Combination of Computational New Approach Methodologies for Enhancing Evidence of Biological Pathway Conservation Across Species

P. Schumann[1], C. Rivetti[2], J. Houghton[2], B. Campos[2], G. Hodges[2], and C. LaLone[3]

[1]Oak Ridge Institute for Science and Education, Duluth, MN
[2]Safety and Environmental Assurance Centre, Unilever, Colworth Science Park, Sharnbrook, Bedfordshire, UK
[3]U.S. Environmental Protection Agency, Office of Research and Development, Mid-continent Ecology Division, Duluth, MN

Peter Schumann | schumann.peter@epa.gov | 414-687-7552

Introduction

- **Objective**: To demonstrate the value of combining the use of two computational New Approach Methodologies (NAMs), the Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) tool and the Genes to Pathways – Species Conservation Analysis (G2P-SCAN) tool, for supporting cross-species predictions of chemical susceptibility.
- **Motivation**: The rate at which new chemicals are developed presents a challenge for keeping pace with thorough safety evaluations[1]. This creates a need to develop NAMs, that are by nature non-animal based, giving rise to a Next Generation Risk Assessment (NGRA) paradigm.
- **Methods**:
  - SeqAPASS uses protein sequence information to evaluate chemical target conservation across species to support predictions of species susceptibility to chemical exposure[2].
  - G2P-SCAN uses biological pathway, gene orthology, and protein family information to support predictions of pathway conservation across 7 species: humans (*H. sapiens*), rats (*R. norvegicus*), mice (*M. musculus*), zebrafish (*D. rerio*), fruit flies (*D. melanogaster*), roundworms (*C. elegans*), and yeast (*S. cerevisiae*)[3].
  - Case examples were used to demonstrate the combined use of these tools for three toxicologically relevant molecular targets: estrogen receptor 1 (ESR1), peroxisome proliferator-activated receptor alpha (PPARA), and gamma-aminobutyric acid type A receptor subunit alpha1 (GABRA1).

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)

SeqAPASS (https://seqapass.epa.gov/seqapass/) v6.1

Chemical Molecular Target in Sensitive Species ≅

Compare to *millions* of proteins from *thousands* of species!

Greater similarity → Greater likelihood that chemical can act on the protein
Line of Evidence: Predict Potential Chemical Susceptibility Across Species

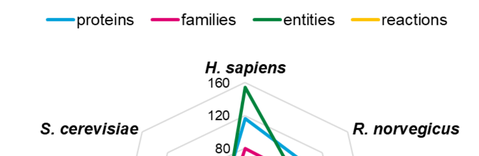
Chemical Name	Target	Target Gene Symbol	Level 1 Query Species (protein accession)	Level 2 Domain (domain accession)	Level 3 Template Species (protein accession)	Level 3 Amino Acids
Oxybenzone; Butylparaben; Dibutyl phthalate; Diethylstilbestrol	estrogen receptor 1	ESR1	Homo sapiens (NP_000116.2)	Ligand binding domain of Estrogen receptor (cd06949)	Homo sapiens (NP_000116.2)	Butylparaben (353E, 394R, 404F); Diethylstilbestrol (343M, 353E, 394R, 404F, 524H)
2-Ethylhexanoic acid	peroxisome proliferator-activated receptor alpha	PPARA	Homo sapiens (NP_005027.2)	Ligand binding domain of peroxisome proliferator-activated receptors (cd06932)	NA	NA
Topiramate	gamma-aminobutyric acid receptor subunit alpha-1 precursor	GABRA1	Homo sapiens (NP_001121116.1)	Neurotransmitter-gated ion-channel ligand binding domain (pfam02931)	NA	NA

- Summary of SeqAPASS information used in Level 1 – 3 evaluations for the three case example targets.

Genes to Pathways – Species Conservation Analysis (G2P-SCAN)

ESR1 → **G2P-SCAN** → reactome → PANTHER, UniProt, InterPro

- Gene orthologs
- Proteins
- Protein Families
- Reactions
- Entities

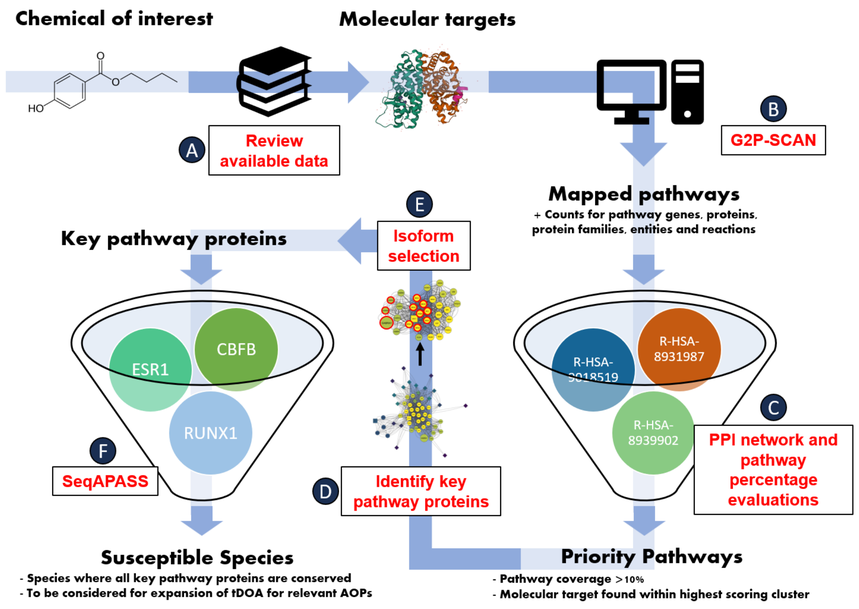


Estrogen-dependent gene expression (R-HSA-9018519)

- Summary of pathway count data for the *ESR1* mapped Reactome pathway "Estrogen-dependent gene expression" (R-HAS-9018519) across 7 common model organisms.
- The similarities in count data in addition to gene orthology and protein family identities supports inferences of pathway conservation across each species

Greater similarity → Greater likelihood that pathway is conserved in the species
Line of Evidence: Predict Potential Conservation of Pathway Across Species

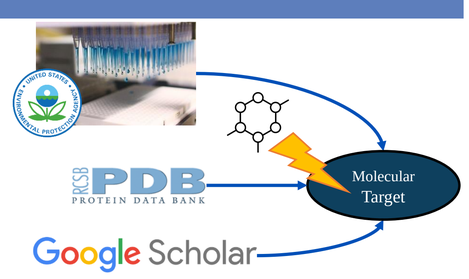
Overview of the Combined Approach



- Diagram showing the approach of combining the use of SeqAPASS and G2P-SCAN tools to support cross-species predictions of chemical susceptibility though inferences of pathway conservation.
- Abbreviations: taxonomic domain of applicability (tDOA); adverse outcome pathways (AOPs); molecular complex detection (MCODE); protein-protein interaction (PPI).

A – Identifying Targets

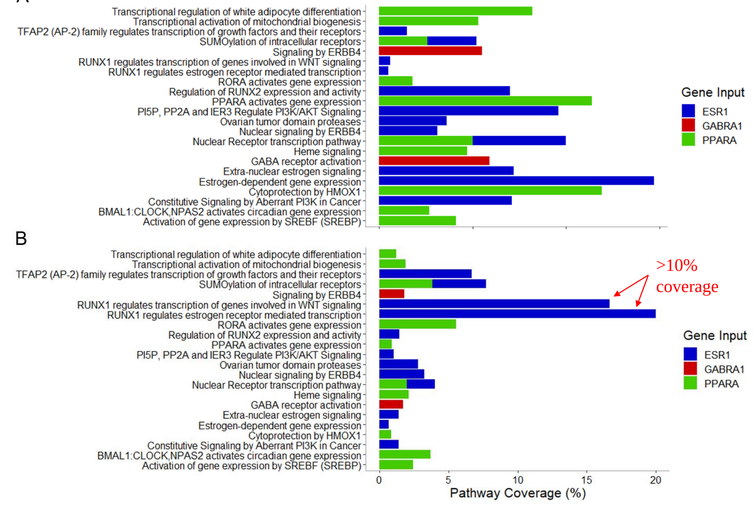
- EPA HTP tPOD
- EPA ToxCast Bioactivity
- RCSB PDB
- Literature



- **HTP tPOD** = High-throughput transcriptomic point of departure
- **EPA** = Environmental Protection Agency
- **RCSB PDB** = Research Collaboratory for Structural Bioinformatics Protein Data Bank

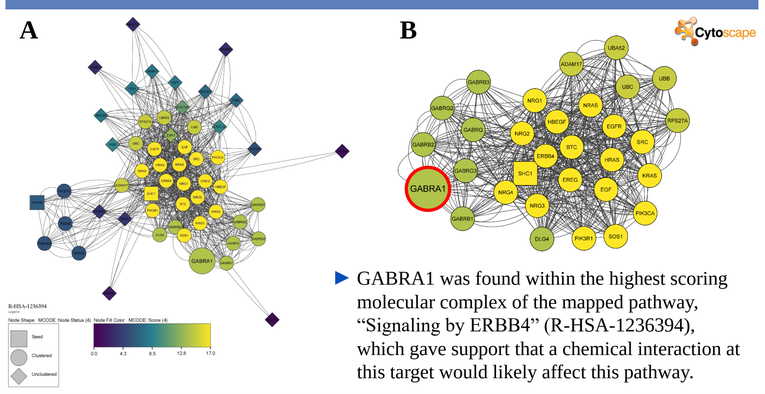
- All potential targets were evaluated as valid only if there was clear literature evidence to support a direct interaction with the compound of interest. Direct interactions were considered as having evidence of binding or alteration of protein activity in a concentration-dependent manner.

B – G2P-SCAN Results



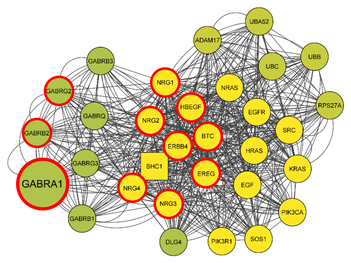
- (**A**) Bar plot of gene counts from all Reactome pathways that were mapped using either *ESR1*, *PPARA*, or *GABRA1* as a G2P-SCAN input. (**B**) Bar plot of the pathway coverage percentage for each mapped Reactome pathway.

C – Pathway Prioritization



- GABRA1 was found within the highest scoring molecular complex of the mapped pathway, "Signaling by ERBB4" (R-HSA-1236394), which gave support that a chemical interaction at this target would likely affect this pathway.

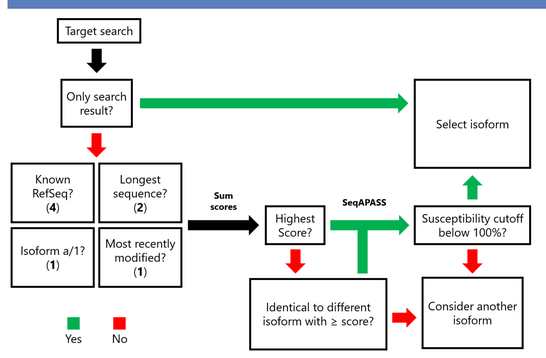
D – Pathway Network Analysis



MCODE Clusters	MCODE Node Status	MCODE Score 4	name
Cluster 1	Clustered	17	NRG4
Cluster 1	Clustered	17	HBEGF
Cluster 1	Clustered	17	NRG3
Cluster 1	Clustered	17	NRG2
Cluster 1	Clustered	17	BTC
Cluster 1	Clustered	17	NRG1
Cluster 1	Clustered	17	ERBB4
Cluster 1	Clustered	17	EREG
Cluster 1	Clustered	14	GABRB2
Cluster 1	Clustered	14	GABRG2
Cluster 1	Clustered	14	GABRB1
Cluster 1	Clustered	14	GABRQ
Cluster 1	Clustered	14	GABRG3
Cluster 1	Clustered	14	GABRB3

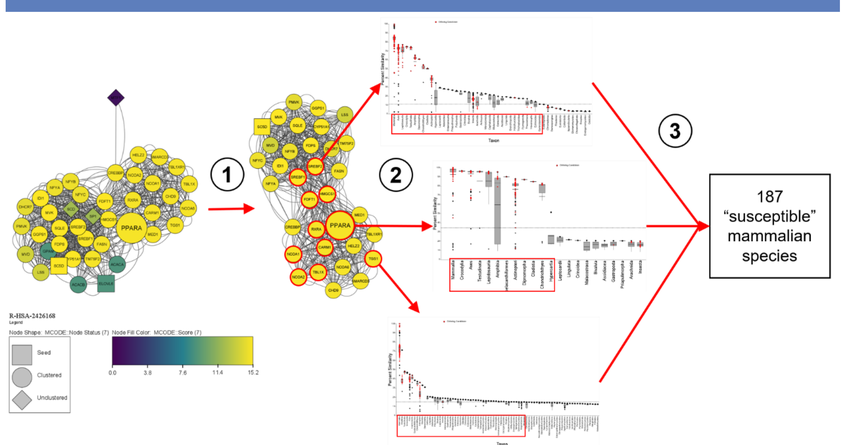
- The top ten scoring proteins within the highest scoring MCODE cluster (i.e., the most interconnected proteins within the network according to the MCODE algorithm) that were also directly connected to the molecular target in the network were used for further analysis. Filtering the proteins in this way helps balance the analysis efficiency with the prediction confidence.

E – Isoform selection



- This weighted decision tree was designed to help determine what protein isoform to choose for SeqAPASS evaluations.

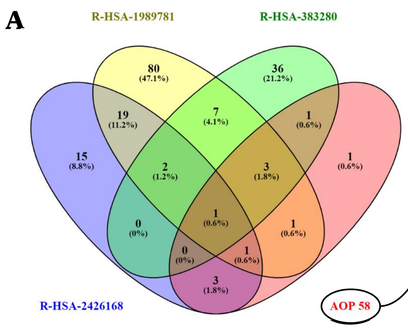
F – SeqAPASS Evaluations



- (**1**) The top ten scoring proteins connected to PPARA within the PPI network for the "Activation of gene expression by SREBF" (R-HSA-2426168) pathway are identified and then (**2**) evaluated using SeqAPASS. (**3**) The susceptible species across each of the resulting protein lists were merged.

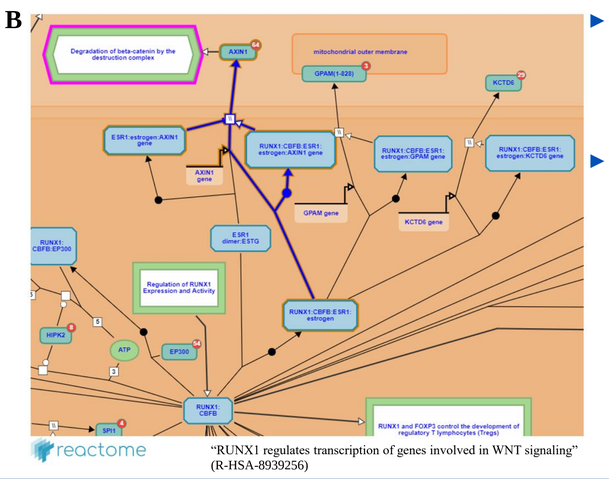
Overlaps with AOPs

- A toxicological context for the mapped Reactome pathways could be derived by comparing them with AOPs that also involve the target.



Common genes in AOP 58 and PPARA-mapped Reactome pathways	Genes found only in AOP 58
FASN	*MLXIPL*
ACACA	
PPARA	
SCD	
SREBF1	
NR1H3	
PPARG	
NR1H2	
NR1I3	
CD36	

- (**A**) The Venn diagram illustrates the overlap in gene identities between AOP (58): NR1I3 (CAR) suppression leading to hepatic steatosis and three mapped pathways using PPARA as the query.



"RUNX1 regulates transcription of genes involved in WNT signaling" (R-HSA-8939256)

- Functional overlaps can be assessed by comparing the biological outcomes of mapped pathways and AOP KEs or KERs.
- The pathway shown in **B** highlights the connections between ESR1 activity and the activation of the pathway "Degradation of beta-catenin by the destruction complex" (R-HSA-195253), which is associated with a KE of AOP 167 (i.e., *SIX1* expression) that describes the development of endometrial carcinoma

Summary & Conclusions

- In combination, it was demonstrated through the use of three case examples that these tools can be used to:
  - Expand the prediction of biological pathway conservation across all species with relevant protein data (several additional lines of evidence are generated through the use of G2P-SCAN with respect to the 6 model organisms)
  - Aid in the prediction of cross-species chemical susceptibility
  - Potentially extend the biologically plausible tDOA of relevant AOPs
  - Provide additional biological information to help further characterize certain KEs and KERs
- The three case examples used here helped identify areas for improvement for this approach:
  - Additional factors such as life stage, life history, biological sex, and toxicokinetic factors like absorption, distribution, metabolism, excretion (ADME) may be incorporated to yield a more complete understanding of the chemical exposure and resulting biological impacts
  - Incorporation of quantitative pathway information through the use of (high-throughput) transcriptomics or proteomics would allow for pathway topology evaluations for molecular target and pathway prioritizations

References / Acknowledgements

(1) Judson, R., Richard, A., Dix, D. J., Houck, K., Martin, M., Kavlock, R., Dellarco, V., Henry, T., Holderman, T., Sayre, P., Tan, S., Carpenter, T., & Smith, E. (2009). The toxicity data landscape for environmental chemicals. *Environmental Health Perspectives*, *117*(5), 685–695. https://doi.org/10.1289/EHP.0800168

(2) LaLone, C. A., Villeneuve, D. L., Lyons, D., Helgen, H. W., Robinson, S. L., Swintek, J. A., Saari, T. W., & Ankley, G. T. (2016). Sequence alignment to predict across species susceptibility (seqapass): A web-based tool for addressing the challenges of cross-species extrapolation of chemical toxicity. *Toxicological Sciences*, *153*(2), 228–245. https://doi.org/10.1093/toxsci/kfw119

(3) Rivetti, C., Houghton, J., Basili, D., Hodges, G. and Campos, B. (2023), Genes-to-Pathways Species Conservation ANalysis (G2P-SCAN): enabling the exploration of conservation of biological pathways and processes across species. Environ Toxicol Chem. Accepted Author Manuscript. https://doi.org/10.1002/etc.5600