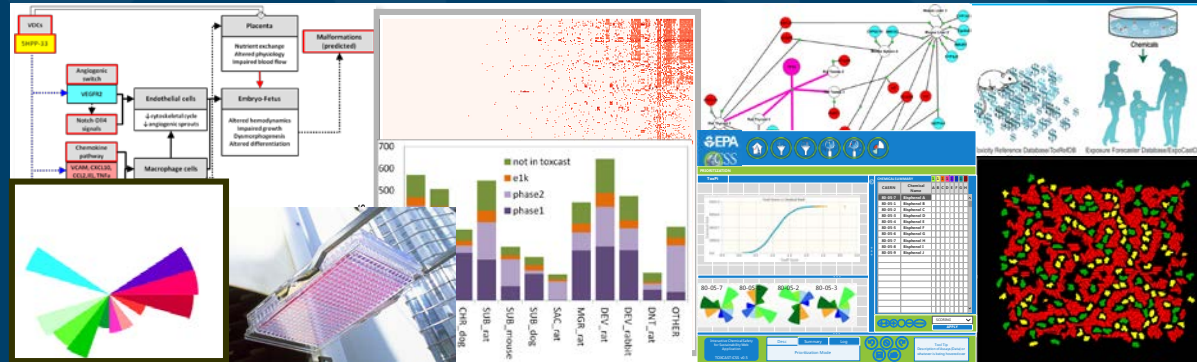


Navigating Through the Minefield of Read-Across Tools and Frameworks: An Update on Generalised Read-Across (GenRA)



Grace Patlewicz, George Helman*, Imran Shah

National Center for Computational Toxicology
*ORISE Fellow

The views expressed in this presentation are those of the authors and do not necessarily reflect the views or policies of the U.S. EPA

Conflict of Interest Statement

No conflict of interest declared.

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. EPA

Outline



- Background and Definitions
- Workflow for category development and read-across
- Current tools and approaches
- Uncertainty assessment in read-across
- Quantifying uncertainties and Assessing Performance of read-across
- From research to implementation
- 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)

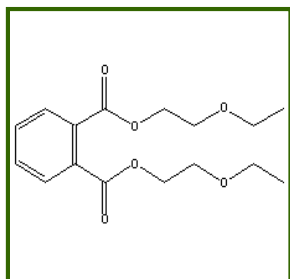
Definition: Read-across

Known information on the property of a substance (**source**) is used to make a prediction of the same property for another substance (**target**) that is considered "similar" i.e. endpoint & often study specific

	Source chemical	Target chemical
Property		

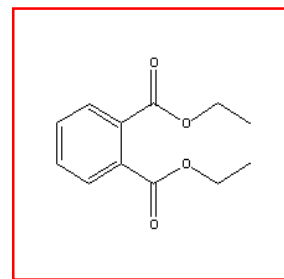
● Reliable data

○ Missing data



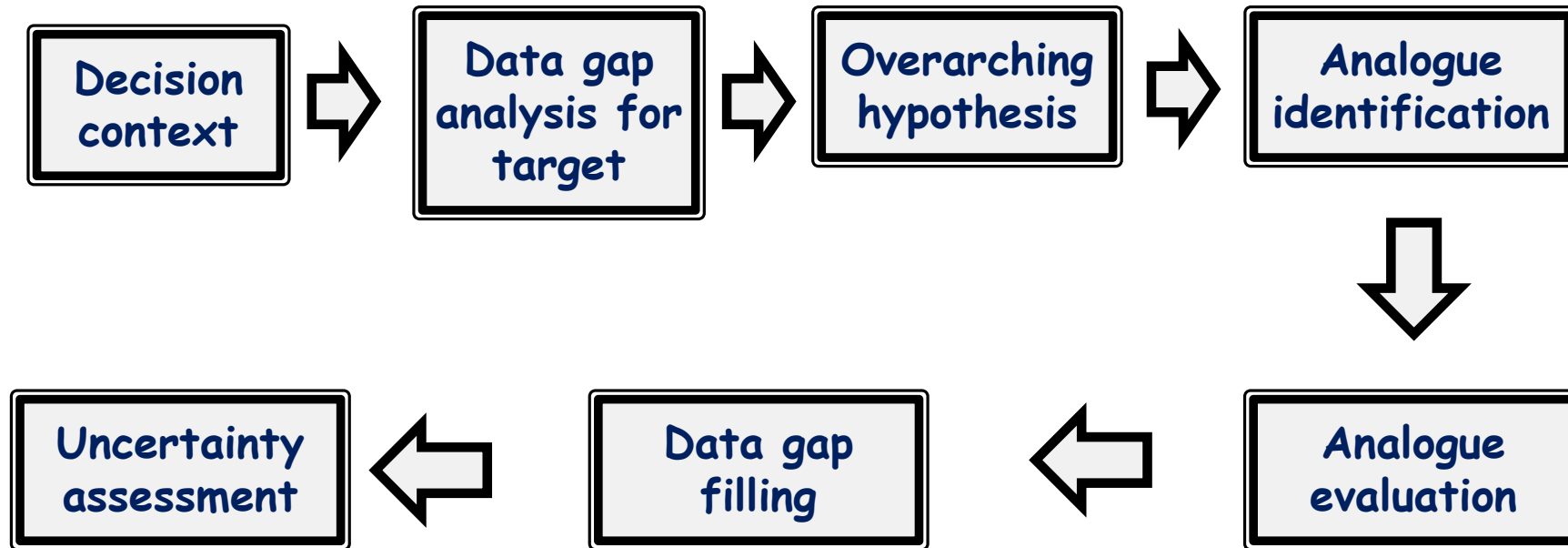
Known to be
harmful

Acute oral
toxicity?

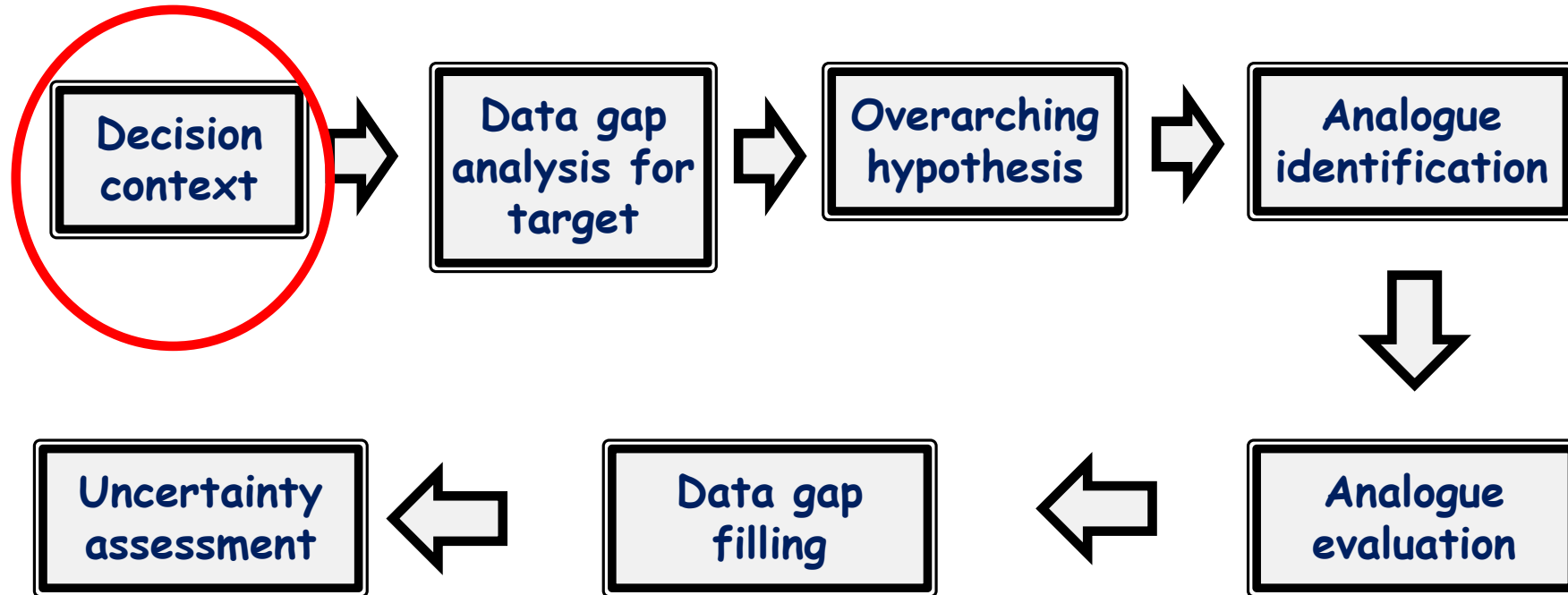


Predicted to be
harmful

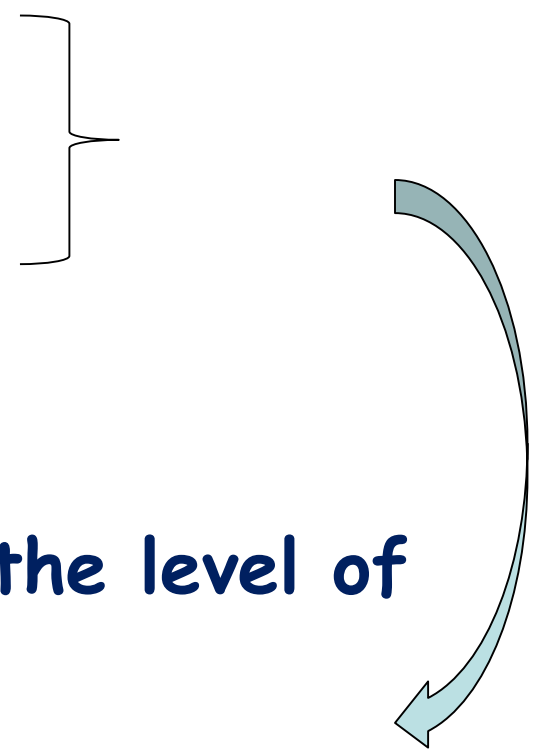
The Category Workflow



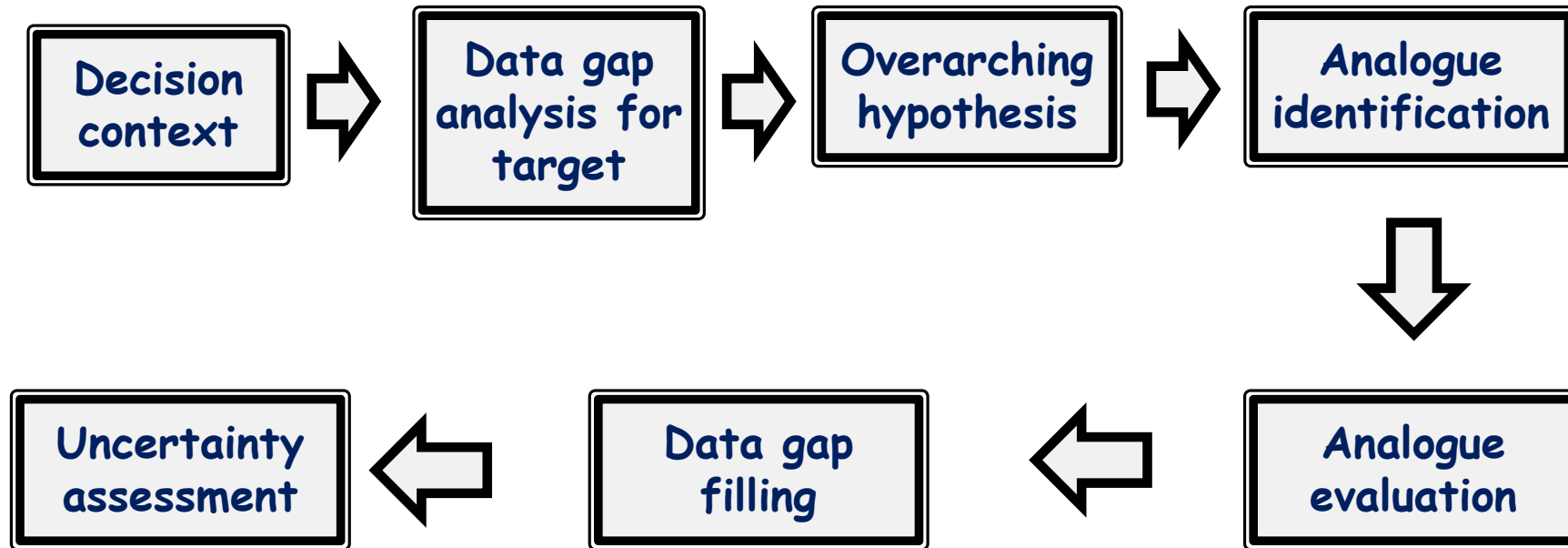
The Category Workflow



Decision Context

- Prioritisation, e.g. PMN
 - Screening level hazard assessment
 - Risk Assessment, e.g. PPRTV
- 
- Different decision contexts will dictate the level of uncertainty that can be tolerated

The 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

QSAR Toolbox 3.4.0.17 [Document]

Source substances

Target

Endpoint specific Similarity rationale

Data gap

Protein binding by OASIS
Protein binding by OECD
Protein binding potency
Supernatants
Toxic hazard classification by Cramer (ext)
Toxic hazard classification by Cramer (orig)
Ultimate biodeg
Biodeg BioHC half-life (Biowin)
Biodeg primary (Biowin 4)

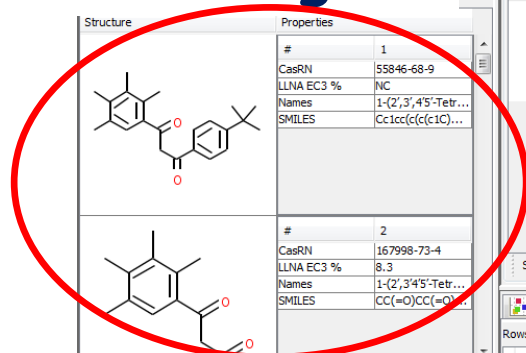
Defined Categories
Document
[481] AN2<AND>AN2 >> Michael addition to

Computational Toxicology

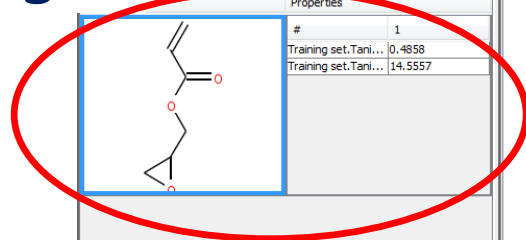
Structure	1	2	3	4	5	6	7	8	9
Immunotoxicity									
Irritation / Corrosion	(101/275)	M: not irritating, moderately irritating, n...	M: not irritating, no...	M: corrosive, corro...		M: irritating, corros...	M: slightly irritating...	M: moderately irrit...	
Neurotoxicity	(10/15)								
Photoinduced Toxicity									
Repeated Dose Toxicity	(69/6204)	M: 300 mg/kg bw/day (nominal), 0.5 mg/L	M: 15 mg/kg bw/d...	M: 10 mg/kg bw/d...		M: 55 mg/kg bw/d...	M: ≥124 mg/kg bw...	M: 20 mg/kg/day, ...	M: 3.33 mg/kg/c
Sensitisation									
Respiratory Tract	(1/1)								
Skin									
In Chemo									
In Vitro	(18/114)					M: 4.55 mg/L, 11.7...	M: <121 mg/L, <1...	M: sensitising, <4...	
In Vivo									
Alternative Methods	(1/1)								
Buehler Test	(5/5)						M: not sensitising		
Combined Intracutaneous and Topical S...	(1/1)								
Pattern	(1/1)								
e Adjuvant Test	(2/2)						M: NOT_SPECIFIED	M: not sensitising	
Lymph Node Assay	(12/14)						M: sensitising		
Sensitisation Test	(8/8)								
Lymph Node Assay	(1/1)						M: not sensitising	M: not sensitising	
Sensitisation Test	(46/64)		M: not sensitising...	M: sensitising		M: NOT_SPECIFIED	M: not sensitising...	M: sensitising	
Sensitisation Test	(4/6)					M: 4E3 µg/cm2, 1...		M: 400 µg/cm2, 1...	
Human Patch Test and Guinea Pig Mag...	(1/1)								
LLNA									
EC3	(20/31)					M: Positive	M: Positive	M: Negative	M: Positive
Maximization Test and Observations of ...	(1/1)								M: sensitising
Miscellaneous	(44/62)						M: Positive, Positiv...	M: Positive, Positive	
Modified Draize Test	(1/1)								
Modified Maximization Test	(1/1)								
Mouse Ear Swelling Test	(4/4)								
Mouse Local Lymphnode Assay (LLNA)									
Skin Sensitisation	(45/4)								
No Data	(1/1)						M: sensitising, NO...	M: sensitising	M: sensitising
Open Epicutaneous Test	(5/5)							M: not sensitising...	

Analogue identification & evaluation within Toxmatch

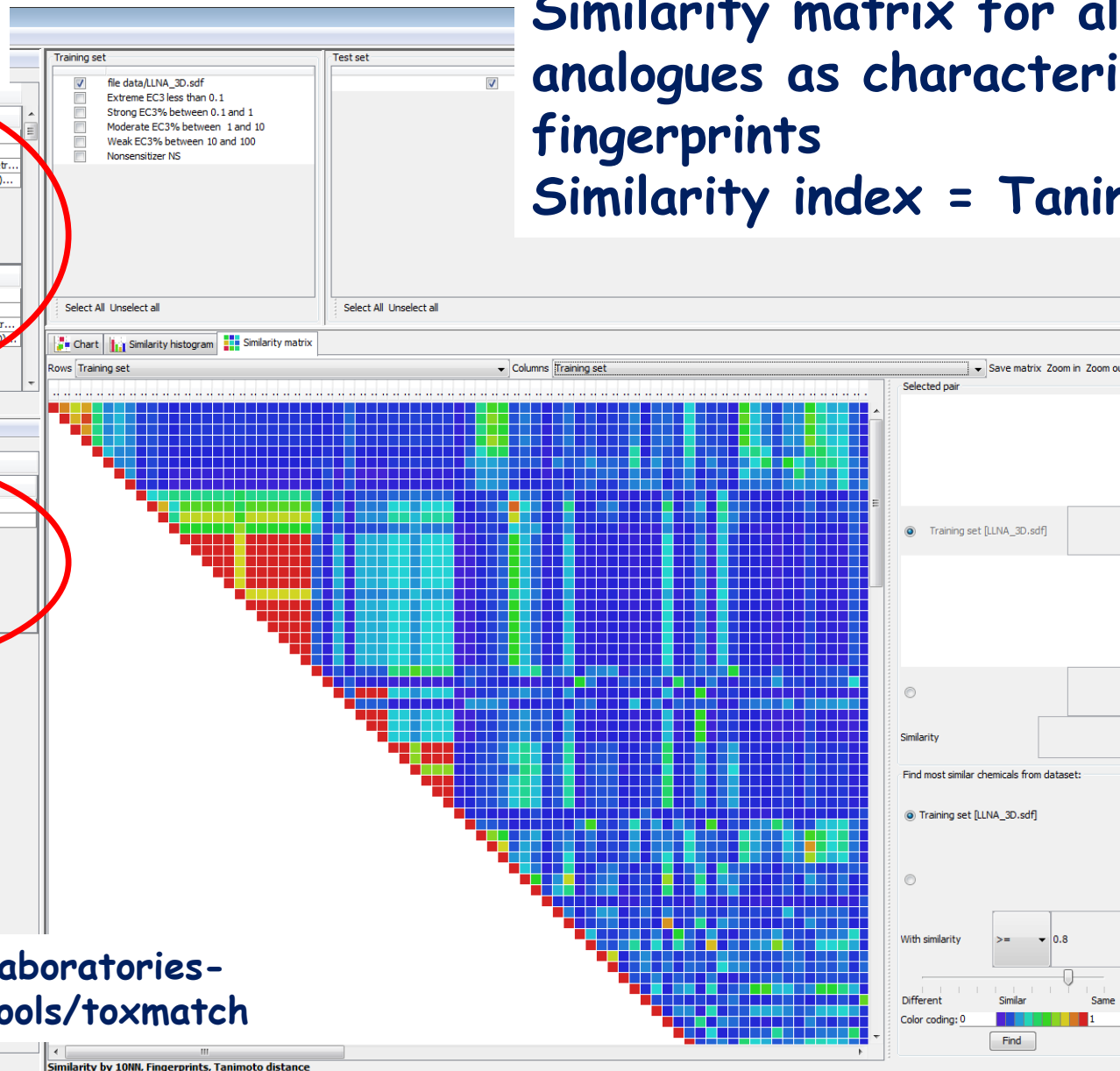
Source analogues



Target



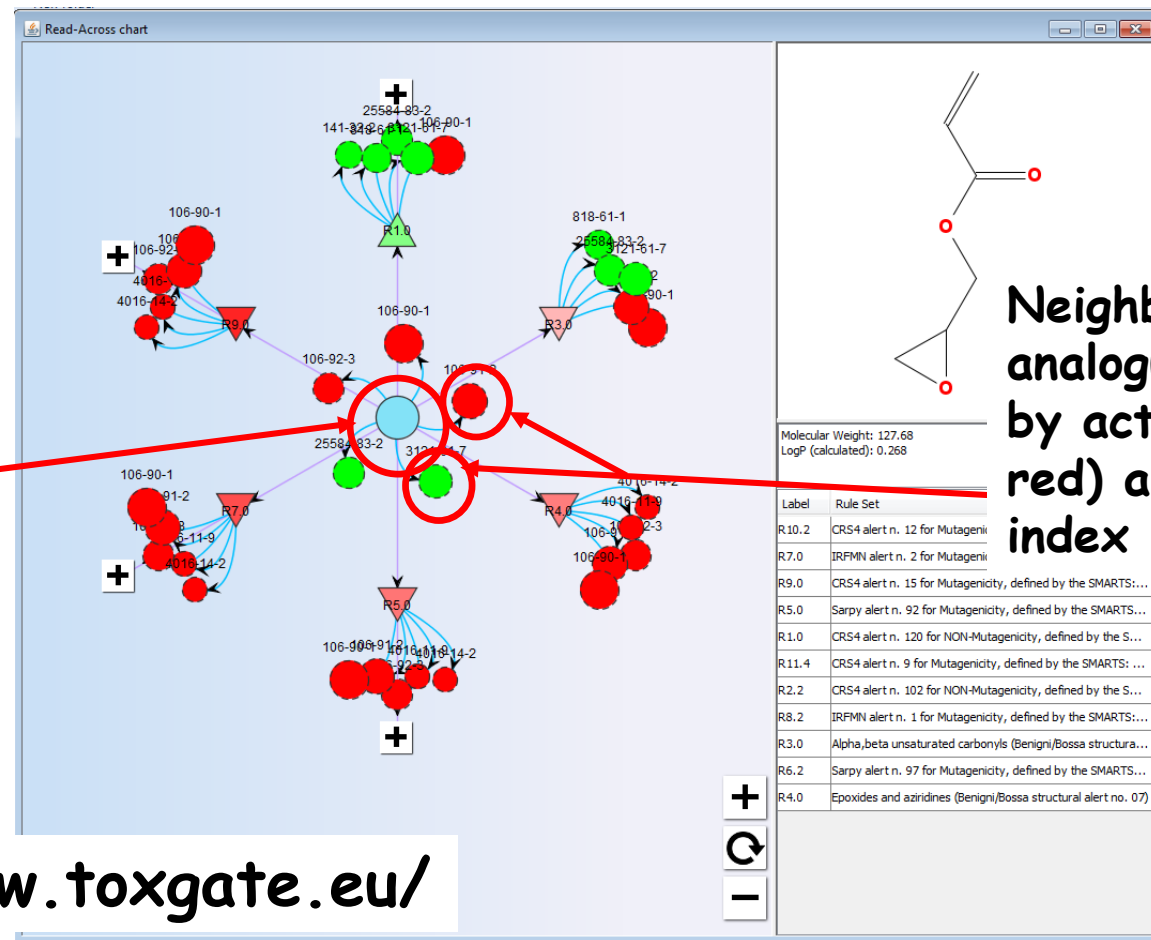
Similarity matrix for all source analogues as characterised by fingerprints
 Similarity index = Tanimoto distance



https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/qsar_tools/toxmatch

ToxRead

Target



Neighboring source analogues, colour coded by activity (positive = red) and by similarity index

<http://www.toxgate.eu/>

Selected Read-Across Tools – Review paper

Computational Toxicology 3 (2017) 1–18



ELSEVIER

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

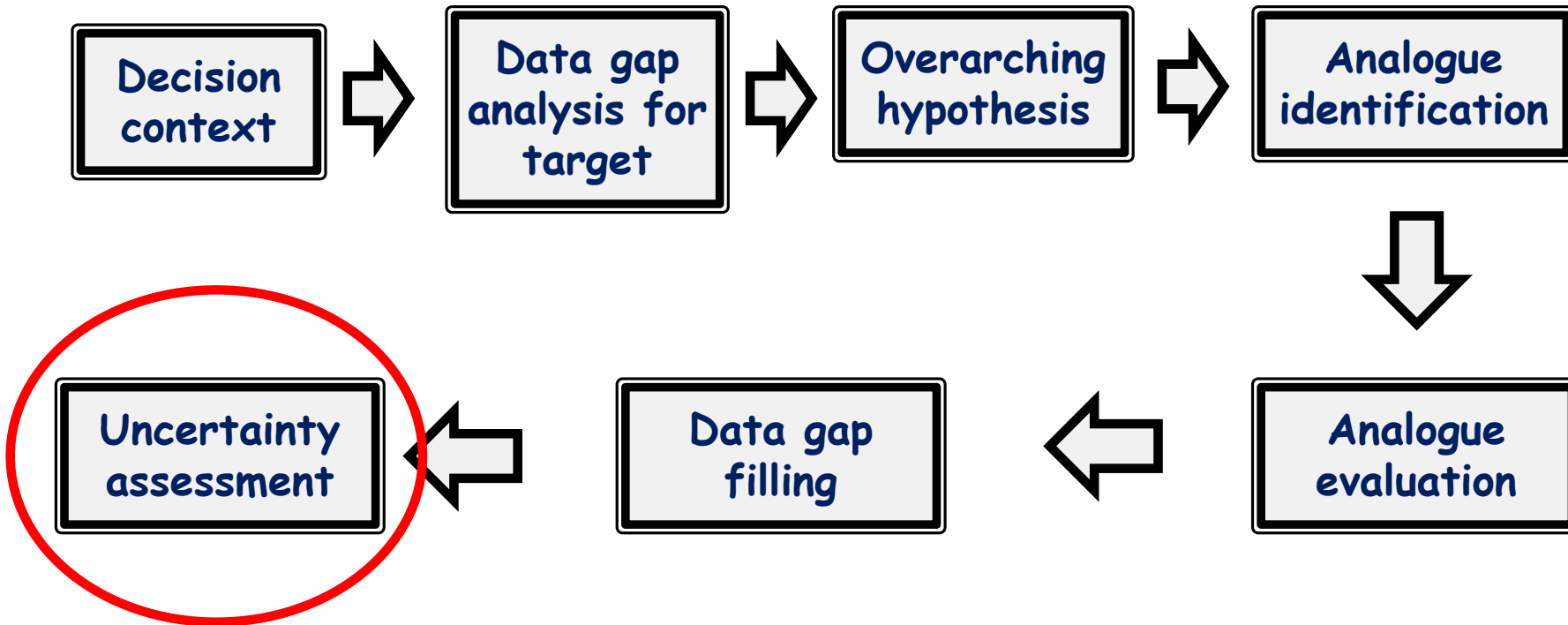
Nearest neighbor

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.

The Category Workflow



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

Sources of Uncertainty (cont'd)

- Overarching hypothesis/similarity rationale - how to identify similar analogues and justify their similarity for the endpoint of interest
- Address the dissimilarities and whether these are significant from a toxicological standpoint e.g. ToxDelta
- Presence vs. absence of toxicity
- Toxicokinetics

Uncertainty Assessment

- A number of publications exist that can guide the construction and assessment of categories and use of read-across
 - Guidance and examples (OECD (2014), ECHA (2008), ECETOC (2012))
 - Frameworks for identifying analogues (e.g., Wu et al (2010), Patlewicz et al (2013))
 - Frameworks for assessing read-across (Blackburn and Stuard (2014), Patlewicz et al (2014), Patlewicz et al (2015), ECHA - RAAF (2015), Schultz et al (2015), Ball et al (2016))

Uncertainty assessment

- However read-across acceptance relies on a subjective expert assessment
- There is no objective measure of read-across performance
- Different approaches have been explored to characterise uncertainties both qualitatively and quantitatively
- E.g. Blackburn and Stuard (qualitative), Molecular Networks (quantitative), EPA NCCT (quantitative and generalisable)

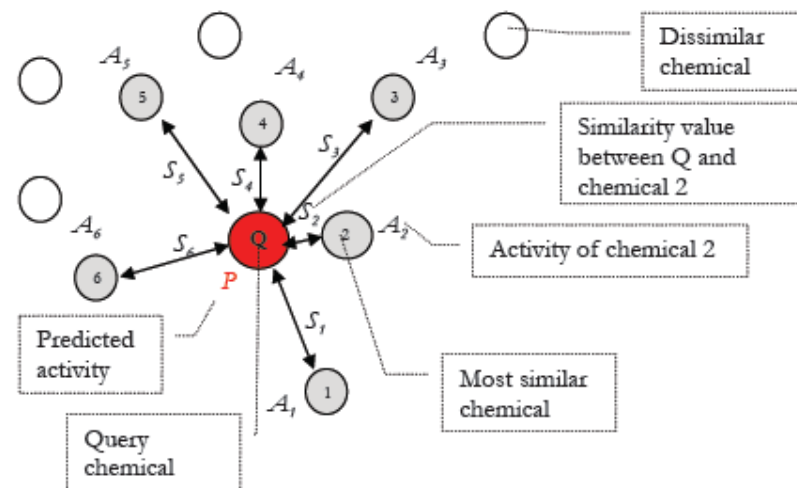
Quantifying Uncertainty & Assessing Performance of Read-Across

- GenRA (Generalised Read-Across) is a “local validity” approach
- Predicting toxicity as a similarity-weighted activity of nearest neighbors based on chemistry and bioactivity descriptors
- Generalised version of Chemical-Biological Read-Across (CBRA) developed by Low et al (2013)
- Systematically evaluates 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:

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



GenRA - Approach

I. Data

1,778 Chemicals
3,239 Structure descriptors (chm)
820 Bioactivity assays (bio)
ToxCast
574 Apical outcomes (tox)
ToxRefDB

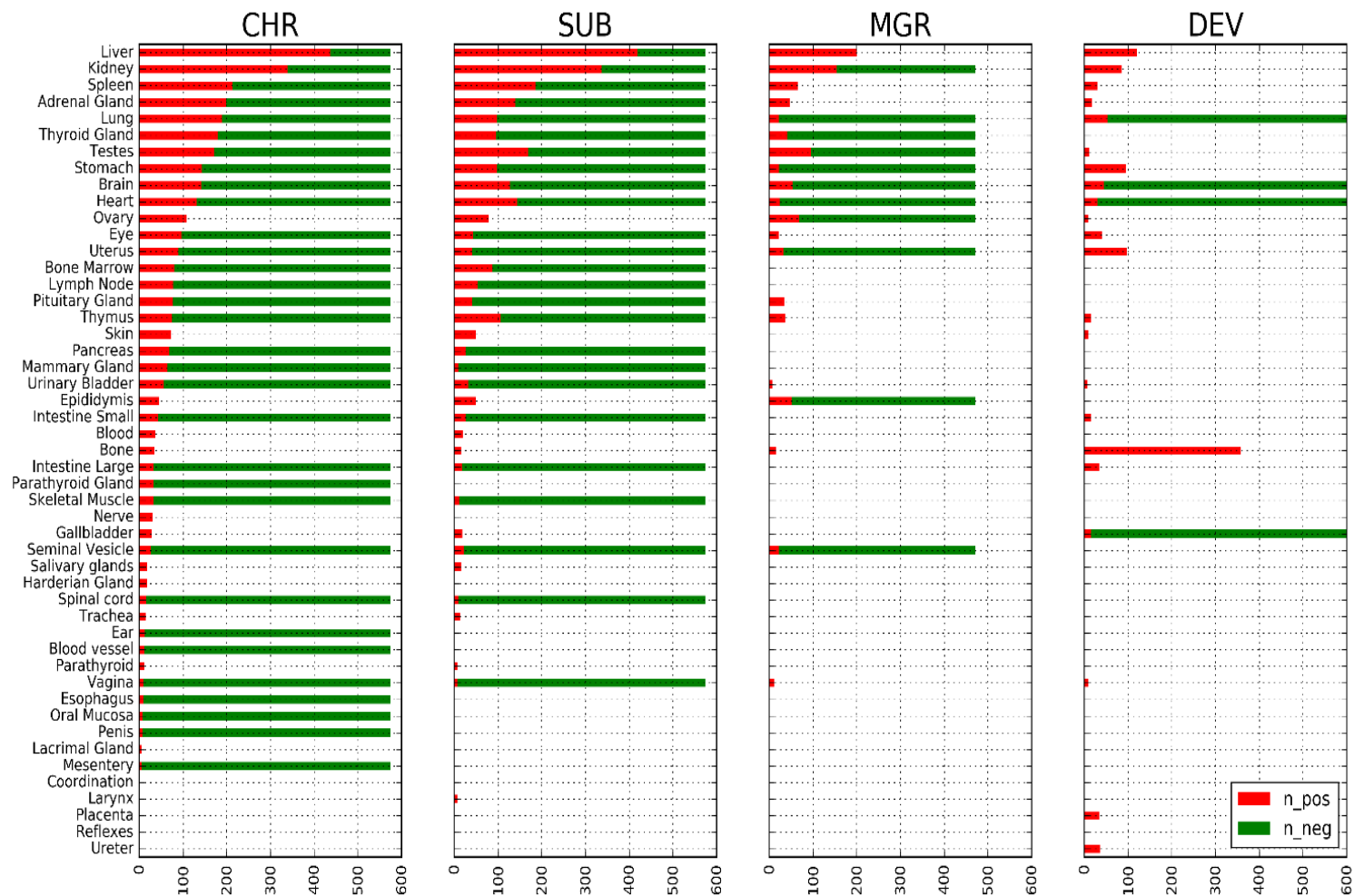
II. Define Local neighborhoods

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

III. GenRA

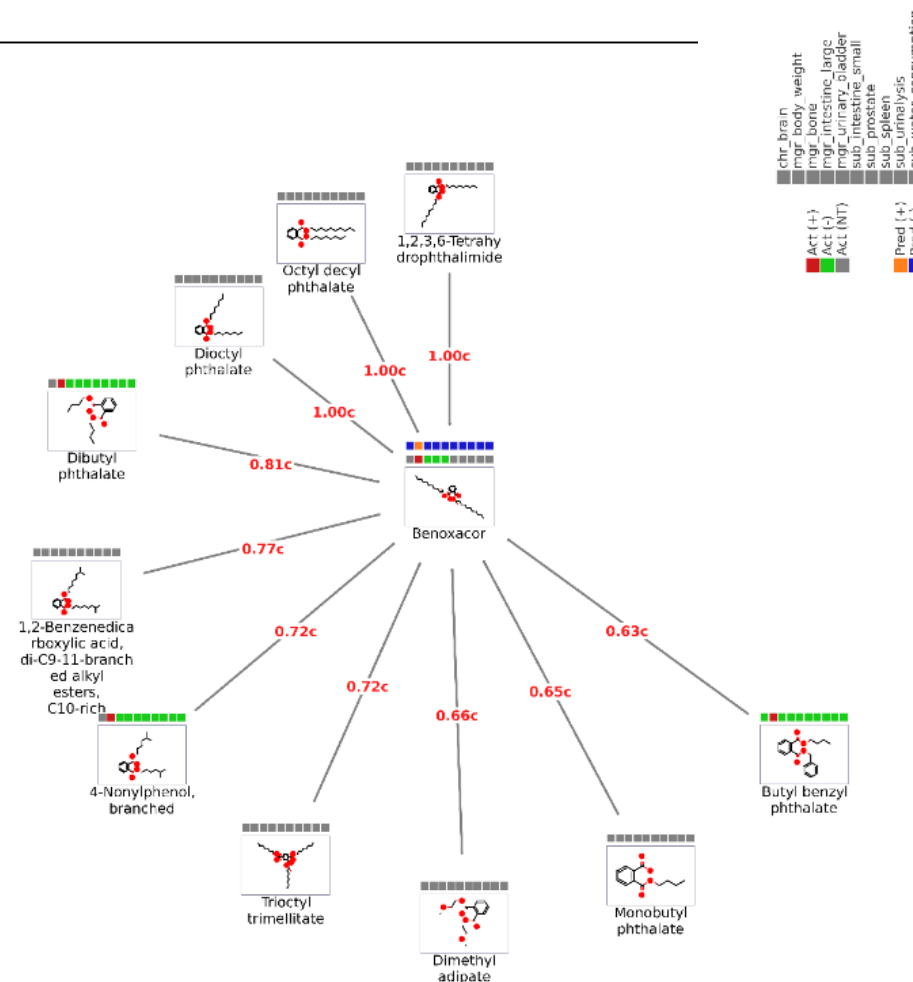
Use GenRA to predict apical outcomes in local neighborhoods
Evaluate impact descriptors (chm, bio, bc) on prediction
Quantify uncertainty

GenRA - Toxicity Data from ToxRefDB

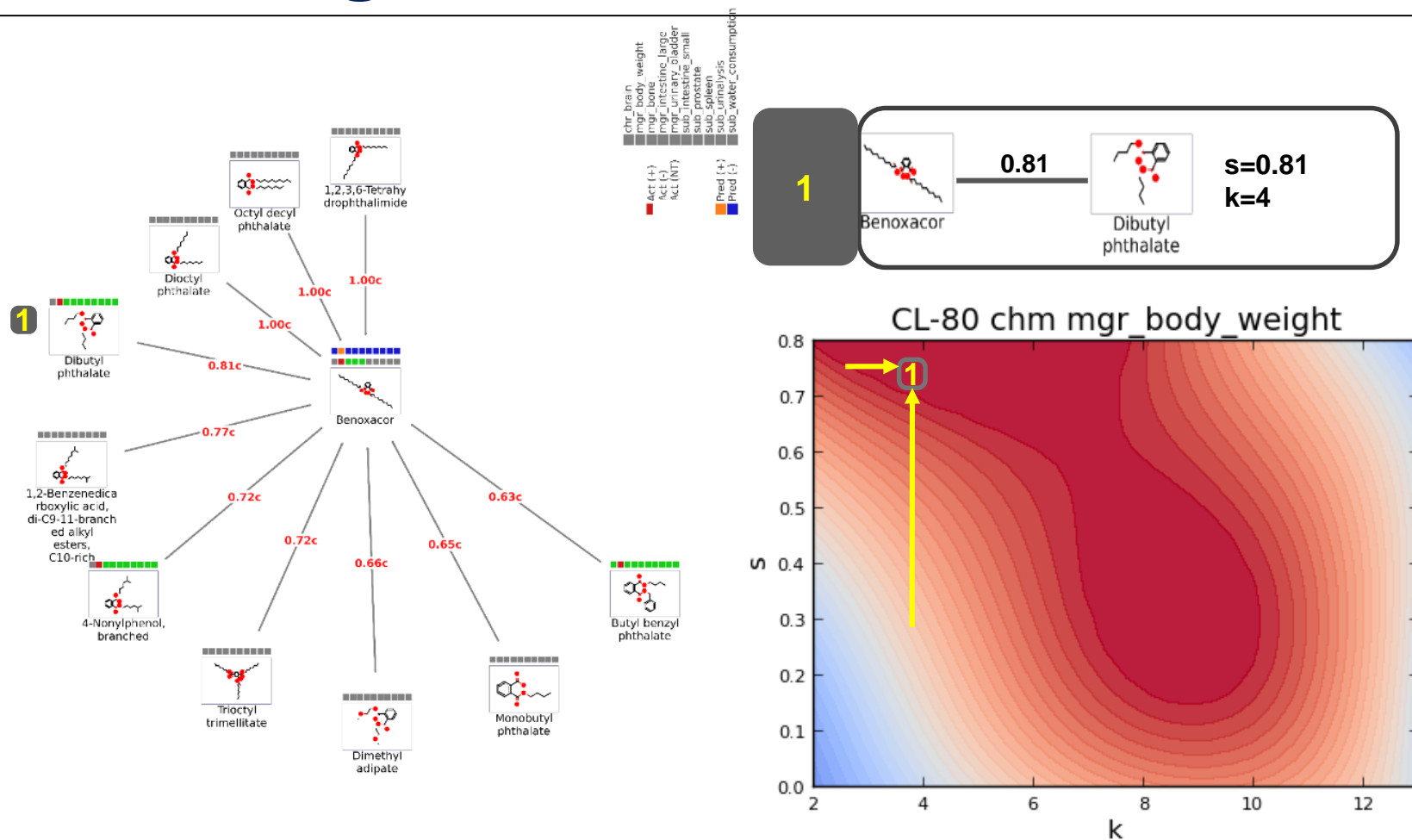


GenRA – Performance in Each Cluster

- Use GenRA to predict the similarity weighted toxicity scores for each
 - Toxicity type (β)
 - Descriptor = {chm, bio, bc} (α)
 - No. of nearest neighbors (k)
 - Similarity score threshold (s_{ij}^{α})
- Calculate performance by comparing predicted y^{tox} and true x^{tox} for all chemicals using area under ROC curve (AUC)
- Results: {cluster, α , β , k , s , AUC}



GenRA - Analysing Local Neighborhood of a Chemical




GenRA – Insights and Next Steps

- The approach enabled a performance baseline for read-across predictions of 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. TK similarity, metabolic similarity, reactivity similarity...

From research to implementation: GenRA prototype

- Intent is to integrate objective read-across functionality as part of ongoing dashboard efforts see <https://comptox.epa.gov/dashboard>
- A limited release of GenRA is currently undergoing internal beta testing
- A video tutorial and help manual has been created to explain the approach and how to use the tool

From research to implementation


United States
Environmental Protection
Agency
Home
Advanced Search
Batch Search
Lists

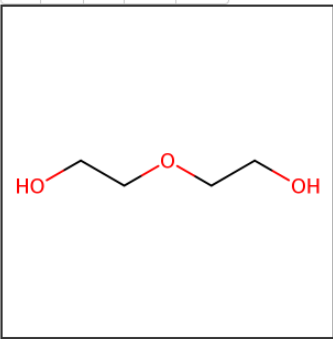
Search Chemistry Dashboard

Chemistry Dashboard
Submit Comment
Share
Copy
Aa
Aa
Aa

Diethylene glycol

111-46-6 | DTXSID8020462

Ⓢ Searched by Approved Name: Found 1 result for 'Diethylene glycol'.



Wikipedia

Diethylene glycol (DEG) is an organic compound with the formula (HOCH2CH2)2O. It is a colorless, practically odorless, poisonous, and hygroscopic liquid with a sweetish taste. It is miscible in water, alcohol, ether, acetone, and ethylene glycol. DEG is a widely used solvent. It can be a contaminant in consumer products; this has resulted in numerous epidemics of poisoning since the early 20th century.... [Read more](#)

Intrinsic Properties

Structural Identifiers

Related Compounds (Beta)

Presence in Lists

Record Information

Chemical Properties
Env. Fate/Transport
Synonyms
External Links
Toxicity Values (Beta)
Exposure
Bioassays
Similar Molecules (Beta)
Literature
Comments

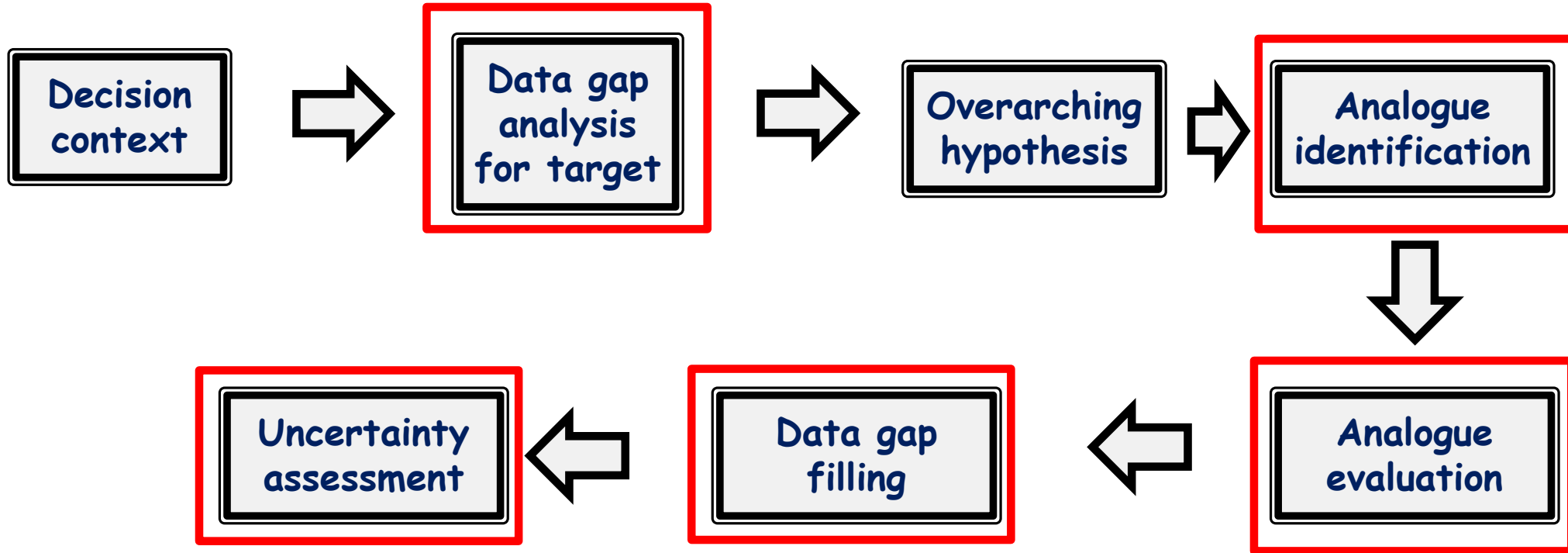
Summary

LogP: Octanol-Water
Water Solubility
Density
Melting Point
Boiling Point

Download as:
TSV
Excel
SDF

Property	Average		Median		Range		Unit
	Experimental	Predicted	Experimental	Predicted	Experimental	Predicted	
LogP: Octanol-Water	-	-1.24 (4)	-	-1.24	-	-1.47 to -0.941	-
Water Solubility	9.42 (1)	10.9 (3)	9.42	10.9	9.42	8.06 to 15.2	mol/L
Density	-	1.11 (1)	-	1.11	-	-	g/cm³
Melting Point	-10.2 (5)	-1.09 (3)	-10.0	-1.09	-10.4 to -10.0	-13.2 to 9.00	°C

GenRA prototype development



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 availabl e	X	X	X For Ames & BCF	NA
Data gap analysis	NA	X	X Data matrix can be exporte d	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

Working interface

GenRA (Beta) Chemical Properties Synonyms External Links Env. Fate/Transport Toxicity Values (Beta) Bioassays Exposure Literature Similar Molecules (Beta) Comments

NN By: K: Sel by:

Summary:

Grp: By:

Chemical structure diagram showing 2-Methoxyethanol at the center, connected to various related compounds including Triethylene glycol, 2-Butoxyethanol, Ethylene glycol, 1,2-Propylene glycol, N,N-Diethylethanol, 2-Methyl-1-propanol, 2-Chloroethanol, Dimethylaminoethanol, Isopentyl alcohol, and 2-(Hexyloxy)ethanol.

Summary table (tox_txrf):

	tox_txrf	chm_mrgn	bio_txrf	bio_txct
2-Methoxyethanol	6	22		
Triethylene glycol				
2-Butoxyethanol				
Ethylene glycol				
2-(Hexyloxy)ethanol				
Isopentyl alcohol				
Dimethylaminoethanol				
2-Chloroethanol				
2-Methyl-1-propanol				
N,N-Diethylethanol				
1,2-Propylene glycol				

Read-across table:

	1,2-Propylene glycol	N,N-Diethylethanol	2-Methyl-1-propanol	2-Chloroethanol	Dimethylaminoethanol	Isopentyl alcohol	2-(Hexyloxy)ethanol	Ethylene glycol	2-Butoxyethanol	Triethylene glycol	2-Methoxyethanol
DEV:Body Weight											
DEV:Bone											
SUB:Clinical Signs											
SUB:Kidney											
SUB:Liver											
SUB:Mortality											
DEV:Mortality											
DEV:Uterus											
DEV:Kidney											
DEV:Food Consumption											
DEV:Clinical Signs											

Run GenRA Min+: Min-: Filter by: Sim wt ☐ Export

Grid interface where windows are dynamically updated in subsequent windows

GenRA (Beta) Chemical Properties Synonyms External Links Env. Fate/Transport Toxicity Values (Beta) Bioassays Exposure Literature Similar Molecules (Beta) Comments

NN By: **chm_mrgn** K: 10 Sel: **tox_txrf** Summary: Grp: **tox_txrf** By: **tox_fp** Read-across

Similarity context

Chemical structure similarity graph showing 2-Methoxyethanol at the center, surrounded by related compounds like Triethylene glycol, 2-Butoxyethanol, Ethylene glycol, 1,2-Propylene glycol, N,N-Diethylethanolamine, 2-(Hexyloxy)ethanol, Isopentyl alcohol, 2-Methyl-1-propanol, 2-Chloroethanol, and Dimethylaminoethanol.

Heatmap data (Left):

	tox_txrf	chm_cf	bio_txrf	bio_txrf
2-Methoxyethanol	6	22		
Triethylene glycol				
2-Butoxyethanol				
Ethylene glycol				
2-(Hexyloxy)ethanol				
Isopentyl alcohol				
Dimethylaminoethanol				
2-Chloroethanol				
2-Methyl-1-propanol				
N,N-Diethylethanolamine				
1,2-Propylene glycol				

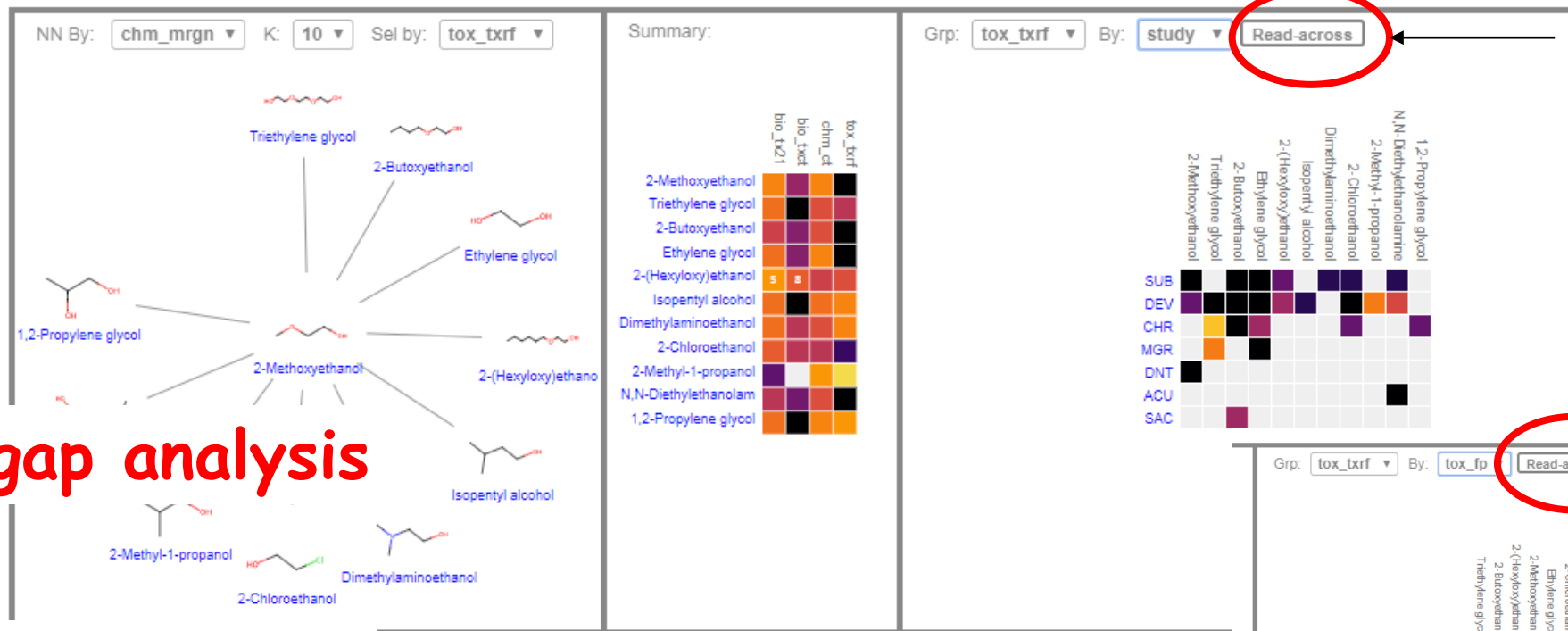
Heatmap data (Right):

	1,2-Propylene glycol	N,N-Diethylethanolamine	2-Methyl-1-propanol	2-Chloroethanol	Dimethylaminoethanol	Isopentyl alcohol	2-(Hexyloxy)ethanol	Ethylene glycol	2-Butoxyethanol	Triethylene glycol	2-Methoxyethanol
DEV:Body Weight											
DEV:Bone											
SUB:Clinical Signs											
SUB:Kidney											
SUB:Liver											
SUB:Mortality											
DEV:Mortality											
DEV:Uterus											
DEV:Kidney											
DEV:Food Consumption											
DEV:Clinical Signs											

Run GenRA Min+: 0 Min-: 0 Filter by: Enter text Sim wt Export

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

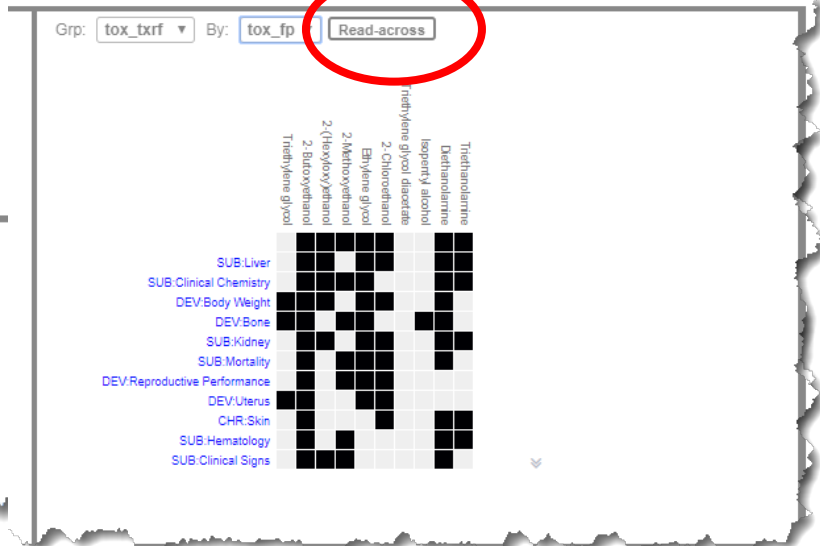
To initiate data matrix view



Data gap analysis

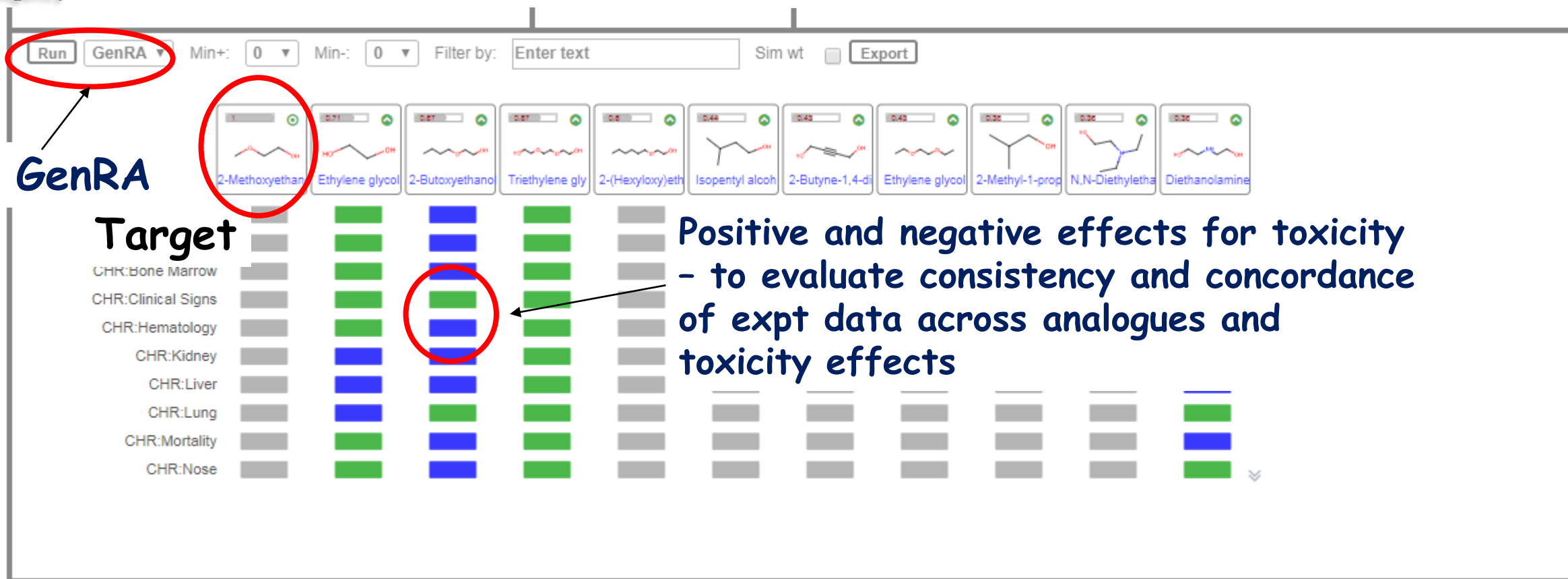
View data quantity by type

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

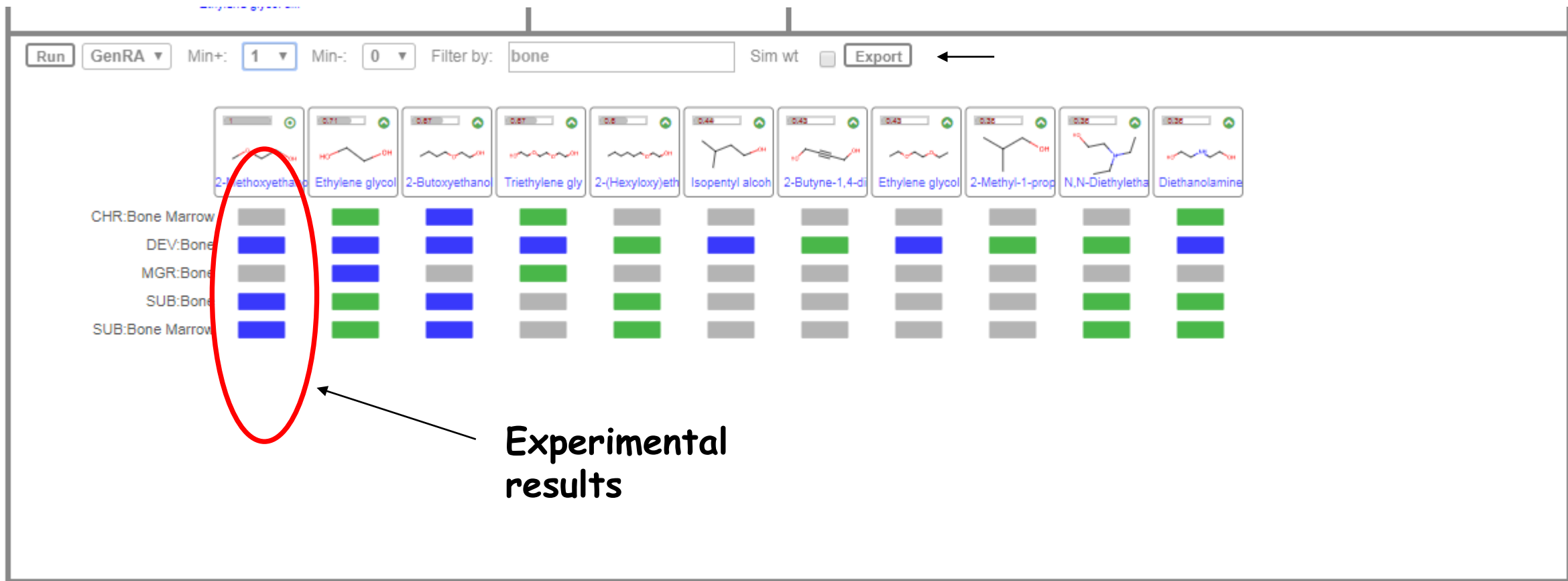


Analogue evaluation using data matrix view

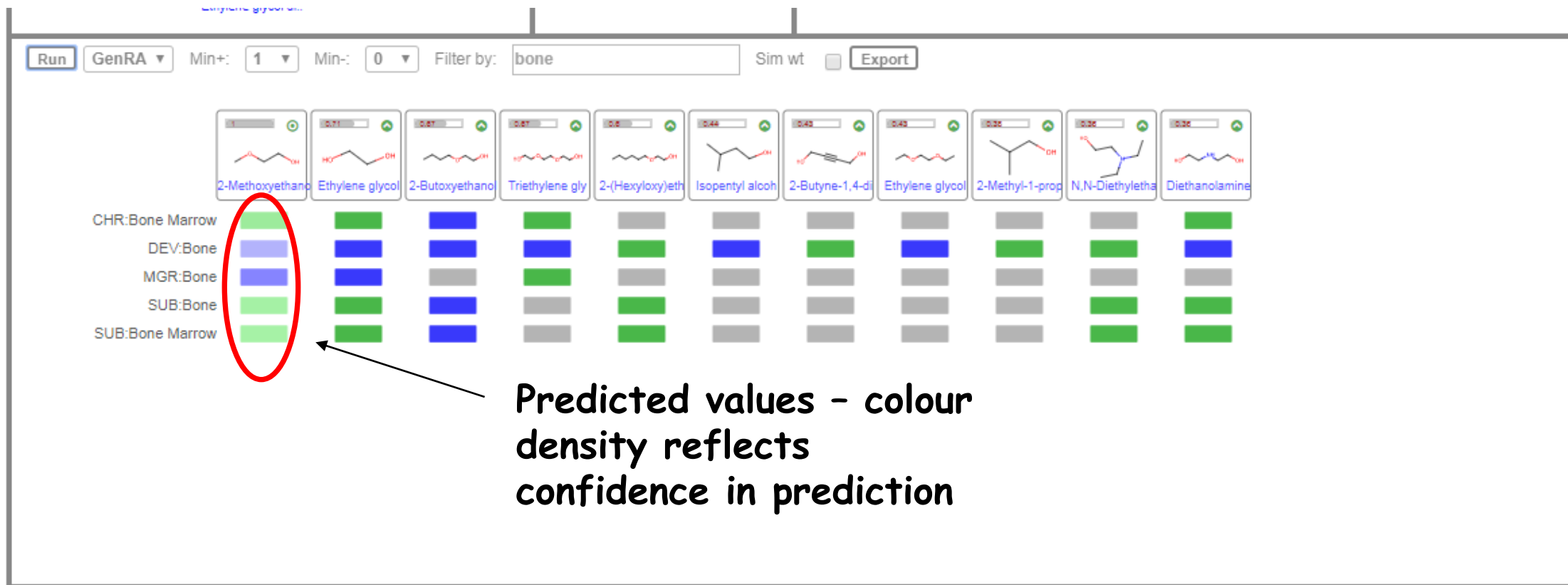
Run GenRA



Data gap filling using GenRA within data matrix



Data gap filling using GenRA within data matrix



Exported results using GenRA

[illegible]

Demo

Summary

- Still many challenges remain in read-across – what information is relevant to integrate and ways in which that integration can be performed
- Quantifying the uncertainty of read-across prediction is a critical issue
- Have illustrated the research directions being taken within NCCT and work to implement these into practical tools

Acknowledgements

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