Capturing a Comprehensive Picture of Biological Events From Adverse Outcome Pathways in the Drug Exposome

Qier Wu[1], Youcef Bagdad[1], Olivier Taboureau[2] and Karine Audouze[1]*

[1] INSERM U1124, CNRS ERL3649, Université de Paris, Paris, France, [2] INSERM U1133, CNRS UMR8251, Université de Paris, Paris, France

Background: The chemical part of the exposome, including drugs, may explain the increase of health effects with outcomes such as infertility, allergies, metabolic disorders, which cannot be only explained by the genetic changes. To better understand how drug exposure can impact human health, the concepts of adverse outcome pathways (AOPs) and AOP networks (AONs), which are representations of causally linked events at different biological levels leading to adverse health, could be used for drug safety assessment.

Methods: To explore the action of drugs across multiple scales of the biological organization, we investigated the use of a network-based approach in the known AOP space. Considering the drugs and their associations to biological events, such as molecular initiating event and key event, a bipartite network was developed. This bipartite network was projected into a monopartite network capturing the event–event linkages. Nevertheless, such transformation of a bipartite network to a monopartite network had a huge risk of information loss. A way to solve this problem is to quantify the network reduction. We calculated two scoring systems, one measuring the uncertainty and a second one describing the loss of coverage on the developed event–event network to better investigate events from AOPs linked to drugs.

Results: This AON analysis allowed us to identify biological events that are highly connected to drugs, such as events involving nuclear receptors (ER, AR, and PXR/SXR). Furthermore, we observed that the number of events involved in a linkage pattern with drugs is a key factor that influences information loss during monopartite network projection. Such scores have the potential to quantify the uncertainty of an event involved in an AON, and could be valuable for the weight of evidence assessment of AOPs. A case study related to infertility, more specifically to "decrease, male agenital distance" is presented.

Conclusion: This study highlights that computational approaches based on network science may help to understand the complexity of drug health effects, with the aim to support drug safety assessment.

Keywords: network science, AOP, bipartite network, drug-AOP, AON, NAM

INTRODUCTION

Nowadays, it is established that the increase of civilization diseases such as obesity or type 2 diabetes is not only explained by genetic changes in individuals, but may also be due to exposure to environmental factors. The concept of the exposome was proposed by Wild (1) and refined by Rappaport (2), which refers to the totality of the environmental exposures an individual is exposed to, during their entire lifetime from conception until death. The exposome aims at capturing all non-genetic factors such as physical stressors, biologicals, psychological, and social stresses, as well as chemical exposure. The exposome is therefore complementary to the genome, and the integration of both, representing the exposome-genome paradigm will be very useful to identify predictive markers for disease prevention at the population level (3). Humans are daily exposed to a high number of hazardous chemicals (drugs, toxicants, pollutants, and nutrients) from various sources such as cosmetics, diet, or medical treatments. The potential effect of the chemical exposome may contribute to disease risks and adverse outcomes, as it has been established in various biomonitoring and epidemiological studies from European projects (e.g., HELIX, HEALS, and HBM4EU) (4–6).

Even if drugs have been primarily designed to treat specific diseases, they may also influence health effects by displaying adverse effects. For example, aspirin and paracetamol are the common drugs to treat pain and fever. However, both drugs can lead to hepatoxicity with high doses (7). Analgesic use has also been associated with endocrine and reproductive effects (8). Many drugs interact with multiple biological targets, as referred to the polypharmacy action of drugs, and these unintended actions could cause adverse effects (9). Polypharmacy remains one of the major challenges in drug development as the modes of action of many drugs are not completely understood. To overcome this limitation, network pharmacology, which considers the drug action across multiple biological layers, is a promising approach (10). A recent review investigated the drug–exposome interaction, which shows how drugs can interact with other chemicals to which humans are exposed, and therefore how they may impact and influence human health (11). For example, chemotherapeutics can negatively be impacted by drug-exposome interactions, resulting in less drug efficacy, an increase of drug resistance, and adverse effects.

Emerging technological advances (e.g., high throughput screening (HTS) screening, omics) have supported the generation of the high number of data, allowing the development of computational studies, at cells and organs levels, to investigate biological mechanisms and to predict potential adverse effects of chemicals (including drugs). It is now possible to develop models based on omics data obtained for drugs or other toxicants (12). Therefore, exposomics can help in identifying novel chemical exposure-health associations, which create opportunities to prioritize relevant chemicals for risk assessment (13). Several systems toxicology models were developed based on network science to assess the chemical exposome effects on human health, some being applied to predict the effects of endocrine disruptors or persistent organic pollutants (14–18).

As phenotypes are the results of the complex interplay between chemicals and genetic factors, understanding the interactions between chemicals and diseases is essential. Recently, the tool "phexpo" was designed to explore bidirectional chemical and phenotype interactions (19). Also, the HExpoChem webserver allows exploring and predicting human health risks from various chemical exposure sources (20). All these integrative and computational approaches are complementary to existing approaches (pharmacogenomics), and combined together could be very useful for drug assessment. Yet, there is still a need to improve our understanding of the adverse effects of drugs in humans and the biological pathways they may perturbto fully assess their real impacts on humans.

Recently, a new concept has emerged in the field of toxicology, which is adverse outcome pathways (AOP) (21). AOPs are linear representations of biological perturbations [molecular initiating events (MIEs) and key events (KEs)] at different levels of the biological organization caused by a stressor (that could be a chemical), which lead to an adverse outcome (AO). The Organization for Economic Cooperation and Development (OECD) has classified MIEs and AOs as two specific KEs in AOPs. When evidences exist, KEs are linked together through Key Event Relationships (KERs), that are the causal linkages between two events. Even if AOPs are chemical agnostics, they are recognized as a tool for risk assessment of chemical effects to better understand their modes of action, and therefore could be also translated to the drug exposome. KEs are not uniquely associated with one AOP but can be shared between several AOPs, leading therefore to AOP networks (AONs) that reflect the real complexity of biological systems (22).

Using the concept of AOP, we proposed a systems toxicological approach to investigate, in a systematic way, linkages between drugs and biological events (MIE, KE, and AO) with the development of a network-based model. First, a bipartite network was generated by integrating known drug–event associations extracted from various data sources. This bipartite network was then transformed into an event–event monopartite network to explore the AOP space of drugs. A recently published method measuring the uncertainty and information loss of network was applied to assess the conservation of the information from the bipartite to the monopartite network (22), and the obtained results in our study are discussed below.

MATERIALS AND METHODS

The developed systems biology approach, to explore the drug effects in the biological space of AOPs, is a multistep procedure. A workflow of the proposed strategy is illustrated in **Figure 1**.

Data Set

Drug-event associations were extracted from two complementary datasets: the AOP-wiki database (https://aopwiki.org/, as of August 2020) (23) and the U.S. Environmental Protection Agency web-based CompTox Chemistry dashboard (https://

FIGURE 1 | Workflow of the procedure to assess AOPs in the drug exposome. A multistep procedure was developed: (1) Data compilation: drug–event associations were extracted from various data sources (solid lines). Several event types were considered, which are molecular initiating events (MIEs) (purple circles), key events (KEs) (light green circles), and adverse outcomes (AOs) (red circles); (2) Bipartite network development: a bipartite drug-event network was created based on the compiled data. To enrich the created network, KE relationships (KERs) (dash lines) extracted from the AOP-wiki database were identified, allowing to add other KEs (green circles); (3) Generation of the monopartite network: if two events shared at least one drug in the bipartite network, a link was created between these two events in the monopartite network. The transformation of the bipartite network to a monopartite network was done by using the "guilt by association" principle; (4) Network analysis: the loss of information from the bipartite network to the monopartite network was quantified by calculating the "increase of uncertainty" and the "loss of coverage."

comptox.epa.gov/dashboard/, as of March 2021) (24). The AOP-wiki database is a data source developed by the OECD. From this data source, 316 AOPs and 1,393 relevant events (i.e., MIE, KE, and AO) that are known to be linked to 615 stressors were extracted. All defined AOPs were taken into consideration, even if they are not yet validated by the OECD. The CompTox database contains a wide range of data related to chemical toxicity including, AOP information, human exposure, and other kwnoledge for over 88,300 chemicals. Direct and putative associations between 22,038 chemicals and 103 AOPs are stored in this database (25). As both databases provided information related to various stressor types (drugs, pollutants, etc) we matched the drug names and CAS numbers obtained from the DrugBank (26) and the DrugCentral databases (27) to be able to keep only information related to drugs, without loosing or duplicating knowledge. Consequently, we were able to extract unique drug-event and drug-AOP associations from these two data sources (Comptox and AOP-wiki).

To enrich our list of drug-AOP associations, an additional manual curation of the literature using the PubMed database (https://pubmed.ncbi.nlm.nih.gov/) (as of September 2020) was performed. The searches were done using the exact terms of interest for which we were missing links from the previous data integration step. As an example, cooccurence between ciprofibrate and activation, PPAR alpha was investigated, and identified publications were fully read for validation.

Development of the Drug-Event Bipartite Network

A drug-event bipartite network was created using the R software (version 3.5.23) based on the previously compiled data from the different data sources. Each node in the network represents either a drug or an event (MIE, KE, or AO). To enrich the connections between drugs and events, and to have a more comprehensive mapping of the drugs in the AOP space, KERs were investigated using the AOP-wiki database (as of September 2020). Globally, if a KER exists between two events (MIE, KE, AO) that are present in the bipartite network, then the biological events linking this KER was selected and included to the network.

Creation of the Monopartite Network of Events in the Drug Space

To analyze the relationships of events targeted by drugs, the bipartite network was projected into a monopartite network using the guilt by association principle (28), i.e., linking events through the shared connections between drugs, MIE, KE, and AO. With this projection, some information might be lost. Therefore, some scores measuring the uncertainty and the loss of coverage of the monopartite network were developed. For better visualization of the obtained complex networks, we exported both networks (bipartite and monopartite) to Cytoscape V3.7.2 which is an open-source library for network analysis and

visualization (29). This tool allows displaying of an interactive representation of the drug–event and event–event networks.

Analysis of the AOP Monopartite Network

The relevance of the relations between nodes (events) might be underestimated when generating a monopartite network, and quantifying the importance of associations between pairs of events could be of interest for AOPs construction (**Figure 2**). To evaluate this issue in the developed AOP network, we applied the approach of loss of information as defined previously by Vogt and Mestres (30) and implemented two measures that are the "increase of uncertainty" and the "loss of coverage".

Increase in Uncertainty

The concept of "uncertainty" came from the Shannon entropy (31). It measures how many associations between drugs and events are spread over all possible connections in a bipartite network. In the bipartite network of the present study, the associations between two types of nodes (i.e., drugs and events) define a "linkage pattern." After a projection to a monopartite network, only pairs of events having at least one common drug will be kept. Using the Shannon entropy theory, the distribution and the occurrence of each linkage pattern in the bipartite network can then be estimated in the projected monopartite network providing a measure of the increase in uncertainty (or loss of information) as presented in the equation below:

$$H^{before} = - \sum_{i=1}^{|L^{G^b}|} p(I_i) \cdot [\ln p(I_i)], \ I_i \in L^{G^b}$$

$$H^{after} = - \sum_{i=1}^{|cliques(G^m)|} p(C_i) \cdot [\ln p(C_i)],$$

$$\Delta H = H^{after} - H^{before}$$

where H^{before} indicates the uncertainty of the linkage patterns of events in bipartite network (G^b); L^{G^b} represents the linkage patterns in the bipartite network including events; and p (I_i) means the probability of each linkage pattern among all the linkage patterns in the bipartite network. After projection into the monopartite network G^m, H^{after} is used to estimate the uncertainty of G^m. Since G^m is the undirected uncertain network, if a set of nodes are fully interconnected in a subnetwork of G^m, this subnetwork is defined as "clique". A clique in the uncertain network tends to be a deterministic subnetwork (32). Here, $|cliques(G^m)|$ indicates all the cliques in the monopartite network. $p(C_i)$ represents the probability of each clique in G^m. ΔH describes quantitatively information in G^m compared with G^b. The unit of increase in uncertainty ΔH (nats) is based on the natural logarithm. The lower the nats value, the lower amount of information is lost in the monopartite network.

Compared with the calculation of H^{before} for which all possible linkage patterns (L^{G^b}) in the bipartite network are considered, H^{after} focus on all possible cliques ($|cliques(G^m)|$) in monopartite network. For a monopartite network with a huge number of events, it is difficult to count manually the number of cliques. Thus $|cliques(G^m)|$ can be represented as the number of subcliques for a maximal clique in G^m. A "maximal clique" is a clique that cannot be extended by adding another node of the network. For example, in **Figure 2**, the cliques in G^{m1} are {A},{A,B},{A,C}, {A,B,C}, while the maximal clique is only {A,B,C}. The number of events included in a maximal clique is defined as the "size of a maximal clique." Thus the size of maximal clique in G^{m1} is 3 (i.e., {A,B,C}). Then, for the calculation of subcliques, we applied the formulation: Number of subcliques for a maximal clique $= 2^{|c|}$, where $|c|$ is denoted as the size of a maximal clique (30). For example, with G^{m1}, it suggests that there exists $2^{|3|}$ possible subcliques in the monopartite network. Although the calculation of the number of subcliques by these formulas can lead to bias compared with the reality (8 subcliques vs. 3 subcliques), it can still provide approximate values by calculating the maximal clique for the monopartite network of large size.

In our work, we calculated two kinds of uncertainty, H^{before}, both in terms of global network uncertainty and nodes (events) uncertainty. $H^{before}_{network}$ is defined as the initial uncertainty of the global network. The calculation of $H^{before}_{network}$ takes into account all the linkage patterns in the bipartite network. Thus $H^{before}_{Gb1} = - \left[\ln\left(\frac{1}{2}\right) \times \left(\frac{1}{2}\right) + \ln\left(\frac{1}{2}\right) \times \left(\frac{1}{2}\right)\right] = 0.69$ nats. At the difference, H^{before}_{nodes} (the initial uncertainty in terms of nodes) only considers the linkage patterns related to a specific node (event) for the calculation. For example, to calculate H^{before}_C for C in G^{b1}, only linkage pattern related to the event "A,B,C" was considered in the calculation. Thus, $H^{before}_C = - \left[\ln\left(\frac{1}{2}\right) \times \left(\frac{1}{2}\right)\right] = 0.35$ nats. Similarly, after the projection to the monopartite network, $H^{after}_{network}$ will include all the cliques for the calculation i.e., $H^{after}_{Gm1} = - \left[\ln\left(\frac{1}{8}\right) \times \left(\frac{1}{8}\right) \times 8\right] = 2.08$ nats. For the nodes, H^{after}_{nodes} considers the maximal clique contained in a specific node. So in theory, $H^{after}_C = - \left[\ln\left(\frac{1}{8}\right) \times \left(\frac{1}{8}\right) \times 8\right] = 2.08$ nats for the event C in G^{m1}. Overall, the resulting $\Delta H_{network} = 1.39$ nats represents an increase in uncertainty for the global network during the transformation from G^{b1} to G^{m1}. The H_{nodes} for C ($\Delta H_C = 1.73$ nats) indicates that linkage patterns related to the C event has an increase in uncertainty of 1.73 nats after translating into cliques.

A high ΔH means that the monopartite network projection of the bipartite network has lost information during the transformation, suggesting an increase of uncertainty on the linkage pattern between events. Also, for the nodes, a high ΔH (like for C) depicts a low linkage of this event compared with the other events inside the clique. Therefore, these ΔH values can quantify the reliability of a connection between an event and other events depending on the drug–event information integrated into the original network.

Loss of Coverage

In addition to the loss of information estimation, it is also possible to determine the degree of involvement of each edge during the projection, which indicates the original properties of the linkage pattern lost during monopartite network projection. This is the "loss of coverage," and it is calculated using the
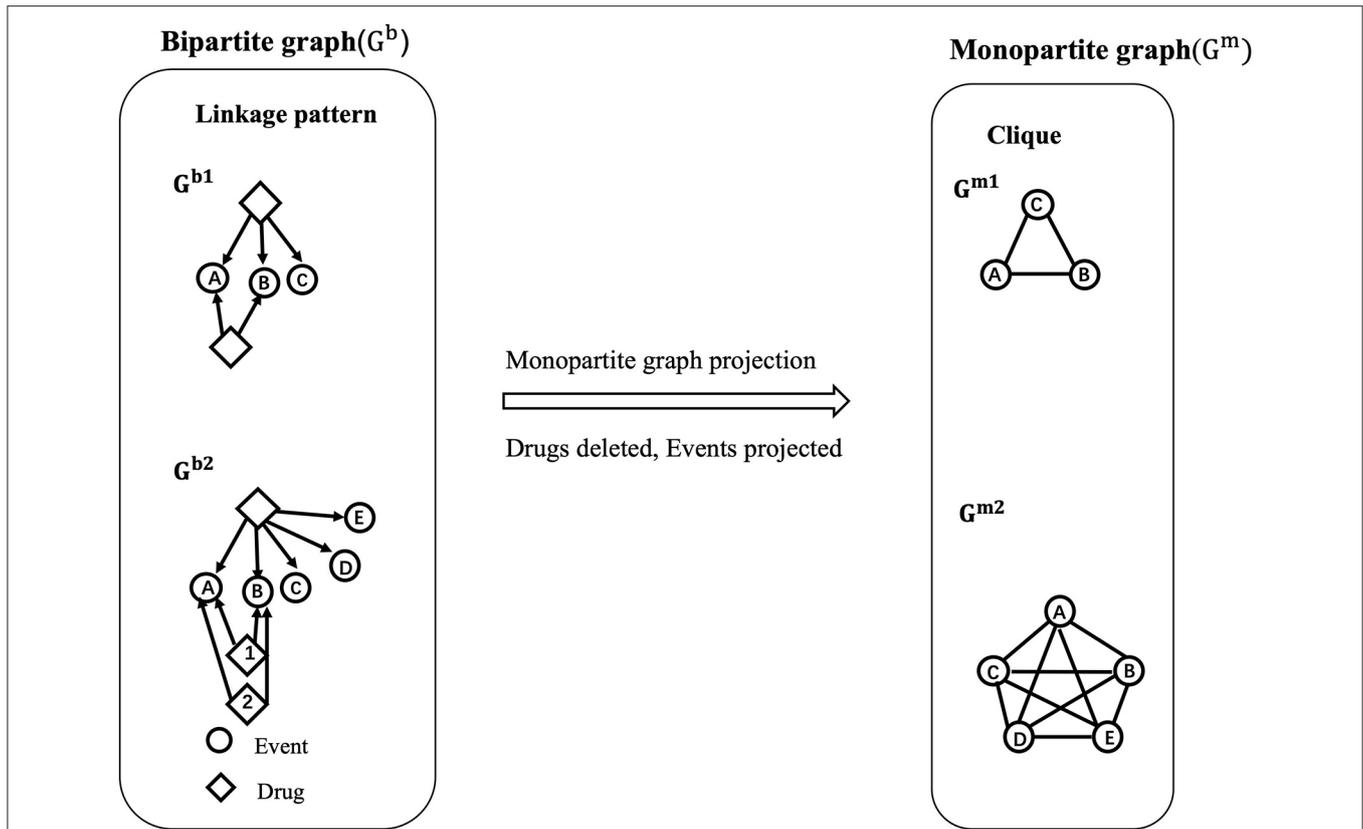
FIGURE 2 | Illustration of bipartite network projection to a monopartite network. In a bipartite network G^b, containing two types of nodes (events represented as circles and drugs as diamonds), the associations between drugs and events are defined as "linkage patterns". For example, in G^{b1}, the A, B, C events are connected to a drug 1 *via* 3 edges, which define a linkage pattern ABC with a linkage profile of 1 (occurrence of 1). In the meantime, the A and B events are both connected to drug 2 one time, therefore defining another linkage pattern AB with a linkage profile of 1. Similarity, in G^{b2}, the linkage pattern AB has a linkage profile of 2 as it appears two times (drugs 1 and 2). The linkage profile of linkage pattern ABCDE is 1 (drug 3). Then, from the bipartite network, a projection to a monopartite network G^m is performed for the events. If two events share at least one common drug in the bipartite network, an association between these two events is generated. Finally, G^{b1} and G^{b2} were transformed into G^{m1}, G^{m2}, respectively. As observed, G^{m1} and G^{m2} are present as "cliques" meaning that all the events in G^{m1} and G^{m2} are fully interconnected. It is showed that the linkage pattern AB appears twice in G^{b2} while oit is overlapping as an edge AB in G^{m2}, thus resulting in the information loss.

formulations below:

$$COV_{Edge} = \frac{\sum^{\left|L_{vw}^b\right|} n_i \cdot \frac{2}{I_iV}}{\sum^{\left|L_{vw}^b\right|} n_i}, \text{ with } COV_{Edge} \in (0, 1)$$

$$\neg COV_{Edge} = 1 - COV_{Edge}, \text{ with } \neg COV_{Edge} \in (0, 1)$$

The term COV_{Edge} is the coverage for each edge between events v and w in G^m. L_{vw}^b is the amount of linkage patterns, I_iV contains both events v,w in the bipartite network. Here, n_i is the occurrence of I_iV. $\neg COV_{Edge}$ indicate the loss of coverage for the edge between event v and w in G^m. A high $\neg COV_{Edge}$ (maximum is 1) indicates that the edge between the two events presented in the monopartite network cover few commun information (for example few number of drugs) and a $\neg COV_{Edge}$ close to 0, which means that many drugs are linked to two events in the bipartite network. For example, to calculate $COV_{Edge(A,B)}$ for G^{m1}, the linkage patterns containing A and B are "AB" and "ABC". Both linkage patterns appear once in the G^{b1}. So, for $COV_{Edge(A,B)} =$

$\frac{\left[1 \times \left(\frac{2}{2}\right) + 1 \times \left(\frac{2}{3}\right)\right]}{1 + 1} = 0.83$ $\neg COV_{Edge(AB)} = (0.17)$. Doing a similar analysis between B and C, we obtained a $\neg COV_{Edge(BC)} = (0.33)$ in G^{m1}. So, the coverage is more important between events A and B than between B and C.

Finally, we computed the loss of coverage for nodes $\neg COV_{Node}$ in the monopartite projection. It considers the average of all the edges in G^{m1} that are linked to a node. For example, the event C is linked to the event B and A. Therefore, the loss of coverage for the event C is $\neg COV_C = \frac{\neg COV_{Edge(B,C)} + \neg COV_{Edge(A,C)}}{2} = 0.33$. This value can then be compared with the loss of coverage for all other events within the aim to assess the coverage associated with each event in the monopartite network.

Visualization and Analysis of the Loss Information of the Developed Network

In both the methods, an increase in uncertainty and loss of coverage were implemented under python 3.7 and R, using the R package Reticule as an interface provider between R and Python

(https://cran.r-project.org/web/packages/reticulate/reticulate. pdf) (R version 3.5.3). All implemented methods and applications in this work are available on GitHub (https://github.com/ QierWU/drug-AOP). Results were visualized and analyzed in Cytoscape V3.7.2.

RESULTS

Drug–Event Data

Drug–event associations were compiled from two complementary data sources and merged to obtain unique relationships. From the CompTox database, we were able to collect 248 drugs associated with 23 unique MIEs (as of March 2021) (**Supplementary Table S1**). From the AOP-wiki database, we extracted 117 unique drugs having at least one corresponding event (as of August 2020) (**Supplementary Table S1**). However, 46 among these 117 drugs were only connected to an AOP, with no information concerning the event linkage (MIE, KE, AO). For example, meclofenamic acid was associated with the AOP 152 'Interference with thyroid serum binding protein transthyretin and subsequent adverse human neurodevelopmental toxicity," but no KE was specified. Therefore, to develop a more complete network, we manually explored the literature using the PubMed database to enrich the connection between the drugs and the 11 events known to be involved in the AOP 152 (**Supplementary Table S2**). For meclofenamic acid, 1,158 publications were identified. Then, we searched among these abstracts, those mentioning KE of interest. As an example, the MIE 957 "Binding, Transthyretin in serum," which is know to be linked to the AOP 152 (33) was retrieved in 5 out of the 1,158 abstracts. With this curation, we were able to decipher the linkage between 46 drugs and 18 events *via* 41 publications. As a result, drugs were connected to 89 unique events (52 MIEs, 31KEs, 6 AOs).

In addition, to enrich the drug–event associations network, we used KERs information that link event pairs from the AOP–wiki database (as of September 2020) (**Supplementary Table S3**). For example, the MIE 718 "Binding, Tubulin" and the AO 728 "Increase, Aneuploid offspring" were present in the developed bipartite network (**Supplementary Table S1**). From the AOP–wiki, both are also part of the AOP 106 "Chemical binding to tubulin in oocytes leading to aneuploid offspring" (https:// aopwiki.org/aops/106, as of September 2021). Therefore, events involved in AOP106 were added into our bipartite network. This step allowed to integrate information related to 13 KEs. Overall, all obtained information were merged to keep only unique information to build the bipartite network, which consists of 321 drugs, 66 MIE, 44 KE, and 6 AO (**Supplementary Table S4**).

Development of the Drug–Event Bipartite Network

All the 321 drugs and their direct associations to events were connected through 981 edges in the bipartite network (**Figure 3**). All the events were classified into three groups based on their different event types (MIE, KE, and AO). Interestingly, many drugs were linked to at least one nuclear receptor: 134 drugs were connected to the MIE 1181 "Activation, Estrogen

receptor," 131 drugs were connected to the MIE 1134 "Irreversible inhibition of hepatic VKOR by binding of AR at tyrosine 139, Failure to cycle vitamin K epoxide to vitamin K to form vitamin K hydroquinone," and 110 drugs were connected to the MIE 245 "Activation, PXR/SXR." Among the most linked drugs, propylthiouracil, used to treat hyperthyroidism was connected to the maximum number of events (12 events). The antifungal agent, clotrimazole, was connected to 11 events, and the antiinflammatory drug, betamethasone, was connected to 9 events. It was also observed that 134 drugs were only associated with a unique event (e.g., Sertraline, Topiramate). Globally, the drug–event network emphasized that most of the studies focused on interactions between drugs and adverse protein targets.

Monopartite Network of Events in the Drug Space

The developed monopartite network projection from the drug-AOP bipartite network is represented in **Figure 4**. It was built based on the assumption that if two events in the drug-event network shared at least one common drug or event, a link between these two events can be depicted in the monopartite network. Fourteen events (1AO, 2KEs, 11MIEs) were linked to a unique drug and were not included in the monopartite projection (as these KEs were not interacting with other KEs). At the end, 102 events were connected through 515 edges.

Considering the increase in uncertainty (**Figure 4A**), nodes with thicker widths indicated more information loss. The MIE 9 "Activation, 5HT2c" and MIE 1009 "Inhibition, Deiodinase 1" had the highest increase in uncertainty (7.60 nats), reflecting that these two events have a linkage to a unique (or few) drug(s), but this (or these) drug(s) are connected to many events. The high value of an increase in uncertainty is related to the maximum size of linkage pattern associated with the MIE 9 or MIE 1009 in the bipartite network. In fact, the MIE 9 was only connected to the drug "clove leaf oil" in the bipartite network. Moreover, this drug was also linked to 10 other MIEs. It suggested that 2^{10} possible linkage patterns relevant to the MIE 9 could exist in the bipartite network. Obviously, the number of linkage patterns related to the MIE 9 projected in the monopartite network is less. Thus, the uncertainty for the linkage patterns relevant to the MIE 9 increased largely after monopartite network projection. On the contrary, it is observed that the KE 1614 "Decrease, AR activation" had the lowest increase in uncertainty (1.34 nats), meaning that this KE had a small size of linkage patterns in the bipartite network, thus resulting in low uncertainty in the monopartite network. Regarding loss of coverage (**Figure 4B**), the nodes with thicker width and purple edges had high information loss. The value of $\neg COV_{Edge}$ gives the coverage loss between two nodes in a monopartite network, whereas the value of $\neg COV_{Node}$ defines the mean of $\neg COV_{Edge}$ associated with a node. For example, the edge between the MIE 9 and the MIE 18 "Activation, AhR" has a value of $\neg COV_{Edge(MIE\ 9,\ MIE\ 18)} = 0.82$. Such $\neg COV_{Edge}$ can be considered as high because it is close to 1. On the opposite, the MIE 167 "Activation, LXR" and the MIE 468 "Inhibition, PPAR alpha" shared 10 drugs. The loss of coverage will be lower with
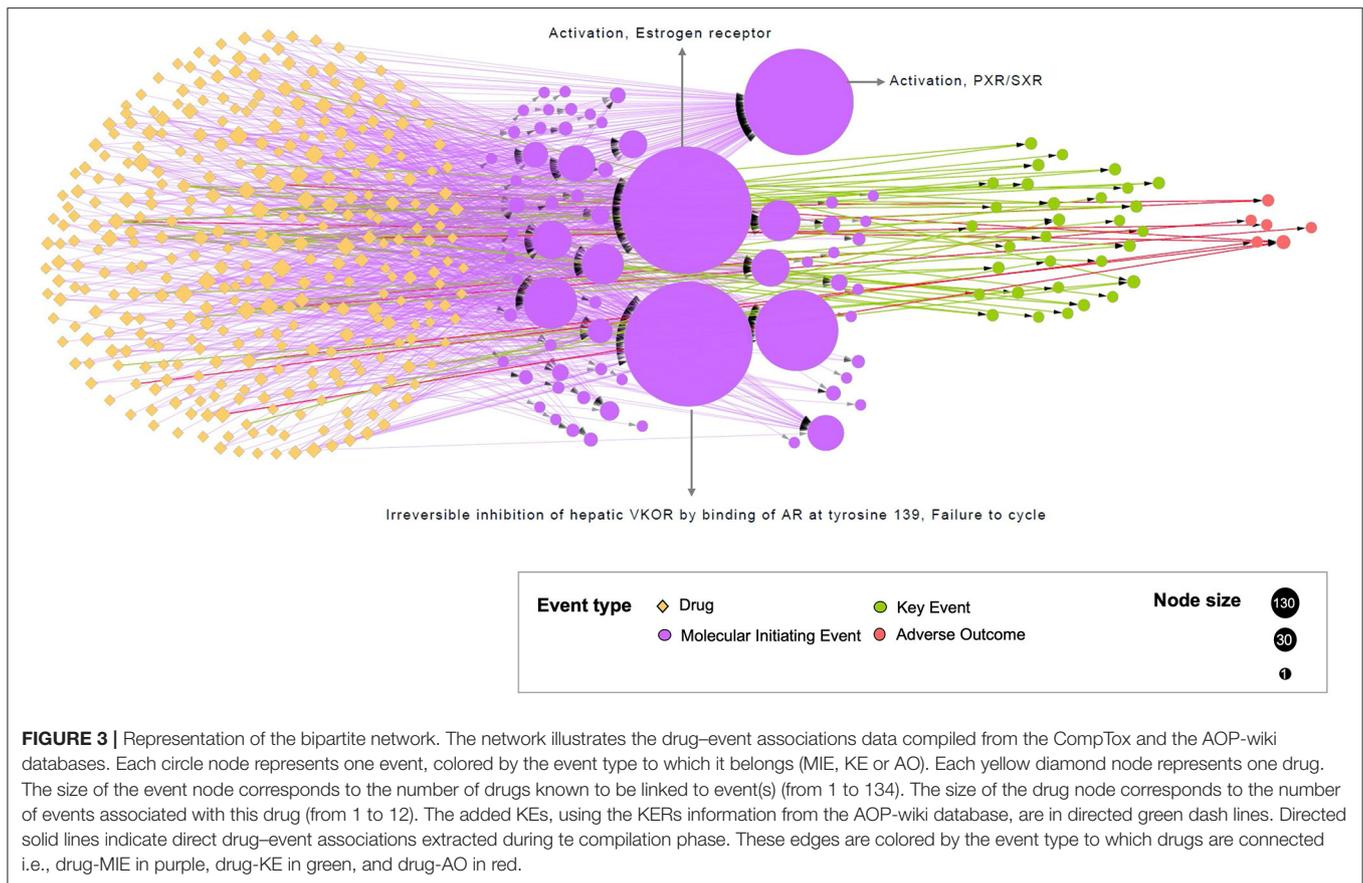
FIGURE 3 | Representation of the bipartite network. The network illustrates the drug–event associations data compiled from the CompTox and the AOP-wiki databases. Each circle node represents one event, colored by the event type to which it belongs (MIE, KE or AO). Each yellow diamond node represents one drug. The size of the event node corresponds to the number of drugs known to be linked to event(s) (from 1 to 134). The size of the drug node corresponds to the number of events associated with this drug (from 1 to 12). The added KEs, using the KERs information from the AOP-wiki database, are in directed green dash lines. Directed solid lines indicate direct drug–event associations extracted during te compilation phase. These edges are colored by the event type to which drugs are connected i.e., drug-MIE in purple, drug-KE in green, and drug-AO in red.

$\neg COV_{Edge\ (MIE\ 167,\ KE\ 468)} = 0.75$, leading to a higher confidence about the relationship between these two events. Similarly, the drug propylthiouracil, which targets the MIE 1009, is also linked to 10 other events, among them the MIE 18 "Activation, AhR" and the MIE 245 "Activation, PXR/SXR." These two MIEs have 19 drugs in common, resulting in a weak loss of coverage ($\neg COV_{Edge(MIE\ 18,\ MIE\ 245)} = 0.61$) between these two events, and a higher chance that a drug impacting one of this event might has also an effect on the second event (**Supplementary Figure S1**). Similarly, for the edge between MIE 1181 "Activation, Estrogen receptor" and MIE 1543 "Mitochondrial Complex IV inhibition," the $\neg COV_{Edge(MIE\ 1181,\ MIE\ 1543)} = 0.33$, meaning a limited loss of coverage.

Additionally, a loss of information can be computed on each node ($\neg COV_{Node}$). $\neg COV_{Node}$ is the mean of $\neg COV_{Edge}$ associated with a node. For example, the MIE 9 and the MIE 18 are connected with a $\neg COV_{Edge(MIE\ 9,\ MIE\ 18)} = 0.82$, resulting at the node level to $\neg COV_{MIE\ 9} = 0.82$ and also $\neg COV_{MIE\ 18} = 0.72$. This suggests that the putative associations related to the MIE 9 and the MIE 18 in the monopartite network simplified the real situation of connections in the bipartite network.

Finally, we noticed a loss of coverage equal to zero between the two events, the MIE 1529 "Blockade, L-Type Calcium Channels" and the MIE 593 "Inhibition, Ether-a-go-go (ERG) voltage-gated potassium channel." It means there is no loss of information

during the projection. This suggests a very strong relationship between these two events. However, in very specific cases, the loss of coverage might be equal to zero when a linkage pattern has a size of two, and the linkage profile is more than once. In this case, loss of coverage does not reflect the reality. For example, the KE 1614 "Decrease, AR activation" and AO 1688 "decrease, male anogenital distance" shared with two common KEs that are very similar (KE 1687 "decrease, transcription of genes by AR" and KE 286 "Altered, Transcription of genes by AR"). The size of the linkage pattern as well as the linkage profile is two. As these associations are merged as one unique edge in the monopartite network, the value of loss of coverage at the edge is 0. To better illustrate the use of this events network, a case study related to infertility is presented in **Supplementary Material**.

In a global overview, the information loss for the full monopartite network and each type of event (MIE, KE, and AO) are summarized in **Table 1**. The initial uncertainty ($H_{network}^{before}$) for the network is 5.00 nats, and for the nodes (H_{nodes}^{before}) is 0.05 nats. Among the subnetworks of H_{nodes}^{before}, the MIE nodes showed the largest initial uncertainty value (0.08 nats), whereas the KE and AO showed the lowest value (0.03 nats). This indicated that the linkage patterns related to nodes in the MIE subgroup had a higher initial uncertainty compared with the AO and KE subgroups. All the nats values for the 3 subgroups are close to 0 and so the initial uncertainty at node level is

FIGURE 4 | Uncertainty increase **(A)** and loss of coverage **(B)** of the monopartite network. (A) Illustrates the increase in uncertainty, and **(B)** represents the loss of coverage, both for the monopartite network of events. Each circle node represents one event, colored by the event type to which it belongs (MIE, KE, AO). Solid edges indicate event–event associations. In **(A)**, the width of nodes increase according to the value of the increase in uncertainty. In **(B)**, the width of nodes is raised with the value of loss of coverage. The dark color of the edges shows the highest value of loss of coverage.

	$H1^{before}_{network}$	$H2^{before}_{nodes}$	$H3^{after}_{network}$	$H4^{after}_{nodes}$	$\Delta H5_{network}$	$\Delta H6_{nodes}$	$\neg COV7_{Edge}$	$\neg COV8_{Node}$
Network	5.00	0.05	11.28	5.06	6.28	5.01	0.67	0.58
AO	–	0.03	–	5.82	–	5.79	0.68	0.64
KE	–	0.03	–	4.18	–	4.15	0.62	0.50
MIE	–	0.08	–	5.19	–	5.11	0.68	0.59

Adverse outcome (AO), key event (KE), and molecular initiating event (MIE).

low. After the monopartite projection, the uncertainty at the global network ($H^{after}_{network}$) and nodes (H^{after}_{nodes}) is 11.28 and 5.06 nats, respectively. This means that the uncertainty increased by 6.28 nats ($\Delta H_{network}$) for the monopartite network. A 5.01 nats increase is observed in terms of nodes (ΔH_{nodes}). For subnetworks, the increase in uncertainty (ΔH_{nodes}) is of 5.79 nats for the AO and 4.15 nats for the KE. This reflects that the projection procedure has more impact on the AO than the KE.

The average loss of coverage for edges ($\neg COV_{edge}$) and nodes ($\neg COV_{node}$) after projection in the monopartite network are 0.67 and 0.58 respectively. Higher values of $\neg COV_{edge}$ and $\neg COV_{node}$ are observed for AO compared with KE and MIE, revealing a higher loss of information for AO.

$H1^{before}_{network}$: initial uncertainty in terms of the network, $H2^{before}_{nodes}$: initial uncertainty in terms of nodes. The value of each subgroup was an average of H^{before} for each node in the given subgroup; $H3^{after}_{network}$: uncertainty after projection for the network; $H4^{after}_{nodes}$: uncertainty after projection for nodes. Value of each subgroup was average of H^{after} for each node in the given subgroup; $\Delta H5_{network}$: increase in uncertainty for the network; $\Delta H6_{nodes}$: increase in uncertainty for nodes; $\neg COV7_{Edge}$: loss of coverage at the level of edge. For the value of each subgroup, considering all the edges that link nodes belonged to the target subgroup, the average of loss of coverage for each edge was calculated; $\neg COV8_{Node}$: loss of coverage for nodes. The value of each subgroup was the average of loss of coverage for each node in the given subgroup.

DISCUSSION

Assessment of the chemical toxicity at an early stage remains a tremendous challenge for the pharmaceutical industries and the regulatory agencies. To have a better picture of the effects of the chemical exposome on human health, new and innovative concepts and methods are needed. Even if various systems toxicological models have been developed to decipher chemical toxicity in humans, there is still a need to increase our understanding (34, 35).

In this study, we proposed a network-based approach that supports the concept of integrated approaches to testing and assessment (IATA) proposed by the OECD (36). Based on our method, we explored the potential biological mechanisms of drugs that lead to adverse effects by analyzing AON that result from AOPs. Network-based modeling has the advantage of being a chemical structural information free and can be used

to establish toxicological profiling for substances. The developed process is part of a new approach methodology (NAM), which supports the use of alternative methods to animal testing for the identification of biological perturbation from chemical exposure (37).

Here, we first developed a bipartite network model to capture the behaviors of drug–event associations described by the AOP concept based on the assumption that the real-world complex networks have bipartite structures (38, 39). The bipartite network was then projected into a monopartite network to analyze the relationship between the two events. However, this translation is often accompanied by a loss of information (40), and we decided to quantify the loss of information and the coverage obtained within this events network.

The generated events network was evaluated using a method that quantifies the reduction with probability in two ways: "increase of uncertainty" and "loss of coverage." Results indicated that the level of information loss during monopartite projection mainly depends on the number of events involved in a linkage pattern (also called the size of linkage pattern), which is illustrated by a case study with the health effect "infertility" (**Supplementary Material**). Therefore, such methods allow to identify uncertainties and data gaps between the different events involved in an AON, and also to quantify the confidence of a relation between the two events.

Although this method is able to illustrate links between drugs and AOPs, one limitation of the developed network is that it relies on existing information. Some causal linkages may have been overlooked or disregarded because of missing or incomplete information. Morever such modeling study should take into consideration the so-called "Matthew effect," that reflects the difference, in terms of available knowledge (database, literature), between the very well investigated drugs and the ones less studied (41). Therefore, an extension of such model could be done by integrating more data, including quantitative information, and to fully evaluate the toxicological effects of drugs. One way could be to adapt existing tools such as AOP-helpFinder, which uses artificial intelligence and text mining to automatically screen the literature, to compile more data (42, 43). For example, this tool was successfully applied to a set of pesticides by automatically identifying among the 32 million abstracts present in the PubMed database, links between pesticides and events (44). Another way would be to gather information from publicly available databases, which should become more accessible within the next few years following the interoperability of the data based on the goals established by Findable, Accessible, Interoperable, and Reusable

(FAIR) data principles (45). There is a need to implement standardization of data to facilitate the use of multiple data sources, as a lack of interporability between all existing databases increases the difficulties for developping integrative models.

Other data types could also be considered to have a more complete model as possible, reflecting therefore the biological "true." As an example, environmental chemicals, pharmacogenomics, or drug–drug interactions could also be taken into consideration.

All these data should allow the development of predictive models and emerging tools. Nevertheless, the challenge will be on how to integrate all these heterogeneous data (from various species, different technologies...) in a common platform (46). Finally, our study relies on the available names of the KEs extracted from existing databases, and some KEs may have a similar biological meaning (e.g., in AOP-wiki, KE1608 "Increase, Oxidative DNA damage" and KE1634 "Increase, Oxidative damage to DNA"). Therefore, there is a need to harmonize the KE nomenclature to develop the best predictive models possible.

CONCLUSIONS

A network-based model was developed to explore in a systematic manner how drugs may be linked to biologic events (MIE, KE, and AO), and to quantify the uncertainty of a relation between events within an AOP. This method could be of interest as a weight of evidence (WoE) evaluation and could be complementary to the systematic WoE analysis provided so far (47). Overall our analysis highlights that such systems' toxicology science can help to improve the knowledge of interactions between drugs and biological systems, which can be of great interest for animal-free next generation risk assessment (NGRA).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

QW: data curation, formal analysis, methodology, result analysis, and drafted the manuscript. YB: software development. OT: design the study, result analysis, and revised the paper. KA: design the study, result analysis, revised the paper, and supervision. All authors contributed to the article and approved the submitted version.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2021.763962/full#supplementary-material

REFERENCES

1. Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev.* (2005) 14:1847–50. doi: 10.1158/1055-9965.EPI-05-0456

2. Rappaport SM, Smith MT. Environment and disease risks. *Science.* (2010) 330:460–1. doi: 10.1126/science.1192603

3. Barouki R, Audouze K, Coumoul X, Demenais F, Gauguier D. Integration of the human exposome with the human genome to advance medicine. *Biochimie.* (2018) 152:155–8. doi: 10.1016/j.biochi.2018.06.023

4. Vineis P, Chadeau-Hyam M, Gmuender H, Gulliver J, Herceg Z, Kleinjans J, et al. The exposome in practice: design of the EXPOsOMICS project. *Int J Hyg Environ Health.* (2017) 220:142–51. doi: 10.1016/j.ijheh.2016.08.001

5. Vrijheid M, Slama R, Robinson O, Chatzi L, Coen M, van den Hazel P, et al. The human early-life exposome (HELIX): project rationale and design. *Environ Health Perspect.* (2014) 122:535–44. doi: 10.1289/ehp.1307204

6. Bopp SK, Barouki R, Brack W, Dalla Costa S, Dorne J-LCM, Drakvik PE, et al. Current EU research activities on combined exposure to multiple chemicals. *Environ Int.* (2018) 120:544–62. doi: 10.1016/j.envint.2018.07.037

7. Altabakhi IW, Anderson J, Zito PM. Acetaminophen/Aspirin/Caffeine. In : *StatPearls.* Treasure Island (FL): StatPearls Publishing. Available at: http://www.ncbi.nlm.nih.gov/books/NBK513274/ (accessed July 9, 2021).

8. Kristensen DM, Mazaud-Guittot S, Gaudriault P, Lesné L, Serrano T, Main KM, et al. Analgesic use— prevalence, biomonitoring and endocrine and reproductive effects. *Nat Rev Endocrinol.* (2016) 12:381–93. doi: 10.1038/nrendo.2016.55

9. Cheng F, Kovács IA, Barabási A-L. Network-based prediction of drug combinations. *Nat Commun.* (2019) 10:1197. doi: 10.1038/s41467-019-09186-x

10. Boezio B, Audouze K, Ducrot P, Taboureau O. Network-based approaches in pharmacology. *Mol Inform.* (2017) 36:1700048. doi: 10.1002/minf.201700048

11. Pristner M, Warth B. Drug–exposome interactions: the next frontier in precision medicine. *Trends Pharmacol Sci.* (2020) 41:994–1005. doi: 10.1016/j.tips.2020.09.012

12. Niedzwiecki MM, Walker DI, Vermeulen R, Chadeau-Hyam M, Jones DP, Miller GW. The exposome: molecules to populations. *Annu Rev Pharmacol Toxicol.* (2019) 59:107–27. doi: 10.1146/annurev-pharmtox-010818-021315

13. Sillé FCM, Karakitsios S, Kleensang A, Koehler K, Maertens A, Miller GW, et al. The exposome—a new approach for risk assessment. *ALTEX.* (2020) 37:3–23. doi: 10.14573/altex.2001051

14. Wu Q, Achebouche R, Audouze K. Computational systems biology as an animal-free approach to characterize toxicological effects of persistent organic pollutants. *ALTEX.* (2020) 37:287–99. doi: 10.14573/altex.1910161

15. Taboureau O. El M'Selmi W, Audouze K. Integrative systems toxicology to predict human biological systems affected by exposure to environmental chemicals. *Toxicol Appl Pharmacol.* (2020) 405:115210. doi: 10.1016/j.taap.2020.115210

16. Taboureau O, Audouze K. Human environmental disease network: a computational model to assess toxicology of contaminants. *ALTEX.* (2017) 34:289–300. doi: 10.14573/altex.1607201

17. Karthikeyan BS, Ravichandran J, Mohanraj K, Vivek-Ananth RP, Samal A, A. curated knowledgebase on endocrine disrupting chemicals and their biological systems-level perturbations. *Sci Total Environ.* (2019) 692:281–96. doi: 10.1016/j.scitotenv.2019.07.225

18. Ruiz P, Perlina A, Mumtaz M, Fowler BA. A systems biology approach reveals converging molecular mechanisms that link different POPs to common metabolic diseases. *Environ Health Perspect.* (2016) 124:1034–41. doi: 10.1289/ehp.1510308

19. Hawthorne C, Simpson DA, Devereux B, López-Campos G. Phexpo: a package for bidirectional enrichment analysis of phenotypes and chemicals. *JAMIA Open.* (2020) 3:173–7. doi: 10.1093/jamiaopen/ooaa023

20. Taboureau O, Jacobsen UP, Kalhauge C, Edsgärd D, Rigina O, Gupta R, et al. HExpoChem: a systems biology resource to explore human exposure to chemicals. *Bioinformatics.* (2013) 29:1231–2. doi: 10.1093/bioinformatics/btt112

21. Vinken M. The adverse outcome pathway concept: a pragmatic tool in toxicology. *Toxicology.* (2013) 312:158–65. doi: 10.1016/j.tox.2013.08.011

22. Villeneuve DL, Angrish MM, Fortin MC, Katsiadaki I, Leonard M, Margiotta-Casaluci L, et al. Adverse outcome pathway networks II: network analytics. *Environ Toxicol Chem.* (2018) 37:1734–48. doi: 10.1002/etc.4124

23. OCDE. Users' handbook supplement to the guidance document for developing and assessing adverse outcome pathways. *OECD Series on Adverse Outcome Pathways No:1.* Paris: OECD Publishing (2018). doi: 10.1787/5jlv1m9d1g32-en

24. Williams AJ, Grulke CM, Edwards J, McEachran AD, Mansouri K, Baker NC, et al. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. *J Cheminform.* (2017) 9:61. doi: 10.1186/s13321-017-0247-6

25. Pittman ME, Edwards SW, Ives C, Mortensen HM, AOP-DB. A database resource for the exploration of adverse outcome pathways through integrated association networks. *Toxicol Appl Pharmacol.* (2018) 343:71–83. doi: 10.1016/j.taap.2018.02.006

26. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* (2018) 46:D1074–82. doi: 10.1093/nar/gkx1037

27. Ursu O, Holmes J, Bologa CG, Yang JJ, Mathias SL, Stathias V, et al. DrugCentral 2018: an update. *Nucleic Acids Res.* (2019) 47:D963–70. doi: 10.1093/nar/gky963

28. Hou L, Chen M, Zhang CK, Cho J, Zhao H. Guilt by rewiring: gene prioritization through network rewiring in Genome Wide Association Studies. *Hum Mol Genet.* (2014) 23:2780–90. doi: 10.1093/hmg/ddt668

29. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* (2003) 13:2498–504. doi: 10.1101/gr.1239303

30. Vogt I, Mestres J. Information loss in network pharmacology. *Mol Inform.* (2019) 38:e1900032. doi: 10.1002/minf.201900032

31. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J.* (1948) 27:379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x

32. Mukherjee AP, Xu P, Tirthapura S. Enumeration of maximal cliques from an uncertain graph. *IEEE Trans Knowl Data Eng.* (2017) 29:543–55. doi: 10.1109/TKDE.2016.2527643

33. Tenenbaum A, Motro M, Fisman EZ. Dual and pan-peroxisome proliferator-activated receptors (PPAR) co-agonism: the bezafibrate lessons. *Cardiovasc Diabetol.* (2005) 4:14. doi: 10.1186/1475-2840-4-14

34. Hartung T, FitzGerald RE, Jennings P, Mirams GR, Peitsch MC, Rostami-Hodjegan A, et al. Systems toxicology: real world applications and opportunities. *Chem Res Toxicol.* (2017) 30:870–82. doi: 10.1021/acs.chemrestox.7b00003

35. Nie W, Lv Y, Yan L, Chen X, Lv H. Prediction and characterisation of the system effects of aristolochic acid: a novel joint network analysis towards therapeutic and toxicological mechanisms. *Sci Rep.* (2015) 5:17646. doi: 10.1038/srep17646

36. Sakuratani Y, Horie M, Leinala E. Integrated approaches to testing and assessment: OECD activities on the development and use of adverse outcome pathways and case studies. *Basic Clin Pharmacol Toxicol.* (2018) 123:20–8. doi: 10.1111/bcpt.12955

37. Bopp SK, Kienzler A, Richarz A-N, van der Linden SC, Paini A, Parissis N, et al. Regulatory assessment and risk management of chemical mixtures: challenges and ways forward. *Crit Rev Toxicol.* (2019) 49:174–89. doi: 10.1080/10408444.2019.1579169

38. Guillaume J-L, Latapy M. Bipartite structure of all complex networks. *Inf Process Lett.* (2004) 90:215–21. doi: 10.1016/j.ipl.2004.03.007

39. Guillaume J-L, Latapy M. Bipartite graphs as models of complex networks. *Physica A: Stat Mech Appl.* (2006) 371:795–813. doi: 10.1016/j.physa.2006.04.047

40. Pavlopoulos GA, Kontou PI, Pavlopoulou A, Bouyioukos C, Markou E, Bagos PG. Bipartite graphs in systems biology and medicine: a survey of methods and applications. *GigaScience.* (2018) 7:1–31. doi: 10.1093/gigascience/giy014

41. Grandjean P, Eriksen ML, Ellegaard O, Wallin JA. The Matthew effect in environmental science publication: a bibliometric analysis of chemical substances in journal articles. *Environ Health Glob.* (2011) 10:96. doi: 10.1186/1476-069X-10-96

42. Carvaillo JC, Barouki R, Coumoul X, Audouze K. Linking bisphenol S to adverse outcome pathways using a combined text mining and systems biology approach. *Environ Health Perspect.* 127:047005. doi: 10.1289/EHP4200

43. Rugard M, Coumoul X, Carvaillo J-C, Barouki R, Audouze K. Deciphering adverse outcome pathway network linked to bisphenol F using text mining and systems toxicology approaches. *Toxicol Sci.* (2020) 173:32–40. doi: 10.1093/toxsci/kfz214

44. Jornod F, Rugard M, Tamisier L, Coumoul X, Andersen HR, Barouki R, et al. AOP4EUpest: mapping of pesticides in adverse outcome pathways using a text mining tool. *Bioinformatics.* (2020) 36:4379–81. doi: 10.1093/bioinformatics/btaa545

45. Watford S, Edwards S, Angrish M, Judson RS, Paul Friedman K. Progress in data interoperability to support computational toxicology and chemical safety evaluation. *Toxicol Appl Pharmacol.* (2019) 380:114707. doi: 10.1016/j.taap.2019.114707

46. Pawar G, Madden JC, Ebbrell D, Firman JW, Cronin MTD. In silico toxicology data resources to support read-across and (Q)SAR. *Front Pharmacol.* (2019) 10:561. doi: 10.3389/fphar.2019.00561

47. Becker RA, Ankley GT, Edwards SW, Kennedy SW, Linkov I, Meek B, et al. Increasing scientific confidence in adverse outcome pathways: application of tailored bradford-hill considerations for evaluating weight of evidence. *Regul Toxicol Pharmacol.* (2015) 72:514–37. doi: 10.1016/j.yrtph.2015.04.004