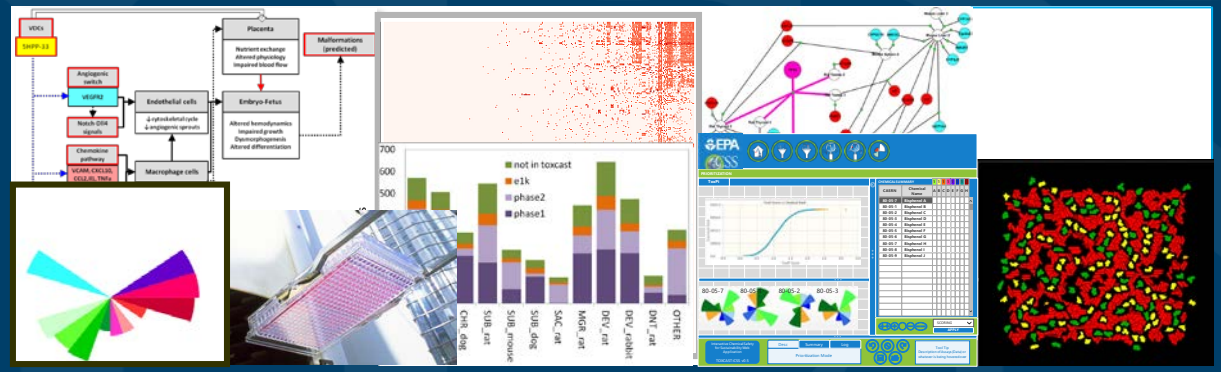
Systematic Approaches to Biological/Chemical Read-Across for Hazard Identification



Grace Patlewicz, George Helman*, Imran Shah

National Center for Computational Toxicology
***ORISE Fellow**

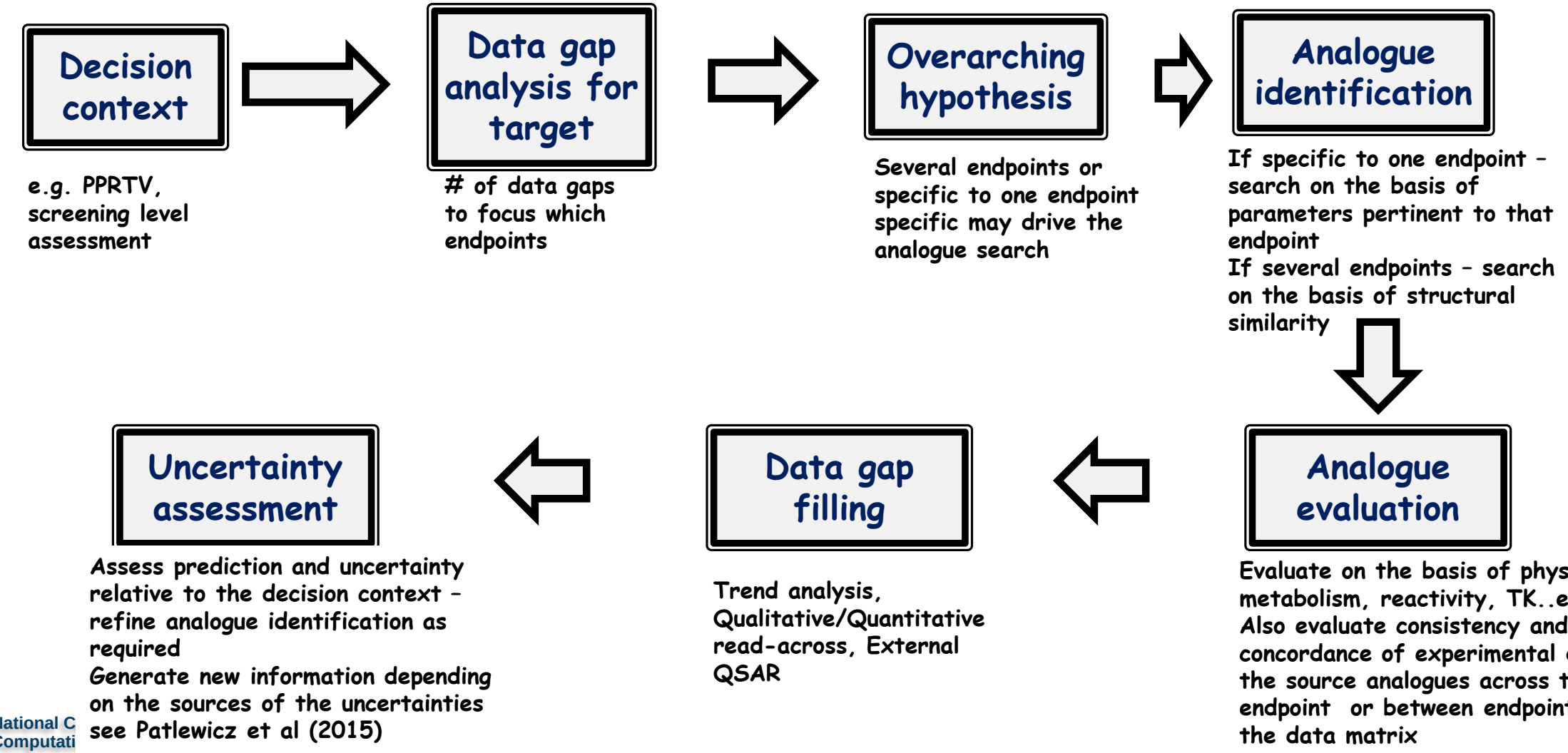
Outline

- Background, Concepts and Definitions
- Category workflow and selected tools for read-across
- Uncertainty assessment in read-across
- Quantifying Uncertainty & Assessing Performance of Read-Across
- From Research to Implementation
- Ongoing research
- Summary

Background & definitions

- Read-across describes one of the data gap filling techniques used within analogue and category approaches
- "Analogue approach" refers to grouping based on a very limited number of chemicals (e.g. target substance + source substance)
- "Category approach" is used when grouping is based on a more extensive range of analogues (e.g. 3 or more members)

Category Workflow



Selected Read-Across Tools

Tool	AIM	ToxMatch	AMBIT	OECD Toolbox	CBRA	ToxRead
Analogue identification	X	X	X	X	X	X
Analogue Evaluation	NA	X	X by other tools available	X	X	X For Ames & BCF
Data gap analysis	NA	X	X Data matrix can be exported	X Data matrix viewable	NA	NA
Data gap filling	NA	X	User driven	X	X	X
Uncertainty assessment	NA	NA	NA	X	NA	NA
Availability	Free	Free	Free	Free	Free	Free

Selected Read-Across Tools – Review

Computational Toxicology 3 (2017) 1–18

Contents lists available at ScienceDirect

Computational Toxicology

journal homepage: www.elsevier.com/locate/comtox

Navigating through the minefield of read-across tools: A review of in silico tools for grouping

Grace Patlewicz [a,*], George Helman [a,b], Prachi Pradeep [a,b], Imran Shah [a]

[a] National Center for Computational Toxicology (NCCT), Office of Research and Development, US Environmental Protection Agency, 109 TW Alexander Dr, Research Triangle Park (RTP), NC 27711, USA
[b] Oak Ridge Institute for Science and Education (ORISE), Oak Ridge, TN, USA

ARTICLE INFO

Article history:
Received 29 March 2017
Received in revised form 22 May 2017
Accepted 25 May 2017
Available online 29 May 2017

Keywords:
Category approach
Analogue approach
Data gap filling
Read-across
(Q)SAR
Trend analysis
Nearest neighb

ABSTRACT

Read-across is a popular data gap filling technique used within analogue and category approaches for regulatory purposes. In recent years there have been many efforts focused on the challenges involved in read-across development, its scientific justification and documentation. Tools have also been developed to facilitate read-across development and application. Here, we describe a number of publicly available read-across tools in the context of the category/analogue workflow and review their respective capabilities, strengths and weaknesses. No single tool addresses all aspects of the workflow. We highlight how the different tools complement each other and some of the opportunities for their further development to address the continued evolution of read-across.

Published by Elsevier B.V.

Selected Read-Across Tools

Tool	AIM	ToxMatch	AMBIT	OECD Toolbox	CBRA	ToxRead
Analogue identification	X	X	X	X	X	X
Analogue Evaluation	NA	X	X by other tools available	X	X	X For Ames & BCF
Data gap analysis	NA	X	X Data matrix can be exported	X Data matrix viewable	NA	NA
Data gap filling	NA	X	User driven	X	X	X
Uncertainty assessment	NA	NA	NA	X	NA	NA
Availability	Free	Free	Free	Free	Free	Free

Sources of Uncertainty

- Analogue or category approach? (# analogues)
- Completeness of the data matrix – no. of data gaps
- Data quality for the underlying analogues for the target and source analogues
- Consistency of data across the data matrix – concordance of effects and potency across analogues
- Address the dissimilarities and whether these are significant from a toxicological standpoint
- Toxicokinetics
- ……..

Uncertainty assessment

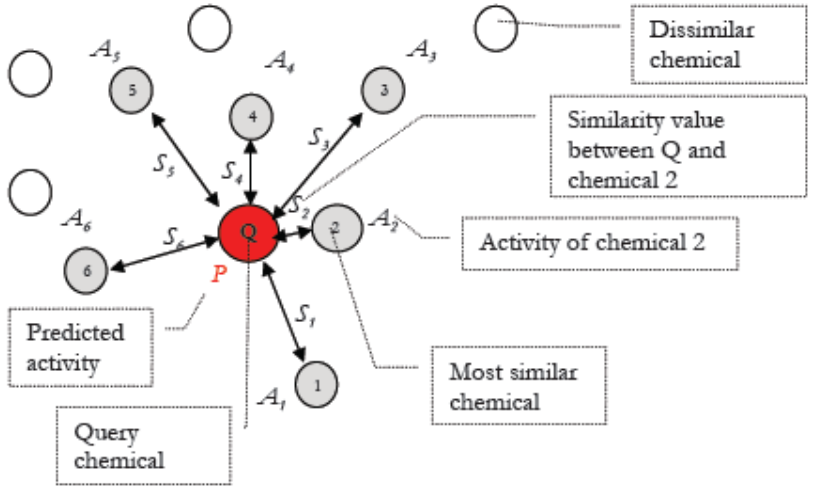
- There are several frameworks which aim to identify, document and address the uncertainties associated with read-across inferences/predictions
  - Blackburn & Stuard (2014)
  - Patlewicz et al (2015)
  - Schultz et al (2015)
  - ECHA RAAF (2015)

- However read-across acceptance relies on a subjective expert assessment

- There is no objective measure of read-across performance and acceptance

Quantifying Uncertainty & Assessing Performance of Read-Across

- **GenRA (Generalised Read-Across)** is a "local validity" approach
- Predicting toxicity as a similarity-weighted activity of nearest neighbours based on chemistry and/or bioactivity descriptors
- Generalised version of Chemical-Biological Read-Across (CBRA) developed by Low et al (2013)
- Goal: to systematically evaluate read-across performance and uncertainty using available data



$$y_i^{\beta,\alpha} = \frac{\sum_j^k s_{ij}^{\alpha} x_j^{\beta}}{\sum_j^k s_{ij}^{\alpha}}$$

Jaccard similarity:

$$s_{ij} = \frac{\sum_l (x_{il} \wedge x_{jl})}{\sum_l (x_{il} \vee x_{jl})}$$

$\alpha \in \{chm, bio, bc\}$

$\beta \in \{bio, tox\}$

y_i = predicted activity of chemical (c_i)

x_j^{β} = activity of c_j in β

s_{ij}^{α} = Jacccard similarity between $x_i^{\alpha}, x_j^{\alpha}$

k = up to k nearest neighbours

GenRA - Approach

I. Data

1,778 Chemicals
3,239 Structure descriptors (chm)
820 Bioactivity hitcall (bio) ToxCast

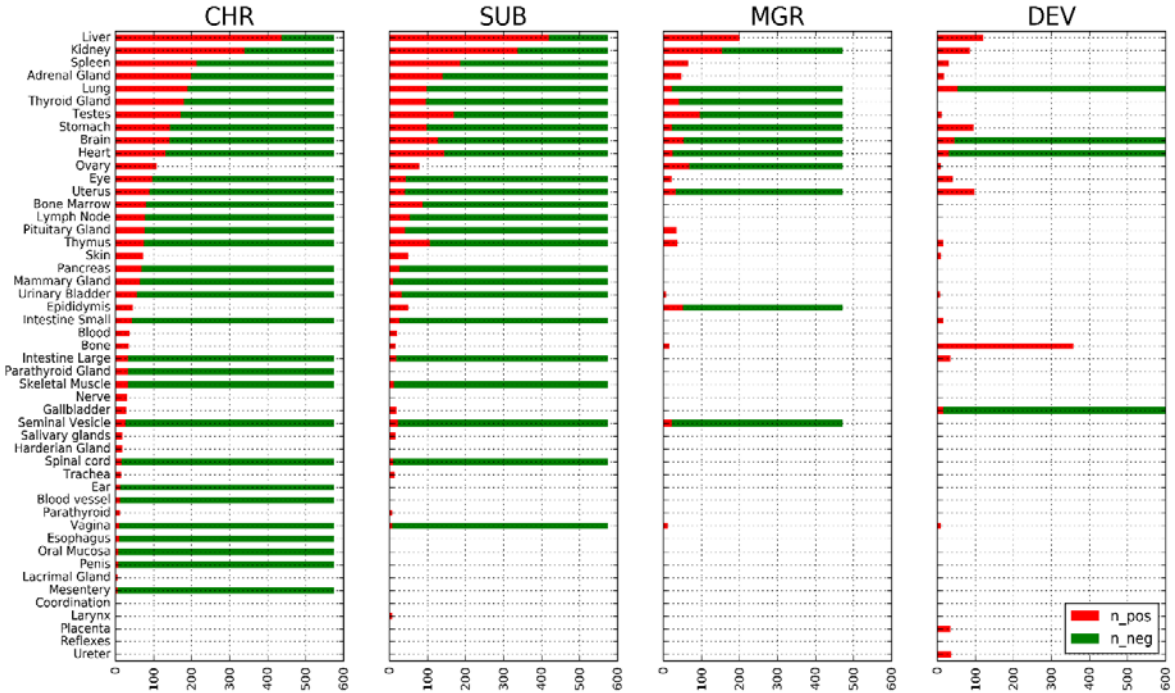
574 toxicity effects (tox) ToxRefDB

II. Define Local neighbourhoods

Use K-means analysis to group chemicals by similarity
Use cluster stability analysis
~ 100 local neighbourhoods

III. GenRA

Use GenRA to predict toxicity effects in local neighbourhoods
Evaluate impact of structural and/or bioactivity descriptors on prediction
Quantify uncertainty



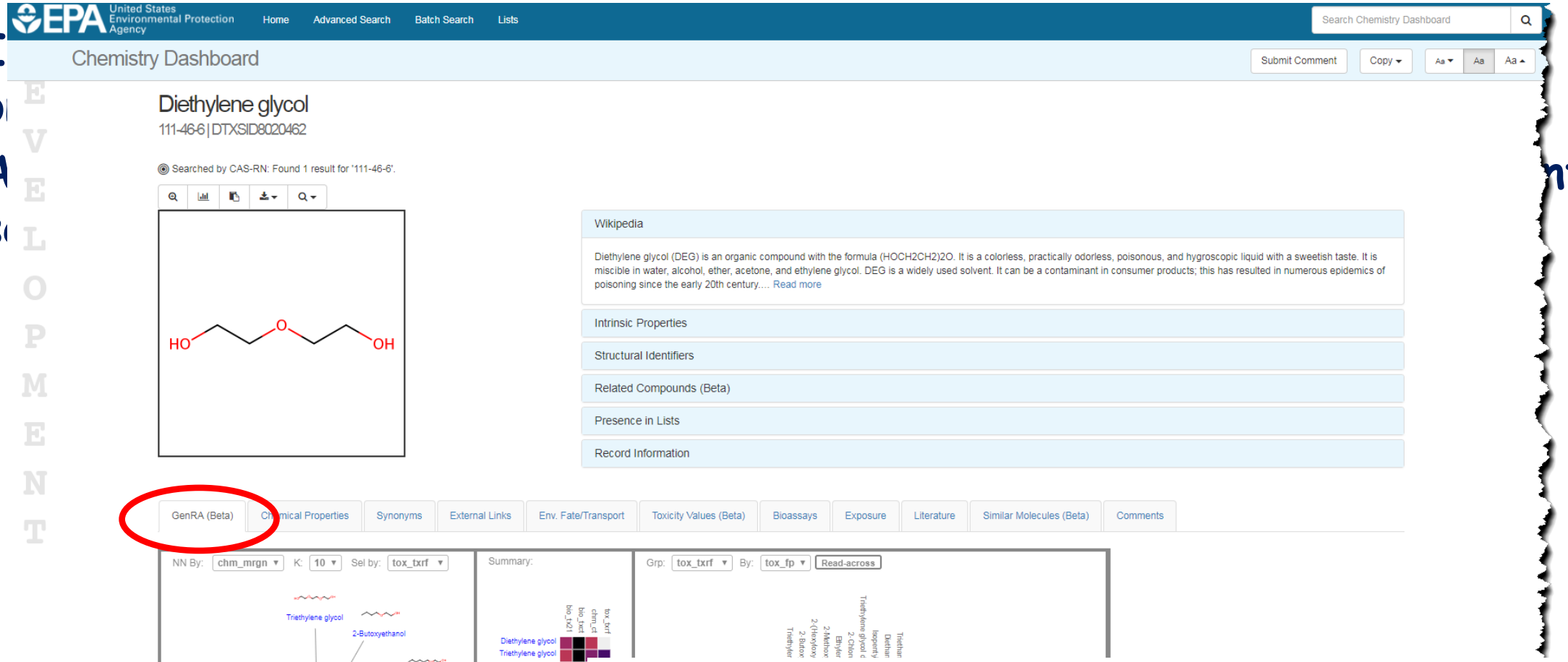
GenRA – Insights and Next Steps

- The approach enabled a performance baseline for read-across predictions of toxicity effects within specific study outcomes to be established but was still context dependent on the endpoint and the chemical

- Ongoing analysis:
- Consideration of other information to refine the analogue selection – e.g. physicochemical similarity, TK similarity, metabolic similarity, reactivity similarity…

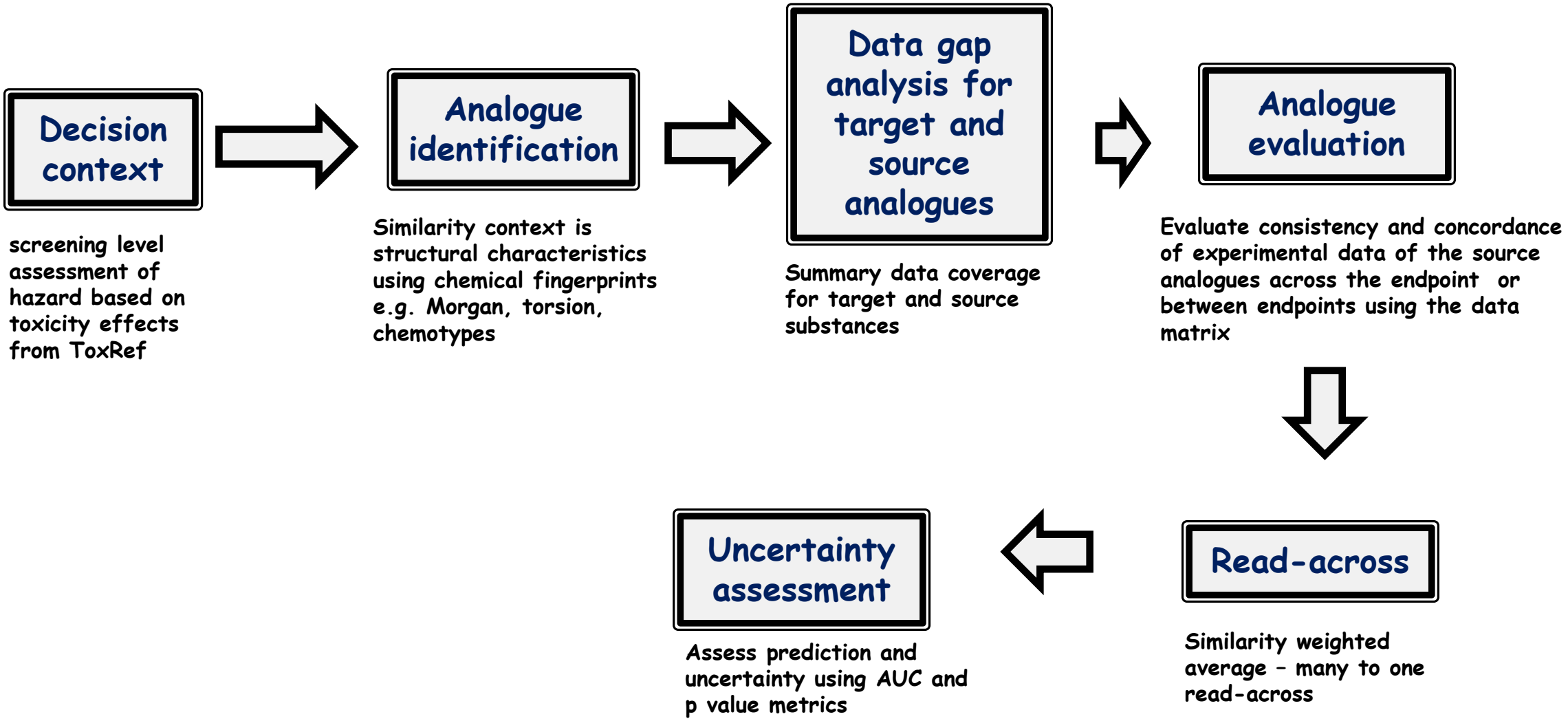
From research to implementation: GenRA prototype

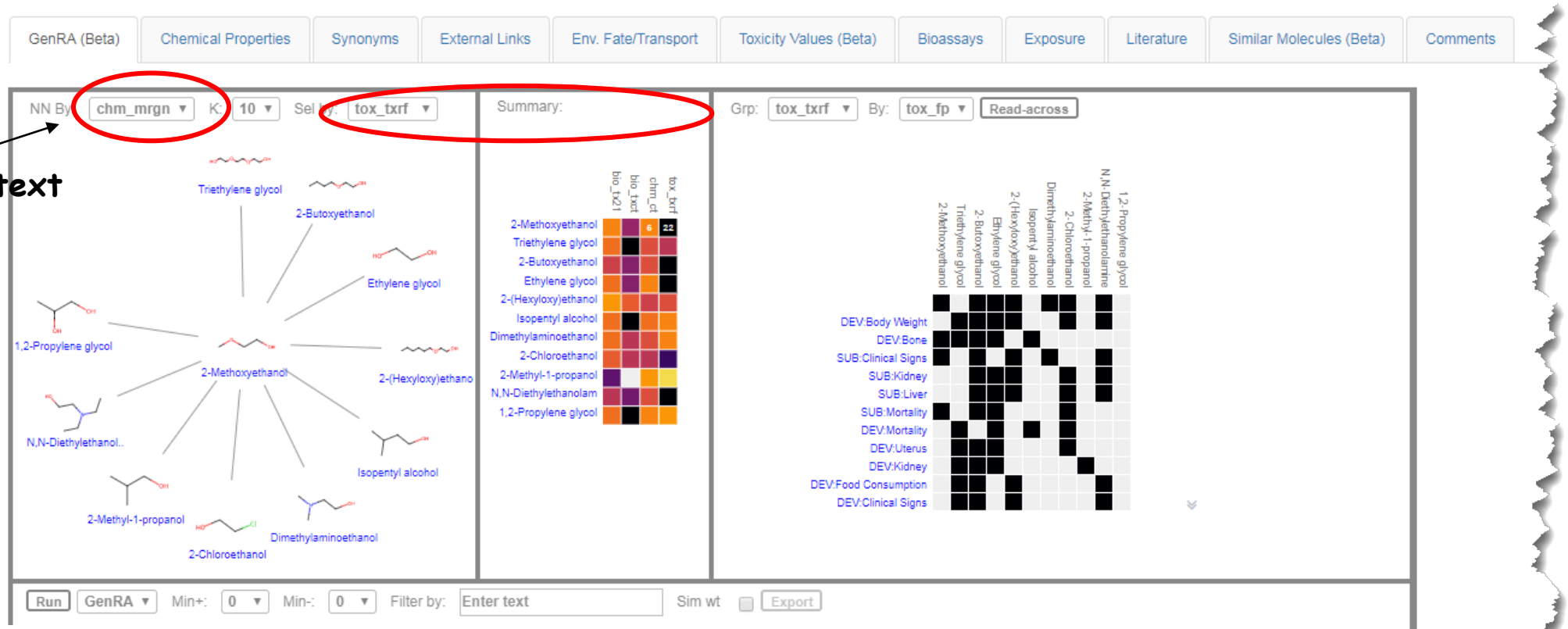
- I
- o
- A
- S



Grid interface where selections are dynamically updated in subsequent windows

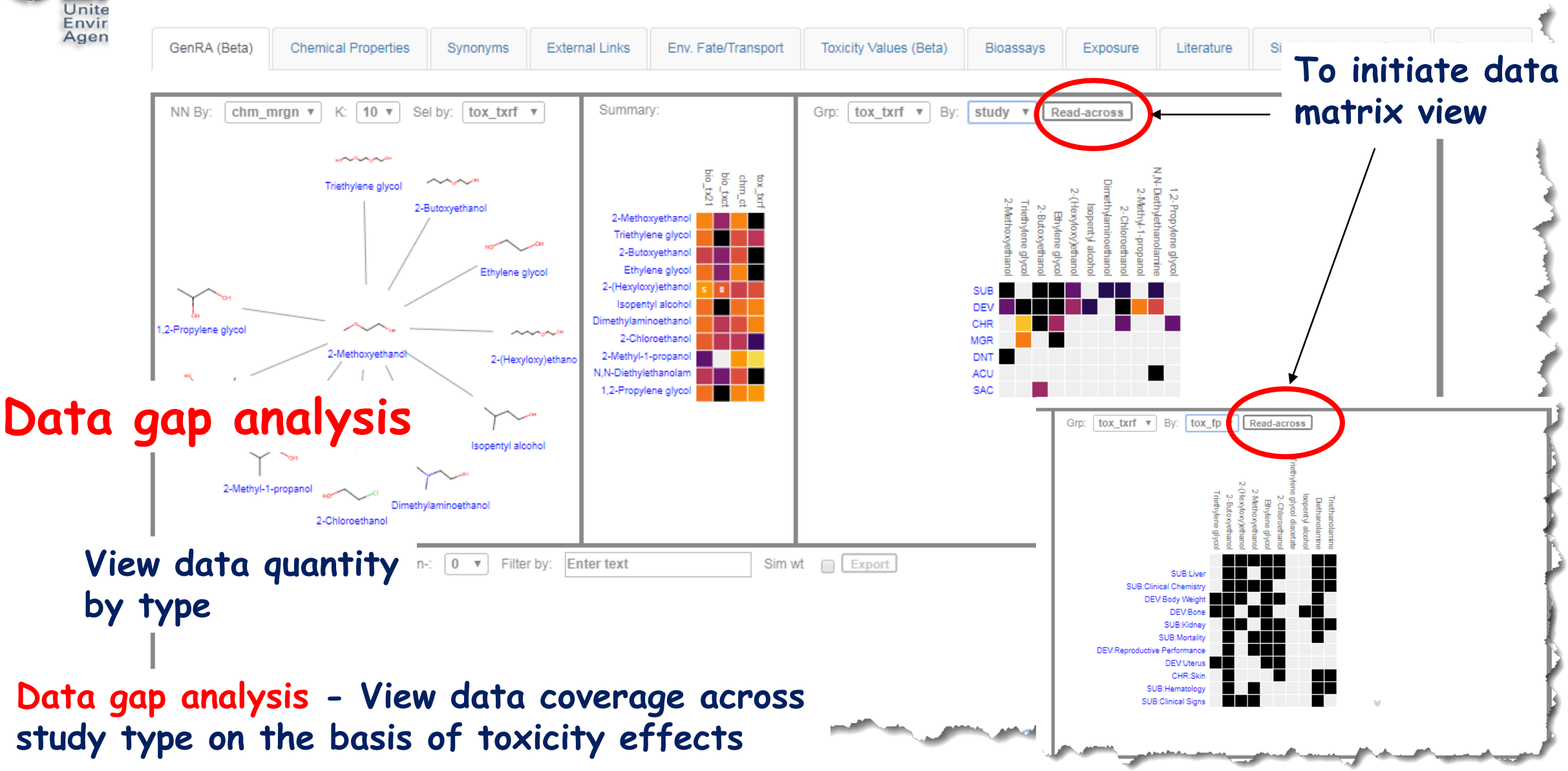
Current Category Workflow in GenRA





Similarity context

Analogue identification:
Search for source analogues on the basis of chemical fingerprints, filtered by availability of in vivo data

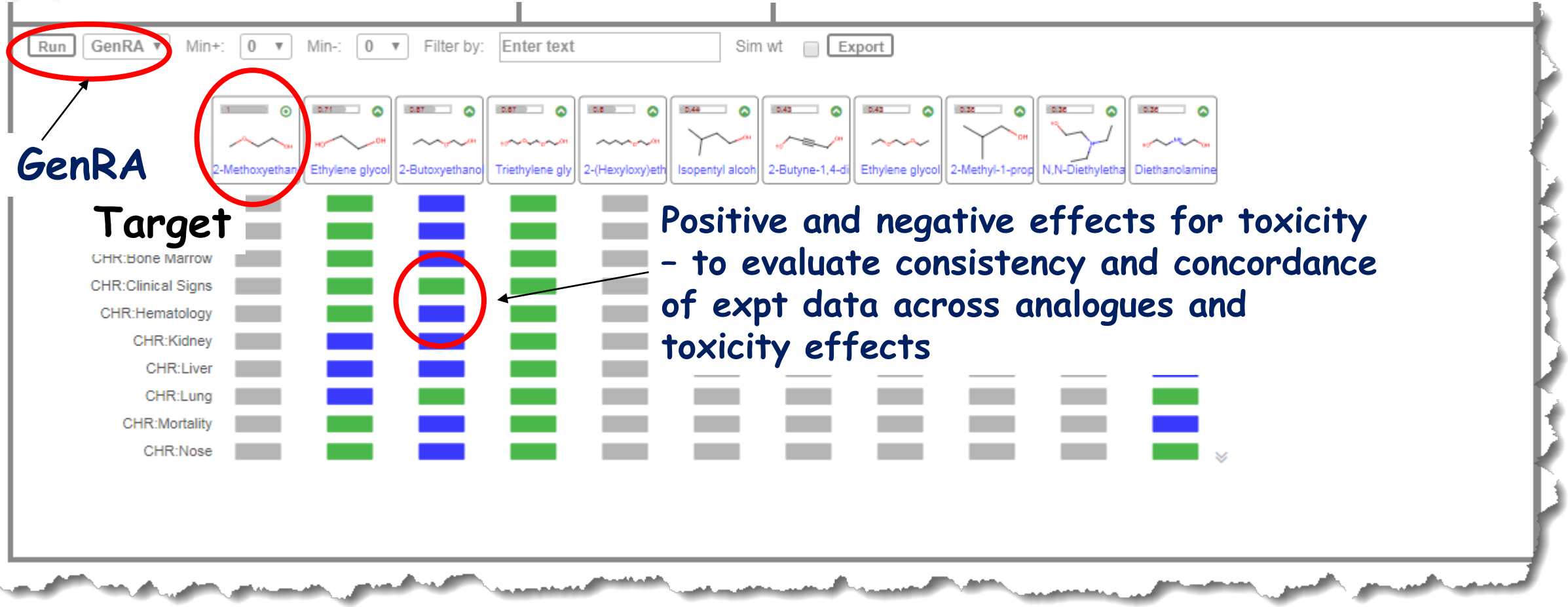


Data gap analysis

View data quantity by type

Data gap analysis - View data coverage across study type on the basis of toxicity effects

Analogue evaluation using data matrix view



Run GenRA

Data gap filling using GenRA within data matrix

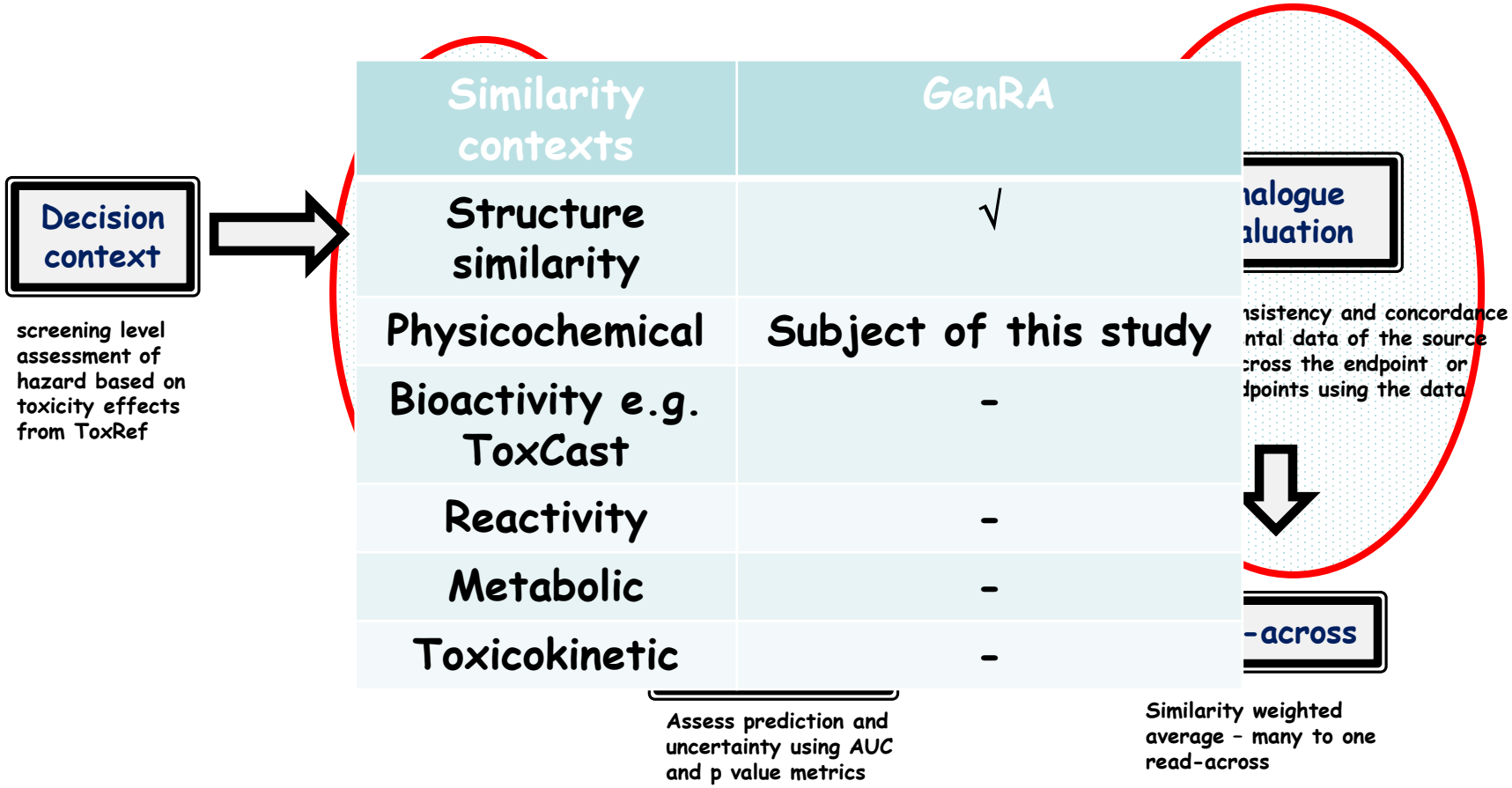
Export to a csv file

A	B	C	D	E	F	G	H	I	J	K
cls	target	analog	analog	analog	analog	analog	analog	analog	analog	
label	2-Methoxyethanol	Ethylene glycol	2-Butoxyethanol	Triethyler	2-(Hexylo	Isopentyl	2-Butyne-	Ethylene g	2-Methyl-1-propanol	
dsstox_cid	DTXCID804182	DTXCID40597	DTXCID904097	DTXCID60	DTXCID60	DTXCID70	DTXCID90	DTXCID30	DTXCID601759	
casrn	109-86-4	107-21-1	111-76-2	112-27-6	112-25-4	123-51-3	110-65-6	629-14-1	78-83-1	
jaccard	1	0.714285714	0.666666667	0.666667	0.6	0.444444	0.428571	0.428571	0.375	
CHR:Bone Marrow	GenRA Neg Act=0 (0.326) AUC=0 p=0.685	no_effect	125.000 ppm	no_effect	no_data	no_data	no_data	no_data	no_data	
DEV:Bone	GenRA TP Act=1 (1) AUC=0 p=1(50.000 ppm)	750.000 mg/kg/day	100.000 ppm	5630.000	no_effect	0.500 pp	no_effect	100.000 r	no_effect	
MGR:Bone	GenRA Pos Act=1 (0.517) AUC=0 p=0.51	1333.330 mg/kg/day	no_data	no_effect	no_data	no_data	no_data	no_data	no_data	
SUB:Bone	GenRA FN Act=0 (0.483) AUC=0 p=0.66(546.000 mg/kg/day)	no_effect	500.000 ppm	no_data	no_effect	no_data	no_data	no_data	no_data	
SUB:Bone Marrow	GenRA FN Act=0 (0.483) AUC=0 p=0.65(297.000 mg/kg/day)	no_effect	62.500 ppm	no_data	no_effect	no_data	no_data	no_data	no_data	

Selected Read-Across Tools

Tool	AIM	ToxMatch	AMBIT	OECD Toolbox	CBRA	ToxRead	GenRA
Analogue identification	X	X	X	X	X	X	X
Analogue Evaluation	NA	X	X by other tools available	X	X	X For Ames & BCF	NA
Data gap analysis	NA	X	X Data matrix can be exported	X Data matrix viewable	NA	NA	X Data matrix can be exported
Data gap filling	NA	X	User driven	X	X	X	X
Uncertainty assessment	NA	NA	NA	X	NA	NA	X
Availability	Free	Free	Free	Free	Free	Free	Beta for Internal testing

Ongoing research



Physchem Similarity Context

- Important context of similarity in read-across
- Models "bioavailability"
- Properties selected: Lipinski Rule of 5 (LogP, MW, # HB donors/acceptors)
- Two approaches investigated as a means to identify source analogs and evaluate their predictive performance relative to GenRA:

Approach 1: "Filter"

Subcategorise from a set of analogues identified based on structural similarity

Common approach

Approach 2: "Search Expansion"

"Frontload" both structure and physchem into analogue identification

Novel approach

Approaches considered

United States Environmental Protection Agency

Approach 1: Filter

- Similarity search using Jaccard distance of Morgan chemical fingerprints to find source analogues. (Default of 10 nearest neighbours (k=10))

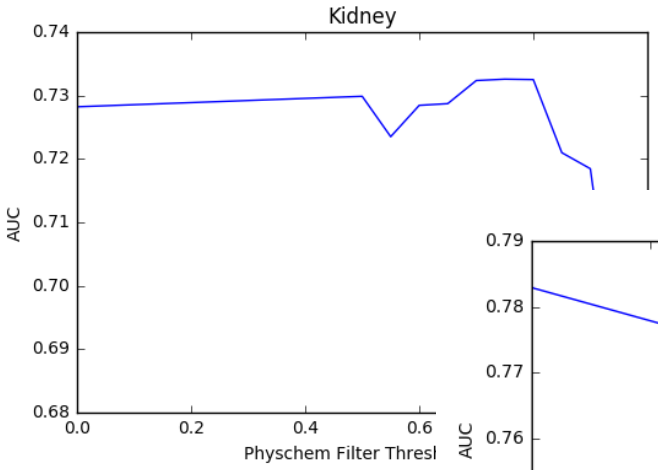
- Calculate physchem similarity between target and source analogues using a generalised Jaccard similarity metric

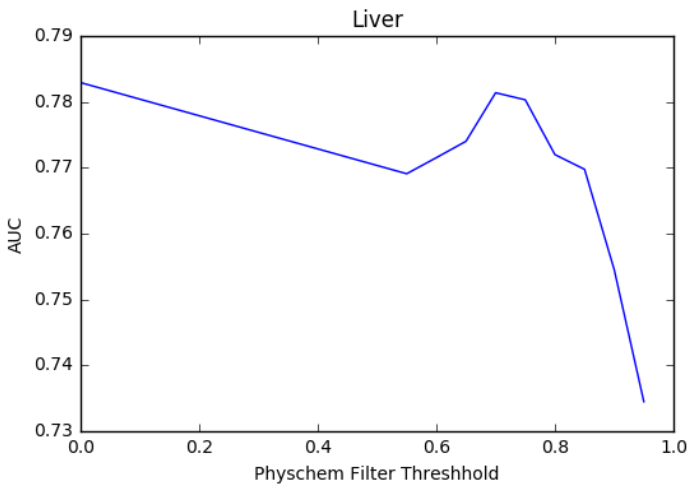
- Reduce neighbourhood based on the physchem similarity threshold

Approaches considered

Approach 1: Filter

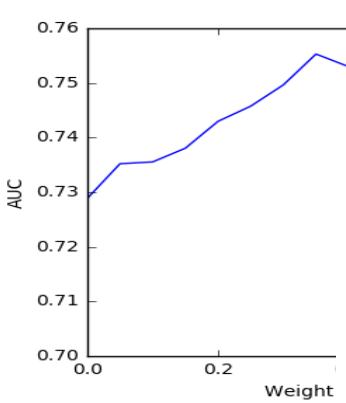
- This approach did not perform as well as GenRA for the entire dataset, nor did it significantly improve any target organ predictions.

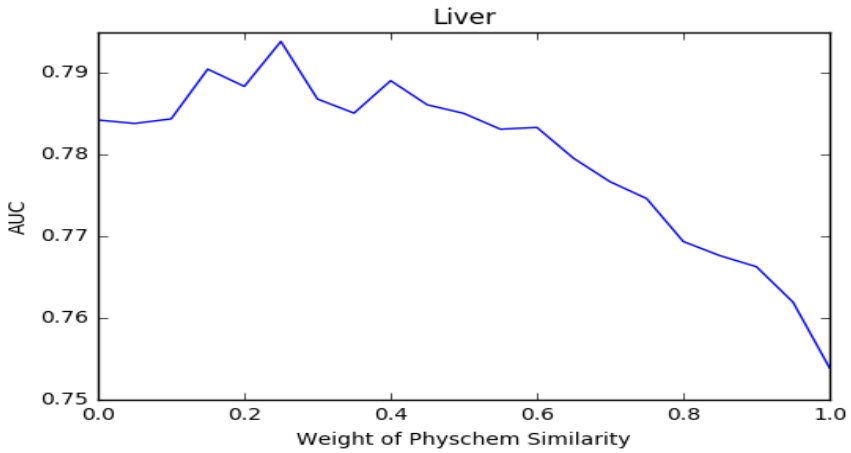




Approach 2: Search Expansion

- This approach shows a small improvement over baseline for entire dataset, but large improvement in certain organs.
- Target organ predictions that were significantly improved: Intestine Large, Intestine Small, Mammary Gland, Pancreas, Ureter, Urinary Bladder

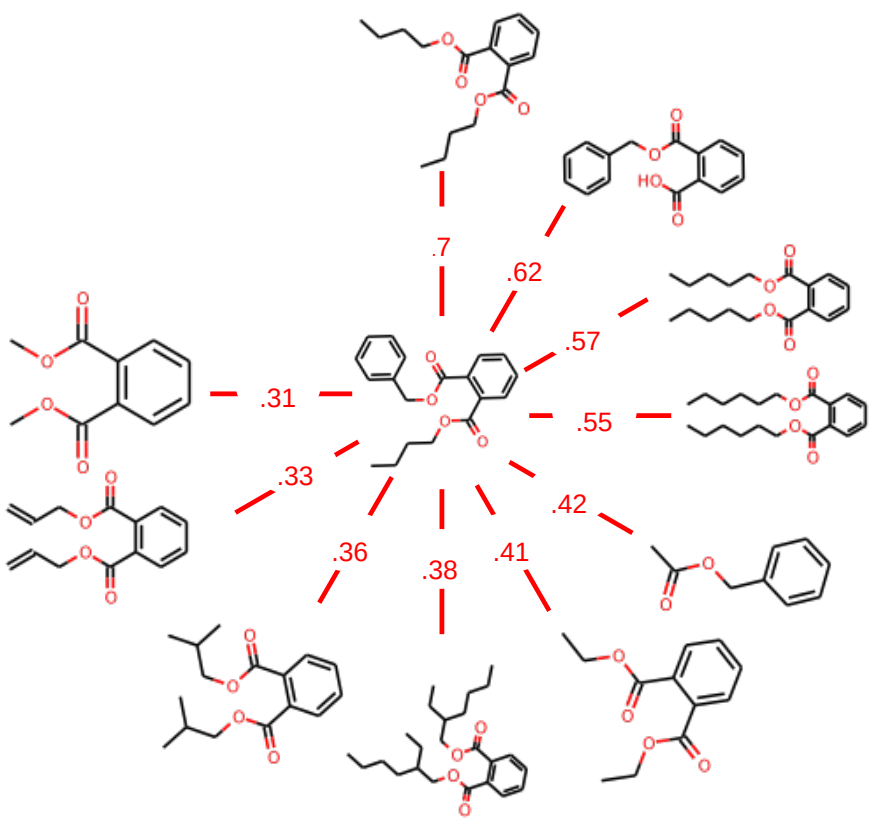




Target organ toxicity predictions aggregated over study type to organ level

Case Study: Butyl Benzyl Phthalate

GenRA: Baseline



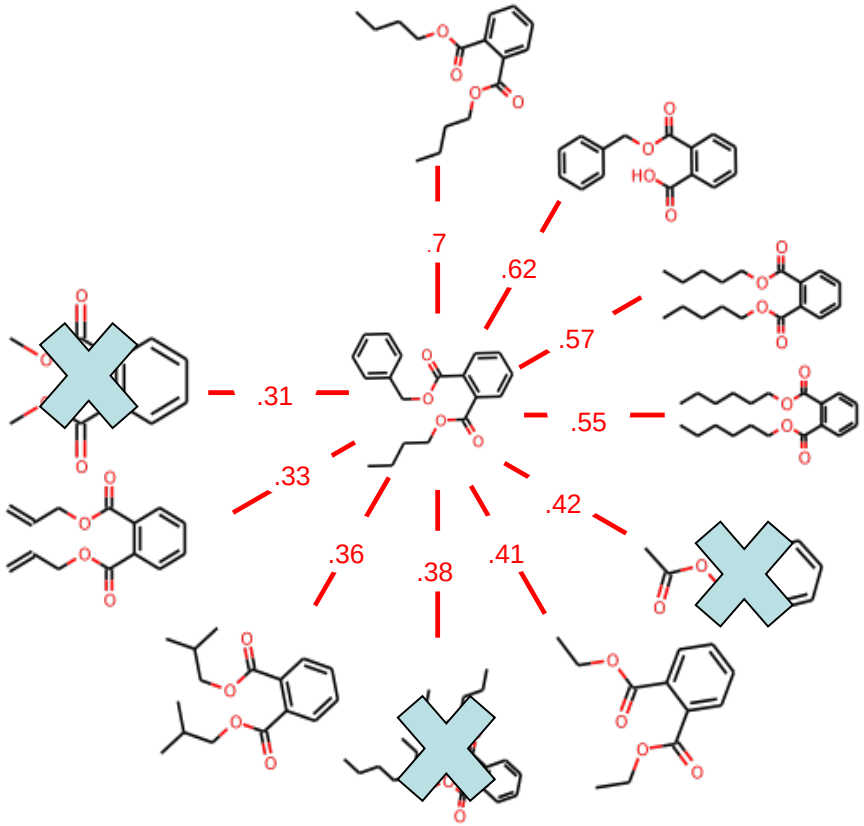
Endpoint	Baseline Prediction
Body Weight	.78
Clinical Chemistry	.27
Food Consumption	0
Hematology	0
Kidney	.27
Liver	1
Mortality	.27
Pancreas	.27
Prostate	0
Skin	.27
Spleen	0
Tissue NOS	0
Urinary Bladder	0

- Chronic studies
- All true positive effects
- Predictions between 0 and 1
- Higher prediction indicates more and stronger positive neighbours

Case Study: Butyl Benzyl Phthalate

Approach 1: Filter

Filter out chemicals with physchem similarity <0.8



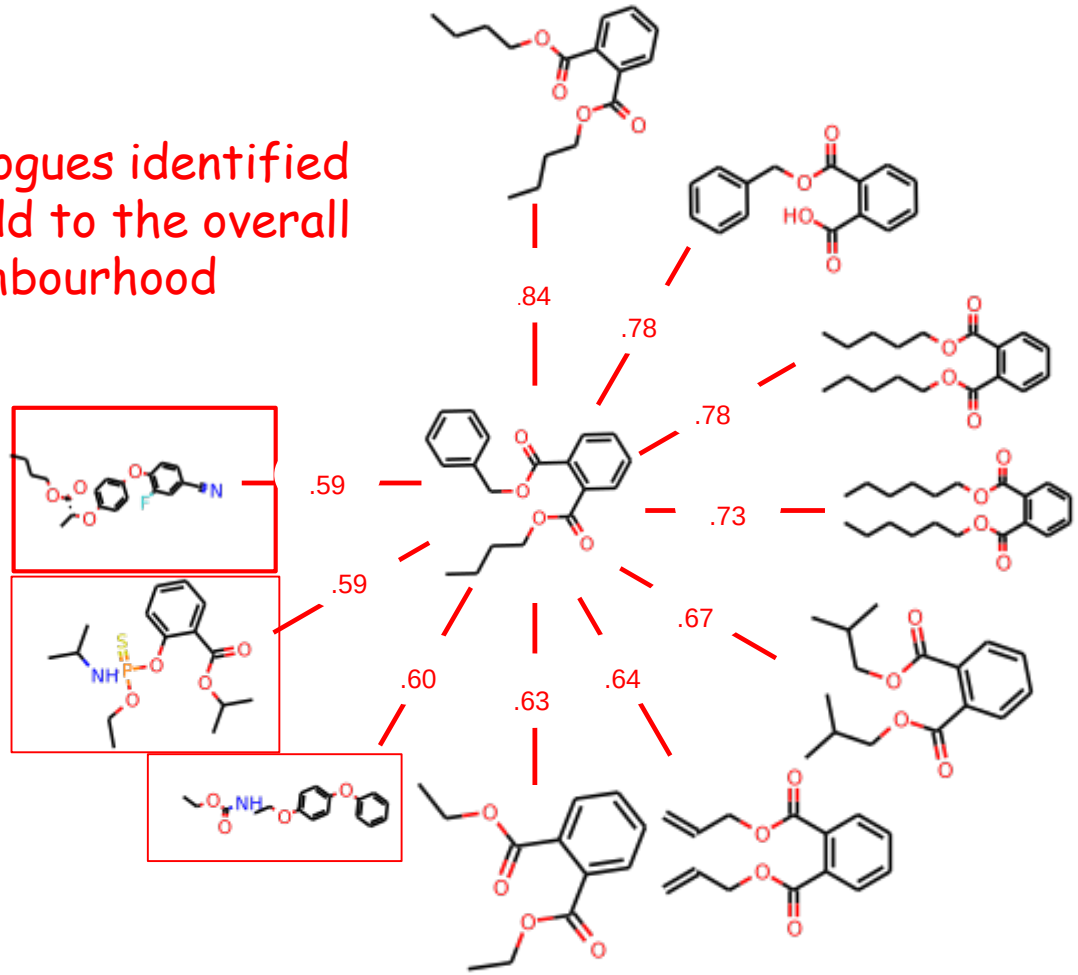
Endpoint
Body Weight
Clinical Chemistry
Food Consumption
Hematology
Kidney
Liver
Mortality
Pancreas
Prostate
Skin
Spleen
Tissue NOS
Urinary Bladder

- Filtering overturns incorrect predictions for 4 endpoints.
- BUT if filtering is too stringent, significant analogues are excluded resulting in a worse performance c.f original GenRA baseline

Case Study: Butyl Benzyl Phthalate

Approach 2: Search Expansion

New
Analogues identified to add to the overall neighbourhood



Endpoint	Baseline Prediction	Structure + Pchem Prediction
Body Weight	.78	.79
Clinical Chemistry	.27	.60
Food Consumption		
Hematology		
Kidney		
Liver		
Mortality		
Pancreas		
Prostate		
Skin		
Spleen		
Tissue NOS		
Urinary Bladder	0	0

- Adding phys-chem to similarity search overturns incorrect predictions for 2 endpoints
- Improves many others

Summary

- Many challenges still remain in read-across
- Quantifying the uncertainty of read-across prediction is a critical issue
- Work in NCCT is focused on systematic and objective approaches to read-across development, evaluation and application
- Established a "baseline" GenRA approach and highlighted progress to implement this approach into a practical tool
- Illustrated the impact on performance that physicochemical similarity can have on analogue identification & evaluation as part of ongoing research

Acknowledgements

- Rusty Thomas
- Imran Shah
- George Helman
- Tony Williams
- Prachi Pradeep
- Richard Judson
- Jeff Edwards
- Chris Grulke