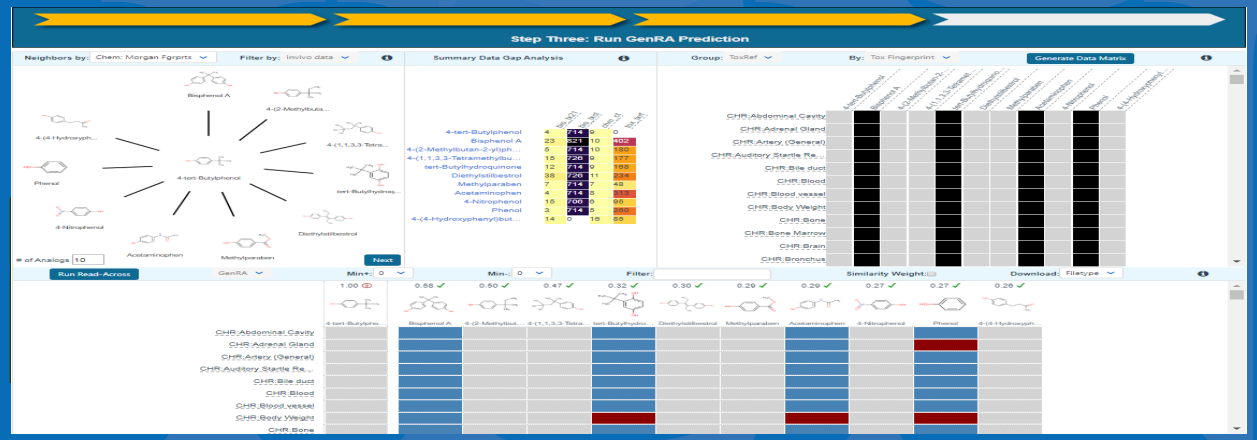
United States Environmental Protection Agency

Filling Metabolism Data Gaps in Read-across



Matthew Boyce, Brian Meyer, Vicente Samano, Chris Grulke, Lucina Lizarraga*, Grace Patlewicz

Center for Computational Toxicology & Exposure (CCTE), US EPA
*Center for Public Health and Exposure Assessment (CPHEA), US EPA

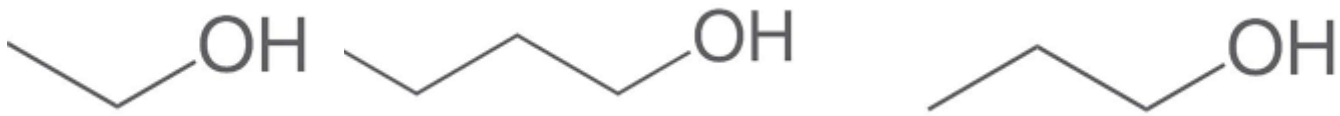
Outline

- Read-across definition
- Generalized Read-across (GenRA): Addressing shortcomings of current read-across approaches
- Contexts of similarity for read-across: Metabolic similarity
- Comparing and contrasting different data streams that provide metabolism information
- *In silico* metabolism: Assessing performance and coverage of tools
- Summary & Next steps

What is Read-Across?

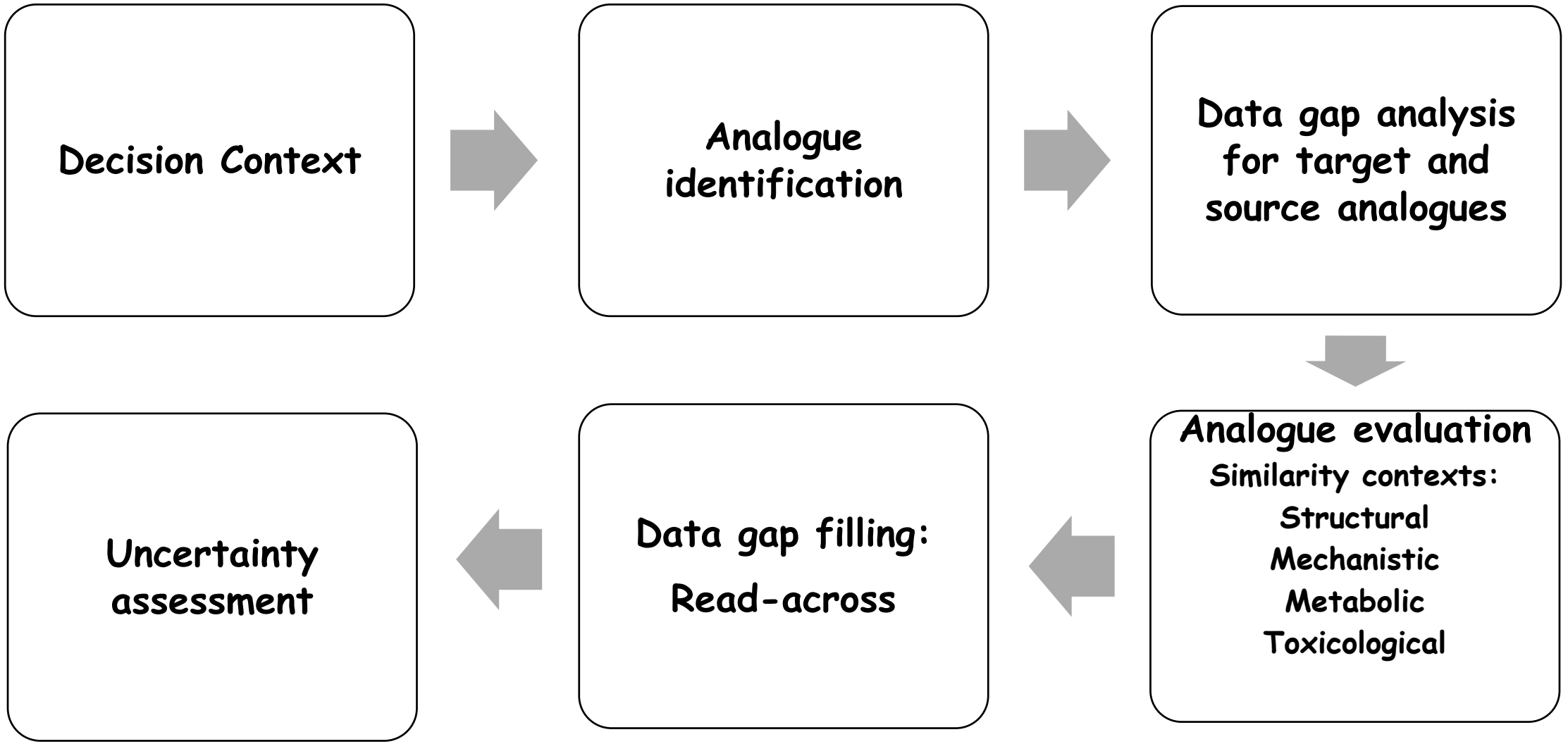
- Read-across describes the method of filling a data gap whereby a chemical with existing data values (source analogue) is used to make a prediction for a ‘similar’ chemical (target).

Source **Target**



	Ethanol	Butanol		Propanol
Property 1	●	●	→	○
Property 2	●	●		●
Property 3	●	●	→	○

Generic Read-across workflow

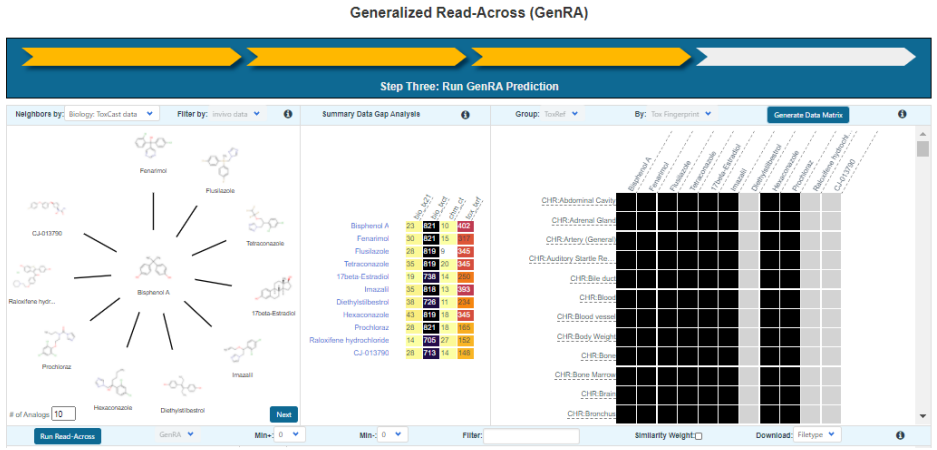


Ongoing issues with read-across

- Although there is much guidance for developing read-across assessment, acceptance remains an issue, not helped since read-across still remains a subjective, expert driven assessment.
- One issue thwarting acceptance relates to the “uncertainty of the read-across prediction”.
- As such there have been many efforts to identify the sources of uncertainty in read-across, characterize them in a consistent manner and identify practical strategies to address and reduce those uncertainties.
- Notable in these efforts have been the development of frameworks for the assessment of read-across, evaluating the utility of New Approach Methods (NAMs).
- Quantifying uncertainty and performance of read-across is still a need

Generalized Read-Across (GenRA)

- GenRA is a data-driven approach to read-across that uses a similarity weighted activity of source analogues to generate predictions. This objective approach enables uncertainty of read-across predictions to be quantified and performance to be assessed.



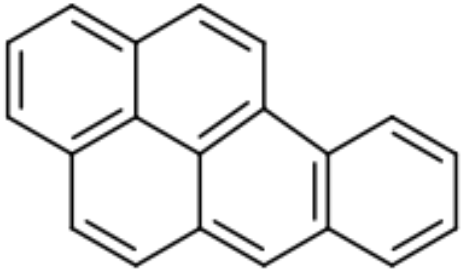
First implemented in the CompTox Chemicals DashBoard (Beta, 2018)

- Current focus is to extend the scope of GenRA to factor other contexts of similarity such as **metabolic similarity** and quantifying its contribution to toxicity predictions in conjunction with mechanistic and structural similarity.

Metabolism is an important similarity context

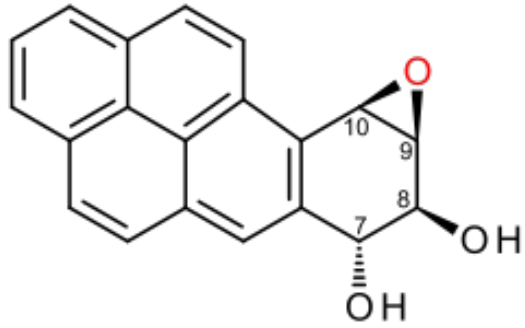
- Through metabolic activation, compounds can see a significant increase in toxicity which is not captured by the parent structure.

Example: Phase 1 metabolism of benzo[a]pyrene yields an epoxide ring that allows it to bond to DNA



Benza[a]pyrene

CYP1A1/CYP1B1 CYP1A1/CYP1B1



Benzo[*a*]pyrene-7,8-dihydrodiol-9,10-epoxide

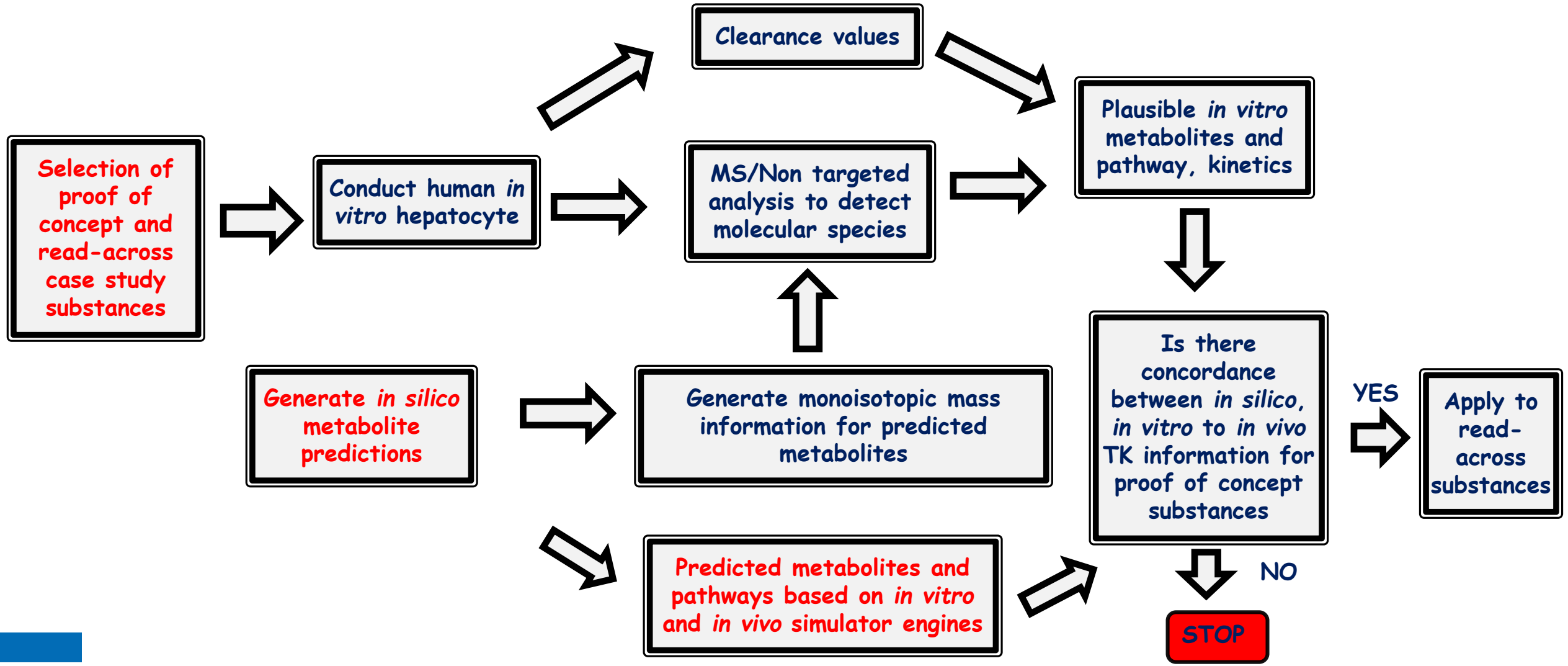
Comparing and contrast different sources of metabolism information

- Aim to investigate the concordance between *in vivo*, *in vitro* and *in silico* metabolism information and how it can be utilized to assess metabolic similarity for read-across
- Data streams
  - *In vitro*:
    - Perform *in vitro* rate and human hepatocyte study to determine intrinsic clearance
    - Apply analytical spectroscopy (MS) for the detection of molecular species and non-targeted analysis for metabolite identification
  - *In silico*
    - Use third party expert systems for the prediction of potential metabolites and their pathways to facilitate MS analysis
  - *In vivo*
    - Extract data in the peer reviewed literature

Comparing and contrast different sources of metabolism information

- Analysis:
  - Evaluate concordance of *in vitro* metabolism data relative to existing experimental data
  - Evaluate concordance of *in silico* metabolism to both *in vitro* metabolism and *in vivo* metabolism data for proof-of-concept substances
  - Use the predicted and experimental metabolism data to determine which source analogue(s) are valid for case study read-across candidates

Overall Project Workflow



Workflow for the in-silico vs experimental metabolism comparison

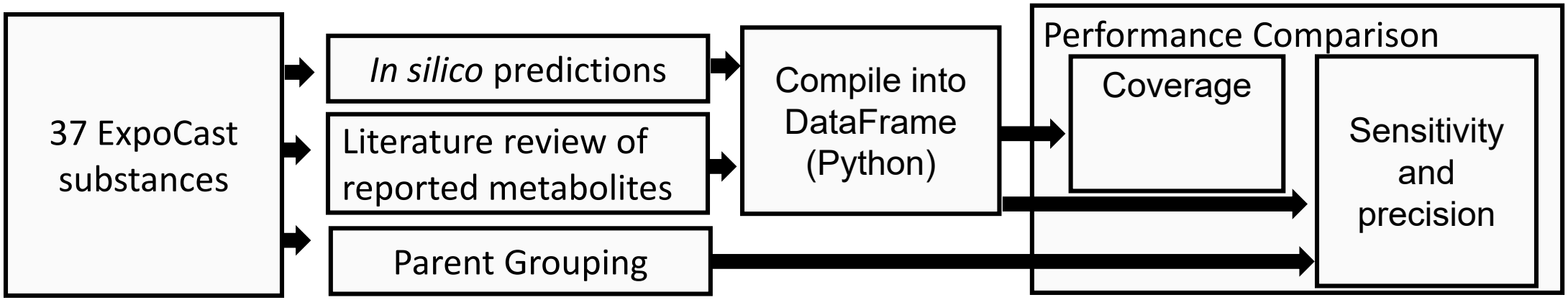
Gather	Store	Analyze
• Mine the literature for *in vivo* and *in vitro* data • Use *in silico* tools to generate predictions	• Update or add metabolites to EPA's chemical registration database (DSSTox) with relevant information (e.g., study reference, species, biotransformation pathway, parent compound, rate information) • Map biotransformation relationships between entities	• Compare performance and coverage of the *in silico* tools relative to experimental data collected

In silico tools for metabolite predictions

- *In silico* tools can provide a rapid and efficient method of predicting metabolites for compounds that lack published research.
- There are a number of metabolism prediction tools. Examples include: MetaPrint 2D, Meteor Nexus, TIMES, the simulators contained within the OECD Toolbox, Symga and Biotransformer.
- Some are freely available, others such as TIMES and Meteor are commercial.
- Few studies have been performed to directly compare the performance of these tools.

Evaluating *in silico* tools

37 proof of concept substances were selected from the ToxCast library—these compounds represented a broad spectrum of pharmaceutical, agrochemical, and industrial chemistries.



EPA United States Environmental Protection Agency

Selected *in silico* tools

	TIMES*	QSAR ToolBox	Meteor	BioTransformer	SyGMa	CTS
Developer	LMC	OECD and ECHA	Nexus	Wishart Group	Riddler & Wagener	EPA
Availability	Commercial	Public	Commercial	Public	Public	Public
Knowledge base	Expert + Statistical Score	Expert + Statistical Score	Expert + Statistical Score	Expert + ML	Expert	-
Customizable Met.	Yes	No	Yes	Yes	Yes	Limited
Interface	GUI	GUI/API	GUI‡	API/CMD Prompt/Terminal	API/Phyton	WebApp
Available Modules	Rat liver (*in vivo*), Rat liver (S9, *in vitro*), Lung (mammal), Gut (mammal), Autoxidation	Autoxidation, Hydrolysis, Rat (S9, *in vitro*), Rat (*in vivo*)	Mammals (Dog, Human, Mouse, Rat)	Human (Liver, Gut), Microbial	Human (Liver)	Human
#Predictions	211 (*vitro*), 459 (*vivo*)	312	459	827	5215	472

*Two modules were used separately for this work: rat *in vivo*, rat *in vitro*

‡ Batch mode requires command prompt or terminal

Literature review

Identification of known metabolites for all 37 ToxCast compounds:

-Extracted metabolites from 49 papers (prioritized primary articles)

-Identified 438 metabolites across all compounds

-Species were recorded, but not considered in performance comparison

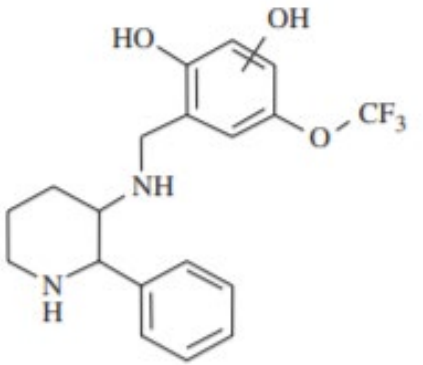
Metabolites were registered into EPA's DSSTox chemical registration system to generate specific identifiers (DTXSID/DTXCIDs) to facilitate subsequent data analysis

-Metabolism pathways captured using Proceeding/Preceding structures

-SMILES and InChI Keys were queried and retrieved for downstream analysis

Accounting for structural isomers

CP-122,721 metabolite



20 Isomers reported in literature

- Prediction software generates discrete structures, which need to be reconciled against literature for evaluation

- Requires enumeration of each potential metabolite to generate InChI Key
- Generated 585 Markush children

Accounting for isomeric structures

Analysis Workflow

1) Read and clean data from reported data and readout for each prediction software
2) Standardize SMILES for each prediction by removing stereochemistry and generate InChI keys
4) Assign Boolean value to indicate which in silico tool was associated with the prediction
5) Merge data to allow other information to be added, such as SMILES, molecular formula, molecular weight, etc.).

Example output:

	DTXSID	Metabolite_INCHIKEY	ToolBox	Meteor	BioTransformer	TIMES_InVivo	TIMES_InVitro	SyGMa	CTS	Reported	Metabolite DTXSID	Markush
3042	DTXSID0020151	UYSKXFUQLZBPKP-UHFFFAOYSA-N	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	NA	False
2963	DTXSID0020151	FLWQJXIDAGIGBN-UHFFFAOYSA-N	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	NA	False
3027	DTXSID0020151	NPZVFJGEBZDVBM-UHFFFAOYSA-N	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	NA	False
2964	DTXSID0020151	WWOBNORICRFIOK-UHFFFAOYSA-N	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	NA	False
2965	DTXSID0020151	CHLHAPIXOIWPHN-UHFFFAOYSA-N	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	NA	False
...	...	...	...	...	...	...	...	...	...	...	...	...
5526	DTXSID9047251	METSLMTUEASVPB-UHFFFAOYSA-N	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	NA	False
5527	DTXSID9047251	QFTGRERVWULBTD-UHFFFAOYSA-N	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	NA	False
5528	DTXSID9047251	YGJPATIPUUSKED-UHFFFAOYSA-N	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	NA	False
5508	DTXSID9047251	XZKSAJUKOYBZML-UHFFFAOYSA-N	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	NA	False
5962	DTXSID9047251	HICABXJERDJNAW-UHFFFAOYSA-N	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	NA	False

Quantifying performance: metrics used

Metric of model similarity

Coverage:

How well does model A match the predictions of model B?

$$\frac{Predictions\ A\ \cap Predictions\ B}{Predictions\ B}$$

Metrics of model performance

Sensitivity:

Does the model cover all reported metabolites?

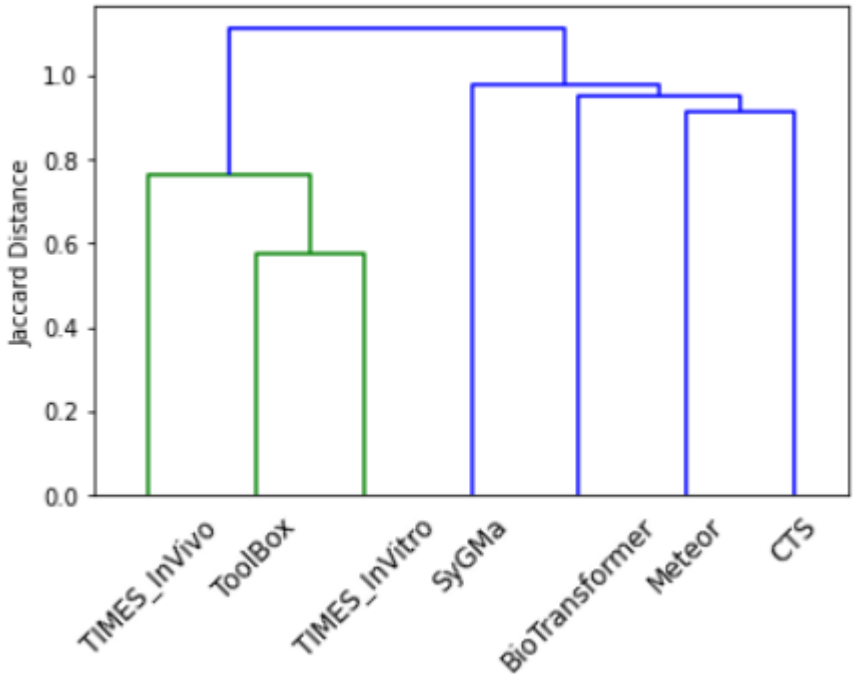
General equation: $\frac{True\ Predictions}{All\ Reported}$

Precision:

Are the predicted metabolites true?

General equation: $\frac{True\ Predictions}{All\ Predictions}$

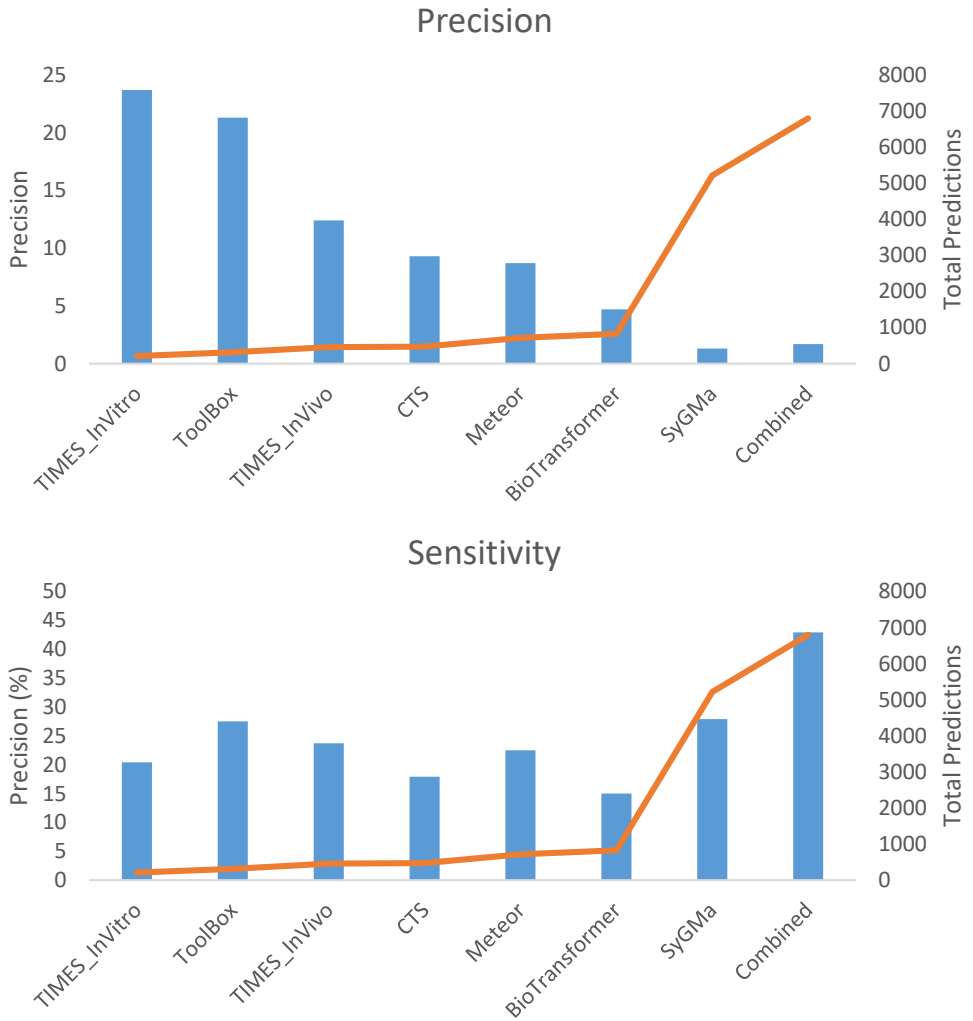
Comparing the *in silico* tools: relative coverage



Model	%Coverage of other software							
	ToolBox	Meteor	BioTransformer	TIMES In Vivo	TIMES In Vitro	SyGMa	CTS	Avg.
ToolBox	-	15.13	7.62	40.74	74.41	2.86	27.33	28.02
Meteor	34.39	-	9.07	25.93	33.65	4.60	19.92	21.26
BioTransformer	20.06	10.50	-	10.24	23.70	4.31	17.16	14.33
TIMES In Vivo	59.55	16.67	5.68	-	61.61	3.99	14.41	26.99
TIMES In Vitro	50.00	9.94	6.05	28.32	-	2.13	16.74	18.86
SyGMa	47.45	33.61	27.21	45.32	52.61	-	32.42	39.77
CTS	41.08	13.17	9.79	14.81	37.44	2.93	-	19.87

Comparing in silico performance: Precision

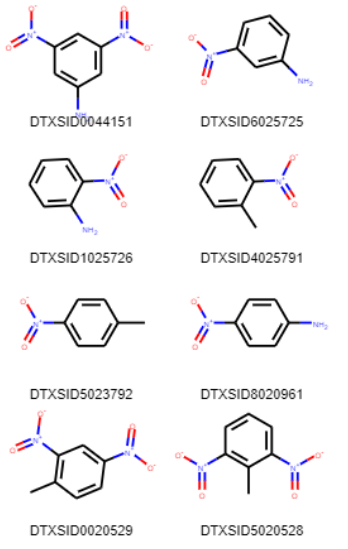
Model	Total	Unique	Precision	Sensitivity
ToolBox	314	12	21.3	27.5
Meteor	714	436	8.7	22.5
BioTransformer	827	570	4.7	15.0
TIMES_InVivo	459	122	12.4	23.7
TIMES_InVitro	211	10	23.7	20.4
SyGMa	5215	4667	1.3	27.9
CTS	472	252	9.3	17.9
Combined	6799	-	1.7	42.9



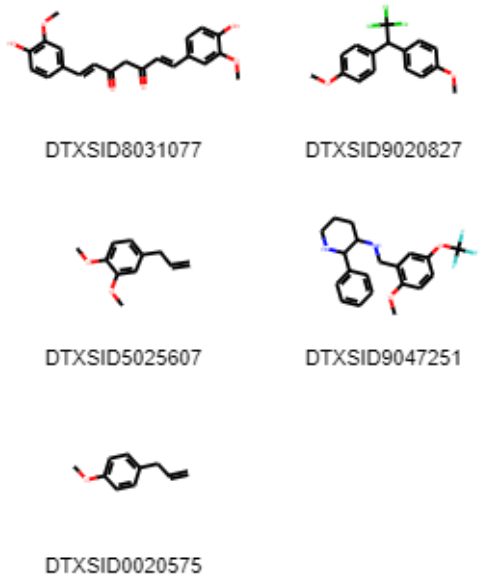
Assessing ‘local’ performance

To compare performance differences relative to the parent compound, five groupings were generating using ClassyFire classifications (a structural/functional group hierarchy) combined with clustering approaches

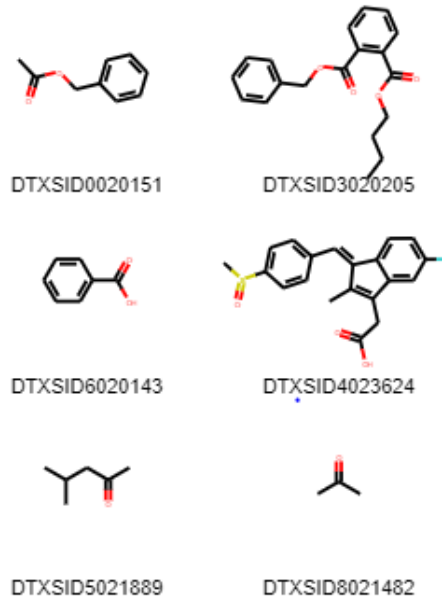
Group 1
Nitrobenzenes



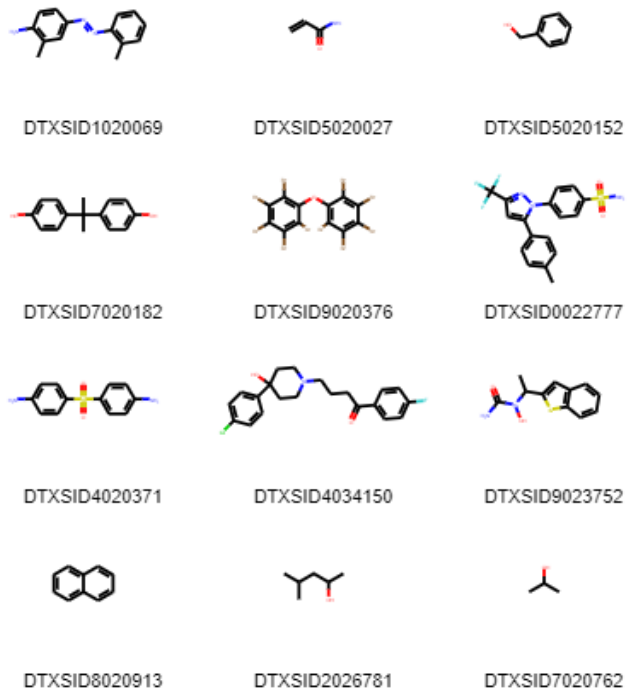
Group 2
Methoxybenzene



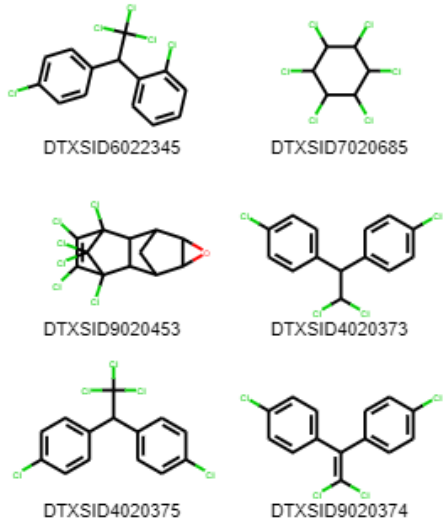
Group 3
Ketone/Benzoyl



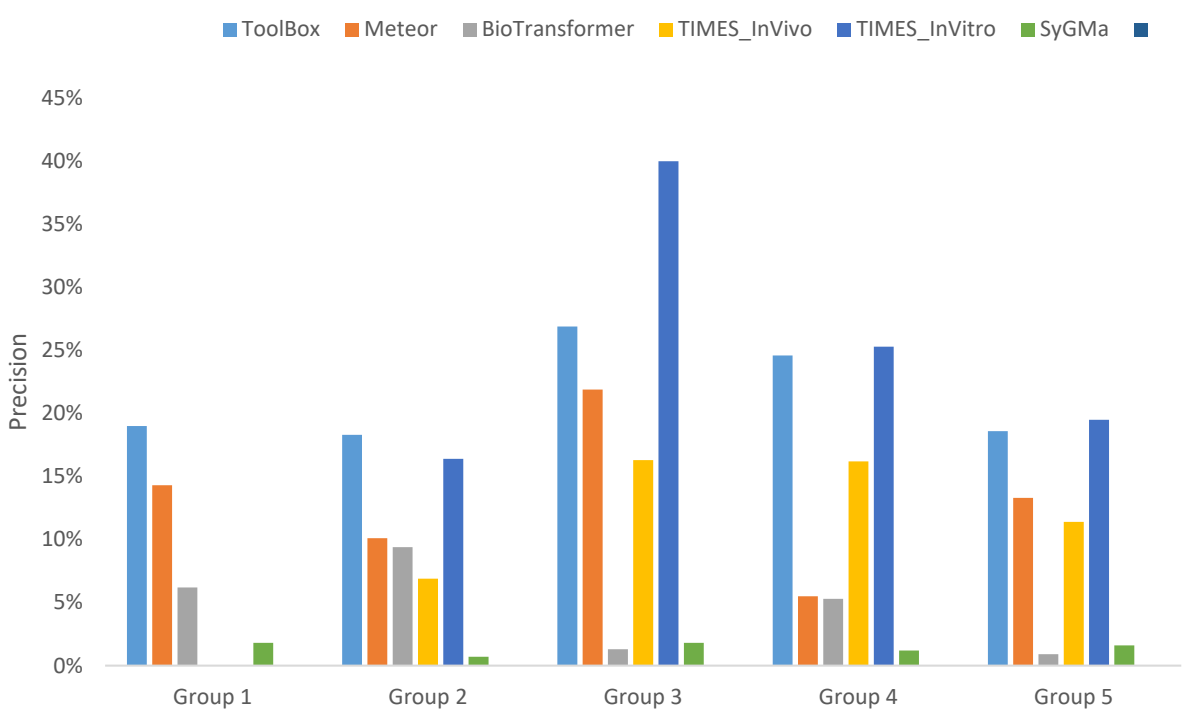
Group 4
Misc

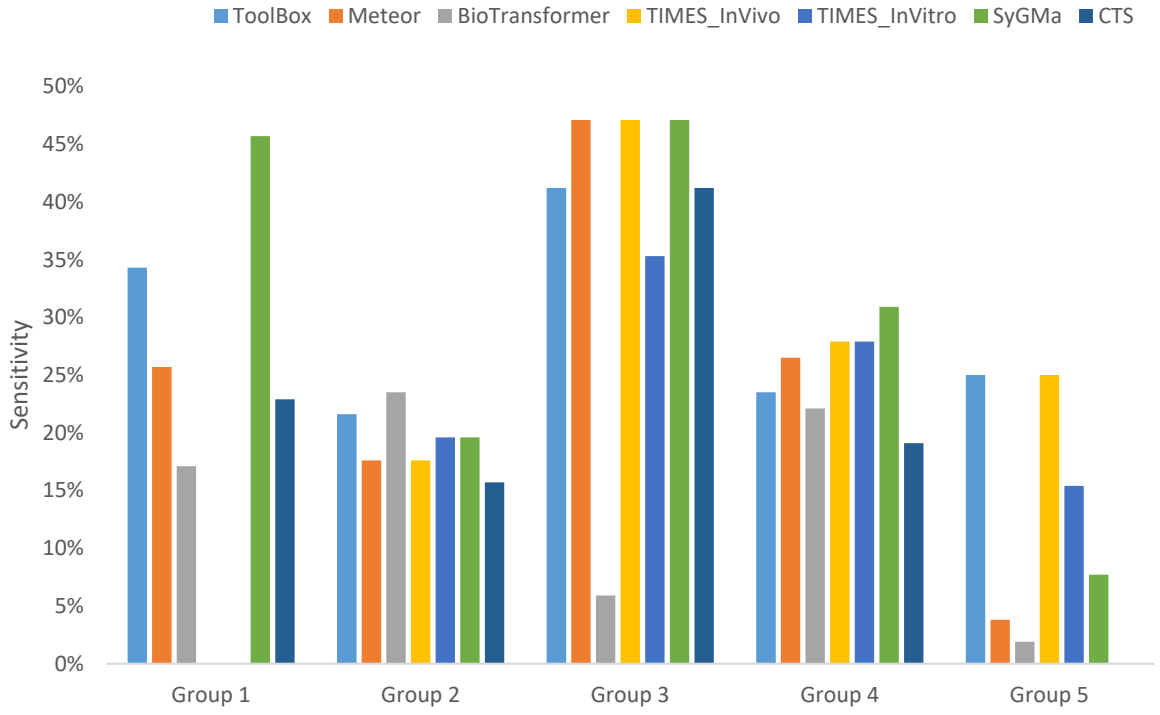


Group 5
Poly-chlorinated



‘Local’ chemical space performance





Summary

- Metabolic similarity is an important component in evaluating analogue suitability within a read-across.
- Approaches to characterize and quantify metabolic similarity is needed
- A proof of concept study is ongoing to compare and contrast different metabolism information sources and evaluate their utility for read-across amongst other purposes.
- Specific *in vitro* data has been generated and is currently being evaluated.
- Predictions have been generated using a selection of *in silico* tools.
- Experimental data has been extracted from the literature.

Summary

- The performance of *in silico* metabolite prediction tools has been evaluated.
    - Coverage was calculated to provide relative comparisons between each tool and provided a metric for prediction similarity
    - Sensitivity and precision were determined by comparing predictions against metabolites reported in literature
    - Using ClassyFire classifications, model performance could be evaluated relative to the 'local' chemical space of the starting compounds
- A manuscript to summarize the *in silico* evaluation is in preparation.
- Next steps include:
    - Evaluating the concordance of *in vitro* data generated relative to the literature data collected and the *in silico* tools
    - Generate *in silico* predictions for a larger number of substances
    - Investigate how to codify and quantify the metabolism information from 1 or more of the *in silico* tools for the purposes of read-across using the GenRA approach