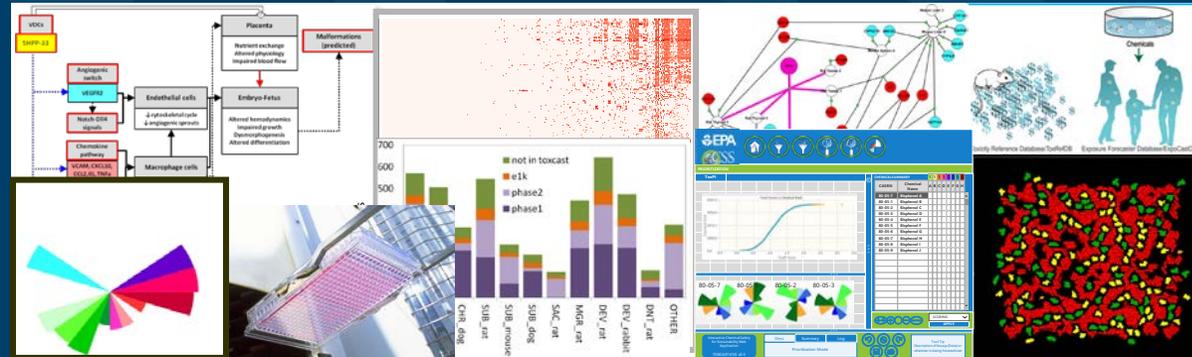


Actualizing research into practical tools: A case study of GenRA, a new read-across tool



George Helman^{1,2}, Imran Shah², Grace Patlewicz²

¹Oak Ridge Institute for Science and Education (ORISE), Oak Ridge, TN, USA

²National Center for Computational Toxicology, US EPA, RTP, NC, USA

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

- Definitions
- Introduction to the Generalized Read-Across (GenRA) approach
- Current progress on web tool development
- Future work on web tool development
- Conclusions

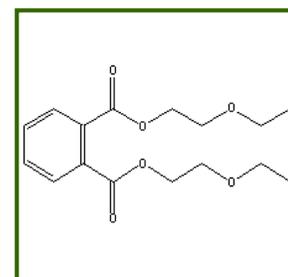
Definitions: Read-across

- Read-across describes the method of filling a data gap whereby a chemical with existing data values is used to make a prediction for a ‘similar’ chemical.
- A target chemical is a chemical which has a data gap that needs to be filled i.e. the subject of the read-across.
- A source analogue is a chemical that has been identified as an appropriate chemical for use in a read-across based on similarity to the target chemical and existence of relevant data.

	Source chemical	Target chemical
Property	●	○

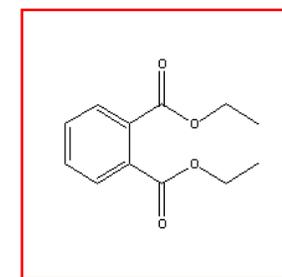
● Reliable data

○ Missing data



Known to be harmful

Acute toxicity?



Predicted to be harmful

GenRA - Introduction

- GenRA (Generalized Read-Across) is a “local validity” approach predicting toxicity as a similarity-weighted activity of source analogues based on chemistry and/or bioactivity descriptors. (Shah et al, 2016)
- Generalized version of Chemical-Biological Read-Across (CBRA) developed by Low et al (2013)
- **Goal:** to establish an objective performance baseline for read-across and quantify the uncertainty in the predictions made.
- Tested and compared chemical, bioactivity (hitcalls from ToxCast) descriptors as well as a hybrid of the two.
- Underlying data used was taken from ToxRefDB, a collection of repeated dose toxicity study types e.g. chronic, multigeneration, developmental, subchronic etc
- Toxicity effects within those study types were recorded as binary outcomes

Algorithm

1. Select chemical/biological fingerprint to characterize target and source analogues. The Jaccard similarity metric is used to quantify the pairwise similarity
2. Choose the number of source analogues, k (where k = the number of nearest neighbors)
3. Filter on the basis of *in vivo* toxicity data for source analogues
4. Read-across prediction for target chemical is the outcome of a similarity-weighted activity calculation from the nearest neighbors
5. Perform Receiver Operating Characteristic (ROC) to identify the threshold where the best performance is produced for a given s and k.

Jaccard similarity:

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

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

Results of GenRA

Study	Cases*	chm	bio	bc	chm > bio bc	bio > chm bc	bc > chm bio	Any
chr	1115	97 (9%)	133 (12%)	67 (6%)	75 (7%)	107 (10%)	21 (2%)	203 (18%)
dev	712	63 (9%)	138 (19%)	50 (7%)	43 (6%)	117 (16%)	7 (1%)	167 (23%)
dnt	84	16 (19%)	26 (31%)	15 (18%)	5 (6%)	20 (24%)	2 (2%)	27 (32%)
mgr	668	78 (12%)	126 (19%)	56 (8%)	46 (7%)	106 (16%)	4 (1%)	156 (23%)
rep	34	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
sac	128	2 (2%)	7 (5%)	6 (5%)	1 (1%)	6 (5%)	2 (2%)	9 (7%)
sub	922	107 (12%)	140 (15%)	78 (8%)	64 (7%)	122 (13%)	21 (2%)	212 (23%)
All	3663	363 (10%)	570 (16%)	272 (7%)	234 (6%)	478 (13%)	57 (2%)	774 (21%)

*cluster-study-effect combination

Tabulation of high-performing scores (p<0.01)

Next steps - GenRA

- Ongoing analysis:

- Consider

physicoc

similarity

- Potency

- Integratio

https://cc

Chemistry Dashboard | GENRA

© Searched by Approved Name: Found 1 result for '2-methoxyethanol'.

Wikipedia
2-Methoxyethanol, or methyl cellosolve, is an organic compound with formula C₃H₈O₂ that is used mainly as a solvent. It is a clear, colorless liquid with an ether-like odor. It is in a...[Read more](#)

Intrinsic Properties

Structural Identifiers

Linked Substances

Presence in Lists

Record Information

Quality Control Notes

GenRA (Beta) | Chemical Properties | Synonyms | Bioassays | Exposure | Analytical | Comments | Similar Compounds | Hazard | Links | ADME (Beta) | Executive Summary (Beta)

NN By: chm_mrgn | K: 10 | Sel by: tox_txf | Summary: | Grp: tox_txf | By: tox_fp | Read-across

Triethylene glycol

2-Butoxyethanol

2-Methoxyethanol

Diethylene glycol mo

1,2-Propanediol

N,N-Diethylethanamine

2-Methyl-1-propanol

2-Chloroethanol

Dimethylamine

Isopentyl alcohol

2-Hydroxyethyl alcohol

Ethylene glycol

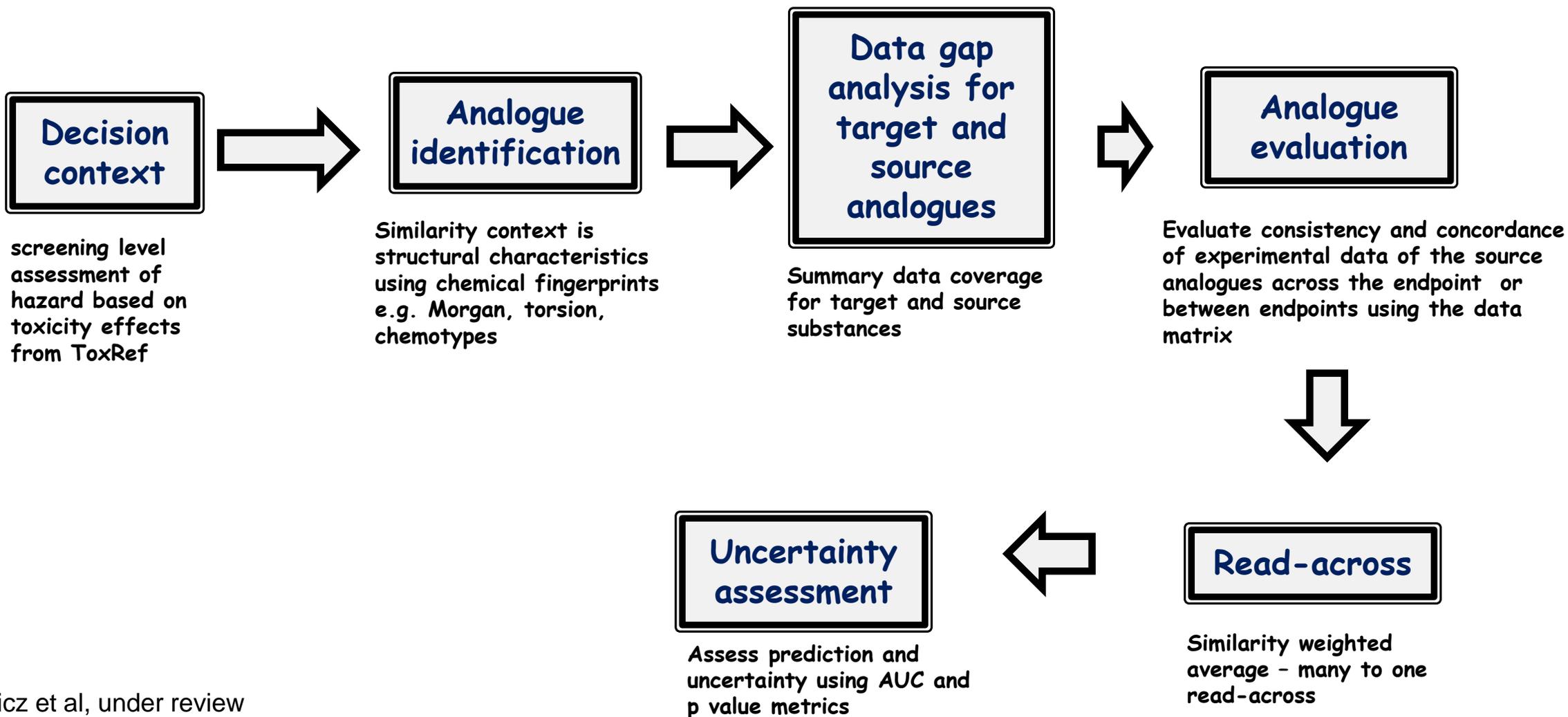
2-Ethoxyethanol

Triethylene glycol

2-Methoxyethanol

e

Category Workflow and its alignment to GenRA



GenRA: Web Tool Development

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

Frame 1

Frame 2

	tox_bxf	chm_of	bio_tox	bio_tox1
2-Methoxyethanol	Black	Black	Black	Black
Diethylene glycol mo	Black	Black	Black	Black
2,5,8,11-Tetraoxatri	Black	Black	Black	Black
2,5,8,11,14-Pentaoxa	Black	Black	Black	Black
2-[2-(2-Methoxyethox	Black	Black	Black	Black
Ethylene glycol dime	Black	Black	Black	Black
Pentaethylene glycol	Black	Black	Black	Black
Tetraethylene glycol	Black	Black	Black	Black
2-Ethoxyethanol	Black	Black	Black	Black
Diethylene glycol	Black	Black	Black	Black
2-Isopropoxyethanol	Black	Black	Black	Black

Frame 3

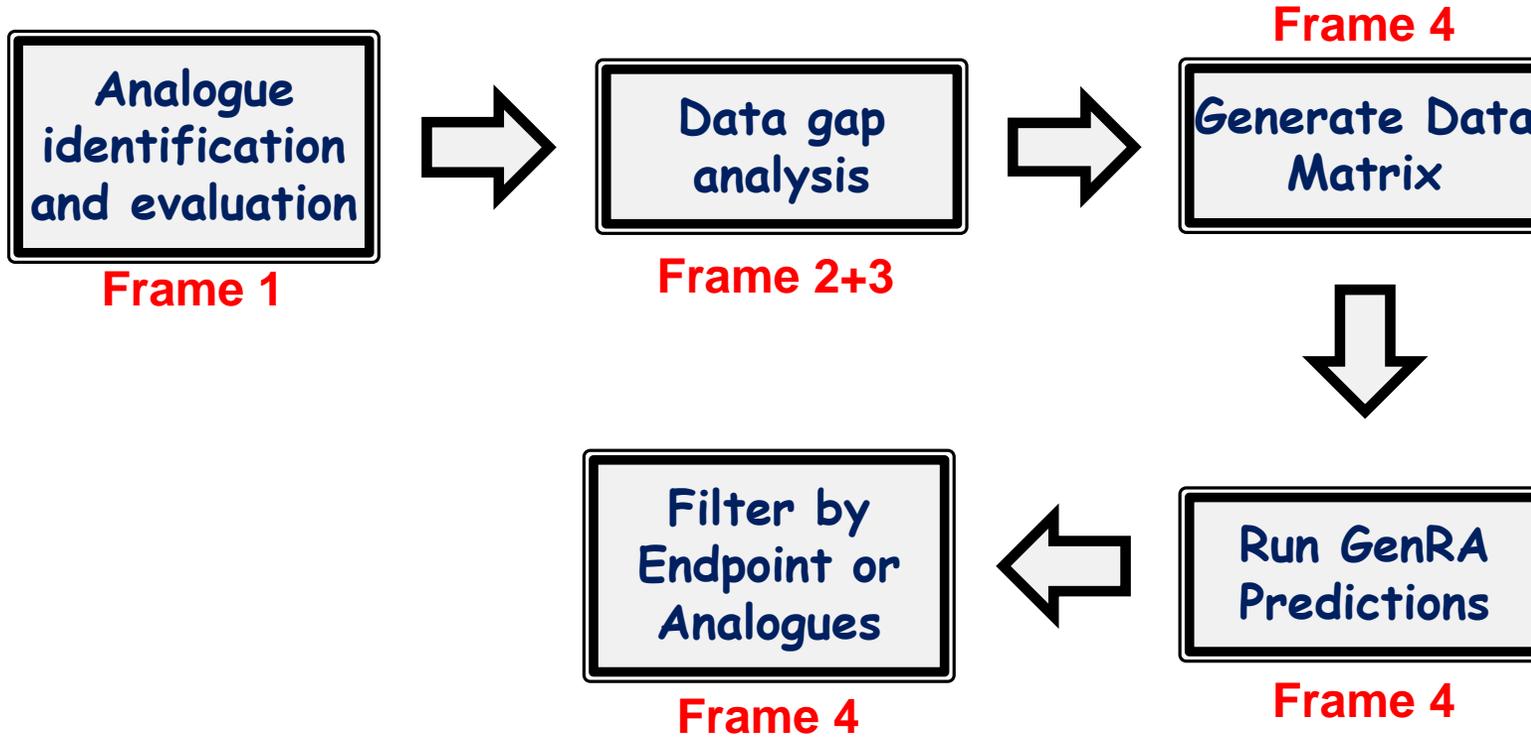
	1,2-Propylene glycol	N,N-Diethylethanolamine	2-Methyl-1-propanol	2-Chloroethanol	Dimethylaminoethanol	Isopentyl alcohol	2-(Hexyloxy)ethanol	Ethylene glycol	2-(Hexyloxy)ethanol	Triethylene glycol	2-Methoxyethanol
DEV:Body Weight	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black
DEV:Bone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

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

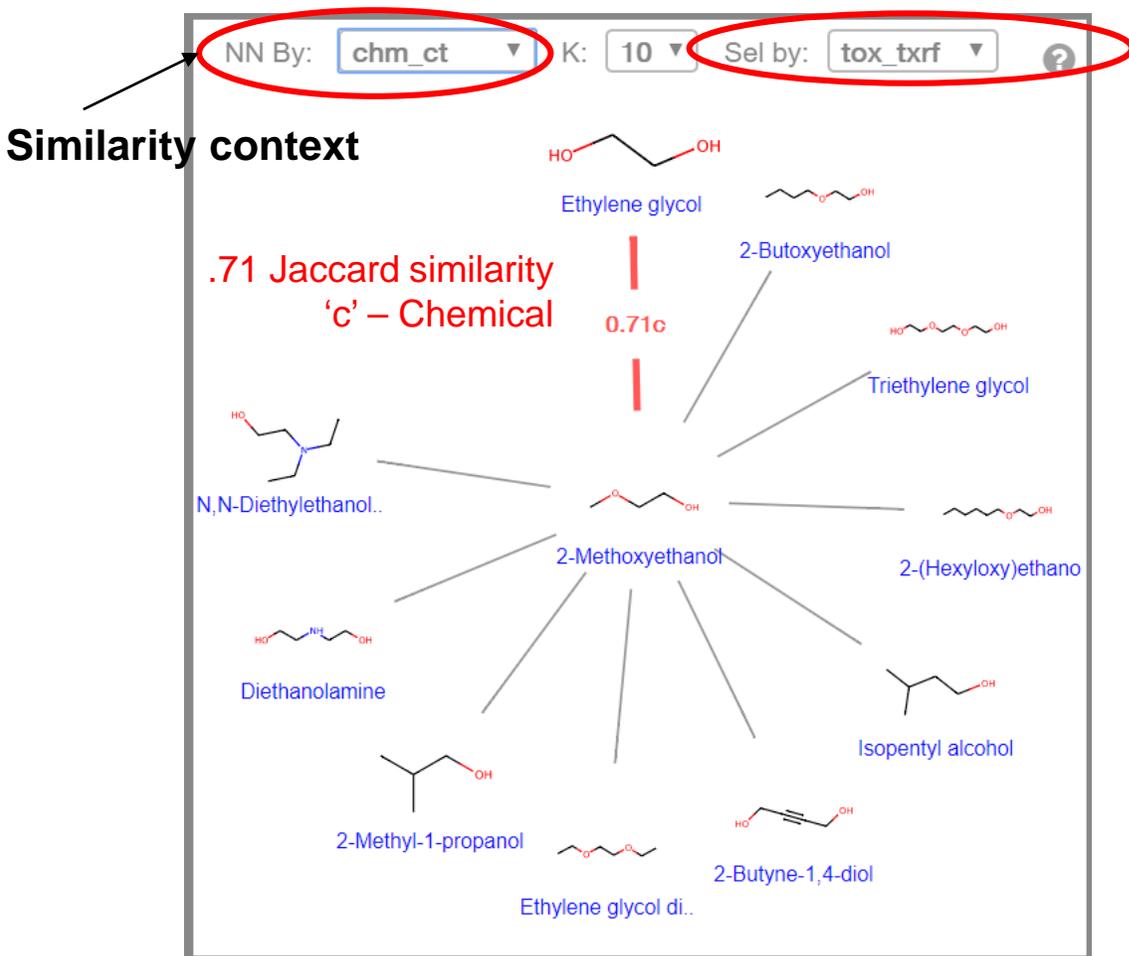
Frame 4

	2-Methoxyethanol	Triethylene gly	2-Butoxyethanol	Ethylene glycol	2-(Hexyloxy)eth	Dimethylaminoe	Isopentyl alcohol	2-Chloroethanol	2-Methyl-1-prop	N,N-Diethyleth	1,2-Propylene g
CHR:Adrenal Gland	Green	Green	Blue	Green	Grey	Grey	Grey	Green	Grey	Grey	Green
CHR:Blood	Green	Green	Green	Green	Grey	Grey	Grey	Green	Grey	Grey	Blue
CHR:Body Weight	Green	Green	Blue	Green	Grey	Grey	Grey	Blue	Grey	Grey	Blue
CHR:Bone Marrow	Green	Green	Blue	Green	Grey	Grey	Grey	Green	Grey	Grey	Green
CHR:Clinical Chemistry	Green	Green	Green	Green	Grey	Grey	Grey	Green	Grey	Grey	Blue
CHR:Hematology	Green	Green	Blue	Green	Grey	Grey	Grey	Green	Grey	Grey	Blue
CHR:Kidney	Green	Green	Blue	Blue	Grey	Grey	Grey	Green	Grey	Grey	Green
CHR:Liver	Green	Green	Blue	Blue	Grey	Grey	Grey	Green	Grey	Grey	Green
CHR:Lung	Green	Green	Green	Blue	Grey	Grey	Grey	Blue	Grey	Grey	Green
CHR:Mortality	Green	Green	Blue	Green	Grey	Grey	Grey	Blue	Grey	Grey	Green

Web Tool Workflow

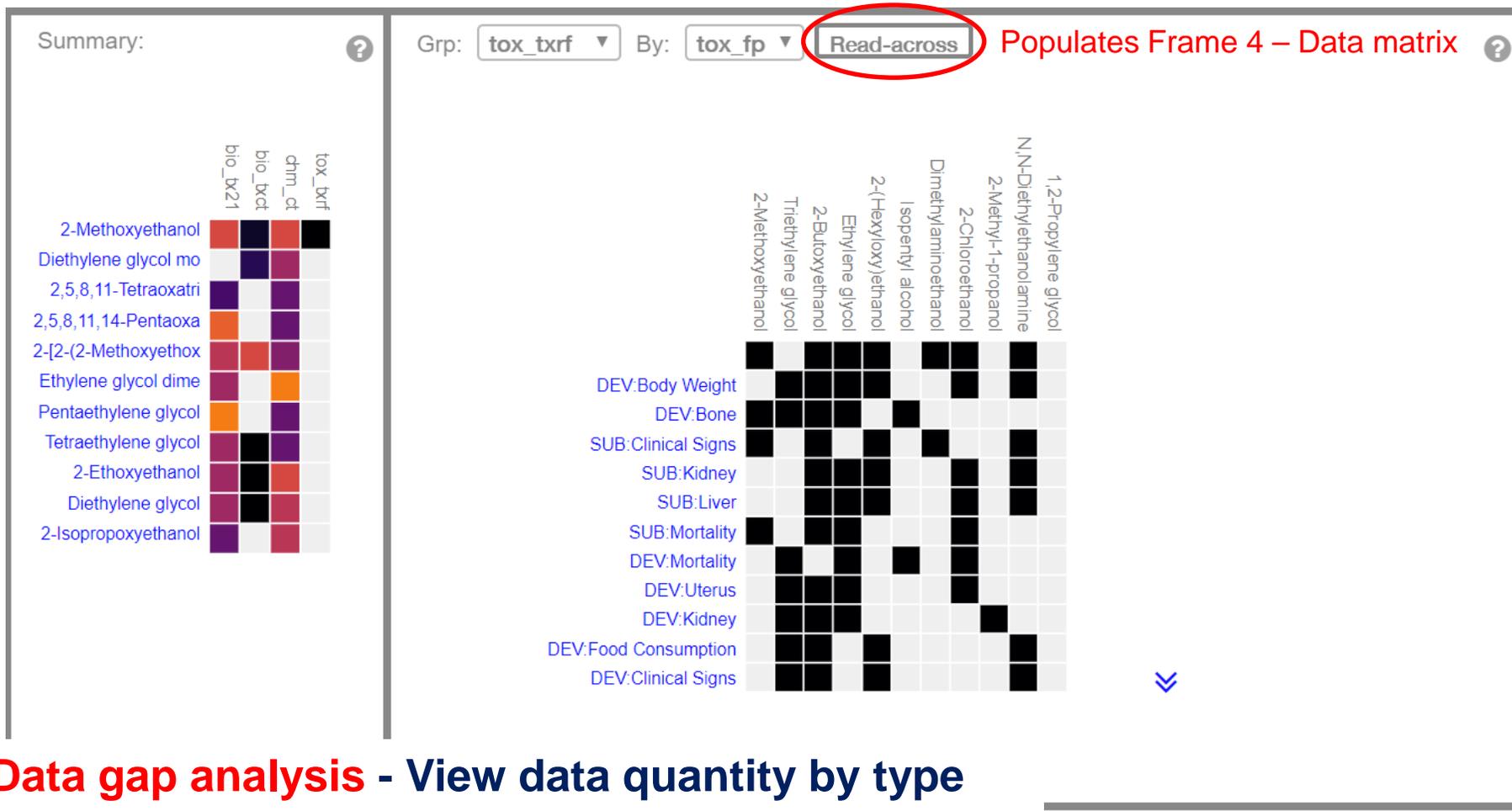


Analogue Identification and Evaluation



- ‘NN By’ option determines fingerprint of choice i.e. search for source analogues on the basis of different chemical fingerprints.
- ‘K’ option determines the number of source analogues to search for
- ‘Sel by’ option subsets chemical space to search for source analogues that have certain types of data, by default *in vivo* data.

Data Gap Analysis



- ‘Grp’ option controls the data that is being shown in the heatmap.
- ‘By’ option allows you to pivot the heatmap to look at different views of the same data.

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

Generate Data Matrix

Source analogues

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

Generates GenRA prediction

Target

	2-Methoxyethanol	Triethylene glycol	2-Butoxyethanol	Ethylene glycol	2-(Hexyloxy)ethanol	Dimethylaminoethanol	Isopentyl alcohol	2-Chloroethanol	2-Methyl-1-propanol	N,N-Diethylethanolamine	1,2-Propylene glycol
CHR:Adrenal Gland		Green	Blue	Green				Green			Green
CHR:Blood		Green	Green	Green							Blue
CHR:Body Weight		Green	Blue	Green				Blue			
CHR:Bone Marrow		Green	Blue	Green				Green			
CHR:Clinical Chemistry		Green	Green	Green				Green			
CHR:Hematology		Green	Blue	Green				Green			Blue
CHR:Kidney		Green	Blue	Blue				Green			Green
CHR:Liver		Green	Blue	Blue				Green			Green
CHR:Lung		Green	Green	Blue				Blue			Green
CHR:Mortality		Green	Blue	Green				Blue			Green

'Active' = blue
green = 'Inactive'

- 'Min+' and 'Min-' options control minimum number of pos and neg values in displayed endpoints.
- 'Sim wt' option changes width of cells to scale of the similarity of the neighbor.
- 'Export' will export data matrix to CSV

Run GenRA Predictions

Min+: Min-:
 Filter by:

 Sim wt
?

	1	0.37	0.32	0.31	0.28	0.26	0.26	0.25	0.24	0.23	0.22
2-Methoxyethanol											
Triethylene gly											
2-Butoxyethanol											
Ethylene glycol											
2-(Hexyloxy)eth											
Dimethylaminoe											
Isopentyl alcoh											
2-Chloroethanol											
2-Methyl-1-prop											
N,N-Diethyletha											
1,2-Propylene g											
CHR:Adrenal Gland											
CHR:Blood											
CHR:Body Weight											
CHR:Bone Marrow											
CHR:Clinical Chemistry											
CHR:Hematology											
CHR:Kidney											
CHR:Liver											
CHR:Lung											
CHR:Mortality											

GenRA Predictions

Filter by Endpoint or Analogues

Min+:
 Min-:
 Filter by:

	2-Methoxyethano	Triethylene gly	2-Butoxyethano	Ethylene glycol	2-(Hexyloxy)eth	Dimethylaminoe	Isopentyl alcoh	2-Chloroethano	2-Methyl-1-prop	N,N-Diethyletha	1,2-Propylene g
CHR:Liver	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
DEV:Liver	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
SUB:Liver	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
MGR:Liver	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
DNT:Liver	<input type="checkbox"/>										
SAC:Liver	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Provide user with flexibility to deselect analogues or focus in on specific toxicity effects or study types of interest

Future Work: Physicochemical Similarity Context

- Searching on the basis of physicochemical information in addition to structure can improve GenRA performance (Helman et al, 2018)
- Physicochemical information will be incorporated into the analogue identification and evaluation steps of the workflow.
- The ‘Lipinski Rule of 5’ properties (LogP, MW, #HB donors, #HB acceptors) were used in the analysis undertaken.

Future Work: Point of Departure Prediction

- Algorithm readily extendible to continuous data.

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

- Preliminary results appear promising for predicting acute oral rat LD50 data.

Conclusions

- GenRA is an objective algorithmic approach for generating read-across predictions.
- A web tool has been developed to front-end the GenRA workflow
- The web tool will be made available on NCCT's CompTox dashboard soon.

Acknowledgements

- Imran Shah
- Grace Patlewicz
- Tony Williams
- Jeff Edwards
- Chris Grulke