ANNOTATING ADVERSE OUTCOME PATHWAYS TO ORGANIZE TOXICOLOGICAL INFORMATION FOR RISK ASSESSMENT

Cataia L. Ives

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the Department of Environmental Science & Engineering.

Chapel Hill
2016

Approved by:

Stephen Edwards

Rebecca Fry

Jacqueline MacDonald Gibson

ABSTRACT

Cataia L. Ives: Annotating Adverse Outcome Pathways to Organize Toxicological Information for Risk Assessment
(Under the direction of Stephen Edwards)

The Adverse Outcome Pathway (AOP) framework connects molecular perturbations with organism and population level endpoints used for regulatory decision-making by providing a conceptual construct of the mechanistic basis for toxicity. Development of an AOP typically begins with the adverse outcome, and intermediate effects connect the outcome with a molecular initiating event amenable to high-throughput toxicity testing (HTT). Publicly available controlled vocabularies were used to provide terminology supporting AOP's at all levels of biological organization. The resulting data model contains terms from 22 ontologies and controlled vocabularies annotating currently existing AOP's. The model provides the ability to attach evidence in support of the AOP, supports data aggregation, and promotes the development of AOP networks. Long term, this structured description of the AOP will enable logical reasoning for hazard identification and for dose-response assessment. Case studies showcase how the model informs AOP development in the context of chemical risk assessment.

ACKNOWLEDGEMENTS

I would like to thank my research advisor, Stephen Edwards, PhD., for his support and mentorship throughout this project, as well as for encouraging collaborative opportunities that allowed me to learn more about a field that is new and challenging. I would also like to thank my academic advisor, Rebecca Fry, Ph.D., for her advice and collaboration on this project, as well as the other member of my committee, Jackie Macdonald Gibson, Ph.D. for her insights. I'm grateful to my colleagues at the US EPA ORD- Environmental Protection Agency National Health and Environmental Effects Research Laboratory, for their time and expert opinions as this project came together. Also, I'd like to thank my professors, who in these two years opened me to ways of thinking about environmental health. I'd like to thank the ESE student services faculty for always being there when I needed academic support.

Most of all, I'd like to thank my parents, my Carolina friends and colleagues, and even my Duke friends for their support throughout graduate school.

The research described has been partially supported by the U.S. Environmental Protection Agency. The views expressed are of the authors and do not necessarily reflect the views or policies of the U.S Environmental Protection Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACToR	Aggregate Computational Toxicology Resource
AFB1	Aflatoxin B1
AHF	Altered Hepatic Foci
AO	Adverse Outcome
AOP	Adverse Outcome Pathway
AOPKB	Adverse Outcome Pathway Knowledgebase
AOPO	Adverse Outcome Pathway Ontology
API	Application Program Interface
AR	Androgen Receptor
ASA	Anatomical Structural Abstraction
At	Anatomy taxonomy
ATA	Anatomical Transformation Abstraction
BAO	BioAssay Ontology
BDNF	Brain-derived neurotrophic factor
BEL	Biological Expression Language
BFO	Basic Formal Ontology
BRENDA	Braunschweig Enzyme Database
BTO	Brenda Tissue Ontology
Ca^{2+}	Calcium ion
CARO	Common Anatomy Reference Ontology
CAS	Chemical Abstracts Service Registry Number
CHEBI	Chemical Entities of Biological interest

CL	Cell Ontology
CLP	Classification and Labelling
COSMOS	Integrated In Silico Models for the Prediction of Human Repeated Dose Toxicity of Cosmetics to Optimise Safety
cpAOP	Computationally Predicted Adverse Outcome Pathway
CTD	Comparative Toxicogenomics Database
DAG	Directed Acyclic Graphs
DAS	Direct-attached Storage
DAVID	Database for Annotation, Visualization, and Integrated Discovery
DEP	Diesel Exhaust Particles
DSSTox	Distributed Structure-Searchable Toxicity
EBI	European Bioinformatics Institute
EFO	Experimental Factor Ontology
EMBL	European Molecular Biology Laboratory
ENCODE	Encyclopedia of DNA Elements
ER	Estrogen receptor
ERDC	US Army Engineer Research & Development Center
ERO	Eagle-I Resource Ontology
EU JRC	European Commission's Joint Research Centre
FDA	Food and Drug Administration
FIFRA	Federal Insecticide, Fungicide and Rodenticide Act
FMA	Foundational Model of Anatomy
GEO	Gene Expression Omnibus

Glu	L-glutamate
GO	Gene Ontology
HapMap	International HapMap Project
HCC	Hepatocellular Carcinoma
HCS	High Content Screening
HPO	Human Phenotype Ontology
HUPO-PSI	Human Proteome Organization Proteomics Standards Initiative
HTML	Hyper Text Markup Language
HTS	High Throughput Screening
iAs	Inorganic Arsenic
IATA	Integrated Approaches to Testing and Assessment
ICD	International Classification of Diseases
ICN	International Code of Nomenclature for algae, fungi, and plants
ICNB	International Code of Nomenclature of Bacteria
ICZN	International Commission on Zoological Nomenclature
IMPC	International Mouse Phenotyping Consortium
IMPReSS	International Mouse Phenotyping Resource of Standardised Screens
InChI	IUPAC International Chemical Identifier
INN	International Nonproprietary Name
INSDC	International Nucleotide Sequence Database Collaboration
IRIS	Integrated Risk Information System
IUPAC	International Union of Pure and Applied Chemistry
KE	Key Event

KEGG	Kyoto Encyclopedia of Genes and Genomes
KER	Key Event Relationship
LINCS	Library of Integrated Network-based Cellular Signatures
MA	Mouse Anatomy Ontology
MESH	Medical Subject Headings
MGI	Mouse Genome Informatics
MI	Protein-protein interaction
MIE	Molecular Initiating Event
Mk	Metaknowledge components
MOA	Mode-of-Action
MOD	Model Organism Database
MP	Mammalian Phenotype
NAS	National Academies of Sciences
NCBI	National Center for Biotechnology Information
NCCT	National Center for Computational Toxicology
NCEA	National Center for Environmental Assessment
NCGC	National Institutes of Health Chemical Genomics Center
NIH	National Institutes of Health
NLM	National Library of Medicine
NMDAR	N-methyl-D-aspartate receptors
NRC	National Research Council
NTP	National Toxicology Program
OBI	Ontology for Biomedical Investigations

OBO	The Open Biological and Biomedical Ontologies Foundry
OECD	Organization for Economic Co-operation and Development
OGG	Ontology for Genes and Genomes
Oort	OBO Ontology Release Tool
OLS	Open Lookup Service
OWL	Web Ontology Language
PATO	Phenotypic Quality Ontology
PBPK	Physiologically based pharmacokinetic model
PCO	Population and Community Ontology
PDBeChem	Protein Data Bank in Europe
PPARα	Peroxisome Proliferator-Activated Receptor
PR	Protein Ontology
RDF	Resource Description Framework
REST	Representational State Transfer
RO	Relations Ontology
QSAR	Quantitative structure-activity relationship
RDF	Resource Description Framework
REACH	Regulation, Evaluation, Authorization and Restriction of Chemicals
SCR	Supplementary Concept Record
SEURAT-1	Towards the Replacement of In Vivo Repeated Dose Systemic Toxicity Testing
SMILES	Simplified Molecular-Input Line-Entry System
SO	Sequence Ontology
STAR	Steroidogenic Acute Regulatory Protein

TaxID	Taxonomy ID
TFHA	Task Force Hazard Assessment
Tox21	Toxicity Testing in the 21st Century
ToxCast	Toxicity ForeCaster
ToxNET	Toxicology Data Network
ToxRefDB	Toxicity Reference Database
TSPO	Translator Protein
Uberon	Uber Anatomy Ontology
UMLS	Unified Medical Language System
UNC	University of North Carolina
UniProtKB	UniProt Knowledgebase
URI	Uniform Resource Identifier
U.S. EPA	United States Environmental Protection Agency
W3C	World Wide Web Consortium
WoE	Weight of Evidence
WNT	National Coordinators Test Guidelines
XAO	*Xenopus* Anatomy Ontology
XML	Extensible Markup Language
ZFA	Zebrafish Ontology

CHAPTER 1: INTRODUCTION, THE ADVERSE OUTCOME PATHWAY FRAMEWORK, OBJECTIVES, AND HYPOTHESES

Introduction

Traditional toxicity testing has focused on *in vivo* testing for whole-organism outcomes of exposure, such as a clinical sign or pathologic state resulting from exposure to a toxicant (OECD, 2013a). Traditional testing is costly, time-consuming, and uses live animal models; while alternative *in vitro* and *in silico* approaches increase the efficiency of testing and reduce the use of live animals. Today, there are 82,000 chemicals in industry, and 700 new ones introduced each year; as opposed to 62,000 chemicals in 1979 (as cited in Krewski et al., 2010, p.13). The need to assess an ever-increasing number of chemicals while reducing animal use and decreasing cost and time of testing, requires a plan for predictive toxicity testing and for organizing the data created in a manner useful for regulatory assessment.

In a 2007 report entitled, *Toxicity Testing in the 21st Century*, the National Academies of Sciences described the future of toxicity testing to support human health risk assessment (NRC, 2007). The report envisions "increased efficiency in toxicity testing and decreased animal usage by transitioning from current expensive and lengthy *in vivo* testing with qualitative endpoints to *in vitro* toxicity pathway assays on human cells or cell lines using robotic high-throughput screening with mechanistic quantitative parameters" (Krewski et al., 2010). Some limitations that the NRC committee concluded needed to be addressed were: 1) The need to focus resources on the evaluation of the most sensitive adverse effects of exposures of greatest concern, rather than fully characterizing all adverse effects regardless of risk assessment needs. 2) Setting priorities as

a component of any testing strategy designed to address a large number of chemicals. 3) Major gaps in current toxicity-testing approaches, and the promise of emerging screening technologies. 4) Evaluation of testing strategies with respect to value of information in regards to depth, breadth, animal welfare, and conservation (Krewski et al., 2010).

Programs in the USA and in Europe have focused on addressing some of these limitations. The Regulation, Evaluation, Authorization and Restriction of Chemicals (REACH) legislation in 2007 required information on chemical safety while limiting animal testing in favor of alternative testing methods (Commission, E.) The Tox-21 (Attene-Ramos et al., 2013) and SEURAT-1 (Berggren et al., 2015) initiatives favor the use of high throughput, *in vitro* testing methods for chemical safety evaluation over traditional *in vivo* animal toxicity studies (Collins et al., 2008). The EU COSMOS project addresses the safety assessment needs of the cosmetics industry without the use of animals.

Joint efforts by the EPA's National Center for Computational Toxicology (NCCT), National Toxicology Program (NTP), and National Center for Advancing Translational Sciences (NCATS) address increased needs for chemical screening based on alternative predictive methods that are target-specific, mechanism-based, biological observations *in vitro* (Collins et al., 2008). *In vitro* methods examine chemical effects at molecular targets on the basis of toxicity pathways, "cellular response pathways that, when sufficiently perturbed in an intact animal, are expected to result in adverse health effects" (NRC, 2007). An example pathway perturbation is the response of cells to oxidative stress caused by exposure to diesel exhaust particles (DEP) within a cellular response network. *In vitro* exposures to DEP lead to dose-related activation of a set of biological pathways. DEPs activate normal adaptive signaling pathways associated with maintaining homeostasis; however, when oxidant exposure is

sufficient to overwhelm these adaptive processes, these toxicity pathways then lead to adverse health effects (NRC, 2007). Thus, the NRC report describes testing these pathways through *in vitro* assays that measure "critical mechanistic end points involved in the induction of overt toxic effects rather than the effects themselves" (as cited in Ankley et al., 2010).

In vitro assays, or experiments, can be adapted for high-throughput, the primary analysis method for drug discovery. Once an assay has been adapted to a fully automated system, 10,000 chemicals per week can be screened (Attene-Ramos et al., 2013). Assays test the effect of a perturbing agent in a biological system, measuring effects of the agent translating the perturbation into a detectable signal to arrive at endpoint(s) that quantify the extent of the perturbation (Abeyruwan et al., 2014). Typical drug discovery methods test compounds at one concentration between 2 and 10 uM, which can lead to high rates of false negatives. Newer methods test all compounds at as many as fifteen concentrations, from ~5 nm to ~100 uM, to generate a concentration-response curve. This approach produces lower false positive and false negative results (Collins et al., 2008). HTS of a large set of chemicals can be used to prioritize chemicals for further testing, by comparing the profile of positive assay results for a test chemical with known toxicity "signatures" based on reference chemicals of known toxicity. U.S. EPA's ToxCast project has assessed nearly 2,000 chemicals in 700 assays covering a range of endpoints to prioritize chemicals for further testing (USEPA, 2013), with the eventual goal of screening thousands of environmental chemicals.

Current toxicity databases vary in breadth and depth of information, with many structure-searchable chemical databases existing and a lack of uniform guidelines regarding their content and outputs. PubChem contains information on chemical structure and HTS assay endpoints for more than 10 million unique compounds, and represents a compilation of multiple data sources

(Judson et al., 2008). EPA structure-searchable databases include DSSTox, ToxRefDB, and ACToR. ToxRefDB houses screening data from the ToxCast effort (Dix et al., 2007). Aggregated Computational Toxicology Resource (ACToR) is a database and set of software applications from multiple sources including *in vitro* and *in vivo* assay results to support the ToxCast effort (Judson et al., 2008). ToxCast chemical data is publicly available through the interactive Chemical Safety for Sustainability Dashboards (iCSS) (USEPA, 2013). Additional informatics tools such as DSS Tox (Distributed Structure Searchable Toxicity), located at www.epa.gov/ncct/dsstox/ provide a structure-searchable database of chemicals linked to physicochemical and toxicological data, facilitating linking HTS data to historical toxicological test results (Collins et al., 2008). Systems such as ToxNET store literature citations and summary toxicity information (Martin, 2008).

Computational approaches to assessing priority environmental chemicals might be used to predict chemical properties, estimate biological activity and potency, and build quantitative structure activity relationship (QSAR) models that predict biological activity from molecular structure (Krewski et al., 2010). Physiologically based pharmacokinetic (PBPK) models allow dose extrapolation from *in vitro* conditions to projected *in vivo* exposure to provide a quantitative model for humans and animals (Krewski et al., 2010). Dose response and extrapolation modelling integrate mechanistic and dosimetric information about toxicity into mathematical terms, resulting in a quantitative model allowing for dose and interspecies extrapolation (Conolly, 2002).

Collins et al. describes several initiatives that are transforming toxicology (2008). Progress has been made in several areas including the examination of genetic diversity of human and animal responses to known toxicants, such as the International HapMap project

(http://www.hapmap.org/) to evaluate differential sensitivity of human cell lines. The Host Susceptibility Program tests for the genetic basis for differences in disease response in various mouse models. Cell lines from these models will be evaluated for differential sensitivity to HapMap compounds, with an ultimate goal of establishing *in vitro* signatures of *in vivo* rodent and human toxicity tests (2008). Virtual tissue models such as the v-Liver project use computational models to simulate how chemicals may affect human development, reducing the use of animal models (Shah & Corton, 2010).

The NRC report envisions that traditional *in vivo* tests will most likely continue to be used to evaluate the formation of metabolites and for some mechanistic aspects of target-organ responses to environmental agents (Krewski et al., 2010). *In vitro* tests will comprise the bulk of toxicity tests and may be conducted using human cell lines, animal cell lines transfected to express human genes and proteins, and metabolic screens with human enzymes (Krewski et al., 2010). In order to support HTS and other emerging alternatives to animal testing, a framework is needed to organize disparate information from these information sources. The adverse outcome pathway (AOP) framework provides a way to connect perturbations of biological processes at the molecular and cellular levels with apical endpoints useful in regulatory decision-making (Ankley et al., 2010).

The Adverse Outcome Pathway Framework

In the context of the alternative testing methods described and changes in the way that toxicity pathways are evaluated, Ankley, et al. proposed the Adverse Outcome Pathway (AOP) framework as a conceptual construct that portrays existing knowledge concerning the linkage between a direct molecular initiating event, the molecular interaction between a xenobiotic and a specific biomolecule, and an adverse outcome at a biological level of organization relevant to

risk assessment (2010). While toxicity pathways focus on molecular-level initating events and the resulting cellular responses that can be monitored *in vitro*, the AOP framework is intended to connect these events to an adverse outcome relevant to regulatory assessment (Ankley et al., 2010). This can then provide the mechanistic context for defining the toxicity signatures used to identify chemical toxicities from HTS results and links those toxicity signatures to a defined regulatory endpoint.

The AOP definition is also distinct from similar pathway-based concepts in the field of predictive toxicology, mechanism of action and mode of action. Mechanism of action is defined as "a complete and detailed understanding of each and every step in the sequence of events that leads to a toxic outcome" (ECETOC, 2007). Mode of action (MOA) describes a "biologically plausible series of key events leading to" an adverse effect (Meek et al., 2014).The AOP differs from these concepts in that it is chemical agnostic and is focused more on the interpretation of the *in vitro* and in silico toxicity predictions than assembly of mechanistic information for a specific chemical (OECD, 2013a). Villeneuve et al. outline a number of core principles underlying AOP development intended to address uncertainties and foster consistency (2014a, 2014b).

The AOP Development Programme was launched in 2012 and defines a workflow for developing, reviewing, and publishing an AOP. A project for development is proposed at OECD, reviewed by the Extended Advisory Group for Molecular Screening and Toxicogenomics (EAGMST), and, if approved, included in a workplan for development. Upon development of draft AOP's, the AOP's then return to OECD for revisions and consultation. The final document is submitted to the WNT (National Coordinators Test Guidelines) and the TFHA (Task Force

Hazard Assessment) for endorsement. The final endorsed document is declassified and published as an OECD report (OECD, 2013a).

A Users' Handbook Supplement by the OECD EAGMST provides guidance for the systematic development and assessment of AOP's (OECD, 2013b). The molecular initiating event explains how the chemical being assessed interacts with a biological target. Identifying the final adverse event, or "adverse outcome", involves determining the most relevant mechanistic information and is often associated with *in vivo* toxicological test endpoints. Strategies for AOP development can either be top-down beginning with the AO, bottom up beginning with the MIE, or start with a middle entity in the AOP depending on the use case (Villeneuve et al., 2014a). Data from *in vitro* and test methods can be used here to identify key events and provide scientific evidence supporting the AOP. Scientific evidence could include: structural alerts for chemicals involved in the initiating event, *in chemico* methods measuring chemical-biological interactions, *in vitro* assays for cellular response, *ex vivo* and *in vivo* mechanistic tests and *in vivo* tests that measure endpoints directly relevant to the adverse outcome of interest (Figure 1). AOP integrates events at different levels of biological organization: molecular, cellular, tissue, organ, organism, and population-level effect (OECD, 2013b).

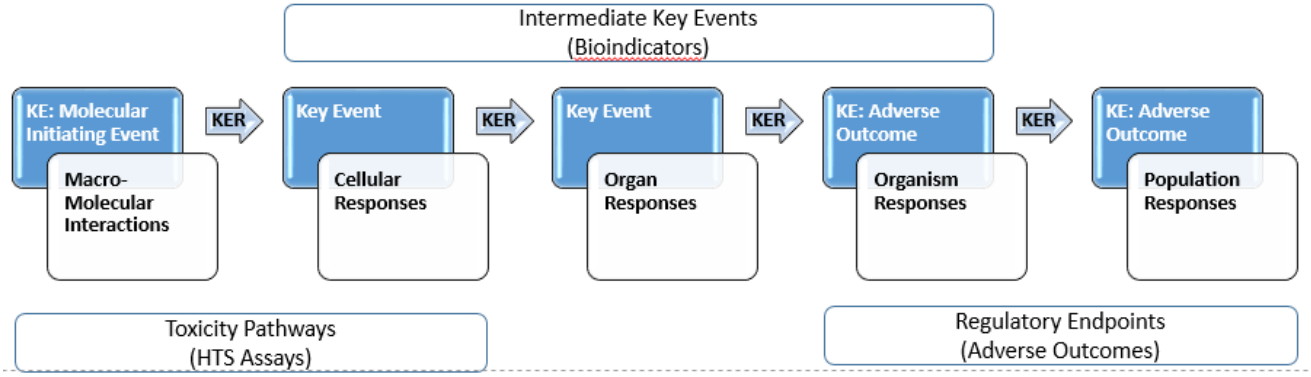

Figure 1. Framework of an AOP beginning at the MIE and ending at the AO, showing KE's at levels of biological organization from Macromolecular to Population, connected by KER's. HTS Assays, Bioindicators, and Adverse Outcomes represent attaching scientific evidence to KE's. (Source: Stephen Edwards).

A molecular initiating event describes the perturbation resulting from interaction of a chemical or non-chemical stressor with a biological target (OECD, 2013a; Villeneuve et al., 2014a). Many MIE's are defined in terms of covalent binding to protein or DNA. In contrast to this electrochemical interaction, enzyme-receptor binding is usually a non-covalent interaction based on affinity for the target. Understanding of how the properties of the drug or chemical, such as bioavailability, structural requirements, and metabolic transformation affect affinity for the target helps define this interaction. Often, a list of known "chemical initiators" (OECD, 2013a; Villeneuve et al., 2014a) will be supplied to provide evidence for the initiating event. For example, 17a- ethynyl estradiol is a prototypical ER agonist. Groups of chemicals with structural features known to trigger the MIE could be supplied as well. A single MIE can trigger multiple different AOPs. For example, binding of estrogen-like compounds to the estrogen receptor can result in several different AOs driven by multiple AOPs. Endogenous estrogen binding to the estrogen receptor also represents an intermediate key event in several AOPs that are initiated by perturbations in estrogen synthesis or regulation. A single molecular target may be active at different sites and therefore represent multiple separate MIEs. An example of this is covalent protein binding, which can result in skin sensitization or liver fibrosis depending on whether it occurs in keratinocytes or hepatocytes.

The adverse outcome usually represents an outcome relevant to regulatory decision-making, either an accepted protection goal or a common apical endpoint in an established study (OECD, 2013b; Villeneuve et al., 2014a). The AO may include long term health endpoints, such as cancer, and short term, local effects such as skin sensitization. Associated with an AO are descriptions of its biological state, how it can be measured, and its taxonomic applicability (OECD, 2013b; Villeneuve et al., 2014a).

A key event describes a change in biological state that is both measurable and essential to the progression of a defined biological perturbation leading to a specific adverse outcome (OECD, 2013b; Villeneuve et al., 2014a). KE's represent points along the pathway that can be used for monitoring the progression along the AOP. KE's are defined in terms of their biological context (tissue, taxa, or life stage), are specific, and are measurable (OECD, 2013b; Villeneuve et al., 2014b). This may include citation of methods used in their detection and the level of confidence in those measurements.

KE's are connected to one another through key event relationships (KERs), a directed relationship between an upstream and a downstream KE (2013b; Villeneuve et al., 2014a). The definition facilitates extrapolation from the known state of the upstream event to the predicted state of the downstream event. Scientific evidence supporting the KER is described with respect to qualitative and quantitative supporting evidence. The biological rationale or "biological plausibility" for a relationship between two KE's may be referenced. Empirical support, specific evidence for the relationship, should be cited, as well as uncertainties or inconsistencies in the relationship. Finally, quantitative relationships between KE's may be defined in terms of correlations, response-response relationships, or dose dependent relationships (OECD, 2013b; Villeneuve et al., 2014b).

The utility of the AOP Framework in hazard assessment is that the AOP framework provides the biological context and supporting weight of evidence (WoE) to facilitate the interpretation of data from *in vitro* and alternative testing methods (Becker et al., 2015, OECD, 2013b). The evaluation process for assessing confidence in supporting information as a basis for regulatory application of AOP's includes explicit evaluation of the evidence supporting each key event relationship and defining the domain of applicability for the AOP in terms of sex, life-

stage, taxa, and other aspects of biological context. The relative level of confidence for each key event relationship in the AOP is assessed based on biological plausibility, empirical support, and quantitative understanding. Level of confidence can be high, moderate or low weight of evidence based on extent of available experimental data. The extent of support for biological plausibility based on a mechanistic (structural or functional) relationships between the upstream KE and downstream KE for each KER should be considered. In addition, the extent of support for essentiality of each of the KEs is based on direct evidence from experimental studies. A defining question is, "Are downstream KEs or the AO prevented if an upstream KE is blocked?" Extent of empirical support for each of the KERS and the overall AOP, as related to concordance of dose response relationships between upstream and downstream KEs. Finally, the degree of quantitative understanding for each KER in the overall AOP is assessed as weak, moderate, or strong (OECD, 2013b).

An AOP represents a pragmatic unit of development and evaluation, and considering AOPs as a common network connected by shared KEs provides the basis for regulatory decision making (OECD, 2013b). The ultimate objective of AOP development is to support inference or extrapolation of information from one KE to another, allowing the creation of an AOP network based on multiple interacting AOPs sharing KEs or KERs. Assessment of supporting information for the overall hypothesized AOP in the Network View represents the degree of confidence in weight of evidence for both essentialities of the KEs and biological plausibility and empirical support for KERs (OECD, 2013b). These AOP networks act as functional units of prediction representing the interaction in response to mixtures or toxicants with multiple biological responses (Villeneuve et al., 2014a). Such a network, based on confidence in qualitative and quantitative elements of the AOP, could be applied to development or refinement

of test guidelines, development of Integrated Approaches to Testing and Assessment (IATA), which integrate and weigh data from different test methods for hazard identification and assessment of chemicals and address data gaps (Tollefsen et al., 2014), development of Quantified Structure Activity Relationships (QSARs), and screening level hazard assessments or risk assessments (OECD, 2013a).

The first AOP endorsed by the OECD was the AOP for Skin Sensitization Initiated by Covalent Binding to Proteins, which describes the regulatory hazards known as human allergic contact dermatitis or rodent contact hypersensitivity, a well-studied adverse outcome (2012). According to Landesmann, the MIE is covalent interaction of a target chemical or metabolite with cysteine or lysine residues (Landesmann, 2016). The second KE, taking place in the keratinocyte, includes inflammatory responses as well as gene expression associated with particular cell signaling pathways. The third KE is activation of dendritic cells, assessed by expression of specific cell surface markers, chemokines and cytokines. Dendritic cells subsequently mature and migrate out of the epidermis to the local lymph node where they display major histocompatibility complex molecules, which include part of the hapten-protein complex to naïve T-cells. This induces the fourth KE, T-cell proliferation, as measured in the current most applied *in vivo* test for skin sensitization, the murine Local Lymph Node Assay. The adverse outcome at the organ level is the acquisition of sensitization (Landesmann, 2016). This AOP resulted in OECD Test Guidelines based on non-animal methods for skin irritation. This AOP also forms the mechanistic basis for development of IATA for chemicals potentially causing skin sensitization (Landesmann, 2016).

Due to the shift from *in vivo* toxicity testing to rapid *in vitro* methods, there is an increased need to efficiently identify and develop new AOP's (Oki et al., 2016). A centralized

resource easily accessible to regulatory decision-makers is needed to provide structured information supporting AOP's. To meet these needs, the AOP knowledgebase (AOP-KB) (https://aopkb.org/) provides the repository for AOPs developed under the OECD program (Oki et al, 2016). The AOP-KB represents a joint collaboration between the OECD, the US Environmental Protection Agency (USEPA), the European Commission's Joint Research Centre (JRC), and the US Army Engineer Research & Development Center (ERDC). A resource for the scientific community to share, develop, and discuss AOP's, the project consists of four independently developed platforms- AOP-Wiki, Effectopedia, AOPXplorer, and Intermediate Effects Database (AOP-KB, 2016). The AOP-Wiki supports collaborative development of AOP descriptions, in an encyclopedia-style text format, and captures the evidence supporting AOP's (https://aopwiki.org). Stakeholders can build AOP's by entering them with linked information regarding MIE's, KE's, AO's, and Chemical Initiators. The Effectopedia provides detailed development of structured and computational AOP's via a graphical interface, including the ability to store quantitative models (http://effectopedia.org/). The AOPXplorer provides for visualization of attribute networks to discover and explore AOPs in a broader context (http://aopxplorer.org/). The Intermediate Effects Database puts chemical-related AOP components in a regulatory context, connecting AOP's to chemical-specific information (Edwards et al., 2016). AOP-KB additionally will support third party applications and plugins (AOP-KB, 2016).

AOP-KB supports all stages of AOP development. Information needs vary for AOP's in terms of data needs and confidence required, depending on the level of detail needed for a given AOP (Edwards et al., 2016). AOP-KB provides knowledge management for AOP's, including identifying data in a computable standardized format, thus facilitating data integration and

information sharing, including evidence integration (Oki et al., 2016). Current challenges for the AOP-KB are the need to exchange information among modules, standardize naming of KE's in support of AOP networks, develop computational methods to define AOP's accelerating AOP development, build models from AOP's, and provide the ability to attach evidence in support of the AOP, thus supporting data aggregation from HTS assays and experiments.

Objectives

The generation of vast amounts of toxicity data and the need to organize these data in a manner useful for risk assessors' efforts at prioritizing decision making regarding particular chemicals, has resulted in data management projects for the structured organization of these data for improved machine-readability and contribution to the knowledgebase for AOP development; for example, the annotation of HTS assay data from EPA's ToxCast (Phuong, 2014). As such, the AOP framework provides a scaffold onto which this data can be organized. Current challenges for AOP development include the need to support data at different levels of granularity from the molecular level up to the organism and population levels. Tying AOP components to existing biological ontologies and controlled vocabularies, will contribute to the body of knowledge for AOP development. Providing the ability for machine-learning to these components will assist developers of tools and software for browsing and querying AOP information.

The objective of this research is two-fold: 1) To provide data to support the AOP Framework at the higher levels of biological organization, by promoting a structured understanding of the supporting biological data and filling data gaps in the framework and 2) Using the resulting data model and approach to provide data from existing ontologies and controlled vocabularies for annotating a group of AOP's under development from EPA's

National Center for Environmental Assessment (NCEA), forming the basis of an Integrated Risk Information System (IRIS) assessment. One of these AOP's, a case example of an AOP for metal exposure culminating in adverse pregnancy outcomes, will demonstrate the formal AOP development process for a reproductive AOP that incorporates this standardized terminology.

1) *To provide data to support the AOP Framework at the higher levels of biological organization, by promoting a structured understanding of the supporting biological data and filling data gaps in the framework.* An ontology describes "a formal representation of a set of concepts within a knowledge domain and the relationships between those concepts" (Hardy, 2012b). Biological ontologies are developed in order to analyze domain knowledge and to provide consistent descriptions and machine-readability across the domain. AOP's at the tissue, organ, individual, and population level will be examined using existing biological ontologies, including Gene Ontology (GO), Mammalian Phenotype Ontology (MPO), and Uberon; and non-ontologies including controlled vocabularies like Medical Subject Headings (MESH).

   Literature review of a formal sub-set of ontologies will be conducted in order to select the most fitting vocabularies by their characteristics at the molecular, cellular, tissue, organ, organism, and population levels and to fit the data needs of existing AOP's. Once the ontologies are selected, literature review of priority existing AOP's in the AOP-Wiki (https://aopwiki.org/) will be conducted, and KE's matched to their supporting ontological entity for seventeen priority AOPs. For instance, a KE of reduced testosterone synthesis by steroidogenic tissue would be referenced to a specific ontology containing steroid hormone diseases related to testosterone. A data model will be chosen to best represent this data, the output being a table mapping AOP entities to ontology

entities (classes). Mapping KE terms from individual AOP's to controlled vocabularies and ontologies will result in the formation of a network of interconnected AOP's based on overlapping KE's or nodes as individual AOPs are entered into the AOP-KB. A review of case studies or specific AOP's will be conducted, to consider how ontological data in these AOP's provide support for the key events in the AOP and to show how the resulting data model works in representing this information.

2) *Using the resulting data model and approach to provide terms from existing ontologies and controlled vocabularies for annotating a group of AOP's under development from EPA's National Center for Environmental Assessment (NCEA), forming the basis of an IRIS assessment.* Using the data model formulated in the first part, entities from the selected ontologies and controlled vocabularies will be mapped to KE's in hypothesized AOP's for an IRIS assessment for inorganic Arsenic for NCEA. A case study of developing an AOP for environmental exposure will utilize the AOP-Wiki development tool to develop an AOP for metal exposure leading to adverse pregnancy outcomes as part of the IRIS assessment.

Hypotheses

1) Mapping AOP components to ontology classes provides a common syntax for AOP components, generating a data structure that promotes machine readability and use by software applications. Additionally, mapping these components furthers development of computationally predicted AOP's (cpAOP's) and development of the AOPKB by addressing ontological data gaps at the higher levels of biological organization.
2) The AOP framework describes mechanistic knowledge concerning the characterization of developmental effects resulting from metal exposure, furthering risk assessment for

metals with potential environmental health effects. Applying the resulting ontological framework from the first part of the project, will provide ontological data for the NCEA AOP's to support integrated risk assessment.

CHAPTER 2: THE USE OF ONTOLOGIES IN PREDICTIVE TOXICOLOGY AND LITERATURE REVIEW OF SELECTED ONTOLOGIES

The Use of Ontologies in Predictive Toxicology

According to Hardy, predictive toxicology requires the development of a framework of open, public, computable, standardized ontologies and vocabularies to support applications required by *in silico*, *in vitro*, and *in vivo* toxicology methods (2012a). An ontology is a formal representation of a set of concepts within a knowledge domain, and the relationships between those concepts (2012b). It represents a shared vocabulary in any domain, be it biomedical sciences, library science, or philosophy. The vocabulary might include terms used for categorizing content, building labeling systems, or creating a database schema (2012b). It is different from a controlled vocabulary, in that a controlled vocabulary provides a set of terms, while an ontology may include a controlled vocabulary and define the properties and relationships of those terms (2012b). In informatics, a database schema is an "ontology" for a database; in that it provides a description of the structure of the data. Using controlled vocabularies and ontologies in the domain of toxicology allows data to be organized and combined with metadata and vocabularies for study terms and experimental protocols (2012b).

In this context, development of ontologies applicable for specific uses requires an understanding of computer organization and retrieval systems. According to Robinson & Bauer, the elements required for computer reasoning include a formal knowledge representation language, a means for reasoning in the language, and a set of representations about a domain expressed in the language (2011). A formal language consists of a syntax, or rules for

constructing sentences; and semantics, or a specification for how the sentences relate to the domain (2011). One example application is Biological Expression Language (BEL), which represents scientific findings in a computable format by capturing causal and correlative relationships in the context of the system in which relationships were observed, the supporting publications cited, and the process of curation (Hourani, 2012). BEL is intended as a knowledge capture and interchange medium, with a use-neutral language facilitating knowledge assembly by applications (Hourani, 2012).

A second example is the use of web-based systems for the organization and retrieval of biological data (Baclawski & Niu, 2006). Development of the Semantic Web in 1999 enabled a framework in which computers are capable of acting as intelligent agents, analyzing the meaning or 'semantic content' of data on the web (Robinson & Bauer, 2011). A layer above the World Wide Web, which represents data as HTML pages referenced by hyperlinks, sites on the Semantic Web function as collections of semantically defined data which can be either read by other servers or used to generate Web pages for users (Robinson & Bauer, 2011). The Semantic Web provides reasoning and retrieval abilities, enabling automated processing, and annotates resources allowing the integration of different data sources (Baclawski & Niu, 2006). As an example, information available via the semantic web could be used for the prediction of protein function or for sequence similarity analysis (Baclawski & Niu, 2006).

Ontologies available via the Semantic web can be used for the functional analysis of microarray data, network modeling, and semantic similarity analysis and clinical diagnostics (Robinson & Bauer, 2011). Hardy envisions a "toxicology ontology roadmap" for the development of an integrated toxicology ontology (2012a). In order to support alternative *in vitro* and computational methods to traditional animal testing, and make available the largest possible

knowledge base of previous toxicology findings, development of such an ontology will support data management, model building, integrated analysis, validation and reporting to support applications. The roadmap includes the promise of universal access to high quality experimental data, improved storage, exchange and use of information from experiments, integrated use of physical, biological, and chemical techniques and data, integration of interdisciplinary and translational concepts, standardized ontologies, computational research and prediction, and stepwise testing strategies and integrated testing schemes (2012a). The uses of ontology in predictive toxicology are in standardizing and organizing data for use cases (Hardy, 2012b). According to Hardy, "ontology and controlled vocabulary may be used to define biological effects and entities, systems and their components and interactions, algorithms and models, pathways, and other useful conceptual entities for supporting complex reasoning about concepts, questions, and data" (2012b). In the context of predictive toxicology, the applications of ontology include integration of *in vivo* and *in vitro* data from diverse sources, Quantitative Structure Activity Relationship (QSAR), and regulatory purposes such as REACH or classification and labeling (CLP) (Hardy, 2012b). Toxicology ontology should represent animal traditional experimental data as evidence, and may represent study terms and experimental protocols (Hardy, 2012b).

Ontology makes information readily available for search and algorithmic processing. With the advent of the World Wide Web and increasing amounts of publicly available data generated by high-throughput methods, came the development of multiple biological ontologies; one of the first and best-known efforts is the Gene Ontology (GO), which contains "annotations" or links between data and structured vocabularies for human and model organism genomes (Robinson & Bauer, 2011). GO terms are manually curated and verified on the basis of published

experimental results (Robinson & Bauer, 2011).

An example use of the Gene Ontology is for the functional annotation of gene expression microarray data. Expression microarray experiments can be annotated with GO terms, and clustered expression patterns for GO terms can be viewed for correlation of biological process, molecular function, and cellular component to a particular gene or gene product (Ashburner et al., 2000). GO also can be used to examine sequence and functional conservation for orthologous genes through the transfer of annotations from experimentally tractable model organisms to less tractable organisms based on gene and protein sequence similarity (Ashburner et al., 2000).

The success of GO was based in large part on its adherence to a set of key principles. In an effort to answer the question of how to best represent biomedical data and information in a format commonly accessible to biologists, in 2001 Michael Ashburner and Suzanna Lewis developed the OBO Foundry, a central repository of life-science ontologies (Smith et al., 2007). OBO ontologies must be open, developed in a collaborative effort, use unambiguously defined common relations, and provide clear delineation of scope and mutual reuse (Smith et al., 2007). As of 2011, the library consisted of 99 ontologies covering the domains of biological process and molecular function, chemical entities, experimental investigations, anatomy, and disease (Hardy, 2012b).

Another use of ontology in predictive toxicology is the annotation of experimental data, such as use of ontologies and controlled vocabularies for the annotation of assay data from EPA's ToxCast effort (Phuong, 2014). Launched in 2007, ToxCast uses automated chemical screening technologies, or high throughput-screening assays, to screen cells for biological activity with potential toxic health effects (USEPA, 2013). Previous efforts have used the BioAssay Ontology (BAO) to annotate ToxCast assay endpoints for assays stored in PubChem,

enabling analysis of trends among annotated assays and observation of variances among screening data (Schurer et al., 2011; Vempati et al., 2014).

Use of ontology for information access, organization, and retrieval requires the use of logic and semantics, enabling automated reasoning. Ontology construction involves selection of the entities or objects to be included, definition of a controlled vocabulary, and recognition of relationships, interactions, and hierarchies (Bard, 2005). Ontology solves some of the difficulties in representing knowledge hierarchies using relational databases; by enabling recursive searching or searches within the data itself, and by representing information as a set of structured links with rules that apply up or down the hierarchy. Ontologies connect terms, usually represented as nodes, with a relationship or edge in an assertion. For example, an IS-A relationship would be represented as 'the cardiovascular system IS A MEMBER OF THE CLASS OF organ systems'. Relationships can be directed (one-way), or undirected (reciprocal). The set of relationships represented graphically is a hierarchical tree or graph. Directed relationships are often used to generate directed acyclic graphs (DAGs), in which nodes have more than one 'parent' node. These relationships allow inferences to be made either by a user or computationally (Bard, 2005).

Construction of a bioinformatics ontology involves design dependent on envisioning its purpose; identifying what it will be used for, its scope, its user community, the length of time it will be used and what domain it will cover (Baclawski & Niu, 2006). Construction begins with choice of relationships to be used, assembly of terms, and then building the ontology using a computer program with a compiler, beginning with the root term and adding subordinate terms with connecting relationships (Bard, 2005). Subsequently, a language that will work with the chosen compiler program should be chosen. The major ontology languages used today can be

classified as basic XML, XML topic maps, and Semantic Web languages including RDF and OWL depending on the intended use. While some languages are preferable for expressing information as logical statements, others present data in an organized structure either in the form of frames or graphs (Baclawski & Niu, 2006). XML is the most basic and widely supported language, while OWL supports the Semantic Web and offers the most flexibility but is the least supported by applications (Robinson & Bauer, 2011). Resource Description Framework (RDF) is the data model on which OWL and the Semantic Web are based (Robinson & Bauer, 2011). RDF describes resources in triples, providing a common framework for integration of web data (Robinson & Bauer, 2011). Uniform Resource Identifiers (URIs) link information about RDF resources to other resources, enabling web representation (Baclawski & Niu, 2006). RDF data can be represented in a graph-based mode, enabling data integration (Robinson & Bauer, 2011). Protein databases, such as UniProtKB, commonly use RDF/XML in which RDF data is represented in an XML file format (Robinson & Bauer, 2011). Web Ontology Language (OWL), extends the use of RDFs with more powerful capabilities for constructs and inference rules (Robinson & Bauer, 2011). While most OBO Foundry member ontologies are developed for the biomedical domain, the OWL (Web Ontology Language), was developed as a way to define structured ontologies for web-based interoperability in any domain (Robinson & Bauer, 2011). Conversion between one language and another represents a challenge requiring special computer programs for transformation (Baclawski & Niu, 2006).

According to Robinson and Bauer, the predominant language among bio-ontologies is the OBO format, developed by the Gene Ontology Consortium and used by the OBO Foundry as a standard language among ontologies (2011). OBO format ontologies are compact, easy to read and easy to parse (2011). The OBO format begins with headers containing general information,

then a list of stanzas; each begins with a keyword and contains key value lines containing specific information such as the GO id, name, and definitions; one file can contain 30,000 stanzas:

[Term]
id: GO:0000031
name: mannosylphosphate transferase activity
namespace: molecular_function
def: "Catalysis of the transfer of a mannosylphosphate group from one compound to another." [GOC:jl]
is_a: GO:0016740 ! transferase activity

Each tag-labeled pair contains a name, a colon, and a value. For the above example, the name *id* contains the value "GO:000031"; *id* is the unique identifier of the entity of a nucleotide sequence accessible through GenBank. Terms may also contain relations such as *is_a*; in the above example GO:0000031 *is_a* subclass of GO:0016740 ! transferase activity (2011).

Several free and open-source programs are available for building and editing ontologies, the most common of which is the program Protégé developed by Stanford University (Robinson & Bauer, 2011). Choice of an editor involves choosing which language the program supports. Other notable tools such as the EMBL-EBI Open Lookup Service (OLS) provide a web interface for browsing, searching, and accessing OBO file content (Mayer et al., 2014). Additionally publicly available browsers that aggregate terms from multiple ontologies and make their information available on the Web include Ontobee (http://www.ontobee.org/) and NCBO Bioportal (http://bioportal.bioontology.org/), while specific ontologies may have their own search tools, such as GO's AMIGO2 (Ashburner et al., 2000).

Review of Selected Ontologies

OBO Foundry Ontologies

Several of the most common OBO Foundry member ontologies were chosen for review

for their domain, content, and level of biological resolution (OBO Foundry, 2016a). Emphasis was placed on OBO Foundry ontologies for their uniformity and consistency of use. Several non-Foundry ontologies and controlled vocabularies were also reviewed.

OBO Foundry candidate ontologies are built on a common top-level ontology, the Basic Formal Ontology (BFO) and use a common set of relations defined in the Relation Ontology (RO) (Zheng et al., 2013). The BFO doesn't contain domain terms from the sciences, but provides concepts to support domain ontologies for information retrieval, analysis, and integration for scientific research (IFOMIS, 2015). BFO provides a hierarchy of upper-level abstract classes, from which classes in domain-specific ontologies can inherit, linking together independently developed ontologies within the OBO Foundry (Mungall et al., 2011). The BFO divides entities into objects and processes, called "continuants" and "occurrents" respectively (Robinson & Bauer, 2011). Continuants represent entities that continue to exist over time and preserve their identity despite change; for example, an organism or cell; while occurrents happen, unfold or develop through time, for example, the process of meiosis (Robinson & Bauer, 2011).

The Relation Ontology (RO) provides a set of relations for standardization across OBO Foundry ontologies and ontologies in the wider OBO library (OBO Foundry, 2016e). Several other upper-level or application ontologies are the Experimental Factor Ontology (EFO), Brenda Tissue Ontology (BTO) and Eagle-I Resource Ontology (ERO) concerned respectively with the domains of experiments, anatomy, and research resources (OBO Foundry, 2016a). The BTO (http://www.BTO.brenda-enzymes.org) provides a standardized representation of all tissue terms from taxonomic groups covering animals, plants, fungi and prokaryotes connecting to the BRENDA enzyme database (Gremse et al., 2010). The BTO contains more than 4600 different

anatomical structures, tissues, cell types and cell lines classified according to the format of the Gene Ontology Consortium and organized as a directed acyclic graph (DAG) (Gremse et al., 2010). The EFO provides a systematic description of experimental variables in the European Bioinformatics Institute (EBI) databases; facilitating consistent annotations and pulling together classes from multiple reference ontologies including disease, cell line, cell type and anatomy (NCBO Bioportal, 2016b).

Gene Ontology

The purpose of the Gene Ontology (GO) project is to produce a dynamic, controlled vocabulary that can be applied to all eukaryotes in a species-independent manner (Ashburner et al., 2000). GO consists of ontologies or defined terms and structural relationships, and annotations or associations between gene products and terms (2000). GO genes and gene products are assigned to three independent ontologies: molecular function, biological process, and cellular component or location (Ashburner et al., 2000). Within the three ontologies, GO terms are related by formally defined *is_a* and *part_of* relationships (Baclawski & Niu, 2006).

According to Ashburner et al., the complete sequencing of several model organism systems including the budding yeast, nematode, and fruit fly, have accelerated the availability of genomic information (2000). Although molecular sequences are more readily available for more species, the way in which biologists conceptualize shared biological elements hasn't kept pace with sequencing. Current systems of nomenclature for genes and their products remain divergent, even when there are underlying similarities. Functional conservation of genes and gene products requires a common language for comparison between species. A comparison between the complete genomes of the budding yeast and worm revealed a large fraction of genes displaying orthology; with approximately 12% (~18,000) of worm genes encoding proteins

whose biological roles could be inferred from their similarity to orthologs in yeast, comprising approximately 27% (~5,700) of yeast genes.

Annotations, or descriptions supported by evidence, associate a specific gene with a term in the ontology, and are either manually curated or generated through predictive methods (Robinson & Bauer, 2011). A 'GO annotation' describes the association between an ontological class and a gene product, including references to evidence in literature supporting the association (Ashburner et al., 2000). Evidence-supported annotations may describe the biological roles of individual genomic products such as genes, proteins, RNA, and complexes) by tying them to ontologies (Ashburner et al., 2000). Annotations are useful for functional prediction based on patterns of annotation, in that if annotations for two attributes occur together in a database, a gene holding one attribute is likely to hold the other attribute too (King et al. 2003).

According to Ashburner et al., GO provides information on genes for over 460,000 eukaryotic species (2000). The information in Gene Ontology derives from three model organism databases: FlyBase, MouseGenome Informatics (MGI), and the *Saccharomyces* Genome Database (SGD). Each term or 'node' in GO is linked to other information, such as gene and protein databases. The utility of GO is the ability to organize, describe, query, and visualize biological knowledge despite changes and updates.

According to Ashburner et al., the biological process ontology describes "the biological objective to which a gene or gene product contributes" (2000). A biological process consists of an ordered assembly of molecular functions, often involving a chemical or physical transformation, for example, 'cell growth and maintenance' or 'translation'. Molecular function ontology describes the biochemical activity of a gene product, including specific binding to ligands or structures; example functional terms include 'enzyme' or 'adenylate cyclase'.

Cellular component ontology describes the location in the cell where a gene product is active; such as 'ribosome' or 'proteasome.' Biological process, molecular function, and cellular component comprise attributes of genes and gene products, reflecting a one-to-many relationship such that one gene or protein may play a role in multiple pathways, processes, or interact with multiple cell components and locations (Ashburner et al., 2000).

An example of a GO annotation is the term 'DNA metabolism'. Metabolism is a biological process carried out by mostly shared elements in eukaryotes (Ashburner et al., 2000). The biological process ontology pertaining to DNA metabolism contains nodes or terms that can have multiple parents, for instance, the term 'DNA ligation' has parent terms of 'DNA-dependent DNA replication', 'DNA repair' and 'DNA recombination' (Ashburner et al., 2000).

According to the Gene Ontology Consortium, GO ontologies and annotations are publicly available through http://www.geneontology.org (2014). Through AMIGO2, an open-source set of tools for querying and browsing GO data, users can perform GO enrichment analysis, search for GO terms, annotations, and metadata across species. The site also includes documentation for generating ontology terms and use of logical definitions. GO is available for download in the OBO flat file format, RDF-XML format, and GO annotation file formats (Gene Ontology Consortium, 2015a). Additionally, many programs have been developed for profiling gene function based on the GO, such as the DAG-Edit, GenMAPP, GoMiner, and DAVID tools (Baclawski & Niu, 2006).

Chemical Entities of Biological Interest (ChEBI)

According to Hastings et al., ChEBI (Chemical Entities of Biological Interest) accessible at http://www.ebi.ac.uk/chebi is a database and ontology of low molecular weight chemical entities of biological interest (2013). Molecules encoded by the genome; nucleic acids, proteins

and peptides derived from protein, are not included. Annotations emphasize immunology, natural products and metabolites in many species. High quality chemical reference data are needed for computer modeling in fields of predictive toxicology, metabolic modeling and the search for disease biomarkers, and ontologies such as ChEBI make the semantics of such information available and computable.

According to EMBL-EBI, ChEBI specifies relationships between molecular entities or classes of entities, and their parents or children (2016). ChEBI uses terminology from the International Union of Pure and Applied Chemistry (IUPAC) and nomenclature from the International Union of Biochemistry and Molecular Biology. Data are drawn from four sources: The Integrated relational Enzyme database of the EBI (IntEnz), Kyoto Encyclopedia of Genes and Genomes (KEGG) Compound, PDBeChem database, and ChEMBL database. A ChEBI page contains the following data fields: ChEBI Identifier, ChEBI Name, ChEBI ASCII Name, star rating, structure, formula, charge, average mass, ChEBI Ontology, IUPAC name, International Nonproprietary Name (INN), synonyms, brand name, database links, registry number such as CAS, citations, and cross-references to biological and chemical databases.

Database entries are manually annotated, with links to external databases such as Rhea and Reactome which use ChEBI identifiers to refer to chemicals in biological context. Where possible, a molecular graph is provided accompanied by chemical structural graphs in InChI, InChIKey, and SMILES (Robinson & Bauer, 2011). ChEBI is also cross-referenced to PubChem (Hastings et al., 2013). The database is aligned with other chemical involving processes in ontologies, such as the Open Biomedical Ontologies (OBO) Foundry and Gene Ontology (Hastings et al., 2013). Of special interest is annotation of compounds deemed ‘natural products’, or secondary metabolites, relevant in drug discovery and metabolism research (Hastings et al.,

2013). Data have been drawn from primary literature sources; publications identifying a particular metabolite in a given species (Hastings et al., 2013).

Using the ChEBI website, data is open and freely available for download (Robinson &Bauer, 2011). Users may search by substructure or similarity of compounds (Hastings et al., 2013). The ChEBI ontology may be displayed in 'tree view' in an interactive graph-based visualization beginning with the chemical entity up to the root of the ontology (Hastings et al., 2013). Roles and structural relationships are displayed separately from the graph-based browser (Hastings et al., 2013). The graph visualization is based on the JavaScript InfoVis toolkit (http://thejit.org/) and mapped to other ontologies available as an OWL file (Hastings et al., 2013).

An example of a natural product in ChEBI would be caffeine (Robinson & Bauer, 2011):

[Term]
id: CHEBI: 27732
name: caffeine
synonym: "C8H10N4O2" RELATED FORMULA [KEGG COMPOUND:]
synonym: "Cn1cnc2nc(C)c(=O)n(C)c(=O)c12" RELATED SMILES [ChEBI:]
is_a: CHEBI:27134 ! trimethylxanthine
relationship: has_role CHEBI:25435 ! mutagen
relationship: has_role CHEBI:35337 ! central nervous system stimulant
relationship: has_role CHEBI:38809 ! ryanodine receptor modulator

Caffeine is one of the superclass, trimethylxanthine. Synonyms describe the molecular formula and SMILES representation of caffeine.

Ontology for Biomedical Investigations (OBI)

According to Brinkman et al., Ontology for Biomedical Investigations (OBI), available at (http://purl.obolibrary.org/obo/obi/2009-11-02/obi.owl) addresses the need for a standardized terminology for descriptions of biological and clinical investigations (2010). It is developed collaboratively by representatives of 19 global biomedical communities and is an OBO- Foundry

member ontology. Biomedical experimental descriptions are usually stored as free text without a standardized terminology, creating challenges in comparison, reproduction, and analysis. Application of OBI to biomedical investigations facilitates interpretation of the experimental process, as well as increasing computational processing and integration with the Semantic Web. OBI represents all phases of experimental processes and the entities involved in preparing for, executing, and interpreting those processes. OBI terms span biomedical and technological domains, representing all phases of experimental processes, and the entities involved in those processes, such as study designs, protocols, instrumentation, and analyses regardless of the field of study. It also represents roles and functions used in biomedical investigations.

OBI is written in OWL and an example OBI class describes culturing cells that do not express the CD8 cell surface receptor using the 'owl:complementOf:' keyword to restrict the properties of the class representing cells that are the complement of cells that express the CD8 receptor (Robinson & Bauer, 2011):

```
<owl:intersectionOf rdf:parseType="Collection">
  <owl:Class rdf:about="CL:cell"/>
  <owl:Class>
    <owl:complementOf>
      <owl:Restriction>
        <owl:someValuesFrom>
          <owl:Class rdf:about="#CD8 receptor"/>
        </owl:someValuesFrom>
        <owl:onProperty>
          <owl:TransitiveProperty rdf:about="#has_part"/>
        </owl:onProperty>
      </owl:Restriction>
    </owl:complementOf>
  </owl:Class>
</owl:intersectionOf>
```

Brinkman et al. describes a vaccine protection investigation use case, demonstrating how

entities and relations involved in experimental processes can be modeled in a biomedical domain using OBI (2010). A vaccine protection investigation describes an experiment that examines how efficiently a vaccine induces protection against a virulent pathogen. Upper-level classes in OBI describe experimental processes, material entities, roles, and inputs and outputs regarding the investigation. For example, the class 'administering substance *in vivo'* describes experimental processes. Subtypes of this parent class include 'vaccination' and 'pathogen challenge', describing the type of experiment being conducted. The definitions of classes in this use case help validate OBI's design and granularity extending from the molecular to higher level biomedical investigations.

Protein Ontology (PRO)

Increasingly, journals require that the data underlying a proteomics study should be made public through journal websites or publicly available repositories for MS (mass spectrometry)-based proteomics data (Mayer et al., 2014). The use of a standardized format in proteomics eases the comparison of data from different sources and reproducibility of results (Mayer et al., 2014). Ontologies in proteomics support flexible definition of semantics of the represented data, making the data computer-readable and accessible for analysis and data mining using software tools (Mayer et al., 2014). The most important ontologies in the proteomics domain are used by the XML-based proteomics standards defined by the HUPO PSI working groups (Mayer et al., 2014). The Protein Ontology (PRO; http://proconsortium.org), formally defines protein entities, complexes, and interrelations (Natale et al., 2014). Protein entities corresponding to single amino acid chains are categorized into family, gene, sequence and modification metaclasses (Natale et al., 2014). Entities may be defined as either organism-agnostic or organism-specific metaclasses (Natale et al., 2014). Members of orthologous genes are recognized on the basis of similarity in

protein sequences, allowing the grouping of genes that are evolutionarily conserved (Robinson & Bauer, 2011). According to Natale et al., some of the uses of PRO include "to define dendritic and hematopoietic cell types, describing biological processes, flagging protein entities mentioned in literature, and capturing information isolated from literature in text mining workflows" (2013).

PRO is an OBO Foundry ontology, using the standardized definitions and relationships from BFO and RO, and it is interoperable with other resources such as UniProtKB and GO. As of 2013, PRO mappings covered 12 GO reference organisms (Natale et al., 2014). PRO is available for download in the OBO file format (Natale et al., 2014).

Sequence Ontology (SO)

Sequence Ontology, located at https://github.com/The-Sequence-Ontology/SO-Ontologies defines sequence features used in biological sequence annotation. Annotating a genome provides information about the genomic sequence and the sequence of molecules derived from the genome to coordinates on biological sequence (Mungall et al., 2011). The Sequence Ontology provides "a structured controlled vocabulary for sequence annotation, for the exchange of annotation data and the description of sequence objects in databases" (The OBO Foundry, 2016).

SO describes the features of biological sequence for both genomic and derived sequences (Mungall et al., 2011). Genome annotation locates genomic sequence information to a linear representation of a chromosome, using coordinates to capture the sequence of information (Mungall et al., 2011). Examples of SO classes include biological features, such as 'binding_site' and 'exon' (Sequence Ontology, 2009). Biomaterial features are intended for experimental use, such as 'aptamer' and 'PCR_product' (Sequence Ontology, 2009). Experimental features describe the results of an experiment (Sequence Ontology, 2009).

SO describes how sequences change over the course of genomic and post-genomic processes (Mungall et al., 2011). SO increases the interoperability of model organism databases, providing a standardized set of terms and relationships with which to describe genomic annotations and enable automated reasoning (Robinson & Bauer, 2011). One use of the SO to the biological community is in providing a structured vocabulary for describing nucleic acid sequence; for example, the BioDAS server for sharing biological data the description of primary annotations of nucleic acid sequence; for example, the annotations shared by a DAS server (Sequence Ontology, 2009).

The SO is an OBO Foundry ontology and its classes have been aligned with the BFO to allow for computable definitions and interoperability with other Foundry ontologies (Robinson & Bauer, 2011). Contributors to SO include the GMOD community, model organism databases such as WormBase, FlyBase, and Mouse Genome Informatics, and institutes such as the Sanger Institute and EBI (NCBO, 2016c). It currently contains 2301 classes. SO is available in the OWL format (NCBO, 2016c).

Cell Ontology (CL)

The cell ontology, located at http://cellontology.org, was developed as a controlled, structured vocabulary for cell type that is non-organism specific, covering cell types from the prokaryotic, fungal, and animal kingdoms (Robinson & Bauer, 2011). Model organisms included in CL are: human, mouse, *Drosophila*, *Caenorhabditis*, zebrafish, *Dictyostelium discoideum*, *Arabidopsis*, fungi, and prokaryotes (Robinson & Bauer, 2011). CL is linked to anatomical structures in Uberon, biological processes in GO, and other entities in ChEBI, PR, and PATO; in turn multiple ontologies provide logical definitions for CL terms (The OBO Foundry, 2016a). Uses of the CL include providing a common reference framework for annotating experimental

metadata from functional genomics studies such as the ENCODE project (The OBO Foundry, 2016a) and gene expression databases such as Array Express (Robinson & Bauer, 2011). CL is also used in Library of Integrated Network-based Cellular Signatures (LINCS), a project to generate multidimensional data sets including biochemical, genome-wide transcriptional, and phenotypic cellular response signatures to a variety of small molecule and genetic perturbations to create a widely applicable systems biology resource (Schurer, 2011).

CL is an OBO Foundry member ontology with files provided in OWL and OBO formats (The OBO Foundry, 2016a). An example class in CL would be the relationship of 'B cell' with its subsuming parent term 'lymphocyte of B lineage' through *is_a* relationships; while a 'pro-B cell' is related to the parent class 'B cell' through a *develops_from* relationship (Robinson & Bauer, 2011, p.161).

Foundational Model of Anatomy (FMA)

Foundational Model of Anatomy is located at http://sig.biostr.washington.edu/projects/fm/. According to the Structural Informatics Group at the University of Washington, FMA provides a reference ontology representative of the "types and relationships necessary for the symbolic representation of the phenotypic structure of the human body in a form that is understandable to humans and is also navigable, parseable, and interpretable by machine-based systems" (2007). The lack of a generalizable, computable representation of anatomy led to ontology developers creating divergent controlled vocabularies and ontologies from their own viewpoints (Rosse & Mejino, 2007). FMA differs from other anatomical domain ontologies and vocabularies in that it provides anatomical information designed to serve the needs of any user group, and to be reused by any application (Rosse & Mejino, 2007). Additionally, FMA encodes anatomical knowledge in a way that supports

computational analysis and the development of anatomy applications targeting specific user groups. These user groups may include knowledge modelers and developers of applications for education, clinical medicine, electronic health records, biomedical research, and health care delivery and management (Rosse & Mejino, 2007).

One of the largest computer-based knowledge sources in the biomedical sciences, FMA comprises 75,000 classes and over 120,000 terms, as well as over 2.1 million relationship instances from 168 relationship types (University of Washington Structural Informatics Group, 2007). The anatomy taxonomy (At) comprises the dominant class, containing types of biological macromolecules, cells and their parts, tissue portions, organs and organ parts, organ systems, and body regions (University of Washington Structural Informatics Group, 2007). Other components include the Anatomical Structural Abstraction (ASA), the Anatomical Transformation Abstraction (ATA), and the Metaknowledge (Mk) components (University of Washington Structural Informatics Group, 2007). Anatomical structures are defined in terms of a *genus* or 'material anatomical entity' and subsequent assertions are *differentiae* (University of Washington Structural Informatics Group, 2007). An example annotation in FMA is the definition of the organ, "Heart" as "Organ with cavitated organ parts, which is continuous with the systemic and pulmonary arterial and venous trees" (Robinson & Bauer, 2011).

FMA is an OBO Foundry member ontology. It is a frame-based ontology, where each frame stores all the information about a named type viewable with the Protégé ontology editor (Robinson & Bauer, 2011). It has also downloadable files available in the OWL and OBO formats (Robinson & Bauer, 2011).

Uberon

Located at http://uberon.org, the Uber Anatomy Ontology (Uberon) facilitates the unification of multiple anatomy ontologies into a multi-species ontology for anatomics (Haendel et al., 2014).Anatomics describes the analysis, computer formulation and use of the *anatome*, or the complete set of tissues and organs in a model organism. There are about 15 anatomical ontologies currently available for a range of model organisms, including the mouse, *Drosophila, C. elegans,* and zebrafish (Bard, 2005). Databases for human adult and developmental anatomies are available but there is as yet no central resource for human data (Bard, 2005). Phenomics, or phenotype-based analyses, play an important role in understanding gene function and the genetic and epigenetic contributions to developmental, behavioral, evolutionary and morphological variation (Haendel et al., 2014). Model organism databases (MODs) aggregate genetics and genomics information from specific model organisms (Haendel et al., 2014). Single-species model organism anatomy ontologies have been developed to annotate and query these gene expression and phenotype data for specific organisms and multi-species anatomy ontologies have been developed for multiple taxons (Haendel et al., 2014). Such single-species ontologies include the adult mouse anatomy ontology (MA), adult human ontology (FMA), zebrafish ontology (ZFA) and *Xenopus* anatomy ontology (XAO) (Haendel et al., 2014). However, each of these model organism anatomies is categorized into an individual ontology with no connection between ontologies, making querying and searching challenging (Mungall et al., 2012). Uber Anatomy Ontology (Uberon) is a metazoan, multi-species anatomy ontology that aims to integrate multiple taxon specific vertebrate ontologies (Haendel et al., 2014).

According to Bard, anatomics is challenging to format in a computer-comprehensible manner because of the complexity of relationships between tissues in an organism (2005).

Precision concerning the actual number of tissues and organs in an organism presents a challenge. Specifying anatomical relationships such as *part of*, *develops from*, and *is-a* requires being explicit as to the knowledge that relationships contain, and preventing errors of inference.

As of 2014, Uberon contained 12241 classes representing high level categories from the Common Anatomy Reference Ontology (CARO) (Haendel et al., 2014). Classes include anatomical systems such as 'nervous system', organs such as 'heart', tissues such as 'adipose tissue' and appendages such as 'pelvic girdle' (Mungall et al., 2012). Ontology entities are referred to by classes and relations, referred to as axioms (Mungall et al., 2012). Relationships were implemented by applying a common set of design and modeling patterns in Uberon for how anatomy parts relate, such as the representation of the 'quadrate-articular joint' as *connected_to* the 'quadrate' and 'articular' classes (Haendel et al., 2014).

Uberon files are available in Open Biomedical Ontologies (OBO) format and OWL. Annotations from source ontologies were manually curated (Haendel et al., 2014). An example would be the Uberon class 'lung' and its relation to ontologies outside of Uberon. 'Lung' is present in Uberon as a superclass including lung classes from mouse and human anatomy ontologies (Mungall et al., 2012). The Uberon term 'alveolus of lung' *is_a* 'alveolus' *part_of* 'lung' (Mungall et al., 2012).

Mammalian Phenotype (MP)

The Mammalian Phenotype (MP) ontology, located at http://www.informatics.jax.org/searches/MP_form.shtml is used in the annotation of phenotype data from high-throughput screens (Smith & Eppig, 2015). Information from these screens can be applied to human disease by relating data from mouse phenotype to human phenotype (Robinson & Bauer, 2011). The International Mouse Phenotyping Consortium (IMPC) is using

MP in the annotation of data from large scale high-throughput mouse knockout phenotyping projects (Smith & Eppig, 2015). Many screens look for deviants in specific phenotype areas and map these phenodeviants to molecular mutations (Smith & Eppig, 2015). The phenotype data can be used to infer information about gene function, expression and biological pathways (Smith & Eppig, 2015).

According to Smith & Eppig, MP is manually curated with data from large-scale phenotype datasets and from published literature (2015). MP files are in OWL format and converted OBO format and edited using Protégé-4.3 software.

MP has been expanded to be used as the common data standard for annotation of large scale mouse phenotype data sets and data import to MGI without curation (Smith & Eppig, 2015). MP terms are assigned to IMPC parameters by anatomical system, 'Cardiovascular system' and ontology terms captured in IMPReSS, a database and web portal used to track phenotyping procedures used by IMPC. For example, test results for Heart Weight [IMPC_HWT_001] are captured by MP terms 'abnormal heart weight' [MP:0004857], 'increased heart weight' [MP:0002833] and 'decreased heart weight' [MP:0002834]. Relationships such as *is-a* or *part-of* are denoted; for example, 'abnormal cardiovascular system morphology' is-a 'cardiovascular system phenotype' (Mouse Genome Database, 2016). IMPC provides a RESTful, or web-browser accessible, interface to mouse alleles, experimental results and genotype-phenotype associations (Smith and Eppig, 2015). According to Smith & Eppig, "MGI will remain the source of global mouse phenotype data integration from large and small scale data sets, contributions and literature" (2015). Since most human diseases are caused by other functional mutations, information imported from MGI contains information on other type of mutations than those caused by knockouts (Smith and Eppig, 2015). MGI data also contain

data on knockout alleles exhibiting a prenatal or perinatal lethal phenotype; processes driving prenatal growth and differentiation that may help identify origins of developmental disease and congenital defects (Smith and Eppig, 2015).

Human Phenotype Ontology

Human Phenotype Ontology (HPO) is publicly available at http://human-phenotype-ontology.github.io/. Created in 2007 (Köhler et al., 2014), its purpose is "to provide a standardized vocabulary of phenotypic abnormalities encountered in human disease" (Human Phenotype Ontology, 2016a). The human phenome describes phenotypic features found in disease; abnormalities describing the signs, symptoms, and manifestations of disease (Robinson & Bauer, 2011). In the past, databases have offered an incomplete representation of the human phenome; by either focusing on genetic mutations (Robinson & Bauer, 2011), or failing to agree on common semantics for phenotype data (Köhler et al., 2014). HPO addresses these issues by providing a set of terms or annotations and relations derived from the medical literature and external databases (Human Phenotype Ontology, 2016a).

Providing computable definitions for phenotypic abnormalities can prove useful for differential diagnosis in the clinical setting. Identifying the same disease phenotype in multiple affected patients, can aid in discovering genetically-linked diseases (Köhler et al., 2014). In this way, queries can search the relationships between ontology terms to return diseases related to the term (Robinson & Bauer, 2011, p. 251).

The current content of HPO is 11,000 terms and 115,000 annotations for hereditary diseases, and about 4000 annotations for common diseases (Human Phenotype Ontology, 2016a). Individual terms describing phenotypic anomalies are connected by is-a relationships in a Directed Acyclic Graph (DAG) format, in which one term can have multiple parents (Human

Phenotype Ontology, 2016a). Terms have a unique ID, label, and textual definition (Human Phenotype Ontology, 2016b).

For example, the entry for "Epibulbar dermoids" has the unique ID [HP:0001140] and text definition, "An epibulbar dermoid is a benign tumor typically found at the junction of the cornea and sclera(limbal epibullar dermoid)" (Human Phenotype Ontology, 2016b).Terms are respectively mapped to four ontologies: phenotypic abnormality, clinical modifier, mortality/aging, and mode of inheritance (Human Phenotype Ontology, 2016b). HPO classes are interoperable with other model organism databases. Human diseases are cross-listed to terms in OMIM, Orphanet, and DECIPHER (Human Phenotype Ontology, 2016b). HPO is cross-referenced with external ontologies and vocabularies including MESH and International Classfication of Diseases (ICD) (Köhler et al., 2014).

HPO is available in OBO and OWL-formats as generated by the Oort, the OBO ontology release tool (Köhler et al., 2014). HPO maintains an integration system for managing releases (Köhler et al., 2014). HPO is linked to external databases and model organism databases (Köhler et al., 2014); providing annotations to clinical descriptions in Online Mendelian Inheritance in Man (OMIM), Orphanet, and DECIPHER (Human Phenotype Ontology, 2016b). The Phenomizer, https://compbio.charite.de/phenomizer/, uses HPO for clinical diagnostics (Human Phenotype Ontology, 2016a). HPO also has a web browser located at http://compbio.charite.de/hpoweb/showterm?id=HP:0001197. (Human Phenotype Ontology, 2016a).

An example annotation defines the HPO term "Hypoglycemia". The disease class is linked to terms from PATO and FMA (Köhler et al., 2014):

Class: Hypoglycemia
EquivalentTo:
'decreased concentration'
and towards some 'glucose'
and inheres_in some 'portion of blood'
and qualifier some 'abnormal'

PATO

The Phenotype and Trait Ontology (https://github.com/pato-ontology/pato/) is an ontology of descriptive terms for use in phenotype annotation (OBO Foundry, 2016d). PATO links "phenotypic abnormalities in humans and model organisms to ontologies for anatomy, biochemistry, cell type and components, pathology, as well as molecular functions and processes in a way that would enable an integrative computational analysis of phenotypic abnormalities and disease" (Robinson & Bauer, 2011). One of the uses of PATO is in the gene driven approach to model organism discovery; determining the relationships between mouse phenotype and human disease and looking for common elements across both species to understand the disease manifestation in humans (Robinson & Bauer, 2011).

According to Robinson & Bauer, PATO is constructed as a single hierarchy of phenotypic qualities with 22090 terms (2011). It's used with other ontologies of 'quality-bearing entities' including FMA, GO, and CL. A PATO annotation consists of a combination of PATO terms with terms from these ontologies. For instance, a PATO annotation usually consists of a combination of a quality-bearing entity and a quality. The PATO annotation to describe a "red-eye" phenotype in Drosophila combines the PATO quality term *red* with the existing Drosophila gross anatomy entity term *eye* from Mouse Anatomy ontology (2011):

E=FBbt:eye Q= PATO:red

PATO is written independently of a database schema or format; annotations are in pheno-

syntax or pheno-xml (OBO Foundry, 2016d). PATO is an OBO Foundry ontology with files available in OWL and OBO formats (OBO Foundry, 2016d).

PCO

According to PCO, Population and Community Ontology (PCO) (https://github.com/PopulationAndCommunityOntology/pco), "describes material entities, qualities, and processes related to collections of interacting organisms such as populations and communities" (2016). It's applicable in the fields of behavioral studies and ecology. An OBO Foundry ontology, it communicates with the Basic Formal Ontology (BFO) and is compatible with GO, PATO, RO, and NCBITaxon. PCO classes describe evolutionary processes, organismal interactions, and ecology experiments. It is available in an OWL format file.

NCBI Taxon

According to Federhen, the purpose of the NCBI Taxonomy, located at http://www.ncbi.nlm.nih.gov/taxonomy is indexing the domain of sequences for all of the organisms in the publicly available databases of the International Nucleotide Sequence Database Collaboration (INSDC) (2012). The NCBI Taxonomy Project began in 1991 with the design of the Entrez information retrieval system, the first database to link nucleotide sequences and protein sequences together with relevant literature in a single resource. The partnership known as International Nucleotide Sequence Database Collaboration (INSDC) contained the nucleotide sequence databases GenBank, EMBL and DDBJ. NCBI Taxonomy evolved out of an agreement among INSDC members to resolve issues of taxonomic nomenclature and classification before the release of new sequence data.

The current NCBI Taxonomy provides a set of names and classifications in the structure of a phylogenetic taxonomy, representing the evolutionary relationships in the evolutionary life

history tree (Federhen, 2002). The taxonomy includes monophyletic groups only, in which group members are closely related (Federhen, 2012). An NCBI Taxonomy entry includes the primary or "scientific name" and associated secondary names (Federhen, 2012). As of 2011, the taxonomy database contained 234,991 species with formal names, 405,546 with informal names (Federhen, 2012). The database contains 111110 prokaryotic species, 221,263 eukaryotic species, and 95 extinct species with formal scientific names (Federhen, 2012). Within the taxonomy database, three nomenclature codes describe animals (ICZN); plants, algae and fungi (ICN); and prokaryotes (ICNB). An additional fourth code describes viruses, and each code has a set of rules for publication of new taxonomic names (Federhen, 2012). The taxonomy database serves as an entry point to Entrez, a system to link nucleotide and protein sequences with abstracts from published literature (Federhen, 2012).

NCBI Taxonomy database is stored in TAXON, a SyBase relational database (Federhen, 2002). Database entries are "taxons" represented by nodes (Handbook). Taxons are represented by unique "taxids", or identifiers (Federhen, 2002). Taxids are stable and persistent; they can be deleted and merged but not reused (Federhen, 2012). The path from the root node to a taxon within the tree is a taxon's "lineage," while the "subtree" represents the collection of nodes below a taxon (Federhen, 2002). The database is publicly available through three resources: Taxonomy Browser http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi, Taxonomy domain of Entrez http://www.ncbi.nlm.nih.gov/taxonomy and the Taxonomy ftp site ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/ (Federhen, 2012). The TaxBrowser "provides a hierarchical view of the classification from any particular place in the taxonomy" (Federhen, 2002). The Browser provides information about publicly-available sequence entries; excluding those entries that have not yet been released (Federhen, 2002).

For example, the taxonomy browser taxon-specific display page for the entry 'Mammalia' contains the fields: Name, Taxonomy ID, GenBank common name, Inherited blast name, Rank, Genetic code, other names, full lineage; links to other Entrez records; and a Comments and References section containing links out to other resources (Federhen, 2012). An example use of NCBI Taxon would be using the NCBI Taxonomy browser to view the hierarchical display for the family 'Hominidae', consisting of four genera: 'Gorilla', 'Homo', 'Pan', and 'Pongo' with six species-level names (Gorilla gorilla, Homo sapiens, Pan paniscus, Pan troglodytes, Pongo pygmaeus, and Pongo sp.) and 2 subspecies (Federhen, 2002). The browser shows a taxon-specific display for 'Hominidae' with the ability to view lineage above and below 'Hominidae' as well as links to information in related databases (Federhen, 2002).

The ontology representation of NCBI taxon is the NCBI taxonomy database translated into OBO/OWL format, available at https://github.com/obophenotype/ncbitaxon (The OBO Foundry, 2016c). Each taxon is taken as an OBO/OWL class with individual organisms forming instances of the class (The OBO Foundry, 2016c).

Controlled Vocabularies

These represent vocabularies and ontologies that are not currently members of the OBO Foundry.

MESH/UMLS

Medical Subject Headings (MeSH), located at https://www.nlm.nih.gov/mesh, is one of the oldest controlled vocabularies for the domain of medical informatics, providing the controlled vocabulary thesaurus for the National Library of Medicine (NLM) (USNLM, 2015). Common terminology is important in clinical practice, and a problem facing medical informatics is the consistent representation of patients, treatments, and outcomes (Nelson, 2009). In 1986, the

NLM began a project to build the Unified Medical Language System (UMLS), a repository integrating biomedical vocabularies from various sources, in order to integrate medical literature in a computer-accessible format (USNLM, 2015). The UMLS integrates names, concepts, and relations from more than 100 biomedical vocabularies, including NCBI, GO, and MeSH (Baclawski & Niu, 2006). The MESH browser consists of a set of terms that assist in locating descriptors in a hierarchical structure (Baclawski & Niu, 2006). Today, MESH contains 13 levels, 27,883 descriptors, with more than 87,000 entry terms, 232,000 supplementary concept records (SCRs), indexing articles from 5400 biomedical journals (USNLM, 2015). Descriptors help locate MESH headings, which are assigned to a citation from medical literature including Medline and PubMed (Nelson, 2009). Examples of broad headings in MeSH are "Anatomy" and "Mental Disorders," while more specific headings at narrow levels of hierarchy, are "Ankle" and "Conduct Disorder" (USNLM, 2015). "Vitamin C" is an entry term to the heading "Ascorbic Acid" (USNLM, 2015). SCRs are specific chemicals, diseases, and drug protocols assigned to descriptors connected to citations to enable user queries tied to the descriptor (USNLM, 2015). The MeSH thesaurus is small in relation to the entirety of UMLS terms, which has browsers independent of MeSH (Baclawski & Niu, 2006).

One use of MeSH is for formulating queries for keyword search or for generating knowledge representations (Baclawski & Niu, 2006). MeSH is organized according to a user model, based on bibliographic retrieval, and an information model in which headings are assigned to citations (Nelson, 2009). Representational integrity means that a single term can be represented in only one way (Nelson, 2009). According to Nelson, "MeSH is successful due to the careful examination of the scope of its use, demanding of itself a clear statement of its mission, regular examination of feedback from its user group, and conscious conceptual

evolution in response to the use environment" (2009). Additionally, indexing procedures ensure that MeSH terminology is kept up-to-date (Nelson, 2009). MeSH is available in XML format from the site https://www.nlm.nih.gov/mesh/filelist.html and in Research Description Framework (RDF) format at https://id.nlm.nih.gov/mesh/ (USNLM, 2015).

OpenTox

The OpenTox Framework, located at http://www.opentox.org, aims to promote data integration and machine accessibility to toxicity data by standardizing chemical and toxicity databases, improving interoperability between toxicity resources, and providing a representation suitable for modeling algorithms (Tcheremenskaia et al., 2012). The project was funded by the EU Seventh Framework Program for predictive toxicology modeling and application development (Tcheremenskaia et al., 2012). OpenTox provides toxicity information regarding endpoints considered by Registration Evaluation and Authorization of Chemicals (REACH) legislation, outlining alternative methods to animal models for chemical testing (Hardy et al., 2010).

Data organization and retrieval consists of selection of toxicological endpoints to be included, and definition of the type and extent of information for each endpoint and their relationships and hierarchies (Hardy et al., 2012b). OpenTox provides a set of ontologies available through REST web services that support tools and APIs (Tcheremenskaia et al., 2012). Within OpenTox, the included ontologies are a)*Toxicological ontology* –toxicological endpoints; b) *Organs system and Effects ontology* – addressing organs, targets/examinations in *in vivo* studies; c) *ToxML ontology* –conversion of the ToxML schema to ontology format; d) *ToxLink*– ToxCast assays ontology e) *OpenTox ontology*– representation of OpenTox framework components e) Algorithms ontology and f) *OpenToxipedia*- collaborative resource for entry and

editing of toxicology terms (Hardy et al., 2012b). The Toxicological ontology is the main ontology, available in OWL format, containing the study types which can be mapped to toxicity datasets: carcinogenicity, *in vitro* bacterial mutagenesis, *in vivo* micronucleus, repeated dose toxicity, and aquatic toxicity (Hardy et al., 2012b). The organs system ontology provides non-endpoint specific diagnostic features for 12 organ systems for rodent *in vivo* studies (Hardy, 2012b). OpenTox ontology provides an OWL representation for the core components- the datasets, features, tasks, algorithms, models, and validation- of OpenTox (Tcheremenskaia et al, 2012). OpenTox is constructed using OBO Foundry principles and using the OWL DL language edited by Protégé OWL editor (Hardy et al., 2012b).

OpenTox components are made available through standardized REST web services, in which compounds, data sets, and predictive methods have a unique URI used to retrieve its associated Resource Description Framework (RDF) representation (Tcheremenskaia et al., 2012). OpenTox is available at three levels aimed at different use cases: a user interface for access to predictions, data, and models; an interface for new model development; and the public OpenTox API for new model development and integration (Hardy et al., 2010). APIs support the integration of chemical and toxicity data, and the development of end user-oriented tools (Hardy collaborative). One application of OpenTox is the support of data mining and cheminformatics applications for modelling Quantitative Structure Activity Relationships (QSARs) (Hardy et al., 2010). Another is the development of distributed web services, two of which are ToxPredict, which given a chemical structure reports toxicities, and ToxCreate which, given a dataset predicts and validates a toxicity model (Hardy et al., 2010).

AOP-Ontology

The adverse outcome pathway ontology (AOPO), located at https://github.com/DataSciBurgoon/aop-ontology, is specific to managing data for AOP's (DataSciBurgoon, 2015a). The purpose of the ontology is to provide: a) Semantic NoSQL way of accessing and managing Adverse Outcome Pathways (AOP) b) Ways and means to facilitate computation on data within the OECD AOP-KB c) Ways and means to standardize the lingua franca of the OECD AOP-KB d) The data framework to enable the development of artificial intelligence for predictive toxicology e) A central repository of toxicological knowledge to facilitate rapid and more efficient risk assessment.

AOPO is "capable of performing first order logic inferences", coupling "key events to assays to phenotypes" (DataSciBurgoon, 2015b). It enables queries; for instance, "At what concentration is chemical X likely to cause leukemia based on data from Tox 21 assays?" (DataSciBurgoon, 2015b). Other queries might identify chemicals associated with a disease, or develop a test battery for assays associated with disease (DataSciBurgoon, 2015b). The core AOPO is restricted to a formal description of the AOP and its components (key events and key event relationships). By design, it doesn't include any biological information. The biological information is inherited from previously established ontologies such as ChEBI, Human Phenotype Ontology, and Bioassay Ontology. This structure makes the long-term management of the AOPO relatively straightforward as the AOP components are static and the biological information will be maintained separately by the outside ontology efforts.

The ontology consists of an OWL file to facilitate use in semantic web and web data (DataSciBurgoon, 2015a). Ultimately, AOP-Ontology will contribute to the AOP-Knowledgebase by providing a unified controlled vocabulary for AOP components.

CHAPTER 3: CREATING AND APPLYING A DATA MODEL TO ANNOTATE ADVERSE OUTCOME PATHWAYS (AOP'S)

Introduction

With the advent of alternative approaches to traditional toxicological assessments of environmental chemicals, namely the use of high throughput (HTS) and high content (HCS) screening assays, came the generation of vast amounts of data and the need to organize such data to support regulatory decision-making. The AOP framework provides the biological context and supporting scientific evidence for organizing such alternative data. The objective of this chapter is to demonstrate the results and application of a common controlled vocabulary to address the data needs of the AOP-Knowledgebase, located at (http://aopkb.org). First, we review the data needs of currently approved AOP's and AOP's under development at levels of biological organization from molecular to population. Second, we define a data model to represent these needs. Third, we designate a common set of ontologies and controlled vocabularies at levels of biological organization relevant to the AOP, to annotate key terms from AOP entities. Finally, we conduct a cross-analysis applying ontology terms to AOP entities. Using the results of this mapping, we further define properties that apply to the data model. Case examples illustrate how the data model represents AOP entities at different levels of biological resolution. Future directions for application to the AOP-KB and AOP Framework as a whole are detailed.

Data resources and retrieval tools publicly available on the World Wide Web provide biologists with access to a variety of information including genomic, chemoinformatic, biological pathway, phenomic, ontological, metabolomic and toxicogenomic data (Oki et al.,

2016). While some sources are specific to one data type, others serve as aggregators or curators of information from disparate data sources (Oki et al., 2016). Publicly available data sources that may be used to drive AOP development include repositories storing multiple types of information. Macromolecular sequence databases, such as GenBank and SWISS-PROT, contain information about nucleotides or protein sequences and may be used in the discovery of novel genes, identification of homologous genes, and detection of polymorphisms (Pandey & Lewitter, 1999). Databases such as Gene Expression Omnibus (GEO) provide storage and retrieval of data sets from high throughput gene expression and genomic hybridization experiments (Edgar et al., 2002). Comparative Toxicogenomics Database (CTD) includes toxicogenomic connections relating chemicals/drugs, genes/proteins, diseases, taxa, phenotypes, GO annotations, pathways, and interaction modules (Davis et al., 2015). Model organism databases such as Mouse Genome Informatics (MGI), FlyBase, and Wormbase organize species-specific genomic information. Pathway resources include Kyoto Encyclopedia of Genes and Genomes (KEGG), "for understanding higher-order functional meanings and utilities of the cell or organism from its genome information" (Baclawski & Niu, 2006). These data sources inform the different levels of biological organization (molecular, cellular, tissue, organ, organism, and population-level) that populate the AOP.

AOP's integrate diverse data types from *in vitro, in vivo, in silico,* and *in chemico* tests supporting KE's at each level of biological organization of the AOP, providing a mechanistic understanding of a chain of biological events to inform regulatory decision-making regarding a chemical. AOP developers should be concerned with identifying the data needs of AOP's and how to integrate this information in a usable format. According to Oki et al., the biological space covered by existing AOP's is lacking, and the current process for AOP development is unable to

address this issue in an acceptable time frame (2016). Computational approaches can speed up the process of AOP development by avoiding the rate-limiting step of expert derivation (Oki et al., 2016). One difficulty is that the incorporation of data from a variety of approaches has resulted in an assortment of diverse terms and definitions, leading to confusion among scientists and organizations.

The OECD recommended development of a standardized set of terminologies to assist in the understanding of the AOP concept and in applying terms relevant to the AOP concept in developing QSAR's and chemical categories to advance the use of predictive techniques in assessments (OECD, 2013a). Use of a common vocabulary providing a cohesive description of events happening across different datasets of interest will facilitate data integration and identification of pathways developed from different datasets (Oki et al., 2016). Providing centralized databases for cross-reference of terms and other information, allowing for identification and aggregation of shared information, will contribute to further development of AOP's in a network framework (Oki et al., 2016). Using a standardized set of ontologies and controlled vocabularies will improve the ability to attach evidence in support of the AOP, supporting aggregation of data from HTS assays and experiments (Edwards, personal communication, July 8, 2016). This will speed up the rate of AOP development by reducing the computational challenge of translating information across datasets in individual ontologies (Oki et al., 2016).

The existing AOP-Ontology provides a set of terms and relationships with which to manage data on AOP's and facilitate computation on data in the AOP-KB (DataSciBurgoon, 2015a); however, it excludes text portions of AOP's such as KE names, KE descriptions, level of biological organization, and the biological context in which a KE occurs. A data model to

provide structure for these entities external to AOP-Ontology, will provide computability for these components.

Methods

A review of OBO Foundry member ontologies obtained from the OBO Foundry website at http://www.obofoundry.org/ was conducted in order to choose a minimum list of ontologies and controlled vocabularies (Table 1) from which to draw ontology terms or "classes".

Table 1. Minimum list of ontologies selected from the review of all ontologies and controlled vocabularies, ordered by level of biological organization.

Granularity	Ontology	Domain
Molecular	Ontology for Biomedical Investigations (OBI)	experiments
	Sequence (SO)	biological sequence
	ChEBI	biochemistry
	Protein Ontology (PR)	proteins
	GO	molecular function & gene products
	Molecular Process Ontology (MOP)	molecular function
	Protein-protein interaction (mi)	experiments- protein-protein interaction
Cellular	Cell Ontology (CL)	in vitro cell type
	Cell Line Ontology (CLO)	cell line information
	BRENDA tissue/enzyme source (BTO)	source of an enzyme- tissues, cell lines, cell types, cell culture
Tissue/Organ	Foundational Model of Anatomy (FMA)	anatomy- vertebrate
	Uberon	anatomy-metazoa
	Mammalian phenotype (MP)	phenotype
	Human Phenotype Ontology (HP)	Human phenotype
	vertrebrate trait (vt)	traits of vertebrates
	Phenotypic Quality (PATO)	phenotypic quality
Individual	neuro behavior ontology (NBO)	behavioral processes and phenotypes
Population	Population and Community Ontology (PCO)	population entity
	Experimental Factor Ontology (EFO)	experimental factor
	NCBI Taxon (ncbitaxon)	taxonomy
Non-OBO	AOP-Ontology	AOP components
Controlled vocabulary	UMLS/Medical Subject Headings (MeSH)	experiments

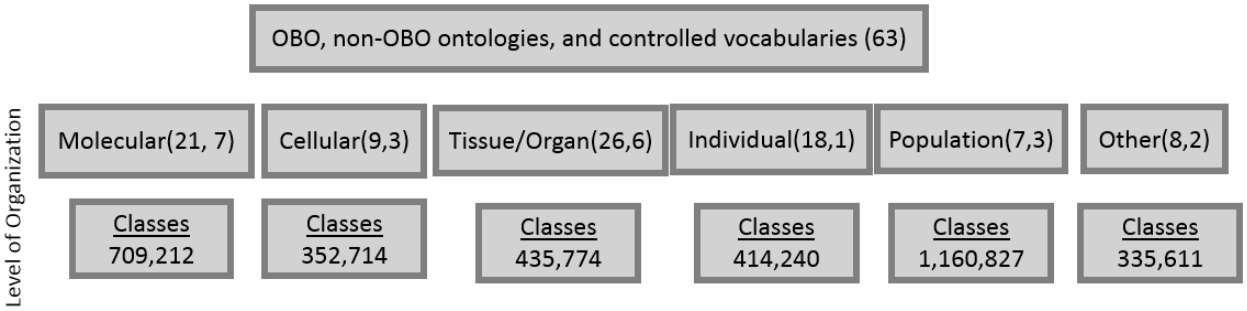

Figure 2. Review of ontologies and controlled vocabularies at each level of biological organization to select the minimum list. Numbers in () represent (ontologies reviewed at each level, ontologies selected at each level). Classes represents the total possible number of ontology classes or terms contained in the selected ontologies at that level of organization. The number of terms at each level was duplicated for ontologies representing more than one level of organization to ensure that the maximum number of possible terms was represented.

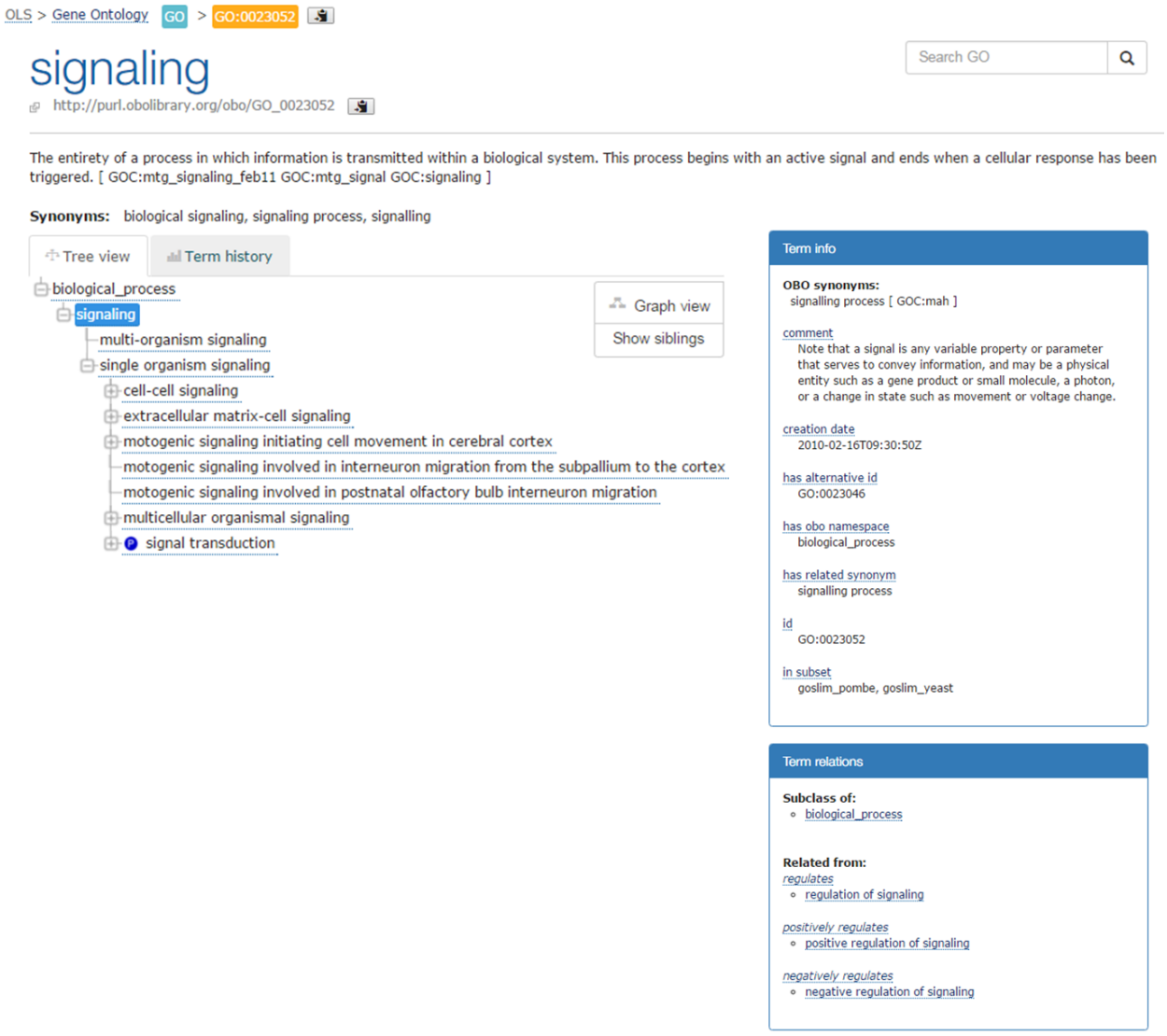

Figure 3. Ontology Lookup Service (OLS) output for a search on a GO term (signaling). Tree View of the GO term, its term info, and term relations are displayed.

OBO Foundry ontologies were chosen as the most broadly accepted and syntactically uniform ontologies available when possible. Ontologies were reviewed based on published literature for their domain of study, their taxonomic applicability, syntax and available format for download, the level of biological organization that they cover from molecular to population, and whether the ontology imported or aggregated other data sources. The upper level Relations

Ontology (RO) and Basic Formal Ontology (BFO) were excluded, as it was assumed that they contain relations common to OBO Foundry member ontologies (Brinkman et al., 2010).

An initial review was conducted of the OBO Foundry ontologies (Appendix A), a few ontologies that were not OBO Foundry members, as well as several controlled biomedical vocabularies from the literature. The final selection (Table 1) of ontologies and vocabularies was chosen based on literature review of those containing entities at a level of biological organization applicable to AOPs; as well as characteristics and chosen vocabularies used in projects involving construction of new frameworks drawing on existing ontologies, such as the Beta Cell Genomics project (Zheng et al., 2013) and PubChem RDF project (NCBI, 2015). This selection represented the minimum number of ontologies that could accurately represent the biological space of key events in the AOP Framework.

After selection of ontologies, key event terms from AOP Wiki (https://aopwiki.org/wiki/index.php/Main_Page) were mapped to classes from the chosen ontologies. For 17 priority AOP's in the categories "Currently under OECD EAGMST Review" and "Open for General Comments", and for 6 AOP's in the "EAGMST Approved" category, KE pages from the Wiki were manually reviewed for text query phrases. Query phrases representative of key events from these pages were entered into three ontology browser sites: Ontobee (http://www.ontobee.org/), NCBO Bioportal (http://bioportal.bioontology.org/), and the EMBL-EBI Ontology Lookup Service (http://www.ebi.ac.uk/ols/beta/index). From the list of "hits" on the search browser output, matching ontology classes (entities) were reviewed for their match to query terms and the appropriate ontology class was chosen (Figure 3). When possible, the least specific class available to accurately portray the information was selected, to promote reusability of terms and minimize the number of classes chosen.

The AOPO (https://github.com/DataSciBurgoon/aop-ontology) was extended to include a data structure (Figure 4) that best captures the results of the matching process between ontology classes and KE terms. Ontology classes were aligned with a data model containing the fields: Process, Object, Action, and Context. Biological process was defined as a recognized series of events or molecular functions representing the key action happening in the event as defined by GO (Gene Ontology Consortium, 2015b). Object represents the biological object to which the process is occurring, a chemical entity or cellular component such as 'vitellogenins' or 'hepatic stellate cell'. Action terms representing directionality were chosen from the following: Increased, Decreased, Altered, Delayed, Accelerated, or Unknown (Crofton, K., personal communication, April 8, 2016). KE's occurring at the Molecular and Cellular level of biological organization were assigned a Cellular Context, and terms at the Tissue and Organ level were assigned an Organ Context based on either the cell or the organ where the event was applicable, and terms at the Individual or Population level were not assigned a context. The data fields of Process, Object, and Context were linked to terms from external ontologies.

Results were compiled and analyzed using Excel. The choice of ontology terms used was validated through review by and consultation with experts and AOP authors (Burgoon, L., Edwards, S., Villeneuve, D., personal communication, April 8, 2016).

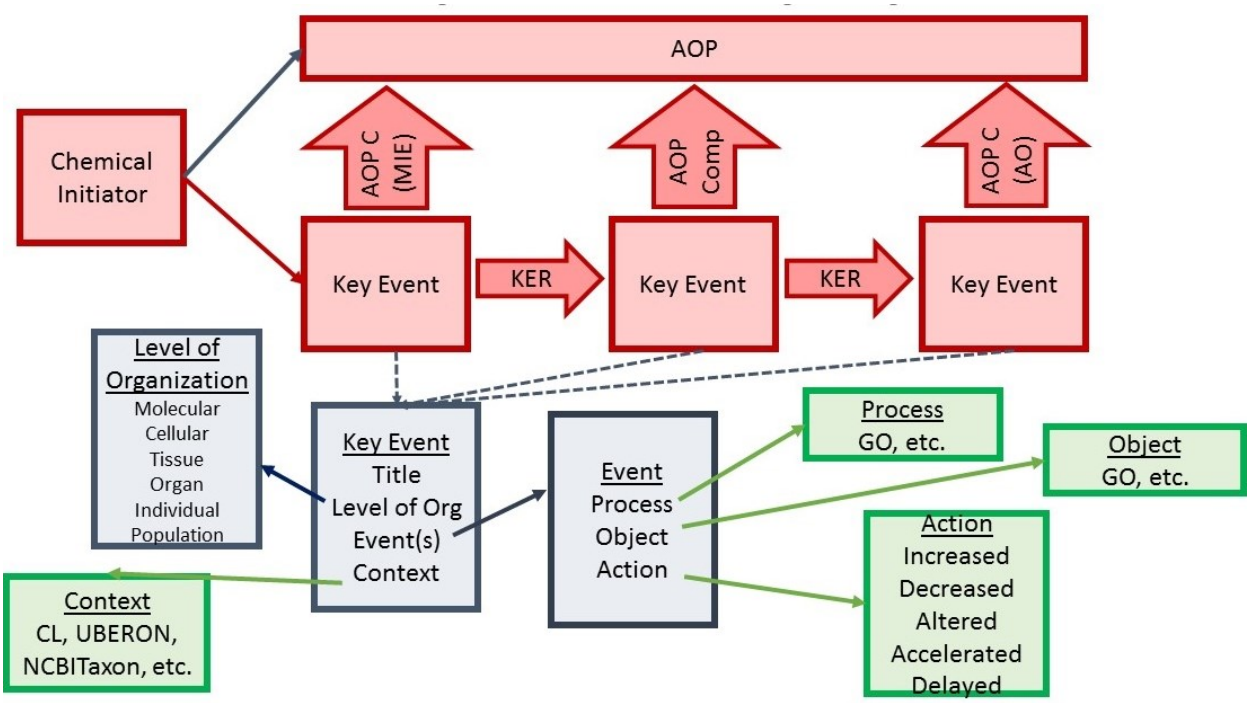

Figure 4. Schematic for how the data model extends the AOP Ontology. Components contained in the existing AOP-Ontology are in red. The data model provides Process, Object, Action, and Context (green), which are external but linked to AOP Ontology via the annotation of KE's. Entities in blue are part of the AOP-Wiki application. Uberon, CL, and GO represent examples of ontologies that these data fields will import from.

Results

A review was conducted of 64 ontologies and controlled vocabularies, 49 of them which were OBO Foundry member ontologies (Appendix A). From this, it was determined that a final list of 22 ontologies and vocabularies was the minimum necessary to represent key events at the molecular, cellular, tissue/organ, individual, and population levels (Table 1). The total number of possible terms at each level represents a subset of the total number of possible terms of all the ontologies reviewed (Figure 2) as calculated using estimates of number of classes from NCBO Bioportal and Ontobee (Burgoon, L., Edwards, S., Villeneuve, D., personal communication, April 8, 2016). Calculation of the possible number of total terms did not account for those ontologies that import terms from other ontologies, so may include overlap. Of the OBO

Foundry ontologies selected, two of these were OBO-Foundry application ontologies, BRENDA tissue/enzyme source (BTO) and Experimental Factor Ontology (EFO) and the remainder were OBO Foundry reference ontologies. One controlled vocabulary, Medical Subject Headings (MeSH), was selected due to its widespread use even though it is not a formal ontology. Ontologies were ordered by level of biological organization according to the minimum level of organization that they represent. For example, while GO includes entities at the cellular level, for the purposes of this review it was considered a molecular level ontology. Finally, in the domain of experiments, the Ontology for Biomedical Investigations (OBI), Protein-protein interaction (MI), and Experimental Factor Ontology (EFO) were included.

The AOPO forms the basis for incorporation of other ontologies as the resulting minimum list of ontologies (Table 1) extends the AOPO and the ontologies it currently inherits. The data model (Figure 4) capturing the results of the matching process between ontology classes and KE terms provides an organizational framework for mapping of KE terms to ontology classes from the minimum list. As shown, the selected ontologies extend the currently existing components of the AOP-Ontology and the ontologies that it currently imports.

The resulting data model developed to extend AOPO provides representation for KE entities, including Events as represented by terms obtained from the selected ontology list (Table 1). Using a text match approach mapping terms from KE's to ontology classes from the selected list, and the data properties needed to represent this information, resulted in the data model.

You should include a paragraph here describing the AOP ontology and how it forms the basis for incorporating all the other ontologies. You can use Figure 4 as the basis for this discussion. I would then discuss the data model again since not everyone reads the methods. That will set up the discussion below.

Mapping the results of search queries for the 23 AOP's to ontology classes from the selected list, resulted in a total of 173 key events, 155 of them unique, being mapped to 467 ontology classes, 209 of them unique as defined from the selected ontology list. Three terms could not be matched to an ontology class or were undefined (Appendix B). This shows that more than half of ontology classes were duplicates, and that only a small part of each ontology was needed to accurately describe the chosen key events.

For example, the KE "17beta-estradiol synthesis by ovarian granulosa cells, Reduction" can be condensed into the query terms "synthesis," "17beta-estradiol," and "ovarian granulosa cells" from the KE name and the text description on the wiki page for the event at https://aopwiki.org/wiki/index.php/Event: 3 (2016), resulting in a mapping to ontology classes as:

Process: *'biosynthetic process'* (GO: GO_0009058)
Object: *'17beta-estradiol'* (CHEBI: CHEBI_16469)
Context: *'granulosa cell'* (CL: CL_0000501)
Action: '*Decreased*.'

Resulting from the mapping of key event terms to ontology classes, came to define properties for the data structure of Processes, Objects, Context, and Actions. The following properties were arrived at to better define requirements for each data field within the structure:

(1) *Multiple and blank fields*

One issue that arose was the necessity of using multiple ontology classes to define key events. It was decided that multiple classes would be allowed for Processes and Objects. Context was restricted to from 0 to 1 context entry; in other words, context is optional but cannot be multiple. Finally, blank fields are allowed for Objects and Contexts, but not for Processes and Actions.

(2) Excluded entities

Based on the definition of an Event, it was decided that relationships, chemicals, and experiments or assays would be excluded from this mapping. These components are represented in AOP-Ontology (DataSciBurgoon), and this data structure was intended to be external to the AOP-Ontology. An exception was the representation of chemical entities; a chemical was included when it was an endogenous metabolite as opposed to exogenous.

(3) Specificity of ontology class

When a query resulted in multiple "hits," it was determined that the least specific ontology class to accurately represent the Event should be selected. This allowed for maximal reuse of classes while accurately portraying the Event.

(4) Consideration of taxonomic applicability

Efforts were made to match the chosen ontologies based on taxonomic applicability as defined in the Event. When an Event included multiple taxons, an ontology that was minimally taxon-specific was chosen.

(5) Restrictions on level of organization

Some ontologies and vocabularies contain classes at multiple levels of biological organization. GO's 'biological_process' sub-ontology, for example, contains classes describing processes at the tissue, organ, and organism levels (NCBO, 2016a). Where possible, we tried to restrict the choice of ontology based on the level of organization it represented.

Further examples of these data properties are illustrated in the following case studies.

Case Example 1: Aromatase inhibition leading to reproductive dysfunction (in fish)

This AOP is one of three related AOP's operating through the shared KE of impaired vitellogenesis, leading to reduced fecundity and declining population levels in fish (Ankley et al., 2010). The MIEs for these AOP's are: ER antagonism, aromatase inhibition, and AR agonism (Ankley et al., 2010). This AOP's taxonomic applicability is the fathead minnow (*Pimephales promelas*), although the AOP is applicable to other oviparous vertebrates as well (Becker et al., 2015). The MIE is inhibition of cytochrome p450 aromatase (cyp19a1), and associated KE1 is reduced 17beta-estradiol synthesis by ovarian granulosa cells. KE2 is a reduction in plasma 17beta-estradiol concentrations; KE3, reduced transcription and translation of vitellogenin in the liver; KE4, reduced plasma vitellogenin concentrations; KE5, reduced vitellogenin accumulation into oocytes and oocyte growth/development; KE6, reduced cumulative fecundity and spawning; and the AO, declining population trajectory (Villeneuve, 2016). In ovarian granulosa cells, aromatase catalyzes the conversion of testosterone to 17beta-estradiol (E2), which stimulates production of vitellogenin in the liver. Hepatic vitellogenin is transported via circulation to the ovaries for uptake. Vitellogenin is necessary for oocyte development, and aromatase inhibitors cause a reduction in egg production leading to population decline. There is strong biological plausibility, empirical support, and weight of evidence for the AOP as a whole, making it applicable for chemical assessment and regulatory purposes (Becker et al., 2015).

Applying the data structure for mapping key event terms to ontology classes, allows the visualization of how the data structure applies to key events at each level of biological organization (Figure 4). 'AOP 25' illustrates the data properties and discussion points regarding choice of the relevant class for each process, object, and context, and selection of the appropriate

Action term. Illustrating data property (1), the mapping for KE3 allows for multiple terms to describe the process field, accurately portraying the complexity of key events:

Process: *GO: 'gene expression'* and *GO: 'translation'*
Object: *MESH: 'vitellogenins'*
Context: *CL: 'hepatocyte'* and modified by
Action: '*Decreased*.'

Secondly, for the KE6 and the AO, the Object and Context fields are blank. Allowing a blank object is useful when a process fully captures the KE without need for an object; the context at the Organism and Population level is defined by the different applicability domains (species, sex, and life stage) that are captured for all KEs, so no separate Context term is needed. For KE6,

Process: *PATO: 'fecundity'*
Object: *UBERON: 'egg'*
Context: n/a
Action: '*Decreased*.'

For the AO:
Process: *PCO: 'population growth rate'*
Object: n/a
Context: n/a
Action: '*Decreased*.'

Illustrating property (2), for KE2, is the inclusion of the object '17beta-estradiol' from ChEBI. Since E2 is an endogenous chemical entity produced by synthesis in granulosa cells, it was not excluded from the mapping:

Process: *GO: 'biosynthetic process'*
Object: *CHEBI: '17beta-estradiol'*
Context: *CL: 'granulosa cell'*
Action: '*Decreased*.'

Finally, according to property (4), it should be noted that this AOP's taxonomic applicability is specific to fish, so mappings to ontologies such as MP and HP were excluded.

This AOP also illustrates several of the discussion points regarding the mapping of terms. For the MIE, the class chosen to best represent the key event term "Inhibition" was 'catalytic activity' with the Action modifier of 'Decreased.' When the key event text included "enzyme inhibition" or "enzyme activation", it was determined that 'catalytic activity' was a more accurate representation than 'signaling,' though both terms convey the transmission of information within a biological system resulting in a cellular response:

Process: *GO: 'catalytic activity'*
Object: *PRO: 'cytochrome p450 19A1'*
Context: *CL: 'granulosa cell'*
Action: '*Decreased*.'

A second discussion point is how best to describe processes involving transcription and translation. For the KE3, the key event text describes transcription of vitellogenin genes, being regulated by estrogens via their action on specific nuclear receptors (Villeneuve, 2016). 'Gene expression' was chosen to describe transcription of genes specific to vitellogenin, and 'translation' to describe the synthesis of the protein vitellogenin. In the GO hierarchy, 'gene expression' is a parent class for 'transcription, DNA-templated' and for 'translation', so 'translation' is the more specific of the two classes. 'Translation' was determined to be necessary to adequately describe the key event:

Process: *GO: 'gene expression'* and *GO: 'translation'*
Object: *MESH: 'vitellogenins'*
Context: *CL: 'hepatocyte'*
Action: '*Decreased*.'

The third discussion point was for the KE2 and KE4, the use of the process class 'blood circulation' twice within one AOP. It was decided that reuse of processes within an AOP should

be allowed, as long as the same Process, Object, Action, and Context mapping of terms didn't exist within non-identical key events. For KE2:

Process: *GO: 'blood circulation'*
Object: *CHEBI: '17beta-estradioll'*
Context: *Uberon: 'blood plasma'*
Action: '*Decreased*.'

and KE4:

Process: *GO: 'blood circulation'*
Object: *MESH: 'vitellogenins'*
Context: *Uberon: 'blood plasma'*
Action: '*Decreased*.'

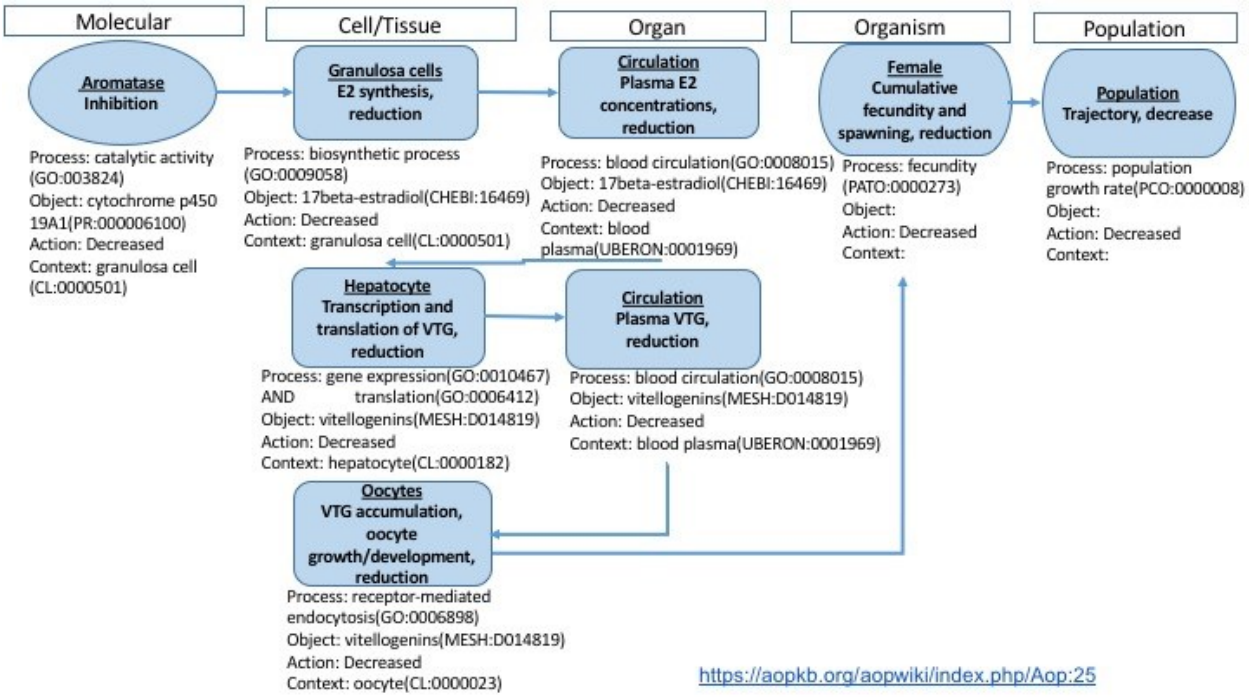

Figure 5. Graphical view of "Aromatase inhibition leading to reproductive dysfunction (in fish)" with key event terms mapped to ontology terms. CL, CHEBI, GO, MESH, PATO, PCO, PR, UBERON represent ontologies and their associated ID's.

Case Example 2: Chronic binding of antagonist to N-methyl-D-aspartate receptors (NMDARs) during brain development induces impairment of learning and memory abilities.

N-methyl-D-aspartate receptors (NMDARs) are a member of a group of ionotropic receptors activated by the neurotransmitter, L-glutamate (Glu) (Sachana et al., 2016). Activation of NMDAR in the hippocampus of the developing rat brain results in synaptic functioning associated with learning and memory processes. Long term potentiation as a result of NMDAR activation results in increased synaptic strength, plasticity, and memory formation in the hippocampus. NMDAR enhances release of brain-derived neurotrophic factor (BDNF) resulting in neuronal survival, differentiation, and synaptogenesis. Blockage of NMDAR by chemical substances such as lead (Pb^{2+}), is associated with disrupted neuronal network function and impaired learning and memory in the developing brain.

The MIE, NMDARs, Binding of Antagonist, is followed by KE1, Inhibition of NMDARs. Following KEs are: KE2, Decreased calcium influx; KE3, Reduced release of BDNF; KE4, aberrant dendritic morphology; KE5, reduced presynaptic release of glutamate; KE6, Cell death; KE7, decreased synaptogenesis; KE8, decreased neuronal network function; and the AO of Impaired learning and memory.

KE5 illustrates data property 2. The object term 'Glutamate' is an endogenous entity, and was included despite the exclusion of chemicals from the mapping (Figure 5):

Process: *GO: 'neurotransmitter secretion'*
Object: *ChEBI: 'glutamate'*
Context: *Uberon: 'blood plasma'*
Action: '*Decreased*.'

This AOP brings up multiple points of discussion. First, using the structure to map the MIE and KE1 resulted in an identical mapping for both key events, with 'signaling' representing both of the event text terms "Inhibition" and "Binding." As in the previous case example, shared

processes are included within a single AOP, but identical terms for Process, Object, Action, and Context shouldn't result within an AOP. It was determined that one event at the molecular level should be chosen as the MIE, and that a discussion should be taken up with the AOP author to remove the extraneous key event. 'Signaling' provides the best process term to accurately portray this event, as inhibition and antagonism both can be described by increased or decreased signaling of the NMDA- receptor. For KE1,

Process: *GO: 'signaling'*
Object: *GO: 'NMDA glutamate receptor complex'*
Context: *Uberon: 'hippocampus'*
Action: '*Decreased*.'

Secondly, this AOP brings up a discussion regarding the selection of terms for inter- and intracellular transport. GO defines 'transport' as "The directed movement of substances into, out of or within a cell, or between cells, or within a multicellular organism by means of some agent such as a transporter or pore" (Ashburner et al., 2000). Subclasses of GO: 'transport' include 'extracellular transport', 'intracellular transport', 'intercellular transport', and 'vesicle-mediated transport.' KE2 describes the directed movement of Ca^{2+} into the cell, where it functions as a signaling molecule to regulate synapse and neuronal cell function (Sachana et al., 2016). It was determined that for the majority of key events in the mapping, 'transport' was adequate. For KE2,

Process: *GO: 'transport'*
Object: *ChEBI: 'Calcium (2+)'*
Context: *CL: 'neuron'*
Action: '*Decreased*.'

For KE3, the key event text describes the transcription and release of BDNF from glutamatergic neurons (Sachana et al., 2016). The process 'gene expression' was chosen to describe BDNF expression and transcription of the protein. GO: 'gene expression' is a parent

class for GO: ‘transcription, DNA-templated’ and for GO: ‘translation,’ so it was determined that this process description would be sufficient to include the key event term “transcription”:

Process: *GO: ‘gene expression’*
Object: *PR: ‘brain-derived neurotrophic factor’*
Context: *CL: ‘neuron’*
Action: ‘*Decreased*.’

KE5, KE7, and KE8 describe key events relating to synaptic function. For KE5, based on the key event text “reduced presynaptic release of Glutamate,” it was determined that ‘neurotransmitter secretion’ was the most applicable process. In GO, ‘Neurotransmitter secretion’ is a subclass of ‘synaptic transmission,’ but it was appropriate to make this specification in order to include ‘Glutamate’ as the object. For KE7, GO: ‘synapse assembly’ was a synonym for the query term “synaptogenesis” from the key event text. GO: ‘synapse assembly’ is a subclass of GO: ‘cellular component assembly’ (Ashburner et al., 2000). For KE8, the query term “neuronal network communication” from the key event did not have an exact match in GO, but it was determined that the event was referring to glutamatergic neurotransmission (Sachana et al., 2016), so the process term was determined to be GO: ‘synaptic transmission.’ For KE5:

Process: *GO: ‘neurotransmitter secretion’*
Object: *ChEBI: ‘glutamate’*
Context: *Uberon: ‘blood plasma’*
Action: ‘*Decreased*.’

For KE7,
Process: *GO: ‘synapse assembly’*
Object: *GO: ‘synapse’*
Context: *CL: ‘neuron’*
Action: ‘*Decreased*.’

For KE8,
Process: *GO: 'synaptic transmission'*
Object: n/a
Context: *Uberon: 'brain'*
Action: '*Decreased*.'

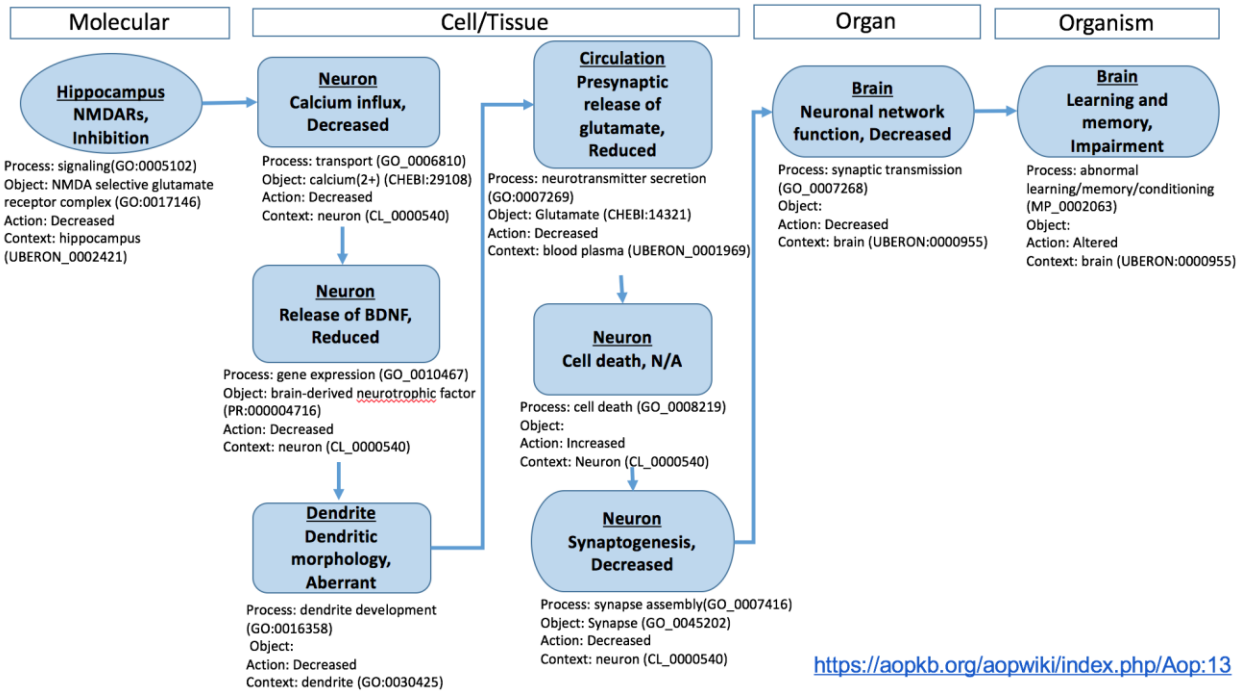

Figure 6. Graphical view of "Chronic binding of antagonist to N-methyl-D-aspartate receptors (NMDARs) during brain development induces impairment of learning and memory abilities" with key event terms mapped to ontology terms. CL, CHEBI, GO, MP, PATO, PR, UBERON represent ontologies and their associated ID's.

Case Example 3: PPARa activation in utero leading to impaired fertility in males

PPARα is a ligand-activated nuclear receptor, a transcription factor in the steroid/thyroid/retinoid family of receptors (Nepelska et al., 2016). Activation of PPARα in the tissues responsible for fatty acid catabolism affects lipid metabolism. The key events of this AOP are: the MIE, PPARα activation; KE1, steroidogenic acute regulatory protein (STAR) decrease; KE2, translator protein (TSPO) decrease; KE3, reduced cholesterol transport in mitochondria; KE4, reduced testosterone synthesis; KE5, reduced testosterone levels; and the AO of male

reproductive tract malformation and impaired fertility. This is a reproductive AOP applicable only in males. Possible applications of the AOP are in identification of endocrine disrupting chemicals.

Examining the mapping for this AOP (Figure 6), KE1 and KE2 exhibit data property (1). For both STAR and TSPO proteins, the key event text identifies gene expression and reduced tissue levels of the protein, so the representative classes were 'gene expression' and 'accumulation.' For KE1:

Process: *GO: 'gene expression'* and *PATO: 'accumulation'*
Object: *PR: 'STAR'*
Context: *Uberon: 'Leydig cell'*
Action: '*Decreased*.'

and KE2:
Process: *GO: 'gene expression'* and *PATO: 'accumulation'*
Object: *PR: 'translocator protein'*
Context: *Uberon: 'Leydig cell'*
Action: '*Decreased*.'

KE4 brings up an example of selecting the correct class relating to "transport" as described earlier. The key event text describes reduced transport of cholesterol from intracellular stores to the inner mitochondrial membrane (Nepelska et al., 2016). GO: 'mitochondrial transport' is a subclass of GO: 'transport,' and as 'transport' was the more general and reusable term this was selected for the process field:

Process: *GO: 'transport'*
Object: *ChEBI: 'cholesterol'*
Context: *Uberon: 'Leydig cell'*
Action: '*Decreased*.'

For KE5 and KE6, the challenge of how to correctly phrase reduced "testosterone synthesis or steroidogenesis", followed by a reduction in circulating testosterone levels. The

terms GO:‘biosynthetic process’ and GO: ‘blood circulation’ were chosen for the process field, and the object for both was CHEBI: ‘testosterone.’ For KE5:

Process: *GO: ‘biosynthetic process’*
Object: *ChEBI: ‘testosterone’*
Context: *Uberon: ‘Leydig cell’*
Action: ‘*Decreased*.’

and for KE6,
Process: *GO: ‘blood circulation’*
Object: *ChEBI: ‘cholesterol’*
Context: *Uberon: ‘Leydig cell’*
Action: ‘*Decreased*.’

The AO illustrates data property (1). AO is described by GO: ‘fertility’ and an action term of ‘Decreased,’ so having an object was not necessary. Since this KE was at the organism level, by the definition of context, a context was not necessary.

For the AO:
Process: *PATO: ‘fertility’*
Object: n/a
Context: n/a
Action: ‘*Decreased*.’

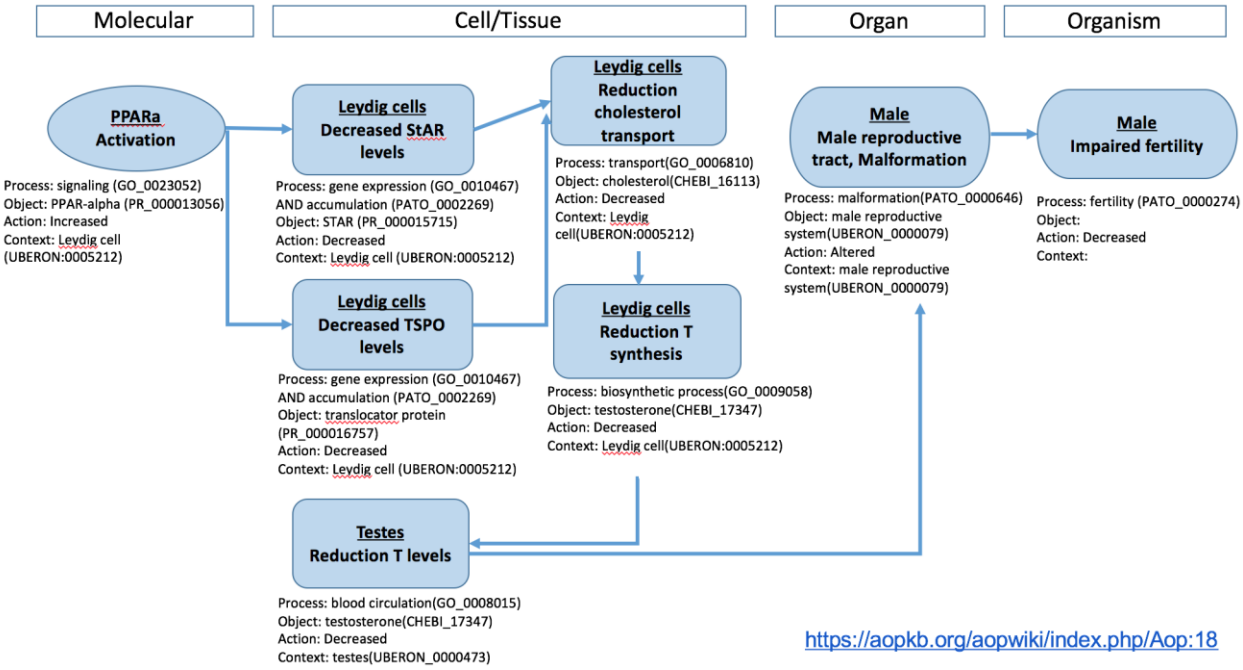

Figure 7. Graphical view of "PPARa activation in utero leading to impaired fertility in males" with key event terms mapped to ontology terms. CHEBI, GO, MP, PATO, PR, UBERON represent ontologies and their associated ID's.

Case Example 4: AFB1: Mutagenic Mode-of-Action leading to Hepatocellular Carcinoma (HCC)

Aflatoxin B1 (AFB1) is an example chemical for the induction of hepatocellular carcinoma (HCC) via a mutagenic mode of action (Pottenger et al., 2016). In gene mutation assays, AFB1 induces mutations in codon 249 of the p53 gene in humans, responsible for cancer etiology of HCC (Pottenger et al., 2016). The MIE for this AOP is the formation of pro-mutagenic DNA adducts. KE1, metabolism of AFB1 to produce reactive electrophiles; KE2, mutation induced in critical genes; KE3, clonal expansion/cell proliferation to form altered hepatic foci (AHF); KE4, insufficient or misrepair of DNA adducts; and the AO, hepatocellular carcinoma and tumorigenesis (Figure 7).

KE1 illustrates data property (2). Since AFB1 is the chemical initiator in this AOP, although it induces mutations in gene mutation assays it is intended as a case example of multiple chemicals that could act through this mode of action resulting in HCC (Pottenger et al.,

2016). As an exogenous chemical entity, it was excluded from the mapping, and the object for this mapping was left blank. The AO, also has a blank object, because no modifiers were needed for the process 'hepatocellular carcinoma.' For KE1:

Process: *GO: 'metabolic process'*
Object: n/a
Context: *Uberon: 'liver'*
Action: '*Increased*.'

The MIE for this AOP illustrates data property (3) of how to choose the most representative term illustrating the formation of pro-mutagenic DNA adducts by chemical metabolites. The parent class of GO: 'DNA alkylation' is GO: 'DNA modification.' However, in order to include key events involving GO: 'protein alkylation', GO: 'cellular macromolecule metabolic process' was the overarching parent term. Thus, in order to remain sufficiently descriptive without losing connotation, the more specific term 'DNA alkylation' was chosen to represent this process. Likewise, FMA: 'nuclear DNA' rather than FMA: 'DNA' was chosen to represent the object for this event based on the key event text:

Process: *GO: 'DNA Alkylation'*
Object: *FMA: 'nuclear DNA'*
Context: *Uberon: 'hepatocyte'*
Action: '*Altered*.'

Another discussion point brought up by KE2, was the choice of class for the key event terms "mutation", "heritable mutation", and "induced mutation". Although this event describes an induced mutation in a codon for p53 gene in humans, it was decided that choosing a process representative of "mutation" should be maximally reusable for other key events. In querying the select list of ontologies, it was found that EFO: 'induced mutation' was a subclass of EFO: 'genetic modification,' but 'genetic modification' did not accurately describe the key event terminology. Therefore, the more general MI: 'mutation' was chosen to represent this key event:

Process: *MI: 'mutation'*
Object: *PR: 'cellular tumor antigen p53'*
Context: *Uberon: 'hepatocyte'*
Action: '*Altered*.'

Finally, this AOP specified the genetic mutation in human p53 gene, but the AO of HCC occurs in multiple species including birds, fish and mammals (Pottenger et al., 2016). According to property (4), the term EFO: 'hepatocellular carcinoma' was chosen rather than HP: 'hepatocellular carcinoma' or MP: 'hepatocellular carcinoma':

Process: *EFO: 'hepatocellular carcinoma'*
Object: n/a
Context: *Uberon: 'liver'*
Action: '*Increased*.'

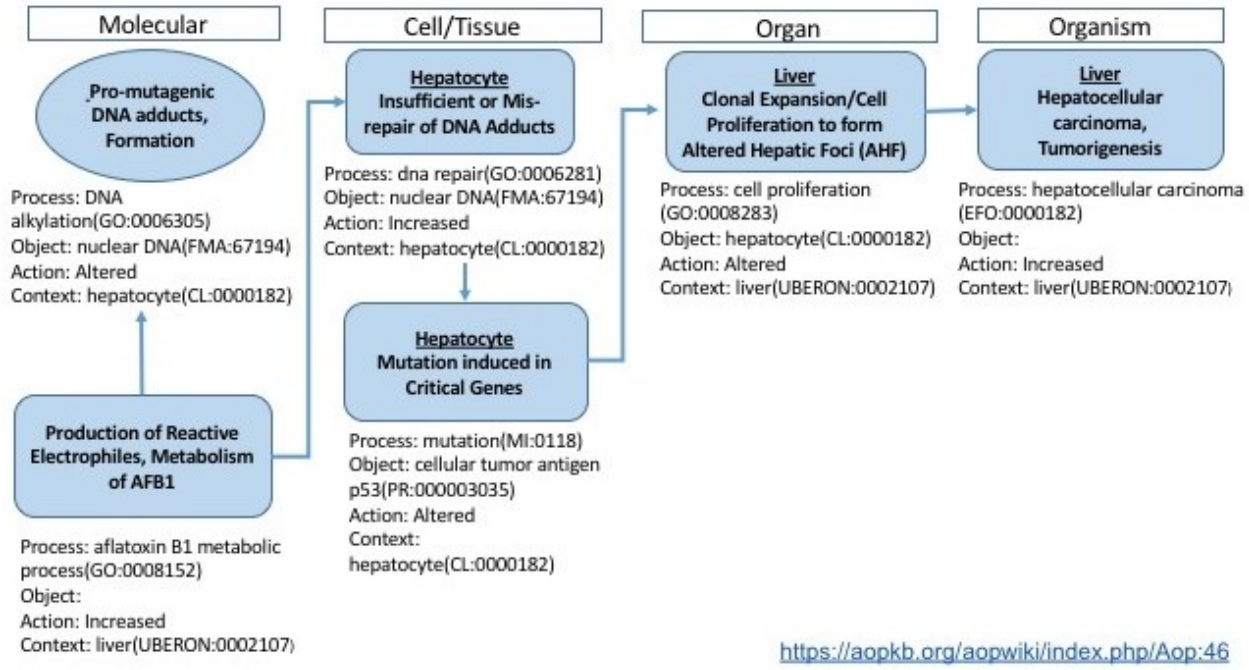

Figure 8. Graphical view of "AFB1: Mutagenic Mode-of-Action leading to Hepatocellular Carcinoma (HCC)" with key events mapped to ontology terms. CL, EFO, FMA, GO, MI, PR, UBERON represent ontologies and their associated ID's.

Discussion

The benefits of applying our data model to the AOP-KB are that it provides the ability to attach scientific evidence in support of the AOP, including assays and biomarkers linked to KE's. The model supports aggregation of data from HTS assays and experiments, supporting KE's. It supports AOP networks, which can then be used for further AOP identification and evaluation (Oki et al., 2016). It also provides for the evaluation of taxonomic relevance of KE's (Oki et al., 2016).

The network framework of AOP's relies on the identification of shared AOP components, KE's and KER's, reusable between different AOP's. Application of ontologies and controlled vocabularies provides a uniform description for events across datasets in a formalized language (Oki et al., 2016), promoting querying and data mining capabilities. Selecting a set of ontologies and a structured approach to their use, as applicable to the AOP framework at all levels of biological organization, will enhance cpAOP development as a basis for expert-derived AOP's. Also, the set of ontologies and data framework arrived at through this project will inform user queries for the definition of KE's, enhancing the capabilities of the AOP-Wiki application.

Hundreds of ontologies and controlled vocabularies representing different domains are publicly available on the web, and reviewing these for a standardized set to inform the AOP framework at the molecular, cellular, tissue, organ system, and organism levels of biological organization led to a standardized set of ontologies and vocabularies from which terms were queried. Mapping text terms from KE's in AOP-Wiki to ontology terms from this minimum set, showed that the number of terms necessary to identify KE's was a minimal set of the total number of terms represented by these ontologies. However, one limitation in restricting to this minimum list was that several ontologies, especially GO and MP, import from ontologies

external to this list and contain duplicate terms. Additionally, ontologies such as GO that covered more than one level of biological organization were counted for both levels of organization (Figure 2), and this may lead to duplicate terms. Possible duplicates will be accounted for in the AOP-Wiki by ensuring that the imported subset of ontology terms, stored in a database, from which users can choose does not contain duplicates.

Several of the ontologies from the minimum set were selected based on their characteristics and literature review, but not actually used in the mapping. These included Sequence Ontology (SO), Ontology for Genes and Genomes (OGG), and Ontology for Biomedical Investigations (OBI). This minimum set may change in the future, as ontologies useful at that level of organization are added and subtracted from the set. Analysis of the ontology terms used in the mapping showed that more than half of them were reused either in duplicate KE's throughout the mapping, or in nonduplicate KE's.

The data model with fields of Process, Object, Action, and Context for tying KE entities to ontologies was effective for the majority of AOP's examined. Benefits of this approach were that it effectively described the majority of AOP's reviewed, with few KE's not mappable to this framework. The definition of data properties for each field, led to the formulation of restrictions and allowances for ontology terms in the mapping for better definition of the data framework. These properties will inform user choices of terms in the AOP-Wiki application.

The process of mapping KE entities to ontology terms also led to multiple discussions about the selection of terms and specific issues resulting from the mapping, such as terms that could not be found, or were undefined by authors. A discussion with experts, and bringing up these issues with individual authors of the AOP, led to decisions about these issues and refinement of terms in the mapping.

One limitation was the manual curation of the terms selected; through the process of manual selection of KE text terms from AOP-Wiki, input into an ontology browser, and selection of the appropriate matching ontology class. Several tools available for ontology extraction include Ontodog, a web-based tool that can retrieve a set of terms and reasoners from an ontology file (Zheng et al., 2013). If the set of ontologies chosen for this project were to be integrated into a unique ontology, and for better integration with AOP-Ontology, software editors such as Protégé and Hermit should be used for integration, enrichment, and checking for logical consistency (Zheng et al., 2013). Ontology validation would ensure that the selected ontology classes are usable for other needs and applications.

Finally, the selection of case examples from the mapping highlights the usability of this framework for individual AOP's and to further the AOP network view through the identification of shared components. Providing an annotation containing Process, Object, and Context terms from ontologies as well as manually identified Action terms for each KE at levels of biological organization encompassed by AOP's, showed the variety in the types of information conveyed by the framework. Additionally, it highlights how the framework might be used to guide future AOP development.

In addition to its use for the AOP framework, the future use of this data model for the AOP-Wiki application is through the selection of ontologies to inform user queries. Currently, Wiki users select Taxonomic Applicability for a KE from a widget drop-down menu populated by NCBITaxon, and they select Organ Applicability from a menu populated by Uberon. Future capabilities could include the ability to define properties of KE's using the fields Process, Object, Action, and Context as derived from imported files from the selected set of ontologies and controlled vocabularies.

CHAPTER 4: APPLICATION OF THE DATA MODEL TO AOP DEVELOPMENT FOR REGULATORY ASSESSMENT

Introduction

In this chapter, we apply the data model developed previously to map ontology terms to AOP's for the risk assessment of inorganic Arsenic. We discuss the benefits and limitations of this structured approach, and provide detailed instances of discussion points that the mapping brought up. Finally, we discuss future directions for this model and how it will be applied to the AOP-KB in the future.

Inorganic Arsenic (iAs) is a commonly occurring environmental contaminant with adverse health effects, with more than 100 million people exposed worldwide to arsenic-contaminated water exceeding WHO's recommended limit of 10 ug/L (WHO, 2004). It has been associated with chronic health conditions in adults including cardiovascular disease, peripheral vascular disease, neurological effects, diabetes mellitus, and cancer in the lungs, liver, urinary bladder, prostate and skin (NRC, 2001; Rahman et al., 2009b). Exposure during pregnancy has been associated with maternal and fetal health effects (Bailey et al., 2014). Additionally, prenatal and early childhood exposure has been associated with delayed or postnatal health effects including increased rates of mortality in young adults from noncancerous diseases (bronchiectasis, acute myocardial infarction) and cancers of the lung, urinary bladder, larynx, and liver (Farzan et al., 2013).

IRIS (Integrated Risk Information System) is the source of toxicity information for the EPA, state and local health agencies, federal agencies, and international health organizations

(USEPA, 2015). EPA uses the program in setting national standards and cleaning up hazard sites.

According to EPA, the role of IRIS is in the first two steps of risk assessment: hazard identification and dose-response assessment (USEPA, 2015). Hazard identification involves identification of health hazards associated with response to a chemical. Dose-response assessment involves the quantitative relationship between a chemical exposure and a credible health hazard, and these quantitative relationships are used to derive toxicity values. The toxicity values derived are for health effects resulting from chronic exposure. They include the oral reference dose (RfD), the amount of a chemical one can ingest daily over a lifetime not anticipated to cause noncancer health effects, compared to exposure estimate in mg/kg-day; and the inhalation reference conference (RfC), the concentration of a chemical one can ingest daily over a lifetime not anticipated to cause noncancer health effects, compared to exposure estimate in mg/m^3. Additional values include cancer descriptors, describing levels of carcinogenicity to humans; oral slope factor, the estimate of increased cancer risk from oral exposure to 1 mg/kg-day for a lifetime; inhalation risk unit, the estimate of increased cancer risk from oral exposure to 1 mg/m^3 over a lifetime to lifetime cancer risk.

According to EPA, in 1985, the EPA created agency-wide workgroups, the carcinogen risk assessment verification endeavor workgroup (CRAVE) and RfC/RfD workgroup in order to reach agency consensus on scientific positions for health effects resulting from chronic oral/inhalation exposure (USEPA, 2015). IRIS was created to provide an internal consensus on the agency's positions in an internally accessible database regarding health effects. The database became available in 1988, and eventually available on the internet, now IRIS assessments use the (Health and Environmental Research Online) HERO database. In December 2015, IRIS released

its first multi-year agenda, identifying top priority chemical assessments with highest public health impacts for use in decision-making.

According to EPA, the IRIS process for developing human health risk assessments consists of: draft development, agency review, interagency science consultation, public comment & external peer review, revision, final agency review/interagency science discussion, and final assessment (USEPA, 2015). Draft development applies principles of systematic review for internal scoping, problem formulation of scientific questions, and applying the principles of systematic review to identify pertinent studies, evaluate study methods and quality, and derive toxicity values. Agency review involves review and revision of the draft assessment based on received comments. Interagency science consultation includes review of the draft assessment by federal agencies and revision based on received comments. The draft assessment is then released for public review and comment, and for external peer review. The assessment is revised according to peer review and public comments. The final agency review and interagency science discussion involves revision by EPA program offices, federal agencies, and the Executive Office. The final assessment is posted to the IRIS website at https://www.epa.gov/iris.

According to the NRC, problem formulation and protocol development are part of systematic reviews (2014). A broad literature search is used to identify evidence from human, animal, and mechanistic studies for an association between chemicals and health outcomes. Studies are evaluated and evidence is integrated on the basis of standardized approaches and guidelines. Finally, hazard identification and dose-response assessment are conducted, from which toxicity values are derived. RfC, RfD, and unit risks are used with exposure assessments to derive quantitative risk estimates. The IRIS assessment process is continuously undergoing updates and changes, as recommendations for future directions and improvements are made.

The IRIS assessment for inorganic Arsenic is currently being reassessed, with an updated assessment being developed for both cancer and noncancer health effects (USEPA, NCEA, 1988). The IRIS assessment for Arsenic was initially published in 1988, with major landmarks being the 1999 and 2001 NRC reports evaluating EPA drinking water standards for inorganic Arsenic and the 2010 release of a draft assessment for cancer health effects following oral exposure (USEPA, NCEA, 1988). Scoping and problem formulation began in 2012 for reassessment.

The most current (2002) IRIS chemical assessment summary (USEPA, NCEA, 1988) provides information regarding the oral RfD and carcinogenicity assessment. The oral RfD, an estimate of daily exposure likely to be without deleterious effects over a lifetime, is based on the assumption that thresholds exist for certain toxic effects such as cellular necrosis, expressed in mg/kg-day, and given with an uncertainty factor. For inorganic Arsenic (CASRN 7440-38-2), there was no clear consensus on the oral RfD among scientists. For the critical effects of hyperpigmentation, keratosis, vascular complications, and human chronic oral exposure, the NOAEL, LOAEL, and RfD were given as .009 mg/L converted to .0008 mg/kg-day, .17 mg/L converted to .014 mg/kg-day, and 3E-4 mg/kg-day of Arsenic, respectively. An external panel found that "it is clear from epidemiological studies that arsenic is a human carcinogen via oral and inhalation routes" and that "one important mode of action is unlikely to be operative for arsenic" (ERG, 1997). Weight-of-evidence evaluation found sufficient evidence from human carcinogenicity data; increased lung cancer mortality observed in multiple human populations exposed through inhalation, and increased mortality from multiple internal organ cancers (liver, kidney, lung, and bladder), and skin cancer, from populations consuming drinking water high in

Arsenic. There was no consistent demonstration of carcinogenicity in test animals for arsenic (USEPA, NCEA, 1988).

The second part of the carcinogenicity assessment, quantitative estimates of risk from oral and inhalation exposure, presents three risk estimates: oral slope factor, unit risk, and drinking water or air concentration (USEPA, NCEA, 1988). Drinking water or air concentration was reported in concentration (ug/L or ug/cu.m) for cancer risks of 1 in 10,000, 1 in 100,000, or 1 in 1,000,000. Finally, discussion of confidence of carcinogenicity and oral exposure for the studies discussed is included in the assessment.

Inorganic Arsenic is currently being reassessed and is in the draft development stage of assessment (USEPA, 2014a). Problem formulation for the toxicological review for cancer and noncancer effects follows an Assessment Development Plan (ADP), containing scoping information and assumptions from partners and stakeholders (USEPA, 2014b). Following the ADP, a literature search for health effects is conducted using natural language processing on the basis of similarities of titles and abstracts. Evidence from epidemiological and animal toxicity studies are summarized, and risk of bias evaluated. Finally, a literature search for mode of action information is conducted, and hypothesized Modes of Action summarized.

As defined by EPA, mode of action describes "a sequence of key events and processes, starting with the interaction of an agent with a cell, proceeding through operational and anatomical changes, and resulting in cancer formation [or other adverse outcomes]" (USEPA, 2005). MOA's form the basis for AOP development when data are sufficient. For IRIS assessments, qualitative MOA analyses are conducted to inform causal determinations for health effects. For causal health effects, mechanistic and susceptibility information is organized into an

AOP. These AOP's will then be used to inform dose-response analyses. If data are insufficient to support an AOP, then mode of action may be used to organize data (USEPA, 2014b).

Potential MOA's for Arsenic as identified in a preliminary draft assessment include: cytotoxicity and regenerative proliferation, generation of reactive oxygen species and depletion of antioxidant enzymes followed by oxidative stress, and alteration of epigenetic mechanisms such as DNA damage (USEPA, 2014b). In evaluating hypothesized MOA's for Arsenic, some considerations relevant across multiple MOA's are the metabolism of inorganic arsenic from pentavalent arsenic (As V) to trivalent metabolites (As III), and the oxidative methylation of trivalent species to pentavalent methylarsonic acid (MMA) and dimethylarsinic acid (DMA). Secondly, in looking at how Arsenic might affect populations at risk due to cumulative effects cotoxic and interactive effects should be examined, including life stage, nutrition, genetics, sex, pre-existing disease, smoking, alcohol consumption, and exposure to mixtures. Arsenic-induced health outcomes affect males and females differentially (NRC, 2013). Additionally, Arsenic has cotoxic effects with metals such as cadmium (Huang et al., 2009).

To support the toxicological review of Arsenic for draft development, EPA's NCEA is working with external organizations to develop AOP's related to several health outcomes for Arsenic, namely bladder cancer, basal cell carcinoma, renal cell carcinoma, squamous cell carcinoma, diabetes, and adverse pregnancy outcomes (I. Cote, personal communication, 2016) based on hypothesized MOA's (USEPA, 2014b). Initially, a causal determination for cancer (bladder, lung, skin, liver, and kidney) and noncancer (CVD, skin lesions, adverse pregnancy outcomes, diabetes, immune effects, neurodevelopmental effects, and respiratory) endpoints was conducted. AOP's will be constructed for those endpoints determined to be causal. AOP development will consist of integrating information regarding the pathophysiology of specific

diseases with arsenic specific data from literature or from NCBI BioSystems (http://www.ncbi.nlm.nih.gov/biosystems/) (I. Cote, personal communication, 2016). Cytoscape, a software package for integrating, visualizing, and analyzing network data (http://www.cytoscape.org/), will be used for developing and visualizing these AOP's.

Methods

The definition of an AOP has a defined starting point, the MIE, and ending point, the AO of regulatory significance (OECD, 2012a). In this study, low birth weight was selected as the AO, and the production of reactive oxygen species was selected as the MIE. Identification of intermediate KE's between the MIE and the AO was achieved by a literature review including preliminary assessments of inorganic Arsenic modes of action. Following the definition of KE's as intermediate steps at different levels of biological organization that are toxicologically relevant to the AO and experimentally quantifiable (OECD, 2012a). For development of an AOP describing adverse pregnancy outcomes, a literature review of publications and IRIS draft development materials for inorganic Arsenic was conducted (USEPA, 2014b). AOP development was conducted using the methods outlined in the Users' Handbook Supplement (OECD, 2013b). Consultation with experts was conducted in choosing the KE's in the AOP and in reviewing the ontology mapping chosen (R. Fry, personal communication, 2016). Ontology terms were mapped to the KE's of the adverse pregnancy outcome AOP's following the methods outlined in Chapter 3.

A graphical representation was used to summarize the data including the MIE, KE's and AO in a linear flow diagram (Ankley et al. 2010, OECD, 2012a). The graphical representation allows the visualization of the sequence of events in the AOP at different levels of biological organization (OECD, 2012a). Finally, AOP evaluation of newly established AOP's should be

conducted via the implementation of Bradford Hill criteria for weight of evidence assessment (Hill, 1965; OECD, 2012a) according to OECD guidelines (OECD, 2012a). Weight of evidence assessment was not conducted as part of this study.

Diagrams in Cytoscape of hypothesized modes of action for the outcomes under assessment were obtained from a collaborator (Source: Sciome, LLC). It was assumed that the events in MOA's obtained from the preliminary draft materials were synonymous with the key events of the draft AOP's created. Health outcomes being assessed were: bladder cancer, basal cell carcinoma, renal cell carcinoma, squamous cell carcinoma, and diabetes. Key events (nodes), relationships, and level of biological organization of events (biochemical, macromolecular, cellular/tissue, organism/population) were obtained from the diagrams. Using the final selection (Table 1) of ontologies and vocabularies chosen in Chapter 3, key event terms from the hypothesized MOA's were mapped to terms from the chosen ontologies. Query phrases representative of key events from these pages were entered into three ontology web browser sites: Ontobee (http://www.ontobee.org/), NCBO Bioportal (http://bioportal.bioontology.org/), and the EMBL-EBI Ontology Lookup Service (http://www.ebi.ac.uk/ols/beta/index). From the list of "hits" on the search browser output, matching ontology classes (entities) were reviewed for their match to query terms and the appropriate ontology class was chosen. When possible, the least specific class available to accurately portray the information was selected, to promote reusability of terms and minimize the number of classes chosen.

The data model described in Chapter 3, containing the fields: Process, Object, Action, and Context and the data properties as described in Chapter 3 were used in the mapping of ontology terms to KE terms. Results were compiled and analyzed using Excel. The mapping of key events to ontology terms was reviewed for accuracy by experts in the field.

Results

Case examples were chosen to illustrate the structured approach of the data model, and its utility to the AOP-KB. These examples were chosen to illustrate application of the AOP development process to create AOP's based on hypothesized MOA's, and because of the direct applicability of this set of hypothesized MOA's to risk assessment for a chemical known to cause human health effects. The case examples bring up discussion points of how the data model works and its limitations for specific Processes, Objects, Actions, and Contexts.

AOP Case Example 1: Low Birth Weight

Exposure to inorganic Arsenic during pregnancy has been associated with maternal and fetal outcomes, including preterm birth, low birth weight (full term infant <2.5kg), reduced postnatal growth and mortality defined as spontaneous abortion, stillbirth, and infant death (USEPA, 2014b).

Two AOP's were outlined for adverse pregnancy outcomes (Figure 8, Figure 9), with the shared AO of low birth weight with the difference being that one of the AOP's includes preterm birth as an AO. For the AO of low birth weight, occupational and environmental studies have shown an association between arsenic exposure through drinking water and low birth weight. Epidemiological studies have shown an association between arsenic exposure through drinking water during pregnancy and impaired fetal and infant growth and survival (NRC, 2013). Drinking water exposures in populations included high (>100 ug/L) exposure and lower exposures (up to 40 ug/L). Cross-sectional, cohort, and ecological studies (Ahamed et al., 2006; Ahmad et al., 2001; Chakraborti et al., 2003; Gelmann et al., 2013; Guan et al., 2012; Kippler et al., 2012; Kwok et al., 2006; Mukherjee et al., 2005; Rahman et al., 2005; Vall et al., 2012; Gardner et al., 2013; Hopenhayn et al., 2003; Huyck et al., 2007; Rahman et al., 2009; Saha et

al., 2012; McDermott et al., 2014; Aelion et al., 2012; Myers et al., 2010; Yang et al., 2003) show a consistent inverse association between maternal arsenic exposure and birth size, particularly birth weight. Two studies (Kwok et al. 2006; Myers et al. 2010) do not demonstrate an association with low birth weight.

Working backwards from the AO, the tissue/organ level KE was identified as 'Altered placental/fetal development and function.' Evidence cited in the literature includes impaired placental function, including poor placental development as a result of placental trophoblast migration; and impaired placental vasculogenesis leading to reduced nutritional uptake by fetus and reduced birth weight. Placental effects could be via gene changes; upregulation of sFLT1 in female cord blood, a protein that inhibits placental angiogenesis, upregulation of AQP9, encodes membrane transporter contributing to arsenic uptake, and decrease in ENPP2 associated with decreased birth weight. Outcomes of placental effects as a result of maternal arsenic exposure include impaired function of the placenta, impaired vasculogenesis and gene changes. The effect of arsenic on the developing fetus could be due to effects on the placenta or effects within the fetus itself; whether maternal arsenic is taken up by the placenta and the fetus is unclear. Studies of effects on the placenta found that placental trophoblast migration is reduced by arsenic, causing poor placental development (Li and LochCaruso, 2007). Other studies showed that arsenic impaired placental vasculogenesis in pregnant mice, which could reduce nutritional uptake by the fetus and lead to low birth weight (He et al. 2007; Coffin et al. 2006). Other studies (Remy et al., 2014) found an association between upregulation of soluble fms-like tyrosine kinase-1 (SFLT1) in human cord blood, a protein that inhibits placental angiogenesis. Fei et al. (2013) found an association between maternal arsenic exposure and placental upregulation of

aquaporin 9 (AQP9), a gene encoding a transporter that contributes to arsenic uptake. A related decrease in ENPP2 was associated with decrease in birth weight.

Working backwards, the next KE was determined to be "Altered gene expression of inflammation-related genes (NFKB, TNF, GCR) in placental/fetal cells" and "Altered gene expression of growth-related genes (KCNQ1) in placental/fetal cells." Studies have shown that iAs can act as an immunomodulatory agent *in utero*, causing the increased expression of proinflammatory genes in newborn cord blood (Ahmed et al., 2011; Fry et al., 2007). The authors found a correlation between arsenic exposure and increased expression of genes related to DNA damage and oxidative stress in cord blood, but did not find an association between these effects and pregnancy outcomes. Studies in human placentas have shown the increased expression of TNF-alpha and IFN-gamma (Ahmed et al., 2011) and increased intracellular H_2O_2 in placental cell lines (Massrieh et al., 2006). Another study showed a correlation between iAs metabolites in maternal urine (U-tAs) and increased TNF-related inflammatory proteins in newborn cord blood (Bailey et al., 2014). Another showed gene expression of disease-related genes in the glucocorticoid receptor pathway (Rager et al., 2014). Mo et al. found down regulation of n-channel genes CACNA1, KCNH2, KCNQ1 and KCNE1 in human cardiomyocytes by Arsenic exposure (2011).

Inflammation, DNA damage, and epigenetic alterations were chosen as the molecular level KE's for this AOP. Evidence for the oxidative stress MOA for Arsenic includes studies of the effects on mouse embryonic cells showing that oxidative stress causes impaired fetal growth. Toxicological studies on mouse embryonic cells show that Arsenic induced oxidative stress (Ren et al., 2014; Singh et al., 2010; Zhang et al., 2010), cell death, and DNA damage (Mirkes and Little, 1998); but specific pathways by which stress and DNA damage affect pre and

postnatal growth are not clear (USEPA, 2014b). At the molecular level, the MIE was chosen as "Production of Reactive Oxygen Species, Increased" based on this hypothesized oxidative stress MOA.

Bollati and Baccarelli (2010) and Jirtle and Skinner (2007) found an association between prenatal exposure to toxicants such as iAs and altered disease risk in adulthood, providing evidence for epigenetics through altered expression of key genes. Martin et al. found altered CpG methylation activated TGF-β-associated gene expression associated with preeclampsia (2015). Rojas et al. assessed changes in 5-methylcytosine methylation, and found that DNA methylation levels of genes were associated with birth outcomes, in a pregnancy cohort exposed prenatally to Arsenic in drinking water (2015). Howe et al. found sex-dependent correlations of altered post-translational histone modifications in rodents associated with arsenic levels in drinking water (2016).

Mapping the results of search queries for KE's in these pregnancy AOP's to ontology classes from the selected minimum list as described in Chapter 3 (Table 1), resulted in the mapping listed in Appendix C.

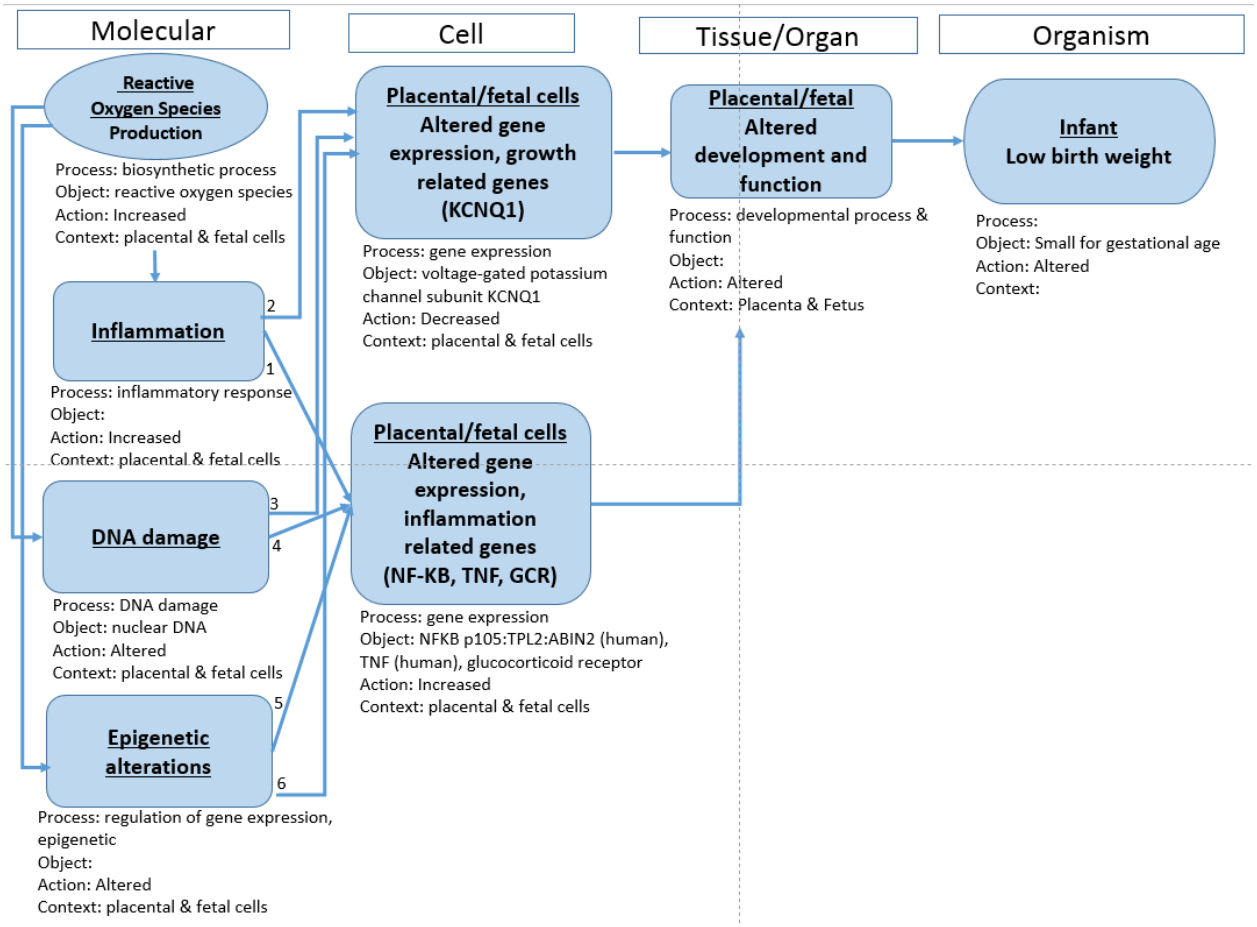

Figure 9. Graphical view of "Reactive Oxygen Species Production leading to Low Birth Weight" with KE terms mapped to ontology terms. & is used to represent instances in which more than one ontology class was used to represent a data entity.

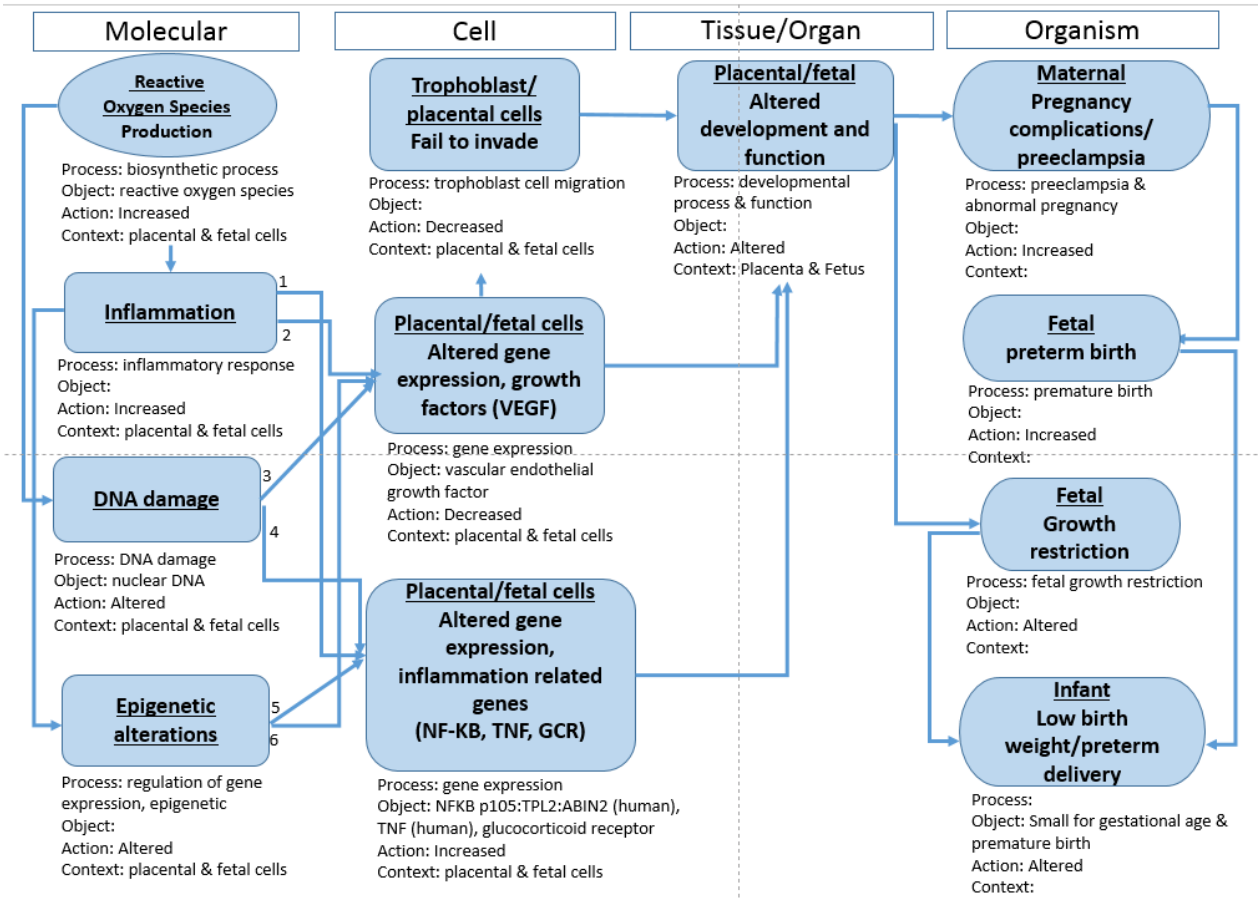

Figure 10. Graphical view of "Reactive Oxygen Species Production leading to Low Birth Weight/Preterm Delivery" with KE terms mapped to ontology terms. & is used to represent instances in which more than one ontology class was used to represent a data entity.

AOP Case Example 2: Low Birth Weight and Preterm Birth

KE's not shared by the two AOP's and unique to the second AOP with the AO of low birth weight/preterm delivery, include at the cellular level, 'Trophoblast cells fail to invade,' at the organism level, 'preeclampsia/pregnancy complications,' 'preterm birth,' and 'fetal growth restriction'. Li and Loch-Caruso found an association between Arsenic exposure and the reduction of placental trophoblasts, which could lead to poor placental development (2007). Oxidative stress, inflammation, and differential gene expression have been previously associated with preeclampsia (Ahmed et al., 2011). Laine et al. examined the interactive effects of placental Cd, Se, and Zn from a pregnancy cohort in the US in association with preeclampsia; and found that essential metals may play a protective effect in reducing the odds of Cd-associated

preeclampsia (2015). According to Harrington et al., "the exact etiology of conditions like preeclampsia and the effects of fetal exposure to toxic metals has not been determined, making the assessment of trace element levels crucial to the elucidation of the causes of conditions like preeclampsia (2016)." No study was found in this literature review for direct effects between prenatal arsenic exposure and preeclampsia.

Preterm birth, describing infants born alive before 37 weeks (WHO, 2015), is a common cause of infant morbidity and mortality. Inflammation during pregnancy, especially proinflammatory cytokines, are hypothesized to be associated with preterm birth but results are inconsistent (Lyon, D et al., 2010). Cross-sectional studies examined low birth weight and preterm birth/delivery (Ahamed et al., 2006; Chakraborti et al., 2003; Mukherjee et al., 2005; Rahman et al., 2005). Ahmad et al. (2001), found consistent positive observations across studies in regions with high environmental Arsenic in Bangladesh and India. Ecological studies (Aelion et al., 2012; Myers et al., 2010; Yang et al., 2003) on low birth weight and preterm birth associated with high arsenic in soil, drinking water, and well water, reported varying results.

Ahmed et al. details Arsenic exposure during pregnancy associated with oxidative stress and inflammation and leptin in the placenta, a hormone involved in transfer of nutrients from the placenta to the developing fetus (2011). Leptin expression is increased during hypoxia and intrauterine conditions such as diabetes, preeclampsia, and intrauterine growth retardation (Spranger et al. 2008; Iwagaki et al. 2004; Mise et al. 1998). "Fetal growth restriction" was the KE chosen to describe decreased growth at the organism level resulting from poor or altered placental development.

Mapping of ontology terms to KE's in Hypothesized MOA's for Arsenic

Mapping the results of search queries for KE's in the hypothesized MOA's for outcomes associated with iAs exposure (bladder cancer, basal cell carcinoma, renal cell carcinoma, squamous cell carcinoma, and diabetes) to ontology classes from the selected minimum list as described in Chapter 3 (Table 1), resulted in the of 100 KE's, 94 of them unique, being mapped to 300 ontology classes, 129 of them unique (Appendix C). There were 3 instances in which the query term from the KE could not be found using the ontology browsers chosen, and mappings these KE's were undefined or incomplete. Over half of ontology classes were duplicates, and only a small subsection of the total number of terms available in each ontology (Figure 2) was needed to accurately describe the chosen KE's.

For example, for the MOA for bladder cancer, KE "FGFR Activation" can be condensed into the query terms "gene expression," and "FGFR" from the text description, with the modifier "Increased" and in the context of a "urothelial cell", resulting in a mapping to ontology classes as:

Process: *'signaling'* (GO: 0023052)
Object: *'fibroblast growth factor receptor'* (PR: 00000134)
Context: *'urothelial cell'* (CL: 0000731)
Action: '*Increased*.'

Discussion

Development of AOP's describing adverse pregnancy outcomes (Figure 8, Figure 9) began with accumulating studies and evidence supporting hypothesized potential MOA's at different levels of organization. Following the AOP development guidelines outlined by OECD (OECD, 2013b), it was determined that two AOP's would best represent the AO's of low birth weight and that a separate AOP detailing a preeclampsia MOA resulting in low birth weight and associated preterm birth would be most representative. Working backward, KE's developed at

each level of organization were created with the intention of maximum overlap between the two AOP's, with the only difference being that the second AOP contained the KE's of 'Trophoblast cells fail to invade,' 'preeclampsia/pregnancy complications,' 'preterm birth,' and 'fetal growth restriction.' Oxidative stress resulting in the production of ROS was determined to be the MIE for both AOP's, based on evidence for oxidative stress MOA for iAs.

Evidence from primary literature was cited to support the KE's developed; however a full Weight of Evidence Assessment according to Bradford Hill Criteria (OECD, 2013b) would need to be conducted in order to fully assess confidence in these AOP's. For several of the KE's, evidence from studies were either insufficient or inconclusive; as the mechanisms of KE's and their relationships are not yet fully understood. Assessing evidence for these AOP's has involved the aggregation of epidemiological from case-control, cross-sectional, and ecological cohorts and animal toxicity study types. One limitation to providing direct evidence for the KE of 'Altered placental/fetal development and function,' for example, was that the majority of studies examined gene and protein levels in newborn cord blood and not the placenta or fetus itself. Evidence from different study types is also complicated by the variety of environmental mediums (well water, drinking water, and soil) studied for Arsenic exposure, and temporality of exposure (prenatal and postnatal).

Mapping ontology terms to KE's for both pregnancy outcomes and hypothesized MOA's for cancer and noncancer outcomes, led to a number of discussion points for the mapping. Several of the discussion points made included:

1) Representation of gene/protein families

For example, for the Event: "RAS Activation," the object chosen was: 'GTPase HRAS' (PR: PR_000029705). Using the OLS Browser (EBI-EMBL OLS), in the PR ontology, there is no

more general class of "RAS proteins" but the user must specify either hRAS, RRAS, or MRAS (EBI-EMBL OLS). Thus, in order to choose the more general class we chose to use MESH rather than PR, the default for gene expression, since MESH contained a class for 'ras Proteins' (MESH: MESH_D018631).

Process: '*signaling'* (GO: GO_0023052)
Object: '*ras Proteins'* (MESH: MESH_D018631)
Action: '*Increased.'*
Context: '*urothelial cell'* (CL_0000731)

2) Differentiation between gene expression and regulation of gene expression

It was necessary to differentiate between key events in which genes were activated, and in which a specific receptor or factor was regulating gene expression. Thus, 'gene expression' (GO: GO_0010467) was chosen for gene activation, while 'regulation of gene expression' (GO: GO_0010468) was chosen to represent regulation of gene expression by a specific receptor or factor. For the Event: "Loss of p53 function," the mapping was:

Process: '*regulation of gene expression'* (GO: GO_0010468)
Object: '*cellular tumor antigen p53'* (PR: PR_000003035)
Action: '*Decreased.'*
Context: '*basal cell'* (CL: CL_0000646)

3) Representation of pathways.

Although there is a specific ontology, the Pathway Ontology, for the representation of pathways, it was decided that pathway activation should be represented in terms of the individual factors or genes being activated. Thus, a key event of "PI3K AKT Activation" would be represented as separate genes:

Process: '*signaling'* (GO: GO_0023052)
Object: '*phosphatidylinositol 3-kinase complex'* (GO: GO_0005942) AND 'AKT Kinase' (PR: PR_000029189)
Action: '*Increased.'*
Context: '*urothelial cell'* (CL: CL_0000731)

4) *Representation of specific cell types within the 'Context' field.*

In mapping the pregnancy AOP's, cellular level KE's occurred in placental and fetal cells. A query for "placental cell" in the OLS Browser resulted in multiple placental cell types, including 'placental pericyte' (CL: CL_2000078), 'placental hematopoietic stem cell' (CL: CL_0002359), or 'placental epithelial cell' (CL: CL_0002577). A search for "fetal cell" similarly returned very specific fetal cell types, but no general parent class for "fetal cell." Generation of a mapping for this context necessitates a conversation with AOP authors about how to best define the cellular context for these events.

5) *Multiple ways to represent phenotype.*

For the KE "Altered placental/fetal development and function," the event could either be represented as two processes 'developmental process' (GO: GO_0032502) and 'function' (MI: MI_0613) in the context of the placenta/fetus, or the KE could be represented as a phenotype of 'abnormal placental development' (MP: MP_ MP_0001712). This illustrates that there are multiple ways to represent the same KE, and the need for restrictions to ensure that KE's are uniformly represented.

Process: '*developmental process'* (GO: GO_0032502) AND '*function'* (MI: MI_0613)
Object:
Action: '*Altered.'*
Context: '*placenta'* AND *'fetus'* (UBERON: UBERON_0001987, FMA: FMA_63919)

6) *Terms for which a query returned no results.*

The KE "Glucolipotoxicity" returned no results when input into any of the browser sites used. Follow up would consist of consulting the Arsenic MOA workgroup to find alternate terms for representing this process.

These discussion points highlight the utility of the data properties and restrictions discussed in Chapter 3, as well as future needs for the data model. The utility of the model in attaching ontology terms to KE's, thus making AOP's computable, is the benefit of using this structured approach to annotate KE's. In conducting these mappings, potential limitations and possible biases in the methods used should be considered as discussed in Chapter 3. Attaching evidence for these hypothesized MOA's in the form of assays and biomarkers, will inform the hazard identification and dose response steps of the IRIS assessment process.

CHAPTER 5: GENERAL DISCUSSION, LIMITATIONS, AND FUTURE DIRECTIONS

The AOP Framework provides a construct to organize mechanistic data supporting chemical regulatory assessment. It supports the use of *in vitro* HTS methods to increase the number of chemicals tested and reduce the cost and time of testing as well as the use of live animals. Additionally, it supports attaching evidence from both traditional toxicity tests and new data streams including data from genomics, chemoinformatics, pathway, phenomics, ontology, metabolomics, and toxicogenomics (Oki et al., 2016). It accomplishes all this by connecting the toxicity pathway as defined by the NRC (NRC, 2007) to adverse effects at the higher levels of biological organization (Ankley et al., 2010), which correspond with criteria for regulation of environmental pollutants.

The AOP development process accounts for AOP's with varying levels of documentation, ranging from AOP's resulting from a limited review of the scientific literature, to more formally documented AOP's resulting from an extensive review in which gaps in confidence are well documented (Villeneuve et al., 2014a). Thus, the development process accounts for uncertainties, inconsistencies and data gaps in supporting evidence.

The AOP-KB benefits knowledge management and computation on data supporting AOP's at all levels of confidence. As the modules of the AOP-KB evolve to meet the needs identified by the OECD, including open access, standardized representation of data, and consistency in reporting (Oki et al., 2016), the need for a shared set of chemical, biological, and toxicological ontologies emerges as a way to unify information across the AOP-KB and provide

computational capabilities for components that are currently not computable, supporting steps in the AOP development process.

Developing the data model of Process, Object, Action, and Context to extend the AOP-Ontology (Figure 4) provides computability for AOP entities that are currently represented in text form in the AOP-Wiki module of the AOP-KB. The data model 1) Provides the ability to attach scientific evidence in support of the AOP, including assays and biomarkers linked to key events. 2) Supports aggregation of data from HTS assays and experiments, supporting KE's. 3) Supports AOP networks, which can then be used for further AOP identification and evaluation (Oki et al., 2016). 4) Provides for evaluation of taxonomic relevance of KE's (Oki et al. 2016).

The properties of the data model, including restrictions and allowances described earlier, depend on its design. One desired property that arose in creating this model, was allowing multiple Events within one KE. For example, if a KE can have more than one Event, each with its inherent Process, Object, and Action, then a KE can have multiple Process, Object, and Action pairings. For example, for the KE of "Low birth weight/preterm delivery," implementing this property would allow 'Low birth weight' and 'Preterm delivery' to each have its own set of Process, Object, Action, and Context, while remaining part of the same KE as an intermediate event in a series of events leading to the AO. This allows more flexibility for the author of the AOP in determining the correct level of abstraction for each KE without artificial restrictions introduced by the need for a more structured description of the biology.

For the case examples analyzed here, multiple AOP's have the same KE's, promoting a network view in which KE's are nodes and relationships are edges (Villeneuve et al., 2014a). In other words, multiple "events" in an MOA collapse into one KE in an AOP. For instance, in the hypothesized MOA for bladder cancer as a result of iAs exposure, multiple events describe the

activation of genes through signaling pathways (FGFR, ERBB2, HRAS, RAS, VEGF, PI3K-AKT, MAPK, JAK-STAT). For the purposes of an AOP, a KE might be named "Altered gene expression of pro-inflammatory genes" with the Event mapping of Process, 'gene expression,' Object 'pro-inflammatory genes', and Action 'Altered.' Allowing more resolution would give the ability to specify the specific gene name and 'Increased' or 'Decreased' expression for the individual gene while retaining the properties of the Event. This proposed, more granular entity was named a "Sub-Event" and will be implemented in future versions of the data model. Thus, an Event can have multiple Sub-Events, each with its own Process, Object, and Action terms. The granularity of Key Event entities will proceed from least specific to most specific, from Key Event, to Event, to Sub-Event, to Process-Object-Action mapping.

Another desired property that arose in creating the data model was the placement of the Context data field. It was decided that Context should be a property of Key Events, so that a Key Event must have one and only one Context. Secondly, this property allows reuse of Events in other biological contexts. For example, for a hypothetical KE entitled 'activation of proinflammatory genes and proteins', the KE would occur in the context of 'urothelial cell' for the Bladder Cancer AOP, and in 'kidney cell' for the Renal Cell Carcinoma AOP. Thus, having the Event not tied to Context, allows Events to be reused in different Contexts. If the Context is part of a Key Event, then its properties are the same as the KE and independently editable by the author of the individual KE.

These properties differentiate between the user-driven, AOP Framework view of AOP's and their relationships and components, and the computer language needed to interpret these relationships and components. The data model will contribute to the computational models underlying all phases of AOP development (Villeneuve et al., 2014a). Computationally predicted

AOP's, as the basis for expert-driven AOP's (Oki et al., 2016) which are then reviewed and used for regulatory decision-making, require a structured data approach to tie the AOP Framework to computable tools and applications. Future applications of the semantics of this model to the AOP-Wiki application within the AOP-KB will determine the database structure of AOP components and the properties of that structure. The choice of how ontology terms populate fields on the web page that users enter and edit, and how these terms populate from one entity to another, will be determined by the developer's design. As new releases are made and the Wiki is updated, having set properties will ease the addition of new capabilities to the Wiki.

Application of the data model to case studies of existing high-priority AOP's within AOP-Wiki, as well as to AOP's supporting a draft IRIS assessment for inorganic Arsenic, demonstrate the applicability of the model to these use cases. Shared biological, chemical, and toxicological ontologies will be useful not only to the Wiki, but across all components of the AOP-KB, providing computability to support AOP development for regulatory decision-making. Addressing the needs for the future of toxicity as outlined by the NRC's 2007 report (Ankley et al., 2010) promoting the use and accessibility of the resources in the AOP-KB, and participation in the AOP development process by both regulators and scientists, will inform testing strategies for adverse outcomes of regulatory significance.

APPENDIX A: REVIEW OF ONTOLOGIES AND CONTROLLED VOCABULARIES FOR THEIR CHARACTERISTICS

Data Source	Domain	Taxon	Source Location	Level of Biological Organization*					
				M	C	T	O	I	P
OBO Foundry Ontologies									
Anatomical Entity Ontology (AEO)	anatomy	all kingdoms	https://github.com/obophenotype/human-developmental-	X	X	Y	Y	X	X
Basic Formal Ontology (BFO)*	OBO Foundry		http://ifomis.org/bfo/						
Beta Cell Genomics (BCGO)	beta cell genomics studies	mouse (Mus)	https://github.com/obi-bcgo/bcgo	Y	Y	X	X	X	X
biological collections ontology (BCO)	biodiversity data		https://github.com/tucotuco/bco						
BRENDA tissue/enzyme source (bto)*	enzyme source	organisms	http://www.brenda-enzymes.info	X	Y	Y	Y	X	X
cardiovascular disease ontology (cvdo)	health		https://code.google.com/p/cvdo/	X	X	X	Y	Y	X
Cell Line Ontology (CLO)	in vitro cell line		http://www.clo-ontology.org	X	Y	X	X	X	X
Cell Ontology (CL)	cells	animals	http://cellontology.org/	X	Y	Y	Y	X	X
Chemical Entities of Biological Interest (CHEBI)	biochemistry		http://www.ebi.ac.uk/chebi	Y	X	X	X	X	X
clinical measurement ontology (cmo)	clinical	human (Homo sapiens)	http://rgd.mcw.edu/rgdweb/ontology/search.html	X	X	X	Y	Y	X
Common Anatomy Reference Ontology (CARO)*	anatomy		https://github.com/obophenotype/caro	X	X	Y	Y	X	X
eagle-I resource ontology (ERO)*	resources		http://code.google.com/p/eagle-i/						
environment ontology (EnVO)	environment		http://environmentontology.org/	X	X	X	X	X	Y
FlyBase Controlled Vocabulary	phenotype	Drosophila	http://purl.obolibrary.org/obo/fbcv	X	X	X	X	Y	X
Foundational Model of Anatomy (FMA)	anatomy	human (Homo sapiens)	http://sig.biostr.washington.edu/projects/fm/index.html	X	X	Y	Y	X	X
gazeteer (GAZ)	geography		http://gensc.org/gc_wiki/index.php/GAZ_Project	X	X	X	X	X	Y
GO	biology	all kingdoms	http://purl.obolibrary.org/obo/go.owl	Y	Y	X	X	X	X
Human developmental anatomy, abstract (ehdaa2)	anatomy	human-developmental	http://genex.hgu.mrc.ac.uk/	X	X	Y	Y	X	X
Human Disease Ontology (DOID)	disease	human (Homo sapiens)	http://diseaseontology.sourceforge.net	X	X	X	Y	Y	X
Human Phenotype Ontology (HP)	phenotype	human (Homo sapiens)	http://www.human-phenotype-ontology.org/	X	X	Y	Y	Y	X
Information Architect Ontology (IAO)	information		https://github.com/information-artifact-ontology/IAO/						
Mammalian phenotype (MP)	phenotype	mammals	http://www.informatics.jax.org/searches/MP_form.shtml	X	X	Y	Y	Y	X
Molecular Process Ontology (MOP)	molecular process		http://purl.obolibrary.org/obo/mop.owl	Y	X	X	X	X	X
Mouse Anatomy (MA)	anatomy	mouse (Mus)	https://github.com/obophenotype/mouse-anatomy-ontology	X	X	Y	Y	X	X
Mouse gross anatomy and development, abstract (emapa)	anatomy	mouse (Mus)	http://emouseatlas.org/	X	X	Y	Y	X	X
Mouse gross anatomy and development, timed (emap)	anatomy	mouse (Mus)	http://emouseatlas.org/	X	X	Y	Y	X	X
NCBI Taxon	taxonomy	all kingdoms	https://github.com/obophenotype/ncbitaxon	X	X	X	X	X	Y
Neuro Behavior Ontology (NBO)	behavioral phenotypes	organisms	http://code.google.com/p/behavior-ontology	x	x	x	x	Y	x
NCI Thesaurus (NCIt)	cancer		https://cabig.nci.nih.gov/concepts/EVS	Y	Y	Y	Y	Y	X
Ontology for Biomedical Investigations (OBI)	experiments		http://obi-ontology.org	Y	X	X	X	X	X
ontology for general medical science (ogms)	medicine	human (Homo sapiens)	http://code.google.com/p/ogms/	X	X	X	Y	Y	X
ontology of biological attirbutes (oba)	phenotype	all kingdoms	https://github.com/obophenotype/bio-attribute-ontology	X	X	X	Y	Y	X
Ontology of genes and genomes (ogg)	genes and genomes of biological organisms	organisms	http://ogg.googlecode.com/	Y	X	X	X	X	X
Pathway Ontology (PW)	pathways	rat(Rattus)	http://rgd.mcw.edu/rgdweb/ontology/search.html	Y	X	X	X	X	X
Phenotypic Quality (PATO)	phenotype		https://github.com/pato-ontology/pato/	X	X	Y	Y	Y	X
Population and Community Ontology (PCO)	populations and communities		https://github.com/PopulationAndCommunityOntology/pco	X	X	X	X	X	Y
Protein modification (PSI-MOD)	protein modification		http://www.psidev.info/MOD	Y	X	X	X	X	X
Protein Ontology (PRO)	proteins		http://proconsortium.org	Y	X	X	X	X	X
Protein-protein interaction (mi)	experiments		http://psidev.sf.net/	Y	Y	X	X	X	X
rat strain (rs)	rat strain	rat(Rattus)	http://rgd.mcw.edu/rgdweb/search/strains.html	X	X	X	X	Y	X
Relation Ontology (RO)*	relations		https://github.com/oborel/obo-relations/						
RNA Ontology (RnaO)	RNA sequence		https://github.com/bgsu-rna/rnao	Y	X	X	X	X	X
Sequence types and features (SO)	biological sequence	all	https://github.com/The-Sequence-Ontology/SO-Ontologies	Y	X	X	X	X	X
software ontology (swo)	software		www.ebi.ac.uk/efo/swo						
Uberon	anatomy	animals- vetrebrate	http://uberon.org	X	X	Y	Y	X	X
units of measurement (UO)	units of measure		https://github.com/bio-ontology-research-group/unit-ontology						
Vertebrate Trait (VT)	vertebrate trait	vertebrates	http://purl.obolibrary.org/obo/vt.owl	X	X	Y	Y	X	X

Data Source	Domain	Taxon	Source Location	Level of Biological Organization*					
				M	C	T	O	I	P
Xenopus anatomy and development	anatomy	Xenopus laevis	http://www.xenbase.org/anatomy/xao.do?method=display	X	X	X	Y	Y	X
Zebrafish anatomy and development	anatomy		http://zfin.org/zf_info/anatomy/dict/sum.html	X	X	X	Y	Y	X
Non-OBO Foundry									
AOP-Ontology	adverse outcome pathways		https://github.com/DataSciBurgoon/aop-ontology	Y	Y	Y	Y	Y	Y
BAO	biological assay		http://bioassayontology.org/	Y	X	X	X	X	X
basic formal ontology (bfo)	upper level		http://ifomis.org/bfo/						
Experimental Factor Ontology (EFO)*	experiments		http://www.ebi.ac.uk/efo/	X	X	Y	Y	Y	X
OpenTox Toxicology	toxicological endpoint		http://www.opentox.org/dev/Ontology	Y	X	X	X	X	X
OpenTox Organs and Effects	organ system	rodents	http://www.opentox.org/dev/Ontology	X	X	X	Y	X	X
Controlled Vocabulary									
UMLS/Medical Subject Headings (MeSH)	biomedical information	human (Homo sapiens)	https://www.nlm.nih.gov/mesh	Y	Y	Y	Y	Y	Y
International Classification of Diseases, Clinical Modification (ICD9CM)	clinical		http://www.cms.hhs.gov/	X	X	Y	Y	X	X
Systematized Nomenclature of Medicine- Clinial Terms (SNOMEDCT)	clinical		http://ihtsdo.org/	X	X	X	Y	Y	X
Common Terminology Criteria for Adverse Events (NCI CTAE)	cancer drugs		http://purl.bioontology.org/ontology/CTCAE	X	X	X	X	Y	X
Medical Dictionary for Regulatory Activities (MedDRA)	clinically validated international medical terminology dictionary (and thesaurus) used by regulatory authorities in the pharmaceutical industry during the regulatory process		http://www.meddra.org/	Y	Y	Y	Y	Y	Y
Human Interaction Network Ontology (HINO)	INO extension for the domain of human interaction networks.		http://purl.bioontology.org/ontology/HINO	X	X	X	X	X	X
Online Mendelian Inheritance in Man (OMIM)	Health/Disease, traits and phenotypes		http://purl.bioontology.org/ontology/OMIM	Y	X	X	X	X	X
WikiPathways	biological pathway		http://www.wikipathways.org/index.php/WikiPathways	Y	X	X	X	X	X
FlyBase	Drosophila		flybase.org	Y	X	Y	Y	Y	X

*Molecular, Cellular, Tissue, Organ, Individual, Population
Y = definitely covers this level of organization.
X = definitely doesn't covers this level of organization
*upper level data source

APPENDIX B: MAPPING OF AOP'S TO ONTOLOGY CLASSES FROM THE MINIMUM LIST IN TABLE 1

AOP	Long name	Level of Organisatio	Key event name	Process	Ontology_ID	Object	Ontology_ID	Action	Context	Ontology_ID
25	**Aromatase inhibition leading to reproductive dysfunction (in fish)**	Molecular	Aromatase, inhibition	catalytic activity	GO_0003824	cytochrome P450 19A1	PR_000006100	Decreased	granulosa cell	CL_0000501
		Cellular	17beta-estradiol synthesis by ovarian granulosa cells, Reduction	biosynthetic process	GO_0009058	17beta-estradiol	CHEBI_16469	Decreased	granulosa cell	CL_0000501
		Organ	concentrations, Reduction	blood circulation	GO_0008015	17beta-estradiol	CHEBI_16469	Decreased	blood plasma	UBERON_0001969
		Cellular	Transcription and translation of vitellogenin in liver, Reduction	gene expression	GO_0010467	vitellogenins	MESH_D014819	Decreased	hepatocyte	CL_0000182
				translation	GO_0006412					
		Organ	Plasma vitellogenin concentrations, Reduction	blood circulation	GO_0008015	vitellogenins	MESH_D014819	Decreased	blood plasma	UBERON_0001969
		Cellular	Vitellogenin uptake into oocytes and oocyte growth/development,	receptor-mediated endocytosis	GO_0006898	vitellogenins	MESH_D014819	Decreased	ooctye	CL_0000023
		Organism	Reduction	fecundity	PATO_0000273	egg	UBERON_0007379	Decreased		
		Population	Population trajectory, Decrease	population growth rate	PCO_0000008			Decreased		
16	**Acetylcholinesterase inhibition leading to acute mortality**	Molecular	Acetylcholinesterase (AchE), Inhibition	catalytic activity	GO_0003824	acetylcholinesterase	PR_000003626	Decreased	synapse	GO_0045202
		Molecular	Accumulation	accumulation	PATO_0002269	acetylcholine	CHEBI_15355	Increased	synapse	GO_0045202
		Organ	Increased	atrioventricular block	MP_0010519			Altered	cardiac muscle tissue	UBERON_0001133
				heart rate	EFO_0004326			Decreased	cardiac muscle tissue	UBERON_0001133
		Organ	Respiratory distress/arrest, Increased	respiratory distress	MP_0001954			Increased	respiration organ	UBERON:0000171
		Organ	Induction	ataxia	MP_0001393			Increased	skeletal muscle tissue	UBERON_0001134
				induced hyperactivity	MP_0001399			Increased	skeletal muscle tissue	UBERON_0001134
				paralysis	MP_0000753			Increased	skeletal muscle tissue	UBERON_0001134
		Organism	Mortality, Increased	mortality	EFO_0004352			Increased		
		Organism	Population trajectory, Decreased	population growth rate	PCO_0000008			Decreased		
22	**AHR1 activation leading to developmental abnormalities and embryolethality (in birds)**	Molecular	AHR, Activation	signaling	GO_0023052	aryl hydrocarbon receptor	PR_000003858	Increased	hepatoctye, cardiac muscle cell	CL_0000182, CL_0000746
		Molecular	UDP-GT, Up Regulation	catalytic activity	GO_0003824	UDP-glucuronosyltransferase 1-9	PR_000017056	Increased	hepatoctye, cardiac muscle cell	CL_0000182, CL_0000746
		Molecular	Regulation	catalytic activity	GO_0003824	glutathione S-transferase	PR_000013420	Increased	hepatoctye, cardiac muscle cell	CL_0000182, CL_0000746
		Tissue	Thyroxine (T4) in serum, Decreased	blood circulation	GO_0008015	thyroxine	CHEBI_30660	Decreased	serum	UBERON_0001977
		Tissue	Retanoids, Decreased	blood circulation	GO_0008015	retinoids	CHEBI_26537	Decreased	serum	UBERON_0001977
		Molecular	AHR nuclear translocator (ARNT)-dependent pathways, Altered	protein dimerization activity	GO_0046983	AHR	PR_000003858	Altered	hepatoctye, cardiac muscle cell	CL_0000182, CL_0000746
						ARNT	PR_000004303	Altered	hepatoctye, cardiac muscle cell	CL_0000182, CL_0000746
		Tissue	Oxidative Stress, Increase			oxidative stress	MP_0003674	Increased	hepatoctye, cardiac muscle cell	CL_0000182, CL_0000746
		Cellular	Arachadonic acid epoxide production, Increase	arachidonic acid metabolic process	GO_0019369	epoxyicosatrienoic acid	CHEBI_64007	Increased	hepatoctye, cardiac muscle cell	CL_0000182, CL_0000746
		Molecular	CYP1A5, Up Regulation	gene expression	GO_0010467	cytochrome p-450 cyp1a5	MESH_C108015	Increased	hepatoctye, cardiac muscle cell	CL_0000182, CL_0000746
		Cellular	Uroporphyrinogen oxidation, Increase	oxidation	MOP_0000568	uroporphyrinogen	CHEBI_27258	Increased	hepatoctye, cardiac muscle cell	CL_0000182, CL_0000746
		Organ	Altered	developmental process	GO_0032502	cardiovascular system	UBERON_0004535	Altered	Heart	UBERON_0000947
				abnormal cardiovascular system physiology	MP_0001544	cardiovascular system	UBERON_0004535	Altered	Heart	UBERON_0000947
		Organism	Uroporphyria, n/a	porphyria	MP_0005654	porphyrin	CHEBI_8337	Altered	liver	UBERON_0002107
		Organism	Embryotoxicity, N/A	embryonic lethality	MP_0008762	embryo	UBERON_0000922	Increased		

AOP	Long name	Level of Organisation	Key event name	Process	Ontology_ID	Object	Ontology_ID	Action	Context	Ontology_ID
27	**Cholestatic Liver Injury induced by Inhibition of the Bile Salt Export Pump (ABCB11)**	Molecular	Bile Salt Export Pump (ABCB11), Inhibition	transport	GO_0006810	bile salt export pump	PR_O95342	Decreased	hepatocyte	CL_0000182
		Cellular	Transcriptional change, Activation of specific nuclear receptors	gene expression	GO_0010467	bile acid receptor	PR_000011396	Increased	hepatocyte	CL_0000182
						nuclear receptor subfamily 1 group I r	PR_000011397			
						nuclear receptor subfamily 1 group I r	PR_000011398			
		Cellular	Pathological condition, Bile accumulation	accumulation	PATO_0002269	bile	UBERON_0001970	Increased	bile canaliculi	UBERON_0001283
		Cellular	Cytokine, Release	secretion	GO_0046903	cytokine	FMA_84050	Increased	hepatocyte	CL_0000182
		Cellular	Inflammation, Increase	inflammatory response	GO_0006954			Increased	hepatocyte	CL_0000182
		Tissue	Reactive oxygen species, Production	biosynthetic process	GO_0009058	reactive oxygen species	CHEBI_26523	Increased	hepatocyte	CL_0000182
		Cellular	Oxidative Stress, Increase			oxidative stress	MP_0003674	Increased	hepatocyte	CL_0000182
		Organism	Pathology, Cholestasis	cholestasis	HP_0001396			Increased	liver	UBERON_0002107
46	**AFB1: Mutagenic Mode-of-Action leading to Hepatocellular Carcinoma (HCC)**	Tissue	Pro-mutagenic DNA Adducts, Formation	DNA alkylation	GO_0006305	nuclear DNA	FMA_67194	Altered	hepatocyte	CL_0000182
		Organism	Production of Reactive Electrophiles, Metabolism of AFB1	metabolic process	GO_0008152			Increased	liver	UBERON_0002107
		Cellular	of DNA Adducts, Insufficient or Mis-repair	dna repair	GO_0006281	nuclear DNA	FMA_67194	Increased	hepatocyte	CL_0000182
		Molecular	Induced in Critical Genes, Mutation	mutation	MI_0118	cellular tumor antigen p53	PR_000003035	Altered	hepatocyte	CL_0000182
		Cellular	to form Altered Hepatic Foci (AHF), Clonal Expansion/Cell Proliferation	cell proliferation	GO_0008283	Hepatocyte	CL_0000182	Altered	liver	UBERON_0002107
		Organism	Hepatocellular carcinoma, Tumorigenesis			hepatocellular carcinoma	EFO_0000182	Increased	liver	UBERON_0002107
30	**Estrogen receptor antagonism leading to reproductive dysfunction**	Molecular	Estrogen receptor, Antagonism	signaling	GO_0023052	estrogen receptor	PR_000007204	Decreased	hepatocyte	CL_0000182
		Tissue	Transcription and translation of vitellogenin in liver, Reduction	gene expression	GO_0010467	vitellogenins	MESH_D014818	Decreased	hepatocyte	CL_0000182
				translation	GO_0006412					
		Organ	Plasma vitellogenin concentrations, Reduction	blood circulation	GO_0008015	vitellogenins	MESH_D014819	Decreased	blood plasma	UBERON_0001969
		Tissue	Vitellogenin uptake into oocytes and oocyte growth/development, Reduction	receptor-mediated endocytosis	GO_0006898	vitellogenins	MESH_D014819	Decreased	ooctye	CL_0000023
		Organism	Cumulative fecundity and spawning, Reduction	fecundity	PATO_0000273			Decreased		
				spawning	NCIT_C82476			Decreased		
		Population	Population trajectory, Decrease	population growth rate	PCO_0000008			Decreased		
15	**Alkylation of DNA in male pre-meiotic germ cells leading to heritable mutations**	Molecular	DNA, Alkylation	DNA alkylation	GO_0006305		CHEBI_16991	Increased	male germ cells	CL_0000015
		Molecular	Mutations, Increase	mutation	MI_0118	DNA	CHEBI_16991	Increased	male germ cells	CL_0000015
		Cellular	Insufficient or incorrect DNA repair, N/A	dna repair	GO_0006281	DNA	CHEBI_16991	Altered	male germ cells	CL_0000015
		Organism	Heritable mutations in offspring, Increase	mutation	MI_0118	DNA	CHEBI_16991	Increased		

AOP	Long name	Level of Organisation	Key event name	Process	Ontology_ID	Object	Ontology_ID	Action	Context	Ontology_ID
23	**Androgen receptor agonism leading to reproductive dysfunction**	Molecular	Androgen receptor, Agonism	signaling	GO_0023052	androgen receptor	PR_000004191	Increased	theca cell	CL_0000503
		Cellular	Testosterone synthesis by ovarian theca cells, Reduction	biosynthetic process	GO_0009058	testosterone	CHEBI_17347	Decreased	theca cell	CL_0000503
		Cellular	17beta-estradiol synthesis by ovarian granulosa cells, Reduction	biosynthetic process	GO_0009058	17beta-estradiol	CHEBI_16469	Decreased	granulosa cell	CL_0000502
		Organ	Plasma 17beta-estradiol concentrations, Reduction	blood circulation	GO_0008015	17beta-estradiol	CHEBI_16469	Decreased	blood plasma	UBERON_0001969
		Organ	Transcription and translation of vitellogenin in liver, Reduction	gene expression	GO_0010467	vitellogenins	MESH_D014819	Decreased	hepatocyte	CL_0000182
				translation	GO_0006412					
		Organism	Cumulative fecundity and spawning, Reduction	fecundity	PATO_0000273			Decreased		
		Tissue	Plasma vitellogenin concentrations, Reduction	blood circulation	GO_0008015	vitellogenins	MESH_D014819	Decreased	liver	UBERON_0002107
		Tissue	Vitellogenin uptake into oocytes and oocyte growth/development, Reduction	receptor-mediated endocytosis	GO_0006898	vitellogenins	MESH_D014819	Decreased	ooctye	CL_0000023
		Organism	Gonadotropins, circulating concentrations, Reduction	blood circulation	GO_0008015	luteinizing hormone, follicle stimulating hormone	FMA_74642, CHEBI_	Decreased	pituitary gland	UBERON_0000007
		Population	Population trajectory, Decrease	population growth rate	PCO_0000008			Decreased		
6	**Antagonist-binding causing stabilization of co-repressor (SMRT or N-CoR) to PPARalpha Ligand Binding Domain causing downstream starvation-like body-weight loss**	Molecular	Antagonist-binding causing stabilization of co-repressor (SMRT or N-CoR) to PPARalpha ligand binding domain, Binding as antagonist	signaling	GO_0005102	PPAR-alpha	PR_000013056	Decreased	nucleus	GO_0005634
		Molecular	PPARalpha transactivation of gene expression, Decreased	regulation of gene expression	GO_0010468	PPAR-alpha	PR_000013056	Decreased	nucleus	GO_0005634
		Organ	Mitochondrial Fatty Acid Beta Oxidation, Decreased	fatty acid beta-oxidation	GO_0006635	fatty acid	CHEBI_35366	Decreased	liver	UBERON_0002107
		Organ	Ketogenesis (production of ketone bodies), Decreased	ketone biosynthetic process	GO_0042181	ketone body	CHEBI_73693	Decreased	liver	UBERON_0002107
		Tissue	Circulating Ketone Bodies, Not Increased	blood circulation	GO_0008015	ketone body	CHEBI_73693	Altered	blood plasma	UBERON_0001969
		Organ	Peroxisomal Fatty Acid Beta Oxidation of Fatty Acids, Decreased	fatty acid beta-oxidation	GO_0006635	fatty acid	CHEBI_35366	Decreased	peroxisome	GO_0005777
		Organ	Catabolism of Muscle Protein, Increased	catabolism	GO_0009056	muscle protein	MESH_D009124	Increased		
		Organism	Body Weight, Decreased	muscle atrophy	GO_0014889	skeletal muscle tissue	UBERON_0001134	Increased		
				body weight loss	EFO_0005245			Increased		
38	**Protein Alkylation leading to Liver Fibrosis**	Molecular	Protein, Alkylation	protein alkylation	GO_0008213	protein	CHEBI_16541	Increased	hepatocyte	CL_0000182
		Cellular	Cell death, N/A	cell death	GO_0008219	hepatocyte	CL_0000182	Increased	hepatocyte	CL_0000182
		Cellular	Hepatic macrophages (Kupffer Cells), Activation	macrophage activation	GO_0042116	Kupffer cell	CL_0000091	Increased	hepatocyte	CL_0000182
		Cellular	TGFbeta1 expression, Up Regulation	gene expression	GO_0010467	TGF-beta	PR_000000046	Increased	parenchymal cell	FMA_84638
		Cellular	Stellate cells, Activation	cell activation	GO_0001775	hepatic stellate cell	CL_0000632	Increased	hepatocyte	CL_0000182
		Tissue	Collagen, Accumulation	collagen metabolic process	GO_0032963	collagen	CHEBI_3815	Increased	liver	UBERON_0002107
		Organ	Liver fibrosis, N/A	fibrosis	MESH_D005355	extracellular matrix structural constituent	GO_0005201	Increased	liver parenchyma	UBERON_0001280

AOP	Long name	Level of Organisation	Key event name	Process	Ontology_ID	Object	Ontology_ID	Action	Context	Ontology_ID
48	**Binding of agonists to ionotropic glutamate receptors in adult brain causes excitotoxicity that mediates neuronal cell death, contributing to learning and memory impairment.**	Molecular	Inotropic glutamate receptors, Binding of agonist	signaling	GO_0005102	ionotropic glutamate re	GO_008328	Increased	neuron	CL_0000540
		Cellular	Mitochondrial dysfunction, N/A	biosynthetic process	GO_0009058	reactive oxygen species	CHEBI_26523	Increased	mitochondria	GO_0005739
		Cellular	Cell death, N/A	cell death	GO_0008219			Increased	neuron	CL_0000540
		Tissue	Neurodegeneration, N/A	neurodegeneration	MP_0002229	neuron	CL_0000540	Increased	brain	UBERON_0000955
		Molecular	NMDARs, Overactivation	signaling	GO_0005102	NMDA selective glutam	GO_0017146	Increased	hippocampus	UBERON_0002421
		Cellular	Calcium influx, Increased	store-operated calcium entry	GO_0002115	calcium(2+)	CHEBI_29108	Increased	cell	CL_0000000
		Tissue	Neuronal network function in adult brain, Decreased	synaptic transmission	GO_0007268	neural network	FMA_74616	Decreased	brain	UBERON_0000955
		Tissue	Neuroinflammation, N/A	cell activation	GO_0001775	glial cell	CL_0000125	Increased	brain	UBERON_0000955
		Organism	Learning and memory, Impairment	abnormal learning/memory/conditioning	MP_0002063			Altered		
13	**Chronic binding of antagonist to N-methyl-D-aspartate receptors (NMDARs) during brain development induces impairment of learning and memory abilities**	Molecular	NMDARs, Inhibition	signaling	GO_0005102	NMDA selective glutam	GO_0017146	Decreased	hippocampus	UBERON_0002421
		Cellular	Calcium influx, Decreased	transport	GO_0006810	calcium(2+)	CHEBI_29108	Decreased	neuron	CL_0000540
		Molecular	Release of BDNF, Reduced	gene expression	GO_0005102	brain-derived neurotro	PR_000004716	Decreased	neuron	CL_0000540
		Cellular	Dendritic morphology, Aberrant	dendrite development	GO_0009058			Altered	dendrite	GO_0030425
		Molecular	Presynaptic release of glutamate, Reduced	neurotransmitter secretion	GO_0008219	Glutamate	CHEBI_14321	Decreased	blood plasma	UBERON_0001969
		Cellular	Cell death, N/A	cell death	MP_0002230			Increased	neuron	CL_0000540
		Cellular	Synaptogenesis, Decreased	synapse assembly	GO_0005103	Synapse	GO_0045202	Decreased	neuron	CL_0000540
		Organ	Neuronal network function, Decreased	synaptic transmission	GO_0012421			Decreased	brain	UBERON_0000955
		Organism	Learning and memory, Impairment	abnormal learning/memory/conditioning	GO_0017574			Altered	brain	UBERON_0000955
40	**Covalent Protein binding leading to Skin Sensitisation**	Molecular	Protein, Covalent Binding	protein binding	GO_0005515	electrophilic reagant	CHEBI_59739	Increased	skin cell	FMA_84783
		Cellular	Keratinocytes, Activation	cell activation	GO_0001775			Increased	keratinocyte	CL_0000312
		Cellular	Dendritic Cells, Activation	cell activation	GO_0001775			Increased	dendritic cell	CL_0000451
		Tissue	T-cells, Activation/Proliferation	cell activation	GO_0001775	T-cell	CL_0000084	Increased	lymph node	UBERON_0000029
				cell proliferation	GO_0008283	T-cell	CL_0000084	Increased	lymph node	UBERON_0000029
		Organ	skin, sensitisation	allergic contact dermatitis	MESH_D017449				skin of body	UBERON_0002097

AOP	Long name	Level of Organisation	Key event name	Process	Ontology_ID	Object	Ontology_ID	Action	Context	Ontology_ID
18	**PPARa activation in utero leading to impaired fertility in males**	Molecular	PPAR alpha, Activation	signaling	GO_0023052	PPAR-alpha	PR_000013056	Increased	Leydig cell	UBERON_0005212
		Cellular	Steroidogenic acute regulatory protein (STAR), Decrease	gene expression	GO_0010467	STAR	PR_000015715	Decreased	Leydig cell	UBERON_0005212
				accumulation	PATO_0002269	STAR	PR_000015715	Decreased	Leydig cell	UBERON_0005212
		Cellular	Translator protein (TSPO), Decrease	gene expression	GO_0010467	translocator protein	PR_000016757	Decreased	Leydig cell	UBERON_0005212
				accumulation	PATO_0002269	translocator protein	PR_000016757	Decreased	Leydig cell	UBERON_0005212
		Cellular	Cholesterol transport in mitochondria, Reduction	transport	GO_0006810	cholesterol	CHEBI_16113	Decreased	Leydig cell	UBERON_0005212
		Cellular	Testosterone synthesis , Reduction	biosynthetic process	GO_0009058	testosterone	CHEBI_17347	Decreased	Leydig cell	UBERON_0005212
		Tissue	testosterone level , Reduction	blood circulation	GO_0008015	testosterone	CHEBI_17347	Decreased	Testes	UBERON_0000473
		Organ	Male reproductive tract, Malformation	malformation	PATO_0000646	Male reproductive syste	UBERON_0000079	Altered	Male reproductive system	UBERON_0000079
		Organism	Fertility, impaired	fertility	PATO_0000274			Decreased		
7	**PPARγ activation leading to impaired fertility in adult female**	Molecular	PPAR gamma, Activation	signaling	GO_0023052	PPAR-gamma	PR_000013058	Increased	granulosa cell	CL_0000501
		Cellular	17beta-estradiol synthesis by ovarian granulosa cells, Reduction	biosynthetic process	GO_0009058	17beta-estradiol	CHEBI_16469	Decreased	granulosa cell	CL_0000501
		Cellular	Aromatase (Cyp19a1), reduction in ovarian granulosa cells	gene expression	GO_0010467	cytochrome P450 19A1	PR_000006100	Decreased	granulosa cell	CL_0000501
				accumulation	PATO_0002269					
		Organ	Plasma 17beta-estradiol concentrations, Reduction	blood circulation	GO_0008015	17beta-estradiol	CHEBI_16469	Decreased	blood plasma	UBERON_0001968
		Organism	Fertility, impaired	fertility	PATO_0000274			Decreased		
		Organism	ovarian cycle, irregularities	ovulation cycle	GO_0042698			Altered		
42	**Xenobiotic Induced Inhibition of Thyroperoxidase and Subsequent Adverse Neurodevelopmental Outcomes in Mammals**	Molecular	Thyroperoxidase, Inhibition	catalytic activity	GO_0003824	thyroid peroxidase	PR_000016584	Decreased	thyrocyte	BTO_0003736
		Molecular	Thyroid hormone synthesis, Decreased	biosynthetic process	GO_0009058	thyroid hormone	CHEBI_60311	Decreased	thyrocyte	BTO_0003736
		Organ	Thyroxin (T4) in neuronal tissue, Decreased	blood circulation	GO_0008015	thyroxine	CHEBI_30660	Decreased	brain	UBERON_0000955
		Tissue	Thyroxin (T4) in serum, Decreased	blood circulation	GO_0008015	thyroxine	CHEBI_30660	Decreased	serum	UBERON_0001977
		Tissue	Hippocampal gene expression, Altered	regulation of gene expression	GO_0010468			Altered	hippocampus	UBERON_0002421
		Tissue	Hippocampal anatomy, Altered	abnormality of brain morphology	HP_0012443			Altered	hippocampus	UBERON_0002421
		Tissue	Hippocampal function, Decreased	synaptic transmission	GO_0007268			Decreased	hippocampus	UBERON_0002421
		Organism	Cognitive Function, Decreased	abnormal learning/memory/conditioning	MP_0002063			Altered		
10	**Binding to the picrotoxin site of ionotropic GABA receptors leading to epileptic seizures**	Molecular	iGABAR chloride channel, Binding at picrotoxin site	signaling	GO_0023052	GABA receptor comple	GO_1902710	Increased	neuron	CL_0000540
		Cellular	Chloride conductance, Reduction	transport	GO_0006810	Chloride ion	CHEBI_17996	Decreased	neuron	CL_0000540
		Tissue	A paroxysmal depolarizing shift, Occurrence	membrane depolarization	GO_0051899			Altered	brain	UBERON_0000955
		Cellular	Neuronal synaptic inhibition, Reduction	negative regulation of synaptic transmission	GO_0050805	GABA-A receptor compl	GO_1902711	Decreased	neuron	CL_0000540
		Tissue	Amplified excitatory postsynaptic potential, Generation	excitatory postsynaptic potential	GO_0060079			Increased	brain	UBERON_0000955
		Organism	Epileptic seizure, Occurrence	epilepsy	EFO_0000474			Increased		

Under Development										
AOP	Long name	Level of Organisation	Key event name	Process	Ontology_ID	Object	Ontology_ID	Action	Context	Ontology_ID
57	**AhR activation leading to hepatic steatosis**	Molecular	AHR, Activation	signaling	GO_0023052	aryl hydrocarbon recep	PR_000003858	Increased	hepatoctye, cardiac muscle	CL_0000182, CL_0000746
		Cellular	VLDL secretion, Suppression	secretion	GO_0046903	very-low-density lipopr	CHEBI_39027	Decreased	hepatocyte	CL_0000182
		Organ	Mitochondrial fatty acid beta-oxidation, Inhibition	fatty acid beta-oxidation	GO_0006635	fatty acid	CHEBI_35366	Decreased	liver	UBERON_0002107
		Organ	Fatty acid, Accumulation	accumulation	PATO_0002269	fatty acid	CHEBI_35366	Increased	liver	UBERON_0002107
		Cellular	PCK1 expression (control point for glycolysis/gluconeogenesis pathway), Decreased	gene expression	GO_0010467	PCK1	MGI_97501	Decreased	hepatocyte	CL_0000182
		Cellular	Triglyceride, Accumulation	accumulation	PATO_0002269	triglyceride	CHEBI_17855	Increased	hepatocyte	CL_0000183
		Molecular	CD36, Up Regulation	gene expression	GO_0010467	Cd36	MGI_107899	Increased	hepatocyte	CL_0000183
		Molecular	SCD-1, Increased	gene expression	GO_0010467	Scd1	MGI_107899	Increased	hepatocyte	CL_0000183
		Cellular	FA Influx, Increased	transport	GO_0006810	fatty acid	CHEBI_35366	Increased	hepatocyte	CL_0000183
		Molecular	LDLR (low density lipoprotein receptor), Up Regulation	gene expression	GO_0010467	low density lipoprotein	PR_000009744	Increased	hepatocyte	CL_0000183
		Cellular	LDL uptake, Increased	endocytosis	GO_0006897	low density lipoprotein	CHEBI_39026	Increased	hepatocyte	CL_0000183
		Molecular	CYP1A1, Up Regulation	gene expression	GO_0010467	Cyp1a1	MGI_88588	Increased	hepatocyte	CL_0000183
		Organ	Liver lipid, Accumulation	hepatic steatosis	MP_0002628			Increased	liver	UBERON_0002107
29	**Estrogen receptor agonism leading to reproductive dysfunction**	Molecular	Estrogen receptor, Agonism	signaling	GO_0023052	estrogen receptor	PR_000007204	Increased		
		Organism	Cumulative fecundity and spawning, Reduction	fecundity	PATO_0000273			Decreased		
		Tissue	Plasma vitellogenin concentrations, Increase	blood circulation	GO_0008015	vitellogenins	MESH_D014819	Increased	blood plasma	UBERON_0001969
		Organism	Vitellogenin synthesis in liver, Increase	gene expression	GO_0010467	vitellogenins	MESH_D014819	Increased	hepatocyte	CL_0000182
				translation	GO_0006412					
		Organ	Renal pathology due to VTG deposition, Increase	kidney disease	EFO_0003086			Increased	kidney	UBERON_0002113
		Population	Population trajectory, Decrease	population growth rate	PCO_0000008			Decreased		
		Organism	Reproductive behaviour, Altered	reproductive behavior	GO_0019098			Altered		
		Organism	Larval development, Altered	larval development	GO_0002164			Altered		
		Organism	Reproductive organs, Impaired development of	developmental process	GO_0032502	reproductive organ	UBERON_0003133	Altered		
4	**Ecdysone receptor (EcR) activation leading to mortality in *Daphnia magna***	Molecular	Ecdysone receptor, Activation	signaling	GO_0023052	ecdysone receptor holo	GO_0008230	Increased	epidermal cell	CL_0000362
		Organism	Pre-mature molting, Induction	molting cycle process	GO_0022404			Altered	dermal skeleton	UBERON_0010364
		Organ	Chitin synthesis, Induction	chitin biosynthetic process	GO_0006031	chitin	CHEBI_17029	Increased	dermal skeleton	UBERON_0010364
		Organ	Chitin degradation and resorption, Induction	chitin degradation	GO_0006032	chitin	CHEBI_17029	Increased	dermal skeleton	UBERON_0010364
		Population	Population, Decline	population growth rate	PCO_0000008			Decreased		
		Organism	Fecundity, Reduction	fecundity	PATO_0000273			Decreased		
		Organism	Mortality, Increase	mortality	EFO_0004352			Increased		
28	**Cyclooxygenase inhibition leading to reproductive failure**	Molecular	Cyclooxygenase activity, Inhibition	catalytic activity	GO_0003824	coproporphyrinogen-III	PR_000005826	Decreased		
		Tissue	Prostaglandin E2 concentration, Reduction	blood circulation	GO_0008015	prostaglandin E2	CHEBI_606564	Decreased	blood plasma	UBERON_0001969
		Tissue	Ca and HCO3 transport to shell gland, Reduction	transport	GO_0006810	calcium ion	CHEBI_29108	Decreased	blood plasma	UBERON_0001969
						hydrogen carbonate ior	CHEBI_17544			
		Tissue	Eggshell thickness, Reduction	chorion-containing eggshell formation	GO_0007304			Altered	ovarian follicle cell	CL_0000477
		Tissue	Gap, N/A							
		Organism	Reproductive failure, N/A	reproduction	GO_0000003			Altered		

APPENDIX C: MAPPING OF HYPOTHESIZED AOP'S TO ONTOLOGY CLASSES FROM THE MINIMUM LIST IN TABLE 1 FOR ARSENIC HEALTH OUTCOMES

Health Out-come	Level of Organi-sation	Key Event	Process	Ontology_ID	Object	Ontology_ID	Action	Cellular/ Organ Context	Ontology _ID
Low Birth Weight	Macro-molecular	Reactive Oxygen Species, Production	biosynthetic process	GO_0009058	reactive oxygen species	CHEBI_ 26523	Increased	placental cell, fetal cell*	
	Macro-molecular	Inflammation	inflammatory response	GO_0006954			Increased	placental cell, fetal cell*	
	Macro-molecular	DNA damage	DNA damage	MESH_ D004249	nuclear DNA	FMA_67194	Altered	placental cell, fetal cell*	
	Macro-molecular	Epigenetic alterations	regulation of gene expression, epigenetic	GO_ 0040029			Altered	placental cell, fetal cell*	
	Cell ular	Altered gene expression, growth related genes (KCNQ1)	gene expression	GO_0010467	voltage-gated potass-ium channel subunit KCNQ1	PR_0000007 28	Decreased	placental cell, fetal cell*	
	Cell ular	Altered gene expression, inflammation related genes (nfkb, tnf, gcr)	gene expression	GO_0010467	NFKB p105:TPL2:A BIN2 (human)	PR_0000284 44	Increased	placental cell, fetal cell*	
			gene expression	GO_0010467	Tumor necrosis factor alpha (human)	PR_P01375	Increased	placental cell, fetal cell*	

Health Out-come	Level of Organi-sation	Key Event	Process	Ontology_ID	Object	Ontology_ID	Action	Cellular/ Organ Context	Ontology _ID
			gene expression	GO_0010467	glucocorticoid receptor	PR_000011406	Increased	placental cell, fetal cell*	
	Organ	Altered development and function	developmental process	GO_0032502			Altered	placenta, fetus	UBERON_0001987, FMA_63919
			function	MI_0613			Altered	placenta, fetus	UBERON_000198, FMA_63919
	Organism	Low birth weight			small for gestational age	HP_0001518	Altered	placental cell, fetal cell*	
Low Birth Weight / pre term delivery	Macro molecular	Reactive Oxygen Species	biosynthetic process	GO_0009058	reactive oxygen species	CHEBI_26523	Increased	placental cell, fetal cell*	
	Macro molecular	Inflammation	inflammatory response	GO_0006954			Increased	placental cell, fetal cell*	
	Macro molecular	DNA damage	DNA damage	MESH_D004249	nuclear DNA	FMA_67194	Altered	placental cell, fetal cell*	

Health Out-come	Level of Organi-sation	Key Event	Process	Ontology_ID	Object	Ontology_ID	Action	Cellular/ Organ Context	Ontology _ID
	Macro molecu lar	Epigenetic alterations	regulation of gene expression, epigenetic	GO_0040029			Altered	placental cell, fetal cell*	
	Cell ular	Fail to invade	regulation of gene expression, epigenetic	GO_0061450			Decreased	placental cell, fetal cell*	
	Macro molecu lar	Altered gene expression, growth factors	gene expression	GO_0010467	Vascular endothelial growth factor	EFO_ 0003276	Decreased	placental cell, fetal cell*	
	Macro molecu lar	Altered gene expression, inflammation related genes	gene expression	GO_0010467	NFKB p105:TPL2:A BIN2 (human)	PR_0000284 44	Increased	placental cell, fetal cell*	
			gene expression	GO_0010467	Tumor necrosis factor alpha (human)	PR_P01375	Increased	placental cell, fetal cell*	
			gene expression	GO_0010467	Glucocorti coid receptor	PR_0000114 06	Increased	placental cell, fetal cell*	
	Organ	Altered development and function	developmental process	GO_0032502			Altered	placenta, fetus	UBERON_ 0001987, FMA_639 19
			function	MI_0613			Altered	placenta, fetus	UBERON_ 0001987,

Health Out-come	Level of Organi-sation	Key Event	Process	Ontology_ID	Object	Ontology_ID	Action	Cellular/ Organ Context	Ontology _ID
									FMA_639 19
	Organ ism	Preeclampsia/p reg nancy complications			preeclampsia	HP_0100602	Increased		
					abnormal pregnancy	MP_ 0009661	Increased		
	Organ ism	Preterm birth			premature birth	HP_0001622	Increased		
		Fetal growth restriction			fetal growth restriction	EFO_000049 5	Altered		
	Organ ism	Low birth weight/preterm delivery			small for gestational age	HP_0001518	Altered		
					premature birth	HP_0001622	Altered		
Bladde r Cancer	Macro mole cular	Reactive Oxygen Species	biosynthetic process	GO_0009058	reactive oxygen species	CHEBI_ 26523	Increased	urothelial cell	CL_00007 31
	Macro molecu lar	FGFR_Activatio n	signaling	GO_0023052	fibroblast growth factor receptor	PR_ 00000134	Increased	urothelial cell	CL_00007 31
	Macro molecu lar	ErbB2_Activatio n	signaling	GO_0023052	receptor tyrosine-protein kinase erbB-2	PR_0000020 82	Increased	urothelial cell	CL_00007 31

Health Out-come	Level of Organi-sation	Key Event	Process	Ontology_ID	Object	Ontology_ID	Action	Cellular/ Organ Context	Ontology _ID
	Macro molecu lar	hRAS_Activatio n	signaling	GO_0023052	GTPase Hras	PR_0000297 05	Increased	urothelial cell	CL_00007 31
	Macro molecu lar	RAS_Activation	signaling	GO_0023052	ras Proteins	MESH_D018 631	Increased	urothelial cell	CL_00007 31
	Macro molecu lar	p53_Mutation	mutation	MESH_D0091 54	cellular tumor antigen p53	PR_0000030 35	Altered	urothelial cell	CL_00007 31
	Macro molecu lar	VEGF_Activatio n	signaling	GO_0023052	vascular endothelial growth factor A	PR_0000172 84	Increased	urothelial cell	CL_00007 31
	Macro molecu lar	P13K_AKT_Acti vation	signaling	GO_0023052	phosphatidyli nositol 3-kinase complex	GO_ 0005942	Increased	urothelial cell	CL_00007 31
			signaling	GO_0023052	AKT kinase	PR_0000291 89	Increased		
	Macro molecu lar	MAPK_Activatio n	catalytic activity	GO_0003824	mitogen-activated protein kinase	PR_0000000 19	Increased	urothelial cell	CL_00007 31
	Macro molecu lar	MMP_Stimulati on	catalytic activity	GO_0003824	Matrix Metalloprote inases	MESH_D020 782	Increased	urothelial cell	CL_00007 31
	Macro molecu lar	CDKN2A	gene expression	GO_0010467	CDKN2A gene translation product	PR_0000290 97	Increased	urothelial cell	CL_00007 31

Health Out-come	Level of Organi-sation	Key Event	Process	Ontology_ID	Object	Ontology_ID	Action	Cellular/ Organ Context	Ontology _ID
	Macro molecu lar	COX-2	gene expression	GO_0010467	prostaglandi n G/H synthase 2	PR_0000134 28	Increased	urothelial cell	CL_00007 31
	Macro molecu lar	Rb1_Mutation	mutation	MESH_D0091 54	retinoblasto ma-associated protein	PR_0000137 73	Altered	urothelial cell	CL_00007 31
	Macro molecu lar	JAK-STAT_Activatio n	signaling	GO_0023052	janus kinase	PR_0000257 48	Increased	urothelial cell	CL_00007 31
			signaling	GO_0023052	STAT dimer	PR_0000279 35			
	Cell ular_ Tissue	Growth_Factor _and_Cytokine _Activation	signaling	GO_0023052	cytokine	FMA_84050	Increased	urothelial cell	CL_00007 31
	Cell ular_ Tissue	Cell_Proliferatio n	cell proliferation	GO_0008283	bladder tumor	EFO_ 0000294		urotheliu m	UBERON_ 0000365
	Cel lular_ Tissue	Angiogenesis_S urvival_Metast asis	angiogenesis	GO_0001525			Increased	urinary bladder	UBERON_ 0001255
			cell survival	MESH_D0024 70			Increased		
			neoplasm metastasis	MESH_D0093 62			Increased		
	Organis m_Pop ulation	Bladder Cancer							

Health Out-come	Level of Organi-sation	Key Event	Process	Ontology_ID	Object	Ontology_ID	Action	Cellular/ Organ Context	Ontology _ID
	Molecu lar	Thiol Binding, Skin Accumulation	binding	GO_0005488	thiol group	CHEBI_ 29917	Increased	basal cell	CL_00006 46
	Cell ular_ Tissue		accumulation	PATO_ 0002269			Increased	skin of body	UBERON_ 0002097
Basal Cell Carci noma	Macro-molecu lar	Reactive Oxygen Species	biosynthetic process	GO_0009058	reactive oxygen species	CHEBI_ 26523	Increased	urothelial cell	CL_00007 31
	Organis m_Pop ulation	Basal Cell Carcinoma			Basal Cell Carcinoma	MPATH_234	Increased		
	Macro molecu lar	Signal Transduction Pathway Crosstalk	cell commun ication	GO_0007154			Altered	basal cell	CL_00006 46
	Macro molecu lar	SMO/Su(Fu) Mutations	mutation	MI_0118	SUFU	PR_ 000003324	Increased	basal cell	CL_00006 46
	Macro molecu lar	Loss of p53 Function	regulation of gene expression	GO_0010468	cellular tumor antigen p53	PR_0000030 35	Decreased	basal cell	CL_00006 46
	Macro molecu lar	PTCH1 Mutations	mutation	MI_0118	PTCH1	PR_0000134 12	Increased	basal cell	CL_00006 46
Renal Cell Carci noma	Cell ular_ Tissue	Cell Proliferation	cell proliferation	GO_0008283	neoplasm	EFO_ 0000616	Increased	kidney	UBERON_ 0002113

Health Out-come	Level of Organi-sation	Key Event	Process	Ontology_ID	Object	Ontology_ID	Action	Cellular/ Organ Context	Ontology _ID
	Organism_Population	Renal Cell Carcinoma			Renal Cell Carcinoma	EFO_ 0000681	Increased		
	Cellular_ Tissue	Cell Proliferation and Survival	cell proliferation	GO_0008283	neoplasm	EFO_ 0000616	Increased	kidney	UBERON_ 0002113
			cell survival	MESH_D002470			Increased	kidney	UBERON_ 0002113
	Macromolecular	S6K	gene expression	GO_0010467	ribosomal protein S6 kinase beta-1	PR_000000123	Increased	kidney cell	CL_10000497
	Macromolecular	mTOR:Raptor	gene expression	GO_0010467	regulatory-associated protein of mTOR	PR_000013735	Increased	kidney cell	CL_10000497
	Macromolecular	Rheb	gene expression	GO_0010467	GTP-binding protein Rheb	PR_000013975	Increased	kidney cell	CL_10000497
	Macromolecular	TSC2/1	gene expression	GO_0010467	testis-expressed sequence 37 protein	PR_000016710	Increased	kidney cell	CL_10000497
	Macromolecular	PI3K	gene expression	GO_0010467	phosphatidylinositol 3-kinase complex	GO_ 0005942	Increased	kidney cell	CL_10000497
	Macromolecular	PTEN	gene expression	GO_0010467	phosphatidylinositol 3,4,5-trisphosphat	PR_000028746	Increased	kidney cell	CL_10000497

Health Out-come	Level of Organi-sation	Key Event	Process	Ontology_ID	Object	Ontology_ID	Action	Cellular/ Organ Context	Ontology _ID
					e 3-phosphatase and dual-specificity protein phosphatase PTEN				
	Macro molecu lar	ERK	gene expression	GO_0010467	ephrin type-B receptor 2	PR_0000071 30	Increased	kidney cell	CL_10000 497
	Macro molecu lar	Akt	gene expression	GO_0010467	AKT kinase	PR_0000291 89	Increased	kidney cell	CL_10000 497
	Cell ular_ Tissue	Angiogenesis	angiogenesis	GO_0001525	blood vessel	UBERON_00 01981	Increased	kidney interstitiu m	UBERON_ 0005215
	Cell ular_ Tissue	Cell Migration and Invasion	cell migration	GO_0016477	neoplasm	EFO_ 0000616	Increased	kidney interstitiu m	UBERON_ 0005215
			neoplasm invasiveness	MESH_D0093 61					
	Cell ular_ Tissue	Loss of ECM	extracellular structure organization	GO_0043062	intercellular matrix	FMA_9672	Altered	kidney cell	CL_10000 497
	Cell ular_ Tissue	Cystogenesis			cystic	PATO_0001 673	Increased	kidney interstitiu m	UBERON_ 0005215
	Cell ular_ Tissue	Dysregulated Signaling	regulation of signaling	GO_0023051			Decreased	kidney cell	CL_10000 497

Health Out-come	Level of Organi-sation	Key Event	Process	Ontology_ID	Object	Ontology_ID	Action	Cellular/ Organ Context	Ontology _ID
	Macro molecu lar	VEGF PGF	gene expression	GO_0010467	vascular endothelial growth factor	EFO_ 0003276	Increased	kidney cell	CL_10000 497
			gene expression	GO_0010467	placental growth factor	PR_0000126 05	Increased	kidney cell	CL_10000 497
	Macro molecu lar	MMP 2 MMP9	gene expression	GO_0010467	MMP2	PR_0000104 79	Increased	kidney cell	CL_10000 497
			gene expression	GO_0010467	MMP9	PR_0000104 91	Increased	kidney cell	CL_10000 497
	Cell ular_ Tissue	Loss of Primary Cilia			primary cilium	GO_ 0072372	Decreased	kidney interstitiu m	UBERON_ 0005215
	Macro molecu lar	HIF Translocation and Signaling Activation	translocation	SO_0000199	hypoxia-inducible factor 1-alpha	PR_0000085 55	Altered	kidney cell	CL_10000 497
			signaling	GO_0023052	hypoxia-inducible factor 1-alpha	PR_0000085 55	Increased		
	Macro molecu lar	Loss of VHL			loss of function variant	SO_0002054	Altered	kidney cell	CL_10000 497
	Cell ular_ Tissue	Cell Differentiation & Oncogenesis	cell differentiatio n	GO_0030154			Altered	squamous epithelium	UBERON_ 0006914

Health Out-come	Level of Organi-sation	Key Event	Process	Ontology_ID	Object	Ontology_ID	Action	Cellular/ Organ Context	Ontology _ID
					tumorigenesis	MP_ 0002006	Increased		
Squamous Cell Carcinoma	Macro molecular	cMyc	gene expression	GO_0010467	c-myc protein	PR_000000084		squamous cell	FMA_66769
	Macro molecular	Activation of STAT Pathways	STAT cascade	GO_0097696			Increased	squamous cell	FMA_66769
	Organism_Population	Squamous Cell Carcinoma			Squamous Cell Carcinoma	MP_ 0004207	Increased		
	Macro molecular	Loss of Rb	gene expression	GO_0010467	retinoblastoma-associated protein	PR_000013773	Decreased		
	Cellular_ Tissue	Cell Survival/Evasion of Apoptosis	cell survival	MESH_D002470			Increased	squamous cell	FMA_66769
			negative regulation of apoptotic process	GO_0043066			Increased	squamous cell	FMA_66769
	Macro molecular	PI3K/Akt	signaling	GO_0023052	phosphatidylinositol 3-kinase complex	GO_ 0005942	Increased		
					AKT kinase	PR_000029189			

Health Out-come	Level of Organi-sation	Key Event	Process	Ontology_ID	Object	Ontology_ID	Action	Cellular/ Organ Context	Ontology _ID
	Macro molecu lar	RAF/Mek/Erk	signaling	GO_0023052	RAF proto-oncogene serine/threo nine-protein kinase	PR_0000032 44	Increased	squamous cell	FMA_667 69
					dual specificity mitogen-activated protein kinase 1	PR_0000101 25	Increased		
					ephrin type-B receptor 2	PR_0000071 30	Increased		
	Macro molecu lar	Activation of Ras Pathways	signaling	GO_0023052	ras Proteins	MESH_D018 631	Increased	squamous cell	FMA_667 69
	Macro molecu lar	Loss of p16	gene expression	GO_0010467	CDKN2A gene translation product	PR_0000290 97	Decreased	squamous cell	FMA_667 69
	Macro molecu lar	Loss of p53	gene expression	GO_0010467	cellular tumor antigen p53	PR_0000030 35	Decreased	squamous cell	FMA_667 69
	Cell ular_ Tissue	Actinic Keratoses	actinic keratosis	EFO_0002496			Increased	epithelium	UBERON_ 0000483
	Macro molecu lar	C to T or CC to TT mutations	C_to_T_transi tion	SO_1000011			Increased	squamous cell	FMA_667 69

Health Out-come	Level of Organi-sation	Key Event	Process	Ontology_ID	Object	Ontology_ID	Action	Cellular/ Organ Context	Ontology _ID
	Macro molecu lar	Polymorphisms in p53 (codon 72) or H-ras (codon 27)	substitution	SO_1000002	cellular tumor antigen p53	PR_0000030 35	Increased	squamous cell	FMA_667 69
					GTPase Hras	PR_0000297 05			
Diabet es	Organis m_Pop ulation	Diabetes			diabetes mellitus	EFO_ 0000400	Increased		
	Cell ular_ Tissue	Beta Cell Exhaustion	abnormal cell physiology	MP_0005621			Altered	beta cell	FMA_857 04
	Individ ual	Insulin Resistance			insulin resistance	EFO_ 0002614	Increased		
	Individ ual	Systemic Proinflammator y Response	inflammatory response	GO_0006954			Increased		
	Macro molecu lar	Increased TNF-alpha			tumor necrosis-factor alpha	PR_0000001 34	Increased	beta cell	FMA_857 04
	Cell ular_ Tissue	Beta Cell Apoptosis	type B pancreatic cell apoptotic process	GO_0097050			Increased	beta cell	FMA_857 04
	Cell ular_ Tissue	Impaired FA Oxidation Ectopic Lipid Deposition and Lipotoxicity*	fatty acid oxidation	GO_0019395			Altered	beta cell	FMA_857 04

Health Out-come	Level of Organi-sation	Key Event	Process	Ontology_ID	Object	Ontology_ID	Action	Cellular/ Organ Context	Ontology _ID
			accumulation	PATO_ 0002269	fatty acid	CHEBI_ 35366	Increased	beta cell	FMA_857 04
	Macro molecu lar	Decreased Insulin Signaling	signaling	GO_0023052	insulin	PR_0000090 54	Decreased	beta cell	FMA_857 04
	Cell ular_ Tissue	Beta Cell Dysfunction	abnormal cell physiology	MP_0005621			Altered	beta cell	FMA_857 04
	Macro molecu lar	Increased IL-6, CRP	gene expression	GO_0010467	interleukin-6	PR_0000013 93	Increased	beta cell	FMA_857 04
			gene expression	GO_0010467	C-reactive protein	PR_0000058 97	Increased	beta cell	FMA_857 04
	Cell ular_ Tissue	Altered Skeletal Muscle Mitochondria Beta Oxidation and Mitochondria Number	fatty acid beta-oxidation	GO_0006635	Mitochon drion	GO_ 0005739	Altered	skeletal muscle tissue	UBERON_ 0001134
					abnormal mitochon drion morphology	MP_ 0006035	Altered	skeletal muscle tissue	UBERON_ 0001134
	Cell ular_ Tissue	Accumulation and Activation of Macrophages	accumulation	PATO_ 0002269	macrophage	CL_0000235	Increased	skeletal muscle tissue	UBERON_ 0001134
			cell activation	GO_0001775					

Health Out-come	Level of Organi-sation	Key Event	Process	Ontology_ID	Object	Ontology_ID	Action	Cellular/ Organ Context	Ontology _ID
	Cell ular_ Tissue	Oxidative and ER Stress			oxidative stress	MP_ 0003674	Increased	beta cell	FMA_857 04
			response to ER stress	GO_0034976	endoplasmic reticulum	GO_0 005783	Altered	beta cell	FMA_857 04
	Cell ular_ Tissue		Amyloid Deposition	MPATH_34			Increased	skeletal muscle tissue	UBERON_ 0001134
	Cell ular_ Tissue	Dysregulation of Mitochondrial Oxidation of Beta Fatty Acids	regulation of fatty acid beta-oxidation	GO_0031998	beta-amino-fatty acid	CHEBI_ 59754	Decreased	skeletal muscle tissue	UBERON_ 0001134
	Init iator**	Obesity	obesity	EFO_0001073			Increased		
	Cell ular_ Tissue	Glucoliptoxicity *							
	Cell ular_ Tissue	Ectopic Lipid Deposition in Skeletal Muscle and Liver*							
	Cell ular_ Tissue	Expansion of Adipose Tissue	adipose tissue development	GO_0060612			Increased	pancreas	UBERON_ 0001264
	Macro molecu lar	Alterations DNA Methylation or Histone Methylation	dna methylation	GO_0006306			Altered	beta cell	FMA_857 04

Health Out-come	Level of Organi-sation	Key Event	Process	Ontology_ID	Object	Ontology_ID	Action	Cellular/ Organ Context	Ontology _ID
			histone methylation	GO_0016571			Altered	beta cell	FMA_857 04
	Init iator**	High Fat Diet	high fat diet	EFO_0002757					

*undefined ** these were labeled as initiating factors (not a level of organization of the AOP) as part of hypothesized MOA's

REFERENCES

17beta-estradiol synthesis by ovarian granulosa cells, Reduction. Retrieved from https://aopwiki.org/wiki/index.php/Event:3

(ECETOC), E. C. f. E. a. T. o. C. (2007). *Intelligent Strategies in Ecotoxicology: Mode of action approach for specifically acting chemicals.* Retrieved from Brussels, Belgium:

(IFOMIS), I. f. F. O. a. M. I. S. (2015). Basic Formal Ontology (BFO): Home. Retrieved from http://ifomis.uni-saarland.de/bfo/

(MGD), M. G. D. (2016). term detail: abnormal cardiovascular system morphology. 6.02. Retrieved from http://www.informatics.jax.org

(NCBI), N. C. f. B. I. (2015). PubChem RDF Release Notes. V 1.5.2beta. Retrieved from https://pubchem.ncbi.nlm.nih.gov/rdf/

(NCBO), N. C. f. B. O. (2016a). Gene Ontology. Retrieved from http://purl.bioontology.org/ontology/

Abeyruwan, S., Vempati, U. D., Kucuk-McGinty, H., Visser, U., Koleti, A., Mir, A., . . . Schurer, S. C. (2014). Evolving BioAssay Ontology (BAO): modularization, integration and applications. *J Biomed Semantics, 5*(Suppl 1 Proceedings of the Bio-Ontologies Spec Interest G), S5. doi:10.1186/2041-1480-5-s1-s5

Adverse Outcome Pathway Wiki. (2016). Retrieved from https://aopwiki.org/wiki/index.php/Main_Page

Aelion, C. M., Davis, H. T., Lawson, A. B., Cai, B., & McDermott, S. (2012). Associations of estimated residential soil arsenic and lead concentrations and community-level environmental measures with mother–child health conditions in South Carolina. *Health & Place, 18*(4), 774-781. doi:http://dx.doi.org/10.1016/j.healthplace.2012.04.005

Agency, U. S. E. P., & Development, O. o. R. a. (2013). Toxicity Forecaster (ToxCast). *Science in Action: Innovative Research for a Sustainable Future*. Retrieved from http://www.epa.gov/sites/production/files/2013-12/documents/toxcast-fact-sheet.pdf

Ahamed, S., Sengupta, M. K., Mukherjee, S. C., Pati, S., Mukherjeel, A., Rahman, M. M., . . . Chakraborti, D. (2006). An eight-year study report on arsenic contamination in groundwater and health effects in Eruani village, Bangladesh and an approach for its mitigation. *J Health Popul Nutr, 24*(2), 129-141.

Ahmad, S. A., Sayed, M. H., Barua, S., Khan, M. H., Faruquee, M. H., Jalil, A., . . . Talukder, H. K. (2001). Arsenic in drinking water and pregnancy outcomes. *Environ Health Perspect, 109*(6), 629-631.

Ahmed, S., Mahabbat-e Khoda, S., Rekha, R. S., Gardner, R. M., Ameer, S. S., Moore, S., . . . Raqib, R. (2011). Arsenic-associated oxidative stress, inflammation, and immune disruption in human placenta and cord blood. *Environ Health Perspect, 119*. doi:10.1289/ehp.1002086

Ankley, G. T., Bennett, R. S., Erickson, R. J., Hoff, D. J., Hornung, M. W., Johnson, R. D., . . . Villeneuve, D. L. (2010). Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environmental Toxicology and Chemistry, 29*(3), 730-741. doi:10.1002/etc.34

Antonogeorgos, G., Panagiotakos, D. B., Pitsavos, C., Papageorgiou, C., Chrysohoou, C., Papadimitriou, G. N., & Stefanadis, C. (2012). Understanding the role of depression and anxiety on cardiovascular disease risk, using structural equation modeling; the mediating effect of the Mediterranean diet and physical activity: the ATTICA study. *Ann Epidemiol, 22*. doi:10.1016/j.annepidem.2012.06.103

Ardalan, B., Subbarayan, P. R., Ramos, Y., Gonzalez, M., Fernandez, A., Mezentsev, D., . . . Ganjei-Azar, P. (2010). A phase I study of 5-fluorouracil/leucovorin and arsenic trioxide for patients with refractory/relapsed colorectal carcinoma. *Clin Cancer Res, 16*. doi:10.1158/1078-0432.ccr-09-2590

Arlinghaus, A., Lombardi, D. A., Willetts, J. L., Folkard, S., & Christiani, D. C. (2012). A structural equation modeling approach to fatigue-related risk factors for occupational injury. *Am J Epidemiol, 176*. doi:10.1093/aje/kws219

Arsenic in Drinking Water. (1999). Washington, DC: National Academy Press.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., . . . Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet, 25*(1), 25-29. doi:10.1038/75556

B., H. A. (1965). The environment and disease: Association or causation? *Proc. R. Soc. Med., 58*, 295-300.

Baclawski, K., & Niu, T. (2006). *Ontologies for Bioinformatics*. Cambridge, Mass.: The MIT Press.

Bada, M., Eckert, M., Evans, D., Garcia, K., Shipley, K., Sitnikov, D., . . . Blake, J. A. e. a. Concept annotation in the CRAFT corpus. *BMC Bioinformatics*, 9, 161.

Bailey, K. A., Laine, J., Rager, J. E., Sebastian, E., Olshan, A., Smeester, L., . . . Fry, R. C. (2014). Prenatal arsenic exposure and shifts in the newborn proteome: interindividual differences in tumor necrosis factor (TNF)-responsive signaling. *Toxicol Sci, 139*(2), 328-337. doi:10.1093/toxsci/kfu053

Baja, E. S., Schwartz, J. D., Coull, B. A., Wellenuis, G. A., Vokonas, P. S., & Suh, H. H. (2012). Structural equation modeling of the inflammatory response to traffic air pollution. *J Expo Sci Environ Epidemiol, 23*. doi:10.1038/jes.2012.106

Bard, J. B. L. (2005). Anatomics: the intersection of anatomy and bioinformatics. *Journal of Anatomy, 206*(1), 1-16. doi:10.1111/j.0021-8782.2005.00376.x

Barry, H., Gordana, A., Philip, C., Dominic, C., David, C., & Ian, D. (2012a). Food for thought … A toxicology ontology roadmap. *ALTEX, 29*(2), 129-137. doi:10.14573/altex.2012.2.129

Becker, R. A., Ankley, G. T., Edwards, S. W., Kennedy, S. W., Linkov, I., Meek, B., . . . Barton-Maclaren, T. S. (2015). Increasing Scientific Confidence in Adverse Outcome Pathways: Application of Tailored Bradford-Hill Considerations for Evaluating Weight of Evidence. *Regul Toxicol Pharmacol, 72*(3), 514-537. doi:10.1016/j.yrtph.2015.04.004

Berggren, E., Amcoff, P., Benigni, R., Blackburn, K., Carney, E., Cronin, M., . . . Worth, A. (2015). Chemical Safety Assessment Using Read-Across: Assessing the Use of Novel Testing Methods to Strengthen the Evidence Base for Decision Making. *Environ Health Perspect, 123*(12), 1232-1240. doi:10.1289/ehp.1409342

Bioportal, N. (2016b). Experimental Factor Ontology- Summary. Retrieved from https://bioportal.bioontology.org/ontologies/EFO

Bioportal, N. (2016c). Sequence Types and Features Ontology. Retrieved from https://bioportal.bioontology.org/ontologies/SO

Bollati, V., & Baccarelli, A. (2010). Environmental epigenetics. *Heredity (Edinb), 105*(1), 105-112. doi:10.1038/hdy.2010.2

Borgert, C., Quill, T., McCarty, L., & Mason, A. (2004). Can mode of action predict mixture toxicity for risk assessment? *Toxicology and Applied Pharmacology, 201*, 85-96.

Brinkman, R. R., Courtot, M., Derom, D., Fostel, J. M., He, Y., Lord, P., . . . Zheng, J. (2010). Modeling biomedical experimental processes with OBI. *J Biomed Semantics, 1 Suppl 1*, S7. doi:10.1186/2041-1480-1-s1-s7

Burgoon, L., Edwards, S., & Villeneuve, D. (Apr 8, 2016).

Caldwell, K. L., Jones, R. L., Verdon, C. P., Jarrett, J. M., Caudill, S. P., & Osterloh, J. D. (2009). Levels of urinary total and speciated arsenic in the US population: National Health and Nutrition Examination Survey 2003–2004. *J Expo Sci Environ Epidemiol, 19*. doi:10.1038/jes.2008.32

Carbonell-Barrachina, A. A., Wu, X., Ramirez-Gandolfo, A., Norton, G. J., Burlo, F., Deacon, C., & Meharg, A. A. (2012). Inorganic arsenic contents in rice-based infant foods from Spain, UK, China and USA. *Environ Pollut, 163*. doi:10.1016/j.envpol.2011.12.036

Chakraborti, D., Mukherjee, S. C., Pati, S., Sengupta, M. K., Rahman, M. M., Chowdhury, U. K., . . . Basu, G. K. (2003). Arsenic groundwater contamination in Middle Ganga Plain, Bihar, India: a future danger? *Environ Health Perspect, 111*(9), 1194-1201.

Chien, C. W., Ho, I. C., & Lee, T. C. (2009). Induction of neoplastic transformation by ectopic expression of human aldo-keto reductase 1C isoforms in NIH3T3 cells. *Carcinogenesis, 30*. doi:10.1093/carcin/bgp195

Coffin, J., Greenwell, R. J., Brooks, D. M., Calderon, L. G., & Beall, H. D. (2006). Vascular dysmorphogenesis causes placental insufficiency and miscarraige in arsenic exposed mice. *Toxicological Sciences, 90*(1), 74.

Collins, F. S., Gray, G. M., & Bucher, J. R. (2008). Toxicology. Transforming environmental health protection. *Science, 319*(5865), 906-907. doi:10.1126/science.1154619

Commission, E. (2006a). concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). *Corrigenda of Regulation*(EC 1907/2006).

Commission, E. (2006b). Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). *Official Journal of the European Union*(EC 1907/2006).

Committee to Review the, I. P., Board on Environmental Studies and, T., Division on Earth and Life, S., & National Research, C. (2014) *Review of EPA's Integrated Risk Information System (IRIS) Process*. Washington (DC): National Academies Press (US)
Copyright 2014 by the National Academy of Sciences. All rights reserved.

Concha, G., Vogler, G., Lezcano, D., Nermell, B., & Vahter, M. (1998). Exposure to inorganic arsenic metabolites during early human development. *Toxicol Sci, 44*. doi:10.1093/toxsci/44.2.185

Consortium, T. G. O. (2014). Gene Ontology Consortium: going forward. *Nucleic Acids Res, 43*(Database issue), D1049-1056. doi:10.1093/nar/gku1179

Consortium, T. G. O. (2015a). Documentation: File Format Guide. CC-BY 4.0. Retrieved from http://geneontology.org/page/file-format-guide

Consortium, T. G. O. (2015b). Ontology Related Information/Biological Process Ontology Guidelines. Retrieved from http://geneontology.org/page/biological-process-ontology-guidelines

COSMOSTOX. COSMOS: Integrated In Silico Models for the Prediction of Human Repeated Dose Toxicity of COSMetics to Optimise Safety. Retrieved from http://www.cosmostox.eu/home/welcome/

Cote, I. (2016, Jan 12).

Cote, R., Reisinger, F., Martens, L., Barsnes, H., Vizcaino, J. A., & Hermjakob, H. (2010). The Ontology Lookup Service: bigger and better. *Nucleic Acids Res, 38*(Web Server issue), W155-160. doi:10.1093/nar/gkq331

Council, N. R. (2007). *Toxicity testing in the 21st Century: A vision and a strategy.* Washington, DC, USA: The National Academies Press.

Council, N. R. (2014). *Review of EPA's Integrated Risk Information System (IRIS) Process*. Washington, DC: The National Academies Press.

Crofton, K. (2016).

Damiano, A. E. (2011). Review: Water channel proteins in the human placenta and fetal membranes. *Placenta, 32*. doi:10.1016/j.placenta.2010.12.012

Damiano, A. E., Zotta, E., & Ibarra, C. (2006). Functional and molecular expression of AQP9 channel and UT-A transporter in normal and preeclamptic human placentas. *Placenta, 27*. doi:10.1016/j.placenta.2005.11.014

DataSciBurgoon. (2015a). aop-ontology/README.md. Retrieved from https://github.com/DataSciBurgoon/aop-ontology/blob/master/README.md

DataSciBurgoon. (2015b). aop-ontology/AOPO Overview Slide.png. Retrieved from https://github.com/DataSciBurgoon/aop-ontology/blob/master/AOPO%20Overview%20Slide.png

David, R. J., & Collins, J. W. (1997). Differing birth weight among infants of U.S.-born blacks, African-born blacks, and U.S.-born whites. *N Engl J Med, 337*. doi:10.1056/nejm199710233371706

Davis, A. P., Grondin, C. J., Lennon-Hopkins, K., Saraceni-Richards, C., Sciaky, D., King, B. L., . . . Mattingly, C. J. (2015). The Comparative Toxicogenomics Database's 10th year anniversary: update 2015. *Nucleic Acids Res, 43*(Database issue), D914-920. doi:10.1093/nar/gku935

Development, O. f. E. C. a. (2013a). *Guidance Document on Developing and Assessing Adverse Outcome Pathways*. (ENV/JM/MONO(2013)6). Paris: OECD.

Development, O. f. E. C. a. (2013b). *Users' Handbook Supplement to the Guidance Document for Developing and Assessing AOPs*. (ENV/JM/MONO(2013)6). Paris: OECD Retrieved from http://aopkb.org/common/AOP_Handbook.pdf.

Diehl, A. D., Augustine, A. D., Blake, J. A., Cowell, L. G., Gold, E. S., Gondre-Lewis, T. A., . . . Nijnik, A. e. a. (2011). Hematopoietic cell types: prototype for a revised cell ontology., 44, 75-79.

Dix, D. J., Houck, K. A., Martin, M. T., Richard, A. M., Setzer, R. W., & Kavlock, R. J. (2007). The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol Sci, 95*(1), 5-12. doi:10.1093/toxsci/kfl103

Drobná, Z., Del Razo, L. M., García-Vargas, G. G., Sánchez-Peña, L. C., Barrera-Hernández, A., Stýblo, M., & Loomis, D. (2012). Environmental exposure to arsenic, AS3MT polymorphism and prevalence of diabetes in Mexico. *Journal of exposure science & environmental epidemiology, 23*. doi:10.1038/jes.2012.103

Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res, 30*(1), 207-210.

Edwards, S. (July 8, 2016).

Edwards, S. W., Tan, Y. M., Villeneuve, D. L., Meek, M. E., & McQueen, C. A. (2016). Adverse Outcome Pathways-Organizing Toxicological Information to Improve Decision Making. *J Pharmacol Exp Ther, 356*(1), 170-181. doi:10.1124/jpet.115.228239

EMBL-EBI. (2016). ChEBI:About ChEBI. Retrieved from https://www.ebi.ac.uk/chebi/aboutChebiForward.do

Farzan, S. F., Karagas, M. R., & Chen, Y. (2013). In utero and early life arsenic exposure in relation to long-term health and disease. *Toxicol. Appl. Pharmacol.*

Federhen, S. (2012). The NCBI Taxonomy database. *Nucleic Acids Research, 40*(D1), D136-D143. doi:10.1093/nar/gkr1178

Federhen S. In: McEntyre, J., Ostell J, editors. (2002 Oct 9 [Updated 2003 Aug 13].). *The Taxonomy Project*. Bethesda (MD): National Center for Biotechnology Information Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK21100/.

Fei, D. L., Koestler, D. C., Li, Z., Giambelli, C., Sanchez-Mejias, A., Gosse, J. A., . . . Robbins, D. J. (2013). Association between In Utero arsenic exposure, placental gene expression, and infant birth weight: a US birth cohort study. *Environmental Health, 12*(1), 1-8. doi:10.1186/1476-069x-12-58

Fei, D. L., Li, H., Kozul, C. D., Black, K. E., Singh, S., Gosse, J. A., . . . Robbins, D. J. (2010). Activation of Hedgehog signaling by the environmental toxicant arsenic may contribute to the etiology of arsenic-induced tumors. *Cancer Res, 70*. doi:10.1158/0008-5472.can-09-2898

Ferm, V. H., & Carpenter, S. J. (1968). Malformations induced by sodium arsenate. *J Reprod Fertil, 17*. doi:10.1530/jrf.0.0170199

Foundry, T. O. (2016a). The OBO Foundry. Retrieved from http://obofoundry.org/

Foundry, T. O. (2016b). Cell Ontology. Retrieved from http://obofoundry.org/ontology/cl.html

Foundry, T. O. (2016c). NCBI organismal classification. Retrieved from http://obofoundry.org/ontology/ncbitaxon.html

Foundry, T. O. (2016d). Phenotypic quality. Retrieved from http://obofoundry.org/ontology/pato.html

Foundry, T. O. (2016e). Relations Ontology. Retrieved from http://obofoundry.org/ontology/ro.html

Foundry, T. O. (2016f). Sequence Types and Features. Retrieved from http://www.obofoundry.org/ontology/so.html

Fry, R. (April 2016).

Fry, R. C., Navasumrit, P., Valiathan, C., Svensson, J. P., Hogan, B. J., Luo, M., . . . Samson, L. D. (2007). Activation of inflammation/NF-kappaB signaling in infants born to arsenic-exposed mothers. *PLoS Genet, 3*. doi:10.1371/journal.pgen.0030207

Gardner, R. M., Kippler, M., Tofail, F., Bottai, M., Hamadani, J., Grandér, M., . . . Vahter, M. (2013). Environmental Exposure to Metals and Children's Growth to Age 5 Years: A Prospective Cohort Study. *American Journal of Epidemiology, 177*(12), 1356-1367. doi:10.1093/aje/kws437

Gehlhaus, M. W., 3rd, Gift, J. S., Hogan, K. A., Kopylev, L., Schlosser, P. M., & Kadry, A. R. (2011). Approaches to cancer assessment in EPA's Integrated Risk Information System. *Toxicol Appl Pharmacol, 254*(2), 170-180. doi:10.1016/j.taap.2010.10.019

Geiss, G. K., Bumgarner, R. E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D. L., . . . Dimitrov, K. (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol, 26*. doi:10.1038/nbt1385

Gelmann, E. R., Gurzau, E., Gurzau, A., Goessler, W., Kunrath, J., Yeckel, C. W., & McCarty, K. M. (2013). A pilot study: The importance of inter-individual differences in inorganic arsenic metabolism for birth weight outcome. *Environmental Toxicology and Pharmacology, 36*(3), 1266-1275. doi:http://dx.doi.org/10.1016/j.etap.2013.10.006

Gilbert-Diamond, D., Cottingham, K. L., Gruber, J. F., Punshon, T., Sayarath, V., Gandolfi, A. J., . . . Karagas, M. R. (2011). Rice consumption contributes to arsenic exposure in US women. *Proc Natl Acad Sci U S A, 108*. doi:10.1073/pnas.1109127108

Gong, G., & O'Bryant, S. E. (2012). Low-level arsenic exposure, AS3MT gene polymorphism and cardiovascular diseases in rural Texas counties. *Environ Res, 113*. doi:10.1016/j.envres.2012.01.003

Gremse, M., Chang, A., Schomburg, I., Grote, A., Scheer, M., Ebeling, C., & Schomburg, D. (2011). The BRENDA Tissue Ontology (BTO): the first all-integrating ontology of all organisms for enzyme sources. *Nucleic Acids Res, 39*(Database issue), D507-513. doi:10.1093/nar/gkq968

Group, U. o. W. S. I. (2007). Foundational Model of Anatomy: About FMA. Retrieved from http://sig.biostr.washington.edu/projects/fm/AboutFM.html

Group, E. R. (1997). *Report on the expert panel on arsenic carcinogenicity: review and workshop.* Retrieved from

Guan, H., Piao, F., Zhang, X., Li, X., Li, Q., Xu, L., . . . Yokoyama, K. (2012). Prenatal Exposure to Arsenic and Its Effects on Fetal Development in the General Population of Dalian. *Biological Trace Element Research, 149*(1), 10-15. doi:10.1007/s12011-012-9396-7

Haendel, M., Balhoff, J., Bastian, F., Blackburn, D., Blake, J., Bradford, Y., . . . Mungall, C. (2014). Unification of multi-species vertebrate anatomy ontologies for comparative biology in Uberon. *Journal of Biomedical Semantics, 5*(1), 21.

Hall, M., Gamble, M., Slavkovich, V., Liu, X., Levy, D., Cheng, Z., . . . Graziano, J. (2007). Determinants of arsenic metabolism: blood arsenic metabolites, plasma folate, cobalamin, and homocysteine concentrations in maternal-newborn pairs. *Environ Health Perspect, 115*.

Hardy, B., Apic, G., Carthew, P., Clark, D., Cook, D., Dix, I., . . . Yang, C. (2012b). Toxicology ontology perspectives. *Altex, 29*(2), 139-156.

Hardy, B., Douglas, N., Helma, C., Rautenberg, M., Jeliazkova, N., Jeliazkov, V., . . . Escher, S. (2010). Collaborative development of predictive toxicology applications. *J Cheminform, 2*(1), 7. doi:10.1186/1758-2946-2-7

Harrington, J. M., Young, D. J., Fry, R. C., Weber, F. X., Sumner, S. S., & Levine, K. E. (2016). Validation of a Metallomics Analysis of Placenta Tissue by Inductively-Coupled Plasma Mass Spectrometry. *Biol Trace Elem Res, 169*(2), 164-173. doi:10.1007/s12011-015-0431-3

Hastings, J., de Matos, P., Dekker, A., Ennis, M., Harsha, B., Kale, N., . . . Steinbeck, C. (2013). The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Res, 41*(Database issue), D456-463. doi:10.1093/nar/gks1146

He, W., Greenwell, R. J., Brooks, D. M., Calderon-Garciduenas, L., Beall, H. D., & Coffin, J. D. (2007). Arsenic exposure in pregnant mice disrupts placental vasculogenesis and causes spontaneous abortion. *Toxicol Sci, 99*(1), 244-253. doi:10.1093/toxsci/kfm162

Hood, R. D., & Bishop, S. L. (1972). Teratogenic effects of sodium arsenate in mice. *Arch Environ Health, 24*. doi:10.1080/00039896.1972.10666051

Hopenhayn, C., Ferreccio, C., Browning, S. R., Huang, B., Peralta, C., Gibb, H., & Hertz-Picciotto, I. (2003). Arsenic exposure from drinking water and birth weight. *Epidemiology, 14*(5), 593-602. doi:10.1097/01.ede.0000072104.65240.69

Hourani, J. (2012). Biological Expression Language (BEL) Overview. *OpenBEL Wiki.* Retrieved from http://wiki.openbel.org/display/BLD/Biological+Expression+Language+%28BEL%29+Overview

Howe, C. G., Liu, X., Hall, M. N., Slavkovich, V., Ilievski, V., Parvez, F., . . . Gamble, M. V. (2016). Associations between Blood and Urine Arsenic Concentrations and Global Levels of Post-Translational Histone Modifications in Bangladeshi Men and Women. *Environ Health Perspect*. doi:10.1289/ehp.1510412

Huang, M., Choi, S., Kim, D., Kim, N., Park, C., Yu, S., . . . Park, J. (2009). Risk assessment of low-level cadmium and arsenic on the kidney. *J Toxicol Environ Health A, 72*, 1493-1498.

Huyck, K. L., Kile, M. L., Mahiuddin, G., Quamruzzaman, Q., Rahman, M., Breton, C. V., . . . Christiani, D. C. (2007). Maternal arsenic exposure associated with low birth weight in Bangladesh. *J Occup Environ Med, 49*(10), 1097-1104. doi:10.1097/JOM.0b013e3181566ba0

Iwagaki, S., Yokoyama, Y., Tang, L., Takahashi, Y., Nakagawa, Y., & Tamaya, T. (2004). Augmentation of leptin and hypoxia-inducible factor 1alpha mRNAs in the pre-eclamptic placenta. *Gynecol Endocrinol, 18*(5), 263-268.

Iwasawa, Y., Fujii, T., Nagamatsu, T., Kawana, K., Okudaira, S., Miura, S., . . . Taketani, Y. (2009). Expression of autotaxin, an ectoenzyme that produces lysophosphatidic acid, in human placenta. *Am J Reprod Immunol, 62*. doi:10.1111/j.1600-0897.2009.00715.x

Jackson, B. P., Taylor, V. F., Karagas, M. R., Punshon, T., & Cottingham, K. L. (2012). Arsenic, organic foods, and brown rice syrup. *Environ Health Perspect, 120*. doi:10.1289/ehp.1104619

Jirtle, R., & Skinner, M. (2007). Environmental epigenomics and disease susceptibility. *Nat Rev Genet, 8*, 253-262.

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics, 8*. doi:10.1093/biostatistics/kxj037

Judson, R., Richard, A., Dix, D., Houck, K., Elloumi, F., Martin, M., . . . Wolf, M. (2008). ACToR--Aggregated Computational Toxicology Resource. *Toxicol Appl Pharmacol, 233*(1), 7-13. doi:10.1016/j.taap.2007.12.037

Karagas, M. R., Andrew, A. S., Nelson, H. H., Li, Z., Punshon, T., Schned, A., . . . Kelsey, K. T. (2012). SLC39A2 and FSIP1 polymorphisms as potential modifiers of arsenic-related bladder cancer. *Hum Genet, 131*. doi:10.1007/s00439-011-1090-x

King, O., Foulger, R., Dwight, S., White, J., & Roth, F. (2003). Predicting gene function from patterns of annotation. *Genome Res, 12*, 656-664.

Kippler, M., Wagatsuma, Y., Rahman, A., Nermell, B., Persson, L.-Å., Raqib, R., & Vahter, M. (2012). Environmental exposure to arsenic and cadmium during pregnancy and fetal size: A longitudinal study in rural Bangladesh. *Reproductive Toxicology, 34*(4), 504-511. doi:http://dx.doi.org/10.1016/j.reprotox.2012.08.002

Kline, R. B. (2004). *Principles and Practice of Structural Equation Modeling*. New York: Guilford Press.

Krewski, D., Acosta, D., Jr., Andersen, M., Anderson, H., Bailar, J. C., 3rd, Boekelheide, K., . . . Zeise, L. (2010). Toxicity testing in the 21st century: a vision and a strategy. *J Toxicol Environ Health B Crit Rev, 13*(2-4), 51-138. doi:10.1080/10937404.2010.483176

Kwok, R. K., Kaufmann, R. B., & Jakariya, M. (2006). Arsenic in drinking-water and reproductive health outcomes: a study of participants in the Bangladesh Integrated Nutrition Programme. *J Health Popul Nutr, 24*(2), 190-205.

Köhler, S., Doelken, S. C., Mungall, C. J., Bauer, S., Firth, H. V., Bailleul-Forestier, I., . . . Robinson, P. N. (2014). The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Research, 42*(D1), D966-D974. doi:10.1093/nar/gkt1026

Laine, J. E., Ray, P., Bodnar, W., Cable, P. H., Boggess, K., Offenbacher, S., & Fry, R. C. (2015). Placental Cadmium Levels Are Associated with Increased Preeclampsia Risk. *PLoS One, 10*(9), e0139341. doi:10.1371/journal.pone.0139341

Landesmann, B., & Dumont, C. (2016). Covalent Protein Binding Leading to Skin Sensitisation. *AOP:40.* Retrieved from https://aopwiki.org/wiki/index.php/Aop:40

Li, C.-S., & Loch-Caruso, R. (2007). Sodium arsenite inhibits migration of extravillous trophoblast cells in vitro. *Reproductive Toxicology, 24*(3–4), 296-302. doi:http://dx.doi.org/10.1016/j.reprotox.2007.06.002

Li, C. S., & Loch-Caruso, R. (2007). Sodium arsenite inhibits migration of extravillous trophoblast cells in vitro. *Reprod Toxicol, 24*(3-4), 296-302. doi:10.1016/j.reprotox.2007.06.002

Lyon, D., Cheng, C. Y., Howland, L., Rattican, D., Jallo, N., Pickler, R., . . . McGrath, J. (2010). Integrated review of cytokines in maternal, cord, and newborn blood: part I--associations with preterm birth. *Biol Res Nurs, 11*(4), 371-376. doi:10.1177/1099800409344620

Magdalini Sachana, S. M., Anna Price. (2016). Chronic binding of antagonist to N-methyl-D-aspartate receptors (NMDARs) during brain development induces impairment of learning and memory abilities. Retrieved from https://aopwiki.org/wiki/index.php/Aop:13

Malgorzata Nepelska, S. M., Brigitte Landesmann. (2016). PPARalpha activation in utero leading to imparied fertility in males. Retrieved from https://aopwiki.org/wiki/index.php/Aop:18

Marsit, C. J., Maccani, M. A., Padbury, J. F., & Lester, B. M. (2012). Placental 11-beta hydroxysteroid dehydrogenase methylation is associated with newborn growth and a measure of neurobehavioral outcome. *PLoS One, 7*. doi:10.1371/journal.pone.0033794

Martin, E., Ray, P. D., Smeester, L., Grace, M. R., Boggess, K., & Fry, R. C. (2015). Epigenetics and Preeclampsia: Defining Functional Epimutations in the Preeclamptic Placenta Related to the TGF-beta Pathway. *PLoS One, 10*(10), e0141294. doi:10.1371/journal.pone.0141294

Martin, M. T. (2008). *Classification of Chemicals Based on Structured Toxicity Information.* (Master of Science), University of North Carolina at Chapel Hill, Chapel Hill, NC. Retrieved from https://cdr.lib.unc.edu Carolina Digital Repository database.

Masci, A. M., Arighi, C. N., Diehl, A. D., Lieberman, A. E., Mungall, C., Scheuermann, R. H., . . . Cowell, L. G. (2009). An improved ontological representation of dendritic cells as a paradigm for all cell types. (pp. 10, 70): BMC Bioinformatics.

Massrieh, W., Derjuga, A., & Blank, V. (2006). Induction of endogenous Nrf2/small maf heterodimers by arsenic-mediated stress in placental choriocarcinoma cells. . *Antioxid Redox Signal, 8*, 53-59.

Masuda, A., Fujii, T., Iwasawa, Y., Nakamura, K., Ohkawa, R., Igarashi, K., . . . Yatomi, Y. (2011). Serum autotaxin measurements in pregnant women: application for the differentiation of normal pregnancy and pregnancy-induced hypertension. *Clin Chim Acta, 412*. doi:10.1016/j.cca.2011.06.039

Mayer, G., Jones, A. R., Binz, P. A., Deutsch, E. W., Orchard, S., Montecchi-Palazzi, L., . . . Eisenacher, M. (2014). Controlled vocabularies and ontologies in proteomics: overview, principles and practice. *Biochim Biophys Acta, 1844*(1 Pt A), 98-107. doi:10.1016/j.bbapap.2013.02.017

McDermott, S., Bao, W., Aelion, C. M., Cai, B., & Lawson, A. B. (2014). Does the metal content in soil around a pregnant woman's home increase the risk of low birth weight for her infant? *Environmental Geochemistry and Health, 36*(6), 1191-1197. doi:10.1007/s10653-014-9617-4

Medicine, U. S. N. L. o. (2015). Fact Sheet: Medical Subject Headings (MeSH). Bethesda, Maryland: National Libary of Medicine.

Meek, M. E., & Armstrong, V. C. (2007). The Assessment and Management of Industrial Chemicals in Canada *Risk Assessment of Chemicals* (pp. 591-621). Netherlands: Springer.

Meek, M. E., Boobis, A., Cote, I., Dellarco, V., Fotakis, G., Munn, S., . . . Vickers, C. (2014). New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. *J Appl Toxicol, 34*(1), 1-18. doi:10.1002/jat.2949

Meharg, A. A., Sun, G., Williams, P. N., Adomako, E., Deacon, C., Zhu, Y. G., . . . Raab, A. (2008). Inorganic arsenic levels in baby rice are of concern. *Environ Pollut, 152*. doi:10.1016/j.envpol.2008.01.043

Milton, A. H., Smith, W., Rahman, B., Hasan, Z., Kulsum, U., Dear, K., . . . Ali, A. (2005). Chronic arsenic exposure and adverse pregnancy outcomes in bangladesh. *Epidemiology, 16*. doi:10.1097/01.ede.0000147105.94041.e6

Mirkes, P., & Little, S. (1998). Teratogen-induced cell death in postimplantation mouse embryos: differential tissue sensitivity and hallmarks of apoptosis. *Cell Death Differ, 5*, 592-600. doi:http://dx.doi.org/10.1038/sj.cdd.4400390

Mise, H., Sagawa, N., Matsumoto, T., Yura, S., Nanno, H., Itoh, H., . . . Nakao, K. (1998). Augmented placental production of leptin in preeclampsia: possible involvement of placental hypoxia. *J Clin Endocrinol Metab, 83*(9), 3225-3229. doi:10.1210/jcem.83.9.5117

Mo, J., Xia, Y., Wade, T. J., DeMarini, D. M., Davidson, M., & Mumford, J. (2011). Altered gene expression by low-dose arsenic exposure in humans and cultured cardiomyocytes: assessment by real-time PCR arrays. *Int J Environ Res Public Health, 8*(6), 2090-2108. doi:10.3390/ijerph8062090

MS, A.-R., N, M., R, H., S, M., M, I., RJ, K., . . . M, X. (2013). The Tox21 robotoic platform for the assessment of environmental chemicals--from vision to reality. *Drug discovery today, 18*, 716-723.

Mukherjee, S. C., Saha, K. C., Pati, S., Dutta, R. N., Rahman, M. M., Sengupta, M. K., . . . Asad, K. A. (2005). Murshidabad—One of the Nine Groundwater Arsenic-Affected Districts of West Bengal, India. Part II: Dermatological, Neurological, and Obstetric Findings. *Clinical Toxicology, 43*(7), 835-848. doi:10.1080/15563650500357495

Mungall, C., Torniai, C., Gkoutos, G., Lewis, S., & Haendel, M. (2012). Uberon, an integrative multi-species anatomy ontology. *Genome Biology, 13*(1), R5.

Mungall, C. J., Batchelor, C., & Eilbeck, K. (2011). Evolution of the Sequence Ontology terms and relationships. *J Biomed Inform, 44*(1), 87-93. doi:10.1016/j.jbi.2010.03.002

Myers, S. L., Lobdell, D. T., Liu, Z., Xia, Y., Ren, H., Li, Y., . . . Mendola, P. (2010). Maternal drinking water arsenic exposure and perinatal outcomes in Inner Mongolia, China. *Journal of Epidemiology and Community Health, 64*(4), 325-329. doi:10.1136/jech.2008.084392

Natale, D. A., Arighi, C. N., Blake, J. A., Bult, C. J., Christie, K. R., Cowart, J., . . . Wu, C. H. (2014). Protein Ontology: a controlled structured network of protein entities. *Nucleic Acids Res, 42*(Database issue), D415-421. doi:10.1093/nar/gkt1173

Nelson, S. J. (2009). *Medial Terminologies that Work: The Example of MeSH.* Paper presented at the ISPAN 2009, Kaoshiung, Taiwan.

NRC. (2001). *National Research Council: Arsenic in the Drinking Water: 2001 Update*. Retrieved from Washington, D.C.:

NRC. (2013). *Critical aspects of EPA's IRIS assessment of inorganic arsenic.* Retrieved from

OECD. (2012a). Proposal for a template and guidance on developing and assessing the completeness of adverse outcome pathways. Paris, France.

OECD. (2012b). *The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins*. (ENV/JM/MONO(2012)10/PART1). Paris: OECD.

Office, G. A. (2005). *Chemical regulation: Options exist to improve EPA's ability to assess health risks and manage its chemical review program.* Washington, DC: U.S. Government Accounting Office Retrieved from http://www.gao.gov/new.items/d05458.pdf.

Oki, N. O., Nelms, M. D., Bell, S. M., Mortensen, H. M., & Edwards, S. W. (2016). Accelerating Adverse Outcome Pathway Development Using Publicly Available Data Sources. *Current Environmental Health Reports*, 1-11. doi:10.1007/s40572-016-0079-y

Okudaira, S., Yukiura, H., & Aoki, J. (2010). Biological roles of lysophosphatidic acid signaling through its production by autotaxin. *Biochimie, 92*. doi:10.1016/j.biochi.2010.04.015

Ontology, H. P. About. Retrieved from http://human-phenotype-ontology.github.io/about.html

Ontology, H. P. Documentation. Retrieved from http://human-phenotype-ontology.github.io/documentation.html

Ontology, P. a. C. PCO. Retrieved from https://github.com/PopulationAndCommunityOntology/pco

Ontology, S. (2009). SO: The Sequence Ontology Project. Retrieved from http://www.sequenceontology.org/index.html

Ontology, T. N. C. f. B. (2005-2016). Sequence Types and Features Ontology. Retrieved from https://bioportal.bioontology.org/ontologies/SO

Organization, W. H., & Cancer, I. A. f. R. o. (2004). Some drinking-water disinfectants and contaminants, including arsenic *IARC monographs on the evaluation of carcinogenic risks to humans. vol. 84* (Vol. 84). France: IARC Press.

Pandey, A., & Lewitter, F. (1999). Nucleotide sequence databases: a gold mine for biologists. *Trends Biochem. Sci., 24*, 276-280.

Persad, A. S., & Cooper, G. S. (2008). Use of epidemiologic data in Integrated Risk Information System (IRIS) assessments. *Toxicol Appl Pharmacol, 233*(1), 137-145. doi:10.1016/j.taap.2008.01.013

Phuong, J. (2014). *Structured Application of Biological Ontologies to Annotate High-Throughput Screening Assays and Their Targets of Activity.* UNC- Chapel Hill, Chapel Hill, NC. Retrieved from https://cdr.lib.unc.edu Carolina Digital Repository database.

Pi, J., Qu, W., Reece, J. M., Kumagai, Y., & Waalkes, M. P. (2003). Transcription factor Nrf2 activation by inorganic arsenic in cultured keratinocytes: involvement of hydrogen peroxide. *Exp Cell Res, 290*. doi:10.1016/s0014-4827(03)00341-0

Pottenger, L. H., Schoeny, R., & Moore, M. (2016). AFB1: Mutagenic Mode of Action leading to Hepatocellular Carcinoma (HCC). Retrieved from https://aopwiki.org/wiki/index.php/Aop:46

Rager, J. E., Yosim, A., & Fry, R. C. (2014). Prenatal exposure to arsenic and cadmium impacts infectious disease-related genes within the glucocorticoid receptor signal transduction pathway. *Int J Mol Sci, 15*(12), 22374-22391. doi:10.3390/ijms151222374

Rahman, A., Vahter, M., Ekstrom, E. C., Rahman, M., Golam Mustafa, A. H., Wahed, M. A., . . . Persson, L. A. (2007). Association of arsenic exposure during pregnancy with fetal loss and infant death: a cohort study in Bangladesh. *Am J Epidemiol, 165*. doi:10.1093/aje/kwm025

Rahman, A., Vahter, M., Smith, A. H., Nermell, B., Yunus, M., El Arifeen, S., . . . Ekström, E.-C. (2009a). Arsenic exposure during pregnancy and size at birth: a prospective cohort study in Bangladesh. *Am J Epidemiol, 169*. doi:10.1093/aje/kwn332

Rahman, L. A., Hairi, N. N., & Salleh, N. (2008). Association between pregnancy induced hypertension and low birth weight; a population based case–control study. *Asia Pac J Public Health, 20*. doi:10.1177/1010539507311553

Rahman, M. M., Ng, J. C., & Naidu, R. (2009b). Chronic exposure of arsenic via drinking water and its adverse health impacts on humans. *Environmental Geochemical Health, 31(Suppl. 1)*, 189-200.

Rahman, M. M., Sengupta, M. K., Ahamed, S., Chowdhury, U. K., Hossain, M. A., Das, B., . . . Chakraborti, D. (2005). The magnitude of arsenic contamination in groundwater and its health effects to the inhabitants of the Jalangi—one of the 85 arsenic affected blocks in West Bengal, India. *Science of The Total Environment, 338*(3), 189-200. doi:http://dx.doi.org/10.1016/j.scitotenv.2004.06.022

RB, C. (2002). The Use of Biologically Based Modelling in Risk Assessment. *Toxicology*, 181-182, 275-279.

Remy, S., Govarts, E., Bruckers, L., Paulussen, M., Wens, B., Hond, E. D., . . . Schoeters, G. (2014). Expression of the sFLT1 Gene in Cord Blood Cells Is Associated to Maternal Arsenic Exposure and Decreased Birth Weight. *PLoS ONE, 9*(3), e92677. doi:10.1371/journal.pone.0092677

Ren, K., Li, X., Yan, J., Huang, G., Zhou, S., Yang, B., . . . Lu, C. (2014). Knockdown of p66Shc by siRNA injection rescues arsenite-induced developmental retardation in mouse preimplantation embryos. *Reprod Toxicol, 43*, 8-18. doi:10.1016/j.reprotox.2013.10.008

Robinson, P. N., & Bauer, S. (2011). *Introduction to Bio-ontologies*. Boca Raton, FL: Taylor & Francis Group, LLC.

Rojas, D., Rager, J. E., Smeester, L., Bailey, K. A., Drobna, Z., Rubio-Andrade, M., . . . Fry, R. C. (2015). Prenatal arsenic exposure and the epigenome: identifying sites of 5-methylcytosine alterations that predict functional changes in gene expression in newborn cord blood and subsequent birth outcomes. *Toxicol Sci, 143*(1), 97-106. doi:10.1093/toxsci/kfu210

Rosen, B. P. (2002). Biochemistry of arsenic detoxification. *FEBS Lett, 529*. doi:10.1016/s0014-5793(02)03186-1

Ross, K. E., Arighi, C. N., Ren, J., Huang, H., & Wu, C. H. (2013). Construction of protein phosphorylation networks by data mining, text mining and ontology integration: analysis of the spindle checkpoint. *Database (Oxford)*, bat038.

Rosse, C., & Mejino, J. L. V. J. (2007). The Foundational Model of Anatomy Ontology. In D. D. A. Burger, and R. Baldock, editors (Ed.), *Anatomy Ontologies for Bioinformatics: Principles and Practice* (Vol. 6, pp. 59-117). London: Springer.

Saha, K. K., Engstrom, A., Hamadani, J. D., Tofail, F., Rasmussen, K. M., & Vahter, M. (2012). Pre- and postnatal arsenic exposure and body size to 2 years of age: a cohort study in rural Bangladesh. *Environ Health Perspect, 120*(8), 1208-1214. doi:10.1289/ehp.1003378

Schurer, S. C., Vempati, U., Smith, R., Southern, M., & Lemmon, V. (2011). BioAssay ontology annotations facilitate cross-analysis of diverse high-throughput screening data sets. *J Biomol Screen, 16*(4), 415-426. doi:10.1177/1087057111400191

Shah, I., & Corton, C. (2010). The EPA Liver Project. *EPA Science Inventory.* Retrieved from http://cfpub.epa.gov/si/si_public_record_Report.cfm?dirEntryId=231189

Shinkai, Y., Sumi, D., Toyama, T., Kaji, T., & Kumagai, Y. (2009). Role of aquaporin 9 in cellular accumulation of arsenic and its cytotoxicity in primary mouse hepatocytes. *Toxicol Appl Pharmacol, 237*. doi:10.1016/j.taap.2009.03.014

Singh, S., Greene, R., & Pisano, M. (2010). Arsenate-induced apoptosis in murine embryonic maxillary mesenchymal cells via mitochondrial-mediated oxidative injury. *Birth Defects Res A Clin Mol Teratol, 88*, 25-34.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., . . . Lewis, S. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotech, 25*(11), 1251-1255.

Smith, C. L., & Eppig, J. T. (2015). Expanding the mammalian phenotype ontology to support automated exchange of high throughput mouse phenotyping data generated by large-scale mouse knockout screens. *J Biomed Semantics, 6*, 11. doi:10.1186/s13326-015-0009-1

Smolders, R., Bartonova, A., Boogaard, P. J., Dusinska, M., Koppen, G., Merlo, F., . . . Schoeters, G. (2010). The use of biomarkers for risk assessment: Reporting from the INTARESE/ENVIRISK Workshop in Prague. *Int J Hyg Environ Health, 213*. doi:10.1016/j.ijheh.2010.05.006

Spranger, R. M., Hartner, A., Schild, R. L., Stuppy, A., Knerr, I., Rascher, W., & Dotsch, J. (2008). Leptin expression in primary trophoblast cells in response to incubation with the serum of preeclamptic women. *Hypertens Pregnancy, 27*(3), 267-277. doi:10.1080/10641950802174870

Tcheremenskaia, O., Benigni, R., Nikolova, I., Jeliazkova, N., Escher, S. E., Batke, M., . . . Hardy, B. (2012). OpenTox predictive toxicology framework: toxicological ontology and semantic media wiki-based OpenToxipedia. *J Biomed Semantics, 3 Suppl 1*, S7. doi:10.1186/2041-1480-3-s1-s7

Tokumura, A., Kanaya, Y., Miyake, M., Yamano, S., Irahara, M., & Fukuzawa, K. (2002). Increased production of bioactive lysophosphatidic acid by serum lysophospholipase D in human pregnancy. *Biol Reprod, 67*. doi:10.1095/biolreprod.102.004051

Torres-Avila, M., Leal-Galicia, P., Sanchez-Pena, L. C., Del Razo, L. M., & Gonsebatt, M. E. (2010). Arsenite induces aquaglyceroporin 9 expression in murine livers. *Environ Res, 110*. doi:10.1016/j.envres.2009.08.009

Tsukaguchi, H., Shayakul, C., Berger, U. V., Mackenzie, B., Devidas, S., Guggino, W. B., . . . Hediger, M. A. (1998). Molecular characterization of a broad selectivity neutral solute channel. *J Biol Chem, 273*. doi:10.1074/jbc.273.38.24737

Tudor, C. O., Arighi, C. N., Wang, Q., Wu, C. H., & Vijay-Shanker, K. (2012). The eFIP system for text mining of protein interaction networks of phosphorylated proteins. . *Database (Oxford)*, bas044.

USEPA. (2005). *Guidelines for carcinogen risk assessment. (EPA/630/P-03/001F)*. Washington, DC: USEPA: Risk Assessment Forum Retrieved from http://www.epa.gov/cancerguidelines/.

USEPA. (2014a). Arsenic, inorganic. Retrieved from https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?substance_nmbr=278#fragment-2

USEPA. (2014b). *Draft Development Materials for the Integrated Risk Information System (IRIS) Toxicological Review of Inorganic Arsenic*. (EPA/630/R-14/101). Washington, D.C.

USEPA. (2015). Basic Information about the Integrated Risk Information System. Retrieved from https://www.epa.gov/iris/basic-information-about-integrated-risk-information-system

USEPA, N. C. f. E. A. (1988). *Integrated Risk Information System (IRIS), Chemical Assessment Summary, Arsenic, inorganic*. Retrieved from https://cfpub.epa.gov/ncea/iris/iris_documents/documents/subst/0278_summary.pdf.

USEPA, O. o. R. a. D. (2013). Toxicity Forecaster (ToxCast).

Vahter, M. (2009). Effects of arsenic on maternal and fetal health. *Annu Rev Nutr, 29*. doi:10.1146/annurev-nutr-080508-141102

Vall, O., Gómez-Culebras, M., Garcia-Algar, O., Joya, X., Velez, D., Rodríguez-Carrasco, E., & Puig, C. (2012). Assessment of Prenatal Exposure to Arsenic in Tenerife Island. *PLoS ONE, 7*(11), e50463. doi:10.1371/journal.pone.0050463

Vempati, U. D., Chung, C., Mader, C., Koleti, A., Datar, N., Vidovic, D., . . . Schurer, S. C. (2014). Metadata Standard and Data Exchange Specifications to Describe, Model, and Integrate Complex and Diverse High-Throughput Screening Data from the Library of Integrated Network-based Cellular Signatures (LINCS). *J Biomol Screen, 19*(5), 803-816. doi:10.1177/1087057114522514

Villeneuve, D. (2016). Aromatase inhibition leading to reproductive dysfunction (in fish). Retrieved from https://aopwiki.org/wiki//index.php/Aop:25

Villeneuve, D. L., Crump, D., Garcia-Reyero, N., Hecker, M., Hutchinson, T. H., LaLone, C. A., . . . Whelan, M. (2014a). Adverse Outcome Pathway (AOP) Development I: Strategies and Principles. *Toxicological Sciences, 142*(2), 312-320. doi:10.1093/toxsci/kfu199

Villeneuve, D. L., Crump, D., Garcia-Reyero, N., Hecker, M., Hutchinson, T. H., LaLone, C. A., . . . Whelan, M. (2014b). Adverse outcome pathway development II: best practices. *Toxicol Sci, 142*(2), 321-330. doi:10.1093/toxsci/kfu200

WHO. (2006). *Guidelines for drinking water quality. First addendum to 3rd addition*. Retrieved from Geneva, Switzerland:

WHO. (Nov 2015). Preterm birth. Retrieved from http://www.who.int/mediacentre/factsheets/fs363/en/

Wild, C. P. (2009). Environmental exposure measurement in cancer epidemiology. *Mutagenesis, 24*. doi:10.1093/mutage/gen061

Wilhelm-Benartzi, C. S., Houseman, E. A., Maccani, M. A., Poage, G. M., Koestler, D. C., Langevin, S. M., . . . Marsit, C. J. (2012). In utero exposures, infant growth, and DNA methylation of repetitive elements and developmentally related genes in human placenta. *Environ Heal Perspect, 120*. doi:10.1289/ehp.1103927

Yang, C.-Y., Chang, C.-C., Tsai, S.-S., Chuang, H.-Y., Ho, C.-K., & Wu, T.-N. (2003). Arsenic in drinking water and adverse pregnancy outcome in an arseniasis-endemic area in northeastern Taiwan. *Environmental Research, 91*(1), 29-34. doi:http://dx.doi.org/10.1016/S0013-9351(02)00015-4

Ye, X. (2008). Lysophospholipid signaling in the function and pathology of the reproductive system. *Hum Reprod Update, 14*. doi:10.1093/humupd/dmn023

Yoshino, Y., Yuan, B., Kaise, T., Takeichi, M., Tanaka, S., Hirano, T., . . . Toyoda, H. (2011). Contribution of aquaporin 9 and multidrug resistance-associated protein 2 to differential sensitivity to arsenite between primary cultured chorion and amnion cells prepared from human fetal membranes. *Toxicol Appl Pharmacol, 257*. doi:10.1016/j.taap.2011.09.006

Yu, X., Robinson, J. F., Gribble, E., Hong, S. W., Sidhu, J. S., & Faustman, E. M. (2008). Gene expression profiling analysis reveals arsenic-induced cell cycle arrest and apoptosis in p53-proficient and p53-deficient cells through differential gene pathways. *Toxicol Appl Pharmacol, 233*. doi:10.1016/j.taap.2008.09.016

Yue, J., Yokoyama, K., Balazs, L., Baker, D. L., Smalley, D., Pilquil, C., . . . Tigyi, G. (2004). Mice with transgenic overexpression of lipid phosphate phosphatase-1 display multiple organotypic deficits without alteration in circulating lysophosphatidate level. *Cell Signal, 16*. doi:10.1016/j.cellsig.2003.08.012

Z, X., C, M., A, R., & Y, H. (July 28-30, 2011). *Ontobee: A Linked Data Server and Browser for Ontology Terms.* Paper presented at the Proceedings of the 2nd International Conference on Biomedical Ontologies (ICBO), Buffalo, NY, USA.

Zhang, C., Liu, C., Li, D., Yao, N., Yuan, X., Yu, A., . . . Ma, X. (2010). Intracellular redox imbalance and extracellular amino acid metabolic abnormality contribute to arsenic-induced developmental retardation in mouse preimplantation embryos. *J Cell Physiol, 222*, 444-455. doi:http://dx.doi.org/10.1002/Jcp.21966

Zheng, J., Manduchi, E., & Stoeckert, C. (2013). *Development of an Application Ontology for Beta Cell Genomics Based on the Ontology for Biomedical Investigations*. Paper presented at the International Conference on Biomedical Ontologies (ICBO 2013), Montreal, Canada. http://www2.unb.ca/csas/data/ws/icbo2013/papers/research/icbo2013_submission_31.pdf