Bioinformatic integration of in vivo data and literature-based gene associations for prioritization of adverse outcome pathway development

Jason Brown[1], Sean Watford[1], Katie Paul Friedman[2]

[1]Oak Ridge Institute for Science Education Student Contractor, Oak Ridge, TN, 37831 [2]U.S. Environmental Protection Agency, National Center for Computational Toxicology, Research Triangle Park NC USA

1756/ P132
March 10-14, 2019
Society of Toxicology Annual Meeting Baltimore, MD

Jason Brown | brown.jason@epa.gov | 919-541-5098

www.epa.gov

Abstract

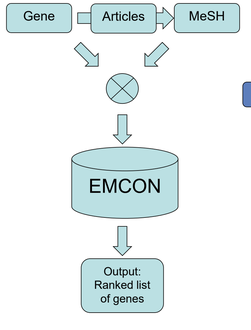
Adverse outcome pathways (AOPs) describe a sequence of events, beginning with a molecular initiating event (MIE), proceeding via key events (KEs), and culminating in an adverse outcome (AO). A challenge for use of AOPs in a safety evaluation context has been identification of MIEs and KEs relevant for AOs observed in regulatory toxicity studies. In this work, we implemented a bioinformatic approach that leverages mechanistic information in the literature and the AOs measured in regulatory toxicity studies to prioritize putative MIEs and/or early KEs for AOP development relevant to chemical safety evaluation. The US Environmental Protection Agency Toxicity Reference Database (ToxRefDB, v2.0) contains effect information for >1000 chemicals curated from >5000 studies or summaries from sources including data evaluation records from the US EPA Office of Pesticide Programs, the National Toxicology Program (NTP), peer-reviewed literature, and pharmaceutical preclinical studies. To increase ToxRefDB interoperability, endpoint and effect information were cross-referenced with codes from the United Medical Language System, which enabled mapping of in vivo pathological effects from ToxRefDB to PubMed (via Medical Subject Headings or MeSH). This enabled linkage to any resource that is also connected to PubMed or indexed with MeSH. A publicly available bioinformatic tool, the Entity-MeSH Co-occurrence Network (EMCON), uses multiple data sources and a measure of mutual information to identify genes most related to a MeSH term. Using EMCON, gene sets were generated for endpoints of toxicological relevance in ToxRefDB linking putative KEs and/or MIEs. The Comparative Toxicogenomics Database was used to further filter important associations. As a proof of concept, thyroid-related effects and their highly associated genes were examined, and demonstrated relevant MIEs and early KEs for AOPs to describe thyroid-related AOs. The ToxRefDB to gene mapping for thyroid resulted in >50 unique gene to chemical relationships. Integrated use of EMCON and ToxRefDB data provides a basis for rapid and robust putative AOP development, as well as a novel means to generate mechanistic hypotheses for specific chemicals. ***This abstract does not necessarily reflect U.S. EPA policy.***

Database resource of *in vivo* adverse outcomes: ToxRefDB

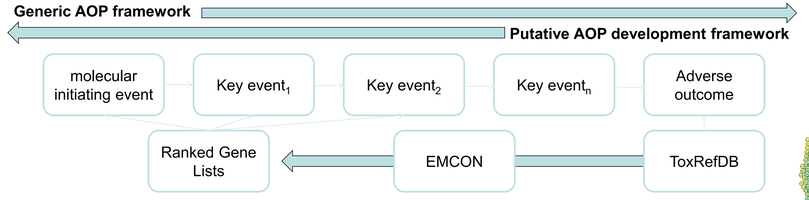
- ToxRefDB contains AO information for >5000 studies and > 1000 substances.
- Initial hypothesis: MIEs for AOP development can be prioritized on the basis of AOs observed in ToxRefDB.
- Do critical effects from *in vivo* toxicology studies indicate putative MIE?
- Frequency of critical effects for all study types in ToxRefDB suggests that bodyweight and food consumption dominate.
- How can we identify putative MIEs of interest that link to AOs relevant to safety evaluation?

Bioinformatic tool to link phenotype to genes involved

- Genes may indicate targets for MIEs. Bioinformatic tools may help us connect ranked gene lists (based on literature association) to AOs.
- EMCON incorporates multiple literature-based database resources to yield gene-MeSH associations. [Watford et al. 2018, Novel application of normalized pointwise mutual information (NPMI) to mine biomedical literature for gene sets associated with disease: Use case in breast carcinogenesis]
- 7 resources integrated: gene2pubmed, GeneRIF, CTD, UniProt/SwissProt, Reactome, Rat Genome Database, Mouse Genome Informatics to develop a network of naïve GeneID-MeSH associations.
- Normalized MeSH term mapping (map parents to all children, such that more specific terms show up less often and less specific terms are more promiscuous).
- Ranking of GeneID-MeSH associations using normalized pointwise mutual information
- One use case for applying EMCON is to understand putative MIE targets (indicated by genes) for AOs of interest (indicated by MeSH terms).



A Putative AOP development framework

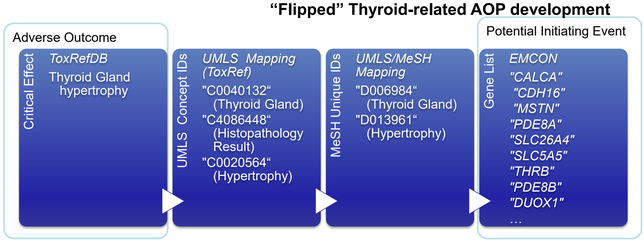


- The development of a putative AOP framework included flipping the generic AOP framework and starting with AOs fromToxRefDB.
- Starting with ToxRefDB increases the likelihood of toxicological importance for putative AOPs.
- ToxRefDB 2.0 endpoints and effects are mapped to UMLS codes for bioinformatic integration and semantic interoperability
- Made use of the newly-mapped standardize language and EMCON to create a linkage between critical effect and a list of genes as potential initiating events

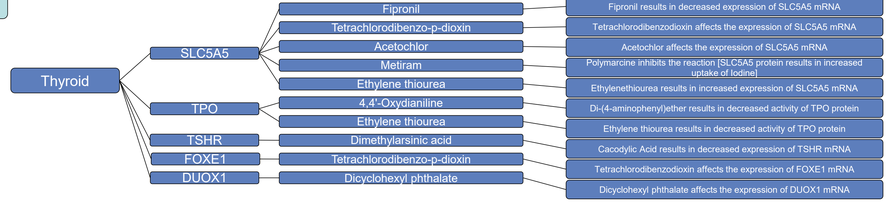
Thyroid-specific proof of concept

- Thyroid, a tissue for which the putative MIEs are well understood, was used as a proof of concept
- Mapped gene relationships were filtered through The Comparative Toxicogenomics Database to get a list of relevant chemical-gene interactions

"Flipped" Thyroid-related AOP development



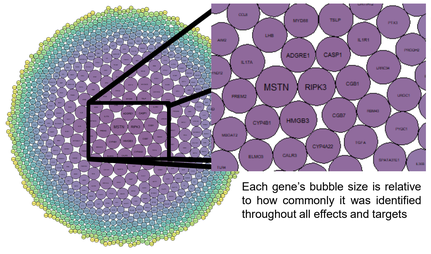
Cross-referencing ToxRef-EMCON with CTD reveals relevant associations



- Top associations in the intersection appear to be relevant, generally relating to previously understood AOPs
- This proof of concept suggests that this workflow may be a reasonable way to prioritize AOPs.
- What about tissues or endpoints for which there is little curated information to cross-reference for relevance?

ToxRefDB-EMCON intersection with ToxCast targets

Most Common Global Genes



Each gene's bubble size is relative to how commonly it was identified throughout all effects and targets

- n = # effects where gene was returned by EMCON
- total # effects for each endpoint target
- local freq = how common was the gene returned in this specific endpoint target (n/total)
- global_freq = % endpoint targets that had this gene returned by EMCON (out of all 88 targets)
- freq_ratio = local/global frequency; indicates whether the gene is expected to be specific to the endpoint_target

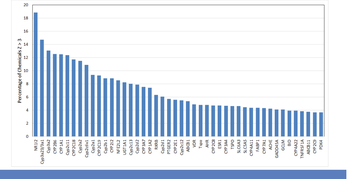
Most specific genes identified

Endpoint target	gene	n	Total effects	local_freq	global_freq	freq_ratio
body weight	PYY	17	17	1	0.011364	88
pituitary gland	POU1F1	16	16	1	0.011364	88
adrenal gland	PNMT	31	32	0.96875	0.011364	85.25
urinary bladder	CHRM3	31	32	0.96875	0.011364	85.25
urinary bladder	UPK2	31	32	0.96875	0.011364	85.25
body weight	TMEM18	16	17	0.941176	0.011364	82.82
adrenal gland	ARMC5	30	32	0.9375	0.011364	82.5
pituitary gland	HESX1	15	16	0.9375	0.011364	82.5
prostate	HOXB13	15	16	0.9375	0.011364	82.5
prostate	MSMB	15	16	0.9375	0.011364	82.5

Ranked intersection of ToxRef-EMCON and top intended targets from ToxCast

- ToxCast is a high-throughput screening bioactivity data resource.
- Intended targets with higher hit rates, free from cytotoxicity, may be the most interesting targets that we have information regarding xenobiotic perturbation
- Cross-reference ToxCast targets with MIEs from the EMCON-ToxRef approach to bolster confidence that we are selecting important targets of regulatory importance

Hit-rate (cytotoxicity-filtered) for ToxCast targets



Conclusions

- Integrated use of EMCON and ToxRefDB data provides a novel basis for rapid and robust putative AOP development.
- The ToxRefDB to gene mapping for thyroid resulted in >50 unique gene to chemical relationships. This proof-of-concept demonstrates for a tissue with well-known MIEs and available curated information that EMCON is effective, returning known MIEs.
- The ToxRefDB-EMCON approach may also inform hypothetical reference chemical lists.
- Combining high-frequency (and specific) targets from ToxCast with the ToxRefDB-EMCON ranked targets provides an informative list of genes for MIE discovery that are likely perturbed by xenobiotics.
- ToxRefDB-EMCON may provide an approach for understanding how to predict phenotypes in ToxRefDB using HTTr data.

U.S. Environmental Protection Agency
Office of Research and Development