Research Article

Development of a Network of Carcinogenicity Adverse Outcome Pathways and its Employment as an Evidence Framework for Safety Assessment

Alex N. Cayley, Robert S. Foster, Emma Hill, Steven Kane, Grace Kocks, Alun Myden, Daniel Newman, Susanne A. Stalford, Jonathan D. Vessey, Reza Zarei and Antonio Anax F. de Oliveira

Lhasa Limited, Leeds, West Yorkshire, United Kingdom

Abstract

The traditional paradigm for safety assessment of chemicals for their carcinogenic potential to humans relies heavily on a battery of well-established genotoxicity tests, usually followed up by long-term, high-dose rodent studies. There are a variety of problems with this approach, not least that the rodent may not always be the best model to predict toxicity in humans. Consequently, new approach methodologies (NAMs) are being developed to replace or enhance predictions coming from the existing assays. However, a combination of the data arising from NAMs is likely to be required to improve upon the current paradigm, and consequently a framework is needed to combine evidence in a meaningful way. Adverse outcome pathways (AOPs) represent an ideal construct on which to organize this evidence. In this work, a data structure outlined previously was used to capture AOPs and evidence relating to carcinogenicity. Knowledge held within the predictive system Derek Nexus was extracted, built upon, and arranged into a coherent network containing 37 AOPs. 60 assays and 351 *in silico* alerts were then associated with KEs in this network, and it was brought to life by associating data and contextualizing evidence and predictions for over 13,400 compounds. Initial investigations into using the network to view knowledge and reason between evidence in different ways were made. Organizing knowledge and evidence in this way provides a flexible framework on which to carry out more consistent and meaningful carcinogenicity safety assessments in many different contexts.

1 Introduction

The current paradigm of carcinogenicity safety assessment in many industries relies heavily on rodent animal studies, more specifically the 2-year rodent bioassay (Wolf et al., 2019). While this model has served human health protection well for many years, it has limitations that need to be addressed in a modern safety assessment setting. It is time-consuming, expensive, requires a large number of animals to be sacrificed and, most importantly, may not be the most predictive of human risk, which is the ultimate species of interest in many cases (Cohen, 2004; Boobis et al., 2016; Berry, 2017; Doe et al., 2019). Additionally, in some settings, the throughput of chemicals requiring assessment can be large, making the current approach untenable (Guyton et al., 2009).

In response, a number of new approach methodologies (NAMs) have been and are being developed with the aim to predict carcinogenic risk in humans more accurately, using fewer animals and at a lower cost, both in terms of time and money (Cohen, 2004). Furthermore, in some contexts, it has been proposed that the significant amount of knowledge already generated as part of certain risk assessments prior to these long-term animal studies may be used in a weight of evidence (WoE), thus negating the value in carrying out further animal studies (ICH S1B(R1), 2021).

One challenge with these approaches, however, is that, in general, a range of NAMs and existing data are required to replace the rodent studies adequately. These disparate pieces of evidence must be combined in a logical way to form an integrated approach to testing and assessment (IATA) and reach a conclusion relevant to carcinogenicity safety assessment. Therefore, a framework is required to contextualize this information and assess how the results relate to one another. The concept of adverse outcome pathways (AOPs) (Ankley et al., 2010) has

Received January 31, 2022; Accepted May 2, 2022;
Epub May 13, 2022; © The Authors, 2023.

ALTEX 40(1), 34-52. doi:10.14573/altex.2201311

Correspondence: Alex N. Cayley, PhD
Lhasa Limited
Granary Wharf House, 2 Canal Wharf
Leeds, West Yorkshire, LS11 5PS, UK
(alex.cayley@lhasalimited.org)

been suggested as the ideal construct to fulfil this function in a more general context (OECD, 2017).

AOPs represent a method of capturing knowledge of the mechanisms by which an adverse event may occur following perturbation of a biological system. This is achieved by creating a knowledge graph associating causally related biological key events (KE), starting at the molecular initiating event (MIE), where a stressor (usually a chemical) perturbs a biological component to start a chain of events, and ending in an adverse outcome (AO). The KEs should all be measurable and are linked to one another through key event relationships (KERs). The pathways usually represent knowledge through events occurring at different levels of biological complexity, starting with events at the molecular level, through the cellular and organ level, to the individual or population. The pathways should capture evidence supporting the assertions relating to each of the components and the context in which they are relevant, with information supporting the KERs being particularly important. AOPs represent an ideal structure for capturing knowledge relating to toxicity as they allow understanding of mechanisms leading to toxicity to be captured in a transparent way, the applicability of the pathways within different contexts (e.g., species, sex, life stage) to be assessed, as well as provide the potential for data and predictions to be contextualized and related in a meaningful way to support better decision-making (Ball et al., 2021; OECD, 2017).

In fact, AOPs have already found some practical application for organizing knowledge and evidence in the development of defined approaches for the assessment of skin sensitization (OECD, 2021) and have been suggested as useful constructs in the organization of knowledge in many domains, including carcinogenicity (Sasaki et al., 2020; Lynch et al., 2019; Jacobs et al., 2016, 2020; Heusinkveld et al., 2020; Stalford et al., 2021; Arnesdotter et al., 2021; Johansson et al., 2020).

For the approach to be useful in carcinogenicity safety assessment, existing knowledge of AOPs relating to cancer needs to be captured and associated with evidence that can be used in IATAs. Methods of using this evidence in the context of AOPs to reach meaningful and transparent conclusions can then be developed. Not only are single AOPs required but also an understanding of how these individual AOPs interact in a network to lead to a response (Knapen et al., 2018; Ball et al., 2021). Therefore, a coherent network of AOPs relating to carcinogenicity is required in order to make good decisions and understand any knowledge gaps when making assessments.

There are already several publications (Helm et al., 2020; Nymark et al., 2021; Hill and Conolly, 2019) and repositories capturing AOPs relating to cancer (AOPWiki[1]), as well as collaborative projects focused on delineating AOPs associated with specific aspects of cancer (Sasaki et al., 2020; Lynch et al., 2019; Jacobs et al., 2016; Heusinkveld et al., 2020). In addition to these more recent activities, knowledge relating to the different modes of action (MoAs) leading to cancer have been documented in the public literature for many years (Cohen et al., 2019), and this knowledge has been captured and extended in expert rule-based predictive systems such as Derek Nexus produced by Lhasa Limited for over 30 years (Derek Nexus[2]). The information contained in this predictive system relates structural alerts for specific compound classes with the evidence and hypotheses thought to explain their toxicological activity. While the knowledge is not delineated directly in the AOP format, it represents a well-curated wealth of public and private knowledge of pathways leading to carcinogenicity that may be harvested and converted into the AOP format to be added to the public knowledge and leveraged in original ways.

With this in mind, the knowledge captured in Derek Nexus relating to carcinogenicity was used as a starting point to build an integrated network of AOPs for this endpoint. Additional work was then undertaken to refine the AOPs and address potential gaps in the network. In addition to the AOPs, relevant evidence sources (assays and (Q)SAR models) were linked to these AOPs in the appropriate places in order that they could be used to contextualize information on specific individual compounds and bring the AOPs to life for the purposes of carcinogenic safety assessment. A recently developed data structure for capturing AOPs and the evidence associated with them within a network was used to store and reason with this knowledge, and the prototype software program described in the work was further developed in order to expose and manipulate the data (Ball et al., 2021).

Initial investigations were made into how this approach might be used to profile data sets and group compounds by their potential mechanisms leading to carcinogenicity, as well as how reasoning between evidence on this framework may aid in making more specific safety assessments for carcinogenicity according to current and future guidance.

2 Materials and methods

Assessing alerts from Derek Nexus

Alerts associated with endpoints relating to carcinogenicity were selected from the Derek Nexus 2020.1 knowledgebase. These included alerts associated with both genotoxic and non-genotoxic mechanisms (as defined in Jacobs et al., 2020) that may lead to carcinogenicity and comprised alerts affiliated with the following endpoints: mutagenicity, chromosome damage, non-specific genotoxicity, estrogen receptor modulation, 5α reductase inhibition, and carcinogenicity (either directly or through Derek Nexus reasoning, and including alerts associated with photo-activated mechanisms).

The commentary associated with each of the alerts for these endpoints was analyzed, and any information captured in these comments pertaining to the MoA leading to the toxicity observed for the chemical class was converted into KEs using a consis-

[1] https://aopwiki.org/

[2] https://www.lhasalimited.org/products/vitic.htm

tent approach to name the KEs, assigning each KE appropriate process, object, and action terms, and linking the KE concepts to the appropriate ontologies, as described in the work carried out by Ives et al. (2017). This standardization in the capturing of the KEs ensured consistency during AOP network development and reduced the likelihood of different KEs being captured that describe the same concept.

Building an AOP network from literature review

With the information on the KEs associated with chemical mechanisms relating to carcinogenicity having been extracted from the Derek Nexus knowledgebase, the KEs identified were compiled into a skeleton network and the initial extrapolation made linking them to the adverse outcome of carcinogenicity.

Literature reviews were then undertaken to validate the initial associations and extrapolations that had been made and expand on the events in the pathways in order to form full AOPs. Care was again taken at this stage to standardize the structure and terminology used in KE names and ensure concepts linking different pathways into a network were captured consistently and at an appropriate level to avoid compromising the integrity of the network. The evidence used to identify the KEs and KERs was primarily based on biological plausibility of associations (OECD, 2018), although empirical evidence and essentiality were invoked where appropriate. This evidence was captured within the relevant objects of the data structure in a commentary and included references and links to the primary literature from which the knowledge was taken. In addition, knowledge of the context in which a KE, KER or AOP is applicable, including the species, sex or life stage of relevance along with the context of the cell or organ in which the events or relationships occur, were captured in a systematic way within the data structure for future use. This included evidence both for and against applicability for a given context.

While the level at which the KEs for the AOPs were captured was generally dictated by the common guiding principles of AOP development, recommending the KEs captured represent "critical steps or check-points along the path to adversity, which are both measurable and have potential predictive value" (OECD, 2018), it was also deemed important to capture some of the more detailed knowledge relating to the more specific aspects of the biological pathways being perturbed described in the public literature. It was felt that this detail will be useful as NAMs develop, especially those relating to omics, measuring gene or protein expression and biomarkers relevant to carcinogenicity pathways. With the knowledge captured in this detail, the concept of grouping more specific KEs into more general KE groups (KEGs) of related events was developed as a way of capturing this relevant knowledge while not overwhelming the user when presenting it. This is a concept being put into practice after being described and implemented as part of a previous publication (Ball et al., 2021).

Associating evidence with AOPs

Once a detailed AOP network relating to carcinogenicity had been developed, subsequent research was undertaken to identify existing and emerging assays to be associated with the appropriate places on the AOPs. Derek Nexus was again used as a data source for the assays selected in conjunction with the toxicity database Vitic[3], developed by Lhasa Limited, and the assays relating to carcinogenicity captured in that database. Assays explicitly selected to fulfil regulatory guidance such as ICH S1 and ICH S2 (ICH S1B, 1997; ICH S2(R1), 2011) were prioritized for inclusion as well as those with associated OECD guidance. However, a more general literature survey was also undertaken to identify new and emerging assays that may become part of a weight of evidence approach to carcinogenicity assessment in the future (Jacobs et al., 2020; Bryce et al., 2016; Hendriks et al., 2012). Particular attention was paid to associating binding assays relating to targets pertinent to carcinogenicity assessment (Jacobs et al., 2020).

The selected assays were then associated with the most relevant KEs within the AOP network through the concept of assay measurements, these being different types of observation that can be made for any given assay, and it is possible that each observation type may measure a specific KE.

In addition to the assays and measurements, the concept of assay exceptions, as described by Ball et al. (2021), was also associated with the various assays where knowledge was available. This applied predominantly to well-established assays for which compounds from a particular chemical class, those acting by specific mechanisms or having certain structural properties, are thought not to be assessed well by an assay, either through under- or overprediction of their toxicity.

Following association of the relevant assays with the network, data for individual compounds, predominantly captured in the toxicity database Vitic[3], were connected to the assays where data was available. In the main part, the data for each assay represented a categorical call for each compound and measurement combination. The categorical calls were generated by combining individual study and protocol results captured in the database using a defined set of rules, where a conservative call was made to reach an overall conclusion (the most positive result observed being taken in preference)[4].

Since most of the pathways delineated in the AOP network were derived from knowledge captured in Derek Nexus alert comments, it was then relatively straightforward to associate the individual Derek Nexus alerts with the relevant KEs on the network. All Derek Nexus alerts in the 2020.1 knowledgebase associated with the endpoints of carcinogenicity (either directly or through Derek Nexus reasoning), mutagenicity, chromosome damage, and non-specific genotoxicity (including photo-activated) were assessed for their association with the KEs in the network. Where possible, the alerts were associated with specific KEs. This allowed for specific knowledge relating to potential

[3] https://www.lhasalimited.org/products/vitic.htm

[4] https://www.lhasalimited.org/publications/summation-of-toxicity-data-in-vitic/3918

MoA to be captured for a prediction along with the toxicity which this may lead to. In instances where an alert could not be associated with a particular MoA, and therefore KE, due to limited knowledge for this compound class, the alert was solely associated with the AO of relevance to the endpoint being predicted, in order that this information was not lost.

Capturing knowledge within a common data structure

The knowledge and associations derived from this work were all captured in the data structure and prototype software described by Ball et al. (2021). The work is currently being transferred into the software, Kaptis[5], produced by Lhasa Limited.

Analyzing knowledge on an AOP framework

With knowledge relating to carcinogenic potential for many compounds now contextualized on an AOP network, the data could be analyzed within this context. Preliminary work to profile data sets according to the likely AOPs they activate was undertaken. A general method was developed allowing for knowledge to be selected, an evidence base to use in the assessment chosen, and an overall conclusion drawn. This aimed to highlight the power of capturing and using carcinogenicity knowledge in this framework and begin able to investigate how it might be combined in different ways for different use cases to support decision-making in chemical safety assessment.

Two different data sets were used to undertake these investigations and act as test sets. The first data set was built based on results from chronic rodent carcinogenicity studies in an attempt to represent the current paradigm and knowledge space. The initial data were extracted from the toxicity database Vitic[3]. An overall carcinogenicity assignment per compound was generated based on combining results from individual studies and taking a conservative assessment of the findings such that a tumor being found in any one study would lead to an assignment of carcinogenicity for the compound. This resulted in a data set of 2420 compounds, with 1211 having a positive call for tumor formation and 1078 being assigned as negative. 131 compounds where the results were inconclusive or equivocal were removed from the test set, producing a final test set of 2289 compounds. The second data set was derived from the compounds for which an IARC categorization was available (IARC, 2019), with an assessment of their carcinogenic potential to human having been made based on the data available. The categories defined by IARC in these classifications were used in the analysis. Compounds assigned to IARC categories 1 and 2a and 2b are thought to be human-relevant carcinogens in some respect, with varying degrees of evidence and confidence associated with their assignment. In the first instance, all of these categories were considered a positive call for human-relevant carcinogenicity. Compounds in IARC category 3 were removed from the analysis since the carcinogenic potential to humans of this category is unclear for various reasons. This category includes compounds with positive results in animals with a lack of human relevance but also compounds for which the assignment is just due to a limitation in data, and it is difficult to differentiate the different reasons for chemicals being placed in this class (IARC, 2019). From the data provided by IARC (IARC list of classification[6]), 1084 unique agents were identified, and unique resolvable chemical structures could be determined for 810 of these after structure standardization. 393 of these had at least one category 1, 2a or 2b classification, 408 had a category 3 classification, and 9 had no individual classification as they had been assessed as part of a wider chemical class. The 393 compounds assigned as category 1, 2a or 2b were taken forward in the analysis.

3 Results

3.1 Assessing alerts from Derek Nexus

Derek Nexus is an expert rule-based system for the prediction of toxicity. The knowledge base underlying the predictions provided by the software is composed of structural alerts, example compounds, and reasoning rules linking the alerts and examples to toxicity endpoints with specific reasoning levels (Derek Nexus[2]). The structure activity relationships (SARs) represented in the structural alerts were developed by experts from a variety of public literature and confidential data. The knowledge base is particularly well developed for the toxicity endpoints relating to genotoxicity (mutagenicity, chromosome damage), skin sensitization, carcinogenicity, hepatotoxicity, and developmental and reproductive toxicity, although other target organ endpoints are also represented within the knowledge base. Derek Nexus largely predicts toxic hazard, which may or may not develop into risk depending on exposure conditions. Several endpoints are developed primarily based on defined assay results, so in these cases the predictions are closely related to predictions for these assays.

When an alert is activated, the user is presented with the toxicological endpoint with which the alert is associated and the level of belief (reasoning level) with which toxicity is thought to be observed in different species, along with a written rationale outlining the association between the compound class and the endpoint, the evidence on which it is based, and any assumptions that have been made (Fig. 1A). Often this commentary will contain information and references relating to the proposed mode of action (MoA) thought to lead to the toxicity observed for the compound class. This knowledge is what made the Derek Nexus alerts such a good starting point for AOP development as it was possible to convert the knowledge captured in this commentary first into KEs and then AOPs for each alert (Fig. 1B), which could then be standardized and, following literature review, developed into a network.

In the example shown in Figure 1, the compound haloperidol activates an alert associating the butyrophenone chemical class with the toxicity endpoint of carcinogenicity in Derek Nexus. It should be noted that this compound is associated with the end-

5 https://www.lhasalimited.org/products/kaptis.htm

6 IARC, List of Classifications: Agents classified by the IARC Monographs, Volumes 1-129. Last updated: 2021-07-22 02.00pm (CEST). https://monographs.iarc.who.int/list-of-classifications

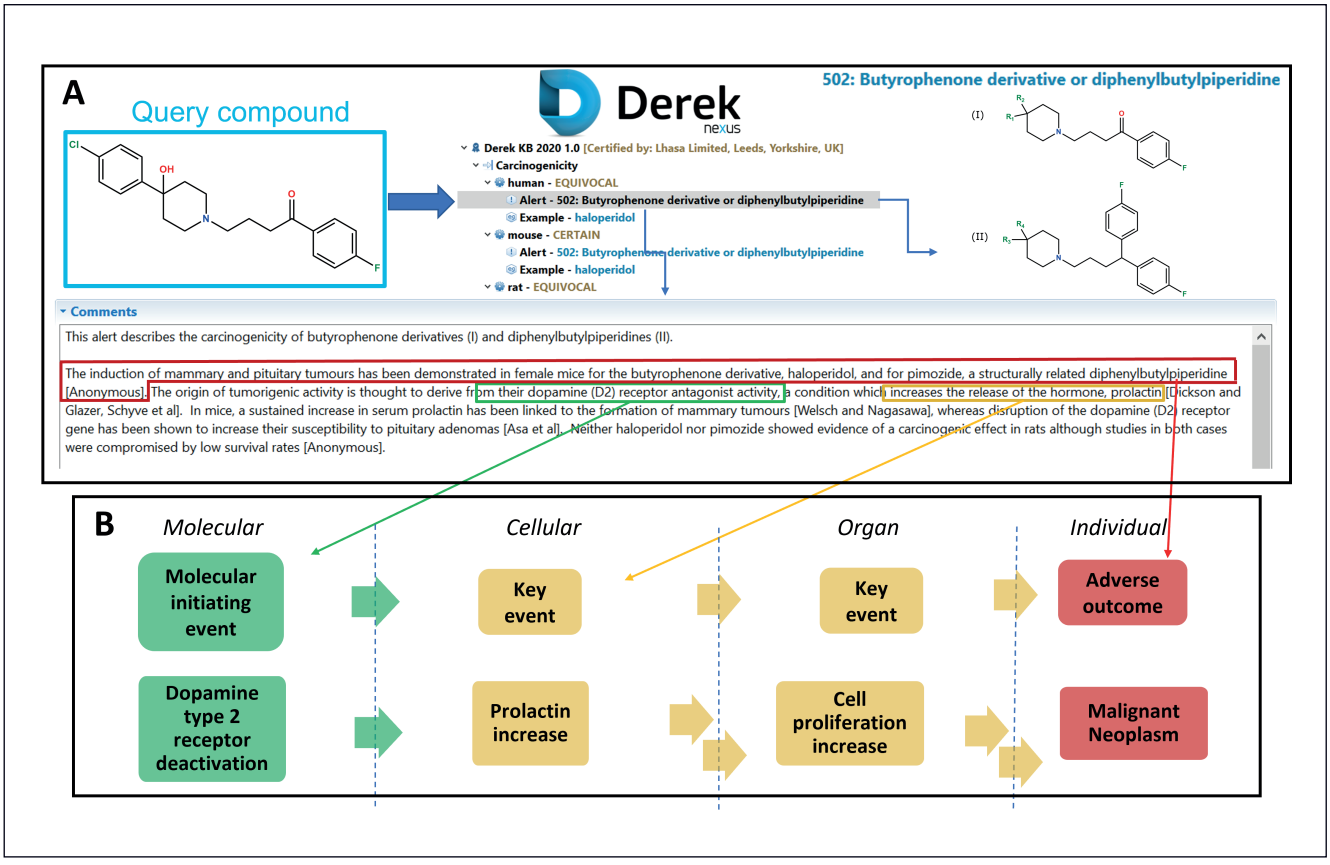

Fig. 1: Converting knowledge captured in the alert comments of a Derek Nexus carcinogenicity alert into KEs for use in AOP network development

point with different levels of confidence (reasoning levels) for different species. There is a strong association in the mouse, while for other rodents and humans the association is weaker. The reasoning behind these differences is reflected in the comments associated with the alert. These comments discuss different results obtained in multiple species as well as describing the MoA by which the compound class may cause carcinogenic activity. In this case, the ability of the compounds to act as agonists of the dopamine type 2 receptor, leading to increases in the levels of prolactin, has been linked to the formation of mammary tumors in mice. Evidence relating to the species extrapolation associated with this MoA as well as the results obtained in other species results in the reasoning levels predicted, and information relating to this is supplied in the alert comments. The relationship between dopamine receptor agonism and the generation of malignant neoplasms has not been observed in the same way in humans (Lichtermann et al., 2001; Wang et al., 2002). It is also easy to see how these pieces of information can be taken from the comments and transcribed directly into KEs in a rudimentary AOP, where deactivation of the dopamine type 2 receptor leads to a prolactin increase which in turn leads to cancer. The pathway can also be supported by the references used to make the assertions and the species relevance also captured and referenced within the AOP data structure (Ball et al., 2021).

The knowledge base is particularly well developed for endpoints associated with carcinogenicity, and the alerts relating to the endpoints of mutagenicity, chromosome damage, and non-specific genotoxicity as well as carcinogenicity were all assessed to help build the network. This led to 351 alerts in total being considered when looking for KEs and pathways associated with carcinogenicity, with the large majority of these alerts being related to the endpoints of mutagenicity, chromosome damage, and carcinogenicity (Tab. 1). While there was a small minority (5.4% overall) of alerts for which no information on the MoA was known, and so a KE could not be assigned, some level of mechanism had been proposed and captured for most of the alerts, and for many, more than 1 KE could be associated with the same alert (average of 1.3 KEs per alert). These represented cases where toxicity may have been caused by multiple MIEs for a given chemical class and those where more detailed information on KEs in a single pathway were delineated and could be captured. The number of unique KEs identified was much lower than the number of alerts investigated due to the fact that many alerts in Derek Nexus may cause toxicity via the same MoA, and therefore will translate into the same KEs. While there is more diversity in MoAs (and therefore KEs) leading to carcinogenicity, there are only relatively few mechanisms which will lead to mutagenicity, involving direct damage

Tab. 1: Endpoints and number of alerts associated with carcinogenicity in the Derek Nexus 2020.1 knowledgebase
Number of alerts where a more specific MIE/KE can be associated with the alert.

Derek Nexus endpoint	Number of alerts assessed[c]	Number of alerts for which at least one KE could be assigned	Number of unique KEs identified
Carcinogenicity[a,b]	108	101	34
Mutagenicity[b]	151	143	12
Chromosome damage[b]	96	90	15
Non-specific genotoxicity[d]	5	5	2

[a] Including alerts associated with carcinogenicity through reasoning; [b] including alerts requiring photoactivation; [c] a single alert may be counted multiple times for different endpoints with which it is associated; [d] including alerts built using data from assays where results indicate genotoxicity but cannot be easily assigned as leading to a mutagenic or chromosome damaging outcome (e.g., unscheduled DNA synthesis (UDS) assay)

to DNA (Tab. 1). There are more MoAs leading to chromosome damage, of which mechanisms leading to mutagenicity are a subset, hence the relative increase in KEs identified for this endpoint (Tab. 1). In the case of the alerts where a MoA could not be assigned, the alerts were linked directly with KEs relating to the toxicity endpoint they were predicting (e.g., inherited DNA mutation, chromosome damage), so the knowledge captured by these alerts was not lost.

3.2 Building an AOP network through literature review

After associating the knowledge captured in the Derek Nexus alerts with KEs and linking this to the endpoint of carcinogenicity, a more thorough review of the literature was undertaken in order to test and support these associations as well as expanding upon the pathways where appropriate and making sure the knowledge was integrated into a coherent network. This work resulted in 38 MIEs being identified associated with 37 different AOPs. As might be expected, for many AOPs there were multiple routes by which a MIE might lead to the AO of malignant neoplasm or a potentially related adverse event, and so 375 pathways were delineated in the network described by the interaction of 142 KEs (10.1 pathways and 16.1 KEs on average for each AOP, supporting the assertion that user-defined AOPs are rarely purely linear entities or linear AOPs (LAOPs) (Pollesch et al., 2019)). This large discrepancy between the number of pathways delineated in the AOP network and the number of individual AOPs also highlights why viewing this knowledge as an interconnected network has great benefits, both because the definition of a single AOP is a rather artificial boundary given the multiple pathways which usually define a single AOP and the fact that pathways between different AOPs will often be interconnected (Fig. 2A).

It should also be noted that these numbers represent the pathways captured at their highest level, where knowledge relating to more specific events associated with biological pathways had been grouped into KEs more relevant to events usually measured at the level of the AOP. Without grouping, 325 events had been identified and the number of MIEs increased to 47. One example where grouping was employed is shown in Figure 3. Events in the AOPs relating to certain KEs leading to cancer endpoints have been investigated in some detail in the literature. For example, the activation of these receptors has been shown to affect multiple cell signaling pathways following heterodimerization, leading to up- or downregulation of specific messenger proteins and pathways (Bayly et al., 1994; Huang et al., 2005; Lien et al., 2013; Tian et al., 2011; Columbano et al., 2005; Kodama et al., 2011; Robbins et al., 2016). While perturbation of the cell signaling pathways associated in Figure 3 may not solely be caused by interaction with these specific MIEs, capturing knowledge of very specific event associations, such as Gadd45beta increase (Kodama et al., 2011), as potential biomarkers may be useful when relating the pathways to NAMs. Indeed, it is likely that a combination of different biomarkers will be required to implicate a liability for cancer occurring via a specific MoA, and these may include markers from the cell signaling pathways as well as those both up- and downstream of these events, with temporal relationships also perhaps being important when associating data from NAMs with the pathways. When visualizing an AOP or network of AOPs and making decisions, it may, however, be more desirable to see and digest the information at a higher level and interrogate in more depth as required. Therefore, under the proposed KEGs concept the user would be able to navigate through these different views and the evidence would be associated appropriately within the group (Fig. 3). The human relevance of the various pathways may also be taken into consideration when grouping and displaying the AOPs. The pathways represented in Figure 3, for instance, may have different amounts of evidence associating them with different species (including human), and the user may wish to group further or filter based on this species relevance.

There were 48 KEGs defined in the carcinogenicity network and these contained 147 different KEs. This approach to knowledge grouping may be particularly useful when evidence relating to more general concepts associated with cancer, such as the key characteristics (KC) of cancer, needs to be captured and reasoned with. It means that data or predictions related to KCs can be associated with the relevant group in the network and the interplay between the different KCs captured and reasoned between, in addition to the more specific knowledge (Smith et al., 2016).

Furthermore, by annotating the individual KEs with appropriate terms relating to the biological concepts they represent

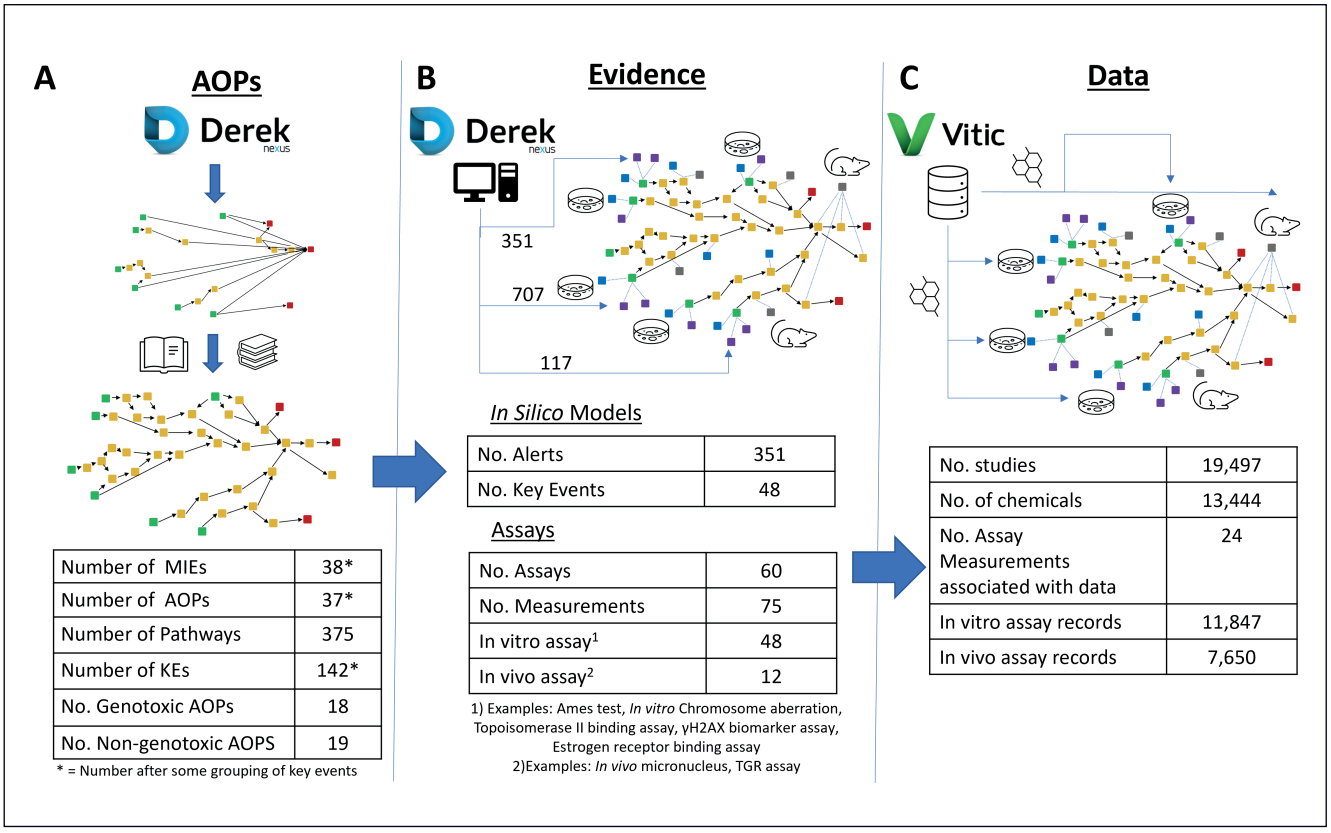

Fig. 2: Numbers of different AOP components and evidence held within the carcinogenicity AOP network developed

within the construct of the process, object, action, and context (POAC) outlined by Ives et al. (2017) and linking these terms to appropriate ontologies, it is possible to view different parts of the carcinogenicity AOP network in different levels of detail according to the preference of the use case in question. For example, in Figure 3, the general AOP starting with nuclear receptor binding, heterodimerization, and activation in the liver is delineated. This representation, in fact, summarizes multiple individual AOPs which relate to interaction with individual nuclear receptors. The processes and objects linked to the individual KEs and their association in the ontologies linked to these individual KEs might be used to view these individual AOPs at a higher level where all AOPs relating to activation of individual nuclear receptors in the liver (CAR, RAR, RXR, PPAR) can be viewed as a single pathway where the more general relating term of nuclear receptor activation in the liver can be used as a grouped concept. An example of how the separate AOPs may be grouped in this way within the entire network is illustrated in Figure 3. This shows how the network may be viewed and interacted with at multiple levels by using ontology labels and KEGs to condense part of the network. The type of visualization shown may fit very well with the level of AOP outlined in the IATA described by Jacobs et al. (2020) for non-genotoxic carcinogenicity assessment and to allow different views of this knowledge. Work is still required on how to establish useful levels of the ontologies at which to represent groups of AOPs.

As well as the KEs and KERs documented within the literature for a given pathway, the species relevance for which the associations were applicable were also captured from the knowledge available. Both evidence for and against the applicability of different KERs and entire AOPs in different species was stored within the data model in a weight of evidence. This led to the majority of the AOPs being initially assigned as being applicable to mammals in general (mostly based on rodent data but with reason to believe they are more widely applicable). Currently, 3 (1.78%) KERs and 4 (9.8%) AOPs had evidence indicating that they were not applicable to humans and were assigned against this species with moderate strength. The knowledge used to reach these conclusions was referenced and stored within the data model.

3.3 Associating evidence with AOPs

With a network of AOPs having been developed, effort was then put into associating relevant evidence sources to the appropriate places on the pathway using the data model described previously (Ball et al., 2021) (Fig. 2B). 60 assays were associated with 52 different KEs using the concept of assay measures in order to take into account the fact that an assay may measure multiple different things and that each measure may be associated with a different KE. While well-established assays were captured and those with OECD guidelines prioritized, newer emerging *in vitro* assays, particularly those relating to binding

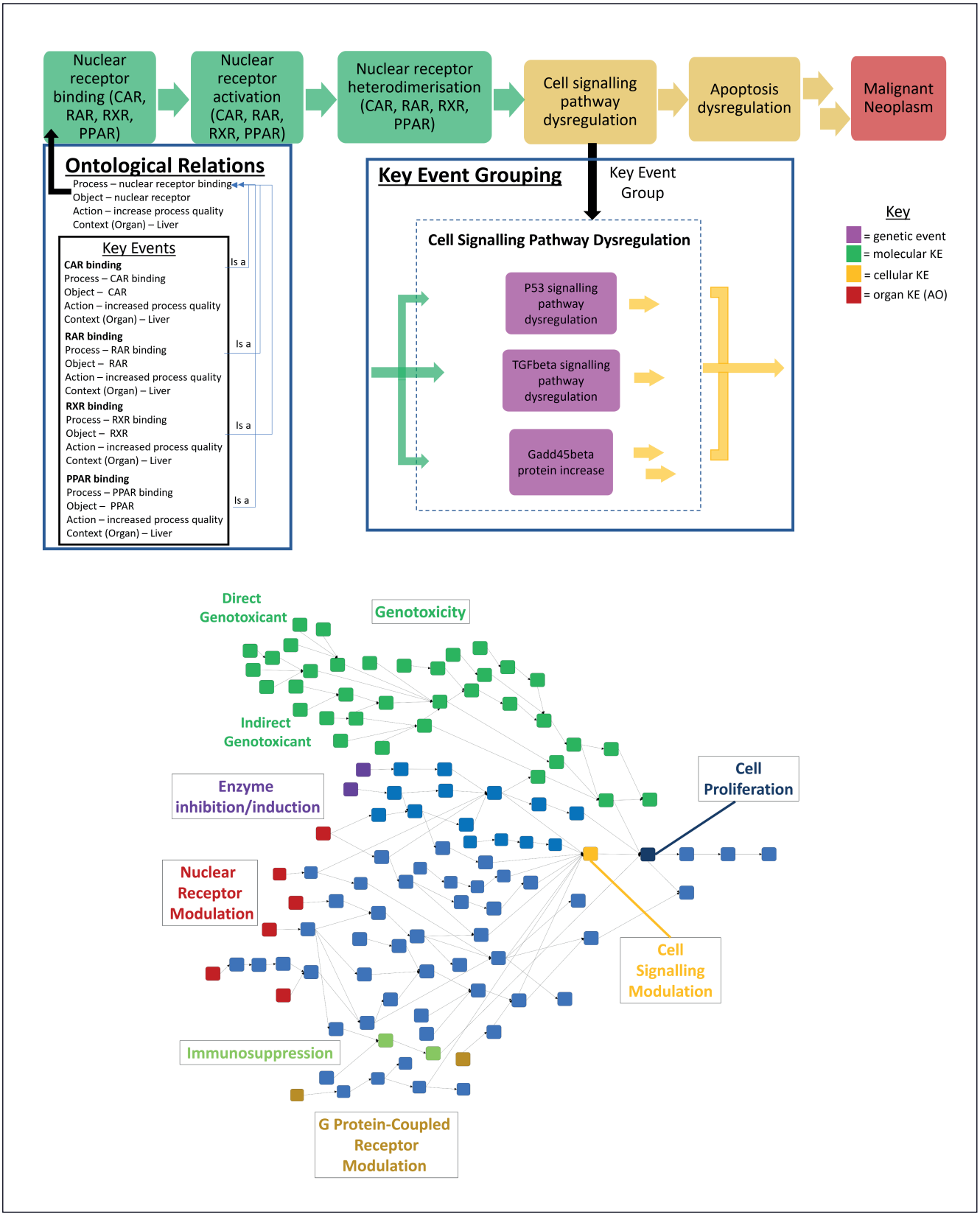

Fig. 3: Key event groups (KEGs), an example of how knowledge could be represented and visualized at different levels of detail within pathways

or biomarker measures, were also captured (Bryce et al., 2016; Dix et al., 2007; Hendriks et al., 2012), which resulted in more *in vitro* assays (48) being associated with the network than *in vivo* assays (12). At the same time as capturing knowledge on assays, information about the limitations of these assays was also captured. This was described in terms of assay exceptions (Ball et al., 2021). For each assay there may be an applicability domain, and often the limitations of a given assay for a certain area of chemical space or mechanism are well known (Jacobs et al., 2020). For example, chemicals from the acid halide class are known to give unreliable results in the Ames test as a result of their direct reactivity with some of the solvents used in the test, either activating or deactivating them (Amberg et al., 2015). As a consequence, results from this assay for this compound class should be treated with a degree of caution, and such information was captured in the assay exception table. These types of assay limitations have also been described by Jacobs et al. (2020) in the consideration of assay performance considerations (category 2) during their assessment of assays relating to non-genotoxic carcinogenicity for use in an IATA. To date, 13 assay exceptions have been captured for the assays relating to carcinogenicity assessment, with the majority being linked to compound class and the limitations of particular assays run under specific protocols to predict these compound classes.

In addition to the predictivity of a given assay and measurement for a particular compound class, it is also important to take into consideration the more general predictivity of the measure (and hence KE) for the AO of interest in order to combine and weight the evidence accordingly (Jacobs et al., 2020). For example, hypertrophy is a histopathological finding from repeat-dose studies, which, according to the proximity of the association on the AOP, would be closely associated with cancer. However, this biomarker has been found to have a relatively weak association with the AO and therefore should be weighted and combined with other evidence to reach a conclusion (Sistare et al., 2011). Therefore, capturing the sensitivity and specificity of the measure for the AO of interest is important information that can be used in a number of roles and is an area of current research within our network.

As well as assay findings, evidence relating a compound to a KE may also come from *in silico* predictions. As described previously, Derek Nexus alerts were used as a starting point to build the network, and so these were the first predictive models to be associated with the network. It was possible to associate 351 alerts with 48 different KEs in the network. Since all alerts in Derek Nexus have an associated toxicity endpoint in the software, the alerts could be associated with KEs on the pathways even if a clear MoA had not been established for that class and, as a result, the KE relating to the toxicity endpoint was used in these instances. Where a MoA had been established (and used to develop the AOP network), the alert could be associated not only with the toxicity endpoint but also with KEs relating to the MoA suggested, thus giving more context to the predictions being made. Using this method meant that nearly all the alerts investigated could be associated with the network in some form or another.

With knowledge captured in this way, we were then able to bring the network to life using experimental data along with the predictions to help associate chemical structures with the assays and KEs within our network. To this end, we employed the toxicity database Vitic[3] to associate assay study data with the appropriate assays on our AOP network. This required selecting data from Vitic and then generating an overall call per compound for each individual assay measure. The method used to generate this overall call was generally a conservative approach, in most cases taking a positive result or finding from any individual study as an overall positive result while also taking some consideration of the protocol employed[4]. The Vitic Lhasa Summary call table was used to access the data at this level for the assays available (Ames mutagenicity, *in vitro* chromosome aberration test, *in vitro* micronucleus test), and for the remainder calls were generated by synthesizing the individual study data outside of Vitic. Most of the data collected to this point provides a categorical call of activity for a given measurement, although associating more continuous data may be investigated further in the future. This work resulted in 19,400 studies being associated with our network, covering approximately 13,400 chemicals. 24 measurements from different assays have been associated with data. While the majority of the data is associated with *in vitro* assays, it was generally the new and emerging assay types for which data was lacking, and these will be populated in the future as evidence and evidence sources become available.

This knowledge was all captured within a database structure visualized in the prototype software described in our previous publication (Ball et al., 2021). The data and functionality is currently being transferred into a full program, Kaptis[5], where it can be visualized, interrogated, and manipulated according to the user's preference. The knowledge can be viewed in a single AOP view or as a network of KEs, and the evidence associated with the pathways can be seen alongside it.

3.4 Analyzing knowledge on an AOP framework

While structuring knowledge of carcinogenicity around the framework of an AOP network is academically appealing, it is also important that knowledge structured in this way can be accessed and capitalized on in order to make better decisions in carcinogenic safety assessment. Digitalizing this information, making the associations explicit within a database, and embedding this in a software tool goes some way to making the knowledge framework accessible and useful. Therefore, initial investigations were undertaken on how to leverage knowledge of carcinogenicity captured in this way. One key benefit of having knowledge associated with a given chemical structured in terms of the AOPs is that profiling data sets in this way will give context and inform not only on how closely these pathways associate with the adverse outcome of carcinogenicity but also allow a more meaningful grouping of compounds, and extrapolation of activity between compounds in specific AOP profile groups. In addition, the framework allows a more meaningful and consistent method of combining data so that reliable and transparent decisions can be made using the evidence available, and any new evidence types can be incorporated into the framework, as long

as it is clear where on the AOPs it should be associated. This allows carcinogenicity assessment to be moved towards a more integrated approach, as proposed by IATAs, which can incorporate all relevant knowledge and can develop and evolve as the technology does, avoiding the limited flexibility, applicability, and responsiveness of some traditional testing strategies. Examples of using the knowledge captured in both these ways are discussed below.

3.4.1 Profiling a compound/data set using AOPs

Knowledge of the MoA by which a compound may cause carcinogenicity can have a profound effect on decisions made relating to its carcinogenic potency, human risk, and subsequent actions to be taken when carrying out a safety assessment. With this in mind, an initial experiment with our network was undertaken to profile both a data set describing rodent carcinogenic potential of a chemical set, derived from the toxicity database Vitic[3], as well as a dataset describing the carcinogenic potential and human relevance of chemicals as assessed by IARC (IARC, 2019), described in Section 2. The compounds were assigned according to the AOPs by which they may cause their carcinogenic effects. The AOPs leading to the potential AOs relating to carcinogenicity were assigned for each compound based on those where the MIE of the AOP could be associated, either through a prediction or data, with the structure and where there was further evidence supporting, or at least no other evidence along the pathway contradicting, this association.

The results of the profiling are shown in Figure 4 (where KEs that were activated by > 20 compounds are displayed on the graph). As can be seen, the AOPs activated by both data sets are heavily biased towards those acting through a genotoxic mechanism and, more specifically, through electrophilic reaction with DNA nucleobases. This may indicate a feature of the test sets and could be used to profile the toxicological space captured in these data. This result would, in fact, not be unexpected given that, historically, evidence and carcinogenicity testing has been focused on a toxicity space that relates to genotoxicity, and this is a category of carcinogen that has been studied and tested widely using alternative methods to the chronic rodent carcinogenicity study, positive results from which would prompt further testing and labelling (ICH S2(R1), 2011). However, it should also be noted that, because of this focus on genotoxic mechanisms and the prevalence of *in silico* modelling and testing carried out in this area, the bias towards these AOPs in the profiling may be a result of an increase in labelling from this type of evidence rather than other AOPs not being relevant for the compounds in these data sets. It is not clear whether the compounds activating these AOPs may also act through non-genotoxic mechanisms which are not represented in the evidence (prediction or data) or whether the carcinogenic compounds that do not activate any AOPs as part of this profiling are actually carcinogenic via non-genotoxic mechanisms. If this is the case, the limitations are in the evidence rather than the classes of compound that have been tested. This analysis of AOPs activated at a data set level is useful in establishing a general profile for the data set. Furthermore, this type of analysis may also allow better profiling of individual compounds. Knowledge of the AOP profile for a single compound will enable it to be grouped with others with the same profile, giving a better indication of other compounds that should be deemed as "similar" to the primary compound of interest when carrying out read-across approaches[7]. Knowledge of these profiles could also dictate the best method of measuring similarity in a particular instance (for example where reactive mechanisms may be in operation, a fragment-based similarity approach may be more appropriate, while a binding MoA may benefit from a pharmacophoric fingerprint method).

While the data presented in Figure 4 represents a profiling approach to AOP assignment rather than a true prediction, it can also be seen that the balance of carcinogenicity classification for each AOP differs, with some AOPs being more predictive of positive carcinogenicity findings than others.

We were interested in exploring this work further to develop our network from a profiling tool to one on which decisions around chemical carcinogenicity safety assessments could be made through a meaningful combination of all the evidence available. To this end, investigations were begun into reasoning between the evidence in the context of an AOP.

3.4.2 Reasoning between evidence in the context of an AOP

In order to make carcinogenicity assessment decisions and utilize the context which an AOP framework brings, a method of reasoning between this evidence must be developed to enable all evidence to be taken into consideration and the right weighting to be applied so that a prediction can be made based on the WoE.

In our previous publication, we suggested that there may be multiple different ways in which evidence may be combined within the context of an AOP in order to reach a conclusion. This ranged from a simple conservative approach, where any one single positive finding would be enough to assign a compound as a carcinogen to more complicated methods based on weighting each piece of evidence on the AOP and combining these weighted results, along with undercutting arguments if appropriate (Ball et al., 2021). In the current work we looked to build these different approaches into a framework on which they could be used and different types of reasoning could be applied depending on the use case in question.

An initial general reasoning method was conceived whereby evidence was resolved at the various levels of the AOP and built up to reach an overall conclusion. First, the knowledge required to be used when making an assessment was filtered from the wider knowledge contained in the AOP network, allowing for the problem to be addressed in a particular use case to be formulated

[7] Draft EFSA Scientific Committee guidance document on scientific criteria for grouping chemicals into assessment groups for human risk assessment of combined exposure to multiple chemicals.
https://connect.efsa.europa.eu/RM/s/publicconsultation/a0c1v00000HnXIB/pc0014

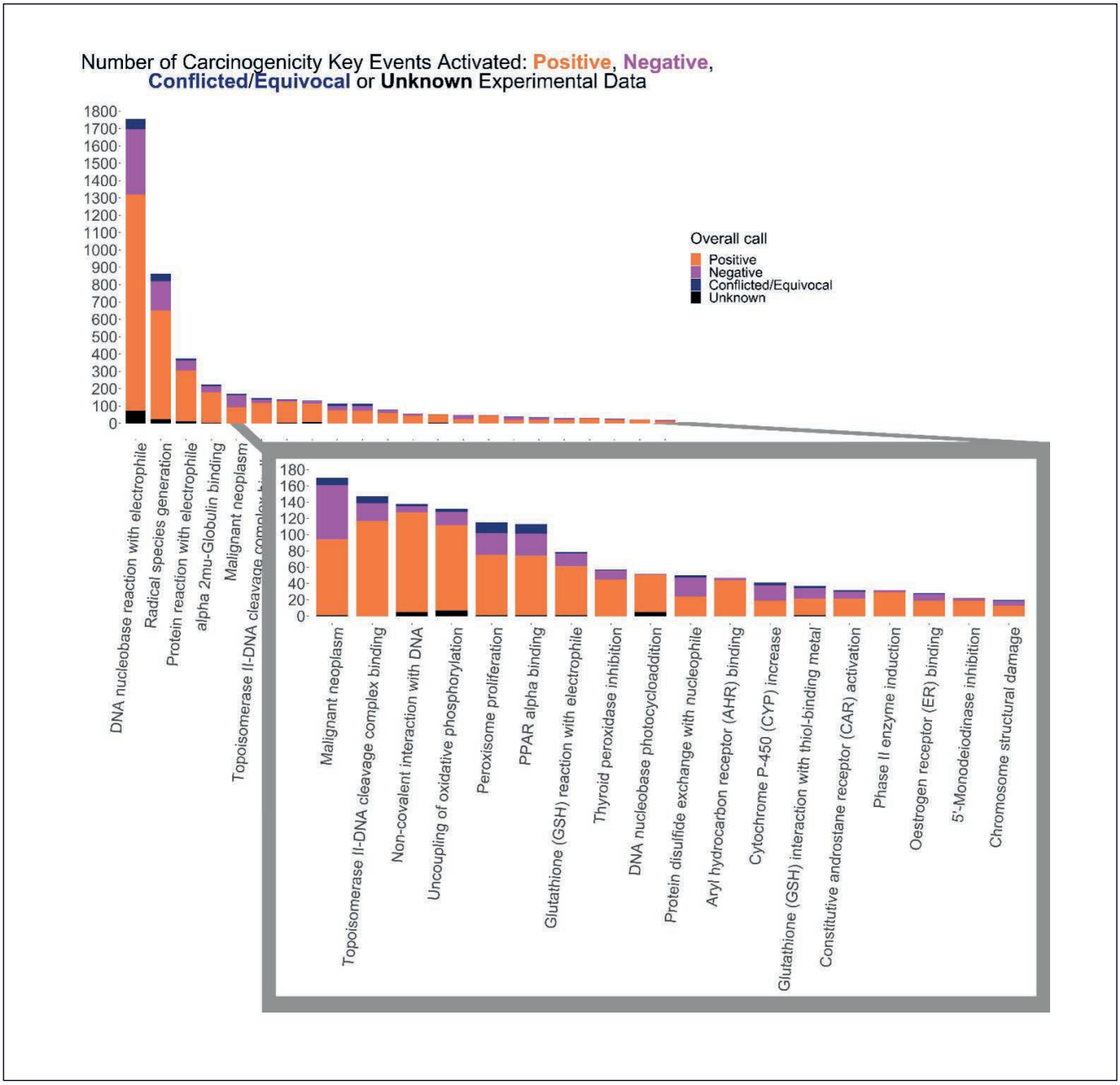

Fig. 4: Graph profiling distribution of rodent carcinogens according to the KEs with which they have been associated, based on the Derek Nexus alerts activated

KEs activated by > 20 compounds are displayed on the graph.

(Knapen et al., 2018). Potentially any facet of the network could be used to filter the network, but in this case a prediction about carcinogenicity is being made, so all AOPs in the network relating to carcinogenicity were used in making the decision (Fig. 5, step 1). Next, the types of evidence to be used in making the decision are selected. In this case we used all available evidence in the WoE, but a minimum may be selected or different types used as this use case becomes better defined or different evidence sources are approved (Fig. 5, step 1). For example, if the decision being made using this approach was within a regulatory context, a specific set of approved evidence may be labelled and selected appropriately at this stage, whereas for use cases such as chemical prioritization a broader evidence base may be used.

Once the AOP network and evidence to be used is defined, the result on each KE is resolved. This may be achieved using a simple reasoning approach, taking the most conservative result from

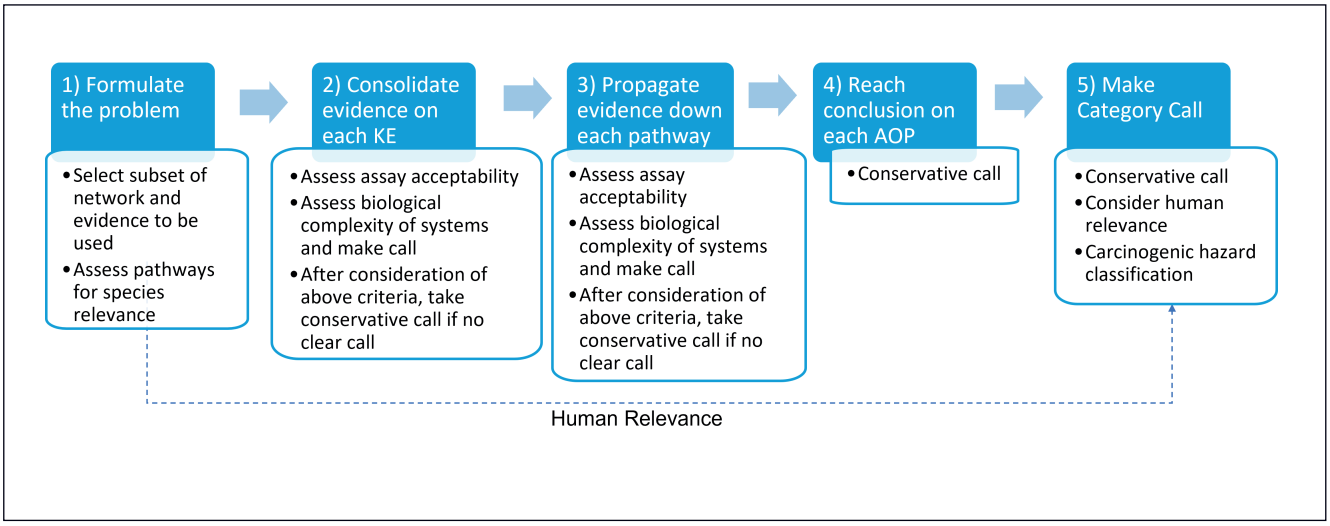

Fig. 5: Reasoning approach employed to test decision-making using an AOP framework

all the evidence available for that KE, or through more complex reasoning based on weighting the evidence associated with the KE. In the current work, the biological complexity of the test systems as well as any assay exceptions encoded in the database that may be associated with the data were taken into consideration when reaching a conclusion (Fig. 5, step 2). *In vivo* test results on a KE were taken in preference to *in vitro* and these in preference to an *in silico* prediction where there were no exceptions in the assay results identified. These WoE criteria may be expanded on or changed in the future, depending on the requirements of the individual use cases.

The signal from each KE is then propagated down the pathways and a call made on each individual AOP (Fig. 5, step 3). Again, this might be achieved in multiple different ways, and in this case the rules on propagating a signal were based on assigning activity to the KERs and giving these a lower weighting than evidence attached directly to the KEs. It may be more or less useful in different use cases to keep the concept of AOPs as individual units within the network, and it may be more useful in certain instances to carry out reasoning and view the results purely from a network of pathways perspective. However, in the current work, we kept the concept of individual AOPs as well as viewing the knowledge as a network.

Once each individual AOP/pathway classification has been made, an overall classification can then be assigned based on combining these results. This was done using a conservative method in this case (a positive conclusion on any one AOP/pathway giving rise to a positive signal overall) as this was felt to be the most appropriate way of combining any potential MoA leading to toxicity (Fig. 5, step 4).

Finally, some consideration of the human relevance of the prediction was made. This was achieved by querying the database for the human relevance of all AOPs/pathways implicated in leading to a positive prediction. If all AOPs implicated in producing the positive result were thought to be unlikely to be applicable in humans, then the prediction was deemed not to be human-relevant and the prediction annotated with this information. In the future it may be possible to expand this to more specific species predictions.

This represents a relatively general approach to reasoning, and it would be expected that it would be possible to follow these general principles while changing the specific approaches to reaching a conclusion and assessing the WoE at each level, taking into account the reproducibility of the evidence and the weighting of the individual pathways in the network, according to each specific use case (Ball et al., 2021).

Both the Vitic rodent carcinogenicity data set and the IARC classification data set described in the methods section were then processed against the AOP network using the data and prediction evidence associated with it along with the reasoning model described. Overall categorical classifications were made for each compound using the evidence and following this logic. For simplicity, the chemicals were assigned as likely human-relevant carcinogens (positive) or non-carcinogens (negative), and chemicals fitting in a third category of showing evidence for carcinogenic hazard in rodents that is not relevant to humans (positive – not human-relevant). This third category of prediction was included as a positive prediction when assessing against the Vitic rodent carcinogenicity data but was excluded from the analysis of the IARC human-relevant carcinogens (Tab. 2).

The initial findings show that there is good sensitivity for picking up potential carcinogens using this approach. The results against the IARC data set display a very good sensitivity against these human-relevant carcinogens, and this sensitivity increases slightly as more evidence is taken into account in the form of *in vitro* and *in vivo* data (Tab. 2A). The same is also true of the Vitic rodent data set, where a good sensitivity is increased as the amount of evidence used in the prediction increases (Tab. 2B). As a result, it is fair to say that coverage of this reasoning ap-

Tab. 2: Profiling a rodent carcinogenicity data set using a reasoning method based around an AOP framework

A. IARC Human Relevant Data Set

Data	Sens. (%)	TP
Derek Nexus[a]	76	291
AOP – *in silico* only	85	323
AOP – *in silico* and *in vitro*	85	326
AOP – *in silico*, *in vitro* and *in vivo*	87	332

B. Vitic Rodent Relevant Data Set

Data	Sens. (%)	TP	FP	TN	FN	Spec. (%)	BA (%)	MCC
Derek Nexus[a]	68	818	384	694	393	64	66	0.32
AOP – *in silico* only	77	930	529	549	281	51	64	0.29
AOP – *in silico* and *in vitro*	78	942	573	505	269	47	62	0.26
AOP – *in silico*, *in vitro* and *in vivo*	79	961	596	482	250	45	62	0.26

[a] Results from processing against all carcinogenicity alerts in Derek Nexus version 6.1.0

proach against known carcinogens and the pathways leading to their toxicity is good.

However, the specificity of the results against the Vitic data set show that while the approach can identify carcinogens, it may lack the ability to differentiate signals which will propagate to produce a carcinogenic outcome with those that will not in its assignments. This may be partly related to the amount of data used to make each assignment, and it might be predicted that the specificity of the predictions will increase as a function of the amount of evidence used in the WoE. However, this is not supported by the findings in the Vitic data set where, along with the increase in sensitivity with increasing evidence used to reach a conclusion, there is a concurrent decrease in specificity. It is important to note that there is a relatively small number of compounds for which the assignments are based on multiple evidence types compared to those based on a prediction alone, and so it is hard to draw too much from this decrease in specificity at this time. Further investigations will be made to understand these results and how this confidence in predictions with different WoEs might be captured along with how the information might be used to make suggestions on next steps in carrying out a safety assessment.

When comparing the predictions obtained using this method with the results obtained from the prediction of carcinogenicity using the expert rule-based system Derek Nexus, there is an improvement in sensitivity using this approach for both data sets. This is not unexpected given the fact that the AOP approach uses all appropriate Derek predictions in its evidence synthesis. This includes predictions for carcinogenicity but also those from related endpoints that may not yet have been converted into alerts for carcinogenicity and, as a consequence, a broader coverage is achieved.

It should also be noted that the results presented in Table 2 do not represent a true validation of this approach since many of the compounds used in the profiling data sets will also have been used to help develop the network and models. As a consequence, the data sets are more a test of the coverage of the approach for known carcinogens and a way of identifying strengths and limitations of the initial reasoning approach. This may also explain why using AOPs with *in silico* predictions already produces such a high sensitivity for both data sets, given that a lot of data will have been used when training these models.

While overall performance metrics are useful in assessing the strengths and gaps in this approach, it is also useful to see how results are reached and presented for individual compounds in order to assess how the knowledge structure can aid in the interpretation of these results and allow human expert review, interrogation of the conclusion, and determination of the next steps in any safety assessment. Two examples of the results for individual compounds selected from the IARC data set are shown in Figure 6.

In the first example, safrole is predicted overall as a potential carcinogenic hazard based on the data and predictions available and using a conservative approach to resolve the evidence (a conflicted classification is deemed as requiring further investigation to resolve and therefore assigned as positive overall). This conservative approach aligns with the IARC classification of 2B for this compound (IARC, 1987). Looking at the prediction in more detail, it can be seen that the evidence indicates that the AOPs associating this compound with cancer relate to its activation to an electrophilic species and reaction with DNA, as indicated by an *in silico* prediction. This initial hypothesis is supported to some degree by *in vitro* assay data

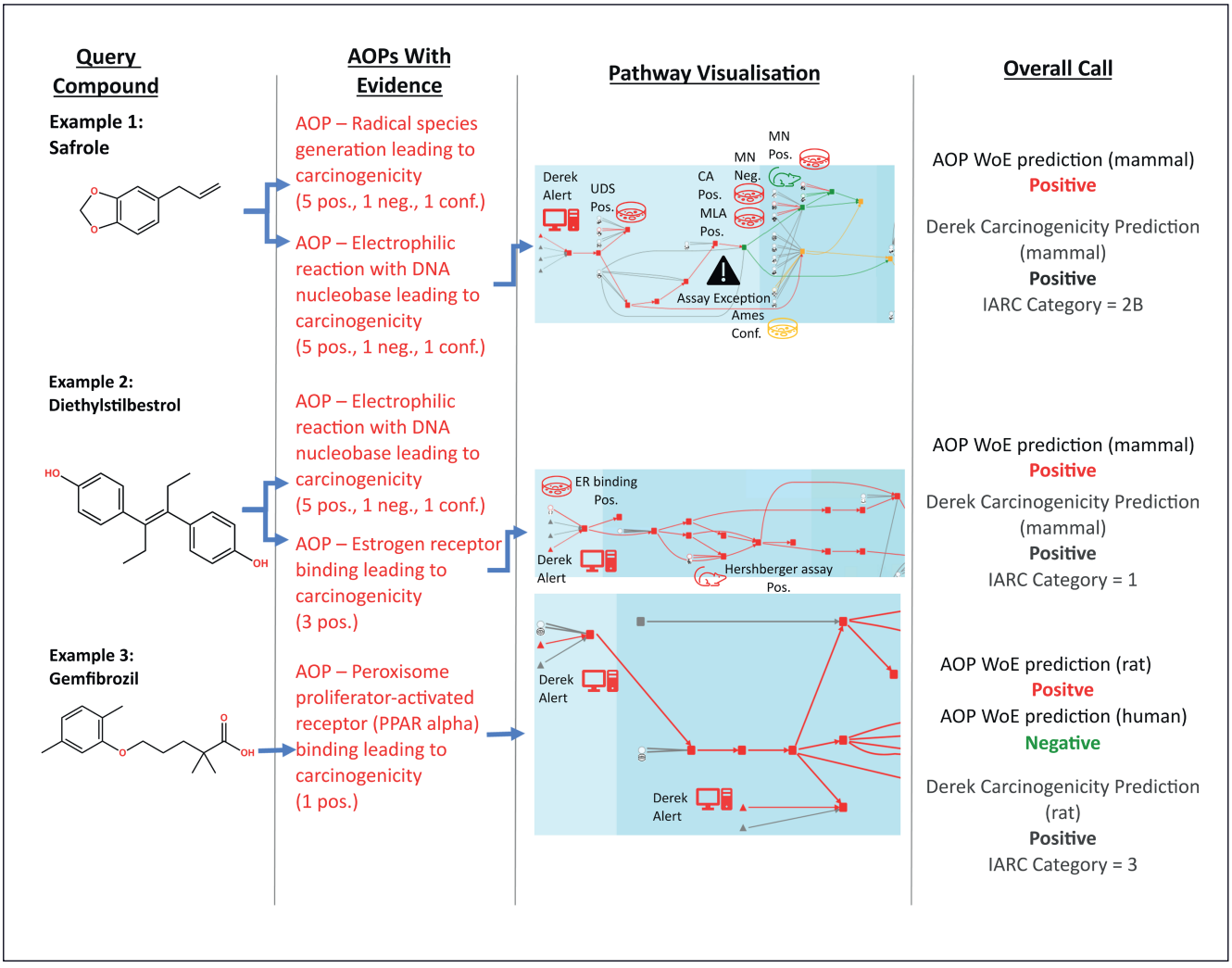

Fig. 6: Example assessments for selected compounds using the AOP reasoning method developed

further along the AOP, which generally shows positive results for this compound. The evidence is connected to KEs leading from the relevant MIE, some directly associated with the pathway leading to the AO of cancer and others to KEs which act as biomarkers to events to this pathway (e.g., structural chromosome damage, micronucleus formation). Most *in vitro* evidence is complimentary, giving positive results and adding to the weight of evidence supporting the pathway. However, there are instances where negative or conflicted findings have been observed for the compound, both *in vitro* and *in vivo*. While these types of conflicts make decision-making more difficult, displaying data on the knowledge framework in this context makes it easy to identify these conflicts in data and interrogate the reason for them more closely in order to reach an expert conclusion that may agree with or overturn the automated one. There is a conflicted finding in the Ames test for this compound associated with the KE of inherited DNA mutation. In this case, the assay for this compound has been flagged as having an exception for this compound class and underpredicting the hazard caused by it. This exception is the result of encoding knowledge relating to inadequacies of the secondary metabolic capabilities in the Ames test to bioactivate the chemical class to their reactive species (Howes et al., 1990; Swanson et al., 1979). With the compound being recognized as part of this class and the conflicted result, the exception flag has been associated with this data point and a potential problem with underprediction has been highlighted. The negative result obtained in the *in vivo* micronucleus test acts to overrule the positive *in vitro* findings associated with assays measuring the same key events. However, since there are other pathways for which the findings are less clear, and the hazard has not been shown to be directly negated (the pathway leading directly to point mutations and on to cancer), the hazard flag remains where the two pathways meet and the underpredicting conflicted result in the Ames test leads to the conflicted outcome propagating down the pathway. All of this information

encoded into the reasoning leads to the positive overall call, as the hazard identified cannot be completely negated. The great advantage of this approach is that it is easy to interrogate where this prediction came from and test the assumptions and evidence leading to both the overall call and the mechanism by which the compound may cause cancer.

The second example in Figure 6 shows the results obtained from the profiler for diethylstilbestrol. A positive overall call is also provided in this instance based on the evidence available. However, in this case there are multiple types of AOPs that the evidence indicates are liable to produce this positive overall result and should be further investigated. One of these relates to the potential for the compound to cause genotoxicity, driven by both a prediction indicating metabolic activation to a reactive species and data supporting this MoA. The second pathway with positive evidence associated is an AOP relating to the potential of this compound to activate the estrogen receptor. Evidence indicating this AOP includes a prediction of the interaction with the estrogen receptor from an *in silico* model along with binding data from an *in vitro* assay. Evidence coming from results in the Hershberger assay, associated with a KE further down the pathway, also support the propagation of the signal down the pathway. In this case, the evidence for this compound is overwhelmingly positive, indicating multiple different pathways leading to toxicity, and supports the IARC classification of category 1 for this compound (IARC, 2012). Since all the AOPs identified as having the potential to be in operation for this compound are thought to have a degree of human relevance, the positive prediction is also thought to be relevant to humans.

In the third example in Figure 6, results from the profiler for gemfibrozil are shown. This compound produces a positive result overall in the profiler for the rat due to its association with the AOP linking peroxisome proliferator-activated receptor alpha (PPARα) activation with cancer in rats. Since this is the only AOP with which a positive signal is associated with this compound and due to the fact that this AOP has been annotated as having a strong WoE for occurring in rats but available evidence indicates the MoA is unlikely to be relevant to humans (Cunningham et al., 2010; Klaunig et al., 2003), the overall WoE assessment indicates a negative prediction in humans. This example highlights how contextualizing evidence on AOPs can allow the species relevance of any findings to be considered easily and transparently to make species-relevant assessments and arguments.

4 Discussion

The carcinogenicity AOP network developed, the evidence added to this network, and the general reasoning method described have already shown great promise in being able to aid carcinogenicity safety assessment. There are areas where this approach provides clear advantages over current paradigms, and there is also great scope for continued improvement using the framework. The advantages of this approach along with areas for continued development are discussed below.

4.1 AOPs allow a greater understanding of MoAs leading to carcinogenicity

Understanding the MoA by which a chemical may cause cancer is an important aspect of any safety assessment, and this knowledge can be used to make better decisions on any next steps to take and how likely a hazard is to translate into a human risk. Contextualizing evidence onto AOPs allows for all the evidence supporting a particular MoA to be viewed and combined into a WoE quickly and easily. Knowledge of how individual pieces of evidence fit together in a particular MoA leading to carcinogenicity can be useful in determining the species relevance of any findings and how this information might be used to argue a specific case in carcinogenicity assessment.

Of course, in order for information on the MoAs to be provided, it is necessary that the network of AOPs leading to cancer is as comprehensive as possible in capturing the knowledge available. The sensitivity obtained against the test sets employed in this work (Tab. 2) indicates that the AOP network developed is already reasonably comprehensive in terms of capturing public knowledge relating to MoAs leading to cancer. However, there is still some room for improvement, and future work will involve filling gaps in pathways identified from test set analysis and delineating new AOPs in this area from public knowledge. There may also be pathway knowledge or additional data in private repositories that should be further integrated in the future. Likewise, in a number of use cases, encoding knowledge of the association of a chemical with its intended target and the association of this target with cancer also needs to be captured. Since the number of potential targets is vast, it may be necessary that these pathways are built *ad hoc* during individual assessments and subsequently stored and built upon. It may also be useful if the searching and structuring of data required to build these pathways as new knowledge becomes available is at least partially automated in the future.

Another key advantage in using AOPs to understand the MoA leading to findings that can be associated with carcinogenicity is that the human relevance of any findings can be assessed and the information used in decision-making. The examples of knowledge relating to human relevance of specific interactions and chains of biological events occurring in rodents are exemplified in the knowledge captured in Figure 1 and the example of reasoning given in the third example in Figure 6.

4.2 AOPs improve the ability to incorporate all available evidence into a safety assessment

Contextualizing evidence on an AOP and using the general reasoning approach outlined in this work allows for all available evidence to be used in a carcinogenicity safety assessment providing it can be associated with a particular KE and this KE can be integrated into the carcinogenicity network. This is particularly important as we move towards IATAs and away from animal models in carcinogenicity assessment. With this approach, the most relevant evidence coming from emerging NAMs can be used appropriately to build a WoE in order to make a decision. These NAMs are developing and improving constantly,

and there is no set list of evidence that will remain in perpetuity. Therefore, it is important that any method of carrying out an assessment can absorb these new developments and improve as a consequence. The framework described in this work allows this flexibility and can be built upon to consider all sources of evidence appropriately.

One area in which the current approach requires more work is in the specificity of the conclusions that are reached. While a good coverage of all potential hazard pathways is important and a conservative approach to assessment is beneficial in many cancer assessment domains, it is also important that the approach taken is not overly sensitive, labelling everything as a potential carcinogen. This work has shown that the network has a good coverage of published mechanisms leading to cancer (as demonstrated by the high sensitivity against the test sets in Tab. 2), whilst the specificity is currently low and needs to be improved.

One way to improve specificity might be to decide a minimum amount or WoE on which a prediction can be made or make the degree of extrapolation clear in the result. However, our initial findings (Tab. 2) indicate that purely increasing the amount of evidence without other measures may not increase specificity on its own. It may be that, as well as measuring the quantity of the evidence used to make a prediction, more consideration of the metadata coming from the evidence and generation of additional assay exception information would help in determining the quality of the evidence too, which would help in determining the weight that should be assigned to it.

Another way of increasing specificity might be to use additional information coming from a read-across approach to supplement any model predictions or data and allow the expert carrying out the review easy access to the information in order that they can assess and reach their own conclusion. Indeed, it is likely that expert review would improve the outcome of any prediction regardless of other methods employed. It is a useful and important aspect of this approach that AOPs and the presentation of the evidence on them allows for an easier and more transparent interpretation of the evidence base. This enrichment through expert review is exemplified with the combination of *in silico* models to assess the mutagenic potential of pharmaceutical impurities in the context of the ICH M7 guideline (Jayasekara et al., 2021).

The implementation of quantitative AOPs (qAOPs) could be another way to improve the specificity of this type of approach. qAOPs allow the description of the quantitative relationship between one KE and another and how the response in one is reflected by a quantitative change in the following event (Spinu et al., 2020). Capturing this knowledge in an AOP, either through experimental data or predictions, may allow thresholds to be set that must be met before one KE propagates a signal to the next, leading to toxicity. Using this information in assessments may allow a more specific prediction, where it is easier to establish when an identified hazard may develop into a risk. While it is not currently obvious how this will be achieved, it will be an important progression of AOP assessment in the future and will aid in the specificity of this approach (Sewell et al., 2018).

There is already the ability within the AOP data structure described in this work to capture the species relevance of the pathways and KEs described. This feature of AOPs is useful in carcinogenicity assessment as it allows human relevance to be easily identified and tested with any hypotheses generated. Further work is required to capture human relevance with the network, and this may be achieved through expert knowledge or analysis of data associated with the pathways. In addition to species relevance, tumor location and type are also important considerations in many carcinogenicity assessments. This information can be captured to some degree within the context terms associated with the KEs on the AOPs. However, more work is required in determining how the relevant location propagates down the AOP, and a conclusion on the location and how this relates to tumor type is determined and presented.

4.3 AOPs improve consistency of interpretation of evidence

As well as improving the ability to capture all different aspects of knowledge relating to carcinogenicity and combine them, AOPs and the reasoning framework presented here also allow this information to be combined and visualized in a transparent and consistent way. The data model and software being developed mean that it is easy to see where and how different evidence associates with the pathways (Fig. 6) and how this evidence combines and propagates down a pathway (Fig. 6). The different levels of complexity to which the pathways can be grouped through KEGs will mean that the user is not overwhelmed by the detail of the information captured while still being able to interrogate it (Fig. 3). Having a defined reasoning approach that is transparent and flexible also means that any conclusion that has been reached can easily be interrogated and the logic leading to the conclusions understood from the individual pieces of evidence upwards (Fig. 5). This is particularly useful if an expert review of the conclusion is to be carried out or a submission reviewed by a regulator. With the results set out in this way, it will also be easier to identify next steps in any carcinogenicity assessment, for example, how to determine which assay to run next or the direction of further investigations into what may have caused an observed finding. This is an area of future development for our work.

Given the broad potential of this approach to carcinogenicity assessment and the possible improvements already identified, investigations are currently underway into how this AOP network in combination with the general reasoning approach described might be used to help make carcinogenicity assessment decisions in more specific use cases. The first of these is the proposed draft addendum to the ICH S1B guidance (ICH S1B (R1), 2021). In this addendum it is suggested that various aspects of the knowledge and findings obtained earlier in the drug development process might be used in a WoE to judge whether the rat 2-year bioassay would add value to the assessment and therefore whether it should be run. However, the addendum deliberately leaves flexibility in how the evidence should be combined in a logical and transparent way, and this is where the AOP network could prove useful (Stalford et al., 2021).

5 Conclusions

AOPs represent a promising method of contextualizing evidence for many different uses in safety assessment, including carcinogenicity. However, to ensure they reach their full potential, work needs to be carried out to capture available knowledge in the form of pathways and to integrate these pathways so they can be used to interact with relevant evidence. This work has gone some way to fulfilling these requirements and has shown how a coherent network of AOPs relevant to carcinogenicity can be built that allows interaction with evidence coming from experimental assays and *in silico* predictions in a meaningful way. Knowledge of species relevance and tissue specificity relating to the pathways was also captured along with any potential limitations in the evidence being associated. Framing knowledge, data, and predictions in this way allowed for the development and application of a flexible and general reasoning approach to reach more relevant conclusions of the carcinogenic potential of a compound based on the evidence available. The data model and the method by which the information is presented within the software developed allow the user to easily understand and interrogate any conclusion that has been reached. The reasoning method is flexible enough that different use cases relating to carcinogenicity assessment can be absorbed. With knowledge organized and digitalized in this way, it will be easier to manipulate it in multiple different ways to answer a wide range of different questions relating to chemical carcinogenicity safety assessment. Selecting the appropriate knowledge domain, along with the evidence to be used and methods of combining this evidence to reach a conclusion, will provide a flexible method of assessment in many different contexts. Future work will involve developing this approach to cancer assessment further by taking into consideration additional factors that can be used in a WoE when carrying out cancer assessments to improve results. Solutions for specific use cases will be targeted, and the approach will be developed from a categorical hazard assignment to a risk assessment tool for carcinogenicity.

References

Amberg, A., Harvey, J. S., Czich, A. et al. (2015). Do carboxylic/sulfonic acid halides really present a mutagenic and carcinogenic risk as impurities in final drug products? *Org Process Res Dev 19*, 1495-1506. doi:10.1021/acs.oprd.5b00106

Ankley, G. T., Bennett, R. S., Erickson, R. J. et al. (2010). Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ Toxicol Chem 29*, 730-741. doi:10.1002/etc.34

Arnesdotter, E., Spinu, N., Firman, J. et al. (2021). Derivation, characterisation and analysis of an adverse outcome pathway network for human hepatotoxicity. *Toxicology 459*, 152856. doi:10.1016/j.tox.2021.152856

Ball, T., Barber, C. G., Cayley, A. et al. (2021). Beyond adverse outcome pathways: Making toxicity predictions from event networks, SAR models, data and knowledge. *Toxicol Res 10*, 102-122. doi:10.1093/toxres/tfaa099

Bayly, A. C., Roberts, R. A. and Dive, C. (1994). Suppression of liver cell apoptosis in vitro by the non-genotoxic hepatocarcinogen and peroxisome proliferator nafenopin. *J Cell Biol 125*, 197-203. doi:10.1083/jcb.125.1.197

Berry, C. (2017). The failure of rodent carcinogenesis as a model for Man. *Toxicol Res 7*, 553-557. doi:10.1039/c7tx00283a

Boobis, A. R., Cohen, S. M., Dellarco, V. L. et al. (2016). Classification schemes for carcinogenicity based on hazard-identification have become outmoded and serve neither science nor society. *Regul Toxicol Pharmacol 82*, 158-166. doi:10.1016/j.yrtph.2016.10.014

Bryce, S. M., Bernacki, D. T., Bemis, J. C. et al. (2016). Genotoxic mode of action predictions from a multiplexed flow cytometric assay and a machine learning approach. *Environ Mol Mutagen 57*, 171-189. doi:10.1002/em.21996

Cohen, S. M. (2004). Human carcinogenic risk evaluation: An alternative approach to the two-year rodent bioassay. *Toxicol Sci 80*, 225-229. doi:10.1093/toxsci/kfh159

Cohen, S. M., Boobis, A. R., Dellarco, V. L. et al. (2019). Chemical carcinogenicity revisited 3: Risk assessment of carcinogenic potential based on the current state of knowledge of carcinogenesis in humans. *Regul Toxicol Pharmacol 103*, 100-105. doi:10.1016/j.yrtph.2019.01.017

Columbano, A., Ledda-Columbano, G. M., Pibiri, M. et al. (2005). Gadd45beta is induced through a CAR-dependent, TNF-independent pathway in murine liver hyperplasia. *Hepatology 42*, 1118-1126. doi:10.1002/hep.20883

Cunningham, M. L., Collins, B. J., Hejtmancik, M. R. et al. (2010). Effects of the PPARα agonist and widely used antihyperlipidemic drug gemfibrozil on hepatic toxicity and lipid metabolism. *PPAR Res 2010*, 681963. doi:10.1155/2010/681963

Doe, J. E., Boobis, A. R., Dellarco, V. et al. (2019). Chemical carcinogenicity revisited 2: Current knowledge of carcinogenesis shows that categorization as a carcinogen or non-carcinogen is not scientifically credible. *Regul Toxicol Pharmacol 103*, 124-129. doi:10.1016/j.yrtph.2019.01.024

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

Guyton, K. Z., Kyle, A. D., Aubrecht, J. et al. (2009). Improving prediction of chemical carcinogenicity by considering multiple mechanisms and applying toxicogenomic approaches. *Mutat Res 681*, 230-240. doi:10.1016/j.mrrev.2008.10.001

Helm, J. S. and Rudel, R. A. (2020). Adverse outcome pathways for ionizing radiation and breast cancer involve direct and indirect DNA damage, oxidative stress, inflammation, genomic instability, and interaction with hormonal regulation of the breast. *Arch Toxicol 94*, 1511-1549. doi:10.1007/s00204-020-02752-z

Hendriks, G., Atallah, M., Morolli, B. et al. (2012). The ToxTracker assay: Novel GFP reporter systems that provide mechanistic insight into the genotoxic properties of chemicals. *Toxicol Sci 125*, 285-298. doi:10.1093/toxsci/kfr281

Heusinkveld, H., Braakhuis, H., Gommans, R. et al. (2020). Towards a mechanism-based approach for the prediction of nongenotoxic carcinogenic potential of agrochemicals. *Crit Rev Toxicol 50*, 725-739. doi:10.1080/10408444.2020.1841732

Hill, III T. and Conolly, R. B. (2019). Development of a novel AOP for Cyp2F2-mediated lung cancer in mice. *Toxicol Sci 172*, 1-10. doi:10.1093/toxsci/kfz185

Howes, A. J., Chan, V. S. and Caldwell, J. (1990). Structure-specificity of the genotoxicity of some naturally occurring alkenylbenzenes determined by the unscheduled DNA synthesis assay in rat hepatocytes. *Food Chem Toxicol 28*, 537-542. doi:10.1016/0278-6915(90)90152-d

Huang, W., Zhang, J., Washington, M. et al. (2005). Xenobiotic stress induces hepatomegaly and liver tumors via the nuclear receptor constitutive androstane receptor. *Mol Endocrinol 19*, 1646-1653. doi:10.1210/me.2004-0520

IARC (1987). International Agency for Research on Cancer (IARC) – Summaries & Evaluations – Safrole, Isosafrole, and Dihydrosafrole. *INCHEM Suppl. 7*, 71. https://inchem.org/documents/iarc/vol10/safrole.html

IARC (2012). Pharmaceuticals – A review of human carcinogens. *IARC Monogr Eval Carcinog Risks Hum 100A*, 175. https://inchem.org/documents/iarc/iarcmono/v100aiarc.pdf

IARC (2019). IARC Monographs on the Identification of Carcinogenic Hazards to Humans. https://monographs.iarc.who.int/wp-content/uploads/2019/07/Preamble-2019.pdf

ICH S1B (1997). ICH Harmonised Tripartite Guideline: Testing for Carcinogenicity of Pharmaceuticals S1B. *International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use*. https://database.ich.org/sites/default/files/S1B%20Guideline.pdf

ICH S1B(R1) (2021). ICH Harmonised Guideline: Addendum to the Guideline on Testing for Carcinogenicity of Pharmaceuticals. *International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use*. https://database.ich.org/sites/default/files/ICH_S1BR1_Step2_DraftGuideline_2021_0510.pdf.

ICH S2(R1) (2011). ICH Harmonised Tripartite Guideline: Guidance on Genotoxicity Testing and Data Interpretation for Pharmaceuticals Intended for Human Use S2(R1). *International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use*. https://database.ich.org/sites/default/files/S2%28R1%29%20Guideline.pdf

Ives, C., Campia, I., Wang, R. L. et al. (2017). Creating a structured AOP knowledgebase via ontology-based annotations. *Appl In Vitro Toxicol 3*, 298-311. doi:10.1089/aivt.2017.0017

Jacobs, M. N., Colacci, A., Louekari, K. et al. (2016). International regulatory needs for development of an IATA for non-genotoxic carcinogenic chemical substances. *ALTEX 33*, 359-392. doi:10.14573/altex.1601201

Jacobs, M. N., Colacci, A., Corvi, R. et al. (2020). Chemical carcinogen safety testing: OECD expert group international consensus on the development of an integrated approach for the testing and assessment of chemical non-genotoxic carcinogens. *Arch Toxicol 94*, 2899-2923. doi:10.1007/s00204-020-02784-5

Jayasekara, P. S., Skanchy, S. K., Kim, M. T. et al. (2021). Assessing the impact of expert knowledge on ICH M7 (Q)SAR predictions. Is expert review still needed? *Regul Toxicol Pharmacol 125*, 105006. doi:10.1016/j.yrtph.2021.105006

Johansson, H. K. L., Damdimopoulou, P., van Duursen, M. B. M. et al. (2020). Putative adverse outcome pathways for female reproductive disorders to improve testing and regulation of chemicals. *Arch Toxicol 94*, 3359-3379. doi:10.1007/s00204-020-02834-y

Klaunig, J. E., Babich, M. A., Baetcke, K. P. et al. (2003). PPARalpha agonist-induced rodent tumors: Modes of action and human relevance. *Crit Rev Toxicol 33*, 655-780. doi:10.1080/713608372

Knapen, D., Angrish, M. M., Fortin, M. C. et al. (2018). Adverse outcome pathway networks I: Development and applications. *Environ Toxicol Chem 37*, 1723-1733. doi:10.1002/etc.4125

Kodama, S. and Negishi, M. (2011). Pregnane X receptor PXR activates the GADD45beta gene, eliciting the p38 MAPK signal and cell migration. *J Biol Chem 286*, 3570-3578. doi:10.1074/jbc.M110.179812

Lichtermann, D., Ekelund, J., Pukkala, E. et al. (2001). Incidence of cancer among persons with schizophrenia and their relatives. *Arch Gen Psychiatry 58*, 573-578. doi:10.1001/archpsyc.58.6.573

Lien, S. C., Wei, S. Y., Chang, S. F. et al. (2013). Activation of PPAR-α induces cell cycle arrest and inhibits transforming growth factor-β1 induction of smooth muscle cell phenotype in 10T1/2 mesenchymal cells. *Cell Signal 25*, 1252-1263. doi:10.1016/j.cellsig.2013.01.021

Lynch, A. M., Eastmond, D., Elhajouji, A. et al. (2019). Targets and mechanisms of chemically induced aneuploidy. Part 1 of the report of the 2017 IWGT workgroup on assessing the risk of aneugens for carcinogenesis and hereditary diseases. *Mutat Res 847*, 403025. doi:10.1016/j.mrgentox.2019.02.006

Nymark, P., Karlsson, H. L., Halappanavar, S. et al. (2021). Adverse outcome pathway development for assessment of lung carcinogenicity by nanoparticles. *Front Toxicol 3*, 653386. doi:10.3389/ftox.2021.653386

OECD (2017). Guidance Document for the Use of Adverse Outcome Pathways in Developing Integrated Approaches to Testing and Assessment (IATA). *OECD Series on Testing and Assessment, No. 260*. OECD Publishing, Paris. doi:10.1787/44bb06c1-en

OECD (2018). Users' Handbook supplement to the Guidance Document for developing and assessing Adverse Outcome Pathways. *OECD Series on Adverse Outcome Pathways, No. 1*. OECD Publishing, Paris. doi:10.1787/5jlv1m9d1g32-en

OECD (2021). Guideline No. 497: Defined Approaches on Skin Sensitisation. *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing, Paris. doi:10.1787/b92879a4-en

Pollesch, N. L., Villeneuve, D. L. and O'Brien, J. M. (2019). Extracting and benchmarking emerging adverse outcome

pathway knowledge. *Toxicol Sci 168*, 349-364. doi:10.1093/toxsci/kfz006

Robbins, D., Cherian, M., Wu, J. et al. (2016). Human pregnane X receptor compromises the function of p53 and promotes malignant transformation. *Cell Death Discov 2*, 16023. doi:10.1038/cddiscovery.2016.23

Sasaki, J. C., Allemang, A., Bryce, S. M. et al. (2020). Application of the adverse outcome pathway framework to genotoxic modes of action. *Environ Mol Mutagen 61*, 114-134. doi:10.1002/em.22339

Sewell, F., Gellatly, N., Beaumont, M. et al. (2018). The future trajectory of adverse outcome pathways: A commentary. *Arch Toxicol 92*, 1657-1661. doi:10.1007/s00204-018-2183-2

Sistare, F. D., Morton, D., Alden, C. et al. (2011). An analysis of pharmaceutical experience with decades of rat carcinogenicity testing: Support for a proposal to modify current regulatory guidelines. *Toxicol Pathol 39*, 716-744. doi:10.1177/0192623311406935

Smith, M. T., Guyton, K. Z., Gibbons, C. F. et al. (2016). Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. *Environ Health Perspect 124*, 713-721. doi:10.1289/ehp.1509912

Spinu, N., Cronin, M. T. D., Enoch, S. J. et al. (2020). Quantitative adverse outcome pathway (qAOP) models for toxicity prediction. *Arch Toxicol 94*, 1497-1510. doi:10.1007/s00204-020-02774-7

Stalford, S. A., Cayley, A. N. and de Oliveira, A. A. F. (2021). Employing an adverse outcome pathway framework for weight-of-evidence assessment with application to the ICH S1B guidance addendum. *Reg Toxicol Pharmacol 127*, 105071. doi:10.1016/j.yrtph.2021.105071

Swanson, A. B., Chambliss, D. D., Blomquist, J. C. et al. (1979). The mutagenicities of safrole, estragole, eugenol, trans-anethole, and some of their known or possible metabolites for Salmonella typhimurium mutants, *Mutat Res 60*, 143-153. doi:10.1016/0027-5107(79)90178-7

Tian, J., Huang, H., Hoffman, B. et al. (2011). Gadd45β is an inducible coactivator of transcription that facilitates rapid liver growth in mice. *J Clin Invest 121*, 4491-4502. doi:10.1172/JCI38760

Wang, P. S., Walker, A. M., Tsuang, M. T. et al. (2002). Dopamine antagonists and the development of breast cancer. *Arch Gen Psychiatry 59*, 1147-1154. doi:10.1001/archpsyc.59.12.1147

Wolf, D. C., Cohen, S. M., Boobis, A. R. et al. (2019) Chemical carcinogenicity revisited 1: A unified theory of carcinogenicity based on contemporary knowledge. *Regul Toxicol Pharmacol 103*, 86-92. doi:10.1016/j.yrtph.2019.01.021

Conflict of interest

The authors declare that they have no conflicts of interest.

Data availability

The data and knowledge presented in this study are currently being developed within software under the direction of a consortium facilitated by Lhasa Limited. As a consequence, the data are not currently publicly available outside of this consortium.