Automated Network Assembly of Mechanistic Literature for Informed Evidence Identification to Support Cancer Risk Assessment

Bernice Scholten,[1,2] Laura Guerrero Simón,[1] Shaji Krishnan,[1] Roel Vermeulen,[2] Anjoeka Pronk,[1] Benjamin M. Gyori,[3] John A. Bachman,[3] Jelle Vlaanderen,[2*] and Rob Stierum[1*]

[1]Research Group Risk Analysis for Products in Development, The Netherlands Organisation for applied scientific research, Utrecht, Netherlands
[2]Institute for Risk Assessment Sciences, Utrecht University, Utrecht, Netherlands
[3]Laboratory of Systems Pharmacology, Harvard Medical School, Boston, Massachusetts, USA

BACKGROUND: Mechanistic data is increasingly used in hazard identification of chemicals. However, the volume of data is large, challenging the efficient identification and clustering of relevant data.

OBJECTIVES: We investigated whether evidence identification for hazard assessment can become more efficient and informed through an automated approach that combines machine reading of publications with network visualization tools.

METHODS: We chose 13 chemicals that were evaluated by the International Agency for Research on Cancer (IARC) *Monographs* program incorporating the key characteristics of carcinogens (KCCs) approach. Using established literature search terms for KCCs, we retrieved and analyzed literature using Integrated Network and Dynamical Reasoning Assembler (INDRA). INDRA combines large-scale literature processing with pathway databases and extracts relationships between biomolecules, bioprocesses, and chemicals into statements (e.g., "benzene activates DNA damage"). These statements were subsequently assembled into networks and compared with the KCC evaluation by the IARC, to evaluate the informativeness of our approach.

RESULTS: We found, in general, larger networks for those chemicals which the IARC has evaluated the evidence to be strong for KCC induction. Larger networks were not directly linked to publication count, given that we retrieved small networks for several chemicals with little support for KCC activation according to the IARC, despite the significant volume of literature for these specific chemicals. In addition, interpreting networks for genotoxicity and DNA repair showed concordance with the IARC KCC evaluation.

DISCUSSION: Our method is an automated approach to condense mechanistic literature into searchable and interpretable networks based on an *a priori* ontology. The approach is no replacement of expert evaluation but, instead, provides an informed structure for experts to quickly identify which statements are made in which papers and how these could connect. We focused on the KCCs because these are supported by well-described search terms. The method needs to be tested in other frameworks as well to demonstrate its generalizability. https://doi.org/10.1289/EHP9112

Introduction

Risk assessment of chemicals is commonly based on toxicological or epidemiological studies. Mechanistic studies can be used to complement animal or epidemiological data to inform mechanisms of toxicity, dose–response assessment, and hazard identification (National Academies of Sciences, Engineering, and Medicine 2017). However, it is generally recognized that summarizing mechanistic data is challenging, in part because of the large diversity of study types and the high volume of available studies (EFSA 2018). At present, there is still no generally accepted procedure to structure, analyze, and interpret mechanistic studies in an efficient way (Guyton et al. 2018; Wikoff et al. 2019). Further, the process of evaluating available mechanistic data, including reading manuscripts and evaluating the associated data, is labor intensive.

There is growing interest in using machine learning and other approaches to reduce the human burden in screening studies for relevance and to facilitate systematic review processes (Howard et al. 2016). For example the "Sciome Workbench for Interactive computer-Facilitated Text-mining" (SWIFT)-review tool has been developed to identify and visualize whether the currently available data for a chemical of interest is rich or poor. The Table Builder and Health Assessment Workplace Collaborative (HAWC) are tools meant to share results of systematic review searches and risk-of-bias assessments and include possibilities for data analyses (Shapiro et al. 2018). Other researchers have applied a bioinformatics approach to structure and analyze mechanistic data. Carvaillo et al. (2019), for example, combined text mining and systems biology by creating a tool, the adverse outcome pathway (AOP)-helpFinder, that enriches AOPs. This tool could assist risk assessors in identifying relevant associations between certain chemicals of interest and AOP components. Guha et al. (2016) combined information on chemical structure with database integration and automated text mining and, as such, prioritized agents for hazard identification.

Here, we propose an approach for identification and prioritization of data and knowledge for use in hazard characterization of chemicals that combines text mining with network visualization tools. We apply our approach within the context of the International Agency for Research on Cancer (IARC) *Monographs* program for the evaluation of carcinogens, which evaluates mechanistic data using a well-defined framework and ontology: the 10 key characteristics of carcinogens (KCCs).

The KCCs have been recently identified in a series of IARC workshops (Smith et al. 2016). The IARC has used mechanistic data to strengthen conclusions on carcinogen classifications since 1991 (IARC 1992) but developed the 10 KCCs to create a more systematic method for the evaluation of mechanistic data to support hazard assessment for carcinogens. The KCCs comprise the properties of known human carcinogens (e.g., having genotoxic or immunosuppressive properties) and data on these characteristics can support the evidence of carcinogenicity (Smith et al. 2016). To retrieve information based on the KCCs from the scientific literature, the IARC *Monographs* staff developed a working list of search terms for the KCCs (Guyton et al. 2018)

Table 1. Ten key characteristics of carcinogens (KCCs) and corresponding search terms (taken from Guyton et al. 2018).

KCC	Description (label)[a]	Query	Search term[b]
1	Is electrophilic or can be metabolically activated (electrophilic)	1	"pharmacokinetics"[MeSH Terms] OR "pharmacokinetics"[Subheading] OR "absorption"[MeSH Terms] OR "distribution"[Title] OR "excretion"
2	Is genotoxic (genotoxic)	2	("Mutation"[Mesh] OR "Cytogenetic Analysis"[Mesh] OR "Mutagens"[Mesh] OR "Oncogenes"[Mesh] OR "Genetic Processes"[Mesh] OR "genomic instability"[MesH] OR chromosom* OR clastogen* OR "genetic toxicology" OR "strand break" OR "unscheduled DNA synthesis" OR "DNA damage" OR "DNA adducts" OR "SCE" OR "chromatid" OR micronucle* OR mutagen* OR "DNA repair" OR "UDS" OR "DNA fragmentation" OR "DNA cleavage")
3	Alters DNA repair or causes genomic instability (DNA repair)		
4	Induces epigenetic alterations (epigenetics)	3	"rna"[MeSH] OR "epigenesis, genetic"[MesH] OR rna OR "rna, messenger"[MeSH] OR "rna" OR "messenger rna" OR mrna OR "histones"[MeSH] OR histones OR epigenetic OR miRNA OR methylation
5	Induces oxidative stress (oxidative stress)	4	"reactive oxygen species"[MeSH Terms] OR "reactive oxygen species"[All Fields] OR "oxygen radicals"[All Fields] OR "oxidative stress"[MeSH Terms] OR "oxidative"[All Fields] OR "oxidative stress"[All Fields] OR "free radicals"[All Fields]
6	Induces chronic inflammation (inflammation)	5	inflamm* or immun* or chemokine or cytokine or leukocyte or white blood cell
7	Is immunosuppressive (immunosuppressive)		
8	Modulates receptor-mediated effects (receptor-mediated)	6	"Hormones, Hormone Substitutes, and Hormone Antagonists"[Mesh] OR "Endocrine Disruptors"[Mesh] OR "Thyroid Hormones"[Mesh] OR "Estrogens"[Mesh] OR "Progesterone"[Mesh] OR "Receptors, Estrogen"[Mesh] OR "Receptors, Androgen"[Mesh] OR "Receptors, Progesterone"[Mesh] OR "Receptors, Thyroid Hormone"[Mesh] OR "Receptors, Aryl Hydrocarbon"[Mesh] OR "Peroxisome Proliferator-Activated Receptors"[Mesh] OR "constitutive androstane receptor"[Supplementary Concept] OR "farnesoid X-activated receptor"[Supplementary Concept] OR "liver X receptor"[Supplementary Concept] OR "Retinoid X Receptors"[Mesh]
9	Causes immortalization (immortalization)	7	"Cell Transformation, Neoplastic"[Mesh] OR "Cell Proliferation"[Mesh] OR apoptosis OR "necrosis"[MeSH] OR "DNA Replication"[Mesh] OR "Cell Cycle"[Mesh] OR brdu OR thymidine OR angiogenesis
10	Alters cell proliferation, cell death or nutrient supply (cell dynamics)		

[a]In parentheses, a label is provided for reference to the specific KCC in this paper.
[b]Truncated search terms are identified by an asterisk.

(Table 1). In 2019, the Preamble to the IARC *Monographs*, which outlines procedures on scientific review and evaluation of carcinogenic hazards, was updated; the KCCs are now used as the basis for the evaluation of mechanistic data (Samet et al. 2020; IARC 2019).

We explored the use of the Integrated Network and Dynamical Reasoning Assembler (INDRA) (Gyori et al. 2017) coupled to the Reach natural language processing system (Valenzuela-Escárcega et al. 2018). INDRA aims to aggregate claims about causal biological and chemical mechanisms extracted by text-mining tools into a mechanistic *in silico* model. The type of *in silico* model can be defined *a priori* based on an evaluation framework. We imported the results of the literature search based on the queries by Guyton et al. (2018) into INDRA. INDRA retrieves so-called causal assertions (i.e., statements in which an entity, such as a small molecule or a protein, interacts with or regulates another entity, such as a protein or biological process) from the literature extracted by Reach and performs a series of assembly steps to transform these relationships into networks. It needs to be mentioned that the term causal in the context of the computational INDRA environment does not automatically imply biological or toxicological causality but, rather, is related to the strength of a computationally inferred value between entities [creating a belief score (BS)].

Importantly, we do not present an approach for an automated hazard characterization. We compared the obtained networks with the IARC's evaluation to assess the informativeness of our approach to synthesize the available evidence into a predefined ontology (i.e., KCCs) and to suggest prioritization of KCCs for certain chemicals. We interpreted correspondence between the expert evaluation and our automated approach for informed evidence identification to indicate usefulness of our approach as a first step in evidence synthesis.

Methods

Figure 1 displays a comparison between our approach (Figure 1A), and the approach by the IARC (Figure 1B), together with a potential application of our approach to aid in identifying and prioritizing relevant information for full-text review of included studies. We chose 13 chemicals that have been mechanistically evaluated in IARC *Monographs* 112–125 and were classified in different IARC carcinogen categories: benzene (1), pentachlorophenol (1), dichlorodiphenyltrichloroethane (DDT; 2A), hydrazine (2A), diazinon (2A), glyphosate (2A), malathion (2A), melamine (2B), parathion (2B), pyridine (2B), allyl chloride (3), β-picoline (3), and coffee (3).

Compounds were selected if they fulfilled the criteria of being evaluated by the IARC, that is, from Monograph 112 onward, based on the potential for induction of the KCCs (see "IARC evaluations for evidence of KC activation" in the Supplemental Material), hereafter referred to as IARC evidence. After evaluating the assembled data, the IARC classified the evidence on the basis of collective expert judgment as strong, moderate, weak, or no evidence. These classifications are based on various criteria, as outlined in the IARC's Instructions for Authors (IARC 2017).

We started our approach with a literature search using the PubMed database, based on the predefined query search terms for the KCCs (Table 1; Guyton et al. 2018). Per query, the search terms were combined with the chemical name(s), as referred to by the IARC, contained within the article title and a date limitation from 1 January 1900 to 1 March 2020 in the following format: "*chemical name*[Title] AND (*Guyton search terms*) AND (1900/01/01[PDat]: 2020/03/01[PDat])." Note that three chemicals were evaluated later (allyl chloride, pentachlorophenol, and β-picoline), hence the search term for these chemicals was extended to October 2021. Each of these searches returned a list
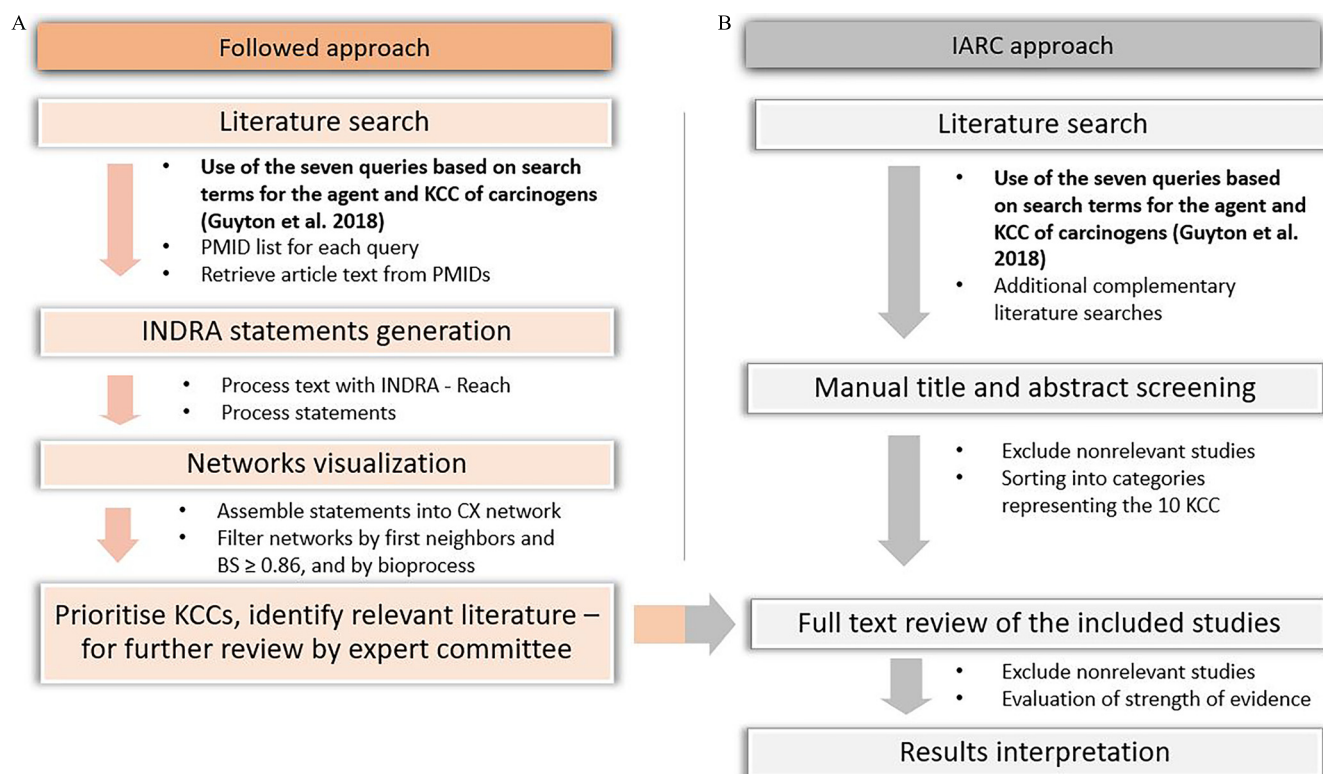
Figure 1. Two different approaches to investigate and summarize mechanistic data: (A) the followed approach and (B) the IARC approach. Note: BS, belief score; CX, a network exchange format; IARC, International Agency for Research on Cancer; INDRA, Integrated Network and Dynamical Reasoning Assembler; KCCs, key characteristics of carcinogens; PMID, PubMed identifier.

of PubMed identifiers [IDs (PMIDs)], each corresponding to an article. We used INDRA's literature retrieval module to obtain the full-text content or the abstract corresponding to each PMID returned by these searches. When available, the full text was retrieved from PubMed Central or Elsevier [an application programming interface (API) key was used to get access to this content via the Elsevier Text and Data Mining API]. For PMIDs, for which the full-text content was not available, the abstract was retrieved.

The retrieved text content (for each of the chemicals: seven lists of article texts based on seven searches, spanning the 10 different KCCs) were then processed with Reach, an open-source natural language processing system for the biomedical literature that is able to read and extract mechanistic descriptions of biological processes from text (Valenzuela-Escárcega et al. 2018). Reach is a type of event extraction system for biology that can detect and normalize information about putative interactions among biological entities and processes (Ananiadou et al. 2010). The system can recognize agents (e.g., proteins, bioprocesses, chemicals), link these to corresponding identifiers in knowledge bases [including UniProt, InterPro, Human Metabolome Database (HMDB), PubChem, and Gene Ontology (GO)], and extract events or interactions (e.g., multiple types of positive or negative regulation). To run Reach, we used the *reach* module of INDRA (specifically, the *process_text* and *process_nxml_str* functions), which provides a Python interface to running the Reach system and processing its extractions into INDRA Statements (see Gyori 2017). INDRA Statements represent a hypothetical, potentially causal influence relation between two agents (e.g., a chemical and a bioprocess). For example, for the sentence "Hydroquinone induces extensive apoptosis in the cells", the generated INDRA Statement reads "Activation(hydroquinone(PUBCHEM:785), apoptotic process (GO:0006915))," which represents that hydroquinone (recognized with the database identifier PUBCHEM:785) activates apoptosis (recognized with the database identifier GO:0006915). (The assignment of database identifiers to entity texts is known as named entity normalization or simply as grounding.)

We thus obtained a list of "raw" (i.e., unprocessed) INDRA Statement objects gathered from Reach after extracting relations from the content of each article's text retrieved in the previous step (Figure 2). Every raw statement contains a set of attributes that includes all information necessary to identify the given putative mechanism and its participants being represented. Each statement also has an evidence attribute that contains additional provenance information and annotations, for instance, text content references (i.e., PMIDs), and the specific sentence from which the statement was extracted, as well as, for instance, whether the sentence was recognized as a hypothetical statement. The evidence here and below is referred to as technical evidence emerging from text mining, which is not automatically and directly equal to biological/toxicological evidence in the context of hazard assessment.

After obtaining all raw statements from the retrieved text of each query, we applied several knowledge assembly steps using the *assemble_corpus* module of INDRA, with the aim of filtering, improving, deduplicating, and calculating BSs for the statements before assembling them into a network.

This assembly process consists of the following steps:
1. Filtering out hypothetical statements with the function *filter_no_hypothesis()*: The Reach system labels a statement as hypothesis when the evidence text for that statement contains one of the default words likely to belong to a hypothesis (e.g., test, consider, predict, speculate, suggest, theorize). This step removes all statements that have been labeled as hypotheses.
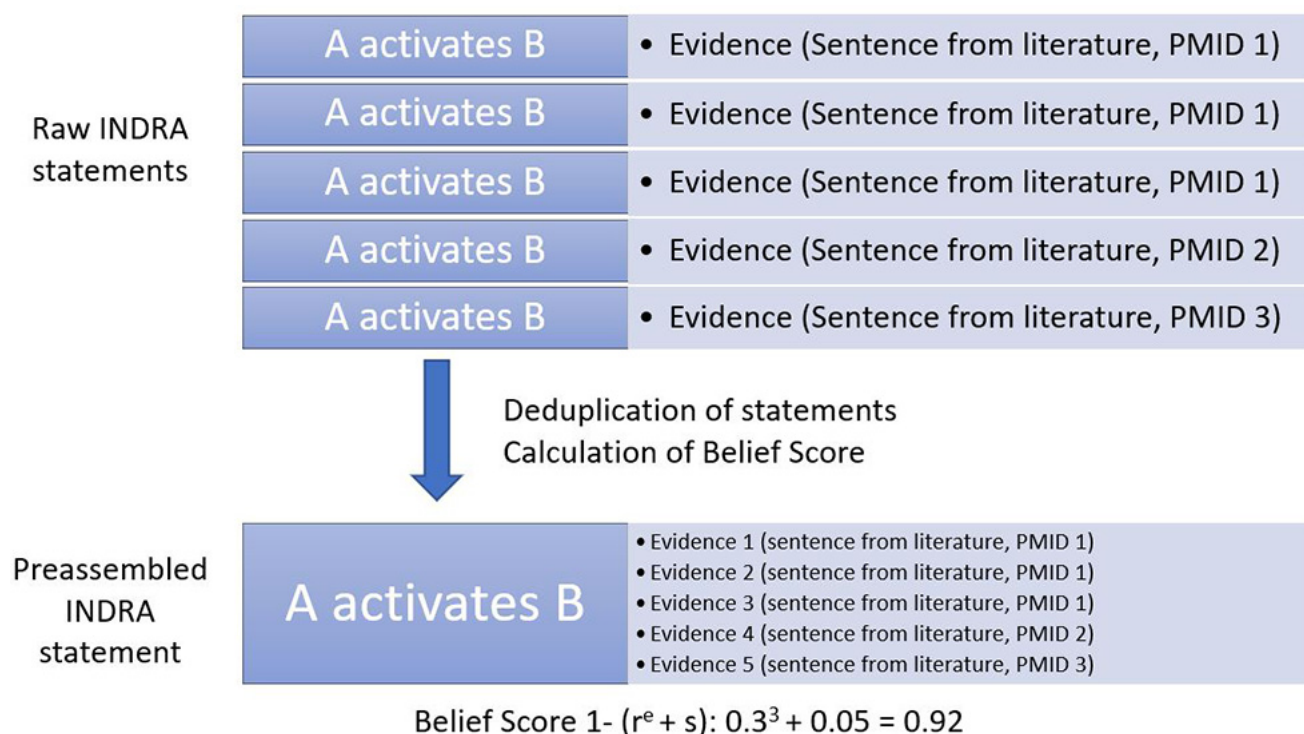
Figure 2. Integrated Network and Dynamical Reasoning Assembler (INDRA) preassembly of statements and calculation of a belief score (BS). The default preassembly function would count in this example five evidences (e), thus the BS would be $1 - (r^e + s) = +0.05 = 0.95$. The function that calculates BS starts with empirical estimates of the prior random (r) and systematic error (s) rates of reading systems. In the present paper Reach was used, and its default built-in values are $r = 0.3$ and $s = 0.05$. With the correction we applied (referred to as deduplication of statements), evidences 1, 2, and 3 are counted as one because they were retrieved from the same paper, hence $e = 3$, resulting in a BS of 0.92. Note: AIM2, absent in melanoma 2; Annexin-V-FLUOS, Annexin-V-fluorescence; Casp1, Caspase 1; DSB, double strand breaks; H2AX, phosphorylated H2AX; r, empirical estimate of the prior random rate; and s, empirical estimate of the systematic error rates; TET, ten-eleven-translocation.

2. Entity renormalization using the function *map_grounding()*: Entities from reading systems, such as Reach, are often incorrectly normalized (i.e., an incorrect database ID is assigned to them). INDRA integrates both expert-curated maps to improve entity normalization and machine-learned models [calling on the Python package Acromine-based Disambiguation of Entities From Text context (Adeft) (Steppi et al. 2020)] to choose between competing senses of ambiguous acronyms (e.g., "IR" can refer to the insulin receptor but also to ionizing radiation). INDRA also standardizes IDs [e.g., when available, it provides equivalent IDs for PubChem compounds in Chemical Entities of Biological Interest (ChEBI), Chemical Abstracts Service (CAS), ChEMBL and other databases]] and the names of agents to their standard names [e.g., HUGO Gene Nomenclature Committee (HGNC) gene symbols, GO labels].

3. Filtering out agents without associated database identifiers with the function *filter_grounded_only()*.

4. Running preassembly with the function *run_preassembly ()*: In the last step of assembly, statements are deduplicated (equivalent statements are merged into one statement) and the associated evidence is gathered in an evidence list (Figure 2). Subsequently, the BS are calculated by INDRA. For each INDRA Statement, the BS is a numerical value between 0 and 1, calculated as a function of the Statement's supporting evidence. The function that calculates BS starts with empirical estimates of the prior random (r) and systematic error (s) rates of reading systems. In the present paper Reach was used, and its default built-in values are

$r = 0.3$ and $s = 0.05$. Coming from a single source, the error probability and BS are as follows:

$$error\ probability = r^e + s,$$
$$Belief\ Score = 1 - error\ probability,$$

with e being the number of pieces of evidence for that statement. Thus, assuming an assembled statement has four pieces of evidence, the BS would be $1 - (0.3^4 + 0.05) = 0.94$ (Figure 2). Hence, the BS is based on the amount of evidence, that is, the more evidence supporting the statement, the higher the BS (Figure 2). These can be supportive, but do not directly refer to a "belief" by, for example, toxicological experts in the cumulative scientific community in the context of KCC hazard assessment.

To avoid counting repeated sentences from the same paper as distinct appearances of the same assertion, we counted sentences coming from the same paper as constituting only a single claim for the purpose of BS calculation. However, if a single paper provided evidence for different KCCs, all these evidence were taken into account. For each query, the number of PMIDs for which INDRA Statements were obtained, was compared with the total number of PMIDs retrieved (expressed as a percentage within parentheses; Table 2).

All programming steps were executed in the environment Spyder (version 3.3.6) and Python (version 3.7.4). The Python script can be found at https://github.com/bernice493/INDRA_hazard. All information related to INDRA was retrieved from https://indra.readthedocs.io/en/latest/.

Once all the steps of this process were finished, the statements were assembled into a network using INDRA's *assemblers.cx*

Table 2. Processed articles for each query.

KCC	Benzene (1)	Penta-chlorophenol (1)	DDT (2A)	Diazinon (2A)	Hydrazine (2A)	Glyphosate (2A)	Malathion (2A)	Melamine (2B)	Parathion (2B)	Pyridine (2B)	Coffee (3)	β-Picoline (3)	Allyl chloride (3)
KCC 1 (electrophilic)	501	139	281	60	69	68	99	71	168	370	173	8	3
	(30%)	(4%)	(35%)	(35%)	(26%)	(50%)	(30%)	(41%)	(33%)	(38%)	(31%)	(13%)	(0%)
	(14%)[a]	(7%)	(9%)	(7%)	(9%)	(32%)	(15%)	(20%)	(21%)	(25%)	(21%)	(0%)	(0%)
KCC 2 and 3 (genotoxic and DNA repair)	740	107	130	39	126	183	84	39	51	619	191	0	7
	(42%)	(7%)	(39%)	(38%)	(47%)	(46%)	(44%)	(46%)	(27%)	(35%)	(52%)	(0%)	(14%)
	(28%)	(10%)	(22%)	(38%)	(25%)	(37%)	(38%)	(36%)	(22%)	(34%)	(21%)	(0%)	(0%)
KCC 4 (epigenetics)	255	61	108	29	48	99	41	19	36	173	149	6	1
	(41%)	(8%)	(58%)	(55%)	(19%)	(58%)	(59%)	(37%)	(33%)	(34%)	(46%)	(17%)	(0%)
	(37%)	(16%)	(47%)	(59%)	(17%)	(55%)	(54%)	(53%)	(36%)	(34%)	(32%)	(0%)	(0%)
KCC 5 (oxidative stress)	332	135	67	99	84	147	121	38	47	350	235	0	4
	(42%)	(4%)	(60%)	(39%)	(54%)	(61%)	(38%)	(39%)	(43%)	(30%)	(47%)	(0%)	(0%)
	(37%)	(10%)	(49%)	(56%)	(39%)	(63%)	(71%)	(71%)	(49%)	(33%)	(32%)	(0%)	(0%)
KCC 6 and 7 (inflammation and immunosuppressive)	577	38	96	46	46	52	58	53	23	247	248	5	3
	(35%)	(11%)	(36%)	(46%)	(33%)	(62%)	(62%)	(37%)	(22%)	(38%)	(40%)	(20%)	(0%)
	(26%)	(16%)	(34%)	(48%)	(30%)	(38%)	(57%)	(36%)	(26%)	(30%)	(39%)	(0%)	(0%)
KCC 8 (receptor mediated)	74	35	215	23	35	46	25	16	33	206	172	0	0
	(33%)	(3%)	(46%)	(48%)	(20%)	(58%)	(36%)	(50%)	(30%)	(31%)	(31%)	(0%)	(0%)
	(35%)	(20%)	(40%)	(61%)	(14%)	(33%)	(52%)	(44%)	(58%)	(31%)	(26%)	(0%)	(0%)
KCC 9 and 10 (immortalization and cell dynamics)	240	42	65	28	54	60	43	48	19	338	169	2	2
	(42%)	(19%)	(55%)	(54%)	(46%)	(53%)	(37%)	(33%)	(26%)	(57%)	(34%)	(50%)	(0%)
	(49%)	(12%)	(48%)	(54%)	(24%)	(47%)	(58%)	(56%)	(37%)	(40%)	(25%)	(0%)	(0%)
TOTAL unique PMIDs associated with chemical name appearing in Title and KCC queries[b]	1,933	427	788	233	368	506	360	215	320	1,820	1,069	18	16

Note: DDT, dichlorodiphenyltrichloroethane; INDRA, Integrated Network and Dynamical Reasoning Assembler; KCCs, key characteristics of carcinogens; PMID, PubMed identifier.
[a]First line within each row: total number of processed articles; second line, in parenthesis "()": the percentage (%) of processed articles; third line: the percentage (%) of PMIDs for which INDRA Statements were retrieved.
[b]The TOTAL number shows the unique papers emerging from the KCC-specific queries (not counting the repeated ones in more than one query).

module. These networks were generated in CX format and visualized in Cytoscape (version 3.7.2): a bio-informatics software platform for visualizing molecular interaction networks (https://cytoscape.org/) and publicly available via National Data Exchange (Table S1). The resulting networks consist of nodes (rectangles), which represent biological entities, and edges (arrows between the nodes), which represent proposed biological or chemical interactions/mechanisms between these entities. Nodes are colored on the basis of the type of entity they represent: bioprocess (orange), chemicals (green), proteins (light blue), protein family (dark blue), and others (being nodes not classified into one of the before mentioned entities, gray). The edges can indicate different events, that is, activation, inhibition, complex formation, negative amount regulation, positive amount regulation, and posttranslational modification, as implied by the underlying INDRA Statements.

To reduce network complexity for further visual and statistical analysis (addressing sizes and support of the different networks), the following network filtering steps were taken: *a*) the chemical of interest and its first neighbors (i.e., directly adjacent nodes) were selected and retained, *b*) only the nodes connected by edges with a BS ≥ 0.86 (two or more pieces of evidence) were retained (note that this is an arbitrary cut-off), and *c*) the nodes or group of nodes not connected to the main network (containing the chemical of interest as central node) were removed (see "Filtering networks" in the Supplemental Material). We also filtered the networks on the basis of the classes: bioprocesses and other processes, for KCC 2 (genotoxic) and KCC 3 (DNA repair), for example (see "Filtering networks" in the Supplemental Material). This filtering helps to focus the attention on potentially relevant biological processes that are possibly directly influenced by the chemical in question.

After creating and filtering the networks, additional quantitative descriptive network information was collected, including the number of nodes and edges. This information represents numerical values describing the network size. In addition, the sum of the BSs (sBS) obtained from all statements within each of the networks (seven networks per chemical, 10 chemicals) was calculated. This provides information on the overall abundance of claims supporting the edges contained within the network. The Wilcoxon rank test was used to compare the sBS metric with the IARC *Monographs* working group (i.e., IARC evidence) classifications.

Aside from serving as input to the network analyses, all original INDRA Statements are also collated in a list. After the assembly step, the statements are saved in JavaScript Object Notation (json) format. These files can be visualized with a json viewer. We used Python json2table software because such information can be retrieved on, for example, the PubMed ID and the original sentence of the paper on which the evidence is based.

Results

The number of analyzed articles for each query, the percentage of articles (i.e., PMIDs) for which INDRA Statements were produced in relation to the retrieved articles, and the percentages of PMIDs with open access, are displayed in Table 2. INDRA Statements were extracted from $\sim 30\%$ of all papers retrieved based on the KCC input, indicating that a reasonable amount of input literature contained information suited to automated processing.

From the INDRA Statements, 91 networks [i.e., seven networks (KCC 1, 2/3, 4, 5, 6/7, 8, and 9/10) for 13 compounds] were created (Table 3). Table 3 provides a comparison between the KCC network size (i.e., the number of nodes and edges) and network support (i.e., sBS of the edges), and the actual evaluation for KCC activation by the IARC (see "IARC evaluations for evidence of KC activation" in the Supplemental Material).

A large network represents that there is (potentially) a richer body of mechanistic literature for that chemical discussed in the context of that KCC. In general, higher sBS were observed for those compounds for which the IARC has proposed strong evidence for induction of KCCs (Figure 3). This is also corroborated by Wilcoxon statistical analysis. Across all chemicals and KCCs, INDRA-derived sBS tended to be significantly higher for KCCs for which the IARC has classified the KCC evaluation as strong or moderate evidence than for those for which the IARC evidence was rated "weak" or "no" [median (interquartile range) = 3.5 (13.7) vs. 0.9 (2.2); $p = 0.0003$], but there was considerable overlap.

From Table 3, we did notice for a number of KCCs the evidence was strong according to the IARC but that the networks and the sBS were small. In those cases, the number of PMIDs (Table 2) was also low. Conversely it does not appear that a larger number of PMIDs, resulting from the KCCs and chemical-specific queries, always results in larger networks, that is, although benzene has the highest number of PMIDs (1,933) and the largest networks, the networks for pyridine and coffee are considerably smaller even though these compounds have the second (1,820) and third (1,069) highest number of PMIDs, respectively. Most compounds show a network for KCC 4 (epigenetic alterations) although the IARC concluded for all compounds but coffee that there was no sufficient evidence to evaluate the induction of this KCC (Table 3).

By using the filtering option on bioprocesses and other processes, we did not limit ourselves to the first neighbors of the compound of interest, hence allowing us to investigate relations between events, also further away from the compound of interest (see "Filtering networks" in the Supplemental Material; Figure S1). An example is given for benzene (Figure 4), where benzene activates DNA damage, whereas DNA damage, in turn, can activate cell death. Further necrotic cell death is associated with activation of an inflammatory response. This relation (cell death activates an inflammatory response) is also described in the AOP wiki databases (ID 1776). In addition, the process of how the disruption of the cell cycle can lead to apoptotic processes (Figure 4) is described in the AOP wiki databases (ID 1712).

Figures S2–S13 show the outcome of "Bioprocess and 'other' filtering" for each of the chemicals for KCC 2 (genotoxicity) and 3 (DNA repair) (originating from Query 2). We see that for all those compounds (benzene, DDT, hydrazine, diazinon, glyphosate, malathion, parathion, pentachlorophenol), for which the IARC evaluated the evidence to be moderate or strong for the activation of this specific KCC, terms (both from "Bioprocesses" and "other") related to genotoxicity and DNA repair did appear in the networks. The terms include, for example, DNA damage and (inhibition of) DNA repair. Conversely, considering the chemicals for which the IARC evaluated the evidence to be weak or absent (melamine, pyridine, allyl chloride, β-picoline, and coffee), these two terms were not observed in the networks for pyridine and coffee (nor for allyl chloride and β-Picoline, for which no network could be created), but only in the network for melamine.

Discussion

In this work, we investigated whether evidence identification for chemical hazard assessment could be supported using an automated, computational approach. As an example, we explored the use of this approach for identifying KCCs as used in the evaluation of mechanistic evidence in the IARC *Monograph* program. Using text mining and network analysis approaches (i.e., INDRA), we found concordance between computationally inferred networks strength (high BS) and the IARC KCC evaluations, especially for those compounds for which the IARC has evaluated the evidence to be strong for KCC induction. As such, our example application

Table 3. Network statistics for 13 chemicals in relation to IARC classifications.

Chemical	KCC 1 (electrophilic)	KCC 2 and 3 (genotoxic and DNA repair)	KCC 4 (epigenetics)	KCC 5 (oxidative stress)	KCC 6 and 7 (inflammation and immunosuppressive)	KCC 8 (receptor-mediated)	KCC 9 and 10 (immortalization and cell dynamics)
Benzene (1), n = 1,933	***[a]	***	?[a]	***	***	***	***
	16/19 (16.73)[b]	43/70 (62.25)	18/21 (18.38)	17/22 (19.75)	21/23 (20.31)	5/4 (3.50)	22/32 (28.83)
Penta-chloro-phenol (1), n = 427	***	?	***	***	**[a]	***	**
	2/1 (0.86)	3/2 (1.51)	5/4 (2.81)	5/4 (3.44)	0	3/3 (1.95)	2/2 (1.51)
DDT (2a), n = 788	?	**	?	***	*[a]	*	*
	4/3 (2.64)	8/7 (6.17)	13/12 (10.52)	11/12 (10.68)	5/6 (5.43)	16/18 (15.63)	7/7 (6.35)
Diazinon (2a), n = 233	?	***	?	***	*[a]	*	*
	0	3/2 (1.78)	2/1 (0.86)	12/13 (11.54)	3/2 (1.78)	2/1 (0.86)	3/2 (1.78)
Glyphosate (2a), n = 506	*	***	?	***	*	*	***
	3/2 (1.72)	10/12 (10.68)	16/17 (15.14)	25/29 (26.01)	4/3 (2.66)	3/2 (1.78)	7/7 (6.11)
Hydrazine (2a), n = 368	*	***	?	***	?	*	***
	0	2/1 (0.92)	2/1 (0.86)	4/3 (2.64)	2/1 (0.86)	0	0
Malathion (2a), n = 360	*	***	?	***	***	***	***
	2/1 (0.86)	8/9 (7.91)	7/6 (5.22)	27/48 (43.05)	16/16 (14.00)	3/3 (2.64)	15/16 (14.11)
Melamine (2b), n = 215	?	*	?	?	***	?	?
	2/1 (0.86)	4/3 (2.58)	2/1 (0.92)	12/13 (11.70)	3/2 (1.78)	2/1 (0.86)	9/9 (8.04)
Parathion (2b), n = 320	*	**	?	*	*	*	***
	11/12 (10.38)	0	0	3/2 (1.78)	0	6/8 (6.88)	0
Pyridine (2b), n = 1,820	?	*	?	*	**	?	?
	0	3/2 (1.72)	5/5 (4.45)	0	0	3/2 (1.78)	2/1 (0.92)
Coffee (3), n = 1,069	?	*	*	*	*	*	*
	2/1 (0.86)	0	2/1 (0.86)	7/8 (7.04)	12/13 (20.36)	3/6 (5.45)	0
Allyl chloride (3), n = 16	***	?	?	?	?	?	*
	0	0	0	0	0	0	0
β-Picoline (3), n = 18	?	*	?	?	?	?	?
	0	0	0	0	0	0	0

Note: DDT, dichlorodiphenyltrichloroethane; IARC, International Agency for Research on Cancer; KCCs, key characteristics of carcinogens.
[a]The symbols refer to the IARC evaluation: ***, strong evidence that KCC is induced; **, moderate evidence for induction of KCC; *, no or weak evidence of the induction of KCC; ?, no adequate data for an evaluation to be made. It regularly occurred that the IARC evaluations differed for the various KCCs, contained within one literature query. For example, for KCC 2 the evidence could have been weak, whereas for KCC3 the evidence was strong. Because the two KCCs are combined, we chose to always use the stronger evidence (in this example we marked the box "strong"). To view the full networks, see Table S1, where URLs for each chemical network are provided.
[b]The numbers in each cell represent number of nodes/edges and sum of belief score (sBS) of the edges are within parentheses.

suggests that compounds with larger networks and higher sBS scores, could be prioritized for hazard identification, making the process of evidence identification for hazard assessment more efficient and transparent.

The output of our approach generates an inventory of available studies, as well as a categorization of data in the form of networks. These generated networks can further be filtered to retrieve information on mechanisms of action by filtering only on bioprocesses (Figure 4; Figures S2–12). This type of visualization can be used as tool to assist in the interpretation of the literature for mechanistic evaluation of compounds within the KCC framework (i.e., informed evidence identification).

Recently Barupal et al. (2021) published a study on prioritizing cancer hazard assessments for IARC *Monographs* using an integrated approach of database fusion and text mining. The authors also used the KCCs as input but, unlike the investigation we conducted, Barupal et al. (2021) mainly looked at publication count, as well as coverage across 34 different databases relevant to cancer, for an agent. Our approach is different in that we are not only identifying possibly relevant literature by the sheer counting of numbers of publications per chemical (Table 2), but that our approach also uses a systems biology–inspired text-mining environment (i.e., INDRA) to extract data from the individual articles and compile these data into potentially meaningful biological networks, describing the possible relations between biomolecules and chemicals, bioprocesses in the context of KCCs (i.e., informed evidence identification). Thus, our work expands beyond the evaluation of publication density or coverage of toxicological content in databases. Importantly, we observed that the number of publications derived from the KCC-specific literature queries (which is driven by general scientific interest in

the chemical) proved not to be an accurate indication for potential KCC activation, at least as inferred here from automated network assembly.

Although promising, using our automated computational approach has several limitations that should be kept in mind when interpreting the results. Stringent filtering on BS, for example, retaining only results with a BS ≥0.86, can exclude relevant results reported in a single study because only one single study could point out a relevant result that now might be discarded. For example, if we consider an unfiltered network for KCC 9, 10 for parathion (KCC 9, 10–Query 7; Figure S13), we observe connections between parathion and apoptotic process, as well as between parathion and cell population proliferation. Both statements have a BS of 0.65, indicating that single studies contribute to these statements. Both processes are linked to KCC 9 and 10 and, according to the IARC, parathion indeed induces KCC 9 and 10; however, this observation would have gone unnoticed upon more stringent filtering. So, for smaller networks it might thus be worthwhile to also investigate the larger, unfiltered networks. Conversely, our network analysis does not distinguish between positive and negative regulation when filtering by BS, so it can occur that a network is large but contains processes that are actually favorable, for example, inhibition of DNA damage. The potential directionality can be further investigated by displaying inhibition vs. activation statements (an example is given for benzene in Figure S14). Last, we applied a filtering step for hypothetical statements by excluding statements that contain certain signaling words such as "suggest." However, using the word "suggest" is sometimes preferred, particularly in human studies, to avoid the use of causal language. Hence, filtering statements with reference to "suggest" can potentially exclude data from articles that use the wording "suggest"
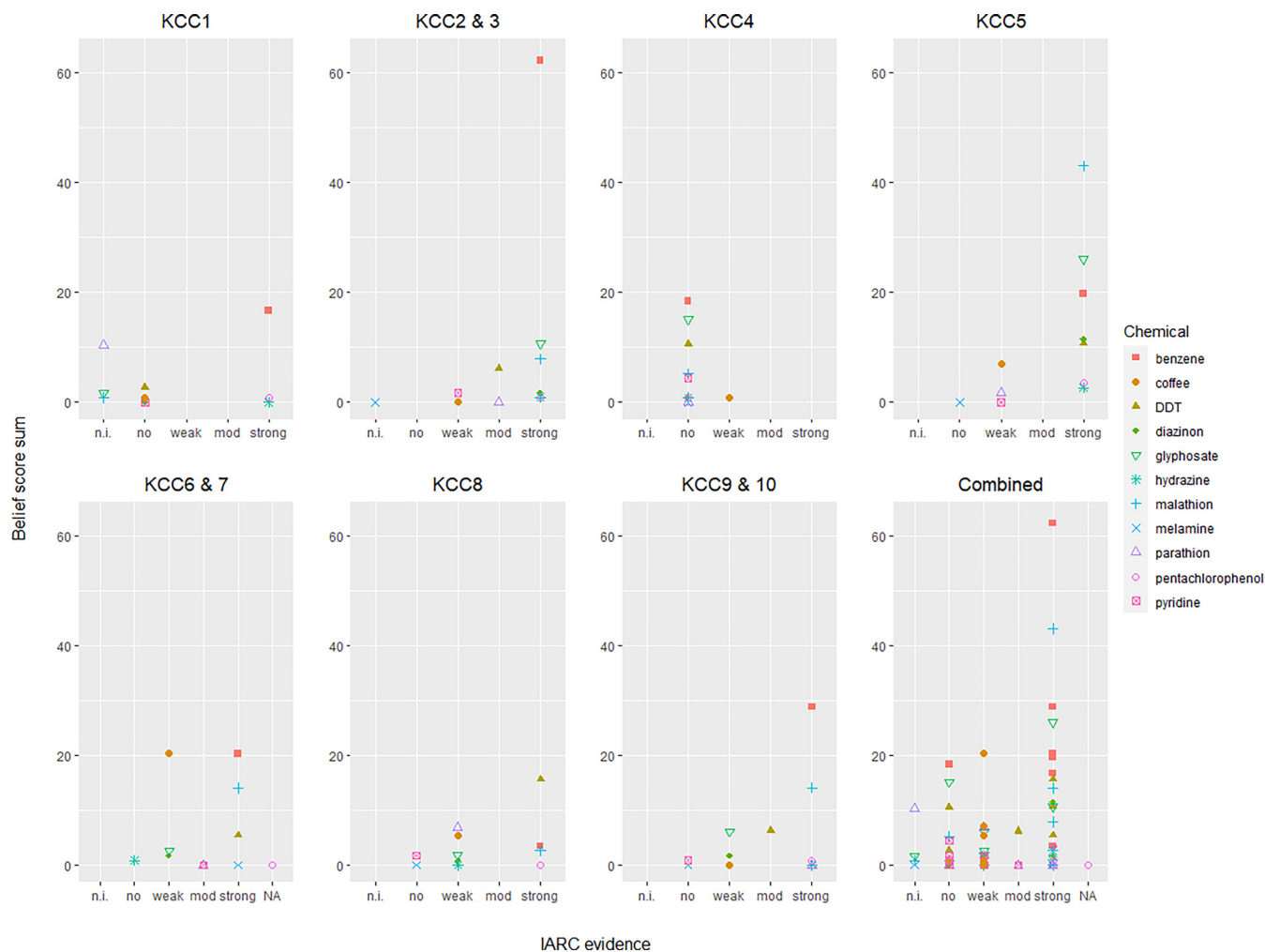
Figure 3. Relation of IARC evidence for key characteristics of carcinogens (KCCs) result vs. sum of the belief scores (sBS) from all filtered edges of a network. Note: IARC, International Agency for Research on Cancer; DDT, dichlorodiphenyltrichloroethane; KCC1, electrophilic; KCC2, genotoxic; KCC3, DNA repair; KCC4, epigenetics; KCC5, oxidative stress; KCC6, inflammation; KCC7, immunosuppressive; KCC8, receptor-mediated; KCC9, immortalization; KCC10, cell dynamics; mod, moderate; NA, not applicable; n.i., not sufficient data available for evaluation.

avoiding the use of causal language, and bias toward articles that inappropriately use causal language.

In other cases, the network showed potentially relevant findings, however not specifically for the KCC for which the network was originally created. An example is parathion, Query 1, which, according to the generated network, can activate cell death, modify testosterone, or inhibit acetylcholinesterase (Figure S15). Given that KCC 1 is on electrophilicity, the findings from Figure S15 would, for example, fit better under KCC 10, which refers to cell death.

Furthermore, we have not evaluated the selected studies' informativeness or study quality after the filtering steps. Relevant questions, such as whether the observed mechanisms can also operate in humans, *in vitro* vs. *in vivo* models, the quality of the studies, biological significance of mechanistic end points, whether evidence is consistent within and among KCCs, for example, were not considered yet. Of course, this can be adopted in the process, that is, to modify the initial PubMed query (e.g., select only human studies), but this requires experts to stratify or limit the evidence base to *a priori* domains or quality assessments.

The composition of the literature query as input, in our case, the search terms by Guyton et al. (2018), is quite influential when retrieving the PMIDs. This was most notable for KCC 4 (induces epigenetic alterations). For many compounds (all but coffee) we see that the IARC states that for this specific KCC there is not sufficient

data available for an evaluation. However, we regularly observe large networks for KCC 4 (Table 3; e.g., benzene, DDT). We discovered that this may be due to the description of Guyton's queries for Query 3/KCC 4: the query includes the terms "rna" or "rna, messenger" [because noncoding RNAs are recognized epigenetic alterations (Chappell et al. 2016)]. However, this resulted, for our computational approach, mainly in the activation of events such as DNA damage, DNA damage check, or cell survival. These statements do not match examples of relevant evidence according to the IARC's instructions (IARC 2017), which, for KCC 4, should involve, for example, terms associated with DNA methylation or histone modification. When we adjust the search term for this specific query by leaving out the 'rna' term, we see that the adjusted networks are much smaller, together with a reduction in BS (Figure S16).

We noticed that the percentage of PMIDs for which we received INDRA Statements was moderate (Table 2). A search on a number of PMIDs for which we retrieved no INDRA Statements showed that some papers (mostly older ones) had no abstract or the study was non-English. For our particular case study, we retrieved full papers when open access and relied on abstracts for others. We did this to make the methodology as open as possible for use by scientists in the hazard assessment process, and we conject that the most important results of a study would be made available in the abstract and, as such, the impact of not
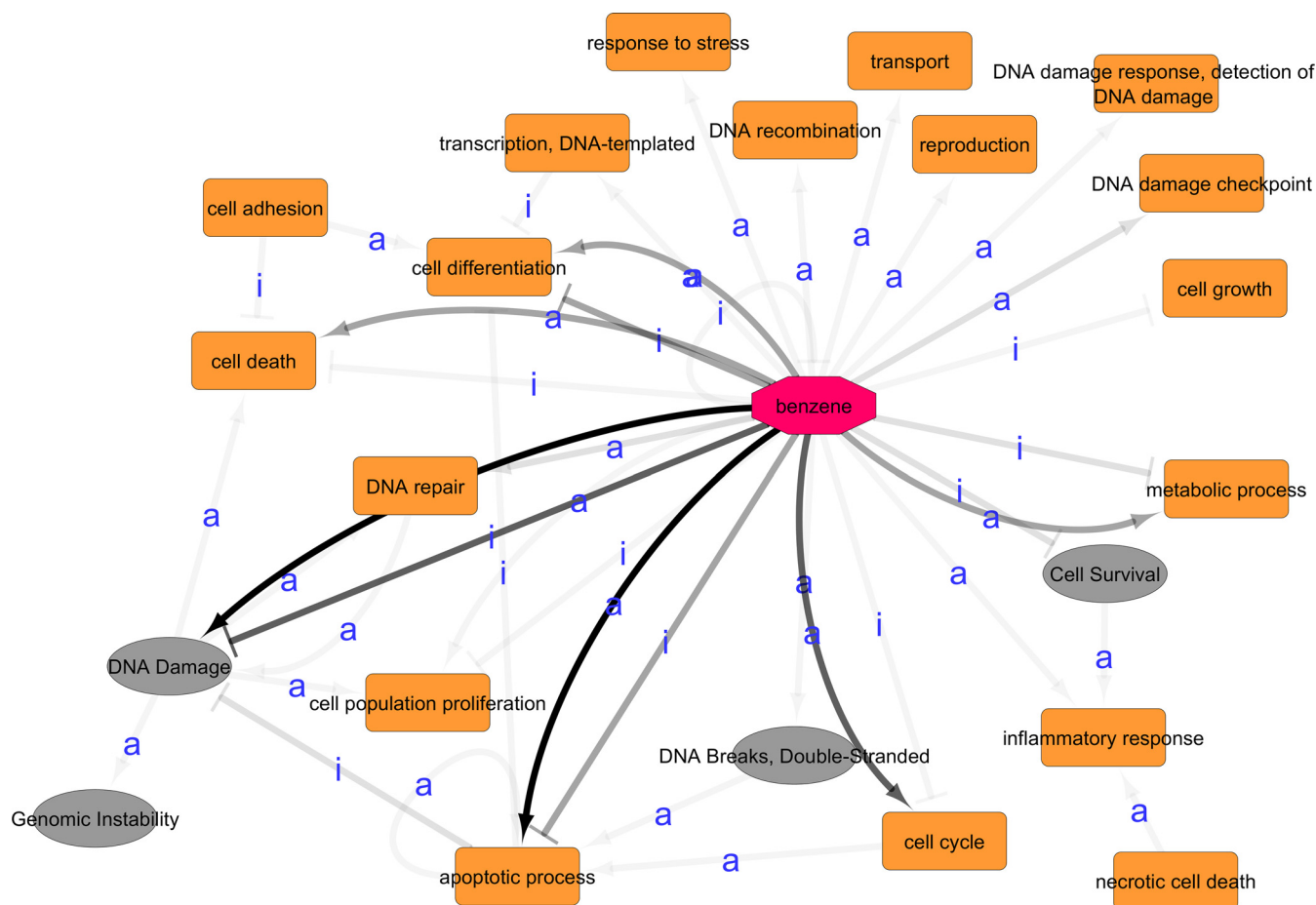
Figure 4. Benzene network for key characteristics of carcinogens (KCCs) on genotoxicity (KCC 2) and DNA repair (KCC 3), filtered on bioprocess (orange rounded rectangle-shaped box) and other processes (gray ellipse-shaped box). A darker arrow represents a higher belief score (BS): black: BS 0.95, dark gray: BS 0.94, gray: BS 0.92, light gray: BS 0.86, transparent gray: BS 0.65. Arrows: activation, an arrow with a raised edge: inhibition. Note: a, activation; i, inhibition.

having full access to all papers might be limited. However, this does illustrate that although we used the same search terms as the IARC working group, the evidence base [i.e., the selected studies to either generate networks (for our approach) or to evaluate the evidence (for the IARC working group)] was not identical for both approaches. We focused specifically on the IARC and the KCCs because these provide well-defined search terms for identifying literature but we recognize other institutes (e.g., National Toxicology Program Report on Carcinogens, U.S. Environmental Protection Agency) also include mechanistic data in their hazard assessment on carcinogens, including adaptations of the KCC literature queries (NTP 2016).

Last, we did not manually annotate papers for which relations are relevant and then check which of these the reading system (in our case, Reach) can pick up. The closest relevant evaluation as to the performance of Reach was done by Glavaški and Velicki (2021), who found a good accuracy of Reach but noted the extraction performance could be improved.

Our approach does not claim to fully automate and replace manual evaluation of mechanistic literature as is done in hazard identification, such as the IARC *Monographs* program. Instead, it could potentially be helpful in the prioritization of chemicals in relation to KCCs for further review, that is, to identify and create a network-based inventory of available studies, the content of which is to be further evaluated by an expert committee. Even though our findings are not directly generalizable outside the IARC framework, there is no reason to assume that our approach would not work well in other (noncancer) hazard identification

programs using a well-defined framework for the evaluation of mechanistic data such as the KCCs. Future work should also focus on strategies to qualitatively or quantitatively assess the strength of the evidence that is provided in the mechanistic literature. This would require identifying those study characteristics that are typically used by experts to define study quality and developing approaches to systematically extract these from identified publications in an automated way.

Acknowledgments

References

Ananiadou S, Pyysalo S, Tsujii J, Kell DB. 2010. Event extraction for systems biology by text mining the literature. Trends Biotechnol 28(7):381–390, PMID: 20570001, https://doi.org/10.1016/j.tibtech.2010.04.005.

Barupal DK, Schubauer-Berigan MK, Korenjak M, Zavadil J, Guyton KZ. 2021. Prioritizing cancer hazard assessments for IARC *Monographs* using an integrated approach of database fusion and text mining. Environ Int 156:106624, PMID: 33984576, https://doi.org/10.1016/j.envint.2021.106624.

Carvaillo JC, Barouki R, Coumoul X, Audouze K. 2019. Linking bisphenol S to adverse outcome pathways using a combined text mining and systems biology approach. Environ Health Perspect 127(4):047005, PMID: 30994381, https://doi.org/10.1289/EHP4200.

Chappell G, Pogribny IP, Guyton KZ, Rusyn I. 2016. Epigenetic alterations induced by genotoxic occupational and environmental human chemical carcinogens: a systematic literature review. Mutat Res Rev Mutat Res 768(Apr–Jun):27–45, PMID: 27234561, https://doi.org/10.1016/j.mrrev.2016.03.004.

EFSA (European Food Safety Authority). 2018. EFSA Scientific Colloquium 23—Joint European Food Safety Authority and Evidence-Based Toxicology Collaboration Colloquium Evidence integration in risk assessment: the science of combining apples and oranges 25–26 October 2017 Lisbon, Portugal. EFSA Support Publ 15(3):1396E, https://doi.org/10.2903/sp.efsa.2018.EN-1396.

Glavaški M, Velicki L. 2021. Humans and machines in biomedical knowledge curation: hypertrophic cardiomyopathy molecular mechanisms' representation. BioData Min 14(1):45, PMID: 34600580, https://doi.org/10.1186/s13040-021-00279-2.

Guha N, Guyton KZ, Loomis D, Barupal DK. 2016. Prioritizing chemicals for risk assessment using chemoinformatics: examples from the IARC Monographs on Pesticides. Environ Health Perspect 124(12):1823–1829, PMID: 27164621, https://doi.org/10.1289/EHP186.

Guyton KZ, Rusyn I, Chiu WA, Corpet DE, van den Berg M, Ross MK, et al. 2018. Application of the key characteristics of carcinogens in cancer hazard identification. Carcinogenesis 39(4):614–622, PMID: 29562322, https://doi.org/10.1093/carcin/bgy031.

Gyori BM, Bachman JA, Subramanian K, Muhlich JL, Galescu L, Sorger PK. 2017. From word models to executable models of signaling networks using automated assembly. Mol Syst Biol 13(11):954, PMID: 29175850, https://doi.org/10.15252/msb.20177651.

Howard BE, Phillips J, Miller K, Tandon A, Mav D, Shah MR, et al. 2016. SWIFT-Review: a text-mining workbench for systematic review. Syst Rev 5(May):87, PMID: 27216467, https://doi.org/10.1186/s13643-016-0263-z.

IARC (International Agency for Research on Cancer). 1992. *Occupational Exposure to Mists and Vapours from Strong Inorganic Acids; and Other Industrial Chemicals. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 54.* Lyon, France: IARC.

IARC. 2017. Instructions for Authors for the Preparation of Drafts for IARC Monographs. Last updated 2 February 2021. http://monographs.iarc.fr/ENG/Preamble/instructions.php [accessed 3 January 2020].

IARC. 2019. Preamble to the IARC Monographs on the Identification of Carcinogenic Hazards to Humans. Amended January 2019. https://monographs.iarc.who.int/wp-content/uploads/2019/07/Preamble-2019.pdf [accessed 8 May 2020].

National Academies of Sciences, Engineering, and Medicine. 2017. *Using 21st Century Science to Improve Risk-Related Evaluations*. Washington, DC: National Academies Press.

NTP (National Toxicology Program). 2016. *Standard Search Strings for Literature Database Searches: Appendix to the Draft Handbook for Preparing Report on Carcinogens Monographs*. https://ntp.niehs.nih.gov/ntp/roc/handbook/rochandbookappendix_508.pdf [accessed 4 March 2020].

Samet JM, Chiu WA, Cogliano V, Jinot J, Kriebel D, Lunn RM, et al. 2020. The IARC *Monographs*: updated procedures for modern and transparent evidence synthesis in cancer hazard identification. J Natl Cancer Inst 112(1):30–37, PMID: 31498409, https://doi.org/10.1093/jnci/djz169.

Shapiro AJ, Antoni S, Guyton KZ, Lunn RM, Loomis D, Rusyn I, et al. 2018. Software tools to facilitate systematic review used for cancer hazard identification. Environ Health Perspect 126(10):104501, PMID: 30392397, https://doi.org/10.1289/EHP4224.

Smith MT, Guyton KZ, Gibbons CF, Fritz JM, Portier CJ, Rusyn I, et al. 2016. Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. Environ Health Perspect 124(6):713–721, PMID: 26600562, https://doi.org/10.1289/ehp.1509912.

Steppi A, Gyori BM, Bachman JA. 2020. Adeft: acromine-based disambiguation of entities from text with applications to the biomedical literature. J Open Source Softw 5(45):1708, PMID: 32337477, https://doi.org/10.21105/joss.01708.

Valenzuela-Escárcega MA, Babur Ö, Hahn-Powell G, Bell D, Hicks T, Noriega-Atala E, et al. 2018. Large-scale automated machine reading discovers new cancer-driving mechanisms. Database 2018:bay098, PMID: 30256986, https://doi.org/10.1093/database/bay098.

Wikoff DS, Rager JE, Chappell GA, Fitch S, Haws L, Borghoff SJ. 2019. A framework for systematic evaluation and quantitative integration of mechanistic data in assessments of potential human carcinogens. Toxicol Sci 167(2):322–335, PMID: 30423162, https://doi.org/10.1093/toxsci/kfy279.