

# Building Scientific Confidence in Read-Across: Progress in using HT Data to inform Read-Across Performance



From Assay to Assessment: Incorporating High Throughput Strategies into Health and Safety Evaluations

41st Annual Winter Meeting of The Toxicology Forum

**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
- Identifying the sources of uncertainties associated with 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 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).

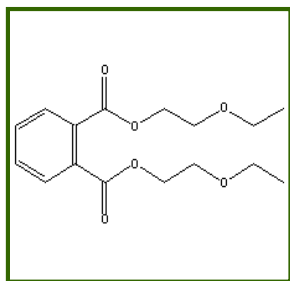
# Definition: Read-across

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		

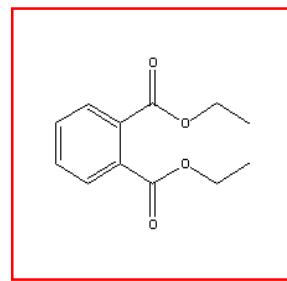
● Reliable data

○ Missing data



Known to be  
harmful

Acute fish  
toxicity?



Predicted to be  
harmful

# Background

---

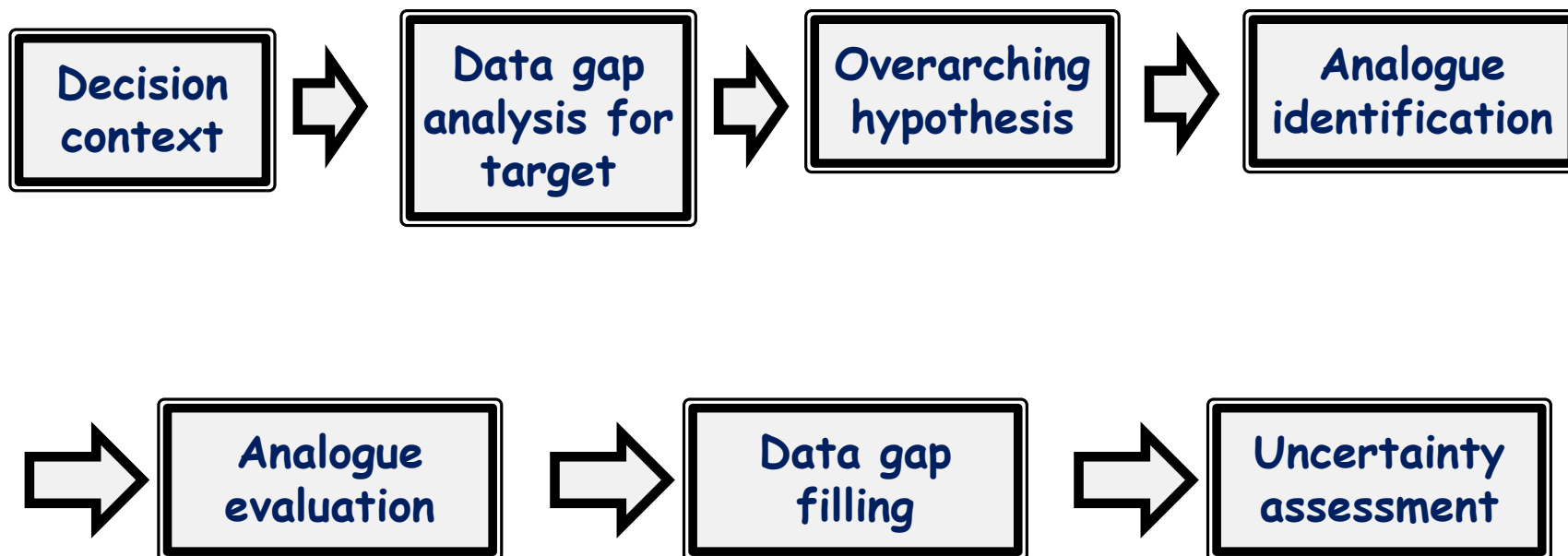
- Category and analogue approaches have been used in the regulatory arena for many years..
- Technical Guidance was first developed by the US EPA in support of the US HPV Challenge Program in 1998
- Same guidance was embedded into the OECD Manual for the Assessment of Chemicals used as part of the OECD HPV programme
- Guidance was updated in 2007 as part of the preparations to the EU REACH regulation

# Background

---

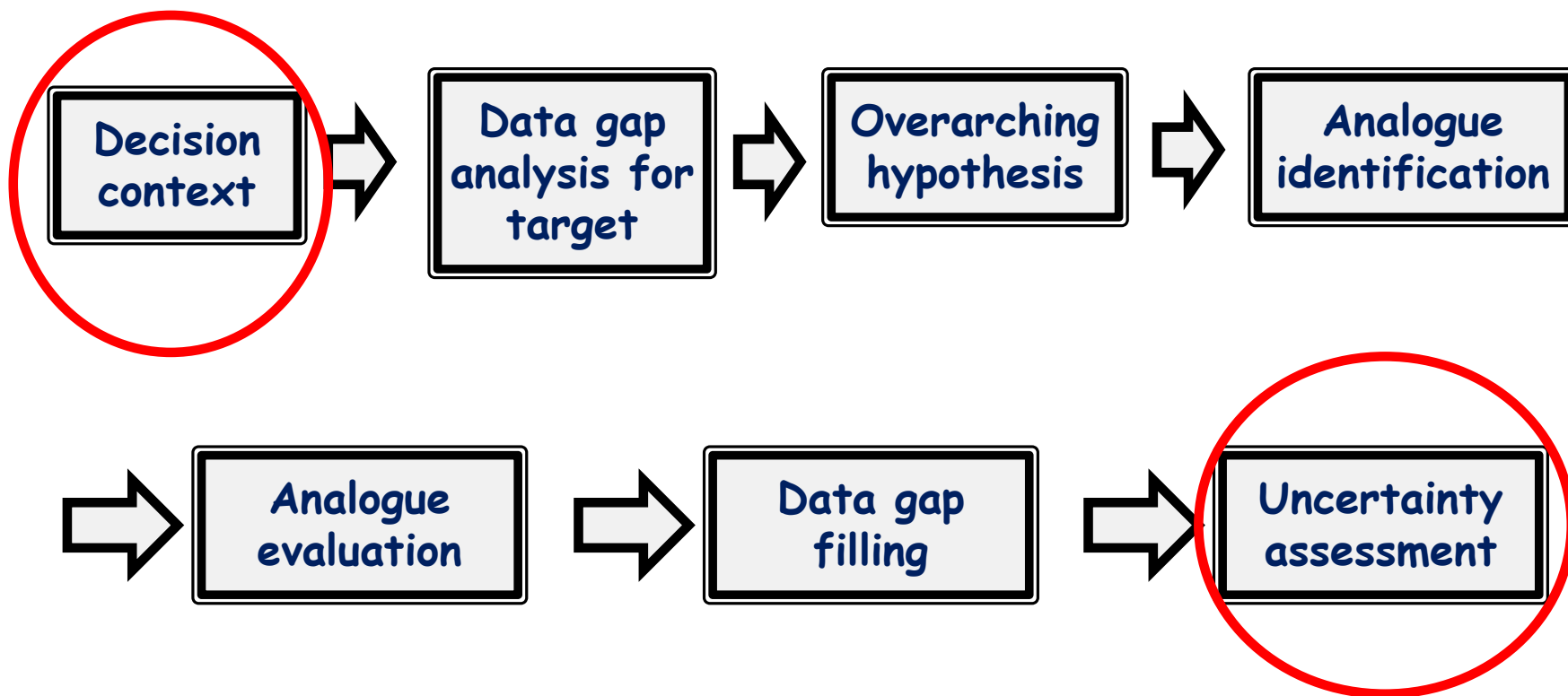
- The last few years has seen a shift towards exploiting mechanistic understanding into hazard and risk assessment relying on HT/HC data, AOP frameworks etc.
- OECD Technical guidance for grouping was revised in 2014
- Currently OECD case studies to develop AOP-informed IATA where read-across forms the basis are anticipated to be helpful in revising the guidance even further

# The Category Workflow



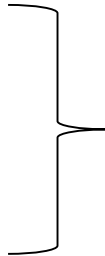


# The Category Workflow



# 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



# 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:

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

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

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

$$y_i = \text{predicted activity of chemical } (c_i)$$

$$x_j^{\beta} = \text{activity of } c_j \text{ in } \beta$$

$$s_{ij}^{\alpha} = \text{Jaccard similarity between } x_i^{\alpha}, x_j^{\alpha}$$

$$k = \text{up to } k \text{ nearest neighbours}$$

# 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

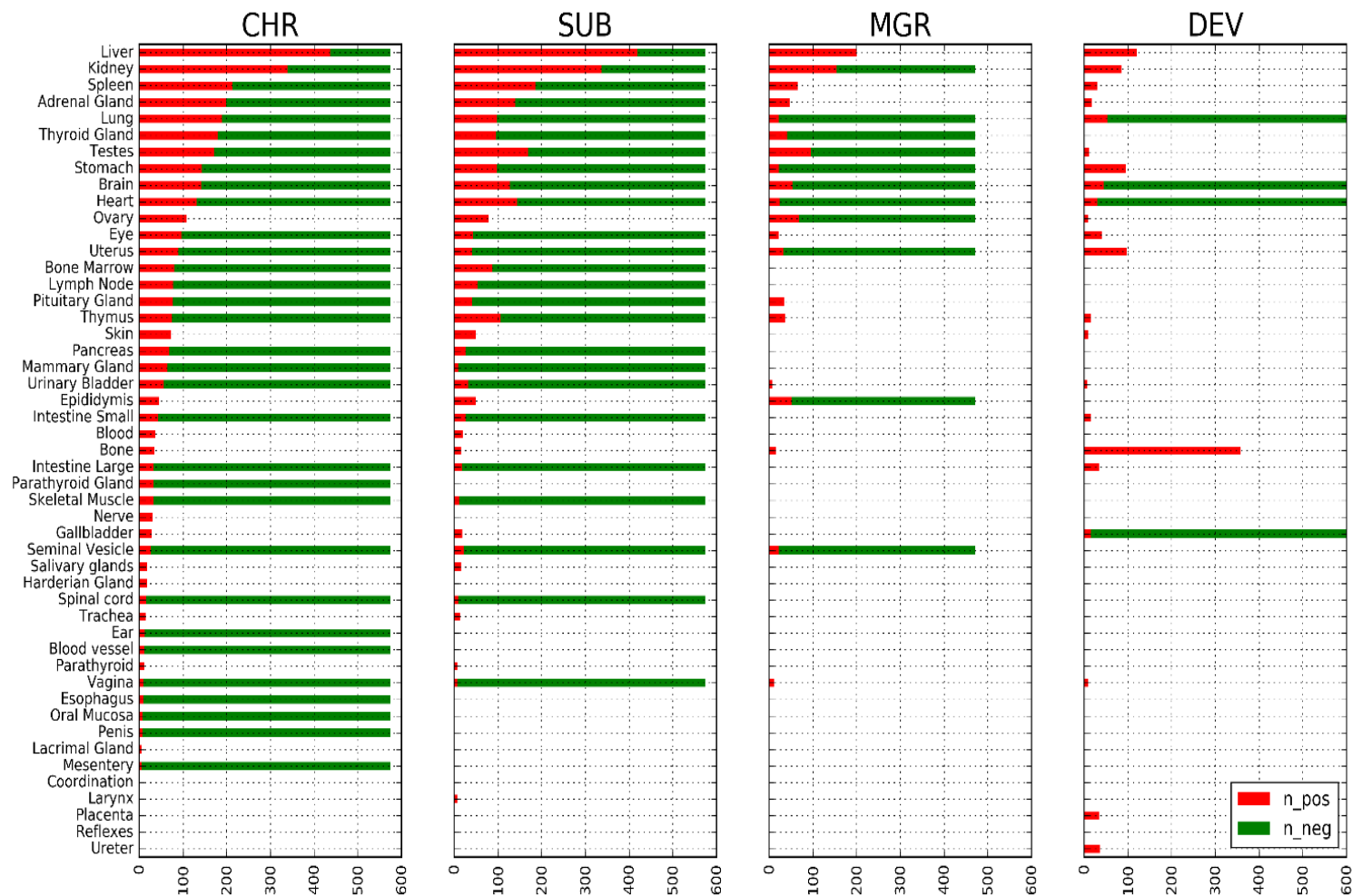
Use K-means analysis to group chemicals by similarity  
Use cluster stability analysis  
~ 100 local neighborhoods

## 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



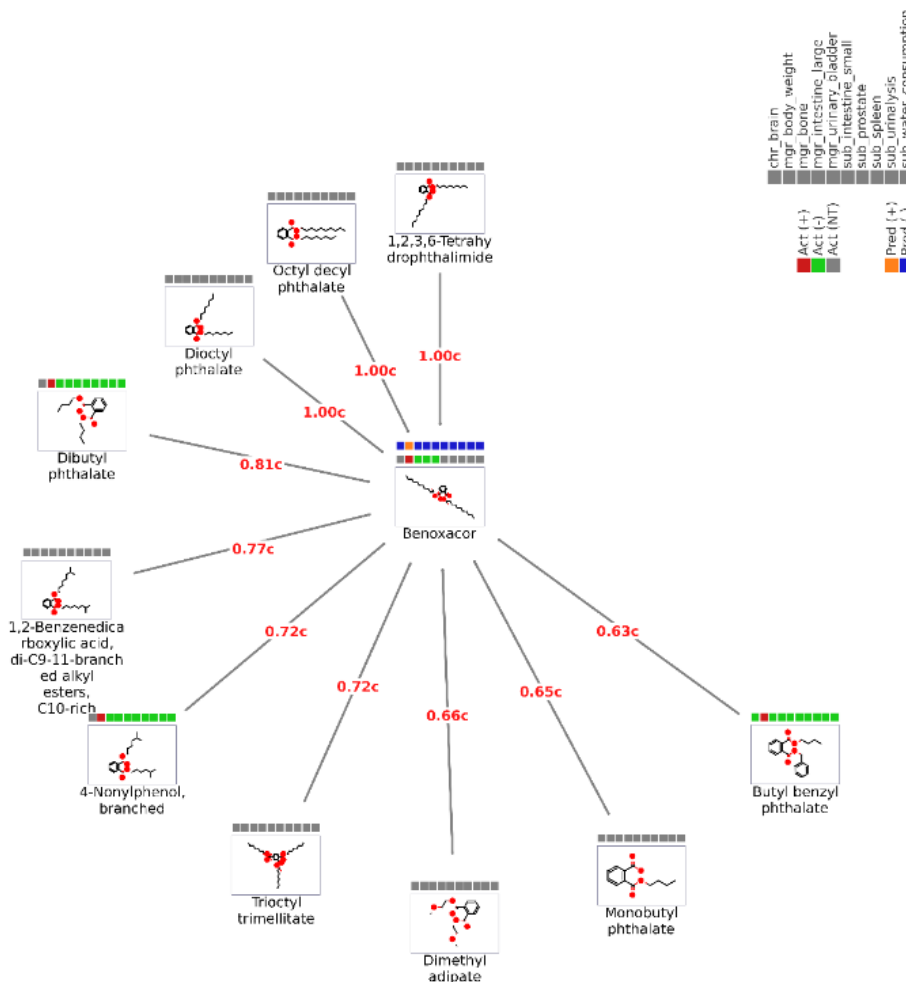
# GenRA: Clustering chemicals

---

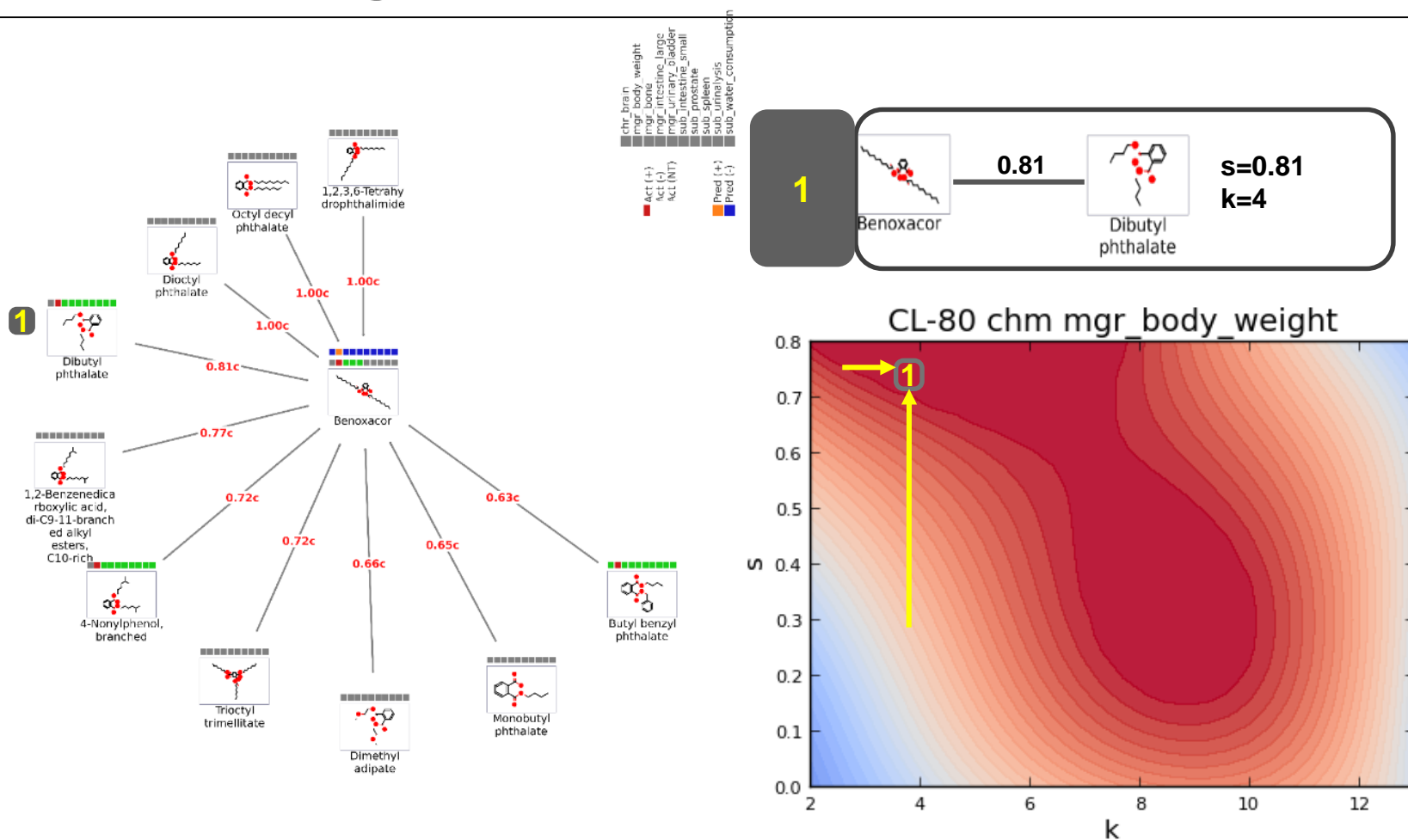
Shah et al, 2006

# GenRA – Performance in Each Cluster

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



# GenRA - Analysing Local Neighborhood of a Chemical



# GenRA – Insights and Next Steps

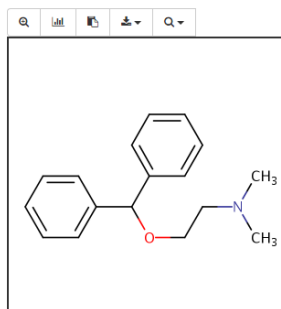
---

- 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...

## Diphenhydramine

58-73-1 | DTXSID4022949 ⓘ

ⓘ Searched by Synonym: Found 1 result for 'Benchlor'.



### Wikipedia

**Diphenhydramine** is an antihistamine mainly used to treat allergies. It is also used for insomnia, symptoms of the common cold, tremor in parkinsonism, and nausea. It is used by mouth, injection into a vein, and injection into a muscle. Maximal effect is typically around two hours after a dose and effects can last for up to seven hours.

Common side effects include sleepiness, poor coordination, and an upset stomach. Its use is not recommended in babies. There is no clear risk of harm... [Read more](#)

### Intrinsic Properties

### Structural Identifiers

### Record Information

Chemical Properties

Synonyms

External Links

Product Composition

Bioassays

Exposure

Analytical

Literature

Comments

### Summary

Download as:

TSV

Excel

SDF

LogP: Octanol-Water

Water Solubility

Density

Melting Point

Boiling Point

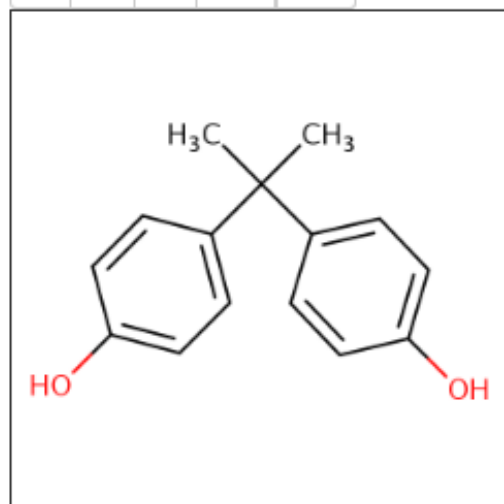
Property	Average		Median		Range		Unit
	Experimental	Predicted	Experimental	Predicted	Experimental	Predicted	
LogP: Octanol-Water	3.27 (1)	3.02 (4)	3.27 to 3.27	3.02	3.27	2.22 to 3.72	-
Water Solubility	-	1.07e-02 (4)	-	1.07e-02	-	9.69e-04 to 3.10e-02	mol/L
Density	-	1.03 (1)	-	1.03	-	-	g/cm <sup>3</sup>
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

## Bisphenol A

80-05-7 | DTXSID7020182

🔍 Searched by Approved Name: Found 1 result for 'bisphenol A'.



Insert GenRA Tab and  
Appropriate Sub-tabs



### Wikipedia

**Bisphenol A (BPA)** is an organic synthetic compound with the chemical formula Cc1ccc(O)cc1C(C)c2ccc(O)cc2. It is soluble in organic solvents, but poorly soluble in water. It has been in common use since the 1960s. BPA is employed to make certain plastics and epoxy resins. BPA-based plastics are used in a wide range of consumer products, including polycarbonate plastic bottles, food storage containers, and thermal paper receipts.

### Intrinsic Properties

### Structural Identifiers

### Record Information

Chemical Properties

External Links

Synonyms

Env. Fate/Transport

Bioassays

Exposure

Analytical

Literature

Similar Molecules

Comments

Summary

Download as:

TSV

Excel

SDF

# Tabs of associated views..

Chemical Properties

External Links

Synonyms

Env. Fate/Transport

Bioassays

Exposure

Analytical

Literature

Similar Molecules

Comments

Summary

Download as:

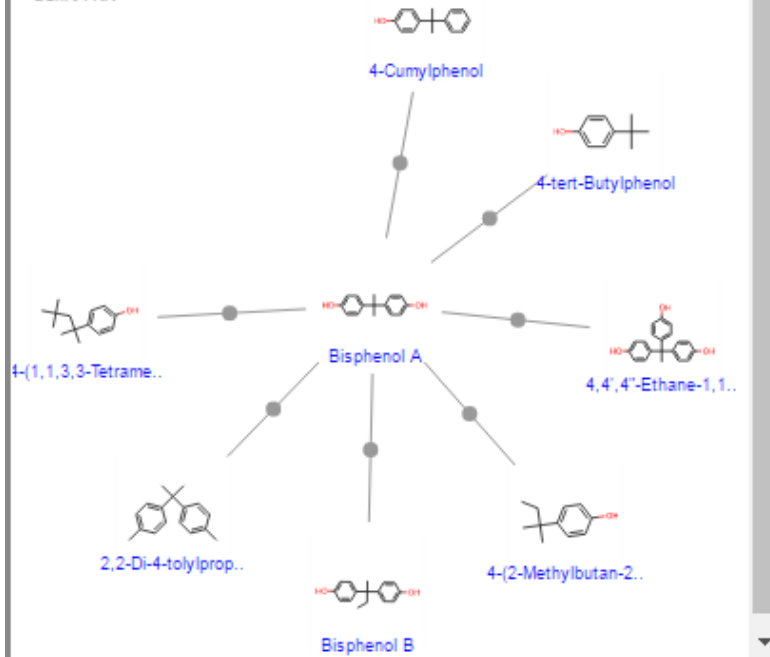
TSV

Excel

SDF

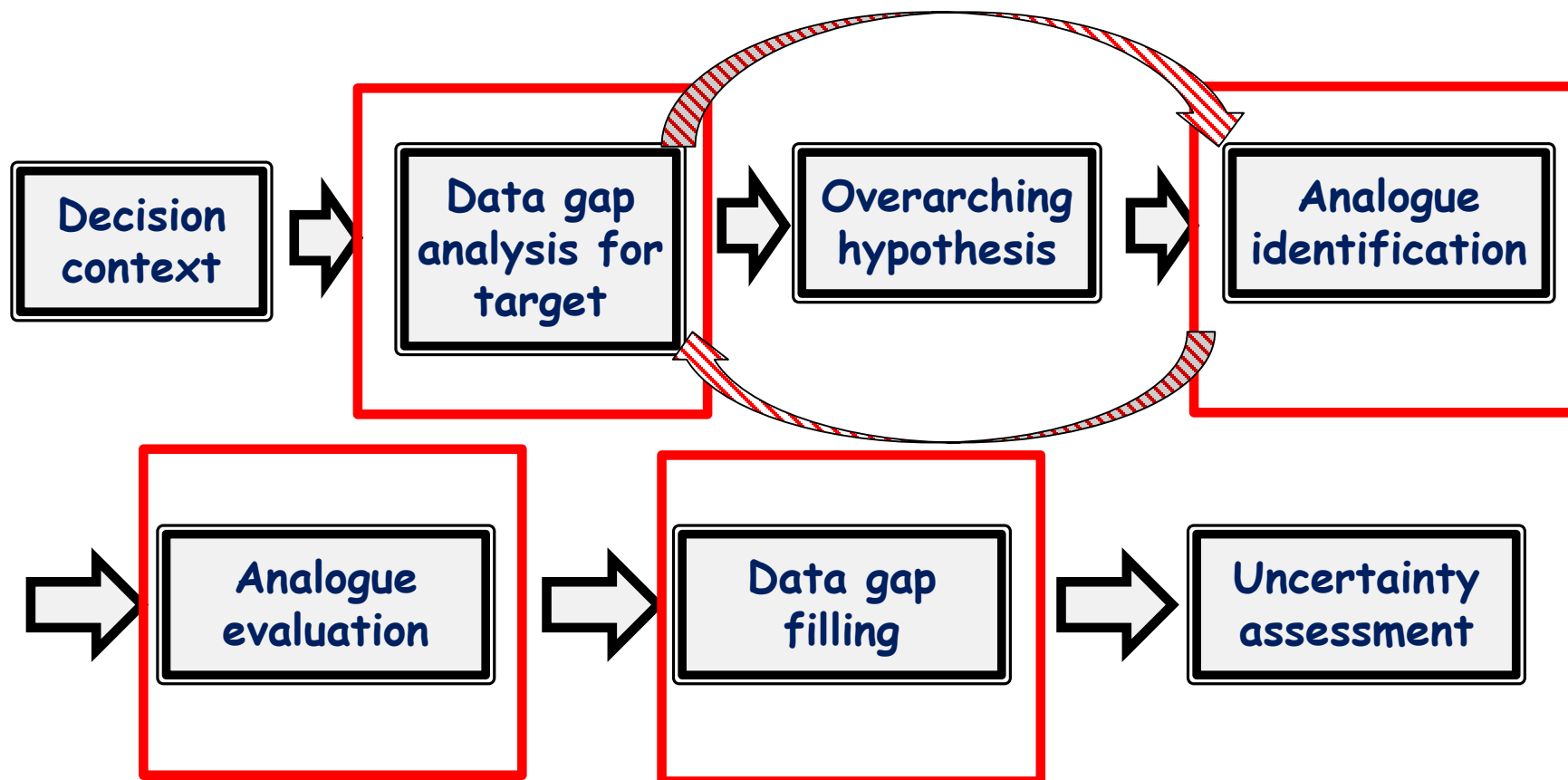
NN By:  K:  Sel by:

GenRA NN





# 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

---

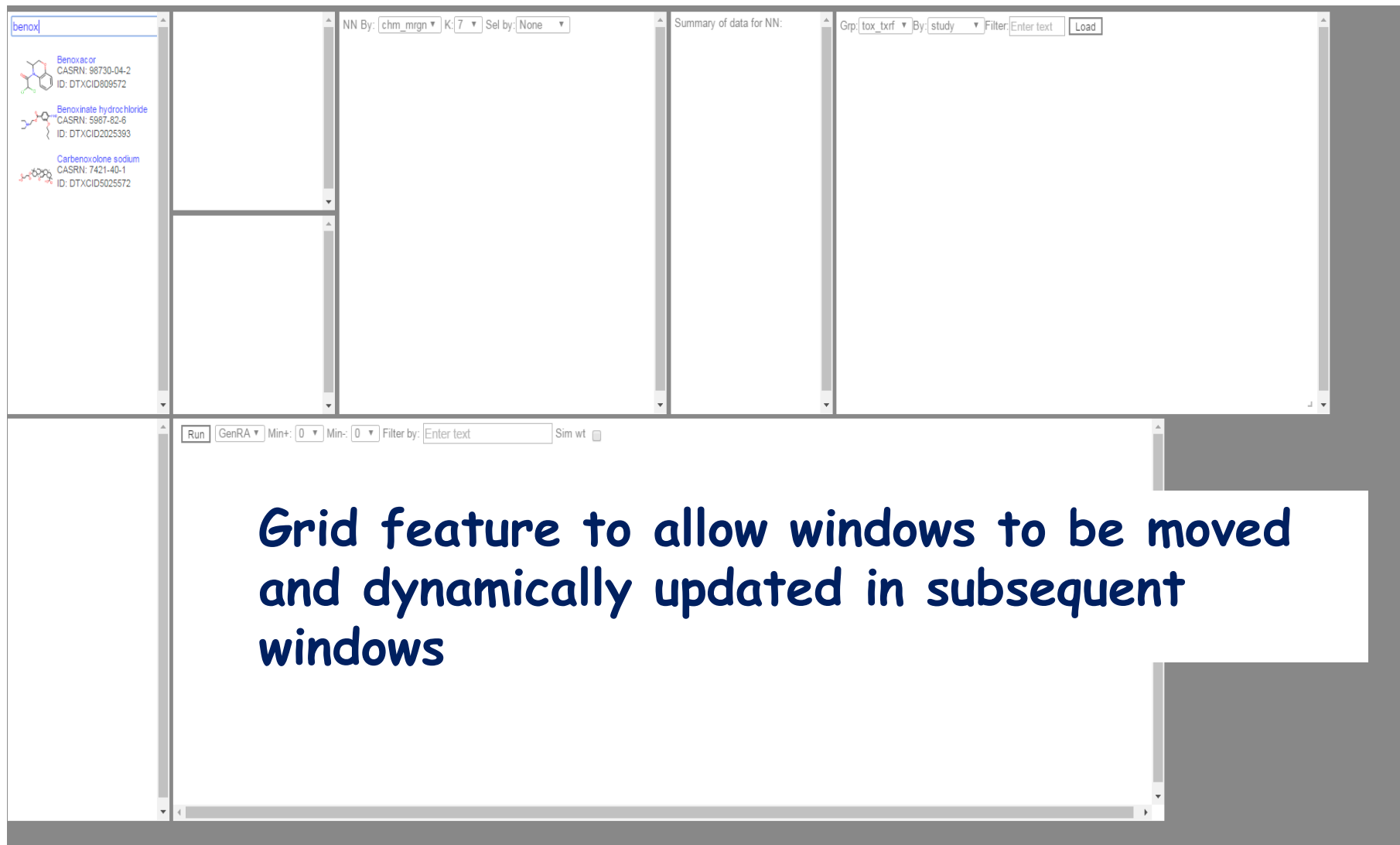
- **Analogue evaluation**
  - Launch a data matrix of the source analogues and target with the available data, colour coded by presence and absence of effects and dose descriptor information to enable an evaluation of consistency and concordance across category members and across toxicity effects
- **Data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



The screenshot displays a software interface with a grid of windows. The top-left window is titled 'benox' and lists three chemical compounds with their CASRN and DTXCID values:

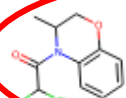
- Benoxacor**  
CASRN: 98730-04-2  
ID: DTXCID809572
- Benoxinate hydrochloride**  
CASRN: 5987-82-6  
ID: DTXCID2025393
- Carbenoxolone sodium**  
CASRN: 7421-40-1  
ID: DTXCID5025572

The top-right window is titled 'Summary of data for NN:' and contains a search bar with the text 'benox'. Below the search bar, there are dropdown menus for 'NN By: chm\_mrgn', 'K: 7', and 'Sel by: None'. To the right of these menus are fields for 'Grp: tox\_txrf', 'By: study', and a 'Filter' field with the text 'Enter text'. A 'Load' button is located to the right of the 'Filter' field.

The bottom window is a large area with a 'Run' button, a 'GenRA' dropdown, and two 'Min+' and 'Min-' dropdowns, both set to '0'. There is also a 'Filter by:' field with the text 'Enter text' and a 'Sim wt' checkbox.

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

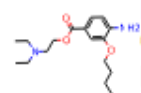
benox



Benoxacor

CASRN: 98730-04-2

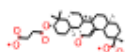
ID: DTXCID809572



Benoxinate hydrochloride

CASRN: 5987-82-6

ID: DTXCID2025393

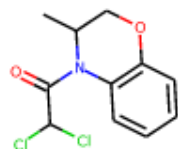


Carbenoxolone sodium

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

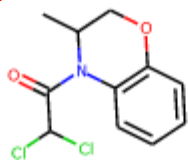
GenRA ▼

Min+: 0 ▼

Min-: 0 ▼

Filter by: Enter text

Benoxacor | 98730-04-2



dsstox\_cid:DTXCID809572  
dsstox\_sid:DTXSID3029572  
gsid:29572  
inchi\_key:PFJJMJDEVDPNE-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

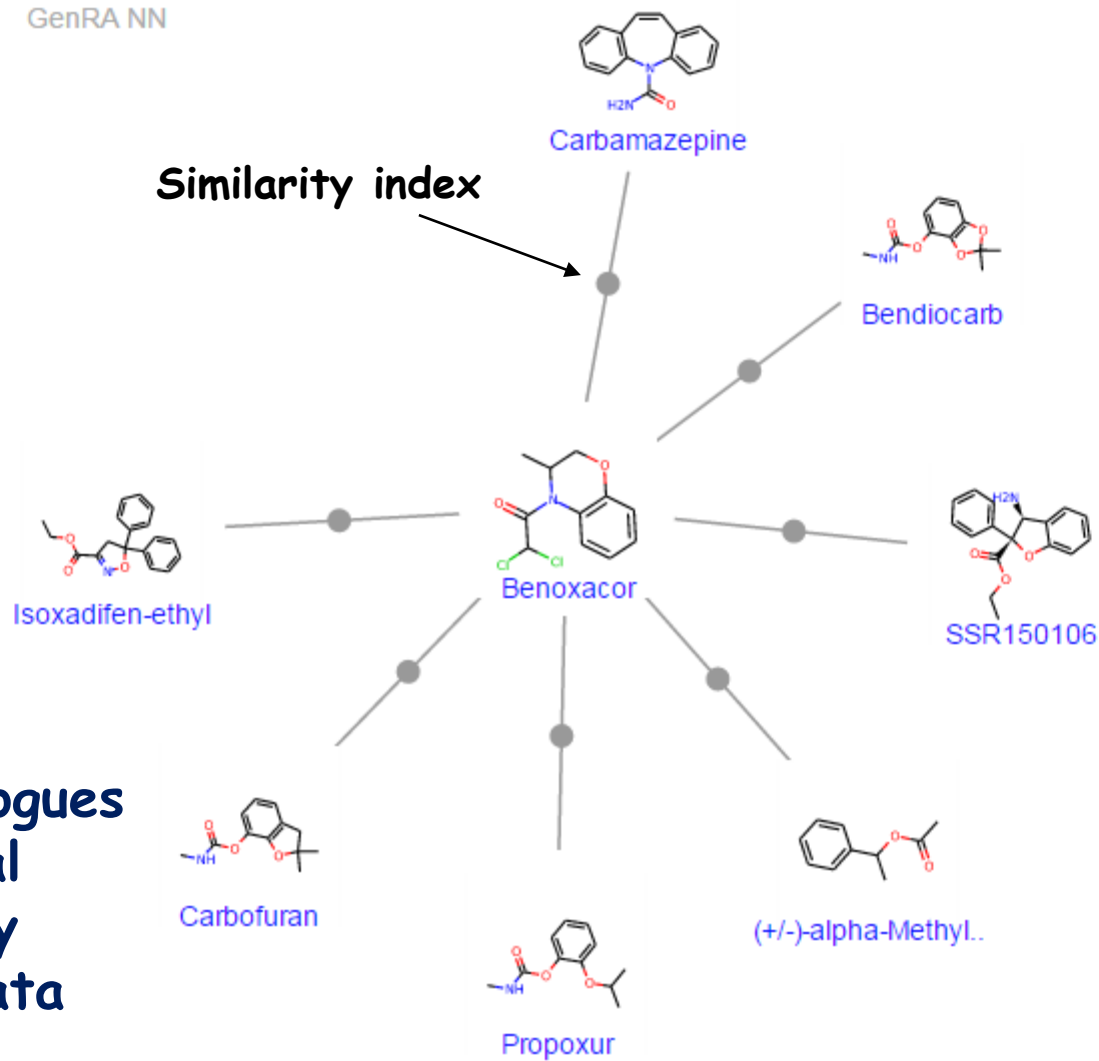
NN By: chm\_mrgn

K: 7

Sel by: tox\_txrf

GenRA NN

Similarity index



**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

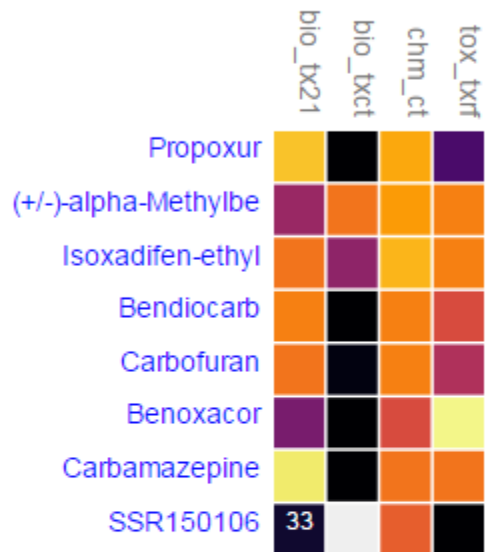
Sim wt



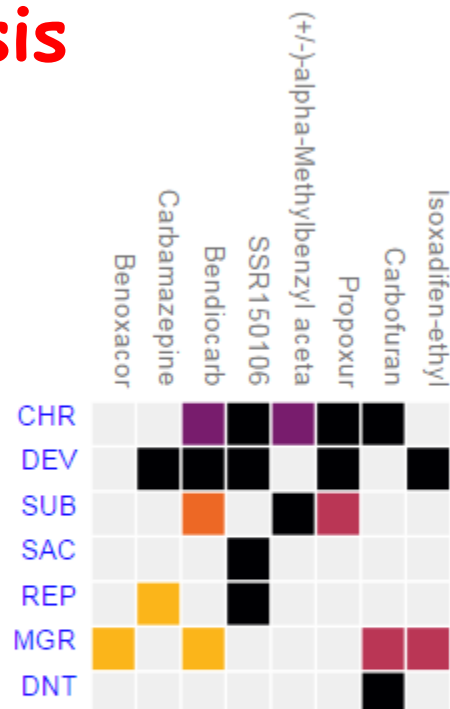
Summary of data for NN:

Grp:  By:  Filter:

## Data gap analysis



View data quantity  
by type



View data coverage  
across study types



# Working interface

benox

Benoxacor

CASRN: 98730-04-2

ID: DTXCID809572

Benoxinate hydrochloride

CASRN: 5987-92-6

ID: DTXCID2025393

Carbenoxolone sodium

CASRN: 7421-40-1

ID: DTXCID5025572

Benoxacor | 98730-04-2

dsstox\_oid: DTXCID809572

dsstox\_sid: DTXSID3029572

gsid: 20572

inchi\_key: PFJJMJDEVLPNE-UHFFFAOY

lupac: 2,2-dichloro-1-(3-methyl-2,3-dihydro

mol\_weight: 200.1165486308594

pubchem\_sid: 62306

smiles: CCTOC2=C(C=CC=C2)N1C(=O)C

tag: TOXCST,TOXREF,CTD

NN By: 

chm\_mrgn

K: 7

 Sel by: 

tox\_txf

GenRA NN

Carbamazepine

Bendiocarb

SSR150106

(+/-)-alpha-Methyl.

Propoxur

(+/-)-alpha-Methylb

Isoxadifen-ethyl

Carbofuran

Benoxacor

Summary of data for NN:

	tox_txf	chm_mrgn	bio_toc	bio_toc1
Propoxur				
(+/-)-alpha-Methylb				
Isoxadifen-ethyl				
Bendiocarb				
Carbofuran				
Benoxacor				
Carbamazepine				
SSR150106				

Grp: 

tox\_txf

 By: 

study

 Filter: 

Enter text

Load

Chr

Dev

Sub

Sac

Rep

MGR

DNT

Benoxacor

Carbamazepine

Bendiocarb

SSR150106

Propoxur

Carbofuran

Isoxadifen-ethyl

(+/-)-alpha-Methylbenzyl acetate

Run

GenRA

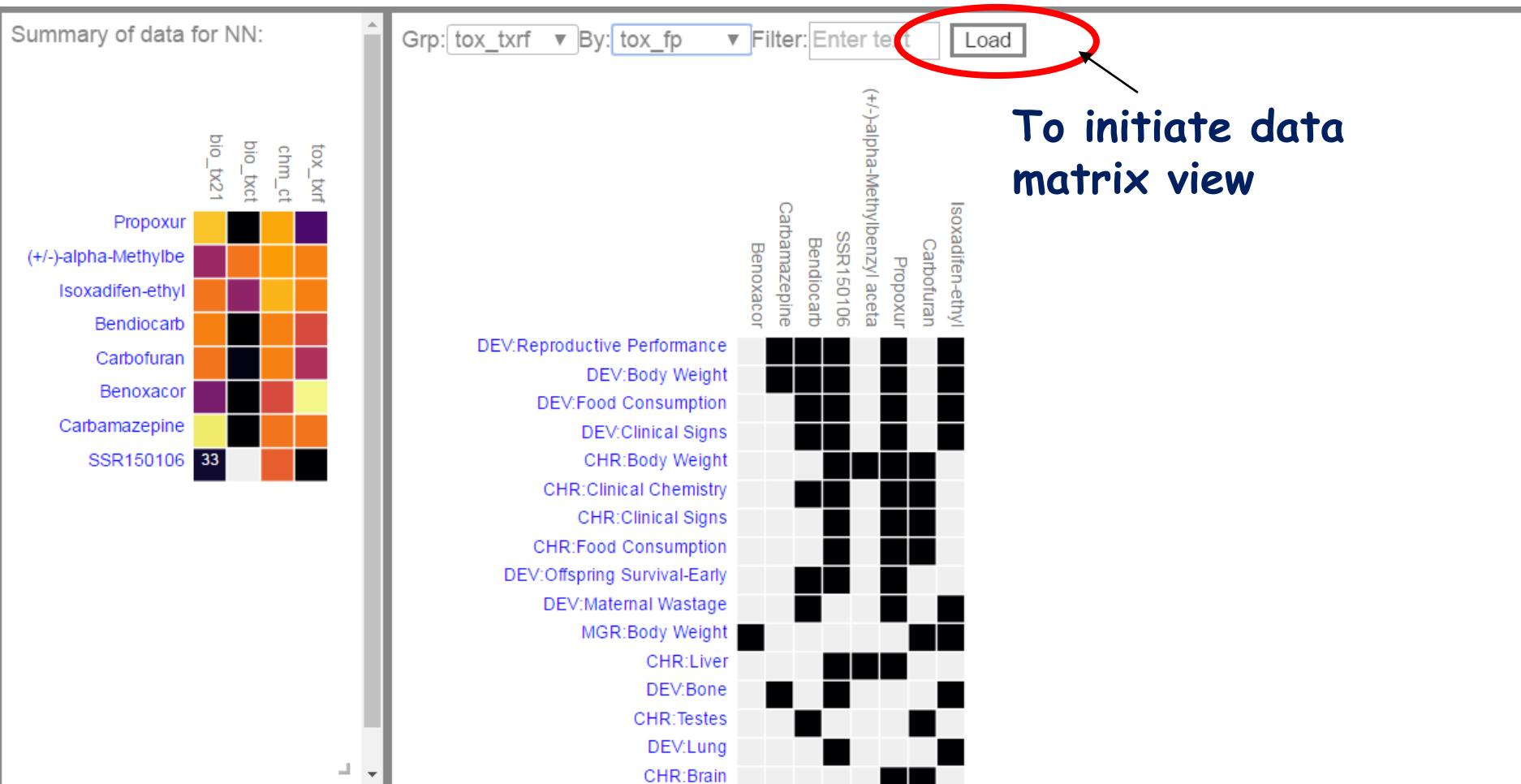
Min+: 0

Min-: 0

Filter by: 

Enter text

Sim wt ☐



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

# Analogue evaluation using data matrix view

Run

GenRA

Min+: 0

Min-: 0

Filter by: Enter text

Sim wt ☐

Target

Run GenRA

Benoxacor

Carbamazepine

Bendiocarb

SSR150106

(+/-)-alpha-Met

Propoxur

Carbofuran

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

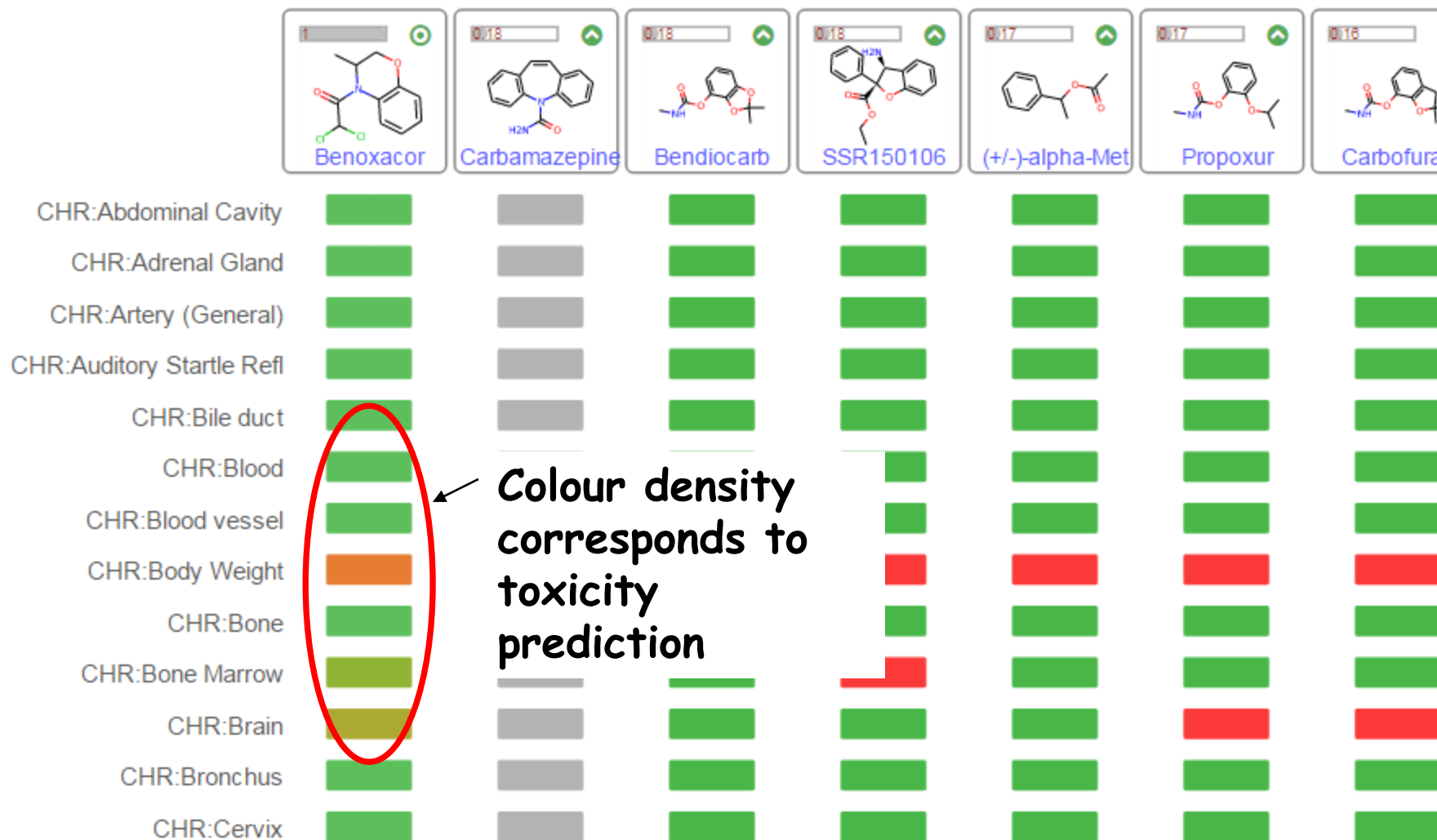
CHR:Bronchus

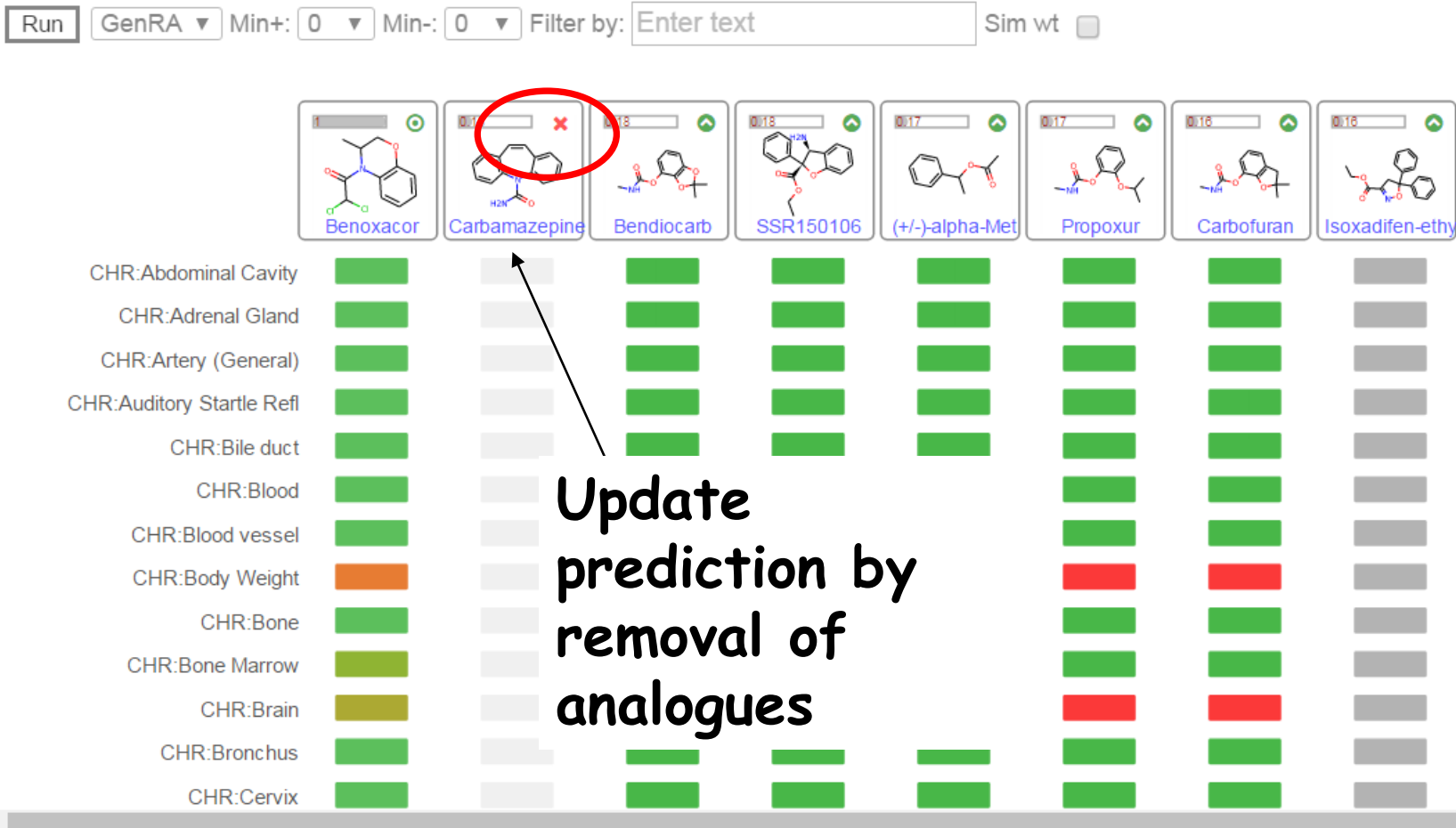
CHR:Cervix

Positive and negative effects for toxicity  
- to evaluate consistency and concordance  
of expt data across analogues and  
toxicity effects



# Data gap filling using GenRA within data matrix





# 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
- .....