Integration of Adverse Outcome Pathway Information into the Biomedical Data Translator

A RENCI Technical Report
TR-25-01

Meisha Mandal
RTI International, Research Triangle Park, North Carolina, USA

Patricia Ceger
RTI International, Research Triangle Park, North Carolina, USA

Karamarie Fecho
Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA; Copperline Professional Solutions, LLC, Pittsboro, North Carolina, USA

Virginia Hench
RTI International, Research Triangle Park, North Carolina, USA

Gaurav Vaidya
Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

James P. Balhoff
Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

How to cite this paper: Mandal, M., Ceger, P., Fecho, K., Hench, V., Vaidya, G., & Balhoff, J.P. (2024). *Integration of Adverse Outcome Pathway Information into the Biomedical Data Translator.* Renaissance Computing Institute. https://doi.org/10.7921/76KE-BY69

Corresponding author: Meisha Mandal, RTI International, 3040 E. Cornwallis Road, P.O. Box 12194, Research Triangle Park, NC 27709; Voice: (919)-541-6161; Email: mmandal@rti.org.

renci

www.renci.org

Integration of Adverse Outcome Pathway Information into the Biomedical Data Translator

Meisha Mandal,[1] Patricia Ceger,[1] Karamarie Fecho,[2,3] Virginia Hench,[1]
Gaurav Vaidya,[2] James P. Balhoff[2]

[1]RTI International, Research Triangle Park, North Carolina, USA
[2]Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[3]Copperline Professional Solutions, LLC, Pittsboro, North Carolina, USA
November 2024

ABSTRACT

In this report, we describe the development and application of a semi-automated method for integrating mechanistic information from adverse outcome pathways (AOPs) into the Biomedical Data Translator ("Translator") ecosystem (Biomedical Data Translator Consortium, 2018; Fecho, Thessen, et al., 2022) as causal activity models (CAMs). These efforts provide a foundation for integrating AOPs and information on associated chemical safety regulatory endpoints and methods of measurement into the biomedical observations already within the Translator knowledge graph.

Keywords: *adverse outcome pathways (AOPs); AOP Event Components; Biomedical Data Translator; causal activity models; GO-CAMs; Integrated Clinical and Environmental Exposures Service (ICEES); Reasoning Over Biomedical Objects linked in Knowledge Oriented Pathways (ROBOKOP)*

INTRODUCTION

AOPs and Gene Ontology (GO)-CAMs are frameworks for modeling sequential and causal relationships between biological mechanisms in a manner that enhances computational reasoning of biological information. The Biomedical Data Translator ("Translator") program has created a diverse knowledge graph-based biomedical question-answering system through the integration and semantic harmonization of hundreds of knowledge sources, including clinical knowledge sources (Biomedical Data Translator Consortium, 2018; Fecho, Thessen, et al., 2022). Once AOPs are converted to the GO-CAM model and ingested into the broader Translator ecosystem, mechanistic information from AOPs can be used to enhance and explain insights derived from the Translator system. We describe a use case detailing how AOP integration into the Translator system can provide mechanistic explanations for clinical observations demonstrating a connection between exposure to particulate matter and exacerbation of asthma symptoms.

Adverse Outcome Pathways

The AOP framework was developed within the ecotoxicology field to link methods for measuring molecular and cellular events to adverse outcomes observed at the individual and population levels, thus offering a way to link *in vitro* testing results to potential impacts on wildlife to support chemical risk assessment (Ankley et al., 2010). Since being introduced in 2010, the AOP framework has been embraced by the wider toxicology community and many AOPs include biomechanistic pathways relevant to biomedical and clinical outcomes.

The AOP-Wiki is a knowledge base and repository for AOPs. Development of an AOP involves aggregation and organization of methods, evidence, and biological plausibility information supporting the core entities of an AOP, the key events (KEs), and the key event relationships (KERs). AOPs have an inherent graph-like quality where KEs resemble nodes and KERs serve as graph edges. KERs serve as causal linkages between an upstream KE and a downstream KE, with a default "leads to" relationship established between KEs within each KER.

Importantly, the AOP framework was not originally designed to support computability, but recent advances associated with the AOP-Wiki are advancing the computability of the AOP framework. The event component (EC) construct was introduced to the AOP-Wiki as a way of defining KEs in computable terms. ECs are modeled on the Resource Description Framework (RDF) triple (subject-predicate-object) format, consisting of a biological process, a biological object, and an action term. Each EC term is defined using terms from ontologies and controlled vocabularies (Ives et al., 2017). In the latest version of the AOP-Wiki (version 2.7), the EC biological process field includes terms that come from phenotype ontologies, like the Mammalian Phenotype Ontology (MPO) and the Vertebrate Trait Ontology (VTO), so there is a conflation between biological process and phenotype. However, a related effort exploring the application of the AOP framework for text annotation purposes has established distinct definitions for AOP EC processes versus AOP EC phenotypes.

In alignment with FAIR principles (Wilkinson et al., 2016), content from the AOP-Wiki is available in downloadable formats to facilitate analysis and information integration across the scientific community. TSV-formatted files are available for KEs, KERs and ECs. These files provide the ontology details for all EC terms in each KE, as well as the sequence of KEs (derived

from KERs), allowing for reconstruction of AOPs from ontology terms, an approach used within this report.

The Gene Ontology (GO)

Ontologies are structured vocabularies consisting of a set of terms and their relationships to one another within a particular domain. Ontologies are used to organize and classify information in a manner that makes complex language concepts amenable to computational reasoning and analysis. GO is an ontology dedicated to conceptualizing and describing gene function across multiple taxa and species in the tree of life (Aleksander et al., 2023). GO consists of three independent aspects to describe genes and gene products: biological processes, molecular functions, and cellular components (Ashburner et al., 2000).

GO Causal Activity Models

GO-CAMs build upon the GO framework to define causal relationships between distinct gene product activities by providing a structured model to link GO annotations, thereby providing a more complete representation of a biological system (Thomas et al., 2019). A standard GO annotation is a minimal and discrete piece of biological knowledge about a single gene containing the gene product name, a description of the gene product's function, and supporting scientific evidence. In contrast, a GO-CAM annotation describes causal relationships between GO annotations, meaning they can serve to describe the relationships between genes and their functions. The connections between GO terms within a GO-CAM are explicitly defined using the Relations Ontology (RO) (GitHub, n.d.-c), an ontology for standardizing descriptions of entity relationships across ontologies in the Open Biological and Biomedical Ontology (OBO) Foundry (Jackson et al., 2021). Whereas traditional annotations are usually flat associations, GO and GO-CAM annotations are based on the Web Ontology Language (OWL) (W3C, 2012), which supports logical consistency checks and semantic queries across biological ontologies. In this way, GO-CAMs provide a schema that can be scaled to define complex biology in computable terms.

Integrated Clinical and Environmental Exposures Service (ICEES)

ICEES contains electronic health record data that have been integrated at the patient level with environmental exposures data such as socioeconomic exposures and airborne pollutant exposures (Fecho et al., 2019). The service is openly available via an OpenAPI interface, command-line cURL request, and Neo4j or Translator Reasoner Standard API (TRAPI) (GitHub, n.d.-d) graph query request exposed in a regulatory-compliant manner and provided to users in semi-aggregated form. ICEES has been used to derive insights into asthma and other common pulmonary disorders, primary ciliary dyskinesia and related rare pulmonary disorders, and several other disorders (Fecho, Ahalt, Appold, et al., 2022; Fecho, Ahalt, Knowles, et al., 2022; Lan et al., 2021).

Reasoning Over Biomedical Objects linked in Knowledge Oriented Pathways (ROBOKOP)

The ROBOKOP system is an open-source knowledge graph (KG), a user interface and query tool, and a collection of harmonized and interoperable knowledge sources, represented as KGs in a service called Automat and accessible via application programming interfaces (APIs) (Bizon et al., 2019; Morton et al., 2019). ROBOKOP has been applied to suggest biological

mechanisms to explain real-world associations between workplace exposures and immune-mediated diseases, hypothesize a “clinical outcome pathway” to explain the effectiveness of nifedipine in the treatment of cardiovascular disease, and apply knowledge on treatments for the common disease psoriatic arthritis to suggest drugs for repurposing for the treatment of the rare disease chronic granulomatosis (Bizon et al., 2019; Fecho et al., 2019; Korn et al., 2022).

NCATS Biomedical Data Translator ("Translator")

ICEES and ROBOKOP were developed as part of the Biomedical Data Translator program. The Translator system has been applied to suggest novel treatments for cyclic vomiting disease, biological mechanisms to explain the clinically observed relationship between Crohn’s disease and Parkinson’s disease, drug candidates for repurposing in the treatment of drug-induced liver injury, and a variety of other translational insights and discoveries (Biomedical Data Translator Consortium, 2018; Fecho, Thessen, et al., 2022; Foksinska et al., 2022;).

The Translator architecture is both federated and hierarchical. When querying Translator, the query is sent to Translator’s autonomous relay system (ARS), which parses the query and distributes it to autonomous relay agents (ARAs). In turn, the ARAs distribute the query to select knowledge providers (KPs). KPs consist of domain-specific KGs, which are built on public databases and knowledge sources for multiple domains, serving as expert sources for that domain. ARAs are reasoning tools that access the KGs provided by different KPs and combine the data from many domains into a new KG. To respond to a query, the ARA creates this integrated graph, then finds a path through the new graph to provide answers to the query. Answers are returned to the ARS, where they are compiled, ranked, and displayed for the user. Of relevance to this report, both ICEES and ROBOKOP serve as Translator KPs.

Motivation for Transforming AOPs into CAMs in Support of ICEES, ROBOKOP, and the Translator Ecosystem

As part of the Translator program, we implemented a software pipeline to integrate the set of GO-CAMs curated by the Gene Ontology Consortium into a knowledge provider component of the Translator system, the CAM-KP. This pipeline applies an OWL reasoner to precompute implied relationships within the GO-CAMs and uses term mappings to convert the GO-CAM relationships into a KG that is compliant with the Biolink Model (Unni et al., 2022), the lingua franca of Translator. Because AOPs already include links to representative terms from bio-ontologies for each of their constituent parts, we can transform an AOP into a CAM built from those ontology terms (“CAM-ification”). By doing so, we can use our CAM pipeline to link the AOPs into the Translator system.

There are multiple reasons to integrate AOPs into the Translator system. First, doing so would allow us to formally connect different (previously unlinked) CAMs and thereby link molecular and cellular perturbations to adverse outcomes, with an emphasis on developing a common model for clinical assertions. Second, once developed, these AOPs will be used to integrate phenotypic data with molecular or cellular data represented via CAMs, with causal relationships based on existing biological knowledge and ordering of events providing the appropriate temporal context. Third, these integrated data can be provided to ARAs, along with the proper spatial and temporal context, to aid user interpretation of the results.

Herein, we describe our efforts to transform AOPs into a CAM framework to create a new Translator KP that can be used to derive causal mechanistic insights into clinical

observations derived from ICEES, ROBOKOP, and other Translator KPs. AOP transformation into CAMs makes information curated by AOP authors accessible through the larger-scale ICEES and ROBOKOP applications and the broader Translator ecosystem, thereby enabling complex querying of the data, as well as integration with the large base of information already accessible through ICEES, ROBOKOP, and the Translator program.

METHODS AND RESULTS

As stated previously, Ives et al. (2017) described an effort to apply ontology-based EC annotations to AOP KEs to make KEs more computable. The EC Process and Object entities describe the normal biology that is perturbed as part of the KE; the Action represents the perturbation that occurs (e.g., "increased" in a case where enzyme activity is inappropriately upregulated). A fourth EC has been proposed, Phenotype, to describe the observable characteristics of the individual, tissue, organ, cell, or biological system being perturbed (Hench et al., 2024). In the AOP-Wiki, ECs for each KE can be viewed in the Key Event Components table on each KE page. ECs, KEs, and KERs can be downloaded from the AOP-Wiki for uses such as those represented in Figure 1 and described as follows.

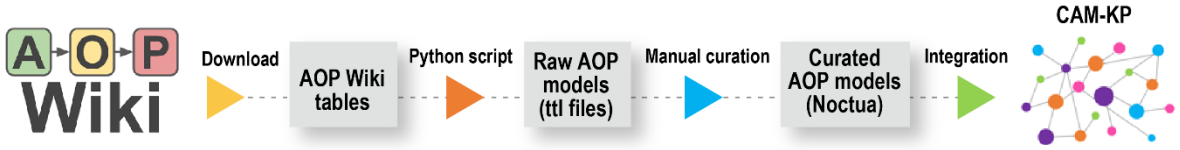

Figure 1. ***A representation of the workflow used to integrate AOPs from the AOP-Wiki into the ICEES CAM-KP.*** **AOPs are downloaded from the AOP-Wiki as tables, and Turtle (.ttl) models are generated using a Python script. These models are imported into Noctua and manually curated. Curated models are integrated into the CAM-KP.**

CAM-ification of AOPs

1. Download tables.

As the first step in the CAM-ification of AOPs, we downloaded the EC, KE, and KER tables from the AOP-Wiki download page (Society for the Advancement of Adverse Outcome Pathways, n.d.). Using the information in the AOP-Wiki tables (Figure 2), AOPs can be modeled as networks of ontology terms.

TABLE OF EVENTS			
	Key Events	**Key Event Relationships**	**Key Event Components**
File name	aop_ke_mie_ao.tsv	aop_ke_ker.tsv	aop_ke_ec.tsv
Description	Tab delimited file containing basic information relating to key events.	Tab delimited file containing basic information relating to key event relationships.	Tab delimited file containing basic information relating to key event components.
Columns	• aop id • key event id • key event type • key event name • aop id	• aop id • upstream event id • downstream event id • relationship id • direct or indirect relationship • evidence for relationship • quantitative understanding of relationship.	• aop id • key event id • action • object source • object ontology id • object term • process source • process ontology id • process term

Figure 2. ***Information contained in each downloadable AOP-Wiki table.***

2. Convert to Turtle files.

Next, we used a Python script that takes the downloaded tables from Step 1 as inputs and outputs a Terse RDF Triple Language (Turtle, or ttl) file (W3C, 2014) representing a model of the relationships in the AOP (Figure 1). In this process, each KE is converted into triple format (subject-predicate-object) based upon the ontology-based EC terms. For each AOP, the order of KEs is determined by the KER table. This table is then used in combination with the EC table to create relationships in the Turtle file.

During the creation of the Turtle file, we developed an automated process that was based on a decision tree (described in detail in the next section) to assign relationships within and between KEs. All edges/relationships in the Turtle file are selected from the Relations Ontology (RO) (GitHub, n.d.-c). The KE processing sequence is based upon the order of KEs within an AOP, determined by the sequence in the KER table. Each EC associated with a KE becomes a triple with relationship edges defined in the Turtle file, connecting an EC's object to its process/phenotype.

KER-derived relationships between KEs also inform creation of new CAMs based on ECs. Consistent with the default 'leads to' relationship inherent to all KERs, a CAM relationship is defined between the process/phenotype term(s) of an upstream KE and the object term(s) of the downstream KE. In cases where a process/phenotype term was not available in the upstream KE, the relationship would be established between the upstream KE object and the downstream KE object. Similarly, in cases where KEs do not have an object, the process/phenotype term is

used to connect to the preceding KE. In summary, connections between KEs are made using the following guidelines (in decreasing order of priority).

1) Upstream Process/Phenotype >> Downstream Object
2) Upstream Object >> Downstream Object
    - This is used if the upstream KE does not have a Process/Phenotype.
3) Upstream Object >> Downstream Process/Phenotype
    - This is used if the upstream KE does not have a Process/Phenotype and the downstream KE does not have an object.

3. Implementation of a decision tree during Turtle file creation (GitHub, n.d.-a).

To implement the decision tree, each relationship is structured as term1 <relates_to> term2. Term 1 and term 2 can be the object/process terms from the same KE or the connection between two KEs (e.g., term 1 is KE1's process/phenotype term, and term 2 is KE2's object term). The terms are then used to traverse the decision tree, ending in one or more RO terms representing suggested relationships to connect the terms. During the curation process, human reviewers resolve the edges with multiple suggested RO terms.

Decision points in the tree (Figure 3) are based on the following (listed in order of application):

- Are term 1 and term 2 objects or processes in the KE? (orange decision points)
- What is the source ontology of the term? (purple decision points)
    - Although the AOP does not distinguish between processes and phenotypes, we used the source ontology to categorize each process term as a process or phenotype. For example, process terms from phenotype ontologies (MP, Human Phenotype Ontology, VT) are classified as phenotypes. Process terms from GO are classified as processes.
    - All object terms are entities and can originate from any ontology.
- Action terms (green decision points)
    - When defining relationships between adjacent KEs, action terms for the two KEs can help distinguish between positive or negative effects. For example, if both action terms are the same, this would indicate a positive effect (e.g., positive regulation), whereas different actions would indicate a negative effect (e.g., negative regulation).
- Text mining (green decision points)
    - The text of a term can help distinguish between process and phenotype if the source ontology is not on the list of known phenotype ontologies or process ontologies. Following are examples of this:
        - If the term contains "osis," it is likely a process term, not a phenotype term.
        - The presence of "biosynthesis" or "generation" would suggest that the relationship is "output of."
        - If term 1 is contained within term 2, that would suggest a relationship of "enables" (e.g., enzyme A enables enzyme A's activity).

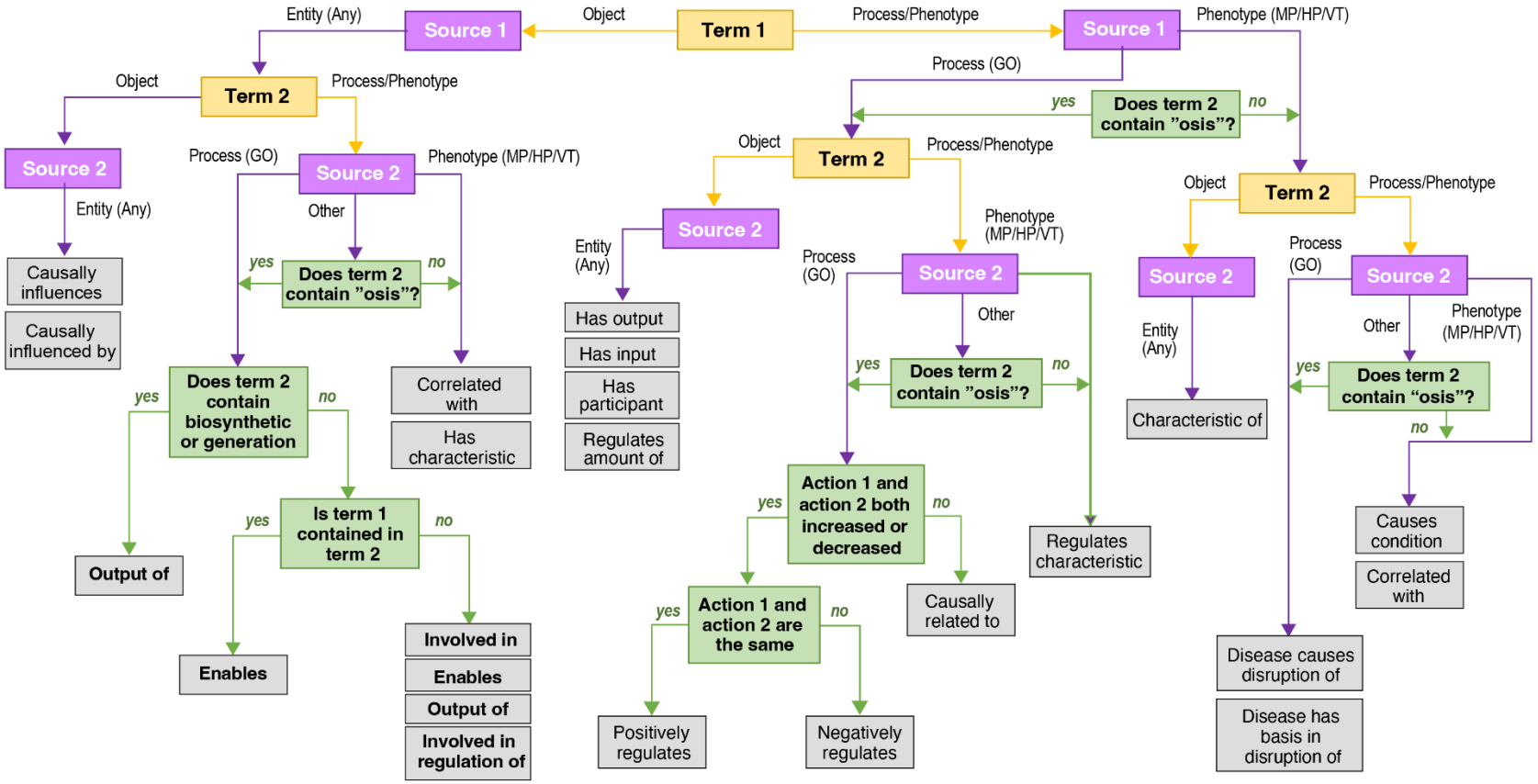

Figure 3. ***A diagram of the decision tree used to assign relationship ontology (RO) terms to edges between two KE terms (Term 1 and Term 2). The orange boxes and arrows represent decision points based on the KE term type (i.e., object or process/phenotype). The purple boxes represent decisions based on the source ontology for the KE term (i.e., GO, MPO, HPO). The green boxes and arrows are decision points based on text mining (i.e., does Term 2 contain "osis") and KE action terms. Note that the endpoints on the decision tree can be a single RO term or multiple RO terms.***

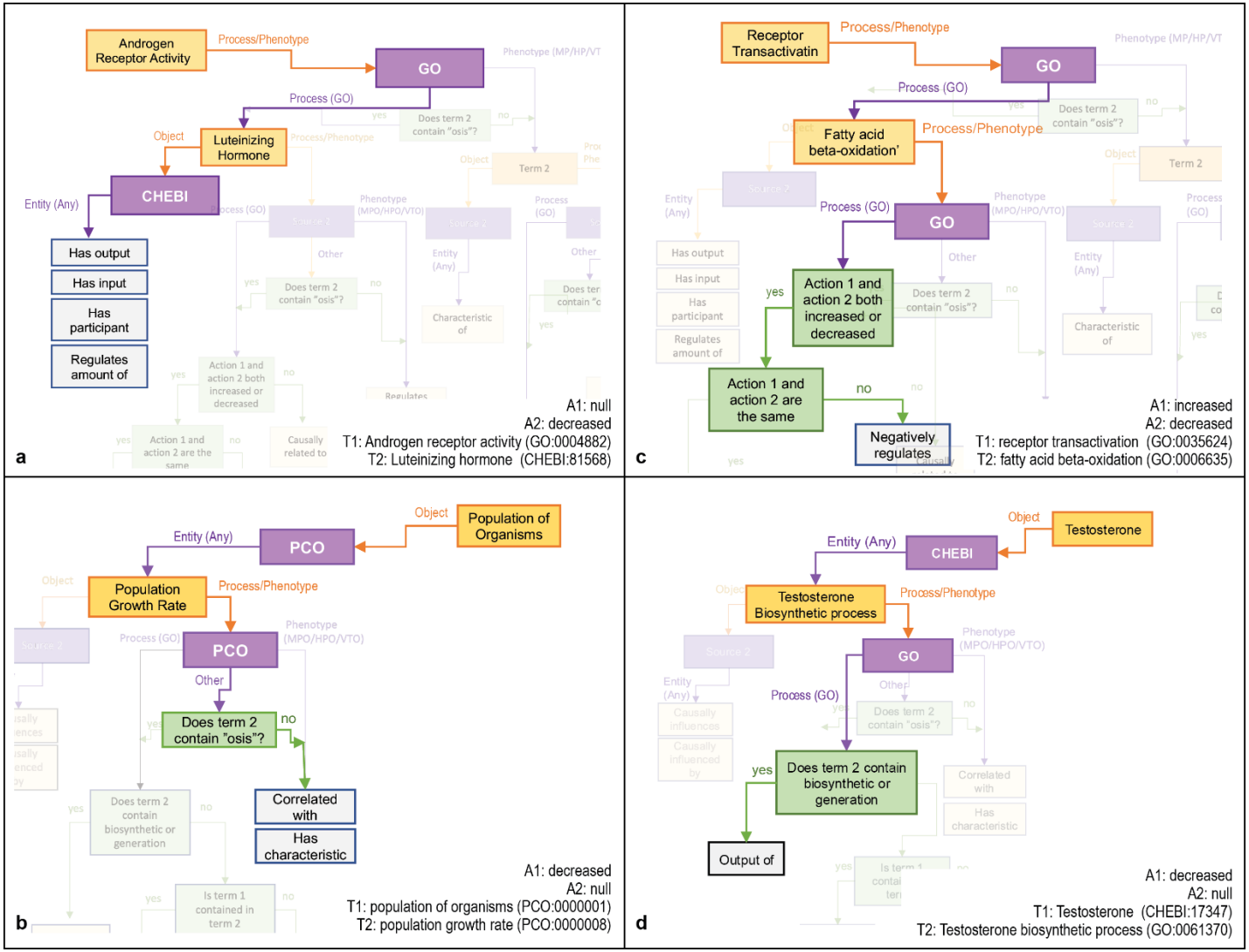

Figure 4. ***Examples of using the decision tree to choose relationship ontology terms.*** **T1 and T2 (orange boxes) are the KE object and process/phenotype terms; A1 and A2 are the KE action terms. a) Relationship connecting "Androgen Receptor Activity" to "Luteinizing Hormone". b) Relationship connecting "Population of Organisms" to "Population Growth Rate". c) Relationship connecting "Receptor Transactivation" and "Fatty acid beta-oxidation". d) Relationship connecting "Testosterone" and "Testosterone biosynthetic process".**

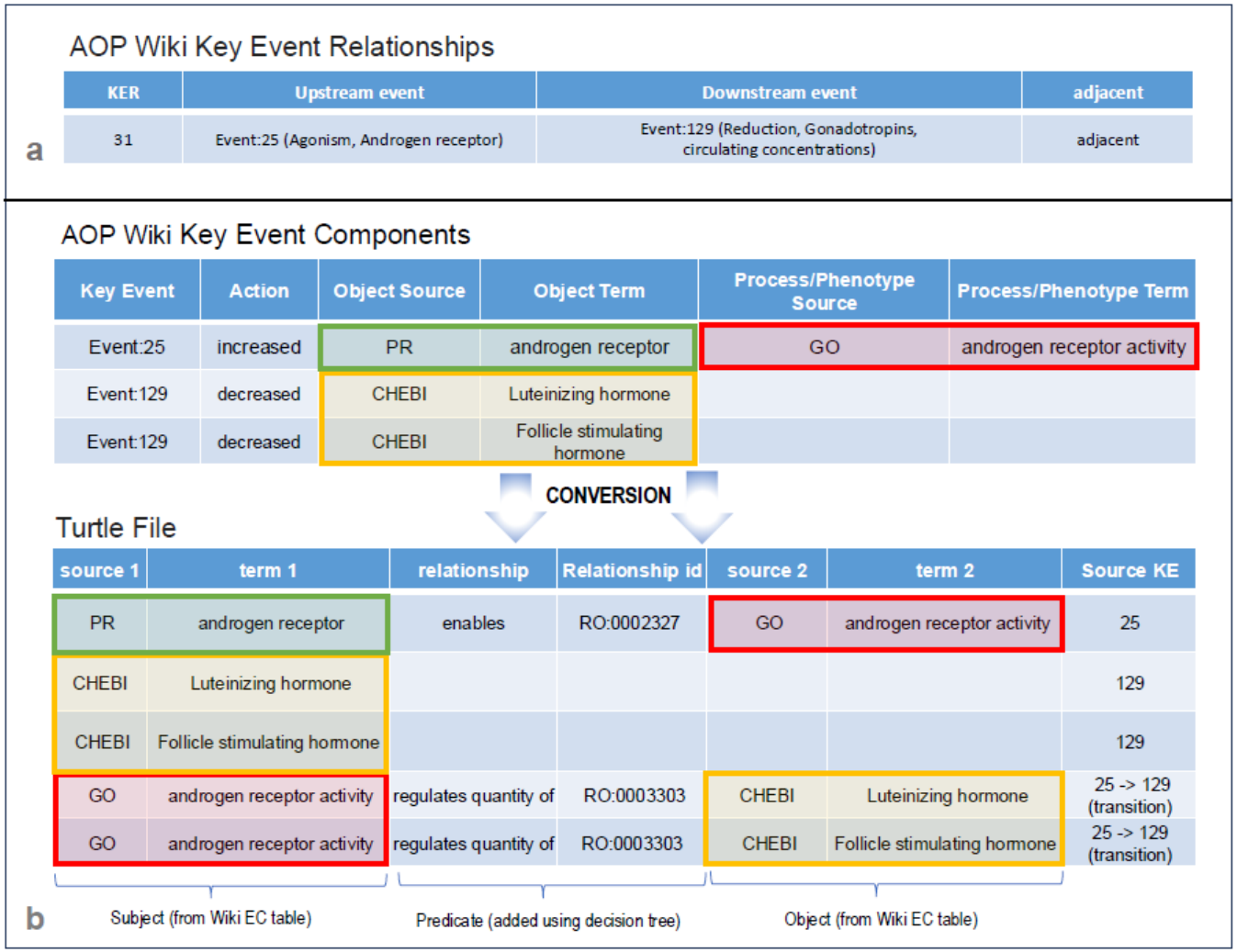

Figure 5. ***An example of the generation of a Turtle representation of two KEs in AOP 23 using the AOP-Wiki tables and the decision tree (see Figure 3).*** **a) The KE Relationships table in panel a indicates that KE 129 directly follows KE 25. b) The decision tree is used to convert the AOP to a Turtle representation. For intra-KE relationships, each row in the KE Components table is used to generate a row in the Turtle model using the decision tree. KE object terms are linked to process/phenotype terms with an RO predicate (rows 1–3). Inter-KE relationships connecting KE 25 and KE 129 are then generated. Specifically, the KE 25 process/phenotype term and the KE 129 object terms are linked with RO predicates (rows 4–5).**

Curation

Once the GO-CAM AOP models were generated, we used a custom installation of the GO modeling and visualization tool, Noctua (GitHub, n.d.-b), to view the models. One member of our team manually curated each model; a different member of our team then reviewed the model. During the curation process, curators review all the nodes (represented in Figure 6 as boxes) and edges (represented in Figure 6 as connecting lines) created by the "CAM-ification" algorithm and clean up the models (Figure 6b). They ensure that the term assigned to each edge is the "best fit" RO term. In addition to correcting nodes that the curator deems incorrect, they also replace edges that are very generic with more specific terms if possible. Additionally, before curation, some node pairs have multiple edges connecting them because the decision tree has multiple terms at some endpoints (Figure 6a). In those cases, curators decide which RO term is the "best fit" and remove the others. The curator may also decide the "best fit" is another RO term that was not "suggested" by the decision tree (e.g., the curator may remove "subject participant in" and replace it with the more specific "input of"). In this case, existing edges are removed, and a new edge is created with the curator-selected term.

To improve the consistency of our curation process, we developed a standard operating procedure (SOP) containing guidelines for curators. One main purpose of the SOP is to assist curators in selecting the "best fit" term during the curation process. The SOP contains general guidelines based on EC type, as well as some narrower guidelines that apply to specific cases. For EC type-based guidelines, the SOP generally contains examples of RO terms that can connect components of different types. The SOP helps facilitate the process; however, the curator ultimately decides which RO term is the "best fit" for an edge. Some examples are outlined as follows:

EC type-based guidelines
Process-to-process edge
    Suggested RO terms: positively regulates, negatively regulates
Phenotype-to-phenotype edge
    Suggested RO terms: positively correlated with, causally influences, causes condition
Entity-to-entity edge
    Suggested RO terms: bearer of, regulates activity or quantity of, location of

After curation, a different curator reviews the model. Once reviewed, the model is set to "Production" status in Noctua. Models represented in RDF are then ingested into CAM-KP by the CAM pipeline, which maps from an instance-based representation of a biochemical process (e.g., an instance of chemical X upregulates an instance of gene Y in the cytoplasm of liver cells) to a higher-order representation (in this example, that chemical X upregulates gene Y in the cytoplasm of liver cells). The pipeline also converts RO-based predicates into simpler Biolink predicates and converts information about where the reaction takes place into Biolink qualifiers. The pipeline produces a tab-delimited file containing subject and object nodes, predicate (as a Biolink predicate), and qualifiers (as a JSON string of qualifier/value pairs). This data format is then ingested by a Translator tool called ORION (GitHub, n.d.-e), which further normalizes node identifiers and repackages the data as a Neo4j database, which is then hosted by a second Translator tool called Plater. The Plater-hosted Neo4j database of the CAM-KP KG can be queried by any other Translator tool via the TRAPI Translator query API.

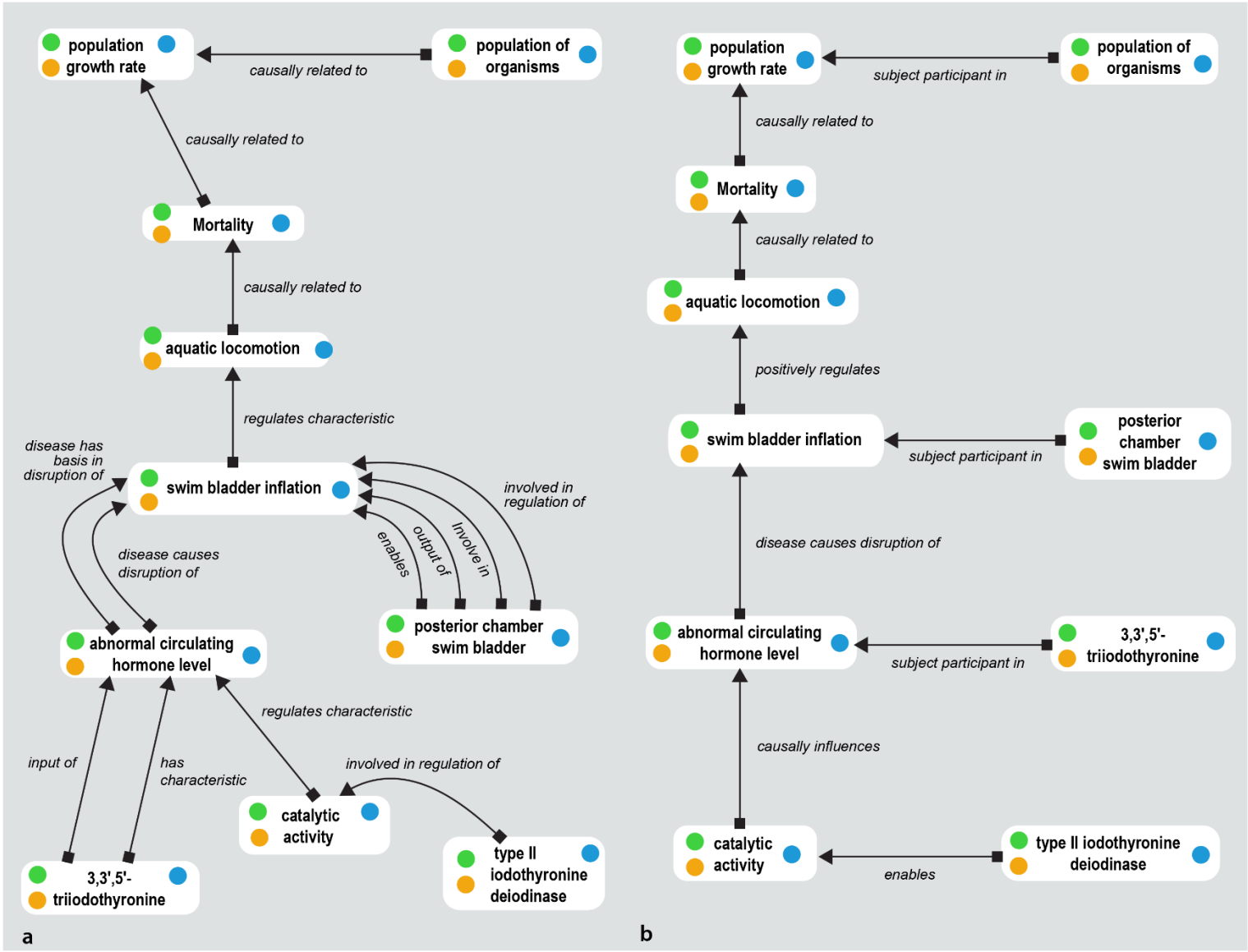

Figure 6. ***Model of AOP 155 displayed in Noctua before (a) and after (b) manual curation.*** **Note that before curation, multiple edges are connecting some pairs of nodes. After curation, generally only one edge exists between any pair of connected nodes.**

CAM-KP Testing

To facilitate testing and promote external use, we built a CAM-KP user interface to directly query the underlying Neo4j database used by the CAM-KP Plater instance.

Our testing process began with an AOP model as represented as a GO-CAM in Noctua. We could provide the URL of this model to CAM-KP Frontend, which would retrieve the CAM-KP representation of the same model from CAM-KP as tables of relationships and edges, which we could then compare with the original AOP model to ensure the following:

1. All nodes from the AOP model were included in the Biolink model.

2. All edges from the AOP model were included in the Biolink model and connected to the correct nodes.

3. The edges correctly describe the relationships between nodes when compared with those in the AOP model.

Initially, testing consisted of querying CAM-KP for known edges to see whether they were correctly returned by CAM-KP, but we quickly realized that three error classes needed to be identified and fixed:

- CAMs are expressed using RO predicates, but CAM-KP (like all Translator KPs) uses Biolink predicates. Thus, we had to test that we were correctly mapping from RO to Biolink predicates.
- Like all Translator KPs, CAM-KP was normalized to a certain subset of all possible node identifiers, so identifiers that cannot be normalized by the Translator Node Normalization tool will be ignored by Translator. Where a node identifier from a CAM could not be normalized, we had to identify the type of identifier we needed to have integrated into Translator, then work with the appropriate Translator team to integrate it.
- Like all Translator KPs, CAM-KP needed to be able to make general statements about named entities (e.g., [protein X] upregulates [gene Y]). However, CAMs are instance-based (i.e., [an instance of protein X] upregulates [an instance of gene Y]), so these instance-based statements needed to be generalized during this transformation. We needed to ensure that the CAM-KP representation of an edge expressed the same meaning as that same edge in the CAM, although it was more broadly generalized than the CAM edge.

CASE STUDY: EXPOSURE TO PARTICULATE MATTER AND ASTHMA EXACERBATIONS

Our driving use case for the work described here involved an effort to integrate ICEES, ROBOKOP, and CAM-AOP-KP to gain mechanistic insights into the role of environmental exposures on health outcomes.

Specifically, we initially queried ICEES to explore the relationship between airborne pollutant exposures and asthma exacerbations among approximately 141,000 UNC Health patients with a diagnosis of asthma or another common pulmonary disorder (Fecho et al., 2019). As previously reported (Fecho, Ahalt, Appold, et al., 2022), we found that patients exposed to

relatively high levels of particulate matter ≤ 2.5 microns in diameter (PM2.5) are more likely to experience asthma exacerbations than patients exposed to relatively low levels of particulate matter (**Figure 7**). This relationship held for two measures of asthma exacerbations: annual visits to the emergency department or inpatient appointments for respiratory issues and annual prescriptions or administrations for prednisone.

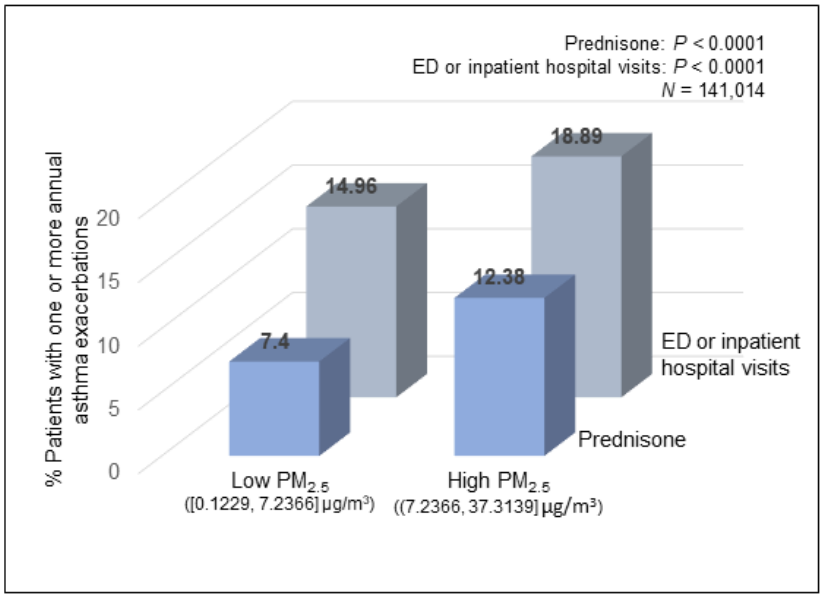

Figure 7. ***Three-dimensional graph of ICEES-derived data depicting the percentage of patients with one more annual asthma exacerbations (reported as emergency department or inpatient hospital visits or prednisone use) and low or high exposure to PM2.5.***

We then queried ROBOKOP to determine whether ROBOKOP could suggest biological mechanisms to explain the relationship between airborne PM2.5 exposure and asthma exacerbations. We iteratively queried ROBOKOP to demonstrate a relationship between particulate matter, toll-like receptor 4 (TLR4), myeloid differentiation primary response protein 88 (MYD88), interleukin-1 (IL-1) receptor binding, IL-1β, and asthma. When we explored the edges or predicates connecting each node in the linear pathway, we identified the following path suggestive of a putative AOP linking particulate matter to asthma:

> *particulate matter – increases response to – TLR4 – directly physically interacts with – MYD88 – catalyzes – IL-1 receptor binding – catalyzes – IL-1β – is related to – asthma*

In natural language, the path translates loosely as follows:

> *Particulate matter increases the response to TLR4, which directly physically interacts with MYD88, which catalyzes IL-1β binding to the IL-1 receptor, which is related to asthma.*

We note that other edges were present in the ROBOKOP subgraph. For instance, there were several edges between IL-1β and asthma: "related to"; "genetically associated with (3 edges)"; and "occurs together in literature with". The "genetically associated with" edges were from three primary knowledge sources, DisGeNet, Comparative Toxicogenomics Database, and DISEASES, and contributed by the aggregator knowledge source, Pharos. However, those edges did not contain supporting publications or any information indicating a causal genetic

relationship between IL-1β and asthma. As such, we focused on the less specific "related to" relationship.

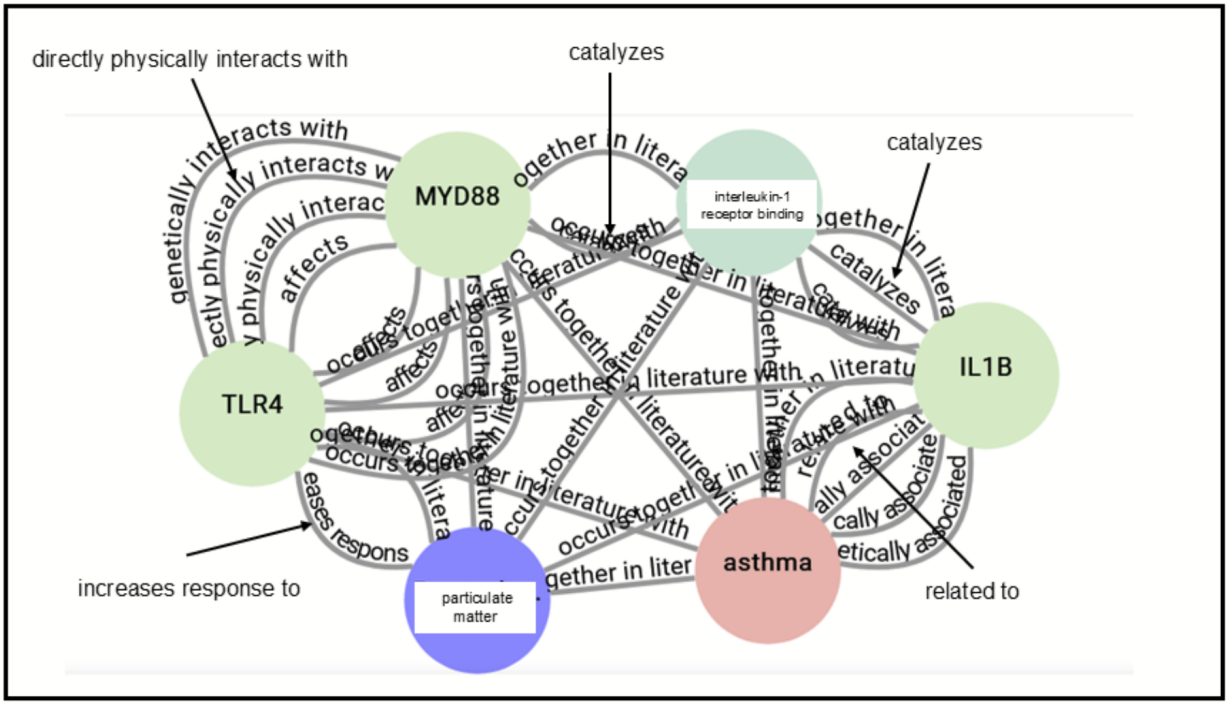

Figure 8. ***ROBOKOP results asserting, in part, that particulate matter increases the response to TLR4, which directly physically interacts with MYD88, which catalyzes IL-1β binding to the IL-1 receptor, with IL-1β being genetically associated with asthma.***

Recognizing that the putative AOP generated by ROBOKOP was high level and somewhat incomplete, we consulted the AOP-Wiki to determine whether we could identify more granular causal relationships that might explain the connection between particulate matter and asthma, focusing on the TLR4/IL-1 signaling pathway. We identified the OECD-endorsed AOP #277 (Kimura et al., 2023) for providing additional details on the IL-1 receptor signaling pathway and a layer of granularity to the ROBOKOP findings. Specifically, AOP-Wiki #277 stated:

> *Decreased IL-2 activity leads to decreased NF-κβ signaling, which leads to decreased T cell activation, which leads to decreased T cell–dependent antibody response.*

However, when we explored the underlying AOP-Wiki knowledgebase (KB), we found that AOP #277 did not have ECs for all KEs, so we couldn't represent the entire AOP using terms from ontologies and controlled vocabularies thereby limiting computability and motivating the current work.

Indeed, after we transformed AOP #277 from the AOP-Wiki KB into a CAM and added it to the CAM-AOP-KP, we were able to directly query the CAM-AOP-KP and obtain additional causal details to support the putative AOP that ROBOKOP provided in response to the ICEES clinical observation of an association between exposure to airborne particulate matter and asthma exacerbations. Indeed, queries of CAM-AOP-KP demonstrated in natural language that:

IL-1 activity positively regulates NF-κβ signal transduction, which positively regulates T cell activation, which positively regulates immunosuppression.

Combined, we were able to generate an expanded, putative AOP that, while still incomplete, asserts the following:

Particulate matter (high-level exposure) increases response to TLR4, which directly physically interacts with MYD88, which catalyzes IL-1B receptor binding to IL-1 receptor, which is related to asthma (exacerbations) and (presumably) "is related to" or "is associated with" or "enables" or "participates in" IL-1 activity, which positively regulates NF-κβ signal transduction, which positively regulates T cell activation, which positively regulates immunosuppression, which (presumably) "is related to" or "is associated with" asthma (exacerbations).

Thus, through the integration of three sources of "knowledge" (ROBOKOP, AOP-Wiki, and CAM-AOP-KP), we have provided a more complete putative AOP than any one source could provide alone to explain the real-world ICEES observation of an association between exposure to high levels of particulate matter and asthma exacerbations (Figure 9). One missing link is that between IL1-B binding to the IL-1 receptor and IL-1 activity, which presumably reflects an activation of IL-1 signaling in the form of "related to" or "associated with", or "enables", or "participates in" relationships, as defined by Biolink Model (Glen, n.d.). A second missing link is between immunosuppression and asthma. While the second missing link as a causal link may seem counterintuitive, especially for certain subtypes of asthma such as allergic asthma, several explanations are possible. First, the immune system functions along with other body systems to maintain homeostasis within the body (Goldstein, 2019). As such, apparent paradoxes may actually reflect factors not encapsulated by the knowledge reported here such as duration of an environmental exposure or "dose". Second, and perhaps more significant, the role of the immune system, and T cells in particular, in asthma is complex, with evidence for and against T cell involvement in contributing to asthma exacerbations (Bryant & Muehling, 2022). For instance, a suppression of regulatory T cell activity can lead to an increased inflammatory response and thus asthma exacerbations (Zhang et al., 2022). We are reviewing additional sources of knowledge, including the AOP-Wiki, to complete the putative AOP that we identified through the integration of ICEES, ROBOKOP, AOP-Wiki #277, and the CAM-AOP version of AOP-Wiki #277.

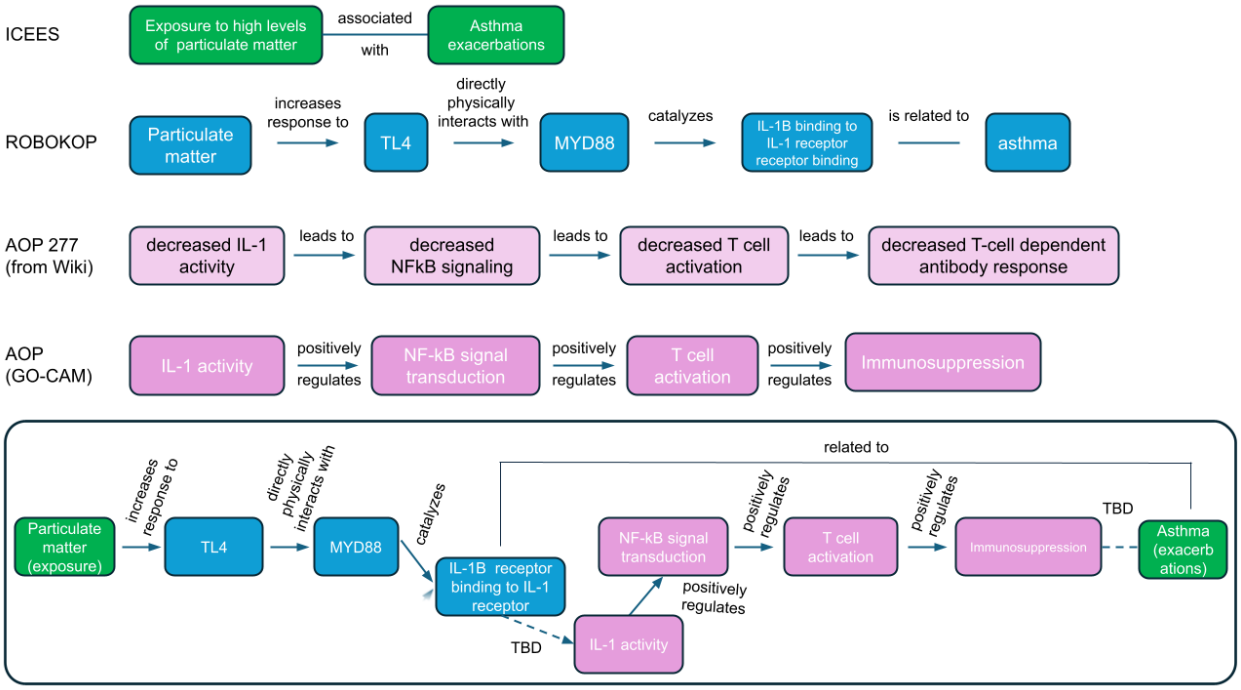

Figure 9. ***A depiction of how the information obtained from the different sources fits together in the use case. TBD edges are presumably "related to" or "associated with" or "enables" or "participates in."***

BENEFITS FOR AOP DEVELOPMENT

The work carried out here offers significant benefits for the AOP community, especially those working to advance AOP network analytics and computational approaches to AOP development. By building upon the AOP-Wiki's EC concept, the AOP CAM-ification process introduced here allows for complex biological mechanisms to be modeled with high resolution in a manner that can be applied to causal relationships between KEs and to mechanistic steps covered under a single KE. In this way, modeling the biological complexity of AOPs using GO-CAMs could help reveal areas of interaction between AOPs and accelerate AOP network development. For example, while exploring the use case above, our team identified areas of overlap shared between AOP #277 and an AOP still under development, AOP #39, which describes a process for respiratory tract sensitization. Both AOPs describe perturbations of T-cell based immune responses, but since they do not share any KEs in common, they would not fit into a classically defined AOP network, which requires the occurrence of shared KE's between two or more AOPs (Knapen et al., 2018). Importantly, the free text descriptions of the upstream KEs reveal how both AOPs describe perturbations of cytokine signaling mechanisms, even though each AOP leads to a different immune-based adverse outcome. While more work is needed to bridge gaps between the AOP framework and AOP CAM-ification, such efforts could greatly benefit efforts to expand AOPs and new approach methods in specific domains, especially those relevant to immunotoxicity (Snapkow et al., 2024).

Another small but tangible benefit of this effort to the AOP community came in the form how CAM-ification of AOP #277 exposed some gaps that our team was able to remedy. As mentioned above, when AOP #277 first drew our attention, it did not have ECs for all KEs. Our team was able to identify a new EC for KE #1700, Suppression of T cell activation, make a

recommendation to the contact author of AOP #277 and add the new EC to the AOP-Wiki. The new EC for KE #1700 has interleukin-1 receptor activity, GO:0004908 as the biological process term, interleukin-1 receptor type I, PR:000001363 as the biological object, and "decreased" as the action term.

SUMMARY

In summary, we developed a semi-automated method for converting AOP models into GO-CAM models and subsequently integrating the information into the larger Translator ecosystem. The method involves an automated conversion step, manual curation, and integration into Translator.

AOPs contain links between molecular and cellular events and adverse outcomes. ICEES contains data on environmental exposures and clinical outcomes, and ROBOKOP can be used to suggest mechanisms by which exposures and outcomes may be associated. By ingesting information from AOPs (as CAM-KP-AOPs) into Translator, we aimed to expand beyond the ICEES clinical observations and ROBOKOP mechanistic paths to provide causal details and relationships.

We successfully CAM-ified 44 AOP models into GO-CAMs and incorporated them into a new Translator KP, CAM-AOP KP. This included all 31 "Endorsed" and 13/17 "Approved and Under Review" AOPs. The remaining AOPs contained in the AOP-Wiki were still under development and were not complete enough to convert to GO-CAMs.

In the case study presented here, we integrated information from ROBOKOP, ICEES, and CAM-AOP-KP and used it to delineate a potential mechanistic pathway connecting exposure to high levels of particulate matter with asthma exacerbations in a real-world clinical cohort. Specifically, particulate matter increases the response to TLR4, which directly physically interacts with MYD88, which catalyzes IL-1B receptor binding to the IL-1 receptor. Decreased IL-1 activity leads to decreased NF-kβ signaling, which leads to decreased T cell activation, which leads to immunosuppression, which is associated with asthma.

Collectively, our findings suggest that further development of AOPs, coupled with CAM-ification and incorporation into Translator via CAM-AOP KP, may be useful for the generation of hypotheses linking environmental exposures through mechanistic events to clinical events or outcomes.

REFERENCES

Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N. L., Hill, D. P., Lee, R., Mi, H., Moxon, S., Mungall, C. J., Muruganugan, A., Mushayahama, T., Sternberg, P. W., Thomas, P. D., Van Auken, K., … Westerfield, M. (2023). The Gene Ontology knowledgebase in 2023. *Genetics*, *224*(1). https://doi.org/10.1093/genetics/iyad031

Ankley, G. T., Bennett, R. S., Erickson, R. J., Hoff, D. J., Hornung, M. W., Johnson, R. D., Mount, D. R., Nichols, J. W., Russom, C. L., Schmieder, P. K., Serrano, J. A., Tietge, J. E., & 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. https://doi.org/10.1002/etc.34

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Biomedical Data Translator Consortium. (2018). Toward a universal biomedical data translator. *Clinical and Translational Science*, *12*(2), 86–90. https://doi.org/10.1111/cts.12591

Bizon, C., Cox, S., Balhoff, J., Kebede, Y., Wang, P., Morton, K., Fecho, K., & Tropsha, A. (2019). ROBOKOP KG and KGB: Integrated knowledge graphs from federated sources. *Journal of Chemical Information and Modeling*, *59*(12), 4968–4973. https://doi.org/10.1021/acs.jcim.9b00683

Bryant, N., & Muehling, L. M. (2022). T-cell responses in asthma exacerbations. *Annals of Allergy, Asthma & Immunology*, *129*(6), 709–718. https://doi.org/10.1016/j.anai.2022.07.027

Fecho, K., Ahalt, S. C., Appold, S., Arunachalam, S., Pfaff, E., Stillwell, L., Valencia, A., Xu, H., & Peden, D. (2022). Development and application of an open tool for sharing and analyzing integrated clinical and environmental exposures data: Asthma use case. *JMIR Formative Research*, *6*(4), e32357. https://doi.org/10.2196/32357

Fecho, K., Ahalt, S. C., Knowles, M., Krishnamurthy, A., Leigh, M., Morton, K., Pfaff, E., Wang, M., & Yi, H. (2022). Leveraging open electronic health record data and environmental exposures data to derive insights into rare pulmonary disease. *Frontiers in Artificial Intelligence*, *5*, 918888. https://doi.org/10.3389/frai.2022.918888

Fecho, K., Pfaff, E., Xu, H., Champion, J., Cox, S., Stillwell, L., Bizon, C., Peden, D., Krishnamurthy, A., Tropsha, A., & Ahalt, S. C. (2019). A novel approach for exposing and sharing clinical data: The Translator Integrated Clinical and Environmental Exposures Service. *Journal of the American Medical Informatics Association*, *26*(10), 1064–1073. https://doi.org/10.1093/jamia/ocz042

Fecho, K., Thessen, A. E., Baranzini, S. E., Bizon, C., Hadlock, J. J., Huang, S., Roper, R. T., Southall, N., Ta, C., Watkins, P. B., Williams, M. D., Xu, H., Byrd, W., Dančík, V., Duby, M. P., Dumontier, M., Glusman, G., Harris, N. L., Hinderer, E. W., … The Biomedical Data Translator Consortium. (2022). Progress toward a universal biomedical data translator. *Clinical and Translational Science*, *15*(8), 1838–1847. https://doi.org/10.1111/cts.13301

Foksinska, A., Crowder, C. M., Crouse, A. B., Henrikson, J., Byrd, W. E., Rosenblatt, G., Patton, M. J., He, K., Tran-Nguyen, T. K., Zheng, M., Ramsey, S. A., Amin, N., Osborne, J., & Might, M. (2022). The precision medicine process for treating rare disease using the artificial intelligence tool mediKanren. *Frontiers in Artificial Intelligence*, *5*. https://doi.org/10.3389/frai.2022.910216

GitHub. (n.d.-a). *NCATS_Translator_MM*. https://github.com/mam288/NCATS_Translator_MM/blob/7bee36b75564d7bdad94b08ffd5f08b3a79f6df5/scripts/write_ttl.py

GitHub. (n.d.-b). *Noctua*. https://github.com/geneontology/noctua

GitHub. (n.d.-c). *OBO Relation Ontology*. https://github.com/oborel/obo-relations

GitHub. (n.d.-d). *ReasonerAPI*. https://github.com/NCATSTranslator/ReasonerAPI

GitHub. (n.d.-e). *RobokopU24: ORION*. https://github.com/RobokopU24/ORION

Glen, A. (n.d.). *Visualizations of the Biolink Model*. http://tree-viz-biolink.herokuapp.com/predicates

Goldstein, D. S. (2019). How does homeostasis happen? Integrative physiological, systems biological, and evolutionary perspectives. *American Journal of Physiology—Regulatory Integrative and Comparative Physiology*, *316*(4), R301–R317. https://doi.org/10.1152/ajpregu.00396.2018

Hench, G., Edwards, S. W., Lynn, S. G., & Villeneuve, D. (2024). *Integrated application and transformation of the adverse outcome pathway (AOP) framework for modeling evidence and advancing new approach methodologies (NAMs)*. Poster presentation at Intelligent Systems for Molecular Biology (ISMB), Montreal, Quebec, Canada.

Ives, C., Campia, I., Wang, R.-L., Wittwehr, C., & Edwards, S. (2017). Creating a structured AOP knowledgebase via ontology-based annotations. *Applied In Vitro Toxicology*, *3*(4), 298–311. https://doi.org/10.1089/aivt.2017.0017

Jackson, R., Matentzoglu, N., Overton, J. A., Vita, R., Balhoff, J. P., Buttigieg, P. L., Carbon, S., Courtot, M., Diehl, A. D., Dooley, D. M., Duncan, W. D., Harris, N. L., Haendel, M. A., Lewis, S. E., Natale, D. A., Osumi-Sutherland, D., Ruttenberg, A., Schriml, L. M., Smith, B., … Peters, B. (2021). OBO Foundry in 2012: Operationalizing open data principles to evaluate ontologies. *Database*. https://doi.org/10.1093/database/baab069

Kimura, Y., Aiba, S., Ashikaga, T., Ohishi, T., & Kushima, K. (2023). *AOP 277: Impaired IL-1R1 signaling leading to impaired T-cell dependent antibody response*. AOP-Wiki. https://aopwiki.org/aops/277

Knapen, D., Angrish, M. M., Fortin, M. C., Katsiadaki, I., Leonard, M., Margiotta-Casaluci, L., Munn, S., O'Brien, J. M., Pollesch, N., Smith, L. C., Zhang, X., & Villeneuve, D. L. (2018). Adverse outcome pathway networks I: Development and applications. *Environmental Toxicology and Chemistry*, *37*, 1723–1733. https://doi.org/10.1002/etc.4125

Korn, D., Thieme, A. J., Alves, V. M., Yeakey, M., Borba, J. V. V. B., Capuzzi, S. J., Fecho, K., Bizon, C., Edwards, S. W., Chirkova, R., Colvis, C. M., Southall, N. T., Austin, C. P., Muratov, E. N., & Tropsha, A. (2022). Defining clinical outcome pathways. *Drug Discovery Today*, *27*(6), 1671–1678. https://doi.org/10.1016/j.drudis.2022.02.008

Lan, B., Haaland, P., Krishnamurthy, A., Peden, D. B., Schmitt, P. L., Sharma, P., Sinha, M., Xu, H., & Fecho, K. (2021). Open application of statistical and machine learning models to explore the impact of environmental exposures on health and disease: An asthma use case. *International Journal of Environmental Research and Public Health*, *18*(21), 11398. https://doi.org/10.3390/ijerph182111398

Morton, K., Wang, P., Bizon, C., Cox, S., Balhoff, J., Kebede, Y., Fecho, K., & Tropsha, A. (2019). ROBOKOP: An abstraction layer and user interface for knowledge graphs to support question answering. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btz604

Snapkow, I., Smith, N. M., Arnesdotter, E., Beekmann, K., Blanc, E. B., Braeuning, A., Corsini, E., Sollner Dolenc, M., Duivenvoorde, L. P. M., Sundstøl Eriksen, G., Franko, N., Galbiati, V., Gostner, J. M., Grova, N., Gutleb, A. C., Hargitai, R., Janssen, A. W. F., Krapf, S. A., Lindeman, B., … Barouki, R. (2024). New approach methodologies to enhance human health risk assessment of immunotoxic properties of chemicals—a PARC (Partnership for the Assessment of Risk from Chemicals) project. *Frontiers in Toxicology*, *6*. https://doi.org/10.3389/ftox.2024.1339104

Society for the Advancement of Adverse Outcome Pathways (SAAOP). (n.d.). *AOP-Wiki: Downloads*. https://aopwiki.org/info_pages/5

Thomas, P. D., Hill, D. P., Mi, H., Osumi-Sutherland, D., Van Auken, K., Carbon, S., Balhoff, J. P., Albou, L.-P., Good, B., Gaudet, P., Lewis, S. E., & Mungall, C. J. (2019). Gene Ontology Causal Activity Modeling (GO-CAM) moves beyond GO annotations to structured descriptions of biological functions and systems. *Nature Genetics*, *51*(10), 1429–1433. https://doi.org/10.1038/s41588-019-0500-1

Unni, D. R., Moxon, S. A. T., Bada, M., Brush, M., Bruskiewich, R., Caufield, J. H., Clemons, P. A., Dancik, V., Dumontier, M., Fecho, K., Glusman, G., Hadlock, J. J., Harris, N. L., Joshi, A., Putman, T., Qin, G., Ramsey, S. A., Shefchek, K. A., Solbrig, H., … Zisk, T. (2022). Biolink Model: A universal schema for knowledge graphs in clinical, biomedical, and translational science. *Clinical and Translational Science*, *15*(8), 1848–1855. https://doi.org/10.1111/cts.13302

W3C. (2012, December 11). *OWL 2 web ontology language document overview (second edition)*. https://www.w3.org/TR/owl2-overview/

W3C. (2014, February 25). *RDF 1.1 turtle*. https://www.w3.org/TR/turtle/

Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Bonino da Silva Santos, L., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*. https://doi.org/10.1038/sdata.2016.18

Zhang, J., Zou, Y., Chen, L., Xu, Q., Wang, Y., Xie, M., Liu, X., Zhao, J., & Wang, C. Y. (2022). Regulatory T cells, a viable target against airway allergic inflammatory responses in asthma. *Frontiers in Immunology*, *13*. https://doi.org/10.3389/fimmu.2022.902318