A Data Fusion Pipeline for Generating and Enriching Adverse Outcome Pathway Descriptions

OXFORD

SOT | Society of Toxicology
www.toxsci.oxfordjournals.org

ToxSci 20 Years

TOXICOLOGICAL SCIENCES, 162(1), 2018, 264–275

doi: 10.1093/toxsci/kfx252
Advance Access Publication Date: November 15, 2017
Research Article

A Data Fusion Pipeline for Generating and Enriching Adverse Outcome Pathway Descriptions

Penny Nymark,[*,†] Linda Rieswijk,[‡,§] Friederike Ehrhart,[‡] Nina Jeliazkova,[¶] Georgia Tsiliki,[‖,|||] Haralambos Sarimveis,[‖] Chris T. Evelo,[‡] Vesa Hongisto,[†] Pekka Kohonen,[*,†] Egon Willighagen,[‡,1] and Roland C. Grafström[*,†,1,2]

[*]Institute of Environmental Medicine, Karolinska Institutet, 171 77 Stockholm, Sweden; [†]Department of Toxicology, Misvik Biology, 20520 Turku, Finland; [‡]Department of Bioinformatics, NUTRIM, Maastricht University, 6200MD Maastricht, The Netherlands; [§]Division of Environmental Health Sciences, School of Public Health, University of California, 94720-7360 Berkeley, California, United States; [¶]IdeaConsult Ltd, 1000 Sofia, Bulgaria; [‖]School of Chemical Engineering, National Technical University of Athens, 157 80 Athens, Greece; and [|||]Institute for the Management of Information Systems, ATHENA Research and Innovation Centre, 151 25 Athens, Greece

1These authors contributed equally to this study.
[2]To whom correspondence should be addressed. E-mail: roland.grafstrom@ki.se.

ABSTRACT

Increasing amounts of systems toxicology data, including omics results, are becoming publically available and accessible in databases. Data-driven and informatics-tool supported pipeline schemas for fitting such data into Adverse Outcome Pathway (AOP) descriptions could potentially aid the development of nonanimal-based hazard and risk assessment methods. We devised a 6-step workflow that integrated diverse types of toxicology data into a novel AOP scheme for pulmonary fibrosis. Mining of literature references and diverse data sources covering previous pathway descriptions and molecular results were coupled in a stepwise manner with informatics tools applications that enabled gene linkage and pathway identification in molecular interaction maps. Ultimately, a network of functional elements coupled 64 pulmonary fibrosis-associated genes into a novel, open-source AOP-linked molecular pathway, now available for commenting and improvements in WikiPathways (WP3624). Applying *in silico*-based knowledge extraction and modeling, the pipeline enabled screening and fusion of many different complex data types, including the integration of omics results. Overall, the taken, stepwise approach should be generally useful to construct novel AOP descriptions as well as to enrich developing AOP descriptions in progress.

Key words: Adverse Outcome Pathways; pathway analysis; pulmonary fibrosis; predictive toxicology; bioinformatics; alternatives to animal testing.

Data-driven bioinformatics-based modeling is likely to capture the dynamic nature of biological processes and serve for hypothesis generation and mechanistic insight into pathobiological effects induced by toxic agents. Pathway analysis based on systems toxicology and omics data, and integration with phenotypical endpoint data, ultimately for generating toxicity predictive tools, can be expected to decrease the need for animal experiments in toxicity testing (Hartung, 2016; Thomas *et al.*, 2013). Helpful to this aim, diverse tools and curated molecular pathway templates, such as WikiPathways (Kutmon *et al.*, 2016), the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al.*, 2017), and Reactome (Fabregat *et al.*, 2016) permit visualization and computational analysis of omics data. Yet, the results from pathway analyses often require further

bioinformatics analysis and expert-based interpretation (Herwig *et al.*, 2016). For risk assessors, the absence of established protocols and need for expertise is an obstacle for using omics and other systems biology data in current risk assessment practice (Sturla *et al.*, 2014).

The concept of Adverse Outcome Pathways (AOP) enables incorporation of mechanistic toxicity data in risk assessment (Vinken, 2013). An AOP is a simplified schema of the central physiological events leading to a pathological state, and uses existing knowledge to causally connect a molecular initiating event (MIE) to key events (KE), associative events (AE), and an adverse outcome (AO; Vinken, 2013). The AOP-associated events can be linked to toxicological tests and assays, which provide a basis for integrated testing strategies predictive of adverse outcome (Burden *et al.*, 2015). For example, pathway analysis from transcriptomics data have been mapped to AOP-associated events and used as indicators of disease (Labib *et al.*, 2015; van der Veen *et al.*, 2014). Additionally, the Comparative Toxicogenomics Database (CTD) contains manually curated multi-species information from peer-reviewed literature on chemical/(nano)particle-gene-disease relationships reported from a diverse array of controlled laboratory experiments (Grondin *et al.*, 2016). Furthermore, life science databases and database tools such as GeneMANIA and ConsensusPathDB contain information on interactions between genes and their functional groupings (Kamburov *et al.*, 2013; Warde-Farley *et al.*, 2010). Coupled to the recent interest of regulatory authorities to broadly consider weight-of-evidence (WoE) data and new approach methodologies (NAMs) (ECHA, 2016), interactive and sequential application of such tools promises to facilitate the use of systems toxicology data in risk assessment work.

Here, we describe and test the idea of generating an ordered, potentially generalizable data fusion pipeline for applying diverse types of information and data collections for constructing AOP-linked molecular pathway descriptions. As case study we chose pulmonary fibrosis, a toxicology-relevant serious chronic respiratory disease induced by a wide variety of environmental and occupational exposures, including silica, asbestos, coal dust, and metals (Mossman and Churg, 1998; Todd *et al.*, 2012). Useful to the idea of enabling a "big data and informatics tools-driven" pipeline, the main key events underlying lung fibrosis are well-studied, involving oxidative stress, TGF-β signaling, disturbed extracellular matrix (ECM) production/degradation, and inflammatory processes (Labib *et al.* 2015; Vietti et al., 2016). Furthermore, studies applying nanomaterials and the generation of omics data have added dimensions to the knowledge on the mechanisms of pulmonary fibrosis (Clippinger *et al.*, 2016; Dong and Ma, 2015; Dymacek *et al.*, 2015; Poulsen *et al.*, 2013; Snyder-Talkington *et al.*, 2013, 2015, 2016). On these premises, we built a knowledge mining and informatics-dependent workflow for capturing existing knowledge from literature and databases to a potentially risk assessment work-applicable AOP description. Finally, we demonstrate significant outcome of the approach from applying the substantiated AOP-description to an independent study of nanomaterial-induced pulmonary fibrosis.

MATERIALS AND METHODS

Mining of literature. Pulmonary fibrosis is a well-studied human disease and the cumulative number of publications between 1975 and 2010 reached beyond 17 000 papers, as reviewed by Todd *et al.* (2012), citing over 300 papers. We focused our search for general schemes of pulmonary fibrosis development and progression to recent reviews, including Todd *et al.* (2012), as well as 6 other reviews (Boyles *et al.*, 2014; Kinnula *et al.*, 2005; Manresa *et al.*, 2014; Miller *et al.*, 2014; Rockey *et al.*, 2015; Vancheri, 2013). In addition, 2 reviews relevant to AOP development for pulmonary fibrosis were identified (Labib *et al.* 2015; Vietti et al. 2016).

Database searching and bioinformatics processing. The widely cited databases Comparative Toxicogenomics Database (CTD; Davis *et al.*, 2017), Ingenuity Pathway Analysis (IPA; Krämer *et al.*, 2014), Online Mendelian Inheritance in Man (OMIM; Amberger *et al.*, 2015), Causal Biological Network Database (CBN; Boué *et al.*, 2015), KEGG (Kanehisa *et al.* 2017), and Reactome (Fabregat *et al.* 2016) were found to represent pulmonary fibrosis and were further explored for information associated with the disease.

The online interaction prediction software GeneMANIA (Warde-Farley *et al.* 2010) was used to identify genetic and physical interaction links (based on curated experimentally validated interactions from the Biological General Repository for Interaction Datasets) between pulmonary fibrosis-associated genes selected for further development. The network was visualized using Cytoscape v3.4.0 (Shannon et al. 2013), by importing the tab-delimited GeneMANIA network file directly and applying various basic Cytoscape features, such as the circular layout option. The interaction information was also used to define groups of genes, eg, genes showing interactions with 1 main key interactor were grouped together. The overrepresentation analysis tool in ConsensusPathDB (Kamburov *et al.* 2013) was used to analyze for pathway enrichment among the total set of genes selected for further analysis and for groups of genes showing strong interactions. WikiPathways, KEGG, and Gene Ontologies level 5 Biological Processes (with a minimum overlap of 2 genes with the input list and p value cutoff at .01) were prioritized as source databases, in that order. The top most significant pathway(s) were assigned to the gene groups and used to draw conclusions on the genes' involvement in molecular processes.

Pathway construction in PathVisio. PathVisio was used to draw an interactive biological/molecular pathway in a standard GPML format, which includes data integration (via the BridgeDb framework; van Iersel *et al.*, 2010) and enables use of cross database identifiers, such as MIM (Molecular Interaction Maps) standard interactions, external references (eg, PubMed), and annotations (Kutmon *et al.*, 2015; van Iersel *et al.*, 2008). The pulmonary fibrosis pathway was designed by creating basic relationships between the cause (oxidative stress, inflammation, and coagulation disturbances) and effect (pulmonary fibrosis). Gene lists of interest, selected for further development, were imported as Human Ensembl IDs using the MAPPBuilder add-on (http://www.pathvisio.org/plugin/mappbuilder-plugin/). Information from the database search described earlier was then used to build connections, interactions, and groupings between the different elements of the pathway (genes, pathways, and general scheme elements) using MIM standards. In addition, where relevant, computational links to corresponding pathways in the WikiPathways and KEGG databases were generated, by using database specific pathway identifiers. A tutorial on the technical aspects of building a pathway can be accessed through the eNanoMapper website (http://www.enanomapper.net/enm-tutorials) or FigShare (Ehrhart *et al.*, 2017).

Linkage to an AOP. A putative AOP for pulmonary fibrosis and an associated transcriptomics data analysis (Labib *et al.* 2015) were used to assign associated AOP events (MIEs, KEs, and AEs) to the

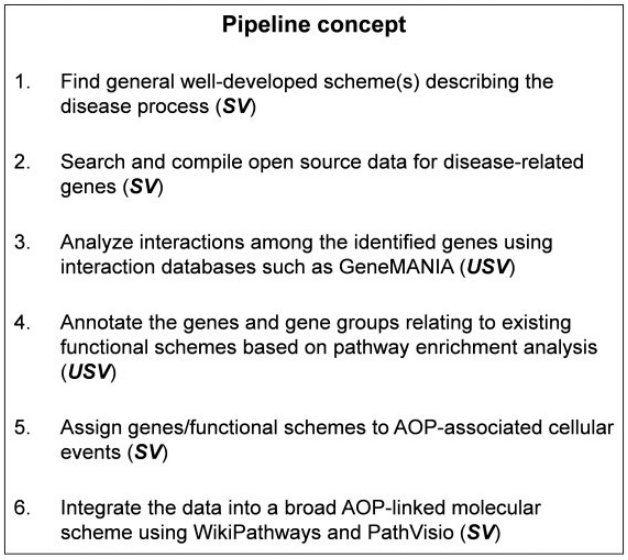

Figure 1. The pipeline concept for constructing a bioinformatics employable Adverse Outcome Pathway (AOP)-linked molecular pathway description based on literature, databases, and bioinformatics-supported tools. The workflow involves initially supervised (SV) information review followed by unsupervised (USV) informatics-based processing and fusion of the results, and finally supervised association and visualization in relation to existing AOP descriptions.

previously identified genes and gene groups in the initial scheme of the pathway. A description of each event (MIE, KE, AE, and AO), as described in the original paper (Labib *et al.* 2015) was also added as a legend next to the pathway itself to allow for easy linkage to the cellular processes described for each event.

Independent data analysis. Publically available transcriptomics data from a study on the effects of multi-walled carbon nanotubes (Mitsui MWCNT-7) at 3 doses (18, 54, and 162 μg per mouse, through intratracheal instillation) on the lungs of C57BL/6 mice (GSE46998) was retrieved from the Gene Expression Omnibus database (https://www.ncbi.nlm.nih.gov/geo/). Differentially expressed genes were obtained from Supplementary Data files in the original paper (Poulsen *et al.* 2013). Ensembl IDs and fold changes were imported into PathVisio and visualized with color-codes in the mouse version of the newly constructed pathway (WP3632).

RESULTS

Establishing a Workflow for the Construction of an AOP-Linked Molecular Pathway

Based on identified sources of knowledge and data, a 6-step workflow concept for the construction of an AOP-linked molecular pathway was established (Figure 1). The pipeline was tested to build a data fusion-based pathway for pulmonary fibrosis and the steps were applied as follows:

1. General scheme describing pulmonary fibrosis

The review by Todd *et al.* (2012) provided a suitable recent overview of pulmonary fibrosis as a basis for the pathway, listing >300 references and providing a general scheme of the development and progression of the disease, with linkage to the most central molecular events. In addition, the review by Vietti *et al.* (2016) was also applied to generate the initial scheme, due to its

A

	CBN	IPA	CTD	OMIM	KEGG	Reactome
CBN	99	IPA				
IPA	13	91	CTD			
CTD	16	15	64	OMIM		
OMIM	1	5	4	12	KEGG	
KEGG	0	6	6	3	6	Reactome
Reactome	0	1	1	1	1	1
Total overlap with the other databases	24/99	24/91	26/64	5/12	6/6	1/1

B

Database(s)	Total nr of genes	Gene names
CTD IPA KEGG OMIM Reactome	1	SFTPA2
CTD IPA KEGG OMIM	2	TERT, SFTPA1
CBN CTD IPA OMIM	1	TGFB1
CBN CTD IPA	4	TNF, IL13, SPP1, IL6
CTD IPA KEGG	3	SFTPC, MUC5B, TERC
CBN CTD	11	IL4, FGF2, EGF, CEBPB, SMAD7, ELN, TIMP1, CCL2, TGFA, CTGF, CXCL8
CBN IPA	8	YY1, AGTR1, COL3A1, RETNLB, FGFR1, PPARG, IL10, SRC
CTD IPA	4	RTEL1, IGF1, PLAU, PARN
IPA OMIM	1	C5
CBN	75	ACE, BMP4, LGALS3, WNT3A, CDH1, ROCK1, MIR29B1, FOS, ACTA2, MKL1, JUNB, DCN, COL4A1, VCAN, ILK, MAPK14, WNT1, ITGB8, COL4A2, TGIF1, WNT5B, ATF2, SMAD4, TGFBR2, STAT6, MAPK8, SMAD2, PTCH1, TGFBR3, EFEMP2, F2R, CCL18, THBS1, JUN, DVL3, MAPK3, VIM, ETS2, WNT10B, WNT7B, PLAUR, AGTR2, BGN, SHH, S1PR2, HSPG2, AKT1, APCS, GSK3B, SMO, COL1A2, MAPK1, RHOA, GLI1, F2, COL1A1, SRF, FZD8, WISP1, SMAD3, FZD6, PTGS2, OLR1, MYOCD, EGR1, CTNNB1, EGFR, MIR15B, TGFBR1, SP1, ITGB3, SMURF2, NOX4, FN1, DKK1
IPA	67	IL17A, PPARD, AGO1, HCK, NFKBIA, DDR2, IFNG, DDR1, SERPINE1, TSLP, TGFB2, WBSCR22, KDR, HS6ST1, CD46, PDGFRB, IFNGR1, PIK3CB, FLT1, FAM111B, LCK, CRLF2, CDH11, SLC7A11, PDE5A, IL23A, EDNRA, ICAM1, AGO2, CSF1R, HPRT1, PTGIR, ABL1, CRBN, FGFR2, FLT4, MMP8, FGFR3, KIT, LYN, EDNRB, IL6R, S100A4, PDGFRA, IL25, SPHK1, CD40LG, CD55, IL17RA, RORC, GSS, IFNGR2, RET, LPAR2, AGER, VHL, NFKB1, HMGA2, NPPB, P2RY12, PTGER1, CXCL17, FOXM1, FLT3, NR3C1, C3, FAS
CTD	38	MMP2, MT2, CYSLTR2, PDGFA, EDN1, CCL11, CCL5, PTX3, OBFC1, HMOX1, CCR2, CMA1, GREM1, DSP, CCL3, PDGFB, ELMOD2, CSF3, IL5, MECP2, FGF1, DPP9, CALCA, HGF, SERPINA1, CCR3, BMP7, FGF7, CXCL2, IL1B, FAM13A, CSF2, CCL4, NFE2L2, ATP11A, IL12B, MMP9, SKIL
OMIM	7	FEOM3A, CFM1, SM2, FEOM3, PHOX2A, KIF21A, CFTR

Figure 2. Overlap between pulmonary fibrosis-associated genes in 6 databases; CBN—Causal Biological Network Database, IPA—Ingenuity Pathway Analysis, CTD—Comparative Toxicogenomics Database, OMIM—Online Mendelian Inheritance in Man, KEGG—Kyoto Encyclopedia of Genes and Genomes, and Reactome. A, Numbers of genes found in pairwise comparisons of the databases. B, Sorted table with overlapping genes identified in the databases. See Supplementary Table 1 for more details.

description of alterations on molecular level related to pulmonary fibrosis or fibrogenic changes. Central signaling pathways and genes identified in the reviews included MAPK and NFKB signaling, inflammasome activation, *TNFA*, *IL1B*, *IL8* (*CXCL8*), *CCL2*, *CCL3*, *IL4*, *IL13*, *CTGF*, *IGF*, *TGFB1*, and *PDGF* and *SMAD* genes.

2. Assessment of databases with information on genes involved in pulmonary fibrosis

Pulmonary fibrosis was represented by variable numbers of genes (range: 1–99) in CBN, IPA, CTD, OMIM, KEGG, and Reactome (February 2016; Figure 2; Supplementary Table 1). CBN provided an unstructured network of molecules causally associated with pulmonary fibrosis, as determined from literature, including 99 genes. IPA featured a set of 91 pulmonary fibrosis-related genes, although without description of the underlying origin of the gene connections. CTD featured 64 genes with direct literature evidence of association (as opposed to predicted) to pulmonary fibrosis (Supplementary Table 1). In the database access point Enrichr (Kuleshov *et al.* 2016; http://amp.pharm.mssm.edu/Enrichr/), OMIM listed 12 tissue-independent genes associated with fibrosis. KEGG provided links to other biological pathways that relate to idiopathic (unknown origin) pulmonary fibrosis (KEGG disease pathway H01299) and listed 6 genes commonly affected by mutations in familial pulmonary fibrosis, including references. Reactome included a pathway entitled "Defective SFTPA2 causes idiopathic pulmonary fibrosis (IPF)" involving only the 1 gene—*SFTPA2*. None of these databases provide a structured pathway or network of the genes.

CTD showed a large overlap (41%, 26/64 genes) with other databases (Figure 2A; Supplementary Table 1), involving 12 of the central genes and gene families as determined by the

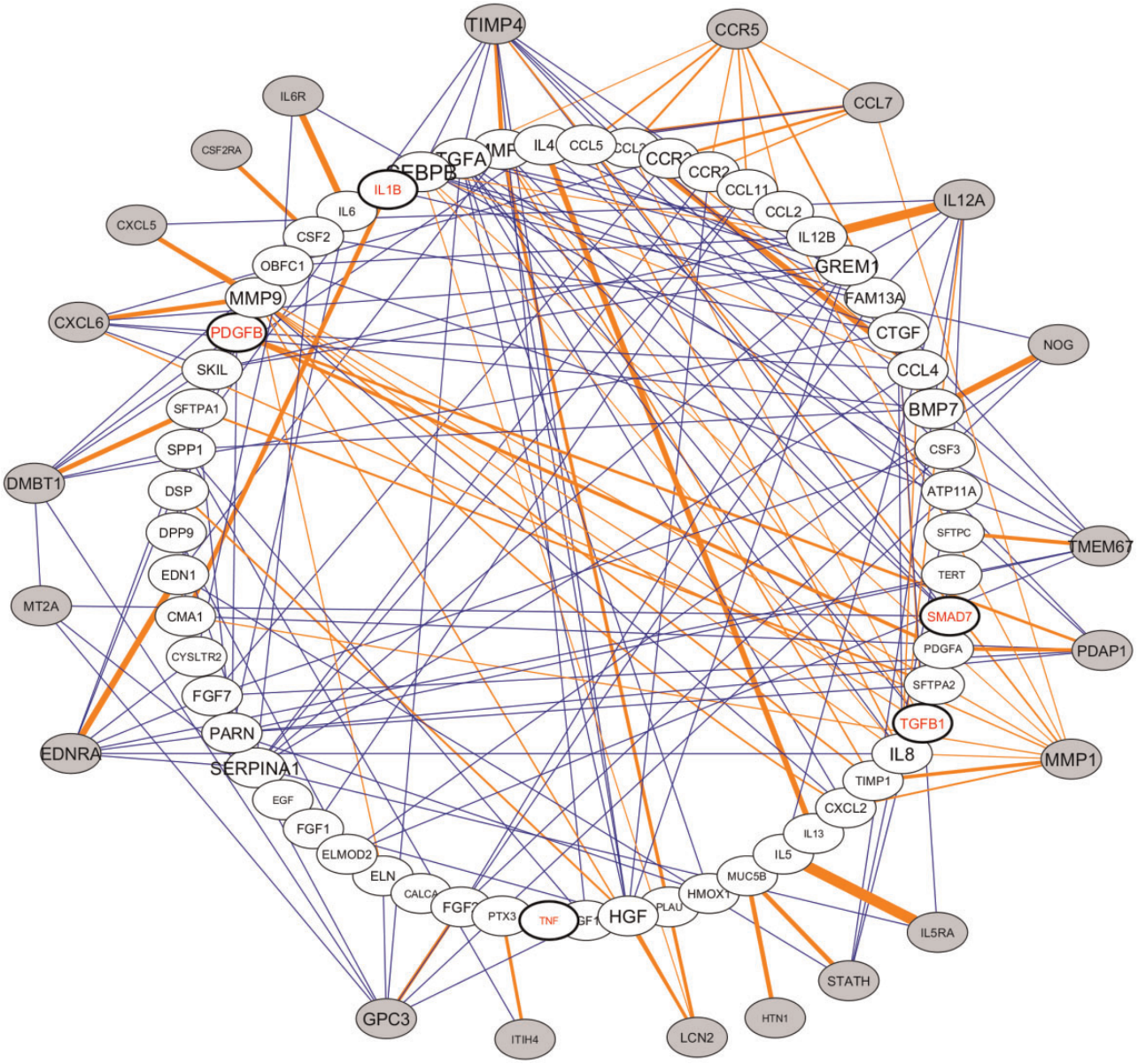

Figure 3. Interaction network between 59 of the 64 CTD genes associated with pulmonary fibrosis, based on analysis in GeneMANIA. Genetic (blue online) and physical (orange online) interactions, based on experimental evidence, are shown. The width of the edges (lines) reflects the frequency that the interaction is described in the data sources used by GeneMANIA. White nodes correspond to the pulmonary fibrosis associated genes and gray nodes are intermediate genes added by the software based on functional similarity (these genes were not included in the pathway). The font size of the gene name indicates the relative number of interactions with other genes in the network (range: 1–14 interactions). Central genes identified based on literature reviews and used as focus genes for connections in the novel pathway are indicated by red font (online). Refer to Supplementary Table 3 for details on the interactions.

literature search described earlier (Todd *et al.* 2012; Vietti et al. 2016). None of the other databases showed such coverage of the literature-based genes. CTD also provided information on 53 chemicals/(nano)particles (with direct evidence) and 51 pathways (≥3 genes involved) linked to pulmonary fibrosis (Supplementary Table 2). The top 15 (inference score >30) indications included asbestos (both serpentine and amphibole), bleomycin, particulate matter (coal ash, dust, smog, and smoke), lipopolysaccharides, cobalt, mustard gas, didecyldimethylammonium, sulfasalazine, paraquat, titanium dioxide, and metformin. Carbon nanotubes were listed among the top 25 agents (inference score 14.68) with direct evidence of correlation with pulmonary fibrosis. Comparing databases, the 4 most common genes, found in 4 to 5 of the 6 databases were *SFTPA2*, *SFTPA1*, *TGFB1*, and *TERT* (Figure 2B). Thirty-five genes were indicated in 2 or more databases (Figure 2B). Due to the richness of evidence and involvement indications, the further AOP-enrichment effort for pulmonary fibrosis was based on adding outstanding CTD-derived genes, ie, 52 of 64 genes without pulmonary fibrosis/AOP association according to previous review work.

3. *Interactions between genes associated with pulmonary fibrosis*

Using GeneMANIA, genetic and physical interactions were identified between 59 of the 64 CTD-derived genes (92%) (Figure 3 and Supplementary Table 3). Genes were grouped according to

Table 1. Relationships Between AOP-Related Events, Biological Pathways Identified From Omics Data (Labib *et al.*, 2015) and Genes/Gene Groups (Described by Pathways) Independently Identified in the Pulmonary Fibrosis Pathway

	Event	Description (Labib *et al.*, 2015)	Labib *et al.*, 2015 (Omics Data)[a]	Pulmonary Fibrosis Pathway[b] (WP3624)	No. of Genes
Cellular responses	**MIE**	*Cellular sensing/danger signals*	Toll-like receptor signaling	Toll-like receptor signaling (WP75)	4
	KE1	*Induction of inflammatory cytokines/acute phase response*	Chemokine signaling	Chemokine signaling (KEGG hsa04062)	14
			IL1R1-regulated signaling (*IL1B*, *IL6*, CXCL-1, *CCL2*, *CCL5*, and *SAA3*)	Cytokines and inflammatory response (WP530) *IL1B*	
			NFKB signaling	NFKB pathway (KEGG hsa04064)	
	KE2	*Chronic/persistant inflammation*	NFKB signaling	NFKB pathway (KEGG hsa04064)	–
	KE3	*Activation of TH2/M2 cells/interleukin/growth factor signaling*	TGFB signaling	*TGFB1* TGFB signaling (WP560) Chondrocyte differentiation (WP474) Matrix metalloproteinases (WP129) *PDGFB*	31
			HGF signaling	Differentiation pathway (WP2848)	
			IL-4 signaling	Cytokines and inflammatory response (WP530)	
			Chemokine signaling	Chemokine signaling (KEGG hsa04062)	
			T Helper Cell Differentiation	Leukocyte/Myeloid cell differentiation (GO: 0045637/GO: 1902105)	
	KE4	*Fibroblast/myofibroblast proliferation*	ERK/MAPK Signaling	MAPK signaling (WP382) p38 MAPK (WP400)	12
			TGFB signaling	TGFB signaling (WP560) *TGFB1* Chondrocyte differentiation (WP474) Differentiation pathway (WP2848) *PDGFB*	
	AE1	*Oxidative stress*	NRF2-mediated Oxidative Stress Response	NRF2 signaling (WP2884)	7
			Production of Nitric Oxide and Reactive Oxygen Species in Macrophages	ROS production (CHEBI: 26523) Regulation of reactive oxygen species metabolism (GO: 2000377)	
	AE2	*Cellular death*	Apoptosis Signaling	Apoptosis (WP254)	–
Organ responses	**KE5**	*ECM deposition*	Hepatic Fibrosis/Hepatic Stellate Cell Activation	Excess deposition and/or reduced turnover of ECM	–
	AE3	*Tissue injury*	HIF1A signaling	Collagen production	–
	AO	*Pulmonary fibrosis*	–	Pulmonary fibrosis	–

Some of the 64 genes are involved in > 1 event.
[a]Based on canonical pathways in IPA.
[b]Identified based on the 64 CTD-derived genes and independent of the original study by Labib *et al.* (2015).

key effector genes, such as *CCL2*, *TGFB1*, *PDGFB* as identified from literature, and *CEBPB* and *SERPINA1* based on their high numbers of interactor genes, 14 and 10, respectively (Supplementary Table 3). The process of grouping the genes was an interactive expert-based task applying both the gene-interaction and pathway enrichment analyses (see below).

4. *Functional schemes among the genes associated with pulmonary fibrosis*

The genes were initially grouped based on enrichment analysis of the complete set of 64 pulmonary fibrosis associated genes, where the most significantly enriched pathways were "Cytokines and Inflammatory Response" (WP530; 9/64 genes; *p* value 9.10E-25) and "Differentiation Pathway" (WP2848; 6/64 genes, *p* value 5.70E-10). The genes in both of these pathways were grouped and computationally linked to their respective WikiPathway IDs. In addition, groups of genes generated based on the interactions, as described earlier, were further analyzed separately for enriched pathways (WikiPathways and KEGG) and Gene Ontologies (GO). The most significant pathway or in case of no enriched pathways, GO biological process, was assigned to the gene group.

5. *Assignment of genes/gene groups to AOP-associated events for pulmonary fibrosis*

The next step in the workflow involved linkage of the genes/gene groups independently identified in the previous steps to a recently published putative AOP on pulmonary fibrosis (Labib *et al.* 2015). Labib *et al.* (2015) used transcriptomics data from nanomaterial-exposed mice that developed pulmonary fibrosis, and correlated pathway enrichment analysis results with MIEs, KEs, and AEs based on expert judgement. These correlations were used in the current study to facilitate the linking of AOP-related events to the components (genes/gene groups associated with molecular processes/pathways) of the pulmonary fibrosis pathway. All cellular events, ie, the MIE,

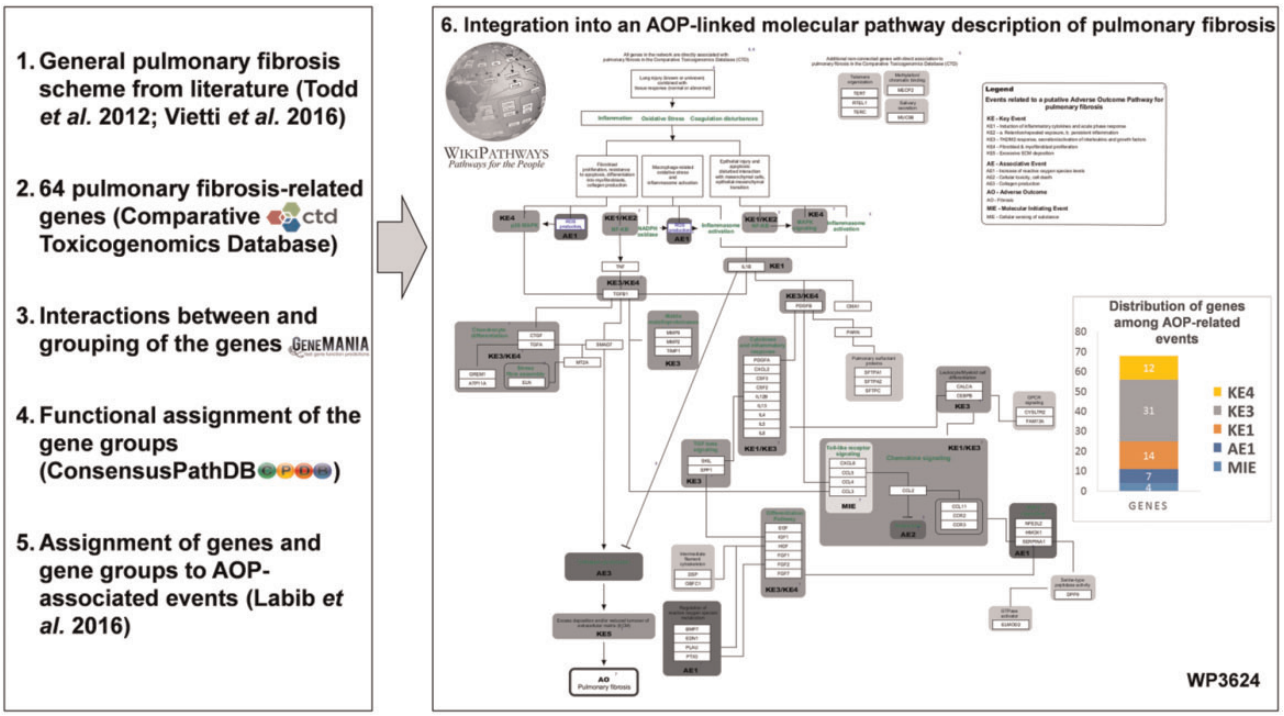

Figure 4. The novel constructed bioinformatically employable AOP-linked molecular pathway (human version WP3624) describing pulmonary fibrosis. The pathway is openly accessible at http://www.wikipathways.org/instance/WP3624.

KEs 1–4, and AEs 1–2 could be readily mapped and associated directly to the components of the pathway (Table 1). In addition, the tissue events KE5, AE3, and the final AO were also assigned to general scheme elements of the pathway.

6. Integrate the data into a broad AOP-linked molecular scheme for pulmonary fibrosis

An initial scheme of the disease was drawn in PathVisio, based on step 1 in the workflow, and included the following events: Lung injury and tissue response followed by inflammation, oxidative stress, and coagulation disturbances leading to (1) fibroblast proliferation, resistance to apoptosis, differentiation into myofibroblasts, and collagen production, (2) macrophage-related oxidative stress and inflammasome activation, and (3) epithelial injury and apoptosis, disturbed interaction with mesenchymal cells, and epithelial-mesenchymal transition (Todd *et al.* 2012; Vietti et al. 2016). The effects in the 3 main cell types together lead to excess deposition and/or reduced turnover of ECM and consequently to pulmonary fibrosis.

Based on step 2 in the workflow, the 64 genes from CTD were brought into PathVisio and listed beside the general scheme. Connections between interacting genes were drawn starting with the most central genes as assessed by the reviews, ie, *TNF*, *TGFB1*, *IL1B*, *SMAD7*, and *PDGFB* (Todd *et al.* 2012; Vietti et al. 2016). The connections were simplified and preferably drawn between groups of genes defined through key interactors, as described earlier in steps 3 and 4. Lines in the pathway that reach all the way to a gene within a group indicate interaction with that specific gene, whereas lines connected to a box surrounding a group of genes indicate interaction with all or most of the genes in the group. Genes within groups in the pathway often also show interactions among each other, which are not shown directly, but can be explored by clicking on the associated pathway (the titles of the gene groups displayed in green when exploring the pathway in PathVisio), eg, "Cytokines and inflammatory response" (WP530). Clicking leads the user to a pathway describing the related molecular process constructed by other researchers in the community. Unconnected genes (as determined by the GeneMANIA results) were listed at the top of the scheme. The precise interactions between all genes can also be explored in Supplementary Table 3.

At this point, the workflow had enabled generation of a pathway that was hypothesized to encompass available, significant information on the molecular representation of pulmonary fibrosis. The upper elements of the pathway represent the general scheme of pulmonary fibrosis, ie, the main cellular changes involved in the disease progression, whereas the elements below represent the underlying genes/mechanisms leading to the final adverse outcome included at the bottom of the pathway (Figure 4).

The next independent integration step, based on step 5 in the workflow, included the linkage between genes and gene groups in the pathway to AOP-related events. The relationships between the events and omics-based pathways, described and identified by Labib *et al.* (2015), and the genes/groups of genes (related to pathways) identified in the pulmonary fibrosis pathway are shown in Table 1 and described in detail in the following.

MIE for Pulmonary Injury Leading to Fibrosis. In Labib *et al.* (2015), the MIE is described as "cellular sensing of substance" and is linked to pathogen associated and endocytosis signaling pathways, such as Toll-like receptor signaling and caveolar/clathrin-mediated endocytosis signaling. In the pulmonary fibrosis pathway (constructed based on steps 1–4 in the workflow), the MIE was associated with the 4 genes, *CXCL8*, *CCL3*, *CCL4*, and *CCL5*, which had been grouped according to their membership in the pathway "Toll-like Receptor Signaling" in step 4 (WP75; Table 1).

The MIE within the pulmonary fibrosis pathway is directly linked with KE1 (inflammatory response), KE3 (activation of interleukins and growth factors), and KE4 ([myo]fibroblast activation), as well as with AE2 (cellular death) demonstrating the dynamic network with several feedback loops involved (Labib *et al.* 2015).

Cellular Responses for Pulmonary Injury Leading to Fibrosis. KE 1–4. KE1 is described as "induction of inflammatory cytokines and acute phase response," mainly through IL-1R1-mediated signaling, involving *IL1B*, *IL6*, and *CCL2* (Labib *et al.* 2015). In the current pathway, KE1 is associated with *IL1B*, the NFKB pathway (KEGG hsa04064), Cytokine and inflammatory response (WP530) including the genes *PDGFA*, *CXCL2*, *CSF3*, *CSF2*, *IL12B*, *IL13*, *IL4*, *IL5*, *IL6*, and Chemokine signaling (WP3929) with the genes *CCL2*, *CCL11*, *CCR2*, *CCR3*. In total, KE1 involves 14 genes (Table 1). KE2 occurs in case of "retention or repeated exposure and/or by persistent inflammation" and is also associated with the NFKB pathway. KE2 is not represented by genes in the current pathway, because none of the the genes chosen for constructing the pathway (the 64 CTD genes) were directly related to NFKB signaling. KE3 is described as "activation of T helper type 2 cells and M2 macrophages, as well as interleukin and growth factor signaling," and was associated with *PDGFB*, *TGFB1*, TGFB signaling (WP366), 3 cell differentiation pathways (the general Differentiation pathway [WP2848], the Chondrocyte differentiation [WP474—Endochondrial ossification] and Leukocyte/myeloid cell differentiation [GO: 0045637/GO: 1902105]), matrix metalloproteinases (WP129) and cytokine/chemokine signaling (WP530 and KEGG hsa04062). In total, KE3 is represented by 31 genes (Table 1). KE4 is described as "fibroblast and myofibroblast proliferation" and is related to MAPK signaling, growth, and cell cycle regulation according to Labib *et al.* (2015), and thus associated with 2 MAPK-related pathways (WP382 and WP400), 2 of the differentiation pathways described earlier, as well as *PDGFB*, *TGFB1*, and TGFB signaling. In total, KE4 is represented by 12 genes (Table 1).

AE 1–2. AE1, described as "oxidative stress" by Labib *et al.* (2015), is mapped to the genes *BMP7*, *EDN1*, *PLAU*, *PTX3*, (involved in Regulation of reactive oxygen species metabolism—GO: 2000377), *NFE2L2*, *HMOX1*, and *SERPINA1* (involved in NRF2 signaling—WP2884; Table 1). In total, AE1 is represented by 7 genes (Table 1). AE2, described as cellular death, is mapped to the Apoptosis pathway (WP254), but is not represented by genes in the pathway.

Organ Responses for Pulmonary Injury Leading to Fibrosis. KE 5, AE 3, and the AO. The organ response events AE3, KE5, and the final AO are not directly associated with genes in the pathway, but nevertheless linked to the pathway as general scheme elements. AE3, ie, "tissue injury" as a consequence of repeated exposure and persistent inflammation (KE2) leading to cytokine/growth factor signaling (KE3), was linked to HIF1A signaling by Labib *et al.* (2015). HIF1A signaling promotes formation of ECM through activation of collagen genes (Gilkes *et al.*, 2013), which led us to assign AE3 to Collagen production (WP2798). KE5 is described as "excessive ECM deposition and/or a reduced turnover," due to repeated cycles of the preceding KEs 1–4 and thus an imbalance in the metabolism of ECM proteins, including collagen. In the new pathway, *IL1B* expression, involved in the initial inflammatory response (KE1), may inhibit collagen production (AE3), as described by Todd *et al.* (2012). However, collagen production might be largely mediated by *TGFB1* and TGFB signaling (involved in KE3 and 4) and the effects of IL1B are then overruled (Todd *et al.* 2012). The final AO is described as pulmonary fibrosis.

The resultant pathway has been made openly available for use in the WikiPathways database under CC0 license (Figure 4; WP3624). The complete history of the different construction steps can be viewed in the database, providing transparency and additional guidance for future building of pathways related to other diseases/AOPs. In addition, import of data and export of the pathway, or gene list, is possible in several formats and can be extended by using additional plugins in, eg, PathVisio (Kutmon *et al.* 2015). The pathway has also been converted into a mouse variant (WP3632). Furthermore, in the interest of the systems biology research on nanomaterials, the pathway has been included in the newly established WikiPathways portal for nanomaterial-related or -specific pathways to facilitate community-based annotation of biological/molecular pathways related to nanosafety (http://nanomaterials.wikipathways.org/).

Visualization of Data in the Newly Constructed Pathway for Pulmonary Fibrosis

A transcriptomics data set representing the pulmonary effects of MWCNT-exposure was chosen to demonstrate, in independent data, the usefulness of the pipeline-derived novel AOP description. Fold-changes for significantly differentially expressed genes (FDR adjusted $p < .05$) were obtained from the original study; 773 for the low dose, 120 for the mid dose, and 1712 for the high dose (Poulsen *et al.* 2013). PathVisio analysis of the data using the newly constructed pathway showed, in total, differential regulation of 14 of the 64 pulmonary fibrosis-associated genes (Figure 5). Two of four genes (50%) associate with the MIE, 3 of 7 (43%) with the AE1, 4 of 14 (29%) genes with the KE1, 5 of 31 (16%) genes relate to KE3, and 1 of 12 genes (8%) relate to KE4. Based on the pipeline starting point, ie, the reviews by Todd *et al.* (2012) and Vietti et al. (2016), 11 of overall 14 genes would have gone undetected as pulmonary fibrosis AOP-associated, including *CCL5*, *CCL11*, and *IL6*, which have been found at increased levels in bronchioalveolar lavage samples from pulmonary fibrosis patients (Emad and Emad, 2007a,b).

DISCUSSION

A pipeline to generate AOP-linked molecular pathway descriptions in support of WoE biological data and NAMs was developed under a 6-step protocol (cf. Figure 1). The workflow is both data-driven and informatics- and visualization-tools dependent, involving initially supervised information review followed subsequently by unsupervised informatics-based processing and fusion of the combined results into molecular interaction networks. The networks were then in turn assigned and plotted relating to earlier AOP descriptions for pulmonary fibrosis, leading ultimately to a further substantiated AOP description. Assessment in independent data indicated that the approach successfully had enriched the AOP description with MIE and KE-related molecular data. Relevant expertise is naturally a key dimension to the work process, yet the tools, and the suggested order of applying the tools generate a structured means of combining supervised and unsupervised data management (cf. Figure 1). We consider the composite pipeline concept therefore to have strong potential to aid the many AOP descriptions that are known for being in current development phase.

The pathways and AOP descriptions generated from following the workflow are currently openly available for inspection

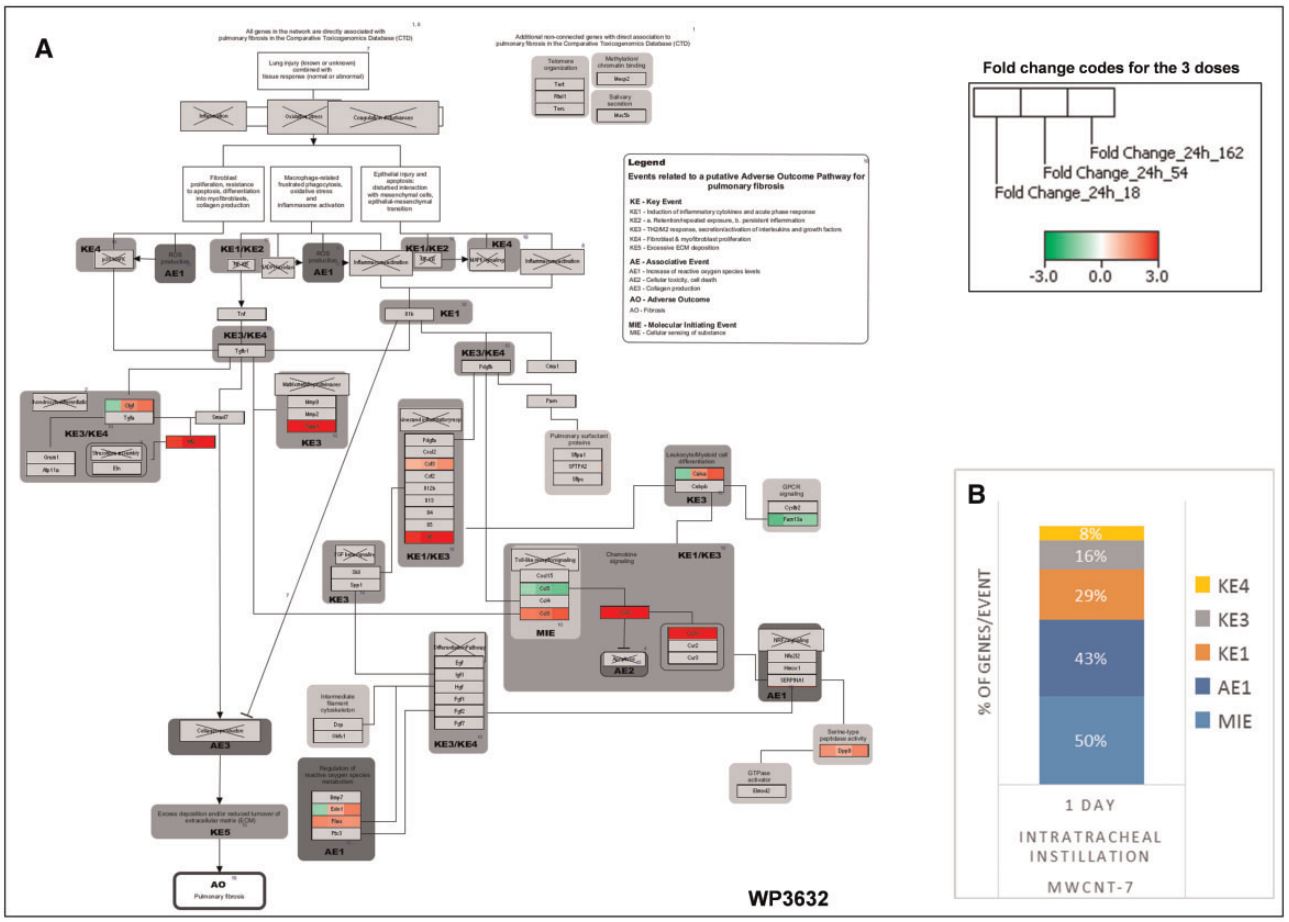

Figure 5. Visualization of a transcriptomics data set, representing the effects of MWCNT exposure (at 3 doses) on the lungs after 1 day, in the AOP-linked molecular pathway for pulmonary fibrosis (mouse version WP3632). Red and green (online) indicates up- and down-regulation, respectively, and the regulation can be seen for each dose separately, as indicated by the diagram at the upper right corner (A). Deregulation of all represented AOP-related events can be seen, including the MIE, AE1, KE1, KE3, and KE4 (B). Crossed boxes indicate associated pathways, which can be explored in relation to the analyzed data by clicking (in PathVisio).

and input, with interactive schemes useful for both intuitive visualization and computational analysis of diverse omics data. The pathways are applicable as multivariate biomarkers for hypothesizing and/or predicting the final adverse outcomes related to specific exposures. They allow a user to dig deeper into the underlying mechanisms through links with associated knowledge (genes/pathways) and thus, provides transparency with regard to the information that they build upon. The linked data in such pathways enable structured computational analyses useful for integration of diverse data, which is common within the multidisciplinary field of toxicology and particularly within risk assessment. Linked data also allow for informed semantic querying in databases, eg, in this case for looking up agents that activate certain genes or pathways (Jeliazkova *et al.* 2015). This is particularly useful when coupled to a structured harmonized terminology, ie, an ontology. An AOP-specific ontology has recently been described and provides a basis for future AOP-linked database searches (Burgoon, 2017). Overall, we see the open source approach to the pipeline concept being pivotal to future AOP developments, as further data and knowledge thereby becomes easily incorporated under the existing protocol and overall stepwise construction approach.

The pipeline concept was successfully used to generate an AOP-linked molecular pathway description of the occupationally relevant disease pulmonary fibrosis. The resultant pathway model deepens the understanding of involvement of genes and molecular mechanisms underlying the disease by revealing how 64 pulmonary fibrosis associated genes are interlinked, which functions and existing pathways the genes relate to, and, more importantly, which of the AOP-associated events that are potentially critical (cf. Figure 4). In this case, the genes were derived from CTD, which provided a firm connection between the genes and the disease (including literature references), as well as to agents associated with the disease. The molecular pathway description was built primarily based on the 64 genes and the analyses of their interactions and functional associations (cf. Figure 1, steps 2–4). Nevertheless, interestingly, all AOP-associated events in the previously described putative AOP for the disease, from the MIE to the AO (Labib *et al.* 2015), corresponded well and could be readily linked to the established components (genes/gene groups) of the pathway (cf. Figure 1, step 5).

The pulmonary fibrosis pathway was constructed to represent a dynamic network, in contrast to the higher level linear directed flow of processes or interactions typical for an AOP. For example, some of the AOP-related events are represented by the same genes or gene groups, eg, KE1 and KE3 both relate to cytokine and chemokine signaling, largely due to the dynamic nature of pulmonary fibrosis, which progresses through cycles

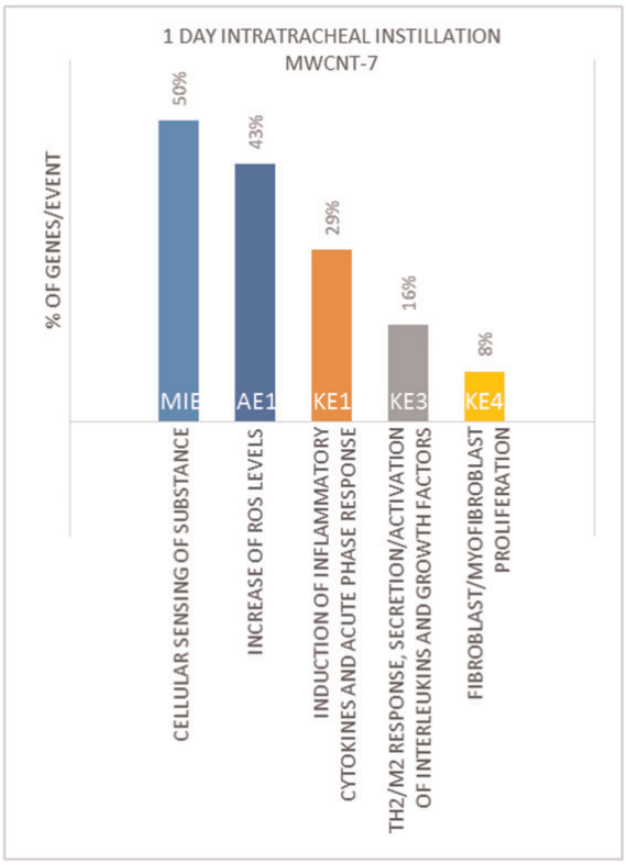

Figure 6. Differentially expressed genes in MWCNT-exposed mice 1-day post-exposure show strong involvement in early pulmonary-fibrosis related events (MIE, AE1, and KE1), with decreases in later events (KE3 and 4).

of repeated tissue injury and persistent inflammation (Labib *et al.* 2015). In addition, autocrine control of growth factors adds to these feedback loops and thus, KE3 and KE4 were also represented by overlapping sets of genes involved in differentiation pathways. Future evidence may specify the order and association of certain molecular pathways with specific AOP-related events. On the other hand, some of the AOP-related cellular events were not linked to any genes in the pathway, ie, KE2 and AE2. This may indicate areas (or events) in need of further research to pinpoint the underlying molecular patterns. Such efforts would strengthen the direct evidence of association in databases such as CTD, giving reason for incorporation of further key genes in the pathway. Finally, AE3, KE5, and the final AO are described as tissue or organ level responses, and although they are not currently represented by associated genes in the new pathway, they are included at the bottom to enable future computational linkage. Future linkage to, eg, collagen genes with AE3 or KE5 can be foreseen.

The independent data analyzed to visualize applicability of the pipeline concept included transcriptomics data from MWCNT-exposed mice, which rapidly develop pulmonary fibrosis (Porter *et al.*, 2010; Poulsen *et al.* 2013). The overall analysis implicated deregulation of all AOP-related events represented by genes in the pathway, ie, the MIE, AE1, KE1, KE3, and KE4. Interestingly, these genes were indicated already 1-day post-exposure (cf. Figure 5), although evident fibrotic changes as detected by histopathological analysis were seen much later (7-day post-exposure; Porter *et al.* 2010). Furthermore, a strong involvement of genes associated with the early events, ie, the MIE (50% of genes deregulated), the AE1 (43%), and the KE1 (29%), and less involvement of the later events KE3 (16%) and KE4 (8%) could be observed. Although the low number of altered genes prevents statistical analysis, these results may reflect the early time point, where the MIE can be assumed to be the most strongly affected, whereas KE4 is just about to kick in (Figure 6). The fact that also AE1 was so strongly affected (43%) already at this point, indicated its potentially strong involvement in the initiating event and progress toward the next KEs (Figure 6). Finally, the new AOP description enabled identification of 11 pulmonary fibrosis AOP-related genes which would not have been identified based on the previous descriptions of the disease (Todd *et al.* 2012; Vietti et al. 2016). Overall, this independent analysis supported that transcriptomics data captures early effects of the exposure, and in particular before the AO becomes histologically detectable.

The pulmonary fibrosis pathway serves as a detailed example for the generation of AOP-linked molecular pathways for other diseases. Fibrosis can occur in most organs and involves largely similar molecular responses (Rockey *et al.* 2015). Therefore, besides the generalizability of the proposed workflow, the knowledge gained in this case study could be used as a basis for the development of similar pathways descriptive of fibrosis in other organs, eg, liver fibrosis, currently described as an established (OECD reviewed) AOP (https://aopwiki.org/aops/38). Recent thinking on novel ways of toxicity testing argues for the application of omics data and AOP-linked molecular pathways (Kohonen *et al.* 2017). A number of current initiatives are strongly focused on the development and use of high-throughput data generation and omics methodology (Collins *et al.*, 2017; Merrick *et al.*, 2015). Such data are directly applicable to AOP-targeted pathway analyses, including, (1) disease-linked overrepresentation analysis, (2) disease-linked enrichment analysis among differentially expressed genes, (3) integration into AOP-based testing strategies as multivariate biomarkers (Tollefsen *et al.*, 2014), (4) application as disease-linked descriptors in (quantitative) structure-activity relationships ((Q)SAR) approaches (Tsiliki et al. 2017), and finally, based on all the above, enable (5) disease-linked grouping and read across among toxic agents through implementation into omics-based toxicity predictive tools. In relation to this, our laboratory recently described a Predictive Toxicogenomics Space (PTGS) scoring concept that gives any omics analysis a predictive toxicity effect (Kohonen *et al.*, 2017). Such analyses could be part of tiered workflows involving, in addition to omics modeling, high-throughput screening for ranking and prioritizing, and high-content analysis for validation of the omics-based AOP-linked hypotheses (Grafström *et al.*, 2015). Thus, although the pipeline-derived AOP descriptions are currently primarily qualitative, future directions will operate to incorporate increasingly quantitative approaches (Grafström *et al.* 2015; Kohonen *et al.* 2017).

The suggested AOP-enriching approach is likely applicable from different angles depending on the data available for the specific disease under investigation. The case study chosen to test the approach was based on genes identified from CTD, and although the database is one of the most comprehensive gene-disease relationship sources, future efforts may refine the identification of key disease-related genes to be based on multiple sources of information. Furthermore, validation, especially from a regulatory point of view, will require large-scale data sets covering at least 30–50 compounds or nanomaterials per AOP, and preferably hundreds, for sufficient statistical power to

test sensitivity, specificity, and coverage (ie, application domain). In addition, these data should preferably cover broad dose ranges, although tiered testing strategies in conjunction with high-content screening data may be able to reduce that need (Grafström *et al.* 2015). The required large-scale and mainly *in vitro*-based gene expression data sets can be expected to be generated as part of future efforts that utilize high-throughput transcriptomics technologies similar to the TempO-seq Tox21 Phase III L1500+ or the Broad Institute Connectivity Map L1000 platform (Andersen *et al.* 2015; Collins *et al.* 2017; Subramanian et al. 2017). Ultimately, transcriptomic benchmark dose-response analyses may utilize WikiPathways resources such as this one, the PTGS or other high-quality targeted gene set collections as part of a validated scoring concept workflow tailored to each AOP (Dean *et al.* 2017; Kohonen *et al.* 2017).

In summary, we present an open source, diverse data, and bioinformatics tool-dependent pipeline concept that lies at the heart of accepting WoE, NAMs, and AOP-linkage for risk characterization work. Taking pulmonary fibrosis as case study, the analysis pointed to both known and novel gene associations, although it served primarily to demonstrate the functionality of using a mixed supervised versus unsupervised but fixed protocol for developing AOP descriptions. We believe such an approach to be gaining momentum within risk assessment practices, and in fact the generalized pipeline described here, might be suitable for standardized use to generate descriptions of further toxicity-relevant diseases. Molecular pathways linked to AOPs provide clarity in the complexity of biology, and provide a much-needed basis for decisions on further testing and validation needs of central events. With the implementation of novel AOP methodology such as that described herein, future cost-effective early phase routine safety testing practices will increasingly ensure that chemicals and engineered nanomaterials are generated and applied under a safe innovation approach.

SUPPLEMENTARY DATA

Supplementary data are available at *Toxicological Sciences* online.

FUNDING

This work was supported by funding from the European Union's Seventh Framework Programme under the grant agreement No. 604134 (eNanoMapper), and Horizon 2020 research and innovation programme under grant agreements No. 646221 (NanoReg2) and No. 686239 (caLIBRAte). Further support was received from the Vinnova-Eurostars project ToxHQ (grant No. E9698), the Swedish Cancer and Allergy Fund, the Swedish Research Council and the Swedish Fund for Research without Animal Experiments. The sole responsibility of this publication lies with the authors. The European Union is not responsible for any use that may be made of the information contained therein.

ACKNOWLEDGMENTS

The partners in the EU-funded projects are acknowledged for inspiring discussions. The Institute of Environmental Medicine at Karolinska Institute is acknowledged for the research facilities and creating an inspiring environment.

REFERENCES

Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F., and Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM(®)), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* **43**, D789–D798.

Andersen, M. E., McMullen, P. D., and Krewski, D. (2015). Developing tools for defining and establishing pathways of toxicity. *Arch. Toxicol.* **89**, 809–812.

Boué, S., Talikka, M., Westra, J. W., Hayes, W., Di Fabio, A., Park, J., Schlage, W. K., Sewer, A., Fields, B., Ansari, S., *et al.* (2015). Causal biological network database: A comprehensive platform of causal biological network models focused on the pulmonary and vascular systems. *Database* **2015**, bav030.

Boyles, M., Stoehr, L., Schlinkert, P., Himly, M., and Duschl, A. (2014). The significance and insignificance of carbon nanotube-induced inflammation. *Fibers* **2**, 45.

Burden, N., Sewell, F., Andersen, M. E., Boobis, A., Chipman, J. K., Cronin, M. T. D., Hutchinson, T. H., Kimber, I., and Whelan, M. (2015). Adverse Outcome Pathways can drive non-animal approaches for safety assessment. *J. Appl. Toxicol.* **35**, 971–975.

Burgoon LD. (2017). The AOPOntology: A semantic artificial intelligence tool for predictive toxicology. *Appl In Vitro Toxicol.* **3**(3), 278–281.

Clippinger, A. J., Ahluwalia, A., Allen, D., Bonner, J. C., Casey, W., Castranova, V., David, R. M., Halappanavar, S., Hotchkiss, J. A., Jarabek, A. M., *et al.* (2016). Expert consensus on an in vitro approach to assess pulmonary fibrogenic potential of aerosolized nanomaterials. *Arch. Toxicol.* **90**, 1769–1783.

Collins, A. R., Annangi, B., Rubio, L., Marcos, R., Dorn, M., Merker, C., Estrela-Lopis, I., Cimpan, M. R., Ibrahim, M., Cimpan, E., *et al.* (2017). High throughput toxicity screening and intracellular detection of nanomaterials. *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* **9**, e1413.

Davis, A. P., Grondin, C. J., Johnson, R. J., Sciaky, D., King, B. L., McMorran, R., Wiegers, J., Wiegers, T. C., and Mattingly, C. J. (2017). The comparative toxicogenomics database: Update 2017. *Nucleic Acids Res.* **45**, D972–D978.

Dean, J. L., Zhao, Q. J., Lambert, J. C., Hawkins, B. S., Thomas, R. S., and Wesselkamper, S. C. (2017). Editor's highlight: (2017) application of gene set enrichment analysis for identification of chemically induced, biologically relevant transcriptomic networks and potential utilization in human health risk assessment. *Toxicol. Sci.* **157**, 85–99.

Dong, J., and Ma, Q. (2015). Advances in mechanisms and signaling pathways of carbon nanotube toxicity. *Nanotoxicology* **9**, 658–676.

Dymacek, J., Snyder-Talkington, B. N., Porter, D. W., Mercer, R. R., Wolfarth, M. G., Castranova, V., Qian, Y., and Guo, N. L. (2015). mRNA and miRNA regulatory networks reflective of multi-walled carbon nanotube-induced lung inflammatory and fibrotic pathologies in mice. *Toxicol. Sci.* **144**, 51–64.

ECHA. (2016). *Topical Scientific Workshop – New Approach Methodologies in Regulatory Science*. https://echa.europa.eu/documents/10162/22816069/tsws_background_document_en.pdf/53102457-43f9-4709-8e4f-cf2b58458fbf.

Ehrhart, F., Kutmon, M., Willighagen, E., Hanspers, K., Pico, A., Rieswijk, L., and Nymark, P. (2017). *Creating Biological Pathways for WikiPathways using PathVisio*. Figshare. https://doi.org/10.6084/m9.figshare.4668349.v1

Emad, A., and Emad, Y. (2007a). Relationship between eosinophilia and levels of chemokines (CCL5 and CCL11) and IL-5 in

bronchoalveolar lavage fluid of patients with mustard gas-induced pulmonary fibrosis. *J. Clin. Immunol.* **27**, 605–612.

Emad, A., and Emad, Y. (2007b). Levels of cytokine in bronchoalveolar lavage (BAL) fluid in patients with pulmonary fibrosis due to sulfur mustard gas inhalation. *J. Interferon. Cytokine Res.* **27**, 38–43.

Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S., *et al.* (2016). The Reactome pathway Knowledgebase. *Nucleic Acids Res.* **44**, D481–D487.

Gilkes, D. M., Bajpai, S., Chaturvedi, P., Wirtz, D., and Semenza, G. L. (2013). Hypoxia-inducible factor 1 (HIF-1) promotes extracellular matrix remodeling under hypoxic conditions by inducing P4HA1, P4HA2, and PLOD2 expression in fibroblasts. *J. Biol. Chem.* **288**, 10819–10829.

Grafström, R., Nymark, P., Hongisto, V., Spjuth, O., Ceder, R., Willighagen, E., Hardy, B., Kaski, S., and Kohonen, P. (2015). Toward the replacement of animal experiments through the bioinformatics-driven analysis of 'Omics' data from human cell cultures. *Altern. Lab. Anim.* **43**, 325–332.

Grondin, C. J., Davis, A. P., Wiegers, T. C., King, B. L., Wiegers, J. A., Reif, D. M., Hoppin, J. A., and Mattingly, C. J. (2016). Advancing exposure science through chemical data curation and integration in the comparative toxicogenomics database. *Environ. Health Perspect.* **124**, 1592–1599.

Hartung, T. (2016). Making big sense from big data in toxicology by read-across. *Altex* **33**, 83–93.

Herwig, R., Hardt, C., Lienhard, M., and Kamburov, A. (2016). Analyzing and interpreting genome data at the network level with ConsensusPathDB. *Nat. Protoc.* **11**, 1889–1907.

Jeliazkova, N., Chomenidis, C., Doganis, P., Fadeel, B., Grafström, R., Hardy, B., Hastings, J., Hegi, M., Jeliazkov, V., Kochev, N., *et al.* (2015). The eNanoMapper database for nanomaterial safety information. *Beilstein J. Nanotechnol.* **6**, 1609–1634.

Kamburov, A., Stelzl, U., Lehrach, H., and Herwig, R. (2013). The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res.* **41**, D793–D800.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361.

Kinnula, V. L., Fattman, C. L., Tan, R. J., and Oury, T. D. (2005). Oxidative stress in pulmonary fibrosis: A possible role for redox modulatory therapy. *Am. J. Respir. Crit. Care Med.* **172**, 417–422.

Kohonen, P., Parkkinen, J., Willighagen, E., Ceder, R., Wennerberg, K., Kaski, S., and Grafström, R. (2017). A transcriptomics-driven gene space accurately predicts liver cytopathology, including drug-induced liver injury. *Nat. Commun.* **8**, 15932.

Krämer, A., Green, J., Pollard, J., and Tugendreich, S. (2014). Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523–530.

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., *et al.* (2016). Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97.

Kutmon, M., Riutta, A., Nunes, N., Hanspers, K., Willighagen, E. L., Bohler, A., Mélius, J., Waagmeester, A., Sinha, Sravanthi, R., Miller, R., *et al.* (2016). WikiPathways: Capturing the full diversity of pathway knowledge. *Nucleic Acids Res.* **44**, D488–D494.

Kutmon, M., van Iersel, M. P., Bohler, A., Kelder, T., Nunes, N., Pico, A. R., Evelo, C. T., and Murphy, R. F. (2015). PathVisio 3: An extendable pathway analysis toolbox. *PLoS Comput. Biol.* **11**, e1004085.

Labib, S., Williams, A., Yauk, C., Nikota, J., Wallin, H., Vogel, U., and Halappanavar, S. (2015). Nano-risk Science: Application of toxicogenomics in an adverse outcome pathway framework for risk assessment of multi-walled carbon nanotubes. *Part Fibre Toxicol.* **13**, 15.

Manresa, M. C., Godson, C., and Taylor, C. T. (2014). Hypoxia-sensitive pathways in inflammation-driven fibrosis. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **307**, R1369.

Merrick, B. A., Paules, R. S., and Tice, R. R. (2015). Intersection of toxicogenomics and high throughput screening in the Tox21 program: An NIEHS perspective. *Int. J. Biotechnol.* **14**, 7–27.

Miller, J., Thompson, J., and Shukla, A. (2014). Asbestos-induced oxidative stress in lung pathogenesis. In *Systems Biology of Free Radicals and Antioxidants* (I. Laher, Ed.), pp. 1587–1610. Springer Berlin Heidelberg, Berlin, Heidelberg.

Mossman, B., and Churg, A. (1998). Mechanisms in the pathogenesis of asbestosis and silicosis. *Am. J. Respir. Crit. Care Med.* **157**, 1666–1680.

Porter, D. W., Hubbs, A. F., Mercer, R. R., Wu, N., Wolfarth, M. G., Sriram, K., Leonard, S., Battelli, L., Schwegler-Berry, D., and Friend, S. (2010). Mouse pulmonary dose- and time course-responses induced by exposure to multi-walled carbon nanotubes. *Toxicology* **269**, 136–147.

Poulsen, S. S., Jacobsen, N. R., Labib, S., Wu, D., Husain, M., Williams, A., Bøgelund, J. P., Andersen, O., Købler, C., Mølhave, K., *et al.* (2013). Transcriptomic analysis reveals novel mechanistic insight into murine biological responses to multi-walled carbon nanotubes in lungs and cultured lung epithelial cells. *PLoS One* **8**, e80452.

Rockey, D. C., Bell, P. D., and Hill, J. A. (2015). Fibrosis—a common pathway to organ injury and failure. *N. Engl. J. Med.* **372**, 1138–1149.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2013). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504.

Snyder-Talkington, B., Dymacek, J., Porter, D., Wolfarth, M., Mercer, R., Pacurari, M., Denvir, J., Castranova, V., Qian, Y., and Guo, N. (2013). System-based identification of toxicity pathways associated with multi-walled carbon nanotube-induced pathological responses. *Toxicol. Appl. Pharmacol.* **272**, 476–489.

Snyder-Talkington, B. N., Dong, C., Porter, D. W., Ducatman, B., Wolfarth, M. G., Andrew, M., Battelli, L., Raese, R., Castranova, V., Guo, N. L., *et al.* (2016). Multi-walled carbon nanotube-induced pulmonary inflammatory and fibrotic responses and genomic changes following aspiration exposure in mice: A 1 year post-exposure study. *J. Toxicol. Environ. Health A* **79**, 352–366.

Snyder-Talkington, B. N., Dong, C., Zhao, X., Dymacek, J., Porter, D. W., Wolfarth, M. G., Castranova, V., Qian, Y., and Guo, N. L. (2015). Multi-walled carbon nanotube-induced gene expression in vitro: Concordance with in vivo studies. *Toxicology* **328**, 66–74.

Sturla, S. J., Boobis, A. R., FitzGerald, R. E., Hoeng, J., Kavlock, R. J., Schirmer, K., Whelan, M., Wilks, M. F., and Peitsch, M. C. (2014). Systems toxicology: From basic research to risk assessment. *Chem. Res. Toxicol.* **27**, 314–329.

Subramanian, A., Narayan, A., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., and Gould, J. A. (2017). Next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *bioRxiv* 136168.

Thomas, R. S., Philbert, M. A., Auerbach, S. S., Wetmore, B. A., Devito, M. J., Cote, I., Rowlands, J. C., Whelan, M. P., Hays, S. M., Andersen, M. E., *et al.* (2013). Incorporating new technologies into toxicity testing and risk assessment: Moving from 21st century vision to a data-driven framework. *Toxicol. Sci.* **136**, 4–18.

Todd, N., Luzina, I., and Atamas, S. (2012). Molecular and cellular mechanisms of pulmonary fibrosis. *Fibrog. Tissue Rep.* **5**, 11–11.

Tollefsen, K. E., Scholz, S., Cronin, M. T., Edwards, S. W., de Knecht, J., Crofton, K., Garcia-Reyero, N., Hartung, T., Worth, A., and Patlewicz, G. (2014). Applying Adverse Outcome Pathways (AOPs) to support Integrated Approaches to Testing and Assessment (IATA). *Regul. Toxicol. Pharmacol.* **70**, 629–640.

Tsiliki, G., Nymark, P., Kohonen, P., Grafström, R., and Sarimveis, H. (2017). Enriching nanomaterials omics data: An integration technique to generate biological descriptors. *Small Methods* **1**, 1700139.

van der Veen, J. W., Soeteman-Hernández, L. G., Ezendam, J., Stierum, R., Kuper, F. C., and van Loveren, H. (2014). Anchoring molecular mechanisms to the adverse outcome pathway for skin sensitization: Analysis of existing data. *Crit. Rev. Toxicol.* **44**, 590–599.

van Iersel, M. P., Kelder, T., Pico, A. R., Hanspers, K., Coort, S., Conklin, B. R., and Evelo, C. (2008). Presenting and exploring biological pathways with PathVisio. *BMC Bioinformatics* **9**, 399–399.

van Iersel, M. P., Pico, A. R., Kelder, T., Gao, J., Ho, I., Hanspers, K., Conklin, B. R., and Evelo, C. T. (2010). The BridgeDb framework: Standardized access to gene, protein and metabolite identifier mapping services. *BMC Bioinformatics* **11**, 5.

Vancheri, C. (2013). Common pathways in idiopathic pulmonary fibrosis and cancer. *Eur. Respir. Rev.* **22**, 265–272.

Vietti, G., Lison, D., and van den Brule, S. (2016). Mechanisms of lung fibrosis induced by carbon nanotubes: Towards an Adverse Outcome Pathway (AOP). *Part Fibre Toxicol.* **13**, 11.

Vinken, M. (2013). The adverse outcome pathway concept: A pragmatic tool in toxicology. *Toxicology* **312**, 158–165.

Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C. T., *et al.* (2010). The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* **38**, W214–W220.