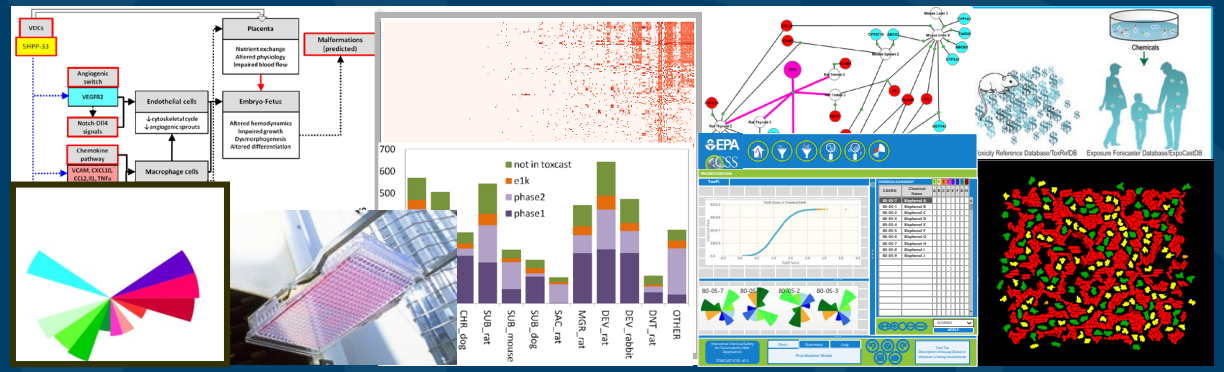
GenRA: From research and implementation to practical application



SOT 2017 Workshop: Application of Data from New Approaches in Regulatory and Product Safety Decisions
Tuesday 13th March 2018

Grace Patlewicz
National Center for Computational Toxicology

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

EPA United States Environmental Protection Agency

Outline

- Definitions
- Frameworks for Development & Assessment of category/analogue approaches to facilitate read-across
- Proposal for a Harmonised framework for read-across
- GenRA approach – "Baseline" GenRA
- Recent refinements to GenRA
- Practical case study example
- Take home messages

Definitions: Chemical grouping approaches

- Read-across describes one of the techniques for filling data gaps in either the analogue or category approaches i.e. not to be confused with the "analogue approach"

- "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).

Frameworks for developing category/analogue approaches

OECD/ECHA

- ECETOC (2012) Patlewicz et al (2013; 2015)

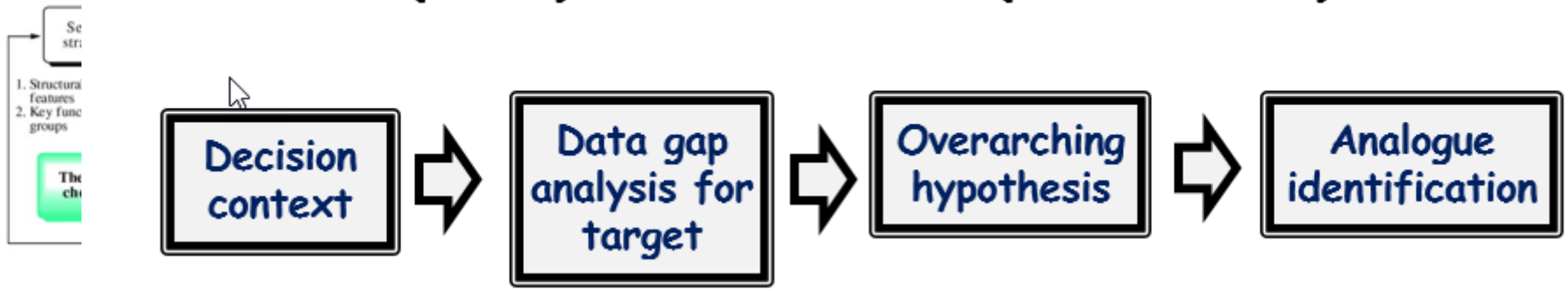
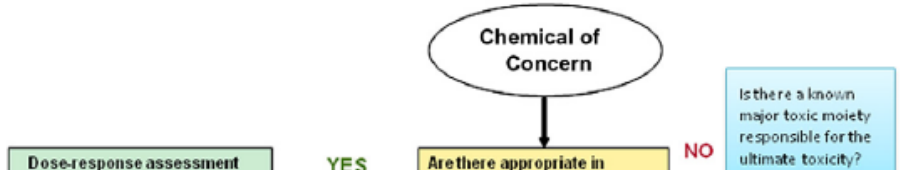
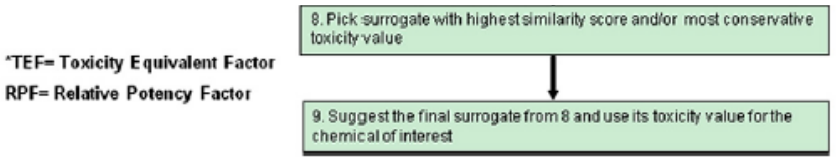


Figure 2 - Stepwise approach to an analogue approach





*TEF= Toxicity Equivalent Factor
RPF= Relative Potency Factor

Fig. 1. Tiered surrogate approach.

Summary highlights of read-across development frameworks

Framework	ECHA	OECD	Wu	Wang	Patlewicz
Context	REACH	International regulatory purposes	Product Stewardship	Quantitative risk assessment	Regulatory purposes/Product stewardship
Approach	Analogue/Category – aim is to fill an endpoint specific study. Focused on structural similarity as a starting point Approach is more hypothesis driven	Analogue/Category – a generalisation of the ECHA approach	Analogue Systematic stepwise evaluation of analogue suitability based on structure, reactivity, p-chem and metabolism	Analogue Approach is based on a WOE assessment from structure, ADME and toxicity considerations	Analogue Stepwise approach considering general (pchem, reactivity, metabolism) and endpoint specific considerations
Terms of reference	Target/Source	Target/Source	Substance of interest/Analogue	Chemical of Concern/Surrogate	Analogue/Category
Scope	Endpoint specific	Endpoint specific	Systematic stepwise evaluation of analogue suitability based on structure, reactivity, p-chem and metabolism Most sensitive/relevant endpoint – focused on repeated dose toxicity endpoints; quantitative risk assessment	Approach is based on a WOE assessment from structure, ADME and toxicity considerations. "Best" surrogate is selected from a set of candidates based on most similar and most conservative toxicity value	Approach is aimed to identify source analogues that can be used to address as many endpoints as appropriate, even though the read-across prediction itself is justified on an endpoint per endpoint basis and some source analogues might be excluded from the prediction itself if they are not appropriate for specific endpoints of

Frameworks for Assessing Read-across

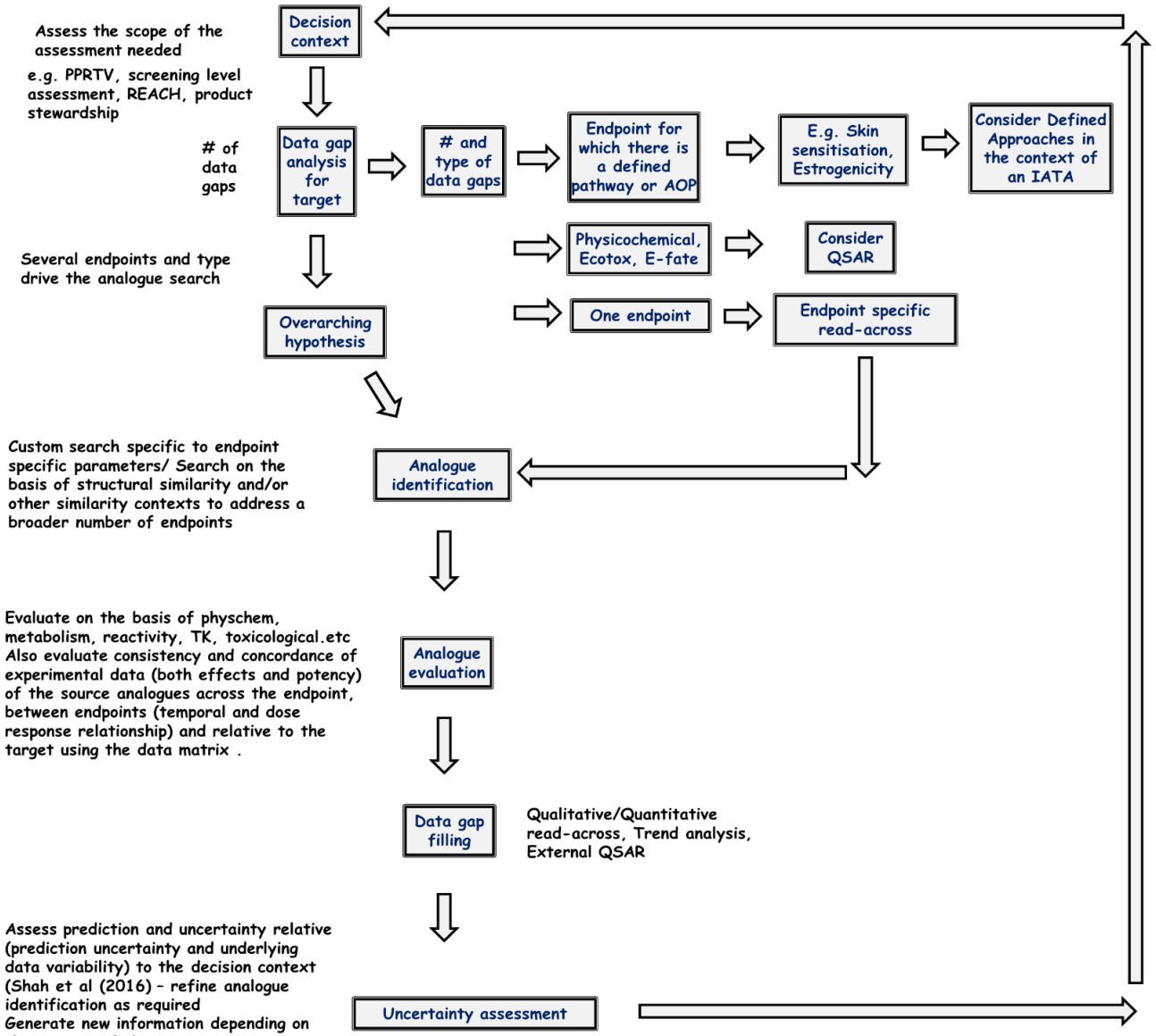
- Blackburn & Stuard
- Patlewicz et al (2015)
- Schultz et al (2015)
- ECHA RAAF (2015, 2017)

- These aim to identify, document and address the uncertainties associated with read-across inferences/predictions

Summary highlights of read-across assessment frameworks

Framework	ECHA RAAF (2017)	Blackburn and Stuard (2014)	Patlewicz et al (2015)	Schultz et al (2015)
Context	REACH	Product Stewardship	Regulatory purposes & Product stewardship	Regulatory purposes & Product stewardship
Scope	Analogue/Category	Analogue/Category	Analogue/Category	Analogue/Category
Framework	Scenarios addressing analogue (2) and category (4) approaches as described above Each scenario is associated with a number of assessment elements (AE) (both common and scenario specific).	Framework addresses 3 aspects: analogue suitability (covered in Wu et al, 2010); data quality of the analogues; consistency of the data across the analogues and relative to the target	Identifies the sources of uncertainty in relationship to the data and similarity context	Different scenarios are articulated to frame up to 11 different similarity criteria. factors proposed to evaluate mechanistic relevance and completeness of the read-across
Grading scale	Each AE is scored by an	Low - High gradings which	None - possible strategies to	Low to High but no default

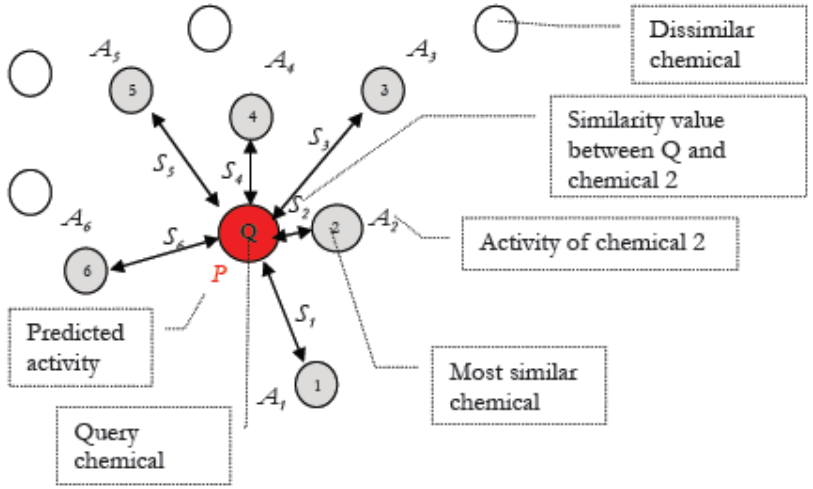
A harmonised read-across workflow



Proposed in Patlewicz et al., 2018 *under review*

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

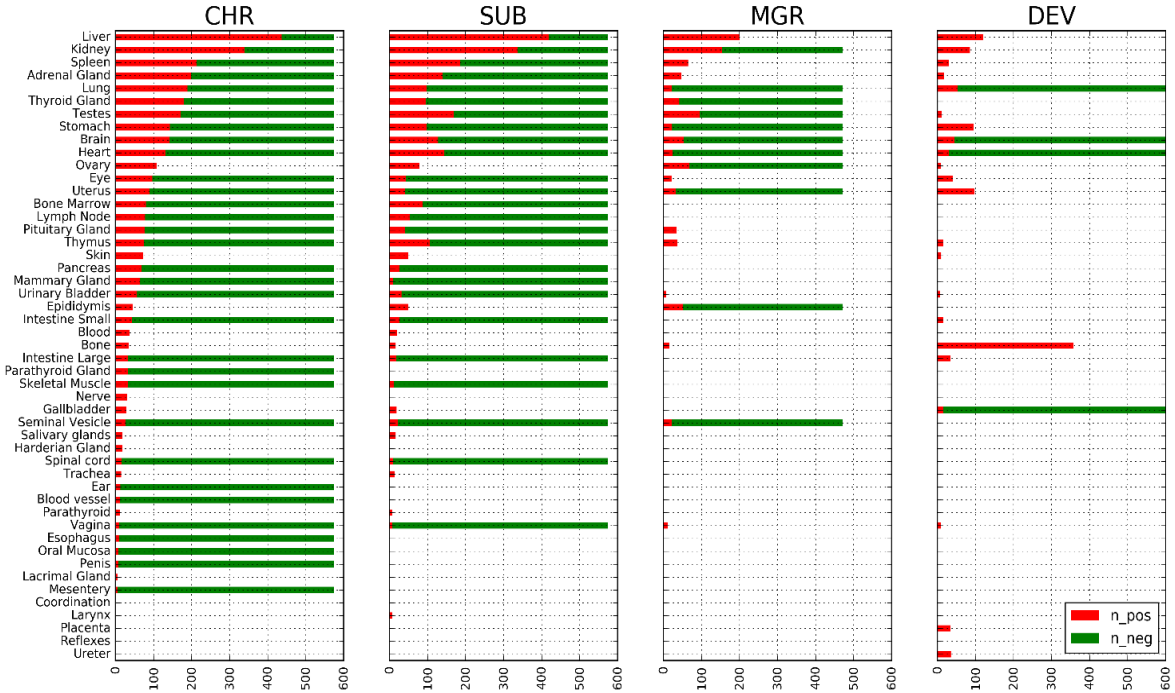
574 toxicity effects (tox) ToxRefDB

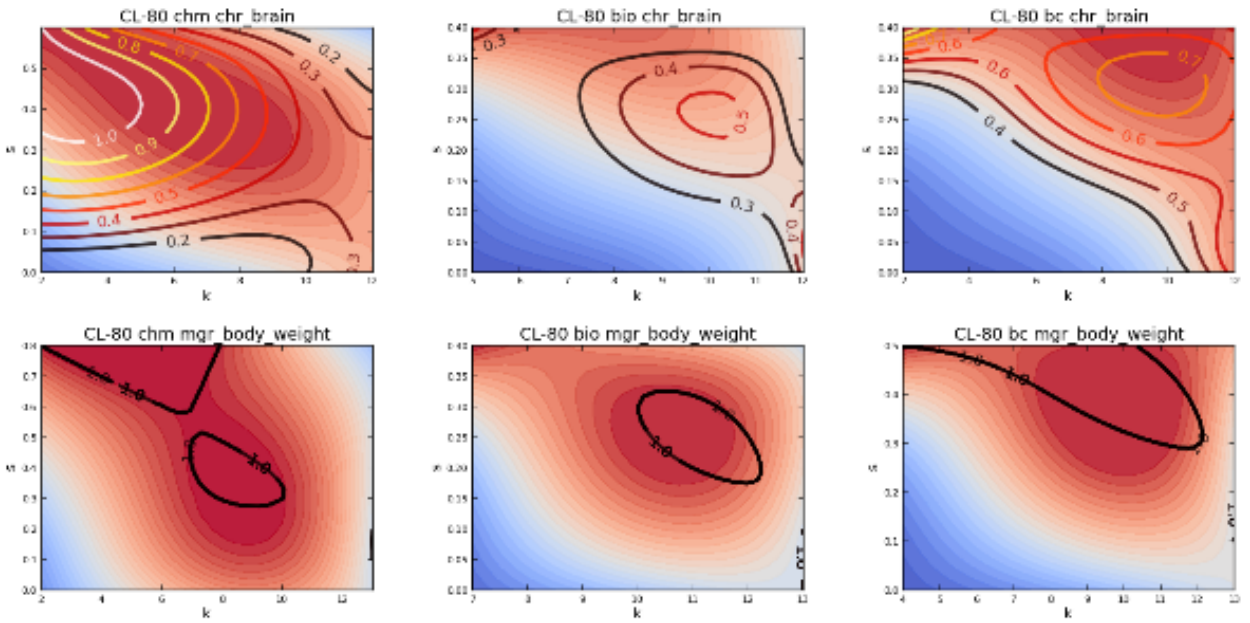
II. Define Local neighbourhoods

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

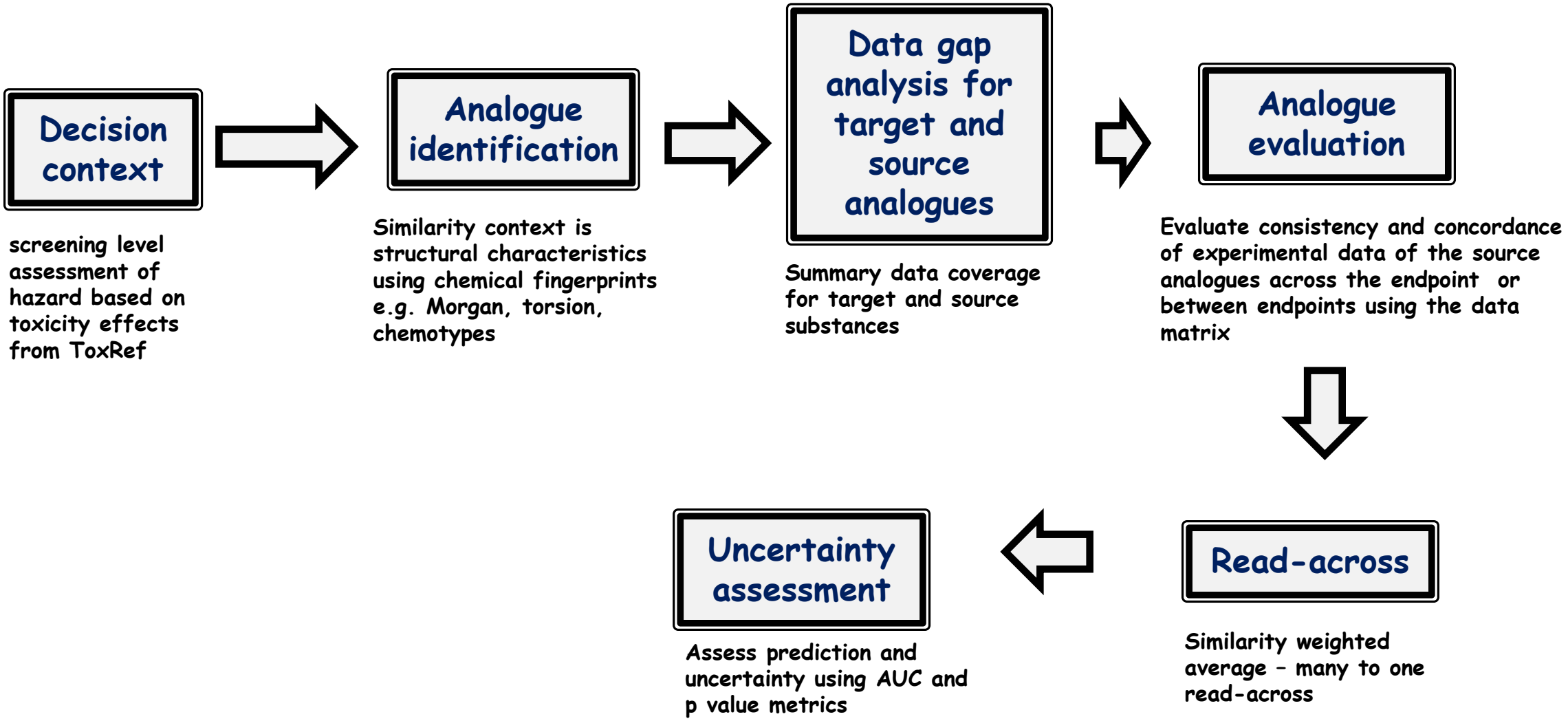
III. GenRA

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





Current Category Workflow in GenRA



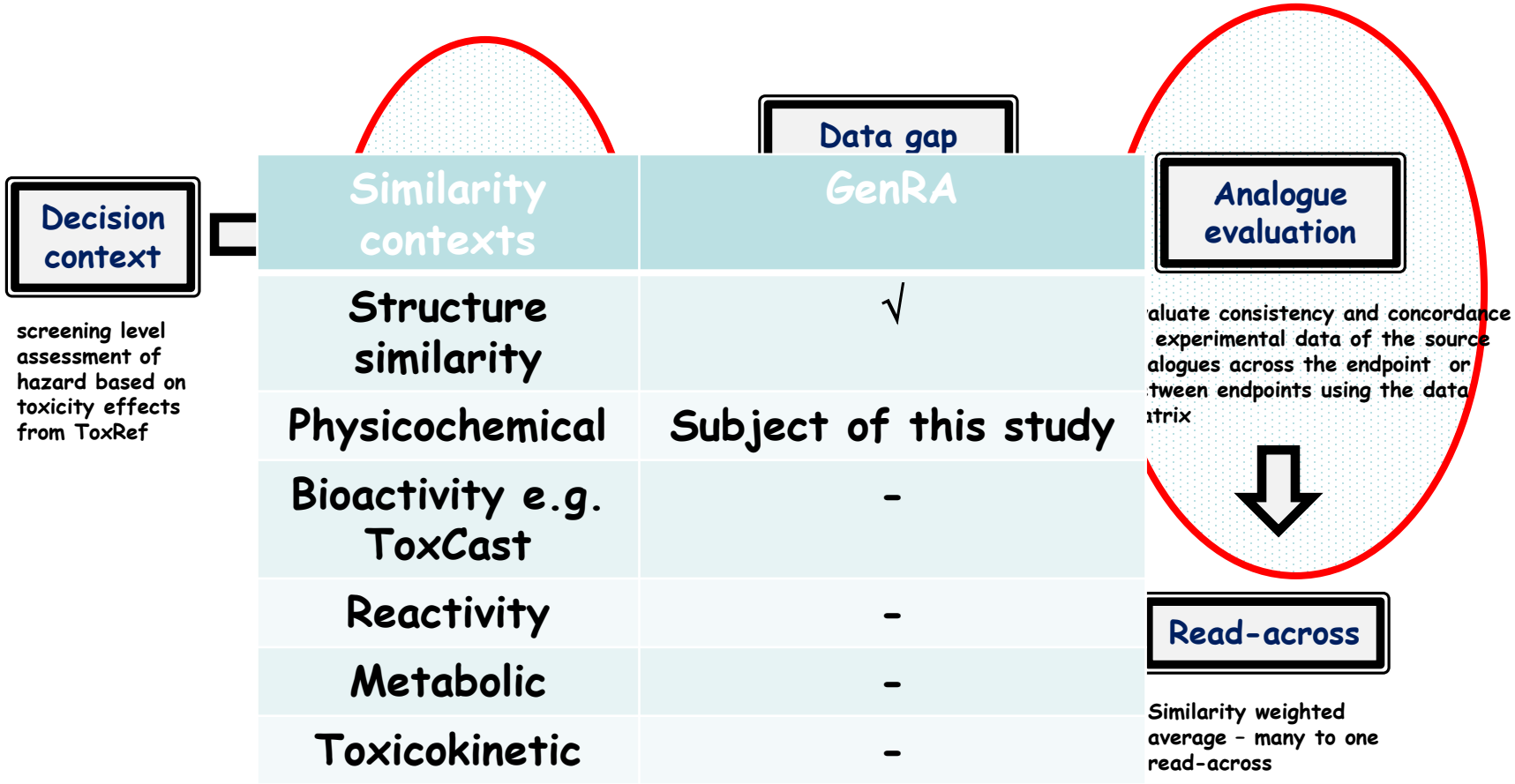
GenRA tool in development

GenRA – Outcomes 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…
- Dose response information to refine scope of prediction beyond binary outcomes..

Refinements to the GenRA approach



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:

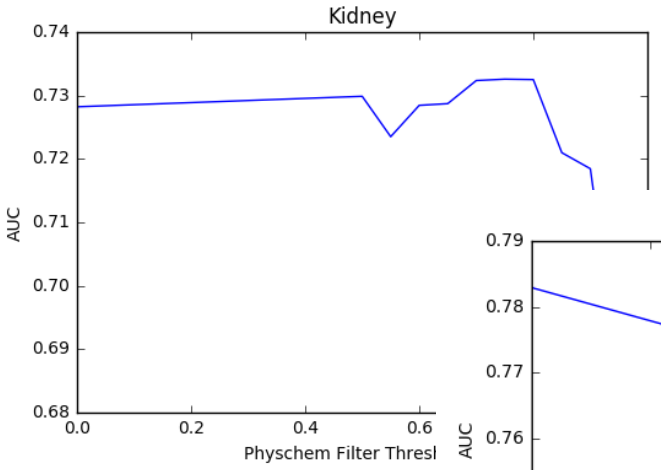
Approach 1: "Filter"	Approach 2: "Search Expansion"
Subcategorise from a set of analogues identified based on structural similarity	"Frontload" both structure and physchem into analogue identification
Common approach	Novel approach

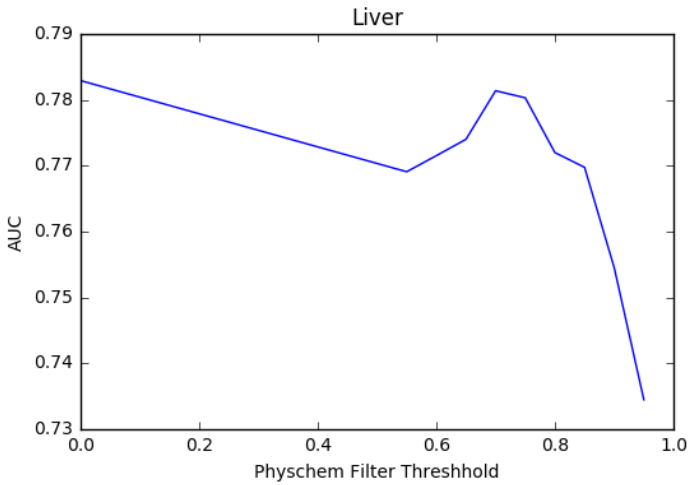
Approaches considered

United States Environmental Protection Agency

Approach 1: Filter

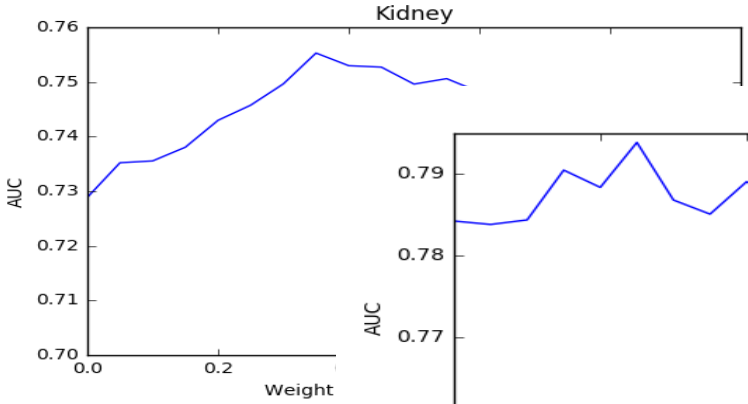
- This approach did not perform as well as GenRA for the entire dataset, nor did it significantly improve any target organ predictions.

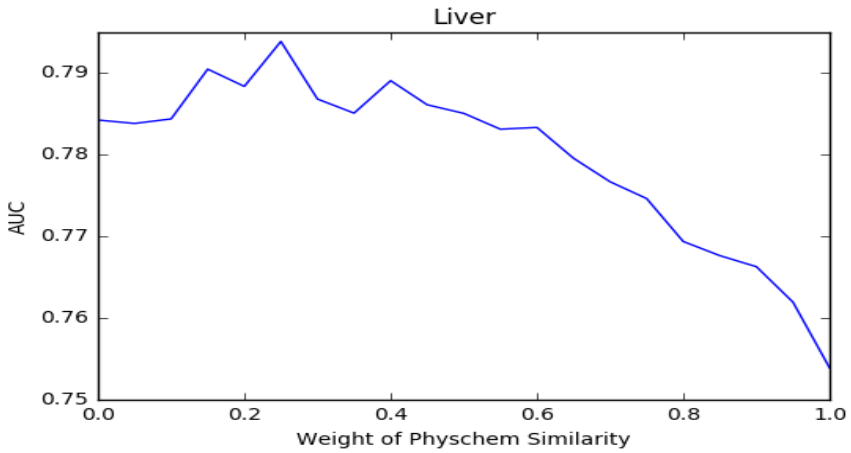




Approach 2: Search Expansion

- 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

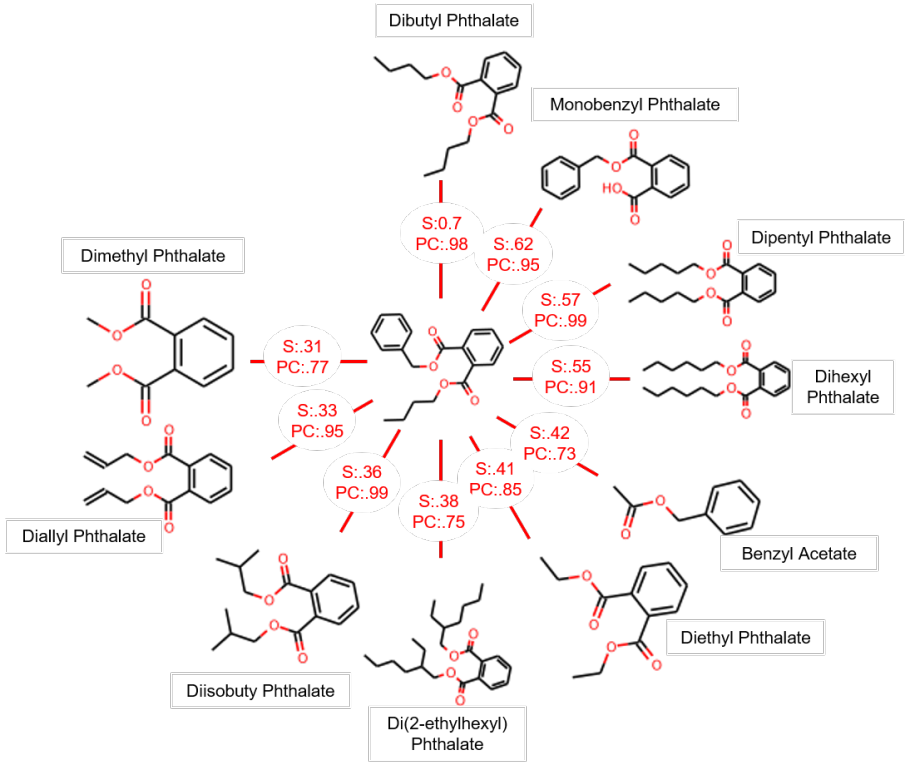




Target organ toxicity predictions aggregated over study type to organ level

Case Study: Butyl Benzyl Phthalate

GenRA: Baseline



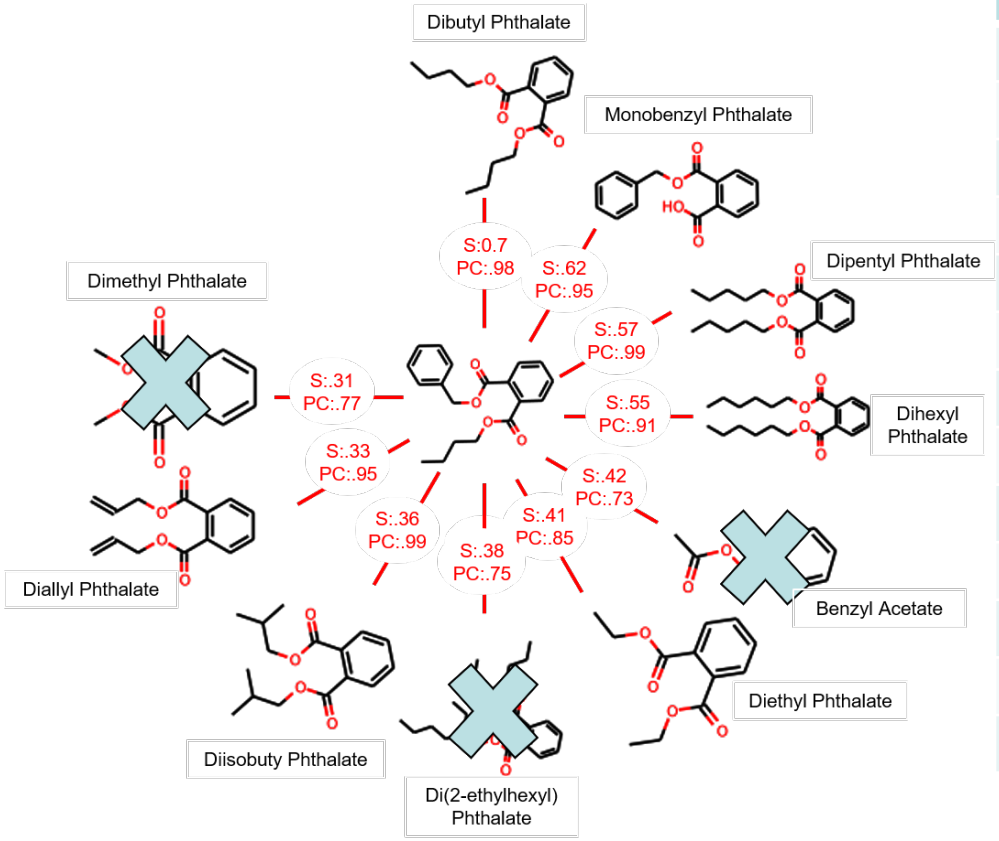
Endpoint	Baseline Prediction
Body Weight	.78
Clinical Chemistry	.27
Food Consumption	0
Hematology	0
Kidney	.27
Liver	1
Mortality	.27
Pancreas	.27
Prostate	0
Skin	.27
Spleen	0
Tissue NOS	0
Urinary Bladder	0

- 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



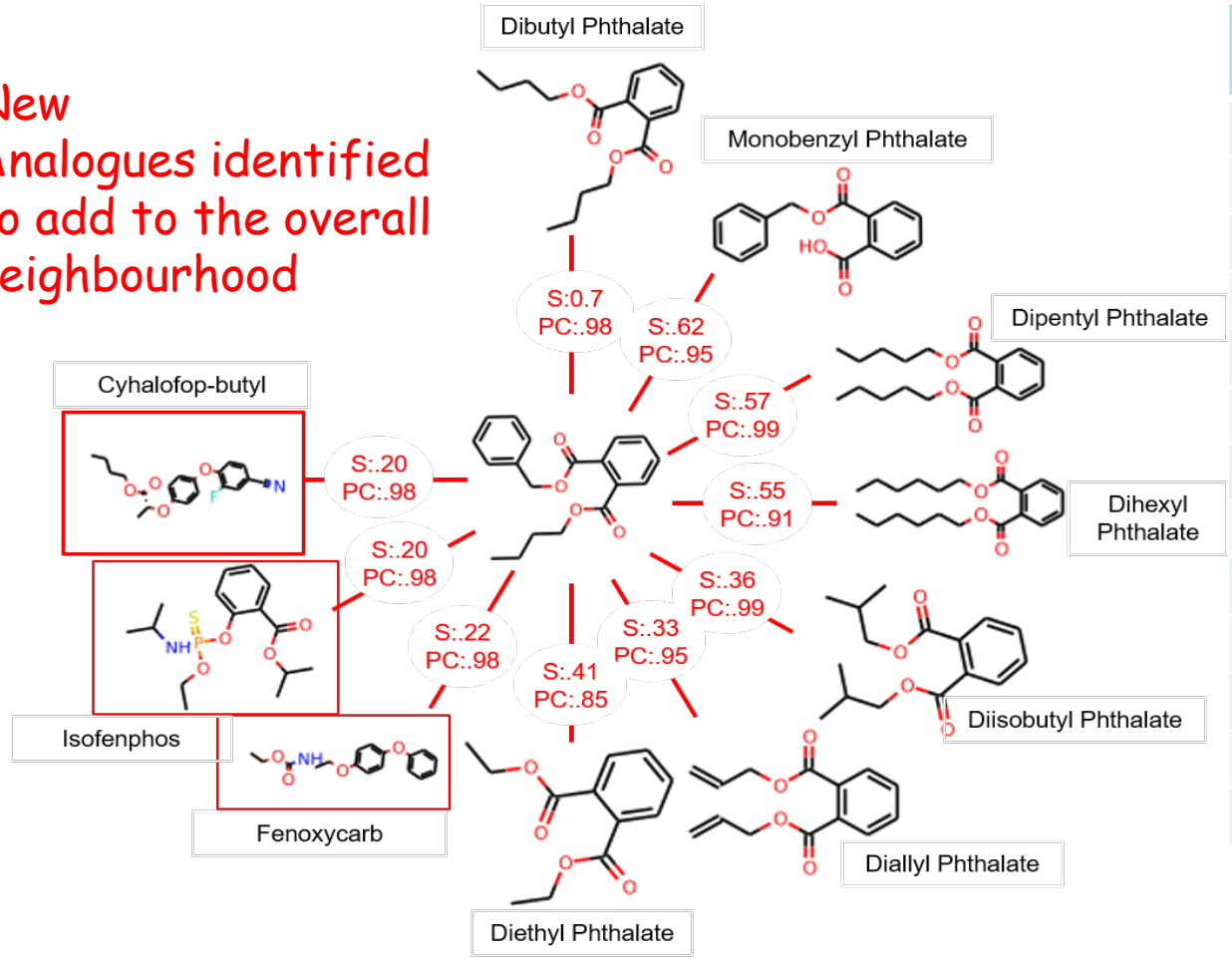
Endpoint
Body Weight
Clinical Chemistry
Food Consumption
Hematology
Kidney
Liver
Mortality
Pancreas
Prostate
Skin
Spleen
Tissue NOS
Urinary Bladder

- 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

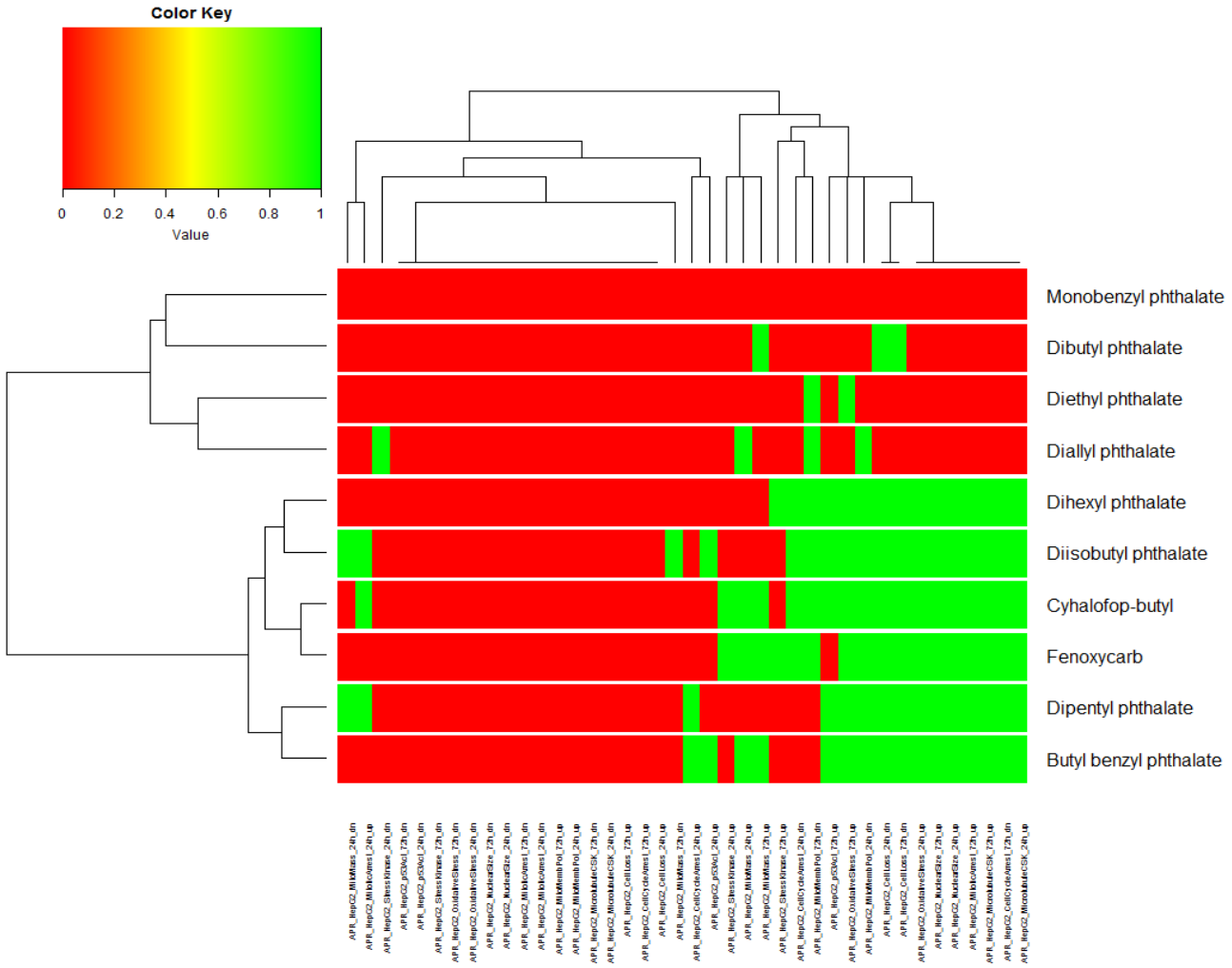


Endpoint	Baseline Prediction	Structure + Pchem Prediction
Body Weight	.78	.79
Clinical Chemistry	.27	.60
Food Consumption		
Hematology		
Kidney		
Liver		
Mortality		
Pancreas		
Prostate		
Skin		
Spleen		
Tissue NOS		
Urinary Bladder	0	0

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

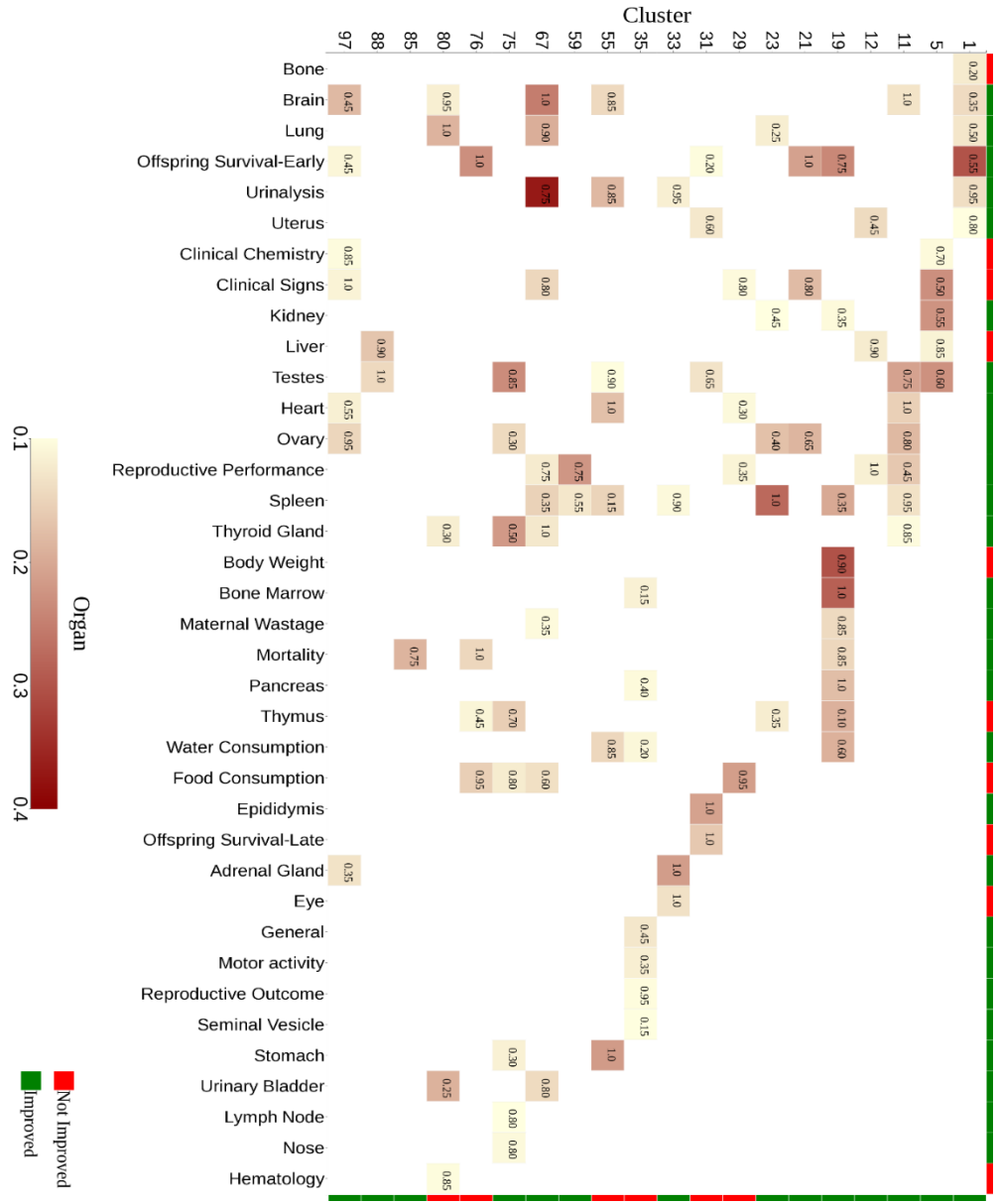
Case Study: Butyl Benzyl Phthalate

Approach 2: Search Expansion



- Are the non phthalate analogues plausible from a biological similarity?
- Heatmap of ToxCast bioactivity profiler from one (Apredica) technology
- From a qualitative perspective – these non phthalates exhibit similarity wrt their bioactivity profile to the target and other source pthalates

Search expansion in practice

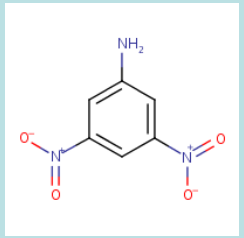
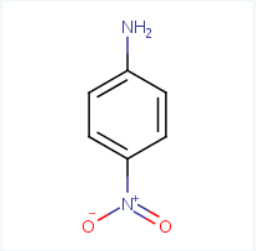


1) Identify target chemical

2) Perform Data gap analysis

3) Use cluster/organ key to guide selection of the optimal physicochemical threshold to use in source analogue identification for a specific toxicity effect of interest

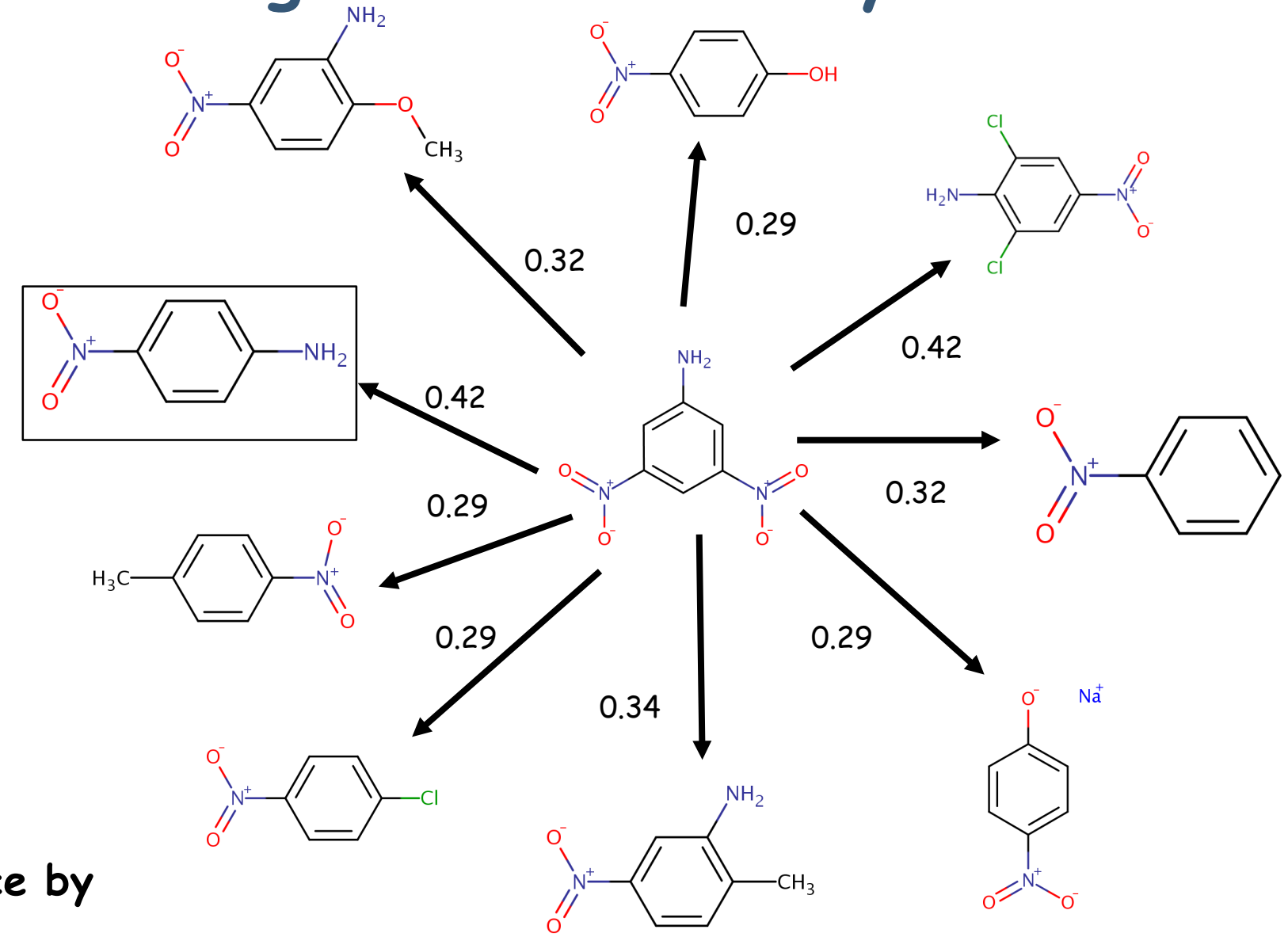
GenRA in practice for a chemical of interest to Superfund: Case study 2

Target chemical	Proposed source analogue	Primary similarity rationale
Structural		
3,5-Dinitroaniline 	4-Nitroaniline 	Considerations for chemical class, structural moiety, reactivity, metabolism and toxicity were used to refine the pool of analogues. Selection of the source analogue is based on availability of toxicity values, duration of the principal study and health protectiveness of the adopted POD, given the commonalities in the toxicokinetic and toxicity profile for all the candidates.

*Taken from SOT poster 2874/P399 from Lizarraga et al

Source analogues identified by GenRA

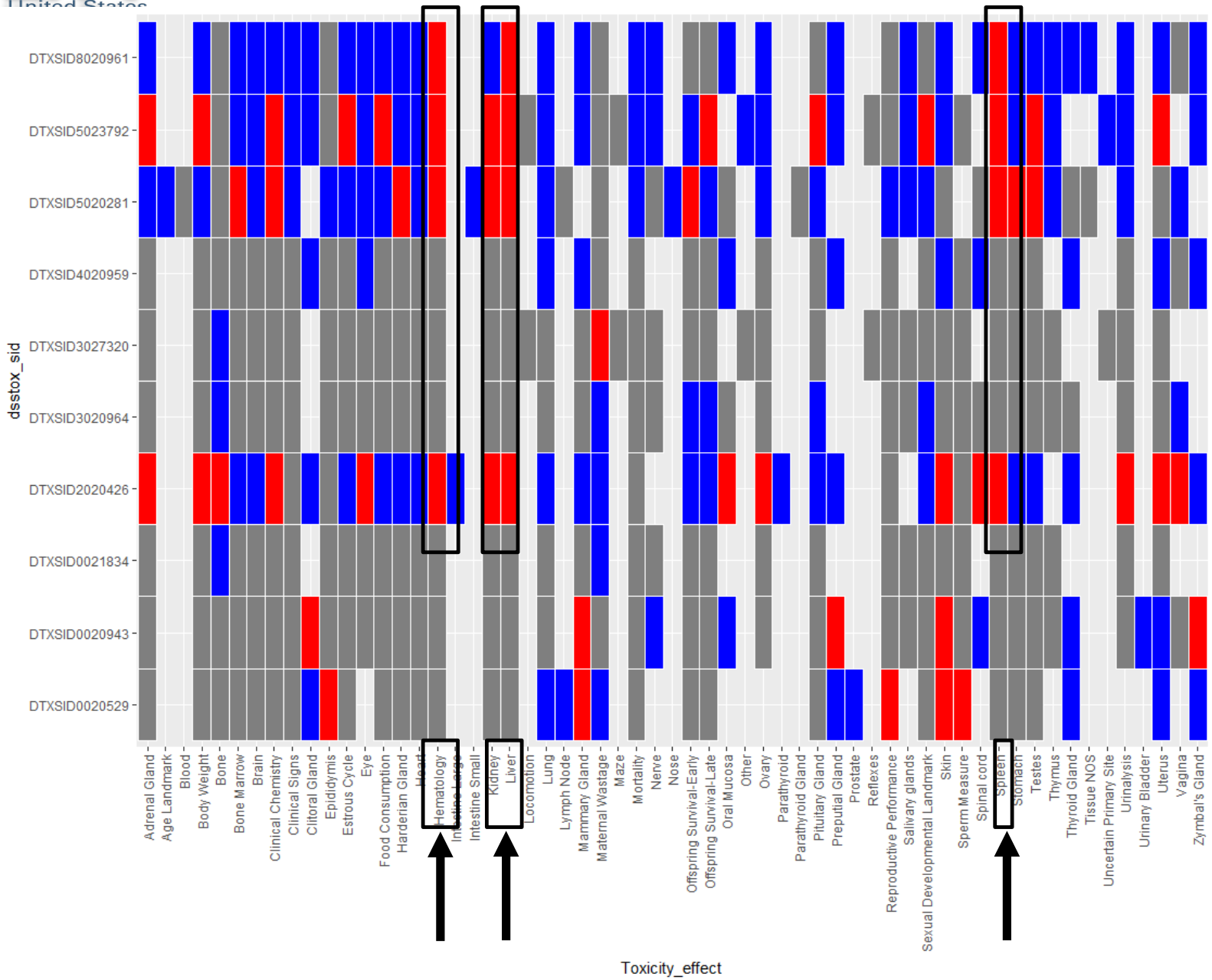
DTXSID	PREFERRED_NAME
DTXSID0020943	2-Methoxy-5-nitroaniline
DTXSID0021834	4-Nitrophenol
DTXSID0044151	**3,5-Dinitroaniline***
DTXSID2020426	Dicloran
DTXSID3020964	Nitrobenzene
DTXSID3027320	Sodium 4-nitrophenolate
DTXSID4020959	2-Methyl-5-nitroaniline
DTXSID5020281	1-Chloro-4-nitrobenzene
DTXSID5023792	4-Nitrotoluene
DTXSID8020961	**4-Nitroaniline****



*= Target ** = Proposed source by expert judgement

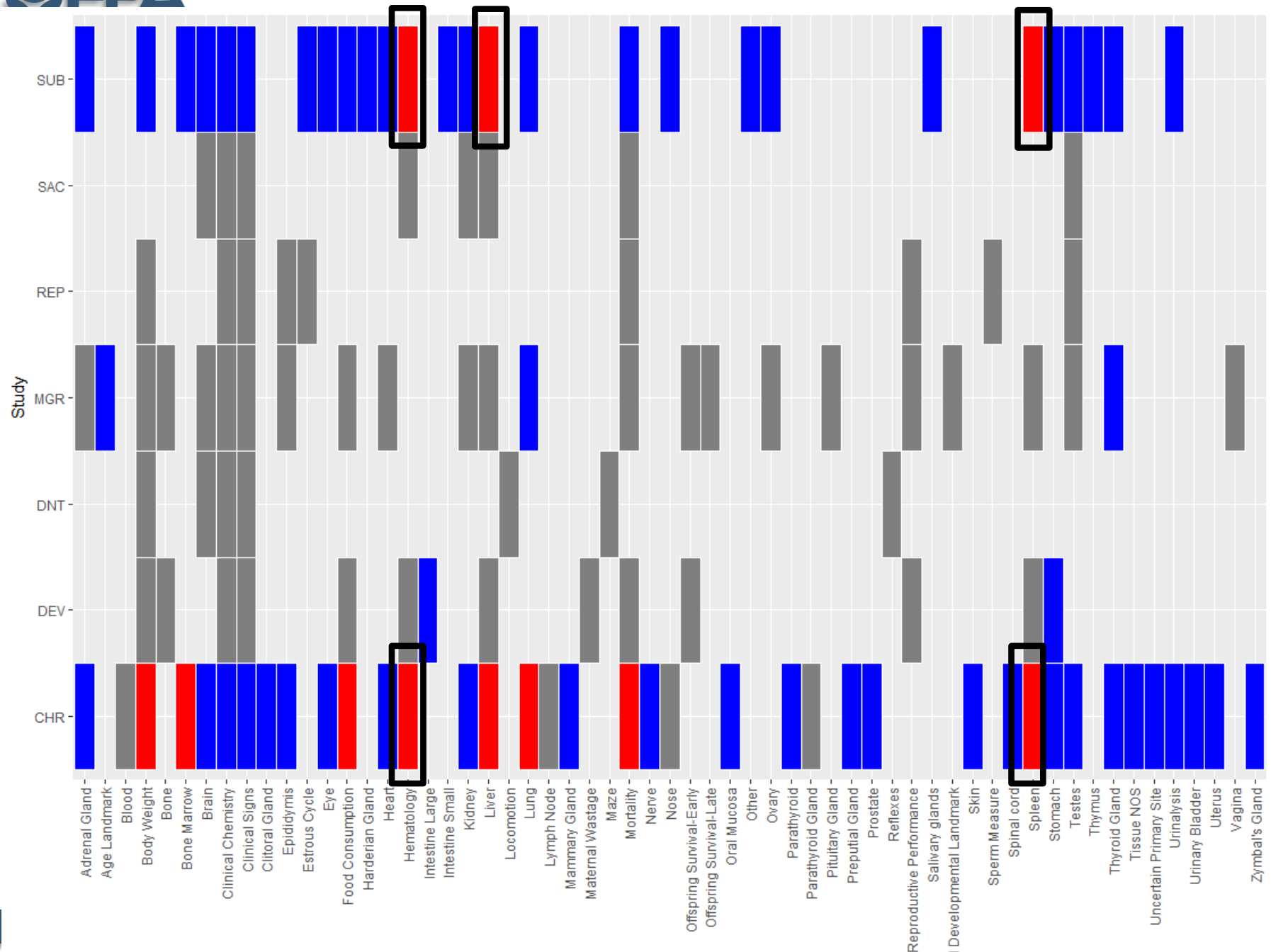
Analogues characterised by Morgan fingerprints

Data matrix for source substances



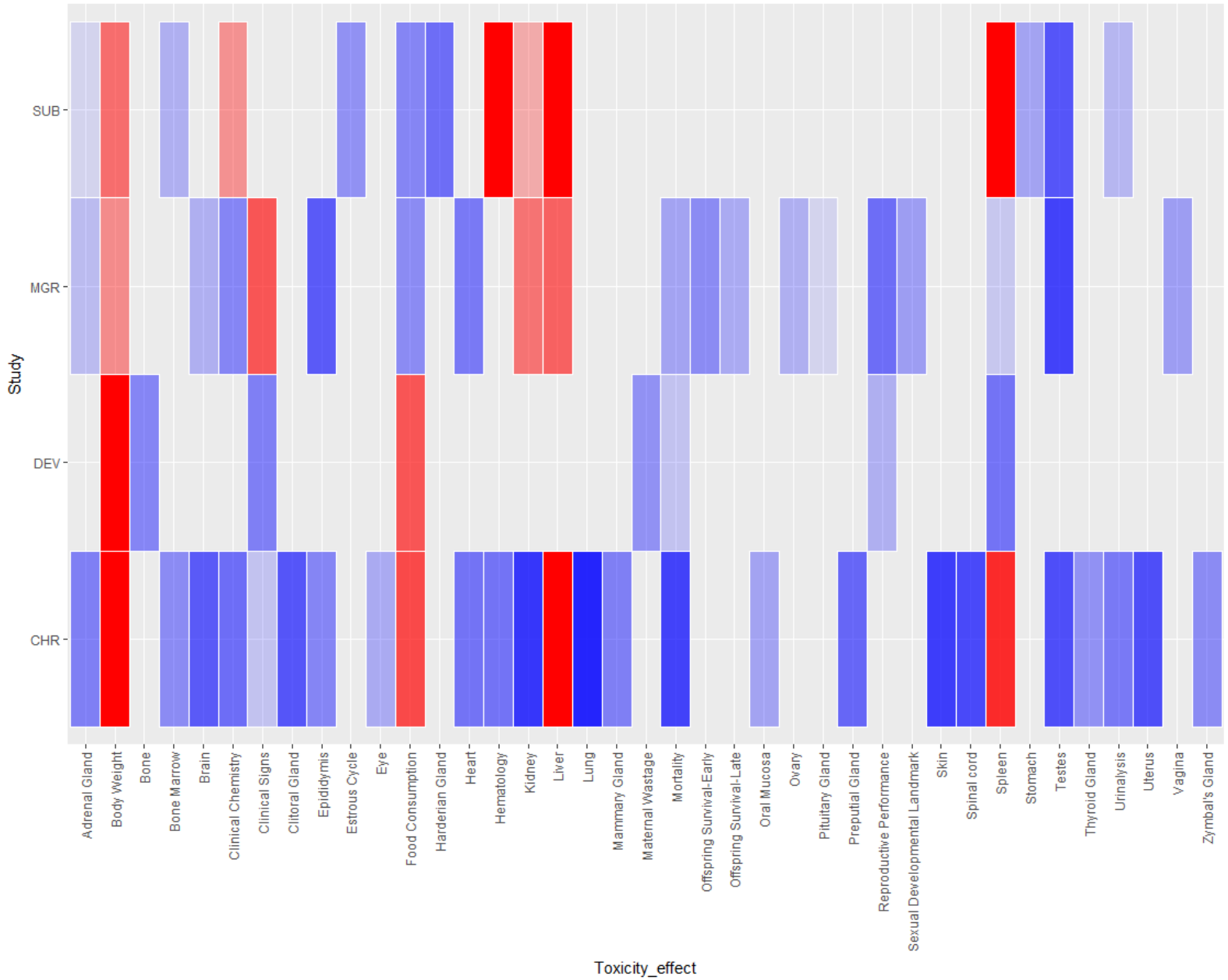
Common 'positive' effects across analogues include haematology, liver, kidney and spleen effects

Data matrix for source substances



Haematology, liver, kidney and spleen effects most readily observed in Subchronic and chronic studies across the source analogues

GenRA predictions for 3,5-dinitroaniline

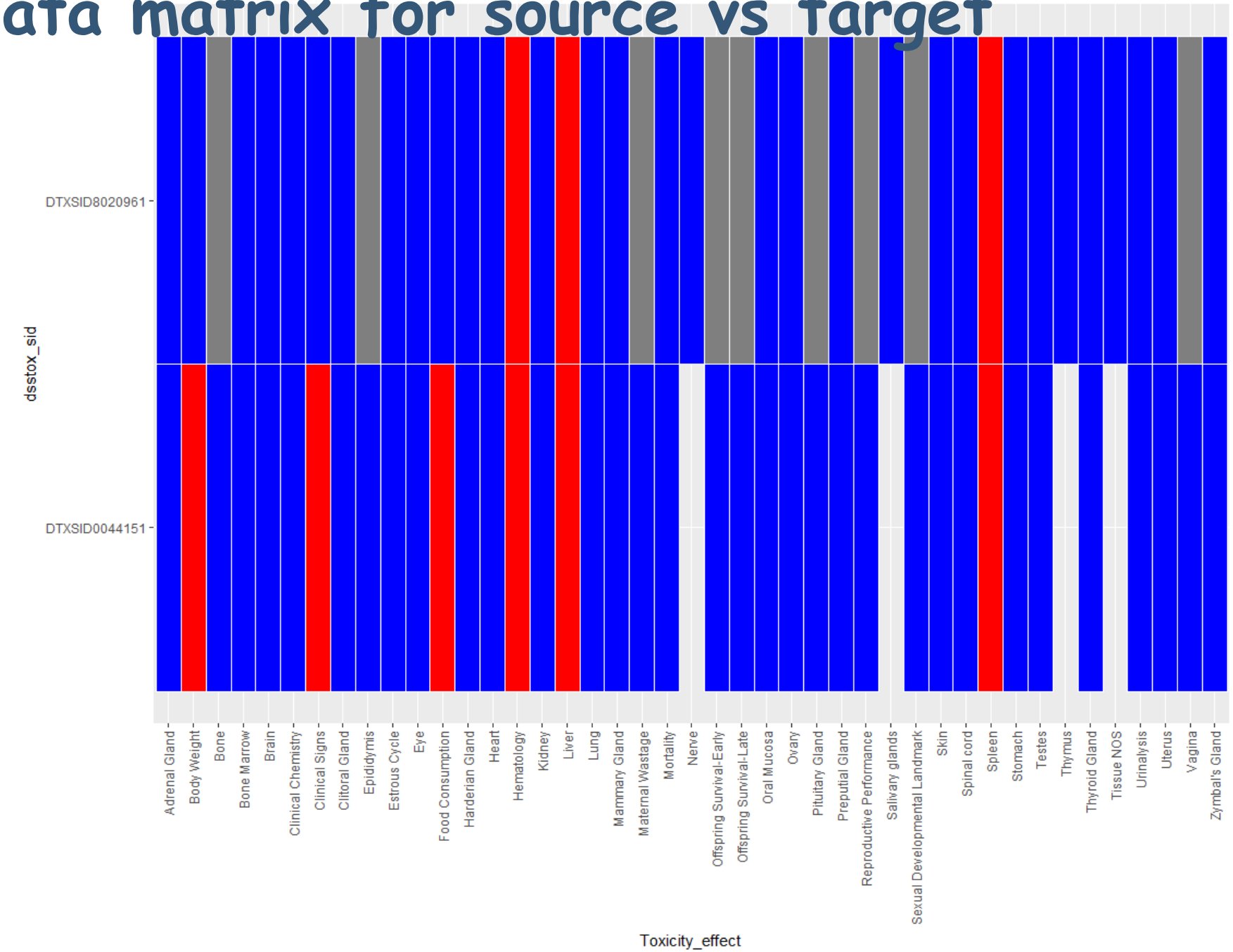


Important to take into account confidence in the predictions

Data matrix for source vs target

Proposed source analogue

3,5-dinitroaniline



Take home messages

- Several read-across development & assessment frameworks – highlighted a handful. These are largely consistent with each other.
- Proposed a harmonised framework for read-across
- GenRA developed is aligned with this framework – public tool in development
- Initial GenRA (baseline) considers structural similarity but current work has evaluated the quantitative impact of physicochemical similarity (as it relates to bioavailability)
- Highlighted the practical application of physchem similarity using a case study chemical in practice
- Illustrated how GenRA baseline has been applied in concert with expert read-across approaches for Superfund substances (see SOT poster)

Acknowledgements

- Many but in particular..
- NCCT
- Imran Shah
- George Helman
- Tony Williams
- Jeff Edwards

- NCEA
- Lucy Lizarraga
- Jason Lambert
- Scott Wesselkamper