ADVANCED SCIENCE Open Access
www.advancedscience.com

Toxicogenomics Data for Chemical Safety Assessment and Development of New Approach Methodologies: An Adverse Outcome Pathway-Based Approach

*Laura Aliisa Saarimäki, Jack Morikka, Alisa Pavel, Seela Korpilähde, Giusy del Giudice, Antonio Federico, Michele Fratello, Angela Serra, and Dario Greco**

Mechanistic toxicology provides a powerful approach to inform on the safety of chemicals and the development of safe-by-design compounds. Although toxicogenomics supports mechanistic evaluation of chemical exposures, its implementation into the regulatory framework is hindered by uncertainties in the analysis and interpretation of such data. The use of mechanistic evidence through the adverse outcome pathway (AOP) concept is promoted for the development of new approach methodologies (NAMs) that can reduce animal experimentation. However, to unleash the full potential of AOPs and build confidence into toxicogenomics, robust associations between AOPs and patterns of molecular alteration need to be established. Systematic curation of molecular events to AOPs will create the much-needed link between toxicogenomics and systemic mechanisms depicted by the AOPs. This, in turn, will introduce novel ways of benefitting from the AOPs, including predictive models and targeted assays, while also reducing the need for multiple testing strategies. Hence, a multi-step strategy to annotate AOPs is developed, and the resulting associations are applied to successfully highlight relevant adverse outcomes for chemical exposures with strong in vitro and in vivo convergence, supporting chemical grouping and other data-driven approaches. Finally, a panel of AOP-derived in vitro biomarkers for pulmonary fibrosis (PF) is identified and experimentally validated.

1. Introduction

Mechanistic aspects of chemical exposures have been long exploited in the context of academic research, resulting in the emergence of toxicogenomics and systems toxicology as independent fields.[1,2] Although the mechanistic insight gained through the technologies employed in academia has been valued as supporting evidence in the regulatory setting, its incorporation into the regulatory framework is to date hindered by concerns related to the robustness and reproducibility of such data and its analysis.[3] At the same time, the growing need for faster, cheaper, and more ethical approaches for chemical safety assessment have made mechanistic toxicology central for clarifying aspects important to regulatory decision making. Furthermore, uncovering exposure related mechanistic properties is emerging as a fundamental approach for the design of new drugs and chemicals.[4,5] Hence, multiple high-end research initiatives are underway to drive the shift from traditional animal-based

L. A. Saarimäki, J. Morikka, A. Pavel, S. Korpilähde, G. del Giudice, A. Federico, M. Fratello, A. Serra, D. Greco
Finnish Hub for Development and Validation of Integrated Approaches (FHAIVE)
Faculty of Medicine and Health Technology
Tampere University
Arvo Ylpön katu 34, Tampere 33520, Finland
E-mail: dario.greco@tuni.fi

A. Serra
Tampere Institute for Advanced Study
Tampere University
Kalevantie 4, Tampere 33100, Finland

D. Greco
Institute of Biotechnology
University of Helsinki
P.O.Box 56, Helsinki, Uusimaa 00014, Finland

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/advs.202203984

© 2022 The Authors. Advanced Science published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/advs.202203984

assessment of apical toxicity endpoints toward in vitro and in silico approaches supported by mechanistic evidence.[6–8]

Adverse outcome pathways (AOP) emerged as models to organize biological mechanisms into causally linked sequences of multi-scale events to support chemical risk assessment.[9] AOPs have since expanded beyond the limits of toxicology, showing their applicability in organizing mechanisms of disease progression and adverse health outcomes,[10,11] and could even be applied to assess beneficial effects of therapies. The mechanisms depicted by AOPs comprise a sequence of events that progress from the molecular initiating event (MIE) toward an adverse outcome (AO) through intermediate steps, key events (KEs), with biological complexity increasing as the AOP progresses. Individual KEs are connected by key event relationships (KER) that verbally explain the causal link between the events and provide context for the pathway.

The AOP concept quickly attracted attention due to its potential in tackling one of the major challenges in the shift away from traditional toxicology: deciphering systemic and long-term outcomes of chemical exposures without the use of animal experiments. While significant efforts still need to be made toward this goal, AOPs encompass the means to systematically guide the integration of in vitro-based evidence into the risk assessment framework.[12] AOPs provide the grounds for various predictive approaches, read-across, and the development of targeted assays and new approach methodologies (NAMs), as also suggested by regulatory agencies and international organizations, such as the OECD.[8] Furthermore, the construction of AOPs can help identify gaps in knowledge and guide resources toward mechanisms in need of further investigation, or alternatively, reveal connections that have not been previously characterized.[13]

Concurrently with the development of the AOP framework, the role of omics data in elucidating biomarkers and mechanisms of action (MOA) of chemical exposures and diseases has become more prominent.[14–18] Omics data have been used to support the development of AOPs, especially through the identification of molecular targets and mechanisms.[19–23] However, full exploitation of omics-based evidence in the context of AOPs is hindered by the complication of linking molecular data to complex biological events, affecting both the development and the application of AOPs. Furthermore, while the value of omics data in answering questions of regulatory importance is recognized, the complexity of its interpretation and the lack of standardization in analysis and reporting have hampered widespread regulatory acceptance of omics-based evidence.[24] Bypassing these challenges could broaden the application of AOPs, support the interpretation of complex omics data, and further aid in the development of the concept toward quantitative models and assays. While molecular assays based on arbitrarily selected reporter genes have been proposed (e.g., ToxCast assays), there is an urgent need to develop new data-driven unbiased molecular assays for reliable and efficient mechanistic safety assessment of chemicals.

Here we hypothesized that rigorous curation of molecular events associated with AOPs could facilitate the implementation of omics-based evidence to 1) guide the interpretation of omics data readout, 2) support the development of new AOPs, 3) identify and fill gaps in knowledge, and 4) transfer AOP-based knowledge into robust assays to support chemical safety assessment.

Well-curated gene ontologies, pathways, and biological processes are used to interpret omics results and their translation into biologically relevant information. While some KEs can be easily crosslinked with such terms and their associated genes, the annotation of complex KEs taking place at a higher level of biological organization (e.g., at the tissue- or organism-level) is a more demanding task. This requires knowledge regarding ontologies and the biological events themselves. For instance, generic annotations are helpful for categorizing KEs, but without the intention of modeling KEs using the associated gene sets, they will likely not reach the level of granularity required for such a task. This is currently reflected in the annotations provided in the AOP-Wiki repository (aopwiki.org). The annotation of KEs to selected ontologies is included as an option in AOP-Wiki. However, the coverage of the annotations is currently low and has not been intended for modeling the KEs using the gene sets associated with their annotations.

Previous efforts to curate external annotation have shown the potential of the approach.[25,26] However, these have either remained at the level of abstract associations or focused on individual examples.[27,28] Hence, systematic, fit-for-purpose, and up-to-date annotation linking KEs to curated gene sets has not yet been established. To this end, we applied an integrated strategy for defining gene-KE-AOP associations through systematic curation. We show the applicability of our strategy for evaluating potential AOs of chemical exposures, and for the identification of AOP-driven biomarkers that can inform the development of target assays and novel approaches to chemical hazard characterization.

2. Results and Discussion

We developed an integrated approach to systematically associate curated gene sets to the KEs and AOPs. Our approach combines natural language processing (NLP) techniques with manual curation to link relevant biological processes and pathways, as well as their associated genes, to KEs of AOPs relevant for human health risk assessment. The resulting gene-KE-AOP connections enable the modeling of KEs and AOPs through gene-level data, which further introduces novel ways to benefit from the AOP concept. We applied this approach to generate an AOP fingerprint for a known profibrotic exposure in vivo and in vitro and finally combined the annotation to a framework for prioritizing KE- and AOP-associated genes to guide the discovery of biomarkers and reporter genes. The complete approach described in the following sections is summarized in **Figure 1**.

2.1. The Majority of KEs can be Successfully Annotated to Curated Gene Sets

At the time of retrieving the data from the AOP-Wiki repository (November 2020), a total of 289 AOPs and 1131 distinct KEs were identified. However, after eliminating the AOPs for which taxonomic applicability was either not available nor in the scope of our investigation, 176 AOPs and 856 unique KEs remained, forming a total of 1245 unique AOP-KE pairs (specific KEs). Although the AOP-Wiki houses selected annotations for some of

 © 2022 The Authors. Advanced Science published by Wiley-VCH GmbH

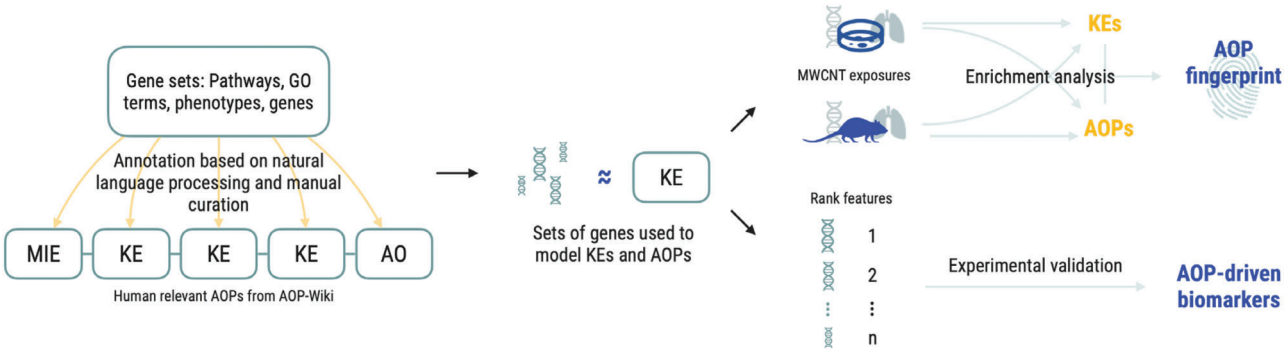

Figure 1. Overall scheme of the study. Established gene sets were annotated to KEs of the AOPs relevant for human health risk assessment. The resulting gene sets were then used to model the KEs. The validity of the annotation was evaluated using gene signatures of exposures with known adverse outcomes. Finally, we combined the approach with a gene prioritization framework resulting in the identification of AOP-driven biomarkers for pulmonary fibrosis.

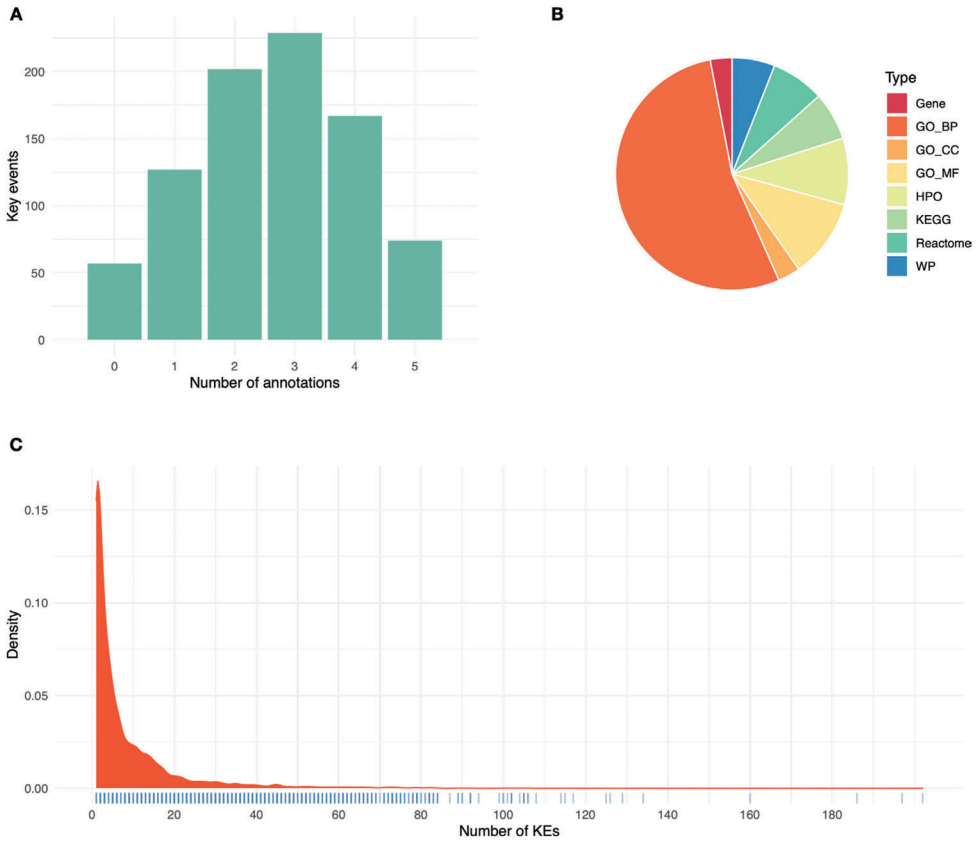

Figure 2. Descriptive analysis of the KE annotation. A) Bar plot describing the number of annotated terms per KEs. B) Pie chart expressing the proportions of different annotation types. C) Density distribution of the number of KEs each gene is annotated to.

the KEs, majority of them were considered not to be specific enough for our purpose (i.e., KEs describing the dysregulation of a specific gene annotated to terms such as "gene expression"). Additionally, as the existing annotations only cover a part of the KEs, we decided to consistently curate the annotation of all KEs. As a result, 799 unique KEs mapped to 175 AOPs received a curated annotation. The KEs were treated as individual instances, hence the same KE mapped to multiple AOPs was always annotated to the same term(s). A summary of the number of terms annotated to the KEs is presented in **Figure 2**A along with the proportions of the different term sources (Figure 2B). GO biological processes (GO_BP) represent most of the mapped annotations, followed by GO molecular functions (GO_MF) and human phenotype ontology (HPO). Since up to five annotations were provided for the KEs, the final gene sets used from herein comprise the union of the genes mapped to each annotated term. This structure

 © 2022 The Authors. Advanced Science published by Wiley-VCH GmbH

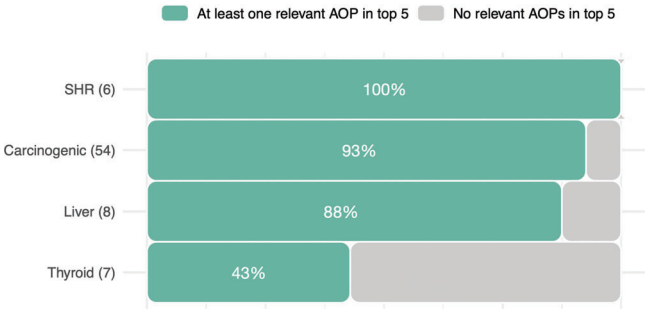

Figure 3. Bar plot representing the proportion of chemicals with relevant AOPs among the top five enriched AOPs based on the chemical classification. Number in brackets after the category name refers to the number chemicals in each category while the percentage on the bars reflects the proportion of chemicals in each category highlighting relevant AOPs. SHR stands for sex hormone receptor agonist.

allowed improved specificity, while also providing the possibility to further refine the gene sets using the hierarchical order implemented where applicable. The size of the gene sets associated to each KE range from one to 5990 genes, with a median value of 81 genes. Consequently, when AOPs are modeled by combining the gene sets associated to their KEs, the gene set sizes range from 15 to 5992 with the median size being 752 genes.

In total, the annotations comprise 15 825 genes. While the majority of genes are annotated to less than 5 KEs (9044 genes), 1434 genes have more than 20 KEs associated to them, and 50 genes have more than 80 associated KEs (Figure 2C). Although these numbers can be affected by annotation bias, for example, certain genes are better researched and annotated than others, they can also guide the selection of AOP-driven biomarkers when specificity is of importance.

2.2. AOP Enrichment Highlights Relevant Adverse Outcomes Associated to Chemicals

We tested the ability of our novel annotations to highlight relevant AOPs by analyzing a set of curated reference chemicals as defined by EU Reference Laboratory for alternatives to animal testing (ECVAM) and National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). We focused on four categories of chemicals defined by their toxicity properties to include hepatotoxic and carcinogenic agents as well as thyroid disruptors and sex hormone receptor (estrogen receptor—ER, and androgen receptor—AR) agonists. For each of the selected chemicals, we retrieved a list of associated genes from the Comparative Toxicogenomics Database (CTD),[29] resulting in a final set of 75 chemicals (File S1, Supporting Information).

First, we identified AOPs related to each of the selected categories (i.e., AOPs related to carcinogenesis, hepatotoxicity, sex hormones, and thyroid disruption) among all the AOPs. We then evaluated the prevalence of these relevant AOPs among the five most significantly enriched AOPs for each chemical. The results suggest that the enrichment approach successfully highlights AOPs of relevance for each group of chemicals (**Figure 3**). All sex hormone receptor agonists had at least one relevant AOP among the top five enriched, while the proportions varied from 43% (thyroid disrupters) to 93% (carcinogens) in the other categories (Figure 3).

In the group of carcinogenic chemicals, 93% of the compounds evaluated had cancer-related adverse outcomes among the top enriched AOPs. In fact, the group of carcinogens had the highest proportion of relevant AOPs at the top as compared to the others (median four out of five compared to the median of two out of five in the other groups). However, it should be noted that AOPs related to cancer are among the most represented group of AOPs, and cancer-related genes are generally highly researched and annotated, which may introduce a level of annotation bias that should be recognized.

The remaining four carcinogenic chemicals (7%) that showed no cancer AOPs among the top enriched AOPs were *N*-nitrosodiethanolamine, *N*-nitrosomorpholine, phenacetin, and tetrachloroethylene. *N*-nitrosomorpholine and *N*-nitrosodiethanolamine are both nitrosamines whose suspected AOs besides carcinogenesis include non-alcoholic steatohepatitis.[30] Indeed, both compounds contained hepatic steatosis related AOPs among the top five enriched AOPs (File S1, Supporting Information). Tetrachloroethylene (perchloroethylene, PCE) is a chlorocarbon solvent used in dry-cleaning and other degreasing applications.[31] AOPs with the most significant enrichment for PCE were also related to hepatic adverse outcomes. Although neurotoxicity is one of the most frequent AOs associated with PCE exposure, hepatotoxicity has also been reported.[31] Our results documenting liver steatosis are supported by biopsy-based evidence of liver disease, both in human and animal models, in settings of high occupational exposures.[32] Last, phenacetin is a drug that was widely used as pain medication until it was withdrawn from the market across the globe due to increasing evidence of carcinogenicity and renal toxicity.[33] The most enriched AOPs for phenacetin included immune related AOPs "Immune mediated hepatitis" (Aop:362) and Aop:277 titled "Inhibition of IL-1 binding to IL-1 receptor leading to increases susceptibility to infection". Although there is no described association between phenacetin and IL-1 or immuno-toxicity, it is known that they both play a role in paracetamol-associated liver toxicity, which is the main metabolite of phenacetin.[33,34]

In the case of the known liver toxicants, hexaconazole was the only compound not highlighting AOPs associated with liver toxicity among the top enriched AOPs. Hexaconazole is a widely used triazole fungicide. It acts by blocking sterol biosynthesis via inhibition of cytochrome P450.[35] Hexaconazole was considered as a Group C-Possible Human Carcinogen by the US EPA due to increased incidence of benign Leydig cell tumors in rats (https://www3.epa.gov/pesticides/chem_search/hhbp/R000356.pdf). Moreover, it was found to affect the reproduction of female rats.[35] The top enriched AOPs correctly identify this signature. Furthermore, the top two pathways "HMG-CoA reductase inhibition leading to decreased fertility" and "Modulation of adult Leydig cell function subsequent to decreased cholesterol synthesis or transport in the adult Leydig cell" both suggest a decrease in cholesterol levels by inhibition of the HMG-CoA reductase. Drugs inhibiting this enzyme, such as statins, are known to possibly cause liver damage.[36]

Known thyroid toxicants performed poorest in our analysis. Bifenthrin, malathion, permethrin, and simazine did not capture

 © 2022 The Authors. Advanced Science published by Wiley-VCH GmbH

thyroid related AOPs among the top five enriched. All these compounds have been widely used in agriculture as herbicides or pesticides. Agrochemicals represent a significant class of endocrine disrupting chemicals, albeit through varying mechanisms. It is now accepted that many of these molecules may mimic the interaction of endogenous hormones with nuclear receptors, such as estrogen, androgen, and thyroid hormone receptors.[37] Indeed, bifenthrin has already been reported as an endocrine-disrupting compound by blocking the binding of endogenous hormones.[38] In our framework, its anti-estrogenic activity emerges as the most enriched AOP (File S1, Supporting Information). Malathion is an organophosphate pesticide that is known for its low acute toxicity and rapid degradation.[39] In this light, it is not listed as a primary thyroid disrupting chemical, and its toxicity has been associated with the inhibition of acetylcholinesterase activity on nerve impulse.[39] Recent studies, however, demonstrated that malathion acts as an endocrine disruptor, both in vitro and in vivo.[40,41] Our results support these findings, highlighting the Aop:165: "Anti-estrogen activity leading to ovarian adenomas and granular cell tumors in the mouse" as well as Aop:112: "Increased dopaminergic activity leading to endometrial adenocarcinomas." Furthermore, Moore et al. demonstrated that malathion exposure at higher concentrations induces cytotoxic and genotoxic effects in HepG2 through oxidative stress, which can finally lead to liver cancer.[39] Similarly, our framework highlights both the "PPARalpha-dependent liver cancer" and "Cyp2E1 activation leading to liver cancer" AOPs. Simazine is a triazine herbicide whose use has been banned in most European countries for nearly two decades.[42] Simazine has now been recognized, similarly to the other compounds, as an endocrine disrupter.[42] Interestingly, the enrichment analysis for simazine highlighted AOPs related to the development of adenomas and carcinomas through endocrine disrupting activities (e.g., Aop:107 titled "Constitutive androstane receptor activation leading to hepatocellular adenomas and carcinomas in the mouse and the rat") as well as direct disruption of the GnRH pulse (File S1, Supporting Information). Although multiple in vivo and in silico evidence also indicate permethrin as possible endocrine disruptor,[43,44] no endocrine related pathways are present in the top enriched AOPs. However, this framework was able to highlight the modulating effect of permethrin on the lipid metabolism. It has been demonstrated that in HepG2 cells, permethrin increases lipogenesis and decreases beta oxidation, possibly contributing to the development of NAFLD.[45] Indeed, the "Inhibition of fatty acid beta oxidation leading to nonalcoholic steatohepatitis (NASH)" AOP is statistically enriched in our results.

Together, these results highlight relevant AOPs modeled by our curated gene sets to be enriched by the genes associated to the compounds, suggesting that our framework is able to support robust mechanistic and data-driven chemical grouping as well as the identification of potential AOs using chemical-gene associations.

2.3. Our Annotation Enables Grouping of KEs Resulting in Improved Modeling of the AOP Network

In order to fully unleash the potential of mechanistic toxicology, more informative testing strategies need to be developed that can monitor specific phases of the exposure-bio interactions and mechanisms. To this end, we defined accurate sets of genes capable of modeling specific KEs and AOPs. However, one of the challenges observed in the AOP-Wiki is the redundant semantics in the naming of KEs. While creating a new KE can be meaningful in many cases (e.g., the same biological process taking place in a distinct organ or a tissue), unnecessary redundancy can lead to challenges in the application of the AOP-based knowledge. This is especially true when modeling AOPs as a network and using such representation to identify hidden connections and to perform read-across analysis.[10,46–51]

Hence, we hypothesized that KEs could be grouped based on the degree of similarity of their associated gene lists. We calculated the similarity of the KEs based on their annotated gene sets and grouped together those mapped to identical sets of genes (Jaccard Index (JI) = 1). This resulted in the identification of 637 groups of varying sizes. These groups were characterized by four main concepts: 1) truly duplicated KEs due to distinct semantics; 2) same biological event in multiple biological systems; 3) subsequent KEs mapped to the same terms due to inadequate specificity; and 4) opposite regulation of the same biological event (i.e., increased vs decreased signaling).

Here, the grouping based on identical gene sets was selected due to the nature of the downstream application and statistical considerations (i.e., to avoid multiple testing against the same gene set in enrichment analysis). However, a parallel approach with varying cut-off values for similarity could be implemented to cluster KEs more roughly and to define specific categories of events. Similarly, further refinement of the KE clusters could help to enhance the AOP network by removing redundant nodes which, in turn, could reveal hidden links.

The potential of the KE grouping was showcased using a subgraph formed by considering the AOPs related to pulmonary fibrosis (PF). PF is a chronic lung disease characterized by tissue damage and scarring that impairs lung function.[52] A range of environmental exposures, including certain chemicals, drugs, radiation, and nanomaterials (most notably carbon nanotubes), as well as infectious diseases have been identified as causative agents for PF.[52–54] Moreover, the COVID-19 pandemic has raised concerns about increasing rates of PF.[55–57] Understanding the disease mechanisms can help in the development of strategies to treat and prevent the disease, and to control and modulate the exposures that contribute to its pathogenesis and progression. Furthermore, it can serve as the foundation for developing targeted assays for evaluating profibrotic potential of chemical exposures.

Six AOPs related to PF were available in the AOP-Wiki at the time of data retrieval (**Figure 4**A). These distinct AOPs characterize multiple pathways leading to the same AO. Together, these AOPs comprise 30 KEs, which form a connected graph when modeled as a network (Figure 4C). However, several redundancies were observed among the KEs. For instance, the AO was expressed either as *lung fibrosis* (Event:1276) or *pulmonary fibrosis* (Event:1458). Hence, the application of the similarity-based grouping resulted in 23 distinct KEs (Figure 4B) that were then used as the basis for merging the KE nodes in the PF network (Figure 4D).

The PF AOPs formed a connected network, indicating that each of the individual AOPs shared at least one KE with one or

 © 2022 The Authors. Advanced Science published by Wiley-VCH GmbH

A

AOP	AOP description
Aop:173	Substance interaction with the lung resident cell membrane components leading to lung fibrosis
Aop:206	Peroxisome proliferator-activated receptors γ inactivation leading to lung fibrosis
Aop:241	Latent Transforming Growth Factor beta1 activation leads to pulmonary fibrosis
Aop:319	ACE2 inhibition leading to lung fibrosis
Aop:347	Toll-like receptor 4 activation and peroxisome proliferator-activated receptor gamma activation leading to pulmonary fibrosis
Aop:382	Angiotensin II type 1 receptor (AT1R) agonism leading to lung fibrosis

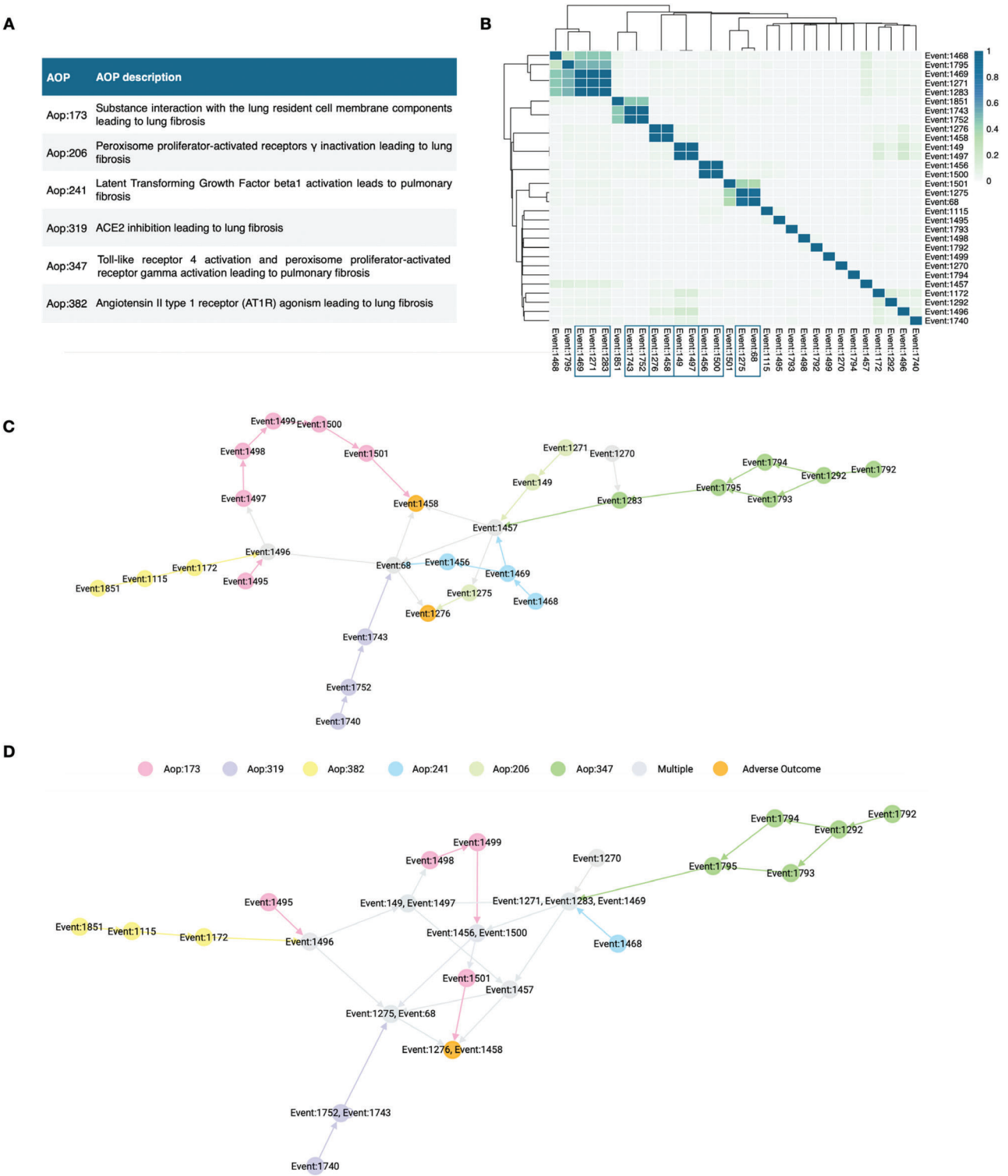

Figure 4. A) Table presentation of pulmonary fibrosis (PF) AOPs identified in the current study. B) Heatmap representing the Jaccard index-based similarity of the PF KEs as per their associated gene sets. Values close to zero (light gray) correspond to a low similarity between distinct KEs, while the increasing similarity is expressed with the color changing through green to blue. C) Graph presentation of the PF AOPs using their original KEs. Distinct colors denote the KEs of individual AOPs, gray nodes are KEs shared by multiple AOPs, and orange nodes correspond to the shared AOs. D) Graph presentation of the PF AOPs after KE grouping. The number of shared (gray) nodes has now increased, and the duplicated AO has been grouped into one distinct AO (orange).

 © 2022 The Authors. Advanced Science published by Wiley-VCH GmbH

Figure 5. AOP fingerprint of Mitsui-7 exposure in vitro and in vivo. Size of the bubble reflects the proportion of significantly enriched KEs in an AOP, while to color denotes the FDR-adjusted *p*-value in a negative logarithmic scale (i.e., the higher the number, the smaller the *p*-value). The AOPs are organized by the enrichment *p*-value from the in vivo data.

more of the other AOPs. However, as the duplicated KEs were merged, the similarities between the AOPs became more evident. This is evidenced by the increasing number of shared KEs in the graph after merging (the gray nodes in Figure 4D) as compared to the original graph (Figure 4C). Furthermore, the merging revealed Aop:206 to be fully contained within the other AOPs.

The refinement of the AOP network through KE grouping simplifies the network while also enhancing the robustness of the KE relationships, depicted by the connections between the nodes. This process, in fact, removes redundant nodes, which supports the application of AOP networks in AOP research and risk assessment. Furthermore, as duplicated events are removed, the true influence of each node can be assessed more robustly through network analytics.

This example demonstrates the effect of KE redundancy and the potential of data-driven grouping of the KEs. While manual assessment and grouping would be achievable for a limited number of AOPs at a time, doing it AOP-Wiki wide would be a massive undertaking. Here, we show how our curated gene-KE-AOP connections can help guide the grouping and hence enhance network-based approaches in AOP research. Furthermore, our results suggest that it is often possible to identify genes that can successfully represent multiple similar key events.

2.4. The AOP Fingerprint of Multi-Walled Carbon Nanotubes Converges In Vitro and In Vivo

Toxicogenomics has supported the development of mechanistic toxicology and further enhanced the possibility to obtain relevant information from in vitro studies, which could reduce the need for animal experimentation.[58–60] Here, we tested the hypothesis that toxicogenomic data generated in two independent in vitro and in vivo exposure models would converge on a robust set of relevant AOPs. We focused on Mitsui-7, a known hazardous long and rigid multi-walled carbon nanotube (MWCNT). The airways provide the most prominent route of exposure to this nanomaterial, and it is best characterized for its lung-related AOs, including PF.[61–64] Hence, we selected data derived from a lung exposure to the MWCNT in mice,[65] and an in vitro dataset with exposures on four cell lines representative of the human airways.[59,66] These cell lines include differentiated THP-1 cells as a model of macrophages, A549 representing alveolar basal epithelial cells, BEAS-2B as bronchial epithelial cells, and MRC9 as a model of lung fibroblasts. Differentially expressed genes (DEGs) from all experimental conditions in vivo and each cell line in vitro were obtained from Saarimäki et al.[67] and merged into a single MOA in vivo and in vitro, respectively.

We then performed enrichment analysis against both the AOPs and the KEs separately in order to evaluate the coverage of distinct KEs. We used the proportion of significantly enriched KEs to further filter the significant AOPs. This led us to identify 33 significant AOPs from the in vivo data, while 12 resulted significant from the in vitro exposure. These results were defined as the specific AOP fingerprint for the exposures, and it is presented in **Figure 5**.

Despite the distinct sizes of the AOP fingerprints, ten of the 12 AOPs enriched in vitro were also included in the in vivo fingerprint. Moreover, the top enriched AOPs were shared and ranked similarly between in vivo and in vitro when ranked by the

 © 2022 The Authors. Advanced Science published by Wiley-VCH GmbH

smallest adjusted *p*-value. The AOP enriched with the most significant *p*-value in both instances was Aop:173 titled "Substance interaction with the lung resident cell membrane components leading to lung fibrosis" (Figure 5). The in vivo data set was able to capture seven of the eight KEs of the AOP as significantly enriched, while three out of the eight KEs were enriched in vitro. Interestingly, Aop:173 has been specifically developed with robust evidence from MWCNT exposures, and multiple types of carbon nanotubes are listed as known stressors for the AOP (https://aopwiki.org/aops/173). The second AOP (Aop:171), on the other hand, describes the induction of pleural/peritoneal mesotheliomas by chronic cytotoxicity in rats. Like most AOPs used in this study, Aop:171 is still under development and lacks information on potential stressors. However, mesothelioma is a well-known AO of asbestos exposure, a fibrous silicate mineral whose adverse effects have often been used as a warning example for MWCNTs.[68] Indeed, similarities in their MOA have been extensively investigated.[61,69,70]

The in vitro AOP fingerprint captures effects such as frustrated phagocytosis, oxidative stress, cytotoxicity, and immune activation, which have all been reported as consequences of this type of exposure and contribute to the pathogenic nature of Mitsui-7.[61,62,64,71] Similarly, the profibrotic effects are highlighted with the multiple PF AOPs enriched. These effects are also observed in the in vivo AOP fingerprint. However, the in vivo fingerprint further highlights various AOPs outside the respiratory system, which is less apparent in vitro. While AOs beyond the immediate exposure site are feasible, many of these could likely be accounted for by the different effects of similar transcriptomic signatures in different biological systems (e.g., multiple AOPs related to gastric ulcer formation could reflect similar mechanisms of surfactant disturbance in two distinct exposure sites). On the other hand, the AOPs unique to the in vitro fingerprint, Aop:277 and Aop:263 (Figure 5), reflect the specific effects of the Mitsui-7 exposure on the immune system. Such specific signals can be easily masked in the in vivo system, where a large array of cell types is affected by the exposure.

It is worth noting that the exposures selected for the analysis had diverse set ups and a notable difference in the size of the combined MOA (863 DEGs in vitro versus 3540 in vivo). While data from multiple cell lines were selected to capture effects besides immune cell activation in vitro, we were not able to match the dose and time course set up present in the in vivo dataset. However, we wanted to include this long-term exposure to evaluate whether it would result in broader coverage over the KEs of AOPs. Furthermore, histopathological evaluation from the same in vivo exposure set up has shown fibrosis in the lung from the day 7 onward,[72] suggesting that a whole PF AOP could be covered with this data. Indeed, all but the MIE (Event:1495) of Aop:173 were enriched in vivo. The high proportion of enriched KEs in the in vivo data supports the modeling of KEs with relevant gene sets and the use of toxicogenomic evidence for the development of AOPs, as well as the evaluation of potential AOs of chemical exposures. Likewise, we show that the analysis of toxicogenomic data against robustly annotated AOP framework supports a high degree of in vitro to in vivo extrapolation and further supports the inclusion of toxicogenomics-based evidence for regulatory purposes. The concept of the AOP fingerprint can be easily adapted to evaluate other chemical exposures and AOs. With nearly 16 000 genes mapped to the KEs in our curation, they are expected to cover most of the human genome. Hence, the AOP fingerprint provides a robust framework for meaningful interpretations also for chemicals and phenotypes that may be less characterized.

2.5. KE-Associated Gene Sets Guide the Selection of Biomarkers

We showed that our KE-linked gene sets provide a robust way of evaluating potential outcomes of chemical exposures from transcriptomics data. This observation alone can help to guide chemical testing and grouping. However, to support the development of target assays and integrated approaches, specific reporter genes and markers need to be identified.

Selection of transcriptional biomarkers and reporter genes only based on differential expression from experimental data gives little context or reference to the AO of interest. Even if a certain exposure is known to induce a specific endpoint, there is no indication whether the measured deregulation could be associated with the phenotype of interest. On the other hand, prioritizing the features in the context of the KEs or whole AOPs could shed light on the importance and specificity of the feature regarding the phenotype. This, in turn, can guide the selection of potential biomarkers even in the absence of experimental data. Hence, we implemented a universal and customizable framework for the prioritization of the KE-associated genes to drive the identification of AOP-informed biomarkers and used it to identify AOP-driven biomarkers for PF. The shortlisted marker genes were then screened by RT-qPCR in an in vitro model of human macrophages exposed to bleomycin, a well-known profibrotic chemical.[73]

First, we defined characteristics for optimal biomarkers based on the Bradford Hill criteria, originally defined to evaluate causality in epidemiological research,[74] but later adopted to other research fields as well.[75] Our newly defined characteristics, their Bradford Hill counterparts, and short descriptions of the consideration of each step in the selection process are summarized in **Table 1**.

The prioritization and selection of the candidate biomarkers considered three main aspects: 1) the social life of genes, that is, some genes (gene products) are more influential than others; 2) specificity regarding the endpoint of interest; and 3) experimental evidence suggesting the genes respond to a relevant exposure. The ranking of the genes was based on the first two considerations, while the experimental evidence was included to guide the selection of candidate genes for RT-qPCR validation from the ranked list. This enabled a flexible selection process that would be applicable even in the absence of experimental data. At this stage, we also considered the biological feasibility of the target genes given the selected macrophage model as well as a broad coverage over the PF KEs.

As a result, we obtained a list of 25 candidates out of the original 2075 genes related to PF (**Table 2**). Although we focused on the genes in the top 10%, we further included genes ranking lower to obtain a broader coverage over the PF KEs. Genes that are specific to individual KEs might rank low when the individual lists are combined. Hence, we considered the expression patterns from the experimental data as well as the specificity scores

 © 2022 The Authors. Advanced Science published by Wiley-VCH GmbH

Table 1. We defined characteristics for optimal biomarkers based on the Bradford Hill criteria. The characteristics were then implemented into the prioritization and selection protocol, and further to the evaluation of the prioritized genes.

Bradford Hill	Our characteristic	Method/Assessment
Consistency (reproducibility)	Reproducibility	Selection considers evidence from previous profibrotic exposures
Strength (effect size)	Amplitude	Significant alteration of the expression as compared to control
Experiment	Measurable	Transcriptional biomarkers measurable by qPCR; selected genes need to be expressed in the model
Biological gradient (dose-response relationship)	Dose-responsive	Benchmark dose modeling to evaluate dose-response
Coherence	In vitro to in vivo extrapolation	Experimental evidence from in vitro and in vivo[a)]
Analogy	Predictive (of the outcome of interest)	Selection based on the KEs preceding the AO of interest
Specificity	Specificity	Gene ranking based on the specificity score
Plausibility	(Biological) plausibility	The AOP framework provides a plausible context; supporting evidence; selection of the organism
Temporality	Temporality	Transcriptional alteration follows the exposure; selection of the model organism[b)]
–	GLP-method	RT-qPCR
–	Influence	Centrality measures from human protein-protein interaction and gene regulatory networks

[a)] The biomarkers selected here are targeted for the development of non-animal assays for toxicological assessment. Hence the coherence to in vivo set ups is not evaluated experimentally. However, in vivo data was used for the selection of the markers to provide context of the systemic response [b)] Temporality in the Bradford Hill criteria refers to a clear distinction of the exposure happening prior to the outcome. Here, we considered temporality by observing transcriptional changes post exposure as well as in the selection of the model organism. Macrophages have a crucial role in the initiation of the profibrotic response preceding the outcome, fibrosis.

and ranks in the individual KEs. This also allowed us to evaluate whether genes ranked higher would perform better than others.

We could detect the expression of 22 of the candidate genes at one or more of the evaluated time points, and six of the detected genes showed significant alteration as compared to the unexposed control samples (Table 2). Finally, five of these genes were altered in a dose-dependent manner: CXCL2 and CCL7 at 24 h, IL8 (CXCL8) at 24 and 72 h, and MMP19 and TGFBI at 72 h. All but TGFBI of these genes were among the top 10% in the global PF rank (Table 2). Although we were not able to fit a dose-dependent model on the highest ranked gene, SMAD7, a suggestive trend could be appreciated in its expression pattern (Figure S4, Supporting Information, panel 18/6H). The expression of each of these genes was upregulated as compared to the controls (Figure S4, Supporting Information).

The central role of TGF-beta signaling is well-established in PF,[76] but neither of the TGFB genes tested (TGFB1 and TGFB3) showed significant change in expression in our setup. Indeed, TGF-beta is activated through a complex cascade of events, where the inactive form of the protein is activated by other effectors post-translationally,[77] making members of the TGF-beta family a more attractive target for protein-based biomarker assessment over gene expression. At the same time, we did observe upregulation of SMAD7 and TGFBI which are both activated by TGF-beta,[78,79] suggesting the induction of TGF-beta signaling. The protein encoded by TGFBI is involved in the extracellular matrix (ECM), and it has been shown to bind type I collagen, resulting in thicker fibers and further affecting macrophage polarization toward the M2 type.[80] Indeed, bleomycin has been suggested to polarize macrophages toward M2 (often referred to as the anti-inflammatory type), which have been shown to drive the development of PF through their ability to promote myofibroblast differentiation.[81,82] Many of our suggested biomarkers are chemokines that mediate immune responses. IL8 and CXCL2 are best characterized as neutrophil attractants, while CCL7 targets a wide variety of leukocytes.[83–85] Indeed, prolonged inflammation, combined with persistent M2 macrophage activation, supports pathogenesis of fibrosis,[86] and a mixed status of M1/M2 macrophage activation has been previously associated with carbon nanotube-induced PF in vivo.[87] Similarly, multi-walled carbon nanotubes have been shown to induce the polarization of macrophages toward such mixed status of M1/M2 polarization.[88,89] MMP19 is a member of the matrix metalloproteinase family involved in ECM remodeling.[90] MMPs have been extensively characterized in the context of PF,[91,92] and MMP19 specifically has been suggested as a key regulator of PF in mice and humans.[93]

Although macrophages alone cannot capture all the KEs of PF, our model is able to highlight the key steps of macrophage involvement in PF. The temporality of the expression of our suggested biomarkers is supportive of the events leading to the development of fibrosis, where the initial inflammation is followed by type M2 macrophage activation that together contribute to the development of a profibrotic microenvironment and responses in other cells in the tissue.[86]

NAMs are urgently needed to reduce animal testing while providing robust evidence to support chemical safety assessment. Although alternative methods have been successfully developed to capture acute effects, modeling long-term outcomes of the exposures, such as fibrosis, in vitro is still a challenge. Here, we propose a panel of five genes CXCL2, CCL7, IL8, MMP19, and TGFBI as AOP-derived robust biomarkers of PF to be

 © 2022 The Authors. Advanced Science published by Wiley-VCH GmbH

Table 2. Genes selected for qPCR validation. Green = yes, white = no.

Gene (rank)	Time point	Detected (Amplification)	Deregulated (ANOVA)	Dose-dependent
	6 h			
SMAD7 (1)	**24 h**			
	72 h			
	6 h			
CXCL2 (3)	**24 h**			
	72 h			
	6 h			
SPP1 (18)	**24 h**			
	72 h			
	6 h			
CCL2 (19)	**24 h**			
	72 h			
	6 h			
TGFB1 (23)	**24 h**			
	72 h			
	6 h			
IL8 (33)	**24 h**			
	72 h			
	6 h			
LOX (48)	**24 h**			
	72 h			
	6 h			
PLOD2 (74)	**24 h**			
	72 h			
	6 h			
MMP7 (80)	**24 h**			
	72 h			
	6 h			
CXCL10 (91)	**24 h**			
	72 h			
	6 h			
CCL7 (93)	**24 h**			
	72 h			
	6 h			
MMP9 (105)	**24 h**			
	72 h			
	6 h			
LTBP4 (112)	**24 h**			
	72 h			
	6 h			
FN1 (116)	**24 h**			
	72 h			
	6 h			
GDF15 (153)	**24 h**			
	72 h			
	6 h			
MMP19 (179)	**24 h**			
	72 h			

(*Continued*)

 © 2022 The Authors. Advanced Science published by Wiley-VCH GmbH

Table 2. (Continued).

Gene (rank)	Time point	Detected (Amplification)	Deregulated (ANOVA)	Dose-dependent
	6 h			
PTX3 (220)	**24 h**			
	72 h	■		
	6 h	■		
TGFB3 (297)	**24 h**			
	72 h			
	6 h			
LTBP3 (335)	**24 h**			
	72 h	■		
	6 h	■		
TWIST1 (609)	**24 h**	■		
	72 h	■		
	6 h	■		
TGFBI (727)	**24 h**	■		
	72 h	■	■	■
	6 h	■		
CTSK (759)	**24 h**	■		
	72 h	■		
	6 h	■		
RCN3 (1592)	**24 h**	■		
	72 h	■		
	6 h			
RSAD2 (1596)	**24 h**			
	72 h	■		
	6 h	■		
PLK3 (2027)	**24 h**			
	72 h			

successfully measured in a model of human macrophages in vitro after short exposure time.

3. Conclusion

Mechanistic toxicology encompasses the means for faster, cheaper, and more ethical chemical safety assessment. However, to unleash the full potential of mechanistic evidence also in the regulatory framework, robust approaches to build confidence toward toxicogenomics are urgently needed. Here, we presented an integrated approach that links toxicogenomics with the concept of AOPs and proved its applicability to chemical grouping and development of data-driven NAMs. We introduced the AOP fingerprint, a concept for evaluating potential systemic outcomes of chemical exposures through unbiased interpretation of toxicogenomics data. Our analysis points to a consistent AOP fingerprint of MWCNTs extrapolated from both in vitro and in vivo experiments. Finally, we identified and experimentally validated a panel of robust AOP-derived in vitro biomarkers for PF.

Our results suggest that combining the regulatory-supported AOP framework with toxicogenomics through a rigorous mapping of the MOA of chemicals into KEs/AOPs can facilitate the inclusion of omics derived evidence in regulatory evaluations. The outcome of our analysis in the form of the AOP fingerprint provides a clear and easily understandable way to summarize complex omics data while providing robust statistical evaluation that can support regulatory decisions. Moreover, the possibility to use the framework suggested in this manuscript as the foundation for developing data-driven molecular assays further opens new possibilities for faster regulatory acceptance of novel alternative methods and NAMs.

4. Experimental Section

Definition of Knowledge Graph-Based Data Structure: A knowledge graph-based data structure was established by expanding the previously introduced framework, the Unified Knowledge Space (UKS) by the authors.[94,95] A detailed description and a full list of integrated data sources are provided in the Supporting Information. The so formed data structure was utilized throughout the study as described in the following sections.

Annotation of Key Events: A multi-step strategy comprising NLP and manual curation was applied to annotate KEs to established gene sets through pathways, phenotypes, and gene ontologies. The annotation strategy is summarized in Figure S2, Supporting Information.

 © 2022 The Authors. Advanced Science published by Wiley-VCH GmbH

Computational prioritization of KE annotations: To match the descriptions of key events and gene sets an NLP pipeline (Figure S2, Supporting Information) was developed. The pipeline performed several operations to extract the informative terms from both the descriptions of a key event and a gene set that were scored to assess the degree of matching between the two entities. In detail, first, the raw text was converted to lower case and all punctuation symbols were removed. Second, concepts that span multiple words in the text description were replaced by a single word expressing the same concept to strengthen the matching quality (e.g., the concept "positive regulation" was replaced with the single word "upregulated"). Third, the text was split into tokens which were further processed one by one. Fourth, each token corresponding to a stop word in the English language was dropped. Stop words refer to the most common words in a language that do not bring additional meaning (e.g., for the English language common stop words include "in", "the", "of", "from"). Fifth, different declinations of the same concept were mapped to their root term (e.g., plurals were converted to singulars, the terms "increased" and "increasing" were both mapped to "increase"). This same procedure was also used to standardize several styles to write the same symbol (e.g., "pparα" and "ppar-alpha" map both to "ppar-alpha"). After these preprocessing steps, each gene set and key event was represented by a set of token words, for example, {upregulate, ppar-alpha}. However, the frequency of each token word across the descriptions of genes and key events was not the same, and hence, the informative value of rare terms was higher than the informative value of more common tokens. This was taken into account by weighting each token by its inverse document frequency (IDF), that is, the weight was inversely proportional to the number of gene sets and key events that contain that token. Finally, a weighted version of the Jaccard Index (JIW) was employed to match gene sets and the key events, using the IDF as weights (i.e., each token that was shared between a gene and a key event did not account for 1 as in the standard JI, but it contributed its IDF weight to the matching score) and the matching gene sets for each key event was sorted in descending order.

Manual curation and refinement of annotations: Next, the results of the computational prioritization were manually evaluated for correct context and accuracy. Manual curation was used for gap filling and refinement of the annotations. In detail, the top five matches retained from the NLP-based approach were evaluated, and inaccurate or spurious matches were discarded. In case no matches from the computational prioritization were deemed suitable, a manual search related to the name of the KE was performed on relevant databases (WikiPathways,[96] HPO,[97] KEGG,[98] Reactome,[99] and GO [100]). For molecular level KEs, where the alteration of an individual gene was described, either the main functions of the gene were selected, or the gene was directly annotated to the ensemble identifier of the said gene. More generic annotations (i.e., annotation of a KE describing the alteration of a gene to a functional term tightly related to that gene) were prioritized to increase the size of the relevant gene sets.

The matches for KEs expressing the increase/activation or decrease/repression of a biological progress were further organized based on the hierarchy of the terms by prioritizing the most generic but suitable term followed by increasing specificity when multiple annotations of various specificities were available. For instance, Ke:1457 called "induction, epithelial mesenchymal transition" was annotated to the following terms: 1) epithelial to mesenchymal transition (GO:0 001837); 2) regulation of epithelial to mesenchymal transition (GO:00 10717); and 3) positive regulation of epithelial to mesenchymal transition (GO:00 10718). The curated KE-gene set links were added to the UKS so that for each key event entity its top five matches were added, while the matching level was stored as an edge attribute. This allowed to either combine multiple mappings for a key event or to filter for specific mapping levels. Since the KE-gene set mappings were always the same for the same KE, these relationships were added to the *Key Event* entities and not to the *Specific Key Event* entities, which reduced complexity of the knowledge graph as well as reduced needed storage space. The information, however, could still be retrieved from the UKS via its connecting paths.

Gene set retrieval: The genes corresponding to the matched terms were retrieved by matching the term names to their exact identifiers and querying the UKS for human genes associated with the terms. For phenotypes (HPO and KEGG disease), only genes with a link in the original database were included by filtering by the source for the connection. In cases where no human genes were linked to the annotated GO term, the mouse and rat genes associated and converted them to human orthologs using Ensembl,[101] which were then used as the corresponding gene sets. When no genes of human, mouse, or rat were associated with the original term, the annotation match was discarded and considered unsuccessful. Once gene sets to all original terms were defined, the gene sets were merged to obtain the final set of genes corresponding to each KE in this study.

Enrichment Analysis of Reference Chemical-Associated Gene Sets: To evaluate the ability of this framework to highlight relevant adverse outcomes from chemical associated gene signatures, lists of reference chemicals were retrieved from the ECVAM reference chemical library[102] and the NICEATM website (https://ntp.niehs.nih.gov/whatwestudy/niceatm/resources-for-test-method-developers/refchem/index.html). From the resources provided by ECVAM, a hepatotoxic chemical list that had clear distinctions between positive and negative compounds was selected. This list was based on the work from EPA's Virtual Liver project (https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=166616&Lab=NCCT), and was provided as a downloadable Excel-file (./CHELIST/CheLIST__EPA_VLIVER.xlsx) by ECVAM. From NICEATM, the list of chemicals with characterized thyroid activity (specified as "ACTIVE" in the listing produced based on a previous publication by Wegner et al.) was selected.[103] AR and ER agonists were selected from the lists of in vitro reference chemicals provided on the website. These listings had been previously published in Kleinstreuer et al.[104] and Browne et al.,[105] respectively. Finally, carcinogenic compounds were identified from the list containing chemicals that were either known carcinogens or reasonably anticipated to be human carcinogens (RAHC) based on the 14th report on Carcinogens (RoC classifications) provided by NICEATM. The chemicals from each of the reference lists were then matched to the list of chemicals obtained from the CTD[29] through name-based matching or by the provided CAS identifiers, resulting in the final lists of reference chemicals for each endpoint used in this study. Chemical-gene links originating from the CTD were retrieved from the UKS and only chemicals with 50–1000 associated genes were included in the enrichment analysis. This filter was applied to minimize the false discovery rate and spurious matches in the enrichment analysis.

Enrichment analysis was performed using the Fisher's exact test as implemented in the function *enrich* from R package bc3net[106] for each chemical associated gene set against the list of AOP-related genes (i.e., the union of all the genes associated to all the KEs of the AOP). Enrichment *p*-values were adjusted using the false discovery rate (FDR) correction. AOP was considered significantly enriched with FDR-corrected p-value < 0.01.

KE Clustering and Construction of the Pulmonary Fibrosis Network: Similarities between the gene sets associated to each KE were evaluated by calculating the JI between all pairs of KEs (size of the intersection divided by the size of the union of the gene sets). The resulting similarity matrix was then transformed into a distance matrix and used to group the KEs using hierarchical clustering as implemented in the function *hclust* in R package stats, specifying the agglomeration method as "complete." The number of clusters was defined so that only KEs with the same gene sets associated to them (JI = 1) were assigned to the same group. The grouping obtained in this manner was used to perform the enrichment against KEs to avoid multiple testing against the same gene set as well as to enhance the network presentation of the PF AOP network. The unweighted PF AOP network was generated using gephi[107] by importing a graphml file generated with the function *graph_from_edgelist* from R package igraph.[108] KE groups from the clustering were added as attributes to the nodes and used to merge redundant nodes in gephi. Similarly, AOPs each KE is associated to were added as attributes and used to color the nodes.

Characterization of the AOP Fingerprints: *Transcriptomics data*: In vivo and in vitro transcriptomics data from MWCNT (Mitsui-7) exposures were selected from a previously published collection by Saarimäki et al.[67] The original data sets are available under GEO accession number GSE29042 (in vivo) and ArrayExpress entry EMTAB6396 (in vitro), while the

 © 2022 The Authors. Advanced Science published by Wiley-VCH GmbH

preprocessed data is available on Zenodo (https://doi.org/10.5281/zenodo.6425445). The in vivo data set comprised multiple doses and time points, while the in vitro data contained a single dose and time point exposure on four distinct cell lines representative different cell types of the lung. In each case, DEGs (filtered by an absolute fold change > 1.5 and Benjamini and Hochberg adjusted p-value < 0.05) obtained from Zenodo (https://doi.org/10.5281/zenodo.6425445) for each distinct comparison (i.e., combination of each dose and time point versus control in vivo and separate cell lines in vitro) were pooled together to obtain a distinct MOA of the exposure in vivo and in vitro, respectively.

AOP fingerprints: To produce the AOP fingerprint for the MWCNT exposures, enrichment analysis was performed using the Fisher's exact test as implemented in the function *enrich* from R package bc3net[106] separately against the AOP-associated gene lists and the KE-associated gene lists (KEs linked to the same set of genes were grouped to avoid multiple tests against the same set). An AOP was considered significantly enriched when the AOP itself and at least 33% (or minimum of 2 KEs when the length of the AOP was less than six) of its KEs were enriched with an FDR-corrected p-value < 0.05.

Selection and Testing of AOP-Driven Biomarkers: *Gene prioritization*: All human protein-protein interaction (PPI) edges were extracted from the UKS and used to create a robust gene–gene network. PPI/gene (product) interaction information could vary across data sources as well as the covered genes may differ. In addition, there was an innate bias in the data, where more data sources were available for "more investigated" genes and gene products. Because of this, it was decided not to apply a global threshold on how many sources need to support an edge,[109] but instead, a local threshold was applied. This ensured that also less investigated genes and gene products will be retained in the final robust gene–gene network, but their edges were less penalized by the number of supporting edges, than highly covered gene (product) nodes. For each node, the mean number of sources supporting all its connecting edges was estimated and only edges with at least the "mean number of sources" for a node were retained, which needed to be true for at least one of the nodes making up an edge. This was only done for GENE nodes, which were flagged as protein coding in Ensembl. The final robust human gene–gene network, consisted of 20 260 nodes, 806 250 edges, a network density of 0.0039. Due to the significant lower number of available sources for transcription factor—gene (product) data, all available sources were kept and scored equally. These edges were used to create a directed gene–gene network, consisting of 18 754 nodes, 363 649 edges and a network density of 0.001. On the so created gene–gene networks, for each node its degree, betweenness, eigenvector, and closeness centrality were estimated with NetworkX.[110] These measures were then used to rank the genes linked to the KEs in the context of individual KEs. The gene list for each KE was ranked based on each of the centrality measures (degree, betweenness, closeness, and eigenvector centrality) individually from most central to the least. The ranked lists were combined using the Borda method as implemented in the function *Borda* in R package TopKLists.[111]

Biomarker selection: The gene centrality-based ranking was then supplemented by a specificity ranking for the KEs of AOPs related to PF. A specificity score for the genes in the context of the KEs by dividing the occurrence of the gene in the KEs of PF AOPs by their occurrence in the KEs of other AOPs were calculated. A similar score was calculated at the level of AOPs (occurrence in PF AOPs/occurrence in other AOPs), as universal PF biomarkers were the target of the identification (i.e., prioritizing those that would be present in as many of the six PF AOPs as possible) while also being as specific as possible to PF. These ranks were again combined by the function *Borda* from R package TopKLists,[111] and a final round of the Borda method was applied to combine the lists of genes from each KE into one PF rank. The final rank was complemented with experimental evidence. It was assessed whether the genes were differentially expressed in the Mitsui-7 exposures in vivo and in vitro. It was also evaluated whether they were dose-dependently altered in the in vivo data as well as in an additional in vitro data set on Mitsui-7 exposure of a THP-1 macrophage model (originally published in Saarimäki et al.[58] and the preprocessed data available as GSE146708 in the previously published collection[67] available in https://doi.org/10.5281/zenodo.6425445). The dose-response modeling of the in vivo (GSE29042) and in vitro (GSE146708) datasets was performed by following the strategy implemented in the BMDx tool.[112] Particularly, for each gene present in the dataset, multiple models were fitted including linear, second order polynomial, hill, power, and exponential model. For each gene, the optimal model was selected as the one with the lowest Akaike Information Criterion (AIC). Genes with an optimal model with lack-of-fit p-values lower than 0.1 were removed from the analysis. The effective doses (BMD, BMDL, and BMDU) were estimated under the assumption of constant variance and by using a BMRF factor of 1.349 (corresponding to a minimum of 10% of difference with respect to the controls). Genes were further filtered based on the predicted doses. Genes with BMD or BMDU values extrapolated higher than the highest exposure dose were filtered. Moreover, genes whose ratio between the predicted doses was higher than the suggested values (BMD/BMDL> 20, BMDU/BMD> 20, and BMDU/BMDL> 40) were removed from the analysis. Genes passing the filters were considered to be dose-dependently altered. At this stage, the measurability and feasibility of the gene in the selected macrophage model was also considered. For instance, numerous collagen-encoding genes were ranked high, but would not be a meaningful target in a macrophage model. Moreover, a high coverage of PF KEs and the selection of genes with high specificity scores were emphasized. With these considerations, a subset of the genes with the following priority was selected: 1) genes that were deregulated both in vivo and in vitro, with most emphasis on dose-dependency; 2) genes that were deregulated in vitro, with most emphasis on dose-dependency; and 3) genes that were not significantly differentially expressed but were dose-dependent. Finally, after this initial selection driven by the rank and experimental evidence, additional candidate biomarkers that had a lower rank but were specific to KEs that would otherwise not have been covered by the selected candidates were included.

Cell culture: THP-1 cells (DSMZ no.: ACC 16) were grown in RPMI 1640 (Gibco, #21 875) + 10% inactivated FBS (Gibco, #10 270). Cells were cultivated in 75 cm^2 culture flasks at 37 °C with a humidified atmosphere of 5% CO_2. For all experiments, cells were seeded at a density of 1×10^5 cells mL^{-1} in 96 well plates and differentiated for 48 h with 25 nM PMA (phorbol-12-myristate-13-acetate, Sigma-Aldrich, #P1585). Cells were then left to rest for 24 h in fresh media containing no PMA prior to bleomycin exposures.

Cell viability assay: THP-1 cells were exposed to 0–10 µg mL^{-1} of bleomycin ready-made solution (Sigma-Aldrich, #B7216) and 0–100 mg mL^{-1} of Triclosan (Sigma-Aldrich, #72 779), for 6, 24 and 72 h. A WST-1 assay was then used to measure cell viability. Briefly, 10 µL of cell proliferation reagent WST-1 (Roche, #11 644 807 001) was added to each well. Cells were left to incubate with WST-1 for 3 h in a 37 °C, 5% CO2 incubator. Absorbance at 450 nm was then measured with a Spark microplate reader (Tecan). Results of the cell viability assay are available in File S2 and Figure S3, Supporting Information.

RT-qPCR: For each time point of 6, 24 and 72 h, THP-1 cells were exposed to 0, 2.5, 5, 10 and 100 µg mL^{-1} of bleomycin ready-made solution (Sigma-Aldrich, #B7216). Media was removed and cells were washed briefly with 50 µL of PBS. 100 µL of lysis buffer from the QIAGEN RNeasy mini kit (Qiagen, #74 104) was added to each well to lyse the cells. Three wells (300 µL) were pooled to create one sample, and there were five samples for each concentration at each time point. Total RNA was then extracted from these samples using the QIAGEN RNeasy mini kit (Qiagen, #74 104). DNase treatment was performed using TURBO DNA-free Kit (ThermoFisher, #AM1907) according to the manufacturer's protocol. cDNA was synthesized from 100 ng of RNA, using the high-capacity cDNA reverse transcription kit (Thermo Fisher Scientific, #4 368 813), according to manufacturer's instructions. Expression levels of target genes were determined by qRT-PCR using CFX96 Touch Real-Time PCR Detection System (BioRad) with 10 µL of iQ Multiplex Powermix (Bio-Rad, #1 725 849), 5 µL of cDNA diluted fivefold, 2.5 µL of nuclease-free (NF) water (not DEPC-Treated, ThermoFisher, #AM9930) in a 20 µL reaction, together with 2.5 µL of single (1 µL assay + 1.5 µL NF water) or multiplexed (0.5 µL of each assay) PrimePCR Probe Assays (Bio-Rad) as followed with single or multiplex reactions grouped in parentheses and formatted as *Gene*/UniqueAssayID:

 © 2022 The Authors. Advanced Science published by Wiley-VCH GmbH

(*ACTB*/qHsaCEP0036280), (*SMAD7*/qHsaCEP0050142, *MMP9*/qHsaCIP0028098, *GDF15*/qHsaCEP0051579, *CTSK*/qHsaCIP0030907, *PLOD2*/qHsaCEP0052848), (*CXCL2*/qHsaCEP0058163, *LTBP4*/qHsaCEP0024931, *TGFB3*/qHsaCEP0058244, *RCN3*/qHsaCEP0057804, *MMP7*/qHsaCEP0052037), (*SPP1*/qHsaCEP0058179, *FN1*/qHsaCEP0050873, *LTBP3*/qHsaCEP0053782, *RSAD2*/qHsaCIP0031596, *CCL7*/qHsaCEP0058033), (*IL8*/qHsaCEP0053894, *MMP19*/qHsaCEP0051244, *TWIST1*/qHsaCEP0051221, *PLK3*/qHsaCIP0027687, *CXCL10*/qHsaCEP0053880), (*LOX*/qHsaCEP0050731, *PTX3*/qHsaCEP0033071, *TGFBI*/qHsaCEP0058394, *CCL2*/qHsaCIP0028103, *TGFB1*/qHsaCIP0030973).

Fold change (FC) values from RT-qPCR data were calculated using the comparative CT($2^{-(ddCt)}$) method.[113] The FC values were log2 transformed ($\log_2$(FC)). For each gene and for each concentration, an outlier detection was performed by removing all the samples with $\log_2$(FC) values above or below the 75th and 25th percentiles of the distribution. C_t values, dC_t values, FC values and $\log_2$(FC) values are available in File S2, Supporting Information, along with ANOVA tables and tukey HSD posthoc test results. The $\log_2$FC expression of the genes as compared to the untreated controls are plotted in Figure S4, Supporting Information.

Dose-dependent modeling: A dose-response analysis of the $\log_2$(FC) values derived from the PCR experiments was performed. For each gene, multiple models were fitted, including linear, hill, power, polynomial, exponential, log-logistic, Weibull, and Michaelis–Mentel models. The optimal model was selected as the one with the lowest AIC. The BMD estimation was performed under the assumption of constant variance. The BMR was identified by means of the standard deviation approach with a BMRF of 1.349. Only genes with lack-of-fit *p*-value >0.10 and with estimated BMD, BMDL and BMDU values were considered relevant.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This work received funding from the EU Horizon 2020 project NanoSolveIT (grant agreement no. 814572), the Academy of Finland (grant agreement no. 322761), and the European Research Council (ERC) programme, Consolidator project ARCHIMEDES (grant agreement no. 101043848). L.A.S. was supported by the Emil Aaltonen Foundation and A.S. by the Tampere Institute for Advanced Study.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

Keywords

adverse outcome pathways, biomarkers, new approach methodologies, toxicogenomics

Received: July 11, 2022
Revised: November 9, 2022
Published online: December 7, 2022

[1] M. D. Waters, J. M. Fostel, *Nat. Rev. Genet.* **2004**, *5*, 936.
[2] Z. Liu, R. Huang, R. Roberts, W. Tong, *Trends Pharmacol. Sci.* **2019**, *40*, 92.
[3] P. Marx-Stoelting, A. Braeuning, T. Buhrke, A. Lampen, L. Niemann, M. Oelgeschlaeger, S. Rieke, F. Schmidt, T. Heise, R. Pfeil, R. Solecki, *Arch. Toxicol.* **2015**, *89*, 2177.
[4] J. Paananen, V. Fortino, *Brief Bioinf.* **2020**, *21*, 1937.
[5] A. Mech, S. Gottardo, V. Amenta, A. Amodio, S. Belz, S. Bøwadt, J. Drbohlavová, L. Farcal, P. Jantunen, A. Małyska, K. Rasmussen, J. Riego Sintes, H. Rauscher, *Regul. Toxicol. Pharmacol.* **2022**, *128*, 105093.
[6] A. M. Richard, R. Huang, S. Waidyanatha, P. Shinn, B. J. Collins, I. Thillainadarajah, C. M. Grulke, A. J. Williams, R. R. Lougee, R. S. Judson, K. A. Houck, M. Shobair, C. Yang, J. F. Rathman, A. Yasgar, S. C. Fitzpatrick, A. Simeonov, R. S. Thomas, K. M. Crofton, R. S. Paules, J. R. Bucher, C. P. Austin, R. J. Kavlock, R. R. Tice, *Chem. Res. Toxicol.* **2021**, *34*, 189.
[7] D. Krewski, D. Acosta, M. Andersen, H. Anderson, J. C. Bailar, K. Boekelheide, R. Brent, G. Charnley, V. G. Cheung, S. Green, K. T. Kelsey, N. I. Kerkvliet, A. A. Li, L. Mccray, O. Meyer, R. D. Patterson, W. Pennie, R. A. Scala, G. M. Solomon, M. Stephens, J. Yager, L. Zeise, Staff Of Committee On Toxicity Test, *J. Toxicol. Environ. Health, Part B* **2010**, *13*, 51.
[8] OECD, *Guidance Document for the Use of Adverse Outcome Pathways in Developing Integrated Approaches to Testing and Assessment (IATA)*, OECD Series on Testing and Assessment, Vol. *260*, OECD Publishing, Paris **2017**.
[9] G. T. Ankley, R. S. Bennett, R. J. Erickson, D. J. Hoff, M. W. Hornung, R. D. Johnson, D. R. Mount, J. W. Nichols, C. L. Russom, P. K. Schmieder, J. A. Serrrano, J. E. Tietge, D. L. Villeneuve, *Environ. Toxicol. Chem.* **2010**, *29*, 730.
[10] L.-A. Clerbaux, N. Amigó, M. J. Amorim, A. Bal-Price, S. Batista Leite, A. Beronius, G. F. G. Bezemer, A.-C. Bostroem, A. Carusi, S. Coecke, R. Concha, E. P. Daskalopoulos, F. De Bernardi, E. Edrosa, S. W. Edwards, J. Filipovska, N. Garcia-Reyero, F. N. E. Gavins, S. Halappanavar, A. J. Hargreaves, H. T. Hogberg, M. T. Huynh, D. Jacobson, J. Josephs-Spaulding, Y. J. Kim, H. J. Kong, C. E. Krebs, A. Lam, B. Landesmann, A. Layton, et al., *ALTEX* **2022**, *39*, 322.
[11] P. Nymark, M. Sachana, S. B. Leite, J. Sund, C. E. Krebs, K. Sullivan, S. Edwards, L. Viviani, C. Willett, B. Landesmann, C. Wittwehr, *Front. Public Health* **2021**, *9*, 638605.
[12] K. E. Tollefsen, S. Scholz, M. T. Cronin, S. W. Edwards, J. De Knecht, K. Crofton, N. Garcia-Reyero, T. Hartung, A. Worth, G. Patlewicz, *Regul. Toxicol. Pharmacol.* **2014**, *70*, 629.
[13] S. Halappanavar, S. Van Den Brule, P. Nymark, L. Gaté, C. Seidel, S. Valentino, V. Zhernovkov, P. HÃ¸Gh Danielsen, A. De Vizcaya, H. Wolff, T. Stöger, A. Boyadziev, S. S. S. Poulsen, J. B. Sørli, U. Vogel, *Part. Fibre Toxicol.* **2020**, *17*, 16.
[14] V. Fortino, P. A. S. Kinaret, M. Fratello, A. Serra, L. A. Saarimã¤Ki, A. Gallud, G. Gupta, G. Vales, M. Correia, O. Rasool, J. Ytterberg, M. Monopoli, T. Skoog, P. Ritchie, S. Moya, S. VÃ¡Zquez-Campos, R. Handy, R. Grafström, L. Tran, R. Zubarev, R. Lahesmaa, K. Dawson, K. Loeschner, E. H. Larsen, F. Krombach, H. Norppa, J. Kere, K. Savolainen, H. Alenius, B. Fadeel, et al., *Nat. Commun.* **2022**, *13*, 3798.
[15] J. Montaner, L. Ramiro, A. Simats, S. Tiedt, K. Makris, G. C. Jickling, S. Debette, J.-C. Sanchez, A. Bustamante, *Nat. Rev. Neurol.* **2020**, *16*, 247.
[16] A. Federico, A. Serra, M.y K. Ha, P. Kohonen, J.-S. Choi, I. Liampa, P. Nymark, N. Sanabria, L. Cattelani, M. Fratello, P. A. S. Kinaret, K. Jagiello, T. Puzyn, G. Melagraki, M. Gulumian, A. Afantitis, H. Sarimveis, T.-H. Yoon, R. Grafström, D. Greco, *Nanomaterials (Basel)* **2020**, *10*, 903.

 © 2022 The Authors. Advanced Science published by Wiley-VCH GmbH

[17] P. A. S. Kinaret, A. Serra, A. Federico, P. Kohonen, P. Nymark, I. Liampa, M.y K. Ha, J.-S. Choi, K. Jagiello, N. Sanabria, G. Melagraki, L. Cattelani, M. Fratello, H. Sarimveis, A. Afantitis, T.-H. Yoon, M. Gulumian, R. Grafström, T. Puzyn, D. Greco, *Nanomaterials* **2020**, *10*, 750.
[18] A. Serra, M. Fratello, L. Cattelani, I. Liampa, G. Melagraki, P. Kohonen, P. Nymark, A. Federico, P. A. S. Kinaret, K. Jagiello, M.y K. Ha, J.-S. Choi, N. Sanabria, M. Gulumian, T. Puzyn, T.-H. Yoon, H. Sarimveis, R. Grafström, A. Afantitis, D. Greco, *Nanomaterials* **2020**, *10*, 708.
[19] E. K. Brockmeier, G. Hodges, T. H. Hutchinson, E. Butler, M. Hecker, K. E. Tollefsen, N. Garcia-Reyero, P. Kille, D. Becker, K. Chipman, J. Colbourne, T. W. Collette, A. Cossins, M. Cronin, P. Graystock, S. Gutsell, D. Knapen, I. Katsiadaki, A. Lange, S. Marshall, S. F. Owen, E. J. Perkins, S. Plaistow, A. Schroeder, D. Taylor, M. Viant, G. Ankley, F. Falciani, *Toxicol. Sci.* **2017**, *158*, 252.
[20] E. J. Perkins, E. A. Woolard, N. L. Garcia-Reyero, *Front. Toxicol.* **2022**, *4*, 786057.
[21] S. Labib, A. Williams, C. L. Yauk, J. K. Nikota, H. Wallin, U. Vogel, S. Halappanavar, *Part. Fibre Toxicol.* **2016**, *13*, 15.
[22] Y. Jin, G. Qi, Y. Shou, D. Li, Y. Liu, H. Guan, Q. Zhang, S. Chen, J. Luo, L. Xu, C. Li, W. Ma, N. Chen, Y. Zheng, D. Yu, *J. Hazard. Mater.* **2022**, *425*, 128041.
[23] R. Guan, N. Li, W. Wang, W. Liu, X. Li, C. Zhao, *Ecotoxicol. Environ. Saf.* **2022**, *234*, 113387.
[24] L. A. Saarimäki, G. Melagraki, A. Afantitis, I. Lynch, D. Greco, *Nat. Nanotechnol.* **2022**, *17*, 17.
[25] M. Martens, T. Verbruggen, P. Nymark, R. Grafström, L. D. Burgoon, H. Aladjov, F. T. Andón, C. T. Evelo, E. L. Willighagen, *Front. Genet.* **2018**, *9*, 661.
[26] M. Martens, C. T. Evelo, E. L. Willighagen, *Appl. In Vitro Toxicol.* **2022**, *8*, 2.
[27] P. Nymark, L. Rieswijk, F. Ehrhart, N. Jeliazkova, G. Tsiliki, H. Sarimveis, C. T. Evelo, V. Hongisto, P. Kohonen, E. Willighagen, R. C. Grafström, *Toxicol. Sci.* **2018**, *162*, 264.
[28] K. Jagiello, S. Halappanavar, A. Rybińska-Fryca, A. Willliams, U. Vogel, T. Puzyn, *Small* **2021**, *17*, 2003465.
[29] A. P. Davis, C. J. Grondin, R. J. Johnson, D. Sciaky, J. Wiegers, T. C. Wiegers, C. J. Mattingly, *Nucleic Acids Res.* **2021**, *49*, D1138.
[30] M. Tong, A. Neusner, L. Longato, M. Lawton, J. R. Wands, S. M. de la Monte, *J. Alzheimers Dis.* **2009**, *17*, 827.
[31] D. M. Ceballos, K. M. Fellows, A. E. Evans, P. A. Janulewicz, E. G. Lee, S. G. Whittaker, *Front. Public Health* **2021**, *9*, 638082.
[32] M. Cave, K. C. Falkner, C. McClain, in *Zakim and Boyer's Hepatology: A Textbook of Liver Disease*, Elsevier, Amsterdam **2012**, pp. 476–492.
[33] W.-H. Chow, L. M. Dong, S. S. Devesa, *Nat. Rev. Urol.* **2010**, *7*, 245.
[34] O. Krenkel, J. C. Mossanen, F. Tacke, *Hepatobiliary Surg. Nutr.* **2014**, *3*, 331.
[35] M. Jia, Y. Wang, D. Wang, M. Teng, J. Yan, S. Yan, Z. Meng, R. Li, Z. Zhou, W. Zhu, *Chemosphere* **2019**, *226*, 520.
[36] W.-S. Lee, J. Kim, *Mol. Cell. Toxicol.* **2019**, *15*, 9.
[37] G. R. Warner, V. E. Mourikes, A. M. Neff, E. Brehm, J. A. Flaws, *Mol. Cell. Endocrinol.* **2020**, *502*, 110680.
[38] S. M. Brander, G. He, K. L. Smalling, M. S. Denison, G. N. Cherr, *Environ. Toxicol. Chem.* **2012**, *31*, 2848.
[39] P. D. Moore, C. G. Yedjou, P. B. Tchounwou, *Environ. Toxicol.* **2010**, *25*, 221.
[40] J. Xiong, L. Tian, Y. Qiu, D. Sun, H. Zhang, M. Wu, J. Wang, *Drug Chem. Toxicol.* **2018**, *41*, 501.
[41] B. Lal, M. K. Sarang, P. Kumar, *Gen. Comp. Endocrinol.* **2013**, *181*, 139.
[42] F. Grasselli, S. Bussolati, R. Ramoni, S. Grolli, G. Basini, *Anim. Reprod.* **2018**, *15*, 3.
[43] W. Tu, C. Xu, Y. Jin, B. Lu, C. Lin, Y. Wu, W. Liu, *Aquat. Toxicol.* **2016**, *175*, 39.
[44] S.-S. Kim, R.-D.a Lee, K.-J.o Lim, S.-J. Kwack, G.-S. Rhee, J.i-H. Seok, G.-S. Lee, B.-S. An, E.-B. Jeung, K.-L. Park, *J. Reprod. Dev.* **2005**, *51*, 201.
[45] J. S. Yang, W. Qi, R. Farias-Pereira, S. Choi, J. M. Clark, D. Kim, Y. Park, *Food Chem. Toxicol.* **2019**, *125*, 595.
[46] J. Ravichandran, B. S. Karthikeyan, A. Samal, *Sci. Total Environ.* **2022**, *826*, 154112.
[47] N. Spînu, M. T. D. Cronin, J. Lao, A. Bal-Price, I. Campia, S. J. Enoch, J. C. Madden, L. Mora Lagares, M. Novič, D. Pamies, S. Scholz, D. L. Villeneuve, A. P. Worth, *Comput. Toxicol.* **2022**, *21*, 100206.
[48] T. Zhang, S. Wang, L. Li, A.n Zhu, Q.i Wang, *Sci. Total Environ.* **2022**, *824*, 153932.
[49] E. Arnesdotter, N. Spinu, J. Firman, D. Ebbrell, M. T. D. Cronin, T. Vanhaecke, M. Vinken, *Toxicology* **2021**, *459*, 152856.
[50] D. Knapen, M. M. Angrish, M. C. Fortin, I. Katsiadaki, M. Leonard, L. Margiotta-Casaluci, S. Munn, J. M. O'brien, N. Pollesch, L. C. Smith, X. Zhang, D. L. Villeneuve, *Environ. Toxicol. Chem.* **2018**, *37*, 1723.
[51] D. L. Villeneuve, M. M. Angrish, M. C. Fortin, I. Katsiadaki, M. Leonard, L. Margiotta-Casaluci, S. Munn, J. M. O'brien, N. L. Pollesch, L. C. Smith, X. Zhang, D. Knapen, *Environ. Toxicol. Chem.* **2018**, *37*, 1734.
[52] V. J. Thannickal, G. B. Toews, E. S. White, J. P. Lynch Iii, F. J. Martinez, *Annu. Rev. Med.* **2004**, *55*, 395.
[53] J. C. Bonner, *Toxicol. Pathol.* **2007**, *35*, 148.
[54] T. Wynn, *J. Pathol.* **2008**, *214*, 199.
[55] C. F. Mcgroder, D. Zhang, M. A. Choudhury, M. M. Salvatore, B. M. D'souza, E. A. Hoffman, Y. Wei, M. R. Baldwin, C. K. Garcia, *Thorax* **2021**, *76*, 1242.
[56] P. A. S. Kinaret, G. Del Giudice, D. Greco, *Nano Today* **2020**, *35*, 100945.
[57] B. J. Hama Amin, F. H. Kakamad, G. S. Ahmed, S. F. Ahmed, B. A. Abdulla, S. H. Mohammed, T. M. Mikael, R. Q. Salih, R. K. Ali, A. M. Salh, D. A. Hussein, *Ann. Med. Surg.* **2022**, *77*, 103590.
[58] L. A. Saarimäki, P. A. S. Kinaret, G. Scala, G. Del Giudice, A. Federico, A. Serra, D. Greco, *NanoImpact* **2020**, *20*, 100274.
[59] G. Scala, P. Kinaret, V. Marwah, J. Sund, V. Fortino, D. Greco, *NanoImpact* **2018**, *11*, 99.
[60] P. Kinaret, V. Marwah, V. Fortino, M. Ilves, H. Wolff, L. Ruokolainen, P. Auvinen, K. Savolainen, H. Alenius, D. Greco, *ACS Nano* **2017**, *11*, 3786.
[61] E. M. Rydman, M. Ilves, E. Vanhala, M. Vippola, M. Lehto, P. A. S. Kinaret, L. Pylkkänen, M. Happo, M.-R. Hirvonen, D. Greco, K. Savolainen, H. Wolff, H. Alenius, *Toxicol. Sci.* **2015**, *147*, 140.
[62] P. Kinaret, M. Ilves, V. Fortino, E. Rydman, P. Karisola, A. LÃ¤Hde, J. Koivisto, J. Jokiniemi, H. Wolff, K. Savolainen, D. Greco, H. Alenius, *ACS Nano* **2017**, *11*, 291.
[63] S. S. Poulsen, N. R. Jacobsen, S. Labib, D. Wu, M. Husain, A. Williams, J. P. BøGelund, O. Andersen, C. KøBler, K. Mølhave, Z. O. Kyjovska, A. T. Saber, H. Wallin, C. L. Yauk, U. Vogel, S. Halappanavar, *PLoS One* **2013**, *8*, e80452.
[64] L. Rahman, N. R. Jacobsen, S. A. Aziz, D. Wu, A. Williams, C. L. Yauk, P. White, H. Wallin, U. Vogel, S. Halappanavar, *Mutat. Res., Genet. Toxicol. Environ. Mutagen.* **2017**, *823*, 28.
[65] N. L. Guo, Y.-W. Wan, J. Denvir, D. W. Porter, M. Pacurari, M. G. Wolfarth, V. Castranova, Y. Qian, *J. Toxicol. Environ. Health, Part A* **2012**, *75*, 1129.
[66] G. Scala, V. Marwah, P. Kinaret, J. Sund, V. Fortino, D. Greco, *Data Brief* **2018**, *19*, 1046.
[67] L. A. Saarimäki, A. Federico, I. Lynch, A. G. Papadiamantis, A. Tsoumanis, G. Melagraki, A. Afantitis, A. Serra, D. Greco, *Sci. Data* **2021**, *8*, 49.
[68] M. Barbarino, A. Giordano, *Cancers (Basel)* **2021**, *13*, 1318.

 © 2022 The Authors. Advanced Science published by Wiley-VCH GmbH

[69] S. Rittinghausen, A. Hackbarth, O. Creutzenberg, H. Ernst, U. Heinrich, A. Leonhardt, D. Schaudien, *Part. Fibre Toxicol.* **2014**, *11*, 59.
[70] S. S. Gupta, K. P. Singh, S. Gupta, M. Dusinska, Q. Rahman, *Nanomaterials (Basel)* **2022**, *12*, 1708.
[71] M. S. P. Boyles, L. Young, D. M. Brown, L. Maccalman, H. Cowie, A. Moisala, F. Smail, P. J. W. Smith, L. Proudfoot, A. H. Windle, V. Stone, *Toxicol. In Vitro* **2015**, *29*, 1513.
[72] D. W. Porter, A. F. Hubbs, R. R. Mercer, N. Wu, M. G. Wolfarth, K. Sriram, S. Leonard, L. Battelli, D. Schwegler-Berry, S. Friend, *Toxicology* **2010**, *269*, 136.
[73] G. Izbicki, M. J. Segel, T. G. Christensen, M. W. Conner, R. Breuer, *Int. J. Exp. Pathol.* **2002**, *83*, 111.
[74] A. B. Hill, *Proc. R. Soc. Med.* **1965**, *58*, 295.
[75] K. M. Fedak, A. Bernal, Z. A. Capshaw, S. Gross, *Emerg. Themes Epidemiol.* **2015**, *12*, 14.
[76] X. Yue, B. Shan, J. A. Lasky, *Curr. Enzyme Inhib.* **2010**, *6*, 67.
[77] J. P. Annes, J. S. Munger, D. B. Rifkin, *J. Cell Sci.* **2003**, *116*, 217.
[78] P. Nummela, J. Lammi, J. Soikkeli, O. Saksela, P. Laakkonen, E. Hölttä, *Am. J. Pathol.* **2012**, *180*, 1663.
[79] X. Yan, Z. Liu, Y. Chen, *Acta Biochim. Biophys. Sin.* **2009**, *41*, 263.
[80] S. Bachy, Z. Wu, P. Gamradt, K. Thierry, P. Milani, J. Chlasta, A. Hennino, *iScience* **2022**, *25*, 103758.
[81] J. Hou, J. Shi, L. Chen, Z. Lv, X. Chen, H. Cao, Z. Xiang, X. Han, *Cell Commun. Signaling* **2018**, *16*, 89.
[82] H. Wang, Y. Gao, L. Wang, Y. Yu, J. Zhang, C. Liu, Y. Song, H. Xu, J. Wang, H. Lou, T. Dong, *J. Adv. Res.* **2022**, https://doi.org/10.1016/j.jare.2022.04.012.
[83] M. E. Hammond, G. R. Lapointe, P. H. Feucht, S. Hilt, C. A. Gallegos, C. A. Gordon, M. A. Giedlin, G. Mullenbach, P. Tekamp-Olson, *J. Immunol.* **1995**, *155*, 1428.
[84] K. De Filippo, A. Dudeck, M. Hasenberg, E. Nye, N. Van Rooijen, K. Hartmann, M. Gunzer, A. Roers, N. Hogg, *Blood* **2013**, *121*, 4930.
[85] J. W. Cheng, Z. Sadeghi, A. D. Levine, M. S. Penn, H. A. Von Recum, A. I. Caplan, A. Hijaz, *Cytokine* **2014**, *69*, 277.
[86] T. T. Braga, J. S. H. Agudelo, N. O. S. Camara, *Front. Immunol.* **2015**, *6*, 602.
[87] J. Dong, Q. Ma, *Nanotoxicology* **2018**, *12*, 153.
[88] P. A. S. Kinaret, G. Scala, A. Federico, J. Sund, D. Greco, *Small* **2020**, *16*, 1907609.
[89] J. Meng, X. Li, C. Wang, H. Guo, J. Liu, H. Xu, *ACS Appl. Mater. Interfaces* **2015**, *7*, 3180.
[90] S. Loffek, O. Schilling, C.-W. Franzke, *Eur. Respir. J.* **2011**, *38*, 191.
[91] A. Pardo, S. Cabrera, M. Maldonado, M. Selman, *Respir. Res.* **2016**, *17*, 23.
[92] V. J. Craig, L. Zhang, J. S. Hagood, C. A. Owen, *Am. J. Respir. Cell Mol. Biol.* **2015**, *53*, 585.
[93] G. Yu, E. Kovkarova-Naumovski, P. Jara, A. Parwani, D. Kass, V. Ruiz, C. Lopez-Otã-N, I. O. Rosas, K. F. Gibson, S. Cabrera, R. Ramã-Rez, S. A. Yousem, T. J. Richards, L. J. Chensny, M. Selman, N. Kaminski, A. Pardo, *Am. J. Respir. Crit. Care Med.* **2012**, *186*, 752.
[94] A. Federico, M. Fratello, G. Scala, L. Möbus, A. Pavel, G. Del Giudice, M. Ceccarelli, V. Costa, A. Ciccodicola, V. Fortino, A. Serra, D. Greco, *Cancers (Basel)* **2022**, *14*, 2043.
[95] A. Pavel, G. Del Giudice, A. Federico, A. Di Lieto, P. A. S. Kinaret, A. Serra, D. Greco, *Brief Bioinf.* **2021**, *22*, 1430.
[96] M. Martens, A. Ammar, A. Riutta, A. Waagmeester, D. N. Slenter, K. Hanspers, R. A. Â Miller, D. Digles, E. N. Lopes, F. Ehrhart, L. J. Dupuis, L. A. Winckers, S. L. Coort, E. L. Willighagen, C. T. Evelo, A. R. Pico, M. Kutmon, *Nucleic Acids Res.* **2021**, *49*, D613.
[97] S. Köhler, M. Gargano, N. Matentzoglu, L. C. Carmody, D. Lewis-Smith, N. A. Vasilevsky, D. Danis, G. Balagura, G. Baynam, A. M. Brower, T. J. Callahan, C. G. Chute, J. L. Est, P. D. Galer, S. Ganesan, M. Griese, M. Haimel, J. Pazmandi, M. Hanauer, N. L. Harris, M. J. Hartnett, M. Hastreiter, F. Hauck, Y. He, T. Jeske, H. Kearney, G. Kindle, C. Klein, K. Knoflach, R. Krause, et al., *Nucleic Acids Res.* **2021**, *49*, D1207.
[98] M. Kanehisa, *Nucleic Acids Res.* **2000**, *28*, 27.
[99] M. Gillespie, B. Jassal, R. Stephan, M. Milacic, K. Rothfels, A. Senff-Ribeiro, J. Griss, C. Sevilla, L. Matthews, C. Gong, C. Deng, T. Varusai, E. Ragueneau, Y. Haider, B. May, V. Shamovsky, J. Weiser, T. Brunson, N. Sanati, L. Beckman, X. Shao, A. Fabregat, K. Sidiropoulos, J. Murillo, G. Viteri, J. Cook, S. Shorser, G. Bader, E. Demir, C. Sander, et al., *Nucleic Acids Res.* **2022**, *50*, D687.
[100] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, *Nat. Genet.* **2000**, *25*, 25.
[101] F. Cunningham, J. E. Allen, J. Allen, J. Alvarez-Jarreta, M.â R. Amode, I. M. Armean, O. Austine-Orimoloye, A. G. Azov, I. Barnes, R. Bennett, A. Berry, J. Bhai, A. Bignell, K. Billis, S. Boddu, L. Brooks, M. Charkhchi, C. Cummins, L. Daâ Rinâ Fioretto, C. Davidson, K. Dodiya, S. Donaldson, B. Elâ Houdaigui, T. Elâ Naboulsi, R. Fatima, C. G. Giron, T. Genez, J. G. Martinez, C. Guijarro-Clarke, A. Gymer, et al., *Nucleic Acids Res.* **2022**, *50*, D988.
[102] J. Sund, P. Deceuninck, EURL ECVAM Library of Reference Chemicals. European Commission, Joint Research Centre (JRC), Brussels, Belgium **2021**, http://data.europa.eu/89h/92614229-d020-4d96-941c-c9604e525c9e
[103] S. Wegner, P. Browne, D. Dix, *Reprod. Toxicol.* **2016**, *65*, 402.
[104] N. C. Kleinstreuer, P. Ceger, E. D. Watt, M. Martin, K. Houck, P. Browne, R. S. Thomas, W. M. Casey, D. J. Dix, D. Allen, S. Sakamuru, M. Xia, R. Huang, R. Judson, *Chem. Res. Toxicol.* **2017**, *30*, 946.
[105] P. Browne, R. S. Judson, W. M. Casey, N. C. Kleinstreuer, R. S. Thomas, *Environ. Sci. Technol.* **2015**, *49*, 8804.
[106] R. De Matos Simoes, F. Emmert-Streib, *PLoS One* **2012**, *7*, e33624.
[107] M. Bastian, S. Heymann, M. Jacomy, *Proc. Int. AAAI Conf. on Web Soc. Media* **2009**, *3*, 361.
[108] G. Csárdi, T. Nepusz, *InterJournal, Complex Syst.* **2006**, *1695*, 1.
[109] A. Pavel, A. Serra, L. Cattelani, A. Federico, D. Greco, *Methods Mol. Biol.* **2022**, *2401*, 161.
[110] A. A. Hagberg, D. A. Schult, P. J. Swart, in *Proc. of the 7th Python in Science Conf*, (Eds: G. Varoquaux, T. Vaught, J. Millman), SciPy 2008, Pasadena, CA **2008**, pp. 11–15.
[111] M. G. Schimek, E. Budinskã¡, K. G. Kugler, V. Å Vendovã¡, J. Ding, S. Lin, *Stat. Appl. Genet. Mol. Biol.* **2015**, *14*, 311.
[112] A. Serra, L. A. Saarimã×Ki, M. Fratello, V. S. Marwah, D. Greco, *Bioinformatics* **2020**, *36*, 2932.
[113] K. J. Livak, T. D. Schmittgen, *Methods* **2001**, *25*, 402.

 © 2022 The Authors. Advanced Science published by Wiley-VCH GmbH