Informatics Approaches in the Biological Characterization of Adverse Outcome Pathways

Maureen Pittman[1], Carlie LaLone[2], Dan Villeneuve[2], Gerald Ankley[2], Stephen Edwards[3], Holly Mortensen[4]

[1] EPA Student Contractor, Research Cores Unit, NHEERL/ORD/US EPA Research Triangle Park, NC 27711, USA; [2]Mid-Continent Ecology Division, NHEERL/ORD/US EPA, Duluth, MN 55804, USA; [3]Integrated Systems Toxicology Division, NHEERL/ORD/US EPA, Research Triangle Park, NC 27711, US; [4]Research Cores Unit, NHEERL/ORD/US EPA Research Triangle Park, NC 27711, USA.

Introduction

Adverse Outcome Pathways (AOPs) are a conceptual framework that characterize toxicity pathways by a series of mechanistic steps from a molecular initiating event to population outcomes. This framework may help to direct risk assessment, by aiding in computational prioritization of chemicals, genes, and tissues relevant to an adverse health outcome. One goal of EPA's Chemical Safety and Sustainability AOP Discovery and Development Project is to establish biological domains of applicability across diverse taxonomic groups, in order to identify which species might be adversely affected by a given pathway. We have developed a computational workflow implementing biological pathway annotations, gene expression networks, and relational database methods to make comparisons between diverse organisms to assess the taxonomic applicability of an AOP. The case studies explored here are: Estrogen Receptor (ER) Antagonism/Aromatase Inhibition Leading to Reproductive Dysfunction, Aryl Hydrocarbon Receptor (AHR) Activation Leading to Cardiotoxicity/ Uroporphyria, and Acetylcholinesterase (AChE) Inhibition Leading to Acute Mortality.

Case Studies

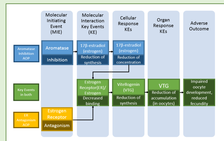


Figure 1: ER Antagonism and Aromatase Inhibition Leading to Reproductive Dysfunction. While the two pathways are initiated by different events, they unite at a single Key Event (KE): "Decreased binding of ER and Estrogen," leading to identical downstream effects of reduced vitellogenin (VTG) egg-yolk synthesis and resulting reproductive dysfunction.

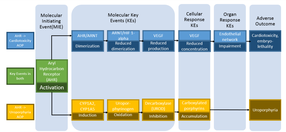


Figure 2: **AHR Activation leading to Cardiotoxicity** and **AHR Activation leading to Uroporphyria**. In the Cardiotoxicity pathway, sustained dimerization of AHR and its nuclear translator leads to reduced production of VEGF, a protein important in endothelial and cardiac development. In the Uroporphyria pathway, AHR induces expression of CYP1A2/CYP1A5 genes, leading to the inhibition of a decarboxylase enzyme and the accumulation of highly-carboxylated porphyrins.

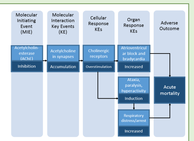


Figure 3: AChE Inhibition leading to Acute Mortality. In this AOP, accumulation of acetylcholine in the synapses leads to overstimulation of cholinergic receptors, leading to respiratory issues and cardiac distress.

Comparison of Pathway Annotation Data

Based on the molecular Key Events (KE) of each case study identified on the AOP wiki[1], we created a list of AOP-related protein orthologs. For seven model species (human, rat, alligator, frog, chicken, zebrafish, and fruit fly) annotated pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG[2]) containing any of these Key Event orthologs were selected and compared across organisms.

Using the full list of genes from AOP-relevant KEGG pathways (**Table 1**), the presence and absence of gene orthologs was compared across species using a simple binary similarity equation (**Figure 4**). Human was used as the reference organism in each case for consistency. We find the unexpected result that Chicken shares fewer orthologs to human than does Zebrafish and Frog, which is presumably due to lack of annotation and is inconsistent with phylogenic knowledge (**Figure 5**).

Case Study	Pathway ID	Pathway name
ER Antagonism/ Aromatase Inhibition	path:ko04915	Estrogen signaling pathway
	path:ko04917	Prolactin signaling pathway
	path:ko04919	Thyroid hormone signaling pathway
	path:ko04961	Endocrine and other factor-regulated calcium reabsorption
	path:ko05205	Proteoglycans in cancer
	path:ko00140	Steroid hormone biosynthesis
	path:ko04913	Ovarian steroidogenesis

Table 1: List of KEGG pathways identified in the ER/Aromatase AOPs. The list of all genes participating in these pathways becomes the AOP "expanded gene list."

Figure 4: Binary Similarity Scores for AOP pathway genes – equation and results by AOP.

$$s = \frac{a}{a + \frac{1}{2}(b + c)}$$

S = similarity
a = number of unique genes contained in both species' AOP expanded gene list
b = number of genes present in human but not the query organism
c = number of genes present in the query but not human

ER/Aromatase Binary Similarity

Organism	Binary Similarity Score (to human)
Human	1
Rat	0.98071
Alligator	0.81966
Frog	0.81623
Zebrafish	0.81211
Chicken	0.78806
Fly	NA

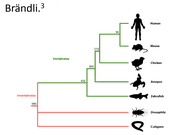
AHR Activation Binary Similarity

Organism	Binary Similarity Score (to human)
Human	1
Rat	0.98192
Alligator	0.78864
Frog	0.78452
Zebrafish	0.77969
Chicken	0.75535
Fly	0.60469

AChE Inhibition Binary Similarity

Organism	Binary Similarity Score (to human)
Human	1
Rat	0.98037
Alligator	0.67699
Frog	0.65031
Chicken	0.64013
Zebrafish	0.63846
Fly	0.04323

Figure 5: Phylogenetic tree showing the evolutionary relationships between the main animal models. Divergence time in millions of years (Mya). Adapted from Wheeler & Brändli.[3]



Objectives

1. Characterize pathway level concordance between model species.
2. Where available, analyze RNA-seq expression data to determine groups of highly co-expressed groups of genes, then compare the makeup of these groups across organisms.
3. Combine the above with public data into a database relating chemical, pathway, disease, and ontological information to each AOP case study.

Comparison of Gene Expression Networks

In order to compare the level of transcription of AOP-relevant genes across organisms, we created gene coexpression networks using publically-available RNA-seq data[4,5]. Coexpression networks are created by measuring the pairwise correlation of expression between all genes in a dataset, assigning distance based on these correlation measures, and defining modules of highly-coexpressed genes. We used expression data for five model organisms: human, mouse, chicken, frog, and fruit fly.

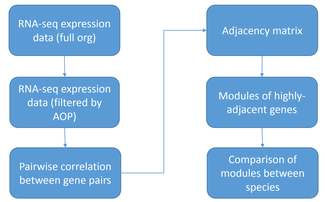


Figure 6: Workflow to construct gene coexpression networks.

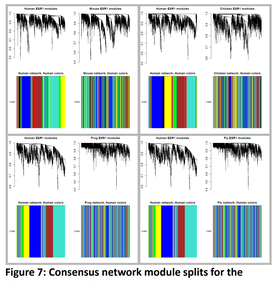


Figure 7: Consensus network module splits for the ER/Aromatase AOPs.

For each human-to-query organism comparison, the KEGG pathway genes common to both sets were selected based on NCBI HomoloGene[6] mapping, and the Weighted Gene Coexpression Network Analysis (WGCNA)[7] methodology was used to calculate expression correlation. The resulting topological distance values of the human reference set were used to cluster genes hierarchically and create modules of genes with similar expression patterns. Module assignments were transposed onto the query organism to examine how well the modules were preserved (**Figure 7**).

For the AHR Activation AOP, chicken showed a wide range of module preservation (**Table 2**). Genes in the yellow module were coexpressed similarly between Human and Chicken, but the expression patterns of genes in the black module showed significant divergence. A diagram showing the distribution of modules across the Chicken co-expression network is shown in **Figure 8**, where edge lengths indicate connectivity score (a measure of expression correlation between two genes). Expression connectivity in the black module (as defined by the human reference) is not reflected well in the chicken expression network, seen in the relatively scattered nature of black nodes. Contrast this with the highly-preserved yellow module, which shows more defined clustering.

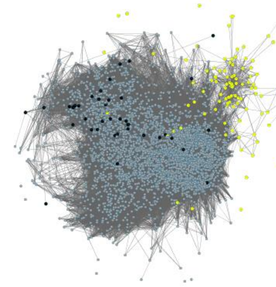


Figure 8: Chicken AHR AOP coexpression network.

Module Color	Preservation	Highly enriched Gene Ontology terms
black	poor	Endothelial cell development (GO:2001028, GO:0001938, GO:0010595)
blue	high	Peroxisome activity (GO:0016558, GO:0006625, GO:0072663)
brown	high	Membrane transport (GO:0090383, GO:0015991, GO:0015988)
green	moderate	Nucleotide biosynthetic processes (GO:1900373, GO:0030810)
magenta	high	ATP synthesis and transport (GO:0042776, GO:0006122, GO:0015986)
pink	high	Acetyl-CoA biosynthetic processes (GO:0061732, GO:0006086)
red	high	Xylulose biosynthetic processes (GO:1901159, GO:0051167, GO:0019640)
turquoise	poor	Pyrimidine biosynthetic processes (GO:0009174, GO:0009130)
yellow	very high	Catabolic processes (GO:0006572, GO:0046113, GO:0006559)

Table 2: Chicken/Human module preservation and enriched ontological terms for the AHR case study.

Creation of the AOPdb

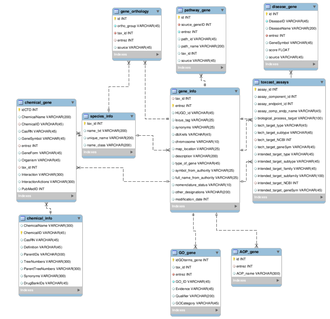


Figure 9: AOPdb Relational Database Schema

The AOPdb schema illustrated in **Figure 9** represents a relational database aggregating multiple types and sources of data related to AOPs through gene information. Database tables are derived from over 20 sources, including gene information from NCBI[8], disease information from DisGeNET[9], orthology information from KEGG[2], HomoloGene[6], and metaPhOrs[10], and chemical information from CTD.[11]

With these connections, we can query for the presence of AOP gene orthologs across multiple data sources, as well as investigate associations between genes, chemicals, and diseases for the discovery and development of new AOPs.

Analysis of ER/Aromatase pathways showed a taxonomic relevance cutoff at the vertebrate level, where the non-vertebrate model (Fruit Fly) was shown to lack orthologs for the ER and thus was predicted to be not-susceptible. Using the gene-orthology relationships in the AOPdb, we were able to further refine our susceptibility prediction, excluding all mammal model organisms based on their lack of a functional ortholog for vitellogenin (VTG), the egg yolk protein.

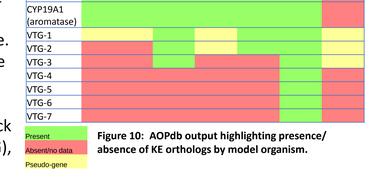


Figure 10: AOPdb output highlighting presence/ absence of KE orthologs by model organism.

Disease Name	Hit Count
Tobacco Use Disorder	18
Weight Gain	16
Nicotine Dependence	15
Alcoholic Intoxication, Chronic	11
Schizophrenia	10
Seizures	8
Alzheimer's Disease	7
Malignant neoplasm of lung	7
Obesity	7
Myasthenia Gravis	6
Substance-Related Disorders	5
Major Depressive Disorder	5
Bipolar Disorder	5
Mental disorders	5

Table 3: Diseases associated with KE genes for the AChE AOP.

The AOPdb can also be used to interrogate gene-disease and chemical-gene associations. From the list of KE genes participating in the AChE Inhibition case study, diseases were selected and sorted by the number of related AOP genes (**Table 3**). Hits for tobacco and alcohol abuse are explained by the involvement of nicotinic receptors in AChE pathways. We also see multiple hits for mental disorders, including depression, bipolar, and schizophrenia. Studies of the effects of organophosphate pesticides, shown in the AOPdb to have strong associations with the acetylcholinesterase gene, have found correlations between pesticide exposure and negative neurobehavioral effects.[12]

Conclusions/Future Directions

We have presented a computational workflow to assess mechanistic and functional similarity of AOP activity across species using curated pathway information, gene expression analysis, and aggregation of public databases. For ER/Aromatase Inhibition Leading to Reproductive Dysfunction, we exclude from our susceptibility prediction invertebrates and mammals lacking KE orthologs, finding a potential susceptibility range of vertebrate egg-laying species. For the AHR Activation case study, we find significant differences in expression of genes relating to endothelial development between chicken and other vertebrates, suggesting that the Cardiotoxicity pathway may progress differently across these species. For AChE Inhibition Leading to Acute Mortality, we find broad similarity across all model animal species and mine the AOPdb to show potential evidence for a new AOP: AChE disturbances leading to mental disorders. Moving forward, we hope to use coexpression networks to further investigate these AOP case studies across organisms. We continue to mine the AOPdb for further hypothesis discovery for the development of Adverse Outcome Pathways.

References
1. AOP wiki, http://aopwiki.org.The Organisation for Economic Co-operation and Development (OECD). Accessed October 2016.
2. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44, D457-D462.
3. Wheeler & Brandli (2009). Simple vertebrate models for chemical genetics and drug discovery screens: lessons from zebrafish and Xenopus. Dev Dyn 238:1287-1308.
4. Bastian F., Parmentier G., Roux J., Moretti S., Laudet V., Robinson-Rechavi M. (2008) Bgee: Integrating and Comparing Heterogeneous Transcriptome Data Among Species.in DILS: 5109:124-131.
5. Obayashi T, Okamura Y, Ito S, Tadaka S, Motoike IN, Kinoshita K. (2009). COXPRESdb: a database of comparative gene coexpression networks of eleven species for mammals. Nucleic Acids Res. 41:D1014-20.
6. Wheeler,D.L., Church,D.M., Edgar,R., et al. (2005) Database resources of the National Center for Biotechnology Information: update. Nucleic Acids Res., 32, D39–D45.
7. Peter Langfelder and Steve Horvath (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559.
8. Donna Maglott, Jim Ostell, Kim D. Pruitt and Tatiana Tatusova. (2007). Entrez Gene: gene-centered information at NCBI. Nucl. Acids Res.doi: 10.1093/nar/gkl993.
9. Piñero et al. (2015). DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database 2015:bav028.
10. Leszek P. Pryszcz, Jaime Huerta-Cepas and Toni Gabaldón. (2010). MetaPhOrs: orthology and paralogy predictions from multiple phylogenetic evidence using a consistency-based confidence score. Nucl. Acids Res. doi: 10.1093/nar/gkq953.
11. Davis, Allan et al. (2008). Comparative Toxicogenomics Database: a knowledgebase and discovery tool for chemical–gene–disease networks. Nucl. Acids Res. 37 (suppl 1): D786-D792.
12. Saeedi Saravi SS, Dehpour AR.(2016). Potential role of organochlorine pesticides in the pathogenesis of neurodevelopmental, neurodegenerative, and neurobehavioral disorders: A review. Life Sci. 2016 Jan 15;145:255-64.

Contact: Dr. Holly Mortensen
US Environmental Protection Agency
mortensen.holly@epa.gov