scientific reports

OPEN

Network-based integrated analysis for toxic effects of high-concentration formaldehyde inhalation exposure through the toxicogenomic approach

Doo Seok Kang[1], Nahyun Lee[2], Dong Yeop Shin[1], Yu Jin Jang[1], Su-Hyon Lee[3], Kyung-Min Lim[2], Yeon-Soon Ahn[4], Cheol Min Lee[5✉] & Young Rok Seo[1✉]

Formaldehyde is a colorless, pungent, highly reactive, and toxic environmental pollutant used in various industries and products. Inhaled formaldehyde is a human and animal carcinogen that causes genotoxicity, such as reactive oxygen species formation and DNA damage. This study aimed to identify the toxic effects of inhaled formaldehyde through an integrated toxicogenomic approach utilizing database information. Microarray datasets (GSE7002 and GSE23179) were collected from the Gene Expression Omnibus database, and differentially expressed genes were identified. The network analyses led to the construction of the respiratory system-related biological network associated with formaldehyde exposure, and six upregulated hub genes (*AREG*, *CXCL2*, *HMOX1*, *PLAUR*, *PTGS2*, and *TIMP1*) were identified. The expression levels of these genes were verified via qRT-PCR in 3D reconstructed human airway tissues exposed to aerosolized formaldehyde. Furthermore, *NRARP* was newly found as a potential gene associated with the respiratory and carcinogenic effects of formaldehyde by comparison with human in vivo and in vitro formaldehyde-exposure data. This study improves the understanding of the toxic mechanism of formaldehyde and suggests a more applicable analytic pipeline for predicting the toxic effects of inhaled toxicants.

Formaldehyde is a toxic chemical with a colorless, pungent odor, and high reactivity[1]. The common toxic effects of formaldehyde include irritation and damage to contacted tissues[2]. Formaldehyde induces various levels of genotoxic effects, such as N^2-hydroxymethyl-dG adducts, DNA–protein crosslinks, and chromosomal damage[3,4]. The International Agency for Research on Cancer (IARC) has classified formaldehyde as a human carcinogen that causes nasopharyngeal cancer and leukemia[1]. Despite its toxic effects, formaldehyde is ubiquitous in the environment because of its wide use in resin production, as a chemical intermediate, in disinfectants, medical fields, and preservatives, among other commercial purpose[2]. Humans are exposed to very low concentrations of environmental formaldehyde in daily life, primarily through inhalation[5]. Unlike the exposure in daily life, occupational exposure occurs at a wide range of concentrations, as much as 20.94 mg/m^3 in an anatomical laboratory and 60.77 ppm in a factory, in extreme cases[2,6,7]. Therefore, people working in high-risk occupational environments are likely to be exposed to high concentrations for an extended time[8]. In addition, if a formaldehyde spill/leakage accident occurs in these occupational environments, human exposure to unusually very high concentrations of formaldehyde in the air causes external symptoms and underlying genetic modifications.

[1]Department of Life Science, Dongguk University Biomedi Campus, Goyang, Gyeonggi 10326, Republic of Korea. [2]College of Pharmacy, Ewha Womans University, Seoul 03760, Republic of Korea. [3]R&D Institute, Biosolution Co., Ltd., Seoul 01811, Republic of Korea. [4]Department of Preventive Medicine and Institute of Occupational and Environmental Medicine, Wonju College of Medicine, Yonsei University, Wonju, Gangwon 26426, Republic of Korea. [5]Department of Chemical and Biological Engineering, College of Natural Science and Engineering, Seokyeong University, Seoul 02173, Republic of Korea. ✉email: cheolmin@skuniv.ac.kr; seoyr@dongguk.edu
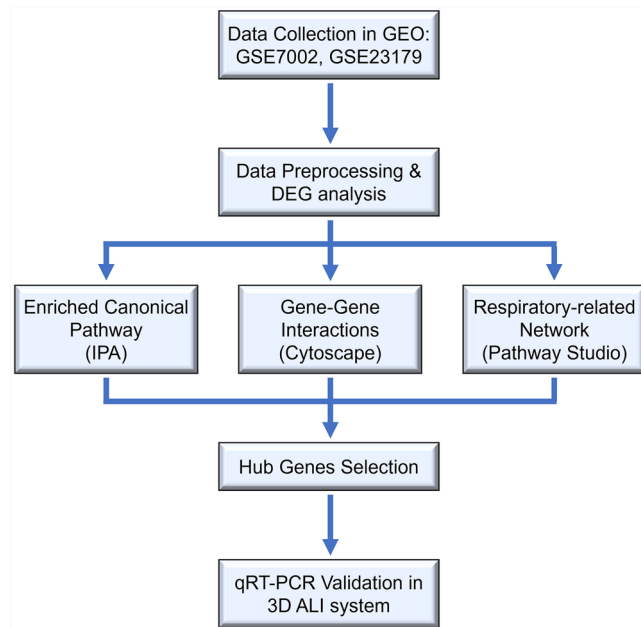
Figure 1. Flow chart of the steps used in this study.

High-throughput technologies, such as microarray and RNA-sequencing, allow the analysis of large amounts of gene expression data. Gene Expression Omnibus (GEO), established by the National Center for Biotechnology Information (NCBI), is the largest public database that archives and distributes several types of high-throughput genomic data[9]. A comprehensive analysis of existing experimental data using bioinformatic tools can improve the understanding of biological phenomena. Various network strategies can be applied to information gathered from numerous independent studies for performing bioinformatic analysis to elucidate complex biological interactions[10,11].

To synthetically explore the toxic mechanisms of formaldehyde inhalation, two microarray datasets (GSE7002 and GSE23179) derived from formaldehyde inhalation studies in rats under identical in vivo experimental conditions were collected from the GEO database. By integrating and expanding the interpretation of the original datasets[12,13], we investigated the toxic effects of formaldehyde exposure on the respiratory system by the toxicogenomic approach. After data preprocessing, differentially expressed genes (DEGs) were obtained, and hub genes associated with the respiratory effect of formaldehyde exposure were identified through the network-based bioinformatics approaches. To resemble airway in vivo exposure, we employed a three-dimensional (3D) reconstructed human airway mucosa model using a custom-made aerosol exposure system. The cytotoxicity of aerosolized formaldehyde was estimated, and the expression levels of selected hub genes were validated via air–liquid interface (ALI) culture. Our integrated analysis could improve understanding of the toxic effects of formaldehyde exposure on the respiratory system. An overview of the methodology applied in this study is shown in Fig. 1.

Materials and methods

Data collection and preprocessing. Formaldehyde inhalation gene expression profile datasets were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). Two gene expression profiles (GSE7002 and GSE23179) derived from the GPL1355 platform [Rat230_2] Affymetrix Rat Genome 230 2.0 Array were downloaded. Among all datasets, samples with the high-exposure concentration (15 ppm) and different exposure periods (6 h, 5 days, 4 weeks, and 13 weeks) were selected to explore the pattern of time-dependent changes induced by formaldehyde. The selected dataset consisted of 37 control samples and 16 formaldehyde-exposed samples. The raw data were normalized using the robust multi-array average method in the affy package (version 1.62.0) in R software[14]. To remove non-biological batch effects generated in the data merging process, the ComBat function in the sva package (version 3.38.0) was used[15].

Clustering analysis and DEGs screening. After preprocessing, Principal component analysis (PCA) was conducted to examine the pattern of gene expression profiles derived from different datasets. DEGs between control samples and formaldehyde-exposed samples were filtered using the limma R package (version 3.40.6)[16]. The threshold criteria were |fold change|> 1.5 and false discovery rate adjusted p-value < 0.05. The expression patterns of the common DEGs in all exposure groups were visualized through the hierarchical clustering heatmap.

Functional enrichment pathway analysis.　The enrichment pathways for the common DEGs were analyzed using the Ingenuity Pathway Analysis (IPA) software (Qiagen, Germany). The IPA is based on the Ingenuity Knowledge Base, a large, curated biomedical database. The core analysis of IPA provides diverse biological information, including canonical pathways, upstream regulators, diseases, biofunctions, and sub-networks. Statistical significance is calculated by Fisher's exact test coupled with the z-score algorithm[17].

Network analysis and hub gene selection.　Pathway Studio version 12.3, a commercial text mining-based biological network analysis software, was used to identify hub genes associated with the toxic effects of formaldehyde inhalation exposure[18]. Pathway Studio enables researchers to explore biological interactions extracted from a vast number of studies and visualize these interactions through the entity and connectivity of each relation. In Pathway Studio, "entity" signifies a node, such as a gene, disease, and cell process, and "relation" signifies a biological interaction between two entities. Each relation is expressed as a connectivity (edge) on the network and supported by the number of reference studies[19]. "Direct Interactions," "Expand Pathway," and "Shortest Path" algorithms were applied to construct the functional network of the common DEGs. Topological parameters between gene–gene interactions were calculated using the NetworkAnalyzer in Cytoscape software (version 3.7.2)[20].

3D reconstructed human airway mucosa model and aerosolized formaldehyde exposure.　A 3D reconstructed human airway mucosa model, SoluAirway, and its media were purchased from Biosolution Co., Ltd. (Seoul, Korea). SoluAirway was placed in a 6-well plate filled with culture media (0.9 mL/well) and pre-incubated at 37 ℃ and 5% CO_2 overnight. Formaldehyde (16%, methanol-free, ultrapure) was purchased from Polysciences, Inc. (Warrington, PA, USA) and diluted in distilled water. According to a previous study's protocol[21], aerosol exposure to SoluAirway was conducted for 2 min using a medical nebulizer NE-U150 (OMRON Healthcare, Japan) in the experimentally designed chamber to treat 100 μL of formaldehyde and distilled water (control). Then, the tissues were incubated for 24 h. After the aerosol exposure, the apical surfaces of the tissues were washed four times with PBS (400 μL). Tissue viability was measured by the 3-(4,5-dimethyl-thiazol-2-yl)-2,5 diphenyltetrazolium bromide (MTT) assay. The tissues were transferred to 24-well plates filled with MTT solution (1.0 mg/mL, 200 μL), and the tissues were filled with MTT solution (100 μL). After incubation at 37 °C and 5% CO_2 for 3 h, MTT solution was eliminated, and the tissues were submerged in isopropanol (2.0 mL) for 3 h, protected from light. Extracted formazan was transferred to a 96-well plate (200 μL/well), and absorbance was measured at 570-nm wavelength.

Quantitative real-time reverse transcription PCR (qRT-PCR).　The membranes of SoluAirway were cut with a surgical blade, placed on a 6-cm plate, soaked in PBS, and the tissues were harvested using scrapers. Total RNA was extracted using an RNeasy kit (Qiagen, Germany), and the RNA quality was evaluated using the Agilent 2100 Bioanalyzer (Agilent Technologies, USA). cDNA was synthesized from extracted RNA (500 μg) using the ImProm-II Reverse Transcription System (Promega, USA), according to the manufacturer's instructions. qRT-PCR was conducted using TB Green Premix Ex Taq (Takara Bio, Japan) in Rotor-Gene Q (Qiagen, Germany) under the following thermal cycling conditions: initial denaturation at 95 °C for 5 min, 40 cycles of 3 steps, including 95 °C for 15 s, 60 °C for 15 s, and 72 °C for 30 s. Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) was used to normalize the gene expression levels according to the $2^{-\Delta\Delta CT}$ method[22]. The sequences of primers for qRT-PCR are described in Supplementary Table S1.

Statistical analysis.　Statistical analyses were performed using R software (version 3.6.3). All experimental data are presented as the mean ± standard error of the mean (SEM). Significant differences between the formaldehyde exposure group and control were estimated by Student's t-test at p-value < 0.05.

Results

Clustering patterns of transcriptomic profiles and DEGs identification.　The experimental conditions of the two GEO datasets (GSE7002 and GSE23179) were designed almost identically: 8-week-old male rats, whole-body inhalation exposure, 6 h/day of formaldehyde exposure, nasal epithelial cells. To acquire more robust gene expression profiles, we merged 37 control and 16 formaldehyde-exposed samples in the two GEO datasets and corrected batch effects arising from different sampling dates. PCA result indicated that formaldehyde-exposed samples were distinct from control samples, regardless of the GEO datasets (Fig. 2a).

DEGs were identified for each formaldehyde exposure group compared to the control group (Fig. 2b). There were 815 genes significantly expressed at 6 h exposure (upregulated: 354, downregulated: 461), 2,308 genes at 5 days (upregulated: 1,342, downregulated: 966), 2,095 genes at 4 weeks (upregulated: 1,118, downregulated: 977), and 3,621 genes at 13 weeks (upregulated: 1,712, downregulated: 1,909). Among all exposure groups, 258 genes were identified as common DEGs (Fig. 2c). These common DEGs showed a mostly consistent up/downregulated expression pattern at each exposure period, except for 7 genes (Fig. 2d).

Enriched canonical pathways associated with formaldehyde exposure.　To interpret the meaning of the observed gene expression changes, IPA was used to identify canonical pathways for the common DEGs filtered by human genes. The top 10 canonical pathways were selected based on the − log(p-value), and the activation z-score pattern of the pathways was considered (Fig. 3, Table 1). Among the canonical pathways, p53 Signaling (z-score = 1, $p = 4.07 \times 10^{-4}$), ErbB Signaling (z-score = 1.13, $p = 5.37 \times 10^{-4}$), and NRF2-mediated Oxidative Stress Response (z-score = 0.378, $p = 4.47 \times 10^{-3}$) exhibited positive activity patterns, and Ferropto-
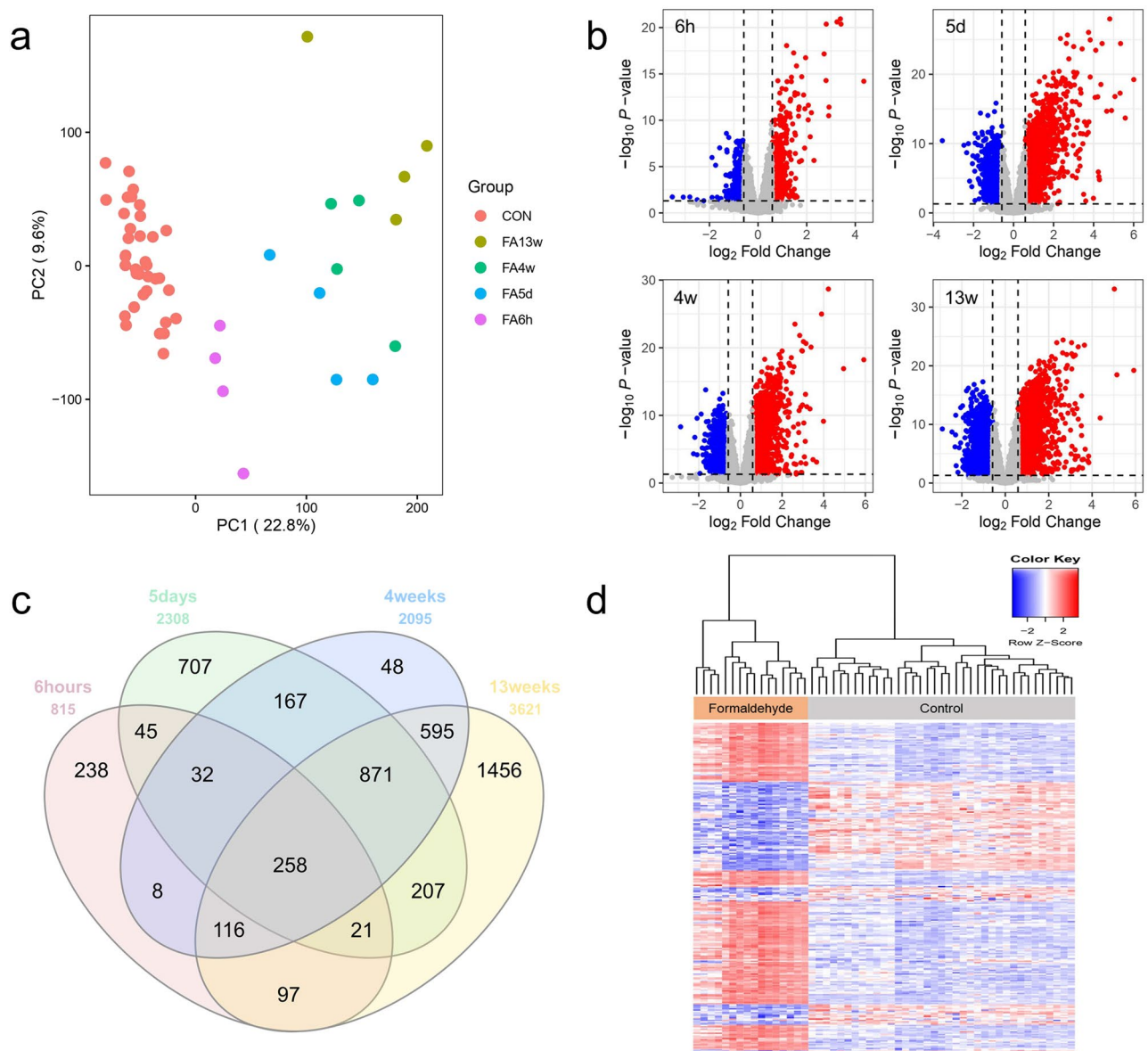
Figure 2. Clustering patterns of transcriptomic profiles and DEGs identification (**a**) Batch effect corrected-PCA result. PCA showed a distinct difference between the control and the exposure groups. (**b**) Volcano plot results of DEGs by each exposure groups. (**c**) Venn diagram of the DEGs among the formaldehyde exposure groups. (**d**) Hierarchical clustering heatmap of the common DEGs. All analyses were performed using R software (version 3.6.3).

sis Signaling Pathway (z-score = $-$1.41, $p = 4.47 \times 10^{-4}$) exhibited negative activation. No activity patterns were observed in other canonical pathways.

Respiratory-related network associated with formaldehyde exposure. To identify the hub genes related to formaldehyde exposure, we used Pathway Studio to perform network analysis on the common DEGs that were most sensitive to formaldehyde exposure. We focused our analysis on the respiratory effects of inhaled formaldehyde exposure, so we filtered the genes with expression information in the major respiratory organs (from nose to lung) among the common DEGs (Fig. 4a). From these genes, we selected the top 24 genes with edge degree number ≥ 10 based on the betweenness centrality of gene–gene interactions derived from the Pathway Studio database information (Fig. 4b, Table 2). In addition, we extracted major cell processes and respiratory diseases associated with formaldehyde exposure based on the minimum number of references (≥ 3) (Fig. 4c). A biological network for respiratory effects of formaldehyde exposure was constructed with comprehensive consideration of the results of gene expression values, the centrality of gene–gene interactions, and chemical–gene–disease associations (Fig. 4d). These genes were highly associated with diseases, including lung cancer, asthma, and pneumonia, and with oxidative stress, inflammatory response, and immune response, as cell processes. Finally, six hub genes (*AREG*, *CXCL2*, *HMOX1*, *PLAUR*, *PTGS2*, and *TIMP1*) and major entities were selected
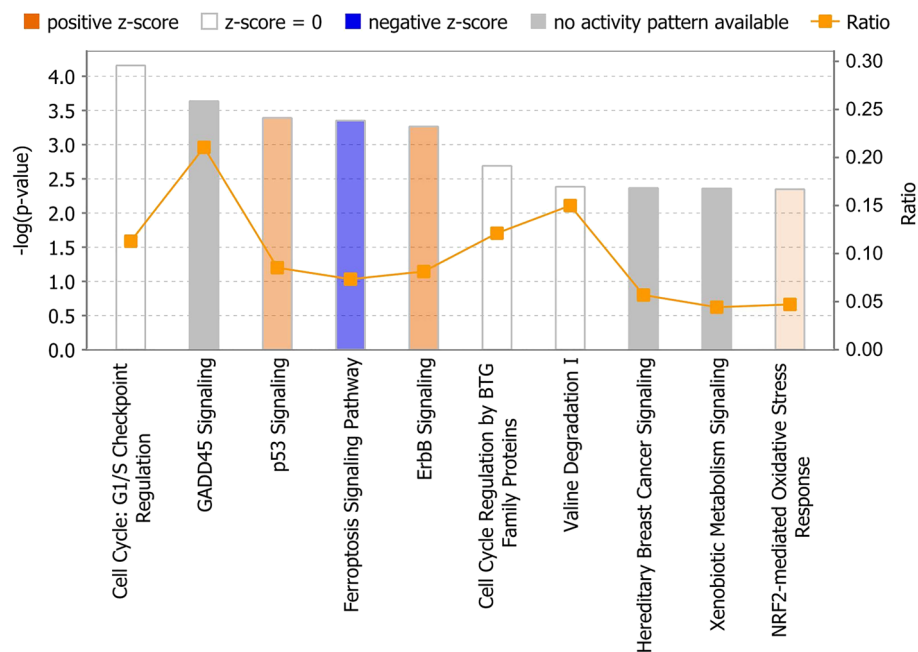
Figure 3. IPA top 10 canonical pathways for the common DEGs. Canonical pathways were sorted by − log(p-value). A positive z-score (orange) denotes pathway activation, and a negative z-score (blue) denotes pathway inhibition. Ratio refers to the percentage of DEGs among the total number of genes that make up the pathway.

Ingenuity canonical pathways	− log(p-value)	Ratio	z-score	Molecules
Cell Cycle: G1/S Checkpoint Regulation	4.16	0.113	0	*BTRC, CCNE2, CDKN1A, GNL3, MYC, NRG1, PAK1IP1*
GADD45 Signaling	3.64	0.211	NULL	*CCNE2, CDKN1A, GADD45B, PCNA*
p53 Signaling	3.39	0.0854	1	*CDKN1A, COQ8A, GADD45B, GNL3, PCNA, PIK3C2G, SFN*
Ferroptosis Signaling Pathway	3.35	0.0734	− 1.414	*CDKN1A, GCH1, H2AX, HMOX1, HSPB1, PRKAG2, SLC1A5, SLC7A11*
ErbB Signaling	3.27	0.0814	1.134	*AREG, EREG, HBEGF, MAP2K3, MAP2K6, NRG1, PIK3C2G*
Cell Cycle Regulation by BTG Family Proteins	2.69	0.121	NULL	*CCNE2, NOCT, PPM1L, PPP2R2B*
Valine Degradation I	2.39	0.15	NULL	*ACADSB, BCAT2, BCKDHB*
Hereditary Breast Cancer Signaling	2.37	0.0569	NULL	*CDKN1A, GADD45B, H2AX, NPM1, PIK3C2G, SFN, SMARCA2*
Xenobiotic Metabolism Signaling	2.36	0.0442	NULL	*FMO1, FMO2, GSTM2, HMOX1, MAP2K3, MAP2K6, MAP3K8, PIK3C2G, PPM1L, PPP2R2B*
NRF2-mediated Oxidative Stress Response	2.35	0.0471	0.378	*CBR1, FMO1, FOSL1, GSTM2, HMOX1, MAFF, MAP2K3, MAP2K6, PIK3C2G*

Table 1. List of top 10 canonical pathways for the common DEGs. *DEG* differentially expressed gene.

through careful examination of reference studies (Fig. 5a). Moreover, functional changes related to formaldehyde exposure were predicted; activation of cell signaling (NF-κB, ERK1/2), cytokine release, histone crosslinks, and degradation of antioxidant functions (superoxide dismutase, glutathione peroxidase).

Validation of hub genes in 3D air–liquid interface (ALI) system. To simulate in vivo inhalation exposure, we employed the 3D reconstructed human airway model SoluAirway and exposed it to aerosolized formaldehyde using a commercially available medical nebulizer. The concentration of formaldehyde exposure was set by tenfold serial dilutions (14.7 mg/mL (≈ 500 mM) to 50 μM), based on a previous study[21]. The MTT assay result showed that the tissue viability was > 90% at concentrations up to 5 mM and dropped significantly at concentrations above 50 mM (Supplementary Fig. S1). Among the sub-cytotoxic concentrations (> 90% viability), 0.5 mM (500 μM) was determined for qRT-PCR, considering that the 15-ppm data of the GEO dataset used is almost equivalent to 500 μM. The expression levels of the six hub genes were upregulated in the DEG results of all exposure periods and were validated using qRT-PCR. The mRNA expression of these genes examined in the 3D ALI system tended to be upregulated except *HMOX1* (Fig. 5b).
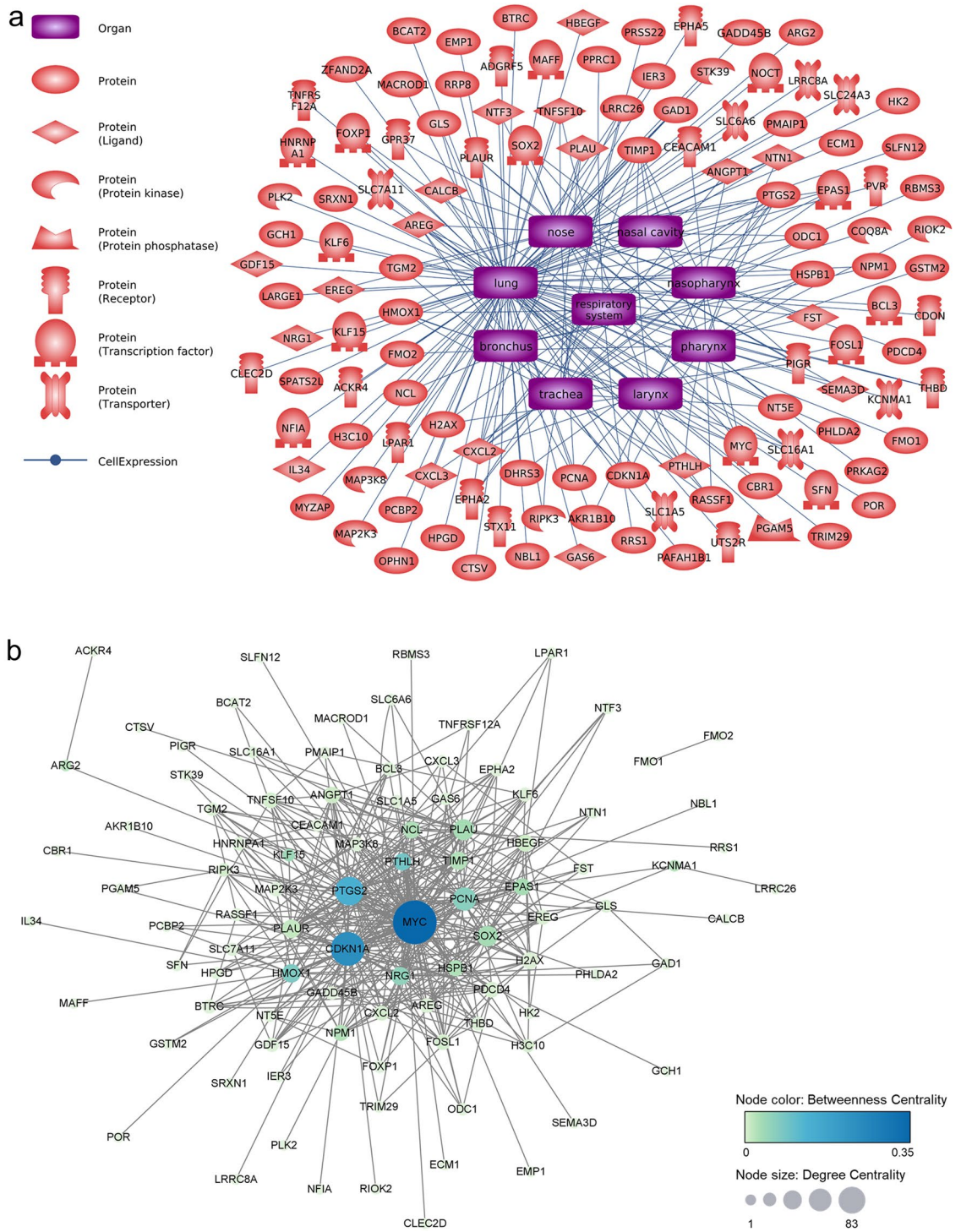
Figure 4. Biological networks for respiratory effects of formaldehyde exposure. (**a**) Genes with respiratory system expression information. Among the common DEGs, only genes with expression information in major respiratory organs were selected using Pathway Studio (version 12.3). (**b**) Gene–gene interaction network. Topological parameters among the genes were calculated using Cytoscape (version 3.7.2). (**c**) Formaldehyde-related major cell processes and respiratory diseases. (**d**) Respiratory system-related biological network associated with formaldehyde exposure. The biological network was constructed based on the results of the previous network analysis. Red and blue highlights, respectively, indicate up- and downregulated DEGs.
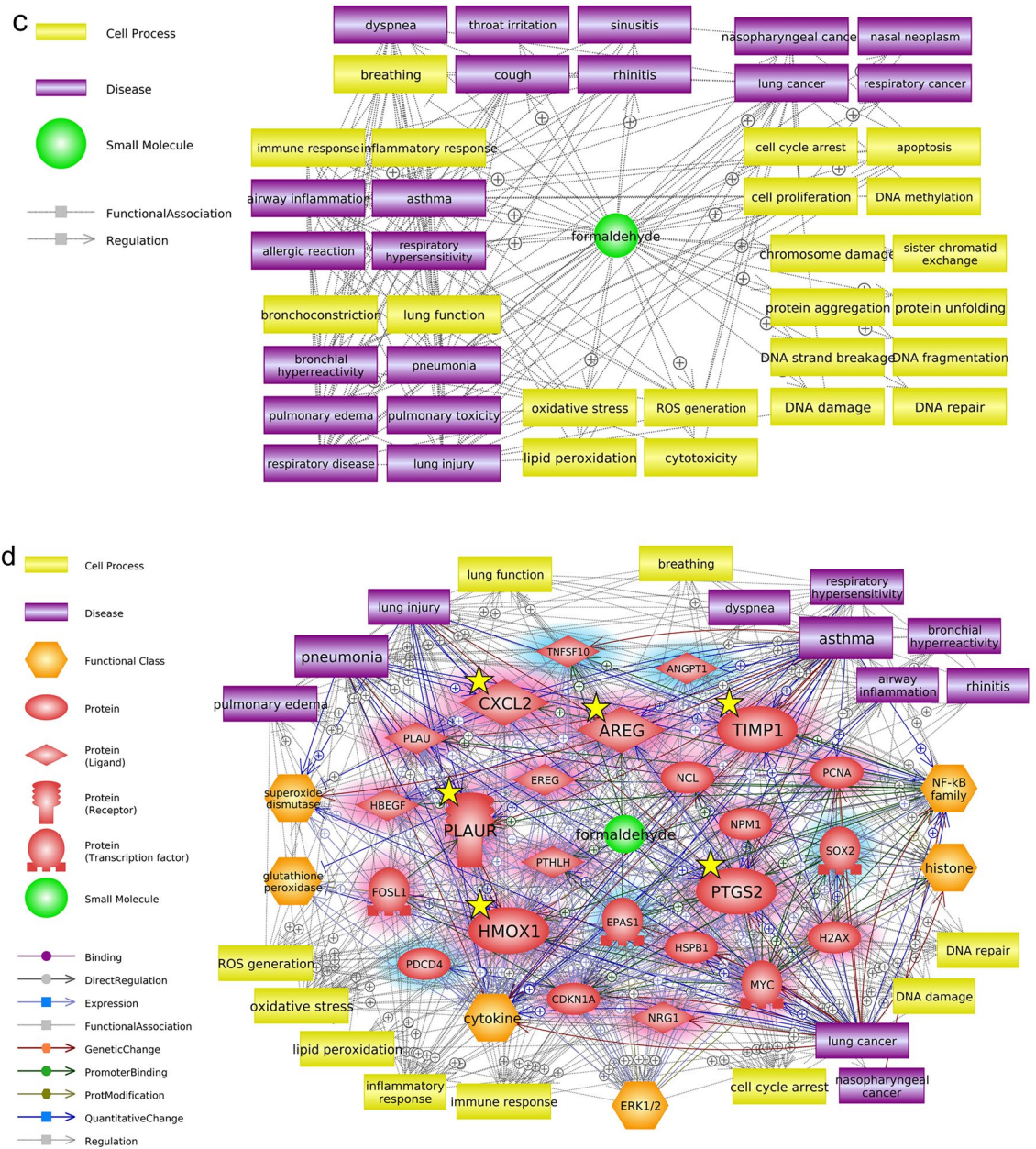
Figure 4. (continued)

Discussion

Formaldehyde is widely used in various fields and poses a high toxicity concern. In the early 1980s, the carcinogenicity of 5.6 and 14.3 ppm formaldehyde was confirmed in rats and mice for 2 years of exposure[23]. These concentrations are very high and unlikely to be encountered in daily life[24]. Therefore, our results using 15-ppm exposure data reflect high-exposure scenarios, such as chemical accidents and chronic occupational exposure to formaldehyde. Inhaled formaldehyde mainly affects the upper respiratory system, but a certain amount can be deposited directly into the lower respiratory system during oronasal breathing[2]. In animal, epidemiologic, and pulmonary function test studies, inhaled formaldehyde caused respiratory symptoms, with asthma identified as one of the most common diseases to arise due to exposure in indoor and occupational environments[25–29]. However, detailed knowledge of the mechanisms of respiratory disease onset associated with genetic changes in response to formaldehyde exposure is still lacking. This study explored integrated transcriptomic changes resulting from formaldehyde inhalation exposure.

Utilizing public genomic data, we acquired significant genetic profiles of formaldehyde exposure. We identified hub genes related to respiratory effects through network analyses and qRT-PCR validation via the 3D ALI aerosol exposure system (Figs. 4, 5). AREG, a ligand of epidermal growth factor receptor (EGFR, ErbB1), was the most strongly expressed, with a fold change > 20 in all exposure groups, and was enriched in ErbB Signaling. AREG activates cell signaling pathways, including Ras/Raf/MEK/ERK and PI3K/AKT, and cellular events, such

Gene name	Betweenness centrality	Degree centrality	Gene name	Betweenness centrality	Degree centrality
MYC	0.3520	83	TIMP1	0.0259	19
CDKN1A	0.2357	58	NPM1	0.0247	14
PTGS2	0.1379	45	PLAUR	0.0137	20
PTHLH	0.0874	17	ANGPT1	0.0092	15
HMOX1	0.0773	20	HBEGF	0.0082	16
PCNA	0.0656	32	CXCL2	0.0060	11
NRG1	0.0575	19	FOSL1	0.0054	13
SOX2	0.0343	25	TNFSF10	0.0050	13
PLAU	0.0338	26	PDCD4	0.0042	13
EPAS1	0.0329	16	EREG	0.0032	10
NCL	0.0273	16	AREG	0.0026	11
HSPB1	0.0268	14	H2AX	0.0022	16

Table 2. Centrality information of selected genes.

as cell proliferation and apoptosis[30]. Elevated AREG expression is associated with airway inflammatory diseases, such as chronic obstructive pulmonary disease (COPD), asthma[31,32], and several types of cancer[33]. HMOX1 is the major downstream antioxidant protein of the NRF2-mediated Oxidative Stress Response pathway and contributes to inhibiting Ferroptosis Signaling, an iron-dependent cell death accompanied by the accumulation of lipid reactive oxygen species[34,35]. HMOX1 is generally known to defend against oxidative and inflammatory damages in diverse diseases, and there are results of overexpressed HMOX1 in some respiratory patients[36–39]. However, the qRT-PCR result did not match the expression tendency of the microarray DEG results. This may be due to inherent differences in the tissues used in the bioassays as well as the exposure environment[40]. PTGS2 is not expressed in most tissues in the normal state, but its expression is induced by diverse stimuli, including cytokines, such as epidermal growth factor, interleukin-1, and tumor necrosis factor, and during an inflammatory response[41]. PTGS2 expression and cytokines were increased in the respiratory epithelium and alveolar macrophages of asthmatics[42,43]. Persistent increased *PTGS2* expression is thought to be induced by the inflammatory response after formaldehyde exposure[44,45]. PLAUR has been identified as an asthma, pneumonia, and COPD susceptibility gene[46,47]. PLAUR expression was elevated in the bronchial biopsy of asthmatics and is required for epithelial wound repair[48]. TIMP1 mainly functions as an important endogenous inhibitor in regulating matrix metalloproteinases. An imbalance of the MMP9/TIMP ratio is associated with the pathogenesis of asthma and lung diseases[49]. Sputum of asthmatics showed increased mRNA expression of *TIMP1*[50]. CXCL2 regulates normal and asthmatic airway smooth muscle cell migration[51]. During pulmonary inflammation, the mRNA expression of *Cxcl2* was increased in rats[52]. Taken together, selected hub genes could be important markers for respiratory diseases caused by inhaled toxicants. Furthermore, people with defects in these genes will be more vulnerable to formaldehyde exposure even if they are not exposed to high concentrations[53].

Additionally, we compared our common DEGs with in vivo (human) and in vitro formaldehyde exposure data (GSE27263: nasal biopsy of volunteers exposed to formaldehyde up to 0.7 ppm for 4 h/day over 5 days; GSE21477: nasal epithelial cells exposed to 200 μM formaldehyde for 4 h). Interestingly, *NRARP* was identified as the only gene that showed increased expression in all formaldehyde-exposed data used, and 9 and 39 common genes were identified in human in vivo and in vitro data, respectively. Apart from the network analysis results, the up-regulation of the *NRARP* gene was also validated (Supplementary Fig. S2). Although the functions of NRARP and its association with formaldehyde have not been fully understood, it interacts with Notch and WNT signaling in angiogenesis[54] and presents dual pro- or antitumor activity[55]. The *Notch1* gene positively regulates *Nrarp*[56] and is upregulated in the long-term exposure (5 days to 13 weeks) DEG results. NRARP stimulates cell proliferation by negatively modulating p21/Rb-dependent cell cycle arrest[57]. Although the significance between the *NRARP* gene and respiratory diseases was not analyzed in the network analyses of our study, respiratory effects and cell cycle regulation-related carcinogenic effects of *NRARP* would be valuable in future formaldehyde studies.

Traditionally, inhalation toxicity studies are based on a strict animal inhalation test that is costly, time-consuming, technically challenging, and unethical. Many in vitro studies have relied on two-dimensional (2D) monocultures in submerged conditions that cannot reflect the real in vivo cellular environments and exposure conditions. Efforts to reduce the gap between animal testing and 2D cell culture models and reflect in vivo physiological conditions have led to advanced in vitro studies, such as 3D cell culture and microfluidics/micro-engineering technologies[58]. Additionally, studies incorporating a sophisticated aerosol exposure system (e.g., VITROCELL) and 3D reconstructed airway tissue to reflect actual exposure are considered ideal in studying inhalation toxicity[59]. For this reason, we employed a 3D reconstructed human airway model and exposed it to aerosolized formaldehyde to investigate the toxic effects of inhaled formaldehyde. Although consideration of interspecies differences and further studies on the reliability of 3D in vitro data will be needed, our integrated approach suggests an applicable analytic pipeline for predicting the toxic effects of inhaled toxicants.

This study performed a toxicogenomic in silico analysis based on in vivo data and validated the result in the 3D ALI system. However, our approach should be viewed as a prioritization method for further toxicity testing due to the limitations of data mining. For instance, the quality of data mining analysis is dependent on the databases of the software used[60]. Moreover, there could be false-positive results and missing interactions due to potential bias toward well-studied interactions[40,61]. Finally, further studies are needed to identify all possible
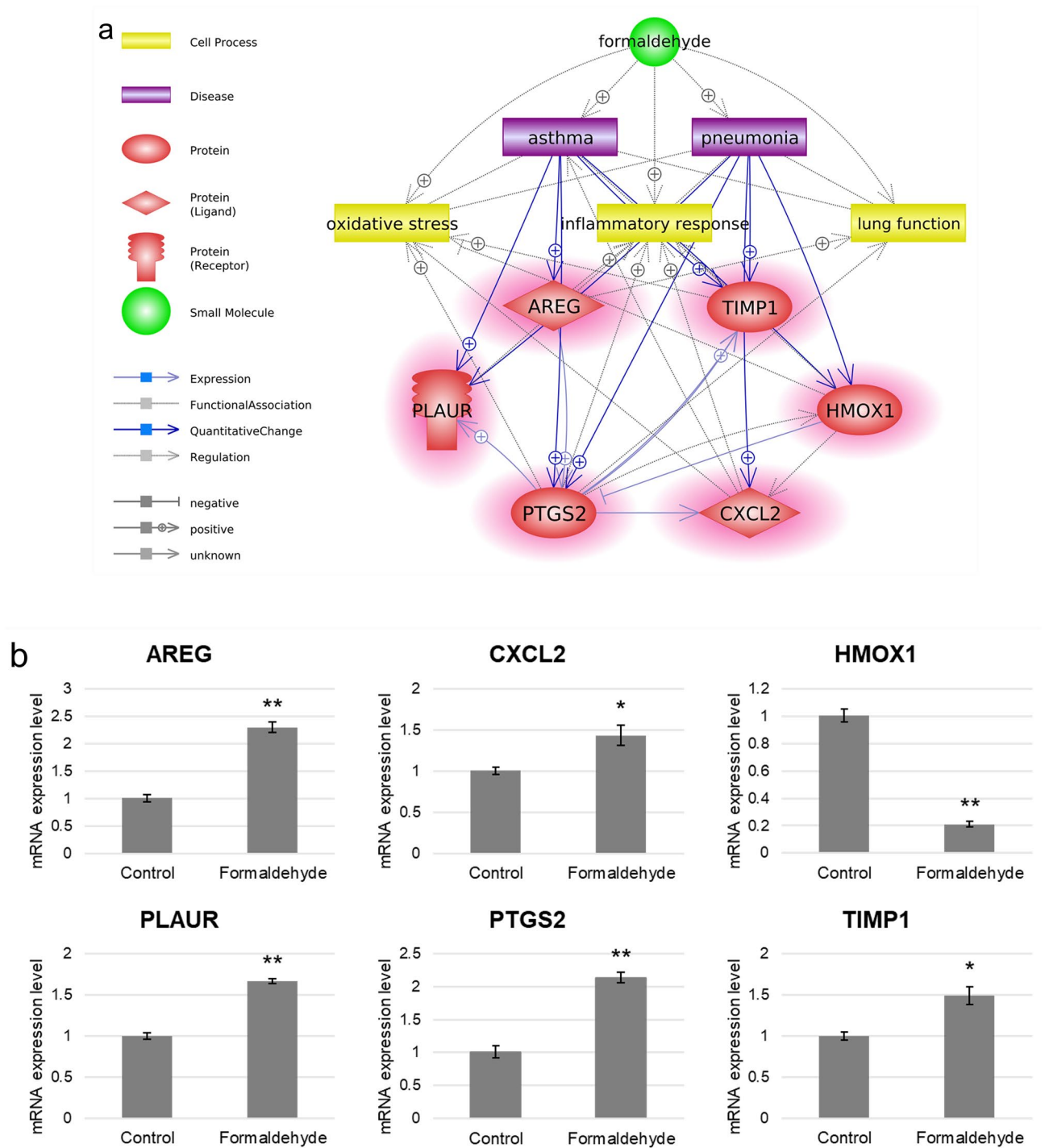
Figure 5. Validation of hub genes in SoluAirway. (**a**) Summarized core interactions of hub genes using Pathway Studio (version 12.3). Red highlights indicate upregulated DEGs. (**b**) Validation of the expression levels of hub genes using qRT-PCR. Error bars represent ± SEM. Single asterisk (*) and double asterisks (**) indicate p-value < 0.05 and < 0.01, respectively.

molecular changes by formaldehyde exposure, considering other types of omics data and other factors, such as detailed dose–response relationships, duration, and sensitivity of the exposed individuals[62–64].

Conclusion

In summary, through a series of bioinformatic analyses of DEGs derived from public in vivo studies, hub genes related to respiratory diseases associated with formaldehyde exposure were identified. Several studies have evaluated lung damages caused by formaldehyde inhalation in rats[65,66]. Formaldehyde toxicity is considered to affect the entire respiratory system, not just the upper respiratory system. Considering these points, the significance

of our toxicogenomic approach is that it can identify the respiratory effects of inhaled toxicants, serve as an early predictive alarm before serious lung diseases occur, and provide the possibility for identified genes to be used as biomarkers for clinical diagnosis and therapy. In addition, our results will contribute to improving the understanding of the toxic mechanism of formaldehyde, such as the construction of an adverse outcome pathway: suggestion of putative key events or adverse outcomes at diverse molecular levels, based on canonical pathways, cell processes, and diseases; inference of causal interaction; predictive assessment of carcinogenicity[67,68].

Data availability

The datasets used during the current study are available in the NCBI Gene Expression Omnibus (accession number: GSE7002, GSE23179, GSE27263, and GSE21477). The other data generated or analyzed during the current study are available from the corresponding author upon reasonable request.

References

1. International Agency for Research on Cancer. Chemical agents and related occupations. *IARC Monogr. Eval. Carcinog. Risks Hum.* **100**, 9–562 (2012).
2. National Toxicology Program. Final report on carcinogens background document for formaldehyde. *Rep. Carcinog. Backgr. Doc.* i-512 (2010).
3. Costa, S. *et al.* Genotoxic damage in pathology anatomy laboratory workers exposed to formaldehyde. *Toxicology* **252**, 40–48. https://doi.org/10.1016/j.tox.2008.07.056 (2008).
4. Swenberg, J. A. *et al.* Formaldehyde carcinogenicity research: 30 years and counting for mode of action, epidemiology, and cancer risk assessment. *Toxicol. Pathol.* **41**, 181–189. https://doi.org/10.1177/0192623312466459 (2013).
5. Nandan, A., Siddiqui, N. A., Singh, C. & Aeri, A. Occupational and environmental impacts of indoor air pollutant for different occupancy: A review. *Toxicol. Environ. Health Sci.* **13**, 303–322. https://doi.org/10.1007/s13530-021-00102-9 (2021).
6. Stewart, P. A., Cubit, D. & Blair, A. Formaldehyde levels in seven industries. *Appl. Ind. Hygiene* **2**, 231–236. https://doi.org/10.1080/08828032.1987.10388468 (1987).
7. Tang, X. *et al.* Formaldehyde in China: production, consumption, exposure levels, and health effects. *Environ. Int.* **35**, 1210–1224. https://doi.org/10.1016/j.envint.2009.06.002 (2009).
8. Rim, K.-T. Exposure of chemical mixtures at work and their application to the prevention of occupational disease. *Toxicol. Environ. Health Sci.* **13**, 91–99. https://doi.org/10.1007/s13530-021-00087-5 (2021).
9. Clough, E. & Barrett, T. The gene expression omnibus database. *Methods Mol. Biol.* **1418**, 93–110. https://doi.org/10.1007/978-1-4939-3578-9_5 (2016).
10. Barabasi, A. L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113. https://doi.org/10.1038/nrg1272 (2004).
11. Managbanag, J. R. *et al.* Shortest-path network analysis is a useful approach toward identifying genetic determinants of longevity. *PLoS ONE* **3**, e3802. https://doi.org/10.1371/journal.pone.0003802 (2008).
12. Andersen, M. E., Clewell, H. J. 3rd., Bermudez, E., Willson, G. A. & Thomas, R. S. Genomic signatures and dose-dependent transitions in nasal epithelial responses to inhaled formaldehyde in the rat. *Toxicol. Sci.* **105**, 368–383. https://doi.org/10.1093/toxsci/kfn097 (2008).
13. Andersen, M. E. *et al.* Formaldehyde: Integrating dosimetry, cytotoxicity, and genomics to understand dose-dependent transitions for an endogenous compound. *Toxicol. Sci.* **118**, 716–731. https://doi.org/10.1093/toxsci/kfq303 (2010).
14. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264. https://doi.org/10.1093/biostatistics/4.2.249 (2003).
15. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883. https://doi.org/10.1093/bioinformatics/bts034 (2012).
16. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47. https://doi.org/10.1093/nar/gkv007 (2015).
17. Kramer, A., Green, J., Pollard, J. Jr. & Tugendreich, S. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* **30**, 523–530. https://doi.org/10.1093/bioinformatics/btt703 (2014).
18. Nikitin, A., Egorov, S., Daraselia, N. & Mazo, I. Pathway studio–the analysis and navigation of molecular networks. *Bioinformatics* **19**, 2155–2157. https://doi.org/10.1093/bioinformatics/btg290 (2003).
19. Kang, D. S. *et al.* Formaldehyde exposure and leukemia risk: a comprehensive review and network-based toxicogenomic approach. *Genes Environ.* **43**, 13. https://doi.org/10.1186/s41021-021-00183-5 (2021).
20. Assenov, Y., Ramirez, F., Schelhorn, S. E., Lengauer, T. & Albrecht, M. Computing topological parameters of biological networks. *Bioinformatics* **24**, 282–284. https://doi.org/10.1093/bioinformatics/btm554 (2008).
21. Lee, N. *et al.* Local toxicity of biocides after direct and aerosol exposure on the human skin epidermis and airway tissue models. *Toxics* **9**. https://doi.org/10.3390/toxics9020029 (2021).
22. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408. https://doi.org/10.1006/meth.2001.1262 (2001).
23. Kerns, W. D., Pavkov, K. L., Donofrio, D. J., Gralla, E. J. & Swenberg, J. A. Carcinogenicity of formaldehyde in rats and mice after long-term inhalation exposure. *Cancer Res.* **43**, 4382–4392 (1983).
24. World Health Organization. *WHO Guidelines for Indoor Air Quality: Selected Pollutants*. WHO Guidelines Approved by the Guidelines Review Committee. p. 103–56. ISBN 9789289002134 (2010).
25. Schachter, E. N., Witek, T. J. Jr., Tosun, T., Leaderer, B. P. & Beck, G. J. A study of respiratory effects from exposure to 2 ppm formaldehyde in healthy subjects. *Arch. Environ. Health* **41**, 229–239. https://doi.org/10.1080/00039896.1986.9938338 (1986).
26. Malaka, T. & Kodama, A. M. Respiratory health of plywood workers occupationally exposed to formaldehyde. *Arch. Environ. Health* **45**, 288–294. https://doi.org/10.1080/00039896.1990.10118748 (1990).
27. Lino dos Santos Franco, A. *et al.* Pulmonary neutrophil recruitment and bronchial reactivity in formaldehyde-exposed rats are modulated by mast cells and differentially by neuropeptides and nitric oxide. *Toxicol. Appl. Pharmacol.* **214**, 35–42. https://doi.org/10.1016/j.taap.2005.11.014 (2006).
28. Uthiravelu, P., Saravanan, A., Kumar, C. K. & Vaithiyanandane, V. Pulmonary function test in formalin exposed and nonexposed subjects: A comparative study. *J. Pharm. Bioallied Sci.* **7**, S35-39. https://doi.org/10.4103/0975-7406.155787 (2015).
29. Yu, L. *et al.* Association between indoor formaldehyde exposure and asthma: A systematic review and meta-analysis of observational studies. *Indoor Air* **30**, 682–690. https://doi.org/10.1111/ina.12657 (2020).
30. Berasain, C. & Avila, M. A. Amphiregulin. *Semin. Cell Dev. Biol.* **28**, 31–41. https://doi.org/10.1016/j.semcdb.2014.01.005 (2014).

31. de Boer, W. I. *et al.* Expression of epidermal growth factors and their receptors in the bronchial epithelium of subjects with chronic obstructive pulmonary disease. *Am. J. Clin. Pathol.* **125**, 184–192. https://doi.org/10.1309/W1AX-KGT7-UA37-X257 (2006).
32. Ogata-Suetsugu, S. *et al.* Amphiregulin suppresses epithelial cell apoptosis in lipopolysaccharide-induced lung injury in mice. *Biochem. Biophys. Res. Commun.* **484**, 422–428. https://doi.org/10.1016/j.bbrc.2017.01.142 (2017).
33. Busser, B., Sancey, L., Brambilla, E., Coll, J. L. & Hurbin, A. The multiple roles of amphiregulin in human cancer. *Biochim. Biophys. Acta* **119–131**, 2011. https://doi.org/10.1016/j.bbcan.2011.05.003 (1816).
34. Loboda, A., Damulewicz, M., Pyza, E., Jozkowicz, A. & Dulak, J. Role of Nrf2/HO-1 system in development, oxidative stress response and diseases: An evolutionarily conserved mechanism. *Cell. Mol. Life Sci.* **73**, 3221–3247. https://doi.org/10.1007/s00018-016-2223-0 (2016).
35. Li, J. *et al.* Ferroptosis: past, present and future. *Cell Death Dis.* **11**, 88. https://doi.org/10.1038/s41419-020-2298-2 (2020).
36. Horvath, I. *et al.* Raised levels of exhaled carbon monoxide are associated with an increased expression of heme oxygenase-1 in airway macrophages in asthma: A new marker of oxidative stress. *Thorax* **53**, 668–672. https://doi.org/10.1136/thx.53.8.668 (1998).
37. Harju, T., Soini, Y., Paakko, R. & Kinnula, V. L. Up-regulation of heme oxygenase-I in alveolar macrophages of newly diagnosed asthmatics. *Respir. Med.* **96**, 418–423. https://doi.org/10.1053/rmed.2001.1283 (2002).
38. Mumby, S. *et al.* Lung heme oxygenase-1 is elevated in acute respiratory distress syndrome. *Crit. Care Med.* **32**, 1130–1135. https://doi.org/10.1097/01.ccm.0000124869.86399.f2 (2004).
39. Fredenburgh, L. E., Perrella, M. A. & Mitsialis, S. A. The role of heme oxygenase-1 in pulmonary disease. *Am. J. Respir. Cell Mol. Biol.* **36**, 158–165. https://doi.org/10.1165/rcmb.2006-0331TR (2007).
40. BARALIć, K. *et al.* Potential genomic biomarkers of obesity and its comorbidities for phthalates and bisphenol A mixture: In silico toxicogenomic approach. *Biocell* **46**, 519–533. https://doi.org/10.32604/biocell.2022.018271 (2022).
41. Williams, C. S., Mann, M. & DuBois, R. N. The role of cyclooxygenases in inflammation, cancer, and development. *Oncogene* **18**, 7908–7916. https://doi.org/10.1038/sj.onc.1203286 (1999).
42. Profita, M. *et al.* Increased prostaglandin E2 concentrations and cyclooxygenase-2 expression in asthmatic subjects with sputum eosinophilia. *J. Allergy Clin. Immunol.* **112**, 709–716. https://doi.org/10.1016/s0091-6749(03)01889-x (2003).
43. Peebles, R. S. Jr. Prostaglandins in asthma and allergic diseases. *Pharmacol. Ther.* **193**, 1–19. https://doi.org/10.1016/j.pharmthera.2018.08.001 (2019).
44. Lee, S.-H. *et al.* Anti-inflammatory effect of Rosa laevigata extract on in vitro and in vivo model of allergic asthma via the suppression of IgE and related cytokines. *Mol. Cell. Toxicol.* **16**, 119–127. https://doi.org/10.1007/s13273-019-00063-8 (2020).
45. Zhang, W. Dapk1 promoted inflammation of infantile pneumonia by p38MAPK/NF-κB signaling pathway. *Mol. Cell. Toxicol.* **17**, 297–304. https://doi.org/10.1007/s13273-021-00136-7 (2021).
46. Portelli, M. A. *et al.* Genome-wide protein QTL mapping identifies human plasma kallikrein as a post-translational regulator of serum uPAR levels. *FASEB J.* **28**, 923–934. https://doi.org/10.1096/fj.13-240879 (2014).
47. Wrotek, A. & Jackowska, T. The role of the soluble urokinase plasminogen activator (suPAR) in children with pneumonia. *Respir. Physiol. Neurobiol.* **209**, 120–123. https://doi.org/10.1016/j.resp.2014.12.018 (2015).
48. Stewart, C. E., Nijmeh, H. S., Brightling, C. E. & Sayers, I. uPAR regulates bronchial epithelial repair in vitro and is elevated in asthmatic epithelium. *Thorax* **67**, 477–487. https://doi.org/10.1136/thoraxjnl-2011-200508 (2012).
49. Chaudhuri, R. *et al.* Low sputum MMP-9/TIMP ratio is associated with airway narrowing in smokers with asthma. *Eur. Respir. J.* **44**, 895–904. https://doi.org/10.1183/09031936.00047014 (2014).
50. Cataldo, D. D. *et al.* Matrix metalloproteinases and tissue inhibitors of matrix metalloproteinases mRNA transcripts in the bronchial secretions of asthmatics. *Lab. Invest.* **84**, 418–424. https://doi.org/10.1038/labinvest.3700063 (2004).
51. Al-Alwan, L. A. *et al.* Differential roles of CXCL2 and CXCL3 and their receptors in regulating normal and asthmatic airway smooth muscle cell migration. *J. Immunol.* **191**, 2731–2741. https://doi.org/10.4049/jimmunol.1203421 (2013).
52. Huang, S., Paulauskis, J. D., Godleski, J. J. & Kobzik, L. Expression of macrophage inflammatory protein-2 and KC mRNA in pulmonary inflammation. *Am. J. Pathol.* **141**, 981–988 (1992).
53. Miller, M. C. 3rd., Mohrenweiser, H. W. & Bell, D. A. Genetic variability in susceptibility and response to toxicants. *Toxicol. Lett.* **120**, 269–280. https://doi.org/10.1016/s0378-4274(01)00279-x (2001).
54. Phng, L. K. *et al.* Nrarp coordinates endothelial Notch and Wnt signaling to control vessel density in angiogenesis. *Dev. Cell* **16**, 70–82. https://doi.org/10.1016/j.devcel.2008.12.009 (2009).
55. Pinto, I. *et al.* NRARP displays either pro- or anti-tumoral roles in T-cell acute lymphoblastic leukemia depending on Notch and Wnt signaling. *Oncogene* **39**, 975–986. https://doi.org/10.1038/s41388-019-1042-9 (2020).
56. Imaoka, T. *et al.* Overexpression of NOTCH-regulated ankyrin repeat protein is associated with breast cancer cell proliferation. *Anticancer Res.* **34**, 2165–2171 (2014).
57. Korn, C. & Augustin, H. G. Mechanisms of vessel pruning and regression. *Dev. Cell* **34**, 5–17. https://doi.org/10.1016/j.devcel.2015.06.004 (2015).
58. Huh, D., Hamilton, G. A. & Ingber, D. E. From 3D cell culture to organs-on-chips. *Trends Cell Biol.* **21**, 745–754. https://doi.org/10.1016/j.tcb.2011.09.005 (2011).
59. Neilson, L. *et al.* Development of an in vitro cytotoxicity model for aerosol exposure using 3D reconstructed human airway tissue; application for assessment of e-cigarette aerosol. *Toxicol. In Vitro* **29**, 1952–1962. https://doi.org/10.1016/j.tiv.2015.05.018 (2015).
60. Živančević, K. *et al.* Elucidating the influence of environmentally relevant toxic metal mixture on molecular mechanisms involved in the development of neurodegenerative diseases: In silico toxicogenomic data-mining. *Environ. Res.* **194**, 110727. https://doi.org/10.1016/j.envres.2021.110727 (2021).
61. Harris, S. M. *et al.* Identification of environmental chemicals targeting miscarriage genes and pathways using the comparative toxicogenomics database. *Environ. Res.* **184**, 109259. https://doi.org/10.1016/j.envres.2020.109259 (2020).
62. Suvorov, A. *et al.* Unbiased approach for the identification of molecular mechanisms sensitive to chemical exposures. *Chemosphere* **262**, 128362. https://doi.org/10.1016/j.chemosphere.2020.128362 (2021).
63. Baralić, K. *et al.* Probiotic reduced the impact of phthalates and bisphenol A mixture on type 2 diabetes mellitus development: Merging bioinformatics with in vivo analysis. *Food Chem. Toxicol.* **154**, 112325. https://doi.org/10.1016/j.fct.2021.112325 (2021).
64. Bozic, D. *et al.* Predicting sulforaphane-induced adverse effects in colon cancer patients via in silico investigation. *Biomed. Pharmacother.* **146**, 112598. https://doi.org/10.1016/j.biopha.2021.112598 (2022).
65. Murta, G. L. *et al.* Oxidative effects on lung inflammatory response in rats exposed to different concentrations of formaldehyde. *Environ. Pollut.* **211**, 206–213. https://doi.org/10.1016/j.envpol.2015.12.054 (2016).
66. Liu, Q. P. *et al.* Formaldehyde inhalation triggers autophagy in rat lung tissues. *Toxicol. Ind. Health* **34**, 834–841. https://doi.org/10.1177/0748233718796347 (2018).
67. Kim, H. S. *et al.* Suggestions for applications of toxicogenomic approaches in the adverse outcome pathway of 2,4-dinitrotoluene. *Toxicol. Environ. Health Sci.* **12**, 109–118. https://doi.org/10.1007/s13530-020-00054-6 (2020).
68. Rim, K.-T. Application of the adverse outcome pathway framework to predict the toxicity of chemicals in the semiconductor manufacturing industry. *Mol. Cell. Toxicol.* **17**, 325–345. https://doi.org/10.1007/s13273-021-00139-4 (2021).

Author contributions

Y.R.S. and D.S.K. designed the study. D.S.K., N.L., D.Y.S., Y.J.J., and S.H.L. performed the experiments. D.S.K analyzed the data and wrote the manuscript. K.-M.L., Y.S.A., C.M.L., and Y.R.S. reviewed the manuscript.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-09673-0.

Correspondence and requests for materials should be addressed to C.M.L. or Y.R.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.