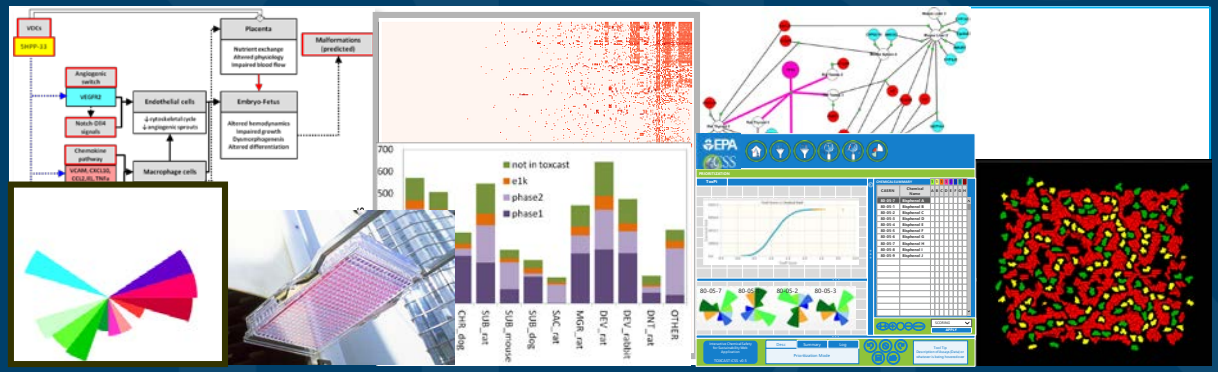
United States Environmental Protection Agency

Current and Future approaches for Read-Across: From Research to Practical Implementation



GlobalChem 2017: Friday 24th February 2017

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

Outline

- Background & 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)

Background & definitions

- The grouping assumes chemicals are similar.."the quality of being alike" ……

- A chemical category is a group of chemicals whose physico-chemical and human heath 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).

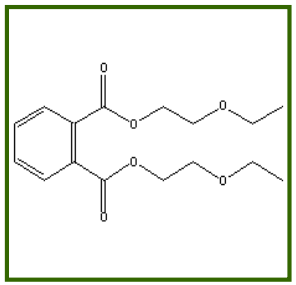
Definition: Read-across

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	●	○

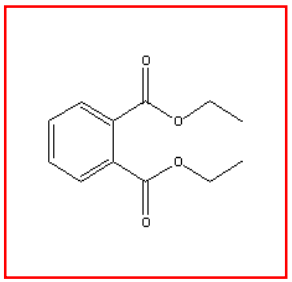
● Reliable data

○ Missing data



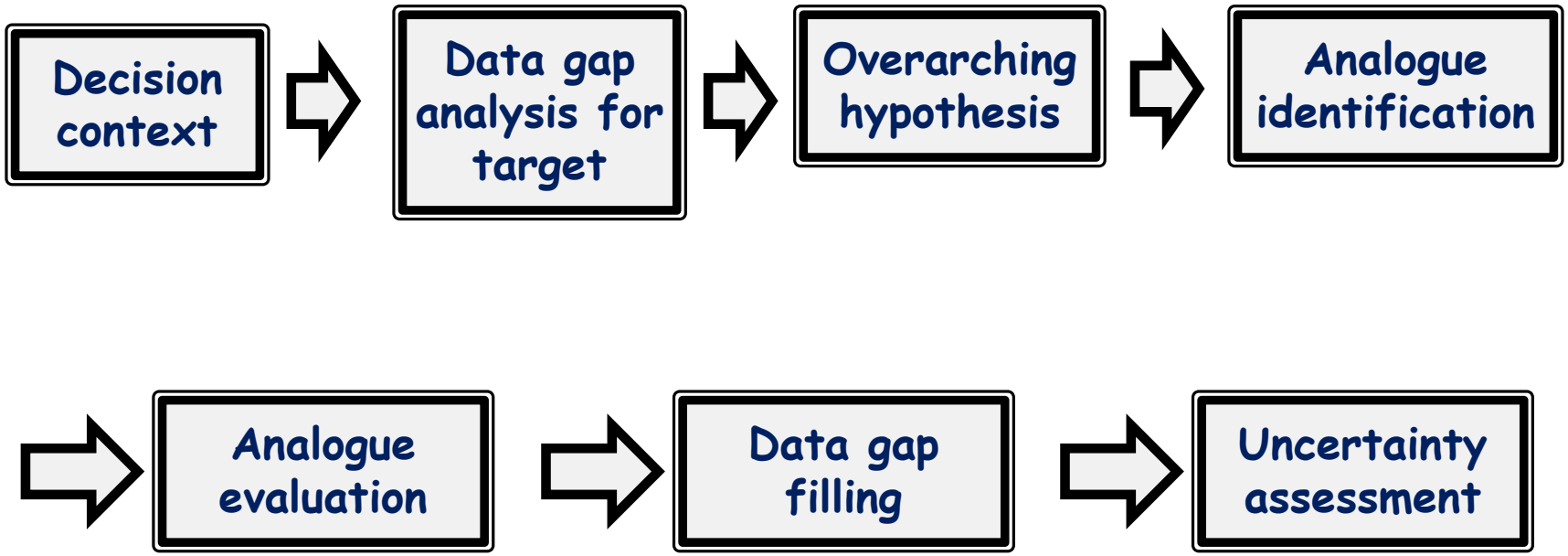
Known to be harmful

Acute fish toxicity?

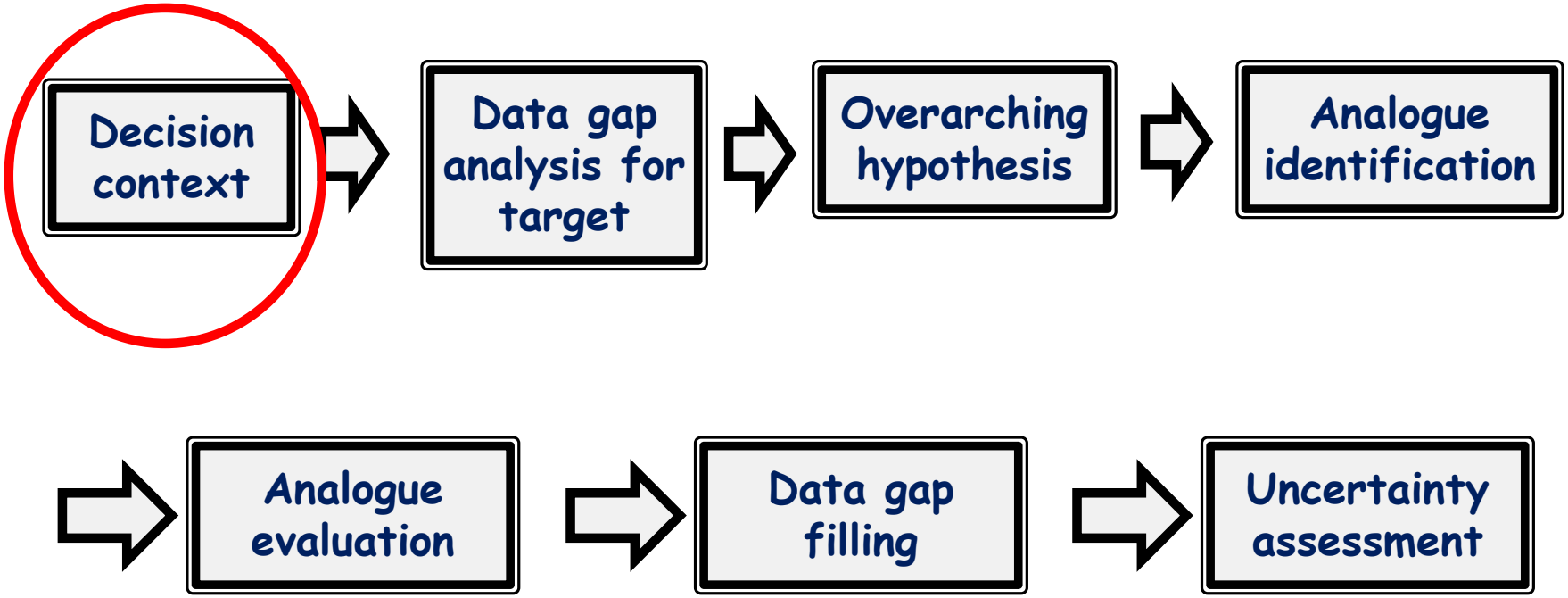


Predicted to be harmful

The Category Workflow

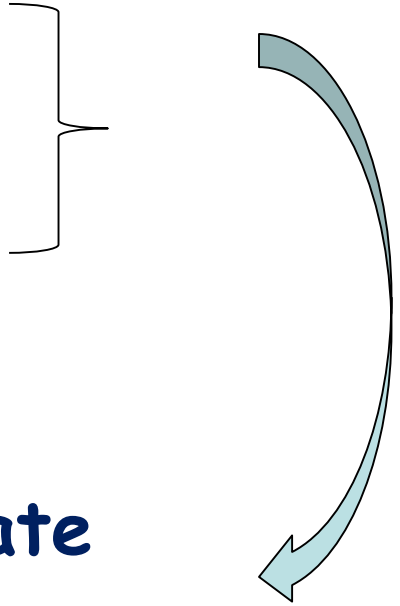


The Category Workflow



Decision Context

- Prioritisation, e.g. PMN
- Screening level hazard assessment
- Risk Assessment, e.g. PPRTV

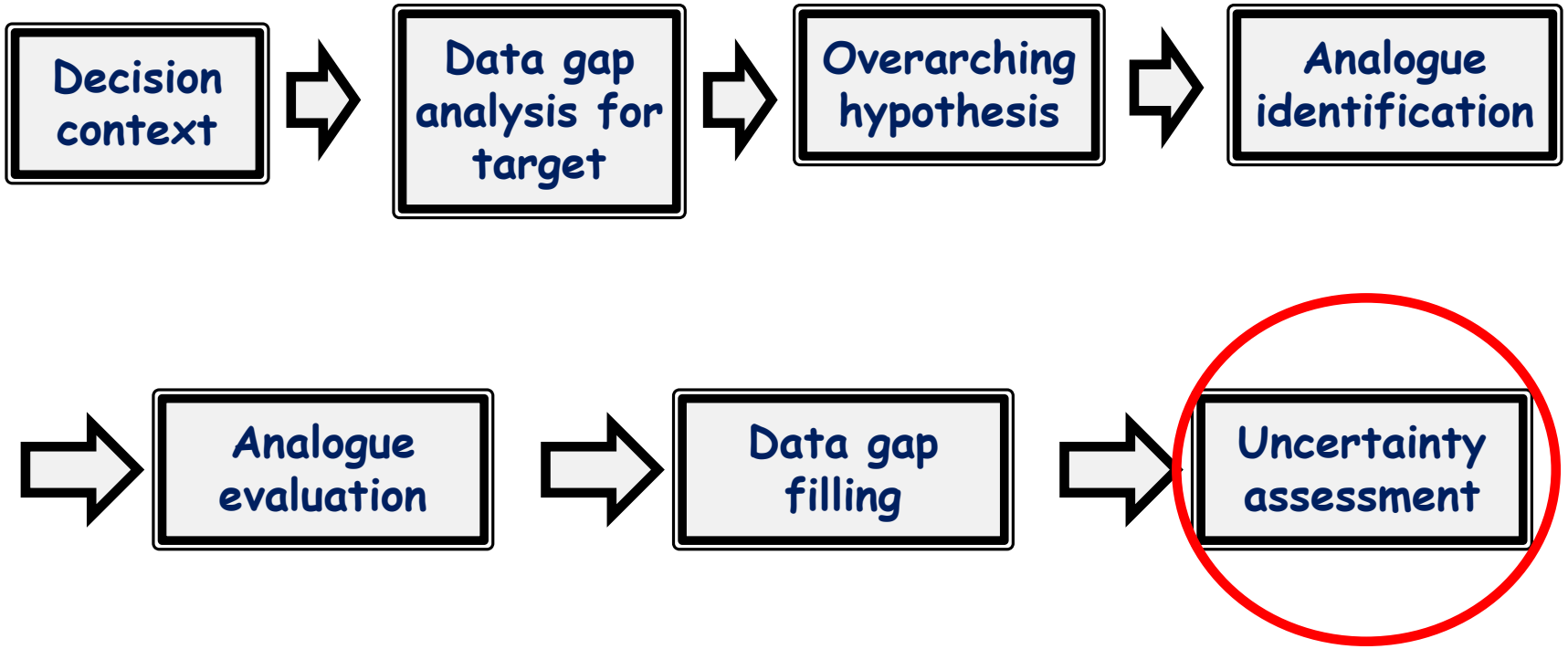


- Different decision contexts will dictate the level of uncertainty that can be tolerated

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

The Category Workflow



EPA United States Environmental Protection Agency

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
- 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
- Uncertainty assessment is qualitative in nature
- There is no objective measure of read-across performance

Critical need is an objective measure of uncertainty in a read-across prediction

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
- 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:

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

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

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

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 assays (bio)
ToxCast
574 Apical outcomes (tox)
ToxRefDB

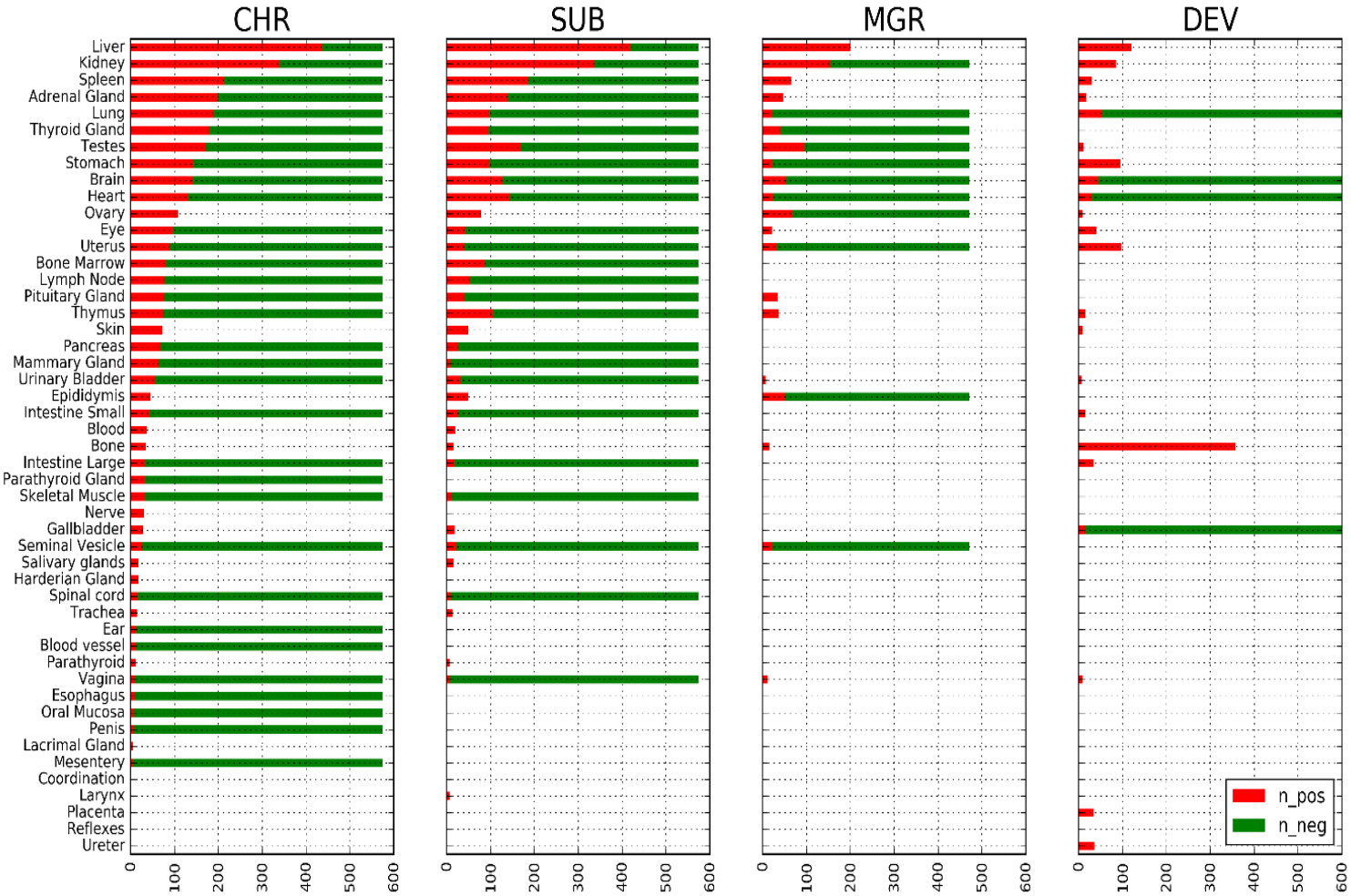
II. Define Local neighborhoods

Us 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 neighbor hoods
Evaluate impact descriptors (chm, bio, bc) on prediction
Quantify uncertainty

GenRA - Toxicity Data from ToxRefDB

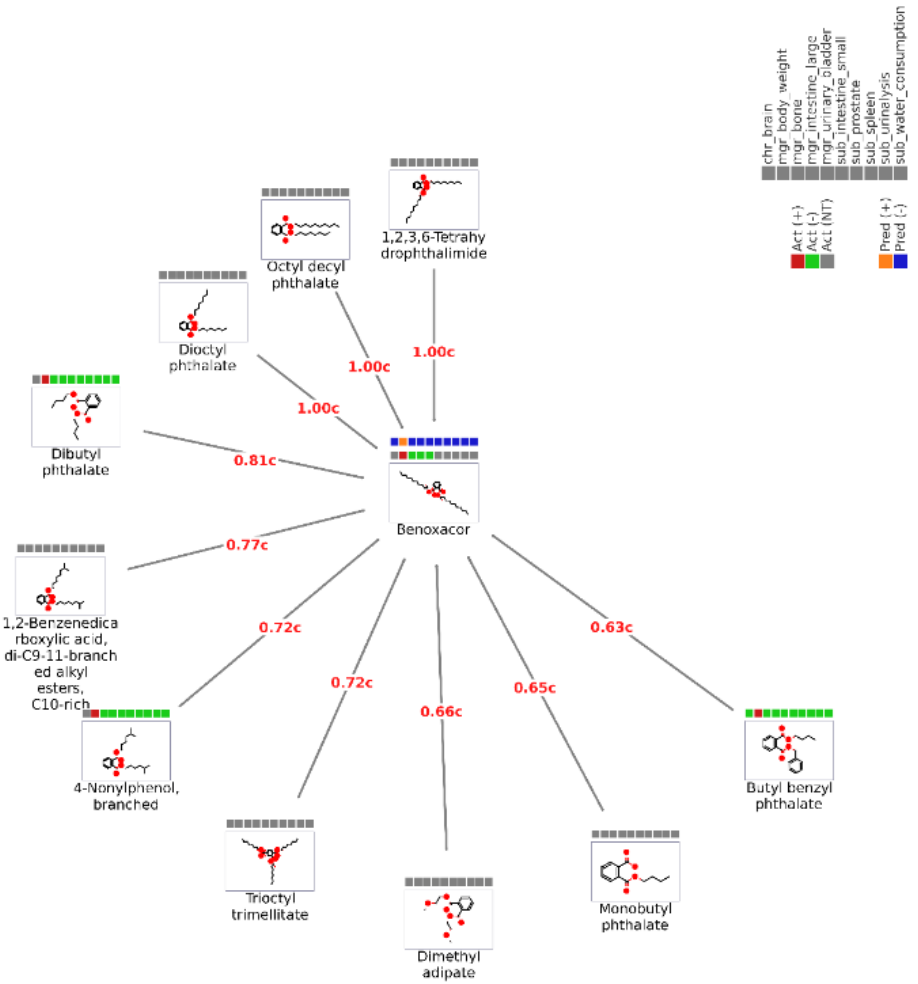


GenRA: Clustering chemicals

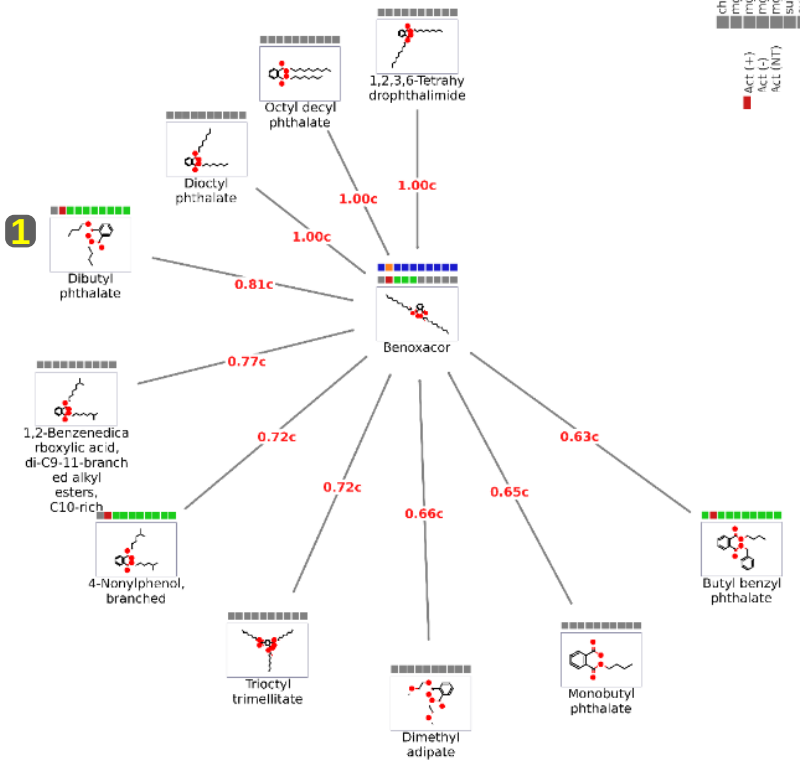
Shah et al, 2006

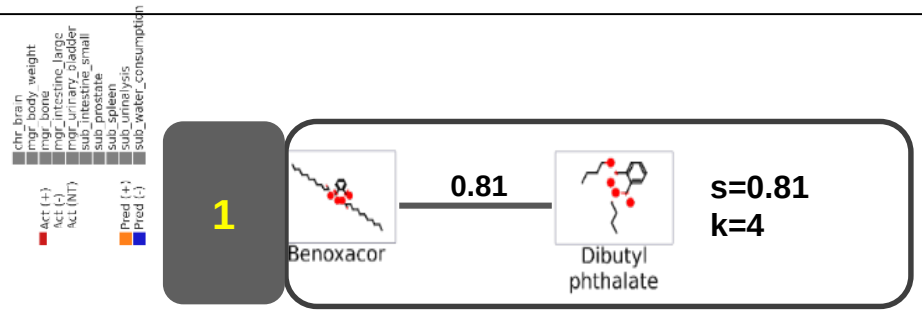
GenRA – Performance in Each Cluster

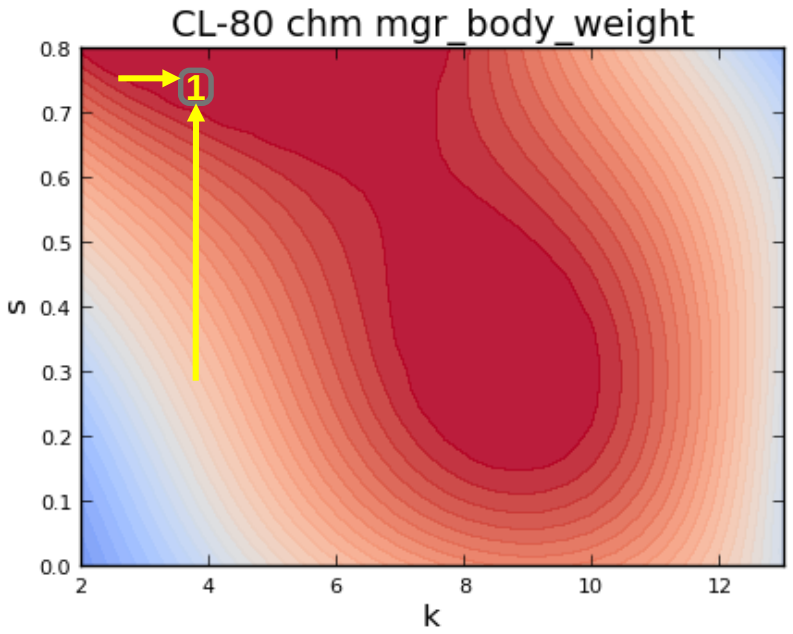
- Use GenRA to predict the similarity weighted toxicity scores for each
  - Toxicity type (β)
  - Descriptor ={chm,bio,bc} (α)
  - No. of nearest neighbors (k)
  - Similarity score threshold (s_{ij}^{α})
- Calculate performance by comparing predicted y^{tox} and true x^{tox} for all chemicals using area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



GenRA - Analysing Local Neighborhood of a Chemical



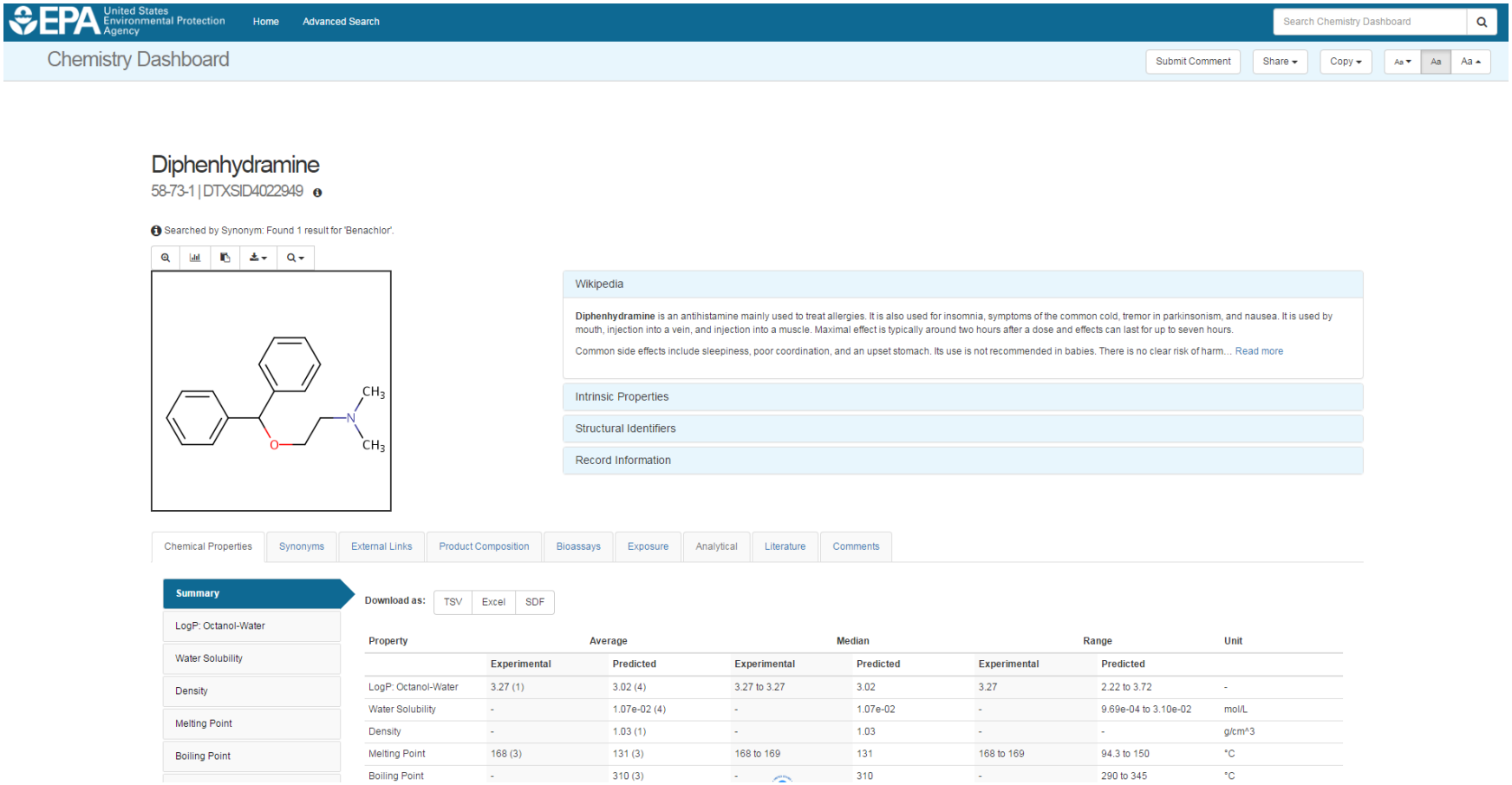




GenRA – Insights and Next Steps

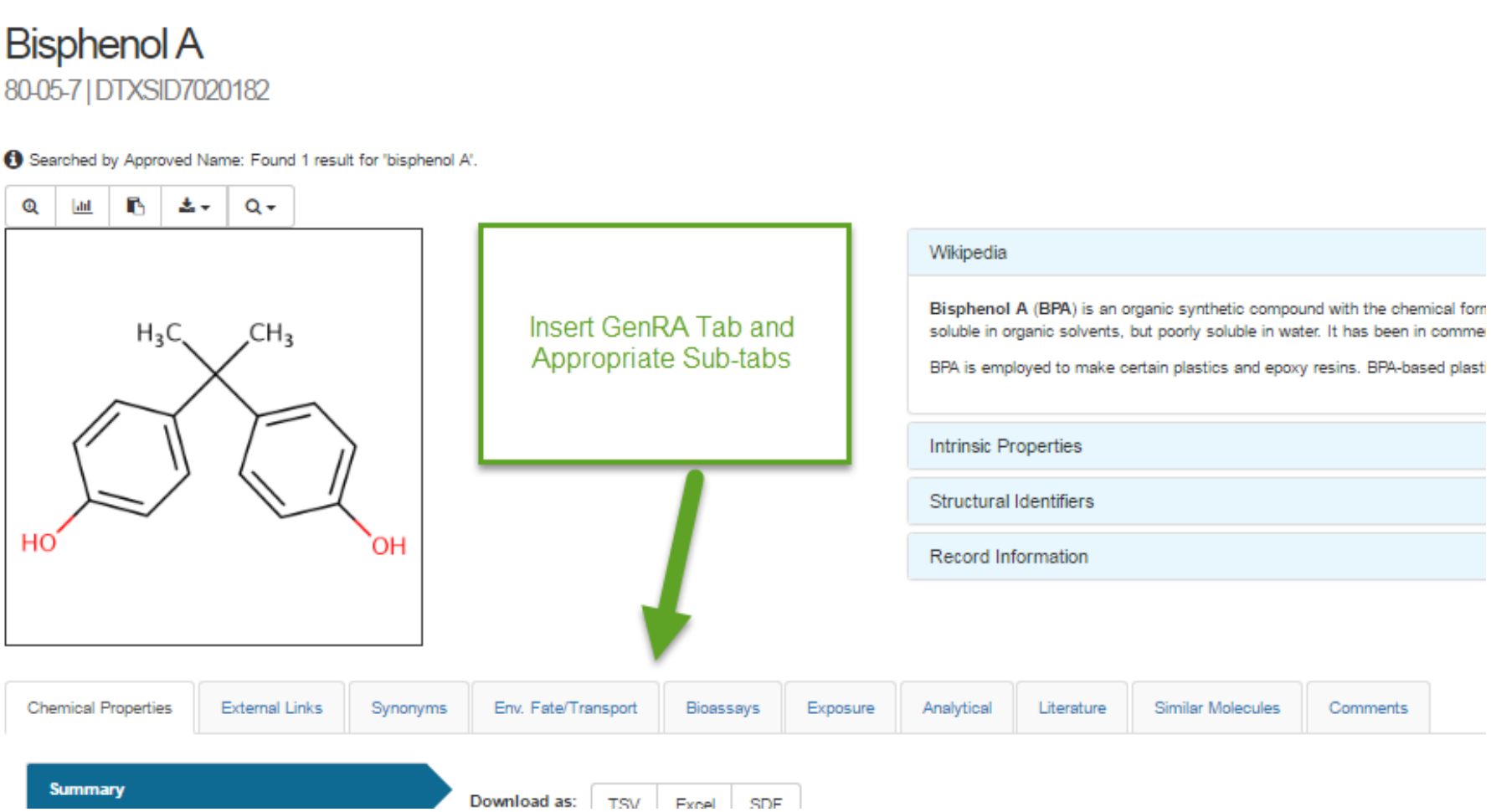
- The approach enabled a performance baseline for read-across predictions of specific study outcomes to be established but was still context dependent on the endpoint and the chemical
- Bioactivity descriptors were often found to be more predictive of *in vivo* toxicity outcomes
- Ongoing analysis:
- Consideration of other information to refine the analogue selection – e.g. TK similarity, metabolic similarity, reactivity similarity…

From research to implementation

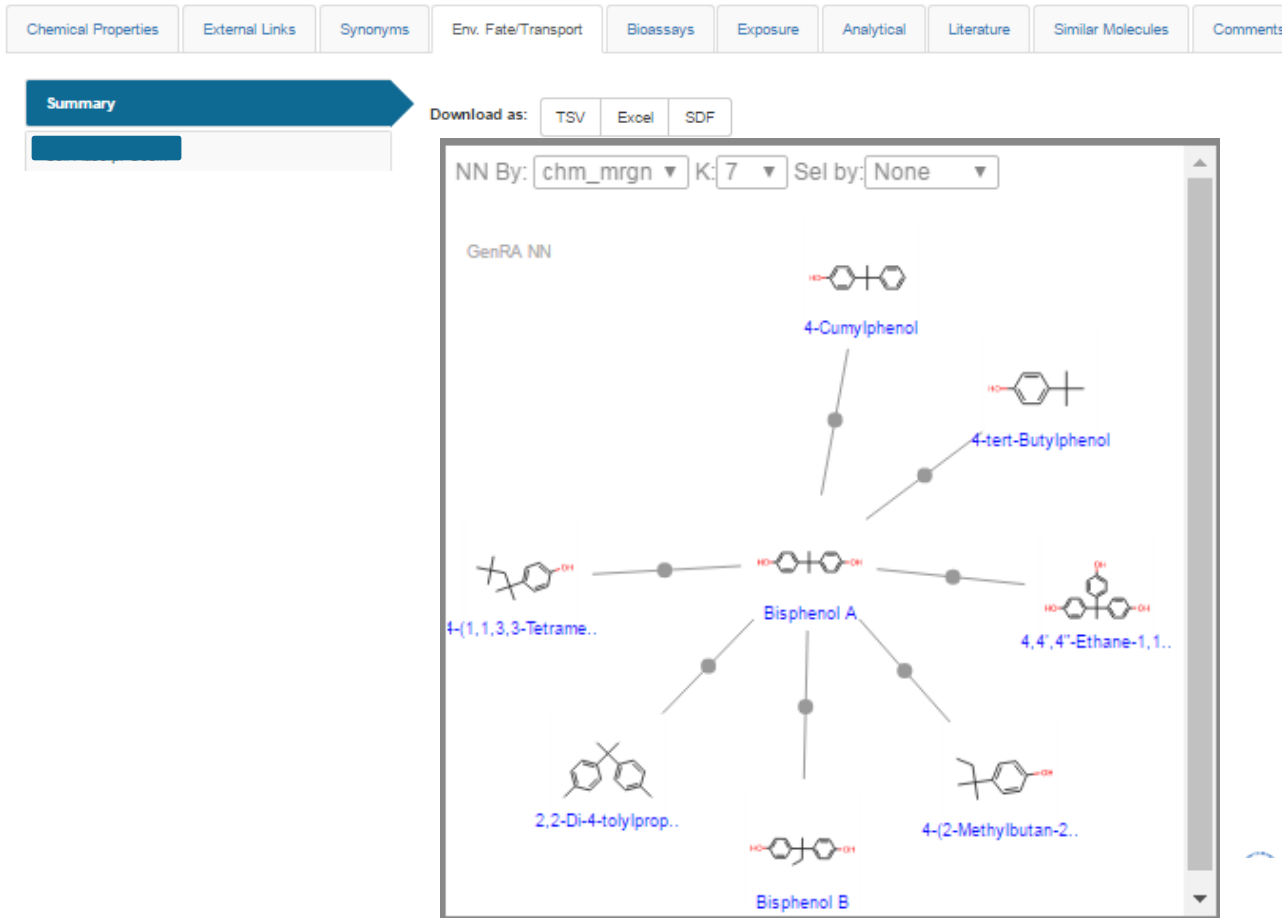


Property	Average		Median		Range		Unit
	Experimental	Predicted	Experimental	Predicted	Experimental	Predicted	
LogP: Octanol-Water	3.27 (1)	3.02 (4)	3.27 to 3.27	3.02	3.27	2.22 to 3.72	-
Water Solubility	-	1.07e-02 (4)	-	1.07e-02	-	9.69e-04 to 3.10e-02	mol/L
Density	-	1.03 (1)	-	1.03	-	-	g/cm^3
Melting Point	168 (3)	131 (3)	168 to 169	131	168 to 169	94.3 to 150	°C
Boiling Point	-	310 (3)	-	310	-	290 to 345	°C

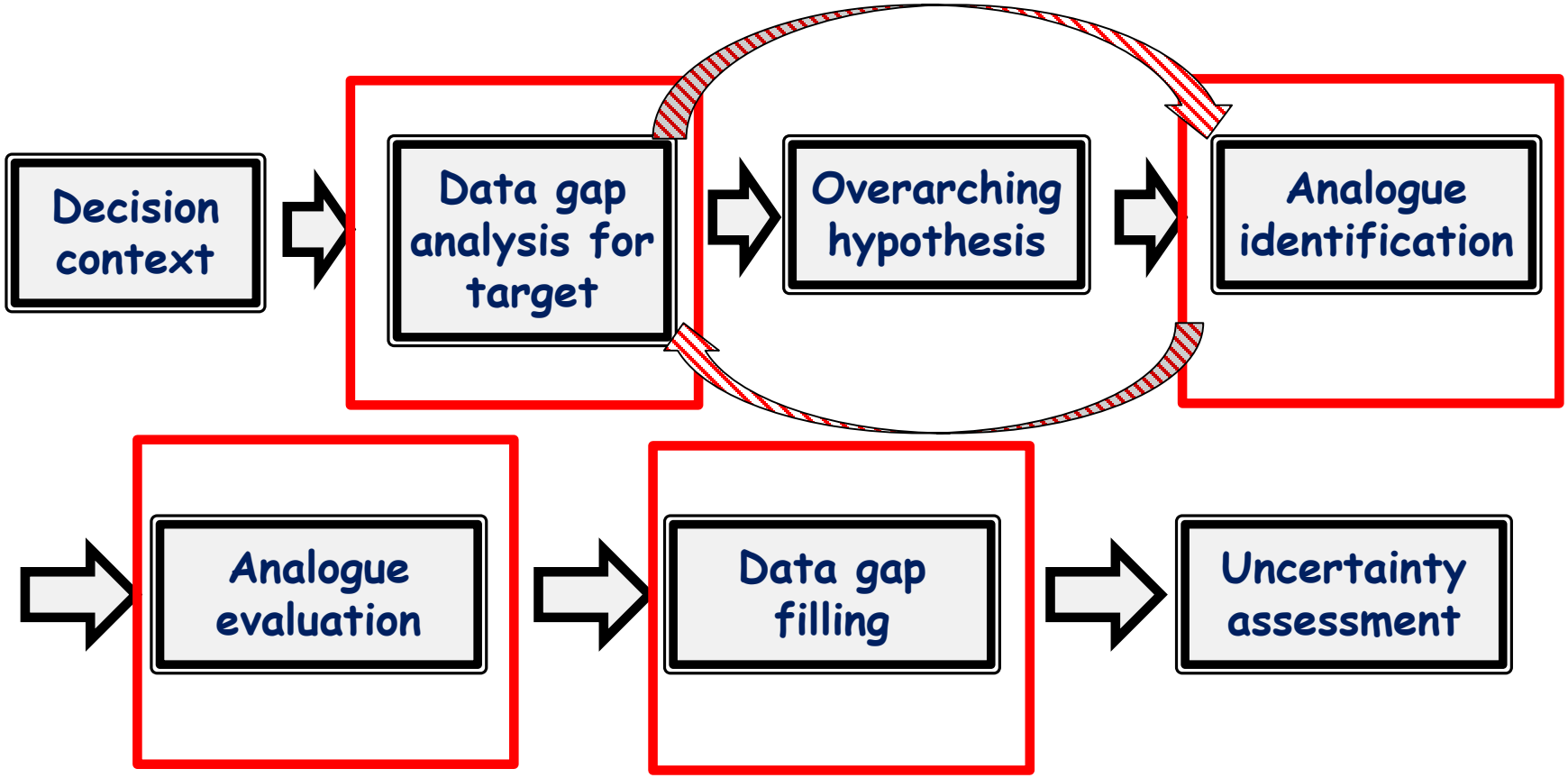
Basic Integration via GenRA tab



Tabs of associated views..



GenRA prototype development



GenRA prototype development

- **Analogue identification**
  - Identifying potential source analogues based on various approaches to characterise chemical structural information or bioactivity profile from ToxCast data
- **Data gap analysis**
  - Exploring the data availability for the target and source analogues to determine whether the source analogues are a promising starting set

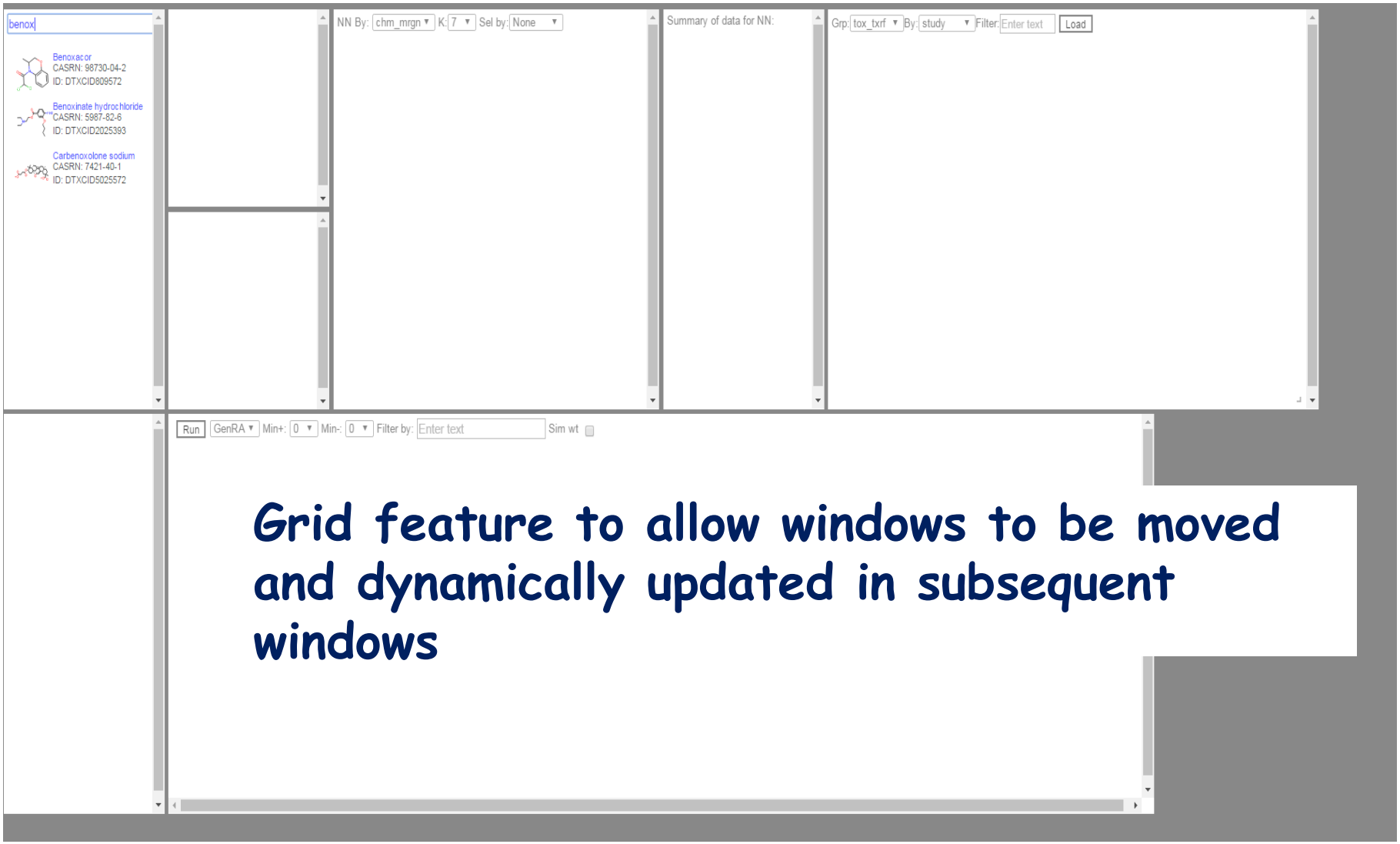
GenRA prototype development

- **Analogue evaluation**
  - Launch a data matrix of the source analogues and target with the available data, colour coded by presence and absence of effects and dose descriptor information to enable an evaluation of consistency and concordance across category members and across toxicity effects
- **Data gap filling**
  - Perform a GenRA (read-across) prediction. Arbitrary thresholds chosen to determine positive, negative and indeterminate calls. Overladen with experimental outcomes to compare predicted/actual where available

GenRA prototype development

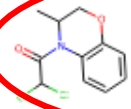
- **Uncertainty assessment**
  - Uncertainty assessment is a feature currently being developed for implementation

Initial interface

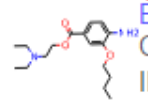


Grid feature to allow windows to be moved and dynamically updated in subsequent windows

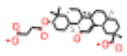
benox



Benoxacor
CASRN: 98730-04-2
ID: DTXCID809572

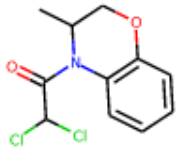


Benoxinate hydrochloride
CASRN: 5987-82-6
ID: DTXCID2025393



Carbenoxolone sodium
CASRN: 7421-40-1
ID: DTXCID5025572

Benoxacor | 98730-04-2



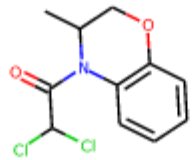
dsstox_cid:DTXCID809572
dsstox_sid:DTXSID3029572
gsid:29572
inchi_key:PFJJMJDEVDLPNE-UHFFFAOY
iupac:2,2-dichloro-1-(3-methyl-2,3-dihydro-
mol_weight:260.1165466308594
pubchem_cid:62306
smiles:CC1COC2=C(C=CC=C2)N1C(=O)C
tag:TOXCST,TOXREF,CTD

NN By: chm_mrgn ▼ K: 7 ▼ Sel by: None

Introduce and select a target chemical

Run GenRA ▼ Min+: 0 ▼ Min-: 0 ▼ Filter by: Enter text

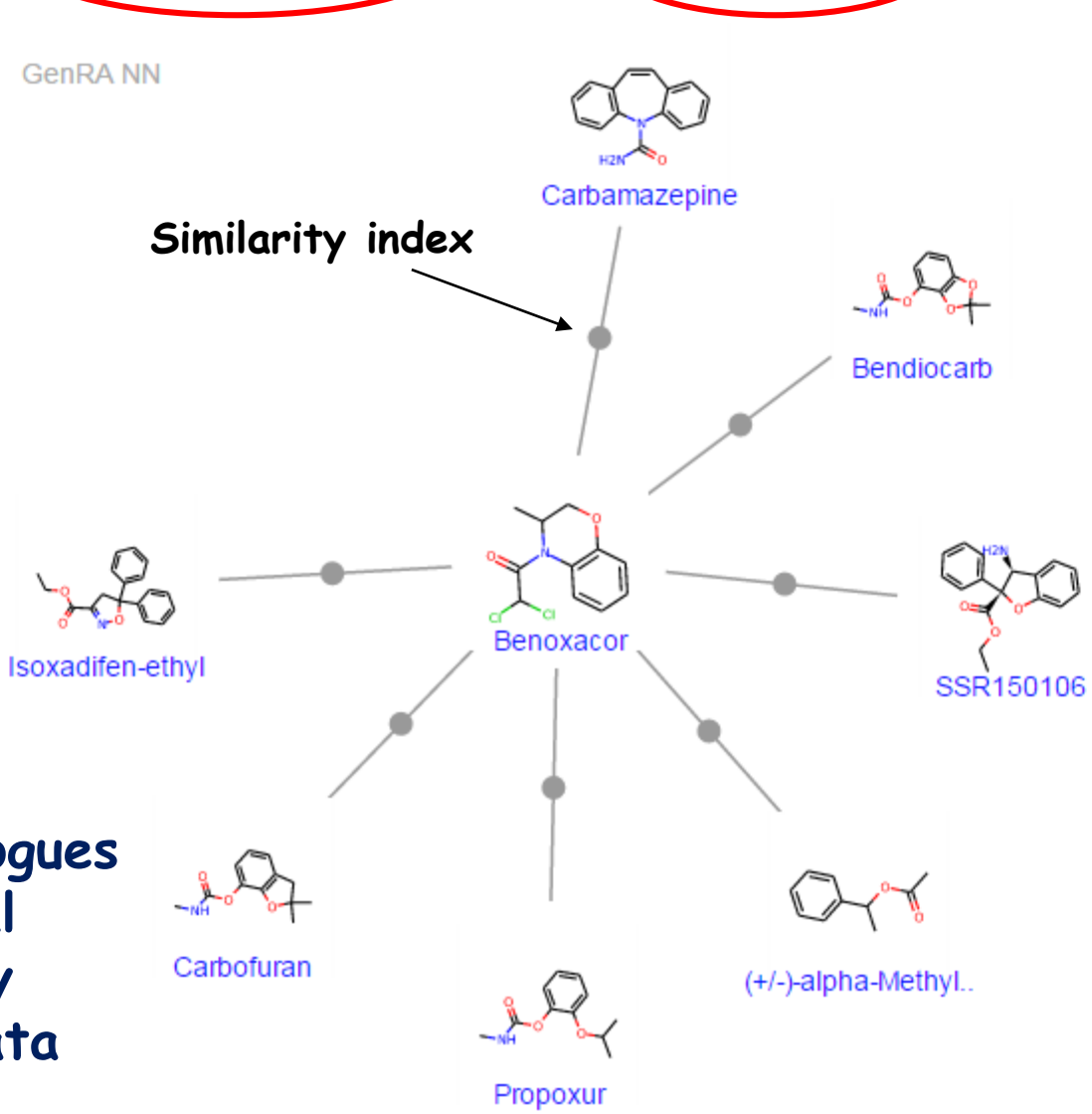
Benoxacor | 98730-04-2



dsstox_cid:DTXCID809572
dsstox_sid:DTXSID3029572
gsid:29572
inchi_key:PFJJMJDEVDLPNE-UHFFFAOY
iupac:2,2-dichloro-1-(3-methyl-2,3-dihydro-
mol_weight:260.1165466308594
pubchem_cid:62306
smiles:CC1COC2=C(C=CC=C2)N1C(=O)C
tag:TOXCST,TOXREF,CTD

NN By: chm_mrgn ▼ K: 7 ▼ Sel by: tox_txrf ▼

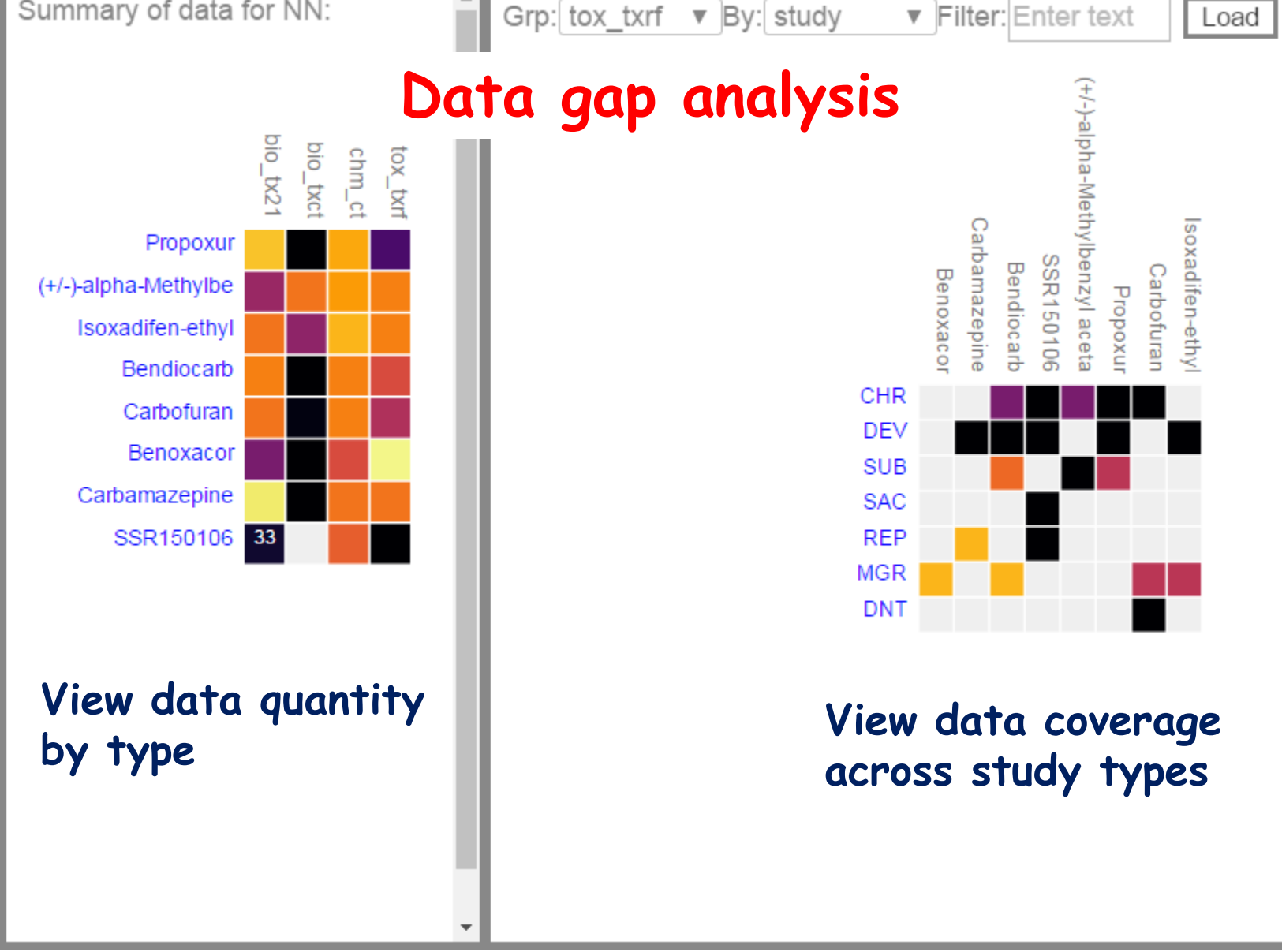
GenRA NN



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

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

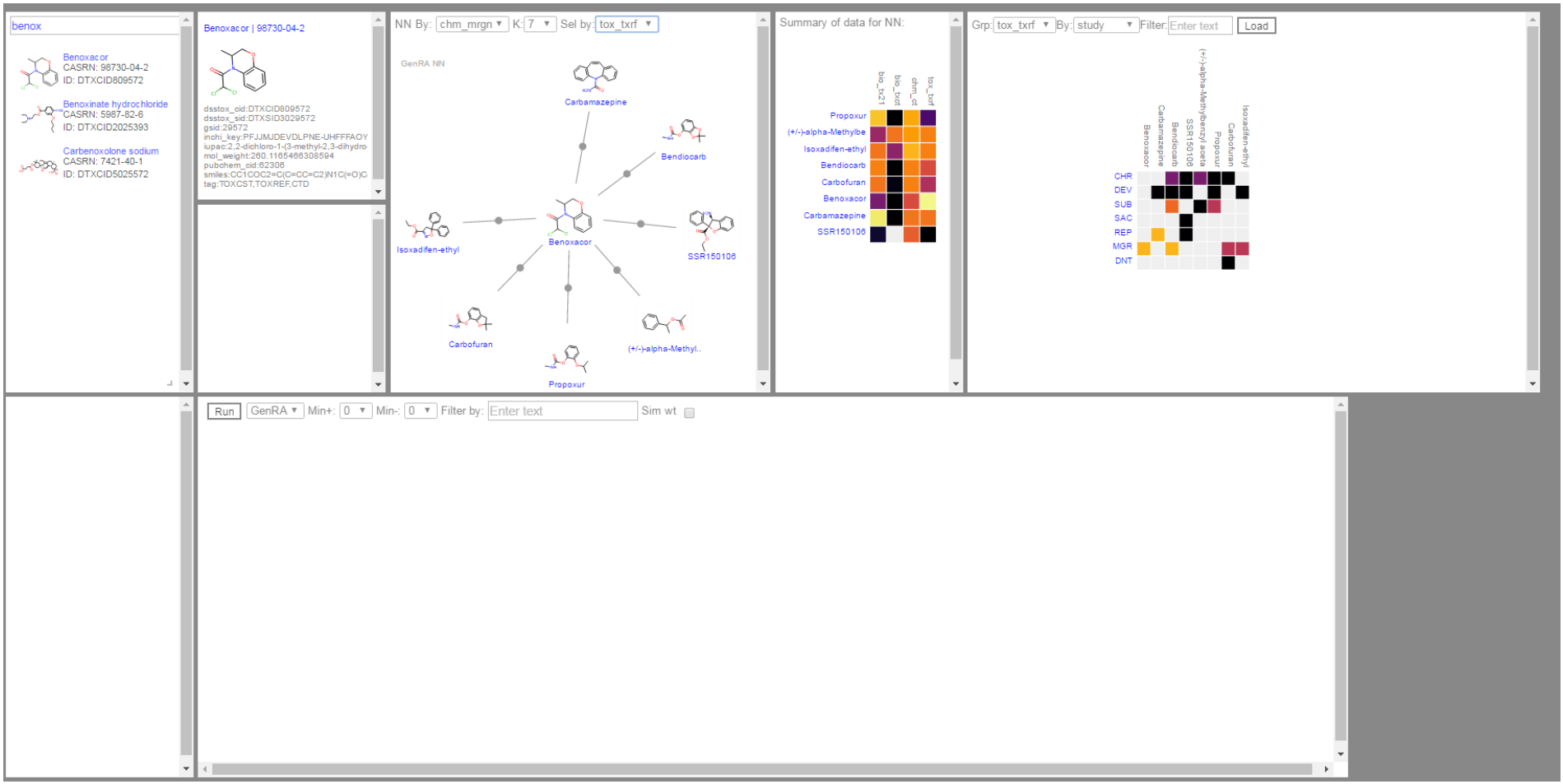
Data gap analysis



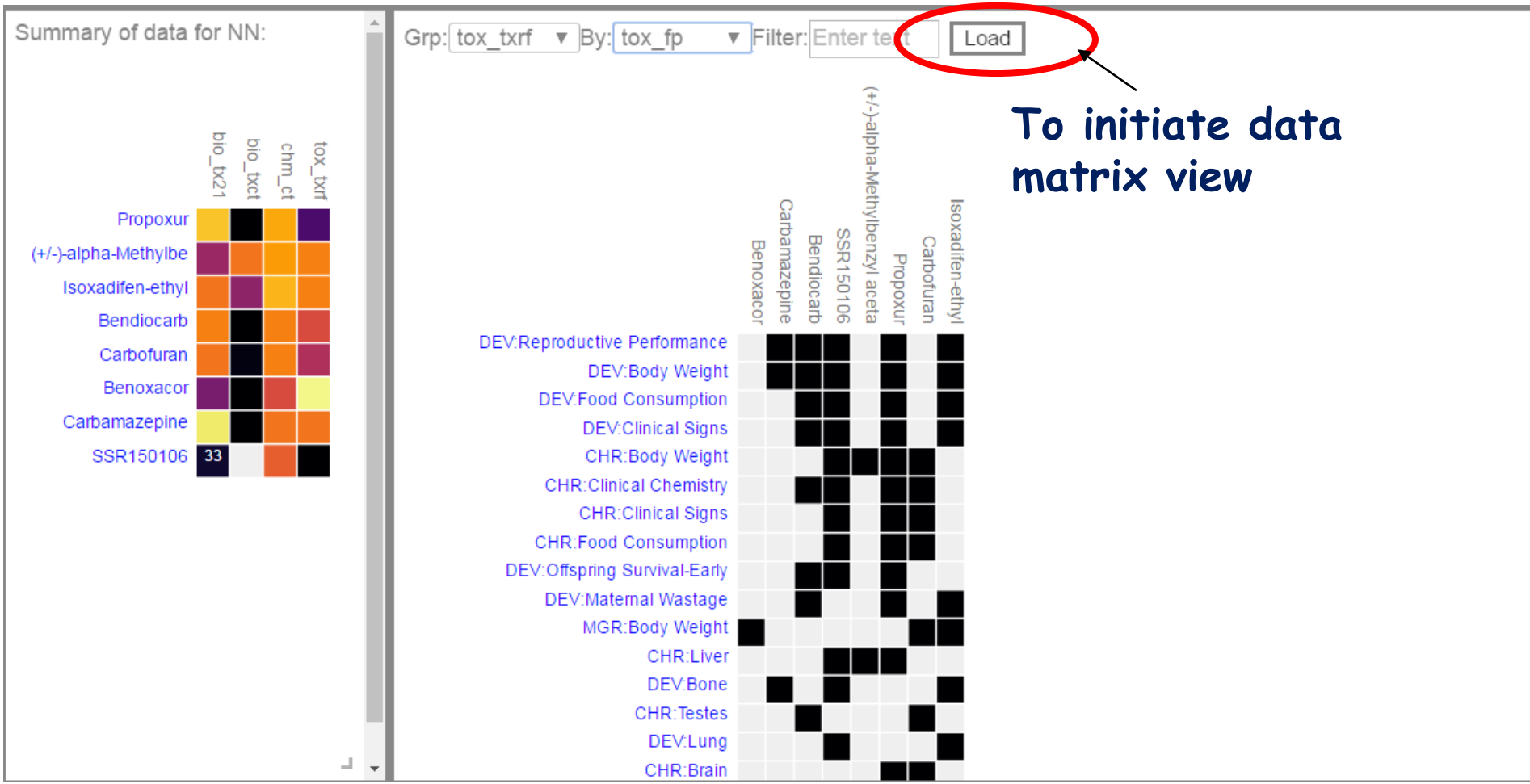
View data quantity by type

View data coverage across study types

Working interface



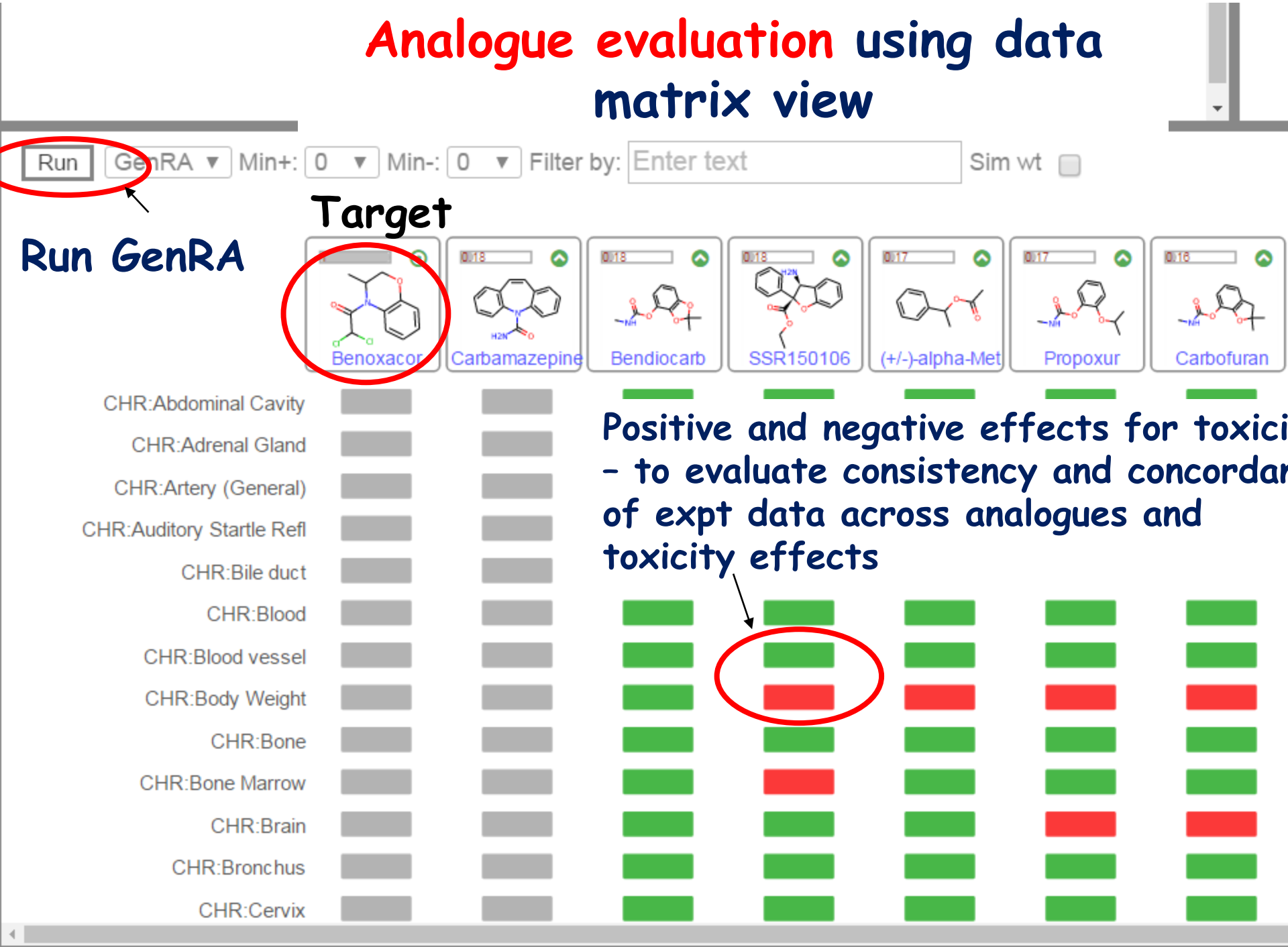
United States Environmental Protection Agency



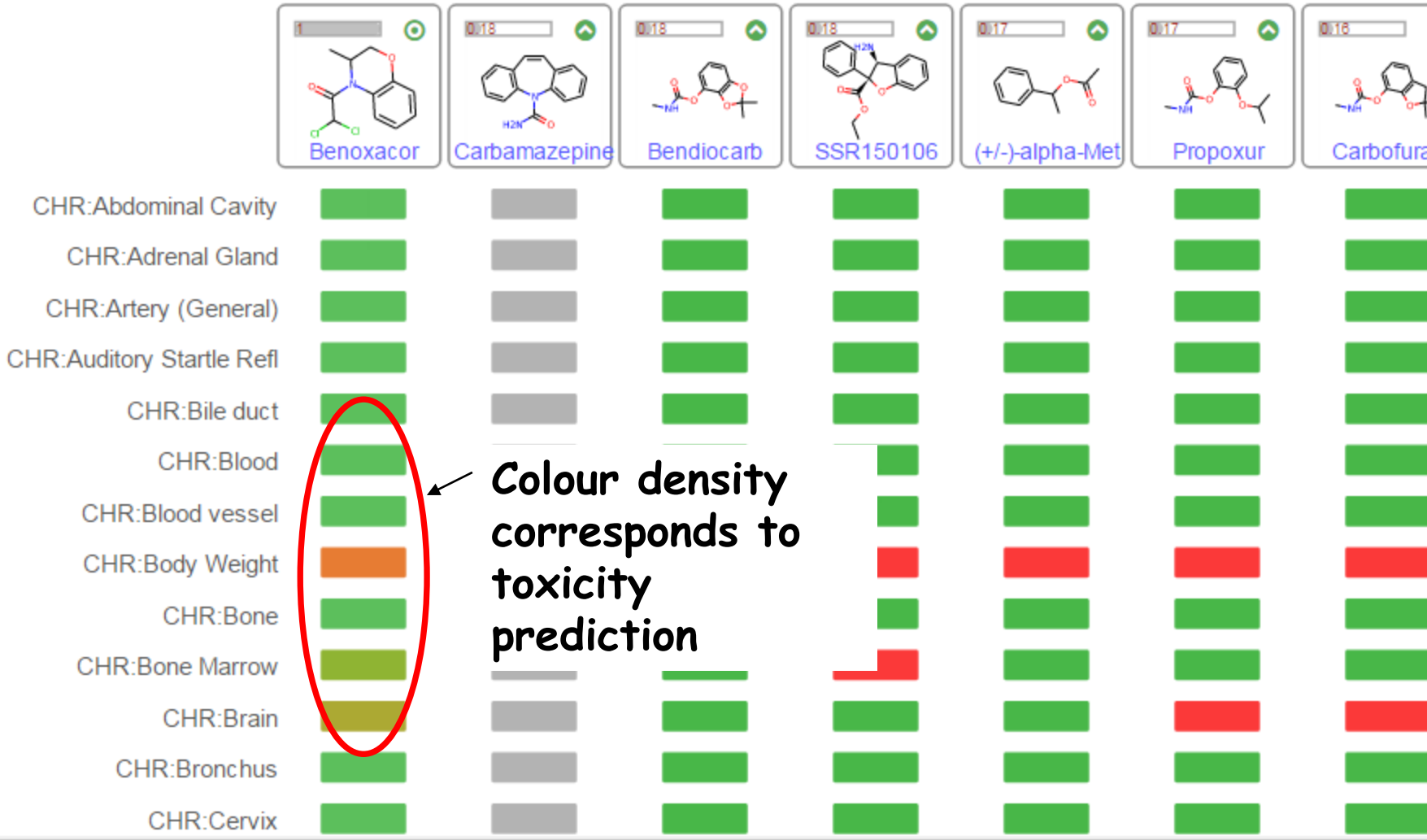
To initiate data matrix view

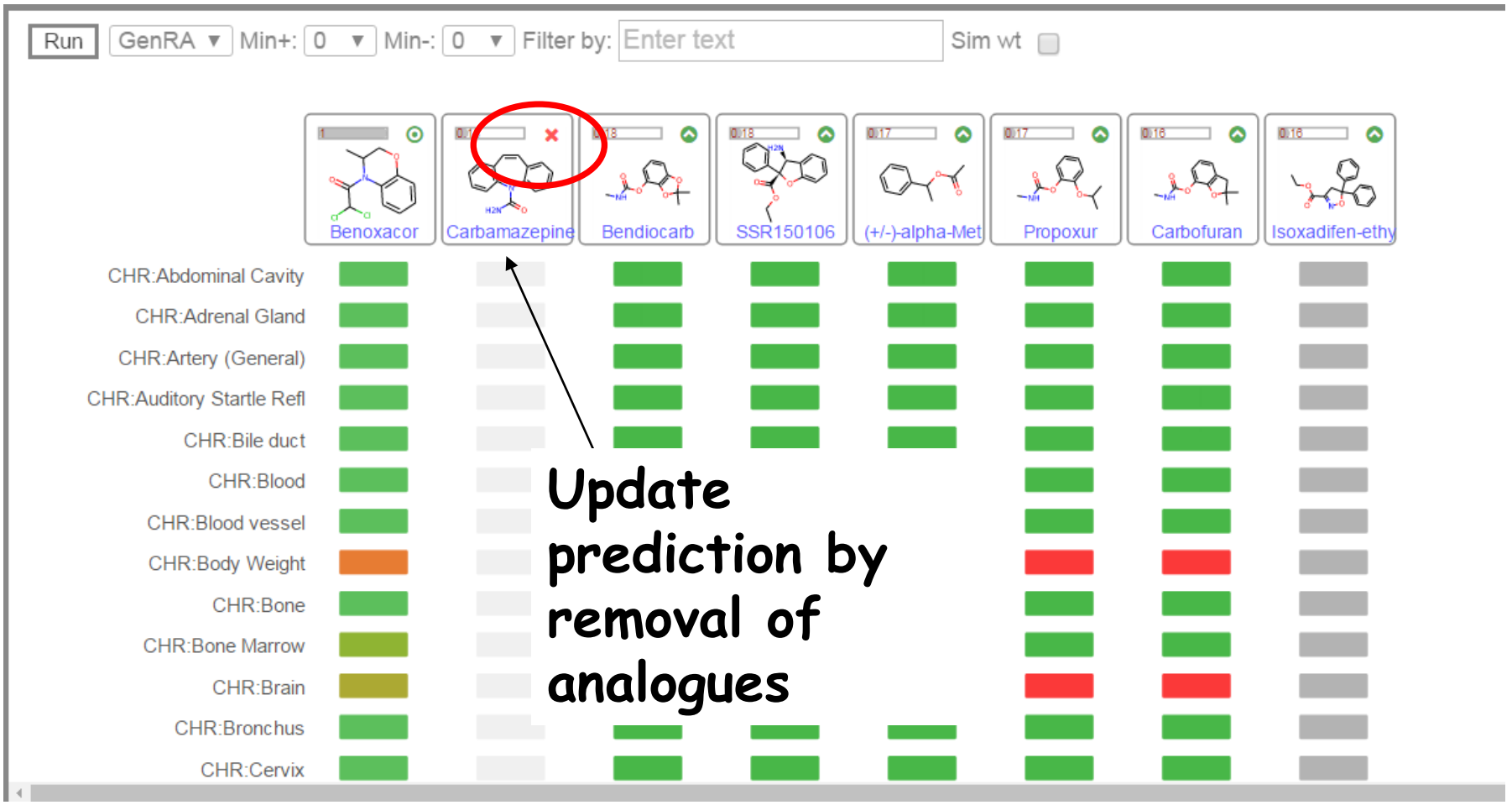
Data gap analysis - View data coverage across study type on the basis of toxicity effects

Analogue evaluation using data matrix view



Data gap filling using GenRA within data matrix




Run
GenRA
Min+: 0
Min-: 0
Filter by: Enter text
Sim wt
Benoxacor
Carbamazepine
Bendiocarb
SSR150106
(+/-)-alpha-Met
Propoxur
Carbofuran
Isoxadifen-ethy
CHR:Abdominal Cavity
CHR:Adrenal Gland
CHR:Artery (General)
CHR:Auditory Startle Refl
CHR:Bile duct
CHR:Blood
CHR:Blood vessel
CHR:Body Weight
CHR:Bone
CHR:Bone Marrow
CHR:Brain
CHR:Bronchus
CHR:Cervix
Update prediction by removal of analogues

Summary

- Still many challenges remain in read-across
- Quantifying the uncertainty of read-across prediction is a critical issue
- Have illustrated the research directions being taken and work to implement these into practical tools

Acknowledgements

- Imran Shah
- George Helman
- Tony Williams
- Jeff Edwards
- Richard Judson
- Chris Grulke
- Ann Richard
- ……