7.39	5.10×10^{-7}	Heart disease, cardiomyopathy, inflammatory skin disease, *etc.*
Q16850	Lanosterol 14-alpha demethylase	3.44	1.44×10^{-5}	Metabolic disease, retinal dystrophy, *etc.*
Q16853	Membrane primary amine oxidase	2.50	1.06×10^{-6}	Vascular disease, cerebravascular disorder, stroke, *etc.*
P10696	Alkaline phosphatase, placental-like	2.35	1.14×10^{-16}	Lymphoma, neurodegenerative disease, acute myeloid leukemia, *etc.*
P35869	Aryl hydrocarbon receptor	2.22	1.80×10^{-17}	Cardiovascular disease, Chrohn's disease, lung disease, *etc.*

Highest-weighted human protein targets in Linear Discriminant Analysis projection

Uniprot	Protein	Weight (absolute)	Associated pathologies
P25789	Proteasome subunit alpha type-4	64.5	Amyloidosis, immune system disease, lymphoma, *etc.*
P28070	Proteasome subunit beta type-4	52.9	Amyloidosis, immune system disease, lymphoma, *etc.*
P32239	Gastrin/cholecystokinin type B receptor	40.6	Digestive system disease, peptic ulcer, *etc.*
P60900	Proteasome subunit alpha type-6	29.9	Amyloidosis, immune system disease, lymphoma, *etc.*
P35346	Somatostatin receptor type 5	29.3	Digestive system disease, Haemorrhage, diarrhoea, *etc.*

ToxCast AC_{50} values show little separation power



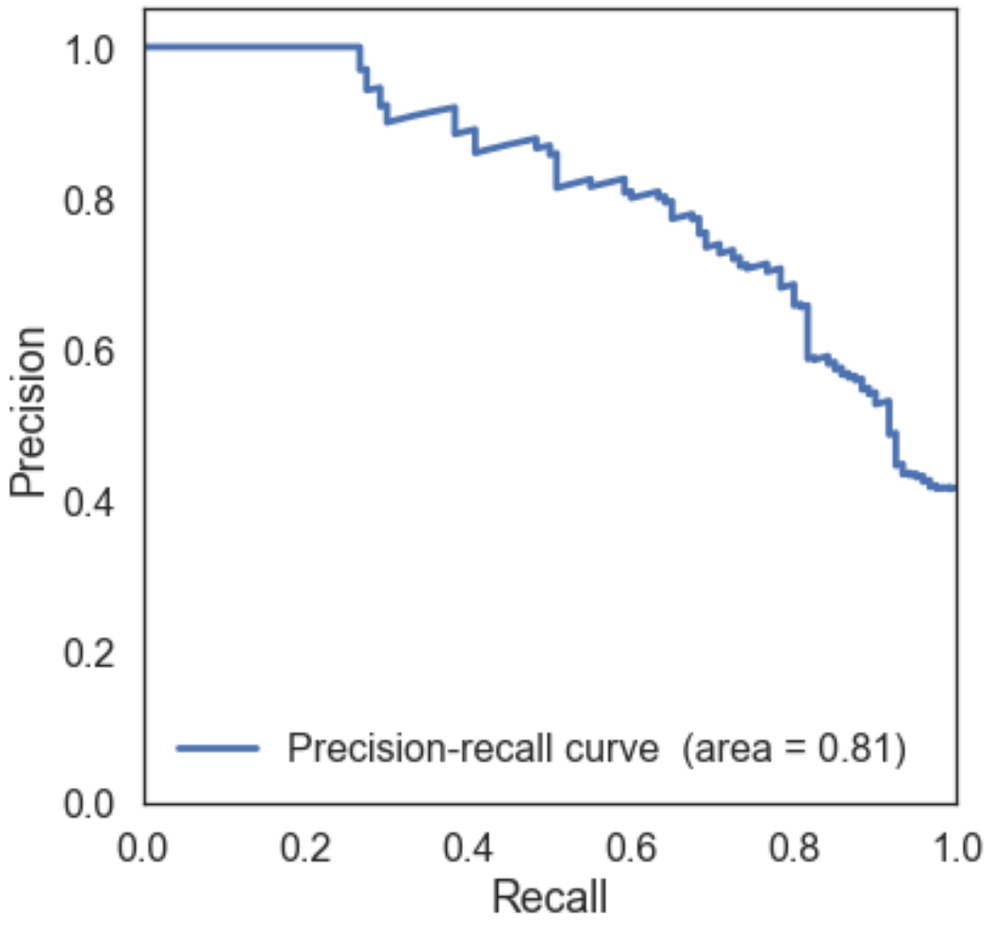


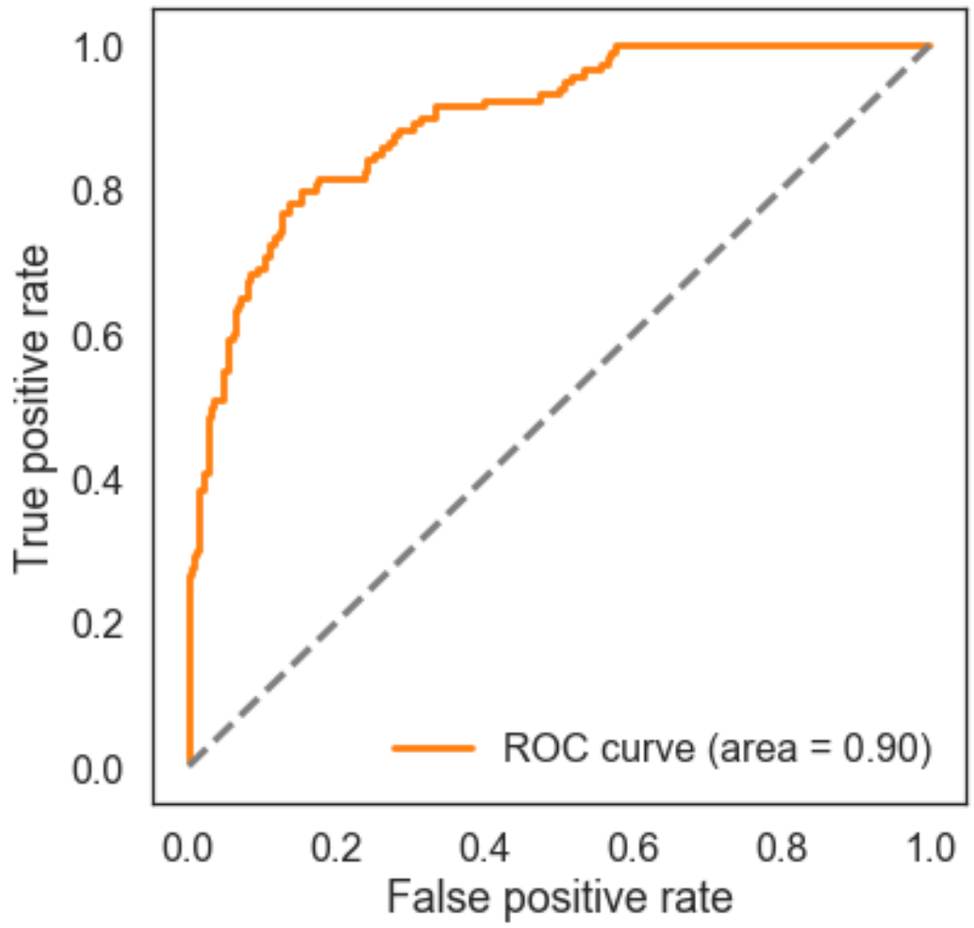
Modelling results

Modelling workflow

- 20% of data reserved as validation set, rest used as training set.
- On training set, for each descriptor set:
    - **Random Forest** hyperparameter optimisation (5-fold cross validation) –
        maximising area under the ROC curve
    - Threshold optimisation (5-fold cross validation) using selected hyper parameters –
        maximising correct classification rate (CCR)
    - Final Random Forest classification model fitted
- On validation set, for each model:
    - Precision-recall and ROC curves plotted
    - Sensitivity, specificity and CCR at optimum threshold calculated

Performance of physico-chemical Random Forests model

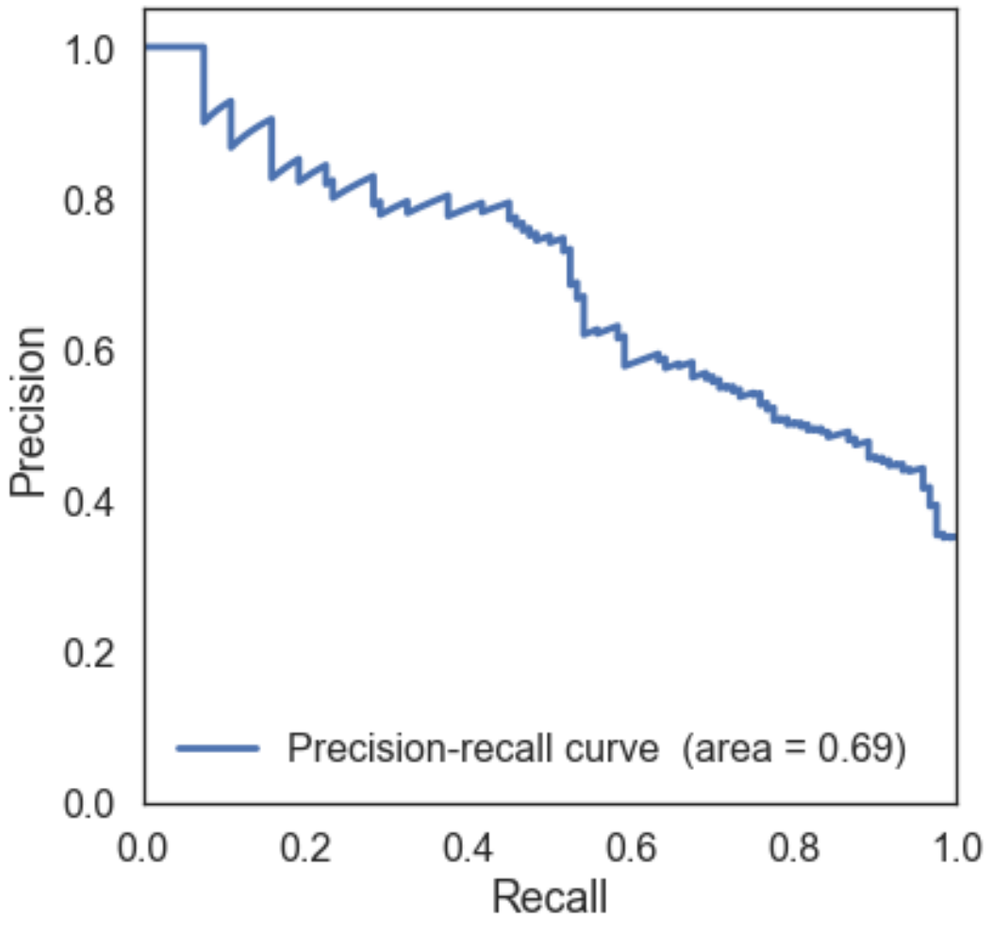


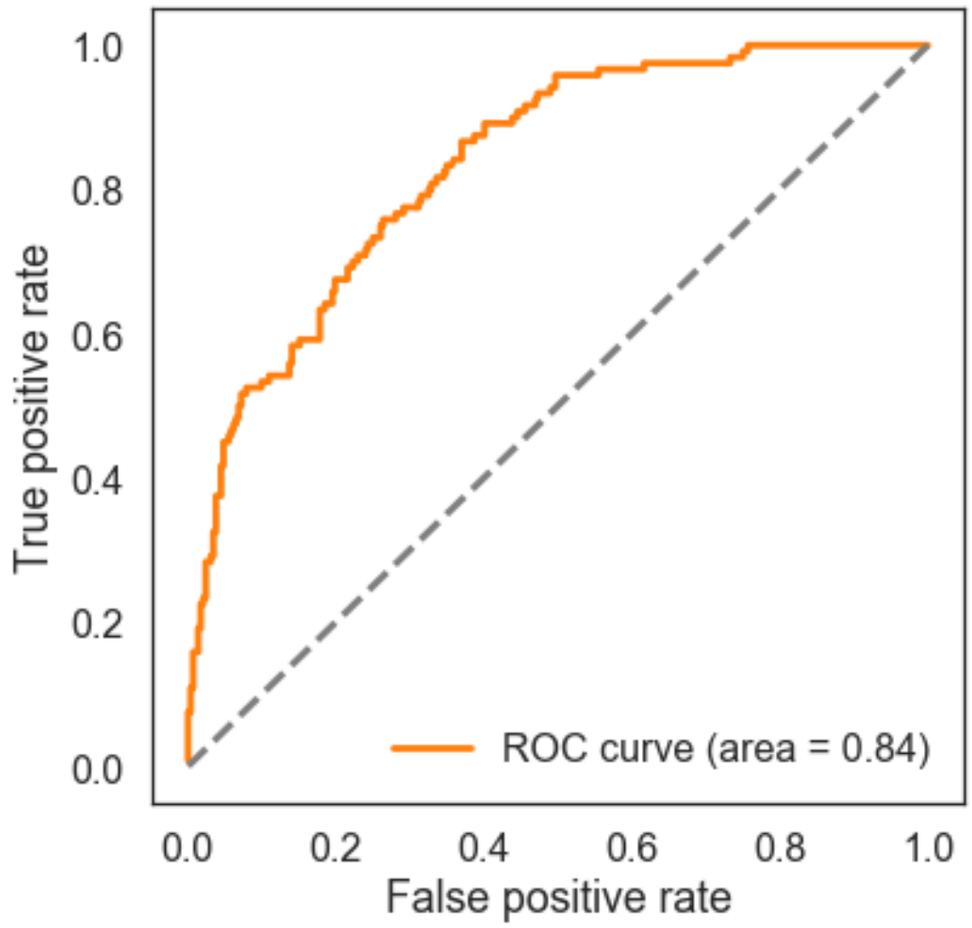


At (training-set determined) optimal threshold:

Sensitivity: 0.84 **Specificity:** 0.75 **Correct classification rate:** 0.80

Performance of protein target Random Forests model

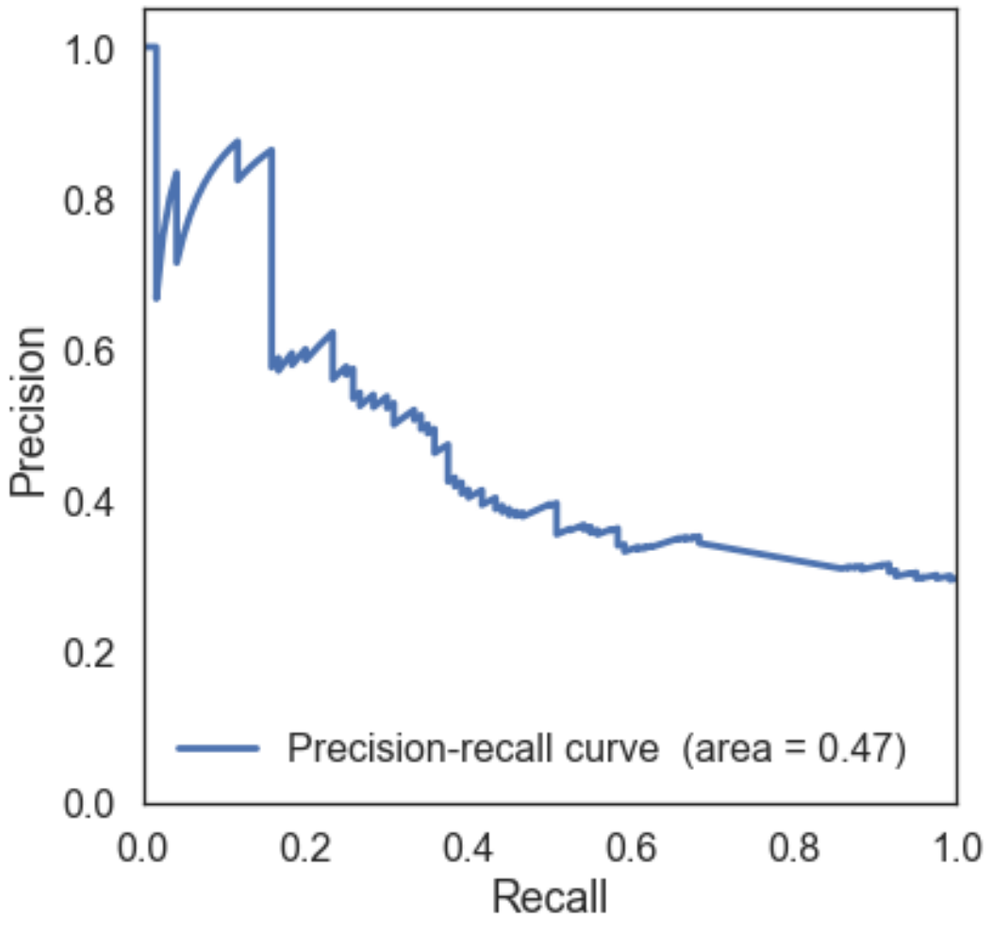


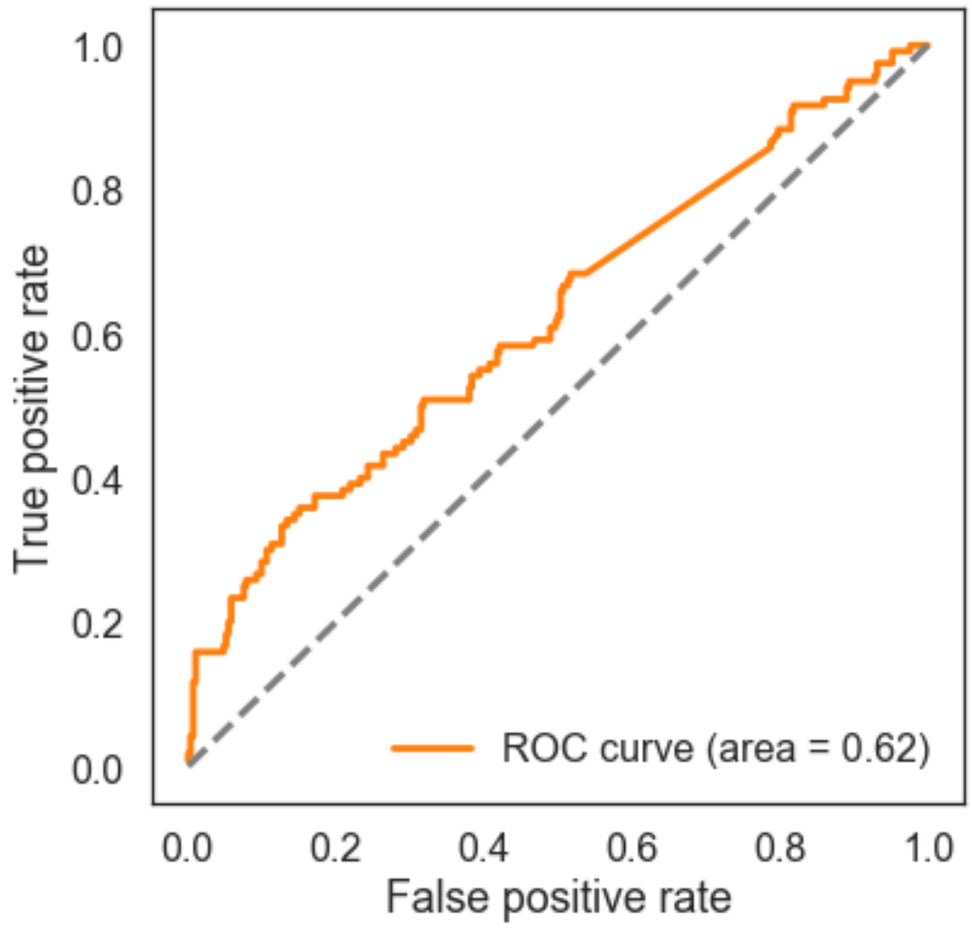


At (training-set determined) optimal threshold:

Sensitivity: 0.71 **Specificity:** 0.76 **Correct classification rate:** 0.74

Performance of ToxCast AC_{50} Random Forests model

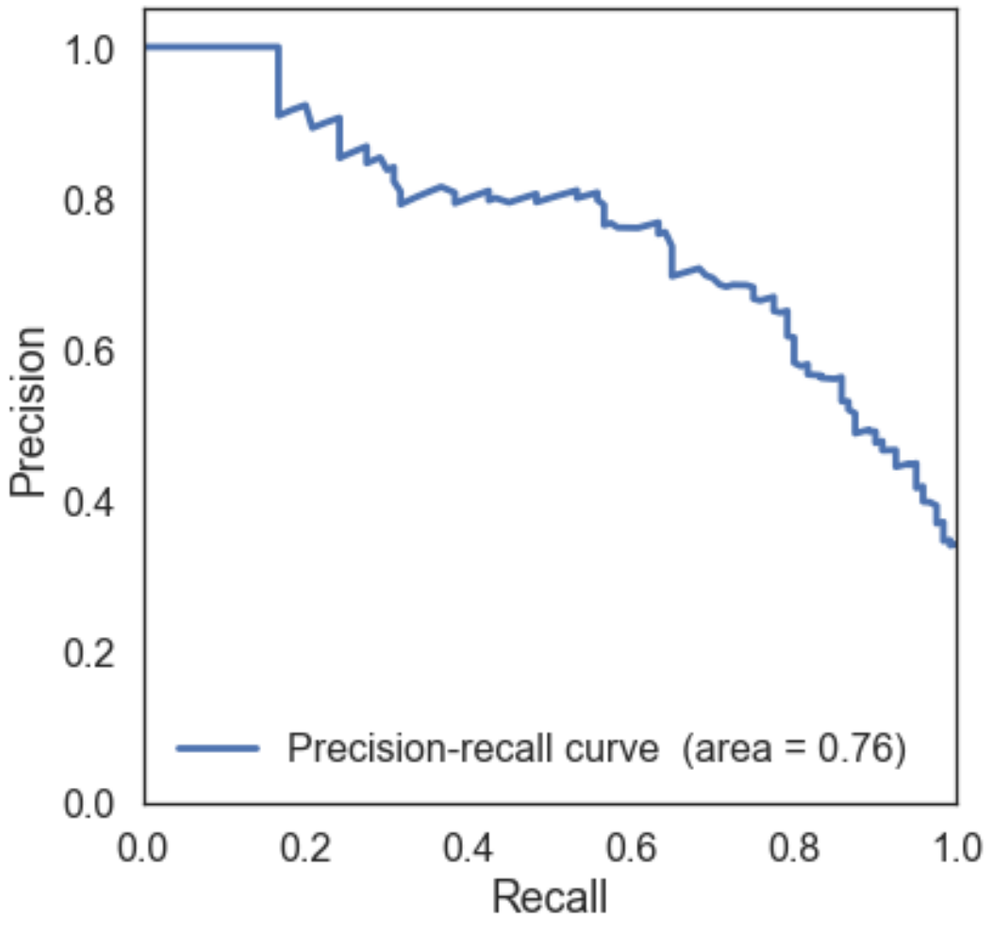


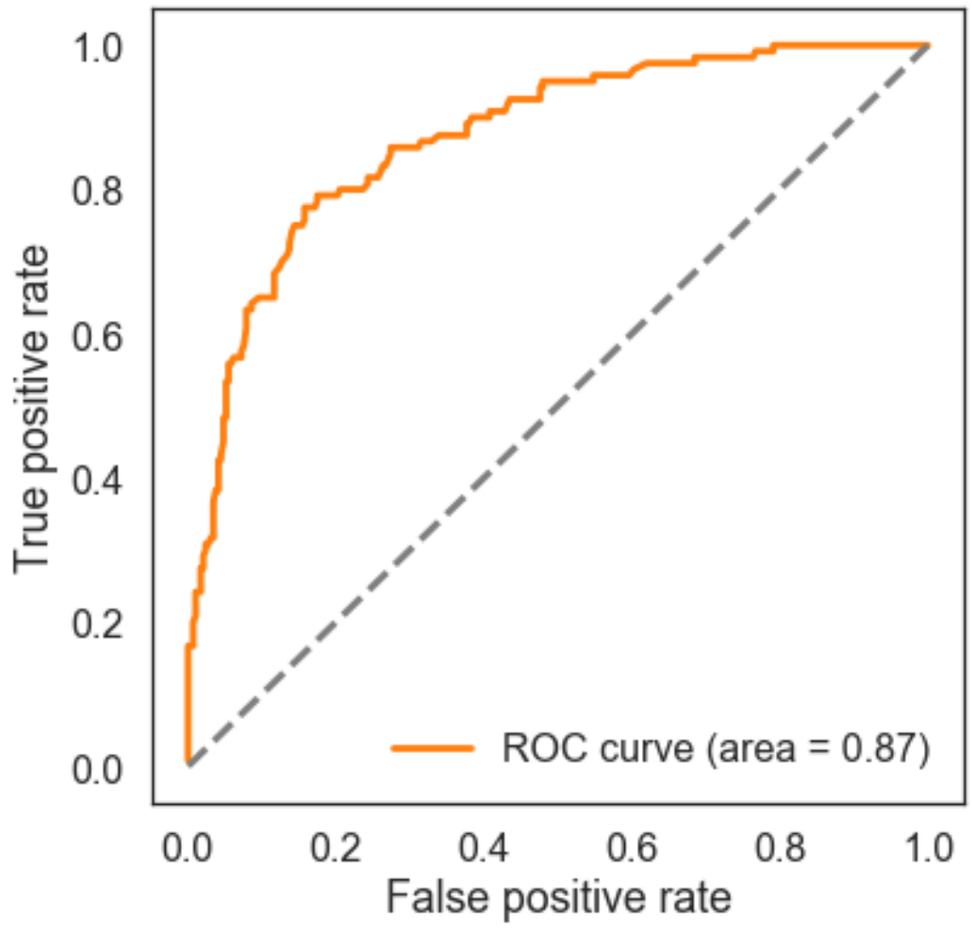


At (training-set determined) optimal threshold:

Sensitivity: 0.68 **Specificity:** 0.46 **Correct classification rate:** 0.57

Performance of physico-chemical *and* protein target Random Forest model





At (training-set determined) optimal threshold:

Sensitivity: 0.79 **Specificity:** 0.81 **Correct classification rate:** 0.80

Performance summary

	PR AUC	ROC AUC	Sensitivity	Selectivity	CCR
Physico-chemical	**0.81**	**0.90**	**0.84**	0.75	**0.80**
Protein target	0.69	0.84	0.71	0.76	0.74
ToxCast AC50	0.47	0.62	0.68	0.46	0.57
Combination physics-chemical and protein target	0.76	0.87	0.79	**0.81**	**0.80**

Model interpretation?

- MOE-generated physico-chemical descriptors can be challenging to interpret
- e.g. highest-importance features in physico-chemical model:

  “Relative negative partial charge”, “mean atom information content”, “number of nitrogen atoms”, “0th GCUT descriptor using atomic contribution to molar refractivity”, “molecular mass density” *etc.*

- In contrast, protein target features are intrinsically suggestive of a specific biological activity
- e.g. highest-importance protein target features in combination physico-chemical and protein target model:

  “5-hydroxytryptamine receptor 3B”, “Metabotropic glutamate receptor 1”, “P2Y purinoceptor 1”, “Prostaglandin E2 receptor EP2 subtype”, "Vitamin D3 receptor”, *etc.*

- Future work can be done using more sophisticated feature analyses to generate potential modes of action for individual predictions

Conclusions

Conclusions

- The GHS represents a rich source of acute dermal, inhalation and oral toxicity data for use in predictive toxicology studies
- Acute oral toxicity, as encoded by the GHS system, is not well screened by the *FAFDrugs4* toxicophore-based screen; however, ToxAlert's "reactive, unstable, toxic" toxicophore set does show a weak relationship with this toxicity
- Toxicity classes derived from GHS data are partially separable in chemical and protein-target space and may be predicted using Random Forests models
- Predictive models could be created using physico-chemical and protein target descriptors. However, qHTS data from the ToxCast/Tox21 assays could not be successfully employed in GHS class prediction.

Acknowledgements

- Members of the Bender group:
    - Avid Afzal
    - Samar Mahmoud
    - Lewis Mervin
    - Andreas Bender

- Philip Judson, Lhasa Ltd.

- CEFIC-LRI for funding