http://www.orcid.org/0000-0003-3863-9689

EPA United States Environmental Protection Agency

Practical Applications of New Approach Methodologies (NAM) in Chemical Safety Assessment: Applying US EPA Tools/Approaches in Practice

Physicians Committee for Responsible Medicine

NURA™

Generalized Read-across (GenRA) – Data driven systematic read-across approaches

Grace Patlewicz

Center for Computational Toxicology and Exposure, US-EPA, RTP, NC

10/23/2020

Happy Mole Day!

Outline

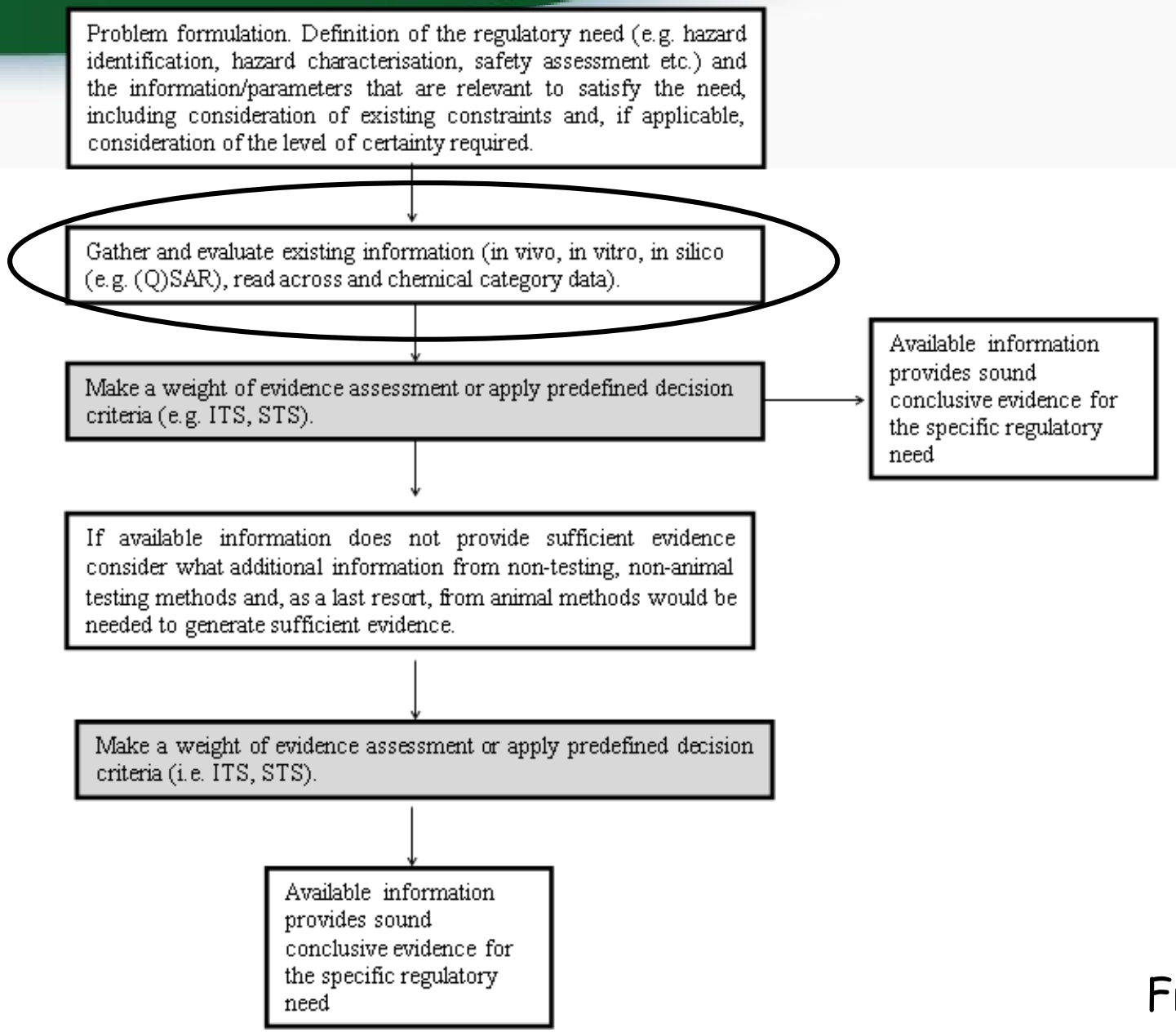
- Where does read-across fit in an integrated testing and assessment approach (IATA)
- What is read-across?
- Read-across tools and frameworks?
- How is read-across evolving? A GenRA perspective
- GenRA workflow
- Example using GenRA
- Ongoing work
- Summary remarks
- Acknowledgements

Where do read-across approaches fit within IATA?

- Integrated Approaches to Testing and Assessment (IATA)

- “A tiered approach to data gathering, testing, and assessment that integrates different types of data (including physicochemical and other chemical properties as well as *in vitro* and *in vivo* toxicity data). When combined with estimates of exposure in an appropriate manner, the IATA provides predictions of risk.”

General framework of an IATA



From OECD

Definitions: Chemical grouping approaches

EPA United States Environmental Protection Agency

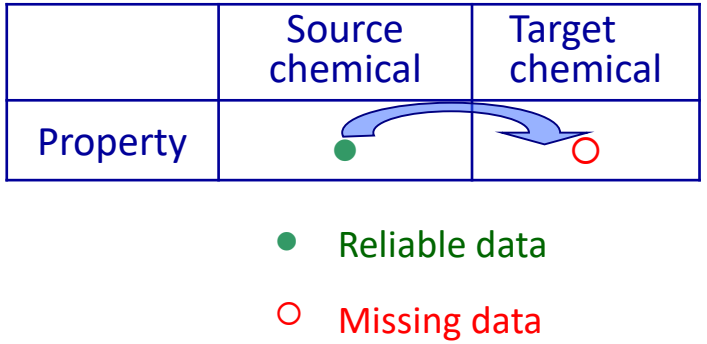
"Analogue approach" refers to grouping based on a very limited number of chemicals (e.g. target substance + source substance)

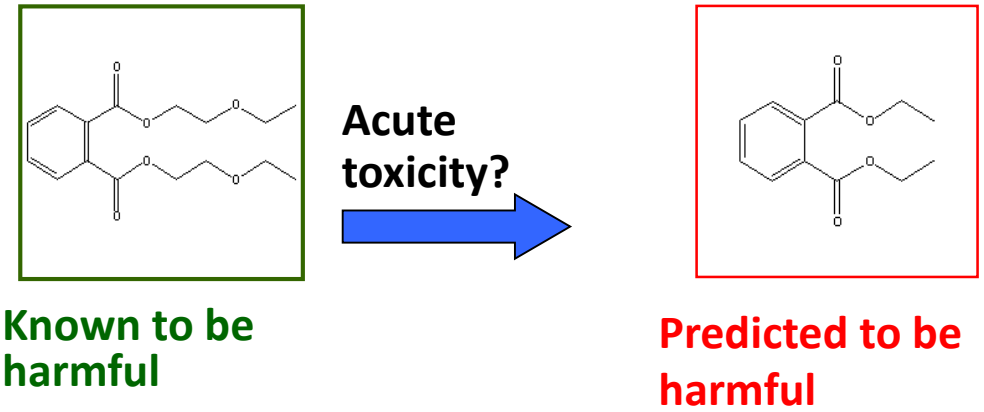
"Category approach" is used when grouping is based on a more extensive range of analogues (e.g. 3 or more members)

A chemical category is a group of chemicals whose physico-chemical and human health and/or environmental toxicological and/or environmental fate properties are likely to be similar or follow a regular pattern as a result of structural similarity (or other similarity characteristics).

Definition

- 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.





Selected read-across tools

Too

Ana
ider

Ana
Eva

Dat
ana

Dat

Unc
ass

Ava

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

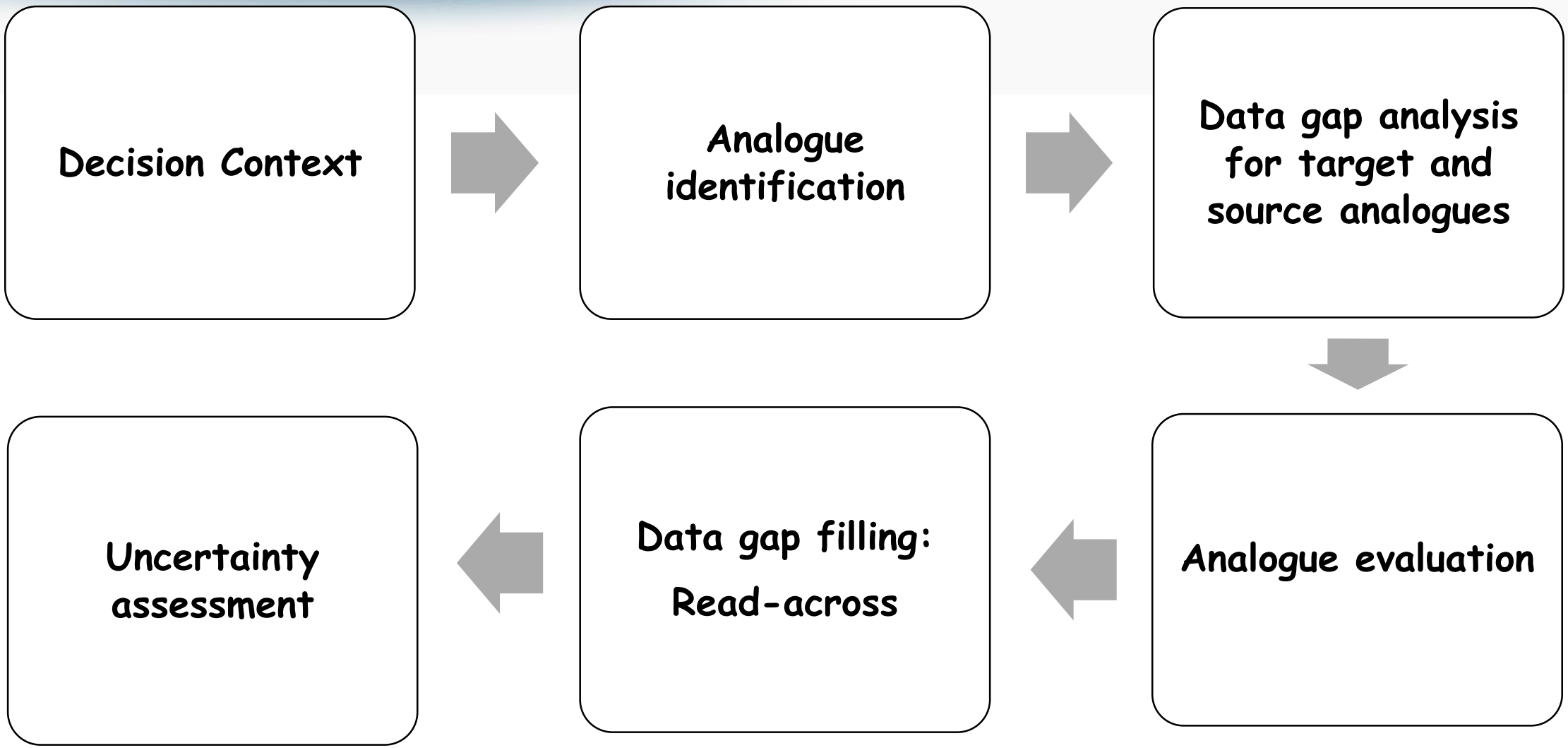
Keywords:
Category approach
Analogue approach
Data gap filling
Read-across
(Q)SAR
Trend analysis
Nearest neighb

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.

Read-across workflow



A harmonized hybrid read-across workflow

Contents lists available at ScienceDirect

Computational Toxicology

journal homepage: www.elsevier.com

Journal Cover Image

Navigating through the minefield of read-across frameworks: A commentary perspective

Grace Patlewicz[a, *], Mark T.D. Cronin[b], George Helman[a, c], Jason C. Lambert[d], Lucina E. Lizarraga[d], Imran Shah[a]

[a] National Center for Computational Toxicology (NCCT), Office of Research and Development, US Environmental Protection Agency (US EPA), 109 TW Alexander Dr, Research Triangle Park (RTP), NC 27711, USA
[b] School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK
[c] Oak Ridge Institute for Science and Education (ORISE), 1299 Bethel Valley Road, Oak Ridge, TN 37830, USA
[d] National Center for Evaluation Assessment (NCEA), US Environmental Protection Agency (US EPA), 26 West Martin Luther King Dr, Cincinnati, OH 45268, USA

A harmonized hybrid read-across workflow

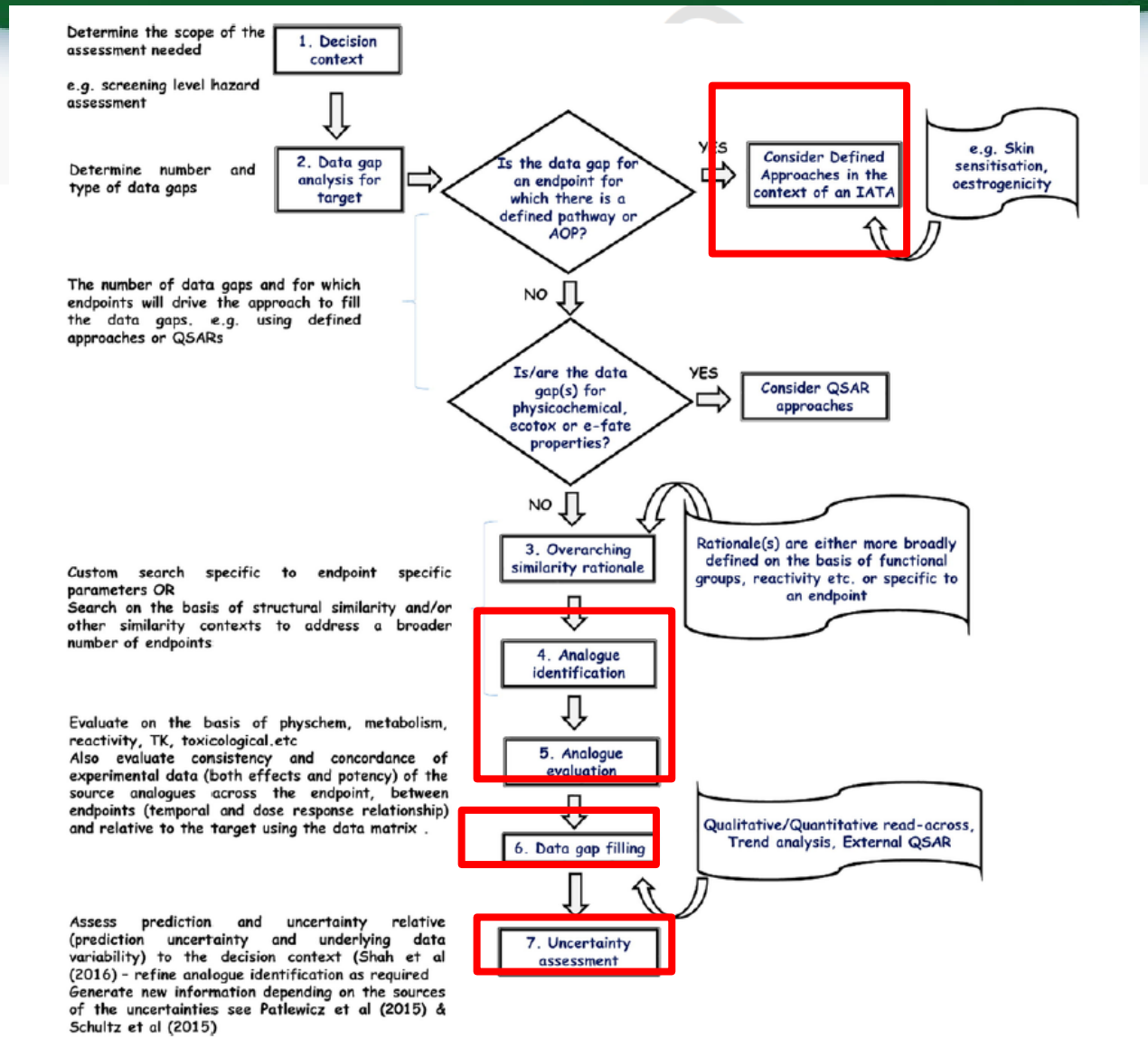


Fig. 9. A harmonised hybrid development and assessment framework.

Where do other new approach data streams fit? E.g. mechanistic data from ToxCast

How should we transition to data-driven approaches? moving away from subjective expert driven assessments.

What about characterizing the uncertainty of the predictions made?

GenRA (Generalized Read-Across)

- Predicting toxicity as a similarity-weighted activity of nearest neighbors based on chemistry and bioactivity descriptors (Shah et al, 2016)
- Generalized version of the 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

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

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 $\boldsymbol{x}_i^{\alpha}, \boldsymbol{x}_j^{\alpha}$

k = up to k nearest neighbours

Read-across workflow in GenRA v1.0

EPA United States Environmental Protection Agency

Decision Context

Screening level assessment of hazard based on toxicity effects from ToxRefDB v1

Analogue identification

Similarity context is based on structural characteristics

Data gap analysis for target and source analogues

Analogue evaluation

Evaluate consistency and concordance of experimental data of source analogues across and between endpoints

Read-across

Similarity weighted average – many to one read-across

Uncertainty assessment

Assess prediction and uncertainty using AUC and p value metrics

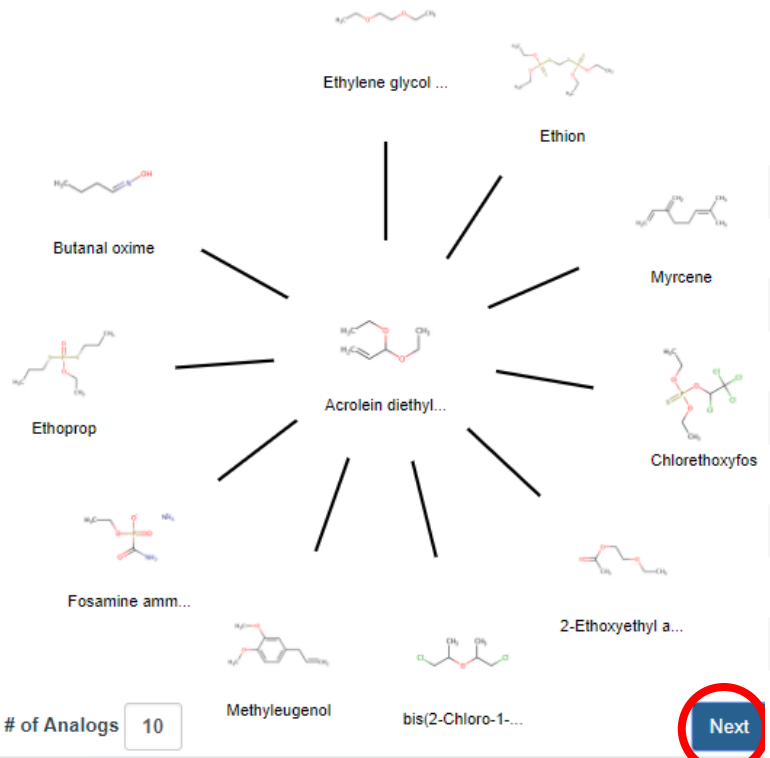
GenRA tool web-based tool

GenRA tool in practice

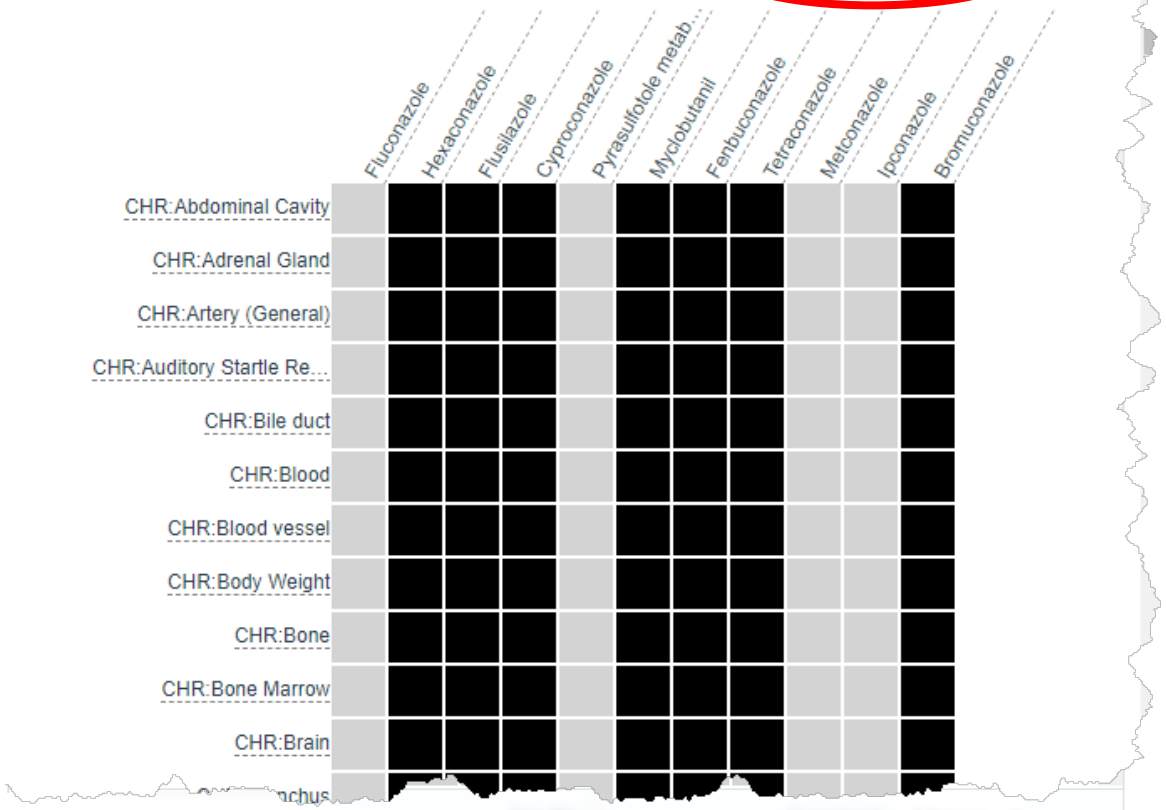
GenRA

Step Two: Data Gap Analysis & Generate Data Matrix

Neighbors by: Chem: Morgan Fgrprts | Filter by: invivo data | Summary Data Gap Analysis | Group: ToxRef | By: Tox Fingerprint | Generate Data Matrix

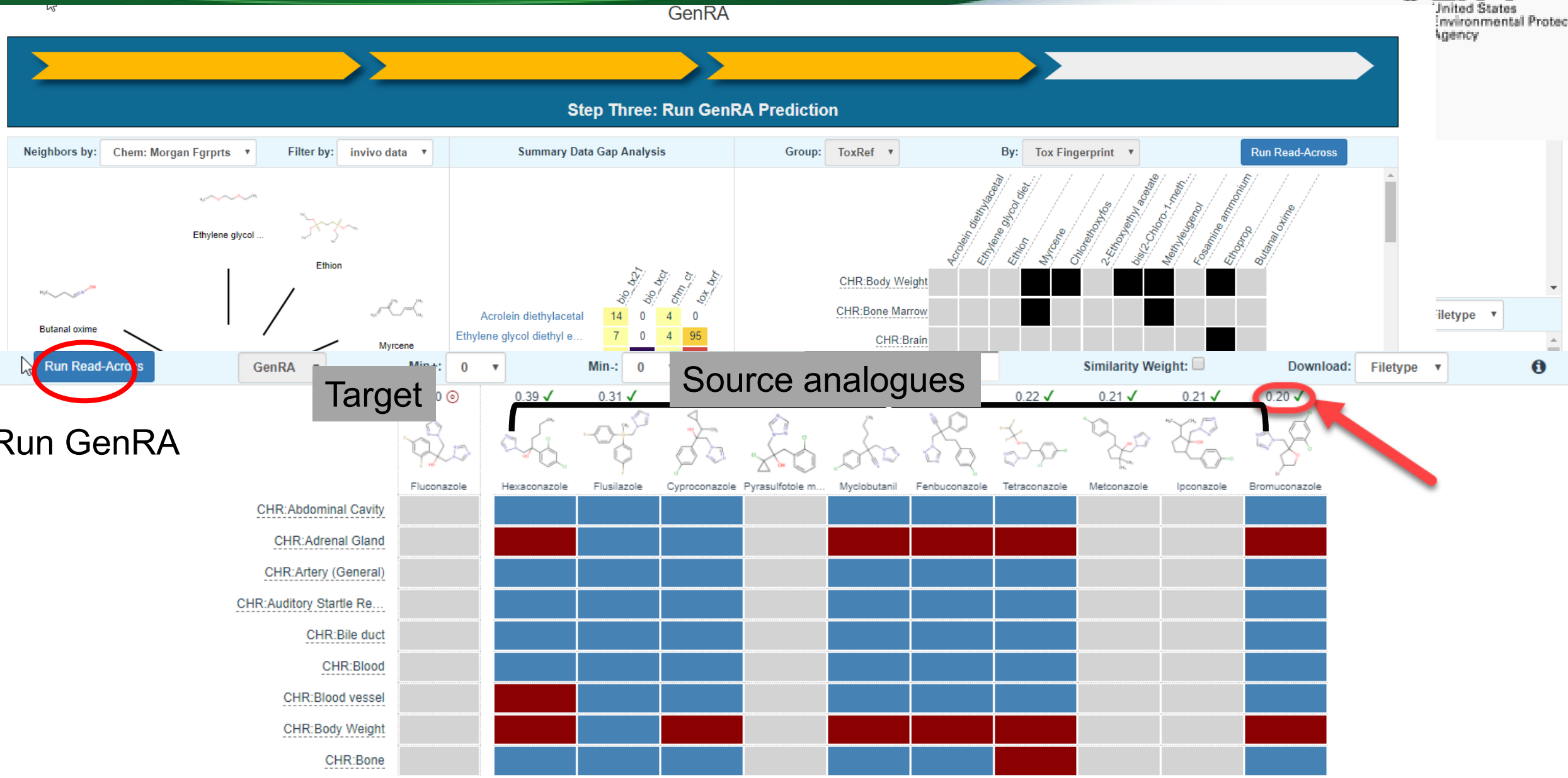


	bio_tx21	bio_txct	chm_ct	tox_txrf
Fluconazole	3	714	15	0
Hexaconazole	43	819	18	345
Flusilazole	28	819	9	345
Cyproconazole	14	819	16	408
Pyrasulfotole metabolite ...	0	0	18	234
Myclobutanil	15	818	15	345
Fenbuconazole	34	819	17	345
Tetraconazole	35	819	20	345
Metconazole	35	215	15	82
Ipconazole	46	232	16	180
Bromuconazole	24	277	13	345



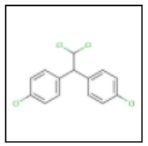
Data gap analysis

GenRA tool in practice



Run GenRA

GenRA tool in practice



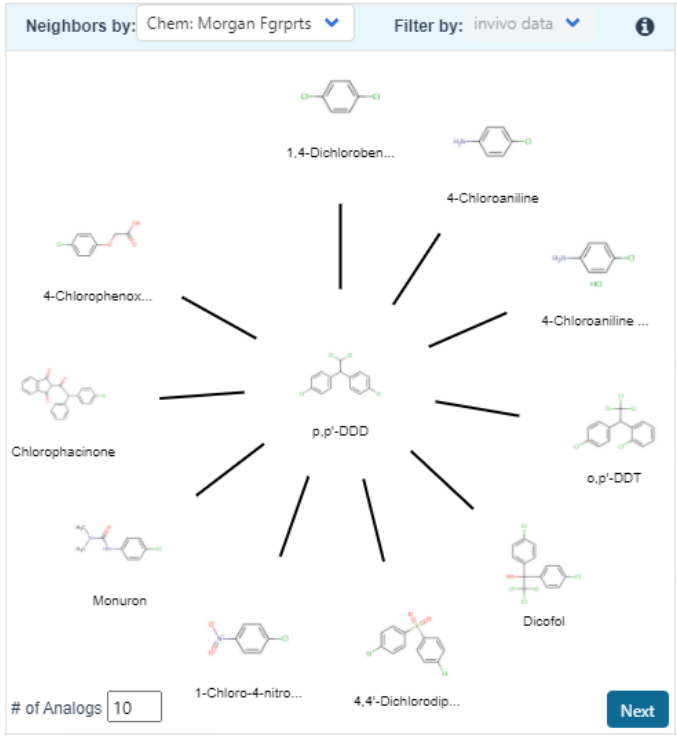
p,p'-DDD
72-54-8 | DTXSID4020373
Searched by DSSTox Substance Id.

Search for a chemical and click on the GENRA link on the lefthand panel

- DETAILS
- EXECUTIVE SUMMARY
- PROPERTIES
- ENV. FATE/TRANSPORT
- HAZARD
- SAFETY
- ADME
- EXPOSURE
- BIOACTIVITY
- SIMILAR COMPOUNDS
- **GENRA (BETA)**
- RELATED SUBSTANCES
- SYNONYMS
- LITERATURE
- LINKS
- COMMENTS

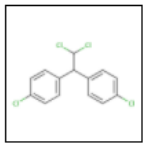
Generalized Read-Across (GenRA)

Step One: Analog Identification and Evaluation



Radial plot with Target chemical of interest in the centre and source analogues (similar) ordered clockwise by decreasing similarity (Jaccard)

GenRA tool in practice



p,p'-DDD

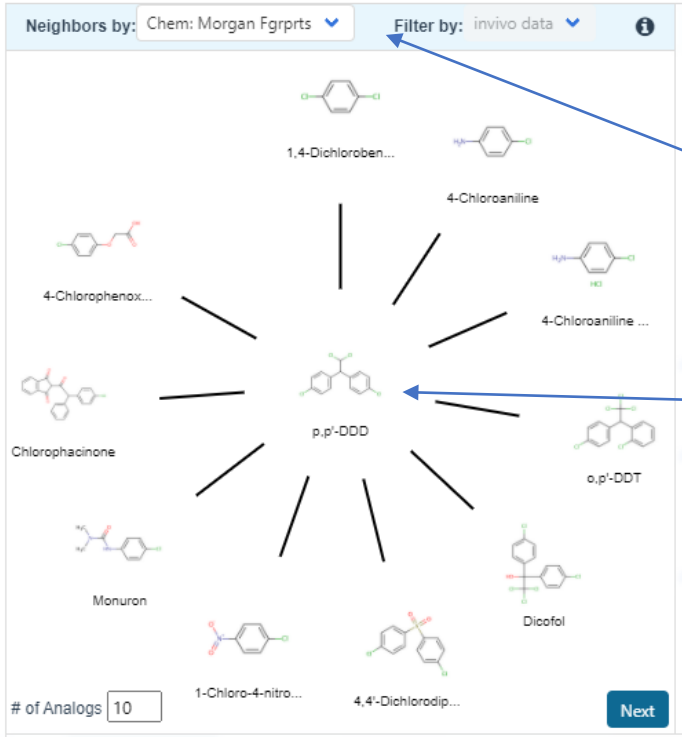
72-54-8 | DTXSID4020373

Searched by DSSTox Substance Id.

- DETAILS
- EXECUTIVE SUMMARY
- PROPERTIES
- ENV. FATE/TRANSPORT
- HAZARD
- SAFETY
- ADME
- EXPOSURE
- BIOACTIVITY
- SIMILAR COMPOUNDS
- **GENRA (BETA)**
- RELATED SUBSTANCES
- SYNONYMS
- LITERATURE
- LINKS
- COMMENTS

Generalized Read-Across (GenRA)

Step One: Analog Identification and Evaluation



Similarity context – Chemical Fingerprints (Morgan)

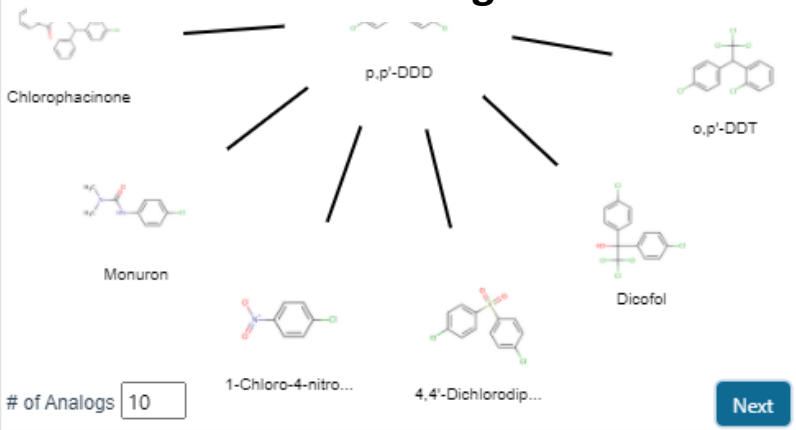
Target chemical of interest

GenRA tool in practice

United States Environmental Protection Agency

Step Two: Data Gap Analysis & Generate Data Matrix

- **How data poor is my target and what data exists for the source analogues identified**
- **Do they address the data gaps of interest for the target chemical?**



# of Analogs 10

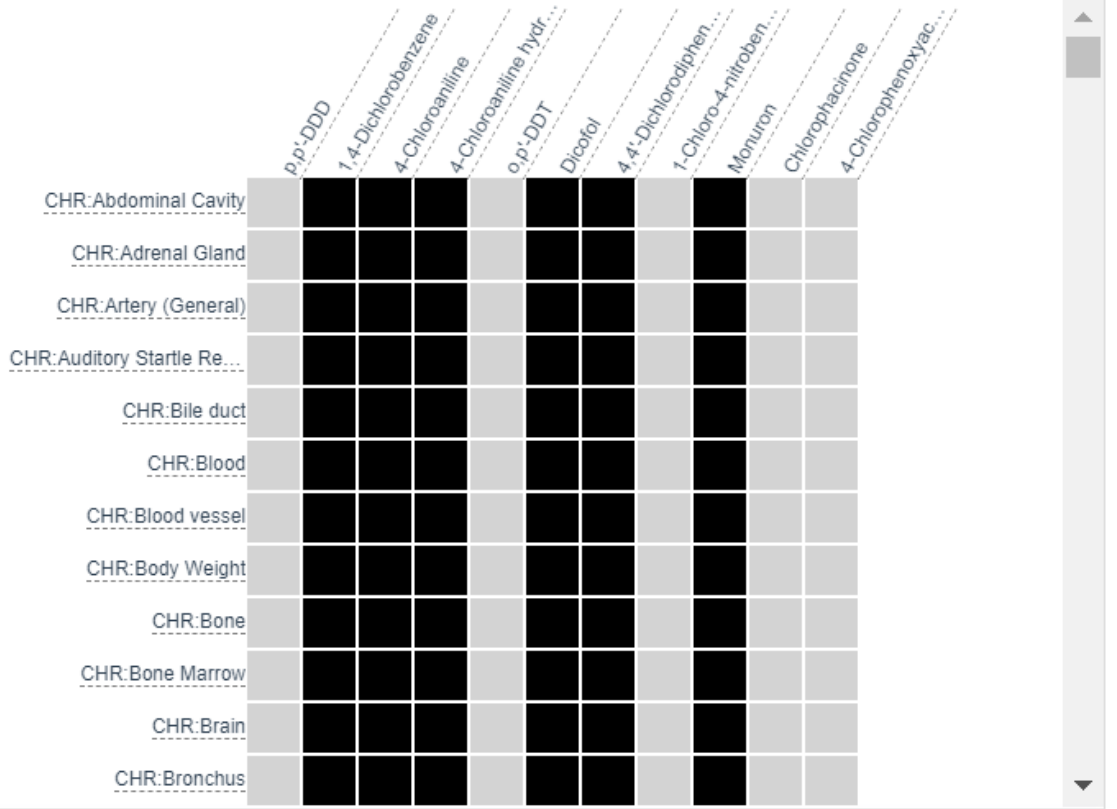
Next

Neighbors by: Chem: Morgan Fgrprts | Filter by: invivo data

Summary Data Gap Analysis

	bio_tx21	bio_txct	chm_ct	tox_txrf
p,p'-DDD	42	714	10	0
1,4-Dichlorobenzene	7	714	4	345
4-Chloroaniline	6	714	6	83
-Chloroaniline hydrochl...	17	0	7	168
o,p'-DDT	37	726	12	177
Dicofol	40	818	17	345
4,4'-Dichlorodiphenyl sul...	9	271	5	168
1-Chloro-4-nitrobenzene	10	674	5	167
Monuron	12	714	7	168
Chlorophacinone	51	234	19	95
4-Chlorophenoxyacetic ...	9	232	8	180

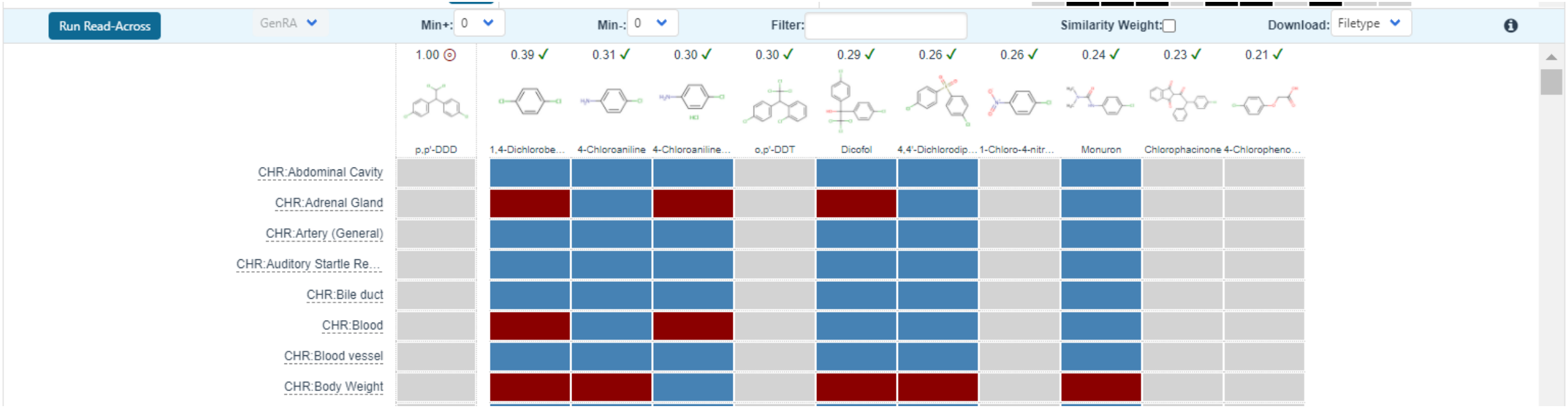
Group: ToxRef | By: Tox Fingerprint | Generate Data Matrix



GenRA tool in practice

What is the consistency and concordance across my source analogues?
Should I deselect analogues from consideration from the entire set of predictions?
Should I consider subcategorizing the analogues selected?

Toxicity data represented as binary outcomes – red (positive), blue (negative), grey (no data)



GenRA tool in practice

First column is updated with predictions

GenRA Tool in practice

- Database underpinning GenRA v1.0: ToxRefDB v1
  - Different study types and effects within them are predicted e.g. chronic_liver is annotated as CHR_liver
  - Negative results – assume that if a particular guideline study was conducted but the effects were not reported than a chemical would be negative for that particular effect for that type of guideline study
  - Positive results – min dose at which toxicity effects are observed in a study
- Prediction: Similarity weighted activity
- Performance is categorized by the AUC of the ROC
  - The significance was empirically estimated by constructing a null distribution by permuting the toxicity values 100 times and calculating the fraction of times the AUC was more extreme than what would be observed by chance (this is reported as the p-value).

GenRA Tool in practice

- Ability to export the predictions into a tsv/excel file
- Output is shown – mirrors the data matrix view but doses replace the colour coding

role	target		analog		analog		analog		analog
preferred	p,p'-DDD		1,4-Dichlo		4-Chloroa		4-Chloroa		o,p'-DDT
dsstox_sid	DTXSID40		DTXSID10		DTXSID90		DTXSID40		DTXSID60
molecular	320.03		147		127.57		164.03		354.48
similarity	1		0.391304		0.310345		0.3		0.295455
CHR:Abdo	GenRA Ne		no_effect		no_effect		no_effect		no_data
CHR:Adrer	GenRA Pos		600mg/kg,		no_effect		18mg/kg/c		no_data
CHR:Arter	GenRA Ne		no_effect		no_effect		no_effect		no_data
CHR:Audit	GenRA Ne		no_effect		no_effect		no_effect		no_data
CHR:Bile d	GenRA Ne		no_effect		no_effect		no_effect		no_data
CHR:Blooc	GenRA Ne		150mg/kg,		no_effect		3mg/kg/da		no_data
CHR:Blooc	GenRA Ne		no_effect		no_effect		no_effect		no_data
CHR:Body	GenRA Pos		300mg/kg		375mg/kg		no_effect		no_data

GenRA Tool in practice

- Output can be analyzed in different ways

role	target		analog
preferred name	p,p'-DDD		1,4-Dichlorobenzene
dsstox_sid	DTXSID4020373		DTXSID1020431
molecular weight	320.03		147
similarity	1		0.391304348
CHR:Abdominal Cavity	GenRA Neg Act=0 (0) AUC=0 p=1		no_effect
CHR:Adrenal Gland	GenRA Pos Act=1 (0.546) AUC=0 p=0.975		600mg/kg/day
CHR:Artery (General)	GenRA Neg Act=0 (0) AUC=0 p=1		no_effect
CHR:Auditory Startle Reflex Habituation	GenRA Neg Act=0 (0) AUC=0 p=1		no_effect
CHR:Bile duct	GenRA Neg Act=0 (0) AUC=0 p=1		no_effect
CHR:Blood	GenRA Neg Act=0 (0.386) AUC=0 p=0.95		150mg/kg/day
CHR:Blood vessel	GenRA Neg Act=0 (0) AUC=0 p=1		no_effect
CHR:Body Weight	GenRA Pos Act=1 (0.832) AUC=0 p=0.8		300mg/kg/day
CHR:Bone	GenRA Neg Act=0 (0) AUC=0 p=1		no_effect
CHR:Bone Marrow	GenRA Neg Act=0 (0.168) AUC=0 p=0.85		no_effect
CHR:Brain	GenRA Neg Act=0 (0) AUC=0 p=1		no_effect
CHR:Bronchus	GenRA Neg Act=0 (0) AUC=0 p=1		no_effect

GenRA Tool in practice

- Rank order positive results based on AUC and p values
- Look at the distribution of positive vs negatives predictions
- Explore what effects are being identified for the source analogues – consider identifying the underlying data for source analogues (elsewhere on the Dashboard) – is there a critical effect that is driving the toxicity that should be compared with the target chemical predictions?
- ……
- Depends on the decision context and the level of uncertainty that can be tolerated.

GenRA – Ongoing refinements

- Consideration of other information to define and refine the analogue selection & evaluation – e.g. physicochemical similarity, metabolic similarity, reactivity similarity, mechanistic similarity (transcriptomics similarity, phenotypic profiling similarity)
- Quantitative predictions of toxicity e.g. LOAEL, LD50
- Read-across to predict other *in vitro* endpoints to supplement *in vitro-in vivo* extrapolations

Summary remarks

- Read-across is an important component of an IATA
- Read-across acceptance for regulatory purposes remains an issue – mainly due to the difficulties of addressing residual uncertainties and the fact that read-across is a subjective expert driven assessment.
- GenRA is an attempt to move towards an objective read-across approach where uncertainties and performance can be quantified.
- GenRA v1.0 establishes a baseline in performance. The approach relies on chemical descriptors to predict binary toxicity values but work is ongoing to characterize other contexts of similarity (e.g. mechanistic, reactivity, metabolism) and quantify their contribution in predicting *in vivo* toxicity outcomes.
- GenRA v1.0 exists as an app within the Dashboard to facilitate a workflow approach to make read-across predictions. A python package (genra-py) will be released soon to facilitate batch processing using user specific datasets.

Acknowledgements

- Imran Shah – co-collaborator
- George Helman – former student
- Tony Williams – Project Owner for the Dashboard

Contact

Grace Patlewicz
CCTE, US EPA Office of Research and Development,
patlewicz.grace@epa.gov
ORCID: https://orcid.org/0000-0003-3863-9689